



HAL
open science

Validating an emerging tool based on auditory scale for decoding users' emotional desires

Weihua Lu, Jean-François Petiot, Kun-Pyo Lee, Haihua Zhu

► **To cite this version:**

Weihua Lu, Jean-François Petiot, Kun-Pyo Lee, Haihua Zhu. Validating an emerging tool based on auditory scale for decoding users' emotional desires. *Journal of Mechanical Science and Technology*, 2018, 32 (9), pp.4245-4254. 10.1007/s12206-018-0823-3 . hal-02004435

HAL Id: hal-02004435

<https://hal.science/hal-02004435v1>

Submitted on 18 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Validating an emerging tool based on auditory scale for decoding users' emotional desires†

Weihua Lu^{1,2,*}, Jean-François Petiot³, Kun-pyo Lee² and Haihua Zhu¹

¹College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

²Department of Industrial Design, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea

³Ecole Centrale de Nantes - LS2N - UMR CNRS 6004, Nantes 44321, France

Abstract

More and more businesses are realizing the vital role of emotions in heightening the attractiveness and adding value to products. Due to the perceptual differences between designers and users, it is a key challenge for designers to transfer a user's emotional desires into technical and design specifications, particularly cultural differences that may lead to ambiguous responses. We conducted research on revealing intangible emotions via quantifiable method, in order to bridge the discrepancy between designers' intentions and users' decoding. An emerging non-verbal tool was proposed, auditory parameter method (AP), which considers a set of auditory stimuli as scale to assess the emotions elicited by products. The objective was to investigate and validate whether the AP method is resistant to interference in cross-cultural and inter-ethnic application. Comparative experiments were organized in parallel with both Chinese and French participants, considering conventional verbal assessment semantic differential method (SD) as control group. The statistical results indicated that the auditory scale of AP method has slight intercultural differences in perception and interpretation.

Keywords: Affective design; Auditory parameter method; Non-verbal; Quantitative approach; User emotion

1. Introduction

In the current marketplace, it is necessary but no longer sufficient to offer good functioning products [1, 2]. Consumer research has shown that the emotional quality of products plays an important role for differential advantage in the marketplace [3-5]. By the time we become consciously aware of what we like, our unconscious has already decided based on multisensory information received. The conscious mind then fabricates a logical story to justify the decision.

It is a challenge for design engineers to transfer users' emotional desires into technical and design specifications. A clear understanding of user perceptions in the design process is a crucial component for a creative and innovative approach towards the realization of a new market-successful product [6]. There are many differences in the perception of designers and users for the same product [7]. Usually, it seems troublesome to discuss emotional responses with users because there is a lack of proper vocabulary to express subtle affective responses.

Measurement of user emotion becomes an interesting drive for product design. How to measure and understand user emo-

tions has attracted the attention of many researchers. As a consequence, both the number of reported instruments and the diversity in approaches to measure emotions are abundant. Today's instruments range from simple pen-and-paper rating scales to dazzling high-tech equipment that measures brain waves or eye movements.

Self-report measurement instruments find a balance between having a reliable method and not disturbing the experience, which are considered as conventional and efficient techniques to quantify user's emotional desires. This kind of measurement technique can be summarized in verbal and non-verbal approaches depending on the type of assessment scales.

A classical verbal approach, such as the semantic differential method (SD) [8], consists in defining a list of attributes, and carrying out user-tests in which the user must assess the product on measurement scales. The verbal nature entails some limitations. First, expressing the affective needs in verbal descriptions tends to be imprecise and ambiguous [9], since user emotions are subtle and complicated impressions. Second, verbal approaches have problems when applied to different cultures and ethnic groups [10]. Translation into other languages might be difficult and result in distortion. Third, the verbal nature is unworkable for young children and illiterate individuals. Finally, the configuration of the term list is product-related. This technique allows the subjects to focus

*Corresponding author Tel : +86 25 84892551, Fax : +86 25 84891501

E-mail address: weihua.lu.ecn@gmail.com

† Recommended by Associate Editor Ki-Hoon Shin

© KSME & Springer 2018

on the expected dimensions since the categories are already provided, so a new set of assignments has to be proposed for each application [11].

Researchers commit to finding more intuitive and understandable assessment scales for developing measurement instruments. The common challenge is that the subjects have to understand the assessment scales and easily reach a consensus in their interpretations, which influences the effectiveness and reliability of the instrument [9]. Norman noted that the visual and behavioral states exist in a subconscious space, while the verbal state exists in a conscious space, which means that sensorial processes (right brain) input in an intuitive way, whereas verbal processes (left brain) input in an analytical way [12]. Therefore, using language-independent scales, like animation, picture, or sound for measuring emotion brings a higher reliability.

The auditory parameter method (AP) is a non-verbal instrument which proposes an original extension to the audio modality [13]. AP method enables quantification of users' emotional responses via interesting and easy association tasks (sound to product image) via a programmed software interface. The set of auditory stimuli is defined as generic assessment scales, interpreted by some perceptual variables. The emotions suggested by the auditory samples are assumed to be representative of the emotions suggested by the products.

In the previous study [13], the audio modality was explored as a means to elicit user emotion, and a new audio-based user emotion measurement tool AP method was defined. The application case concerned eyeglass frame, because it is an interesting product from a connotative point of view. The proposed AP method was also compared with the conventional approach SD method (control group) using experimental study. The statistical results indicated that the AP method demonstrates interesting qualities to reveal the users' desires for form and transmits to technical specifications for designers, which are presented in terms of emotional responses.

