Distance Control of Virtual Sound Source Using Parametric and Dynamic Loudspeakers

Masato Nakayama* and Takanobu Nishiura†
* Osaka Sangyo University, Osaka, Japan
E-mail: nakayama@ise.osaka-sandai.ac.jp Tel: +81-72-875-3001
† Ritsumeikan University, Shiga, Japan
E-mail: nishiura@is.ritsumei.ac.jp Tel/Fax: +81-77-561-5075

***Abstract*—We propose a new range control of sound image using parametric and dynamic loudspeakers. In the proposed method, a parametric loudspeaker generates a closed sound source, and a dynamic loudspeaker generates a distant sound source. More specifically, the proposed method uses panning of sound sources between the listener's head and surround loudspeakers.**

I. Introduction

Three-dimensional sound field reproductions [1], [2], [3], [4] have been proposed in order to realize a sound field with high presence for theaters, virtual live stages, and telecommunications.

As one of these reproductions, a binaural system [5] has been proposed with dummy-head microphones and headphones. In the binaural system, the head-related transfer function (HRTF) is used, which is the transfer function between a sound source and the listener's ears. Head-related transfer functions are measured under various conditions in advance, and the binaural sound is created by convolving a dry source and the HRTFs. The binaural sound is presented to the listener using headphones. Although this system is effective, it requires the measurement of several HRTFs because personal HRTFs are used, the measurement of which is very difficult.

On the other hand, a transaural system [6], which uses numerous remote loudspeakers, has been proposed. A transaural system can reproduce a sound field by designing the multi-channel inverse filter between remote loudspeakers and the ears. However, this requires very expensive computation as well as a large system with numerous loudspeakers.

Therefore, surround sound has been used for practical sound field reproduction. Various surround systems, such as stereo systems, Dolby 5.1ch surround systems [7], 22.2ch surround systems [8], which are channel-based systems, have been proposed. Channel-based systems can represent sound sources at the positions of loudspeakers and so require numerous loudspeakers. On the other hand, object-based systems, such as Dolby Atmos [9], have been proposed. This system has multiple audio objects, and audio objects are positioned as virtual sound sources based on amplitude panning [10]. These surround systems can represent the localization of virtual sound sources in terms of angle. However, localization in terms of distance is also important to represent the virtual sound source. Conventional localization for distance has been proposed as vector-based amplitude panning, which controls the sound pressure level [11]. The performance of the conventional method is insufficient because the cue of distance sensation includes not only the sound pressure level but also the Doppler effect [12] and the direct-to-reverberant ratio (DRR) [13].

In the present paper, we therefore propose distance control of a virtual sound source based on DRR reproduction using parametric and dynamic loudspeakers. A parametric loudspeaker can achieve super directivity using a rectilinear ultrasonic wave. Its emitted sound has a high DRR, i.e., a large-amplitude direct wave and low-amplitude reverberations. Thus, the parametric loudspeaker causes the virtual sound source to be perceived closer to the listener. On the other hand, the remote dynamic loudspeaker causes the virtual sound source to be perceived far from the listener because its emitted sound has a low DRR. The proposed method uses the difference in the DRR between parametric and dynamic loudspeakers. In the proposed method, amplitude panning is applied to input signals of parametric and dynamic loudspeakers so as to obtain a suitable DRR for the distance of the virtual sound source. As a result, the proposed method can control the distance sensation of the virtual sound source in the space between a remote dynamic loudspeaker and the listener. Finally, we confirmed the effectiveness of the proposed method through evaluation experiments.

II. Conventional virtual sound source positioning based on amplitude panning

Conventional surround systems, which position the virtual sound source based on amplitude panning for multiple loudspeakers [10], have been proposed. Figure 1 shows the arrangement of real sound sources and a virtual sound source in the conventional method. As shown in Fig. 1, when real sound sources (loudspeakers) $A, B, andC$ are placed in a triangle arrangement, virtual sound source P is represented by interpolation coefficients for each loudspeaker, as follows:

$$\alpha_A = \sqrt{\frac{PA'}{AA'}},\ \alpha_B = \sqrt{\frac{PB'}{BB'}},\ \alpha_C = \sqrt{\frac{PC'}{CC'}}, \qquad (1)$$

where α_A, α_B, and α_C are the gains of input signals to loudspeakers $A, B, andC$, respectively. The conventional method controls the acoustic center using multiple loudspeakers.

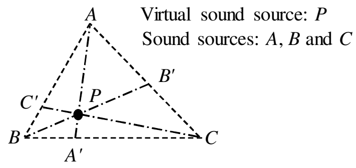

Fig. 1. Arrangement of real sound sources and a virtual sound source in the conventional method.

