Audio-visual Experiment for 3D Audio System Based on Multiple Vertical Panning

Toshiyuki Kimura and Hiroshi Ando
Universal Communication Research Institute
National Institute of Information and Communications Technology
Seika-cho, Soraku-gun, Kyoto, Japan

***Abstract*—** In this paper, in order to evaluate the audio-visual performance of the proposed three-dimensional (3D) audio system, which is based on Multiple Vertical Panning (MVP) method and matches to the glasses-free 3D video display system in which the size of screen is very large, the audio-visual experiment was designed by using the loudspeaker array in which eighty-two loudspeakers were placed on the upper and lower sides of the 200-inch screen. The results of this experiment show that the proposed 3D audio system was effective as compared with a conventional system such as stereophonic audio because viewers could always feel the synthesized sound images at the position of the 3D video object when the 3D sound of the proposed system was presented with 3D video.

Keywords- *ultra-realistic communication; 3D audio system; vertical panning; glasses-free 3D video display system; audio-visual experiment*

I. INTRODUCTION

At the NICT, a glasses-free 3D video technique using a projector array has been proposed and a multiview 3D video display system, in which the size of a screen is 70 inches [1] or 200 inches [2], has been developed. However, the developed system presents only visual sensations. In order to achieve a more realistic sensation, a system must be developed that can present multimodal sensations. Particularly, a 3D audio system must be developed that corresponds with a large-screen multiview 3D video display system in order to present realistic auditory sensations.

We have proposed the novel 3D audio system based on multiple vertical panning (MVP) method [3]. The basic configuration of the proposed system is shown in Figure 1. First, as shown in the left-hand side of Figure 1, two loudspeakers (called "vertically panned loudspeakers") are placed at the upper and lower sides of the position of the 3D object. If a sound is played from two loudspeakers using the panning between two loudspeakers (called "vertical panning"), viewers perceive that there is a sound image between two loudspeakers. Second, as shown in the right-hand side of Figure 1, 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 3D objects depicted by the 3D video display system, regardless of the viewing position.

This paper describes the audio-visual experiment used to evaluate the audio-visual performance of the proposed system. The results of this experiment indicate that the proposed system is effective as compared with conventional 3D audio systems such as stereophonic audio.

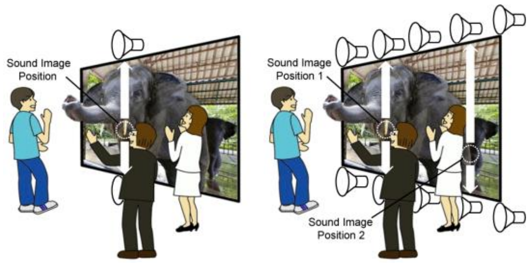

Figure 1. Basic configuration of the 3D audio system

II. AUDIO-VISUAL EXPERIMENT

A. *Environment and conditions*

The experiment was performed in a room where a 200-inch rear-projection visual screen was set up. Two projectors for the 2D video of the left and right eyes are set up behind the screen. Because polarization plates are set up in front of the projectors, viewers can see the 3D video by wearing the polarization glass. The reverberation time of the room was 258 ms, and the background noise level had an A-weighted level of 41 dB.

Loudspeakers were manufactured by mounting a loudspeaker unit (Fostex: FE103En) on a loudspeaker enclosure (width: 11 cm, depth: 25 cm, height: 11 cm). In order to place manufactured loudspeakers at the upper and lower sides of the screen densely in the horizontal direction, eighty-two loudspeakers were placed as shown in Figure 2. The total width of the two loudspeaker arrays was 4.51 m (= 11 cm × 41). The viewing distance was set at 5.2 m from the screen. Two viewing positions were set at a front position from the screen and at a lateral position, which was 2 m to the left of the front position. The height of two viewing positions was set to 1.4 m at the ear position of viewers. In addition, in order to compare the proposed system with the conventional system, two loudspeakers were placed at the left and right sides of the screen. The height of these loudspeakers was also 1.4 m. The sound pressure level was set to an A-weighted level of approximately 70 dB in the front viewing position.

