172nd Meeting of the Acoustical Society of America

Honolulu, Hawaii

28 Nov - 2 Dec 2016

Engineering Acoustics: Paper 5aEA3

Effect of the horizontal panning on sense of presence in three-dimensional audio system using multiple vertical panning

Toshiyuki Kimura
Faculty of Engineering, Tohoku Gakuin University, Tagajo, Miyagi, Japan; t-kimura@m.ieice.org

Hiroshi Ando
Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita, Osaka, Japan; h-ando@nict.go.jp

We previously proposed a three-dimensional (3D) audio system using the multiple vertical panning (MVP) method to develop a 3D audio system that matches a multi-view 3D video display system (REI display). In this paper, in order to apply our proposed method to the teleconference system, we performed two audio-visual psychological experiments and evaluated the effect of the horizontal panning on the sense of presence in our proposed method. In the first experiment, although viewers cannot discriminate the differences of the sense of presence even if the horizontal panning was added to our proposed method, it is difficult to reduce the number of loudspeakers from ten. In the second experiment, viewers cannot discriminate the differences of the sense of presence even if the number of loudspeakers is reduced to four in the condition where loudspeakers are not placed at four vertex positions of the rectangle display. Thus, we found that the performance was adequately maintained if four loudspeakers were placed except four vertex positions of the rectangle display when the teleconference system based on our proposed method is constructed.

1. INTRODUCTION

Ultra-realistic communication techniques have been investigated at NICT.[1] Their applications will enable more realistic forms of communication (e.g., 3D television and 3D teleconferencing) than those currently offered by conventional video and audio techniques (4K television and 5.1-channel audio).

At NICT, a glasses-free 3D video technique using a projector array has been proposed, and a multi-view 3D video display system (REI display) with a 200-inch screen has also been developed[2] whose basic configuration is shown in Figure 1. Parallax videos are projected to a Fresnel lens by projector units, which are components of the projector array. These parallax videos are only projected in the horizontal direction because of the diffusion characteristics of a diffuser screen placed in front of the Fresnel lens (a small diffusion angle in the horizontal direction and a wide diffusion angle in the vertical direction). If viewers view parallax videos from a particular horizontal viewing position, since right and left eyes of viewers view different videos corresponding to the horizontal viewing position, viewers can view 3D videos based on stereoscopic vision. This system allows several viewers simultaneously view natural 3D objects based on each particular horizontal viewing position without special glasses.
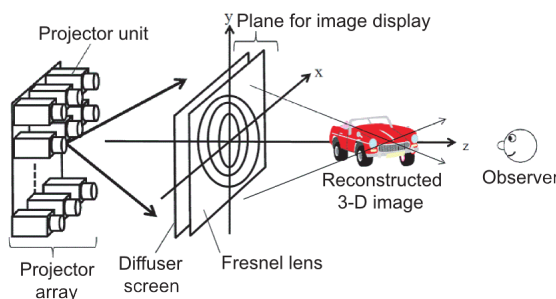

Figure 1: Basic configuration of a large-screen multiview 3D video display system (REI display).[3]

We previously proposed a 3D audio system using the multiple vertical panning (MVP) method to develop a 3D audio system that matches our developed REI display system. The results of the audio-visual psychological experiment indicated that our proposed system was effective with such conventional audio systems as stereophonic audio[4] and that the number of loudspeakers can be reduced to ten in our proposed method.[5] The basic configuration of our proposed system is shown in Figure 2. First, as shown on its left-hand side, two loudspeakers are placed at the upper and lower sides of the 3D object's position depicted in the screen by the developed 3D video display system. If a sound is played from two loudspeakers using the panning between them (vertical panning), viewers can perceive a sound image between the two loudspeakers. If their sound pressure level difference is properly adjusted, because the sound-playing devices are only two loudspeakers placed at the upper and lower sides of the screen (vertically panned loudspeakers), multiple viewers can perceive a sound image at the position of the 3D object, regardless of their viewing position. Second, as shown on the right-hand side of Figure 2, the sound image positions are also expanded in the left-right direction by placing multiple vertically panned loudspeakers at the upper and lower sides of the screen. As a result, multiple viewers can simultaneously experience the
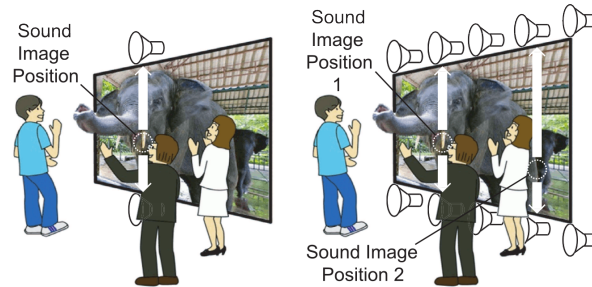
Figure 2: Basic configuration of a 3D audio system using multiple vertical panning (MVP) method.[4]

sound images at the position of the 3D objects depicted by the 3D video display system, regardless of their viewing position.

In our previous studies,[4,5] we assumed that the recording microphones were placed at the neighborhood of sound sources and that the sound source signals and the position information of sound sources were transmitted. However, when our proposed method is applied to the teleconference system in which many people participate, it is difficult to place microphones in the neighborhood of the sound sources and the amount of the transmitted data increases according to the number of participants. To solve these problems, it is necessary to develop the recording microphone array where the transmission of the position information of participants can be omitted.

