

Sense of Presence and Verisimilitude of Audio–Visual Contents: Effects of Sounds and Playback Speeds on Sports Video

Akio HONDA^{1,*}, Takayuki KANDA², Hiroshi SHIBATA³, Shuichi SAKAMOTO²,
Yukio IWAYA⁴, Jiro GYOBA⁵ and Yôiti SUZUKI²

¹Faculty of Humanities, Yamanashi Eiwa College, Kofu 400-8555, Japan

²Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan

³Faculty of Medical Science and Welfare, Tohoku Bunka Gakuen University, Sendai 981-0943, Japan

⁴Faculty of Engineering, Tohoku Gakuin University, Tagajo 985-8537, Japan

⁵Graduate School of Arts and Letters, Tohoku University, Sendai 980-8576, Japan

Received November 29, 2014; final version accepted May 11, 2015

Multimedia users all over the world hope to experience more natural and realistic audio–visual contents such as landscapes, music, and sports. This study investigated whether sounds and playback speeds of sports video (a golf swing scene) were able to modify viewers' perceived sense of presence and verisimilitude. Previous studies revealed different characteristics between the sense of presence and verisimilitude during perception of audio–visual contents. For this study, we recorded a video of the golf swing scene with a high-speed camera and manipulated the playback speeds and impact sound of the video for use as experimental stimuli. Results showed that the sense of presence was increased when the scene was presented at normal playback speed with or without sound conditions. In contrast, verisimilitude showed different patterns with respect to the playback speeds between these sound conditions. Our findings indicate that viewers' sense of presence and verisimilitude have different sensitivity to the contingent sounds and different temporal characteristics.

KEYWORDS: playback speed, sense of presence, verisimilitude, sound, slow motion effect

1. Introduction

Multimedia users all over the world hope to experience more natural and realistic audio–visual contents such as landscapes, music, and sports. Recent advancement of information and communications technologies has raised expectations for systems that provide highly natural and realistic experiences [1–3]. In fact, users can receive high perceptual reality from advanced multimedia systems such as Ultra High Definition Television (UHDTV) and Three-Dimensional Television (3DTV) [4]. Although recent progress of these technologies has been remarkable, it is also important to examine and clarify how *Kansei*¹ information is perceived from audio–visual contents. Findings from these investigations will encourage and support further technological innovation of information and communications technologies. From this viewpoint, our research group has been experimentally examining the determinants of sense of presence and verisimilitude in audio–visual contents of various types [4–6].

In *Kansei* studies, the sense of presence is the most commonly used index to evaluate audio–visual contents and virtual reality systems. The sense of presence is defined as the subjective experience of being in a given place or environment, even when one is physically situated in another [7]. Verisimilitude, another type of perceived reality index, is defined as the appearance of being true or real [4]. The sense of presence is related to the impression of a scene. Therefore, it naturally affects a person's evaluation of both a figure and the ground. Another evaluation measure based mainly on figure information is desired to advance multimedia systems and contents in harmony with human *Kansei* if one remembers that the sensory information related to the figure and information related to the ground are processed in the human brain in different manners and in different regions [8]. Therefore, the authors have examined verisimilitude and its characteristics: particularly the difference from presence. Teramoto *et al.* [5] first demonstrated that verisimilitude is distinguishable from the sense of presence. Moreover, it is deeply convolved with the aesthetic impression of a scene. In their experiments, audio–visual contents including *Shishi-odoshi* (a traditional Japanese tiny drum made of bamboos tapped by natural water drops) were used as stimuli. Their experiment 1 investigated the effects

*Corresponding author. E-mail: honda@yamanashi-eiwa.ac.jp

This work was supported by Ministry of Education, Culture, Sports, Science and Technology Grant-in-Aid for Specially Promoted Research (No. 19001004), a Grant-in-Aid for Scientific Research (A) (No. 24240016), and (B) (No. 26280078).

¹This Japanese word is sometimes translated as “affectiveness,” but it has wider and deeper meaning. Therefore, it is often used as an English word as it is here.

of the field size of view and the sound pressure level of the stimulus. Their results demonstrated that a higher degree of verisimilitude was observed with the middle field size of view with the original sound pressure level of the stimulus than with small and large view size, whereas a larger field size of view with a higher sound pressure level elicited a higher sense of presence. They also investigated the effect of temporal asynchrony between the audio–visual stimuli produced by *Shishi-odoshi* in their experiment 2. The results revealed that the range of the temporal window of the audio–visual stimuli necessary for high verisimilitude differed from that for high presence.

