PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Influence of Full-Body Vibration Adapted to Foreground Components on High-Level Perception of Reality

Shota Abe[1]; Zhenglie Cui[2]; Shuichi Sakamoto[3], Yôiti Suzuki[4], and Jiro Gyoba[5]

[1] Graduate School of Information Sciences, Tohoku University.

[1,2,3,4] Research Institute of Electrical Communication, Tohoku University.

[5] Graduate School of Arts and Letters, Tohoku University

ABSTRACT

To develop advanced multimedia communications systems, understanding how humans perceive reality from the media presented by the systems is important. Various indexes can be used to evaluate the sense of reality. While the sense of presence is an evaluation index that includes background information, the sense of verisimilitude focuses on foreground information. By adding appropriate sensory information related to the foreground or background components, these perceptual realities can be enhanced. Based on this concept, we investigated the effect of vibration information added to audio-visual content. In this study, vibration information was generated from a sound that included rich information about the content. We generated nine vibration types by adjusting the cutoff frequency and the carrier frequency of the sound. The results showed that higher verisimilitude was observed when vibration closely connected to foreground components was added to a scene. Moreover, under this condition, the sense of presence was hardly affected even when the vibration was added to the content. These results suggest that enhancing realism is possible by artificially generating vibrations from sound if the auditory signal is appropriately processed.

Keywords: Multimodal perception, Perceived reality, Full-body vibration, Sense of presence, Sense of verisimilitude

1. INTRODUCTION

With the advancement of information technology during recent years, there is an increased demand for advanced multimedia systems that enable users to easily share a “sense of being in another space” and “sense of having a true experience.” To realize such advanced systems, not only the effect of audiovisual information, but also that of other sensory information, such as touch, smell, and vibration, should be considered and effectively used. Thus, it is important to understand how humans perceive reality from multisensory information.

[1] shota.abe.r4@dc.tohoku.ac.jp

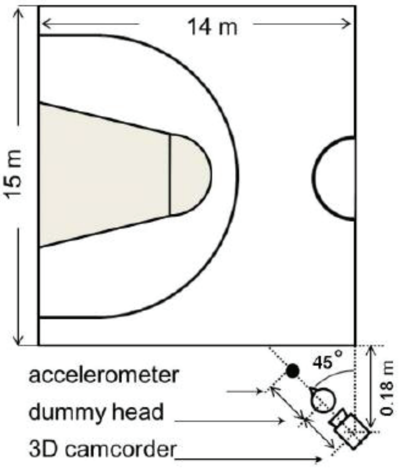

Figure 1 – The recording scenery. Figure 2 – Recording environment.

We investigated the contribution of full-body vibration on the perceived reality of multisensory content [1,2]. In our previous study, we examined how the vibration intensity affects perceived reality from multisensory content [3]. As perceived reality indexes, the senses of presence and verisimilitude [4,5] were measured. The sense of presence is defined as the subjective experience of being in one place or environment when one is physically situated in another place. Because this sense mainly concerns the place where the observers are present, the sense of presence is an evaluation index that includes background information. On the other hand, the sense of verisimilitude is defined as the trueness of appearance and quality of the presented objects; it is an evaluation index that focuses on foreground information. Results have showed that the sense of presence monotonically increases with an increase in vibration intensity up to a value which is much higher than the natural level. This tendency is also observed for other physical parameters; e.g., view angle and sound pressure level [6,7]. In contrast, the level of the sense of verisimilitude saturates at an intensity near the natural value. Thus, these results indicated that the senses of presence and verisimilitude have different characteristics, suggesting that enhancing both senses would be important for creating a natural and realistic virtual reality environment.

Although full-body vibration is useful to enhance the perceived reality, it is difficult to obtain appropriate vibration information. Nearly all media content which people can easily obtain includes only audio-visual information. Moreover, to measure actual vibration information of the scene, special equipment has to be installed at the recording environment. However, it is quite difficult to measure actual vibration information at the contents. Therefore, in order to add effective vibration information to the contents, it is necessary to create vibration information from other sensory information. Because sound information is closely related to vibration information, some researchers have proposed methods of generating vibration from sound [9, 10]. One of these methods is to extract low frequency components of the audio signal as vibration information. The result showed that generated vibration can enhance the perceived reality. However, it is still unknown which characteristics of generated vibration strongly affect perceived reality. It is likely that the effect of vibration characteristics on the sense of presence is different from that on the sense of verisimilitude because the senses of presence and verisimilitude have different perceptual characteristics. Previous studies have suggested that the temporal characteristic of amplitude and the frequency of generated vibration affect perceived reality [9,11]. Therefore, in this study, we investigated how these parameters are related to the senses of presence and verisimilitude perceived from the content.

