Visual Similarity Assessment for Product Aesthetic Properties Using Single Reference Training

SEOK YOUNG HWANG[1], JUSEONG KIM[1], and KICHEOL PAK[1]
[1]International Design School for Advanced Studies, Hongik University, Seoul 04066, South Korea

Corresponding author: Seok Young Hwang (e-mail: y0ung@g.hongik.ac.kr).

ABSTRACT Traditional product aesthetic evaluation requires extensive preference labeling for each design, limiting scalability and increasing costs. This study presents a novel single reference training approach that enables aesthetic preference assessment through visual similarity to a single highly preferred product. Using 30 home audio speakers, we collected aesthetic preference ratings from 44 participants on a 7-point Likert scale. Four deep learning approaches—Pre-trained CNN, Auto-encoder, Siamese Network, and Triplet Network—measured visual similarity between the most preferred reference products and remaining samples. Results demonstrated significant correlations between visual similarity and aesthetic preference: the Triplet Network achieved Pearson correlation coefficient $r = 0.448$ ($p = 0.013$), while Pre-trained CNN approach yielded $r = 0.478$ ($p = 0.008$). After filtering high-variance products, correlations substantially improved to $r = 0.738$ ($p < 0.001$). Principal component analysis of embedding vectors revealed interpretable aesthetic dimensions, with specific components significantly correlating with novelty ($r = 0.51 - 0.61$), harmony ($r = 0.36 - 0.37$), dynamics ($r = 0.55 - 0.65$), and complexity ($r = 0.42 - 0.56$). Interaction models of multiple principal components increased explanatory power ($R^2 = 0.4065$), demonstrating that aesthetic preferences emerge from complex relationships among visual features. The findings prove that single reference training effectively extracts interpretable aesthetic properties from embedding spaces without explicit labeling. Our results suggest that highly preferred products contain multiple aesthetic preference-related properties, and deep learning models can successfully extract features corresponding to these properties. This approach provides an alternative aesthetic assessment from labor-intensive individual evaluation to efficient similarity-based inference, offering a methodology with reduced data requirements for product design evaluation and enabling evaluation of new designs.

INDEX TERMS design evaluation, design engineering, aesthetics, deep learning, product design

I. INTRODUCTION

A. AESTHETICS IN PRODUCT DESIGN AND CURRENT EVALUATION METHODS

A good product is usually defined as one that includes a sufficient level of technology to satisfy consumers and a visually pleasing exterior design. Products that meet these elements influence consumers' purchasing decisions. As the pace of technological advancement has accelerated, the technological level of many products encountered in daily life has become standardized, contributing to a wider range of choices for consumers regarding good products [1].

As the technological differences between products gradu-ally decrease, it becomes increasingly important to differentiate products in aspects other than technology [2]. Hekkert [3] emphasized aesthetic appeal as one of these differentiating factors, and Raghubir & Greenleaf [4] empirically confirmed that the proportion of product packaging affects consumer cognition, preference, and purchase intention.

The theoretical foundations of aesthetic preference research include Berlyne's [5] inverted U-shaped model and Loewy's [6] M.A.Y.A (Most Advanced Yet Acceptable) principle. Berlyne explored the relationship between complexity, novelty, uncertainty of visual stimuli and hedonic responses, while Loewy suggested that the balance between innovation

and familiarity is important for product success. Based on these models, Thurgood et al. [7] empirically verified the interaction effect of typicality and novelty in product design, and Hung & Chen [8] experimentally demonstrated the impact of various dimensions of novelty on aesthetic preference in chair designs. Other researchers have explored the association between aesthetic preference and various product properties such as complexity, harmony, unity, timeliness, and novelty [3], [8]–[11]. Table 1 presents the dimensions of product aesthetics as defined by Ellis [12] and Brunel [13].

While these theoretical frameworks provide a foundation for understanding aesthetic properties, evaluating them in practice remains challenging. Previous studies have explored various approaches to measuring aesthetic preferences of products. Subjective evaluation methods such as consumer surveys and interviews have been widely used, including efforts to conceptualize and measure the importance of visual aesthetics in consumer decision-making [14]. To complement these approaches, objective techniques like Eye-Tracking have been introduced, enabling researchers to analyze visual attention through physical responses [15], [16]. Other studies have attempted to derive product-related features by mining consumer-generated content such as online reviews [17]. More recently, Liu et al. [18] proposed an aesthetic measurement approach that attempts to quantify design elements, while Lai et al. [19] presented a method incorporating heterogeneous data for evaluating product appearance. However, gaze concentration does not necessarily indicate aesthetic preference, and limitations still exist in quantitatively extracting visual properties and establishing a direct connection between such properties and user preferences.

B. DEEP LEARNING APPROACH IN DESIGN EVALUATION

Deep learning is a methodology that effectively learns complex relationships between inputs and outputs by accurately allocating credit for learning performance across multiple computational stages within artificial neural networks [20]. These computational stages generally consist of nonlinear functions, enabling high-dimensional representation learning and inference through deep neural network structures.

The advancement of deep learning technology has opened new possibilities for product design evaluation. Krizhevsky et al. [21] demonstrated the superior image classification performance of CNNs, and Sharif Razavian et al. [22] proved that features extracted from Feature Extraction models show excellent performance in various recognition tasks. In the field of product design evaluation, Wu et al. [23] developed a model that predicts design award outcomes from product images using DCNN, while Burnap et al. [24], [25] proposed a model that predicts product aesthetic preferences based on a large image dataset labeled with preference scores through a human-machine hybrid approach. However, existing deep learning approaches have mostly been limited to predicting consumer preferences by training on numerous images or predicting preference scores by labeling each product with preference scores. This approach involves inputting a photo

(x) to predict a preference score ($\hat{y}$) and confirming the correlation (r) with the actual preference score (y).

$$\hat{y} = f(x), \quad r = \text{Corr}(y, \hat{y}) \tag{1}$$

Rather than relying on existing approaches that require extensive preference labeling and large training datasets, our research explores an alternative framework based on single reference training. While previous methods typically focus on supervised learning with comprehensive aesthetic databases, our approach examines visual similarity analysis combined with aesthetic property extraction through principal component analysis. This methodological difference leads to several contributions to the field of aesthetic evaluation: First, we explore a single reference training approach that reduces the need for extensive preference labeling, which may help decrease evaluation costs in certain contexts. Second, we provide interpretable analysis of aesthetic properties through principal component analysis of embedding vectors, examining how deep learning models capture aesthetic dimensions such as novelty, harmony, dynamics, and complexity. Third, we investigate whether visual similarity to highly preferred products can predict aesthetic preferences across different product categories. Finally, our approach offers potential applicability to new product evaluation with limited additional training requirements.

II. METHODOLOGY

To investigate the relationship between visual similarity and aesthetic preference through single reference training, we developed a comprehensive experimental framework. Our approach encompasses stimuli selection, participant evaluation, deep learning model development, and similarity measurement techniques.

A. OUR APPROACH

This research departs from conventional deep learning approaches by training models using only a single product with the highest preference rating. Our methodology investigates whether visual similarity to this reference product correlates with aesthetic preference, and when significant relationships are found, we analyze the embedding vectors to identify aesthetically relevant properties such as novelty, harmony, dynamics, and complexity. Therefore, when visual similarity is expressed as $S^{\text{Model}}(j, \text{ref})$ and aesthetic preference as $P(j)$, the relationship between these two factors can be mathematically represented as follows:

$$P(j) = \alpha_{\text{ref}} \times S^{\text{Model}}(j, \text{ref}) + \beta_{\text{ref}}, \quad j \in \{1, 2, \dots, 30\} \tag{2}$$

where α_{ref} is the correlation coefficient representing the strength of the relationship between visual similarity and aesthetic preference, and β_{ref} is the intercept term of the linear regression model.

TABLE 1. Dimensions of product aesthetics [12], [13]

Properties	Content
Simplicity/Complexity	The subjective difficulty in evaluating the stimulus due to the number of visual parts, and the degree of differentiation of these different parts.
Harmony	The similarity or concordance among the various parts of a product's visual design, with respect to elements like shape, size, and color. It can also refer to the degree to which the stimulus fits with its surroundings.
Balance	The sense of equilibrium can be influenced by the shape and relative locations of the visual design, the apparent spatial depth of the design elements and the degree of isolation of the parts of the design.
Unity	Degree of oneness of the design, i.e. the degree to which all the elements of the stimulus get combined to create a whole.
Dynamics	The degree to which there exists a perception of motion and tension in the design of the stimulus.
Timeliness/Style	The subjective perception of the extent to which the design represents current fashionable trends vs. traditional, old-fashioned properties. It is based on the recurrent features of the design.
Novelty	Perception that the product design and aesthetics are new to the world. That they constitute a new experience for the focal consumer.
Gestalt	The integrated aesthetic evaluation. The evaluation of the product as a whole, without a necessary analysis of each subpart. The whole and the sum of the subparts may be different.

B. RESEARCH PROCESS

This research employs the following approach to explore the relationship between aesthetic preference for products and similarity:

1) **Selection of experimental stimuli and collection of data**: Home audio speakers were selected as experimental stimuli, and data was collected through participants' evaluations of aesthetic preference and visual properties. Electronic products are particularly suitable for this study's purpose as they represent a product category where aesthetic plays an important role in purchase decisions [26].

2) **Construction of various deep learning models**: Based on aesthetic preference evaluation results, the most preferred product was identified and used as a reference to build various deep learning models. Four different methodological approaches—Pretrained CNN, Auto-encoder, Siamese Network, and Triplet Network—were utilized to compare the effects of feature learning and similarity measurement.

3) **Similarity measurement and comparative analysis**: The trained deep learning models were used to measure the similarity between the reference product and the remaining products, and these measurements were comparatively analyzed with aesthetic preference evaluation results. In particular, this study explores whether the features of a single reference product alone can predict the aesthetic preference of other products.

4) **Properties analysis**: For models where a relationship between similarity and preference was confirmed, deeper analysis was conducted to explore which properties influence aesthetic preference. This helps to understand how property dimensions like novelty is related to the features learned by the model.

This approach is an attempt to connect the subjectivity and objectivity of aesthetic evaluation, presenting a new evaluation methodology utilizing artificial intelligence technology in the field of product design. It may open the possibility of predicting the aesthetic preference of other products using only a few highly preferred products, which could present a new quantitative approach to product design evaluation and development processes.

C. STIMULI

Thirty speakers were selected for the survey. Each speaker was chosen from different brands (companies), and products with various design configurations were selected through Focus Group Interview (FGI). The selected products are shown in Table 2.

A survey was conducted on aesthetic preference for each of the 30 speakers using a Likert scale from 1 to 7. Additionally, according to Berlyne's [5] theory, one of the prominent theories on aesthetic preference and visual stimulation mentioned above, participants were also asked to rate novelty from 1 to 7. The novelty scores will be used for correlation analysis with embedding vectors later in the study.

