inter.noise HAMBURG 2016

Effects of vibration information on the senses of presence and verisimilitude of audio–visual scenes

Shuichi SAKAMOTO[1]; Zhenglie CUI[2]; Tomoko OHTANI[3]; Yôiti SUZUKI[4]; Jiro Gyoba[5]

[1,2,3,4] Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, Japan

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

ABSTRACT

To enhance people's immersive experience from high-definition multimedia communications systems, one of the important sensory modalities is vibration. We investigated the effects of full-body vibration information on perceived reality from multimodal contents. Specifically, the relation between full-body vibration amplitude and the degree of perceived reality from audio–visual contents was measured, together with the changes of perceived reality when vibrational information was presented asynchronously against audio–visual information. To create multimodal content including vibration, we recorded sound, moving pictures, and ground vibration when a train passed just beside the measurement point. Observers' ratings of the senses of presence and verisimilitude were used as indicators of perceived reality. Results revealed that the sense of presence increased monotonically according to the full-body vibration amplitude, while the sense of verisimilitude exhibited a saturating curvilinear tendency. Moreover, the perceived subjective reality in terms of both senses decreased when full-body vibration was presented asynchronously against audio–visual information. Especially, the observed degradation was steeper when full-body vibration was presented after audio–visual information than when it was presented before audio–visual information. These results indicated clearly that perceived reality from multimodal contents can be increased by presenting full-body vibration of an appropriate amount and timing.

Keywords: Vibration information, Audio–visual contents, Sense of reality, multimodal information processing   I-INCE Classification of Subjects Number(s): 13.4, 49.1, 63.7

1. INTRODUCTION

Good understanding of the multimodal information processing of humans is crucially important for investigating advanced multimedia systems. Although almost all researchers have specifically examined the effects of audio–visual information (1), other sensory modalities also play important roles in forming people’s perceptual reality. Vibration information is one of the important sensory information for enhancing people's immersive experience from high-definition multimedia systems. Actually, some entertainment systems, such as those used in motion rides, driving simulators, and movie theaters, present rich vibration information using various motion platforms. Although vibration is often regarded as noise from unsuspected noise sources in inter-noise community, vibrational information is used empirically and effectively used to provide rich multimodal information from the systems.

Some researchers have reported the effects of vibration information on perceptual reality from audio–visual contents. They pointed out that a comprehensive model of human spatial hearing can only be formulated with the inclusion of human perception of self-motion. Moreover, vibration information enhances perceived reality in musical scenes presented by an audio–visual display (2, 3). They presented full-body vibration according to music using a 3-DOF platform installed at the floor of

[1] saka@ais.riec.tohoku.ac.jp
[2] sai@ais.riec.tohoku.ac.jp
[3] ohtani.tomoko@noc.geidai.ac.jp, present affiliation is Tokyo University of the Arts
[4] yoh@riec.tohoku.ac.jp
[5] gyoba@m.tohoku.ac.jp

the experimental room. The same motion platform was used to investigate the relation between the amplitude of full-body vibration and perceived subjective reality from multimodal contents (4). Results revealed that perceived reality was enhanced by adding full-body vibration. This result is consistent with those found from previous studies (1-3). However, the observed tendency was slightly different according to the indicators used to evaluate the perceived reality.

This study investigated the effects of full-body vibration information on perceived reality from multimodal contents consisting of sound, moving image and vibration. Specifically, the relation between full-body vibration amplitude and the degree of perceived reality from audio–visual contents was measured, together with changes of perceived reality when vibrational information was presented asynchronously against audio–visual information. Audio–visual asynchrony affects various perceptual phenomena (5–9). The effect of asynchrony between auditory information and self-motion is expected to affect multimedia system quality. Yairi *et al.* reported that the latency between the timing of the listener's movement and that of the presentation of appropriate sound should be smaller than around 45 ms, which is the detection threshold of the latency for a virtual auditory displays (10). Such asynchrony is expected to affect perceptual reality from multi-modal contents. We investigated the changes of perceived subjective reality when vibration information was presented asynchronously against audio–visual information.

2. INDICATOR OF PERCEIVED REALITY (4)

The sense of reality includes many meanings (1, 11, 12). Therefore, it is difficult to ask observers directly about the amount of perceived reality. We hypothesized that perceived reality is divisible mainly into two parts: the reality obtained by the object or target and that obtained by the environment around the people. The former and latter, respectively, may also be calledas the "figure" and the "background" of a scene. If the former reality is increased, then people might perceive the object or target as the real object, even if it is an imitation. If the latter reality is increased, then people might perceive it as though they were inside the environment. The importance of each reality is therefore dependent on the contents presented by the multimodal display. Based on this hypothesis, we selected two indicators to evaluate the perceived reality.

