Effect of the Horizontal Panning in 3D Audio System Based on Multiple Vertical Panning

Toshiyuki Kimura
Faculty of Engineering
Tohoku Gakuin University
Tagajo, Miyagi, Japan

Hiroshi Ando
Universal Communication Research Institute
NICT
Soraku-gun, Kyoto, Japan

Abstract—We previously proposed a 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 an audio-visual psychological experiment and evaluated the effect of the horizontal panning on the sense of presence in our proposed method. We found that the performance is maintained even if the hyperdirectional microphone array is applied as the recording unit when the teleconference system is constructed based on our proposed method.

Keywords- *multi-view 3D video display; multiple vertical panning; Scheffe's paired comparison; sound location; sound movement*

I. INTRODUCTION

At NICT, a multi-view 3D video display system (REI display) with a 200-inch screen has also been developed [1]. 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 and indicated that our proposed method was effective with such conventional audio as stereophonic [2]. We also showed that the number of loudspeakers can be reduced to ten in our proposed method [3].

In our previous studies, we assumed that the recording microphones were placed at the neighborhood of sound sources and that the position information of sound sources was 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 position information increases according to the number of participants. To solve these problems, it is necessary to develop the recording unit which can avoid the need to transmit the participants' position information.

We assume the hyperdirectional microphone array as the recording unit in order to apply our proposed method to the teleconference system. In this case, if there are no sound sources toward the direction of hyperdirectional microphones, neighboring microphones simultaneously record a sound. As a result, it is predicted that a horizontal panning adds to the recorded signals. In this paper, the effect of the horizontal panning on the sense of presence is evaluated by the audio-visual psychological experiment.

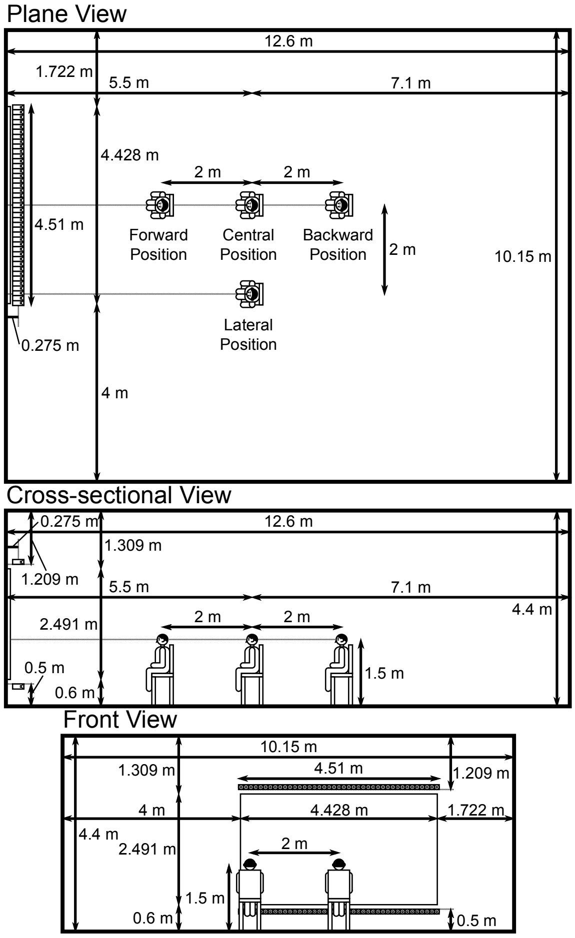

Figure 1. Position of viewers, screen and the loudspeaker array in the audio-visual experiment.

II. AUDIO-VISUAL EXPERIMENT

A. Environment and Conditions

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. Because polarization plates are placed in front of the projectors, viewers can see 3D video by wearing polarization glasses. The room's reverberation time was 402 ms, and the background noise level had an A-weighted level of 38 dB.

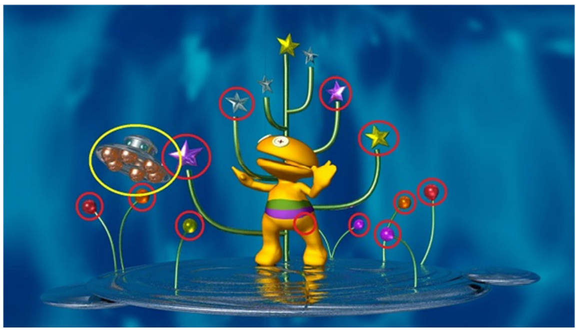

Figure 2. 3D video used in the audio-visual experiment.

