Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **1**, January **2016**

Effects of the Relation between Reproduced Sound Pressure Levels and Viewing Angles on the Sense of Presence in Audio-visual Content

Kenji Ozawa and Mayuko Mitsuki

Faculty of Engineering
University of Yamanashi
4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan
ozawa@yamanashi.ac.jp

Received August, 2015; revised October, 2015

Abstract. *This study deals with the simplest signal processing of amplitude control of audio signals in audio-visual (AV) content. Our previous investigations showed that reproduced sound pressure levels (SPLs) have considerable effects on the sense of presence of audio signals in the AV content: A reproduced SPL that is higher than the original level yields the maximum presence when a sound is presented to a listener with its corresponding moving picture. This phenomenon was observed when the viewing angle of a sound source object was the same as that in the original field. In a practical situation, however, people can position themselves freely to observe a reproduced AV content item, resulting in a different viewing angle from the original one. In this study, we examined the effects of the relation between reproduced SPLs and viewing angles on the sense of presence for eight AV content items. Five stimuli with different SPLs were prepared for each content item, and the subjects of the experiment compared their presence using the paired comparison technique. The results showed that the effects of the relation on the presence of audio signals were not significant. We can explain this result by the size constancy of vision.*
Keywords: Sense of presence, Sound pressure level, Viewing angle, Audio-visual content, Sheffé's paired comparison.

1. **Introduction.** The term *presence* is widely used to evaluate the performance of audio-visual (AV) content and systems. The literal meaning of presence is a subjective experience of being in a place or environment while physically being situated in another [1]. According to this definition, the sense of presence is determined by the property of an AV system used, i.e., how accurately the recoding and reproduction system reproduces the physical characteristics of the original field. However, our preliminary studies on auditory presence, which is the sense of presence evoked by auditory stimuli, revealed that the perceived presence depends on the reproduced content items even if one recording and reproduction system is used [2, 3]. Thus, the sense of presence must be considered from two aspects: *system presence*, which is determined by the characteristics of an AV system used, and *content presence*, which depends on the characteristics of a reproduced content item [3].

To date, many advanced AV systems have been proposed to enable the perception of a higher-degree presence. The methodologies used to evaluate the sense of presence evoked from these AV systems also play important roles in the evolution of AV content and systems. Hence, it is beneficial for users and designers of content or systems to quantitatively determine the sense of presence that a reproduced content item evokes. Although many researchers have discussed the structure of the sense of presence [1, 4, 5, 6, 7, 8], models that evaluate the degree of presence have yet to be established. Thus, our research group has been developing a model that estimates the sense of presence based on the properties of AV content items [9, 10, 11]. However, further consideration is required to obtain an accurate estimation.

One problem for the accurate estimation is amplitude control of an audio signal because this simple signal processing has considerable effects on the sense of presence in an AV content item [3, 12, 13]. Our previous investigations revealed that a reproduced sound pressure level (SPL) that was higher than the original one yielded the maximum presence when a sound was presented to a listener with its corresponding moving picture [12, 13]. This phenomenon was observed under the condition that the viewing angle of a sound source object was the same as that in the original field. In a practical situation of reproducing an AV content item, however, people can position themselves freely to observe a reproduced AV content item. This implies that the viewing angle can be different from the original one. Thus, we need to examine the effects of the relation between reproduced SPLs and viewing angles on the sense of presence of AV content items. To investigate this relation, in this study, we conducted subjective experiments on the sense of presence in two conditions of the viewing angles. A simple assumption is that a higher SPL is required to achieve the maximum presence when the viewing angle becomes larger than the original one.

2. **Method and apparatus.**

2.1. **Recording and reproduction of the stimuli.** Eight AV content items from our day-to-day lives were recorded and used as stimuli that lasted 20 s each. They were classified into three groups (Table 1) with respect to the location of sound sources relative to the listener (a dummy head in the original field and a subject in the reproduction field):

Group I: Sound sources are moving.

Group II: Sound sources are moving with the listener.

Group III: Sound sources are stationary.

