Holistic Processing Affects Surface Texture Perception: Approach from Japanese Sound Symbolic Words

Jinhwan Kwon1, Tatsuki Kagitani1, Maki Sakamoto1

1 Department of Informatics,
The University of Electro-Communications, Tokyo, Japan
E-mail: sakamoto@inf.uec.ac.jp

The human visual system is able to perceive not only the macrostructure (form and shape) of a surface, but also its microstructure (texture). Some evidence suggests that microstructural characteristics are processed independently of macrostructural features. However, the human visual system can interpret a variety of information about the physical world, enabling the recognition and semantic categorization of complex visual scenes at a glance. This remarkable perceptual ability relies heavily on holistic processing, which is achieved by estimating the global statistical summary of an image. On the other hand, texture is an important source of information for distinguishing between artificial and naturally occurring surfaces in images. In addition, it is reported that Japanese sound symbolic words are useful to express fine differences in texture and synesthetic characteristics. However, there is no evidence comparing the characteristics of surface texture perception between whole- and part-based images using sound symbolic words. The objective of the present study was to examine whether sound symbolic words for describing the surface texture perception differs between whole-based images related to the holistic processing and part-based images. In Experiment 1, we examined the effect of whole-based images in surface texture perception using sound symbolic words. In Experiment 2, we examined the effect of part-based images in surface texture perception using sound symbolic words. The results revealed that the sensory and symbolic descriptors differed in texture perceptions between whole-based and part-based image processing. These findings suggest that sound symbolic words can describe differences in surface texture between whole-based and part-based images at a fine resolution.

***Journal of Cognitive Science* 18-3: 321-340, 2017**
Date submitted:04/11/17 Date reviewed:08/22/17
Date confirmed:09/11/17

Introduction

Vision allows us to interpret a wide variety of information about the physical world. Although visual scenes are often complex, we are typically able to recognize and understand the meaning of a scene at a glance, even with exposure times as short as 20 ms [11, 12]. Furthermore, humans have the ability to rapidly semantically categorize the meaning of an image [13]. This remarkable perceptual ability is heavily dependent on holistic processing [14]. A number of psychophysical studies have reported that the perception of a scene is achieved by the processing of global image features, estimated by global statistical summary of the image [15-17]. Although Texture is determined by the microstructure of surfaces, in contrast to macrostructural form and shape information, texture is an important factor in the characterization of global image features [14, 18-21]. The current study focused on differences in texture perception between whole- and part-based visual processes.

Texture is an important source of information for distinguishing between artificial and naturally occurring surfaces in images [1]. Texture typically refers to softness, smoothness, slipperiness and other qualities of a surface, and is an important property in the field of haptics [2-5]. Humans are able to perceive various textural properties of surfaces from visual information such as colors, dots, lines, edges and spatial density, and the role of vision in texture perception has been the topic of substantial research [6, 7]. For example, Lederman et al. (1986) reported that when participants were asked to judge the spatial density of a surface texture, visual inputs were weighted more heavily than tactile inputs [8]. In addition, visual texture has been implicated in the perception of visual complexity, emotional content and aesthetics [9, 10].

In recent years, there has been a growing research interest in the relationship between sound symbolism and perceptual matching [21-30]. Sound symbolism is defined as a property of certain words that have a direct link between sound (phonological form) and perceptual (or semantic) meaning [31-38]. For example, Köhler (1929) reported a relationship between non-words and object shapes, revealing that participants preferred to match some nonsense words (e.g., “maluma”) with curvy rounded shapes and other (e.g., “takete”) to spiky angular shapes [39]. Recent research suggests that this process, referred to as the “bouba/kiki effect”, operates in a similar way to the correspondence between sound symbolism and visual perception [24, 25]. In particular, several studies have specifically examined crossmodal correspondence and sound symbolism in the Japanese language [40-42]. Japanese sound symbolic words typically have a strong and systematic association with sensations [43], and commonly refer to the tactile or visual perception of surface texture. For example, “nuru-nuru” indicates sliminess, while “sara-sara” indicates dryness and smoothness, and “zara-zara” indicates dryness and roughness. In this study, we focused on the Japanese sound symbolic words to confirm the differences in texture perception between holistic and part-based visual processes.

In addition, Doizaki et al. (2017) proposed a method for estimating the fine impression of sound symbolic words [68]. Specifically, this system can evaluate sound symbolic words as quantitative adjectives by calculating subjective impressions of sound symbolic words on the basis of the impressions evoked by each phoneme from a quantitative rating database. This method for quantifying qualitative data uses quantification theory class I (a type of multiple regression analysis), calculating the degree to which each phoneme contributes to each rating scale. Furthermore, the estimated ratings of sound symbolic words enable us to visualize a tactile perceptual space. The system is based on a database of sound-meaning association that can convert a sound symbolic word expressing tactile sensations into multidimensional ratings of adjective words. That is, the system can calculate evaluations in terms of 26 pairs of fundamental texture scales, such as roughness, hardness, and warmth. Therefore, we used the multidimensional rating system to analyze Japanese sound symbolic words.