Emotion measurement tools may have limits in cross-culture use, just as the aesthetics and experienced preferences differ somewhat from culture to culture [14]. Due to the inevitable perception gaps and ambiguity of the cognitive process, the emotion created by a piece of music may be different between individuals. To propose a more public acceptance pool of music samples, an iterative design process is needed for continuous improvement. The current paper is a follow-up study, with comparative experiments duplicated in both China and France, to investigate and validate whether the AP method is resistant to interference in cross-cultural and interethnic context. Considering the continuity of the study, verbal protocol SD method was conducted as control group, and the pre-defined product pool eye-glass frames were used to illustrate the application.

In the present paper, the organization of comparative experiments is described in Sec. 2. The analysis and discussion of the results are presented in Secs. 3 and 4. Finally, conclusions and perspectives are drawn in Secs. 5 and 6.

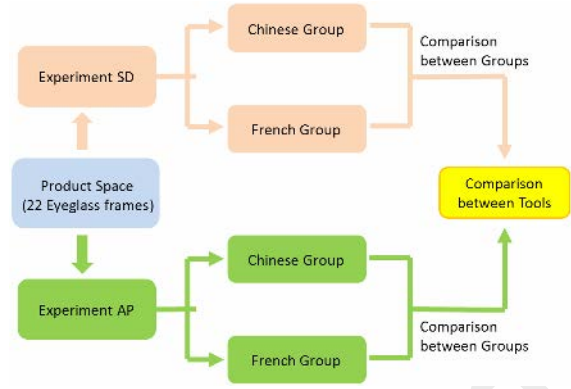


Fig. 1. Organization of the comparative experiments.

2. Material and method

We employed two parallel surveys with Chinese and French participants based on both auditory parameter method (AP) and semantic differential method (SD), shown in Fig. 1. The product under study was eyeglass frames, a product space of 22 samples was proposed for evaluation to both groups. Statistical methods allowed a comparison of the results between groups and tools.

2.1 Product space

From a design perspective, eyeglass frames are very interesting products, which integrates functionality, aesthetics, and individuality. Particularly, eyeglass frames transmit to the owner a controlled affective image [15, 16].

Described in preliminary research [13], full-frame samples were selected and geometrical details have been simplified into four shape factors with different levels to control the design variables: frame leg width (wide vs. narrow), rim profile (8 forms), rim aspect ratio (high vs. wide), and rim thickness (thick vs. thin).

An experimental design with 22 samples was defined and proposed, illustrated in Fig. 2. For the purpose of focusing experimental participants' attention on shape only, no information was regarding to material, color, and texture.

2.2 SD experiment

The SD experiment was carried out with the semantic differential method. The process consists of two parts: The selection of semantic attributes and the layout of the scale [17].

2.2.1 Selection of semantic attributes

Eight pairs of semantic attributes (see Table 1) were selected by analyzing previous publications in Kansei Engineering [18], in particular on the perception of eyeglass frames [19], with their Chinese and French translations listed.

2.2.2 Layout of the scale

For each semantic attribute, an assessment was imple-

Table 1. List of the 8 semantic attributes in the SD experiment.

No.	Attributes		
	English	Chinese	French
1	Feminine – masculine	女性的-男性的	Féminine - masculine
2	Common - special	普通的-特别的	Commune - particulière
3	Original - dull	新颖的-无趣的	Originale - banale
4	Delicate - rough	精美的-粗糙的	Recherchée - basique
5	Funny - serious	趣味的-庄重的	Drôle - sérieuse
6	Obedient - rebellious	顺从的-桀骜不驯的	Sage - rebelle
7	Modern - retro	时髦的-复古的	Moderne - rétro
8	Smart - ordinary	讲究的-平凡的	Élégante - vulgaire

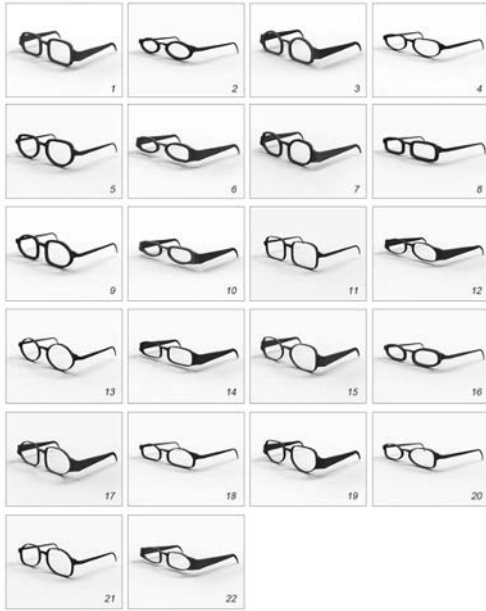


Fig. 2. Product space of the 22 samples.

mented on an unstructured scale (0-100) to get a continuous value. Subjects were advised to take their time in making the assessment on a graphical interface.

After evaluating the samples by the participants, the 22 designs of the product space can be described by a matrix, named V_{SD} , which represents the average value of the assessment according to all the attributes (dimension $m \times p$ matrix, m refers to the number of products, p is the number of semantic attributes).

2.3 AP experiment

2.3.1 Mechanism of AP experiment

The AP method has been exhaustively described in Lu and Petiot 2014 [13]. We consider here only the principles of this method. For more information, the reader is advised to refer to the initial publication.

