Product Recommendation Method based on Onomatopoeia Expressing Texture

Ryuichi Doizaki[†] **Saki Iiba**[†] **Takashi Abe**[‡] **Takayuki Okatani**[‡] **Maki Sakamoto**[†]
[†]Graduate School of Informatics and Engineering,
The University of Electro-Communications
[‡]Graduate School of Information Sciences,
Tohoku University
sakamoto@inf.uec.ac.jp

Abstract

With the widespread use of online shopping in recent years, consumer search requests for products have become more diverse. Previous web search methods have used adjectives as input by consumers. However, given that the number of adjectives that can be used to express textures is limited, it is debatable whether adjectives are capable of richly expressing variations of product textures. In Japanese, tactile experiences are easily and frequently expressed by onomatopoeia, such as "*fuwa-fuwa*" which indicates a soft and light sensation. Onomatopoeia are useful for understanding not only material textures but also a user's intuitive, sensitive, and even ambiguous feelings evoked by materials. In this study, we propose a system to recommend products corresponding to product textures associated with Japanese onomatopoeia based on their symbolic sound associations between the onomatopoeia phonemes and the texture sensations. Our system quantitatively estimates the texture sensations of onomatopoeia input by users, and calculates the similarities between the users' impressions of the onomatopoeia and those of product pictures. Our system also suggests products which best match the entered onomatopoeia. An evaluation of our method revealed that the best performance was achieved when the SIFT features, the colors of product pictures, and text describing product pictures were used; Specifically, precision was 66 for the top 15 search results. Our system is expected to contribute to online shopping activity as an intuitive product recommendation system.

Keywords: Product Recommendation Method, Texture, Onomatopoeia, Color, Sound Symbolism

1 Introduction

With the widespread use of online shopping in recent years, consumer search requests for products have become more diverse. Not only are the designs of products, such as their colors and shapes, important factors to consumers, but so are the texture sensations of products, such as tactile sensations. Previous studies have proposed web search methods based on the tactile sensations of the material textures of products [1], [2]. These previous studies propose methods of recommending products to consumers through the input of adjectives such as "soft" or "smooth." However, there is a wide variety of products which can be described by the adjective "soft," such as towels, mouton (sheepskin) fur, rice cakes, and slime, all of which may have soft textures but can differ in their softness. An adjective cannot describe specific tactile information such as the kind or degree of softness of an object. Moreover, the number of adjectives for texture sensations is limited, and thus it is debatable whether such adjectives are capable of richly expressing product texture variations. Therefore, this paper proposes a method of using onomatopoeia based on the symbolic sound associations between onomatopoeia phonemes and texture sensations. Japanese is known to have a large number of onomatopoeic words to express texture sensations. In particular, tactile sensations are easily and frequently expressed by onomatopoeia. The softness of towels, mouton fur, rice cakes, and slime may be expressed, respectively, by "*fuwa-fuwa*" "*moko-moko*" "*mochi-mochi*" and "*funya-funya*". In addition, onomatopoeic expressions such as "*sube-sube*" and "*tsuru-tsuru*" are often used to express the sensation of a smooth surface, and "*neba-neba*" and "*jime-jime*" are often used to refer to sticky sen-

sations. Based on such associations, we propose a new method of estimating texture sensations which can be expressed by onomatopoeia that are suitable as input as consumer search requests.

2 Previous Studies and Purpose of Our Study

2.1 Effectiveness of onomatopoeia in tactile perception

The existence of synesthetic associations between sounds and sensory experiences, referred to as sound symbolism in the literature, has been demonstrated over the decades [3], [4]. It is also known that sensory-sound correspondence can be found not only in words referring to visual shapes, as demonstrated in the landmark studies [4], [5], which identified "mal/mil" and "bouba/kiki" for round and sharp shapes, respectively, but also in words referring to tactile, smell, and taste sensations. [6] has investigated the sound symbolism associations between the phonemes of Japanese sound symbolic words for expressing tactile sensations (i.e., onomatopoeia) and subjective evaluations of comfort or discomfort for touched objects. For example, onomatopoeic words expressing a sense of smoothness often use the consonant /s/ in the first syllable, as in "*sara-sara*" while those expressing roughness often use /z/ in the first syllable, as in "*zara-zara*". [6] also notes that onomatopoeia comprise a class of words whose speech sounds are systematically associated with sensory experiences, and recently they have been used in psychophysical studies as a unique index of sensation categories. [7] has also found that onomatopoeia can be used to differentiate various tactile sensations more easily than adjectives, and that onomatopoeia is thus useful for evaluating minute differences in texture sensations.

