Sharp-sound-image Construction Method Using Multichannel Sound System with Optimal Parametric Loudspeaker Arrangement

Yuna Harada*, Naoto Shimada*, Haonan Wang*, Kenta Iwai*, Masato Nakayama†, and Takanobu Nishiura*
* Ritsumeikan University, Shiga, Japan
E-mail: {is0426sh@ed, is0310hp@ed, h-wang@fc, iwai18sp@fc, nishiura@is}.ritsumei.ac.jp
† Osaka Sangyo University, Osaka, Japan
E-mail: nakayama@ise.osaka-sandai.ac.jp

***Abstract*—It is difficult to construct a sharp sound image using electro-dynamic loudspeakers (EDLs) due to wide directivity and high reverberation. A parametric array loudspeaker (PAL) can produce sharper directivity than an EDL due to the straightness of ultrasound. Therefore, a sharp sound image can be constructed with PALs. We previously proposed a sharp-sound-image construction method involving multiple PALs in addition to a conventional 22.2 surround sound system. Experimental results indicated that this method can construct a sharp sound image with high sound quality at any desired position in a horizontal plane. However, the optimal arrangement of PALs is insufficient in the horizontal and vertical planes. Therefore, we propose a sharp-sound-image construction method that involves a multichannel sound system with two optimal PAL arrangements to obtain more accurate sound localization. We conducted objective and subjective experiments to evaluate the proposed method, and the experimental results indicate the effectiveness of the method.**

I. Introduction

Three-dimensional (3-D) sound-field reproduction systems providing a high presence have attracted attention with the popularization of 3-D content. Conventional systems, such as channel-based systems [1], transaural systems [2], and object-based systems [3], [4], [5], have commonly use electro-dynamic loudspeakers (EDLs). In object-based systems, multiple audio objects are produced as sound images on the basis of amplitude panning [4]. Amplitude panning makes it possible to create sound fields through the arbitrary placement of any number of loudspeakers. Apparent source width (ASW) [6] is defined as the acoustical width of the sound source that is perceived by a listener. A smaller ASW produces an impression of a point-like source such as bees and birds. In contrast, a larger ASW makes an impression of a diffuse sound source such as rain and wind. Hence, a high presence can be provided using ASW control. However, the sound image constructed with EDLs is diffuse due to the reverberation when constructing a sharp sound image. In this paper, we focus on parametric array loudspeakers (PALs) [7]. A PAL can produce sharper directivity than an EDL by using the straightness of ultrasound. A PAL can construct a sharp sound image because the listener is mostly hearing the direct sound from the loudspeaker [8].

We previously proposed a sharp-sound-image construction method [9] with high-precision sound localization and high sound quality by combining a 22.2 surround sound system [1], [10] with PALs. With this method, PALs are used to create sharp sound images, and EDLs are used to reproduce low-frequency sounds, which are difficult to reproduce with PALs. The arrangement of the EDLs is also the same as a 22.2 surround sound system. A 22.2 surround sound system consists of EDLs with 22 regular channels and 2 low-frequency effect (LFE) channels, which has been standardized in ITU-R.BS.2051 [10] (Fig. 1). However, a sound image is constructed using two PALs with this method [9]. Amplitude panning was proposed for stereo sound systems with PALs [11], [12]. Our previous method experimental results indicated that this method [9] can construct a sharper sound image with high sound quality at the desired position in a horizontal plane. However, it is insufficient for the optimal PAL arrangement in the horizontal and vertical planes. This suggests that the performance of sound localization can be improved with optimal PAL arrangements by considering the characteristics of both EDLs and PALs.

We propose a sharp-sound-image construction method that involves a multichannel sound system with two optimal PAL arrangements. One arrangement (TRI-arrangement, where TRI = triangle) is an optimal arrangement for both horizontal and vertical planes. The other arrangement (RECT-arrangement, where RECT = rectangle) is an optimal arrangement for the horizontal plane. In the TRI-arrangement, the number of PALs in the upper layer is reduced compared with the middle layer so that the diagonal distance of the PALs is short. To improve the performance of sound localization, we place more PALs in the top layer in the RECT-arrangement. A preliminary experiment confirmed the optimal PAL arrangement in the horizontal plane and determines RECT-arrangement. In our evaluation experiment, we confirmed the effectiveness of the proposed method with these two arrangements.

II. Conventional Stereo Sound System with PALs

A stereo system with two PALs was proposed for accurate sound localization [11], [12]. Fig. 2 shows an overview of this

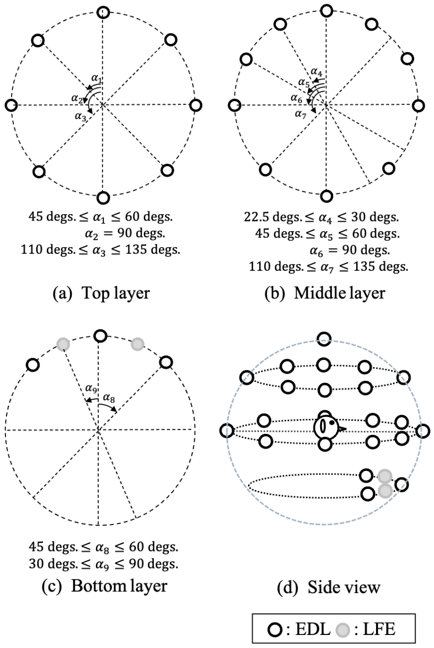


Fig. 1: Loudspeaker arrangement in 22.2 surround sound system.

conventional system. It renders a sound image by stereo amplitude panning. Stereo amplitude panning controls the inter-aural time difference (ITD) and inter-aural level difference (ILD) of the observed signals by weighting the emitting signals of the PALs [13], [14].

