Audio Engineering Society

Convention Paper 10610

Presented at the 152nd Convention
2022 May, In-Person and Online

Degradation in reproduction accuracy due to sound scattered by listener's head in local sound field synthesis

Izumi Tsunokuni[1] and Yusuke Ikeda[1]

[1]*Tokyo Denki University*

Correspondence should be addressed to Izumi Tsunokuni (`21ud02@ms.dendai.ac.jp`)

ABSTRACT

The purpose of this study is to investigate the phenomenon of degradation in the accuracy of local sound field synthesis (LSFS) due to the sound scattered by a listener's head. In conventional sound field synthesis (SFS) methods, the degradation in accuracy due to a listener's head is negligible, because the degradation are smaller at the low reproducible frequencies than the discretization artifacts of synthesized sound field. As LSFS method synthesizes the sound field only to a narrow extent at higher frequencies which is not considered in the conventional methods, how degraded the reproduction accuracy due to scattered sound in LSFS must be investigated. We conducted simulation experiments, using a rigid sphere for modeling the sound scattered by the head, using two LSFS methods: local wave field synthesis with virtual secondary sources (LWFS-VSS), and the pressure-matching method. The following two points were investigated: (i) The dependency of degradation on the frequency of sound and reproduction position; and (ii) the relationship between the virtual source distance and reproduction accuracy. The results showed that the degradation in the accuracy at the position opposite to the virtual source became larger as the frequency increased. Regarding the distance of the virtual source, when the source was placed near the listener's head, the reproduction accuracy was significantly low. Specifically, in the case of LWFS-VSS, as the virtual source approached the head, the reproduction accuracy became more degraded compared with the no-scattering condition.

1 Introduction

Various spatial audio techniques have been proposed with headphones and multiple loudspeakers [1, 2]. Audio systems based on psychoacoustic effects, such as the stereophony, and multi-channel surround systems [3], are widely introduced in movie theaters and home theaters. However, a sweet spot where the listener can correctly perceive the sound image is restricted to a very narrow area. In contrast, sound field synthesis (SFS), such as Wave Field Synthesis (WFS) [4] and Higher order Ambisonics (HOA) [5], control a physically accurate sound field with multiple loudspeakers (secondary sources). An interior closed by a loudspeaker array generally become the reproduction area in SFS. There are two types of SFS methods to derive the driving functions of secondary sources, namely, analytical and numerical methods [2].

Typical analytical SFS methods, such as WFS and HOA, derive the driving function on the basis of the assumption that there are no acoustic scattering objects like a listener's head in the room. Thus, it is required to compensate for room reverberation. In numerical methods, such as pressure matching (PM) method [6] and BoSC [7], the sound field can be controlled when the transfer functions from the secondary sources to the controlled points are measured in advance. These numerical methods also ignore the sound reflection and scatter by a listener's head in the synthesized sound field.

In [8], Ahrens *et al.* reported a sound field scattered by a listener's head in SFS. The listener's head was modeled by a rigid sphere. The simulation using a rigid sphere as the model of the sound reflection from a listener's head showed that the degradation in synthesized plane-wave field is sufficiently small to be ignored at frequencies lower than 1 kHz. At frequencies over 2 kHz, the discretization artifacts are larger than the degradation caused by the reflection due to the listener's head in the conventional analytical SFS method.

In recent years, local SFS methods have been proposed to control a sound field around a listener's head at higher frequencies using large-scale loudspeaker systems [9, 10]. In [9], Spors *et al.* proposed local wave field synthesis using virtual secondary sources (LWFS-VSS), which are generated using focused sources inside a real loudspeaker array. Since the virtual secondary sources can be arranged at shorter intervals compared to the real one, the sound field can be synthesized at higher frequency regardless of the discretization artifacts that are caused by the secondary source spacing. Similarly, Tsunokuni *et al.* proposed the PM method at higher frequencies from small number of measurements by modeling the room impulse responses based on the equivalent source method [11, 12]. Therefore, the effect of head reflection at higher frequencies is required to be revealed. In addition, since LSFS makes it possible to set the distance of the desired sound source close to the listener, the effect of head reflection due to the source distance is also required to be revealed.