2.1 Sense of Presence

The sense of presence is defined as the subjective experience of being in one place or environment even when one is physically situated in another place (13). This is the most popular indicator used to evaluate virtual reality (VR) systems and environment. Many researchers have examined cues for the sense of presence. Teramoto *et al.* reported that the sense of presence is used not only for evaluating the VR environment, but also for the expression of the existing "real" experience (14). Moreover, they showed that the sense of presence is dominated mainly by the absolute amount of stimuli and that it is dominantly related to background components in a scene (12, 15).

2.2 Sense of Verisimilitude

The sense of verisimilitude is the appearance of being true or real. We extend the meaning of this indicator as the existence of essence of target object in a scene, although not in actual fact. We have investigated the mechanisms of perceiving the sense of presence and the sense of verisimilitude using audio–visual contents (12, 15). Results suggested that the sense of verisimilitude corresponds to appreciation of foreground components in multimodal content.

3. EXPERIMENT1

3.1 Stimulus

To create multimodal content including vibration, we recorded sound, moving pictures, and ground vibration when a train passed just beside the measurement point. Figure 1 shows the recording setup. A dummy head (SAMRAI; Koken Co. Ltd.) with binaural microphones (4101; Brüel & Kjær) inside its ears and a DV camera (AG-DVX100A; Panasonic Inc.) were set on the ground. Binaural microphones were connected to audio inputs of DV camera via an amplifier (2639; Brüel & Kjær). Binaural signals were recorded at the sampling frequency of 48 kHz. Two acceleration pickups (VM-80; RION Co. Ltd.) were fixed on both sides of the wooden board on the ground. When the vibration information was


Figure 1 – Recording setup


Figure 2 – One scene of the multi-modal movie used in the experiment.

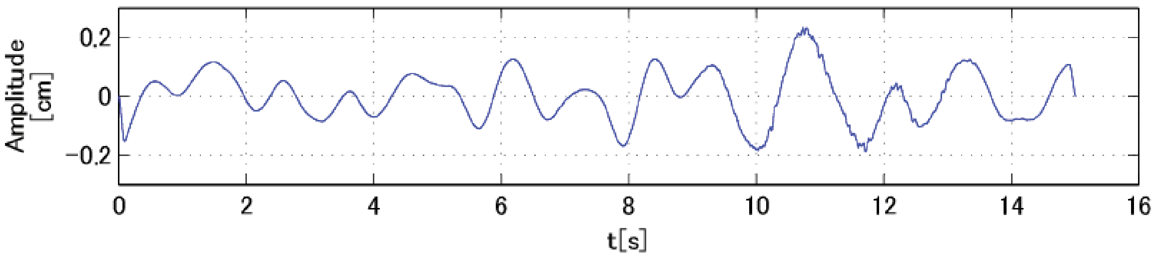

Figure 3 – Recorded vibration information

recorded, an approximately 60-kg weight was put on the board. These recording systems were set beside the JR local line (Tōhoku Honsen) near Iwakiri station. The total duration of the stimulus was 15 s. The maximum amplitude of the recorded vibration was around 0.2 cm. L_{eq} was 77.3 dB and the maximumsound pressure level was 106.7 dB. Figure 2 shows one scene of a recorded multimodal movie. Figure 3 shows an example of the recorded vibration information.

3.2 Experimental Setup

Figure 4 portrays the experimental setup. Visual information was presented using a DLP projector (PDG-DHT100JL; Sanyo Electric Co. Ltd.) on a 150 inch screen (Stewart Sound Screen) set in front of a observer. The field of view was 22.6 deg (horizontal) × 17.0 deg (vertical). The distance between the observer and the screen was around 4 m. Auditory information was presented binaurally via the headphones (HDA-200; Sennheiser Electronic). The presented sound pressure level was the same as the recorded level. Full-body vibration was provided via a motion platform (D-BOX Mastering Motion). Only 1-DOF vibration (up/down) was presented during the experiment.