In Step 1 of Fig. 2, the PALs generate an amplitude modulated (AM) wave. The AM wave is radiated into the air at high sound pressure then self-demodulates to an audible sound due to the nonlinearity of the air. The demodulated sound inherits the straightness of ultrasound; therefore, the PAL can produce sharp directivity and construct a sharp sound image [8]. The AM wave is expressed as

$$u_{\mathrm{AM}}(t) = \{1 + \alpha s(t)\} u_{\mathrm{C}}(t), \tag{1}$$

where t denotes time, $s(t)$ denotes the input signal, $u_{\mathrm{AM}}(t)$ denotes the AM wave, $u_{\mathrm{C}}(t)$ denotes the carrier wave, and α $(0 < \alpha \leq 1)$ denotes the amplitude modulation factor.

In Step 2, the sound image is constructed at the desired position by amplitude panning. The weighting factors for each PAL are calculated on the basis of the positions of the listening point and PALs. These factors can be indicated as

$$W_{\mathrm{L}} = \frac{\sin\frac{\varphi}{2} + \sin\theta}{\sqrt{2(\sin^2\frac{\varphi}{2} + \sin^2\theta)}}, \tag{2}$$

$$W_{\mathrm{R}} = \frac{\sin\frac{\varphi}{2} - \sin\theta}{\sqrt{2(\sin^2\frac{\varphi}{2} + \sin^2\theta)}}, \tag{3}$$

where W_{L} and W_{R} denote the gain factors for the left and right channels, respectively, φ denotes the angle between the

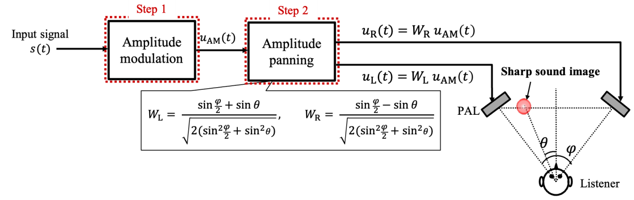


Fig. 2: Conventional stereo sound system with two PALs.

two PALs, and θ denotes the angle between the listening point and sound image. Then, from (1), (2), and (3), the emitting signals of each PAL are expressed as

$$u_{\mathrm{L}}(t) = W_{\mathrm{L}} u_{\mathrm{AM}}(t), \tag{4}$$

$$u_{\mathrm{R}}(t) = W_{\mathrm{R}} u_{\mathrm{AM}}(t), \tag{5}$$

where $u_{\mathrm{L}}(t)$ and $u_{\mathrm{R}}(t)$ denote the emitting signals of the left and right channels, respectively.

The sound observed at the listening point can be expressed by the sum of demodulated sounds from each channel, which can be indicated as:

$$y_i(t) = \mathbb{D}[g_{i,\mathrm{L}}(t), u_{\mathrm{L}}(t)] + \mathbb{D}[g_{i,\mathrm{R}}(t), u_{\mathrm{R}}(t)], \tag{6}$$

where $y_i(t)$ denotes the observed signal at the positions of each ear, i $(i \in \{\mathrm{L}, \mathrm{R}\})$ denotes the index of the left and right ear, $g_{i,\mathrm{L}}(t)$ and $g_{i,\mathrm{R}}(t)$ denote the impulse response between each channel and each ear, and $\mathbb{D}[\cdot,\cdot]$ denotes the self-demodulation process of the AM wave emitted from the PAL.

In theory, the conventional system can construct a sharper sound image at the desired position between two PALs. However, the directivity of a PAL differs from that of an EDL; therefore, it is necessary to set the loudspeakers in an optimal arrangement when constructing a system using PALs.

III. Proposed Sharp-sound-image Construction Method Using Multichannel Sound System with Optimal Parametric Loudspeaker Arrangement

Fig. 5 shows an overview of the multichannel sound system with the proposed method. The proposed method consists of N $(N \leq 22)$ PALs and M $(M \leq 22)$ EDLs. The PALs are used to construct the sound image. Low-frequency sound is difficult to reproduce with PALs; thus EDLs are used for this. The sound sources for low-frequency compensation are designed by weighting the input signal in the frequency domain so that the amplitude-frequency characteristic of the total perceived sound becomes flat. This low-frequency compensation makes it possible to achieve a high sound quality while constructing sharp sound images at the desired position.

The proposed method uses two optimal PAL arrangements for constructing a sound image in the front upper hemisphere, i.e., TRI-arrangement and RECT-arrangement. As shown in Fig. 3, the TRI-arrangement uses the same arrangement as in a 22.2 surround sound system. The number of PALs in the upper layer is reduced compared with the middle layer so that the diagonal distance of the PALs is short. In TRI-arrangement, nearby PALs form a triangle. We increase the number of the PALs in the top layer in the RECT-arrangement on the basis of the results from a preliminary experiment. As shown in Fig. 4, the RECT-arrangement uses the same arrangement in both the top and middle layers as in the middle layer of a 22.2 surround sound system. PALs are placed more closely together to improve the performance of sound localization in the horizontal plane. In this arrangement, nearby PALs form an approximate rectangle. The EDLs are arranged in the same manner.

