Subjective Evaluation of Reproduction Method for Frontal Channels of 22.2 Multichannel Sound over a Direct-View Display

Takehiro Sugimoto (member)[†], Satoshi Oode (member)[†], Yasushige Nakayama (member)[†],
Hiroyuki Okubo (member)^{††}

Abstract

A reproduction method for the 11 frontal channels of a 22.2 multichannel sound system (22.2 ch) over a direct-view display is investigated. NHK is planning to broadcast 8K Super Hi-Vision with 22.2 ch in a pilot broadcasting in 2016. A loudspeaker array frame has been developed for integration into the direct-view display as a home reproduction system and is used as a reproduction device of 22.2 ch. Localization of the frontal channels reproduced by the real loudspeakers is subjectively evaluated in terms of localization accuracy by estimating the difference from the intended localization for an 85-inch display. Three reproduction methods to synthesize the channels on the display are proposed and evaluated by comparison with a real loudspeaker placed at an exact channel position. Appropriate reproduction methods for the 11 frontal channels of 22.2 ch were decided from the results of a subjective evaluation.

Key words: Localization, Sound image, 22.2 multichannel sound system, 8K Super Hi-Vision, Direct-view display.

1. Introduction

NHK is currently developing 8K Super Hi-Vision (SHV) as a next-generation television system that will convey a far stronger sense of reality [1]. 8K SHV audio is presented with a 22.2 multichannel sound system (22.2 ch) composed of 24 channels three-dimensionally distributed in three layers [2–4]. 8K SHV satellite broadcasting is scheduled to be launched in 2020. For the popularization of 8K SHV broadcasting, the quality of the reproduction device, i.e., the display and loudspeaker system, is a key factor. A world's first direct-view display for 8K SHV was developed with liquid crystal display (LCD) technology in 85 inch [5]. The size was selected so that the display can be readily brought into the home. Concerning 22.2 ch, it fundamentally uses separately placed discrete loudspeakers. However, it is sometimes difficult to install multichannel audio at home. As a matter of fact, even discrete 5.1 ch has hardly been accepted in the current market. Hence, we have been developing a home reproduction method using a loudspeaker array frame (LAF) integrated into

a direct-view display [6–9].

The reproduction methods of 22.2 ch with an LAF can be roughly classified into two groups: those for the 11 frontal channels around and on the display, and those for the 11 side and rear channels. The side and rear channels are auditorily imitated by binaural reproduction over loudspeakers [10]. A major issue of the reproduction methods for the 11 frontal channels is the three channels placed on the display, because a conventional loudspeaker cannot be equipped on the surface of a direct-view display.

Several studies on the presentation of a sound image on a display have been previously reported. Ozawa and Furuya investigated the localization accuracy of a synthesized sound image using four loudspeakers placed on each corner of a display [11]. Their result showed that the central part of the display, which is far from each loudspeaker, cannot easily synthesize a stable sound image. Furuya *et al.* also studied the synthesis of a sound image using four loudspeakers placed at the vertices of a square [12]. The sound image localization was reported to be unstable in the case that the intended position was distant from each loudspeaker, particularly at the central part of the loudspeaker arrangement. Kimura and Ando studied the localization of a sound image on a large screen by using a vertical panning method be-

Received August 8, 2014; Revised September 24, 2014; Accepted October 21, 2014

[†] Science & Technology Research Laboratories, NHK
(1-10-11, Kinuta, Setagaya-ku, Tokyo, 1578510 Japan)

^{††} NHK Engineering System, Inc.
(1-10-11, Kinuta, Setagaya-ku, Tokyo, 1578540 Japan)

tween two loudspeaker arrays set on the top and bottom edges of the large display [13]. Although their result showed that a sound image synthesized by vertical panning did not localize at an intended position without a video, the studied display size (200 inch) and geometrical configuration (5.2 m distance) were too large for use at home.

The purpose of this paper is to establish an appropriate reproduction method for the 11 frontal channels of 22.2 ch over the display. Localization of the frontal channels reproduced by the real loudspeakers was subjectively evaluated in terms of localization accuracy by estimating the difference from the intended localization for an 85-inch display. Three reproduction methods to synthesize the channels on the display were proposed and evaluated by comparison with a real loudspeaker placed at an exact channel position.