III. Conventional pin-spot audio with a parametric loudspeaker

A parametric loudspeaker achieves super directivity through rectilinearity of the ultrasonic sound, and uses an amplitude-modulated (AM) wave generated by modulating the amplitude of an ultrasonic sound (carrier wave) with an audible sound [14]. An amplitude-modulated wave $x_{\mathrm{AM}}(t)$ is formulated, as follows:

$$\begin{aligned} x_{\mathrm{AM}}(t) &= A_{\mathrm{C}}\cos(2\pi f_{\mathrm{C}} t) \\ &\quad + \frac{mA_{\mathrm{S}}}{2}\cos(2\pi(f_{\mathrm{C}}+f_{\mathrm{S}})t) \\ &\quad + \frac{mA_{\mathrm{S}}}{2}\cos(2\pi(f_{\mathrm{C}}-f_{\mathrm{S}})t), \end{aligned} \tag{2}$$

where t is time, f_{C} is the frequency of the carrier wave, A_{C} is the maximum amplitude of the carrier wave, f_{S} is the frequency of the audible sound, A_{S} is the maximum amplitude of the audible sound, and m $(m \leq 1)$ is the modulation depth. In the case of $m > 1$, overmodulation occurs, and the AM wave is distorted.

From, Eq. (2), the AM wave consists of a carrier frequency (f_{C}), a sum frequency ($f_{\mathrm{C}} + f_{\mathrm{S}}$), and a difference frequency ($f_{\mathrm{C}} - f_{\mathrm{S}}$). The sum and difference frequencies are referred to as frequencies of the sideband wave. The audible sound is demodulated by nonlinear interaction in the air because the parametric loudspeaker emits an intense AM wave [14].

IV. Distance Control of a Virtual Sound Source Using Parametric and Dynamic Loudspeakers

In the present paper, we propose distance control of a virtual sound source using parametric and dynamic loudspeakers. Figure 2 shows an overview of the proposed method. In the proposed paper, we describe the process flow in the case of one parametric loudspeaker and two dynamic loudspeakers (front LR loudspeakers). Figure 3 shows the arrangement of real sound sources and a virtual sound source in the proposed method. As shown in Fig. 3, the proposed method controls the virtual sound source (VSS) in the target area between loudspeakers and the listener. From Fig. 3, simple amplitude panning is derived, as follows.:

$$\alpha_L = \sqrt{\frac{PL'}{LL'}},\ \alpha_R = \sqrt{\frac{PR'}{RR'}},\ \alpha_O = \sqrt{\frac{PO'}{OO'}}, \tag{3}$$

where α_L, α_R, and α_C are the gains of input signals to dynamic loudspeakers L and R and parametric loudspeaker

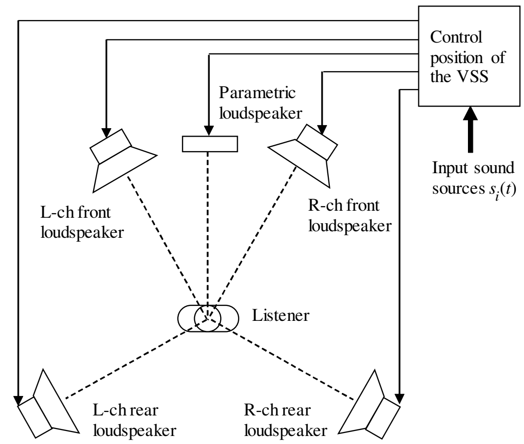

Fig. 2. Overview of the proposed method.

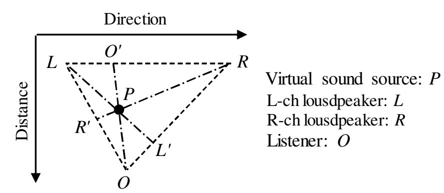

Fig. 3. Arrangement of real sound sources and a virtual sound source in the proposed method.

O, respectively. In the present paper, the acoustic center of the parametric loudspeaker is placed at Position O, although the parametric loudspeaker is placed in front of the listener. This is because the listeners perceive the acoustic center near themselves due to the high DRR. Here, the correction parameter should be calculated in order to approximate the real DRR. Thus, the interpolation coefficients of the proposed method for the i-th sound source are as follows:

$$\begin{aligned} \alpha_{L_i} &= \sqrt{\frac{P_i L_i'}{L_i L_i'}},\ \alpha_{R_i} = \sqrt{\frac{P_i R_i'}{R_i R_i'}}, \\ \alpha_{O_i} &= \gamma\sqrt{\frac{P_i O_i'}{O_i O_i'}}, \end{aligned} \tag{4}$$

where γ is the correction parameter of the DRR and is calculated from a preliminary experiment of the real DRR.

