Acoust. Sci. & Tech. 40, 4 (2019)

PAPER

Evaluation of the extension and coloration of multiple listening zones synthesized by the shared field reproduction system

Taku Shimizu[1,2,*], Jorge Treviño[1], Shuichi Sakamoto[1], Yôiti Suzuki[1] and Tomohiko Ise[3]

[1]*Graduate School of Information Sciences and Research Institute of Electrical Communication, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai, 980–8577 Japan*
[2]*Research fellow of the Japan Society for the Promotion of Science (DC), Japan*
[3]*Alpine Electronics, Inc., 20–1, Yoshima Industrial Park, Iwaki, 970–1192 Japan*

(*Received 12 July 2018, Accepted for publication 1 March 2019*)

Abstract: A conventional sound field synthesis re-creates a highly accurate 3D sound space. However, it is difficult to present spatial sound to multiple listeners because an impractical number of loudspeakers is required for a single synthesized field to cover them all. The present research introduces a new approach, termed "shared sound field synthesis," to multi-zone reproduction, which allows multiple users to simultaneously listen to the spatial sound corresponding to a single sound field. The proposed method considers a new kind of multi-zone transfer function in the spherical harmonic domain to characterize the output of each loudspeaker in the reproduction array. These are calculated using a phase compensation formula to account for sound propagation between zones. Numerical simulations show that the proposed method can consistently achieve a two-zone reproduction with a consistent distortion level below -6 dB and is stable with respect to the relative position between the listening zones, except when the zones and target source are aligned. The size of the low distortion is similar to that of the single-zone synthesis obtained through high-order Ambisonics. Moreover, the proposed method dose not proved significant coloration at low frequency or near the center of the listening zones.

Keywords: Sound field control, Sound field synthesis, Multi-zone sound reproduction, Spatial sound, Spherical acoustics

PACS number: 43.60.Pt, 43.60.Sx [doi:10.1250/ast.40.241]

1. INTRODUCTION

Recently, audio systems have advanced to present highly realistic 3D spatial sound. Their technologies can be classified into three groups: binaural techniques [1,2], which control sound pressure at ears; surround sound techniques [3], which present spatial sound by driving loudspeakers in the direction of the sound source; and sound field synthesis [4–6], which creates a sound field physically, accounting for sound propagation in space.

This research considers sound field synthesis, as this approach offers the users some freedom of motion. As long as the users remain inside the listening zone, there is no need for tracking or individualization. In particular, this paper focuses on the problem of presenting spatial sound to multiple listeners. Conventional sound field synthesis methods, such as high-order Ambisonics (HOA) [6] are not optimal for presenting spatial sound to multiple users as this would require controlling the sound field over a region large enough to fit all the users. This implies that the method will also reproduce an accurate sound field in regions where no listener is present. The proposal is to re-create the target sound field inside multiple disconnected regions, one for each listener. That is, the multi-zone reproduction of a shared sound field.

Multi-zone sound field reproduction methods have been proposed in [7,8]. The goal of these methods, however, is not to provide a shared experience for multiple users but to create personal listening areas. To synthesize multiple sound fields by using multi-zone sound field methods, the sound field contents for all areas are required. Although the same content is used for all regions, this is not enough to achieve a shared sound field synthesis system because it may lead to sudden and unnatural changes in sound

*e-mail: shimizu@ais.riec.tohoku.ac.jp

pressure where two or more listening zones are close to each other. The reproduction of such a discontinuous sound field is very demanding in terms of the required number of loudspeakers. To solve this problem, we proposed an extension to HOA called "shared sound field synthesis" [9]. The current paper provides a detailed evaluation of this method and is organized as follows. Section 2 introduces the concept of a shared sound field and its difference from other sound field synthesis methods. In Sect. 3, we formulate the concrete signal processing method based on HOA by applying the concept of the shared sound field. Section 4 provides the evaluation of the performance of the proposed method through computer simulations. Finally, Sect. 5 presents our conclusions.

2. CONCEPT OF SHARED SOUND FIELD

This section introduces the concept of the shared sound field. The goal of this research is to present a single sound space with multiple listeners. If there are multiple listeners, conventional sound field synthesis methods require an impractical number of loudspeakers for controlling a vast area to cover all the listeners. In the case of HOA, the number of loudspeakers increases with the distance between listeners because the synthesized zone obtained by HOA is a single connected sphere. However, the listeners need not to be concentrated inside a small region; they may be distributed sparsely over an extended area. Furthermore, it is not necessary to control the sound field in regions where no listener is present, and focusing only on the neighborhoods of the users can result in a more optimal use of the available loudspeakers. This is illustrated in Fig. 1.