The objective of the present study was to examine whether sound symbolic words for describing the surface texture perception differs between whole-based images related to the holistic processing and part-based images. In Experiment 1, we examined responses when participants were instructed to use sound symbolic words to describe the surface texture of whole images of various materials. In Experiment 2, we examined the use of sound symbolic words for describing the surface texture of part-based images produced by cropping sections of the images used in Experiment 1. Then, we analyzed the obtained sound symbolic words using the multidimensional rating system of sound symbolic word.

Materials and Methods

1. Task Design

We performed two psychophysical experiments using images from the Flickr Material Database (FMD) as visual stimuli [51]. In Experiment 1, whole FMD images were presented, and participants answered spontaneously and freely using sound symbolic words to describe the surface textures shown in the images. We analyzed participants' responses by evaluating sound symbolic words as quantitative adjectives [46]. In Experiment 2, cropped sections of the images used in Experiment 1 were presented to another group of participants. Participants were instructed to describe the surface texture of each image.

2. Ethics Statement

The experimental protocol in this study was approved by the Ethics Committee of the University of Electro-Communications, Tokyo, Japan. Participants were recruited from the University of Electro-Communications using a poster. All subjects provided written informed consent prior to the experiments and were paid an allowance for their participation.

3. Participants

100 participants (25 women and 75 men, mean age = 22.1 years) participated in Experiment 1. 100 participants (25 women and 75 men, mean age = 20.6) took part in Experiment 2. In both experiments, participants were divided into 10 groups. Participants were not informed of the purpose of the experiment, and reported no known abnormalities in speech or in vision.

4. Apparatus and Stimuli

The experimental stimuli used in this study were obtained from the FMD (http://people.csail.mit.edu/celiu/CVPR2010/FMD/) (Sharan et al. 2014), which is one of the major stimulus sets used in vision research. 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. Each image contains surfaces that belong to a single material category in the foreground. A range of images was selected to provide a variety of illumination conditions, compositions, colors, textures, surface shapes, material sub-types, and object associations. Since the FMD was constructed with the specific goal of capturing the natural range of material appearances, the surfaces depicted in the images each belong to a specific material category, and not any of the others. We selected 10 images from each material category (fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood), and categorized them into 10 groups. As a result, 1,000 FMD images were classified into 10 groups of 100 images each. Figure 1 shows an example of the FMD image stimuli. Each group of visual stimuli were presented for each participant group. To produce the stimuli for Experiment 2, we conducted an experiment to mark the part of the visual stimulus participants focused on when describing the surface texture. Ten participants participated in this experiment, we cropped each image section that three or more participants marked. Figure 2 shows an example of cropped image stimuli and we used the cropped images in Experiment 2. Since the average size of the image sections marked by participants was approximately 100 pixels, we cropped square images of 150 × 150 pixels. Consequently, we obtained a total of 1,946 image samples, and classified them into 10 groups. Each group of visual stimuli was presented to each participant group.

Fig 1. An example of FMD images used in Experiment 1.

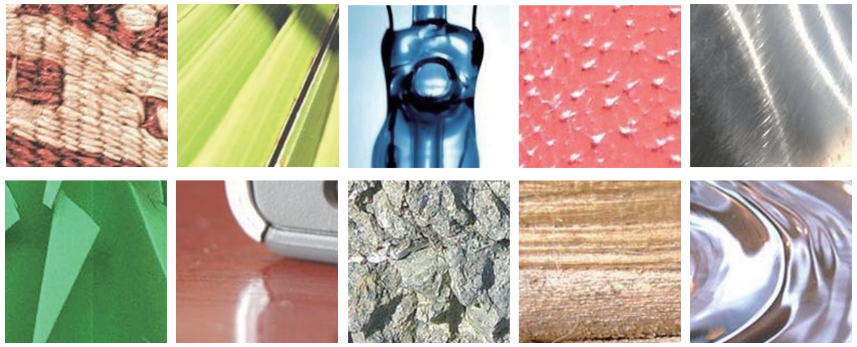

Fig 2. An example of cropped images used in Experiment 2.

5. Procedure

In Experiment 1, each trial was conducted in an isolated test room under controlled lighting conditions. Participants were kept at a viewing distance of approximately 50 cm from a touch panel display showing the visual stimuli. The visual stimuli were presented vertically at eye-height, in a random order, using the slideshow function of Microsoft Powerpoint 2010. During the test, participants were instructed to answer spontaneously and freely with 1–6 sound symbolic words expressing the texture of each material. At the same time, they were asked to circle the part of the visual stimuli they focused on while describing the surface texture. An example answer is shown in Figure 3. The sound symbolic word in the left cell is 'mosa-mosa,' which refers to a hairy and thick texture. The sound symbolic word in the middle cell is 'fusa-fusa', which refers to a bushy and thick texture.