On the other hand, the sense of presence was adequately expressed even if the horizontal position of the sound sources was discretized with five steps and their vertical position was discretized with two steps in our previous study.[5] Thus, the recording microphone array for our proposed method can be constructed if microphones are placed at the position of discretized sound sources and respective microphones capture the sound of the surrounding area of the placed position.

As the microphone array for recording, we propose the hyperdirectional microphone array shown in Figure 3 in order to apply our proposed method to the teleconference system. Hyperdirectional microphones are placed the position of corresponding loudspeakers. The direction of hyperdirectional microphones is the same as the radiation direction of loudspeakers. Because the gain of the opposite direction of hyperdirectional microphones is little or none, the howling between the hyperdirectional microphones and loudspeakers is suppressed. Because the hyperdirectional microphones are placed at the upper and lower sides of the display, viewers can view 3D videos without coming hyperdirectional microphones into the viewers' eyesight. The number of transmitted channels is the same as that of loudspeakers in the proposed array.

In the proposed hyperdirectional microphone array, neighboring microphones simultaneously record a sound if there are no sound sources toward the direction of hyperdirectional microphones. As a result, the horizontal and vertical panning of sound sources is added to the recorded signals. Although there is no effect of the vertical panning because our proposed method utilizes the vertical panning, the effect of the horizontal panning is not evaluated. In this paper, the effect of the horizontal panning on the sense of presence is evaluated by the audio-visual psychological experiments.

On the other hand, if the horizontal panning is added, because the horizontal position of sound images is accurately expressed by stereophonic in all viewing positions, the horizontal position of
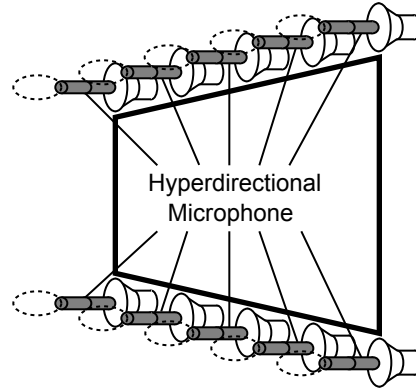
Figure 3: Basic configuration of a hyperdirectional microphone array for MVP method.

sound images becomes equal to that of 3D objects in the 3D video display system. As a result, There is the possibility of the reduction of the required number of loudspeakers and hyperdirectional microphones for the construction of the teleconference system. In the experiments, the effect of the horizontal panning on the number of loudspeakers is also evaluated by the sound conditions in which the number of loudspeakers is less than ten (the minimum number of loudspeakers in the previous study[5]).

2. AUDIO-VISUAL EXPERIMENT 1

In this section, to evaluate the effect of the horizontal panning, we perform an audio-visual psychological experiment based on the condition used in the previous study.[5]

A. ENVIRONMENT AND CONDITIONS

Although a REI display should be applied in this experiment, the REI display had been exhibited at the Knowledge Capital[6] of Grand Front Osaka[7] daily from 10 am to 9 pm. Because it was difficult to perform the experiment on the time except an exhibition considering the overload of viewers for the experiment, the REI display could not be used in this experiment. Our experiment was performed in a conference room where a 200-inch rear-projection visual screen was set up. Two projectors for the 2D video of the left and right eyes were placed behind the screen. Polarization plates were placed in front of the projectors so that viewers can see 3D video by wearing polarization glasses based on stereoscopic vision. If viewers don't move their viewing position, the 3D video by this screen is equal to that by the REI display. The room's reverberation time was 402 ms, and the background noise level had an A-weighted level of 38 dB.

We placed 82 loudspeakers in the room (Figure 4) in the forward position 0.275 m from the screen because they could not be placed over and under the screen, which was attached to the wall. The loudspeakers were made by mounting a loudspeaker unit (Fostex: FE103E) on a loudspeaker enclosure (width: 11 cm, depth: 25 cm, height: 11 cm). Considering the proper viewing distance in the developed large-screen multi-view 3D video display system (±2 m centered 5.5 m), three viewing positions (forward, central, and backward) were set 3.5, 5.5, and 7.5 meters from the screen. The viewing width of the developed system was 2 m across, centered around the front
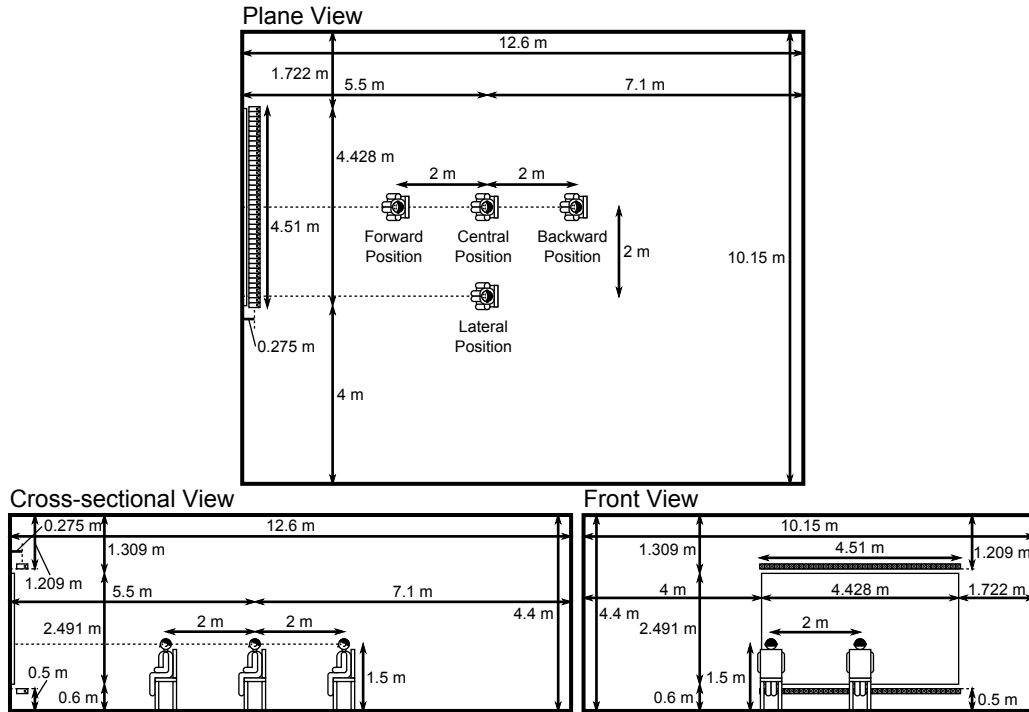
Figure 4: Position of viewers, screen and the loudspeaker array in first audio-visual experiment.