The rest of this paper is organized as follows. Section 2 outlines 22.2 ch. The loudspeaker arrangement of the channels around the display is provided in Section 3. The three reproduction methods for the three channels on the display are proposed in Section 4. Methodology of the subjective evaluation is described in detail in Section 5. Sections 6 and 7 present our experimental study, and Section 8 provides a discussion of the results. Finally, a brief summary of our main findings is given in Section 9.

2. 22.2 multichannel sound system

The essential concept of 22.2 ch is to provide a highly immersive sense of presence and reality with a compatibility to existing multichannel audio [2].

Figure 1 shows the channel mapping of 22.2 ch [3] and the center of the listening area by the figure of

Fig. 1 22.2 multichannel sound system.

a listener. The system consists of 24 discrete channels three-dimensionally distributed in three layers. The top layer has nine channels that are placed above the display. The middle layer has ten channels at the height of the listener's ears. The bottom layer has three channels, including two low frequency effects (LFE) channels, usually placed below the display. As shown in Fig. 1, 22.2 ch has 11 frontal channels around and on the display. The other 11 channels are placed around the listener to provide the sensation of sound above the listener and to the sides and rear.

3. Loudspeaker arrangement of the channels around the display

In 22.2 ch, eight channels are placed around the display as shown in Fig. 1. They are labeled as TpFL, TpFC, TpFR, FL, FR, BtFL, BtFC, and BtFR. Here we installed loudspeakers corresponding to these channels in contact with the edge of the display, as shown in Fig. 2, aiming at market acceptance by integrating the loudspeakers with the display. TpFL and TpFR are respectively placed at the top left and top right of the display, and TpFC is placed at the center of the top edge. FL and FR are placed at half the height on both vertical sides. BtFL, BtFC, and BtFR are placed on the bottom edge similarly to the loudspeakers on the top edge.

The reasons why we employed a loudspeaker corresponding to each channel are as follows:

- 22.2 ch specifies the stable localization of frontal sound over the entire image [2]. Thus, unstable sound image obtained by the conventional amplitude panning adopted in two-channel stereophonic system [14] is not preferable. We concluded that a four-loudspeaker system on every corner [11,12] would not be suitable for 22.2 ch.

- In the current market, loudspeakers for televisions

Fig. 2 Loudspeaker arrangement of the channels around the display.

are required to be as small as possible. However, a small loudspeaker is apt to suffer from a lack of sound pressure. To ensure sufficient sound pressure, a loudspeaker corresponding to each channel is desirable from the viewpoint of obtaining a sufficient peak margin.

4. Reproduction methods for the channels on the display

We studied three reproduction methods for the three channels on the display, i.e., FLc, FC, and FRc. Figure 3 illustrates the three proposed methods.

- (i) Amplitude panning by a horizontal pair of loudspeakers (horizontal pair).
- (ii) Amplitude panning by a vertical pair of loudspeakers (vertical pair).
- (iii) Quasi-wavefront synthesis (QWS) by the LAF.

Details of each method are given in the following sections.

4.1 Horizontal pair

The horizontal pair is operated by conventional left-right amplitude panning based on the tangent law adopted in two-channel stereophonic system [15]. The tangent law is formulated as follows:

$$\frac{\tan \phi}{\tan \phi_0} = \frac{g_l - g_r}{g_l + g_r}, \quad (1)$$

where ϕ_0 is the azimuthal angle of each loudspeaker from the listener's front, ϕ is the offset azimuthal angle of the sound image from the listener's front, g_l and g_r are the gain factors distributing the signal levels to left and right loudspeaker. The horizontal pair uses the

same loudspeaker units as FL and FR in Fig. 2.

4.2 Vertical pair

Three pairs of loudspeaker units for FLc, FC, and FRc are placed on the top and bottom edges as indicated in Fig. 3. The vertical pair is also operated by the amplitude panning based on the tangent law formulated as follows:

$$\frac{\tan \varphi}{\tan \varphi_0} = \frac{g_t - g_b}{g_t + g_b}, \quad (2)$$

where φ_0 is the elevation angle of each loudspeaker from the listener's front, φ is the offset elevation angle of the sound image from the listener's front, g_t and g_b are the gain factors distributing the signal levels to top and bottom loudspeaker. Because FLc, FC, and FRc are positioned at the middle of the vertical pair, $g_t = g_b$ is adopted in every case.

4.3 QWS by LAF

We have proposed QWS by the LAF as a reproduction method for the frontal channels of 22.2 ch [6, 7]. The concept is to modify the Rayleigh integral, which is an extension of the Kirchhoff-Helmholtz integral over a boundary plane surface [16, 17], in order to apply to the LAF.