In this study, we investigate the degradation of the synthesized sound field due to the reflection by the listener's head. In particular, we investigate the degradation at higher frequencies and its dependance on the distance of the disired sound source. Differently from [8], the sound field is synthesized at frequencies above 2 kHz using local SFS methods. The sound field is synthesized in the following two local SFS method: the PM method [6] as a numerical method and LWFS-VSS [9] as an analytical method. These two methods can reproduce the sound field without depending on the interval of the secondary source, namely, the sound field at higher frequencies can be considered. In addition, the desired source is assumed to be a point source. In [8], the sound source was assumed to be a plane wave; thus, the deterioration of scattered sound field by the reproduced source position is not still revealed. In particular, we focus on the relationship between the distance of the desired sound source and the degradation of sound field synthesis. The simulation experiments were conducted using a circular secondary source array and a rigid sphere model the listener's head.

2 Sound field scattered by a listener's head

We consider an acoustically-rigid sphere of radius a is centered at the origin of the coordinate system, which we regard as the listener's head. In this paper, the Cartesian coordinates (x,y,z) are related to the spherical coordinates (r,α,β) with $x = r\cos\alpha\sin\beta, y = r\sin\alpha\sin\beta$, and $z = r\cos\beta$.

When the listener's head is included in the synthesized sound field, the resulting field can be decomposed as follows [2],

$$S(\mathbf{x},\omega) = S_{\text{in}}(\mathbf{x},\omega) + S_{\text{scat}}(\mathbf{x},\omega), \tag{1}$$

where $S_{\text{in}}(\mathbf{x},\omega)$ and $S_{\text{scat}}(\mathbf{x},\omega)$ denote the incident sound field by the secondary sources and the scattered sound field by the head, respectively, and ω denotes angular frequency.

The incident sound field from a point source positioned at $\mathbf{x}_l$ is described using the spherical harmonics by

$$S_{\text{in}}(\mathbf{x},\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \breve{S}_n^m(\mathbf{x}_l,\omega) j_n(kr) Y_n^m(\beta,\alpha), \tag{2}$$

$$\breve{S}_n^m(\mathbf{x}_l,\omega) = (-i)kh_n^{(2)}(kr_l)Y_n^{-m}(\beta_l,\alpha_l), \tag{3}$$

where, $j_n(\cdot)$ n-th order spherical Bessel function, k the wave number, $Y_n^m(\cdot)$ spherical harmonic of n-th degree and m-th order, i the imaginary unit, and $h_n^{(2)}$ n-th order spherical Hankel function of second kind.

In the sound field synthesis using multiple secondary sources, the incident sound field $S_{\text{in}}(\mathbf{x},\omega)$, or synthesized sound field with no head's reflection, is described by the summation of the sound fields radiated from the secondary sources. Thus, the incident sound field to be synthesized $S_{\text{in}}(\mathbf{x},\omega)$ is obtained as follows,

$$S_{\text{in}}(\mathbf{x},\omega)=\sum_{l=1}^{L}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}D_l(\omega)\breve{S}_n^m(\mathbf{x}_l,\omega)j_n(kr)Y_n^m(\beta,\alpha), \tag{4}$$

where $D_l(\omega)$ denotes the driving function of the l-th secondary source positioned at $\mathbf{x}_l\,(l=1,\dots,L)$. The driving function $D_l(\omega)$ of the secondary source is derived by the conventional SFS methods, such as WFS, HOA, and LWFS-VSS [4, 5, 9].