Experiment 2 followed the same procedure as Experiment 1, except for the following changes. We cropped image stimuli determined by participants' responses to circle the part of the visual stimuli they focused on while describing the surface texture in Experiment 1 (see Figure 2 for example stimuli). We used the cropped image stimuli in Experiment 2. Each group of visual stimuli was presented to each participant group. Participants were instructed to answer spontaneously and freely with 1–6 sound symbolic words describing the texture of the material shown in each image. An example answer is shown in Figure 4. The sound symbolic word in the left cell is 'gowa-gowa,' which refers to a coarse and stiff texture. The sound symbolic word in the middle cell is 'zara-zara', which refers to a dry and rough texture.

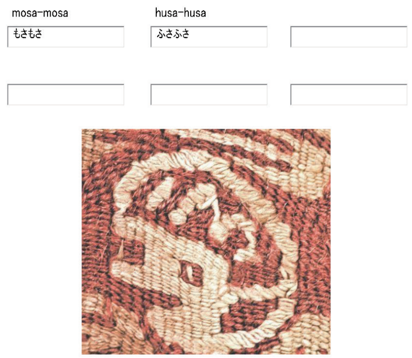

Fig 3. An example response to an image stimulus in Experiment 1.

Fig 4. An example response to an image stimulus in Experiment 2.

6. Data analysis

We have proposed a system that can convert a SSW in Japanese into quantitative ratings in multiple texture-based dimensions (26 pairs of adjectives) [19]. In the system, when a word that intuitively expresses a texture is input into the text field, information equivalent to evaluations against the 26 pairs of touch adjectives is obtained based on an analysis of the sounds of the word. To estimate the quantitative information of every possible SSWs, we built a database of sound symbolic associations for each phoneme with the 26 pairs of adjectives through psychological experiments. The system can be predicted by combining the evaluations of each phoneme in these words, which are potential advantages of our system [68]. However, the adjectives are not sufficient to describe the visual perception of the texture of objects, because visual texture is so strongly involved in visual complexity, emotions and aesthetic perception, as high-level perceptual attributes [9, 10]. In the current study, we used 37 pairs (total 43 pairs) of adjectives appropriate for evaluating the visually-perceived texture of objects (see Table 1) [52-54]. To expand the scope of the quantitative rating database of sound symbolic words, we used an impression-rating methodology (using the semantic differential (SD) method), in which we measured the relationship between phonological features and impression ratings to create a quantitative rating database. The participants were 78 native Japanese speakers aged 20–24 years (51 males and 27 females). The vision-rating scale included 37 pairs of adjectives appropriate for evaluating the perceived texture of objects (see Table 1). These scales were used in a psychological experiment and then applied to the method we developed, both of which are described below. Using this method, we calculated subjective impressions of sound-symbolic words on the basis of the impressions evoked by each phoneme. Thus, the experimental stimuli are required to include all varieties of Japanese phonemes, including basic phonemes (consonants /C/ and vowels /V/), and special phonemes (syllabic nasals /N/, choked sound and the assimilated sound with small “tsu” in Japanese /Q/, long vowels /R/, and adverbs ending in /Li/) [55]. We created various combinations of sounds to examine the effects of sound order, determining whether first-syllable and second-syllable phonemes evoked

different subjective impressions (e.g., the difference between “kasa” and “saka”). We combined all sounds in the Japanese syllabary (from /a/ to /n/) to create two-syllable expressions (i.e., /aa/, /ai/, . . . , /wan/, /nn/). We obtained a total of 11,075 words, including those containing repeated two-syllable sound-symbolic expressions (e.g., /aa-aa/, /ai-ai/). Moreover, we added 3,509 words including every type of special phoneme (e.g., /fuwari/ and /peQtari/). From these 14,584 words, we selected 312 words that were judged by three participants as sound-symbolic expressions for describing texture sensations. Importantly, the 312 selected word stimuli covered every possible phoneme type. This enabled the system to estimate the meaning of every possible type of onomatopoeia, because Japanese-speakers commonly create new onomatopoeic expressions by combining phonemes to express intuitive feelings. Participants were presented with sound-symbolic expressions that conveyed texture impressions as well as visual rating scales for evaluation. Using a 7-point SD scale, participants responded in regard to the extent they felt each word related to each scale. The questionnaire presented the stimuli in a random order. Participants were divided into six groups of 13, so the calculations included 13 people per sound-symbolic expression. Thus, each participant responded using 52 sound-symbolic expressions chosen from 312 expressions. Participants were unaware of the purpose of the experiments, and had no expert knowledge about linguistics. They were not trained to answer this type of questionnaire, and reported their intuitive impressions.