Two types of sounds (white noise and speech) were used as the sound source. The experimental conditions in this experiment are shown in Figure 3. The gray loudspeakers denote the loudspeaker from which a sound is not replayed in each condition. In both conditions, grid points and lines are always presented to the screen as a 2D image. In the proposed system condition (a), two loudspeakers placed at the upper and lower sides of the screen were selected according to the horizontal position of the presented sound images, and the sound calculated from the sound source signal, $s(n)$, was

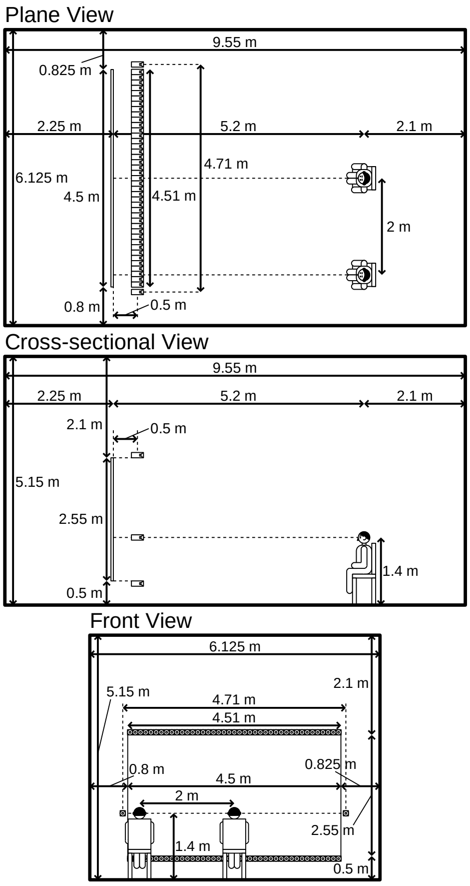

Figure 2. Position of viewers, screen, and the loudspeaker array

replayed from two selected loudspeakers according to the following equations:

$$x_U(n) = a_U s(n),\ x_D(n) = a_D s(n), \tag{1}$$

where $x_U(n)$ and $x_D(n)$ denote the sound signals replayed from two loudspeakers of the upper and lower sides, respectively, and a_U and a_D (the gain coefficients in each sound signal) are calculated from the level difference ,ΔA [dB], as follows:

$$a_U = 10^{\Delta A/20} \Big/ \sqrt{10^{\Delta A/10}+1},\ a_D = 1 \Big/ \sqrt{10^{\Delta A/10}+1}. \tag{2}$$

In this experiment, the level difference, ΔA, was set according to the past study [3] as follows:

$$\Delta A = P_V + 0.1437/0.1065, \tag{3}$$

where P_V denotes the vertical position of sound images. The height of the sound images is the same as that of the ear position of the viewers if P_V is -0.33. In the stereo condition (b), according to the conventional stereophonic audio, the sound calculated from the sound source signal, $s(n)$, was replayed from two loudspeakers on the left and right as follows:

$$x_L(n) = a_L s(n),\ x_R(n) = a_R s(n), \tag{4}$$

where $x_L(n)$ and $x_R(n)$ denote the sound signals replayed from two loudspeakers on the left and right side, respectively, and a_L and a_R (the gain coefficients in each sound signal) are calculated according to the tangent law [4] as follows:

$$a_L = (1-x) \Big/ \sqrt{2(1+x^2)},\ a_R = (1+x) \Big/ \sqrt{2(1+x^2)}, \tag{5}$$

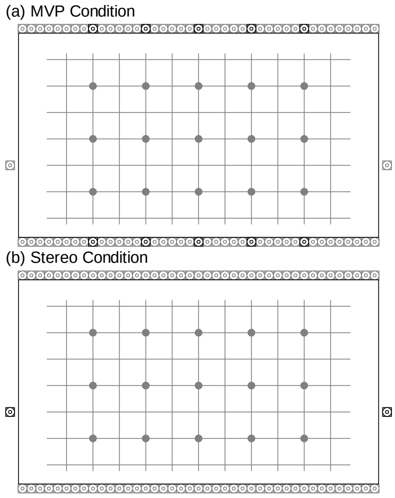

Figure 3. Sound conditions used in the audio-visual experiment

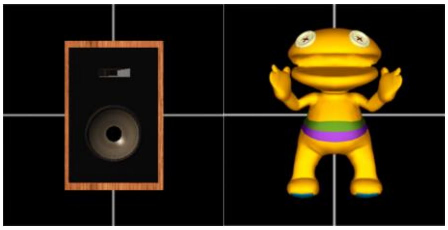

Figure 4. 3D videos (Left: white noise, Right: speech).

where x = tan θ/tan θ_0 and θ and θ_0 denote the angles of the sound images and the two loudspeakers on the left and right in the front viewing position, respectively. In this experiment, the angles were set as follows:

$$\tan\theta / \tan\theta_0 = P_H / 2.355, \tag{6}$$

where P_H denotes the horizontal position of a sound image. If P_H is 0, the sound image is placed at the center of the screen.

