# Virtual Sound Source Rendering Based on Distance Control to Penetrate Listeners Using Surround Parametric-array and Electrodynamic Loudspeakers

Takuma Ekawa*, Masato Nakayama* and Toru Takahashi*
* Osaka Sangyo University, Osaka, Japan
E-mail: nakayama@ise.osaka-sandai.ac.jp Tel: +81-72-875-3001

***Abstract*—This paper proposes a new virtual sound source rendering method based on distance control to penetrate listeners using surround parametric-array loudspeakers (PALs) and electrodynamic loudspeakers (EDLs). Our method utilizes direct-to-reverberation ratio (DRR) control of room impulse responses and in-head localization based on interaural level difference (ILD) control with a virtual headphone. For distance control between the listener and the remote loudspeakers, we employ the DRR based on the amplitude panning for surround PALs and EDLs. For distance control centered on the listener, we use ILD control based on amplitude panning for the virtual headphones using surround PALs. Additionally, the type of control used is switched at a certain boundary distance, and gain coefficients are calculated for the positions and times of the moving sound images to achieve effective virtual sound source rendering. We also report on the results of evaluation experiments conducted to confirm the effectiveness of the proposed method.**

## I. Introduction

Three-dimensional sound field reproductions [1], [2], [3], [4] have been proposed as ways to achieve sound fields with high presence levels in theaters, on virtual live stages, and for telecommunications purposes. Such methods include binaural systems [5] based on microphones and headphones mounted in dummy heads, in which the head-related transfer function (HRTF) provides the handover function between the sound source and the listener's ears. In this method, HRTFs are measured under various conditions in advance, and binaural sounds are created by convolving a dry source and the HRTFs. The binaural sounds are then presented to the listeners via headphones. Although these systems are effective, they require the measurement of multiple personal HRTFs, which is very difficult. In contrast, transaural systems [6], which use numerous remote loudspeakers, reproduce sound fields by setting up appropriate multi-channel inverse filters between the remote loudspeakers and the listener's ears. However, these systems require expensive computations as well as numerous loudspeakers. Therefore, at present, surround-sound audio systems [7], which include channel-, object-, and scene-based systems, provide the most practical way to achieve sound fields with high presence levels. To date, proposed channel-based stereo systems include 5.1 ch and 22.2 ch surround systems [8][9]. However, those systems still present sound sources at the loudspeaker positions and thus require numerous loudspeakers to achieve effective sound fields. Separately, object-based systems, such as that proposed by Dolby Atmos [10], use numerous loudspeakers to present multiple sound objects that are positioned as virtual sound sources (VSSs) based on amplitude panning [11], [12], while scene-based systems, such as high-order ambisonic (HOA) [13] and wave field synthesis (WFS) systems [1], [14] present VSSs using numerous loudspeakers. However, in cases involving small-scale loudspeaker setups, these surround systems have distance-related problems when presenting VSSs. To address this problem, we previously focused on the direct-to-reverberant ratio (DRR) [15], which is a distance localization cue, and proposed a distance control method for VSSs that uses small numbers of parametric-array loudspeakers (PALs) and electrodynamic loudspeakers (EDLs) [16]. PAL systems[17], which can achieve superior directivity using rectilinear ultrasonic waves, emit sounds with high-amplitude direct waves and low-amplitude reverberations (i.e., high DRRs), which cause their VSSs to be perceived as being closer to the listener. In contrast, the emitted sounds for EDL systems have low DRRs, which cause their VSSs to be perceived as being farther away from the listeners. In our previously proposed method, we utilized the DRR differences between PAL and EDL loudspeakers to present VSS movement. However, that method had difficulty presenting a VSS capable of penetrating the listeners because the moving VSS was stopped at an out-of-head position. In this paper, penetrating the listeners is defined as the VSS movements connecting an out-of-head position to another out-of-head position through an in-head position. This can provide experience of augmented reality (AR) beyond reality in sound AR contents including the tele-existence systems, AR movies and AR attractions.

With that issue in mind, this paper proposes a new VSS rendering technique based on distance control for surround PALs and EDLs that is capable of penetrating listeners. To accomplish this, we utilize DRR control of room impulse responses and in-head localization based on the interaural level difference (ILD), which is a near field distance localization cue, to control virtual headphones. To facilitate distance control between the listeners and the surround loudspeakers, the proposed method employs DRR control based on amplitude panning for surround PALs and EDLs. For distance control that penetrates the listener, the proposed method employs ILD control using surround PALs based on amplitude panning for the virtual headphone. To achieve VSS rendering, the type of

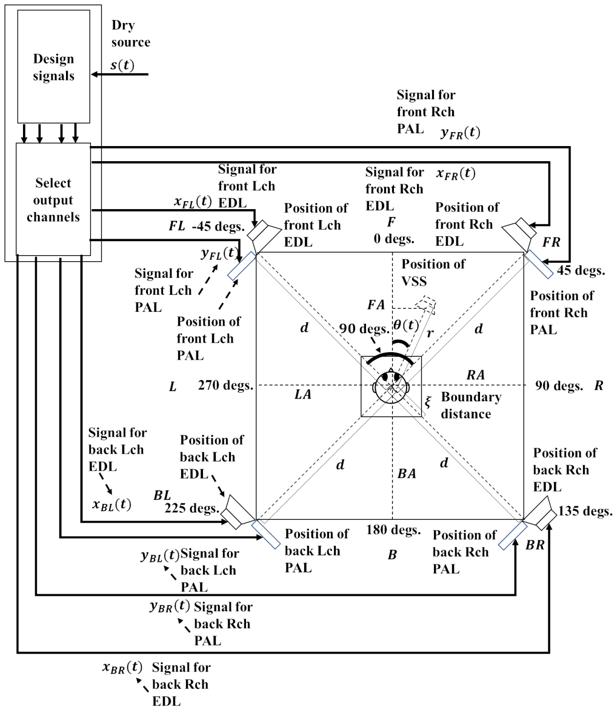



Fig. 1. Overview of the proposed method.

control used is switched at a certain boundary distance, and gain coefficients are calculated for the positions and times of the VSS movements. The effectiveness of the proposed method was confirmed through evaluation experiments.

