NAIST-IS-DD0961010

Doctoral Dissertation

Sound Field Reproduction Integrating Multi-Point Sound Field Control and Wave Field Synthesis

Noriyoshi Kamado

February 2, 2012

Department of Information Processing
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Noriyoshi Kamado

Thesis Committee:

Professor Kiyohiro Shikano	(Supervisor)
Professor Kazushi Ikeda	(Co-supervisor)
Associate Professor Hiroshi Saruwatari	(Co-supervisor)

Sound Field Reproduction Integrating Multi-Point Sound Field Control and Wave Field Synthesis[†]

Noriyoshi Kamado

Abstract

Multichannel sound field reproduction (SFR) methods are promising for improving the quality and interactivity of acoustic telecommunication experience. In this dissertation, robust and high-quality SFR is addressed with a new proposed method integrating of multi-point sound field control and wave field synthesis (WFS). In SFR, the competing requirements of a wide listening area and a high reproduction accuracy arise. In this dissertation, a new method of meeting these requirements sis proposed: the null space of a generalized inverse matrix is dedicated to a compensation filter of a wave field outside the control points by imposing WFS into multi-point sound field control. The results of the experiments show that the proposed method can meet these competing requirements under realistic acoustical conditions. In applying the proposed method to the reproduction of commercially available sound contents, the localization information of each sound source is required as a cue of the primary sources generated in WFS. Therefore, in this dissertation, a new sound field coding method and an interactive controller for audio object localization are proposed. The developed controller enables a listener to intuitively operate displayed audio objects with a touch pen. The results of the experiments clarify that the controller enables the listener to change the localization of audio objects without sound degradation if the gain operation is not extreme. From these findings, we can conclude that these proposed methods have wide applicability in an actual environment.

[†]Doctoral Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0961010, February 2, 2012.

Keywords:

Sound field reproduction, Sound field coding, Multi-point controlled inverse filtering, Wave field synthesis, Generalized singular value decomposition

Contents

Chapter 4
Evaluation in actual environment of DMCWS

Chapter 5
Hybrid wave field synthesis

List of Figures

List of Tables

CHAPTER 1

PROLOGUE

1.1. Background

The application of electronic communication over large distances is a rapidly evolving technological field, and these applications are already a natural part of our everyday lives. In the field of acoustical digital communications, these technologies, particularly the application of internet protocol phones or smart mobile phones, provide a wide range of usability. The ultimate aim of these applications is to perfectly convey the characteristics of natural hearing over the entire spatial and frequency domains from one side to another. However, the present communication experiences provided by the above technologies do not satisfy the required naturalness compared with that of a real face-to-face conversation because recent technological developments have focused on only sound quality.

One of the important reasons for the degeneracy of naturalness is that the technologies mentioned above pay no particular attention to the spatial localization information of sound sources during both recording and playback. This defective information becomes more apparent in simultaneous communication with two or more people on the other side, e.g., online conference. Multi-channel sound field coding and multichannel sound field reproduction are prospective candidate solutions to these problems, they use a large number of microphones and loudspeakers to improve the quality of these experiences.

1.2. Multichannel sound field reproduction

1.2.1 Stereophonic-based system

Until 1930, sounds were reproduced in a monaural manner. The sound field reproduction system was first introduced and developed by Blumlein [1] as a two-channel stereo system, so-called the two-channel "*Stereophonic*". The typ-

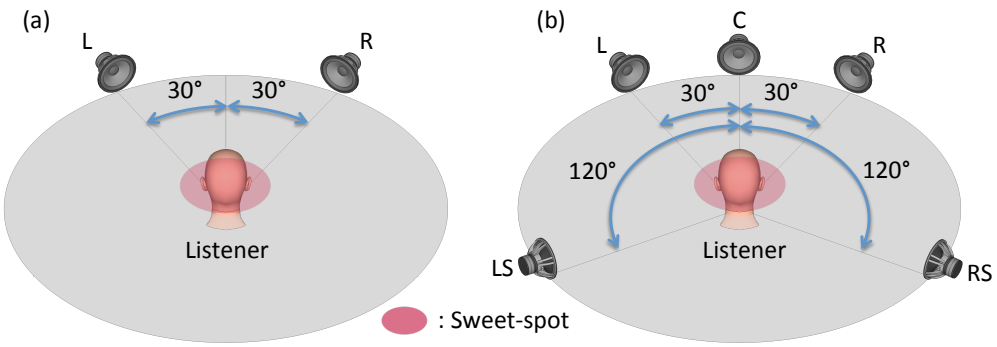


Figure 1. Typical loudspeaker setups and sweet-spots of (a) two-channel stereo system and (b) 5.1-channel surround sound system.

ical loudspeaker setup of the Stereophonic is shown in Fig. 1 (a). With this stereophonic-based system, the location of each recorded original sound source is perceived in different directions of the real playback loudspeakers near the original direction of the recorded sound source. The recorded original sound source is called the "*primary source*" and the reproduced perceived sound image of the primary source is called the "*virtual source*." Also, the reproducible ability of the virtual source is called "*sound localization*." Therefore, the possibility of the reproducible direction of virtual sources was extended to some extent in comparison with previous monaural systems. Details of the histories of the two-channel Stereophonic are described in [2].

The reproducible directions of virtual sources are restricted within the area of the baseline of two-channel loudspeakers. Therefore, after the development of a two-channel stereophonic system, this system was then extended to a four-channel *Quadraphonic* system to enable full-circle spatial reproduction [3]. The loudspeakers are located at the four corners of the listening space in the Quadraphonic system. However, the sound localization performance was poor in the front half-plane, the formats were more expensive to produce than those of the standard two-channel stereo, and the playback required additional loudspeakers and custom-designed decoders and amplifiers. Hence, the Quadraphonic was quickly abandoned.

The "*surround system*" using a larger number of channels, such as 5.1 and 7.1, was proposed to compensate for the Quadraphonic sound localization [4, 5], so-called the "*multichannel stereophonic*." The typical loudspeaker setup of the 5.1 surround system is shown in Fig. 1 (b). The first surround sound system was provided by Disney as the discrete six-channel system so-called "*Fantasound*" in "*Fantasia*" in 1941 [6]. The direction with the poorest sound quality in the Quadraphonic is the frontal localization direction. Therefore, the surround system uses more loudspeakers in front of the listener to improve frontal localization. In addition, loudspeakers on the lateral side (which are only used in the 7.1ch system) and behind the listening position are used for providing ambience and side or rear sound localization. The red area located at the center of the circle of the listening area, shown in Fig. 1, is called the "*sweet-spot*." The reproduced sound field provides correct localization and ambience effects only at this small sweet-spot position and the area outside of the sweet-spot cannot be controlled using a stereophonic-based system.

To summarize the above-mentioned facts, all stereophonic-based multichannel sound field reproduction systems generally suffer from the same problems as follows.

1. The system suffers from having a very small sweet-spot.

2. The system can only reproduce virtual sources between the loudspeaker positions or farther, but not at the gap between the listener and the loudspeakers, i.e., the perceptual distance control of the virtual source.

1.2.2 Acoustic window and wave field synthesis

In parallel with the development of multichannel stereophonic systems, extensive investigations of the above-mentioned problems have been conducted. The first successful approach to extending the sweet-spot was introduced by Snow in 1934 [7] and the proposed system recorded sound sources by nine microphones and played the signals through nine similarly located loudspeakers in another remote room in real time. Interestingly, the best experimental results were provided by

connecting each of the microphones to the corresponding loudspeakers directly, and the virtual source localization almost matches the localization of the primary source. The behavior of this reproduction system seems to be like a real mechanical window between two experimental rooms. Therefore, we refer to this mechanism as "*virtual acoustic opening*," and the reproduction system was called the "*acoustic window*" between two remote rooms. The sound-emitting room that generates the sound field is called the "*primary sound field*," the sound-receiving room that receives the reproduced sound field is called the "*secondary sound field*," and the reproduction loudspeakers are called the "*secondary sources*." Figure 2 shows an overview of the concept of the acoustic window [8]. Unfortunately, the concept of the acoustic window was soon abandoned owing to the enormous bandwidth necessary to send a large number of audio signals in the 1930s. In recent years, however, with the advent of powerful multichannel perceptual coders (like MPEG4), this scheme has become much more feasible and the acoustic opening concept has been reviewed [9, 10, 11].

On another front, the concept of the acoustic window seems to be like the "*Huygens principle* [12]." The Huygens principle states that the propagation of a wave through a medium can be formulated by summing up the contributions of all of the secondary sources positioned along a wavefront. Figure 3 shows an overview of the Huygens principle [13]. In Fig. 3, the primary source generates a wavefront that propagates along the surface, which is called the "*primary wavefront*." The Huygens principle indicates that a primary wavefront can be generated by an infinite number of secondary sources along any position defined by the passage of the primary wavefront, which is called the "*secondary wavefront*." This secondary wavefront is perfectly accurate outside the field delimited by the secondary source distribution. The secondary sources therefore serve to relay sound propagation, and can reproduce the primary wavefront in the absence of the primary source as well as the acoustic window concept.

The mathematical descriptions of the Huygens principle can be found through the Kirchhoff-Helmholtz integral. *Wave field synthesis* (WFS) [14, 15, 16] is premised on an anechoic reproduction environment and offers a large listening area with a flexible high perceptual reproduction quality for multiple listeners

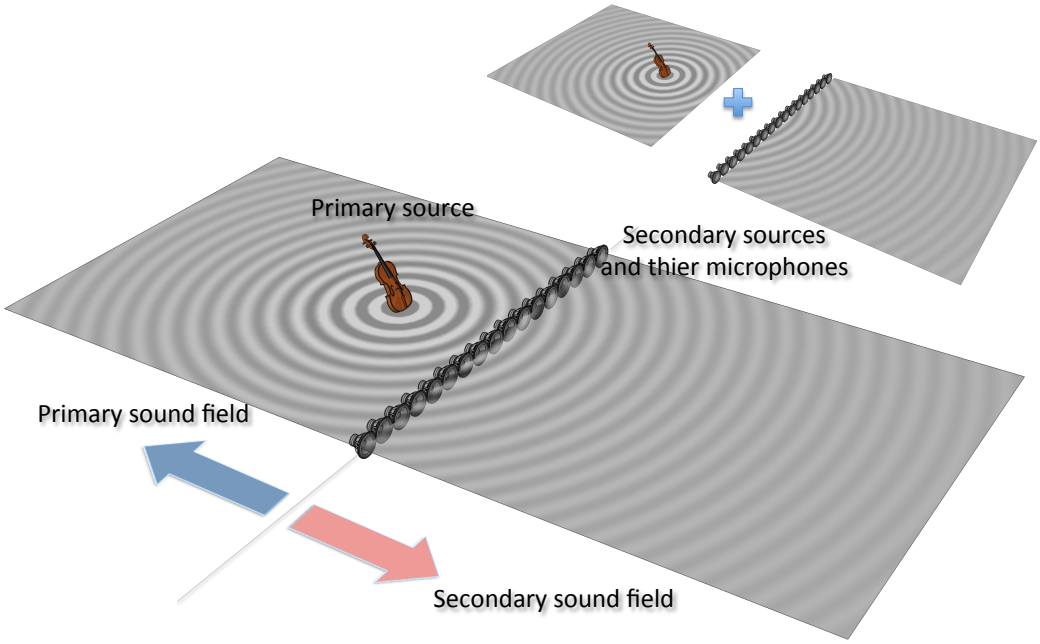


Figure 2. Concept of acoustic window.

based on the Kirchhoff-Helmholtz integral. However, its reproduction accuracy decreases in practice owing to the inherent disadvantages of the physical inaccuracies found in the Kirchhoff-Helmholtz integral and room reverberation [17] in WFS. Therefore, WFS cannot accurately reproduce the sound field at the sweet-spot.

1.2.3 Multi-point controlled sound field reproduction

In parallel with the development of WFS, in recent years, research interest in multi-point controlled robust sound field reproduction methods operated in a reverberant environment has been increasing. These methods reproduce the multi-point sound pressure at the observation microphones in the secondary sound field and that of wavefronts. The locations of the microphones for observing the controlled pressure are called "*control-points*." Many of the systems based on multi-point controlled sound field reproduction were developed using inverse filtering.

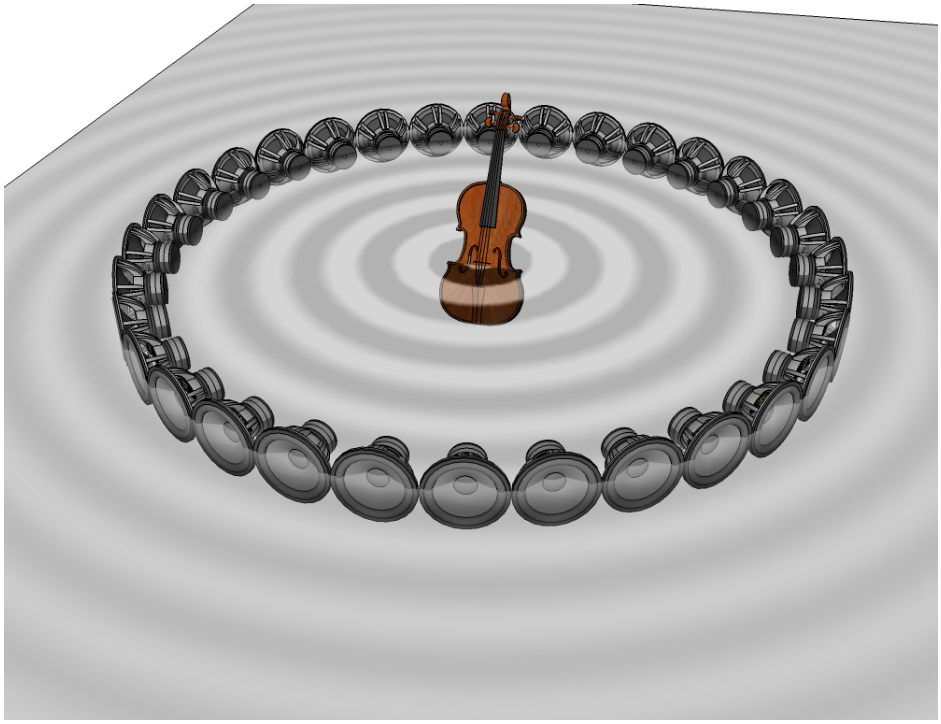


Figure 3. Huygens principle: wavefronts emitted by the primary source (violin) behave as an ensemble of secondary sources located along an equiphase surface of the primary source.

The *transaural system* [18, 19] based on the multiple input/output inverse theorem (MINT) [20] is a typical system aimed at reproducing binaural recording to achieve realistic sensation, where an inverse filter compensates for transfer functions at the user's ears including room reverberation. However, outside the control points (sweet-spot), the inverse filter does not compensate for transfer functions satisfactorily, and a method based on the inverse filtering of multi-point controlled reproduction does not work well with the user's movements. *Multi-point controlled wavefront synthesis* (MCWS) [21, 22, 23, 24, 25] is a technique of solving this problem using a number of control points located in the listening area. Although the theory of the WFS has been well studied, the optimal control-point geometry and the behavior of the synthesized secondary wavefront within and above the frequency band-limit in MCWS have not been investigated so far.

In MCWS, ensuring a large sweet-spot by placing a sensor array on the control points fills the reproduction space with a large number of sensors, which is impractical. With the practical placement of the control points, the potential problem in the reproduction area persists in MCWS because the wavefront formed outside the control points is entirely different from that formed in the primary sound field, and sound localization degrades considerably when the user moves from the controlled area. To mitigate the effect of the listener's movement in such multi-point control-based methods, an improved inverse filter design has been proposed [26, 27]. This method can maintain the perceptual direction of sound outside the sweet-spot, while keeping the accuracy at the control points by substituting arbitrary subspace components of MINT with a simple emphasis filter that drives the nearest loudspeaker to the primary source. However, this emphasis filter cannot compensate for the perceptual distance, and the reproducible position of the primary source becomes circumscribed only at the loudspeaker position. Therefore, this framework still requires further improvement in accuracy as well as a flexible method of reproducing the arbitrary primary wavefront regardless of the loudspeaker position, a problem mentioned in the previous stereophonic section.

