Tactile Experience Is Evoked by Visual Image of Materials: Evidence from Onomatopoeia

Maki Sakamoto (sakamoto@inf.uec.ac.jp)
Department of Informatics, The University of Electro-Communications
1-5-1, Chofugaoka, Chofushi, Tokyo 182-8585, Japan

Tatsuki Kagitani (k1330014@edu.hc.uec.ac.jp)
Department of Informatics, The University of Electro-Communications
1-5-1, Chofugaoka, Chofushi, Tokyo 182-8585, Japan

Ryuichi Doizaki (r.doizaki@uec.ac.jp)
Department of Informatics, The University of Electro-Communications
1-5-1, Chofugaoka, Chofushi, Tokyo 182-8585, Japan

Abstract

Human beings get a lot of information from a picture based on what we see and our background knowledge. However, many computer vision researches are heavily dependent on the use of image features and have paid little attention to background knowledge we use in texture processing. The present study explores the degree to which onomatopoeia evoked by visual images is affected by the multimodal experience-based knowledge such as tactile experience. In Experiment 1 participants saw original complete images of Flickr Material Database (FMD) and answered onomatopoeia for expressing their textures and in Experiment 2 participants saw cut out images and answered onomatopoeia for expressing their textures. We obtained 17487 onomatopoeic words (1827 types) from experiment 1 and 30138 onomatopoeic words (2442 types) from experiment 2. We counted the number of types of onomatopoeia evoked by each image. Result showed that original image evoked significantly more variety of onomatopoeia than cut-off image. This result suggests that human texture evaluations based on the original complete images of FMD are affected more easily by experience-based knowledge about the material. Furthermore, we showed that image whose material category is relatively easy to recognize evokes significantly frequently tactile onomatopoeia than image whose material category is hard to recognize.

Keywords: Visual image; Texture; Tactile experience; Onomatopoeia

Introduction

We human beings get a lot of information from a picture based on what we see and our background knowledge. However, many researches are dedicated to letting the computer extract efficient and effective visual features and building models from them rather than human background knowledge. The most common visual features include color, texture and shape, etc. and many image annotation and retrieval systems have been constructed based on these features (As a review for the recent researches of computer vision, see Tian, 2013). These researches are heavily dependent on the use of image features and have paid little attention to background knowledge we use in texture processing.

In recent years, there has been a growing interest in material perception that requires multimodal information (Bergmann & Kappers, 2007; Gaissert & Wallraven, 2012). Interplay between visual and tactile senses is sometimes required for precise material perception (Baumgartner et al., 2013). Bergman and Kappers (2007) reports that participants were able to retrieve similar information through visual and tactile modalities for surface roughness. As for visual and tactile representation of material properties, some studies point out material properties obtained with the tactile modality (i.e., hardness and roughness) are crucial (Klatzky & Lederman, 1987). On the other hand, visual dominance in assessing building materials was reported (Wastiels et al. 2013). Abe et al. (2012) also argues that humans appear to be able to fairly accurately sense the surface quality only from visual inputs, although it is the most closely related to tactile sensations, and proposes a method for estimating the quantitative values of some attributes associated with surface qualities of an object, such as glossiness and transparence, from its image. Baumgartner et al. (2013) overviews that, whereas a great deal of research has been conducted to investigate both uni- and bimodal shape perception, the perception of materials and material qualities has only recently received more attention in vision research (Fleming et al., 2003; Motoyoshi, 2010; Motoyoshi et al., 2007). Furthermore, very little attention has been paid to the multimodal experience-based knowledge which might affect the perception of materials and material qualities in vision research. The present study explores whether human texture evaluations based on visual images are affected by the multimodal experience-based knowledge.