Table 1. List of adjective pairs

pair of adjectives		pair of adjectives	
warm	cool	simple	complex
hard	soft	fashionable	unfashionable
smooth	rough	masculine	feminine
slippery	sticky	intense	calm
bumpy	flat	loud	plain
wet	dry	western-style	Japanese-style
glossy	nonglossy	Luxurious	Austere
bright	dark	relieved	uneasy
stretchy	nonstretchy	good	bad
firm	fragile	impressive	unimpressive

sharp	dull	happy	sad
elastic	nonelastic	stable	unstable
thick	thin	comfortable	uncomfortable
heavy	light	eccentric	ordinary
regular	irregular	natural	artificial
repulsive	nonrepulsive	familiar	unfamiliar
clean	dirty	like	dislike
strong	weak	static	dynamic
sharp	mild	pleasant	unpleasant
modern	old-fashioned	positive	negative
fresh	annoying	young	old
elegant	vulgar		

Impression-rating predictive model

On the basis of the hypothesis that visual impressions associated with sound-symbolic expressions can be determined by the sound symbolism of each expression, we created an impression-rating predictive model as follows. Equation (1) predicts the value as a simple linear sum of the impact of each phoneme on the impression created by the expression, as a quantitative value.

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

Here, $\hat{Y}$ represents a predictive rating value of sound-symbolic words on a particular rating scale. X1 – X13 represent the category quantity (the degree of impact of each phoneme on the predictive rating value) for each phoneme. X1 – X6 represent the consonant category, voiced/semi-voiced, palatalized, lowercase vowel, vowel and medial indicator for the first mora (the smallest sound unit in Japanese), respectively, and X7 – X12 represent the consonant category, voiced/ semi-voiced, palatalized, lower case vowel, vowel, and end of a word indicator for the second mora, respectively. X13 represents the presence or absence of repetition. The detailed relationships between variables and phonemes are shown in Table 2.

This method produced 24,336 items of data (43 rating scales × 312 expressions × 13 participants). We then calculated the average rating value for each scale multiplied by each expression. By employing the average rating values of each sound-symbolic expression as the objective variables

and variation of phonemes used in the expression as the predictor variables, we conducted mathematical quantification theory class I. This method for quantifying qualitative data uses a type of multiple regression analysis, calculating the degree to which each phoneme contributes to each rating scale. Table 3 shows examples of the analysis results for each scale. From Equation (1) and Table 3, the rating values for each sound-symbolic expression can be determined by totaling the category values for each phoneme in the expression. For example, the expression “pika” is composed of the first mora /pi/ (/h/ semi-voiced sound /i/) and the second mora /ka/ (/ k/ /a/). Therefore, the value of the “bright–dark” scale is estimated by the following equation. For example, because evaluated values are rated on a 7-point scale, an estimated value of 2.91 indicates that “pika” is associated with the impression of brightness.

$\hat{Y}$ = /p/ + /i/ + /k/ + /a/ + Const.
= /h/ (X1) + semi-voiced sound (X2) + absence (X3) + /i/ (X4) + absence (X5) + absence (X6) + /k/ (X7) + absence (X8) + absence (X9) + /a/ (X10) + absence (X11) + absence (X12) + absence (X13) + Const.
= (-0.38) + (-0.66) + (0.06) + (0.44) + (-0.03) + (0.03) + (-0.35) + (-0.08) + (-0.02) + (-0.05) + (-0.03) + (0.00) + (0.12) + (3.86)
= 2.91.

The multiple-correlation coefficient R between the predicted values and average rating values (actual values) was used as an indicator of prediction accuracy. For the six pairs of scales, the R values ranged from 0.8 to 0.9. We thus considered our model to be appropriate for estimating sound-symbolic word impressions evaluated by people.

Table 2. Correspondence between variables and phonemes.

First mora	Second mora	Phonological characteristics	Phonemes
X1	X7	Consonants	/k/, /s/, /t/, /n/, /h/, /m/, /y/, /r/, /w/ or absence
X2	X8	Voiced sounds/p-sounds	Presence or absence
X3	X9	Contracted sounds	Presence or absence
X4	X10	Vowels	/a/, /i/, /u/, /e/, /o/
X5	X11	Semi-vowels	/a/, /i/, /u/, /e/, /o/ or absence
X6	X12	Special sounds	/N/, /Q/, /R/, /Li/ or absence
	X13	Repetition	Presence or absence

7. Results

We obtained 17,487 sound symbolic word tokens (1,827 sound symbolic word types) in Experiment 1, and 30,138 sound symbolic word tokens (2,442 sound symbolic word types) in Experiment 2. We analyzed the sound symbolic word tokens using our new methodology. When we input the sound symbolic word tokens into the system, the estimated rating values for 43 impression rating scales (i.e., 43 adjectives) were obtained by the impression-rating predictive model. We calculated the mean value of the impression rating values obtained from all sound symbolic words in Experiments 1 and 2, respectively. We then used Welch's t-tests to examine differences between the results of Experiments 1 and 2 (see Table 3). The analysis revealed significant differences in 20 scales of 43 impression rating scales between Experiments 1 and 2.