3D videos shown in Figure 4 were simultaneously presented in addition to the sound. The equipment for playing 3D videos was synchronized with audio-playing devices by Longitudinal Timecode (LTC). The 3D videos of the loudspeaker (left-hand side of Figure 4) and the character (right-hand side of Figure 4) were presented when the sound sources were white noise and speech. 3D videos were depicted by MAYA software [5]. The proper viewing distance and the parallax of 3D videos were 5.5 m and 0.0625 m, respectively. Thus, four experimental conditions listed in Table 1 were set in this experiment.

TABLE 1. EXPERIMENTAL CONDITIONS

Index	Sound	3D Video
(I)	Stereo	Sound only
(II)	Proposed system	Sound only
(III)	Stereo	Sound & video
(IV)	Proposed system	Sound & video

B. *Design and procedure*

Twelve subjects (age: 21-40, six males and six females), whom the audibility was normal in daily life, participated as viewers in this experiment. The flowchart of the audio-visual experiment is shown in

Figure 5. The experiment was divided into four sessions for viewing positions and sound sources. The presented orders of the sessions were counterbalanced in all viewers. Ten practice trials and one hundred main trials were performed in each session. During the main trials, rest periods were allowed after every set of fifty trials. The presentation orders of the trials were randomized for each viewer. The position of the sound images and the detail of the practice and main trials are shown in Tables 2 and 3, respectively.

TABLE 2. SOUND IMAGE POSITIONS

Index	P_H	P_V	Index	P_H	P_V	Index	P_H	P_V
1	-1.32	0.66	6	-1.32	0	11	-1.32	-0.66
2	-0.66	0.66	7	-0.66	0	12	-0.66	-0.66
3	0	0.66	8	0	0	13	0	-0.66
4	0.66	0.66	9	0.66	0	14	0.66	-0.66
5	1.32	0.66	10	1.32	0	15	1.32	-0.66

TABLE 3. PRACTICE AND MAIN TRIALS

	Element	3D Video
Practice (10)	= 2 conditions	(II) & (IV) of Table 1
	× 5 positions	1, 5, 8, 11 & 15 in Table 2
Main (100)	= [1 condition	(I) in Table 1
	× 5 positions	P_H = -1.32, -0.66, 0, 0.66, 1.32
	+ 3 conditions	(II)-(IV) in Table 1
	× 15 positions]	1-15 in Table 2
	× 2 repetitions	

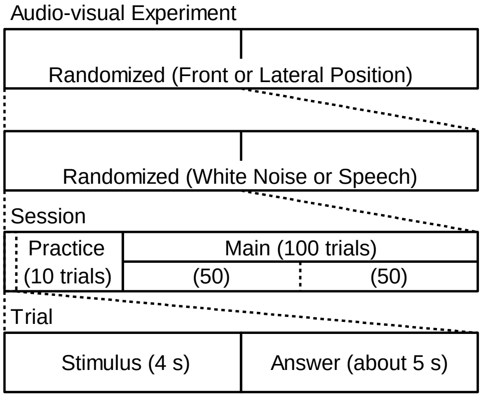

Figure 5. Flowchart of the audio-visual experiment

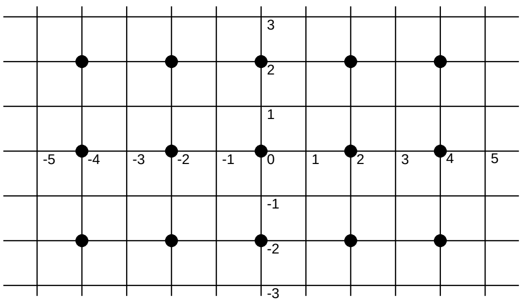

Figure 6. Relation between perceived positions and answer grids

The viewers were instructed to report the perceived position of the sound images by listing the indexes of the positions in an answer sheet. Note that the viewers were instructed to gaze at a 3D object when the 3D video and the sound were presented. The relation between the perceived position and the answer grids is shown in Figure 6. This grid corresponds to the grid lines and points shown in Figure 3. The viewers could choose a horizontal index of 11 patterns (from -5 to 5) and a vertical index of 7 patterns (from -3 to 3). If viewers perceived multiple sound images in the trials, the viewers could list multiple indexes in an answer sheet. The viewers were allowed to move their heads and upper bodies freely while listening to the sounds.