The principle of the auditory parameter method is based on a multimodal matching, which corresponds to an association between a representative auditory stimulus and a picture of a product. The task of the participant is to select among a set of sounds, the audio stimulus that best fits the picture presented, from an emotional point of view.

Given the limited short term memory of humans for sounds, only three auditory stimuli are proposed at each selection phase. To get more accurate information, four successive selection stages are proposed to the subject, explained in Fig. 3 (sounds are represented by forms, applied to each design of the product space). Under simple random sampling without replacement, a participant has to select the most representative auditory stimulus for the proposed designed product in the first three phases. In the fourth phase, the participant has to choose the most representative one again among the three previous selected options. Finally, for each product sample, there are nine different auditory stimuli proposed by the algorithm. Two stimuli are selected once, and only one is selected twice.

The results of the selection process of the subjects during the AP experiment are recorded in a frequency matrix, named F (dimension $m \times n$ matrix, m refers to the number of objects of the product space, n is the number of sounds in the bank), which represents the number of times a given sound has been associated with a given product.

2.3.2 Definition of the auditory stimuli bank

The auditory stimuli bank is composed of a set of generic sounds designed for the measurement of user emotions.

Two versions of the auditory stimuli have been investigated: A selection from the international affective digitized sounds [20, 21], which is a collection of 137 ecological sounds, and a bank of music samples defined by the authors [13].

In a previous study, we showed that a version of ecological sound was sufficient to provide significant results, which agreed with empirical evidence, and validated by a statistical test [20]. However, the ecological sounds are context dependent and may evoke emotions that are interrelated with a scenario (e.g., baby crying may always arouse stress). Compared to ecological sounds, music has some prominent advantages. Music and design are highly interrelated in the emotional representation of both creating and evaluating processes. The emotions conveyed by music are often more complex than simple rules, which is the same as emotions elicited by a design. Moreover, music has more abundant choices for the effect of the combination of different elements of a musical sequence.

We adopted a bank of music samples described by four perceptual dimensions in Table 2, which were determined after sorting task and multidimensional scaling analysis, described in Lu and Petiot 2014 [13]. The perceptual dimensions are given in a matrix, named P (dimension $n \times k$ matrix, n refers to the number of sounds, and k is the number of variables).

Table 2. List of the 4 variables in the AP experiment.

No.	Perceptual dimensions
1	Funny, lively - serious, calm
2	Provocative, odd - serene, mellow
3	Depressed, calm - technological, excitant
4	Dynamic, fashion - peaceful, relaxed

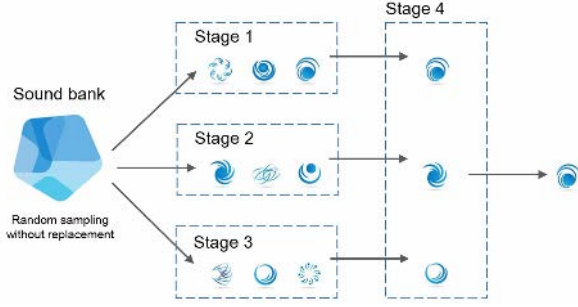


Fig. 3. Illustration of the four-stage selection procedure.

Finally, the objects of the product space are described by the matrix V_{AP} , the product of the frequency matrix F with the variables matrix P of sounds ($V_{AP} = F * P$, $m * k$ matrix, m refers to the number of products, k is the number of variables).

2.4 Subjects and procedure

Eighty-eight college students, including 44 Chinese and 44 French, participated in these two experiments. They were undergraduate or graduate students of large public universities in China and France, between the ages of 18 and 24. They had similar educational specialties: Mechanical engineering and industrial design. There was no particular requirement on musical-background for the participants.

The two experiments were conducted in a calm computer room to avoid any external disturbance, using headphones. In a short introduction, the subjects were informed about the purpose of the experiments, and given an explanation of the functioning of the interfaces. They were told that they could take as much time as they needed, but to avoid too much reflection because we were interested in their spontaneous emotional responses.

For purpose of controlling carry-over effects, Williams Latin Square was conducted for presentation order. The experiment was fully balanced because the number of subjects (88 subjects) is a multiple of the number of products (22 products).

3. Results of the SD experiment

The average rating of the 22 eyeglass frames, obtained by the SD experiment, was computed for both Chinese and French groups (matrices V_{SDC} and V_{SDF}). In this case, the eyeglass frames were described by eight pairs of attributes.

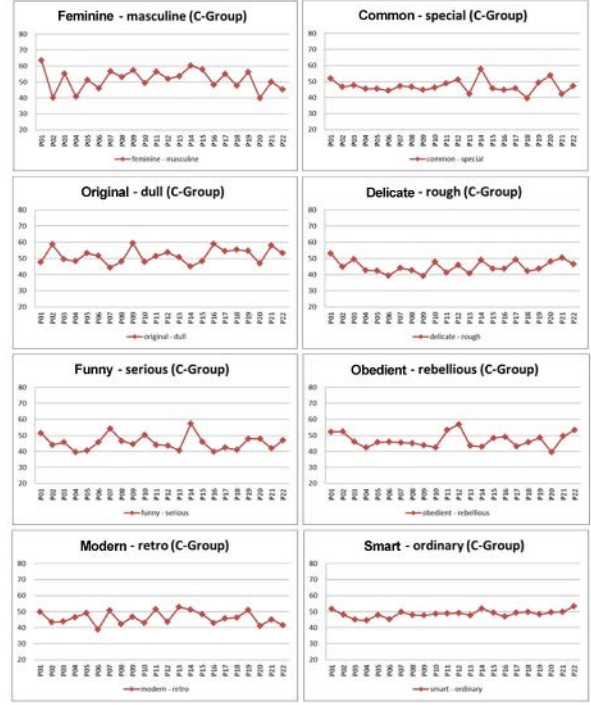


Fig. 4. Mean value of the assessments for each semantic attribute (Chinese group).

