Effect of the number of loudspeakers on sense of presence in 3D audio system based on multiple vertical panning

Toshiyuki Kimura and Hiroshi Ando
Universal Communication Research Institute, National Institute of Information and Communications Technology, Kyoto, Japan.

Summary
NICT has developed a multiview 3D video display system with a 200-inch screen. In this system, several viewers can simultaneously view natural 3D objects from their respective viewing positions without special glasses. We have proposed a 3D audio system using multiple vertical panning, which is compatible with a large-screen multiview 3D video display system, that has following four features: 1) Multiple viewers can simultaneously feel the sound images at the position of the 3D objects depicted by the 3D video display system regardless of the viewing position; 2) Viewers do not need to wear hearing devices; 3) There are no sound devices between the projector array and the viewing position; 4) No recording microphones need to be placed between the projector array and the viewing position. However, because the loudspeakers are densely placed at the horizontal direction of the screen in this proposed system, their number must be reduced for practical use. In this paper, we performed an audio-visual psychological experiment to develop a practical system. 3D videos were presented to viewers wearing glasses and six sound stimuli were randomly presented by 4, 6, 10, 12, 22, and 42 loudspeakers. Nine viewers compared the sound stimuli according to two evaluation criteria: the position and the movement of sound images. Our results show that viewers failed to discriminate the difference of the location and the movement of sound images with ten or more loudspeakers. We conclude that a practical system can be constructed with our proposed system with ten loudspeakers.

PACS no. 43.60.Dh, 43.66.Qp

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 multiview 3D video display system with a 200-inch screen has been developed [2]. Its 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). Since this system allows viewers to view parallax videos from a horizontal position, several viewers can simultaneously view natural 3D objects based on each particular viewing position without special glasses. However, the developed system only presents visual sensations. To achieve realistic auditory sensations, a 3D audio system must be developed that corresponds to a large-screen multiview 3D video display system.

We have proposed a novel 3D audio system using the multiple vertical panning (MVP) method to develop a 3D audio system that matches our developed 3D video display system. The results of our audio-visual psychological experiment indicated that our proposed system was effective with such conventional audio systems as stereophonic audio [4]. However, because the loudspeakers are densely placed horizontally to its screen, their number is enormous. Thus, we must reduce them before our proposed system can be put into practical use. In this paper, we evaluated the effect of the number of loudspeakers on the sense of presence by performing an audio-visual psychologi-
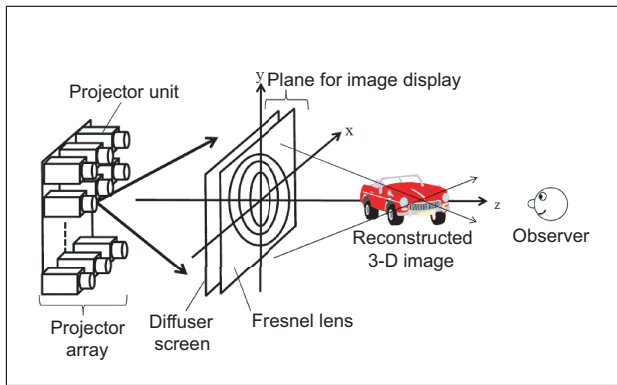
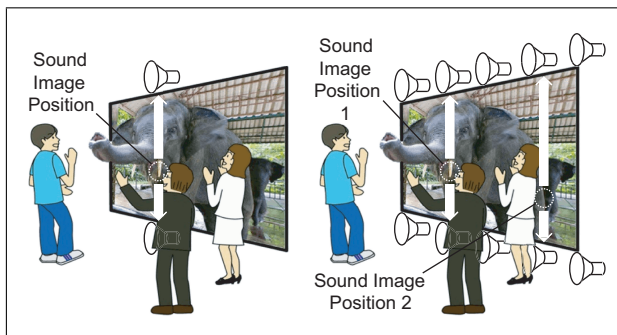Figure 1. Basic configuration of a large-screen multiview 3D video display system [3].

Figure 2. Basic configuration of multiple vertical panning (MVP) method [4].

cal experiment and developed a practical audio-visual system using our proposed system.

2. Diagram of proposed system [4]

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 are expected to 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), we expect that multiple viewers can perceive a sound image at the position of the 3D object, regardless of the 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 feel the sound images at the position of the 3D objects depicted by the 3D video display system, regardless of the viewing position.