Humans categorize sensory inputs using words, and words are an important index in investigating what they perceived sensory inputs. Japanese language has a word class called "onomatopoeia" or "mimetics" that has been used to express vivid sensations in everyday life (e.g., "sara-sara" represents a dry and smooth sensation, and "zara-zara" represents a dry and rough sensation). Japanese onomatopoeia has a strong and systematic association with sensations (Hamano, 1998), although the existence of sound symbolic words has been demonstrated in many languages of the world (e.g., Jespersen, 1922; Köhler, 1929; Newman, 1933; Sapir, 1929; Taylor, 1963; Werner & Wapner, 1952; Wertheimer, 1958, for early studies) and, to varying extent in a wide variety of languages (e.g., Brown, Black, & Horowitz, 1955; Davis, 1961; Emeneau, 1969; Hinton, Nichols, & Ohala, 1994; Klank, Huang, & Johnson, 1971; Nuckrolls, 1999; Voeltz & Kilian-Hatz, 2001). Since Japanese, compared to other languages, has a large number of onomatopoeia for expressing texture sensations, the relationship between onomatopoeia and tactile or visual sensations has been studied. Watanabe et al. (2012) investigated tactile sensations by analyzing onomatopoeia used to expressing tactile sensations and subjective evaluations of comfort/discomfort for touched objects. Yoshino et al. (2013a, b) investigated the metal texture design by analyzing onomatopoeia used to express visual sensations. Through psychological experiments where participants were asked to look at a pair of imitation and real materials without touching them and answer sound symbolic words associated with them, Yoshino et al. (2013a) showed that real materials were significantly more easily associated with onomatopoeia than imitation materials. Yoshino et al. (2013a) also showed that non-experts tend to respond tactile onomatopoeic words significantly more frequently than by experts. This result suggests that although experts engaged in metal texture design focus on visual design of metal surface, non-experts, namely people in general, perceive material properties recalling experiences through touch. Since Japanese onomatopoeia can be used to express tactile sensation as well as visual sensations, the present study focuses on onomatopoeia to explore whether human texture evaluations based on visual images are affected by the multimodal experience-based knowledge such as tactile experience.

Method

In this study we use Flickr Material Database (FMD) (http://people.csail.mit.edu/celiu/CVPR2010/FMD/) (Sharan et al. 2014), that is one of major dataset frequently used in vision researches. The FMD consists of color photographs of surfaces belonging to one of ten common material categories: fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood. There are 100 images in each category, 50 close-ups and 50 regular views. Each image contains surfaces belonging to a single material category in the foreground and was selected manually from approximately 50 candidates to ensure a variety of illumination conditions, compositions, colors, textures, surface shapes, material sub-types, and object associations. Since FMD was constructed with the specific goal of capturing the natural range of material appearances, it is clear, for example, that surfaces belong to the specific material category and not any of the others. We hypothesize, therefore, that human texture evaluations based on visual images of FMD, which are used in previous vision researches, are affected by the multimodal experience-based knowledge and evoke onomatopoeia related not only to vision but also tactile sensation. We hypothesize that the onomatopoeic words evoked by original FMD images are different from those evoked by a visual image cut out from original complete image because the influence of the multimodal experience-based knowledge is expected to be reduced in the cut out image.

For the purpose described in the above, we performed two psychophysical experiments. In Experiment 1 participants saw complete images of FMD and answered onomatopoeia for expressing their textures and in Experiment 2 participants saw cut-off images and answered onomatopoeia for expressing their textures. Experiment 1 and 2 were approved by the Ethics Committee of the University of Electro-Communications, Tokyo, Japan.

Experiment 1

Participants 100 people participated (25 women and 75 men, mean age 22.08). The participants were not informed of the purpose of the experiment, and they had no known abnormalities in speech or in vision. They visited a laboratory at the University of Electro-Communications. All participants provided written informed consent prior to the experiments. Documents about the experimental procedures and written informed consents were presented to the ethics committee.

Materials 1000 FMD images were classified into 10 material groups. Each group of materials consisted of 100 samples (10 materials were selected from each 10 material categories of FMD. 100 participants were classified into 10 groups. As a result, 100 materials were presented to each participant.

Procedure The experiment was conducted in an isolated test room at the university under controlled lighting conditions. Participants were kept at a distance of about 50 cm from the touch panel display showing the materials. The materials were presented vertically at eye-height by the slideshow function of Microsoft office powerpoint 2010 in a random order.