2. Experiment

2.1 Stimulus

As an experimental stimulus, a moving picture 180 s in duration was cut from recorded material consisting of high-definition(HD) video, binaural stereo sound, and vibrations of a basketball game scene (3-on-3 play). Fig.1 shows a scene of the recording. Full HD video cameras (AG-3DA1, Panasonic Inc.) a dummy head (SAMRAI, Koken Co., Ltd.), and a vibration meter using a

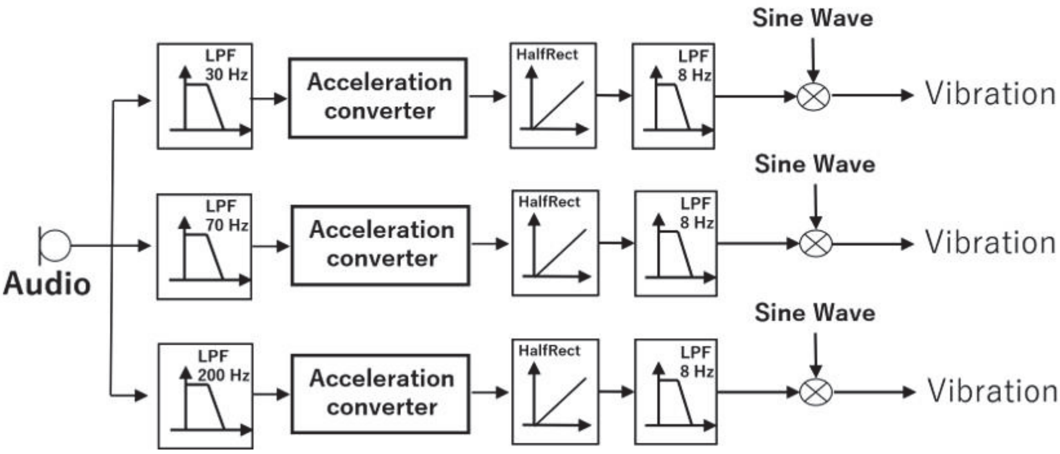

Figure 3: Schematic diagram of signal processing

piezoelectric accelerometer (VM-80, RION Co., Ltd.) were set up at the recording position as recording equipment. Fig. 2 shows the schema of the recording environment, including the location of the audio-visual recording equipment and accelerometer. In the figure, the size and location of the game area are drawn. The three-dimensional(3D) spatial sound was binaurally recorded via two microphones (4101, Brüel & Kjær) inside the two ears of the dummy head. The outputs of the binaural microphones were connected to the audio inputs of the digital video camera via an amplifier (2639, Brüel & Kjær). The sampling frequency of the binaural signals was 48 kHz. The quantization resolution was 16 bit. The accelerometer was fixed on the floor. The accelerometer and microphones outputs were connected to an analog-to-digital (AD) converter (ONOSOKKI, DS-0264). The vibration and sound signals were synchronously recorded to a PC. With this signal, the recorded vibration and audio-visual content was synchronized. After the recording, we edited the video, sound, and ground vibration to create a multisensory content to be used as the stimulus.

For the vibration stimulus during the experiment, we used the signal at frequencies lower than 70 Hz, because of the upper limit of the frequency characteristics, to reflect human sensitivity for the vibration defined in ISO 2631-1 (1997) [12], as well as the playback performance of the motion platform (D-BOX, D-BOX Mastering Motion). To produce the vibration information used for the "original vibration condition" as a reference condition, the sensed-vibration signal was presented via the motion platform after a low-pass filter (FIR filter, 128th-order) at a cutoff frequency of 70 Hz was applied.