In the proposed system, viewers do not need to wear hearing devices. Because the loudspeakers are placed at the upper and lower sides of the screen, there are no sound devices between the projector array and the viewing position. Since this system only has to directly record the speech of the participants in teleconferences and does not restrict the position of the recording microphones, it does not need to place recording microphones between the projector array and the viewing position.

3. Audio-visual experiment

3.1. 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.

We placed 42 loudspeakers in the room (Figure 3). 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 multiview 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.

The 3D video used in this experiment is shown in Figure 4. In it, the UFO (inside the yellow oval in Figure 4) that plays a sound is moving about the screen every five seconds. When it touches the stars and balls (inside the red circles in Figure 4), 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 Figure 5. The gray loudspeakers denote the loudspeaker from which
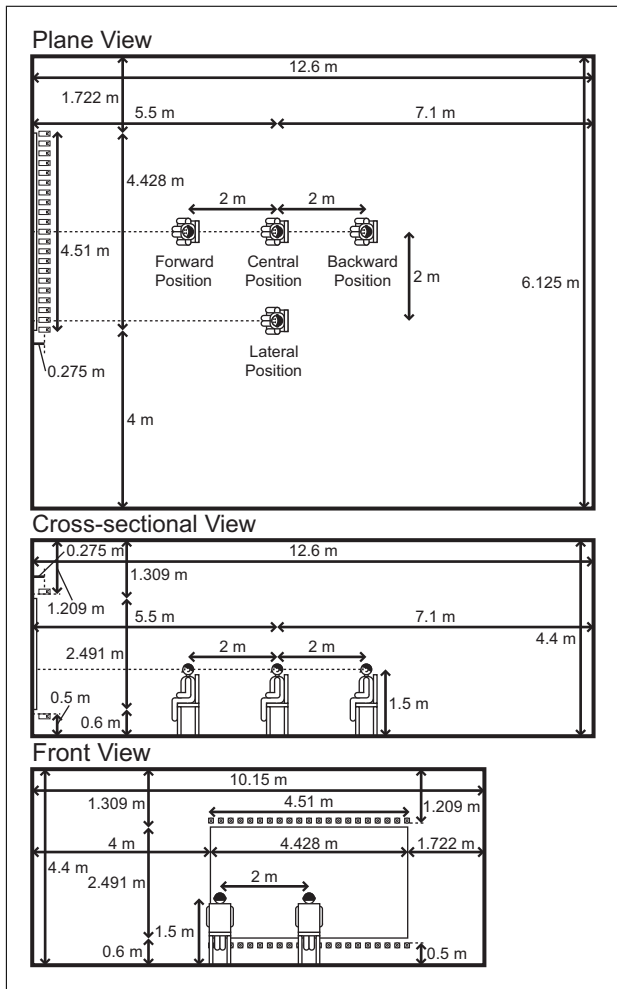
Figure 3. Position of viewers, screen, and loudspeaker array in audio-visual experiment.
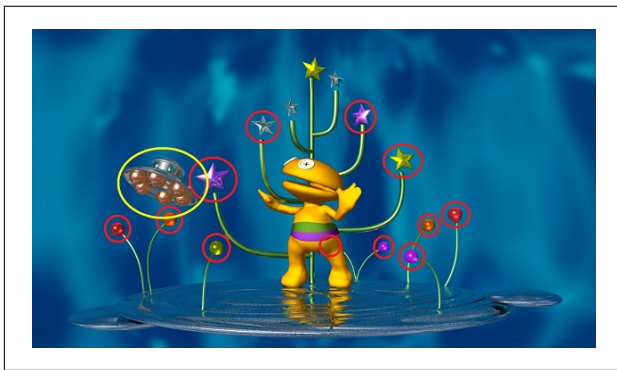

Figure 4. 3D video used in audio-visual experiment.