Figure 4 is a schematic view of QWS by the LAF. r_0 is the listening position, r_s is the position of an imaginary sound source, and r_l denotes the position of each loudspeaker unit. The operation principle of QWS by the LAF is based on the Rayleigh I integral, which drives a monopole sound source using a sound pressure gradient at the position of each monopole sound source [16]. By modifying the Rayleigh I integral, we define the operation principle of QWS by the LAF as follows using a line integral:

Fig. 3 Three reproduction methods investigated in this study: (i) horizontal pair (red), (ii) vertical pair (green), and (iii) QWS by the LAF (blue).

Fig. 4 Schematic view of QWS by the LAF based on the Rayleigh I integral.

$$p(r_0) \equiv -\frac{1}{2\pi} \oint_C \frac{\partial}{\partial z} p(r_l) \cdot \frac{e^{-ik|r_0-r_l|}}{|r_0-r_l|} ds, \quad (3)$$

where p is the sound pressure, k is the wave number, C is the integral path along the LAF, z is the vertical direction of the display surface, and ds is a line element. In Eq. (3), $\frac{\partial}{\partial z} p(r_l)$ corresponds to the sound pressure gradient at each loudspeaker unit r_l and $\frac{e^{-ik|r_0-r_l|}}{|r_0-r_l|}$ corresponds to the sound wave at the listening position r_0 . Moreover, we assume $k|r_l-r_s| \ll 1$ under the geometrical condition that the imaginary sound source is close to the display for simplification.

By discretizing Eq. (3) using a discrete loudspeaker unit position indicated with a subscript n and converting the integral to a summation, we obtain the sound pressure at the listening position as

$$p(r_0) \simeq \frac{Kp_s}{2\pi} \sum_{n=1}^m \left(\frac{\cos \theta_n}{|r_{l_n}-r_s|^2} e^{-ik|r_{l_n}-r_s|} \times \frac{e^{-ik|r_0-r_{l_n}|}}{|r_0-r_{l_n}|} \right), \quad (4)$$

where K is a constant of proportionality used to adjust the sound pressure at the listening position and m is the number of loudspeaker units. In Eq. 4, the deformation

$$\frac{\partial}{\partial z} p(r_{l_n}) = \frac{\cos \theta_n}{|r_{l_n}-r_s|^2} p_s \cdot e^{-ik|r_{l_n}-r_s|} \quad (5)$$

is adopted, where p_s is the sound pressure of the imaginary sound source and θ_n is the angle between the direction from the imaginary sound source to the listening position and the direction from the imaginary sound source to each loudspeaker unit.

We obtain a concrete operation procedure for the QWS by the LAF from Eq. (4) as to apply the signal of an imaginary sound source with gain $\frac{\cos \theta_n}{|r_{l_n}-r_s|^2}$ and delay $\frac{|r_{l_n}-r_s|}{c}$ to each loudspeaker unit, where c is the sound velocity.

The Rayleigh integral is in principle valid in the case that the boundary plane is an infinite and an infinite number of loudspeakers are uniformly distributed. Then, we measured the wavefront of QWS by the LAF with a microphone array to evaluate the synthesized wavefront [7]. Although a moiré due to interference among the superimposed sound waves from each loudspeaker unit was observed in the measured wavefront, a secondary wavefront was approximately synthesized as an envelope of superimposed primary sound waves. In this paper, the localization synthesized by QWS by the LAF will be subjectively evaluated.

5. Methodology of subjective evaluation

5.1 Procedure and stimuli

Listeners were required to plot the center of a perceived sound image on an answer sheet on which the loudspeaker units and guide grids were printed. We used a full-band white noise with a duration of 1 s as a stimulus. The stimulus was repeated three times with 1 s intervals and 5 s was given to listeners to answer. Thus, each trial had a duration of 10 s. The listeners evaluated two different trial sequences which were randomized in each sequence to counterbalance the effect of the trial order. All stimuli were presented at a sound pressure level of 70 dB (A-weighted).

5.2 Evaluation room and equipment

We carried out the subjective evaluation in an acoustic evaluation room at NHK. The design of the room strictly adheres to Rec. ITU-R BS.1116-1 [18]. The reverberation time is 0.38 s at 500 Hz, the room dimensions are 6.4 m (W) $\times$ 8.0 m (D) $\times$ 4.5 m (H), and the room complies with NR-10.