To use the temporal characteristic of amplitude as the parameter, the cutoff frequency of the low pass filters was changed when artificially generated vibrations were converted from the auditory signal. We generated the vibration conditions via the process shown in Fig.3. First, the binaural signal recorded by the dummy head was averaged to obtain a monophonic signal. Low-pass filters (FIR filter, 128th-order) at cutoff frequencies of 30, 70, and 200 Hz were applied to this signal. These synthesized signals were differentiated twice to convert to vibration acceleration. To extract the temporal fluctuation of the amplitude of the synthesized vibration acceleration, an 8 Hz low-pass filter (FIR filter, 128th-order) was applied to the acceleration after applying half-wave rectification. As a carrier signal of the vibration acceleration, sine waves at a frequency of 20, 40, or 60 Hz were used and multiplied to the extracted vibration acceleration envelope. The synthesized vibrations levels were equalized to those of the original vibration in terms of the vibration level, resulting in vibration levels of 84.5 dB for all conditions.

The nine conditions previously described are referred to as the Fc30 Cf20, Fc30 Cf40, Fc30 Cf60, Fc70 Cf20, Fc70 Cf40, Fc70 Cf60, Fc200 Cf20, Fc200 Cf40, and Fc200 Cf60 conditions. In addition to these nine conditions, the original condition was added as a reference condition. Consequently, the experiment was conducted under 10 conditions in total.

2.2 Observers

In this experiment, 18 healthy adults with normal or corrected-normal vision and normal hearing (10 male, 8 female, mean age = 21.8 years, *SD* = 1.6) participated. They were divided into two groups

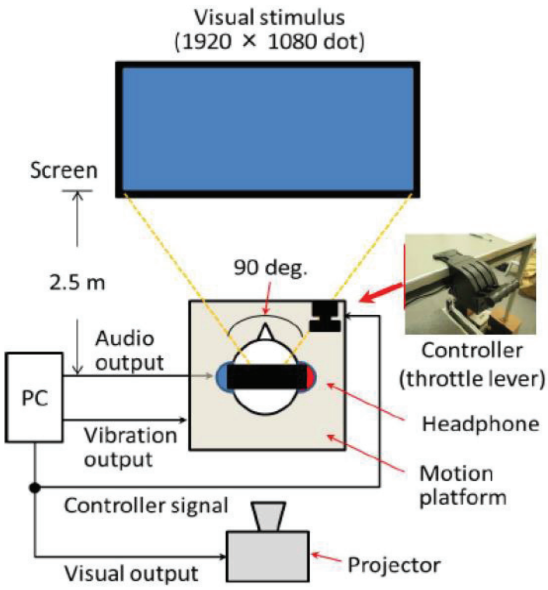

Figure 4：Experimental setup

one to judge the sense of presence and the other to judge the sense of verisimilitude. Both the sense-of-presence group and the sense-of-verisimilitude group consisted of five males and four females. All observers were right-handed.

2.3 Experimental Setup

Fig.4 shows the experimental setup. All experiments were conducted in a soundproof room. The observers were asked to stand directly on the motion platform and observe the video. The visual stimulus (resolution: 1920 × 1080 pixels, frame rate: 30 fps) was presented using a Digital Light Processing (DLP) projector (PDG-DHT100JL, SANYO Co. Ltd.) on a screen (Stewart Sound Screen) set 2.5 m in front of the observer. The field of horizontal view was 90 degrees. The audio stimulus was presented at the same sound pressure level, that is equal to that at the time of measurement, binaurally via headphones (HDA-200, Sennheiser Electronic). The equivalent A-weighted sound pressure level(L_{Aeq}) was approximately 60 dB. The full-body vibration stimulus was provided via a motion platform (D-BOX Mastering Motion). Only one degree of freedom (1DOF) of vibration (in the perpendicular direction) was presented during the experiment. In addition, a throttle lever controller (Throttle Quadrant, SAITEK) was attached on the right-hand side of the observers. The controller smoothly moved only in the vertical direction, and it was also possible to hold the lever in a stationary position. The movable range was from 0 to 90 degrees and the minimum movable angle was 0.5 degrees.