D. DESIGNER VS. NON-DESIGNERS AND AESTHETICS - NOVELTY

A total of 44 answers (16 males & 28 females; age range: 20–30s) from Hongik University were collected. The sample consisted of 34 design majors and 10 non-design majors. First, the study examined whether there were differences in the standard for judging aesthetic preference between design majors and non-design majors.

A t-test was conducted on the differences between the preference scores of design majors and non-design majors for each product. The analysis results for aesthetic preference scores were t-statistic = -0.04, $p = 0.9657$, and similarly, the analysis results for novelty scores were t-statistic = -1.44, p-value = 0.1514. Accordingly, confirming that there was no difference in responses between design majors and non-design majors, and the response data from all participants regardless of major was used in subsequent research. The survey data are shown in Tables 3, 4, and 5. Table 3 presents

TABLE 2. Images of selected speakers

No.	Image	No.	Image	No.	Image
1		11		21	
2		12		22	
3		13		23	
4		14		24	
5		15		25	
6		16		26	
7		17		27	
8		18		28	
9		19		29	
10		20		30	

FIGURE 1. Relationship between novelty and aesthetic preference

the comparison of mean scores between design majors and non-design majors for both aesthetic preference and novelty ratings, along with t-test statistics confirming no significant differences between groups. Table 4 displays the descriptive statistics for aesthetic preference scores across all 30 products, including mean, variance, and standard deviation values. Products 5 and 13 achieved the highest preference scores and were selected as reference products for subsequent deep learning model training. Table 5 presents the corresponding novelty evaluation results for each product, which will be used for correlation analysis with embedding vectors in later sections of the study.

Additionally, as participants were also asked to provide novelty scores for each speaker, the study aimed to determine whether the product category follows the inverted U-shaped model, showing a quadratic relationship with novelty. If this relationship is significant and if a meaningful correlation between similarity and preference is confirmed, it would imply that the features extracted by the deep learning models may be related to elements that determine novelty. Conversely, if there is no significant relationship between novelty and aesthetic preference, it suggests the possibility that the properties related to aesthetic preference for speakers and such product features may not be limited to novelty alone.

Linear, quadratic, and cubic polynomial relationship analyses were conducted based on aesthetic preference and novelty scores, with results shown in Table 6. The evaluation metrics used are: R^2 (coefficient of determination) measuring the proportion of variance explained by the model, MSE (Mean
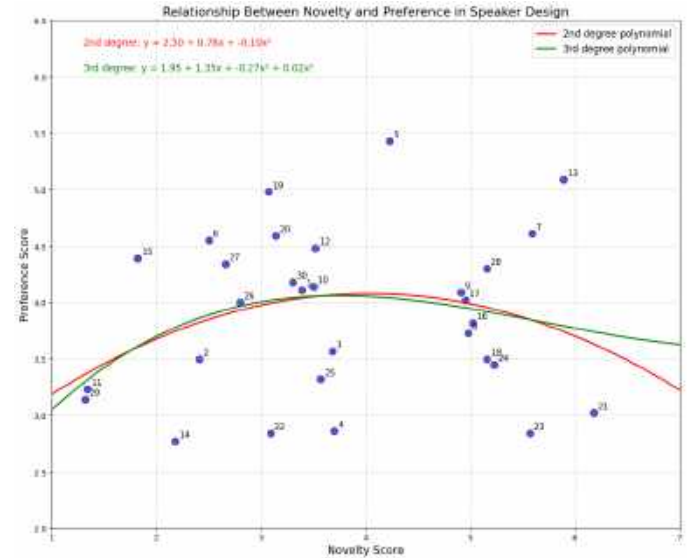
Squared Error) representing the average squared differences between predicted and actual values, and MAE (Mean Absolute Error) indicating the average absolute differences between predictions and observations. No significant relationships were derived from any of the models. However, compared to the linear model, the quadratic model showed an improvement in R^2 from 0.0088 to 0.0701 and a 6.18% decrease in MSE, while comparing the quadratic and cubic models showed only a 4.14% increase in R^2 and a 0.31% decrease in MSE. This indicates that the quadratic model is the most suitable explanatory model. An exploration of why Berlyne's inverted U-shaped model did not show explanatory power for these products will be addressed shortly after. Figure 1 shows the relationship between novelty and aesthetic preference.

E. CREATING DEEP LEARNING MODELS

The survey results showed that products #5 and #13 had the highest preference scores at 5.43 points (var = 2.02) and 5.09 points (var = 1.95) respectively. To account for cross-validation between models and assuming that different visual elements from the two products' features could influence aesthetic preference, training was conducted not only with product #5 (ranked first in aesthetic preference) but also with product #13 (ranked second) separately.

All product images were preprocessed using consistent procedures to ensure reproducibility across different model architectures. Images were resized to 224 x 224 pixels using TensorFlow's load_img function with target_size parameter, and converted to arrays using img_to_array [29]. Pixel values were normalized to the range [0,1] by dividing by 255. No data augmentation techniques were applied to preserve the authentic visual characteristics essential for aesthetic evaluation.

We implemented and compared four deep learning-based

TABLE 3. Survey Evaluation

	Design major mean	Non-design major mean	t-statistic	p-value
Aesthetic preference	3.90	3.90	-0.04	0.9657
Novelty	3.75	3.93	-1.44	0.1514

TABLE 4. Scores of aesthetic preferences by product

No.	Mean	Var	Std dev	No.	Mean	Var	Std dev
1	4.11	2.38	1.54	16	3.82	2.38	1.54
2	3.50	2.53	1.59	17	4.02	2.95	1.72
3	3.57	2.34	1.53	18	3.50	2.44	1.56
4	2.86	2.12	1.46	19	4.98	2.12	1.45
5[a]	5.43	2.02	1.42	20	4.59	1.92	1.39
6	4.55	2.16	1.47	21	3.02	2.67	1.64
7	4.61	2.99	1.73	22	2.84	2.46	1.57
8	3.73	2.76	1.66	23	2.84	2.65	1.63
9	4.09	3.57	1.89	24	3.45	3.42	1.85
10	4.14	1.98	1.41	25	3.32	1.94	1.39
11	3.23	2.88	1.70	26	4.00	2.05	1.43
12	4.48	1.88	1.37	27	4.34	2.70	1.64
13[b]	5.09	1.95	1.39	28	4.30	3.28	1.81
14	2.77	2.41	1.55	29	3.14	2.77	1.66
15	4.39	2.43	1.56	30	4.18	3.04	1.74

[a] Product with highest preference score

[b] Product with second highest preference score

TABLE 5. Scores of novelty by product

No.	Mean	Var	Std dev	No.	Mean	Var	Std dev
1	3.29	2.61	1.62	16	5.02	1.33	1.15
2	2.41	2.15	1.47	17	4.95	2.23	1.49
3	3.68	2.18	1.47	18	5.16	2.65	1.63
4	3.70	2.68	1.64	19	3.07	1.88	1.37
5	4.23	1.71	1.31	20	3.14	2.03	1.42
6	2.50	2.91	1.70	21	6.18	1.64	1.28
7	5.59	1.16	1.06	22	3.09	3.43	1.85
8	4.98	3.28	1.81	23	5.57	1.88	1.37
9	4.91	2.88	1.70	24	5.23	2.60	1.61
10	3.50	2.07	1.44	25	3.57	1.83	1.35
11	1.34	0.46	0.68	26	2.80	1.33	1.15
12	3.52	2.12	1.45	27	2.66	1.49	1.22
13	5.89	0.85	0.92	28	5.16	2.37	1.54
14	2.18	2.48	1.57	29	1.32	0.36	0.60
15	1.82	1.18	1.08	30	3.30	1.93	1.39

TABLE 6. Model Evaluation

	R^2	MSE	MAE
Linear	0.0088	0.492	0.60
Quadratic	0.0701	0.462	0.549
Cubic	0.0730	0.460	0.544

approaches for measuring image similarity. Since this re-

search prioritizes the exploration of relationships between aesthetic preference and visual properties over the development of highly robust models for reference products, both the choice of deep learning architectures and hyperparameter settings were based on researcher discretion rather than systematic optimization or rigorous selection criteria. As the primary objective is to investigate whether meaningful relationships exist between these factors, the focus remains on identifying potential correlations rather than achieving optimal model performance. As stated previously, no separate data such as preference scores were labeled, and only the products themselves were trained. The theoretical background and operating principles of each model are as follows:

1) Pre-trained CNN
The Pre-trained CNN model utilizes pre-trained Convolutional Neural Networks (CNN) to extract feature representations from images without additional training. This study adopted the ResNet50 architecture pre-trained on the ImageNet dataset [27]. The ResNet50 model consists of 50 layers with residual connections, configured with weights='imagenet', include_top=False, and pooling='avg' parameters. The final classification layer was removed, and the global average pooling layer output was used directly, generating 2048-dimensional feature vectors for each input image. Visual similarity between images is quantified by measuring the cosine similarity between these extracted feature vectors.

2) Auto-Encoder
The Auto-encoder is an unsupervised learning neural network consisting of an encoder and a decoder architecture. The encoder follows a convolutional structure with three encoding blocks: Conv2D(32, 3x3) - MaxPooling2D(2x2) - Conv2D(64, 3x3) - MaxPooling2D(2x2) - Conv2D(128, 3x3) - MaxPooling2D(2x2), reducing the 224x224x3 input to a compressed latent representation. The decoder mirrors this structure using transposed convolutions: Conv2D(128, 3x3) - UpSampling2D(2x2) - Conv2D(64, 3x3) - UpSampling2D(2x2) - Conv2D(32, 3x3) - UpSampling2D(2x2) - Conv2D(3, 3x3, activation='sigmoid'), reconstructing the original image dimensions. All convolutional layers use ReLU activation except the final output layer which uses sigmoid activation to ensure pixel values remain in the [0,1] range.

3) Siamese Network
The Siamese Network is designed to directly learn similarity relationships between two images through a twin ar-

chitecture consisting of two parallel neural networks that share identical weights. Each branch utilizes a ResNet50 backbone (without pre-trained weights, trainable=False) followed by GlobalAveragePooling2D, Dense(1024, activation='relu'), and Dense(128, activation='relu') layers, producing 128-dimensional feature vectors. The L1 distance (Manhattan distance) between these feature vectors is computed and fed into a final Dense(1, activation='sigmoid') layer for binary similarity classification. The network is trained using binary cross-entropy loss to distinguish between similar (same category) and dissimilar (different category) image pairs.

4) Triplet Network

The Triplet Network employs a metric learning approach that learns from triplets of images: an anchor image, a positive image (from the same category as the anchor), and a negative image (from a different category). The network architecture consists of three identical branches sharing weights, each comprising a ResNet50 backbone (trainable=False) followed by GlobalAveragePooling2D, Dense(1024, activation='relu'), Dense(512, activation='relu'), and Dense(128) layers. The final 128-dimensional embeddings are L2-normalized to unit length to ensure stable distance computations. The triplet loss function with margin $\alpha = 0.3$ minimizes the distance between anchor-positive pairs while maximizing the distance between anchor-negative pairs, creating a semantically meaningful embedding space where visually similar products cluster together.