Sounds in their original sound fields were recorded binaurally using a dummy head (Koken, SAMRAI), which was a mannequin similar in size and acoustical properties to real humans. Two microphones (B&K, 4134) with C-type couplers were mounted in the ears of the dummy head, and the sounds were recorded using a digital video camera (Panasonic, DJ100) with an accuracy of 16 bits and a sampling frequency of 48 kHz. The sounds were presented to the subjects via headphones (Stax, SR-Lambda). The SPLs of the content items in their original fields are listed in Table 1 as "Original L_{Aeq} (A-weighted equivalent SPL)."

It can be expected that the sounds at the ears of a listener in an original sound field are reproduced correctly when the binaurally recorded sounds are presented to the ears with headphones. If no correction was made prior to the presentation, however, two problems can occur: The first one is that the acoustical characteristics of the outer ears are convolved redundantly in recording and presentation [14], and the other is that the frequency characteristics of the reproduced sounds are different from those of the original ones because the headphones have their own characteristics. To overcome these problems, a correction was made to remove the frequency characteristics of the outer ears in the reproduction and to calibrate the frequency characteristics of the headphones by digital signal processing as follows [14]: First, the headphones used in this experiment were worn

Table 1. Eight content items and their original sound pressure levels denoted in L_{Aeq}. The results of AVOVA are shown as marks associated with the significance levels, where the meanings of the factors M, M×I, C, O, and O×I are given in Table 2. The upper and lower marks in each cell denote the significance levels for the 5-m and 1.25-m conditions, respectively.

Group	Content itme	Original L_{Aeq} [dB]	Results of ANOVA					$Y(0.05)$
			M	M×I	C	O	O×I	
I: Sound Sources are moving.	Passing train (3 m)	85.0	** **	** **	– –	– *	– –	0.28 0.25
	Passing train (150 m)	63.8	** **	** **	* *	– –	– –	0.28 0.26
	Passing train (app. 1 km)	62.8	** **	** **	– *	– –	– –	0.29 0.31
II. Sound sources are moving with the listener.	First-person view from a moving car	85.6	** **	** **	– *	– –	– –	0.25 0.29
	First-person view on a playground slide	72.0	** **	** **	– –	– **	– –	0.25 0.22
III. Sound sources are stationary.	Waterfall	82.4	** **	** **	– **	– *	– –	0.28 0.26
	Murmuring river (2 m)	76.7	** **	** **	– –	– –	– *	0.23 0.29
	Murmuring river (50 m)	69.7	** **	** **	– *	– –	– –	0.29 0.28

– not significant, * p <0.05, ** p <0.01

on the dummy head, and the transfer function from the input terminal of the headphone to the microphone mounted in the ear of the dummy head was measured. Second, the inverse characteristics of the transfer function were calculated. Finally, the inverse characteristics were convolved to the recorded sounds to obtain the stimuli used in the experiment.

Moving pictures corresponding to the sounds were simultaneously recorded using the video camera in the format given by the National Television System Committee (NTSC). These pictures were presented to a subject using a 50-inch display (Sony, KL-9200J). During recoding, zooming was adjusted so that the reproduced angle of a visual field was the same as the original field for a subject sitting 5 m away from the display. Figure 1 shows sample still images captured from the visual signals.

2.2. **Experimental procedure.** The aim of the experiment was to observe the effects of reproduced SPLs on the sense of presence under two conditions of the viewing angle of the corresponding moving picture. Hereafter, the two conditions are referred to as 5-m and 1.25-m conditions in which the distances between the display and a subject are 5 m and 1.25 m, respectively. Fourteen undergraduate students, six female and eight males with normal audio and visual perception, participated in this experiment. They were divided into two subject groups of Groups 1 and 2. The subjects in Group 1 first evaluated the

(a) Passing train (3 m) (b) Passing train (app. 1 km)

Figure 1. Sample still images of the visual stimuli. In panel (b), the train is indicated with an orange oval.

sense of presence under the 1.25-m condition and then under the 5-m condition, whereas Group 2 subjects evaluated in the reverse order. There was at least 1-week period between the experiments using the 5-m condition and those using the 1.25-m condition.