Similarly, the sound reflection by a rigid sphere of radius a corresponding to a point source positioned at $\mathbf{x}_l$ is described by

$$S_{\text{scat}}(\mathbf{x},\omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\hat{S}_n^m(\mathbf{x}_l)h_n^{(2)}(kr)Y_n^m(\beta,\alpha), \tag{5}$$

$$\hat{S}_n^m(\mathbf{x}_l,\omega)=-\frac{j_n'(ka)}{h_n^{(2)'}(ka)}\breve{S}_n^m(\mathbf{x}_l,\omega), \tag{6}$$

where the prime denotes differentiation with respect to the argument [2, 13]. Thus, in SFS, the scattered sound field $S_{\text{scat}}(\mathbf{x},\omega)$ by the rigid sphere is represented as follows,

$$S_{\text{scat}}(\mathbf{x},\omega)=\sum_{l=1}^{L}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}D_l(\omega)\hat{S}_n^m(\mathbf{x}_l)h_n^{(2)}(kr)Y_n^m(\beta,\alpha). \tag{7}$$

3 Simulation experiments and results

3.1 Conditions

To investigate how a listener's head distorts the synthesized sound field depending on the frequency and distance of the sound source, the following two simulation experiments were conducted:

Expt. 1 Comparison of reproduction accuracies based on the distribution of errors in each frequency.

Expt. 2 Comparison of reproduction accuracies by changing the distance of the desired point source.

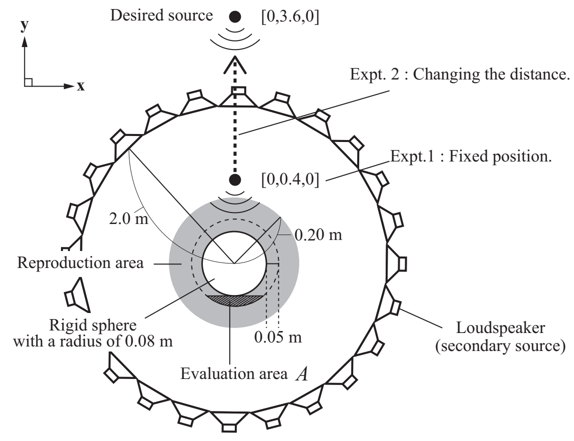

Fig. 1: Arrangement of experiment. The local reproduction area was colored gray. The rigid sphere and secondary source array were centered around the coordinate origin. In experiment 1, the desired source was fixed at the position of $(0,0.4,0)$. In experiment 2, the position of the desired source is changed from $(0,0.4,0)$ to $(0,3.6,0)$ by 0.2 m.

The local sound field was synthesized by two methods: LWFS-VSS and the PM methods. In LWFS-VSS, driving functions for virtual secondary sources were derived by near field compensation higher order ambisonics (NFC–HOA) [14]. In the PM method, driving functions of secondary sources were derived by the inverse of transfer function matrix between loudspeakers and control points based on the regularized least squares method, with the penalty parameter 0.01 [15].

Figure 1 and Table 1 show the simulation conditions of both synthesized methods. A circular secondary source array with a radius of $R_{\text{ss}}=2.0$ was used to synthesize the sound field. The synthesized local area was within a circle with a radius of 0.20 m at the origin. The intervals of virtual secondary sources and matching points were 0.01 m in length. Thus, the numbers of VSSs and MPs were $N_{\text{vss}}=126$ and $N_{\text{mp}}=1000$, respectively.

The reflection of listener's head is modeled by a rigid sphere with a radius of 0.08 m. The center of rigid sphere was located at the origin of the coordinate system. For simulation, the spherical harmonic expansion in Eq.(4) and (7) must be truncated, and the truncation order decides the radius of region to be simulated[13, 2]. In this experiments, we use 80-th