viewing position of the screen when the viewing distance was 5.5 m. Thus, an additional viewing position (lateral) was set at a lateral position 2 m to the left of the central position. The height of all the viewing positions was set to 1.5 m at the ear position of the viewers. The sound pressure level was set to an A-weighted level of approximately 70 dB in the central viewing position.

The 3D video (30 fps) used in this experiment is shown in the left-hand side of Figure 5. In it, the UFO (inside the yellow oval) that plays a sound is moving about the screen every five seconds (i.e., 150 frames). When it touches the stars and balls (inside the red circles), the sound of the stars and balls is played at their positions. The position and movement of the UFO, the stars, and the balls in the 3D video are shown in the right-hand side of Figure 5. The start and end points of the arrows denote the UFO position between the frames (1/30 seconds) in the 3D video. The gray stars and circles denote the position of the stars and balls in the 3D video. The sounds of the UFO, the stars, and the balls are the materials of the sound effects published on a website.[8] The time waveform and spectrogram of the sounds are shown in Figure 6. The proper viewing distance and the parallax of the 3D video are 5.5 m and 0.0625 m, respectively. Because the 3D viewing videos change based on the viewing positions in our developed 3D video display system, we also changed the presented 3D videos in this experiment based on the viewing positions.

The sound conditions are shown in Figure 7. The gray loudspeakers denote the loudspeaker from which a sound was not replayed in each condition. The black arrows on the screen denote the horizontal and vertical panning in the replayed vertically panned loudspeakers. Sound conditions (e) and (f) are the condition where viewers could not discriminate the differences of the sense of presence even if the number of loudspeakers was increased in the previous study.[5] On the other hand, sound condition (a) is the condition where viewers discriminated the differences of the sense
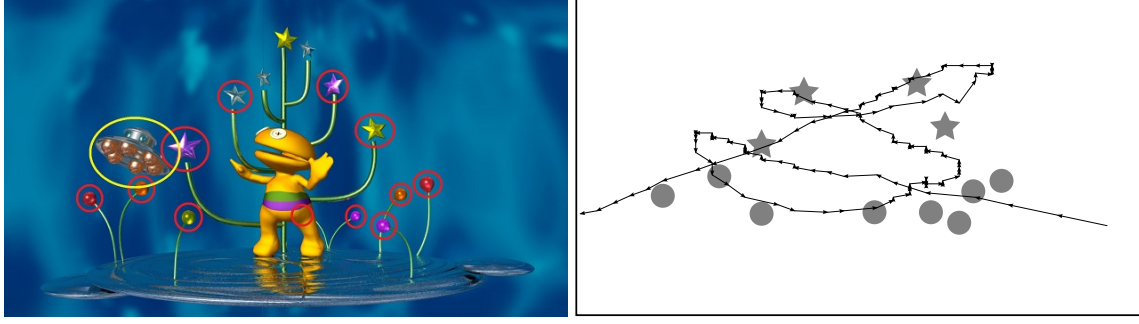
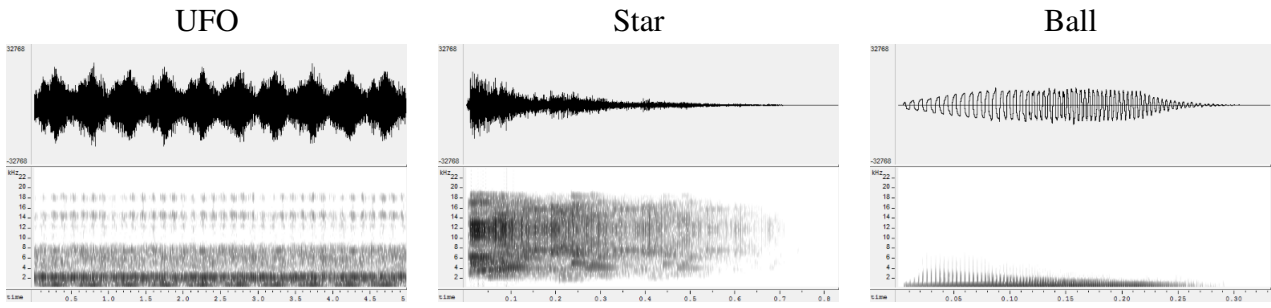**Figure 5: 3D video and position and movement of UFO, stars, and balls in first audio-visual experiment.**

UFO	Star	Ball

Figure 6: Time waveform and spectrogram of UFO, star and ball sounds in first audio-visual experiment.

of presence in the previous study.[5] These conditions are set to evaluate the viewers' ability to discriminate the sense of presence in this experiment.