3.3 Experimental Procedure

For this study, two experiments were conducted to elucidate the effects of vibration on perceived reality from multimodal contents. One experiment was conducted to investigate the effect of amplitude of vibration information on perceived reality. Although similar experiments were conducted in author's earlier study (4), the contents differed from those used here. Therefore, this experiment was examined to generalize the effect of the amplitude which was obtained in the earlier study. Another experiment was conducted to analyze the effect of asynchrony between the vibration information and audio–visual information.

3.3.1 Experiment 1: The effect of the vibration amplitude on perceived reality

Nine men and one woman (mean age=24.6 years, SD=5.6), all with normal hearing acuity and normal/corrected-normal vision, participated in the experiment. All knew the purpose of the experiments. They stood directly on the motion platform during experiments. The vibration amplitude was changed as the experimental parameter from 0 (without vibration) to two times as large as the original amplitude at 0.25 steps. Each amplitude vibration was combined to the original audio–visual stimulus and was presented four times to the observers. Therefore, 36 multimodal stimuli were presented to each observer. The order of presentation of these amplitudes was randomized. Before the experiment, the following instructions were given to the observers to clarify the difference between

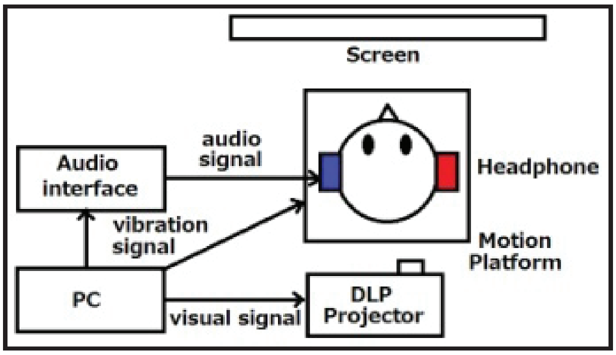

Figure 4 – Experimental setup

two indicators:

<Sense of presence>
Please rate the degree to which they felt they were now just in the presented place.
<Sense of verisimilitude>
Please rate the degree to which they felt as if presented object were real.

These instructions were almost as same as those given to observers in earlier studies (12, 15).

To evaluate each indicator, all 36 multimodal stimuli were presented twice to the observers. These 36 stimuli were divided to four sessions. Each session consisted of nine trials. Therefore, eight sessions were conducted in total. Observers were requested to rate the perceived subjective reality from 0 (low) to 6 (high). The order of the indicator was also counterbalanced among observers. The obtained data were normalized based on the observers' average and were converted to a *z*-score for comparison of scores assigned to each perceived subjective reality.

3.3.2 Experiment 2: The effect of asynchrony of information on perceived reality

Eight men and two women (mean age=23.6 years, SD=4.2), all with normal hearing acuity and normal/corrected-normal vision, participated in the experiment. Six of ten observers participated in Exp.1 (five men and one woman). All knew the experiment purpose. The setup used for Exp. 1 was used for this experiment. The lag between vibration information and audio–visual information was changed as the experimental parameter to ±800, ±400, ±200, ±100, and 0 ms. A positive value signifies that audio–visual information was started before vibration information. In all conditions, audio–visual information was presented synchronously. Each stimulus was presented to the observers four times. Therefore, 36 multimodal stimuli were presented to each observer. The order of presentation of these amplitudes was randomized. The same instructions as for Exp.1 were given to observers to evaluate senses of presence and verisimilitude.

To evaluate each indicator, all 36 multimodal stimuli were presented twice to the observers. These 36 stimuli were divided to four sessions. Each session consisted on nine trials. Therefore, eight sessions were conducted in total. Observers rated the perceived subjective reality from 0 (low) to 6 (high). The order of the indicator was also counterbalanced among the observers. The obtained data were normalized based on the observers' average and were converted to a z-score for comparison of scores of each perceived subjective reality.