In each condition, the experiments were carried out separately for each of the eight AV items using the method of Scheffé's paired comparison [15] modified by Ura [16]. Five sounds whose SPLs were changed in 3- to 12-dB steps were prepared for each content item. Two sounds, randomly selected from the five sounds, were presented to a subject sequentially with an interval of 3 s. The accompanying moving picture was always the same, irrespective of the SPLs. The subjects compared the two sounds (namely, A and B, in the presentation order) in terms of their presence. They were instructed to give rating judgment on a seven-point category scale (from "-3" to "$+3$"). One end (-3) of the scale corresponded to the case when sound A had much higher presence than sound B; the other end ($+3$) corresponded to the opposite case; and the midpoint (0) corresponded to the case when both A and B had equivalent presence. The number of comparisons for each item was 20 ($= {}_5P_2$), because every possible permutation was presented to cancel out any time order error. The total number of comparisons was 320 ($= 2$ conditions $\times$ 8 items $\times$ 20 comparisons) for each subject.

3. Experimental results.

3.1. Group data.

Table 2 gives an example of the results of analysis of variance (ANOVA) for all subjects. The main effect is highly significant, which indicates that the sense of presence changes depending on the reproduced SPLs. Besides the main effect, the interaction between the main effect and the individual is statistically significant as well. It is not usually sufficient to discuss the main effect if the interaction is significant. However, as given in Table 2, the F_0 value, which is the ratio of unbiased variances between a factor and the error, of the main effect is much larger than that of the interaction. Thus, let us first discuss the main effect for the group data in this section and then consider the interaction in the following section.

The significance levels of the factors are summarized in Table 1, and the two aforementioned effects, i.e., the main effect and the interaction between the main effect and the individual, are highly significant for all content items. However, it is difficult to find any systematic trends in the significance levels for other factors. This suggests that the type of stimuli with respect to the movement of a sound source relative to the listener is not a

TABLE 2. Example of the results of ANOVA for the 5-m condition of stimulus *Passing train* (3 m). SS, $d.f.$, variance and F_0 mean sum of square, degree of freedom, unbiased variance, and the F value, respectively.

Factor	SS	$d.f.$	Variance	F_0	Significance
Main effect (M)	470.5	4	117.6	147.5	**
Main effect × Individual (M × I)	209.9	52	4.0	5.1	**
Combination effect (C)	5.8	6	0.97	1.2	–
Order effect (O)	0.032	1	0.032	0.040	–
Order × Individual (O × I)	8.1	13	0.62	0.78	–
Error	162.6	204	0.80		
Total	857	280			

– not significant, * p <0.05, ** p <0.01

key factor when considering the effects of the reproduced SPLs on the sense of presence, although the stimuli were classified into three groups in the present study.

To visualize the effects of the reproduced SPLs on the sense of presence, relative presence scores for the five sounds of each AV content item were calculated by applying the scaling procedure [15, 16] to the data for each content item and each condition separately. Figure 2 shows the derived scores. The 5% yardstick, $Y(0.05)$, for each data set is given in Table 1.The yardstick indicates that the difference in the presence scores between two stimuli is statistically significant beyond the 5% level if the difference is larger than $Y(0.05)$. In all data sets, the effects of the reproduced SPLs are significant in the yardstick test as well as in the results of ANOVA. Incidentally, it is noteworthy that the presence data of the two conditions in a graph cannot be quantitatively compared since the scores of each condition are denoted in relative scales. In other words, only tendencies as a function of SPL can be compared.

The results show good agreement with our previous investigations [12, 13] in the following two points: The first one is that the reproduced SPL at the maximum presence is higher than that in the original field, and the other one is that, for Group II items, monotone increase is observed, or the maximum presence is yielded at a much higher level of SPL than the original SPL.

To examine the effects of the viewing angle on the sense of presence, Pearson product-moment correlation coefficients were calculated for the relative presence scores between the 5-m and 1.25-m conditions for each item. High correlations are obtained for all items, as shown in Fig. 2, indicating that the effect of the viewing angle is small.

3.2. **Individual data.** To examine the interaction between the main effect and the individual, the scaling procedure was applied to individual data. As a result, there are three types of graph shifts between the two conditions, and examples are shown in Figure 3. Figures 3(a) and 3(c) show that the graphs shift toward higher and lower SPLs, respectively, when the distance between the display and the subject is changed from 5 m to 1.25 m. In contrast, Figure 3(b) shows that there is no shift.