Honda *et al.* [6] investigated whether these results [5] could be more generalized. They conducted an experiment using a clip of classical orchestral music, which revealed that verisimilitude was more sensitive to audiovisual synchronicity than to the display size, whereas the sense of presence was more sensitive to the spatial size than to the temporal property. These findings well corresponded with those of the previous study [5], which differentiated verisimilitude from the sense of presence. Recently, Sakamoto *et al.* [4] examined the relation between full-body vibration amplitude and the sense of presence and verisimilitude in multi-modal contents. They demonstrated that the sense of presence increased monotonically according to the full-body vibration amplitude, whereas verisimilitude exhibited a saturating curvilinear tendency. All of these findings [4–6] indicate that the sense of presence and verisimilitude have different temporal characteristics.

This study investigated whether other temporal characteristics, i.e. sounds and playback speeds, can modify the viewers' perceived sense of presence and verisimilitude. Our group had already examined the sense of presence and verisimilitude with audio–visual contents related to landscape [5] and music [6]. The present study specifically examines a sports scene for the following reasons: Audio-visual contents of sports have often used “sound effects (e.g., expanded or contracted sounds)” and “playback speed effects (e.g., slow or quick motions).” However, few studies have experimentally demonstrated the effects of sounds and playback speeds on sports video. Some studies have examined the effects of playback speeds of educational materials on learners' understanding [9,10]. Their results revealed that viewers feel comfortable watching sports videos at slower-than-normal speeds [11], but such studies are unclear about the relation between viewers' perceived sense of presence and verisimilitude and the audio–visual playback speeds. Particularly, interactions between the sounds and the playback speeds have not been clarified in previous studies. In fact, it remains unclear whether the sounds corresponding to the playback speed have different effects on viewers' perceived sense of presence and verisimilitude.

This study investigated whether playback speeds and contingent sounds in a golf video can modify viewers' perceived sense of presence and verisimilitude. Golf is a widely enjoyed sports activity that is popular with all ages worldwide. The audio–visual contents of this sport often include golf impact scenes shown at various playback speeds (e.g., a slow motion scene of tee shot). For this study, we recorded a video of a golf swing scene with a high-speed camera and manipulated the playback speeds and impact sounds of the video to use these materials as experimental stimuli. Participants were asked to evaluate the sense of presence and verisimilitude for each stimulus. These procedures enabled us to measure the exact change of viewers' perceived sense of presence and verisimilitude with various sounds and playback speeds of the golf swing video. We predicted that the sense of presence and verisimilitude have different temporal characteristics with respect to the playback speeds and different sensitivity to the contained sounds.

2. Methods

2.1 Participants

In this experiment, 31 healthy adults with normal vision and audition (including two authors, TK and HS) participated. All the participants, who were affiliated with Tohoku University, were divided randomly into three groups: 11 participants (4 men, 7 women) in the expanded and contracted sound condition group, 10 participants (6 men, 4 women) in the normal sound condition group, and 10 participants (7 men, 3 women) in the no-sound condition group. We used the between-subject design to avoid exposure effects whereby repeated exposure to an object increases preference for that object [12].

2.2 Apparatus and stimuli

As the stimulus used for this study, we took a golf swing scene of a male Tohoku Fukushi University student hitting a tee shot at Izumi International Golf Club, located in Izumi-ku, Sendai, Miyagi Prefecture. We used a high-speed digital camera (MEMRECAM fix K4; NAC Image Technology Inc.) with 1280×720 pixel resolution and a frame rate of 1000 fps. Camera parameters such as the focal length and the camera height were chosen to enable comprehensive control of the object movement. Concretely, the camera was set 4.75 m from the object at 99.3 cm height. The focal lengths were set to 28 mm using a zoom lens (AI AF-S Zoom Nikon ED 28–70 mm F2.8D; Nikon Corp.). To record the impact of the golf club with a ball, we used a preamplifier (NEXUS Range of Conditioning Amplifiers Type 2690; Brüel & Kjær) and a microphone (4189; Brüel & Kjær). Part of the experimental stimulus is presented in Figure 1. The microphone was placed 1 m ahead and 95 cm left of the camera at 120-cm height. The sound was recorded as PCM data with sampling frequency of 48,000 Hz. The instantaneous peak sound pressure level of the impact sound was 103.6 dB.