4. RESULTS

4.1 Experiment 1

Figure 5 presents results of Exp. 1 as a function of the vibration amplitude. Error bars denote standard errors. Compared with the no-vibration condition, all perceived subjective reality was increased by adding vibration. However, tendencies differed slightly between indicators. The sense of presence increased monotonically according to the amplitude of full-body vibration, whereas the sense of verisimilitude showed a saturating curvilinear tendency. The highest verisimilitude score was obtained around the amplitude that was physically equivalent to the original amplitude. This tendency is completely consistent with the results of an earlier study (4). The results of ANOVA revealed that the interaction between the indicators and the vibration amplitude is statistically significant ($F(8,72)=4.94$, $p<.01$). The analysis of a simple main effect revealed that the simple main effect of the

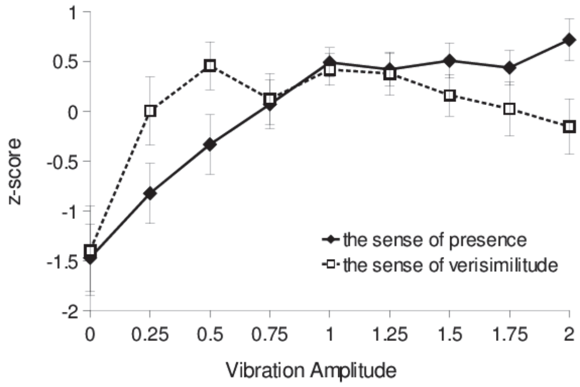

Figure 5 – Perceived reality as a function of the vibration amplitude

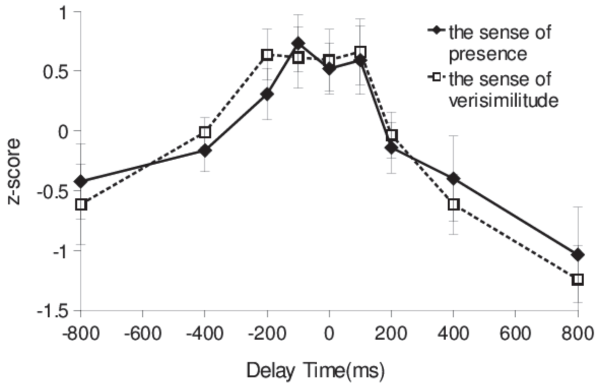

Figure 6 – Perceived reality as a function of the lag between audio-visual and vibration information

vibration amplitude is statistically significant for both indicators (sense of presence: $F(8,144)=9.24$, $p < .01$; sense of verisimilitude: $F(8,144)=5.41$, $p<.01$). A multiple comparison test (Ryan's method, $p< .05$) revealed that the perceived reality of the no-motion condition is lower than those of other conditions (sense of presence, 0.5 or above; sense of verisimilitude, 0.25 or above). Moreover, the perceived sense of presence in the 0.25 condition is statistically lower than that in 1.0 and 2.0 conditions.

4.2 Experiment 2

Figure 6 presents results of Exp. 2 as a function of the lag between vibration and audio–visual information. Error bars denote standard errors. When the lag between audio–visual and vibration information increases, the perceived reality decreases. The highest perceived reality is observed when all information is presented synchronously. The degradation slope is steeper when audio–visual information is presented before vibration information than that when vibration information is presented before audio–visual information. The results of ANOVA revealed that the lag effect is statistically significant ($F(8,72)=11.27$, $p<.05$). Multiple comparison tests (Ryan's method, $p< .05$) revealed a significant difference between the scores of −800 and −200-100, −200-100 and 400, and −400-200 and 800 ms. The effect of indicators was not statistically significant ($F(1, 72)=0.00$, n.s.).

5. DISCUSSIONS

The results of Exp. 1 demonstrated that the perceived subjective reality is increased by adding vibration. Moreover, the obtained score of the sense of presence increases monotonically according to the vibration amplitude, whereas the score of the sense of verisimilitude shows saturating curvilinear tendency. Results of an earlier study suggested that the sense of presence is strongly related to the physical amount of the stimulus, while the peak of the sense of verisimilitude is observed around the original physical amount of the stimulus (4). The results of the current study suggest that this tendency is generally independent of the presented sensory modality or the combination of the modalities. When people perceive the sense of verisimilitude, the prototype of the scene which people remember in mind is expected to play an important role. Abe *et al.* reported that the image intrinsic to a sound source, evoked by verbal and visual information, affects the evaluation of the sound (16). By comparing the actual scene with the prototype, people would evaluate the sense of verisimilitude. Perhaps for that reason, the peak is observed at the amount of the original stimulus, which suggests that same relationship between these senses and physical strength of the stimulus might exist when people perceive subjective reality.

The results of Exp. 2 demonstrated that the perceived subjective reality depends on the amount of asynchronicity between vibration and audio–visual information. The time-window size of the integration between audio–visual and vibration information is estimated to be around 600 ms (from −400 to 200 ms),which is a wider range than that of audio–visual integration at the point of the sense of presence, which is estimated as being around 300-450 ms (12). Although this size is expected to depend on the presented contents, the size of integration related to vibration information seems larger than that of other sensory information.