We placed 82 loudspeakers in the room (Fig. 1). They were placed 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: FE103En) 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 (5.5 m), three viewing positions (forward, central, and backward) were set at 3.5, 5.5, and 7.5 meters from the screen. The viewing width of the developed system is 2 m across, centered around the front viewing position of the screen when the viewing distance is 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.

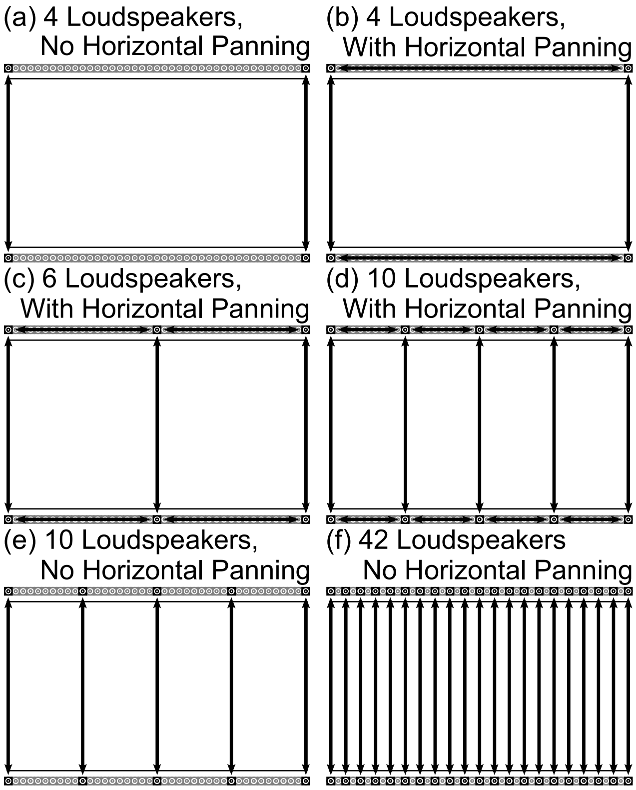

Figure 3. Sound conditions used in the audio-visual experiment.

The 3D video used in this experiment is shown in Fig. 2. In it, the UFO (inside the yellow oval in Fig. 2) that plays a sound is moving about the screen every five seconds. When it touches the stars and balls (inside the red circles in Fig. 2), the sound of the stars and balls is played at their positions. 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 the 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 Fig. 3. The gray loudspeakers denote the loudspeaker from which a sound was not replayed in each condition. The black arrows on the screen denote the direction of horizontal and vertical panning. 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 [3]. On the other hand, sound condition (a) is the condition where viewers discriminated the differences of the sense of presence in the previous study [3]. 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 (=$(m\text{-}1)/F_v$) 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:

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

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:

$$x_U(n) = a_U w(n) s(n),\ x_D(n) = a_D w(n) s(n)$$

$$\left(n = \frac{F_s}{F_v}(m-1), \ldots, \frac{F_s}{F_v} m + LF_s \right), \tag{3}$$

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:

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

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 [2] 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 [2]. Thus, α (=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:

$$\begin{Bmatrix} x_{U1}(n) \\ x_{D1}(n) \end{Bmatrix} = \cos\left(\frac{\pi}{2} \frac{P_H - P'_H}{\Delta d_H} \right) w(n) s(n) \begin{pmatrix} a_U \\ a_D \end{pmatrix},$$

$$\begin{Bmatrix} x_{U2}(n) \\ x_{D2}(n) \end{Bmatrix} = \cos\left(\frac{\pi}{2} \frac{P_H - P''_H}{\Delta d_H} \right) w(n) s(n) \begin{pmatrix} a_U \\ a_D \end{pmatrix},$$

$$\left(n = \frac{F_s}{F_v}(m-1), \ldots, \frac{F_s}{F_v} m + LF_s \right), \tag{7}$$

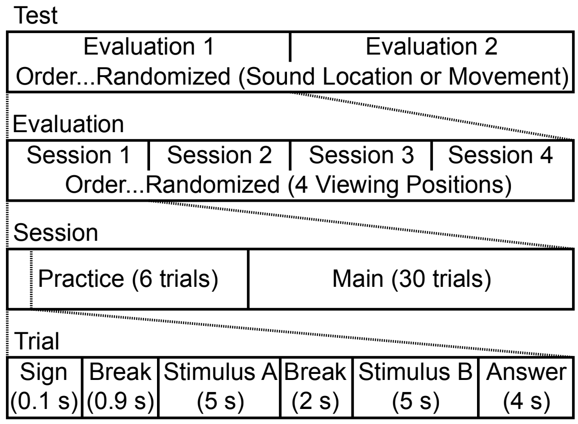

Figure 4. Flowchart of the audio-visual experiment.

