Multisensory Shitsukan perception

Waka Fujisaki*

Department of Psychology, Faculty of Integrated Arts and Social Sciences, Japan Women’s University, 1–1–1 Nishi-Ikuta, Tama-ku, Kawasaki, 214–0014 Japan

Abstract: Shitsukan is a Japanese word that means “a sense of quality.” Shitsukan includes visual qualities such as “glossiness” and “translucency”; acoustic qualities such as “brightness” “sharpness” and “pitch”; tactile qualities such as “roughness” and “hardness”; aspects of materials themselves such as “glass,” “cloth,” “wood,” “stone,” “metal,” and “pearl”; and affective properties such as “prettiness,” “fragility,” “expensiveness,” “preference,” “naturalness,” and “genuineness.” Thus, a wide range of concepts has been examined with respect to the Shitsukan perception. It is also important to note that Shitsukan perception is not merely the processing of information input through various sensory modalities; it also results from multimodal, adaptive, and active processes including prediction, decision-making, body motor control, and sensory-motor feedback. In this review, I would like to introduce the following three studies that my collaborators and I have recently conducted. 1) Auditory modulation of material properties of food by pseudo-mastication feedback sound generated from electromyogram signal; 2) Perception of the material properties of wood based on vision, audition, and touch; and 3) The rules of audiovisual integration in human perception of materials.

Keywords: Shitsukan, Pseudo-mastication feedback sound, Food texture perception, Material properties of wood, Audiovisual integration rules

PACS number: 43.66.Lj [doi:10.1250/ast.41.189]

1. MULTISENSORY SHITSUKAN PERCEPTION

Shitsukan is a Japanese word that means “the sense of quality.” A wide range of concepts has been examined with respect to the Shitsukan perception. Shitsukan includes visual qualities such as “glossiness” [1–4] and “translucency” [5,6]; acoustic qualities such as “brightness,” “sharpness,” and “pitch” [7–13] tactile qualities such as “roughness” and “hardness” [14–17]; aspects of materials themselves such as “glass,” “cloth,” “wood,” “stone,” “metal,” and “pearl” [18–21] and affective properties such as “prettiness,” “fragility,” “expensiveness,” “preference,” “naturalness,” and “genuineness” [18,22–24].

It is also important to note that Shitsukan perception is not merely the processing of information input through a variety of sensory modalities; it also results from multimodal, adaptive, and active processes including prediction, decision-making, body motor control, and sensory-motor feedback.

* e-mail: fujisakiw@fc.jwu.ac.jp

2. AUDITORY MODULATION OF MATERIAL PROPERTIES OF FOOD BY PSEUDO-MASTICATION FEEDBACK SOUND GENERATED FROM ELECTROMYOGRAM SIGNAL

Perception of food textures is an optimal subject area to investigate multisensory Shitsukan processes because sensory feedback arising from active eating behavior contributes greatly to food texture perception.

Several studies have used auditory feedback of mastication sounds to influence participants’ perception of food textures. For instance, Zampini and Spence [17] reported that the perception of crispness and staleness of potato chips can be altered by varying the loudness or frequency composition of the sound of the first bite. Their work received the Ig Nobel prize for nutrition in 2008. However, this epoch-making study was limited in the sense that it only examined crispy foods (e.g., potato chips or crackers) that can produce distinct mastication sounds.

To examine not only brittle food but any food that have different physical properties, other researchers tried to identify the mastication onset, and by using it as a trigger, they played various pre-recorded mastication sounds. For
example, Masuda and Okajima [25] detected mastication onset using small microphones placed in participants’ ear canals; Koizumi et al. [26] detected mastication onset from jaw motion using photo reflectors. However, these methods can synchronize only the onset, but not the offset or biting force. Thus, their methods were also limited in that they were not able to provide perfectly synchronized sounds.

Endo, Ino, and Fujisaki [27] recently found a way to solve these problems. In order to achieve temporal synchrony, they developed a technique to convert the masseter muscle’s electromyogram (EMG) waveforms during mastication to acoustic signals and provide them to participants as real-time feedback. Their method enables the feedback of perfectly synchronized mastication sounds for foods with any type of physical properties.

EMG sound feedback has the following advantages. First, the EMG signal is an electrical waveform, and it can be readily interpreted as a sound. Second, onset and offset of mastication sound are synchronized because sound is produced only when chewing force is applied. Third, EMG signal strength is proportional to mastication force; therefore, mastication force is proportional to sound volume. Fourth, mastication EMG sound is subjectively similar to the natural mastication sound.