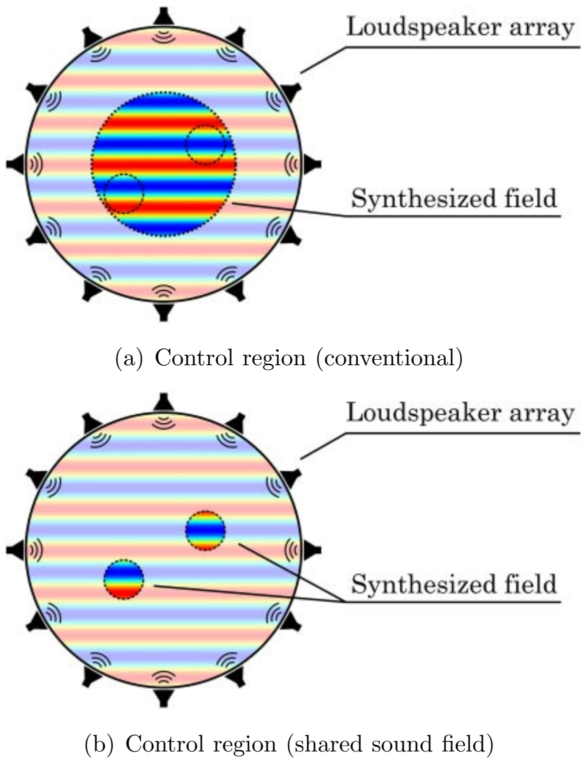


Fig. 1 (a) Regions to cover multiple listeners, which are controlled during conventional sound field synthesis and (b) the shared sound field synthesis approach. Faint colors show the underlying (target) sound field while bright colors show the reproduction areas.

An underlying shared sound field is assumed for all regions to prevent sound pressure mismatches. A direct application of existing multi-zone reproduction methods without the shared field assumption can lead to unnatural jumps in sound pressure between zones close to each other, as shown in Fig. 2(a). Under the assumption of a shared field, the target for each zone is adjusted according to its position, as shown in Fig. 2(b), thus avoiding the discontinuities and resulting in a smooth field with reduced inter-zone interference. Moreover, in the case of the overlapping the zones applied the multi-zone reproduction methods, the zones cannot be produced due to the synthesis of different sound pressures in the same region. However, the zones based on the shared field are free from the limitation as the results of the adjustment according to the listeners' positions.

In this study, we focus on the synthesis of a plane wave field. Although we evaluate only a single plane wave, sound field that consists of multiple plane waves can be synthesized by the superposition of these plane waves. This is possible because sound propagation is linear in the domain considered in our study. Moreover, the proposed method can be coupled with plane wave expansion to present any sound field that admits such decomposition.

3. SHARED SOUND FIELD SYNTHESIS BASED ON HOA

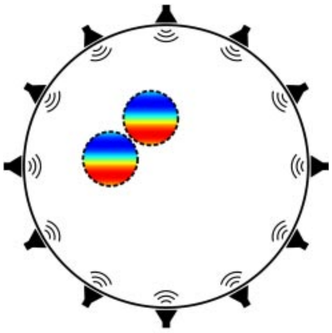
This section introduces the details of the application of a shared sound field; the method is named "shared sound field synthesis." Moreover, before the introduction, we summarize the synthesis of HOA. In the present paper, we assume a spherical coordinate system, as shown in Fig. 3. Therefore, the Cartesian coordinates $\boldsymbol{r} = [x, y, z]^{\mathrm{T}}$ are converted from spherical coordinates as follows:

$$\begin{aligned} x &= r\cos\theta\cos\varphi, \\ y &= r\sin\theta\cos\varphi, \\ z &= r\sin\varphi, \end{aligned} \tag{1}$$

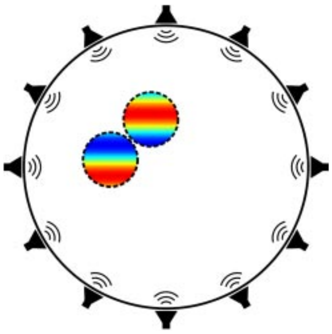
where r, θ, and φ denote the radial distance, azimuth angle, and elevation angle of the position $\boldsymbol{r}$; and superscript T represents the transpose.

3.1. Conventional HOA

The synthesis of conventional HOA [6] is obtained by inverting the linear system composed of all transfer



(a) Multi-zone sound field reproduction



(b) Shared sound field synthesis

Fig. 2 (a) The comparison of the synthesis of multiple fields through the conventional multi-zone approach and (b) the shared sound field synthesis approach. The fields obtained by the conventional multi-zone sound field reproduction may lead to unnatural jumps in sound pressure when the fields are close to each other (a), while such a jump is not observed with the assumption of a shared field (b).

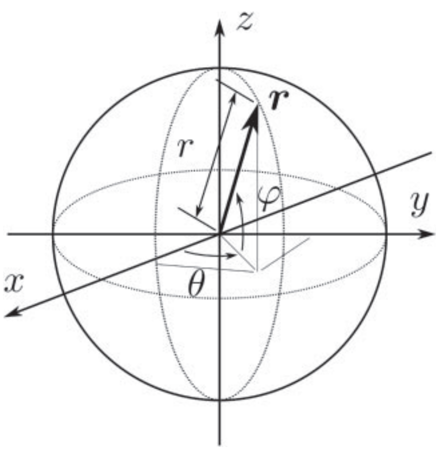


Fig. 3 Spherical coordinate system used in the present study.

functions between each loudspeaker and the listening position. The linear system can be expressed as follows:

$$\mathrm{C}\boldsymbol{w}^{(\mathrm{HOA})} = \boldsymbol{B}, \tag{2}$$

where

$$\boldsymbol{w} = [w_1^{(\mathrm{HOA})}, w_2^{(\mathrm{HOA})}, \ldots, w_l^{(\mathrm{HOA})}, \ldots, w_L^{(\mathrm{HOA})}]^{\mathrm{T}}, \tag{3}$$

$$\boldsymbol{B} = [B_{0,0}, B_{1,-1}, B_{1,0}, \ldots, B_{n,m}, \ldots, B_{N,N}]^{\mathrm{T}}. \tag{4}$$

