

# Multiple Vertical Panningを用いた立体音響システムにおいて音源位置の離散化が臨場感に及ぼす影響

木村 敏幸<sup>\*1</sup>      安藤 広志<sup>\*2</sup>

Effect of Discretization of Sound Source Position on Sense of Presence in 3D Audio Systems  
Using Multiple Vertical Panning

Toshiyuki Kimura<sup>\*1</sup> and Hiroshi Ando<sup>\*2</sup>

**Abstract** – NICT has developed a multi-view 3D video display system with a 200-inch screen (REI display). In this system, several viewers can simultaneously view natural 3D objects from their respective viewing positions without special glasses. We proposed a 3D audio system using multiple vertical panning (MVP), which is compatible with REI display systems. However, the practical realization of our proposed system is obstructed by the assumption that the position of its sound sources is continuous. In this paper, to develop a practical system, we evaluated the effect of the discretization of sound source positions on the sense of presence by performing two audio-visual psychological experiments where the horizontal and vertical positions of sound sources were discretized. 3D videos were presented to viewers wearing glasses in both experiments. The results of the first experiment showed that viewers failed to discriminate differences of the sense of presence with ten or more loudspeakers. In the second experiment, viewers failed to discriminate differences of the sense of presence even if the vertical position of the sound sources was discretized with two steps. We conclude that a practical system using our proposed system can be constructed with five horizontal and two vertical steps of sound source positions.

**Keywords** : multi-view 3D video display, multiple vertical panning, Scheffé’s paired comparison, sound location, sound movement

## 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 Fig. 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 characteris-

tics 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.

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]. The basic configuration of our proposed system is shown in Fig. 2. First, as shown on its left-hand side, two loudspeakers are placed at the upper and lower sides of

<sup>\*1</sup>東北学院大学工学部

<sup>\*2</sup>国立研究開発法人情報通信研究機構ユニバーサルコミュニケーション研究所

<sup>\*1</sup>Faculty of Engineering, Tohoku Gakuin University

<sup>\*2</sup>Universal Communication Research Institute, National Institute of Information and Communications Technology

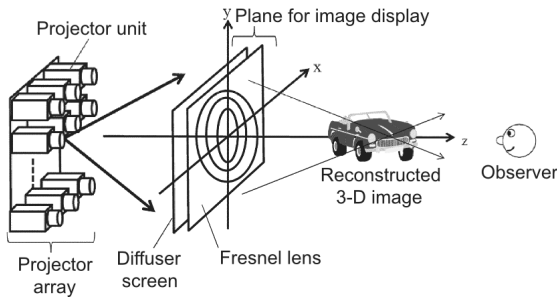


図1 大画面裸眼立体映像ディスプレイ (REI) の基本構成 [3]

Fig.1 Basic configuration of a large-screen multi-view 3D video display system (REI display) [3].

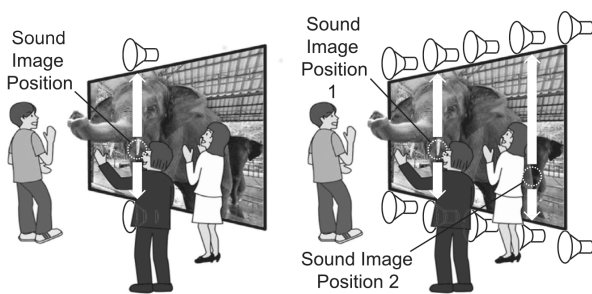


図2 Multiple Vertical Panning (MVP) 方式による立体音響システムの基本構成 [4]

Fig.2 Basic configuration of a 3D audio system using multiple vertical panning (MVP) method [4].

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

In our proposed system, viewers do not need to wear listening 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 participants in teleconferences and does not restrict the positions of microphones, they do not need to be placed between the projector array and the viewing position.