During the test, participants were asked to answer one to six onomatopoeic words expressing the texture of each material. At the same time, they were asked to circle the part of the image material shown on the touch panel display which they focused on to express the texture of the material. They were allowed to mark as many as they like. An answer sample is shown in Figure 1. The onomatopoeia input into the left cell is ‘zara-zara,’ which means dry and rough texture. The onomatopoeia input in the middle cell is ‘gotsu-gotsu’, which means stiff and harsh texture.

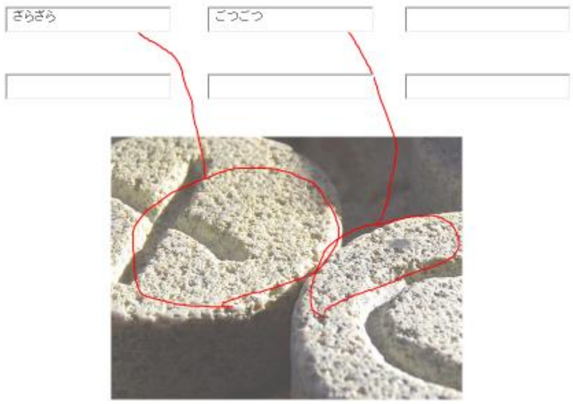

Figure 1: An answer sample of an image material

Another answer sample is shown in Figure 2. The onomatopoeia input into the left cell is ‘mosa-mosa,’ which means hairy and thickly texture. The onomatopoeia input in the middle cell is ‘husa-husa’, which means bushy and thick texture.

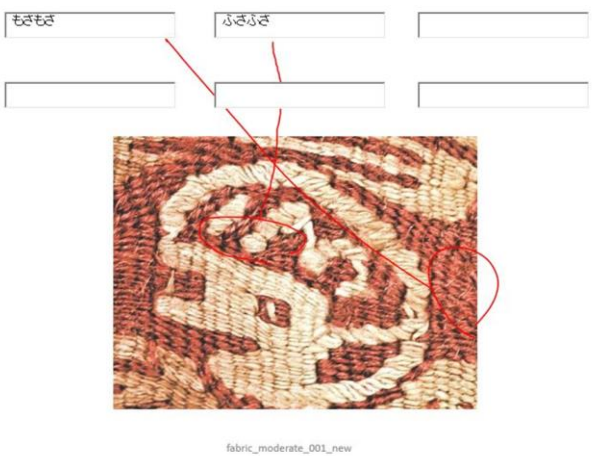

Figure 2: Another answer sample of an image material

Experiment 2

Participants 100 people participated (25 women and 75 men, mean age 20.59). The participants had not participated for experiment 1 and were not informed of the purpose of the experiment. They had no known abnormalities in speech or in vision. They visited a laboratory at the University of Electro-Communications. All participants provided written informed consent prior to the experiments. Documents about the experimental procedures and written informed consents were presented to the ethics committee.

Materials We cut out the part that more than 3 participants marked from the complete image material used in experiment 1 as exemplified in Figure 2. Since the average size of marked part was about a few 100 pixels, we cut the square image in 150 x 150 pixels. 1946 image samples were classified into 10 groups. Since we cut out one to three parts from the complete image material, each group of image materials consisted of 160 to 200 samples. 100 participants were classified into 10 groups. As a result, 160 to 200 image materials were presented to each participant. In Figure 3, the left is the cut-out image of the complete material shown in Figure 1 and the right is the cut-out image of the complete material shown in Figure 2.

Figure 3: examples of cut-out image

Procedure The experiment was conducted in an isolated test room at the university under controlled lighting conditions. Participants were kept at a distance of about 50 cm from the touch panel display showing the materials. The materials were presented vertically at eye-height by the slideshow function of Microsoft office powerpoint 2010 in a random order.

During the test, participants were asked to answer one to six onomatopoeic words expressing the texture of each material. The answer sample is shown in Figure 4. The left onomatopoeia input into the left cell is ‘gowa-gowa,’ which means coarse and stiff texture. The onomatopoeia input in the middle cell is ‘zara-zara’, which means stiff and harsh texture.