Here, $w_l^{(\mathrm{HOA})}$ represents the HOA signal weight of the l-th loudspeaker located at $\boldsymbol{r}_l = [x_l, y_l, z_l]^{\mathrm{T}}$ $(l \in \{1, 2, \ldots, L\})$. Furthermore, coefficient $B_{n,m}$ is characterized by the target sound field and truncated up to maximum order N. If the target sound field is a plane wave, the coefficients are described as follows [6]:

$$B_{n,m} = 4\pi i^n Y_{n,m}^*(\theta^{(\mathrm{src})}, \varphi^{(\mathrm{src})}), \tag{5}$$

where $Y_{n,m}$ represents a spherical harmonic function of order n and degree m. Superscript $*$ denotes the complex conjugate, and i is the complex number. $\theta^{(\mathrm{src})}$ and $\varphi^{(\mathrm{src})}$ denote the azimuth and the elevation of incidence direction of the plane wave, respectively, and C denotes the transfer function matrix, which consists of elements $\mathrm{C}_{n,m;l}$, as follows:

$$\mathrm{C}_{n,m;l} = -ikh_n^{(2)}(kr_l)Y_{n,m}^*(\theta^{(\mathrm{src})}, \varphi^{(\mathrm{src})}), \tag{6}$$

where $h_n^{(2)}$ is a spherical Hankel function of the second kind of order n, k represents the wavenumber and r_l, θ_l, and φ_l denote the radial distance, azimuth angle, and elevation angle of the l-th loudspeaker depending on $\boldsymbol{r}_l$. This equation is slightly different from that proposed in [6] because we adhered to the $\exp(i\omega t)$ sign convention, where ω denotes angular frequency and t represents time.

3.2. Proposed Method

The schema of the proposed shared field approach is shown in Fig. 4. The assumption of a single sound field shared among all listening positions eliminates the need to synthesize unnatural sound fields, and can therefore reduce the necessary number of loudspeakers.

The proposal follows the same ideas used in the conventional HOA but modifies Eq. (2) to achieve multi-zone reproduction of a shared sound field. To this end, two changes were introduced to this equation: a phase compensation for expansion coefficients $B_{n,m}$ and a set of multi-zone transfer functions replacing single zone C.

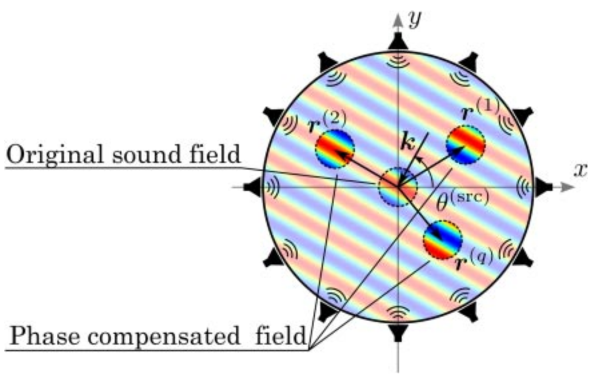


Fig. 4 Proposed shared field approach.

3.2.1. Phase compensation

A phase compensation is needed to match the sound field description at each listening position to the shared sound field. This considers the finite speed of sound; as all the listening positions share the same sound field, those closer to the sound source possess an advanced phase compared to those farther from it. Furthermore, this ensures that the listening positions close to one another will not result in abrupt changes in the synthesized sound field. As observed from listening position located at $\boldsymbol{r}^{(q)}$ ($q \in \{1, 2, \ldots, Q\}$), the target sound field is represented as follows:

$$\begin{aligned} p(\boldsymbol{r} + \boldsymbol{r}^{(q)}) &= \exp(-i\,\boldsymbol{k}^{\mathrm{T}}(\boldsymbol{r} + \boldsymbol{r}^{(q)})) \\ &= \exp(-i\,\boldsymbol{k}^{\mathrm{T}}\boldsymbol{r}^{(q)})p(\boldsymbol{r}), \end{aligned} \tag{7}$$

where $\boldsymbol{k}$ denotes the wavevector of the encoded plane wave, as shown in Fig. 4, which is given by the following formula:

$$\begin{aligned} \boldsymbol{k} = -[&k \cos\theta^{(\mathrm{src})} \cos\varphi^{(\mathrm{src})}, \\ &k \sin\theta^{(\mathrm{src})} \cos\varphi^{(\mathrm{src})}, \\ &k \sin\theta^{(\mathrm{src})}]^{\mathrm{T}}. \end{aligned} \tag{8}$$

This equation provides the q-th listening-position coefficients $\boldsymbol{B}^{(q)}$ as follows:

$$\boldsymbol{B}^{(q)} = \exp(-i\,\boldsymbol{k}^{\mathrm{T}}\boldsymbol{r}^{(q)})\boldsymbol{B}. \tag{9}$$

3.2.2. Multi-zone transfer function

The formulation of the proposal considers that the sound fields for all listening positions must be simultaneously generated by a single set of loudspeakers. However, conventional HOA considers the synthesis only at the center of the loudspeaker array. Therefore, all of the transfer functions between each loudspeaker and each of the listening positions must be considered simultaneously. To this end, a new multi-zone transfer function, $\mathrm{C}^{(q)}_{n,m;l}$, is defined for each listening position q as follows:

$$\mathrm{C}^{(q)}_{n,m;l} = -ikh^{(2)}_n(kr^{(q)}_l)Y^*_{n,m}(\theta^{(q)}_l, \varphi^{(q)}_l). \tag{10}$$

Here, $r^{(q)}_l$, $\theta^{(q)}_l$, and $\varphi^{(q)}_l$ represent the position of loudspeaker l as seen from listening position q.