Through recent studies, typical multi-point controlled sound field reproduction methods are classified into two groups depending on whether the methods use a null-space of the spatial impedance matrix of the secondary sound field to solve the filter of the robust sound field reproduction.

1.3. Multichannel sound field coding and interactive sound field control

On another front, the increase in the number of audio channels additionally increases the amount of audio data to be transferred by such an acoustic window system between two remote locations. Sending a huge amount of audio data from tens of microphones over the same connection that is normally used for only a couple of channels becomes impossible. The problem has been investigated, and coding methods of audio transfer have been proposed [9]. In addition, in apply-

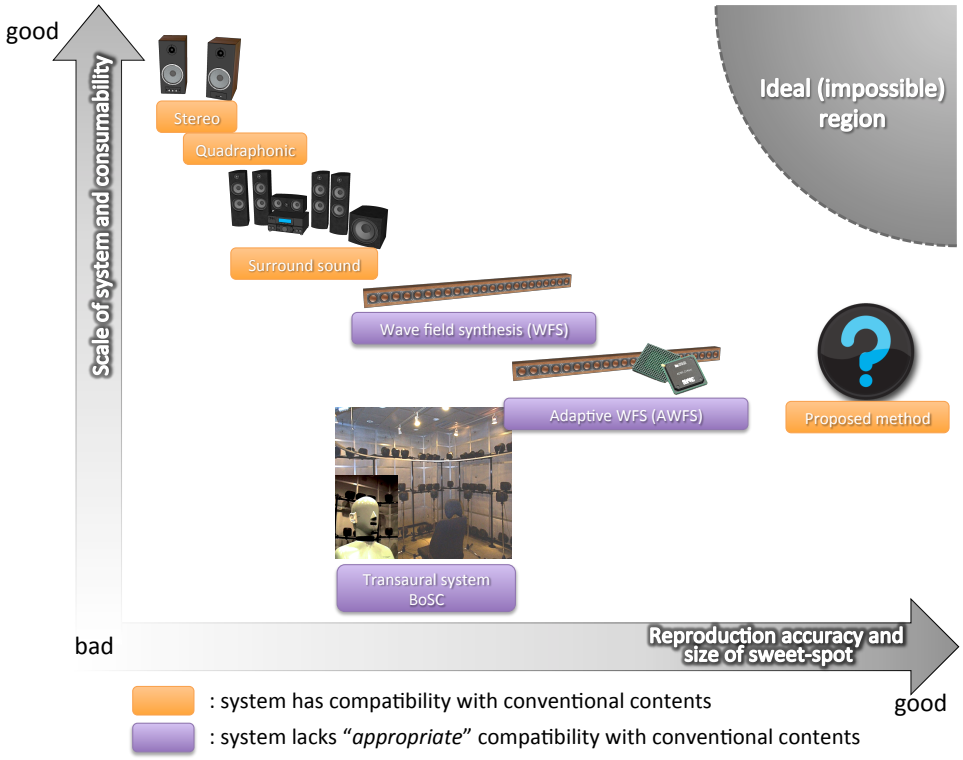


Figure 4. Orientation of conventional methods and target region of proposed method from the perspective of triaxial scale. 1. Scale and censurability of system. 2. Reproduction accuracy and size of sweet-spot. 3. Compatibility with conventional contents. BoSC and AWFS will be discussed later.

ing WFS to commercially available sound contents, the localization information of each sound source is required as a cue of the primary sources generated in WFS. This means that a method of analyzing and decomposing the sound sources is indispensable for the total system of multichannel sound field reproduction. Alternatively, the reproduction of music with a high sense of reality has been superceded by the development of sound reproduction systems with a *creative sense of reality*, in which the user can operate each sound object individually in audio signals [28]. As a system for achieving the above-mentioned problems related to sound field coding and a *creative sense of reality*, MPEG Spatial Audio Object Coding (MPEG-SAOC) is being devised by the Moving Picture Experts Group

(MPEG) [29, 30, 31, 32]. MPEG-SAOC basically requires separated audio object sources. Such sources are, however, difficult to obtain for common-end users because separated sources are not generally supplied via normal music distribution media such as compact discs, in which all sources are mixed. Thus, such a codec is difficult to use.

1.4. Objective and Overviews of this dissertation

Figure 4 shows the orientation of conventional methods summarized in the previous sections and the target region of the proposed method from the perspective of triaxial scale. My aim in this dissertation is to extend the current multichannel sound field coding and multichannel sound field reproduction frameworks to a new system with the flexibility of user experiments, as shown in Fig. 4. In particular, I have addressed the following issues.

1. To evaluate the performance of sound reproduction systems, a real-apparatus-based sound measurement system is developed in a real acoustical environment.

2. To achieve more accurate reproduction, I investigate the optimization of the MCWS, and perform a comparative study of MCWS and WFS.

3. A new hybrid sound reproduction method combining MCWS and null-space-imposed WFS is proposed for achieving both requirements of a high accuracy in the sweet-spot and a large listening area.

4. In applying WFS to commercially available sound contents, the localization information of each sound source is required as a cue for the primary sources generated in WFS. This means that a method of analyzing and decomposing the sound sources is indispensable for the total system of multichannel sound reproduction.

The dissertation is organized as follows.

First, I describe the MCWS and WFS as basic frameworks of the sound field reproduction in Chapter 2. Next, I investigate the optimal control-point geometry and the behavior of the synthesized secondary wavefront in MCWS, and I evaluate the effectiveness of MCWS through comparison with WFS in Chapter 3. Since the wavefront has the frequency characteristics of actual audio applications, I determine the spatial spectrum characteristics of MCWS in an actual environment. Hence, I derive the spatial spectrum characteristics from the impulse responses at each observation point and measure the MCWS wavefront using a wavefront measurement system [33] in an actual environment in Chapter 4.

In Chapter 5, to mitigate the trade-off problem between the accuracy and reproducible region caused by the practical location of the control-points of the sound field reproduced in both conventional methods, I propose a new robust sound field reproduction method that enables high-accuracy reproduction with a large listening area by integrating our previously proposed multi-point sound field reproduction method and WFS with practical control-point locations. In the proposed method, the wavefront for the desired spatial cue outside the sweet spot is derived by approximating the WFS-synthesized wavefront, but the sweet-spot sound is perfectly preserved. The efficacy of the proposed method is ascertained by objective evaluation through computer-based simulations and wavefront measurement. The mathematical relationship and differences between the conventional multi-point-controlled-based method and the proposed method by a new interpretation, which is given by the generalized singular value decomposition (GSVD) [34, 35, 36, 37] in Chapter 6.

In Chapter 7, I propose a new audio coding framework enabling the localization operation of audio objects (e.g., vocal, guitar and drums) by the temporal quantization of spatial information and introduce an audio object controller developed for interactive sound field reproduction using audio object localization based on spatial representative vector operations. An audio object is represented by the centroid vector of a spatial clustering algorithm, similarly to k-means [38]. Since the angle of the centroid expresses the direction of the spatial image of the audio object, we can operate the localization of audio objects individually by steering the spatial representative centroid vector of interest. In addition, the proposed au-

dio object controller enables the listener to operate an initial sound field and build a new virtual sound field by performing sound field operations on a stereo mixed source. The main contribution of this dissertation is the bridging of the theories of multi-point control-based sound field reproduction and WFS, while achieving accurate sweet-spot reproduction and large-area wavefront generation.

Finally, in Chapter 8, I conclude this thesis and clarify and discuss the remaining problems.

Chapter 2

Basic theory of conventional sound field reproduction

2.1. Introduction

In this chapter, WFS and MCWS are described theoretically and the equations used for numerical calculations are derived in detail. It will be shown that the physical and mathematical descriptions of WFS and MCWS. In addition, secondary source driving functions are derived that allows for a spatial and frequency reconstruction of the primary wave field in the secondary sound field. In this dissertation, these methods are based on linear loudspeaker array.

2.2. WFS

The geometric configuration and parameters in WFS are depicted in Fig. 5, where $S_P(\omega)$ and $S_{Sn}(\omega)$ denote the spectra of the primary and the nth secondary sources, respectively, on the x-y horizontal plane.

The spectrum of the nth secondary source, which synthesizes the primary spherical wavefront, $S_{Sn}(\omega)$ is expressed as [39]

$$\begin{aligned} S_{S_n}(\omega) &= Q_n^{(\mathrm{WFS})}(\omega) S_P(\omega) \\ &= S_P(\omega) \sqrt{\frac{\mathrm{sign}(\zeta)k}{2\pi j}} C(\zeta) \frac{\exp(\mathrm{sign}(\zeta) jkr_{P_n})}{\sqrt{r_{P_n}}} \cos\theta_{P_n} \Delta x, \end{aligned} \tag{1}$$

where $Q_n^{(\mathrm{WFS})}(\omega)$ is the wavefront synthesis filter, j is the imaginary unit, k is the wave number (ω/c), c is the sound velocity, ω is the angular frequency, Δx is the interelement interval among the secondary sources, r_{Pn} is the distance between the primary source and the nth secondary source, and θ_{Pn} is the angle between the y-axis and the line connecting the nth secondary and primary sources. $C(\zeta)$ is a function that compensates for the level of mismatch due to the stationary phase

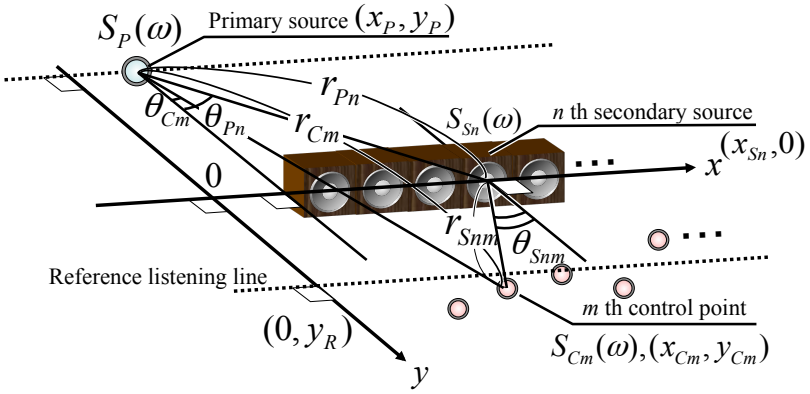


Figure 5. Configuration of WFS and MCWS.

approximation along the x-direction [40], which is a function of only the reference listening distance y_R and is given as

$$C(\zeta) = \sqrt{\frac{\zeta}{\zeta - 1}}, \tag{2}$$

$$\zeta = \frac{y_R}{y_P}. \tag{3}$$

Outside the reference listening line, the level of the sound field is expressed in [41] as

$$Att_{S_S}(y) = \sqrt{\frac{|y_R|}{|y|}} \sqrt{\frac{|y| + |y_P|}{|y_R| + |y_P|}} \frac{1}{|y - y_P|}. \tag{4}$$

2.3. MCWS

The geometric parameters of MCWS are shown in Fig. 5. MCWS controls the spatial spectra at the control points, which are located on the x-y horizontal plane in front of the secondary sources, and generates the desired wavefront. Here, $S_{Cm}(\omega)$ denotes the secondary wavefront spectrum at the mth control-point position. Also, θ_{Cm} is the angle between the y-axis and the line connecting the mth control point and the primary source, θ_{Snm} is the angle between the y-axis and the line connecting the mth control point and the nth secondary source, r_{Cm} is the

spatial distance between the mth control point and the primary source, r_{Snm} is the spatial distance between the mth control point and the nth secondary source, N is the number of secondary sources, and M is the number of control points.

Here, we derive the spectrum of the secondary source $S_{Sn}(\omega)$, which synthesizes the primary spherical wavefront. The transfer function between the nth secondary monopole source and the mth control point, $H_{nm}(\omega)$, is written as

$$Z_{nm}(\omega) = \frac{\exp(-jkr_{Snm})}{r_{Snm}}. \tag{5}$$

From Eq. (5), we define the transfer function matrix

$$\boldsymbol{Z}(\omega) = \begin{bmatrix} Z_{1,1}(\omega) & Z_{2,1}(\omega) & \cdots & Z_{N,1}(\omega) \\ Z_{1,2}(\omega) & Z_{2,2}(\omega) & \cdots & Z_{N,2}(\omega) \\ & \vdots & & \\ Z_{1,M}(\omega) & Z_{2,M}(\omega) & \cdots & Z_{N,M}(\omega) \end{bmatrix}. \tag{6}$$

We write the secondary wavefront spectrum vector at the mth control-point position as

$$\boldsymbol{S_C}(\omega) = \boldsymbol{Z}(\omega)\boldsymbol{S_S}(\omega), \tag{7}$$

where

$$\boldsymbol{S_C}(\omega) = [S_{C1}(\omega), S_{C2}(\omega), \cdots, S_{CM}]^{\mathrm{T}}, \tag{8}$$

$$\boldsymbol{S_S}(\omega) = [S_{S1}(\omega), S_{S2}(\omega), \cdots, S_{SN}]^{\mathrm{T}}, \tag{9}$$

and the superscript T denotes the transpose of the vector or matrix. If the primary wavefront spectrum is equal to the secondary wavefront spectrum at the control-point position, Eq. (7) can be transformed into

$$\boldsymbol{S_C}(\omega) = \boldsymbol{W}(\omega)S_P(\omega), \tag{10}$$

where

$$\boldsymbol{W}(\omega) = \left[\frac{e^{-jkr_{C1}}}{r_{C1}}, \quad \frac{e^{-jkr_{C2}}}{r_{C2}}, \quad \cdots, \quad \frac{e^{-jkr_{CM}}}{r_{CM}} \right]^{\mathrm{T}}. \tag{11}$$

From Eqs. (7) and (10) and the Moore-Penrose (MP) generalized inverse matrix of $\boldsymbol{Z}(\omega)$, $\boldsymbol{Z}^{+}(\omega)$, we obtain the secondary source spectrum vector $\boldsymbol{S_S}(\omega)$ with a wavefront synthesis filter of MCWS $\boldsymbol{Q}^{(\mathrm{MC})}(\omega)$ in the form,

$$\boldsymbol{S_S}(\omega) = \boldsymbol{Q}^{(\mathrm{MC})}(\omega)S_P(\omega) = \boldsymbol{Z}^{+}(\omega)\boldsymbol{W}(\omega)S_P(\omega). \tag{12}$$

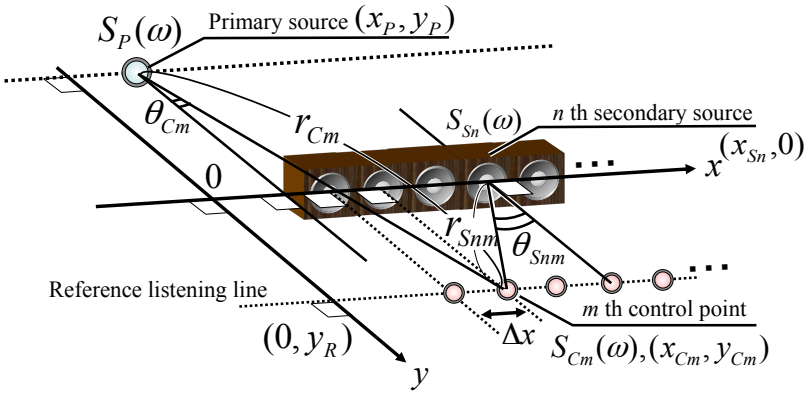


Figure 6. Configuration of DMCWS.

2.4. Directly-aligned MCWS

In MCWS, there is one typical case in which each control point is located on the reference listening line of WFS which is parallel to the x-axis and intersecting the y-axis at y_R (its geometric parameters are depicted in Fig. 6). In this case, the wavefront synthesis method is called DMCWS (Directly-aligned MCWS), named after its control-point geometry.

2.5. Summary of this Chapter

In this chapter, the theory of conventional sound field reproduction methods WFS, MCWS and DMCWS were described with the geometric configuration and parameters.