Figure 4: An answer sample of an image material

Results and Discussion

We obtained 17487 onomatopoeic word tokens (1827 onomatopoeic word types) from experiment 1 and 30138 onomatopoeic word tokens (2442 onomatopoeic word types) from experiment 2. By analyzing the data obtained from the two experiments, we testified our hypothesis that the onomatopoeic words are different between those evoked by a visual image cut out from a whole and those evoked by its original complete image because the influence of the multimodal experience-based knowledge is expected to be reduced in the cut-off image. We also testify the hypothesis that human texture evaluations based on the original complete images are affected by the multimodal experience-based knowledge and evoke onomatopoeia related not only to vision but also tactile sensation.

Difference between onomatopoeia evoked by an original image and that evoked by a cut-off image

We compared the type of onomatopoeia associated with original images and that associated with image cut out from the original images. Table 1 shows onomatopoeia associated with original image No. 1, which is given in Figure 2 and onomatopoeia associated with image cut off from the original image No. 1, which is given in Figure 3.

Table 1: Onomatopoeia associated with original image No. 1 and onomatopoeia associated with image cut off from it.

Onomatopoeia from original image No.1	Onomatopoeia from cut-off image of No.1
Gasa-gasa	Gowa-gowa
Gishi-gishi	Mohu-mohu
Gowa-gowa	Mosa-mosa
Gyuu-gyuu	Tiku-tiku
Husa-husa	Zara-zara
Huwa-huwa	Zowa-zowa
Keba-keba	
Mishi-mishi	
Moko-moko	
Mosa-mosa	
Tiku-tiku	
Toge-toge	
Zara-zara	

In experiment 1, participants circled the part of the image material which they focused on to express the texture of the material and in experiment 2, participants answered onomatopoeia expressing the texture of the part of the material cut off from original image. It means that participants answered onomatopoeia expressing the texture of the same part of material. However, Table 1 shows that onomatopoeic words evoked by original image are different from those evoked by cut-off image. It indicates that onomatopoeia answered by participants has the more variety for original image than for cut-off image. We counted the number of types of onomatopoeia evoked by each image. Result showed that the average number of types of onomatopoeia evoked by original image was 11.60, while the average number of types of onomatopoeia evoked by cut-off image was 10.84. T-test (two-tailed, the alpha level .05) showed that original image evoked significantly more variety of onomatopoeia than cut-off image. This result suggests that human texture evaluations based on the original complete images of FMD are affected more easily by experience-based knowledge about the material. As we see from Figure 5 showing Fabric samples of FMD, we can recognize material category of images. The original FMD image, which is easy to recognize material category such as fabric, glass, paper, and so on, evoked more variety of onomatopoeic words than the cut-off image, which shows only a part of material, its texture and is hard to recognize material category. We believe that this is because participants unconsciously used the knowledge about the material and remembered multi-modal experiences about the material. Practically, onomatopoeia evoked from original image listed in Table 1 includes those based on multimodal experience. For example, 'Gasa-gasa' is used to describe a dry and rough skin (tactile sensation) or a rustling sound (auditory sensation). 'Gishi-gishi' is used to describe a dry and hard uncomfortable skin (tactile sensation) or creaking sound (auditory sensation). '

Tactile experiences are evoked by visual image

Although very little attention has been paid to the multimodal experience-based knowledge which might affect the perception of materials and material qualities in vision research, some researches on visual perception of texture mention the importance of memories about textiles derived from experiences of viewing, touching, and wearing them. Lee and Sato (2001), for example, investigated the mechanism of texture perception and pointed out that past visual and tactile experiences and memories influenced significantly texture perception although participants used only the sense of sight to evaluate the surface of a textile. We, therefore, focus on influence of the tactile experience in the evaluation of visual image texture. We hypothesized that the original FMD image, which was easy to recognize material category such as fabric, glass, paper, and so on, would evoke tactile onomatopoeia than the cut-off image, where material category was hard to recognize.

