

Springer Series in Design and Innovation 56

Sangeun Jin · Jeong Ho Kim ·
Yong-Ku Kong · Jaehyun Park ·
Myung Hwan Yun *Editors*

Proceedings of the 22nd Congress of the International Ergonomics Association, Volume 4

Better Life Ergonomics for Future
Humans (IEA 2024)

Editor-in-Chief

Francesca Tosi, *University of Florence, Florence, Italy*

Series Editors

Claudio Germak, *Politecnico di Torino, Turin, Italy*

Francesco Zurlo , *Politecnico di Milano, Milan, Italy*

Zhi Jinyi, *Southwest Jiaotong University, Chengdu, China*

Marilaine Pozzatti Amadori, *Universidade Federal de Santa Maria, Santa Maria, Brazil*

Maurizio Caon , *University of Applied Sciences and Arts, Fribourg, Switzerland*

Springer Series in Design and Innovation (SSDI) publishes books on innovation and the latest developments in the fields of Product Design, Interior Design and Communication Design, with particular emphasis on technological and formal innovation, and on the application of digital technologies and new materials. The series explores all aspects of design, e.g. Human-Centered Design/User Experience, Service Design, and Design Thinking, which provide transversal and innovative approaches oriented on the involvement of people throughout the design development process. In addition, it covers emerging areas of research that may represent essential opportunities for economic and social development.

In fields ranging from the humanities to engineering and architecture, design is increasingly being recognized as a key means of bringing ideas to the market by transforming them into user-friendly and appealing products or services. Moreover, it provides a variety of methodologies, tools and techniques that can be used at different stages of the innovation process to enhance the value of new products and services.

The series' scope includes monographs, professional books, advanced textbooks, selected contributions from specialized conferences and workshops, and outstanding Ph.D. theses.

The volumes of the series are single-blind peer-reviewed.

Keywords: Product and System Innovation; Product design; Interior design; Communication Design; Human-Centered Design/User Experience; Service Design; Design Thinking; Digital Innovation; Innovation of Materials.

How to submit proposals

Proposals must include: title, keywords, presentation (max 10,000 characters), table of contents, chapter abstracts, editors'/authors' CV.

In case of proceedings, chairmen/editors are requested to submit the link to conference website (incl. relevant information such as committee members, topics, key dates, keynote speakers, information about the reviewing process, etc.), and approx. number of papers.

Proposals must be sent to: series editor Prof. Francesca Tosi (francesca.tosi@unifi.it) and/or publishing editor Mr. Pierpaolo Riva (pierpaolo.riva@springer.com).

Sangeun Jin · Jeong Ho Kim · Yong-Ku Kong ·
Jaehyun Park · Myung Hwan Yun
Editors

Proceedings of the 22nd Congress of the International Ergonomics Association, Volume 4

Better Life Ergonomics for Future Humans
(IEA 2024)

Editors

Sangeun Jin
Industrial Engineering
Pusan National University
Busan, Korea (Republic of)

Jeong Ho Kim
Oregon State University
Corvallis, OR, USA

Yong-Ku Kong
Industrial Engineering
Sungkyunkwan University
Jangan-Gu, Suwon-Si, Gyeonggi-do, Korea
(Republic of)

Jaehyun Park
Department of Industrial and Management
Engineering
Incheon National University
Incheon, Korea (Republic of)

Myung Hwan Yun
Department of Industrial Engineering
Seoul National University
Gwanak-gu, Seoul-t'ukpyolsi, Korea
(Republic of)

ISSN 2661-8184

ISSN 2661-8192 (electronic)

Springer Series in Design and Innovation

ISBN 978-981-95-0288-2

ISBN 978-981-95-0289-9 (eBook)

<https://doi.org/10.1007/978-981-95-0289-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Singapore Pte Ltd. 2025

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

If disposing of this product, please recycle the paper.

Preface

This proceedings book is a publication of selected excellent papers presented at the World Congress on Ergonomics, the International Ergonomics Association 2024 (IEA 2024), held in Jeju Island, Republic of Korea, in August 2024. Representatives from over 50 countries, totaling approximately 1,600 ergonomics professionals, attended the event to present their outstanding research findings in the field of ergonomics and share their insights. About 1100 abstracts and full papers were presented, and all were subjected to a stringent peer review process that ensured both anonymity and double-blinding. Only submissions meeting the established standards were selected for inclusion in the conference proceedings.

The proceedings aim to encompass a broad spectrum of theories and applications within human factors and ergonomics (HFE). This multidisciplinary field is dedicated to understanding the interactions between humans and various systems, products, environments, and technologies. The volume covers a diverse range of scientific areas, including engineering, design, human biology, psychology, and so forth. It explores the application of human factors and ergonomics across various domains such as healthcare, medicine, transportation, aviation, automotive, construction, mining, manufacturing, agricultural engineering, education, and sustainable development.

The themes of the IEA 2024 conference and the structure of the papers included in each volume of the proceedings are as follows:

IEA 2024 Conference Theme: **“Better Life Ergonomics for Future Humans”**

Volume I: Affective Design (I), Aging (I), Agriculture (I), AI (I), Anthropometry (I), ATWAD (I), Aviation (I), Bio-signal (I), Building and Construction (I).

Volume II: Building and Construction (I), Design (I), DHM (I), Gender (I), Healthcare (I), HMI (I), Macro-ergonomics (I), Manufacturing (I).

Volume III: Manufacturing (I), Musculoskeletal (I), Product (I), Resilience (I), Robotics (I), Safety and Health (I), Slips (I), Sustainable Development (I), Training Education (I).

Volume IV: Training Education (I), Transport (I), UI/UX (I), Visual (I), VR/AR Metaverse (I), Others (I).

Volume V: Others (I), Affective Design (II), Aging (II), Agriculture (II), AI (II), Anthropometry (II), ATWAD (II), Bio-signal (II), Design (II), DHM (II), Healthcare (II), HMI (II), Macro-ergonomics (II), Manufacturing (II), Musculoskeletal (II), Product (II), Resilience (II), Safety and Health (II).

Volume VI: Safety and Health (II), Slips (II), Sustainable Development (II), Training Education (II), Transport (II), UI/UX (II), Visual (II), VR/AR Metaverse (II), Others (II), Affective Design (III), AI (III), Anthropometry (III), ATWAD (III), Bio-signal (III), Building and Construction (III), Design (III), DHM (III), Healthcare (III), Macro-ergonomics (III), Manufacturing (III), Musculoskeletal (III), Product (III), Robotics (III), Safety and Health (III), Space (III), Sustainable Development (III), Training Education (III), Transport (III), UI/UX (III), Visual (III), Others (III).

Organization

Editor-in-Chief of the Proceeding Book (Volumes I–V)

Name & Title	Address	Email & URL
Yong-Ku Kong Professor, Ph.D., CPE.	2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, 16419, Republic of Korea Dept. of Industrial Engineering, Sungkyunkwan University	ykong@skku.edu https://sites.google.com/site/skkuergospace/

Editors of the Proceeding Book (Volumes I–V)

Name & Title	Address	Email & URL
Myung Hwan Yun Professor, Ph.D.	Building #39, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea. Dept. of Industrial Engineering, Seoul National University	mhy@snu.ac.kr https://his.snu.ac.kr/
Jaehyun Park Professor, Ph.D.	120 Neungdon-ro, Gwangjin-gu, Seoul 05029, Republic of Korea Dept. of Industrial Engineering, Kunkuk University	jaehpark@inu.ac.kr https://sites.google.com/site/uvlabinu/
Sangeun Jin Associate Professor, Ph.D.	2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241, Republic of Korea Dept. of Industrial Engineering, Pusan National University	sangeunjin@pusan.ac.kr https://hse.pusan.ac.kr/hse/index.do
Jeong Ho Kim Associate Professor, PhD.	212 Adriance Lab Rd., College Station, TX 77843-1266, USA. Environmental and Occupational Health, Texas A&M University	Jay.kim@oregonstate.edu https://health.oregonstate.edu/labs/oeb

Contents

Training Education (I)

Micro-credentials Framework Design for Ready-to-Use Industrial Engineering Internship Students and Fresh Graduates	3
<i>Twin Yoshua R. Destyanto and Chandra Dewi Kurnianingtyas</i>	
Students' Mental Health and Experiences in the Blended Learning Setup: A Semester-Long Case Study	10
<i>Lorelie C. Grepo</i>	
Research on Interactive Guidance Design in Metaverse Education and Training Systems	17
<i>Shan Hu, Bingxin Zhang, Jitong Zhang, Weitao Jiang, and Lingda Rong</i>	
Identification of Technical and Ergonomic Requirements for Full-Scale Simulation with Extracorporeal Membrane Oxygenation	33
<i>Steffen Kuklinski, Simon König, Maximilian Kehmann, and Claus Backhaus</i>	
How Well a Simulation-Training Program Prevents WMSDs Among Nursing Staff?	39
<i>Bouhafs Mebarki, Brahim Zaoui, Mohamed Mokdad, and Imed Eddine Mebarki</i>	
An Anthropotechnological Analysis of Training Innovations in PROSUB: The Role of Computer-Based Training (CBT) in Achieving Operational Readiness for Brazil's Riachuelo-Class Submarine Crews	45
<i>Henrique Ribeiro Menezes, Júlio César Bispo Neves, and Mario Cesar Rodriguez Vidal</i>	
Development of Integrated Ergonomics Intervention for Employees with Signs and Symptoms of Musculoskeletal Disorders: Health Belief Model and ADDIE as Valuable Tools	51
<i>Nor Haslina Mohd, Anita Abdul Rahman, Shamsul Tamrin, Hezery Harun, and Ng Yee Guan</i>	
Methods of Assessment of Students in Active Learning Strategies in Occupational Health and Safety Teaching in Engineering	57
<i>Italo Rodeghiero Neto, Thiago Arndt Schimit, Lucas Gomes Miranda Bispo, and Fernando Gonçalves Amaral</i>	

The Use of Prototyping as an Experience for Teaching Ergonomics and Universal Design	64
<i>Sabine De Paris, Mariana Maciel Pasa, Angela Weber Righi, and Vanessa Goulart Dornéles</i>	
Mental Workload on Engineering Students: A Comparison Between Remote and Presential Teaching Using NASA-TLX	70
<i>João Victor Razzera, Eduardo Duarte Pena, and Angela Weber Righi</i>	
Challenges in Mobilizing Workers' Experience and Development in a Technological Transition Project	76
<i>Marta Santos, Cláudia Pereira, Maria Antónia Cadilhe, and Liliana Cunha</i>	
A Technological Implementation to Improve Maritime Firefighting Training or an Opportunity to (Re)Think Training Contents and Project Management	82
<i>Annaëlle Soulet, Lauriane Roger, and Catherine Delgoulet</i>	
Experiences in Teaching Ergonomics as Part of an Undergraduate Course on Occupational Safety and Health: A Case Study from the National School of Public Health in Medellín, Colombia	88
<i>Yaniel Torres</i>	
Enhancing Workers' Safety in Manufacturing: A Human-Centered Approach in Interaction Design of Digital Training Tools	96
<i>Lucia Vigoroso, Günter Alce, Luca Boccadutri, and Federica Caffaro</i>	
Transport (I)	
The Effect of Augmented Reality Head-up Display Interface Layout on Driving Behavior: A Systematic Literature Review	105
<i>Jingjing Cao, Yanling Zuo, Peggy Wang, and Jia Zhou</i>	
How to Make Autonomous Driving More Acceptable and Safer: Using an Augmented Highlighting Display	113
<i>Minwoo Cho and Dongwook Hwang</i>	
Vision and Brain Dynamics: Investigate Controllers' Performance during Emergency Situations in the Next Era of Air Traffic Control	119
<i>Donglin Li, Fan Li, Yanjie Zhang, Shanshan Feng, and Su Han</i>	
The Acceptability of Automated Vehicles by Other Road Users: Effects According to the Type of Vehicles	130
<i>Guelle Kevin, Erdinger Nicolas, and Metayer Natacha</i>	

The Influence of Advanced Driver Assistance Systems on Fatigue Development Among Long-Haul Truck Drivers	136
<i>Chunxi Huang, Zuyuan Wang, and Dengbo He</i>	
Methods for Evaluating the Comfort of Vehicle Seating Systems - A Literature Review	143
<i>Weisheng Jiang, Zhongqi Liu, and Qianxiang Zhou</i>	
Human Factors and Ergonomic Issues in Autonomous Shipping	149
<i>Hongtae Kim</i>	
Creating and Assessing Automotive Telltales for Preventing Drunk Driving	155
<i>Daehui Kim and Jaehyun Park</i>	
A VR Study: How Should an Automated Vehicle Communicate with Multiple Pedestrians?	161
<i>Yuchen Liu, Niklas Grabbe, and Klaus Bengler</i>	
Evolved Comfort in the Era of Autonomous Driving: Emic/Etic Approach to Reformulate It	167
<i>Saeedeh Mosaferchi, Alireza Morteza pour, and Alessandro Naddeo</i>	
Differences in Fatigue Levels and Driving Performance Based on Circadian Rhythms and Scene Transitions	173
<i>Po Chun Nien, Pei Hsuan Wu, Ting Hsuan Wang, and Yung Ching Liu</i>	
Exploring Driving Comfort and Vehicle Usability in Designing E-Cargo Bikes for Last Mile Delivery as Sustainable Urban Mobility Solutions	180
<i>Alessandra Rinaldi and Jonathan Lagrimino</i>	
Competencies and Technological Support for Remote Operator of Automated Vehicles: A Literature Review and Perspectives	187
<i>Sarah Schwindt-Drews, Jan-Philipp Frenzel, Ole Hans, Bettina Abendroth, and Jürgen Adamy</i>	
An Ergonomics Study on the Risks and Difficulties Faced by Drivers of Oil Tank Trucks in the Sahara Desert	193
<i>Mourad Semmani, Bouhafis Mebarki, and Mohamed Mokdad</i>	
Identifying Effective Visual Searching for the Recognition of Extraordinary Events during the Acceleration Process of Railway Driving	199
<i>Daisuke Suzuki, Takaharu Koike, and Ryo Kawai</i>	

Emotional Eyes for Automated Vehicles: Investigating Design Dimensions
and Pedestrian Acceptance 208
Nina Theobald, Felix Friedrich, Philip Joisten, and Bettina Abendroth

Classifying Rapid Deceleration Events for Older Drivers 215
Misako Yamagishi

Which Eye-Tracking Measure is Better for Examining Pedestrian Crossing
Behavior? 222
Putra A. R. Yamin and Jaehyun Park

UI/UX (I)

Implementing Affective Considerations for Design Pattern Selection
for Website Homepages 231
*Carissa M. Araco, Maria Mikaela A. Arriola,
Reinabelle Angelique Y. Gonzales, Antonio Lorenzo L. Sahagun,
and Ronaldo V. Polancos*

Novel GUI for the Biology Sector Based on HCD Approach 237
Alessia Brischetto and Mattia Pistolesi

Enhancing Comprehension and Trust in Privacy Policies: A User-Centric
Design Approach 244
*Raymond Christian M. Coronel II, Jose Antonio T. Conde,
Caila Jean C. Yap, Angelo Gabriel A. Reyes, Ezekiel L. Bernardo,
and Jazmin Tangsoc*

Integrating Cognitive Ergonomics and UI Enhancements for Efficient
Appointments and Elevated Patient Satisfaction in Metro Manila Hospitals 250
*Justine Bless Costales, Aaron Raphael Ricardo, Sheila Mae Carungay,
and Francispito Quevedo*

Explanatory Modelling of Factors Affecting the Adoption of Self-service
Technology 256
*Jaeyeon Park, Giannina Anais Rodas, Josh Adrian Dyaico,
Cedric Nathaniel Yu, Ezekiel L. Bernardo, and Jazmin C. Tangsoc*

Exploring the Impact of Virtual Instructor’s Attributes on Cognitive Load
in Health Short Videos 262
Yichen Guo, Mengli Yu, and Zhenkun Zhou

A Research Pathway for Interaction Design of Home-Based Elderly Care
Service Robots Based on Scenarios 268
Canqun He, Pengjie Shi, and Xiaomeng Sun

Designing Paediatric Hospital Navigation: A Human-Centred Mobile App Solution	274
<i>Ester Iacono, Virginia Piombino, and Francesca Tosi</i>	
Time's Up!: Assessing the Impact of Timer and Warning Alerts in Online Learning Management System towards Student's Satisfaction and Perceived Usefulness	280
<i>Beatriz Anne T. De Jose, Danielle Alexandria A. Pasia, Ezekiel L. Bernardo, Jazmin C. Tangsoc, and Charmine Sheena Saflor</i>	
Mobile Layout Design for Effective Exploration of Food Labeling Information	286
<i>Yunah Kang and Youngjae Im</i>	
Anthropomorphic is Always Better? Effects of Financial Robo-Advisor Types on Investors' Online Finance Investment Behavior	292
<i>Xinyue Kang, Fanxuan Wang, and Pei-Luen Patrick Rau</i>	
Developing a User-Centric Stepwise Comment Moderation AI Service: Balancing Free Expression and Respectful Online Interaction	299
<i>Jonghan Kim, Suhwan Jo, Woosung Jung, Manseo Kim, and Sung Park</i>	
Effective Design of Gauge Cluster for Improving Information Recognition	306
<i>Sungju Kim, Seeung Oh, and Youngjae Im</i>	
Designing the User Interface for an Elderly-Friendly Taxi-Hailing Mobile App: A Preliminary Study	312
<i>Yura Kim, Dongha Kim, Saeyeon Kim, Yeonji Kim, Gyuhee Kim, Chaewon Park, and Woojoo Kim</i>	
Framework for Understanding User Experience of Foldable Smartphones	319
<i>Yankuan Liu, Lili Dong, and Pei-Luen Patrick Rau</i>	
Cognitive Analyses for Printing Tasks Using DTF Technology: A Mental Workload and Human Error Approach	325
<i>Aide Aracely Maldonado-Macías, Gabriela Pérez Potter, and Mónica Gabriela Gutiérrez-Hernández</i>	
Optimizing eGov Philippines Application for Users in Manila City: A Study on Enhancing Engagement, Accessibility, and User Experience through Usability Testing and Web Scraping	331
<i>Janina Elyse A. Reyes, Rapeh Jahaziel L. Galao, Joshua J. Ordinario, Vanessa L. Santillan, and Hannagene C. Vidal</i>	

Generating Design Guidelines for Alternative Medication Management
Tools from Medical Cannabis Patients 338
*Deepthi Rajagopal, Douglas Bruce, Elissa Foster, Olayeke Adelakun,
and Enid Montague*

Research on the Emotional Vocabulary Used in the User Interface
of Generative AI-Based Music Production System 344
Yeongju Seo and Kiechan Namkung

Exploring Usability and User Value of the Electric Powered Wheelchair
for the Elderly and Physically Disabled 350
Sujin Seong, Yubin Ahn, Pakeeza Abid, and Jaehyun Park

Proposing Heuristic Guidelines for Auditory Interfaces in Medical Device
Usability Evaluation 356
Junghoon Shin, Kyungwoo Kwon, and Kiechan Namkung

Research on a Water-Drawing Robot Using Appropriate Technology:
Focusing on Interaction Methods and User Experience 364
Yunjeong Song, Jieun Sa, Jaehoon Cha, Juseong Kim, and Kicheol Pak

Urban Mobility Platforms and Trends Towards a MaaS Scenario 372
*Mariana Menezes O. Sousa, Manuela Quaresma,
Carlos Eduardo Félix da Costa, Claudia Mont’Alvão, and João Bonelli*

Usability Evaluation of the Fuel Management System for a Flag Carrier
Airline 378
*Alma Maria Jennifer Gutierrez, Macario Cordel,
Merlin Teodosia Suarez, and Christine Diane Ramos*

Enhancing Hospitality Service Experiences: Exploring Context-Aware
Systems Based on Smart Accommodation Framework 384
Yuming Wang, Chaemoon Yoo, Kuan Zhang, and Younghwan Pan

Examining the Population Stereotypes of Everyday Objects With
Rotational Controls 390
Rachma Widaningtyas, Fitri Trapsilawati, and Ardiyanto Ardiyanto

Applied Research on Gamification-Based Intelligent Interaction Design
of Smart Traffic Digital Twin Model 396
*Tianlun Yang, Zuyao Wang, Georgios Kapogiannis, Byung-Gyoo Kang,
Hu Zhenyu, Wu Yanhui, Shengyuan Liu, and Yuchang Zhang*

Visual (I)

Visuo-Cognitive Cartography, A New Tool for the Analysis of Eye-Tracking Data	411
<i>Denis I. Chavez</i>	
Supervision of Automated Production Systems: The Effects of Task Demands, Time on Task, and Task Reward on Visual Attention Allocation in a Process Monitoring Task	418
<i>Sebastian Pütz, Alexander Mertens, and Verena Nitsch</i>	
Dynamic Signs: Statistical Extraction Of Ergonomic Requirements of Moving Signs from Interviews and Questionnaires for Various Users	424
<i>Reiko Sakata, Saki Matsui, Atsushi Shimada, Naoki Furuhashi, Nana Ito, Ken Sagawa, Hiroshi Watanabe, and Hiroyasu Ujike</i>	
Dynamic Signs: Effects of Position and Age on the Detection of Signs During a Dual Task	431
<i>Hiroshi Watanabe, Nana Ito, Ken Sagawa, Hiroyasu Ujike, Reiko Sakata, and Naoki Furuhashi</i>	

VR/AR Metaverse (I)

User-Centred Development of an AR Application for the Remote Support of a Heart-Lung Support System	439
<i>Romy Alm, Simon König, and Claus Backhaus</i>	
A Framework Proposal for Augmented Reality Considering Human Errors in Sequential Tasks	445
<i>Moacyr Machado Cardoso Júnior, Henrique Costa Marques, and Marcos Vaz de Lima</i>	
Immersive and Interactive Digital Twins to Train Engineering Students	450
<i>Sylvain Fleury</i>	
Immersion in the Shadows: Physiological Markers of Flow Experience in a Virtual Reality Survival Horror Game	454
<i>Sanjram Premjit Khanganba and Anantha Ubaradka</i>	
User Errors in Augmented Reality Assisted Preoperative Planning	461
<i>Simon König and Claus Backhaus</i>	

Optimizing Button Size to Enhance User Experience for Augmented Reality Interactions in Mobile Scenarios	467
<i>Hojeong Lee, Yeahyoung Jeon, Keisuke Kudo, Paul Ssemakula, Seonghyeok Park, and Shuping Xiong</i>	
Examining the Influence of Metacognitive Monitoring Feedback in an AR Learning Environment	473
<i>Sara Mostowfi, Jung Hyup Kim, Ching-Yun Yu, Siddarth Mohanty, Fang Wang, Danielle Oprean, and Kangwon Seo</i>	
Virtual Reality for the Human-Centred Design of Assistive Devices	480
<i>Ivan de Souza Rehder, Moacyr Cardoso Machado Junior, Edmar Thomaz da Silva, and Emilia Villani</i>	
Quantification of Vergence–Accommodation Conflict for Virtual Reality Learning Content	486
<i>Takashi Shibata and Shugo Kasuya</i>	
A Review: Identification of Menu Types, Interaction Tasks, and Performance Metrics in Virtual Environment Menu Interaction	493
<i>Benedikta A. H. Siboro and Chiuhsiang Joe Lin</i>	
Others (I)	
Earable-Based Posture Inference and Correction System for Deskwork Behavior Transformation	503
<i>Yilizati Abulaiti, Shoko Kimura, and Guillaume Lopez</i>	
Inclusive Dwelling for Inclusive Living Communities in the Current Eco-Socio-Technical Changing Scenario	510
<i>Erminia Attaianese</i>	
Author Index	517

Training Education (I)



Micro-credentials Framework Design for Ready-to-Use Industrial Engineering Internship Students and Fresh Graduates

Twin Yoshua R. Destyanto^(✉)  and Chandra Dewi Kurnianingtyas 

Industrial Engineering Department, Universitas Atma Jaya Yogyakarta, Yogyakarta 55281,
Indonesia

twin.destyanto@uajy.ac.id

Abstract. The gaps between skills trained to Industrial Engineering (IE) internship level students and fresh graduates in a university with needs in industry were found. Certain basic skills need to be trained by the users in their entry period as interns or fresh graduates. Therefore, this study aimed to design a framework of micro-credential training for IE students based on the needs of their prospective users. The framework was designed by following the usability engineering lifecycle involving ten stakeholders, including students/alumni users (human resources and other department staff and managers), students who pass internship programs, fresh graduates, and IE curriculum experts. This study results in a framework of micro-credentials for IE students with three main scopes of skills comprising IE body of knowledge skills, communication and critical thinking skills, and attitude-related skills. The findings may help both the IE department and the industry to maintain harmony between the outcome of learning and the needs of the industry. This study is a case study in an IE department in a university; in the future, the framework needs to be tested in IE departments across some universities to improve the current framework prototype.

Keywords: micro-credential · usability engineering · industrial engineering

1 Introduction

The rapid development of technology in this era helps industries to achieve higher productivity and improve efficiency. These advantages require professional workers who are skillful enough to solve complex problems industries face. Moreover, companies expect professionals to understand and be capable of solving those problems without ignoring society [1, 2].

On the other hand, the Industrial Engineering (IE) program in certain universities tries to empower their students to meet the requirements of enterprises nowadays [1]. The curriculum is designed to cover as much of all industrial and system engineering bodies of knowledge as possible. Therefore, the students and fresh graduates are expected to become skillful enough to work by having a holistic understanding of system concepts in the industry.

However, from the point of view of companies as graduates' users, it is found that there are gaps between the trained and the needed skills [3]. Consequently, recruiters should hold various training courses to adjust the student's or fresh graduates' abilities to the required skills. The additional training can be skilled-based, including engineering drawing, business process, and occupational-related training, or leadership and communication-related training comprising public-speaking and human-capital management-related skills.

Higher education institutions then include various micro-credential programs to minimize this gap. Universities try to empower students with essential practical skills that meet and align with current or future technology development in industries. The micro-credentials held by academic institutions and industries provide shorter hours of learning [1] that focus on specific skills or abilities and are typically provided online [4]. Once the students finish the credential, the training organizer usually digitally provides the certificate or badge [5, 6].

The various frameworks related to the design of micro-credentials programs have been developed to minimize the mentioned gap and are continuously being improved. For example, [7], as academicians and industry partners, designed micro-credentials for students in engineering programs that focus on practical entrepreneurship-related skills. Micro-credential frameworks were also developed to adjust the skills needed to deal with ecological and environmental issues [8]. Micro-credentials programs can also be developed based on the needs of the industry to have employees with good attitudes, e.g., integrity [4] and leadership abilities [9]. Using a learning analytics approach, [1] designed a micro-credential framework to minimize the university-industry gap. However, none were designed holistically by involving the current students, fresh graduates, users (e.g., industries), and human capital professionals as the stakeholders to improve the micro-credential framework designs iteratively. Most frameworks are developed based on past designs and theories but lack stakeholder involvement during the design steps. Therefore, this study aimed to design a micro-credential framework, especially for IE students, based on the needs of their prospective users and the experience of current students and fresh graduates. The frameworks should not only fit from the users' points of view but also students' motivation and interests.

2 Methods

The micro-credentials framework was designed iteratively following a usability engineering lifecycle [10] to make the framework become a user-centered design. It starts by defining the users and stakeholders and then understanding their needs and criteria by conducting in-depth interviews with the users. The competitive advantages were achieved by collecting the micro-credential design used at high-education levels. The next phase was setting the main goals to make the framework usable by making the bridge from the found gap. The iterative design process was initiated by brainstorming to produce the first framework and continuously improving it by involving users, stakeholders, and experts in the IE curriculum to produce the final framework.

2.1 Study Participants

This study took the case from the IE Department in Universitas Atma Jaya Yogyakarta (UAJY), Indonesia, accredited by the Indonesian Accreditation Board for Engineering Education (IABEE) and is the official representative of the Washington Accord. The “products” in this system (education institution) are the current IE Department students and the fresh graduates recruited by industries as interns and full-time workers. Considering this, we defined the primary users of the products as the stakeholders from industries that have experience recruiting and working with IE students and fresh graduates. Therefore, this study conducted an in-depth interview involving ten stakeholders. The primary users were represented by 3 human resources (HR) staff and managers from industries that ever-recruited UAJY’s students and fresh graduates. The other stakeholders were 3 IE students who had finished their internship program and 3 fresh graduates working in the manufacturing and service industries to dig into their experiences related to the gap between skills trained and the needed skills during their internship and work as fresh graduates. An IE curriculum expert also gave feedback on the second prototype. These in-depth interviews were recorded verbatim and aimed to collect the ideas and needs of the participants involved.

2.2 Interviews

The interviews were conducted with the 10 mentioned participants by using online calls via WhatsApp. Participants consented to this study and verbally agreed to be recorded (audio). They also permitted their verbatim to be analyzed.

The interviewer asked the interviewees to mention and explain their roles in the company(s) in which they were recruited for the user group. Further questions were designed based on specific themes comprising (1) how the skills trained by the IE department are compared to the required skills in the industry field; (2) additional training provided by the industry to both interns and fresh graduates; (3) expected additional skill-based training provided by IE departments from users, students, and fresh graduates point of views. The expert group was also asked to improve the second prototype to produce the final one. The questions were classified based on participant group as presented in Table 1.

Table 1. The questions asked of interviewees in this study

Participant group	Questions
Users (e.g., HR staff and managers, another department manager)	a. When recruiting IE students for the internship program, what skills do you expect they already have? b. When recruiting IE fresh graduates for entry-level positions, what skills do you expect they already have? c. After the students were recruited, what kind of training did you give to equip the interns from IE students, and what difficulties were faced? d. After the IE fresh graduates were recruited, what training did you give to equip them, and what difficulties were faced? e. Please explain your expectation-related skills; the IE department must train students before entering the internship program. f. Please explain your expectation-related skills; the IE department needs to train students before they graduate from the bachelor program.
Students who just finished their internship program	a. When doing your internship program, what job/project did you receive? b. What trained skills can be applied to do that job/project? c. What additional training was given by the company? d. What other skills do you think need to be equipped by IE departments before students enter the internship program, based on the difficulties you faced?
Recruited fresh graduates	a. What is your current job? b. What trained skills (from the IE department) can be applied to that job/project? c. What additional training was given by the company? d. What other skills do you think need to be equipped by IE departments before graduation, based on your difficulties?
IE curriculum expert	Please suggest improvements for the second stage of the framework prototype to make it more suitable for IE knowledge

2.3 Data Analysis

The transcripts of interviewees' answers were open-coded using NVivo [11]. After that, the codes were categorized based on the theme. The first version of the prototype of the micro-credential frameworks was produced from those categories through brainstorming and abstraction.

3 Results and Discussion

3.1 Industry-IE Department Gaps

Based on the data, it is found that the IE department stakeholders seem more focused on how they can equip the students with skills related to IE Body of Knowledge (BOK). The department has provided certification and micro-credentials programs related to that knowledge, comprising engineering drawing, business process management, data analysis certifications, etc. Those pieces of training are essential for students to be taught and trained. However, by digging into the information from industry stakeholders, the bulk of training is insufficient compared to the needed skills in the industry.

The industry environment does not only require engineering-related thinking and skills but also communication and attitude-related skills. The lack of communication abilities, comprised of reporting, presentations, global language skills, and interpersonal skills in intergenerational environments, is potent in restricting the students and fresh graduates from performing optimally. Moreover, the wrong attitude, i.e., impoliteness, lateness on the due date, and disobedience to the rules and leaders' directions, may affect performance and career promotion opportunities. These phenomena indicate a wide gap between industry and IE departments, both directly and indirectly. Therefore, minimizing the gap by remodeling the micro-credentials framework is crucial.

3.2 Usable Micro-credentials Framework

The iterative process in this study results in a micro-credentials framework for IE students with three main scopes. The first scope is IE BOK-based skill micro-credential certification comprising Enterprise Resource Planning tools operation ability (e.g., Odoo, SAP), computer-aided design software skills (e.g., SolidWorks, AutoCAD), statistical and data visualization skill (e.g., Microsoft Excel, Power BI, Minitab), and Human Resource Management basic knowledge certification. The second scope is communication and critical thinking skills comprising presentation skills (e.g., Microsoft PowerPoint), public speaking ability, English and other global language abilities (e.g., Mandarin, Korean, Japanese), critical-thinking training, and art-based designing (e.g., Adobe Illustrator, Corel Draw). The third main scope is attitude-related training, as used in [4], which comprises leadership, character building (e.g., introduction to DISC knowledge), basic manners, and politeness that overcome intergenerational issues. This "usable" micro-credential framework is described in Fig. 1.

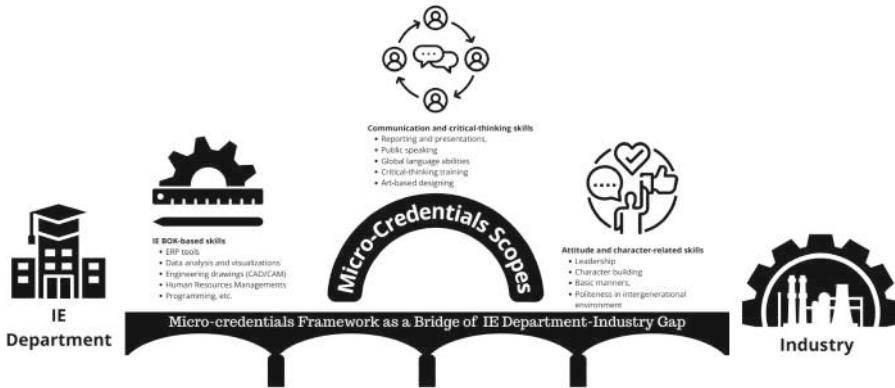


Fig. 1. Usable micro-credential framework

The framework of the micro-credentials for IE students must cover the three main scopes mentioned above. Moreover, using real cases from industries will support reducing the gap between the skills produced from micro-credential certification and the users' needs in industries. The findings may help both the IE department and the industry to maintain harmony between the outcome of learning and the needs of the industry.

4 Conclusion

To minimize the gap between the skills required by the IE internship students and fresh graduates with the skills required by the industry, the IE department needs to design a comprehensive skill-based training with at least cover three main scopes comprising IE BOK-based, communication, and critical thinking, and attitude-related skills training. This study is a case study in an IE department in a university; in the future, the framework needs to be tested in IE departments in some universities to improve the current framework prototype.

References

1. Messaoud, S.B., Ilahi, M., Cheniti-Belcadhi, L.: Towards a personalized micro-credentials approach based on learning analytics for reducing the gap university-industry. In: 2022 IEEE Global Engineering Education Conference (EDUCON), pp. 1448–1454. IEEE, Tunis, Tunisia (2022)
2. Leng, J., et al.: Industry 5.0: Prospect and retrospect. *J. Manuf. Syst.* **65**, 279–295 (2022)
3. Raj, R., Singh, A., Kumar, V., Verma, P.: Achieving professional qualifications using micro-credentials: A case of small packages and big challenges in higher education. *Int. J. Educ. Manag.* **38**(4), 916–947 (2024)
4. Ruddy, C., Ponte, F.: Preparing students for university studies and beyond: A micro-credential trial that delivers academic integrity awareness. *J. Aust. Libr. Inf. Assoc.* **68**, 56–67 (2019)
5. Ifenthaler, D., Bellin-Mularski, N., Mah, D.-K.: *Foundation of Digital Badges and Micro-Credentials*. Springer International Publishing, Switzerland (2016)

6. Lemoine, P.A., Richardson, M.D.: Micro-credentials, nano degrees, and digital badges: New credentials for global higher education. *Int. J. Technol. Educ. Mark. (IJTEM)*. **5**, 36–49 (2015)
7. Eager, B., Cook, E.: Micro-credentialing of entrepreneurship education in a practice-based undergraduate engineering context. *Entrep. Educ. Pedagogy*. **3**, 352–363 (2020)
8. Miller, K.K., de St, J., Jorre, T.: Digital micro-credentials in environmental science: An employer perspective on valued evidence of skills. *Teach. High. Educ.* **29**, 1058–1074 (2024)
9. Hunt, T., Carter, R., Zhang, L., Yang, S.: Micro-credentials: The potential of personalized professional development. *Dev. Learn. Organ.: Int. J.* **34**, 33–35 (2020)
10. Möller, S.: Usability engineering. In: *Quality Engineering: Quality of Communication Technology Systems*, pp. 55–72. Springer, Berlin, Heidelberg (2023)
11. Elo, S., Kyngäs, H.: The qualitative content analysis process. *J. Adv. Nurs.* **62**, 107–115 (2008)



Students' Mental Health and Experiences in the Blended Learning Setup: A Semester-Long Case Study

Lorelie C. Grepo^(✉)

University of the Philippines Diliman, Quezon City 1101, Philippines
lcgrepo@up.edu.ph

Abstract. Educational institutions implemented changes in the learning setup due to the COVID-19 pandemic. While some universities returned to in-person classes, some adopted hybrid setups. This study investigates the 3rd year Industrial Engineering student's perceived workload, stress, and fatigue throughout a semester in a blended learning mode, transitioning from a fully remote setup. Student experiences were correlated with their mental health and academic performance. Sixty-six participants completed the 20-week study during the 2nd semester of AY 2022–2023. Results show that students' workload, stress, fatigue, and mental health differed significantly over time. The outcomes were not statistically different between males and females. Depression and anxiety were related to students' perceived stress and fatigue, stress was related to the students' perceived workload and fatigue, and the perceived academic performance was related to the perceived workload, stress, and fatigue. Recommendations for further studies include expanding the study to a wider population, increasing the sample size, and examining the implications of having breaks during the semester.

Keywords: Mental health · Workload stress and fatigue · Blended learning

1 Introduction

College students experience stress and fatigue due to school work and activities, eventually affecting their well-being [1–3]. With the COVID-19 pandemic, studies have shown increased stress and poorer well-being of students [4] related to the changes in learning setup. In the Philippines, schools were forced to adopt a fully remote learning setup in 2020 during the lockdowns. It took about 2 years for universities to return to an in-person learning setup. From the experiences in the remote setup, schools now take advantage of technological advancements to improve learning. Some adopted a blended learning setup. Student experiences with blended learning, particularly at the University of the Philippines Diliman, have yet to be understood.

Understanding student workload is important as it is one of the major stressors for students [5, 6] and has been shown to correlate with their perceived stress [7] and fatigue, both physically and mentally [2, 8]. Studies also revealed a significant difference in the

reported perceived experiences of male and female students. In UP Diliman, the same was found with female students experiencing higher stress and fatigue levels in the remote setup [9]. High stress and fatigue levels were correlated with negative consequences, including poor well-being, unhealthy behaviors, poor performance, poor mental health, and poor general health [1, 6, 7, 10–12].

This study investigates student experiences in the blended learning setup for a whole semester, particularly the perceived workload, perceived stress, perceived fatigue, mental health, and perceived academic performance. A similar study was previously done on the students' experiences in the fully remote setup, describing how their workload, stress, fatigue, and academic performance varied throughout the semester [9]. Filipino students in one university indicated that stress significantly affected their performance in the remote setup [13]. Since the university transitioned to requiring face-to-face meetings, but not 100%, this study aims to investigate its impact on the students, considering varying modes for different classes students take.

2 Methods

A repeated-measure design elicited student experiences through weekly online surveys, complemented by qualitative data to describe student experiences and coping mechanisms. This paper reports on the results of the quantitative data.

Third-year Industrial Engineering students taking the Cognitive Ergonomics course during the Second Semester of the Academic Year 2022–2023 participated in this study. Students taking the course were informed about the study, and those who indicated interest in participating were asked to sign consent forms. Students who completed the 20-week study were given extra credits for the course.

The perceived workload was measured using the Subjective Workload Assessment Technique (SWAT) [14] and the NASA-TLX [15] and perceived stress with the Perceived Stress Scale [16], created to measure the stress experienced by general population. Perceived fatigue was measured using the 12-item Piper Fatigue Scale [17]. The 21-item Depression, Anxiety, and Stress Scale short version (DASS-21) [18] was used to measure students' mental health during the semester, including depression, anxiety, and stress, which are scored individually and then categorized into normal, mild, moderate, severe, and extremely severe levels. Finally, academic performance was measured using a single item rated from 1 (worst) to 10 (best).

All surveys were administered online through Qualtrics. A pre-study survey collected demographics and established baseline measures. Throughout the semester (Weeks 1 to 20), including the breaks (weeks 8 to 9), the survey link is sent to the participants every Saturday, with reminders for completion. The last survey included an evaluation of the entire semester in the blended learning setup.

Data were analyzed using Microsoft Excel and R version 4.3.2. Data were cleaned and checked for response validity and completeness. Descriptive statistics using frequency, means, and variances were calculated. Each variable was tested using a linear mixed model with random effects to check whether they differed between weeks in the semester and which weeks were significantly different. Sex was included in the model to check whether the outcomes were different between male and female students. Linear mixed

models with random effects were also used to evaluate relationships between variables. Each model was checked for assumptions of normality, independence, and homoscedasticity. Finally, the r-squared was calculated to quantify the variances accounted for by the linear mixed model.

3 Results and Discussion

Sixty-six (66) students completed the 20-week study, with 50% male and 50% female subjects. The participants' mean age is 21, with a standard deviation of 1.

Perceived Workload. Using a linear mixed model with random effect, results show that the perceived workload measured using SWAT differed significantly across weeks. Weeks 8 and 9 have negative estimates (-0.2 and -0.03), indicating reduced workload scores on these periods relative to the baseline at week 0. These weeks corresponded to the Lenten holidays and the reading break, respectively. Weeks 2 and 20, however, were not statistically different from the baseline. The workload level did not differ between male and female students. Individual students had varied baseline workloads with a random effects coefficient of 0.6969. Figure 1 shows the NASA-TLX scores for each subscale over the 20-week study period.

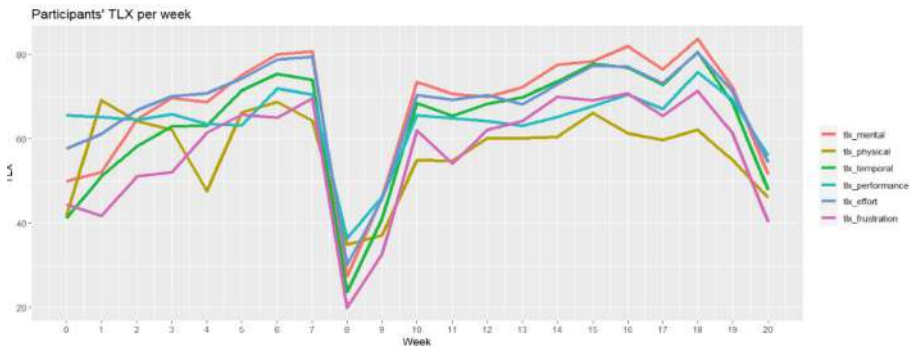


Fig. 1. Perceived workload measured using the NASA-TLX

Perceived Stress. The students reported stress symptoms they experienced weekly. On average, more than 50% of the participants reported fatigue, feeling overwhelmed, headache, anxiety, lack of motivation or focus, sleep problems, and restlessness. The students' perceived stress scores are shown in Fig. 2: thin black lines represent the individual responses, thicker blue and pink lines distinguish male and female students' average scores, respectively, and thicker black line represents the weekly average score, with trends similar to the workload measures. Most students experienced moderate stress, with weeks 8 and 9 reflecting lower levels. Based on the linear mixed model with random effects, there were statistical differences in stress across weeks, particularly weeks 1, 8, 9, and 20, compared to Week 0. Results indicated reduced stress reflected by the negative coefficients. Week 1 was the beginning of classes, understandably lighter than

the other weeks (-2.19); week 8 was the Lenten break (-5.59); week 9 was the reading break (-3.05); week 20 was the week after the finals (-2.76). There was no significant difference in the stress of male and female students.

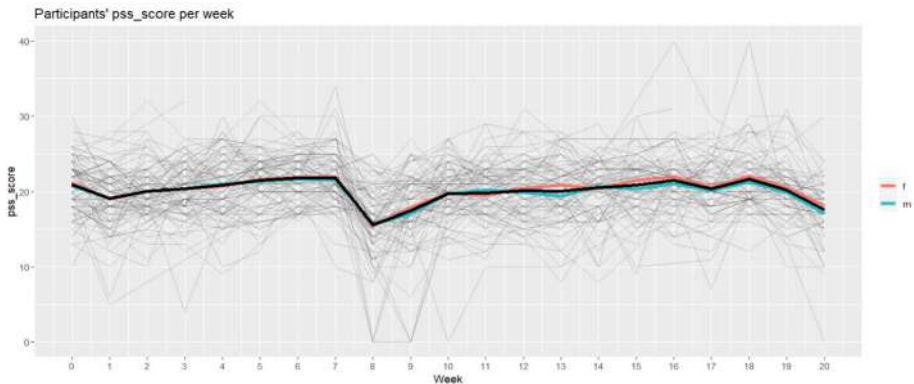


Fig. 2. Perceived stress scores

Perceived Fatigue. Most students experienced moderate to severe fatigue levels except in weeks 8, 9, and 20. Based on the t-values from the linear mixed model with random effects where the week is set as the fixed factor, participants as the random factor, and the PFS score as the outcome, results indicate a statistical difference between weeks relative to the baseline measure, except weeks 1, 2, and 20. Only weeks 8 and 9 had negative estimates (-2.16 and -0.85 , respectively), showing reduced reported fatigue scores relative to the baseline. All other weeks had higher fatigue scores than week 0. Sex was not significant in the model.

Well-being. The participants' well-being was measured using the short version of the Depression, Anxiety, and Stress Scale (DASS-21). The 3 subscales, depression, anxiety, and stress, are scored separately based on the different items in the questionnaire. The average scores per subscale over 20 weeks of the study are shown in Fig. 3. It is noticeable that stress was scored the highest among the three, while anxiety was scored the lowest. However, the general trend of having lower scores on weeks 8 and 9 is also apparent on all 3 measures.

Academic Performance. The students rated their perceived academic performance weekly using a scale from 1 (worst performance) to 10 (best performance). Based on a linear mixed model with academic performance as the outcome, weeks as a fixed effect, and participants as a random effect, all weeks were significant except week 20. The perceived academic performance was below the baseline throughout the semester, with negative estimates for all weeks. Weeks 5, 8, and 9 had the biggest score reduction, possibly due to reduced academic-related activities.

Relationships between variables. A linear mixed model with random effects, with stress as outcome and workload and fatigue as predictors, shows that workload ($t = 6.82$)

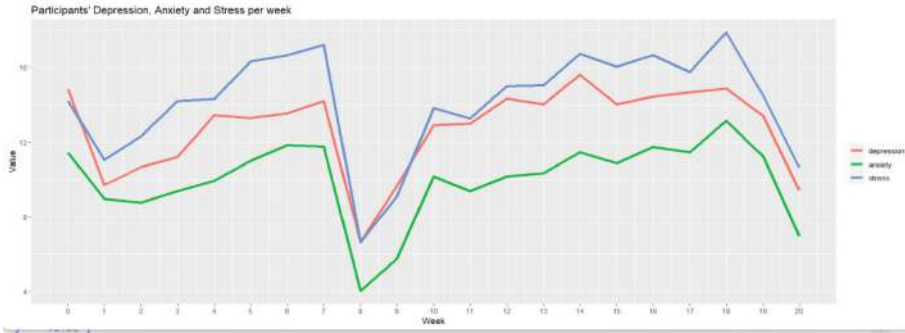


Fig. 3. DASS scores

and fatigue ($t = 7.80$) are related to stress. With fatigue as the outcome, workload ($t = 17.81$) and stress ($t = 7.94$) are both significant. Sex was not significant in both models. Each subscale was modeled for mental health as an outcome with fixed factors: workload, stress, and fatigue, while participants were treated as random factors. For depression, stress, and fatigue, intercepts were statistically significant ($p < .05$). Workload and sex were insignificant. For anxiety, only the stress and fatigue scores were significant ($p < .05$). The workload and fatigue were both significant factors affecting stress. For all 3 subscales, the reported levels did not differ between male and female students. Lastly, the relationship between workload, stress, fatigue, and academic performance was significant ($p < .05$).

4 Conclusions

Perceived workload, stress, fatigue, mental health, and academic performance varied across weeks in the semester. The lowest scores were reported on weeks 8 and 9 during the semester, falling on the Lenten and reading breaks. Aside from these two weeks, week 1 and week 20 were also significant in some outcomes, reflecting the lighter load at the beginning and end of the semester. This study shows the benefit of having breaks during the semester, which help students “cope” with their workload and academic requirements. Unlike other studies showing differences between male and female students, this study did not find statistical differences in the participants’ reported scores but highlighted individual differences. Finally, only perceived stress and perceived fatigue were significantly related to depression and anxiety. All factors were significant for stress. Results indicate a relationship between multiple outcomes measured over time, which can be further analyzed, particularly the dynamics between them. Further research can improve this study by including a more heterogeneous population and increasing the sample size. Examining the effects of breaks and learning setup more deeply can help formulate policies that enhance learning experiences and well-being.

Acknowledgments. The author would like to thank Dr. Aura C. Matias for the Renato Bernaldes Matias Professorial Chair in Industrial Engineering Award granted through the University of the

Philippines Diliman College of Engineering and UPERDFI. The author also acknowledges the University of the Philippines Diliman Office of the Vice Chancellor for Research and Development (OVCRD) for the research dissemination grant that made the presentation of the paper at the IEA Conference possible. Finally, the author thanks all the student participants who completed the weekly surveys and shared their experiences in the study.

References

1. Kizhakkeveetil, A., Vosko, A.M., Brash, M., Ph, D., Philips, M.: A: Perceived stress and fatigue among students in a doctor of chiropractic training program. *J. chiropr. educ.* **31**(1), 8–13 (2017)
2. Aldiabat, K., Matani, N., Le Navenec, C.: Mental health among undergraduate university students: A background paper for administrators, educators and healthcare providers. *Univ Jour Public Health.* **2**(8), 209–214 (2014)
3. Palmer, L.: The relationship between stress, fatigue, and cognitive functioning. *Coll. Stud. J.* **47**(2), 312–325 (2013)
4. Son, C., Hegde, S., Smith, A., Wang, X., Sasangohar, F.: Effects of COVID-19 on college students' mental health in the US: Interview survey study. *J. Med. Internet Res.* **22**(9), e21279 (2020)
5. Dy, M., Espiritu-Santo, K., Ferido, M., Sanchez, R.: Stressors and stress responses of Filipino college students (2019)
6. Britz, J., Pappas, E.: Sources and outlets of stress among university students: Correlations between stress and unhealthy habits. *URJHS.* **9**, 1 (2010)
7. Yang, C., Chen, A., Chen, Y.: College students' stress and health in the COVID-19 pandemic: the role of academic workload, separation from school, and fears of contagion. *PLoS One.* **16**(2), e0246676 (2021)
8. Sy, M., Irinea, M., Malabuyoc, F., Sobrevilla, D., Gumasing, M.: Effects of academic workload to the mental fatigue of students during online class. In: 12th Annual International Conference on Industrial Engineering and Operations Management. IEOM Society International. IEOM Conf Proc, Istanbul, Turkey (2022)
9. Grepo, L.: Students' perceived workload, stress, fatigue and performance during the remote learning setup: A semester-long study. In: Nazir, S. (ed.) *Training, Education, and Learning Sciences.* AHFE Open Access, vol. 109. AHFE Int, USA (2023)
10. Koch, A.: How academic and extracurricular workload and stress impacts the mental & physical health of college students. *Honors Theses*, 165. (2018)
11. Canillo, G., et al.: PM is the key: perceived stress and mental health as the key indicator of wellbeing of selected college online students in the philippines during COVID-19 pandemic. *Psychol. Educ.* (2022)
12. Lee, S., Wuertz, C., Rogers, R., Chen, Y.: Stress and sleep disturbances in female college students. *Am J Health Behav.* **37**(6), 851–858 (2013)
13. Cahapay, M., Rotas, E.: Remote learning stress, adaptive coping strategies, and academic performance of Filipino university students amid COVID-19 crisis. *Comput. Child.* **1**(1), 6 (2022)
14. Reid, G., Nygren, T.: The subjective workload assessment technique: a scaling procedure for measuring mental workload. *Adv in Psych.* **52**, 185–218 (1988)
15. Hart, S., Staveland, L.: Development of NASA-TLX: results of empirical and theoretical research. *Adv in Psych.* **52**, 139–183. North-Holland (1988)
16. Cohen, S., Kamarck, T., Mermelstein, R.: A global measure of perceived stress. *J. Health Soc. Behav.* **24**(4), 385–396 (1983)

17. Reeve, B., et al.: The Piper Fatigue Scale-12 (PFS-12): Psychometric findings and item reduction in a cohort of breast cancer survivors. *Breast Cancer Res. Treat.* 2012. **136**(1), 9–20 (2012)
18. Lovibond, S., Lovibond, P.: *Manual for the Depression Anxiety & Stress Scales*, 2nd edn. Psychology Foundation, Sydney (1995)



Research on Interactive Guidance Design in Metaverse Education and Training Systems

Shan Hu^(✉), Bingxin Zhang, Jitong Zhang, Weitao Jiang, and Lingda Rong

Wuhan University of Technology, Wuhan 430068, Hubei, China
2028621@qq.com

Abstract. This study focuses on the interactive guidance design methods in metaverse education and training systems, analyzing the main factors affecting learners' cognitive load levels and training learning outcomes in the metaverse education and training process. The study classifies six interactive guidance elements into two groups based on the amount of information conveyed from visual and auditory dimensions and sets them in two training modes: learning and practice. Using a 2×2 multifactorial controlled experiment, the study analyzes test scores, test completion times, and subjective cognitive load scale data of participants to examine the differences in training effectiveness and cognitive load levels among learners with different prior experiences in training modes composed of different interactive guidance elements. The experimental results indicate that the high-experience group outperformed the low-experience group in test performance across both training modes. For high-experience learners, there was no significant difference in test performance after training in both modes, but learners trained in the learning mode had slightly lower cognitive load levels than normal during the training process.

Keywords: Metaverse Education and Training · Cognitive Load Theory (CLT) · Interactive Guidance Design · Prior Experience · Interactive Guidance Elements

1 Introduction

In 2021, following a series of events such as the listing of Roblox, the first stock of the metaverse concept, and Facebook's official rebranding to Meta to fully embrace the metaverse, a wave of global interest in metaverse development surged. Metaverse education and training is the result of the deep integration of metaverse technology and educational training activities. The distinctive features of metaverse technology, such as technological integration, virtual reality, and immersive experience, are fully reflected in educational training activities.

Currently, some scholars have provided their research ideas for the design and development of metaverse education and training systems based on their respective fields. For example, Yoo et al. [1] connected facial recognition algorithms with the teaching management data system and developed a metaverse language education and training system suitable for different learning stages. Jagatheesaperumal et al. [2] and Pręgoska et al.

[3] explored the application of extended reality and IoT technology in metaverse education and skills training. Although education and training methods based on metaverse technology have been researched and applied in various professional fields, there is no unified standard for the instructional design of metaverse education and training systems. The rationality of the teaching process and its adaptability to different types of learners are still in the exploratory stage.

1.1 Cognitive Load Theory

The concept of cognitive load can be traced back to the 1950s [4, 5]. In 1988, Sweller provided a comprehensive description of the concept of cognitive load based on the Resource Limitation Theory and Schema Theory, and developed the Cognitive Load Theory (CLT) upon this foundation. Sweller defined cognitive load as the level of mental effort required to process given information [6]. This theory aims to explain how the information processing load generated during the learning process affects learners' ability to process new information and construct new knowledge.

According to the Cognitive Load Theory, the cognitive architecture of humans consists of working memory (short-term memory) and long-term memory, as well as the relationship between the two. The capacity and duration of working memory are limited, while the information capacity of long-term memory is unlimited. During the learning process, actively processing and integrating material information through working memory allows it to be stored in long-term memory. Additionally, working memory can access information stored in long-term memory without capacity limitations. Due to the limited capacity of human cognitive processing, only a limited number of information elements can be processed at once. When unnecessary demands are imposed on the cognitive system from the external environment, cognitive load increases. If the cognitive load becomes too high, learners may experience confusion, anxiety, or even abandon learning tasks. However, when the learning materials are relatively simple for the learners, the level of cognitive resources invested decreases, leading to feelings of boredom or laxity, which indirectly cause learners to experience distraction and lose motivation during the learning process.

In existing studies on cognitive load, some scholars have explored the impact of individual learner characteristics on cognitive load and learning outcomes. For example, Lee et al. [7] conducted a comparative experiment to study the differences in learning outcomes of learners with different spatial abilities in a desktop virtual reality learning environment, and provided explanations and analyses based on cognitive load theory. Other scholars have explored the impact of learning methods or instructional formats on learners' cognitive load. For example, Andersen et al. [8] studied the effects of distributed and concentrated training modes on participants' cognitive load by measuring their secondary task response times during virtual reality mastoidectomy training. Richardson et al. [9] compared the effects of 3D augmented reality maps and 2D topographic maps on learners' cognitive load and learning outcomes. Current research on metaverse education and training primarily focuses on aspects such as immersion and subjective emotions of learners during the learning process. However, studies on the changes in cognitive load

and its influencing factors in metaverse training environments are relatively few. Therefore, this paper focuses on the impact of learners' individual differences and training modes in metaverse education and training systems on learners' cognitive load.

1.2 Interactive Guidance Design

The concept of "guidance design" originally comes from urban architectural design. Guidance design is not just about signage design; it is also about designing media that help people establish richer, deeper relationships with spaces and events. For this reason, the term "guidance" has been interactive since its inception.

The metaverse differs from real-world learning environments in that multimodal information elements such as images, audio, and video are received by learners through various interactive methods (visual, auditory, tactile interactions) in the digital space. However, the multimodal feedback [10] in digital space, while providing a more comprehensive and detailed experience and enhancing the interaction between the system and learners, can lead to cognitive overload for learners. This can cause anxiety or laxity during the learning process, making it difficult to capture key learning points and absorb all the information provided by the digital space. Negut et al. [11] believe that tasks in virtual reality have higher cognitive complexity. Li Yueqi et al. [12] validated this viewpoint by comparing users' reading comprehension under paper reading, mobile reading, and virtual reality reading conditions. Therefore, interactive guidance design has strong reference value for the rationality of teaching processes and the intuitiveness of content presentation in metaverse education and training systems. Well-designed guidance elements can guide learners to quickly identify and understand task objectives, effectively reduce learning pressure, and improve training outcomes.

2 Method and Experiment

2.1 Method of Dividing Interactive Guidance Elements

Information volume refers to the extent to which users obtain information from learning materials. In metaverse education and training systems, learning materials convey information to learners through visual, auditory, tactile, and olfactory sensory channels, and the amount of information conveyed varies depending on different wayfinding elements and their combinations. Previous studies have shown [13] that the amount of information in augmented reality environments can affect learners' cognitive load and memory. This study starts from the visual and auditory dimensions, dividing the interactive guidance elements in the training system into three types: contour lines, progress, and audio, and categorizes each type of interactive guidance element into two groups based on the amount of information conveyed, as shown in Table 1.

At key points in the training, the system provides operation prompts and feedback in the form of displaying the contour lines of the target object visually, as shown in Fig. 1a. Previous studies have shown that dynamic indication can clearly convey the importance of information and attract observers' attention more effectively [14, 15], while static indication is more conducive to cognitive learning [16, 17].

Table 1. Classification of interactive guidance elements.

Sensory dimension	Interactive guidance elements	Classification	Information volume
Vision	Contour lines guidance	Dynamic contour line indication	High
		Static contour line indication	Low
	Progress guidance	Taskbar progress indication	High
		Timeline progress indication	Low
Audition	Audio guidance	Human voice indication	High
		Sound indication	Low

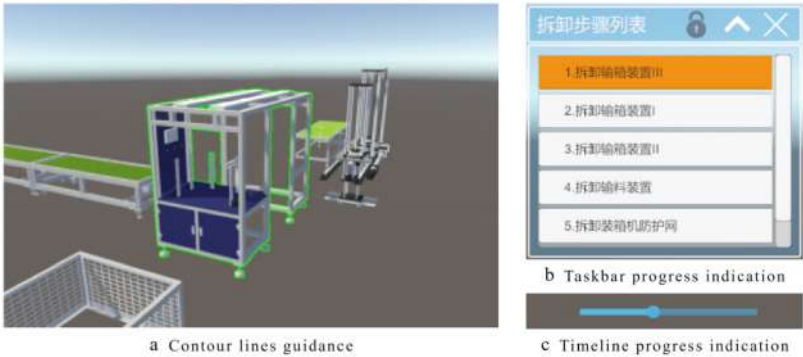


Fig. 1. Contour lines guidance and progress guidance.

Progress guidance refers to conveying information such as the current progress of tasks, the amount completed and remaining tasks, and their proportions to users in the human-computer interface. The presentation mode of progress elements in the interface significantly affects users' perception of time intervals [18]. The system includes two types of progress guidance: taskbar progress guidance and timeline progress guidance. The taskbar progress guidance will use text information in the interface taskbar to sequentially explain each operation, and provide feedback on the completed, ongoing, and pending states of the operation steps using different colors, as shown in Fig. 1b; the timeline progress guidance is represented by a progress bar without text information, only indicating the proportion of completed and pending operation steps, as shown in Fig. 1c.

Audio guidance refers to the system providing learners with task instructions, operation methods, or feedback results through human voice or sound effects. Human voice

is particularly effective in conveying complex information and facilitating learning [19], while sound effects can enhance the perception of spatial information and complement or reinforce visual information [20, 21]. Therefore, this study categorizes audio guidance into two types: human voice and sound. Voice indication provide learners with descriptive text-based instructions in the form of simulated human voices, while sound indication provide auditory feedback to learners during and after their operation processes based on the operation results. The volume of audio materials used in the system remains consistent and is not affected by the distance in the virtual space.

2.2 Experiment

The experiment adopted a 2×2 factorial design, providing two types of assembly training tasks: “learning-exam” and “practice-exam”. The independent variables were Factor 1: training mode (learning mode, practice mode) and Factor 2: participants’ mechanical assembly experience (high, low). The dependent variables were assessment completion time and assessment scores. The experimental results were analyzed and discussed using SPSS 26.0, combined with subjective measurement scales of cognitive load.

The Experimental Equipment and Materials. A 27-inch LCD monitor with a refresh rate of 60 Hz and a resolution of 1920*1080 was used as the display device in this experiment, as shown in Fig. 2. The experimental materials included an automatic packing machine assembly training system developed by the Information and Interaction Design Research Institute team at Hubei University of Technology. This system is based on the Unity 3D engine and X-Dreamer plugin, developed using the C# language.



Fig. 2. Experimental environment.

Training Models. Based on the classification of interactive guidance elements, the metaverse assembly training system is set up with three types of modes: “Learning”, “Practice”, and “Assessment”.

The learning mode consists of interactive guidance elements such as dynamic contour indication, taskbar progress, and human voice indication, which convey more information. This mode requires the system to guide the learner step by step to complete

the equipment disassembly process. However, due to program limitations, the learner's operational freedom is low, and the interaction with the system is weak. The interactive interface is shown in Fig. 3a.

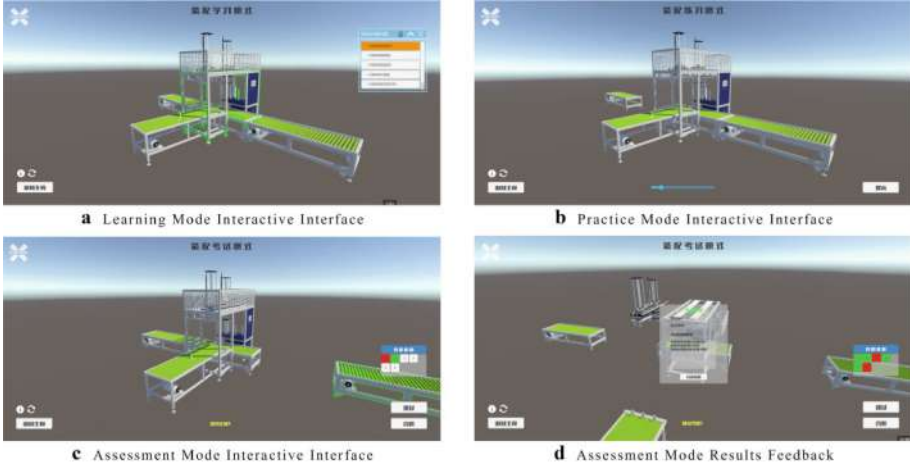


Fig. 3. Training model interactive interface.

The practice mode consists of interactive guidance elements such as static contour indication, timeline progress indication, and sound indication, which convey less information. The system does not provide specific step-by-step guidance but gives feedback on the correctness of the learner's actions based on their specific operations. In practice mode, learners have greater operational freedom and more space for independent thinking and exploration, but the amount of information they can obtain is less compared to the "learning" mode. The interactive interface is shown in Fig. 3b.

The assessment mode is a direct way to test learners' learning outcomes. This mode requires learners to follow steps to complete the equipment disassembly process within a specified time. During the assessment, there are no text or sound indication, and answers can only be submitted or skipped using the "Answer" or "Skip" buttons. The system will play the corresponding animation of the equipment based on the learner's operation and move on to the next step. There are 25 operation steps in the assessment mode, each worth 4 points. Incorrect answers or skips will result in a deduction, and the total score is 100 points. After the assessment, the system records the completion time and score. The assessment mode interface and score feedback are shown in Fig. 3c and d.

Subjective Cognitive Load Measurement Scale. The WP self-assessment scale [22] is a subjective workload assessment scale proposed by Tsang and Velazquez based on the multiple resources model of workload. The Multiple Resources Model of Mental Load divides the completion of learning tasks by learners into four stages, each stage consuming two different mental resources [23]. Compared to other internationally recognized subjective cognitive load measurement tools like Pass [24] and NASA-TLX (National Aeronautics and Space Administration-Task Load Index)[25], the WP self-assessment

scale has advantages in terms of convergent validity, concurrent validity, diagnosticity, and sensitivity in measuring cognitive load for assembly tasks [26]. Therefore, the WP Self-Rating Scale was selected as the measurement tool for cognitive load in this experiment. The operational definitions of each dimension of the scale are provided in Table 2.

Table 2. Operational definition of the WP self-assessment scale.

Dimension	Operational definition
Central Processing Resources	Mental resources consumed during task selection and execution
Response Resources	Mental resources required to respond to the task
Spatial Coding Resources	Mental resources required to respond to the task
Linguistic Coding Resources	Mental resources consumed for linguistic activities during the task
Visual Reception Resources	Mental resources consumed for visual information processing
Auditory Reception Resources	Mental resources consumed for auditory information processing
Operational Resources	Influence of physical operation on cognitive load during the task

Participants. This experiment selected a total of 48 undergraduate students from the School of Industrial Design and the School of Mechanical Engineering at Hubei University of Technology as participants. Twenty basic questions were chosen from the “Mechanical Assembly Technology” textbook [27] to design a questionnaire, testing whether the participants had any mechanical equipment assembly experience before receiving simulation training. The full score of this test questionnaire is 100 points, and a passing score of 60 points is used to divide the participants into two categories: high assembly experience (≥ 60 points) and low assembly experience (< 60 points). Subsequently, participants from both categories were required to participate in training tasks that included either learning mode or practice mode. Among the high assembly experience participants, those who received training tasks that included learning mode (Task 1) were classified as group A_1 , and those who received training tasks that included practice mode (Task 2) were classified as group A_2 . Participants with low assembly

Table 3. Experimental grouping situation.

	Task 1	Task 2
High assembly experience participants	$A_1(n = 9)$	$A_2(n = 9)$
Low assembly experience participants	$B_1(n = 15)$	$B_2(n = 15)$

experience who received Tasks 1 or 2 were classified as groups B₁ and B₂, respectively. The specific grouping is shown in Table 3.

To ensure that the comparison data between different training modes in high and low assembly experience groups are not influenced by differences in assembly test scores, independent sample T-tests were conducted on the assembly test scores of participants in the A₁, A₂, B₁, and B₂ groups. The test results indicate that there is no significant difference in the average assembly test scores between the A₁ and A₂ groups, as well as between the B₁ and B₂ groups ($P > 0.05$). The results are shown in Table 4.

Table 4. Results of independent samples T-tests.

Group	The assembly test scores	T	P
A ₁ (n = 9)	72.22 ± 6.18	−0.157	0.877
A ₂ (n = 9)	71.67 ± 8.66		
B ₁ (n = 15)	45.67 ± 4.95	0.504	0.618
B ₁ (n = 15)	46.67 ± 5.88		

The Experimental Tasks. To ensure the validity of the experimental data, participants in both learning mode and practice mode are required to complete the same training content. During the training process, the occurrence frequency, nodes, and duration of appearance of contour lines guidance and audio guidance (progress guidance continuously displayed in the interactive interface without special instructions) remain consistent between the two modes, with differences only in presentation format. Training content and the distribution of interactive guide elements is shown in Fig. 4.

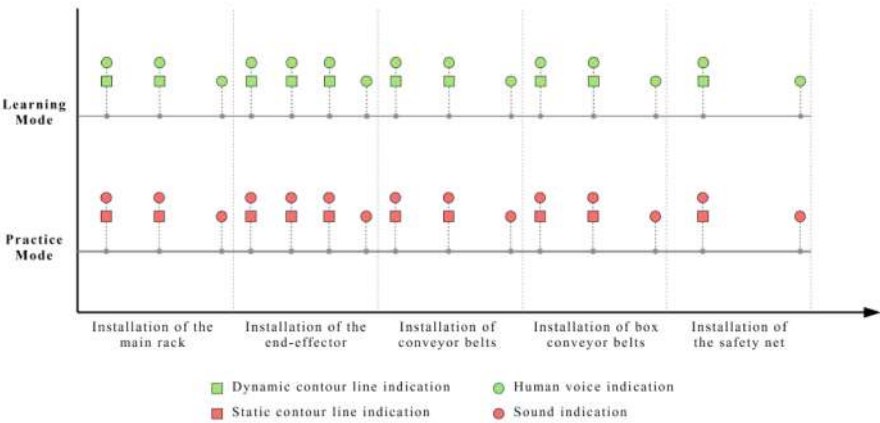


Fig. 4. Interactive guidance elements in the learning and practice model.

Two training tasks were set up in this experiment. Task 1 required participants to understand the specific disassembly and assembly process of the equipment step by step through the learning mode. After completing all disassembly and assembly steps, they directly entered the assessment mode for testing. Task 2 required participants to directly engage in the practice mode for free exploration and practice. After completing all assembly operation steps in the practice mode, they entered the assessment mode for testing. Before entering the assessment mode for both tasks, participants were required to rest for five minutes to reduce the impact of working memory on test performance during learning or practice. After the assessment, participants were asked to fill out the WP self-assessment scale based on the actual training process. After the experiment, some participants were invited for interviews to understand their experience.

3 Results

3.1 Task Performance Data

First, a one-way ANOVA and post-hoc multiple comparisons were conducted on the assessment scores of groups A_1 , A_2 , B_1 , and B_2 . The results are shown in Table 5.

Table 5. Results of one-way ANOVA and post-hoc multiple comparisons of assessment scores data for the four groups of participants.

Group	Assessment scores	F	P	Post-hoc multiple comparisons
$A_1(n = 9)$	84.00 ± 4.90	13.10	< 0.05	$A_1 > B_1^{***}$; $A_1 > B_2^{***}$ $A_2 > B_1^{***}$; $A_2 > B_2^{***}$ $B_1 > B_2^{***}$; $A_1 \approx A_2$
$A_2(n = 9)$	84.89 ± 6.25			
$B_1(n = 15)$	77.07 ± 6.32			
$B_2(n = 15)$	70.67 ± 7.04			

Note: *** indicates $P < 0.01$; ** indicates $P < 0.05$; * indicates $P < 0.1$

Analyzing with assessment scores as the dependent variable, the assessment scores of participants in groups A_1 and A_2 were significantly higher than those of groups B_1 and B_2 , $P = 0.000 < 0.01$. The assessment scores of group B_1 were also significantly higher than those of group B_2 , $P = 0.000 < 0.01$. There was no significant difference in the assessment scores between groups A_1 and A_2 , $P = 1.000 > 0.1$. Then, a one-way ANOVA and post-hoc multiple comparisons were conducted on the assessment completion time data of groups A_1 , A_2 , B_1 , and B_2 . The results are shown in Table 6.

Using assessment completion time as the dependent variable, the analysis shows that the assessment time for participants in group A_1 was lower than that of group B_1 , $P = 0.075 < 0.1$, and significantly lower than that of group B_2 , $P = 0.000 < 0.01$. The assessment time for participants in group A_2 was lower than that of group B_1 , $P = 0.032 < 0.05$, and significantly lower than that of group B_2 , $P = 0.000 < 0.01$. The assessment time for group B_1 was also lower than that of group B_2 , $P = 0.049 < 0.05$. There was no significant difference in assessment time between groups A_1 and A_2 , $P = 1 > 0.1$.

Table 6. Results of one-way ANOVA and post-hoc multiple comparisons of assessment completion time data for the four groups of participants.

Group	Assessment completion time	F	P	Post-hoc multiple comparisons
A ₁ (n = 9)	181.36 ± 7.75	136.29	< 0.05	A ₁ > B ₁ *; A ₁ > B ₂ *** A ₂ > B ₁ **; A ₂ > B ₂ *** B ₁ > B ₂ **; A ₁ ≈ A ₂
A ₂ (n = 9)	179.94 ± 7.51			
B ₁ (n = 15)	201.77 ± 9.33			
B ₁ (n = 15)	234.65 ± 5.56			

Note: *** indicates $P < 0.01$; ** indicates $P < 0.05$; * indicates $P < 0.1$

3.2 Subjective Cognitive Load Scale Data

To ensure the scientific validity and rationality of the scale data, the reliability of the WP self-assessment scale data for 48 participants was first tested through Cronbach’s α coefficient analysis, and the results are shown in Table 7.

Table 7. Reliability test of the WP self-assessment scale.

Name of the scale	Number of items	Sample size	Cronbach’s α coefficient
the WP self-assessment scale	8	48	0.900

After confirming the good overall reliability of the scale, one-way ANOVA was conducted to compare whether there were significant differences in cognitive load among the four groups of participants in each dimension and overall cognitive load level. The results are shown in Table 8.

Table 8. One-way ANOVA of WP self-assessment scale dimensions and overall cognitive load values for the 4 groups of participants.

Dimension	A ₁ (n = 9)	A ₂ (n = 9)	B ₁ (n = 15)	B ₂ (n = 15)	F	P
Central Processing Resources	3.44 ± 0.88	5.56 ± 0.88	5.00 ± 0.93	8.27 ± 0.88	62.72	0.000
Response Resources	3.67 ± 1.00	5.33 ± 1.00	5.07 ± 1.03	7.40 ± 1.18	25.289	0.000
Spatial Coding Resources	2.44 ± 0.88	4.78 ± 0.97	4.93 ± 0.88	6.60 ± 0.99	37.366	0.000
Linguistic Coding Resources	3.78 ± 0.97	4.56 ± 0.53	5.53 ± 0.74	3.80 ± 0.78	15.888	0.000

(continued)

Table 8. (continued)

Dimension	A ₁ (n = 9)	A ₂ (n = 9)	B ₁ (n = 15)	B ₂ (n = 15)	F	P
Visual Reception Resources	4.89 ± 0.78	5.78 ± 1.09	6.00 ± 0.76	6.60 ± 0.99	6.823	0.001
Auditory Reception Resources	4.33 ± 0.71	5.11 ± 0.78	6.00 ± 0.93	6.33 ± 0.90	12.302	0.000
Operational Resources	3.11 ± 1.05	4.00 ± 1.12	4.87 ± 0.74	6.47 ± 1.30	21.231	0.000
Total Cognitive Load	3.67 ± 0.57	5.02 ± 0.51	5.34 ± 0.58	6.50 ± 0.53	51.016	0.000

The analysis reveals significant differences among the participants in the seven dimensions of the WP self-assessment scale and the total cognitive load value. Post-hoc multiple comparisons show that in dimensions such as central processing resources, response execution resources, and spatial encoding processing resources, as well as in the total cognitive load value, the cognitive resource utilization of groups A₂ and B₁ is significantly higher than that of group A₁ and significantly lower than that of group B₂, with no significant difference between groups A₂ and B₁; in the language encoding resource dimension, the cognitive resource utilization of groups A₁, A₂, and B₂ is significantly lower than that of group B₁, with no significant differences among the three groups; in the dimensions of visual reception resources, auditory reception resources, and operational resources, the cognitive resource utilization of groups A₁, A₂, B₁, and B₂ shows significant differences, exhibiting an increasing trend. Finally, using assembly test scores as the independent variable and cognitive load total value as the dependent variable, a simple linear regression analysis was conducted to examine the influence of participants' previous experience levels on cognitive load under both learning and practice modes. The results are shown in Tables 9 and 10.

Table 9. Table of regression coefficients for learning patterns.

Model	Standardized coefficient β		Standardized coefficient	t	Significance
	B	Standard error	Beta		
Constant	8.057	0.426		18.914	0.000
Assembly test scores	−0.058	0.007	−0.855	−7.742	0.000
F			59.931		
P			< 0.001		
R ²			0.731		
Dependent variable: Total cognitive load					

Table 10. Table of regression coefficients for practice patterns.

Model	Standardized coefficient β		Standardized coefficient	t	Significance
	B	Standard error	Beta		
Constant	9.909	0.509		19.484	0.000
Assembly test scores	−0.069	0.009	−0.853	−7.677	0.000
F			58.935		
P			< 0.001		
R ²			0.716		
Dependent variable: Total cognitive load					

The analysis from Table 9 indicates that in the learning mode, participants’ assembly test scores significantly influence their cognitive load total values during the training process, with $P = 0.000 < 0.05$ and a regression coefficient of -0.058 . This suggests that participants’ previous experience levels significantly negatively impact their cognitive load total values. Specifically, when participants’ assembly test scores increase by 1 point, their cognitive load total values during the training process decrease by 0.058. The analysis from Table 10 indicates that in the practice mode, participants’ assembly test scores significantly influence their cognitive load total values during the training process, with $P = 0.000 < 0.05$ and a regression coefficient of -0.069 . This suggests that participants’ previous experience levels significantly negatively impact their cognitive load total values. Specifically, when participants’ assembly test scores increase by 1 point, their cognitive load total values during the training process decrease by 0.069.

4 Discussion

4.1 Previous Experience

According to the single-factor analysis of task performance data, it is evident that the pre-training mechanical assembly experience of the subjects and the design of the training mode in the training system will significantly affect the training assessment scores and completion time. Subjects with some mechanical assembly experience before training generally perform better in assembly operation accuracy after training compared to those lacking assembly experience, and they also spend less time completing the operations. Furthermore, based on the one-way linear regression analysis of subjective cognitive load questionnaire data in the two modes, whether it is learning mode or practice mode training, the subjects’ prior experience will significantly negatively affect their cognitive load level during the training process. In other words, the higher the subjects’ prior experience, the more comprehensive their understanding of mechanical assembly knowledge, and the lower their cognitive load level during the training process.

4.2 Training Mode

In this experiment, the different training modes did not significantly affect the training assessment scores and completion time of participants with higher mechanical assembly experience. However, analysis of the WP self-assessment scale data revealed that participants undergoing learning mode training generally had lower levels of cognitive load. Interviews conducted after the experiment revealed that participants in this group felt bored and frustrated with the training process and experienced a lack of concentration. Conversely, participants undergoing practice mode training had relatively normal levels of cognitive load and were more actively exploring assembly methods guided by interactive elements in the system. They also demonstrated better concentration during the learning process.

For participants with lower assembly experience, compared to practice mode training, learning mode training with richer guided information and step-by-step operations allows them to complete assembly tasks more quickly and accurately in the assessment mode. Additionally, according to the analysis of the WP self-assessment scale data, participants undergoing practice mode training generally have higher levels of cognitive load, particularly evident in dimensions such as central processing resources, response execution resources, and spatial coding processing resources. Post-experiment interviews revealed that participants undergoing practice mode training felt confused about assembly tasks during the training process. Some participants believed that the training system provided insufficient prompt information to help them complete assembly operations, leading to anxiety and tension due to difficulty in understanding task instructions. On the other hand, participants undergoing learning mode training felt that the step-by-step instructional approach enabled them to effectively acquire assembly knowledge and skills. Some participants who had never been exposed to mechanical assembly showed great interest in the metaverse-style assembly training.

4.3 Interactive Guidance Elements

The difference in the amount of information presented by different types of interactive guiding elements is the fundamental reason for the different levels of cognitive load experienced by participants with different prior experiences in the two training modes.

Comparing the data of the WP self-assessment scale dimensions between the A₁ and A₂ groups, it can be observed that participants with higher mechanical assembly experience consume slightly less psychological resources in the information input stage (dimensions of visual reception resources and auditory reception resources) during learning mode training compared to the normal level in practice mode. This indicates that elements such as dynamic contour prompts, taskbar progress prompts, and voice prompts in learning mode, while conveying more information through visual and auditory sensory channels, fail to effectively attract the attention of highly experienced participants. Conversely, elements in practice mode, such as static contour prompts, timeline progress prompts, and sound prompts, which convey only a small amount of key information, can effectively attract the attention of highly experienced participants to prompt information during the training process. Additionally, highly experienced participants undergoing

learning mode training consume significantly fewer psychological resources in dimensions such as central processing resources, response resources, spatial coding resources, and language coding resources compared to practice mode. This further indicates that dynamic contour prompts, taskbar progress prompts, voice prompts, and other interactive guiding elements, while conveying more information, mostly consist of repetitive or ineffective information for highly experienced participants. As a result, their investment of psychological resources in information and encoding processing stages is lower than normal, leading to an inability to maintain a focused learning state during the training process and a lack of immersive experience.

Comparing the data of the WP self-assessment scale dimensions between the B_1 and B_2 groups, it can be observed that low-experience participants undergoing practice mode training consume slightly fewer psychological resources in the dimensions of visual reception resources and auditory reception resources compared to the learning mode, but consume significantly higher psychological resources in dimensions such as central processing resources, response resources, and spatial coding resources, indicating that the limited information conveyed by static contour prompts, timeline progress prompts, and sound prompts in practice mode can effectively attract the attention of low-experience participants, but is insufficient to help them understand the system's task instructions, leading to an overload of cognitive load during information and encoding processing stages. Additionally, low-experience participants undergoing practice mode training consume significantly fewer psychological resources in the dimension of language coding resources compared to normal level, because sound prompts and timeline progress prompts convey less language information compared to voice prompts and taskbar progress. Overall, interactive guiding elements such as static contour prompts, timeline progress, and sound prompts in practice mode cannot provide sufficient information to low-experience participants to ensure their effective mastery of training content, while in learning mode, interactive guiding elements such as dynamic contour prompts, taskbar progress prompts, and voice prompts can help low-experience participants effectively learn through visual and auditory sensory channels.

5 Conclusion

The experimental results show that factors such as the participants' prior experience and the combination of interactive guiding elements in the training mode affect the participants' cognitive load levels and training effectiveness during the training process. Therefore, when designing interactive guiding elements for the instructional content of a metaverse education training system for learners with different prior experiences, it is crucial to first clarify the design objectives to ensure that different types of learners can benefit from the training process.

For learners with low levels of experience, clear and targeted operational guidance should be provided. The selection of interactive guiding elements should be based on the principles of complete and clear information transmission and timely feedback, aiming to help learners understand and master the instructional content more comprehensively. For learners with high levels of experience, a certain degree of free and open exploration space should be provided. The selection of interactive guiding elements should be based

on the principles of concise and accurate information transmission and the avoidance of irrelevant information interference, aiming to create an open and smooth learning experience and stimulate learners' interest. In addition, the metaverse education training system should provide tests related to the instructional content before training to assess learners' prior experience levels and recommend targeted training courses. This helps learners develop a scientific and reasonable training plan, thereby achieving training goals more efficiently.

Although this study has made some explorations in the research and design of metaverse education training systems, there are still some shortcomings. First, due to the limitations of the experimental conditions, the sample size of the participants in this study was relatively small, which may affect the reliability and generalizability of the results. Secondly, measuring participants' cognitive load solely through self-assessment scales has a certain degree of subjectivity, which may differ from their actual cognitive load. Therefore, in future studies, the sample size will be expanded to gain a more comprehensive understanding of the impact of metaverse education training systems on different learners. Additionally, objective physiological indicators such as heart activity, brain activity, and eye activity will be considered to more objectively assess learners' cognitive load levels, thereby improving the reliability and accuracy of the research results.

References

1. Yoo, G.-S., Chun, K.: A study on the development of a game-type language education service platform based on metaverse. *dcs*. **22**, 1377–1386 (2021)
2. Jagatheesaperumal, S.K., Ahmad, K., Al-Fuqaha, A., Qadir, J.: Advancing education through extended reality and internet of everything enabled metaverses: Applications, challenges, and open issues. (2022).
3. Pregowska, A., Osial, M., Gajda, A.: What will the education of the future look like? How have Metaverse and Extended Reality affected the higher education systems? *Metaverse Basic Appl. Res.* **3**, 57 (2023)
4. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97 (1956)
5. Peterson, L., Peterson, M.J.: Short-term retention of individual verbal items. *J. Exp. Psychol.* **58**, 193–198 (1959)
6. Sweller, J.: Cognitive load during problem solving: Effects on learning. *Cogn. Sci.* **12**, 257–285 (1988)
7. Lee, E.A.-L., Wong, K.W.: Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Comput. Educ.* **79**, 49–58 (2014)
8. Andersen, S.A.W., Mikkelsen, P.T., Konge, L., Cayé-Thomasen, P., Sørensen, M.S.: Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope*. **126**(2), E74–E79 (2016)
9. Richardson, R., Sammons, D., Delparte, D.: Augmented affordances support learning: Comparing the instructional effects of the augmented reality sandbox and conventional maps to teach topographic map skills. *J. Interact. Learn. Res.* (2017)
10. Liu, W.: *Human-Machine Fusion: Beyond Artificial Intelligence*. Tsinghua University Press (2021)

11. Neguț, A., Matu, S.-A., Sava, F.A., David, D.: Task difficulty of virtual reality-based assessment tools compared to classical paper-and-pencil or computerized measures: A meta-analytic approach. *Comput. Hum. Behav.* **54**, 414–424 (2016)
12. Yueqi, L., Liwei, W., Ran, L., Xin, L.: An experimental study on the comparison of the effectiveness and immersion experience of virtual reality and traditional reading comprehension. *Intell. Theory Pract.* **46**, 127–135 (2023)
13. Lim, T., Bong, J., Kang, J., Dennen, V.: The effects of information volume and distribution on cognitive load and recall: Implications for the design of mobile marker-less augmented reality. *Educ. Technol. Int.* **20**(2), 137–168. Presented at the October 31 (2019)
14. Shan, H., Mingyu, Z., Huan, X.: Research and application of visual cognitive differences in interactive guide design. *Packag. Eng.* **40**, 12–17 (2019)
15. Jiang, W., Zhang, B., Sun, R., Zhang, D., Hu, S.: A study on the attention of people with low vision to accessibility guidance signs. *J Multimodal User Interfaces.* **18**, 87–101 (2024)
16. Wagner, I., Schnotz, W.: Learning from static and dynamic visualizations: What kind of questions should we ask? In: *Learning from Dynamic Visualization*, pp. 69–91. Springer International Publishing, Cham (2017)
17. Suyatna, A., Anggraini, D., Agustina, D., Widyastuti, D.: The role of visual representation in physics learning: Dynamic versus static visualization. *J. Phys. Conf. Ser.* **909**, 012048 (2017)
18. Branaghan, R.J., Sanchez, C.A.: Feedback preferences and impressions of waiting. *Hum. Factors.* **51**, 528–538 (2009)
19. Hafen, R.P., Henry, M.J.: Speech information retrieval: A review. *Multimedia Systems.* **18**, 499–518 (2012)
20. Rocchesso, D.: Audio effects to enhance spatial information displays. In: *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission.* IEEE Comput. Soc
21. Rigas, D., Yu, H., Memery, D.: Experiments using speech, non-speech sound and stereophony as communication metaphors in information systems. In: *Proceedings 27th EUROMICRO Conference. 2001: A Net Odyssey*, pp. 383–390. IEEE Comput. Soc, Warsaw, Poland (2001)
22. Tsang, P.S., Velazquez, V.L.: Diagnosticity and multidimensional subjective workload ratings. *Ergonomics.* **39**, 358–381 (1996)
23. Egeth, H., Kahneman, D.: Attention and effort. *Am. J. Psychol.*, 339 (1975)
24. Paas, F.G.W.C., Van Merriënboer, J.J.G.: Instructional control of cognitive load in the training of complex cognitive tasks. *Educ. Psychol. Rev.* **6**, 351–371 (1994)
25. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: *Hancock, P.A., Meshkati, N. (eds.) Advances in Psychology*, pp. 139–183, North-Holland (1988)
26. Qi-Wei, Y., Ding-Zhong, F., Xinsheng, X., Qi-Heng, C., Hao-Jie, H.: A lego simulation-based approach to measurement of brain load in assembly operations. *Ind. Eng. Manag.* **23**, 53–59 (2018)
27. Bing, X.: *Mechanical Assembly Technology*. China Light Industry Press (2014)



Identification of Technical and Ergonomic Requirements for Full-Scale Simulation with Extracorporeal Membrane Oxygenation

Steffen Kuklinski¹✉, Simon König¹, Maximilian Kehmann¹, and Claus Backhaus^{1,2}

¹ University of Applied Sciences Münster, Centre for Ergonomics and Medical Engineering, Bürgerkamp 3, 48565 Steinfurt, Germany
steffen.kuklinski@fh-muenster.de

² Technical University Berlin, Institute of Psychology and Ergonomics, Fasanenstraße 1, 10623 Berlin, Germany

Abstract. Intensive care medicine treats patients who may suffer from severe cardiac or pulmonary disorders. For these patients, extracorporeal membrane oxygenation (ECMO) can be a life-saving therapy. The application of ECMO requires an interdisciplinary and highly specialised team. The treatment must therefore be practised as a team. However, there are currently no suitable simulators available that enable interdisciplinary ECMO trainings. In this study, ergonomic and technical requirements for the development of a full-scale ECMO simulator are analysed. The development of the requirements is based on the process of user-centred product development. It is structured into (1.) carrying out expert interviews to identify relevant work processes, (2.) the analysis and visualisation of the work processes and (3.) deriving requirements together in an expert workshop based on the visualised work processes. A total of 15 work processes were identified as relevant and then observed open and non-participatory. On this basis, 39 requirements were derived. These requirements now need to be implemented in a prototype and tested by additional users.

Keywords: full-scale simulation · medical training · extracorporeal membrane oxygenation · patient safety · process analysis

1 Situation

Intensive care medicine treats patients who may suffer from severe cardiac or pulmonary disorders. Conventional medication and ventilation therapies are not sufficient for severe disorders. For these patients, extracorporeal membrane oxygenation (ECMO) can be a life-saving therapy. For the application of ECMO, large blood vessels in the thighs or neck must be cannulated to gain access to the patient's cardiovascular system. The treatment often takes place during resuscitation and requires the cooperation of various specialists from the professions of intensive care medicine, cardiothoracic surgery, anaesthesia and clinical perfusion [1].

The treatment must therefore be practised as a team. However, there are no suitable simulators available to enable interdisciplinary trainings with ECMO. They can only simulate specific aspects of the treatment, such as resuscitation of the patient or the cannulation of the blood vessels for ECMO [2].

In this study, requirements for an ECMO simulator are identified, that enable co-operation of various medical and medical-technical professions.

2 Methods

The development of the ECMO simulator is based on a user-centred development process [3]. The approach is structured as follows.

2.1 Expert Interviews

Experts from the participating professions are interviewed in a guided manner about the relevant work processes in the care of ECMO patients.

2.2 Analysis of the Work Processes

Work processes identified in the interviews are observed in an open, non-participatory manner in a hospital. Workflows that are not accessible for real observation (emergency cases) are simulated in a training centre. The workflows are illustrated in the form of flowcharts [4].

2.3 Deriving Requirements

The flowcharts created were validated in an expert workshop and analysed regarding the technical and medical requirements for the ECMO simulator. The requirements are differentiated into essential and optional requirements and compiled in a list of requirements.

3 Results

3.1 Results of the Expert Interviews

A total of 6 clinical experts from the fields of cardiothoracic surgery, anaesthesia, intensive care medicine and clinical perfusion were interviewed and 15 work processes were identified as particularly relevant and included in the following analysis.

3.2 Results of the Analysed Work Processes

There were 4 full-shift observations in an internal intensive care unit, 5 full-shift observations in the department of clinical perfusion and 1 work shift in a cardiac arrest team which were documented in the form of process flow analyses. In total 14 of the 15 identified work processes were analysed. An implantation of an ECMO could not be observed and was analysed in a simulation instead. This took place in the training centre of Münster University Hospital with the participation of the professions of intensive care medicine, cardiothoracic surgery, anaesthesia and clinical perfusion.

Figure 1 shows the clinical users during the simulation while securing the airway with the simulator HAL S3000 (Gaumard, Canada), as well as while implanting the arterial cannula in the ECMO trainer (QleSim, Germany).



Fig. 1. Simulation of emergency treatment with the HAL S3000 simulator (right) and cannulation of an artery on the ECMO trainer simulator (left).

The simulators are not designed for full emergency implantation and do not allow the critical care team to practice their team performance. The analysed work processes are visualised in flowcharts. Figure 2. shows an example of the flowcharts for the emergency implantation.

3.3 Derived Requirements

The flowcharts were validated with 7 experts from the fields of cardiothoracic surgery, anaesthesia and intensive care medicine, clinical perfusion and medical training. A total of 39 requirements for the simulator were derived from the flowcharts of which 21 are essential and 18 are optional. Table 1 shows the essential requirements.

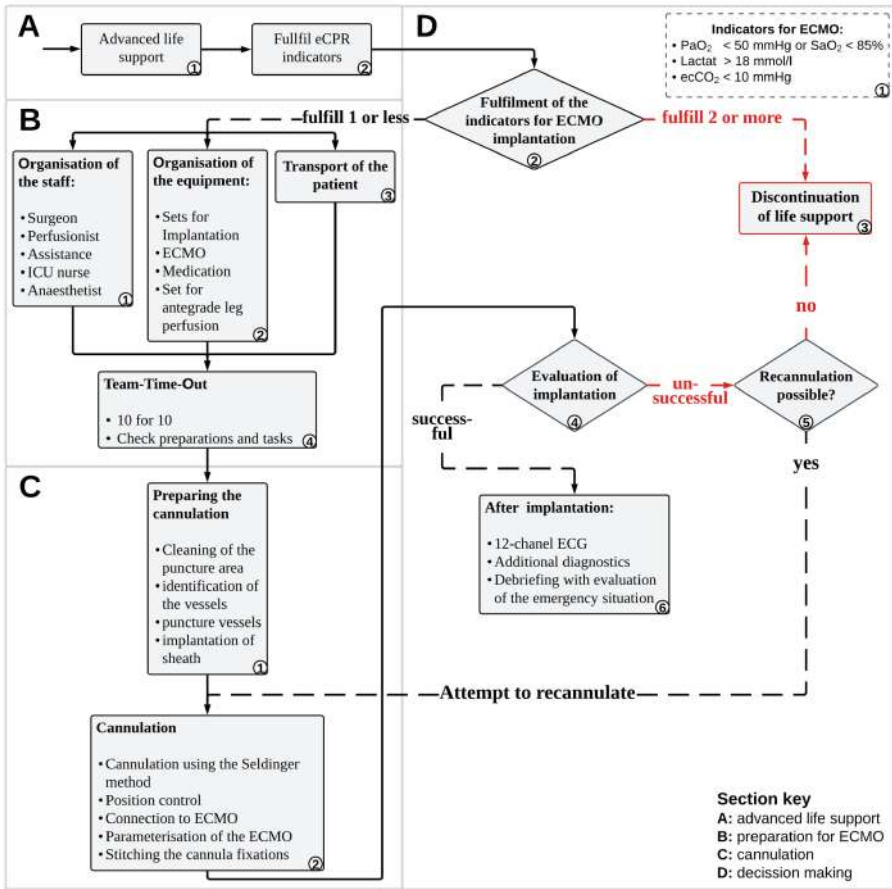


Fig. 2. Process for the emergency implantation of an ECMO in case of cardiac arrest: fast implantation of cannulas into large blood vessels to temporarily replace the heart or lung function.

Table 1. Essential requirements for an ECMO simulator.

No.	Requirement	Value/description
1	Puncturable vessels in the thighs and neck	Thighs: 1 $\times$ artery, 1 $\times$ Vein Neck: 1 $\times$ vein
2	Physiologically correct flow directions	Art. in the thigh: away from the heart Vein in the thigh: towards the heart Neck vein: towards the heart

(continued)

Table 1. (*continued*)

No.	Requirement	Value/description
3	Position of the blood vessels in the thigh	Depth of the artery: 25 mm Depth of the vein: 35 mm Distance between the vessels: 10 mm
4	Position of the blood vessel on the neck (centre, right)	Depth of the vein: 20 mm Distance between the vessels: 3 mm
5	Vessel diameter	$\geq 3/8$ in
6	Vessel length	≥ 0.5 m
7	Required size of the puncture area	70 × 130 mm
8	Puncture surface material	Vessels must be detectable by sonography
9	Adjustable pulsation in the art. Vessels	0–225 mmHg
10	Pulsation frequency	0–140 bpm
11	Flow	0–6.5 l/min
12	Variable mean Vascular pressures	Vene: 0–200 mmHg Artery: 0–200 mmHg
13	Parameter output on the ECMO console	p _{Art} : 0–200 mmHg p _{vein} : 0–200 mmHg Flow: 0–6.5 l/min
14	Contactless control	Tablet PC
15	Emergency manual control	Flow, pulsation, pressure must be adjustable on the simulator
16	Adjustable vital signs on a patient monitor	ECG, SP0 ₂ , artery blood pressure waveform, respiratory rate
17	Compressible thorax	Compression depth: 70 mm at 600 N
18	Intubation of the airway	Oral tube, size 6.5, 7.0, 7.5 Laryngeal mask, size 3, 4, 5
19	Connection for invasive blood pressure monitoring	Puncture site on the wrist (radial artery)
20	Intravenous medication	Puncture area for intravenous medication
21	Connection of a defibrillator	Defibrillator dummy is sufficient

4 Discussion

Based on the process of user-centred product development, an interdisciplinary team of doctors, technicians and nurses derived a total of 39 requirements for a patient simulator for interdisciplinary training with ECMO.

Interdisciplinary team training improves patient safety by allowing clinical staff to practise life-threatening situations in a controlled environment. The training increases team performance by optimising communication, coordination and effective teamwork [5]. To develop the simulator, an iterative approach with clinical users was carried out. The requirements identified now need to be implemented in a prototype.

A limitation of the study is the restriction to 1 university hospital. The results were discussed with experts from additional hospitals. ECMO therapy is mostly standardised due to its invasiveness [6]. We assume that the requirements are transferable. The prototype must therefore be tested with clinical users in additional hospitals.

References

1. Nagaoka, E., et al.: Efficacy of multidisciplinary team approach with extracorporeal membrane oxygenation for COVID-19 in a low volume ECMO center. *Artif. Organs.* **45**(9), 1061–1067 (2021)
2. Trummer, G., et al.: Training module extracorporeal life support (ECLS): Consensus statement of the DIVI, DGTHG, DGfK, DGAI, DGIIN, DGF, GRC and DGK2021. *Med Klin Intensivmed Notfmed.* **116**(7), 605–608 (2021)
3. International Organization for Standardization (ISO): Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO 9241-210:2019). Geneva: ISO (2019)
4. Hering, E.: Programme flow chart according to DIN 66001. In: *Software-Engineering*. Vieweg+Teubner Verlag, Wiesbaden (1984)
5. Salas, E., Wilson, K.A., Burke, C.S., Priest, H.A.: Using simulation-based training to improve patient safety: what does it take? *Jt. Comm. J. Qual. Patient Saf.* **31**(7), 363–371 (2005)
6. Association of the Scientific Medical Societies in Germany (AWMF): S3 guideline 011/021: Use of extracorporeal circulation (ECLS/ECMO) in cardiovascular failure (2021)



How Well a Simulation-Training Program Prevents WMSDs Among Nursing Staff?

Bouhafs Mebarki¹ , Brahim Zaoui¹ , Mohamed Mokdad² ,
and Imed Eddine Mebarki^{3,4} 

¹ Laboratory of Ergonomics and Risks Prevention, University of Oran 2,
31000 Bir El Djir, Algeria

mebarki.bouhafs@univ-oran2.dz

² Department of Psychology, University of Bahrain, Zellaq 1054, Bahrain

³ Centre Hospitalier de Sarrebourg, Sarrebourg 57400, France

⁴ University Hospital of Oran, Oran 31000, Algeria

Abstract. Training has a great place in ergonomics. Through training, ergonomics can increase the capabilities of individuals and their potentials to be able to adapt positively to the work place and to maximize the effectiveness of the work components (machines, work systems, tools, etc.) through training. One of the well-known ergonomic training methods is the simulation. It is powerful, as it combines the advantages of out-of-work training methods with the positives of on-the-job training methods.

The study's goal was to show how well a simulation-based training program might avoid WMSDs and help emergency and intensive care nursing staff members strengthen their motor skills. To collect the data, the Rapid Entire Body Assessment method [1] was applied on a sample of ten nurses (age $X = 43$ years, professional experience $X = 7$ years). Before training intervention, results showed that the majority of the nurses in the sample had postures that were characterized as having a high degree of WMSDs risk and necessitating an immediate intervention. However, after training intervention, results demonstrated the efficiency of the suggested training program to lower WMSDs risks and to improve the nurses' motor skills. Hence, to be efficient, the medical and healthcare training programs should incorporate ergonomics principles.

Enhancing the learning process by strengthening the acquired ergonomics, best practices during the training and eliminating the residual behaviors of the wrong acquired practices of patient handling, was put forth as a recommendation.

Keywords: Patients Handling Tasks · Caregiving Staff · Training Program · WMSDs

1 Introduction

Patients' load handling tasks are among the most strenuous tasks that are characterized by high level of risk for caregivers and patients alike, as patients are in a pain suffering state, they are bulky, have no handles, cannot always be held close to the body and can

be unpredictable [2]. One of the well documented hazards in healthcare giving sector is the prevalence of work musculoskeletal disorders (WMSDs) among health workers, especially nursing staff [3, 4].

Nursing duties like cleaning patients, wounds' treatment, dressing change, changing linen, helping patient for their morning walks and physical movements, etc., are among the daily caregiving duties which involve many patients handling tasks, like: Standing and sitting the patient, Sitting the patient to the edge of the bed, Walking with the patient, Rolling the patient in bed, etc.

Such tasks are physically demanding, involving awkward postural situations and injury risks for caregivers and patients alike, particularly in the absence or lack of the appropriate patient handling equipment, like Hoists, Slings, Transfer boards, Bath and shower equipment, Bed and chair equipment, Handling belts, Slide sheets, ambulance stretchers. Furthermore, some of the existing equipment are not well manipulated/adjusted to the required patient-handling task by the nursing staff.

Absence of ergonomics notions on the prevention of WMSDs, as a part of healthcare givers educational programs, particularly nursing staff, is a major hurdle in preventing WMSDs in health care professions. In the absence of such elementary requirements to prevent WMSDs among healthcare giving staff, and the lack of research based evidences, patient handling techniques are based on opinion rather than scientific evidence, making way to the use of traditional or controversial techniques, in spite of recommendations against these methods, as has been advanced by Hignett et al.(2007) [5].

To tackle the problem of WMSDs among patient handling nursing staff, interventions reported in the literature [4] recommend the implementation of four complementary solutions strategy (i) training concerned staff members for patient handling technics (ii) providing specialized equipment (iii) implementing policies and procedures to reduce work-related musculoskeletal disorders and (iv) support and follow up the strategy.

2 Aims and Questions

The study's goal was to show how well a simulation-based training program might avoid WMSDs and help emergency and intensive care nursing staff members strengthen their motor skills. The main questions are:

- Are the program's simulation elements adequate for a preventative approach of WMSDs?
- Does the training program help nursing staff members develop the necessary motor skills?

3 Method

3.1 Sample

Ten nurses working in intensive care unit at the university hospital of Oran, Algeria, were used as the sample for the research. The nurses were 43 years old on average (SD = 4.3 years). They have an average of 7 years of professional experience (SD = 2.4). They all took part in a training regimen designed to stave off work musculoskeletal disorders.

3.2 Tool and Procedures

The Rapid Entire Body Assessment (REBA) method [1] was used to gather and analyze the data during two stages: the pre-training stage, when a training program was conceived, and the post-training stage, when the training program was put into practice and evaluated. At the end of the training.

program, trainees were asked (1) if the training was useful for the development of their patients' handling skills, (2) and if it helped reducing WMSDs complaints.

3.3 Training Program

A training program is conceived to train a sample of 10 nursing staff members on ergonomics best practices and technics of patients handling tasks, which were inspired from relevant ergonomics and occupational health and safety sources [2, 6, 7]. Patients were in a mixed of fully and partially dependent states. The training protocol was carried out during the morning shift when most of the daily caregiving duties, which involve patients handling tasks, took place. The program was a mixed of simulated and realistic context situations, which consisted of two parts, a theoretical part (1/3rd of the training time), conducted in two morning sessions, and a practical part (2/3rd of the training time) conducted along four sessions of two to three hours each.

The theoretical part of the training program consisted of two PowerPoint Presentations on ergonomics principals and guideline of safe patients handling methods, technics and equipment, with reference to the established materiel [2, 8], followed by discussion and further detailed explanations.

The practical part the training program was conducted along four sessions of two to three hours each.

Session 1: Two ergonomists (searchers) took part in practical demonstrations on how to carry out patient's handling in the following ten situations, which are daily faced by the nursing staff, especially during their morning shift:

Standing and sitting the patient	Moving the patient in bed
Sitting the patient to the edge of the bed	Sitting to sitting transfers
Walking with the patient	Repositioning the patient in a chair
Rolling the patient in bed	Assisting a fallen patient
Sitting the patient up in bed	Moving a patient from laying surface

Session 2: A peer training session, in which one trainee acted as a patient and the other trainee acted as a trainer, carried out patient's handling exercise of the above-mentioned ten situations, under the supervision of two searchers.

Session 3: During the following day, a training evaluation session (post-training test) took place in real circumstances, in the intensive care unit of the hospital, during the whole morning shift, with a mean time of 8 hours, and a brake time of one hour at launch time, and a few sporadic pauses of 10 to 15 minutes each. Most patients handling tasks

carried out by the caregivers of the study covered all the above mentioned ten situations, as intensive care nursing duties are usually required during morning shift, after a long night for patients in bed.

Session 4: Was organized during the last day of the training program, in which all ten nurses who participated in the training program were present, besides the participation of two members of the managerial staff of the hospital in which the training took place, who closely followed the training and were eagerly interested in its outcome. After having a feedback from the trainees on the outcome of the training program, if it was useful for the development of their patients’ handling skills, and if it helped reducing their WMSDs complaints, a discussion was launched among participants.

4 Results

The pre-training tests (Table 1) showed that, the majority of the nurses in the sample adopted postures that are characterized, as having high levels of WMSDs risk, and necessitating an immediate intervention action. While the post-training test (Table 2) showed, a remarkable improvement towards adopting correct postures in performing the above mentioned ten patients’ handling tasks in real work situations, which is a positive indicator. On the other hand, a feedback from the trainees, at the end of the training, revealed a consensus among trainees that the training program did enhance their patients’ handling skills and reduced their WMSDs complaints.

Table 1. Risk levels and required action to prevent WMSDs in ten patients handling tasks among nursing staff of intensive care unit (the pre-training REBA scores).

Tasks	Action level	Risk level	Action (including further assessment)
1. Standing and sitting the patient	3	High	Necessary soon
2. Sitting the patient to the edge of the bed	3	High	Necessary soon
3. Walking with the patient	3	High	Necessary soon
4. Repositioning the patient in a chair	3	High	Necessary soon
5. Rolling the patient in bed	4	Very High	Necessary now
6. Sitting the patient up in bed	4	Very High	Necessary now
7. Moving the patient in bed	4	Very High	Necessary now
8. Sitting to sitting transfers	4	Very High	Necessary now
9. Assisting a fallen patient	4	Very High	Necessary now
10. Moving a patient from one laying surface to another	4	Very High	Necessary now

Table 2. Risk levels and required action to prevent WMSDs in ten patients handling tasks among nursing staff of intensive care unit (the post-training REBA scores).

Tasks	Action level	Risk level	Action (including further assessment)
1. Standing and sitting the patient	1	Low	May be Necessary
2. Sitting the patient to the edge of the bed	1	Low	May be Necessary
3. Walking with the patient	1	Low	May be Necessary
4. Repositioning the patient in a chair	1	Low	May be Necessary
5. Rolling the patient in bed	2	Medium	Necessary
6. Sitting the patient up in bed	2	Medium	Necessary
7. Moving the patient in bed	2	Medium	Necessary
8. Sitting to sitting transfers	2	Medium	Necessary
9. Assisting a fallen patient	2	Medium	Necessary
10. Moving a patient from one laying surface to another	2	Medium	Necessary

5 Discussion

It is well established that patient handling training enhances skills and prevents WMSDs among caregiving staff [4, 5] and reduces the incidence of caregivers injury and improve patient outcomes. Training can significantly reduce stressful trunk postures, by raising awareness of the physical strains that frequently occur during a shift, by changes in work practices and by redesigning the work environment, as has been pointed out by Nourollahia et al. (2018) [9]. The present patient handling training program is in line with these findings.

In nursing education, simulation gives students the chance to hone their abilities in a secure setting while being watched over by a supervisor. They practice common nursing activities in a hospital-like setting using the most advanced mannequins before using these abilities in actual clinical settings.

Disseminating ergonomics and human factors practices and principles in healthcare settings was one of the goals realized by the present intervention. Participants in the training program, beside the management representatives who followed the training sessions and participated in discussions, were convinced of the usefulness and cost-effectiveness of ergonomics interventions, a conclusion, which is well established throughout the ergonomics human factors community.

Although, the apparent realization of the training program goals, its limitation are multiple. Acquiring new skills and knowledge is a continuous process. A five days training program, cannot easily change all the long-term wrong practices. That trainees of caregiving are accustomed to, and became a part of their daily behavior, like the adoption of awkward postures, lifting patient's in the absence of the appropriate equipment, or the absence of "how to lift training", though, many researchers question its effectiveness in the prevention of WMSDs [3].

These practices are anchored, well rooted and became reflexes in the day to work activities. To replaces them by the correct ergonomics practices, some actions should take place. One of which is to implement a collective work organization of two-person teams to realize patient handling tasks, especially in the absence/inadequacy of handling

equipment. Enhancing the learning process by strengthening the acquired ergonomics, best practices during the training and eliminating the residual behaviors of the wrong acquired practices of patient handling.

6 Conclusion

This study illustrated the significant contribution that simulation may make to ergonomics. A well-conceived training program is deemed to attain its intended goals, through a continuous process of evaluation, readjustment and reinforcement of learned skills. The new best-practiced skills learned through training, should replace wrong/unwanted skills, which were acquired throughout time, by the use of traditional and controversial techniques.

References

1. Hignett, S., McAtamney, L.: Rapid entire body assessment (REBA). *Appl. Ergon.* **31**(2000), 201–205 (2000). [https://doi.org/10.1016/S0003-6870\(99\)00039-3](https://doi.org/10.1016/S0003-6870(99)00039-3)
2. ISO 11228-1:2021, Ergonomics-Manual handling-Part 1: Lifting, lowering and carrying. <https://www.iso.org/standard/76820.html> (2021). (retrieved: 11/02/2024)
3. Oakman, J., Lambert, K.A., Rogerson, S., Bell, A.: We know it doesn't work: Why do we still use how to lift training for the prevention of musculoskeletal disorders? *Int. J. Ind. Ergon.* **100**(2024), 103542 (2024). <https://doi.org/10.1016/j.ergon.2023.103542>
4. Asuquo, E.G., Murphy Tighe, S., Bradshaw, C.: Interventions to reduce work-related musculoskeletal disorders among healthcare staff in nursing homes; An integrative literature review. *Int. J. Nurs. Stud. Adv.* **3**, 100033 (2021). <https://doi.org/10.1016/j.ijnsa.2021.100033>
5. Hignett, S., et al.: Implementation of the manual handling directive in the healthcare industry in the European Union for patient handling tasks. *Int. J. Ind. Ergon.* **37**, 415–423 (2007). <https://doi.org/10.1016/j.ergon.2007.01.003>
6. EU-OSHA: E-fact 28 - Patient handling techniques to prevent MSDs in health care. EU-OSHA – European Agency for Safety and Health at Work. Available from: <https://osha.europa.eu/en/publications/e-fact-28-patient-handling-techniques-prevent-msds-health-care> (2008)
7. Salas, E., Burke, C.S.: Simulation for training is effective when *Qual. Saf. Health Care.* **11**(2), 119–120 (2002) <https://10.1136/qhc.11.2.119>
8. OSHA 3182-3R: Guidelines for Nursing Homes Ergonomics for the Prevention of Musculoskeletal Disorders. Available from: https://www.osha.gov/sites/default/files/publications/final_nh_guidelines.pdf (2009)
9. Nouroollahia, M., Afsharib, D., Dianat, I.: Awkward trunk postures and their relationship with low back pain in hospital nurses. *Work.* **59**, 317–323 (2018). <https://doi.org/10.3233/WOR-182683>



An Anthropotechnological Analysis of Training Innovations in PROSUB: The Role of Computer-Based Training (CBT) in Achieving Operational Readiness for Brazil's Riachuelo-Class Submarine Crews

Henrique Ribeiro Menezes¹ , Júlio César Bispo Neves² ,
and Mario Cesar Rodriguez Vidal¹ 

¹ Federal University of Rio de Janeiro (COPPE), Rio de Janeiro, RJ, Brazil

² Federal Fluminense University (TEP), Niteroi, Brazil

juliobispo@id.uff.br

Abstract. This study investigates strategies to optimize knowledge retention and operational performance among Brazil's Riachuelo-Class Submarines (RCS) crews during the France-Brazil technology transfer under PROSUB. Using a Grounded Theory approach, we analyzed semi-structured interviews with submariners and instructors, complemented by anthropotechnological frameworks, to evaluate how all training instances mitigate skill gaps, enhance adaptive capacity in high-complexity naval environments, and ensure effective technology transfer. Results identified four key categories: (1) Cross-Cultural Communication Barriers impacting knowledge transmission, (2) Technical Capability Building via immersive simulations, (3) Organizational Adaptation to new roles and technologies, and (4) Technological Trust in Computer-Based Training (CBT) systems. Findings underscore CBT's critical role in maintaining crew readiness amid evolving technological and operational demands.

Keywords: Ergonomics · training · anthropotechnology · simulation · Grounded Theory

1 Introduction

Brazil's maritime territory, critical to national sovereignty and economic activity (accounting for 80% of export-import trade), demands advanced naval capabilities. The PROSUB initiative addresses this through strategic partnerships, such as the Brazil-France submarine agreement, which prioritizes crew training as a cornerstone of technological sovereignty.

The overall objective of the research is to present a framework that identifies factors influencing RCS Crew performance based on insights from relevant focus groups. In parallel, the specific objectives are to inventory technical resources available in the

defined context, explain knowledge gaps throughout the initial training phase, and deepen learning regarding innovation in the specific phase.

This study aims to establish solid ergonomic modeling to facilitate structured diagnoses. To that end, it follows the methodological flow presented by Vidal [1] and focuses on anthropotechnological concepts taught by Wisner [2].

Building upon the preliminary findings of Menezes, Neves, and Vidal [3], this paper extends the analysis by addressing emergent challenges in simulator-based training, knowledge, and new technology transfer, including the four new core categories that emerged from the analyzed data.

2 Ergonomics and Training in Technology Transfer

Ergonomics adapts work processes to human capabilities to improve efficiency, safety, and well-being. In training the RCS Crew, this discipline is essential to drive positive change and improve the interaction between the team and the technology and organizational processes during technology transfer.

Wisner [2] shows, through studies in anthropotechnology, some origins of the failures, partial or total, of some technology transfer experiences, highlighting, for example, insufficient staff training and difficulty in the form of conversation and understanding [4]. In the present study, within the scope of the technology transfer agreement between Brazil and France, some original conditions similar to those pointed out by Wisner were listed, such as difficulty in communication and understanding due to language barriers and cultural differences related to professional education.

In the scenario of indispensable continuity in the Technological Transfer process, the ability to absorb technology results from managerial and technical capital accumulated over years and various processes, not undone by the disaggregation of teams or loss of technical memory [5].

This work aims to invest in the search for suggestions for maintaining this knowledge acquired in technological absorption during the RCS Crew's training course, highlighting practices in simulators. Computer-Based Training (CBT), one of the simulation phases in the course, is a type of E-learning that uses computers to provide instructional content, which deserves full emphasis in this research since CBT consists of an effective concept of retaining knowledge, providing updates, and improvement training.

3 Methodology

The Grounded Theory (GT) methodology was applied with instrumentation from an anthropotechnological assessment. GT combines a set of steps to achieve theorization with an explicit approach [5]. According to Birks and Mills [6], GT encompasses a series of essential methods such as initial coding and categorization of data; concurrent data generation or collection and analysis; writing memos; theoretical sampling; constant comparative analysis using inductive and abductive logic; theoretical sensitivity; intermediate coding; identifying a core category; and advanced coding and theoretical integration.

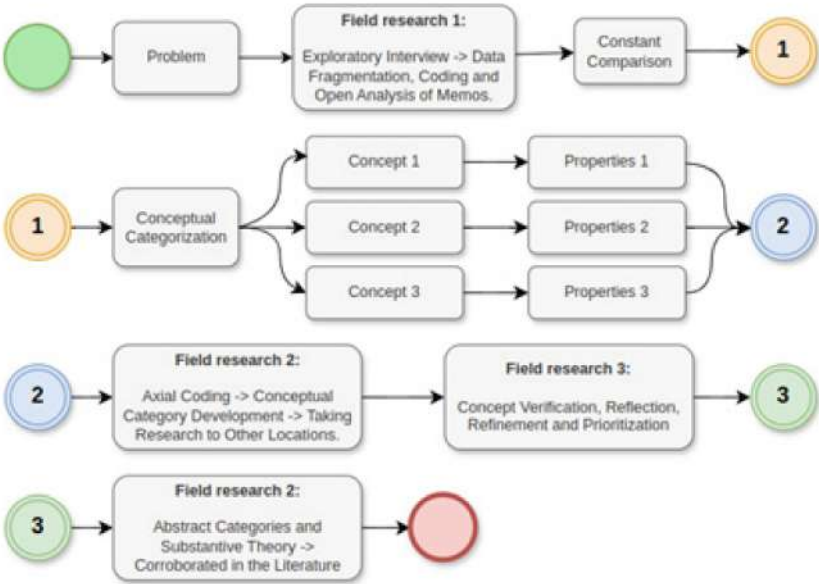


Fig. 1. Flow adopted for applying Grounded Theory and defining the substantive theory.

Table 1. Example of the initial coding.

Interviewed - I(n)	Raw Data (transcript)	Initial Coding (line by line)
101	The military goes on board to make fine adjustments but will be ready if he does a good simulator stage.	Informing that CBT leaves the military ready.
103	At the end of the course, we had the opportunity to put submarine sailing into practice.	Finishing with submarine training at sea.
104	The operations had little contact with the systems simulator (CBT).	Informing the degree of contact between the Op. and the CBT.
105	One of the difficulties, initially, was the language barrier because they were not yet used to the Portuguese language.	Identifying that the initial barrier was communication.

We opted for the flow shown in Fig. 1, in which the theory evolves during the research process, shaped by the interaction between analysis and data collection (see Table 1). After constant comparison, the collected data was constantly compared and organized, reduced, and grouped into meaningful categories. The main category obtained

and aligned with the purpose of the research was ‘Simulator Training’, as well as other emerging categories that will be discussed in the final results.

After completing this categorization stage, a theoretical model was obtained, with the appropriation of knowledge, guidance for future actions, and strengthening of simulator instruction, especially CBT, essential to keeping the RCS Crew qualified.

The methodology applied allowed us to identify the main category and its existing relationships with the other emerging categories (see Table 2).

Table 2. Example of the theoretical coding process.

Interviewed - I(n)	Raw Data (transcript)	Initial Coding (line by line)	Focused Coding	Theoretical Coding
101	The military goes on board to make fine adjustments but will be ready if he does a good simulator stage.	Informing that CBT leaves the military ready.	Guarantee of Qualification in CBT	Simulator Training
103	At the end of the course, we had the opportunity to put submarine sailing into practice.	Finishing with submarine training at sea.	Ease of training on board	Technical capability building
104	The operations had little contact with the systems simulator (CBT).	Informing the degree of contact between the Op. and the CBT.	CBT few available	Simulator Training
105	One of the difficulties, initially, was the language barrier because they were not yet used to the Portuguese language.	Identifying that the initial barrier was communication.	Difficulty in communicating with instructors	Transmission of knowledge

4 Results

Data was collected from the Crew’s experiences of qualified submariners in the RCS and their perceptions about technology transfer in the PROSUB. As a result of the methodological path adopted, the central category, “Simulator Training,” highlighted diving, tactical, and especially CBT simulators. Trainees retain knowledge of the new class of submarine, which is essential during class change, and CBT minimizes gaps in specialized knowledge (see Fig. 2). Four emerging categories related to “Simulator Training” were obtained:

- a) **Transmission of Knowledge:** This category considered the instruction, communication, and didactics of French instructors (former military personnel trained in Conventional Submarine Classes).
- b) **Technical Capability Building:** Related to Crew qualification, directly linked to technological transfer and explaining typical onboard situations. It consists of prior training of Brazilians based on the history of the current class of submarines, which optimizes operational technology transfer.
- c) **Technical-Organizational Innovation:** Built by the technological differences of transitioning from the Tupi Class Submarine (TCS) to RCS and the beginning of CBT. Organizational needs were transformed and adapted throughout the process, teams were reconfigured, and new functions were created.
- d) **Effect on Final Training:** This category refers to the changes or impacts the training program produces on the Crew and instruction. At the end of this category, we have the new concept of CBT training, the interaction of enlisted personnel when traveling, the impact of control technology, and systemic analysis of the target's movement.

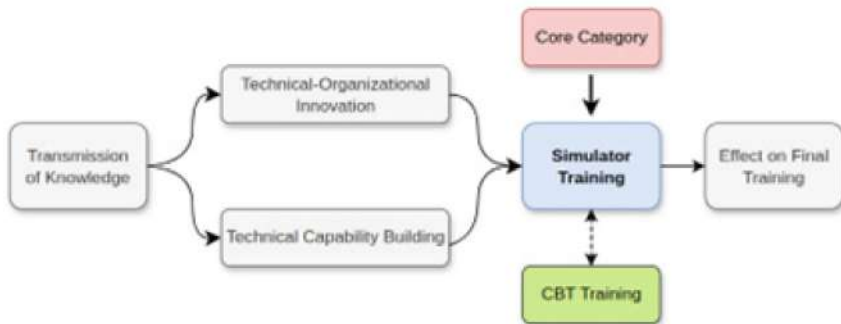


Fig. 2. Substantive theory generated from data analysis with Grounded Theory - the basis for consolidating ergonomics demand.

5 Discussion

Identifying patterns in the Crew's perception of technological transfer processes allowed us to understand the discursive movement and the order-disorder-interactions-organization tetralogy, including retroactions, recursions, self-organization, concepts, attitudes, beliefs, and experiences.

The emerging categories identified several gaps in the training process, emphasizing the importance to the role of Computer-Based Training (CBT) to maintain the Crew's operational readiness which an advantage of CBT over traditional methods is that it provides effective preparation for crew members because it offers scalable learning opportunities, reducing procedural errors in a virtual environment.

6 Conclusions

This study addressed the main challenges in technology transfer under PROSUB, including training, technology, and organization, with crew training being a central theme of the program. In the context investigated, the availability of technical resources provided various opportunities to strengthen management initiatives to improve instruction in simulators. In this work, simulator training involves conducting more frequent and targeted sessions for crew members using computer-based learning (CBT) to ensure the retention of knowledge and enhancing dexterity in responding to incidents in a range from minor to severe damage.

Identifying the Crew's prior knowledge deficiencies is fundamental to improving the process during the initial technology transfer phase. In parallel, adopting human factors and ergonomics concepts regarding organizational and cognitive aspects and considering anthropotechnological "warnings" that could facilitate the technological transfer process is also a future suggestion.

These actions could deepen the Naval Force's understanding and improve the mastery of technological innovations during the technological transfer process in current and future cooperation agreements.

References

1. Vidal, M.C.: Guia para análise ergonômica do trabalho na empresa (AET): Uma metodologia realista, ordenada e sistemática. Editora Virtual Científica, Rio de Janeiro (2003)
2. Wisner, A.: The irruption of new technologies: A new challenge for ergonomics and anthropotechnology. *J. Hum. Ergol. (Tokyo)*. **24**(1), 45–55 (1995 Jun) ISSN: 03008134
3. MENEZES, H.R., NEVES, J.C.B., Vidal, M.C.: Technology transfer with an operational focus on ergonomic demand: Training of the Riachuelo class crew of the Brazilian Navy. *Rev. Ação Ergon.* **17**(2), e202302 (2023. ISSN 2965-7318). <https://doi.org/10.4322/rae.v17n2.e202318.en>
4. Masculo, F.S., Vidal, M.C.: Ergonomia: Trabalho adequado e eficiente. Elsevier, Brasil (2013)
5. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publishing Company, Chicago (1967)
6. BIRKS, M., MILLS, J.: Grounded Theory: A Practical Guide, 2nd edn. SAGE, Londres (2015) ISBN 978-1-4462-9577-9



Development of Integrated Ergonomics Intervention for Employees with Signs and Symptoms of Musculoskeletal Disorders: Health Belief Model and ADDIE as Valuable Tools

Nor Haslina Mohd^{1,2}(✉), Anita Abdul Rahman³, Shamsul Tamrin², Hezery Harun⁴, and Ng Yee Guan²

¹ Environmental and Occupational Health Program, School of Health Sciences, Universiti Sains, Kubang Kerian, Malaysia
nhaslina@usm.my

² Department of Environmental and Occupational Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

³ Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

⁴ Department of Orthopedic Surgery, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

Abstract. Background: To date, the burden of disabling musculoskeletal pain and injuries (musculoskeletal disorders, MSDs) arising from work-related causes in many workplaces remains substantial. Quay crane operators were no exception from sustaining MSD. However, no structured protocol is available to identify and apply an intervention for quay crane operators' problems at the workplace. This paper describes the structured development, implementation, and planning for evaluating an ergonomics intervention for quay crane operators with the signs and symptoms of MSD. The ergonomics intervention was developed using the health belief model (HBM) and ADDIE model as a framework.

Methods: The principles of HBM and ADDIE were applied to combine theory and evidence in the development, implementation, and planning of the evaluation of an integrated ergonomics workplace intervention aimed at reducing the level and intensity of signs and symptoms of MSD

Discussion: The development of the ergonomics intervention according to the HBM and ADDIE principles resulted in a structured ergonomics intervention tailored explicitly to employees with signs and symptoms of MSD. Increased scores in knowledge, attitude, and practice (KAP) and decreased VAS pain scores were proposed as a result of behavioral change.

Conclusion: The HBM and ADDIE frameworks are promising methods for developing interventions tailored to a specific target group in the field of occupational health.

Keywords: Ergonomics Intervention · Health Belief Model · ADDIE model

1 Introduction

Musculoskeletal Disorders (MSDs) significantly influence workers' health, welfare, and productivity (Crawford et al. 2020). These disorders impact muscles, tendons, ligaments, nerves, and other soft tissues, resulting in pain, discomfort, and limitations in movement. Employees afflicted with MSDs experience a diminished quality of life and increased absenteeism, which in turn affects efficiency and productivity.

Managing MSDs is vital for maintaining a healthy workforce and the depreciation of adverse effects. Effective management of MSDs involves strategies such as risk assessment (Department of Occupational Safety and Health, Ministry Of Human Resource et al. 2017), ergonomic design (Karwowski 2005), training (Denadai et al. 2021), and the utilization of assistive devices (Haque and Kumar 2015). Comprehending workplace dynamics and employee needs is crucial for successfully preventing and enhancing productivity.

Ergonomics, a well-established approach for managing MSDs (Wilson 2000), focuses on enhancing the compatibility between individuals and their work environment. Ergonomics aims to prevent discomfort, fatigue, and injuries from repetitive tasks by considering factors such as posture, movement, and physical stress. Appropriate ergonomics can mitigate the risk of developing MSDs, and ergonomic interventions can alleviate symptoms (Punnett et al. 2009). It is essential to consider both workplace and individual factors when managing MSDs. Consequently, some studies suggest that while ergonomics intervention alone may not always be significantly effective, combining ergonomics with behavior change is recommended to enhance effectiveness (Angus et al. 2013).

The Health Belief Model (HBM) facilitates behavior changes through interventions (Saunders et al. 2013). The HBM posits that individuals are more likely to adopt health-related behaviors if they perceive themselves as susceptible to a health threat and recognize the benefits of taking specific actions. Applying HBM principles empowers individuals to make informed decisions for improved health, such as musculoskeletal health.

Additionally, researchers are considering the application of the ADDIE model in designing interventions. This model, which stands for Analyse, Design, Develop, Implement, and Evaluation (ADDIE), provides a systematic approach to developing and implementing effective learning programs (Robert Maribe Branch 2009). The ADDIE model promotes a comprehensive view of program development, ensuring that all aspects, from analysis to evaluation, are considered. Adhering to this structured approach increases the likelihood of successfully managing MSDs within an organization.

The HBM focuses on personal perceptions, beliefs, and motivations related to health behaviors, while the ADDIE model provides a systematic framework for designing, implementing, and evaluating interventions. This comprehensive approach combines individual motivation (HBM) with systematic program design (ADDIE) to enhance the effectiveness of interventions in managing health issues.

This paper proposes the development of an ergonomics intervention utilizing the HBM and ADDIE models.

In the conducted research, investigators carried out three distinct types of analyses: a needs analysis, an individual analysis, and a contextual analysis. These analyses assessed

the constructs of the HBM, including perceived threats, perceived barriers, perceived benefits, cues to action, and self-efficacy.

Needs Analysis: This analysis delineated and cataloged all the challenges or problems. The objective was to comprehend the gaps and issues necessitating resolution.

Individual Analysis: This analysis scrutinized the individuals' characteristics, background, and other pertinent factors. It also identified their extant knowledge and skills. These factors could influence their learning experience, and the findings would assist in customizing instruction to their needs.

Contextual Analysis: This analysis evaluated any constraints or limitations in the organization and workplace environment. It considered all available training resources, such as technology, materials, and personnel. It also identified external factors that could impact the training program, like industry trends and legal stipulations.

2 Method

2.1 Step 1: Analysis

Before establishing a program's objectives, it is paramount to comprehend the context, encompassing the participants and their milieu. This comprehension serves as the foundation for the subsequent phase in the ADDIE model, which ensures the alignment of the instructional design with the needs of the individuals and the organization's objectives. During this phase, the intervention's broad and specific goals were ascertained.

2.2 Step 2: Designing

In the subsequent phase, a comprehensive plan was formulated for the entire program, encompassing the sequence of topics and activities.

The selection of an appropriate intervention is of paramount importance. It depends on the participants' age, qualifications, and job characteristics. This ensures the effective transfer of knowledge. The intervention could take the form of workshops, e-learning modules, or coaching sessions.

The intervention was devised based on the subject matter and the principles of the HBM. The materials were crafted considering individuals' perceived risk, severity, and benefits. The content revolved around MSD, comprehension of the nature of work, risk factors, prevention through ergonomic practices, and awareness and improvement of posture. The HBM constructs in the analysis employ several tools to assess knowledge, severity, and other elements. These tools are also utilized post-intervention to ascertain any changes and to measure the success of the intervention.

2.3 Step 3: Develop Content

In the third step, the actual materials were developed. These resources encompassed a variety of formats, such as videos, slides, scripts, handouts, interactive exercises, and

modules. They were made accessible in digital and physical formats to facilitate behavioral modification. The content was formulated based on the analyzed HBM constructs. It encapsulated ergonomic practices, stretching and strengthening exercises, posture enhancement, emphasis on benefits, reduction of barriers, and implementation strategies.

After the development process, the materials underwent rigorous testing and refinement. A preliminary test was conducted with a select group, feedback was solicited, and necessary modifications were implemented. This ensured that the materials were not only accurate and precise but also congruent with the objectives of the program.

2.4 Step 4: Implementation

Upon the development of the MSD management program, it was disseminated across the organization. In the initial phase, data and feedback were systematically gathered to scrutinize the participants' engagement with the program and adherence to the preventive measures. This feedback served as a critical input for the iterative refinement of the program as required.

2.5 Step 5: Evaluation

The fifth step, the program was evaluated from two perspectives. The first perspective, formative evaluation, examined whether participants applied what they learned. The second perspective, summative evaluation, assessed the program's effectiveness for the participants. These evaluations helped to improve the program continuously.

In this stage, it was also checked whether the participants understood the severity and Susceptibility of the disease, the benefits of prevention, how to overcome barriers, and how to take preventive actions. The program's overall impact was also assessed, including changes in the prevalence of MSD, pain levels, knowledge levels, attitudes, and practices. This assessment was then used for the next step by the employer.

3 Discussion

The theoretical framework, in conjunction with the analytical process, aids in identifying pivotal components and strategies pertinent to the population under study. This structured methodology paves the way for efficacious interventions, laying a robust design foundation. Theories play a crucial role in enabling researchers to comprehend the underlying mechanisms and forecast the impact of interventions, thereby ensuring an approach rooted in evidence. This process amplifies the probability of successful management of Musculoskeletal Disorders (MSD).

The intervention delineated in this study was conceptualized to effect behavioral modifications. Research substantiates that such behavioral alterations can augment the effectiveness of MSD management. Incremental changes in behavior culminate in superior outcomes.

Given the recurrent nature of MSDs, focusing on behavioral modifications can mitigate the risk of recurrence. Such changes can also bolster workplace ergonomics and posture awareness, alleviating muscle and joint tension. The cognitive-behavioral approach

equips individuals to manage pain, adhere to guidelines, and execute self-treatments. This ensures patients' compliance with treatment plans and exercise routines, indirectly enhancing their physical function.

This paper's combination of theory and model has successfully interpreted participant needs and identified prerequisites. Considerations have been made for participant needs such as knowledge, skills, and training pertinent to existing MSD management and prevention.

In the preliminary step, the analysis furnished insights into the current state of the population under study. Based on this information, the primary objective was delineated, which is to curtail the level of MSD.

Furthermore, the analysis provided insights into all Health Belief Model (HBM) constructs that were intended to be identified. Comprehending the risks concomitant with MSDs is of paramount importance. The process of enlightening individuals about the prevalence, etiology, and repercussions of MSDs can shape their perceived susceptibility. When individuals personally contend with an MSD, experiencing pain, discomfort, or work-related limitations, they are more inclined to perceive themselves as susceptible. Even in the absence of an MSD, comparing themselves with others who have endured an MSD can also influence perceived susceptibility. The exposure to information disseminated through media, healthcare providers, or peers can mold the perception of vulnerability.

In the context of MSDs and the interventions to be administered for managing MSDs in this demographic, it is essential to deliberate on the risk factors. This entails providing information on the risk factors inherent in this occupation and strategies to address these risk factors. In addition to risk factors, it is crucial to communicate the symptoms that are likely to manifest so that individuals are equipped with the knowledge to manage and prevent them.

References

- Angus, K., Cairns, G., Purves, R., Stuart, B., MacDonald, L., Gordon, R.: Systematic Literature Review to Examine the Evidence for the Effectiveness of Interventions that use Theories and Models of Behaviour Change: Towards the Prevention and Control of Communicable Diseases. ECDC, Stockholm (2013) <http://ecdc.europa.eu/en/publications/Publications/health-communication-behaviour-change-literature-review.pdf>
- Åström, M., Thet Lwin, Z.M., Teni, F.S., Burström, K., Berg, J.: Use of the visual analogue scale for health state valuation: A scoping review. In Qual. Life Res. **32**(10), 2719–2729. Springer Science and Business Media Deutschland GmbH (2023). <https://doi.org/10.1007/s11136-023-03411-3>
- Branch, R.M.: *Instructional Design: The ADDIE Approach*. Springer, New York, NY (2009)
- Crawford, J.O., et al.: Musculoskeletal health in the workplace. Best Pract. Res. Clin. Rheumatol. **34**(5), 101558 (2020). <https://doi.org/10.1016/j.berh.2020.101558>
- Denadai, M.S., Alouche, S.R., Valentim, D.P., Padula, R.S.: An ergonomics educational training program to prevent work-related musculoskeletal disorders to novice and experienced workers in the poultry processing industry: A quasi-experimental study. Appl. Ergon. **90**(July 2020), 103234 (2021). <https://doi.org/10.1016/j.apergo.2020.103234>
- Department of Occupational Safety and Health, Ministry of Human Resource, M., Malaysia, D. of O. S. A. H. M. O. H. R., Department of Occupational Safety and Health, Ministry of

- Human Resource, M., & Malaysia, D. of O. S. A. H. M. O. H. R: Guidelines On Ergonomics Risk Assessment At Workplace. <http://www.dosh.gov.my/index.php/en/competent-person-form/occupational-health/guidelines/ergonomic/2621-01-guidelines-on-ergonomics-risk-assessment-at-workplace-2017/file> (2017).
- Haque, M.S., Kumar, M.: Ergonomics, biomechanics & musculoskeletal disorder- A review. *Int. J. Electr. Electron. Eng.* **07**(December), 132–138 (2015)
- Karwowski, W.: Ergonomics and human factors: The paradigms for science, engineering, design, technology and management of human-compatible systems. *Ergonomics*. **48**(5), 436–463 (2005). <https://doi.org/10.1080/00140130400029167>
- Punnett, L., Cherniack, M., Henning, R., Morse, T., Faghri, P.: A conceptual framework for integrating workplace health promotion and occupational ergonomics programs. *Public Health Rep. (Washington, D.C. : 1974)*. **124**(4_suppl1), 16–25 (2009). <https://doi.org/10.1177/00333549091244S103>
- Robert Maribe Branch: Instructional Design: The ADDIE Approach (2009)
- Saunders, G.H., Frederick, M.T., Silverman, S., Papesh, M.: Application of the health belief model: Development of the hearing beliefs questionnaire (HBQ) and its associations with hearing health behaviors. *Int. J. Audiol.* **52**(8), 558–567 (2013). <https://doi.org/10.3109/14992027.2013.791030>
- Shariat, A., Tamrin, S.B.M., Arumugam, M., Ramasamy, R.: The Bahasa Melayu version of Cornell musculoskeletal discomfort questionnaire (CMDQ): Reliability and validity study in Malaysia. *Work*. **54**(1), 171–178 (2016). <https://doi.org/10.3233/WOR-162269>
- Wilson, J.R.: Fundamentals of ergonomics in theory and practice. *Appl. Ergon.* **31**(6), 557–567 (2000). [https://doi.org/10.1016/S0003-6870\(00\)00034-X](https://doi.org/10.1016/S0003-6870(00)00034-X)



Methods of Assessment of Students in Active Learning Strategies in Occupational Health and Safety Teaching in Engineering

Italo Rodeghiero Neto^(✉) , Thiago Arndt Schimit ,
Lucas Gomes Miranda Bispo , and Fernando Gonçalves Amaral 

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
italorneto@gmail.com, thiago.arndt.schimit@gmail.com,
lucasbispoprod@gmail.com, amaral@producao.ufrgs.br

Abstract. The teaching of Occupational Health and Safety (OHS) requires modern teaching methodologies to assist in analyzing and understanding issues within the professional scope of future practitioners. Throughout the application of these strategies, it is necessary to assess how measure the skills acquired by the student described in the literature. This study investigates the forms of knowledge assessment that should be applied when active learning strategies are used in this field. A narrative literature review identified five primary forms of student knowledge assessment: learning styles, students' perception of learning, pre-and post-tests, existing tools and methodologies, and the creation of assessment instruments. The assessment methods vary from questionnaires for quantitative analysis to interviews for qualitative analysis. Additionally, approaches to evaluating students' learning styles were observed, such as Kolb's Experiment, the Index of Learning Styles (ILS), and Myers-Briggs indicators. Although gaps were identified in recommending specific forms of competence assessment in OHS within engineering, OHS curriculum components in various engineering disciplines underscore the importance of future research in this combined area with education.