3.2.3. Proposed method

To simplify the derivation of the proposed method, we first considered single-zone synthesis at the listening position. For this, Eq. (2) was modified as follows:

$$\mathrm{C}^{(q)}\boldsymbol{w}^{(q)} = \boldsymbol{B}^{(q)}, \tag{11}$$

where

$$\boldsymbol{w}^{(\mathrm{q})} = [w^{(q)}_1, w^{(q)}_2, \ldots, w^{(q)}_l, \ldots, w^{(q)}_L]^{\mathrm{T}}. \tag{12}$$

Here, $w^{(q)}_l$ represents the signal weight of the loudspeaker l for listening position q. By substituting Eq. (9) into (11), the equation for the single-zone synthesis of listening zone q is expressed as follows:

$$\mathrm{D}^{(q)}\boldsymbol{w}^{(\mathrm{q})} = \boldsymbol{B}, \tag{13}$$

where

$$\mathrm{D}^{(q)} = \exp(i\,\boldsymbol{k}^{\mathrm{T}}\boldsymbol{r}^{(q)})\mathrm{C}^{(q)}. \tag{14}$$

By using Eq. (14), we aim to minimize the average error across all listening positions; this leads to the formation of the following linear system:

$$\mathrm{S}\boldsymbol{w}^{(\mathrm{s})} = \boldsymbol{B}, \tag{15}$$

where

$$\mathrm{S} = \frac{1}{Q}\sum_{q=1}^{Q}\mathrm{D}^{(q)}, \tag{16}$$

$$\boldsymbol{w}^{(\mathrm{s})} = [w^{(\mathrm{s})}_1, w^{(\mathrm{s})}_2, \ldots, w^{(\mathrm{s})}_l, \ldots, w^{(\mathrm{s})}_L]^{\mathrm{T}}. \tag{17}$$

Here, $w^{(\mathrm{s})}_l$ denotes the signal weight of the loudspeaker l derived by the shared sound field synthesis. This linear system can be inverted to derive the signal weights $\boldsymbol{w}^{(\mathrm{s})}$ as follows:

$$\boldsymbol{w}^{(\mathrm{s})} = \mathrm{S}^{+}\boldsymbol{B}. \tag{18}$$

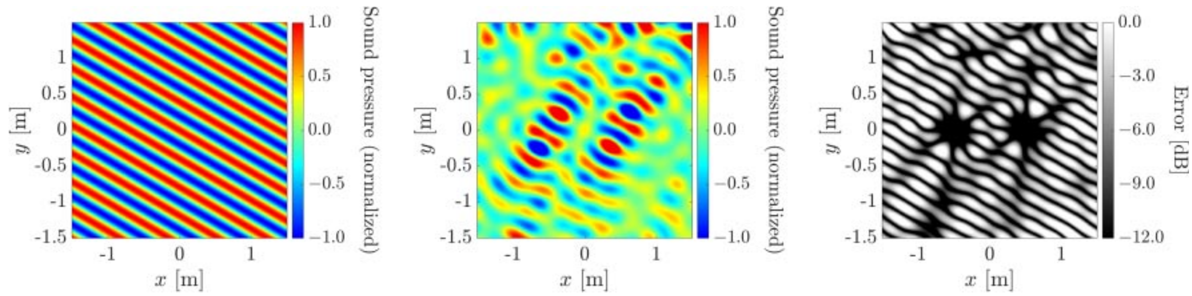
Here, S^+ is the pseudo-inverse of S, which is the equivalent to the decoder in a conventional HOA.

4. NUMERICAL SIMULATION

In the present paper, we provide evaluations for the proposed method by considering the case of the reproduction of a plane wave, the amplitude of which is normalized to unity, in two zones. The azimuth and elevation angles of the incidence direction of the plane wave are 60° and 0°, respectively. The source frequency is 1 kHz. The spherical harmonic expansion characterizing the target sound field was calculated up to order 4. The reproduction system consists of a 192-channel loudspeaker array distributed on the surface of the sphere with a radius of 2.5 m. The loudspeaker distribution was obtained by solving the Thomson problem, which is to determine the position of electrons on a sphere by minimizing the potential energy by using the steepest descent method [10]. The center of the loudspeaker array is at the origin of the coordinates. The two listening positions are located −0.5 m (position 1) and 0.5 m (position 2) away from the origin on the x-axis; thus, $\boldsymbol{r}^{(1)} = [-0.5, 0, 0]^{\mathrm{T}}$, $\boldsymbol{r}^{(2)} = [0.5, 0, 0]^{\mathrm{T}}$. Moreover, the source frequency, incidence direction, and listening positions may be changed, depending on simulations.