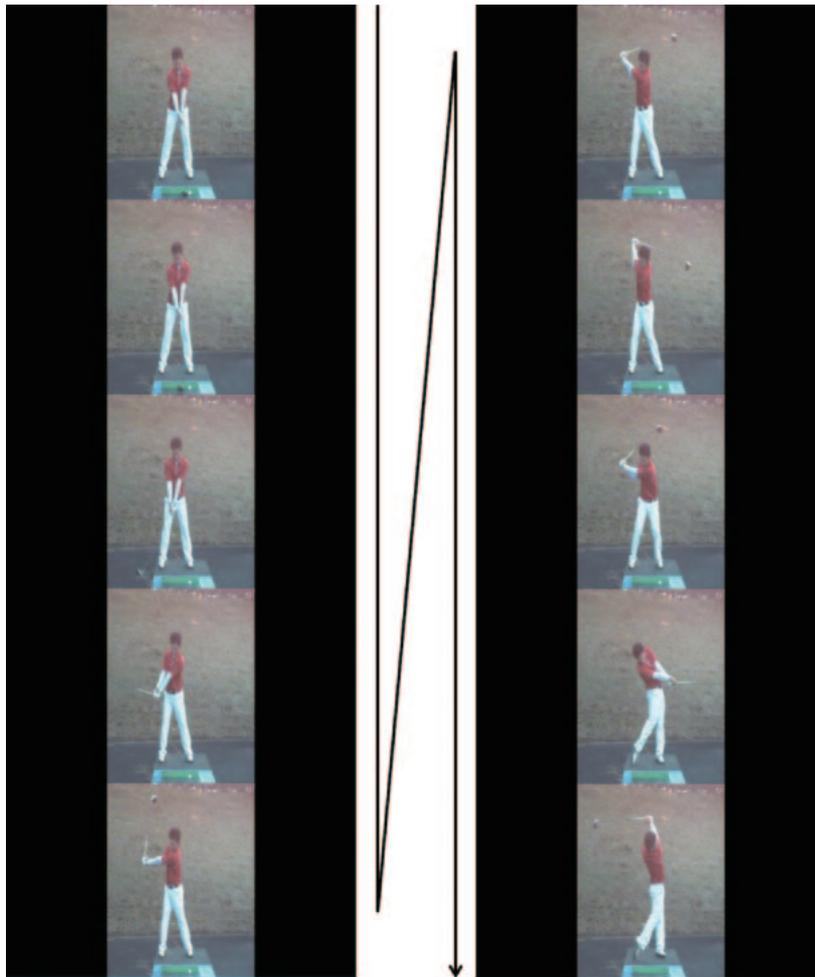


Fig. 1. Golf swing scene is used as the stimulus.

The video and sound signal were stored respectively as JPEG files and WAV files on different PCs. The video was image-processed using Matlab (the MathWorks Inc.). The video was taken as continuous still images into a video editor, Avid (Avid Technology Inc.) to produce a WMV file with 640×480 pixel resolution. To produce video files with different speeds, we selected some still images from the original sequence of 1,000 JPEG files that had been directly outputted by the high-speed digital camera. We presented them at the same frame rate (30 fps) to produce our experimental stimulus. According to this process, we produced seven videos with different playback speeds (seven playback speed factors: double-speed (2), normal (1), half-speed ($1/2$), quarter-speed ($1/4$), one-eighth ($1/8$), one-sixteenth ($1/16$), and one-thirty-second ($1/32$)).

In this study, we investigated not only the effects of playback speeds on the sports video, but also of sound. For expansion and contraction of sound signals, using the resample function of Matlab, we created and used the sound stimulus corresponding to the playback speed conditions. Then, in the expanded and contracted sound condition group, the playback speeds of video images and sound were synchronized under all conditions. In the normal sound condition group, the timings between the impact sounds and the impact scenes were synchronized. However, different from the former group, real time sound corresponded to each video image. In other words, in this group, the playback speeds between the video image and sound were only synchronized with the video image of the normal playback speed. The sound pressure levels of the stimulus corresponded to the recorded impact sound. In the no-sound condition group, the sound stimulus was not presented to participants.