Keywords: Assessment methods · active learning · OHS · teaching · engineering

1 Introduction

Expository and traditional approaches still prevail in engineering education, particularly in universities in emerging countries [1]. There is a need for more studies that aim to integrate Occupational Health and Safety (OHS) with different teaching methods, such as active learning strategies, which makes it challenging for professors to implement and modernize their classes [2]. However, it is already established in the literature that active learning strategies are the most recommended in engineering education [3].

Active learning fosters the development of a range of essential technical competencies, ensuring long-term knowledge retention. It also guarantees the development of transversal and professional skills in students, such as teamwork, ethics, social responsibility, communication, and problem-solving [4, 5]. Despite recognizing the benefits of

active learning in OHS education, studies have been superficially applied, focusing on well-known strategies [6]. This does not allow for an investigation of effectiveness based on the perceptions of students, professors, or even the job market [2]. Consequently, due to the number of studies that only analyze teaching strategies, one of the crucial limitations lies in the assessment strategies employed to measure the competencies developed in this process.

Given the importance of OHS topics in the professional realm, this work holds significant relevance, considering the need for a literature review encompassing the synthesis of assessment methodologies in this domain. Analyzing these methods is pertinent because, by understanding their benefits, advantages, and implementation challenges, educators can identify which methods best suit the assessment of developed knowledge, drawing from the perspectives of current works by various researchers. Besides, there is no adequate tool using active learning to provide feedback to students [7].

This study aims to investigate the assessment methodologies used in teaching Occupational Health and Safety within engineering. To achieve this objective, a literature review was conducted to answer the question: “What are the forms of student knowledge assessment when active learning strategies are used, and how can we classify them?”

2 Methodology

This study encompasses a narrative literature review. This type of review provides comprehensive and descriptive syntheses, written in a readable and broad manner, on a particular topic of interest [8]. Initially, we developed the research question previously presented, seeking ways to assess students’ knowledge after using active learning. Based on this theme, we searched the Scopus database for works that resemble the topic. Subsequently, we made a selection based on titles, abstracts, and keywords, followed by a full-text reading of the selected articles. In total, we selected 17 articles that addressed the formulated research question. Finally, the authors analyzed the content of the selected articles to address the main research question. For this purpose, a content analysis of the articles was conducted, following the approach outlined by [9]. In addition to conceptualizing assessment methodologies, the study proposes a classification based on their similarities.

3 Results

Based on what was observed in the articles of this review, a classification of assessment methods within the OHS teaching environment in engineering is presented. Table 1 shows five main groups: learning styles, student learning perception, pre-posttest assessment, existing tools and methodologies, and creation of assessment instruments.

In most cases, students are evaluated based on their perceptions of learning [4, 10], present their developed questionnaires for assessing knowledge perception, which were conducted with students through surveys followed by quantitative analysis [11], used a specific questionnaire to learn about human factors, and they also identified the importance of studying these concepts based on students’ perceptions.

Table 1. Methods of assessment for active learning methodologies

Methods of Assessment	Learning Styles	Index Learning Style (ILS)	Tortorella and Cauchick-Miguel (2018); Tortorella et al. (2020)
		Kolb Inventory	Borrego et al. (2013); Davidovitch et al. (2006)
		Myers-Briggs Questionnaire	Borrego et al. (2013)
	Student Learning Perception	Authored by	Behm et al. (2014); Bisantz and Paquet (2002); Oviedo-Trespalcacios et al. (2015)
		Already published in the literature	Cerezo-Narváez et al. (2019); Kuo et al. (2019); Page and Stanley (2014)
	Evaluation using pre-post testing	Perception-based evaluation	Bisantz and Paquet (2002); Din and Gibson Jr. (2019)
		Knowledge-based evaluation	Behm et al. (2014); Kuo et al. (2019); Oviedo-Trespalcacios et al. (2021)
	Utilization of existing tools and methodologies	Interviews	Kuo et al. (2019); McNeill et al. (2016); Page and Stanley (2014)
		Concept mapping	Besterfield-Sacre et al. (2004)
		Eye-tracking	Ahmed et al. (2021)
		Bloom's Taxonomy	Fernández and Ceacero-Moreno (2021)
	Creation of assessment instruments	Competency-based Assessment (CBA)	Ahmed and Sayed (2020)
		Industrial Learning Pilot Events (ILPE)	Kiritsis et al. (2013)

[12] use a simple questionnaire on knowledge perception based on the Kirkpatrick evaluation model [13]. employed two questionnaires already available in the literature: the Motivated Strategies for Learning Questionnaire (MSLQ) to assess motivation

perception and the Abbreviated Torrance Test for Adults (ATTA) to assess students' creativity [6]. present another questionnaire on self-assessment of knowledge based on results expected by the Accreditation Board for Engineering and Technology (ABET).

A series of studies utilized assessment at two-time points: before and after an active methodology approach. In the works of [11, 14], two assessments of knowledge and understanding of concepts are made with similar time intervals [4, 10, 13]. used the time interval between assessments to measure improvement in students' knowledge perception. Some researchers used existing tools and methodologies in the literature to assess their students. Regarding interviews, [13] used focus groups to evaluate qualitative questions and report their experiences [15]. conducted interviews to identify how students solve problems in disciplines based on their reports [6]. also used interviews to engage students and community members in dialogue about pro-proposed resolutions.

[16] applied skills assessment in participants' problem-solving, using eye-tracking and the generated mapping [17]. use concept maps to assess their students' knowledge. For this assessment, essential criteria are analyzed and then converted into rubrics [18]. also assessed the skills developed by students based on Bloom's Taxonomy. This taxonomy consists of a hierarchical organization structure of educational objectives, divided into six cognitive domains for different levels of learning.

Other authors developed their evaluative methodologies, such as [19], who created the Industrial Learning Pilot Events (ILPE). The methodology has a qualitative-quantitative bias based on multiplying different weights and factors related to developing attitudes, skills, knowledge, and competencies [20]. used an assessment style based on three pillars (summative, formative, and behavioral), in which the assessed factors are competencies developed by students, termed Competency-based Assessment (CBA).

This review generally observed three ways of assessing students' learning styles. The researched authors could observe how students learn concepts best, according to different approaches [21]. employed learning styles presented by the Kolb Experiment, and [22] used profiles to compare with classroom knowledge [23, 24]. used the Index Learning Style (ILS) questionnaire by Felder and Silverman as a way to assess student-learning performance. Additionally, the use of Myers-Briggs indicators, mentioned in [21] work, was observed. These indicators provide essential information about participants' strengths and weaknesses, guiding their development.

4 Discussion

Incorporating active learning instead of traditional teaching ensures student satisfaction and meets the actions expected by companies from this professional [25]. Even considering the various advantages and benefits in the literature, active learning methodologies still need to improve their adoption, mainly due to the assessment of the knowledge acquired by students and the feedbacks [7]. Most studies measured knowledge by applying questionnaires and tests to perceive what was learned [10, 12]. Some authors created methodologies that allow evaluation of how and what the student learns [19, 20]. However, [14] still argue that more efforts are needed to assess the knowledge and skills developed while applying the methodologies.

Considering interaction, using learning style methodologies was recognized as essential to understanding the class better and identifying the student profile [21, 24]. Although

not widely disseminated and used, activities involving practice bring different benefits to the class and the student-teacher relationship [2, 3]. Demonstrations of activity plans that elucidate the skills intended to be developed were also not found in the analyzed studies. This highlights a gap regarding practical skills, as many teachers need work to base their activity plans on. The area of health and safety at work is little observed in teaching research in engineering. It was found that the concepts related to the area are presented only as benefits and contributions to a productive system [6, 25].

The focus of these studies is not directly related to OHS. Thus, it is possible to affirm the existence of few studies with detailed analyses of the application of active learning methodologies in the studied area. It is also worth noting that the same finding is identical among all the analyzed studies regarding teaching practice with active learning in health and safety at work in engineering. This research not identified longitudinal studies, which bring the benefit of observing the knowledge and skills acquired over time and their variations.

5 Conclusion

This research was an initial study investigating different methods of knowledge assessment for developing a proprietary tool to use in classes that apply active learning. The analyzed studies offer a comprehensive overview of the different approaches to student knowledge assessment about OHS in engineering. These approaches highlight assessments based on student perception to more traditional knowledge tests and qualitative interviews. Despite efforts to categorize and evaluate these methodologies, challenges persist in determining the most effective and comprehensive approach to assess OHS competencies in engineering students. This gap underscores the ongoing need for research in OHS assessment in engineering education. This ensures that students are adequately prepared for workplace safety challenges throughout their careers.

References

1. Nguyen, H., Wu, L., Fischer, C., Washington, G., Warschauer, M.: Increasing success in college: Examining the impact of a project-based introductory engineering course. *J. Eng. Educ.* **109**(3), 384–401 (2020). <https://doi.org/10.1002/jee.20319>
2. Rodeghiero Neto, I., Amaral, F.G.: Teaching occupational health and safety in engineering using active learning: A systematic review. *Saf. Sci.* **171**, 106391 (2024). <https://doi.org/10.1016/j.ssci.2023.106391>
3. Freeman, S., et al.: Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci.* **111**(23), 8410–8415 (2014). <https://doi.org/10.1073/pnas.1319030111>
4. Oviedo-Trespalacios, O., Angarita, L.P., Maestre-Meyer, M., Correa, C.B.: Building the life-long learning competence in undergraduate engineering students with a laboratory practice in learning curve. *Procedia Soc. Behav. Sci.* **174**(12), 2021–2026 (2015). <https://doi.org/10.1016/j.sbspro.2015.01.870>
5. Lima, R.M., Villas-Boas, V., Soares, F., Carneiro, O.S., Ribeiro, P., Mesquita, D.: Mapping the implementation of active learning approaches in a school of engineering – the positive effect of teacher training. *Eur. J. Eng. Educ.* Online version. **49**(5), 945–964 (2024). <https://doi.org/10.1080/03043797.2024.2313541>

6. Page, L., Stanley, L.: Ergonomics service learning project: Implementing an alternative educational method in an industrial engineering undergraduate ergonomics course. *Hum. Factors Ergon. Manuf. Serv. Ind.* **24**(5), 544–556 (2014)
7. Grohs, J.R., Kirk, G.R., Soledad, M.M., Knight, D.B.: Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. *Think. Skills Creat.* **28**, 110–130 (2018). <https://doi.org/10.1016/j.tsc.2018.03.003>
8. Minichiello, A., Caldwell, L.: A narrative review of design-based research in engineering education: Opportunities and challenges. *Stud. Eng. Educ.* **1**(2), 31–54 (2021)
9. Bardin, L.: *Análise de Conteúdo*, 1ª edn, p. 288 p. ISBN: 9724415066. Edições 70, Lisbon (2015)
10. Behm, M., Culvenor, J., Dixon, G.: Development of safe design thinking among engineering students. *Saf. Sci.* **63**, 1–7 (2014). <https://doi.org/10.1016/j.ssci.2013.10.018>
11. Bisantz, A.M., Paquet, V.L.: Implementation and evaluation of a multi-course case study for framing laboratory exercises. *J. Eng. Educ.* **91**(3), 299–307 (2002). <https://doi.org/10.1002/j.2168-9830.2002.tb00707.x>
12. Cerezo-Narváez, A., Córdoba-Roldán, A., Pastor-Fernández, A., Aguayo-González, F., Otero-Mateo, M., Ballesteros-Pérez, P.: Training competences in industrial risk prevention with Lego® serious play®: A case study. *Safety.* **5**(4), 81 (2019). <https://doi.org/10.3390/safety5040081>
13. Kuo, H., Tseng, Y., Yang, Y.: Promoting college student's learning motivation and creativity through a STEM interdisciplinary PBL human-computer interaction system design and development course. *Think. Skills Creat.* **31**, 1–10 (2019). <https://doi.org/10.1016/j.tsc.2018.09.001>
14. Din, Z., Gibson Jr., G.: Serious games for learning prevention through design concepts: An experimental study. *Saf. Sci.* **115**, 176–187 (2019). <https://doi.org/10.1016/j.ssci.2019.02.005>
15. McNeill, N., Douglas, E., Koro-Ljungberg, M., Therriault, D.J., Krause, I.: Undergraduate Students' Beliefs about Engineering Problem Solving. *J. Eng. Educ.* **105**(4), 560–584 (2016). <https://doi.org/10.1002/jee.20150>
16. Ahmed, A., Hurwitz, D., Gestson, S., Brown, S.: Differences between professionals and students in their visual attention on multiple representation types while solving an open-ended engineering design problem. *J. Civ. Eng. Educ.* **147**(3) (2021). [https://doi.org/10.1061/\(ASCE\)EJ.2643-9115.0000044](https://doi.org/10.1061/(ASCE)EJ.2643-9115.0000044)
17. Besterfield-Sacre, M., Gerchak, J., Lyons, M.R., Shuman, L.J., Wolfe, H.: Scoring concept maps: An integrated rubric for assessing engineering education. *J. Eng. Educ.* **93**(2), 105–115 (2004). <https://doi.org/10.1002/j.2168-9830.2004.tb00795.x>
18. Fernández, P., Ceacero-Moreno, M.: Study of the training of environmentalists through gamification as a university course. *Sustainability.* **13**(4), 2323 (2021). <https://doi.org/10.3390/su13042323>
19. Kiritsis, D., Bufardi, A., Mavrikios, D., Knothe, T., Szigeti, H., Majumdar, A.: A competence-based industrial learning approach for factories of the future. *Educ. Inf. Technol.* **18**, 331–350 (2013). <https://doi.org/10.1007/s10639-012-9247-3>
20. Ahmed, A., Sayed, K.: Development of competency-based training system in Assiut- ITEC: A case study. *J. Competency Based Educ.* **5**(3), 1–12 (2020). <https://doi.org/10.1002/cbe2.1217>
21. Borrego, M., Karlin, J., McNair, L.D., Beddoes, K.: Team effectiveness theory from industrial and organizational psychology applied to engineering student project teams: A research review. *J. Eng. Educ.* **102**(4), 472–512 (2013). <https://doi.org/10.1002/jee.20023>
22. Davidovitch, L., Parush, A., Shtub, A.: Simulation-based learning in engineering education: Performance and transfer in learning project management. *J. Eng. Educ.* **95**(4), 289–299 (2006). <https://doi.org/10.1002/j.2168-9830.2006.tb00904.x>

23. Tortorella, G. L., Miorando, R., Fettermann, D., Mendoza, D. T.: An empirical investigation on learning and teaching lean manufacturing. *Educ. Train.* 62 (3), 339–354 (2020). <https://doi.org/10.1108/ET-11-2018-0232>
24. Tortorella, G.L., Cauchick-Miguel, P.: Combining traditional teaching methods and PBL for teaching and learning of lean manufacturing. *IFAC Pap.* **51**(11), 915–920 (2018). <https://doi.org/10.1016/j.ifacol.2018.08.465>
25. Huang, J., Pan, W., Liu, Y., Wang, X., Liu, W.: Engineering design thinking and making: Online transdisciplinary teaching and learning in a covid-19 context. In: Markopoulos, E., Goonetilleke, R., Ho, A., Luximon, Y. (eds.) *Advances in Creativity, Innovation, Entrepreneurship and Communication of Design. AHFE 2020. Advances in Intelligent Systems and Computing*, vol. 1218, pp. 267–274. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51626-0_19



The Use of Prototyping as an Experience for Teaching Ergonomics and Universal Design

Sabine De Paris^(✉) , Mariana Maciel Pasa , Angela Weber Righi ,
and Vanessa Goulart Dornéles

Federal University of Santa Maria, Santa Maria, RS, Brazil

sparis.arq@gmail.com, mariana_pasa@hotmail.com, {angela.w.righi,
vanessa.g.dorneles}@ufsm.br

Abstract. In Architecture and Engineering courses, teaching Ergonomics and Universal Design (UD) requires a combination of theoretical and practical approaches, employing methodologies that incorporate the latest market technologies. Specifically in Anthropometry, leveraging digital fabrication and 3D printing helps students enhance their experimentation and problem-solving abilities. This study aims to describe the development of instructional materials through prototyping, providing better support and improving the understanding of these concepts. The prototype, optimized through testing and anthropometric percentiles, demonstrates user-environment interaction in the classroom, integrating everyday elements while utilizing digital manufacturing to enhance the teaching of Ergonomics and Universal Design through innovative methodologies.

Keywords: Ergonomics · Universal Design · Prototyping

1 Introduction

Understanding user-centered spaces requires a comprehensive knowledge of human physical dimensions, such as standing and sitting height and work dimensions, which involve how the body interacts within a given space [1]. Additionally, the interaction between users and spaces is influenced by cultural factors and individual values, which shape how the space is perceived [2, 3]. Among the fields that explore different types of users, Ergonomics and Universal Design (UD) stand out, as both consider sizing aspects for products and built environments.

Ergonomics involves studying the human body's interaction with its environment, aiming to optimize the relationship with every day and work objects [4]. UD consists of designing products, buildings, and open spaces that accommodate users' diverse spatial needs and limitations [5]. Ergonomics and UD are increasingly prominent in educating new professionals in Architecture and Engineering courses, necessitating the integration of diverse perspectives into the teaching-learning process [6, 7]. Emerging production methods tied to digital technology, including rapid prototyping and digital manufacturing, introduce new avenues for progress and innovation in learning and knowledge assimilation [8].

The literature includes studies that employ modeling and additive manufacturing to produce medical products such as prosthetics, models for ergonomic tests, furniture, tactile maps, and architectural models [9, 10]. However, only some studies utilize anthropometric models for teaching Ergonomics and UD. The models available on the market, such as wooden mannequins, depict the articulation of movements and the proportionality of human body dimensions. Nevertheless, models with anthropometric percentiles are scarcely available. Furthermore, everyday elements such as doors, stairs, and furniture, which serve as fundamental teaching examples for Ergonomics and UD, are often not represented as physical models suitable for practical teaching. Therefore, this article aims to present the process of developing teaching materials for Ergonomics and UD, utilizing prototyping and 3D printing, in a manner that provides support as teaching aids without requiring substantial investment resources.

2 Theoretical Framework

Digital manufacturing enables the production of various personalized elements linked with digital design and virtual shape manipulation. Through rapid prototyping, tangible artifacts can be created and analyzed [11]. Among the most utilized types of rapid prototyping are additive manufacturing and subtractive manufacturing. Additive manufacturing involves layering material, for example, using a 3D printer, while subtractive manufacturing entails material removal, often through processes like CNC machining and laser cutting [12]. In the case of 3D printing, educational applications of the technology include supporting classroom activities, artifact production, and developing assistive technologies [13]. In Architecture and Engineering programs, rapid prototyping aids the design process as both a creation and application tool, bridging the digital and physical realms. Prototyping fosters experimentation, enhancing problem-solving skills among students as they detail and plan their artifacts [14].

Scientific exploration into models derived from 3D scans of real bodies, subsequently digitized for replication on mannequins, holds the promise of streamlining the production of products and environments for various demographic groups [15–17]. The Mandesso Design firm, led by industrial engineer Manu Alvarez, has crafted 29 digital human models spanning the 5th, 50th, and 95th percentiles across a spectrum of ages, from infants to adults. Unfortunately, access to the data from Alvarez's research remains unavailable [18]. Alternatively, other digitally accessible models, though lacking scientific grounding, are freely available on dedicated 3D printing platforms such as Thingiverse.

Even with available models and research, more educational resources need to be tailored to teaching Ergonomics and UD. Consequently, creating teaching materials suitable for diverse courses and higher education institutions is imperative for knowledge dissemination.

3 Methodology

The development stages of the present study are: Stage 1 - Literature review; Stage 2 - Purpose of the material; Stage 3 - Modeling; and Stage 4 - Prototyping. Stage 1 involved research in journals and scientific events on Ergonomics and UD education in Architecture and Engineering courses. Additionally, the literature review explored the incorporation of prototyping in education, along with its implications for experimentation and problem-solving.

After analyzing the literature and examining models used in scientific case studies as part of Stage 2, we found that the ergonomic mannequins currently on the market generally represent the dimensions and scope of the human body. This makes it possible to enhance the use of mannequins in architecture and engineering courses through the practical representation of ergonomic percentiles. To apply this study to our context, we used the percentiles from the book by [19], which refer to the Brazilian population.

In Stage 3, the authors explored modeling tools based on their previous knowledge. Among the available tools are BIM software and SketchUp, which have demonstrated limitations in modeling organic shapes. After conducting modeling tests on the human body, the chosen software was Rhinoceros, which met the requirements for the prototype.

Stage 4 of our research involved printing the models developed in Stage 3. For this, we used PLA filament (Polylactic Acid Biopolymer) and a Creality printer, specifically the CR-10S Pro model. The software we utilized for this process was Ultimaker Cura. These tools were made available to us within the facilities of the Digital Fabrication Laboratory at the Technology Center of the Federal University of Santa Maria (UFSM), also known as “Fábrica CT”. Throughout Stages 3 and 4, we conducted thorough analyses of improvements, as prototyping tests necessitate the practical verification of printed models and their respective digital models.

4 Results and Discussion

The digital modeling of the first prototype was based on commercially available wooden mannequins, which utilize nails and springs in their joints. After printing the first model at a 1:10 scale, it was observed that production became essentially artisanal due to the use of materials other than filament and the necessity of tools such as bench drills, complicating reproduction. The second prototype was derived from an articulated model available in the library of the 3D printing model website Thingiverse [20]. This model features a fitting system that functioned adequately for the study; however, its replication would require usage rights and it did not accommodate some of the desired positions, such as reach and standing/sitting. Both prototypes, one and two, are depicted in Fig. 1.

For the third prototype, building on the knowledge acquired from the previous prototypes, the development of an original model was initiated using Rhinoceros software. This software facilitates the modeling of diverse shapes and the interaction and association of different objects, enabling the achievement of the desired objectives for the prototype. Due to the digital modeling and 3D printing process, the prototype required several fitting and movement tests to ensure correct functionality (Fig. 2).



Fig. 1. First and second prototype



Fig. 2. Printing prototype fit tests

During the movement tests for range of motion and sitting/standing positioning, parts of the prototype frequently broke due to insufficient resistance and stability. Consequently, achieving compatibility between the digital model and the printing software settings necessitated fine adjustments of thicknesses, fitting dimensions, proportionality, and the positioning of the print itself to ensure maximum possible resistance.

With adjustments in stability and durability, the final prototype was printed at both 1:20 and 1:10 scales (Fig. 3), using the anthropomorphic dimensions of the 50th percentile of the male body [21]. It was determined that the 1:10 scale is more suitable for the proposed study, as it ensures reinforcement in the spherical structures, preventing breakage during use.

In general, it should be noted that a deeper understanding of the Rhinoceros software is necessary to streamline the modeling process and mitigate errors. Identifying issues related to object fitting, intersections, shape proportionality, and sizing are crucial aspects to consider during digital modeling. Furthermore, configuring the printing software,



Fig. 3. Final prototype

which encompasses layer thickness, supports, positioning, and filament type, is essential for achieving satisfactory results.

5 Conclusions

The prototype developed for this study, utilizing 3D printing, embodies the practical implementation of dimensions corresponding to a male body's 50% anthropometric percentile. In ergonomics and UD, the methodologies employed to fabricate the model and conduct 3D printing assessments necessitated thorough investigation and experimentation to determine the most effective approaches for achieving innovative material outcomes. The final prototype successfully illustrates a range of functionalities and maintains stability in both seated and standing positions, thereby fulfilling the research objectives. Future endeavors aim to broaden the prototype's applicability and encompass printing at various scales for additional desired percentiles (5%, 50%, and 95%) for both male and female subjects.

A limitation arises from the learning curve required to grasp various software applications for modeling and printing. Specifically concerning printing, variations occur based on the type of 3D printer employed, which may utilize either generic software (such as Ultimaker Cura) or proprietary software.

Encouraging student engagement in the creation of elements and fostering reflection on the interaction among diverse human bodies, their dynamics, and their capacities are not just academic exercises. They are essential for active pedagogy and preparing students for the professional arena. The practical application and comprehension of various user types and their interactions with the environment serve to heighten students' awareness of prevalent barriers in daily life and mitigate those that do not ensure comprehensive inclusion. This research, therefore, has significant implications for the field of ergonomics and industrial design.

Acknowledgment. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

1. Panero, J., Zelnik, M.: Dimensionamento humano para espaços interiores. Editora GG, Barcelona (2008)
2. Hertzberger, H.: Lições de arquitetura. Martins Fontes, São Paulo (1999)
3. Hall, E.: A dimensão oculta. Martins Fontes, São Paulo (2005)
4. Dul, J., et al.: A strategy for human factors/ergonomics: Developing the discipline and profession. *Ergonomics*. **55**(4), 377–395 (2012)
5. Mace, R.L., Hardie, G.J., Place, J.P.: Accessible environments: Toward universal design. Center for Universal Design, NC (1996)
6. Olguntürk, N., Demirkan, H.: Ergonomics and universal design in interior architecture education. *METU J. Fac. Archit.* **26**(2), 123–138 (2009)
7. Özeren, Ö., Özeren, E.B., Top, S.M., Qurraie, B.S.: Learning-by-Doing using 3D printers: Digital fabrication studio experience in architectural education. *J. Eng. Res.* **11**(3), 1–6 (2023)
8. Soomro, S.A., Casakin, H., Georgiev, G.V.: Sustainable design and prototyping using digital fabrication tools for education. *Sustainability*. **13**(3), 1196 (2021)
9. Kermavnar, T., Shannon, A., O'sullivan, L.W.: The application of additive manufacturing/3D printing in ergonomic aspects of product design: A systematic review. *Appl. Ergon.* **97**, 103528 (2021)
10. Howeid, D.R., Arafat, Z.: The impact of using 3D printing on model making quality and cost in the architectural design projects. *Int. J. Appl. Eng. Res.* **12**(6), 987–994 (2017)
11. Beaudouin-Lafon, M., Mackay, W.E.: Prototyping tools and techniques. In: *The human-computer interaction handbook*, pp. 1043–1066. CRC Press, Florida (2012)
12. Mathur, R.: 3D printing in architecture. *Int. j. innov. sci. eng. technol.* **3**(7), 583–591 (2016)
13. Ford, S., Minshall, T.: Where and how 3D printing is used in teaching and education. *Addit. Manuf.* **25**, 131–150 (2019)
14. Greenhalgh, S.: The effects of 3D printing in design thinking and design education. *J. Eng. Des. Technol.* **14**(4), 752–769 (2016)
15. Šajatović, B.B., Petrak, S.: Application and advantage of 3D body scanning in the ergonomic design of a sitting workplace. In: Salopek Čubrić, I., Čubrić, G., Jambrošić, K., Jurčević Lulić, T., Sumpor, D. (eds.) *Proceedings of the 9th International Ergonomics Conference. ERGONOMICS 2022. Lecture Notes in Networks and Systems*, vol. 701, pp. 37–44. Springer Nature Switzerland, Cham (2022)
16. Alemany, S., Valero, J., Ballester, A.: Advanced processing of 4D body scanning technology for the ergonomic design of products and environments. *Ergon. Des.* **31**(3), 30–37 (2023)
17. Pagliari, C., Montalti, A., Frizziero, L., Liverani, A.: Enhancing ergonomic comfort: A study on customized cushion design using 3d scanning and additive manufacturing. *Results Eng.* **25**, 104256 (2025)
18. Mandesso. Mandesso Design, Manu Alvarez Homepage. <https://mandesso.com/> (2024). Accessed 21 May 2024
19. Másculo, F.S., Vidal, M.C.: *Ergonomia: Trabalho Adequado Eficiente*. Elsevier, Rio de Janeiro (2011)
20. Makermachine. Poseable mannequin. Thingiverse Homepage. <https://www.thingiverse.com/thing:101451> (2024). Accessed 21 May 2024
21. Másculo, F.S., Vidal, M.C.: *Ergonomia: Trabalho adequado eficiente*. Elsevier, Rio de Janeiro (2011)



Mental Workload on Engineering Students: A Comparison Between Remote and Presential Teaching Using NASA-TLX

João Victor Razzera^(✉) , Eduardo Duarte Pena , and Angela Weber Righi

Federal University of Santa Maria, Santa Maria 1000, Brazil

joao.razzera@acad.ufsm.br

Abstract. In-person academic activities post-COVID-19 pandemic have different characteristics than those experienced pre-pandemic. Thus, understanding the cognitive aspects of this activity for students in these various scenarios and focusing on the mental workload becomes a strategy for improving teaching conditions. This study aimed to analyze the mental workload of engineering students at a public institution in southern Brazil during remote teaching (2021) and after returning to face-to-face activities (2023) using NASA-TLX and focus groups, which are related to cognitive performance. For this, a remote questionnaire was sent to all students enrolled at the institution in both periods. The main results indicate increased students' mental workload, from 73,40 (2021) to 75,18 (2023). Furthermore, the factors that most influenced mental workload during remote teaching were mental demand (83,30) and, during in-person teaching, effort (71,40). These findings reinforce the need to pay attention to constant changes and their consequences in the teaching-learning process, aiming for strategies compatible with the dynamic context currently experienced.

Keywords: Mental workload · NASA-TLX · Teaching-learning process

1 Introduction

The COVID-19 pandemic has brought about the digital transformation of higher education. During the pandemic period, teaching moved online due to social isolation, surprising academics who, until then, used in-person teaching strategies. Understanding the cognitive aspects of exact science students was already a pre-pandemic concern [1], mainly due to the intensification of work demands and time dedicated to academic development. These concerns increased during the pandemic, with the rapid transition to remote learning at all educational levels [2].

With the return to face-to-face activities post-pandemic, adopting a new teaching model was necessary, combining practices prior to 2020 with those adopted during the pandemic [3]. Therefore, analyzing the mental workload of students and its possible impacts, aiming to improve both the quality of teaching and the well-being of those involved, becomes fundamental.

Faced with this scenario, the present study aimed to: compare the mental workload of engineering students from a public institution in the South of Brazil during remote teaching (2021) and after the return of in-person activities (2023).

2 Research Method

This study was carried out in two distinct stages: the first in 2021, during the COVID-19 pandemic and using remote teaching, and the second moment in the year 2023, with face-to-face teaching taking place since 2022. Engineering students from a public institution in southern Brazil participated in the study. The same tools were applied at both data collection times, as shown in Fig. 1. The main tools used were an online questionnaire based on [4] and focus groups.

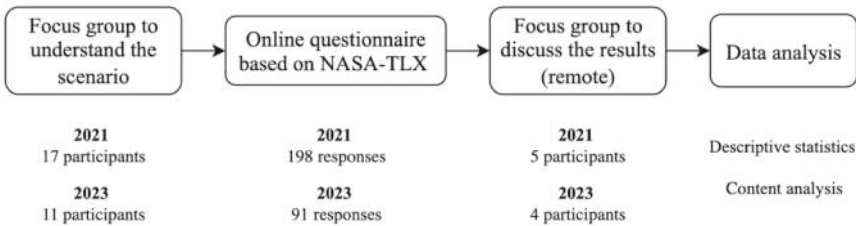


Fig. 1. Main information about the research method used.

The two stages were conducted with similar structures but with different student adherence, both in the quantitative and qualitative research approaches. It is believed that this occurred due, in the case of the study carried out in 2021, to the fact that students were in social isolation due to the COVID-19 pandemic. Because of this, there was a greater tendency to participate in remote activities, as there were no other possibilities at the time. With the scenario of returning to face-to-face, in the year 2023, participation in the research was more limited, as with face-to-face activities occurring commonly and students involved in them, there needed to be more time to respond to the survey with online questionnaires.

It is worth noting that the students participating in 2021 and 2023 are not necessarily the same, as the invitation was directed to all students enrolled at the institution during the collection periods, and participation was voluntary. However, in the 2023 collection, the inclusion criterion was that students had participated in the remote class period in 2021.

Furthermore, the first focus group was held remotely in 2021, while in 2023, it was segmented into two moments, one in person and the other remote, totaling 11 active participants. Regarding the questionnaire, distribution was maintained online via Google Forms and dissemination via email and social networks.

The data analysis was conducted using descriptive statistics for quantitative data and content analysis for qualitative data, ensuring a comprehensive understanding of the findings. Importantly, the research was approved by the institution’s ethics committee, underscoring the adherence to ethical precepts and the integrity of the research process.

3 Results and Discussion

In general, the results obtained in the two years of the research (2021 and 2023) were similar, with the general average for mental workload obtaining values of 73,40% and 75,18%, respectively. The specific values of each construct that make up this measure, for each of the years of the research, can be seen in Table 1. It can be seen that the year 2023 (face-to-face activities) showed a tendency towards higher values than the previous year of the study (remote activities).

Table 1. NASA-TLX main results.

	Mental D.	Physical D.	Temporal D.	Effort	Performance	Level of frustration	Mental Workload
2021	77,15	34,55	70,00	70,71	69,32	70,08	73,40%
2023	82,30	41,10	68,30	71,40	65,20	63,50	75,18%

Factors that may have contributed to this result were elucidated mainly through focus groups and responses to the questionnaire, the main ones being many demands in daily life, assessments in test format, and loss of content due to the remote period of activities throughout the year pandemic.

Aiming to understand the data better, some correlations were made between the results of the NASA-TLX demands and characteristics of the participants at the time of the research, questioned in an initial part of the form (Table 2), namely: i) current semester; ii) number of subjects taken; and iii) extracurricular activities carried out.

Table 2. Main correlated data.

	Mental D.		Physical D.		Temporal D.		Effort		Performance		Level of frustration	
	2021	2023	2021	2023	2021	2023	2021	2023	2021	2023	2021	2023
i) 5th to 7th semester	77,61	91,28	34,63	45,26	72,69	74,74	67,69	74,87	69,93	67,18	73,58	66,54
ii) 8th semester or more	78,40	75,48	35,94	37,98	71,60	63,46	72,08	68,85	72,83	63,75	68,87	61,15
ii) 1 to 4 subjects	77,17	78,41	33,48	41,59	65,22	69,77	67,39	66,59	68,48	57,95	70,22	63,64
ii) 5 to 7 subjects	76,71	83,96	36,07	45,00	68,80	67,60	70,30	73,65	71,07	66,56	70,73	63,54
ii) 8 subjects or more	77,41	82,38	31,61	31,67	74,11	68,33	67,50	71,43	70,18	69,76	68,57	63,10
iii) One extracurricular activity	79,41	78,80	38,83	36,57	72,50	68,24	71,65	67,59	73,78	60,46	70,37	61,02
iii) More than one extracurricular activity	79,36	86,67	27,87	44,76	70,96	71,43	67,45	77,86	67,98	71,90	73,19	69,05
iii) No extracurricular activities	71,58	88,13	32,98	51,56	65,09	64,38	67,02	75,94	67,89	72,50	67,02	64,38

It can be seen that, for mental demand (Fig. 2), there was a considerable increase for the group of students who were in the middle of their academic journey (5th to 7th period), a fact that can be justified due to their entry into university during the pandemic period, having to adapt again to the face-to-face routine in 2023. In addition to the fact that this group experienced a process of changing the conditions of the university offline to entirely online upon entry, and completion of this started another adaptation, but this time in a reverse flow, that is, from online to offline conditions. Despite this adaptation, there was no significant change concerning temporal demand (Fig. 3), with all groups maintaining similar results in both years of the study.

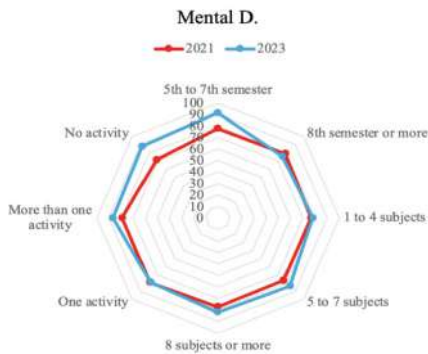


Fig. 2. Mental demand 2021 × 2023.

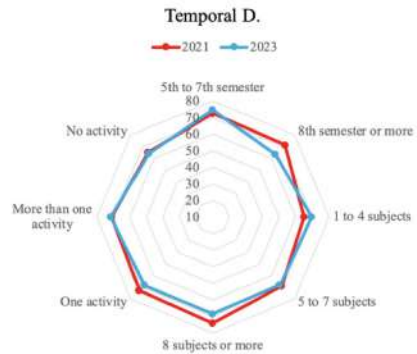


Fig. 3. Temporal demand 2021 × 2023.

Physical demand (Fig. 4) was the construct that showed the most significant difference between the 2021 and 2023 values. The values were higher in practically all correlated groups, emphasizing the difference in students who practice more than one extracurricular activity (27,87 in 2021 and 44,76 in 2023). With the support of qualitative data, it is understood that travel to activity locations and in-person experiences at the university may be the factors that most influence this criterion. Similar behavior was evidenced with the effort construct (Fig. 5). However, it is worth highlighting the importance of physical activity for general well-being, as regular practice promotes an improvement in self-knowledge, body image, and self-esteem, as well as cognitive and socialization functions, and also reduces levels of stress and anxiety [5].

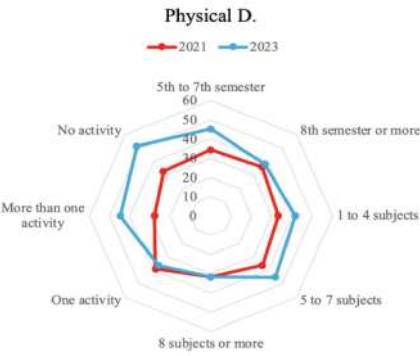


Fig. 4. Physical demand 2021 × 2023.

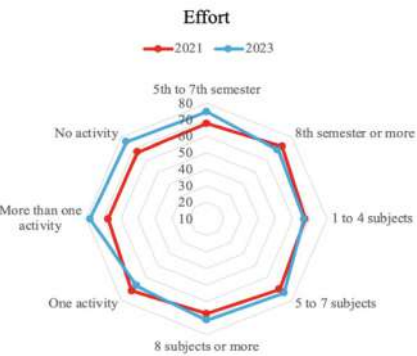


Fig. 5. Effort 2021 × 2023.

In Fig. 6, it is observed that the performance showed a drop or persistence in the values in the analyzed groups. In other words, students realize that their performance in face-to-face teaching is lower than in the remote period. However, the research could not identify apparent aspects that justify such values. And there was no comparison with

secondary data, such as academic records, that could relate to whether the students’ perception is expressed in the grades obtained in 2023.

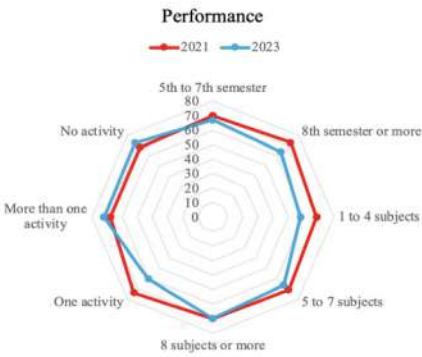


Fig. 6. Performance 2021 × 2023.

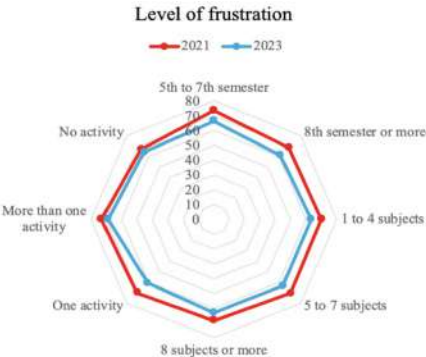


Fig. 7. Level of frustration 2021 × 2023.

As the last construct analyzed, the level of frustration (Fig. 7) showed a reduction in practically all related groups, which the interaction and sociability can influence resumed with face-to-face contact among the academic community. Face-to-face contact encourages greater acceptance among students, as highlighted through focus groups of the study in question, with a view to greater ease of communication and understanding.

4 Conclusion

This study aimed to compare the mental workload of engineering students from a public institution in the South of Brazil during remote teaching (2021) and after the return of in-person activities (2023). For this, a remote questionnaire using NASA-TLX and focus groups was applied.

The findings of this study are significant, indicating a notable increase in the mental workload of engineering students, from 73,40 (2021) to 75,18 (2023). The factors that most influenced this mental load were mental demand (83,30) and effort (71,40), underscoring the importance of these aspects in the student’s academic journey.

It’s important to acknowledge the limitations of this research, particularly the number of responses obtained through the questionnaire, which was 198 participants in 2021 and 91 responses in 2023. The low participation in the focus groups also contributed to a limited diversity of opinions. These limitations are crucial to consider for future studies, ensuring a more comprehensive understanding of the mental workload of students.

Studies similar to this should be carried out, mainly with a larger audience, covering other courses and teaching centers at the institution, to obtain more accurate results concerning the university context. To further improve understanding of the scenario over time, replicating the study at regular intervals, such as in 2025 and 2027, for example, would be valuable.

References

1. Alam, S., Hameed, A.: Teaching concerns in higher education: Impact of covid-19 in pedagogy. *J. Educ. Cult. Soc.* **14**(1), 318–332 (2023)
2. Ramadhani, D.S., Muslimah, E., Bagaskara, W.B., Fitriadi, R., Sakti, A.S.: The impact of COVID-19 pandemic on mental workload and fatigue of students during online and offline practicum. *SHS Web Conf.* **189**, 01001 (2024)
3. Oliveira, E.N., Aguiar, R.C., Oliveira, M.T., Cordeiro, S.E., Queiroz, T.L.: Benefícios da atividade física para saúde mental. *Saúde Coletiva.* **8**, 126–130 (2011)
4. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Adv. Psychol.* **52**, 139–183 (1988)
5. Estrada-Araoz, E.G., Paredes-Valverde, Y., Quispe-Herrera, R., Larico-Uchamaco, G.R., Paricahua-Peralta, J.N.: Examining the relationship between academic burnout and the university student's engagement: A cross-sectional study on the re-turn to face-to-face classes. *J. Law Sustain. Dev.* **11**(2), e424–e424 (2023)



Challenges in Mobilizing Workers' Experience and Development in a Technological Transition Project

Marta Santos^(✉) , Cláudia Pereira , Maria Antónia Cadilhe ,
and Liliana Cunha 

Center for Psychology at University of Porto, Faculty of Psychology and Educational Sciences,
University of Porto, Porto, Portugal
marta@fpce.up.pt

Abstract. Studies drawing attention to the importance of considering the different users, including operators, at all stages of a technology introduction project are becoming increasingly frequent. However, other studies point out that operators are often only included in reskilling training courses to deal with the technologies introduced. In this qualitative study carried out in a company in the Portuguese industrial sector, we sought to analyze a technological transition in an inbound warehouse (use-case pallet tracking) and understand the challenges posed by mobilizing the experience of workers throughout the process. The results show that operators were only taken into account during the implementation phase and that no formal moments of development were associated with the technology. However, casuistic factors (regarding collective work and leadership) were identified, contributing to the general perception that it was a successful process. Intentionalizing these factors in future projects could contribute to work sustainability and the development of workers.

Keywords: Technological Transition · Work Analysis and Training · Experience

1 From Reskilling to the Possibility of Mobilizing Experience

Following the publication of the Ergonomic Work and Analysis and Training network's participation in the IEA Congress in Toronto [1], some challenges were identified for future reflection. Among them was the need to “qualify the sustainability conditions and criteria of learning situations in training and at work” (p. S149), associated with technological and environmental changes that have an impact on workplaces.

Several documents [2–4] outline the challenges associated with Industry 4.0 and point to the skills gap that threatens to delay companies' ability to introduce new technologies into their contexts. The solution to this gap seems to involve these workers in upskilling and reskilling processes so they can be better prepared to work with new technologies. The discussion in these documents is, therefore, no longer based on the more catastrophist view that anticipated the disappearance of thousands of jobs as a result of

the technological revolution, but rather on its profound change. This change will require workers to get involved in initial or continuous qualification training processes to update their qualifications. In the words of Silva [5], who systematizes the position of these documents, “Workers will either ‘adapt’ to technological imperatives or else they will have to face the obsolescence of their know-how, which is becoming more and more labile as the capabilities of the machine advance” (p. 47).

Therefore, the reference to learning in the context of technological transition processes is not innovative. However, as the analysis carried out by Barcellini et al. [6] points out, in these processes, workers' participation is often limited to training sessions, which are designed without necessarily taking into account previous experience and skills developed because they are perceived as being out of step with the new needs of interacting with technology.

However, if “the industry of the future must be built around people and their know-how” [7, p. 40], can we limit their participation to training moments when all the decisions have already been made? Moreover, can we undervalue the contributions that result from acquired work experience, knowing how important it is for carrying out current and future work [8, 9]?

2 Digitalization of an Inbound Warehouse Activity: A Pallet Tracking Use-Case

2.1 Case Study Conducted

This research is an exploratory case study [10], seeking an in-depth understanding of a project to digitalize work in an inbound warehouse.

Qualitative and participatory methodologies, anchored in the analysis of work activity, were used to anticipate forms of activity that may emerge in I4.0 environments [11], associated with the possibilities of mobilizing experience and developing skills.

In this study participated 10 Logistics operators, ages between 26 and 57 years old ($M = 40$) and a seniority between 1 and 26 years old; 1 Team Leader (TL); 1 Project Manager (PM); and 1 Human Resources (HR) team member.

2.2 Context of the Study

The study is part of a technological transition project developed by a supplier of technology and services company from the industrial sector in Portugal.

As the project is still in progress, at this stage, our study focuses on one of the use-cases that the company intends to test, corresponding to a pallet tracking system in the inbound warehouse. This use-case aims to introduce a technology that enables the reorganization of the warehouse, making it more reliable and agile, ensuring product traceability.

The inbound warehouse work process is divided into three parts: unloading goods and recording software, labelling the products, and storage. The technologies developed corresponded to integrating a sensor for identification and creating a pallet ID and, consequently, software to support product verification and labelling. Two workstations

interact differently with these technologies. Workstation 2 (product verification and labelling) directly interacts with the software and perceives the work pace's imposition derived from the sensor's registration of the palette at Workstation 1.

2.3 Data Collection and Analysis

This section presents the data collection and analysis process carried out so far.

Data collection corresponded to the analysis of micro demographic data and documents relating to the function and process of introducing the technological system (TS); Meetings with PM and TL (3) and with an HR team member (1); and observations of Logistics operators' work activity (logistics warehouse).

Data analysis consisted of the descriptive analysis of micro-demographic data, the production of tables/schemes summarizing the observations, the synthesis of the verbalizations collected. The *Enabling Collaborative Situation* (ECS) framework [12] was used to systematize and present the analysis of the work situation.

The results that will be presented focus on the analysis of those involved during the process of introducing the new technology into the warehouse, the current work situation of the operators at workstation 2, and their training and learning processes.

3 Activity Evolution in Interaction With Technology and Workers' Experience

3.1 Management of the Process of Change Through the Introduction of New Technology

From the reconstitution of the technology implementation process in the meetings with the PM and TL, it emerged that the operators were only involved in the implementation phase of the first version of the technology (May 2022), being absent from the decision-making and design periods (beginning in 2018).

The figure of the TL in the design and implementation phases stands out for his knowledge of the operators' activities and close involvement with them.

The absence of an assessment of the future usefulness of the technology and monitoring of the implementation process was attenuated precisely by the TL's knowledge of the work activity in the warehouse and the proximity recognized by the operators.

3.2 Analysis of the Work Situation According to the ECS Framework

Concerning the first ECS criterion [12]—*learning a new and more efficient way of doing things*—the operators recognize that working with this new software allows them to get the job done faster and with fewer errors. However, they identify moments when the software is unstable: “*Work goes faster if [the software] is working well. Otherwise it delays more than it helps*” (Log. Op. 3). This instability leads operators to keep redundant operating modes, which they have from previous work experience, to make sure that they don't miss any critical information or make any mistakes.

Regarding the second ECS criterion—to increase the available possibilities and ways of doing things –, quality instructions guiding the workflow were identified. However, operators have the autonomy to introduce variations to speed up the work process, with room to manage non-conforming situations. The operators feel that their activity is not hindered by technology and is supported by the leadership (*“The guideline is to leave it alone when a [purchase order] has no data, but what we do is call the buyer or leave a message”* (Log. Op. 4).

In the third criterion—adjust the human-machine couple attributes according to the evolution of situations over time—in addition to the possibilities of autonomy to get the job done, the operators have the opportunity to participate in improving the system: they make suggestions which are analyzed and implemented according to their expertise (*“The inputs [for improvement] were always from the operators”* (TL); *“The operation always has to be part of [the implementation]”* (TL)). The autonomy granted to the operators comes from the management's recognition of their experience and ability to solve problems, particularly those related to improving technology: *“When the ideas are from the operation, they stand out a lot”* (TL). The operators perceive the work pace as the result of the quantity of products they receive from the pallet registration at Workstation 1 and not as an imposition by the software. In fact, when the software is working well, the pace of digital processing is slower than that of the worker, and there is a need to slow down to do a job well done.

3.3 Training and Learning Issues in the Analysis of the Work Situation

In addition, the study made it possible to highlight issues from the point of view of the conditions created for the learning and training of operators in the technological transition.

It was clear from the start that there were no formal training moments for using the new software. Still, the learning took place between colleagues, according to the needs they felt, and with the support of TL: *“I learned from my colleagues”* (Log. Op. 2); *“He [TL] came to work with the team to explain how it was done. He had a night with the team on shift”* (Log. Op. 1). *In fact, the company's choice seems to favour a close relationship between leadership and operations since the training is designed by a department associated with the plant's operations and not the people management department.*

However, it was also found that in the current situation, the activity is perceived as a learning opportunity due to (i) the possibility given to workers to manage problems when interacting with the system: *“Everything helps us to grow. If everything is easy we don't have any difficulties, but if there are deviations from the process it will help us grow”* (Log. Op. 4); (ii) the perception that this is a variable and challenging job, which requires new and adapted responses: *“you never do the same thing here and that's why I like this job”* (Log. Op. 2).

This would not be possible without the mobilization of previous individual experience and the use of an experienced work collective, visible in the recognition of the specificities of suppliers and purchase orders that determine the way of interacting with the software or the different products that require different handling and determine the forms of spatial and temporal organization in the interaction with the software.

4 Challenges and Criteria for the Development of the Activity and Workers

The analysis of this use-case allows for identifying some of the challenges that, although not new, make it possible to understand the difficulties workers face in technological transition projects and which can condition the ‘use’ of their know-how and experience throughout the transition process.

Firstly, we can identify the moment they were involved in the transition project. In this use-case, as pointed out in other studies [6], the operators were only considered when they needed to interact with the system to produce. In other words, they were part of the actual use of the technology. One possible consequence of this late introduction of the operators may have been the inability to find a technological alternative to the need felt by the operators at station 2 to slow down their work pace to keep up with the machine.

Alongside this challenge, the absence of a formal training and skills development strategy for operators to develop new ways of carrying out the activity was also identified. The project managers did not foresee the upskilling of these operators, contrary to what is widely believed [2–4]. Although learning took place with colleagues and TLs, there was no planning or intervention on the part of the design team to anticipate any difficulties the operators might have or to develop skills for using the system.

Despite this, the overall positive perception of the introduction of the new system seems to be due to some conditions that reveal themselves as ways to take into account the know-how of workers in technological transition processes and that can contribute to the sustainability of learning situations in training and at work [1]:

- possibility of mobilizing and developing the knowledge and skills they have, making use of their work experience, mainly because there was no disruption concerning their previous work;
- recognition by the TL of the activity carried out by the operators in interacting with the technology and solving problems arising from their work in the warehouse;
- perception by workers that the software introduced is functional and allows for the development of new ways of working and taking into account the evolution of work situations;
- having the possibility to make suggestions for improving the system and to use it collectively in a process that contributes to the development of shared skills.

While these aspects seem to us, as they do to other authors, to be essential for the sustainability of the work [8, 13], in this use-case they are the result of a favourable, but casuistic, combination of factors. However, in order to contribute to this sustainability, it would be important that these conditions were intentionalized during the project.

The operators’ contribution to the success of this use-case is clear, even though they were only integrated during the use phase. As pointed out by Sgarbossa et al. [14], the underperformance of some aspects of the new system could have been minimized if the operators had participated in the different decision-making moments.

Acknowledgements. This work was supported by Portuguese National Funding from the Agency for Investment and Foreign Trade of Portugal, E.P.E. (AICEP) (SI I&DT (Co-promotion) n.º

182852) and by national funding from the Portuguese Foundation for Science and Technology (UIDB/00050/2020).

References

1. Delgoulet, C., Santos, M.: Ergonomic work analysis and training: past, present and future. *Work*. **73**, S141–S152 (2022). <https://doi.org/10.3233/WOR-211267>
2. WEF—World Economic Forum: The Future of Jobs Report 2020. Geneva, Switzerland, WEF (2020) https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf
3. OECD: OECD Skills Outlook 2019: Thriving in a Digital World. In: OECD Publishing (2019), Paris, <https://doi.org/10.1787/df80bc12-en>
4. AAA—Association for Advancing Automation: Work in the Automation Age: Sustainable Careers Today and Into the Future [White Paper]. AAA. <https://www.automate.org/a3-content/whitepaper-work-in-the-automation-age> (2017)
5. Silva, D.: Entre mim e a máquina: O papel da experiência profissional em contextos em mudança pela automação e os seus contributos para pensar o futuro do trabalho. Doctoral Thesis. FPCEUP, Porto-Portugal (2024)
6. Barcellini, F., Béarée, R., Benckekroun, T.H., et al.: Promises of industry 4.0 under the magnifying glass of interdisciplinarity: revealing operators and managers work and challenging collaborative robot design. *Cogn. Tech. Work*. **25**, 251–271 (2023). <https://doi.org/10.1007/s10111-023-00726-6>
7. AIF: Le guide des technologies de l'industrie du futur: enjeux et panorama des solutions. Fédération des industries mécaniques (2018), France, http://www.industrie-dufutur.org/content/uploads/2018/03/Guide-des-Technologies_2018_V3.pdf
8. Volkoff, S., Gaudart, C.: Working Conditions and “Sustainability”: Converting Knowledge into Action. Rapport de Recherche 92. CEE (2015). <https://ceet.cnam.fr/publications/rapports-de-recherche/rapports-de-recherche-2015-954187.kjsp?RH=1507626803290>
9. Buchmann, W., Zara-Meylan, V.: The role of professional experience at the articulation of sustainable prevention for the environment and the work: Examples from two evolving sectors, landscape maintenance and horticulture. *Laboreal*. **19**(1) (2023). <https://doi.org/10.4000/laboreal.20434>
10. Yin, R.: Case Study Research and Applications: Design and Methods, 6th edn. Los Angeles, SAGE Publications (2014)
11. Barcellini, F.: Quelles conceptions de la coopération humains-robots collaboratifs. *Activités*. **17**(1), 1–28 (2020). <https://doi.org/10.4000/activites.5007>
12. Compan, N., et al.: The impact of Enabling Collaborative Situations in AR-assisted conignment tasks. *Front. Ind. Eng.* **1**, 1258241 (2023). <https://doi.org/10.3389/fieng.2023.1258241>
13. Vendramin, P., Valenduc, G., Molinié, A., et al.: Sustainable Work and the Ageing Workforce—A Report Based on the Fifth European Working Conditions Survey: Eurofound. Publications Office (2012) <https://data.europa.eu/doi/10.2806/42794>
14. Sgarbossa, F., Grosse, E.H., Neumann, W.P., Battini, D., Glock, C.H.: Human factors in production and logistics systems of the future. *Annu. Rev. Control*. **49**, 295–305 (2020). <https://doi.org/10.1016/j.arcontrol.2020.04.007>



A Technological Implementation to Improve Maritime Firefighting Training or an Opportunity to (Re)Think Training Contents and Project Management

Annaëlle Soulet¹ (✉), Lauriane Roger², and Catherine Delgoulet¹

¹ Conservatoire national des arts et métiers, 75005 Paris, France
annaëlle.soulet@lecnam.net

² Conservatoire national des arts et métiers, Hérouville-Saint-Clair 14200, France

Abstract. Fire is a major risk, the human, material, and environmental costs of which can have catastrophic consequences for the health and integrity of many living beings. The management of fire-related incidents on ships is a major preoccupation, forcing human beings, and sailors in particular, to develop new techniques and technologies to mitigate these incidents. These new techniques and technologies take the form of firefighting organizations and prevention tools. Based on intervention research, our paper provides an analysis of the notion of internal and reciprocal relevancy to training. Introducing this notion enables us to move away from a techno-centric approach to rethink learning situations and training content, by focusing on the activity of learners and trainers in simulated situations, i.e. in training situations. On the analysis of mediation, the results show the interactions played a major role in learning situations. These results lead us to discuss the relevance of introducing a technology into maritime fire-fighting training and invite to consider other vectors of transformation as the dimension of the collective health and safety in training courses.

Keywords: Work analysis · Training relevancy · Social mediations

1 Introduction

At sea, fire is a catastrophic event. To minimize the number of casualties and property damage in these situations, training is seen as a key to prevent accident and enables maritime firefighters to experience a wide range of fire situations [1]. Among the solutions to continually improve these trainings. The use of simulation via immersive technologies is highlighted [2]. Virtual reality technologies are promoted in training as a means of accelerating learning by putting learners into simulation [3]. However, this “technocentric” vision, i.e., guided by the prospect of innovation, is at odds with the challenges posed by the reality of work activity [4]. Training activities are unconsidered upstream of design projects [4]. This lack of consideration compromises the relevance of the training provided, which makes it a major issue [5].

In this paper, the relationship between the conduct of learning situations and the design of training is examined by focusing on the social mediations of learners and trainers to characterize learning and task completion processes. Based on intervention research, conducted in a maritime firefighting training school in France, the demand was to identify the benchmarks to be transposed to design a new simulator in five years' time in the form of an augmented reality. Concerned with the articulation of training and work dimensions, this demand is questioned from the angle of the notions of internal and reciprocal relevance between these training devices and the work situations that learners will encounter in real work situations [6]. The discussion considers the place to the preservation of the collective in training course.

2 Entry by Technology Turned Into an Analysis of Relevance

Upstream of the simulator design process, stakeholders' expectations were to transpose knowledge to design the future simulator based on the current training. Transposing means transforming what takes place between learned and practical knowledge to invest it in a later stage in a creative process [7]. In training, the transposition of knowledge integrates a didactic dimension characterized by a selection organized by the trainers between scholarly and practical knowledge [8]. The knowledge requires an analysis to identify the relevance of the current training system for subsequent transformation. Relevance is defined as the elements that are adapted to the training situation and help to connect what is learned with what can be applied in the workplace [6]. This concept is presented in two distinct registers: i) Internal relevance corresponds to the adequation between the determinants of the training situation (organizational, material and human resources, characteristics of learners and trainers, and relationships between them) and the development conditions [9]; ii) Reciprocal relevance relates to the properties of training situations that can sustain the development by organizing a dialogue between training and work [9]. Underlining a process of collaborative development, the articulation of these registers depends on the analysis of current training. Analyze the current training means asking what forms of social mediation promote learning [10]. According to Vidal-Gomel (2022) [6], social mediation takes five distinct forms: i) Formative mediations provided by a member of the current situation, and which guides the learner in the realization of its operations [6]; ii) Epistemic mediations correspond to the knowledge provided to enable learners to acquire knowledge about the object and its characteristics [11]; iii) Pragmatic mediations based on an objective of achieving an action [12]; iv) Heuristic mediations refer to the relationships that learners develop to promote the creation of references [13]; v) Collaborative mediations dedicate to sharing knowledge, references, constraints between actors to collectively elaborate strategies to act subsequently in situation [6].

In the maritime firefighting training, simulations are based on social mediations to achieve the objective be able to act within a collective organization to extinguish a fire and evacuate smoke [14]. As a result, the health protection aspect appears to be secondary, or even missing [15]. Therefore, we are questioning how adopting registers of relevance, by focusing on social mediations, can help us understand what is at stake in training, build links with work, and bring about a transformation of the current situation and the project management?

3 Method

3.1 Characteristics of the Maritime Firefighting Training

The learner population was aged between 23 and 47 with a wide range of navigation experience, from no training at all to in-depth knowledge of the operations involved. Seven trainers with between 10- and 20-years' experience, and an average age of 45, supervise the training. Taking place over four weeks, the objective is to train the professionals to be able to coordinate actions to act on a fire in less than twenty minutes. This training includes theoretical part on the characteristics of the formal organization and regulatory policies, as well as a two-week practical part in which eight simulations are carried out each day. In these simulations, ten participants, are mobilized in a space representing a ship. Each learner is initiated to play a role and communicate by radio. Two of them coordinate the actions of the other eight learners as they move through space to extinguish the fictitious fire. Two trainers supervise and assess the coordinators. At the end, the conduct of operations is debated between the trainers and learners.

3.2 Examine Social Mediations in Relation to Relevance Registers

To understand the internal and reciprocal relevance, an analysis of the activity was carried out, including an examination of the training requirements, observations and working groups [16]. On the first hand, the exploratory and systematic observations were realized during learning situations over a 4-month period. The aim was to understand the organization of learning situations and their determinants. We chose to stand next to the two learners coordinating the operations and filmed the twenty-minute simulations and debates. In this way, we had access to communications between learners and trainers. As a result, 157 films were gathered.

On the other hand, the verbal communications in learning situations were analyzed jointly with eight learners and seven trainers regarding five social mediations (formative, collaborative, epistemic, pragmatic, heuristic) and relevance registers [8]. At the same time, social mediations were tracked temporally, using an activity chronicle. This collective analysis provokes a dialogue between the participants promoting the sharing of experiences, identifying the variety of learning situations, deducing links with real situations and distinguishing ways to improve training. All working groups were filmed with the consent of the participants and transcribed.

4 Social Mediation Questioning the Relevance of Training

4.1 Information Process Organized to Promote Collaborative Mediations

Learning is mediated by learners seeking to collectively solve the problem. Mediations evolve according to the phases of the situation. During the first five minutes, the mediations are predominantly heuristic (32%), epistemic (35%) and collaborative (28%). The remaining 5% correspond to pragmatic mediations. For 55 learners out of 60, communications are induced to share common references to collectively establish strategies. Between five and ten minutes, mediations are mostly pragmatic (representing 64% of

the learners). Strategies are formulated by stating specific goals for action on the fire. Between ten and twenty minutes, mediations are oriented towards collaboration (77%). This collaboration is based on heuristic (13%) and epistemic (10%) mediations, which involve monitoring to react to unforeseen events. Three learners, who have the most experience of the eight respondents, report that they are unable to use the references they have established in their professional activities, which, in their view, limits their ability to participate. On the contrary, five novice learners express that being in binomial with an experienced allows to understand how to communicate, how fire is represented and to build temporal references. According to experienced learners, collaborative mediation is in the majority because they “need to discuss effective strategies that go beyond the operational rules defined in the reference manuals”. Sixty-one percent of collaborative mediations between experienced and novice learners lead to conflicts of objectives that can compromise the progress. As a result, experienced mobilize strategies deviate from the prescribed, while novice learners respect the rules. Mediations are constrained by the emission of divergent cues and depend on individual experience. This variability of strategies for coordinating actions gives rise to controversy, particularly between experienced and trainers. Trainers decide which strategies are acceptable on the basis on the training standards. Therefore, experienced are limited in their ability to apply their acquired strategies in real-life situations due to the assessment framework. Even though the trainers acknowledge the problems faced by these learners, the evaluation problems prevail over the arguments. Mobilizing the internal relevancy, the simulation promotes mutual learning between learners. This training neglects the experiential knowledge developed in practice and partially adapted to the variability of the audience.

4.2 The Trainer’s Role: A Guarantor of Strategy to Safeguard the Collective

The prescribed role of trainers is to assess learners during simulations. However, results show that they do more. Trainers distribute instructions and add variables to help the situation evolve. On average, 32% of social mediations initiated by trainers are aimed at learner coordinators. Most mediations were addressed to learners dealing with the fire. Among the mediations addressed to learner coordinators, 64% of formative and 33% of pragmatic are identified, corresponding to moments when trainers demonstrate gestures to be reproduced to check the equipment worn, and provide temporal, linguistic and strategic cues to check the health of other learners. In addition, analyses of mediations addressed to learners acting on fire shows an average of 96% of heuristic mediations, which correspond to the addition of variables. Trainers add these variables based on the strategies recognized by the learners. In this case, the variables added are often a casualty or the loss of a crew member. According to trainers, “fictitiously experiencing of losing a colleague or the need to evacuate an injured person helps to create strategic cues that will be a basis for protecting the health of their colleagues in real-life work situations”. The mediations have two objectives: to encourage learners to reconfigure their strategies to adapt to changes in the situation, and to ensure the collective is safe throughout the simulation period. Taking the view of reciprocal relevance, trainers deviate from the prescription by adding that “the success of an operation is not extinguishing the fire but returning with all the members”.

5 Discussion and Conclusion

The aim of this paper was to examine how the adoption of relevance registers, coupled with an analysis of social mediations, supports a transformation from technocentric project management to health-related consideration to build experience. Analysis of learners' social mediations show that learning takes place in a unilateral way, where the novice learns through the presence of a more experienced learner. Referring to the theory of the zone of proximal development, learning is possible by social mediation with a more experienced third party [17]. Discussing the variability of reference points is tantamount to apprehending different visions of conducting operations, but also to learning to evolve within a social environment [18]. This training does not recognize the diversity of the characteristics of learners. Elaborating choices based on anticipation, the experienced provide the continuity of action to the achievement of the pre-defined stages of the training reference. During the debriefing, trainers express understanding of these strategies but keep to the prescribed reference to note. Trainers tend to disregard experience under the guise of treating learners fairly [19]. Developing oneself in situations means being able to mobilize one's power to act, to invest oneself in them, to realize oneself in them [20]. The lack of diversity of reference points limits the power to act of professional learners. To overcome this constraint, we sought to identify all the information derived from professional practice and transpose it into learning situations. The purpose was to provide each learner with individualized development within a collective learning space [21]. However, this training program places a strong emphasis on health preservation. Beyond this focus, this teaching method ensures a link with real-life work situations. Vocational training often focuses on technical aspects and underestimated the dimensions of health [22]. Considering the "act on health" dimensions ensures a pragmatic knowledge to integrate professionally [23]. In this regard, the dimensions learned respond to a transmission of partial practical knowledge. Neglecting the experience can have adverse effects such as reviving trauma or deteriorating relationships between members [6]. By questioning the existing, the training content has been reworked to integrate the dimensions of health preservation that appear essential to establish a link work and training. As part of the implementation of augmented reality in training, the consideration of experience allows to configure systems capable of encouraging collaborative learning by ensuring communication opportunities between learners, but also by accepting that the references are varied and individual. Augmented reality must be able to go beyond predefined prescription scenarios to provide all relevant references. This prescription often amounts to validating steps in a specific order. This successive validation process limits the acquisition of collaborative and practical learning. To conclude, technological implementation requires to adapt to the characteristics of the participants and identifying singular objectives to ensure collaborative learning for each learner.

References

1. Baalisampang, T.: Review and analysis of fire and explosion accidents in maritime transportation. *Ocean Eng.* **158**, 350–366 (2018)
2. Williams-Bell, F.M.: Using Serious Games and Virtual Simulation for Training in the Fire Service: A Review. *Fire. Technol.* **51**, 553–584 (2014)

3. Müller, S.: Attributes of Expert Anticipation Should Inform the Design of Virtual Reality Simulators to Accelerate Learning and Transfer of Skill. *Sports Med.* **53**, 301–309 (2023)
4. Béguin, P.: Designing for instrument-mediated activity. *Scand. J. Inform. Syst.* **12**, 173–190 (2000)
5. Våpenstad, C.: Lack of transfer of skills after virtual reality simulator training with haptic feedback. *Minim. Invasive Ther. Allied Technol.* **26**(6), 346–354 (2017)
6. Vidal-Gomel, C.: Analysing relationships between work and training in order to prevent psychosocial risks. *Saf. Sci.* **145**, 105517 (2022)
7. Dusi, N.: Translating, Adapting, Transposing. *Appl. Semiot./Sémiotique Appliquée.* **24**(9), 82–94 (2010)
8. Vidal-Gomel, C.: What ergonomics, professional didactics and training practices bring to educational design? *Activités.* **8**(2), 115–149 (2011)
9. Delgoulet, C.: Soutenir la circulation de l'expérience au travail. *Apprentissage et transmission de l'expérience en situation de travail: Dialogue entre formation d'adultes et ergonomie*, Presses universitaires de Rouen et du Havre, pp. 41–56 (2021)
10. Donato, R.: A sociocultural perspective on language learning strategies: The role of mediation. *Mod. Lang. J.* **78**(4), 453–464 (1994)
11. Drew, P.: Epistemics in social interaction. *Discourse Stud.* **20**(1), 163–187 (2018)
12. Arundale, R.: Toward a pragmatics of relating in conversational interaction. *J. Pragmat.* **179**, 19–25 (2021)
13. Béguin, P.: Working, learning and designing through simulation. In: 11th European Conference on Cognitive Ergonomics: Cognition Culture and Design, pp. 5–13. European Association of Cognitive Ergonomics, Catania (2002)
14. Samurcay, R.: Didactical use of simulations. *Le travail humain.* **61**(4), 333 (1998)
15. Seyedin, H.: Health, safety, and education measures for fire in schools: a review article. *J. Edu. Health Promo.* **9**(1), 121–128 (2020)
16. Guérin, F., Laville, A.: Understanding and transforming work: the practice of ergonomics, 1st edn. Anact, Lyon (2007)
17. Van Meter, P.: The Role of Theory in the Study of Peer Collaboration. *J. Exp. Educ.* **69**(1), 113–127 (2000)
18. Mugny, G.: Socio-cognitive conflict and structure of individual and collective performances. *Eur. J. Soc. Psychol.* **8**, 181–192 (1978)
19. Cau-Bareille, D.: Training, age and technological change: difficulties associated with age, the design of tools, and the organization of work. *Work.* **41**(2), 127–141 (2012)
20. Clot, Y.: *Travail et pouvoir d'agir*, 1st edn. Presses universitaires de France, Paris (2008)
21. Falzon, P.: *Constructive Ergonomics*, 2nd edn. CRC Press, New York (2014)
22. Ouellet, S., Vézina, N.: Activity analysis and workplace training: an ergonomic perspective. In: Filliettaz, L., Billett, S. (eds.) *Francophone Perspectives of Learning Through Work. Professional and Practice-based Learning*. Springer, Cham (2015)
23. Gastmyer, C.L.: The impact of the affordable care act on health education: perceptions of leading health educators. *Health Promot. Pract.* **15**(3), 349–355 (2014)



Experiences in Teaching Ergonomics as Part of an Undergraduate Course on Occupational Safety and Health: A Case Study from the National School of Public Health in Medellín, Colombia

Yaniel Torres^(✉) 

National School of Public Health, Universidad de Antioquia, Medellín 050010, Colombia
yaniel.torres@udea.edu.co

Abstract. This study evaluated the impact of introducing ergonomics as part of an occupational safety and health (OSH) course for undergraduate students at the National School of Public Health in Medellín, Colombia. A total of 12 student projects corresponding to two academic semesters of 2023 were evaluated regarding the application and appropriation of concepts related to ergonomics. Seven of the 12 projects (58%) explicitly identified the different components of the work system. Concerning the three domains of ergonomics, 100% of the projects identified aspects of physical ergonomics. Only six (50%) and five (42%) projects identified aspects of organizational and cognitive ergonomics. The results indicate that physical ergonomics remains the domain most frequently associated with ergonomics. In addition, cognitive and organizational aspects of work are not easily visible for students who are not experts in ergonomics, as they require a greater understanding of the work system. Including ergonomics in the teaching of undergraduate students can increase awareness of the importance of ergonomics as a discipline. Our results show that in a modest amount of time, the appropriation and application of concepts associated with ergonomics can be achieved, albeit limited in scope.

Keywords: Ergonomics Teaching · Undergraduate Programs · Occupational Safety and Health · Student Projects

1 Introduction

The need for the public and professionals from various fields to be aware of ergonomics and its wide range of goals and activities has been recognized for a long time [1]. Different societal stakeholders must realize how the principles of ergonomics can help design and evaluate tasks, jobs, products, environments, and systems while considering people's needs, abilities, and limitations [2]. In this way, the well-being of people is put at the center while optimizing the system's performance (quality, productivity, safety).

One possible way to increase the awareness of ergonomics and its benefits is to include ergonomics as a subject within the curriculum of some undergraduate programs. At the National School of Public Health of Universidad de Antioquia in Medellín, Colombia, students of two undergraduate programs (1) Healthcare Administration and (2) Environmental and Sanitary Administration, have historically received a course in Occupational Safety and Health (OSH). Unfortunately, until recently, this course excluded ergonomics as a topic. From 2020 onwards, the topic of ergonomics was included in this OSH course as an initiative proposed by Prof. Yordán Rodríguez, Ph.D., an internationally recognized scholar in the field of ergonomics. Thus, eight of the 64 hours of the OSH course are dedicated to ergonomics: six hours of lectures and two hours of student project presentations. In this modest time, students are taught the concept of ergonomics, the concept of work system, and the three domains of ergonomics (physical, cognitive, and organizational) through concrete examples. As an evaluation, students are instructed to present in groups a case in which they describe a work system and its components and identify some problems from the perspective of ergonomics and its three domains. Students are also asked to outline some solutions to the extent possible.

Until now, no formal assessment has been made of these student projects concerning learning the concepts taught. This study aimed to evaluate the impact of teaching ergonomics to undergraduate students as part of an OHS course on the application and appropriation of concepts taught. This is based on the assessment of student projects.

2 Methodology

The year 2023 was selected as the reference period, which comprised two academic semesters. We analyzed 12 ergonomic projects presented by students of two undergraduate programs as part of their OSH course. This involved 51 undergraduate students, 24 students (5 projects) from the first semester, and 27 students (7 projects) from the second semester. For the ergonomic projects, students had to use a real case of a work system to which they had easy access. They were instructed to (1) identify the elements of the work system according to the model developed by [3], which includes person(s), tasks, tools/technology, organization, internal environment, and external environment; (2) identify problems related to the domains of ergonomics (physical, cognitive, or organizational); and (3) propose some potential solutions.

For analysis purposes, information was extracted from the cases presented by the students in their ergonomic projects. A table was created with a summary of the relevant information, including the project title, the sector of activity concerned and type of business, the ergonomics domains that were brought up (physical, cognitive, or organizational), problems identified, and proposed solutions. An overall balance of the projects was then made based on the information extracted as evidence of the impact of ergonomics teaching on the application and appropriation of relevant ergonomic-related concepts.

3 Results

A summary of the ergonomics projects used in this study is presented in Table 1. Five of the 12 projects revolved around administrative office work; two were related to the healthcare sector and two were related to the restoration sector. The remaining three projects corresponded to software engineering, automotive parts sales and warehousing, and electrical engineering services. Seven of the 12 projects (58%) explicitly identified the different components of the work system as defined in the model proposed by [3]. Regarding the domains of ergonomics, 100% of the projects identified aspects of physical ergonomics. Only six (50%) and five (42%) projects identified aspects of organizational and cognitive ergonomics. Six cases (50%) presented in the ergonomic projects came from small or very small companies, three of which were in the informal sector (25%).

In all projects, several problems were identified that needed to be addressed. Some of the issues identified included poor lighting, limited workspace, poor organization of the workplace, exposure to heat and deficiencies in ventilation, ambient noise, and imbalances in the workplace (working heights, chairs with poor ergonomics, inadequate postures, manual handling). Students could identify possible errors in two cases, No. 5 and No. 10 (Table 1). In eight cases, the proposed solutions were very basic or were not made explicit.

In nine of the 12 student projects, possible solutions to the ergonomics issues identified were not addressed, and very basic solutions were proposed. Only three projects covered, to some extent, the potential solutions.

4 Discussion

Results of the study of the students' projects showed that physical ergonomics remains the domain most frequently associated with ergonomics. This can have several explanations. For example, [4] conducted a study on the social representation of ergonomics among 21 Colombian executives and found that "the vision of ergonomics is more oriented toward micro-ergonomics in the domain of physical ergonomics, showing a limited knowledge of the scope of current ergonomics." In addition, the application of ergonomics in industrially developing countries (IDCs) has focused, with more emphasis, on the physical elements of work: use of tools, design of workstations, manual material handling, and evaluation of work environments, especially in tropical climates [5]. Unsurprisingly, posture, movements, and chairs are the usual things that come to mind when students are asked about ergonomics at the beginning of the course.

In addition, cognitive and organizational aspects of work are not easily visible for students who are not experts in ergonomics, as they require a greater understanding of the work system or a deeper inquiry. For example, communication, teamwork, time management, and information processing issues require further investigation. Physical aspects, including environmental aspects, are more evident. The above can explain the difference in the relative presence of these domains among student projects.

The marked representation of office work likely reflects the ease of access to these work environments (e.g., Some students work in offices). It is simply easier to access an office than a construction site. Despite this, the projects generally reflected various

Table 1. Summary of ergonomics projects. Abbreviations: W.S. (Work System), P (Physical), C (Cognitive), O (Organizational)

Project Title	Sector of activity	W.S.	P	C	O	Identified problem	Proposed solutions
1. Skin care products commercialization	Administrative work in the formal sector. Small business	X	X	–	X	Lighting, space, and ventilation problems. Prolonged work in an office chair	Very basic solutions
2. Higher education institution	Administrative office work. Formal sector. Large company	–	X	–	–	Lighting and postural problems associated with working with computers	Basic postural hygiene recommendations
3. Beauty salon (Manicurist work)	Manicurists work in the formal sector. Very small business	X	X	X	X	Critical postures, lighting conditions, and high external noise	Improve workload and work environment. Redesign chair
4. Maintenance of household appliances	Administrative work in the informal sector. Very small business	–	X	–	–	Work in precarious conditions, plastic chairs and glass tables unsuitable for office work. Residential space	Very basic solutions
5. Medical examination laboratory.	Laboratory work in the health sector. Large company	–	X	X	X	Poor storage of blood samples. Poor ergonomics of lab workstations. Reduced spaces. Possibility of error.	Possible solutions were not explicitly addressed

(continued)

Table 1. *(continued)*

Project Title	Sector of activity	W.S.	P	C	O	Identified problem	Proposed solutions
6. Integral solutions in electrical engineering	Office work and workshops. Formal sector. Very small business	–	X	–	X	Poor office ergonomics, disorganized tool and wiring storage area, poor lighting.	Several possible solutions were discussed
7. Higher education institution	Administrative office work. Formal sector. Large company	X	X	X	–	Limited space. Background noise from outside. Limited space. Obstacles in circulation spaces	Very basic solutions
8. Software engineering	Software developer. Formal sector. Medium-sized company	–	X	X	–	Lighting, ventilation and background noise problems. Chair with poor ergonomics	Very basic solutions
9. Work in a cafeteria	Restoration sector. Formal sector. Large company	X	X	–	–	Limited working space. Exposure to heat	Possible solutions were not explicitly addressed
10. Hospital sterilization center	Healthcare sector Formal sector. Large company	X	X	X	–	Handling of heavy loads. Poor design of workstation. Noise and lighting problems. Possibility of errors	Possible solutions were not explicitly addressed

(continued)

Table 1. (*continued*)

Project Title	Sector of activity	W.S.	P	C	O	Identified problem	Proposed solutions
11. Sales and storage of automotive parts	Automotive parts warehouse. Informal sector. Very small business	X	X	–	X	Work surface height, chair with poor ergonomics, workflow deficiencies, labor informality, lack of storage space	Possible solutions were not explicitly addressed
12. Bakery products and pastries	Restoration sector. Informal sector. Very small business	X	X	–	X	Exposure to heat, poor lighting, standing work, and lack of space (private house). Labor informality	Very basic solutions

working contexts, as shown in Fig. 1. The sectors of activity chosen by the students reflect Colombia's reality. A significant percentage of labor informality exists, and small and very small businesses face situations of precarity [6].

A positive aspect is that in at least two projects, students could identify factors influencing the possibility of error occurrence. Both cases were related to the healthcare sector: a medical examination laboratory and a hospital sterilization center (case No. 5 and case No. 10 in Table 1). This is encouraging, considering that some students are undergraduates in hospital administration and that patient safety is a fundamental element in managing health services. Patient safety focuses on preventing adverse events to avoid harm to patients, including medical errors, intrahospital infections, and patient falls [7].

The fact that only three of the 12 projects covered the possible solutions in any way is entirely logical. The exercise was not aimed at proposing solutions; instead, they were instructed to identify some solutions as far as possible. However, this seems difficult for undergraduate students with no formal training in ergonomics and only a few hours of basic instruction.



Beauty center (case No. 3)



Automotive parts repair (case No. 11)



Sterilization center (case No. 10)



Education institution (case No. 2)

Fig. 1. Examples of work contexts covered in student ergonomics projects

5 Conclusion

The inclusion of ergonomics in the teaching of undergraduate students can increase their awareness of this important discipline. Our results show that in a modest amount of time, the appropriation and application of concepts associated with ergonomics can be achieved, although the scope is limited. Overall, the students were able to approximate various work contexts and identify different work system components from an ergonomic perspective. They also identified ergonomic issues that need to be addressed. While essentially focusing on physical aspects, they also considered environmental elements (noise, lighting, heat) and, in some cases, even factors contributing to errors.

Ergonomics can significantly contribute to meeting challenges in designing and managing work systems. To this end, raising awareness among different social actors is essential. To achieve a greater impact of ergonomic discipline, it is necessary to include it in various educational programs at both undergraduate and graduate levels. The experience presented here can be replicated and expanded to other undergraduate programs where appropriate.

References

1. Martin, D.W., Wogalter, M.S.: The Exposure of Undergraduate Students to Human Factors/Ergonomics Instruction. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. **41**(1), 470–473 (1997)

2. Dul, J., et al.: A strategy for human factors/ergonomics: developing the discipline and profession. *Ergonomics*. **55**(4), 377–395 (2012)
3. Holden, R.J., et al.: SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*. **56**(11), 1669–1686 (2013)
4. Gómez-Salazar, L.: Representaciones sociales de la Ergonomía en personal directivo. *Revista Venezolana de Gerencia*. **27**(98), 435–451 (2022)
5. Huck-Soo, L., Richardson, S.: Ergonomics in industrially developing countries: a literature review. *J. Hum. Ergol.* **41**(1–2), 1–16 (2012)
6. Torres, R.M.S.: Poverty and labor Informality in Colombia. *IZA J. Labor Policy*. **10**(1) (2020)
7. WHO. Patient safety: a global health priority. 2019.; <https://www.who.int/patientsafety/policies/global-health-priority/en/> (2020). Accessed on 23 May 2020



Enhancing Workers' Safety in Manufacturing: A Human-Centered Approach in Interaction Design of Digital Training Tools

Lucia Vigoroso¹  , Günter Alce² , Luca Boccadutri³, and Federica Caffaro¹ 

¹ Department of Education, University of Roma Tre, 20, Rome, Italy
lucia.vigoroso@uniroma3.it

² Design Sciences, Lund University, P.O. Box 118, 221 00 Lund, Sweden

³ Olivero Impianti, Strada del Santuario, 74-76, 12045 Fossano, CN, Italy

Abstract. Digital innovation offers promising avenues for accessible, engaging, and user-friendly training solutions, crucial for promoting safety among manufacturing workers. However, optimizing these solutions demands a profound understanding of workers' needs and preferences in interacting with these digital systems. The ongoing “DigiSafeMan” project aims to develop a prototype of a digital training tool tailored to manufacturing workers, adopting a Human-Centered Design approach. The present study focused on context and user needs analysis, involving the examination of secondary data on accidents and training activities at a selected firm in the Piedmont region (Italy), as well as semi-structured interviews with 26 workers to explore their risk perceptions, training priorities, and feedback on game features and design. Results highlighted prevalent injuries and the need to improve in current safety training methods. Participants prefer interactive, visually engaging materials, emphasizing video content and intuitive interfaces. Concerns arose about potential motion sickness from a 360° panorama image. This research phase laid the foundation for a human-centered, innovative, and sustainable digital safety training tool.

Keywords: Human-Centered Design · Interaction design · Manufacturing industry

1 Introduction

Digital innovations have offered great opportunities in recent years to the manufacturing industry in the three pillars of economic, environmental and social sustainability. More specifically, regarding the social aspects, new technologies help to reduce the number of fatal and non-fatal accidents at work, and improve workers' physical health, safety, and spirit [1]. Innovation has also allowed the development of more engaging training methods, involving workers in continuous discussions on the knowledge acquired and behaviours performed. These digital tools are designed to be as interactive as possible, by paying particular attention to the user interface (UI) and user experience (UX) in the interaction design [2] and providing users with continuous stimuli and feedback [3].

Furthermore, integrating game elements into the interface, like challenges and rewards, fosters experiential learning and boosts workers' engagement, motivation, and satisfaction [4]. Also, through a "microteaching" approach, knowledge "in pills" and brief factual information can be conveyed through active and interactive exercises and quizzes [5], reinforcing the memorization of specific work operations [6]. To be sustainable for workers, digital training tools should be designed through an ergonomic human-centered approach [7] which accounts for workers' needs, perceptions and interest about adoption.

1.1 Context and Aim of the Study

The manufacturing industry is still responsible for most fatal and non-fatal accidents worldwide. According to the last available statistics [8], in 2021, in Europe, the manufacturing sector accounted for 19.0% and 14.4% of the total non-fatal and fatal accidents, respectively, occurred in all economic activities. Adequate safety training accessible on digital devices, incorporating effective digital interfaces with gamified elements, can significantly promote safe behaviors among workers, help reducing the perception of work-related stress and enhancing engagement with safety practices and injury prevention [9]. Thus, the overall aim of the "DigiSafeMan" (DIGItal SAFETy training for MANufacturing) research project is to develop a prototype of a digital training tool able to create an engaging game experience to convey safety information to production workers of a manufacturing plant in the Piedmont region (north-west Italy). This contribution describes the preliminary findings of the ongoing project and focuses on the analysis of the context and users' needs and requirements.

2 Material and Methods

2.1 Participants

The company involved in the project (Olivero Impianti) provides its clients with a comprehensive service of design, construction, assembly, and maintenance of industrial plants for any type of industry. Twenty-six workers took part in the investigation. All participants were male, had a mean age of $M = 32.8$ years ($SD = 9.74$, range: 18–57 years) and a mean working experience of $M = 12.6$ years ($SD = 10.79$, range: 1–31 years). Regarding the job positions within the firm, 34.6% were technicians, 34.6% were assemblers/maintainers, and 30.7% specialized as machine operator. With regard to education, 23.08% of the participants had completed junior high school, 73.08% had high school-level education and 3.84% had a university degree.

2.2 Instruments and Procedure

For context analysis, secondary data on 5-year accidents and training activities in the firm was analyzed, and workers and trainers were observed during safety training courses. For users' needs analysis, participants were administered a short questionnaire collecting socio-demographic data (i.e., gender, age, years of experience in manufacturing work) and information about their previous experience with digital games, in terms of hours

per day (open question) and interest in using this type of safety tool if available on the market (four-point scale, from 1 = “not interested at all”, to 4 = “very interested”).

Then, semi-structured interviews were administered in two parts. Each interview lasted approximately 45–50 minutes. The first part explored workers’ risk perceptions and what safety contents should be emphasized during training courses; the second part provided feedback regarding game features/mechanics and components of the design system. Also, the participants tested selected demos of safety applications and safety games available online (<https://www.safetyapp.co.uk/app/demo>), and the Thinking Aloud technique [10] was adopted to collect data on their cognitive processes while interacting with the demos. The interviews and testing phase were audio-recorded. Then, each record was transcribed verbatim and subjected to qualitative content analysis. Miro visualization software (<https://www.miro.com>) was also used to organize the data into categories and visualize them in an abstraction diagram. Overall, the qualitative data collected allowed us to understand the context and limitations of existing similar tools and to identify the behaviors and mental models of potential users [11]. Regarding the quantitative data, descriptive statistics in terms of frequencies of all the investigated variables were computed.

3 Results

3.1 Context Analysis

Eyes (25%), hands (25%), and ankles (10.72%) were detected as the most frequently injured body parts. Safety training courses were provided through lectures, in which slides were used to support the trainer’s presentation. However, the trainer kept workers engaged by asking questions, fostering discussion, and prompting them to share real-life experiences. When possible, a practical demonstration of the correct use of safety equipment was provided. Please note that the first paragraph of a section or subsection is not indented.

3.2 Users’ Needs Analysis

Game experience Regarding workers’ prior experience with digital games, 61.54% did not usually play games, 15.38% usually played less than 30 minutes per day, 7.70% played about 30–60 minutes per day, and 15.38% usually play one hour or even more. Regarding their interest in using a digital tool when available, 19.27% were not interested/little interested in using it, whereas 80.77% were interested/very interested in using it.

Workers’ risk perceptions and safety contents Results from the first part of the interviews can be summarized as follows

- 1) risks were often underestimated, leading to inattention and indifference: *“Often, habit and the monotony of work operation can lead you to perform them incorrectly, even if you’ve always done them safely”, “When you become confident in something, you tend to overlook the danger and pay less attention”;*

- 2) more real-life case studies and consequences of wrong behaviors should be shared during training: *"The experience of a near-miss, whether personally lived or seen with your own eyes, can be shocking"*;
- 3) what Personal Protective Equipment (PPE) should be used in what kind of operation should be better clarified: *"At the workplace, there should be a strong emphasis on using Personal Protective Equipment correctly and educating what are inappropriate clothes when operating with machineries"*;
- 4) the relevance of keeping protective gears in place should be emphasized, *"In a rush, a worker uses a machine and doesn't keep protective gears in place as it should be. It is dangerous!"*.

Usability aspects and games The participants' responses from the second part of the interviews can be grouped in 4 different sections, referring to the preferred channel, the preferred/main characteristics, UX/UI and game features/mechanics.

The preferred channel to convey the safety messages was videos, since some concepts might be better explained through dynamic images rather than written text. "Videos convey a lot of details. Instead with images I may focus on specific details". However, this did not imply that the text should be eliminated, as some workers have emphasized the importance of having both channels complement each other: "The text content is ok. But I would have some specific interactive images instead of the usual static picture", "If I don't want to watch a video, but I'm obliged to watch it...I would also like a textual description. Both the video and text should complement each other".

Regarding the preferred characteristics, participants explained that written texts should not be too long, whereas, videos should have a maximum duration of 2–3 minutes and should explain how to perform operations safely, not providing theoretical information only: *"Duration of 2-3 minutes is fine, if it can be done in less, even better"*, *"The video should have a track so I can easily watch from a specific point in the video"*, *"Videos are okay for theoretical aspects, but it would be nice to have videos about safety procedures"*. Also, music in the background is useless and, a voice description may be better.

About the UX/UI, workers mentioned that the adoption of strategic colors could make the interface more engaging and can maximize the intuitiveness: *"Color code like traffic lights, red, yellow and green, to indicate risk levels"*, *"I prefer the dark theme"*. Adoption of bigger texts and icons was highly recommended as well as better internal links between pages to enhance the user flow: *"Maybe bigger text size and no small buttons. Starting at the age of 40, it begins to become difficult to read on devices"*, *"It must have a drop-down menu"* and, *"The horizontal scrolling would be interesting to be implemented"*.

Regarding the game features/mechanics, participants expressed preferences for a realistic graphic style to represent the shop floor, while the interactivity proposed by "find the error" games was appreciated. While some participants positively perceived the interactive 360° panoramas, others expressed uncertainty, noting that such images may induce seasickness: *"I find the cartoon image more suitable for kids"*, *"Use green or red signs in case of correct or incorrect responses"*, *"Find the error" mechanic quite immediate at a glance*, *"The demo game with the 360-degree rotation is more engaging and interesting"*, *"This is a bit more complicated for me [360° panoramas]"*.

I'm already getting seasick. Anyway, it is better than a video". Quizzes were widely appreciated, whereas the time feature was the least enjoyed game mechanic: *"The idea of the questionnaire at the end is nice, like when you do surveys on Instagram or social networks", "The idea of implementing quiz is okay, but there shouldn't be a timer; it makes me feel anxious"*.

4 Discussion

Overall, the results indicated that workers perceived the current safety training activities as "too stiff", listing the machinery-related risks without adequately explaining safety procedures or the actual consequences of incorrect behaviors. The need for more interactive training methods emerged. This evidence fits with previous results indicating the learning-by-doing method as more effective, and it paves the way for developing new training interventions based on users' actual needs [12].

Considering their emotional feedback and responses, the workers' insights will be pivotal for creating an interactive and gamified digital product. Indeed, adding gamified components may cause confusion and may require a period of adaptation and appropriation. Positive emotional feedback toward e-learning and games is crucial, as it boosts self-confidence, motivation and enjoyment, and helps overcome the fear of being judged, allowing users to make mistakes without fear of being laughed at [13].

Feedback on interactive 360° panoramas aligns with previous findings. While this interaction can trigger more engagement and guide users' attention, it can also be more demanding and may cause motion sickness, preventing users from perceiving its added value [14]. Finally, the appreciation for simple game format, such as quizzes, is encouraging as previous research has demonstrated that various online quiz tools and quiz apps are highly effective in enhancing users' learning experiences.

It should be acknowledged that gender was not considered in the user's needs analysis since participants were only males. This aligns with the statistics reporting that women are underrepresented in the manufacturing industry [15] but would deserve further exploration. In the following phases, according to the Human Centered Design approach [7], the prototype of the digital training tool will be developed from the concept design stage through scenarios development and mock-ups, and its usability will be tested with workers representing human variability in terms of gender, age, provenance, and work experience.

5 Conclusion

The first phase of the "DigiSafeMan" research project provided pivotal information for developing a human-centered, innovative, and sustainable digital safety training tool. The to-be-developed prototype intends to serve as a pilot experience, highlighting the transferability of the research methods and product to other manufacturing contexts and the possible socio-economic benefits of more sustainable safety training in the medium and long term.

References

1. Braccini, A., Margherita, E.: Exploring organizational sustainability of Industry 4.0 under the triple bottom line: The case of a manufacturing company. *Sustainability*. **11**(1), 36 (2019)
2. Norman, D.A.: *The Design of Everyday Things*, Revised and Expanded edn. Basic Books, Hachette Book Group, New York (2013)
3. Ardanza, A., Moreno, A., Segura, Á., de la Cruz, M., Aguinaga, D.: Sustainable and flexible industrial human machine interfaces to support adaptable applications in the Industry 4.0 paradigm. *Int. J. Prod. Res.* **57**(12), 4045–4059 (2019)
4. Isabelle, D.A.: Gamification of entrepreneurship education. *Decis. Sci. J. Innov. Edu.* **18**(2), 203–223 (2020)
5. Dwinggo Samala, A., Bojić, L., Bekiroğlu, D., Watrinhos, R., Hendriyani, Y.: Microlearning: Transforming education with bite-sized learning on the go—insights and applications. *Int. J. Interact. Mobile Techno. (IJIM)*. **17**(21) (2023)
6. Hesse, A., Ospina, P., Wieland, M., Yepes, F.L., Nguyen, B., Heuwieser, W.: Microlearning courses are effective at increasing the feelings of confidence and accuracy in the work of dairy personnel. *J. Dairy Sci.* **102**(10), 9505–9511 (2019)
7. ISO: International Organization for Standardization. ISO 9241-210:2019. Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts. International Organization for Standardization, Geneva, Switzerland, 2016
8. Eurostat (2024). Accidents at Work—Statistics by Economic Activity. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics_by_economic_activity#:~:text=In%202021%2C%20there%20was%20a,19.0%20%25%20of%20the%20total%3B (2024). Accessed 6 May 2024
9. Lovreglio, R., Duan, X., Rahouti, A., Phipps, R., Nilsson, D.: Comparing the effectiveness of fire extinguisher virtual reality and video training. *Virt. Reality*. **25**(1), 133–145 (2021)
10. Lewis, C.: Using the "Thinking-Aloud" Method in Cognitive Interface Design. IBM TJ Watson Research Center (1982)
11. Cooper, A., Reimann, R., Cronin, D., Noessel, C.: *About Face, The Essentials of Interaction Design*, fourth edn. John Wiley & Sons Inc, Indianapolis, Indiana (2014)
12. Zhang, X.S., Xie, H.: Learning by doing approach in the internet environment to improve the teaching efficiency of information technology. *Phys. Procedia*. **24**, 2231–2236 (2012)
13. Allal-Chérif, O., Lombardo, E., Jaotombo, F.: Serious games for managers: Creating cognitive, financial, technological, social, and emotional value in in-service training. *J. Busi. Res.* **146**, 166–175 (2022)
14. Broeck, M.V.D., Kawsar, F., Schöning, J.: It's all around you: Exploring 360 video viewing experiences on mobile devices. *Proceedings of the 25th ACM international conference on Multimedia*, 762–768 (2017)
15. EIT Manufacturing. (2021). Promoting Gender balance in the manufacturing sector. <https://www.eitmanufacturing.eu/news-events/news/promoting-gender-balance-in-the-manufacturing-sector/> (2024). Accessed 6 May 2024

Transport (I)



The Effect of Augmented Reality Head-up Display Interface Layout on Driving Behavior: A Systematic Literature Review

Jingjing Cao¹, Yanling Zuo¹, Peggy Wang², and Jia Zhou¹ (✉)

¹ School of Management Science and Real Estate, Chongqing University, Chongqing, People's Republic of China
Jiazhou@cqu.edu.cn

² California PATH, University of California, Berkeley, CA 94804, USA

Abstract. Augmented Reality Head-Up Display technology (AR-HUD), which combines graphical information with real-world scenes, brings users a better interaction experience. However, too much virtual information on AR-HUD might distract the driver's attention and even cause traffic accidents. To explore this, how to present the diverse information prompts clearly to users and balance driving performance with the interface design becomes particularly important. In this study, we first analyze the product of AR-HUD automotive interaction interfaces in the current market. Then, our search results yielded 29 articles and discussed and summarized the influences of AR-HUD on driving behavior from the perspectives of humans, machines, and the environment. Finally, the paper provides the potential design guidelines for AR-HUD. The results showed that information based on AR elements, such as number, display color, transparency, and so on, could be automatically adjusted in different scenarios, and the design also requires displaying the types of information to match the demands of different task levels during driving. Additionally, subsequent design should consider the principles of visual attention allocation to offer augmented information types as quickly as possible to avoid dangers.

Keywords: Augmented Reality Head-up display (AR HUD) · driving behavior · inference layout · interaction design

1 Introduction

With the growth of Vehicle-to-Everything (V2X) technology, autonomous or self-driving cars have emerged as the era's mainstream, and intelligence and user-friendly redefine the automotive industry's human-machine Interaction (HMI) experience. HUD (Head-Up Display) has long been employed as a form of automotive display system, with introductions such as C-HUD [1] and W-HUD [2], which have to some extent improved driving safety and enhanced the driving experience. Subsequently, with the development of Augmented Reality (AR) and the emergence of the metaverse, Augmented Reality Head-Up Display technology (AR-HUD) [3], which provides AR graphics to overlay

information onto the real world, has been deployed and rapidly iterated and upgraded. The global sales of vehicles equipped with AR-HUD will reach 5,300,000 units in 2025 [4].

Although AR-HUD enables vehicles to blend real-world environments with virtual graphics and overlays environmental information such as road conditions [5] and navigation [6] onto the real driving scenes, additional information might divert drivers' attention and increase the likelihood of traffic accidents [9]. For instance, too many AR graphics could lead to visual confusion, which causes the road signs in other directions to be ignored, resulting in hidden driving dangerous [10]. A large amount of driving information displayed on the windshield could block the driver's field of view, and frequently changing information on the AR HUDs might distract the driver's attention from crucial road elements [14], and even cause driving accidents while driving in dangerous situations [20]. Additionally, some researchers proposed that when drivers paid too much attention to the information displayed on the windshield, they tended to ignore the real-world driving environment [9]. Although HUD was designed to enhance safety and reduce cognitive load [8] information overload might make it difficult for drivers to quickly identify and process critical information, increasing the cognitive load while driving and even causing traffic accidents [9]. Because of its real-time interaction, AR-HUD technology also needs drivers to switch cognitively between physical and virtual information spaces, requiring time and cognitive effort [10].

Therefore, researchers must fully understand what essential information should be displayed based on the current driving tasks, and how to design the proper location and arrangement of elements in the windshield. In this systematic review paper, a broad discussion and survey of AR-HUD research is summarized from human factors and interaction design. We try to provide guidelines for designing the AR-HUD interface layout to improve the safety of drivers and other road users.

2 Method

Based on meta-analysis, we used the keywords AR, HUD, Automobile, and Information Display Design to review original articles using electronic databases systematically. Only the articles written in English from 2010 through 2024 were considered, and 29 published articles were included in our review, as shown in Fig. 1.

3 Result and Discussion

3.1 Current AR-HUD on the Market

The interface information of the AR-HUD system provides in-vehicle driving [11], entertainment [12], convenient information [13], safety, and navigation. Most of the products in the current market generally include the types of information [16] and their presentation mode (multi-window or single-window layout) are different. These informational advantages help reduce the workload for drivers [14], enabling them to better focus on driving tasks in dangerous situations [14], 90% of all the information used in driving is visual [15], and visual support is paramount for Advanced Driver Assistance

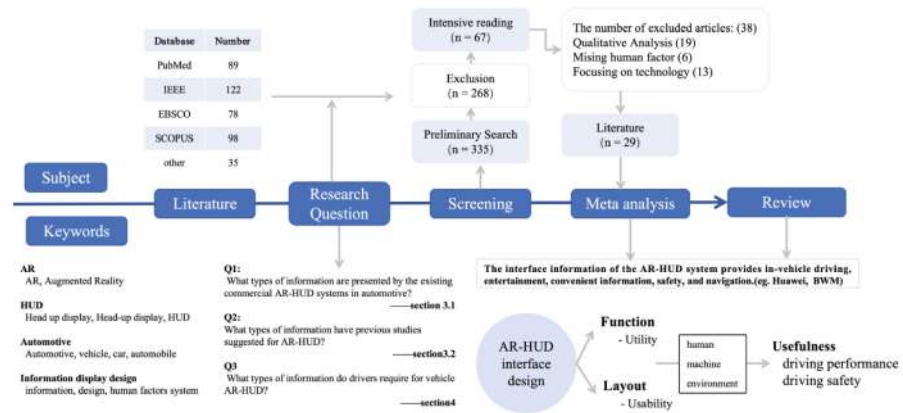


Fig. 1. Methodological framework used to inform the research questions and guide the review

Systems (ADAS). In-vehicle industries have been exploring the information framework of these visual elements. Therefore, this paper provides two examples to show the layout characteristics of several AR-HUD products on the market, as detailed in Table 1.

Table 1. Current AR-HUD on the market & Interface Feature

Product	AR-HUD Interface Elements	AR HUD Interface layout
BMW	Vehicle speed, cruise control, navigation, diagnostic control, music, phone, speed limits sign, radio stations, lane keeping assist, vehicle status alerts, and adaptive cruise contro.	A large-scale display with panoramic vision; Capable of perceiving multiple situational factors; The type of information is personalized according to the driver’s interests
Huawei	Engine speed, vehicle speed, fuel level, navigation (road name, distance, arrows, driving direction), lane changing guidance, front vehicle distance warning, safety assistance driving (car dashboard warning lights, night vision technology / enhanced rain-fog display), and infotainment	Providing six different display interface modes

3.2 Information on the AR-HUD

Human Factors. Many Individual characteristics might influence users’ preferences and usage patterns of technology. Therefore, considering gender differences in the design of AR-HUD could be beneficial for enhancing user experience [17]. Age is

also an important factor, and affects cognitive abilities and vision [18]. Regional cultural differences and road conditions might necessitate AR-HUD to display different types of information or present information in specific ways [19]. Driving experience is also a significant factor. Experienced users might adapt more quickly to complex HUD systems, while novice users might prefer simplified and intuitive displays [21]. Individuals' lifestyles and driving styles might influence drivers' preferences and understanding of AR-HUD information. Some researchers conducted a literature review to identify which information affects driving behavior in HUD, and HUD information types should be divided into three main aspects: urgency, general driver information (GDI), and In-Vehicle Information System (IVIS) [17, 22, 23].

Machine. The design of the AR-HUD human-machine interface (HMI) is also a crucial factor influencing driver interaction preferences [24]. We discuss the interface design from three dimensions: content, recognition, and understanding.

Content of Visual Display Information. The content of visual display information primarily relates to the quantity of symbols in the HUD, referring to the total number of various symbols, icons, pointers, and others, displayed on the screen [25]. The drivers focused on completing the driving task by using visual information. Therefore, the interface layout should be designed to strictly control the number of elements and reduce the number of non-driving-related information types [22]. A simulator study found that most drivers did not want HUD graphics to appear in their focal areas while driving. Instead, they preferred to place the HUD on the right, below, or left of their viewing areas [26].

Recognition of Visual Display Information. The style, color, transparency, position, and size of elements in AR HUD could influence the driver's attention and cognitive ability [20, 24, 28]. It is essential to guide the driver's attention effectively, ensuring that the driver can accurately understand and respond to information while driving [27]. The choice of these colors could influence the driver's emotions and behaviors. Transparency also needs to consider the clarity of information to avoid interfering with the driver's line of sight [27].

Understanding of Visual Display Information. Because of the continuous increase in automotive information system content and the complex driving scenarios, it is necessary to understand the semantics and meanings of various types of information before designing. AR-HUD requires a longer projection distance and a larger field of visual angle [29]. AR extends the three-dimensional world to overlay graphics directly onto the real world [30], aligning the information displayed on the windshield with the real world seen by the driver [31]. Combining real-world objects or locations with their intrinsic information in in-vehicle could get more compact the structure of information and enhance perception.

Environment. Due to the multitude of information types, the external environment information is divided into two parts for discussion: safety information and navigation information.

External Safety Information. External safety information includes factors such as safety alerts, hazard warnings, night vision, and future state predictions. In AR-HUD, these

elements’ display is achieved through a 3D object recognition module [32]. By quantitative analysis, researchers find that the external safety information supports the detection and avoidance of dangers in the environment to reduce the drivers’ cognitive load and improve cognitive efficiency, and also enhances visual road perception and understanding to improve the detection and avoidance of dangers in in low visibility conditions (such as rain and snow, and dark [33].

Navigation Information: Navigation information includes navigation instructions (such as maps, routes, arrows, and so on), traffic signs, front vehicle distance, and so on. It provides real-time guidance by AR-HUD interface or navigation applications. Navigation information helps drivers predict future states of vehicles and the environment to enhance situational awareness and reduce driver workload, thereby improving driving performance and safety [34].

4 Discussion and Conclusion

To address the question of what information and how it should be presented to the driver by AR-HUD, this study conducts a literature review to explore the design of the AR-HUD interface about driving behavior, considering the perspectives of humans, machines, and the environment. Therefore, we summarize some guidelines for reference, as shown in Table 2.

Table 2. The guideline for information of AR-HUD interface design

Key	AR-HUD Interface Elements
Amount of Information	Hick’s Law states that the time required to make a decision.Based on this, the number of information is limited to less than six and present the most critical information should be only one in AR-HUD
Necessary Information	Speedometer and navigation instructions as necessary information should be positioned in AR-HUD. Waring information should be adaptively presented according to real-time road conditions . Entertainment information are briefly displayed in AR-HUD view, and ASAD information should be conveyed by voice and light flashing
Transparency & Color	Maintain a high brightness between1200:1 and 1500:1 , and color contrast ratio up to 80,000:1 Use vivid and easily distinguishable colors for different types of information. -- speed limit could be shown in red , navigation in green , and warnings in yellow

(continued)

Table 2. (continued)

Key	AR-HUD Interface Elements
Location & Size	Different types of information could be hierarchically displayed to ensure drivers could access necessary data in real time. -- navigation instructions typically should be placed at the top or near the right-hand side of the screen for a clear understanding of upcoming road plans, while speed and others could be displayed at the bottom of the HUD . Automatic eye height identification should be considered in AR-HUD design. To ensure a certain imaging distance and size, AR-HUD need a wide field of view (FOV) and a long virtual image distance (VID). --the VID should be longer than 7.5 m , FOV should be larger than 10° × 3°
Simplified & Readable elements	The elements in the AR-HUD interface should be straightforward and uncomplicated to provide clear information and reduce drivers' cognitive load. Th elements should try to use graphic design, such as symbol and number , instead of words to represent information, and also combine voice with symbols to better help driver get information about driving. --The font is recommended to choice Roboto , each type of information in AR-HUD display could not be focused on for more than 2 seconds , and the recommended time for HUD animation is less than 500 ms

Based on these recent research studies in the design of AR-HUD interface, it is essential to gain a clear understanding of the diverse and changing user information needs and usage contexts, and more human factors research is necessary to improve the user interface.

References

1. Singh, I.: Optical design and performance evaluation of a dual-beam combiner head-up display. *Opt. Eng.* **35**, 813 (1996)

2. Qin, Z., Lin, F.-C., Huang, Y.-P., Shieh, H.-P.D.: Maximal acceptable ghost images for designing a legible windshield-type vehicle Head-Up Display. *IEEE Photonics J.* **9**, 1–12 (2017)

3. Park, H.S., Kim, K.-H.: AR-based vehicular safety information system for forward collision warning. In: Shumaker, R., Lackey, S. (eds.) *Virtual, Augmented, and Mixed Reality. Applications of Virtual and Augmented Reality*, pp. 435–442. Springer International Publishing, Cham (2014)

4. Ren L: Special research on automobile HUD industry., <https://www.vzkoo.com/document/202211102fecae061dcee2ed96d309462.html> (2024). Accessed 20 April 2024

5. Smith, M., Gabbard, J.L., Burnett, G., Doutecheva, N.: The effects of augmented reality Head-Up Displays on drivers' eye could patterns, performance, and perceptions: *International Journal of Mobile. Hum. Comput. Interact.* **9**, 1–17 (2017)
6. Hwang, Y., Park, B.-J., Kim, K.-H.: The effects of augmented-reality head-up display system on the perception of precautionary situation. In: 2015 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1146–1148. IEEE, Jeju Island, South Korea (2015)
7. Bram-Larbi, K.F., et al.: Reducing driver's cognitive load with the use of artificial intelligence and augmented reality. In: *HCI International 2021 - Late Breaking Papers: HCI Applications in Health, Transport, and Industry*, pp. 235–245 (2021)
8. Schneider, M., Bruder, A., Necker, M., Schluesener, T., Henze, N., Wolff, C.: A field study to collect expert knowledge for the development of AR HUD navigation concepts. In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, pp. 358–362. ACM, Utrecht Netherlands (2019)
9. Gabbard, J.L., Mehra, D.G., Swan, J.E.: Effects of AR display context switching and focal distance switching on human performance. *IEEE Trans. Vis. Comput. Graph.* **25**, 2228–2241 (2019)
10. Li, S., Wang, D., Zhang, W.: Effects of interface complexity of head-up display on drivers' driving performance and self-perception. In: *Presented at the International Conference on Applied Human Factors and Ergonomics (AHFE)* (2021)
11. Gabbard, J.L., Fitch, G.M., Kim, H.: Behind the glass: driver challenges and opportunities for AR automotive applications. *Proc. IEEE.* **102**, 124–136 (2014)
12. Lee, J.M., Ju, D.Y.: Prioritization analysis for contents sensibility evaluation of the future mobility. *KOSES.* **21**, 3–16 (2018)
13. Park, K., Jung, E.S., Im, Y.: Affective design of automotive center fascia display: Focused on regional preference. *JESK.* **38**, 279–289 (2019)
14. Kim, H., Gabbard, J.L.: Assessing distraction potential of augmented reality Head-Up Displays for vehicle drivers. *Hum. Factors.* **64**, 852–865 (2022)
15. Summala, H., Lamble, D., Laakso, M.: Driving experience and perception of the lead car's braking when looking at in-car targets. *Accid. Anal. Prev.* **30**, 401–407 (1998)
16. Haeuslschmid, R., Pflöging, B., Alt, F.: A Design Space to Support the Development of Windshield Applications for the Car. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5076–5091. ACM, San Jose California USA (2016)
17. Park, M., Kim, H.: Deduction & Weight analysis of scenario-based variable information elements for front window display (Head-Up Display) in a Car. *J. Digi. Des.* **12**, 459–468 (2012)
18. Currano, R., Park, S.Y., Moore, D.J., Lyons, K., Sirkin, D.: Little road driving HUD: heads-up display complexity influences drivers' Perceptions of Automated Vehicles. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, Yokohama Japan (2021)
19. Merenda, C., Suga, C., Gabbard, J., Misu, T.: Effects of vehicle simulation visual fidelity on assessing driver performance and behavior. In: 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 1679–1686. IEEE, Paris, France (2019)
20. Guo, H., Zhao, F., Wang, W., Jiang, X.: Analyzing drivers' attitude towards HUD system using a stated preference survey. *Adv. Mech. Eng.* **6**, 380647 (2014)
21. Park, K., Im, Y.: Ergonomic Guidelines of Head-Up Display user interface during semi-automated driving. *Electronics.* **9**, 611 (2020)
22. Ma, X., Jia, M., Hong, Z., Kwok, A.P.K., Yan, M.: Does Augmented-Reality Head-Up Display help? A preliminary study on driving performance through a VR-simulated eye movement analysis. *IEEE Access.* **9**, 129951–129964 (2021)

23. Teng, J., Wan, F., Kong, Y.: Exploring enhancement of AR-HUD visual interaction design through application of intelligent algorithms. *Int. J. Infor. Techno. Syst. App.* **16**, 1–24 (2023)
24. Li, X., Rong, J., Li, Z., Zhao, X., Zhang, Y.: Modeling drivers' acceptance of augmented reality head-up display in connected environment. *Displays*. **75**, 102307 (2022)
25. Yang, H., Zhang, J., Wang, Y., Jia, R.: Exploring relationships between design features and system usability of intelligent car human-machine interface. *Robot. Auton. Syst.* **143**, 103829 (2021)
26. Herbert, A.M., Baschnagel, J., Gordon, M.E., McDonald, Z.: Augmented reality HUDs: warning signs and driver situation awareness. *J. Vis.* **17**, 40 (2017)
27. De Oliveira Faria, N.: Evaluating automotive augmented reality head-up display effects on driver performance and distraction. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 553–554. IEEE, Atlanta, GA, USA (2020)
28. Yu, K., Du, D., Yu, D., Zhi, J., Wang, Y., Jing, C.: Effects of a color gradient and emoji in AR-HUD warning interfaces in autonomous vehicles on takeover performance and driver emotions. *Traffic Inj. Prev.*, 1–10 (2024)
29. Zhan, T., Yin, K., Xiong, J., He, Z., Wu, S.-T.: Augmented reality and virtual reality displays: perspectives and challenges. *iScience*. **23**, 101397 (2020)
30. Park, H.S., Park, M.W., Won, K.H., Kim, K., Jung, S.K.: In-Vehicle AR-HUD system to provide driving-safety information. *ETRI J.* **35**, 1038–1047 (2013)
31. Heo, J., Lee, H., Yoon, S., Kim, K.: Responses to take-over request in autonomous vehicles: effects of environmental conditions and cues. *IEEE Trans. Intell. Transport. Syst.* **23**, 23573–23582 (2022)
32. Park, B.-J., Lee, J.-W., Yoon, C., Kim, K.-H.: Augmented reality and representation in vehicle for safe driving at night. In: 2015 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1261–1263. IEEE, Jeju Island, South Korea (2015)
33. Wiesner, C.A., Ruf, M., Sirim, D., Klinker, G.: 3D-FRC: depiction of the future road course in the Head-Up-Display. In: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 136–143. IEEE, Nantes (2017)
34. Pampel, S.M., Lamb, K., Burnett, G., Skrypchuk, L., Hare, C., Mouzakitis, A.: An investigation of the effects of driver age when using novel navigation systems in a Head-Up Display. *Presence Teleop. Virt.* **27**, 32–45 (2018)



How to Make Autonomous Driving More Acceptable and Safer: Using an Augmented Highlighting Display

Minwoo Cho  and Dongwook Hwang  

Kwangwoon University, Nowon-Gu, Seoul 01897, Korea
dongwookkr@kw.ac.kr

Abstract. The study explores the implementation of highlighting display in autonomous vehicles windshield to enhance reaction, acceptance and workload. The display colors the surface of surrounding vehicles based on their safety levels, aiming to improve occupants' understanding of the autonomous system's awareness. Research investigates the effectiveness of the highlighting display in reducing anxiety and workload, increasing acceptance (especially safety), and enhancing response capabilities in takeover situations. Highlighting display enhanced reaction efficiency and had a positive psychological impact on users. However, this did not reduce the workload, which seems to be caused by the highlighting display reducing the sense of depth.

Keywords: Autonomous Vehicle · Highlighting Display · Trust · Forward Windshield Display

1 Introduction

When an occupant engages in non-driving related activities during the SAE Level 3 automated driving situations, the occupant must be capable of resuming control of the vehicle in a critical takeover scenario and switching attentional allocation for rapidly reconstructing lost awareness. In order to effectively design the takeover control and maximize the occupants' perceived safety, it is important to understand cognitive and attentional factors that affect an occupant's capacity to promptly resume control.

In order to identify those factors, this study introduces a new concept of a forward windshield display for autonomous vehicles which integrates an augmented highlighting visualization onto the forward leading vehicles. This highlighting display is designed to improve occupants' situational awareness and facilitate prompt responses to the surrounding context, ultimately cultivating the occupant's confidence and safety in autonomous vehicles. This study also investigated the impact of the highlighting display in terms of mental workload, response, acceptance and perceived usefulness.

2 Literature Review

2.1 Autonomous Vehicle and Takeover Request

Autonomous driving research has focused on human-machine interactions during takeover requests (TOR) when the system deems the situation too complex. Studies have measured mental workload as an indicator of these interactions [1], and examined response times based on the duration of autonomous driving [2, 3]. Research has also explored information delivery systems for TOR, including multi-modal designs with acoustic, visual, and haptic feedback [4]. Visual feedback is effective; however, it can be overlooked by inattentive drivers, prompting investigation of ambient lighting designs to notify even distracted users [5, 6]. Considering information clarity and response time, visual feedback remains the most effective approach.

The TOR methods for autonomous vehicles are directly linked to safety, as users' reluctance to embrace these technologies stems from concerns about safety and trust [7]. Studies have shown that drivers expect to have various user interface (UI) elements, including hazard information, traffic regulations, and driving behavior explanations [8]. This highlights that drivers would prefer to have feedback on surrounding dangers and hazard avoidance due to their hesitancy to fully trust autonomous systems. Also, it is known that displaying the vehicle's status and pathways plays a crucial role in fostering trust among other road users, such as bicyclists [9]. Similarly, research has found that an AR visualization system which displays and highlights outer objects not only enhances perceived safety but also decreases cognitive load [10]. This may highlight the notion that in the context of autonomous driving, the invisibility of the vehicle's decisions to both occupants and other road users can engender anxiety, and showcasing the vehicle's state plays a crucial role in building confidence [11].

Building on these foundations, this study proposes a highlighting display feature that assigns colors to vehicles based on their safety levels. This feature aims to reduce occupant's anxiety, increase acceptance, and decrease workload. By assessing workload and response accuracy, the study seeks to verify whether the highlighting display feature can improve occupants' response capabilities, particularly in critical situations.

3 Methods

3.1 Participants

A total of 62 students voluntarily participated in the study after seeing the recruitment announcement. Among them, 42% (26 participants) were male, and 58% (36 participants) were female, with an average age of 22.8 years ($SD = 4.08$). All participants had no prior experience with virtual reality (VR) and signed a consent form before the experiment. Among the participants, 58% did not wear glasses, 29% wore glasses, and 13% wore contact lenses.

3.2 Procedure and Stimuli

This study investigated the impact of windshield highlighting display in terms of reaction time, perceived acceptance, usefulness, and workload in a VR autonomous driving



Fig. 1. VR driving simulation setup.



Fig. 2. Surface highlighting display on the forward vehicle.

simulation as shown in Figs. 1 and 2. The simulation was developed using Unity editor (2020.3.37f1) and featured a realistic virtual environment and vehicle controls. High-lighting display shows color-coded vehicles which may provoke accidents based on their distance, with closer objects appearing in red and farther objects in blue (see Fig. 3).

The experiments were conducted to demonstrate the effect of the presence of high-lighting display. The participants drove in a VR setting and during the driving scenario,

they have experienced critical driving situations twice: TOR with and without the highlighting display. The order of the conditions was counterbalanced to mitigate order effects. Each driving session lasted approximately 1.5 min and included a sudden braking event and a takeover situation with measuring participants reaction time, requiring manual control (see Fig. 3). Participants completed surveys after each session to assess their perceived acceptance towards autonomous driving systems, perceived usefulness of the highlighting display and their workload.

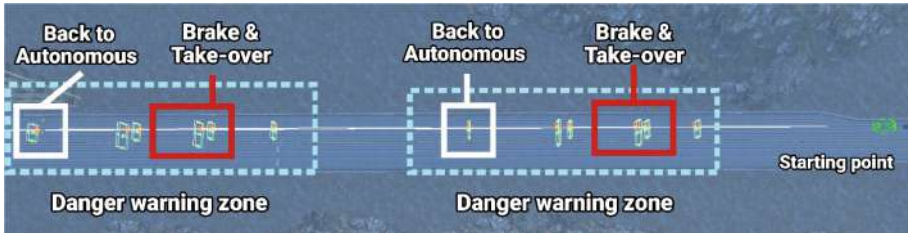


Fig. 3. The overview of driving simulation. Blue dotted line is danger warning section. Red square is braking and take-over point. White square is back to autonomous point.

3.3 Measurements

For computing the participants' actual reaction time, the time for the onset of a critical situation until pressing the brake pedal is measured. For perceived acceptance, the Acceptance of Autonomous Vehicles (AVAM) questionnaire [12] was implemented. The questionnaire items include expectation, self-efficacy, anxiety levels, intention to use, and perceived safety. Also, the perceived usefulness (functional usefulness) was measured using a seven-item survey [13] for examining how psychological factors affecting acceptance of autonomous driving technology, such as anxiety and safety. To assess whether the highlighting display function reduced the workload for participants, the NASA-TLX survey was conducted. Additionally, participants' levels of motion sickness were measured using a modified version of the Simulator Sickness Questionnaire (SSQ). The survey items were based on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree), and the average scores were calculated.

4 Results

In this study, a Paired t-test was conducted to examine the effect of highlighting display as compared with the absent of highlighting display. A statistically significant improvement in reaction time was found with the highlighting display. The response time was better when the highlighting display was present ($M = 1.11$, $SD = 0.45$) compared to when it was absent ($M = 1.66$, $SD = 0.59$), $t(29) = -4.83$, $p < 0.001$. Statistically significant differences were observed across all AVAM subscales. Expectation was higher when the highlighting display was present ($M = 6.47$, $SD = 0.60$) compared to when it was

absent ($M = 5.63$, $SD = 1.25$), $t(29) = 3.90$, $p < 0.001$. Attitude was higher when the highlighting display was present ($M = 6.27$, $SD = 0.78$) compared to when it was absent ($M = 5.73$, $SD = 1.13$), $t(29) = 2.54$, $p < 0.05$. Self-efficacy was higher when the highlighting display was present ($M = 6.48$, $SD = 0.76$) compared to when it was absent ($M = 5.98$, $SD = 1.05$), $t(29) = 2.79$, $p < 0.01$. Anxiety was lower when the highlighting display was present ($M = 2.74$, $SD = 1.11$) compared to when it was absent ($M = 3.27$, $SD = 1.28$), $t(29) = -2.70$, $p < 0.05$. Intention to use was higher when the highlighting display was present ($M = 6.20$, $SD = 0.89$) compared to when it was absent ($M = 5.73$, $SD = 1.29$), $t(29) = 2.25$, $p < 0.05$. Perceived safety was higher when the highlighting display was present ($M = 5.58$, $SD = 0.90$) compared to when it was absent ($M = 5.14$, $SD = 1.02$), $t(29) = 3.16$, $p < 0.01$. A statistically significant difference was also found in perceived usefulness, with higher usefulness when the highlighting display was present ($M = 6.10$, $SD = 0.72$) compared to when it was absent ($M = 5.55$, $SD = 0.93$), $t(29) = 4.22$, $p < 0.001$. In the NASA-TLX, a significant difference was observed only in the performance subscale. Performance was higher when the highlighting display was present ($M = 6.27$, $SD = 0.83$) compared to when it was absent ($M = 5.80$, $SD = 1.06$), $t(29) = 2.31$, $p < 0.05$.

For the SUS, both the absence ($M = 81.79$, $SD = 13.97$) and presence ($M = 85.92$, $SD = 10.76$) of the highlighting display indicated no usability issues. At SSQ, total severity score was modest with no major concerns ($M = 21.44$, $SD = 24.13$).

5 Discussion

In summary, the primary findings of the research indicate that the highlighting display feature positively impacts users' cognition and reactions. In terms of response efficiency, the existence of highlighting displays demonstrated significant improvements in response time. Also, using the highlighting display impact on all AVAM items and also significantly influenced perceived usefulness. This indicates that highlighting display did provide sufficient psychological impact or assistance to users. However, in the NASA-TLX items, only the performance item showed significant improvement. This suggests that participants did not benefit from the workload with the highlighting display despite an increase in overall performance. This result may be affected by 3D depth perception. Humans perceive depth in 3D images through overlapping cues and protrusion along the Z-axis, with Z-axis protrusion and movement being particularly crucial for depth perception [14, 15]. The highlighting display, by covering the vehicle's surface, reduces contrast and depth cues, which making the vehicle appear flatter. This could require occupants to exert more effort in judging the distance and extent of protrusion and movement of the vehicle ahead, potentially leading to additional cognitive work. However, high visibility of the highlighting display's function might have influenced users psychologically by making it easier for them to recognize the existence and impact of the feature.

Overall, this research investigates how highlighting display features contribute to improving users' response efficiency in autonomous vehicles. It revealed that highlighting display have a positive impact on user acceptance, and psychological responses in autonomous vehicles. The highlighting display did not significantly reduce workload, but it effectively increased user acceptance and perceived usefulness.

Acknowledgment. This research was conducted with a Research Grant from Kwangwoon University in 2025.

References

1. Chen, W., Sawaragi, T., Hiraoka, T.: Driver's mental workload measurement concerning cognitive channels. In: 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE) (2020)
2. Gold, C., Damböck, D., Lorenz, L., Bengler, K.: "take over!" how long does it take to get the driver back into the loop? In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, pp. 1938–1942 (2013)
3. Funkhouser, K., Drews, F.: Reaction times when switching from autonomous to manual driving control. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 60, pp. 1854–1858 (2016)
4. Olaverri-Monreal, C., Jizba, T.: Human factors in the design of human–machine interaction: an overview emphasizing V2X communication. *IEEE Trans. Intell. Veh.* **1**, 302–313 (2016)
5. Morales-Alvarez, W., Sipele, O., Léberon, R., Tadjine, H.H., Olaverri-Monreal, C.: Automated driving: a literature review of the take over request in conditional automation. *Electronics* **9**, 2087 (2020)
6. Capalar, J., Olaverri-Monreal, C.: Hypovigilance in limited self-driving automation: Peripheral visual stimulus for a balanced level of automation and cognitive workload. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) (2017)
7. Muir, B.M., Moray, N.: Trust in automation. part II. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*. **39**, 429–460 (1996)
8. Lindemann, P., Lee, T.-Y., Rigoll, G.: An explanatory windshield display interface with augmented reality elements for urban autonomous driving. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). (2018)
9. Hou, M., Mahadevan, K., Somanath, S., Sharlin, E., Oehlberg, L.: Autonomous vehicle-cyclist interaction: peril and promise. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020)
10. Colley, M., Eder, B., Rixen, J.O., Rukzio, E.: Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (2021)
11. Müller, T., Colley, M., Dogru, G., Rukzio, E.: AR4CAD: Creation and Exploration of a Taxonomy of Augmented Reality Visualization for Connected Automated Driving. *Proc. ACM Hum. Comput. Interact.* **6**, 1–27 (2022)
12. Hewitt, C., Politis, I., Amanatidis, T., Sarkar, A.: Assessing public perception of self-driving cars. In: Proceedings of the 24th International Conference on Intelligent User Interfaces (2019)
13. Panagiotopoulos, I., Dimitrakopoulos, G.: An empirical investigation on consumers' intentions towards autonomous driving. *Transp. Res. Part C: Emerg. Technol.* **95**, 773–784 (2018)
14. Park, J.: The influence of depth on visual attention in stereoscopic images. *J. Digital Des.* **10**, 441–450 (2010)
15. Park, J.: The influence of movement direction on visual attention in stereoscopic images. *J. Digital Des.* **9**, 227–238 (2009). Author, F.: Article title. *Journal* 2(5), 99–110 (2016)



Vision and Brain Dynamics: Investigate Controllers' Performance during Emergency Situations in the Next Era of Air Traffic Control

Donglin Li¹ , Fan Li^{1,2} , Yanjie Zhang¹ , Shanshan Feng³ , and Su Han¹

¹ Department of Aeronautical and Aviation Engineering,
The Hong Kong Polytechnic University, Hong Kong, China
fan-5.li@polyu.edu.hk

² Research Centre Data Science AI, The Hong Kong Polytechnic University, Hong Kong, China

³ Wecar Technology Co., Ltd., Shenzhen, China

Abstract. With emerging forms and conventional air traffic operating in the Next Era of Air Traffic Control (ATC) mixed environment, airspace is more complicated and increased emergencies. Air Traffic Controller (ATCO) must manage any situation effectively to ensure safety. Such challenges make human performance in handling emergency situations a critical factor in determining whether an accident occurs or the extent of its severity. This study aims to utilize functional Near-Infrared Spectroscopy (fNIRS) and eye-tracking technology to investigate the performance of ATCOs when handling emergency situations within the Next Era ATC mixed environment and recognize the connection between the biomarkers and performance levels. Sixteen participants engaged in an ATC task and managing emergencies by using a high-fidelity simulation platform. An ATCO expert assessed their performance and categorized them into two groups. The preliminary results show the significant differences in brain and vision dynamics between well performing and poor performing groups. The well performers exhibited elevated brain activation relative to their poor performing counterparts and displayed higher fixation entropy. The findings demonstrate a unique correlation between brain activity, visual behavior patterns, and performance levels in response to emergency situations. By leveraging fNIRS and eye-tracking technologies, we can provide quantifiable insights into performance efficiency. This approach could establish a foundational framework for further exploratory research into ATC performance assessment and training strategies.

Keywords: Emergency management · Human performance · Next Era of Air Traffic Control · fNIRS · Eye-tracking

1 Introduction

The aviation industry is experiencing an upward trend, with transport volume expected to reach a record high. The potential for a low-altitude economy is significantly propelling the expansion of Unmanned Aerial Vehicles (UAVs) and other forms of air mobility, and space operations have emerged as popular trends in human exploration as well [1, 2].

This transformation has directly increased the structural complexity of airspace which makes it more congested [3]. Conflicts and other emergency situations may become more frequent. A review of Federal Aviation Administration (FAA) data shows that the frequency of hazardous situations involving commercial aircraft and other aircraft has increased significantly compared to the past [4], this will become a norm in the Next Era of Air Traffic Control (ATC) environment. Special emergency situations in this context may introduce new human errors, and these errors upgrading easily leads to severe consequences [5, 6]. Maintaining and optimizing human performance becomes a crucial compensation mechanism. Effective understanding of human behavior is one of the key strategies to control human performance, which is a foundational assurance for safety in the ATC system.

Human performance reflects the quality and efficiency of task execution, with ATCOs being central to both identifying and managing all risks within the system. However, they are also the most variable element, influenced by many factors such as vigilance and attention [7]. In emergency situations, maintaining vigilance will avoid loss of situation awareness, impaired vigilance leads to a degradation of the monitoring processes involved in supervisory tasks [8]. In case lose the overall understanding of the situations, the “out-of-the-loop (OOTL)” phenomenon appearing that contributes to termination of the interaction between the controller and the system [9]. In addition, concentrating attention enables controllers to sift through a vast amount of input and focus on key information to process effectively [10]. ATCOs need to face with many information from different sources, but assigning the correct priorities is the essence of ATC work [11]. This places high demands on attention allocation, where reasonable allocation can help improve overall work efficiency and reduce the occurrence of errors. These fundamental human behaviors could indicate various states and provide crucial feedback for performance.

Recent scientific literature has shown that incorporating neurophysiological measures in behavioral research enhances the reliability and sensitivity of findings, providing stronger validation compared to traditional methods such as questionnaires [12]. The eye movement and brain activity are effective representations for revealing human behavioral patterns. Eye movement not only reflects states such as vigilance, fatigue, situational awareness, and detection failures [13–15], but it can also be quantified as a reference. There are numerous eye movement features that have been identified, such as blink rate, saccade, and fixation are directly related to cognitive states, depth of information processing, and concentration of attention [16]. Especially in ATC environments, controllers must process large amounts of information and allocate their attention appropriately. The fixation reflects an individual’s sustained attention to a specific visual target, and its variations have been proven to distinguish operations between high-level athletes and professional medical practitioners [17, 18]. Entropy in information theory provides a unique perspective to analyze ATCO’s visual strategies when performing tasks, and it quantifies the complexity and randomness of eye movement [19]. Researchers have found that driver’s eye movement entropy can be possible to accurately evaluate their attention allocation and information processing efficiency under different load conditions [20], this makes fixation entropy a powerful feature to explain human visual behavior. Functional near-infrared spectroscopy (fNIRS) is a non-invasive neuroimaging technology

extensively used in neuroscience, noted for its excellent temporal and spatial resolution. It allows for real-time monitoring and analysis of brain oxygen-dependent reactions, enabling inferences about an operator's brain behavior patterns during complex tasks [21]. fNIRS can be used in a passive Brain-Computer Interface (BCI) that monitors the occurrence of the OOTL phenomena [22], especially in emergency situations to ensure vigilance is maintained. The patterns of brain behavior exhibited by fNIRS have been shown to help classify an operator's ability to perform a specific task and evaluate different groups' cognitive abilities [23]. This highlights the unique advantages of fNIRS in understanding human cognitive and behavior states.

This study aims to explore the relationship between vision and brain dynamics behavior and the ATCO's performance. We analyzed the blood oxygen levels measured in 63 channels in fNIRS and obtained activation responses in each brain region. At the same time, we also applied the theory of eye movement entropy to analyze the participants' visual behavior. Based on this internal correlation, we can select the specific behavior patterns to establish a metric to distinguish between well and poorly handling performance in emergencies and laying the foundation for objective evaluation structure.

2 Methodology

2.1 Participants

Sixteen students (3 females and 13 males) participated in the experiment. Their age ranged from 18 to 26 years ($M = 21.4$, $SD = 1.63$) and they had normal or corrected-to-normal vision. The information consent was obtained from all participants prior to the data collection. They received coupons as compensation upon completion of the experiment. This study was conducted in strict adherence to the ethical guidelines approved by the institutional review board of the Hong Kong Polytechnic University (HSEARS20211117002-01).

2.2 Apparatus

In this experiment, brain signals were measured using a floor-standing fNIRS device (NirScan, manufactured by Danyang Huichuang Medical Equipment Co., Ltd., PR China) that offers a sampling frequency of 11 Hz and operates at wavelengths of 760 nm and 850 nm. It has 63 channels, as shown in Fig. 1. Eye movement data were collected by the Gazepoint 3 eye tracker, operating at a sampling frequency of 60 Hz. This is a desk-based, contactless eye-tracking device that requires a connection to a computer. The Euroscope software is a high-fidelity radar approach simulator and has the same interface as ATCO using in their realistic work. The Fig. 1 shows the whole experiment setting.

2.3 Design

In the experimental process, participants wore the fNIRS headpiece on their heads and sat in front of a computer equipped with the eye tracker. The eye tracker device needs to

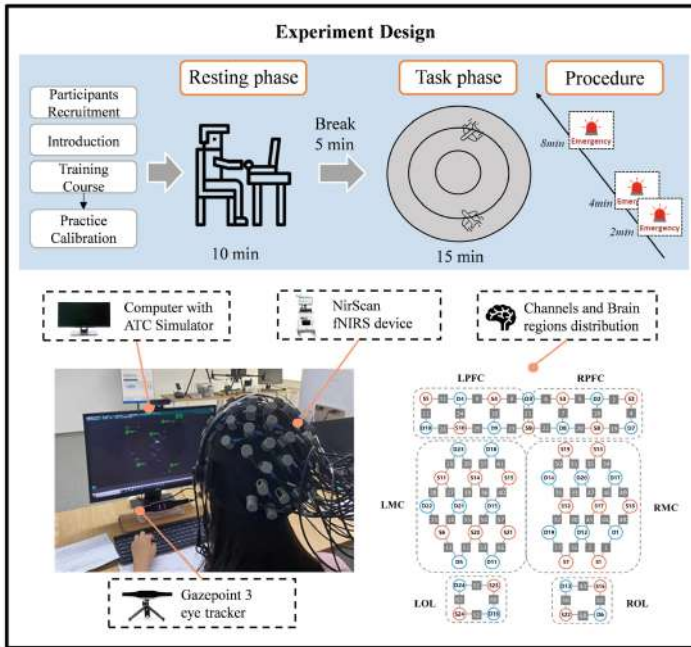


Fig. 1. The ATC experiment setting and design plot

aim at the eyes of the subjects at an upward angle. The overall experiment involved three phases, the training phase, resting phase, the task state phase, lasting for 45 minutes roughly, Fig. 1 shows the whole experiment process.

Before the commencement of the experiment, three training ATC sessions were provided. The ATC task includes managing several flights and three random emergency situations. In the designed scenario, the airspace is characterized by a diverse array of activities, with the emergence of various stakeholder forms all intertwined. These elements align seamlessly with the Next Era ATC environment. The emergency situations include UAV interference, engine failure, and cabin decompression. Furthermore, to aid the participants in navigating handling these emergency situations, an emergency checklist was prepared as a reference tool.

For assessing the participants' emergency situations handling performance, a senior ATCO expert rated the performance across the following five critical dimensions: radio-telephony communication, control consciousness, coordination, adaptability, and compliance with procedures. Each of these sections is worth 10 points and is scored by an ATCO expert. Those five selected dimensions ensure the most thorough assessment of emergency management skills. According to the evaluation ratings from the ATCO expert, the participants were divided into two groups, namely well and poor handling groups.

2.4 Data Analysis

In the subjective performance processing phase, we summarized the total scores of all participants on the performance evaluation table. Based on the scores, the participants were divided into two groups which are named well and poor. We conducted a statistical report of the mean scores and standard deviation of the two groups on each assessment indicator and total scores. For objective measures, one minute was specifically selected for metric analysis during each emergency.