Chapter 3

Computer-Simulation based verification of DMCWS

3.1. Introduction

The secondary wavefront spectrum vector in DMCWS contains the control line geometry, and its optimal geometry has not been elucidated completely. Hence, I study its geometry through a wavefront calculation in this chapter. In addition, I compare DMCWS and WFS through computer-simulation-based experiments in terms of the amplitude, phase, and attenuation of the synthesized secondary wavefront spectrum.

3.2. Optimized control point geometry of MCWS

3.2.1 Calculation conditions

The conditions of the wavefront calculation are shown in Table 1. The diaphragm radius b and secondary source distance Δx mimic those of the Soundevice SD-0.6 loudspeaker shown in Fig. 7. The spatial aliasing frequency in front of the center of the array of the loudspeakers is given as

$$f_{alias} = \frac{c}{2\Delta x} = \frac{340.64}{2 \times 0.12} \simeq 1417 \ \text{[Hz]}, \tag{13}$$

where Δx denotes the distance between the elements of the secondary loudspeaker array and c denotes the sound velocity. The inter-aural time differences (ITD) which are obtained by evaluated wavefront band frequencies below 1600 Hz are major cues for sound source localization in case of conflicting cues of ITD, inter-aural time differences (ILD), head and torso [42, 43]. The geometric parameters in the wavefront calculation are illustrated in Fig. 8. The height of an evaluating horizontal plane is set to $z_O = 1.22$ m at the ear level of the sitting position for

Table 1. Wavefront calculation conditions

PARAMETER	VALUE
Temperature	15℃
Sound velocity c	340.64 m/s
Evaluated wavefront band frequencies	20–1600 Hz
	(10 Hz interval)
Spatial aliasing frequency	1417 Hz
Primary source geometry (x_P, y_P, z_P)	(1.2, -0.1– -1.0,1.22) m
Secondary source and	
Control point interval Δx	0.12 m
Diaphragm radius b	0.05 m
Number of secondary sources N and	
Control points M	16
y-coordinate of reference listening line y_R	0.1–2.0 m
Height of evaluating horizontal plane z_O	1.22 m
Height of secondary sources	1.22 m

a normal listener. In addition, the number of secondary sources is determined by the hardware limitation. In terms of the control area, the more control points, the better. However, under the terms of $N < M$, the rank-deficient problem occur in the transfer function matrix $\boldsymbol{H}(\omega)$ and the reproduction accuracy becomes diminished. Therefore, we set the number of control points to the maximum value which satisfies $N \geq M = 16$.

3.2.2 Method of calculating secondary wavefront

The secondary source and observation point geometric parameters are shown in Fig. 9. Equation (14) defines $S_O(\omega)$, which denotes the spectrum of the secondary wavefront at the observation point, given as

$$S_O(\omega) = \sum_{n=1}^{N} \left[S_{Sn}(\omega) G(\theta_{On}, \omega) \frac{\exp(-\mathrm{j}kr_{On})}{r_{On}} \right]. \tag{14}$$

Figure 7. Soundevice SD-0.6 loudspeaker, whose diaphragm radius and source distance were assumed in the experiment.

The secondary sources are circular vibration planes on an infinite baffle whose directional characteristic is

$$G(\theta, \omega) = \frac{2J_1(kb \sin\theta)}{kb \sin\theta}, \tag{15}$$

where $J_1(\cdot)$ is the Bessel function of the first kind and b is the diaphragm radius of each circular vibration plane.

3.2.3 Evaluation criterion of secondary wavefront

$E_{wf}(y_P, y_R)$ is the criterion used for evaluating the reproduced wavefront accuracy,

$$\begin{aligned} &E_{wf}(y_P, y_R) \\ &= 10 \log_{10} \frac{\displaystyle\sum_{i,j} \sum_{\omega} \{|\mathrm{SWF}(i, j, \omega)| - |\mathrm{PWF}(i, j, \omega)|\}^2}{\displaystyle\sum_{i,j} \sum_{\omega} |\mathrm{PWF}(i, j, \omega)|^2} \ [\mathrm{dB}], \end{aligned} \tag{16}$$

where $\mathrm{PWF}(i, j, \omega)$ is a function of the primary wavefront spectrum at the observation point (i, j);

$$\mathrm{PWF}(i, j, \omega) = \frac{\exp\left(-\mathrm{j}kr_{O(i,j)}\right)}{r_{O(i,j)}}, \tag{17}$$

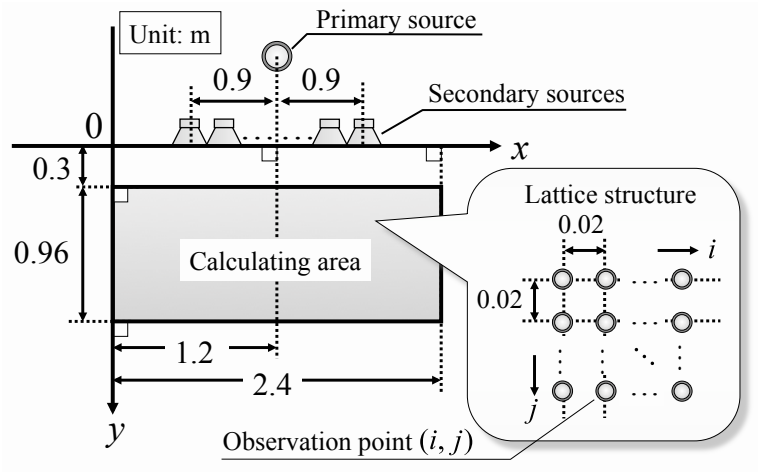


Figure 8. Geometric parameters in wavefront calculation.

where $r_{O(i,j)}$ is the spatial distance between the primary source and the observation point (i, j). SWF(i, j, ω) is a function of the DMCWS secondary wavefront spectrum at the same position. Here, $\sum_{\omega}$ is the summation over ω in the evaluation frequency band, and $\sum_{i,j}$ is the summation over the observation position (i, j). In this paper, we define the optimal control line geometry y_{Ropt} as

$$y_{Ropt}(y_P) = \underset{y_R}{\operatorname{argmin}}\, E_{wf}(y_P, y_R), \tag{18}$$

where $\operatorname{argmin}_{y_R} \cdot$ denotes the minimization function. We calculate the optimal control line geometry y_{Ropt} with numerical simulation.

3.2.4 Calculation results

Figure 10 shows the results of the calculation, where the contour lines show values of $E_{wf}(y_P, y_R)$ with an interval of 2 dB. Figure 11 shows y_P for the optimized y_R, and the corresponding value of $E_{wf}(y_P, y_R)$. Figure 11(a) shows that when y_R is optimized, the y-coordinate y_{Ropt} ranges from 0.6 to 0.7 m for the synthesized secondary wavefront. Also, Fig. 11(b) shows an increase in the evaluation criterion E_{wf} with the primary source y-coordinate y_P, and that the value of y_P which minimize E_{wf} in Fig.11(b) is 0.1 m. Hence, we define the control line coordinate

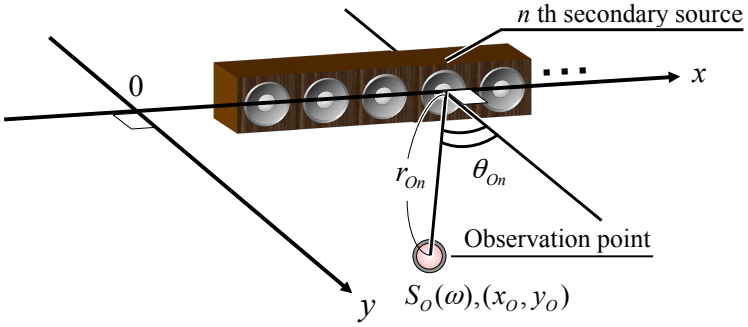


Figure 9. Secondary source and observation point geometric parameters.

.

y_{Ropt} as 0.6 m when the primary y-coordinate y_P is 0.1 m according to the condition shown in Fig. 8. We decided to use these values in the subsequent computer simulations.

3.3. Computer-Simulation based comparison of DM-CWS and WFS

3.3.1 Calculation conditions

The wavefront calculation conditions are listed in Table 2 and the other conditions are the same as those in Table 1 and Fig. 8.

3.3.2 Evaluation criteria of secondary wavefront

The evaluation criteria $E_A(i, j)$ and $E_P(i, j)$ used to evaluate the complex amplitude and phase errors of the secondary wavefront are respectively defined as

$$E_A(i, j) = 10 \log_{10} \frac{\sum_{\omega} \{|\mathrm{WF}(i, j, \omega)| - |\mathrm{PWF}(i, j, \omega)|\}^2}{\sum_{\omega} |\mathrm{PWF}(i, j, \omega)|^2} \ [\mathrm{dB}], \tag{19}$$

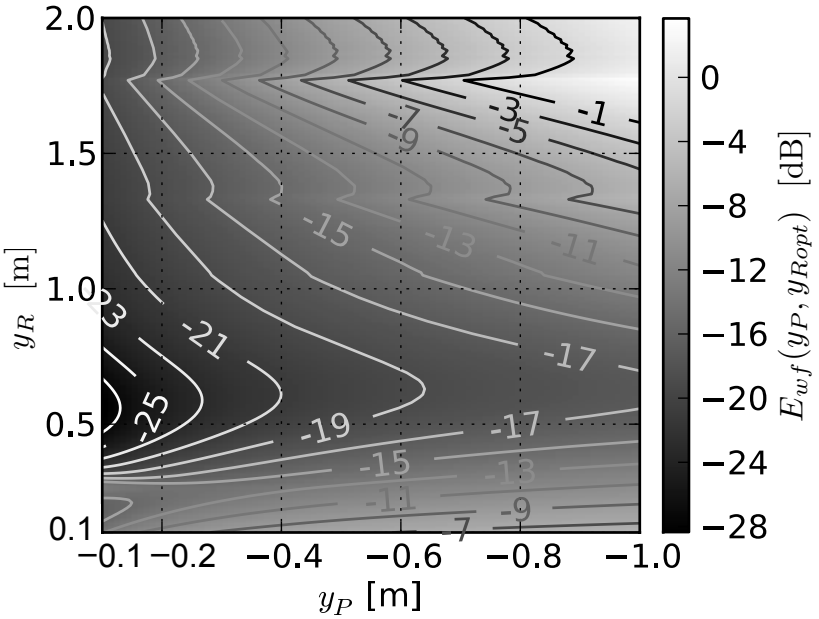


Figure 10. Contour plot of $E_{wf}(y_P, y_R)$.

$$E_P(i,j) = 10\log_{10}\frac{1}{K}\sum_{\omega}\left|\frac{1}{\pi}\arg\left(\frac{\mathrm{PO}(\mathrm{WF}(i,j,\omega))}{\mathrm{PO}(\mathrm{PWF}(i,j,\omega))}\right)\right| \ [\mathrm{dB}], \tag{20}$$

where $\mathrm{WF}(i,j,\omega)$ denotes the secondary wavefront synthesized by DMCWS or WFS, K is the number of frequency bins and $\mathrm{PO}(\cdot)$ denotes the phase function given by

$$\mathrm{PO}(x) = \frac{x}{|x|}, \tag{21}$$

where x is a complex-valued variable.

3.3.3 Calculation results

Figures 12 and 13 show the amplitude error E_A of WFS and DMCWS that of the evaluated wavefront band frequencies is 20–1400 Hz and 20–1600 Hz, respectively. The values of E_A are given on the contour lines, and the intervals between the contour lines are 1.0 dB in Figs. 12(a) and 12(c) and 4.0 dB in Figs. 13(b) and

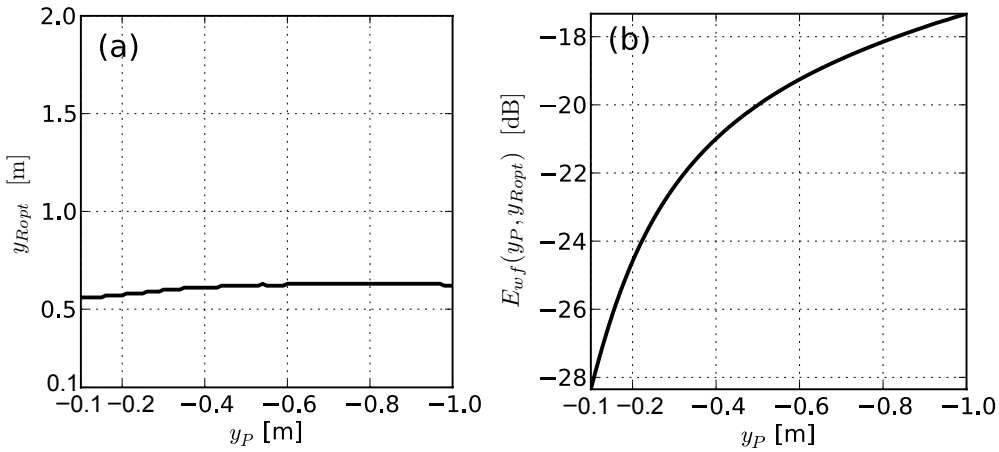


Figure 11. (a) Resultant optimized control line y_{Ropt} for various y_P and (b) calculation results for evaluation value E_{wf} on the line with a y-coordinate of y_{Ropt}.

13(d). As can be seen in Figs. 12 and 13, the amplitude error of WFS is serious by comparison with that of DMCWS both for the upper-limit of the evaluation frequency band which are below and above the spatial aliasing frequency (1417 Hz). In contrast, in Figs. 12(b), 12(d), 13(b) and 13(d), the amplitude error of DMCWS is from −8 to −20 dB, which is small in comparison with that of WFS, and is generally smallest in the vicinity of the control points. The amplitude error difference between the results of WFS and that of DMCWS indicates 60–120 dB around the control points.

Figures 14 and 15 show the phase error E_P of WFS and DMCWS that of the evaluated wavefront band frequencies is 20–1400 Hz and 20–1600 Hz, respectively. The values of E_P are given on the contour lines, and the intervals between the contour lines are 1.0 dB in Figs. 14(a), 14(c), 15(a) and 15(c) and 2.0 dB in Figs. 14(b), 14(d), 15(b) and 15(d). According to Figs. 14 and 15, there is a significant phase error E_P by comparison with that of DMCWS, similar to the amplitude error shown in Figs. 12 and 13, in the case of WFS. In contrast, there is an extremely small phase error in DMCWS, as shown in Figs. 14(b), 14(d), 15(b) and 15(d).

In Figs. 12 and 13, the amplitude error and the phase error is smaller than that

Table 2. Calculation conditions used in evaluation of DMCWS and WFS

PARAMETER	VALUE
Primary source geometry (x_P, y_P, z_P)	$(1.2, -0.1, 1.22), (1.2, -0.7, 1.22)$ m
Control line y-coordinate y_R	0.6 m
Evaluated wavefront	
band frequencies	20-1400 and 20-1600 Hz
	(10 Hz interval)
Reference listening distance	
y-coordinate y_R	0.6 m

of WFS.

We next calculate the attenuation to examine the characteristic of the wavefront amplitude error. Figure 16 shows the attenuations of the primary wavefront and the WFS and DMCWS secondary wavefronts, as well as Att_{S_S} (see Eq. (4)) in front of the primary source at the upper-limit frequency of the evaluation band (1600 Hz). To determine the effect of the y-coordinate of the primary source y_P on wavefront attenuation, we calculate the attenuation for y_P of -0.1 and -0.7 m, as shown in Figs. 16(a), (c) and 16(b), (d), respectively. The attenuation plot for WFS undergoes greater fluctuation than the other attenuation plots in Figs. 16(c) and (d), because the evaluation frequency of 1600 Hz is above the spatial aliasing frequency (1417 Hz). On the other hand, the amplitude in DMCWS undergoes little fluctuation compared with that in WFS, suggesting the applicability of DMCWS in frequency bands higher than the spatial aliasing frequency. Also, Fig. 16 shows that the attenuation in DMCWS is very similar to Att_{S_S} rather than that of the primary sound source. This result implies the possibility that spatial decay [44] occurs in DMCWS in the same way as it does in WFS, suggesting the ease of estimating the amplitude error in DMCWS.