Figure 5 shows the LAF for the 85-inch LCD. The size of the LAF is 2.3 m (W) $\times$ 1.4 m (H) $\times$ 0.08 m (D). It uses 16 units with 70 mm diameter and 86 loudspeaker units with 35 mm diameter. The large units were placed at intervals of 540 mm in the horizontal array as indicated by the red dashed circles in Fig. 5. At half height of each vertical array, a large unit was also placed. These large units were used in horizontal and vertical pairs. The small units were arranged with intervals of 60 mm. The QWS used 102 loudspeaker units in total. As a real sound source of FLc, FC, and FRc, we fabricated a loudspeaker with a large loudspeaker unit as indicated by the blue dashed circles.

We did not use a curtain to conceal the loudspeaker unit because even perforated curtain was found to in-

Fig. 5 LAF for 85-inch LCD display.

duce diffraction at its edge affecting the localization of a sound image in a preliminary experiment. To reduce the ventriloquism effect [19], which usually enhances localization on visible object, we added several dummy loudspeakers both inside and outside of the LAF.

5.3 Listeners

A total of 18 listeners participated in the evaluation, three females and 15 males in their 20s to 50s. They were engaged in audio or speech processing research except for one student who was studying audio signal processing. The listeners performed the evaluation alone. The height of the ears was adjusted to the center of the LAF. The listeners were asked not to move their heads during the evaluation, but their heads were not fixed to the chair.

6. Subjective evaluation I: Localization of the frontal channels of 22.2 ch reproduced by the real loudspeakers

In this section, we evaluate the localization accuracy of the frontal channels of 22.2 ch reproduced by the corresponding real loudspeaker indicated in Fig. 5 to examine an ideal condition.

Figure 6 shows the perceived localizations evaluated in front of FC, whose coordinate on the grid is (0, 0), at a distance of 1.9 m. Each mean value is presented as a circle with a 95% confidence interval (CI) in both the horizontal and vertical directions. A pink circle shows the actual location of a real loudspeaker both in Sections 6 and 7. The geometrical relation between the frontal channels and the listener is provided in the polar coordinate system in Appendix. We can estimate the localization accuracy from the length of the 95% CI and the closeness to the pink circle.

Figure 7 shows the perceived localizations evaluated in front of FLc (−2, 0) at a distance of 1.9 m.

It was found from Figs. 6 and 7 that most of the channels of the middle and bottom layers are upwardly perceived compared with the height of the reproducing loudspeaker. In contrast, the channels of the top layer localize at the position of the loudspeaker. In addition, the 95% CI in the vertical direction tends to be longer for the lower channels. Reasonable localization accuracy in the horizontal direction is achieved regardless the listening position.

Fig. 6 Perceived localizations evaluated in front of FC at a distance of 1.9 m.

Fig. 7 Perceived localizations evaluated in front of FLc at a distance of 1.9 m.

7. Subjective evaluation II: Localization of the synthesized channels on the display

In this section, we compare the localization accuracy of the three reproduction methods proposed in Section 4 with a real loudspeaker placed at each channel. Note that the localization accuracy is evaluated only for FLc, FC, and FRc, so that the outside of them is not covered in this study.

7.1 Evaluation in front of FC

Figures 8(a)-(c) show perceived localizations evaluated in front of FC (0, 0) at a distance of 1.9 m. A mean value is presented as a dot with a 95% CI in both the horizontal and vertical directions. A pink circle shows the actual location of a real loudspeaker.

In Fig. 8(a), the vertical pair and QWS by the LAF are found to localize at a similar position to the real

Fig. 8 Perceived localizations evaluated in front of FC channel at a distance of 1.9 m. The intended localizations are (a) FC, (b) FLC, and (c) FRc.

loudspeaker because the 95% CIs overlap each other. However, the horizontal pair is slightly shifted from the real loudspeaker.

Figure 8(b) shows that the horizontal and vertical pairs can localize at FLC as intended. QWS by the LAF localizes near the left array, which is a different from the other results. The reason why QWS by the

Fig. 9 Perceived localizations evaluated in front of FLC channel at a distance of 1.9 m. The intended localizations are (a) FC, (b) FLC, and (c) FRc.

LAF does not provide an appropriate localization will be discussed later in Section 8.

Figure 8(c) is a symmetrical configuration to Fig. 8(b) and the result is also symmetrical; thus, this result can be understood similarly to Fig. 8(b).

