https://doi.org/10.2352/issn.2169-2629.2020.28.20

Investigating Effects of Visual and Auditory Adaptation on Metallic Material Appearance

Takumi Nakamura [1], Yagi Daichi [1], Kuangzhe Xu [2], Toshihiko Matsuka [2], Keita Hirai [1]
1) Department of Imaging Sciences, Chiba University; 1-33, Yayoicho, Inage Ward, Chiba-shi, Chiba, 263-8522 Japan.
2) Graduate School of Humanities and Social Sciences, Chiba University; 1-33, Yayoicho, Inage Ward, Chiba-shi, Chiba, 263-8522 Japan.

Abstract
In this paper, we investigated the effects of visual and auditory adaptation on material appearance. The target in this study was metallic perception. First, participants evaluated CG images using sounds and other images. In the experiment, we prepared metallic stimulus under various adaptation conditions with different combinations of metal image, non-metal image, metal sound, and non-metal sound stimuli. After these adaptations, the participants answered "metal" or "non-metal" after viewing a displayed reference image. The reference images were generated by interpolating metal and non-metal images. Next, we analyzed the results and clarified the effects of visual, auditory, and audiovisual adaptations on the metallic perception. For analyzing results, we used a logistic regression analysis based on Bayesian statistics. From the analysis results, we found visual and auditory adaptation effects. On the other hand, we did not find the cross-modal effects of audiovisual adaptation. Finally, we created a model of the linear sum of the visual and audio adaptation effects on metallic material appearance.

Introduction
Material perception is an essential factor that determines the value of everything. For example, material appearance such as glossiness, transparency, roughness and warmth show the state of things. Material appearance cannot be represented by color or shape. For this reason, in recent years, many researchers have been investigated the relationship between physical laws (such as the reflection characteristics of the object surface) and human perception of material surfaces (such as for interpreting how humans perceive material surfaces.) Pellacini et al. [1] investigated the relationship between physical parameters and the glossiness of objects by experiments and proposed a new light reflection model for CG (Computer Graphics). The model developed by the research-based on psychophysics reproduces the glossiness.

On the other hand, Fleming et al. [2] clarified to what extent lighting conditions contribute to material appearance (especially the reflection characteristics of the object surface) by an evaluation experiment using the surface reflectance matching task. From various viewpoints, many researchers approach the study of material appearance; such as a study that clarified the effect of the shape of an object on the material appearance [3], and a study to explain that humans could recognize materials quickly and accurately using an image database containing 1000 images [4]. All these studies deal with visual material perception (in other words, material appearance).

When perceiving materials, humans usually combine not only visual information but also information of multiple senses, which is called cross-modal perception; such as the feeling of touching the object (tactile information) and the sound of hitting the object (auditory information). Based on this knowledge, in recently, researchers have actively studied material perception in multi-sensory cross-modal. Fujisaki et al. experimented on the material perception concerning the vision and audition [5]. They also investigated the material perception in the three senses of vision, audition, and touch [6]. Rodrigo et al. [7] discovered that perceiving a texture in a cross-modal manner can perceive the material surface richer than perceiving it with a single sensation. Thus, the material perception that integrates senses tends to be different from visual perception alone.

Besides, the color adaptation effect is generally well known. In recent years, there is an adaptation effect in visual material appearance. For example, when a participant adapts to its 3D shape and material, an illusion occurs in which the 3D shape and material appear to change, which is called "object aftereffect illusion" [8, 9, 10]. This research suggests that the object aftereffect illusion can occur even for material appearances such as gloss and unevenness. However, material appearance in the adaptation of senses except visual is not fully verified. Thus, it is unclear how the adaptation except for visual effects material appearance.

Therefore, in this research, we focused on not only visual but also auditory material perception. We investigated the effects of the adaptation of audiovisual stimuli on material appearance through subjective evaluation experiments. Particularly. We focused on metallic material appearance, which is an essential factor in much industrial use. We conducted subjective evaluation experiments, using sound and CG images. Finally, we analyzed the experimental results to clarify the adaptation effects of audiovisual perception on metallic material appearance.