In Step 1 of Fig. 5, amplitude modulation is carried out and the AM wave is generated, as shown in (1). In Step 2, the sound image is constructed at the desired position using multiple PALs that are closest to the desired sound image position. When the sound image is constructed at the position between two PALs, the signals are emitted from these two PALs. In this case, the emitting signal for the first PAL $u_1(t)$ can be calculated in the same manner as (4), and the emitting signal for the other PAL $u_2(t)$ can be calculated in the same manner as (5). In Step 3, the low-frequency compensation sound $s_{\mathrm{LPF}}(t)$ is designed by convolving the low-pass filter with $s(t)$. The $s_{\mathrm{LPF}}(t)$ is expressed as

$$s_{\mathrm{LPF}}(t) = h_{\mathrm{LPF}_{\Omega_{\mathrm{C}}}}(t) * s(t), \tag{7}$$

where $*$ denotes the convolution operator, and $h_{\mathrm{LPF}_{\Omega_{\mathrm{C}}}}(t)$ denotes a low-pass filter, with cutoff frequency Ω_{C}. In Step 4, the compensation sound is divided and emitted from M EDLs. The sound images are then constructed at the desired position with high sound quality. The emitting signals for the EDLs can be expressed as

$$s_m(t) = \beta \frac{s_{\mathrm{LPF}}(t)}{\sqrt{M}}, \tag{8}$$

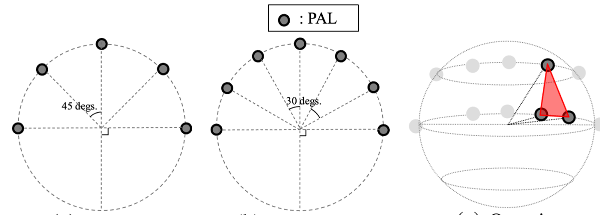
where $s_m(t)$ denotes the input signal for m-th $(1 \leq m \leq M)$ EDLs, and β denotes the gain normalization parameter for the PALs and EDLs and should be measured in advance to flatten the power spectrum of the sound image. Finally, the sound observed at the listening point is expressed as

$$y_i(t) = \sum_{n=1}^{N} \mathbb{D}[g_{i,n}(t), u_n(t)] + \sum_{m=1}^{M} f_{i,m}(t) * s_m(t). \tag{9}$$

where $f_{i,m}(t)$ is the impulse response between m-th EDL and each ear. Therefore, the proposed method can construct sharper sound images at the desired position with high sound quality.

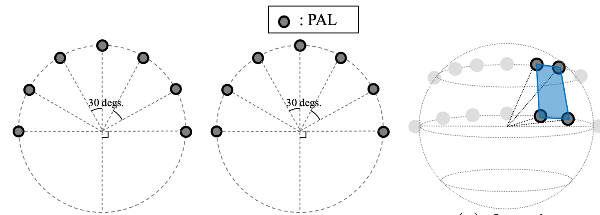
IV. Preliminary Experiment for Determining Optimal PAL Arrangement in Horizontal Plane

With our previous method, a sound image is constructed using two PALs. Therefore, it can be considered that the positions of the PALs and distance between them affect the sharpness of the sound image and accuracy of sound localization. From studying the optimal arrangement of the PALs, we carried out a preliminary experiment in the simplest scenario, i.e., constructing a sound image in the horizontal plane.



(a) Top layer (b) Middle layer (c) Overview

Fig. 3: TRI-arrangement.



(a) Top layer (b) Middle layer (c) Overview

Fig. 4: RECT-arrangement.

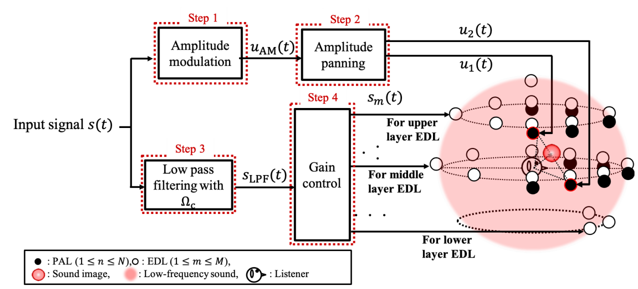


Fig. 5: Proposed method with 12 PALs and 22 EDLs.

A. Experimental Conditions

Table I lists the experimental conditions, and Table II lists the equipment used in this preliminary experiment. Fig. 6 illustrate the loudspeaker arrangement and experimental conditions of the preliminary experiment. We compared the sound images constructed with two EDLs or PALs with the real sound source at the sound image's position in the horizontal plane, as follows:

- REAL: Real sound source (a single EDL) is placed at the position of the sound image in Fig. 6
- EDL-2CH: EDLs are placed in the loudspeaker positions shown in Fig. 6
- PAL-2CH: PALs are placed in the loudspeaker positions shown in Fig. 6

In Fig. 6, θ_{s} denotes the direction of the sound image, which was set to 0–15 degs. in steps of 5 degs. in the preliminary experiment, d denotes the distance between the loudspeaker and listening point, which was set to 1.5, 1.9, and 2.5 m, and φ denotes the angle between two loudspeakers, which was set to 15–90 degs. in steps of 15 degs. The distance between two loudspeakers l can be calculated as

$$l = 2d \sin\left(\frac{\varphi}{2}\right). \tag{10}$$

The performance of sound localization was evaluated with the ILD and ITD by comparing the real sound source and sound images constructed from loudspeakers. The sharpness of the sound image was also evaluated with inter-aural cross-correlation (IACC) [15], [16]. The ILD is the difference in the level of acoustic signals arriving at the left and right ears, which is calculated as

$$\mathrm{ILD} = 10\log_{10}\frac{\sum_{k=0}^{K-1} s_{\mathrm{O,L}}^2(k)}{\sum_{k=0}^{K-1} s_{\mathrm{O,R}}^2(k)}, \tag{11}$$

where k denote time index, K denotes signal duration, and $s_{\mathrm{O,L}}(k)$ and $s_{\mathrm{O,R}}(k)$ denote the signals observed at the left and right ears, respectively. A larger ILD suggests that the sound pressure difference is more significant between two ears so that the sound image is easier to be perceived on the left.