The simulation area, in which sound pressures are calculated, is an x–y plane with $z = 0$ m; the spatial resolution is one hundredth of the wavelength. The synthesized field obtained by the proposed method $p^{(\mathrm{s})}$ is calculated as follows:

$$p^{(\mathrm{s})}(\boldsymbol{r}, t) = \mathrm{Re}(\mathcal{F}^{-1}[S(\boldsymbol{r}, k)]), \tag{19}$$



(a) Ideal field (b) Synthesis field (c) Synthesis error

Fig. 5 Sound pressure field obtained using the proposed method with two-zone synthesis of a 1 kHz plane wave with incidence direction from 60° azimuth and 0° elevation at $\alpha = 0$. The two listening positions are located -0.5 and 0.5 m away from the center of the loudspeaker array on the x-axis, respectively.

where $\mathcal{F}^{-1}$ represents the inverse Fourier transform in the time domain. S denotes the spatial component of the synthesized sound pressure calculated by the following formula:

$$S(\boldsymbol{r}, k) = \sum_{l=1}^{L} w_l^{(s)} \frac{\exp(-ik|\boldsymbol{r} - \boldsymbol{r}_l|)}{4\pi|\boldsymbol{r} - \boldsymbol{r}_l|}. \tag{20}$$

This is equivalent to the Fourier transform of the sound pressures for each observation point $\boldsymbol{r}$ in the time domain. Therefore, these values indicate the frequency response of the synthesized sound pressure. In this study, the signal weights $w_l^{(s)}$ were calculated for each incidence direction and the frequency of the target source.

For evaluating the performance of our proposed method, the error between synthesized sound pressure $p^{(s)}$ and ideal sound pressure $p^{(i)}$ observed at $\boldsymbol{r}$ was measured as follows:

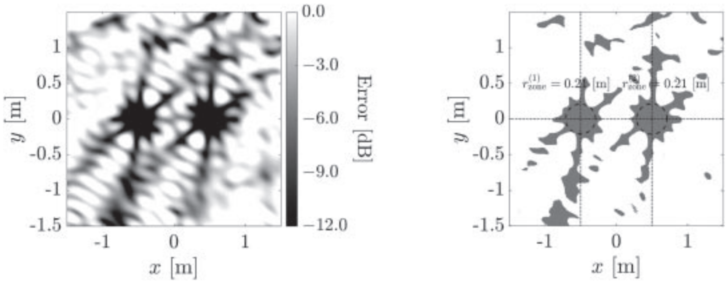
$$E(\boldsymbol{r}) = \max_t [10 \log_{10}\{|p^{(s)}(\boldsymbol{r}, t) - p^{(i)}(\boldsymbol{r}, t)|^2\}],$$
$$\left(t = \frac{\alpha}{fA},\ \alpha = 0, 1, \ldots, A - 1\right), \tag{21}$$

where A denotes the number of evaluated instants and f represents the frequency of target sound, and t denotes the sampled time at the rate of $1/fA$ for calculating sound pressures. Moreover, operator $\max_t\{-\}$ gives the maximum value, which a function attains for set t. This evaluation method considers a varying reproduction field as a function of time. In our simulation, A was set at 5. Moreover, we define the listening zone as a circle in the x–y plane ($z = 0$) in which the error level calculated by Eq. (21) is below -6 dB. The center of the listening zones are set to each of the listening positions; their radii $r_{\text{zone}}^{(q)}$ are maximized while excluding the area in which the error is more than or equal to -6 dB. The listening zones of position 1 and position 2 are defined as zone 1 and zone 2, respectively.

4.1. Evaluation of Proposed Method

First, the relationship between synthesis accuracy and the position of listening zones and incident direction of the target source was analyzed. Figure 5 shows the simulation results obtained by applying the proposed method when $\alpha = 0$. Figures 5(a) and 5(b) show the sound pressure of the target and synthesized fields, respectively, and Fig. 5(c) shows the magnitude of the synthesis error between the sound pressures in Figs. 5(a) and 5(b). Moreover, Fig. 6 shows the synthesis error between the field synthesized by the proposed method and target field, calculated using Eq. (21) at all instants of time. Figure 6(a) shows the magnitude of synthesis error between the ideal and synthesized fields. Figure 6(b) shows the listening zones obtained by the synthesis error in Fig. 6(a). The results of Fig. 6(b) show that the radii of the listening zones are 0.21 m. As shown in Fig. 6, the synthesis errors at the two zones in the neighborhood of the listening positions indicate small values, showing the validity of the proposed method.

Furthermore, we analyzed the size of the listening zones while changing the incidence direction of the plane wave. In the simulation, the azimuth angle of the incident direction was varied from 0° to 359° at intervals of 1°. Figure 7 shows the size of the listening zones as a function of the incidence direction of sound. As shown in the figure, the radii of the target zones scarcely change except for the azimuth angles of 0° and 180°. Figure 8 shows the sound pressure field synthesized for the plane wave incoming from 0°. Although Fig. 8(b) shows that the errors in the regions near the listening zones are relatively small, these errors are larger than those in the other incidence directions shown in Fig. 6(a). Moreover, Fig. 8(c) shows that the listening zone disappears at some areas around the listening positions. The sound pressure field for the incidence angle of 180° behaves almost the same as that for the incidence angle of 0°. These results show that the listening zones shrink significantly when the zones and source are aligned. These results are a consequence of the target field being considered in our study. Our proposed method drives the



(a) Synthesis error calculated using Eq. (21)

(b) Low error regions

Fig. 6 Synthesis error between the field synthesized by the proposed method and target field calculated at all instants of time. The target source is a 1 kHz plane wave with an incidence direction of 60° azimuth and 0° elevation. The number of time instants is 5. The two listening positions are located −0.5 and 0.5 m away from the center of the loudspeaker array on the x-axis, respectively. The low error regions are defined as regions at which the error level is below −6 dB, as shown in Fig. 6(b). These position are illustrated by the intersections of the dotted lines in Fig. 6(b). The dashed-line circles in Fig. 6(b) shows the edges of the listening zones. The radii of the listening zones are $r^{(1)}_{\text{zone}} = 0.21$ and $r^{(2)}_{\text{zone}} = 0.21$ m.