F. IMAGE TRAINING

For products #5 and #13, the four learning approaches described above were implemented, creating a total of 10 models. Approximately 35 images were used for each, with examples shown in Table 7.

Images for training were carefully curated to ensure consistent quality and minimize noise that could interfere with aesthetic feature learning. Images were selected from official product websites and professional studio photography, prioritizing clean backgrounds with minimal visual distractions. Images were collected from different angles and distances to capture comprehensive visual characteristics while maintaining authentic product representation. No data augmentation techniques were applied to preserve the genuine aesthetic properties essential for preference evaluation. Model training employed the Adam optimizer with a learning rate of 0.0001 for all trainable models and validation splits of 0.2 [31].

Training was conducted for each model under the conditions shown in Table 8. The Triplet Network used batch 16, epoch 25; the Siamese Network used batch 16, epoch 20; the Auto-encoder used batch 16, epoch 50 [35]. Throughout all training processes, early stopping was set to prevent model overfitting, configured to stop training if there was no loss improvement greater than 0.001 during a specified patience period [36]. For Pre-trained CNN, there were no separate batch and epoch settings as it utilized an existing learning

TABLE 7. Example images used for training

	Image 1	Image 2	Image 3	Image 4	Image 5
No. 5					
No. 13					

TABLE 8. Batch and epochs for each model

Model	Batch	Epoch
Triplet Network	16	25
Siamese Network	16	20
Auto-encoder	16	50
Pre-trained CNN	-	-

TABLE 9. Stopped epoch and val loss of No.5 training model

Model	Epoch	Val loss
Triplet Network	6/25	< 0.001
Siamese Network	15/20	0.001
Auto-encoder	37/50	0.008
Pre-trained CNN	-	-

TABLE 10. Stopped epoch and val loss of No.13 training model

Model	Epoch	Val loss
Triplet Network	6/25	< 0.001
Siamese Network	15/20	0.001
Auto-encoder	46/50	0.005
Pre-trained CNN	-	-

architecture [22], [37]. In cases where negative images were needed, images from the 29 products among the 30 products (excluding the product being trained) were utilized. The ending epoch for each training and the measured loss are shown in Tables 9 and 10.

G. SIMILARITIES OF EACH MODEL

After the models were generated, the similarity between model #5 and the remaining products, as well as the similarity between model #13 and the remaining products, was examined. The results are shown in Tables 11 and 12. For this examination, 15 images of each product, including the datasets of products #5 and #13 which would serve as the maximum similarity reference points, were tested against each deep learning model.

Similarity scores were calculated using model-specific methods: cosine similarity between feature vectors for Pre-trained CNN, inverse reconstruction error for Auto-encoder, inverse L1 distance for Siamese Network, and inverse Euclidean distance in embedding space for Triplet Network. All similarity values were converted to percentage scores (0-

100%).

III. RESULTS

We present our experimental findings in three main areas: correlation analyses between image similarity and aesthetic preference, principal component analysis of embedding vectors, and exploration of relationships with various aesthetic properties.

A. CORRELATION BETWEEN IMAGE SIMILARITY AND AESTHETIC PREFERENCE

These similarities were used to analyze the correlation between preference scores from the survey and similarity through Pearson and cosine similarity, with results shown in Tables 13 and 14. Since product similarity ranges from 0 to 100% and preference scores range from 1 to 7, cosine similarity tests were conducted to analyze the directionality of each data vector [39]. Permutation tests were done with 1000 iterations.

For the model trained on product #5, the Triplet Network model showed a significant positive correlation with $r = 0.448$, $p = 0.013$, and the cosine similarity yielded a result of $\cos(\theta) = 0.983$, $p = 0.012$.

$$P_5(j) = \alpha_5 \times S_{5\text{Triplet}}(j, 5) + \beta_5, \quad j \in \{1, 2, \ldots, 30\} \quad (3)$$

where $\alpha_5 \approx 0.448$ ($p = 0.013$), $\alpha_5 > 0$

The correlation coefficient α_5 was obtained from Pearson correlation analysis between the Triplet Network similarity scores and aesthetic preference ratings for all 30 products, while β_5 represents the intercept term.

For the model trained on product #13, the Pre-trained CNN model showed a significant correlation with $r = 0.478$, $p = 0.008$, and the cosine similarity yielded a result of $\cos(\theta) = 0.988$, $p = 0.002$.

$$P_{13}(j) = \alpha_{13} \times S_{13\text{FE}}(j, 13) + \beta_{13}, \quad j \in \{1, 2, \ldots, 30\} \quad (4)$$

where $\alpha_{13} \approx 0.478$ ($p = 0.008$), $\alpha_{13} > 0$

Similarly, α_{13} was derived from Pearson correlation analysis between the Pre-trained CNN similarity scores and aesthetic preference ratings, with β_{13} as the intercept term. Therefore, for product #5 in the Triplet Network model and for product #13 in the Pre-trained CNN model, we successfully identified significant linear relationships between visual similarity to the reference products and the aesthetic preference scores of the remaining products. Figures 2 and 3 show the correlation plots.

After this, assuming that products with high variance in aesthetic preference or novelty scores show polarization due to respondents' personal preferences, a reanalysis was conducted after excluding 9 products (number 6, 7, 8, 9, 17, 22, 24, 28, and 30) that had variance of 2.9 or higher in aesthetic preference and novelty scores, defined as $J' = J \setminus \{6, 7, 8, 9, 17, 22, 24, 28, 30\}$. The results are shown in
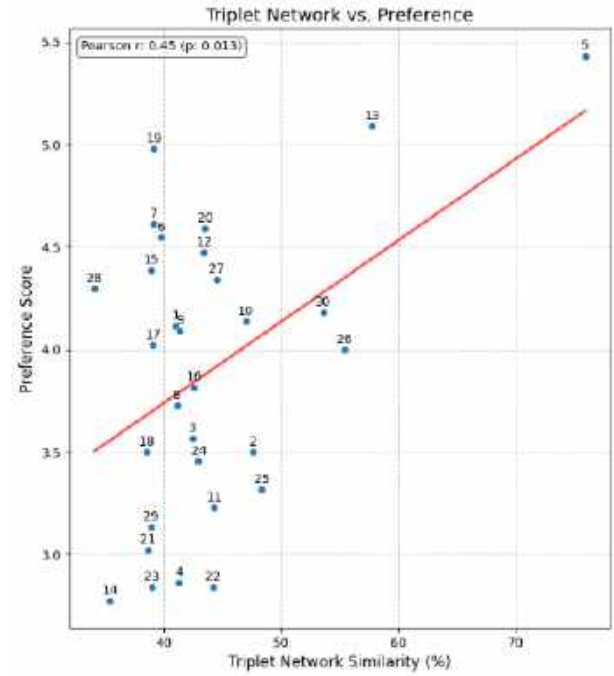

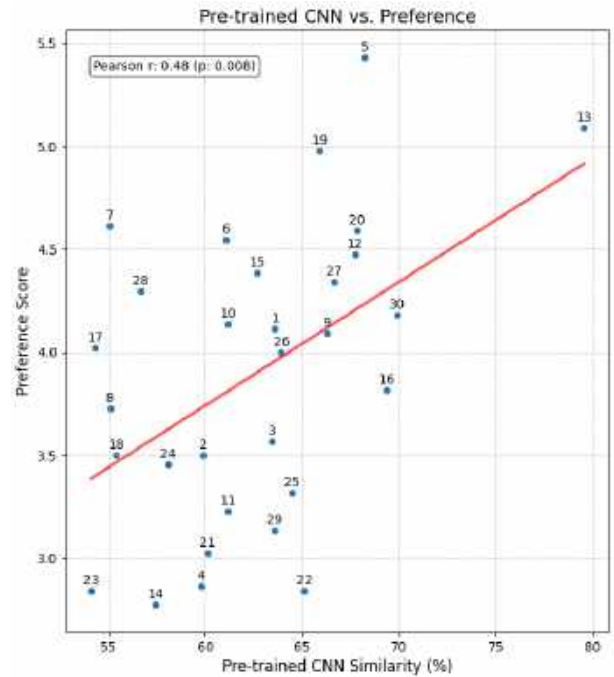**FIGURE 2.** Similarity (%) and aesthetic preference for No.5 Triplet Network

FIGURE 3. Similarity (%) and aesthetic preference for No.13 Pre-trained CNN

Tables 15 and 16. As can be seen in the table, it was discovered that the correlation coefficients with aesthetic preference increased and p-values decreased across the models. Figures 4 and 5 show the filtered correlation plots.

TABLE 11. Similarity results No.5 models vs. 29 speakers (%)

No.	Triplet Network	Siamese Network	Auto-encoder	Pre-trained CNN
5	75.88	95.22	75.65	81.32
1	41.00	87.74	66.71	55.57
2	47.63	93.49	72.71	69.30
3	42.52	90.31	45.28	65.35
4	41.30	91.90	71.70	65.33
6	39.80	89.07	30.21	58.18
7	39.19	89.68	70.12	57.03
8	41.15	91.53	62.84	61.84
9	41.36	91.63	55.95	63.32
10	46.98	88.00	30.01	63.90
11	44.33	89.04	72.57	57.08
12	43.46	87.26	76.99	58.25
13	57.72	91.94	76.79	66.86
14	35.45	86.11	39.91	50.19
15	38.96	90.05	68.15	59.14
16	42.61	92.39	46.00	63.03
17	39.12	89.00	74.96	54.93
18	38.56	89.84	58.32	59.70
19	39.20	87.73	65.83	59.09
20	43.55	89.92	78.04	64.82
21	38.70	87.70	80.36	57.26
22	44.23	88.94	42.55	57.34
23	39.00	90.78	73.31	57.61
24	43.00	92.46	80.64	64.72
25	48.37	92.62	66.12	70.64
26	55.37	92.96	62.08	69.10
27	44.50	88.20	62.49	60.53
28	34.12	87.09	0.00	50.44
29	38.91	87.56	55.26	55.15
30	53.59	92.29	60.77	69.11

$$J' = J \setminus \{6, 7, 8, 9, 17, 22, 24, 28, 30\} \qquad (5)$$

$$\forall j \in J' : P(j) = \alpha_5^* \times S_{5^{\text{Triplet}}}(j, 5) + \beta_5$$

$$\alpha_5^* \approx 0.597 \quad (p = 0.004) \qquad (6)$$

$$\forall j \in J' : P(j) = \alpha_{13}^* \times S_{13^{\text{FE}}}(j, 13) + \beta_{13}$$

$$\alpha_{13}^* \approx 0.738 \quad (p < 0.001) \qquad (7)$$

These equations demonstrate the substantial improvement in correlation strength when high-variance products are excluded from the analysis. Equation 5 defines the filtered dataset by removing products with high response variability, while Equations 6 and 7 show the resulting enhanced correlations. The improved correlation coefficients ($\alpha_5^* = 0.597$ and $\alpha_{13}^* = 0.738$) indicate that products with polarized aesthetic preferences may introduce noise in the similarity-preference relationship, and filtering these products reveals stronger underlying patterns between visual similarity and aesthetic preference.

