Listening test for three-dimensional audio system based on multiple vertical panning

Toshiyuki Kimura and Hiroshi Ando[1]

[1] *Universal Communication Research Institute,*

National Institute of Information and Communications Technology,

2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto JAPAN

Abstract: In this paper, the novel three-dimensional (3D) audio system is proposed. The proposed system 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 vertical position of sound images is synthesized by the panning between two loudspeakers placed at the top and bottom of screen. The horizontal position of sound images is controlled by the position of two loudspeakers. By the proposed system, multiple listeners can simultaneously feel the sound images at the position of 3D objects depicted by the video display system. In order to evaluate the auditory performance of the proposed system, the listening test was designed by using the loudspeaker array in which twenty-seven loudspeakers were aligned on the vertical line. Sound images were synthesized by the panning between two loudspeakers placed at the top and bottom of the loudspeaker array. Twelve listeners listened to a sound and reported the position of synthesized sound images. As a result, it was indicated that listeners could feel the synthesized sound images at the position between two loudspeakers placed at the top and bottom of the loudspeaker array.

Keywords: 3D audio system, Vertical panning, Listening test

1. Introduction

Ultra-realistic communications techniques have been investigated in NICT[1]. If, by applying these techniques, realistic three-dimensional (3D) video and audio can appear in a 3D space, this will enable more realistic forms of communication (e.g., 3D television and 3D teleconference) than those currently offered by conventional video and audio techniques (HD video and 5.1-channel audio).

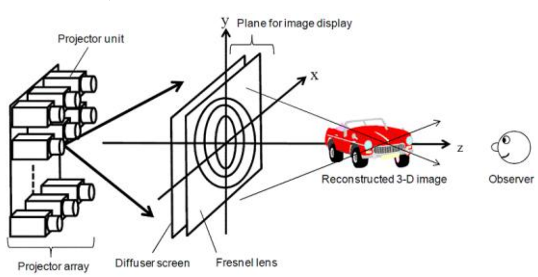

Figure 1 – Basic configuration of a large glasses-free 3D video display system[2]

In NICT, a glasses-free 3D video technique using projector array has been proposed and a large glasses-free 3D video display system, in which the size of a screen is 70 inches[2] and 200 inches[3], has been developed. The basic configuration of the developed system 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 to the horizontal direction

[1] {t-kimura,h-ando}@nict.go.jp

due to the diffusion characteristics of a diffuser screen placed in front of the Flesnel lens (small diffusion angle in the horizontal direction and wide diffusion angle in the vertical direction). As a result, since this system allows observers to observe parallax videos according to the horizontal position, several people can observe natural 3D objects at the same time according to each person's particular viewing position without needing special glasses.

In this paper, the novel 3D audio system based on Multiple Vertical Panning (MVP) method is proposed in order to match to the developed 3D video display system.

2. Diagram of proposed system

The basic configuration of the proposed system is shown in Figure 2. First, as shown in the left side of Figure 2, two loudspeakers are placed at the top and bottom of the position of a 3D object. Two loudspeakers placed at the top and bottom of the screen because the depth of 3D objects depicted by the developed 3D video display system is close to the position of the Flesnel lens in Figure 1. If a sound is played from two loudspeakers by the panning between two loudspeakers (called "vertical panning"), listeners feel that a sound image is played between two loudspeakers. According to appropriate sound pressure level differences, multiple listeners can feel that a sound image is played at the position of the 3D object regardless of listening position of listeners because sound playing equipments are only two loudspeakers.

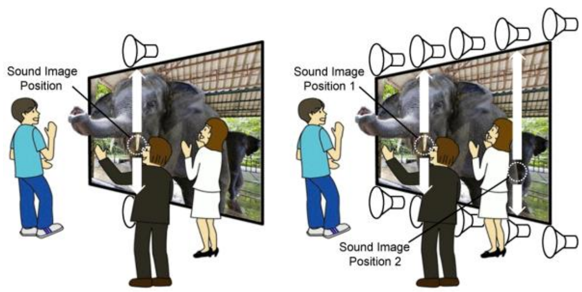

Figure 2 – Basic configuration of the proposed 3D audio system

Second, as shown in the right side of Figure 2, sound image positions are also expanded to the left-right direction by placing multiple loudspeaker pairs at the top and bottom of the screen. As a result, multiple listeners can simultaneously feel the sound images at the position of 3D objects depicted by the video display system regardless of listening position. In the proposed system, listeners do not need to wear headphones. Since loudspeakers are placed at the top and bottom of the screen, there are no devices between the projector array and the screen in the proposed system. In case of the application as 3D teleconference system, because this system only has to directly record the speech of talkers and does not restrict the position of recording microphones, it does not need to place recording microphones between the screen and the listening position in the proposed system.