However, our proposed system assumes that the position of the sound sources is continuous in it. This assumption causes various problems when a practical system is constructed. For example, if the horizontal position of the sound sources is continuous, constructing the system is very expensive with an enormous number of loudspeakers because it needs to densely place vertically panned loudspeakers in the screen's horizontal direction. On the other hand, it needs to transmit the position information of the sound sources when our proposed system is applied to broadcasting and telecommunications systems. If the position of the sound sources is continuous, the amount of the transmitted bit-rate enormously increases because it needs to transmit the continuous position information based on the movement of the sound sources. To solve these problems, we need to reduce the number of loudspeakers and the transmitted bit-rate of the position of the sound sources by discretizing their horizontal and vertical positions.

The position of the sound sources shifts from the position of the 3D objects to the horizontal and vertical directions by discretizing the horizontal and vertical positions of the sound sources. If the viewers cannot perceive the shift of the positions of the sound sources, the performance of our proposed system is expected to be maintained even if the position of the sound sources shifts to the horizontal and vertical directions. The phenomenon where viewers cannot perceive the shift of the position of sound sources is known as the ventriloquism effect. However, in a past study of it [5], the video of sound sources (speaker) is static and the position of sound sources only shifts to the horizontal direction. On the other hand, we assume that the video of the sound sources is moving and their position shifts to the vertical direction when our proposed system is applied to broadcasting and telecommunications systems. Thus, it is difficult to predict the effect of the discretization of the horizontal and vertical positions of sound sources based on a past study of the ventriloquism effect.

In this paper, to construct a practical system using our proposed system, we performed two audio-visual psychological experiments where the horizontal and vertical positions of sound sources were discretized. In the first experiment, to reduce the number of loudspeakers, the horizontal position of the sound sources was discretized, and the effect of the number of loudspeakers on the sense of presence was evaluated. In the second experiment, to reduce the bit-rate transmitting the vertical position of the sound sources, we evaluated the effect of the discretization of the vertical position of sound sources on the sense of presence.

## 2 Audio-visual experiment 1

In this section, to evaluate the effect of the discretization of the horizontal position of sound sources, we performed an audio-visual psychological experiment where the number of loudspeakers is reduced and evaluated the effect of the number of loudspeakers on the sense of presence.

### 2.1 Environment and conditions

Although a REI display should be applied in this experiment, the REI display has 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 42 loudspeakers in the room (Fig. 3) 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

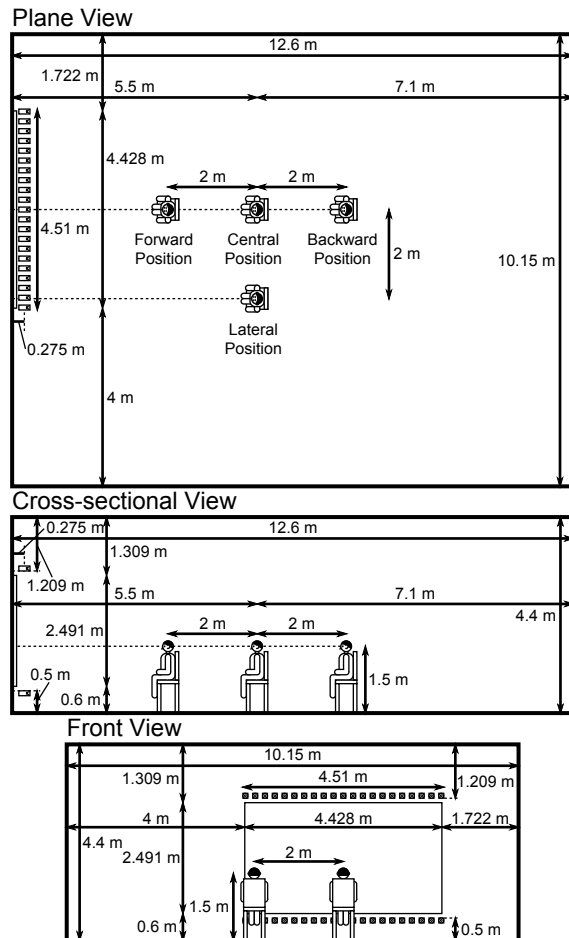


図3 視聴覚実験1における視聴者、スクリーン及びスピーカアレイの配置

Fig. 3 Position of viewers, screen, and loud-speaker array in first audio-visual experiment.