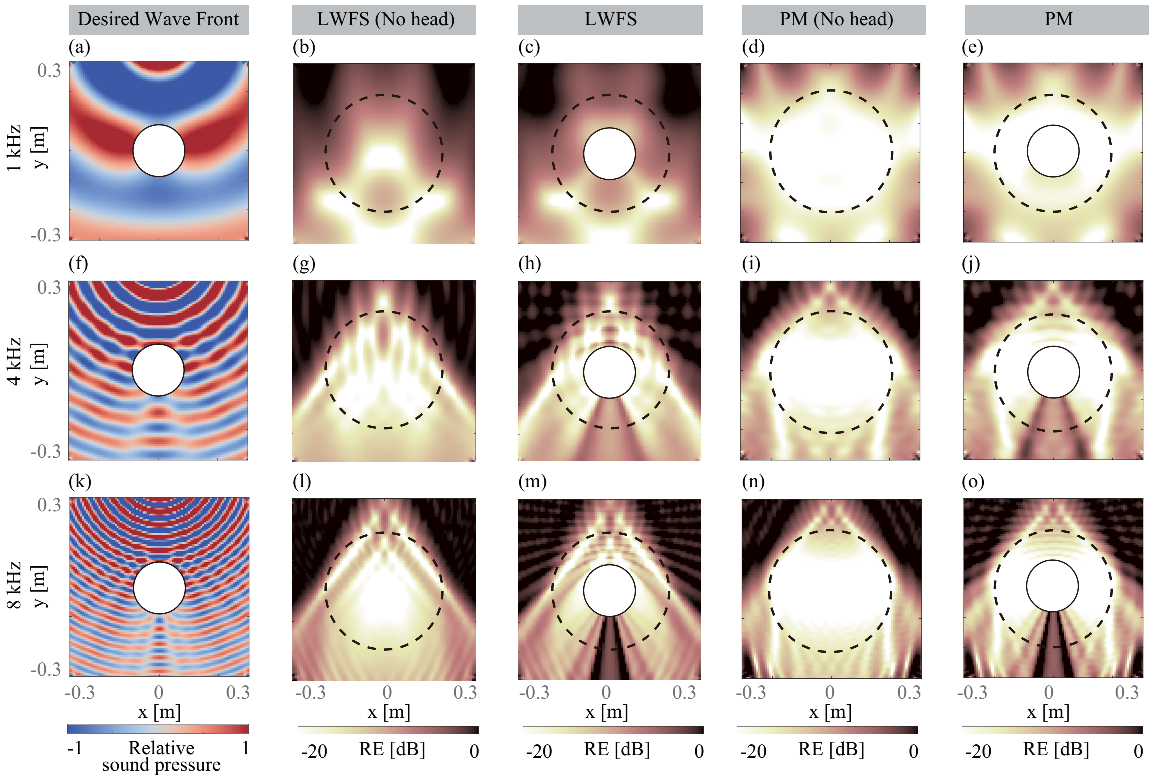

Fig. 2: Distributions of relative sound pressures and reproduction error (RE) at 1, 4, and 8 kHz by using LWFS-VSS and PM methods. (a), (f), and (k) show the desired sound fields; (b), (c), (g), (h), (l), and (m) show REs by LWFS-VSS without and with a rigid sphere; (d), (e), (i), (j), (n), and (o) show those by the PM method. The black dashed circle indicate reproduction area. The desired source was positioned at $(0, 0.4, 0)$. The center of the rigid sphere was position at $(0, 0, 0)$. The radius of the virtual secondary source array is $R_{\text{vss}} = 0.2$ m. The radius of the reproduction area for the PM method is $R_{\text{pm}} = 0.2$ m. Here, RE, defined by Eq.(8), was calculated at each point.

truncation order, which is large enough to simulate the sound field around the rigid sphere up to 8 kHz.

To evaluate the reproduction accuracy, the reproduction error (RE) was defined as follows:

$$\text{RE}(\omega) = 10\log_{10} \frac{\sum_{e=1}^{N_{\text{ep}}} |S(e,\omega) - \hat{S}(e,\omega)|^2}{\sum_{e=1}^{N_{\text{ep}}} |S(e,\omega)|^2}, \quad (8)$$

where S is the complex sound pressure of desired sound field, $\hat{S}$ the complex sound pressure of reproduced sound field with the reflection of the rigid sphere, and the number of evaluation points was $N_{\text{ep}} = 448$. The values of sound pressures were calibrated using the desired sound pressure at the origin, and the RE was evaluated in the X–Y plane. The simulation experiments were conducted by using SFS toolbox (ver. 2.4.2) [16] in MATLAB.

