High-Presence Sharp Sound Image Based on Sound Blending Using Parametric and Dynamic Loudspeakers

Naoto Shimada¹, Kenta Iwai,² Masato Nakayama³,⁴ and Takanobu Nishiura²

¹ Graduate School of Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan
E-mail: is0310hp@ed.ritsumei.ac.jp
² College of Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan
E-mail: {iwai18sp@fc, nishiura@is}.ritsumei.ac.jp
³ Research Organization of Science and Technology, Ritsumeikan University, Shiga 525-8577, Japan
⁴ College of Design Technology, Osaka Sangyo University, Osaka 574-8530, Japan
E-mail: nakayama@ise.osaka-sanyo.ac.jp

Abstract

Previously, a virtual sound image that is produced with multiple electro-dynamic loudspeakers (EDLs) was diffused due to the reverberation nature of the room. A parametric array loudspeaker (PAL) can produce a sharper sound image than an EDL. However, it is difficult to produce low-frequency sound with a PAL, and the sound quality of the virtual sound image also deteriorates. Although the sound quality of the virtual sound image can be improved with a subwoofer, the sound image moves to the direction of the subwoofer. In this paper, we propose a high-presence sharp sound image based on sound blending using PALs and EDLs.

1. Introduction

Recently, 3-D surround sound systems have been proposed to realize a sound field with high presence with multiple electro-dynamic loudspeakers (EDLs), such as a channel-based system [1], transaural system [2], and object-based systems [4–6]. In object-based systems, multiple audio objects are produced as virtual sound images on the basis of amplitude panning. However, the produced sound image is diffused due to the reverberation nature of the room when producing a sharp sound image such as a point-like source. In this paper, we focused on the parametric array loudspeaker (PAL) [7–11]. The PAL can achieve sharper directivity due to straightness of the ultrasonic wave, and it has lower reverberation than the EDL. Therefore, the PAL can produce a sharp sound image [7, 8]. However, it is difficult to produce low-frequency sound lower than 1 kHz with a PAL and the sound quality of the virtual sound image deteriorates. Although the sound quality of the virtual sound image can be improved with a subwoofer, the sound image moves to its direction.

In this paper, we propose a high-presence sharp sound image based on sound blending using PALs and EDLs. In the proposed system, sharp sound images are produced with multiple PALs in addition to the conventional multichannel system. Moreover, by distributing the gain of low-frequency sound to the surrounding EDLs, the sharp sound images are produced at the desired position with high sound quality.

2. High-Presence Sharp Sound Image Based on Sound Blending Using PALs and EDLs

The proposed system consists of N PALs and M EDLs. Figure 1 shows an overview of the proposed system using 4 PALs and 22 EDLs. In the proposed system, PALs are used for production of sharp sound images, and EDLs are used for sound quality compensation. The PAL utilizes an amplitude-modulated wave [9]. Step 1 in Fig. 1 shows the production of sound images and their localization control by amplitude panning [5, 7, 8]. The input signals for PALs are expressed as

\[ u_n(t) = \alpha_n \{1 + \beta s(t)\} c(t) \]  

where \( u_n(t) \) denotes the input signal for the \( n \)-th \( (1 \leq n \leq N) \) PAL, \( t \) is the time index, \( \alpha_n \) is the gain of input signal for the \( n \)-th PAL, \( \beta \) (\( 0 < \beta \leq 1 \)) is the amplitude modulation factor, \( s(t) \) is the input signal without any processing, and \( c(t) \) is the carrier wave. Step 2 shows the process of designing low-frequency sound \( s_{LPF}(t) \) to improve the sound quality of the sound image by convolving the low-pass filter with \( s(t) \).
Figure 2: Experimental arrangement for activated loudspeaker with each condition

\[ s_{LPF}(t) = s(t) * h_{LPF,1C}(t) \]  

(2)

where * denotes the convolution operator, \( h_{LPF,1C}(t) \) is the low-pass filter, and \( \Omega_C \) is the crossover frequency between PALs and EDLs that is obtained by taking the frequency into account, which is difficult to produce with PALs. Step 3 shows the process of distributing the gain of \( s_{LPF}(t) \) to \( M \) EDLs and emitting from them. Then, the sound images are produced at the desired position with high sound quality. The input signals for EDLs are expressed as

\[ s_m(t) = \gamma \frac{s_{LPF}(t)}{M} \]  

(3)

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

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

(4)

where \( i \in \{ L, R \} \) denotes the index of the left and right ear, \( y_i(t) \) is the observed signal at each ear, \( f_{i,n}(t) \) is the impulse response between the \( n \)-th PAL and each ear, \( D[\cdot, \cdot] \) is the demodulation process of the amplitude-modulated sound emitted from the PAL, and \( g_{i,m}(t) \) is the impulse response between the \( m \)-th EDL and each ear. Therefore, the proposed system can produce sharp sound images at the desired position with high sound quality.