We decided that if the correlation coefficient is less than 0.8, the graphs are regarded to be shifted. Table 3 gives the numbers of subjects for the three types of shift, where "lower shift" and "higher shift" mean that the graphs seem to shift toward lower and higher SPLs, respectively, when the distance changed from 5 m to 1.25 m. The tendencies are different between the subject groups: Group 1 subjects show higher or no shift, whereas Group 2 subjects show all three types of shift (Table 3).

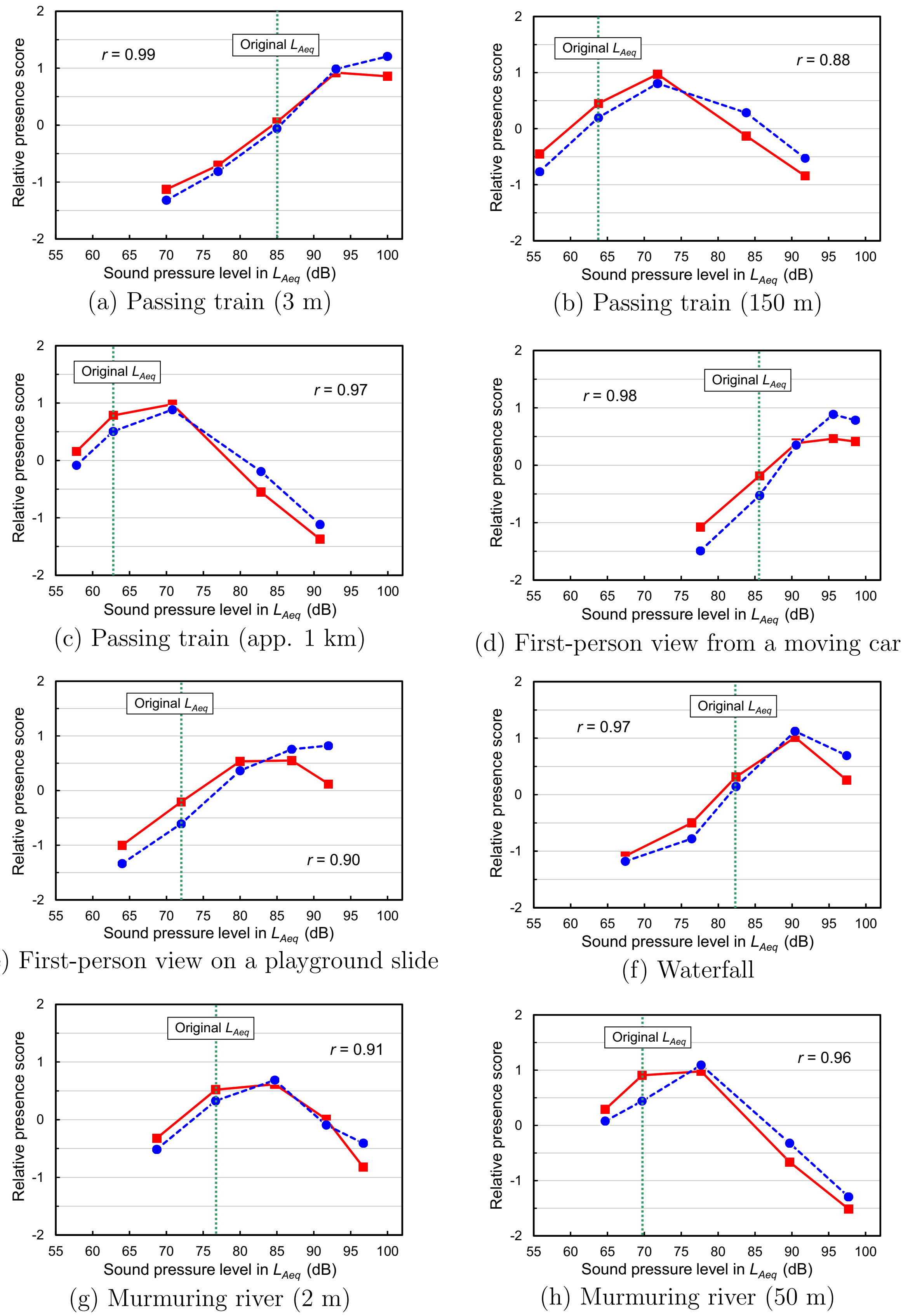
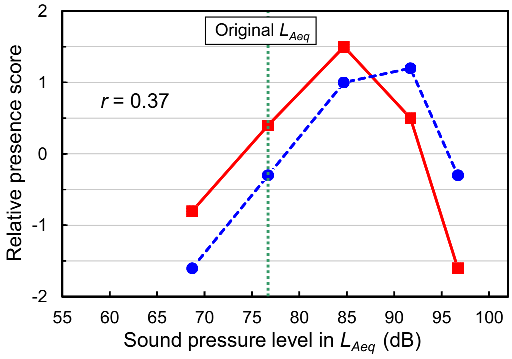
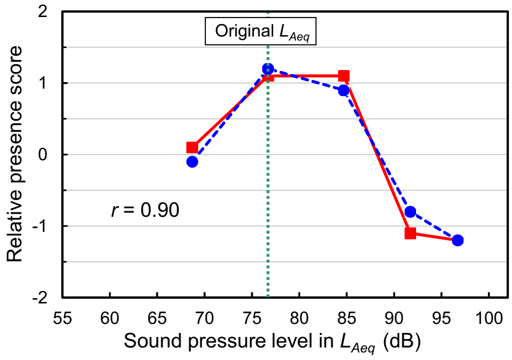