Subjective Evaluation Experiment
In the experiment, we verified whether the visual, auditory, and audiovisual adaptation effects on the visual perception of metallic. We used CG images and sounds as experimental stimuli. After various adaptation conditions, we asked the participants to evaluate the intermediate image of metal and non-metal, whether they felt metal or non-metal. Experimental instruments were a headphone for sound stimuli, a display for visual stimuli (images), and a mouse for their response.

Visual and Audio Stimulus
We used PBRT (Physically Based Rendering Tool), which is a RGB-based renderer and created visual stimuli. The stimuli were high-quality CG images by the Naive Ray Tracing method. Rendering metal images and non-metal images in raw-data with PBRT, we converted these raw-images into tif-images for display.

There were two types of object shapes, buddha and sphere. Both they had two types, the stimuli for visual adaptation of participants and the stimuli for evaluation.

There were two types of visual adaptation stimuli: a metal image and a non-metal image. The reflection model used to create the metal images was the Torrance-Sparrow model. It had reflection characteristics of silver and set the roughness of specular reflection to 0.0016. A non-metal image was a matte (diffuse) object with a reflectance calculated from the refractive index of silver.

The evaluation stimuli were intermediate image between the metal image and the non-metal image. Then, We created intermediate images with linear interpolation of the metal and non-metal tif-images. There were 17 kinds of mixture ratios; metal image: non-metal image = 0.1: 0.9, 0.15: 0.85, …, 0.85: 0.15, 0.9: 0.1, which were 0.05 increments.

Figure 1 and Figure 2 show some of the visual stimuli. When the shape is buddha, Figure 1 shows a metal image (a), a non-metal image (b), and an intermediate image (c) in which the metal image and the non-metal image mixes at 0.5: 0.5. When the shape is a sphere, Figure 2 shows a metal image (a), a non-metal image (b), and an intermediate image (c) in which the metal image and the non-metal image mixes at 0.5: 0.5. Figures 1 (a), (b) and Figure 2 (a), (b) are adaptation stimuli, and Figures 1 (c) and Figure 2 (c) are part of the evaluation stimuli. The display showed these images at 960 × 720px.

For auditory stimuli, on the other hand, we used a metallic sound and a non-metallic sound, which were free materials on the web. We chose these synthetic sounds; a sound like hitting metal, and a sound like hitting a plastic.

Experiment Procedure

In the experiment, we used nine adaptation conditions for every two shapes (buddha, sphere) of visual stimulus. After looking at these nine adaptation images, the participants answered whether the displayed evaluation stimulus (an intermediate image between a metallic image and non-metallic image) was metallic or non-metallic. Table 1 shows the nine adaptation conditions, and Figure 3 shows the flow of the experiment. The order of experiments was Experiment 1 to 4, and the adaptation conditions changed A to D. Experiment 1 was a non-adaptation condition. Experiments 2A and 2B were visual adaptations. Experiments 3A and 3B were auditory adaptations. Experiments 4A, 4B, 4C and 4D were audiovisual adaptation conditions. For example, Experiment 4A was a case of adapting a metal image and a metal sound. The adaptation time was 15 seconds in all experiments. While 15 seconds, we kept displaying an image and playing a single sound every second. The adaptation was performed for each evaluation stimulus each time. We displayed evaluation stimulus for 0.5 seconds after adaptation. The evaluation stimulus was presented randomly in each experiment so that the participants will not remember the evaluation. We conducted these series of experiments for three times. Therefore, each participant evaluated a total of 918 trials (= 17 (stimuli) × 9 (conditions) × 2 (shapes) × 3 (times)).