3.2 Degradation of reproduced sound field by a listener's head at higher frequencies (Experiment 1)

To evaluate the dependency of degradation on the frequency, the distributions of REs were compared at multiple frequencies. The desired sound field was a spherical wave radiated from a source at position $(0, 0.4, 0)$. Figure 2 shows the distributions of relative sound pressures and the REs for the sound field scattered by the rigid sphere at 1, 4, and 8 kHz.

As shown in Fig. 2 (b)–(e), there was no significant degradation on RE caused by reflection by the head at

Table 1: Simulation conditions for LWFS-VSS and PM method.

LWFS-VSS	Number of Secondary Sources	84
	Radius of Secondary Source array [m]	2.0
	Number of Virtual Secondary Sources	126
	Radius of Virtual Secondary Source array [m]	0.2
PM	Number of Matching Points	1000
	Radius of reproduction area [m]	0.2

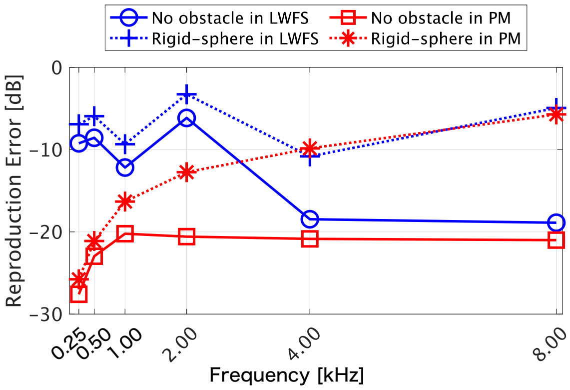

Fig. 3: REs at 0.25, 0.5, 1, 2, 4, and 8 kHz. The spherical wave emitted from $(0,0.4,0)$ was synthesized. The evaluation area A is shown in Fig. 1. Blue and Red lines show REs by LWFS-VSS and PM methods, respectively.

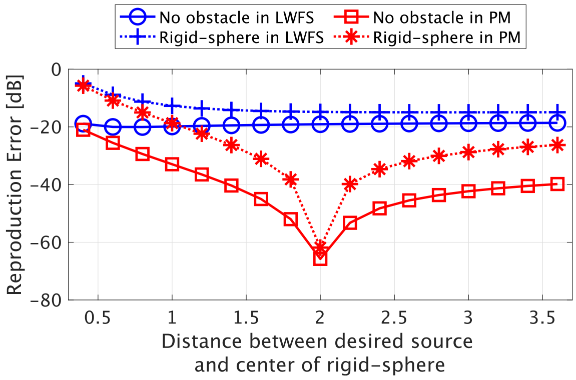

Fig. 4: REs at 8 kHz when the position of the virtual source was changed from $(0,0.4,0)$ to $(0,3.6,0)$ by 0.2 m. Blue and red lines show REs with LWFS-VSS and PM methods, respectively.

1 kHz. However, the overall REs with LWFS-VSS at 1 kHz were larger than those in the PM method. This is because the relationship between reproduction accuracy and frequency depends on the size of the reproduction area in the LWFS-VSS[17]. Thus, when the size of area is small, reproduction accuracy at lower frequency is degraded regardless of the head's scattering in the LWFS-VSS.

When the frequency reaches 4 kHz in Figs. 2 (g)–(j), the REs of LWFS-VSS and PM method at the opposite side to the desired source were degraded by approximately 8 and 15 dB, respectively, compared with REs with no sphere.

As shown in Fig. 2(l)–(o), the REs in the same side were degraded more to approximately 0–5 dB at 8 kHz. Therefore, in both the reproduction methods the REs were more degraded as the frequency increased, especially on the opposite side to the desired source.

Ahrens *et al.* [8] reported that when a plane wave is synthesized in 2.5-dimensional conditions, the distortion of the sound field due to the scattering object results in the occurrence of shadowing in the synthesized sound field. Our simulation results showed the similar shadowing. As illustrated in Fig. 2, it was revealed that the degradation of the synthesized sound field due to the scattering object becomes narrower and worse at higher frequencies. These results indicate that LSFS can synthesize the sound field with high accuracy even at higher frequencies; however, the reproduction accuracy is significantly degraded on the opposite side to the desired source owing to the reflection by the head.