3.4. Summary of this Chapter

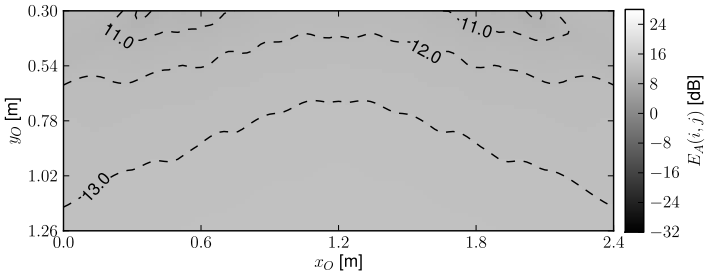
In this Chapter, the optimal control-point geometry and the behavior of the synthesized secondary wavefront in MCWS was investigated. It has been shown that

the accuracy of the synthesized secondary wavefront is related to the control-point coordinates and the wavefront measurement results of DMCWS through the numerical calculations.

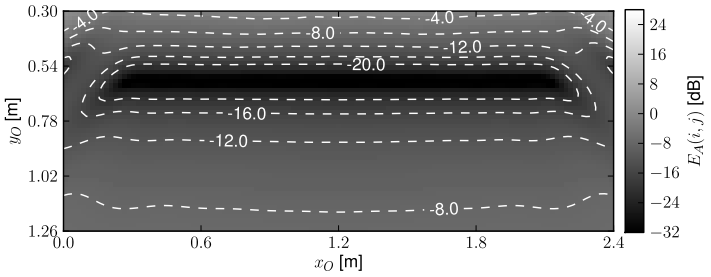
Numerical wavefront calculations clarified that the optimum y-coordinate of the directly aligned control point is y_R = 0.6–0.7 m in the target area of our sound field reproduction system.

Next, the effectiveness of DMCWS was evaluated through the comparison with WFS. Numerical wavefront calculations obtained WFS and DMCWS wavefronts compared using this range of y_R that DMCWS has a larger listening area with fewer amplitude and phase errors than WFS, whereas they have a similar attenuation error.

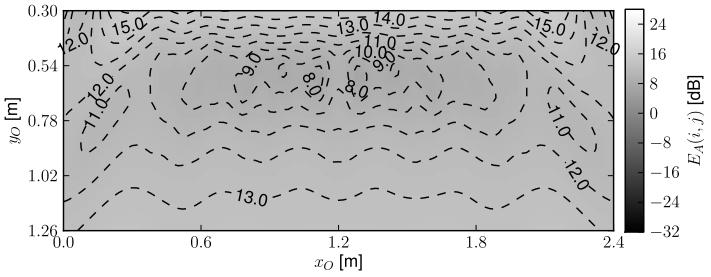
From these findings, the results clarified that the advantageousness of DMCWS compared with WFS.



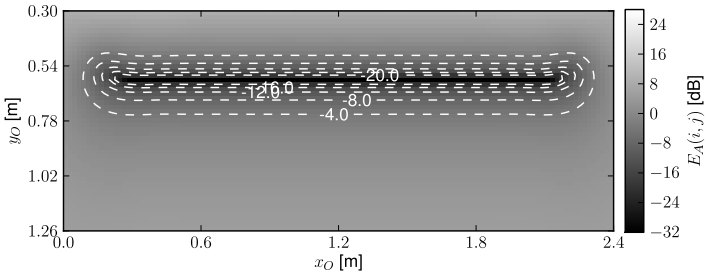
(a) Calculated amplitude error E_A for WFS.



(b) Calculated amplitude error E_A for DMCWS.

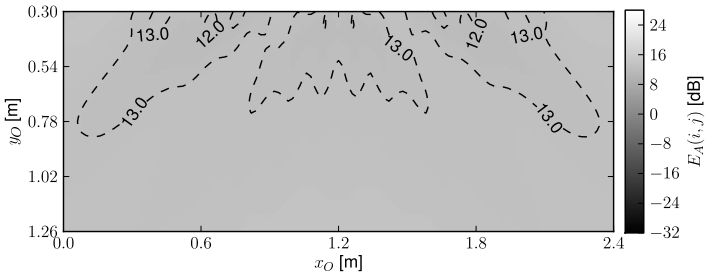


(c) Calculated amplitude error E_A for WFS.

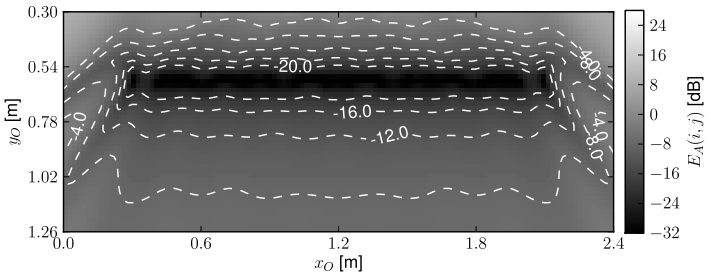


(d) Calculated amplitude error E_A for DMCWS.

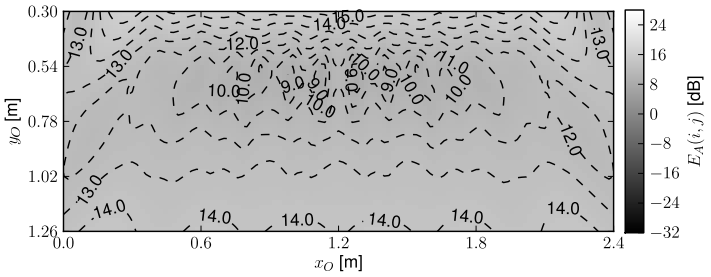
Figure 12. Comparison of the calculated E_A of WFS and DMCWS. The evaluated frequency bands are 20–1400 Hz. The primary source geometry of (a) and (b) are $(x_P, y_P, z_P) = (1.2, -0.1, 1.22)$, and (c) and (d) are $(x_P, y_P, z_P) = (1.2, -0.7, 1.22)$.



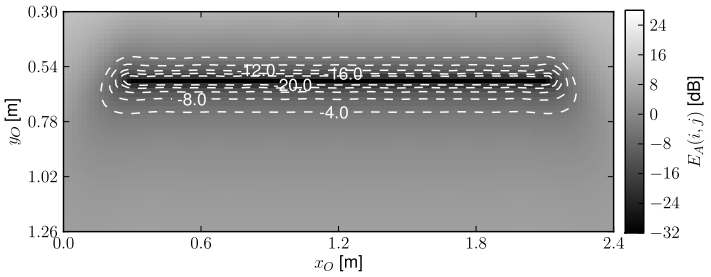
(a) Calculated amplitude error E_A for WFS.



(b) Calculated amplitude error E_A for DMCWS.

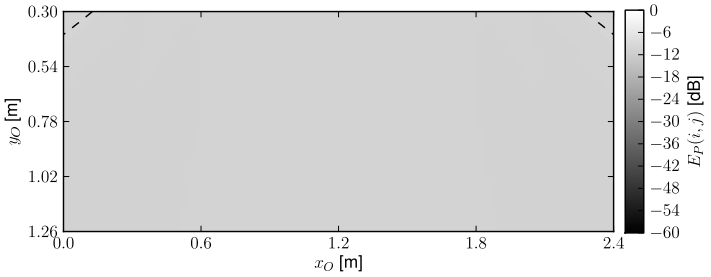


(c) Calculated amplitude error E_A for WFS.

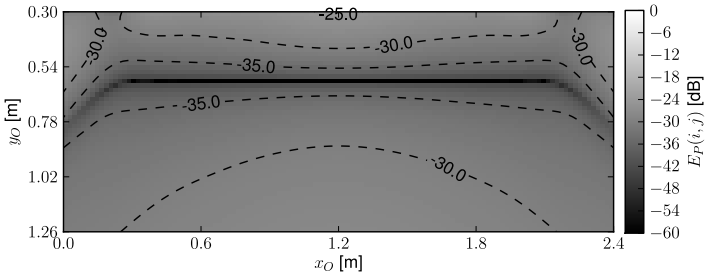


(d) Calculated amplitude error E_A for DMCWS.

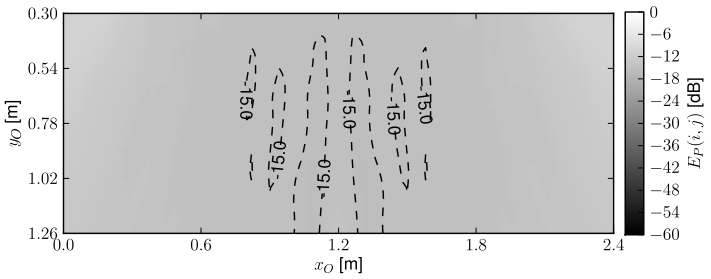
Figure 13. Comparison of the calculated E_A of WFS and DMCWS. The evaluated frequency bands are 20–1600 Hz. The primary source geometry of (a) and (b) are $(x_P, y_P, z_P) = (1.2, -0.1, 1.22)$, and (c) and (d) are $(x_P, y_P, z_P) = (1.2, -0.7, 1.22)$.



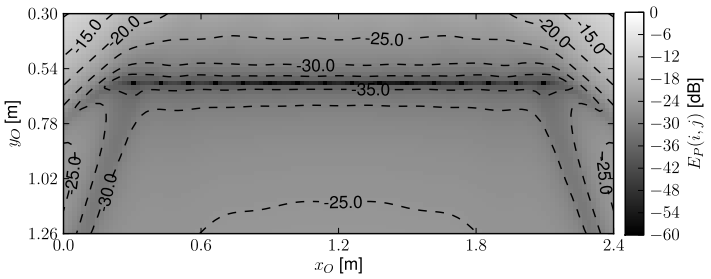
(a) Calculated phase error E_P for WFS.



(b) Calculated phase error E_P for DMCWS.

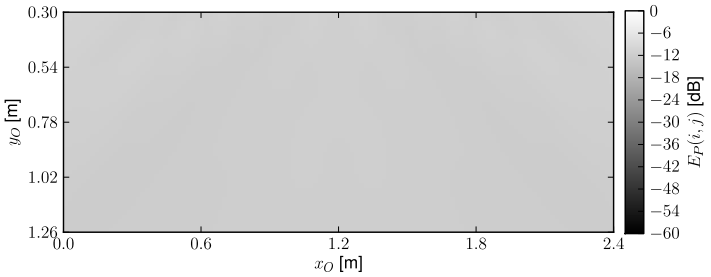


(c) Calculated phase error E_P for WFS.

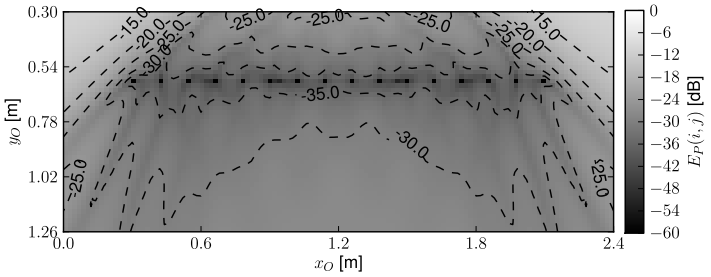


(d) Calculated phase error E_P for DMCWS.

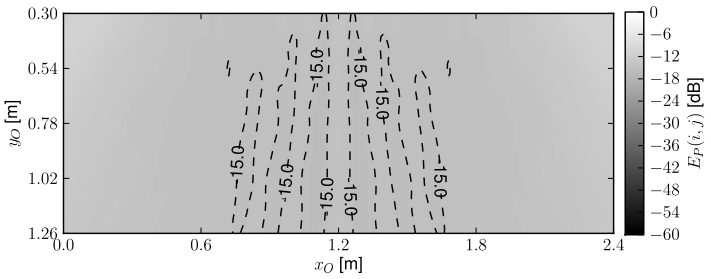
Figure 14. Comparison of the calculated E_P of WFS and DMCWS. The evaluated frequency bands are 20–1400 Hz. The primary source geometry of (a) and (b) are $(x_P, y_P, z_P) = (1.2, -0.1, 1.22)$, and (c) and (d) are $(x_P, y_P, z_P) = (1.2, -0.7, 1.22)$.



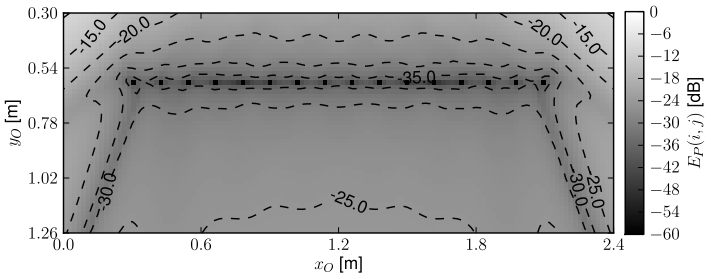
(a) Calculated phase error E_P for WFS.



(b) Calculated phase error E_P for DMCWS.

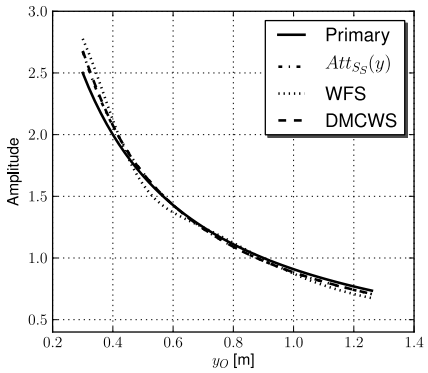


(c) Calculated phase error E_P for WFS.

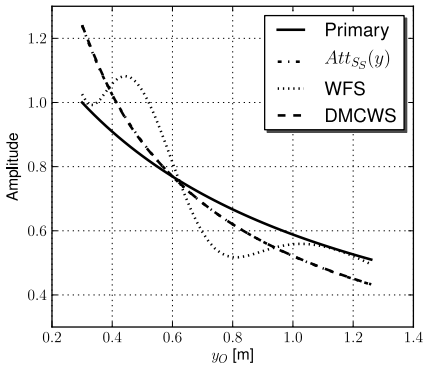


(d) Calculated phase error E_P for DMCWS.

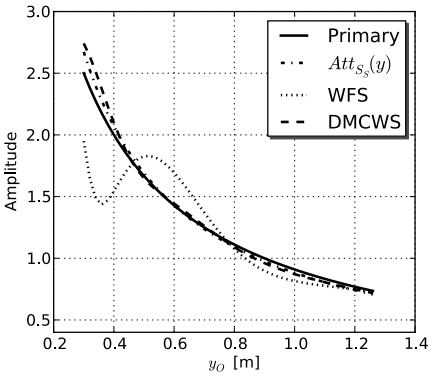
Figure 15. Comparison of the calculated E_P of WFS and DMCWS. The evaluated frequency bands are 20–1600 Hz. The primary source geometry of (a) and (b) are $(x_P, y_P, z_P) = (1.2, -0.1, 1.22)$, and (c) and (d) are $(x_P, y_P, z_P) = (1.2, -0.7, 1.22)$.



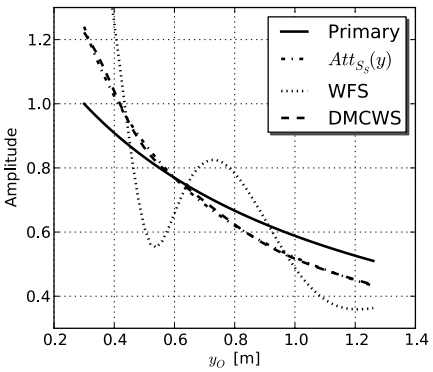
(a) Primary source is located at $(x_P, y_P, z_P) = (1.2, -0.1, 1.22)$ m and wavefront frequency is 1400 Hz.



(b) Primary source is located at $(x_P, y_P, z_P) = (1.2, -0.7, 1.22)$ m and wavefront frequency is 1400 Hz.



(c) Primary source is located at $(x_P, y_P, z_P) = (1.2, -0.1, 1.22)$ m and wavefront frequency is 1600 Hz.



(d) Primary source is located at $(x_P, y_P, z_P) = (1.2, -0.7, 1.22)$ m and wavefront frequency is 1600 Hz.

Figure 16. Comparison of attenuation in front of primary source.