7.2 Evaluation in front of FLC

Figures 9(a)-(c) show perceived localizations evalu-

Table 1 Summary of the results obtained in Section 7. “✓” denotes *acceptable* and “N.A.” denotes *not acceptable*. “Center” corresponds to the listening position in front of FC, “Left” corresponds to FLc, and “Right” corresponds to FRc.

Reproduced channel	FC		FLc		FRc	
Listening position	Center	Left	Center	Left	Center	Left
Horizontal pair	✓	N.A.	✓	N.A.	✓	N.A.
Vertical pair	✓	✓	✓	✓	✓	✓
QWS by LAF	✓	✓	N.A.	N.A.	N.A.	N.A.

ated in front of FLc $(-2, 0)$ at a distance of 1.9 m.

The horizontal pair does not localize at FC at all as shown in Fig. 9(a). It is because the horizontal pair does not appropriately work in case that the listening position is horizontally different from the original position at which the amplitude panning was adjusted. QWS by the LAF has a comparatively long vertical 95% CI.

Figure 9(b) shows that the vertical pair almost overlaps with the real loudspeaker. On the other hand, the horizontal pair and QWS by the LAF localize on the left array, which is far from the intended position FLc.

Figure 9(c) shows the evaluation result when the intended localization and the listening position are placed on opposite sides of the center of the display. The horizontal pair is markedly shifted from the real loudspeaker. In addition, the 95% CI in the horizontal direction is very long, which implies that the synthesized sound image was unstable. The vertical pair overlaps with the real loudspeaker. Although QWS by the LAF localizes near the real loudspeaker, it does not overlap with the real loudspeaker. Thus, there is a significant difference between the localization of QWS by the LAF and that of a real loudspeaker.

7.3 Summary

Table 2 shows a summary of the results obtained in this section. We comprehensively judged each acceptability from the distance between the synthesized localization and the real loudspeaker and the overlap of 95% CI between the synthesized localization and the real loudspeaker. It is concluded from Table 2 that the vertical pair is the most appropriate method for the channels on the display among the proposed methods.

8. Discussion

8.1 Frontal channels of 22.2 ch reproduced by the real loudspeakers

The result obtained in Section 6 proved that the relative geometrical relation among the frontal channels

of 22.2 ch is maintained in the reproduction by the real loudspeakers. Each sound image can be separately perceived because the 95% CIs do not overlap with each other. In conclusion, the channels around the display are reproduced by the provided loudspeaker arrangement with sufficient separation.

The localization of channels in the middle and bottom layer is upwardly perceived. A similar phenomenon was reported by Tanaka *et al.* [20], who studied the localization accuracy of sound images at various angles of depression using a real loudspeaker. According to their experiment, a real loudspeaker placed at -40° in the median plane was perceived at a position of 5° higher than the loudspeaker itself. Furuya *et al.* synthesized a sound image among four loudspeakers placed at the vertices of a square [12]. They reported that an imaginary sound source synthesized below the horizontal plane is perceived at a higher position than intended. Previous studies lead us to consider that a sound image synthesized below the horizontal plane is apt to be perceived at a higher position than both the real loudspeaker and the intended position of the imaginary sound source. In any case, further studies on the perception of the sound images are still needed.

8.2 Synthesized channels on the display

The horizontal pair in Figs. 9 can be readjusted according to the shift of the listening position. However, the condition that the appropriate listening position is limited is unvarying in principle. This is the reason why we did not examine the readjustment of the horizontal pair.

Focusing on the vertical pair, the 95% CI is longer in the vertical direction than in the horizontal direction. This demonstrates that an imaginary sound source synthesized by a vertical pair is to some extent unstable in the vertical direction. Furuya *et al.* [12] and Barbour [21] also reported a tendency for a sound image synthesized by a vertical pair to be unstable. In the meantime, Kimura and Ando proved that the insta-

bility of a sound image synthesized by a vertical pair was markedly reduced by presenting the image together with a video [13]. This is the so-called ventriloquism effect [19], and thus, we can expect the improved localization of the channels on the display when 22.2 ch is simultaneously reproduced with 8K SHV video.

Another concern is that the vertical localization is shifted if the listening position is biased in the vertical direction. The ventriloquism effect will work provided the shift in the listening position is small. However, an appropriate localization will not be obtained if the listening position shifts to the edge of the display or outside the display. This issue strongly depends on the display size and the listening position, implying that a continuous study is still needed.