2.4 Experimental Procedure

As previously mentioned, we set 10 vibration conditions: the lowpass filtered (Fc = 70 Hz) measured vibration (original vibration condition) and nine artificially generated vibration conditions**.** The presentation of these 10 conditions was randomized for each observer. To avoid misunderstanding of the senses of presence and verisimilitude, we showed the observers the definitions of these terms as follows:

Sense of presence: the sensation of feeling like you are there

Sense of verisimilitude: the sensation of feeling that what you see looks real

These definitions were similar to those provided to the observers in our earlier studies [4]. To investigate the temporal characteristic of the two senses, we asked the observers to continuously report their perceived magnitude of presence or verisimilitude, depending on the group that the observers belonged to, over the entire presentation, with the flow. as follows: At the beginning, the fixation point was shown on the screen set in front of the observers. Thereafter, one of the 10 experimental stimuli was presented for 180 seconds. Observers were asked to quantify the sense of presence or verisimilitude in real time, using the throttle lever. Observers were asked to relate the maximum angle

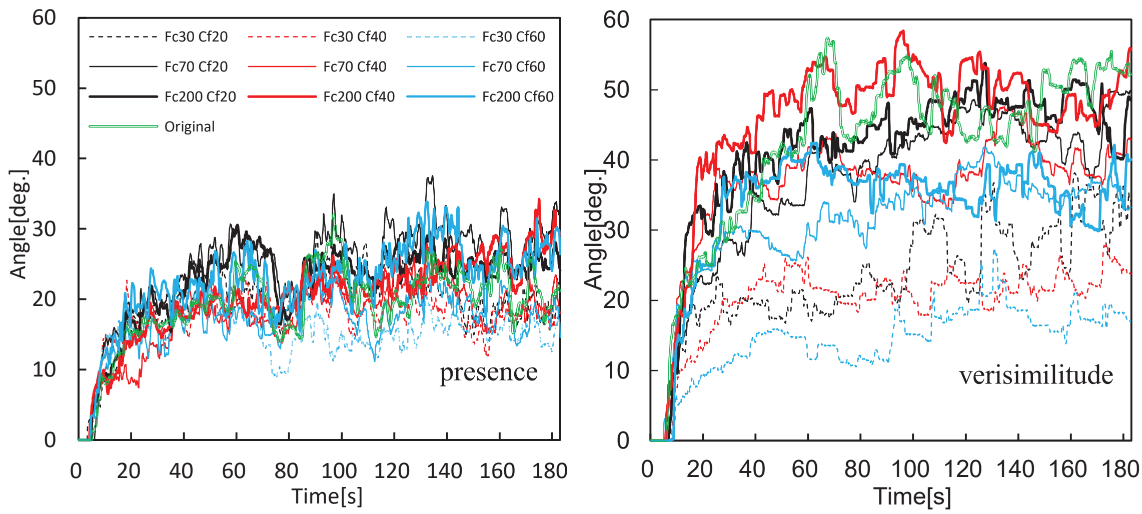

Figure 5: Magnitude change in each sense

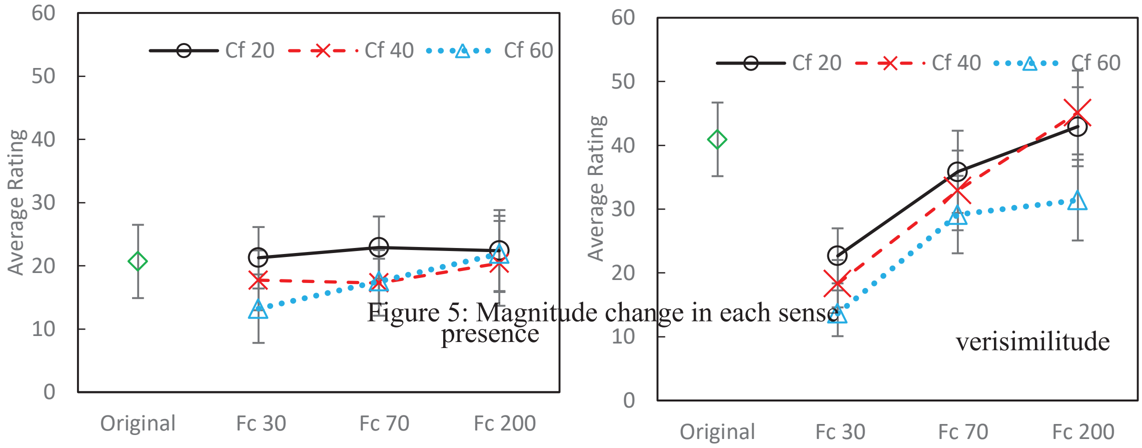

Figure 6: Average evaluation scores during the presentation

to the maximum quantity of presence/verisimilitude that observers experienced in daily life, and the minimum angle to the condition in which observers did not feel any sense of presence/verisimilitude at all. After each trial, observers were also requested to rate the overall sense of presence or verisimilitude on a scale from 0 (low) to 6 (high). Each observer participated in only one session, rating either the sense of presence or the sense of verisimilitude (1 × 10 vibration conditions), and each vibration condition was randomly presented for each observer.

3. Results

3.1 Change in each sense of reality over time