a sound is not replayed in each condition. 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 $F_v(=30 \text{ fps})$ and $m(=1,...)$ denote the frame rate and the frame index of the video signals. $P_H(= -2.214\sim2.214)$ and $P_V(= -1.2455\sim1.2455)$ 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
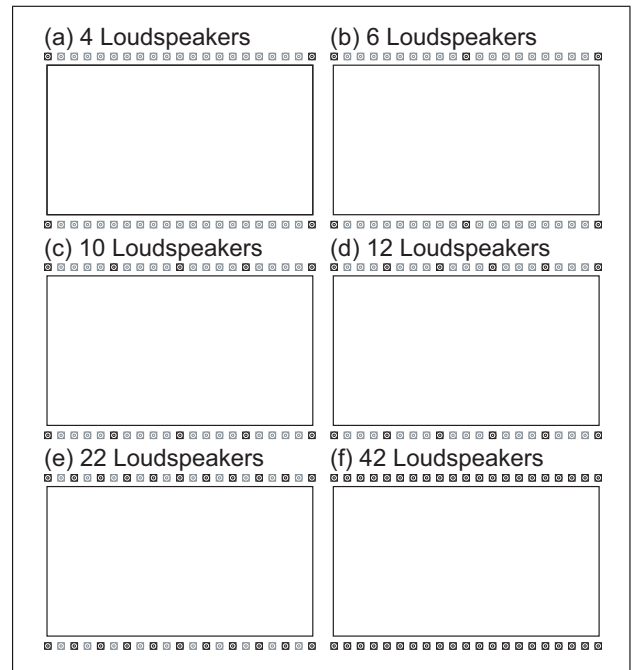

Figure 5. Sound conditions used in audio-visual experiment.

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:

$$P'_H = \Delta d \, \text{round} \left(\frac{P_H + 2.2}{\Delta d} \right) - 2.2, \qquad (1)$$

where $P'_H(= -2.2, ..., 2.2)$ denotes the horizontal position of the two selected loudspeakers. Δd denotes their right-and-left intervals. In this experiment, the Δd values are 4.4, 2.2, 1.1, 0.88, 0.44 and 0.22 m in an order corresponding to the sound conditions shown in Figure 5.

Second, the sound calculated from the sound source signal, $s(n)$, is replayed from two selected loudspeakers:

$$x_U(n) = a_U w(n) s(n), \qquad (2)$$
$$x_D(n) = a_D w(n) s(n), \qquad (3)$$
$$\left(n = \frac{F_s}{F_v}(m-1), ..., \frac{F_s}{F_v}m + L F_s\right)$$

where $F_s(=48 \text{ kHz})$ 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:

$$w(n) = \begin{cases} \frac{1}{LF_s}\{n - \frac{F_s}{F_v}(m-1)\} \\ \quad (n = \frac{F_s}{F_v}(m-1), ..., \frac{F_s}{F_v}(m-1) + LF_s) \\ 1 \quad (n = \frac{F_s}{F_v}(m-1) + LF_s, ..., \frac{F_s}{F_v}m) \\ -\frac{1}{LF_s}(n - \frac{F_s}{F_v}m) + 1 \\ \quad (n = \frac{F_s}{F_v}m, ..., \frac{F_s}{F_v}m + LF_s) \end{cases},$$

(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:

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

(5)

$$a_D = \frac{1}{\sqrt{10^{\frac{\Delta A}{10}} + 1}}.$$

(6)

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

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

(7)

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.

3.2. Design and procedure

Nine subjects (ages: 29–39, five males and four females) with normal audibility participated as viewers in this experiment. Scheffé's paired comparison [5] was applied as an evaluation method. This experiment's flowchart is shown in Figure 6. 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 (Figure 4) 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 (Figure 4) 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 5 ((a), (b), and (f)). The permutations of the six sound conditions shown in Figure 5 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 I. The viewers were allowed to freely move their heads and upper bodies while listening to the sounds.
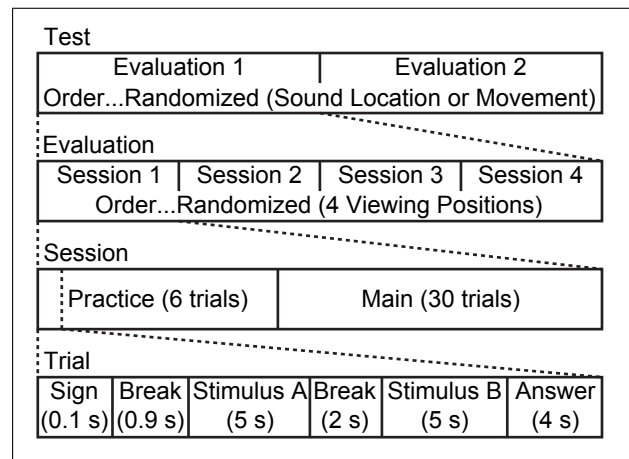

Figure 6. Flowchart of audio-visual experiment.

Table I. 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

3.3. 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, backward viewing position). Thus, since there are significant differences among the number of loudspeakers, 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 Figures 7-8. The error bars denote 95% confidence intervals based on a yardstick. In the sound condition (a), the average grades are significantly lower than other sound conditions except the 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 difference between the 3D object and the sound image. Viewers correctly discriminated the differences of the sense of presence.