C. *Results and discussion*

After reducing the answers where viewers listed multiple indexes, the averages of the horizontal and vertical indexes were calculated from the answer indexes of viewers. Results of the averages in each experimental condition are shown in Figs. 7-10. Error bars of horizontal and vertical directions denote the 95% confidence interval of the averages of the horizontal and vertical indexes. Note that results of white noise source are only plotted. Because the gray circles denote the presented positions of the sound images, it is shown that viewers accurately feel sound images at the presented position if the perceived positions of sound images are close to the gray circles.

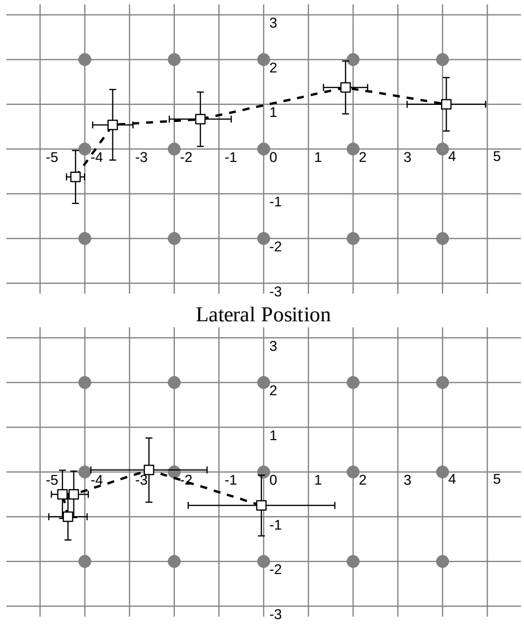

Figure 7. Results of condition (I)

When the sound is only presented in the stereo condition, although the horizontal localized positions of the sound images are generally accurate in the front viewing position, the vertical localized positions of the sound images are higher than the input position. In the lateral viewing position, the horizontal and vertical positions of the sound images are not accurately localized. This is attributed to the fact that the stereophonic system assumes that viewers listen to a sound in the front viewing position.

On the other hand, in the proposed system, the horizontal localized accuracy of the sound images is improved in the lateral viewing position. The vertical localized accuracy of sound images is improved in the front viewing position. Thus, when only the sound is presented in the proposed system, it is indicated that the localized accuracy of the sound images is improved as compared with the stereophonic system at any viewing position.

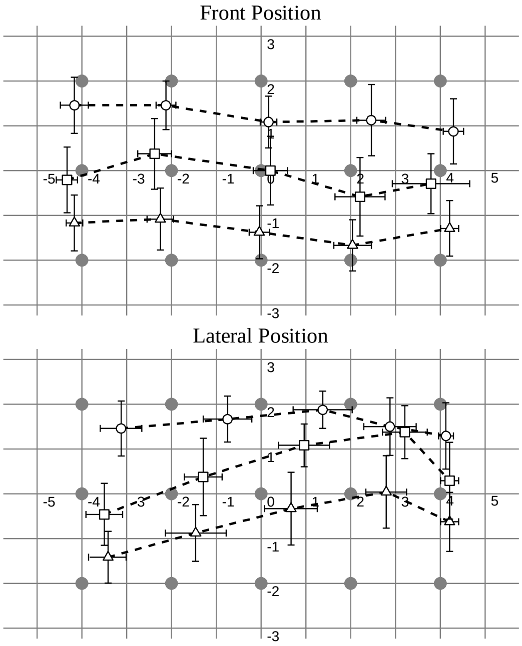

Figure 8. Results of condition (II)