Fig. 5 shows the mean values of the senses of presence and verisimilitude as a function of time for all observers for each of the stimuli. Although there were small differences in the evaluation scores among all the conditions for the sense of presence, the differences in the evaluation scores for the sense of verisimilitude were large among the various conditions. Fig.6 shows the average evaluation scores during one stimulus presentation for each of the ten vibration conditions for the two senses. Comparing the difference in the evaluation scores between each condition, while there was a large difference for the sense of verisimilitude, there was a smaller difference for the sense of presence than that for the sense of verisimilitude. For the sense of presence, the evaluation scores of the Fc200 and Cf20 conditions scored as high as the original condition. On the other hand, for the sense of verisimilitude, only two artificially generated vibration conditions scored as high as the original condition. Moreover, the evaluation scores of all Fc conditions decreased along with a decrease in the cutoff frequency.

To analyze the relationship between the cutoff frequency and the carrier frequency of the artificially generated vibration, the two-way ANOVA was separately conducted for the sense of

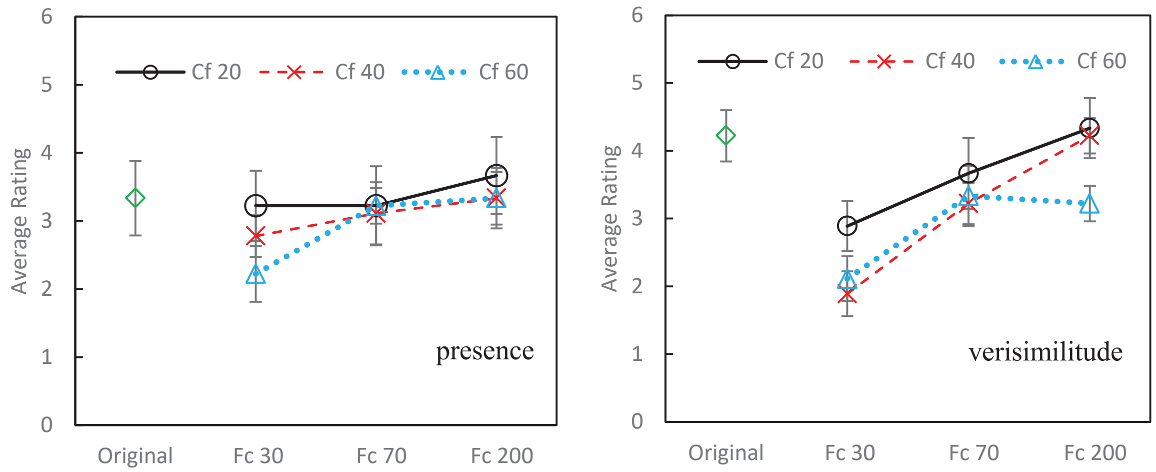

Figure 7 : Total impression of the content after viewing

presence and the sense of verisimilitude with the three Fc conditions and the three Cf conditions as the factor. Results for the sense of presence show no significant differences in the main effects of the Cf and Fc conditions and interaction. On the other hand, a significant difference ($F_{(8,\ 16)} = 15.9$, $p < .005$) was observed in the main effects of the Fc conditions for the sense of verisimilitude, while there was no significant difference in the main effects of Cf conditions and interaction. The results of the multiple comparisons (Ryan's method, $p < .05$) on the main effects of the Fc conditions showed that the evaluation scores of the Fc70 and Fc200 conditions are statistically significantly higher than those of the Fc30 conditions for the sense of verisimilitude. These results indicate that the temporal characteristic of amplitude and frequency of the artificially generated vibration did not appreciably affect the evaluation score of the sense of presence. However, for the sense of verisimilitude, the temporal characteristics of amplitude affected the evaluation score and it is necessary to choose an appropriate cutoff frequency when vibration is generated artificially from sound.