2.3 Procedure

All experiments were conducted in a dark room to avoid possible undesirable effects. Participants were asked to wear headphones (HDA 200; Sennheiser Electronic GmbH & Co. KG) and to fix their head on the chin rest. Auditory information was presented via the headphones. The presented sound pressure level was the same as the recorded level. Visual information was presented on the CRT monitor (CPD-G520, 60 Hz; Sony Corp.). The frame rate of generated video files was 30 fps. To match the frame rate to with the refresh rate of the CRT monitor, all video files were presented by zero order hold. The observation distance was 73.7 cm. The visual angle was 18.4 deg. Participants were

asked to evaluate the sense of presence and verisimilitude of each trial using a visual analog scale from “not at all (0)” to “very strongly felt (6).” To avoid confusion and misunderstandings between the sense of presence and verisimilitude, we used instructions such as “sense of presence (you feel like being there)” and “verisimilitude (you feel that what you see looks real nature)” presented with the evaluation scale. A counterbalanced design was used to control for the possible effects of order of rating: half of participants were asked to rate the sense of presence before the verisimilitude for each trial, whereas the other half of participants were asked to evaluate the verisimilitude before the sense of presence for each trial. After practice trials, all participants performed four sessions of seven trials per session. The order of presentation of trials was randomized for each participant.

3. Results

Previous studies did not investigate whether the sound corresponding to the playback speed has different effects on the viewers’ perceived sense of presence and verisimilitude. To examine the difference between two sound conditions, we conducted two-way factorial analyses of variance by sound (2: normal sound condition, extended and contracted sound condition) \times playback speed (7: 2, 1, 1/2, 1/4, 1/8, 1/16, 1/32) on each average rating scale value of the sense of presence and verisimilitude. Results show that, for both the sense of presence and for verisimilitude, only the main effect of playback speed was significant (Presence: $F(6, 114) = 27.03, p < 0.001$; Verisimilitude: $F(6, 114) = 20.16, p < 0.001$). The main effect of sound and the interaction of sound and replay speed were not significant. Therefore, we combined the extended and contracted sound condition and the normal sound condition to use them as the sound condition. Figure 2 shows the average ratings between the sound condition and the no-sound condition.

We conducted a two-way factorial analysis of variance on sound (2: sound condition, no-sound condition) \times playback speed (7: 2, 1, 1/2, 1/4, 1/8, 1/16, 1/32) with dependent variables of the average rating scale value of the sense of