The sounds played at 3D object position (P_H, P_V) at time $T = \frac{m-1}{F_v}$ were synthesized by the following procedure. Note that F_v(=30 fps) and m(=1,...) denote the frame rate and the frame index of the video signals. P_H(=-2.2~2.2) and P_V(=-1.25~1.25) denote the horizontal and vertical positions of the presented 3D object. If P_H is 0, the horizontal position corresponds to the screen's horizontal central position. The height of the sound images is the same as that of the ear position of the viewers if P_V is -0.3455.

First, based on the horizontal position of the presented 3D object, P_H, two loudspeakers placed at the upper and lower sides of the screen are selected:

$$P_H' = \Delta d_H \, \mathrm{round} \left(\frac{P_H + 2.2}{\Delta d_H} \right) - 2.2, \tag{1}$$

where $P_H'(= -2.2, ..., 2.2)$ denotes the horizontal position of the two selected loudspeakers. Δd_H denotes their right-and-left intervals. In this experiment, the Δd_H values are 4.4 m in sound conditions (a) and (b), 2.2 m in sound condition (c), 1.1 m in sound conditions (d) and (e), and 0.22 m in sound condition (f). If the horizontal panning is added (sound conditions (b)-(d) in this experiment), two loudspeakers for the horizontal panning are additionally selected:

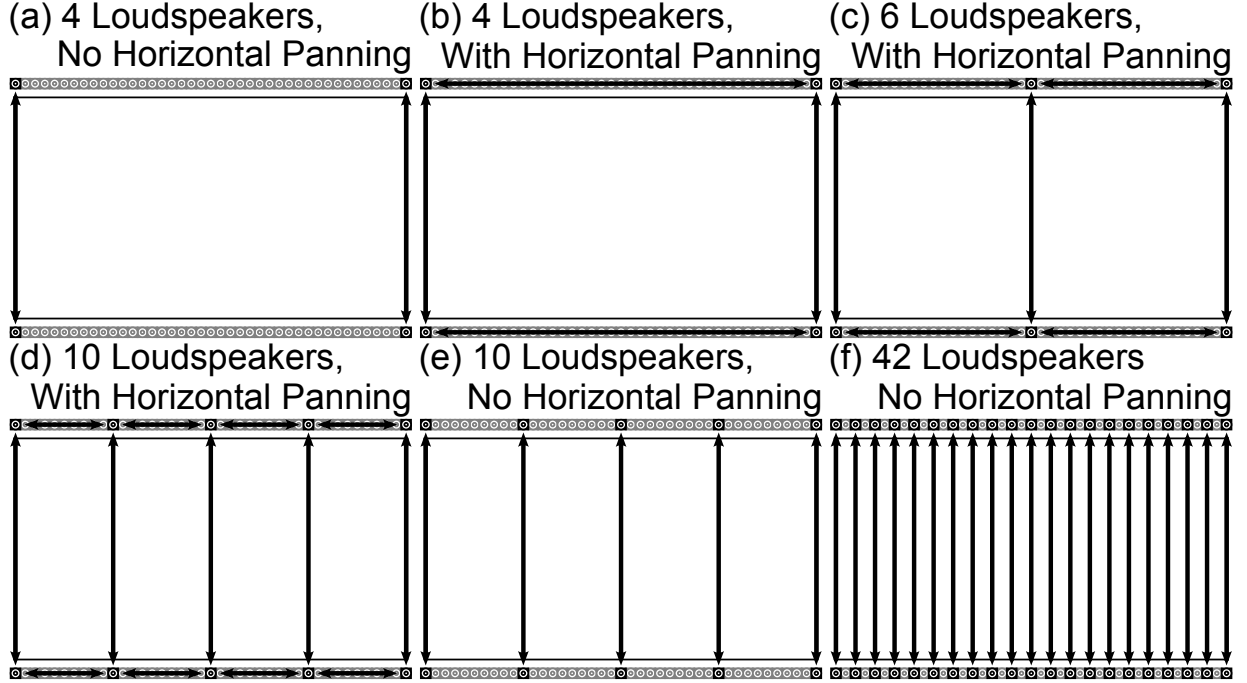$$P_H'' = P_H' + \mathrm{sign}(P_H - P_H')\Delta d_H, \tag{2}$$

Figure 7: Sound conditions in first audio-visual experiment.

where P_H'' denotes the horizontal position of two additionally selected loudspeakers for the horizontal panning.