B. PCAS OF MODELS' EMBEDDING VECTORS

While a positive correlation between image similarity determined by deep learning models and aesthetic preference for each product was identified, questions remain as to why different models explained this relationship for products #5 and #13. To investigate this, the correlation between novelty scores and deep learning models was analyzed to determine whether each model identifies novelty or if other properties exist. Embedding vectors for the similarity of product #5's Triplet Network model and product #13's Pre-trained CNN model, which had significant correlations with aesthetic preference, were extracted and reduced to 10 dimensions through principal component analysis [40], [41]. PCA is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while preserving the maximum variance, allowing identification of the most important patterns in the data. Tables 17 and 18 show the PCA results. Principal component analysis was performed on mean embedding vectors (30 products × embedding dimensions), reducing dimensionality to 10 components. The "For each" column shows the explained variance ratio

TABLE 12. Similarity results No.13 models vs. 29 speakers (%)

No.	Triplet Network	Siamese Network	Auto-encoder	Pre-trained CNN
13	76.14	96.21	80.63	79.55
1	41.20	87.86	60.71	63.60
2	36.57	91.54	65.59	59.90
3	38.35	89.95	34.33	63.43
4	32.47	90.33	67.47	59.78
5	45.72	94.38	76.39	68.24
6	37.76	89.33	12.55	61.07
7	31.42	89.90	65.45	55.06
8	33.68	90.26	57.33	55.10
9	46.74	94.73	57.06	66.28
10	35.86	87.32	14.16	61.16
11	39.48	88.87	69.34	61.18
12	42.86	88.12	73.82	67.73
14	36.62	86.56	43.61	57.43
15	36.01	90.38	65.27	62.68
16	54.28	95.79	46.58	69.36
17	33.54	88.42	75.84	54.32
18	34.08	88.71	53.41	55.40
19	35.91	89.83	64.85	65.89
20	37.29	90.50	75.04	67.82
21	36.47	87.60	79.63	60.10
22	43.65	90.33	47.90	65.11
23	33.23	89.83	71.32	54.10
24	39.70	92.19	78.21	58.07
25	41.30	91.88	63.35	64.46
26	38.58	90.69	52.69	63.93
27	33.73	88.32	59.21	66.66
28	35.15	87.68	0.00	56.65
29	39.80	89.70	52.15	63.57
30	38.27	91.83	54.91	69.92

TABLE 13. No.5 models' analysis metrics

Model	Pearson	Cosine Similarity
Triplet Network[a]	$r = 0.448, p = 0.013$	$\cos(\theta) = 0.983, p = 0.012$
Siamese Network	$r = 0.136, p = 0.475$	$\cos(\theta) = 0.984, p = 0.224$
Auto-encoder	$r = -0.001, p = 0.997$	$\cos(\theta) = 0.938, p = 0.494$
Pre-trained CNN	$r = 0.319, p = 0.086$	$\cos(\theta) = 0.985, p = 0.047$

[a] Model with $p < 0.05$

TABLE 14. No.13 models' analysis metrics

Model	Pearson	Cosine Similarity
Triplet Network	$r = 0.309, p = 0.097$	$\cos(\theta) = 0.974, p = 0.050$
Siamese Network	$r = 0.300, p = 0.107$	$\cos(\theta) = 0.985, p = 0.057$
Auto-encoder	$r = 0.018, p = 0.923$	$\cos(\theta) = 0.932, p = 0.486$
Pre-trained CNN[a]	$r = 0.478, p = 0.008$	$\cos(\theta) = 0.988, p = 0.002$

[a] Model with $p < 0.05$

for individual components, while "Cumulative" shows the cumulative explained variance. The structure of the entire data matrix before reduction consisted of rows representing

individual product categories (1, 2, ..., 30), composed of the mean values of embedding vectors from 15 images belonging to each category. The columns represent each dimension of

TABLE 15. No.5 models' analysis metrics without products var > 2.9

Model	Pearson	Cosine Similarity
Triplet Network[a]	$r = 0.597, p = 0.004$	$\cos(\theta) = 0.985, p = 0.002$
Siamese Network	$r = 0.202, p = 0.381$	$\cos(\theta) = 0.982, p = 0.188$
Auto-encoder	$r = 0.218, p = 0.342$	$\cos(\theta) = 0.969, p = 0.182$
Pre-trained CNN[a]	$r = 0.456, p = 0.038$	$\cos(\theta) = 0.985, p = 0.014$

[a] Model with $p < 0.05$

TABLE 16. No.13 models' analysis metrics without products var > 2.9

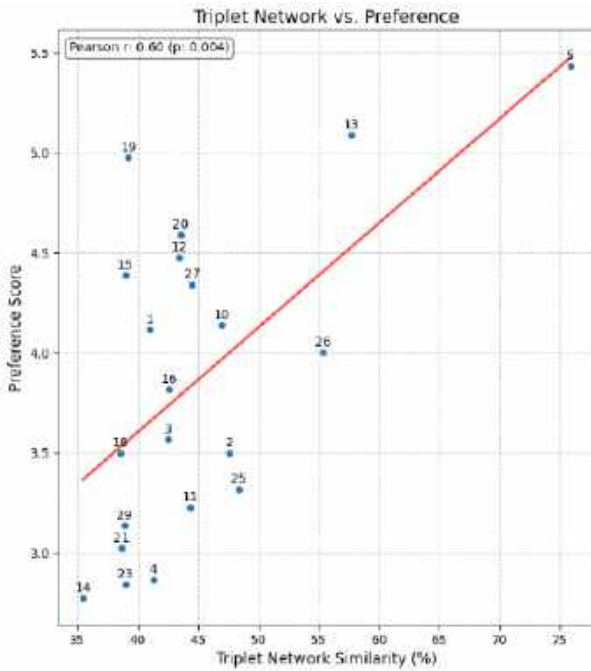| Model	Pearson	Cosine Similarity
Triplet Network[a]	$r = 0.444, p = 0.044$	$\cos(\theta) = 0.975, p = 0.027$
Siamese Network	$r = 0.414, p = 0.062$	$\cos(\theta) = 0.983, p = 0.033$
Auto-encoder	$r = 0.212, p = 0.356$	$\cos(\theta) = 0.960, p = 0.177$
Pre-trained CNN[a]	$r = 0.738, p < 0.001$	$\cos(\theta) = 0.990, p < 0.001$

[a] Model with $p < 0.05$

FIGURE 4. Similarity (%) and aesthetic preference (scores) without products with var > 2.9 for No.5
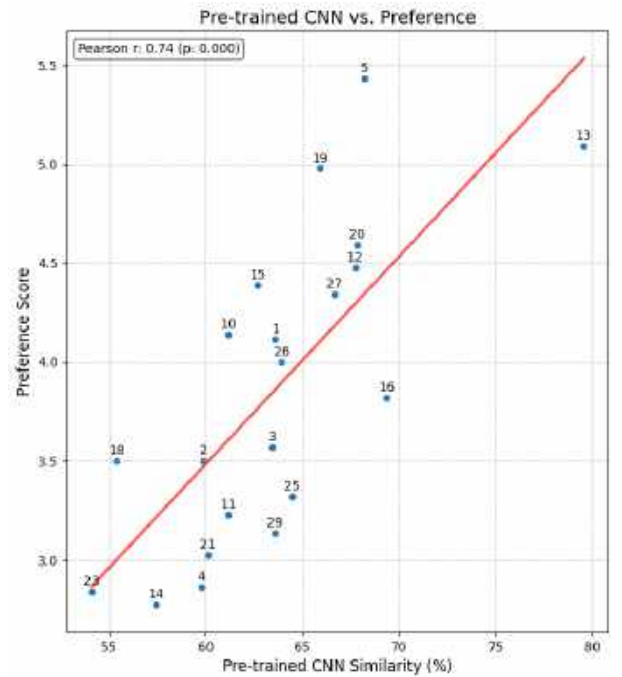

FIGURE 5. Similarity (%) and aesthetic preference (scores) without products with var > 2.9 for No.13

the embedding space. Therefore, the overall structure of the data matrix consists of 30 rows and as many columns as the embedding dimensions for each model.

Based on the principal component analysis results, specific principal components showed significant correlations with the absolute value of differences in product novelty scores. The correlation coefficients in Tables 19 and 20 were calculated using Pearson correlation analysis between principal component values and the absolute novelty score differences from the reference product. A negative correlation between the absolute value of novelty score differences and principal components $\text{Corr}(|N(j) - N(\text{ref})|, PC^i_{\text{ref}}(j)) < 0$ suggests that

the principal component is related to features that define the novelty of the reference product. Here, PC^i_{ref} represents the i-th principal component value of the reference product (#5 or #13) model, $N(j)$ is the novelty score of the j-th product, $N(\text{ref})$ is the novelty score of the reference product, $|N(j) - N(\text{ref})|$ is the absolute value of the difference between the two scores, and Corr refers to the Pearson correlation coefficient. As shown in Tables 19 and 20, the analysis revealed that PC^7_5 from product #5 showed a significant negative correlation with novelty score differences ($r = -0.407, p = 0.029$), while PC^2_{13} from product #13 showed a negative correlation ($r = -0.619, p < 0.001$).

TABLE 17. No.5 Triplet Network PCA

Component	For each	Cumulative
PC1	0.269	0.269
PC2	0.169	0.438
PC3	0.090	0.528
PC4	0.085	0.613
PC5	0.063	0.676
PC6	0.040	0.716
PC7	0.035	0.751
PC8	0.033	0.784
PC9	0.032	0.816
PC10	0.024	0.840

TABLE 18. No.13 Pre-trained CNN PCA

Component	For each	Cumulative
PC1	0.169	0.169
PC2	0.139	0.308
PC3	0.104	0.412
PC4	0.096	0.508
PC5	0.058	0.566
PC6	0.049	0.615
PC7	0.043	0.658
PC8	0.041	0.699
PC9	0.032	0.731
PC10	0.030	0.761

TABLE 19. Correlation between No.5 Triplet Network's PCs and absolute novelty difference

Component	Correlation	p-value
PC1	-0.243	0.204
PC2	-0.335	0.076
PC3	-0.157	0.415
PC4	-0.012	0.951
PC5	-0.127	0.511
PC6	-0.030	0.878
PC7[a]	-0.407	0.029
PC8	-0.026	0.892
PC9	-0.019	0.922
PC10	0.082	0.674

[a] Principal component with $p < 0.05$

These results indicate that features constituting novelty in each product can be explained by certain principal components. It suggests that other principal components may reflect properties other than novelty.