We evaluated the effect of the number of loudspeakers 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). With four or six loudspeakers, sometimes the average grades were significantly lower (5% level) than the basic sound condition. On the other hand, we found no
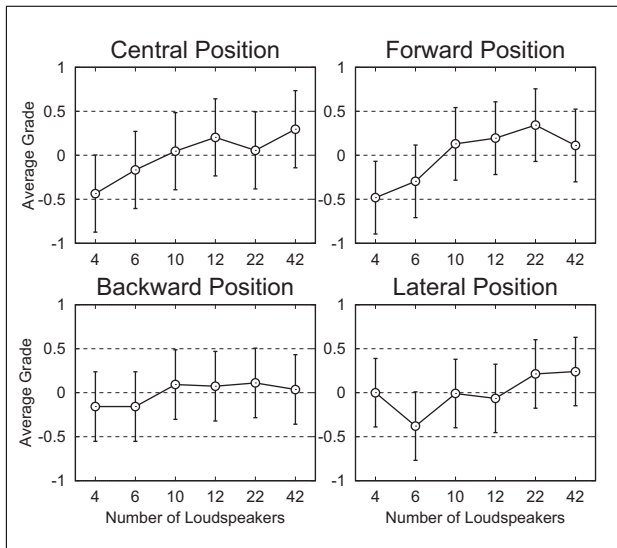
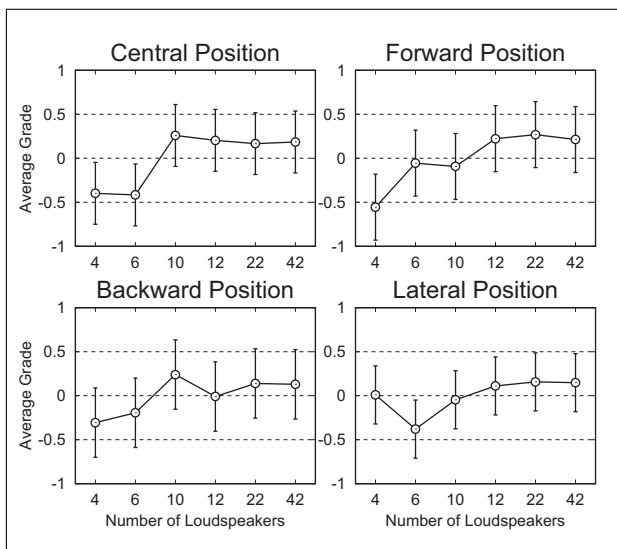Figure 7. Results of audio-visual experiment: sound location.

Figure 8. Results of audio-visual experiment: sound movement.

significant differences among the basic sound conditions in all the sessions with 10, 12, 22, or 42 loudspeakers. When the number of loudspeakers equals or exceeds ten, viewers cannot discriminate the differences of the location and the movement of sound images even with more loudspeakers. Thus, it is considered that the number of loudspeakers can be reduced to ten when an audio-visual system is based on our proposed system.

4. Conclusion

In this paper, to reduce the required number of loudspeakers in our previously proposed system, we performed an audio-visual psychological experiment and evaluated the effect of the number of loudspeakers on the sense of presence. We found that viewers could not discriminate the differences of the sense of presence with ten or more loudspeakers.

To construct a practical recording method for our proposed system, future work must perform an additional audio-visual psychological experiment and evaluate the effect of the number of discretized vertical positions of sound sources on the sense of presence.

References

[1] K. Enami : Research on Ultra-realistic Communications. ECTI Transactions on Electrical Engineering, Electronics, and Communications **6** (2008) 22-25.

[2] S. Iwasawa and M. Kawakita: Quantifying Capabilities of the Prototype 200-inch Automultiscopic Display. Proc. International Conference on 3D Systems and Applications (2012) 105-109.

[3] M. Kawakita, S. Iwasawa, G. Sabri and N. Inoue: Development of Glasses-free 3D Video System. Proc. International Universal Communication Symposium, (2011) 323-327.

[4] T. Kimura and H.Ando: 3D Audio System Using Multiple Vertical Panning for Large-screen Multiview 3D Video Display. ITE Transactions on Media Technology and Applications **2** (2014) 33-45.

[5] H. Scheffé: An analysis of variance for paired comparisons. Journal of the American Statistical Association **47** (1952) 381-400.