Finally, the output signals are obtained by the proposed method, as follows:

$$\begin{aligned} x_L(t) &= \sum_{i=1}^{I} \alpha_{L_i} s_i(t), \\ x_R(t) &= \sum_{i=1}^{I} \alpha_{R_i} s_i(t), \\ x_O(t) &= \sum_{i=1}^{I} \alpha_{O_i} \beta(1 + s_i(t))c(t), \end{aligned} \tag{5}$$

TABLE I
EXPERIMENTAL EQUIPMENT.

Parametric loudspeaker	MITSUBISHI, MSP-50E
Dynamic loudspeaker	FOSTEX, FE83En
Power amplifier	VICTOR, PS-A2002
Microphone	SONY, ECM-88B
A/D, D/A converter	RME, Fireface UFX

TABLE II
EXPERIMENTAL CONDITIONS.

Reverberation time	T_{60} = 650 ms
Sampling frequency	192 kHz
Quantization	16 bits

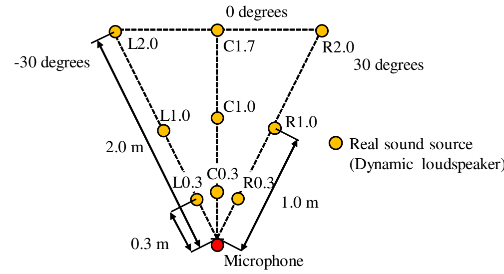

Fig. 5. Experimental arrangement for the real sound source (loudspeaker) in objective evaluation.

where $s_i(t)$ is the input signal with the i-th sound source, β is the gain normalization parameter of the parametric and dynamic loudspeakers, c(t) is the carrier wave of ultrasonic sound, $x_L(t)$ is the output signal of the L-ch loudspeaker, $x_R(t)$ is the output signal of the R-ch loudspeaker, and $x_O(t)$ is the output signal of the parametric loudspeaker. The parameter β should be measured in advance.

Even in case of a greater number of dynamic loudspeakers, the acoustic center can be calculated in the same manner. The proposed method can freely control the distance of the VSS using simple gain control for parametric and dynamic loudspeakers.

V. OBJECTIVE EVALUATION EXPERIMENT

A. Experimental conditions with objective evaluation

In order to confirm the effectiveness of the proposed method, we measure the impulse response using the proposed method and the real sound source. From the impulse response, the DRRs are calculated using the proposed method and the real sound source. The DRR value was calculated as follows:

$$\mathrm{DRR} = 10\log_{10}\frac{\sum_{n=0}^{T_\mathrm{d}-1} o_e^2(n)}{\sum_{n=T_\mathrm{d}}^{T'-1} o_e^2(n)}, \tag{6}$$

where $o_e(n)$ is the evaluated impulse response, T_d(= 7 ms) is the length of direct sound, and T' is the length of $o_e(n)$. A higher DRR value indicates that the virtual sound source was perceived close to the listener.

Tables I, II, and III show the experimental equipment, the experimental conditions, and the experimental conditions for the parametric loudspeaker, respectively.

We evaluate two conditions as follows:

- Real: Results of the real sound source (loudspeaker).
- Proposed: Results of the proposed method.

Figure 4 shows the experimental arrangement for the proposed method, and Figure 5 shows the experimental arrangement for a real sound source (loudspeaker).

Impulse responses are measured with the TSP method [15]. A sound source is a time-stretched pulse (signal length: 2^{19}sample, bandwidth: 0 to 8 kHz). In this experiment, γ is 1, and β is adjusted from the actually measured sound pressure level.

TABLE III
EXPERIMENTAL CONDITIONS FOR THE PARAMETRIC LOUDSPEAKER.

Carrier frequency	40 kHz
Modulation method	AM-DSB

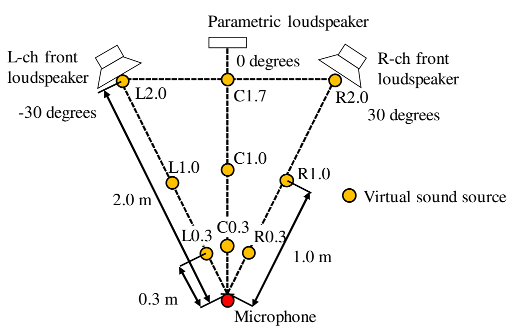

Fig. 4. Experimental arrangement for the proposed method in objective evaluation.