3.1 Result of the Chinese group of the SD experiment

3.1.1 Mean value of Chinese group by the SD experiment

The mean value of Chinese group obtained by the SD experiment is shown in Fig. 4. The average assessment towards 22 product samples was around the median value (50) for almost all terms, especially for the attribute “smart-ordinary”. The approximate flat arithmetic average values of Chinese group were uninformative regardless of the differences between the product samples, which did not allow an efficient characterization of the differences between the eyeglass frames.

3.1.2 Two-way ANOVA results of Chinese group of the SD experiment

To estimate the agreement inside the group, the assessments from Chinese participants of each semantic attribute were analyzed by two-way ANOVA with the two factors “subject” and “product”. As listed in Table 3, for most of the attributes, except “feminine - masculine”, the effect of the explanatory variable (subject) is significant ($p < 0.05$) and the effect of the product is not significant ($p > 0.05$). That means the ratings for these attributes are mainly influenced by the difference between subjects, rather than the differences among product samples. Only one attribute, feminine-masculine, obtains a significant product effect, which is a sign of perceived differences between the samples. The agreement inside the Chinese group is weak, and the average value is not representative of the assessments of the group. The individual evaluations of

Table 3. Two-way ANOVA result of each semantic attribute of Chinese group.

Attribute	Factor	Fisher F-test	P-value	Significant
Feminine - masculine	Subject	$F(43, 903) = 2.92$	< 0.0001	Yes
	Product	$F(21, 903) = 3.18$	< 0.0001	Yes
Common - special	Subject	$F(43, 903) = 1.95$	< 0.0001	Yes
	Product	$F(21, 903) = 1.51$	0.067	No
Original - dull	Subject	$F(43, 903) = 2.55$	< 0.0001	Yes
	Product	$F(21, 903) = 1.17$	0.268	No
Delicate - rough	Subject	$F(43, 903) = 3.84$	< 0.0001	Yes
	Product	$F(21, 903) = 1.23$	0.217	No
Funny - serious	Subject	$F(43, 903) = 3.83$	< 0.0001	Yes
	Product	$F(21, 903) = 1.80$	0.015	No
Obedient - rebellious	Subject	$F(43, 903) = 3.58$	< 0.0001	Yes
	Product	$F(21, 903) = 1.60$	0.043	No
Modern - retro	Subject	$F(43, 903) = 4.28$	< 0.0001	Yes
	Product	$F(21, 903) = 1.17$	0.269	No
Smart - ordinary	Subject	$F(43, 903) = 3.69$	< 0.0001	Yes
	Product	$F(21, 903) = 0.36$	0.997	No

the Chinese participants are too diverse, which have important conflicting opinions towards particular product sample. We consider that there are three main causes driving this.

Understanding obstacles: Attributes, the assessment scales in the SD experiment, are pure antonymous adjective pairs, which are usually not enough for solid understanding. Many understanding obstacles are caused by cultural factors rather than the language itself. Taking “common” and “ordinary” for instance, they mainly have the same meaning “普通的” (people or things are not special or different in any way) in Chinese, which leads to difficulty in translation. The weak difference can be appreciated along with the context in phrase or sentence. Compared to “common”, “ordinary” puts particular emphasis to the contrary of “superior”, and it has a little derogatory sense, for instance, “ordinary-looking”.

Value conflict: Chinese consumers are just at the crossroads of modernism and traditionalism, facing both Chinese and Western cultures. Although changes and conflicts are happening, the traditional segments and values still remain. For some people, they satisfy the fashionable society with rapid changes, while for others, they remain conservative values. It's difficult to have all the participants to have the same interpretation towards the Kansei pairs, such as “modern - retro”, “obedient - rebellious”, and “common - special” against this background.

Regional disparity: Given China's expanse and large population, it has strong regional cultures (e.g., the rich coastal area and western depressed area), as well as sharp differences in urban and rural cultures associated with income, dialect, and custom. Although the participants are college students at similar ages, they are from different regions with different backgrounds, which causes variation in assessments, even

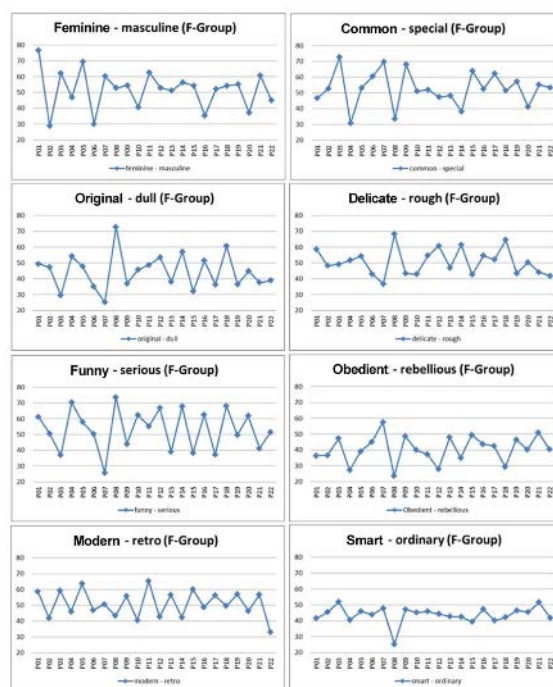


Fig. 5. Mean value of the assessments for each semantic attribute (French group).

opposite views and interests.