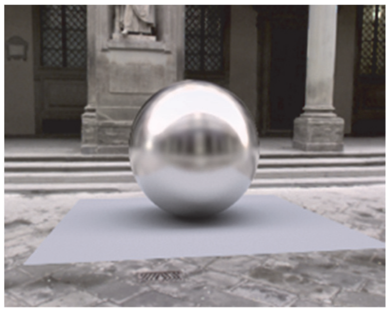
(a) non-metallic image

(b) metallic image

(c) metal : non-metal = 0.5: 0.5

Figure 1. *Sample images of the visual stimulus of the buddha*

(a) non-metallic image



(b) metallic image

(c) metal : non-metal = 0.5: 0.5

Figure 2. *Sample images of the visual stimulus of the sphere*

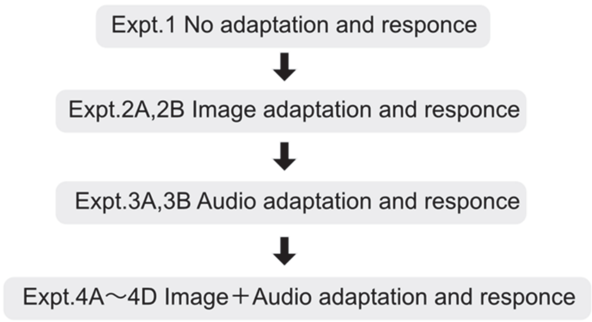


Figure 3. *The experiment flow.*

Table1. Adaptation condition. The experiment (Expt.) number is described. Visual adaptation had three conditions: no adaptation, metal image adaptation, non-metal image adaptation. Auditory adaptation had three states: no adaptation, metallic sound adaptation, non-metallic sound adaptation. Besides, the image shape had two types: buddha and sphere.

Adaptation Condition		In Audio		
		No Sound	Metallic Sound	Non-Metallic Sound
In Visual	No Image	Expt.1	Expt.3A	Expt.3B
	Metal Image	Expt.2A	Expt.4A	Expt.4B
	Non-Metal Image	Expt.2B	Expt.4C	Expt.4D

Experiment Environment

To use sound in the experiment, we tried to eliminate noise from the outside. The display was ColorEdge CG-221 BK (EIZO), and the headphone was ATH-M50x (audio-technica). The viewing distance was 70 cm. In each experiment, the participants were ten men in their 20s. Figure 4 shows how we experimented with an audiovisual adaptation. The participant watched the adaptation stimulus on the display while listening to the sound with headphones. The experiment environment was actually in a dark room.

Subjective Evaluation Experiment

Results and Discussions for Visual Adaptation

Figure 5 shows the number of metal responses for each evaluation stimulus with no adaptation (Expt.1), metal image adaptation (Expt.2A), and non-metal image adaptation (Expt.2B). The yellow line shows the result of no adaptation (Expt.1), the blue line shows the result of metal image adaptation (Expt.2A), and the red line shows the result of non-metal image adaptation (Expt.2B). Compared to no adaptation, metal image adaptation had a smaller number of metal responses. On the other hand, in the non-metal image adaptation, the number of metal responses increased compared to the case without adaptation. When performing t-test at a significance level of 5%, we observed significant differences between no adaptation, metal image adaptation, and non-metal image adaptation. In other words, visual adaptation seemed to affect the perception of metallic feeling in the opposite direction of the adaptation stimulus (such as visual aftereffect on color and motion perception). This result shows the same tendency as the one of Motoyoshi [8, 9].

Figure 4. *The experiment of audiovisual adaptation. In the actual experiment, the room light was turned off.*