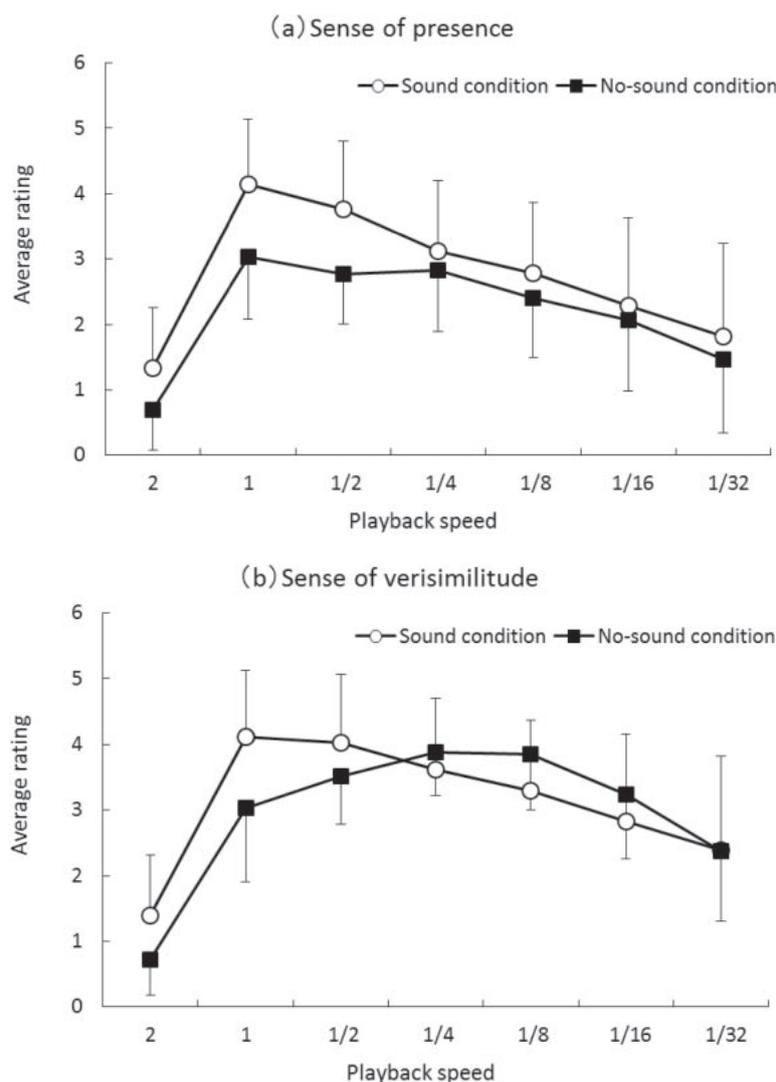


Fig. 2. Average rating scale value of the sense of presence and verisimilitude.

presence. The average ratings and standard deviations are presented in Figure 2a. A significant main effect of the playback speed was observed ($F(6, 174) = 33.97, p < 0.001$). Post-hoc analyses (Ryan's, $p < 0.05$) revealed that the sense of presence of the normal playback speed is higher than of other playback speeds ($ps < 0.05$, except for half-speed). In addition, the sense of presence of the twofold playback speed is lower than of other conditions ($ps < 0.05$). The sense of presence under the sound condition ($M = 2.75$) is slightly higher than that under the no-sound condition ($M = 2.18$), but the main effect of sound was not significant ($F(1, 29) = 3.69, p = 0.06$). Interaction of the sound and replay speed was not significant.

A two-way factorial analysis of variance of the average rating scale value of verisimilitude was performed similarly, considering the sound (2: sound condition, no-sound condition) and the playback speed (7: 2, 1, 1/2, 1/4, 1/8, 1/16, 1/32). The average ratings and standard deviations are presented in Figure 2b. The sound and playback speed interaction was significant ($F(6, 174) = 2.96, p < 0.01$). Simple main effect analysis revealed the effect of the sound condition under normal playback speed ($F(1, 203) = 7.62, p < 0.01$), the effect of playback speed condition under the no-sound condition ($F(6, 174) = 19.33, p < 0.001$), and the effect of playback speed under the sound condition ($F(6, 174) = 14.85, p < 0.001$). Simple main effects of playback speeds under the no-sound condition and the sound condition were apparent. Therefore, Ryan's multiple comparison tests ($p < 0.05$) were conducted. Results revealed that viewers' perceived verisimilitude without sound was increased when the playback was slightly slower than the normal playback speed (i.e., one-fourth speed and one-eighth speed) ($ps < 0.05$). Moreover, the results showed that the verisimilitude with sound became the highest at normal playback speed ($ps < 0.05$, except for half-speed and quarter-speed), which resembles the result for the sense of presence.

4. Discussion

Recent progress in information and communications technologies has been remarkable. Therefore, it is important to examine how we perceive *Kansei* information from audio–visual contents and how to provide more natural and realistic experiences to users. Our group has already examined the sense of presence and verisimilitude with audio–visual contents related to landscapes [5] and music [6], but has not yet studied them for sports. This study investigated whether the playback speeds and contingent sounds of sports video can modify the viewers' perceived sense of presence and verisimilitude because few studies have conducted psychometric examination of the effects of such parameters on viewers' *Kansei* for sports video. We recorded a video of a golf swing scene using a high-speed camera. Then we manipulated the playback speeds and impact sound of the video to use them as experimental stimuli. In the experiment, participants were asked to evaluate both the sense of presence and verisimilitude for the sports video. Results show that the sense of presence was highest at the normal playback speed, irrespective of the sound conditions, i.e., normal speed or expanded/compressed and present or absent. Moreover, verisimilitude was higher at normal replay speed under the conditions with sound and at a slightly slower playback speed under the condition without sound. Importantly, our results indicate that the sense of presence and verisimilitude have different temporal characteristics in visual-only contents (no-sound condition). In contrast, previous studies [4–6] have demonstrated that the sense of presence and verisimilitude have different temporal characteristics in audio–visual contents. Therefore, our results did not correspond well with those findings [4–6], but provided additional findings showing that the sense of presence and verisimilitude exhibit different sensitivity to the contingent sounds and the playback speed. The discordance in findings between these studies might be related to the unnaturalness resulting from playback speed effects in audio–visual contents. However, few studies have examined the viewers' *Kansei* in sports video [11]; even fewer have explored the adequate sound corresponding to the playback speed. Future studies must investigate the interactions between sounds and playback speeds for various audio–visual contents. Moreover, the following two points should be considered.