Table 3. Results of mean values of 43 impression rating scales in Experiments 1 and 2. The gray shaded areas represent differences that reached statistical significance.

43 adjective scales	Mean values (Experiment 1)	Mean values (Experiment 2)	P-value	43 adjective scales	Mean values (Experiment 1)	Mean values (Experiment 2)	P-value
bright-dark	-0.058	-0.059	0.570	simple-complex	-0.138	-0.130	0.000 ***
warm-cool	0.046	0.039	0.000 ***	like-dislike	0.055	0.053	0.357
thick-thin	-0.013	-0.018	0.096	slippery-sticky	-0.119	-0.109	0.000 ***
relieved-uneasy	0.130	0.127	0.055	sharp-dull	0.089	0.107	0.000 ***
good-bad	0.075	0.073	0.267	static-dynamic	0.208	0.208	0.895
impressive-unimpressive	-0.148	-0.148	0.835	fashionable-unfashionable	0.076	0.084	0.000 ***
happy-sad	-0.018	-0.021	0.145	pleasant-unpleasant	-0.062	-0.063	0.593
stable-unstable	0.184	0.185	0.405	masculine-feminine	0.050	0.055	0.008 **
comfortable-uncomfortable	0.128	0.124	0.145	elastic-nonelastic	0.223	0.214	0.002 **
hard-soft	0.071	0.084	0.000 ***	glossy-nonglossy	0.116	0.103	0.000 ***
regular-irregular	0.058	0.059	0.595	strong-weak	0.009	0.011	0.272
clean-dirty	0.029	0.030	0.629	bumpy-flat	-0.071	-0.072	0.664
modern-old-fashioned	-0.048	-0.043	0.000 ***	smooth-rough	0.033	0.026	0.025 *
eccentric-ordinary	0.033	0.034	0.782	stretchy-nonstretchy	0.204	0.198	0.003 **
fresh-annoying	0.123	0.129	0.002 **	intense-calm	0.011	0.024	0.000 ***
natural-artificial	0.060	0.059	0.265	loud-plain	0.072	0.078	0.000 ***
familiar-unfamiliar	-0.076	-0.079	0.130	positive-negative	-0.051	-0.053	0.418
wet-dry	0.172	0.166	0.055	Western-style-Japanese-style	-0.047	-0.042	0.000 ***
sharp-mild	0.031	0.049	0.000 ***	young-old	-0.115	-0.113	0.492
heavy-light	0.109	0.104	0.053	luxury-cheap	0.137	0.141	0.004 **
elegant-vulgar	0.096	0.100	0.033 *	**repulsive-nonrepulsive**	0.034	0.032	0.544
firm-fragile	0.085	0.082	0.010 *				

8. Discussion

The current study investigated the relationship between the use of sound symbolic words and the perception of texture between holistic and part-based image processes. We developed a new system for analyzing sound symbolic words by calculating subjective impressions of sound-symbolic words on the basis of the impressions evoked by each phoneme from a quantitative rating database. Thus, if different sound symbolic words are used for texture impression analysis, the texture impression values would also be expected to differ. However, if similar sound symbolic words were used in Experiments 1 and 2, the texture impression values would also be similar. The results revealed significant differences in texture impression between holistic and part-based image processes. These findings have a number of implications for understanding texture perception in the context of holistic and part-based visual processing, which are discussed below.

9. Differences between holistic and part-based image processing

The current results indicate that holistic processing influenced the perception of surface texture. In particular, there were significant differences in the 20 scales such as “warm-cool”, “hard-soft”, “slippery-sticky”, “elastic-nonelastic”, “glossy-nonglossy”, “firm-fragile”, “smooth-rough”, “stretchy-nonstretchy”, “sharp-mild”, “simple-complex”, “sharp-dull”, “fashionable-unfashionable”, “masculine-feminine”, “modern-old-fashioned”, “fresh-annoying”, “elegant-vulgar”, “intense-calm”, “loud-plain”, “Western-style-Japanese-style”, “luxury-cheap”. Fenko et al. (2010) reported that adjectives can be divided into three categories: sensory descriptors (e.g., hard, red, noisy); symbolic descriptors (e.g., interesting, expensive, modern); and affective descriptors (e.g., pleasant, beautiful). The current results show that the sensory and symbolic descriptors differed in texture perceptions between holistic and part-based image processing. Texture is typically considered to be a microstructural property of surfaces, processed independently of macrostructural properties such as form and shape. Although the basic elements of texture perception are deeply related to the sensation of touch, various textural properties of a surface can be