We analyzed how much tactile onomatopoeia was evoked from the complete image of material compared to cut-off image showing only the texture of material. Since, as described in Introduction, Japanese onomatopoeic words

Figure 5: Fabric samples of FMD (Sharan et al., 2014)

Table 2: list of tactile onomatopoeia

Onomatopoeia	Explanations (examples)	Onomatopoeia	Explanations(examples)
sara-sara	dry and smooth	pasa-pasa	dry and powdery
tsuru-tsuru	slippery and smooth	huni-huni	soft and limp
sube-sube	smooth, silky, velvet hand	puri-puri	springy and soft
huwa-huwa	soft, light, and fluffy	kishi-kishi	creak ex. hair creaks.
zara-zara	texture of coarse paper	husa-husa	bushy and rich ex. feather
gowa-gowa	coarse and stiff sheets	chiku-chiku	ex. the undershirt scratches
gotsu-gotsu	rugged and scraggy rock	mohu-mohu	fluffy, warm ex. blanket
mochi-mochi	skin like a rich cake	howa-howa	fluff of clouds
poko-poko	texture like bubbling water	puru-puru	soft and elastic
beta-beta	grisly and sticky	shari-shari	crunch crunch
moko-moko	lumpy and fluffy surface	peta-peta	pasty
huka-huka	soft and fluffy	gishi-gishi	strongly creaking
gasa-gasa	dry and rough skin	beto-beto	sticky and greasy
nuru-nuru	slimy	jori-jori	ex. mustache
suru-suru	smooth	nume-nume	smooth, slimy and shining
kasa-kasa	desiccated skin	tsubu-tsubu	dots on the surface
shaka-shaka	mixture of smooth and rough textures	zaku-zaku	crunch through the snow
gunya-gunya	limp and soft	shori-shori	crispy and light
puni-puni	squishy, but comfortable	sawa-sawa	rustling
kori-kori	crunchy	mosa-mosa	sluggish
butsu-butsu	pimples on the surface	hunya-hunya	soft, flaccid and weak
boko-boko	uneven and nubby		

are sound-symbolic, we made onomatopoeic words by combining all Japanese syllables (105 syllables). We created onomatopoeic words in two-syllables-repeated form (e.g., /saka-saka/, /saki-saki/, /saku-saku/, /sake-sake/, /sako-sako/, and /sakari-sakari/), and added some special phonemes used in Japanese onomatopoeic words. Finally we got 14,584 onomatopoeic words. From the considerably large number of words, three Japanese native speakers selected 307 words (including novel words) that can be acceptable as tactile onomatopoeic words. 307 onomatopoeic words combined with the word "touch" were tested by using Google search queries. As a result, top 43 search results given in Table 2 were selected as frequently used tactile onomatopoeia.

We counted the number of tactile onomatopoeia included in onomatopoeia evoked by complete FMD image and cut-off image showing only a part of material. The result is given in Table 3.

Table 3: Number of tactile onomatopoeia obtained from experiment 1 and 2

Onomatopoeia	Experiment 1	Experiment 2	Sum
Tactile onomatopoeia	8125	13310	21435
Others	9362	16828	26190
Sum	17487	30138	47625

The result of Chi-square tests showed that image that is relatively easy to recognize material category evokes significantly frequently tactile onomatopoeia than image that is hard to recognize material category, $\chi2(1)=3.28$, $p<.01$. This result suggests that past tactile experiences and memories are thought to significantly influence texture perception when participants evaluate the texture of material whose category is easy to recognize despite using only the sense of sight.

Conclusion

Many researches have been dedicated to letting the computer extract efficient and effective visual features and building models from them rather than human background knowledge. FMD is one of the most frequently used dataset by such researches. Our study, however, showed that past tactile experiences and memories significantly influence texture perception when participants evaluate the texture of material appearing in original FMD image. This finding suggests that future vision research using material image should consider influences of past tactile experiences on texture perception.

Acknowledgement

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas "Shitsukan" (No. 23135510 and 25135713) from MEXT, Japan.