## II. VSS rendering based on distance control to penetrate listeners using PALs and EDLs

### A. Overview of the proposed method

As stated above, this study proposes a new VSS rendering method based on distance control to penetrate listeners using surround PALs and EDLs. Figure 1 shows an overview of the proposed method, in which we use four PALs and four EDLs. In this figure, the F section covers the $0°$ directions, the L section covers the $270°$ directions, the R section covers the $90°$ directions, and the B section covers the $180°$ directions. The F section is split between FL $(-45°)$ and FR $(45°)$, while the B section is split between BL $(225°)$ and BR $(135°)$. Additionally, FA is the front area $(-45° \sim 45°)$, RA is the right area $(45° \sim 135°)$, BA is the back area $(135° \sim 225°)$, and LA is the left area $(225° \sim 315°)$. Furthermore, $t$ is time, $r(t)$ is the VSS distance, and $\theta(t)$ is the VSS direction. Finally, $r(t)$ and $\theta(t)$ are the input parameters. In this method, the VSS can be positioned in a location where it is surrounded by the FL, FR, BR, and BL, as shown in Fig. 1. The proposed method switches two processes for the out-of-head and in-head positions at a boundary distance of $\xi$. For the out-of-head positions, four PALs and four EDLs are utilized to control the DRR. For the in-head positions, the ILD is controlled by four PALs. The output signals are projected from the VSS coordinates of each area to the FA, which is their intended destination. Next, to reinforce projection to the VSS coordinates of each area, the output channels of the designed signals are selected. The proposed method designs output signals by calculating the gain coefficients for the positions and times of the VSS movements. Thus, in a moving VSS, the gain coefficients for the output signals become a weighting function in the time domain. Section II-B describes the output signal design and the output channel selection process for the out-of-head position, while Sec. II-C describes the output signal design and the output channel selection process for the in-head position.

### B. Output signal design and output channel selection for the out-of-head position.

First, the projection of the VSS direction from each area to the FA is calculated as follows:

$$\theta_F(t) = \begin{cases} \theta(t) & \text{if } FA: (-45° \leq \theta(t) < 45°), \\ \theta(t) - 90°, & \text{if } RA: (45° \leq \theta(t) < 135°), \\ \theta(t) - 180°, & \text{if } BA: (135° \leq \theta(t) < 225°), \\ \theta(t) - 270°, & \text{if } LA: (225° \leq \theta(t) < 315°), \end{cases} \tag{1}$$

where $\theta_F(t)$ is the projected VSS direction and $\theta(t)$ is the original VSS direction shown in Fig. 2. The out-of-head condition is as follows:

$$\frac{\xi \cos 45°}{\cos(\theta_F(t))} \leq r(t) \leq \frac{d \cos 45°}{\cos(\theta_F(t))}, \tag{2}$$

where $\xi$ is the boundary distance, $d$ is the distance between the listener and the remote loudspeaker, and $r(t)$ is the VSS distance.

Figure 2 shows an overview of the output signal design for the out-of-head condition. As can be seen, the output signals for the out-of-head position are designed so that the VSS is located in the FA and two PALs and two EDLs are utilized for the FA. The weighting functions for the VSS direction are calculated as follows:

$$g_L(t) = \sqrt{\frac{45° - \theta_F(t)}{90°}}, \tag{3}$$

$$g_R(t) = \sqrt{\frac{45° + \theta_F(t)}{90°}} \tag{4}$$

where $g_L(t)$ and $g_R(t)$ are the weighting functions for the directions of the left- and right-side loudspeakers, respectively. The weighting functions for the VSS distance are calculated as follows:

$$\alpha_F(t) = \sqrt{\frac{r(t)\cos(\theta_F(t)) - \xi \cos(45°)}{(d - \xi)\cos(45°)}}, \tag{5}$$

$$\beta_F(t) = \sqrt{\frac{d \cos(45°) - r(t)\cos(\theta_F(t))}{(d - \xi)\cos(45°)}}, \tag{6}$$

where $\alpha_F(t)$ and $\beta_F(t)$ are the weighting functions for the distance of the EDLs and PALs, respectively.

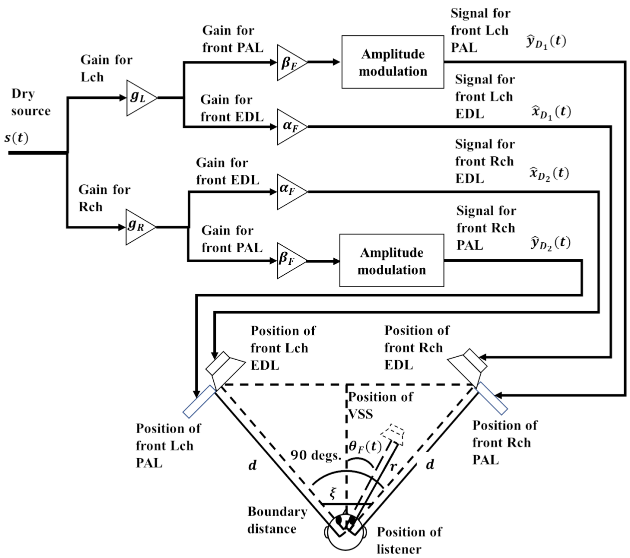



Fig. 2. Overview of front area output signal design.