B. Experimental results obtained by objective evaluation

Figures 6, 7, and 8 show the impulse responses obtained using the real sound source at Positions C0.3, C1.0, and C1.7, respectively. Figures 9, 10, and 11 show the impulse responses obtained using the proposed method at Positions C0.3, C1.0, and C1.7, respectively. From Figs. 6 through 11, we confirmed that the impulse responses obtained using the proposed method tend to be similar those obtained using the real sound source.

Table IV shows the DRRs calculated from impulse responses for the real sound source and the proposed method. From Table IV, we confirmed that the proposed method can approximate the DRR of the real sound source because the error of the DRR is approximately 3 dB.

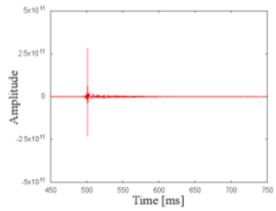

Fig. 6. Impulse response with the real sound source at Position C0.3.

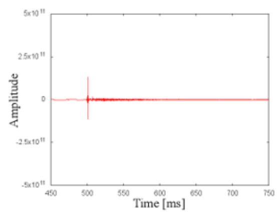

Fig. 7. Impulse response with the real sound source at Position C1.0.

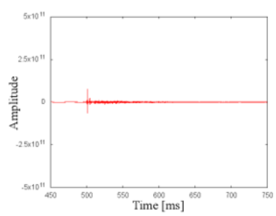

Fig. 8. Impulse response with the real sound source at Position C1.7.

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

	Real	Proposed
L0.3	15.5 dB	13.4 dB
C0.3	15.7 dB	13.4 dB
R0.3	17.1 dB	13.1 dB
L1.0	7.2 dB	10.6 dB
C1.0	6.3 dB	9.8 dB
R1.0	6.9 dB	9.6 dB
L2.0	1.2 dB	2.4 dB
C1.7	0.9 dB	1.1 dB
R2.0	1.0 dB	1.5 dB

VI. SUBJECTIVE EVALUATION EXPERIMENT

A. Experimental conditions with subjective evaluation

In order to confirm the effectiveness of the proposed method, we evaluate the localization performance for angle and distance using the proposed method and a real sound source. In subjective evaluation, the experimental equipment and conditions are the same as shown in Tables I through III. The sound source is white Gaussian noise (bandwidth: 0 to 8 kHz). Sound sources are randomly presented to seven subjects at two times per position. Subjects answer the direction and angle of the virtual sound source. The step size of angle is 30 degrees, and steps of the distance are 0.3, 1.0, and 2.0 m.

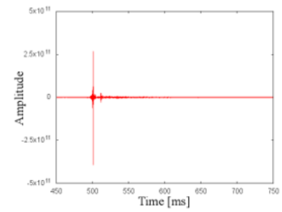

Fig. 9. Impulse response with the proposed method at Position C0.3.

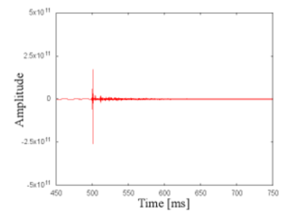

Fig. 10. Impulse response with the proposed method at Position C1.0.

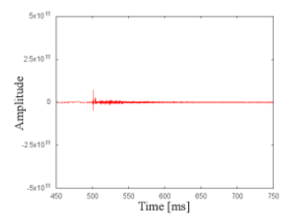

Fig. 11. Impulse response with the proposed method at Position C1.7.

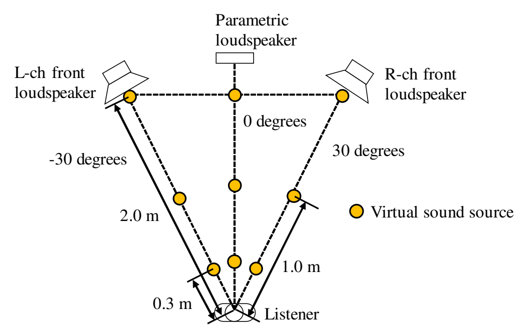

Fig. 12. Experimental arrangement for the proposed method in subjective evaluation.

B. Experimental results with subjective evaluation

Figure 14 shows the frequency distribution of answers with the angle of the real sound source, and Fig. 15 shows the frequency distribution of answers with the distance of the real sound source. Figure 16 shows the frequency distribution of answers with the angle of the proposed method, and Fig. 17 shows the frequency distribution of answers with the distance of the proposed method. Table V shows the correct answer rates for angle and distance.