Chapter 4

Evaluation in actual environment of DMCWS

4.1. Introcution

The numerical evaluation results in the previous chapter clarify the effectivity of the wavefront synthesized with DMCWS. In this chapter, we propose a wavefront measurement method using the spatial spectrum characteristics obtained from the impulse response at each observation point, and we measure the wavefront to verify the numerical evaluation results of DMCWS in actual environment.

4.2. Spatial spectrum characteristics obtained from impulse responses

In this section, we propose a wavefront measurement method using the spatial spectrum characteristics obtained from the impulse response at each observation point.

The wavefront spectrum characteristics (Eq. (14)) at the observation point $S_O(\omega)$ are expressed below in vector form:

$$\begin{aligned} S_O(\omega) &= \boldsymbol{Q}^{\mathrm{T}}(\omega)\boldsymbol{S_S}(\omega) \\ &= \boldsymbol{Q}^{\mathrm{T}}(\omega)\boldsymbol{H}^{+}(\omega)\boldsymbol{P}(\omega)S_P(\omega) \\ &= S_{arr}(\omega)S_P(\omega), \end{aligned} \tag{22}$$

where

$$\boldsymbol{Q}(\omega) = [Q_1(\omega), Q_2(\omega), ..., Q_N(\omega)]^{\mathrm{T}}, \tag{23}$$

$$Q_n(\omega) = G_{Sn}(\theta_{On}, \omega)\frac{\exp(-\mathrm{j}kr_{On})}{r_{On}}. \tag{24}$$

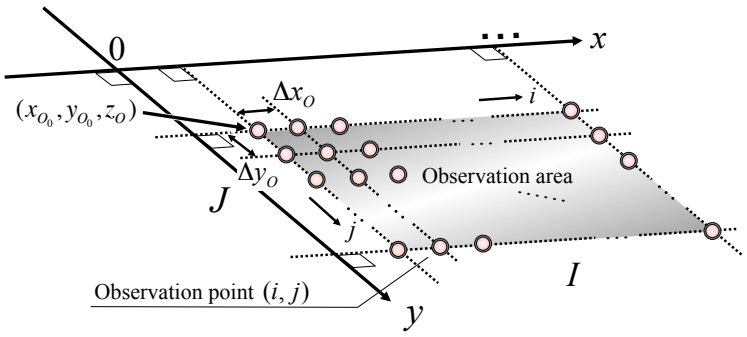


Figure 17. The observation area arranged in a reticular pattern of observation point over a horizontal plane $z = z_O$.

$Q_n(\omega)$ is the radiation characteristic of the nth sound source at the azimuth angle θ for angular frequency ω, and the loudspeaker array characteristic $S_{arr}(\omega)$ is assumed to be the radiation characteristic of a single sound source.

Hence, we can estimate the spectrum characteristic of a loudspeaker array $S_{arr}(\omega)$ using an acoustic impulse response measurement method with a signal input to a primary source $S_P(\omega)$ and an output signal at the observation point $S_O(\omega)$. The set of spatial spectrum characteristic of loudspeaker array $S_{arr}(\omega)$ in the observation area around the construct wavefront, since we can estimate the spatial spectrum characteristic of synthesized secondary wavefront to measure the set of $S_{arr}(\omega)$.

Figure 17 shows the observation area arranged in a reticular pattern of observation point. In Fig. 17, the set of spectrum characteristic of synthesized secondary wavefront at each observation point in matrix form:

$$\boldsymbol{S}_{arr}(\omega) = \begin{bmatrix} S_{arr(1,1)}(\omega) & \cdots & S_{arr(1,J)}(\omega) \\ S_{arr(2,1)}(\omega) & \cdots & S_{arr(2,J)}(\omega) \\ \vdots & \ddots & \vdots \\ S_{arr(I,1)}(\omega) & \cdots & S_{arr(I,J)}(\omega) \end{bmatrix}, \tag{25}$$

where I and J are the total numbers of spatial sampling index of i and j, respectively.

From Eq. (25), we obtain the temporal wavefront using inverse fourier trans-

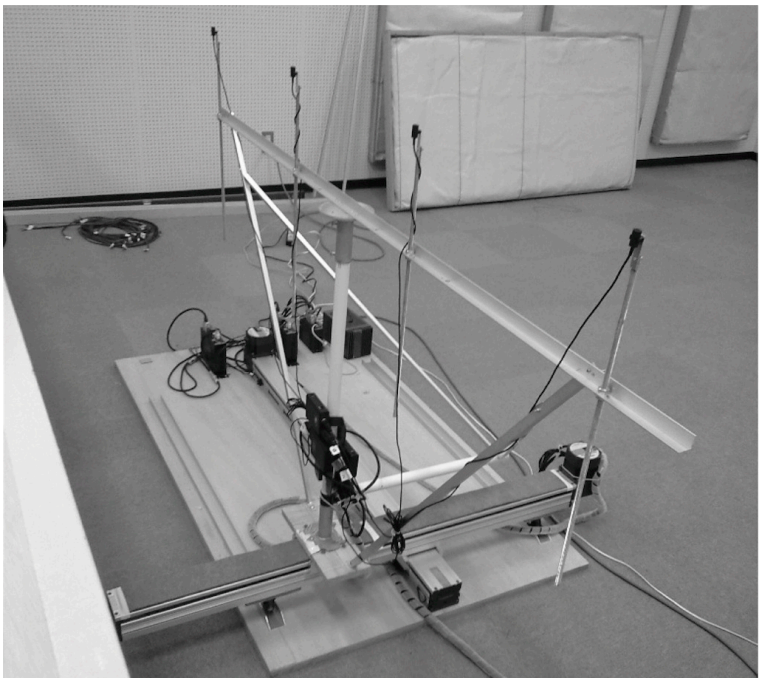


Figure 18. Wavefront measurement system consists of electric two axes actuator and microphone array.

form;

$$\boldsymbol{s}_{\mathrm{arr}}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{S}_{arr}(\omega)e^{j\omega t}d\omega = \begin{bmatrix} s_{arr(1,1)}(t) & \cdots & s_{arr(1,J)}(t) \\ s_{arr(2,1)}(t) & \cdots & s_{arr(2,J)}(t) \\ \vdots & \ddots & \vdots \\ s_{arr(I,1)}(t) & \cdots & s_{arr(I,J)}(t) \end{bmatrix}. \tag{26}$$

In this paper, we estimate $S_{arr}(\omega)$ using the M-sequence method [45] to measure the acoustic impulse response.

4.3. Wavefront measurement system

Figure 18 shows the wavefront measurement system for visualization of the wavefront obtained by Eq. (25).

The measuring microphones are placed on the electric two axes actuator so that the microhpones can move around the horizontal plane in front of the loudspeaker array and measure the spectrum characteristics $S_{arr(i,j)}(\omega)$. In addition, we use a linear microphone array to save the amount of time spent for measurement. The interval of the microphones is 0.48 m and the total number of microphones is 4, i.e., microphone array width is 1.44 m, and the microphones are audio-technica ATM14a omnidirectional microphones. The electric actuator has 0.96 m range of movement on each axis. As a result, the width (x-axis) and height (y-axis) of the observation area are 2.4 m and 0.96 m, respectively.

Figure 19 shows the procedure used to construct resultant all wavefronts of all the observation areas from measured wavefront at each observation area. The wavefront constructed in all the observation areas, shown in Fig. 19 (b), can be obtained as an overlap of each wavefront measured from adjacent microphones, shown in Fig. 19 (a).

4.4. Wavefront measurement conditions

Table 3 shows the wavefront measurement conditions. The control line y-coordinate y_C is set to 0.6 m. The order L of M-Sequence is determined by the relation between measurement room's reverberation time T_r and sampling frequency f_s as

$$L \geq \lceil \log_2(f_s T_r) \rceil, \tag{27}$$

where $\lceil \cdot \rceil$ denotes ceiling function. Table 3 and Eq. (27) show $L \geq \lceil \log_2(48000 \times 0.3) \rceil = 14$, and consequently we use L=15 in this paper.

4.5. Results of wavefront measurement

Figures 20 and 21 show the calculate and measured wavefronts obtained using DMCWS, respectively. Compared with Fig. 20, we can see a clearer interference pattern in Fig. 21. Figures 22 and 23 show the calculated and measured frequency-amplitude characteristics in front of the center of the loudspeaker array, respectively. In Fig. 23, a pattern due to the interference can also be observed.

Table 3. Wavefront measurement conditions

PARAMETER	VALUE
Measurement room	Nagaoka University of Technology Acoustically isolated room W 5.3×H 9.5×D 4.3 m Reverberation time T_r 0.3 s
Room temperature	15℃
Sound velocity c	340.64 m/s
Primary source geometry (x_P, y_P, z_P)	(1.2,-0.1,1.22) m
Wavefront drawing sample	31 samples
Sampling frequency f_s	48 kHz
Wavefront synthesis method	DMCWS
Synchronous addition count	40 cycles
Secondary source	Soundevice SD–0.6
Secondary source order N	16
Source distance Δx	0.12 m
Spatial aliasing frequency f_{alias}	1417 Hz
Measured wavefront band frequencies	200~24000 Hz
Height of evaluating horizontal plane z_O	1.22 m
Height of secondary sources	1.22 m

4.6. Effects of room reflection

In the soundproof room used in this measurement, we consider that the wave reflected by the wall affects the secondary synthesized wavefront. The reflection surface nearest to the source is the floor of this room. The relation between the direct and reflected waves generated by the primary source is obtained by the *image method* [46], where the line $y = 0$ corresponds to the floor. In the image method

Table 4. Calculation conditions for primary source wavefront and the wavefront reflected from the floor surface

PARAMETER	VALUE
Room temperature	15℃
Sound velocity c	340.64 m/s
Sampling frequency f_s	48 kHz
Primary sound source model	Point source

the reflected wave is generated by an imaginary source located at the reflection of the primary source in the line $y = 0$. In the image method, we can regard the reflected wave as a direct wave generated from this imaginary point source. Using the image method, we can consider the effect of waves reflected from the floor, and we calculate the primary wavefront using this method. Figure 24 and Table 4 show the calculation conditions.

Figure 25 shows the frequency characteristics of the wavefront calculated with the primary source and the reflected wave obtained by the image method. This result shows that the wave reflected from the floor surface causes an interference pattern in the frequency characteristics, which tends to broaden as the observation point becomes more distant from the primary source. Next, we consider that the secondary wavefront affects the wave reflected from the wall surface. Figure 26 shows an impulse response as an example including the wave reflected from the wall surface and the definitions of T_D and T_{R1} which are the time of arrival at the observation point to the primary source and the arrival time interval of the first early reflection wavefront at the observation point, respectively. Therefore, we remove the wave reflected from all of the wall surface in room using hanning

window given by the following equation:

$$w(t) = \begin{cases} 0.5 - 0.5\cos\left(2\pi\dfrac{t-(T_D-T_{R1})}{2T_{R1}}\right) \\ \qquad (\text{if } T_D - T_{R1} \le t \le T_D + T_{R1}), \\ 0 \qquad (\text{otherwise}), \end{cases} \tag{28}$$

where the arrival time of the direct wave from the primary source at the observation point, T_D, is given by

$$T_D = \underset{t}{\text{argmax}}(s_{arr(x_O,y_O)}(t)) \tag{29}$$

where $\text{argmax}_t(\cdot)$ denotes the function that indicates the specific variable t to maximize the subject. Figure 27 shows the geometry of the primary source, the secondary source and the imaginary secondary source obtained by the image method. In Fig. 27, z_S denotes the distance from the floor surface to the secondary source. From this figure, T_{R1} is obtained as follows:

$$T_{R1} = \frac{r_{IO} - r_{SO}}{c}, \tag{30}$$

where r_{SO} and r_{IO} are written as

$$r_{SO} = \sqrt{\left(x_O - \frac{y_O x_P - x_O y_P}{y_O - y_P}\right)^2 + {y_O}^2}, \tag{31}$$

$$r_{IO} = \sqrt{{r_{SO}}^2 + 4{z_S}^2}. \tag{32}$$

Figure 28 shows the frequency characteristics of the secondary wavefront in front of the primary source after the removal of the wavefront reflected from all of the wall surface in room. In this figure, the interference pattern shown in Fig. 23 is mitigated by the proposed window function for removal of the wavefront reflected from the wall surface. Therefore, it is predictable that the interference pattern is due to the wavefront reflected from the wall surface.

Figure 29 shows the secondary wavefront after the removal of the wavefront reflected in front of the primary sound source. By comparing this figure with

Fig. 21, we can conclude that the reflected wavefront caused the interference observed in the measurement results because the wavefront interference is less evident in Fig. 29 is mitigated. In addition, DMCWS is performed above the WFS aliasing frequency (1417 Hz) in Fig. 29. Then, we evaluate the amplitude and phase error of measured wavefront using E_A and E_P in evaluated frequency band. Figure 30 shows the results of E_A and E_P with measured wavefront after removal of reflected wavefront. The lower limit of the evaluated frequency 200 Hz is determined by hardware (loudspeaker) limitation. The values of E_A and E_P are given on the contour lines, and the intervals between the contour lines are 1.0 dB in Figs. 30(a), 30(c). According to Fig. 30(a) and 30(c), the results of the amplitude and phase error E_A and E_P differ from that of the numerical simulations. In this paper, the filter calculation of DMCWS is premised on an anechoic reproduction environment. Therefore, the secondary wavefront accuracy decreases in practice owing to the inherent disadvantages of physical inaccuracies found in the DMCWS model characteristics described in the Sec. 2.2.

4.7. Effect of loudspeaker frequency characteristics on measured wavefront

To find the cause of the errors which are shown in the Fig. 30, we compare the frequency characteristics of the calculated, the measured after the removal of the reflected wavefront and that of the loudspeaker (Soundevice SD–0.6) at the observation point of $(x_O, y_O, z_O) = (1.2, 0.3, 1.22)$ m in Fig. 31. According to the results, the measured characteristics are similar to those of the secondary source loudspeaker in the low-frequency subband with frequencies of up to 1600 Hz. Thus, the amplitude and phase error of the measured wavefront which shown in Fig. 30 can be attributed to the frequency-amplitude characteristics of the loudspeaker.

4.8. Summary of this Chapter

In this chapter it has been shown that the accuracy of the wavefront measurement results of DMCWS in an actual environment.

Our wavefront measurement system and our algorithm using impulse responses measured in an acoustically isolated room clarified that the measured wavefront is affected by the mural-reflected wave and the frequency-amplitude characteristics of the secondary source loudspeaker. In addition, DMCWS can be performed above the spatial aliasing frequency of WFS according to the results of measurement.

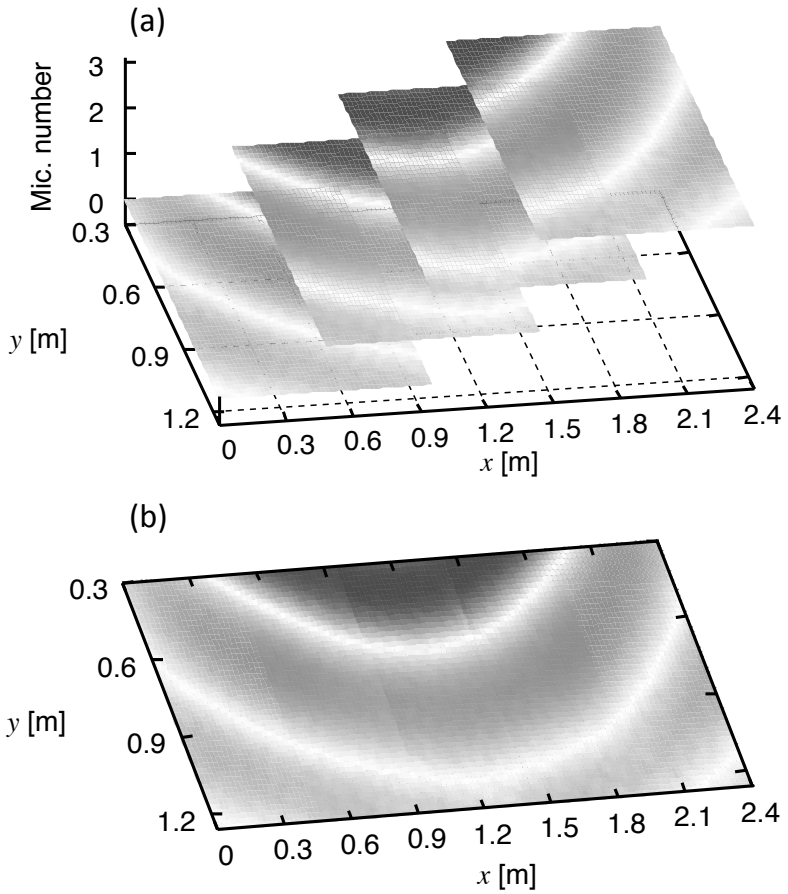


Figure 19. The procedure used to construct all wavefront of all the observation areas from measured wavefront at each observation area. (a) Before coupling the wavefront measured by each microphone observation area. (b) The wavefront after coupling process.