multi-view 3D video display system ( $\pm 2$  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 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 Fig. 4. In it, the UFO (inside the white 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 black circles),

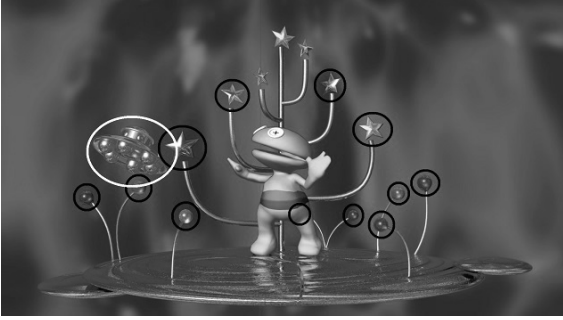


図4 視聴覚実験1にて用いた立体映像  
Fig.4 3D video in first audio-visual experiment.

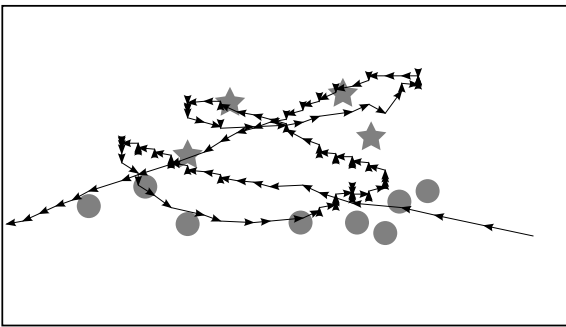


図5 立体映像におけるUFO, 星, ボールの位置及び動き  
Fig.5 Position and movement of UFO, stars, and balls in 3D video.

the sound of the stars and balls is played at their positions. The time waveform and spectrogram of the sounds of the UFO, the stars, and the balls are shown in Figs. 6-8. The position and movement of the UFO, the stars, and the balls in the 3D video are shown in Fig. 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 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 Fig. 9. The gray loudspeakers denote the loudspeaker from which a sound was not replayed in each condition. The black arrows on the screen denote the vertical panning in the replayed vertically panned loudspeakers. The sounds played at 3D object position ( $P_H, P_V$ ) at

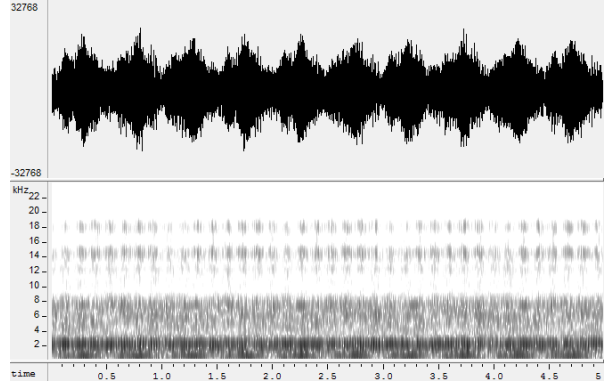


図6 視聴覚実験1にて用いたUFO音の時間波形及びスペクトログラム  
Fig.6 Time waveform and spectrogram of UFO sound in first audio-visual experiment.

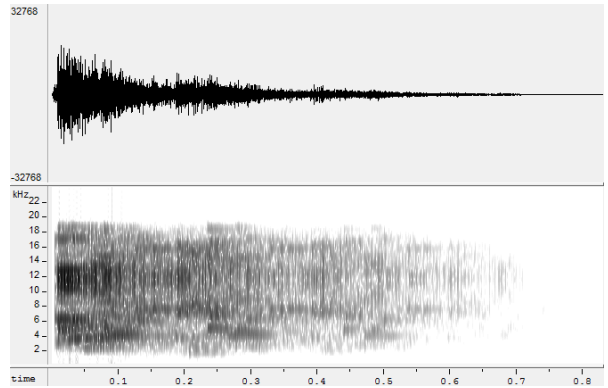


図7 視聴覚実験1にて用いた星音の時間波形及びスペクトログラム  
Fig.7 Time waveform and spectrogram of star sound in first audio-visual experiment.