To calculate the degree of degradation, the REs of the evaluation area A, depicted in Fig. 1 which were significantly distorted, were compared at several frequencies by using Eq.(8). Fig. 3 shows the results of REs at 0.25, 0.5, 1, 2, 4 and 8 kHz.

As shown in Fig. 3, at 0.25, 0.5, and 1 kHz in both of the LWFS-VSS and the PM methods, the differences of REs between a head and no head condition were less than 5 dB. However, in the PM method, as the frequency became higher than 2 kHz, the REs were increasingly degraded owing to scattering by the rigid sphere. In LWFS-VSS, the similar degradation was observed at over 4 kHz. It is difficult to synthesize the sound field at frequencies lower than 2 kHz in LWFS-VSS with a radius of 0.2 m, because the reproduction accuracies depend on the size of the listening area [17]. At 8 kHz, in the both methods, REs were degraded by approximately 15 dB owing to the head's scattering.

Therefore, at higher frequencies over 2 kHz, the synthesized sound fields at the opposite side of the desired source have non-negligible degradation due to the reflection by the head. Therefore, when the desired source is on the axis passing through both ears, this degradation in reproduction accuracy may lead to distortions in the location and timbre of the synthesized sound source.

3.3 Distortion of reproduced sound field when varying the distance of the desired source (Experiment 2)

In this experiment the effect the distance of the desired source on the reproduction accuracy was investigated by comparing REs on the opposite side to the desired source. The simulation conditions are depicted in Fig. 1. Figure 4 shows the REs of evaluation area A while changing the desired source position from $(0,0.4,0)$ to $(0,3.6,0)$ by 0.2 m along the positive y-axis.

Comparing REs of no obstacle and rigid-sphere condition in Fig. 4, the degradation of reproduction accuracies with LWFS-VSS were largest—approximately 15 dB when the desired source was closest to the head. In LWFS-VSS, as the distance of the desired source increased, the degradation due to reflection by the head gradually decreased. When the distance of the desired source was within 1.0 m, the degradation due to the reflection gradually decreased from approximately 15 dB to 5 dB. When the distance of the desired source was more than 2.0 m, the degradation by the reflection became less than 5 dB in LWFS-VSS. In other words, as the desired wave front approached a plane wave, by having the desired source far from the head, the degradation became smaller and stable in LWFS-VSS.

In the PM method, when the desired source was closest, the degradation by the reflection was approximately 15 dB, which was nearly the same as in the LWFS-VSS method. When the distance of the desired source was smaller than 2.0 m, the accuracy of reproduction improved as the source moved away from the head. However, the degradation caused by the reflection by the head did not significantly change with the distance of the desired source. Thus, the degradation of REs were approximately 15 dB excluding at $(0,2.0,0)$.

At $(0,2.0,0)$ of the desired source position, the REs became extremely small, and the degradation of the REs due to the reflections by the head became less than 5 dB. This is because the distance of the desired source from the origin match that of the secondary sources. Then, when the distance of the desired source was more than 2.0 m, REs increased again, and the degradation was similarly constant at approximately 15 dB.

In the PM method, degradation of reproduction accuracies was caused by the difference between the transfer functions of secondary sources with and without reflections by the head. Thus, the REs did not depend on the distance of the desired source because the difference of the transfer functions from secondary sources to the control points were not changed by the position of the desired source.

In the previous research by Ahrens *et al.* [8], the degradation by the listener's head was discussed about a plane wave. Within this experiment, the distance of the desired source affected the reproduction accuracy only in the LWFS-VSS method. Thus, to keep the reproduction accuracy higher in the LWFS-VSS method, the desired source must be kept far from the listener's position especially when the source is located at the axis of both ears. In the PM method, the degradation of reproduction accuracy was constantly more than 10 dB regardless of the position of the desired source. Thus, the distance of the desired source is not a factor in maintaining a high reproduction accuracy.