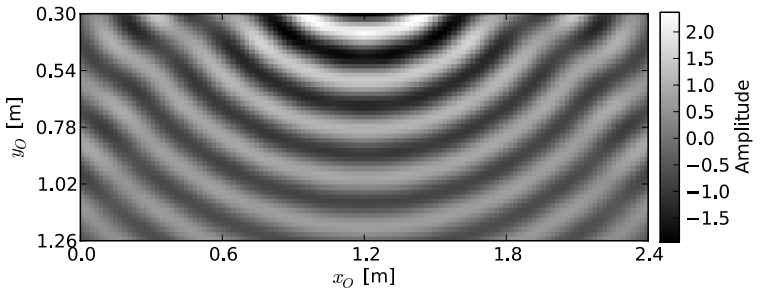


Figure 20. Calculated secondary wavefront (1600 Hz).

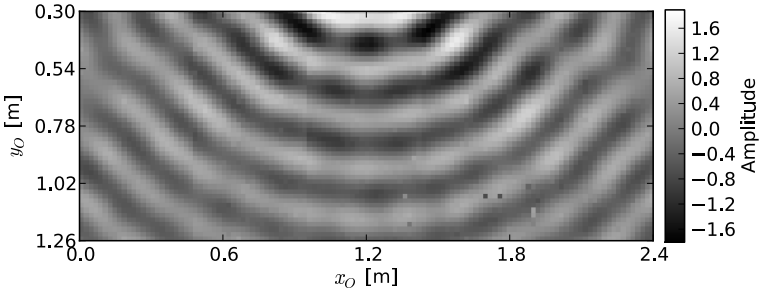


Figure 21. Measured secondary wavefront (1600 Hz).

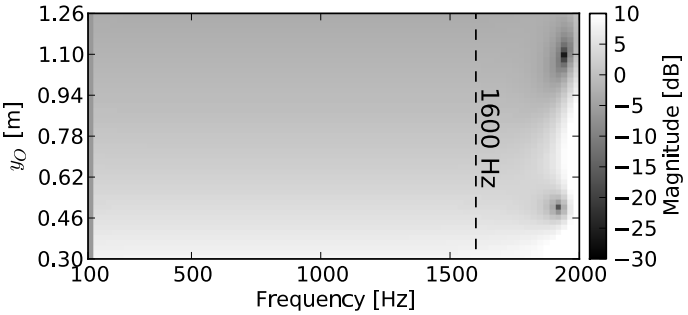


Figure 22. Calculated frequency amplitude characteristics in front of center of loudspeaker array.

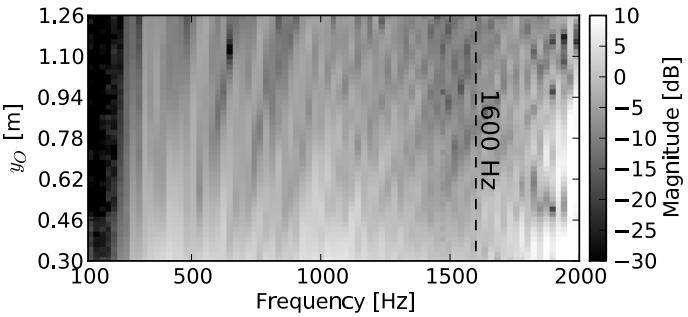


Figure 23. Measured frequency amplitude characteristics in front of center of loudspeaker array.

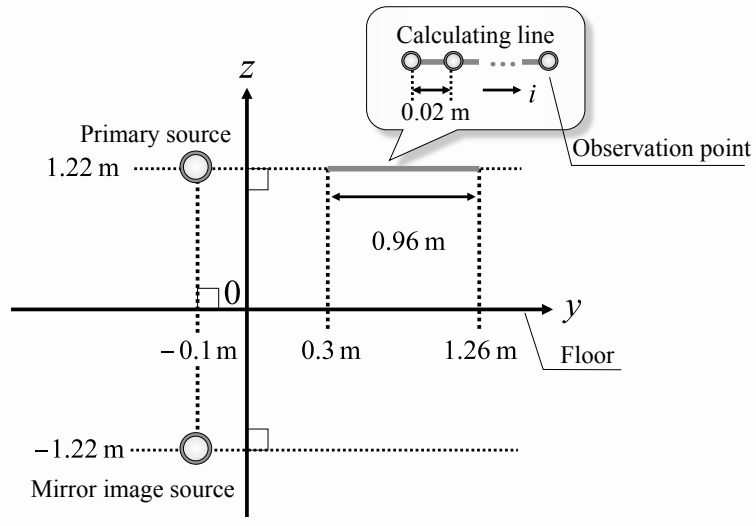


Figure 24. Calculation conditions used to determine the effect of the wave reflected from the floor surface.

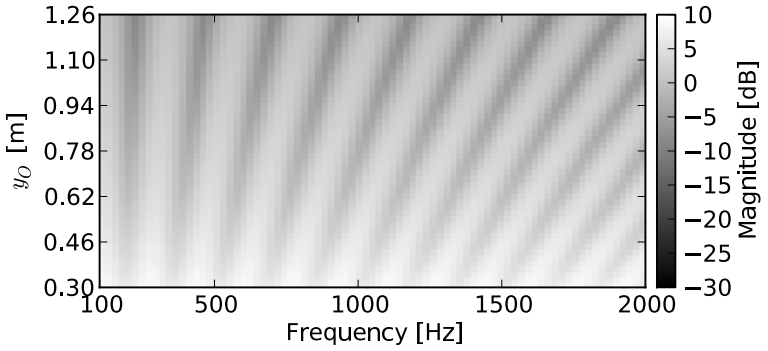


Figure 25. Frequency characteristics of primary source wavefront and the wavefront reflected from the floor surface calculated with the image method.

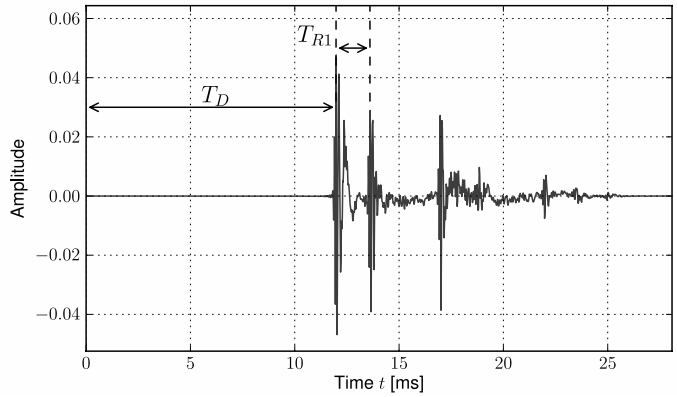


Figure 26. An impulse response including the wave reflected from the surface which is nearest to the source and the definitions of T_D and T_{R1}.

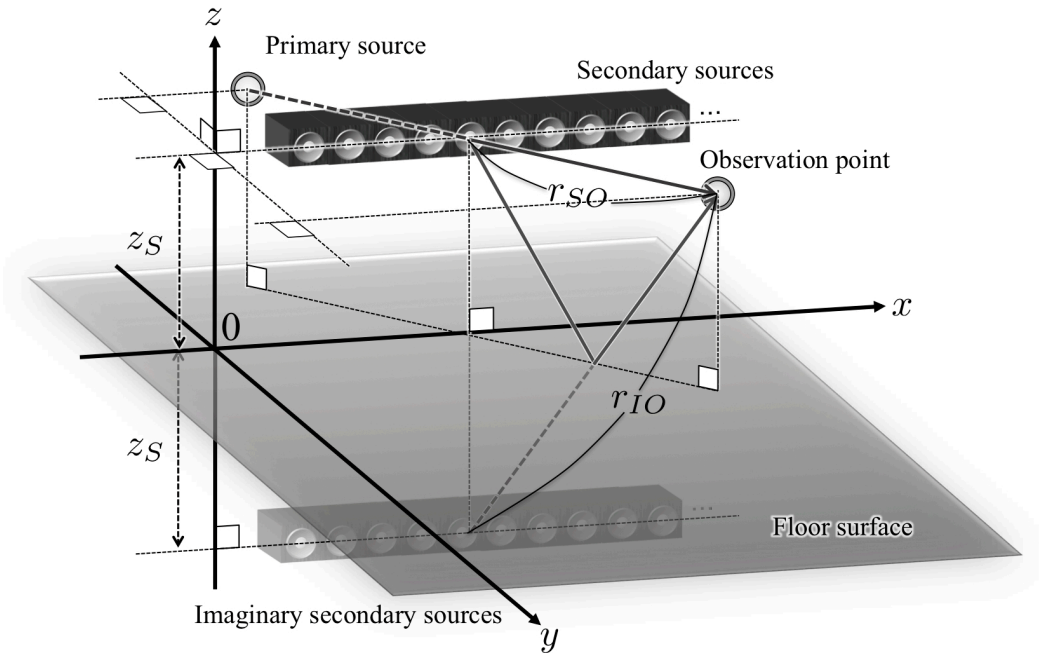


Figure 27. Geometry of the primary source, the secondary sources and the imaginary secondary sources with the image method.

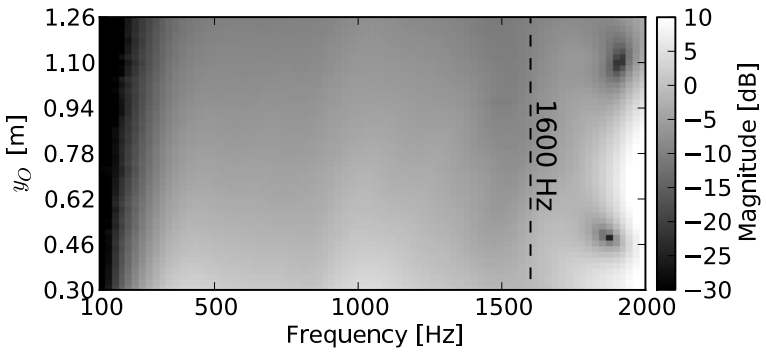


Figure 28. Frequency-amplitude characteristics of the secondary wavefront in front of the primary sound source after the removal of the wavefront reflected from the wall surface.

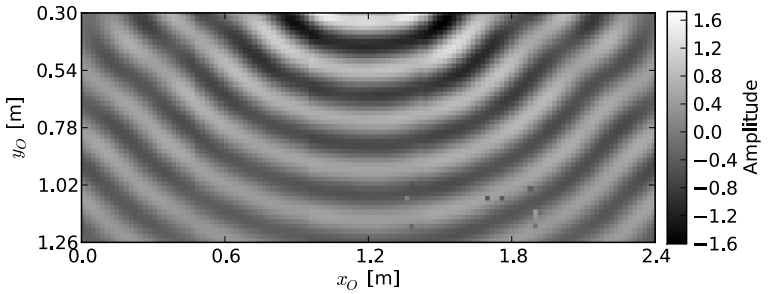
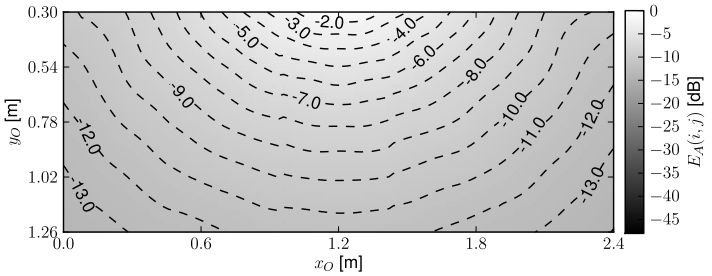
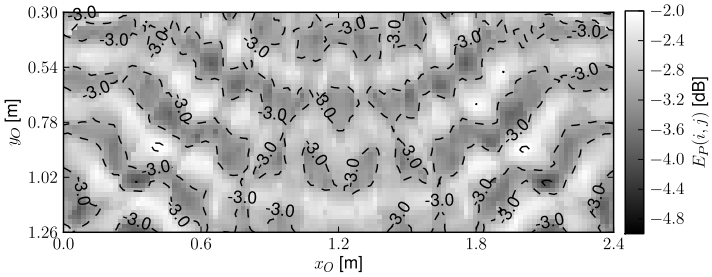


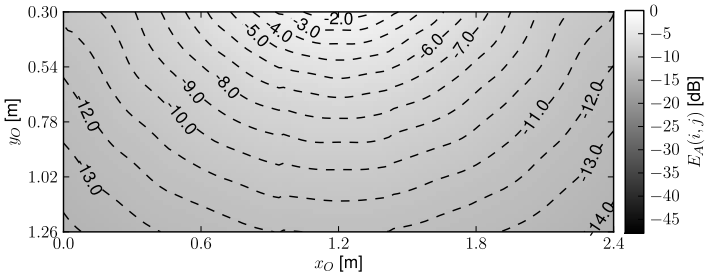
Figure 29. Secondary wavefront after removal of reflected wavefront(1600 Hz).



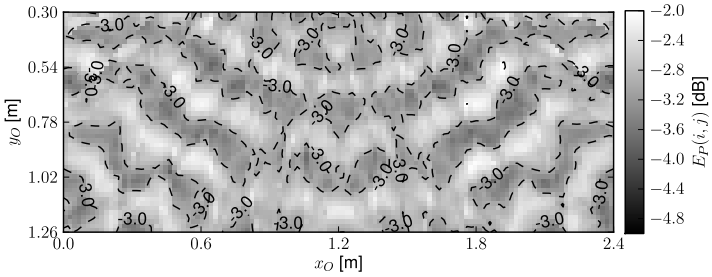
(a) E_A of the measured secondary wavefront (200–1400 Hz).



(b) E_P of the measured secondary wavefront (200–1400 Hz).



(c) E_A of the measured secondary wavefront (200–1600 Hz).



(d) E_P of the measured secondary wavefront (200–1600 Hz).

Figure 30. Results of E_A and E_P with measured wavefront. Evaluated frequency bands are (a), (b) 200–1400 Hz and (c), (d) 200–1600 Hz.

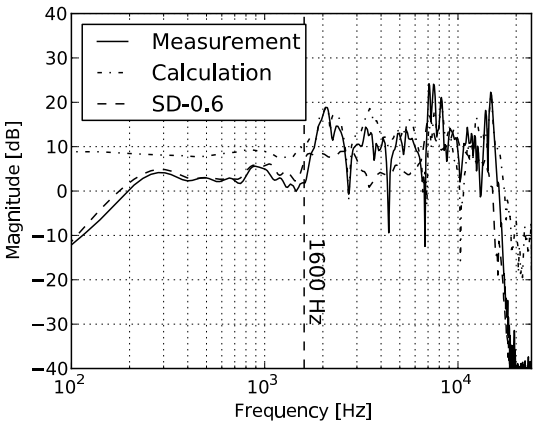


Figure 31. Comparison of frequency characteristics of the loudspeaker and measured wavefront.

Chapter 5

Hybrid wave field synthesis

5.1. Introduction

In this Chapter, I propose a new method of balancing the listening area and reproduction accuracy using an inverse filter of the room acoustics. To develop an expression for the proposed filter, we used the secondary source driving filter of WFS instead of the filter used in conventional studies. By using WFS, the proposed method overcomes the compensation limitation of auditory distance and azimuth perception outside the control points. The efficacy of the proposed method is ascertained by objective evaluation through computer-based simulations and wavefront-measurement.