6. CONCLUSIONS

This study investigated the relation between full-body vibration amplitude and the degree of perceived reality from audio–visual contents, together with changes of perceived reality when vibrational information was presented asynchronously against audio–visual information. In the experiments, the senses of presence and verisimilitude were used as indicators of perceived reality. Results revealed that the sense of presence increases monotonically according to the full-body vibration amplitude, whereas the sense of verisimilitude exhibits a saturating curvilinear tendency; the tendencies are consistent with our previous findings. Moreover, the estimated time-window size to integrate presented sensory information including vibration seems larger than that of audio–visual integration. These results indicate that the perceived reality from multimodal contents can be effectively increased by presenting full-body vibration with appropriate amounts and timing.

ACKNOWLEDGEMENTS

Portions of this report were presented at the 15th International Multisensory Research Forum (IMRF) (2014). This work was supported by the National Institute of Information and Communications Technology (NICT) of Japan and Grant-in-Aid for Scientific Research (B) (No. 26280067). We wish to thank Ms. Emi TAKAHASHI for her help with analysis of the results.

REFERENCES

1. Meyer GF, Wang LT, Timson E, Perfect P, White MD Objective fidelity evaluation in multisensory virtual environments: Auditory cue fidelity in flight simulation. PLoS ONE 2012: 7(9): e44381.
2. Kim S, Sakanashi H, Martens WL, Multimodal representation of electronic piano sound for piano players. Proc. INTER-NOISE 2011; September 2011; Osaka, Japan 2011. (6-page manuscript)
3. Woszczyk W, Cooperstock J, Roston J, Martens WL, Shake, rattle, and roll: Getting immersed in multisensory, interactive music via broadband networks. J. Audio Eng. Soc. 2005: 53: 336-344.
4. Sakamoto S, Hasegawa G, Honda A, Iwaya Y, Suzuki Y, Gyoba J, Body vibration effects on perceived reality with multi-modal contents. ITE Trans. on MTA 2014: 2(1): 46-50.
5. Summerfield Q, Lipreading and audio–visual speech perception. Philos Trans R Soc Lond B Biol Sci. 1992: 335(123): 71-78.
6. Navarra J, Hartcher-O'Brien J, Piazza E, Spence C, Adaptation to audiovisual asynchrony modulates the speeded detection of sound. PNAS 2009: 106(23): 9169-9173.
7. Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW, Neural processing of asynchronous audiovisual speech perception. NeuroImage 2010: 49: 3308-3318.
8. Sugita Y, Suzuki Y, Audiovisual perception: Implicit estimation of sound-arrival time. Nature 2003: 421: 911.
9. Fujisaki W, Shimojo S, Kashino M, Nishida S, Recalibration of audiovisual simultaneity. Nature Neurosciences 2004: 7: 773–778.
10. Yairi S, Iwaya Y, Suzuki Y, Estimation of detection threshold of system latency of virtual auditory display. Applied Acoustics 2007: 68(8): 851-863.
11. Slater M, Steed A, McCarthy J, Maringelli F, The influence of body movement on subjective presence in virtual environments. Human Factors 1998: 40(3): 469–477.
12. Teramoto W, Yoshisa K, Hidaka S, Asai N, Gyoba J, Sakamoto S, Iwaya Y, Suzuki Y, Spatio-temporal characteristics responsible for high vraisemblance. Journal of the Virtual Reality Society of Japan 2010: 15(3): 483-486. (in Japanese)
13. Witmer BG, Singer MJ, Measuring presence in virtual environments: A presence questionnaire. Presence 1998: 7(3): 225–240.
14. Teramoto W, Yoshida K, Asai N, Hidaka S, Gyoba J, Suzuki Y, 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.(in Japanese)
15. Honda A, Kanda T, Shibata H, Asai N, Teramoto W, Sakamoto S, Iwaya Y, Suzuki Y, Gyoba J, Determinants of senses of presence and verisimilitude in audio–visual contents. Journal of the Virtual Reality Society of Japan 2013: 18(1): 93-101. (in Japanese)
16. Abe K, Ozawa K, Suzuki Y, Sone T, Comparison of the effects of verbal versus visual information about sound sources on the perception of environmental sounds, Acta Acustica united with Acustica 2006: 92: 51-61.