3. Evaluation Experiments

To evaluate the effectiveness of the proposed system, we conducted the evaluation experiments on sound quality, sharpness of sound image, and performance of sound localization. In the experiment on sound quality, we evaluated the improvement of the sound quality with the proposed system by log spectral distortion (LSD) [12]. In the experiment on sharpness of sound image, we calculated the cross-power spectrum phase (CSP) coefficient [13] and evaluated the sharpness of sound image produced with the proposed system by using the difference between the maximum and the average values of the CSP coefficient (\( D_{CSP} \)). In the experiment on performance of sound localization, we evaluated the performance of sound localization by using the interaural level difference (ILD) [14] and the inter-aural cross-correlation (IACC) [15]. Figure 2 shows the experimental arrangement of activated loudspeakers with each system. In this experiment, when producing a sound image with the loudspeaker placed in front of recording point shown in this figure, we evaluated under the following conditions.

- **PS-1P0D**: Proposed system with 1 PAL and 0 EDLs (\( N = 1, M = 0 \), Fig. 2(a))
- **PS-1P1Dn**: Proposed system with 1 PAL and 1 EDL near PAL (\( N = 1, M = 1 \), Fig. 2(b))
- **PS-1P2Dn**: Proposed system with 1 PAL and 2 EDLs near PAL (\( N = 1, M = 2 \), Fig. 2(c))
- **PS-1P1Df**: Proposed system with 1 PAL and 1 EDL far from PAL (\( N = 1, M = 1 \), Fig. 2(d))
- **CS-1D**: Conventional system with 1 EDL (Fig. 2(e))
3.2 Evaluation experiments on sharpness of sound image

3.2.1 Evaluation conditions on sharpness of sound image

In the experiment on sharpness of the sound image, we calculated the CSP coefficient and evaluated the sharpness of the sound image produced with the proposed system by $D_{CSP}$. The CSP coefficient is calculated as

$$
CSP(t) = IDFT \begin{bmatrix} S_i(k)S_j(k)^* \end{bmatrix} \left[ \frac{|S_i(k)|}{|S_j(k)|} \right] \tag{6}
$$

where CSP(t) denotes the CSP coefficient, $t$ is the time index, $IDFT$ is the function of inverse discrete Fourier transform, $S_i(k), S_j(k)$ are the spectra of the signals observed with the $i$-th and $j$-th microphone, and $*$ indicates the complex conjugate operator. Also, $D_{CSP}$ is derived from

$$
D_{CSP} = \max_{-\frac{f_{sd}}{2} \leq t \leq \frac{f_{sd}}{2}} \{ CSP(t) \} - \frac{1}{2\pi} \left\{ \left( \frac{f_{sd}}{c} \sum_{t=-f_{sd}/c}^{f_{sd}/c} CSP(t) \right) - \max_{-\frac{f_{sd}}{2} \leq t \leq \frac{f_{sd}}{2}} \{ CSP(t) \} \right\} \tag{7}
$$

where $c$ is the sound velocity, and $d$ is the microphone interval. In this experiment, the larger value of $D_{CSP}$ means that the system can produce a sharper sound image. The CSP coefficient was measured using the observed signals recorded by two microphones. The microphone interval was set to 0.14 m ($d = 0.14$ m), and the direction of the microphones was rotated 45 degs. left around the center of two microphones. The other experimental conditions, equipment and arrangement were the same as shown in Tables 1, 2 and Fig. 3.

3.2.2 Evaluation results on sharpness of sound image

Figure 5 shows the experimental results for $D_{CSP}$ with each system. From this figure, we confirmed that the proposed system can produce sharper sound image than that of the conventional system because the $D_{CSP}$ with the proposed systems are larger than that of CS-1D. On the other hand, the $D_{CSP}$ with PS-1P1Dn, PS-1P2Dn, and PS-1P1Df are slightly smaller than that of PS-1P0D because of using EDLs in these proposed systems.

3.3 Evaluation experiments on performance of sound localization

3.3.1 Evaluation conditions on performance of sound localization

In the experiment on performance of sound localization, we evaluated the performance of sound localization by the ILD and the IACC. The ILD is calculated as

$$
ILD = 10 \log_{10} \frac{\sum_{t=0}^{T-1} s_l^2(t)}{\sum_{t=0}^{T-1} s_r^2(t)} \tag{8}
$$

where $T$ denotes the length of the signal, and $s_l(t)$ and $s_r(t)$ are the signals observed at the left and right ears, respectively.
Here, the large positive value of the ILD means that the sound pressure level of the observed signal at the left ear is larger than that at the right ear. Also, the IACC is calculated as

\[
IACC = \max_{-1 \leq \tau_A \leq 1} \left( \frac{\sum_{t=0}^{T-1} s_R(t)s_L(t + \tau_A)}{\sqrt{\sum_{t=0}^{T-1} s_R^2(t)} \sqrt{\sum_{t=0}^{T-1} s_L^2(t)}} \right)
\]  

(9)

where \(\tau_A\) is the index of the interaural time difference. If the IACC has a large value, it means that the correlation between the observed signal at the left ear and right ear is large. In this experiment, if the difference between the ILD and IACC with PS-1P0D is smaller, it means that the sound localization of PAL can be maintained with the system. The ILD and IACC are measured using the observed signals recorded by a binaural microphone (3DIO, Free Space Pro II). The microphone direction was rotated 45 degs. left to the loudspeaker placed in front of them. The other experimental conditions, equipment and arrangement were the same as shown in Tables 1, 2 and Fig. 3.