3. Listening test

3.1 Experimental environment and condition

The listening test was performed in the ATR variable reverberation room[4]. A reverberation time can be changed from 140 ms (all absorption) to 1030 ms (all reflection) in this room as shown in Figure 3 by rotating the cylinders and ceiling louvers that are components of walls. A background noise level was A-weighted level of 14 dB (when the reverberation time was 140 ms) and 22 dB (when the reverberation time was 1030 ms).

Figure 3 – Image of experimental environment (Left: Reverberation time 140 ms, Right: Reverberation time 1030 ms)

Twenty-seven loudspeakers were placed in the vertical line as shown in Figure 4. Loudspeakers were manufactured by mounting a loudspeaker unit (Fostex: FE103En) on a loudspeaker enclosure (width: 11 cm, depth: 25 cm, height: 11

cm). The total height of the loudspeaker array was 2.97 m (=11 cm × 27). The listening position was set at the 5.5 m distance from the loudspeaker array according to the appropriate viewing distance in the developed 3D video display system[3]. The height of the listening position was set to 1.485 m at the ear position of listeners. The sound pressure level in the listening position was set to A-weighted level of about 70 dB.

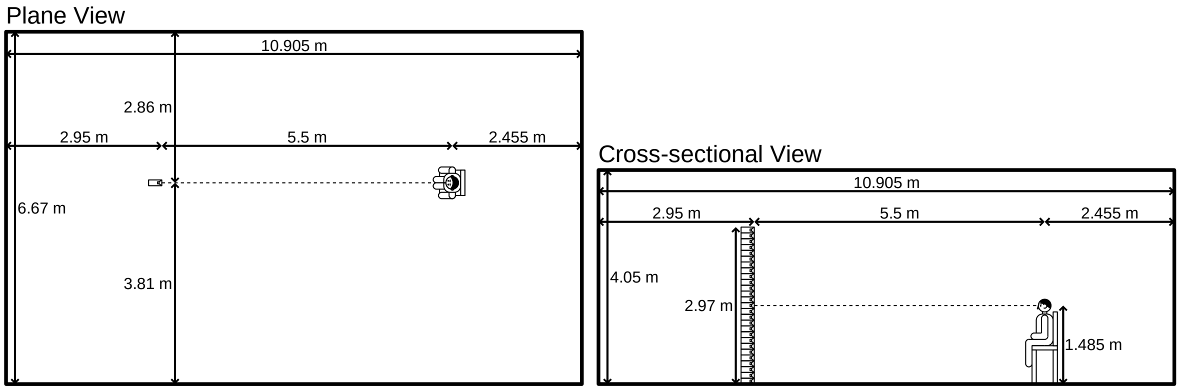

Figure 4 – Position of the listener and the loudspeaker array in the listening test

The experimental conditions in the listening test are shown in Figure 5. The gray loudspeakers denote the loudspeaker from which a sound is not replayed in each condition. In the panning condition (a), the sound calculated from the sound source signal $s(n)$ were replayed from two loudspeakers placed at the up and down side in the loudspeaker array according to following equations:

$$x_U(n) = a_U\, s(n),\; x_D(n) = a_D\, s(n), \qquad (1)$$

where $x_U(n)$ and $x_D(n)$ denote the sound signals replayed from two loudspeakers of the up and down side and a_U and a_D ($a_U^2+a_D^2=1$) denote the gain coefficients in each sound signal. If the level difference ΔA [dB] is defined as follows:

$$\Delta A = 20\log_{10}\{x_U(n) / x_D(n)\},\; = 20\log_{10}\{a_U / a_D\}, \qquad (2)$$

a_U and a_D are calculated as follows:

$$a_U = 10^{\Delta A/20} \Big/ \sqrt{10^{\Delta A/10}+1}, \quad a_D = 1 \Big/ \sqrt{10^{\Delta A/10}+1}\,. \qquad (3)$$

In this test, the level difference ΔA was set from –15 dB to 15dB at the interval of 1 dB. In the control condition (b), the sound source signal $s(n)$ was replayed from one loudspeaker selected from a group of thirteen loudspeakers.