To improve the robustness against a shift of the user's position in MCWS and MINT, Miyabe, et al. have proposed an inverse filter design method in which a wavefront radiated from the loudspeaker nearest to the primary source is inserted in the subspace not spanned for the reproduced signal space (sweet-spot-signal space) in the inverse filter matrix [26, 27]. This method can approximately provide a sound field even outside the control points without any degradation of reproduction at the control points (sweet spot). However, this method has a major disadvantage in the localization reproduction ability outside the sweet spot because the reproducible primary source position is quantized within the loudspeaker position, i.e., we cannot compensate for the perceptual distance outside the sweet spot.

In contrast, WFS can reproduce an arbitrary wavefront that compensates for the perceptual distance. However, as described in introduction of this thesis, WFS cannot accurately generate sound pressure at the control points owing to the existence of room reverberation and the approximation in the theory.

The above-mentioned facts imply that the advantages and disadvantages of our previous method and WFS are complementary; this motivates us to propose an approach combining multi-point sound field reproduction and WFS in this chapter.

In the proposed method, the wavefront outside the sweet spot is derived from an approximation of the WFS-synthesized wavefront, and we insert it in the subspace in the inverse filter matrix. As the result, we can simultaneously achieve the following: (a) we can reproduce perfect sound pressures in the sweet spot (control points), which are not disturbed by the WFS wavefront, and (b) we can perceive the approximated wavefront reproduced by WFS outside the sweet spot. Such a behavior of the integration of two complementary methods in the proposed method, I call the proposed method "*hybrid wave field synthesis* (HWFS)." The detailed algorithm of HWFS is described below.

5.2. Proposed method

Utilizing singular value decomposition, the generalized inverse matrix $\boldsymbol{Z}^{-}(\omega)$ of the transfer impedance matrix $\boldsymbol{Z}(\omega)$ can be denoted as

$$\boldsymbol{Z}^{-}(\omega) = \underbrace{\boldsymbol{V}(\omega)}_{(N\times N)} \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{S}(\omega) \end{bmatrix}}_{(N\times M)} \underbrace{\boldsymbol{U}^{\mathrm{H}}(\omega)}_{(M\times M)}, \tag{33}$$

where the superscript H denotes the complex conjugate transposition of a matrix, $\boldsymbol{V}(\omega)$ and $\boldsymbol{U}(\omega)$ are the unitary matrices whose columns are the right and left singular vectors of $\boldsymbol{Z}(\omega)$, and $\boldsymbol{\Lambda}(\omega)$ is

$$\boldsymbol{\Lambda}(\omega) = \mathrm{diag}\left[\lambda_1(\omega), ..., \lambda_M(\omega)\right], \tag{34}$$

where λ_m is expressed with the singular values σ_m of $\boldsymbol{Z}(\omega)$ as

$$\lambda_m(\omega) = \begin{cases} \frac{1}{\sigma_m(\omega)} & (\text{if } \sigma_m(\omega) \neq 0), \\ 0 & (\text{otherwise}), \end{cases} \tag{35}$$

where the singular values $\sigma_m(\omega)$ have non-negative elements appearing in non-increasing order such that $\sigma_1(\omega) \geq \cdots \geq \sigma_M(\omega)$. The MP generalized inverse matrix $\boldsymbol{Z}^{+}(\omega)$ can be obtained by setting $\boldsymbol{S}(\omega)$ to be a zero matrix. However, the MP-type inverse filter is specific to the reproduction at the control points and the

reproduction cannot be guaranteed outside the control points. Thus, the sound localization degrades considerably when the user moves from a controlled area.

Next, to approximate $\boldsymbol{T}(\omega)$, which is the wavefront control filter outside the control points in the subspace (or nullspace) of $\boldsymbol{Z}^{-}(\omega)$ with arbitrary components $\boldsymbol{S}(\omega)$ in Eq. (33), obtain the generalized inverse matrix $\boldsymbol{Z}^{-}(\omega)$ closest to $\boldsymbol{T}(\omega)$. We utilize the Frobenius norm as the distance measure and we obtain $\boldsymbol{Z}^{-}(\omega)$ to minimize $F(\omega) = \|\boldsymbol{Z}^{-}(\omega) - \boldsymbol{T}(\omega)\|_{\mathrm{F}}$. Since the Frobenius norm is not changed by the multiplication of unitary matrices, $F(\omega)$ can be rewritten as

$$\begin{aligned} F(\omega) &= \|\boldsymbol{V}^{\mathrm{H}}(\omega)(\boldsymbol{Z}^{-}(\omega) - \boldsymbol{T}(\omega))\boldsymbol{U}(\omega)\|_{\mathrm{F}} \\ &= \left\| \begin{bmatrix} \boldsymbol{\Lambda}(\omega) - \boldsymbol{V}^{\mathrm{H}}_{\mathrm{span}}(\omega)\boldsymbol{T}(\omega)\boldsymbol{U}(\omega) \\ \boldsymbol{S}(\omega) - \boldsymbol{V}^{\mathrm{H}}_{\mathrm{null}}(\omega)\boldsymbol{T}(\omega)\boldsymbol{U}(\omega) \end{bmatrix} \right\|_{\mathrm{F}}, \end{aligned} \tag{36}$$

where $\boldsymbol{V}_{\mathrm{span}}(\omega)$ is a matrix composed of the first M columns of $\boldsymbol{V}(\omega)$. Since $\boldsymbol{\Lambda}(\omega)$ is a constant matrix, $F(\omega)$ can be minimized if and only if $\boldsymbol{S}(\omega) - \boldsymbol{V}^{\mathrm{H}}_{\mathrm{null}}(\omega)\boldsymbol{T}(\omega)\boldsymbol{U}(\omega) = 0$; thus, the optimal inverse filter is obtained as follows by setting $\boldsymbol{S}(\omega) = \boldsymbol{V}^{\mathrm{H}}_{\mathrm{null}}(\omega)\boldsymbol{T}(\omega)\boldsymbol{U}(\omega)$ in Eq. (33):

$$\begin{aligned} \boldsymbol{Z}^{-}_{\mathrm{opt}}(\omega) &= \underset{\boldsymbol{Z}^{-}(\omega)}{\mathrm{argmin}} F(\omega) \\ &= \boldsymbol{V}(\omega) \begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{V}^{\mathrm{H}}_{\mathrm{null}}(\omega)\boldsymbol{T}(\omega)\boldsymbol{U}(\omega) \end{bmatrix} \boldsymbol{U}^{\mathrm{H}}(\omega). \end{aligned} \tag{37}$$

Next, we design the filter to guarantee the sound field accuracy outside the control points. As a method of generating the desired wavefront, WFS was introduced in this study. From Eq. (1), the spectrum at the control points can be written in terms of the impedance matrix $\boldsymbol{Z}(\omega)$ as

$$\boldsymbol{S}_C(\omega) = \boldsymbol{Z}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)S_P(\omega), \tag{38}$$

$$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega) = \left[Q^{(\mathrm{WFS})}_1(\omega), \ldots, Q^{(\mathrm{WFS})}_N(\omega)\right]^{\mathrm{T}}. \tag{39}$$

Equally, the spectrum of MCWS at the same control points can be written as

$$\boldsymbol{S}_C(\omega) = \boldsymbol{Z}(\omega)\boldsymbol{Q}^{(\mathrm{MC})}(\omega)S_P(\omega) = \boldsymbol{Z}(\omega)\boldsymbol{Z}^{+}(\omega)\boldsymbol{W}(\omega)S_P(\omega). \tag{40}$$

Equation (39) is equivalent to Eq. (40) because WFS and MCWS synthesize identical primary wavefronts, and from the equivalence of these equations, the filter $\boldsymbol{T}(\omega)$ has to satisfy the condition $\boldsymbol{Q}^{(\mathrm{WFS})}(\omega) = \boldsymbol{T}(\omega)\boldsymbol{W}(\omega)$. Therefore, the filter $\boldsymbol{T}(\omega)$ is obtained as

$$\boldsymbol{T}(\omega) = \boldsymbol{Q}^{(\mathrm{WFS})}(\omega)\boldsymbol{W}_{\mathrm{ortho}}(\omega), \tag{41}$$

where $\boldsymbol{W}_{\mathrm{ortho}}(\omega)$ is an orthonormal vector of $\boldsymbol{W}(\omega)$ obtained as follows

$$\boldsymbol{W}_{\mathrm{ortho}}(\omega) = \frac{1}{M}\left[\frac{1}{W_1}, \cdots, \frac{1}{W_M}\right]. \tag{42}$$

From Eqs. (37) and (41) the filter of proposed method $\boldsymbol{Q}^{(\mathrm{HWFS})}(\omega)$ that compensates the limitation of reproduction accuracy outside the control points with WFS and that of the secondary source driving function $\boldsymbol{S}_S(\omega)$ are obtained as follows:

$$\begin{aligned}\boldsymbol{S}_S(\omega) &= \boldsymbol{Q}^{(\mathrm{HWFS})}(\omega)S_P(\omega) \\ &= \boldsymbol{V}(\omega)\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{V}_{\mathrm{null}}^{\mathrm{H}}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)\boldsymbol{W}_{\mathrm{ortho}}(\omega)\boldsymbol{U}(\omega) \end{bmatrix}\boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega)S_P(\omega). \end{aligned} \tag{43}$$

5.3. Practical implementation of proposed method

In practice, the directly-calculated inversion of the spatial impedance matrix $\boldsymbol{Z}(\omega)$, which is described in Eq. (33), causes divergence of the secondary source driving function $\boldsymbol{S}_S(\omega)$ due to the matrix dimensionality and potentially ill-conditioning of $\boldsymbol{Z}(\omega)$. A fundamental methods to circumvent the ill-conditioning of $\boldsymbol{Z}(\omega)$ which is called *regularization*. The regularization is used to derive a new problem with a well-conditioned rank deficient matrix $\hat{\boldsymbol{Z}}(\omega)$. A fundamental result of this problem is obtained by the SVD of $\hat{\boldsymbol{Z}}(\omega)$ that is the closest rank-R_T approximation of $\boldsymbol{Z}(\omega)$ in Eq. (33). This method is called TSVD [47, 48]. The TSVD expansion of $\boldsymbol{Z}(\omega)$ is given by

$$\hat{\boldsymbol{Z}}(\omega) = \sum_{m=1}^{R_T} \sigma_m(\omega)\boldsymbol{u}_m(\omega)\boldsymbol{v}_m^{\mathrm{H}}(\omega) \quad (R_T \leq M), \tag{44}$$

where R_T is the desecrate regularization parameter, $\boldsymbol{u}_i(\omega)$ is the ith left singular vector of $\boldsymbol{Z}(\omega)$ which is the column of $\boldsymbol{U}(\omega)$, $\boldsymbol{v}_i(\omega)$ is the ith right singular vector

of $\boldsymbol{Z}(\omega)$ which is the column of $\boldsymbol{V}(\omega)$. In addition, the rank deficient generalized inverse matrix of $\boldsymbol{Z}(\omega)$ can be written as follows:

$$\hat{\boldsymbol{Z}}^{-}(\omega) = \boldsymbol{V}(\omega) \begin{bmatrix} \hat{\boldsymbol{\Lambda}}(\omega) \\ \hat{\boldsymbol{S}}(\omega) \end{bmatrix} \boldsymbol{U}^{\mathrm{H}}(\omega), \tag{45}$$

where $\hat{\boldsymbol{\Lambda}}(\omega)$ is

$$\hat{\boldsymbol{\Lambda}}(\omega) = \mathrm{diag}\left[\hat{\lambda}_1(\omega), \ldots, \hat{\lambda}_M(\omega)\right], \tag{46}$$

$$\hat{\lambda}_m(\omega) = \begin{cases} \frac{1}{\sigma_m(\omega)} & (\text{if } \sigma_m(\omega) > \epsilon\sigma_1(\omega)), \\ 0 & (\text{otherwise}), \end{cases} \tag{47}$$

where ϵ is a condition number of truncation and the desecrate regularization parameter R_T is defined by Eq. (47) as $\underset{m}{\mathrm{argmax}} \left\{\hat{\lambda}_m(\omega) \neq 0\right\}$. From Eq. (45), the secondary source driving functions of proposed method are described in Eq. (43) can be rewritten as

$$\begin{aligned} \hat{\boldsymbol{S}}_S(\omega) &= \boldsymbol{Q}^{(\mathrm{HWFS})}(\omega) S_P(\omega) \\ &= \boldsymbol{V}(\omega) \begin{bmatrix} \hat{\boldsymbol{\Lambda}}(\omega) \\ \hat{\boldsymbol{V}}_{\mathrm{null}}^{\mathrm{H}}(\omega) \boldsymbol{Q}^{(\mathrm{WFS})}(\omega) \boldsymbol{W}_{\mathrm{ortho}}(\omega) \boldsymbol{U}(\omega) \end{bmatrix} \boldsymbol{U}^{\mathrm{H}}(\omega) \boldsymbol{W}(\omega) S_P(\omega), \end{aligned} \tag{48}$$

where $\hat{\boldsymbol{V}}_{\mathrm{null}}(\omega)$ is a matrix composed of the last $M - R_T$ columns of $\boldsymbol{V}(\omega)$.

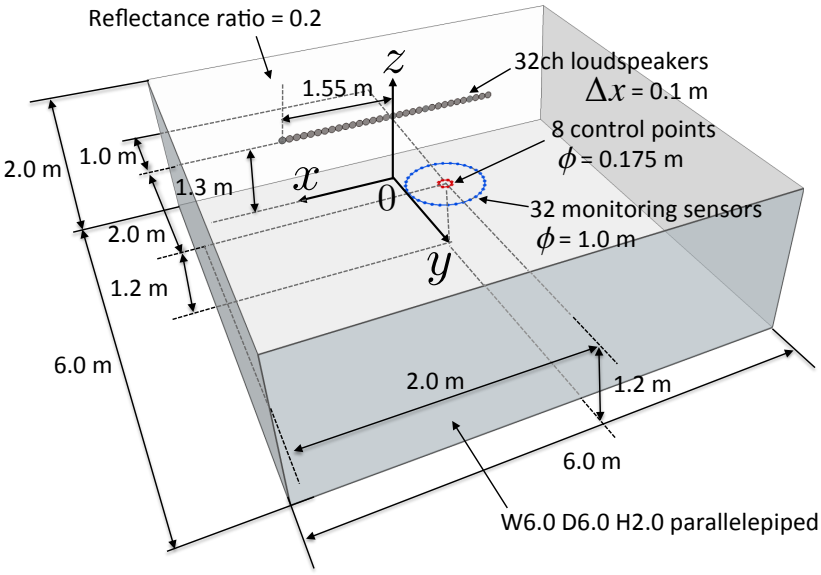


Figure 32. Configurations of loudspeaker array, control points, monitoring sensors and listening room.

5.4. Computer-based simulation and discussion

5.4.1 Simulation conditions and evaluation criteria

To illustrate the properties of the proposed method, the frequency domain and spatial domain descriptions of the synthesized wavefront are used for numerical simulations. The configuration of the simulation system is shown in Fig. 32. The simulation was conducted via 32ch linear-array loudspeakers for reproduction in a free-field and listening room situation. The truncation parameter ϵ is set to 10^{-2}. The reflected waves under the room condition are regarded as direct waves generated from imaginary sources obtained by the *image method* [46]. In this numerical simulation, we use the imaginary sources up to the second order. To show the wide applicability of the proposed method in the spatial domain, we evaluate the sound field reproduction error

$$E_S(x, y, \omega) = \frac{|\mathrm{SWF}(x, y, \omega) - \mathrm{PWF}(x, y, \omega)|^2}{|\mathrm{SWF}(x, y, \omega)|^2}, \tag{49}$$

where $\mathrm{PWF}(x, y, \omega)$ denotes the primary wavefront radiated by primary sources and $\mathrm{SWF}(x, y, \omega)$ denotes the synthesized secondary wavefront.