8.3 QWS by LAF

The reason why QWS by the LAF did not provide an appropriate localization is considered as follows. QWS by the LAF algorithm aims to distribute the signal level to each loudspeaker unit in accordance with the inverse square law. As mentioned in Section 4.3, the Rayleigh integral is valid in case that the boundary plane is infinite and an infinite number of loudspeakers are uniformly distributed. However, the LAF acts similarly to a combination of one vertical pair and one horizontal pair whose crosspoint corresponds to each imaginary sound source, in which the signal level distribution should follow the inverse law rather than the inverse square law. The observed localization error is due to this inconsistency.

8.4 Conclusive loudspeaker arrangement

Figure 10 illustrates the conclusive loudspeaker arrangement of the reproduction method for the 11 frontal channels of 22.2 ch over a display. A total of 12 loudspeakers are placed around the display, which are composed of eight loudspeakers for the channel around the display and four additional loudspeakers to synthesize FLc and FRc using the amplitude panning by the vertical pair of loudspeakers.

9. Conclusion

We subjectively evaluated a reproduction method for the 11 frontal channels of 22.2 ch in terms of localization accuracy. The results indicated that the channels around the display can be reproduced with real loudspeakers in contact with the edge of the display and that the three channels on the display can be stably localized by a vertical pair of loudspeakers for the 85-inch display. The obtained method is expected to be valid

Fig. 10 Loudspeaker arrangement of the reproduction method for the 11 frontal channels of 22.2 ch over a display.

for the display size close to 85 inch.

Acknowledgments

The LAF was jointly developed with Foster Electric Co. Ltd. The authors would like to thank Mr. Y. Morita and Mr. H. Sato for their implementation of the LAF.

Appendix

Following tables show the positions of the frontal channels of 22.2 ch in the polar coordinate system when the listener is in front of FC and FLc. r is the distance, ϕ is the azimuthal angle and φ is the elevation angle.

app.Table 1 Positions of the frontal channels in the polar coordinate system when the listener is in front of FC (Figs. 6 and 8).

Channel	r [m]	ϕ [°]	φ [°]
FL	2.2	30	0
FR	2.2	330	0
FC	1.9	0	0
FLc	2.0	16	0
FRc	2.0	344	0
TpFL	2.3	30	16.8
TpFR	2.3	330	16.8
TpFC	2.0	0	19.3
BtFC	2.0	0	-19.3
BtFL	2.3	30	-16.8
BtFR	2.3	330	-16.8

References

- 1) Recommendation ITU-R BT.2020, "Parameter values for ultra-high definition television systems for production and international programme exchange", International Telecommunication Union, Geneva (2012)
- 2) K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando: "A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)", SMPTE J. Vol. 117, No. 4, pp. 40-49 (2008)

app.Table 2 Positions of the frontal channels in the polar coordinate system when the listener is in front of FLc (Figs. 7 and 9).

Channel	r [m]	ϕ [°]	φ [°]
FL	2.0	16	0
FR	2.5	319	0
FC	2.0	344	0
FLc	1.9	0	0
FRc	2.2	330	0
TpFL	2.1	16	18.6
TpFR	2.6	319	14.7
TpFC	2.1	344	18.6
BtFC	2.1	344	-18.6
BtFL	2.1	16	-18.6
BtFR	2.6	319	-14.7