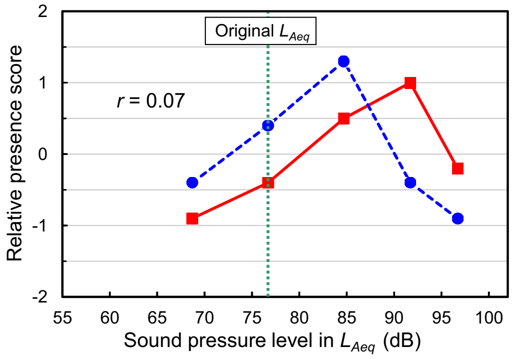
FIGURE 2. Relative presence score as a function of reproduced SPLs. Solid and dashed lines represent the results of the 5-m and 1.25-m conditions, respectively. Correlation coefficients between the two conditions are indicated by r.



(a) Example of "shift to higher level" subjects.



(b) Example of "no shift" subjects.



(c) Example of "shift to lower level" subjects.

FIGURE 3. Examples of individual results of the content item "Murmuring river (2 m)." Solid and dashed lines represent the results of the 5-m and 1.25-m conditions, respectively.

Our assumption is that a higher SPL is required to achieve the maximum presence when the viewing angle becomes larger than its original one. The higher shift is the case in which the assumption holds.

4. Discussion.

4.1. Presence score as a function of SPL. For most of the AV content items, relative presence scores show single-peaked graphs as a function of SPL, and the SPLs at the maximum presence are higher than the original ones. The reason for this phenomenon was explained in our previous reports [12, 13]. When a subject listens to a sound with its corresponding moving picture, the primary criterion of presence should be whether the

Table 3. The number of subjects who showed lower, higher, or no shift when the condition was changed from 5 m to 1.25 m.

	Content item	Lower shift	No shift	Higher shift
Subject group 1	Passing train (3 m)	0	5	2
	Passing train (150 m)	0	3	4
	Passing train (1 km)	0	2	5
	First-person view from a moving car	0	4	3
	First-person view on a playground slide	0	4	3
	Waterfall	0	4	3
	Murmuring river (2 m)	0	3	4
	Murmuring river (50 m)	0	4	3
	Subtotal	0	29	27
Subject group 2	Passing train (3 m)	0	7	0
	Passing train (150 m)	0	5	2
	Passing train (1 km)	3	4	0
	First-person view from a moving car	1	5	1
	First-person view on a playground slide	2	4	1
	Waterfall	1	6	0
	Murmuring river (2 m)	2	3	2
	Murmuring river (50 m)	1	5	1
	Subtotal	10	39	7
Total		10	68	34

loudness of the sound matches the moving picture. Thus, this results in a single-peaked pattern, indicating that the factor of *evaluation* involved in AV presence is dominant where an adjective pair of "matched–mismatched" has the largest factor loading on this factor [4]. However, around the original SPL, there is a permissible SPL range in which a listener can accept a combination of a given sound and the corresponding moving picture as natural. In this range, the criterion should be that a more powerful sound yields a higher presence because of the factor of *psychological loading* [4], to which an adjective pair "powerful–powerless" is closely related.