First, we clarified that the participants felt higher levels of the sense of presence under conditions corresponding to real space (with presence of sound) and real time course (normal playback speed). The sense of presence is defined as the subjective experience of being in a given place or environment, even when one is physically situated in another [7]. Our results corresponded well with those of previous studies [4–6]. The sense of presence exhibited accurate correspondence to spatiotemporal reality among others. To increase the sense of presence further, our previous results [4–6] demonstrated that it is effective to increase the physical intensity of audio–visual contents (e.g., with a larger field size of view and a larger sound pressure level in more realistic spatiotemporal conditions).

A second noteworthy result is that the verisimilitude varies according to the playback speed between the sound condition and the no-sound condition. This phenomenon has not been examined systematically in previous studies. Regarding sports video, however, Ono and Sugawara [11] reported comparable results. In their experiment, participants were asked to evaluate their favorite playback speed and the smoothness of the object's movement using soccer and baseball monochrome video materials. Their results showed that the evaluation of favorability was highest at the one-eighth playback speed and that the smoothness of movement was also highest at the same playback speed. This experiment used a high-speed digital camera having 292×220 pixel resolution with the maximum frame rate of 500 fps. This study, using no reference to the sound, is expected to be equivalent to the no-sound condition in our study. Therefore, although some technical and methodological differences exist between our study and their study [11], their

results, in line with ours, suggest that when sports videos were presented without sound and at a slower playback speed (one-fourth or one-eighth), viewers might feel “favorable” and sense “smoothness of movement.” In other words, their feeling of “being close to the essence” of sports might be intensified under such conditions. Consequently, the perception of smooth movement associated with slow motion is expected to work toward the improvement of the verisimilitude. Future studies must be conducted to confirm the observed “slow motion effect” conditions in more audio–visual contents of various kinds. Regarding the sports videos with sound, our results show that the sense of presence and verisimilitude were commonly highest at the normal playback speed. However, as we described above, the interactions between sounds and playback speeds in various audio–visual contents are not clear. Consequently, it remains a possibility that a newly developed sound can produce changes in the pattern of the viewers’ sense of presence and verisimilitude in audio–visual contents.

Despite the importance of our results, the findings should be interpreted in light of some limitations. The number of participants in our study was small. They might not be representative of the entire population of multimedia users. There also exist large differences between laboratory experimental environments such as those used in this study and viewing environments in daily life situations. Future studies must be conducted to investigate the characteristics of various age groups under more natural settings.

5. Conclusion

This study investigated whether playback speeds and contingent sounds in sports video can modify viewers’ perceived sense of presence and verisimilitude. We recorded a video of a golf swing scene using a high-speed camera and manipulated the playback speeds and impact sounds of the golf video. Previous studies [4–6] revealed that different characteristics exist, including a temporal characteristic on stimulus onset asynchrony, between the sense of presence and verisimilitude during perception of audio–visual contents. The results of this study demonstrated that the sense of presence was increased when the scene was presented at normal playback speed with or without sound conditions. In contrast, verisimilitude showed different patterns related to the playback speeds between these sound conditions. These findings indicate that viewers’ sense of presence and verisimilitude have different sensitivity to the contingent sounds and different temporal characteristics on the playback speed. Particularly, our results showed that viewers’ perceived verisimilitude without sound was increased when the playback speed was slower than the normal playback speed. This “slow motion effect” is similar to results obtained by Ono and Sugawara [11]. Future studies must be undertaken with various age groups under more realistic settings to confirm the observed “slow motion effect” in other audio–visual contents.