3.2 Results of the French group of the SD experiment

3.2.1 Mean value of French group by the SD experiment

The average assessment towards 22 product samples of the French group by the SD experiment was calculated as well, shown in Fig. 5. The ratings of each product sample are differentiable and significant compared to the Chinese group. The French participant of the SD experiment could reach a consensus towards the same product, and their assessments are distinguishable, which allow an efficient characterization of their preference towards the specific design samples.

3.2.2 Two-way ANOVA results of the French group of the SD experiment

To estimate the agreement inside the French group, we analyzed the assessments of each semantic attribute by two-way ANOVA with the two factors “subject” and “product” (see Table 4). For all the semantic attributes, the effect of the explanatory variables, subject and product, is significant. The higher F-test value of the factor “product” compared to the factor “subject” is a sign of the discriminability of the product samples.

3.3 Comparison between the groups of the SD experiment

The estimation between Chinese and French groups was analyzed with both univariate and multivariate statistical methods.

Table 4. Two-way ANOVA result of each semantic attribute of French group.

Attribute	Factor	Fisher F-test	P-value	Significant
Feminine - masculine	Subject	F(43, 903) = 2.62	< 0.0001	Yes
	Product	F(21, 903) = 9.55	< 0.0001	Yes
Common - special	Subject	F(43, 903) = 3.91	< 0.0001	Yes
	Product	F(21, 903) = 9.75	< 0.0001	Yes
Original - dull	Subject	F(43, 903) = 2.19	< 0.0001	Yes
	Product	F(21, 903) = 9.25	< 0.0001	Yes
Delicate - rough	Subject	F(43, 903) = 2.87	< 0.0001	Yes
	Product	F(21, 903) = 4.90	< 0.0001	Yes
Funny - serious	Subject	F(43, 903) = 2.38	< 0.0001	Yes
	Product	F(21, 903) = 12.18	< 0.0001	Yes
Obedient - rebellious	Subject	F(43, 903) = 2.04	< 0.0001	Yes
	Product	F(21, 903) = 6.29	< 0.0001	Yes
Modern - retro	Subject	F(43, 903) = 2.65	< 0.0001	Yes
	Product	F(21, 903) = 4.94	< 0.0001	Yes
Smart - ordinary	Subject	F(43, 903) = 3.49	< 0.0001	Yes
	Product	F(21, 903) = 3.28	< 0.0001	Yes

3.3.1 Univariate analysis of the SD experiment

We calculated the Kendall coefficient between the groups for the average values of each semantic variable, shown in Table 5. The correlations between Chinese group and French group are mostly low. Only the assessments according to the semantic attributes “feminine - masculine” and “modern - retro” are significantly correlated ($p < 0.05$).

3.3.2 Multivariate analysis of the SD experiment

We calculated another metric to assess the agreement between multidimensional ratings: the RV coefficient [22]. The coefficient was 0.206, which measures the closeness between the matrices V_{SDC} and V_{SDF} . It was compared later to the value obtained with the AP experiment.

To estimate the consensus between groups, generalized procrustes analysis (GPA) [23] was carried out to define the degree of agreement between the two different configurations (Chinese group and French group), and to assess the similarity between families of terms used to describe the products. GPA is a multivariate technique commonly used in sensory analysis to produce a consensual representation from a set of different individual data matrices (called configurations), and to represent the consensus via principal components analysis (PCA).

GPA was conducted to explore the consensus between average matrices V_{SDC} and V_{SDF} . The consensus ratio R_c between the two groups was 59 %. It signifies that after transformation of the data, 59 % of the total variance is represented by the consensus configuration. The statistical test (permutation test) on R_c indicates that this consensus is significant with a confidence = 90.2 %. The general agreement between the two groups is then not very important.

Table 5. Kendall coefficient between the groups for the average values of each semantic attribute and significance test (p-value).

Attribute	Kendall coefficient (p-value)
Feminine - masculine	0.515 (p < 0.05)
Common - special	-0.134 (p = 0.4)
Original - dull	-0.004 (p > 0.9)
Delicate - rough	0.013 (p > 0.9)
Funny - serious	-0.074 (p = 0.65)
Obedient - rebellious	0.013 (p > 0.9)
Modern - retro	0.403 (p < 0.05)
Smart - ordinary	-0.169 (p = 0.28)

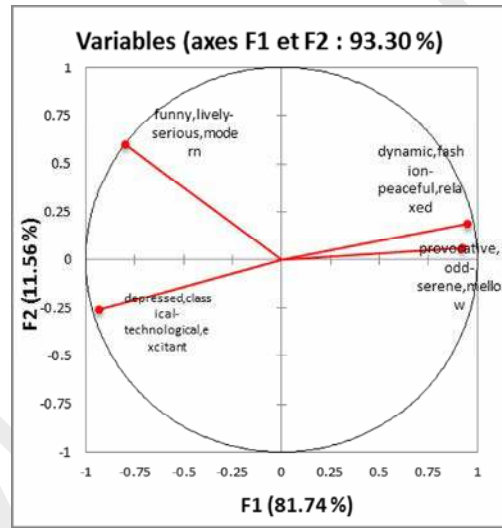


Fig. 6. PCA of the matrix V_{APC} (loadings of the variables - AP experiment - Chinese group).