4 Summary

In this study, we investigated the degradation of reproduction accuracy in LSFS by a listener's head located inside the reproduction area. From the simulation experiments involving a rigid sphere, it was revealed that reproduction accuracy was significantly degraded in the higher frequencies at the side opposite of the desired source position. As for the distance of the desired source, the reproduction accuracy in LWFS-VSS is increasingly degraded by the reflection by the head as the desired source approaches the head. In the PM method, the degradation of reproduction accuracy by the listener's head was constant at approximately 10 dB, regardless of the distance of the desired source.

Therefore, in local SFS, the following attentions should be considered when making sound field reproduction contents: The desired sound source should be located away from the vicinity of the listener's ears. Alternatively, if the desired sound source is to be placed near the ear, it is necessary to compensate for head reflections. For example, the transfer functions between secondary sources and control points are measured, including the head reflections. In future studies, we will conduct evaluation experiments to investigate the degradation of auditory perception by head reflection.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number 17KT0142 and 22K12099.

References

[1] Niwa, K., Koizumi, Y., Kobayashi, K., and Uematsu, H., "Binaural sound generation corresponding to omnidirectional video view using angular region-wise source enhancement," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2852–2856, 2016.

[2] Ahrens, J., *Analytic Methods of Sound Field Synthesis*, Springer-Verlag Berlin Heidelberg, 2012.

[3] Hamasaki, K., Nishiguchi, T., Okumura, R., Nakayama, Y., and Ando, A., "A 22.2 Multichannel Sound System for Ultrahigh-Definition TV (UHDTV)," *SMPTE Motion Imaging Journal*, 117(3), pp. 40–49, 2008.

[4] Berkhout, A. J., "Holographic approach to acoustic control," *AES: Journal of the Audio Engineering Society*, 36(12), pp. 977–995, 1988.

[5] Daniel, J., Moreau, S., and Nicol, R., "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," *Audio Engineering Society Convention 114*, 2003.

[6] Miyoshi, M. and Kaneda, Y., "Inverse Filtering of Room Acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(2), pp. 145–152, 1988.

[7] Shiro, I., "A principle of active control of sound based on the Kirchhoff-Helmholtz integral equation and the inverse system theory," *Journal of the Acoustical Society of Japan*, 53(9), pp. 706–713, 1997.

[8] Ahrens, J. and Spors, S., "On the Scattering of Synthetic Sound Fields," in *Audio Engineering Society Convention 130*, 2011.

[9] Spors, S. and Ahrens, J., "Local Sound Field Synthesis by Virtual Secondary Sources," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, 2010.

[10] Hahn, N., Winter, F., and Spors, S., "Local Wave Field Synthesis by Spatial Band-Limitation in the Circular/Spherical Harmonics Domain," in *Audio Engineering Society Convention 140*, 2016.

[11] Tsunokuni, I., Kurokawa, K., and Ikeda, Y., "Pressure-matching-based 2D sound field synthesis with equivalent source array," in *Proceedings of the 23rd International Congress on Acoustics*, 2019.

[12] Tsunokuni, I., Kurokawa, K., Matsuhashi, H., Ikeda, Y., and Osaka, N., "Spatial extrapolation of early room impulse responses in local area using sparse equivalent sources and image source method," *Applied Acoustics*, 179, p. 108027, 2021, ISSN 0003-682X.

[13] Gumerov, N. A. and Duraiswami, R., *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions.*, Elsevier Science, 2005.

[14] Daniel, J., "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format," in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, 2003.

[15] Poletti, M. A., "Improved Methods for Generating Focused Sources Using Circular Arrays," in *Audio Engineering Society Convention 133*, 2012.

[16] Wierstorf, H. and Spors, S., "Sound Field Synthesis Toolbox," in *132nd Convention of the Audio Engineering Society*, 2012.

[17] Daniel, J., Moreau, S., and Nicol, R., "Further Investigations of High-order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," *Journal of the audio engineering society*, 2003.