In their study, Endo et al., [27] investigated whether pseudo-mastication sound feedback generated from EMG signals can modulate the perception of food textures. Thirty adults participated in this study. Five commercially available nursing care foods packaged in retort pouches were used. They employed a two-factor, within participants design and used five different foods and assessed sound feedback with EMG signals put on and off. Participants responded using a questionnaire containing two sheets: Questionnaire 1 evaluated material properties and Questionnaire 2 compared with-sound and no-sound conditions and evaluated their general impression.

The results indicated that perceived chewiness and roughness were changed by the EMG sound feedback. Additionally, the participants felt more excited and involved in dining experiences. Thus, EMG sound feedback significantly affects perceptions of properties such as chewiness and roughness. It also appears to affect higher-order food perceptions such as “exciting” and “engaging dining experience.” Furthermore, they found that by comparing conditions with and without sound feedback, for some foods, EMG sound feedback appeared to promote the illusory perception of an increased number of ingredients in the consumed foods. Thus the participants felt that there were more ingredients in the food when they received the pseudo-mastication feedback sounds while eating.

Endo et al. [27] conducted the experiment with healthy adults. However, such a system is considered to be necessary for elderly persons whose function of chewing or swallowing is declining. Thus, Endo et al. [28] conducted a similar experiment with healthy elderly participants. This study also examined the influences of texture inhomogeneity on the effects of chewing sound modulation. Three kinds of nursing care foods in two food process types (minced-/puréed-like foods for inhomogeneous/homogeneous texture respectively) were used as sample foods. A pseudo-chewing sound presentation system using EMG signals was used to modulate chewing sounds. Thirty healthy elderly individuals participated in the experiment. In two conditions, with and without the pseudo-chewing sound, participants rated the taste, texture, and feelings evoked in response to sample foods. The results showed that inhomogeneity strongly influenced the perception of food texture. Regarding the effects of the pseudo-chewing sound, taste was less influenced, the perceived food texture tended to change in the minced-like foods, and the evoked feelings changed in both food process types. Though there were some food-related differences in the effects of the pseudo-chewing sound, the presentation of the pseudo-chewing sounds was more effective in foods with an inhomogeneous texture.

As for the next step, Endo, Kaneko, Ino, and Fujisaki [29] improved the system, which could change the chewing sound to the intended one without changing the envelope of EMG signals but changing the career sound. In their system, as illustrated in Fig. 1, first, the EMG was measured from masseter; then the EMG signal envelope was extracted with an analog integration circuit. The amplitude modulation of base chewing sounds was realized with a digital audio volume controller and micro-computer. The base chewing sounds were stored in a digital audio player, and one of the base chewing sounds was played. Each base chewing sound had to be continuous. These continuous sounds were made using the actual chewing sounds. To record the actual chewing sounds, earphones designed for binaural recording were used; outward-facing microphones were installed bilaterally in the earphone housing, and air-conducted chewing sounds were recorded at ear level. Signals corresponding to silent periods were removed and those corresponding to sounds were extracted and connected. Finally, the base chewing sound of 1-min duration was played continuously with the repeat playback function of the digital audio player.

This system enabled the presentation of a wide variety of mastication sounds synchronously with chewing motion. Tasks for the future include system miniaturization and adaptation to clinical situations. In addition to clinical use, it may be used for entertainment and suggesting new food experiences.
3. PERCEPTION OF THE MATERIAL PROPERTIES OF WOOD BASED ON VISION, AUDITION, AND TOUCH

When separately evaluating material properties (Shitsukan) of the same target objects using vision, audition, and touch, would the judgments of affective properties of materials be similar if the target objects were the same, even if the sensory modalities were different?

According to Gibson [30], sensory stimulation is registered by a set of perceptual systems that are directly responsive to amodal invariants. For example, a fire is a source of four kinds of stimulation: sound, odor, heat, and light. Each type of stimulation specifies the same event, and each alone specifies the event. Gibson [30] claims that the four kinds of stimulus information and the four perceptual systems are equivalent. Therefore, the perception of fire is simply the reception of information; the perception will be the same regardless of which system is activated, even though the conscious sensations will be different.

Material perception (Shitsukan) involves a wide range of concepts, perceptual properties, and affective properties that are intermixed. Thus, from a Gibsonian point of view, one can hypothesize that all three modalities (vision, audition, and touch) will provide similar results for the perception of affective properties, which are considered to evoke relatively higher levels of processing.