4. Results of the AP experiment

The characterization of the 22 eyeglass frames obtained with the AP experiment was computed for both Chinese and French groups (matrices V_{APC} and V_{APF}). The eyeglass frames were described by four perceptual parameters.

4.1 PCA results of Chinese group of the AP experiment

The matrix V_{APC} was analyzed using standardized principal components analysis (PCA) to uncover the underlined structure of the data. The first two factors of PCA represent 93.3 % of variance. These two factors are considered as sufficient to represent the differences between the products. The loadings of the variables (see Fig. 6) are useful for interpreting the meaning of the factors. Factor 1 is mainly created by the variables “dynamic, fashion - peaceful, relaxed”, “provocative, odd - serene, mellow”, and “depressed, calm - technological, excitant”. The variable “funny, lively - serious, calm” contributes to factor 2.

Illustrated on the factorial planes of the Chinese group in Fig. 7 are the coordinates of the 22 eyeglass frames. The simi-

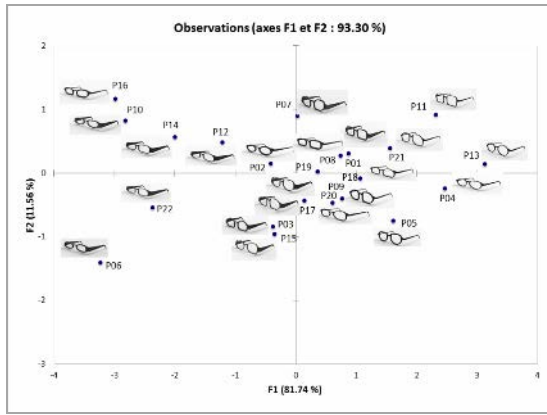


Fig. 7. PCA of the matrix V_{APC} (locations of the eyeglass frames - AP experiment - Chinese group).

larities and differences between samples are determined by the distances between them. In accordance with the distribution of the eyeglass frames, it's obvious that the designs with thin rim and narrow frame leg have an important score on the positive direction of factor 1, which evokes “mellow, peaceful, classical, and relaxed” feelings. The contrary design schemes with wide rim and wide frame leg are located on the opposite direction, which looks like more fashion, dynamic, and technological. In terms of factor 2, the eyeglass frame samples P03 and P15 are more funny and lively among the 22 samples. The sample P06 is particular, with extremely high value on “technological, excitant, fashion, and lively”.

4.2 PCA result of French group of the AP experiment

The matrix V_{APF} was analyzed using standardized PCA as well. The first two factors of PCA represent 87.8 % of variance, which is sufficient to represent the differences among the eyeglass frames (see Fig. 8). Compared to the Chinese group, the interpretation of the factors is very similar. The first factor is merely reversed; the second factor is almost the same, mainly created by the variable “funny, lively - serious, calm”.

The locations of the 22 eyeglass frames on the factorial planes of the French group are given in Fig. 9. We can notice on this graph that the two groups produced very similar results; the locations of the eyeglass frames are similar on the two first factors. The designs with thin rim and narrow frame leg are distributed on the opposite pole of the first factor, opposite to the designs with wide rim and wide frame leg. The eyeglass frame P06 is an extreme sample on the first factor, similar to the result of Chinese group. The eyeglass frames P03 and P15 have important scores on the second factor, which look like particularly funny and lively.

4.3 Comparison between the groups for the AP experiment

To examine the consensus between groups of the AP experiment and compare with the SD experiment, the same univariate and multivariate statistical analyses were conducted on

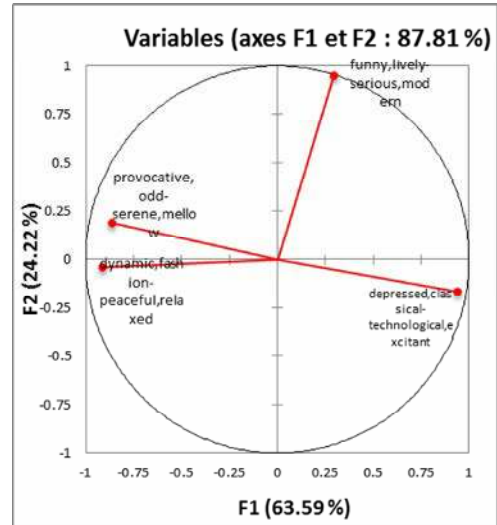


Fig. 8. PCA of the matrix V_{APF} (loadings of the variables - AP experiment - French group).

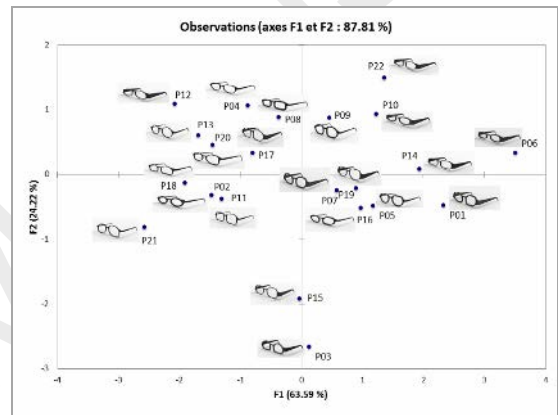


Fig. 9. PCA of the matrix V_{APF} (locations of eyeglass frames - AP experiment - French group).

matrices V_{APC} and V_{APF} .