The fNIRS data were processed using the Homer 3 software package [24], which incorporated several internal functions within MATLAB version R2020a. Among the parameters measured by fNIRS, the concentration of oxygenated hemoglobin (HbO) is strongly correlated with neural activity. Thus, the changes in the HbO concentration were considered the best indicator of brain activity. The wavelet method effectively reduced outliers, physiological noise, motion artifacts due to breathing, blood pressure, and heartbeat. The raw light intensities were initially converted into optical densities. The bandpass filter with a cutoff frequency from 0.01 to 0.08 Hz further cleansed the data. Following the principles of the modified Beer-Lambert Law (MBLL), which were then further processed to obtain changes in HbO concentration. The detailed notation is shown in eq. (1).

$$\Delta OD_{HbO}^{\lambda} = \epsilon_{HbO}^{\lambda} \cdot \Delta c_{HbO} \cdot d \cdot DPF^{\lambda} \quad (1)$$

The $\Delta OD_{HbO}^{\lambda}$ is the optical density refers to the change in light absorption. The ϵ_{HbO}^{λ} represents the absorption coefficients of HbO. λ represents the wavelength of the two kinds of near-infrared light emitted, d represents the distance between the transmitter and the receiver sensor, DPF is the differential path factor, and Δc_{HbO} represent the relative concentration changes of HbO. After data pre-processing, the variation of HbO parameters was calculated between the resting state and the task state. Using a block paradigm and averaging to obtain the hemodynamic response function (HRF), we employed the generalized linear model (GLM) to fit and analyze the HRF function. Equation (2) and (3) describe in detail the steps of data fitting through a GLM and illustrate the model parameters derived from this process. The parameter beta value as the activation coefficient, which represents the level of activation that represents the brain's behavioral pattern. We mapped all channels to six brain regions to analyze activation. The brain regions were divided into Left Prefrontal Cortex (LPFC), Right Prefrontal Cortex (RPFC), Left Motor Cortex (LMC), Right Motor Cortex (RMC), Left Occipital Lobe (LOL), Right Occipital Lobe (ROL), as depicted in Fig. 1. To facilitate a more intuitive comparison of these values, beta coefficients corresponding to channels within the same cortical region were average.

$$Y = X\beta + \varepsilon \quad (2)$$

$$\beta = X^*Y, \quad X^* = (X^T X)^{-1} X^T \quad (3)$$

The eye movement data was outputted by Gazepoint 3 software. The fixation behavior-based entropy theory was examined. The entropy of fixations may be measured with

the Shannon entropy equation. The fixation entropy H_g is computed as eq. (4). Where X represents the total number of fixations, denoted as n , and each fixation is indexed by i , indicating its location within X . The variable p represents the probability of the controller's fixation at each specific position i . To calculate the fixation entropy, we first discretized the radar screen into 64 blocks. The fixation frequency for each block can then be directly obtained at any given time from the data.

$$H_g(X) = -\sum_{i=1}^n p(i) \cdot \log_2 p(i) \quad (4)$$

In the process of statistical analysis, a normality test was performed first to evaluate the distribution characteristics of the data. If the data is normally distributed and satisfies assumptions like homogeneity of variance, then a parameter test is applied, such as the independent sample t-test. If the data does not conform to a normal distribution, a non-parametric test like the Mann-Whitney U test is used. Figure 2 described the specific data process workflow briefly.

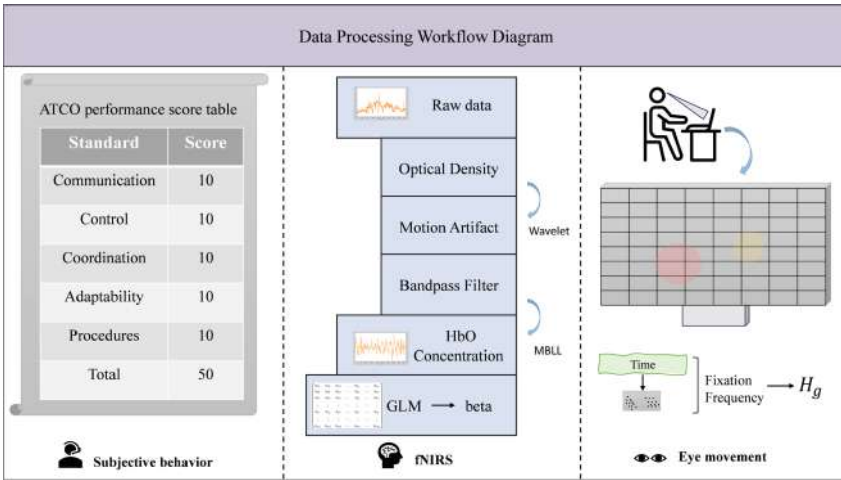


Fig. 2. The data processing technical workflow

3 Result

3.1 Subjective Performance Measurement

In the participant's assessment, the performance during emergency situations was evaluated based on five critical dimensions. The well performing group exhibits strong emergency management skills compared to poor performing group. Table 1 shows the scores of the two groups in each evaluation dimension and sum score of emergency management efficiency.

Table 1 The detailed score of Well and Poor performing group

Category	Well group		Poor group	
	Mean	Standard deviation	Mean	Standard deviation
Communication	5.75	1.03	3.75	0.71
Control	5.50	0.53	4.25	0.46
Coordination	5.50	0.93	4.40	0.46
Adaptability	6.00	0.35	4.13	0.53
Procedure	5.85	0.53	4.70	0.83
Total	35.06	3.04	24.63	2.33

3.2 Brain Dynamics

In the hemodynamic responses within the brain, the beta coefficient represented the estimated prediction of the HbO signals in HRF. It served as a crucial indicator of cortical activation intensity, and it was mapped in six brain regions to analysis. After conducting statistical analysis, no significant differences were found between the two groups. The summarized data are presented in Table 2, illustrating the regional mean and standard deviation values which were measured in two groups differentiated by performance levels. In response to emergency situations, the well performing group exhibits broader activation across a range of brain regions and higher beta values. In the well performing group, the highest value was 0.6513 in the LPFC, and the lowest value was 0.6398 in the LOL. In the poor performing group, the highest value was 0.5897 in the LPFC, and the lowest value was 0.5007 in the LOL. Overall, the poor performers exhibited lower levels of global brain activation compared to their well performing counterparts. Figure 3 displays visual brain maps illustrating the changes in activation across different brain regions for two groups during the experiments. Each channel is denoted by a small circle on the maps. The color intensity indicates the activation changes: a redder color signifies a greater increase in activation, while a bluer color indicates decrease in activation. The maps are presented in both front (a) and rear (b) views for the well performing group, and front (c) and rear (d) views for the poor performing group.

3.3 Vision Dynamics

An independent sample t-test was used to analyze the data, which conformed to normality assumptions, revealing significant differences between the two groups ($p = 0.028$). Figure 4 illustrates the fixation entropy, comparing the mean values, standard deviations, and data distributions between the well performing and poor performing groups. Specifically, the well performing group exhibited higher entropy. In contrast, the poor performing group showed lower values of fixation entropy overall.

Table 2. Relevant data of 6 brain regions with beta value between Well and Poor performing group

Brain Regions	Number channels	Well group		Poor group	
		Mean	Standard deviation	Mean	Standard deviation
LPFC	10	0.6513	0.3259	0.5897	0.3357
RPFC	10	0.6507	0.3239	0.5642	0.3515
LMC	18	0.6503	0.3224	0.5392	0.3374
RMC	18	0.6488	0.3227	0.5875	0.3497
LOL	4	0.6398	0.3298	0.5007	0.3309
ROL	4	0.6446	0.3211	0.5547	0.3239

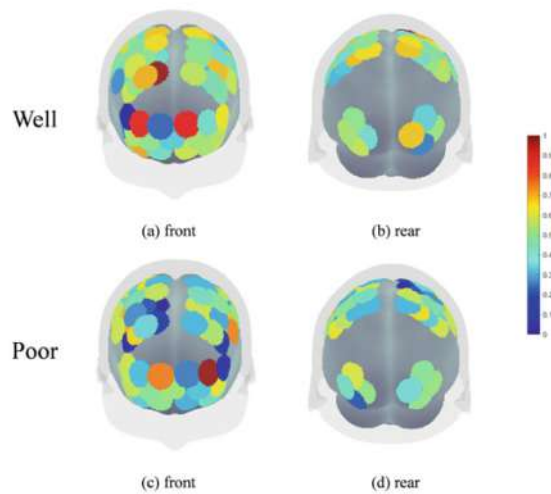


Fig. 3. The brain activation of between the Well and Poor performing participants

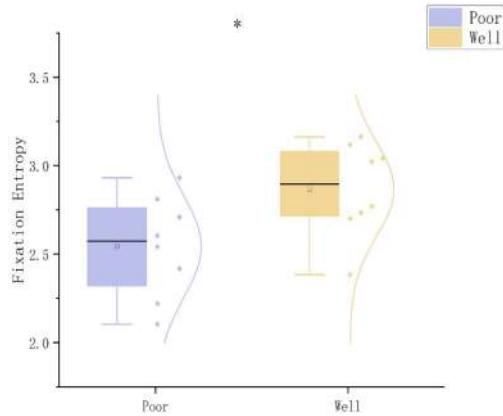


Fig. 4. The fixation entropy between Well and Poor performing group

4 Discussion

This study investigates the brain and vision behavior dynamics of ATCOs using fNIRS and eye-tracking technology while they execute ATC tasks, particularly focusing on emergency scenarios. The aim is to identify the underlying connection between biomarkers and performance levels. This experiment was conducted using a high-fidelity ATC approach simulator, with efforts made to replicate realistic working conditions. Analysis of the data revealed significant differences in brain activation and vision behavior between the two groups categorized based on performance levels. These findings revealed the unique mechanism of ATCOs in emergency situations, contributing to a deeper understanding of the controller's performance.

For the brain dynamics patterns, there are differences in well performing and poor performing ATCOs. The result shows the more widespread and intense activation in every region of the well performing group, with higher beta values than the poor performing group. This brain pattern may imply that the ATCOs who are performing well show more effective utilization of these cognitive resources in emergency situations. During the emergency, ATCOs are exposed to more warnings and incoming auditory information, and the widespread activation of the brain would help in the efficient flow of information through the system [25]. On the other hand, the decrement of activation in the poor performing group discloses deficits in these crucial cognitive abilities, which eventually make the ATCOs perform poorly during the emergency and fail to understand the whole picture of the environment. In addition, the poor performing group showed low activation in the LOL region. These regions are associated with visual information processing [26], which is a possible reason to miss the key warnings further leading to detection failures. For vision dynamics, the fixation entropy serves as a robust measure of the variability and complexity of fixation patterns. The higher entropy observed in the well performing group suggests that they have more diverse scanning strategies and the ability to spread their attention across the whole picture efficiently [27]. The ATC environment needs to process multiple visual information quickly and efficiently, which also indicates that they

have enough situational awareness to make the right decisions. Conversely, the lower fixation entropy values seen in the poor performing group indicate a more constrained or repetitive pattern. This corresponds to the poor performance of the above brain behavior patterns in visual information processing. This may reflect their less-than-optimal use of visual attention resources, which may make it difficult for them to navigate complex visual environments, resulting in lower performance [28].

These findings contribute to a better understanding of the bio-mechanisms associated with performance in complex tasks. Future research could further explore the causal relationships between specific patterns of vision and brain dynamics and performance outcomes. The difference in the brain and vision behavior can be used as a distinguishing biomarker of ATCO's performance that can design a more effective training program. The experiment had insufficient participants, and they were all students and not professional ATCOs so that potentially leading to suboptimal data quality in certain metrics. Increasing the number of ATCOs participating in the future would help confirm the generalizability of our findings. Additionally, another aspect of future research could explore other potential physiological and behavioral indicators, such as heart rate variability, electroencephalogram (EEG), or galvanic skin response, to gain a more comprehensive understanding of ATCOs' performance in emergency situations.

5 Conclusion

In summary, this study provided an insightful comprehension of the ATCO's performance and vision and brain dynamics under emergency situations, considering the complexity and diversity in the Next Era ATC airspace environment. We conducted a real ATC task with a high-fidelity simulator, leveraging fNIRS and eye tracker to collect the data. After the analysis, we observed significant differences in vision and brain dynamics based on varying performance levels. It can not only enhance our understanding of biomarkers of ATCOs' performance in emergency situations, but more meaningfully, it offers a practical foundation for improving training strategies and assessment performance. By applying brain and vision dynamics to the aviation industry, this study offered a new perspective on understanding human performance in high-risk ATC environments.

Acknowledgement. This research study is partially supported by the Hong Kong Polytechnic University (P0038933 & P0038827), by Research Centre Data Science AI (P0046166). The manuscript is proofread by ChatGPT 4.0.

References

1. IATA, Global Outlook for Air Transport. 2023
2. Mohsan, S.A.H., et al.: Unmanned aerial vehicles (UAVs): practical aspects, applications, open challenges, security issues, and future trends. *Intell. Serv. Robot.* **16**(1), 109–137 (2023)
3. Djokic, J., Lorenz, B., Fricke, H.: Air traffic control complexity as workload driver. *Trans. Res. Part C: Emerging Techno.* **18**(6), 930–936 (2010)
4. FAA. Aviation Safety Reporting System (ASRS). 2024.; <https://www.asias.faa.gov/apex/f?p=100:20:::NO:::>

5. Danaher, J.W.: Human error in ATC system operations. *Hum. Factors*. **22**(5), 535–545 (1980)
6. Malakis, S., Kontogiannis, T., Kirwan, B.: Managing emergencies and abnormal situations in air traffic control (part I): Taskwork strategies. *Appl. Ergon.* **41**(4), 620–627 (2010)
7. Isaac, A.R., Ruitenbergh, B.: *Air traffic control: human performance factors*. Routledge (2017)
8. Li, F., et al.: Hybrid data-driven vigilance model in traffic control center using eye-tracking data and context data. *Adv. Eng. Inform.* **42** (2019)
9. Di Flumeri, G., et al.: Brain–computer interface-based adaptive automation to prevent out-of-the-loop phenomenon in air traffic controllers dealing with highly automated systems. *Front. Hum. Neurosci.* **13**, 296 (2019)
10. Duncan, J.: Visual search and visual attention. In: *Attention and performance XI*, pp. 85–106. Routledge (2016)
11. Hopkin, V.D.: *Human factors in air traffic control*. CRC Press (2017)
12. Suzuki, Y., Wild, F., Scanlon, E.: Measuring cognitive load in augmented reality with physiological methods: A systematic review. *J. Comput. Assist. Learn.* **40**(2), 375–393 (2024)
13. Li, F., et al.: Hierarchical eye-tracking data analytics for human fatigue detection at a traffic control center. *IEEE Trans. Hum.-Mach. Syst.* **50**(5), 465–474 (2020)
14. Li, Z., Li, Z., Li, F.: Visual Attention Analytics for Individual Perception Differences and Task Load-Induced Inattentional Blindness. In: *International Conference on Human-Computer Interaction*. Springer (2023)
15. Yoong, V.R.Q., Using eye tracking metrics to measure and predict situation awareness of air traffic controllers. 2020
16. Mengtao, L., et al.: Leveraging eye-tracking technologies to promote aviation safety-a review of key aspects, challenges, and future perspectives. *Saf. Sci.* **168**, 106295 (2023)
17. Aksum, K.M., et al.: What do football players look at? An eye-tracking analysis of the visual fixations of players in 11 v 11 elite football match play. *Front. Psychol.* **11**, 562995 (2020)
18. Mohamad Shahimin, M., Razali, A.: An eye tracking analysis on diagnostic performance of digital fundus photography images between ophthalmologists and optometrists. *Int. J. Environ. Res. Public Health.* **17**(1), 30 (2020)
19. Gray, R.M.: *Entropy and Information Theory*. Springer Science & Business Media (2011)
20. Schieber, F., Gilland, J.: Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, CA, Los Angeles, CA (2008)
21. Wilcox, T., Biondi, M.: fNIRS in the developmental sciences. *Wiley Interdiscip. Rev. Cogn. Sci.* **6**(3), 263–283 (2015)
22. Gateau, T., et al.: Real-time state estimation in a flight simulator using fNIRS. *PLoS One.* **10**(3), e0121279 (2015)
23. Yang, D., et al.: Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: An fNIRS study. *Front. Hum. Neurosci.* **13**, 317 (2019)
24. Huppert, T.J., et al.: HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* **48**(10), D280–D298 (2009)
25. Baker, J.M., et al.: fNIRS measurement of cortical activation and functional connectivity during a visuospatial working memory task. *PLoS One.* **13**(8), e0201486 (2018)
26. Wang, Z., et al.: Effects of three different rehabilitation games' interaction on brain activation using functional near-infrared spectroscopy. *Physiol. Meas.* **41**(12), 125005 (2020)
27. Lanini-Maggi, S., et al.: Assessing how visual search entropy and engagement predict performance in a multiple-objects tracking air traffic control task. *Comput. Hum. Behav. Reports.* **4**, 100127 (2021)
28. Shiferaw, B.A., et al.: Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Sci. Rep.* **8**(1), 2220 (2018)



The Acceptability of Automated Vehicles by Other Road Users: Effects According to the Type of Vehicles

Guelle Kevin^{1(✉)}, Erdinger Nicolas^{1,2}, and Metayer Natacha¹

¹ Department of Human Factors and Economics of Sustainable Mobility, VEDECOM Institute, Versailles, France

kevin.guelle@vedecom.fr

² Virtual Center Laval, Arts et Métiers Institute–Arts et Métiers ParisTech, Changé, France

Abstract. Automated Vehicles (AVs), also known as driverless vehicles, are a category of vehicles in which some or all driving decisions are made without human intervention. In addition to the benefits enabled by AVs, such as access to a dedicated vehicle fleet and sharing among users, AVs must be considered in interaction with all forms of mobility: other vehicles, notably motorists (Other Drivers), active mobility such as Cycling or Scootering, and Pedestrians as well. We are interested in the potential interaction, in a future mobility environment, between these different users and AVs through the dissemination of videos and through mental projection. Indeed, it is necessary to anticipate the impact of AVs deployment in the road environment, not only from the perspective of AVs users, but also from the standpoint of other road users who will be required to share this common space and interact together. This study focuses on the acceptability of these other users regarding different types of AVs (i.e., Shuttle, Droid, RoboTaxi, Cyber Valet). We are interested in the assessment made by these users in terms of interaction preferences. Our results show an overall positive evaluation following the observation of these AVs, with preferences varying according to the type of users. Notably, regarding the question of interaction priority, Other Drivers consider themselves less prioritized in their interactions with AVs than Cyclists or Pedestrians.

Keywords: Automated vehicles · Acceptability · Other road users

1 Context and Interests of this Study

This study is part of a research context focusing on the Safety and Acceptability of Autonomous Mobility (SAM project). Within this project, this study aims to investigate the acceptability of other users revolving around various types of AVs in their environment. We are interested in the projected interaction between these other users and these vehicles, in particular to understand the ergonomics integration needs.

2 Theoretical Anchorage: From Situated Acceptability to Mental Projection

Acceptability typically refers to the “degree” of integration and appropriation of an object in a usage context [1]. We can define integration (frameworks of acceptance) as how the product fits into the user’s activities and how it contributes to transforming those activities. On the other hand, we can define appropriation as how the user personally invests in the object, such as extent the object motivates the user to act upon it rather than merely endure it. Moreover, situated acceptability seeks to study what “technology allows/facilitates”, what it “obliges one to do” but also what it “prevents one from doing” or “no longer like before” (p. 12) [2]. In the context of this study, the goal is to understand how AVs projected into environments can modify travel habits, especially depending on the type of users (i.e., Other Drivers, Cyclists and Pedestrians).

Ergonomics also addresses the question of the future by integrating the notion of probable future activity and desirable future activity [3, 4]. Thus, inferring future activity is based on known specifics of the future situation [5]. Our study falls within the realm of future-oriented cognition as well: we examine the cognitive projection of the person into the future by simulating a scenario of realistic activity specific to them [5–7]. The originality lies in projecting participants into a probable universe to generate an exchange based on lived experience [5]. The objective of this study is then to advance knowledge about AVs, particularly to support the designers of these vehicles and cooperation between humans and machines [8].

3 Problem Statement: The Acceptability based on the Type of AV

This study responds to a general research problem formulated as follows: What is **the impact of the type of AVs** (i.e., Shuttle, Droid, RoboTaxi, Cyber Valet) on their acceptability? This question is also complemented by several variables, including the type of users (i.e., Other Drivers, Cyclists and Pedestrians) and the moment when acceptability is assessed (i.e., before the confrontation with the AVs, just after, a few weeks later).

We can hypothesize that the **type of AVs will induce changes in the dimensions of acceptability depending on the assessment time** (before, just after and/or few weeks after confrontation with the AVs). More specifically, we can expect an evolution of acceptability between the *a priori* evaluation and the *a posteriori* evaluation following the viewing of videos and mental projection. Among other things, we can consider that the interaction of a pedestrian will not be the same as that of a cyclist, in the sense that the deviations in the AV’s trajectory or safety will not be perceived in the same way.

4 Methodology

4.1 Types of projected AVs and protocol

The study focuses on the projection of different types of AVs: the Cyber Valet (automated functionality allowing the vehicle to park independently), the RoboTaxi, the Logistics Droid and fully automated Shuttle (Figure 1):



Fig. 1. The AVs presented in the form of videos: from left to right, the Valet Parking, the RoboTaxi, the Droid and the Shuttle.

These AVs are visually projected in the form of videos. The complete protocol is as follows (Table 1):

Table 1. Research Protocol: Various temporalities of the study

1—Before projection of AVs video	2—Just after projection of AVs video	3—A few weeks after projection of AVs video
- Acceptability assessment (= <i>a priori</i> acceptability)	- Acceptability assessment (= <i>a priori</i> acceptability) - Mental projection of interaction with the different types of AVs - Interaction ratings with AVs: score + reasoning	- Acceptability assessment

4.2 Materials and Participants

The assessment of AVs was conducted using a questionnaire with a 6-points Likert scale (from 1—completely disagree to 6—completely agree). This questionnaire was the same between the three steps of the study (i.e., the respondents completed it 3 times). Between step 1 and step 2, the 4 AVs (i.e., Shuttle, Droid, RoboTaxi, Cyber Valet) were presented to participants in the form of short videos (between 1 and 2 min). The AVs were shown moving in varied road environments and interacting with several other road users. A projection interview was also conducted to immerse participants in interaction with the different AVs. Lastly, a scale ranging from 1 (low desire for interaction) to 10 (high desire for interaction) was presented to participants to rank the AVs according to their interaction preference.

49 voluntary participants took part in this study (26 females, 23 males, average age 42.75 ± 13.17 years) divided into three groups: 17 Pedestrians (average age 43.9 ± 13 years), 17 Cyclists (average age 38 ± 15 years), and 15 Other Drivers (average age 48.9 ± 11 years).

5 Results

5.1 Preference and Evaluation of AVs

The results highlight that, once the AV videos were presented and participants projected themselves into interaction with them, all AVs were evaluated very positively (average rating of 8 out of 10) (Table 2).

Table 2. Results of AVs assessment

	Shuttle	Droid	RoboTaxi	Cyber Valet
Mean (<i>SD</i>)	8.25 (1.49)	7.61 (2.19)	8.02 (2.02)	8.10 (1.86)
Min. / Max.	4 / 10	3 / 10	1 / 10	2 / 10

Furthermore, by using a multifactorial ANOVA, no significant results were found between the different objects type of AV ($p > .05$), nor based on the type of user ($p > .05$): all AVs were evaluated in the same way regardless of whether participants were Pedestrians, Cyclists or Other Drivers.

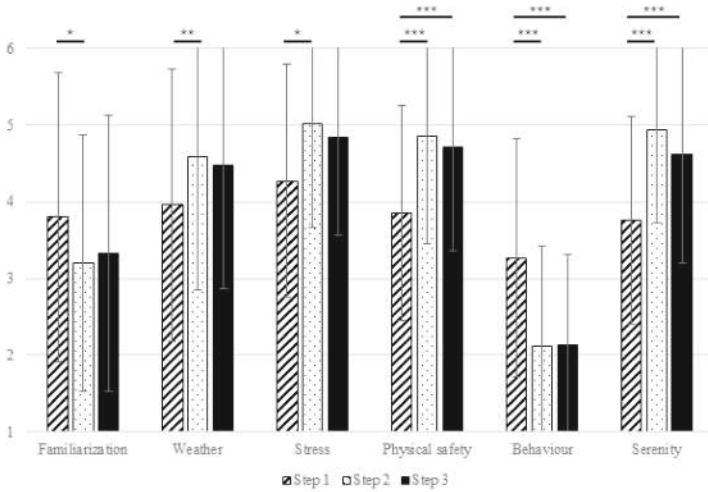
While participants generally did not report preferring interaction among the different AVs, the type of user had an impact on the two preferred AVs with which participants wished to interact. Thus, Pedestrians and Cyclists would prefer to interact first with the Shuttle (respectively, in 69% and 77% of cases, the Shuttle is classified in the first 2 choices of interaction), while Other Drivers would prefer to interact with the Robo-Taxi (ranked 67% in one of the first two places while this proportion is only 38% for Pedestrians).

Three arguments per vehicle were expressed to explain each classification, these arguments were grouped by lexical field into categories to facilitate analysis. It appears that the arguments expressed by Other Drivers are more expressed in favour of usefulness (100%), while other categories are more contrasted depending on the role of the participant (e.g., 17.6% of the arguments are in favour of ecology for Pedestrians, while they represent 6.7% for Other Drivers and 11.8% for Cyclists). Some differences are important, such as safety issues for pedestrians and cyclists which are present in their arguments for 82.4% of them, compared to 66.7% for Other Drivers, or the trust appears to be a more significant theme for Cyclists (82%) compared to Pedestrians (64%) or Other Drivers (47%). The themes mentioned to express preferences for interaction with AVs are not distributed in the same way depending on the type of other users.

5.2 Differential Evaluations based on Participant Profiles and for all AVs

Depending on the type of other road users, Pedestrians, Cyclists, and Other Drivers do not evaluate certain acceptability items in the same way. Thus, regarding the question of interaction priority, Other Drivers consider themselves less prioritized in their interactions with AVs ($M = 2.07$, $SD = 1.32$) than Cyclists ($M = 3.24$, $SD = 1.68$) or Pedestrians ($M = 3.88$, $SD = 1.76$) ($F(2) = 10,197$, $p < .001$, $\eta^2 = .307$); whereas no difference appears between the latter two ($p > .05$).

Depending on the type of moment of the acceptability assessment, some differences appear (Fig. 2).



Notes. * $p < .05$; ** $p < .01$; *** $p < .001$. Familiarization: "I would need time to become familiar with the presence of VA"; Weather: "I would feel more reassured if AVs are used during bad weather conditions compared to vehicles with drivers"; Stress: "Coming across a VA on my journey will not be stressful"; Physical safety: "I would not feel physically vulnerable when having to interact with a VA"; Behaviour: "I wouldn't know what to do with a VA object"; Serenity: "I would be calm when interacting with a VA object that does not have a driver".

Fig. 2. Effect of assessment time on AVs acceptability.

Furthermore, it appears that merely exposure to AVs videos is sufficient to alter respondents' perception durably and positively on several aspects (physical safety, behaviour, serenity), but this will not be sufficient for other aspects where, without reactivation of exposure, evaluations will return to their pre-viewing level (e.g., need to familiarize with the AV, perceived stress).

6 Limitations and Next Steps

While the assessment of AVs is positive regardless of the type of AVs, this study represents a first step in understanding the potential differences in interaction among user types based on the proposed AVs and identifying the main challenges to address based on user types (e.g., anticipated misbehaviours of Cyclists and Pedestrians, the need for communication prior to the introduction of AVs).

To further deepen this work, we plan to evaluate interactions in real-world environments to account for the variety of profiles (e.g., people with disabilities) and confirm, or refute, the results obtained through simple projection with videos. Differentiated studies based on the typology of the territory are also envisaged to measure differences between rural and urban areas, for example. Finally, we aim to study acceptability also in terms of the services offered, as this study focused on a very general vision of AVs without contextualization around specific use cases.

Acknowledgement. This work was supported by the SAM project (Safety and Acceptability Driving and Mobility). Officially launched on June 20, 2019, the SAM project is the result of the EVRA (Experiment of automated vehicles on open roads) call for projects as part of the Investments for the Future Program (PIA) which is at the heart of the national strategy for the development of automated vehicles presented by the Government in May 2018. The project was co-funded by the French Agency for Ecological Transition (Ademe).

References

1. Loup-Escande, E., Loup, G.: Designing acceptable emerging technologies: what contribution from ergonomics? *Theor. Issues Ergon. Sci.* **22**(5), 581–602 (2020). <https://doi.org/10.1080/1463922X.2020.1836568>
2. Bobillier-Chaumon, M.-E.: L'acceptation située des technologies dans et par l'activité: premiers étayages pour une clinique de l'usage. *Psychologie du travail et des organisations*. **22** (2016). <https://doi.org/10.1016/j.pto.2016.01.001f>
3. Sagot, J.-C., Gouin, V., Gomes, S.: Ergonomics in product design: safety factor. *Saf. Sci.* **41**, 137–154 (2003)
4. Daniellou, F.: Simulating future work activity is not only a way of improving workstation design. *ATWAD*. (2007). <https://doi.org/10.4000/activites.1704>
5. Martin, A.: Méthodes d'ergonomie prospective pour l'exploration des besoins futurs: applications à l'hydrogène énergie pour l'habitat, PhD Thesis. University of Lorraine (2021)
6. Schacter, D.L., Benoit, R.G., Szpunar, K.K.: Episodic future thinking: Mechanisms and functions. *Curr. Opin. Behav. Sci.* **17**, 41–50 (2017). <https://doi.org/10.1016/j.cobeha.2017.06.002>
7. Hollis-Hansen, K., O'Donnell, S.E., Seidman, J.S., Brande, S.J., Epstein, L.H.: Improvements in episodic future thinking methodology: Establishing a standardized episodic thinking control. *PLoS One*. **14** (2019). <https://doi.org/10.1371/journal.pone.0214397>
8. Caroly, S., Hubaut, R., Guelle, K., Landry, A.: Le travail digital, un enjeu pour les psychologues du travail: L'exemple de l'industrie 4.0. *Le Journal des psychologues*. **367**, 27–32 (2019). <https://doi.org/10.3917/jdp.367.0027>



The Influence of Advanced Driver Assistance Systems on Fatigue Development Among Long-Haul Truck Drivers

Chunxi Huang¹ , Zuyuan Wang², and Dengbo He^{1,2,3}

¹ The Hong Kong University of Science and Technology, Hong Kong, China
dengbohe@hkust-gz.edu.cn

² The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

³ HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Shenzhen, People's Republic of China

Abstract. To alleviate the workload of long-haul truck drivers and subsequently reduce their fatigue, companies have begun exploring the implementation of advanced driver assistance systems (ADAS) on heavy trucks to aid drivers with certain driving tasks. However, prior research conducted in the passenger car domain has suggested that drivers utilizing ADAS are more susceptible to experiencing fatigue in comparison to manual driving. Therefore, it is imperative to examine the impact of ADAS on the fatigue levels of truck drivers. Based on over 120 h of naturalistic driving experiments, this study employs various eye-tracking metrics (e.g. blink frequency), response time, and subjective ratings to compare the fatigue levels of long-haul heavy truck drivers when driving manually versus driving with the assistance of ADAS. The results indicate that the fatigue levels of long-haul truck drivers are lower when operating vehicles with the assistance of ADAS in comparison to manual driving.

Keywords: Fatigue · ADAS · Truck drivers · Prolonged driving tasks · Manual driving

1 Introduction

Driving fatigue refers to a state in which the driver's level of wakefulness decreases and the degree of drowsiness increases [1]. Driving fatigue widely exists in daily driving and it is not only one of the main causes of rear-end collisions but also the cause of 20%–30% of fatal accidents [2], posing a great threat to road traffic safety. For example, in China, 1.75 million road traffic accidents occurred nationwide in 2023, 15% of which were caused by fatigue driving, accounting for the highest proportion [3].

Driving fatigue is more prominent in logistics industries with high timeliness requirements (such as express delivery). Typically, long-haul truck drivers need to continuously drive heavy trucks traveling long distances (up to 800–1200 km in a single trip) for a long time (about 10–14 h) to deliver goods. These drivers are prone to driving fatigue, causing great potential threats to road traffic safety [4, 5]. For example, a questionnaire study

found that 27.8% of heavy truck drivers often feel fatigued during driving [6]. Friswell and Williamson found that compared with light truck drivers, heavy truck drivers are more likely to experience driving fatigue [7]. Another questionnaire study conducted in China also found that excessive driving time is the leading cause of truck driver fatigue [8].

To alleviate the driving fatigue of long-haul heavy truck drivers during long-distance driving, some companies (e.g., Zhijia, TuSimple) have tried to deploy advanced driver assistance systems (ADAS) on heavy trucks. However, existing studies in the passenger car domain have found that with ADAS, drivers were more likely to experience fatigue compared to manual driving (e.g., [9–12]). Therefore, to ensure long-haul heavy truck drivers' safety when using ADAS, it is imperative to understand the influence of ADAS on heavy truck drivers' fatigue development during long-haul driving. Our study serves as one of the first studies that analyzed the fatigue development of long-haul truck drivers based on real-world data.

2 Methods

2.1 Apparatus

Experiment Vehicle and Experiment Route. A Jie Fang J7 truck (total length: 18 m, total weight: 31.75 t) equipped with SAE L2 ADAS was used in the experiment, which runs on a real operation route of an express company in China, from Suzhou, Jiangsu, to Xiaogan, Hubei, totaling 802 km, of which 785 km was highway. The ADAS includes various functions, including adaptive cruise control, lane centering control, lane departure warning, within-lane trajectory fine-tuning (adjusting lateral position slightly to avoid obstacles), and automatic lane changing (changing lanes when drivers initiate the turning light). The average driving time was 10.8 hours with a standard deviation of 0.36 h. The experiments were conducted on weekdays to avoid the potential influence of traffic flow variations on weekends and holidays. The departure time was set at around 7.30 am to avoid potential sleep-related fatigue

Eye Tracking. A dashboard-mounted remote eye tracking system (Smart Eye Pro with three cameras, Sweden) was used to capture drivers' eye movement behaviors. To ensure accurate gaze capture, gaze calibration was re-conducted whenever the driver left the vehicle for a short break during the experiment. Figure 1 presents the experiment vehicle and the eye tracker settings

Driving Response Task (DRT). The DRT can measure drivers' alertness levels. When the experimenter initiated the stimuli manually, the speaker would emit a "beep" sound with a frequency of 4000 Hz, a duration of 0.5 s, and a loudness of 95 dB. The driver was instructed to press the button as soon as possible when heard the sound, and the response time was recorded. To avoid the influence of too frequent DRT on driver fatigue, the DRT was conducted every two hours. In total, 5 DRTs were conducted during the 10 hours of driving for each day of the experiment

Subjective Ratings. To measure drivers' subjective sleepiness during driving, we collected drivers' oral responses to the Karolinska Sleepiness Scale (KSS) [13]. To avoid



Fig. 1. The experiment vehicle and the eye tracker settings

the influence of too frequent subjective responses on fatigue development, the KSS rating was conducted every hour. In total, 10 KSS scores were obtained during the 10 h of driving for each day of the experiment

2.2 Participants and Experiment Procedures

To ensure safety during the use of ADAS on real roads, after screening, three professional long-haul truck drivers (all male, average age 43 years old) who have used ADAS for more than two years participated in this experiment. Each driver conducted 2 rounds of driving experiments over 4 days. In each round of experiments, a driver drove manually one day and with the help of ADAS the other day, on a round trip between Suzhou and Xiaogan. The order of the two drives (i.e., with and without ADAS and destinations of drives) was balanced. To avoid possible fatigue accumulation, the gap between two rounds of experiments for each driver was at least 5 days.

Before the experiment, the experimenter introduced the experimental procedures, answered relevant questions, and obtained written informed consent. Before the experiment each day, we calibrated the eye-tracking system. Then the driver manually drove the vehicle to the highway toll station and then they drove with or without ADAS, depending on the experimental condition. The moment that the vehicle passed the toll station was set as time 0 when the collection of eye movement data, DRT data, and subjective rating data began.

2.3 Data Processing and Statistical Analysis

Data Processing. For the eye tracking data, four metrics, i.e., blink frequency, blink duration, pupil diameter, and percentage of eye closure (PERCLOS), were extracted following previous driving fatigue-related studies (e.g., [9–12]). For each metric, we calculated the mean value in each hour during the experiment. The raw response time in DRT and KSS scores were used. For eye-tracking metrics, to account for individual differences, we assumed that truck drivers were less fatigued in the first hour of the experiment, so the eye-tracking metrics in the first hour of the experiment were used as the baseline to normalize the changes of corresponding metrics during the rest of the day.

Dependent Variable and Independent Variables. Four eye-tracking metrics, DRT response time, and KSS scores were selected as dependent variables. The independent variables included driving time (i.e., continuous), driving mode (i.e., ADAS versus manual), and the two-way interaction effect between driving time and driving mode.

Statistical Models. Six linear mixed-effects models were fitted using SAS OnDemand for Academics, the models accounted for repeated measures from drivers using the generalized estimating equations method. The significance level was set as 0.05, and only significant results were reported in the following section

3 Results

3.1 Eye Tracking-Based Metrics

The driving mode significantly influenced drivers' blink duration ($F(1, 103) = 54.87$, $p < .0001$), blink frequency ($F(1, 103) = 33.88$, $p < .0001$), pupil diameter ($F(1, 103) = 13.72$, $p = .0003$), and PERCLOS ($F(1, 103) = 56.69$, $p < .0001$). More specifically, compared to driving manually, drivers had higher pupil diameter ($\Delta = 17.61\%$, 95% confidence intervals (95%CI): $[8.18\%, 27.04\%]$, $t(103) = 3.70$, $p = .0003$), lower blink frequency ($\Delta = -24.07\%$, 95%CI: $[-32.27\%, -15.87\%]$, $t(103) = -5.82$, $p < .0001$), lower blink duration ($\Delta = -9.28\%$, 95%CI: $[-11.77\%, -6.81\%]$, $t(103) = -7.41$, $p < .0001$), and lower PERCLOS when driving with ADAS ($\Delta = -130.76\%$, 95%CI: $[-165.20\%, -96.32\%]$, $t(103) = -7.53$, $p < .0001$). Figure 2 presents the post-hoc comparisons of eye-tracking metrics between different driving modes.

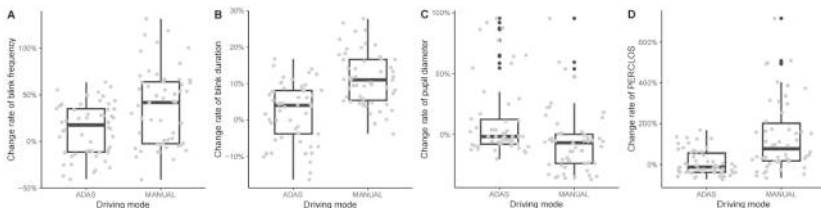


Fig. 2. Post-hoc comparisons of eye-tracking metrics between two driving modes

3.2 Response Time and KSS Score

A significant interaction between driving mode and driving time was found for DRT response time ($F(1, 54) = 4.38$, $p = .04$, Fig. 3). In manual driving, for every 2-hour increase in the driving time, drivers' response time increased by 0.032 s (95%CI: $[0.02, 0.05]$, $t(54) = 4.30$, $p < .0001$). With the increase in driving time, the marginal effect of driving mode on drivers' response time increased, which means that the benefit of ADAS would be more obvious with the increase in driving time.

Driving mode ($F(1, 115) = 66.23$, $p < .0001$) and driving time ($F(1, 115) = 167.20$, $p < .0001$) significantly influenced drivers' subjective KSS scores, and no significant

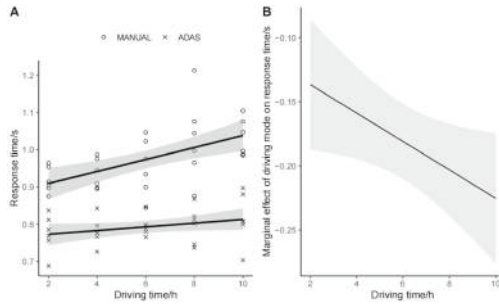


Fig. 3. Post-hoc comparisons of response time: A—The interaction effect between driving time and driving mode; B—The marginal effect of driving mode on response time.

interaction effect was observed. Figure 4 presents the comparison of KSS scores between different driving modes and different driving times. In manual driving conditions, drivers' KSS scores were higher than that when driving with ADAS ($\Delta = 1.05$, 95%CI: [0.79, 1.31], $t(103) = 8.14$, $p < .0001$). For each 1-hour increase in driving time, the driver's subjective KSS score increased by 0.29.

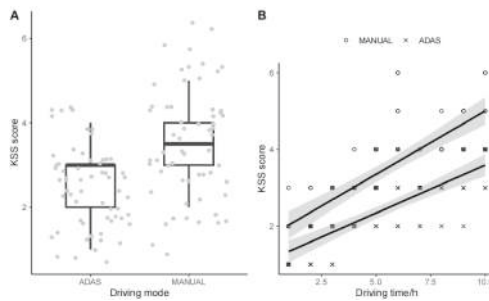


Fig. 4. Post-hoc comparisons of KSS score: A—Comparison of KSS scores between different driving modes; B—The relationship between KSS score and driving time

4 Discussion

Based on over 120 h of data from a real-world driving experiment on open roads, for the first time, this study explores the impact of ADAS on long-haul truck drivers' fatigue development. Several fatigue metrics indicated that the average fatigue level of the drivers when driving with ADAS is lower than that of manual driving. In addition, drivers' subjective KSS scores when driving with ADAS were also significantly lower than those in manual driving. These results show that ADAS can reduce the fatigue level of truck drivers during long-haul driving. This finding is different from the conclusions obtained among passenger car drivers in previous studies. We propose two possible reasons: (1) as professional drivers, heavy truck drivers are more able to appropriately adjust their

attention allocation in driving tasks through self-discipline and work responsibility [14], thereby avoiding passive fatigue caused by ADAS; (2) Most of the existing research on passenger car drivers were based on simulators, and the participants lacked the stimulation from the real environment and experienced a risk that was lower than real-world driving [15], which made the simulation drive boring and fatigue-inducing.

Finally, it should be noted that, although we have tried to collect more data through repeated measures, the findings of our study should be carefully interpreted due to the limited participants. Future research can validate our findings with more participants.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (Grant No. 52202425) and the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083).

References

1. Wörle, J., Metz, B., Othersen, I., Baumann, M.: Sleep in highly automated driving: Takeover performance after waking up. *Accid. Anal. Prev.* **144**, 105617 (2020)
2. Murata, A.: Proposal of a method to predict subjective rating on drowsiness using physiological and behavioral measures. *IIE Trans. Occup. Ergon. Hum. Factors.* **4**, 128–140 (2016)
3. Traffic Management Bureau: Ministry of Public Security. China road traffic accidents annual statistical report, China (2023)
4. Di Milia, L.: Shift work, sleepiness and long distance driving. *Transp. Res. Part F Traffic Psychol. Behav.* **9**, 278–285 (2006)
5. Romo, A., Hernandez, S., Cheu, R.L.: Identifying precrash factors for cars and trucks on interstate highways: mixed logit model approach. *J. Transp. Eng.* **140** (2014)
6. Perttula, P., Ojala, T., Kuosma, E.: Factors in the fatigue of heavy vehicle drivers. *Psychol. Rep.* **108**, 507–514 (2011)
7. Friswell, R., Williamson, A.: Comparison of the fatigue experiences of short haul light and long distance heavy vehicle drivers. *Saf. Sci.* **57**, 203–213 (2013)
8. Meng, F., et al.: Driving fatigue in professional drivers: A survey of truck and taxi drivers. *Traffic Inj. Prev.* **16**, 474–483 (2015)
9. Jarosch, O., Bellem, H., Bengler, K.: Effects of Task-Induced Fatigue in Prolonged Conditional Automated Driving. *Hum. Factors J. Hum. Factors Ergon. Soc.* **61**, 1186–1199 (2019)
10. Körber, M., Cingel, A., Zimmermann, M., Bengler, K.: Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manuf.* **3**, 2403–2409 (2015)
11. Vogelpohl, T., Kühn, M., Hummel, T., Vollrath, M.: Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accid. Anal. Prev.* **126**, 70–84 (2019)
12. Vogelpohl, T., Kühn, M., Hummel, T., Gehlert, T., Vollrath, M.: Transitioning to manual driving requires additional time after automation deactivation. *Transp. Res. Part F Traffic Psychol. Behav.* **55**, 464–482 (2018)
13. Kaida, K., et al.: Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clin. Neurophysiol.* **117**, 1574–1581 (2006)

14. Chen, T., Sze, N.N., Newnam, S., Bai, L.: Effectiveness of the compensatory strategy adopted by older drivers: Difference between professional and non-professional drivers. *Transp. Res. Part F Traffic Psychol. Behav.* **77**, 168–180 (2021)
15. Mayhew, D.R., Simpson, H.M., Wood, K.M., Lonero, L., Clinton, K.M., Johnson, A.G.: On-road and simulated driving: Concurrent and discriminant validation. *J. Saf. Res.* **42**, 267–275 (2011)



Methods for Evaluating the Comfort of Vehicle Seating Systems - A Literature Review

Weisheng Jiang^{1,2}, Zhongqi Liu^{1,2}, and Qianxiang Zhou^{1,2}(✉)

¹ School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China

zqxg@buaa.edu.cn

² Beijing Advanced Innovation Centre for Biomedical Engineering, Beihang University, Beijing 102402, China

Abstract. This review assesses methods for evaluating vehicle seating comfort, which is crucial due to prolonged fixed postures by occupants. These postures, aggravated by vehicle and environmental vibrations, cause significant psychological and physiological discomfort. Such conditions diminish work efficiency, reduce working hours, and adversely affect occupants' health, potentially leading to various occupational diseases. The study defines comfort, explores influencing factors via the Human-Machine-Environment Systems (HMES) framework, and examines current evaluation techniques and technological approaches based on different types of comfort. It suggests potential enhancements to seating comfort, including ergonomic modifications and material innovations. Furthermore, the research incorporates biomechanical analyses of human lower limbs under sedentary conditions, identifying critical gaps in existing studies on vehicle seating comfort. This examination aids in pinpointing deficiencies and outlining future research directions. The paper concludes by advocating for in-depth studies that merge ergonomic design with biomechanical insights, aiming to refine vehicle seating systems. These enhancements strive to meet human physiological needs better, thus improving overall comfort and reducing health hazards.

Keywords: Evaluation methods · Seating comfort · Static comfort · Dynamic comfort · Human-Machine-Environment systems

1 Introduction

In prolonged driving scenarios, occupants are required to maintain fixed postures amidst mechanical vibrations, which can lead to muscle contraction and decreased blood flow in the lower extremities. Such static postures decrease comfort and may induce physical and psychological injuries. The body weight supported by the seat generates a pressure gradient from the ischial tuberosities outward, exerting undue pressure on the blood vessels and nerves in the muscle groups of the gluteal and posterior thigh regions, which are particularly sensitive to compression. This seating configuration can distort blood vessels and alter hemodynamics, impacting circulation and tissue perfusion, potentially

causing numbness, swelling, and conditions such as low back pain, varicose veins, or peripheral arterial disease (PAD). Despite the emphasis on safety in current vehicle seat designs, comfort often remains secondary. This oversight underscores the necessity of integrating human factors engineering to enhance seating comfort, presenting a vital area for future research.

2 Definition and Influencing Factors

Comfort in vehicle seating is multifaceted and often ambiguous, with evaluations frequently subject to significant subjective bias. Current research into comfort evaluation systems is limited. Hertzberg et al. define comfort as the absence of discomfort, emphasizing a dichotomous relationship [1]. Lueder et al. expand on this by suggesting that comfort encompasses physical and psychological aspects modeled through an interplay of forces and constraints [2]. Within the Human-Machine-Environment Systems (HMES) framework, the interaction between the vehicle's design and the environmental conditions plays a crucial role in influencing occupant comfort [3]. This interaction is illustrated in Fig. 1, which depicts how ergonomic principles are applied to optimize comfort in vehicle seating.

3 Evaluation Method for Vehicle Seating Comfort

The assessment of vehicle seating comfort involves various methodologies broadly classified into subjective, objective, and combined approaches. Subjective methods, deeply rooted in personal perceptions, often employ scales that vary significantly between individuals due to personal sensitivities and psychological factors. This inherent subjectivity can lead to consistently standardized results across different populations or conditions [4].

Objective methods provide a more standardized and replicable approach by quantifying physical data. These methods include using advanced technologies such as multi-signal fusion, which integrates data from multiple sensors to give a detailed picture of the seating environment. Additionally, behavioral data analysis offers insights into how occupants interact with their seating over time, revealing long-term trends and comfort issues. Among the most frequently used objective evaluation techniques are seat pressure measurement, which directly records the pressure distribution that a seat exerts on an occupant's body; bioelectrical information analysis, which measures physiological responses [5]; driving posture and motion analysis, which evaluates the ergonomic suitability of seat design in maintaining optimal body posture [6]; and finite element modeling (FEM), which simulates physical interactions between the human body and the seat to predict potential discomfort and stress points [7].

These methods are precious in different states of vehicle use—stationary, operational, and dynamic. Dynamic conditions, characterized by the vehicle's vibration environment, are influenced by the amplitude and frequency of the vehicle's movements. In high-vibration scenarios, dynamic comfort evaluation becomes critical, as excessive vibration can lead to significant discomfort and even health issues over time [8].

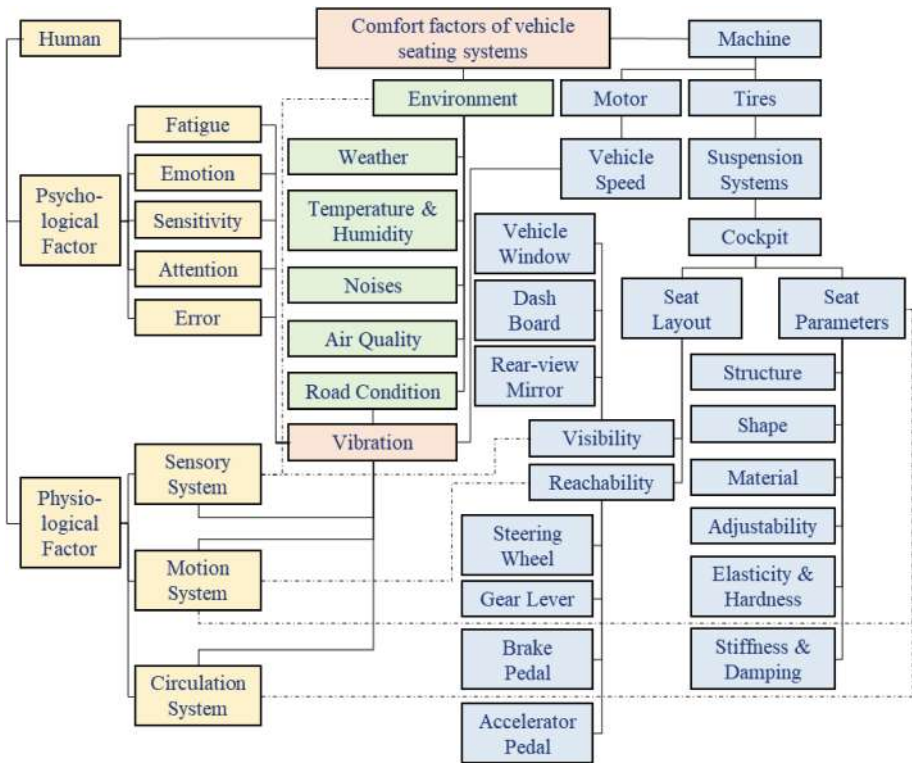


Fig. 1. Factors influencing the comfort of vehicle seating systems

Static comfort assessment focuses on a non-moving vehicle's subjective and objective comfort experience. This involves assessing the initial comfort and how well the seat supports prolonged sitting without causing discomfort. For optimal static comfort, the seat must provide adjustable features to fit different body types and adequate support that promotes healthy posture and evenly distributes body pressure to minimize stress points.

Techniques for evaluating static comfort include direct methods like in-vivo seat pressure measurement, which provides immediate feedback on the pressure distribution characteristics of a seat. Indirect methods involve sophisticated equipment such as optical instruments to monitor changes in the partial pressure of oxygen and other blood flow parameters in the lower limbs, which are critical in assessing the circulatory impacts of seating [9–11]. Ultrasound equipment is also employed to measure hemodynamic parameters in vivo, providing data that can be used to simulate the arterial vasculature of the lower limbs in various seating postures. These simulations help researchers understand the potential risks for conditions like peripheral arterial disease (PAD), which can be exacerbated by poor seating design [12].

Operational comfort assessments extend these static evaluations by incorporating real-time analyses under operational conditions, using biomechanical finite element

models of the musculoskeletal system. These models help understand the mechanical forces at play during vehicle operation. They are complemented by surface electromyographic (sEMG) data, which provides insights into muscle activity and fatigue as direct indicators of comfort or discomfort [7].

Dynamic comfort, also known as vibration comfort, evaluates the impact of environmental vibrations on the human body. Adjustments to vehicle tires and suspension systems are standard measures to mitigate these effects. However, such adjustments must carefully balance improvements in comfort with potential impacts on vehicle stability and safety. In the realm of human-machine-environment systems (HMES), vibrations are categorized into local (affecting specific body parts) and whole-body (affecting the entire body) vibrations, with the latter often having a more profound impact due to its extensive influence on occupant comfort.

The global discourse on dynamic comfort evaluation features a variety of methodologies proposed by researchers. For instance, the Ride Comfort Law (RCL) introduced by Janeway calculates comfort based on tolerance levels to vibration. At the same time, Pradko's Absorbed Power (AP) method evaluates how energy is absorbed and transferred through the body, offering a dynamic perspective on comfort. Additionally, the NASA-developed Standard Uncomfortableness Index Rule and the ISO standards provide widely respected and followed frameworks in both academic and industrial circles [13–15].

In sum, evaluating human comfort in vehicle seating is an intricate process involving a blend of subjective assessments and objective measurements, guided by established scientific methods and enhanced by the latest technological advancements. This thorough approach ensures that every aspect of seat design is optimized to enhance comfort, mitigate discomfort, and prevent potential health issues, thereby aligning with the highest standards of ergonomic research and practice [16–19].

4 Program to Improve Comfort

In the human-machine-environment systems (HMES), discomfort primarily arises from prolonged, fixed postures. To alleviate this, the human body passively shifts to redistribute pressure at the seat interface, reducing discomfort. Dynamic seating, which allows for periodic and gradual adjustments in body posture, has proven effective in enhancing comfort [20, 21]. The most widely adopted solution for increasing comfort involves incorporating a support system within the backrest that conforms to the natural curves of the spine and body [22, 23]. Enhancing shock absorption can also be achieved by altering the seat's padding material [24]. Additionally, many drivers use thin, soft cushions for extra comfort. Moreover, gas-mediated adjustable seat cushions have been developed to periodically vary the pressure distribution between the human body and the seat, simultaneously improving elasticity, reducing firmness, and enhancing thermal comfort for greater occupant comfort.

5 Conclusion

Based on existing literature, this paper summarizes the factors influencing vehicle seat design from a human-machine-environment perspective. It also analyzes the shortcomings of current research based on the comfort evaluation system of driving posture, exemplifying the analytical methods and technical means currently applied for physiological signal measurement. Evaluation lags behind design and production because the current comfort evaluation system can only partially dominate but needs to collect subjective and objective data feedback from evaluators in experiments to obtain evaluation conclusions. With the development of automotive technology and the change of driving modes, the factors within the HMES grow increasingly complex, placing higher demands on evaluating driving comfort. In the future, multi-physiological signal fusion technology should be applied to analyze human mechanical parameters and physiological signals multi-dimensionally and objectively, aiming to establish a model that quantifies subjective feelings and forms a comprehensive comfort evaluation system.

References

1. Hertzberg, H.T.E.: The human buttocks in sitting: pressures, patterns, and palliatives. *SAE Trans.*, 39–47 (1972)
2. Lueder, R.K.: Seating comfort: a review of the construct in the office environment. *Hum. Factors*. **25**(6), 701–711 (1983)
3. Sammonds, G.M., Fray, M., Mansfield, N.J.: Effect of long-term driving on driver discomfort and its relationship with seat fidgets and movements (SFMs). *Appl. Ergon.* **58**, 119–127 (2017)
4. Andersson, B.J., Ortengren, R., Nachemson, A., Elfström, G.: Lumbar disc pressure and myoelectric back muscle activity during sitting. IV. Studies on a car driver's seat. *J. Rehab. Med.* **6**(3), 128–133 (1974)
5. Park, S.J., Kim, C.B., Kim, C.J., Lee, J.W.: Comfortable driving postures for Koreans. *Int. J. Ind. Ergon.* **26**(4), 489–497 (2000)
6. Li, W., Zhang, M., Lv, G., Han, Q., Gao, Y., Wang, Y., Li, Z.: Biomechanical response of the musculoskeletal system to whole body vibration using a seated driver model. *Int. J. Ind. Ergon.* **45**, 91–97 (2015)
7. Ebe, K., Griffin, M.J.: Qualitative models of seat discomfort including static and dynamic factors. *Ergonomics*. **43**(6), 771–790 (2000)
8. De Looze, M.P., Kuijt-Evers, L.F., Van Dieën, J.A.A.P.: Sitting comfort and discomfort and the relationships with objective measures. *Ergonomics*. **46**(10), 985–997 (2003)
9. Verver, M.M., Van Hoof, J., Oomens, C.W.J., Wismans, J.S.H.M., Baaijens, F.P.T.: A finite element model of the human buttocks for prediction of seat pressure distributions. *Comput. Methods Biomech. Biomed. Engin.* **7**(4), 193–203 (2004)
10. Van Geffen, P., Reenalda, J., Veltink, P.H., Koopman, B.F.: The effects of a dynamic tubular support on ischial buttock load and pattern of blood supply. *IEEE Trans. Neural Syst. Rehabil. Eng.* **18**(1), 29–37 (2010)
11. Hou, J., Li, X., Li, Z., Yin, L., Chen, X., Liang, F.: An in vivo data-based computational study on sitting-induced hemodynamic changes in the external iliac artery. *J. Biomech. Eng.* **144**(2) (2022)
12. Xu, G.X., Gu, S.Q., Meng, D.H., Wang, H.X., Li, J.A.: Effect of muscle bio-resonance on walking efficiency: A primary study. *Chinese J. Rehab. Med.* **23**(12), 1092–1094 (2008)

13. Griffin, M.J., Whitham, E.M.: Duration of whole-body vibration exposure its effect on comfort. *J. Sound Vib.* **48**(3), 333–339 (1976)
14. Wang, P., Luo, P., Zhang, C., Liu, C., Yang, X.T., He, C.Q.: Research progress on the effect of whole-body vibration training on cardiopulmonary function. *Chinese J. Rehab. Med.* **32**(1), 117–120 (2017)
15. Mechanical vibration and shock-evaluation of human exposure to whole-body vibration-part 1. ISO2631-1 (1997)
16. Wu, X., Rakheja, S., Boileau, P.: Distribution of human-seat interface pressure on a soft automotive seat under vertical vibration. *Int. J. Ind. Ergon.* **24**(5), 545–557 (1999)
17. Le, P., Rose, J., Knapik, G., Marras, W.S.: Objective classification of vehicle seat discomfort. *Ergonomics.* **57**(4), 536–544 (2014)
18. Nahvi, H., Fouladi, M.H., Nor, M.M.: Evaluation of whole-body vibration and ride comfort in a passenger car. *Int. J. Acous. Vib.* **14**(3), 143–149 (2009)
19. Kolich, M., White, P.L.: Reliability and validity of a long term survey for automobile seat comfort. *Int. J. Veh. Des.* **34**(2), 158–167 (2004)
20. Van Geffen, P., Reenalda, J., Veltink, P.H., Koopman, B.F.: Decoupled pelvis adjustment to induce lumbar motion: A technique that controls low back load in sitting. *Int. J. Ind. Ergon.* **40**(1), 47–54 (2010)
21. Van Veen, S., Vink, P.: Posture variation in a car within the restrictions of the driving task. *Work.* **54**(4), 887–894 (2016)
22. Carcone, S.M., Keir, P.J.: Effects of backrest design on biomechanics and comfort during seated work. *Appl. Ergon.* **38**(6), 755–764 (2007)
23. Donnelly, C.J., Callaghan, J.P., Durkin, J.L.: The effect of an active lumbar system on the seating comfort of officers in police fleet vehicles. *Int. J. Occup. Saf. Ergon.* **15**(3), 295–307 (2009)
24. Kolich, M., Essenmacher, S.D., McEvoy, J.T.: Automotive seating: the effect of foam physical properties on occupied vertical vibration transmissibility. *J. Sound Vib.* **281**(1–2), 409–416 (2005)



Human Factors and Ergonomic Issues in Autonomous Shipping

Hongtae Kim^(✉)

Korea Research Institute of Ships and Ocean Engineering, Daejeon 34103, Republic of Korea
hongtae.kim@kriso.re.kr

Abstract. Autonomous shipping offers benefits such as preventing marine accidents, improving fuel efficiency, and reducing costs. However, it also brings e job loss, task changes, and security issues. Remote operators' situational awareness is a new human factor and ergonomic issue that arises as shore-based control precedes the full operation of autonomous ships. This paper examines the major human factors and ergonomic issues to be considered in autonomous shipping by surveying the current state of the technology and classifying it by field, with a special focus on human factors and ergonomic aspects.

Keywords: Autonomous shipping · Maritime autonomous surface ship · Remote operator · Crew training · Human error

1 Introduction

Recently, in line with the era of the 4th Industrial Revolution, the shipping industry has actively engaged in discussions on autonomous shipping. Autonomous shipping refers to using unmanned vessels to transport goods across bodies of water without direct human oversight on board.

At the 98th session of the International Maritime Organization (IMO) Maritime Safety Committee (MSC) in June 2017, a Maritime Autonomous Surface Ship (MASS) was defined as a ship capable of operating independently without human intervention at various levels of automation. By 2021, IMO conducted a Regulatory Scoping Exercise (RSE) to evaluate how existing regulations apply to different levels of MASS automation, finalizing this in May 2021 for MSC.

The IMO compared training and certification requirements for MASS operators with the International Convention on Standards of Training, Certification, and Watch-keeping for Seafarers (STCW), proposing additional competencies [1]. It was decided to develop Knowledge, Understanding, and Proficiency (KUP) guidelines as part of sub-committee on Human Element, Training, and Watchkeeping (HTW) to establish minimum competency standards for seafarers [2].

The era of a fully upgraded MASS may render a seafarer's role futile given the operator's offshore management of the ship, but various human element issues are likely to remain. This paper introduces an overview of a MASS and analyses the main human element issues that need consideration in operating a MASS.

2 Human Elements in MASS and Autonomous Shipping

A MASS is critical to the maritime industry in the Fourth Industrial Revolution. It is a game-changer in shipping, port logistics, shipbuilding, and equipment industries, accelerating digitization, platformization, and intellectualization in related sectors. While autonomous shipping has positive aspects such as preventing marine accidents, improving fuel efficiency, and reducing costs, it also has negative aspects such as job reduction, task changes, and security problems [3–5].

The introduction of MASS through various projects globally aims to optimize operations, enhance resource management, and improve marine logistics, with the EU's MUNIN, Finland's AAWA, Japan's NYK, China's UCSDA, and Korea's KASS projects each developing advanced autonomous technologies and aiming for significant milestones in unmanned ship development by 2035 [6–11].

2.1 Research Trends

Many human factors studies indicate the crucial need for humans to monitor automated operations, even with highly automated systems. However, there is limited research on human factors issues in autonomous shipping.

New job opportunities for highly qualified seafarers, such as controlling autonomous ships remotely, are emerging as automation in maritime operations increases, particularly for high-level officers who are less likely to be automated [12–17]. Improved seafarer education and training are necessary to adapt to new technologies and decision support systems [18–21], and the introduction of novel concepts in Maritime 5.0 will further address the evolving needs in autonomous shipping [22].

Research shows that human errors by Remote Operators (ROs) still occur in fully unmanned autonomous ships, with new error types emerging from interactions between manned and unmanned vessels, highlighting the need for Human Factors Engineering (HFE)-oriented simulator-based studies to address these challenges and solutions [23–26]. Information overload, situation awareness, skill degradation, and boredom uniquely impact ROs' performance [27]. Research on ROs' Health, Safety, and Environmental (HSE) aspects is limited, and future studies should address issues such as mental stress, fatigue, computer-related injuries, and musculoskeletal diseases from sedentary work.

The MUNIN project identified situational awareness and decision-making issues for Remote Operation Centers (ROC) operation teams, leading to improved ROC design [28, 29]. Also, research on autonomous ship systems has emphasized the need for interdisciplinary efforts in human-AI interaction design [30]. Research on autonomous vehicles has progressed more rapidly than on autonomous ships, highlighting the need to consider the driver's experience as automation increases [31].

Gap analysis and SWOT-AHP analysis were used to identify factors affecting the safe operation of Vessel Traffic Services (VTS) with autonomous ships [32]. Enhancing VTS operators' (VTSOs) skills, including situational awareness, understanding COLREGs, and communication with ROs, was recommended [33]. A real-time collision risk assessment support system was proposed to improve situational awareness for VTSOs and MASS ROs in near-collision situations [34].

2.2 Major Issues Related to the Human Elements Required for Autonomous Shipping

Although research on human factors and ergonomics in autonomous shipping has been less active than technical aspects, their importance is increasing. Since the 2020s, several review papers have identified 13 human factor issues affecting interactions in remote ship operations [35–38].

We reviewed 38 articles and reports selected as highly relevant for this study and derived five research fields through thematic analysis and human element discussions in IMO. The distribution of articles is presented in Fig. 1, and the thematic analysis results are shown in Fig. 2.

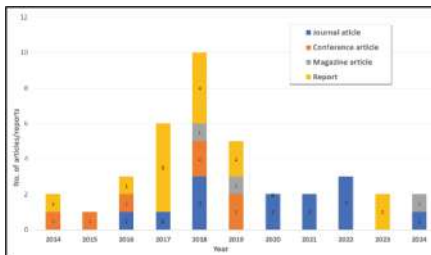


Fig. 1. Distribution of selected articles and reports.

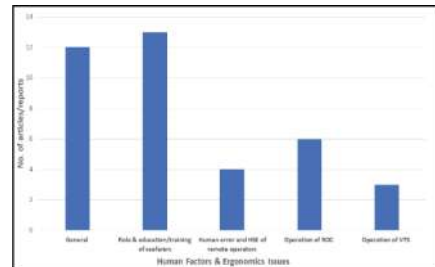


Fig. 2. Thematic analysis of the articles and reports.

The largest articles and reports were distributed in the ‘Roles and education/training of seafarers’ field, focusing on the benefits and expected effects of autonomous shipping and MASS and issue analysis. Figure 3 shows the results of the four categories of human elements in autonomous shipping, excluding the ‘General’ field.

Role and Education/Training of Seafarers. As autonomous ships are developed in stages, seafarers’ roles will diminish, and they will ultimately manage and operate ships from shore. The operational and technical conditions of ships will change, affecting crew manning standards. Experts with maritime knowledge will be needed in various professional fields related to autonomous ships, such as law, IT, and public administration. Remote operators and remote masters in ROC will require competencies in technical knowledge, safety management, navigation and situational awareness, communication and teamwork, and legal and regulatory knowledge.

Human Error and HSE of Remote Operators. Remote control and situational awareness are crucial for autonomous ships. ROs face challenges such as information overload and visual information recognition, requiring user-centered design and high-quality information. Human roles remain important even with fully unmanned ships, and minimizing errors involves using Human Machine Interface (HMI) technologies. Automation reduces mundane tasks but increases cognitive demands and health and safety risks for ROs due to time pressure and communication needs.

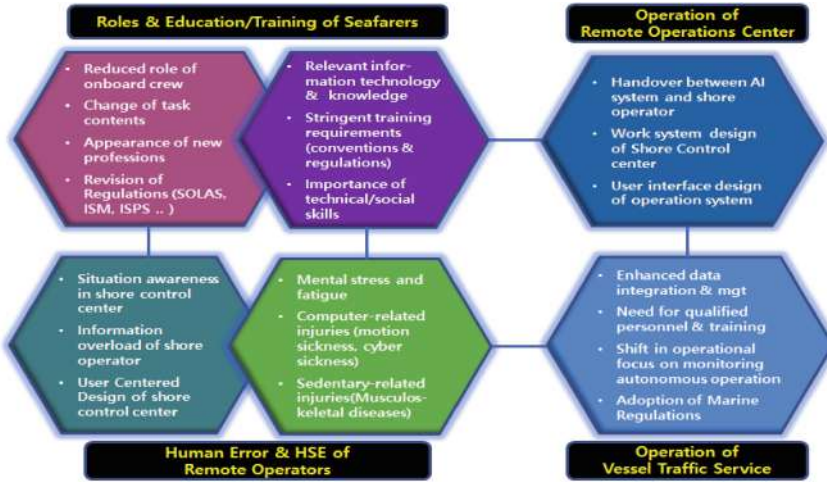


Fig. 3. Four categories of human elements in autonomous shipping.

Operation of ROC. Three considerations for efficient and safe ROC operation are the design and testing of work systems, technologies, and procedures to support humans, ensuring system reliability and safety; strategies for foreseeable situations, principles-based control measures, and considerations of system resilience, reliability, and sustainability; and requirements for each voyage stage, cyber-social-technical systems, technical-cultural aspects, human well-being, and incident reporting.

Operation of VTS. The introduction of autonomous ships will change the role of VTS, requiring new operating procedures, technologies, and secure communication between ships, ROs, and VTS controllers. Additionally, training on new VTS systems and emergency control transfer regulations is necessary to ensure the safe operation of MASS through internal and external cooperation.

3 Conclusions and Future Remarks

While autonomous systems aim to eliminate human error in accidents, human decision-making risks persist, shifting errors from ship operators to ROs. Although marine accidents may decrease in an autonomous environment, secondary accidents could occur without recovery and emergency measures due to the absence of seafarers on board.

Efforts should continue to focus on the human element, including changes in seafarers' roles, education and training, software development, and HMI technologies for ROCs. It is essential to systematize seafarers' knowledge for ROs, ensure a smooth interface between autonomous ships and ROCs, and identify and evaluate human factors affecting work performance.

Realizing autonomous ships requires intellectualization and virtualization, focusing on technologies like autonomous navigation, AI-based situation awareness, navigation control integration, and collaborative group navigation. Success depends on consensus

on cybersecurity, big data, workforce training, job responsibility for accidents, and AI ethics. Revising the curriculum for autonomous ship operations and addressing social and cultural aspects, such as responsibility and compensation in marine accidents and AI malfunctions, is also essential.

Acknowledgments. This research was supported by the Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (20200615).

References

1. IMO: Bridging Competency Gaps for MASS Operators in Alignment with the STCW Framework. MSC/ISWG/MASS 2/3/1 (2023)
2. IMO: Draft Report of the MSC on its 107th Session. MSC 107/WP.1/Rev.1 (2023)
3. Rolls Royce: The Ear of Ship Intelligence. <http://www.rolls-royce.com> (2016)
4. Bureau Veritas: Guidelines for Autonomous Shipping, BV Guidance Note. Bureau Veritas Marine & Offshore, Paris, France (2017)
5. DNV-GL: Autonomous and Remotely Operated Ships. DNV-GL Class Guideline. DNV, Høvik, Norway (2018)
6. Laurinnen, M.: Advanced Autonomous Waterborne Applications Initiative. AAWA Seminar, Helsinki, Finland (2016)
7. MUNIN: Research in maritime autonomous systems project Results and technology potentials. 7th Framework Programme Grant Agreement No 314286 (2017)
8. JSMEA: Activities of Smart Ship Application Platform 2 Project (SSAP2). Japan Ship Machinery and Equipment Association (2017)
9. Seo, Y.S.: A technical prospects and trends in MASS. KEIT PD Issue Rep. **18**(4), 57–75 (2018)
10. Park, H.R., Park, H.S., Kim, B.R.: A Study on the Policy Directions related to the Introduction of Maritime Autonomous Surface Ship (MASS). KMI Rep. (2018)
11. Korea Autonomous Surface Ship (KASS) Project Homepage. <https://kassproject.org/en> (2024). Accessed 21 May 2024
12. Burmeister, H.C., Bruhn, W.C., Rødseth, Ø.J., Porathe, T.: Can Unmanned ships improve navigational safety?. In: Transport Research Arena, Paris (2014)
13. Lloyd's Register: Cyber Enabled Ships ShipRight Procedure. A Lloyd's Register guidance document (2017)
14. Kim, K.S.: Impacts of maritime autonomous surface ship on the seafarer and seafarer's position. In: Proceedings of Policy Seminar on Maritime Autonomous Surface Ship. Korea Maritime Institute, Busan, South Korea (2018)
15. WMU: Transport 2040:Automation, Technology, Employment - The Future of Work. World Maritime University, Malmö (2019)
16. Choi, J.H., Lee, S.I.: Interpretive Consideration relating to the Ship's Definition of the Unmanned Ship. Ocean Policy Res. **33**(1), 171–191 (2018)
17. Chang, H. J.: Promising jobs in the era of the 4th Industrial Revolution. MK News (2019)
18. Shanghai Maritime University: Research on the Impacts of Marine Autonomous Surface Ship on the Seafarer's Career and MET. Technical Report (2018)
19. Sencila, V., Kalvaitiene, G.: Autonomous Shipping and New Challenge for Maritime Education and Training. Lithuanian Maritime Academy, Klaipeda (2019)
20. Hone, R.N.: Training standards and accreditation. In: Proceedings of 4th Maritime Autonomous Systems Regulatory Conference. UK Chamber of Shipping, UK (2019)

21. Lloyd, J.: Marine autonomous ship regulation next step – people. In: Proceedings of 4th Maritime Autonomous Systems Regulatory Conference. UK Chamber of Shipping (2019)
22. Shahbakhsh, M., Emad, G.R., Cahoon, S.: Industrial revolutions and transition of the maritime industry: The case of Seafarer's role in autonomous shipping. *Asian J. Shipp. Logist.* **38**(1), 10–18 (2022)
23. Porathe, T., Prison, J., Man, Y.: Situation awareness in remote control centres for unmanned ships. In: Proceedings of Human Factors in Ship Design & Operation. The Royal Institution of Naval Architects, UK (2014)
24. Ahvenjärvi, S.: The human element and autonomous ships. *Int. J. Mar. Navig. Saf. Sea Transp.* **10**(3), 517–521 (2016)
25. Zhang, D., Han, Z., Zhang, K., Zhang, J., Zhang, M., Zhang, F.: Use of hybrid causal logic method for preliminary hazard analysis of maritime autonomous surface ships. *J Mar Sci Eng.* **10**(6), 725 (2022)
26. Karvonen, H., Martio, J.: Human factors issues in maritime autonomous surface ship systems development. In: Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships (ICMASS). IOP Conference Series: Materials Science and Engineering, Busan, Korea (2018)
27. Ramos, M.A., Utne, I.B., Mosleh, A.: On factors affecting autonomous ships operators performance in a shore control center. In: Proceedings of 14th Probabilistic Safety Assessment and Management. International Association for Probabilistic Safety Assessment and Management (IAPSAM), USA (2018)
28. Man, Y., et al.: From desk to field - human factor issues in remote monitoring and controlling of autonomous unmanned vessels. In: 6th International Conference on Applied Human Factors and Ergonomics, Las Vegas, USA (2015)
29. Man, Y., et al.: Human factor issues during remote ship monitoring tasks: An ecological lesson for system design in a distributed context. *Int. J. Ind. Ergon.* **68**, 231–244 (2018)
30. Veitch, E., Alsos, O.T.: A systematic review of human-AI interaction in autonomous ship systems. *Saf. Sci.* **152** (2022)
31. Park, H.: A review on autonomous vehicles focused on human factors. *J. Ergon. Soc. Korea.* **36**(5), 421–435 (2017)
32. Baldauf, M., Mejia Jr., M., Chong, J., Amphanthong-paphakul, T.: Conventional vessels and maritime autonomous surface ships – studying, operational challenges of future traffic scenarios. In: 20th International Navigation Simulator Lecturers Conference, Auckland (NZ) (2018)
33. Chong, J. C.: Impact of maritime autonomous surface ships (MASS) on VTS operations. World Maritime University Dissertations (2018)
34. Yoo, Y., Lee, J.K.: Collision risk assessment support system for MASS RO and VTSS support in multi-ship environment of vessel traffic service area. *J Mar Sci Eng.* **9**, 1143 (2021)
35. Fan, C., Wrobel, K., Montewka, J., Gil, M., Wan, C., Zhang, D.: A framework to identify factors influencing navigational risk for maritime autonomous surface ships. *Ocean Eng.* **202**, 1–15 (2020)
36. Kari, R., Steinert, M.: Human factor issues in remote ship operations: Lesson learned by studying different domains. *J Mar Sci Eng.* **9**(4), 385 (2021)
37. Kim, H., Yang, Y.-H., Joung, T.-W., Lee, S.: Human factor and ergonomic issues in maritime autonomous surface ships. In: Proceedings of ERGOSHIP 2021 Conference, vol. 2021, Busan, Korea
38. Li, X., Yuen, K.F.: A human-centred review on maritime autonomous surfaces ships: Impacts, responses, and future directions. *Transp. Rev.* **1–20** (2024)



Creating and Assessing Automotive Telltales for Preventing Drunk Driving

Daehui Kim¹ and Jaehyun Park²(✉)

¹ Department of Industrial and Management Engineering, Incheon National University, Incheon, South Korea

² Department of Industrial Engineering, Konkuk University, Seoul, South Korea
jaehpark@konkuk.ac.kr

Abstract. The issue of driving under the influence of alcohol has become a pressing concern, posing a significant threat to road safety and traffic regulations. In recent years, several technological solutions have been developed to prevent drunk driving. Such systems include those that prevent vehicles from starting if alcohol is detected. This study aimed to develop and evaluate vehicle telltales to be used with these technologies. A total of 56 individuals, comprising 30 South Koreans and 26 Americans, participated in the evaluation. The evaluation employed indicators of visibility and intuitiveness. The analysis identified symbols consistently selected in the upper group, regardless of country, and characterized their attributes. Nevertheless, certain telltales exhibited cultural differences in interpretation and effectiveness. The procedures and findings of this study are expected to serve as a valuable reference for the design of telltales related to drunk driving prevention technologies. The cross-cultural evaluation underscores the significance of incorporating cultural nuances into the development of universally understood safety telltales, which contributes to enhancing global road safety through improved visual communication in-vehicle technologies.

Keywords: Drunk Driving Prevention Technology · Vehicle Telltale · Cultural Difference · Cognitive Evaluation

1 Introduction

Drunk driving is recognized as a serious problem threatening road safety and traffic regulations in contemporary society [1, 2]. Accidents caused by drunk driving highlight the need for the implementation of drunk driving prevention technologies. Consequently, various methods of prevention technologies for drunk driving are being developed. Recent technological advancements are providing a new dimension to the prevention of drunk driving in the driving environment. Particularly, the innovation of technology detecting and preventing drunk driving at the onset of vehicle operation is emerging as a crucial component of next-generation safety systems [3]. The advancement of such technology can offer drivers substantially enhanced safety. Currently, drunk driving prevention technology is being partially utilized, but it is anticipated that in the future, it will be implemented in all vehicles. Consequently, the telltale indicating the intoxicated state of the driver is assuming an even more critical role.

The telltale for vehicles is one of the important elements that provide visual alerts for various situations that occur during driving. Telltales are usually located in the vehicle's interior cluster and are indicated by small icons or lights. Additionally, telltales provide drivers with simple and clear information about engine status, safety system operation, and other aspects before driving. Non-standardized symbols can confuse drivers [4]. Therefore, it emphasizes the need for standard design guidelines for vehicle telltales.

Previous studies have evaluated symbols related to vehicles using cognitive indicators. One study introduced a rating methodology utilizing five indicators (concreteness, complexity, meaningfulness, familiarity, and semantic distance) to quantify symbol characteristics [5]. Another study evaluated graphic symbols and icons for public information signs using five criteria: visibility, intuitiveness, simplicity, aesthetic consistency, and psychological constraint [6]. Additionally, a third study defined and evaluated cognitive characteristics using simplicity, familiarity, concreteness, meaningfulness, and semantic closeness as the primary evaluation criteria for Dynamic Lane Departure Warning telltales [7]. The common features of the previous studies are as follows: they evaluated whether users accurately interpreted the meaning of symbols and assessed the visibility of symbols. The common metrics used in previous studies demonstrate their significance in assessing whether users interpret and understand symbols correctly. Therefore, this study derived two evaluation metrics from this perspective. The first metric is Visibility, and the second metric is Intuitiveness. In other words, they were used as important indicators of user (driver) cognition and interaction. In this study, symbols and telltales were considered synonymous.

The primary objective of this study is to conduct a systematic investigation into the creation of telltales that can serve as standards for drunk driving prevention technology and their effectiveness. It particularly focuses on technologies that detect drunk driving at the onset of vehicle operation. The study analyzed the evaluation results to establish standards for telltales suitable for drunk driving prevention technology and to derive insights into effective telltale design.

2 Telltale Design

In this study, 30 Korean university students were guided through task-based design to create vehicle telltales related to alcohol consumption. The experimenters briefed the participants beforehand on the purpose and evaluation criteria of the vehicle telltales. The evaluation criteria for the experiment were visibility and intuitiveness. Furthermore, as part of the instructions, an example telltale (washer fluid low warning light) was provided to reference the telltale's size and thickness. Subsequently, participants were instructed to design a (5 cm × 5 cm) black square to serve as the background resembling the vehicle cluster. Participants were to choose colors and shapes for their designs beyond the given example and reference. The following figure depicts the 30 telltales to be evaluated (Fig. 1).



Fig. 1. The South Korean participants created telltales numbered 1 through 30.

3 Telltale Evaluation

3.1 Participants and Evaluation Procedure

In this study, 30 university students from South Korea, with an average age of 22.5 years ($SD = 1.5$, $Min = 20$, $Max = 26$), and 26 residents from the Texas region in the United States, aged between 20 and 30 with an average age of 23.2 years ($SD = 3.6$, $Min = 20$, $Max = 35$), participated in the symbol evaluation. A total of 56 individuals participated in the experiment. Among the South Korean participants, 23 were male and 7 were female. All 25 participants from the United States had driving experience, with an average driving experience of 6.7 ± 3.8 years. The gender distribution among the United States participants was 19 males, 6 females, and one participant who chose not to disclose their gender. The evaluation of the 30 telltales utilized two criteria: visibility and intuitiveness, each rated on a 0–10 Likert scale. Subsequently, the scores from the evaluations of visibility and intuitiveness were used to classify the telltales into upper and lower groups based on their ratings.

3.2 Telltale Evaluation Results

The evaluation results were analyzed through technical evaluation analysis, Two-factor mixed design ANOVA, and Simple effects analysis. Figures 2 and 3 depict the data results for visibility and intuitiveness indicators for Korean and American participants using box plots.

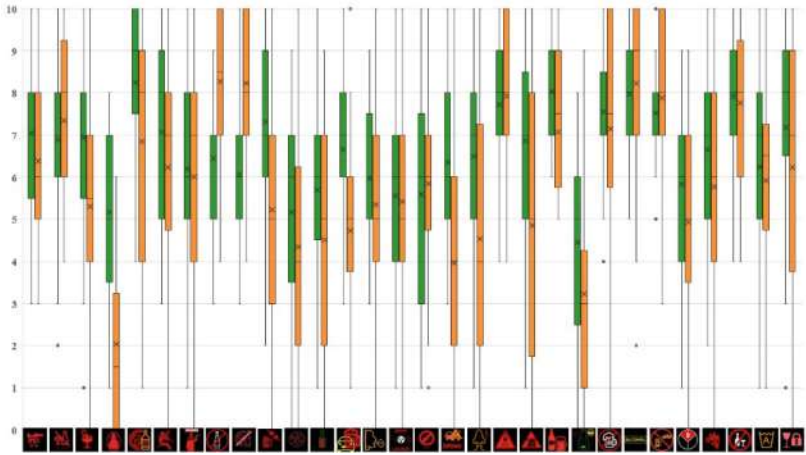


Fig. 2. Box plot of ‘Visibility’ scores for Korea (Green) and the United States (Orange).

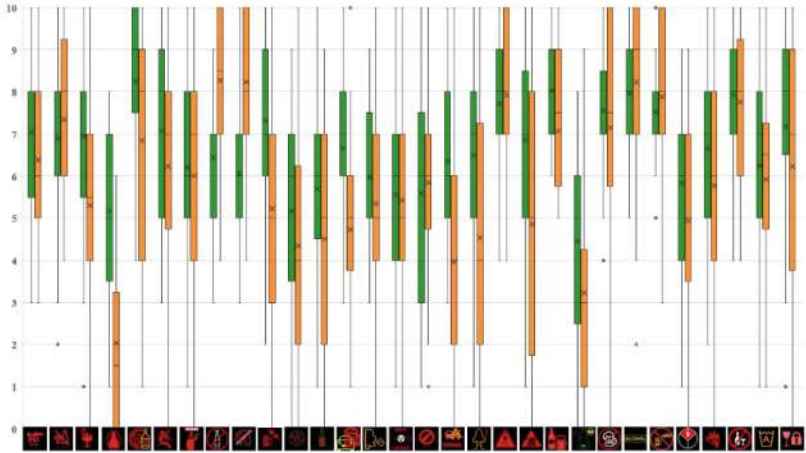


Fig. 3. Box plot of ‘Intuitiveness’ scores for Korea (Green) and the United States (Orange).

Through these box plots, the central tendency and dispersion of the evaluation scores for each telltale can be observed, providing an overall assessment of the results. Overall,

there is a larger difference in evaluation scores between countries in the intuitiveness indicator compared to the visibility indicator.

The two-factor mixed-design ANOVA was conducted with visibility and intuitiveness as dependent variables. Post-hoc analysis was performed using the LSD method, with the significance level set at 0.05. Based on the post-hoc analysis results, upper and lower groups were selected for each indicator (with telltales obtaining the highest and lowest scores compared and considering them as the same group if there were no significant differences). Finally, shared upper and lower groups were selected by comparing the upper and lower groups for each indicator.

Telltales 5, 25, and 28 were selected in the shared upper group. The common characteristics of these three telltales are as follows: Firstly, they all used the shape of a bottle or a glass, closely related to the purpose of the telltale, representing alcohol. However, they were used in different ways. Additionally, they all utilize a red slash ('/') symbol to convey the meaning of prohibition or warning. Telltales 5 and 25 use the symbol of a car, symbolizing the vehicle's steering wheel or body.

In the shared lower group, tell-tales 4 and 22 were selected. Telltales 4 and 22 share common characteristics: they both lack completeness in the images (the bottle shape can be interpreted as a regular bottle rather than one containing alcohol, and they lack detail). Overall, they use many thin lines, significantly lacking consistency. Additionally, both telltales use a red X symbol to convey the meaning of prohibition or warning. Telltale 22 is characterized by being unclear and complex in its interpretation.

4 Discussion

Through the preceding experimental results, we can summarize the characteristics of preferred telltales from a cognitive perspective as follows:

There was no significant difference between countries in terms of visibility indicators. However, there was a somewhat significant difference between countries regarding intuitiveness indicators. Both countries aimed for telltales with clear interpretations. This was further examined in detail through simple effects analysis.

Cultural differences between the two countries were evident in some telltales. In the case of the United States, tell-tales such as 3, 10, and 18, which used special (somewhat less intuitive and metaphorical) shapes like keys or plugs, or those like 18 and 20, which used exclamation marks or triangle symbols, and 12, 21, and 29, which only depicted alcohol (bottles, glasses), as well as 6, 7, 14, and 17, which depicted people or parts of the body, received lower cognitive evaluation scores. Conversely, South Korea tended to rate text-based telltales, such as 7, 15, 17, 20, and 24, slightly higher than in the United States. Contrary to expectations of significant cultural differences, there was no substantial difference between countries in the choice of shapes representing alcohol.

Ultimately, preferred telltales from a cognitive perspective are as follows: Firstly, cultural differences should be minimized. To minimize cognitive side effects due to cultural differences, it is necessary to develop icon-based telltales rather than text-based ones. Icon-based telltales are known to be more suitable for casual users as they are easier to learn and use [5]. In this case, universally recognized and familiar symbols should be used, avoiding metaphorical symbols and opting for intuitive ones. Additionally, as the

number of symbols used increases, users may perceive the interpretation of meaning as more complex [8], so it is necessary to use a minimal number of symbols.

The recommended telltales are numbers 5, 25, and 28. Their characteristics can be summarized as follows. Firstly, they utilize familiar and widely recognized shapes to minimize the impact of cultural differences [9, 10]. Secondly, they reduce visual complexity by employing a minimal number of shapes. Research suggests that fewer symbols result in better comprehension [8]. Therefore, for alcohol-related driving prevention telltales, it is advisable to use only three shapes: a bottle or glass for “Alcohol,” a car body or steering wheel for “Anti-drive,” and a slash (‘/’) for “Prohibited.”

5 Conclusion

As shown in the results of this study, cultural differences were observed in specific telltales. The unique cultural aspects of each country need to be carefully considered when developing telltale standards. Additionally, telltales should employ intuitive and simple symbols consistently. It is advisable to use a minimal number of shapes to convey the maximum amount of meaning, except for the prohibition symbol, which should consist of two shapes.

The process and findings of this experiment can serve as valuable guidance for designing telltales related to drunk driving prevention technologies and for creating future telltales.

References

1. Edwards, G.: Alcohol policy and the public good. *Addiction*. **92**(3s1), 73–80 (1997)
2. Baker, S.P.: *The Injury Fact Book*. Oxford University Press, USA (1992)
3. Ramanath, T.S., Sudharsan, A., Udhayaraj, U.F.: Drunken driving and rash driving prevention system. In: 2010 International Conference on Mechanical and Electrical Technology, pp. 599–603. IEEE, Singapore (2010, September)
4. Zulkefli, M. T. (2019). Automotive dashboard indicator: Consequences of different types of indicator to the driver
5. Maguire, M.C.: A review of human factors guidelines and techniques for the design of graphical human-computer interfaces. *Comput. Graph.* **9**(3), 221–235 (1985)
6. Han, S., Nah, K.: A study on the application of graphic signal design principles in the pictorial cover of Korean standard information network public information. *J. Digit. Des.* **10**(3), 453–462 (2010)
7. Maddahi, H., Pouyakian, M., Ghomsheh, F.T., Piri, L., Osqueizadeh, R.: Design and cognitive evaluation of 6 dynamic lane departure warning symbols. *Traffic Inj. Prev.* **17**(8), 842–847 (2016)
8. Shen, Z., Zhang, L., Li, R., Liang, R.: The effects of icon internal characteristics on complex cognition. *Int. J. Ind. Ergon.* **79**, 102990., ISSN 0169-8141 (2020). <https://doi.org/10.1016/j.ergon.2020.102990>
9. Cahill, M.C.: Interpretability of graphic symbols as a function of context and experience factors. *J. Appl. Psychol.* **60**(3), 376 (1975)
10. Vukelich, M., Whitaker, L.A.: The effects of context on the comprehension of graphic symbols. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **37**(8), 511–515 (1993)



A VR Study: How Should an Automated Vehicle Communicate with Multiple Pedestrians?

Yuchen Liu^(✉) , Niklas Grabbe , and Klaus Bengler 

Technical University of Munich, Boltzmannstrasse 15, 85748 Garching, Germany
yuchen.liu@tum.de

Abstract. In the near future, automated vehicles (AVs) should take on the responsibility to communicate with other road users, such as pedestrians. However, in complex urban scenarios, AV's external Human Machine Interfaces (eHMIs) may lead to confusion and danger while communicating with multiple road users. The present virtual reality (VR) study investigated the interaction between an AV and two pedestrians. 31 participants were instructed to cross an unsignalized two-lane street with the existence of an additional pedestrian. The approaching AV exhibited 3 behaviors and communicated with an eHMI concept. The two pedestrians started crossing on different roadsides with different lateral distances from the approaching AV. Each participant experienced 24 trials. Participants generally found the eHMI concept useful. However, misunderstandings and feelings of uncertainty were revealed, especially when participants were unsure whether the communication was intended for them or the other pedestrian. The other pedestrian's crossing behavior was regarded as a signal for a safe crossing; thus, a second brake for the farther pedestrian was wrongly expected. Therefore, considering the reciprocal affection among pedestrians on their crossing behavior, when there is more than one pedestrian in the traffic environment, AV should be able to communicate with multiple pedestrians and dynamically adjust the driving behavior and communication content.

Keywords: Automated driving · Vulnerable road users · external HMI · Multi-agent interaction · Virtual reality

1 Introduction

To ensure the successful integration of automated vehicles (AVs) into urban traffic, the challenge lies in establishing effective and efficient communication between AVs and other road users. While AVs' external Human Machine Interfaces (eHMIs) and dynamic Human Machine Interfaces (dHMIs) show the potential to convey AV's intentions, their performance in complex scenarios, particularly in urban areas with multiple road users, remains a concern [1, 2]. The potential misunderstanding about the addressed subjects might lead to an unsafe interaction and a lower acceptance of AV. Additionally, social dynamics among human road users can lead to reciprocal affection and influence their behavior when interacting with AVs [3]. This study focuses on analyzing the behavior of two pedestrians when interacting with an AV and evaluating the AV's communication concept.

2 Objectives

The main goal of this study was to investigate the interaction between an AV equipped with an eHMI concept and two pedestrians within an unsignalized crossing scenario. In this study, one pedestrian was controlled by the human participant, while the other was represented by a pre-programmed avatar referred to as the Confederate. Two research questions were defined in this study:

- How does the presence and behavior of the Confederate influence the crossing behavior of the human pedestrian when interacting with the AV?
- To what extent can AV's communication concept enhance safety and efficiency in situations where two pedestrians are in close proximity during a crossing?

3 Methodology

3.1 Virtual Environment

The virtual environment used in this study was implemented using Unity3D. Participants wearing a wired HTC Vive headset moved around in a $6 \times 10 \text{ m}^2$ labor. The scenario included an urban connecting road with two oncoming lanes of right-hand traffic. Two pedestrians were positioned laterally 3 m from each other across the road. An AV approached with an initial speed of 30 km/h from the same direction in all trials.

3.2 Communication Concept of AV

The AV was programmed to exhibit three different behaviors (*Brake_for_AV_Side* / *Brake_for_Opposite_Side* / *No_Brake*) and was equipped with an eHMI developed by [4], including two displays located under the blinkers and a 360° LED strip (see Fig. 1). A green arrow conveyed the AV's braking information and indicated that the pedestrian could cross before the AV. An orange hand indicated the AV would reserve the right-of-way and the pedestrian should wait. The 360° LED strip with a pulsating illumination symbolized the AV's detection of the addressed pedestrian.



Fig. 1. Illustration of the eHMI concept when the AV: a) brakes for the AV side; b) brakes for the opposite side; c) drives through.

3.3 Confederate Position and Behavior

The Confederate, presented by a pre-programmed avatar, was the other pedestrian starting with a 3 m lateral distance opposite to participants from the other roadside in the scenario. The start position of the Confederate was therefore determined by the participants' start position (*Player_Far* / *Player_Near* from the approaching AV, and *Player_on_AV_Side* / *Player_on_Opposite_Side*). To investigate the Confederate's influence on participants, the Confederate was programmed to exhibit two behaviors (*Cross* / *Not cross*). In trials when the Confederate performed the crossing behavior, the Confederate would initiate the crossing 1 s after AV initiated its communication and with a speed of 1.23 m/s.

3.4 Experimental Design

Twenty-four trials were implemented based on the *AV Behavior*, *Confederate Behavior*, *Player Side*, and *Player Distance* to the AV and experienced by each participant, while the order of trials was randomized. The study was approved by the Ethics Commission of the Technical University of Munich.

3.5 Dependent Variables

Dependent variables were measured with objective and subjective metrics, including:

- Effectiveness (*Number of Accidents* / *Feeling of Certainty* / *Number of Effectiveness Errors*);
- Efficiency (*Crossing Initiation Time* / *Number of Efficiency Errors*);
- Satisfaction (*Acceptance* / *Trust in Automation*).

In addition, participants were asked to fill out a questionnaire that included multiple qualitative questions.

4 Results

A total of 31 participants took part in the study. Of these, 14 were female and 17 were male. The age range was from 19 to 84 years ($M = 38.26$, $SD = 20.22$).

4.1 Crossing Effectiveness and Efficiency

The trials when the pedestrian on the AV side was nearer to the AV were found to have lower effectiveness and efficiency (See Fig. 2). The distribution of participants' *Feelings of Certainty* about their crossing behavior in all 24 Trials is shown in Fig. 3. A Kruskal-Wallis test was conducted to reveal the differences among the groups. Subsequently, post-hoc pairwise comparisons using the Dunn test with Bonferroni correction were performed. The results showed that *Player_Far* led to a significantly lower *Feeling of Certainty* compared to *Player_Near* ($p = 0.0202$), and *Brake_for_AV_Side* led to a significantly lower *Feeling of Certainty* than the other two AV Behaviors ($p < 0.001$).

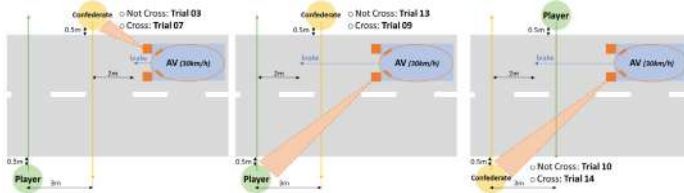


Fig. 2. Illustration of the scenario: a) Trial 03 & 07; b) Trial 09 & 13; c) Trial 10 & 14.

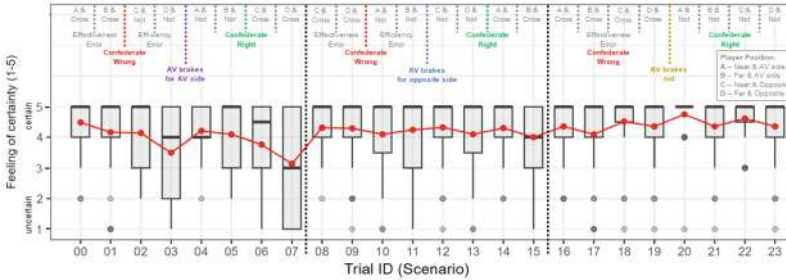


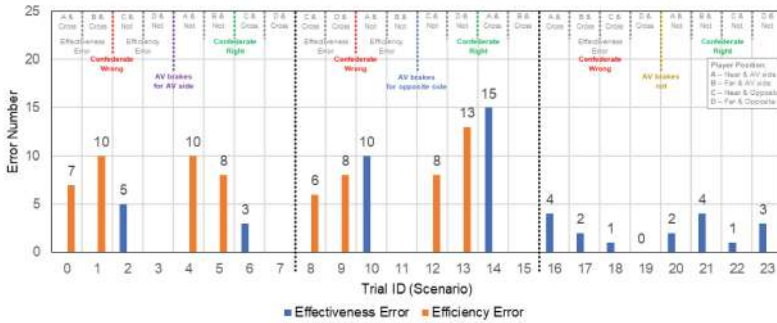
Fig. 3. Results of *Feeling of Certainty*.

No significant difference in the *Feeling of Certainty* was detected across different levels of *Confederate Behavior* ($\chi^2 = 0.50916$, $df = 1$, $p = 0.4755$) and *Player Side* ($\chi^2 = 1.2353$, $df = 1$, $p = 0.2664$).

Among all 24 Trials, participants showed the lowest *Feeling of Certainty* about the crossing behavior in Trial 07, when participants started from the furthest position to the AV while AV braked only for the Confederate (See Fig. 2). With the lowest *Feeling of Certainty* about their decision, 6 participants crashed with the AV in Trial 07.

Meanwhile, in Trial 09 and Trial 13, when participants started from the same position as Trial 07 with the furthest distance to the approaching AV and the AV braked for them, plenty of participants made an efficiency error (See Fig. 4). They crossed the road until the AV had entirely stopped in front of them. For crossings, when the AV braked for participants and no efficiency error occurred, a Wixcocon test was conducted to compare the *Crossing Initiation Time*. No significant difference was found for *Player Distance* ($W = 3702$, $p = 0.4027$) and *Confederate Behavior* ($W = 3364.5$, $p = 0.8284$). However, a significantly lower *Crossing Initiation Time* was performed in trials of *Brake_for_Opposite_Side* than *Brake_for_AV_Side* ($W = 4171$, $p = 0.01903$), although participants also made more efficiency errors. Based on a deeper look at the trajectory files and the post-experimental questionnaire, it was found that when starting from the opposite roadside, many participants stayed and waited for a while on the road before entering the AV lane, which was confirmed as a risky crossing behavior.

Another risky crossing behavior that happened in the study was effectiveness error, which referred to a crossing before the AV, although the AV did not brake for the participants. Most effectiveness errors occurred when participants started from the nearest position to the AV (Trial 10 and 14) (See Fig. 4). Especially in Trial 14, almost half of



the participants made an effectiveness error and crossed the road before the AV, although the AV braked only for the Confederate (See Fig. 2).

5 Discussion

behavior of participants, the existence of an additional pedestrian has affected the perception of participants on the AV. In the post-experimental qualitative questionnaire, participants reported the influence of the Confederate on them. The crossing behavior of the Confederate was regarded as a signal for a safe crossing. The interaction between the Confederate and the AV has altered AV's driving behavior and thus influenced the interaction between participants and the AV. The critical situation happened when the pedestrian on the AV side was also laterally nearer to the AV. A second brake for the opposite farther pedestrian was wrongly expected, especially when the nearer pedestrian initiated the crossing after AV braked for them.

6 Conclusion

In summary, considering the influence of the additional pedestrian on the pedestrian's crossing behavior, when there is more than one pedestrian in the traffic environment, AV equipped with an eHMI concept should communicate with both pedestrians, be aware of their behavior, and be able to dynamically adjust its driving behavior and communication content when risky crossing behavior is detected. The authors recommend further experiments with this paradigm to investigate multiparty effects on crossing decisions.

Acknowledgments. This work is a result of the joint research project STADT:up. The project is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), based on a decision of the German Bundestag. The author is solely responsible for the content of this publication.

References

1. Dey, D., van Vastenhoven, A., Cuijpers, R.H., Martens, M., Pfleging, B.: Towards scalable eHMIs: Designing for AV-VRU communication beyond one pedestrian. In: 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 274–286. ACM, New York, NY, USA (2021). <https://doi.org/10.1145/3409118.3475129>
2. Hübner, M., Feierle, A., Rettenmaier, M., Bengler, K.: External communication of automated vehicles in mixed traffic: Addressing the right human interaction partner in multi-agent simulation. *Transport. Res. F: Traffic Psychol. Behav.* **87**, 365–378 (2022). <https://doi.org/10.1016/j.trf.2022.04.017>
3. Jiang, Y., et al.: Acting together: Joint pedestrian road crossing in an immersive virtual environment. *ACM Trans. Appl. Percept.* **15**(2), 1–13 (2018). <https://doi.org/10.1145/3147884>
4. Hübner, M., Baude, M., Bengler, K.: Pedestrian groups matter: Unraveling their impact on pedestrian crossings when interacting with an automated vehicle. In: 35th IEEE Intelligent Vehicles Symposium. IEEE, Jeju Island, Korea (2024). <https://doi.org/10.1109/IV55156.2024.10588543>



Evolved Comfort in the Era of Autonomous Driving: Emic/Etic Approach to Reformulate It

Saeedeh Mosaferchi^{1b}, Alireza Morteza pour^{1b}, and Alessandro Naddeo^{(✉) 1b}

University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
anaddeo@unisa.it

Abstract. The concept of driving comfort has evolved significantly with advancements in automobile technology and the emergence of self-driving vehicles. This study uses the emic/etic approach to redefine driving comfort. While manufacturers have successfully addressed physical comfort (e.g., vibration control and noise reduction), modern comfort now encompasses new aspects, which will continue to evolve with the widespread adoption of self-driving cars. In this study, an attempt has been made to achieve a comprehensive definition of comfort while using autonomous vehicles, according to some interviews which were conducted with related experts. This study is inspired by the emic/etic approach. This approach is a successful model in presenting new concepts in the psychology (personality, etc.), which has recently been tested in the interaction of humans with technology. In the preliminary examination of the results, comfort first and foremost refers to more than just physical comfort and absence of vibration. The study addresses new issues like meeting passengers' technical demands and gaining their trust in self-driving cars' decision-making. Increasing research is focusing on comfort in autonomous vehicles, particularly human-technology interaction. Expectations for self-driving cars differ from conventional ones, with people wanting autonomous vehicles to ensure comfort and safety through human-like driving patterns and to build trust by making reliable decisions, while humans monitor the driving.

Keywords: Autonomous vehicles · Emic and etic · Human-machine interaction · Trust to technology · Physical comfort

1 Introduction

1.1 Background and Objectives

There is a greater emphasis on assessing passenger comfort as a result of several developments that have occurred recently. This is due to the fact that a lot of daily tasks these days require cars, and people spend a lot of time in them, whether they are their own or someone else's [1]. Many factors motivate the improvement of riding comfort. In addition to increasing safety and dramatically lowering passenger weariness, a pleasant environment also improves driving performance. The health and wellbeing of passengers will also be improved, as musculoskeletal disorders will be avoided [2]. On the other hand, a number of variables influence how comfortable cars are. The perceived comfort

of individuals in an automobile can be greatly influenced by various factors such as thermal comfort, the absence of strange or annoying noises from the engine or interior, the brand of the car manufacturer, the design of the seat's structure or cover material, vibration-free operation, high-quality materials, and many more [3–6].

Table 1. Search Strategy for the literature review.

All synonyms: (“Car” OR “automobile” OR “vehicle” OR “automated car*” OR “autonomous car*” OR “self-driving” OR “self-driving” OR “self-driving” OR “driverless” OR “vehicle automation” OR “automated driving” OR “autonomous vehicle*” OR “automated vehicle*”)	AND	All synonyms: (“Comfort” OR “discomfort” OR “dis-comfort” OR “UX” OR “ride experience” OR “user experience” OR “pressure” OR “driving experience”)

Table 2. Emic-Etic classification of comfort parameters.

Etic	Same value	Emic
Static vs dynamic muscle comfort	Vibration	The quality of multimedia system
Annual maintenance cost	Posture of users /Joint angles	Negotiation with others (pedestrians, bicyclists, other cars, etc.)
Acoustic comfort	Vehicle Kinematic	The sense of car control/ Car mastering/ Controlling the car speed
The role of brands in comfort induction (Waymo vs. Mercedes)	Annual service cost	Having various entertainment systems/ facilities available combine with the above
	Amount of available cargo space	Screen size
	Quality of interior parts (quality of manufacturing)	Driving style (humanlike driving)/ Lane change timing/ pattern
	Quality of exterior parts (quality of manufacturing)	The clarity of the interfaces’ function (usability)
	Passenger safety (perceived safety)	Cyber security (Privacy of personal information)
	Necessary car features such as temperature adjustment, map reader (temperature comfort)	Visual context (window to car body ratio)

(continued)

Table 2. (continued)

Etic	Same value	Emic
	Trust in car performance (Dynamic Trust)	Communication strategy (explainability of car-human interaction)
	A sense of distinction and uniqueness	Access to rare cues such as olfactory and haptic cues
	Hedonic comfort (Aesthetic parameters)	Level of involvement in the loop (situational awareness)
	Controller layout and reachability	Availability of predictive feedback
	Initial trust	Confidence in an external supervision presence
	Peace of mind	Compatibility with Lifestyle (e.g., doing make up, online shopping, etc.)/ human behaviour detection
		clarity of responsibilities (liability)
		Social interactivity condition
		Sensory conflict (Coordination between visual perception and balance)
		Availability of fuel stations (comparison of electric fuel stations and gas/gasoline)

Car technology has advanced significantly over time and with the arrival of industrial revolutions. In particular, the development of self-driving automobiles has led to the definition of many characteristics pertaining to passenger comfort paragraphs, however, are indented [7, 8]. For instance, prior to the development of high-tech automobiles like electric cars, the most significant elements influencing end users' comfort were things like the interior noise level of the vehicle and the body vibration of the occupants [9–11]. In general, it can be claimed that physical and biomechanical approaches were considered earlier [12]. While the elements that determine human comfort have also been modified to suit them, automobiles now have a variety of interfaces installed, including touch and multipurpose displays, as well as a plethora of features like cruise control, automated parking, and lane change. Thus, new technological end users' demands have been raised in recent years [13] (Table 1 and 2).

By conducting numerous interviews and meetings with experts in the field of car comfort, the researchers in this study attempted to extract a number of factors affecting the comfort of car manufacturers. Their position was established from the perspective

of the team of experts based on the thorough review conducted prior to the interviews, regarding the emic-etic approach.

2 Methodology

2.1 Emic/etic Approach

Started from psychology, emic/etic approach is a cross-cultural analogy concept that implies on a difference between old-fashioned and modern view to a parameter. Here it means that comfort can be considered as an evolving concept. In the past decades some physical parameters such as vibration of the car, angles of the seat, reachability to buttons, etc., define the main part of the comfort in the cars. But nowadays with introducing autonomous vehicles, some new parameters also must be considered to define comfort including trust to decision making of the car, cyber security, and so on.

With considering the evolving pattern of comfort, the current study was developed based on emic/etic approach. This approach was also used in other studies related to human-technology interaction [14].

2.2 Literature Review

A comprehensive search was carried out in the Scopus and Web of Science databases and then completed by Google Scholar to identify the variables influencing the “comfort of passengers in a car”, and the resulting publications were then reviewed. Additionally, the following is the search strategy that was employed to reach relevant articles. After screening all abstracts, 50 papers were skimmed completely to extract the effective parameters on people’s comfort while using a car, in addition to defining the trend of relevant researchers. Finally, groups of statements were prepared and discussed by the research team and raised in the meetings and interviews which were held with the expert panel.

2.3 Interviews with Expert Panel Members

Some academic and lay experts (Mechanical engineering, Human factors and Ergonomics, Safety) who were involved and experienced in car comfort, were invited to the interview. Meetings and interviews were held in separate settings to avoid the influence of expert opinions and potential bias. After explaining the study goals and introduction of emic-etic approach, the list of statements was shown to the expert panel. 38 impressive parameters were discussed in influencing end users’ comfort in the vehicle and were put in a column based on 3 options (Etic = significant in old cars, Same value for old and new cars, Emic = significant in new cars). The etic column they selected, suggested that the parameter was taken into account in older cars and had a higher bearing on the comfort of the occupants. As far as experts’ perceptions of comfort in old and new cars (including self-driving cars Level 4 or 5) are concerned, the “same value column” denotes the same relevance of that characteristic. As the emic was chosen, it became apparent that this characteristic had a bigger effect on passengers’ comfort in more recent models, and even some experts believed that designers and automakers had overlooked it.

3 Results

The findings of this study revealed significant insights into the factors influencing passenger comfort in vehicles. Expert panel members confirmed gradual changes in effective parameters on end users' comfort, but they added some statements in addition to dividing some of them into 2 separate factors, as well as rewriting a few words. In fact, according to the experts involved in the study, there has been a significant shift in the parameters influencing how comfortable car passengers consider themselves. Because of the differences in the nature and capabilities of automobiles in the past, designers and automakers are now debating and analysing issues that were either overlooked or deemed insignificant. At the end, 38 parameters were defined as effective agents on people's comfort while using a car, regarding emic-etic approach, which are mentioned below:

4 Discussion and Conclusion

After introducing the concept of emic/etic to the experts, they also agreed with the research team that the concept of comfort, from the past to the present, has made a significant difference in different generations of cars and is still evolving. Similar to these changes, researchers in other studies also believe that comfort in self-driving cars has differences with this concept in traditional cars [15].

In conventional vehicles, comfort is commonly related to physical Ergonomics, but new concept of comfort is encompassed from parameters which are more inclined to cognitive Ergonomics. Passing from physical Ergonomics concerns such as vibration, reachability of interfaces, adjustability of the seat into cognitive Ergonomics parameters such as trust to AVs, transparency, cyber security and sense of controllability are discussed and agreed in the current study. As the old parameters have been evident for decades, the importance of the new parameters is cited in the recent publications [16, 17].

5 Limitations and Future Works

The limited number of panel members is one of the study's limitations. The richness of the findings would be improved by increasing the number of participants and their national diversity. For the future studies, the current team is developing another study to check the weight of each parameter in comparison to another ones.

Acknowledgements. The authors would like to extend their sincere gratitude to the esteemed experts who generously gave their time and brilliant opinions to shape and improve this work.

References

1. Da Silva, M.G.: Measurements of comfort in vehicles. *Meas. Sci. Technol.* **13**(6), R41 (2002 Apr 30)
2. Ormuž, K., Muftić, O.: Main ambient factors influencing passenger vehicle comfort. *Ergonomics*. **21**, 77–84 (2004 Oct)
3. Nor, M.J., Fouladi, M.H., Nahvi, H., Ariffin, A.K.: Index for vehicle acoustical comfort inside a passenger car. *Appl. Acoust.* **69**(4), 343–353 (2008 Apr 1)
4. Mosaferchi, S., Califano, R., Naddeo, A.: Shed light on the path of human-machine interaction in autonomous vehicles: where did we come from? Part II, search strategy, findings and perspectives. In: *International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering*, pp. 327–334. Springer Nature Switzerland, Cham (2023 Sep 6)
5. Jiang, Z., Streit, D.A., El-Gindy, M.: Heavy vehicle ride comfort: Literature survey. *Int. J. Heavy Veh. Syst.* **8**(3–4), 258–284 (2001 Jan 1)
6. Danca, P., Vartires, A., Dogeanu, A.: An overview of current methods for thermal comfort assessment in vehicle cabin. *Energy Procedia*. **1**(85), 162–169 (2016 Jan)
7. Dillen, N., Ilievski, M., Law, E., Nacke, L.E., Czarnecki, K., Schneider, O.: Keep calm and ride along: Passenger comfort and anxiety as physiological responses to autonomous driving styles. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. Association for Computing Machinery, New York NY United States (2020 Apr 21)
8. Wu, Z., Liu, Y., Pan, G.: A smart car control model for brake comfort based on car following. *IEEE Trans. Intell. Transp. Syst.* **10**(1), 42–46 (2008 Nov 5)
9. Huang, H., Huang, X., Ding, W., Yang, M., Fan, D., Pang, J.: Uncertainty optimization of pure electric vehicle interior tire/road noise comfort based on data-driven. *Mech. Syst. Signal Process.* **15**(165), 108300 (2022 Feb)
10. Du, X., Sun, C., Zheng, Y., Feng, X., Li, N.: Evaluation of vehicle vibration comfort using deep learning. *Measurement*. **1**(173), 108634 (2021 Mar)
11. Silveira, M., Pontes Jr., B.R., Balthazar, J.M.: Use of nonlinear asymmetrical shock absorber to improve comfort on passenger vehicles. *J. Sound Vib.* **333**(7), 2114–2129 (2014 Mar 31)
12. Gross, C.M., Goonetilleke, R.S., Menon, K.K., Banaag, J.C., Nair, C.M.: The biomechanical assessment and prediction of seat comfort. In: *Hard Facts About Soft Machines*, pp. 231–253. CRC Press (2020 Dec 17)
13. Wang, C., Zhao, X., Fu, R., Li, Z.: Research on the comfort of vehicle passengers considering the vehicle motion state and passenger physiological characteristics: Improving the passenger comfort of autonomous vehicles. *Int. J. Environ. Res. Public Health*. **17**(18), 6821 (2020 Sep)
14. Babamiri, M., Heidarimoghadam, R., Ghasemi, F., Tapak, L., Morteza pour, A.: Personality traits and willingness to use a robot: extending emic/etic personality concept. *Int. J. Soc. Robot.* **16**, 1–1 (2024 Apr)
15. Peng, C., et al.: Conceptualising user comfort in automated driving: Findings from an expert group workshop. *Transp. res. interdiscip. perspect.* **1**(24), 101070 (2024 Mar)
16. Su, H., Brooks, J., Jia, Y.: Development and evaluation of comfort assessment approaches for passengers in autonomous vehicles. *SAE Int. J. Adv. Curr. Pract. Mobil.* **5**. (2023-01-0788), 2068–2077 (2023 Apr 11)
17. Kaszas, D., Roberts, A.C.: Comfort with varying levels of human supervision in self-driving cars: Determining factors in Europe. *Int. J. Transp. Sci. Technol.* **12**(3), 809–821 (2023 Sep 1)



Differences in Fatigue Levels and Driving Performance Based on Circadian Rhythms and Scene Transitions

Po Chun Nien^(✉), Pei Hsuan Wu, Ting Hsuan Wang, and Yung Ching Liu

National Yunlin University of Science and Technology, Yunlin, Taiwan
jimmynien0510@gmail.com

Abstract. This study aims to investigate the effects of different chronotypes, time periods, and contextual transitions on driving fatigue and performance. A total of 16 participants were recruited for a 2 (chronotype) * 2 (time period) * 4 (context) mixed-factorial design experiment. Participants were required to react to sudden events while driving normally and to record their fatigue levels every 5 min. The results of the study indicate that under fatigue conditions, night-oriented drivers exhibit a greater attentional load in driving behavior. Furthermore, based on several driving indicators, it may be suggested that complex road environments demand increased attention. Transitioning from driving in complex scenes to monotonous scenes shows a temporary increase in fatigue levels, indicating a potential influence of contextual transitions on fatigue levels.

Keywords: circadian rhythm · driving fatigue · fatigue transfer

1 Introduction

Fatigued driving is a significant traffic safety concern in road accidents. According to the National Highway Traffic Safety Administration (NHTSA), there were 91,000 crashes related to fatigued driving in 2017, resulting in 50,000 injuries and nearly 800 fatalities (NHTSA, 2017). Moreover, from 2017 to 2021, fatigued driving accounted for 1.8% of fatal crashes (NHTSA, 2023). Prolonged driving is a major cause of fatigue-related accidents on highways or main roads, as it severely impairs drivers' alertness and performance. Fatigue can also result from sleep deprivation or irregular lifestyles [12].

Circadian rhythms refer to the 24-hour cycle of physiological processes, and individual differences cause people to perform differently at various times of the day [7]. These differences can be categorized using the Morningness-Eveningness Questionnaire (MEQ), MEQ scores range from 16 to 86. Scores of 59–86 indicate morning types, 16–41 indicate evening types, and the remaining scores indicate intermediate types [8]. This time preference affects people's behavior [1, 6], as fatigue, lack of concentration, and decreased alertness are more likely to occur during the biological low points [2, 10]. Studies have shown that morning types are more likely to engage in safe driving behaviors and have fewer driving errors and violations compared to evening types [3].

Hwang et al. measured drivers' fatigue using a driving simulator with long, monotonous highway scenarios. The results indicated that prolonged driving time is a potential cause of fatigue-related accidents. Another study investigating the impact of different road conditions on drivers' driving and cognitive behaviors found that transitions from complex to simple road conditions exacerbated driver fatigue [9]. Therefore, fatigue transfer is another factor to consider that can influence driving behavior [11, 13].

Therefore, the aim of this study is to investigate whether individuals with different chronotypes exhibit better performance and lower levels of fatigue during their corresponding time periods. Additionally, we seek to examine whether driving fatigue intensifies as situations transition from complex to monotonous due to fatigue transfer.

2 Methodology

2.1 Participants

Sixteen participants were recruited, consisting of 8 morning types (with an average MEQ score of 67.6) and 8 evening types (with an average MEQ score of 34.8). Their ages ranged from 19 to 24 years, with an average age of 22.5 years ($SD = 1.1$). All participants held valid driver's licenses and had normal or corrected-to-normal vision.

2.2 Equipment

This study used a vision tester (SHIN-NIPPON CP-500) to measure whether participants' vision met the required conditions. The driving simulator used in the experiment was the STISIM Model 300, paired with a car body (VOLVO DL340). Driving scenarios were displayed on a front screen using an EPSON projector (EB-2065).

2.3 Participant Tasks

During the driving simulation experiment, participants were required to follow all traffic rules, including maintaining a speed limit of 60 km/h and staying in the outermost lane.

Divided Attention Task: On the left and right sides of the driving scenario, red diamond icons were displayed, which randomly changed to red triangles at various locations on the road. When the icon changed, participants had to press the corresponding button on the steering wheel. This task aimed to assess differences in attention response among drivers.

Subjective Sleepiness Rating Task: Every 5 min after the start of the drive, researchers asked participants to rate their current level of fatigue using the Stanford Sleepiness Scale (SSS). Participants had to respond with a score representing their current fatigue level (ranging from 1 to 7), with higher scores indicating greater fatigue.

2.4 Road Scenarios

The experiment took place in the Driving Simulation Laboratory at the National Yunlin University of Science and Technology. The driving scenarios were divided into two time

periods: daytime and nighttime. The total length of the road environment was 70 km, and the overall driving time was approximately 70 min.

To increase driver fatigue, the first 40 min consisted of a monotonous and straight road environment without curves. The remaining 30 min were divided into two scenarios: one transitioned from complex urban roads to monotonous rural roads, while the other transitioned from monotonous rural roads to complex urban roads. Each scenario lasted for 15 min and included various road traffic events, such as vehicles running red lights from the right, oncoming vehicles making left turns, and pedestrians crossing the road from the right. These events occurred randomly during the scenarios.

2.5 Experimental Design

The experiment utilized a 2 (driver chronotype: morning type, evening type) * 2 (time period: daytime, nighttime) * 4 (scenario: initial monotonous, later complex, initial complex, later monotonous) mixed factorial design. Driver chronotype and scenario were between-subject factors, aiming to investigate fatigue levels and driving performance among drivers with different chronotypes and fatigue transitions in different scenarios. Time period was a within-subject factor, aiming to investigate fatigue levels and driving performance on different road segments during different times of the day.

2.6 Experimental Procedure

After signing the consent form, participants familiarized themselves with the driving simulator for 5 min, followed by a 5-min rest before starting the experiment. The experiment was conducted over two days, with sessions scheduled at 8 AM and 10 PM. After completing the experiment, participants filled out a feedback form and were then allowed to leave. Each session lasted approximately 90 min.

2.7 Data Analysis

The collected data for driving behavior performance includes: (1) longitudinal speed, (2) longitudinal acceleration, (3) throttle longitudinal acceleration, (4) lane deviation, (5) braking longitudinal acceleration, (6) lateral acceleration, and (7) steering wheel angle. The data are collected via the simulator.

For the divided attention task, the calculation methods for correctness rate and response time are as follows: (1) Correctness rate: Number of correct responses divided by the total number of tasks. (2) Response time: Time taken by the participant to press the button after the divided attention stimulus changes. Response times exceeding 5 s are not counted. (3) The Stanford Sleepiness Scale (SSS) fatigue score is verbally recorded every five minutes.

3 Results

The analysis was conducted using IBM SPSS Statistics 22 software. Differences in driving behavior and subjective evaluations among different chronotypes in different time periods and scenarios were explored. Post-hoc analyses were performed using the Least Significant Difference (LSD) test, with a significance level set at $\alpha < 0.05$.

3.1 Fatigue Levels

Figure 1 displays the fatigue levels of drivers at fourteen positions while driving on the road. There is a significant difference in fatigue levels every five minutes ($p = 0.000$), with the highest fatigue point reached at 40 min. The fatigue level during scenario transitions shows a temporary increase as drivers transition from complex to monotonous scenes (Fig. 2). Additionally, there is a significant difference in fatigue levels between daytime and nighttime driving ($p = 0.009$), with higher fatigue levels observed during daytime compared to nighttime (Fig. 3).

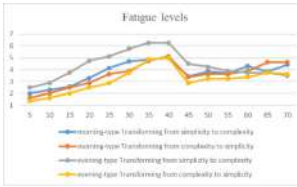


Fig. 1. scenario transitions and time periods.

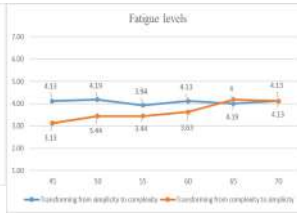


Fig. 2. scenario transitions.

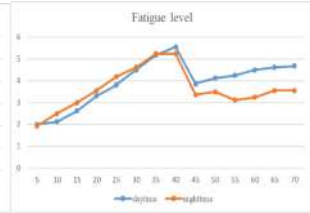


Fig. 3. time periods.

3.2 Driving Behavior

From the analysis of variance, it is found that there are no significant differences in the mean longitudinal speed, mean longitudinal acceleration, longitudinal acceleration of the throttle, lane deviation, and longitudinal braking acceleration.

The analysis of variance for average lateral speed shows significant differences in chronotype ($F(1,48) = 5.066$, $p = 0.029$) and context ($F(3,48) = 8.550$, $p = 0.000$), with no significant interaction effect. Post-hoc analysis reveals that night types are significantly higher than morning types ($p = 0.029$); the average lateral speed in the complex context is significantly higher than in the monotone context ($p = 0.043$); the average lateral speed in the complex context after transitioning from the monotone context is significantly higher ($p < 0.000$), while it significantly decreases when transitioning from the complex context to the monotone context ($p = 0.001$).

In the analysis of variance for steering wheel angle, significant differences are observed in context ($F(3,48) = 4.116$, $p = 0.008$), with no significant interaction effect. Post-hoc analysis indicates that the steering wheel angle is significantly higher in the complex context compared to the monotone context ($p = 0.024$), significantly higher in the monotone context compared to the complex context ($p < 0.011$), significantly increases when transitioning from the monotone context to the complex context ($p = 0.017$), and significantly decreases when transitioning from the complex context to the monotone context ($p = 0.015$).

3.3 Divided Attention

The analysis of variance for reaction time in the divided attention task shows no significant differences in chronotype, time of day, or context.

The analysis of variance for accuracy in the divided attention task reveals an interaction effect between chronotype and time of day ($F(1,48) = 5.453, p = 0.024$). Fig. 4 is a line graph illustrating the interaction effect on divided attention accuracy, showing that morning-type participants have higher accuracy during the day compared to the night, while evening-type participants have higher accuracy at night compared to the day.

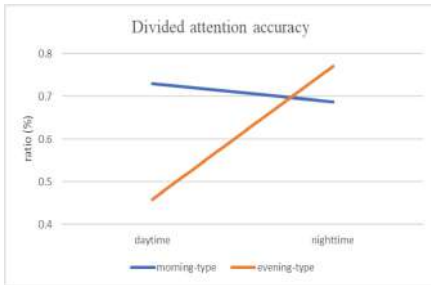


Fig. 4. line graph of divided attention accuracy.

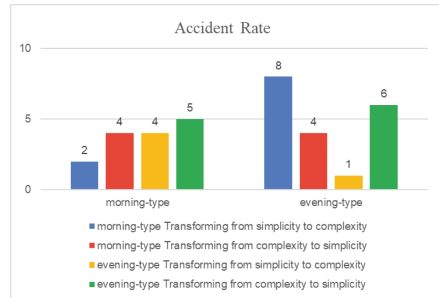


Fig. 5. Bar chart of accidents. rate.

3.4 Accident Rate

Figure 5 shows that morning-type participants have fewer accidents during the day than at night, with the highest number of accidents occurring in the complex-to-monotonous transition at night. Conversely, evening-type participants have fewer accidents at night compared to the day, with the highest number of accidents occurring in the monotonous-to-complex transition during the day. Figure 6 shows that morning-type participants

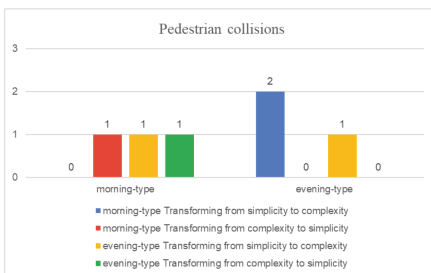


Fig. 6. Bar chart of pedestrian collisions.

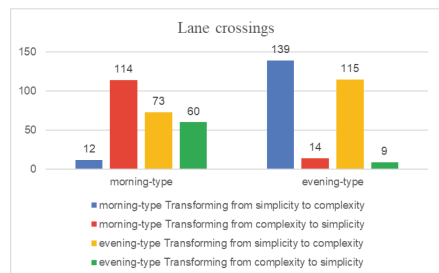


Fig. 7. Bar chart of lane crossings.

have fewer pedestrian collisions during the day compared to at night, while evening-type participants have fewer pedestrian collisions at night compared to during the day. Figure 7 shows that morning-type subjects have fewer lane crossings during the day compared to at night, while evening-type subjects have fewer car accidents at night compared to during the day.

4 Discussion

In the first 40 min of this study, the scenario was designed to be monotonous and straight without any curves to induce fatigue in the participants. Figure 1 shows that after the start of the drive, the fatigue levels of the participants increased over time, reaching a peak at the 40-min mark. However, when the scenario switched, a decrease in fatigue levels was observed.

4.1 Impact of Chronotype and Time Period on Fatigue Driving Behavior and Performance

According to Table 1, there is a significant difference in lateral acceleration between different chronotypes. The average lateral acceleration of evening-type drivers is greater than that of morning-type drivers, indicating that evening-type individuals are more prone to sudden left or right deviations while driving. This suggests that evening-type drivers experience a higher attentional load regardless of whether they are driving in the morning or at night.

Figure 3 shows that morning-type drivers exhibit more stable driving performance during the day compared to evening-type drivers. Morning-type drivers are more attentive to their surroundings and drive more cautiously [4]. Conversely, evening-type drivers demonstrate better alertness at night than morning-type drivers, indicating that each chronotype performs better and experiences lower fatigue levels during their corresponding optimal times of day [5].

4.2 Impact of Different Road Scenarios on Fatigue Driving Behavior and Performance

Although subjective fatigue levels during scenario transitions did not show significant differences, variations were observed. When driving from complex to monotonous scenarios, there was a slight increase in fatigue levels (Fig. 2), suggesting that changes in driving environments have a potential impact on fatigue levels. Lateral acceleration also showed significant differences across different scenarios. Lateral acceleration significantly increased when transitioning from monotonous to complex roads and decreased when transitioning from complex to monotonous roads. This indicates that driving in complex road environments requires more focused attention. Drivers tend to be more cautious due to the presence of pedestrians and other vehicles.

Additionally, similar results were found for steering wheel angle variations, further confirming that drivers exhibit higher alertness in complex environments. This increased vigilance likely stems from the need to navigate more carefully in the presence of surrounding pedestrians and vehicles.

5 Conclusions

5.1 Recommendations and Applications

Similar to findings from previous studies, this research offers the following recommendations: Long-distance drivers should consider driving during times that correspond to their circadian type to avoid potential risks. Considering fatigue transfer effects, driving from a complex road environment to a relatively monotonous environment may exacerbate fatigue. Therefore, during such transitions, drivers should take appropriate measures, such as taking breaks or adjusting their driving patterns, to reduce the physiological burden of prolonged driving. Transportation safety authorities should also consider the impact of fatigue transfer when planning roads. They could install relevant warning signs or messages on monotonous rural roads or highways to remind drivers to be aware of potential fatigue transfer and to remain alert, thereby reducing the risk of fatigue-related accidents.

References

1. Adan, A., Archer, S.N., Hidalgo, M.P., Di Milia, L., Natale, V., Randler, C.: Circadian typology: A comprehensive review. *Chronobiol. Int.* **29**(9), 1153–1175 (2012)
2. Anderson, C., Horne, J.A.: Driving drowsy also worsens driver distraction. *Sleep Med.* **14**(5), 466–468 (2013)
3. Bergomi, M., Vivoli, G., Rovesti, S., Bussetti, P., Ferrari, A., Vivoli, R.: Role of some psycho-physiological factors on driving safety. *Ann Ig.* **22**(5), 387–400 (2010)
4. Correa, Á., Molina, E., Sanabria, D.: Effects of chronotype and time of day on the vigilance decrement during simulated driving. *Accid. Anal. Prev.* **67**, 113–118 (2014)
5. Del Rio-Bermudez, C., Diaz-Piedra, C., Catena, A., Buella-Casal, G., Di Stasi, L.L.: Chronotype-dependent circadian rhythmicity of driving safety. *Chronobiol. Int.* **31**(4), 532–541 (2014)
6. Dijk, D.J., Duffy, J.F., Czeisler, C.A.: Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *J. Sleep Res.* **1**(2), 112–117 (1992)
7. Gillette, M.U., Abbott, S.M.: Basic mechanisms of circadian rhythms and their relation to the sleep/wake cycle. In: *Neuroendocrine correlates of sleep/wakefulness*, pp. 19–40. Springer US, Boston, MA (2006)
8. Horne, J.A., Brass, C.G., Petitt, A.N.: Circadian performance differences between morning and evening ‘types’. *Ergonomics.* **23**(1), 29–36 (1980)
9. Liu, Y.C., Wu, T.J.: Fatigued driver’s driving behavior and cognitive task performance: Effects of road environments and road environment changes. *Saf. Sci.* **47**(8), 1083–1089 (2009)
10. Natale, V., Cicogna, P.: Circadian regulation of subjective alertness in morning and evening ‘types’. *Personal. Individ. Differ.* **20**(4), 491–497 (1996)
11. Philip, P., et al.: Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep.* **28**(12), 1511–1516 (2005)
12. Saleh, A.M., Awadalla, N.J., El-masri, Y.M., Sleem, W.F.: Impacts of nurses’ circadian rhythm sleep disorders, fatigue, and depression on medication administration errors. *Egypt. J. Chest Dis. Tuberc.* **63**(1), 145–153 (2014)
13. Ting, P.H., Hwang, J.R., Doong, J.L., Jeng, M.C.: Driver fatigue and highway driving: A simulator study. *Physiol. Behav.* **94**(3), 448–453 (2008)
14. Zhao, X., Rong, J.: The relationship between driver fatigue and monotonous road environment. In: Wang, W., Wets, G. (eds.) *Computational intelligence for traffic and mobility*, vol. 8, pp. 19–36. Atlantis Press, Paris (2012)



Exploring Driving Comfort and Vehicle Usability in Designing E-Cargo Bikes for Last Mile Delivery as Sustainable Urban Mobility Solutions

Alessandra Rinaldi^(✉)  and Jonathan Lagrimino 

Innovation in Design and Engineering Laboratory, Department of Architecture,
University of Florence, Via Sandro Pertini 93, 50041 Calenzano, Florence, Italy
{alessandra.rinaldi, jonathan.lagrimino}@unifi.it

Abstract. To address the sustainable urban mobility challenge, the European Union adopted a strategic approach that goes beyond replacing traditional vehicles with electric vehicles and instead proposes new mobility paradigms, focused on users' needs and their multimodal, automated, and connected travel experience [1].

The sector of light vehicles represents one of the most promising solutions due to lots of advantages: sustainable and healthy alternative to single occupancy car issues; unbeatable low energy requirements and consumption; higher social accessibility; space saving [2]. Regarding the various urban mobility solutions, the need to design light 2–3 wheels vehicles for the last mile delivery, to increase safety and driveability by using new materials, sensors and vehicle connectivity systems stands out.

This sector is a great opportunity for the development of new logistics models and the management of last mile goods traffic, that is constantly growing. Light 2–3 wheels vehicles offer flexibility and advantages for goods delivery solutions compared to conventional electrified vehicles [3] and are a competitive solution in urban areas.

In this context, within the Next Generation EU program, the National Center for Sustainable Mobility (MOST), Light and Active Mobility Spoke, has the objective of designing an innovative, ergonomics and safe e-cargo bike, considering technical and infrastructural characteristics, and driver behaviors. The aim is to introduce a small fleet of e-cargo bikes for last mile delivery of mail and parcels in Italy.

The paper presents the results of the activity carried out in the first year of three by the “Vehicle & User” WP.

Keywords: sustainable mobility · light vehicles · active mobility · e-cargo bike

1 Introduction

Road transport, for mobility of people and goods, is responsible for 23% of greenhouse gas (GHG) emissions in Europe and of urban air pollution [4, 5]. The challenge of sustainable mobility requires the development of new paradigms, that must aim at

decongestion of the urban context through the accessibility for a wide range of users. This requires a design capacity oriented towards innovation and immediately usable and highly scalable solutions in different social contexts [2].

The sector of light vehicles and active mobility of people and goods represents a promising solution and is part of new and sustainable lifestyles and an alternative form of mobility. It aims at encouraging a new organization of everyday activities. From the point of view of vehicle design, it is necessary to create new design paradigms and a new generation of flexible and modular safe vehicles. From the point of view of the user, user-centred personalized mobility experience solutions, that include improving health, accessibility, and usability, need to be investigated. From a social and economic point of view, it is necessary to demonstrate the ability to safely use L-vehicles in an extended range of scenarios and extend the use range [2].

One of the challenges is the design of light 2–3 wheels vehicles for last mile delivery of goods, to increase safety and efficiency using new materials, sensors, and vehicle connectivity systems (V2X). Scientific research reveals how these can be a great opportunity for the development of new logistics models and for the management of last mile goods traffic, the flow of which is constantly growing [6]. Light 2–3 wheels vehicles offer flexibility and advantages for goods delivery solutions compared to conventional electrified vehicles [3], presenting themselves as a more competitive solution, for example for service in restricted or pedestrian areas of historic centers [7].

In this context, within the PNRR Next Generation EU program, the National Center for Sustainable Mobility [2], Light and Active Mobility Spoke (Spoke 5), has the objective of designing an innovative e-cargo bike, considering technical and functional characteristics, and driver behavior, to be introduced as a small fleet for last mile delivery of mail and parcels in Italy.

2 Research Objectives and Activities

The “Vehicle and User” WP of the project aims to: (i) explore innovative technological solutions to make a tailor-made cargo-bike and ensure sustainability throughout its life-cycle; (ii) broaden the target user base through the integration of customized solutions and the implementation of new features focused on users and their needs; (iii) design a modular vehicle to adapt to different contexts of use; (iv) increase safety, comfort and vehicle usability (effectiveness, efficiency and satisfaction) for a wide cluster of users. Specifically, this paper reports the results of the first phase of analysis, carried out according to the development of the design concepts, and focused on the identification and definition of ergonomic, technological, of comfort and usability requirements in user-vehicle interaction. The research activities described below have been conducted through the following tasks.

Preliminary Analysis and Definition of the Detailed Features of E-Cargo Bikes. A market analysis and classification by types of commercial setups and of e-cargo bikes (e-cargo bike models analyzed: > 40) were conducted to determine the main layouts and technical-formal characteristics of the vehicles.

Expert Evaluation of Ergonomics and Product Usability Aspects. This activity has been focused on the evaluation of e-cargo bike features for each category identified on the

first task. The main aspects explored are those that generate the driving user experience: i) postural aspects, vehicle frame layout and adaptability to the anthropometric user's dimension; ii) usability of the vehicle concerning modularity applied to the loading and unloading of goods, accessibility to the box and vehicle interaction with the driver and environment; iii) vehicle maintenance, and scalability in different contexts.

Definition of E-Cargo Bike Layout for the Specific Case Study Being Researched by MOST- Spoke 5. Based on the considerations of the technical analysis and classification, and on the specific stakeholder's needs has been individuated the most promising vehicle layout to the development of the e-cargo bike design for the last mile delivery of parcels and letters. The identified layout is the basis for the development of the project up to TRL 8–9.

3 Results

3.1 Identification and Classification of the E-Cargo Bike Typologies

The e-cargo bikes analyzed have been classified on specific technical-formal characteristics (number of wheels, load position and volume, drivability and riding posture, safety, visibility, etc.). Next, the 5Ws method “*Why, Who, Where, When, What*”, has been adopted to define correlation between vehicle's features with parameters that influence their selection offering a comprehensive overview of the many aspects involved in designing a cargo bike in relation to specific user and context needs. This activity therefore has provided a solid basis for the creation of useful tools for various actors involved in the design and selection of e-cargo bikes.

Specifically, the analyzed and collected data were used to design an interactive tool, aimed at designers or users, to guide them in evaluating, defining and choosing the most suitable vehicle layout, based on who is the user, why, where, when, and how the e-cargo bike is used (see Fig. 1). The purpose of the tool is to raise awareness among users and designers on issues relating to ergonomics, which are not well known in the field of e-cargo bikes, even though they are most often work tools [8].

The tool outlines the potential functions of a virtual assistance platform that, using the 5 W parameters, collects feedback directly from users to provide a customized selection of the best available e-cargo bike options. The software not only guides users in their choice of most suitable light vehicle, but is also conceived to guide designers in the definition of a new one, optimizing their choices based on specific physical, contextual, or usage needs, guiding them in defining practical and effective solutions to improve the accessibility of e-cargo bikes, enabling a wide range of users to identify the most suitable product for their mobility needs.

3.2 Design Opportunities and Constraints in E-Cargo Bike Vehicles

The design of e-cargo bikes must consider several key factors related to both human factors and the context of use to ensure an optimal user experience for a wide range of users and in relation to a specific intended use. The research has identified the constraints

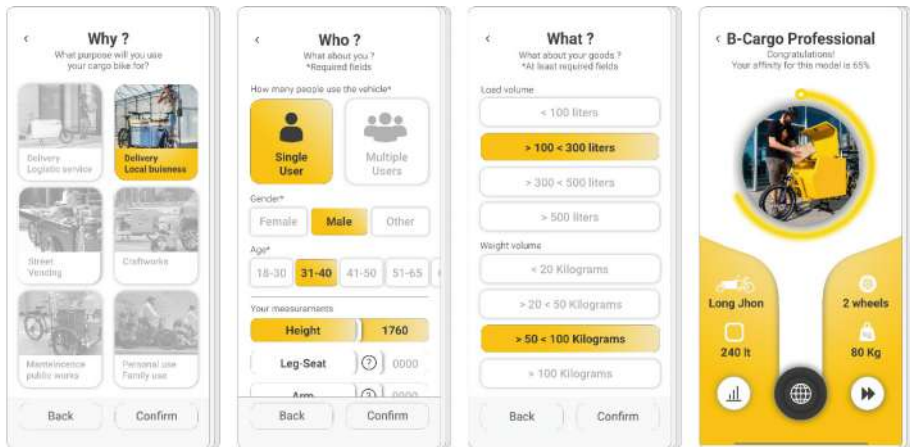


Fig. 1. Some screenshots of the tool for selecting and evaluating an e-cargo bike depending on the users and their needs.

and characteristics of the vehicle model that guide design choices based on the user's needs and the characteristics of the context of use, to improve the UX. Specifically, the following results emerged.