The output signals for the FA are calculated as follows:

$$\hat{y}_{D_1}(t) = \beta_F(t)g_L(t)(1+s(t))c(t), \quad (7)$$

$$\hat{y}_{D_2}(t) = \beta_F(t)g_R(t)(1+s(t))c(t), \quad (8)$$

$$\hat{x}_{D_1}(t) = \alpha_F(t)g_L(t)s(t), \quad (9)$$

$$\hat{x}_{D_2}(t) = \alpha_F(t)g_R(t)s(t), \quad (10)$$

where $c(t)$ is the ultrasonic carrier wave, $s(t)$ is the input signal of the audible sound, $\hat{y}_{D_1}(t)$ is the output signal for the PAL in the FL direction, $\hat{y}_{D_2}(t)$ is the output signals for the PALs in the FR direction, $\hat{x}_{D_1}(t)$ is the output signal for the EDL in the FL direction, and $\hat{x}_{D_2}(t)$ is the output signal for the EDL in the FR direction. These outputs achieve DRR control for both the distance and the amplitude direction panning. To achieve back projection to the VSS coordinates of each area, the following output channels are selected for the designed signals:

- $x_{FL}(t)$: EDL in the FL direction.
- $x_{FR}(t)$: EDL in the FR direction.
- $x_{BL}(t)$: EDL in the BL direction.
- $x_{BR}(t)$: EDL in the BR direction.
- $y_{FL}(t)$: PAL in the FL direction.
- $y_{FR}(t)$: PAL in the FR direction.
- $y_{BL}(t)$: PAL in the BL direction.
- $y_{BR}(t)$: PAL in the BR direction.

For the FA of the out-of-head position, the output channels are selected as follows:

$$x_{FL}(t) = \hat{x}_{D_1}(t), \quad (11)$$

$$x_{FR}(t) = \hat{x}_{D_2}(t), \quad (12)$$

$$y_{FL}(t) = \hat{y}_{D_1}(t), \quad (13)$$

$$y_{FR}(t) = \hat{y}_{D_2}(t), \quad (14)$$

$$x_{BL}(t) = x_{BR}(t) = y_{BL}(t) = y_{BR}(t) = 0. \quad (15)$$

For the RA of the out-of-head position, the output channels are selected as follows:

$$x_{FR}(t) = \hat{x}_{D_2}(t), \quad (16)$$

$$x_{BR}(t) = \hat{x}_{D_1}(t), \quad (17)$$

$$y_{FR}(t) = \hat{y}_{D_2}(t), \quad (18)$$

$$y_{BR}(t) = \hat{y}_{D_1}(t), \quad (19)$$

$$x_{FL}(t) = x_{BL}(t) = y_{FL}(t) = y_{BL}(t) = 0. \quad (20)$$

For the BA of the out-of-head position, the output channels are selected as follows:

$$x_{BL}(t) = \hat{x}_{D_1}(t), \quad (21)$$

$$x_{BR}(t) = \hat{x}_{D_2}(t), \quad (22)$$

$$y_{BL}(t) = \hat{y}_{D_1}(t), \quad (23)$$

$$y_{BR}(t) = \hat{y}_{D_2}(t), \quad (24)$$

$$x_{FL}(t) = x_{FR}(t) = y_{FL}(t) = y_{FR}(t) = 0. \quad (25)$$

For the LA of the out-of-head position, the output channels are selected as follows:

$$x_{FL}(t) = \hat{x}_{D_1}(t), \quad (26)$$

$$x_{BL}(t) = \hat{x}_{D_2}(t), \quad (27)$$

$$y_{FL}(t) = \hat{y}_{D_1}(t), \quad (28)$$

$$y_{BL}(t) = \hat{y}_{D_2}(t), \quad (29)$$

$$x_{FR}(t) = x_{BR}(t) = y_{FR}(t) = y_{BR}(t) = 0. \quad (30)$$

### *C. Output signal design and output channel selection for the in-head position*

The projection of the VSS direction from each area to the FA is the same as in the case of out-of-head position shown in Eq. (1). The out-of-head condition is as follows:

$$0 \leq r(t) < \frac{\xi \cos 45^\circ}{\cos(\theta_F(t))}. \quad (31)$$

Figure 3 shows an overview of the output signal design for the in-head position. As shown in Fig. 3, the output signals for the in-head position are designed under the assumption that the VSS is located in the front area. For the in-head position, four PALs are utilized but no EDLs are used. The weighting functions for the VSS direction are calculated using Eqs. (3) and (4) in the same way as for the out-of-head position. For the in-head position, the weighting functions for the VSS distance are calculated as follows:

$$\gamma_F(t) = \sqrt{\frac{\xi \cos(45^\circ) + r(t)\cos(\theta_F(t))}{2\xi \cos(45^\circ)}}, \quad (32)$$

$$\gamma_B(t) = \sqrt{\frac{\xi \cos(45^\circ) - r(t)\cos(\theta_F(t))}{2\xi \cos(45^\circ)}}, \quad (33)$$

where $\gamma_F(t)$ and $\gamma_B(t)$ are the weighting functions for the PALs in the front and back directions, respectively. The output