time  $T = \frac{m-1}{F_v}$  were synthesized by the following procedure. Note that  $F_v (=30 \text{ fps})$  and  $m (=1, \dots)$  denote the frame rate and the frame index of the video signals.  $P_H (= -2.2 \sim 2.2)$  and  $P_V (= -1.25 \sim 1.25)$  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:

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

where  $P'_H (= -2.2, \dots, 2.2)$  denotes the horizontal position of the two selected loudspeakers.  $\Delta d$  denotes their right-and-left intervals. In this experiment, the

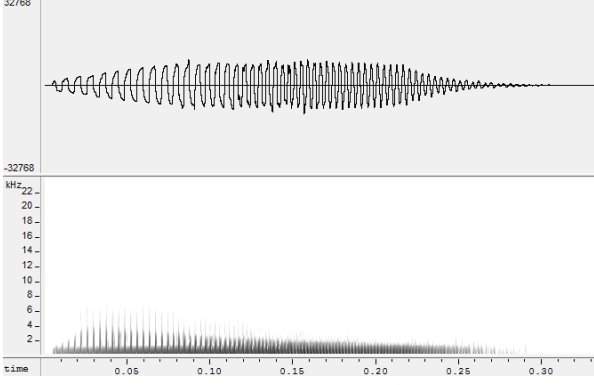


図8 視聴覚実験1にて用いたボール音の時間波形及びスペクトログラム  
Fig.8 Time waveform and spectrogram of ball sound in first audio-visual experiment.

$\Delta d$  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 Fig. 9.

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

$$x_U(n) = a_U w(n) s(n), \quad (2)$$

$$x_D(n) = a_D w(n) s(n), \quad (3)$$

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

where  $F_s (=48 \text{ kHz})$  and  $n (=0, \dots)$  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. The sound source signal is the materials of the sound effects published on a website [8].  $w(n)$  denotes the window function of the sound signals, defined as follows:

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

where  $L (=1 \text{ ms})$  denotes the cross-fade time of the window function.  $a_U$  and  $a_D$  (the gain coefficients in each sound signal) are calculated from level differ-

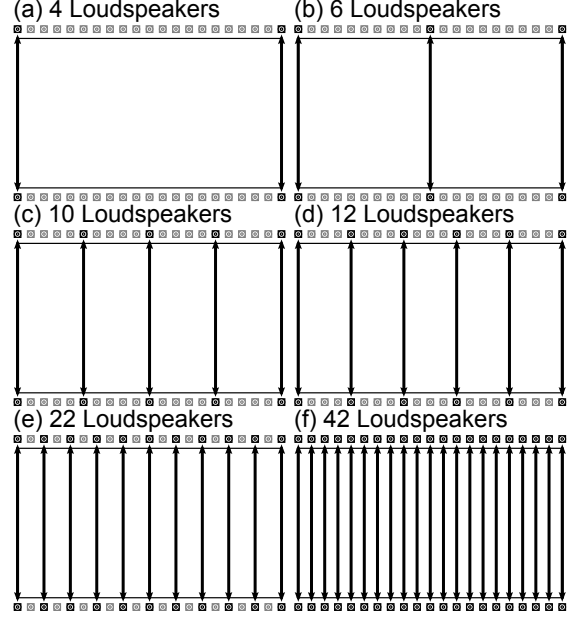


図9 視聴覚実験1にて用いた音条件  
Fig.9 Sound conditions in first audio-visual experiment.

ence  $\Delta A$  [dB] as follows:

$$a_U = \frac{10^{\frac{\Delta A}{20}}}{\sqrt{10^{\frac{\Delta A}{10}} + 1}}, \quad (5)$$

$$a_D = \frac{1}{\sqrt{10^{\frac{\Delta A}{10}} + 1}}. \quad (6)$$

In this experiment, level difference  $\Delta A$  was based on a previous study [4]:

$$\Delta A = \frac{\alpha P_V + 0.1437}{0.1065}. \quad (7)$$

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

## 2.2 Design and procedure

Nine subjects (ages: 29–39, five males and four females) with normal stereoscopic acuity and normal audibility participated as viewers in this experiment. Scheffé's paired comparison [9] was applied as an evaluation method. This experiment's flowchart is shown in Fig. 10. 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. 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 (Fig. 4)