determined from visual information [8, 56]. In addition, visual perception of texture is heavily involved in the perception of visual complexity, emotional content and aesthetics [9, 10]. However, in the current study, participants’ impressions of texture differed in the sensory and symbolic descriptors between holistic processing and part-based image processing conditions. These findings indicate that the interaction between texture perception and holistic processing may affect the sensory and symbolic dimensions.

10.The relationship between sensory dominance and adjectives

In the current study, we investigated whether texture perception differed between holistic and part-based image processes using Japanese sound symbolic words. Visual texture is implicated not only in low-level perception but also in the perception of high-level features, such as visual complexity, emotions and aesthetics [9, 10, 60]. In the current study, we used 43 pairs of appropriate adjectives for describing the visual texture of objects, including low- and high-level visual characteristics [52-54]. However, there were significant differences in the sensory descriptors between holistic processing and part-based image processing conditions. According to previous studies, the sensory descriptors representing basic textural properties, such as softness, smoothness, slipperiness, is strongly involved in haptics [2-5]. Furthermore, Fenko et al. (2010) reported that touch is the dominant sensory modality in sensory descriptors related to texture perception and vision is the dominant sensory modality in symbolic and affective descriptors. The current results indicate that despite sensory descriptors (e.g., hard, sharp, rough) representing tactile dominance, the descriptors are implicated in the differences in texture perception between holistic and part-based image processing. These findings have implications regarding the mechanisms by which holistic processing affects sensory descriptors in surface texture perception.

How can the differences appear on the scales of tactile dominance? One possibility is that experience-based knowledge of material affected texture perception in the current study. Although material properties obtained via the tactile modality (i.e., hardness and roughness) are important, previous studies have reported that humans are able to perceive various textural properties of surfaces from visual information [6, 7]. In particular, several studies reported the importance of unsupervised learning of crossmodal associations through everyday experience [63-65]. In addition, recent studies have indicated that haptic or visuo-haptic experience influences the representation of neural activity patterns in material perception [66, 67]. Therefore, the interaction between holistic processing and the experience-based knowledge may affect visual perception of surface texture.

The present study examined whether the use of sound symbolic words to describe texture perception differs between holistic and part-based image processes. The results revealed several differences in the reported impressions of texture using sound symbolic words between holistic and part-based image processing. These findings suggest that holistic processing may influence sensory and symbolic descriptors in surface texture perception, and that sound symbolic words can describe differences in surface texture at a fine resolution. Moreover, the current findings suggest that the interaction between holistic processing and the experience-based knowledge may affect visual perception of surface texture. However, the limitation of this study is that it was studied only in Japanese. Therefore, there is a need to be investigated in other languages in the future. These results indicate that future developmental studies should further investigate the relationship between sound symbolic words and perceptual properties.

References

1. Adelson, E. H., & Bergen, J. R. 1991. The plenoptic function and the elements of early vision. In Landy M, Movshon JA (Eds.), Computational models of visual processing, 1–20. Cambridge: MIT Press.
2. Jones, L. A., & Lederman, S. J. 2006. Human hand function. Oxford University Press.
3. Lederman, S. J., & Klatzky, R. L. 2009. Haptic perception: A tutorial. Attention, Perception & Psychophysics 71, 1439–1459.
4. Bensmaia, S. J. 2009. Texture from touch. Scholarpedia 4, 7956.
5. Tiest, W. M. B. 2010. Tactual perception of material properties. Vision Research. 50, 2775–2782.
6. Whitaker, T. A., Simões-Franklin, C., & Newell, F. N. 2008. Vision and touch: independent or integrated systems for the perception of texture?. Brain Res 1242, 59–72.

