Automatic Estimation of Affective Impressions from a Single Sound-symbolic Word and Word-based Visualization of Perceptual Space

Maki SAKAMOTO*

The University of Electro-Communications, 1-5-1 Chofugaoka, Chofui-shi, Tokyo 182-8585, Japan
Maki.sakamoto@uec.ac.jp

Abstract: The semantic differential (SD) was mainly developed by the US psychologist Charles E. Osgood in 1950s. Since then, it has been widely applied to capture the affective and cognitive factors of respondents' attributions to selected concepts and objects on a multidimensional level. This method asks respondents to analyze their affective impression and perceptual experiences one by one. On the other hand, Gestalt psychology implies that human understands external stimuli as whole rather than the sum of their parts. Therefore, we proposed a system that can automatically estimate multidimensional ratings of affective impressions of objects from a single sound-symbolic word that has been spontaneously and intuitively expressed by a user. When a user inputs a sound-symbolic word into the system, the system refers to a database of phonemes and their auditory impressions and calculates ratings in terms of fundamental scales of affective and perceptual experiences. In this paper I will outline the advantage of our method in visualizing our affective impressions of objects and our perceptual space.

Keywords: *Affective impression, Sound-symbolic word, System construction, Visualization, Perceptual space*

1. INTRODUCTION

Sounds can be expressed by sound symbolic words (hereafter, SSWs) such as tick-tock and ding-dong. Japanese people frequently verbalize the perceived auditory information from environmental sounds by SSWs. In our previous research [1], we asked participants to describe 60 environmental sounds by SSWs and evaluate the affective impression of the 60 environmental sounds using the semantic differential (SD) method with a seven-point scale (very pleasant, +3; pleasant, +2; somewhat pleasant, +1; neither pleasant nor unpleasant, 0; somewhat unpleasant, -1; unpleasant, -2; very unpleasant, -3).We analyzed the relationship between the pleasant/unpleasant evaluation and the phoneme of the SSW for each stimulus. The result showed that pleasant/unpleasant evaluation of sounds was associated with the phoneme of the SSW.

A sensory-sound correspondence can be found not only in SSWs expressing environmental sounds but also in words referring to visual shapes. This was demonstrated in landmark studies (e.g., buba/kiki for round and sharp shapes in studies by Ramachandran and Hubbard [2], respectively). We have investigated the sound symbolic associations in touch, specifically the association between the phonemes of Japanese SSWs for expressing tactile sensations and subjective evaluations of comfort/discomfort for touched objects.

Compared with other languages, Japanese has a large number of SSWs for tactile sensations, and associations between the phonemes of Japanese SSWs and typical categories of tactile sensations can be observed [3]. Most Japanese SSWs expressing tactile sensations consist of two-syllable repetitions (C1V1C2V2-C1V1C2V2, where C and V indicate a consonant and vowel, respectively, e.g., "sara-sara"), and the sound of the first syllable (C1V1, e.g., "sa" for "sarasara") is strongly associated with evaluations of tactile sensations. For instance, "sara-sara" and "zara-zara," which are different only in the first syllable of the repeated unit, denote completely different tactile sensations. Whereas the former is used for expressing a smooth and pleasant sensation, the latter is used for expressing a rough and unpleasant sensation.

The affective impression of various modalities such as sound, vision, and touch by an SSW is associated with phonological features of the SSW. Japanese SSWs make up a large body of vocabulary that can express the complex and minute qualities of multiple sensory experiences.

2. SYSTEM TO QUANTIFY AFFECTIVE IMPRESSION OF SOUND-SYMBOLIC WORDS

The features described in section 1 led us to the idea that we can calculate the multidimensional ratings of a word by integrating each phoneme's impression, and we have

proposed a system that can convert an SSW in Japanese into quantitative ratings in multiple sensory dimensions (26 pairs of tactile adjectives were adopted in the system [4]). In this paper I will outline the system published in [4] and the advantage of our method in visualizing our affective impressions of objects and our perceptual space.