2.2 Quantification of affective texture sensations expressed by onomatopoeia

[8] described the sound symbolic relationship between Japanese phonemes and meanings (e.g., the vowel /i/ is associated with sensations of sharpness). [9] quantified the relationship between Japanese phonemes and their perceptions as shown in Table 1, and constructed a model to predict the mental images evoked by two-syllable repetitive onomatopoeia. Based on the model proposed by [9], [10] developed a system to estimate the mental image of various onomatopoeia input by users, as shown in Figure 1.

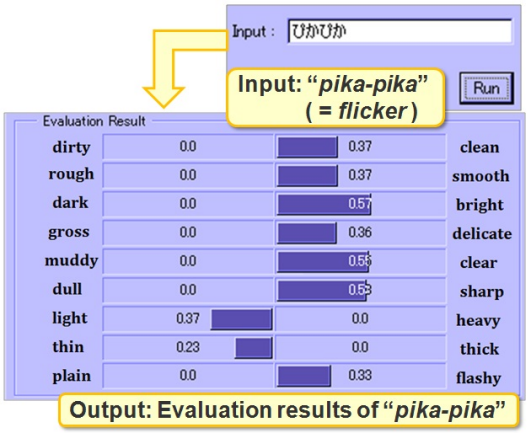

Figure 1. System for evaluating onomatopoeia sensations

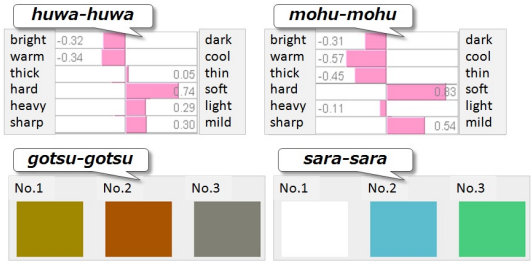

Figure 2. System for evaluating onomatopoeia tactial perceptions

[11] developed a system to evaluate tactile images associated with onomatopoeia based on sound symbolism, as shown in Figure 2. This system quantifies the mental images of input onomatopoeia using 43 adjective scales related to tactile sensations, and then recommends colors which match the impressions associated with the onomatopoeia based on the work of [12], which reported on the influences between the colors of a material and its tactile perceptions.

3 Procedure

3.1 Overview of our system

Figure 3 shows an overview of our system, which quantifies the mental images of textures by analyzing the phonological characteristics comprising the input onomatopoeia, and then recommends products which match the onomatopoeia based on their ranking obtained by

Table 1. Relationship between phonemes and images

Scale	Consonants /t/	/n/	/r/	Sounds voiced sounds	p-sounds
bright / dark	0.09	-0.21	1.21	-1.29	0.69
clear / dull	0.00	-0.45	0.77	-2.47	0.11
heavy / light	0.02	0.42	-0.30	2.58	0.01

calculating the similarities between the texture mental images of the onomatopoeia and the actual texture features of the product pictures. Because our system evaluates mental texture im-

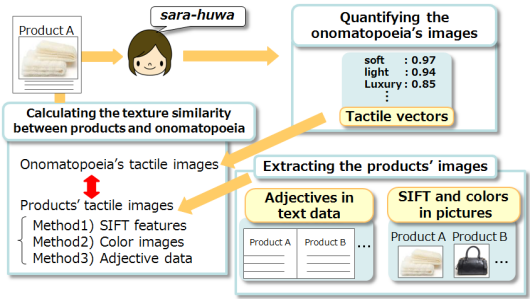

Figure 3. Overview of our system

ages expressed by onomatopoeia based on their phonological characteristics, we began our study by creating a "phoneme image" database. These phoneme image data specify what mental images are associated with each Japanese phoneme in relation to semantic differential (SD) scales. We conducted a psychological experiment to collect the data.