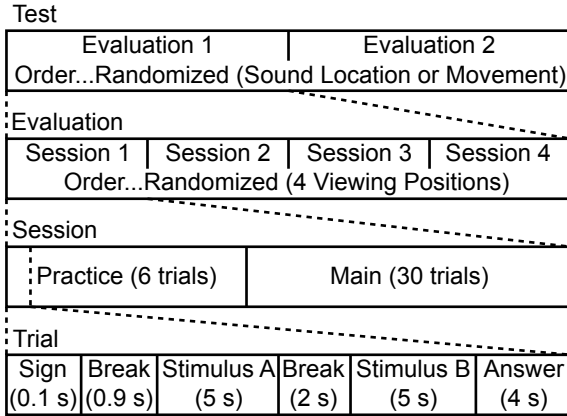


図 10 視聴覚実験 1 の流れ図

Fig. 10 Flowchart of first audio-visual experiment.

表 1 Scheffé の一対比較法の評定

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

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. 9(a), (b), and (f). The permutations of the six sound conditions shown in Fig. 9 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.

### 2.3 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

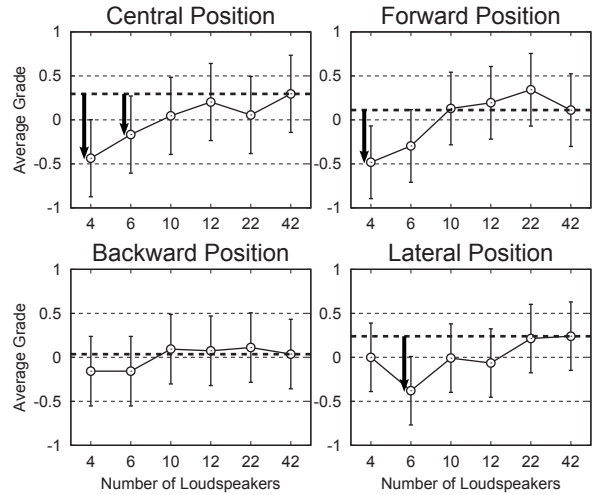


図 11 視聴覚実験 1 の結果 (音の位置)

Fig. 11 Results of first audio-visual experiment: sound location.

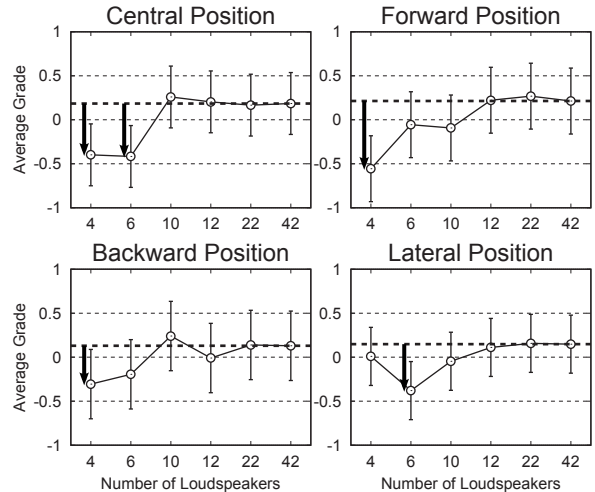


図 12 視聴覚実験 1 の結果 (音の動き)

Fig. 12 Results of first audio-visual experiment: sound movement.

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 Figs. 11 & 12. The error bars denote 95% confidence intervals based on a yardstick.