Driving Comfort and Seating Posture. Driving comfort is a fundamental factor that must be considered in the design of a bike. The concept of comfort finds its real meaning in the term “adaptation”. The “adapted” bike expresses the highest degree of riding comfort for the user. An adapted bicycle allows the rider to maintain an appropriate posture, increasing the quality of physical performance and preventing injury through the adjustment of its components [9].

- The decision between a sit-on or sit-in position should consider key features of the delivery journey, such as route length, number of stop-and-go instances, and other relevant factors (see Fig. 2)
- There is no universally “correct” posture. Cycling posture varies based on individual needs and context. The “dutch bike position” offers even weight distribution on the saddle, while the “racing bike position” focuses on aerodynamic efficiency by leaning forward onto the handlebars (see Fig. 2).
- To ensure comfort during riding for a wide range of users, adopting a unisex adjustable frame structure based on specific anthropometric measurements is crucial. User-friendly adjustment systems for the saddle-handlebar-crank triangle help prevent incorrect postures linked to physical discomfort, enhancing user awareness and knowledge, and reducing health risks [10].
- Maintaining a proper posture isn't enough to prevent discomfort during long rides. Riders often improvise postures to ease strained muscles. Designing easy posture change systems for different ride phases is crucial for addressing this issue [11, 12].

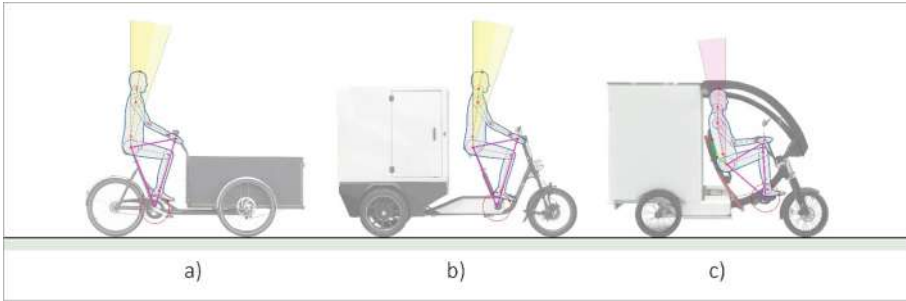


Fig. 2. Drive position and postural range variation: a) Sit-on position on front load (2, 3-wheels), b) Sit-on position and rear load (2, 3.4 wheels), c) Sit-in position and rear load (3, 4-wheels).

Modularity and Scalability. Modular product design is a modern approach that allows for the development of highly flexible devices that can adapt to completely different scenarios without affecting the user experience. In cargo bike cases, specifically, modularity is understood according to two distinct aspects:

- **Vehicle modularity:** is the design of a modular vehicle that allows the development of a horizontally and/or vertically differentiated product line based on the use of a limited number of interchangeable components. This approach allows companies to meet the needs of a business that, in the case of freight transport, is highly targeted and diversified in its needs.
- **Box-container modularity:** The concept of box modularity investigates two distinct aspects: i) logistics problems related to the management and movement of goods between different hubs [7]; ii) designing a box for specific use: the concept of modularity moves to the position's choice of the payload volume and to the “inside” organization aim to maximize the volume capacity in relation to the characteristics of the contents.

3.3 Design Concept of the E-Cargo Bike Case Study

The layout developed for the e-cargo bike case study is the synthesis of the exploration and generation of design concepts, evaluated and selected as possible viable ways identified by the results of the previous task, crossed with the user needs and constraints provided by the company on which the case study is based. The layout includes considerations relating to aspects of traditional ergonomics, anthropometry and biomechanics of posture and movement, and considerations relating to user-vehicle interaction, necessary to configure the vehicle for maximum driving comfort and to optimize actual and perceived usability and user experience.

To define light vehicle's design an approach focused on exploring modularity has been adopted as strategy to increase flexibility, adaptability and scalability of the vehicle. The study of modularity aspects has been divided into two distinct phases. The first phase has evaluated the possibility of developing the concept of a 2 in 1 vehicle to configure an e-bike and an e-cargo bike through the combination of several common modules.

On the second phase the research activity has been focused on developing specific solutions for the cargo box. In this regard the study of modularity moved “inside” the cargo box to maximize the volume capacity in relation to the characteristics of the contents, interpreting the ability to manage the cargo box according to the amount of cargo volume, also adapting it to the different stages of delivery. Specifically, in case study, the internal volume of the vehicle has been designed to enhance user interaction on delivery phases ideating some shelves and drawers which slide on rails placed at various heights and that be pulled out to reconfigure the box’s loading volume. (see Fig. 3)

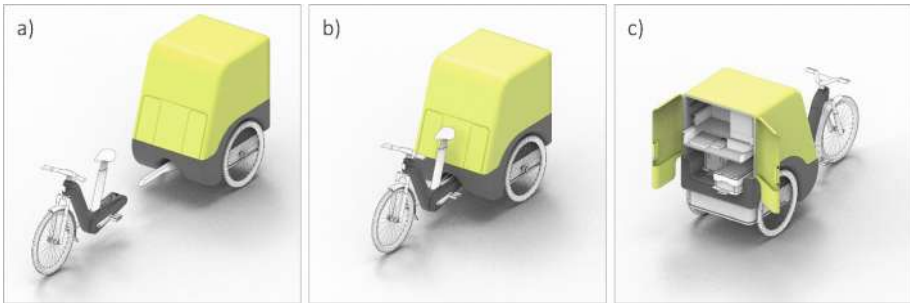


Fig. 3. Modular vehicle concept: a), b) Vehicle modularity, c) Box-container modularity.

4 Conclusion

The use of e-cargo bikes in the urban environment as a work tool, as well as for personal use, is rapidly expanding. To our knowledge, there are no in-depth studies in scientific literature regarding ergonomics, safety, driveability, and usability for a large cluster of users. Taking these aspects into account within the design process permits to optimise the user experience, improving comfort, safety and interaction with the vehicle.

In this regard, our project aims to give a contribution to these aspects, to provide guidelines for designers of new vehicles and to guide users in choosing vehicle best suited to their needs highlighting the crucial role of ergonomics and product usability aspects in the design of micro-mobility solutions.

The in-depth study of these issues is a fundamental step towards the design of light vehicles that meet the challenges of urban mobility, with the aim of promoting and enhancing the transition to more sustainable, accessible and inclusive modes of transport. During the first year the research led to the results described above, which provided the basic research for the next phase of product development for a specific case study. The next objectives are to create 2 prototypes to be tested, and then report all the feedback within guidelines and design scenarios, following an iterative process.

Acknowledgements. The results are part of the activity of the Italian National Center for Sustainable Mobility (MOST), Light and Active Mobility Spoke (Spoke 5), funded by: Next Generation EU program, Ministry of University and Research (MUR), Italia Domani Program - National Recovery and Resilience Plan (PNRR).

The article is the result of the joint work of the authors; the contribution of each author is the following: Alessandra Rinaldi: Research objectives and activities, Results (3.1, 3.2); Jonathan Lagrimino: Introduction, Results (3.3).

References

1. European Commission. <https://www.consilium.europa.eu/it/policies/clean-and-sustainable-mobility/> (2024). Accessed 21 Feb 2024
2. MOST, National (Italy) Center for Sustainable Mobility.: <https://www.centronazionalemost.it/> (2024). Accessed 16 May 2024
3. Oliveira, C.M.D., et al.: Sustainable vehicles-based alternatives in last mile distribution of urban freight transport: A systematic literature review. *Sustainability*. **9**(8), 1324 (2017)
4. European Commission: European Strategy for Low-Emission Mobility. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 2016, COM (2016). 501 Final
5. Narayanan, S., Antoniou, C.: Electric cargo cycles-A comprehensive review. *Transp. Policy*. **116**, 278–303 (2022)
6. Schomakers, E.M., et al.: Analysis of the potential of a new concept for urban last-mile delivery: Ducktrain. *Transp. Res. Interdiscip. Perspect.* **14**, 100579 (2022)
7. Bachofner, M., et al.: City logistics: Challenges and opportunities for technology providers. *J. Urban Mobil.* **2**, 100020 (2022)
8. IDEE Lab, <https://www.behance.net/labidee> (2024). Accessed 21 May 2024
9. Hsiao, S.W., et al.: Applying riding-posture optimization on bicycle frame design. *Appl. Ergon.* **51**, 69–79 (2015)
10. Bernate, J.A., et al.: Bike fitting para el rendimiento y la prevención de lesiones en ciclistas de ruta aficionados y profesionales. *Viref Rev. Educ. Fís.* **9**(1), 63–71 (2020)
11. Tosi, F.: *Ergonomia e Design, Design per l'Ergonomia*. Franco Angeli, Milan (2018)
12. Tosi, F., Belli, A., Rinaldi, A., Tucci, G.: The intermodal bike: Multi-modal integration of cycling mobility through product and process innovations in bicycle design. *Work*. **41**(Supplement 1), 1501–1506 (2012)



Competencies and Technological Support for Remote Operator of Automated Vehicles: A Literature Review and Perspectives

Sarah Schwindt-Drews¹(✉), Jan-Philipp Frenzel¹, Ole Hans², Bettina Abendroth¹,
and Jürgen Adamy²

¹ Institute of Ergonomics and Human Factors, Technical University of Darmstadt,
Otto-Berndt-Straße 2, 64289 Darmstadt, Germany
sarah_selina.schwindt@tu-darmstadt.de

² Department of Control Methods and Intelligent Systems, Technical University of Darmstadt,
Landgraf-Georg Straße 4, 64283 Darmstadt, Germany

Abstract. Automated vehicles at SAE Level 4 and 5 no longer demand an on-board driver, but may require the assistance of remote operators in certain situations. Consequently, it is imperative to examine the qualifications and competencies of remote operators to ensure effective performance. Based on a comprehensive literature review, this paper compares the conventional driving task with the tasks of re-mote operators and identifies relevant qualifications and competencies. Further-more, this paper identifies potential for technological support to assist the remote operator in the perception and assessment of information, as well as in the execution of actions. Since the results are solely derived from a literature analysis, they require further evaluation through subject tests.

Keywords: Remote Operator · Competencies · Technical Support · Remote Control · Automated Vehicles

1 Problem Statement and Objective

Automated vehicles at SAE levels 4 and 5 are no longer required to have an on-board driver [1]. Nevertheless, these vehicles may encounter circumstances that necessitate the intervention of a human operator. As observed by [2], humans are significantly better at rapidly comprehending and interpreting complex situations. In particular, they are able to anticipate the behavior of other road users with great accuracy. Furthermore, in many situations, automated vehicles are unable to perceive information that humans take for granted [3]. Therefore, remote operation represents an appropriate transitional technology towards autonomous driving. In addition, according to the German Law on Automated Driving [4], the operation of vehicles with automated driving functions on public roads demands the presence of a technical supervisor as the responsible individual. Consequently, a trained remote operator (RO) must be able to control the vehicle from a control center in the event of limitations or malfunctions [5].

Since the remote control of vehicles differs significantly from conventional driving it requires specific competencies. Therefore, the objective of this paper is to compare the conventional tasks of a driver and the task of a RO, identify the necessary competencies for a RO, and explore potential technological supports for their tasks.

2 Methodology

In order to achieve this objective, a systematic literature review was conducted. The TUFind and ScienceDirect databases were searched for papers addressing either the design of the workstation or the tasks of remote operators for automated vehicles. Following the screening of the title and abstract, 74 papers were identified as suitable for further analysis. Subsequently, a comparison was conducted between the conventional driving task and the remote operator driving task, based on the existing concepts described in the identified literature. Afterwards, competency requirements for remote operators were derived, taking into account the legal framework, economic efficiency, and consumer acceptance. Finally, based on user-centered requirements for the human-machine interface, the potential for technological support for remote operators is identified in order to ensure safe and efficient teleoperation.

3 Results

Conventional Driving Task vs. Remote Operator Driving Task

The conventional driving task primarily involves the processing of information, whereby data is continuously received and transformed into actions [6]. In accordance with the SAE [1], the driver is thereby responsible for performing a range of operational and tactical tasks, which depend on the automation level of the vehicle. In general, the conventional driver is responsible for controlling the vehicle's lateral and longitudinal motion through steering, acceleration, and deceleration [6]. Furthermore, the driver must detect, recognize, classify, prepare, and execute responses to objects and events, plan maneuvers, and increase conspicuity through lighting or signaling [6].

In contrast to conventional drivers, ROs are not continuously involved in the driving process of the automated vehicles. They do not observe the vehicle continuously but are only involved when it reaches its system limits or when errors occur [7]. Accordingly, they must rapidly form an appropriate situation awareness in order to decide whether to approve the proposed maneuver or to take direct or indirect control of the vehicle [7]. Once they have taken control, they perform the same tasks as conventional drivers. Moreover, the RO is responsible for communicating with individuals inside and outside the vehicle. This may include providing information to passengers or authorities, such as police, fire brigade or ambulance service [7], or requesting data about the environment also called Operational Design Domain (ODD) [8]. A systematic approach to include the RO Performance into the definition of an ODD in urban environment has been outlined by [8].

In particular, the sudden shift into the situation and the local distance to the vehicle present challenges to the remote operator's information intake and processing, especially for perception and recognition of important information. This can be explained by the

combined sequence and resource model, which includes the three phases of information intake, information processing in the narrower sense, and information output [9]. Each of these phases depends on the individual characteristics and attention level of the driver and determine the operation of the workstation and thus also indirectly the control of the vehicle. Figure 1 illustrates the interdependencies within the remote operator-workstation-vehicle-environment system according to [6].

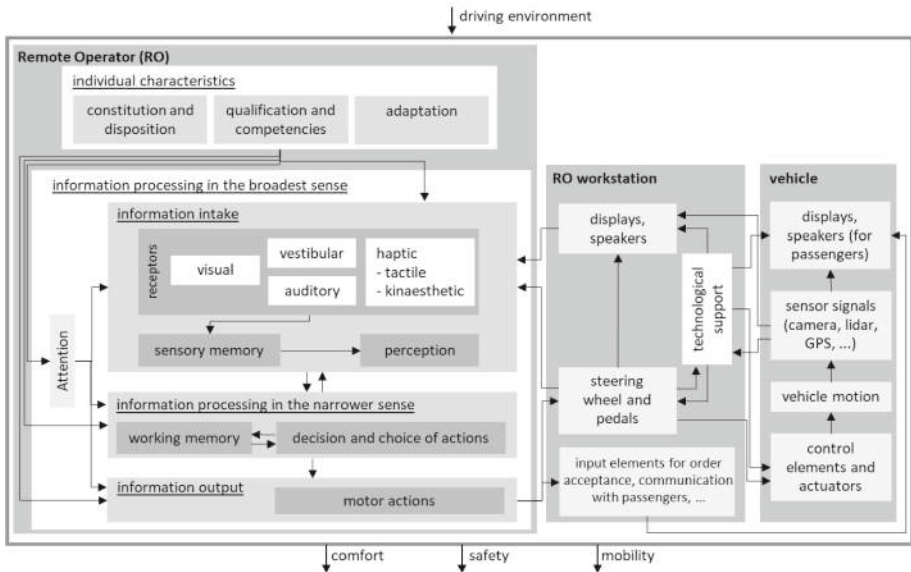


Fig. 1. Remote Operator-workstation-vehicle-environment system model (according to [6])

In conventional driving, 80 to 90% of traffic-relevant information is received visually, both from the environment and the vehicle's human-machine interface [6]. In particular, the speed at which the vehicle approaches an obstacle is calculated visually with the help of the viewing angle speed. The extent of visual perception is thereby constrained by both the field of view and the observation time. The remaining traffic relevant information is received via auditory, vestibular and haptic receptors [6]. In addition to the detection of acoustic warning signals, acoustic perception plays a role in the assessment of the distance of other vehicles [6]. Moreover, binaural hearing has an impact on spatial orientation. Vestibular and haptic perception are essential for perceiving vehicle speed and acceleration [6].

The majority of concepts for RO workstations (e.g [10, 11].) do not include motion platforms, which precludes any possibility of perceiving the dynamic situation via vestibular or haptic receptors. Furthermore, spatial orientation via acoustic signals is limited, as these are only transmitted via speakers or headphones. Consequently, the perception of traffic-relevant information is even more focused on the visual channel compared to conventional driving, which, according to the theory of multiple resources

[12], makes it even more difficult to detect the various signals and information. Moreover, the field of view of the RO is also severely restricted due to the display of visual information on the monitors of the RO workstation as well as the camera positioning on the vehicle. Furthermore, there are latencies and short-term failures in the signal transmission between vehicle and workstation [13]. Consequently, ROs must possess specific competencies in the field of information perception and assessment and should be provided with technological support in these areas.

Required Qualifications and Competencies of Remote Operators

As previously described, information processing is influenced by the individual characteristics of a person, which can be divided into the following categories: constitution and disposition, qualification and competencies, and adaptation [9] (see Fig. 1).

The constitutional and dispositional characteristics that determine suitability for the role of RO include, in particular, personality traits such as a willingness to take risks, emotional stability, and impulsiveness, which are associated with an increased risk of accidents [6]. Furthermore, ROs must not be susceptible to motion sickness.

The qualifications and competencies that ROs should possess include, in particular, sensorimotor skills and experience [6]. Due to the aforementioned limitations of the RO's tasks and workstation, ROs should have sensorimotor skills that enable them to perceive speed and spatial orientation, control selective attention and quickly develop situational awareness. In addition, ROs should gain sufficient experience through training before beginning their employment to improve not only their steering skills, but also their ability to identify and assess risks. Furthermore, experience allows for the development of operational routines, which means that actions require less attention, allow for high processing speed, and enable quick and flexible response to situational changes [6].

A relevant adaptation characteristic for ROs is task-related fatigue [9], which can be caused by both overload and underload. Due to the waiting time between requests for assistance from the vehicles [7] and the sudden shift into complex situations, ROs are at risk of being exposed to a constant change between overload and underload. As a consequence, reception, perception, and coordination disorders, as well as a decrease in attention, can occur, which may affect all stages of human information processing and the safety of the overall system. Consequently, ROs must be able to cope with sudden increases in load.

Potentials for Technological Support

Technological support offers great potential for optimizing communication between the vehicle and the RO via the RO workstation. It can support the perception and assessment of information as well as the execution of actions.

The driver's ability to perceive relevant information is largely determined by their level of attention [6]. As ROs primarily receive visual information, it is advisable to support important visual information with acoustic, vestibular or haptic signals. Additionally, a remote-controlled camera on the vehicle can enhance the driver's perception of the surrounding environment. To enhance situational awareness, telepresence should also be optimized. For instance, the use of head-mounted displays can improve situational awareness and reduce cognitive load [14]. However, the use of such devices could also increase the risk of motion sickness and physical strain due to their weight. In order

to intensify the feeling of presence, another modality should be continuously transmitted in addition to the visual modality [15]. For example, the reproduction of acoustic signals from the vehicle environment can improve the feeling of speed and the localization of other road users, especially if these are transmitted as 3D audio [2]. Furthermore, the transmission of vehicle movement via motion platforms can be employed to enhance telepresence and mitigate motion sickness. Force feedback steering wheels and pedals can also be utilized to optimize the sensation of presence [16].

To further assist the assessment of information augmented reality can provide visual support, such as projecting the vehicle's trajectory to improve situational awareness and compensate for latency impairments [17]. Moreover, an overhead view of the vehicle can be beneficial in assisting spatial orientation and the utilization of assistance systems, such as blind spot or distance warning systems. This is particularly the case when ROs operate vehicles with different dimensions without having an accurate understanding of them.

In order to assist the RO in the execution of actions, haptic shared control concepts (e.g. [18].) can be implemented, which, for example, support steering movements. Furthermore, passengers and pedestrians should be automatically and continuously informed about the vehicle status without the necessity of initiating communication from the RO. Additionally, requests from passengers or pedestrians should be withheld in critical situations to avoid overloading the RO.

4 Discussion and Conclusion

This paper compares the conventional driving task with the driving task of a remote operator and presents a comprehensive overview of the competencies required by the remote operator job profile. Furthermore, the paper describes how technology can assist remote operators in their tasks. However, since the results are solely derived from a literature analysis, they require further evaluation. Accordingly, subject tests should assess both the requisite competencies and the impact of technological support, individually and in combination.

Acknowledgments. This paper is the result of the joint research project STADT:up. The research project is supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK) based on a decision of the German Bundestag. The authors are solely responsible for the content of this publication.

References

1. SAE International (2021). Surface Vehicle Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_202104). https://www.sae.org/standards/content/j3016_202104/
2. Diermeyer, F., Gnatzig, S., Chucholowski, F., Tang, T., Lienkamp, M.: Der Mensch als Sensor - Der Weg zum teleoperierten Fahren. In: AAET - Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel, pp. 119–135. ITS Niedersachsen, Braunschweig (2011)

3. Beiker, S.: Unsettled Issues in Remote Operation for On-road Driving Automation. Silicon Valley Mobility. SAE International, Pennsylvania, USA (2021)
4. Gesetz zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes – Gesetz zum autonomen Fahren (2021) <https://dip.bundestag.de/vorgang/gesetz-zur-%C3%A4nderung-des-stra%C3%9Fenverkehrsgesetzes-und-des-pflichtversicherungsgesetzes-gesetz/273887>
5. Hans, O., Radlak, K., Adamy, J.: Human-factor focused application of STPA to remotely driven vehicles. In: 2024 IEEE World Forum on Public Safety Technology (WFPST), pp. 120–125. IEEE, Herndon, VA, USA (2024)
6. Abendroth, B., Bruder, R.: Capabilities of humans for vehicle guidance. In: Winner, H., Hakuli, S., Lotz, F., Singer, C. (eds.) Handbook of Driver Assistance Systems. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-09840-1_1-1
7. Schwindt, S., Heller, A., Theobald, N., Abendroth, B.: Who will drive automated vehicles? - usability context analysis and design guidelines for future control centers for automated vehicle traffic. In: Human Factors in Transportation. 14th International Conference on Applied Human Factors and Ergonomics, vol. 95, pp. 99–109. Fenerbahçe University, Istanbul, Türkiye (2023)
8. Hans, O., Avezum, M., Borysov, S., Ross, H., Adamy, J.: Operational design domain qualification framework for remotely driven vehicles in urban environment. In: 2023 IEEE International Automated Vehicle Validation Conference (IAVVC), pp. 1–6. IEEE, Austin, TX, USA (2023)
9. Schlick, C., Bruder, R., Luczak, H.: Arbeitswissenschaft. Springer-Verlag, Berlin and Heidelberg (2018)
10. Schrank, A., Walocha, F., Brandenburg, S., Oehl, M.: Human-centered design and evaluation of a workplace for the remote assistance of highly automated vehicles. Cogn Tech Work. **26**, 183–206 (2024)
11. Kettwich, C., Schrank, A., Oehl, M.: Teleoperation of highly automated vehicles in public transport: User-centered design of a human-machine interface for remote-operation and its expert usability evaluation. MTI. **5**(5), 26 (2021)
12. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance, 3rd edn. Prentice-Hall, Upper Saddle River, NJ (2000)
13. Neumeier, S., Wintersberger, P., Frison, A., Becher, A., Facchi, C., Riener, A.: Teleoperation: The holy grail to solve problems of automated driving? Sure, but latency matters. In: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. Association for Computing Machinery, pp. 186–197. Association for Computing Machinery, New York NY United States (2019)
14. Georg, J., Feiler, J., Diermeyer, F., Lienkamp, M.: Teleoperated driving, a key technology for automated driving? comparison of actual test drives with a head mounted display and conventional monitors. In: 2018 21st International Conference on Intelligent Transportation Systems, pp. 3403–3408. IEEE, Maui, HI, USA (2018)
15. Chen, J., Haas, E., Barnes, M.: Human performance issues and user interface design for teleoperated robots. IEEE Trans, Syst, Man Cybern. **37**(6), 1231–1245 (2007)
16. Neumeier, S. (2023). Enabling Teleoperated Driving in Everyday's Traffic Scenarios [, Technische Universität München]. BibTeX. <https://mediatum.ub.tum.de/1695113>
17. Dybvik, H., Løland, M., Gerstenberg, A., Slåttsveen, K.B., Steinert, M.: A low-cost predictive display for teleoperation: Investigating effects on human performance and workload. Int. J. Hum.-Comput. Stud. **145**, 102536 (2021)
18. Marcano, M., Diaz, S., Perez, J., Irigoyen, E.: A review of shared control for automated vehicles: Theory and applications. IEEE Trans. Human-Mach. Syst. **50**(6), 475–491 (2020)



An Ergonomics Study on the Risks and Difficulties Faced by Drivers of Oil Tank Trucks in the Sahara Desert

Mourad Semmani¹ , Bouhafs Mebarki² , and Mohamed Mokdad³ 

¹ University of Ahmed Draia Adrar, Adrar, Algeria
mouradopsy@gmail.com

² Laboratory of Ergonomics, University of Oran 2 Oran, Oran, Algeria

³ College of Arts, University of Bahrain Zallaq, Zallaq, Bahrain

Abstract. Truck driving is among the most dangerous businesses in the world. Truck drivers are the people who transport goods from one place to another all over the world. These drivers face many risks and accidents that may lead to their death.

Among truck drivers, we refer in this paper to oil tanker drivers carrying oil and oil byproducts. Petroleum tanker drivers are exposed to the dangers related to the hazardous materials they carry. This study aims to shed light on the various risks the petroleum truck drivers are exposed to in the Algerian desert.

In completing this study, the researchers relied on the descriptive survey method, as it is the most appropriate method for such topics. Data was collected from members of the research sample (35 drivers) using a questionnaire prepared specifically for this purpose.

The average age of the sample members was 47 years (SD: 6.05). All drivers were married (100%). Their average professional experience was (14.3) with a SD of (4.22). The results revealed that truck drivers are exposed to five main risk categories: Health Risks, Physical Risks, Chemical Risks, Mental Load, and Organizational Risks.

Regarding the risks the drivers are exposed to, they are varied and many.

Although Truck driving work is sometimes, considered as indoors work, done in a confined space, which is the cabin. However, in reality it is an outdoor work in remote areas. To treat risk factors, which affect truck drivers' health, a continuous ergonomics improvement process should be implemented.

Keywords: Oil Truck Drivers · Algerian Desert · Transportation Safety · Risks

1 Introduction

Despite continuous technological progress and widespread awareness of best safety practices, not all work is safe for human life. Every job has its own risks. Many professions and jobs are risky, threatening workers with health problems and serious injuries, which may sometimes cost them their lives.

Truck driving is among the most dangerous businesses in the world. Truck drivers are the people who transport goods from one place to another all over the world. These drivers face many risks and accidents that may lead to their death. The reason for this is that they perform this job for long working hours a day, which exposes them to fatigue. This may lead to deadly accidents. They may also have heavy goods with them that affect their control of the truck, causing accidents [1].

The literature on risk factors which are related to occupational health problems of truck drivers, mention the prolonged sitting, the whole-body vibration, physical and psychological fatigue and related health problems such as work musculoskeletal disorders [2–4].

Among truck drivers, we refer to oil tanker drivers carrying oil and oil byproducts. In addition to the risks to which conventional truck drivers are exposed, petroleum tanker drivers are exposed to the dangers related to the hazardous materials they carry [5]. Thus, oil tanker truck disasters are well documented [6]. Results of such disasters are especially catastrophic in low- and low-middle-income countries, where, almost all deaths (94%) occurred (Ewbank, et al. 2019).

Driving oil tankers in the Sahara Desert conditions, is even more risky then, driving on conventional roads and highways, for reasons like driving long distance on unpaved roads, in extreme physical conditions (dust, high summer temperatures and low winter temperatures). Deprivation/lack of the necessary technical assistance and road servicing in case of material failing truck breakdowns, and accidents, particularly between oil fields to the main road connections. Beside the high-risk factors related to truckers work, work environment have a great impact on the ability of the driver to properly carryout the job, and compromises his performance [8].

1.1 Objectives of This Study

This study aims to shed light on the various risks to which petroleum truck drivers are exposed to in the Algerian desert. This study also seeks to answer the following questions:

1. What are the demographic variables of drivers? (Age, educational level, social status, professional experience).
2. What risks are oil truck drivers exposed to?
3. How to manage these risks?

2 Methodology

In completing this study, the researchers relied on the descriptive survey method, as it is the most appropriate method for such topics.

The researchers also relied on a sample of 35 drivers who start their work from the land transport station in the state of Adrar in southwestern Algeria. From this station, drivers move their trucks to other states, especially considering that it is the center for exporting oil abroad. Long distances, 1500 to 2000 km are traveled each round, mostly, on a rough road network and in poor road conditions (sandstorm, high temperature during hot season and cold temperature during winter, etc....).

Data was collected from members of the research sample using a questionnaire. The questionnaire consists of three parts:

- Questionnaire title and instructions.
- Demographic data include age, social status, and professional experience.
- The work-related phrases (05) revolve around the risks that oil truck drivers are exposed to at work, and risks management.

3 Results

3.1 What are the Demographic Variables of Drivers?

(Age, social status, professional experience). It was found that the average age of drivers was (47) years, with a standard deviation of (6.05). As regard to marital status of truck drivers, it was found that all drivers are married and support entire families. It was also found that drivers have an average professional experience of (14.3) with a standard deviation of (4.22).

3.2 What Risks are Oil Truck Drivers Exposed to in Sahara Desert Algeria?

It has been shown that truck drivers are exposed to the following risks:

- **Health Risks:** Truck Drivers complained about a number of elements that threaten health, including work accidents, occupational diseases, sitting position, and a lot of static work.
- **Mental Load:** They also complained about the amount of mental effort driving a truck usually caused by permanent information collection, constant vigilance, and psychological constraints (delivery times, repair of truck breakdowns far from any technical assistance, etc.).
- **Physical Risks:** Drivers also pointed out that the roads they cross, although they are not all paved, are in need of constant maintenance. Drivers agreed that working in the extremely cold winter and extremely hot summer, vehicle vibration, noise, dust increases their suffering.
- **Chemical Risks:** They also complained against chemicals at work such as exhaust gas, atmospheric pollution, petrol transported.
- **Organizational Risks:** Truck drivers complained against organizational risks such as night/day work, irregular work hours, extended work hours, irregular meal intake, sleeping in the truck, etc., working alone while driving.

3.3 How to Manage These Risks?

To overcome some of these risks, oil truck drivers organized themselves into work groups of 3 to 5 truckers, undertaking the journey from the oil filling point to the distribution point, in a fleet of 3 to 5 trucks. These work organization practices are essential for the survival of the group members in the harsh conditions of the desert, where cooperative work organization and mutual aid is prevailing on all other forms of work organization.

During the journey, the truck fleet has its well-known rest stops, generally near water points (Oasis and wells), or at service stations (if available). The rest stops are places

where truckers prepare and eat their collective hot meals/ beverage (Fig. 1 and Fig. 2), stock up on water, do some self-cleaning and truck minor maintenance, and sleep for irregular periods (depending on the time schedule of the trip).



Fig. 1. A stop for rest.



Fig. 2. Truck drivers share a meal together.

In cases of truck breakdowns or accidents (Fig. 3 and Fig. 4), all the fleet has to stop and participate in breakdown repair operations, where mutual aid is well-appreciated (experience of seniors, force and energy of young truckers).



Fig. 3. Aftermath of oil tanker accident.



Fig. 4. Truck break down repair work [9].

4 Discussion

Taking into account the results of the demographic variables, obtained above, it is clear that the average age of truck drivers is 47 years, they are all married and support their families and children. Their average professional experience exceeds (14) years, which are important for the benefit of those who do this work because this work requires a lot of patience, social support and physical fitness, flexibility and constant attention. Age, marital status and length of professional experience of the driver is an indicator of the professional health, as older drivers are the most marked age category by the performance degradation factors, like functional decline in sensory, cognitive and physical abilities related to the safe operation of a vehicle on the road, as stipulated by Waller [10]. On the other hand, some researchers [11] argue that older truck drivers have the experience

and skill to be safe on the road, higher seat belt use rates compared to other age groups. These compensatory strategies act as protective factors in the health and safety of the older workforce.

Taking into consideration the many risks of this profession (health, physical, chemical, mental and organizational Risks), the sample members of the present study are well affected by the disadvantages of the profession (occupational diseases, WMSDs due to sitting, static work and whole body vibration, atmospheric pollution, etc.). On the other hand, work teams of our sample are well age balanced, as the drivers of each truck fleet is generally a mix of old, medium and young age. Such work groups are experience and energy balanced work teams organization, which boosts the level of satisfaction and performance of the team members, as has been advocated by Gander et al. [12].

Regarding the risks to which drivers are exposed, they are varied and many because, from one perspective, it seems that it is a work that takes place indoors, but in reality it is outdoor work in remote areas. On the other hand, it is a work done in a confined space, which is the cabin.

Effective management to confront the risks to which truck drivers transporting petroleum are exposed includes working in three areas:

First, the design of the truck cabin as a multifunction space (working and sleeping space). Thus, the inclusion the right ergonomics design criteria is of most importance priority.

Secondly, road paving and maintaining should be at the forefront. Roads in the desert are the biggest obstacle faced by truck drivers. Driving on unpaved Sandy roads will add new risks to working in this profession. The appropriate design of tanks that transport petroleum in such extreme conditions is also very important.

Third, the individual driver. Emphasis is placed on training, especially continuous truck driving training and training on solution of frequent mechanical truck failure incidents. In addition, physical exercise, health monitoring, and psychological well-being are necessary as suggested by Benstowe [13].

5 Conclusions

Oil delivery truck drivers have received little attention from researchers, despite the significance of their work of moving oil from the production fields to the delivery stations. Because of the continued dangers of their work, ergonomics needs to pay greater attention. Bringing attention to the work conditions of this segment of the work force is a first step towards improving their professional life and the competitiveness of oil distribution sector.

References

1. Toscano, G.: Dangerous jobs. *Compens. Work. Cond.* **2**(2), 57–60 (1997)
2. Nazerian, R., Korhan, O., Shakeri, E.: Work-related musculoskeletal discomfort among heavy truck drivers. *Int. J. Occup. Saf. Ergon.* **26**(2), 233–244 (2020)
3. Peters, S.E., et al.: Working conditions influencing drivers' safety and well-being in the transportation industry: "On board" program. *Int. J. Environ. Res. Public Health.* **18**(19), 10173 (2021)

4. Pritchard, E., et al.: Impact of work and coping factors on mental health: Australian truck drivers' perspective. *BMC Public Health*. **23**(1), 1090 (2023)
5. Kaka, B., Hassan, A., Fatoye, F., Sooknunan, S.M.: Prevalence and pattern of work related musculoskeletal disorders among petroleum tanker drivers in kano. *Physiotherapy*. **107**, e183–e184 (2020)
6. Van Kooij, E., et al.: Responding to major burn disasters in resource-limited settings: Lessons learned from an oil tanker explosion in Nakuru, Kenya. *J. Trauma Acute Care Surg*. **71**(3), 573–576 (2011)
7. Ewbank, C., Gupta, S., Stewart, B.T., Kushner, A.L., Charles, A.: A systematic review of oil tanker truck disasters: Identifying prevention targets. *Burns*. **45**(4), 905–913 (2019)
8. Al-Mekhlafi, A.B.A., Isha, A.S.N., Al-Tahitah, A.N., Kineber, A.F., Al-Dhawi, B.N.S., Ajmal, M.: Modelling the impact of driver work environment on driving performance among oil and gas heavy vehicles: SEM-PLS. *Safety*. **9**(3), 48 (2023)
9. Semmani, M.: Pictures of fuel truck drivers, taken while on the road. March 2024. (2024)
10. Waller, P.F.: The older driver. *Hum. Factors*. **33**(5), 499–505 (1991)
11. Farrow, A., Reynolds, F.: Health and safety of the older worker. *Occup. Med*. **62**(1), 4–11 (2012)
12. Gander, F., Gaitzsch, I., Ruch, W.: The relationships of team role-and character strengths-balance with individual and team-level satisfaction and performance. *Front. Psychol*. **11**, 566222 (2020)
13. Benstowe, S. J.: Long driving hours and health of truck drivers. (2008)



Identifying Effective Visual Searching for the Recognition of Extraordinary Events during the Acceleration Process of Railway Driving

Daisuke Suzuki¹ , Takaharu Koike², and Ryo Kawai³

¹ Railway Technical Research Institute, 2-8-38 Hikari-cho Kokubunji-shi, Tokyo, Japan
suzuki.daisuke.55@rtri.or.jp

² Emovis Corporation, 3-20-12-702 Matsugaya Taito-ku, Tokyo, Japan

³ Hokkaido Railway Company, 1-1 Kita11, Nishi15, Chuo-ku, Sapporo, Japan

Abstract. This study aimed to quantitatively investigate the influence of effective visual searching behaviors on recognizing extraordinary events during the acceleration process of railway driving. Simulator training data was analyzed, including eye movements data from 128 actual drivers. The given driving scenario was to recognize a major extraordinary event (a subsidence of a railway track) after dealing with four minor events. Participants who braked before passing the subsidence were identified as part of the recognizing group, and those who did not brake until after passing the subsidence were identified as part of the nonrecognizing group. Logistic regression analysis revealed that the possibility of recognizing subsidence increased by 1.41 times when the percentage of gazing at the landscape increased by 1%, and the possibility increased by 1.21 times when the percentage of gazing at the railway track increased by 1%. The percentage of gazing at the landscape had the largest influence. The regression model with these variables can classify the participant group, with the percentage of correct classifications being 86.7%.

Keywords: Eye Movement · Railway Driver · Logistic Regression Analysis

1 Introduction

Railway operational companies often conduct vocational training sessions using a driving simulator [1]. To address extraordinary events appropriately, it is crucial to properly recognize the situation in front of a railway vehicle. Thus, identifying effective visual searching for the recognition of extraordinary events is important for enhancing railway safety.

The authors [2] previously examined an effective visual searching process during coasting at high velocity, approximately 90 km/h. Furthermore, a previous study [3] investigated an effective visual searching process while accelerating at a low velocity, approximately 30 km/h, showing that drivers who could recognize extraordinary

events gazed widely with short gazes. Another previous study [4] showed that overlooking extraordinary events could be attributed to excessively concentrating on gazing straight forward. Therefore, using the same data of previous studies [3, 4], this study aimed to quantitatively identify the influence of effective visual searching on recognizing extraordinary events during the acceleration process at low velocity, approximately 30 km/h.

2 Methods [3, 4]

2.1 Participants and Driving Scenario

We analyzed the simulator training data from actual drivers. Figure 1 shows the railway-driving simulator (Mitsubishi Precision Company, Limited) for vocational training. An eye tracker equipped with the simulator was used to measure the eye movement of the drivers. Calibration was conducted before measurements were obtained. The sampling rate was 30 Hz. The front view appeared on a 42-inch display with a width of 930 mm and a height of 520 mm. The distance between the display and the driver's eye point was 900 mm. The visual angle of the display was 54.7° and 32.2° .



Fig. 1. Railway-driving simulator for vocational training



Fig. 2. Appearance of the subsidence

In total, 128 professional drivers from a railway operational company participated in our study as a part of their annual vocational training sessions; the ages of the participants ranged from 24 to 65 years (mean 41.1 years, standard deviation 12.3 years). Driving experience ranged from 0 to 31 years (mean 14.3 years, standard deviation 7.5 years). We explained about the purpose of this study and how to use the data before using the vocational training data.

In the given driving scenario, participants were tasked with recognizing a major extraordinary event (subsidence of a railway track; Fig. 2) after dealing with four minor events to produce a situation that was difficult for the participants to recognize the major extraordinary event. The four minor events were (1) a stop signal just before a station, (2) a signal substituted by a station's staff, (3) railway workers present beside a railway track, and (4) an instruction to drive slowly (Fig. 3).

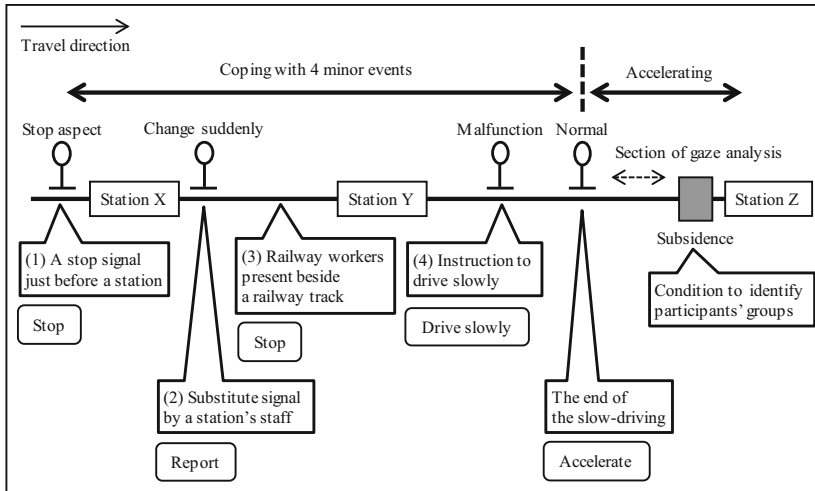


Fig. 3. Outline of the driving scenario

After dealing with the four minor events, the participants accelerated the simulated train from a driving velocity of 15 km/h, passing through the location of the subsidence at 30 km/h. The participants were asked to brake just after recognizing the subsidence. The driving course was an imaginary course. Participants who braked the simulated train before passing the subsidence were identified as part of the recognizing group (RG), whereas those who did not brake before passing the subsidence were identified as part of the nonrecognizing group (NG).

2.2 Analysis Item

Gaze analysis began approximately 25 s before passing through the subsidence, where the participants had not yet recognized it. Eye movement data were classified as a “gaze” if four or more subsequent frames (0.133 s) were observed with the same point of regard. The gaze was extracted using Sight Tracker Editor software (Emovis Corporation). The duration and location of gaze were measured based on the gaze extracted. The gaze location refers to the horizontal angle of gaze. Zero degrees indicate the center of the display, the right side shows positive values, and the left side shows negative values. The gaze range represents the standard deviation of gaze location.

In addition, we divided the front scenery at which the participants gazed into five parts based on a previous study [5] in the railway field, analyzing gaze targets. The five parts included straight forward, landscape, railway track, signal, and others (Fig. 4). We calculated the ratio of gaze duration on each gaze target against the total gaze duration of each participant, which is referred to as the gaze distribution ratio.

We defined the ratio of detected gaze duration to the total duration a participant gazed at the display in front of them as the “detected ratio.”, selecting participants with a detected ratio of greater than 60% for analysis of eye movement data. Of the 128 participants, 59 were chosen for eye movement analysis.

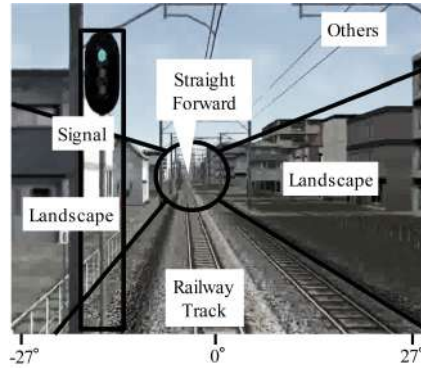


Fig. 4. Representation of gaze targets

Regarding statistical analysis, to choose candidates of the explanatory variables for the logistic regression analysis, correlation ratios were first calculated between the participant group and gaze duration, gaze location, gaze range, gaze targets, participants' age, and driving experience. Second, a correlation analysis was performed to clarify the correlation between the explanatory variable candidates, selecting the explanatory variables for the logistic regression analysis based on the results of the correlation analysis. Finally, a logistic regression analysis was performed to identify the quantitative influence of visual behaviors on the recognition of extraordinary events. The objective variable was the participant group.

To verify the accuracy of the logistic regression analysis, the data were divided into two groups: data for learning and data for testing. A regression model was identified using the data for learning, and the accuracy of the model was verified using the data for testing that were not used for identifying a regression model.

3 Results

3.1 Classification of Participant Group

The number of participants who recognized the subsidence (RG: recognizing group) was 85 (66.4%), and the number who did not (NG: nonrecognizing group) was 43 (33.6%). Among the 59 participants whose gaze was analyzed, 39 (66.1%) were in the RG, and 20 (33.9%) were in the NG. Based on these results, the data of 10 participants in the RG and 5 in the NG who were randomly selected were used for testing. The data of the other participants, 29 in the RG and 15 in the NG were used in logistic regression analysis to identify a regression model.

3.2 Results of the Calculation of Correlation Ratios

Table 1 shows the results of the calculation of the correlation ratios. Correlation ratios indicated that the participant group was significantly correlated with gaze range, percentage of gazing straight forward, percentage of gazing at the landscape, percentage of

gazing at the railway track, age, and driving experience. These variables were selected as candidates of the explanatory variables for logistic regression analysis.

Table 1. Results of correlation ratios between participant group and gaze behaviors.

	NG	RG	Correlation ratio	<i>F</i> value	<i>p</i> value
	Mean (SD)	Mean (SD)			
1. Mean of gaze duration (s)	0.70 (0.40)	0.77 (0.64)	0.00	0.17	0.69
2. SD of gaze duration (s)	0.71 (0.46)	0.88 (0.98)	0.01	0.39	0.54
3. Mean of gaze location (°)	−2.69 (1.77)	−3.39 (1.72)	0.04	1.59	0.21
4. SD of gaze location (°) (gaze range)	3.5 (1.56)	4.91 (1.90)	0.12	5.70	0.02*
5. Percentage of gazing straight forward (%)	85.23 (8.33)	59.27 (12.67)	0.55	51.21	0.00*
6. Percentage of gazing at the landscape (%)	6.90 (7.08)	22.65 (8.67)	0.47	36.73	0.00*
7. Percentage of gazing at the railway track (%)	6.56 (5.35)	16.52 (11.71)	0.19	9.71	0.00*
8. Percentage of gazing at the signals (%)	0.29 (1.11)	0.52 (1.32)	0.01	0.33	0.57
9. Percentage of gazing at others (%)	1.03 (2.81)	1.04 (2.45)	0.00	0.00	0.98
10. Age (years)	46.80 (13.15)	37.14 (11.11)	0.14	6.59	0.01*
11. Driving experience (years)	17.47 (6.88)	11.97 (7.09)	0.13	6.07	0.02*

3.3 Results of the Correlation Analysis

Table 2 shows the results of the correlation analysis with the candidates of the explanatory variables. The numbers in the column headings correspond to the same row numbers.

Considering each correlation, we determined the explanatory variables for the logistic regression analysis. The percentage of gazing straight forward was not selected because it was correlated with the percentage of gazing at the landscape and railway track. Furthermore, driving experience was not chosen as an explanatory variable because it was significantly correlated with age. Thus, the explanatory variables for the logistic regression analysis were gaze range, percentage of gazing at the landscape, percentage of gazing at the railway track, and age.

3.4 Results of the Logistic Regression Analysis

Table 3 shows the results of the logistic regression analysis. We calculated the variance-information factors (VIFs) to evaluate multicollinearity. There is a possibility of multicollinearity if the maximum VIF is greater than 10 or if the average VIF is considerably

Table 2. Results of the correlation analysis.

	1.	2.	3.	4.	5.	6.
1. Gaze range	1.00	−0.46*	0.54*	0.14	−0.24	−0.22
2. Percentage of gazing straight forward		1.00	−0.74*	−0.72*	0.30	0.27
3. Percentage of gazing at the landscape			1.00	0.10	−0.32*	−0.32*
4. Percentage of gazing at the railway track				1.00	−0.14	−0.09
5. Age					1.00	0.92*
6. Driving experience						1.00

Numbers in the column headings correspond to the same numbers in the rows.

greater than 1 [6]. In this study, the maximum VIF among the explanatory variables was 1.48. We determined that there was no problem of multicollinearity.

Table 3. Results of the logistic regression analysis.

Explanatory variable	Partial regression coefficient	Standardized partial regression coefficient	Odds ratio	<i>p</i> value
Gaze range	−0.27	−0.51	0.76	0.62
Percentage of gazing at the landscape	0.34	3.77	1.41	0.01*
Percentage of gazing at the railway track	0.19	2.07	1.21	0.01*
Age	−0.08	−0.97	0.93	0.17
Constant	−1.88		0.15	0.57

Note: *Significant at the 5% level

The results of the logistic regression analysis revealed that the recognition of subsidence was significantly associated with the percentage of gazing at the landscape and the percentage of gazing at the railway track. The following findings were obtained from the odds ratio: With all other variables being constant, the possibility of recognizing subsidence increased by 1.41 times when the percentage of gazing at the landscape increased by 1%. The possibility increased by 1.21 times when the percentage of gazing at the railway track increased by 1%. The standardized partial regression coefficient showed

that, among the four variables, the percentage of gazing at the landscape had the largest influence.

3.5 Results of the Verification of Regression Model

Prediction values were calculated using the regression model (Table 3) with the data of 10 participants in the RG and 5 in the NG for testing. A prediction value of more than 0.5 indicated the RG and a prediction value of <0.5 indicated the NG. Table 4 shows the percentages of correct classifications with the data for testing. The percentage of correct classifications was 80.0% in the NG, 90.0% in the RG, and 86.7% in total. The validity of the regression model was verified.

Table 4. Percentages of correct classifications with data for testing.

		Participant group		Total
		NG	RG	
Prediction	NG	4	1	
	RG	1	9	
Correct classifications (%)		80.0	90.0	86.7

Note: Value for classification: 0.50

4 Discussion

4.1 Correlation Ratio and Correlation Analysis

The results of the correlation ratios revealed a positive correlation between the participant group and gaze range. The results indicated that a broader gaze range contributed to the recognition of the subsidence. Our results support those of previous studies [3, 7, 8], which indicated that experienced drivers and railway drivers who could recognize extraordinary events had a wider spread of visual searching in the horizontal axis than less-experienced drivers did.

In addition, as the results of the correlation analysis, the percentage of gazing straight forward was negatively correlated with the percentage of gazing at the landscape and railway track. The area of gazing straight forward meant gazing at the railway track far away. The area of gazing at the railway track meant gazing at the railway track near the participant's vehicle. The results indicated that the participants who spent too much time concentrating in the straight forward direction could not distribute their attention to the wide range of gaze targets. Those drivers decreased their intention to allocate their attention properly.

4.2 Logistic Regression Analysis

The results of the logistic regression analysis revealed that drivers who recognized the subsidence paid attention to the landscape and railway track as well as the straightforward direction. These results backed up the findings of previous studies [9, 10], which indicated that experienced drivers scanned a variety of potential hazards. While accelerating from low velocity after dealing with minor events, railway drivers must pay attention to various wide-gaze targets. However, railway driving requires drivers to gaze at signals and level crossing. The highest mean percentage of gazing by the RG was for straight forward, which was 59.27%. Importantly, railway drivers not only to gaze straight forward but also to pay attention to various targets.

Considering the results of the correlation analysis and logistic regression analysis, excessive concentration on gazing straight forward due to acceleration manipulation can cause drivers to overlook an extraordinary event around the railway track. These results are consistent with the findings of previous studies [4, 11], which indicated that novices tended to spend a lot of time looking at irrelevant regions.

4.3 Comparison of the Way of Gazing at High Velocity

The present study and previous studies [3, 4], which investigated the participant's method of gazing at low velocity (approximately 30 km/h), indicated that gazing straight forward too much was one of the causes of overlooking an extraordinary event. On the contrary, the previous study by the authors [2], which examined the way of gazing at high velocity (approximately 90 km/h), revealed that gazing straight forward can contribute to the recognition of an extraordinary event, whereas railway drivers who gazed at a variety of targets tended to overlook an extraordinary event. According to these results, effective visual searching varies based on velocity. It is important that railway drivers use a visual searching process in which they gaze at a variety of targets largely near their own vehicles and gaze straight forward far away in accordance with velocity. Future studies should investigate the proper use of visual searching and the most effective visual searching method at middle velocity.

5 Conclusion

This study quantitatively identified the influence of effective visual searching on the recognition of extraordinary events at low velocity (approximately 30 km/h) during the acceleration process of railway driving. Logistic regression analysis revealed that, with all other variables being constant, the possibility of recognizing subsidence increased by 1.41 times when the percentage of gazing at the landscape increased by 1%, and the possibility increased by 1.21 times when the percentage of gazing at the railway track increased by 1%. The percentage of gazing at the landscape had the largest influence among the four variables. The regression model with these variables can classify the participant group, with the percentage of correct classifications being 86.7%.

References

1. Endoh, H., Omino, K.: Practical system for implementing vocational training program for improving train driver skills for coping with abnormal situations. *Q. Rep. RTRI*. **54**, 237–242 (2013)
2. Suzuki, D., Yamauchi, K., Matsuura, S.: Identifying the effects of visual searching by railway drivers upon the recognition of extraordinary events. In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (eds.) *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, vol. VI, pp. 165–173. Springer, Cham (2019)
3. Suzuki, D., Yamauchi, K., Matsuura, S.: Effective visual behavior by railway drivers upon the recognition of extraordinary events. *Q. Rep. RTRI*. **60**, 286–291 (2019)
4. Suzuki, D., Koike, T., Kawai, R.: Visual searching of railway drivers for recognizing extraordinary events -Consideration of acceleration process from low velocity driving. *Jpn. J. Ergon.* **60**(4), 221–230 (2024) (In Japanese)
5. Luke, T., Brook-Carter, N., Parkes, A.M., Grimes, E., Mills, A.: An investigation of train driver visual strategies. *Cogn. Technol. Work*. **8**, 15–29 (2006)
6. Netter, J., Wasserman, W., Kutner, M.H.: *Multicollinearity, Influential Observations, and Other Topics in Regression Analysis-II, Applied Linear Regression Models*, pp. 377–416. Richard D Irwin Inc., Homewood (1983)
7. Underwood, G., Chapman, P., Bowden, K., Crundall, D.: Visual search while driving: Skill and awareness during inspection of the scene. *Transp. Res. F*. **5**, 87–97 (2002)
8. Konstantopoulos, P., Chapman, P., Crundall, D.: Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accid. Anal. Prev.* **42**, 827–834 (2010)
9. Garay-Vega, L., Fisher, D.L., Pollatsek, A.: Hazard anticipation of novice and experienced drivers: Empirical evaluation on a driving simulator in daytime and nighttime conditions. *Transp. Res. Rec.* **2009**, 1–7 (2007)
10. Borowsky, A., Shinar, D., Oron-Gilad, T.: Age, skill, and hazard perception in driving. *Accid. Anal. Prev.* **42**, 1240–1249 (2010)
11. Vansteenkiste, P., Lenoir, M., Bourgois, J.G.: Gaze behaviour of experienced and novice beach lifeguards – An exploratory in situ study. *Appl. Cogn. Psychol.* **35**, 251–257 (2021)



Emotional Eyes for Automated Vehicles: Investigating Design Dimensions and Pedestrian Acceptance

Nina Theobald^(✉) , Felix Friedrich , Philip Joisten , and Bettina Abendroth

Technical University of Darmstadt, Department of Mechanical Engineering,
Institute of Ergonomics and Human Factors, Otto-Berndt-Straße 2, 64287 Darmstadt, Germany
{nina.theobald, felix.friedrich,
bettina.abendroth}@tu-darmstadt.de

Abstract. An approach to increase acceptance of automated vehicles (AVs) using emotional eyes of AVs for pedestrian-AV interaction is presented. In an online survey conducted in Germany with 245 participants, key design dimensions for an abstract emotional eye design and the acceptance of AVs expressing Ekman's six basic emotions in pedestrian-AV interactions were examined. Results show that the design dimensions of shape, color and motion enhance emotion recognition in abstract eye designs, with specific dimensions being more effective for certain emotions. While Disgust is the least, Anger and Happiness are the most accepted emotions of AVs by pedestrians. Beyond the attitudes toward AV-expressed emotions studied here, the findings indicate that emotionalizing AVs influences pedestrians' affective and behavioral outcomes in interactions, warranting further studies in more ecologically valid settings.

Keywords: Automated Vehicle · Design · Emotion · External Human-Machine Interface · Human-Machine Interaction · Pedestrian

1 Introduction

Insufficient acceptance of automated vehicles (AVs) by other road users challenges their introduction into road traffic [1]. To improve acceptance, research in human-robot interaction (HRI) proposes using anthropomorphic features and conveying social skills of robots, e.g. by letting them express emotions [2]. However, the context dependency of anthropomorphism in HRI requires a reevaluation of AVs expressing emotions [3].

Aligning with street etiquette, human-shaped social signals like emotional facial expressions could improve AV acceptance. Yet, external human-machine interfaces (eHMIs) for pedestrian-AV interaction mostly utilize light, text or symbols to communicate an AV's status, perception and intention [4].

Thus, in line with contemporary robot designs and recognizing the importance of driver gaze for pedestrian safety [5], this study investigates how key design dimensions influence emotion conveyance via abstract eye design and the acceptance of an AV expressing emotions on its avatar head in interaction with pedestrians.

2 Methodology

An online survey was conducted in Germany with 253 participants with eight responses excluded due to participants completing too quickly or being colorblind ($N = 245$; 50.61% male, 49.39% female; Mean = 30.48 years, Standard Deviation = 12.75 years).

Due to prevailing concerns about the Uncanny Valley Effect, especially in Europe [6], a rather abstract, non-human-like eye design was chosen. For each of Ekman's six basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise) [2, 7] specifications for the design dimensions shape, color and motion were used to create three eye designs building on each other: design 1: shape; design 2: shape and color; design 3: shape, color and motion (see Fig. 1) [8]. Studies on emotional robot faces [9, 10], popular representations like emoticons and human face depictions of the six basic emotions from [11] served as design templates. Particular attention was given to eye-related action units from Ekman's Facial Action Coding System (FACS) [11]. In addition, FACS action units for other facial elements as well as typical human postures and body movements associated with emotion expression were incorporated into the eye designs. Literature on color-emotion associations guided color selection [12, 13]. Context-independent findings on the intensity, frequency and orientation associated with emotions were also considered [13]. In case of multiple design options, traffic-related associations were the main criterion for the final design decision.



Fig. 1. Eye designs representing Ekman's six basic emotions (from left to right: Neutral, Anger, Disgust, Fear, Happiness, Sadness, Surprise) based on shape, color and motion.

In the first study section, participants watched eighteen videos of the eye designs devoid of vehicle context, were tasked with assigning the correct emotions to them using a forced-choice paradigm and optionally provided critiques and improvement suggestions to the designs. First, the six videos of eye design 1, second, the six videos of eye design 2 and third, the six videos of eye design 3 were shown [8]. Each set of six videos was presented in a randomized order, with each video starting with the neutral eye design before transforming into the emotional eye design.

The second study section comprised the description of six brief user scenarios with interactions between a pedestrian and a multifunctional, anthropomorphic AV "EDAG CityBot" [14] in road traffic, each accompanied by a video created in a virtual reality environment showing the AV expressing the situation-specific emotion using eye design 3 on its avatar head (see Fig. 2) [8]. Participants rated the acceptance of each emotional expression based on three items concerning its perceived ease of use ("I find that the [emotion] pair of eyes makes it easier to recognize the AV's concern."), perceived usefulness ("I find the [emotion] pair of eyes useful to recognize the AV's concern.") and attitude toward using ("The [emotion] pair of eyes should be used in the future as an aid for AVs to communicate with pedestrians.") in pedestrian-AV interaction using a 7-point Likert scale [15].

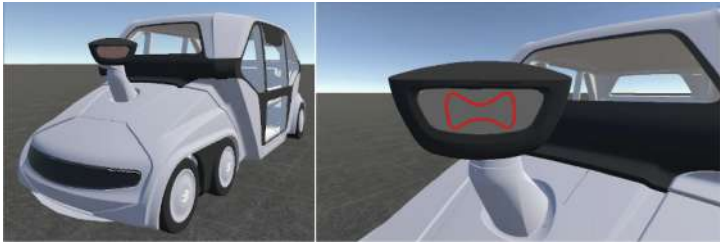


Fig. 2. The multifunctional, anthropomorphic AV “EDAG CityBot” is in its function of passenger transport and expresses the emotion Anger on its avatar head.

3 Results

The results of the first study section focusing on emotion recognition in abstract eye designs involved 206 participants who watched all eighteen videos. A Chi-Square Test of Independence revealed a significant association between design and correct recognition for all emotions (see Table 1). Maximum recognition mostly occurred in design 3. While recognition rates for the emotions Disgust, Happiness and Sadness uniformly increased with the addition of design dimensions, the recognition rate for Fear showed a marked increase with the addition of the design dimension motion in design 3. However, for Anger and Surprise, correct recognition peaked with the design dimensions shape and color in design 2, diminishing with the addition of motion in design 3.

Table 1. Results of correct emotion recognition, Chi-Square Tests of Independence with Cohen’s w and Post-Hoc Tests (n = 206).

Emotion	Emotion Recognition (# _c = number of correct responses)						Chi-Square Test of Independence			Post-Hoc Test* (Critical value = ± 2.64)		
	Design 1		Design 2		Design 3		$\chi^2(2)$	<i>p</i> (two- sided)	w	Standardized Residuals		
	# _c	%	# _c	%	# _c	%				De- sign 1	De- sign 2	De- sign 3
Anger	139	67.5	187	90.8	174	84.5	38.74	< .001	0.25	-6.01	4.41	1.59
Disgust	60	29.1	105	51.0	123	59.7	41.08	< .001	0.26	-6.16	1.54	4.62
Fear	35	17.0	37	18.0	162	78.6	218.42	< .001	0.59	-7.56	-7.21	14.78
Happiness	71	34.5	102	49.5	106	51.5	14.39	< .001	0.15	-3.77	1.54	2.23
Sadness	112	54.4	134	65.0	166	80.6	32.21	< .001	0.23	-4.59	-0.60	5.19
Surprise	173	84.0	175	85.0	157	76.2	6.32	< .05	0.10	1.03	1.47	-2.50

* Alpha level adjusted using Bonferroni correction.

Overall, compared to the other emotions, Disgust and Happiness had lower maximum recognition rates with the design dimensions of Happiness being criticized the most and the design dimensions of Disgust being criticized the second most. The color of Anger was the only design dimension that was not criticized by any participant.

Participants’ improvement suggestions for emotion recognition included integrating pupils (n = 17), eyebrows (n = 7), a mouth (n = 7), familiar emoticons or symbols (n

= 6) and using a more human-like, realistic eye design ($n = 2$) or acoustic signals ($n = 8$).

The results of the second study section focusing on the acceptance of an AV expressing emotions involved 212 participants who watched all six videos. The descriptive statistics of the ratings per acceptance item per emotion are shown in Table 2.

Table 2. Descriptive statistics of the ratings per acceptance item per emotion ($n = 212$).

Emotion	Ease of use				Usefulness				Attitude toward using			
	M	SD	Mdn	Skew	M	SD	Mdn	Skew	M	SD	Mdn	Skew
Anger	5.34	1.59	6.00	-0.91	5.35	1.64	6.00	-0.98	4.49	1.97	5.00	-0.31
Disgust	4.29	1.81	5.00	-0.18	4.27	1.74	5.00	-0.31	3.96	1.85	4.00	0.09
Fear	4.81	1.78	5.00	-0.66	4.85	1.77	5.00	-0.74	4.41	1.76	5.00	-0.38
Happiness	5.09	1.77	5.50	-0.83	5.16	1.80	6.00	-0.92	4.98	1.85	5.00	-0.73
Sadness	5.03	1.55	5.00	-0.70	5.00	1.57	5.00	-0.72	4.62	1.75	5.00	-0.34
Surprise	4.71	1.76	5.00	-0.51	4.78	1.70	5.00	-0.53	4.42	1.83	5.00	-0.26

Legend: M = Mean, SD = Standard Deviation, Mdn = Median, 1 = strongly disagree, 7 = strongly agree.

Friedman tests conducted per acceptance item revealed that emotion acceptance in pedestrian-AV interaction differs significantly between the six emotions (ease of use: $\chi^2(5) = 77.05$, $p < .001$, Kendall's $W = 0.07$; usefulness: $\chi^2(5) = 86.20$, $p < .001$, Kendall's $W = 0.08$; attitude toward using: $\chi^2(5) = 67.78$, $p < .001$, Kendall's $W = 0.06$). Subsequent Post-Hoc Sign Tests (see Table 3) found that Disgust scored significantly worse than all emotions on the three acceptance items. While Anger scored highest for ease of use and usefulness, Happiness scored second-highest for usefulness and highest for attitude toward using, each with significant differences from other emotions.

In the free-text responses, Anger ($n = 3$), Disgust ($n = 3$) and Fear ($n = 2$) were categorized as critical, questionable or unexpected emotions. AVs should only exhibit friendly and neutral responses to environmental cues, refraining from expressing opinions ($n = 2$). Doubts were raised regarding the necessity and value of external emotional communication with AVs ($n = 8$), citing concerns about interpreting emotions ($n = 1$), the artificiality of machine emotions ($n = 2$) and the potential for an AV's emotional responses to prompt deliberate human misbehavior aimed at testing and manipulating the AV ($n = 3$). Broader concerns included AI surveillance with comparisons to dystopian scenarios in science fiction films featuring powerful embodied AIs ($n = 2$).

Table 3. Significant results of Post-Hoc Sign Tests with alpha level adjusted using Bonferroni correction reported with Vargha-Delaney A (VDA) effect size per acceptance item (n = 212).

Item	Emotion 1	Emotion 2	Statistic	df	<i>p</i>	VDA
Ease of use	Anger	Disgust	115	154	< .001	0.68
		Fear	94	141	< .01	0.61
		Surprise	98	135	< .001	0.64
	Disgust	Fear	46	136	< .01	0.40
		Happiness	39	146	< .001	0.34
		Sadness	40	143	< .001	0.35
		Surprise	48	146	< .001	0.38
Usefulness	Anger	Disgust	116	150	< .001	0.69
		Fear	91	136	< .01	0.61
		Sadness	82	123	< .01	0.50
		Surprise	92	134	< .001	0.62
	Disgust	Fear	45	142	< .001	0.38
		Happiness	38	151	< .001	0.32
		Sadness	39	148	< .001	0.34
		Surprise	50	153	< .001	0.38
	Happiness	Surprise	94	148	< .05	0.59
Attitude toward using	Anger	Disgust	87	136	< .05	0.59
		Happiness	50	135	< .05	0.42
	Disgust	Fear	40	129	< .001	0.38
		Happiness	31	142	< .001	0.31
		Sadness	38	141	< .001	0.35
		Surprise	44	137	< .001	0.38
	Fear	Happiness	44	137	< .001	0.38
	Happiness	Sadness	85	133	< .05	0.59
		Surprise	89	140	< .05	0.59

4 Discussion

The study demonstrates that using the design dimensions of shape, color and motion is effective in conveying Ekman's six basic emotions via an abstract eye design, with specific dimensions being more effective for certain emotions. It emphasizes the dependence of correct emotion recognition on the existence of conventional and intuitive expression-emotion relationships and the importance of their adherence in design [2], subject to cultural variations that may affect result transferability [16].

Inconsistencies with studies on the recognition rates of emotions and effect of design dimensions may stem from the chosen level of detail of eye designs and facial elements utilized to convey emotions [9, 17]. Overall, the emotions Disgust and Happiness were the least recognized in this study, possibly due to the focus on eye expressions, while research shows that these emotions are better conveyed by lower facial features [9].

The acceptance of emotions varies, with Disgust being the least accepted, Anger and Happiness being the most accepted emotions of AVs, highlighting the influence of perceived legitimacy of emotions in road traffic on their acceptance [9, 17]. Moreover, the futuristic user scenarios [8], describing interactions between an anthropomorphic AV in currently unknown use cases and a pedestrian, might have limited the elicitation of vicarious experiences and justification of the AV's emotional expressions [17].

Discrepancies between participants' pragmatic ratings and emotional free-text responses, especially for low-valence emotions like Anger, suggest the potential impact

of emotionalizing AVs on human emotional responses [9] and behaviors [18], with interpersonal mechanisms at play [19]. While the online survey effectively captured attitudes toward emotional expressions of AVs, future research could investigate and reveal their impact on human affective and behavioral outcomes in virtual reality which allows for live observation and direct interaction, thus improving ecological validity while ensuring participant safety [2, 3, 17].

Acknowledgments. This research was funded by research project “Campus FreeCity – real lab for the research of a networked fleet of modular robot vehicles”, carried out at the request of the Federal Ministry for Digital and Transport (BMDV), under research project No. 45KI151091. The authors are solely responsible for the content of the paper.

References

1. Prattico, F.G., Lamberti, F., Cannavo, A., Morra, L., Montuschi, P.: Comparing state-of-the-art and emerging augmented reality interfaces for autonomous vehicle-to-pedestrian communication. *IEEE Trans. Veh. Technol.* **70**(2), 1157–1168 (2021)
2. Wang, Y., Hespanhol, L., Tomitsch, M.: How can autonomous vehicles convey emotions to pedestrians? A review of emotionally expressive non-humanoid robots. *Multimodal Technol. Interact.* **5**(12), 1–19 (2021)
3. Roesler, E., Manzey, D., Onnasch, L.: A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Sci. Robot.* **6**(58), 1–31 (2021)
4. Dey, D., et al.: Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles’ external human-machine interfaces. *Transp. Res. Interdiscip. Perspect.* **7**(100174), 1–24 (2020)
5. Onkhar, V., Bazilinskyy, P., Dodou, D., de Winter, J.: The effect of drivers’ eye contact on pedestrians’ perceived safety. *Transport. Res. F: Traffic Psychol. Behav.* **84**, 194–210 (2022)
6. Castelo, N., Sarvary, M.: Cross-cultural differences in comfort with human-like robots. *Int. J. Soc. Robot.* **14**(8), 1865–1873 (2022)
7. Ekman, P.: An argument for basic emotions. *Cognit. Emot.* **6**(3–4), 169–200 (1992)
8. Theobald, N., Friedrich, F., Joisten, P., Abendroth, B.: Emotional eyes for automated vehicles: Investigating design dimensions and pedestrian acceptance - Supplementary material [Data set]. *TUdatalib*. (2024). <https://doi.org/10.48328/tudatalib-1445>
9. Herdel, V., Kuzminykh, A., Hildebrandt, A., Cauchard, J.R.: Drone in love: Emotional perception of facial expressions on flying robots. In: Kitamura, Y., Quigley, A., Isbister, K., Igarashi, T., Bjørn, P., Drucker, S. (eds.) *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–20. ACM, NY (2021)
10. Pollmann, K., Tagalidou, N., Fronemann, N.: It’s in your eyes. In: Alt, F., Bulling, A., Döring, T. (eds.) *Proceedings of Mensch und Computer 2019*, pp. 639–642. ACM, NY (2019)
11. Ekman, P., Friesen, W.V., Hager, J.C.: *Facial Action Coding System*. Research Nexus, Salt Lake City, Utah (2002)
12. da Pos, O., Green-Armytage, P.: Facial expressions, colours and basic emotions. *Colour: Des. Creat.* **1**(1), 1–20 (2007)
13. Johnson, D.O., Cuijpers, R.H., van der Pol, D.: Imitating human emotions with artificial facial expressions. *Int. J. Soc. Robot.* **5**(4), 503–513 (2013)
14. EDAG CityBot. <https://www.edag-citybot.de/en/> (2024). Accessed 30 May 2024
15. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results. Doctoral Dissertation, MIT (1985)

16. Choungourian, A.: Color preferences and cultural variation. *Percept. Mot. Skills*. **26**, 1203–1206 (1968)
17. Wang, Y., Wijenayake, S., Hoggenmüller, M., Hespanhol, L., Worrall, S., Tomitsch, M.: My eyes speak: Improving perceived sociability of autonomous vehicles in shared spaces through emotional robotic eyes. In: *Proceedings of the ACM on Human-Computer Interaction*, pp. 1–30. ACM, NY (2023)
18. Mehrabian, A., Russel, J.: *An Approach to Environmental Psychology*. The MIT Press, Cambridge, Massachusetts (1974)
19. Knutson, B.: Facial expressions of emotion influence interpersonal trait inferences. *J. Nonverbal Behav.* **20**, 165–182 (1996)



Classifying Rapid Deceleration Events for Older Drivers

Misako Yamagishi^(✉)

Toyota Transportation Research Institute, Toyota-city, Aichi, Japan
yamagishi@tttri.or.jp

Abstract. Rapid deceleration events (RDEs) are employed as an index in the driving as assessment of older drivers using naturalistic driving data. The differences between simple RDE (SRDE) and trigger RDE (TRDE) have been analyzed using naturalistic driving data and human characteristics of 84 drivers aged between 50 and 83. The SRDEs and TRDEs were then extracted from the driving data. In this study, SRDEs were extracted based on only acceleration data, while TRDEs were extracted in consideration of acceleration and situations regarded as triggers for braking behavior. The results of the negative binomial and mixed Poisson regression analyses demonstrated that SRDEs up to a detection threshold of 0.5 g can predict the occurrence of TRDEs. However, the relevant factors differed between the two RDEs for the 0.5 and 0.7 g detection thresholds. The study revealed that SRDEs exhibited an increase with age, whereas this trend was not observed for TRDEs. The constructed model suggests that TRDE indices are not simply a scaled-down SRDE. In driving assessments of older drivers based on naturalistic driving, RDEs should be applied in accordance with the specific usage of the assessment and the available information to detect the events.

Keywords: Older drivers · Rapid Deceleration events · Naturalistic driving

1 Introduction

For older drivers, collecting information and obtaining self-awareness about their driving is important to ensure safe driving. Consequently, regular observation and evaluation using naturalistic driving monitoring with drive recorders is useful. The rapid deceleration events (RDEs) index represents a driving evaluation indicator for naturalistic driving. RDEs are detected based on acceleration variation, utilizing an accelerometer mounted on the driving monitoring system, such as drive recorders [1–3]. These are recognized as driving behaviors that require improvement, as they are not only lead to traffic violations but also because the brief duration and distance between braking events increases the likelihood of collision with objects. Nevertheless, not all RDEs result in high-risk conditions. Some studies have adopted RDEs selected based on consideration of the surrounding situation [4], which raises the question of whether RDEs detected based on different information should be considered identical indicators.

This study addresses two different types of RDEs, namely a simple RDE (SDRE) and a trigger RDE (TRDE). An SRDE is extracted automatically based only on acceleration

data, which is similar to detection using probe vehicle data. A TRDE is extracted through the combination of acceleration and video data. The selection is based on images of the surroundings when an RDE occurs, which is similar to the evaluation of near-crash events. Consequently, TRDEs are constrained to incidents involving objects that trigger deceleration. In other word, the cause of deceleration can be readily identified and represents behavior aimed at avoiding contact with an object. By contrast, SRDEs can occur without conflict objects for various reasons, including driving styles and poor braking skills.

This study investigated the differences between SRDEs and TRDEs in terms of the relationships between the events and the human characteristics associated with RDEs. If a TRDE is a part of an SRDE, it can be predicted by an SRDE. Furthermore, if a TRDE is attributable to a diminished probability of the occurrence of an SRDE, the human characteristics related to the TRDE would be similar to those of an SRDE, with the exception of the degree of influence. In addition, Eby et al. [3] compared two RDEs with different detection thresholds (0.35 and 0.75 g) for older drivers and concluded that the implications of RDEs vary according to the levels of the detection threshold. Thus, this study aggregated the number of SRDEs and TRDEs by varying the detection threshold and investigated the influences of detection thresholds on these RDEs.

2 Methods

2.1 Participants and Human Characteristics

Data from older drivers in the Data Repository for Human Life-Driving Anatomy (DAHLIA) constructed and managed by the Nagoya Center of Innovation (COI) project was utilized. The DAHLIA construction process, including data collection, was initiated after approval from the Ethics Committee of the Institute of Innovation for Future Society at Nagoya University. Oral and written informed consent was obtained from each participant. Data obtained from 84 drivers aged 50–83 years (mean = 68.5 years, SD = 7.7; male = 53, female = 31) from the DAHLIA were analyzed. The participants, who resided in Aichi Prefecture in Japan, held a legal license to drive and drove privately owned vehicles regularly. The DAHLIA includes various human characteristics. For cognitive function assessment, this study used the Mini-Mental State Examination [5] related to general cognitive performance; the Trail Making Test (TMT) [6] to evaluate executive functions, visual search speed, and set shifting; and the Advanced Industrial Science and Technology's Cognitive Aging Test [7] comprising a Working memory task, visual Attention task, Planning task, and Task-switching task.

For visual function assessment, the static visual and kinetic visual acuities (SV and KV) were measured using a visual acuity meter (AS-4F, KOWA Co., Ltd), and contrast sensitivity (CS) was evaluated using the CSV-1000 (VECTORVISION®). Additionally, the effect of glare on contrast sensitivity was also evaluated (CS with glare).

For subjective assessment, the participants were given a driving style questionnaire (DSQ) [8] that covers driver attitudes, habits, and beliefs regarding common driving tasks. This comprised questions about eight factors: self-confidence in driving skills, hesitation for driving, impatience driving, methodical driving, preparatory maneuvering, cars as self-expression, moodiness in driving, and anxiety about accident.

2.2 Naturalistic Driving Data

The naturalistic driving data of the participants was recorded by installing drive recorders (BU-DRHD431, Yupiteru Co., Ltd) in their private vehicles. The device collected various data at a frequency of 10 Hz during driving, including video data in front of the vehicle, global positioning system (GPS) data, distance driven, GPS-based velocity, and acceleration. Three types of data were collected: continuous records per trip, continuous text data except footage data per trip, and event records of the 10-s period before and after a large sudden acceleration variation in each trip.

First, the number of days and distance driven per month for each participant was calculated from the continuous text data. Subsequently, months with 20 or more recorded days were selected, and the average distance was calculated. The data analyzed was selected one month approximate to average distance and before and after the month from the first year when participants enrolled in this study, which was adopted three months for the analyses.

The number of SRDEs was calculated by extracting data points that exceeded the given detection thresholds (0.3–0.8 g) for the longitudinal acceleration of the continuous text data and counting the points where the velocity decayed after the data point. TRDEs employed organized event records, because the continuous text data lacked video data to corroborate the situations of the events. First, decelerating scenes were selected from the event records and categorized using the same detection thresholds as the SRDEs. Subsequently, a trained evaluator viewed the video data of the selected events and selected those that included objects, such as vehicles, pedestrians, bicycles, or traffic signals, which are considered triggers for deceleration. The evaluator repeated the process in order to compensate for the initial selection.

3 Results

3.1 RDE Counts and Demographics

Participants with missing values in the naturalistic driving and human characteristics data were excluded from the analyses, resulting in a final sample size of 54 participants. Table 1 presents the SRDEs and TRDEs by longitudinal acceleration levels, along with their correlation with age.

3.2 Relationship Between SRDEs and TRDEs

After extending the data longitudinally to include data from after the first year, negative binomial and mixed Poisson regression analyses (with participant, residence, and year as random effects) were conducted with SRDEs as explanatory variables and TRDEs as objective variables. The negative binomial and the mixed Poisson regression analyses were significant for SRDEs with 0.4, 0.5, and 0.6 g. The results of TRDE with 0.5 and 0.7 g are listed in Table 2 as examples of significant and non-significant models.

Table 1. The number of SRDEs and TRDEs per 10,000 km by longitudinal acceleration levels and their correlation with age (p value significance level 0.01: **, 0.05: *, and 0.10: †)

Detection threshold	SRDE (per 10,000 km)				TRDE (per 10,000 km)			
	Mean	SD	Correlation with Age		Mean	SD	Correlation with Age	
0.3 g	3664.43	3875.64	0.26	†	22.44	34.57	0.23	†
0.4 g	623.63	833.13	0.28	**	22.09	33.51	0.23	†
0.5 g	114.24	156.66	0.30	**	19.59	30.33	0.23	†
0.6 g	26.91	39.98	0.30	**	10.17	16.53	0.26	†
0.7 g	8.81	17.3	0.24	†	4.85	8.77	0.25	†
0.8 g	3.06	9.34	0.14		2.3	4.45	0.20	

Table 2. Comparison of predictive models for 0.5 and 0.7 g of TRDEs

	TRDE: 0.5 g				TRDE: 0.7 g			
	Negative binomial		Mixed Poisson		Negative binomial		Mixed Poisson	
	β	p value	β	p value	β	p value	β	p value
Intercept	−6.529	0.000	−7.005	0.000	−7.598	0.000	−8.338	0.000
SRDE	0.013	0.007	0.011	0.000	0.021	0.391	0.044	0.054
AIC	687.02		659.02		402.56		372.82	
Deviance	165.39		162.56		137.71		99.07	

3.3 Human Characteristics Associated with RDEs

The analysis revealed that the trend of the results changes near the 0.5 g detection threshold. Consequently, negative binomial and mixed Poisson regression analyses (with residence as the random effect) were conducted for SRDEs and TRDEs at 0.5 and 0.7 g. The explanatory variables were cognitive functions, visual functions, and driving styles, and the model performance was evaluated in terms of AIC (Akaike Information Criterion) and deviance. The 0.5 g detection threshold was based on the SRDE model of the mixed Poisson regression analysis and the TRDE model of the negative binomial regression analysis. The 0.7 g detection threshold was based on the mixed Poisson regression analysis models for both RDEs (Table 3).

For 0.5 g, Working memory was a significant variable for SRDEs and TRDEs, while other significant variables differed for both RDEs. For 0.7 g, all significant variables differed between SRDEs and TRDEs. Comparing 0.5 and 0.7 g, SRDEs had fewer relevant variables at 0.7 g than at 0.5 g. By contrast, for TRDEs similar factors to SRDE model listed as significant variables. Age was significant in the SRDE model, but not in the TRDE model. In the SRDE model, the 0.5 g was associated with cognitive functions

Table 3. Results for each analysis for SRDEs and TRDEs with 0.5 and 0.7 g

	0.5 g			0.7 g				
	SRDE model		TRDE model	SRDE model			TRDE model	
	<i>B</i>	p value	<i>B</i>	p value	<i>B</i>	p value	<i>B</i>	p value
Intercept	−15.146	0.000	−10.213	0.014	−16.190	0.001	−17.290	0.004
Age	0.147	0.000	0.023	0.580	0.148	0.001	0.057	0.337
Sex	0.479	0.048	−0.720	0.172	0.395	0.249	−1.309	0.127
TMT part A	−0.013	0.375	−0.066	0.044	−0.017	0.461	−0.091	0.106
TMT part B	0.013	0.017	0.002	0.867	−0.002	0.864	0.026	0.223
Working memory	0.174	0.000	−0.163	0.028	−0.041	0.488	−0.204	0.049
Attention	−0.075	0.019	0.111	0.080	−0.012	0.676	0.187	0.022
Planning	0.405	0.085	−0.204	0.425	0.225	0.386	−0.650	0.075
Task-switching	0.102	0.000	−0.015	0.656	0.040	0.201	0.006	0.889
SV	0.279	0.558	−2.211	0.054	−1.202	0.230	−1.797	0.290
KV	−1.127	0.442	−1.614	0.385	1.380	0.442	−3.163	0.239
CS	−0.967	0.415	0.207	0.899	1.908	0.177	1.776	0.372
CS with glare	1.030	0.182	0.537	0.648	0.315	0.755	−1.372	0.372
DSQ: Confidence	−0.356	0.138	0.110	0.800	−1.275	0.000	−0.905	0.196
DSQ: Hesitation	−1.031	0.000	0.538	0.138	−0.224	0.464	0.499	0.331
DSQ: Impatience	−0.559	0.065	1.349	0.027	0.000	0.999	2.441	0.033
DSQ: Methodical	−0.680	0.003	0.633	0.097	0.340	0.210	1.225	0.041
DSQ: Preparatory	1.767	0.000	0.323	0.248	0.648	0.010	0.262	0.472
DSQ: Moodiness	1.210	0.000	−1.046	0.096	0.196	0.650	−2.028	0.086
DSQ: Anxiety	−0.563	0.000	0.473	0.156	−0.886	0.003	0.266	0.618
Distance group	−1.292	0.000	−0.428	0.461	−0.534	0.235	0.903	0.357
AIC	408.45		227.17		162.63		130.26	
Deviance	35.38		47.36		39.73		24.89	

and driving styles, while the 0.7 g was mainly associated with driving styles. The 0.5 and 0.7 g TRDE models included cognitive functions and driving styles.

4 Discussion

This study used SRDEs and TRDEs derived from naturalistic driving data of older drivers and examined the differences between them. If TRDEs are part of SRDEs, they are considered to be equivalent. Furthermore, if the TRDE is attributable to a mere reduction in the probability of occurrence of an SRDE, the human characteristics related to a TRDE would be similar to an SRDE, with differences in the degree of influence. The results indicated that SRDE below the 0.5 g detection threshold are plausible variables to predict the TRDEs. Moreover, investigating the relationship between RDEs and human characteristics, for 0.5 g model, some of the variables involved in SRDEs were also related to TRDEs. In contrast, for the 0.7 g model, SRDEs and TRDEs exhibited different models of the relevant variables. These results provide partial support for the predictions described in the Introduction section.

The difference between the detection thresholds of 0.5 and 0.7 g for SRDEs and TRDEs was examined. The results indicate that TRDEs were part of SRDEs and that the likelihood of the occurrence of TRDEs could be predicted by SRDEs for 0.5 g. SRDEs were related to four cognitive functions and five driving styles. This is considered a feature of SRDEs in compositional situations in which the events occur. Compared with TRDEs, SRDEs were linked to a greater number of driving styles, including a tendency to harsh brake at a stop. Habitual harsh braking makes it more challenging to anticipate the appearance of sudden objects in the moment. Consequently, if drivers who brake harshly can be distinguished from the balance of SRDEs and TRDEs, it may be possible to provide guidance and advice to modify the behaviors of older drivers. For 0.7 g, the probability of occurrence of TRDEs could not be predicted by SRDEs, and the variables associated with SRDEs and TRDEs were completely different. These results indicate that SRDEs may serve as indicators of driving style, while TRDEs can function as an index that combined two key aspects of drivers: their cognitive functions and driving styles. With regard to cognitive functions, a lower Working memory score was identified as a significant predictor of RDEs, aligning with the conclusions of Eby et al. [3]. Furthermore, the data indicated a correlation between the frequency of TRDEs and a high level of Impatience. Impatience is a concept related to impulsiveness, and this result is consistent with the findings reported by Owsley et al. [9], who stated that subjects reporting driving violations were more likely to have high impulsivity scores.

With regard to the relationships between age and RDE, SRDEs can be considered an indicator that is directly affected by aging, whereas TRDEs are not strongly associated with age. However, the fact that TRDEs were indicated to be related to TMT part A (Table 3), Working memory, and Attention as affected by age raises the possibility that age may be indirectly related to TRDE. Thus, network models incorporating driving-related factors are essential for elucidating the impact of TRDEs on the driving behavior of older drivers.

However, this study used the number of events as a measure of RDEs, which implies that the aspects of any given event were considered identical across all detection thresholds. A more detailed analysis of each RDE, including the time and distance from the start of braking to the end of deceleration, may prove useful in distinguishing between mild events and more critical events. To clarify the significance of RDEs and propose them as an effective driving assessment indicator would assist in the comprehension of

the meaning and significance of the RDEs of various drivers, thereby contributing to safe driving for older drivers.

Acknowledgement. This study was conducted with the support of the Japan Science and Technology Agency as a part of the COI program. I would also like to express my gratitude to the members who provided their cooperation with the data collection and analysis.

References

1. Keay, et al.: Older drivers and rapid deceleration events: Salisbury Eye Evaluation Driving Study. *Accid. Anal. Prev.* **58**, 279–285 (2013)
2. Chevalier, et al.: Predictors of older drivers' involvement in rapid deceleration events. *Accid. Anal. Prev.* **98**, 312–319 (2017)
3. Eby, et al.: Factors related to rapid deceleration events among a large cohort of older drivers. *Transp. Res. Part F: Traffic Psychol. Behav.* **66**, 76–86 (2019)
4. Seacrist, et al.: Near crash characteristics among risky drivers using the SHRP2 naturalistic driving study. *J. Saf. Res.* **73**, 263–269 (2020)
5. Yakushiji, et al.: Norms of the Mini-Mental state Examination for Japanese subjects that underwent comprehensive brain examinations: the Kashima Scan Study. *Intern. Med.* **53**(21), 2447–2453 (2014)
6. Hioka, et al.: Usefulness of a novel higher brain dysfunction screening test for evaluating higher brain function in healthy persons. *J. Med. Investig.: JMI.* **64**(3.4), 280–285 (2017)
7. Sato, et al.: Relations between elderly drivers cognitive functions and their compensatory driving behaviors. In: *Humanist 5th Conference* (2016)
8. Ishibashi, et al.: Indices for characterizing driving style and their relevance to car following behavior. In: *SICE annual conference 2007*, pp. 1132–1137. IEEE (2007)
9. Owsley, et al.: Impact of impulsiveness, venturesomeness, and empathy on driving by older adults. *J. Saf. Res.* **34**(4), 353–359 (2003)



Which Eye-Tracking Measure is Better for Examining Pedestrian Crossing Behavior?

Putra A. R. Yamin^{1,2}  and Jaehyun Park³  

¹ Department of Industrial & Management Engineering, Incheon National University, Incheon, South Korea

² Department of Industrial Engineering, Bandung Institute of Technology, Bandung, Indonesia

³ Department of Industrial Engineering, Konkuk University, Seoul, South Korea
jaehpark@konkuk.ac.kr

Abstract. Extensive research has been done on external human-machine interfaces, and a growing number of studies have been conducted measuring eye tracking. This study investigated different eye-tracking measures and their relation to pedestrian crossing initiation time in virtual reality. The objective is to understand further the applications and differences between measuring saccade and fixation for examining pedestrian crossing behavior. The pedestrian crossing initiation time, saccade count, fixation length, and fixation count were examined with the HTC Pro Eye virtual reality eye-tracking headset ($N = 30$). Significant differences were found between eHMI type, crossing initiation time, and all eye metrics (fixation length, fixation count, and saccade count). Interaction effects were found between crossing initiation time and saccade count. Saccade count was found to be the most appropriate eye-tracking measure for pedestrian crossing studies, followed by fixation length and then fixation count. Furthermore, it is not necessary to include both saccade and fixation. This study provides recommendations for future eye-tracking research, especially in human-vehicle interaction.

Keywords: external human-machine interface (eHMI) · automated vehicle · vehicle-pedestrian interaction · eye-tracking

1 Introduction

External human-machine interfaces (eHMIs) have been proposed to compensate for the lack of communication between automated vehicles and pedestrians. Extensive research has been conducted using crossing time and subjective ratings as the most used measures. However, pedestrian-vehicle interaction studies using eye tracking are limited. Furthermore, the eHMI type and size of the vehicle are also important factors that lead to different results [1, 2]. Previous eHMI with eye-tracking studies used either fixation length (for example, in [3]) or saccade count (see [4, 5]). Both eye metrics provide valuable information. This study investigates the applications and differences between measuring saccade and fixation for examining pedestrian crossing behavior.

The present study aimed to analyze and compare pedestrians' saccade count, fixation length, and fixation count in a crossing situation against automated vehicles. The research

questions of this study are as follows: (1) Is there any correlation between the eye-tracking measures and crossing initiation time? (2) Which eye-tracking measure is better for a pedestrian crossing study? (3) Is it necessary to include both fixation and saccade in a study? And (4) Is it mandatory to include both fixation length and fixation count in a study? This study provides an understanding of the advantages and limitations of eye-tracking metrics used in eHMI research and guides future research.

The findings of this study are expected to inform future eye-tracking eHMI studies and eye-tracking studies involving pedestrian crossing in general.

2 Methodology

2.1 Participants

The study observed 30 individuals, comprising 15 females, with a mean age of 23.46 (standard deviation [SD] = 3.6). All the participants possessed normal vision or used corrective lenses when required. All the participants had prior experience utilizing PMVs like electric bikes, electric scooters, and segways and had encountered other PMVs while commuting as pedestrians. The study offered a gratification of USD 15 to the participants.

2.2 Apparatus

The experimentation was carried out safely and with control using the virtual reality pedestrian simulator through the HTC VIVE Pro Eye head-mounted display (HMD). This device features two OLED displays, each with a resolution of 1440×1600 pixels [6]. The virtual environment and vehicles were created with Unity3D Version 2018.3.7f1 utilizing Westdrive X LoopAR [7], an open VR project designed to evaluate user interaction methods.

The virtual environment resembled an urban traffic scenario with a one-way urban pedestrian with no crosswalk. The virtual road has a cross-section of 3.5 m. Participants could move freely within 15 square meters. The position and orientation of the HMD in the three-dimensional space are determined using the HTC VIVE base stations. The HMD provides recordings of movement with a system-to-system latency of 22 ms. Participants' eye movements were recorded using a head-mounted eye-tracking system. The binocular eye-tracker provided the eye-tracking data at a precision of about $0.5\text{--}1.1^\circ$ of visual angle [6]. The sampling frequency of motion data and eye tracking data was set to 20 Hz. Before the virtual reality experiment, a calibration routine was performed for each subject.

2.3 Experimental Design

The experiment was a within-subjects design with the independent variable eHMI type (no-eHMI, text eHMI "WILL STOP", or green-red light bar eHMI) (Fig. 1) and seven AOIs (Fig. 2). Each participant faced each eHMI type on a small AV approaching at a 60° angle from the target position to cross. Figure 1—right is a screenshot of the VR environment from the top view. The small AV was designed to look like a future



Fig. 1. Left: WILL STOP eHMI, Middle: light bar eHMI Right: Top view of the virtual environment and 60° approach angles of the APMV.

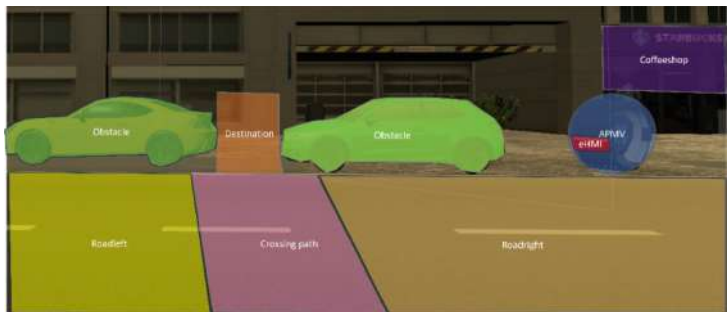


Fig. 2. Designated Area of Interest (AOI)

automated PMV, and its dimensions are 1.8 m $\times$ 0.6 m. The APMV was moving at 15 km/h (4.167 m/s). The dimension of the eHMI was set at 0.25 $\times$ 0.5 m.

The crossing time, saccadic movement, and fixation duration were measured. The crossing time was measured from the moment when the eHMI image appeared on the screen (the tenth second) until the participant stepped on the road to cross. Saccades were defined based on whether the speed of the gaze point exceeded 30°/s and filtered using a Savitzky-Golay filter. The number of saccades during a trial reflects how many eye movements the participants made to reach a decision. Fixation duration is the duration of a participant gazing at the eHMI area of interest. The minimum fixation duration was set to 40 ms [8]. For the trials with no-eHMI, fixation was calculated when the participant is gazing at the front of the APMV.

2.4 Procedure

At the beginning of the experiment, the participants were required to complete a sociodemographic questionnaire and consent to participate in the study. Along with this, a document outlining the goals of the investigation and the sequence of events was also provided. The participants were then put on the HTC Vive Pro Eye HMD, and eye

calibration was carried out next. The eye position and VR surroundings are matched according to the participant’s height. The participants were allotted approximately five minutes for preliminary training in acclimatizing to the VR setting and understanding the pedestrian route across the road. The participant was instructed to imitate pedestrian behavior and be prepared to encounter different types of filler vehicles from various angles within the virtual setting. Furthermore, the participants were informed that conventional vehicles with drivers, AVs without communication systems, and AVs with communication systems could be present. Subsequently, the participants were informed via onscreen instructions that the first trial phase had begun. Their designated task was to cross the road safely. After each condition, participants were allowed a 3-min break for rest. The overall duration of each experiment was 60 min.

3 Results

Significant differences were found between eHMI type, AOIs, and all eye metrics (saccade count, fixation length, and fixation count). Each eye metric has its advantages; however, if one eye measure must be chosen, saccade count was found to be the most appropriate eye-tracking measure for pedestrian crossing studies, followed by fixation length and then fixation count. It is not necessary to include both saccade and fixation. Lastly, it is highly recommended that both fixation length and fixation count be measured at the defined Area of Interest (AOI) if fixation opts as the eye-tracking measure for the eHMI study (Table 1).

Table 1. Crossing time, fixation count, and fixation length on different eHMI types.

eHMI	Crossing Time (ms)	Saccade Count mean	AOI	Fixation Count mean	Fixation Length mean
No-eHMI	2923*	10.43**	Crossing path	0.73	1225
			Destination	2.37	258
			Front of APMV	2.32*	2370
			APMV	6.29***	872
			Roadleft	2.82	237
			Roadright	2.24	374
			Obstacle	0.76	276
			Coffeeshop	0.53	143
WILL STOP	2361*	5.86**	Crossing path	3.73	3427
			Destination	2.49	289
			eHMI	5.32***	4577
			APMV	2.89	411
			Roadleft	0.82*	132

(continued)

Table 1. (continued)

eHMI	Crossing Time (ms)	Saccade Count mean	AOI	Fixation Count mean	Fixation Length mean
			Roadright	1.24*	258
			Obstacle	0.98	314
			Coffeeshop	0.73	164
Lightbar	2763*	11.04**	Crossing path	3.92	3727
			Destination	2.54	297
			eHMI	3.65***	3638
			APMV	2.46	379
			Roadleft	0.92	130
			Roadright	1.44	249
			Obstacle	1.28	352
			Coffeeshop	0.71	170

*p < .05; **p < 0.01; ***p < 0.001

4 Discussion

In all eye-tracking studies, the author’s definition of the eye metric and the fixation/saccade filter is important and affects the study’s results. A saccade occurs when the gaze moves quickly between objects, and a fixation is when the gaze is directed to a specific object for a brief period [9].

The eHMI is the main AOI; however, measuring fixation on the eHMI as the AOI has its limitations. It was found from this study that fixation length would increase and average higher for eHMI types that are difficult to interpret, such as text eHMI. However, after participants started crossing, mixed results of fixation length were observed, which limits the potential for fixation to be a predictor of crossing decisions. One reason is that participants kept fixating on the eHMI while crossing the road when the eHMI type is text-based but less so on light bar eHMI. By comparing average fixation duration across AOIs, one could distinguish areas that were looked at for longer durations than others [9].

Saccade counting on several AOIs provides more insight because it was shown that a high number of saccades, when interacting with a given eHMI, consistently leads to a lower crossing initiation time. Next, we could evaluate the participant’s scan path before crossing, which provides more information on how the participant decided to cross [10].

Eyes precede the actions, but conclusions regarding the emotional reaction to the visual scene cannot be made with eye tracking alone. Other measures, such as facial expression analysis or EEG, can help fill the gaps in future studies.

5 Conclusions

Eye tracking can provide valuable information about the perception of pedestrians, which would be hard to grasp with any other research tool. Saccade count was the most appropriate eye-tracking measure for pedestrian crossing studies, followed by fixation length and fixation count. It is not necessary to include both saccade and fixation. Lastly, it is highly recommended that both fixation length and fixation count be measured at the defined Area of Interest (AOI) if fixation opts as the eye-tracking measure for the eHMI study.

Research in this area has further contributed to our understanding of human–vehicle interaction with eye tracking.

References

1. Guo, J., et al.: External human–machine interfaces for autonomous vehicles from pedestrians' perspective: A survey study. *Sens.* **22**(9), 3339 (Apr. 2022). <https://doi.org/10.3390/s22093339>
2. Lau, M., Jipp, M., Oehl, M.: Toward a holistic communication approach to an automated vehicle's communication with pedestrians: combining vehicle kinematics with external human-machine interfaces for differently sized automated vehicles. *Front. Psychol.* **13** (Jul. 2022). <https://doi.org/10.3389/fpsyg.2022.882394>
3. Bindschadel, J., Krems, I., Kiesel, A.: Interaction between pedestrians and automated vehicles: Exploring a motion-based approach for virtual reality experiments. *Transp. Res. Part F Traffic Psychol. Behav.* **82**, 316–332 (Oct. 2021). <https://doi.org/10.1016/j.trf.2021.08.018>
4. Eisma, Y.B., Reiff, A., Kooijman, L., Dodou, D., de Winter, J.C.F.: External human-machine interfaces: Effects of message perspective. *Transp. Res. Part F Traffic Psychol. Behav.* **78**, 30–41 (Apr. 2021). <https://doi.org/10.1016/j.trf.2021.01.013>
5. Eisma, Y.B., Van Gent, L., De Winter, J.: Should an external human-machine interface flash or just show text? A study with a gaze-contingent setup. *Transp. Res. Part F Traffic Psychol. Behav.* **97**, 140–154 (Aug. 2023). <https://doi.org/10.1016/j.trf.2023.07.005>
6. HTC Corporation, “HTC Pro Eye Includes.” (2022). Accessed 29 Dec 2022. [Online]. Available: <https://www.vive.com/us/product/vive-pro-eye/specs/>
7. Nezami, F.N., et al.: Westdrive X LoopAR: An open-access virtual reality project in unity for evaluating user interaction methods during takeover requests. *Sens.* **21**(5), 1879 (Mar. 2021). <https://doi.org/10.3390/s21051879>
8. Nyström, M., Holmqvist, K.: An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behav. Res. Methods.* **42**(1), 188–204 (Feb. 2010). <https://doi.org/10.3758/BRM.42.1.188>
9. Mahanama, B., et al.: Eye movement and pupil measures: A review. *Front. Comput. Sci.* **3**, 733531 (Jan. 2022). <https://doi.org/10.3389/fcomp.2021.733531>
10. Biassoni, F., Bina, M., Confalonieri, F., Ciceri, R.: Visual exploration of pedestrian crossings by adults and children: Comparison of strategies. *Transp. Res. Part F Traffic Psychol. Behav.* **56**, 227–235 (Jul. 2018). <https://doi.org/10.1016/j.trf.2018.04.009>

UI/UX (I)



Implementing Affective Considerations for Design Pattern Selection for Website Homepages

Carissa M. Araco, Maria Mikaela A. Arriola[✉], Reinabelle Angelique Y. Gonzales, Antonio Lorenzo L. Sahagun, and Ronaldo V. Polancos

De La Salle University-Manila, 1004 Manila, Philippines

Abstract. A large variety of design patterns is found in each design component per webpage that contribute toward the overall affective design of an interface. Usability tools are often used to measure the functionality of a design. Yet, dissatisfaction can persist, which could be caused by the users' emotions when perceiving the visual appearance of the interface they are interacting with. This paper attempts to amalgamate findings from empirical evidence and emotional ratings obtained from users as variables in the Non-Dominated Sorting Genetic Algorithm II (NSGA II) to develop a set of combined design patterns that fit the context of the website. This study explores methods of integrating ergonomics into web page design, given that the age of technology is transforming the world daily.

Keywords: Usability · Web Page Design · Cognitive Ergonomics

1 Introduction

Humankind's growing dependence on technology has resulted in many traditional practices transforming in the name of innovation. Business entities and governing bodies now utilize their online presence to communicate with their customers, using websites as their platform for their transactions. Homepages provide an overview of a website's purpose [1]. It sets the tone of a user's first impression, which means that the visual appeal is crucial to motivate the user to continue interacting with the webpage [2], prevent frustration from arising, and increase the "bounce rate" [2]. Web designers are often constrained by time and budget when designing web pages; hence, designing presentable layouts that effectively communicate the information using cognitive elements. Website homepage design patterns have been assessed for usability and aesthetics at the level of individual design components. However, the combination of the best design patterns of page layouts, screen dynamics, menus, headers and footers, and content layouts have not been simultaneously considered in a webpage design study and measured its effectiveness and efficiency using user emotions and strength of evidence.

2 Web Page Design

2.1 Design Patterns

Websites use F-shaped, Z-shaped, columnar, and grid or magazine page layouts. The F-shaped design conveys brief, concise information. The Z-shape fits the letter Z's orientation, and keywords reduce page content for massive web presentations. The columnar design pattern is known for integrating vertical columns on web pages. It is easy to read and lets readers choose what to see [5]. The grid design is the most attractive, maintains structure, and supports a wide range of information due to its flexible architecture [6]. Users can view all page information on a dynamic screen with page, scroll, and slider navigation. Pages are used for landing pages, event portals, and low-content websites [7]. Scrolling simplifies reading without page components like forward and back. Instead of scrolling, sliders increase website engagement and control [8]. The fixed sticky menu is dynamic since it stays displayed while a person scrolls down a webpage. A single menu is for easy recognition, which enhances comprehension, which is crucial for complex systems. Finally, the double menu style is popular for providing additional information and navigation options. A multi-subunit website's first menu bar is the overall navigation, while the second is subunit-specific [9]. The content arrangement suggests high-intensity concentration from a website grid [10]. Content organized in boxes permits quick page scanning and increases brand recognition [11]. Finally, a banner attracts users and gives vital website or content information [12]. The final design elements are headers and footers that group material. A website's rectangular headers and footers convey brand and identity information.

2.2 Emotion

Emotions are complex psychological and physiological responses to a stimulus or event, including feelings, thoughts, and behavioral reactions. A well-designed user interface can inspire pleasure, contentment, and delight, enhancing user experience and motivation to finish a transaction. However, poorly designed user interfaces can cause frustration, uncertainty, and discomfort, resulting in a bad user experience. Evidence has found a correlation between human emotion and user experience under Human-Computer Interaction (HCI) and Uses and Gratification Theory.

3 Methodology

3.1 Phase 1: Initial Assessment of Websites

To validate the initial design patterns, a particular design pattern must be present on the Philippine government website homepages to be considered in the study. The researchers visited one hundred sixty-one (161) Philippine government websites and collated all the design patterns on each website's homepage. The eligibility of affective design and measure of usability was based on literature that studied the interaction of human emotion with a product design. In addition to the selected variables, the empirical support would take on recommendations from similar studies and utilize standardized tools that ensure data processing and analysis are conducted most accurately.

3.2 Phase 2: Initial and Secondary Survey

The first objective function of the study aims to maximize the “Strength of Evidence,” or the academic backing on each design pattern and its significant influences on website design and user experience.

$$\text{Maximize Strength of Evidence Score} = \sum (e_i \times x_i) \quad (1)$$

$x_i = 1$; if Design Pattern i is chosen.

$= 0$; if Otherwise.

e_i = Evidence Score of Design Pattern i

$$\text{Maximize Emotional Score} = \sum (x_i \times r_{ij}) \quad (2)$$

$x_i = 1$; if Design Pattern i is chosen.

$= 0$; if Otherwise.

r_{ij} = Emotion Rating of Design Pattern i .

The strength of evidence of a solution is computed by multiplying the evidence score by the binary decision variable of design pattern i . The evidence score of each design pattern is defined as the supporting research-based evidence. The scores were determined through a comprehensive review of related literature and identification of the occurrence of the significance of these design patterns stated in previous studies. The related literature was used to select the design pattern elements and their corresponding strength of evidence ratings, most of which came from journals with Quartile Levels of 1, 2, and conference proceedings. The second objective function of the non-dominated sorting genetic algorithm aims to maximize the positive emotion score for each solution. The emotion rating for each emotion was determined through a survey performed. The survey was distributed to adults aged 18-60 modelling the demographic of experienced government website users. 100 responses were collected. An emotion rating Likert scale based on the Geneva Emotion Wheel Literature, was used to allow participants to rate the intensity of their emotions when encountering the design patterns. Twenty (20) emotions were presented in the survey, 10 of which are considered “Positive” emotions, and the other 10 are “Negative” emotions. With these, the ratings for each emotion for all design patterns were used to compute the emotion rating.

3.3 Phase 3: Webpage Design and Prototype Testing

The algorithm used in the study’s methodology is a Multi-Objective Hybrid Non-Dominated Sorting Genetic Algorithm, also known as Hybrid NSGA-II. NSGA-II evolves solutions through mathematical operations while promoting convergence of solution points to an optimal Pareto-Front. When running the algorithm, an initial population and 108 design pattern combinations were created in the solution space. After using the crowding distance elimination, the results of the NSGA-II were trimmed down to the top five design pattern combinations (DPC) for prototype testing. These solutions all carry higher-performing design considerations based on the NSGA-II inputs and parameters (Table 1).

Table 1. Design Pattern Combinations from NSGA-II Result.

Design Pattern	Page Layout	Screen Dynamics	Menu	Content Layout
1	Grid/Magazine Layout	Scrolling	Fixed Sticky Menu	Grid of Cards
2	Grid/Magazine Layout	Scrolling	Fixed Sticky Menu	Boxes
3	Grid/Magazine Layout	Scrolling	Single Menu	Grid of Cards
4	Columnar	Scrolling	Single Menu	Grid of Cards
5	Columnner	Scrolling	Single Menu	Boxes

The generated combinations underwent a prototype with testing parameters such as participants and tasks, detailed in the next sections. Given that there are six prototypes to be tested, one being the original website and five being the generated prototypes, the study required unique groups of participants for each. This removed the possibility of a learning curve between prototypes. The prototype testing used 30 unique participants for each prototype, leading to 180 participants for the test. Based on a preliminary survey conducted before the execution of the test, the age bracket of participants was determined to be from 18 to 25. The participants must also have prior experience with any government website, and the frequency of their usage of these websites was noted. Quant-UX was used to develop, test, and analyze the prototypes. The tasks were selected for all the participants across all prototypes. The basis for choosing the prototype tasks was that these were assumed to be the things a user typically accesses or navigates through on the website. In addition, the group aimed for each task to be designed to target one design component.

4 Methodology

4.1 Phase 1: Initial Assessment of Websites

To analyze the experiment runs of the prototypes, the testing results were standardized and converted into a Single Usability Metric for each prototype. The Single Usability Metric (SUM) score is a value that pertains to the prototypes' overall usability. This evaluation method uses four (4) metrics: Time of Task, Task Completion, Error Rate, and Satisfaction. DPC 5 was compared against DPC 3 and 4. DPC 3 was able to score a 3.442 process sigma, while DPC 4 was able to score 3.037. When converted to the standardized SUM score, DPC 3 and 4 scored a Six Sigma level of 3.0, indicating a 93.32% quality rate and allowable defects per million opportunities of 66,807 errors. Given that DPC 3 and 4 received the same standardized metric score, this suggests that there is no significant difference between the two systems. However, it is noted that DPC 3 received a higher SUM process sigma, implying that the pattern performs best out of all the prototypes created.

The test collected the click interactions alongside the participants' task times. Variance was observed for DPC 5 and 4, where DPC 5 showed a higher mean than DPC 4. Given this, only the content varied between DPC 5 and 4, where DPC 4 was transformed

from the grid of cards to boxes. Although content doesn't have any significant variance among the design pattern combinations, this can still influence how the users see the design pattern as they navigate through the page through the screen dynamics scrolling. As stated in the study of [11], scrolling would ask the user to go through different portions of the website compared to a page mode screen dynamic that would display the contents of the webpage all at once (Table 2).

Table 2. Summary of Results for Each Design Pattern Component.

Design Element		DPC 1	DPC 2	DPC 3	DPC 4	DPC 5
Page Layout	Efficiency	-	-	Best	Worst	-
	Effectiveness	Equal	Equal	Equal	Equal	Equal
	Satisfaction	4th	5th	3rd	1st	2nd
Content	Efficiency	-	-	Best	Worst	-
	Effectiveness	Equal	Equal	Equal	Equal	Equal
	Satisfaction	4th	5th	1st	2nd	3rd
Menu	Efficiency	-	-	Best	Worst	-
	Effectiveness	Equal	Equal	Equal	Equal	Equal
	Satisfaction	4th	5th	3rd	1st	2nd
Screen Dynamics	Efficiency		Worst	Best	-	-
	Effectiveness	-	-	-	Best	Worst
	Satisfaction	4th	5th	1st	2nd	3rd
Header	Efficiency	-	-	-	Worst	Best
	Effectiveness	Equal	Equal	Equal	Equal	Equal
	Satisfaction	4th	5th	2nd	3rd	1st
Footer	Efficiency	-	-	Best	Worst	-
	Effectiveness	Equal	Equal	Equal	Equal	Equal
	Satisfaction	4th	5th	1st	2nd	3rd

** 1 being the highest and 5 being the lowest

4.2 Affective Analysis

The 14 emotions presented during the prototype testing were clustered into positive or negative emotions. Although 14 emotions are presented using the tool PREMO, the main purpose of this is because emotions can't simply be displayed as words of "happy" or "sad" and at the same time can't just be visually represented between two faces [12]. Therefore, as a tool, PREMO allows the user to choose between different visualizations of positive and negative emotions. In the succeeding part of this section, the negative and positive emotions were tested under one-way analysis of variance against each combination of design p to determine which prototype under a specific task gained the highest rating of positive or negative emotions.

The single menu has a shorter task time and a higher positive rating than the other menu design elements, contrary to the NSGA emotion rating, which favored the fixed sticky menu. DPC 2 and 3 have low and high means for on-screen dynamics and scrolling. Task time can also be compared, with DPC 2 having a larger average task duration than DPC 2. Shorter work times result in higher positive ratings. Finally, DPC 2 has a low mean for both the header and footer, while DPC 5 has a high header mean, and DPC 3 has a high footer mean. Since headers and footers are constants in every DPC, users rate them neutrally. This task's variance may be affected by other design patterns. DPC 3 has the greatest positive feeling rating of all 6 analyzed design patterns, including the original, with 3 design elements.

5 Conclusion and Future Perspectives

Transforming web page design by including considerations for the user's emotional aspect is the future of effective web page design. Quantifying user emotion is a multiple-step process that requires testing and considering the webpage's context to implement the appropriate available design pattern. Future studies can explore design patterns that incite negative emotions and consider testing these on other websites.

References

1. Flavian, C., Gurrea, R., Orús, C.: Web design: A key factor for the website success. *J. Syst. Inf. Technol.* **11**(2), 168–184 (2009). <https://doi.org/10.1108/13287260910955129>
2. Garrett, R., Chiu, J., Zhang, L., Young, S.D.: A literature review: website design and user engagement. *Online J. Commun. Media Technol.* **6**(3) (2016). <https://doi.org/10.29333/ojcm/2556>
3. Sutcliffe, A., Namoun, A.: Predicting user attention in complex web pages. *Behav. Inf. Technol.* **31**(7), 679–695 (2012). <https://doi.org/10.1080/0144929x.2012.692101>
4. Wang, M.-Y., Tang, D.-L.: Waterfall flow vs. fixed grid webpage layout design—The Effects depend on the Zhong-Yong thinking style. *Cross-Cult. Des.*, 94–103 (2017). https://doi.org/10.1007/978-3-319-57931-3_8
5. Chen, M.: Reducing web page complexity to facilitate effective user navigation. *IEEE Trans. Knowl. Data Eng.* **32**(4), 739–753 (2020). <https://doi.org/10.1109/tkde.2019.2893242>
6. Noriega, P., Pinto, T., Assunção, J., Lourenço, F., Ungureanu, C., Ribeiro, E.: Web press carousel and his effects over news memorization. *Adv. Ergon. Des.*, 53–61 (2016). https://doi.org/10.1007/978-3-319-41983-1_6
7. Comeaux, D.J.: Web design trends in academic libraries—A longitudinal study. *J. Web Librariansh.* **11**(1), 1–15 (2016). <https://doi.org/10.1080/19322909.2016.1230031>
8. Dost, J.M., et al.: Site in a box: Improving the Tier 3 experience. *J. phy. Conf. Ser.* **898**, 082050–082050 (2017). <https://doi.org/10.1088/1742-6596/898/8/082050>
9. B. Oliveira and C. Teixeira Lopes, “The Evolution of Web Search User Interfaces--An Archaeological ” <https://arxiv.org/abs/2301.08613> (2024.) (Accessed 19 Apr 2024)
10. Hennessy, M.J.: Investigating the effects of header display formats on reading webpages. *Scholarship@Western* (2014) https://ir.lib.uwo.ca/psychd_uht/23/ (Accessed 19 Apr 2024)
11. Gao, Z., Huang, J.: Human-computer interaction emotional design and innovative cultural and creative product design. *Front. Psychol.* **13** (2022). <https://doi.org/10.3389/fpsyg.2022.982303>
12. Katz, E., Blumler, J.G., Gurevitch, M.: Uses and gratifications research. *Public Opin. Q.* **37**(4), 509 (1973). <https://doi.org/10.1086/268109>



Novel GUI for the Biology Sector Based on HCD Approach

Alessia Brischetto^(✉)  and Mattia Pistolesi 

Department of Architecture DIDA, University of Florence, Ergonomics and Design Lab.,
Sandro Pertini 93 street, Calenzano, Florence, Italy
alessia.brischetto@unifi.it

Abstract. In the paper, the authors describe the evaluation and design phases that led to the definition of a novel GUI (Graphic User Interface) designed for a medical device in biology laboratories, i.e. an analytical device for diagnostic investigations that use the serum agglutination technique on a microplate. In particular, some methods of the Human-Centred Design (HCD) approach were used, which made it possible to evaluate the existing state of the interface to be redesigned (phase 1), define the design requirements of the new GUI (phase 2), and finally evaluate the usability levels of the new interface thanks to a usability test session which involved 17 participants with specific training in the biological field (phase 3). The results highlighted the methodology's effectiveness and possible replicability for other products and graphic interfaces to be used in the medical field and beyond. Furthermore, thanks to the methodological approach adopted, it was possible to identify the main criticalities of the end users, the areas of intervention, and the relevant present and future design solutions. Fundamentally, the project of medical devices implies the knowledge of human factors, which focus on the interaction between human beings and the system, as an element to minimize the risks of use and ensure a safer and usable medical device.

Keywords: Human-Centred Design · Ergonomics in Design · Usability · Medical Device · Graphic User Interface

1 Introduction

The ergonomic approach to the project and the usability evaluation and design methods of Human-Centred Design (HCD) are particularly relevant in the healthcare sector, where the focus is on creating human-centred and user-friendly solutions. Inappropriate medical devices (MDs) design puts manufacturers and users at risk. The consequences for the manufacturer are loss of business, legal liability and litigation, while for users, it ranges from lack of treatment efficacy to death. For end users, a well-designed MDs can increase ease of use, reduce frustration when interacting with it, increase treatment efficiency, and reduce the cost of training, the number of incidents and litigation expenses. For manufacturers, the design of fit-for-purpose MDs can increase sales, market margins, product reliability and corporate image, reduce product complaints and returns and reduce or eliminate litigation [1, 2].

The potential offered by modern technologies is changing how humans and machines interact. In order to reduce the probability of error and improve the usability and user experience of products-services-systems intended for healthcare, attention must be paid to the human factor.

HCD is a problem-solving approach that bases its effectiveness on understanding users' needs, preferences and limitations and then proposes new requirements that can be applied to the design. Designers collaborate closely with users at every stage of the process to ensure that the final product is understandable, practical, attractive and innovative [3, 4].

For these reasons, the authors applied some methodologies of the HCD approach, which allowed them to define the requirements for a new Graphical User Interface (GUI) to be used on a medical device for biology laboratories, necessary for the standardisation of microplate serum agglutination tests. The article's authors also dealt with the formal development of the new Autodat medical device.

The article describes only the phases that involved the development of the new GUI.

2 Methodology

The article describes the phases that led to the definition of the new GUI, which is the approach used in another case study [5]. Following the logic of the Double Diamond process [6], the objectives of the research were the evaluation of the existing GUI and the design of the new GUI through co-designing with some figures of the company Dienes Diagnostica Senese.

In particular, phase 1 involved the application of the Hierarchical Task Analysis (HTA) [7, 8] and heuristic evaluation [8] of the existing GUI and a test session with the medical device. This phase allowed the researchers to study the human-computer interaction mode and, in detail, the logical-semantic correlation between what is displayed on the screen and the inner workings of the medical device.

Phase 2 involved the article's authors in the preliminary definition of the new GUI. In this phase, the company's engineers and IT developers discussed, implemented, and adapted the versions produced.

Finally, phase 3 involved a usability testing session of the new GUI, involving several company representatives of varying ages and professional expertise who were not directly involved in the project.

To quantify the Usability and UX levels of the GUI, and considering the data that emerged in phase 1, 4 strategic tasks were identified, (i) including microplate programming, (ii) type of test to be performed, (iii) analysis of preliminary results and (iv) confirmation of results. Usability testing is probably one of the most common methods to assess the usability of a product [9]. Some authors claim that around 80% of usability problems can be detected with 3-5 users [10–12], although a more significant number of sample users guarantees more reliable results. 17 company employees participated in the study (female $n = 12$, male $n = 5$), recruited thanks to the support of some company representatives and perfectly in line with the inclusion and exclusion criteria expressed in the study protocol, below:

- inclusion criteria: age ≥ 18 years; any gender, educational and religious identity; and finally, not having been directly involved in the project of the new Autodat medical device.
- exclusion criteria: direct involvement during the development of the new Autodat medical device.

The usability test was organized as follows: (i) introduction to the usability test, (ii) presentation of the tool, (iii) usability test with the tool, (iv) compilation of a NASA-TLX questionnaire, a multidimensional assessment tool for the subjective workload composed of 6 values (mental demand, physical demand, temporal demand, performance, effort and frustration level) [13, 14]. It was applied to quantify the usability of the new GUI submitted to the operators using the Google Forms platform and, finally, (v) a semi-structured interview [8] to investigate some critical issues experienced by the operators during the usability test.

The results that emerged enabled the research team to plan subsequent implementations and adjustments to the GUI.

3 First Section

Below are the most relevant results.

3.1 Phase 1: Heuristic Evaluation and Hierarchical Task Analysis

The heuristic evaluation, conducted on the previous interface, allowed the authors to study how users used the product, the number of steps they took to achieve a goal and to quantify possible usage errors or gaps to be resolved, such as reduced screen size, the graphic quality of the interface, the arrangement and hierarchization of the information on the screen.

In comparison, the HTA allowed us to analyze and understand all the tasks an operator must perform when interacting with the new GUI. The HTA defined the hierarchy of information to be shown on the video and determined the sequence and logic of how the steps should be performed to use the tool correctly. This allowed us to define a complete and useful flow diagram for identifying potential errors and semantic gaps and defining new project requirements addressed in phase 2.

3.2 Phase 2: Preliminary Definition of the New GUI

Thanks to the web application for interface design and prototyping Figma, it was possible to co-design the new GUI of the new medical device together with the software developers.

This multidisciplinary approach has made it possible to connect the world of biomedical engineering with the world of design and to define an interface based on the needs of end users, i.e. those who will have to use the new product within specific work contexts such as biotech sector, with effectiveness, efficiency and satisfaction of use.

Thanks to this approach, testing some parts of the GUI through operational simulations throughout the conception and design process was possible.

3.3 Phase 3: Usability Test

A usability test session was organized with some significant figures from the Diesse Diagnostica Senese company to quantify the usability levels of the GUI. A session was organized simulating the type of medical device within a controlled environment, with biologists using the instrument for the first time. Therefore, the results emerging from the NASA-TLX questionnaire and the subsequent semi-structured interview indicate a positive reliability of the interface, even for those who find themselves using it for the first time. More specifically, regarding task 1 *microplate programming* (select the test to be performed and insert the sequence and hierarchy of programming the samples to be processed into the plate preview), the results demonstrate that 8 out of 17 users (4 users = value 6; 2 users = value 7; 2 users = value 8) experienced a high “mental demand” on a value scale from 1 to 10 where 1 represents a low value, therefore positive, and 10 a value and therefore negative. As regards the “performance” value, 16 users out of 17 (4 users = value 6; 5 users = value 7; 4 users = value 8; 1 user = value 9; 2 users = value 10) experienced high and therefore negative values. This condition may be motivated by the fact that a good part of the participants did not notice and did not recognize the meaning of the icons on the screen, such as “run”, “host”, “next”, and “test selection”.

If we refer to task 2, *the type of test to be performed* (select the test and the relative reagent), the results are very positive. Only for the “performance” value do any difficulties emerge declared by the group involved in the study. 2 participants declared a value equal to 6, 6 users a value equal to 7, 4 users a value equal to 8 and 3 users a value equal to 10. The high values may have been influenced by the fact that some users did not notice and did not recognize the meaning of the “sample selection” and “test selection” icons.

If we refer to task 3, *analysis of preliminary results* (control and validation of preliminary results before confirmation), the data, although low and therefore positive, still show some difficulties experienced by users. As regards the “mental request”, 2 users declare that they experience a value equal to 6, 2 users a value equal to 7 and 2 users a value equal to 8. For the “performance” value, the declared data are high and, therefore, negative. 1 user declared a value equal to 6, 6 users a value equal to 7, 5 users a value equal to 8, 1 user a value equal to 9 and finally 1 user a value equal to 10. As for the other two tasks, also in this case, the high values may have been conditioned by the fact that a good part of the participants did not notice and did not recognize the meaning of the icons on the screen, such as “confirm all”, “manual check positive” and “manual check all”.

Finally, task 4, *confirm results* focused on confirming the results that emerged from the test. The data that emerged are very similar to tasks 1 and 3. The data concerning the “mental request” indicate that 1 user declared a value equal to 6, 2 users a value equal to 7, 1 user a value equal to 8 and 1 user a value equal to 9. As regards the “performance” value, 13 users out of 17 (1 user = value 6; 4 users = value 7; 6 users = value 8; 1 user = value 9; 1 user = value 10) experienced high and therefore negative values. The major problems that emerged from the sample involved in the study underline that a part of the sample did not notice and did not recognize the outline of the result, the “sample result”, “view result images”, and “add notes” icons (Fig. 1).

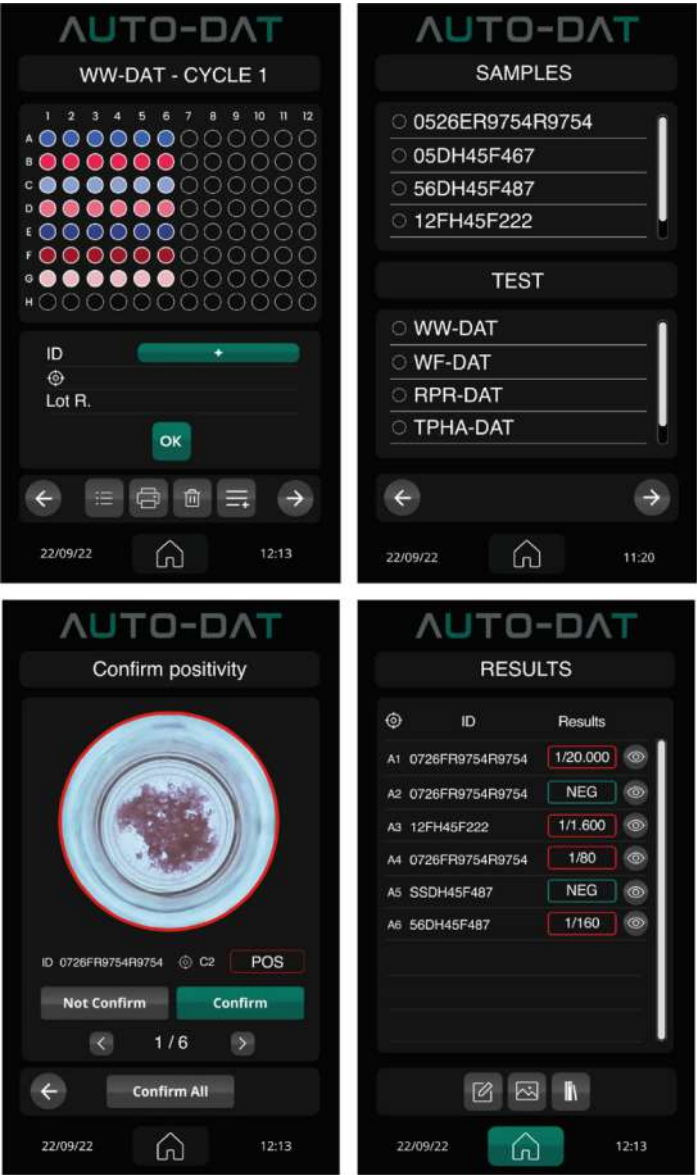


Fig. 1. While the proposed description is accurate, it is overly detailed and partially redundant with the discussion provided in the text. The caption’s purpose is to offer a concise explanation of the new GUI, highlighting its structure and the four main tasks evaluated during usability testing.

4 Discussion

Although the results indicate high usability of the system, they underline future implementations of both the GUI and the tool currently being studied by the authors of the article, including (i) type of screen, (ii) tilt and resistance capacity of screen tilt, (iii) taxonomy of icons and information on the screen.

Furthermore, the study suggests the importance of usability testing as an integral and systemic part of the development process (conception-prototyping-validation), for example, of a medical device. Very often, due to the tight timeframe for launching a product on the market, the usability of a product is determined empirically. The study demonstrates how fundamental the opinion of a small group of possible end users is to identify possible modifications or additions before the product is launched on the market.

This research has limitations. Although scientifically valid, the sample cannot be considered representative of the biotech sector. Further studies will need to be planned and conducted.

Finally, the study highlighted the effectiveness of the methodology and its possible replicability for other products and graphic interfaces to be used in the medical field.

5 Conclusion

The article describes how the authors used some methods of the HCD approach specifically to evaluate and design the usability and User Experience (UX) of a new GUI for the organic sector.

User interface (UI) validation is an iterative process that requires ongoing user involvement and various techniques to ensure the interface is effective, efficient, and satisfactory. Integrating validation into all design phases allows you to identify and resolve problems early, improving the overall quality of the final product. The setting of the study allowed us to determine critical issues and difficulties, as well as collect reflections and suggestions from a segment of users representative of the possible end users of the product. The ergonomic approach and HCD methods represent a strategic factor, especially in a complex sector such as that of medical devices. They can improve the quality and usability of products, allowing the identification of future scenarios and solutions that meet human needs and expectations. Therefore, the design of medical devices must involve knowledge of human factors, which focus on the interaction between the human being and the system, to minimize the risks of use and ensure a safer and more efficient medical device. Usable, even by those preparing to use a system for the first time.

Acknowledgments. The authors would like to thank Diesse Diagnostica Senese which financed research, Alessandra Brogi, Michele Norelli, Thomas Massari, all participants of the company and Professor Francesca Tosi is scientific director of the research.

References

1. Bias, R.G., Mayhew, D.J.: Cost-Justifying Usability. An update for an Internet Age. San Francisco. Morgan Kaufmann publications (2005)
2. Weinger, M.B., Wiklund, M., Gardner-Bonneau, D.: Human factors in medical device design: A handbook for designer. CRC Press, Taylor and Francis group, Boca Raton (2011)
3. Norman, D.: The Design of Everyday Things: Revised and, Expanded edn. Basic Books, New York (2013)
4. Ku, B., Lupton, E.: Health design thinking: creating products and services for better health. MIT Press, Cambridge (2022)
5. Brischetto, A., Pistolesi, M., Fedele, G., Tosi, F.: UX Evaluation of a New Rowing Ergometer: The Case Study of the Technogym “SkillRow”. In: Ahram, T., Falcão, C. (eds.) Advances in Usability, User Experience and Assistive Technology. AHFE 2018. Advances in Intelligent Systems and Computing, vol. 794. Springer, Cham (2019)
6. Design Council., <https://www.designcouncil.org.uk/our-resources/framework-for-innovation/> (2024). Accessed 21 May 2024
7. Stanton, N.A., Young, M.S., Harvey, C.: Guide to methodology in ergonomics. Taylor and Francis, Oxford (2014)
8. Tosi, F., Brischetto, A., Pistolesi, M.: Human-centred design—user experience: Tools and intervention methods. In: Design for Ergonomics. Springer Series in Design and Innovation, vol. 2. Springer, Cham (2020)
9. Rubin, J., Chisnell, D.: Handbook of usability testing: How to plan, Design and Conduct Effective Tests. John Wiley & Sons Inc., Hoboken, NJ (2008)
10. Nielsen J., <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> (2024), Accessed 21 May 2024
11. Preece, J., Rogers, Y., Sharp, H.: Interaction design: Beyond human-computer interaction. John Wiley & Sons, New York (2002)
12. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human-Computer Interaction, 6th edn. Wiley, Hoboken (2023)
13. NASA: Task Load Index (TLX): Paper and Pencil Version. NASA—Ames Research Center, Aerospace Human Factors Research Division, Moffett Field CA (1986)
14. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in Psychology, vol. 52, pp. 139–183, North-Holland (1998)



Enhancing Comprehension and Trust in Privacy Policies: A User-Centric Design Approach

Raymond Christian M. Coronel II, Jose Antonio T. Conde, Caila Jean C. Yap,
Angelo Gabriel A. Reyes, Ezekiel L. Bernardo^(✉) , and Jazmin Tangsoc

Industrial and Systems Engineering Department, Gokongwei College of Engineering, De La
Salle University, Manila, Philippines
ezeziel.bernardo@dlsu.edu.ph

Abstract. In the digital age, the surge in data collection necessitates robust privacy policies to safeguard users' rights. However, existing policies often lack clarity and user-centric design, hindering comprehension and trust. This study investigates the impact of varying privacy policy designs on comprehension and trust. A sample size of 102 participants engaged in an experimental setup evaluating different design factors. Results show the efficacy of blocking control and interactive prompts in enhancing comprehension, while dynamic consent fosters higher trust levels. Incentives did not significantly moderate comprehension outcomes. Findings underscore the importance of clear communication and user-centric design in privacy policies to promote trust and empower users in navigating digital platforms. This study contributes insights for crafting transparent and user-friendly data practices, crucial for fostering a privacy-conscious digital landscape.

Keywords: Privacy policies · trust · comprehension · digital privacy · user-centric design

1 Background of the Study

In the era dominated by the digital revolution, the relentless surge in data collection and processing by tech giants stands out as a defining feature (Romansky, 2019). Just think about it: in 2020 alone, the internet churned out a mind-boggling 64.2 zettabytes of data, and experts predict this number will skyrocket to 180 zettabytes by 2025 (Inghirami, 2021). But with this data gold rush comes a pressing concern: privacy.

Privacy policies aren't just legal jargon; they're the guardians of our online privacy, building trust, ensuring legal compliance, and giving individuals a say in the digital realm (Evans et al., 2022). However, these policies often fail to effectively inform users due to technical jargon and poor presentation, leaving gaps in user understanding.

Now, why do people bother reading privacy policies in the first place? It's not just about dodging data disasters; it's also about feeling in control and informed (Yee & Braver, 2018). Amidst these considerations, another layer is the motivation driving user engagement with privacy policies. External incentives, such as avoiding data misuse, and internal incentives, such as feeling informed and secure, play pivotal roles (Yee & Braver,

2018). Yet, the intricate interplay between motivation and cognition in this context warrants further exploration. This sets the stage for our inquiry: how can privacy policies be designed to enhance comprehension and trust? Our paper aims to address this question by:

- Examining the impact of different privacy policy designs on comprehension and trust.
- Unraveling the relationship between comprehension and trust.
- Understanding the moderating effect of motivation on this relationship.

Understanding the nuanced relationship between comprehension, trust, and privacy policies not only empowers users to make more informed decisions regarding their personal data but also equips service providers with valuable insights for crafting policies that foster trust and confidence among their user-base. This research contributes to the ongoing discourse on data privacy and user empowerment in a digital world.

2 Review of Related Literature

2.1 Privacy Policy Statements and Design

Privacy policies explain data collection, processing, sharing (van Ooijen & Vrabec, 2019), and users' data control rights (Makhdoom et al., 2019). Barth et al. (2022) advocate for user-centric privacy policies, using visualizations to improve comprehension and engagement. This research emphasizes ongoing evolution and experimentation to enhance privacy policy effectiveness, building on Schaub et al.'s (2015) work. The study addresses usability issues by proposing a structured design space for transparency and trust through better privacy disclosures, using user-centered and participatory design methods. Key design components include timing, modality, control, and accessibility to enhance user comprehension and engagement.

Privacy policies inform users about data practices, fostering trust and understanding (Liu et al., 2021). Design considerations like timing, modality, control, and accessibility optimize privacy policy communication (Schaub et al., 2015). Timing strategies provide information at setup, just-in-time notifications, context-dependent notices, periodic reminders, persistent indicators, and on-demand access. Modality diversification uses visual, auditory, haptic, and machine-readable formats to improve engagement and comprehension. Controls, including blocking and non-blocking approaches, empower users to manage privacy preferences. Accessibility ensures users can easily access privacy policies, promoting transparency and user empowerment.

Incorporating these design components into privacy policies enhances user understanding and control over data privacy, fostering trust in service providers. This approach underscores the importance of user-focused privacy policies, contributing to a more transparent and privacy-conscious digital landscape.

2.2 Motivation Through Incentives

Wang et al. (2020) examined the impact of incentives on reading behavior and motivation, building on Kim & Bak (2020), which tied reading motivations to behavior. Wang et al.'s (2020) findings showed that external incentives, beyond monetary rewards, significantly influence reading motivation. For instance, students are more motivated to read for good grades or peer recognition. This highlights the role of intrinsic (personal satisfaction) and extrinsic (external recognition) incentives in driving reading behavior, with implications for educational strategies.

Locher et al. (2019) also found that intrinsic motivation, like enjoyment, combined with extrinsic motivations, such as good grades, improves reading outcomes. Understanding both intrinsic and extrinsic motivations is crucial for influencing behavior and fostering positive attitudes toward digital platforms. This study proposes the Privacy Policy Design Framework (PPDeF) to address these insights (Fig. 1).

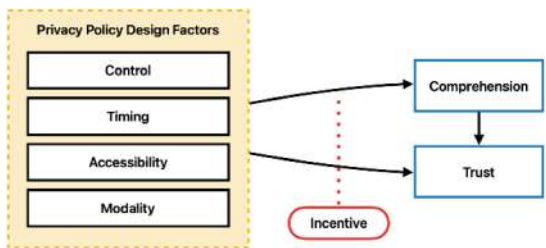


Fig. 1. Privacy Policy Design Framework (PPDeF)

3 Methodology

The methodology involved a series of steps. Firstly, the top design factors influencing social media platforms were determined. An initial survey was then conducted to gauge consumer preferences regarding social media design. The four Privacy Policy Design Factors were binary categorical data, each composed of two levels that determined the design for the respective factor. Utilizing Design of Experiment, 12 combinations for privacy policy designs were identified (four were omitted due to design incompatibility). Subsequently, privacy policy design combinations were created from the variations of the four Privacy Policy Design Factors (refer to Fig. 2 for a sample).

The sample size of the population was determined through power analysis. Pilot testing was conducted to validate the results. Following validation, the pilot test was revised in preparation for the proper experiment. Structural equation modeling was employed for analysis based on predefined objectives. The results were interpreted to generate a proposed framework, which underwent validation analysis. Finally, proposed privacy policy design guidelines were formulated based on the findings.

4 Results and Analysis

The data-gathering phase spanned 21 days with all sessions adhering successfully to the planned methodology, maintaining a cumulative duration of 20 minutes per participant. No issues were raised during the preliminary stages, and ethical concerns were effectively addressed during the onboarding process. A total of 102 samples were gathered, exhibiting no skewness and kurtosis anomalies. Exploratory factor analysis demonstrated that the questionnaire measuring trust loaded significantly, without any reliability, consistency, or validity issues. Furthermore, through 1000 bootstrapped structural equation model runs, a good model fit (Table 1).