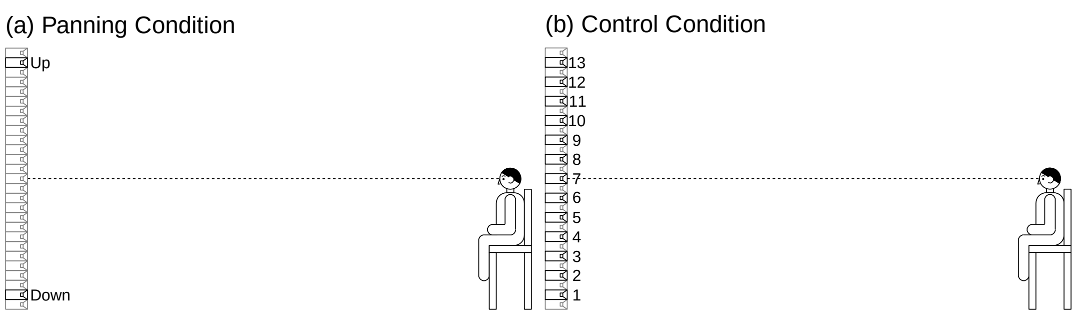

Figure 5 – Experimental conditions used in the listening test

3.2 Experimental design and procedure

Twelve subjects (age: 21-32, six males and six females), whom the audibility was normal in daily life, participated as listeners in this test. Three type sounds (white noise, speech and flute) were used as the sound source. The flowchart of the listening test is shown in the left side of Figure 6. The test was divided into six sessions for reverberation times and sound sources. Twelve practice trials and eighty-eight main trials were performed in each session. During the main trials, rest periods were allowed after every set of forty-four trials. The presentation orders of reverberation times, sound sources and trials were randomized for each listener. The details of the practice and main trials are shown in Table 1.

The listeners were instructed to report the perceived height of sound images by listing the index of heights in an answer sheet. The relation between the perceived height and the answer indexes is shown in the right side of Figure 6.

This index is from 1 to 27 and the height of the loudspeaker, of which the index is 14, is the same as that of the ear and eye of listeners. If listeners felt multiple sound images in the trials, the listeners could list multiple indexes in an answer sheet. The listeners were allowed to move their heads and upper bodies freely while listening to the sounds.

Table 1 – Practice and main trials in the listening test

	Element	Note
Practice (12)	= 7 conditions + 5 positions	ΔA = 0, ±5, ±10 and ±15 in (a) of Figure 5 1, 4, 7, 10 and 13 in (b) of Figure 5
Main (88)	= [31 conditions + 13 positions] ×2 repetitions	ΔA = -15 ~ 15 in (a) of Figure 5 1 ~ 13 in (b) of Figure 5

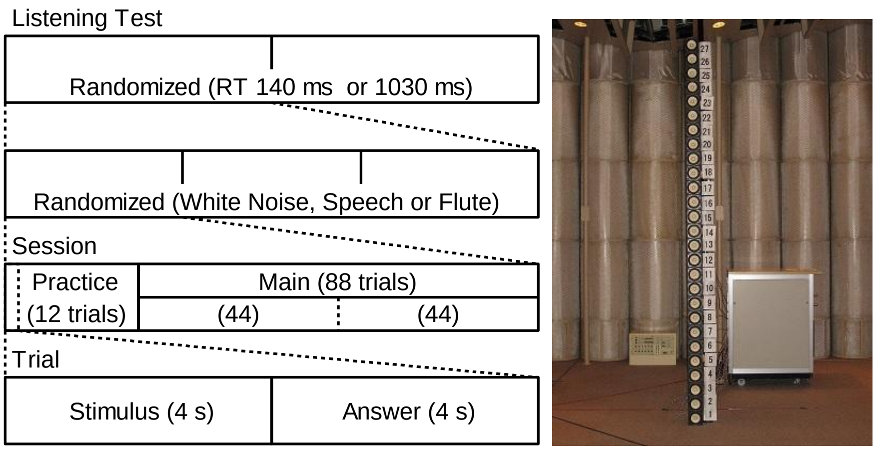

Figure 6 – Flowchart of the listening test (Left), Relation between perceived heights and answer indexes (Right)