Because only the number of loudspeakers in our proposed system varies in the sound conditions of this experiment, the sound condition (f) in which the number of loudspeakers is largest is the condition in which the sense of presence is accurately expressed. On the other hand, to maintain the per-

formance of our proposed system (i.e., viewers can perceive a sound image at the position of the 3D object regardless of their viewing position), it is necessary for viewers to feel the same sense of presence as the sound condition (f) in eight sessions. Thus, we evaluated the effect of the number of loudspeakers on the sense of presence on the basis of the sound condition (f). The bold dashed lines in Figs. 11 & 12 denote the border line based on the sound condition (f). The arrows in Figs. 11 & 12 denote the significant differences among the sound condition (f) at the 5% significant level. When the number of loudspeakers is 10, 12 and 22, we found no significant differences among the sound condition (f) in all session. Thus, 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. However, the average grades were significantly lower (5% level) than the sound condition (f) in the central and backward viewing positions with six loudspeakers. On the other hand, the average grades were significantly lower (5% level) than the sound condition (f) in the central and forward viewing positions and the session (the backward viewing position and the sound movement) with four loudspeakers. Based on the result described above, it is indicated that the performance of our proposed system is not maintained with four or six loudspeakers and that equal to or more than ten loudspeakers are needed to maintain the performance.

### 3 Audio-visual experiment 2

In this section, to evaluate the effect of the discretization of the vertical position of the sound source, we performed an audio-visual psychological experiment where the number of vertical steps is reduced and evaluated the effect of the discretization of the vertical position of the sound source on the sense of presence.

#### 3.1 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 Fig. 13. The gray loudspeakers denote the loudspeaker from which a sound is not replayed in each condition. The black arrows on the screen de-

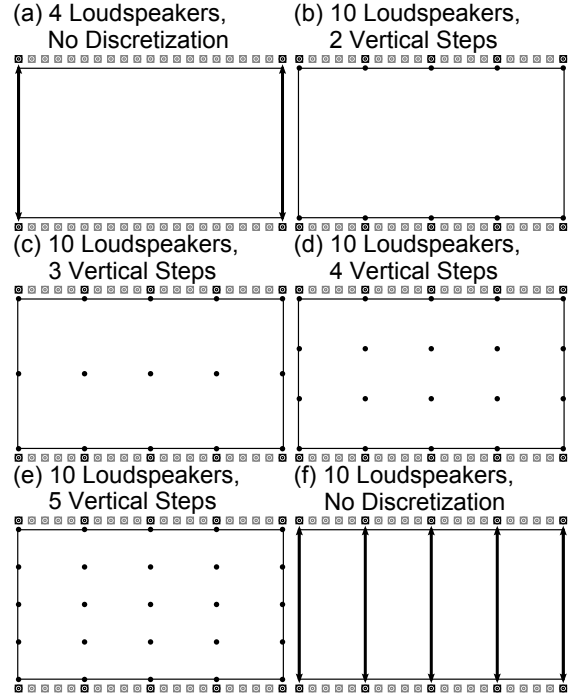


図 13 視聴覚実験 2 にて用いた音条件  
 Fig.13 Sound conditions in second audio-visual experiment.

note the vertical panning in the replayed vertically panned loudspeakers. The black circles denote the sound source position which is discretized to the vertical direction. Sound condition (f) is the condition where viewers could not discriminate the differences of the sense of presence even if the number of loudspeakers was increased 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. Both 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  $\Delta d$  (right-and-left intervals of the two selected loudspeakers) values are 4.4 m in sound condition (a) and 1.1 m in the other sound conditions shown in Fig. 13. In addition, Eq. (7) is modified as follows:

$$\Delta A = \frac{\alpha P'_V + 0.1437}{0.1065}, \quad (8)$$

where  $\alpha (= \frac{2.7}{2.5})$  is a compensating coefficient which is set identically as the first audio-visual experiment described in Section 2.  $P'_V (= -1.25, \dots, 1.25)$  (the discretized vertical position of 3D objects) is defined as



follows:

$$P'_V = \begin{cases} \Delta d_V \text{ round} \left( \frac{P_V + 1.25}{\Delta d_V} \right) - 1.25 & (\Delta d_V \neq 0) \\ P_V & (\Delta d_V = 0) \end{cases}, \quad (9)$$

where  $\Delta d_V$  denotes the interval of the discretized vertical position. In this experiment, the  $\Delta d_V$  values are 0, 2.5, 1.25, 0.833, 0.625, and 0 m in an order corresponding to the sound conditions shown in Fig. 13.