7. Tamura, H., Mori, S., & Yamawaki, T. 1978. Textural features corresponding to visual perception. Syst. Man Cybern. IEEE Trans. 75, 460–473.
8. Lederman, S. J., Thorne, G., & Jones, B. 1986. Perception of texture by vision and touch: Multidimensionality and intersensory integration. Journal of Experimental Psychology: Human Perception & Performance 12, 169-180.
9. Liu, J., Lughofer, E., & Zeng, X. 2015. Aesthetic perception of visual textures: a holistic exploration using texture analysis. psychological experiment, and perception modeling, Front Comput Neurosci. 9, 134.
10. Guo, X., Asano, C. M., Asano, A., Kurita, T., & Li, L. 2012. Analysis of texture characteristics associated with visual complexity perception. Optical review. 19(5), 306–314.
11. Schyns, P. G., Oliva A. 1994. From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. Psychol. Sci. 5, 195–200.
12. Thorpe, S., Fize, D., & Marlot, C. 1996. Speed of processing in the human visual system. Nature. 381, 520–522.
13. Oliva, A. 2005. Gist of the scene. In: Itti L, Rees G, Tsotsos, JK. (Eds.), Neurobiology of Attention. Elsevier, San Diego, CA. 251–256.
14. Oliva, A, & Torralba, A. 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comp. Vis. 42: 145–175.
15. Parkes, L., & Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. 2001. Compulsory averaging of crowded orientation signals in human vision. Nat. Neurosci. 4, 739–744.
16. Ariely, D. 2001. Seeing sets: representation by statistical properties. Psychol. Sci. 12, 157–162.
17. Chong, S. C. 2003. Treisman A. Representation of statistical properties. Vision Res. 43, 393–404.
18. Oliva, A., & Torralba, A. 2002. Scene-centered description from spatial envelope properties. Lecture note in computer science series. Proceedings of the 2nd Workshop on Biologically Motivated Computer Vision, Tubingen, Germany.
19. Rao, A. R., & Lohse, G. L. 1993. Identifying high-level features of texture perception. GMIP. 55, 218–233.
20. Heaps, C., & Handel, C. H. 1999. Similarity and features of natural textures. J. Exp. Psychol. Hum. Percept. Perform. 25, 299–320.
21. Sucevic, J., Jankovic, D., & Kovic, V. 2013. When the sound-symbolism effect disappears: The differential role of order and timing in presenting visual and auditory stimuli. Psychology. 4(7A), 11–18.
22. Parise, C.V., & Spence, C. 2012. Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. Experimental Brain Research. 220, 319-333.
23. Spence, C. 2011. Crossmodal correspondences: a tutorial review. Atten Percept

Psychophys. 73(4), 1–25.
24. Ramachandran, V. S., & Hubbard, E. M. 2001. Synaesthesia—A window into perception, thought and language. Journal of Consciousness Studies. 8, 3–34.
25. Ramachandran, V. S., & Hubbard, E. M. 2003. Hearing colors, tasting shapes. Scientific American. 288, 43–49.
26. Suþeviü, J., Saviü, A. M., Popoviü, M. B., Styles, S. J, & Koviü, V. 2015. Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape-sound-meaning congruence in words, and shape-sound congruence in pseudowords. Brain and language. 145, 11-22.
27. Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. 2014. Crosslinguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. Brain and Language. 128, 18–24.
28. Spence, C., & Gallace, A. 2011. Tasting shapes and words. Food Quality and Preference. 22, 290–295.
29. Spence, C., & Deroy, O. 2013. Tasting shapes: A review of four hypotheses. Theoria et Historia Scientiarum.10, 207–238.
30. Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. 2013. "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. Cognition. 126, 165–172.
31. Nordberg, B. 1986. The use of onomatopoeia in the conversational style of adolescents'. FUMS rapport nr 132, Uppsala University. FUMS. Uppsala.
32. Jakobson, R., & Waugh, L. 1979. The sound shape of language, Harvester Press, Sussex.
33. Boyle, M. W., & Tarte, R. D. 1980. Implications for phonetic symbolism: The relationship between pure tones and geometric figures. Journal of Psycholinguistic Research. 9, 535–544.
34. Holland, M. 1964. Wertheimer M. Some physiognomic aspects of naming, or, maluma and takete revisited. Percept. Motor Skill. 19, 111–117.
35. Lindauer, M. S. 1990. The meanings of the physiognomic stimuli taketa and maluma. Bulletin of the Psychonomic Society. 28, 47–50.
36. Taylor, I. K. 1963. Phonetic symbolism re-examined. Psychological Bulletin. 60, 200–209.
37. Bankieris, K., & Simner, J. 2015. What is the link between synaesthesia and sound symbolism? Cognition. 136, 186–195.
38. Sapir, E. 1929. A study in phonetic symbolism. Journal of Experimental Psychology. 12, 225–239.
39. Köhler, W. 1929. Gestalt psychology. New York: Liveright.
40. Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. 2014. How sound symbolism is processed in the brain: a study on Japanese mimetic words. PLoS ONE. 9: 1–8.