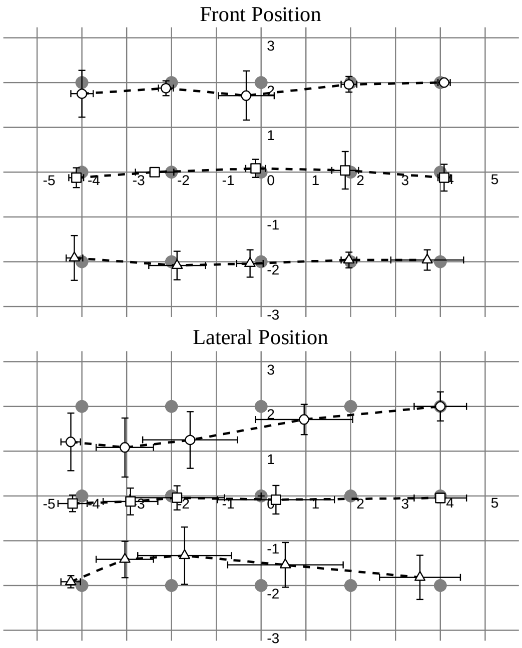

Figure 9. Results of condition (III)

When the sound and 3D video are presented, the localized position of the sound images is generally accurate in the stereo condition because of the ventriloquism effect if the viewing position is frontal. However, in the lateral viewing position, the localized position of the sound images is biased to the left side because the ventriloquism effect does not occur.

On the other hand, the localized position of the sound images is the same as the position of 3D objects in the lateral viewing position in the proposed system. Thus, when the sound and 3D video are presented, it is indicated that the proposed system is effective as compared with a conventional system such as stereophonic audio because viewers can always perceive the sound images at the position of the 3D objects at any viewing position.

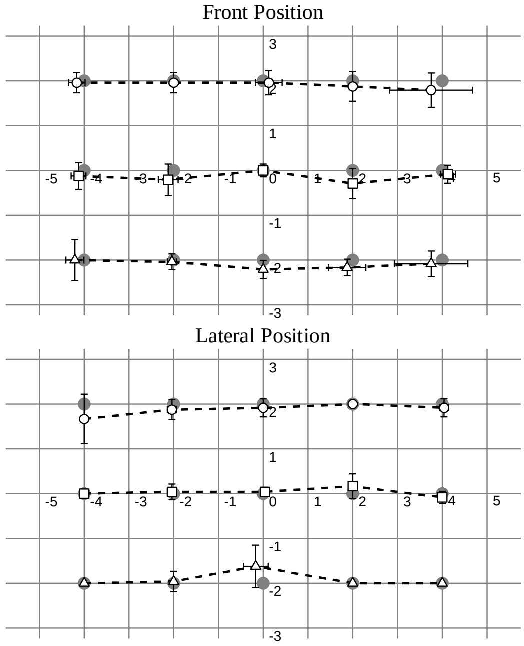

Figure 10. Results of condition (IV)

III. Conclusion

In this paper, in order to evaluate the audio-visual performance of the proposed system, an audio-visual experiment was performed by using a loudspeaker array in which eighty-two loudspeakers were placed at the upper and lower sides of the 200-inch screen. As a result, the proposed 3D audio system was effective as compared with a conventional system such as stereophonic audio because viewers could always perceive the synthesized sound images at the position of the 3D object at any viewing position when the sound of the proposed system was presented with 3D video.

In future work, the feasibility of a practical realization of the proposed system should be studied by reducing the number of loudspeakers and by constructing the method of recording and transmission. The means of expression of the 3D sound distance by changing sound pressure amplitudes should also be developed.

References

[1] M. Kawakita, S. Iwasawa, G. Sabri and N. Inoue, “Development of Glasses-free 3D Video System,” Proc. IUCS, pp. 323-327, October 2011.

[2] H. Ando, M. Kawakita and S. Iwasawa, “The World's Largest 200-inch Glasses-free 3D Display Has Been Successfully Developed,” NICT Press Release, January 2011.

[3] T. Kimura and H. Ando, “Listening Test for 3D Audio System Based on Multiple Vertical Panning,” Proc. Acoustics, No. 1pHT6, pp. 1-6, Hong Kong, China, May 2012.

[4] B. Bernfeld, “Attempts for Better Understanding of the Di-rectional Stereophonic Listening Mechanism,” Proc. 44th Audio Eng. Soc. Convention, C-4, pp.1-24, February 1973.

[5] Website of Maya Software, http://www.autodesk.co.jp/maya.