The ITD is the time difference between the acoustic signals arriving at the left and right ears and is calculated with the IAC function (IACF) [16] as

$$\mathrm{ITD} = \arg\max_{\tau}(\mathrm{IACF}(\tau)), \tag{12}$$

$$\mathrm{IACF}(\tau) = \frac{\sum_{k=0}^{K-1} s_{\mathrm{O,L}}(k) s_{\mathrm{O,R}}(k+\tau)}{\sqrt{\sum_{k=0}^{K-1} s_{\mathrm{O,L}}^2(k)\sum_{k=0}^{K-1} s_{\mathrm{O,R}}^2(k)}}, \tag{13}$$

where τ denotes the time index ($|\tau| \times T_s \leq 1.0 \times 10^{-3}$). Here, T_s denotes sampling period. A larger value of ITD means that sound arrives quickly to the left ear than the right ear so that the sound image is easier to be perceived on the left.

The IACC is the degree of correlation between the arrived signals at the left and right ears, which is calculated as

$$\begin{aligned}\mathrm{IACC} &= \max_{\tau}\left(\mathrm{IACF}(\tau)\right),\\ &= \mathrm{IACF}(\mathrm{ITD}).\end{aligned} \tag{14}$$

A larger IACC means that the method can render a sharper sound image.

TABLE I: Experimental conditions

Reverberation time (T_{60})	350 ms
Ambient noise level (L_{A})	37.8 dB
Sampling frequency	96 kHz
Carrier frequency	40 kHz
Quantization	32 bits
Sound source	White noise (0–8 kHz, 3 s)

TABLE II: Experimental equipment

PAL	MITSUBISHI, PS-50E
EDL	FOSTEX, FE83En
Binaural microphone	3DIO, Free Space Pro II
Power amplifier for PAL	YAMAHA, XM4180
Power amplifier for EDL	YAMAHA, XMV8280
A/D converter	RME, FIREFACE UFX
D/A converter	RME, M-32DA

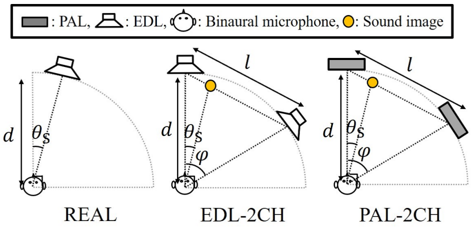


Fig. 6: Experimental loudspeaker arrangements for sharp sound image in horizontal plane (top view).

B. Experimental Results

Figs. 7 and 8 show the results for the ILD and ITD in the horizontal plane under each condition. The performance of sound localization worsened as φ increased while d remained invariant. It also worsened as d increased while φ remained invariant. Considering the relationship among d, φ, and l as shown in (10), a conclusion can be drawn that the performance of sound localization worsens as l increases. From Fig. 7 (e), the performance of sound localization was maintained when $l \leq 1.3$ m. Also, from Figs. 7 (b) and (d), it defused when $l \geq 1.5$ m. Therefore, we can have an experimental approximation that a sharp sound image can be constructed with accurate sound image localization when $l \leq 1.4$ m. The PALs should be placed in a narrower arrangement than the EDLs to achieve an equivalent sound localization. This can be explained by the narrow directivity of a PAL.

Fig. 9 shows the IACC in the horizontal plane under each condition. The IACC of a PAL was larger than that of an EDL under every condition. Therefore, PALs can construct shaper sound images than EDLs. When the sound image is constructed in the direction of a certain loudspeaker (θ_{S} = 0), the IACC of a PAL reaches a higher level than that of an EDL. This is because a PAL's sharp directivity leads to low reverberation. Different from sound localization, the sharpness of a sound image does not show a significant relevance to the loudspeaker arrangement when using PALs. The sharpness of a sound image should be considered as an innate characteristic of a PAL.

From these results, the optimal PAL arrangement in the horizontal plane depends on the distance between loudspeakers. A narrower arrangement can result in more accurate sound localization, while sharpness has a very weak correlation with the arrangement parameters we studied. It should be noted that only one type of PAL was used in this experiment. When using other PALs, the directivity, or radiation area should be considered.

V. Evaluation Experiments for Optimal PAL Arrangement in Horizontal and Vertical Planes

In evaluation experiments, we evaluated the proposed method regarding the sharpness of sound image and performance of sound localization. The two optimal arrangements of the proposed method were compared by experiments.