3.2 Stimuli

3.2.1 SD scales

SD scales included 43 pairs of adjectives appropriate for evaluating the design and texture perceptions of objects, such as bright/dark, warm/cool, thick/thin, smooth/rough, and wet/dry. 43 pairs of adjectives were extracted from previous studies related to evaluation of colors or tactile textures. In our method, there are no correlations between 43 adjectives, because our method can calculate the impression values of each adjective independently.

3.2.2 Onomatopoeia

Our method calculates the subjective impressions of onomatopoeia based on the impressions of each phoneme. Therefore, the experimental stimuli are required to include all varieties of Japanese phonemes, that is, basic phonemes (consonants /C/ and vowels /V/) as well as special phonemes (syllabic nasals /N/, choked sounds /Q/, long vowels /R/, and adverbs ending in /ri/) [13]. We created various combinations of sounds to consider the effects of sound order, that is, to investigate whether first-syllable and second-syllable phonemes could evoke different subjective impressions (e.g., between "*kasa*" and "*saka*"). First, we combined all sounds in the Japanese syllabary (from /a/ to /n/), and then created two-syllable expressions (i.e., /aa/, /ai/, . . . , /wan/, /nn/). We obtained a total of 11,075 words, including those made by repeating two-syllable onomatopoeic expressions (e.g., /aa-aa/, /ai-ai/). Moreover, we added 3,509 words with all types of special phonemes, such as /fuwari/ and /peQtari/. Second, we selected 312 words from these 14,584 words which were judged by 3 participants as onomatopoeic expressions for describing texture sensations. We confirmed that the selected 312 words covered every possible kind of phonemeand thus decided to use those 312 words as our experimental stimuli. Table 2 shows some stimuli examples.

3.3 Experiment: Evaluating onomatopoeia images

To create the phoneme image data, we conducted a psychological experiment to investigate what images could represent the 312 onomatopoeic words. Participants were asked to evaluate images for the 312 words (divided into 6 groups of 52 words) against 43 pairs of adjectives across 7-point SD scales, as shown in Figure 4. Onomatopoeic stimuli were displayed on a computer screen randomly for each participant. In total, 78 Japanese undergraduates aged 20 to 24 (51 males and 27 females, spread across 6 groups of 13 people) participated in the experiment.

After obtaining the evaluations (312 words

Table 2. Examples of onomatopoeia stimuli

unyo-unyo	*chiri-chiri*	*mosha-mosha*
gasa-gasa	*tsubu-tsubu*	*yore-yore*
gito-gito	*tsuru-tsuru*	*washa-washa*
gucho-gucho	*teka-teka*	*uNnyuri*
keba-keba	*toro-toro*	*kusyaR*
gotsu-gotsu	*doro-doro*	*zazaQ*
koro-koro	*nayo-nayo*	*zaraR*
gowa-gowa	*nyupu-nyupu*	*subeQ*
zazazaza	*nucha-nucha*	*doroRri*
zashu-zashu	*necha-necha*	*pahuQ*
sara-sara	*fuka-fuka*	*fuRwari*
zara-zara	*bunyu-bunyu*	*moQsa*
jori-jori	*punyu-punyu*	*boko-boko*
sube-sube	*fuwa-fuwa*	*fowa-fowa*
chiku-chiku	*becho-becho*	*muni-muni*
chapu-chapu	*beto-beto*	*washari*

× 43 scales × 13 participants), we calculated the standard deviations of each value and removed 275 evaluations with a standard deviation of 2.0 or over (i.e., 2 %of the total obtained data). By calculating the average values across participants, we thus collected the subjective impressions of all 312 onomatopoeic words (312 words × 43 scales).