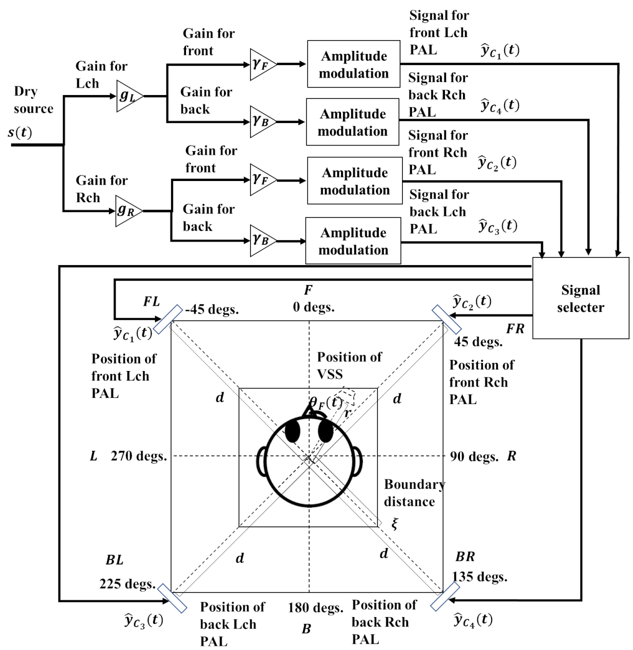



Fig. 3. Overview of the output signal design of the proposed method for the FA of the in-head position.

signals for the in-head position in the FA are calculated as follows:

$$\hat{y}_{C_1}(t) = \gamma_F(t)g_L(t)(1+s(t))c(t), \tag{34}$$
$$\hat{y}_{C_2}(t) = \gamma_F(t)g_R(t)(1+s(t))c(t), \tag{35}$$
$$\hat{y}_{C_3}(t) = \gamma_B(t)g_L(t)(1+s(t))c(t), \tag{36}$$
$$\hat{y}_{C_4}(t) = \gamma_B(t)g_R(t)(1+s(t))c(t), \tag{37}$$

where $\hat{y}_{C_1}(t)$ is the output signal for the PAL in the FL direction, $\hat{y}_{C_2}(t)$ is the output signal for the PAL in the FR direction, $\hat{y}_{C_3}(t)$ is the output signal for the PAL in the BL direction, and $\hat{y}_{C_4}(t)$ is the output signal for the PAL in the FR direction. These output signals achieve both ILD distance control and direction amplitude panning. To facilitate back projection to the VSS coordinates of each area, output channels are selected for the designed signals. Those output channels are as same as those used for the out-of-head position, but no EDLs are used.

$$x_{FL}(t) = x_{FR}(t) = x_{BL}(t) = x_{BR}(t) = 0. \tag{38}$$

For the FA of the in-head position, the output channels are selected as follows:

$$y_{FL}(t) = \hat{y}_{C_1}(t), \tag{39}$$
$$y_{FR}(t) = \hat{y}_{C_2}(t), \tag{40}$$
$$y_{BL}(t) = \hat{y}_{C_4}(t), \tag{41}$$
$$y_{BR}(t) = \hat{y}_{C_3}(t). \tag{42}$$

For the RA of the in-head position, the output channels are selected as follows:

$$y_{FL}(t) = \hat{y}_{C_4}(t), \tag{43}$$
$$y_{FR}(t) = \hat{y}_{C_1}(t), \tag{44}$$
$$y_{BL}(t) = \hat{y}_{C_3}(t), \tag{45}$$
$$y_{BR}(t) = \hat{y}_{C_2}(t). \tag{46}$$

For the BA of the in-head position, the output channels are selected as follows:

$$y_{FL}(t) = \hat{y}_{C_3}(t), \tag{47}$$
$$y_{FR}(t) = \hat{y}_{C_4}(t), \tag{48}$$
$$y_{BL}(t) = \hat{y}_{C_2}(t), \tag{49}$$
$$y_{BR}(t) = \hat{y}_{C_1}(t). \tag{50}$$

For the LA of the in-head position, the output channels are selected as follows:

$$y_{FL}(t) = \hat{y}_{C_2}(t), \tag{51}$$
$$y_{FR}(t) = \hat{y}_{C_3}(t), \tag{52}$$
$$y_{BL}(t) = \hat{y}_{C_1}(t), \tag{53}$$
$$y_{BR}(t) = \hat{y}_{C_4}(t). \tag{54}$$

TABLE I
EXPERIMENTAL EQUIPMENT.

| | |
|---|---|
| PAL | MITSUBISHI, MSP-50E |
| EDL | ELAC, BS302 |
| Power amplifier | YAMAHA, XM4180 |
| Dummy head microphone | NEUMANN, KU100 |
| Microphone | SENNHEISER, MKH8020 |
| Audio interface | RME, Fireface UFX II |

TABLE II
EXPERIMENTAL CONDITIONS.

| | |
|---|---|
| Boundary distance | $\xi$ = 0.25 m |
| Reverberation time | $T_{60}$ = 750 ms |
| Sampling frequency | 96 kHz |
| Quantization | 16 bits |

## III. Objective Evaluation Experiments

To confirm the effectiveness of the proposed method, we used the EDL to evaluate the ILD and DRR using room impulse responses measured via the TSP method [18] at real sound source (RSS) positions. The proposed method measures room impulse responses at the positions described above and recomposites the room impulse responses corresponding to the VSS positions. The ILD is evaluated under both in-head and out-of-head conditions, while the DRR is evaluated solely under the out-of-head condition.