3.2 Overall evaluation of each sense after viewing

Fig. 7 shows the overall evaluation scores in terms of the sense of presence and verisimilitude after viewing the stimulus. The evaluated scores trends for the sense of presence and verisimilitude, which occur as cutoff frequency increases, are similar to those observed for the evaluation scores during the presentation (Fig. 6). To analyze the relationship between the cutoff frequency and carrier frequency of the generated vibration, the two-way ANOVA was separately conducted for the sense of presence and the sense of verisimilitude with the three Fc conditions and the three Cf conditions as the factors. Results for the sense of presence show no significant differences in the main effects of the Cf and Fc conditions and interaction, while a significant difference ($F_{(8,\ 16)} = 18.1$, $p < .005$) was observed in the main effects of the Fc conditions for the sense of verisimilitude. There was no significant difference in the main effects of the Cf conditions and interaction. The results of the multiple comparisons (Ryan's method, $p < .05$) on the main effects of the Fc conditions showed that the evaluation scores of Fc70 and Fc200 conditions are statistically significantly higher than those of the Fc30 conditions. These results indicate that the tendency of evaluation scores during the presentation remined similar to that of the overall scores after viewing.

4. Discussion

In this study, we investigated how the temporal characteristic of amplitude and the frequency spectrum of a generated vibration are related to the senses of presence and verisimilitude perceived from the content. The results showed that there is a significant difference between the Fc30 condition and the Fc70 and Fc200 conditions for the sense of verisimilitude. This result clearly shows that the temporal characteristic of amplitude should be considered when vibration is artificially generated from sound. Since the sense of verisimilitude is an evaluation index that focuses on foreground information, this result suggests that the vibration of the Fc30 condition might not have been appropriately presented with the foreground component in the scene. The main foreground component of this content is a basketball because the most important object in a scene is likely to attract attention [13]. The sound information from the bouncing basketball has a frequency above approximately 50 Hz. Thus, for the Fc30 conditions, there was no vibration information when the basketball was bouncing. Additionally, for the sense of verisimilitude, the evaluation scores of some Fc200 conditions seem to have exceeded those of the Fc70 conditions. Calculating the vibration levels of each Fc condition when the basketball was bouncing, there were differences of approximately 3 dB among the three Fc conditions. This may indicate that presenting a relatively larger vibration with the foreground component compared to that of the background component in the scene leads to higher evaluation scores for the sense of verisimilitude.

For the Fc200 Cf60 condition, although the temporal characteristic of amplitude is the same as under the Fc200 Cf20 and Fc200 Cf40 conditions, the evaluation scores seem to be lower than in the latter. These results showed the frequencies of the artificially generated vibration also influenced the sense of verisimilitude. The reason is that the carrier frequency at 60 Hz might not be inappropriate for this basketball content. Some observers reported a mismatch between the vibration and the scene when players in the content were running. Considering previous research showed that a carrier frequency of 20 Hz was unfit for concert content [10], there are a range of appropriate frequencies for each content.

For the sense of presence, there was smaller difference in the evaluation scores among the conditions than for the sense of verisimilitude. There are two reasons for this result. First, during this experiment, the vibration levels, which reflects the frequency characteristic of human sensitivity to full-body vibration, under all conditions were the same. According to the results of previous research, one characteristic of the sense of presence is that it monotonically increases with an increase in the intensities of presented sensory information. The vibration levels of all conditions were the same during this experiment; thus, there would be no significant differences in the evaluation scores among conditions. From this perspective, if the amplitude of the vibration increased, the sense of presence would be enhanced. However, it is unclear how a change in vibration amplitude related to foreground or background components in a scene affects the perceived reality. Second, the sense of presence is an index related to space and thus suing this index observers would mainly evaluate their environment as a whole. Considering the characteristics of evaluation for the sense of presence, observers might have evaluated the presented vibration regardless of the foreground and background information. As a result, they were not able to well detect the difference among conditions, and the evaluation scores among conditions might have fallen in much narrower ranges than those for the sense of verisimilitude.