4.3.1 Univariate analysis of the AP experiment

The Kendall coefficient was calculated between the groups for the average values of each perceptual parameter (see Table 6). The assessments towards the parameter “depressed, calm - technological, excitant” and “dynamic, fashion - peaceful, relaxed” are significantly different from 0 ($p < 0.05$). Compared to the SD experiment, the coefficients of the AP experiment between groups are higher as a whole. Half of the ratings of the AP experiment between Chinese group and French group are correlated.

4.3.2 Multivariate analysis of the AP experiment

The RV coefficient between the average ratings of the 22 eyeglass frames obtained by the AP experiment between groups (matrices V_{APC} and V_{APF}) was computed for comparison as well. The correlation between the groups is more important with the AP experiment ($RV = 0.39$) than with the SD

Table 6. Kendall coefficient between the groups for the average values of each dimension of the AP experiment.

Parameter	Kendall coefficient (p-value)
Funny, lively - serious, calm	0.152 (p = 0.34)
Provocative, odd - serene, mellow	0.203 (p = 0.2)
Depressed, calm - technological, excitant	0.455 (p < 0.05)
Dynamic, fashion - peaceful, relaxed	0.359 (p < 0.05)

experiment (RV = 0.206).

Though PCA results are interesting, they are unable to assess the agreement between the two groups. GPA was conducted to explore the consensus between matrices V_{APC} and V_{APF} . The consensus ratio R_c between the two groups was 75.9 %. It signifies that after transformation of the data, 75.9 % of the total variance is represented by the consensus configuration. The statistical test (permutation test) on R_c indicates that this consensus is significant (confidence = 99.7 %). The general agreement between groups of the AP experiment is better than those of the SD experiment.

5. Discussion

There are three main possible reasons to explain the differences between groups for the SD experiment.

- (1) Intercultural differences in the perception of the product samples;
- (2) Distortion in translation: The translation of attributes does not represent exactly the same notion in both languages;
- (3) Assessment error of the participants (due to the experimental protocol).

Concerning the AP experiment, there are three main possible reasons to explain the differences between groups as well.

- (1) Intercultural differences in the perception of the product samples;
- (2) Intercultural differences in the perception of the music samples;
- (3) Assessment error of the participants (due to the experimental protocol).

Subjective reports of Chinese group via the SD approach were uninformative, and the correlation between Chinese and French groups was not significant. The similar results of PCA and high consensus examined by univariate and multivariate statistical analyses indicated that the two groups of the AP experiment have similar perception towards the same product samples. The statistical analysis indicated that the results of the AP experiment are more similar between groups than the results of the SD experiment.

If we consider that the assessment error is similar for both experiments, and the perception of the music samples is similar for both groups, given that we obtained similar results with the AP method, the difference observed between the SD experiment is considered to be only due to the effect of the translation of the adjectives. To verify this assumption, the matrices

F concerning the choices of the participants, before multiplying by the matrix P , should be similar. Therefore, we estimated the consensus between matrices F with the Mantel test [24].

We transferred two matrices F into square dissimilarity matrices by computing the Pearson similarity between the rows of each matrix F . Then the two square matrices were examined with Mantel test. As the computed p-value (< 0.0001) is lower than the significance level $\alpha = 0.05$, we should reject the null hypothesis that there is no correlation between the matrices. As a conclusion, we can consider that the associations between music samples and products for Chinese and French groups are similar.

How to represent the user emotional responses to products is a challenging issue in learning design. Sometimes, without any technical experience, the users do not know what they really want until their preferences are violated. In consideration of the characteristics of user emotions, such as intangible, mixed, subtle, and inexpressible, the expression or representation by terms is difficult because they are language and culture dependent.

Subjectivity is a difficulty that cannot be avoided during the user emotion research process for all, because it is subjective per se. An emotion is an internal sensation, unfortunately, all the presently available measuring methods are external methods interpreting different body expressions. The definition of assessment scales is the key point, since it influences the effectiveness of the instrument, the difficulty of the task, application range, type of data, etc. Thus, exploring for a more intuitive universal assessment scale for affective responses elicited by product is endless.

The outcome of this experimental study verified that audio stimuli have superiorities in cross-cultural and interethnic situations unlike languages. Nowadays, globalized markets and the increasing technological sophistication mean more intense competition than ever before. An effective design requires consideration of these influences and should be planned to accommodate the needs of people with diverse cultural backgrounds. The AP method is implemented mostly at a subconscious level, which makes the respondents able to reveal the complex emotion profile of a product in an intuitive manner. This superiority turns the AP method into a reliable emotion measurement technique, in its application for the adaptation of products in overseas markets.

6. Conclusions

We carried out a comparative study to validate the stability and effectiveness of the auditory parameter method (AP) (non-verbal protocol). The experiment based on semantic differential method (SD)(verbal protocol) was considered as experimental control group. The experimental surveys on the same product samples were administered to two panels, with Chinese and French participants.

The statistical analysis indicates that the AP experiment has

more similar assessments between groups than the SD experiment. The data of the French group on the SD experiment revealed the tendency of their preference towards the specific design samples. Unfortunately, the flattened mean values of Chinese group were not informative. This observed difference between groups within the SD experiment was mainly due to the effect of the translation of the attributes.

The statistical results of the AP experiment indicated that both the frequency matrices F and description matrices V of the AP experiment were highly correlated between groups. The proposed 24 music samples of the AP method, serving as generic emotion assessment scale, have slight intercultural differences in the perception and interpretation.