However, monotone increases are found for Group II items in which sound sources are moving with the listener. The sound sources cannot be seen in the accompanying video for these items, because the sound sources of *a moving car* were the engine of the car and its tires, and those of *a playground slide* were its rolling bars rotated by the person's legs and hip. In such cases, it is probably difficult to evaluate the degree of matching between the sound and the video; as a result, the factor of *psychological loading* became predominant. Another explanation is that the lack of information with respect to senses other than audition and vision, e.g., vibration and tactile sensation, was compensated for by higher SPLs.

4.2. **Effects of viewing angles.** As stated in Section 3.1, the effects of the viewing angle, which was changed by the distance between the display and the subject, were not significant for the group data. This phenomenon can be explained by the size constancy, which is a type of visual subjective constancy: Within a certain range, people's perception of one particular object's size does not change, regardless of the changes in the distance or the video size change on the retina [17]. This phenomenon suggests that, from the viewpoint of the sense of presence, we need not change the SPL of a reproduced AV item depending on the distance between a TV and a viewer in practical situations. Although this study assumed a practical living room condition where a viewer tends to sit near a TV, the situation should be similar to a movie theater where the distance between the screen and the audience differs significantly depending on the position of a seat. However, further investigation is required to examine the limitation of the viewing angle where the correct SPL is compensated for by the size constancy. For example, in a visual research, Shibata *et al.* observed the sense of presence for a small display like mobile phones [18]. In future research, we need to examine whether the phenomenon shown in this investigation also holds for such a small display.

For the individual data, Table 3 shows that some results meet our simple assumption: A higher SPL is required to achieve the maximum presence when the viewing angle becomes larger than its original value, as shown in Fig. 3(a). However, the opposite results are also given in Table 3 and Fig. 3(c): A lower SPL is required to achieve the maximum presence. It is clear from Table 3 that the distributions of the higher and lower shifts are asymmetric for the two subject groups. This suggests that there was a memory effect based on the order of the tested experimental conditions. As stated in Section 2.2, Group 1 subjects first evaluated under the 1.25-m condition and then 5-m condition, whereas Group 2 subjects evaluated in the reverse order. Thus, the memory effect shown in Table 3 indicates that the subjects were satisfied with lower SPL in the latter condition (5-m or 1.25-m condition for Group 1 or 2, respectively) for some cases. Although these two conditions were tested at least 1 week apart, there was a possibility that the visual stimuli were memorized by the subjects because one visual stimulus was presented to them 40 times (twice in a pair and the number of the pairs was ${}_5P_2$) in a session. Because these memories of the stimuli might have affected the presence judgment, further experiments are required to clarify the memory effects by changing the interval between these two conditions. However, the memory effect should have been canceled out in Fig. 1 because the number of the subjects was balanced between these two groups.

When we assume that the observed number 10 in the lower shift in Group 2 (Table 3) is because of the memory effect, and if the memory effect is subtracted from the numbers in Table 3 (namely, the number 10 is subtracted from the subtotal numbers of the higher shifts in Group 1 and of the lower shifts in Group 2), the number of the higher shifts is greater than that of the lower shifts. This tendency supports our simple assumption that a higher SPL is required to achieve the maximum presence when the viewing angle becomes larger than its original value, although the tendency is not statistically significant as shown in Fig. 1.

5. **Conclusions.** In this study, we examined the effects of the relation between reproduced SPLs and viewing angles on the sense of presence in AV content items. The experimental results showed that the effects were not significant. We considered that this is explained by the size constancy in vision: Within a certain range, people's perception of one particular object's size does not change, regardless of the changes in the distance or the video size change on the retina [17]. Thus, it is suggested that, from the viewpoint of the sense of presence, we need not change the SPL of a reproduced AV item depending on the distance between a TV and the viewer in practical situations. Further investigations are needed to examine the following two points: The first one is that we need to clarify the limitation of the viewing angle where the correct SPL is compensated for by the size constancy, and the other is how the memory effect of moving pictures influences the sense of presence as a function of SPLs of a content item.