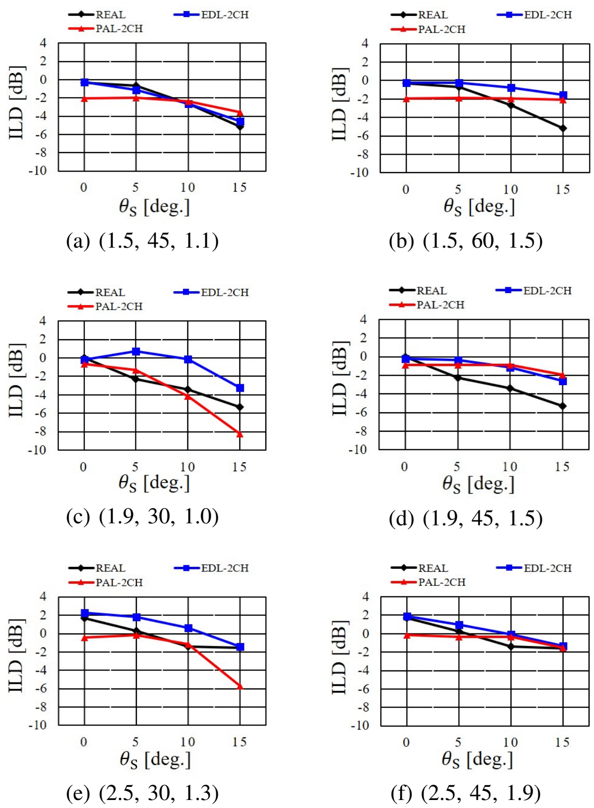


Fig. 7: Preliminary experimental results for ILD in horizontal plane under each condition (d [m], φ [degs.], l [m]).

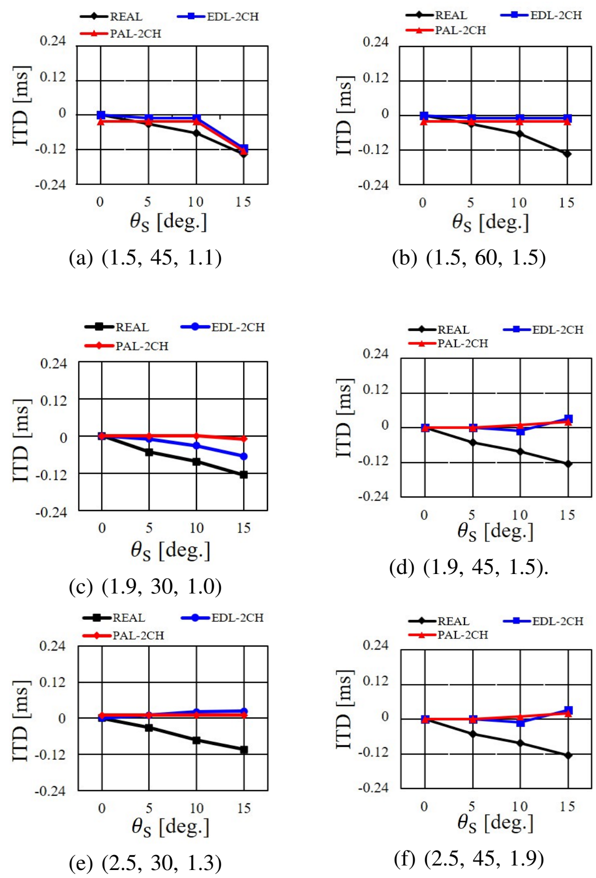


Fig. 8: Preliminary experimental results for ITD in horizontal plane under each condition (d [m], φ [degs.], l [m]).

We consider constructing a sound image in the front upper hemisphere, as a common scenario.

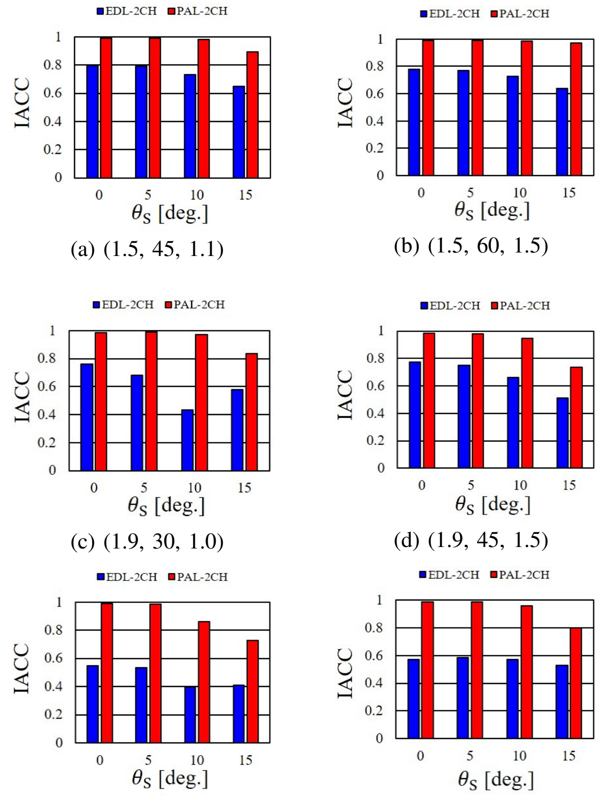
A. Experimental Conditions

Fig. 10 shows the TRI-arrangement, in which the top and middle layers are in the same arrangement as in a 22.2 surround sound system. The diagonal distance of the PALs is short in this arrangement. Fig. 11 shows the RECT-arrangement, in which the top and middle layers are the same as the middle layer of a 22.2 surround sound system. The RECT-arrangement uses more loudspeakers than the TRI-arrangement in the top layer. The upper loudspeakers are placed more densely in the horizontal plane.