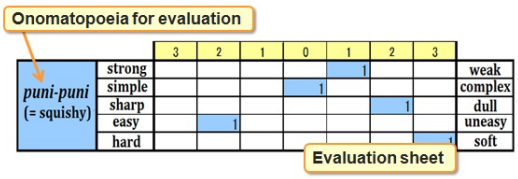

Figure 4. Example of answer sheet in experiment

3.4 Onomatopoeia image estimation model

Our method estimates texture sensations expressed by onomatopoeia based on their phonological characteristics. We designed an "onomatopoeia image estimation model" that quantifies the subjective images of onomatopoeia by calculating a linear sum of the individual phoneme mental images using the following equation (1).

$$\hat{Y} = X_1 + X_2 + \cdots + X_{12} + X_{13} + Const. \quad (1)$$

Here, $\hat{Y}$ represents the mental image estimations of the respective 43 scales for a given onomatopoeia. $X_1 \cdots X_{13}$ show the mental images of the respective phonemes, and $Const.$ represents constant terms. In more detail, $X_1 \cdots X_6$ respectively mean consonant, voiced sound/p-sound, contracted sounds, vowels, semi-vowels and special phonemes in first mora. $X_7 \cdots X_{12}$ respectively mean the same categories in second mora, and X_{13} does repetition.

To obtain the phoneme image data, we applied the onomatopoeia image data used in the experiment to our model, and calculated the images for each phoneme by making use of class I mathematical quantification theory. This theory is a kind of multiple regression analysis using the dummy variable.

For example, the expression "*fuwa-fuwa*" made by repeating /fuwa/, is composed of the first syllable /fu/ (/f/ + /u/) and the second syllable /wa/ (/w/ + /a/). Therefore, the image of "*fuwa-fuwa*" for the "hard/soft" scale is estimated by the following equation. Because the evaluated values are based on 7-points scales, the estimated value 6.28 suggests that "*fuwa-fuwa*" is strongly associated with an image of softness.

$$\begin{aligned} \hat{Y} &= /h/ + /u/ + /w/ + /a/ \\ &\qquad + repetition + Const. \\ &= 0.29 + 0.55 + 0.71 + 0.07 \\ &\qquad + 0.23 + 4.43 \\ &= 6.28 \end{aligned}$$

To evaluate the accuracy of our model, we compared the mental onomatopoeia images estimated by our model with the images of onomatopoeia obtained from participants through our experiment (hereafter, we refer to the former as "estimated images" and the latter as "real images"). We calculated the multiple correlation coefficient $\boldsymbol{R}$ values between the estimated images and real images for the 43 scales. Results showed that for 33 scales, the $\boldsymbol{R}$ values were in the range of 0.8 to 0.9, while for the other 10 scales, the $\boldsymbol{R}$ values were 0.9 or higher. Therefore, we considered our model to be sufficient for estimating onomatopoeia images to be evaluated by humans.

Using equation (1), our system quantifies the onomatopoeia mental images created by consumers for the 43 scales. Thus, we obtain the

vectors for onomatopoeia texture images, that is, the 43 scale values for each image as quantified by the system.

4 Extracting Texture Feelings of Products

In our study, we extracted the design and texture images from actual product pictures and their text descriptions and reviews. To collect pictures to be regarded as products, we used part of the Flickr Material Database (FMD), as described by [14]. The FMD, which was created explicitly for the study of material recognition, consists of 100 pictures of 10 different materials (for a total of 1000 pictures) collected from pictures available on [15]. From these, we chose the five materials of metal, glass, plastic, stone, and wood, for a total of 500 pictures. We also collected 500 text excerpts from the descriptions of the 500 pictures (which we regarded as products in our study). We asked 60 participants to write their impressions of each of the materials in the pictures.

4.1 Method 1: Extracting SIFT features of pictures

The 500 FMD pictures were divided into two sets of 250 pictures each, one set to be used for training and the other for testing. We used the resulting training data for mapping the relationships between the pictures and adjectives. For each set of 250 pictures for training, we manually created sets of related information by the following procedure: (1) two pictures were randomly selected from the set and shown to a participant [16], (2) the participant was asked to answer which picture more represents each adjective. The participant is also allowed to choose "impossible to rank" when the two pictures are considered to be equal in value or when the specified adjective seems unrelated to the pictures.

Using this training data, we calculated the probability of co-occurrence between the adjective scales and the SIFT (Scale-Invariant Feature Transform is an algorithm in computer vision to detect and describe local features in images) feature value for each of the 250 remaining pictures based on the work of [17]. As a result, we obtained the mapping of the relationships between the pictures and adjectives.