Fujisaki, Tokita, and Kariya [22] investigated whether the same affective classifications of materials can be found in three different modalities of vision, audition, and touch, using wood as the target object. They used wood because it is familiar, has an abundance of variation, and contains much visual, auditory, and tactile information.

Fifty participants took part in the experiment involving the three modalities of vision, audition, and touch, in isolation. Twenty-two different wood types including genuine, processed, and fake were perceptually evaluated using a questionnaire consisting of 23 items (12 perceptual and 11 affective).

Visual stimuli were photographs of each test piece. All photographs were presented on a display with the same visual angle relative to the observer. Auditory stimuli (Fig. 2) were recordings of each test piece being struck with a mallet: each test piece was placed on a xylophone from which all boards had been removed and the ball of a mallet was placed to fall on the test piece from a height of 1 cm. All sounds were generated by striking the pieces in this fashion. The recorded sound was edited to include a duration of 1 s and then repeated five times, so that the duration of each auditory stimulus was 5 s. Auditory stimuli were presented through headphones. Tactile stimuli were the original test pieces placed in a small paper box with an opening at the top of the panel. Participants inserted their index finger through the opening to touch the test piece. The paper box was placed to a box with a black curtain drawn around it, so that participants could not see the test pieces.

The results demonstrated that evaluations of the affective properties of wood were similar in all three modalities. The elements of “expensiveness, sturdiness, rareness, interestiness, and sophistication” and “pleasantness, relaxed feelings, and liked–disliked” were separately grouped for all three senses. Their results suggest that the affective material properties of wood are at least partly represented in a supramodal fashion. These results also suggest an association between perceptual and affective properties, which will be a useful tool not only in science, but also in applied fields.
The phenomenon, termed cross-modal correspondence, may result from correlations between multisensory signals learned in daily experience of the natural environment. If so, one would observe cross-modal correspondences in the perception of not only artificial stimuli but also natural objects. Kanaya, Kariya, and Fujisaki [31] reanalyzed part of Fujisaki et al.’s dataset [22] in order to determine the extent to which participants’ evaluations are in line with the hypothesis of cross-modal correspondence regarding vision, audition, and touch. Kanaya et al. [31] compared participant evaluations of three perceptual properties (surface brightness, sharpness of sound, and smoothness) of wood blocks obtained separately via vision, audition, and touch. It was difficult for participants to assign ratings via sensory modalities that are not typically involved in perceiving the relevant property. Therefore, participants were instructed to make speculative ratings in such cases. For example, participants rating the surface brightness of a stimulus in the auditory condition did so by estimating how bright the surface was likely to be purely based on the sound. Significant positive correlations were identified for all properties in the audition-touch comparison. By

Fig. 2 Spectrograms and sound waves of 22 varieties of stimuli. (a) Fourteen varieties of genuine wood from different tree species (b) four varieties of wood from the same species (cedar) processed in different ways (c) four varieties of fake wood (wood-grain sheet attached to non-wood materials) (d) keying device (view from above and from the side) (reproduced from Fujisaki, Tokita & Kariya, 2015, Vision Research).
contrast, no properties exhibited significant positive correlations in the vision-audition comparison. These results suggest that we learn correlations between multisensory signals through experience; however, the strength of this statistical learning is apparently dependent on the particular combination of sensory modalities involved.

4. THE RULES OF AUDIOVISUAL INTEGRATION IN THE HUMAN PERCEPTION OF MATERIALS

What material is perceived when the visual appearance of one material is combined with the impact sound of another, and what are the rules that govern cross-modal integration of material information?

Fujisaki, Goda, Motoyoshi, Komatsu, and Nishida [32] recently found a strong interaction between audiovisual material perceptions; for example, an object appearing to be glass was perceived as transparent plastic when paired with the sound of a pepper being hit. They also found that material-category-likelihood ratings follow a multiplicative integration rule, while material-property ratings follow a weighted average rule; both can be interpreted as optimal Bayesian integration.

In their experiment, visual stimuli were computer-generated movies of a scene in which a human hand hit an object with a small stick. Figure 3 illustrates images for the six material categories they used: glass, ceramic, metal, stone, wood, and bark. They chose these material categories, their textures, and the cylindrical object shape following Hiramatsu et al. [19]. Glass had a transparent body with a glossy surface. The other materials were opaque.

For auditory stimuli, they used the impact sounds of eight real objects (glass, ceramic, metal, stone, wood, vegetable (pepper), plastic, and paper). These were selected from 16 sounds based on the data of a preliminary experiment. Auditory stimuli were created by hitting real objects with a wooden (maple) mallet in a soundproof chamber.