41. Sakamoto, M., & Watanabe, J. 2013. Effectiveness of onomatopoeia representing quality of tactile texture' a comparative study with adjectives. Paper from the 13th National Conference of the Japanese Cognitive Linguistics Association. 473-485 (in Japanese)
42. Watanabe, J., Hayakawa, T., Matsui, S., Kano, A., Shimizu, Y., & Sakamoto, M. 2012. Visualizing tactile material relationships using sound symbolic words. Lecture Notes in Computer Science. 7283. Haptics: Perception, Devices, Mobility, and Communication. 175-182.
43. Hamano, S. 1998. The sound-symbolic system of Japanese. Stanford, CA & Tokyo: CSLI & Kuroshio Publisher.
44. Sakamoto, M., & Watanabe, J. 2013. Effectiveness of onomatopoeia representing quality of tactile texture' a comparative study with adjectives. Paper from the 13th National Conference of the Japanese Cognitive Linguistics Association. 473-485 (in Japanese)
45. Watanabe, J., Hayakawa, T., Matsui, S., Kano, A., Shimizu, Y., & Sakamoto, M. 2012. Visualizing tactile material relationships using sound symbolic words. Lecture Notes in Computer Science. 7283. Haptics: Perception, Devices, Mobility, and Communication. 175-182.
46. Sakamoto, M., Yoshino, J., Doizaki, R., & Haginoya, M. 2015. Metal-like Texture Design Evaluation Using Sound Symbolic Words. International Journal of Design Creativity and Innovation. 4(3-4), 181-194.
47. Sakamoto, M., & Watanabe, J. 2015. Cross-Modal Associations between Sounds and Drink Tastes/Textures: A Study with Spontaneous Production of Sound-Symbolic Words. Chemical Senses. 4(3), 197-203.
48. Evans, K. K., & Treisman, A. 2010. Natural cross-modal mappings between visual and auditory features. Journal of Vision. 10(1): 6:1–12.
49. Nahm, F. K. D., Tranel, D., Damasio, H., & Damasio, A. R. 1993. Cross-modal associations and the human amygdale. Neuropsychologia. 31, 727–744.
50. Westbury, C. 2005. Implicit sound symbolism in lexical access: Evidence from an interference task. Brain and Language. 93: 10–19.
51. Sharan, L., Rosenholtz, R., & Adelson, E. H. 2014. Material Perception: What Can You See in a Brief Glance?. Journal of Vision. 14(9), 12.
52. Baek, S., Hwang, M., Chung, H., & Koo, P. 2008. Kansei factor space classified by information for kansei modeling of image. Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation. 205(2), 874–882.
53. Choi, K., & Jun, C. 2007. A systematic approach to the kansei factors of tactile sense regarding the surface roughness. Appl Ergon. 38, 53–63.
54. Chen, Y. W., Sobue, S., & dan Huang, X. 2008. Mapping function of color image features and human KANSEI, IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 725.

55. Kakehi, H., Tamori, I., & Schourup, L. C. 1996. Dictionary of iconic expressions in Japanese. Berlin: Mouton de Gruyter.
56. Lederman, S. J., & Abbott, S. 1981. Texture perception: Studies of intersensory Lntegration using a discrepancy paradigm. Journal of Experimental Psychology: Human Perception & Performance. 7: 902-91.
57. Schifferstein, H. N. J., & Cleiren, M. P. H. D. 2005. Capturing product experiences: a split-modality approach. Acta Psychol. (Amst.) 118, 293–318.
58. Schifferstein, H. N. J. 2006. The perceived importance of sensory modalities in product usage: a study of self reports. Acta Psychol. (Amst.) 121, 41–64.
59. Lennie, P. 1998. Single units and cortical organization. Perception. 27, 889-935.
60. Kastner, S., de Weerd, P., & Ungerleider, L.G. 2000. Texture segregation in the human visual cortex: a function MRI study. Journal of Neurophysiology. 83, 2453-2457.
61. 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.
62. 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), 31-41.
63. Flanagan, J. R., Bittner, J. P., & Johansson, R. S. 2008. Experience can change distinct size-weight priors engaged in lifting objects and judging their weights. Curr. Biol. 18, 1742–1747.
64. Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. 2007. Simultaneous and independent acquisition of multisensory and unisensory associations. Perception. 36: 1445–1453.
65. Ernst, M. O. 2007. Learning to integrate arbitrary signals from vision and touch. J. Vis. 7(7), 1–14.
66. Goda, N., Yokoi, I., Tachibana, A., Minamimoto, T., & Komatsu, H. 2016. Crossmodal association of visual and haptic material properties of objects in the monkey ventral visual cortex. Current Biology. 26(7), 928-934.
67. Goda, N., Tachibana, A., Okazawa, G., & Komatsu, H. 2014. Representation of the material properties of objects in the visual cortex of nonhuman primates. J Neurosci. 34(7), 2660-2673.
68. R. Doizaki, J. Watanabe, and M. Sakamoto, "Automatic Estimation of Multidimensional Ratings from a Single Sound-symbolic Word and Word-based Visualization of Tactile Perceptual Space," IEEE Transactions on Haptics, vol. 10, no. 2, pp. 173–182, 2017.