Through the above process, we extracted the SIFT feature values for all 500 pictures as texture sensations. These data are regarded as the SIFT vectors for picture texture images, as shown in Figure 5.

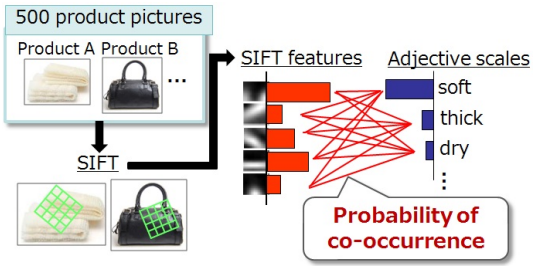

Figure 5. Method 1: Extracting SIFT picture features

4.2 Method 2: Extracting picture color images

In this method, we employed 45 color samples. The color samples were selected by 5 participants as the representative hues and tones among 130 colors, as described in Kobayashi [18].

To collect color image data, we conducted a psychological experiment. Participants were asked to evaluate the images of the 45 colors against 43 pairs of adjectives on 7-points SD scales, as shown in Figure 6. Color stimuli were displayed on a computer screen randomly for each participant. Twenty Japanese undergraduates aged 20 to 24 (10 males and 10 females) participated in the experiment.

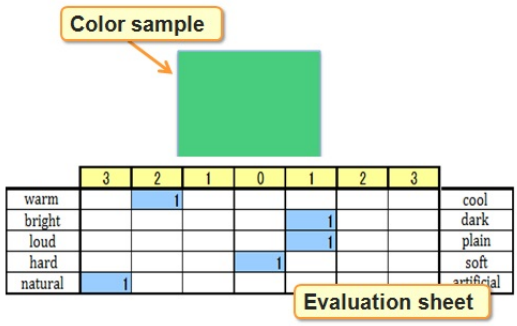

Figure 6. Example of experimental answer sheet

After obtaining the evaluations (i.e., 45 colors $\times$ 43 scales $\times$ 20 participants), we calculated the average values and collected the color image data (45 colors $\times$ 43 scales) as vectors.

We calculated the similarity between the obtained vectors for onomatopoeia texture images

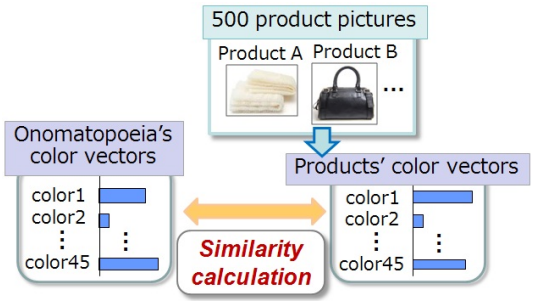

Figure 7. Method 2: Extracting color vectors of pictures

and color image vectors for each color. Furthermore, for the 500 pictures, we calculated the color frequency in each pixel, and each picture was assigned 45 color feature values. Using the process above, we obtained the color feature values as the picture color vectors for the pictures, as shown in Figure 7

4.3 Method 3: Extracting texture sensations from text data

As noted previously, we collected 500 text excerpts describing the products and then asked 60 subjects to write down their subjective impressions of the materials shown in the pictures freely. We assumed the text data to represent product reviews, and then used the collected data in the following analysis.

We extracted the adjectives included in the 43 adjective scales from the collected text data. We used the concept dictionary introduced in [19] and an online thesaurus [20] to extract the relevant adjectives.

We calculated the frequency of the adjectives appearing in the 500 text excerpts. We calculated the importance of the adjective scale results for each product picture using TF-IDF (Term Frequency Inverse Document Frequency is a numerical statistic which reflects how important a word is to a document in a collection or corpus). We obtained the importance of the adjectives in relation to the pictures as adjective vectors for picture texture perceptions.

4.4 Calculating the texture similarity between products and onomatopoeia

For each method mentioned above, product pictures were ranked according to their degree of matching with respect to the onomatopoeia by calculating the cosine similarity between the vectors of onomatopoeia texture perceptions and the vectors of picture texture perceptions.