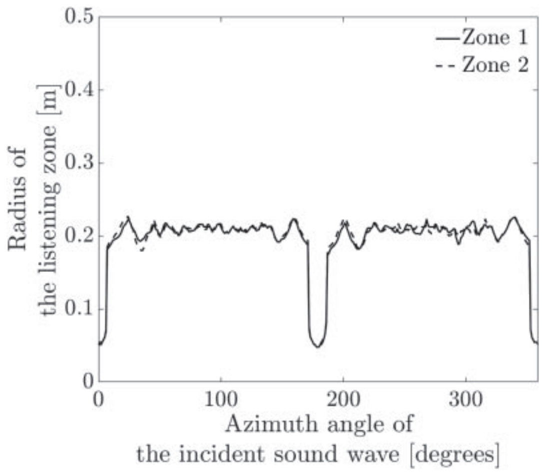


Fig. 7 Radii of the listening zones as a function of the azimuth angle of incidence. The target source is a 1 kHz plane wave with an incidence direction from an elevation of 0° and the azimuth angle is varied from 0 to 359° at intervals of 1°. The two listening positions are located −0.5 and 0.5 m away from the center of the loudspeaker array on the x-axis, respectively. The number of time instants is 5.

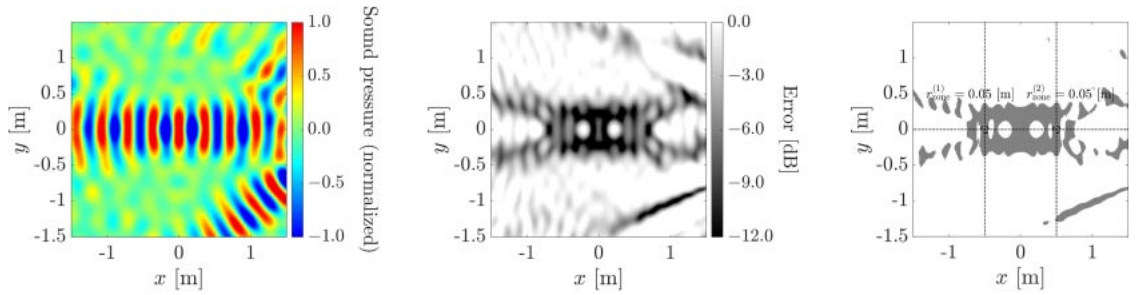
loudspeakers so that the spherical harmonic expansion coefficients for the target sound field match those of a plane wave up to a maximum order; higher orders are not controlled (assumed to be zero). The result for a given listening zone is a target sound field that exhibits planar wavefronts inside the listening zone and vanishing small sound pressures outside it for directions different from that of sound propagation. However, the sound pressure does not vanish in the direction of propagation; it consists of curved wavefronts similar to an outgoing spherical wave. When two listening zones are aligned in the direction of sound propagation, it leads to inconsistent target sound fields between the zones (planar and curved wavefronts). Solving this issue requires controlling higher orders (making them non-zero) through the loudspeakers, which in turn increases system complexity (higher number of loudspeakers).

We then examined the size of the listening zones while changing the distance between listening positions. In the simulation, the distance between the listening positions was varied from 0 m to 4.9 m at intervals of 0.01 m away from the center of the listening position to the origin of the coordinates, i.e., the listening positions are located $\boldsymbol{r}^{(1)} = [-1, 0, 0]$ and $\boldsymbol{r}^{(2)} = [1, 0, 0]$ when the distance is 2 m. The results in Fig. 9 display that the radii of the listening zones is stable with a value of approximately 0.2 m up to a distance of 3.74 m, and then it reduces abruptly. This reduction seems to occur because the listening positions are very close to the loudspeaker array. These results show the robustness of the proposed method in terms of the distance between listening positions, unless they are close to the loudspeaker array.

The next simulation focused on the frequency characteristics of the proposed method. In the simulation, the source frequency was varied in 2,048 uniform steps up to 24 kHz. Figure 10 shows the radii of the listening zones as a function of the frequency of the sound source. The dash-dotted line illustrates expressed the radius of the listening zone obtained by the single-zone HOA [6] when the listening position is located the center of the loudspeaker array. By using HOA, the radius of a listening zone is approximated by [11]

$$R \approx \frac{N}{k}. \tag{22}$$

However, this is calculated for a threshold of −14 dB using a different formula from the one applied in this paper. The method to calculate the synthesis error E_N in [11] (Eq. (15)) is:



(a) Synthesized field (b) Synthesis error calculated using Eq.(21) (c) Low error regions

Fig. 8 Sound pressure field and its error obtained using the proposed method with the two-zone reproduction of a 1 khz plane wave with an incidence direction of 0° azimuth angle and 0° elevation angle. (a) shows the synthesized field when $\alpha = 0$. (b) shows the synthesis error between the synthesized and the ideal field at all instants of time; the low error regions where the error level is below -6 dB are shown in (c). The number of time instants is 5. The two listening positions are located -0.5 and 0.5 m away from the center of the loudspeaker array on the x-axis, respectively. These position are illustrated by the intersections of the dotted lines in (c). The dashed-line circles in (c) shows the edges of the listening zones. The radii of the listening zones are $r^{(1)}_{\text{zone}} = 0.05$ and $r^{(2)}_{\text{zone}} = 0.05$ m.