Second, if the horizontal panning is not added (sound conditions (a), (e) and (f) in this experiment), the sound calculated from the sound source signal, $s(n)$, is replayed from two selected loudspeakers: Second, the sound calculated from the sound source signal, $s(n)$, is replayed from two selected loudspeakers:

$$\begin{cases} x_U(n) & = a_U w(n)s(n) \\ x_D(n) & = a_D w(n)s(n) \end{cases} \left(n = \frac{F_s}{F_v}(m-1), ..., \frac{F_s}{F_v}m + LF_s\right), \tag{3}$$

where F_s(=48 kHz) and n(=0,...) denote the sampling frequency and the sample time of the sound signals and $x_U(n)$ and $x_D(n)$ denote the sound signals replayed from the two loudspeakers of the upper and lower sides. $w(n)$ denotes the window function of the sound signals defined as follows:

$$w(n) = \begin{cases} \frac{1}{LF_s}\{n - \frac{F_s}{F_v}(m-1)\} & \left(n = \frac{F_s}{F_v}(m-1), ..., \frac{F_s}{F_v}(m-1) + LF_s\right) \\ 1 & \left(n = \frac{F_s}{F_v}(m-1) + LF_s, ..., \frac{F_s}{F_v}m\right) \\ -\frac{1}{LF_s}(n - \frac{F_s}{F_v}m) + 1 & \left(n = \frac{F_s}{F_v}m, ..., \frac{F_s}{F_v}m + LF_s\right) \end{cases}, \tag{4}$$

where L(=1 ms) denotes the crossfade time of the window function. a_U and a_D (the gain coefficients in each sound signal) are calculated from the level difference, ΔA [dB], as follows:

$$a_U = \frac{10^{\frac{\Delta A}{20}}}{\sqrt{10^{\frac{\Delta A}{10}} + 1}}, a_D = \frac{1}{\sqrt{10^{\frac{\Delta A}{10}} + 1}}. \tag{5}$$

In this experiment, level difference ΔA was based on a previous study[4] as follows:

$$\Delta A = \frac{\alpha P_V + 0.1437}{0.1065}. \tag{6}$$

The vertical interval of the loudspeakers is 2.7 m in this experiment, but it was 2.5 m in the previous study.[4] Thus, $\alpha(= \frac{2.7}{2.5})$ was set to compensate the differences of the vertical intervals of the loudspeakers.

On the other hand, if the horizontal panning is added (sound conditions (b)-(d) in this experiment), the sound calculated from the sound source signal, $s(n)$, is replayed from four selected loudspeakers:

$$\begin{cases} x_{U1}(n) & = \cos\left(\frac{\pi}{2}\frac{P_H-P'_H}{\Delta d_H}\right)a_U w(n)s(n) \\ x_{D1}(n) & = \cos\left(\frac{\pi}{2}\frac{P_H-P'_H}{\Delta d_H}\right)a_D w(n)s(n) \\ x_{U2}(n) & = \cos\left(\frac{\pi}{2}\frac{P_H-P''_H}{\Delta d_H}\right)a_U w(n)s(n) \\ x_{D2}(n) & = \cos\left(\frac{\pi}{2}\frac{P_H-P''_H}{\Delta d_H}\right)a_D w(n)s(n) \end{cases} \left(n = \frac{F_s}{F_v}(m-1), ..., \frac{F_s}{F_v}m + LF_s\right), \tag{7}$$

where $x_{U1}(n)$ and $x_{D1}(n)$ denote the sound signals replayed from the two selected loudspeakers and $x_{U2}(n)$ and $x_{D2}(n)$ denote the sound signals replayed from the two additionally selected loudspeakers for the horizontal panning.

B. DESIGN AND PROCEDURE

Nine subjects (ages: 27–38, five males and four females) with normal stereoscopic acuity and normal audibility participated as viewers in this experiment. Scheffé' paired comparison[9] was applied as an evaluation method. This experiment's flowchart is shown in Figure 8. First, we set two evaluation criteria: the degree of the coincidence of the sound location and the sound movement. The sound location's degree of coincidence denotes whether viewers feel that the sound of the stars and balls (the left-hand side of Figure 5) is always played at the position of the videos. The degree of the coincidence of the sound movement denotes whether viewers feel that the UFO's sound (the left-hand side of Figure 5) is always moving in concert with the video. We divided our experiment into eight sessions for evaluation criteria and viewing positions and randomized their presented orders for all viewers. Six practice trials and thirty main trials were performed in each session. The six practice trials were permutations of the three sound conditions shown in Figure 7(a), (c), and (f). The permutations of the six sound conditions shown in Figure 7 resulted in thirty main trials. The presentation orders of the trials were randomized for each viewer.

The viewers graded the degree of the coincidence of stimulus B in reference to stimulus A using the 7-step scale shown in Table 1. The viewers were allowed to freely move their heads and upper bodies while listening to the sounds.