We next explain the procedure for constructing the automatic estimation system.

2.1 Experiment to Build Sound-symbolic Database

First, we discuss an experiment that we performed to analyze the sound-symbolic associations of phonemes and 26 pairs of adjectives. In the experiment, participants viewed SSWs displayed on a monitor and rated their impressions in terms of the 26 bipolar adjective scales. Seventy-eight paid participants, aged 20 to 24 years (51 men and 27 women), participated in this experiment. They had no knowledge of linguistics and were all native Japanese speakers. They were unaware of the purpose of the experiments, and informed consent was obtained from all the participants before the experiment started.

To obtain sound-symbolic associations of all Japanese phonemes with the 26 pairs of adjectives, we selected word stimuli that included all varieties of Japanese phonemes. First, we made a list of 14,584 onomatopoeias by combining all sounds in the Japanese syllabary as two syllable forms (e.g., /sara/). Then, we also prepared repetition expressions of two-syllable forms (i.e., /sarasara/). In addition, we made another list of expressions by adding words with all types of special phonemes used in Japanese SSWs (syllabic nasals /N/, choked sounds /Q/, long vowels /R/, and adverbs ending in /Li/) to the two-syllable forms (i.e., /saraLi/). Second, from this large word list, 312 words that could be used as tactile SSW were selected by three experts (including the author) in psychology and linguistics. The trials started with the presentation of an SSW on the monitor, and participants were asked to report how they felt about it on a seven-point SD scale (e.g., for the comfortable–uncomfortable scale, participants selected one of the following seven points: 1, very comfortable; 2, comfortable; 3, slightly comfortable; 4, neither comfortable nor uncomfortable; 5, slightly uncomfortable; 6, uncomfortable; 7, very uncomfortable). The participants responded by pressing one of seven buttons. The time allotted for answering was unlimited, but almost all trials took less than 1 minute. The presentation order of SSWs was randomized among the participants. The orders and polarities of the bipolar scales were also randomized in the answer matrix. The experimental results produced 105,456 items of data (26 rating scales × 312 words × 13 participants). Then, we calculated the averages and standard deviations of the rating values for each scale of all words. The average of the standard deviation was 1.31. Ninety-eight percent of all data (105,181 items of data) were within 2.0 standard deviations of the average.

2.2 Sensory Image Estimation Model

To estimate the sensory impressions of SSWs, we constructed a model in which the following equation was used to predict each rating value:

$$Y = \sum_{i=1}^{13} X_i + Const.$$

(1)

where Y represents the rating values of the 26 scales, and X1–X13 are the corresponding values defined in Table 1. X1–X6 are respectively the mean values of the specific consonant, voiced sound/p-sound, contracted sounds, vowels, semivowels, and special phonemes in the first syllable. X7–X12 are the same categories for the second syllable, and X13 denotes the presence or absence of repetition in the word. Using the average of the rating values as the objective variables and the variation of phonemes as the predictor variables, we conducted class I mathematical quantification theory, which is a type of multiple regression analysis. As shown by Eq. (1), the rating values of an SSW can be calculated by a linear sum of the values (X1–X13) of the word. For example, the expression "sara" is composed of the first mora /sa/ (/s/ + /a/) and the second mora /ra/ (/r/ + /a/). Therefore, the value of the "rough–smooth" scale on a seven-point scale (1 = smooth, 7 = rough) is estimated using the Eq. (2)

Y = /s/ (X1) + absence (X2) + absence (X3) + /a/ (X4) + absence (X5) + absence (X6) + /r/ (X7) + absence (X8) + absence (X9) +/a/ (X10) + absence (X11) + absence (X12)+ absence (X13) + Const.