3.3.2 Evaluation results on sound localization

Figures 6 and 7 show the experimental results for ILD and IACC with each system. From Fig. 7, we confirmed that the sound localization of PAL is maintained by distributing the gain of the low-frequency sound because the IACC with PS-1P2Dn is the closest to that of PS-1P0D. However, it can be seen from Fig. 6 that the ILD with CS-1D is the closest to that of PS-1P0D. This is probably because in the cases of PS-1P1Dn, PS-1P2Dn, and PS-1P1Df, the power of the EDLs was the main factor instead of the power of the PAL, and the power difference between the left and right ears was reduced.

4. Conclusions

In this paper, we proposed sharp sound image production with high sound quality based on sound blending using PALs and EDLs. From the experimental results, we confirmed that the proposed system can produce a sharp sound image with high sound quality. In the future, we aim to further improve the sound quality by finding a suitable parameter for the crossover frequency. Also, we intend to improve the performance of the sound localization by distributing the gain of low-frequency sound to more EDLs.

Acknowledgment

This work was partly supported by JST COL, and JSPS KAKENHI Grant Numbers JP18K19829, JP18K11365, JP19H04142.

References

[1] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama and A. Ando: A 22.2 multichannel sound system for ultra high-definition TV (UHDTV), SMPTE Motion Imaging J., Vol. 117, No. 3, pp. 40-49, 2008.
[2] H. Kurabayashi, M. Otani, K. Itoh, K. Hashimoto and M. Kayama: Development of dynamic transaural reproduction system using non-contact head tracking, 2013 IEEE 2nd Global Conf. Consum. Electron. (GCCE), pp. 12-16, 2013.
[3] D. B. Ward and T. D. Abhayapala: Reproduction of a plane-wave sound field using an array of loudspeakers, IEEE Trans. Speech Audio Process., Vol. 9, No. 6, pp. 697-707, 2001.
[4] Dolby Laboratories, Inc.: https://www.dolby.com/us/en/technologies/home/dolby-atmos.html
[5] V. Pulikki: Virtual sound source positioning using vector base amplitude panning, J. Audio. Eng. Soc., Vol. 45, No. 6, pp. 456-466, 1997.
[6] J. J. Lopez, P. Gutierrez, M. Cobos and E. Aguilera: Sound distance perception comparison between wave field synthesis and vector base amplitude panning, 2014 6th Int. Symp. Commun. Contr. Signal Process. (ISCCSP), pp. 165-168, 2014.
[7] 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 Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC), pp. 1-9, 2012.
[8] M. Nakayama and T. Nishihara: Distance control of virtual sound source using parametric and dynamic loudspeakers, Proceedings of the 2018 Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC), pp. 1262-1267, 2018.
[9] H. K. Schilling, M. P. Givens, W. L. Nyborg, W. A. Pielemeier and H. A. Thorpe: Ultrasonic propagation in open air, J. Acoust. Soc. Am., Vol. 19, No. 1, pp. 222-234, 1946.
[10] C. Shi and W. S. Gan: Development of a parametric loudspeaker, IEEE Potentials, Vol. 29, No. 6, pp. 20-24, 2010.
[11] D. Itofuji, M. Nakayama, T. Nishihara and Y. Yamashita: Weighted double sideband modulation toward high quality audible sound on parametric loudspeaker, 2013 IEEE Int. Conf. Acoust. Speech Signal Process., pp. 843-847, 2013.
[12] L. Rabiner: Fundamentals of Speech Recognition, 1993.
[13] M. Omologo and P. Svazier: Acoustic source location in noisy and reverberant environment using CSP analysis, 1994 IEEE Int. Conf. Acoust. Speech Signal Process., Vol. 2, pp. 273-276, 1994.
[14] S. T. Birchfield and R. Gangisethty: Acoustic localization by interaural level difference, 2005 IEEE Int. Conf. Acoust. Speech Signal Process., Vol. 4, pp. 1109-1112, 2005.
[15] M. Morimoto and K. Iida: Appropriate frequency bandwidth in measuring interaural cross-correlation as a physical measure of auditory source width, Acoust. Sci. and Technol., Vol. 26, No. 2, pp. 179-184, 2005.
[16] ITU-R Rec. BS. 775-2: Multichannel stereophonic sound system with and without accompanying picture, 2006.