Not only the absolute value of novelty differences but also the correlation with novelty scores themselves was examined (Tables 21 and 22). Product #13 was evaluated as having a high novelty score among the entire product group, and just as PC_{13}^2 in the model showed a strong negative correlation with the absolute value of novelty score differences, it also showed a significant positive correlation with the novelty score itself

TABLE 20. Correlation between No.13 Pre-trained CNN's PCs and absolute novelty difference

Component	Correlation	p-value
PC1	-0.150	0.436
PC2[a]	-0.619	< 0.001
PC3	-0.241	0.209
PC4	-0.201	0.295
PC5	-0.147	0.447
PC6	0.144	0.455
PC7	-0.236	0.218
PC8	0.089	0.647
PC9	-0.123	0.524
PC10	0.134	0.490

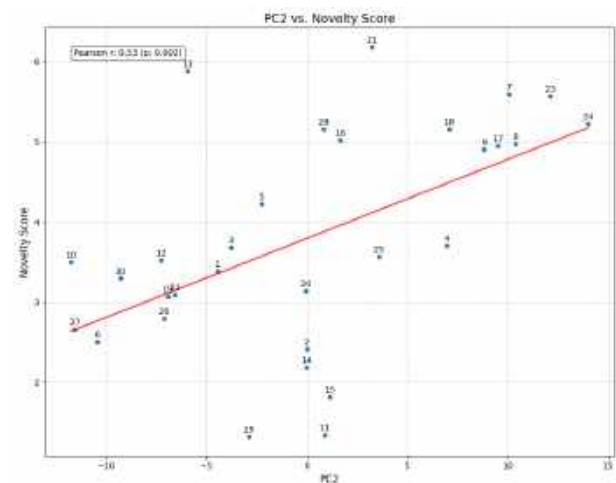[a] Principal component with $p < 0.05$

FIGURE 6. No.13's Pre-trained CNN's PC_{13}^2 and novelty scores

($r = 0.533$, $p = 0.002$). This suggests that PC_{13}^2 captures visual elements that determine novelty as generally perceived regardless of the product. In other words, products with high PC_{13}^2 values are evaluated as having high overall novelty and can be considered to share similar novelty properties with product #13, which showed a high novelty score. Figure 6 shows the relationship between No.13's Pre-trained CNN's PC_{13}^2 and novelty scores.

While in the model of product #5, which showed a relatively lower novelty score than product #13, PC_5^7 showed a significant negative correlation with the absolute value of novelty score differences, it did not show a significant correlation with the novelty score itself. These results suggest that PC_5^7 is a principal component that reflects novelty specific to product #5 rather than general novelty. Instead, PC_5^2 and PC_5^4 showed positive correlations with novelty scores ($r = 0.499$, $p = 0.005$; $r = 0.461$, $p = 0.010$), suggesting that PC_5^2 and PC_5^4 reflect features associated with novelty that have explanatory power in other products beyond product #5. Figures 7 and 8 show these relationships.

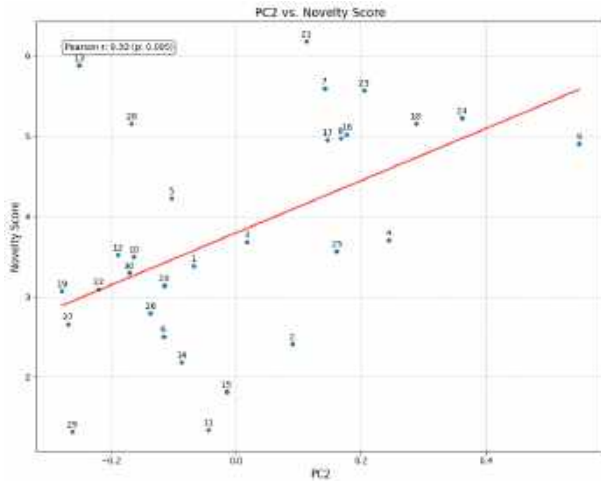This difference suggests that product #13, which received a

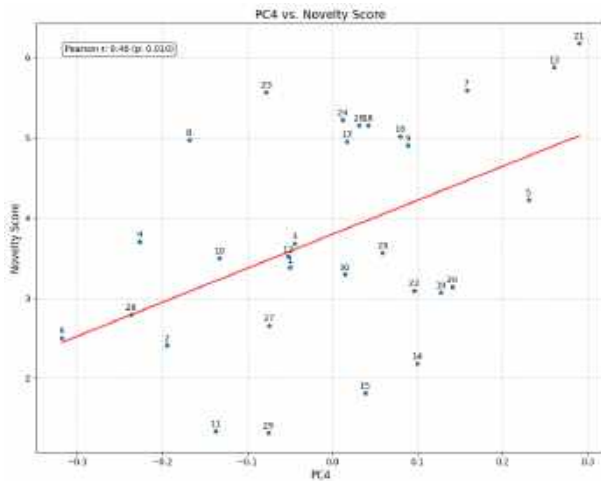**FIGURE 7.** No.5's Triplet Network's PC_5^2 and novelty scores

FIGURE 8. No.5's Triplet Network's PC_5^4 and novelty scores

high novelty score (mean = 5.89), strongly embodies universally recognized novelty elements, while product #5, which showed a relatively lower novelty score (mean = 4.23), appears to have its novelty-reflecting principal components separated into several branches.

In addition, the correlation between each of the 10 principal components and aesthetic preference was also analyzed, with results shown in Tables 23 and 24. Pearson correlation coefficients were computed between each principal component value and aesthetic preference scores across all products. While the comprehensive image similarity yielded significant correlations with aesthetic preference, only significant negative correlations were found with individual principal components, with no positive correlations emerging. This suggests that the relationship between principal components and aesthetic preference may be non-linear or follow a polynomial relationship, and it can be inferred that the cognitive process by which people determine aesthetic preference involves a

TABLE 21. Correlation between No.5 Triplet Network's PCs and novelty scores

Component	Correlation	p-value
PC1	0.021	0.915
PC2[a]	0.508	0.005
PC3	-0.046	0.813
PC4[a]	0.464	0.011
PC5	0.155	0.422
PC6	0.202	0.294
PC7	0.052	0.788
PC8	-0.041	0.832
PC9	0.107	0.579
PC10	-0.215	0.262

[a] Principal component with $p < 0.05$

TABLE 22. Correlation between No.13 Pre-trained CNN's PCs and novelty scores

Component	Correlation	p-value
PC1	0.132	0.494
PC2[a]	0.609	< 0.001
PC3	0.296	0.119
PC4	0.226	0.238
PC5	0.116	0.549
PC6	-0.148	0.445
PC7	0.220	0.251
PC8	-0.084	0.666
PC9	0.141	0.465
PC10	-0.138	0.476

[a] Principal component with $p < 0.05$

TABLE 23. Correlation between No.5 Triplet Network's PCs and aesthetic preference scores

Component	Correlation	p-value
PC1	0.312	0.094
PC2	-0.358	0.052
PC3	-0.128	0.501
PC4	0.255	0.175
PC5	-0.009	0.962
PC6	-0.087	0.647
PC7	0.149	0.433
PC8	0.272	0.146
PC9	0.253	0.178
PC10	-0.057	0.765

more complex process.

C. PCAS AND AESTHETIC PREFERENCE

Using the 10 principal components extracted earlier, we analyzed polynomial regression models for their relationship with preference scores. The individual results for each model's principal components are shown in Tables 25 and 26.

We also explored which interaction model of two principal components best explains preference scores. In model #5, the

TABLE 24. Correlation between No.13 Pre-trained CNN's PCs and aesthetic preference scores

Component	Correlation	p-value
PC1	0.094	0.621
PC2[a]	-0.366	0.047
PC3	-0.052	0.783
PC4	-0.032	0.868
PC5	0.329	0.076
PC6[a]	-0.444	0.014
PC7	-0.010	0.957
PC8	-0.135	0.476
PC9	-0.233	0.216
PC10	-0.149	0.433

[a] Principal component with $p < 0.05$

TABLE 25. Results for #5's Triplet Network PC polynomial analysis

Component	Quadratic R^2	Cubic R^2
PC1	0.214	0.223
PC2	0.168	0.202
PC3	0.029	0.031
PC4	0.104	0.162
PC5	0.046	0.064
PC6	0.018	0.023
PC7	0.038	0.061
PC8	0.074	0.127
PC9	0.098	0.203
PC10	0.050	0.067

TABLE 26. Results for #13's Pre-trained CNN PC polynomial analysis

Component	Quadratic R^2	Cubic R^2
PC1	0.041	0.049
PC2	0.142	0.144
PC3	0.068	0.089
PC4	0.012	0.051
PC5	0.110	0.136
PC6	0.272	0.365
PC7	0.015	0.017
PC8	0.033	0.046
PC9	0.133	0.154
PC10	0.052	0.055

TABLE 27. Results for Interaction model

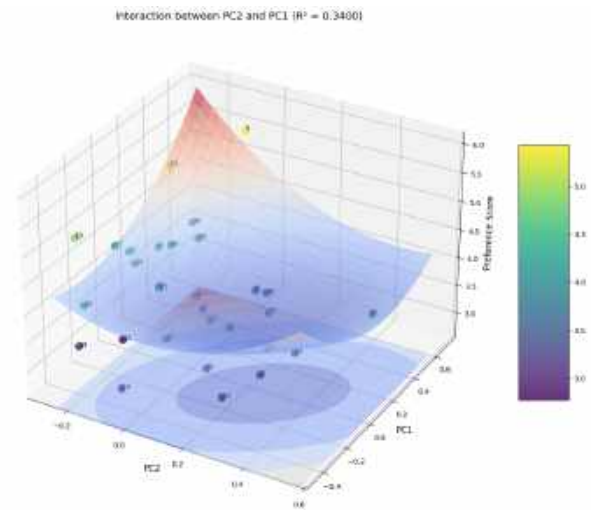| Model	R^2
#5 Triplet Network $PC_5^1 \times PC_5^8$	0.3400
#13 Pre-trained CNN $PC_{13}^2 \times PC_{13}^6$	0.4065

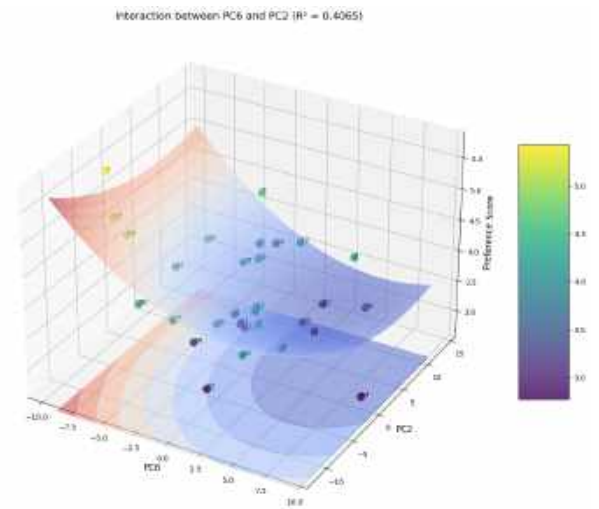**FIGURE 9.** 3D Visualization of Interaction model of #5's $PC_5^1 \times PC_5^8$

FIGURE 10. 3D Visualization of Interaction model of #13's $PC_{13}^2 \times PC_{13}^6$

interaction of PC_5^1 and PC_5^8 yielded $R^2 = 0.3719$, while in model #13, the interaction of PC_{13}^2 and PC_{13}^6 resulted in $R^2 = 0.4065$. While individual principal components did not show clear relationships with preference, we observed that interactions of multiple principal components substantially increased explanatory power. Figures 9 and 10 show 3D visualizations of these interaction models.