3.3 Experimental result and discussion

In order to analyze experimental results, after reducing the answers where listeners listed multiple indexes, the perceived height of sound images was calculated from the answer indexes of listeners according to following equation:

$$H_{\mathrm{per}}\ [\mathrm{m}] = (I_{\mathrm{ans}} - 14) \times 0.11, \qquad (4)$$

where I_{ans} and H_{per} denote the answering index of loudspeakers and the perceived height of sound images. Results of the averages of the perceived height in the panning condition are shown in Figure 7. Error bars denote the 95% confidential interval of average heights. In all conditions, the perceived height of sound images is about 0 m (i.e., the middle point between two loudspeakers placed at the top and bottom of the screen) when the level difference is about 0 dB. Thus, it is shown that listeners feel the synthesized sound images between two loudspeakers by the proposed system. Particularly, it seems that the perceived height of sound images linearly changes relative to the level difference when the level difference is from –3 dB to 9 dB. As the result of linear regressions in the range from –3 dB to 9 dB, following regression lines were obtained in each condition:

$$\begin{aligned}
H_{\mathrm{per}} &= 0.1475\Delta A - 0.4066 \text{ (White noise, Reverberation time 140 ms)},\\
H_{\mathrm{per}} &= 0.1253\Delta A - 0.4499 \text{ (Speech, Reverberation time 140 ms)},\\
H_{\mathrm{per}} &= 0.0784\Delta A - 0.0045 \text{ (Flute, Reverberation time 140 ms)},\\
H_{\mathrm{per}} &= 0.1279\Delta A - 0.0510 \text{ (White noise, Reverberation time 1030 ms)},\\
H_{\mathrm{per}} &= 0.1079\Delta A - 0.1635 \text{ (Speech, Reverberation time 1030 ms)},\\
H_{\mathrm{per}} &= 0.0518\Delta A + 0.2130 \text{ (Flute, Reverberation time 1030 ms)}.
\end{aligned} \qquad (5)$$

Regression lines are shown in Figure 7 at bold lines. It is shown that these lines are correctly estimated in the range from –3 dB to 9 dB of the level difference.