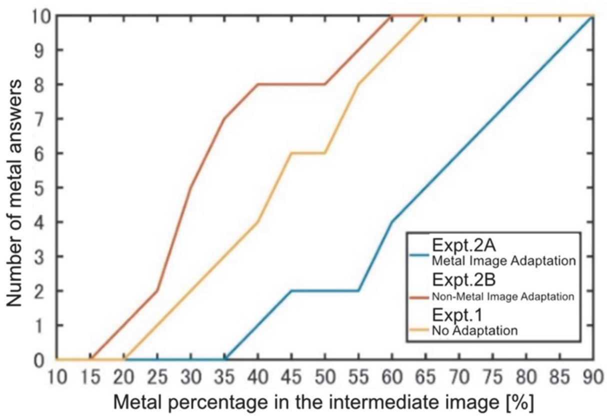


Figure 5. *No Adaptation VS Visual Adaptation*

Results and Discussions for Auditory Adaptation

Figure 6 shows the number of metal responses for each evaluation stimulus in the cases of no adaptation (Expt.1), metal sound adaptation (Expt.3A), and non-metal sound adaptation (Expt.3B). The yellow line shows the result of no adaptation (Expt.1), the blue line shows the result of metallic sound adaptation (Expt.3A), and the red line shows the result of non-metallic sound adaptation (Expt.3B). Compared with no adaptation, the number of metal responses increased in metal sound adaptation. On the other hand, in non-metallic sound adaptation, the number of metal responses was almost the same as no adaptation. When performing t-test at a significance level of 5%, a significant difference was found only between no adaptation and metal sound adaptation. In other words, in auditory adaptation using a metallic sound, the perception of metallic feeling was influenced in the forward direction by the adaptation stimulus. In the non-metal image adaptation, we did not observe any adaptation effect, probably because the auditory stimulus did not match the visual stimulus. Besides, 3 out of 10 participants reported that non-metallic sounds were strange; they are not like real sounds. As a result, the auditory adaptation effect is similar to the priming effect [10].

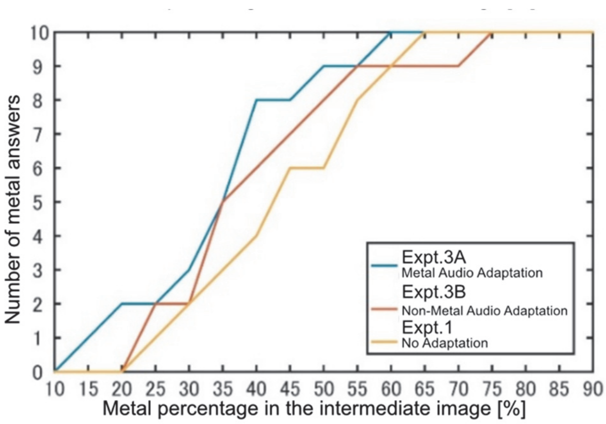


Figure 6. *No Adaptation VS Audio Adaptation*

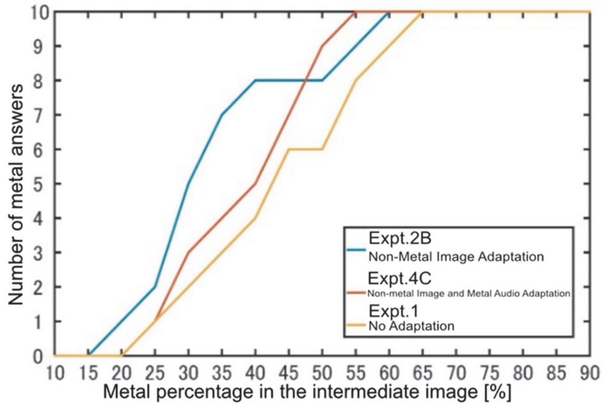


Figure 7. *Non-Metallic Image Adaptation VS Non-Metallic Image and Metallic Audio Adaptation*

Results and Discussions for Audiovisual Adaptation

We only described representative results in this section. Figure 7 shows the number of metal responses to each evaluation stimulus in the non-metal image adaptation (Expt.2B) and non-metal image and metal sound adaptation (Expt.4C). The yellow line is the result of no adaptation (Expt.1), the blue line is the result of non-metal image adaptation (Expt.2B), and the red line is the result of non-metal image and metal sound adaptation (Expt.4C). In both non-metal image adaptation and metal sound adaptation, the number of metal answers increased. Therefore, if there is an interaction between the visual sense and the auditory sense, non-metal image/metal sound adaptation, which is a combination of them, should double the metal answer. However, as shown in Fig. 7, the number of metal responses in non-metal image adaptation and non-metal image/metal sound adaptation were almost the same. Besides, when performing t-test at a significance level of 5%, we did not find a significant difference between non-metal image adaptation and non-metal image/metal sound adaptation. Therefore, there is no interaction between sight and hearing in audiovisual adaptation.

Logistic Regression Analysis

Analysis Method

In the previous section, we examined the effects of visual, auditory, and audiovisual adaptation on the perception of metallic feeling based on the size of the number of metallic responses and the results of the t-test. This section provides a more robust consideration of the visual, auditory, and audiovisual adaptation effects in the perception of metallic feeling by statistical analysis using logistic regression analysis. We also considered the perception of metallic feeling in the difference of image shape.