The vectors of onomatopoeia texture perceptions, calculated by equation (1), comprised 43 values evaluated based on each adjective scale. Here, if a texture image vector of an onomatopoeia f is $\boldsymbol{k}(f)$, and those of a product picture t, which is recommended by each method, is $\boldsymbol{k}(t)$, then the cosine similarity between the two vectors ($\cos(\boldsymbol{k}(f), \boldsymbol{k}(t))$) is given by following equation (2). For all product pictures, we calculated the onomatopoeia cosine values, and recommended the product picture having higher cosine values than the other pictures.

$$\cos(\boldsymbol{k}(f), \boldsymbol{k}(t)) = \frac{\boldsymbol{k}(f) \cdot \boldsymbol{k}(t)}{|\boldsymbol{k}(f)||\boldsymbol{k}(t)|} \qquad (2)$$

We implemented a product recommendation system based on each method as a Java program. Figure 8 shows examples of recommendation results based on each of the three methods.

Figure 8. Recommendation results using each method

5 Evaluation

To verify the efficiency of our system, we used 60 pictures of the 500 product pictures for evaluation, with each picture having one text excerpt obtained in Method 3. Ten subjects were asked to provide responses for onomatopoeia expressing texture perceptions of the materials in each picture. For this task, (1) we chose onomatopoeia corresponding to 10 or more pictures, and also (2) selected onomatopoeia answers provided by at least 3 subjects. Those that satisfied both of these conditions included the following six onomatopoeia: *tsuru-tsuru*, *zara-zara*, *teka-teka*, *sube-sube*, *pika-pika*, *gotsu-gotsu*. These

Table 3. Precision $\boldsymbol{P}$ and reproducibility $\boldsymbol{R}$ in top $\boldsymbol{N}$ search results

Top N search results	**Method 1 SIFT features**		**Method 2 Color images**		**Method 3 Adjectives**		**All Methods 1 + 2 + 3**	
	$\boldsymbol{P}$	$\boldsymbol{R}$	$\boldsymbol{P}$	$\boldsymbol{R}$	$\boldsymbol{P}$	$\boldsymbol{R}$	$\boldsymbol{P}$	$\boldsymbol{R}$
5	0.73^{*1}	0.18^{*2}	0.37	0.08	0.63	0.17	0.63	0.15
10	0.63	0.3	0.37	0.17	0.63	0.34^{*2}	0.67^{*1}	0.33
15	0.64	0.48	0.38	0.26	0.58	0.46	0.66^{*1}	0.5^{*2}
20	0.6	0.6	0.36	0.34	0.58	0.6	0.63^{*1}	0.62^{*2}
25	0.54	0.66	0.36	0.42	0.56	0.72^{*2}	0.57^{*1}	0.72^{*2}
30	0.51	0.75	0.35	0.5	0.52	0.8	0.54^{*1}	0.82^{*2}

were thus categorized as the most accurate onomatopoeia to describe the pictures.

Using the degree of precision (percentage of correct data in the top $\boldsymbol{N}$ search results) and reproducibility (percentage of correct data in the top $\boldsymbol{N}$ search results in among correct data), we compared the performance of each method. In cases of the integration of multiple methods, we calculated the average similarities across all methods. We calculated the precision and reproducibility values for each of the six onomatopoeia, and then obtained the average precision $\boldsymbol{P}$ and average reproducibility $\boldsymbol{R}$ for each, as shown in Table 3. In the table, "*1" means the highest precision value, and "*2" does the highest reproducibility value for the top $\boldsymbol{N}$ search results.

As Table 3 shows, when we integrated all three methods, we obtained the highest values for both precision and reproducibility. We thus found that our system shows the best performance using SIFT features (for product pictures), colors (for product pictures) and adjectives (for text data). When integrating all three methods, precision reached 66 %, and reproducibility reached 82 % in more than top 15 search results. This result shows that our system succeeded in recommending products which cognitively fit consumers' texture perceptions expressed by onomatopoeia.

6 Conclusion

In this study we proposed a system for recommending products which match texture perceptions expressed by onomatopoeia based on sound symbolism. Our system quantitatively estimates the texture perceptions of input onomatopoeic expressions by users, and then calculates the similarities between the texture perceptions expressed by those onomatopoeia and those of product pictures. Finally, our system suggests products to the user which match the perception or sensation expressed by the onomatopoeia.