D. WHAT ABOUT OTHER PROPERTIES?

As we observed above, we were able to identify several principal components of embedding vectors that correlate with novelty. Building on this finding, we sought to explore relationships with additional properties beyond novelty, specifically those outlined in Table 1. From these properties, we selected complexity, harmony, balance, unity, and dynamics for further analysis. Following the same methodology used for novelty assessment, we asked some of the same participants ($N = 15$) to evaluate the identical set of 30 products on a 7-point Likert scale for each of these five additional properties. Using the same participants was to minimize unexpected errors across evaluations. The mean scores for each property
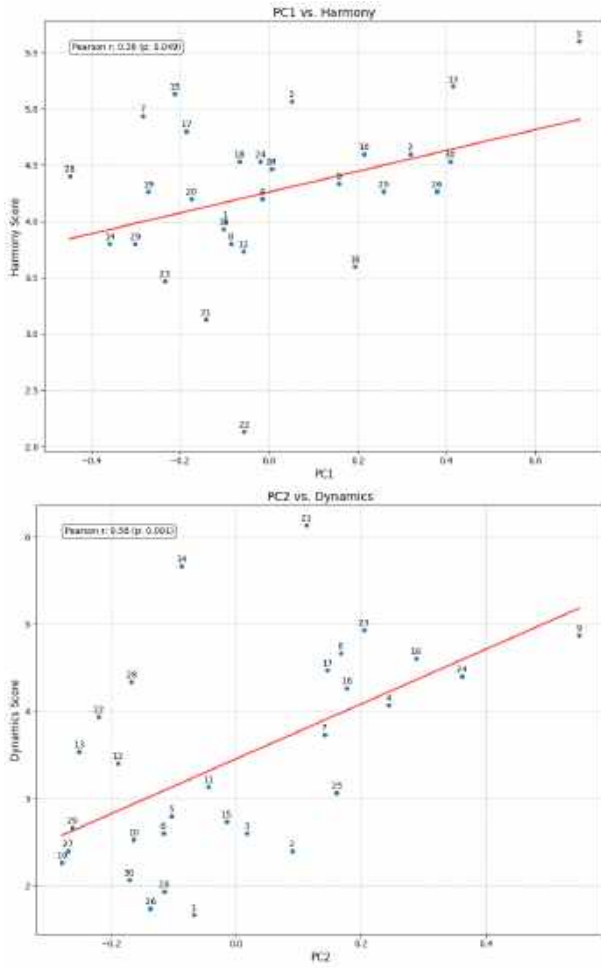
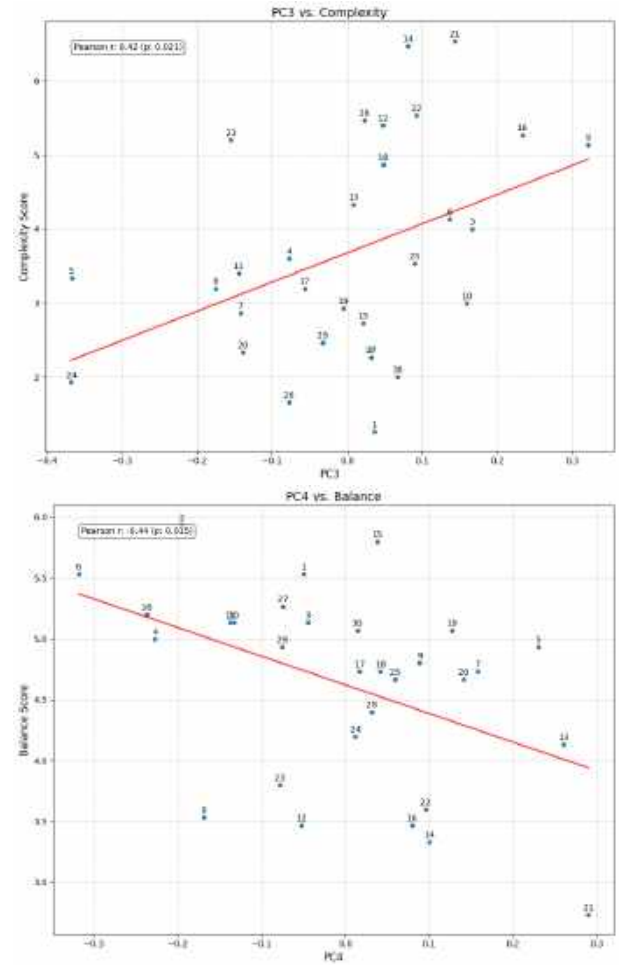**FIGURE 11.** (a) #5's properties: Harmony. (b) #5's properties: Dynamics.

FIGURE 12. (a) #5's properties: Complexity. (b) #5's properties: Balance.

across all products are presented in Table 28. These scores were obtained from 15 participants evaluating all 30 products on a 7-point Likert scale for complexity, harmony, balance, unity, and dynamics.

Based on the response data, we analyzed the correlations between the 10 principal components and various properties again. For the #5 Triplet Network, properties that showed significant relationships ($p < 0.05$) with principal components were PC_5^1 - Harmony ($r = 0.362$, $p = 0.049$), PC_5^2 - Dynamics ($r = 0.555$, $p = 0.001$), PC_5^3 - Complexity ($r = 0.418$, $p = 0.021$), PC_5^4 - Balance ($r = -0.440$, $p = 0.015$) and for the #13 Pre-trained CNN, they were PC_{13}^1 - Harmony ($r = 0.374$, $p = 0.042$), PC_{13}^2 - Dynamics ($r = 0.649$, $p < 0.001$), PC_{13}^4 - Complexity ($r = 0.557$, $p = 0.001$), PC_{13}^9 - Unity ($r = -0.374$, $p = 0.042$). The visualizations of these relationships are shown in Figures 11 through 14.

Following the examination of the polynomial relationship between novelty and aesthetic preference scores as described above, we conducted analogous polynomial analyses for the five properties, examining linear, quadratic, and cubic relationships. For this analysis, we used response data from the

same 15 participants regarding aesthetic preference scores. The analysis revealed that unity was the only property among the five that demonstrated a statistically significant relationship ($p < 0.05$) with aesthetic preference scores. The corresponding graph is presented in Figure 15. While the polynomial relationship between unity and aesthetic preference yielded significant results, given that only the correlation with PC_{13}^9 of #13 showed significance in relation to the principal components, it appears somewhat challenging to regard unity as having substantial explanatory power for overall aesthetic preference.

Furthermore, as we confirmed that each property showed similar response patterns, we analyzed the correlations between properties and visualized these relationships in a heatmap as shown in Figure 16. The highest and significant positive correlations were found between Complexity - Dynamics ($r = 0.762$, $p < 0.001$) and Balance - Unity ($r = 0.707$, $p < 0.001$). For negative correlations, the strongest were Balance - Dynamics ($r = -0.755$, $p < 0.001$) and Complexity - Balance ($r = -0.717$, $p < 0.001$). Here, we can observe that dynamics and novelty were grouped into the same PC, which is consistent with their high correla-

TABLE 28. Mean scores of other properties by product

| No.	Comp	Harm	Bal	Uni	Dyna	No.	Comp	Harm	Bal	Uni	Dyna
1	1.27	4.00	5.53	5.73	1.67	16	5.27	3.60	3.47	3.87	4.27
2	2.27	4.60	5.93	5.87	2.40	17	3.20	4.80	4.73	4.87	4.47
3	4.00	5.07	5.13	4.60	2.60	18	4.87	4.53	4.73	5.93	4.60
4	3.60	4.20	5.00	4.40	4.07	19	2.93	4.27	5.07	4.80	2.27
5	3.33	5.60	4.93	5.47	2.80	20	2.33	4.20	4.67	4.33	1.93
6	4.13	4.47	5.53	4.27	2.60	21	6.53	3.13	2.73	4.13	6.13
7	2.87	4.93	4.73	4.60	3.73	22	5.53	2.13	3.60	2.53	3.93
8	3.20	3.80	3.53	3.87	4.67	23	5.20	3.47	3.80	3.60	4.93
9	5.13	4.33	4.80	5.00	4.87	24	1.93	4.53	4.20	4.87	4.40
10	3.00	4.60	5.13	4.20	2.53	25	3.53	4.27	4.67	4.80	3.07
11	3.40	3.93	5.13	5.53	3.13	26	1.67	4.27	5.20	5.07	1.73
12	5.40	3.73	3.47	3.67	3.40	27	2.27	4.47	5.27	5.07	2.40
13	4.87	4.53	4.73	5.93	4.60	28	5.47	4.4	4.4	5.07	2.4
14	2.93	4.27	5.07	4.80	2.27	29	2.47	3.8	4.93	4.47	2.67
15	2.33	4.20	4.67	4.33	1.93	30	2.00	4.53	5.07	4.93	2.07
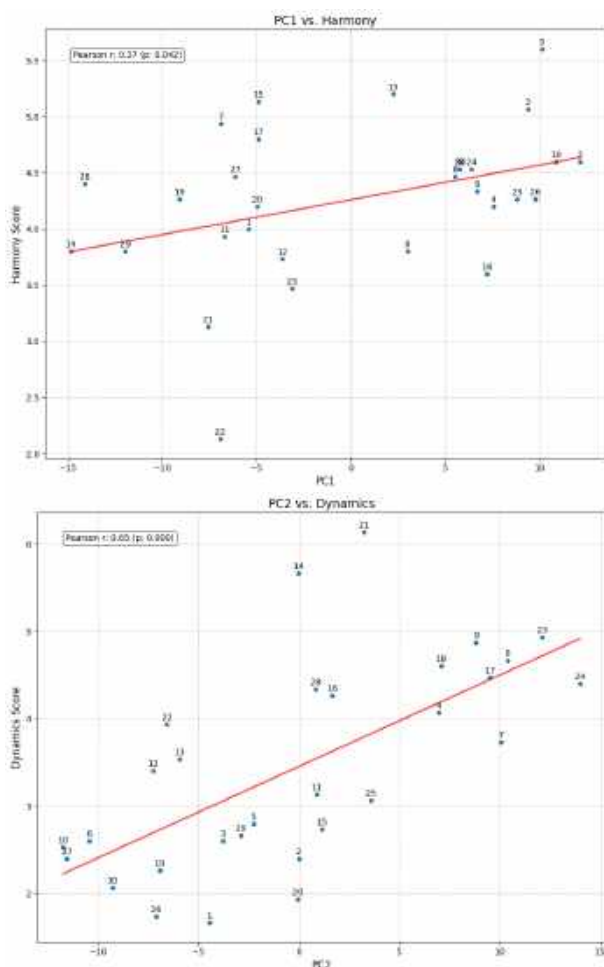
FIGURE 13. (a) #13's properties: Harmony. (b) #13's properties: Dynamics.
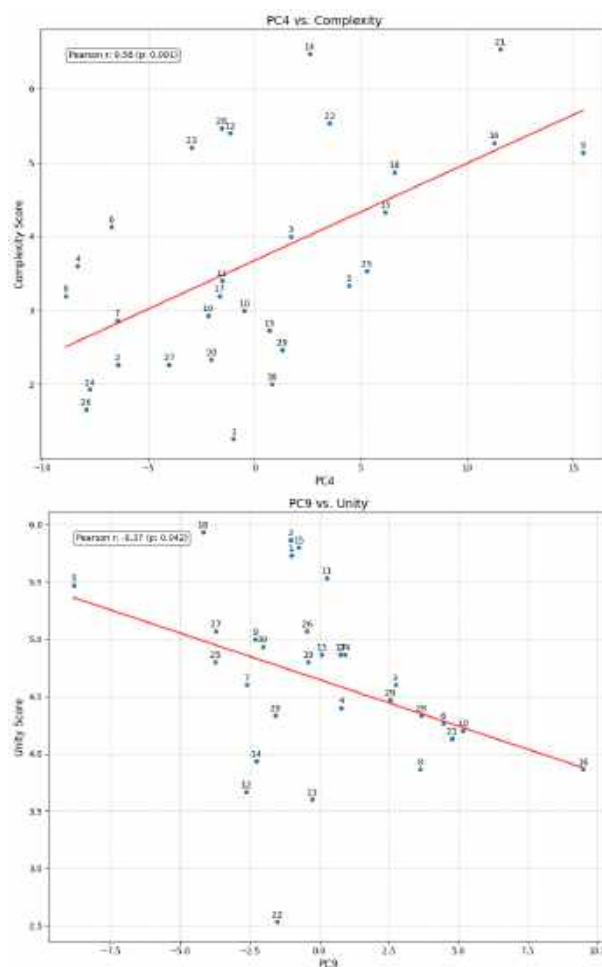