Seven visual stimuli (including a blank image for auditory-only conditions) and nine auditory stimuli (including a silent sound for visual-only conditions) were combined to create 62 combinations of audiovisual stimuli (excluding the combination of a blank image and a silent sound).

They conducted a multiple regression analysis using all the data, including a multiplicative interaction term. The regression equation was as follows.

\[ VA = \beta_0 + \beta_1 V_{\text{only}} + \beta_2 A_{\text{only}} + \beta_3 V_{\text{only}} \times A_{\text{only}} \]

In this equation, VA indicates the predicted audiovisual rating; \( V_{\text{only}} \) and \( A_{\text{only}} \) were obtained visual-only and auditory-only ratings, respectively; and \( V_{\text{only}} \) times \( A_{\text{only}} \) was the product of the two obtained ratings.

Regression weights (\( \beta_1, \beta_2, \) and \( \beta_3 \)) for visual-only, auditory-only, and the interaction terms are shown in Fig. 4. If VA can be explained by a simple weighted average of visual and auditory information, the regression weight of the interaction term, \( \beta_3 \), would be very small or zero. However, their result clearly showed that the value of this interaction term was significantly higher than zero. Therefore, multiplication can be considered an AND operation, which identifies the material category most consistent with both visual-only and auditory-only stimuli.

As with material-category ratings, they conducted a multiple regression analysis, including an interaction term. Regression weights (\( \beta_1, \beta_2, \) and \( \beta_3 \)) for visual, auditory, and audiovisual terms are also shown in Fig. 4. For visual properties, the regression weight for vision (\( \beta_1 \)) was high, and that for audition (\( \beta_2 \)) was very low. For auditory properties, the regression weight for audition (\( \beta_2 \)) was high, and that for vision (\( \beta_1 \)) was nearly zero. For other properties, the regression weights were comparable to those for audition and vision. Importantly, for all types of material properties, the regression weight for the interaction term (\( \beta_3 \)) was nearly zero, contrary to those of the material-category ratings. It should be noted that in other property ratings, the participants had to use two sources of information to infer the property values from a potentially ambiguous stimulus as in the case of material-category
ratings. Thus, audiovisual material-property ratings appear to follow a weighted average rule. The weight was given almost exclusively to vision and audition for visual and auditory properties, respectively. The weight for “other properties” was similar for the two modalities when averaged over the nine properties they used.

5. CLOSING REMARKS

This review mainly introduced the researches that authors have conducted in recent years in the field of multisensory Shitsukan perception. The Shitsukan perception research has been progressively increasing in recent years, and various studies on the multisensory Shitsukan perception also have been carried out. Multisensory perception is a treasure house of Shitsukan research, but experts in the field of audition are not involved as much as those in visual field. It is expected that many studies will be carried out involving experts from the field of audition in the future.

ACKNOWLEDGMENT

WF is currently supported by JSPS KAKENHI Grant Number 17K04514.