- 3) SMPTE ST 2036-2-2008, "Ultra high definition television - audio characteristics and audio channel mapping for program production" (2008)
- 4) Recommendation ITU-R BS.2051, "Advanced sound system for programme production", International Telecommunication Union, Geneva (2014)
- 5) T. Kumakura, M. Shiomi, S. Horino, S. Imai, and S. Mizushima: "Super Hi-Vision 8Kx4K direct-view LCD for next generation TV", Eurodisplay 2011, XXXI International Display Research Conference, 14.5 (2011)
- 6) T. Sugimoto, K. Matsui, and H. Okubo: "A loudspeaker array frame reproducing 22.2 multichannel sound for Super Hi-Vision flat panel display", 2012 NAB BEC Proceedings, pp. 16-21 (April 2012)
- 7) H. Okubo, T. Sugimoto, S. Oishi, and A. Ando: "A method for reproducing frontal sound field of 22.2 multichannel sound utilizing a loudspeaker array frame", The 133rd Convention of the Aud. Eng. Soc., Convention Paper 8714 (October 2012)
- 8) K. Matsui, S. Oishi, T. Sugimoto, S. Oode, Y. Nakayama, H. Okubo, A. Ando, H. Sato, K. Mizuno, and Y. Morita: "Reproduction of 22.2 multichannel sound with FPD-integrated loudspeakers for home use", 20th International Display Workshops (IDW'13), PDP2-1, pp. 679-682 (December 2013)
- 9) S. Oode, K. Matsui, S. Oishi, T. Sugimoto, and Y. Nakayama: "12-loudspeaker system for three-dimensional sound integrated with a flat-panel display", SMPTE J. Vol. 123, pp. 35-43 (2014)
- 10) K. Matsui and A. Ando: "Binaural reproduction of 22.2 multichannel sound with loudspeaker array frame", The 135th Convention of the Aud. Eng. Soc., Convention Paper 8954 (October 2013)
- 11) K. Ozawa and T. Furuya: "Effects of positioning with synthetic sound images on exploration of object in GUI", IPSJ J. Vol. 42, No. 6, pp. 1299-1310 (2001)
- 12) T. Furuya, K. Ozawa, and Y. Suzuki: "Two-dimensional localization of a phantom sound image controlled by the level differences among four loudspeakers in a vertical plane facing a listener", Acoust. Sci. & Tech. Vol. 25, pp. 493-495 (2004)
- 13) T. Kimura and H. Ando: "3D audio system using multiple vertical panning for large-screen multiview 3D video display", ITE Trans. Med. Tech. Appl. Vol. 2, No. 1, pp. 1-13 (2014)
- 14) F. Rumsey: "Spatial Audio", Focal Press, Oxford (2001)
- 15) V. Pulkki: "Virtual sound source positioning using vector base amplitude panning," J. Aud. Eng. Soc. Vol. 45, No. 6, pp. 456-466 (1997)
- 16) D. de Vries: "Wave Field Synthesis", Audio Engineering Society, New York (2009)
- 17) A. Berkhout, D. de Vries, and P. Vogel: "Acoustic control by wave field synthesis", J. Acoust. Soc. Am. Vol. 93, pp. 2764-2778 (1993)
- 18) Recommendation ITU-R BS.1116-1 "Method for subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunication

Union, Geneva (1994)

- 19) J. Sato, K. Fukue, Y. Kinoshita, and K. Ozawa: "Evaluation of the ventriloquism effect in the vertical influenced by audio reproduction systems", The Journal of The Institute of Image Information and Television Engineers, Vol. 63, No. 1, pp. 110-113 (2009)
- 20) Y. Tanaka, H. Hokari, and S. Shimada: "Sound localization accuracy versus depression angle in multichannel reproduction system", IEICE Tech. Rep. EA2009-56 (2009)
- 21) J. Barbour: "Elevation perception: phantom images in the vertical hemi-sphere", 24th International Conference of Aud. Eng. Soc. (2003)

Takehiro Sugimoto received his B.E. and M.E. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1999 and 2001, respectively. He also received a Ph.D. in information processing from Tokyo Institute of Technology, Tokyo, Japan, in 2013. He joined NHK in 2001 and has been working at Science & Technology Research Laboratories since 2004. His research interests are acoustic transducers, audio coding, and three-dimensional audio reproduction. He is currently engaged in MPEG and ARIB standardization.

Satoshi Oode received his B.S. degree in physics from Sophia University and his M.S. degree in computational intelligence and systems science from Tokyo Institute of Technology, Tokyo, Japan, in 1997 and 1999, respectively. He joined NHK in 1999 and worked as a video engineer in the Programs Engineering Center. Since 2001 he has been working at Science & Technology Research Laboratories. His research interests include emotion evoked by music, psychoacoustics, and three-dimensional acoustic space perception.

Yasushige Nakayama received his B.E. and M.E. degrees from the University of Iwate, Iwate, Japan, in 1992 and 1994, respectively. He joined NHK in 1994. He has mainly been engaged in the research and development of a three-dimensional sound system for television. He is currently a senior research engineer of NHK Science & Technology Research Laboratories.

Hiroyuki Okubo received his M.E. degree from Meiji University, Tokyo, Japan, and joined NHK in 1992. He has been on loan to NHK Engineering System, Inc. since 2014 and is currently a chief engineer. He has been engaged in the research and development of a 22.2 multichannel sound system for Super Hi-Vision.