C. RESULTS AND DISCUSSION

An analysis of variance (ANOVA) of this experiment's result was performed based on Scheffé's paired comparison of eight sessions: evaluation criterion (2) × viewing position (4). We found
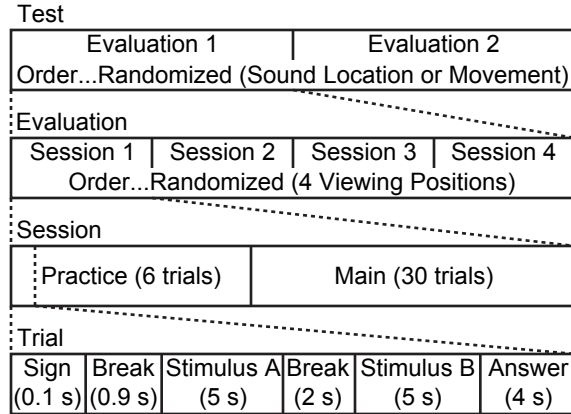
Test	
Evaluation 1	Evaluation 2
Order...Randomized (Sound Location or Movement)	

Evaluation			
Session 1	Session 2	Session 3	Session 4
Order...Randomized (4 Viewing Positions)			

Session	
Practice (6 trials)	Main (30 trials)

Trial					
Sign (0.1 s)	Break (0.9 s)	Stimulus A (5 s)	Break (2 s)	Stimulus B (5 s)	Answer (4 s)

Figure 8: Flowchart of first audio-visual experiment.

Table 1: Scale of Scheffé's paired comparison.

Grade	Judgment
3	Very good
2	Fairly good
1	Little good
0	The same
-1	Little bad
-2	Fairly bad
-3	Very bad

a significant main effect of the sound conditions at a 0.1% level except for one session (sound location, lateral viewing position). Thus, since there are significant differences among the sound conditions, we evaluated their effect based on the average grades calculated in each session.

In each evaluation criterion and viewing position, the average grades of all the sound conditions are shown in Figure. 9. The error bars denote 95% confidence intervals based on a yardstick. In sound condition (a), the average grades are significantly lower than other sound conditions except for one session (sound location, lateral viewing position), since the position of the sound images is biased to the right-and-left sides of the screen and viewers can clearly perceive the position differences between the 3D object and the sound image. Thus, we believe that viewers correctly discriminated the differences of the sense of presence.

We evaluated the effect of the horizontal panning on the sense of presence on the basis of a sound condition in which the average grade is highest in all the sound conditions (basic sound condition). The bold dashed lines in Figure 9 denote the border line based on the basic sound condition. The arrows in Figure 9 denote the significant differences among the basic sound condition at the 5% significant level. There are no significant differences among the basic sound conditions in all the sessions when the sound conditions range from (d) to (f). In other words, even if the
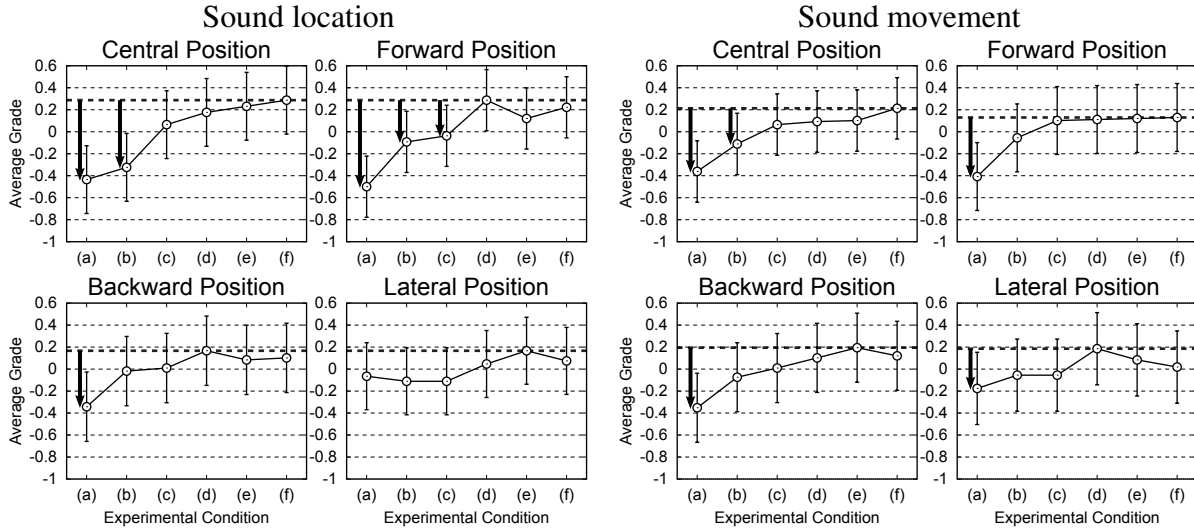
Figure 9: Results of first audio-visual experiment.

horizontal panning is added to the constructed system based on our proposed method, viewers cannot discriminate the differences of the sense of presence. Thus, we believe that the performance is maintained even if the hyperdirectional microphone array is applied when the teleconference system is constructed based on our proposed method.

However, when the number of loudspeakers is four in the sound condition (b), there are significant differences among the basic sound condition in three sessions. When the number of loudspeakers is six in the sound condition (c), there are significant differences among the basic sound condition in one session. Thus, it is difficult to reduce the number of loudspeakers from ten even if the horizontal panning is applied to our proposed method.

3. AUDIO-VISUAL EXPERIMENT 2

In the sound conditions of Section 2, loudspeakers were always placed at four vertex positions of the rectangle display because we assume that our proposed method can play the sound of such conventional audio systems as stereophonic audio. However, when our proposed method is applied.to the teleconference system, it is not necessary to place loudspeakers at four vertex positions of the rectangle display because our proposed method does not play the sound of conventional audio systems. In Section 3, we repeatedly perform an audio-visual psychological experiment and evaluate the effect of the sound conditions where loudspeakers are not placed at four vertex positions of the rectangle display.