As shown in Figs. 10 and 11, we constructed sound images with two PALs placed diagonally. The positions of sound images can be decomposed into the horizontal and vertical directions. The performance of sound localization of the horizontal direction $\psi_{H,S}$ was evaluated with the ILD and ITD by comparing the real sound source and sound images constructed with loudspeakers. The performance of sound localization of the vertical direction $\psi_{V,S}$ is difficult to evaluate through an objective experiment, so we evaluated it through a subjective experiment. The sharpness of the sound image was also subjectively evaluated with the IACC.

In the objective experiment, we used white noise (0–8 kHz, 3 s) and a time-stretched pulse signal [17] (0–8 kHz, 4 s) as the input signal. The other experimental conditions and equipment were the same as those listed in Tables I and II. As shown in Figs. 10 and 11, the position of sound image ψ_S was set to 0–40 degs. in steps of 5 degs.

In the subjective experiment, we let participants answer their perceived direction of the sound image, and evaluate the perceived sharpness of the sound image. The same input signal, experimental conditions, and equipment were set the same as those of the objective experiment. The sound image was presented twice in random order of three positions (positions of each loudspeaker, midpoint of two loudspeakers). The sharpness of the sound image was evaluated using mean opinion score (MOS) [18]. The real sound source was set as a reference and it has a standard score of 3. Table III shows the MOS score for the sharpness of the sound image constructed by loudspeakers with TRI-arrangement and RECT-arrangement.



(a) (1.5, 45, 1.1) (b) (1.5, 60, 1.5)

(c) (1.9, 30, 1.0) (d) (1.9, 45, 1.5)

(e) (2.5, 30, 1.3) (f) (2.5, 45, 1.9)

Fig. 9: Preliminary experimental results for IACC in horizontal plane under each condition (d [m], φ [degs.], l [m]).

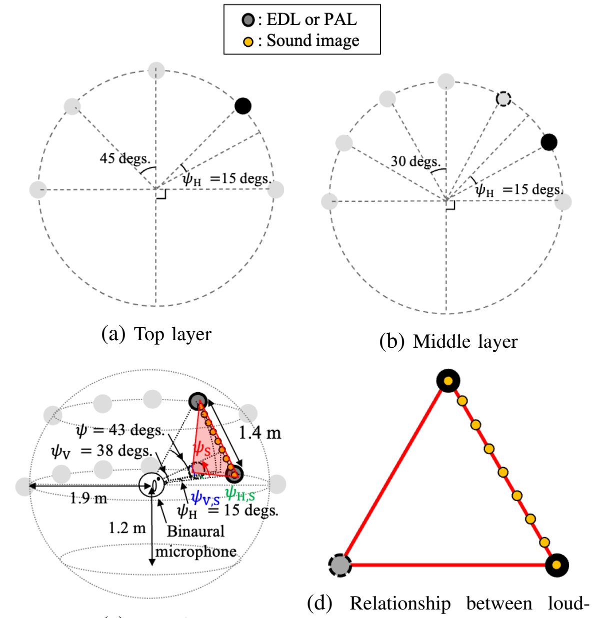
TABLE III: MOS score for sharpness of sound image

Sharpness of sound image	Score
Much sharper	5
Slightly sharper	4
About the Same	3
Slightly lack sharpness	2
Lack sharpness	1

B. Experimental Results from Objective Evaluation in Horizontal Plane

In the objective experiments, the ILD and ITD were measured to evaluate sound localization performance in the horizontal plane. Fig. 12 shows the experimental results for the ILD, and Fig. 13 shows those for the ITD. The IACC was measured to evaluate the sharpness of the sound image in the horizontal plane, and those results are shown in Fig. 14.

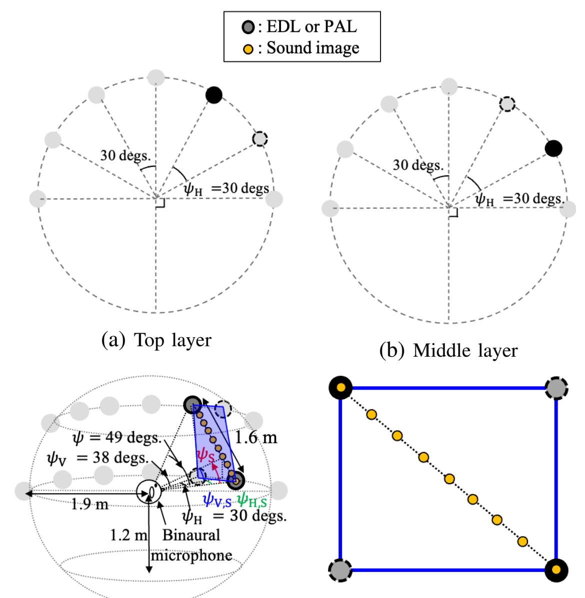
From Figs. 12 and 13, the RECT-arrangement showed better performance for sound localization in the horizontal plane of PALs than the TRI-arrangement. A more significant difference of ILDs can be noticed between constructing a sound image in front ($\psi_{\mathrm{H,S}}$ is small) and constructing a sound image on



(a) Top layer (b) Middle layer

(c) Overview (d) Relationship between loudspeakers and sound images

Fig. 10: Experiment arrangement of loudspeakers (TRI-arrangement).