Using logistic regression analysis (Equations 1 to 6) based on Bayesian statistics, we examined how visual adaptation, auditory adaptation, and image shape affect the participants' evaluation. This method is useful in many analysis and used a lot. (Visual adaptation had three conditions: no adaptation, metal image adaptation, non-metal image adaptation. Auditory adaptation had three states: no adaptation, metallic sound adaptation, non-metallic sound adaptation. Image shape had two types: buddha and sphere.) Bayesian statistics consider all parameters as random variables and assume a probability distribution. Unlike conventional statistics (such as t-tests), the probability that a parameter value is in the interval [a, b] is 95% [11].

Each answer (R_i) followed Bernoulli distribution due to two choices, metallic or non-metallic (Equation 1). We set the probability of answering "metal" to μ_i, and using a logistic function, converted μ_i within the range (0,1) (Equation 2). The metal response y_i was an analysis of variance model that included the three factors (visual adaptation factor (V), auditory adaptation factor (A), and image factor (S)) and all these interactions. (Equation 3).

The coefficient b was a fixed effect, and the superscript indicates the main effect, the second-order communication, and the third-order interaction. For example, b_j^v represents the effect of visual adaptation, j represents metallic images and non-metallic images. $b_{j,k}^{v\times a}$ represents the interaction between visual adaptation and auditory adaptation. This model included the individual differences (r^{subj}) of the participants and the mixture ratio (r^{pic}) of the images as random effects.

We used the Rstan package for parameter estimation. The prior distribution of fixed effect was a normal distribution with mean 0 and standard deviation 10 (Equation 4). The prior distribution of random effects was a normal distribution with mean 0 and standard deviation σ_r (Equation 5). Standard deviation σ_r was assumed to follow the gamma distribution (α=10, β=10) (Equation 6). Since each model was set to the default stan function, the number of chains was 4, the number of thins was 1, the number of iteration steps was 2000, and the number of warmup steps was 1000. The total number of MCMC (Markov chain Monte Carlo) samples obtained was 4000.

$$R_i \sim Bern(\mu_i) \tag{1}$$

$$\mu_i = \frac{1}{1+exp(-y_i)} \tag{2}$$

$$y_i = b_0 + \sum_j b_j^v V_{ij} + \sum_k b_k^a A_{ik} + b^s S_i + \sum_{j,k} b_{j,k}^{v\times a} V_{ij} A_{ik} + \sum_j b_j^{v\times s} V_{ij} S_i + \sum_k b_k^{a\times s} A_{ik} S_i + \sum_{j,k} b_{j,k}^{v\times a\times s} V_{ij} A_{ik} S_i + r_{i,l}^{subj} + r_{i,m}^{pic} \tag{3}$$

$$r \sim Normal(\mu=0,\sigma_r) \tag{4}$$

$$\sigma_r \sim Gamma(\alpha=10,\beta=10) \tag{5}$$

$$b \sim Normal(\mu=0,\sigma=10) \tag{6}$$

Analysis Result

First, we checked Rat ($R^\wedge$), which was one of the convergence judgment indices, to see if the MCMC sample had converged. Rhat was the variance of the parameters of each MCMC sample calculated from the execution results of each chain by Rstan. We considered that "the number of chains was three or more and Rhat <1.1 for all parameters" had converged [12]. As a result of the analysis, Rhat was less than 1.1 for all the coefficients. Therefore, we confirmed that the MCMC samples had converged.

Table 2 shows the coefficients of the logistic regression model that do not include 0 in the 95%, which is the highest density interval (HDI), of the posterior distribution. The most significant feature of the most upper-density range is that it covers the most reliable part of the distribution and the most significant part of the distribution. In other words, it is more reliable than the confidence intervals that are generally assumed to be asymmetric normal distribution [13]. Here, $b^v_{j=NM}$ was a coefficient showing the effect of non-metal image adaptation, $b^v_{j=M}$ was a coefficient indicating the impact of metal image adaptation, and $b^a_{k=M}$ was a coefficient showing the impact of metal sound adaptation. As a result of analyzing, the 95% HDI of the posterior distribution of the coefficients of all second-order interactions and third-order interactions included 0. Thus, we did not find any significant effect. Also, there was no significant difference in the shapes of the two images. On the other hand, from Table 2.3, the value of $b^v_{j=NM}$ was positive. Thus, when adapting to a non-metal image, the rate of responding "the evaluation stimulus were metal" was significantly higher than in the condition without adaptation. Besides, the value of $b^v_{j=M}$ was negative. In other words, when adapting to a metal image, the rate of responding "the evaluation stimulus was metal" was significantly lower than in the condition without adaptation. Furthermore, the value of $b^a_{k=M}$ was positive. In other words, when adapting to metallic sounds, the rate of responding "the evaluation stimulus was metallic" was significantly higher than in the condition without adaptation. However, when adapting to non-metallic sounds, no effect was detected.