FIGURE 14. (a) #13's properties: Complexity. (b) #13's properties: Unity.

tion, whereas none of the other properties were found to be grouped within the same principal component.

The findings covered in this paper can be summarized and visualized as shown in Figures 17 and 18. We've analyzed the
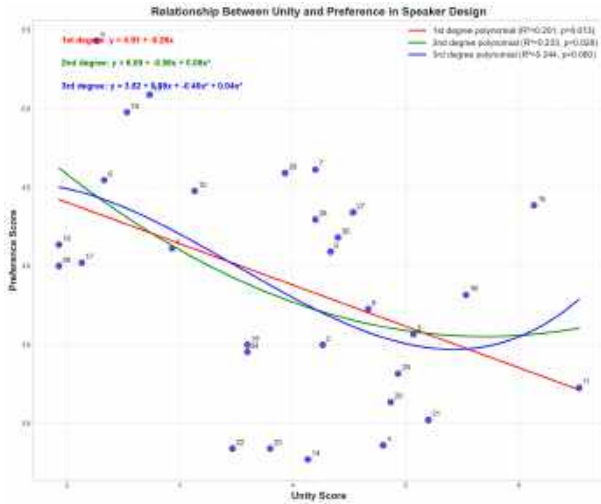
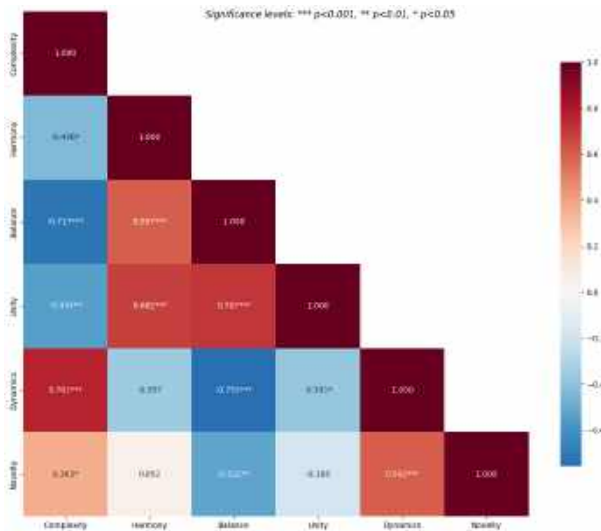**FIGURE 15.** Relationship between unity and aesthetic preference

FIGURE 16. Correlation matrix between properties

correlation between aesthetic preferences and visual similarity, extracted image embedding vectors, reduced them to 10 dimensions through principal component analysis, and then thoroughly examined their correlation with multiple visual properties.

IV. CONCLUSION

Based on prior studies on aesthetic preferences and visual properties, the present research employed deep learning to quantitatively and comprehensively investigate the underlying relationships. First, we examined preferences within the product category of home audio speakers and found that similar preference trends emerged regardless of participants' academic background. Additionally, it was observed that aesthetic preferences did not necessarily follow the inverted-U model, the reason for which was soon explored in PCAs of models' embedding vectors. Second, we developed four

distinct deep learning models for the most preferred product and measured the similarity between this product and other samples using the generated models. A linear relationship between similarity and preference was then examined, and it was found that this relationship became stronger when high-variance products were excluded from the analysis. Subsequently, principal component analysis (PCA) was conducted on the embedding vectors obtained from the deep learning model. This analysis revealed which principal components were associated with perceived novelty, indicating that the deep learning models effectively extracted relevant features related to novelty. However, no significant linear relationships were found between individual principal components and aesthetic preferences. Interaction models among multiple principal components were analyzed to further explore this issue. The findings revealed polynomial relationships between these components and aesthetic preferences, suggesting that such preferences are determined by complex interactions among various elements rather than by any single factor alone. Lastly, through principal component analysis of some properties in Table 1 and embedding vectors, we were able to discover what properties the features extracted by the model incorporate besides novelty.

A. IMPLICATIONS

This research confirmed a correlation between similarity to other products and aesthetic preference by extracting features through a deep learning model trained on only one product. This indicates that the most preferred product contains multiple visual elements that influence aesthetic preference. In contrast to previous researches mentioned above, this research is distinctive in that it trains a model based on the most preferred product without labeling, and then quantitatively evaluates visual elements affecting aesthetic preference through similarity comparisons with other products. Although not shown in this study, heat maps can be output during similarity assessment to visually identify areas of low similarity within each product image. This allows for identification of which parts should be modified in the design process, which parts are related to preference, or which parts are associated with properties.

Previous deep learning approaches to aesthetic evaluation rely heavily on extensive labeling for supervised learning, whereas our method requires minimal data without preference labeling. While Wu et al. [23] and Burnap et al. [24] focus on direct prediction through large labeled datasets, our approach centers on extracting insights through PCA analysis of the relationship between visual similarity and user-provided aesthetic preference scores and visual property ratings. This fundamental difference in methodology can be summarized in Table 29. This comparison illustrates the different approaches to aesthetic evaluation, with our method offering an alternative framework that trades predictive accuracy for reduced data requirements and interpretable aesthetic property analysis.

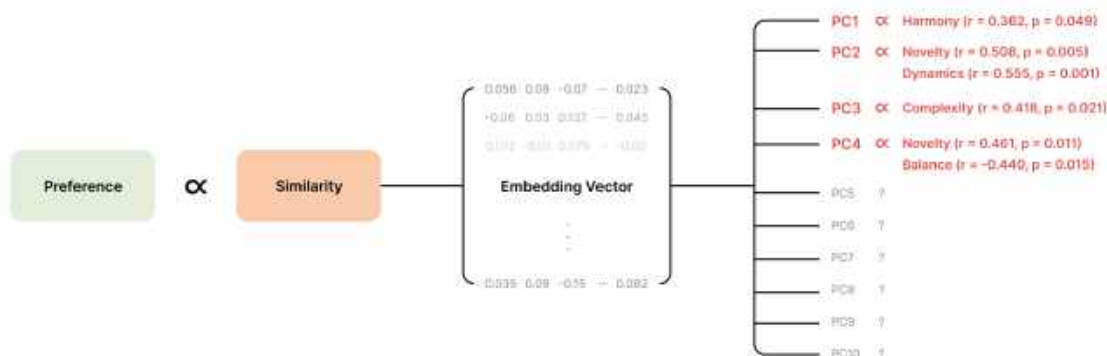Furthermore, we plan to validate this method through inte-

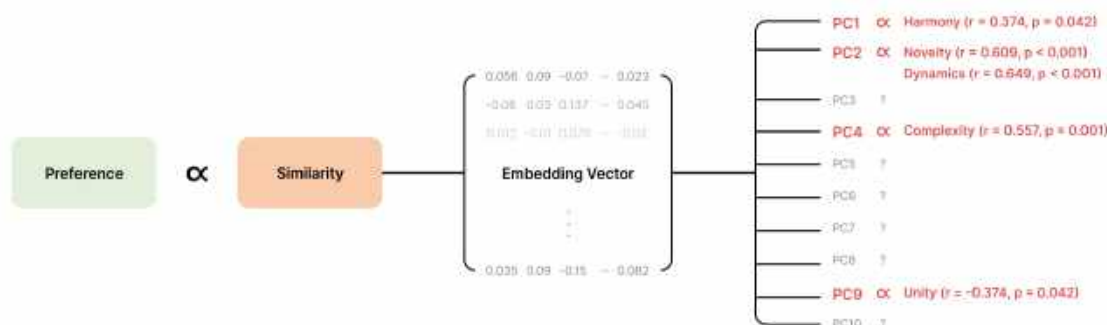**FIGURE 17.** Schematic Diagram of No. 5 Triplet Network Model

FIGURE 18. Schematic Diagram of No. 13 Pre-trained CNN Model

TABLE 29. Comparative Analysis of Aesthetic Evaluation Approaches

Aspect	Wu et al. (2020)	Burnap et al. (2021)	Our Approach
Training Data Requirements	Large dataset with award labels for each product	203 labeled + 180,000 unlabeled images	Single reference product (35 images, limited product diversity)
Labeling Requirements	Extensive: Award status for each product	Extensive: Consumer ratings (7,308 evaluations)	Minimal: No preference labeling required
Model Architecture	Deep CNNs for direct prediction	VAE-GAN hybrid (custom architecture)	Multiple approaches: Pre-trained CNN, Siamese, Triplet, Autoencoder
Evaluation Methodology	Award prediction accuracy	Theme clinic validation + generative quality	Correlation analysis (depends on reference product selection)
Scalability	Limited by award data availability	Limited by consumer evaluation costs	Requires domain-specific validation for new categories

gration into actual design workflows. The proposed practical implementation process would involve the following iterative design cycle: (1) Product category selection and establishment of aesthetic preference benchmarks through user evaluation of existing products for both aesthetic preference and aesthetic properties; (2) Training deep learning models on highly preferred reference products; (3) Initial design concept development and rendering image extraction; (4) Visual similarity assessment and embedding vector analysis to identify aesthetic properties and generate attention heatmaps highlighting design elements; (5) Design refinement based on identified deficiencies or improvement opportunities; and (6) Iterative similarity assessment until design objectives are achieved.