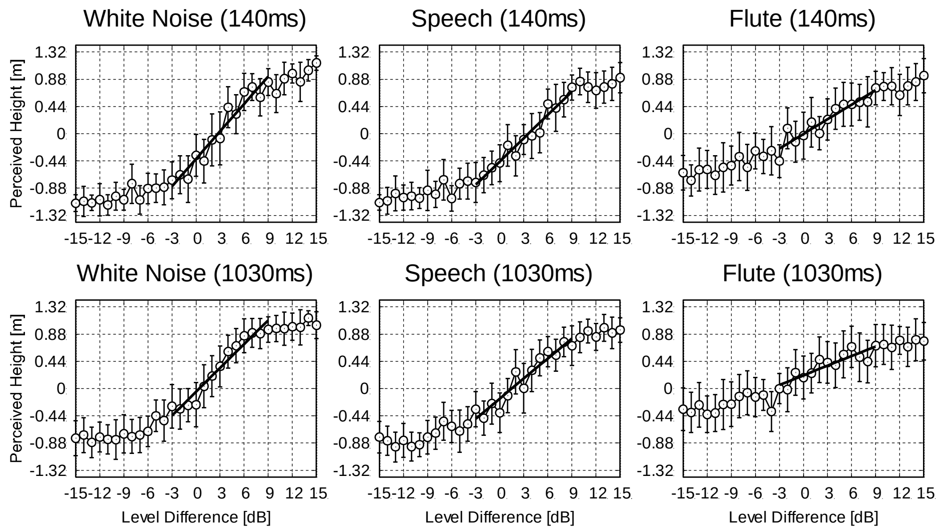

Figure 7 – Results and regression lines in the panning condition

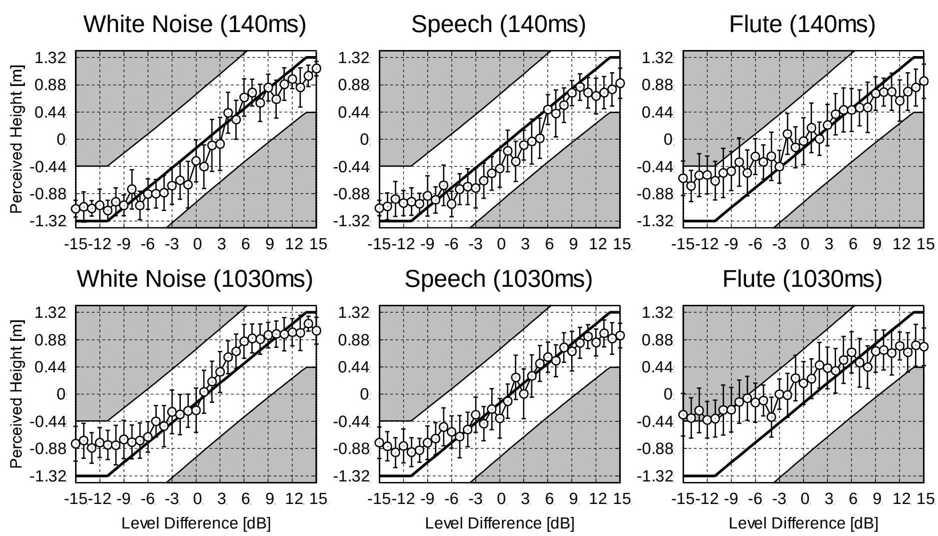

Figure 8 – Panning curves and differential limens in the panning condition

According to averaging the regression lines obtained in six conditions, the panning curve H_{pan} was calculated as follows:

$$H_{\mathrm{pan}} = \begin{cases} -1.32 & (\Delta A < -11.05) \\ 0.1065\Delta A - 0.1437 & (-11.05 \le \Delta A \le 13.74) \\ 1.32 & (\Delta A > 13.74) \end{cases} \tag{6}$$

The differential limens of the perceived height of sound images (DL^{+}_{pan} and DL^{-}_{pan}) were also calculated according to following equations:

$$DL^{+}_{\mathrm{pan}} = \tan\left(\tan^{-1}(H_{\mathrm{pan}} / 5.5) + \phi\right) \times 5.5,$$
$$DL^{-}_{\mathrm{pan}} = \tan\left(\tan^{-1}(H_{\mathrm{pan}} / 5.5) - \phi\right) \times 5.5, \tag{7}$$

where ϕ denotes the differential angle of the perceived height of sound images. The value of ϕ was set to 9 degrees with reference to past studies[5]. The panning curves and differential limens in the panning condition are shown in Figure 8. Gray areas denote the outside of differential limens. If there is the average of the perceived height of sound images in the gray areas, listeners can discriminate the difference between the presented height of sound images according to the panning curve and the perceived height of sound sources presented by videos. In five conditions except the condition (Flute, Reverberation time 1030 ms), since there is no average of the perceived height of sound images in the gray areas, the auditory performance of the panning curve is so high that listeners cannot discriminate the difference between the heights. However, in the condition (Flute, Reverberation time 1030 ms), since there is the average of the perceived height of sound images in the gray areas, listeners may be able to discriminate the difference between the heights.

On the other hand, the results and differential limens in the control condition are shown in Figure 9. Error bars denote the 95% confidential interval of average heights. The differential limens in the control condition (DL^{+}_{ctrl} and DL^{-}_{ctrl}) were calculated according to following equations:

$$\begin{aligned} DL^{+}_{\mathrm{ctrl}} &= \tan(\tan^{-1}(H_{\mathrm{pre}} / 5.5) + \phi) \times 5.5, \\ DL^{-}_{\mathrm{ctrl}} &= \tan(\tan^{-1}(H_{\mathrm{pre}} / 5.5) - \phi) \times 5.5, \end{aligned} \qquad (8)$$

where H_{pre} $(=(I_{\mathrm{pre}}-14)\times 0.11)$ denotes the height of presented sound sources and I_{pre} denotes the index of presented loudspeakers. In five conditions except the condition (Flute, Reverberation time 1030 ms), there is no average of the perceived height of sound images in the gray areas. However, in the condition (Flute, Reverberation time 1030 ms), there is the average of the perceived height of sound images in the gray areas. Thus, listeners may not perceive the height of sound sources itself due to the reverberation time when the sound sources are flute. It needs to evaluate the effect of the reverberation time on the height perception of sound sources by the additional listening test in which the sound source is a flute and the reverberation time varies step-by-step.