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's paired comparison [4] was applied as an evaluation method. This experiment's flowchart is shown in Fig. 4. 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 (Fig. 2) 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 (Fig. 2) 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 Fig. 3(a), (c), and (f). The permutations of the six sound conditions shown in Fig. 3 resulted in thirty main trials. The presentation orders of the trials were randomized for each viewer.

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

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 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 Figs 5 & 6. 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). 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 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 as the recording unit when the teleconference system is constructed based on our proposed method.

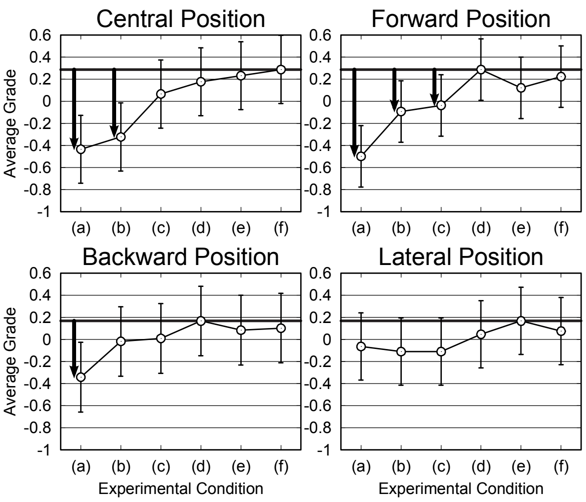

Figure 5. Results of the audio-visual experiment (Sound location).

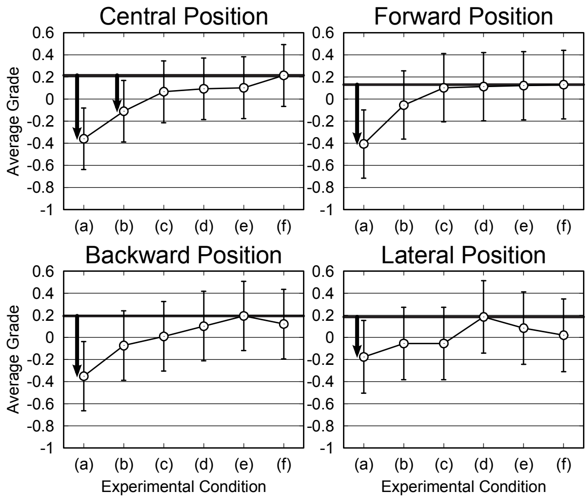

Figure 6. Results of the audio-visual experiment (Sound movement).

On the other hand, in sound conditions (b) and (c), average grades are significantly lower than the basic sound conditions in a subset of sessions. Thus, it is difficult to reduce the number of loudspeakers even if the horizontal panning is applied to our proposed method.

III. CONCLUSION

In this paper, we performed an audio-visual psychological experiment and evaluated the effect of the horizontal panning on the sense of presence in our proposed method. We found that the performance is maintained when the teleconference system is constructed based on our proposed method. However, it is difficult to reduce the number of loudspeakers even if the horizontal panning is applied to our proposed method.

Future work must study the placement of loudspeakers in order to reduce the number of loudspeakers.

ACKNOWLEDGMENT

The authors thank Dr. S. Iwasawa and Dr. S. Enomoto for constructing the environment of our audio-visual experiment and Dr. M. Makino for depicting the 3D videos in it. The audio-visual experiment in this paper was performed with the approval of the ethics committee of the National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

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

[2] 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., vol. 2 no. 1, pp. 33-45, January 2014.

[3] T. Kimura and H. Ando, "Effect of the Number of Loudspeakers on Sense of Presence in 3D Audio System Based on Multiple Vertical Panning," Proc. Forum Acusticum, No. R05D_6, pp. 1-5, September 2014.

[4] H. Scheffe, "An analysis of variance for paired comparisons;" J. Am. Stat. Assoc., vol. 47, pp. 381-400, 1952.