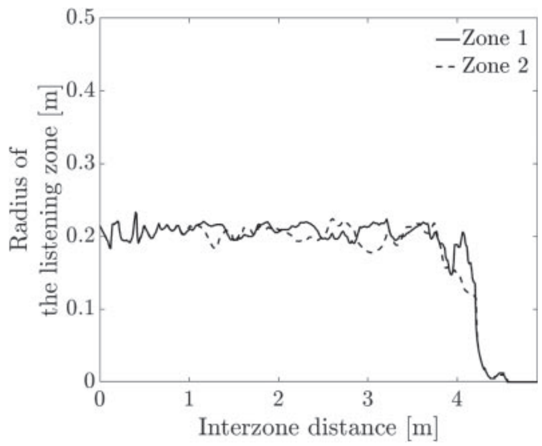


Fig. 9 Radii of the listening zones as a function of the distance between the listening positions. The distance is varied from 0 to 4.90 m at intervals of 0.01 m; the listening positions are located at the same distance from the center of the loudspeaker array on the x-axis. Thus, $\boldsymbol{r}_1 = [-1, 0, 0]^{\mathrm{T}}$ and $\boldsymbol{r}_2 = [1, 0, 0]^{\mathrm{T}}$ when the distance is 2 m. The target source is a 1 kHz plane wave with an incidence direction of 60° azimuth angle and 0° elevation angle. The number of time instants is 5.

$$E_N(kR) = \frac{\displaystyle\iint |S(R,\theta,\varphi;k) - \hat{S}(R,\theta,\varphi;k)|^2 \cos\varphi \, d\varphi d\theta}{\displaystyle\iint |S(R,\theta,\varphi;k)|^2 \cos\varphi \, d\varphi d\theta}. \tag{23}$$

Here, $\hat{S}$ represents the approximate plane wave field after truncating the spherical harmonic expansion at order N. S is the original plane wave field. This method looks at the average error on the surface of a sphere in the frequency domain. In contrast, this paper uses the maximum error at all points inside a sphere of radius R. Further, this error is evaluated in the time domain at multiple instants, with the highest error being used. Therefore, the synthesis error in our manuscript seems to be greater than or equal to that in [11]. The threshold of -6 dB does not correspond to the threshold of -14 dB in [11] directly. To compare the size of the listening zone of our proposed method and HOA, we obtained the radius of the listening zone for conventional HOA, which is simulated with the same condition as that of the proposed method. The listening position of the HOA is defined as the center of the loudspeaker array. This value is drawn with a dash-dotted line in the figure, indicating that the proposed method can synthesize as many sound fields as the HOA. Moreover, while we have used the different method to calculate the listening zone size from that in [11], the values shown in Fig. 10 are similar to the result of Eq. (22); this result is simply a coincidence. Both error metrics are unrelated, as explained above. We chose the more strict maximum time-domain error as our metric to account for potential artifacts due to interference between the listening zones.

Finally, we evaluated the frequency response by applying the proposed method. In the simulation, the source frequency was varied in 2,048 uniform steps up to 24 kHz. Figure 11 shows the magnitudes of the sound pressures synthesized using the proposed method; the observation points are located at 0.0 m (Fig. 11(a)) and 0.1 m (Fig. 11(b)) away from the listening position 1 on the x-axis. These magnitudes are the absolute values of the results calculated by Eq. (21). It is desirable that these values are 0 dB because we assumed that the amplitude of the target plane is normalized to unity. Spherical harmonics-based sound field synthesis methods exhibit large gains at low frequencies [12]. However, as shown in Fig. 11(a),

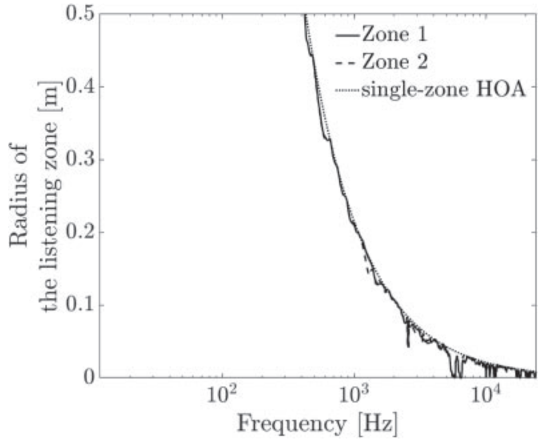
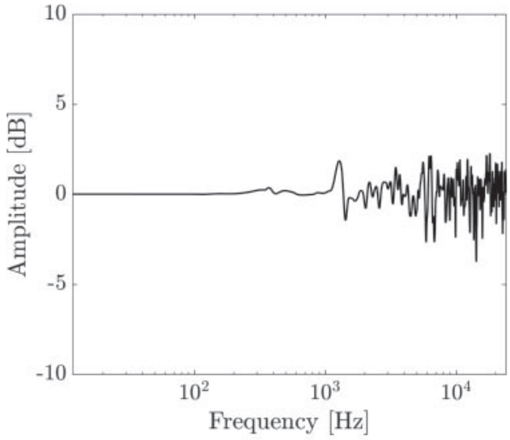
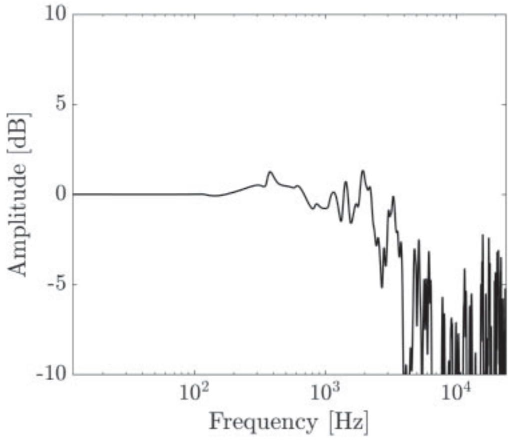