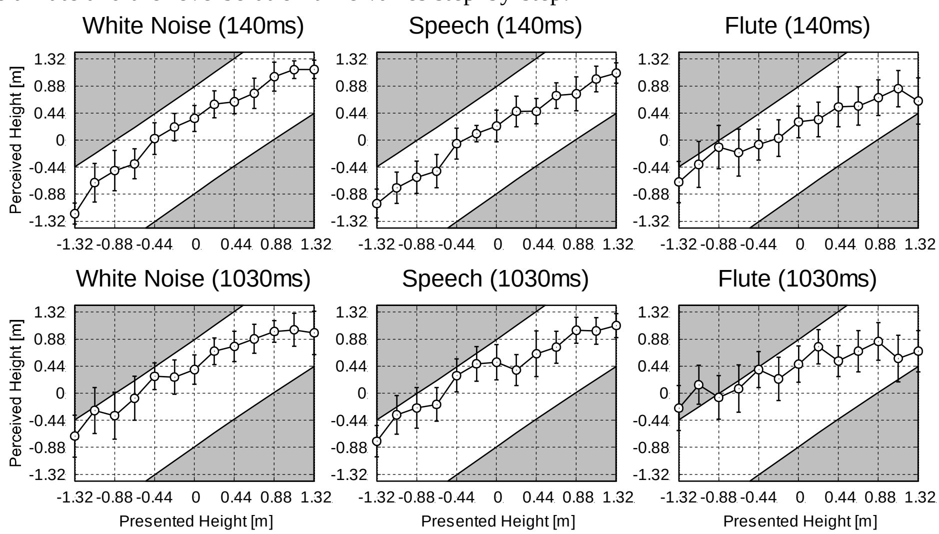

Figure 9 – Results and differential limens in the control condition

4. Conclusion

In this paper, in order to match to the large glasses-free 3D video display system, the novel 3D audio system based on Multiple Vertical Panning (MVP) method was proposed. In order to evaluate the auditory performance of the proposed system, the listening test was designed by using the loudspeaker array in which twenty-seven loudspeakers were aligned on the vertical line. As a result, it was indicated that the auditory performance of the proposed system is so high that listeners cannot discriminate the difference between the perceived heights of sound images in five conditions except the condition (Flute, Reverberation time 1030 ms).

As a future work, it needs to evaluate the effect of the reverberation time on the height perception of sound sources by the additional listening test in which the sound source is a flute and the reverberation time varies. It also needs to develop the audio-visual system in which the proposed system and the large glasses-free 3D video display system are integrated and evaluate the effectiveness of the proposed system in an audio-visual system by performing a test.

References and links

[1] K. Enami, “Research on ultra-realistic communications,” ECTI Trans. Electr. Eng. Electron. Commun., vol. 6, no. 1, pp. 22-25, Feb. 2008.

[2] M. Kawakita, S. Iwasawa, G. Sabri, and N. Inoue, “Development of glasses-free 3D video system,” Proc. IUCS, pp. 323-327, Oct. 2011.

[3] H. Ando, M. Kawakita and S. Iwasawa, “The World's Largest 200-inch Glasses-free 3D Display Has Been Successfully Developed,” NICT Press Release, http://www2.nict.go.jp/pub/whatsnew/press/h22/110125/110125_e.html, Jan. 2011.

[4] T. Hirahara, C. Müller and Y. Tohkura, “Structure and Acoustic Characteristics of the ATR Variable Reverberation Room,” J. Acoust. Soc. Jpn., vol. 48, no. 5, pp. 301-308, May. 1992 (in Japanese).

[5] J. Blauert, "Spatial hearing," pp. 37-50, MIT Press, Cambridge, Mass, revised edition, 1997