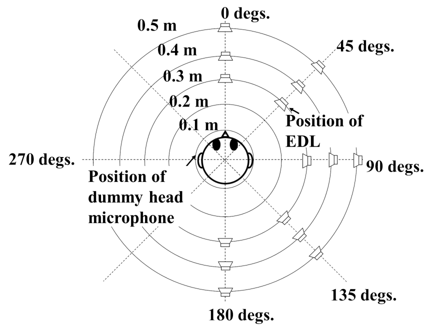
### A. Experimental conditions for objective evaluation with ILD

Tables I, II, and III show the experimental equipment, experimental conditions, and experimental conditions for the PAL, respectively. We conducted evaluations using the following two methods:

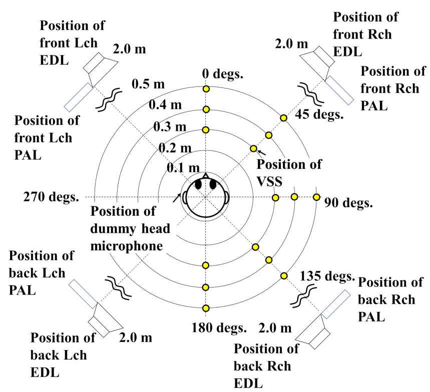
- RSS: RSS using EDL.

TABLE III
PAL EXPERIMENTAL CONDITIONS.

| | |
|---|---|
| Carrier frequency | 40 kHz |
| Modulation method | AM-DSB |




(a) RSS using EDL




(b) VSS with the proposed method

Fig. 4. Experimental arrangement for evaluating ILD under the out-of-head condition.

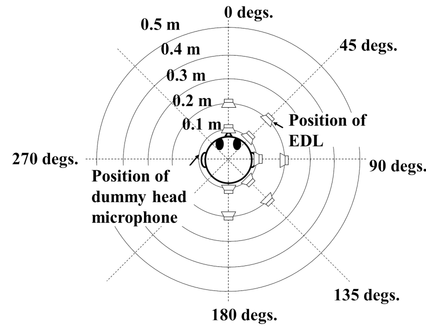
- VSS: VSS using the proposed method.

For ILD evaluations under the out-of-head condition, Figure 4(a) shows the experimental arrangement of the dummy head microphone at the listener position and the RSS positions using the EDL, while Figure 4(b) shows the experimental arrangement of the dummy head microphone at the listener position and the VSS positions using the proposed method.

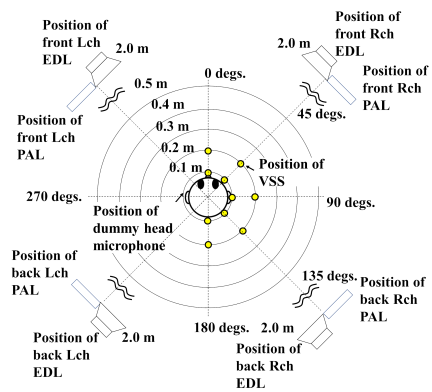
For ILD evaluations under the in-head condition, Figure 5(a) shows the experimental arrangement of the dummy head microphone at the listener position and the RSS positions using the EDL, while Figure 5(b) shows the experimental arrangement of the dummy head microphone at the listener position and the VSS positions using the proposed method. The ILD is calculated as follows:

$$\mathrm{ILD} = 10 \log_{10} \frac{\sum_{t=0}^{T_{\mathrm{D}}-1} p_L^2(t)}{\sum_{t=0}^{T'-1} p_R^2(t)}, \tag{55}$$

where $p_L(t)$ and $p_R(t)$ are room impulse responses measured at the left and right ear positions of the dummy head microphone, $T'$ is the length of the room impulse response. In the out-of-head condition, as the ILD of the RSS (Real) and VSS




(a) RSS using EDL




(b) VSS with the proposed method

Fig. 5. Experimental arrangements for evaluating the ILD under the in-head condition.

(Proposed) become closer, the reproduction performance improves. Additionally, the average absolute value ILD ($\overline{|\mathrm{ILD}|}$) is calculated as follows:

$$\overline{|\mathrm{ILD}|} = \frac{1}{M} \sum_{m=1}^{M} |\mathrm{ILD}_m| \,. \tag{56}$$

where $M (M = 10)$ is the total number of RSS or VSS positions and $\mathrm{ILD}_m$ is the ILD calculated from the room impulse response measured at the $m$-th position of the RSS or VSS. In the in-head condition, as the $\overline{|\mathrm{ILD}|}$ becomes closer to 0 dB, the in-head localization becomes higher.

### *B. Experimental results for objective evaluation with ILD*

Table IV shows the ILDs calculated from measured or recomposited room impulse responses at each RSS and VSS position. From Table IV, it can be seen that in the directions of 0, 45, and 180 degs., good performance was confirmed because the errors between RSS and VSS were about 2 dB. In contrast, compared with the other directions, performance was degraded in the directions of 90 and 135 degs. because the errors between the RSS and VSS exceeded 4 dB. A particularly large error is shown for the condition that includes the PAL in the BA. This suggests that the direct PAL wave is affected by the pinna of the listener's ears. Nevertheless, we confirmed that the proposed method achieves an overall good level of performance.