The outcome of this study provides an effective and reliable tool for product emotion quantification to reveal the true and often complex emotion profile of a product in cross-cultural and interethnic context. The understanding of user emotions and the measurement of emotions become crucial components for a creative and innovative design process. To be able to create a larger database of users' perception, it is essential to shift from standard verbal methods for information gathering towards a novel method. The AP experiment can attract numbers of people from various demographic and cultural backgrounds, and the dissemination of the AP method and its results improve the communication between designer and user.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.51505218), the Fundamental Research Funds for the Central Universities of China (NS2015048) and the National Research Foundation of Korea (Korea-China Young Scientist Exchange Program). The authors appreciate the contributions of the students from Nanjing University of Aeronautics and Astronautics in China and Ecole Centrale de Nantes in France, who took part in the user tests.

References

- [1] D. McDonagh, A. Bruseberg and C. Haslam, Visual product evaluation: Exploring users' emotional relationships with products, *Applied Ergonomics*, 33 (3) (2002) 231-240.
- [2] B. C. Kim and S. H. Park, Development of a piggyback-type transfer assist system to assist caregivers with patients being unable to move themselves, *J. of Mechanical Science and Technology*, 29 (9) (2015) 3761-3769.
- [3] P. W. Jordan and A. Macdonald, Pleasure and product semantics, M. A. Hanson (Editor), *Contemporary ergonomics*, London: Taylor & Francis (1998) 264-268.
- [4] H. M. Khalid, Embracing diversity in user needs for affective design, *Applied Ergonomics*, 37 (4) (2006) 409-418.
- [5] H. Z. Huang, Y. Li, W. Liu, Y. Liu and Z. Wang, Evaluation and decision of products conceptual design schemes based on customer requirements, *J. of Mechanical Science and Technology*, 25 (9) (2011) 2413-2425.
- [6] M. Ding, W. Dong and C. C. Yang, Product color design based on multi-emotion, *J. of Mechanical Science and Technology*, 27 (7) (2013) 2079-2084.
- [7] N. Crilly, A. M. Maier and P. J. Clarkson, Representing artifacts as media: Modelling the relationship between designer intent and consumer experience, *International J. of Design*, 2 (3) (2008) 15-27.
- [8] C. E. Osgood, G. J. Suci and P. H. Tannenbaum, *The measurement of meaning*, University of Illinois Press, Urbana, United States (1957).
- [9] P. M. A. Desmet, Measuring emotions: Development and application of an instrument to measure emotional responses to products, M. A. Blythe, K. Overbeeke, A. F. Monk and P. C. Wright (Editors), *Funology: From usability to enjoyment*, Springer Netherlands (2004) 111-123.
- [10] W. Lu, Contribution to the affective design of products -- an audio-based approach for the study of users' emotions, *Unpublished Ph.D. Dissertation*, Ecole Centrale de Nantes, Nantes, France (2013).
- [11] S. T. W. Schütte, J. Eklund and J. R. C. Axelsson, Nagamachi M. Concepts, methods and tools in Kansei engineering, *Theoretical Issues in Ergonomics Science*, 5 (3) (2004) 214-231.
- [12] D. A. Norman, *Emotional design: Why we love (or hate) everyday things*, Basic Books (2003).
- [13] W. Lu and J. F. Petiot, Affective design of products using an audio-based protocol: Application to eyeglass frame, *International J. of Industrial Ergonomics*, 44 (3) (2014) 383-394.
- [14] H. Maclay and E. E. Ware, Cross-cultural use of the semantic differential, *Behavioral Science*, 6 (3) (1961) 185-190.
- [15] H. Yanagisawa and S. Fukuda, Interactive design support system by customer evaluation and genetic evolution: Application to eye glass frame, *Proc. of the 7th International Conference, KES 2003*, LNAI 2774, Springer-Verlag Berlin Heidelberg (2003) 481-487.
- [16] T. K. Huang, M. Y. Ma and W. C. Tseng, Preference-based analysis of black plastic frame glasses, *Proc. of International Conference on Kansei Engineering and Emotion Research*, Paris, France (2010) 226-235.
- [17] J. Al-Hindawe, Considerations when constructing a semantic differential scale, *La Trobe Working Papers in Linguistics*, 9 (3) (1996) 41-58.
- [18] M. Nagamachi, *Kansei/affective engineering*, Bosa Roca, Taylor & Francis, United States (2010).
- [19] C. H. Lo and C. H. Chu, Affective modelling: Profiling geometrical models with human emotional responses, *Computer Graphics Forum*, 28 (7) (2009) 1811-1820.
- [20] W. Lu and J. F. Petiot, A sound-based protocol to study product emotions, *International J. of Product Development*, 16 (3/4) (2012) 207-224.
- [21] M. M. Bradley and P. J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, *J. of Behavior Therapy and Experimental Psychiatry*, 25 (1) (1994) 49-59.

- [22] P. Robert and Y. Escoufier, A unifying tool for linear multivariate statistical methods: The RV coefficient, *Applied Statistics*, 25 (3) (1976) 257-265.
- [23] J. C. Gower, Generalized procrustes analysis, *Psychometrika*, 40 (1) (1975) 33-51.
- [24] P. E. Smouse, J. C. Long and R. R. Sokal, Regression and correlation extensions of the mantel test of matrix correspondence, *Systematic Zoology*, 35 (4) (1986) 627-632.

Accepted Manuscript