In Figs. 14 through 17, the bubble on the dotted line

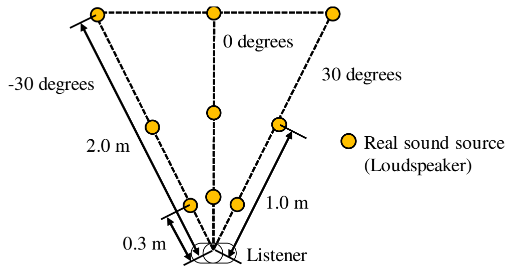

Fig. 13. Experimental arrangement for the real sound source (loudspeaker) in subjective evaluation.

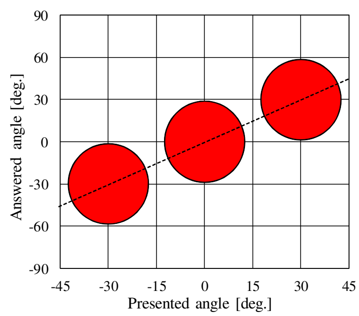

Fig. 14. Frequency distribution of answers with angle for the real sound source.

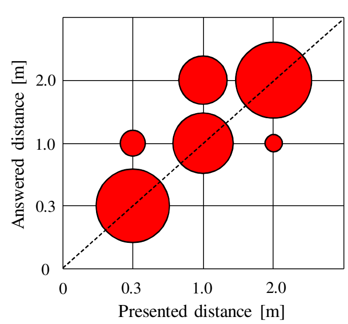

Fig. 15. Frequency distribution of answers with distance for the real sound source.

indicates the correct answer. Larger bubbles indicate higher frequencies of answers. As a result, the performance of angle estimation of the proposed method decreased compared with the real sound source. On the other hand, the performance of distance estimation of the proposed method is approximately equal to that of the real sound source. Table V confirmed the effectiveness of the proposed method.

VII. Conclusions

In the present paper, we proposed a new distance control method of a virtual sound source based on DRR reproduction using parametric and dynamic loudspeakers. As a result of objective and subjective evaluation experiments, the effectiveness of the proposed method was confirmed. In the future, we intend to evaluate the proposed method with rear loudspeakers. In addition, we intend to investigate the suitable parameter of the proposed method for various environments with different reverberation times.

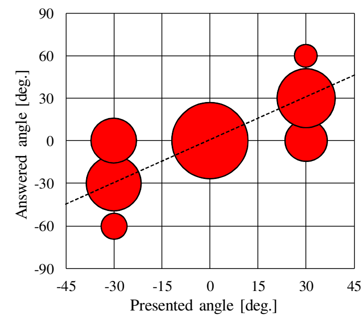

Fig. 16. Frequency distribution of answers with angle for the proposed method.

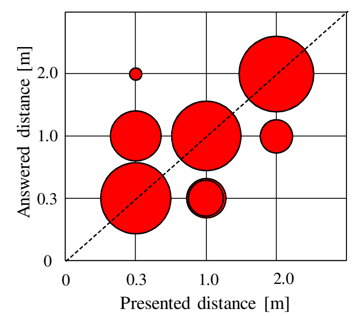

Fig. 17. Frequency distribution of answers with distance for the proposed method.

TABLE V
Correct answer rate for angle and distance.

	Real	Proposed
Angle	100%	71%
Distance	81%	76%

Acknowledgment

The present study was supported in part by the Center of Innovation Program of the Japan Science and Technology Agency, JST and JSPS KAKENHI Grant Number JP18K11365.

References

[1] A.J. Berkout, D. de Vries and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764-2778, 1993.
[2] J. Bauck, D. H. Cooper, "Generalized transaural stereo and applications," *The Journal of the Acoustical Society of America*, 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] Hiroaki Kurabayashi; Makoto Otani; Kazunori Itoh; Masami Hashimoto; Mizue Kayama "Development of dynamic transaural reproduction system using non-contact head tracking," GCCE 2013, pp.12-16, 2013.

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

[8] 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.

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

[10] V. Plukki, "Virtual sound source positioning Using vector base amplitude panning," The Journal of the Audio Engineering Society, vol. 45, no. 6, pp. 456-466, 1997.

[11] 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.

[12] L. Yang, L. Yang and K.C. Ho, "Moving target localization in multistatic sonar using time delays, Doppler shifts and arrival angles," ICASSP 2017, pp. 3399-3403, 2017.

[13] 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. on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1793-1805, 2010.

[14] 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.

[15] A.J. Berkhout, D. de Vries and M.M. Boone, "A new method to acquir impulse responses in concert halls," *The Journal of the Acoustical Society of America*, vol. 68, no. 1, pp. 179-183, 1980.