= (−0.05) + (−0.32) + (−0.05) + (0.46) + (−0.02) + (−0.03) + (−0.14) + (−0.1) + (0.05) + (−2.19) + (0.2) + (−0.02) + (0.05) + (0.01) + (3.75)= 2.05

(2)

Table 1: Correspondences between variables and phonemes			
First syllable	**Second syllable**	**Phonological characteristics**	**Variation of phonemes**
X_1	X_7	Consonants	/k/, /s/, /t/, /n/, /h/, /m/, /y/, /r/, /w/ or absence
X_2	X_8	voiced sounds / p-sounds	presence (/g/, /z/, /d/, /b/, /p/) or absence
X_3	X_9	contracted sounds	presence(/ky/, /sy/, /ty/, /ny/, /hy/, /my/, /ry/, /gy/, /zy/, /by/, /py/) or absence
X_4	X_{10}	Vowels	/a/, /i/, /u/, /e/, /o/
X_5	X_{11}	semi-vowels	/a/, /i/, /u/, /e/, /o/ or absence
X_6	X_{12}	special sounds	/N/, /Q/, /R/, /Li/ or absence
X_{13}		Repetition	presence (ex. huwa-huwa) or absence

The estimated value of 2.05 suggests that "sara" is associated with a smooth impression. The multiple correlation coefficients R2 between the predicted values and the mean values of the actual ratings (the values obtained from the participants) were used as indicators of prediction accuracy. For 20 scales, the R2 values were in the range of 0.80 to 0.90, and for the other six scales, they were higher than 0.90. Therefore, we considered our model to be satisfactory for estimating impressions of SSWs by analyzing the phonemes and forms of the words.

2.3 Advantage of our System

In our system, when a word that intuitively expresses an affective impression is input into the text field, information equivalent to evaluations against, for example, the 26 pairs of texture related adjectives is obtained from an analysis of the sounds of the word. Figure 1 shows example of outputs from our system for "sara-sara". "Sara-sara" shows higher ratings for "slippery," "dry," "nonelastic," "clean," "thin," and "light" feelings. This system enables us to analyze affective impressions with many criteria by expressing the sensation with only a single word, and this idea can be applied to any combination of phonemes (even to newly created words) in Japanese. To estimate the quantitative information of every possible SSW, we built a database of sound-symbolic associations for each phoneme, for example, with the 26 pairs of adjectives through psychological experiments. We described how we constructed a system that can estimate multidimensional rating scores from a single SSW based on sensory-sound associations. Our system can estimate not only SSWs established as Japanese vocabulary, but also newly created novel SSWs.
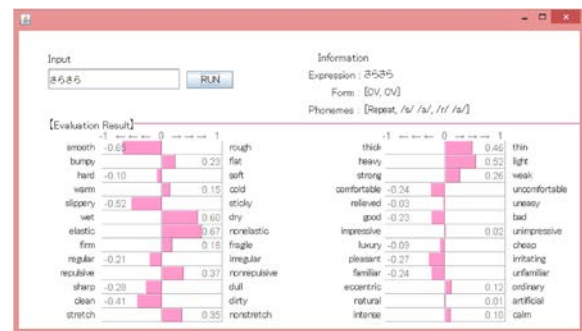

Figure 1: Example of output for "sara-sara"

3. APPLICATION OF SYSTEM

In this section, we introduce a word-based visualization of tactile perceptual space as an application of our system based on our previous research [4].

3.1 Distribution diagram of the sound-symbolic words

We collected Japanese SSWs using a Google search (conducted on June 6th, 2014). We found 76 words for which the number of phrase search results for "touch sensation like ****" was more than 100. We regarded the 76 words as those used frequently to express tactile sensations. We input each of the 76 SSWs into our system and obtained their rating scores. Using the scores of the six basic perceptual dimensions ("hard to soft," "rough to smooth," "bumpy to flat," "sticky to slippery," "wet to dry," and "warm to cold"), we performed a principal component analysis. Then we generated a distribution diagram of the SSWs using the first and second principal components as the horizontal axis and the vertical axis, respectively (Fig. 2). The cumulative contribution ratio of the first and second principal components was 80.0%. In this diagram, SSWs that express closely related sensations are also located close to each other. Additionally, we drew the six pairs of adjectives in the diagram based on the

principal component loadings. The distribution diagram enables us to grasp the tactile perceptual space intuitively through the relationships among SSWs, and allows us to identify the categorization of tactile sensations in a detailed way.