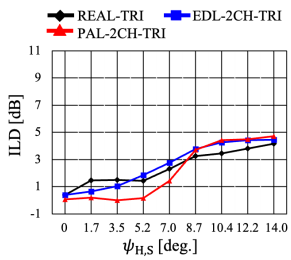


(a) Top layer (b) Middle layer

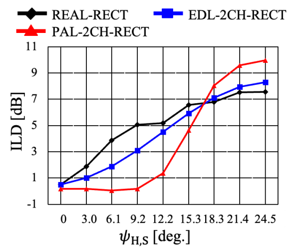
(c) Overview (d) Relationship between loudspeakers and sound images

Fig. 11: Experiment arrangement of loudspeakers (RECT-arrangement).

the side ($\psi_{\mathrm{H,S}}$ is large) when using the RECT-arrangement. Therefore, it can be considered that the RECT-arrangement can provide a marked perception of direction. From Fig. 14, PAL can construct a sharp sound image in both arrangements. Since the real sound source is an omnidirectional EDL, it is

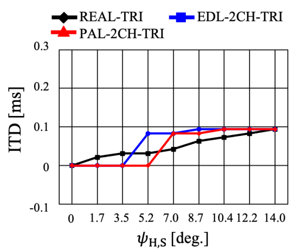


(a) with TRI-arrangement

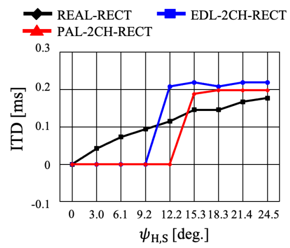


(b) with RECT-arrangement

Fig. 12: ILD results on experiment for direction in horizontal plane.

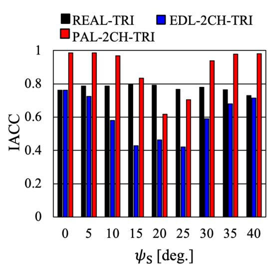


(a) with TRI-arrangement

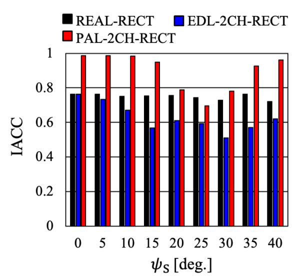


(b) with RECT-arrangement

Fig. 13: ITD results on experiment for direction in horizontal plane.



(a) with TRI-arrangement



(b) with RECT-arrangement

Fig. 14: IACC results on experiment for sharp sound image in horizontal plane.

reasonable to obtain a sharper sound image with PALs than the real sound source. These results indicate that the RECT-arrangement is the optimal arrangement considering only the horizontal plane.

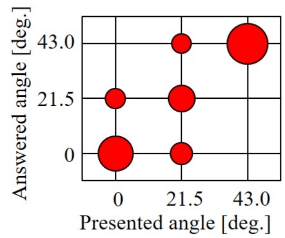
C. Experimental Results from Subjective Evaluation in Horizontal and Vertical Planes

Fig. 15 shows the experimental results of direction perception with each arrangement, and Table IV shows the overall correct answer rate of the direction perception with each system. Fig. 16 shows the experimental results of the sharpness perception with each arrangement.

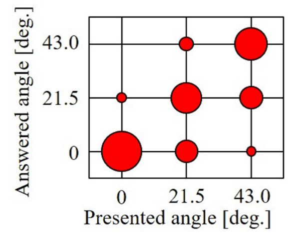
From Fig. 15, the TRI-arrangement was better in performance of sound localization in the horizontal and vertical planes than the RECT-arrangement. This is because the distance between the loudspeakers in the TRI-arrangement (l = 1.4 m) meets the condition of the optimal arrangement derived in Section IV-B. From Table IV, the RECT-arrangement with EDL shows the highest performance. This is because EDL can construct a sound image in both arrangements. Also, it is easy for participants to answer because the distance of

TABLE IV: Overall correct answer rate for sound images in vertical plane

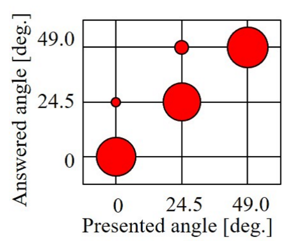
Condition	Correct answer rate [%]
EDL-2CH-TRI	72.9
PAL-2CH-TRI	70.8
EDL-2CH-RECT	93.8
PAL-2CH-RECT	64.6



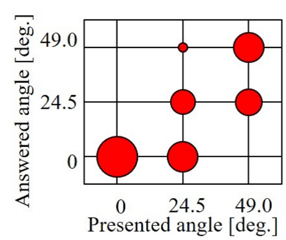
(a) with EDL-2CH-TRI



(b) with PAL-2CH-TRI

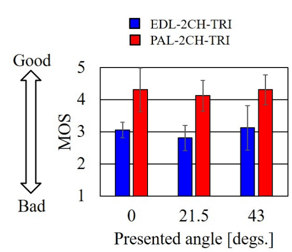


(c) with EDL-2CH-RECT

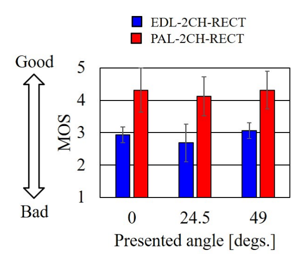


(d) with PAL-2CH-RECT

Fig. 15: Perceived direction of sound image in horizontal and vertical planes.



(a) with TRI-arrangement



(b) with RECT-arrangement

Fig. 16: Perception results for sharp sound image in horizontal and vertical planes.

sound image construction positions in the RECT-arrangement is longer than the TRI-arrangement. From Fig. 16, a PAL can construct a sharp sound image in both arrangements. These results indicate that the TRI-arrangement is the optimal arrangement considering both horizontal and vertical planes.