### 3.2 Design and procedure

Nine subjects (ages: 22–38, five males and four females) with normal stereoscopic acuity and normal audibility participated as viewers in this experiment. Two males and one female 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.

### 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)  $\times$  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 Figs. 14 & 15. 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.

The sound condition (f) is the condition in which the sense of presence is accurately expressed. On the other hand, to maintain the performance of our proposed system (i.e., viewers can perceive a sound image at the position of the 3D object regardless of their viewing position), it is necessary for viewers to

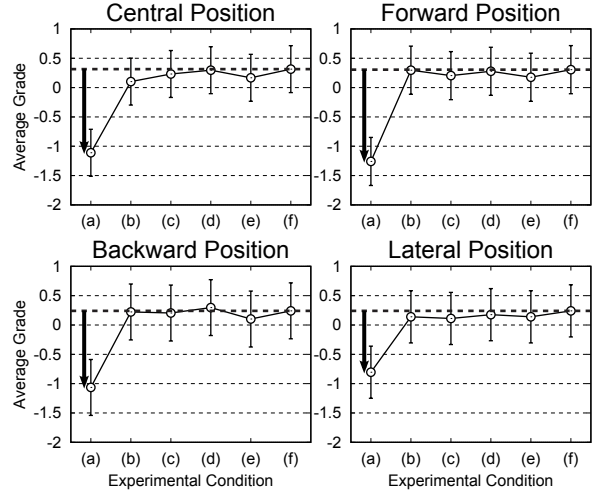


図 14 視聴覚実験 2 の結果 (音の位置)  
Fig. 14 Results of second audio-visual experiment: sound location.

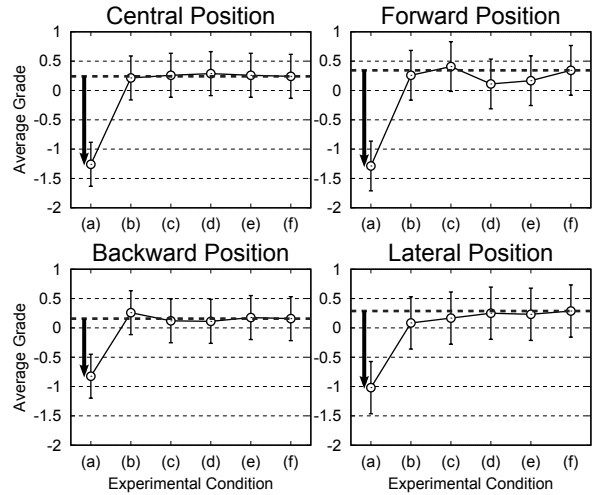


図 15 視聴覚実験 2 の結果 (音の動き)  
Fig. 15 Results of second audio-visual experiment: sound movement.

feel the same sense of presence as the sound condition (f) in eight sessions. Thus, we evaluated the effect of the discretization of the vertical position on the sense of presence on the basis of the sound condition (f). The bold dashed lines in Figs. 14 & 15 denote the border line based on the sound condition (f). The arrows in Figs. 14 & 15 denote the significant differences among the sound condition (f) at the 5% significant level. There are no significant differences among the sound condition (f) in all the sessions when the sound conditions range from (b) to (e). In other words, even if the vertical positions of the sound sources are discretized to two steps, viewers cannot discriminate the differences of the location



and the movement of the sound images. Thus, we believe that the vertical position of the sound sources can be discretized to two steps when an audio-visual system is based on our proposed system.

We consider the reduction of the transmitted bit-rate by discretizing the vertical position of sound sources. Before the reduction, the transmitted bit-rate is 32 bits at least (i.e., float type in C language) because it needs to transmit the vertical position by a real value. On the other hand, because the vertical position of the sound sources can be discretized to two steps, the transmitted bit-rate is 1 bit (i.e., boolean type in C language). Thus, the transmitted bit-rate is reduced to one thirty second.