Additionally, the integration of this evaluation model with generative AI systems presents promising opportunities. By pairing generative models with our aesthetic assessment framework, designers could specify desired similarity scores
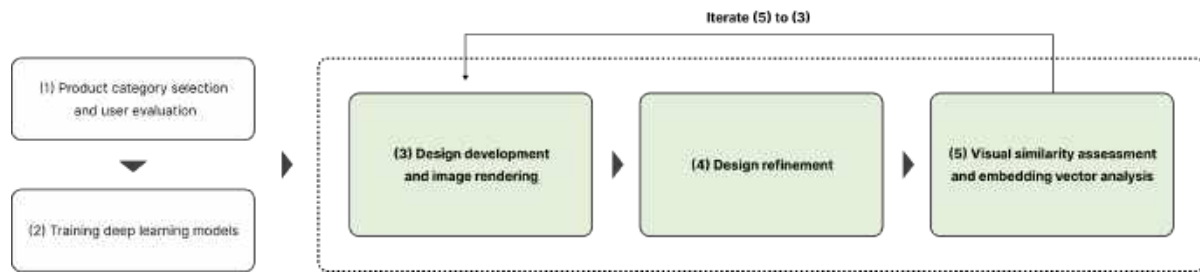
FIGURE 19. Design Process Based on the Methodology of This Research

and aesthetic properties as generation parameters, producing design candidates that meet predetermined aesthetic criteria. The resulting images could then serve as references for design application, enabling a systematic approach to aesthetic-driven design development.

It should also be noted that the identified correlations do not imply causation. There is no design evaluation method that can be used in all cases reflecting individual tastes or perspectives, and such a method is impossible [42]. This approach demonstrates the potential to be used as one of many design evaluation methods. The method proposed in this study cannot be claimed as the optimal design evaluation method. This research was conducted as an exploration of a new design evaluation methodology. Rather than serving as a standalone evaluation tool, it should be used in conjunction with other well-established methods to provide a more comprehensive understanding of design quality. It may serve as a foundation for future studies exploring data-driven approaches to design evaluation.

B. LIMITATIONS

While this research provides valuable insights into aesthetic evaluation through deep learning, several limitations should be acknowledged. The reliance on a single reference product assumes a certain degree of universality in aesthetic preference that may not hold across all design domains or cultural contexts. Our approach inherently carries the aesthetic bias embedded in the reference product selection process. When the reference product represents a highly specialized or atypical aesthetic, the similarity measurements may not generalize well to broader market preferences, potentially leading to biased evaluation outcomes.

Additionally, our study's scope is limited to home audio speakers with 44 participants, which constrains the generalizability of findings. The method's scalability remains unproven, as product categories with different design languages and cultural contexts may respond differently to similarity-based evaluation. Furthermore, the approach may face challenges when applied to large-scale or highly complex products such as automobiles, where multiple design elements and functional constraints interact in more intricate ways than in compact consumer electronics.

Finally, variance analysis reveals challenges in handling polarizing designs. Products showing high variance in prefer-

ence ratings (variance > 2.9 in our study) indicate designs that evoke divided opinions among users, requiring careful consideration within the similarity-based evaluation framework.

C. FUTURE WORK

Our future research should focus on building a robust model through hyperparameter adjustments such as training data quantity, batch size, epoch, kernel, padding, and stride. By tuning these hyperparameters, we aim to investigate whether the linear relationship between similarity and aesthetic preference can be strengthened and further explore their correlation. Additionally, some principal components were identified as being related to certain properties, but there are still principal components that remain unknown. It is necessary to investigate what features these principal components represent in products, and which properties related to aesthetic preference they are associated with.

Second, to address scalability concerns, we plan to conduct comprehensive validation studies across diverse product categories and expanded user populations. This multi-domain validation will examine how different design languages and cultural contexts affect similarity-preference correlations, with particular attention to developing domain-specific adaptations where necessary.

Furthermore, validation of the method as a practical tool will include Focus Group Interviews (FGI) with senior designers and design professionals to gather expert feedback on the method's practical utility and design recommendations. By combining quantitative validation across multiple product domains with qualitative insights from experienced practitioners, we aim to establish both the technical generalizability and practical value of our aesthetic assessment framework.

Lastly, verification is also needed on whether it is possible to predict aesthetic preference of a product through similarity. It is necessary to confirm whether the predicted scores on the linear model of similarity and preference scores for speakers other than the 30 used as stimuli in this study match the actual preference scores, and to determine whether the correlations in property scores remain valid for additional products based on principal components. Future research requires the application of qualitative research methodologies and verification in systematic experimental conditions to understand the causal relationships of properties affecting aesthetic preference.

IEEE *Access*

REFERENCES

[1] P. Kotler, K. L. Keller, F. Ancarani, and M. Costabile, *Marketing management 14/e*. Pearson, 2014.

[2] R. Verganti, *Design driven innovation: changing the rules of competition by radically innovating what things mean*. Harvard Business Press, 2009.

[3] P. Hekkert, "Design aesthetics: principles of pleasure in design," *Psychology science*, vol. 48, no. 2, pp. 157, 2006.

[4] P. Raghubir and E. A. Greenleaf, "Ratios in proportion: what should the shape of the package be?," *Journal of Marketing*, vol. 70, no. 2, pp. 95-107, 2006.

[5] D. E. Berlyne, *Aesthetics and psychobiology*. Appleton-Century-Crofts, 1971.

[6] R. Loewy, *Never leave well enough alone*. JHU Press, 2002.

[7] C. Thurgood, P. Hekkert, and J. Blijlevens, "The joint effect of typicality and novelty on aesthetic pleasure for product designs: Influences of safety and risk," in *Congress of the International Association of Empirical Aesthetics*, 2014.

[8] W. K. Hung and L. L. Chen, "Effects of novelty and its dimensions on aesthetic preference in product design," *International Journal of Design*, vol. 6, no. 2, pp. 81-90, 2012.

[9] M. Kumar and N. Garg, "Aesthetic principles and cognitive emotion appraisals: How much of the beauty lies in the eye of the beholder?," *Journal of Consumer Psychology*, vol. 20, no. 4, pp. 485-494, 2010.

[10] P. H. Bloch, "Seeking the ideal form: Product design and consumer response," *Journal of marketing*, vol. 59, no. 3, pp. 16-29, 1995.

[11] K. Koffka, *Principles of Gestalt psychology*. Routledge, 2013.

[12] S. R. Ellis, "A psychometric investigation of a scale for the evaluation of the aesthetic element in consumer durable goods," The University of Arizona, 1993.

[13] F. F. Brunel, "The psychology of product aesthetics: Antecedents and individual differences in product evaluations," University of Washington, 1998.

[14] P. H. Bloch, F. F. Brunel, and T. J. Arnold, "Individual differences in the centrality of visual product aesthetics: Concept and measurement," *Journal of consumer research*, vol. 29, no. 4, pp. 551-565, 2003.

[15] F. Guo, Y. Ding, W. Liu, C. Liu, and X. Zhang, "Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered," *International Journal of Industrial Ergonomics*, vol. 53, pp. 229-235, 2016.

[16] S. Khalighy, G. Green, C. Scheepers, and C. Whittet, "Quantifying the qualities of aesthetics in product design using eye-tracking technology," *International Journal of Industrial Ergonomics*, vol. 49, pp. 31-43, 2015.

[17] N. Archak, A. Ghose, and P. G. Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," *Management science*, vol. 57, no. 8, pp. 1485-1509, 2011.

[18] P. Liu, K. Wang, K. Yang, H. Chen, A. Zhao, Y. Xue, and L. Zhou, "An aesthetic measurement approach for evaluating product appearance design," *Mathematical Problems in Engineering*, vol. 2020, no. 1, p. 1791450, 2020.

[19] H. Lai, Z. Wu, X. Zhang, H. Liao, and E. K. Zavadskas, "A method for product appearance design evaluation based on heterogeneous data," *Advanced Engineering Informatics*, vol. 57, p. 102089, 2023.

[20] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85-117, 2015.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[22] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 806-813, 2014.

[23] J. Wu, B. Xing, H. Si, J. Dou, J. Wang, Y. Zhu, and X. Liu, "Product design award prediction modeling: Design visual aesthetic quality assessment via DCNNs," *IEEE Access*, vol. 8, pp. 211028-211047, 2020.

[24] A. Burnap, J. R. Hauser, and A. Timoshenko, "Design and evaluation of product aesthetics: A human-machine hybrid approach," Available at SSRN 3421771, 2021.

[25] A. Burnap, J. R. Hauser, and A. Timoshenko, "Product aesthetic design: A machine learning augmentation," *Marketing Science*, vol. 42, no. 6, pp. 1029-1056, 2023.

[26] M. E. Creusen and J. P. Schoormans, "The different roles of product appearance in consumer choice," *Journal of product innovation management*, vol. 22, no. 1, pp. 63-81, 2005.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[28] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504-507, 2006.

[29] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[30] F. Chollet and others, "Keras," GitHub repository, https://github.com/fchollet/keras, 2015.

[31] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[33] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Computer vision–ECCV 2016 workshops*, pp. 850-865, Springer, 2016.

[34] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu, "Learning fine-grained image similarity with deep ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1386-1393, 2014.

[35] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural networks: Tricks of the trade: Second edition*, pp. 437-478, Springer, 2012.

[36] L. Prechelt, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*, pp. 55-69, Springer, 2002.

[37] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Advances in neural information processing systems*, vol. 27, 2014.

[38] A. Shrivastava, A. Gupta, and R. Girshick, "Training Region-Based Object Detectors with Online Hard Example Mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 761-769, 2016.

[39] S. H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, no. 4, pp. 300-307, 2007.

[40] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014*, pp. 818-833, Springer, 2014.

[41] A. Nguyen, J. Yosinski, and J. Clune, "Multifaceted feature visualization: Uncovering the different types of features learned by each neuron in deep neural networks," arXiv preprint arXiv:1602.03616, 2016.

[42] P. Hekkert and H. Leder, "Product aesthetics," *Product experience*, pp. 259-285, 2008.

SEOK YOUNG HWANG received the degree in mechatronics engineering. He is currently a researcher specializing in analyzing human emotional responses to products and their usability through data-driven methodologies. His research interests include physical computing, physical UX, and programming applications in human-computer interaction. He focuses on developing methodologies to understand and measure user experiences with products through quantitative analysis and data-driven approaches.

JUSEONG KIM is a design engineer with research interests in robotics, mobility, and intelligent products. He specializes in solving design-engineering challenges to enhance user experience, with particular focus on usability and physical interaction within human-centered design frameworks. His work bridges the gap between engineering functionality and user-centered design principles in the development of innovative products and systems.

KICHEOL PAK is currently a Professor in the Department of Mechanical and System Design Engineering and the Graduate School of International Design Arts (IDAS) at Hongik University, Seoul, Korea, where he leads the Human-Centered Integrated Design Engineering Laboratory. He has conducted collaborative research with major companies including Hyundai Motor Company, LG Electronics, Samsung Electronics, Embrain Research, and Doosan Mobility Innovation. He serves as the Principal Investigator for government design R&D projects under the Design Innovation Capability Enhancement Program. His research interests focus on Physical UX Design for robots, mobility systems, and innovative products, exploring new values and possibilities for future design through human-centered design and technology convergence.

. . .