A. ENVIRONMENT AND CONDITIONS

The experimental environment and the 3D video used in this experiment are the same as in the first audio-visual experiment described in Section 2.

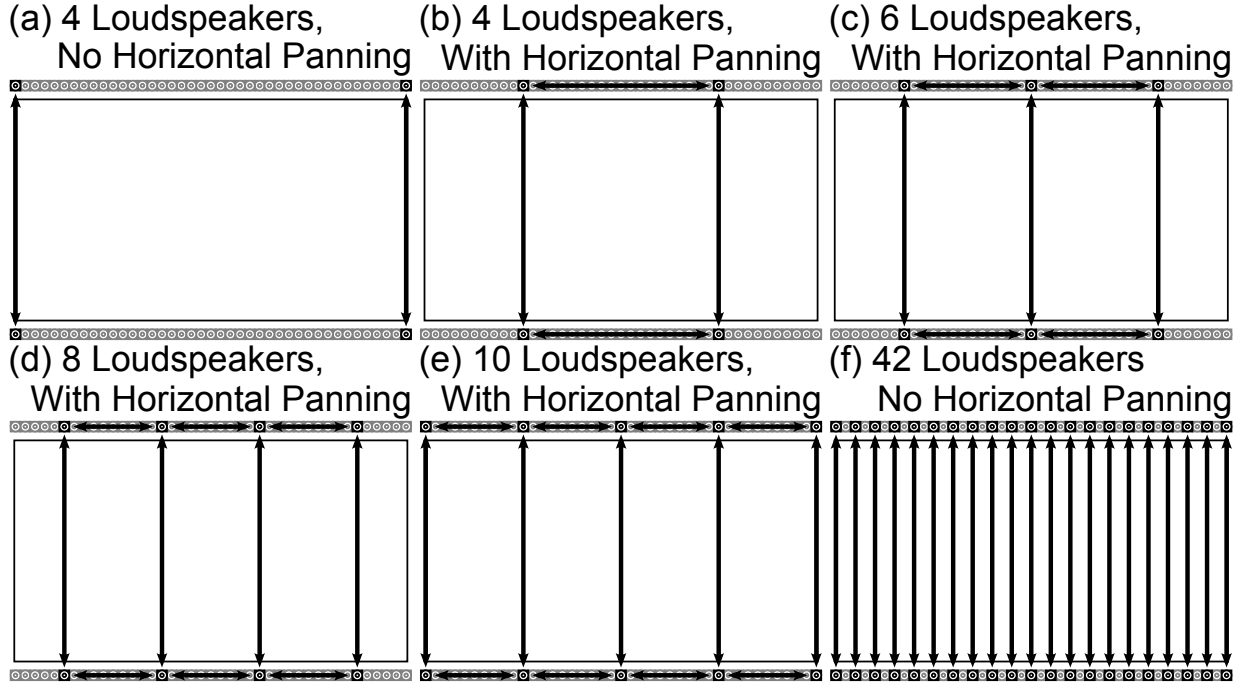The sound conditions used in this experiment are shown in Figure 10. The gray loudspeakers

Figure 10: Sound conditions in second audio-visual experiment.

denote the loudspeaker from which a sound is not replayed in each condition. The black arrows on the screen denote the horizontal and vertical panning in the replayed vertically panned loudspeakers. Sound conditions (e) and (f) are the condition where viewers could not discriminate the differences of the sense of presence in the first audio-visual experiment described in Section 2. On the other hand, sound condition (a) is the condition where viewers discriminated the differences of the sense of presence in the first audio-visual experiment. These conditions are set to evaluate the viewers' ability to discriminate the sense of presence in this experiment.

The synthesis procedure of playing sound in each sound condition is the same as that in the first audio-visual experiment described in Section 2. Note that the Δd_H (right-and-left intervals of the two selected loudspeakers) values are 4.4 m in sound conditions (a), 2.2 m in sound condition (b), 1.43 m in sound condition (c), 1.1 m in sound conditions (d) and (e), and 0.22 m in sound condition (f).