The study found that blocking control design positively impacts comprehension and learning outcomes Haagsman et al. (2020). Interactive prompts also improve comprehension, though increased blocking control may limit user actions and hinder error recovery Yuan & Lau (2018). While altering input presentation timing didn’t affect comprehension, the visual text scroll-through design showed promise. Accessibility design changes didn’t significantly enhance comprehension. Regarding trust, control design might not inherently increase trust, but dynamic consent on privacy policies does (Petrescu & Krishen, 2020). Timing and modality design didn’t consistently affect trust. There’s



Fig. 2. Sample of testbed application created for the experiment with (a) final screen, (b) prompt type accessibility, and (c) visual text, scroll thru with video modality

Table 1. Structural relationship analysis

From	To	Std. Est.	p-Value	Direct Effect?
Control	Comprehension	0.413	< 0.001	Yes
Timing		0.020	0.349	No
Accessibility		−0.026	0.299	No
Modality		0.358	< 0.001	Yes

(continued)

Table 1. (continued)

From	To	Std. Est.	<i>p</i> -Value	Direct Effect?
Control	Trust	0.026	0.907	No
Timing		0.021	0.748	No
Accessibility		0.099	0.212	No
Modality		−0.493	0.005	Yes
Comprehension		0.687	< 0.001	Yes

a positive correlation between comprehension and trust, highlighting the importance of clear communication in building trust. User characteristics recorded such as gender, age, and educational attainment, alongside prior knowledge to privacy may have affected the comprehension and trust of privacy policy design, however it is noted that these characteristics were not utilized in the analysis (Table 2).

Table 2. Global multi-group moderation results

From	DF	Std. Est.	<i>p</i> -Value	Moderated?
Incentive	31	32.344	0.400	No

The hypothesis that incentives moderate the relationship between privacy policy design and comprehension was not supported. Monetary incentives had no direct impact on comprehension, and the moderation model lacked statistical significance. Although incentives increase survey participation, they do not significantly affect comprehension. These findings suggest that while incentives boost participation, their effect on comprehension is unclear, warranting further investigation into how different incentives influence comprehension in various contexts.

5 Conclusion and Recommendation

This study highlights the importance of privacy policies in the digital age and examines how their design affects user comprehension and trust. Our findings suggest that certain design elements, like blocking controls and interactive prompts, can enhance comprehension and trust. However, the relationship between comprehension and trust is complex, and incentives may not directly impact comprehension levels. These results emphasize the ongoing need for research to improve privacy policy design and enhance user empowerment and trust in digital environments.

References

Barth, S., Ionita, D., Hartel, P.H.: Understanding Online Privacy—A Systematic review of privacy visualizations and privacy by design guidelines. *ACM Comput. Surv.* **55**(3), 1–37 (2022). <https://doi.org/10.1145/3502288>

- Evans, B., Frumkin, L., Coopamootoo, K., Jesus, V., Little, S.: Digital technologies, power and control. In: *A Review of How Organisations Can Empower Individuals and Communities to Develop Trust, Uphold Their Privacy and Curate Their Identity in a Secure Digital Environment*. [Spritehub.org](https://oro.open.ac.uk/93746/) (2022). <https://oro.open.ac.uk/93746/>
- Haagsman, M.E., Scager, K., Boonstra, J., et al.: Pop-up questions within educational videos: effects on students' learning. *J. Sci. Educ. Technol.* **29**, 713–724 (2020). <https://doi.org/10.1007/s10956-020-09847-3>
- Inghirami, I.: Taming Corporate Data. ITAIS 2021 Proceedings, vol. 4. Retrieved from <https://aisel.aisnet.org/itaish2021/4>, (2021)
- Kim, D.H., Bak, H.-J.: Reconciliation between monetary incentives and motivation crowding-out: the influence of perceptions of incentives on research performance. *Public Perform. Manag. Rev.* **43**, 1–26 (2020). <https://doi.org/10.1080/15309576.2020.1771387>
- Liu, B., Pavlou, P.A., Cheng, X.: Achieving a balance between privacy protection and data collection: a field experimental examination of a theory-driven information technology solution. *Inf. Syst. Res.* **33**(1) (2021). <https://doi.org/10.1287/isre.2021.1045>
- Locher, F.M., Becker, S., Pfost, M.: The relation between students' intrinsic reading motivation and book reading in recreational and school contexts. *AERA Open*. **5**(2) (2019). <https://doi.org/10.1177/2332858419852041>
- Makhdoom, I., Zhou, I., Abolhasan, M., Lipman, J., Ni, W.: PrivySharing: A Blockchain-Based Framework for Privacy-Preserving and Secure Data Sharing in Smart Cities. *Comput. Secur.* **88**, 101653 (2019). <https://doi.org/10.1016/j.cose.2019.101653>
- Petrescu, M., Krishen, A.S.: The dilemma of social media algorithms and analytics. *J. Mark. Anal.* **8**(4), 187–188 (2020). <https://doi.org/10.1057/s41270-020-00094-4>
- Romansky, R.: A survey of informatization and privacy in the digital age and basic principles of the new regulation. *Int. J. Inf. Technol. Secur.* **1**(11), 95–106 (2019)
- Schaub, F., Balebako, R., Durity, A.L., Cranor, L.F.: *A Design Space for Effective Privacy Notices*, pp. 365–393. Cambridge University Press eBooks (2015). <https://doi.org/10.1017/9781316831960.021>
- van Ooijen, I., Vrabec, H.U.: Does the GDPR Enhance consumers' control over personal data? An analysis from a behavioural perspective. *J. Consum. Policy.* **42**(1), 91–107 (2019). <https://doi.org/10.1007/s10603-018-9399-7>
- Wang, X., Jia, L., Jin, Y.: Reading amount and reading strategy as mediators of the effects of intrinsic and extrinsic reading motivation on reading achievement. *Front. Psychol.* **11** (2020). <https://doi.org/10.3389/fpsyg.2020.586346>
- Wu, K.W., Huang, S.Y., Yen, D.C., Popova, I.: The effect of online privacy policy on consumer privacy concern and trust. *Comput. Hum. Behav.* **28**(3), 889–897 (2012)
- Yee, D.M., Braver, T.S.: Interactions of motivation and cognitive control. *Curr. Opin. Behav. Sci.* **19**, 83–90 (2018). <https://doi.org/10.1016/j.cobeha.2017.11.009>
- Yuan, X., Lau, A.: The effects of blocking control on user comprehension and error recovery. *Int. J. Hum.-Comput. Stud.* **111**, 1–15 (2018). <https://doi.org/10.1016/j.ijhcs.2017.10.011>



Integrating Cognitive Ergonomics and UI Enhancements for Efficient Appointments and Elevated Patient Satisfaction in Metro Manila Hospitals

Justine Bless Costales^(✉), Aaron Raphael Ricardo, Sheila Mae Carungay,
and Francispito Quevedo

Adamson University, Manila, Philippines
justine.bless.costales@adamson.edu.ph

Abstract. This research discusses the scarcity of medical professional and medical facilities, and the importance of utilizing and enhancing telemedicine application in seeking healthcare services for outpatients in Metro Manila. 385 outpatients 18- to 75- years old participated in this study where they expressed low satisfaction in waiting periods and technical difficulties in using the appointment system in existing telemedicine applications. The results highlight the potential advantages of including improved user interfaces and cognitive ergonomics in telemedicine systems. These apps can greatly lower cognitive burden, expedite the scheduling process, and improve patient satisfaction by becoming more intuitive and user-friendly. The study comes to the conclusion that these enhancements can give outpatients more effective access to medical treatments, improving healthcare outcomes in Metro Manila.

Keywords: Appointment System · Cognitive Ergonomics · Outpatients · Telemedicine Application · UI (User Interface)

1 Introduction

Telemedicine, a rapidly growing field that blends healthcare and technology, is utilized to overcome the difficulties the patients face due to distance and lack of specialists nearby [1]. By overcoming geographic barriers and making healthcare more convenient, telemedicine has become an essential part of modern healthcare delivery. The platform has since provided healthcare at the comfort of the patient's home.

Haleem, et al. [2] described that hospital visits can be expensive due to the traveling costs, especially in rural areas. The online healthcare system has continuously evolved since it became popular during the CoViD-19 pandemic. Online medical consultation through technological-means became more apparent due to the restriction of in-person hospital visitations. As an online healthcare system, it has its own limitations in observations and medical diagnosis [3]. However, there are often times that patients are required or feel the need to attend an in-person consultation. Hence, telemedicine developers

made innovations with the platform from being a platform of online consultation to providing other healthcare services such as booking an appointment to see a physician.

The convenience that the system provides was seen as an opportunity to improve in providing healthcare. Healthcare institutions adapted telemedicine systems as a means of communication with patients. It has extended the reach of health care providers giving access to living in rural areas [4]. Telemedicine emphasized its importance by widening its reach to underserved areas providing fair health care assistance.

2 Background

Today, the use of smartphones has become a daily necessity. The frequency of using smartphones for the older generation has an average of 67.6 min per day [5] and for the younger generation has an average of 300 min per day [6]. The lockdown that was implemented during the pandemic restricted people from going outside, especially going to hospitals. This became an opportunity to develop telemedicine applications and the rise of tele-consultations from different social media platforms.

Tele-consultations have since then provided convenience for patients to seek medical needs through online platforms. The development of telemedicine application systems has continued to improve and develop. They added other features such as booking an appointment with the physician. The pandemic had a significant impact on patients' preference on whether they choose to have their consultations in-person or online. On some cases, the patients are required to have their consultations in-person as their medical needs require in-depth check-up. Medical professionals are scarce in the region, but internet services are easily accessible and reasonably priced in the nation.

The study of Stoltzfus, et al. [7] regards the usage of telemedicine to have decreased significant factors needed for medical consultations such as cost, time, and health-care facility use. It has also provided an access to reach patients that has difficulties in travelling to hospitals and with physical disabilities receiving medical care at their convenience.

Appointment systems of outpatient face optimization problems as patient demand rises and resources become limited [8]. Staff shortages, irregular schedules, and strained resources contribute to scheduling problems, long wait times, and patient dissatisfaction. However, bridging the gap in technology and providing accessibility for all remains critical. The rising worldwide healthcare information technology presents an amazing opportunity. Increased investments in technological innovations including telemedicine, remote monitoring, and interconnected platforms can mitigate resource restrictions while also improving the quality of patient care.

However, in order for these solutions to be successful, user interface design must utilize cognitive ergonomics. Telemedicine apps may increase user engagement, improve productivity, and make way for a more efficient healthcare experience. Utilizing the power of mobile technology and embracing digital healthcare innovations are important steps toward modernizing outpatient appointment systems [9]. Hence, telemedicine developers made innovations with the platform from being a platform of online consultation to providing other healthcare services such as booking an appointment to see a physician. The telemedicine application systems continuously evolve to provide

utmost convenience to patients for their booking appointment needs. Chauhan, et al. [10] described the capacity of telemedicine in providing an easy access to receiving medical consultations, treatments, and monitoring.

3 Methods

The primary aim is to quantitatively assess and closely examine the factors in order to do research on them. An essential and precise descriptive research methodology is required to ascertain the research's findings. The objective of the study is to determine the optimal measures that should be implemented to improve hospital outpatient appointment systems.

3.1 Sampling Technique

The research was limited to outpatients of Metro Manila, ages 18 to 75 years, with a latest total population of 8,415,000. Cochran's Formula was utilized to calculate the ideal sample size for the study. With a confidence level of 95%, Z of 1.96, estimated proportion of 0.5, and precision level of 5%, the sample size was calculated to be 385 outpatients.

$$n_0 = \frac{(1.96)^2 \cdot 0.5(1 - 0.5)}{0.05^2} = 385 \quad (1)$$

3.2 Data Gathering Procedure

Researchers utilized Microsoft Forms and physical survey forms to collect data from outpatients in nearby hospitals in Metro Manila. The online and physical survey form will give data on the challenges outpatients face on their appointments during their visits to hospitals. Furthermore, the survey includes suggestions on the improvements they would want in the current systems. The study was limited to 10 common experiences and problems of outpatients with hospitals and their appointment system in Metro Manila in order to determine the solution needed to increase the efficiency of hospitals in the region.

3.3 Statistical Treatment

The researchers tested the reliability of their survey using Cronbach's Alpha. MANOVA was used to assess the statistical significance the variables. Poisson rate was used to determine the occurrences of the dependent variables. A value of $p < 0.05$ was accepted as statistically significant.

Table 1. Poisson Rate for two (2) most occurrent experiences from the survey λ : Poisson rate of “I had to wait for a long-time for my doctor’s appointment” and “It is challenging to reschedule my appointment at my convenience”. Exact Method is used for this analysis.

	N	Total Occurrences	Sample Rate	95% CI for
I had to wait for a long-time for my doctor’s appointment	385	1154	2.99740	(2.82694, 3.175466)
It is challenging to reschedule my appointment at my convenience	385	1121	2.91169	(2.74371, 3.08726)

Table 2. P-Values for “I had to wait for long-time for my doctor’s appointment.” And “It is challenging to reschedule my appointment at my convenience.”

Sample	P-Value
I had to wait for a long-time for my doctor’s appointment	0.000
It is challenging to reschedule my appointment at my convenience	0.000

4 Results and Discussion

The researchers’ survey reliability test resulted to an alpha of 0.8 for the experiences segment suggests that the questions were highly reliable and consistent.

Poisson rate comparison compares the occurrences between “I had to wait for a long time.” And “It is challenging to reschedule my appointment at my convenience.” Table 1 indicates that the two variables are statistically consistent with the data gathered by the researchers. P-values of 0.000 for both variables indicate that there is significant relationship between the dependent and independent variables (Table 2).

Table 3. MANOVA Test for I had to wait for a long-time for my doctor’s appointment.

	Test	DF				
Criterion	Statistic		Approx F	Num	Denom	P
Wilk’s	0.69731		5.877	24	1295	0.000
Lawley-Hotelling	0.39627		6.101	24	1478	0.000
Pillai’s	0.32976		5.600	24	1496	0.000
Roy’s	0.27595					
$S = 4$	$m = 0.5$		$n = 184.5$			

The difference in the occurrences of the two dependent variables in Table 1 suggests that rescheduling difficulties are slightly more satisfactory than the waiting time issues.

Table 4. MANOVA Test for It is challenging to reschedule my appointment at my convenience.

	Test	DF				
Criterion	Statistic		Approx F	Num	Denom	P
Wilk's	0.69369		5.966	24	1295	0.000
Lawley-Hotelling	0.42525		6.547	24	1478	0.000
Pillai's	0.31775		5.379	24	1496	0.000
Roy's	0.38448					
$S = 4$	$m = 0.5$		$n = 184.5$			

Table 3 and Table 4 also suggest that there are significant differences between the two variables regarding waiting times and rescheduling for hospital appointments with P-Values of 0.000. All criteria in both MANOVA tests have a P-Value of 0.000 suggests that there are significant differences in the between the independent variables, patients and the hospitals, across the dependent variables “I had to wait for a long-time for my doctor’s appointment.” and “It is challenging to reschedule my appointment at my convenience.”

The statistical values of all criteria are below 1 which indicates that there is a strong relationship between the independent variables and dependent variables. The values of Approx F further support the significance of MANOVA test.

The results of the study indicate that the dependent variables, rescheduling difficulties and waiting time issues, are both significantly influenced by the independent variables. Long waiting times may be associated with the difficulty in rescheduling as suggested by the results. The findings of these analyses suggest that there could have implications on optimizing the operational efficiency of appointment system of hospitals.

5 Conclusion

The combination of increasing patient demand and restricted resources creates optimization challenges in appointment systems, which subsequently pose difficulties for hospitals providing outpatient treatment. Various reasons contribute to prolonged waiting periods, such as insufficient personnel, unpredictable schedules, and limited resources, ultimately resulting in patient dissatisfaction.

Optimizing human-system interactions in telemedicine applications requires incorporating cognitive ergonomics into their design. By making systems more instinctive and user-friendly, cognitive ergonomics seeks to lessen users’ cognitive burden and provide accessibility to technology for a wider range of patients. Telemedicine apps can be made easier to use, even for non-techies, by adding features like plain language, clear navigation, and visual assistance. Creating an interface that is in line with users’ mental models can make tasks easier to complete and enhance the user experience in general.

The suggested improvements to telemedicine applications have the potential to establish new benchmarks in healthcare delivery because they emphasize both ergonomic design and technology breakthroughs. By making healthcare more easily accessible, effective, and satisfying for outpatients, this strategy hopes to improve healthcare outcomes in Metro Manila.

References

1. Allaert, F., Legrand, N.A., Carime, N., Quantine, C.: Will applications on smartphones allow a generalization of telemedicine? *BMC Med. Inf. Decis. Making* (2020)
2. Haleem, A., Javaid, M., Singh, R.P., Suman, R.: Telemedicine for healthcare: capabilities, features, barriers, and applications. *Sens. Int.* (2021)
3. Romanick-Schmiedl, S., Raghu, G.: Telemedicine – maintaining quality during times of transition. *Nat. Rev. Dis. Primers.* **6**(1), 45 (2020)
4. Haimi, M.: The tragic paradoxical effect of telemedicine on healthcare disparities – a time for redemption: a narrative review. *BMC Med. Inf. Decis. Making* (2023)
5. Busch, P.A., Hausvik, G.I., Ropstad, O.K., Pettersen, D.: Smartphone usage among older adults. *Comput. Hum. Behav.* **121**, 5 (2021)
6. Ahmed, N.: Generation Z's smartphone and social media usage: a survey. *J. Mass Commun.* (2019)
7. Stoltzfus, M., Kaur, A., Chawla, A., Gupta, V., Anamika, F.N.U., Jain, R.: The role of telemedicine in healthcare: an overview and update. *Egypt. J. Internal Med.* (2023)
8. Ala, A., Chen, F.: Appointment scheduling problem in complexity systems of the healthcare services: a comprehensive review. *J. Healthc. Eng.* **2022**, 5819813 (2022)
9. Abdalkareem, Z.A., Amir, A., Al-Betar, M.A., Ekhan, P., Hammouri, A.I.: Healthcare scheduling in optimization context: a review. *Health Technol.* (2021)
10. Chauhan, P., Bali, A., Kaur, S.: Breaking barriers for accessible health programs: the role of telemedicine in a global healthcare transformation (2024)



Explanatory Modelling of Factors Affecting the Adoption of Self-service Technology

Jaeyeon Park, Giannina Anais Rodas, Josh Adrian Dyaico, Cedric Nathaniel Yu, Ezekiel L. Bernardo^(✉) , and Jazmin C. Tangsoc

Industrial and Systems Engineering Department, Gokongwei College of Engineering,
De La Salle University, Manila, Philippines
ezeziel.bernardo@dlsu.edu.ph

Abstract. This paper investigates the complexities of self-service technology (SST) adoption, recognizing the pivotal role of demographics in shaping consumer attitudes and behaviors. Through a comprehensive methodology involving initial interviews, measurement tool development, and statistical analysis, a novel factorization of SST adoption dimensions is revealed, emphasizing utility, engagement, and functionality. Surprisingly, while established constructs like perceived trust remain consistent, the model explains over 77.70 percent of behavioral intention variance, with utility, engagement, functionality, and resistance to change emerging as critical indicators. Moderation analysis highlights the influence of familiarity, technological anxiety, and age, underscoring the importance of tailoring SST experiences to diverse demographic cohorts. These findings not only enrich our understanding of consumer behavior but also provide actionable recommendations for designing user-centric SST systems. Ultimately, this research contributes to the refinement of self-service technologies, ensuring their relevance and efficacy in a rapidly evolving technological landscape.

Keywords: Self-service technology · adoption framework · user-centric design · elderly

1 Background of the Study

In the dynamic landscape of today's technological evolution, the empowerment of individuals through technology stands as a defining characteristic of our modern era. From streamlining daily tasks to revolutionizing entire industries, the progress in computing has endowed machines with capabilities once exclusive to human endeavor. A quintessential example of this phenomenon is the widespread adoption of self-service technology (SST), encompassing systems or facilities enabling consumers to autonomously access services without human intervention (Liu & Hung, 2021).

Originating as early as the late 19th century with the introduction of vending machines, SST has metamorphosed into sophisticated systems like ATMs, self-service gas pumps, and various other automated service provisions (Algarawi & Khan, 2021). This evolution has been propelled by a convergence of factors including the ascent of

consumerism, rising labor costs, and an escalating demand for convenience. However, as SST continues to reshape consumer experiences, it introduces new contexts fraught with challenges and uncertainties, necessitating a closer examination of shifting consumer behaviors (Payne et al., 2018).

Amidst the surge in SST adoption, a conspicuous gap in research emerges concerning the underlying factors influencing consumer acceptance or rejection of these systems. Notably, the influence of age emerges as a critical moderating factor, with older generations potentially exhibiting distinct attitudes and behaviors towards SST compared to younger demographics (Hwang & Nam, 2017; Nunan & Di Domenico, 2019). What are the key determinants driving SST adoption, and how do they intersect with age-related factors? How can these insights inform the design of SST systems that resonate with diverse user demographics? Understanding these dynamics is pivotal not only for optimizing the design and implementation of SST but also for ensuring its relevance and efficacy across diverse user populations.

Therefore, this research endeavors to unravel the intricacies of SST adoption, with a specific focus on elucidating the role of age in shaping consumer attitudes and behaviors, as depicted in the conceptual framework in Fig. 1. By addressing this gap, the aim is to:

- Create and validate framework that organizes the idea of SST adoption.
- Provide insights that pave the way for a more user-centric approach to SST development, ultimately enhancing its acceptance across varying demographic cohorts.
- Offer actionable recommendations that foster greater synergy between technology and its users in the realm of self-service innovation.

To achieve this, the paper is structured as follows: Sect. 2 reviews definitions and past research, Sect. 3 details the methodology, Sect. 4 analyzes the findings, and Sect. 5 concludes with recommendations.

2 Review of Related Literature

To bridge the knowledge gap and meet study goals, the SSTAME (Self-Service Technology Adoption Model for the Elderly) was created. It explores factors influencing SST use intentions, with familiarity, age, and technological anxiety as moderators. Despite the increasing importance of older adults in tech adoption, their needs are often sidelined. Drawing from established models like TAM and UTAUT, this study integrates existing constructs and introduces new ones tailored to SSTs and the elderly.

2.1 Added Constructs

The model introduces essential new constructs for understanding technology adoption among the elderly. Perceived social presence, emphasizing human-like interactions in technology, is rooted in social presence theory (Biocca, 1997; Short et al., 1976; Grewal et al., 2020). Customization tailors interfaces to individual preferences, boosting adoption, particularly among older demographics (Sheehan et al., 2020; Begde et al., 2023; Elimelech et al., 2022).

Perceived convenience, encompassing availability, accessibility, and time savings, significantly influences SST ease of use (Brown, 1989; Bernardo & Tangsoc, 2019) and technology acceptance across contexts (Hong et al., 2020). Perceived trust, related to reliability and data protection, fosters user confidence (Jameel & Kareem, 2022), crucial for designing user-centric, secure SSTs (Chao, 2019). These enrichments to the SST adoption model illuminate the complex factors shaping technology acceptance in society.

2.2 Moderating Variables

The proposed moderating variables in self-service technology (SST) adoption include age, familiarity, and technological anxiety. Age significantly affects technology acceptance, with different generational cohorts showing varying levels of technological familiarity and willingness to adopt new innovations (Chung et al., 2010; Siwela et al., 2022). Familiarity, defined as prior experience or knowledge of SST usage, greatly influences users' comfort, trust, and intention to use such technologies. A lack of familiarity often hinders adoption, particularly among the elderly (Cao et al., 2022; Lee & Lyu, 2018). Technological anxiety, characterized by discomfort or fear when encountering new technology, negatively impacts adoption rates, especially among older demographics, and moderates relationships such as those between facilitating conditions and performance expectancies (Yap et al., 2022; Lee, 2010; Di Giacomo et al., 2019; Deng et al., 2014; Yang et al., 2013; Tsai et al., 2020). Understanding these moderators is crucial for designing user-centric SST systems that cater to the diverse needs and preferences of different demographic groups.

3 Methodology

The methodology involves initial interviews to understand self-service technology (SST) usage among elderly adults. A measurement tool is developed to assess key factors, followed by post-experimental validation. Sample size is determined using Slovin's formula and power analysis. After a pilot test, individuals who have used SSTs were recruited. Data is collected, prepared, and screened. Exploratory factor analysis tests coherence using Maximum Likelihood factor extraction and Promax oblique rotation, followed by confirmatory factor analysis to check for model fit. Variables that showed strong association were grouped to form new factors. Model validation is conducted, and results are interpreted and validated relative to the study objectives. Finally, a final framework and recommendations are formulated, drawing conclusions and providing recommendations for future research and practical implications.

4 Results and Analysis

The sample of 191 participants provided a solid basis for analysis, with no data normality issues. Factor analysis uncovered a new perspective, revealing dimensions like utility, formed from performance and effort expectancy, convenience, and facilitating conditions; engagement and connection, originating from social presence; functionality,

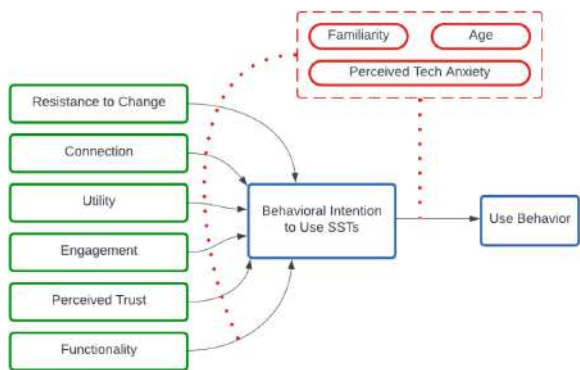


Fig. 1. Validated framework for SST Adoption

stemming from customization, offering significant insight into overlooked aspects of self-service technology (SST) adoption among the elderly. Established factors such as perceived trust and resistance to change remained significant, reinforcing their role in adoption behavior. See Fig. 1 for the validated model, named the Self-Service Technology Adoption Model for the Elderly (SSTAME) (Table 1).

Table 1. Summary of objectives based on path analysis

Direct Relationship					Moderation Effect		
From	To	β	p -value	Supported	Fam	Age	PTA
<i>Resistance</i>	Behavioral Int	-0.168	< 0.001	Yes	No	Yes	No
<i>Connection</i>		0.124	0.080	No	No	No	No
<i>Personalization</i>		0.076	0.149	No	No	No	No
<i>Utility</i>		0.423	< 0.001	Yes	Yes	No	No
<i>Engagement</i>		-0.149	0.004	Yes	Yes	No	No
<i>Trust</i>		0.151	0.003	Yes	No	Yes	No
<i>Functionality</i>		0.265	< 0.001	Yes	Yes	Yes	Yes
<i>Social Influence</i>		0.081	0.108	No	No	No	No
<i>Behavioral Int</i>	Use Behavior	0.890	< 0.001	Yes	No	No	No

Notes: Behavioral Int. = Behavioral Intention; Fam - Familiarity; PTA - Perceived Trust Anxiety; p -value evaluated at 0.05

The model explains over 77.70% of the variance in SST adoption among the elderly, emphasizing its effectiveness. The results of the hypothesized direct constructs show that SST engagement, trust, utility, functionality, and acceptance are pivotal in influencing adoption behavior. Moderation analysis revealed insights into familiarity, technological anxiety, and age. Familiarity significantly influenced utility, engagement, and

functionality's impact on behavioral intention, emphasizing prior experience's importance. Technological anxiety moderated functionality's relationship, highlighting the need to address technology apprehensions among the elderly. Age moderated perceived trust, functionality, and resistance to change, indicating age-related differences in SST perception and adoption.

5 Conclusion

In conclusion, this study delves into the complex dynamics of self-service technology (SST) adoption among the elderly, focusing on how user characteristics shape attitudes and behaviors. It introduces new dimensions such as utility, engagement, and functionality, enriching our understanding of adoption processes. Established factors like perceived trust and resistance to change remain crucial, while SST engagement, utility, and functionality emerge as key drivers of behavioral intentions. Moderation analysis highlights the roles of familiarity, technological anxiety, and age in shaping adoption behaviors. These insights offer practical implications for designing user-centric SST systems for diverse demographic groups. This research paves the way for further exploration and refinement of SST frameworks to stay relevant in the evolving technological landscape.

References

- Algarawi, F.K., Khan, N.: Integrating cloud with self service kiosk: an impact study on society evolution. *Int. J. Interact. Mob. Technol.* **15**, 186–194 (2021)
- Begde, A., et al.: Exploring factors influencing willingness of older adults to use assistive technologies: evidence from the cognitive function and ageing study II. *Inf., Commun. Soc.* 1–18 (2023)
- Bernardo, E., Tangsac, J.C.: Explanatory modelling of factors influencing adoption of smartphone shopping application. *Ind. Eng. Manage. Syst.* **18**(4), 647–657 (2019). <https://doi.org/10.7232/iems.2019.18.4.647>
- Biocca, F.: The cyborg's dilemma: Progressive embodiment in virtual environments. *J. Comput.-Mediated Commun.* **3**(2), JCMC324 (1997)
- Brown, L.G.: The strategic and tactical implications of convenience in consumer product marketing. *J. Consum. Mark.* **6**(3), 13–19 (1989)
- Cao, Z., Xiao, Q., Zhuang, W., Wang, L.: An empirical analysis of self-service technologies: mediating role of customer powerlessness. *J. Serv. Mark.* **36**(2), 129–142 (2022)
- Chao, C.M.: Factors determining the behavioral intention to use mobile learning: an application and extension of the UTAUT model. *Front. Psychol.* **10**, 1652 (2019)
- Chung, J.E., Park, N., Wang, H., Fulk, J., McLaughlin, M.: Age differences in perceptions of online community participation among non-users: an extension of the technology acceptance model. *Comput. Hum. Behav.* **26**(6), 1674–1684 (2010)
- Deng, Z., Mo, X., Liu, S.: Comparison of the middle-aged and older users' adoption of mobile health services in China. *Int. J. Med. Inform.* **83**(3), 210–224 (2014)
- Di Giacomo, D., Ranieri, J., D'Amico, M., Guerra, F., Passafiume, D.: Psychological barriers to digital living in older adults: computer anxiety as predictive mechanism for technophobia. *Behav. Sci.* **9**(9), 96 (2019)
- Elimelech, O.C., et al.: Technology use characteristics among older adults during the COVID-19 pandemic: a cross-cultural survey. *Technol. Soc.* **71**, 102080 (2022). <https://doi.org/10.1016/j.techsoc.2022.102080>

- Hong, S.J., Choi, D., Chae, J.: Exploring different airport users' service quality satisfaction between service providers and air travelers. *J. Retail. Consum. Serv.* **52**, 101917 (2020)
- Hwang, H., Nam, S.: The digital divide experienced by older consumers in smart environments. *Int. J. Consum. Stud.* **41**(5), 501–508 (2017)
- Liu, C., Hung, K.: A multilevel study on preferences for self-service technology versus human staff: insights from hotels in China. *Int. J. Hosp. Manag.* **94**, 102870 (2021)
- Lee, H.-J., Lyu, J.: Exploring factors which motivate older consumers' self-service technologies (SSTs) adoption. Taylor & Francis Online (2019). <https://www.tandfonline.com/doi/full/10.1080/09593969.2019.1575261>
- Nass, C., Steuer, J., Henriksen, L., Dryer, D.C.: Machines, social attributions, and ethopoeia: performance assessments of computers subsequent to "self-" or "other-" evaluations. *Int. J. Hum. Comput. Stud.* **40**(3), 543–559 (1994)
- Nunan, D., Di Domenico, M.: Older consumers, digital marketing, and public policy: a review and research agenda. *J. Public Policy Mark.* **38**(4), 469–483 (2019)
- Payne, L.M., Peltier, J.W., Barger, V.: Enhancing the value co-creation process: artificial intelligence and mobile banking service platforms. *J. Res. Interact. Market.* **15**, 68–85 (2021). 10.110
- Sheehan, B., Jin, H.S., Gottlieb, U.: Customer service chatbots: anthropomorphism and adoption. *J. Bus. Res.* **115**, 14–24 (2020)
- Short, J., Williams, E., Christie, B.: The social psychology of telecommunications (1976)
- Siwela, B., Ebewo, P.E., Faku, E.M.: Customer acceptance of self-service technology at fast service restaurants in Tshwane. *J. Managem. Entrepreneurship Res.* **3**(1), 26–45 (2022). <https://doi.org/10.34001/jmer.2022.6.03.1-26>
- Tsai, T.-H., Lin, W.-Y., Chang, Y., Chang, P., Lee, M.-Y.: Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. *PLoS ONE* **15**(1), e0227270–e0227270 (2020). <https://doi.org/10.1371/journal.pone.0227270>
- Venkatesh, V., Thong, J.Y., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q.* 157–178 (2012)
- Yap, Y.Y., Tan, S.H., Choon, S.W.: Elderly's intention to use technologies: a systematic literature review. *Heliyon* (2022)



Exploring the Impact of Virtual Instructor's Attributes on Cognitive Load in Health Short Videos

Yichen Guo¹, Mengli Yu^{1,2,3(✉)}, and Zhenkun Zhou⁴

¹ School of Journalism and Communication, Nankai University, Tianjin, China
mengliyu@nankai.edu.cn

² Convergence Media Research Center, Nankai University, Tianjin, China

³ Publishing Research Institute, Nankai University, Tianjin, China

⁴ School of Statistics, Capital University of Economics and Business, Beijing, China
zhenkun@cueb.edu.cn

Abstract. Amidst the widespread application of virtual agents in online content production, our study explored the impact of virtual instructors' attributes on users' cognitive load in health short videos. Through a 2 (Formal Clothing Style vs. Casual Clothing Style) × 2 (Formal Linguistic Style vs. Casual Linguistic Style) × 2 (Treatment Theme vs. Lifestyle Theme) experimental design, we found that: (1) a casual clothing style and the theme related to healthy lifestyle contributes to lower extraneous cognitive load; (2) a casual linguistic style and the theme related to healthy lifestyle induce lower intrinsic cognitive load; (3) the interaction between clothing, linguistic and thematic styles did not exhibit a significant effect. Our research expands understanding of the mechanisms by which virtual images affect user cognition in health short video environments. Stakeholders can optimize the design and delivery of health short videos, ultimately enhancing their effectiveness in engaging users in health promotion efforts.

Keyword: Virtual Instructor · Cognitive load · Health short videos

1 Introduction

The rapid advancement of Artificial Intelligence (AI) technology has paved the way for the widespread application of virtual agents (Li et al., 2016). A significant and emerging subgenre is health-related content in short video platforms. Distinct from entertainment-focused content, health short videos often rely on on-screen instructors for effective knowledge dissemination. Given the intricacies of health information, users can leverage virtual instructors that embody the professional image of health practitioners to enhance the effectiveness of information delivery.

As the users of short video platforms encompass a wide range of demographics and interests, delivering health information through these platforms requires ensuring clarity and generality. In the process of watching health short videos, viewers' cognitive load is an important concept. It refers to the cognitive resources expended by individuals

during the processing of information (Sweller, 1988). Previous research has investigated the influence of instructors' appearance and voice on cognitive load in online educational environments (Klebig et al., 2016). However, relative research is limited in several ways. Firstly, studies exploring cognitive load induced by health information have mainly focused on single text or pictures, lacking investigation into the multimodal environment. Secondly, research in multimedia learning environments has primarily concentrated on structured learning environments such as online courses, whereas health short videos offer brief, fragmented, and non-compulsory learning experiences. Thirdly, previous studies have mainly examined real instructors, with limited observations on virtual instructors. Audience perceptions and expectations of virtual entities may differ from those of real individuals, potentially altering cognitive processing. Therefore, there is a need for in-depth exploration of this new subject.

2 Literature Review

2.1 Intrinsic and Extraneous Cognitive Load in Health Short Videos

Cognitive load represents the amount of mental effort required for an individual to process information or perform a task (Sweller, 1988). It can manifest in various forms, including intrinsic cognitive load, extraneous cognitive load, and germane load. In health short videos, intrinsic cognitive load may arise from the complexity of medical concepts or procedures being explained. Extraneous cognitive load may be influenced by the visual and auditory elements of the virtual instructor. For instance, if the instructor's attire is distracting or incongruent with the content being discussed, viewers may experience an increased extraneous cognitive load as they attempt to reconcile the discrepancy. By considering these different forms and reducing extraneous load with the design of visual and auditory elements, health short video producers can enhance viewers' comprehension and engagement.

2.2 Instructors' Attributes and Cognitive Load

Research in multimedia learning has consistently highlighted the significance of visual elements in shaping cognitive processes (Castro-Alonso et al., 2021). The clothing style is one visual factor that can potentially influence learners' cognitive load by modulating attention and perception. We have chosen to categorize clothing styles into formal (professional attire typically worn by doctors) and casual (informal tops and jeans) based on existing research findings. Formal clothes may convey authority and professionalism, potentially reducing extraneous cognitive load by increasing viewers' confidence in the instructor's credibility. Conversely, casual clothes may elicit a more relaxed and approachable perception of the instructor, influencing viewers' extraneous load differently.

Previous research has also demonstrated that expressive attribute can significantly impact cognitive load, and much attention has been devoted to studying the voice of the instructor (Yuan et al., 2021). However, the stylistic aspects of content, including linguistic style, can also contribute to shaping the viewer's cognitive process (Fan et al.,

2023). Formal linguistic style, characterized by precise vocabulary and complex syntax, may increase extraneous cognitive load due to the additional mental effort required to process and comprehend the language. On the other hand, casual linguistic style may reduce cognitive load.

Furthermore, health short videos encompass a diverse range of thematic content, with two common themes being disease treatment and healthy lifestyle. Content related to disease treatment often involves complex medical concepts, procedures, and terminology, which may increase intrinsic cognitive load. Lifestyle-related content tends to focus on practical tips, behavior modifications, and preventive measures, which are more relatable and easily understood. Thus, we incorporated thematic style as an independent variable. As the combination of the above factors may also provide different viewing experiences, we try to explore the interaction between these factors affects user's cognitive load in health short videos.

3 Methodology

We employed a 2 (Formal Clothing Style vs. Casual Clothing Style) $\times$ 2 (Formal Linguistic Style vs. Casual Linguistic Style) $\times$ 2 (Treatment Theme vs. Lifestyle Theme) experimental design. In designing the experimental materials, we utilized the online AI tool *Midjourney* to generate virtual instructors with consistent appearance and background but differing clothing styles (Fig. 1). Drawing from previous literature, we crafted narration scripts in different linguistic styles (Table 1). Subsequently, using the *D-ID* website (<https://studio.d-id.com/>), we generated approximately 1-min-long health short videos based on the input static images and text content.

This experiment was conducted using a laboratory setting, where participants were presented with 8 videos in a randomized order. A total of 30 participants, comprising 12 males and 18 females, with ages ranging from 18 to 26 years (mean = 21.17), were included in the sample. To minimize the influence of prior experience, a pre-test was conducted to assess their baseline knowledge. Following each material, participants rated their cognitive load level through Likert five-point scales, covering both the extraneous and intrinsic aspects of cognitive processing during video consumption (Korbach et al, 2018). Eye movement data were recorded throughout the process. Based on the research of Krejtz et al. (2018), we selected the percentage change in pupil diameter (PCPD) as a crucial eye-tracking measurement indicator.



Fig. 1. Virtual instructors' images used as stimuli

Table 1. Comparison of the Linguistic Styles

	Formal Linguistic Style	Casual Linguistic Style
Start	Formal greetings to the audience	Informal greetings to the audience
Tone	Avoid using modal particles	Use various modal particles
Wording	Use more formal words	Use internet slang
Grammar	Standard	Relatively non-standard

4 Results

We employed analyses of variance (ANOVA) to assess the effects. Our data fulfilled the assumptions of normality, and the assumption of homogeneity of variances was met for all variables ($p > 0.05$). According to our results (Table 2), linguistic style ($F = 13.294$, $p = 0.0003$) and thematic style ($F = 10.883$, $p = 0.001$) were found to have significant effects on viewers' intrinsic cognitive load. Clothing Style ($F = 4.012$, $p = 0.046$) and thematic style ($F = 7.692$, $p = 0.006$) have a significant influence on extraneous cognitive load. Casual Linguistic style, casual clothing style and explanations involving lifestyle factors contributes to lower intrinsic or extraneous cognitive load to viewers. The interaction of these dimensions did not influence cognitive load. Although there were variations in average pupil diameter change rate between different clothing styles and linguistic styles, it's hard to establish their relationship with both intrinsic and extraneous cognitive load.

Table 2. Results of ANOVA

Independent Variables	F_1	P_1	F_2	P_2
Clothing Style	0.271	0.603	4.012	0.046*
Linguistic Style	13.294	0.000***	2.374	0.125
Thematic Style	10.883	0.001**	7.692	0.006**
Clothing Style * Linguistic Style	0.030	0.862	2.374	0.125
Clothing Style * Thematic Style	0.271	0.603	0.214	0.644
Linguistic Style * Thematic Style	0.030	0.862	0.855	0.356

Notes: (1) F_1 , P_1 : Dependent variable = Intrinsic Cognitive Load; F_2 , P_2 : Dependent variable = Extraneous Cognitive Load. (2) * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

5 Discussion

Several factors can explain the observed results. Firstly, when instructor dressed in casual clothes, viewers experience reduced external cognitive load. This phenomenon may stem from viewers' recognition of the instructor's virtual agent identity. Secondly, the lack of significant influence of the instructor's linguistic style on extraneous cognitive load suggests that viewers may prioritize visual cues over linguistic cues in health short videos because they are primarily seeking information and guidance. Surprisingly, while clothing style and thematic style significantly affected extraneous cognitive load ratings, these effects were not reflected in the average pupil diameter change rate. This discrepancy could be attributed to the condensed timeframe of short videos. Viewers are presented with a restricted amount of visual information, which may not sufficiently engage their cognitive processes to the extent where significant changes in eye-tracking behavior are discernible. Additionally, in such brief durations, the volume of essential information users is required to process often falls within their capacity. Thus, redundant, or incongruous information is unlikely to heighten the difficulty of information significantly.

Our research expands understanding of the mechanisms by which virtual images affect user cognition in short video environments. The combination of self-report measures and eye-tracking data provides novel perspectives in the realm of human-computer interaction research. Practically, stakeholders involved in selecting or creating virtual instructor avatars can optimize the design and delivery of health short videos to better cater to user preferences based on our research.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (Grant No. 72404150), Humanity and Social Science Youth Foundation of Ministry of Education of China (Grant No. 22YJC860034), Nankai University Asia Research Center Project (Grant No. AS2310), and the Undergraduate Teaching Reform and Quality Construction Research Project of Tianjin Higher Education Institutions (Grant No. B231005530).

References

- Castro-Alonso, J.C., de Koning, B.B., Fiorella, L., Paas, F.: Five strategies for optimizing instructional materials: instructor-and learner-managed cognitive load. *Educ. Psychol. Rev.* **33**(4), 1379–1407 (2021)
- Fan, A., Van Hoof, H.B., Dou, X., Serrano, A.L.: The impact of linguistic style on customer satisfaction: a cross-cultural empirical study in Ecuador. *Int. Hospitality Rev.* **37**(1), 125–142 (2023)
- Klebig, B., Goldonowicz, J., Mendes, E., Miller, A.N., Katt, J.: The combined effects of instructor communicative behaviors, instructor credibility, and student personality traits on incivility in the college classroom. *Commun. Res. Rep.* **33**(2), 152–158 (2016)
- Korbach, A., Brünken, R., Park, B.: Differentiating different types of cognitive load: a comparison of different measures. *Educ. Psychol. Rev.* **30**(2), 503–529 (2018)
- Krejtz, K., Duchowski, A.T., Niedzielska, A., Biele, C., Krejtz, I.: Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS ONE* **13**(9), e0203629 (2018)

- Li, J., Kizilcec, R., Bailenson, J., Ju, W.: Social robots and virtual agents as lecturers for video instruction. *Comput. Hum. Behav.* **55**, 1222–1230 (2016)
- Sweller, J.: Cognitive load during problem solving: effects on learning. *Cogn. Sci.* **12**(2), 257–285 (1988)
- Yuan, M., Zeng, J., Shang, J.: Would it be better if instructors technically adjust their image or voice in online courses? Impact of the way of instructor presence on online learning. *Front. Psychol.* **12**, 746857 (2021)



A Research Pathway for Interaction Design of Home-Based Elderly Care Service Robots Based on Scenarios

Canqun He^(✉), Pengjie Shi, and Xiaomeng Sun

Hohai University, 210024 Nanjing, Jiangsu, China
hecq@163.com

Abstract. To explore interaction design methods for service robots intended for home-based elderly care, aiming to more accurately capture the needs of elderly users and to facilitate design centered around these needs, thereby enhancing the efficiency of robot interaction design and advancing the development of service robots for elderly care.

Methods: In response to existing issues with home-based elderly care service robots, this study is based on scenario-based design (SBD) and integrates a variety of research methods and tools throughout the process, exploring a re-search pathway for interaction design of service robots for home-based elderly care.

Results: Drawing on the principles of SBD, this research has identified the key components of home-based elderly care scenarios and analyzed the research methods at each stage of service robot interaction design. This approach has led to a systematic and scientific research process, culminating in a cohesive pathway that integrates the aesthetics and interaction design of service robots.

Conclusion: The application of SBD in the human-robot interaction (HRI) of service robots for elderly care aids in precisely understanding the needs of elderly users, promoting user-needs-oriented design, and enhancing the practicality and human-robot interaction experience. Furthermore, it provides new insights for the development of intelligent products targeted at elderly users.

Keywords: SBD · HRI · Home-based elderly care service robots

1 Introduction

The development of service robots for the elderly can support long-term living at home and effectively reduce the pressure of family and societal elderly care. HRI is one of the key factors hindering the widespread adoption of robots.

Taking into account existing research, it is apparent that current advancements in robot-related hardware technology and human-robot interaction techniques are capable of fulfilling, to a certain extent, the usage demands of the elderly within specific contexts. Particularly, the advancement of artificial intelligence technology has continuously enhanced the autonomy of robots, making the master-slave relationship during human-robot interaction more flexible and proactive [1], providing favorable conditions

for the application of service robots for the elderly. However, there is currently a lack of suitable research processes in HRI [2], leading to insufficient acquisition and fulfillment of elderly user needs, low efficiency in the development of prototypes for home-based elderly care service robots, and a lack of systematic research on robot appearance and human-robot interaction.

Therefore, this study incorporates the SBD approach into the research of HRI design for the home-based elderly care population. By integrating a variety of design methods and tools, the research process is systematic and scientific to address the associated issues in HRI design.

2 Concepts and Applications of SBD

Scenario-based design emphasizes shifting the focus of design work from defining system operations to describing how people will use the system to accomplish their tasks [3]. In HCI, scenarios serve as a means that encompass elements such as participants, background information about the participants and assumptions about their environment, the participants' goals or objectives, and sequences of actions and events [4]. In terms of its manifestation, a scenario can be represented through various media and formats, including textual narratives, storyboards, video models, or script prototypes [5]. SBD involves transforming the design object from an interaction system to a series of actions and experiences within human activities [3], which can assist designers in better understanding user needs and behaviors and translating them into specific design solutions and interactive experiences.

Scenarios can be applied at a specific stage within the interaction design process or integrated throughout the entirety of the interaction design workflow. In specific design practices, the types of scenarios continually evolve based on the different stages of the design process. Numerous classification methods have been established to categorize relevant scenarios. Here, scenarios are divided into objective scenarios for gathering user requirements, target scenarios for deriving design solutions, and validation scenarios for assessing the feasibility of design proposals [6].

3 Scenario-Based Interaction Design Approach for Home Care Service Robot

3.1 Content Analysis of Home-Based Aging Scenarios

When utilizing scenarios to study issues related to home-based elderly care, the key is to construct objective scenarios that fully reflect the current state of home care. To create an objective scenario for home-based elderly care, it is necessary to gather information on the characteristics of the home-care population, living environment, daily activities, products usage, and service requirements. Additionally, the interrelationships and dynamic changes among these constituent elements must be considered, followed by presenting the home-based elderly care scenario through visualization methods (see Fig. 1). When analyzing user needs, it is necessary to deconstruct the complex home care scenarios and determine the primary and secondary relationships of the deconstructed

scenarios to more effectively understand user needs. Furthermore, considering the diversity within the elderly population, it is necessary to differentiate target users based on research content to identify the needs of different types of users and provide targeted interaction design methods.



Fig. 1. Scenario elements of home-based aging

3.2 Scenario-Based Interaction Design Process of Home Care Service Robot

The interaction design process for home-based elderly care service robots is primarily divided into three stages: user requirement analysis, design proposal generation and iteration, and design proposal evaluation. Considering the characteristics of elderly users, as well as the diversity of service robot presentation forms and HRI features, it is necessary to select appropriate research methods at different design stages based on SBD. The research content should be appropriately refined to adapt to the interaction design of service robots. Ultimately, this leads to the formation of an interaction design process as shown in Fig. 2.

User Requirements Analysis. Before establishing an objective scenario, it is important to gather scenario information by considering the universal aspects of the elderly population, such as aging, chronic or long-term illnesses, and living conditions. Desk research can be conducted on a large number of home-based elderly care reports to form the foundation for constructing home-based care scenarios. Further collection of information on home-based elderly care life can be achieved through a combination of offline questionnaire surveys, semi-structured interviews, and in-depth interviews. It is crucial

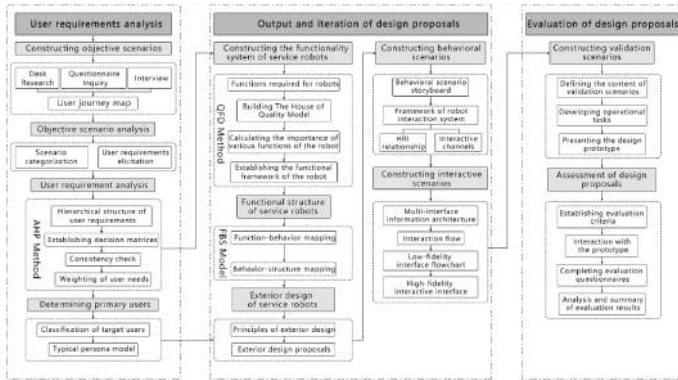


Fig. 2. Interaction design process for home-based elderly care service robots based on scenario

to establish a trust relationship with the elderly during the initial preparation for interviews, to increase their willingness to participate and facilitate the smooth progression of the interviews [7].

User journey maps detail the experiences, personal feelings, and subjective feedback of users as they interact with a product or service through a graphical representation [8]. They help designers understand the factors behind user issues from the user's perspective, allowing for a more precise grasp of user needs. Here, the user journey map is used to depict "A Day in the Life of Home-Based Aging," exploring the contextual information where problems occur in the daily lives of the elderly and condensing typical home-based care scenarios to further summarize and generalize the needs for home-based elderly care.

In order to obtain the expectations of the elderly regarding the participation of robots in their daily lives across various scenarios, as well as the weighted relationships of robot requirements, the Analytic Hierarchy Process (AHP) [9] is introduced. Experts are invited to judge the weight values of various home-based care scenarios and user needs within those scenarios. Concurrently, user research is conducted to categorize elderly users, thereby determining the primary target users for the service robot and establishing user persona models [10] to clarify the direction for subsequent design.

Design Scheme Production and Iteration. Service robots need to leverage their own information acquisition and analysis capabilities to achieve smooth HRI. Moreover, in HRI, the appearance design of the robot has a significant impact on interaction intentions and experiences [11]. Therefore, before establishing target scenarios, it is necessary to determine the information acquisition and processing capabilities that the robot needs to have based on user requirements. Then, according to the established target user characteristics and needs, set principles for the robot's exterior design to produce a service robot design proposal.

The House of Quality (HOQ) is the core of the Quality Function Deployment (QFD) method, a diagrammatic approach that demonstrates the correlation between user needs

and functional features. Using HOQ, user requirements are transformed into robot functions and the importance of each function is calculated. Subsequently, the Function-Behavior-Structure (FBS) model [12] is used to ultimately determine the robot's functional structure. The application of these methods ensures that the design revolves around user needs, which is beneficial for enhancing the feasibility of the final design proposal.

By introducing service robots into home-based elderly care scenarios, envisioning behavior scenarios that meet the key needs of elderly users, and visualizing these scenarios with storyboards, the feasibility of the robot's exterior design proposal can be more intuitively analyzed. Based on the analysis of behavior scenarios, an interaction system framework is established to clarify the interaction medium between the elderly and the robot and the value of scenario information in enhancing the human-robot interaction experience. Various types of interaction interface information architecture and interaction processes are organized, and low-fidelity interface flowcharts are used to construct interaction scenarios, further detailing the information on the interaction interface between users and service robots in specific scenarios for high-fidelity interface design.

Evaluation and Verification of Designing Scheme. Verification scenarios are set up to evaluate the design proposals for home-based elderly care service robots. Considering the complexity of robot technology and the high costs of prototype development, the Wizard of Oz approach can be used to present interaction prototypes, where designers manually control the prototype interface to interact with elderly participants, simulating the scenario of elderly users interacting with robots. In the design of the evaluation scheme, key scenarios are constructed based on the analysis of the home-based elderly care group's needs, and operational tasks are specified to evaluate the interaction design prototype. Choosing the Almere model, designed by Heerink et al. [13], as the foundation for establishing evaluation criteria. This model is specifically tailored for assessing the acceptance of service robots among elderly users [13]. Subsequently, employing Likert scales to create evaluation questionnaires aimed at assessing the performance of design prototypes on key factors influencing the acceptance of robots among elderly users.

4 Discussion and Conclusion

The SBD method provides a new research perspective for the interaction design of home-based elderly care service robots. By summarizing the descriptive and scattered user requirements in the research into contextualized scenarios, an accurate understanding of user needs can be achieved. These scenarios can also be used for communication between designers and target users in subsequent research, focusing the design on user needs. Moreover, conducting interaction design based on scenarios is beneficial for addressing user diversity and the dynamic changes in user needs, encouraging designers to create flexible solutions that adapt to different users and environments.

Based on the SBD process, it analyzes the content that needs to be studied in the interaction design of home-based elderly care service robots. It integrates various research methods and tools into the research process, taking into account the existing issues in research on interaction design for elderly users. This forms a detailed research path, providing new ideas for related interaction product design research.

Acknowledgments. This work is supported by the Ministry of Education's Industry-University Collaborative Education Project (Grant No. 220705329284309).

References

1. Zhang, S., Qin, J., Cao, S., Dou, J.: HRI design research for intelligent house-hold service robots: Teler as a case study. In: 7th International Conference, DUXU 2018, Held as Part of HCI International 2018 Proceedings, Part II, vol. 7, pp. 513–524. Springer International Publishing, Cham, Switzerland (2018)
2. Bo, W.: Research on interactive experience design trends in the era of artificial intelligence. *ART Educ.* **2021**(8), 233–236 (2021)
3. John, M.C.: Five Reasons for Scenario-Based Design[C]//Hawaii International Conference on System Sciences. *Int. Comput.* **13**, 43–60 (2000)
4. Carroll, J.M.: Making Use: Scenario-Based Design of Human-Computer Interactions. MIT press, Cambridge, MA (2003)
5. Go, K., Carroll, J.M.: The blind men and the elephant: views of scenario-based system design[J]. *Interactions.* **11**(6), 44–53 (2004)
6. Yumei, W., Weifeng, H., Jin, T., etl.: Scene theory in product interaction design. *Pkg. Eng.* **38**(6), 76–80 (2017)
7. Binda, J., Wang, X., Carroll John, M.: Recruiting older adults in the wild: reflections on challenges and lessons learned from research experience. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 290–293. ACM, New York, NY, USA (2018)
8. Xiong, D., Wenjie, Z., Shan, L.I.U.: Analysis and research on journey visualization tools in service design. *Zhuangshi.* **2021**(3), 80–83 (2021)
9. Zongming, L., Zeqi, W.: Research on design of intelligent charging pile based on SAPAD-AHP method. *J. Graph.* **44**(2), 380–388 (2023)
10. Chao, Z.: Research on the Construction of Persona Model Based on Product Design. Hu Nan University (2009)
11. Klüber, K., Onnasch, L.: Appearance is not everything - Preferred feature combinations for care robots. *Comput. Hum. Behav.* **128**, 107–128 (2022)
12. Gero, J.S., Kannengiesser, U.: Modelling expertise of temporary design teams. *J. Des. Res.* **4**(2), 0 (2004)
13. Heerink, M., Kröse, B., Evers, V., et al.: Assessing acceptance of assistive social agent technology by older adults: The Almere model. *Int. J. Soc. Robot.* **2**(4), 361–375 (2010)
14. Felding, S.A., Koh, W.Q., Teupen, S., et al.: A Scoping review using the Almere model to understand factors facilitating and hindering the acceptance of social robots in nursing homes. *Int. J. Soc. Robot.* **15**(7), 1115–1153 (2023)



Designing Paediatric Hospital Navigation: A Human-Centred Mobile App Solution

Ester Iacono^(✉) , Virginia Piombino^{}, and Francesca Tosi^{}

Laboratory of Ergonomics and Design (LED), Department of Architecture,
University of Florence, 93 Sandro Pertini Street, 50041, Florence, Italy
`ester.iacono@unifi.it`

Abstract. Hospital communication and wayfinding are crucial challenges, especially in paediatric contexts where orienting oneself in space and managing emotions is particularly difficult. The complexity of hospital environments and the ineffectiveness of orientation systems can cause cognitive stress and difficulties understanding, negatively influencing healthcare quality and operational efficiency. An effective wayfinding system must communicate clear messages, making the hospital space interpretable and familiar, using scientific approaches integrated into the architecture. Technologies such as mobile apps can improve orientation and personalize assistance, but a thorough evaluation of the usability and effectiveness of these systems is necessary. This article addresses the challenges of hospital-paediatric wayfinding by presenting a new mobile app that combines navigation, booking and entertainment. The study, conducted at the Meyer Children's Hospital in Florence, made it possible to define the requirements of a new orientation system for the medical day hospital, improving autonomous orientation and optimizing the user experience. The proposed solution developed thanks to the application of the Human-Centred Design and User Experience approach, aims to reduce stress and improve hospital efficiency, transforming the wayfinding experience into an interactive and engaging process for paediatric patients and their parents.

Keywords: In-hospital wayfinding · Human-Centred Design · User Experience · UI App mobile · Gamification in hospital

1 Introduction

Hospital communication and wayfinding challenges, particularly in paediatric care contexts, are pressing and demand immediate, innovative solutions. The complexity of these environments and the emotional management required make navigation a significant hurdle. Navigation within care spaces can be a source of cognitive stress and difficulties in understanding the space and decoding the signage system [1]. The vastness and complexity of the environments, coupled with the poor usability and ineffectiveness of the orientation system [2] and the complexity of the procedures and management of the child and his family, can lead to significant issues. These include delays in treatment, confusion between patients and healthcare workers, and increased stress and anxiety, all

of which can adversely affect the quality of care and operational efficiency. While hospitals prioritise physical well-being, they often overlook the high levels of stress, anxiety, and uncertainty that pervade this environment. Waiting areas, labyrinthine corridors, and anonymous departments can create a sense of disorientation and frustration, compromising the emotional well-being of patients and their families. This is particularly crucial for children, who need to feel secure and understand their position within the hospital context [3]. Hospitals must implement a comprehensive wayfinding system that provides unambiguous messages, making the space more understandable and familiar. It can be achieved by integrating the scientific approaches of semiotics, cognitive psychology, and design in architecture to design people's movement flows in care spaces through, for example, experiential feedback determined by paths with shapes, colours, and lighting [4]. Integrating technological tools, such as mobile apps, could significantly improve the orientation experience, making navigation more intuitive and efficient, communication more transparent and assistance more personalised [5]. The study conducted at Meyer Children's Hospital by the Ergonomics & Design Laboratory focused on improving hospital orientation through communication design. It aimed to understand how individuals navigate hospital spaces, identify optimal navigation methods for different users, and enhance overall navigation experiences. The primary goal was to develop an efficient navigation service for the Medical Day Hospital, promoting autonomous orientation and improving users' overall experience regarding time efficiency and psychological well-being. Specific objectives included: (i) analysing the navigation space and human-environment interaction; (ii) identifying user needs and critical issues in the navigation flow; (iii) evaluating the effectiveness and usability of the current orientation system; (iv) defining the requirements for a new wayfinding system.

2 Methods

The research at Meyer Children's Hospital applied Ergonomics for Design [6], Human-Centred Design (HCD) [7], and User Experience (UX) [8] methodologies to enhance the wayfinding system in a paediatric hospital setting. It involved several phases: 1) Literature analysis and benchmarking; 2) Evaluation of the current orientation system; 3) Data analysis; 4) Development of the design concept. The study included field investigations and collaboration with healthcare workers from the Medical Day Hospital service, focusing on understanding patients' needs, expectations, and perspectives of communication management roles.

2.1 Literature Analysis and Benchmarking

The study began with a literature review focused on topics related to hospital wayfinding and navigation systems. This search involved platforms like Google Scholar, PubMed, and ResearchGate using specific search strings ("Hospital Wayfinding", "Navigation Application", "Paediatric hospital", and "Gamification"). The goal was to identify and analyse scientific articles that contributed significantly to understanding orientation in hospitals and navigation technologies. The review also included benchmarking analyses of wayfinding systems such as pictogram systems, Bluetooth beacons, indoor GPS

systems, and mobile apps available in the market. This phase laid the groundwork for structuring case studies based on insights from the literature, focusing on structural requirements, environmental interactions, and typographical considerations in paediatric hospital environments at national and international levels.

2.2 Evaluation of the Current Orientation System

Phase 1 was followed by field investigations, including discussions with professionals (engineers, architects) and healthcare workers. These investigations, conducted using HCD methodologies, were aimed at understanding user needs and evaluating the current orientation system's usability. The following activities were conducted:

- *Direct observations* [9] of patients, parents, and healthcare personnel in the Medical Day Hospital involved filming and note-taking. This approach provided deep insights into users' behaviours, thoughts, and sensations during orientation, offering valuable information for redesigning the wayfinding system.
- *Architectural surveys of floors 0–1* of Meyer Hospital involved mapping the entire structure to understand navigation paths and flows from access to discharge. This process included simplifying floor plans and collecting photographic material of elements and reference points for creating cognitive maps.
- *Semi-structured interviews* [10] with diverse experts and professionals enriched the activity analysis, highlighted critical issues and unexpressed needs, and gathered information on user-environment interactions.
- *Personas, Scenarios* [11] and *Task Analysis* [6] were used to represent user and operator activities through storyboards. These tools helped analyze the operational phases of the Day Hospital and identify interaction problems with the environment.

2.3 Data Analysis

The data collected in the previous phase were essential to identifying the main orientation problems within the hospital and outlining new requirements and innovation scenarios. Field investigations have made it possible to establish a solid set of requirements summarised in maps and diagrams, highlighting design indicators to break down perceptive barriers, functional indicators relating to the main features and services of the mobile application, and the main routes for the Medical Day Hospital.

2.4 Development of the Design Concept

The last phase of this research allowed, based on the problems that emerged in the previous phases, the formulation and development of new design solutions, focusing attention on aspects related to usability and user experience. Therefore, the use of *design-orienting scenarios* [12] and brainstorming sessions was fundamental to verify and establish the functionality and requirements of the new wayfinding system.

3 Results

3.1 Evaluation and Analysis of Critical Issues

Field investigations identified significant disorientation among patients and caregivers in a hospital comprising old and new sections with complex connections between areas and wards. The lack of coherent wayfinding makes it challenging to navigate to crucial destinations like the wheelchair collection area, elevators, and certain wards, often necessitating outdoor crossings even in inclement weather. Services dispersed across multiple floors in the Day Hospital result in frequent returns to reception for information. Observations highlighted inadequate support for foreign or visually impaired users and a lack of entertainment options for young patients. For instance, children in the “cow corridor” misunderstand sculpture art and climb on them. Moreover, information and payment totems situated far from examination rooms often hinder discharge procedures. Direct observations and interviews were pivotal in outlining requirements for developing a mobile application tailored to improve orientation within this specific hospital context. The main requirements are:

- *Simple and accessible interface* for all users, with clear communication of paths, tangible feedback on reference points and displays of content diversified for specific categories of users.
- *Educational entertainment* for patients while they wait.
- *Inclusion and involvement of companions* in services that, by regulation, prohibit their access with the patient, using new forms of communication, and respecting the child’s privacy.
- *Complete management of medical services*, such as booking, payment, information to the patient before the visit and directions to the examination rooms.

3.2 My Meyer App

The collected data led to the development of “My Meyer,” an app that can be downloaded at the hospital entrance and connected to the planning of the *Single Booking Centre* (CUP 2.0) through a simple scan of the patient’s health card barcode. A key element in the app’s design was guaranteeing 3 specific activities: *navigation, booking, and entertainment*. The app includes a digital map with the main reference points that users can find directly in the architectural space and a reserved area for booking and managing the clinical dossier, even remotely. Specifically, for developing the app’s User Interface (UI), the following aspects were considered: how content is displayed, the choice of paths, the layout and the wireframe, and the UI interaction styles for specific categories of users. In particular, the main screen (Home) summarises the main functions of the app: a search bar, a “profile” section, proposed paths and wards, a lower menu with 4 icons, including the “map”. The customisable interface offers 3 graphic themes, each with a mascot accompanying and supporting families, reducing loss and loneliness. From the “home” section, you can search for destinations by doctor’s name, room number, or floor or select them directly on the hospital map. A key innovation of the app is its dual navigation system (see Fig. 1): “parent” mode focuses on efficient navigation for completing tasks quickly, while “child” mode engages patients with gamified experiences

and virtual adventures. In this mode, patients can collect prizes and discover “Meyer’s treasure” independently upon reaching their goal. Parents can access the shortest routes, customized with additional services (bar, toilets or lifts) and view the various steps to be taken, such as paying for the visit or being discharged. An “updates” section provides real-time information on the patient’s status and location for companions unable to access sensitive hospital areas, ensuring privacy compliance. Considering inclusivity for users with sensory sensitivities and diverse cultural backgrounds, the app offers accessibility features like voice guides, language options, and adjustable text sizes. The “support” section provides practical assistance, such as wheelchair pickup and elevator locations. In case of issues, a “photo recognition” function uses the camera to identify the patient’s location by capturing reference points in the hospital. “My Meyer” offers a patient-centred experience, combining efficiency, entertainment and safety for better hospital orientation.



Fig. 1. The two navigation options: for the child, a prize-based adventure that guides him along the route (on the left), and for the carer, a classic setting that allows the paths to be displayed on a planimetric map of the hospital section (on the right)

4 Discussions and Conclusions

The results of this study underscore the benefits of using Ergonomics for Design, Human-Centred Design, and User Experience methodologies in healthcare solutions. The study identified scenarios and solutions for a new wayfinding system by analysing user interactions and evaluating existing communication artefacts, including defining necessary functions and front-end requirements. Considering the complexity of the hospital structure examined, the study was limited to a single medical day hospital case; future research will expand to the entire hospital, involving more users over extended periods. The goal will be to develop a beta version of the application for field testing by all hospital users. Limited connectivity in some hospital areas could hinder the app’s effectiveness, impacting hospital orientation and patient experience. Future developments will focus on creating offline solutions for the app and exploring technological alternatives and infrastructure strategies to improve connectivity. This solution will ensure reliable access to digital resources and maximise the app’s utility, enhancing the hospital user experience.

The research demonstrates that communication artefacts in paediatric hospitals are crucial for guiding users and fostering an inclusive, welcoming, and efficient environment that accommodates children's cognitive and emotional needs. Investing in these solutions can enhance patient satisfaction, operational efficiency, and overall well-being in healthcare facilities.

References

1. Pilozzi, F., Torresan, M.: *La Comunicazione Visiva Per La Salute: Wayfinding, Pittogrammi E Health Literacy Nello Spazio Ospedaliero*, 1st edn. FrancoAngeli, Milano (2019)
2. Harper, C., Duke, T., Avera, A., Crosser, A., Jefferies, S., Klisans, D.V.: Exploring hospital wayfinding systems: design guidelines for wayfinding interfaces. In: *International Conference on Applied Human Factors and Ergonomics*, pp. 30–36. Springer International Publishing, Cham (2020)
3. Lambert, V., Coad, J., Hicks, P., Glacken, M.: Young children's perspectives of ideal physical design features for hospital-built environments. *J. Child Health Care.* **18**(1), 57–71 (2014)
4. Dalke, H., et al.: Colour and lighting in hospital design. *Opt. Laser Technol.* **38**(4–6), 343–365 (2006)
5. Smolenaers, F., et al.: User centred development of a smartphone application for wayfinding in a complex hospital environment. In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (eds.) *Proceedings of the 20th Congress of the International Ergonomics Association, IEA 2018. Advances in Intelligent Systems and Computing*, vol. 818, pp. 383–393. Springer International Publishing, Cham (2019)
6. Tosi, F.: *Design for Ergonomics*. Springer, Cham (2020)
7. *Iso 9241–210: Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems*. International Standard Organization, Ginevra (2019)
8. Garrett, J.J.: *The Elements of User Experience: User-Centered Design for the Web and Beyond*. Pearson Education, London (2010)
9. Stanton, N.A., Young, M.S., Harvey, C.: *Guide to Methodology in Ergonomics: Designing for Human Use*, 2nd edn. CRC Press, Boca Raton (2014)
10. Wilson, J.R., Sharples, S.: *Evaluation of Human Work*, 4th edn. CRC press, Boca Raton (2015)
11. Hanington, B., Martin, B.: *Universal Methods of Design Expanded and Revised: 125 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport publishers, Gloucester (2019)
12. Manzini, E., Jégou, F.: Design degli scenari. In: Bertola, P., Manzini, E. (eds.) *Design Multiverso: Appunti Di Fenomenologia Del Design*, pp. 189–207. Edizioni Polidesign, Milão (2004)



Time's Up!: Assessing the Impact of Timer and Warning Alerts in Online Learning Management System towards Student's Satisfaction and Perceived Usefulness

Beatriz Anne T. De Jose, Danielle Alexandria A. Pasia, Ezekiel L. Bernardo^(✉) ,
Jazmin C. Tangsoc, and Charmine Sheena Saflor

Industrial and Systems Engineering Department, Gokongwei College of Engineering,
De La Salle University, Manila, Philippines
ezeziel.bernardo@dlsu.edu.ph

Abstract. In the digital era, online learning is pivotal in education, yet questions remain about its essence and impact. This study investigates the impact of timers and alerts in online learning management systems (LMS) on satisfaction and perceived usefulness, considering the moderating role of user traits. Data from 152 tertiary students experienced with LMS were analyzed using structural equation modeling, revealing that timers significantly influenced satisfaction and perceived usefulness, while warning alerts only affected perceived usefulness. Personality traits, particularly openness to use, moderated these effects, highlighting individual differences in user experience. Qualitative insights emphasized the preference for countdown timers and the adverse effects of early warning alerts on user anxiety. Overall, findings underscore the importance of considering user traits in optimizing LMS design for a student-centric approach, enhancing the effectiveness and acceptance of online learning platforms within the evolving landscape of education.

Keywords: Online learning · Learning management systems (LMS) · Timer · Warning alerts · user experience

1 Background of the Study

As the digital age continues to reshape the landscape of education, one phenomenon stands out: the evolution of online learning. From its humble beginnings, online education has burgeoned into a dynamic force, particularly in the realm of higher education (Hussain and Khan, 2018). Defined as the acquisition of knowledge and skills through various internet-based mediums, online learning, through different learning management systems (LMS) offers a flexible and accessible approach to education, transcending the constraints of time and place (Kaymak and Horzum, 2013). In the wake of the recent global pandemic, the adoption of online learning has surged, serving as a vital lifeline for both educators and students alike.

Yet, amidst this rapid transformation, one must ponder: What truly defines the essence of this digital revolution in education? It's not merely about revolutionizing the learning experience but rather adapting to the ever-accelerating pace of change, while still preserving the essence of meaningful interaction.

Consider, for instance, the subtle yet profound shift in the way quizzes are administered. Gone are the days of pen and paper with clocks are secondary in our disposal; instead, we find ourselves immersed in a world where questions are upfront with timers and alerts right there with us service as a constant reminder of the fleeting nature of time, urging students to channel their focus and efforts effectively. This transition from the traditional to the digital realm underscores a fundamental truth: in education, as in life, adaptation is key.

But amidst the flurry of technological innovations, questions linger. What real impact do these digital interventions have on the learning experience? And, more crucially, how can we ensure that our online learning management systems are not only efficient but also enhance students' overall satisfaction and ease of use? Consider the presence of timers and alerts—do they truly benefit students, or do they harbor potential drawbacks? Moreover, how might a student's openness to new experiences influence their perception of these tools? These questions, though implicit in the transition from traditional systems, demand rigorous investigation.

As encapsulated in the conceptual framework presented in Fig. 1, this paper aims to address these inquiries by examining:

- The effectiveness of timers and alerts in fostering satisfaction and perceived usefulness.
- Exploration of how different designs of these elements might yield varying effects.
- Understanding the human element in the equation, in a form of user's personality, which plays a pivotal role in shaping the online learning experience.

Bridging this understanding gap will optimize tools and guide recommendations for a more student-centric approach, enhancing digital intervention utilization. To achieve this, the paper is structured as follows: Sect. 2 reviews previous studies, Sect. 3 outlines the methodology, Sect. 4 discusses results, and Sect. 5 concludes with operational recommendations.

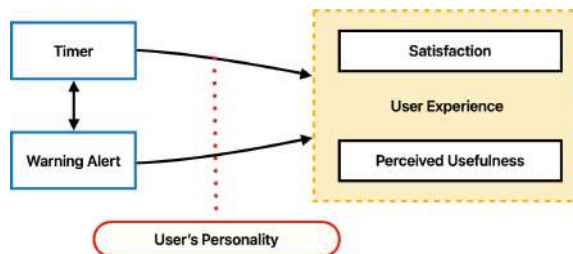


Fig. 1. Conceptual framework illustrating the theorized direct effect of timers and warning alerts in LMS on user experience, measured by satisfaction and perceived usefulness, with user's personality as the moderator of this process

2 Review of Related Literature

2.1 Online Quizzes

Learning Management Systems (LMS) revolutionize education by facilitating resource delivery and enhancing interaction. Online quizzes, a key feature, offer a technological alternative for assessment, but their efficacy depends heavily on interface design (Borisov et al., 2020). Literature highlights the importance of displaying remaining time during assessments (Sek et al., 2012), which improves timely submission and organizational efficiency (Islam Rahman and Islam, 2013). Alerts contribute to better time management and exam performance (Abdel Karim and Shukur, 2016; Ponnle et al., 2014), guided by individual personality traits (Grey et al., 2009). However, empirical evidence on optimal timer and alert configurations remains limited, necessitating further investigation aligned with established standards.

2.2 User Experience

Understanding student perceptions of online learning, including exam interface design, is crucial as online education expands (Barnes, 2017). Variations in student experiences impact satisfaction and perceived usefulness (Barnes, 2017; Jung & Lee, 2018). Satisfaction, driven by responsive services and high-quality materials, significantly influences user acceptance. Institutions must gauge satisfaction to foster positive relationships and optimize the online quiz experience (Tabaku, 2018). Perceived usefulness, crucial for technology adoption, particularly in Learning Management Systems (LMSs), guides system evaluation in online settings (Davis, 1989; Juhary, 2014). Understanding the relationship between interface design, student reactions, and openness to experience is vital for tailoring assessments effectively and refining digital learning environments (Barnes, 2017).

2.3 User's Personality

Personality, defined as a set of underlying traits influencing individuals' behavior, thoughts, and feelings, plays a significant role in technology usage, including engagement with online quizzes (McGeown et al., 2014). Specifically, openness to experience, characterized by traits like imagination and curiosity, correlates with active engagement in intellectual activities and intrinsic motivation for learning (Soto, 2018). However, while literature extensively discusses certain design features of online examination interfaces, such as font styles and color schemes, little attention has been devoted to optimizing timer and alert configurations, crucial elements for efficient exam-taking experiences (Placencia and Muljana, 2019). Emphasizing the need for comprehensive user interface design, particularly in online exam environments, these features hold potential to enhance students' evaluative skills and promote efficient task performance (Olipas and Luciano, 2020). Moreover, existing research lacks exploration into how personality traits, particularly openness to experience, influence user interaction with online exam interfaces, highlighting the importance of adapting interfaces to cater to individual traits and needs, thereby facilitating smoother adoption and promoting usability (Cohen and

Baruth, 2017). Understanding the interplay between personality and interface design is crucial in alleviating the added stress of online exams and ensuring equitable testing procedures for all students.

3 Methodology

The methodology involved several key steps. For the independent variable, a modified LMS was developed, integrating a timer and warning alert. Concurrently, questionnaires were refined to assess user traits for the moderator and perceived usefulness plus satisfaction for the dependent variables. Using a design of experiment approach, factor combinations were determined, and sample size was calculated through power analysis. A pilot test, involving 10% of the sample size, assessed usability and user experience. Participants, tertiary students experienced with LMS, were recruited based on predefined criteria. Data analysis utilized structural equation modeling and f-tests, aligning with predefined objectives. This process led to the formulation of a final framework and recommendations, contributing to understanding LMS design and user experience (Fig. 2).

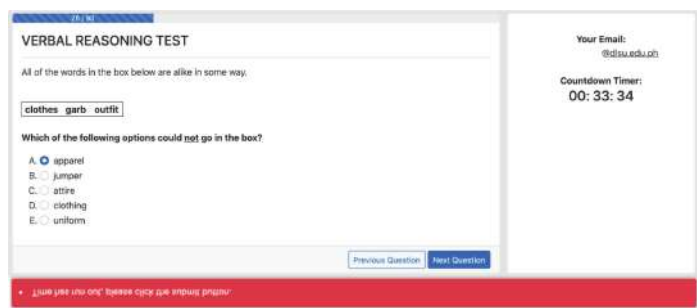


Fig. 2. Sample of testbed quiz page showing a countdown timer & warning alert pop-up banner

4 Results and Analysis

The experiment yielded insightful findings regarding the influence of timer and warning alert features within a modified Learning Management System (LMS), as well as the moderating effect of personality traits on user experience. Over a span of two weeks, data were collected from 152 participants who completed the experiment without encountering normality issues. Exploratory factor analysis confirmed that all latent variables measured through questionnaires loaded correctly onto the proposed structure.

The structural equation modeling (SEM) analysis revealed that the proposed model explained 62% of the variation, indicating a good fit on absolute incremental and parsimonious fit indices. However, out of the seven possible paths tested, only five were supported, suggesting nuanced relationships between the variables under study. Specifically, the use of the timer emerged as significant for both satisfaction and perceived usefulness, whereas the warning alert only loaded onto perceived usefulness, and the interaction between timer and alert was not significant (Table 1).

Table 1. Summary of objectives based on path analysis

Objectives	Path Analysis	β	p -value	Supported
<i>Direct</i>	Timer $\rightarrow$ Satisfaction	0.511	0.025	Yes
	Timer $\rightarrow$ Perceived Usefulness	0.021	0.003	Yes
	Alert $\rightarrow$ Satisfaction	0.432	0.521	No
	Alert $\rightarrow$ Perceived Usefulness	0.333	0.124	No
<i>Interaction</i>	Timer $\leftrightarrow$ Alert	0.092	0.087	No
<i>Moderator</i>	Openness $\diamond$ Satisfaction	0.231	0.023	Yes
	Openness $\diamond$ Perceived Usefulness	0.193	0.012	Yes

Notably, personality traits were found to moderate all paths at a global level, indicating individual differences in how users perceive and interact with LMS features. Qualitative insights gleaned from interviews further elucidated these findings. Participants expressed preference for a countdown timer over alternative formats, citing ease of use and reduced cognitive load. However, early warning alerts were met with apprehension, as they induced anxiety and diverted attention away from the quiz, highlighting the importance of timing and presentation in enhancing user experience.

Moreover, personality traits, particularly openness to use, were found to significantly influence engagement levels, with individuals exhibiting higher levels of openness reporting greater satisfaction and perceived usefulness. These findings underscore the multifaceted nature of user experience within LMS environments, wherein features such as timers and alerts interact with individual characteristics to shape perceptions and outcomes.

5 Conclusion and Recommendation

In conclusion, this study sheds light on how timer and warning alert features in a modified Learning Management System (LMS) impact user experience, with a focus on personality traits. The countdown timer notably enhances satisfaction and perceived usefulness, while early warning alerts may induce anxiety. Personality traits, especially openness to use, moderate these effects, emphasizing the importance of personalized approaches in LMS design. These findings enrich educational technology literature, advocating for tailored strategies to enhance user experience and learning outcomes in digital environments.

References

Abdel Karim, N., Shukur, Z.: Proposed features of an online examination interface design and its optimal values. *Comput. Hum. Behav.* **64**, 414–422 (2016). <https://doi.org/10.1016/j.chb.2016.07.013>

Barnes, C.: An analysis of student perceptions of the quality and course satisfaction of online courses. *J. Higher Educ. Theor. Pract.* **17**(6), 92–98 (2017)

- Borisov, E., Petrenko, E., Zatsarinnaya, E., Romanov, A., Cherkina, V., Sotnikova, L.: Organization of taking exams of university students in online format: Problems and opportunities. In: E3S Web of Conferences, vol. 217, p. 08005. EDP Sciences (2020)
- Cohen, A., Baruth, O.: Personality, learning, and satisfaction in fully online academic courses. *Comput. Hum. Behav.* **72**, 1–12 (2017). <https://doi.org/10.1016/j.chb.2017.02.030>
- Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **13**(3), 319–339 (1989). <https://doi.org/10.2307/249008>
- Grey, I., Healy, O., Leader, G., Hayes, D.: Using a time timer TM to increase appropriate waiting behavior in a child with developmental disabilities. *Res. Dev. Disabil.* **30**(2), 359–366 (2009). <https://doi.org/10.1016/j.ridd.2008.07.001>
- Hussain, Z., Khan, W.: Traditional and Online Learning: Student Perception Survey (2018)
- Islam, M.Z., Rahman, M.M., Islam, M.K.: Online examination system in Bangladesh context. *Int. J. Sci. Environ. Technol.* **2**(3), 351–359 (2013)
- Juhary, J.: Perceived usefulness and ease of use of the learning management system as a learning tool. *Int. Educ. Stud.* **7**(8), 23–34 (2014). <https://doi.org/10.5539/ies.v7n8p23>
- Jung, Y., Lee, J.: Learning engagement and persistence in massive open online courses (MOOCs). *Comput. Educ.* **122**, 9–22 (2018). <https://doi.org/10.1016/j.compedu.2018.02.013>
- Kaymak, Z.D., Horzum, M.B.: Relationship between online learning readiness, structure and interaction of online learning students. *Educ. Sci.: Theor. Pract.* **13**(3), 1792–1797 (2013). <https://doi.org/10.12738/estp.2013.3.1580>
- McGeown, S.P., Putwain, D., Geijer Simpson, E., Boffey, E., Markham, J., Vince, A.: Predictors of adolescents' academic motivation: Personality, self-efficacy and adolescents' characteristics. *Learn. Individ. Differ.* **32**, 278–286 (2014). <https://doi.org/10.1016/j.lindif.2014.03.022>
- Olipas, C.N., Luciano, R.G.: Understanding the impact of using countdown timer on the academic motivation and computer programming anxiety of IT students: The case of a state university in the Philippines. *Int. J. Sci. Technol. Res.* **9**, 235–240 (2020)
- Placencia, G., Muljana, P.S.: The effects of online course design on student course satisfaction. In: Proceedings of the 2019 American Society for Engineering Education Pacific Southwest Section Meeting (2019)
- Ponnle, A.A., Adedeji, K.B., Udekwe, C.N.: Development of a microcontroller based countdown time reminder with visual and aural indicators. *Int. J. Scientific Eng. Technol.* **3**(11), 1380–1385 (2014)
- Sek, Y.-W., Law, C.-Y., Liew, T.-H., Bt Hisham, S., Lau, S.-H., Pee, A.N.B.C.: E-Assessment as a self-test quiz tool: The setting features and formative use. *Procedia. Soc. Behav. Sci.* **65**, 737–742 (2012). <https://doi.org/10.1016/j.sbspro.2012.11.192>
- Soto, C.J.: Big Five personality traits. In: Bornstein, M.H., Arterberry, M.E., Fingerman, K.L., Lansford, J.E. (eds.) *The SAGE Encyclopedia of Lifespan Human Development*, pp. 240–241. Sage, Thousand Oaks, CA (2018)
- Tabaku, E.: Students' perceptions on quality and satisfaction in higher education. In: Springer Proceedings in Business and economics, pp. 797–810 (2018). https://doi.org/10.1007/978-3-319-70377-0_55



Mobile Layout Design for Effective Exploration of Food Labeling Information

Yunah Kang¹ and Youngjae Im^{2,3}✉

¹ Department of Chemical, Environmental and Ergonomics,
Dong-eui University, Busan, South Korea

² Department of Ergonomics Engineering, Dong-eui University, Busan, South Korea
ergoim@deu.ac.kr

³ The Research Institute of UX, Dong-eui University, Busan, South Korea

Abstract. This study investigates how information should be structured for efficient cognitive processing when providing food labeling information on mobile platforms. To determine how to present various food labeling information, a two-step approach is implemented. Firstly, a survey is conducted to evaluate and categorize the importance of 28 food labeling elements. Based on these categories, the mobile layout is designed. Secondly, eye-tracking techniques are implemented to measure performance time, fixation time, and scan path. This helps identify which layout facilitates optimal eye movement. After the experiment, get qualitative feedback by asking about challenges in information exploration. When providing food labeling information in a mobile environment, it is advisable to use a ‘Tap & List’ layout. This is due to its high satisfaction in terms of both interactive information exploration and usability. Moreover, exploring diverse components of mobile devices as experimental variables will be essential for improving screen design.

Keywords: Food Labeling Information · Mobile Layout · Eye-tracking

1 Introduction

There is an ongoing debate about how to display information on product packaging. The area on the product packaging is small. There is a lot of information consumers want or need to know, and many consumers are frustrated by the small print, making it difficult to read the information. To solve this readability problem, reduce the information displayed on the packaging and increase the font size. A typical example is the Korean Ministry of Food and Drug Safety’s (MFDS) “E-label” system. E-labeling uses QR codes on drug or food packaging to ensure that only essential information is displayed on the package. In contrast, the rest of the information is available on mobile. There is a widespread interest in nutrition information on food packaging, but this interest varies by person, context, and product [1]. Consumers can get frustrated when information they want to know is missing, and mobile can help solve this problem. Advantages of the E-label system First, it can increase packaging visibility. Second, it can contain unlimited information and is easily modified when information changes. Third, it can save resources by reducing

waste when packaging information is modified. Various studies on the readability of food labeling information, such as the Olivier Droulers and Mary Alton Mackey studies, point to the need for national and international action [2, 3]. Regarding the user interface, layout plays a huge role in mobile readability and efficient information perception [4, 5]. A significant challenge is presenting a lot of information on mobile so that consumers easily perceive it. Therefore, this study addresses how information should be organized for efficient perception when providing nutrition information in a mobile environment.

2 Method

2.1 Categorizing Food Labeling Information

34 individuals, 16 men and 18 women aged between 20 and 50 participated in an online experiment. Participants rated the importance of 28 food labeling information using a 5-point Likert scale. To help them understand each food labeling information, we provided an image of a familiar product, Maxim Coffee Mix Mocha Gold.

2.2 Experimental Design for Eye-Tracking

Participants. An experiment was conducted on 5 people (2 men in their 20 s, 1 woman in her 20 s, and 1 man in his 40 s). All participants had no eye disease and a corrected visual acuity of 0.8 or better.

Apparatus. Among the various eye trackers, we utilized the mobile software ‘See So Web Analysis.’ The software is compatible with smartphones, PCs, and VR devices, making it accessible without hardware.

Variables. Three levels of screen layout defined the experimental variable: ‘List’, ‘Tap & List’, and ‘Card & List’. The dependent variables included performance time, fixation time, and scan path (Fig. 1).

Procedure. Before starting the experiment, participants understand the whole process and see the layouts. The experiment was conducted to ensure precise eye tracking with the phone slightly lowered and facing downward. The three screen layouts are in randomized order. We also randomly selected two pieces of information from each food labeling category (12 pieces). The post-experiment interview concludes the experiment by asking about the difficulty of each type of navigation and any other peculiarities observed during the experiment.

3 Results

3.1 Results of Categorizing Food Labeling Information

A total of 28 food labeling information was prioritized for importance. The survey was conducted using a 5-point Likert scale. One-way ANOVA analysis was then performed using SPSS. The results showed that the probability of significance (P) was less than 0.05, indicating a difference in importance in at least two groups. In conclusion, food labeling information was divided into six groups (Table 1).

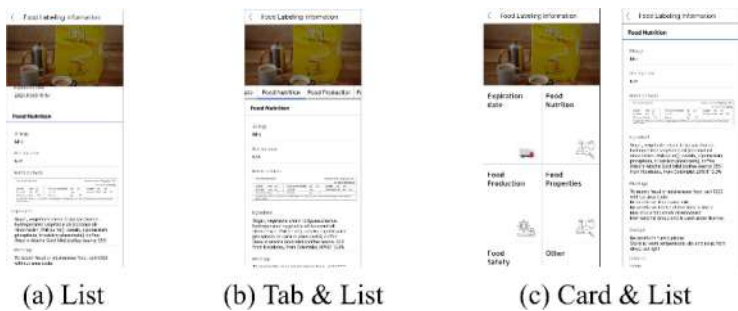


Fig. 1. Three levels of information layout

Table 1. 6 categories of food labeling information

Category	Expiration	Food	Food	Food	Food	Other
Sub-Items	Date	Nutrition	Production	Properties	Safety	
	Sell-by date	Allergy	Discharge mark	Ingredient content	Customer service	report numbers
		Use-by date	Location	Recipes	Food brands	Other labeling
		Nutrition facts	Manufacturing process	Product mark	Product exchange	Site URL
		Ingredient		Quality assurance	Packaging materials	
		Warnings		Food types		
		Storage		Genetically modified		
		Capacity(calories)		Sodium content		
		Product name		Irradiated foods		
		Date of manufacture				

3.2 Results of Eye-Tracking

Total Performance Time. The total performance time is the total time to find all 12-food labeling information ‘List’ (47.46 seconds on average) is the quickest to find information. ‘Card & List’ takes the longest to find information, with an average of 120.84 seconds. A ONE-WAY ANOVA was performed to check the significance of the difference in total performance time by layout type. There was a significant difference in total performance time among the three layouts ($f = 26.211$, $P < 0.5$) (Table 2).

Fixation Time. Fixation time is the amount of time spent looking at the information. ‘List’ shows the lowest fixation time, averaging 38.92. A ONE-WAY ANOVA was

Table 2. ANOVA results of Total Performance Time

Source		Total Performance Time				
		S.S	df	Mean Square	F	Sig.
Information layout	Between Groups	226476.583	2	13238.292	26.22	<.001***
	Within Groups	10602.146	21	504.864		
	Total	37078.730	23			

conducted to determine if the difference in fixation time by layout type was significant. There was a significant difference in fixation time between the three layouts ($F = 24.178$, $p < .05$) (Table 3).

Table 3. ANOVA results of Fixation Time

Source		Fixation Time				
		S.S	df	Mean Square	F	Sig.
Information layout	Between Groups	8124.306	2	4062.153	24.178	<.001***
	Within Groups	3528.174	21	168.008		
	Total	11652.480	23			

Scan path. The scan path measures the direction of travel and occupancy of the eye as it looks at the layout (Fig. 2).

List. The first gaze point is often the top image, and then the gaze moves downward. Repeated up-and-down scrolling occurred to find the 12 pieces of information.

Tab & List. The first gaze point mostly starts in the horizontal category. Participants didn't care about the image. From the initial gaze point, the gaze moves to the 'n' shape. After selecting a category tap, the participant reads its entire content. They move on to the next category if they don't find the information they seek. They scroll up and down within a category less often.

Card & List. Most of the first gaze starts on the image. The scan path varied by participant. They first determined which category the information they were looking for was in and then selected that category. 'Card & List' quickly left the page without scrolling up and down.



Fig. 2. Results of Scan path

4 Discussion

4.1 Categorizing Food Labeling Information

Out of the six categories, four category keywords are most valued by people: Food Nutrition, Food Production, Food properties, and Food Safety. Excluding expiration dates and other information mostly found on the packaging, these are the information that people want to see on mobile. Each sub-item is organized similarly. This finding suggests that consumers can quickly find the information they seek when presented with categories.

4.2 Eye-Tracking

‘List’ scored the highest, but testers preferred the ‘Tab & List’ the most in the post-experiment interviews. This preference is consistent with other studies. Moon’s study explains that a combination of List and Bar(‘Tap & List’) is preferred regarding app screen layout preferences [6]. This is because ‘Tap & List’ involves relatively little scrolling, unlike ‘List’, which consumers find less tedious. ‘Tap & List’ and ‘Card & List’ are also favored because of their active navigation process, which looks at categories first to deduce where the desired information is located. However, ‘Card & List’ is less satisfying because of its long navigation time. A study by Chen C.H. and Zhai W. explains that when presenting information using a combination of text and images, a matrix (grid) is preferred over a list [7]. If organizing a mobile screen for providing information, it is recommended to use a matrix (grid) layout when images and text should be combined.

In this study, food labeling information is presented in text only. The screen layout is best organized in a ‘Tab & List’ format for easy information navigation. Furthermore, if images are added when organizing the screen, the screen should be designed to include a matrix form as appropriate.

5 Conclusion

This study shows how the layout of categories and the information within them makes navigation favorable and easy. First, Food Labeling Information is divided into six categories with similar characteristics. Second, eye-tracking tests and interviews identified the most preferred layout types. Based on the simple score, 'List' is the best, but further qualitative analysis shows that 'Tap & List' is the most preferred. The 'List' is repetitively scrolled up and down. So, the consumer feels that the information is not easily found properly. 'Tab & List' takes longer, but it is better preferred because there is less repetition of scrolling up and down. In addition, comparisons with other studies confirm that the 'Tab & List' design is commonly favored for informational screens.

E-labels have been proposed as a solution to the ongoing issue of product packaging readability. This study is significant because it presented mobile screens that make food labeling information more accessible to find. On the other hand, data collection was limited due to the small number of subjects in the eye-tracking. The results do not reflect GUI or technical elements in layout design, so further research is needed. Although the study was conducted on food labeling information, the same layout structure for informational screens will likely apply to various product groups such as pharmaceuticals and cosmetics.

References

1. Klaus, G., Josephine, M.: A review of European research on consumer response to nutrition information on food labels. *J. Public Health*. **15**, 385–399 (2007)
2. Olivier, D., Jennifer, A.: The legibility of food package information in France: an equal challenge for young and elderly consumers. *Public Health Nutr.* **19**(6), 1059–1066 (2016). <https://doi.org/10.1017/S1368980015002141>
3. Mary, A., Marilyn, M.: Ease of reading mandatory information on Canadian food product labels. *Int. J. Consum. Stud.* **33**(4), 359–520 (2009). <https://doi.org/10.1111/j.1470-6431.2009.00787.x>
4. Park, K.: The effect of content layout in mobile shopping product page on product attitude and purchase intention: Focusing on consumer cognitive responses depending on regulatory focus. *Knowl. Manag. Res.* **23**(2), 193–210 (2022)
5. Kim, H.: A study on the designing user interfaces for mobile web. *Digit Des. Stud.* **10**(2), 65–74 (2010). <https://doi.org/10.17280/jdd.2010.10.2.007>
6. Moon, A.: Study on social commerce App from the UX design. *J. Digit. Interact. Des.* **12**(3), 35–46 (2013)
7. Chen, C., Zhai, W.: The effects of information layout, display mode, and gender difference on the user interface design of mobile shopping applications. *IEEE*. **11**, 47024–47039 (2023). <https://doi.org/10.1109/ACCESS.2023.3274575>



Anthropomorphic is Always Better? Effects of Financial Robo-Advisor Types on Investors' Online Finance Investment Behavior

Xinyue Kang, Fanxuan Wang, and Pei-Luen Patrick Rau^(✉)

Department of Industrial Engineering, Tsinghua University, Beijing, China
rpl@mail.tsinghua.edu.cn

Abstract. Internet financial management is booming. Many wealth management companies have launched a variety of intelligent financial robo-advisors to help users in asset allocation and choice of financial products. This research designed two types of financial robo-advisors, humanized and un-humanized, and investigated the changes of consumers' financial investment behaviors when they face the advice of different financial robo-advisors, and the impact of investors' risk preference through questionnaires. The results showed that compared with the humanized robo-advisor, the un-humanized robo-advisor had a stronger impact on investors, but only found for investors with low-risk preference.

Keywords: Finance robo-advisor · Online finance investment · Un-humanized V.S. humanized

1 Introduction

As China's economy develops, residents' demand for financial management has increased. Due to cultural and historical factors, they prefer saving over investing, which negatively impacts economic growth. The rise of online financial products has helped address this issue, because they have lower transaction costs and reduce risks encouraging higher consumption. This has led to the rapid development of internet financial products, including professional robo-advisor platforms like "Ant Wealth" by Ant Financial. These platforms use big data, quantitative models, and intelligent algorithms to offer asset allocation and investment advice.

Despite their widespread use, the effectiveness of these robo-advisors remains unclear. Studying their impact on investment behavior can guide future financial product design and development.

2 Related Work

Many factors influence investors' behavior. Bucciol and Zarri (2017) analyzed data from the US Health and Retirement Study and found that traits like cynical hostility, agreeableness, and anxiety negatively affect financial risk-taking. Barber and Odean (2001) showed that men trade more excessively than women, driven by overconfidence, which often results in lower profits.

Investors' risk and value perceptions of financial products also impact their decisions. Perceived usefulness is crucial for adopting online financial services, whereas perceived risk is less of a deterrent. When consumers understand that online financial products can enhance returns and payment efficiency, they are more likely to adopt them. Robo-advisors facilitate this understanding. A robo-advisor is an interactive digital platform that guides investors through automated investment consultations using information technology. Similar to human advisors, robo-advisors assess clients' profiles through questionnaires that gauge goals, risk appetite, and return expectations, then provide portfolio advice via intelligent algorithms.

However, robo-advisors face acceptance challenges. Key factors affecting acceptance include the automation level of the service (Rühr et al., 2019), user control over the system (Jung et al., 2018), and system transparency. User characteristics, such as educational background and smart device experience, also influence acceptance. Moreover, perceived usefulness and familiarity with the system impact users' willingness to use it (Belanche et al., 2019).

Using a conversational agent (CA) as a robo-advisor interface, like "Zhi Xiaobao," could offer advantages over traditional interfaces. CAs interact using natural language, respond to commands, and provide assistance with various forms of embodiment (Knote et al., 2018). However, it's not always better to anthropomorphize robo-advisors. Emphasizing ease of use, speed, reliability, and accuracy in product marketing can boost consumer purchase intentions and income more than increasing user dependency and engagement (Tassiello et al., 2021).

Investors' decisions are influenced by personal characteristics and their risk and value perceptions. Intelligent financial assistants help users understand financial products' features, such as historical yields and average volatility. This information enhances the perceived value of financial products, encouraging active investment. So, we have our hypothesis:

H1: The investment decisions of investors are closer to suggestions of financial robo-advisor under the participation of financial robo-advisor.

H2: under participation of financial robo-advisor, the perceived risk of investors for online financial investing will decrease.

At the same time, in reference to historical studies, users in complex tasks have a higher degree of trust in un-humanized systems, and we propose a final hypothesis:

H3: un-humanized financial robo-advisor will increase the likelihood that investors will rely on the advisor more than humanized financial robo-advisor.

3 Method

3.1 Independent and Dependent Variables

We designed two different intelligent financial robo-advisor prototypes. They are the un-humanized intelligent investment system S1 and the humanized intelligent investment assistant "Xiao Bao". The functions of the two systems are the same. They can do risk preference assessment, asset allocation advice, financial knowledge sharing and financial products recommendation, four main functions. The financial knowledge and product information, as well as the financial advice provided, are also consistent.

As shown in Fig. 1, the information is introduced in the same statement from Alipay. S1 simulates the un-humanized interface of Bank of China and other banking APPs. The interface is similar to a typical mobile app interface, presenting information in the form of cards, with a formal and official tone. Users interact by clicking buttons. Xiao Bao interacts with users through natural language conversations and keyword triggers. It has a more intimate tone, like your friend. Participants were randomly asked to read a simulation of the interaction process of one of the systems through a questionnaire.

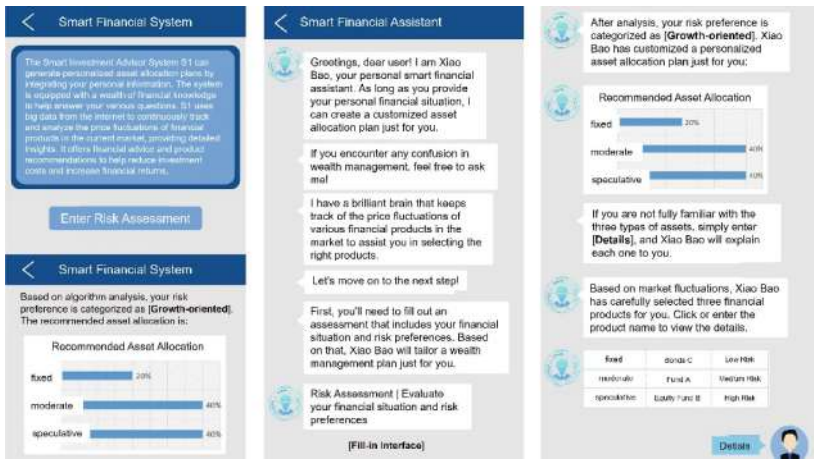


Fig. 1. Two financial robo-advisor model

In terms of dependent variables, we collected participants' asset allocation decisions before and after reading materials as the main behavioral data. They were asked to judge what proportion they would choose in fixed income, moderate and speculative products, if they had 20,000 yuan to invest. In the interaction materials of the two types of intelligent assistants, both assistants indicated that the participants' risk preference measurement results were growth type. Participants are advised to invest in a ratio of 20:40:40.

In terms of subjective variables, we measured users' risk preference, risk perception, perceived validity, trust and satisfaction. The financial risk part of the domain-specific risk taking scale (DOSPRT) was used to measure risk preference (Wu and Cheung, 2014). The measurement of risk perception includes five dimensions, financial risk perception, privacy risk perception, time risk perception, psychological risk perception and comprehensive risk perception, and 16 questions. Trust referred to Jian et al., 2000 and used a scale with 12 questions for measurement. In the data analysis, we calculated the average score of each dimension.

3.2 Participants and Questionnaire

We conducted an inter-group experiment using online questionnaires.

Questionnaire I collected participants' basic information, including gender, age, online finance experience, risk preference, risk perception, and their first round of investment decisions.

A total of 88 participants completed Questionnaire I. Responses with incomplete items, inconsistent answers, or unusual completion times were excluded, leaving 62 valid questionnaires. These participants then received Questionnaire II, which involved reading one of two materials on financial robo-advisor systems to simulate a second round of investment decisions. Questionnaire II also assessed their perceived validity, trust, satisfaction with the robo-advisor, and their risk perception post-experiment.

The participants were all young college students (mean = 23) with a balanced gender ratio. They had some understanding of online financial behavior but generally low financial knowledge. Most participants used Alipay (95.55%) and various banking apps for online wealth management, and a majority (88.71%) had used intelligent investment systems, including risk assessment functions. This familiarity suggests that participants' responses in the experiment could reflect their real reactions and experiences with online financial management systems.

4 Result

To compare the participants' investing decision-making behaviors, we compute the Euclidean distance of subjects' investing decisions and the recommendation provided by robo-advisor. Table 1 shows mean of the allocation proportion on three kinds of financial products including fixed income products, moderate growth products and speculative products.

Table 1. Investing decisions difference between two types of robo-advisors

	Un-humanized			Humanized			Difference of increased amount
	First decision	Second decision	Increased amount	First decision	Second decision	Increased amount	
Fixed income products	42.23	25.90	-16.32**	34.97	40.71	5.74	22.06**
Moderate growth products	34.26	42.29	8.03	36.94	36.81	-0.13	-8.16
Speculative products	23.52	31.81	8.29	28.10	22.48	-5.61	-13.90*
Euc_distance to recommendation	44.13	23.47	-20.65**	40.34	37.16	-3.18	17.47*

* $p = 0.05$; * * $p = 0.01$; * * * $p = 0.001$

Participates' allocation proportions on fixed income products only show significant increasing with the help of un-humanized robo-advisors. The increased amounts of allocation on fixed income products between un-humanized and humanized robo-advisors shows significant difference. When under the un-humanized robo-advisors, the Euclidean distance between participates' investing decision and the provided recommendation significantly decreases while not when under the humanized robo-advisors. Besides, the decreased distance shows significant difference between un-humanized and humanized robo-advisors. In general, the un-humanized robo-advisor has greater influence on participates' investing decisions.

The investing decision-making process may also be affected by the risk preference of participates. Considering 18 as the dividing boundary, participates with risk scores below 18 are considered as low risk preference and vice versa. Therefore, we divided total participates into two group. Group L denotes participates with low-risk preference and Group H denotes participates with high-risk preference. Results demonstrate that the distance between investing decisions and provided recommendation will decrease despite of risk preference and types of robo-advisors.

From Table 2, we know that decreased distance value shows significant difference between un-humanized and humanized robo-advisors in among participates with low-risk preference while not among participates with high-risk preference. The result shows that people with low-risk preference are more likely to rely on un-humanized robo-advisor. Besides, decreased distance value shows significant difference between groups with the help of un-humanized robo-advisor while not with humanized one, which demonstrated that people with low-risk preference are more likely to be affected by un-humanized robo-advisor.

Table 2. Investing decisions difference analysis

	Type A	Type B	Difference of decreased distance
Group L	30.70	4.60	26.09**
Group H	4.75	2.16	2.60
Difference of decreased distance	25.95*	2.45	—

*p = 0.05; * * p = 0.01; * * * p = 0.001

As for perceived risk, only the difference of perceived financial risk shows significance with the participation of un-humanized robo-advisor. There is no significant difference of people's perceptions of robo-advisors between difference types, including effectiveness, satisfaction and trust.

5 Discussion

In this report, we investigated how un-humanized and humanized robo-advisors influence people's investment decisions. Our data analysis confirmed that robo-advisors significantly impact these decisions, with users tending to follow the provided recommendations, supporting hypothesis 1.

We also found that un-humanized robo-advisors have a stronger influence on investors compared to humanized ones, aligning with hypothesis 3. This can be attributed to research on persuasion, which indicates that when tasks are complex, people's reliance on advice is partly based on their belief in the advisor's competence. Investing is inherently complex due to the variety of financial products and the risk of losses. Humanized robo-advisors, through natural language interactions, can appear less capable of handling complex investment tasks. This perception is reinforced by the current limitations in natural language processing technology, making these interactions seem rigid and simplistic. Conversely, un-humanized robo-advisors, with their formal and official interaction patterns, appear more sophisticated and reliable.

Further analysis showed that the stronger influence of un-humanized robo-advisors is primarily seen in individuals with low-risk preferences. These individuals tend to be more cautious and thus place greater trust in what they perceive as more reliable technology. High-risk individuals, being more confident and risk-taking, are less influenced by both types of robo-advisors.

Our findings suggest that investors are differently influenced by various robo-advisors based on their risk preferences. This insight is crucial for wealth management firms. Previous research often highlights the benefits of humanizing technology (Burgoon et al., 2000; Waytz et al., 2014), but our study reveals potential drawbacks in financial advising.

6 Conclusion

Our research found that un-humanized robo-advisors have a stronger impact on investors with low-risk preferences compared to humanized ones. This finding enhances understanding of human-computer interaction and addresses issues in anthropomorphic interactions in finance, offering valuable insights for wealth management companies.

References

- Barber, B. M., Odean, T.: Boys will be boys: Gender, overconfidence, and common stock investment. *Q. J. Econ.* **116**(1), 261–292 (2001)
- Belanche, D., Casaló, L.V., Flavián, C.: Artificial intelligence in FinTech: Understanding robo-advisors adoption among customers. *Ind. Manag. Data Syst.* **119**(7), 1411–1430 (2019). <https://doi.org/10.1108/IMDS-08-2018-0368>
- Buccioli, A., Zarri, L.: Do personality traits influence investors' portfolios? *J. Behav. Exp. Econ.* **68**, 1–12 (2017). <https://doi.org/10.1016/j.socec.2017.03.001>
- Burgoon, J.K., Bonito, J.A., Bengtsson, B., Cederberg, C., Lundeberg, M., Allspach, L.: Interactivity in human-computer interaction: A study of credibility, understanding, and influence. *Comput. Hum. Behav.* **16**(6), 553–574 (2000). [https://doi.org/10.1016/S0747-5632\(00\)00029-7](https://doi.org/10.1016/S0747-5632(00)00029-7)
- Jian, J.-Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* **4**(1), 53–71 (2000). https://doi.org/10.1207/S15327566IJCE0401_04
- Jung, D., Dörner, V., Weinhardt, C., Pusch, H.: Designing a robo-advisor for risk-averse, low-budget consumers. *Electron. Mark.* **28**(3), 367–380 (2018). <https://doi.org/10.1007/s12525-017-0279-9>

- Knote, R., Janson, A., Eigenbrod, L., & Söllner, M. (2018). The what and how of smart personal assistants: Principles and application domains for IS research
- Rühr, A., Berger, B., & Hess, T. (2019). Can I control my robo-advisor? Trade-offs in automation and user control in (digital) investment management.
- Tassiello, V., Tillotson, J.S., Rome, A.S.: “Alexa, order me a pizza!”: The mediating role of psychological power in the consumer–voice assistant interaction. *Psychol. Mark.* **38**(7), 1069–1080 (2021). <https://doi.org/10.1002/mar.21488>
- Waytz, A., Heafner, J., Epley, N.: The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* **52**, 113–117 (2014). <https://doi.org/10.1016/j.jesp.2014.01.005>
- Wu, J., Cheung, H.Y.: Confirmatory factor analysis of DOSPERT scale with Chinese university students. *Psychol. Rep.* **114**(1), 185–197 (2014). <https://doi.org/10.2466/09.03.PR0.114k15w2>



Developing a User-Centric Stepwise Comment Moderation AI Service: Balancing Free Expression and Respectful Online Interaction

Jonghan Kim¹, Suhwan Jo², Woosung Jung², Manseo Kim¹, and Sung Park³✉

¹ Department of Human-Centered Artificial Intelligence, Sangmyung University, Seoul 03016, Republic of Korea

² Department of Electronic Engineering, Sangmyung University, Seoul 03016, Republic of Korea

³ Department of Emotion Engineering, Sangmyung University, Seoul 03016, Republic of Korea
sjpark@smu.ac.kr

Abstract. The rise in online trolling, marked by intentionally posting offensive messages to provoke or disrupt, poses significant challenges to digital interactions. Trolling negatively impacts psychological well-being and disrupts online communities. Our research focuses on AI-based moderation, moving beyond traditional word censorship to accommodate user diversity and mental states in content moderation. We incorporated the discover, define, develop, and deliver stages to create a user-centered comments moderation service. This service was informed and validated through user feedback. Our research included conducting user surveys ($n = 106$) and in-depth interviews (IDI, $n = 9$), revealing a diversity in preferences for comment moderation in existing services. Notably, many users expressed a preference for replacing offensive words with alternatives rather than entirely blocking comments. In response to these findings, we designed and developed a stepwise moderation service using current large language models, such as ChatGPT. This service empowers users to choose their preferred level of comment moderation, striking a balance between free expression and a respectful online environment. We tested a working prototype with potential users ($n = 9$) to evaluate its effectiveness. The results highlight the wide range of user needs and preferences in comment moderation. While there was positive feedback towards progressive comment moderation, it was accompanied by concerns about potential over-moderation and the need to preserve genuine communication. These findings underscore the critical need for nuanced moderation approaches that respect user diversity.

Keywords: sentence purification · comments moderation · large language model · generative AI · AI design · user-centered design

1 Introduction

Trolling is when someone posts messages to upset people, start arguments, or draw others into pointless debates, either subtly or blatantly. This behavior includes sharing unconstructive content designed to elicit reactions, trap targets in pointless arguments, and disturb the normal functioning of online communities [1].

Recently, platforms like YouTube adopted automatic detection and removal of malicious comments, aiming to create a safer online environment. This method helps filter harmful content in public forums or news articles. However, this automatic censorship has drawn criticism from users who feel their freedom of expression is being stifled, leading some to switch platforms or find ways to bypass the censorship algorithms for continued trolling. The challenge lies in balancing the minimization of trolling with preserving user freedom and addressing the limitations of current systems. In light of these findings, we hypothesized that reactions to moderated comments are likely to be diverse and multifaceted. Understanding and addressing this diversity is crucial to ensure that AI moderation systems are inclusive, effective, and sensitive to the needs of all users. To thoroughly investigate this hypothesis and to gain a deeper understanding of the varying user needs and responses, we embarked on a service design and development process.

2 Related Works

Studies on handling malicious comments have explored a range of methods. Such as, censoring or filtering techniques aim to prevent user exposure to such comments by not displaying inappropriate content. Another strategy is the outright disabling of the comment feature, which blocks users from posting comments altogether [2]. More recent research has focused on AI-based moderation, which advances beyond mere word or sentence censorship. Notably, the use of generative AI, particularly large language models (LLMs), has opened new possibilities for analyzing and transforming the nature of malicious comments [3].

Several studies [4–6] have introduced innovative methods for modifying text styles through generative AI. While these techniques hold significant potential, they may not universally meet the diverse needs and preferences of all users. For instance, Park's research [7] demonstrates that users' perceptions and reactions to content can vary greatly based on their mental state and personal experiences. These variations in user experiences underscore the complexity of effectively moderating online content.

Considering these findings, we hypothesized that reactions to moderated comments are likely to be diverse and multifaceted. Understanding and addressing this diversity is crucial to ensure that AI moderation systems are inclusive, effective, and sensitive to the needs of all users. To thoroughly investigate this hypothesis and to gain a deeper understanding of the varying user needs and responses, we embarked on a service design and development process.

3 Methods

3.1 Survey and Interview

We collected subjective responses from participants, aged between 20 to 50 years old, with 89% in their 20 s and the remainder in their 30 s to 50 s. Among them, 65 (61%) were men and 41 (39%) were women, totaling 106 participants.

Questionnaire items were diverse, aimed at gathering user evaluations on existing malicious comments detection methods and comment moderation assessments. Additionally, we conducted in-depth interviews with 11 individuals selected based on their susceptibility to comment influence compared to other survey participants. This diverse group included individuals like a composer who regularly shares their creations online.

3.2 Understanding User Needs

Through surveys and interviews, we have identified user requirements and perceptions in three key areas. Firstly, users express dissatisfaction with the current detection methods, highlighting their negative impact and the need for improvement. Many participants reported feeling discomfort and hostility when encountering online comments, raising concerns about the inadequacy of existing strategies.

Secondly, diverse preferences exist regarding comment moderation. While some users preferred filtering out only profanity, others favored replacing offensive words with alternatives. There was also a preference for modifying the overall tone of comments to make them less aggressive.

Lastly, some users express interest in reading the original malicious comments. For instance, one participant, a professional composer, shared a nuanced perspective on comment moderation, stating, *“I am in favor of moderating malicious comments. However, I also value seeing unfiltered feedback on my work, as such comments, even if critical, often motivate me to strive harder in my creative endeavors.”*

3.3 Designing The Service Concept

We designed and developed a stepwise moderation service using current large language models, such as ChatGPT to address the user needs defined earlier (see Fig. 1).

Each moderation level is characterized by its unique prompt instructions. When a sentence is input, it undergoes pre-processing, which includes removing special characters and normalizing words. After pre-processing, the sentence is sent to the corresponding model section based on the assigned moderation level. In this stage, the model moderates the sentence and generates a revised version. The moderated sentence then undergoes error detection in the post-processing section before being stored in the database. Simultaneously, it is pre-processed for the next moderation level. This cycle continues, with moderated sentences from levels 1 to 3 being sequentially processed and stored in the database. The moderation levels were defined based on user feedback gathered from surveys and interviews, which provided insights into effective moderation of malicious comments (refer to Table 1).

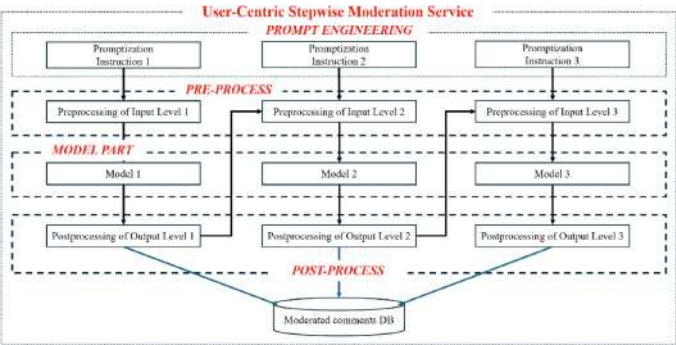


Fig. 1. The streamlined comment generation process of a stepwise moderation service

Table 1 Examples of sentence based on a stepwise moderation service

Moderation levels	Example	
Original comment	You're just a good for nothing jerk!	
Level 1	You're just a good for nothing @@ @!	
Level 2	You're just a good for nothing bro!	
Level 3	You're like a sloth. Put in more effort, please.	

At the “Original Comment” stage, the comment is displayed in its unaltered form. At “Level 1,” slang and swear words are filtered out and replaced with “@@” symbols. Moving to “Level 2,” the sentences containing “@@” symbols from the previous level are supplemented with appropriate words. Finally, at “Level 3,” an AI rephrases the sentences from Level 2, aiming to soften any aggressive or provocative expressions while also adding positive elements to enhance the overall tone. Therefore, users can adjust the moderation level from Level 1 to Level 3 according to their preference. The primary aim of our service is to mitigate psychological harm by moderating harmful content such as criticism, profanity, and offensive language, without resorting to complete removal. This service empowers users to choose their preferred level of comment moderation, striking a balance between free expression and a respectful online environment.

4 User Evaluations

To gather user feedback on our service, we developed a community website that serves as a representative platform. On this site, users went through several steps as part of their interaction with our service. Initially, they read the main articles along with the existing comments. Subsequently, they had the option to adjust the moderation settings from level 1 to level 3, thereby observing the changes in the comments at each level. After interacting with the service, we presented users with a set of questions designed to gather feedback on their experience.

To enhance user engagement and provide a realistic experience, we populated the website with actual articles and comments sourced from the internet. This approach encouraged users to naturally engage with the content, immersing themselves in the experience. Importantly, it allowed users to experiment with different moderation levels, observe how comments were altered at each stage, and determine which level best suited their preferences.

We curated a set of questions aimed at capturing comprehensive user experiences and opinions about our service. We received insightful responses from nine individuals who interacted with the service on our community website.

Following their experience with the service, two distinct response categories emerged. While the majority had a positive reaction to the concept, some expressed negative views, particularly regarding the third level of moderation, which involved softening any aggressive or provocative expressions and adding positive elements to enhance the overall tone. A few examples of participants' responses are as follows:

"Functionally, using it was smooth. I was quite amazed by the step-by-step moderation feature and how well it worked in practice." (Participant #6).

"I like that I can choose the level of moderation for malicious comments, allowing me to select different levels based on my purpose." (Participant #1).

"I think moderation was done well, but there were instances in Level 3 where it changed to praise, altering the meaning significantly." (Participant #4).

When considering potential utilization in daily life, participants primarily focused on the service's suitability for children and creators. A notable number of respondents ($n = 5$) believed that the service could be beneficial in contexts involving children, helping to shield them from inappropriate language in the content they view.

"It would be great if this could be used for YouTube comments that are suitable for children to watch." (Participant #8).

"I think that parents with children are often concerned that their kids might learn inappropriate language when using the internet. Having a service like this could be helpful in addressing such issues." (Participant #9).

Overall, while there was enthusiasm for the innovative approach of our moderation service, there were concerns about over-moderation and the preservation of authentic communication. These insights are invaluable for refining the service, ensuring it strikes the right balance between protecting users and maintaining the integrity of online discourse.

5 Discussion

Most feedback indicated a positive reception towards the concept of progressive comment moderation, suggesting our service could play a role in mitigating the prevalent issue of aggressive online culture. Our approach, which transcends simple censorship, considers both the author's right to expression and the reader's desire for a positive online experience. This method allows users to adjust the moderation level to their preference, thereby preserving freedom of expression while minimizing the harm caused by malicious comments.

Nonetheless, our study faced two primary limitations, leading to some negative feedback. The first was related to inaccuracies in translations and interpretations, particularly at Level 3 moderation, where AI sometimes produced awkward or misunderstood sentences. This issue was compounded by the lack of context, as isolated comments could be misinterpreted without the accompanying main article. For instance, sarcastic comments could be misconstrued as genuine praise. The second limitation concerned the inability to stay current with the latest internet slang and terminology. The AI model, with a cutoff point for its data, failed to recognize newly coined words and phrases, resulting in awkward or inaccurate moderations. This was particularly evident with trending yet mildly toxic words or memes, which were sometimes incorrectly classified and overly moderated, stripping away the original humor or creativity.

Future research could explore measuring the similarity between original and moderated comments to assess effectiveness more accurately. Additionally, further refining the moderation levels could offer more personalized experiences. Future studies might also test the effectiveness of this service in preventing the amplification of malicious comments, known as the echo chamber effect [8]. Despite the challenges, our study provides a foundation for developing more nuanced and effective online comment moderation solutions.

Acknowledgement. This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF2022R1I1A1A01066657).

References

1. Bishop, J.: The psychology of trolling and lurking: The role of defriending and gamification for increasing participation in online communities using seductive narratives. In: Li, H. (ed.) *Virtual Community Participation and Motivation: Cross-Disciplinary Theories*, pp. 160–176. IGI Global, Hershey, USA (2012). <https://doi.org/10.4018/978-1-4666-0312-7.ch010>
2. Lee, S.H.: Biased artificial intelligence: Analyzing the types of hate speech classified by 'cleanbot', NAVER AI for detecting malicious comments. *J. Cybercom. Acad. Soc.* **38**, 33–75 (2021)
3. Axelsen, H., Jensen, J. R., Axelsen, S., Licht, V., Ross, O.: Can AI Moderate Online Communities? *Work in Progress*
4. Atwell, K., Hassan, S., Alikhani, M.: APPDIA: A discourse-aware transformer-based style transfer model for offensive social media conversations (2022)

5. Nogueira dos Santos, C., Melnyk, I., Padhi, I.: Fighting offensive language on social media with unsupervised text style transfer. (2018)
6. Wang, K., Hua, H., Wan, X.: Controllable unsupervised text attribute transfer via editing entangled latent representation. (2019)
7. Park, M., Mcdonald, D. W., Cha, M.: Perception differences between the depressed and non-depressed users in Twitter (2013). www.aaai.org
8. Cinelli, M., et al.: The echo chamber effect on social media (2021). <https://doi.org/10.1073/pnas.2023301118/-/DCSupplemental.y>



Effective Design of Gauge Cluster for Improving Information Recognition

Sungju Kim¹, Seeung Oh², and Youngjae Im^{1,3}(✉)

¹ Department of Ergonomics Engineering, Dong-eui University, Busan, South Korea
ergoim@deu.ac.kr

² Department of Engineering, South Carolina State University, Orangeburg, USA

³ The Research Institute of UX, Dong-eui University, Busan, South Korea

Abstract. This study aims to investigate the impact of analog and digital data types on human information cognition and the effective design of gauge cluster displays. To achieve this, experiments were conducted with analog and digital displays of the gauge cluster. Initially, the experiment involved a computer screen simulation of driving with an eye-tracking device to analyze eye movements. The Gaze-point Analysis software was utilized, which records and analyzes eye movements as participants watched an 85-s driving simulation with either an analog or digital display on the monitor. They took a survey right after the simulation to evaluate their attention and preference for different displays. The result showed that participants had better performance and a stronger preference for the digital display than the analog display, according to the survey result. In conclusion, the digital data type of the gauge cluster display is more effective and efficient because it performs better with shorter eye movements. These findings can contribute to improving gauge cluster designs to increase safety by understanding users' eye movements.

Keywords: Gauge Cluster · Information Recognition · User experience

1 Introduction

Recently, with the development of vehicle technology, the use of in-vehicle displays has increased. As the in-vehicle digital display market gradually expands, the amount of information displayed is increasing, and the form of information is gradually changing dynamically [1]. The gauge cluster, in which visual information transmission is most actively performed concerning vehicle driving, also increases the amount of information and uses various expression methods as it changes from the analog form to the digital display form. This increase in the amount of information and various expression methods can cause cognitive load from the perspective of drivers who need to keep an eye on the front. For that reason, in terms of displaying and delivering information related to vehicle driving, the cluster must be designed to display various information accurately, easily, and quickly to the driver [2]. Looking at previous studies related to gauge clusters, most of them are comparative analyses of visibility according to the complexity of information,

menu structure, and layout [3–5]. Research on the comparative analysis of visibility according to the information display method is insufficient. The main background of this study is the comparative analysis of visibility according to the information display method of gauge clusters that deal with important vehicle driving information. It has been studied to compare digital and analog information about clocks [6] and scales [7] with information display methods similar to gauge clusters, and it also emphasizes the importance of information display methods. When analyzing previous research and trends related to gauge clusters, it was possible to derive the results that there are two main types of information display methods, analog and digital. Therefore, this study aims to investigate the effect of two types of data, analog and digital, on human information perception and the effective design of gauge cluster displays.

2 Survey on Users' Priority of Vehicle Functions

To produce a gauge cluster image for an information display method (analog, digital) comparison experiment, this study investigated the factors that drivers see or consider important in driving situations. The reason for the survey is that at least 20 to 30 pieces of information are marked within a limited area of the instrument panel [8], so there is a limit to handling all this information in the actual driving situation. Among these various pieces of information, this study selected the following six pieces of information as the main vehicle operation information that is most frequently communicated visually. The selection of the six pieces of information was based on the research on digital instrument panel information structure [4].

The survey was conducted with about five minutes of online participation with 40 people with actual driving experience. The image data used in the survey were referenced in the open explanatory data of the domestic Hyundai Motor Company. This image data contains six key components of the instrument panel. 1. Tachometer 2. Speedometer 3. Cooling water thermometer 4. Fuel meter 5. Warning lamp and indicator lamp 6. Driving information screen (trip computer). It consisted of questions to select the most viewed 3 information during actual driving and questions to list priorities from the most important information. As an answer to the first question, the most viewed information among actual driving situations was commonly selected as numbers 2 and 6. The second question, the answer to the priority list, was 2 and 5 in common.

3 Main Test

3.1 Participants

The experiment was conducted with about 15 undergraduate students from South Carolina State University. The selection criteria for participants in the experiment are as follows. Participants need to be 18 years of age or older, have normal or corrected to normal vision, and be free of current or past neurological or psychiatric disorders. Approvals for the study (SCSU IRB #: 2024-01-29-SO) from the IRB of South Carolina State University. Next, a 10 min direct presentation on research and experiments was conducted to help participants understand.

3.2 Apparatus

This study used an eye-tracking device, the Gaze-point GP3-HD, recording at 150 Hz (Gaze-point, Vancouver, British Colombia, Canada) on a small tripod below a 29 inch flat screen monitor (LG, South Korea) at a native 2560 x1080 resolution. The eye tracker was placed at a 45° angle and 50 cm to 70 cm from the participant's eyes, followed by the Gaze-point instruction.

3.3 Design Alternatives and Response Measures

As mentioned in the introduction, most of the previous studies were a lack of research on visibility based on how information is displayed. Therefore, this study analyzed the changes and trends in gauge clusters by dividing them into two types: analog and digital. Figure 1 shows the final two types of gauge clusters using Illustrator, a design tool that draws on gauge cluster design trends from domestic and international automotive companies.



Fig. 1. Final Analog and Digital Cluster Design

The information in the gauge cluster was designed to reflect the results of an earlier survey of 40 experienced drivers about the key components of the gauge cluster. Based on the results, in both questions, the speedometer and trip computer were the most commonly viewed and most important. Therefore, this made dynamic changes to the speedometer and trip computer in the experimental video and made the same changes to the tachometer because it is related to the speedometer. In addition, the highly important warning lights were represented by two types of fuel gauge and coolant temperature gauge, which are considered core vehicle driving information.

The gauge cluster design video for the experiment was created to closely resemble a real driving situation. The video was created using After Effects, a video editing and design tool, and was created to closely resemble real-world driving situations.

This study analyzed the eye movements in the gauge display area using the Gazepoint Analysis software. Before recording any eye movement, an Area of Interest (AOI) was marked in the gauge display area to calculate quantitative eye movement measures. Using the AOI in the gauge display area, this study analyzed FD (Fixation Duration), TFF (Time to First Fixation), FC (Fixation Count), and RC (Revisit Count).

3.4 Task and Procedure

The total experimental time was approximately 1 hour per participant. The purpose of the experiment was to analyze the eye movement of analog and digital gauge cluster

mp4 videos produced using the eye tracker instrument. In the experimental procedure, the participants first filled out the consent form. After that, the eye tracker equipment to be used in this experiment was calibrated to suit the participants. Before taking any task, participants completed an eye calibration using a nine-point grid using the Gazepoint Control software. They repeated until they passed equal to, or above, six from nine points for both eyes. After completing the eye tracker calibration, the participants watched two types (analog & digital) of 85 s experimental videos. Participants took a survey right after the simulation to evaluate their attention and preference for different displays (Fig. 2).



Fig. 2. Example of the Gazepoint GP3-HD and a 29-inch flat-screen monitor

The survey questions consisted of Q1. What was the maximum speed while driving? Q2. How much fuel was left while driving? Q3. What was the highest RPM on a car? Q4. What was the temperature in the gauge on a car? with four. Next, preference questions consisted of Q5. Which one do you prefer (analog or digital display)? with one.

Each experimenter conducted an analog or digital display sequence randomly, and half of the experimenters conducted the experiment in the analog-digital order, and half of the experimenters conducted the experiment in the digital-analog order. At the end of one experiment, participants were asked to answer four survey questions. After receiving the responses, there was a 5 min break. Following the completion of both experiments, a preference survey was conducted.

4 Results

The results of the eye tracking analysis are as follows (Fig. 3).

The average FD (Fixation Duration) is 12.32747 s in the digital display, the average FD is 24.546 s in the analog display. The average TFF (Time to First Fixation) is 0.903467 s in the digital display, but the average TFF is 1.5395333 s in the analog display. While the average FC (Fixation Count) is 33.267 in the digital display, the average FC is 41.53333 in the analog display. The average RC (Revisit Count) is 22.46667 in the digital display, but the average RC is 28.733333 in the analog display. According to the t-tests, FD, TFF, and FC are considered to be extremely statistically significant. However, the RC is not considered to be statistically significant. The heatmaps show that participants pay more attention to the analog display than the digital display (See Fig. 4). Red spots on the analog display indicate high attention, while the digital display has more blue spots, indicating low attention.

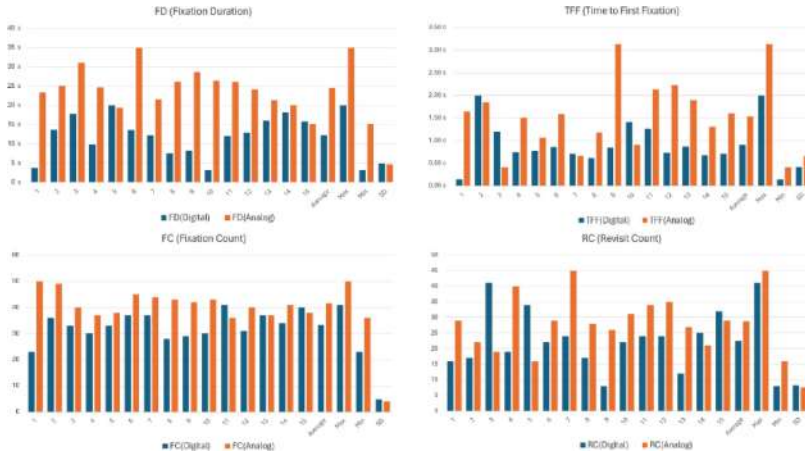


Fig. 3. Comparisons of FD, TFF, FC, and RC in digital and analog display

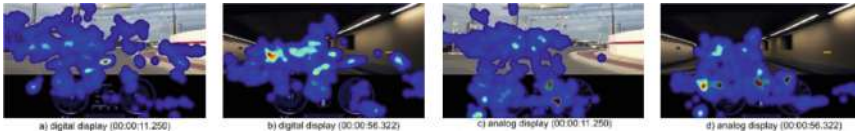


Fig. 4. Heatmaps in digital and analog display

According to the survey results, participants showed 60% performance in the digital display and 40% in the analog display. In addition, participants showed an 80% preference for the digital display and 20% for the analog display.

5 Conclusion

According to the t-test, the results of FD (Fixation Duration), TFF (Time to First Fixation), and FC (Fixation Count) are statistically significant. However, the result of RC (Revisit Count) is not quite statistically significant, meaning there is no apparent difference between digital and analog displays.

First, visual attention can be related to attraction and preference. The digital display has a shorter TFF (Time to First Fixation), meaning it has visual attention faster than the analog display. In addition, according to the survey results, participants showed an 80% preference for the digital display and 20% for the analog display.

Second, visual attention can be related to information processing time. In this regard, there is a study on visual attention to find out the cognitive processing process of animation characters [9]. The digital display has a shorter FD (Fixation Duration) and fewer FC (Fixation Count), meaning it does not need a long time to process information. Also, it should consider this as a driving situation. In addition, the heatmaps visualized participants with longer eye gazes on the analog display than on the digital display. According to the survey results, participants showed 60% performance in digital displays and 40%

in analog displays. Even though participants had a shorter FD (Fixation Duration) and fewer FC (Fixation Count) on the digital display, they showed a better performance in understanding and remembering information in the gauge display.

In conclusion, the survey results showed that participants performed better and preferred the digital display more than the analog display. The eye-tracking analysis results showed a shorter FD, shorter FTT, and fewer FC in the digital display than in the analog display. Therefore, the gauge cluster display's digital data type is more effective and efficient because it performs better with shorter eye movements.

References

1. Gu, B., Ju, D.: Usability evaluation according to instrument cluster GUI elements. In: Korea society of automotive engineers spring conference, pp. 1222–1222 (2017)
2. Koo, S.: An observation on implications of instrument cluster design for aged drivers. In: Korea society of automotive engineers fall conference & exhibition, pp. 784–788 (2016)
3. Choi, J., Ji, Y.: The effect of visual complexity and driver's age on information search performance. *J. HCI Soc. Korea*. (2013)
4. Lee, Y., Park, S.: Study on information structure of digital instrument cluster. In: KSDS conference, pp. 344–349 (2015)
5. Kim, H., Jung, K., Lee, D.: A Study on the Menu Type of Instrument Cluster IVIS. *J. Ergon. Soc. of Korea*. **32**(2), 189–198 (2013). <https://doi.org/10.5143/JESK.2013.32.2.189>
6. Kim, K., Chung, D., Jang, J.: A comparative study of consumers' preference for digital and analog watch design. *J. Digit. Des.* **12**(3), 485–495 (2012)
7. Kim, K., Chung, D., Jang, J.: A study on the comparative evaluation of consumer preferences in digital and analog design - focusing on the consumer eye measurement and analysis of digital and analog style products. *Ind. Des. Stud.* **6**(3), 33–42 (2012)
8. Kim, S.: A study on essential elements for visibility enhancement of vehicle cluster gauge - focused on South Korean sedans - the treatise on the plastic Media, 10(3), 81–88 (2007)
9. Hwang, M., Kwon, M., Park, M., Yin, S.H.: A study on the visual attention of popular animation characters utilizing eye tracking. *J. Korea Cont. Assoc.* **19**(6), 214–221 (2019)



Designing the User Interface for an Elderly-Friendly Taxi-Hailing Mobile App: A Preliminary Study

Yura Kim^(✉), Dongha Kim, Saeyeon Kim, Yeonji Kim, Gyuhee Kim, Chaewon Park,
and Woojoo Kim^{ID}

Division of Liberal Studies, Kangwon National University, 1 Gangwondaehak-gil,
Chuncheon-si 24341, Gangwon-do, South Korea
woojoo.kim@kangwon.ac.kr

Abstract. This study addresses the challenges faced by elderly individuals in accessing transportation services, particularly in the context of Korea's transition to a super-aged society. Despite government efforts, insufficient transportation alternatives and digital literacy pose significant hurdles for elderly users, aggravating their transportation vulnerability. To tackle these issues, this study aims to identify the needs of elderly users regarding taxi-hailing mobile apps, design an elderly-friendly user interface (UI), and demonstrate its superior usability compared to an existing app. A preliminary survey and interviews with 20 elderly participants were conducted to gather demographic information and assess challenges with existing taxi-hailing apps. Based on these findings, a UI of a new elderly-friendly app named *DuriTaxi* was developed and compared with the market-dominating app (*Kakao T*) in usability tests among 20 elderly users. The results indicate that *DuriTaxi* outperformed *Kakao T* in all usability evaluation items, providing clearer text and image sizes, improved color perception and overall consistency. The findings highlighted the importance of intuitive UI design and font readability for elderly users, with *DuriTaxi* receiving positive feedback for addressing these issues. The study emphasized the need to consider elderly users' prior experience and preferences when introducing new technologies. The study underscores the importance of improving technology acceptance and usability among elderly users by prioritizing intuitiveness, readability, and familiarity with existing technologies. These insights can inform the development of taxi-hailing apps to better meet the requirements of elderly users.

Keywords: Taxi-Hailing · Ride-Hailing · Mobile User Interfaces · Usability · Elderly

1 Introduction

Aging problem is a global issue. Based on projections from the World Health Organization, by the year 2030, one out of every six people globally will be 60 years old or older [1]. As the global population ages, the need for technology solutions that cater

Y. Kim and D. Kim—These authors contributed equally to this work.

to elderly users becomes increasingly critical. Elderly individuals with impaired hearing and vision are classified as transportation vulnerable, as they encounter difficulties using both public transportation and private cars. Despite Korean government efforts to encourage voluntary license returns, the effectiveness of such initiatives is hampered by insufficient transportation alternatives, especially in areas with limited infrastructure. Rural elderly residents face challenges with accessing buses due to limited availability and physical barriers, necessitating alternative transportation solutions like taxis.

Some studies have aimed to improve the usability of taxi/ride-hailing apps. Rahman et al. [2] assessed the usability of three ridesharing apps (*Uber*, *Pathao*, and *Obhai*) in Bangladesh through heuristic evaluation and a user study with 60 participants, revealing several usability issues and finding that *Pathao* had the best overall usability. Hsu and Chen [3] investigated the usability of three representative ride-hailing apps (*Uber*, *Lyft*, and *Gojek*), and identified key design elements for optimal user experience, emphasizing the importance of clear signifiers, modular tabs, ad-free modes, streamlined access to frequently used functions, and inline editing capabilities. Zhou [4] evaluated the usability of a Chinese taxi-hailing app (*Didi Chuxing*), identifying issues such as inaccurate icon and text, excessive functional options, poorly placed marketing pop-ups, and unclear location details, and proposed improvement strategies.

In this era of digital transformation, traditional approaches such as phone call-based taxi-hailing are diminishing, causing difficulties for those not proficient in digital technology [5]. Taxi-hailing mobile apps poses a significant challenge for elderly users, with current apps lacking user-friendly features tailored to their needs [6]. Surprisingly, however, while several studies [7, 8] have shown the benefits of ride-hailing adoption by older adults, there has been a severe lack of research aimed at making taxi-hailing or ride-hailing apps more elderly-friendly. To our knowledge, only one study attempted to design age-friendly digital taxi-hailing products, identifying mental gaps in elderly users' interactions, and developing a prototype to bridge these gaps [9].

Therefore, the objective of this study is to identify the needs of the elderly when using a taxi-hailing mobile app, design a UI tailored to those needs, and to validate its usability for elderly users compared to the other typical taxi-hailing mobile app.

2 Preliminary Survey and Interview

A preliminary survey and interviews with 20 elderly people (5 males, 15 females) were conducted (Fig. 1). The age distribution of the participants was as follows: 3 individuals aged 60–69, 9 aged 70–79, and 8 aged 80 or over. Participants' demographic information including gender, age group, education level, living arrangements, health status, and monthly allowance, in addition to their transportation usage, outing frequency, and challenges encountered when using taxi-hailing apps were collected [10].