Furthermore, we applied a hierarchical cluster analysis (Ward's method) to the six values of the outputs of the words to visualize the structure of tactile perceptual space. We chose six categories because the cross-validated value of grouped words reached almost 90% and saturated at six categories. In Fig. 2, the six categories are shown by black circles. The distribution diagram with category circles is useful for developing a new material/product.

3.2 Tactile Material Collection using Word-based Visualization

The tactile materials can be placed on the diagram by using the connections between how the materials are felt and the SSWs. The way in which the SSWs are spatially distributed on the diagram guides people to locate materials precisely and intuitively. People can locate materials directly on the words, or between words referring to impressions of the words. After locating many materials on the diagram, we can easily grasp the relationships between the diagrammed materials. For example, when you need to collect 50 tactile materials that have a wide range of tactile qualities, you can place materials on the diagram and see what kind of tactile quality is lacking. Although material-based visualization can be changed depending on the variety of materials, the word-based diagram remains unchanged regardless of the number and variety of materials, and therefore the diagram is a useful way to explore the relationships between materials.
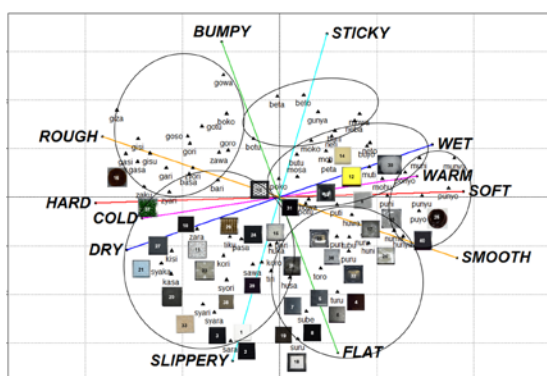

Figure 2: Distribution diagram of 40 tactile materials (see Table A3 in the Appendix). The diagram indicates that the cactegory including "gori-gori," "goso-goso," and "gisi-gisi" and the category including "gunya-gunya," "beto-beto," and "beta-beta" are lacking corresponding materials.

This word-based visualization might be used for product design and communication between designers and customers in the field of product design. If a customer requests tactile feeling by indicating a location on the diagram, a designer can develop a new material/product referring to the location of the SSW, and/or an existing material/product located nearby on the diagram. If this kind of visualization is used for a specific product, such as cosmetics or fabrics, only a collection of words is needed to make a new diagram suitable for the product.

4. CONCLUSION

We described how we constructed a system that could estimate multidimensional rating scores from a single SSW based on sensory–sound associations. Our system can transform any kind of SSW into information equivalent to evaluations against 26 pairs of adjectives. As an application of our system, we introduced word-based visualization of the tactile sensation categories based on the outputs of our system.

REFFERENCES

[1] Jinhwan Kwon and Maki Sakamoto; Visualization of Relation between Sound Symbolic Word and Perceptual Characteristics of Environmental Sounds, American Institute of Physics (AIP) Conference Proceedings, pp. 1-6, 2016.

[2] Vilayanur S. Ramachandran and Edward M. Hubbard; Hearing Colors, Tasting Shapes, Scientific American, 288, pp.43-49, 2003.

[3] Maki Sakamoto and Junji Watanabe; Bouba/Kiki in touch: associations between tactile perceptual qualities and Japanese phonemes, Frontiers in Psychology, 9, pp.1-12, 2018. DOI: 10:3389/fpsyg.2018.00295

[4] Ryuichi Doizaki, Junji Watanabe and Maki Sakamoto; Automatic Estimation of Multidimensional Ratings from a Single Sound-symbolic Word and Word-based Visualization of Tactile Perceptual Space, IEEE Transactions on Haptics, 10, pp.173-182, 2017. DOI: 10.1109/TOH.2016.2615923