Table2. Coefficients those of 95% HDI does not include 0

Coefficient	Average	Mode (MAP)	95%HDI Minimum	95%HDI Maximum
$b^v_{j=NM}$ *Non-Metal Image Adaptation*	0.160	0.151	0.010	0.315
$b^v_{j=M}$ *Metal Image Adaptation*	-0.491	-0.485	-0.674	-0.309
$b^a_{k=M}$ *Metal Audio Adaptation*	0.166	0.165	0.013	0.317

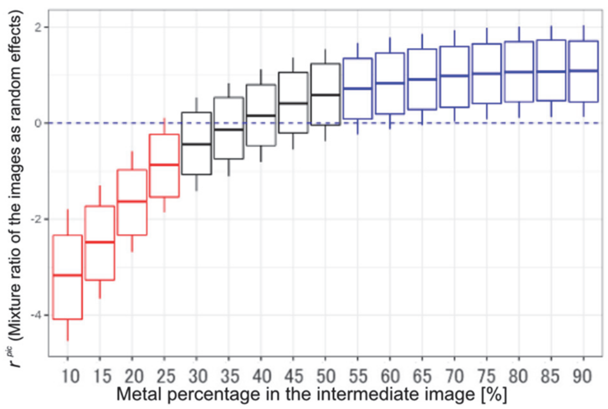


Figure 8. *The effect of mixing ratio of evaluation stimuli*

Figure 8 shows the effect of the mixing ratio of the evaluation stimuli (mixing ratio of metal and non-metal). The horizontal axis shows the metal percentage in the intermediate image. The vertical axis shows the value of r^{pic}. The images (red boxes), which metal percentage is 10% to 25%, show the evaluation stimulus that the probability of answering as metal was significantly low. Besides, the images (blue boxes), which metal percentage is 55% to 90%, show the evaluation stimulus that the probability of answering as metal was significantly higher.

Discussions for Analysis Result

From the analysis results, in visual adaptation and auditory adaptation, we confirmed the adaptation effect only when using metallic sounds. Specifically, in visual adaptation, when compared to non-adaptation conditions, non-metal responses increased in metal image adaptation, and metal responses increased in non-metal image adaptation. In other words, in visual adaptation, the perception of metallic feeling was affected in the opposite direction to the adaptation stimulus. On the other hand, in auditory adaptation, when compared to non-adaptation conditions, metal responses increased only in metal sound adaptation. In other words, in the auditory adaptation of metallic sounds, the perception of metallic feeling was influenced in the forward direction to the adaptation stimulus.

We found no significant effect in the secondary and tertiary interactions of all (visual adaptation factor (V), auditory adaptation factor (A), image factor (S)). Therefore, in audiovisual adaptation, there was no synergistic effect between vision and hearing, and no interaction occurred. Furthermore, there was no significant difference in the shapes of the two images. Therefore, the shapes of the image did not concern the adaptation effect of the perception of metallic feeling.