The results revealed that some elderly experienced accessibility and convenience issues with transportation. Despite taxis being the second most frequently used mode of transportation for the elderly after buses, they were experiencing problems in calling and using taxis. They not only felt the app's UI was overly complex, but also had a strong aversion to learning a new way of using the app, thereby wanting the app to work in a similar way to conventional phone call-based taxi-hailing services. Based on



Fig. 1. Snapshots of conducting the preliminary survey (left) and the usability evaluation (right)

the findings, we developed the UI for a new elderly-friendly taxi-hailing mobile app (*DuriTaxi*) using Figma (Fig. 3).

3 Usability Evaluation

3.1 Methods

Usability tests comparing *DuriTaxi* with the existing Korean market-dominating app, *Kakao T*, were conducted (Fig. 1) among a new group of 20 elderly people (6 males, 14 females; 5 aged 60–69, 6 aged 70–79, and 9 aged 80 or over). The participants experienced Figma prototypes of both *Kakao T* (Fig. 2) and *DuriTaxi* (Fig. 3) by following the given task sequence, and then responded to a usability evaluation questionnaire (Table 1). The questionnaire, adapted from [11], consisted of 11 questions related to visibility, consistency, aesthetics, user control, convenience, and satisfaction, which were rated on a 5-point Likert scale. Some follow-up interview questions were also asked to understand the survey results.



Fig. 2. The Figma prototype of *Kakao T*. The task sequence was “Touch the taxi icon” → “Check one’s current location” → “Search and enter the destination and confirm” → “Select vehicle type and payment method, and tap the call button” → “The information of the called taxi is displayed”

3.2 Results and Discussion

Table 2 and Fig. 4 show the usability evaluation ratings of *Kakao T* and *DuriTaxi*. Compared to *Kakao T*, elderly users responded that *DuriTaxi*’s font size and color were more suitable for information recognition, and that the size of the icons and buttons were



Fig. 3. The Figma prototype of *DuriTaxi*. The task sequence was “Touch the call button or the speaker icon for voice input” → “(For voice input) Call a taxi with voice” → “While the app is calling a taxi, you can tap the cancel button” → “The information of the called taxi is presented”

Table 1. Usability evaluation questionnaire

Categories	No	Questions
Visibility	1	Do font sizes and colors match the perception of information?
	2	Are icons and buttons properly sized and clearly displayed?
Consistency	3	Are screen styles (design and arrangement) designed to be consistent across all screens?

(continued)

Table 1. (continued)

Categories	No	Questions
	4	Do icons give a sense of unity with a consistent style?

(continued)

Table 1. (continued)

Categories	No	Questions
Aesthetic	5	Are texts and icons appropriately harmonized to provide information?

(continued)

more appropriate and clearly displayed. They evaluated *DuriTaxi* had higher screen style consistency and thematic unity of images, and better harmony between text and images with a balanced amount of information per screen. In addition, *DuriTaxi* was easier to use without additional education or support and had higher accessibility to desired menus

Table 1. (continued)

Categories	No	Questions
	6	Does it present the right amount of information on one screen?

(continued)

Table 1. (continued)

Categories	No	Questions
User Control	7	Is it possible to use the desired functions without additional training or assistance?
	8	Is the desired menu and function easily accessible?

(continued)

Table 1. (continued)

Categories	No	Questions
Convenience and Satisfaction	9	Are features and content properly organized?
	10	Do you think it would be easy to call a taxi if you use this application?
	11	Do you think this application will have a positive effect on enhancing the mobility rights of the elderly?

and functions. The composition of functions and content was also evaluated as more appropriate for *DuriTaxi*. Lastly, *DuriTaxi* was perceived easier to call a taxi and more helpful for improving elderly mobility.

Table 2. Mean (M) and Standard Deviation (SD) of usability evaluation ratings

Question No.		1	2	3	4	5	6	7	8	9	10	11
<i>Kakao T</i>	M	1.85	1.80	2.10	2.35	2.10	1.80	1.90	1.80	2.05	1.95	2.05
	SD	0.75	0.70	0.72	0.75	0.85	0.77	0.72	0.62	0.76	0.89	0.83
<i>DuriTaxi</i>	M	4.55	4.60	4.35	4.40	4.40	4.55	4.15	4.35	4.40	4.30	4.45
	SD	0.51	0.50	0.49	0.50	0.50	0.51	0.88	0.49	0.50	0.57	0.51

Overall, *DuriTaxi* received higher ratings in all usability evaluation items (Fig. 4). According to the Wilcoxon Signed Rank Test results, the differences in all items were statistically significant ($p < 0.001$ in all items). The average scores for *DuriTaxi* ranged from 4.15 to 4.60, significantly higher than *Kakao T*'s average scores of 1.80 to 2.35, with a substantial gap of over 2 in all items.

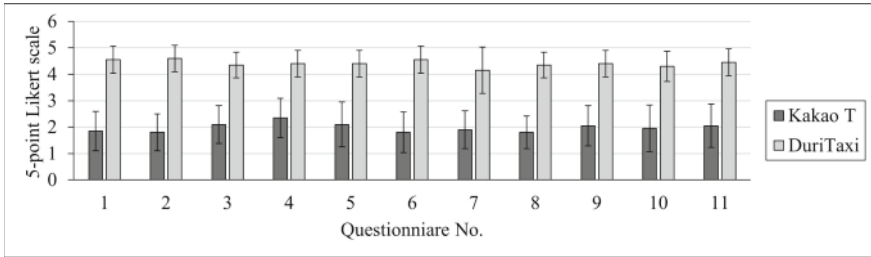


Fig. 4. Usability evaluation ratings of *Kakao T* and *DuriTaxi* (Error bars indicate SD)

In summary, elderly users found *DuriTaxi* to be superior to *Kakao T* in terms of clarity and suitability of text and image sizes, color perception, and overall consistency. *DuriTaxi* provided an appropriate amount of information, allowing easier access and utilization of desired menus, features, and content without additional support. As a result, it was evaluated as more accessible, easier to use, and effective in enhancing taxi-hailing services. According to the interview results, this is attributed to several key factors: *Kakao T*'s complex interface and icon design caused confusion and rejection among elderly users, highlighting the importance of an intuitive and simple user interface. Font size and readability are crucial for elderly users' technology use, and *DuriTaxi* received positive feedback from elderly users by addressing these issues.

It was found that elderly users tend to abandon or avoid new technology when familiarity is low, which is consistent with existing literature [12]. Some participants expressed a preference for traditional phone call-based taxi-hailing services over new technologies during interviews. Thus, considering elderly users' existing experiences and preferences when introducing new technologies is important.

This study is limited in the sense that the prototype was created using Figma, and the feasibility of implementing the app's features was not considered. Also, it was not tested for various usability issues that may arise in actual use. In future work, summative evaluations with tasks and a high-fidelity prototype based on more specific, detailed, and realistic taxi-hailing app usage scenarios could be conducted.

4 Conclusions

The study suggests important directions for improving technology acceptance and usability among elderly users, emphasizing the need for intuitive interfaces, appropriate font, and familiarity to existing technologies. Simplifying procedures, prioritizing readability, and using voice interfaces could make apps more akin to familiar phone call-based taxi hailing services, aligning more closely with elderly users' needs. These findings can help taxi-hailing app developers and designers better understand and meet elderly users' requirements.

Acknowledgements. This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2024-RS-2023-00260267)

References

1. World Health Organization: Ageing and Health., <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (2024). Accessed 30 May 2024.
2. Rahman, M.M., Sarker, A., Khan, I.B., Islam, M.N.: Assessing the usability of ridesharing mobile applications in Bangladesh: An empirical study. In: 2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), pp. 1–6. IEEE (2020)
3. Hsu, Y.-H., Chen, C.-H.: Usability study on the user interface design of ride-hailing applications. In: Marcus, A., Rosenzweig, E., Soares, M.M. (eds.) Design, User Experience, and Usability. HCII 2023. LNCS, vol. 14032, pp. 208–216. Springer, Cham (2023)
4. Zhou, C.: A study on application interface usability evaluation and improvement strategy of Didi Chuxing. J. Brand Des. Assoc. Korea. **21**, 275–286 (2023)
5. Czaja, S.J., Charness, N., Fisk, A.D., Hertzog, C., Nair, S.N., Rogers, W.A., Sharit, J.: Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (create). Psychol. Aging. **21**, 333–352 (2006)
6. Chae, Y.: “I keep getting a shady taxi”...there are no taxi apps for seniors. https://www.hani.co.kr/arti/society/society_general/1063448.html (2024). Accessed 30 May 2024.
7. Mitra, S.K., Bae, Y., Ritchie, S.G.: Use of ride-hailing services among older adults in the United States. Transp. Res. Rec.: J. Transp. Res. Board. **2673**, 700–710 (2019)
8. Irawan, M.Z., Joewono, T.B., Belgiawan, P.F., Chalermpong, S., Thaithakul, P.: Examining the ride-hailing adoption behaviors among older adults in an Indonesian city: The case of Yogyakarta. Transp. Res. Interdiscip. Perspect. **16**, 100729 (2022)
9. Gong, X., Zhang, J., Jiang, B.: The age-friendly design of taxi-hailing digital products based on mental Model. J. Beijing Inst. Technol. **25**, 181–192 (2023)
10. Lee, K., Seomun, J.: A study on the transportation welfare alternatives to enhance the right of mobility for the elderly in rural areas. Korean J. Local Gov. Adm. Stud. **27**, 247–268 (2013)
11. Cho, Y.-R., Paik, J.-K.: Usability test and design proposal for mobile applications to improve muscle strength for the elderly. J. Korea Multi-media Soc. **26**, 313–325 (2023)
12. Awan, M., Ali, S., Ali, M., Abrar, M.F., Ullah, H., Khan, D.: Usability barriers for elderly users in smartphone app usage: An analytical hierarchical process-based prioritization. Sci. Program. **2021**, 1–14 (2021)



Framework for Understanding User Experience of Foldable Smartphones

Yankuan Liu^(✉), Lili Dong^(✉), and Pei-Luen Patrick Rau^(✉)

Department of Industrial Engineering, Tsinghua University, Beijing, China
dongli2012@mail.tsinghua.edu.cn, rpl@tsinghua.edu.cn

Abstract. This study presents a comprehensive framework for understanding the User Experience (UX) of foldable smartphones. Through product research, user case studies, and expert evaluations, this study defines the UX of foldable smartphones and identifies three core elements unique to foldable smartphones: the hinge, the user, and the folding interface. Special emphasis is placed on the hinge's design, which is essential for preserving screen continuity, enhancing device portability, and ensuring user comfort during the folding and unfolding process. Additionally, ergonomic compatibility and psychological acceptability are identified as crucial for the user. Key aspects such as state consistency, intuitive compatibility, proper feedback, and unique features of the interface design are highlighted for facilitating smooth transitions across different folding states. The study's findings are validated through a case study examining the impact of hand size on UX, revealing that users with larger hands exert more pressure and experience greater fatigue. The proposed UX framework holds significant implications for driving innovation in foldable mobile device design, elevating the quality of user experience, and supporting the successful mainstream adoption of this emerging product category.

Keywords: Foldable smartphones · User experience · Human-computer interaction · Hinge

1 Introduction

The latest and rapidly expanding category of mobile computing is the foldable smartphone, which combines the enhanced productivity capabilities of a tablet with the portability of a conventional smartphone [1]. Foldable smartphones can be categorized into three types based on the folding method: inward folding, outward folding, and vertical folding. In contemporary times, prominent smartphone manufacturers including SAMSUNG, HUAWEI, vivo, OPPO, Google and various others have unveiled various types of foldable smartphones including inward folding, outward folding, and vertical folding to ensure different needs [2].

Compared to non-foldable phones, the mechanical structures and novel interaction forms of foldable smartphones have drawn attention from both the industry and academia. An investigation conducted to examine the effects of varying folding angles of foldable

smartphones on users' privacy concerns indicated that the larger screens may potentially give rise to supplementary privacy breaches [3]. Lee, Kyung, Choi, Yi, Kim, Choi and Lee [4] conducted experiments to investigate user preferences for creases and revealed that the dimensions of a user's hand do not have an impact on their inclination. However, the specific characteristics of the task at hand significantly influence the users' selection of the folding method [5]. Specific interface optimizations are necessary for foldable smartphones. A recent investigation conducted by Dai, Zhang, Dai, Liang, Du and Chen [6] revealed that the utilization of split keyboards in the unfolded mode of foldable smartphones yields notable reductions in input errors and typing time. It is noteworthy that while the academic community has undertaken certain research on user experience (UX) pertaining to foldable smartphones, these studies have primarily concentrated on particular facets, thereby leaving a comprehensive UX framework for foldable smartphones unproposed.

There are numerous factors that influence the UX of foldable smartphones; however, there is a lack of comprehensive research conducted by the academic community on these factors. The objective of this study is to develop a UX framework for foldable smartphones. This study focuses on the UX aspects unique to foldable smartphones that non-foldable phones do not possess, proposing essential elements of a UX framework for foldable smartphones, with a particular emphasis on the hinge mechanism through thorough product research, use case studies, and evaluations. In order to ascertain the validity of the proposed framework, a simple case study was designed and conducted. By utilizing the proposed framework for UX in foldable smartphones, service providers can systematically collect data and analyze the manner in which consumers engage with foldable smartphone products. This framework facilitates a more profound understanding of user requirements, acting as an important component to improve the attractiveness and market standing of their products or services.

2 Foldable Smartphone UX Framework

The foldable smartphone UX framework were developed through a four-step process: 1. Surveying existing foldable smartphone products; 2. Deploying a more in-depth examination focusing specifically on user complaints or issues encountered during usage; 3. Conducting an detailed task analysis with five researchers who possess expertise in the field of foldable smartphones and possess a strong understanding of human-computer interaction and UX theories for a duration of two weeks, utilizing the photo diaries method integrated with use cases; 4. Conducting brainstorming sessions with eight experts, consisting of professors, researchers engaged in case studies, and senior designers to derive the proposed UX framework. Finally, the validity of the proposed framework was verified through a case study. This structured research approach ensures theoretical rigor and empirical substantiation for the framework, providing pioneering insights into UX understanding and design guidance tailored to the novel form factor of foldable mobile devices.

As shown in Fig. 1, this study proposes a UX framework for foldable smartphones composed of three core elements: the hinge, the user, and the folding interface, which means the interface optimization for folding and unfolding state. The interdependence

of individual elements within the framework for foldable smartphones can collectively shape the UX.

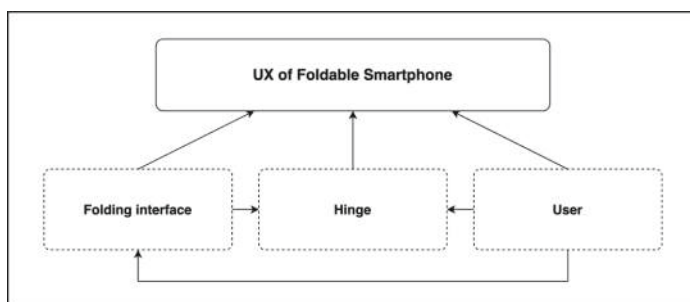


Fig. 1. Foldable smartphone UX framework

The hinge mechanism constitutes the most significant structural differentiation between foldable and non-foldable smartphones. As a critical component, the hinge exerts considerable influence on both the functionality and user experience of foldable devices. Its design facilitates seamless screen continuity when the device is unfolded, while also mitigating the formation of creases, thereby enhancing the overall visual quality. It also contributes to the device's slim profile and durability, making the phone more portable and comfortable to use. The hinge's design ensures ease of operation, aligning with users' natural force habits and can easily lock at key angles for stability. Ergonomic design is tailored to fit user's hand size and grip, optimizing comfort and improving long-term UX. Psychological factors like ease of learning, high error tolerance, and openness to innovation make the technology user-friendly and appealing. In terms of the user interface, state consistency between folded and unfolded modes ensures a smooth transition, enhancing overall interaction. The interface design matches cognitive expectations, making actions intuitive. Feedback mechanisms like sounds or visual signals during folding or unfolding enrich engagement. Fold-exclusive features provide unique utilities, such as specialized apps or functions, enhancing the overall experience by utilizing the foldable design.

3 Case Study: The Impact of Hand Size on the UX of Foldable Smartphones

This case study investigates the influence of users' hand sizes on the UX of foldable smartphones. The independent variable for the experiment is the participants' hand size, categorized into large and small. The dependent variable, the UX of foldable smartphone was measured by pressure applied by the fingers on the smartphone and subjective fatigue.

The foldable smartphone used in the experiment is vivo X Fold2. The pressure sensor for measuring finger pressure on the phone employs the FSR400. The two pressure point locations which are determined through task analysis, and pre-experiment, measure the

force exerted by the left and right thumb of the subject during the task. The Borg Fatigue Scale [7] was used to measure participants' subjective fatigue.

Before the task, participants were given time to engage with the smartphone. The formal experiment consisted of a continuous 3-minute folding task at their self-selected pace. The participants adhered strictly to the prescribed posture ensuring that their fingers maintained contact with the locations of the pressure sensors. Participants provided their current Borg fatigue scale values at the 1st, 2nd, and 3rd min of the folding task. The duration of the experiment ranged from 10 to 15 min.

A total of fourteen participants were recruited, with an equal distribution of genders and an average age of 20.83 ± 1.21 years. the Chinese Adult Body Size GB 10000-88 was utilized to determine the hand size. Participants with hand sizes exceeding the 90th percentile were categorized as large-handed, while those below the 50th percentile were categorized as small-handed. A total of seven participants were classified as large-handed (195.83 ± 10.71 mm), and seven participants were classified as small-handed (170 ± 7.43 mm). Every participant was right-handed.

4 Results

Regarding the pressure exerted by the fingers on the phone, the large-handed participants demonstrated significantly higher pressure values and greater fluctuation amplitude at the two touchpoints compared to the small-handed participants (P value $< .01$). This finding suggests that individuals with larger hand sizes exert greater force during the process of folding and unfolding the phone, resulting in increased physiological fatigue and subsequently leading to a decline in the UX of foldable smartphones. In terms of subjective fatigue, it was observed that large-handed participants experienced higher levels of subjective fatigue compared to small-handed participants, particularly during the third minute (P value $< .01$). Table 1 displays the descriptive results.

Table 1. The results of pressure and fatigue of big and small hand-size users

Variable	Time	Value	
		Big hand-size	Small hand-size
Fatigue	1 min	3.33 ± 1.37	3.50 ± 1.38
	2 min	4.33 ± 1.75	4.50 ± 1.22
	3 min	6.00 ± 2.00	5.17 ± 1.17
Pressure	1 min	78.17 g	48.00 g
	2 min	84.67 g	60.67 g
	3 min	61.17 g	38.83 g

The task lasted for 3 min. The pressure values of the participants with larger hands exhibited a consistent elevation (> 50 g) during the initial phases, followed by a gradual decline to approximately 55 g in the subsequent stages. In contrast, the pressure values

exhibited by the participants with smaller hands remained relatively low (< 50 g) during the initial phases, gradually rising to approximately 45 g in the later stages.

In general, the pressure and subjective fatigue data indicate that the UX of foldable smartphones for individuals with larger hands was initially worse. However, over the course of the task, the level of UX for participants with small hands gradually diminished, while the UX of both groups reached a comparable level.

5 Discussion and Conclusion

The case study outcomes affirm the proposed UX framework's comprehensive and systematic approach to understanding UX with foldable smartphones. The framework facilitated detailed user feedback on specific aspects, such as hinge design, screen continuity, and ergonomic factors, which are pivotal in shaping the overall UX. Device manufacturers can utilize this framework for product planning assessments to achieve foldable smartphones that more closely align with user habits and expectations. Researchers can design experiments based on the variables provided in the framework to conduct scientific inquiries surrounding foldable smartphones.

It should be noted that foldable smartphones are a type of mobile smartphone, and factors that affect the UX of regular phones will also impact the UX of foldable smartphones. Examples include hardware design factors like screen size, resolution, and refresh rate [8, 9], as well as design factors like software fluency and user interface [10, 11]. The framework proposed in this study focuses on factors that are distinct from non-foldable phones and specifically affect the UX of foldable smartphones, such as hinge. Therefore, the general factors are not included in the framework, for this framework is more of a technological UX centered around the hinge, according to Jang and Han [12].

The literature review has revealed that individuals with smaller hand sizes exhibit a higher susceptibility to fatigue and discomfort when engaging in stationary tasks with mobile devices [13]. Nevertheless, the empirical findings from the case study indicate that individuals with larger hands encounter greater fatigue. The potential explanation for this outcome could be attributed to the absence of dedicated attachment points for users' fingers during the folding and unfolding. Hence, participants endeavored to grasp the seams or apply pressure on both sides of the foldable smartphone to exert maximum force while unfolding. Participants with larger hands typically possess thicker fingers, which poses a challenge in gripping the seams or applying pressure on narrower edges, resulting in increased force exertion.

This study developed a UX framework for foldable smartphones, focusing on the hinge, user elements, and the folding interface. The hinge, a central pivot, directly affects device longevity and user engagement. User elements, including ergonomic compatibility and psychological acceptability, address interaction styles and user expectations. The folding interface ensures consistency and user-friendliness across different device states. Validated by a case study that analyzed the impact of hand size on UX, this investigation highlights ergonomic considerations in the design of foldable devices. Through the integration of these findings, this study affirms the comprehensive scope of the proposed framework in capturing the fundamental aspects of UX in foldable smartphones. This establishes a robust basis for further investigation and advancement in this evolving technological field. As foldable smartphones represent a transformative frontier in

mobile computing, this framework lays a crucial foundation for catalyzing user-centric innovations that can unlock the full potential of this burgeoning technology and shape the future landscape of personal digital experiences.

References

1. <https://www.theverge.com/foldables>
2. Huang, D.Z.: Analysis of the Impact of Foldable Mobile Phones Design on People's Lives. In: 2021 Int. Conf. Public Relat. Soc. Sci. (ICPRSS 2021), pp. 1155–1160. Atlantis Press
3. Jang, S.: User Experience Research on UTILization of the Folding Screen on a Foldable Phone for Privacy. (2021)
4. Lee, S., et al.: Where to put the creases? Interactions between hand length, task, screen size, and folding method on the suitability of hand-held foldable display devices. *Ergonomics*. **62**, 723–733 (2019)
5. Tan, Z., Su, Y., Dai, N., Gao, X., Du, L., Chen, Q.: Research on User Needs for Gesture Interaction of Foldable Smartphones: Comparison Between Current and Potential Users. (2023)
6. Dai, M., Zhang, J., Dai, N., Liang, K., Du, L., Chen, Q.: Performance Evaluation of QWERTY Keyboards on Foldable Smartphones: Keyboard Layout and Phrase Complexity. (2023)
7. Borg, G.A.: Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* **14**, 377–381 (1982)
8. Raptis, D., Tselios, N., Kjeldskov, J., Skov, M.B.: Does size matter? Investigating the Impact of Mobile Phone Screen Size On Users' Perceived Usability, Effectiveness and Efficiency. In: Proc. 15th Int. Conf. Hum.-Comput. Interact. Mob. Devices Serv., pp. 127–136. (2013)
9. Zinner, T., Hohlfeld, O., Abboud, O., Hoßfeld, T.: Impact of frame rate and resolution on objective QoE metrics. In: 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX), pp. 29–34. IEEE (2010)
10. Lin, Y.-D., Chu, E.T.-H., Wen, C.-L., Lai, Y.-C., Chen, I.-C.: Benchmarking handheld graphical user interface: Smoothness quality of experience. *Comput. Electr. Eng.* **68**, 76–91 (2018)
11. Anderson, J., McRee, J., Wilson, R.: Effective UI: The Art of Building Great User Experience in Software. "O'Reilly Media, Inc." (2010)
12. Jang, H., Han, S.H.: User experience framework for understanding user experience in blockchain services. *Int. J. Hum.-Comput. Stud.* **158**, 102733 (2022)
13. Gustafsson, E., Coenen, P., Campbell, A., Straker, L.: Texting with touchscreen and keypad phones-A comparison of thumb kinematics, upper limb muscle activity, exertion, discomfort, and performance. *Appl. Ergon.* **70**, 232–239 (2018)



Cognitive Analyses for Printing Tasks Using DTF Technology: A Mental Workload and Human Error Approach

Aide Aracely Maldonado-Macías¹ , Gabriela Pérez Potter² ,
and Mónica Gabriela Gutiérrez-Hernández¹ 

¹ Department of Industrial and Manufacturing Engineering,
Autonomous University of Ciudad Juárez, 32310 Chihuahua, Mexico
amaldona@uacj.mx

² Department of Industrial Design, Autonomous University of Ciudad Juárez, 32310
Chihuahua, Mexico

Abstract. Due to globalization, jobs in the textile industry have undergone significant technological changes and complex human-machine interfaces. Among these changes, the design and printing of textile garments offered greater versatility and printing quality. However, human factors that influence the interaction of this technology are still unknown. Because of this, high levels of mental work and the presence of errors are frequent. In this sense, the use of techniques for the evaluation of the mental workload, as well as the identification of human errors, may improve the process efficiency, user interaction, and usability. By developing a case study, this work uses cognitive evaluation methods to determine the mental workload and human error in the operation of Digital Transfer Film (DTF) printing machines. Five users voluntarily participated in the study. The methodology is made up of three stages. First, the hierarchical task analysis (HTA) was developed, followed by the Task Analysis for Error Identification (TAFEI), and finally, the mental workload was evaluated through the Workload Profile (WP). As a result, 12 illegal transitions were identified and related to errors. Additionally, those tasks that require complex set-ups by human operators to improve printing quality results demand the highest mental workload. Based on the results, it is possible to provide conclusions and recommendations that facilitate the interaction between people and DTF printing machines. The results highlight the need to have defined procedures to improve human interaction, machine functioning and training.

Keywords: Mental workload · HTA · TAFEI · Workload Profile

1 Introduction

Textile printing applies ink to textile surfaces using specialized printing processes and machines. One of the advances has been made in digital printing, through which a greater diversity of colors and dyes can be recreated and excellent printing fastness

(Ragab et al., 2022). Although there are several techniques, their choice depends on many factors: the number of pieces, colors, and dimensions of the image, the characteristics of the fabric to be used and its variety, and the characteristics of the design itself to be printed (Prybeha et al. 2021a); as well as the company's investment capital and objectives for its commercialization. Among the latest launches and novelties for the product customization sector is the technique known as DTF, whose acronym comes from the English name "Digital Transfer Film" or "Direct to Film" (br Importaciones 2022; Bou 2023). Today, there are transfer systems, which include a DTF printer, a roll holder, an adhesive application system, a dryer or drying oven, as well as a mechanism for rolling the finished transfer onto a roll (Prybeha et al., 2021b). However, Akbar et al. (2023) state that, although DTF printing can be faster and more varied than other types of textile printing, there are difficulties in user interaction that influence the result and the tasks to be performed. In this sense, developing analyses aimed at human-artifact interaction can help find opportunities for improvement in cognitive aspects and the configuration of the tasks themselves. Therefore, it is necessary to analyze the work process from the cognitive perspective, applying methods for a deeper analysis of the task, mental workload, and human errors to detect possible problems and deficiencies in the human-artifact interaction; this serves as a basis to improve the user experience and training.

Consequently, there is a variety of methods addressing cognitive analyses. The literature highlights the importance of performing a Hierarchical Task Analysis (HTA). This technique is one of the best-known and most widely used task analyses since its emergence in the late 1960s (Torres-Medina, 2020). HTA can be used previous to mental workload studies. Mental workload refers to the tensions induced in a person by the demands of the cognitive work they perform and represents the proportion of mental resources demanded by a task or set of tasks (Tsang and Velazquez, 1996; Stanton et al., 2013). One of the most commonly used methods to evaluate it is the Workload Profile (WP). It emphasizes dual-task assessments and subjective procedures for assessing mental resources. The method asks the subjects to provide the proportion of attentional resources used after they had experienced all of the tasks to be rated. The tasks to be rated are listed according to the HTA diagram and shown in a list in this order using a rating sheet proposed by Rubio et al., 2004. The workload dimensions used by this method are based on the resource dimensions in the multiple resource model of Wickens (1987): perceptual/central processing, response selection and execution, spatial, processing, verbal processing, visual processing, auditory processing, manual output, and speech output. Subjects rated the dimension that was previously defined and explained to them. In each cell on the rating sheet, subjects provide a number between 0 and 1 to represent the proportion of attentional resources used in a particular dimension for a particular task. A rating of "0" means that the task placed no demand on the dimension being rated; a rating of "1" means that the task required maximum attention. The ratings on the individual dimensions are later summed for each task to provide an overall workload rating (Tsang and Velazquez 1996, Rubio et al., 2004, Durán-Coronado et al. 2019).

In addition, human error is defined as all cases in which a planned sequence of physical activities does not achieve the expected result and when these errors cannot be

attributed to the intervention of a random factor (Stanton et al., 2013). In this case study, human error in complex production processes such as the DTF technique becomes an element of interest. Among the most commonly used methods for the analysis of this factor is Task Analysis for Error Identification (TAFEI), which allows the identification of errors in the use of devices by modeling the interaction in analysis and assumes that the device offers information about its functionality (Durán-Coronado et al. 2019). Based on the above, the objective of this work is to apply cognitive evaluation methods to determine the mental load and human error present in the operation of DTF printing machines through the development of a case study, and this serves as a basis for identifying possible design deficiencies that make it difficult to use.

2 Methodology

2.1 Participants

In the mental workload study and human error, 5 participants (2 technicians and 3 operators) with different levels of experience in using and operating DTF machines were considered.

2.2 Methods

The methodology consisted of three stages. Figure 1 shows the three stages that made up the study. This methodology is based on a similar case study proposed by Durán-Coronado et al. 2019.

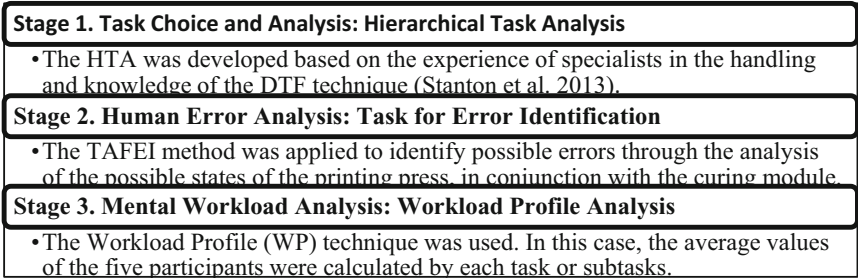
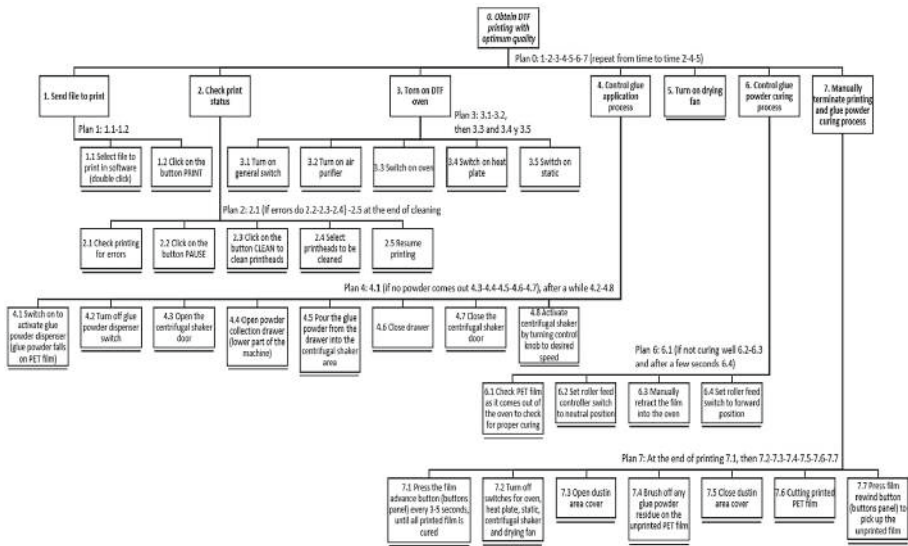


Fig. 1. Stages that shaped the methodology of the case study



Appendix 1. HTA developed for this case study

3 Results

Stage 1 Task Choice and HTA Results. The HTA results are shown in Fig. 1, which reveals the main task of printing the DTF file in the DTF printer on the Audley printer decomposed into seven subtasks and thirty-one sub-sub-tasks.

Stage 2 Human Error Analysis and TAFEI Results. The task analysis for error identification (TAFEI) was conducted from the HTA. The State-Space Diagrams (SSD) were developed, and 15 possible states were obtained from DTF interaction. Subsequently, the transition matrix showed 12 situations where errors, or so-called illegal transitions, can occur. Most are directly related to the “Printing” state and its transition to later non-consecutive stages, where the results are severely affected.

Stage 3 Mental Workload Analysis Results. Mental Workload Analysis. This analysis aims to identify those tasks or subtasks that may represent a source of mental workload for achieving the objective of printing a DTF file. The HTA performed in Stage 1 was used, using the 31 subtasks identified. Once the participants filled out the form, it was processed and generated based on the final sum of each task evaluated by each participant. It indicates that the five subtasks with the highest mental load, based on the calculated averages, in a descendent order were: Subtask 6.4: Place the roller feed switch in the previous position (value 4.56), Subtask 6.2: Place the roller feed controller switch in the neutral position (value 3.84), Subtask 3.4: Turn on the Iron Switch (value 3.40), Subtask 2.1: Display print for errors (value 3.70), Subtask 7.3: Open protective glass cover (value 3.60). Accordingly, those tasks that require complex set-ups by human operators to improve printing quality results demanded the highest mental workload.

4 Conclusions

The implementation of the methodology proposed in this project helped to understand better the task execution and the interaction with DTF technology, as well as to detect those factors that could influence the correct use of it. Therefore, to propose design solutions to the errors detected and the actions that cause a more significant mental workload. Based on the results, these techniques can be beneficial for designing machines for industrial use, where the artifact and the user can develop a more efficient, secure interaction and an improved user experience, minimizing error and avoiding losses of material and time.

Likewise, it is necessary that when interacting with this technology for the first time, the personnel can be adequately trained. Thus, this study can help companies provide and design a guide for the use and troubleshooting of DTF technology, which serves as a basis for human error identification that may exist during the production process, improving human-machine interaction and performance and reducing mental workload. The application of cognitive analysis methods can ensure the correct use of products and interfaces used by people, increasing their usability level and reducing frustration during their use. Although this work presents a methodology to analyze the relevant tasks in the use of DTF, it is necessary to continue deepening the development of these methods towards this activity and execute them with more participants to obtain results of greater scientific rigor. Likewise, it is important to implement these methods with novice and expert users since their experience will provide more specialized information on the subject. Cognitive analyses can help obtain better results regarding the training guidelines for new users and propose changes in design that can benefit companies by improving performance, reducing waste, and enhancing man-machine interaction.

References

- Akbar, A.A.M., Suryadi, G.S., Nur, C.E.L.: Implementasi teknik cetak sablon metode dtf pada kaus dengan desain ilustrasi bertema budaya betawi. *Kreator*. **9**(2), 24–31 (2023). <https://doi.org/10.46961/kreator.v9i2.632>
- Bou, Adria. 2023. “Impresión DTF (directa sobre film): todo lo que necesitas saber”. Printful. el 5 de octubre de 2023. <https://www.printful.com/es/blog/que-es-la-impresion-dtf>
- br Importaciones. 2022. DTF (Direct Transfer Film) en BR Importaciones. br Todo para tu Emprendimiento de Personalización. el 15 de febrero de 2022. <https://brimportaciones.com.uy/blog/post/dtf-lo-ultimo-en-estampado-textil.html>
- Durán-Coronado, A.A., et al.: Análisis cognitivos de carga mental e identificación del error humano para mejorar la experiencia de usuario. *CienciaUAT*. **14**(1), 71–84 (2019) <https://doi.org/10.29059/cienciauat.v14i1.1173>
- Prybeha, Dmytro, et al. 2021a. “Analysis of methods of printing images on textile materials and evaluation of their quality”
- Prybeha, Dmytro, et al. 2021b. “Technology Of Making Thermal Transfers”
- Ragab, M., Othman, H., Hassabo, A.: An overview of printing textile techniques. *Egypt. J. Chem.* **0**((0)), 0–0 (2022). <https://doi.org/10.21608/ejchem.2022.131477.5793>
- Rubio, S., Díaz, E., Martín, J., Puente, J.M.: Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Appl. Psychol.* **53**(1), 61–86 (2004)

- Olga Sebastián García, Ma. de los Ángeles Del Hoyo Delgado. s.f. La carga mental de trabajo. Madr.: Inst. Nac. Segur. Hig. Trab. <https://www.insst.es/documents/94886/96076/carga+mental+de+trabajo/2fd91b55-f191-4779-be4f-2c893c2ffe37>
- Stanton, N.A., et al.: Métodos de factores humanos: Una guía práctica para la ingeniería y el diseño. 2da. Edición. Londres (2013). <https://doi.org/10.1201/9781315587394>
- Torres-Medina, Y.: El análisis del error humano en la manufactura: un elemento clave para mejorar la calidad de la producción. Rev. UIS Ing. **19**(4), 53–62 (2020) <https://doi.org/10.18273/revuin.v19n4-2020005>
- Tsang, P.S., Velazquez, V.L.: Diagnosticity and multidimensional subjective workload ratings. Ergonomics. **39**(3), 358–381 (1996). <https://doi.org/10.1080/00140139608964470>
- Wickens, C.D.: Information processing, decision-making, and cognition. Handb. Hum. Factors, 72–107 (1987)



Optimizing eGov Philippines Application for Users in Manila City: A Study on Enhancing Engagement, Accessibility, and User Experience through Usability Testing and Web Scraping

Janina Elyse A. Reyes^(✉), Rapeh Jahaziel L. Galao, Joshua J. Ordinario, Vanessa L. Santillan, and Hannagene C. Vidal

Technological Institute of the Philippines, Quiapo, Manila, Philippines
jereyes.ie@tip.edu.ph

Abstract. The eGovPH app has transformed how citizens interact with the government in the Philippines by introducing a digital platform that is essential in today's era of technology. However, to meet users' changing needs and expectations, the app's usability and efficiency must be optimized. Researchers are currently conducting a comprehensive study to enhance the application's navigation and accessibility features and address any issues that may negatively impact the user experience. The primary goal of the study is to design an efficient and user-friendly interface for the eGovPH app that can compete in the highly competitive digital landscape. To ensure the effectiveness of the study, real-time insights are obtained from the Google Play Store using web scraping techniques. These insights are then used to create an analytics dashboard, which facilitates informed decision-making processes to enhance the user interface and meet the constantly changing user demand. By identifying and analyzing existing problems, the study aims to provide practical solutions that result in a seamless and intuitive user experience. Specifically, the study will conduct usability testing of the proposed improved user interface mockup of the eGovPH app based on the needs of Manila City residents compared to the current user interface. This interdisciplinary approach enables the comprehensive review of the eGovPH app's development and offers a visionary method of adapting to the rapidly evolving digital landscape. By responding to the changing needs of Manila City residents, the study ensures the sustainability of the application in supporting the provision of government services.

Keywords: e-Governance · Usability Testing · User Interface · and Experience

1 The Problem and its Background

1.1 Introduction

Access to top-quality public services should be a fundamental right for everyone globally [5]. Over the past few years, The Department of Information Communication and Technology in the Philippines has made several notable steps towards solving these many

problems. They have been creating an application that welcomes Electronic Governance to improve how citizens interact with government officials. This new method simplifies all government services into one mobile application called the ‘eGovernment Philippines’ app. eGovernment uses the internet, mobile apps, and cloud-based technology to enable businesses and citizens to operate online [2].

1.2 Statement of the Problem

Significantly, the study seeks to explore the user’s experience that causes dissatisfaction in terms of barriers, challenges, current scope and functionality, and impact improvement.

1.3 Objectives of the Study

The objective of the study is to achieve the following goals: (1) Collect user feedback by conducting surveys and interviews to design a prototype, (2) extract eGovPH app ratings and reviews from the Google Play Store, and (3) develop an analytics dashboard to track user satisfaction and experience.

1.4 Scope and Limitations

The study will evaluate the usability of the eGovPH app in Manila through questionnaires, usability testing, and web scraping. The results will be displayed on an analytics dashboard, and the prototype UI will showcase the test results.

2 Review of Related Literature

eGovernment in the Philippines improves public administration transparency and advances public service provision. The eGovernment Master Plan (EGMP) acknowledges that interoperability and harmonization involve technical and organizational concerns [3]. Personal opinions and corporate culture influence the successful implementation of eGovernment in the Philippines. Web scraping can help automate data acquisition for big data applications, and Python’s BeautifulSoup and Selenium libraries are valuable tools for this [1]. A study developed a web scraper to extract and analyze data from websites like Amazon, capturing product attributes such as name, price, reviews, rating, and links. The study also emphasized an improvement of a user-friendly interface with data visualization techniques [4].

3 Research Methodology

The study’s methodology involves web scraping and usability testing to improve the eGovPH app for users in Manila City.

3.1 Method of Research Used

The study uses qualitative and quantitative methods to evaluate the eGovPH app’s user experience and identify trends in Manila City.

3.2 Respondents of the Study

The study involves 417 Manila City residents aged 18 to 54, comprising 220 males (52.8%) and 197 females (47.2%).

3.3 Sampling Technique

Researchers used Slovin’s formula and a 5% margin of error to determine a sample size of 400 from a population of 14,406,059 from Manila City’s population.

$$n = \frac{N}{1 + N e^2}$$

where:
n = number of samples.
N = total population.
e = margin of error.

3.4 Data Gathering Procedure

The application’s user feedback and preferences were gathered through web scraping, surveys, and usability testing (Fig. 1).

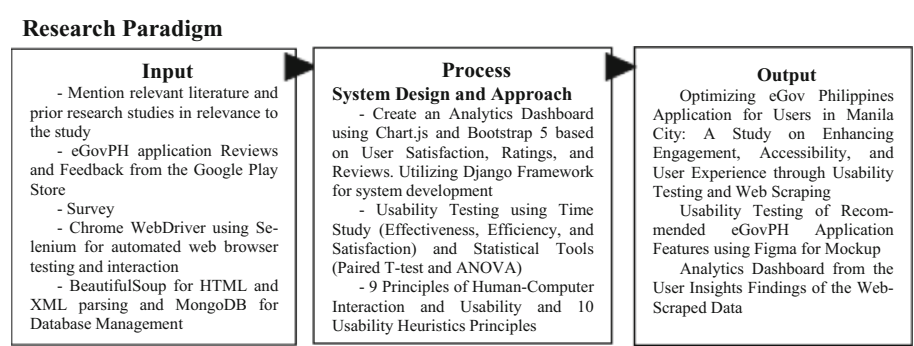


Fig. 1. Conceptual Framework

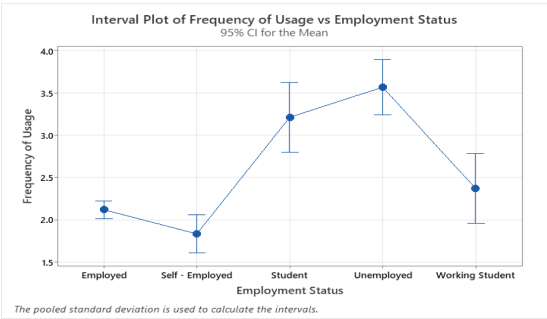


Fig. 2. Interval Plot of Frequency of Usage vs Employment Status

4 Presentation, Analysis, and Interpretation of Data

Frequency of eGovPH app Usage

Occupations affect eGovPH app usage frequency significantly. Employed individuals use the app more often than unemployed, self-employed, or student respondents (Fig. 2).

Usability for the Current and Improved UI eGovPH app

The effectiveness of the Request Government Documents more greatly enhanced with the enhanced User Interface of the eGovPH application. In the current version, there are 32 tasks accomplished and the remaining 8 tasks are unfulfilled. On the other hand, the enhanced paradigm had 37 successful tasks while only 3 out of the total tasks were proven to be incomplete or a failure. All these can be attributed to major advances in design and operation thus resulting in improved human relations and goal accomplishment.

Current and Improved User Interface Effectiveness Goals

This compares the effectiveness of the current and improved user interface of the eGovPH app across four tasks: a document, repeating a document, a new document, and a one-stop-shop service in health services. All tasks were successfully solved in the current version except for 40 requests to address government documents and official records. The enhanced version of this request increased its chances to 37 out of 40 documents when done at the government level. The observed usability problems were also fewer in the new interface, thus suggesting that the system is more user-friendly.

Current and Improved User Interface

This explains shows that the eGovPH app’s UI improvements resulted in better performance. The current version had a time-based user efficiency of 0.0627 goals/sec, 99.19% overall relative efficiency, 95% effectiveness, and 78.81% satisfaction. The improved version achieved 0.0473 goals/sec, 99.64% overall relative efficiency, 98.13% effectiveness, and 85.56% satisfaction. These enhancements indicate a more efficient and satisfying user experience.

Paired T-test Results

Table 1 illustrates that the new eGovPH app met expectations on all tasks, showcasing

technological advancements. Recent updates to the app have improved usability and functionality, validating the recent amendments.

Table 1. Paired T-test result of the Current and Improved eGovPH app

	Local Govt. Documents Task	Renew Govt. Docs Task	Request Govt. Docs Task	One-Stop Health Services
P-Value	0.000	0.225	0.000	0.000
T-Value	45.50	1.23	24.68	-7.05

5 Findings, Recommendations, and Conclusions

5.1 Summary of Findings

The report analyzed 417 users' opinions and preferences and recommended practical app interface improvements. The eGovPH app scored 78.81 in user satisfaction, while the improved version achieved 85.56 on the System Usability Scale (SUS), surpassing the industry standard SUS score of 68. The Analytics Dashboard provides monthly updates on user behaviour and includes charts for trend analysis and ratings.

5.2 Conclusion

The survey, web scraping, and usability tests revealed visibility and navigation issues. Necessary improvements include enhancing visibility, simplifying document requests, removing external links, adding a search function, and enhancing accessibility.

5.3 Mockup eGovPH App

Based on the survey results, the proposed eGovPH app's User Interface has improved features such as One-Stop Healthcare, Electronic Freedom of Information, and access to government documents. The Analytics Dashboard displays monthly data through various charts. For instance, the Line Chart shows that in June 2023, eGovPH received 1059 reviews, while in December 2023, it received 173 reviews. The Sentiment Analysis Pyramid Chart identifies areas for improvement based on user feedback. The Doughnut Chart shows that the user sentiment towards the app has been primarily positive (61.4%), with 24% neutral and 14.6% negative. The Stacked Bar Chart helps analyze the app, identify issues, and evaluate effectiveness. The Radar Chart shows that the app received 3027 ratings from June to December 2023. Of this, 703 reviews were 1-star, and 1678 reviews were 5-star (Fig. 3).

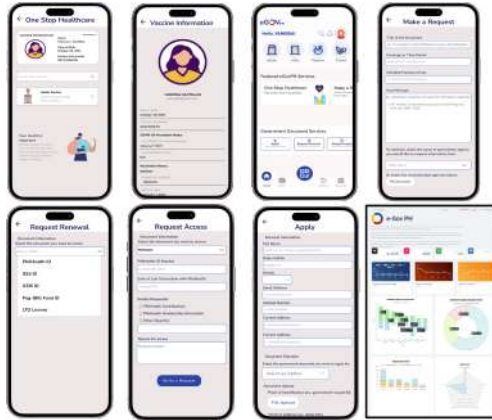


Fig. 3. Proposed eGovPH app Mockup and User Satisfaction Analytics Dashboard

5.4 Recommendation

Researchers have recommended that the eGovPH app should integrate all essential government services to ensure an efficient and improved user experience for the citizens of Manila City to ensure a streamlined and efficient user experience. Additionally, researchers recommend changing the web scraping frequency to bi-weekly intervals, which would enable the app to obtain up-to-date information on user engagement, satisfaction, and trends. To ensure that each app update, including new features and design, meets the highest quality standards, the researchers recommend adding a testing stage. This would allow for thorough app testing before it is released to the public, ensuring that any issues are addressed before they affect users. The researchers also suggest adding a slicer feature to the data visualization software, enabling a detailed monthly review of user satisfaction metrics. This feature would allow users to explore patterns over various months and gain insights into user behavior, ultimately leading to a better understanding of how to improve the app.

References

1. Abodayeh, R., Hejazi, W., Najjar, L., Shihadeh, L., Latif, R.: Web scraping for data analytics: A beautiful soup implementation. In: 2023 Sixth International Conference Women Data Science Prince Sultan University (WiDS PSU), pp. 65–69. IEEE, Riyadh, Saudi Arabia (2023). <https://doi.org/10.1109/WiDS-PSU57071.2023.00025>
2. Basu, S.: E-government and developing countries: an overview. *Int. Rev. Law Comput. Technol.* **18**(1), 109–132 (2004)
3. Borsigue, A. V. (2022). The effective eGovernment adoption in the Philippine local government: A supply-side perspective. Retrieved from: https://www.researchgate.net/publication/364347085_The_Effective_eGovernment_Adoption_in_the_Philippine_Local_Government_A_Supplyside_Perspective_of_Masbate_Province

4. Chang, C., Almaghalsah, H.: Usability evaluation of eGovernment websites: A case study from Taiwan. *Int. J. Data Netw. Sci.* **4**, 127–138 (2020). <https://doi.org/10.5267/j.ijdns.2020.2.004>
5. Sønderskov, M., Rønning, R.: Public service logic: An appropriate recipe for improving serviceness in the public sector? *Adm. Sci.* **11**(3), 64 (2021). <https://doi.org/10.3390/admsci11030064>



Generating Design Guidelines for Alternative Medication Management Tools from Medical Cannabis Patients

Deepthi Rajagopal¹(✉), Douglas Bruce¹, Elissa Foster¹, Olayele Adelakun¹,
and Enid Montague²

¹ DePaul University, Chicago, IL 60604, USA
drajagop@depaul.edu

² University of Toronto, Toronto, ON, Canada

Abstract. Increasingly emerging use of Cannabis for pain management warrants the need for recommendations and guidelines in place for patients to navigate their use effectively. We conducted a secondary data analysis of 11 interviews with patients who consumed marijuana for chronic pain. The purpose of this study was to understand how patients with chronic pain manage their symptoms with cannabis. By applying the user centered design process, we identified pain points and gaps in how different patients procured, used and managed their cannabis use. We identified three distinct use cases and modeled these use cases using experience maps and developed personas describing the stories of medical cannabis users. The journey of patients in learning, obtaining and using medical cannabis were mapped into tasks and subtasks with the help of charts to recognize gaps in overall experience. The reported challenges include: (1) finding information about cannabis for pain vs other medications, (2) finding trusted information, (3) doctors being hesitant to prescribe medical cannabis, (4) cost factors (5) need for monitoring by doctors to advise patients about usage. We developed design guidelines for applications or tools to support patients' use of alternative medicines grounded in patient experiences. The design guidelines address the challenges of the study by helping patients log in the symptoms, side effects and strain information making it easy to track their use of cannabis over time using technology.

Keywords: Medical cannabis · Design research · User centered design

1 Introduction

Medical cannabis patients consistently report using cannabis as a substitute for prescription medications; however, little is known about individuals accessing cannabis through adult-use markets (areas where cannabis is legal to purchase for the general adult population) [1]. In 2017, the National Academies of Science, Engineering and Medicine (NASEM) released a report supporting the effectiveness of using cannabis to treat chronic non-cancer pain among adults [4]. Due to the increasing availability of medical-grade cannabis as a treatment for chronic conditions, it is important to examine

how prospective users navigate the complex and often-difficult process of researching, procuring and successfully using medical cannabis [2]. As of April 2024, Medical cannabis use is allowed in 38 states, three territories and District of Columbia in the US [6]. Over the last years there is a tremendous burst regarding health information supported by several medical oriented webpages, applications and social media [7]. Using an app to collect daily data could offer more detailed, prospective, and dynamic information of cannabis use than traditional paper-based measures [8]. Those with chronic conditions considering medical cannabis as an option may not discuss it with their care providers because of perceived social stigma or fear of legal repercussions [9]. Physicians currently lack dominant narrative environments that would enable them to coherently present how cannabis can be integrated into their clinical practice [10]. With the increase in use of medical cannabis, people are often left to make less informed decisions about the usage.

This paper discusses the process involved in identifying the design opportunities related to medical cannabis use using a user centered approach and guidelines generated for the design of medical cannabis applications. In this study, we focused on analyzing the raw data collected by interviewing medical cannabis users. We processed the raw data of their cannabis use and mapped the user characteristics including demographics, symptoms, usage experience, interaction with healthcare providers, and social concerns. The purpose of this study was to discover the gaps and understand how technological interventions can help users obtain useful information and monitor the use of cannabis.

2 Method

The raw data used for this study were collected from the larger project examining patients' use of medical cannabis for the treatment of their chronic conditions, primary care providers' (PCPs) perspective on medical cannabis, the communications strategies used by patients and providers to discuss medical cannabis use and patient experiences of medical cannabis and more traditional pharmacological treatments for their chronic conditions [3].

2.1 Recruitment

Medical Cannabis patients in Illinois were recruited through flyers distributed to selected medical cannabis dispensaries in the state [2]. Most participants (63.3%) reported that they learned of the study through a dispensary location and the remainder (36.6%) reported that they learned of the study through Facebook [2]. People were eligible if they were 18 years and older, reported having one of the 39 qualifying health conditions for medical cannabis in Illinois, and had consumed medical cannabis in the past 3 months [2]. The final sample included 30 eligible participants.

2.2 Data Collection

Data collected through telephonic interviews were conducted with open ended questions. The protocol focused on four topics: (a) symptom management and pain reduction (e.g., which symptoms do you find that cannabis is helpful in managing?), (b) communication

about chronic conditions and cannabis with PCPs (e.g., Did the stigma of using cannabis affect when you talked to your doctor about it for the first time?), (c) participant perspectives on medicinal uses of cannabis and social stigma (e.g., Can you describe what you understand about the use of cannabis as a medicine?), and (d) views on recreational cannabis use (e.g. What do you think about the use of cannabis for pleasure?) [2].

2.3 Data Analysis

We used a collection of 11 interview transcriptions from the 30 interviews who described similar symptoms and concerns. Focusing on all 30 patients would be too broad for the purposes of this study. The interview transcripts ranged from 16–27 pages. Some key findings from the raw data by Bruce, Foster et al. [3] were

- Participants described a range of approaches to using medical cannabis, including (1) as alternatives to using prescription or over-the-counter medications; (2) complementary use with prescription medications; and (3) as a means for tapering off prescription medications.
- Motives for using MC as an alternative to prescribed opioids included MC (1) acting more quickly; (2) having longer lasting effects; (3) reducing potential harm (from opioids); (4) better managing symptoms; (5) having fewer side-effects.
- Several described experimenting with different MC products and strains and finding certain ones to be much more long lasting than other forms of prescribed pain medication.

User characteristic maps. The findings were processed to map user characteristics using a formulated data capture framework. The data capture framework included: (a) Medical Cannabis use to address health condition (b) details about overall use (c) experience related to medical cannabis procurement and use, (d) user details. Our team of 3 coders used this framework to capture data across 11 interview transcriptions.

User types. Using our user characteristics maps of 11 users, we aggregated captured data in a summary to compare experiences and classify them as different types of use cases. We identified 3 distinct use cases. They differed distinctly on prior exposure, extent of medical cannabis use for conditions or symptoms, and level of monitoring cannabis use on a regular basis.

- Type 1 | Unmoderated and limited use, not monitored by healthcare
- Type 2 | Moderated use, some self-monitoring, not monitored by healthcare
- Type 3 | Moderated and extensive use, monitored by healthcare

Experience maps. After identifying 3 distinct use cases, we used the data we captured to think about key elements of the experience map, as described below:

- *Motivations for use:* Users' medical conditions, related symptoms, medication and side effects, quality of life affected, and effects of cannabis use on conditions.
- *Current cannabis use:* Frequency of use, dosage and routine, modes of delivery.
- *Patient experiences:* Patient's journey of how they decided to use cannabis, how they got certification for use, and how they use cannabis on a regular basis.

Based on these key elements, we created a map template to document the experience of 3 distinct user types. We recorded three unique user type experiences using the template below:

Type 1 user: Unmoderated & limited use, not monitored by healthcare.

- Prior exposure: A patient in this user type was open to using cannabis based on some prior knowledge obtained from being a substance abuse counselor, but did not get opportunities to be guided on use for medical conditions.
- Motivations: Their main aim was to address the negative side effects of using excessive pain medication. They used cannabis to address some symptoms that MC has known usage for, such as nausea and pain management, but later also used it for improved quality of life and stress management
- Extent of use: They only use “smoke” as a delivery mode, and choose to smoke cannabis that was available based on onset of symptoms. Based on their cannabis use, they could reduce intake of over the counter, and prescription medication significantly.
- Monitoring use: There is a missed opportunity to consider a wide variety of strains for individual symptoms, and tracking dosage and use of strains for symptom alleviation.

Personas. Personas were created based on the comprehensive customer journey map generated in the previous step for our three user types. The touchpoints of each user were mapped to understand the goals and motivations of cannabis use. This was brought together with the summary sheet and journey map, identifying the potential roadblocks and pain points in the experience of the users in consumption of cannabis. Then, based on the identified themes and user types in the journey map, we created goal-directed personas for the three types of users. The personas highlight how each of the patient types uses cannabis, their goals, different problems faced by them in terms of consumption or getting prescribed.

Task mapping. Using different parts of the experience maps, we documented tasks that users are performing across different parts of their journeys. Using task-mapping methodology, we broke down each task into sub tasks. We identified multiple ways in which a system could act as an information agent for different types of gaps:

- Gaps arising from the current legal ecosystem and stigma in healthcare: Provide information support with legalization information, ways to procure certification, where to procure certification from.
- Gaps arising from lack of reliable information on health effects: Point users to reliable information about health benefits from other users with similar conditions, provide information on strains and effects

Scenario, feature list and claim analysis. Two scenarios were formulated from user interviews describing the tasks that users would perform using the application. These scenarios informed the feature list and claims analysis was performed to elaborate the pros and cons of the features. The feature list included:

1. Feature to input symptoms
2. Feature to log the prescribed medication
3. Recommendation on the alternative medication/medical cannabis
4. Add/log in daily experience and emotions
5. Send report to physicians

3 Discussion and Future Work

Although our demographics showed a large number of males using cannabis, there was no evidence found if gender plays a role in this. The task maps were used to generate scenarios of medical cannabis use for three types of user groups and their pain points were visualized using story maps. The scenarios and story maps helped in creating a feature recommendation followed by a competitor analysis of other applications in the market. The process helped in building the foundation of an application to help monitor the symptom relief offered by medical cannabis for the management of chronic medical conditions and share information regularly with care providers. The application was development at the ID Lab at DePaul University, Chicago.

4 Conclusion

Use of Medical cannabis seems to help the participants manage their symptoms of medical conditions and negative side effects of prescribed medication. By interviewing the users of medical cannabis, the study gathered patient centered research to find the gaps of the patient experience and uncover difficulties faced by medical cannabis users to manage their chronic pain symptoms effectively. The research data was translated into set of features to develop an application for medical cannabis use management which can be scaled towards the development of alternative medication management and use.

References

1. Bachhuber, M., Arnsten, J.H., Wurm, G.: Use of cannabis to relieve pain and promote sleep by customers at an adult use dispensary. *J. Psychoact. Drugs.* **51**(5), 400–404 (2019). <https://doi.org/10.1080/02791072.2019.1626953>
2. Brady, J.P., Bruce, D., Foster, E., Shattell, M.: Self-efficacy in researching and obtaining medical cannabis by patients with chronic conditions. *Health EDuc. Behav.: Off. Publ. Soc. Public Health Educ.* **47**(5), 740–748 (2020). <https://doi.org/10.1177/1090198120914249>
3. Bruce, D., Brady, J.P., Foster, E., Shattell, M.: Preferences for medical marijuana over prescription medications among persons living with chronic conditions: alternative, complementary, and tapering uses. *J. Altern. Complement. Med. (N. Y. N.Y.)*. **24**(2), 146–153 (2018). <https://doi.org/10.1089/acm.2017.0184>
4. NASEM webpage, <https://www.nationalacademies.org/news/2017/01/health-effects-of-marijuana-and-cannabis-derived-products-presented-in-new-report> last Accessed 19 May 2024.
5. NCBI webpage, <https://www.ncbi.nlm.nih.gov/books/NBK458660/> Accessed 19 May 2024.
6. NCSL webpage, <https://www.ncsl.org/health/state-medical-cannabis-laws> Accessed 19 May 2024.
7. M. Spanakis, S. Sfakianakis, E. G. Spanakis, G. Kallergis and V. Sakkalis, PDCA: An eHealth service for the management of drug interactions with complementary and alternative medicines. 2018 IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI). Las Vegas, NV, USA, 2018;pp. 373–376. doi:<https://doi.org/10.1109/BHI.2018.8333446>. keywords: {Drugs;Databases;Electronic healthcare;Tools;Diseases}.
8. Santesteban-Echarri, O., Haffey, P., Kim, G., Tang, J., Addington, J.: A mobile-based app to monitor cannabis use among youth at clinical high risk (CHR) for psychosis: Feasibility and acceptability of LooseLeaf. *Schizophr. Res.* **222**, 505–506 (2020)

9. Satterlund, T.D., Lee, J.P., Moore, R.S.: Stigma among California's medical marijuana patients. *J. Psychoactive Drugs*. **47**(1), 10–17 (2015). <https://doi.org/10.1080/02791072.2014.991858>
10. Zolotov, Y., Vulfsons, S., Zarhin, D., Sznitman, S.: Medical cannabis: An oxymoron? physicians' perceptions of medical cannabis. *Int. J. Drug Policy*. **57**, 4–10 (2018). <https://doi.org/10.1016/j.drugpo.2018.03.025>



Research on the Emotional Vocabulary Used in the User Interface of Generative AI-Based Music Production System

Yeongju Seo¹  and Kiechan Namkung²  

¹ Department of Smart Experience Design (TED), Kookmin University, Seoul 02707, South Korea

² Department of AI Design, College of Design, Kookmin University, Seoul 02707, South Korea
soundux@kookmin.ac.kr

Abstract. Generative AI-based music production systems are useful tools for generating text, images, and sounds, with particularly active use in the fields of text and image generation. However, most of these AI systems' user interfaces (UI) does not yet sufficiently reflect user experience (UX) to satisfy users. This study aims to analyze the emotional vocabulary that is currently used in generative AI systems and to derive and compare the emotional vocabulary used by users to express the atmosphere of the music through experiments. This serves as a foundational study that demonstrates the necessity for more user centered development of emotional vocabulary.

Keywords: User Interface · Emotional Vocabulary · AI-Based music production System

1 Introduction

AI systems are useful tools for generating text, images, and sounds, and they are actively used, particularly in text and image generation [1].

However, most of these AI systems' user interfaces (UI) does not yet sufficiently reflect user experience (UX) to satisfy the users [2]. User experience encompasses the overall experience of user interactions within a product or a service [3], and the user interface plays a crucial role in utilizing the system as it is a key element in the creative process provided by generative AI systems [4].

The user interface of AI systems consists of various elements, including audio, voice, and text input [5]. Among these, selecting or entering text is essential for generative AI systems to generate appropriate content [6]. Therefore, the users' choice of vocabulary used in the user interface is important.

Generative AI-based music production systems use artificial intelligence algorithms to generate music [7]. The user interface of these systems is designed to allow users to generate music based on their preferences by manipulating parameters such as emotion, tempo, and length through text. Similar to other generative AI systems that rely on text,

this system enables users to create music by adjusting these variables to suit the users’ tastes [7].

Various research on the interaction between generative AI-based music production systems and users indicate that most current user interfaces are very difficult to use in actual music generation scenarios. Moreover, there are few preceding studies on user interfaces and emotions [2]. Since emotion plays a decisive role in music composition [8], the emotional vocabulary provided in the user interface becomes a critical element in generative AI-based music production systems.

This study aims to analyze the emotional vocabulary currently used in generative AI-based music production systems and to derive and compare the emotional vocabulary used by users to express the atmosphere of the music through experiments. This is a foundational study for designing user-centered emotional vocabulary in generative AI-based music production systems.

2 Literature Review

2.1 Current Status of Generative AI-Based Music Production Systems

Current AI systems that generate music by selecting given emotional vocabulary are Soundraw [9], Mubert [10], Evoke Music [11], Aiva [12], Ecrett Music [13], Beatoven [14], and MusicStar [15]. Most of these systems are rule-based systems, which use predefined emotional vocabulary as parameter tools to encode and express specific emotions, thus allowing the creation of musical works [16]. The emotional vocabulary used in each system is shown in Table 1.

Table 1. Emotion Vocabulary Used in Music Production AI System Interface

Music Production AI systems interface	Emotional Vocabulary
Soundraw	Epic, Happy,Hopeful, Angry, Sentimental, and 20 others
Mubert	Fun, Beautiful, Inspirational, Peaceful, and 50 others
Evoke music	Aggressive, Busy, Danger, Frantic, Cold, and 55 others
Aiva	Calm, Excited, Fearful, Tension, Sad
Ecrett music	Chill, Dark, Uplifting, Happy, Serious, and 6 others
Beatoven	Sad, Calm, Motivational, Happy, Scary, and 11 others
MusicStar	Happy, Sad, Blessed, Loved, Adventurous, and 4 others

2.2 Research on the Relationship Between Music and Emotional Vocabulary

Previous studies dealing with emotional vocabulary in music have mostly focused on improving the technical performance of AI music tools. In particular, studies have been conducted on categorical models and dimensional models, which form the theoretical basis for designing and implementing AI-based emotional music production systems [17, 18]. Research on label-based generative models, which use emotion labels to train music, has also been conducted [19, 20].

However, the emotion models and emotion labels covered in these preceding studies have not beneficially impacted user-centric interfaces [19]. Users often experience problems where music with different emotions are generated from their selected emotional vocabulary, significantly lowering the user experience in the creative process.

3 Methods

The experiment was conducted with 15 content designers who has experience with using generative AI-based music production systems. These core users of generative AI-based music production systems are mainly identified as music composers and content designers. The experiment was conducted in three main stages.

First, a focus group interview was conducted with the participants. Participants listened to 34 randomly pre-selected pieces of music, discussed the emotions they felt when listening to the music, and recorded the emotional vocabulary. Then, the derived emotional vocabulary was grouped through card sorting. Card sorting was conducted in two groups using an open card sorting method. Finally, the emotional labels obtained from the card sorting were compared with the emotional vocabulary used in existing AI generative systems (see Fig. 1).

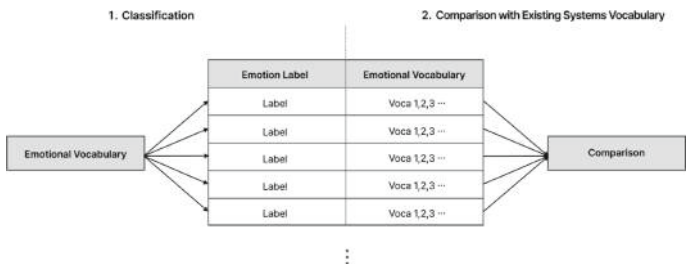


Fig. 1. Emotional Vocabulary classification and detailed comparison procedure

4 Results

The focus group interview resulted in 192 emotional vocabularies for the 34 pieces of music. This indicates that listeners can describe their emotions with various emotional vocabularies for the same music. Through the card sorting of the two groups, 14 labels were derived from Group A and 11 labels from Group B, totaling 25 emotional labels. Among these, 4 common labels—Dull, Powerful, Dreamlike, Exciting—were derived,

including most of the common emotional vocabulary. However, the remaining labels showed differences, indicating that emotional expressions are subjective and can be perceived differently by individuals (Table 2).

Table 2. Emotion Label

	Group A	Group B
Emotion Label	Powerful, Exciting, Dull, Dreamlike, Tranquil, Refreshing, Youthlike, Sweet, Rough, Attractive, Peppy, Overwhelming, Delicate, Sacred	Powerful, Exciting, Dull, Dreamlike, Warm, Dark, Fluttering, Emotional, Exhilarating, Distinctive, Tortuous

The comparison results with the emotional vocabulary used in existing generative AI systems are shown in Table 3.

Table 3. Comparison of Emotional Vocabulary Used in Existing Generative AI Systems with User-Derived Emotional Labels

Music Production AI systems interface	Emotion Vocabulary Count	Representative Vocabulary Match Count(Group A)	Representative Vocabulary Match Count(Group B)	Total Matches with AI System Vocabulary (192 terms)
Soundraw	25	0	1	1
Mubert	44	0	0	10
Evoked music	57	1	2	20
Aiva	5	0	0	1
Ecret music	11	0	1	6
Beatoven	16	0	0	6
MusicStar	9	0	0	1

The analysis of the matching items between these sets showed a low level of agreement and poor matching between the entire derived emotional vocabulary and the representative emotional vocabulary.

5 Conclusions and Limitations

This study confirmed through experiments that the emotional vocabulary used in generative AI-based music production systems are not user-centered. The results are as follows: first, it was found that the emotional vocabulary used in most generative AI-based music production systems differs significantly from the emotional vocabulary used by actual users to express the emotions of music. This discrepancy is a major cause of the system’s inability to accurately capture and reflect users’ intentions and emotions, resulting

in poor user experience. Second, the potential to design user-centered emotional vocabulary was discovered. The experiment conducted with the same emotional vocabulary yielded many common emotional labels and commonly used emotional vocabulary with similar meanings.

This study has several limitations. The sample size of the user group participating in the study is small and not extensive, limiting the generalizability of the research outcomes. Specifically, the individual musical diversity and personal experiences of the participants were not adequately reflected. Additionally, this study focused on a few specific generative AI-based music production systems and did not address all available options, which is somewhat limiting. This suggests that the study results may primarily apply to the tools examined, and different outcomes may be observed with other generative AI-based systems. Furthermore, the detailed relationships between emotional domains arising from musical elements such as harmony, melody, timbre, and chord progression and the emotional vocabulary were not sufficiently considered. This indicates that there was a limitation in precisely analyzing the emotional impact of musical elements on users.

This foundational study demonstrates the need for more user-centered emotional vocabulary development of in the interfaces of generative AI-based music production systems. When interpreting the results of this study, it is necessary to consider several limitations. Future research should focus on specific measures to design user-centered emotional vocabulary and practical methods for their application.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4], as well as a URL [5].

References

1. Gozalo-Brizuela, R., Garrido-Merchán, E. C.: A survey of generative AI applications. arXiv preprint arXiv:2306.02781v2 (2023)
2. Civit, M., Civit-Masot, J., Cuadrado, F., Escalona, M.J.: A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Syst. Appl.* **209**, 118190 (2022). <https://doi.org/10.1016/j.eswa.2022.118190>
3. Yang, B., Wei, L., Pu, Z.: Measuring and improving user experience through artificial intelligence-aided design. *Front. Psychol.* **11**, 595374 (2020). <https://doi.org/10.3389/fpsyg.2020.595374>
4. Torricelli, M., Martino, M., Baronchelli, A., Aiello, L.M.: The role of interface design on prompt-mediated creativity in generative AI. In: *ACM Web Science Conference (WebSci '24)*, May 21–24, Stuttgart, Germany, pp. 1–6. ACM, New York, NY, USA (2024). <https://doi.org/10.1145/3614419.3644000>
5. *Interaction Design Foundation - IxDF*: What is user interface (UI) design? *Interaction Design Foundation - IxDF*. <https://www.interaction-design.org/literature/topics/ui-design> (2024). Accessed 14 Nov 2024
6. Lee, S., Yan, J.: The potential of a text-based interface as a design medium: An experiment in a computer animation environment. *Interact. Comput.* **28**(1), 85–101 (2016). <https://doi.org/10.1093/iwc/iwv045>

7. Zhu, Y., Baca, J., Rekabdar, B., Rawassizadeh, R.: A survey of AI music generation tools and models. arXiv preprint arXiv:2308.12982v1 (2023)
8. Juslin, P.N., Sloboda, J.A.: *Music and Emotion: Theory and Research*, 1st edn. Oxford University Press, Oxford (2001)
9. Soundraw., <https://soundraw.io> (2024). Accessed 14 Nov 2024
10. Mubert., <https://mubert.com> (2024). Accessed 14 Nov 2024
11. Evoke Music., <https://evokemusic.ai> (2024). Accessed 14 Nov 2024
12. Aiva., <https://www.aiva.ai> (2024). Accessed 14 Nov 2024
13. Ecrett Music., <https://ecrettmusic.com> (2024). Accessed 14 Nov 2024
14. Beatoven., <https://www.beatoven.ai> (2024). Accessed 14 Nov 2024
15. MusicStar., <https://www.musicstar.ai> (2024). Accessed 14 Nov 2024
16. Dash, A., Agres, K.R.: AI-based affective music generation systems: A review of methods, and challenges. In: *Proceedings of the XX International Conference*, pp. 1–26. ACM, New York, NY, USA (2022)
17. Williams, D., Kirke, A., Miranda, E.R., Roesch, E., Daly, I., Nasuto, S.: Investigating affect in algorithmic composition systems. *Psychol. Music.* **43**(6), 831–854 (2015). <https://doi.org/10.1177/0305735614543282>
18. Ferreira, L.N., Whitehead, J.: Learning to generate music with sentiment. In: *20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, pp. 384–390. ISMIR, Delft, The Netherlands (2019)
19. Zheng, K., et al.: EmotionBox: A music-element-driven emotional music generation system based on music psychology. *Front. Psychol.* **13**, 841926 (2022). <https://doi.org/10.3389/fpsyg.2022.841926>
20. Zhao, K., Li, S., Cai, J., Wang, H., Wang, J.: An emotional symbolic music generation system based on LSTM networks. In: *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2019)*, pp. 2039–2043. IEEE, New York, NY, USA (2019). <https://doi.org/10.1109/ITNEC.2019.8729352>



Exploring Usability and User Value of the Electric Powered Wheelchair for the Elderly and Physically Disabled

Sujin Seong , Yubin Ahn, Pakeeza Abid, and Jaehyun Park  

Department of Industrial and Management Engineering, Incheon National University, Incheon, South Korea

jaehpark@inu.ac.kr

Abstract. Individuals experiencing various disabilities and aging who rely on Electric Powered Wheelchairs (EPW) for mobility assistance highlight the necessity of User-Centered Design in developing EPWs due to their diverse characteristics. This study evaluated the usability of a developed EPW, considering factors such as aging, prior EPW experience, and usage duration. Usability evaluations were conducted, including the System Usability Scale (SUS) and semi-structured interviews. Results indicate that elderly users rated the EPW more favorably than disabled users, with prior experience and usage duration influencing evaluations, suggesting a critical evaluation based on familiarity and robustness. Disabled users emphasized operational issues, especially regarding footrest height adjustment for long-term severe disability users, while elderly users appreciated its ease of operation. Longer EPW usage enhances user understanding, with previous device experience potentially influencing evaluations. These findings underscore the significance of considering user experience factors in EPW design for diverse user groups.

Keywords: Electric Powered Wheelchairs (EPW) · Usability evaluation · User value · User Experience (UX) · User-Centered Design (UCD)

1 Introduction

Recent advancements in Electric Powered Wheelchairs (EPW) have revolutionized mobility solutions, enhancing independence, comfort, and accessibility for users with disabilities and the elderly [1]. With the increasing demand for EPWs, there is a growing emphasis on evaluating their usability and effectiveness in meeting users' diverse needs [2]. Especially for the elderly and disabled, there were many studies have investigated numerous technological advancements over time to enhance the EPWs' User Experience (UX) [3], usability [4], accessibility [5], and functionality [6].

Despite the advancements in EPW technology, a significant gap persists in research on usability evaluation concerning users' prior experience, usage duration, and age. Addressing the diverse needs of user groups is crucial for enhancing user experience and ensuring EPWs meet their expectations and requirements.

This study seeks to address this gap by examining the usability using several questionnaires and semi-structured interviews of a newly designed EPW among two distinct user groups: long-term users and novices. By comparing these groups, we aim to identify potential differences in their experiences, preferences, and priorities concerning EPW usage. Additionally, our investigation explores whether variations exist in the primary values and priorities attributed to the designed EPW among elderly users and those with disabilities.

Through this comprehensive examination, we aim to provide insights into tailoring EPW design and features better to accommodate different user groups’ diverse needs and preferences, ultimately leading to improved user satisfaction and quality of life.

2 Methods

2.1 Participants

This study recruited experimental participants based on the following conditions: (1) The elderly whose walking ability has decreased due to aging. They were aged 65 or over. (2) Disabled people with physical or cognitive disabilities. They have been diagnosed with a physical or cognitive disability. (1) The elderly group consists of a total of 20 participants (10 males and 10 females) aged 65–94 years (total mean = 71.50 years old; SD = 6.35). (2) The disabled group consists of a total of 14 participants (7 males, 7 females), ranging in age from 32–74 years (total mean = 57.86 years old; SD = 11.36).

In addition, participants were asked about their previous experience with EPW. The elderly group had no prior experience with EPWs, whereas the disabled group primarily used EPWs, thus having prior experience and ownership of mobility devices. The usage duration is detailed in the table below (Table 1), along with the overall demographics of both groups. In addition, the specific disability grade follows the standards set by South Korea.

Table 1. Demographics results for two groups. (*N stands for the number of people)

		Total	Male	Female
Elderly	Age (years)	71.50 ± 6.35	72.38 ± 9.16	70.92 ± 3.23
	Mobility experience	No		
Disabled	Age (years)	57.86 ± 11.36	56.57 ± 14.42	59.14 ± 6.83
	Mobility experience	Yes		
	Usage duration of EPW (N)	5 years or less (5), 10 years or less (4), Over 10 years (4)		
	Disability grade (N)	Grade 1 (2), Grade 2 (4), Grade 3 (2), Grade 4 (2), Grade 5(1), and others (3)		

2.2 Electric Powered Wheelchairs (EPW)

The EPW in this study was the E-Trot, developed by SJ-tech. The joystick located on the right armrest can be used to move forward and backward and left and right to control the direction of rotation [7]. The following figure illustrates the operation of the EPW used in this study (Fig. 1).

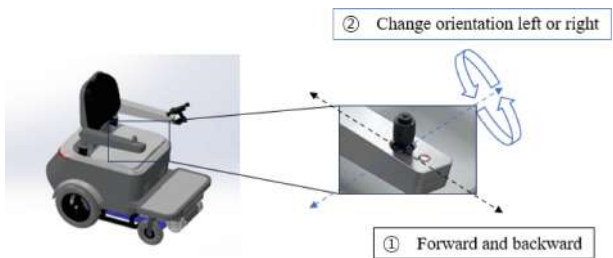


Fig. 1. E-Trot developed by SJ-tech.

Participants in the experiment performed three tasks while wearing electric-powered wheelchairs. The following figure shows Tasks 1, 2, and 3 (Fig. 2). The design ensured that the rear of the parking space was obstructed, simulating real-world conditions as closely as possible.

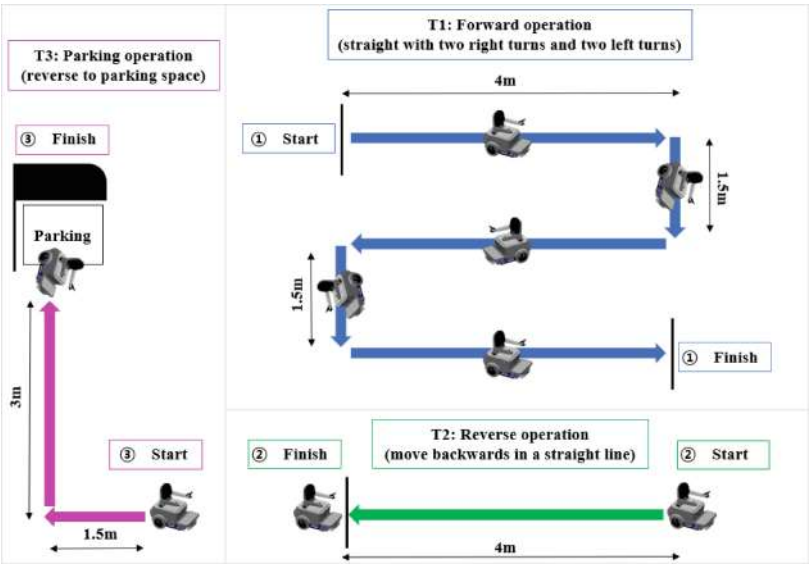


Fig. 2. Description and process of three driving tasks. (*T stands for Task).

2.3 Measurements and Analysis

All usability evaluations were conducted after participants completed the EPW driving task. The System Usability Scale (SUS) was used to assess the usability, using a 5-point Likert scale [8]. After completing all driving tasks and questionnaires, semi-structured interviews were conducted to get feedback on the designed EPW and suggested improvements. ANOVA and additional network analysis were conducted to examine differences based on usage experience.

3 Results

3.1 Comparison of SUS Scores Based on Prior EPW Experience

The SUS scores reflecting all 10 items were compared between the elderly group and the disabled group. As a result, the elderly group exhibited a higher average SUS score than the disabled group ($*p < 0.05$) (Fig. 3).



Fig. 3. Comparison of SUS scores between elderly and disabled individuals.

Subsequently, scores for each of the 10 items on the SUS evaluation were compared. The analysis revealed statistically higher scores for the ease of use and integration items in the elderly group compared to the disabled group ($*p < 0.05$) (Figs. 4 and 5).

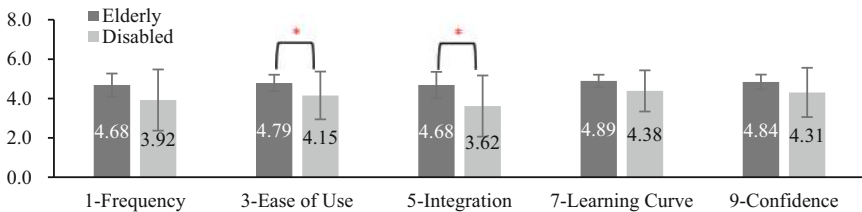


Fig. 4. Each item has an odd number of items that indicate positive questions. $(*p < 0.05)$.

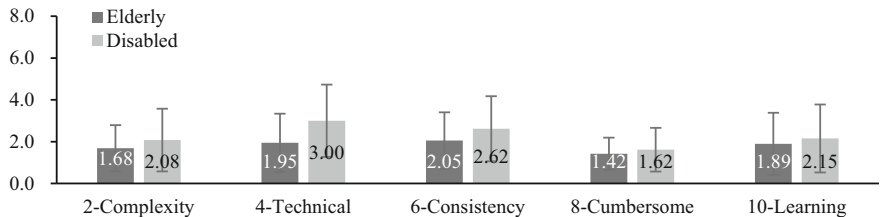


Fig. 5. An even number of items indicate negative questions.

3.2 Network Analysis Based on Prior EPW Experience

To conduct further analysis, network analysis was performed based on prior experience with EPWs. Each specific item of the SUS was designated as individual nodes, with the average reported scores represented as the size of each node. The connectivity between each item was also illustrated with lines to understand their relationships (Fig. 6).

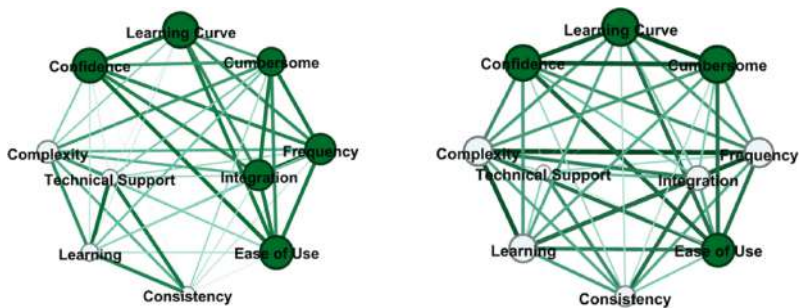


Fig. 6. Comparison of each item through network analysis. There is no EPW prior experience (left) or EPW prior experience (right).

As a result, the group with EPW experience rated Integration and Frequency lower. Furthermore, lower connectivity between nodes was observed, particularly for Frequency, Complexity, and Learning. Conversely, for the top-rated items, such as Confidence, Learning Curve, and Cumbersome, the group with prior EPW experience exhibited more pronounced connectivity.

3.3 Semi-structured Interview

In the disabled group, many comments were related to operation and function, particularly emphasizing the need for smoother operation of the front wheels to enhance driving usability (#S06: “The front wheels should operate smoothly to enhance driving usability”). Concerns were raised regarding the height of footrests (#S13, #S18, #S23: “Footrest height needs to be adjusted”) and handles (#S22: “Handles should be taller”), (#S28, #S29, #S33: “Handles work too sensitive), especially among individuals with severe disabilities, highlighting challenges in driving convenience and ease of use (#S12: “Driving convenience and ease of use is not good enough for people with severe disabilities”).

On the other hand, in the elderly group, feedback indicated that operation was easy, leading to convenient driving experiences (#S26: “Comfortable and easy to operate, making it easy to drive”). Participants generally found the wheelchair operation straightforward and easy (#S08, #S11: “Overall, it was fine and good”).

4 Discussion

Significant differences were observed between elderly and disabled groups, particularly in assessing ease of use and integration, as well as prior experience conditions. Significant differences were evident across items assessing trust, confidence, ease of recovery from malfunctions, and feedback on interactions, indicating varying perceptions between the groups. These findings suggest that factors related to previous experience with EPW influenced evaluation outcomes. As the EPW usage duration increases, users develop a better understanding of the device, and their previous experience with similar devices likely influenced the evaluation of the developed EPW assessment in this study. This trend could be attributed to long-term users’ familiarity with previously used devices, leading them to scrutinize newly developed ones more critically.

5 Conclusion

These insights highlight the critical need for customized design approaches that address the varying requirements of different user groups, ensuring EPW effectively caters to the diverse needs and preferences of users. The findings are also crucial in advancing User-Centered Design practices tailored explicitly for EPW development and implementation. EPWs can be optimized to meet individual users’ unique needs better, resulting in enhanced satisfaction levels and overall quality of life improvements.

Acknowledgments. This work was supported by Incheon National University Research Grant in 2024.

References

1. Candiotti, J.L., et al.: Usability evaluation of a novel robotic power wheelchair for indoor and outdoor navigation. *Arch. phys. med. rehabil.* **100**(4), 627–637 (2019)
2. Wang, H. Development and evaluation of an advanced real-time electrical powered wheelchair controller. Doctoral dissertation, University of Pittsburgh (2012)
3. Gudgeon, S., Kirk, S.: Living with a powered wheelchair: Exploring children’s and young people’s experiences. *Disability and rehabilitation. Assis. technol.* **10**(2), 118–125 (2015)
4. Simpson, R.C., LoPresti, E.F., Cooper, R.A.: How many people would benefit from a smart wheelchair? *J. Rehabil. Res. Dev.* **45**(1), 53–71 (2008)
5. Panchea, A.M., Todam Nguenpang, N., Kairy, D., Ferland, F.: Usability evaluation of the smartwheeler through qualitative and quantitative studies. *Sensors*. **22**(15), 5627 (2022)
6. Whittington, P., Dogan, H.: SmartPowerchair: Characterization and usability of a pervasive system of systems. *IEEE Trans. Hum.-Mach. Syst.* **47**, 500–510 (2017)
7. Seong, SujinSeong/electric-powered-wheelchair. <https://github.com/SujinSeong/electric-powered-wheelchair> (2024)
8. Brooke, J.: SUS-A quick and dirty usability scale. *Usability eval. ind.* **189**(194), 4–7 (1996)



Proposing Heuristic Guidelines for Auditory Interfaces in Medical Device Usability Evaluation

Junghoon Shin¹ , Kyungwoo Kwon¹, and Kiechan Namkung²  

¹ Interaction Design Laboratory TED, Kookmin University, Seoul 02707, Republic of Korea

² Department of AI Design, Kookmin University, Seoul 02707, Republic of Korea
soundux@kookmin.ac.kr

Abstract. This research introduces heuristic guidelines for evaluating auditory interfaces in medical devices, focusing on integrating often-neglected auditory evaluations in usability assessments. Utilizing the IEC 60601–1-8 standard, pivotal auditory components impacting device safety and efficiency were identified and structured into a heuristic framework. The process involved expert user experience professionals using closed card sorting to draft and refine these guidelines, supported by further expert reviews. The study underscores the capability of these guidelines to uncover previously undetected usability and safety issues in medical devices, marking significant progress in incorporating comprehensive auditory interface assessments into usability evaluations. This advancement is vital for enhancing the effectiveness of user-centered medical device designs, providing a structured approach to integrating auditory interface evaluations that have been historically overlooked in the field.

Keywords: Usability · UEQ-MD · Medical device · Auditory Interface · Formative Evaluation · Heuristic Guidelines

1 Background

For many years, the field of medical device design has emphasized systematic safety design to ensure patient safety [1]. The consideration of usability and user experience (UX) in medical device design can enhance efficiency and safety. However, research in this area has not been vigorously pursued [2]. Recently, the importance of user-centered design in medical devices has been increasingly emphasized, supported by the rise in international standards related to usability [3–6]. Most countries now require usability evaluations as a mandatory component to validate these international standards. Currently, usability evaluations predominantly focus on the interaction with visual user interfaces of devices equipped with displays, whereas evaluations related to auditory interfaces are relatively underdeveloped.

The evaluation of auditory interfaces in medical devices involves assessing a range of sound environments, not merely identifying sources of high noise levels [7]. The standard IEC 60601–1-8 [8] addresses auditory and voice alarm signals in medical

electrical equipment, detailing aspects related to safety and effectiveness. However, these guidelines are challenging to apply in real-world usability evaluation settings.

The aim of this research is to assess auditory interfaces in medical devices to facilitate designs that are both efficient and safe. This involves applying heuristic evaluation principles and the Medical Device User Experience Questionnaire (UEQ-MD) derived from previous studies to propose heuristic principles for evaluating auditory interface during formative usability testing of medical devices.

2 Literature Review

2.1 Usability Evaluation Process and Standards for Auditory Assessment

Medical device usability evaluation is classified into formative and summative assessments. It is recommended to conduct at least one formative evaluation before the summative assessment, and it is effective to perform small-scale formative evaluations 2–3 times [9]. Evaluating user interfaces requires significant resources, hence simpler methods like expert review are beneficial during formative evaluations [10].

The IEC 60601–1-8 standard, which mentions specific frequencies, rise and fall times, waveforms, sound levels (dB), pulse widths, repetition rates, and harmonics, addresses the auditory specifications related to medical devices [8]. However, these details mostly cover the acoustic aspects of auditory alarm signals, and methods evaluating emotional responses from users are not provided. Additionally, the specifications allow manufacturers to adjust settings, which complicates the effective evaluation of auditory interfaces in practice.

2.2 Heuristic Evaluation and UEQ

Heuristic evaluation is a widely used expert evaluation method in the field of user experience, where experts use established principles to inspect interfaces and identify usability issues, evaluating aspects such as severity, frequency, and importance [11]. This method is cost-effective and intuitive, suitable for early development stages. The Food and Drug Administration (FDA) has proposed heuristic analysis as an analytical approach to review and evaluate user interactions with devices [12]. While heuristic evaluation is an effective method for preliminary identification of usability issues in medical device interfaces, it is not yet a standard method in medical device usability evaluations [13].

The User Experience Questionnaire (UEQ) serves as a tool for directly and rapidly assessing product user experience, offering modular scales that can be customized for specific research questions [14]. This reliability and validity make it a complement to other evaluation methods [15]. Specifically, the Medical Device User Experience Questionnaire (UEQ-MD) has been developed to reflect the particular characteristics of medical devices [17], comprising scales for dependability, efficiency, Perspicuity, Trust, Result quality, and usefulness. Each scale evaluates aspects of user experience essential for assessing the usability of medical devices.

3 Method

This study aimed to define heuristic evaluation principles for auditory interface in medical device design. To achieve this, eight user experience experts were engaged to identify auditory evaluation factors and conduct closed card sorting. The resulting heuristic guideline were subsequently verified and refined with three experts. The development of heuristic guideline can be validated and improved based on feedback from 3–5 experts [11, 16].

3.1 Extraction of Auditory Interface Evaluation Factors Based on Standards

Evaluation key factors were derived based on the IEC 60601–1-8 standard, which covers auditory and alarm signals for heuristic evaluation. This included considerations of safety and effectiveness as specified in the standard, encompassing auditory elements that significantly impact the safety and performance of medical devices.

3.2 Closed Card Sorting According to UEQ MD Classification

The identified key evaluation elements were classified into six categories as specified by the Medical Device User Experience Questionnaire (UEQ MD) using closed card sorting techniques.

3.3 Expert Evaluation and Validation

The auditory experience of dermatological medical devices was evaluated by three experts (see Table 1). These experts used the derived auditory heuristic principles for evaluation, recording any issues identified during the process. Post-evaluation, feedback was collected to refine and amend the heuristic principles.

Table 1. Profile of Experts.

No.	Occupation	Experience	Education
1	Usability Engineer for MD	8 years	Master’s degree
2	UX Designer	14 years	Doctoral candidate
3	Auditory UX Designer	20 years	Ph.D

Table 2. Auditory Interface Evaluation Factors based on Standards(IEC 60601–1–8).

Scale	Description
Subclause A. General Guidelines and Basis (14)	Clarity of alarms / Perception of urgency and severity / Signal distortion / Consistency of signal quality / Signals according to risk levels / User awareness time / Rapid recovery in case of errors / Consistent signal cycles / Distinct error guidance signals / Awareness of start and end of visual interface operations / Signals considering the operating environment / Simultaneous operation of alternative signals besides auditory signals / Location and device identification capability / Promptness of alarms and signals
Subclause B. Marks and Labels for ME Equipment and ME Systems (3)	Practicality of information / Accessibility of information access / Use of standard terminology
Subclause C. Indicator Symbols (3)	Consistency and integration of visual interface / Help for signals / Guidance for indicators other than signals
Subclause D. Guidelines for Auditory Alarm Signals (1)	Volume, frequency, and alarm frequency adjustment
Subclause E. Speech Alarm Signals (3)	Differentiation between multiple alarm systems / Validity of auditory alarm signals / Priority between signals
Subclause G. Auditory Alarm Signals (3)	Number of auditory and alarm signals / Validity of alarm signals / Operability of emergency signals

4 Result

4.1 Derivation of Auditory Interface Evaluation Factors Based on Standards

The study reviewed the guidelines and annexes of the IEC 60601–1–8 standard [8], assessing the requirements for auditory and alarm signals in medical devices. This review led to the identification of 27 auditory interface evaluation factors (see Table 2).

4.2 Heuristic Evaluation and Expert Validation for Auditory Interfaces

The identified 27 auditory evaluation factors were classified according to the UEQ MD scale, and each factor was defined. After drafting a heuristic guideline for auditory interfaces, an expert evaluation was conducted on medical devices. This evaluation uncovered a total of 32 issues that were difficult to detect with previous methods, indicating that auditory heuristic principles could significantly enhance the usability and safety of user interfaces.

Further, expert reviews suggested the need to re-evaluate the appropriateness of the UEQ MD scale and the classification of evaluation factors. Ultimately, the heuristic

evaluation factors and their detailed definitions for auditory interfaces were refined and presented in Table 3.

Table 3. Evaluation factors of auditory interface classified UEQ MD Scale.

Scale	Evaluation Factors of auditory interface	Definition
Controllability	Guidance for indicators other than signals	Does the system provide signals other than alarms to allow users to quickly identify system-related errors?
	Rapid recovery in case of errors	Are appropriate instructions provided to users in case of errors?
	Volume, frequency, and alarm frequency adjustment	Does adjustment of alarms and signals enable flexible handling in the operating environment?
Efficiency	Number of auditory and alarm signals	Are appropriate alarm signals and the number of signals provided to prevent user confusion?
	User awareness time	Are sounds that are easy to perceive in a medical environment provided to allow users to quickly recognize the device?
Learnability	Distinct error guidance signals	Are error-related signals clearly distinguished from general event signals and delivered clearly?
	Signals according to risk levels Priority between signals Perception of urgency and severity	Is the priority between signals clear according to the risk level, allowing users to perceive urgency and severity?
	Use of standard terminology	Are standard words used to ensure that users can understand the intended messages
	Validity of alarm signals Clarity of alarms	Are alarm signals clear and effectively audible to users?
	Awareness of start and end of visual interface operations	Are signals for starting and ending operations clear?
	Validity of auditory alarm signals	Is the effectiveness of auditory alarm signals (intensity, Voice format, pronunciation, emotional response, etc.) such that users can quickly understand the alarms?

(continued)

Table 3. *(continued)*

Scale	Evaluation Factors of auditory interface	Definition
	Differentiation between multiple alarm systems	Can important alarms be quickly identified in situations where multiple alarms are sounding?
Trust	Operability of emergency signals	Does the device operate quickly in emergency scenarios to respond promptly to urgent situations?
	Consistent signal cycles Consistency of signal quality Signal distortion	Are signals output consistently in terms of cycle and quality to avoid confusion in alarms?
	Promptness of alarms and signals	Do alarms and signals operate promptly without delay?
Result Quality	Simultaneous operation of alternative signals besides auditory signals	Do alternative feedback mechanisms outside of auditory and alarm signals operate accurately?

*(continued)***Table 3.** *(continued)*

Scale	Evaluation Factors of auditory interface	Definition
	Awareness of start and end of visual interface operations	Is the information provided consistent between auditory and visual interfaces?
Usefulness	Signals considering the operating environment	Are alarm sounds clear and audible without interference from ambient noise, allowing them to be distinguished quickly without confusion with alarms from other situations?

(continued)

5 Conclusion

This research developed a heuristic guideline for assessing auditory interfaces in formative usability evaluations of medical devices, analyzing standards and annexes to identify evaluation factors categorized by the previously established UEQ-MD scale. The guidelines, refined through expert evaluations into a questionnaire format, introduce the first system capable of integrating auditory interface assessments into medical device usability evaluations—an area previously unexplored. While implementing these guidelines may expose unforeseen issues, marking a limitation of this study, future research aims

Table 3. (continued)

Scale	Evaluation Factors of auditory interface	Definition
	Help for signals Practicality of information Accessibility of information access	Does the system enhance user understanding of alarms and enable them to take appropriate actions?
	Location and device identification capability	Does the system allow for recognition of location or individual patients through alarms?

to use these guidelines for actual evaluations, seeking to create a more systematic and reliable framework.

References

1. Donaldson, M. S., Corrigan, J. M., & Kohn, L. T. (eds.). (2000). *To Err is Human: Building a Safer Health System*. Washington (DC): National Academies Press (US)

2. Bitkina, O.V., Kim, H.K., Park, J.: Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. *Int. J. Ind. Ergon.* **76**, 102932 (2020)

3. International Organization for Standardization. ISO 13485:2016 Medical devices - Quality management systems - Requirements for regulatory purposes (2016)

4. International Electrotechnical Commission. (2005). IEC 60601-1:2005 Medical electrical equipment - Part 1: General requirements for basic safety and essential performance. Edition 3.0

5. International Electrotechnical Commission: IEC 62366-1:2015/AMD1:2020 Medical devices - Part 1: Application of usability engineering to medical devices. Amendment 1. International Electrotechnical Commission (IEC), Geneva, Switzerland (2020)

6. International Electrotechnical Commission. (2016). IEC TR 62366-2:2016 Medical devices - Part 2: Guidance on the application of usability engineering to medical devices

7. Ruettgers, N., Naef, A.C., Rossier, M., Knobel, S.E., Jeitziner, M.M., Gerber, S.M.: Perceived sounds and their reported level of disturbance in intensive care units: A multinational survey among healthcare professionals. *PLoS One.* **17**(12), e0279603 (2022)

8. International Electrotechnical Commission. (2006). IEC 60601-1-8 Medical electrical equipment - Part 1-8: General requirements for basic safety and essential performance - Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems

9. GE HealthCare: Alarm Management White Paper. Available online at: <https://clinicalview.gehealthcare.com/white-paper/alarm-management-white-paper>. (2020)

10. Ministry of Food and Drug Safety (MFDS). Practical Guide on Common Standards for Electrical and Mechanical Safety of Medical Devices. Available online at: https://www.mfds.go.kr/brd/m_218/view.do?seq=30856 (2017).

11. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Chew, J.C., Whiteside, J. (eds.) *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 249–256. Association for Computing Machinery, New York NY United States (1990)

12. Food and Drug Administration: Applying Human Factors and Usability Engineering to Medical Devices. Guidance for Industry and Food and Drug Administration Staff. Available online at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices> (2011)
13. Roma, M.S.G., de Vilhena Garcia, E.: Medical device usability: Literature review, current status, and challenges. *Res. Biomed. Eng.* **36**, 163–170 (2020)
14. Sularsa, A., Prihatmanto, A.S., Nugroho, E.: Evaluasi User Experiences Produk iDigital Museum dengan Menggunakan UEQ. *J. Teknol. Inf.* **2**(2), 56–62 (2015)
15. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008*, Graz, Austria, November 20–21, 2008. *Proceedings 4*, pp. 63–76. Springer, Berlin Heidelberg (2008)
16. Quiñones, D., Rusu, C., Rusu, V.: A methodology to develop usability/user experience heuristics. *Comput. stand. interfaces.* **59**, 109–129 (2018)
17. Lindemann, M., Weber, H.: UX in Zahlen: Der user experience questionnaire for medical devices. In: *Usability Professionals 23. Gesellschaft für Informatik e.V., Rapperswil*, September 3–6, 2023 (2023). <https://doi.org/10.18420/muc2023-up-455>



Research on a Water-Drawing Robot Using Appropriate Technology: Focusing on Interaction Methods and User Experience

Yunjeong Song^{1(✉)}, Jieun Sa², Jaehoon Cha³, Juseong Kim⁴, and Kicheol Pak⁵

¹ Visual Communication Design, Hongik University, 94, Wausan-ro, Mapo-gu, Seoul, Republic of Korea
hide@hongik.ac.kr

² Design Management, Hongik University, 94, Wausan-ro, Mapo-gu, Seoul, Republic of Korea

³ Mechanical Engineering, Hongik University, 94, Wausan-ro, Mapo-gu, Seoul, Republic of Korea

⁴ Department of Design, Kangwon National University, 1, Gangwon Daehak-gil, Chuncheon, Republic of Korea

⁵ Department of Mechanical and System Design Engineering, Hongik University, 94, Wausan-ro, Mapo-gu, Seoul, Republic of Korea

Abstract. Robots play a crucial role in various industries, and while humanoid robots often come to mind when considering robots that assist humans, their implementation is technically challenging and has limitations to replace human roles. Therefore, research is needed to enhance human experiences through appropriate technology. This study explores and verifies whether the application of appropriate technology, in this case water drawing, can provide positive value to users. It also proposes a modality to improve communication between robots and humans through non-verbal interaction. The study was conducted through a survey showing scenario videos of adult, child, and elderly to the subjects, and having them evaluate how the interaction methods of the water-drawing robot affect likability, satisfaction, and perceived value in various situations. The results showed that the water-drawing robot positively influenced likability, satisfaction, and perceived value, with a high correlation among the three variables. Therefore, the interaction of the water-drawing robot positively affects user experience, and using appropriate technology like water drawing can provide positive experience, suggesting that the adaptation of appropriate technology can enhance valuable experiences and positively impact humans.

Keywords: Water Drawing · Human-Robot Interaction · Appropriate Technology · Non-verbal Communication · Human Factor

This paper was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE)
(P0012725, HRD Program for Industrial Innovation)

1 Introduction

1.1 Research Background

Robotics, which has been primarily used in industrial settings, is advancing in everyday life in forms such as guide robots and serving robots. In the past, robots performed repetitive tasks to assist humans, but with the advancement of technology and AI, the use of robots has expanded to interact with users in various situations. Although humanoid robots could be considered as an ideal form to assist humans, the technologies are still too complicated to be adapted in real life. Therefore, further studies on robots using appropriate technology are necessary.

The term ‘appropriate technology’ originally refers to technology that considers the political, cultural, and environmental conditions of the community where it is used, enabling sustainable production and consumption, and ultimately improving the quality of human life. Generally, it aims to produce products using locally sourced materials suitable for the local environment. On the other hand, in this study, we redefined the term appropriate technology which means a technology that uses low technology and simple mechanisms.

From a previous study [1], a water-drawing robot called WADE was created to provide new experiences and entertainment targeted for amusement parks. This robot can print an image on the ground using water instead of paint or ink, which also enables sustainability since it does not cause any pollution on the surface. This study aims to understand how robots with appropriate technology could enhance user experience.

1.2 Research Objectives

This study empirically analyzes whether the emotional interaction of a water-drawing robot can provide positive value to users. The purpose of this study is to develop a robot that can perform human roles using appropriate technology, thereby enhancing the user experience. Instead of using highly advanced technology, this study focuses on exploring new possibilities in robotics with appropriate technology and proposes practical robots for daily use.

2 Research Questions and Methods

In this study, we reviewed 11 previous studies in the HRI (Human-Robot Interaction) field to identify the characteristics of emotional expression elements in robots. The review showed that most of these studies mainly focused on non-verbal communication (body movements, eye movements, facial expressions, paralinguistic cues, anthropomorphic expressions, etc.).

According to Han Hoon, due to human psychological tendencies, robots are perceived as social entities, making the design of robots’ emotional and intentional expressions crucial [2]. Soo-ah Kim stated that motion language plays a significant role in supplementing or independently conveying meaning, enhancing intimacy, and activating communication [3]. These previous studies confirm that non-verbal expressions significantly impact communication between robots and users. Therefore, this study aims

to investigate whether non-verbal elements using appropriate technology can enhance emotional expression in robots and improve user experience [4–14].

2.1 Research Method (Model and Hypotheses)

The WADE robot can express emotions through non-verbal communication modalities such as eye movements, head movements, and water drawing technology, as mentioned in previous studies. This study aims to evaluate the water drawing technology whether it can be considered as a third interaction modality and have a significant impact on users among the other two (eye movements and head movements). The following Table 1 and Fig. 1 gives a research model and hypotheses are set as follows (see Fig. 1).

Table 1. Research questions and hypotheses.

No.	Research questions and hypotheses.
H1	Does water-drawing positively impact user experience?
H1-1	Anthropomorphized interactions incorporating water-drawing technology will positively influence likability.
H1-2	Anthropomorphized interactions incorporating water-drawing technology will positively influence satisfaction.
H1-3	Anthropomorphized interactions incorporating water-drawing technology will positively influence perceived value.

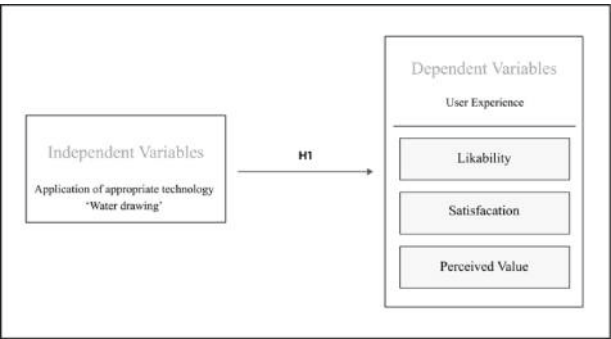


Fig. 1. Research model of a correlation between Independent variables and Dependent variables

2.2 Deriving Evaluation Factor

To derive evaluation factors, studies on usability and user experience evaluations related to robots, products, and services were analyzed, extracting likability, satisfaction, and functionality as applicable evaluation factors. The functionality, in this case, is redefined as perceived value to assess the expected effects as an interaction modality rather than evaluating the water drawing technology itself. The final evaluation factors are organized as follows [15–18] (Table 2).

Table 2. Evaluation Factors.

Variable	Evaluation Criteria (questionnaire)
Likability	Q1. I like the reaction of the robot Q2. The robot felt friendly
Satisfaction	Q3. The robot is expressing an expression that matches each user’s situation. Q4. The water drawing method is satisfactory
Perceived Value	Q5. Robots are performing interactions with appropriate technology Q6. The interaction of the robot meets the expectations

2.3 Experimental Method

The experiment was designed to show three user scenario videos featuring an adult, a child, and an elderly to the subjects, and have them rate the interactions of the water-drawing robot in each scenario on a 5-point Likert scale. Although the WADE robot was designed for entertainment purposes in amusement parks, in this study we changed the environment to a general park setting to evaluate its interactions for more general use (Fig. 2 and Table 3).

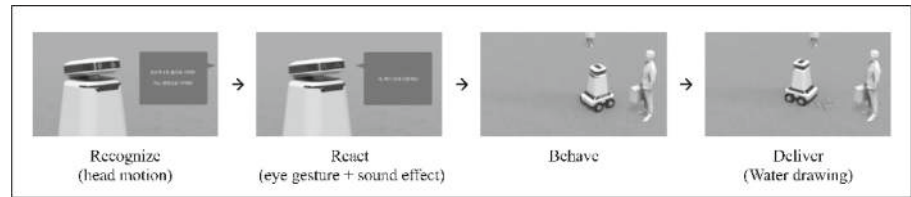


Fig. 2. Interaction sequence of the simulation video

3 Results

3.1 Demographic Analysis and Reliability Verification

The experiment was conducted by online survey May 22 to 27, 2024. Among the 51 participants, 31 were in their 20 s (61%), and 20 were in their 30 s (39%). Gender distribution was similar, with 29 females (57%) and 22 males (43%). In this study, Cronbach’s α coefficient was calculated to verify the internal consistency of the measured factors and ensure their reliability. Factors with a coefficient of 0.6 or higher were considered reliable. The results of the reliability verification indicated that all factors measuring the dependent variables scored 0.6 or higher, confirming that the measurement indicators meet the required reliability standards (Tables 4 and 5).

Table 3. Interaction Scenarios.

Target user group	Non-verbal communication methods	Water-drawing	Application Scene
Adult	Eye movements pseudo-voice nodding	Composite information (arrows + text)	
Child	Eye movements pseudo-voice nodding	Emotional expression (emoticons)	
Elderly	Eye movements pseudo-voice nodding	Graphic information (arrows)	

Table 4. Demographic Analysis.

Variable	Description	Frequency (n = 51)	Percentage (%)
Age group	20 s	31	61
	30 s	20	37
Gender	Male	22	43
	Female	29	57

Table 5. Descriptive Statistics and Reliability Verification.

Variable	Metrics	M	SD	Sk	K	Cronbach's α
Likability	Q1	3.98	.948	−.984	.983	.802
	Q2	4.14	.980	−.949	−.083	
Satisfaction	Q3	3.61	1.060	−.400	−.630	.842
	Q4	3.86	1.096	−.857	.197	
Perceived Value	Q5	3.55	1.270	−.663	−.545	.786
	Q6	3.88	1.089	−.919	−.343	

3.2 Research Question and Hypothesis Verification Results

The correlation analysis of likability, satisfaction, and Perceived value is shown in Table 6. All three variables positively correlated, affecting each other. Satisfaction and Perceived value had a high correlation of 0.845. The average scores of the variables were calculated, with likability scoring the highest at 4.06. The factor asking about the robot's likability specifically scored in the 4-point range. Satisfaction scored 3.74, and Perceived value scored 3.72, both receiving relatively good evaluations.

Table 6. Correlation Analysis.

Variable	Likability	Satisfaction	Perceived Value	M
Likability	1			4.06
Satisfaction	.749**	1		3.74
Perceived Value	.758**	.845**	1	3.72

** $p < .01$

The analysis confirmed that interactions including water drawing positively impacted user satisfaction, likability, and Perceived value. Particularly, water drawing technology, combined with anthropomorphic emotional expression, elicited positive responses in overall likability and satisfaction. Additionally, as a sustainable technology using water, water drawing significantly influenced users' Perceived value evaluation.

4 Conclusion

4.1 Research Results and Discussion

This study aimed to confirm the potential of interaction using appropriate technology to enhance user experiences. The analysis showed that the emotional interaction of the WADE robot with water-drawing technology positively impacted users in terms of likability, satisfaction, and perceived value. The highest average score was for likability, followed by satisfaction and perceived value. These results indicate that anthropomorphic interaction technology, including water drawing, can provide sufficient satisfaction in human-robot interactions. Moreover, providing new experiences through appropriate technology can enhance the quality of human life. Therefore, this study proposes innovative interaction methods for HRI based on the anthropomorphic characteristics of service robots using appropriate technology.

4.2 Research Limitations and Future Directions

Limitation. This study analyzed the positive value of anthropomorphized interactions, including appropriate technology, through empirical research to contribute to improving the quality of human life. Additionally, it focuses on developing human-centered HRI, aiming to contribute to better HRI development in the future. Based on the study results, the following limitations and suggestions for future research are presented:

Future Directions. This study used virtual modeling animations instead of actual robot prototypes for user evaluation. Future studies should address this limitation by implementing real prototypes to assess overall user satisfaction and experience more accurately.

The experiment tested different scenarios and interaction methods for each age group, verifying the value of water-drawing technology. However, the study lacked control groups for comparison. Future research should investigate the impact of water-drawing technology on interaction enhancement by including control groups.

The study was conducted online with a small sample of 51 participants, primarily in their 20 s and 30 s. Future research should involve a larger sample size and broader age range for more accurate results. Additionally, qualitative interviews should be conducted to analyze the impact on user experience and propose practical design suggestions.

References

1. Ko, S.: Design and system layout of water drawing service robot. *HCI Korea*. **35**(2), 1237–1240 (2024)
2. Han, H.: The nonverbal emotional expression of AMR(autonomous mobile robot) based on the circumplex model, Korea (2023)
3. Kim, S., Oh, D.: The Kinesics for emotional expression in the household social robot- Focus on the head movement of the robot. *Korea Soc. Sci. Art*. **35**, 69–82 (2018)
4. Oh, Y.J., Shin, Y.S., Lee, J.H., Kim, J.W.: A study on interaction design of companion robots based on emotional state. *J. Digit. Contents Soc.* **18**(7), 1293–1301 (2017)
5. Kang, B., Jun, S.: User perception of social robot's emotional expression and forms of anthropomorphism. *Arch. des. res.* **35**(2), 137–153 (2022)
6. Park, D.: A study on human-robot's emotional communication through the movement of the eye of a social robot. *J. Korea Soc. Vis. Des. Forum*. **63**(0), 129–138 (2019)
7. Kim, W.: A study on nonverbal communication of social robots - focusing on emotional expression using movement. *Bull. Korean Soc. Basic Des. Art*. **24**(5), 95–110 (2023)
8. Choi, C.: A study on the visual standardization of facial expressions to develop emotional expressions of humanoid robot design. *Korea Des. Knowl. Soc.* **24**, 125–138 (2012)
9. Jung, J.: Optimal human-robot interaction facial expression design based on derivation of expression emotional intensity factors of an autonomous quarantine robot. *Arch. des. res.* **36**(3), 317–339 (2023)
10. Lee, S.: Design of the emotion expression in multimodal conversation interaction of companion robot. *Des. converg. study*. **16**(6), 137–152 (2017)
11. Kim, D.: The effect of expressions of self-consciousness in social robots during social attraction. *J. KIISE*. **46**(7), 653–663 (2019)
12. Son, H.: The effect of parents' technology acceptance factors on their infants' satisfaction and acceptance intention with social robots. *HCI soc. Korea*. **19**(1), 41–51 (2024)
13. Jang, M.: The effect of the use of humor in artificial intelligence on the emotional experience of users, pp. 266–271. *The HCI society of Korea* (2021)
14. Ha, S.: A study on the mechanism of social robot attitude formation through consumer gaze analysis: Focusing on the face of the robot. *Regul. inf. res.* **28**(1), 243–262 (2022)
15. Kim, J.: Satisfaction study on meal assistance robot for people with severely disability and their caregivers- focused on K-QUEST and improvement in the meal assistance. *Proc. Korean Soc. Rehabil. Welf. Eng. Conf.*, 119–121 (2021)

16. Lee, J.: The impact of design innovation on product affinity and intention to Buy-Bluetooth speaker focused on. *J. Korea Inst. Inf. Commun. Eng.* **25**(2), 597–599 (2021)
17. Lee, J.: Tourism images of influencing satisfaction. Revisir intention, willingness to recommend. *Korean Acad. Mark. Sci. Fall Int. Conf.* **2006**(3), 538–555 (2006)
18. Lee, J.: An empirical study of the influence of automobile attributes on overall satisfactions, repurchase intentions, recommendation intentions: Focusing on Chinese consumers. *Int. Bus. Rev.* **14**(1), 19–54 (2010)



Urban Mobility Platforms and Trends Towards a MaaS Scenario

Mariana Menezes O. Sousa^(✉) , Manuela Quaresma ,
Carlos Eduardo Félix da Costa , Claudia Mont'Alvão , and João Bonelli 

Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
mmos@aluno.puc-rio.br

Abstract. This research encompasses an interface pattern analysis of urban mobility platforms from 20 smart cities stated in the Smart City Index 2023. The study identifies main features in mobility apps, exploring how interfaces aim to promote a seamless experience in commuting for citizens and cities. This article aims to present an overview of the current scenario of urban mobility platforms and how they can be designed towards MaaS, considering both services and user experiences. The methodology involves mapping cities using the Smart City Index 2023, urban mobility applications, and analyzing the usability of selected apps. The findings provide a visual repertoire of mobility platform interfaces, from which most follow the same interaction model, with some variations. A discussion on similarities and differences of the user experiences of various urban mobility platforms is also stated, as well as a synthesis of various platforms' features and their main characteristics. Finally, the research allows the suggestion of parameters for building MaaS platforms centered on citizens' usage of public transportation and multimodal travel experiences.

Keywords: Mobility as a Service · User Experience · Platform Usability Analysis

1 Introduction

The concept of Mobility as a Service (MaaS) means “the integration of various forms of transportation services into a single accessible on-demand mobility service” [1]. It consists of an aggregation of multiple public and private transport modes and services, with the aim of promoting the best door-to-door travel offer for its users. MaaS allows people to plan and to use public and private transport services in their journey in a combined way and offers payment for all passes on a single platform.

In the context of smart cities, several topics must be addressed, such as smart economy, smart governance, smart environment, smart mobility, smart population, and smart life [2]. In mobility, various technologies and services can be used to optimize the commute of citizens and thus reduce travel time and the emission of gasses released by combustion-powered vehicles: from services focused on traffic monitoring to assistance to optimize public transport, as well as new services and products that can be offered to

travelers linked to vehicle connectivity with the city's various sensors. These services and technologies, when properly integrated, can lead to continuous mobility, key to a smart sustainable city.

Therefore, smart cities are developing digital platforms to address urban mobility and transportation needs, which offer various transport modes and services and provide door-to-door commuting options. To achieve seamless mobility, there are several opportunities and risks to work on the evolution of urban mobility, in four main layers: 1) the base layer of physical infrastructure (roads, railways, etc.); 2) the second layer of rolling stock (cars, buses, trains, bicycles, etc.); 3) the third layer of data and analysis (ticketing and payments, maps, correspondence, etc.); and 4) the final user interface layer (navigation applications, payment integration, etc.) [3], which is investigated in this research.

In smart cities, where complex transportation systems take place, MaaS platforms are desirable in order to improve mobility efficiency and establish a consistent, accessible and easy-to-use transportation system that benefits citizens and tourists. Nevertheless, the creation of a MaaS platform relies on information from several databases, which is especially challenging for metropolises in which there are varied modes of transport, each with its own system. Communication with citizens through an interface must therefore be one of the main points of contact to be worked on in the urban mobility service, so that it improves the user experience when using complex urban transport networks.

The possibility of dealing with incompatible app structures and user tasks' presentation may become an issue when combining features and mobility service providers to create MaaS solutions that are seamless and adequate to users' expectations. This gap can be filled by comprehending the interface patterns among urban mobility platforms and applications, which may provide a repertoire of how the main user tasks, such as travel planning and payment, are presented. Additionally, it is important to reveal if those user tasks follow the same visual and interaction structure, and if there is a need to increase resemblance to the interface patterns.

In this sense, it becomes necessary to analyze mobility platforms and their developments in MaaS platforms, to achieve goals such as reduction in car traffic, ease of movement by public transport, reduction of sedentary lifestyle through the promotion of active transport and reduction of pollution through sustainable transport alternatives. Therefore, this article aims to present an overview of the current scenario of urban mobility platforms and how they can be designed towards MaaS, considering both services and user experiences. While it is pivotal, further analysis of mobility platforms is needed to understand the similarities and differences in user experiences across them.

2 Methodology

The methodology involves mapping cities using the Smart City Index 2023, urban mobility applications, and analyzing the usability of selected apps, divided in stages of analysis: (1) analysis of cities and classification according to their relevance; (2) survey of platforms from the top 20 cities in the 2023 Smart City Index (SCI) ranking and from other cities with good mobility rating in the SCI and further classification of mobility platforms according to relevance criteria; (3) detailed analysis and comparison of selected mobility applications, considering their main features and interface patterns, as they might serve as a model for building future MaaS platforms.

2.1 Selection of Cities from the Smart City Index 2023

First, we identified cities considered ‘smart’, especially those with complex mobility systems and platforms that could be used as study cases. Therefore, the Smart City Index (SCI) of 2023 – an index of smart cities proposed by the International Institute for Management Development (IMD) – was used as reference for mapping cities. The index evaluates some areas such as health and safety, governance and mobility from several cities, based on citizens’ opinions. The 2023 edition of the index comprises 141 cities. From those, we selected the top 20 as well as cities with a high mobility score.

Initially, we considered the top 20 cities, from which twelve are in Europe – Zurich, Oslo, Copenhagen, Lausanne, London, Helsinki, Geneva, Stockholm, Hamburg, Prague, Amsterdam and Munich –; six are in Asia– Singapore, Beijing, Hong Kong, Seoul, Abu Dhabi and Dubai –, of which the last two are located in the Middle East; and two in Oceania – Canberra and Sydney.

Besides the top 20, other cities were included in our selection based on the following criteria: overall score (alphabetical and numerical), mobility criteria scores, structures and technology scores, presence in previous editions of the index, stable or growing classification and population – we prioritized population above one million due to the higher complexity in mobility it imposes. Each criterion was graded and this process resulted in the addition of 22 cities to our list: Nanjing, Hangzhou, Guangzhou, Shanghai, Zhuhai, Shenzhen, Tianjin, Chongqing, Chengdu, Hanoi, Taipei, Riyadh, Vienna, Jakarta, City of Ho Chi Minh, Mumbai, Bengaluru, Madrid, Mecca, Delhi, Cairo and Hyderabad. Then, the next step was searching for platforms from the selected cities in the AppStore and other platform catalogs.

2.2 Selection of Urban Mobility Platforms

Based on the comparative analysis between SCI 2023 cities, the survey of mobility platforms began with the selected cities as a reference. The method used to find platforms consisted of consulting the App Stores of different countries – especially those where the selected cities are located – and the databases Product Hunt, Mobility Pages, Page Flows, Mobbin, Screenlane, GoodUX Appcues, UX Archive, Tracxn and Design Vault. Thus, 93 platforms were listed, 82 of which were mobile apps, and 17 screen flows. The results were recorded in a table with the following information: coverage; format; rating and reviews in the AppStore from its country of origin or a country with a substantial number of users; features; and tags.

With the information collected, it was possible to measure the coverage of platforms by continent. Therefore, of the 93, 50 cover Asia, 38 America, 38 Europe, 12 Oceania and 11 Africa. It is important to highlight that some platforms have global reach, and therefore, although cities on the American continent did not appear in the survey, platforms that cover America do.

Among the data collected, AppStore ratings and number of reviews were used to identify the most relevant platforms to concentrate efforts on a more detailed analysis of such platforms. Therefore, 27 applications were selected, due to having more than **50,000 evaluations** and a score **greater than or equal to 4.0**, which we classified as most relevant. It is important to state the relationship between those apps and the smart

cities. Except for Ruter, Tube Map, MTR Mobile and the South Korean trio (Subway Korea, Kakao T and Kakao Map), most platforms have an extensive coverage, with international reach, so in the Table 1 we mention only some of the cities covered by each app.

Table 1. Table of urban mobility apps considered most relevant. Source: authors.

<i>Ruter</i> (Oslo)	<i>Tube Map</i> (London)	<i>Zipcar</i> (Taipei, London...)
<i>Subway Korea</i> (Seoul...)	<i>Kakao T</i> (Seoul...)	<i>Kakao Map</i> (Seoul...)
<i>AllTrails</i> (Seoul, Guangzhou, Zurich, Oslo, Copenhagen, Lausanne, London, Helsinki, Geneva, Stockholm, Hamburg, Prague, Amsterdam, Munich, Vienna, Madrid, Singapore, Delhi, Sydney...)	<i>Moovit</i> (Hamburg, Munich, Canberra, Sydney, Shanghai, Beijing, Nanjing, Guangzhou, Chengdu, Hangzhou, Shenzhen, Chongqing, Tianjing, Seoul, Abu Dhabi, Dubai, Madrid, Oslo, London, Prague, Singapore, Stockholm, Geneva, Zurich, Hanoi, Ho Chi Minh, Taipei...)	<i>Uber</i> (Canberra, Zurich, Oslo, Lausanne, London, Helsinki, Geneva, Stockholm, Hamburg, Abu Dhabi, Prague, Amsterdam, Seoul, Dubai, Sydney, Hong Kong, Munich ...)
<i>Grab</i> (Singapore, Jakarta, Hanoi, Ho Chi Minh...)	<i>MTR Mobile</i> (Hong Kong)	<i>Go-Jek</i> (Singapore, Jakarta...)
<i>Careem</i> (Cairo, Dubai, Abu Dhabi...)	<i>Bird</i> (Madrid, Vienna, Abu Dhabi, Helsinki...)	<i>Getaround</i> (Hamburg, Oslo, London...).
<i>Lime</i> (Copenhagen, London, Helsinki, Stockholm, Hamburg, Dubai, Oslo, Prague, Munich, Vienna, Zurich, Seoul, Sydney ...)	<i>Citymapper</i> (London, Hamburg, Munich, Madrid, Stockholm, Copenhagen, Oslo, Helsinki, Prague, Vienna, Zurich, Lausanne, Hong Kong, Singapore, Seoul, Sydney...)	<i>Sixt</i> (Munich, Hamburg, Zurich, Geneva, Oslo, Copenhagen, Canberra, Sydney, Helsinki, London, Stockholm, Amsterdam, Singapore, Abu Dhabi, Dubai, Riyadh ...)
<i>Transit</i> (Sydney, Canberra, Hamburg, Madrid, London...)	<i>Trainline</i> (London, Madrid...)	<i>Ola Cabs</i> (Bengaluru, Delhi, Mumbai, Hyderabad...)
<i>Rapido Bike Taxi</i> (Bengaluru, Delhi, Hyderabad...)	<i>Waze</i> (Zurich, Madrid, Oslo, Lausanne, Copenhagen, London ...)	<i>Didi</i> (滴滴) (Beijing, Canberra, Sydney, Cairo, Hong Kong...)
<i>Blá Blá Car</i> (Prague, Delhi, Mumbai, Bengaluru, Milão, Roma, Hamburg, Munich, Madrid, Amsterdam...)	<i>Google Maps</i> (Zurich, Geneva, Hamburg, Oslo, Sydney, Madrid, Singapore...)	<i>Share Now</i> (Hamburg, Munich, Amsterdam, Copenhagen, Madrid, Vienna...)

2.3 Feature Analysis – Selected Applications

After identifying the 27 most relevant applications, we conducted an analysis of the main features and interface patterns of those mobility applications with comparisons among them. As of the home screen, some present a pinchable map with a field to indicate the desired address, some present vehicles for rental and tourism options and some present transportation filters. Most apps present navigation bars with options related to rides, payment, tickets, account settings, preferences/favorites, rental, tourism, history and/or transport modes. Some have springboards with options regarding rental, transportation modes, maps, train/bus schedules, wallet credits, stores, ticket purchase and news. Navigation drawer menus were also a pattern, with options regarding profile, security, settings, help, payment, coupons/rewards/gift cards, guidance materials, ride history, time tables, line status, news and updates. Some apps have scrolling tabs with tags related to facilities (e.g. cafes, gas stations) and/or transportation modes (e.g. bus, bike, train).

Even though the apps may have differences, it is possible to suggest an ideal model by identifying what they have in common and adding interesting features that may be one of a kind. Therefore, based on the observed features, a urban mobility platform designed towards MaaS should: allow users to plan multimodal trips; track real-time itinerary of nearby buses and trains; compare route options; receive updates on traffic conditions; rent vehicles; request rides by car or motorcycle; save favorite settings/places; check station facilities, products and parking spots; purchase trips that may combine public and private transports; search for tourism options; filter by accessible, active, sustainable, cheapest or fastest routes; and achieve bonuses, rewards and discounts. By encompassing multiple modes and promoting an extensive and planned travel service with payment options, the platform will fit into the MaaS concept.

3 Results and Discussion

The findings of this research point to a pathway within the field of urban mobility, because they provide comparisons among mobility platforms considering the cities in which they were implemented. Also, they contribute to the field of interface design because they present feature standards in efficient mobility services so that new platforms can adequate themselves to the models to which users are already familiar with. Finally, the research contributes to the field of service design, since it analyzes different mobility services and proposes an ideal standard of MaaS application that can be applied in the creation or improvement of mobility services.

The procedures carried out and results obtained in this stage of the research may be used as reference materials for the development of more sustainable MaaS platforms that promote citizens' well-being through active mobility options and also environmental consciousness through informing the user about the carbon emissions emitted by each trip. It is suggested a deeper investigation on features that depend on GPS and the actual usage of transportation in loco, since interaction with those features was limited due to the researchers' geographical location. Such contributions will be valuable to promote mobility applications that are aligned with patterns in smart cities.

4 Conclusion

The study presents an overview of the current scenario of urban mobility platforms and how they can be designed towards MaaS, considering both services and user experiences. The offering of an organized interface repertoire and recommendations on the usability of platforms may contribute to the design of new mobility apps towards MaaS platforms. Therefore, the analysis of smart cities and mobility platforms provides parameters to build MaaS platforms adapted to a seamless user experience.

References

1. MAAS ALLIANCE. Guidelines & Recommendations to create the foundations for a thriving MaaS Ecosystem (2017)
2. Kiritmat, A., et al.: Future Trends and Current State of Smart City Concepts: A Survey. *IEEE Access*. **8**, 86448–86467 (2020)
3. Hannon, E., Knupfer, S., Stern, S., Sumers, B., Nijssen, J.T.: An integrated perspective on the future of mobility, part 3: Setting the direction toward seamless mobility demand. McKinsey & Company: McKinsey Center for Future Mobility (2019)
4. EUROPEAN COMMISSION. Smart cities., https://ec.europa.eu/info/eu-regional-and-urban-development/topics/cities-and-urban-development/city-initiatives/smart-cities_en (2024). Accessed 10 Feb 2024
5. International Institute for Management Development (IMD). Index report 2023 – IMD Business School., <https://www.imd.org/wp-content/uploads/2023/04/smartcityindex-2023-v7.pdf> (2023). Accessed 31 Oct 2023



Usability Evaluation of the Fuel Management System for a Flag Carrier Airline

Alma Maria Jennifer Gutierrez^(✉) , Macario Cordel ,
Merlin Teodosia Suarez^(✉) , and Christine Diane Ramos 

De La Salle University, Manila, Philippines
merlin.suarez@dlsu.edu.ph

Abstract. According to several IATA reports, fuel is one of the airline's most significant expenses, accounting for 18% to 30% of its operating costs. This prompted a flag carrier airline to utilize a commercial fuel management system to improve their planning and decision-making on fuel savings and consumption. As such systems are driven by statistical and simulation models that are intensive on numerical and multiple variables, ensuring that the interface design and user experience are adequately factored in the development for sustained use is critical. This work presents the usability evaluation of a fuel management consumption system designed especially for pilots and flight operation personnel and measured against the System Usability Scale. Results indicate that the users are satisfied with the interface and navigation. However, adherence to Nielsen's heuristics, particularly on consistency and standards, can be improved by revising the position of specific controls. Shortcut gestures can also be created to enhance its flexibility and ease of use.

Keywords: Airline · Fuel Management · Usability

1 Introduction

The COVID-19 pandemic severely impacted the aviation industry, with restrictions and lockdowns leading to operational and financial challenges for airlines. Liquidity became a major issue in 2020 and 2021 due to severely reduced passenger sales and revenue. Lockdown measures and travel restrictions significantly affected airline operations, causing a decline in cash flow. With the crucial financial position of most airline companies, they must undertake cost-saving measures to help ease their economic struggles. Fuel is one of the most significant resources they need, which is also why it is one of its most essential expenses. Consequently, efficient consumption of this resource can significantly lessen the company's expenses, leading to higher income. Rather than using static historical data as reference for applying the fuel saving techniques to conserve fuel, a decision tool that could project a more realistic amount of fuel savings depending on the combinations of fuel savings techniques application and aircraft and environment variables would be more advantageous. As airlines increasingly rely on digital solutions

to optimize fuel efficiency and reduce operational costs, the interface of fuel management systems plays a crucial role in facilitating effective utilization by pilots, engineers, and other relevant stakeholders. However, current literature primarily emphasizes the technical capabilities and data analytics functionalities of these systems, overlooking the importance of user interface (UI) design in enhancing user experience and operational performance. In this work, we designed a user interface for a fuel management system and evaluated the usability aspects of the interface in terms of ease of navigation, information accessibility, task efficiency, and user satisfaction.

2 Related Works

Various tools (Boeing, 2022; General Electric, 2022; FuelerLinx, 2017; OpenAirlines, 2020; i6 Group, 2022; Honeywell Aerospace, 2020; Skymetrix, 2021; StorkJet, 2021; Safety Line, 2022) have been introduced to address the challenges posed by the recognition of fuel as a significant operating expense for airlines, coupled with growing concerns regarding carbon emissions. These solutions offer a range of features, including advanced analytics, data processing capabilities, predictive maintenance insights, and integration with other airline operations. For instance, the Fuel Dashboard (Boeing, 2022) provides comprehensive analytics capabilities and real-time insights into fuel performance, contributing to significant savings for airlines such as Flybe and Kalitta Air. Similarly, Fuel Insight Software (General Electric, 2022) is widely adopted by major airlines like Emirates and All Nippon Airways, offering powerful analytics tools and aiding sustainability initiatives. FuelerLinx (FuelerLinx, 2017) optimizes flight operations and reduces costs for airlines such as Cutter Aviation and X8 Aviation. At the same time, SkyBreathe Analytics (OpenAirlines, 2020) leverages big data and AI to identify fuel-saving opportunities for carriers like Cebu Pacific and Fiji Airways. Moreover, Synthesis6 (i6 Group, 2022) enables airlines like Virgin Atlantic and KLM Airline to control fuel costs at granular levels. At the same time, Forge (Honeywell Aerospace, 2020) supports fuel-saving initiatives for over 100 airlines, leading to increased profitability and sustainability. Skymetrix (Skymetrix, 2021) provides comprehensive fuel and cost management for airlines such as Atlas Air, emphasizing efficiency and profitability. FuelPro (StorkJet, 2021) facilitates fuel efficiency initiatives for Volaris and Wizz Air, resulting in substantial savings and emissions reductions. Signal enhances fuel efficiency through personalized feedback for pilots and management, yielding significant savings and carbon impact reductions for airlines like Etihad Airways and Virgin Atlantic. Lastly, OptiFlight (Safety Line, 2022) offers in-flight guidance to optimize fuel efficiency for carriers such as Transavia and AirFrance, leading to substantial fuel savings and emissions reductions. These systems primarily focus on these fuel management systems' technical functionalities and performance metrics. While these aspects are crucial, a gap exists in exploring the UI design and its impact on user experience and operational performance. Since these systems rely heavily on pilot and ground staff interaction, an intuitive and user-friendly UI is essential. Efficient data visualization, clear communication of actionable insights, and user-friendly reporting tools can significantly enhance user adoption and ultimately lead to better fuel-saving outcomes.

3 Methodology

Usability evaluation measures the effectiveness, efficiency, and satisfaction of a product. In this research, the System Usability Scale (SUS) was used as the reliable tool for assessing the application's usability. SUS is an industry standard on the evaluation of product or services, including software, mobile devices, websites, and applications. This was supplemented with a usability script that details the flow and structure of the usability study. The selection of participants followed purposive sampling based on rank and tenure in the company. Because this project requires that data be collected from human participants, utmost considerations were taken to protect their privacy and interest. Before the usability testing, all participants were given a consent form that complies with the Data Privacy Act. Participation was voluntary, and they have been informed to have the opportunity to withdraw from the testing at any time. To protect their privacy, their data was anonymized, and are only visible to members of the research team. The data is stored on a flash drive protected by a password.

4 System Interface and Usability Results

4.1 Fuel Consumption Management System

The *User Account Management* interface allows Level 1 users to create, modify, or delete Level 2 or 3 FCMS user accounts. They can view the dashboard and print the report as well. Only the Flight Operations Analyst has access to this account. Meanwhile, Level 2 and 3 users can only see their accounts and can change their passwords. Level 2 users (head of Flight Operations Department) can view, query, and print the data and visualizations as a .pdf file. Lastly, the pilots can view and query the data from the dashboard. The *Main Dashboard* screen gives the user a view of actual fuel consumption in the airline's fleet. The user can choose the "operator", aircraft model, and the flight phase (take-off, cruising, and landing). This data is plotted versus the season (whether winter or summer), and the fuel consumption is measured in kilograms, as per industry standard. In addition to this, two pages namely, *Fuel Consumption Prediction* and *Fuel Savings Techniques* are included. The *Fuel Consumption Prediction* page displays the predictive page of the fuel consumption management system. It displays the aircraft type, the plane's model (Boeing or Airbus), the granularity (daily, weekly, or monthly), and the flight phase. It shows the actual fuel consumption and the predictive model, as shown in Fig. 1.

The *Fuel Savings Techniques* page provides insights into future fuel consumption, given the type of aircraft and the different flight phases (take-off, cruising, or landing), along with various strategies or techniques for fuel consumption. It compares the actual and predictive models with the mean, standard error, and mean absolute error for the predictive models.

4.2 User Interface Design of the Fuel Consumption Management System

This study used Nielsen's Usability Heuristics to guide designers and developers in user interface design (Nielsen, 1994), please refer to Fig. 3 for illustration. It consists of ten



Fig. 1. *Fuel Consumption Prediction* page allows users to view the fuel consumption at different aggregation levels, i.e. daily, weekly, and monthly levels for different aircraft models and flight phases.

usability principles that are followed to avoid design problems. The Fuel Consumption Monitoring System (FCMS) prioritizes a user-friendly interface to ensure efficient fuel management. Clear labels and page identification (visibility of system status) help users navigate the software easily. The terminology used throughout the system (match between the system and the world) is approved by users and aligns with industry standards, such as using “1000/kg” for fuel consumption. This consistency extends to the look and feel of all screens (consistency and standards), reducing cognitive load. The single-page interface with drop-down menus (user control and freedom) allows users to make changes quickly. To minimize errors, the system employs several strategies. Confirmation prompts are displayed before critical actions like password changes and user access modifications (error prevention). FCMS utilizes multiple dashboards (recognition rather than recall) to present data in different views: a default view, a comparison of actual vs. predicted data, and the impact of fuel-saving strategies. This reduces the need for users to constantly recall information. The mouse or trackpad is the primary interaction method (flexibility and efficiency of use) since it is a familiar choice that ensures broad user accessibility. The system also adheres to an aesthetic and minimalist design principle (aesthetic and minimalist design). This ensures users only manage one set of information at a time, further reducing cognitive load. Error messages are presented clearly in English using industry-specific terms (help users recognize and recover from errors). Finally, although FCMS is designed for ease of use, a user manual is provided to the airline for additional reference and support (help and documentation).

Following the usability test script provided to navigate and measure interface usability and effectiveness in the above screens, the following average scores resulted out of the SUS evaluation: Table 1

Looking at the averages of the raw SUS resulted in 29.22, and the final SUS was 73.06, which falls on the 70–79 percentile. These results indicate that the fuel consumption management system still needs improvement. The goal is to increase the final score to 96–100 percentile. It should be emphasized that the SUS is not diagnostic and is used to categorize the ease of use of the fuel consumption management system application. Nielsen’s ten usability heuristics can also strengthen the improvement of the application to see if it complies with these heuristics and if it does not.

