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.
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,