Acknowledgments

We appreciate valuable support from the golf club members of Tohoku Fukushi University, especially Mr. Shinji TOMIMURA, Mr. Ryosuke KAJII and Mr. Kikuto KAWASAKI. This research was supported by a Ministry of Education, Culture, Sports, Science and Technology Grant-in-Aid for Specially Promoted Research (No. 19001004), and Grants-in-Aid for Scientific Research (A) (No. 24240016), (B) (No. 26280078), and (c) (No. 26330306).

REFERENCES

- [1] Suzuki, Y., “Verisimilitude and presence of audio–visual contents,” *ITE Technical Report*, **36(44)**: 27–32 (2012) (in Japanese).
- [2] Suzuki, Y., Teramoto, W., Yoshida, K., Asai, N., Hidaka, S., Sakamoto, S., Iwaya, Y., and Gyoba, J., “Reality and presence of audio–visual contents,” *IEICE Technical Report EMM*, **111(74)**: 29–36 (2011) (in Japanese).
- [3] Teramoto, W., Yoshida, K., Asai, Y., Hidaka, S., Gyoba, J., Sakamoto, S., Iwaya, Y., and Suzuki, Y., “What is “reality” in virtual reality?,” *ITE Technical Report*, **36(12)**: 37–44 (2012) (in Japanese).
- [4] Sakamoto, S., Hasegawa, G., Honda, A., Iwaya, Y., Suzuki, Y., and Gyoba, J., “Body vibration effects on perceived reality with multi-modal contents,” *ITE Transactions on Media Technology and Applications*, **2(1)**: 46–50 (2014).
- [5] Teramoto, W., Yoshida, K., Hidaka, S., Asai, Y., Gyoba, J., Sakamoto, S., Iwaya, Y., and Suzuki, Y., “Spatiotemporal characteristics responsible for high “Vraisemblance”,” *Journal of the Virtual Reality Society of Japan*, **15(3)**: 483–486 (2010) (in Japanese).
- [6] Honda, A., Kanda, T., Shibata, H., Asai, Y., Teramoto, W., Sakamoto, S., Iwaya, Y., Gyoba, J., and Suzuki, Y., “Determinants of the sense of presence and verisimilitude in audio–visual contents,” *Journal of the Virtual Reality Society of Japan*, **18(1)**: 93–101 (2013) (in Japanese).
- [7] Witmer, B. G., and Singer, M. J., “Measuring presence in virtual environments: A presence questionnaire,” *Presence: Teleoperators and Virtual Environments*, **7(3)**: 225–240 (1998).
- [8] Gyoba, J., “Figure-ground perception: A review of recent psychological studies,” *The Journal of Institute of Electronics, Information, Communication Engineers*, **74(4)**: 315–320 (1991) (in Japanese).
- [9] Fischer, S., Lowe, R. K., and Schwan, S., “Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system,” *Applied Cognitive Psychology*, **22(8)**: 1126–1141 (2008).
- [10] Meyer, K., Rasch, T., and Schnotz, W., “Effects of animation’s speed of presentation on perceptual processing and learning,” *Learning and Instruction*, **20(2)**: 136–145 (2010).

- [11] Ono, T., and Sugawara, M., “A study on the effect of temporal sampling parameters of a camera on the evaluation of slow-motion video,” *ITE Winter Annual Convention 2001*, **83**: (2001) (in Japanese).
- [12] Zajonc, R. B., “Attitudinal effects of mere exposure,” *Journal of Personality and Social Psychology*, **9(2, 2)**: 1–27 (1968).

Appendix

In additional analysis, we conducted two-way factorial analyses of variance by group (3: normal sound condition, extended and contracted sound condition, no-sound condition) \times playback speed (7: 2, 1, 1/2, 1/4, 1/8, 1/16, 1/32) on each of the average rating scale values of the sense of presence and verisimilitude. Results show that, for both the sense of presence and for verisimilitude, the main effect of playback speed was significant (Presence: $F(6, 168) = 40.20$, $p < 0.001$; Verisimilitude: $F(6, 168) = 33.60$, $p < 0.001$). The main effect of group on the sense of presence was marginally significant ($F(2, 28) = 2.44$, $p = 0.10$). The main effect of group on the sense of verisimilitude was not significant. Furthermore, the interaction of the group and replay speed on the verisimilitude was marginally significant ($F(12, 168) = 1.57$, $p = 0.10$). The interaction of the group and replay speed on the presence was not significant. These results exhibit similar patterns to the basic results mentioned in the main text of this paper.