Fig. 10 Radii of the listening zones as a function of the target source frequency. The target source is a plane wave with an incidence direction of 60° azimuth angle and 0° elevation angle. The result obtained by the single-zone HOA when the listening positions are located at the center of the loudspeaker array is expressed by the dotted line. The source frequency is varied in 2,048 uniform steps up to 24 kHz. The listening position 1 and position 2 are located −0.5 and 0.5 m away from the center of the loudspeaker array on the x-axis, respectively. The number of time instants is 5.



(a) $x = -0.5$ m (on the listening position 1)



(b) $x = -0.4$ m (located 0.1 m away from the listening position 1 on x-axis)

Fig. 11 Magnitudes of the frequency response of the sound pressures synthesized using the proposed method at two points. The listening position 1 and position 2 are located −0.5 and 0.5 m away from the center of the loudspeaker array on the x-axis, respectively. The measurement points are located 0.0 m and 0.1 m away from the listening position 1 on the x-axis, respectively. The target source is a plane wave with an incidence direction of 60° azimuth angle and 0° elevation angle. The source frequency is varied in 2,048 uniform steps up to 24 kHz.

our proposed method is free from this common issue, with a frequency response variation of less than 1 dB up to 1,195 Hz on the listening position 1. Further, as shown in Fig. 11(b), the frequency response on the position located away from the listening position 1 is varied by less than 1 dB up to 375 Hz and by less than 2 dB up to 2,379 Hz.

5. CONCLUSION

In this paper, we introduced the concept of a shared sound field to tackle the problem of presenting sound fields to multiple listeners by using the existing sound field synthesis methods. The proposed method termed "shared sound field synthesis" is an extension of the conventional HOA and multi-zone sound field reproduction. It incorporates a phase compensation stage and uses a multi-zone transfer function to synthesize an underlying shared sound field inside multiple, disconnected regions. We evaluated the proposed method through numerical simulations, and the results showed that the method can synthesize sound fields inside two zones. The zonal accuracy does not change with respect to the separation between zones. The direction of incidence does not cause a rapid change in accuracy, except when the listening zones and target source are aligned. Moreover, we confirmed that the proposed method can synthesize sound fields as accurately as a single-reproduction method using HOA and does not provide a significant gain at low frequency near the listening position.

The proposed method requires the incidence angle of the target plane wave in advance. Therefore, this method cannot be directly applied to natural sound fields such as a recorded sound field. However, this study focuses on sound field synthesis applications such as VR; therefore, these information can be obtained. Nevertheless, the sound field sharing idea can be applied to recorded sound fields as long as they are known at two or more positions, either via multiple recordings or by translating the spherical harmonic expansion coefficients [13].

ACKNOWLEDGMENTS

Part of this work was supported by JSPS Grant-in-Aid for JSPS Research Fellow No. 16J07287, Grant-in-Aid for Scientific Research(A) No. 24240016, and the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University (H26/S1).

REFERENCES

[1] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn.* (*E*), **1**, 167–174 (1980).

[2] I. Toshima, H. Uematsu and T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement: TeleHead," *Acoust. Sci. & Tech.*, **24**, 327–329 (2003).

[3] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama and A. Ando, "A 22.2 multichannel sound system for ultra high-definition TV," *SMPTE Motion Imaging J.*, **117**, 40–49 (2008).

[4] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *Acta Acust. united Acu.*, **85**, 78–87 (1999).

[5] A. J. Berkhout, "A holographic approach to acoustic control," *J. Audio Eng. Soc.*, **36**, 977–995 (1988).

[6] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, **53**, 1004–1025 (2005).

[7] M. Poletti, "An investigation of 2-D multizone surround sound systems," *125th Audio Eng. Soc. Conv.*, pp. 1–9 (2008).

[8] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio Speech Lang. Process.*, **19**, 1711–1720 (2011).

[9] T. Shimizu, J. Trevino, S. Sakamoto and Y. Suzuki, "A multizone approach to sound field reproduction based on spherical harmonic analysis," *Acoust. Sci. & Tech.*, **36**, 441–444 (2015).

[10] V. A. Yudin, "The minimum of potential energy of a system of point charges," *Discrete Math. Appl.*, **3**, 75–81 (1993).

[11] D. B. Ward, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Process.*, **9**, 697–707 (2001).

[12] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," *Proc. Audio Eng. Soc. 23rd Int. Conf.*, 15 pages (2003).

[13] Y. Wang and K. Chen, "Translations of spherical harmonics expansion coefficients for a sound field using plane wave expansions," *J. Acoust. Soc. Am.*, **143**, 3474–3478 (2018).