The evaluation of our system shows that when integrating all three of the proposed methods, the system shows the best performance. Moreover, we were able to verify the validity of our system.

In our system, we chose the five materials of metal, glass, plastic, stone, and wood as search categories. By expanding these product material categories in future work, we expect that our system will correspond to more and more materials required by consumers as they perform online search activities.

To these ends, in future work we aim to examine the practical uses of our system in actual web search situations.

Acknowledgment

This study was supported by Grant-in-Aid for Scientific Research on Innovative Areas "Shitsukan" (No. 23135007) from MEXT, Japan.

References

[1] E. Yano, Y. Kitano, E. Sueyoshi, I. Shinohara, P. Sineenard, and T. Kato. Development of recommendation system with anonymous kansei model. *Information Processing Society of Japan: Database*, 44(SIG8(TOD18)):46–54, 2003.

[2] Y. Murakami, S. Nakamura, and S.Hashimoto. An article search system that learns the kansei of the user. *Proceedings of the 72th National Convention of the Information Processing Society of Japan*, 2:451–452, 2010.

[3] W. Köhler. *Gestalt Psychology*. Liveright Pub., New York, 1929.

[4] E. Sapir. A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3):225–239, 1929.

[5] V. S. Ramachandran and E. M. Hubbard. Synaesthesia a window into perception. *Thought and Language Journal of Consciousness Studies*, 8(12):3–34, 2001.

[6] J. Watanabe, T. Hayakawa, S. Matsui, A. Kano, Y. Shimizu, and M. Sakamoto. Visualization of tactile material relationships using sound symbolic words. *P. Isokoski and J. Springare (Eds.): EuroHaptics 2012*, LNCS 7283(Part II):175–180, 2012.

[7] M. Sakamoto and J. Watanabe. Effectiveness of onomatopoeia representing quality of tactile texture - a comparative study with adjectives. *Proceedings of the The 13th Annual Conference of the Japanese Cognitive Linguistics Association*, pages 242–245, 2012.

[8] S. Hamano. *The Sound-symbolic System of Japanese*. University of Florida, Gainesville, 1986.

[9] N. Fujisawa, F. Obata, M. Takada, and S. Iwamiya. Impression of auditory imagery associated with japanese 2-mora onomatopoeic representation. *Journal of the Acoustical Society of Japan*, 62(11):774–783, 2006.

[10] Y. Shimizu and M. Sakamoto. Creation support system for japanese onomatopoeia based on sound symbolism. *Proceedings of the The 26th Annual Conference of the Japanese Society for Artificial Intelligence*, 2N1-OS-8c-4:1–3, 2012.

[11] S. Iiba, Y. Shimizu, and M. Sakamoto. A method to select colors appropriate for tactile onomatopoeia based on sound symbolism. *Proceedings of the International Workshop on Modern Science and Technology (IWMST2012)*, pages 127–131, 2012.

[12] L. T. Maloney and H. Brainard. Perception of color and material in complex scenes. *Journal of Vision*, 10(9):1, 2010.

[13] H. Kakehi, I. Tamori, and L. C. Schourup. *Dictionary of Iconic Expressions in Japanese*. Mouton de Gruyter, Berlin, New York, 1996.

[14] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. *In Proc. CVPR*, 2010.

[15] Flickr material database. `http://people.csail.mit.edu/celiu/CVPR2010/FMD/`.

[16] Amazon mechanical turk. `https://www.mturk.com/mturk/welcome`.

[17] T. Abe, T. Okatani, and K. Deguchi. Recognizing surface qualities from natural images based on learning to rank. *21st International Conference on Pattern Recognition (ICPR 2012)*, pages 3712–3715, 2012.

[18] S. Kobayashi. *Color Image Scale*. Kodansya, Tokyo, 2001.

[19] T. Yamaguchi. *Japanese Thesaurus - Categorized word dictionary-*. Taishukan, Tokyo, 2003.

[20] Wordnet. `http://nlpwww.nict.go.jp/wn-ja/index.ja.html`.