TABLE IV
ILDs CALCULATED FROM MEASURED OR RECOMPOSITED ROOM IMPULSE RESPONSES.

| Distance, Direction | RSS | VSS |
|---|---|---|
| 0.3 m, 0 degs. | 0.5 dB | 0.1 dB |
| 0.3 m, 45 degs. | -17.3 dB | -15.9 dB |
| 0.3 m, 90 degs. | -19.9 dB | -12.7 dB |
| 0.3 m, 135 degs. | -18.1 dB | -11.2 dB |
| 0.3 m, 180 degs. | 2.0 dB | 0.1 dB |
| 0.4 m, 0 degs. | 0.5 dB | 0.1 dB |
| 0.4 m, 45 degs. | -14.9 dB | -14.5 dB |
| 0.4 m, 90 degs. | -17.7 dB | -11.9 dB |
| 0.4 m, 135 degs. | -15.4 dB | -10.1 dB |
| 0.4 m, 180 degs. | -0.3 dB | -0.1 dB |
| 0.5 m, 0 degs. | 0.7 dB | 0.2 dB |
| 0.5 m, 45 degs. | -15.1 dB | -13.6 dB |
| 0.5 m, 90 degs. | -16.3 dB | -11.3 dB |
| 0.5 m, 135 degs. | -13.5 dB | -9.5 dB |
| 0.5 m, 180 degs. | 0.6 dB | -0.2 dB |

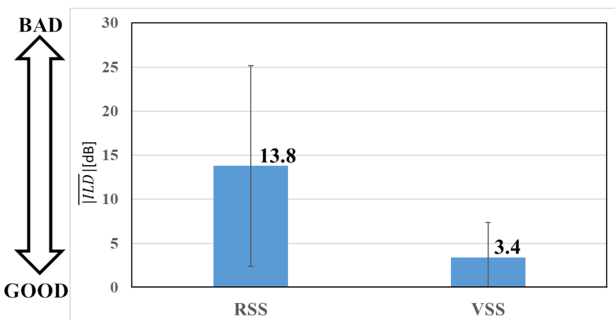


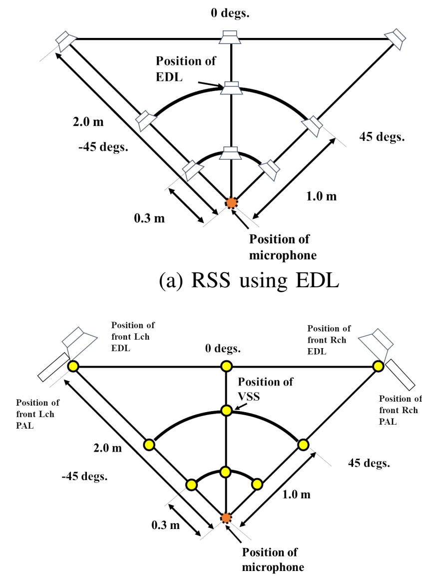

Fig. 6. Results for $\overline{|\mathrm{ILD}|}$ for in-head localization in an objective evaluation experiment.

Figure 6 shows the experimental results for $\overline{|\mathrm{ILD}|}$ for the in-head condition. From Fig. 6, we confirmed that the proposed method achieves a higher level of in-head localization than the real sound source because the $\overline{|\mathrm{ILD}|}$ of the VSS is closer to 0 dB than that of the RSS.

### C. Experimental conditions for objective evaluation with DRR

Tables I, II, and III show the experimental equipment, experimental conditions, and experimental conditions for the PAL, respectively. These conditions are the same as those used in the objective evaluation of the ILD. We also evaluated the DRR under the RSS and VSS conditions.

For DDR evaluations performed under the out-of-head condition, Figure 7(a) shows the experimental arrangement of the dummy head microphone at the listener position and the RSS positions used for the EDL, while Figure 7(b) shows the experimental arrangement of the dummy head microphone at the listener position and the VSS positions used with the proposed method. The DRR is calculated as follows:




(a) RSS using EDL

(b) VSS with the proposed method

Fig. 7. Experimental arrangement for evaluating the DRR under the out-of-head condition.

TABLE V
DIRECT-TO-REVERBERANT RATIOS CALCULATED FROM IMPULSE RESPONSES.

| Distance, Direction | RSS | VSS |
|---|---|---|
| 0.3 m, -45 degs. | 14.6 dB | 16.1 dB |
| 0.3 m, 0 degs. | 13.2 dB | 11.1 dB |
| 0.3 m, 45 degs. | 14.9 dB | 17.0 dB |
| 1.0 m, -45 degs. | 4.6 dB | 3.6 dB |
| 1.0 m, 0 degs. | 4.7 dB | 5.4 dB |
| 1.0 m, 45 degs. | 5.4 dB | 3.6 dB |
| 2.0 m, -45 degs. | 1.9 dB | 1.9 dB |
| 1.4 m, 0 degs. | 2.8 dB | 4.2 dB |
| 2.0 m, 45 degs. | 1.8 dB | 1.8 dB |

$$\mathrm{DRR} = 10\log_{10}\frac{\sum_{n=0}^{T_\mathrm{d}-1} p_e^2(n)}{\sum_{n=T_\mathrm{d}}^{T'-1} p_e^2(n)}, \tag{57}$$

where $p_e(n)$ is the measured or recomposited impulse response, $T_\mathrm{d}(= 7$ ms) is the length of the direct sound, and $T'$ is the length of $p_e(n)$. A higher DRR value indicates that the VSS is perceived as being close to the listener.