#### 4 Conclusion

In this paper, we constructed a practical system using our proposed 3D audio system based on multiple vertical panning (MVP) by performing two audio-visual psychological experiments where the horizontal and vertical positions of sound sources were discretized.

In the first experiment, to evaluate the effect of the discretization of the horizontal position of sound sources, we performed an audio-visual psychological experiment where the number of loudspeakers was reduced and the effect of the number of loudspeakers on the sense of presence was evaluated. The results of the first experiment showed that viewers failed to discriminate differences of the sense of presence with ten or more loudspeakers.

In the second experiment, to evaluate the effect of the discretization of the vertical position of sound sources, we performed an audio-visual psychological experiment where the number of vertical steps was reduced and the effect of the discretization of the vertical position of sound sources on the sense of presence was evaluated. The viewers failed to discriminate the differences of the sense of presence even if the vertical position of the sound sources was discretized with two steps.

Thus, when an audio-visual system is constructed using our proposed system, a practical system can be constructed if the horizontal position of the sound sources is discretized with five steps and their vertical position is discretized with two steps. In these experiments, the used 3D video was one type because it takes a lot of costs to render the 3D video for REI

display. Thus, it is necessary to evaluate the effect on the sense of presence by using other 3D videos in order to verify the generality of the experimental results.

Future work needs to study the expression method of the depth direction by our proposed system. On the other hand, our proposed system assumes that sound is directly recorded by a microphone placed at the neighborhood of the sound sources. However, it is difficult to place the microphone in the neighborhood of the sound sources if the number of sound sources is numerous. Thus, to solve this problem, future work needs to study a recording method using a distantly placed microphone array.

#### Acknowledgment

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

#### 参考文献

- [1] K. Enami: Research on ultra-realistic communications; ECTI Trans. Elect. Eng., Electro., Comm., 6(1), 22-25 (2008.02)
- [2] 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 (2013.06)
- [3] M. Kawakita, S. Iwasawa, G. Sabri and N. Inoue: Development of Glasses-free 3D Video System; Int. Univ. Comm. Symp., pp. 323-327 (2011.10)
- [4] T. Kimura and H. Ando: 3D Audio System Using Multiple Vertical Panning for Large-screen Multi-view 3D Video Display; ITE Trans. on Media Tech. and App., 2(1), 33-45 (2014.01)
- [5] S. Komiyama: Subjective Evaluation of Angular Displacement between Picture and Sound Directions for HDTV Sound Systems; J. Audio Eng. Soc., 37(4), 210-214 (1989.04)
- [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)

(2015年3月2日受付)

[著者紹介]

木村 敏幸 (正会員)



1998年名大・工・物化卒。2000年同大学院・人間情報・修士課程了。2005年同大学院・人間情報・博士課程了。日本学術振興会特別研究員(PD), 名大研究員, 東京農工大特任助手, 情報通信研究機構有期研究員を経て, 2015年より東北学院大・准教授, 現在に至る。ヒューマンインタフェース, 空間知覚, アレイ信号処理に関する研究に従事。博士(学術)。電子情報通信学会, 日本音響学会, ヒューマンインタフェース学会, Audio Engineering Society 各会員。

安藤 広志



1983年京都大学理学部(物理学)卒業。1987年京都大学文学研究科(実験心理学)修士課程修了。1992年MIT脳・認知科学科博士課程修了, Ph.D.(計算神経科学)。同年, (株)国際電気通信基礎技術研究所(ATR)に入社, 人間情報通信研究所研究員, 主任研究員を経て, 2006年より2011年まで認知ダイナミクス研究室室長, 2006年より(独)情報通信研究機構(NICT)ユニバーサルメディア研究センターの超臨場感研究プロジェクトに参画, グループリーダーを経て, 現在, NICTユニバーサルコミュニケーション研究所多感覚・評価研究室室長, 脳情報通信融合研究室副室長, 大阪大学生命機能研究科招聘教授。認知脳科学, 計算神経科学, 多感覚情報処理, 多感覚インタフェースの研究に従事。