D. Summary and Discussion

In the evaluation experiments, we compared the TRI-arrangement and RECT-arrangement. The sharpness of the sound image and sound localization performance were evaluated with objective metrics and subjective perception. The experimental results indicate that the RECT-arrangement can produce more accurate sound localization and construct a sharper sound image than the TRI-arrangement considering the horizontal plane. The subjective perception results indicate that the TRI-arrangement can produce more accurate sound localization than the RECT-arrangement considering both horizontal and vertical planes. Though RECT-arrangement uses more PALs than the TRI-arrangement, as shown in Figs. 10 and 11, the distance between the PALs, which are placed diagonally l is larger than that in the TRI-arrangement. We confirmed that the TRI-arrangement is the optimal arrangement considering both horizontal and vertical planes, and the RECT-arrangement is the optimal arrangement considering only the horizontal plane.

VI. Conclusions

We proposed a sharp-sound-image construction using a multichannel sound system with two optimal arrangements of the PALs. With the proposed method, PALs are used to create a sharp sound image, and EDLs are used to reproduce the low-frequency sounds which are difficult for the PALs to reproduce. We also investigated the optimal arrangement of PALs and introduced two arrangements: TRI-arrangement and RECT-arrangement. We confirmed that the TRI-arrangement is the optimal arrangement considering both horizontal and vertical planes, and the RECT-arrangement is the optimal arrangement considering only the horizontal plane. In the future, we will study the ASW control by using PALs and EDLs to provide more higher presence. For example, we can consider using two-way speakers that use a PAL as a tweeter and EDL as a woofer.

Acknowledgment

This work was partly supported by JST COI, and JSPS KAKENHI Grant Numbers JP19H04142, and JP21H03488.

References

[1] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando, “A 22.2 Multichannel Sound System for Ultrahigh-Definition TV (UHDTV),” SMPTE Motion Imaging Journal, vol. 117, no. 3, pp. 40–49, 2008.

[2] H. Kurabayashi, M. Otani, K. Itoh, M. Hashimoto, and M. Kayama, “Development of Dynamic Transaural Reproduction System Using Non-Contact Head Tracking,” Proceedings of 2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE), pp. 12–16, 2013.

[3] Dolby Laboratories, Inc.: https://www.dolby.com/us/en/technologies/home/dolby-atmos.html

[4] V. Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning,” Journal of the Audio Engineering Society, vol. 45, no. 6, pp. 456–466, 1997.

[5] J. J. Lopez, P. Gutierrez, M. Cobos, and E. Aguilera, “Sound Distance Perception Comparison between Wave Field Synthesis and Vector Base Amplitude Panning,” Proceedings of 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP), pp. 165–168, 2014.

[6] S. Sato, and Y. Ando, “Apparent source width (ASW) of complex noises in relation to the interaural cross-correlation function,” Journal of Temporal Design in Architecture and the Environment, vol. 2, pp. 29–32, 2002.

[7] M. Nakayama and T. Nishiura, “Distance Control of Virtual Sound Source Using Parametric and Dynamic Loudspeakers,” Proceedings of The 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 1262–1267, 2018.

[8] E. L. Tan, W. S. Gan, and C. H. Chen, “Spatial Sound Reproduction Using Conventional and Parametric Loudspeakers,” Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 1–9, 2012.

[9] N. Shimada, K. Iwai, M. Nakayama, and T. Nishiura, “High-Presence Sharp Sound Image Based on Sound Blending Using Parametric and Dynamic Loudspeakers,” Journal of Signal Processing, vol. 24, no. 4, pp. 171–174, 2020.

[10] ITU-R Rec. BS.2051-2, “Advanced sound system for programme production,” International Telecommunications Union, 2018.

[11] S. Aoki, M. Toba, and N. Tsujita, “Sound localization of stereo reproduction with parametric loudspeakers,” Applied Acoustics, vol. 73, no. 12, pp. 1289–1295, 2012.

[12] S. Aoki, K. Shimizu, and K. Itou, “Study of vertical sound image control with parametric loudspeakers,” Applied Acoustics, vol. 116, pp. 164–169, 2017.

[13] J. Blauert, “Spatial hearing: the psychophysics of human sound localization,” MIT Press, 1983.

[14] D. M. Leakey, “Some measurements on the effects of interchannel intensity and time differences in two channel sound systems,” The Journal of the Acoustical Society of America, vol. 31, no. 7 pp. 977–986, 1959.

[15] T. Okano, L. L. Beranek, and T. Hidaka, “Relations among interaural cross-correlation coefficient (IACC E), lateral fraction (LFE), and apparent source width (ASW) in concert halls,” The Journal of the Acoustical Society of America, vol. 104, no. 1, pp. 255–265, 1998.

[16] S. Sato, and Y. Ando, “On the apparent source width (ASW) for bandpass noises related to the IACC and the width of the interaural cross-correlation function (W IACC),” The Journal of the Acoustical Society of America, vol. 105, no. 2, pp. 1234–1234, 1995.

[17] N. Aoshima, “Computer-generated pulse signal applied for sound measurement,” The Journal of the Acoustical Society of America, vol. 69, no. 5, pp. 1484–1488, 1981.

[18] Rec. ITU-R P. 800, “Methods for subjective determination of transmission quality,” 1996.