In summary, we could explain the adaptation effect in the perception of metallic feeling only for the visual adaptation effect and the auditory (metal sound) adaptation effect. That is, the metal answer y_i would be expressed as in Equation 7. Besides, from Equation 7, we found that the audiovisual adaptation effect can be expressed by the linear sum of the visual adaptation effect and the auditory (metal sound) adaptation effect.

$$y_i = b_0 + \sum_j b_j^v V_{ij} + b_M^a A_{iM} + r_{i,l}^{subj} + r_{i,m}^{pic} \quad (7)$$

Conclusions

In this study, we conducted a subjective evaluation experiment using CG images and sounds to investigate the effect of visual and auditory adaptation effects on material appearance. The target of this study was a metallic feeling. From the experimental results, we confirmed that the visual and auditory adaptation effects on the perception of metallic feeling. In visual adaptation, the perception of metallic feeling was affected in the direction opposite to the material of adaptation stimulus. In auditory adaptation, only in the case of metallic sounds, the perception was affected in the same direction. Besides, in audiovisual adaptation, no interaction between visual and auditory was confirmed. These results were similar to the conventional color aftereffects and auditory priming effects. Besides, by logistic regression analysis based on Bayesian statistics, we expressed the adaptation effect of audiovisual by the linear sum of the adaptation effect of visual and auditory (metal sound).

As future work, it is necessary to record the sound of stimulus actually or use real sound data set. We used free synthetic sound effects collected from the Internet in this research, but some participants reported that sounds were not like real ones. Moreover, our target was the static CG image. It is necessary to verify whether the adaptation effect also occurs for dynamic CG. For example, it would be practical to use stimulus with animation with sound in audiovisual adaptation (like hitting an object with a stick). Furthermore, by verifying different shapes and other material appearance factors such as transparency and glossiness, more robust results would be obtained regarding the adaptation effect of visual and auditory perception in material appearance.

References

[1] P. Fabio, J. A. Ferwerda, and D. P. Greenberg, “Toward a psychophysically-based light reflection model for image synthesis”, Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp.55-64, New Orleans, USA, July, 2000.

[2] R. W. Fleming, R. O. Dror, and E. H. Adelson, Real-world illumination and the perception of surface reflectance properties, Journal of Vision, vol.3, no.5 July, 2003.

[3] P. Vangorp, J. Laurijssen, and P. Dutré, The influence of shape on the perception of material reflectance, ACM Transactions on Graphics, vol.26, no.77, July, 2007.

[4] L. Sharan, R. Rosenholtz, and E. Adelson, Material perception: What can you see in a brief glance?, Journal of Vision, vol.9, no.8, August, 2009.

[5] W.Fujisaki, N.Goda, I.Motoyoshi, H.Komatsu, and S.Nishida, Audiovisual integration in the human perception of materials, Journal of Vision, vol.14, no.4, 2014.

[6] W. Fujisaki, M. Tokita, and K.Kariya, Perception of the material properties of wood based on vision, audition, and touch, Vision Research, vol.19, Part B, pp.185-200, April, 2015.

[7] M. Rodrigo, I. Julian, W. Michael, and H. Matthias, "Multimodal Perception of Material Properties, " Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, pp.33-40, Tübingen, Germany, September, 2015.

[8] I. Motoyoshi, Visual aftereffects in 3D shape and material of a single object, Journal of Vision, vol.12, no.9, August, 2012.

[9] I. Motoyoshi, Visual aftereffects in natural object categories, Journal of Vision, vol.13, no.9, July, 2013

[10] D.Schacter, I.Dobbins, D.Schnyer, Specificity of priming: a cognitive neuroscience perspective, Neuroscience 5, 853-862 November, 2014

[11] K. Kruschke, Doing Bayesian data analysis: A tutorial with R, JGAS, and Stan (2nd ed), MA: Academic Press/Elsevier, Burlington , 2014

[12] K.Matsura, Bayesian statistical modeling with Stan and R, M.Ishida, Kyoritsu Publishing, 2016

[13] I.Motoyoshi, Visual perception and image statistics, Cognitive Studies, vol.21, no.3, pp304-313, September, 2014

Author Biography

Takumi Nakamura received the B.E. and Master degrees from Chiba University in March 2018 and March 2020 respectively His research interests are material perception and imaging technologies. Especially, he focused on cross-modal material perception using CG.

Daichi Yagi, presenter of this research at CIC, received the B.E. degree from Chiba University in March 2020. He is currently a student of a master's course program in the Department of Imaging Sciences, Chiba University. His research interests are material appearance measurement and 3D imaging technologies. Especially, he is now constructing a material appearance reconstruction method based on imaging technologies.