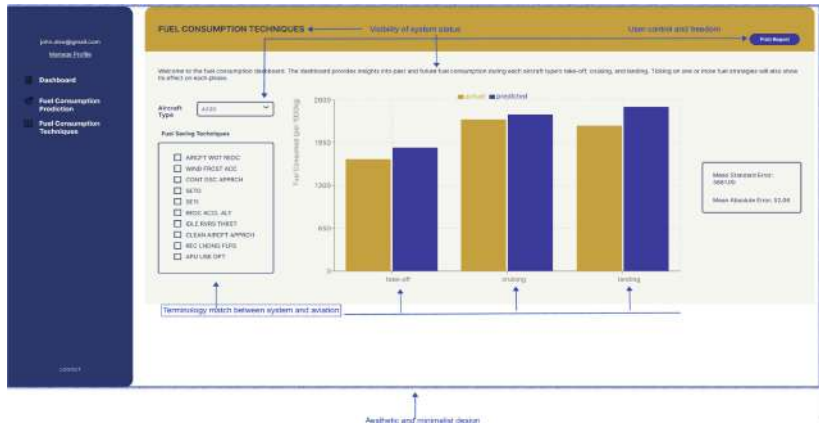


Fig. 3. UI design principles based on Nielsen’s Usability Heuristics, such as visibility of system status, user control and freedom, aesthetic and minimalist design, and terminology alignment with aviation standards, aimed at problem prevention.

Table. 1. SAS Results from Usability Testing.

Participant	Raw SUS	Final SUS	Percentile Range
1	34	85	96–100
2	28	70	35–40
3	30	75	70–79
4	30	75	70–79
5	24	60	0–14
6	28	70	35–40
7	25	62.50	0–14
8	37	92.50	96–100
9	27	67.50	35–40
Average	29.22	73.06	73.06

5 Conclusion

The effect of fuel costs on airline operations and sustainability has always been a challenge, and the COVID-19 pandemic certainly has not helped the airline industry’s financial situation. As the number of flights continue to normalize to pre-pandemic levels, digital technologies and data analytics have been employed by airline companies to make informed decisions using historical fuel consumption of its fleet. Most of these works focus on data preparation and modeling, ignoring the needs of pilots and flight operations personnel to easily engage the data.

To address this, Nielsen's heuristics was used to guide the design of the FCMS's user interface. The final user interface is simple, aesthetically pleasing, flexible, and uses the language of the airline industry. By creating three separate pages for each feature (View Dashboard, Fuel Consumption Prediction Model, and Fuel Savings Techniques), the user is fully aware of the status of the system. Error messages are immediately shown, and prompts provided in decisions like changing the password. There are various ways to continue the project. The placement of buttons and other controls may also be reviewed to lessen the user's confusion, and shortcuts and other gestures may also be added to provide additional interaction methods. More extensive testing can be performed to continue to improve the usability score of FCMS. Other performance measures can be collected, such as the number of clicks and the time it takes to complete a goal, and use an eye-tracking device to determine the areas of interest for a task, and creating a collection of common mistakes to help aid future users. These new metrics can then be analyzed and then used to improve FCMS.

References

- Boeing. 2022 Fuel Dashboard. <https://ww2.jepesen.com/wp-content/uploads/2019/05/fuel-dashboard-fact-sheet.pdf>. Accessed 24 Apr 2022
- General Electric. Increase Fuel Efficiency to **Lower** Costs and Reduce Carbon Emissions. Fuel Insight From GE Digital Aviation Software. https://www.ge.com/digital/sites/default/files/download_assets/fuel-insight-from-ge-digital.pdf. Accessed 25 Apr 2022 (2022)
- FuelerLinx. 2017 The Most Accurate Fuel Planning & Tankering Software in Business Aviation. Fuelerlinx - Aviation Fuel Management Software. <https://www.fuelerlinx.com/Landing2017/index.html#/top>. Accessed 25 Apr 2022
- OpenAirlines. 2020 Fuel Management Software – OpenAirlines. <https://www.openairlines.com/fuel-management-software/>. Accessed 25 Apr 2022
- i6 Group. 2022 Synthesis6 - Commercial Fuel Management. i6 Group. <https://i6.io/synthesis6>., Accessed 23 Apr 2022
- Honeywell Aerospace. 2020 Smarter with Honeywell Forge. Honeywell Aerospace. https://aerospace.honeywell.com/content/dam/aerobt/en/documents/learn/services/flight-services/brochures/N61-2527-000-00_Honeywell_Forge_Airlines_Bro.pdf., Accessed 24 Apr 2022
- Skymetrix. 2021 Atlas Air Automates its Airport and Navigation Charges. Skymetrix., <https://skymetrix.com/atlas-air-automates-its-airport-and-navigation-charges/>. Accessed 24 Apr 2020
- StorkJet. (2021). FuelPro. StorkJet. <https://storkjet.com/fuelpro/>. Accessed 24 Apr 2022
- Safety Line. (2022). Optiflight – Safety Line. Safety Line. <https://www.safety-line.fr/optiflight/#>., Accessed 24 Apr 2022
- Nielsen, J.: Enhancing the Explanatory Power of Usability Heuristics. In: Proceedings of the Human Factors in Computing Systems Conference, CHI '94, pp. 152–158. Association for Computing Machinery, New York, NY (1994)
- Brooke, J.: SUS: a retrospective. J. User Exp. **8**(2), 29–40 (2013)



Enhancing Hospitality Service Experiences: Exploring Context-Aware Systems Based on Smart Accommodation Framework

Yuming Wang^(✉), Chaemoon Yoo, Kuan Zhang, and Younghwan Pan

Kookmin University, Seoul 02707, South Korea
yumio99123@gmail.com

Abstract. Digital technologies enable the hospitality industry to transform the resident experience and service process. This study investigates the components of smart accommodation services and explores user preferences in conjunction with context-aware systems. Through a literature review and framework construction, we conducted a case study of existing hotel accommodation services. The comprehensive results facilitated the development of a standard for evaluating smart levels. Based on focus group interviews, five levels of smart accommodation were defined. It was discovered that the facilities of current hotel smart services are similar, and contextual interaction is considered a high-level smart because it can provide deeper personalized services and utilize richer contextual information to optimize the experience. This study highlights the role of context-aware systems in improving hotel service quality and analyzes the smart level standards. The findings emphasize the importance of comprehensive smart service systems in the hospitality industry, offering insights and reference base data for optimizing the resident experience and emerging trends in hotel services.

Keywords: Hospitality · Resident experience · Context-aware system · Smart accommodation · User Experience

1 Introduction

In the hotel industry, the proliferation of e-commerce and Internet of Things(IoT) systems has revolutionized the way users interact with each other and experience their accommodation. IoTs has emerged in recent years as a paradigm for communication between connected devices, allowing them to share data without human awareness to optimize system performance [1]. The context-aware system enriches the capabilities of intelligent devices complemented with smart applications, providing relevant contextual services to users within the IoT domain [2]. Existing hotel smart accommodation service processes are diverse and complex, with significant differences between different hotels, leading to inconsistent user experiences and varying levels of smart service effectiveness. Context-aware systems have the potential to significantly improve users' experiences and service quality in smart accommodation.

A smart level standard for the hotel industry can serve as a reference framework for practitioners, enabling them to better understand, measure, and optimize their services. Based on the theoretical foundation established by the literature review, we construct a framework and dare multiple case analyses. After collating the framework information, a four-person focus group interview was conducted to define the smart level content of the accommodation.

This study aims to investigate smart services in the hospitality industry, explore their relationship with context-aware systems, and discuss these applications through case studies to identify emerging trends and define smart levels for hotel accommodations.

2 Literature Review

2.1 Context-Aware System in Hospitality

As a concept of ubiquitous computing originally developed from computer science, context-aware systems are the ability to detect and respond to information about users and related IT devices, such as environment, situation, location and other surrounding information [3]. The definition is still evolving but the purpose remains the same: provide the computer with information about the user/device/environment and then adjust the system's response. Context-aware systems focus on delivering the right service to the right user at the right time.

Context-aware systems in hospitality refer to technological solutions that adapt to and respond to the specific needs, preferences, and circumstances of guests within a hospitality setting. For instance, within the realm of mobile computing, hotels are actively promoting contactless technology and real-time service in guest rooms. These systems leverage device sensors and machine learning algorithms to autonomously regulate various configurations by discerning user environments and activities [4].

2.2 Smart Accommodation

Liu Ping (2014) initially defined "smart accommodation" as one of the key facets within the framework of smart travel [5]. Building upon this foundation, Qin Zhicong (2022) further summarized the process of smart accommodation and defined it as six service stages of "Information-Booking-Access-Using-Payment-Sharing" [6]. It's noteworthy that user's goal of hotel patrons extends beyond mere travel, the concept of smart accommodation should break through this limitation to adapt to more scenarios. Moreover, recent research has increasingly emphasized the role of smart homes, which combine multiple forms of interaction to remotely control and monitor various home functions to enhance convenience, comfort, safety, and security [7].

Residences with service systems are not limited to homes, and research in commercial property has yet to be determined. This research will be based on the definition of smart homes and expand the research scope of smart accommodation to full-process service research in the hospitality industry. Therefore, this study defines smart accommodation as a residential space service within various ownership models, providing smart services to users through integrated systems such as the IoTs and context-aware systems.

3 Methodology

3.1 Multiple Case Analyses

Multiple case studies can choose different integration and analysis methods according to the research purpose and needs, as well as the connections and differences between the research objects. This approach enables the exploration of specific phenomena, the testing of theories, and the provision of a broader understanding. We hope to conduct framework research on hotel cases based on items related to smart services.

Mauro Migliardi and Marco Gaudina (2012) studied a social network-like mechanism facilitating user-moderated and controlled insertions into the system's ontology. Context is defined as any information characterizing the situation of an entity. They proposed a system framework of 4 W(What-Where-When-Who), which combines the intelligence of context-aware applications with human needs for contextual suggestions [8]. Their research hypothesized that a system that could remind planned activities at the right time and place could improve the quality of life in modern society. The scenario of smart accommodation services exists within hotels. This article primarily discusses the systematic service content of the processes. Therefore, our analytical framework will be modified to the 3W1H(What-When-Who-How), following is its definition:

What: Service components used by hotels.

When: Service stages of application.

Who: Beneficiaries of smart services and what goals have been achieved.

How: Technological Solutions.

Based on this framework, we selected international hotel groups that have recently been actively developing smart services to conduct multiple case analyses. The hotel groups' chain hotels are distributed in different regions, so we gave priority to hotel areas with relatively complete service systems. They are: Marriott(United States), Inter Continental Hotels Group(Greater China) and Henn na Hotel(Japan). The results are shown in Table 1.

We have analyzed publicly available hotel information, finding that high-end chain hotel groups are the primary adopters of smart services. These services are mainly concentrated in the Access/Deliver and Using stages. In the Access/Deliver stage, smart services simplify processes and enhance convenience through contactless check-in. The Using stage is where smart services are most intensively applied, incorporating situational awareness systems and IoT technologies, emphasizing service timeliness and responsiveness.

The overlap of services in the Access/Deliver, and Using stages underscores the importance of a comprehensive service system. The rise of online applications has further transformed the hotel service process, making the payment stage more flexible and the final service delivery stage more streamlined. Marriott and InterContinental Hotels have developed applications to provide mobile services and cooperate with cloud computing platforms to integrate data, while Henn na Hotel prefers to use the integration of independent smart devices to construct a fully automated robot hotel.

Table 1. Multiple case study [9–14]

Hotel	Smart Components	Service Stage	Beneficiaries and Goal		Technological Solutions
Marriott (United States)	Kiosks	Access/Deliver	User&Staff	Simplified Process	Marriott Bonvoy; ARTIK platform, SmartThings Cloud, Support by Samsung & Legrand; IBM Cloud Pak for Data.
	Mobile Key	Access/Using	User		
	Voice Control	Using	User	Real-time Service	
	Context-aware Regulation	Using	User		
	Hybrid Meeting	Using	User	Working	
	Delivery Robots	Using	Staff	Save Labor	
InterContinental Hotels (Greater China)	Digital Check-in&Check-out	Access/Deliver	User&Staff	Simplified Process	IHG Concerto; Baidu's DuerOS Platform; IHG One Rewards, Google Cloud.
	Mobile Key	Access/Using	User		
	Voice Control	Using	User	Real-time Service	
	Trip Planning	Using	User	Entertainment	
	Delivery Robots	Using	Staff	Save Labor	
Henn na Hotel (Japan)	Kiosks	Access/Deliver	User&Staff	Simplified Process	Kokoro Company Ltd.; Unibo; Chromecast; LG Styler.
	Humanoid Concierge	Using	Staff		
	Delivery Robots	Using	Staff	Save Labor	
	Remote Control	Using	User	Real-time Service	
	Voice Control	Using	User		
	Cleaning Management	Using	User&Staff	Simplified Process	
	Entertainment Devices	Access/Deliver	User&Staff	Entertainment	

3.2 Focus Group Interview

Smart accommodation does not mean fully intelligent or fully automated services. We hope to provide hotels with a more universal smart level framework. In order to refine and validate our smart level framework, based on the findings of the case analysis, we conducted Focus Group Interviews(FGI) with four experts and derived the following division of five smart levels. Details of the interviews are shown in Table 2.

In the standard definition, autonomous cars specify the levels of automation and the framework content to be studied and referenced. Based on this, we redefined the service process and used it to conduct the following open-ended questions to gain insights into the content of different smart levels. After discussion and sorting, the smart level was divided into five levels (Shown in Fig. 1).

The focus group findings indicate that although hotel service processes have been streamlined, all procedures remain in place except for adopting online payments. The

Table 2. FGI Research Summary

Interviewer Information	Interview Questions
P1, UX PhD student	-What do you think is the hotel service process? -What levels do you think the smart level of hotels should be divided into? -How should different levels of intelligence be defined?
P2, UX PhD student	
P3, UX master student	
P4, Smart home practitioners	

	Level 1	Level 2	Level 3	Level 4	Level 5
Name	Manual Operation	Batch Automation Operation	Data-Based Interaction	Dynamic Interaction	Context-Based Interaction
Service Process	Booking--Access--Using--Payment--Deliver		Booking--Access--Using-- ~~Payment~~ --Deliver		
Service Feature	Active Request		Real-time Feedback	Intelligent Response	
User Goal	Basic accommodation needs	Living experience	Incomplete personalized experience	Automated experience	Comprehensive smart experience

Fig. 1. Accommodation Smart Level Standard

paradigm shift of service features from passive acquisition to active provision. This evolution aligns with users’ goals from basic accommodation to seeking diverse and high-quality smart experiences.

4 Conclusion

This paper explores the importance of context-aware systems in smart accommodation, summarizes the current status of smart services in existing hotels through multiple case analyses, and uses focus group interviews to review the smart level of smart accommodation and provide suggestions for services.

In the future, higher-level smart accommodation services will include automated systems that can provide real-time feedback. The context-aware system uses Internet of Things technology to provide technical support for the hotel’s smart services. In terms of enhancing the hotel experience, a comprehensive service system is essential for ensuring spatial affordances, facilitating quick user adaptation, and improving satisfaction and feasibility. Our desk research also found that Building Information Modeling platforms are being utilized during the hotel construction stage. These platforms provide intelligent cloud services and 3D references throughout the entire project lifecycle [15]. The construction of smart accommodation must also consider the system integration stage,

which can reduce the cost and time of future modifications, thereby improving overall efficiency and coordination. In actual cases, even if service updates involve greater equipment overhead, we still hope that our framework can provide a reference for smart service guidelines for hotels at different price points to clarify the development direction of smart services.

References

1. Hallou, A., Fissaa, T., Hafiddi, H., Nassar, M.: Context-aware iot system development approach based on meta-modeling and reinforcement learning: A smart home case study. *Int. J. Online Biomed. Eng. (iJOE)*. **20**(06), 25–42 (2024)
2. Pradeep, P., Krishnamoorthy, S.: The MOM of context-aware systems: A survey. *Comput. Commun.* **137**, 44–69 (2019)
3. Zheng, P., Chen, C.-H., Wang, Z.: Graph-based context-Aware product-service family configuration. In: Zheng, P., Chen, C.-H., Wang, Z. (eds.) *Smart Product-Service Systems*, pp. 117–149. Elsevier (2021) ISBN 9780323852470
4. Gay, G.: *Context-Aware Mobile Computing: Affordances of Space, Social Awareness, and Social Influence*, 1st edn. Morgan & Claypool Publishers, United States (2009)
5. Liu, P., Pan, Y.H.: A Design of Smart Travel Based on City User Experience. In: Rau, P.L.P. (ed.) *Cross-Cultural Design. CCD 2014. Lecture Notes in Computer Science*, vol. 8528. Springer International Publishing, Cham (2014)
6. Zhicong, Q.: A Study on the Intelligent Tourism Based on the Six Elements of Tourism: Focused Hotel Reservation. PhD Dissertation. Kookmin University, Techno Design Graduate School, Seoul (2022)
7. Papagiannidis, S., Davlembayeva, D.: Bringing smart home technology to peer-to-peer accommodation: exploring the drivers of intention to stay in smart accommodation. *Inf. Syst. Front.* **24**, 1189–1208 (2022)
8. Migliardi, M., Gaudina, M.: The 4W (what-where-when-who) project goes social. In: *Proceedings of The Sixth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2012)*, pp. 1–8. IEEE, Palermo, Italy (2012)
9. <https://news.marriott.com/news/2021/04/05/less-hassle-more-to-explore-marriott-international-continues-to-deliver-on-consumer-desire-for-enhanced-contactless-technology>
10. <https://skift.com/2017/11/14/hilton-and-marriott-turn-to-the-internet-of-things-to-transform-the-hotel-room-experience/>
11. <https://www.ihgplc.com/en/news-and-media/news-releases/2018/intercontinental-hotels-and-resorts-launches-artificial-intelligence-rooms-in-greater-china>
12. <https://www.ihgplc.com/en/news-and-media/news-releases/2024/ihg-hotels-and-resorts-builds-a-new-travel-planner-powered-by-google-cloud-ai>
13. https://www.ihg.com/content/us/en/support/mobile?_PMID=99634986&cm_mmc=CJ_-4152654_-7304739&mktgvar=CJ_-4152654_-7304739_-www.ihg.com%2Fcontent_-0f2ddc0c1daa11ef8329008a0a18b8fb&glat=AFFI&setPMCCookies=true&cjevent=0f2ddc0c1daa11ef8329008a0a18b8fb
14. Reis, J., Melão, N., Salvadorinho, J., Soares, B., Rosete, A.: Service robots in the hospitality industry: The case of Henn-na hotel. *Jpn. Technol. Soc.* **63**, 101423 (2020)
15. <https://roboticulized.com/latest-news/2021/03/02/50506/marriott-international-pilots-ai-solutions-to-streamline-hotel-design-process-for-future-properties-in-asia-pacific/>



Examining the Population Stereotypes of Everyday Objects With Rotational Controls

Rachma Widaningtyas, Fitri Trapsilawati, and Ardiyanto Ardiyanto^(✉)

Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Yogyakarta,
Indonesia
ardiyanto@mail.ugm.ac.id

Abstract. This study investigated the population stereotypes of everyday objects with rotational controls. We examined two types of objects: 1) single-function rotational controls, such as door handles and lamp rotary switches, and 2) rotational controls for adjusting output magnitudes, such as volume knobs. The data collection involved 102 Indonesians participants fill out an online questionnaire that primarily queried their stereotypes regarding the operation of the investigated objects. The average percentage of the population stereotypes was about 80%, with majority being statistically significant at $p < 0.001$. The study revealed that everyday objects with single-function rotational controls were predominantly perceived as functioning clockwise, primarily for tasks such as opening doors and turning lights on. Similarly, clockwise rotation of controls for adjusting output magnitudes was associated with positive outcomes, such as higher brightness and volumes.

Keywords: population stereotypes · everyday objects · rotational controls · user experience

1 Introduction

User experience (UX) plays a pivotal role in shaping customer satisfaction with a product. A negative user experience resulting from interactions with products can have detrimental effects on customer retention, impacting the overall business of the product makers [1]. The user experience itself is determined by various factors, including usability, functionality, responsiveness, and population stereotypes. Population stereotypes refer to the tendency of the majority of people to have preconceived notions about control and display relationships before physically interacting with the products [2]. The population stereotypes are influenced by the visual stimuli of the physical attributes, colors, and textures of the products, which aid users in creating an intuitive experience [3]. Therefore, to ensure minimal errors and inefficiencies that lead to positive user experiences, product should have visual stimuli that allow users' correct intuitive interpretation [4].

Everyday objects are products commonly used by individuals in their daily lives to solve various problems. Some of these objects, such as volume controls and faucet levers, feature rotational controls designed to provide specific functions or adjust output magnitudes [5, 6]. To minimize negative user experiences and promote good well-being, it is crucial that these products possess certain properties that can shape accurate stereotypes. However, studies focusing on this aspect have been particularly limited in developing countries, including Indonesia.

This study aimed to investigate population stereotypes among Indonesians regarding everyday objects, specifically those with rotational controls. By understanding these stereotypes, we hope to assist professionals such as industrial designers, architects, and interior designers in designing or selecting everyday products with the appropriate control-output relations. This should lead to positive user experiences, which may result in better customer retention and favorable economic outcomes.

2 Methods

2.1 Participants

One hundred and two participants (53 males and 49 females), predominantly Indonesian college students, volunteered for this study. The average age of the participants was 21.59 ± 3.56 years. The study protocol was approved by the ethical committee of the authors' institution (No. KE/UGM/009/EC/2022).

2.2 Objects

The study investigated everyday objects with rotational controls, some of which were examined in Vu & Sun [5]. These rotational controls were categorized into two types: 1) single-function rotational controls and 2) rotational controls for adjusting output magnitudes. Some everyday objects with single-function rotational controls examined in the study included bathroom locks, door handles, and lamp switch. Meanwhile, some of the everyday objects with rotational controls for adjusting output magnitudes included amplifier rotary knobs and shower knobs. For a visual representation of these objects, please refer to Tables 2 and 3.

2.3 Questionnaires

The data collection for the study was conducted by requesting participants to complete a questionnaire consisting of three parts. The first part comprised an informed consent form. Upon agreeing to participate, they were then asked to provide demographic information. The last part of the questionnaire involved inquiries aimed at determining their stereotypes regarding the operation of the investigated objects. The inquiries generally asked, "What would you do to perform a particular task using the products?" [5]. Participants were also provided with illustrations of the objects as stimuli. The inquiries had three response choices. Typically, the first two choices represented two possible ways of operating the products for particular tasks, such as turning the knobs clockwise and

Table 1. Population stereotype of investigated objects with single-function rotational controls

Stimulus (alphabetically ordered)	Illustrations	Stereotype description	% of responses	p-value
Bathroom stall locks		Turn the locks counter-clockwise to open the doors.	78%	<0.001 **
Cross handle faucets		Turn the handle clockwise or toward the wall to shut off the water.	60%	<0.05*
Forward-facing cross-handle faucets		Turn the handle of faucet 1 clockwise while simultaneously turning the handle of faucet 2 counterclockwise to activate the hot water.	56%	0.488
Horizontal lever-operated faucets		Turn the levers toward us to turn the water on.	83%	<0.001 **
Horizontal door handles		Turn the handle clockwise or downward to open the door.	99%	<0.001 **
Keyholes		Turn the keys clockwise to unlock the door.	56%	0.315
Lamp rotary switches		Turn the switches clockwise to turn the lights on.	99%	<0.001 **

Notes: *) Significant at $p < 0.05$; **) significant at $p < 0.001$

counterclockwise to increase the volume. The last choice of responses was the option to “decline to answer”. Quality control questions were included to ensure participants provided their best answers. The questionnaire was administered in Bahasa Indonesia through an online survey platform (Google Form).

2.4 Data analysis

The frequency of responses for each object was calculated relative to the total number of participants. The population stereotype was established based on the highest percentage of responses observed. Additionally, to assess the statistical significance of the stereotypes, a binomial test was conducted [5].

3 Results

Tables 2 and 3 present the population stereotypes associated with the objects investigated. As shown in the table, more than half of the participants associated the clockwise rotation of several products with single-function rotational controls, such as bathroom locks, door handles, and keyholes, with the action of opening. However, no significant proportion was observed regarding the stereotypes for keyholes ($p = 0.315$). Similarly, almost all participants (99%) associated the clockwise rotation of lamp rotary switches with turning

Table 2. Population stereotype of investigated objects with rotational controls for adjusting output magnitudes

Stimulus (alphabetically ordered)	Illustrations	Stereotype description	% of responses	p-value
Amplifier rotary knobs		Turn the knobs clockwise to increase the volume.	98%	<0.001 **
Camera lens rings		Rotate the lens rings clockwise to zoom the lens.	83%	<0.001 **
DVD player rotary knobs		Turn the knobs counterclockwise to decrease the volume.	94%	<0.001 **
Guitar rotary knobs		Turn the knobs clockwise to increase the volume.	92%	<0.001 **
Projector lens zoom rings		Turn the rings clockwise towards "A" to zoom the lens.	60 %	0.061
Shower knobs		Turn the shower knobs counterclockwise to decrease the water temperature.	81%	<0.001 **
Tally counter reset knobs		Turn the knob counterclockwise to decrease the numbers.	86 %	<0.001 **
Watch crowns (time adjusting knobs)		Turn the knob clockwise to increase the time.	94%	<0.001 **

Notes: *) Significant at $p < 0.05$; **) significant at $p < 0.001$

the lights on ($p < 0.001$). However, rotational controls for water faucets did not entirely conform to the previously observed positive associations with clockwise rotation seen in other products, as more than 50% of participants perceived that rotating the controls clockwise (toward the wall) would shut off the water.

The objects investigated that featured rotational controls for adjusting output magnitudes, such as lamp brightness, volume knobs, and standing desk height, were perceived by more than 75% of participants to increase these magnitudes through clockwise rotation ($p < 0.001$). Conversely, the majority of participants (81%) believed that functions such as lowering water temperature using shower knobs could be achieved by rotating the knob counterclockwise. However, these findings did not extend to projector lens zoom rings, where more participants associated counterclockwise rotation with decreasing magnitudes. Nonetheless, statistical tests did not indicate a significant proportion for these stereotypes ($p = 0.061$).

4 Discussion and Conclusions

This study found that everyday objects with single-function rotational controls were predominantly perceived as functioning clockwise, primarily for tasks such as opening doors and turning lights on. Similarly, clockwise rotation of controls for adjusting output magnitudes was associated with positive outcomes. However, although more than 50% of participants indicated this pattern, some everyday objects with rotational controls in this study did not exhibit significant population stereotypes.

This study indicated that participants typically stereotype the act of increasing the magnitude of specific functions in products with rotating knobs by turning them clockwise, a finding that aligns with similar observations reported in Vu & Sun [5]. This stereotype could be attributed to the participants' familiarity with clockwise rotation, commonly associated with the increase of time on analog clocks or watches.

The average percentage of responses for each object in the study was approximately 82%, which was higher than the average reported in Vu & Sun [5]. This higher percentage could be attributed to the uniformity of the participants in our study, the majority of whom were post-secondary students in Indonesia. In contrast, the population in the other study included non-students, potentially leading to less homogeneous results due to variability in participant characteristics [7].

Although population stereotypes can be identified through a high proportion of responses with significant statistical results, several stereotypes of the objects investigated in this study, such as projector lens zoom rings, could not be conclusively determined. This lack of conclusive data could be attributed to the participants' limited familiarity with the products. Therefore, to ensure a positive user experience, additional interventions need to be adopted, including the addition of signifiers that promote correct mental models to minimize errors and inefficiencies [8].

Although this study presents some important findings, it also has several limitations. This study only included participants aged 18 to 30 years old. Consequently, familiarity with such control mechanisms might be more pronounced compared to older generations [9]. Additionally, the considerable ethnic diversity across Indonesia was not comprehensively accounted in this study [10]. As a result, the generalizability of the study's findings was constrained by these limitations.

Acknowledgement. The authors express their gratitude for the funding support received from the Department of Mechanical and Industrial Engineering at Universitas Gadjah Mada, which enabled the conduct of this research. Additionally, the authors are thankful to the Directorate of Research at Universitas Gadjah Mada for the travel grants that allowed them to attend this conference.

References

1. Yoon, J., Kim, C., Kang, R.: Positive user experience over product usage life cycle and the influence of demographic factors. *Int. J. Des.* **14**(2), 85–102 (2020)
2. Chan, A.H.S., Hoffmann, E.R.: The psychological cost of making control responses in the nonstereotype direction. *Hum. Factors.* **58**, 1173–1186 (2016)
3. Hurtienne, J., Stöbel, C., Weber, K.: Sad is heavy and happy is light: Population stereotypes of tangible object attributes. In: *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pp. 61–68. Association for Computing Machinery, New York, NY, USA (2009)
4. Kee, D.: Stereotype for control-burner relationship of four-burner ranges for Koreans. *Int. J. Ind. Ergon.* **44**, 343–348 (2014)
5. Vu, K.-P.L., Sun, Y.: Population stereotypes for objects and representations: response tendencies for interacting with everyday objects and interfaces. *Hum. Factors.* **61**, 953–975 (2019)
6. Bergum, B.O., Bergum, J.E.: Population stereotypes: An attempt to measure and define. In: *Proceedings of the Human Factors Society Annual Meeting*, pp. 662–665. Sage, Los Angeles, CA (1981)
7. Peterson, R.A.: On the use of college students in social science research: Insights from a second-order meta-analysis. *J. Consum. Res.* **28**, 450–461 (2001)
8. Norman, D.: *The Design of Everyday Things*. Doubleday, New York (1988)
9. Stratton, S.J.: Population research: Convenience sampling strategies. *Prehosp. Disaster Med.* **36**, 373–374 (2021)
10. Ismianti, I., Herianto, H., Ardiyanto, A.: Studi Antropometri Mahasiswa Indonesia Bersuku Batak Dan Jawa. *J. Ergon. Indones. (Indones. J. Ergon.)* **5**, 47–56 (2019)



Applied Research on Gamification-Based Intelligent Interaction Design of Smart Traffic Digital Twin Model

Tianlun Yang^{1,2(✉)}, Zuyao Wang², Georgios Kapogiannis³, Byung-Gyoo Kang⁴,
Hu Zhenyu¹, Wu Yanhui¹, Shengyuan Liu¹, and Yuchang Zhang¹

¹ Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China
tianlun.yang@nbt.edu.cn

² NingboTech University, Ningbo, China

³ AFG with The University of Aberdeen (Qatar Campus), The University of Aberdeen,
Aberdeen, UK

⁴ University of Nottingham Ningbo China, Ningbo, China

Abstract. BIM-based digital twin models are pivotal for the AEC industry, providing essential data for optimizing design, construction, and operations, and enhancing production quality and efficiency. Despite the ISO 19650 framework, the complexity of these models poses challenges for current visualization and interaction technologies, affecting information acquisition efficiency. This study highlights the importance of interaction design in facilitating project coordination and explores intelligent interaction design, which integrates human-computer interaction and new media art theories to bridge the real and digital worlds. Gamification, as a subset of intelligent interaction design, is noted for its potential to increase user engagement and project quality through instant feedback. A prototype for smart traffic digital twin interaction, developed based on these theories and technologies, is presented. Initial discussions on its impacts are provided, with future research planned for deeper analysis and validation.

Keywords: Digital Twin · Intelligent Interaction · Gamification · Information Visualization · Project Management

1 Research Background and Research Question

The development of information management has traversed three distinct stages: the era of paper-based printing media, the computer-based digital era, and the internet-based information era. Throughout these stages, the sophistication of information processing methods has escalated, and the velocity and efficiency of information dissemination have been markedly enhanced. Currently, information management is on the cusp of entering the intelligent era. In this new epoch, artificial intelligence, grounded in digital media and the internet, enables computers to autonomously analyze data, predict human behavior, and anticipate societal needs, thereby augmenting the efficiency of data processing and delivery through the synergy of smart terminals and interactive technologies. Building

Information Modelling (BIM) has been introduced and refined to bolster project information management within the Architecture, Engineering, and Construction (AEC) sectors [1]. Concurrently, Lean management principles have been developed to elevate project quality [2]. The amalgamation of BIM and Lean Management, coupled with advancements in computer software and hardware, allows for the visualization of project data in the AEC industries, thereby enhancing project quality and efficiency.

Michael Grieves first introduced the concept of the digital twin in 2002, exploring its potential for developing and managing complex systems through digital twin simulations [3]. Studies have shown that digital twin models necessitate support from diverse data sources and information to perform the required simulations during project development [4]. Consequently, the Standards, Methods, and Procedures (SMPs) of Building Information Modelling (BIM) offer AEC enterprises a strategic framework for constructing digital twin models, given that BIM encompasses both graphical and non-graphical information to support a multitude of objectives. Despite digital twins offering robust data support for AEC industries with current technologies and methodologies, there persist significant gaps in human-digital model interaction. The ISO 19650 standardizes the collection, processing, and delivery of information for AEC projects [5, 6]; however, the intricacy of digital twin models poses challenges for end-users in manipulation. As a result, the efficiency of extracting necessary information from digital twin models remains suboptimal. Therefore, the aim of this research is to address this gap by enhancing the interaction between humans and the virtual environment.

2 Intelligent Interaction Design and Gamification Theory

In addressing the research gap, this paper explores the theory of intelligent interaction design to identify an effective approach to human-computer interaction that could enhance the efficiency of information acquisition within virtual environments. The paper specifically examines the characteristics of new media art and gamification to provide theoretical underpinnings for the prototype of the proposed smart traffic digital twin model interaction framework.

2.1 New Media Art

New media art emerged at the end of the twentieth century as a novel form of artistic expression. It builds upon traditional art by incorporating modern technologies, including computers, the internet, and smart terminals, among others. New media art offers a new platform for information delivery through the synergy of aesthetics and programming [7]. Over the subsequent decades, through experimentation and iteration, artists have explored how electronic media and digital technology can be applied in artistic creation to innovate the delivery of information [8]. Under the influence of digital technologies, various genres such as virtual reality art, internet art, and interactive art have emerged. Since the twenty-first century, with the advancement of intelligent technologies, new media art has transformed the way people acquire information and perceive the world [9].

New media art significantly impacts information delivery; it utilizes digital technologies and tools to create diverse art forms that present content, thereby overcoming the limitations of traditional methods in data expression [10]. In this way, artists can enable audiences to experience interaction with artworks, allowing them to engage with art in a more intimate manner and gain a deeper understanding of the underlying emotions. Consequently, new media art balances the relationship between art and technology and establishes a connection between humans and society [11]. Furthermore, new media art enhances the quality and efficiency of information delivery and motivates people to consider social development and environmental protection [12]. It spearheads innovation and development in the field of interaction design, showcasing not only the possibilities in artistic expression but also enhancing user experiences with artworks. As an interdisciplinary art form, new media art integrates art creation, digital technologies, and user interfaces to improve the quality of information delivery [13]. By applying digital tools in art creation, the limitations of traditional media in mono-dimensional expression can be overcome, enabling users to explore artworks more thoroughly and have a richer experience in acquiring information.

2.2 Gamification and Interactive Environment

The most important characteristic of gamification is its mechanism of real-time feedback, which can provide users with immediate feedback based on their inputs. This significantly reduces waiting time and enhances the user experience [14]. Gamification theory is an interdisciplinary field, encompassing interactive design, psychology, education, sociology, and more. It primarily investigates how to apply game elements in non-game environments to encourage users to change their participation and social behavior [15]. Early gamification theories, formulated since the 1980s, revealed that gamification can influence human behavior in learning activities [14]. Consequently, motivation and cognition in human information processing have been proposed to demonstrate how gamificative environments can impact interactivity in digital models [16]. With the advancement of information technology, gamification has been increasingly adopted in business environments to improve the efficiency of human-information interaction [17]. Therefore, gamification can be integrated with new media art to construct a new framework aimed at enhancing information acquisition performance in the field of interactive design.

Gamificative environments can enable abstract content to be presented directly through the implementation of visualization and interaction technologies, allowing users to improve their understanding of relevant information during project development [16]. As a result, the project digital model can be utilized more easily by end-users, and information can be acquired more efficiently by project participants. Given that many people enjoy playing games, the deployment of gamificative technologies in non-game environments can stimulate interest and enhance user participation. Building on traditional interactive design, intelligent interaction design incorporates additional information technologies, enhancing user experience by allowing terminal equipment to automatically provide end-users with relevant outputs based on embedded programs. Consequently, the application of intelligent interaction design in gamificative environments can improve the efficiency of information collection, processing, and delivery.

2.3 Integration of New Media Art, Gamification and Intelligent Interaction

New media art introduces new media and innovative methods of information delivery to the field of interaction design, while gamification provides a motivational framework to enhance user participation. Intelligent interaction technologies offer smart solutions for processing information within gamified environments. Based on the findings from literature reviews, this research integrates new media art, gamification, and intelligent interaction technologies to develop a prototype that supports Human-Computer Interaction (HCI) in smart traffic digital twin models. This prototype addresses the current limitations, particularly the low information acquisition efficiencies in digital twin models, offering a novel solution for information management in the AEC industry.

3 Research Design

The research question of this study is how to enhance the efficiency of information acquisition from digital twin models. The research objectives are threefold: A) to identify the problems and issues in current digital twin model interaction solutions; B) to examine the features of new media art, gamification, and intelligent interaction technologies; C) to assess how these features can impact information acquisition in smart traffic digital twin models. Objective A is identified through background study; for Objective B, this research employs a literature review to explore how these features can be integrated to influence digital twin model interaction; for Objective C, this research employs design intervention to construct a prototype of a smart traffic digital twin model interaction framework. Based on the findings from the above objectives, this research formulates hypotheses for future validation.

This research leverages the Unity3D game engine as a platform to integrate graphical and non-graphical information, organizes the structure of model content, and ultimately develops the prototype of an interaction framework through the adoption of intelligent interaction technologies. The C# programming language is used to develop the relevant interactive functions, and the .NET framework is employed to deploy the prototype across different operating systems. In this paper, the prototype is tested exclusively on Windows.

4 Structure of Digital Twin Model Information and Design of Intelligent Interaction Framework

In constructing the framework for a smart traffic digital twin interaction system, this research employs a design intervention method. It selects a road project as a use case scenario to facilitate discussion. The research analyzes the information structure within the road digital twin model, organizes the information based on its priority, and then designs an intelligent interaction framework based on these findings.

4.1 Project Overview

This project, located in the Hangzhou Bay New District of Ningbo, China, aims to establish a smart traffic zone for autonomous driving tests. Along this road, numerous pieces of equipment will be installed to collect and process data that support autonomous driving. This data will be synchronized with the digital twin model to enhance smart traffic management.

4.2 Digital Twin Model Information Structure and Priority

This project encompasses several key information categories: vehicle information, road information, weather information, and urban information. Table 1 below delineates the specific details within each category. Vehicle information comprises Make, Model, Color, Speed, and Lane; road information encompasses traffic lights, guide signs, cameras, sensors, motorways, non-motorways, and green belts; weather information includes visibility, weather conditions, and time; urban information pertains to traffic management.

Table 1. Static and Dynamic Information in Smart Traffic Digital Twin Model

Static Information	Dynamic Information			
Road Information	Road Information	Vehicle Information	Weather Information	Urban Information
Traffic Light Position	Status of Traffic Light	Make	Visibility	Traffic Control
Road Sign Position	Status of Camera	Model	Weather Condition	
Guide Sign Position	Status of Sensor	Colour	Time	
Camera Position		Speed		
Sensor Position		Lane		
Motorway		Vehicle Position		
Non-Motorway				
Green Belt				

These categories are further bifurcated into static and dynamic information, as detailed in (Table 1). Static information is relatively invariant and does not typically change, while dynamic information is subject to constant flux. In the realm of smart traffic digital twin models, risk assessment and prediction of vehicle behavior are of paramount importance. Weather information is pivotal, as daylight, weather conditions, and visibility significantly influence driving status. Moreover, urban information exerts an influence on vehicle behavior, given that traffic control mechanisms can dictate lane

and road usage. Thus, weather and urban information are two critical factors in the traffic simulation process.

The digital twin model’s development is anchored in Building Information Modeling (BIM), which includes both graphical and non-graphical information. This research categorizes the information within the smart traffic digital twin model, as depicted in Table 2. Graphical information includes road and vehicle information, which can be directly visualized within the model. Non-graphical information, which is not directly observable, comprises weather data, vehicle data, and sensor data. BIM’s role in Architecture, Engineering, and Construction (AEC) projects is to render non-graphical information visible, thereby enhancing project comprehension and bolstering the quality and efficiency of decision-making. This research employs visualization technologies to simulate non-graphical information. Table 3 below presents the information being visualized in this project.

Table 2. Graphical and Non-Graphical Information in Smart Traffic Digital Twin Model

Graphical Information		Non-Graphical Information			
Road Information	Vehicle Information	Road Information	Vehicle Information	Weather Information	Urban Information
Traffic Light Position	Color	Traffic Light Status	Speed	Visibility	Traffic Control
Road Sign Position	Make	Camera Status	Lane	Weather Condition	
Guide Sign Position	Model	Sensor Status	Vehicle Position	Time	
Camera Position					
Sensor Position					
Motorway					
Non-Motorway					
Green Belt					

Table 3. Visualization of Non-Graphical Information in Smart Traffic Digital Twin Model

Visualization of Non-Graphical Information			
Road Information	Vehicle Information	Weather Information	Urban Information
Traffic Light Status	Lane	Visibility	Traffic Control
	Vehicle Position	Weather Condition	
		Time	

Non-graphical information includes details on road, vehicle, weather, and urban conditions. For road information, visualization technologies can illustrate the status of traffic lights within the model, offering a visual reference for understanding road management conditions. For vehicle information, the position of vehicles can be identified and displayed in the digital twin model using GPS and GIS data, enabling end-users to ascertain traffic conditions through visualized information. Weather information can be simulated and presented in the digital twin model using visualization technologies, aiding in decision-making regarding the necessity of traffic control measures based on visibility and daylight conditions. For urban information, end-users can label roads or lane control data directly within the digital twin model for immediate presentation and reference. Consequently, visualization technologies enable the efficient presentation and communication of non-graphical information within the digital twin model.

Based on the information requirements of the smart traffic digital twin model, this research proposes a priority ranking for these information categories, as outlined in Table 4. Road information is given the highest priority, as all other information is predicated upon it. Weather information is accorded the next highest priority, as it significantly impacts road usage and, by extension, project decision-making. Considering the high variability of daily weather conditions, integrating weather information into the model can assist project managers in conducting risk assessments under various weather scenarios. Urban information impacts vehicle behavior due to regional maintenance, expansion, and traffic management, which in turn affects driving. Vehicle information is assigned the lowest priority because, despite the large number of automobiles, there is relatively little functional variation among them.

Table 4. Priority of Information in Smart Traffic Digital Twin Model

	High Priority		Low Priority	
	Road Information	Weather Information	Urban Information	Vehicle Information
High Priority	Motorway	Visibility	Traffic Management	Vehicle Position
	Non-Motorway	Weather Condition		Lane
	Traffic Light Position	Time		Speed
	Traffic Light Status			Color
	Road Sign Position			Make
	Guide Sign Position			Model
	Camera Position			
	Camera Status			
	Sensor Position			
	Sensor Status			
Low Priority	Green Belt			

4.3 Design of Intelligent Interaction Framework

Based on the findings from literature reviews and Section 4.2, this research presents the preliminary framework of the interaction system within the smart traffic digital twin model in Table 5.

Table 5. Preliminary Framework of Smart Traffic Digital Twin Model Interaction System

Gamificative Environment (Build a Game-like Environment to Place Road Digital Twin Model)			
	Structured Information (Based on the Priority of Road Information)		
		Design of Interaction System	
			Design of Intelligent Interaction System

The proposed intelligent interaction system framework utilizes a gamificative environment as a platform to integrate both graphical and non-graphical information from the smart traffic digital twin model. It employs a physics engine in Unity3D to simulate and visualize the non-graphical information. By leveraging the self-exploration and instant feedback mechanisms inherent in gamification, the proposed system can motivate individuals, thereby increasing project participation rates. The system organizes the information of the digital twin model according to priority and employs intelligent technologies to assist users in acquiring the necessary information more efficiently.

In this research, the design of intelligent interaction is pivotal to the digital twin model’s information processing, as depicted in Table 6. The core of interaction design involves using a user interface (UI) to extract project information from the model, encompassing both information querying and feedback methods. The intelligent design’s cornerstone is the utilization of a game physics engine to simulate vehicle driving and visualize non-graphical information. By integrating the UI with the game physics engine, the system enables end users to control vehicles from either a first or third-person perspective and view project information in a visualized format, thereby enhancing project understanding.

Table 6. Elements of Smart Traffic Digital Twin Model Intelligent Interaction System

Design of Interaction System	UI Design	Input			
		Output	Intelligent Design	Drive Simulation	
				Visualization of Non-Graphical Information	Weather Condition
					Time/Light

This research has developed intelligent interaction functions for the smart traffic digital twin model, informed by the findings on information structure and acquisition. Fig. 1 illustrates the visualization process of traffic light status. In the virtual environment, lights of different colors can be displayed based on commands input from the physical world's system.



Fig. 1. Visualization of Traffic Light Information

Fig. 2 presents the visualization process of vehicle information. End-users can access vehicle information (e.g., make, model, color) from a panel adjacent to the virtual vehicle within the smart traffic digital twin model, and the developed interaction system facilitates switching between different vehicles for project participants.



Fig. 2. Visualization of Vehicle Information

The research concentrates on the development of intelligent interaction for vehicle driving simulation. Physics engines in Unity3D are utilized to visualize non-graphical information such as weather and lighting, as shown in Fig. 3, and end-users can obtain this information from the digital twin model through a driver's first-person view perspective. The interactive system's innovation lies in its fusion of new media art features and gamification elements, which can pique interest, and its intelligent approach to visualizing abstract information, thereby aiding in the enhancement of project information comprehension.

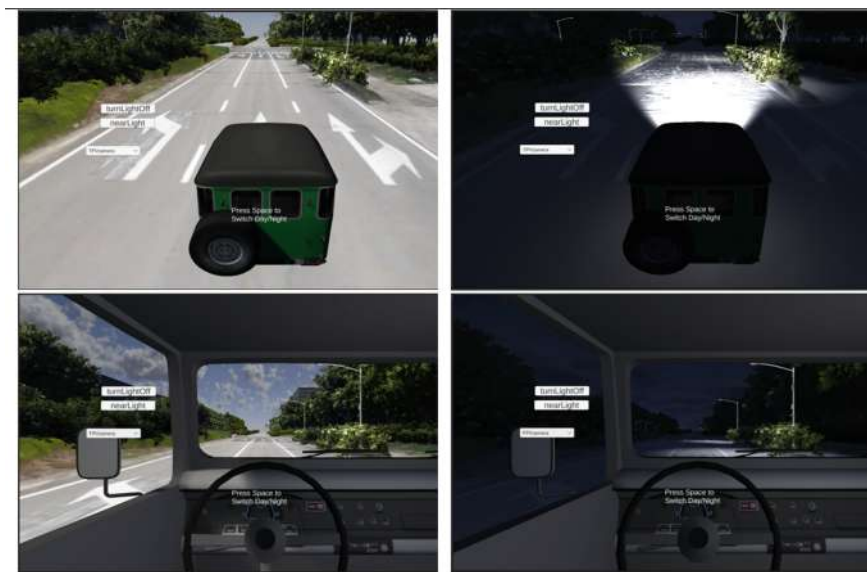


Fig. 3. Intelligent Design (Virtualization of Weather Information and Drive Simulation)

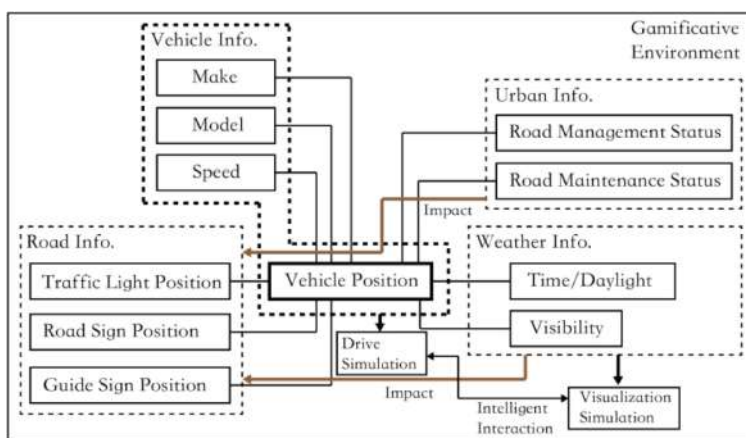


Fig. 4. Conceptual Framework of Digital Twin Intelligent Interaction System

Drawing on the findings from literature review and design intervention, this research proposes a conceptual framework for an intelligent interaction system for smart traffic digital twin models, as displayed in Fig. 4. In the proposed conceptual framework, vehicles are central to the intelligent interaction system, with all other information interacting based on the first-person driving view of the vehicle. Within the intelligent interactive system, weather information and urban information directly affect road information, and through the feedback of road information, end-users can engage in relevant interactions

and acquire necessary information for project assessment and decision-making. As visualization technologies render non-graphical information, project participants can enjoy an enriched experience while interacting with the digital twin model under a gamificative environment, thus potentially increasing the quality and efficiency of smart traffic project management.

5 Impacts from Intelligent Interaction Framework to Digital Twin Model Information Acquisition

The essence of new media art lies in its departure from traditional modes, employing innovative media to convey information and utilizing more engaging methods to facilitate information delivery. Consequently, this research adopts a gamificative environment to develop an interactive system aimed at stimulating user participation in smart traffic digital twin models. By organizing structured information based on their priority and presenting this information from a first-person perspective, users can more directly access the required project information.

The crux of intelligent interaction in this study is the visualization of non-graphical information within the digital twin model. The developed system can simulate and visually present weather, daylight, and visibility conditions through programmed interactive functions, thereby allowing users to conduct comprehensive explorations within the digital twin model. This enhances the quality and efficiency of project information management. The role of smart traffic is to establish a platform by integrating relevant software and hardware for the development of autonomous driving infrastructure. Therefore, in smart traffic projects, there is a need to simulate driving to assess the rationality of the design. Consequently, the prototype developed in this research can support decision-making in smart traffic projects.

The smart traffic digital twin intelligent interaction framework developed in this research enables users to explore virtual content through a first-person car driving view within a gamificative environment. Supported by situated cognition theory, the proposed framework enables users to better understand project information, thereby increasing the efficiency and quality of decision-making. This research formulates three hypotheses based on the findings: 1) A first-person view of interaction in the digital twin model can enhance project understanding; 2) Structuring information based on priority can enhance information interaction efficiency; 3) Visualizing non-graphical information can increase the quality and efficiency of project decision-making.

6 Conclusion, Research Limitation and Future Researches

This paper constructs a prototype of a gamification-based smart traffic digital twin model intelligent interaction system, based on the findings from literature reviews and design interventions, and formulates three hypotheses based on this prototype. A limitation of this research is that these hypotheses have not been empirically validated; therefore, future research will be tasked with assessing and validating these hypotheses. Additionally, another limitation is that the project size and the amount of contained information are

relatively small, which constrains the development of system functions. Consequently, future research will focus on exploring intelligent interactions in a variety of digital twin projects, including those of larger scales.

References:

1. Eastman, C., Teicholz, P., Sacks, R., Liston, K.: *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*. John Wiley & Sons (2011)
2. Koskela, L., Ferrantelli, A., Niiranen, J., Pikas, E., Dave, B.: Epistemological explanation of lean construction. *J. Constr. Eng. Manag.* **145**(2), 04018131 (2019). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001597](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001597)
3. Grieves, M.: Intelligent digital twins and the development and management of complex systems [version 1; peer review: 4 approved]. *Digit. Twin.* **1**(1), 8 (2022). <https://doi.org/10.12688/digitaltwin.17574.1>
4. Yoon, S.: Building digital twinning: Data, information, and models. *J. Build. Eng.* **76**, 107021 (2023). <https://doi.org/10.1016/j.jobee.2023.107021>
5. ISO: BS EN ISO 19650-1: Organization and digitization of information about buildings and civil engineering works, including building information modelling - information management using building information modelling: Concepts and principles. International Organization for Standardization (2018)
6. ISO: BS EN ISO 19650-2: Organization and digitization of information about buildings and civil engineering works, including building information modelling - information management using building information modelling: Delivery phase of the assets. International Organization for Standardization (2018)
7. Grau, O.: New Media Art. obo in Art History (2016). <https://doi.org/10.1093/obo/9780199920105-0082>
8. Gilbert, T.: Looking at digital art: Towards a visual methodology for digital sociology. *Am. Sociol.* **49**(4), 569–579 (2018). <https://doi.org/10.1007/s12108-018-9384-2>
9. Lopreiato, P.: New media art: The liminal space between thinking and perceiving. In: Rezaei, N., Saghezadeh, A. (eds.) *Thinking: Bioengineering of Science and Art*, pp. 397–412. Springer International Publishing, Cham (2022)
10. Gong, Y.: Design methods of visual communication in new media art design. In: Xu, Z., Parizi, R.M., Loyola-González, O., Zhang, X. (eds.) *Cyber Security Intelligence and Analytics, CSIA 2021. Advances in Intelligent Systems and Computing*, vol. 1342. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-70042-3_27
11. Gor, A.: Reimagining the iconic in new media art: Mobile digital screens and chôra as interactive space. *Theory Cult. Soc.* **36**(7–8), 109–133 (2019). <https://doi.org/10.1177/0263276419836481>
12. Xu, X.: The theoretical construction of new media art design and cultural thinking. *Int. Acad. J. Humanit. Soc. Sci.* **1**(2), 56028 (2023)
13. Lughi, G.: Digital media and contemporary art. *Mimesis J.* **3**, 43–52 (2014). <https://doi.org/10.4000/mimesis.686>
14. Gee, J.: What video games have to teach us about learning and literacy. *Educ. Train.* **46**(4) (2004). <https://doi.org/10.1108/et.2004.00446dae.002>
15. Deterding, S.: Gamification in management: Between choice architecture and humanistic design. *J. Manag. Inq.* **28**(2), 131–136 (2019). <https://doi.org/10.1177/1056492618790912>
16. Deterding, S., Khaled, R., Nacke, L.E., Dixon, D.: Gamification: toward a definition. In: *Paper Presented at the CHI 2011 Gamification Workshop Proceedings*, Vancouver, BC, Canada (2011)

17. Dale, S.: Gamification: Making work fun, or making fun of work? *Bus. inf. rev.* **31**(2), 82–90 (2014). <https://doi.org/10.1177/0266382114538350>

Visual (I)



Visuo-Cognitive Cartography, A New Tool for the Analysis of Eye-Tracking Data

Denis I. Chavez^{1,2}(✉) 

¹ Pontificia Universidad Católica del Ecuador, Quito 70525, Ecuador
dichavez@puce.edu.ec

² Human and Artificial Cognitions Laboratory (CHArt, RNSR 200515259U), EPHE – PSL,
93300 Aubervilliers, France

Abstract. This paper proposes visuo-cognitive cartography (CVK) as a new method for the study of human visual interaction. The method discusses a way of characterizing visual strategies through the construction of a complex system of n-dimensional vector spaces that allows a multi-scale analysis (micro, meso and macro) of the instrumentalization processes in different visual interaction situations. To test this method, we used the raw eye movement data of 32 subjects collected during a reading task in a laboratory experience. In the present paper we present the logic of the data pre-processing pipeline for the construction of this cartography of visual and related cognitive strategies as also some possibilities in terms of analysis. The macroanalysis evidences the applicability for cognitive ergonomics and the studies of artifactual interaction with texts. The main findings highlight the potential of CVK and its metrics to analyze a wide range of visuo-cognitive strategies and the value for facilitate the application of statistical techniques for inter- and intra-subject analysis.

Keywords: interaction · eye-tracking · analyze method · cognition · scanpath

1 Introduction and Objectives

Human vision offers valuable data for cognitive ergonomics [1–5]. Previous studies on perceptual processes and visual behavior has uncovered the complexity of visual activities and their link to high-level cognitive processes like linguistic and semantic comprehension [6, 7]. Yarbus [8] demonstrated that a user's objectives dictates their visual strategies. The study of these strategies has grown thanks to the development of portable eye-tracking technologies. However, processing the vast amounts of data they produce and designing experiments to yield reliable and contextualized insights about the cognitive processes, remain a challenge. Studies often analyze areas of interest through average viewing time, stimulus salience, and visual paths influenced by interface design [9]. In reading and language research, eye-tracking has provided valuable insights into linguistic and semantic processing [10–12], although incorporating subjects' objectives remains challenging, affecting visual strategies and complicating analysis in both laboratory and ecological settings. Aiming to tackle these challenges,

this paper proposes a new method for processing and analyzing ocular behavioral data: the *visuo-cognitive cartography (CVK)*. To test this method, we use an experimental case study of ocular behavior during reading with narrative referential changes. This study aims to expand eye-tracking analysis tools, providing empirical evidence to enrich theories for text interaction with potential applications in ergonomics and others.

2 Methods

The ocular behavior data corpus consists of 10 raw files, detailing the eye movements of 32 subjects during a reading task of 26 texts, categorized into training texts (2), target texts with referential change (8), control texts without referential change (8), and distractor texts (8). The texts were presented randomly during the experiment. The 32 subjects were divided into two groups (G1 and G2). The 16 subjects in G1 observed manipulated stimuli in a counterbalanced manner with G2. Target stimulus texts for G1 were control texts for G2, and vice versa [13]. Distractor texts were the same for both groups. The corpus was organized by participants (32), groups (2), and stimulus categories (4), including training texts. Incomplete recordings were removed, reducing the corpus to 22 subjects (11 in each group) who read 21 texts, categorized into control texts without referential change (7), target texts with referential change (7), and distractors (7). Data were filtered by word type area of interest (AOI). Eye movements were collected using a fixed SMI RED 500 system. Texts were displayed on a 22" LCD with a 1680x1050px resolution and 60 Hz refresh rate. Data were retrieved as CSV files from the SMI Begaze (v3.7), containing records on fixations (64,355), saccades (82,811), landing positions (37,071), pauses (91,027), regressions to zones of interest (2956), regressions between zones of interest (6763), initial regression saccades (43,991), return sweeps (5029), and overall reading speed (818).

3 Data Analysis

3.1 Pipeline - Pretreatment of Data for A Visuo-Cognitive Cartography

- **Generation of *n*-dimensional workspaces:** After cleaning and delimiting the ocular database, relevant information from each CSV file was extracted and organized: 462 records with 18 working dimension, combining data from 22 subjects and 21 texts. Each dimension represents the mean of corresponding measures recorded per subject and text. This *n*-dimensional workspace ($d = 18$) allows for the study of subjects' behaviors based on internal (global behavior) and external (stimulus) characteristics.
- **Ocular behavior as a *K*-dimensional vector type:** By compiling average spatial and temporal measures from subjects' readings and organizing them in the *n*-dimensional space, an 18-dimensional ocular behavior vector (OBV) was generated for each participant and text, filling the *n*-dimensional space with *K*-dimensional vectors. This setup facilitates the use of specific algebraic operations within the vector space.
- **Reference ocular vectors:** Two types of reference ocular vectors were defined

1. Global Ocular Behavior Vector (GOBV): $\sum OBVs$;

2. Normalized Ocular Behavior Vector (SBV): $GOBV / \# \text{ of behaviors recorded}$, for each case.

Each oculometric vector, whether OBV, GOBV, or SBV, shares characteristics such as modulus, direction, and orientation. Since all measurements are positive on a real number scale, the vector directions are always positive.

- **Similarity and Distance Measures between Vectors:** Vector spaces allow various operations to be performed between vectors, such as comparing vector length (modulus). This is derived from the generalized Pythagorean theorem for k-dimensional vectors. Another operation consists of calculating the scalar product as magnitude and the angle between vectors as similarity. The latter is used in NLP to find similarities of terms in n-dimensional corpora. Finally, it is possible to calculate another type of distance, based for example on the Manhattan distance.

3.2 Visuo-Cognitive Cartography (CVK) with A Real Corpus

Reference vectors GOBV and SBV were calculated for each participant. A database was then constructed with the OBV for each participant and text. Comparisons between the SBV and OBV for each participant were performed using the similarity measure based on leading cosines (COS). The modulus (MOD) of each OBV and the Manhattan distance (MH) between the GOBV and OBVs for each participant and text were also calculated. This resulted in a new database containing 462 records (22 participants x 21 texts), organized by group (G1, G2) and stimulus category (WC, WOC, D), with comparisons by cosines, modulus, and distances.

4 Results

4.1 Standard Strategy Divergences as an Analysis Metric

Using similarity calculations and reference vectors, the study assessed whether individuals' ocular behavior varies with different text types. The samples did not meet homoscedasticity and normality assumptions (Fig. 1 - A). Therefore, Welch's test, a non-parametric tool, was applied. Results revealed significant differences in ocular behavior based on text type ($t_{(2291)} = 7.20$, p-value < 0.001). Post-hoc analysis with the Games-Howell test indicated marked differences between distractor (D) texts and WC target texts ($t_{(262)} = 3.79$, p-value < 0.001), and between D and WOC texts ($t_{(306)} = 2.48$, p-value < 0.05), but no significant differences between WC and WOC categories (Fig. 1 - B). This figure shows that participants concentrate their cognitive resources according to a standard norm (SBV), adapting their visual strategy to the text type. The standard visual strategy is activated when participants encounter manipulated texts (WC or WOC). With distractor texts, participants' visual strategies deviate from the standard. However, for control or target texts, the strategy aligns more closely with the standard visual strategy. While no significant differences were found between WC and WOC texts, a slight difference and reduced variance in strategies were observed, indicating the similar characteristics of WC and WOC texts make differentiation difficult during reading. These findings suggest that participants use a consistent visual strategy for evaluation texts but adjust their approach when recognizing distractor texts.

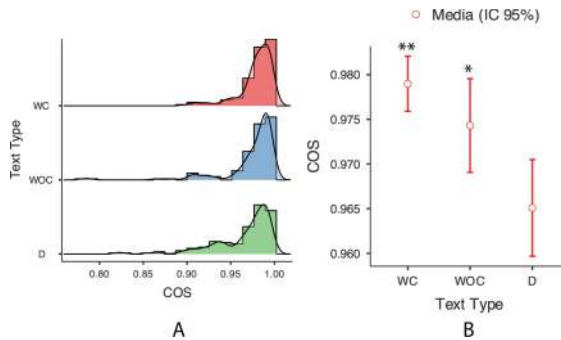


Fig. 1. Similarity (cos) of specific visuo-cognitive strategies by text type. A) vector distances between SBV and OBV.) Effect of text type on distance (COS) between OBV and SBV. 95% confidence interval.

4.2 Number of Cognitive Resources as a Metric for Categorizing Visual Strategies

We calculate cognitive resource utilization by determining the vector magnitude for each subject. The overall mean magnitude was 572.03 (SD = 76.98). To verify if this metric could classify visual strategies, a Shapiro-Wilks test confirmed the normality of the distribution ($w = 0.95$, $p > 0.05$). This stability represents each participant's visual-cognitive strategy and can be used statistically to classify subjects (Fig. 2).

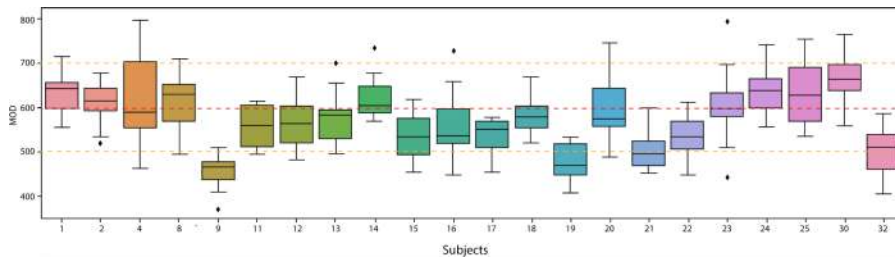


Fig. 2. Distribution of subjects in relation to average ocular behavior.

Figure 3 - A depicted each participant, using the mean OBV modulus for the X-axis and the standard deviation for the Y-axis. Applying the K-means technique grouped participants into four clusters based on the similarity of their visuo-cognitive strategies. (Figure 3 - B). This classification enriches analysis, even without predefined clusters in the experimental setting.

Comparing distances between individual strategies by text type and standard strategies, while considering stability clusters, enhances effect analysis. Figure 4 illustrates OBV-SBV similarity by participant category and text type, showing the stability of visual strategies in three experimental conditions (WC, WOC, D) and referenced by the OBV modulus (K-means - G). Findings revealed that divergence between standard strategies and those for each text type varied by group. The KM-0 group (red) showed greater

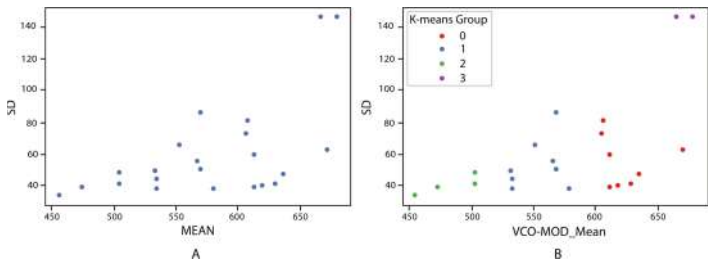


Fig. 3. Subject characterization by ocular behavior stability and classification (K-means).

divergence for WC and D texts than for WOC texts. Other groups exhibited greater divergence in D-type texts. The KM-2 group (green) had reduced divergence for WC texts compared to the standard strategy. The KM-1 (blue) and KM-3 (purple) groups displayed stable divergence for WC and WOC texts but reversed for D texts.

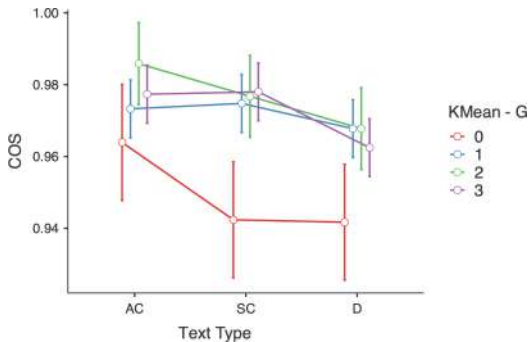


Fig. 4. Effects of text type on similarity (COS) of ocular behaviors.

4.3 Cumulative Visual-Cognitive Behavior as an Analysis Filtering Metric

The OBV vector’s magnitude represents cumulative visual-cognitive behavior during reading. Statistical analysis was conducted using the group variable from the original experiment (G1 and G2), which was assigned randomly. While the homoscedasticity hypothesis was confirmed, the normality hypothesis was not. Adding the text type variable (WC, WOC, D) to the analysis showed no significant differences overall. However, crossing group and text type variables revealed differing behaviors between G1 and G2 (Fig. 5. - A). A post-hoc test found that VCO magnitude behaved similarly for G1 and G2 with WC and D text types (Fig. 5. - C). Significant differences were observed in WOC texts ($t_{(455)} = -5.70$, $p\text{-value} < 0.001$, $d = -0.928$) and D texts ($t_{(455)} = -4.682$, $p\text{-value} < 0.001$, $d = -0.754$), but not for WC texts ($t_{(455)} = -6.22$, $p\text{-value} < 0.001$, $d = 0.583$), indicating similar cognitive resource use strategies for WC texts across groups. This suggests that WC texts generate similar strategies for both groups, while WOC and D texts lead to differentiated strategies due to the structure of presentation stimuli. The

method could potentially allow for more granular analysis by changing the vector space from a macro-scale (document) to a mesoscale (paragraphs and words) or micro-scale (syntags), although this analysis is beyond the current work.

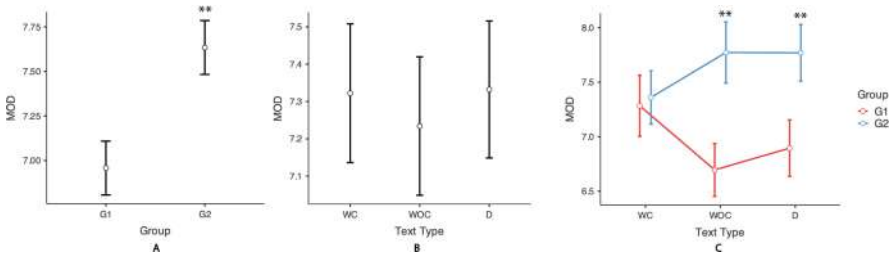


Fig. 5. Effects of visual-cognitive behavior (MOD) by group and text type.

5 Discussion and Conclusions

Our study demonstrates that visuo-cognitive cartography (CVK) can transform eye movement data into an analysis model combining n-dimensional vector space and k-dimensional vectors. This method reveals relationships between ocular behavior and the subject's tasks or actions, enhancing eye-tracking analysis. CVK offers numerous applications across different scenarios, including texts [14], objects, and scenes. Our findings highlight CVK's potential and its metrics in analyzing a wide range of visuo-cognitive strategies. This enables statistical techniques to identify significant differences between subjects, groups, experimental conditions, and moments of experience. CVK also offers filtering capabilities to identify critical analysis points and supports macro, meso, and micro-level analyses. Beyond data processing, CVK organizes ocular behavior information and connects it with underlying cognitive processes. Its metrics help understand the adaptability of cognitive resources for tasks, aligning with the ergonomic methodological framework by understanding activity and regulatory aspects, promising future applications in eye-tracking research.

References

1. Amadiou, F., Bastien, C., Tricot, A.: Les Méthodes On-Line 1: Analyse Des Parcours. In: *Ergonomie Des Documents Électroniques*, vol. 14, pp. 251–270. Presses Universitaires de France, Paris cedex (2008)
2. Bastien, J., Tricot, A.: L'évaluation Ergonomique Des Documents Électroniques. In: *Ergonomie Des Documents Électroniques*, vol. 14, pp. 205–227. Presses Universitaires de France, Paris cedex (2008)
3. Chevalier, A., Tricot, A.: *Ergonomie Des Documents Électroniques*. Presses universitaires de France, Paris (2008)
4. Couronné, T.: *De La Prise D'information Visuelle À La Formation D'impressions: Apports De L'oculométrie Pour L'étude Des Processus De La Perception Et De La Cognition Visuelle Des Objets Manufacturés*. Université Joseph-Fourier-Grenoble I (2007)

5. Desnoyers, L.: 5. La Prise D'information. In: *Ergonomie*, 1^{re} edn, p. 83. Presses Universitaires de France (2004)
6. Rayner, K.: Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **124**, 372–422 (1998)
7. Just, M.A., Carpenter, P.A.: A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* **87**, 329 (1980)
8. Yarbus, A.L.: *Eye movements and vision*. Plenum Press, Michigan (1967)
9. Plużyczka, M.: The first hundred years: A history of eye tracking as a research method. *Appl. Linguist. Pap.* **4**, 101–116 (2018). <https://doi.org/10.32612/uw.25449354>
10. Lévy-Schoen, A.: La flexibilité des saccades et des fixations au cours de la lecture. *Année psychol.* **80**, 121–136 (1980). <https://doi.org/10.3406/psy.1980.28306>
11. Liu, P.-L.: Using eye tracking to understand learners' reading process through the concept-mapping learning strategy. *Comput. Educ.* **78**, 237–249 (2014). <https://doi.org/10.1016/j.compedu.2014.05.011>
12. Rayner, K.: Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **62**, 1457–1506 (2009). <https://doi.org/10.1080/17470210902816461>
13. Ballenghein, U., Baccino, T.: Referential processing during reading: concurrent recordings of eye movements and head motion. *Cogn. Process.* **20**, 371–384 (2019). <https://doi.org/10.1007/s10339-018-0894-1>
14. Chavez, D.: *Ergonomie de la connaissance : outils cognitifs appliqués à l'analyse des activités de connaissances*. Thèse de doctorat. Université Paris sciences et lettres (2022)



Supervision of Automated Production Systems: The Effects of Task Demands, Time on Task, and Task Reward on Visual Attention Allocation in a Process Monitoring Task

Sebastian Pütz^(✉), Alexander Mertens, and Verena Nitsch

Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen,
Germany

{s.puetz,a.mertens,v.nitsch}@iaw.rwth-aachen.de

Abstract. To supervise automated production systems, operators have to allocate their visual attention among a variety of information sources to maintain their understanding of the current system state. In the present study, we investigated how this process is affected by the interplay of different task characteristics (task demands, time on task, and task reward) and their effects on dimensions of the operator's mental state (mental workload and mental fatigue). To this end, 50 participants performed a process monitoring task with ten experimental blocks, which alternated between low and high task demands. In addition, the last two blocks included a task reward manipulation. We measured eye fixation durations and positions as indicators of attention allocation. In line with our hypotheses, the analyses revealed that participants' fixation durations decreased with higher task demands, increased over time on task, and decreased with the addition of task reward. We also found that participants' attentional focus narrowed over time on task and widened again with the addition of task reward. Our results demonstrate how operators can cope with changing task demands by adapting their attention allocation. The results also highlight safety implications of mental fatigue, as both slower shifts in operators' visual attention and restricted attentional focus create the potential for operators to miss critical changes in system state.

Keywords: automation · eye-tracking · fixation

1 Introduction

The progressing digitalization of the industrial sector is expected to lead to a continuous shift of the role of human operators from manual labor to the supervisory control of complex production systems and processes [1]. This shift introduces the well-known human factors challenges of human-automation interaction in production systems, in which human tasks can range from passive monitoring over long periods of time to decision-making under time pressure in the event of critical system states [2]. The resulting task conditions place strain on the operator's mental state, both in terms of fluctuating mental workload due to changes in task demands, and mental fatigue, accumulating over time on task [3].

2 Visual Attention Allocation

With regard to operators' mental state, research on visual attention allocation has followed two primary rationales. On the one hand, visual attention allocation is a key behavioral component of supervisory tasks. Typically, operators have to monitor large amounts of mostly visually presented information and allocate their attention between different sources of information. This allocation process determines what and when information is perceived and processed, and is therefore central to human supervisory performance. On the other hand, researchers have investigated features of visual attention allocation as indicators of operators' mental state [4]. Respective allocation features include temporal features, such as eye fixation duration and the number of fixations per time interval, as well as the spatial distribution of fixations.

Although previous research has investigated operator's visual attention allocation in supervisory process control tasks, most studies have focused on the effect of individual task characteristics or interface design features and have analyzed their effects with respect to a single dimension of the operator's mental state [3]. Thus, the goal of the present study was to examine the interplay of different task characteristics of a process monitoring task, specifically task demands, time on task, and task reward. Furthermore, their effects on the visual attention allocation are discussed in relation to the scientific literature on both mental workload and mental fatigue.

The task characteristic most often studied with respect to how it affects operators' visual attention allocation is task demand, the primary antecedent of mental workload [5]. Interestingly, the relationship between task demands and the duration of eye fixations is strongly depended on the types of demands. When more complex information must be processed, fixation durations tend to increase [6, 7]. In contrast, fixation durations typically decrease when humans must process information from more visual sources or with less time [8, 9]. As a result, treating fixation durations as index of operator's mental workload requires a nuanced analysis of the task.

Unlike mental workload, the effect of mental fatigue on visual attention allocation has received little consideration in supervisory control research [3]. The phenomenon of fatigue is strongly associated with time on task, with mental fatigue increasing over time on task, causing operators to disengage [10, 11]. Given the strong influence of motivational factors, researchers have also examined the relationship between task reward and mental fatigue. Specifically, it has been shown that introducing task reward after performing a mental task for an extended period of time alleviates the task disengagement caused by mental fatigue [10, 11].

3 Hypotheses

For the process monitoring task used in the present study, we hypothesized that increasing mental workload, induced by higher task demands in the form of less predictable process parameter changes, would lead to a faster reallocation of visual attention, whereas mental fatigue would have the opposite effect—H1: “*Fixation durations (a) decrease with higher task demands, (b) increase over time on task, and (c) decrease with the addition of task reward.*” Furthermore, we included an exploratory hypothesis on the spatial distribution

of fixations. We assumed that mental fatigue would be reflected in a narrower attention focus, which we operationalized as the average offset of fixations from the center of the monitoring screen—H2: “*The average offset of fixations from the center of the screen (a) decreases over time on task and (b) increases with the addition of task reward.*”

4 Methods

4.1 Participants

A sample of $N = 50$ participants (28 women; $M_{\text{age}} = 24.34$ years, $SD_{\text{age}} = 3.60$ years) was recruited for the study. Participants received a compensation of 20 €. The study design was approved by the Ethics Committee at RWTH Aachen University.

4.2 Materials

Participants performed a process monitoring task, in which they observed a three-by-five grid of parameter gauges and responded to critical parameter value deviations by pressing an associated alarm button. The task was divided into ten 9-minute blocks, with blocks alternating between a low and a high task demands condition. Task demands were manipulated in terms of the temporal uncertainty of parameter value changes, including both their consistency and the maximum rate.

The final two experimental blocks included a task reward manipulation in which participants received instantaneous feedback on their response times in terms of a visually presented score between 1 and 10 points. In addition, they were instructed that they could earn a monetary bonus of 5 € for reaching a defined performance threshold. In fact, all participants received the bonus at the end of the study.

Participants were seated at a distance of 70 cm from a 27 in. IPS monitor with a resolution of 2560×1440 pixels, which displayed the experimental task. The 15 gauges occupied a 25×32 cm area in the center of the screen. Participants' visual attention allocation was assessed using an FX3 remote eye-tracker. Eye-tracking data were used to identify fixations and to estimate their duration and position using EyeWorks version 3.21 by EyeTracking. The fixation positions were the basis for calculating their offset from the center of the screen.

The data analyzed in this article is derived from the same study reported in a previous publication, which provides additional details on the experimental setup [12].

4.3 Analysis

Data from fixations outside the screen area of the parameter gauges were excluded (1.17% of fixations), resulting in data from 309,042 fixations being submitted for statistical analysis. The research hypotheses were tested using linear mixed models including task demands, time on task, task reward, an interaction between task demands and time on task, and an interaction between task demands and task reward. The models also included a random intercept per participant. Fixation durations were log-transformed to account for the distributional properties of the data.

5 Results

5.1 Mean Fixation Duration

Mean fixation durations are presented in Fig. 1a. The analysis of fixation duration revealed significant main effects of all independent variables as well as a significant interaction between task demands and task reward (s. Table 1). Fixations were shorter in the high task demands condition, became longer over time on task, and became shorter with the addition of task reward. The task reward effect was more pronounced in the high task demands condition.

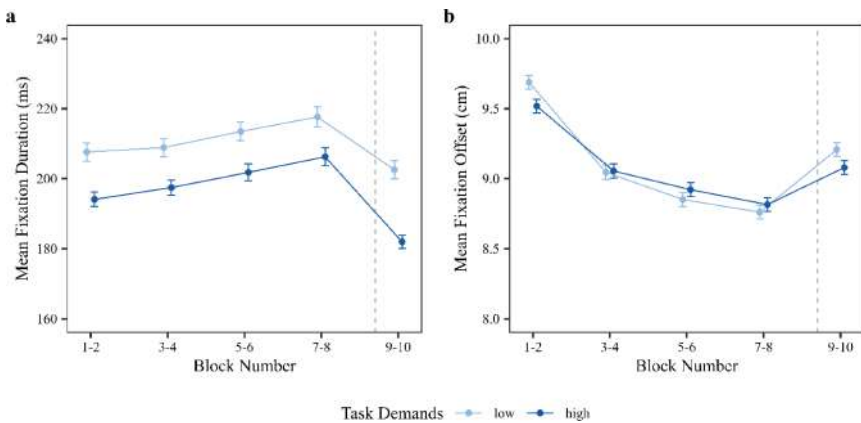


Fig. 1 Mean fixation duration (a) and mean fixation offset from the center of the screen (b) are displayed as a function of task demands and time on task, as well as task reward in blocks 9 and 10. Blocks are grouped in pairs to account for counterbalancing the order of block task demands levels across participants. Error bars indicate 95% confidence intervals

Table 1. Linear mixed models for fixation duration and fixation offset

	<i>B</i>	<i>SE B</i>	95% <i>CI</i>	<i>p</i>
Fixation Duration (log(ms))				
(Intercept)	5.053	0.019	[5.016, 5.090]	
Task Demands	−0.027	0.005	[−0.037, −0.019]	< .001
Time on Task	0.005	0.000	[0.004, 0.007]	< .001
Task Reward	−0.065	0.005	[−0.075, −0.054]	< .001
Task Demands x Time on Task	0.000	0.001	[−0.003, 0.002]	.721
Task Demands x Task Reward	−0.016	0.008	[−0.032, −0.001]	.033
Fixation Offset (cm)				
(Intercept)	9.683	0.154	[9.378, 9.987]	
Task Demands	−0.111	0.032	[−0.174, −0.048]	.001

(continued)

Table 1. (continued)

	<i>B</i>	<i>SE B</i>	<i>95% CI</i>	<i>p</i>
Time on Task	−0.155	0.005	[−0.166, −0.145]	< .001
Task Reward	0.901	0.039	[0.825, 0.977]	< .001
Task Demands x Time on Task	0.034	0.008	[0.019, 0.050]	< .001
Task Demands x Task Reward	−0.314	0.055	[−0.422, −0.206]	< .001

5.2 Mean Fixation Offset

Mean fixation offsets are presented in Fig. 1b. The analysis of fixation offset revealed significant main effects of all independent variables and both interaction terms (s. Table 1). The average distance between fixation positions and the center of the screen was smaller in the high task demands condition, decreased over time on task, and increased in the task reward blocks. Both the time on task effect and the task reward effect were smaller in the high demands condition.

6 Discussion

The statistical analyses supported H1, highlighting the dependence of fixation duration on task demands, time on task, and task reward. Participants responded to the faster and less predictable parameter value changes in the high task demands condition by shifting their visual attention more rapidly. This enabled them to update their knowledge of the current state of the process parameters in a timelier manner. Furthermore, the analyses showed how this strategy is hindered by the onset of mental fatigue. The observed effect demonstrates safety implications of mental fatigue, as slower attentional reallocation may result in operators not noticing critical system states in time.

The analyses also supported H2, with operators’ visual attention becoming more focused on the center of the screen over time on task. The effect was moderate in magnitude, with a decrease in mean fixation offset of about 10% from the first to the last block before the addition of task reward. Importantly, the present study involved only a single monitoring screen and a relatively small number of parameters gauges. Future studies should investigate whether the observed effect is also present, or even more pronounced, in typical multi-screen setups. If so, the effect would represent a further safety implication of mental fatigue, with an increased risk of operators missing important information presented in their peripheral vision.

6.1 Limitations and Outlook

The obtained results are dependent on the examined process monitoring task. While in this comparatively simple monitoring task participants could identify critical system events directly from the sensory input of the parameter gauges, it remains to be investigated how higher cognitive processing demands would influence the observed effects.

Furthermore, the analyses should be extended to multi-task scenarios where participants have to divide their visual attention to balance competing task goals.

Funding. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2023 Internet of Production—390621612.

Data Availability. The dataset and R script for the reported analyses are available in the Open Science Framework repository, <https://doi.org/10.17605/OSF.IO/DRB5P>.

References

1. Spath, D., Braun, M.: Human factors and ergonomics in digital manufacturing. In: Salvendy, G., Karwowski, W. (eds.) *Handbook of Human Factors and Ergonomics*, 5th, pp. 1438–1459. John Wiley & Sons, Hoboken, NJ (2021)
2. Endsley, M.R.: From here to autonomy: Lessons learned from human–automation research. *Hum. Factors*. **59**, 5–27 (2017). <https://doi.org/10.1177/0018720816681350>
3. Pütz, S., Mertens, A., Chuang, L.L., et al.: Physiological measures of operators mental state in supervisory process control tasks: A scoping review. *Ergonomics*. **67**, 801–830 (2024). <https://doi.org/10.1080/00140139.2023.2289858>
4. Tao, D., Tan, H., Wang, H., et al.: A systematic review of physiological measures of mental workload. *Int. J. Environ. Res. Public Health*. **16**, 2716 (2019). <https://doi.org/10.3390/ijerph16152716>
5. Young, M.S., Brookhuis, K.A., Wickens, C.D., et al.: State of science: Mental workload in ergonomics. *Ergonomics*. **58**, 1–17 (2015). <https://doi.org/10.1080/00140139.2014.956151>
6. Qiu, Y., Pan, D., Li, Z., et al.: A human reliability analysis approach based on the concepts of meta-operation and task complexity. In: *Proceedings of the Probabilistic Safety Assessment and Management PSAM*, vol. 12, pp. 1–11 (2014)
7. Wang, Y., Hu, R., Lin, S., et al.: The impact of automation on air traffic controller's behaviors. *Aerospace*. **8**, 260 (2021). <https://doi.org/10.3390/aerospace8090260>
8. Foy, H.J., Chapman, P.: Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Appl. Ergon.* **73**, 90–99 (2018). <https://doi.org/10.1016/j.apergo.2018.06.006>
9. De Rivecourt, M., Kuperus, M.N., Post, W.J., et al.: Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*. **51**, 1295–1319 (2008). <https://doi.org/10.1080/00140130802120267>
10. Hopstaken, J.F., van der Linden, D., Bakker, A.B., et al.: A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*. **52**, 305–315 (2015). <https://doi.org/10.1111/psyp.12339>
11. Hopstaken, J.F., van der Linden, D., Bakker, A.B., et al.: The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. *Biol. Psychol.* **110**, 100–106 (2015). <https://doi.org/10.1016/j.biopsycho.2015.06.013>
12. Pütz, S., Mertens, A., Chuang, L.L., et al.: Physiological predictors of operator performance: The role of mental effort and its link to task performance. *Hum Factors Online First*. **67**(6), 595–615 (2024). <https://doi.org/10.1177/00187208241296830>



Dynamic Signs: Statistical Extraction Of Ergonomic Requirements of Moving Signs from Interviews and Questionnaires for Various Users

Reiko Sakata^{1(✉)}, Saki Matsui¹, Atsushi Shimada¹, Naoki Furuhata¹, Nana Ito², Ken Sagawa², Hiroshi Watanabe², and Hiroyasu Ujike³

¹ Mitsubishi Electric Corporation, Kanagawa, Japan

Sakata.Reiko@ab.MitsubishiElectric.co.jp

² National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

³ Tokyo Information Design University, Tokyo, Japan

Abstract. Older people and people with vision, hearing, physical, and cognitive disabilities were surveyed regarding in-facility moving signs through interviews and questionnaires. Principle component analysis was carried out, by attribute, on the response data indicating “agreement/disagreement” of respondents to items on the list of issues (requirements on signs). The results reveal differences in issues which must be addressed by dynamic signs for each attribute. People with low vision tend to agree on issues related to sign placement, people with hearing disabilities on sign expression, wheelchair users on sign style of movement and use of sound, people with cognitive disabilities on icons and colors composing the sign, older people on sign brightness/contrast, and older people with dementia on simplicity of information. Based on the trends of all attributes, the international standards relating to accessibility of dynamic signs that we are about to propose should emphasize “communicating information simply.”

Keywords: dynamic signs · accessibility · principal component analysis

1 Problem Statement

ISO/TC 159/SC 5/WG 7 is reviewing standards for defining ergonomic requirements of moving signs called “dynamic signs” which change or are animated according to the situation. Basic rules for addressing accessibility are defined in ISO/IEC Guide 71 [1], but there is a need to identify and flesh out requirements for dynamic signs. There are several standards for signs that are easy to understand for people with vision, hearing, and physical disabilities [2–5], there is room for review of sign design for people with cognitive disabilities and older people with dementia.

2 Objective

Older people and various types of disabilities were surveyed regarding in-facility moving signs through interviews and questionnaires. The results reveal differences in issues which must be addressed by dynamic signs (requirements on signs) for each attribute.

3 Methodology

3.1 Preparation of Issue List (Interview Survey)

An interview survey was conducted with 3 persons with low vision, 3 persons with hearing disabilities, 8 persons with cognitive disabilities, a caregiver for older, and a designer of care facilities. They were asked about issues relating to in-facility moving signs. A list of issues (requirements on signs) was prepared with a total of 46 items, classified into five categories (Table 1). Category A is an issue with 12 items pertaining to the desired locations and frequency of posting signs inside facilities. Category B is an issue with 11 items pertaining to the indication methods of moving signs, such as scrolling and flashing, and the use of audio signs. Category C is an issue with 9 items pertaining to sign color and brightness. Category D is an issue with 8 items pertaining to characters and sentences, such as character types, wording, and parallel indication of symbols. Category E is an issue with 6 items pertaining to guidance content.

3.2 Questionnaire Survey

Using a web questionnaire, the list of issues described above was provided to people with low vision (167 persons), hearing disabilities (64 persons), cognitive disabilities (167 persons), wheelchair users (52 persons), and older people (186 persons) (Table 2), and they were asked to indicate, through multiple-choice selection, issue items they agreed with.

3.3 Analysis Method

Principle component analysis was carried out, by attribute, on the response data indicating “agreement/disagreement” of respondents to items on the list of issues (requirements on signs). As a result, it was found that, for all attributes, the contribution of the first principal component exceeded 20%, and the contribution of the second and subsequent components was 10% or less. Thus, in this study, focus was placed only on the first component score as the indicator with the most information on requirements for dynamic signs for each attribute. For each attribute, a standardized value was used for the principal component score, with a mean of 0 and a standard deviation of 1. Issue items for which the absolute value of the principal component score is high are taken as indicating the features of each respondent attribute, and items with an absolute value larger than $\pm 1.645\sigma$, where the range exceeds approximately 95% of the distribution for the standardized principal first component, were identified.

Table 1. List of issues for signs

A. Placement and quantity	
A-1	Want guidance posted at eye height
A-2	Want signs posted at high locations such as overhead so they are visible even during crowding
A-3	Want guidance signs placed on the floor so they are easy to find and directions are easy to understand
. . .	. . .
B. Style of movement and use of sound	
B-1	For guidance signs with a moving display, want the display to repeat or be redisplayed at a fixed interval
B-2	With multilingual guidance, e.g., in four languages, want Japanese to be displayed a little longer
B-3	Want static displays, if possible, not displays with switching lettering, like the video billboards at stations and airports
. . .	. . .
C. Color, illumination, brightness	
C-1	Want guidance signs themselves and their surroundings to not be too glaring, due to LEDs that are too bright, etc
C-2	Want display with high contrast and a clear difference between light and dark
C-3	Want contrast to not be set too high
. . .	. . .
D. Characters and sentences	
D-1	Want signs to use clear and simple wording
D-2	Want character types to be set to a uniform, easy-to-read font
D-3	Want the character font to be Gothic type
. . .	. . .
E. Content of guidance	
E-1	Want guidance on urgent notifications which change from moment to moment, like train changes
E-2	Important information is hard to find, so want highly important guidance to be provided concisely
E-3	Want information that is important for some people, such as those with disabilities, to be provided
. . .	. . .

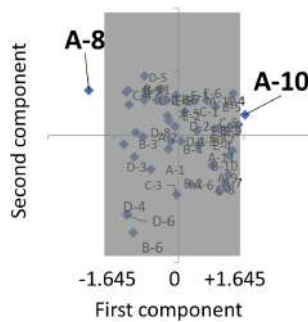
Table 2. Attributes of questionnaire subjects (636 persons total)

Low vision, 167 persons (145 persons with disabilities + 22 cohabiting family members)					
Central visual field only	22 persons	Peripheral visual field only	53 persons	Either left or right visual field	35 persons
Blurring of entire visual field	15 persons	Missing part of visual field	8 persons	Color blindness (protanopia or deuteranopia)	23 persons
Photophobia	4 persons	Night blindness	7 persons		
Hearing disabilities (64 persons)					
Deafness (congenital)		11 persons	Deafness (acquired)		21 persons
Hard of hearing (congenital)		11 persons	Hard of hearing (acquired)		21 persons
Cognitive disabilities, 167 persons (147 persons with disabilities + 20 cohabiting family members)					
Autism	42 persons	Inattentive dominant type	42 persons	Asperger syndrome	21 persons
Hyperactive/impulsive dominant	21 persons	Dyscalculia	11 persons	Dysgraphia	4 persons
Dyslexia	5 persons	Combined type	21 persons		
Wheelchair users (52 persons)					
Older (187 persons)					
Family members of older people with dementia		100 persons	Older people (65 or older)		87 persons

4 Results

For people with low vision, the first component had features in content relating to sign placement (Fig. 1). People with low vision have difficulty even finding the signs themselves, and they stop to understand sign content. Therefore, for dynamic signs in particular, emphasis is placed not just on sign content and expression, but also on consideration of sign placement.

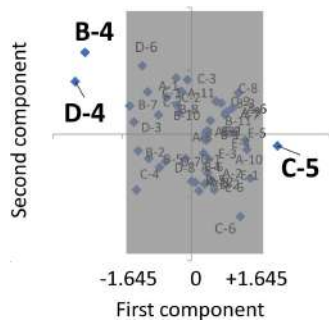
For people with hearing disabilities, the first component had features in content relating to sign expression (Fig. 2). People with hearing disabilities have a high probability of obtaining information only through sight, and thus desire simplicity enabling immediate understanding of the content, and prominent expression such as flashing for signs that communicate danger.



A-8) Want signs placed at locations that easily catch the eye, like facility entrances, or immediately after exiting an elevator, train, etc.

A-10) Want signs placed at locations where people stopping to look at a guidance sign will not block the flow of people

Fig. 1. Distribution of principal component scores for people with low vision.



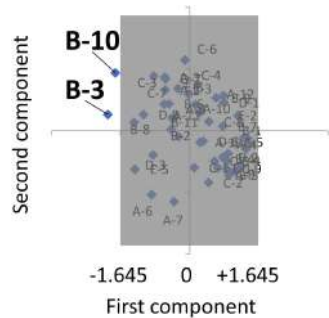
B-4) Want guidance signs indicating danger to be flashing and prominent

C-5) If use of color is meaningful, such as yellow encouraging caution, want the meaning to be more easily understood by using highly saturated colors like safety colors

D-4) Want signs to provide guidance with symbols and pictograms, avoiding the use of text as far as possible

Fig. 2. Distribution of principal component scores for people with hearing disabilities.

For wheelchair users, the first component had features in content relating to sign style of movement and use of sound (Fig. 3). Wheelchair users, who have a low line of sight and find it hard to stop in corridors, only have a short time when they can see signs, so there is a need to avoid, as far as possible, displays on which the presented information changes from moment to moment, and to supplement the display with information provided via audio.



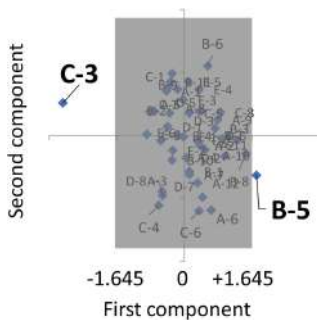
B-3) Want static displays, if possible, not displays with switching text, like the video billboards at stations and airports

B-10) For expansive spaces, like stairs or elevators, want audio signs to indicate the direction to go in

Fig. 3. Distribution of principal component scores for wheelchair users.

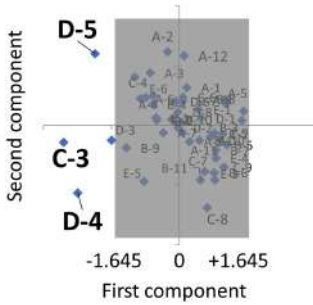
- C-4) Want the ability to discriminate based on color, as with guidance for men's and women's restrooms
- C-5) If use of color is meaningful, such as yellow encouraging caution, want the meaning to be more easily understood by using highly saturated colors like safety colors
- D-6) Want content pared down to a single word or icon on one guidance sign
- D-4) Want signs to provide guidance with symbols and pictograms, avoiding the use of text as far as possible

For older people, the first component had features in content relating to sign brightness/contrast (Fig. 5). Generally speaking, displays with higher contrast have improved visibility, but people like the older have the attribute of sensitivity to display stimulation, so care must be taken to not excessively strengthen stimulation, e.g., by avoiding glare and intense flashing.



- B-5) In the case of flashing guidance signs, want an increased presentation time and clear display contrast
- C-3) Want contrast to not be too high

For older people with dementia, the first component had features in content relating to simplicity of information (Fig. 6). Dementia is an attribute involving difficulties with short-term memory. Those with the condition can actively use graphical symbols they are familiar with, such as those for restrooms, but it is hard for them to memorize and understand the meaning of unfamiliar symbols. Therefore, it is desirable to communicate information simply, e.g., by using text.



- D-4) Want signs to provide guidance with symbols and pictograms, avoiding the use of text as far as possible
- D-5) Want indication incorporating text because it is hard to understand the meaning with symbols or pictograms alone
- C-3) Want contrast to not be too high

Fig. 6. Distribution of principal component scores for older people with dementia.

5 Discussion

There are plans to make an international proposal of standards for ergonomic requirements of dynamic signs. These standards should consider not only design requirements for improving ease of understanding of moving signs, with elements such as contrast, symbols, text, and animation, but also description relating to installation locations so that signs can safely be seen. Also, there is an emphasis on “communicating information simply,” so it will be best if the standard can suggest systematic techniques for switching content and information based on user attributes.

Acknowledgement. These results were achieved through the project “International Standardization Relating to Safety and Accessibility of Dynamic Signs,” part of the Ministry of Economy, Trade and Industry’s initiative to promote acquisition and dissemination of international standards relating to energy efficiency, etc.

References

1. ISO/IEC Guide 71 Guide for Addressing Accessibility in Standards
2. ISO9186-1 Graphical symbols-Test methods-Part1:Methods for Testing Comprehensibility
3. ISO 7001 Graphical symbols—Public guidance symbols
4. ISO 24502 Ergonomics—Accessible design—Specification of Age-Related Luminance Contrast For Coloured Light
5. ISO 24509 Ergonomics—Accessible design—A Method for Estimating Minimum Legible Font Size for People at Any Age



Dynamic Signs: Effects of Position and Age on the Detection of Signs During a Dual Task

Hiroshi Watanabe¹(✉), Nana Ito¹, Ken Sagawa¹, Hiroyasu Ujike², Reiko Sakata³,
and Naoki Furuhata³

¹ National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8566, Japan
h.watanabe@aist.go.jp

² Tokyo Information Design University, Tokyo 132-0034, Japan

³ Mitsubishi Electric Corporation, Kanagawa 247-8501, Japan

Abstract. Dynamic signs have content that changes according to the situation or have a position or shape that changes to enhance visibility. The present study used a dual task method to clarify the characteristics of noticing dynamic signs while walking. Participants were asked to respond to a distractor task presented on the floor when viewing a scene in which they moved through an underground passageway in an immersive virtual reality device, while simultaneously reading dynamic signs presented on the floor, to the front, or on the ceiling. The performance of the participants was evaluated based on the correct response rate for the reading of the dynamic signs. When participants were divided into age groups, an effect of age on the attractiveness of dynamic signs was found. The results also revealed differences in the attractiveness of dynamic signs due to the presentation position, suggesting a more reasonable result than in prior studies.

Keywords: Dynamic Signs · Visibility · Attractiveness · Visual Field · Dual Task · Virtual Reality · Effect of Aging

1 Introduction

Dynamic signs have content that changes according to the situation or have a position or shape that changes to enhance visibility. ISO/TC 159/SC 5/WG 7 brings together various pieces of knowledge to establish an international standard for the ergonomic requirements of dynamic signs. Watanabe et al. [1] accumulated data on overall visibility for people with normal vision, while Itoh et al. [2] focused on contrast characteristics for people with low vision. Data on the attractiveness of dynamic signs remains limited, with Watanabe et al. [3] having conducted preliminary investigations on the noticing of dynamic signs while walking using an immersive virtual reality (VR) device. They found that presentation of signs on the floor and ceiling significantly reduced noticing. However, the reduced detection of signs presented on the floor appeared to depend on the resolution of the display. The present study aimed to clarify this issue by using a new display with higher resolution.

2 Methods

2.1 Participants

To determine the impact of age-related changes in cognitive function on sign reading, 30 participants were recruited from two age groups: 15 in their 20 s (mean age, 22.7 years; standard deviation (SD), 1.19 years), and 15 participants in their 60 s (mean age, 61.2 years; SD, 1.52 years). All participants had normal vision, hearing, and walking ability, and received compensation for their participation. On the day of the experiment, it was verified that participants were in good physical condition using a questionnaire, conducted monocular visual acuity tests using Landolt rings at a distance of 5 m, and performed binocular tests using the Stereo Fly Test (Stereo Optical Co. Inc., Chicago, IL). The Mini-Mental State Examination was administered to assess participants' cognitive function. These preliminary tests indicated that all participants had adequate vision and cognitive function for the experiment. No participants had prior knowledge of the content of the experiment. The experimental protocol was reviewed and given ethical approval by the Ergonomics Committee of the National Institute of Advanced Industrial Science and Technology (AIST). All participants provided written informed consent before starting the experiment.

2.2 Apparatus

Computer graphics imagery was displayed on LED panels (SiliconCore, P1.25-Floor Screen) using a CAVE system (University of Illinois) with the following dimensions and resolutions: front (2.4 m wide $\times$ 2.25 m high; resolution: 1920 px $\times$ 1800 px); sides (1.2 m wide $\times$ 2.25 m high; resolution: 960 px $\times$ 1800 px), and floor (2.4 m wide $\times$ 1.35 m deep, resolution: 1920 px $\times$ 1080 px). The luminance of each panel was 1200 cd/m². Participants wore LCD shutter-type glasses to view the stereoscopic images (frame rate: 120 Hz). The position and orientation of participants' heads were tracked using a Vicon system (Vicon Motion Systems Ltd., Oxford, UK) at a sampling rate of 120 Hz, and the data were transmitted to the CAVE system. This information was used to adjust the rendering, allowing participants to experience undistorted immersion in the virtual space.

2.3 Stimuli

The stimuli used in the experiment were the same as in Watanabe et al. [3]. The experiment was conducted in a VR environment simulating an underground passageway with a T-junction, with three goals (doors) on each side of the passage after the T-junction split (Fig. 1). The main task for the participants was to start moving toward the junction at a speed of 4 km/h from a starting point 15 m before the middle of the junction and just before entering the junction, they needed to obtain information from a dynamic sign (Fig. 2; 2 m wide $\times$ 1.15 m high) presented on the floor, to the front, or on the ceiling, indicating the direction of the goal (left or right) and the distance from the junction (near, medium, or far door). Next, as a dual task during the main task, participants responded to a distractor stimulus (0.25 m wide $\times$ 0.25 m high) presented about 3 m in

front of the dynamic sign, approximately 2.6 s before the dynamic sign appeared. The distractor stimulus consisted of a square, a pentagon, and a rhombus flashing in a random order. Participants were asked to click a mouse button when the rhombus appeared. If they noticed the appearance of the dynamic sign during this dual task, they were to immediately read and interpret the sign's information.

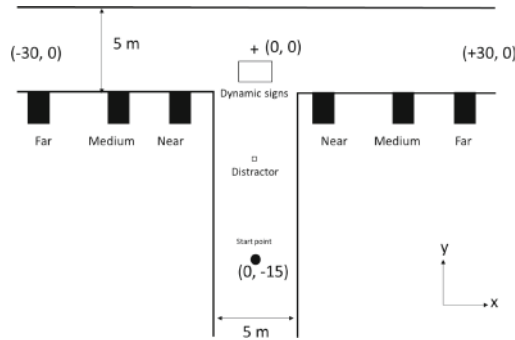


Fig. 1. Layout of the underground passageway used in the experiment.

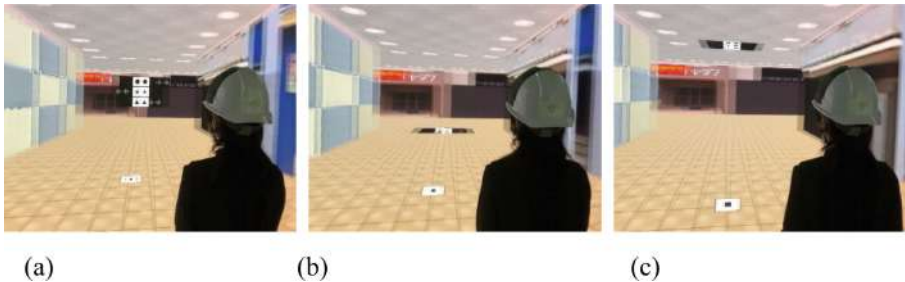


Fig. 2. Presentation positions of the distractor stimulus and the dynamic sign to the front (a), on the floor (b), and on the ceiling (c).

The dynamic signs consisted of circle, triangle, and star markers. Participants were required to always read the direction and position of the star marker from the signs. An arrow moving at a speed of 0.18 m/s was assigned to either the left or right of each marker, and its movement distance and position indicated the distance from the junction. Four types of arrangement conditions with different orderliness were set for the direction and distance information.

Direction In-order, Position In-order (DI-PI) condition: The left-right positions of all arrows are the same, and the distance indications are arranged in a sequence from top to bottom as near, medium, and far or as far, medium, and near (Fig. 3(a)).

Direction In-order, Position Disorder (DI-PD) condition: The left-right positions of all arrows are the same, but the movement distances of the signs are not in an ordered sequence (Fig. 3(b)).

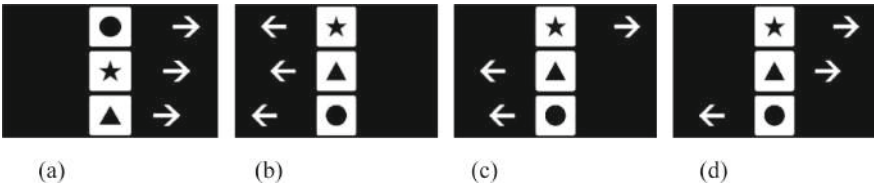


Fig. 3. Examples of four arrangement conditions.

Direction Disorder, Position In-order (DD-PI) condition: The left-right positions of the arrows are mixed, but the distance indications are arranged in a sequence from top to bottom as near, medium, and far or as far, medium, and near (Fig. 3(c)).

Direction Disorder, Position Disorder (DD-PD) condition: In The left-right positions of the arrows are mixed, and the distance indications are also not in an ordered sequence (Fig. 3(d)).

Multiple signs were created by randomizing the circle, triangle, and star mark positions in accordance with the above conditions.

2.4 Procedure

The experimental procedure followed that of Watanabe et al. [3]. After signing the consent form, participants were briefed on the information they should read from the dynamic signs and gained an understanding of direction and distance by solving practice exercises on paper. Next, they acquired an understanding of how to respond to the distractor stimuli using the actual experimental program (without displaying the dynamic signs). Each participant then put on a helmet equipped with head position sensors and stereoscopic glasses, and stood at the rear of the floor area to participate in the main experiment. Each trial began with a cue to start moving forward, and participants first responded to the distractor stimuli. If they noticed the appearance of a dynamic sign during this response, they were instructed to look toward it, read the sign's content (direction and distance), and respond verbally. The experimenter recorded these responses on a keyboard, the participant's movement stopped, and they were returned to the starting point.

Each participant conducted 48 trials as follows. Three types of signs were created by randomly changing the positions of the circular, triangular, and star markers in each of the four arrangement conditions. These were then displayed at four different dynamic sign presentation locations (to the front, on the floor, on the ceiling, and no display), resulting in a total of $4 \times 3 \times 4 = 48$ trials.

3 Results

Three types of signs were created for a single arrangement condition, and after performing the task of reading each one, correct responses of 0%, 33%, 66%, and 100% for the four arrangement conditions were obtained. These rates were determined for combinations of the four arrangement conditions and three presentation positions, defining

each participant's sign reading performance. The effects of combining the presentation position and arrangement condition on correct response rate, along with reaction times under each condition, are shown in Fig. 4, which also shows the results of an analysis of variance. Additionally, the pooled average correct response rates for the front, floor, and ceiling positions were 88%, 89%, and 76% for participants in their 20 s, respectively, and 80%, 72%, and 59% for participants in their 60 s.

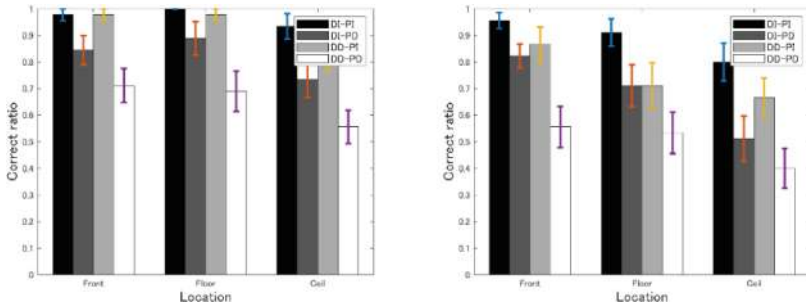


Fig. 4. Correct response rates and standard errors (SE) for each arrangement condition and presentation location (results for participants in their 20 s on the left and participants in their 60 s on the right)

For participants in their 20 s, significant main effects were observed for all conditions (presentation position $F(2, 28) = 8.33, p < .01$; direction $F(1, 14) = 16.72, p < .01$; distance $F(1, 14) = 60.16, p < .01$). Using the Holm method to compare mean values between presentation positions, no significant difference was found between the front and the floor ($\alpha' = 0.05, n.s.$), but a significant difference was found between the front and the ceiling ($\alpha' = 0.025, n.s.$), the floor and the ceiling ($\alpha' = 0.017, n.s.$).

For participants in their 60 s, significant main effects were also observed for all conditions, similar to those for participants in their 20 s (presentation position $F(2, 28) = 8.38, p < .01$; direction $F(1, 14) = 10.96, p < .01$; distance $F(1, 14) = 39.98, p < .01$). Using the Holm method to compare mean values between presentation positions showed the same results as for the participants in their 20 s: no significant difference was found between the front and the floor ($\alpha' = 0.05, n.s.$), but a significant difference was found between the front and the ceiling ($\alpha' = 0.017, n.s.$), and the floor and the ceiling ($\alpha' = 0.025, n.s.$).

4 Discussion and Conclusion

Comparing correct response rates across age groups showed a clear decline in correct responses with age (Fig. 4). However, for signs presented on the ceiling, the significant decrease in correct responses common to both age groups is interesting to note. Also, the correct response rate declined similarly as the arrangement conditions became less orderly in terms of both direction and distance. This suggests that the effect of age uniformly lowers overall performance, and the patterns of responses to presentation

position and orderliness are qualitatively similar. On the other hand, the decrease in correct response rate for signs presented on the floor shown by Watanabe et al. [3] was not replicated in this study regardless of age group, and there was no difference when the sign was presented to the front. The display used in that prior study was a screen projection type, which had lower luminance and contrast than the LED panels used in the present experiment. This difference could have disproportionately influenced the reading of the signs.

In an environment where a wide variety of stimuli and users coexist, effectively attracting attention to particular dynamic signs and conveying information pose significant challenges in the design of dynamic signs for everyday life. The results of this experiment, combined with studies by Watanabe et al. [1] and Itoh et al. [2], which addressed the impact of contrast, presentation time, and animation type on visibility, may be useful for developing guidelines for designing dynamic signs in environments used by people of various ages.

Acknowledgement. This research was conducted with funding from the “FY2023 METI Rule Formation and Market Creation Standardization Promotion Project.”, via Mitsubishi Research Institute, Inc. We thank Shingo Terada of RTC and Hiroko Ojima of AIST for their significant assistance in the recruitment of participants and data acquisition and analysis.

References

1. Watanabe, H., Itoh, N., Ujike, H., Sagawa, K., Sakata, R., Furuhashi, N.: Dynamic signs: Multiple attributes determining visibility. In: Black, N.L., Neumann, W.P., Noy, I. (eds.) *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, Lecture Notes in Networks and Systems, vol. 220, pp. 783–790. Springer, Cham (2021)
2. Itoh, N., Sagawa, K., Watanabe, H., Sakata, R.: Dynamic signs: Appropriate contrast and speed for older adults and low vision. In: Black, N.L., Neumann, W.P., Noy, I. (eds.) *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, Lecture Notes in Networks and Systems, vol. 220, pp. 742–748. Springer, Cham (2021)
3. Watanabe, H., Itoh, N., Ujike, H., Sagawa, K., Sakata, R., Furuhashi, N.: Effect of presentation position on the visibility of dynamic signs. In: *Proceedings of the International Display Workshops (IDW 2022)*, VHF3. The 29th International Display Workshop, Fukuoka (2022)

VR/AR Metaverse (I)



User-Centred Development of an AR Application for the Remote Support of a Heart-Lung Support System

Romy Alm¹(✉), Simon König¹, and Claus Backhaus^{1,2}

¹ University of Applied Sciences Münster, Centre for Ergonomics and Medical Engineering, Bürgerkamp 3, 48565 Steinfurt, Germany

² Technical University Berlin, Institute of Psychology and Ergonomics, Fasanenstraße 1, 10623 Berlin, Germany

Abstract. Intensive care comes with the risk of incidents that can have significant impact on patient health. Extracorporeal membrane oxygenation (ECMO) is important when supporting heart and lung function. However, problems with the device can cause delays in treatment, as perfusionists often need to be on site to troubleshoot. Augmented Reality (AR) provides a solution to bridge these waiting times and increase patient safety. The aim of this study is the user-centered development of an AR-application for improved communication between intensive care nurses and perfusionists via remote assistance. Therefore, intensive care nurses use a Microsoft HoloLens 2 (HL2) and perfusionists a tablet. The application is developed in accordance with VDI guidelines 2221 and 2222. Audio and video transmission as well as holographic markings are identified as system functions using the black box method. The developed solutions are organised using a morphological box and evaluated based on predefined criteria. Unity is used for implementation, so that audio and video transmission can be realised using Unity assets. The graphical user interface is customised for HL2 and tablet, with a focus on functionality and user-friendliness. The application provides remote assistance during ECMO and meets all functional requirements. Technical limitations of HL2, such as difficulties with gesture recognition and hologram placement, have been identified and should be improved. Next steps include the evaluation of user acceptance and functionality in clinical tests as well as iterative application development. Successful deployment could improve work processes in intensive care medicine and reduce the workload on staff.

Keywords: augmented reality · telemedicine · remote support · extracorporeal membrane oxygenation · user-centered development

1 Situation

In intensive care treatment, patient survival is often at risk and a quick and effective treatment is crucial [4, 5]. Extracorporeal membrane oxygenation (ECMO) has an important role to play as it supports heart and lung function. Oxygen-poor, carbon-dioxide-rich

(CO₂) blood is directed into the ECMO, freed of CO₂, enriched with oxygen and pumped back to the patient. Perfusionists are required to solve malfunctions that occur. Usually, they are contacted by phone, as continuous handling of the device is not necessary, which results in waiting times due to physical distance [1].

Augmented reality (AR) has the potential to bridge these waiting times by providing remote assistance for troubleshooting and improving patient safety [4]. AR is the expansion of reality with virtual content that is provided via a display, e.g. through AR glasses such as the HoloLens 2 (HL2) developed by Microsoft Corp (USA). They are controlled by gesture or voice commands so that additional input devices are not required. The integrated development environment Unity is used to develop the application, as it supports the HL2 operating system.

Audio and video transmission are essential for remote assistance. This is complemented by holographic markings and displayed in the user's environment using AR glasses. Remote assistance with AR is already used in various industrial sectors due to its flexibility. Specifically in medicine, AR is primarily tested as a supporting tool in the planning and training of operations, while remote assistance during medical applications has been insufficiently researched so far [10, 11]. Requirements for AR systems in medical technology are strictly regulated in order to prevent health-critical problems as far as possible. These applications thus need to be thoroughly tested before use and tests must prove relevant treatment improvements [5]. The aim of this study is the user-centred development of an AR application for remote assistance during ECMO.

2 Methods

The application is developed in a user-centred process. Requirements analysis and product development are aligned with the guidelines VDI 2221 and VDI 2222 and commence with the definition of the problem and culminate in the development of an ergonomic overall solution [9, 10].

At the start of development, mandatory and facultative system requirements specifying the overall function are collected and recorded in a requirements list. A categorisation of the requirements as well as the allocation of constraints and relevancies allow a greater transparency [7]. For this purpose, references are made to previous studies of expert surveys, conducted in hospitals, on clinical and ergonomic requirements and the generic list of requirements described by Marx and Meister [5].

The overall function is divided into its interactions with the environment using the black box method, while the internal structure is not considered for the time being [6, 9]. The black box is subsequently divided into its most important subfunctions so that the internal structure can be described by a functional structure corresponding to the functional realisation of the system [8]. Based on this functional structure, potential solutions are identified for each subfunction of the system. The potential solutions are visualised using a morphological box. They are assigned line by line to the corresponding subfunctions [9].

By combining the partial solutions, principal overall solutions for the implementation can be derived. In order to identify the most suitable overall solution, the quality of each potential partial solution is evaluated based on the criteria of facultative requirements,

user-friendliness, efficiency and costs. The most suitable partial solution is determined for each subfunction by comparing the partial solutions in a matrix and comparing them with regard to each criterion [2].

Table 1 shows the result when analysing a partial function with three potential solutions concerning a single criterion. The combination of all partial solutions of quality 1 results in the optimal principal overall solution. This is highlighted in the morphological box and finally implemented. Supporting usability tests are carried out with novices during development to evaluate technical and ergonomic requirements.

Table 1. Decision matrix for the quality assessment of the subfunctions

	Partial solution 1	Partial solution 2	Partial solution 3	Σ	Quality
Partial solution 1	–	o	x	1	2
Partial solution 2	x	–	x	2	1
Partial solution 3	o	o	–	0	3

Legend: X better, – equal, O worse

The first stage of the iterative development is completed with the provision of a working solution including graphical user interfaces (GUI).

3 Results

The resulting list of requirements contains 11 mandatory and 8 facultative requirements, which are divided into the categories of functionality, safety and ergonomics.

Using the black box method, the AR-supported video call for remote assistance during the EMCO is defined as the overall function of the application (see Fig. 1). It includes the transmission of video, audio and data signals with the subfunctionalities of signal detection, transmission and output.

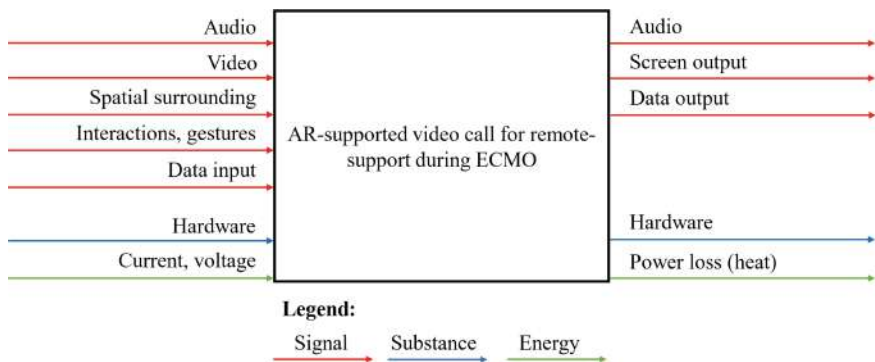


Fig. 1. Black box for the AR supported video call for remote assistance during ECMO.

To implement the overall function, the input signals must comprise audio, video, spatial environment, interactions and data input. The spatial environment, i.e. the recognition of the environment through the system, is exclusively provided by the HL2. The output signals of the system include audio, screen output and data output. To ensure that external signals are recognised and processed correctly, corresponding subprograms are implemented. Depending on the received signals, functions are executed that result in different executions of the outgoing signals. As the application is developed with Unity, solutions can be realised by writing own scripts, integrating external libraries or a combination of both options.

The assessment of the partial solutions results in the implementation of audio and video detection as well as audio and video output using the components integrated in the hardware. Unity assets are used for audio and video switching. Due to the volume and scope of the data to be transmitted, the implementation must be kept to a low level of complexity. The Photon Voice 2 (Photon Engine) asset is used for audio communication and the FMETP Stream 3.0 (Frozen Mist) asset is used for video communication. Interaction is processed via the individualisation and integration of scripts provided by Unity itself, as the response to detected interactions, such as clicks on the tablet, needs to be individualised. Data is transmitted and received through individualised data communication.

The GUI of the HL2 is located in the top left corner of the display to minimise the restriction of the field of vision. There are options to end the call, turn off the microphone and check the connection status. The tablet's GUI has a menu on the left-hand side of the display with the same basic functions (see Fig. 2). Additionally, various colour, marking and removal options can be selected, including customised methods such as marking via freehand drawing, laser pointer and fading in and out of placed markers.

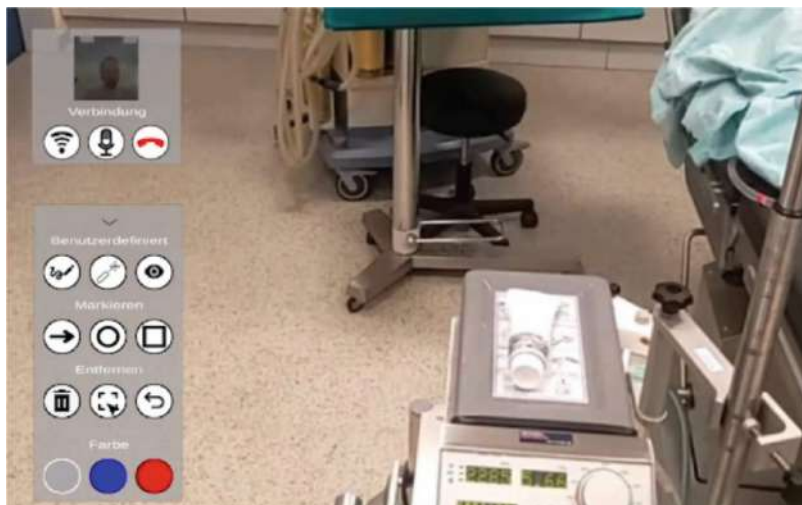


Fig. 2. Expanded interface of the application on the side of the tablet

4 Discussion

The developed solution enables audio and video transmission as well as the placement of holograms for remote assistance during ECMO between the perfusionist (tablet) and intensive care staff (HL2). All mandatory and facultative requirements for functionality and safety are met. During initial usability tests with novices, technical limitations of the HL2 were identified. Interactions via gesture control are occasionally recognised incorrectly, as the sensors and cameras of the HL2 are fixed in their position and cannot capture certain angles of the environment optimally [3]. Materials of surrounding objects can also affect detection. Transparent or dark materials are more difficult to detect, which can cause problems when marking hoses. Furthermore, hologram placement is difficult while moving the HL2. In day-to-day medical practice, this can hardly be avoided due to head movements, which is why precise and fast marking transmission should be prioritised. The combination with voice support makes inaccurate marking less problematic altogether.

The handling of the HL2 interface should be optimised for the use of HL2 in intensive care units. In particular, starting the call needs to be performed quickly so that perfusionists can provide support at a very early stage. This could include simplifying the interface and switching from gesture to voice control or an automatically outgoing call when the glasses are turned on.

During an ongoing iterative development of the application, it should be analysed if the observed technical limitations are acceptable. Usability tests should be carried out with participants corresponding to the real user group in a clinical environment. In that way, the effects of the application usage on work effectiveness and safety can also be investigated.

For the evaluation of the application, a technical introduction can be considered before using the HL2 in practice. This can help to reduce initial operating difficulties through repeated practice, so that the user is already able to use the application effectively in an urgent situation [3]. If the benefits of the AR application are confirmed, this can result in an optimisation of work processes, an increase in the quality of healthcare and a reduction in the workload of clinical staff.

References

1. Al Disi, M., Alsalemi, A., Alhomsy, Y., Bensaali, F., Amira, A., Alinier, G.: Extracorporeal membrane oxygenation simulation-based training: methods, drawbacks and a novel solution. *Perfusion*. **34**(3), 183–194 (2019)
2. Bender, B., Gericke, K.: *Pahl/Beitz Engineering design principles: methods and applications of successful product development*, 9th edn. Springer, Berlin Heidelberg (2021)
3. Brown, D., et al.: *HoloLens Environment Considerations*. <https://learn.microsoft.com/en-us/hololens/hololens-environment-considerations>. Accessed 17 Aug 2023
4. Gocke, P., Elsner, C., Meisheit, B.C., Schneider, H.: *The Digital Hospital: Requirements, Areas Of Action, Implementation*, 1st edn. Medical Scientific Publishing Company, Berlin (2023)
5. Marx, G., Meister, S.: In: Debatin, J.F. (ed.) *The Digital Intensive Care Unit: Areas of Action, Requirements, Implementation*, 1st edn. Medical Scientific Publishing Company, Berlin (2023)

6. Peschges, K.-J.: Developing and Designing in Teams: The Sure Path to Success. Springer Vieweg, Wiesbaden (2015)
7. Reinprecht, S.: Requirements List/Specifications document. <https://netzkonstrukteur.de/konstruktionsmethodik/anforderungsliste>. Accessed 05 Aug 2023
8. VDI 2221: Design of Technical Products and Systems—Model of Product Design. VDI-Guideline 2221 (2019)
9. VDI 2222: Methodic Development of Solution Principles. VDI-Guideline 2222 (1997)
10. Vorraber, W., Gasser, J., Webb, H., Neubacher, D., Url, P.: Assessing augmented reality in production: remote-assisted maintenance with HoloLens. *Procedia CIRP*. **88**, 139–144 (2020)
11. Vostars: Video and Optical See Through Augmented Reality Surgical Systems. <https://www.vostars.eu>. Accessed 06 Aug 2023



A Framework Proposal for Augmented Reality Considering Human Errors in Sequential Tasks

Moacyr Machado Cardoso Júnior¹ (✉) , Henrique Costa Marques¹ ,
and Marcos Vaz de Lima² 

¹ Technological Institute of Aeronautics SJC, São José dos Campos, Brazil
moacyr@ita.br

² Faculdade de Ciências Médicas da Santa Casa de São Paulo, São Paulo, Brazil

Abstract. The framework proposed for augmented reality includes a detailed exploration of a small example within the article. By examining this example, the application of the framework in addressing human errors in sequential tasks will be showcased. The inclusion of this practical illustration further enhances the understanding of how the proposed framework can be effectively implemented in real-world scenarios. This article provides valuable insights into the integration of hierarchical task analysis and the SHERPA technique, offering a comprehensive approach to mitigating human errors and optimizing the performance of augmented reality systems.

Keywords: AR · Human Error · HTA

1 Introduction

Using AR/VR technologies to improve human performance without considering a human-centered perspective can lead to several problems. Firstly, without taking into account the needs, capabilities, and limitations of users, the technology may not align with their cognitive, physical, and sensory abilities. This can result in poor usability, reduced effectiveness, and even potential safety hazards.

Secondly, when AR/VR technologies are implemented without addressing specific problems encountered in industrial settings, they may become a mere novelty or gimmick. Without a clear purpose and problem-solving approach, the technology may not provide significant value or improve productivity. This can lead to wasted resources and a lack of acceptance or adoption by users.

Moreover, the lack of a human-centered perspective can result in a disconnect between the technology and the actual work processes and tasks. AR/VR solutions may not be seamlessly integrated into the existing workflows or may introduce additional complexity, causing disruptions and inefficiencies. Without considering the specific needs and requirements of the industrial setting, the technology may be ill-suited for the context, hindering rather than enhancing performance.

Another potential problem is the risk of overreliance on AR/VR technologies. By solely relying on the technology without considering human factors, there is a danger

of diminishing human skills, decision-making abilities, and situational awareness. This can lead to complacency and a reduction in performance when the technology fails or is not available.

Furthermore, without a human-centered perspective, there may be a lack of consideration for the ethical implications of using AR/VR technologies in industrial settings. Issues such as privacy, data security, and the potential for misuse or abuse of the technology may be disregarded, leading to unintended negative consequences for individuals and organizations.

Inadequate attention to the human-centered perspective can also result in a lack of user engagement and acceptance. Without involving users in the design and development process, their needs and preferences may be overlooked, leading to low user satisfaction and resistance to adoption. This can hinder the successful implementation and utilization of AR/VR technologies in industrial settings.

By employing a hierarchical task analysis approach to AR/VR development, numerous advantages can be realized. Firstly, it enables a systematic understanding of the task structure and hierarchy, allowing for accurate identification of potential areas of improvement. This comprehensive analysis facilitates the mitigation of usability issues and the creation of purpose-driven solutions tailored to address specific problems encountered in industrial settings. Moreover, a hierarchical task analysis approach ensures a seamless integration of AR/VR technologies into existing workflows. By breaking down tasks into smaller, manageable subtasks, potential disruptions are identified and addressed, leading to enhanced efficiency and productivity. This approach also provides insights into the cognitive processes involved, enabling the optimization of AR/VR experiences to align with human capabilities and limitations. Additionally, a hierarchical task analysis approach considers the long-term effects and sustainability of AR/VR technologies. By thoroughly examining the task structure and hierarchy, it allows for the identification of maintenance and compatibility requirements, ensuring the longevity and effectiveness of the implemented solutions. This approach also facilitates the alignment of AR/VR technologies with organizational goals, ensuring that the technology serves as a means to address specific challenges and improve overall performance.

The research question: Are we addressing correctly the problem when designing an AR system as a tool for industry 4.0? What is the main core of its application—prevent human errors?

The framework proposed for augmented reality includes a detailed exploration of a small example within the article. By examining this example, the application of the framework in addressing human errors in sequential tasks will be showcased. The inclusion of this practical illustration further enhances the understanding of how the proposed framework can be effectively implemented in real-world scenarios. This article provides valuable insights into the integration of hierarchical task analysis and the SHERPA technique, offering a comprehensive approach to mitigating human errors and optimizing the performance of augmented reality systems.

2 Theoretical Background

2.1 HTA

Hierarchical Task Analysis (HTA) is a method used to analyze and understand complex tasks by breaking them down into smaller, more manageable subtasks. It was originally described by [1] as a way to study human performance. HTA provides a structured framework for analyzing tasks and identifying the specific actions and decisions required to complete them successfully.

The analysis of tasks is guided by three key principles, as stated by [1]. The first principle emphasizes that a task is considered to consist of an operation, which is defined based on its goal. This goal represents the objective of the system and can be quantified in terms of production units, quality, or other relevant criteria. By defining the operation in terms of its goal, a clear understanding of the desired outcome is established.

The second principle highlights that the operation can be further broken down into sub-operations, each defined by a sub-goal. These sub-goals are measured in real terms by their contribution to the overall system output or goal. This measurement is based on performance standards and criteria, allowing for a tangible assessment of the sub-operations' effectiveness. By breaking down the operation into smaller, measurable sub-operations, a comprehensive analysis of the task can be achieved.

The third principle recognizes the hierarchical relationship between operations and sub-operations. The relationship between them is one of inclusion, indicating that sub-operations are part of the overall operation. While tasks are often structured as procedural sequences, where sub-goals must be attained in a specific order, this is not always the case. The hierarchical nature of the relationship allows for flexibility in the sequence of sub-operations, depending on the specific requirements of the task.

These principles provide a framework for systematically analyzing tasks and understanding the interdependencies between different levels of operations and sub-operations.

One variant of HTA is the tabular format, which provides a structured way to represent the hierarchical structure of the task. In this format, the task hierarchy is represented in a table, with each row representing a subtask and its associated information. The columns in the table represent different attributes of the subtask, such as the action required, the decision to be made, and the information needed [2].

2.2 SHERPA

The Systematic Human Error Reduction and Prediction Approach (SHERPA) method is a valuable tool used in conjunction with Hierarchical Task Analysis (HTA) to identify and mitigate human errors in complex systems. SHERPA aims to systematically analyze and predict potential human errors by considering various factors such as task design, cognitive processes, and environmental influences [3].

In summary, the SHERPA method used in conjunction with HTA provides a human-centered approach to analyze and address potential human errors in AR/VR development. By considering task design, cognitive processes, and environmental influences, designers can optimize the AR/VR experience, minimize errors, and enhance user performance and satisfaction.

3 Case Study

Endoscopic spinal surgery is extremely complex, making the learning curve very long. The rate of complications is high, especially when the surgeon has less experience. In this study we systematically present the main steps of the lumbar recess decompression technique using the endoscopic interlaminar approach as described by Ruetten et al. and point out the crucial points of most common errors. In this way, HTA/SHERPA can greatly contribute to learning and especially to avoiding errors during the procedure. In this study we used the lumbar lateral recess decompression technique, as described by [4], applied in HTA/SHERPA. This procedure consists of a surgical approach to the lumbar spine using an interlaminar endoscopic approach. Briefly, the step by step is as follows: Determine the level to be approached through fluoroscopy (intraoperative radiography); introduction of the endoscope; identification of joints (articular facets); milling of the medial edge of the inferior articular facet of the superior vertebra; “slimming” milling of the medial edge of the superior articular facet of the superior vertebra; medialization of the endoscope and opening of the ligamentum flavum in a lateral direction; removal of the medial edge of the superior articular facet of the inferior vertebra with Kerrinson; observation of neural structures; release of adhesions with gentle manipulation of neural structures; introduction of the bevel into the spinal canal; bevel maneuver, with 180 degree rotation and removal of the root on your shoulder; identification of the intervertebral disc; removal of the herniated disc fragment; return of the bezel to its original position; observation of free neural structures; removal of the endoscope; closure with stitch on the skin; end of the procedure. During surgery, there are several sensitive points where errors are most common, some of which can lead to catastrophic repercussions such as neurological damage or dural tear. It is well known that the majority of these incidents occur due to non-compliance with the surgery step by step as described in the technique. With HTA/SHERPA the aim is to minimize these risks, showing during the training of beginner surgeons the importance of following the sequence originally described.

Figure 1 shows the prototype that is being developed to address the problem of human error identified in HTA/SHERPA analysis.



Fig. 1. Prototype of augmented reality solution.

For SHERPA's consideration we will describe task 3.10 Bezel maneuver on Table 1.

Table 1. SHERPA description of the task “Bezel Maneuver”

Task	Subtask	Error Mode	Error Description	Consequence	Recovery	Probability	Prevention or Mitigation
3.10 Bezel maneuver	3.10.1 Positioning the cannula in the canal	A3	Position the cannula in the “armpit” of the root instead of the shoulder	Neurological damage	Unknown	Unknown	Visualize neural structures after flavectomy
	3.10.2 Cannula rotation	A6	Rotation of the cannula without being able to dislocate the root	Neurological damage	Unknown	Unknown	When realizing that the root has not been dislocated, go back and manipulate the root to release adhesions.
	3.10.3 Shove of neural structures	A4	Pushing the neural structures too much	Neurological damage	Unknown	Unknown	Carry out the maneuver gently and carefully
	3.10.4 Disc exposure and discectomy	A5	Difficulty identifying the disk	Insufficient removal of disc material	Unknown	Unknown	Check with intraoperative radiography, observe free neural structures

References

1. Annett, J., Duncan, K.D., Stammers, R.B., Gray, M.J.: Task Analysis. Department of Employment Training Information Paper 6. HMSO, London (1971)
2. Stanton, N.A.: Hierarchical task analysis: developments, applications, and extensions. *Appl. Ergon.* **37**(1), 55–79 (2006). <https://doi.org/10.1016/j.apergo.2005.06.003>
3. Embrey, D.E.: SHERPA: A Systematic Human Error Reduction And Prediction Approach. American Nuclear Society, United States (1986)
4. Ruetten, S., Komp, M., Merk, H., Godolias, G.: Surgical treatment for lumbar lateral recess stenosis with the full-endoscopic interlaminar approach versus conventional microsurgical technique: A prospective, randomized, controlled study. *J. Neurosurg. Spine.* **10**(5), 476–485 (2009). <https://doi.org/10.3171/2008.7.17634>



Immersive and Interactive Digital Twins to Train Engineering Students

Sylvain Fleury^(✉) 

Arts et Metiers Institute of Technology, LAMPA, 53810 Paris, Change, France
sylvain.fleury@ensam.eu

Abstract. Immersive Digital Twins (IDT) have many advantages for training engineering students in the use of industrial machines. Some of the advantages are practical, such as avoiding the need to buy machines that are very expensive to maintain, saving energy and journeys. There are also advantages in terms of learning, since immersion improves performance for certain types of content. In addition, immersive environments now make it possible to integrate intelligent virtual agents, thanks to the development of language models. These new possibilities will undoubtedly lead to new forms of teaching in engineering schools. Most of the researches are focused on optimizing course situations. However, there is currently a lack of research regarding their impact on the learners independent study, and on the role of the teachers in the use of these devices.

Keywords: Digital Twins · Virtual Reality · Learning

1 Introduction

There are several very different definitions of digital twins [1]. In this work, we refer to digital twins defined as virtual models that interact simulate a physical system and interact with it. Digital twins are more and more used in industry, as well as immersive technologies, defined as technology that offers a high quality or quantity of sensory information to the user [2]. Immersive technologies have three main uses in industry: Process simulation, design and training [3]. Immersive technologies are emerging as a valuable educational resource, although the pedagogical effectiveness of learning in Immersive Virtual Environments (IVE) depend on type of knowledge to acquire, pedagogical method and IVE features. Accessing an engineering qualification requires rigorous training, which imparts a wide spectrum of knowledge and skills. IDT merging digital replica of a machine with the immersive capabilities of virtual reality offers an educational experience that closely mirrors the authentic interaction an engineer would have when operating the actual machine in an industrial setting.

2 Practical Usefulness of IDT for Teaching

IDTs, compared with their physical twins, offer certain practical advantage for teaching. An IDT can be used remotely if a suitable terminal is available, which can reduce learner travel and potentially give students around the world access to training in industrial

equipment. The use of an IDT rather than a physical twin also often results in savings in materials (the material consumed or degraded by the simulated machine), or energy (in the case of foundry systems, for example). Last but not least, the use of IDTs makes it possible to confront machines without endangering either the students or the machine, in cases where certain actions could damage the equipment. This opens the way to safety-oriented training in which teachers can let learners go as far as the accident.

Investing in technological platforms at industrial scale for educational institutions can swiftly escalate into a substantial financial commitment. One strategy to mitigate these expenditures involves collaborative pooling among multiple campuses or educational institutions, and potentially with research laboratories as well. Another problem with physical machine is the lack of availability when there are several users. Having a IVE replication allows to have a lower-cost installation, accessible at all time and from remote locations. While digital technology may not fully substitute the physical environment in the actual production of mechanical parts, conducting preliminary tests to explore various configurations for feasibility validation also serves as a means of saving on costs..

IVE facilitates access to specific work scenarios for individuals with disabilities. While some real workstations may be inaccessible, understanding an operational workstation allows them to better understand upstream tasks such as design or process management. Consequently, IDT presents novel opportunities for individuals previously excluded from such employment prospects.

IVE attracts the curiosity of younger populations, particularly through video games. Engaging young learners in a virtual educational setting depicting industrial work scenarios has the potential to ignite enthusiasm. These lifelike situations can help dispel misconceptions about certain professions, potentially inspiring aspirations for further education in specialized training programs.

3 Pedagogical Effectiveness of IDT

Depending on the type of learning and content, IDTs and more generally IVEs can have a positive, negative or neutral impact compared to non-immersive media. Immersive technologies appear to have potential whenever training requires users to feel a sense of social presence, i.e. the feeling of “being there” with a “real” person [4]. This sensation is supported by stereoscopic 3D, immersion and consistent representation of the people we are with. In this case, individuals’ emotional responses are analog to those measured in real-life situations, which is relevant to train social skills (public speaking...). IVE also has the capacity to generate a feeling of physical presence. A visually realistic environment is an effective way to help learners to familiarize to a specific workplace or to contextualize a learning content.

Through headset and gamepad position tracking, IVE leads users to adopt postures that are more or less equivalent to those they would adopt when carrying out a similar activity in a real situation. It is therefore particularly suitable for learning technical gestures or procedures [5].

Finally, many studies show that course sequences based on immersive technologies often improve course attractiveness and learner satisfaction. For instance, an immersive

anatomy course by [6] did not make any difference to the traditional course regarding anatomy knowledge, but it was perceived as more captivating, enjoyable and motivating for students. In certain instances, enhancing course attractiveness is crucial for trainer satisfaction, student engagement, and institution reputation.

4 Pedagogical Potential of IDT

Over and above the concrete advantages mentioned in the previous section, IVEs generally have potential that could also be of added value for learning. Three-dimensional representations enable the visualization of invisible phenomena within realistic context: Cross-sectional views of machinery in operation or representations of invisible physical phenomena like airflow and heat distribution. IVE offers the added benefit of presenting the visualization of phenomena from a first-person perspective and at a 1:1 scale, thereby facilitating intuitive interaction with the data [7].