Acknowledgment. This study was funded by the National Institute of Information and Communication Technology (NICT), Japan.

REFERENCES

[1] B. G. Witmer and M.J. Singer, Measuring presence in virtual environments: A presence questionnaire, *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[2] K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone, Contents which yield high auditory-presence in sound reproduction, *Kansei Engineering International*, vol. 3, pp. 25–30, 2002.

[3] K. Ozawa and Y. Chujo, Content presence vs. system presence in audio reproduction systems, *Proc. of 2nd Int. Symp. on Universal Communication (ISUC2008)*, Osaka, Japan, pp. 50–55, 2008.

[4] K. Fukue, K. Ozawa, and Y. Kinoshita, Study on multidimensional structure of the sense of presence in audio-visual content, *Trans. of Japan Soc. of Kansei Engineering*, vol. 11, no. 2, pp. 183–192, 2012. (in Japanese)

[5] T. B. Sheridan, Musings on telepresence and virtual presence, *Presence*, vol. 1, no. 1, pp. 120–126, 1992.

[6] T. B. Sheridan, Further musings on the psychophysics of presence, *Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC 1994)*, pp. 1073–1077, San Antonio, USA, 1994

[7] T. Schubert, F. Friedmann, and H. Regenbrecht, The experience of presence: Factor analytic insights, *Presence*, vol. 10, no. 3, pp. 266–281, 2001.

[8] W. Teramoto, K. Yoshida, N. Asai, S. Hidaka, J. Gyoba, and Y. Suzuki, What is "sense of presence": A non-researcher's understanding of sense of presence, *Trans. of the Virtual Reality Soc. of Japan*, vol. 15, pp. 7–16, 2010. (in Japanese)

[9] Y. Kinoshita, K. Fukue, and K. Ozawa, Development of *Kansei* estimation models for the sense of presence in audio-visual content, *Proc. of 2011 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC 2011)*, Anchorage, USA, pp. 3280–3285, 2011.

[10] K. Ozawa, M. Obinata, and Y. Kinoshita, *Kansei* estimation models for the sense of presence in audio-visual content with different audio reproduction methods, *Proc. of 13th ACIS Int. Conf. on Software Engineering, Artificial Intelligence Networking and Parallel/Distributed Computing (SNPD 2012)*, Kyoto, Japan, pp. 567–573, 2012.

[11] K. Ozawa, S. Tsukahara, Y. Kinoshita, and M. Morise, Development of an estimation model for instantaneous presence in audio content, *Proc. of the 4th IEEE Int. Conf. on Consumer Electronics - Berlin (ICCE-Berlin 2014)*, Berlin, Germany, pp. 401–405, 2014.

[12] K. Ozawa and M. Miyasaka, Effects of reproduced sound pressure levels on auditory presence, *Acoustical Science and Technology*, vol. 25, no. 3, pp. 207–209, 2004.

[13] M. Mitsuki, M. Miyasaka, and K. Ozawa, Reproduced sound pressure level yielding the maximum auditory presence: Further study on effects of reproduced SPLs on auditory presence, *Acoustical Science and Technology*, vol. 26, no. 1, pp. 79–81, 2005.

[14] H. Møller, Fundamentals of binaural technology, *Applied Acoustics*, vol. 36, pp. 171–218, 1992.

[15] H. Scheffé, An analysis of variance for paired comparison, *J. American Stat. Assoc.*, vol. 47, pp. 381–400, 1952.

[16] S. Ura, An analysis of a paired comparison experiment, *Quality Control – Journal of Japanese Union of Scientists & Engineers*, vol. 16, pp. 78–80, 1959. (in Japanese)

[17] N. Carlson, *Psychology: the Science of Behavior, 4th Canadian ed.*, Pearson, Toronto, Canada, p. 188, 2010.

[18] T. Shibata, S. Kurihara, T. Kawai, T. Takahashi, T. Shimizu, R. Kawada, A. J. Häkkinen, J. Takatalo, and G. Nyman, Evaluation of stereoscopic image quality for mobile devices using interpretation based quality methodology, *Proc. of SPIE 7237, Stereoscopic Displays and Applications XX*, 8 pages PDF, 2009. (doi:10.1117/12.807080)