### D. Experimental results obtained by objective evaluation on DRR

Table V shows the DRRs calculated from measured or recomposited room impulse responses at each RSS and VSS position. From Table V, we confirmed that the proposed method could approximate the real sound source DRR because the DRR error is approximately 3 dB.

TABLE VI
MOVING SENSATION EVALUATION INDEX WITHOUT LISTENER PENETRATION.

| Score | Impairment |
|---|---|
| 5 | Smooth movement |
| 4 | Movement with some discomfort |
| 3 | Uncomfortable movement |
| 2 | Slight movement |
| 1 | No movement |

TABLE VII
MOVEMENT SENSATION EVALUATION INDEX WITH LISTENER PENETRATION.

| Score | Impairment |
|---|---|
| 5 | Clear in-head penetration |
| 4 | A slight feeling that something is wrong, but with in-head penetration |
| 3 | A clear feeing that something is wrong, but with in-head penetration |
| 2 | Slight in-head penetration |
| 1 | No in-head penetration |

## IV. SUBJECTIVE EVALUATION EXPERIMENT

Next, using the mean opinion score (MOS) [19], we confirmed the effectiveness of the proposed method by evaluating the movement sensations for the VSS with and without head penetration. This evaluation was performed using both the proposed method and by moving the EDL (RSS).

### A. Experimental conditions for subjective evaluation

Tables I, II, and III show the experimental equipment, experimental conditions, and experimental conditions for the PAL, which were the same as those used in the objective evaluation with the ILD. We also conducted movement sensation evaluations under RSS and VSS conditions for five subjects. The movement sensation was evaluated using five-level MOS values under two different conditions: one in which the moving sound source did not penetrate the listener and the other with listener penetration. Table VI, shows the moving sensation evaluation index without listener penetration. Table VII, shows the evaluation index of the moving sensation with listener penetration. Table VIII, shows the experimental conditions used for the subjective evaluations. Two methods and two dry sound sources were evaluated as follows:

- RSS (White): EDL (RSS) is moved manually (white noise).
- RSS (Canon): EDL (RSS) is moved manually (Pachelbel's Canon).
- VSS(White): VSS is reproduced with the proposed method (white noise).
- VSS (Canon): VSS is reproduced with the proposed method (Pachelbel's Canon).

TABLE VIII
EXPERIMENTAL CONDITIONS FOR SUBJECTIVE EVALUATIONS.

| | |
|---|---|
| Dry sound sources | White noise (0 to 8 kHz, 8 s), Pachelbel's Canon (8 s) with electro bass |
| Movement speed (Path and direction) | 0.5 m/s (FR to BL), 0.35 m/s (F to B) |

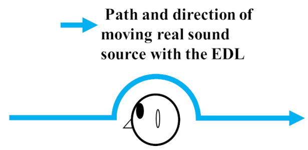
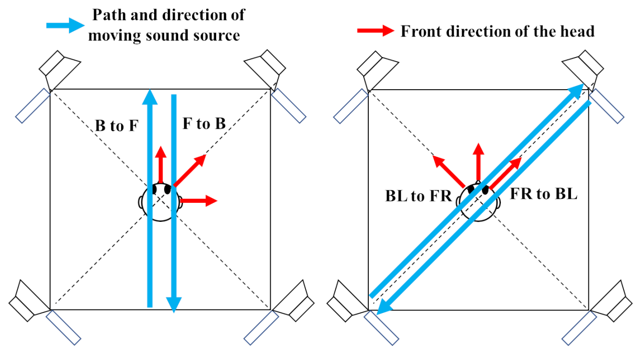



Fig. 8. Path and direction from F to B for manual EDL movement (side view).




(a) Front and back directions (b) Diagonal directions
Fig. 9. Path and direction of RSS or VSS movement and the front direction of the head.

Figure 8 shows the side view of the path and the direction of real sound source movement with the EDL for F to B. As shown in Fig. 8, the EDL is moved in a manner that avoids a collision with the head. Figure 9 shows the paths and directions of RSS or VSS movement and the front direction of the head Figure 10 shows the weighting function with the proposed method when the path and direction are from FR to BL.

### B. Experimental results with subjective evaluation

Figure 11 shows the MOS values for the movement sensation without listener penetration. As shown in Fig. 11, we confirmed that the RSS and VSS provide good movement sensations for the path without listener penetration because both MOS values are above 4.0. This result is particularly noteworthy because it shows that the proposed method can smoothly move the VSS even at the boundary distance.

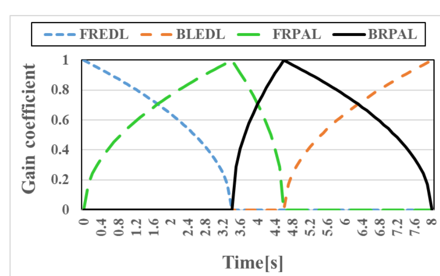



Fig. 10. Weighting function with the proposed method (FR to BL).

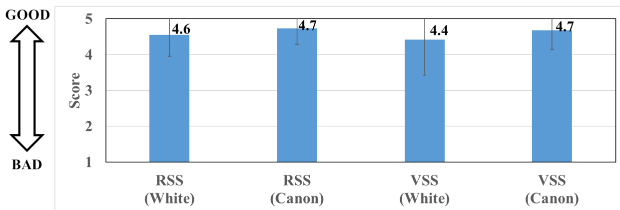



Fig. 11. MOS values for movement sensation without listener penetration.