IVEs can be populated by virtual agents, which can be connected to language models and act as assistants to the learners. Key attributes of effective pedagogical agents include utilizing a human-like voice instead of a synthetic one [8], exhibiting a range of facial expressions rather than maintaining a neutral expression [9], employing specific gestures [10]. Ideally, these agents have to be designed to be consistent with the learning content [11]. For example, Albert Einstein can give a mathematics lesson, or a historical figure can present his/her own story. They have the capacity to transmit knowledge interactively (by answering questions, for example), but they are also a way of generating social engagement [12].

5 Avenues for Future Research

The role of teachers in the use of IVE topic seems crucial but little studied. In most research on IVEs for learning, the authors do not mention teachers, so we do not know what they do. Do they stand by while learners use the IVE? Do they give instructions? Is this activity part of a longer pedagogical sequence? There are several ways to teach in IVE (See Fig. 1). The artificial situations of experiments generally do not answer these questions. And yet, when it comes to deploying digital training systems, teachers are there, they exist and they cannot be ignored. Thus, it is essential to analyze what they can do and to dedicate specific interfaces to them, enabling them to monitor and possibly interact with learners in IVE.

A second major topic is independent study that is an essential factor in the success of one's studies, but it is rarely studied. The positive effects of IVE on the pleasurable and motivating nature of the activity offer a potential that can be interesting in cases where learners have the freedom to do exercises for practice or not to do so. Having access to an attractive practice tool can make a difference to the amount of practice done.



Fig. 1. Illustrations of two possible classroom situations with an immersive digital twin and a group of students supervised by a teacher.

References

1. Semeraro, C., Lezoche, M., Panetto, H., Dassisti, M.: Digital twin paradigm: a systematic literature review. *Comput. Ind.* **130**, 103469 (2021)
2. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B: Biol. Sci.* **364**(1535), 3549–3557 (2009)
3. Fleury, S., Bernard, F., Paquin, R., Blanchard, P., Richir, S.: Augmented and virtual reality simulation in industry. *Computer*. **56**(12), 55–64 (2023). <https://doi.org/10.1109/MC.2023.32833112>
4. Bondesan, P., Fleury, S., Boisadan, A., Richir, S.: Assessing the educational effectiveness of immersive digital twins in engineering education. *Int. J. Des. Innov. Res.* (2023)
5. Stepan, K., et al.: Immersive virtual reality as a teaching tool for neuroanatomy. *Int. forum allergy rhinol.* **7**(10), 1006–1013 (2017)
6. Christmann, O., et al.: Visualizing the invisible: user-centered design of a system for the visualization of flows and concentrations of particles in the air. *Inf. Vis.* **21**(3), 311–320 (2022)
7. Craig, S.D., Schroeder, N.L.: Reconsidering the voice effect when learning from a virtual human. *Comput. Educ.* **114**, 193–205 (2017). <https://doi.org/10.1016/j.compedu.2017.07.003>
8. Bringula, R.P., Fosgate, I.C.O., Garcia, N.P.R., Yorobe, J.L.M.: Effects of pedagogical agents on students' mathematics performance: a comparison between two versions. *J. Educ. Comput. Res.* **56**(5), 701–722 (2018). <https://doi.org/10.1177/0735633117722494>
9. Davis, R.O., Vincent, J., Park, T.: Reconsidering the voice principle with non-native language speakers. *Comput. Educ.* **140**, 103605 (2019). <https://doi.org/10.1016/j.compedu.2019.103605>
10. Oh, C.S., Bailenson, J.N., Welch, G.F.: A systematic review of social presence: definition, antecedents, and implications. *Frontiers in Robotics and AI*, 114 (2018)
11. Schmidt, S., Bruder, G., Steinicke, F.: Effects of virtual agent and object representation on experiencing exhibited artifacts. *Comput. Graph.* **83**, 1–10 (2019). <https://doi.org/10.1016/j.cag.2019.06.002>
12. Pataranutaporn, P., Leong, J., Danry, V., Lawson, A.P., Maes, P., Sra, M.: AI-generated virtual instructors based on liked or admired people can improve motivation and foster positive emotions for learning. In: 2022 IEEE Frontiers in Education Conference (FIE), pp. 1–9 (2022). <https://doi.org/10.1109/FIE56618.2022.9962478>



Immersion in the Shadows: Physiological Markers of Flow Experience in a Virtual Reality Survival Horror Game

Sanjram Premjit Khanganba^{1,2,3,4,5}  and Anantha Ubaradka^{1,2} 

¹ Human Factors & Applied Cognition Lab, Indian Institute of Technology Indore, Indore 453552, India

sanjrampk@iiti.ac.in

² Discipline of Psychology, Indian Institute of Technology Indore, Indore 453552, India

³ Centre for Electric Vehicles and Intelligent Transport Systems, Indian Institute of Technology Indore, Indore 453552, India

⁴ Department of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Indore 453552, India

⁵ Centre of Futuristic Defence and Space Technologies, Indian Institute of Technology Indore, Indore 453552, India

Abstract. The immersive nature of virtual reality (VR) survival horror games enables the utilization of ‘mediated fright’ as an entertaining medium, providing gamers with a realistic yet safe environment for engagement. Despite the inherent fear and dread these genres evoke, examining their effectiveness in captivating user interest is intriguing. This study investigated the degree of immersion in terms of the flow state achieved while playing a VR survival horror game (Resident Evil 4) compared to a VR casual rhythm game (Beat Saber). Although no statistically significant differences were observed, distinct trends in emotional and cognitive responses were evident. The study also examined heart rate variability (HRV) as a physiological indicator of flow experience, finding that HRV may not consistently reflect flow across different types of gameplays. Additionally, this study explored the potential for cybersickness and cognitive workload to disrupt the flow experience during VR gameplay. It was found that higher cognitive demands in Resident Evil 4 led to increased physical discomfort, likely due to the game’s intense nature. However, no such phenomenon was observed among the Beat Saber group.

Keywords: Flow · Virtual Reality · Survival Horror Games · Heart Rate Variability · Cybersickness · Cognitive Workload

1 Introduction

Over the past few decades, the fascination with frightening media content has been considered paradoxical [1]. This paradox arises from the dual nature of fear and enjoyment. Typically, fear evoked by such media is unpleasant and causes a desire to avoid the source of the threat. However, in the context of ‘mediated fright’—such as movies,

books, and video games—people voluntarily expose themselves to these frightening experiences. This willingness to confront fear indicates that underlying psychological mechanisms transform fear into a source of enjoyment. Several theories have been proposed to explain this paradox. One prominent explanation is the concept of excitation transfer [2]. According to this theory, the physiological arousal induced by fear can intensify subsequent positive emotions, such as relief and joy, once the threat is resolved. This heightened emotional experience can make the initial experience of fear more enjoyable in retrospect.

While there were challenges in manipulating horror resolution within other forms of entertainment media, video games have emerged as a compelling medium to engage audiences with mediated fright through interactive agencies [3]. Unlike non-interactive digital media such as movies, video games allow players to influence the storyline and outcomes, thereby offering unique psychological experiences such as flow and immersion [4]. The interactive elements of the video games enhance the richness of experiencing horror, making the emotional and psychological impact more profound. This interactive nature allows a deeper exploration of the fear and enjoyment paradox, as players actively participate in and shape their virtual experiences.

Among various video gaming platforms, virtual reality (VR) is emerging as a significant medium in the entertainment industry [5]. VR is an immersive technology that mimics and supplements the real world with a synthetic one, creating a simulated user environment. This technology transcends traditional screen-based media by immersing users in three-dimensional spaces where they can look around, move, and interact as if they were in the physical world.

One of the most notable applications of VR has been in the genre of first-person survival horror games (or just horror games). These games are particularly well-suited to the immersive capabilities of VR, as they rely heavily on atmosphere and player immersion to evoke fear. With the widespread reception of games such as *Resident Evil*, VR survival horror games are receiving massive reception from gamers. These games distinguish themselves by offering an unparalleled level of interactivity when compared to other game genres. Unlike traditional horror games on consoles and computers, where players observe from a third-person perspective, VR immerses players directly in the virtual environment. This direct engagement allows players to face in-game threats more personally and intensely. This immersion not only enhances their sense of agency [6]—empowering them to act within the virtual environment—but also amplifies their sense of presence [3], defined as the psychological state of being there within the game.

Despite these insights, the research on the psychological impact of VR horror games is still nascent, particularly when considering comparative analyses across different genres. There is also a notable gap in understanding whether experiences such as flow and immersion are associated with heightened emotional states like fear. Previous studies indicate that VR games can foster a flow state [7], reflecting a peak psychological condition where individuals immerse deeply in an intense focus, a sense of control, and profound enjoyment [8]. However, considering the mediated fright experienced during VR survival horror games, it is interesting to see whether players engaged in such games would experience these psychological states.

In this context, the present study aims to investigate how the immersive quality of VR survival horror games facilitates flow experience during gameplay in comparison to a VR casual game and to what extent cybersickness and cognitive workload arise to disrupt this flow. The study also quantifies the flow experience with physiological parameters such as heart rate variability (HRV). Furthermore, this study assesses the causal relationship between cybersickness and cognitive workload across both gaming genres.

2 Methods

2.1 Participants

A total of 32 ($M_{age} = 21.21$ years; $SD = 2.45$) participants were randomly assigned to one of two groups: One group engaged in playing 'Resident Evil 4,' a survival horror game, while the other group played 'Beat Saber,' a casual rhythm game. The participants were not regular gamers. However, most reported playing video games infrequently on the consoles and computers, averaging 1.25 hours per week. Efforts were made to ensure that no participants had prior experience with either gaming genre. All procedures conformed to the Declaration of Helsinki and were approved by the Institute's Human Ethics Committee. Informed consent was taken from each participant before the experiment.

2.2 Measures

The experiences of fear and joy were assessed using three-item self-reporting measures developed by the authors. The items were measured on a five-point Likert scale. The Flow Short Scale [9] was used to measure the subjective flow experience (ten items measured on a seven-point Likert scale). Immersion in the gaming environment was measured using the Immersive Experience Questionnaire (IEQ) developed by Jannet et al [10]. The cognitive workload was measured with the NASA Task Load Index [11], where the participants rated seven items on a 21-point scale. The Cybersickness in Virtual Reality Questionnaire [12] measured the intensity of cybersickness during the gameplay (six items measured on a seven-point Likert scale). Cardiac activity was recorded with the Polar H10 chest band, validated on par with the ECG gold standard. Heart rate parameters were assessed through the mean of RR intervals, HR, and HRV. HRV is quantified using the square root of the mean squared differences between successive RR intervals (RMSSD) and also with the frequency-domain metrics.

2.3 Procedure

The experiment took place in an acoustically treated laboratory, with participants randomly and equally divided to play either Resident Evil 4 or Beat Saber. Following the provision of consent, each participant wore a Polar H10 chest band. The device was then wirelessly connected to the EliteHRV© app on an Android phone using Bluetooth 4.0. Following a two-minute baseline reading, participants were introduced to the Meta Quest 2 HMD interface. Before commencing gameplay, a tutorial was provided to acquaint

them with the game mechanics. Participants were expected to play the respective VR game for roughly 45 min, although the specific protocols varied based on the nature of each game.

3 Result

Table 1 shows the descriptive scores of the examined variables. Further, an independent samples *t*-test was conducted to assess if there were significant differences between groups across the examined variables. Although the differences were not statistically significant, the results indicated that gamers who played Resident Evil reported higher scores on fear experience ($M_{\text{difference}} = 1.31$), cybersickness ($M_{\text{difference}} = 1.50$), and cognitive workload ($M_{\text{difference}} = 34.47$). Conversely, gamers who played Beat Saber scored higher on the flow index ($M_{\text{difference}} = 3.93$), immersion ($M_{\text{difference}} = 5.75$), and experience of joy ($M_{\text{difference}} = 1.06$).

Table 1. Descriptive scores of the examined variables

Variables	<i>M</i>	SD
Age	21.28	2.45
Fear Factor	4.53	2.43
Enjoyment Factor	12.96	1.90
Cybersickness	11	4.31
Cognitive Workload	264.68	92.42
Flow	48.21	6.93
Immersion	154.18	13.71
HR	98.72	18.61
HRV (RMSSD)	43.18	27.12
HRV (LF)	0.10	0.02
HRV (HF)	0.20	0.07

The study also employed Pearson's product-moment correlation to examine the relationships between the examined variables. A strong positive correlation was found between cognitive workload and cybersickness among Resident Evil players ($r = 0.55$, $df = 30$, $p < 0.01$). Additionally, the experience of joy and immersion were strongly correlated ($r = 0.75$, $df = 30$, $p < 0.01$). Psychological flow showed a strong negative correlation with RMSSD ($r = -0.51$, $df = 30$, $p < 0.01$), although no such association was observed with the frequency domain of HRV. On the other hand, among the players of Beat Saber, the flow was strongly correlated with the experience of joy ($r = 0.55$, $df = 30$, $p < 0.01$). Interestingly, no HRV indices were linked with the flow experience among the Beat Saber group, suggesting that HRV may not consistently reflect flow across different gaming contexts. The ordinary least squares regression analysis showed

that neither cybersickness nor cognitive workload predicted flow in either gameplay scenario. Furthermore, regardless of the gaming group, cognitive workload was found to predict cybersickness among players ($\beta = 0.37$, $SE = 0.008$, $t = 2.23$, $p = 0.03$).

4 Discussion

The current study holds significance as it explored differing user experience patterns between players engaged in VR survival horror games compared to those involved in VR casual rhythm games. The findings revealed that while there were no statistically significant differences, distinct trends emerged in the emotional and cognitive responses associated with both VR gaming experiences. Participants who played the survival horror game (Resident Evil) reported higher measures of fear experience, cybersickness, and cognitive workload. In contrast, those who played the rhythm game (Beat Saber) exhibited higher scores on the flow index, immersion, and experience of joy. These trends suggest that the genre of video games significantly influences the emotional and cognitive engagement experienced by players.

Further analysis using Pearson's product-moment correlation provided insight into the relationships between these variables. The results showed that as cognitive demands increase, players experience greater physical discomfort. This could be due to the intense focus and quick decision-making required in a survival horror context, which might elevate stress and physical symptoms of anxiety. Moreover, a strong positive correlation was noted between the experience of joy and immersion, underscoring that greater emotional engagement in the game is linked to a deeper sense of being 'in the game.' This suggests that immersive environments in gaming are successful in enhancing positive emotional experiences, thereby contributing to a more engaging gameplay experience. Therefore, it can be asserted that players of survival horror games demonstrate mediated fright; although they experience fear, their immersion levels will be strongly linked to feelings of joy.

Furthermore, neither fear nor enjoyment factors were associated with flow. Interestingly, psychological flow in Resident Evil players showed a strong negative correlation with RMSSD, a measure of HRV associated with autonomic flexibility. This might reflect the intense, stress-inducing nature of horror games, which can reduce physiological calmness and increase HRV levels under stress. For Beat Saber players, the flow was strongly correlated with the experience of joy, highlighting that rhythmic and music-driven gameplay can effectively enhance positive emotional states, which are essential for achieving flow. This relationship emphasizes the enjoyment factor as a crucial component of flow in less stressful gaming environments compared to the intense experiences of horror games. The absence of significant correlations between HRV indices and flow among Beat Saber players suggests that HRV may not be a reliable indicator of flow across various types of games, possibly due to the varying intensities and emotional impacts of different gaming genres.

5 Conclusion

This study investigated the differences in user experiences between players engaged in a VR survival horror game and a VR casual rhythm game. While no statistically significant differences were found, clear trends in emotional and cognitive responses were noted. It is also revealed that higher cognitive demands in Resident Evil correlated with increased physical discomfort, possibly due to the intense gameplay. Conversely, in Beat Saber, joy was strongly linked to flow, emphasizing the enjoyment factor in less stressful gaming environments. These findings enhance our understanding of how video game genres impact players' psychological and physiological states, informing game design to improve player engagement and satisfaction.

Acknowledgment. The authors express their gratitude to all the participants who took part in the study. Both authors are equal contributors to this work and are designated as co-first authors. Either of them may have the first appearance in the name sequence in the representation of this work.

References

1. Lin, J.-H.T.: Fear in virtual reality (VR): fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Comput. Hum. Behav.* **72**, 350–361 (2017)
2. Zillmann, D.: Transfer of excitation in emotional behavior. *Soc. Psychophysiol. Sourceb.*, 215–240 (1983)
3. Madsen, K.E.: The differential effects of agency on fear induction using a horror-themed video game. *Comput. Hum. Behav.* **56**, 142–146 (2016)
4. Jin, S.-A.A.: I feel present. Therefore, I experience flow: a structural equation modeling approach to flow and presence in video games. *J. Broadcast. Electron Media.* **55**, 114–136 (2011)
5. Pallavicini, F., et al.: What distinguishes a traditional gaming experience from one in virtual reality? An exploratory study. Presented at the advances in human factors in wearable technologies and game design. In: Proceedings of the AHFE 2017 International Conference on Advances in Human Factors and Wearable Technologies, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA, vol. 8, (2018)
6. Jicol, C., et al.: Effects of emotion and agency on presence in virtual reality. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (2021)
7. Bodzin, A., Junior, R.A., Hammond, T., Anastasio, D.: Investigating engagement and flow with a placed-based immersive virtual reality game. *J. Sci. Educ. Technol.* **30**, 347–360 (2021)
8. Csikszentmihalyi, M., Csikszentmihalyi, M.: *Flow: The psychology of Optimal Experience* Harper & Row New York. (1990)
9. Rheinberg, F., Vollmeyer, R., Engeser, S.: Assessment of flow experiences. *Diagn. Motiv. Self-Concept Tests Trends NF.* **2**, 261–279 (2003)
10. Jennett, C., et al.: Measuring and defining the experience of immersion in games. *Int. J. Hum.-Comput. Stud.* **66**, 641–661 (2008)
11. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Advances in psychology*, pp. 139–183. Elsevier (1988)

12. Kourtesis, P., MacPherson, S.E.: An ecologically valid examination of event-based and time-based prospective memory using immersive virtual reality: the influence of attention, memory, and executive function processes on real-world prospective memory. *Neuropsychol. Rehabil.* **33**, 255–280 (2023)



User Errors in Augmented Reality Assisted Preoperative Planning

Simon König¹(✉)  and Claus Backhaus^{1,2}

¹ University of Applied Sciences Münster, Center for Ergonomics and Medical Engineering,
Bürgerkamp 3 , 48565 Steinfurt, Germany
simon.koenig@fh-muenster.de

² Technical University of Berlin, Institute of Psychology and Ergonomics, Fasanenstraße 1,
10623 Berlin, Germany

Abstract. Augmented reality (AR) overlays virtual information into a user's real environment. AR software already exists for preoperative planning. User errors during planning can influence the surgery and put patients at risk. The aim of the study is to identify user errors of AR software for preoperative planning and to evaluate patient safety. A user test was conducted with 10 physicians who tested the AR software by planning a simulated surgery using the HoloLens 2 (Microsoft Corp., USA). The planning was divided into 21 tasks. An experimenter evaluated the execution of these tasks using a 3-step scheme (good, medium, poor). User errors that led to a medium or poor rating were examined in a Failure Mode and Effects Analysis. Each user error was assigned a risk classification: Unacceptable, conditionally acceptable, and acceptable. User acceptance was recorded using the System Usability Scale and interpreted using a benchmark scale. There were 64 user errors, corresponding to an error rate of 30%. Of these, 8 represented an unacceptable risk, and 16 a conditionally acceptable risk. User acceptance was interpreted as sufficient. User errors causing unacceptable risks occurred mainly in tasks requiring high precision. Therefore, preoperative planning should not be carried out completely in AR at present. AR software should currently only be used to visualize medical images.

Keywords: augmented reality · preoperative planning · user errors · patient safety

1 Introduction

For many patients, surgery is the only curative treatment. However, surgery is always invasive and poses a high risk to the patient. Careful planning is necessary to maximize patient safety and the efficiency of the surgery.

Until now, this has been done on a computer using specialized software, enabling medical images such as X-ray images to be examined and processed. Physicians can draw incisions or measure pathological structures such as tumors. A disadvantage of computer-based planning is the accessibility during the surgery [1]. Software already exists that enables preoperative planning with the help of augmented reality (AR).

AR overlays virtual information into the user’s real environment. The visualization often takes place in the user’s field of view on a display that they wear as glasses (head-mounted display; HMD). The AR is operated by using hand gestures that the HMD detects. This operating concept is unfamiliar to most physicians. However, user errors can lead to incorrect planning, affecting the surgery and putting the patient at risk [2]. User errors of AR software for preoperative planning and their effects on patient safety have rarely been researched up to now.

The aim of the study is to identify user errors of AR software for preoperative planning and to evaluate the impact on patient safety.

2 Methodology

The investigators analyzed user errors in a user test followed by a user survey [3]. Ten physicians (4 surgeons, 3 anesthetists, 3 internists) participated in this study. At least five days before the test, the participants received a standardized 90-minute briefing on usage from the manufacturer. In the user test, the participants planned a fictitious surgery using the AR software operated with the HMD HoloLens 2 (Microsoft Corp.). A simulation OR served as the test environment.

The planning of the surgery was divided into 21 tasks [4] (see Table 3). A supervisor observed the execution of the tasks by mirroring the HoloLens 2 screen on a laptop and capturing a video of the test environment using a GoPro Hero 5 (GoPro Inc., USA). During the user test, the supervisor evaluated the participants’ ability to perform the tasks using a 3-stage evaluation scheme [5] (see Table 1).

Table 1. Criteria for evaluating the task.

Rating	Description
Good	- Rapid and user error free execution
Medium	- User error corrected on their own - Execution with delay
Poor	- User error not corrected - Execution with support - Task not executed

Tasks rated as medium or poor were examined in a Failure Mode and Effects Analysis, where each user error was assigned a damage severity and a probability of occurrence [6]. Due to the small sample size, only the severity of damage was used to assess the risk, which was classified as unacceptable, conditionally acceptable, or acceptable [7] (see Table 2).

After completing all tasks, user acceptance was recorded using the System Usability Scale (SUS) [8]. The participants rated their agreement with 10 items on a 5-point Likert scale. Summing up all points and multiplying by 2.5 resulted in a SUS score between

Table 2. Assessment of the risk of user errors based on the severity of damage.

Severity	Description	Risk
High (III)	- Direct impact on the patient - Change in planning results - Loss of patients or planning data	Not acceptable
Moderate (II)	- Indirect impact on the patient - Planning cannot be completed	Conditionally acceptable
Low (I)	- No impact on the patient - Delay or interruption of planning	Acceptable

0 (negative) and 100 (good). The SUS score was interpreted according to Bangor et al. [9] (see Fig. 1).

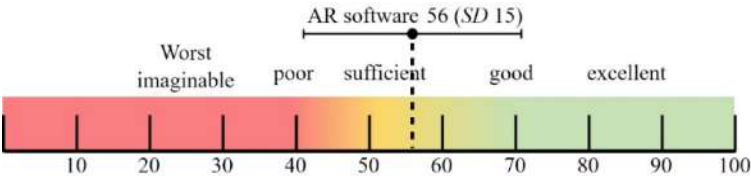


Fig. 1. SUS score of the AR software and benchmark scale.

3 Results

A total of 64 user errors occurred in the user test, corresponding to an error rate of 30%. Table 3 shows the assessment of the participants’ competence and the severity of the observed user errors.

Of the user errors that occurred, 8 represented an unacceptable risk (III), 16 a conditionally acceptable risk (II), and 40 an acceptable risk (I) for the patient. In 5 of the 10 participants, at least one user error occurred that represented an unacceptable risk. Table 4 lists the user errors with unacceptable and conditionally acceptable risks and shows their consequences.

In the user survey, the AR software achieved a SUS score of 56 (± 15), corresponding to sufficient user acceptance (see Fig. 1).

Table 3. Evaluation of the competencies (green, yellow, red) and the severity of damage of the observed user errors (I, II, III).

No.	Task	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
1	Start AR-Software										
2	Select user and log in			I			I				
3	Select patient					I	II				
4	Select case		II								
5	Select medical image		II	I	II		II			I	I
6	Adjust opacity	I	I	II	I	II			I		I
7	Adjust brightness										
8	Adjust contrast	I			I						
9	Adjust threshold value										
10	Select image Slicer				I				II		
11	Examine medical image				I				I		
12	Detect tumor	I	I		III						
13	Create marker		I	I	I			I		I	II
14	Mark tumor	I							I	III	II
15	Select drawing tool	II		II	II	I	II				II
16	Draw in planned cur	I	I	I	III				II		I
17	Select measuring tool										I
18	Measure tumor			III	III				III		III
19	Open planning menu	I					I	I		I	I
20	Save planning										III
21	Log out			I	I		I		I		

Table 4. User errors with not acceptable (III) and conditionally acceptable risks (II) and user error consequences.

No.	Participant	User error	User error consequences
Not acceptable Risk (III)			
12	4	Slicer position set incorrectly	Tumor cannot be viewed completely
14	9	Hologram shifted	Tumor incorrectly marked
16	4		
18	3, 4, 8, 10	Tumor measured imprecisely	Tumor size incorrectly determined

(continued)

Table 4. (continued)

No.	Participant	User error	User error consequences
20	10	Planning is not saved	Loss of planning
Conditionally acceptable Risk (II)			
3	6	Operation takes place on the wrong level (behind the control panel)	Operation of the user interface not possible
4	2		
5	2, 4, 6		
6	3	Incorrect rendering mode activated	Image parameters cannot be set
6	5	Gesture for opening the menu unknown	Menu cannot be opened
10	8	Two image slicers selected at the same time	Generated slices cannot be interpreted
13	10	Operation takes place on the wrong level (behind the control panel)	Operation of the user interface not possible
14	10	Hologram not found	Tumor cannot be marked
15	1, 3, 4, 6, 10	Drawing tool not found	The incision cannot be drawn
16	8	Drawing tool cannot be operated	The incision cannot be drawn

4 Discussion

The study shows that preoperative planning is possible using AR. However, the high error rate indicates usability problems. User errors that represent an unacceptable risk occurred mainly in tasks requiring high precision. In this user test, 6 out of 8 user errors with an unacceptable risk occurred when marking or measuring the tumor or drawing the course of the incision.

User acceptance can be rated as sufficient despite the high error rate, suggesting a high intention to use on the part of the participants [10]. The AR software should therefore be optimized in a user-centered development process [11]. Alternative operating concepts, such as combined gesture and gaze control, should also be investigated. An HMD that enables this operating concept is the Vision Pro (Apple Inc., USA).

5 Conclusion

User errors occur that have an unacceptable impact on patient safety. Preoperative planning should therefore not be carried out completely with AR software at present. We recommend using it only to visualize medical images or to view information during surgery.

References

1. Fardoun, H.M., Al-Ghamdi, A.S., Ciprés, A.P.: Improving surgery operations by means of cloud systems and distributed user interfaces. In: DUI '14: Proceedings of the 2014 Workshop on Distributed User Interfaces and Multimodal Interaction, pp. 31–36 (2014)
2. Atesok, K., Galos, D., Jazrawi, L.M., Egol, K.: A.: Preoperative planning in orthopedic surgery. *Curr. Pract. Evol. Appl., Bull. Hosp. Jt. Dis.* **73**(4), 257–268 (2015)
3. German Institute for Standardization: DIN EN 62366–1: Medical Devices—Part 1: Application of usability engineering to medical devices. Beuth, Berlin (2021)
4. Diaper, D., Stanton, N.: *The Handbook of Task Analysis for Human-Computer Interaction*. (2003)
5. German Institute for Standardization: DIN EN 614–1: Safety of machinery—Ergonomic design principles—Part 1: Terminology and general principles. Beuth, Berlin (2009)
6. International Standard Organization: IEC 60812: Analysis techniques for system reliability—Procedure for failure mode and effects analysis. Beuth, Berlin (2006)
7. International Standard Organization: ISO 14971: Medical Devices—Application of risk management to medical devices. Beuth, Berlin (2019)
8. Brooke, J.: SUS: a quick and dirty usability scale. *Usability Eval. Ind.* **189**, 4–7 (1996)
9. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective scale. *J. Usability Stud.* **4**, 114–123 (2009)
10. Tiawo, A.A., Dowe, A.G.: The theory of user acceptance and use of technology (UTAUT): a meth-analysis review of empirical findings. *J. Theor. Appl. Inf. Technol.* **49**(1), 48–58 (2013)
11. German Institute for Standardization: DIN EN ISO 9241-210: Ergonomics of human-system interaction—Part 210: Human-centered design for interactive systems. Beuth, Berlin (2020)



Optimizing Button Size to Enhance User Experience for Augmented Reality Interactions in Mobile Scenarios

Hojeong Lee^(✉) , Yeahyoung Jeon , Keisuke Kudo , Paul Ssemakula ,
Seonghyeok Park , and Shuping Xiong 

Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu,
Daejeon, Republic of Korea

{leeho, y.jeon, graywizard01, semaxpaul, seonghyeok,
shupingx}@kaist.ac.kr

Abstract. The anticipated widespread use of augmented reality (AR) head-mounted displays (HMDs) as daily mobile devices underscores the importance of button size, a critical factor influencing user interface effectiveness and task performance. While most mobile device interactions occur during walking, prior research on button touch has primarily been conducted in stationary VR or AR HMD environments, often neglecting the dynamic context of mobile use. This study addresses this gap by optimizing button sizes in AR user interfaces specifically for walking conditions. The objective is to identify an ideal button size that maximizes task performance accuracy while minimizing visual obstruction. Utilizing a randomized complete block design, participants interacted with buttons of various sizes (15–55 mm) in an AR setting. Findings indicate that larger buttons enhance task performance though excessively large buttons can cause visual obstruction and dizziness. The optimal size was identified to be 45 mm, balancing design and user experience. This research contributes to human-computer interaction in AR, emphasizing user-centered design in emerging technologies.

Keywords: Augmented Reality · Human-Computer Interaction · Button Size · User Interface Design · User Experience

1 Introduction

Virtual Reality (VR) and Augmented Reality (AR) are rapidly advancing technologies poised to impact daily life. While VR requires users to disconnect from the real world and interact within preset boundaries and computer-generated scenes, AR overlays virtual images onto the real world via head-mounted displays (HMDs), enhancing real-world experiences [1]. As AR HMD technology becomes more portable and flexible, these devices are expected to become as ubiquitous as smartphones, potentially even replacing them. Using AR HMDs as daily communication devices will require users to interact with their surroundings and virtual objects simultaneously in mobile scenarios.

The size of buttons on smartphones and extended reality (XR) devices is one of the main factors of user interface (UI) and significantly affects performance accuracy and usability. Previous research has examined the optimal button sizes for smartphones and VR [2, 3, 6], as well as factors affecting AR button-pressing tasks [4, 5]. Although mobile devices are usually used while walking, previous research on button touch was conducted mainly within stationary VR or AR HMD environments, and the interactive aspect within mobile scenarios has been often overlooked and remained underexplored. Therefore, it is essential to reevaluate the UI interaction in AR settings and assess its usability while walking in the real world, taking the surrounding environment into account. This study aims to optimize button sizes in AR user interfaces specifically for walking conditions.

2 Methods

2.1 Experimental Design

The experiment adopted a randomized complete block design (RCBD), which allowed for the assessment of button sizes (15 mm, 25 mm, 35 mm, 45 mm, and 55 mm) as five treatment levels. Sizes 15 to 45 mm were based on previous studies in static and mobile states [3, 6, 7] and we added 55 mm for walking conditions navigating the real environment. Each participant experienced all treatments in random order, with three trials. Each trial included five sessions presenting all button sizes in randomized sequences to minimize learning or fatigue effects.

For performance evaluation, task completion time, deviation range, and wrong button selections were measured. Unity tracked button presses via index fingertip coordinates. Deviation range was defined as the average Euclidean distance of attempts within 4 cm of the final successful presses [8]. Task completion time was the time from receiving the instruction in the pop-up message to finally pressing the button. For user experience, participants reported their preferences for different button sizes.

2.2 Participants

Ten right-handed adults, 5 males and 5 females (age 21–29, $M = 23.5$, $SD = 2.38$) were recruited. All participants had normal walking capabilities as well as normal eyesight. Participants were asked to fill in the consent form and write down some basic data.

2.3 Experimental Setting

The experiment employed a Microsoft HoloLens 2 and Leap Motion 1 as a robust hand-tracking device. The task scene and data collection system were developed using Unity 3D (2019.3.7f1) and displayed to the HoloLens via a wireless connection.

We designed the scene with 5 buttons displaying different icons aligned in a row at the bottom of the view to optimize participants' walking visibility (Fig. 1). The order of icons was randomized for each attempt. Buttons were positioned 450 mm from the users' eyes based on adult arm length averages [9]. All buttons shared fixed center positions within their size groups to maintain consistency. Task instructions were delivered through



Fig. 1. AR scene with buttons of different sizes: 15, 25, 35, 45, and 55 mm at the distance of 450 mm.



Fig. 2. Participant conducting the button pressing task along with the guideline.

pop-up messages above eye level to simulate realistic alerts. Pop-ups were triggered by the experimenter at random times within the Unity scene.

For the experimental environment, we drew the J-shaped guideline path with 1.5 m of curved path and 2 m of straight path on the floor to represent real-like walking conditions where participants should follow the path. Participants walked following the path while conducting the task. They were instructed to walk at their preferred pace with speed adjustments made if necessary. We asked participants to use only the index finger to collect accurate hand trajectory (Fig. 2).

2.4 Experimental Procedure

First, we calibrated the positions of the virtual and real hands to match well. Participants underwent an initial familiarization with the AR environment and task. Once acclimated, they positioned themselves at the starting point of a walking path. Following the experimenter's verbal instructions, participants were tasked with pressing the correct button indicated by a pop-up prompt. Each session featured five different button sizes in random order, tested across three repetitions (15 trials total). After pressing the button, participants returned to the starting point. Following each trial, participants provided subjective ratings for their button preference. The experiment ended with an interview to gather feedback on the overall experience and UI.

2.5 Statistical Analysis

A repeated-measures analysis of variance (ANOVA) was performed to analyze the main effect of button size on task completion time and deviation range by using Python, and Minitab 21.4.2. Post-hoc analysis was conducted using the Least Significant Difference (LSD). For participants' preferences of best and worst button size and the number of wrong button selections, frequency analysis and Chi-square test were conducted.

3 Results

The results of three performance metrics (task completion time, deviation range, and number of wrong button selections) and subject preferences (most and least preferred sizes) are summarized in Figs. 3 and 4. Details are described as follows.

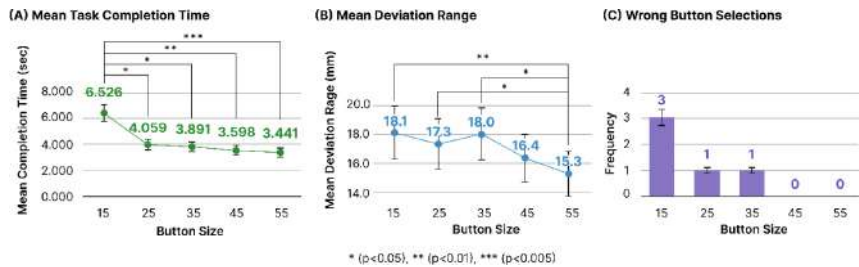


Fig. 3. Performance metrics under different button sizes (A) task completion time; (B) deviation range; (C) frequency of wrong button selections. (* significant between-group differences).

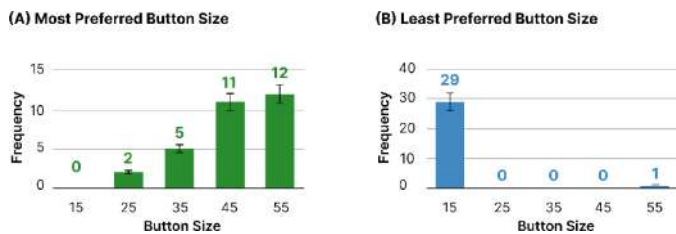


Fig. 4. Frequency distribution of (A) Most preferred button size, (B) Least preferred size.

3.1 Task Completion Time

The ANOVA test revealed a significant effect of button size on task completion time ($F(4, 36) = 3.24, p = .023$). The LSD post-hoc test (Fig. 3 (A)) shows that the smallest button size, 15 mm was statistically significant and different from button sizes 25 mm ($p = .018$), 35 mm ($p = .012$), 45 mm ($p = .006$) and 55 mm ($p = .004$). However, no significant differences were found between other button sizes.

3.2 Deviation Range

The button size had a statistically significant impact on the deviation range ($F(4, 36) = 2.85, p = .038$). The LSD post-hoc test (Fig. 3 (B)) indicates that 55 mm size was statistically significant and different from button sizes 15 mm ($p = .008$), 25 mm ($p = .048$), and 35 mm ($p = .010$). The largest button size (55 mm) resulted in a significantly less deviation compared to the smaller button sizes.

3.3 Wrong Button Selections

A frequency analysis of wrong button selections revealed that the smallest button size (15 mm) had the highest frequency of wrong selections as shown in Fig. 3 (C). This pattern suggests that smaller buttons could potentially increase the frequency of wrong button selections in practice.

3.4 User Preference

As shown in Fig. 4, the participants exhibited a clear preference trend, favoring the larger button sizes (45 mm and 55 mm) over the smaller ones, especially the smallest one (15 mm). These preferences are consistent with the objective performance data. Chi-square analysis revealed a significant difference ($\chi^2(4, 60) = 8.9, p < .001$) in participants' preference ratings for various button sizes. This indicates that certain button sizes are favored over others, reflecting the consistency of user preferences.

4 Discussion

This AR interface design study explored the optimization of UI button sizes to maximize task performance accuracy and enhance user experience while walking in the augmented real world. This focus is particularly relevant given the increasingly mobile usage contexts for AR technologies. Results of task completion time and deviation range indicated an inverse relationship between button size and performance metrics. Results showed that both task completion time and deviation range were significantly affected by button size, with smaller buttons leading to decreased accuracy and longer completion times. This is consistent with the concept that the smaller button size (e.g., 15 mm) may be harder to see and press with precision, leading to skewed results in user preference ratings as it was selected as the worst button in most cases.

On the other hand, the post-experiment interview revealed that the largest button size (55 mm) was too big and occasionally obstructed the user's view, causing fatigue and dizziness. Some participants also experienced dizziness when focusing on the smallest buttons, highlighting the need for future studies to consider measuring mental workload and cybersickness [5].

Prior research in static VR environments identified 20 to 25 mm as the optimal button size [3, 6], 32 mm in static AR environment [9], while the study in mobile VR found 40 mm to be ideal [7]. However, the mobile VR study [7] was conducted on treadmills, which did not require users to navigate real-world surroundings. Our experiment identified 45 mm as the optimal button size for AR interfaces, as it balanced visibility and usability without overwhelming the user's field of view or causing discomfort. This size correlated with high user preference, no wrong button selections, and good performance metrics. These findings align with interaction design principles that emphasize intuitive interactions for efficient navigation in AR settings [10].

This study contributes to the body of knowledge on AR UI design for mobile scenarios in real-world environments, offering evidence-based design recommendations. Our findings support the development of UI guidelines to prevent potential accidents caused

by UIs obstructing the user's view or requiring excessive attention, which can hinder situational awareness while walking. Continuous adaptation of interface design principles is crucial for developing functional, comfortable, and immersive AR experiences.

Future research should involve a larger and more diverse participant pool to ensure more reliable conclusions and to explore various view distances, spacing, and layouts to enhance generalization. This user-centered approach is essential for designing AR interactions that accommodate the diverse conditions of mobile usage.

5 Conclusion

This study explored the optimal button size that maximizes task performance accuracy while minimizing visual obstruction during mobile walking interactions in AR. The results showed that a button size of 45 mm in AR interfaces was optimal, striking a balance between design and user experience, ensuring accurate interaction without overwhelming the user's view or causing discomfort. These findings underscore the importance of user-centered design in AR technology, offering evidence-based design recommendations and valuable insights for future AR interface development.

Acknowledgments. This work was supported by the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning under the Basic Science Research Program (NRF2020R1F1A1048510, NRF2022R1F1A1061045).

References

1. Qiao, X., Ren, P., Dustdar, S., Liu, L., Ma, H., Chen, J.: Web AR: a promising future for mobile augmented reality—State of the art, challenges, and insights. *Proc. IEEE*. **107**(4), 651–666 (2019)
2. Choe, M., Park, J., Kim, H.K.: Effect of target size, location, and input method on interaction in immersive virtual reality. *Appl. Sci.* **11**, 9846 (2021)
3. Park, K., Kim, D., Han, S.H.: Usability of the size, spacing, and operation method of virtual buttons with virtual hand on head-mounted displays. *Int. J. Ind. Ergon.* **76**, 102939 (2020)
4. Kim, W., Xiong, S.: Pseudo-haptic button for improving user experience of midair interaction in VR. *Int. J. Hum.-Comput. Stud.* **168**, 102907 (2022a)
5. Yamin, P.A., Park, J., Kim, H.K., Hussain, M.: Effects of button colour and background on augmented reality interfaces. *Behav. Inform. Technol.* **43**, 1–14 (2023)
6. Zhou, X., Guo, Y., Jia, L., Jin, Y., Li, H., Xue, C.: A study of button size for virtual hand interaction in virtual environments based on clicking performance. *Multimed. Tools Appl.* **82**(10), 15903–15918 (2023)
7. Wang, C.H., Hsieh, C.Y., Yu, N.H., Bianchi, A., Chan, L.: HapticSphere: physical support to enable precision touch interaction in mobile mixed-reality. In: *In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 331–339. IEEE (2019)
8. Kim, W., Xiong, S.: Viewfindervr: configurable viewfinder for selection of distant objects in VR. *Virtual Reality*. **26**(4), 1573–1592 (2022b)
9. Microsoft (2023). Microsoft HoloLens—Microsoft Learn
10. Thakur, A. (2023). The Emergence of Virtual Reality and Augmented Reality in UI/UX Design



Examining the Influence of Metacognitive Monitoring Feedback in an AR Learning Environment

Sara Mostowfi¹(✉), Jung Hyup Kim¹, Ching-Yun Yu², Siddarth Mohanty¹,
Fang Wang³, Danielle Oprean⁴, and Kangwon Seo¹

¹ Department of Industrial and Systems Engineering, University of Missouri, Columbia,
MO 65201, USA

sara.mostowfi@mail.missouri.edu, {kijung, smdqv,
seoka}@missouri.edu

² Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO 65201, USA

cytbn@missouri.edu

³ Department of Engineering and Information Technology, University of Missouri, Columbia,
MO 65201, USA

wangfan@missouri.edu

⁴ Info Science and Learning Technology Department, University of Missouri, Columbia, MO,
United States

opreand@missouri.edu

Abstract. In the context of student learning, investigating effective feedback mechanisms within augmented reality (AR) learning systems is crucial for better understanding and optimizing study behaviors. This study examines the influence of metacognitive monitoring feedback within an AR setting. Our hypothesis suggests that regularly providing students with feedback on their metacognitive monitoring within an AR learning environment has a beneficial effect on their metacognitive state. The results of the study confirm that frequent exposure to such feedback significantly improves scores on the Metacognitive Awareness Inventory. Essentially, there was a marked increase in the inventory scores of participants who received ongoing feedback, in contrast to those who were only given metacognitive monitoring feedback once after the lecture, particularly in the areas of planning, monitoring comprehension, and debugging strategies. This enhancement is achieved by influencing student calibration by directly impacting their metacognitive state.

Keywords: Augmented Reality · Metacognition · Engineering Learning

1 Introduction

According to Flavell [1], metacognition involves managing and controlling one's own thinking processes using strategies like planning, monitoring, and adjusting Norman and Furnes [2]. Metacognition helps individuals understand how they learn, reflect on

their thinking patterns, and guide their personal and professional development [3–5]. It includes organizing approaches for learning, problem-solving techniques, and self-evaluation to correct mistakes and track progress [6, 7]. Metacognitive awareness, defined as regulating one's own thinking, is linked to better learning outcomes [8]. However, in augmented reality (AR) learning environments, the opportunity to cultivate these evaluative skills may be hindered by reduced interaction and feedback. Addressing this issue necessitates an investigation into effective feedback strategies within AR learning platforms to comprehend better and assess students' study practices. Consequently, this research aims to investigate the impact of feedback on metacognitive monitoring in an AR context.

2 Methodology

For this study, we developed a total of fifteen 3D scenes using the Unity Game Engine, with seven modules in lecture 1 and eight modules in lecture 2 [9]. Lecture 1 covers the introduction to biomechanics and shows students how to draw force and moment on different body segments (comprised of an introduction to basic biomechanics knowledge including definitions, concepts of force and moment, static equilibrium, multiple link examples, and center of mass); and lecture 2 which is more challenging than lecture 1 (comprised of a review of lecture 1, free body diagrams on hand, upper arm, lower arm, and trunk segments). Each scene includes a large semicircular blackboard comprising five interconnected panels (as depicted in Fig. 1 Right). This design offers users a convenient and immersive experience in the virtual lecture space when they stand at the center of the scene and face forward. These five panels display various elements, including figures, human avatars, formula calculations, problem statements, and tables of figures (refer to Fig. 1 Left). In the first experiment (experiment #1), participants were required to respond to a question concerning the material they had studied, as well as to rate their confidence in their answers. However, no immediate feedback was provided; they only received a single, comprehensive feedback session after the lecture concluded [9]. In contrast, the second experiment (experiment #2) provided participants with immediate metacognitive monitoring feedback, detailing both their performance and confidence levels, but did not offer the summary feedback at the lecture's end. To evaluate changes in their Metacognitive Awareness Inventory (MAI) scores, participants in both experiments completed the MAI questionnaire following the lecture. For the duration of the experiment, participants were equipped with two devices: A HoloLens for accessing AR learning content and an NFER tag to track their physical location within the setup, which comprised seven distinct learning spots for each lecture [10]. Each spot is a signed place on the ground which has been determined for each module. It means when a student is moving on the marked sign with the table which has an NFER tag, the AR scene will be shown based on the programming connected to the tag on the table (see Fig. 2). The entire experimental process, including material review and the final biomechanics practice on the computer, was designed to last approximately 35–40 min.

The only difference between experiments #1 and #2 is how the question is presented to students. In experiment #2, student confidence level and performance are evaluated on a monitor screen instead of paper. In the metacognitive monitoring feedback, students can see their performance and confidence level results after each module;

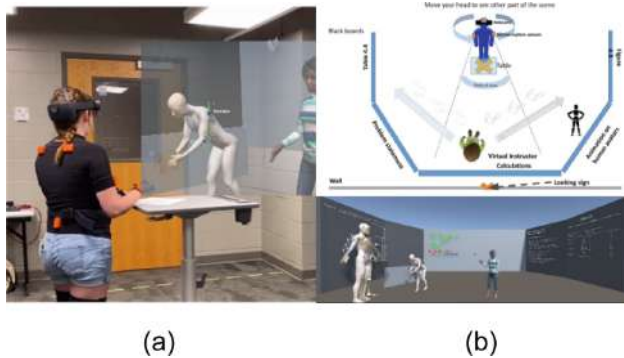


Fig. 1. Description of AR learning system (a): student is watching a scene in the AR environment, (b): arrangement of 3D objects in an AR scene showing the system setup.

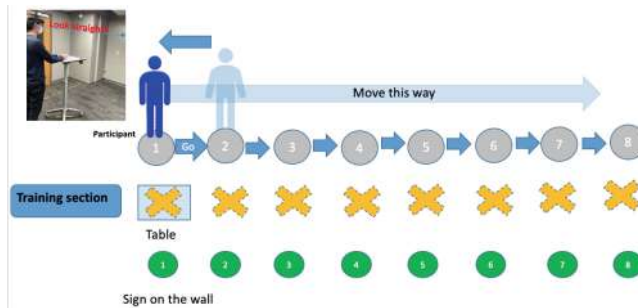


Fig. 2. The learning process during each lecture

however, in the paper-based, students can access their results at the end of each lecture. The metacognitive monitoring feedback is associated with the student's capacity to accurately assess their current metacognitive state, discerning between over-confidence and under-confidence. This feedback mechanism enhances student learning by fostering self-awareness and regulation. The metacognitive monitoring feedback comprises two distinct components (as illustrated in Fig. 3).

- 1) Student's response and the answer to the exercise problem (Left).
- 2) Visual graphs of the confidence and the performance level of the exercise problem (Right).

The first component of the metacognitive monitoring feedback encompasses students' responses to various quizzes and the retrieval of the correct answers for them. The second component involves visual representations, such as graphs, displaying the retrospective confidence judgments (RCJ) scores alongside the scores for the quiz outcome. The RCJ is one of the metacognitive prompting techniques to self-evaluate the learner's confidence levels before knowing actual test scores. For the retrospective confidence judgment, students were supposed to answer: "How well do you think have you performed question 1? (1%—low confidence level, 100%—high confidence level)" [11]. These graphical

representations highlight disparities between students' RCJ and their actual competency in the assigned lab activities. When a student is over-confident (RCJ > competency), the RCJ graph appears larger than the outcome graph. In contrast, an under-confident student (RCJ < competency) exhibits an RCJ graph smaller than the outcome graph. To help students rectify their confidence levels in line with their competency, an asymmetrical triangular shape is employed as a visual aid in both graphs.



Fig. 3. Example of Metacognitive Monitoring Feedback showing the performance and confidence level

3 Result

The paired sample t-test was done between lecture #1 and lecture #2 between MAI subcategory factors for both experiments #1 and #2 separately to see how metacognitive monitoring has changed between two lectures in both experiments. The findings (see Tables 1 and 2) showed that in experiment #1, there was a noticeable MAI difference in Evaluation between lectures #1 and #2. In experiment #2, significant variations were observed in the MAI scores related to Conditional Knowledge, Planning, Information Management Strategies, and Comprehension Monitoring.

Table 1. MAI Comparisons between lectures #1 and 2 in Experiment #1

MAI CATEGORY	LECTURE	N	MEAN	SD ERROR	T-RATIO	P-VALUE
Declarative knowledge	1	31	0.772339	0.02883	0.657439	0.5188
	2	31	0.792500			
Procedural Knowledge	1	31	0.905	0.02672	0.280652	0.7820
	2	31	0.9125			
Conditional Knowledge	1	31	0.790323	0.05871	0.9815	0.3258
	2	31	0.872581			
Planning	1	31	0.689658	0.04720	0.0819	0.7757
	2	31	0.670552			

(continued)

Table 1. (continued)

MAI CATEGORY	LECTURE	N	MEAN	SD ERROR	T-RATIO	P-VALUE
Information Management Strategies	1	31	0.804839	0.03500	0.8205	0.3687
	2	31	0.760000			
Comprehension Monitoring	1	31	0.721339	0.04135	0.1188	0.7315
	2	31	0.701184			
Debugging Strategies	1	31	0.806452	0.02898	0.0128	0.9101
	2	31	0.801806			
Evaluation	1	31	0.650294	0.04566	4.01138	0.0497
	2	31	0.779665			

Table 2. MAI Comparisons between lectures #1 and 2 in Experiment #2

MAI CATEGORY	LECTURE	N	MEAN	SD ERROR	T-RATIO	P-VALUE
Declarative knowledge	1	20	0.8205	0.0289	0.6574	0.5188
	2	20	0.8395			
Procedural Knowledge	1	20	0.905	0.02672	0.28065	0.7820
	2	20	0.9125			
Conditional Knowledge	1	20	0.8685	0.02602	2.36366	0.0289
	2	20	0.93			
Planning	1	20	0.6282	0.06088	3.41003	0.0029
	2	20	0.8358			
Information Management Strategies	1	20	0.67315	0.07077	2.601427	0.0175
	2	20	0.85725			
Comprehension Monitoring	1	20	0.755	0.02563	3.667142	0.0016
	2	20	0.849			
Debugging Strategies	1	20	0.9035	0.04951	0.33326	0.7426
	2	20	0.92			
Evaluation	1	20	0.6195	0.04576	1.9034	0.0722
	2	20	0.7066			

4 Discussion

This research investigated how metacognitive monitoring feedback impacts students' metacognitive state within the AR learning environment. The comparison between lectures #1 and #2 for experiment #1 showed that no significant MAI score improvement between the two lectures (except the evaluation). In experiment #1, students were

informed of their confidence level and overall performance across all modules in the lecture only once at the end of each lecture. Providing feedback only at the end of the lecture series allowed students to evaluate their learning strategies over an extended period. This broad perspective could enhance their ability to assess their learning behaviors thoroughly. Since students are required to monitor their understanding and performance individually, this setting may promote deeper self-assessment practices and strengthen their evaluative skills.

On the other hand, for the results from experiment #2, the comparison between lectures 1 and 2 showed improvements in areas such as Conditional Knowledge, Planning, Information Management Strategies, and Comprehension Monitoring, which are all parts of the regulation of cognition in the MAI. Regulation of cognition involves employing techniques for efficient information management throughout the learning process as well as proactive planning and monitoring activities before engaging in educational activities [12]. With regular access to metacognitive monitoring feedback, students could adjust their learning strategies for each module in real time. This means they might change their approach based on the effectiveness of their current methods, potentially enhancing skills such as information management and planning. In this study, metacognitive monitoring feedback in AR learning environments significantly enhances students' ability to effectively plan, monitor, and adjust their learning strategies, leading to improved comprehension and metacognitive awareness. This research highlights the importance of feedback timing and method, suggesting that continuous, contextually relevant feedback is essential for maximizing learning potential and metacognitive development within immersive educational settings.

Acknowledgements. This research is supported by NSF IIS-2202108.

References

1. Flavell, J.H.: Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *Am. Psychol.* **34**(10), 906 (1979)
2. Norman, E., Furnes, B.: The concept of “metaemotion”: what is there to learn from research on metacognition? *Emot. Rev.* **8**(2), 187–193 (2016)
3. Kim, J.H.: The effect of metacognitive monitoring feedback on performance in a computer-based training simulation. *Appl. Ergon.* **67**, 193–202 (2018)
4. Kim, J.H., et al.: Investigating the effects of metacognition in dynamic control tasks, in human-computer interaction. Design and development approaches. In: 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9–14, 2011, Proceedings, Part I 14. Springer (2011)
5. Guo, W., Kim, J.H.: Using metacognitive monitoring feedback to improve student learning performance in a real-time location-based augmented reality environment. In: IIE Annual Conference. Proceedings. Institute of Industrial and Systems Engineers (IIE) (2020)
6. Flavell, J.H.: Metacognitive aspects of problem solving. In: *The nature of Intelligence*, pp. 231–236. Routledge (1976)
7. Schraw, G., Dennison, R.S.: Assessing metacognitive awareness. *Contemp. Educ. Psychol.* **19**(4), 460–475 (1994)

8. Ostafichuk, P., et al.: Developing metacognition in first-year students through interactive online videos. In: Proceedings of the 127th American Society of Engineering Education Conference, Virtual (2020)
9. Mostowfi, S., et al.: The effect of metacognitive judgments on metacognitive awareness in an augmented reality environment. In: International Conference on Human-Computer Interaction. Springer (2023)
10. Guo, W., Kim, J.H.: Designing augmented reality learning systems with real-time tracking sensors. In advances in neuroergonomics and cognitive engineering. In: Proceedings of the AHFE 2021 Virtual Conferences on Neuroergonomics and Cognitive Engineering, Industrial Cognitive Ergonomics and Engineering Psychology, and Cognitive Computing and Internet of Things, July 25-29, 2021, USA. Springer (2021)
11. Kim, J.H.: The impact of metacognitive monitoring feedback on mental workload and situational awareness. In: Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings 15. Springer (2018)
12. Pintrich, P.R.: A conceptual framework for assessing motivation and self-regulated learning in college students. *Educ. Psychol. Rev.* **16**, 385–407 (2004)



Virtual Reality for the Human-Centred Design of Assistive Devices

Ivan de Souza Rehder^(✉) , Moacyr Cardoso Machado Junior ,
Edmar Thomaz da Silva , and Emilia Villani

Aeronautical Institute of Technology, São José dos Campos, SP, Brazil
ivan@ita.br

Abstract. The development of assistive devices for the blind and visually impaired (BVI) has often overlooked the essential participation of BVI users in the design process, resulting in products that are not user-friendly for them. This paper introduces a virtual reality (VR)-based framework designed to integrate BVI users into the development of assistive technologies actively. Leveraging VR, the framework facilitates immersive and interactive product testing environments where BVI users can directly evaluate and provide feedback on assistive device prototypes. This method allows for real time adjustments and refinements, significantly enhancing the usability of the products. The setup integrates two scenarios, one virtual and one real, each built with identical configurations. As BVI users navigate the real scenario, their interactions inform the virtual scenario in real-time, allowing for immediate adjustments and refinements. This paper evaluates this framework and seeks to determine if human factors can be used to evaluate assistive products and if non-BVI users, when deprived of their vision, can similarly evaluate assistive devices as BVI users. The proposed framework seeks to elevate the practical utility of assistive devices and to include the users in the design process.

Keywords: VR · mental workload · situation awareness · human-centred design

1 Introduction

Virtual reality (VR) can be a tool for designing assistive devices [14]. Despite modern advances in assistive technology, nowadays options frequently do not attend all the needs of the BVI users [1, 10]. This happens because the development process doesn't include blind and visually impaired (BVI) people in the process. This article aims to show a framework that create real-life situations in a virtual setting allowing the users to test at earlier design stages giving the users a central role in designing them. Additionally, this paper evaluates whether human factors can be used to evaluate assistive products and if non-BVI users, when deprived of their vision, can similarly evaluate assistive devices as BVI users. Chiu [3] emphasizes the importance of effective communication within design teams and between designers and users, suggesting that VR can bridge this gap by facilitating more immersive and interactive collaboration, which is critical for understanding and integrating the true needs and feedback of BVI users into the design process.

1.1 Problem Statement

Most assistive technologies are designed without the participation of their users. As a result, many of these products lack practicality, are not intuitive for BVI users, and sometimes even add unnecessary complexity to their lives [2, 3]. Without their participation, the final products may not address the BVI users' reality, leading to dissatisfaction. Including them in the design process ensure the development of effective assistive devices.

1.2 Previous Works

Prior research has significantly contributed to understanding and developing assistive technologies for blind and visually impaired (BVI) users, yet gaps remain in integrating these users into the design process effectively. Studies by Bradley and Dunlop [2, 3] have explored context-aware solutions and wayfinding techniques to aid BVI navigation, highlighting the necessity for user-centric design approaches. Marston's research [11] further advances this discussion by examining the spatial awareness and environmental interaction required for effective navigation aids tailored to BVI users, advocating for more intuitive and responsive technologies. Additionally, Yang et al. [16] have demonstrated how integrated sensor technologies can enhance the mobility and independence of BVI individuals, providing practical insights into the design of more effective assistive devices. These studies underscore the importance of incorporating direct input from BVI users in the development process to ensure that the technologies developed truly meet their needs and preferences.

The potential of virtual reality (VR) in this domain has gained increasing attention. Kirner et al. [9] explored the use of augmented reality (AR) in cognitive rehabilitation, indicating that these technologies could be effectively adapted for BVI assistive devices, providing a basis for more immersive and user-tailored experiences. Similarly, Siu et al. [14] demonstrated how VR can be employed without visual inputs, using haptic and auditory feedback to navigate complex virtual environments. This method underscores the versatility of VR in creating sensory-rich environments that can cater to the needs of BVI users. Moreover, Moerland-Masic et al. [12] illustrated how VR technology can be integrated into complex design processes, such as the development of aircraft cabins, showing the practical applications of VR in enhancing user experience and facilitating detailed user feedback during the development phase. Additionally, Cummings and Bailenson's 2016 research [5] on the psychological effects of VR emphasizes how immersive VR environments can significantly affect user perception and behavior, which is critical for designing interfaces that are intuitive and responsive to the needs of BVI users.

2 Proposed Framework

This paper introduces a design framework that integrates blind and visually impaired (BVI) users into the development process of assistive devices using virtual reality (VR) environments to simulate real world scenarios, enabling BVI users to interact and evaluate prototypes of assistive devices in real time allowing users to provide a feedback that the development team uses to improve the device.

2.1 The Framework Process

This framework follows a five-phase process: 1) Context Definition; 2) Specification; 3) Development; 4) Preliminary Evaluation; and 5) Systematic Evaluation:

- 1) **Context Definition:** This phase focuses on establishing the settings for the VR simulations. Both real and virtual environments are meticulously crafted to mirror the daily challenges faced by BVI users. By aligning the scenarios with real-world contexts, the framework ensures that the testing environment accurately reflects the actual conditions in which the devices will be used.
- 2) **Specification:** During this phase, the specific features of the prototype devices and the parameters of the VR environment are determined. This includes setting the technical and functional specifications of the assistive devices, as well as configuring the virtual environment's dynamics, such as obstacle placement and interaction triggers.
- 3) **Development:** This phase involves the actual creation of the virtual environments and the initial prototypes of the devices. The development is guided by the specifications set in the previous phase, with a focus on creating a seamless and interactive user experience. This stage is iterative, with continuous refinements being made based on preliminary feedback and technical assessments.
- 4) **Preliminary Evaluation** This stage serves as an initial checkpoint to assess the readiness of the VR scenarios and prototype devices for comprehensive testing. It determines whether the environment and devices are optimized for user interaction. If the preliminary feedback indicates the need for modifications, the project returns to the Development phase for adjustments.
- 5) **Systematic Evaluation:** In this final phase, detailed human factors assessments are conducted to refine the prototypes. Cognitive workload, situation awareness, and overall user experience are thoroughly evaluated. This rigorous analysis ensures that the final products are not only technically sound but also aligned with the real-world needs and preferences of BVI users, thereby enhancing usability and satisfaction.

2.2 Human Factors Assessment Techniques

Drawing on the principles outlined by Sanders [13] regarding human factors in engineering and design, the evaluation of assistive devices within the VR framework incorporates advanced techniques to effectively meet the specific needs of BVI users. These techniques include physiological sensors and subjective questionnaires to assess cognitive workload and situation awareness. Physiological responses such as heart rate variability and skin conductivity are monitored using electrocardiograms (ECG) and galvanic skin response (GSR) sensors. Cognitive workload and situation awareness are further evaluated through the NASA Task Load Index (NASA-TLX) and the Situation Awareness Global Assessment Technique [5, 6], which help to determine how effectively users can navigate and perform tasks within the VR environment. Additionally, subjective questionnaires gather detailed feedback on user satisfaction and perceived usability, ensuring that the devices are not only functional but also align with user preferences [12].

3 Evaluation Methodology

The methodology presented here aims to answer the following questions:

- Is it possible to evaluate and compare concepts of an assistive device from a human factors' perspective in a virtual environment?
- Do non-BVI users, when deprived of their vision, similarly evaluate assistive device as BVI users?

For this purpose, the evaluation was performed with the following groups:

- Blind group: composed of 4 participants with ages varying from 26 to 56, all male, three of them graduated and one with ongoing graduation.
- Sighted group: composed of 4 participants with ages varying from 22 to 31, three males and one woman, all graduated.

For the framework to work, both the virtual environment and the real environment must match in positions. Both environments resemble a medical clinic reception, and the participants must complete the following tasks:

1. Clean the hands at the sanitizer totem (COVID-19 procedures);
2. Go to the reception desk to receive a queue number;
3. Go to the waiting area and wait for the number calling;
4. Leave the room when called.

The experiment is organized as follows:

- The participant receives a brief introduction about the guidance method and is given some time to familiarize and train with it. During this step, no data is gathered.
- First round and it's the NASA-TLX
- Return round and its NASA-TLX

There are five different test devices:

- Regular White cane (Reference method);
- Audio guidance;
- Vibration guidance with command (virtual cane);
- Vibration guidance without command (haptic belt);
- Mixture of audio and vibration guidance

4 Results and Discussions

The experiment to answer these questions has the approval of the Brazilian ethics committee under the process number 52080921.2.0000.5503.

4.1 NASA-TLX

About the assistive devices for BVI users:

- Audio and reference methods consistently show lower mental workload scores compared to other methods.
- The return round generally has lower scores than the first round, indicating an adaptation effect or learning curve.

- The standard deviation is reduced in the return round, especially for the haptic belt, suggesting more consistent performance and lower perceived workload upon repeated exposure.

About the difference between BVI and non-BVI users:

- Non-BVI users typically report higher workload scores than BVI users. The influence of the round is approximately the same for both groups.
- The order of preference for the devices differs between BVI and non-BVI users.

An ANOVA test confirms the influence of the round for both sighted and blind participants. For the devices, the p-value for blind participants is lower than the threshold of 0.05, while that for sighted participants is slightly higher.

5 Conclusions

This research has demonstrated the significant benefits of incorporating a virtual reality (VR) framework in the development of assistive devices for the blind and visually impaired (BVI). By actively involving BVI users in every phase of the design and development process, from context definition to systematic evaluation, the framework ensures that the resulting products are not only technically sound but also finely tuned to the real-world needs and preferences of their end users.

The application of human factors assessment techniques, including physiological sensors and subjective questionnaires, has proven invaluable. These techniques provide deep insights into the cognitive workload, situational awareness, and overall user experience, allowing for iterative refinements that enhance device usability and effectiveness.

This framework not only improves a more inclusive design philosophy that could be applied across various fields where user-specific customization is critical. The next steps would involve broader testing with diverse user groups to further validate and refine the framework, ensuring its adaptability and effectiveness in meeting a wide range of BVI needs.

The ultimate goal of this research is to shift the paradigm in assistive technology development towards a more user-centered approach. By doing so, it hopes to significantly improve the quality of life for BVI individuals, granting them greater independence and confidence in navigating their environments.

References

1. Bourne, R., et al.: Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the global burden of disease study. *Lancet Glob. Health.* **9**(2), e130–e143 (2021)
2. Bradley, N.A., Dunlop, M.D.: Investigating context-aware clues to assist navigation for visually impaired people. *Proceedings of Workshop on Building. Interdisciplinary Context-Sensitive Computing*, University of Glasgow, Bridges (2002)
3. Bradley, N.A., Dunlop, M.D.: An experimental investigation into wayfinding directions for visually impaired people. *Pers. Ubiquit. Comput.* **9**(6), 395–403 (2005)

4. Chiu, M.-L.: An organizational view of design communication in design collaboration. *Des. Stud.* **23**(2), 187–210 (2002)
5. Cummings, J.J., Bailenson, J.N.: How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychol.* **19**(2), 272–309 (2016)
6. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: *Proceedings of the Human Factors Society Annual Meeting*, vol. 32, pp. 97–101. Sage Publications Sage CA, Los Angeles, CA (1988)
7. Endsley, M.R.: Automation and situation awareness. In: *Automation and Human Performance: Theory and Applications*, pp. 163–181. CRC Press (2018)
8. Haas, J. K. (2014). A History of the Unity Game Engine
9. Kirner, C., Cerqueira, C., Kirner, T.: Using augmented reality artifacts in education and cognitive rehabilitation. In: *Virtual Reality in Psychological, Medical and Pedagogical Applications 2 Will-Be-Set-by-IN-TECH*, pp. 247–270 (2012)
10. Lozano, C.A., Kaczmarek, K.A., Santello, M.: Electrotactile stimulation on the tongue: Intensity perception, discrimination, and cross-modality estimation. *Somatosens. Mot. Res.* **26**(2–3), 50–63 (2009)
11. Marston, J.R., Loomis, J.M., Klatzky, R.L., Golledge, R.G., Smith, E.L.: Evaluation of spatial displays for navigation without sight. *ACM Trans. Appl. Percept. (TAP)*. **3**(2), 110–124 (2006)
12. Moerland-Masic, I., Reimer, F., Bock, T.M., Meller, F., Nagel, B.: Application of VR technology in the aircraft cabin design process. *CEAS Aeronaut. J.* **13**, 1–10 (2021)
13. Sanders, M.S., McCormick, E.J.: Human factors in engineering and design. *Ind. Robot. Int. J.* **25**(2), 153 (1998)
14. Siu, A.F., Sinclair, M., Kovacs, R., Ofek, E., Holz, C., Cutrell, E.: Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (2020)
15. Wolf, A., Binder, N., Miehling, J., Wartzack, S.: Towards virtual assessment of human factors: A concept for data-driven prediction and analysis of physical user-product interactions. In: *Proceedings of the Design Society: International Conference on Engineering Design*, vol. 1, pp. 4029–4038. Cambridge University Press (2019)
16. Yang, C.H., Hwang, S.L., Wang, J.L.: The design and evaluation of an auditory navigation system for blind and visually impaired. In: *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 342–345. IEEE (2014)



Quantification of Vergence–Accommodation Conflict for Virtual Reality Learning Content

Takashi Shibata¹(✉) and Shugo Kasuya²

¹ School of Information Science and Technology, Tokai University, Hiratsuka-shi,
Kanagawa 259-1292, Japan
tshibata@tokai.ac.jp

² School of Information and Telecommunication Engineering, Tokai University, Minato-ku,
Tokyo 108-8619, Japan

Abstract. When accessing learning content designed using virtual reality (VR) technology, to observe a 3D object in detail, learners tend to bring the object closer to their eyes. However, this can cause visual fatigue because of the vergence–accommodation conflict (VAC) in stereoscopic viewing. While VR technology enables learners to move 3D objects freely, it is difficult to determine at what depth and for how long learners are viewing 3D content. Prolonged viewing at a near distance can result in severe visual fatigue. Consequently, this study aims to quantify VAC when using VR educational content to facilitate a comfortable learning experience. The authors of this study have developed a function to measure the viewing distance, the depth position of a 3D object, and the amount of binocular disparity when viewing VR educational content. The amount of binocular disparity indicates the degree of VAC. Consequently, estimation of the degree of induced visual fatigue becomes possible. Further, the developed function showed good accuracy. Time-series measurements can also be made using this function. Thus, this study provides a method to objectively record how learning with stereo images takes place, and further proposes ways to reduce the associated visual fatigue.

Keywords: Virtual Reality · Stereoscopic Display · Education · Visual Fatigue · Viewing Distance

1 Introduction

In recent years, there have been remarkable developments in information communication technology (ICT) as well as display technology. In addition to smartphones and computer displays, a number of head mounted displays (HMDs), smart glasses, and other displays with new technology are being developed for the use of virtual reality (VR) and mixed reality applications. For instance, although the development of HMDs has been studied for a long time, a major recent change is that HMDs, that have the ability to display high-quality images, are now commercially available. Thus, in addition to being used for academic research and commercial purposes, general users can also use HMDs in their daily lives. The availability of such displays has facilitated the application of VR technology in many areas, including education [1, 2].

In school education, it has become common to use display technology and ICT as learning tools, with each student now using one digital device for learning in Japan. In classrooms, digital devices are used for a variety of learning purposes, including displaying digital textbooks. While digital textbooks perform the same function as that of traditional paper textbooks, they have the advantage of providing users with the benefits of digital information through being connected to the Internet. For instance, Japanese digital textbooks use two-dimensional matrix barcodes called QR codes to facilitate access to video, audio, and other educational materials. Digital devices also allow learners to view VR content such as 360-degree panoramic images. With the development of display technology and use of educational ICT, in addition to the use of tablet PCs currently used in schools, various information technologies utilizing the characteristics of new displays and media have educational applications.

This study describes the use of laptop-type displays for VR content in classrooms and discusses the vergence–accommodation conflict, a problem that arises during observing stereo images in VR. Subsequently, this study proposes a method for quantifying visual fatigue and learning-related behaviors in the use of VR in education.

2 Aspects of Using VR Technology in School Education

2.1 Use of Laptop-Style VR Displays for Learning

The following points need to be considered regarding designing educational content using VR technology [3]: Suitability for showing the intended educational content, supportive of teachers' teaching methods, aiding to achieve learning objectives, and not burdening learners with visual fatigue or other problems caused by the use of VR displays. A typical use of VR in education is to simulate objects or processes that cannot be experienced under normal circumstances, consequently enabling learners to learn effectively through realistic experiences. In this sense, the use of HMDs can be considered highly effective for VR educational applications. However, in school classes, teachers closely view students' learning processes and consider their learning situation by observing how their students are engaged in learning, viewing what they write on the tablet screen or in their notebooks, observing their facial expressions, and talking to them about their learning experiences. Moreover, it is also important for learners to engage in lively discussions with each other while using educational materials.

For the above reasons, in this study, the authors examine how VR technology can be used in school education using a laptop-style VR system. This system was selected as the study apparatus as it is considered appropriate for the current learning situation in classrooms in Japan. In this system, students can view stereo images and freely manipulate and observe the content using a stylus (Fig. 1). It has been shown that stereoscopic viewing clarifies concavity and convexity, making it easier for learners to notice important points of learning [4], thus facilitating more active group discussions [5].

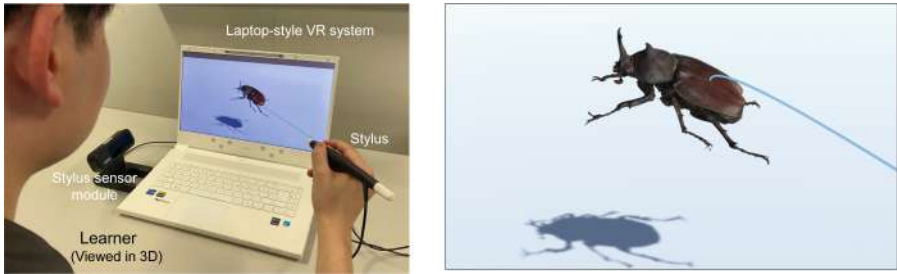


Fig. 1. Laptop-style VR system and 3D object for learning

2.2 Vergence–Accommodation Conflict in VR System

Many current VR systems have a fixed display plane, that creates vergence–accommodation conflict (VAC) during observation (Fig. 2). VAC is known to cause visual fatigue [6, 7]. The greater the degree of the conflict, the worse is the visual fatigue. In conventional three-dimensional (3D) images without interactivity such as 3D movies, VAC can be controlled to some extent by adjusting the stereoscopic effect of content and the viewing environment, thereby reducing visual fatigue. However, in VR-based images, observers can actively move 3D objects included in the content. This is advantageous in learning because observers can bring the 3D object they want to view closer to their eyes. This behavior is commonly observed in learners when learning with VR. However, it is difficult to determine at what depth and for how long learners are viewing content. Consequently, the learner may continue to observe the content using inadvisable behaviors, which can lead to severe visual fatigue. This study attempts to solve this problem by quantitatively measuring the observer’s viewing distance and the observed 3D object’s depth position.

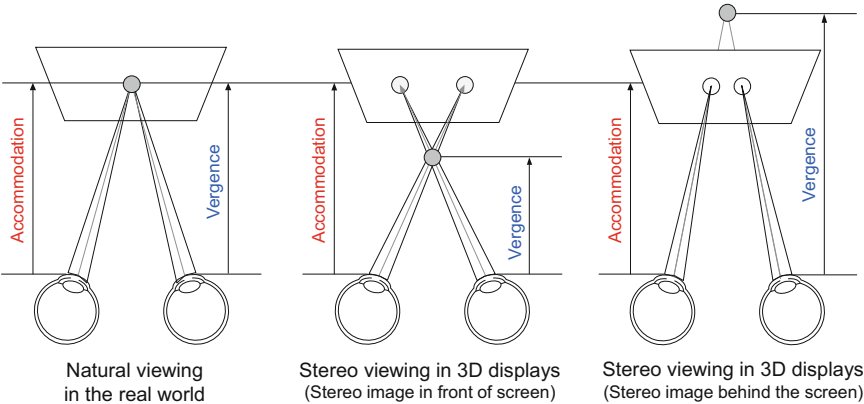


Fig. 2. Vergence–accommodation conflict in stereo viewing.

3 Methods for Quantifying VAC in VR Systems

3.1 VR System and Measurement Items

To quantify VAC, we used the zSpace Inspire (zSpace, Inc.) as the educational VR system. The zSpace Inspire is a laptop-style VR system containing a 15.6-inch autostereoscopic 3D display. Observers can view stereoscopic 3D images without using 3D glasses. They hold a dedicated stylus exhibiting six degrees of freedom, similar to a pen, and can naturally rotate their wrist to move and examine objects in educational materials. The VR system contains two tracking sensors to track the user's eye position and stylus. Since the observer uses the stylus to move the 3D object, it can be assumed that the position of the 3D object indicates the point from where the observer is viewing the object.

We used the Unity game engine (Unity Technologies) to create an experimental 3D computer graphics (3DCG) that was designed to be used as an educational material. The following four items, shown in Fig. 3, were developed in a way that they could be logged to quantify the observer's VR content operation record and observation status. The following parameters are used for this quantification: (1) distance from the eye to the 3D object, (2) distance from the eye to the display (i.e., viewing distance), (3) distance of the 3D object from the screen, and (4) amount of binocular disparity, that is defined by the difference between the vergence angles of the 3D object and screen. This value was used as an index of VAC.

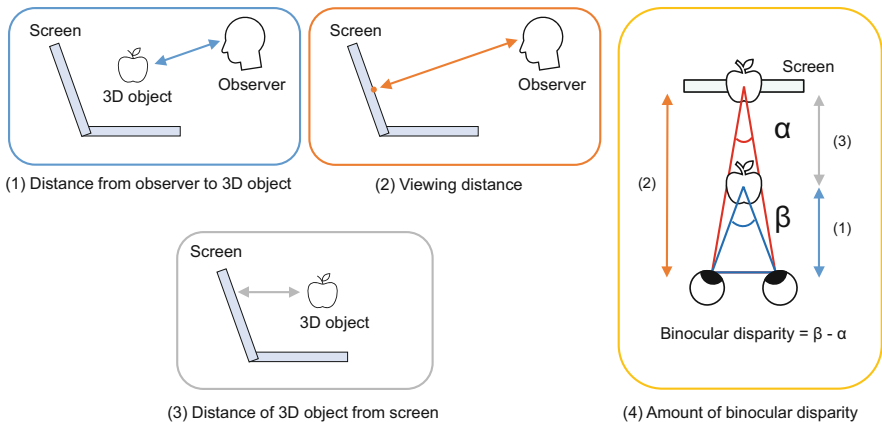


Fig. 3. Values measured during learning while manipulating the 3D object.

The values of (1), (2), and (3) are calculated from the coordinates in the 3DCG. Subsequently, using these values and the average human interpupillary distance (64 mm), the amount of binocular disparity (4) is automatically calculated. The amount of binocular disparity of 0 indicates that the observer is viewing the 3D object on the screen, a positive value indicates that the 3D object is in front of the screen, and a negative value indicates that the 3D object is behind the screen. More importantly, the amount

of binocular disparity indicates the degree of VAC as follows: The larger the amount of binocular disparity, the greater is the degree of visual fatigue [7].

3.2 Verification of Measurement Accuracy

The measurement accuracy was verified after the implementation of the function to measure each value. Measurements were taken with the observer’s viewing distance fixed strictly at 45 cm, 50 cm, and 55 cm using a chinrest. The results showed that the measured viewing distances were 45.08 cm, 50.04 cm, and 54.92 cm, respectively.

3.3 Experimental Measurement

Assuming an educational module of learning of insects, we created a VR learning content that enabled learners to observe the body structure of beetles (Fig. 1). For detailed observations, we used a 3D model of the beetle constructed in high resolution using photogrammetry [8]. The learner was told that the beetle has fine hairs on its body surface and was instructed to observe which parts of the body were covered with these hairs. The learner could move the 3D model freely to better view the beetle. The background of the beetle was a monochromatic image. Observation time was 1 min, and values such as viewing distance and binocular disparity amount were sampled every 2 s.

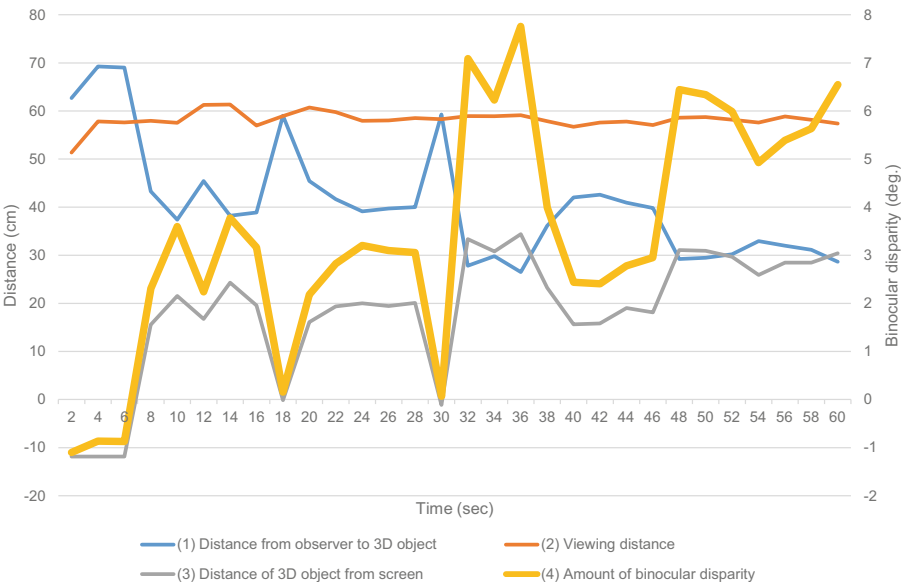


Fig. 4. Measurements of learners’ observational behavior. The numbers in the legend correspond to Fig. 3.

Figure 4 shows the results of the measurements. The viewing distance did not change significantly at approximately 60 cm. Additionally, after 6 s, the 3D object was moved

closer to the learner's eyes. Along with this, the amount of binocular disparity, indicated by a light orange thick line, increased. Considering the amount of binocular disparity, it showed a very large value of approximately 7.8 degrees after 36 s. This is the result of the 3D object being 26 cm closer to the eyes at a viewing distance of 59 cm.

4 Discussions

This study showed it is possible to record time-series changes in the position of the 3D object and viewing distance while using VR content by implementing the necessary functions in the VR system. Thus, this study determined at what depth position the learner was viewing the content and for how long the learner was viewing the content. In experimental measurements, where learners were allowed to freely perform observational learning, the amount of binocular disparity was found to be generally large because the learner brought the 3D object closer to their eyes. Binocular disparity of 1 degree is sometimes used as a criterion for comfort in observing 3D images. For instance, changes in physiological responses have been reported, with longer accommodative responses for observations with binocular disparity exceeding 1 degree [9]. In the present study, there were no significant subjective complaints of visual fatigue after the observation. This may be attributed to short observation periods. Thus, it is inferred that the degree of visual fatigue may be small for short periods of observation. However, it is necessary to carefully consider the time spent learning and the frequency of near-distance viewing, and to contemplate the circumstances under which learners can use VR content safely and comfortably.

Thus, this study made it possible to quantify learning behavior during using VR educational content by considering the way 3D objects are moved by learners; consequently, it determined a method to estimate the visual fatigue caused by VAC. Although this study is specifically intended for the educational use of VR, it can be adapted to other VR applications. In future studies, we aim to explore methods that do not interfere with learning behaviors and do not cause excessive visual fatigue in learners by using data such as the amount of binocular disparity.

Acknowledgments. This work was supported by JST, CREST Grant Number JPMJCR22M4, Japan.

References

1. Al-Ansi, A.M., Jaboob, M., Garad, A., Al-Ansi, A.: Analyzing augmented reality (AR) and virtual reality (VR) recent development in education. *Soc. Sci. Humanit. Open*. **8**(1), 100532 (2023)
2. Guindy, M., Kara, P.A.: Light field visualization for training and education: areview. *Electronics*. **13**(5), 876 (2024)
3. Shibata, T.: Virtual reality in education: how schools use VR in classrooms. In: *Proceedings of the 20th Congress of the International Ergonomics Association*, pp. 423–425 (2018)
4. Shibata, T., Drago, E., Araki, T., Horita, T.: Comparison of 2D and 3D views on educational virtual reality content. *Int. J. Educ. Media Technol.* **15**(1), 14–22 (2021)

5. Shibata, T., Drago, E., Araki, T., Horita, T.: Encouraging collaborative learning in classrooms using virtual reality techniques. In: *Proceedings of EdMedia: World Conference on Educational Media and Technology*, pp. 1577–1582 (2018)
6. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* **8**(3), 1–33 (2008)
7. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: predicting visual discomfort with stereo displays. *J. Vis.* **11**(8), 1–29 (2011)
8. Kano, Y.: Bio-photogrammetry: Digitally archiving coloured 3D morphology data of creatures and associated challenges. *Res. Ideas Outcomes*. **8**, e86985 (2022)
9. Inoue, T., Ohzu, H.: Accommodative responses to stereoscopic three-dimensional display. *Appl. Opt.* **36**(19), 4509–4515 (1997)



A Review: Identification of Menu Types, Interaction Tasks, and Performance Metrics in Virtual Environment Menu Interaction

Benedikta A. H. Siboro^(✉)  and Chiuhsiang Joe Lin 

National Taiwan University of Science and Technology, Section 4 No. 43, Keelung Rd., Da'an Dist., 10607, Taipei City, Taiwan, ROC
cjoelin@mail.ntust.edu.tw

Abstract. This systematic literature review was performed to synthesize findings from previous research on menu interaction in virtual environments. Specifically, previous research published between January 2013 and December 2023 was examined through different stages, which included identification, screening, and eligibility. A total of 32 articles were selected and reviewed, considering various tools such as (1) menu types, (2) tasks assigned, and (3) performance metrics. Consequently, the findings showed that radial menu types were the most popular used tools, particularly for selection tasks. Additionally, usability and task completion time were observed to be the frequently used performance indicators. Finally, the paper concluded with recommendations for future exploration.

Keywords: Menu types · Interaction tasks · Performance metrics · Virtual environment

1 Introduction

The core components of virtual reality (VR) are known as the 3 I's, namely imagination, interaction, and immersion [1]. A crucial aspect in this context is the user interface (UI), which facilitates interaction between humans and computers through both hardware and software [2]. Among these UI elements, the menu plays a fundamental role, enabling users to navigate and select options [3]. Current research on menu interactions within virtual environments focuses on exploring and categorizing menu types, associated tasks, and performance metrics [4–6]. Despite extensive research on menu usage, interaction, and performance evaluation in VR, there is a noticeable gap in analyzing current usage trends and developing an appropriate hierarchical menu structure. Therefore, this research aimed to address the gaps by identifying the most commonly used menu types (RQ1), analyzing the types of tasks in menu interactions (RQ2) and classifying the parameters used to measure performance (RQ3).

2 Methodology

The systematic Literature Review (SLR) method explored the menu types, interaction tasks, and performance parameters adopted in VR [7]. This process was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), which consisted of 4 critical stages [8]. A comprehensive literature search on Scopus and Web of Science was conducted using some terms. Specifically, the terms included (“menu” OR “3D menus”) AND (“task”) AND (“performance”) AND (“virtual environment” OR “virtual reality” OR “virtual immersive environment” OR “immersive virtual environment” OR “stereoscopic”), which initially produced 209 articles. Duplicates were removed using Mendeley software, resulting in 188 articles. Furthermore, specific screening criteria were applied to review the titles and abstracts, excluding articles based on mismatched definitions of ‘menu’, review articles, theses, books, publications before 2013, and those lacking keywords. After reviewing the full texts, 9 more articles were excluded. Finally, out of 53 articles assessed for eligibility, 32 were selected for final analysis after excluding 21 for not meeting inclusion criteria.

3 Result

Research Question (RQ1) What menu types are commonly used in VR interaction? Menu types were classified [4], showing that the layout category, particularly radial or pie, was the most commonly used menu ($n = 12$), as shown in Table 1.

Table 1. Menu types

Menu and References	Frequency
a. Layout	
- Radial/pie/extended pie [9–19]	12
- Panel [17, 20, 21], Ring [22], Grid [23]	5
b. Depth of hierarchy	
- Linear (pull down) [3, 13, 18, 23–25]; Pop-up [3, 26]; Stack [3]	9
c. Position (handheld fixed [23] and floating [23, 27–30])	5
d. Others (mobile and dashboard) [11, 31]	2

Research Question (RQ2) What tasks are commonly performed when using menus in virtual environments? We classified 3 categories interaction tasks [6, 32], and selection received more attention in the research than other tasks, as shown in Table 2.

Research Question (RQ3) What performance metrics are required during menu interaction in a virtual environment? Performance metrics were categorized into 3 [5], and usability was the most popularly used metric, as shown in Table 3.

Table 2. Interaction Tasks

Tasks and References	Frequency
a. Navigation [11, 16, 24, 25]	4
b. Selection and Manipulation	
- Selection [3, 9, 12–21, 23, 24, 26, 27, 29–31, 33–36]	23
- Manipulation [11, 12, 21, 22, 24, 28, 33, 36–38]	10
- Rotation [21, 39]	2
c. System Control and other	
- Scaling [39]; Commands [11, 40]; Translation [9, 20, 39], simulation [36]	7

Table 3. Performance metrics

Performance Metrics and References	Frequency
a. Independent variables	
Selection technique, type/shape, size, menu placement, menu interface, task difficulty, menu geometry, layout, alignment [3, 13, 17–21, 29]	16
b. Dependent variables	
Task completion time [3, 9, 13, 15, 16, 20–25, 28, 29, 33, 38]	17
Error rate [10, 13–15, 18, 19, 23, 28, 29, 33, 38]	11
Interaction, selection, movement, execution, activation, and evaluation time [10, 11, 14, 17–19, 23, 25, 39]	9
Accuracy [11, 24, 37, 38]; Task Effectiveness [9, 40]	6
Success rate [23]; Effective throughput, throughput [14, 15]	3
c. Questionnaires	
Usability, presence, perceived usability (ease to use, efficiency, learnability, ease to remember, awareness, attention, finding tools), preference, user discomfort, user experience, satisfaction [3, 9, 11, 15–19, 21, 23, 25–27, 29–31, 33–36, 38–40]	30
Workload [15, 20, 21, 29, 38]	5
Fatigue [3, 20, 23]; Sickness [30, 34]	5
d. Uncategorized (image, game integration, head movement degree, angle traveled, unnecessary steps, EEG, GUI) [13, 17, 20, 25, 30]	9

4 Discussion

This systematic review examined various menu types with the radial menu being the most commonly used (**RQ1**). The radial menu, classified by its layout, is a handheld tool for users to call up temporary options [4] quickly. This classification is a key factor in determining menu usability by ensuring the necessary appearance for users, while position focuses on the placement and orientation of the menu's reference objects and

depth hierarchy related to menu structure [4, 41]. Typically, the radial menu leverages the 3D spatial nature of VR, offering a more immersive and intuitive way to interact with digital content and contributing to more efficient search times, lower error rates, and quick selection times [11, 19]. However, as the number of items in the pie menu increases, selection times and errors tend to rise. Most research opts for the radial menu as the preferred UI, likely due to several reasons, including (1) ease of use for novices, (2) reduction of human error, and (3) enhancement of accuracy and reliability. Furthermore, some studies have conducted comparisons with other menu types, and the results indicated that the radial menu consistently exhibits superior performance [11, 17, 18].

This result also showed that selection was the most commonly performed task in menu interactions (**RQ2**). Specifically, task selection was implemented in various fields such as games, education, exhibitions, learning, experiments, and training in most articles. Some articles further elaborated on the selection task by incorporating techniques such as head and hand button presses, hand and head pointing, and the virtual belt. The techniques were designed to be user-friendly, focusing on efficiency and error minimization. The selection task, along with specific interaction techniques, is crucial for investigating the effectiveness, efficiency, and user satisfaction when interacting with menu in the virtual environment. Additionally, some articles contained more than one task during interaction in the virtual environment. For instance, Rahim et al. [35] and Nan et al. [24] used selection, navigation, and manipulation to obtain user-defined gestures and to develop Vdesign using hand interaction.

Based on the numerous performance metrics available, only the frequently used ones were selected for presentation in conjunction with the review results. The analysis showed that most articles used usability as a metric for questionnaires and task completion time as a metric for dependent variables (**RQ3**). In the context of this research, usability is the ability at which a system can be understood, learned, adopted, and implemented by the user. Furthermore, usability comprises other concepts such as efficiency, learnability, memorability, errors, and satisfaction. Most research utilized different methods to measure the usability of menu interactions in virtual environments, such as Computer Systems Usability Questionnaire (CSUQ), AttrakDiff, Borg, User Experience Questionnaire (UEQ), and SUS [11, 19, 30, 38, 40].

Usability tests covered a wide range of aspects, including presence, perceived usability, user experience, and discomfort, all of which were related to these concepts. Additionally, task completion time was frequently used by research to compare different menu types, interaction methods (egocentric vs. exocentric), interaction techniques, and menu conditions. Meanwhile, some articles explored metrics beyond these three categories, as shown in Table 3.

The research provides new insights into the design and development of performance measurement in VR. The first is prioritizing the human element, which is essential for crafting engaging and efficient experiences through the determination of appropriate menu design. Human-centred design ensures that virtual environments are tailored to meet user needs, preferences, and behaviours, thereby enhancing usability and user satisfaction. The second is task forms. The more difficult the task, the longer the task completion time and the impact on usability. By timing users' task completion in virtual environments, the impact of different interaction techniques, feedback mechanisms,

or design elements could be evaluated on user performance. Finally, there is a clear opportunity to expand performance metrics in the future, even though performance measurement is predominantly focused on usability and task completion time. This evolution is in line with the increasing complexity of VR tools under development, the diverse interaction techniques used, and various other influencing factors.

5 Conclusion and Future Research

This study reviewed 32 articles to identify menu types, interaction tasks, and performance metrics used in evaluating menu interactions in VR. A radial menu was predominantly used, with user interactions primarily focusing on selection tasks. Additionally, performance was assessed based on usability and task completion time, signifying the importance of human factors, effectiveness, and efficiency in interactions. The research provided insights into the factors to consider when designing menu interfaces in virtual environments and proposed practical guidelines for designers.

The research had several limitations that needed to be addressed in the future. The first was the systematic keyword search that was limited to just two primary sources, capturing only 32 articles and potentially missing relevant exploration from other databases. The second limitation focused on menu types, interaction tasks, and performance metrics. In the future, additional aspects of menu interaction could be explored in VR, such as tools and interaction techniques.

Acknowledgement. This paper is supported by a project from the Taiwan Ministry of Science and Technology (MOST 111-2221-E-011 -074 -MY3).

References

1. Wu, H., Cai, T., Luo, D., Liu, Y., Zhang, Z.: Immersive virtual reality news: A study of user experience and media effects. *Int. J. Hum. Comput. Stud.* **147**, 102576 (2021)
2. Bowman, D.: Interaction Techniques for Common Tasks in Immersive Virtual Environments: Design, Evaluation, and Application. PhD Thesis, Georgia Institute of Technology (1999)
3. Jeong, S., Jung, E.S., Im, Y.: Ergonomic evaluation of interaction techniques & 3D menus for the practical design of 3D stereoscopic displays. *Int. J. Ind. Ergon.* **53**, 205–218 (2016)
4. Hou, S., Thomas, B.H., Lu, X.: VrMenude signer: A toolkit for automatically generating and modifying VR menus. *Proc. IEEE Int. Conf. Virtual Reality*, 154–159 (2021)
5. Samini, A., Palmerius, K.L.: Popular performance metrics for evaluation of interaction in virtual and augmented reality. In: *Proceeding International Conference Cyberworlds*, pp. 206–209 (2017)
6. LaViola, J.J., et al.: 3D User Interfaces Theory & Practice. Pearson Education (2017)
7. Booth, A., Sutton, A., Papaioannou, D.: Systematic Approaches to a Successful Literature Review. Sage Publications Ltd (2016)
8. Moher, D., et al.: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **6**(7), e1000097 (2009)
9. Bustamante, M., et al.: ColorBound: Comparing menu dynamics in virtual reality. In: *Proceeding. IEEE Conf. Virtual Real. 3D User Interfaces*, pp. 914–915 (2022)
10. Hansen, N.T., Hald, K., Stenholt, R.: *IEEE Symp. 3D User Interfaces*, pp. 137–138 (2014)

11. Gebhardt, S., et al.: An evaluation of a smart-phone-based menu system for immersive virtual environments. In: *Proc. IEEE Symp. 3D User Interfaces 2014*, pp. 31–34 (2014)
12. Kwon, J.H., Kim, J.Y., Nam, S.H.: Designing 3D menu interfaces for spatial interaction in virtual environments. *Int. J. Grid Distrib. Comput.* **10**, 31–38 (2017)
13. Lediaeva, I., LaViola, J.J.: Evaluation of Body-referenced graphical menus in virtual environments. In: *Proc. Graphics Interface*, pp. 1–9 (2020)
14. Lubos, P., Bruder, G., Steinicke, F.: Analysis of direct selection in head-mounted display environments. In: *Proc. IEEE Symp. 3D User Interfaces 2014*, pp. 11–18 (2014)
15. Lubos, P., Bruder, G., Ariza, O., Steinicke, F.: Touching the sphere: Leveraging joint-centered kinespheres for spatial user interaction. *Proc. Spat. User Interact.*, 13–22 (2016)
16. Mardanbegi, D., et al.: EyeSeeThrough: Unifying tool selection and application in virtual environments. In: *Proc. 26th IEEE Conf. Virtual Real. 3D User Interfaces*, pp. 474–483 (2019)
17. Monteiro, P., Coelho, H., Gonçalves, G., Melo, M., Bessa, M.: Comparison of radial and panel menus in virtual reality. *IEEE Access.* **7**, 116370–116379 (2019)
18. Santos, A., Aedo, I., Zarraonandia, T., Díaz, P.: A comparative study of menus in virtual reality environments. In: *Proc. ACM Int. Conf. Interact. Surfaces Spaces*, pp. 294–299 (2017)
19. Gebhardt, S., et al.: Extended pie menus for immersive virtual environments. *IEEE Trans. Vis. Comput. Graph.* **19**, 644–651 (2013)
20. Grogna, D., et al.: The impact of drowsiness on in-vehicle human-machine interaction with head-up and head-down displays. *Multimed. Tools Appl.* **77**, 27807–27827 (2018)
21. Shi, R., Zhang, J., Stuerzlinger, W., Liang, H.N.: Group-based Object Alignment in Virtual Reality Environments. *Proceeding ACM Conf. Spat, User Interact* (2022)
22. Mendes, D., et al.: Mid-air modeling with Boolean operations in VR. In: *2017 Proceeding IEEE Symp. 3D User Interfaces*, pp. 154–157 (2017)
23. Saidi, H., Dubois, E., Serrano, M.: Holobar: Rapid command execution for head-worn ar exploiting around the field-of-view interaction. In: *Proc. Conf. Hum. Factors Comput* (2021)
24. Nan, X., et al.: vDesign: A CAVE-based virtual design environment using hand interactions. *J. Multimodal User Interfaces.* **8**, 367–379 (2014)
25. Zielasko, D., et al.: Passive Haptic Menus for Desk-Based and HMD-Projected Virtual Reality. In: *Proceeding IEEE 5th Work. Everyday Virtual Reality*, pp. 5–10 (2019)
26. Wagner, A., Deligiannidis, L.: Showcasing Native Wildflowers in Virtual Reality. In: *Proceeding Int. Conf. Comput. Sci. Comput. Intell*, pp. 1992–1995 (2022)
27. Wegner, K., et al.: Comparison of two inventory design concepts in a collaborative virtual reality serious game. In: *Proc. Computer-Human Interaction*, pp. 323–329 (2017)
28. Fu, Y., Li, Q.: Spatial reference frame-based user interface design in the virtual reality game design. In: *DMSVIVA 2020—Proc. 26th Int. DMS Conf. Vis. Vis. Lang*, pp. 34–36 (2020)
29. Wang, Y., Hu, Y., Chen, Y.: An experimental investigation of menu selection for immersive virtual environments: fixed versus handheld menus. *Virtual Real.* **25**, 409–419 (2021)
30. Schild, J., et al.: Creating and analyzing stereoscopic 3D graphical user interfaces in digital games. In: *Proc. Conf. Hum. Factors Comput. Syst*, pp. 169–178 (2013)
31. Woodworth, J.W., Broussard, D., Borst, C.W.: Designing Tools To Improve Collaborative Interaction in a VR Environment for Teaching Geosciences Interpretation. In: *Proc. Ser. Gesellschaft fur Inform* (2020)
32. Sun, C., Hu, W., Xu, D.: Navigation modes, operation methods, observation scales and background options in UI design for high learning performance in VR-based architectural applications. *J. Comput. Des. Eng.* **6**, 189–196 (2019)
33. Rehman, I.U.: FPSI-Fingertip pose and state-based natural. *Multimed. Tools Appl.* **20711–20740** (2023)
34. Kim, M., Lee, J., Kim, C., Kim, J.: TPVR: User interaction of third person virtual reality for new presence and experience. *Symmetry (Basel).* **10**, 1–15 (2018)

35. Zeng, X., et al.: An interactive VR system based on full-body tracking and gesture recognition. *Opt. Des. Test. VII.* **10021**, 1002129 (2016)
36. Rahim, N., Wook, T.S.M.T., Zin, N.A.M.: User-defined gestures for virtual heritage in three-dimensional user interface: motion Gestures. *Int. J. Recent Tech. Eng.* **7**, 85–92 (2019)
37. Saavedra, M., et al.: Development of a 3D User Interface for Programming a Robotic Arm Using Virtual Reality. *Smart Innov. Syst. Technol.* **255**, 435–446 (2022)
38. Khadka, R., Banic, A.: Effects of egocentric versus exocentric virtual object storage technique on cognition in virtual environments. In: *Proc. IEEE Conf. Virtual Real. 3D User Interfaces*, pp. 205–209 (2020)
39. Lopes, D.S., et al.: Expedition illustration of layer-cake models on and above a tactile surface. *Comput. Geosci.* **90**, 1–9 (2016)
40. Lafci, M. T. et al. An evaluation of hand interaction metaphors for immersive environments. *Proc. 15th Int. Conf. Qual. Multimed.* pp.232–235 (2023)
41. R. Dachsel and A. Hübner, “Three-dimensional menus: A survey and taxonomy,” *Comput. Graph.* **31**(1), 53–65 (2007)

Others (I)



Earable-Based Posture Inference and Correction System for Deskwork Behavior Transformation

Yilizati Abulaiti, Shoko Kimura , and Guillaume Lopez ^(✉) 

Aoyama Gakuin University, Sagami-hara, Japan
guillaume@it.aoyama.ac.jp

Abstract. In contemporary society, the prevalence of posture-related health issues has reached alarming levels, particularly within office and educational environments. In response to these concerns, ergonomics has evolved to include how the individual interacts with the work environment. This study aims to fill gaps in posture correction, using Earables as a pioneering approach to health promotion in a deskwork context. We have developed “EPICS,” a posture inference and correction system combining a multi-sensory earable device and a scalable movement sensor. EPICS can identify seven types of postural movements in real-time, including three poor postures and their transitions through machine learning algorithms. A dual-application mobile platform leverages inference results for immediate inference and feedback. Substantial improvements in posture detection accuracy confirmed the system’s effectiveness, reaching 94% on a 20-subject dataset, showcasing its reliability and robustness to different individuals. Various test scenarios demonstrated the system’s reliability and effectiveness in real-world applications. Survey results indicated that real-time posture correction feedback helps users become more aware of their posture, thus encouraging the development of healthier postural habits, and confirming EPICS’ effectiveness.

Keywords: Earable · Deskwork · Posture · Activity Computing · Behavior Transformation

1 Introduction

1.1 Impact of Posture-Related Issues on Health and Economy

In contemporary society, the prevalence of posture-related health issues has reached alarming levels, particularly within office environments and educational institutions. As modern lifestyles become increasingly sedentary, there is a corresponding surge in the incidence of musculoskeletal disorders (MSDs). Statistics reveal that 70% to 80% of adults in industrialized countries report musculoskeletal discomfort from prolonged sitting, with growing evidence linking such discomfort to sedentary behavior [1]. Recent studies have highlighted that desk-based employees can spend approximately 70% of their workday seated, which is linked to a range of health concerns, including MSDs like lower back pain [2].

Beyond the direct health implications, poor posture and associated disorders carry substantial economic burdens, and interventions to correct poor seating posture and enhance ergonomic arrangements can mitigate these economic costs significantly [3]. The US Department of Labor highlights that MSDs contribute significantly to work and study-related injuries and absences [4]. The World Health Organization (WHO) underscores the high cost of physical inactivity, projecting that almost half a billion people could develop heart disease, obesity, diabetes, or other diseases by 2030 due to insufficient physical activity [5]. McKinsey & Company's analysis further elucidates that improved global health could contribute an additional \$12 trillion to global GDP by 2040, driven by an expanded healthy labor force and increased productivity [6].

1.2 Technological Advances in Posture Management

The contemporary field of ergonomics now encompasses a vast array of studies, from optimizing physical workspaces to enhancing the interaction with digital interfaces. This more holistic approach to ergonomics emphasizes the importance of behavior change alongside physical adjustments [7]. The proliferation of wearable sensors signifies a broader acceptance and integration of these technologies in everyday life, promising to transform rehabilitation, sports performance, and workplace risk assessment [8]. Besides, integrating Artificial Intelligence (AI) in ergonomic design implies a shift towards more data-driven and personalized approaches.

This study addresses the critical issue of posture-related health problems among office workers and students. The primary goal is to fill the gaps in current posture correction methods, enhancing accessibility and user engagement. By developing an innovative posture estimation system suitable for deskwork, it employs the latest wearable technology, including eSense Earables (Nokia Bell Labs [9]) and Movesense sensors (Suunto [10]), in conjunction with a user-friendly mobile application.

2 Related Works

Yoga practice is a popular application target for posture detection technology. Fazil Rishan et al. developed the "Infinity Yoga Tutor," a mobile application for real-time yoga posture detection and correction using facial key points in images and videos and machine learning [11]. Munkhjargal Gochoo et al. developed an IoT-based yoga posture recognition system using low-resolution infrared sensors and deep learning [12]. Their system achieved high accuracy while addressing privacy concerns, demonstrating its potential for non-intrusive health monitoring.

Francesco Pistolesi et al. explored using 2-D LiDAR technology for non-invasive posture monitoring in office environments by tracking six typical leg motions [13]. Leonor Pereira et al. developed an intelligent chair system that can accurately monitor posture and ECG in office settings using sensors embedded in the chair's armrests [14]. A team led by Martina Zaltieri assessed a wearable system using Fiber Bragg Grating (FBG) sensors for recognizing sitting postures and evaluating respiratory rates [15]. The study by Ravi Charan Ailneni et al. examined wearable posture sensors' impact on head

and neck posture, showing significant reductions in neck flexion angles and gravitational moments using wearable sensors, particularly at standing workstations [16].

While advanced ergonomic technologies offer significant benefits, their implementation faces challenges like the complexity and cost of LiDAR systems, the need for multiple sensors, privacy concerns with camera-based systems, limitations of infrared sensors, and issues with wearable technology like user discomfort and maintenance needs. However, they also provide opportunities for enhanced ergonomic assessments, personalized solutions, predictive risk identification, and long-term cost savings through improved health and productivity.

3 EPICS: Eearable-based Posture Inference and Correction System

3.1 EPICS Overview

EPICS combines eSense Earables worn at the ear and Movesense sensor worn at the waist. Both sensors capture gyroscopic data (deg/s) at a sampling rate of 50 Hz, essential for accurate posture detection. The system includes two smartphone application software (App): the Data Collection App and the Posture Estimation App. The former gathers data from the sensors to feed into a machine-learning model, while the latter provides real-time posture analysis and feedback to the user (see Fig. 1).



Fig. 1. Schematic of EPICS overview (left) and picture of data collection set-up (right)

Participants of the experiments (3.1 and 4.1) were informed about the study's purpose and assured of the confidentiality of their responses. Informed consent was obtained, ensuring participants were aware of their voluntary participation.

3.2 Data Collection App and Modelling

This phase captures real-time gyroscopic data, starting with meticulous calibration of the sensors to ensure accuracy. In addition to EPICS sensors, Azure Kinect [17] obtained reference posture quantification from the skeleton data it produces. Twenty participants, aged 20 to 30, performed four postures: normal [18, 19], lean forward [19], hunch back [20], and lean backward [20]. Collected data pre-processing goes through synchronization of head and waist motion data using resampling and interpolation techniques, segmentation to obtain manageable chunks, and calculation of various statistical features

to form feature vectors. A machine learning model processes these vectors and classifies the user’s posture into one normal posture, three bad postures (lean forward, hunch back, lean backward), and the return movements from these three bad postures to normal.

We explored several machine learning algorithms for posture movement identification. Random Forest stands out due to its ensemble approach, which leverages multiple decision trees to improve prediction accuracy and robustness. Table 1 summarizes the performance of each algorithm.

Table 1. Evaluation results of various machine learning models

	Decision Tree	Random Forest	SVM	Logistic Regression	KNN	Naïve Bayes
Accuracy	89.80	94.4%	75.7%	59.0%	77.7%	69.1%
F1 Score	88.9%	93.7%	74.0%	56.5%	75.5%	65.1%
Precision	89.3%	94.0%	77.4%	57.9%	77.1%	73.0%
Recall	88.5%	93.5%	72.2%	58.2%	74.4%	67.0%

3.3 Posture Estimation and Correction App

The App processes the data collected by the two sensors, utilizing the above-selected Random Forest model to predict the user’s posture in real-time, and provides immediate feedback to promote continuous user engagement and posture improvement. If a bad posture is detected, the app issues alerts to prompt correction (see Fig. 2 left). In addition, a scoring system accompanied by emotional graphics motivates users by penalizing bad and rewarding good postures (see Fig. 2 middle & right).

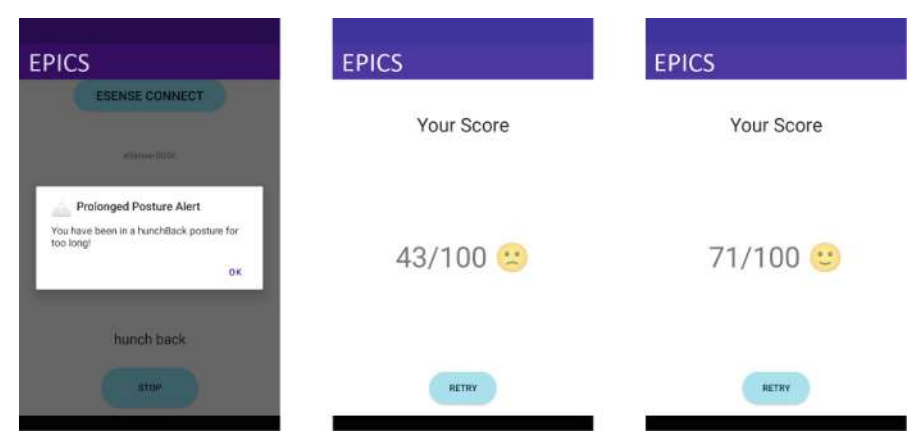


Fig. 2. Screenshots of the posture estimation App feedback examples: bad posture alert (left), session bad score (middle), session good score (right)

4 Posture Detection App User Experiment

4.1 Experiment Design and Methodology

The objective of the user experiment was to assess the App’s ability to detect and accurately correct poor posture in real-time and to gather user feedback on its usability and overall experience. The experiment involved 16 volunteers from 22 to 35 years old. After using the App for 10 to 20 min, the participants answered using a 5-point scale to a qualitative survey about usability, feedback effectiveness, design and layout, overall experience, and scoring system effect on motivation.

4.2 Survey Results

As shown in Table 2, the results suggest that the App’s alert mechanisms are largely effective in prompting users toward posture correction. However, the App’s design and layout received mixed reviews, indicating a need for aesthetic improvements. The scoring system was generally well understood and appreciated for its motivational and engaging qualities. Nevertheless, some users have raised the need for a multimodal alert approach, such as vibration and flashlight alerts. Others suggested that a comparative scoring system that allows viewing progress over time or in relation to others could provide an added motivational boost.

Table 2. Summary of the participants’ subjective assessment for each survey section

	1. not good	2. poor	3. neutral	4. good	5. very good
ease of use	0	0	0	37.5	62.5
feedback effectiveness	0	0	12.5	37.5	50
design & layout	0	20.0	53.3	26.7	0
overall experience	0	0	18.8	18.8	62.4
motivation by scoring	0	0	18.8	56.2	25.0

5 Conclusion and Future Work

EPICS achieves effective assessment of deskwork posture-related habits. The user experiment demonstrated its real-world viability, with participant feedback highlighting its practicality and user-centric design. While the findings are encouraging, we recognize the study’s limitations, such as the user experiment’s sample size and the range of postures analyzed.

Future research endeavors will focus on expanding the capabilities of the posture estimation system. The primary goal is to enhance the system’s ability to recognize a wider array of posture types, thereby increasing its adaptability to various ergonomic requirements in different environments. Furthermore, the project will incorporate an expanded set of feedback methods to enrich user interaction and experience.

References

1. Shariat, A., et al.: Office exercise training to reduce and prevent the occurrence of musculoskeletal disorders among office workers: A hypothesis. *Malays J. Med. Sci.* **23**(4), 54–58 (2016)
2. Edwardson, C.L., et al.: Effectiveness of an intervention for reducing sitting time and improving health in office workers: three arm cluster randomised controlled trial. *BMJ.* **378**, e069288 (2022)
3. van Niekerk, S.M., Louw, Q.A., Hillier, S.: The effectiveness of a chair intervention in the workplace to reduce musculoskeletal symptoms. A systematic review. *BMC Musculoskelet. Disord.* **13**, 145 (2012)
4. Putz-Anderson, V., et al.: Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back (July 1997). https://stacks.cdc.gov/view/cdc/21745/cdc_21745_DS1.pdf, Accessed 20 May 2024
5. World Health Organization (2022): WHO Highlights High Cost of Physical Inactivity in First-Ever Global Report. <https://www.who.int/news/item/19-10-2022-who-highlights-high-cost-of-physical-inactivity-in-first-ever-global-report>, Accessed 20 May 2024
6. McKinsey & Company (2022): How Prioritizing Health Could Help Rebuild Economies. <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/how-prioritizing-health-could-help-rebuild-economies>, Accessed 20 May 2024
7. University Hospitals of Leicester NHS Trust: Reducing Sitting Time to Boost Health in Office Workers (2022, August 17). <https://medicalxpress.com/news/2022-08-boost-health-office-workers.html>, Accessed 20 May 2024
8. Lee, W. (Ed.): Wearable sensors for biomechanics applications [Special issue]. *Sensors*, MDPI (2023). https://www.mdpi.com/journal/sensors/special_issues/wearable_biomechanics
9. Nokia Bell Labs (n.d.): eSense. <https://esense.io/>, last accessed 2024/05/20
10. Suunto (n.d.): Movesense. <https://www.suunto.com/Campaigns/Movesense/>, Accessed 20 May 2024
11. Bhattacharjee, T., Biswas, P.: A wearable sensor based system for real-time monitoring of repetitive wrist movements during occupational tasks. In: 2020 IEEE International Conference on Innovations in Technology for Rehabilitation (ICITR), 1–6, (2020). <https://doi.org/10.1109/ICITR51448.2020.9310832>
12. Gochoo, M., et al.: Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning. *IEEE Internet of Things J.* **6**, 7192–7200 (2019)
13. Pistolesi, F., Baldassini, M., Lazzarini, B.: Are you sitting with good posture? Tracking the position of the legs via 2-D LiDAR. *IEEE Embedded Syst. Lett.* **15**, 137–140 (2023)
14. Pereira, L., Plácido da Silva, H.: A novel smart chair system for posture classification and invisible ECG monitoring. *Sens.* **23**, 719 (2023)
15. Zaltieri, M., et al.: Assessment of a multi-sensor FBG-based wearable system in sitting postures recognition and respiratory rate evaluation of office workers. *I.E.E.E. Trans. Biomed. Eng.* **70**, 1673–1682 (2023)
16. Ailneni, R.C., et al.: Influence of the wearable posture correction sensor on head and neck posture: sitting and standing workstations. *Work.* **62**, 27–35 (2019)
17. Microsoft (2020): Azure Kinect DK documentation. <https://docs.microsoft.com/en-us/azure/kinect-dk/>, Accessed 20 May 2024
18. Mayo Clinic. Office ergonomics: Your how-to guide. <https://www.mayoclinic.org/healthy-lifestyle/adult-health/in-depth/office-ergonomics/art-20046169>, (2024). Accessed 20 May 2024

19. Back Intelligence. 10 Tips for Proper Posture at a Desk—Correct Sitting Posture. <https://backintelligence.com/proper-sitting-posture-at-a-desk/>, (2024). Accessed 20 May 2024
20. Cleveland Clinic. Posture: What it is & Why it Matters for Your Health. <https://my.clevelandclinic.org/health/articles/posture>, (2024). Accessed 20 May 2024



Inclusive Dwelling for Inclusive Living Communities in the Current Eco-Socio-Technical Changing Scenario

Erminia Attaianese^(✉) 

University of Naples Federico II, Via Toledo 402, 80135 Naples, Italy
erminia.attaianese@unina.it

Abstract. The paper discusses the role of inclusive dwellings for inclusive living communities. At the light of vulnerability of contemporary society, and despite the global diffusion of the current interdependent critical phenomena, not everyone is equally at risk, since available social and economic resources and the qualitative performance of places and services, may improve or worse the population's physical and mental health. In this challenging framework, it is paramount that people, especially fragile ones, are supported, also by the built environment, to fit the changes that these phenomena continue to generate. Inclusive design for emergent dwelling introduces an array of topics that are relevant for Inclusive Living Communities (ILCs), by focusing on how to conceive the emergent dwelling to be responsive to different human needs in a lifelong and multigenerational perspective and to assure a low environmental impact as well as “climate proof and pandemic proof” conditions. Three experimental applications of building re-functionalization for inclusive dwellings are also presented.

Keywords: Human-Centred Design · Generation diversity · Emergent dwelling

1 Introduction

Connections between ecological emergencies and social crises are phenomena that favor the development and diffusion of systemic crises. Factors such as climate change or the loss of biodiversity, together with the depletion of resources and the reduction of natural habitats, associated with widespread social, health and environmental inequalities, determine conditions of widespread vulnerability in which the effects of crises are poorly controllable and therefore more devastating, especially in more marginalized contexts [1]. The vulnerability of contemporary society seems having unequivocally demonstrated that the strongly entropic development model based on globalization and accelerated growth creates conditions of systemic imbalance that determine global environmental, social and economic crises, generating risks. These multi scalar and multidimensional poly crisis, may be seen as the key feature of the current scenario, in which pandemic acted not only as detector and accelerator of previously imperceptible aspects and processes already underway, but also as amplifiers of before undervalued issues, unavoidable

today (RIF). The systemic nature of Covid-19 diffusion is indeed a paradigmatic example: from the extreme gravity of the disease, the pervasiveness, and features of health crisis, and the consequent cascade effects, have produced multiple outcomes which went further epidemiological aspects by impacting the anthropic and ecological [2]. According to nonlinear dynamics, the spread has grown exponentially to the point that it affects every aspect of our world, including the most vulnerable and disadvantaged communities, worsening pervasive effects of climate change, that lies on the background [2]. On a global scale, the linear production system based on the quantitative increase of goods and products and on the intensive exploitation of resources has led to a serious ecological and environmental fragility that favors the development of pathogens, highlighting that the common destiny of all human beings is inevitably connected to those of the environment, in a relationship of absolute interdependence [6]. On the individual and collective level, ecological fragility is connected to an extensive social fragility, caused by the persistence of fractures and inequalities that the contemporary economic system is unable to stem, is worsened by global socio-demographic trends, which are connoted by gender diversity, an increasing number of aged population and a persistent flux of immigrants in Western countries [7].

2 Current Issues in Eco-Socio-Technical Changing Scenario

New instances connote the critical scenario of this new normality. The time of the emergency arose as an objective and subjective reality characterized by the coincidence of conditions that are only apparently contradictory, such as distance and proximity, isolation and connection, closure and opening. Distancing and isolation, which represented two of the most severe effects of Covid-19, have imposed alterations of social behaviors commonly regulated by proxemics. The dilatation of the intimate distance and interpersonal communication with family members and loved ones operated a forced unnatural overlapping between personal and social spaces. However, the change of physical interaction did not correspond to an actual social distancing. Conversely, the increased use of internet through social networks have favored the perception of a strong sense of virtual proximity. This allowed, when possible, to continue many of the activities on which the social systems are structured—e.g., schools, universities, working—and contributed to build a network of ‘interpersonal remote relationships’, by which people remained tied to every-day reality, partly shielded from the sense of loneliness and segregation due to lock-down. Ironically, the ensued isolation has generated an intense and generalized ability to connect, be in touch and share, never experienced before. Such trends have gone beyond the need for physical interaction towards the development of strong social bonds that constitute the basis of a real sense of community—a target of ILCs. The resulted ‘social claustrophobia’ due to the need to delimit the houses’ living spaces spurred inhabitants to find new creative forms be in touch with the real world. Public spaces in this scenario assume an essential role, since in many cases they remain the only places in which active life and social interaction may be promoted. Places to rethink however in terms of healthy city, for preventing the onset of diseases, guaranteeing security and safety of citizens, also according with the environmental issue. Rather than mere appurtenant areas, new roles are assumed by balconies, loggias, terraces, courtyards,

gardens, etc. become resources to benefit for people socialization [7]. Similarly, emerging effects of climate change, are already having an impact on everyday life, and this includes health in many forms—i.e., beyond protecting people from weather events to mitigate many of the social determinants for good health, such as livelihoods, equality and access to resources, health and care, and social support. and social support [8].

3 Fragility and Hyper Connection in the Human-Environment-Centered Perspective

In the current scenario fragility and hyper connection emerge as complementary traits connoting directions toward individual and collective sustainable actions can be reoriented, in the light of the paradigm shift that this polycritic era evokes. Fragility involves the theme of care, as the feeling for overcoming the sense of detachment and the loss of meaning, that connotes the lack of respect for what surrounds us [9]. In this perspective, the idea of care goes beyond the private dimension, toward a conscious and responsible participative attitude oriented to protect and support real world, i.e. the other, the community, the environment. Moreover, the care involves the concepts of mutuality and reciprocity. The inter-subjective factors are essential for responding to the current challenges of a multidimensional and interconnected society, where it is no longer possible face problems by isolating and compartmentalizing them. In such scenario, sharing and collaboration become transformative drivers for effectively impact on the socio-technical world. By exchanging knowledge, experiences and resources, flexible and non-hierarchical relational networks can respond in an adaptive way, to these changing arising issues.

4 Inclusive Design for the Emergent Dwelling

Despite the global diffusion of the current interdependent phenomena, not everyone is equally at risk, since available social and economic resources and the qualitative performance of places and services, may improve or worse the population's physical and mental health. In this case, social fragility and inequalities influence the entity of these impacts [3]. In this challenging framework, it is paramount that people, especially fragile ones, are supported to fit the changes that these phenomena continue to generate. Accordingly, architectural, spatial, and functional features of living spaces must respond through an inclusive physical and social environment [4], able to accommodate and sustain fragility and the lifestyle adaptation to this complex scenario. Supported by the adaptive dimension of the relation between inhabitants and places, the concept of emergent dwelling is a reference model for architecture design, since it translates the instances coming from the current eco-socio-technical scenario into expected performance of living places [5]. Adopting a co-evolutive dimension of design of the human habitat, requirements of emergent dwelling are suitable for representing the key conditions which built environment has to assure in its interaction with people. Spaces that have to be inclusive, smart, healthy and user-friendly, for merging people needs with demands coming from social instances, climate issues and environmental protections. A selection of experimental

applications of inclusive design for emergent dwelling in the Southern Italy made by the research group of LEAS Laboratory of University of Naples Federico II is presented as following.

5 Inclusive Dwellings for Inclusive Living Communities: Results of Three Experimental Applications

Flexibility of spaces and generational diversity are the key drivers of a design program for a life-time home, aimed at converting into a multigenerational residential setting an existing building located in the north area of Naples (Italy). Here new functions of social-care and multi-aging living are integrated within the regional health chain, primarily referring, but not limited to, the elderly population, also including the vulnerable users in a life course perspective. Thanks to a co-design process and supported by a multidisciplinary team of physicians, architects ergonomists and social facilitators, environments have been re-conceptualized to be user friendly, for addressing generational diversity. In order to host the changing needs of aged people, students and young migrants flexible solutions have been propose able to create versatile and adaptable settings. Using proactive interventions the building program has been aimed at improving quality of life of occupants, also impacting health and green outcomes through smart environment systems and digital-health solutions. (Fig. 1).

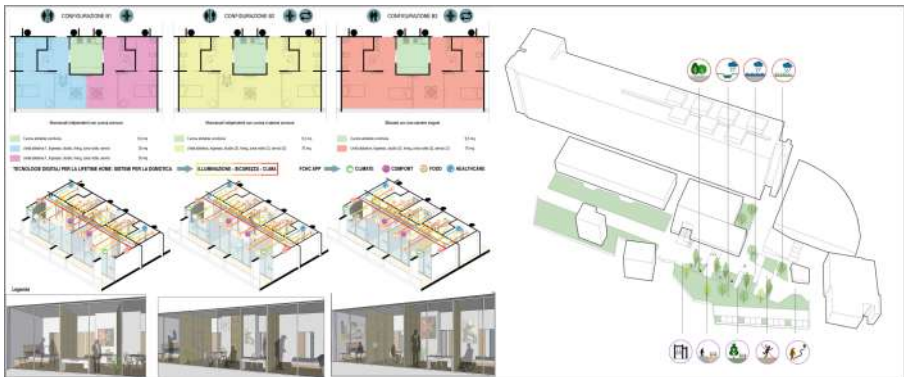


Fig. 1. Details of indoor flexible spaces and climate proof solutions of the Centro Polifunzionale Piscinola-Marianella, Naples (Italy)

Co-housing, low environmental impact of construction process and full accessibility are the main goals of a multifunctional social-welfare center designed by converting a primary school located in Pomigliano d'Arco (Italy). In this case, principles of Universal Design and Design for All have been mainly emphasized for defining a detailed set of inclusive requirements of the spaces oriented to address needs and limitations of fragile people, single-parent and disadvantaged families. Flexible solutions have been proposed to have residential units easy to adapt to the needs of the changing typology of occupants (Fig. 2).

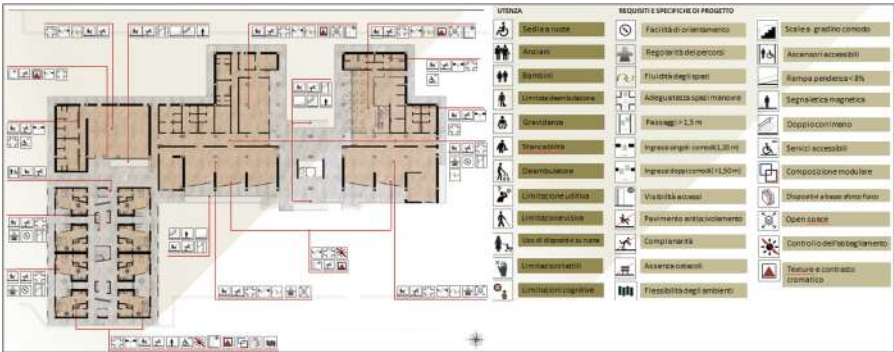


Fig. 2. Layout and the set of inclusive requirements of the multifunctional social-welfare center in Pomigliano d'Arco (NA)

Autonomy and socialization are, finally, the focal points around which the re-functionalization of a center for independent living in the town of Campagna, near Salerno has been developed. Originally design to host physical rehabilitation activities, building functions have been renovated and implemented for multifunctional purposes: flexible co-housing spaces for people with mental diseases and spaces for social activities also open to the community.

The concept of autonomy has been here declined in a wider sense, referring to: people, through the full accessibility and usability of the building; financial sustainability of the hosted activities; energy and resource, with installation of renewable energy and rain water recovery systems; food, with the predisposition of a vegetable garden for food production (Fig. 3).

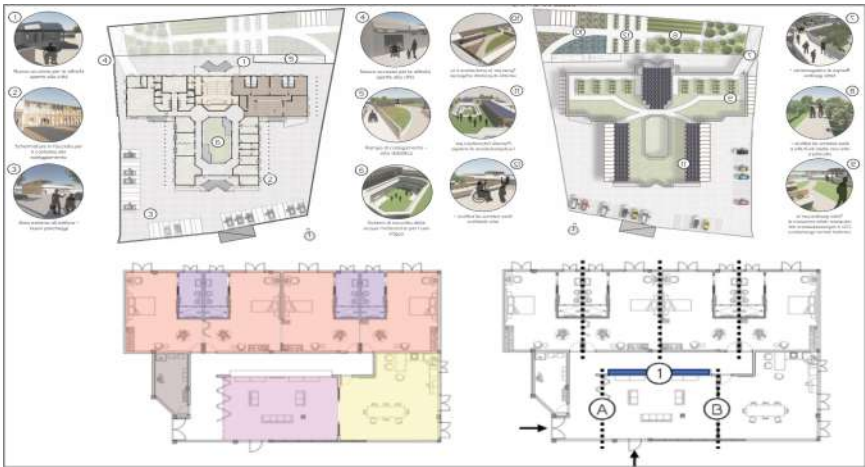


Fig. 3. Layout, visualizations and details of the co-housing flexible spaces of the Center for Independent Living in Campagna (SA)

6 Conclusion

Inclusive design for emergent dwelling introduces an array of topics that are relevant for ILCs. Firstly, the themes of the built environment's responsiveness to different human needs in a lifelong and multigenerational perspective to the spatial scale [6]—i.e. buildings—translate into spatial and technical specifications the needs coming from the application to the building design of the concepts of human variability and diversity, through HCD approaches and the application of Design for All principles and practices. Secondly, the need to ensure living spaces with a low environmental impact as well as 'climate-proof' and 'post-pandemic', ask for the development of environmentally sustainable requirements on the design of living spaces. These indeed entail the reduction of energy needs, a more efficient climate adaptation, higher responsiveness, and self-sufficiency for multi-risk conditions, alongside the use of indoor and outdoor spaces by occupants. Flexible, autonomous, collaborative and cooperative physical and social contexts, able to generate inclusive and interconnected local cultures, capable to generate adaptive reactions to uncertainty, and produce adequate innovations as the current and future realities require.

References

1. Losasso, M.: Interconnected crises and design complexity. *TECHNE—J. Technol. Archit. Environ.* **23**, 7–9 (2022)
2. United Nations Office for Disaster Risk Reduction: Increasing Global Resilience to Systemic Risk: Emerging Lessons from the COVID-19 Pandemic. Office for Disaster Risk Reduction. UNDDR. <https://www.undrr.org/publication/increasing-global-resilience-systemic-risk-emerging-lessons-covid-19-pandemic> (2021)
3. World Health Organization: WHO Housing and health guidelines. World Health Organization, Geneva (2018)
4. Attaianesi, E., Rigillo, M.: Ecological-thinking and collaborative design as agents of our evolving future. *TECHNE—J. Technol. Archit. Environ.* **2**, 97–101 (2021)
5. Attaianesi, E., Illario, M., Rigillo, M.: Emergent dwelling. Requests for designing a human-scale and climate-proof lifetime house. *TECHNE—J. Technol. Archit. Environ.* **24**, 254–263 (2022)
6. Danasekaran, R.: one health: a holistic approach to tackling global health issues. *Indian J. Community Med.* **49**(2), 260–263 (2024)
7. Guzman, J.M.: Past and Future Demographic Trends: Fears, Facts and Policy Implications. CPD 2024. UN Commision on Population and Development (2024)
8. Chiesa, V., Antony, G., Wismar, M., Rechel, B.: COVID-19 pandemic: Health impact of staying at home, social distancing and 'lockdown' measures—A systematic review of systematic reviews. *J. Public Health.* **43**(4), e462–e481 (2021)
9. Dietz, T., Shwom, R.L., Whitley, T.: Climate change and society. *Annu. Rev. Sociol.* **2020**(46), 135–158 (2020)
10. Attaianesi, E.: Fragility and hyperconnection in the post-pandemic scenario by the human-environment centered point of view—Editorial. *Riv. Ital. Ergon.* **23**, I–IV (2021)

Author Index

A

Abendroth, Bettina 187, 208
Abid, Pakeeza 350
Abulaiti, Yilizati 503
Adamy, Jürgen 187
Adelakun, Olayele 338
Ahn, Yubin 350
Alce, Günter 96
Alm, Romy 439
Amaral, Fernando Gonçalves 57
Araco, Carissa M. 231
Ardiyanto, Ardiyanto 390
Arriola, Maria Mikaela A. 231
Attaianese, Erminia 510

B

Backhaus, Claus 33, 439, 461
Bengler, Klaus 161
Bernardo, Ezekiel L. 244, 256, 280
Bispo Neves, Júlio César 45
Bispo, Lucas Gomes Miranda 57
Boccadutri, Luca 96
Bonelli, João 372
Brischetto, Alessia 237
Bruce, Douglas 338

C

Cadilhe, Maria Antónia 76
Caffaro, Federica 96
Cao, Jingjing 105
Cardoso Júnior, Moacyr Machado 445
Carungay, Sheila Mae 250
Cha, Jaehoon 364
Chavez, Denis I. 411
Cho, Minwoo 113
Conde, Jose Antonio T. 244
Cordel, Macario 378
Coronel II, Raymond Christian M. 244
Costales, Justine Bless 250
Cunha, Liliana 76

D

da Silva, Edmar Thomaz 480
De Jose, Beatriz Anne T. 280
de Lima, Marcos Vaz 445
De Paris, Sabine 64
de Souza Rehder, Ivan 480
Delgoulet, Catherine 82
Destyanto, Twin Yoshua R. 3
Dong, Lili 319
Dornéles, Vanessa Goulart 64
Duarte Pena, Eduardo 70
Dycaico, Josh Adrian 256

F

Félix da Costa, Carlos Eduardo 372
Feng, Shanshan 119
Fleury, Sylvain 450
Foster, Elissa 338
Frenzel, Jan-Philipp 187
Friedrich, Felix 208
Furuhata, Naoki 424, 431

G

Galao, Rapeh Jahaziel L. 331
Gonzales, Reinabelle Angelique Y. 231
Grabbe, Niklas 161
Grepo, Lorelie C. 10
Guan, Ng Yee 51
Guo, Yichen 262
Gutierrez, Alma Maria Jennifer 378
Gutiérrez-Hernández, Mónica Gabriela 325

H

Han, Su 119
Hans, Ole 187
Harun, Hezery 51
He, Canqun 268
He, Dengbo 136
Hu, Shan 17
Huang, Chunxi 136
Hwang, Dongwook 113

I

Iacono, Ester 274
 Im, Youngjae 286, 306
 Ito, Nana 424, 431

J

Jeon, Yeahyoung 467
 Jiang, Weisheng 143
 Jiang, Weitao 17
 Jo, Suhwan 299
 Joisten, Philip 208
 Jung, Woosung 299
 Junior, Moacyr Cardoso Machado 480

K

Kang, Byung-Gyoo 396
 Kang, Xinyue 292
 Kang, Yunah 286
 Kapogiannis, Georgios 396
 Kasuya, Shugo 486
 Kawarai, Ryo 199
 Kehmann, Maximilian 33
 Kevin, Guelle 130
 Khangnba, Sanjram Premjit 454
 Kim, Daehui 155
 Kim, Dongha 312
 Kim, Gyuhee 312
 Kim, Hongtae 149
 Kim, Jonghan 299
 Kim, Jung Hyup 473
 Kim, Juseong 364
 Kim, Manseo 299
 Kim, Saeyeon 312
 Kim, Sungju 306
 Kim, Woojoo 312
 Kim, Yeonji 312
 Kim, Yura 312
 Kimura, Shoko 503
 Koike, Takaharu 199
 König, Simon 33, 439, 461
 Kudo, Keisuke 467
 Kuklinski, Steffen 33
 Kurnianingtyas, Chandra Dewi 3
 Kwon, Kyungwoo 356

L

Lagrimino, Jonathan 180
 Lee, Hojeong 467
 Li, Donglin 119

Li, Fan 119
 Lin, Chiuhsiang Joe 493
 Liu, Shengyuan 396
 Liu, Yankuan 319
 Liu, Yuchen 161
 Liu, Yung Ching 173
 Liu, Zhongqi 143
 Lopez, Guillaume 503

M

Maldonado-Macías, Aide Aracely 325
 Marques, Henrique Costa 445
 Matsui, Saki 424
 Mebarki, Bouhafs 39, 193
 Mebarki, Imed Eddine 39
 Menezes O. Sousa, Mariana 372
 Mertens, Alexander 418
 Mohanty, Siddarth 473
 Mohd, Nor Haslina 51
 Mokdad, Mohamed 39, 193
 Mont'Alvão, Claudia 372
 Montague, Enid 338
 Morteza pour, Alireza 167
 Mosafarchi, Saeedeh 167
 Mostowfi, Sara 473

N

Naddeo, Alessandro 167
 Namkung, Kiechan 344, 356
 Natacha, Metayer 130
 Nicolas, Erdinger 130
 Nien, Po Chun 173
 Nitsch, Verena 418

O

Oh, Seung 306
 Oprean, Danielle 473
 Ordinario, Joshua J. 331

P

Pak, Kicheol 364
 Pan, Younghwan 384
 Park, Chaewon 312
 Park, Jaehyun 155, 222, 350
 Park, Jaeyeon 256
 Park, Seonghyeok 467
 Park, Sung 299
 Pasa, Mariana Maciel 64
 Pasia, Danielle Alexandria A. 280

Pereira, Cláudia 76
 Piombino, Virginia 274
 Pistolesi, Mattia 237
 Polancos, Ronaldo V. 231
 Potter, Gabriela Pérez 325
 Pütz, Sebastian 418

Q

Quaresma, Manuela 372
 Quevedo, Francispito 250

R

Rahman, Anita Abdul 51
 Rajagopal, Deepthi 338
 Ramos, Christine Diane 378
 Rau, Pei-Luen Patrick 292, 319
 Razzera, João Victor 70
 Reyes, Angelo Gabriel A. 244
 Reyes, Janina Elyse A. 331
 Ribeiro Menezes, Henrique 45
 Ricardo, Aaron Raphael 250
 Rinaldi, Alessandra 180
 Rodas, Giannina Anaïs 256
 Rodeghiero Neto, Italo 57
 Rodriguez Vidal, Mario Cesar 45
 Roger, Lauriane 82
 Rong, Lingda 17

S

Sa, Jieun 364
 Safflor, Charmine Sheena 280
 Sagawa, Ken 424, 431
 Sahagun, Antonio Lorenzo L. 231
 Sakata, Reiko 424, 431
 Santillan, Vanessa L. 331
 Santos, Marta 76
 Schimit, Thiago Arndt 57
 Schwindt-Drews, Sarah 187
 Semmani, Mourad 193
 Seo, Kangwon 473
 Seo, Yeongju 344
 Seong, Sujin 350
 Shi, Pengjie 268
 Shibata, Takashi 486
 Shimada, Atsushi 424
 Shin, Junghoon 356
 Siboro, Benedikta A. H. 493
 Song, Yunjeong 364
 Soulet, Annaëlle 82

Ssemakula, Paul 467
 Suarez, Merlin Teodosia 378
 Sun, Xiaomeng 268
 Suzuki, Daisuke 199

T

Tamrin, Shamsul 51
 Tangsoc, Jazmin C. 256, 280
 Tangsoc, Jazmin 244
 Theobald, Nina 208
 Torres, Yaniel 88
 Tosi, Francesca 274
 Trapsilawati, Fitri 390

U

Ubaradka, Anantha 454
 Ujike, Hiroyasu 424, 431

V

Vidal, Hannagene C. 331
 Vigoroso, Lucia 96
 Villani, Emilia 480

W

Wang, Fang 473
 Wang, Fanxuan 292
 Wang, Peggy 105
 Wang, Ting Hsuan 173
 Wang, Yuming 384
 Wang, Zuyao 396
 Wang, Zuyuan 136
 Watanabe, Hiroshi 424, 431
 Weber Righi, Angela 64, 70
 Widaningtyas, Rachma 390
 Wu, Pei Hsuan 173

X

Xiong, Shuping 467

Y

Yamagishi, Misako 215
 Yamin, Putra A. R. 222
 Yang, Tianlun 396
 Yanhui, Wu 396
 Yap, Caila Jean C. 244
 Yoo, Chaemoon 384
 Yu, Cedric Nathaniel 256
 Yu, Ching-Yun 473
 Yu, Mengli 262

Z

Zaoui, Brahim	39	Zhang, Yuchang	396
Zhang, Bingxin	17	Zhenyu, Hu	396
Zhang, Jitong	17	Zhou, Jia	105
Zhang, Kuan	384	Zhou, Qianxiang	143
Zhang, Yanjie	119	Zhou, Zhenkun	262
		Zuo, Yanling	105