B. DESIGN AND PROCEDURE

Nine subjects (ages: 22–39, five males and four females) with normal stereoscopic acuity and normal audibility participated as viewers in this experiment. One male also participated in the first audio-visual experiment described in Section 2. Scheffé's paired comparison[9] was applied as an evaluation method. The experiment's flowchart, evaluation criterion, and evaluation procedure are the same as those in the first audio-visual experiment.
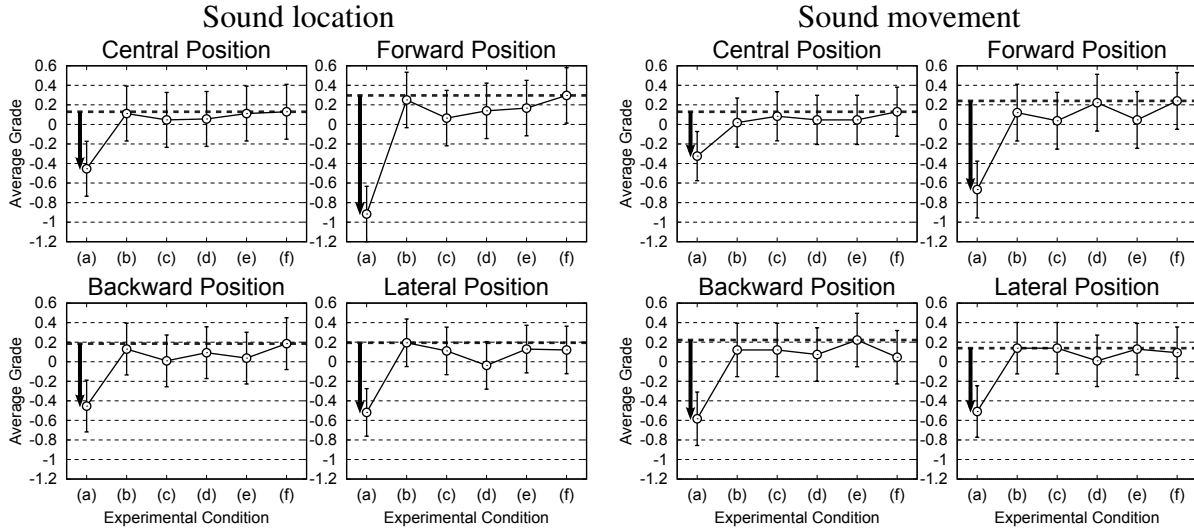
Figure 11: Results of second audio-visual experiment.

C. RESULTS AND DISCUSSION

An analysis of variance (ANOVA) of this experiment's result was performed based on Scheffé's paired comparison of eight sessions: evaluation criterion (2) × viewing position (4). We found a significant main effect of the sound conditions at a 0.1% level in all sessions. Thus, since there are significant differences among the sound conditions, we evaluated their effect based on the average grades calculated in each session.

In each evaluation criterion and viewing position, the average grades of all the sound conditions are shown in Figure 11. The error bars denote 95% confidence intervals based on a yardstick. In sound condition (a), the average grades are significantly lower than other sound conditions, since the position of the sound images is biased to the right-and-left sides of the screen and viewers can clearly perceive the position differences between the 3D object and the sound image. Thus, we believe that viewers correctly discriminated the differences of the sense of presence.

We evaluated the effect on the basis of a sound condition in which the average grade is highest in all the sound conditions (basic sound condition). The bold dashed lines in Figure 11 denote the border line based on the basic sound condition. The arrows in Figure 11 denote the significant differences among the basic sound condition at the 5% significant level. There are no significant differences among the basic sound condition in all the sessions when the sound conditions range from (b) to (f). In other words, when loudspeakers are not placed at four vertex positions of the rectangle display, viewers cannot discriminate the differences of the location and the movement of the sound images even if the number of loudspeakers is reduced to four. Thus, we believe that the number of loudspeakers can be reduced to four if loudspeakers are not placed at four vertex positions of the rectangle display when the teleconference system is constructed based on our proposed system.

4.   CONCLUSION

In this paper, in order to evaluate the effect of the horizontal panning on the sense of presence in our proposed method applied to the teleconference system, we performed two audio-visual psychological experiments.

In the first experiment, the effect of the horizontal panning was evaluated based on the condition used in the past study. As a result, although viewers cannot discriminate the differences of the sense of presence even if the horizontal panning was added to our proposed method, it is difficult to reduce the number of loudspeakers from ten.

In the second experiment, the effect of the horizontal panning was evaluated in the condition where loudspeakers are not placed at four vertex positions of the rectangle display. As a result, viewers cannot discriminate the differences of the sense of presence even if the number of loud-speakers is reduced to four.

Thus, when the teleconference system based on our proposed method is constructed, it was shown that the performance was adequately maintained if four loudspeakers were placed except four vertex positions of the rectangle display.

Future work needs to implement the teleconference system by constructing the hyperdirectional microphone array for recording in our proposed method and to evaluate the performance of the implemented system.

ACKNOWLEDGMENT

REFERENCES

[1] K. Enami, "Research on ultra-realistic communications," *ECTI Trans. Elect. Eng., Electro., Comm.*, **6**, 22-25 (2008).

[2] S. Iwasawa, M. Kawakita and N. Inoue, "REI: an Automultiscopic Projection Display," *Proc. Int. Conf. on 3D System and App.*, Selected paper 1, pp. 1-4 (2013).

[3] M. Kawakita, S. Iwasawa, G. Sabri and N. Inoue, "Development of Glasses-free 3D Video System," *Proceed Int. Univ. Comm. Symp.*, pp. 323-327 (2011).

[4] T. Kimura and H.Ando, "3D Audio System Using Multiple Vertical Panning for Large-screen Multiview 3D Video Display," *ITE Trans. on Media Tech. and App.*, **2**, 33-45 (2014).

[5] T. Kimura and H.Ando, "Effect of Discretization of Sound Source Position on Sense of Presence in 3D Audio Systems Using Multiple Vertical Panning," *Trans. VR Soc. Jpn.*, **20**, 179-188 (2015).

[6] Knowledge Capital, http://kc-i.jp/en/

[7] Grant Front Osaka, http://www.grandfront-osaka.jp/multilingulal/index

[8] Pocket Sound, http://pocket-se.info/.

[9] H. Scheffé, "An analysis of variance for paired comparisons," *J. Am. Stat. Assoc.* **47** 381-400 (1952).