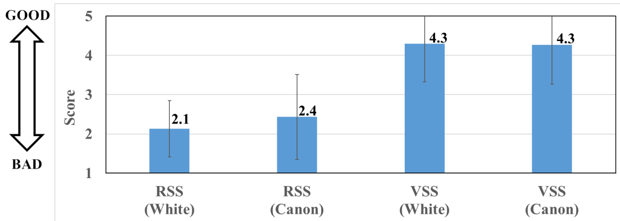



Fig. 12. MOS values for the movement sensation with listener penetration.

Figure 12 shows the MOS values for the movement sensation condition with listener penetration. As shown in Fig. 12, the MOS value for the RSS is about 2.0, and that for the VSS is about 4.0. These results confirm that the VSS is superior to the RSS for the movement sensation condition with listener penetration. We can also see that the MOS value for the RSS is reduced because the RSS moves so that the EDL avoids a collision with the head, as shown in Fig. 8. On the other hand, the VSS MOS value is good because the proposed method can smoothly move the VSS from an out-of-head to an in-head position and from an in-head to an out-of-head position.

## V. Conclusions

In this paper, we proposed a new virtual sound source rendering system based on distance control that can be used to penetrate listeners using surround PALs and EDLs. We also reported on objective and subjective evaluations that were performed and their results, which confirmed the effectiveness of the proposed method. It was particularly noteworthy that, in terms of listener penetration, the movement sensation produced by the proposed method was superior to that achieved when actually moving the sound source. In the future, we intend to extend the proposed method to a three-dimensional surround sound system by including vertical direction elements.

## Acknowledgment

The present study was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP21H03488 and JP19K12056, and by cooperative research in partnership with EQUOS RESEARCH Co., Ltd.

## References

[1] A.J. Berkout, D. de Vries and P. Vogel, “Acoustic control by wave field synthesis,” J. Acoust. Soc. Am., vol. 93, no. 5, pp. 2764-2778, 1993.

[2] J. Bauck, D. H. Cooper, “Generalized transaural stereo and applications,” J. Acoust. Soc. Am., vol. 44, no. 9, pp. 683-705, 1996.

[3] G. Zhang, Q. Huang, K. Liu, “Three-dimensional sound field reproduction based on multi-resolution sparse representation,” ICSPCC 2015, pp. 1-5, 2015.

[4] M. Nakayama, K. Nakahashi, Y. Wakabayashi and T. Nishiura, “Surround Sensation Index Based on Differential S-IACF for Listener Envelopment with Multiple Sound Sources,” Journal of Communication and Computer, vol. 14, pp. 122-128, 2017.

[5] L.S. Zhou, C.C. Bao and M.S. Jia and B. Bu, “Range extrapolation of Head-Related Transfer Function using improved Higher Order Ambisonics,” APSIPA ASC 2014, pp.1-4, 2014.

[6] H. Kurabayashi, M. Otani, K. Itoh, M. Hashimoto, M. Kayama, “Development of dynamic transaural reproduction system using non-contact head tracking,” GCCE 2013, pp.12-16, 2013.

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

[8] Rec. ITU-R BS.775-2, “Multi-channel stereophonic sound system with or without accompanying pucture,” ITU, 2006.

[9] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama and A. Ando, “A 22.2 multichannel sound system for ultrahigh-definition TV,” SMPTE Motion Imaging Journal, pp.40-49, 2008.

[10] Dolby Laboratories, Inc., Dolby Atmos: http://www.dolby.com/us/en/technologies/home/dolby-atmos.html

[11] V. Plukki, “Virtual sound source positioning Using vector base amplitude panning,” J. Audio Eng. Soc., vol. 45, no. 6, pp. 456-466, 1997.

[12] J.J. Lopez, P. Gutierrez, M. Cobos, E. Aguilera, “Sound distance perception comparison between Wave Field Synthesis and Vector Base Amplitude Panning,” ISCCSP 2014, pp. 165-168, 2014.

[13] J. Daniel, R. Nicol and S. Moreau, “Further investigations of High Order Ambisonics and wavefield synthesis for holophonic sound imaging,” Proc. Audio Eng. Soc. 114th Conv., 5788, 2003.

[14] Y. Ogami, M. Nakayama and T. Nishiura, “Virtual Sound Source Construction Based on Radiation Direction Control Using Multiple Parametric Array Loudspeakers,” J. Acoust. Soc. Am., vol. 146, no. 2, pp. 1314-1325, 2019.

[15] Y.C. Lu and M. Cooke, “Binaural estimation of sound source distance via the direct-to-deverberant energy ratio for static and moving sources,” IEEE Trans. ASLP, vol. 18, no. 7, pp. 1793-1805, 2010.

[16] M. Nakayama and T. Nishiura: “Distance Control of Virtual Sound Source Using Parametric and Dynamic Loudspeakers, ” APSIPA ASC 2018, pp.1262-1267, 2018.

[17] M. Chen, Leon. Xu, Y. Cao, Limei. Xu, X. Wang, X. Li, and J. Ma, “Research on an improved amplitude modulation method of audio directional loudspeaker,” ICALIP 2008, pp. 5-9, 2008.

[18] A.J. Berkhout, D. de Vries and M.M. Boone, “A new method to acquir impulse responses in concert halls,” J. Acoust. Soc. Am., vol. 68, no. 1, pp. 179-183, 1980.

[19] D. Ikefuji, H. Tsujii, S. Masunaga, M. Nakayama, T. Nishiura and Y. Yamashita, “Reverberation Steering and Listening Area Expansion on 3-D Sound Field Reproduction with Parametric Array Loudspeaker, ” APSIPA ASC 2014, pp.1-5, 2014.