5. Conclusions

We examined how the temporal characteristic of amplitude and the frequency of generated vibration are related to the senses of presence and verisimilitude perceived from content. In order to consider which characteristics of generated vibration strongly affect perceived reality, we prepared nine conditions by adjusting the cutoff frequency and the carrier frequency when the vibration was generated from sound. The results showed that while there were smaller differences in the evaluation scores among all the conditions for the sense of presence than that for sense of verisimilitude, a significant difference was observed between the Fc conditions for the sense of verisimilitude. These results suggest that for artificially enhancing the sense of presence from sound, it is unnecessary to consider the appropriate cutoff frequency or frequencies of artificially generated vibration. On the other hand, for the generation of vibration enhancing the sense of verisimilitude, the cutoff frequency and frequency spectrum of the generated vibration should be carefully considered.

ACKNOWLEDGEMENTS

A part of this work was supported by the National Institute of Information and Communications Technology (NICT) of Japan and Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research Grant Number 16K12506.

REFERENCES

1. S. Sakamoto, G. Hasegawa, A. Honda, Y. Iwaya, Y. Suzuki, J. Gyoba, Body vibration effects on perceived reality with multi-modal contents, ITE Transactions on Media Technology and Applications, 2014: 2(1), 46-50.
2. S. Sakamoto, Z. Cui, T. Ohtani, Y. Suzuki,. Gyoba, Effects of vibration information on the senses of presence and verisimilitude of audio–visual scenes, INTER-NOISE 2016, 1777-1782.
3. Z. Cui, S. Sakamoto, J. Gyoba, Y. Suzuki, Influence of Visual Depth and Vibration on the High-level Perception of Reality in 3D Contents, ISSN, Volume 8, Number 6, November 2017
4. W. Teramoto, K. Yoshida, S. Hidaka, N. Asai, J. Gyoba, S. Sakamoto, Y. Iwaya, Y. Suzuki, Spatio-temporal characteristics responsible for high "Vraisemblance," Transactions of the Virtual Reality Society of Japan, 2010: 15(3), 483-486.
5. A. Honda, T. Kanda, H. Shibata, N. Asai, W. Teramoto, S. Sakamoto, Y. Iwaya, J. Gyoba, Y. Suzuki, Determinants of senses of presence and verisimilitude in audio-visual contents, Transactions of the Virtual Reality Society of Japan, 2013: 18(1), 93-101.
6. J.D. Prothero, et al., Widening the Field-of-view Increases the sense of presence within immersive virtual environments, Human Interface Technology Laboratory Technical Report, R-95-4, Seattle: University of Washington, 1995.
7. W. Teramoto, et al.: "What is "sense of presence"? A nonresearcher's understanding of sense of presence", Journal of the Virtual Reality Society of Japan, 2010: 15(1), 7-16.
8. K. Ozawa, M. Miyasaka, Effects of reproduced sound pressure levels on auditory presence, Acoustical Science and Technology, 2004: 25(3), 207-209.
9. S. Merchel and M. E. Altinsoy, The influence of vibration on musical experience, Journal of the Audio Engineering Society, 2014: 62 (4), 230-234.
10. Zhenglie Cui, Hiroyuki Yagyu, Shuichi Sakamoto, Yôiti Suzuki, Jiro Gyoba, Effect of Full-Body Vibration Generated from Audio Signal on Perceived Reality. INTER-NOISE 2017, 3003-3010.
11. S. Merchel and M.E. Altinsoy, Vibratory and acoustical factors in multimodal reproduction of concert DVDs, in Haptic and Audio Interaction Design (Springer, Berlin, Germany), pp.119-127, 2009
12. International Organization for Standardization: Mechanical vibration and shock – Evaluation of human exposure to whole-body vibration – Part 1: General requirements, ISO 2631-1, 1997.
13. A.L. Yarbus. Eye movements and vision. NY:Plenum Press.