References

Abe, T., Okatani, T., & Deguchi, K. (2012). Recognizing Surface Qualities from Natural Images Based on Learning to Rank, *Proceedings of International Conference on Pattern Recognition (ICPR)*, 3712-3715.

Baumgartner, E., Wiebel, C. B., & Gegenfurtner, K. R. (2013). Visual and Haptic Representations of Material Properties, *Multisensory Research*, 26, 429-455.

Bergmann Tiest, W.M. & Kappers, A. M. L. (2007). Haptic and Visual Perception of Roughness, *Acta Psychologica*, 124, 177-189.

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The Journal of Abnormal and Social Psychology*, 50, 388-393.

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology*, 52, 259-268.

Emeneau, M. B. (1969). Onomatopoetics in the Indian linguistic area. *Language*, 45, 274-299.

Fleming, R. W., Dror, R. O., Adelson, E. H. (2003) Real-world Illumination and the Perception of Surface Reflectance Properties, *Journal of Vision*, 3(5), 3.

Gaissert, N. & Wallraven, C. (2012). Categorizing Natural Objects: A Comparison of the Visual and the Haptic Modalities, *Experimental Brain Research*, 216, 123-134.

Hamano, S. (1998). *The sound symbolic system of Japanese*. Stanford, CA: CSLI Publications; Tokyo: Kuroshio.

Hinton, L., Nichols, J., & Ohala, J. (Eds.). (1994). *Sound symbolism*. Cambridge, UK: Cambridge University Press.

Jespersen, O. (1922). The symbolic value of the vowel i. *Philologica*, 1, 1-19.

Klank, L. J. K., Huang, Y. H., & Johnson, R. C. (1971). Determinants of success in matching word pairs in tests of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, 10, 140–148.

Klatzky, R. L. & Lederman, S. (1987). There's More to Touch than Meets the Eye: The Salience of Object Attributes for Haptics with and without Vision, *Journal of Experimental Psychology*, 116(4), 356-369.

Köhler, W. (1929) *Gestalt Psychology*. NewYork: Liveright Publishing Corporation.

Lee, W. & Sato, M. (2001). Visual Perception of Texture of Textiles, Color Research and Application, 26(6), 469-477.

Motoyoshi, I. (2010). Highlight-shading Relationship as a Cue for the Perception of Translucent and Transparent Materials, Journal of Vision, 10, 6.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image Statistics and the Perception of Surface Qualities, *Nature*, 447, 206-209.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology*, 45, 53-75.

Nuckrolls, J. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 28, 225-252.

Sapir, E. (1929). A study of phonetic symbolism. *Journal of Experimental Psychology*. 12, 225-239.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Material Perception: What Can You See in a Brief Gance?, *Journal of Vision*, 14(9), 12.

Tian, D. P. & Shaanxi, B. (2013). A Review of Image Feature Extraction and Representation Techniques, *International Journal of Multimedia and Ubiquitous Engineering*, 8(4), 385-396.

Voeltz, F. K. E., & Kilian-Hatz, C. (Eds.). (2001). *Ideophones*. Amsterdam: John Benjamins.

Wastiels L., Schifferstein, H. N. J., Wouters, I., & Heylighen, A. (2013). Touching Materials Visually: About the Dominance of Vision in Building Material Assessment, *International Journal of Design*, 7(2), 31-396.

Watanabe, J. & Sakamoto, M. (2012). Sound Symbolic Relationship between Onomatopoeia and Emotional Evaluations in Taste. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society(CogSci2012)*, 2517-2522.

Yoshino, J., Yakata, A., Shimizu, Y., Haginoya, M, & Sakamoto, M. (2013a). Method of Evaluating Metal Textures by the Sound Symbolism of Onomatopoeia, *Proceedings of the 2nd Asian Conference on Information Systems (ACIS 2013)*, 618-624.

Yoshino, J., Yakata, A., Shimizu, Y., Haginoya, M, & Sakamoto, M. (2013b). Sound Symbolic Words Are More Easily Associated with Real Metal Than Imitation, *Proceedings of the 5th International Congress of International Association of Societies of Design Research (IASDR 2013),* 1471-1477.