REFERENCES

[1] R. W. Fleming, R. O. Dror and E. H. Adelson, “Real-world illumination and the perception of surface reflectance properties,” J. Vis., 3, 347–368 (2003).
[2] I. Motoyoshi, S. Nishida, L. Sharan and E. H. Adelson, “Image statistics and the perception of surface qualities,” Nature, 447(7141), 206–209 (2007).
[3] S. Nishida and M. Shinya, “Use of image-based information in judgments of surface-reflectance properties,” J. Opt. Soc. Am. A Opt. Image. Sci. Vis., 15, 2951–2965 (1998).
[4] H. Tsuda and J. Saiki, “Constancy of visual working memory of glossiness under real-world illuminations,” J. Vis., 18(8), p. 14 (2018).
[5] R. W. Fleming, F. Jakel and L. T. Maloney, “Visual perception of thick transparent materials,” Psychol. Sci., 22, 812–820 (2011).
[6] I. Motoyoshi, “Highlight-shading relationship as a cue for the perception of translucent and transparent materials,” J. Vis., 10(9), p. 6 (2010).
[7] A. Almeida, E. Schubert, J. Smith and J. Wolf, “Brightness scaling of periodic tones,” Atten. Percept. Psychophys., 79, 1892–1896 (2017).
[8] M. Aramaki, M. Besson, R. Kronland-Martinet and S. Ystad, “Controlling the perceived material in an impact sound synthesizer,” IEEE Trans. Audio Speech Lang. Process., 19, 301–314 (2011).
[9] B. L. Giordano and S. McAdams, “Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates,” J. Acoust. Soc. Am., 119, 1171–1181 (2006).
[10] R. L. Klatzky, D. K. Pai and E. P. Krotkov, “Perception of material from contact sounds,” Presence: Teleoperators Virtual Environ., 9, 399–410 (2000).
[11] G. Lemaitre and L. M. Heller, “Auditory perception of material is fragile while action is strikingly robust,” J. Acoust. Soc. Am., 131, 1337–1348 (2012).
[12] R. A. Lutfi and E. L. Oh, “Auditory discrimination of material changes in a struck-clamped bar,” J. Acoust. Soc. Am., 102, 3647–3656 (1997).
[13] R. P. Wildes and W. A. Richards, “Recovering material properties from sound,” in Natural Computation, W. A. Richards, Ed. (MIT Press, Cambridge, 1988).
[14] S. Guest and C. Spence, “What role does multisensory integration play in the visuotactile perception of texture?” Int. J. Psychophysiol., 50, 63–80 (2003).
[15] S. J. Lederman and R. L. Klatzky, “Hand movements: A window into haptic object recognition,” Cogn. Psychol., 19, 342–368 (1987).
[16] S. Okamoto, H. Nagano and Y. Yamada, “Psychophysical dimensions of tactile perception of textures,” IEEE Trans. Haptics, 6, 81–93 (2013).
[17] M. S. Zampini and C. Spence, “The role of auditory cues in modulating the perceived crispness and staleness of potato chips,” J. Sens. Stud., 19, 347–363 (2004).
[18] R. W. Fleming, C. Wiebel and K. Gegenfurtner, “Perceptual
qualities and material classes,” *J. Vis.*, 13(8) (2013).

[19] C. Hiramatsu, N. Goda and H. Komatsu, “Transformation from image-based to perceptual representation of materials along the human ventral visual pathway,” *NeuroImage*, 57, 482–494 (2011).

[20] L. Sharan, R. Rosenholtz and E. H. Adelson, “Accuracy and speed of material categorization in real-world images,” *J. Vis.*, 14(9) (2014).

[21] Y. Tani, T. Nagai, K. Koida, M. Kitazaki and S. Nakauchi, “Experts and novices use the same factors—but differently—to evaluate pearl quality,” *PLoS One*, 9(1), p. e86400 (2014).

[22] W. Fujisaki, M. Tokita and K. Kariya, “Perception of the material properties of wood based on vision, audition, and touch,” *Vision Res.*, 109, 185–200 (2015).

[23] K. E. Overvliet and S. Soto-Faraco, “I can’t believe this isn’t wood! An investigation in the perception of naturalness,” *Acta Psychol.*, 136, 95–111 (2011).

[24] P. Rozin, “The meaning of “natural”: Process more important than content,” *Psychol. Sci.*, 16, 652–658 (2005).

[25] M. Masuda and K. Okajima, “Effects of mastication sound in food texture perception and pleasantness,” *IEICE Tech. Rep.*, 111, 57–62 (2011).

[26] N. Koizumi, H. Tanaka, Y. Uema and M. Inami, “Chewing JOCKEY: Designing user interface to augment food texture with sound AR system,” *Trans. Virtual Real. Soc. Jpn.*, 18, 141–150 (2013).

[27] H. Endo, S. Ino and W. Fujisaki, “The effect of a crunchy pseudo-chewing sound on perceived texture of softened foods,” *Physiol. Behav.*, 167, 324–331 (2016).

[28] H. Endo, S. Ino and W. Fujisaki, “Texture-dependent effects of pseudo-chewing sound on perceived food texture and evoked feelings in response to nursing care foods,” *Appetite*, 116, 493–501 (2017).

[29] H. Endo, H. Kaneko, S. Ino and W. Fujisaki, “An attempt to improve food/sound congruity using an electromyogram pseudo-chewing sound presentation system,” *J. Adv. Comput. Intell. Intell. Inform.*, 21, 342–349 (2017).

[30] J. J. Gibson, *The Senses Considered as Perceptual Systems* (Greenwood Press, Westport, Connecticut, 1966).

[31] S. Kanaya, K. Kariya and W. Fujisaki, “Cross-modal correspondence among vision, audition, and touch in natural objects: An investigation of the perceptual properties of wood,” *Perception*, 45, 1099–1114 (2016).

[32] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu and S. Nishida, “Audiovisual integration in the human perception of materials,” *J. Vis.*, 14(4) (2014).