To evaluate the accuracy of the reproduction at the control points, we compared the normalized quadratic reproduction error of the proposed method with that of the MCWS at the control points using [49, 50]

$$E_{LS}(\omega) = \sum_{x_O, y_O} \frac{|\mathrm{SWF}(x_O, y_O, \omega) - \mathrm{PWF}(x_O, y_O, \omega)|^2}{|\mathrm{SWF}(x_O, y_O, \omega)|^2}, \tag{50}$$

where x_O and y_O denote the coordinates of the observation point, including the locations of the control points and monitoring sensors. We calculate $E_{LS}(\omega)$ at the monitoring sensor location to evaluate the accuracy of the proposed method outside the sweet spot.

5.4.2 Simulation results

Figures 33(a) and 33(b) show the wave fields synthesized by MCWS and the proposed method in the free-field for a monopole source at position $(x_P, y_P, z_P) = (0.0, -1.0, 1.3)$ [m], where the radiated signal frequency is 1600 Hz. The evaluated wavefront frequency of 1600 Hz is the upper limit of major cues for sound source localization. Figures 33(c) and 33(d) show the values of $E_S(x, y, \omega)$ corresponding to Figs. 33(a) and 33(b), respectively. In Figs. 33(c) and 33(d), the dark areas represent high reproduction accuracy with a small error. In addition, the black solid line indicates the level of $E_S = -6.0$ dB, which is suggested to be the acceptable reproduction region limit for comparison purposes [49, 50]. As can be seen in Figs. 33(a) and 33(c), the reproduction error of MCWS is large because the MP-type inverse filter cannot guarantee the correct wavefront outside the control points. In contrast, in Figs. 33(b) and 33(d), the reproduction error of the proposed method is smaller than that of MCWS, and is generally smallest in the vicinity of the control points.

In Figs. 34(a) and 34(b), the $E_{LS}(\omega)$ values of WFS, MCWS and the proposed method at the control points and monitoring sensors are shown. The frequency of 1700 Hz is indicated in these figures, corresponding to the WFS spatial aliasing frequency. In Fig. 34(a), the $E_{LS}(\omega)$ of the proposed method at the control points

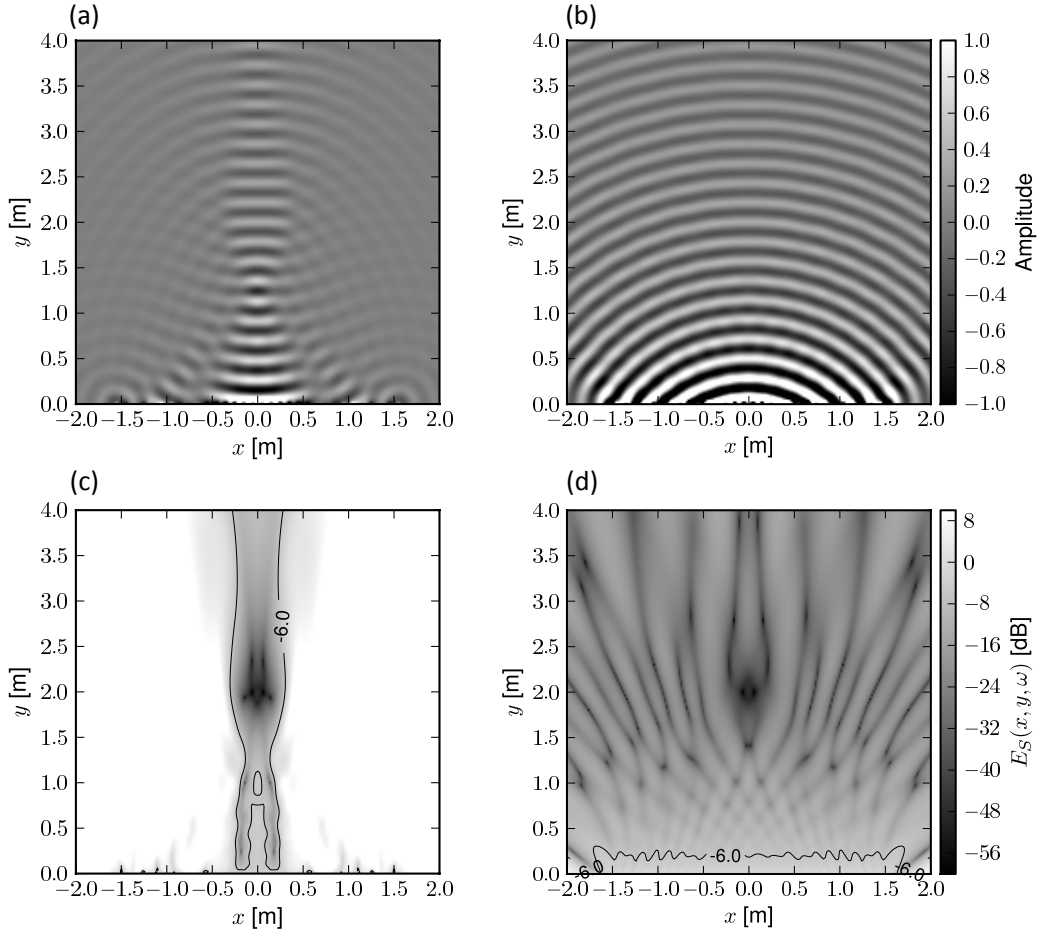


Figure 33. Wavefronts synthesized at 1600 Hz in free-field by (a) MCWS and (b) proposed method. $E_S(\omega)$ values of MCWS and the proposed method are represented in (c) and (d), respectively.

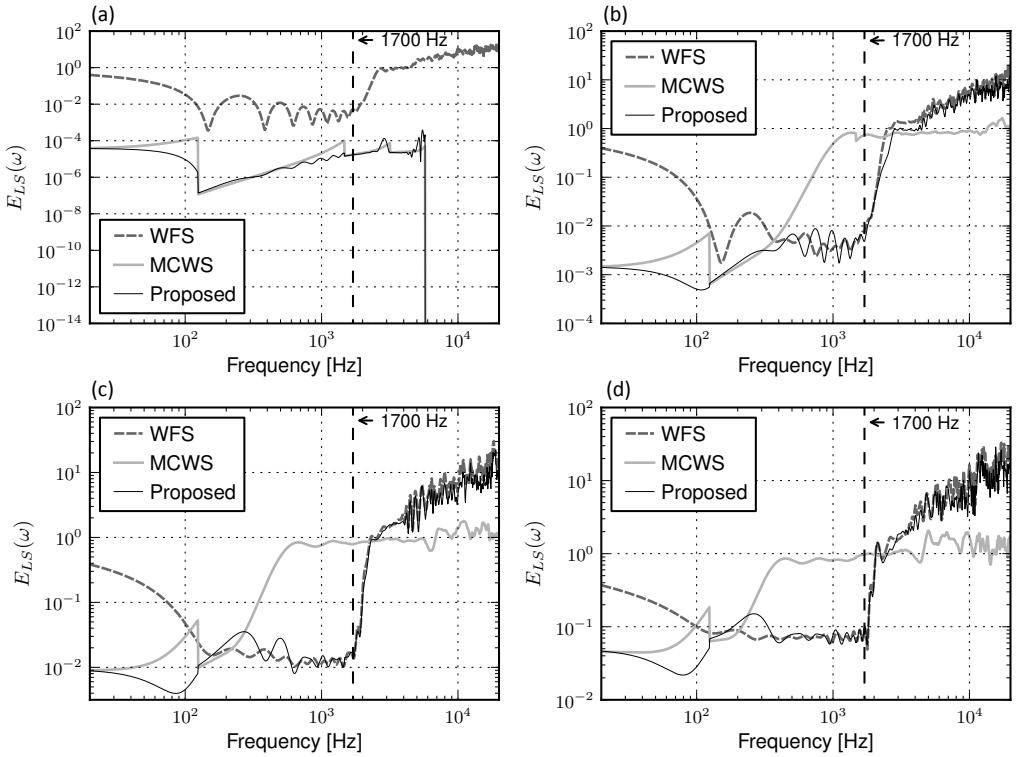


Figure 34. $E_{LS}(\omega)$ values in free-field for conventional methods and proposed method at (a) control points and at monitoring sensors. Diameter of monitoring sensors are (b) $\phi = 1.0$ m, (c) $\phi = 2.0$ m and (d) $\phi = 3.0$ m.

is extremely small compared with that of WFS throughout all the frequencies, and is equivalent to that of MCWS. Thus, the reproduction accuracy at the control points is maintained in the proposed method. In addition, in Fig. 34(b), the $E_{LS}(\omega)$ of the proposed method at the monitoring sensors is smaller than that of MCWS below the spatial aliasing frequency, and is almost equivalent to or smaller than that of WFS. Thus, it is possible to compensate for the lower reproduction accuracy outside the control points using the proposed method. However, in the proposed method, the coloration problem of WFS will occur outside the control points because the error-increasing tendency of the proposed method is similar to that of WFS above the spatial aliasing frequency. In addition, the discontinuity of

$E_{LS}(\omega)$ due to the truncation of singular values in SVD is shown in the results of MCWS and of the proposed method.

Figures 36(a) and 36(b) show the wave fields synthesized by MCWS and the proposed method in the room for a monopole source at position $(x_P, y_P, z_P) = (0.0, -1.0, 1.3)$ [m]. Figures 36(c) and 36(d) show the the $E_S(\omega)$ values for Figs. 36(a) and 36(b), respectively. From Figs. 36(a) and 36(b), we can see that reflected waves in the room disturb the synthesized wavefront. In Figs. 36(c) and 36(d), since interference of room-reflected waves occurs, the acceptable reproduction regions become narrower than the region free-field simulation results; however, the proposed method maintains this region better than with MCWS. Figures 36(a) and 36(b) show the $E_{LS}(\omega)$ values of WFS, MCWS and the proposed method at the control points and monitoring sensors. In Fig. 36(a), the proposed method also maintains the reproduction accuracy at the control points. However, in Fig. 36(b), $E_{LS}(\omega)$ increases outside the control points for all of the methods, and that of the proposed method fluctuates significantly between the results of MCWS and WFS. Thus, as the room reverberation increases, the reproduction errors outside the control points of the proposed method probably approach to the results of MCWS.

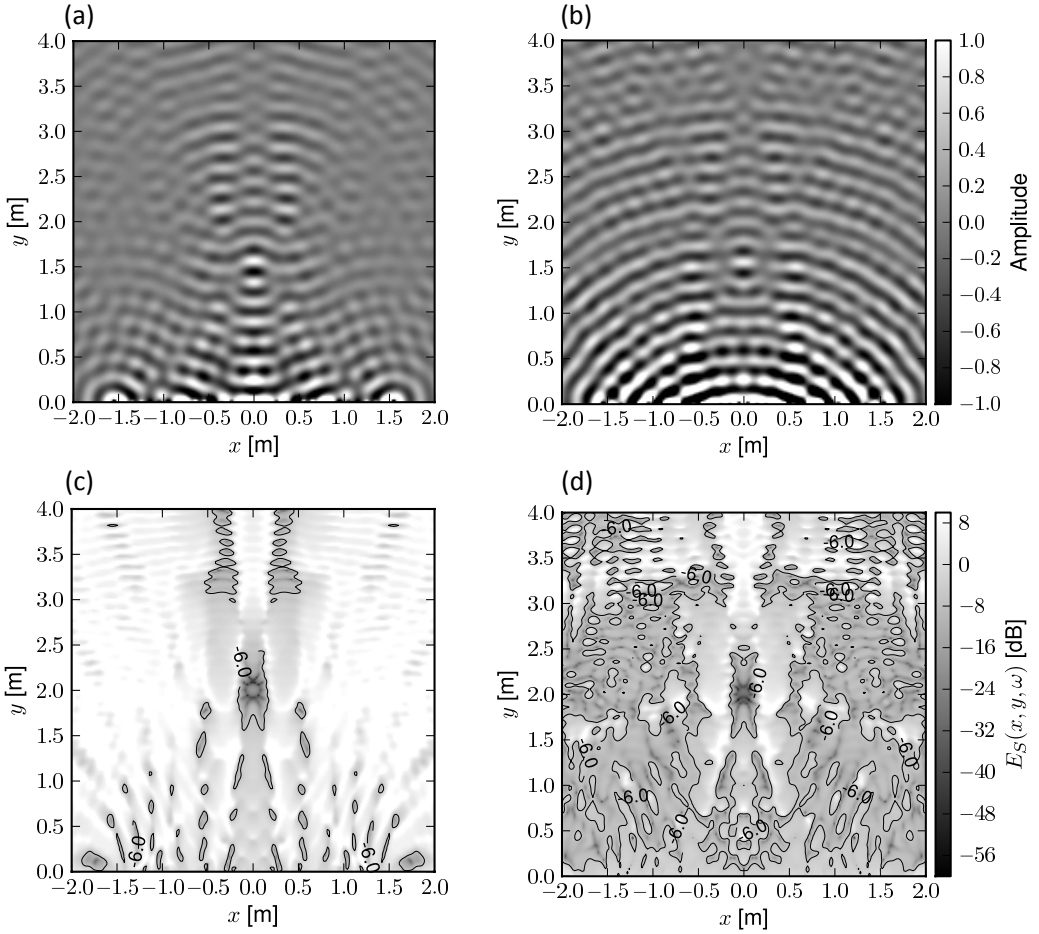


Figure 35. Wavefronts synthesized at 1600 Hz in virtual room by (a) MCWS and (b) proposed method. $E_S(\omega)$ values of MCWS and the proposed method are represented in (c) and (d).

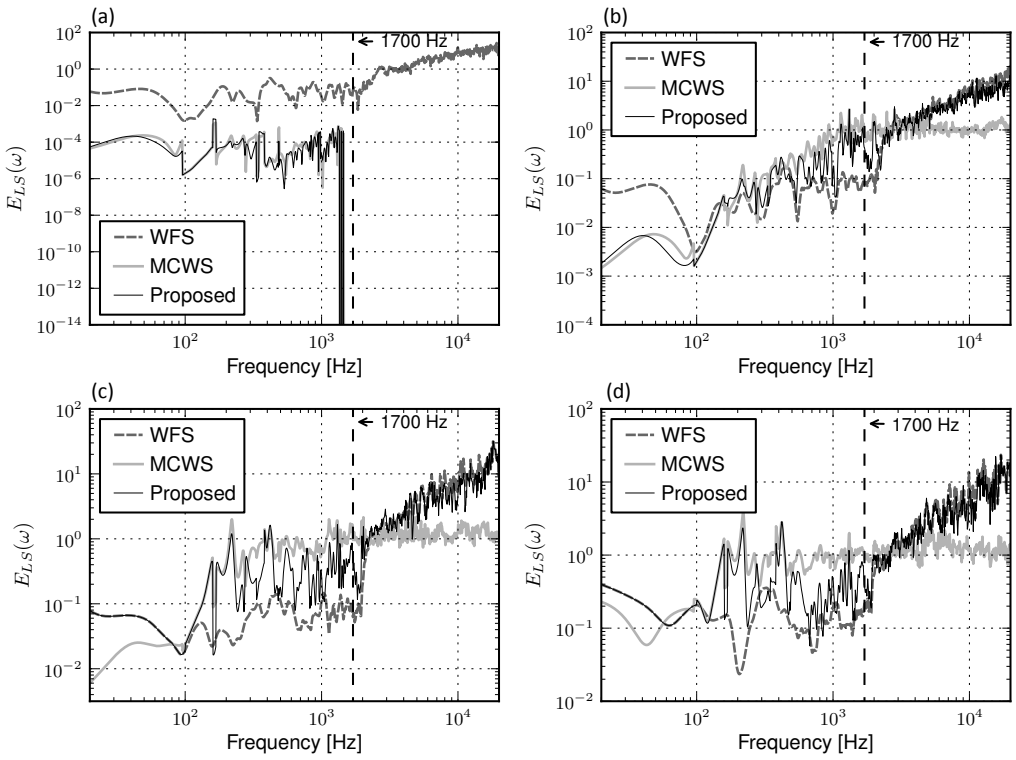


Figure 36. $E_{LS}(\omega)$ values in reverberant room for conventional methods and proposed method at (a) control points and at monitoring sensors. Diameter of monitoring sensors are (b) ϕ = 1.0 m, (c) ϕ = 2.0 m and (d) ϕ = 3.0 m.

5.5. Real-apparatus-based experiment

5.5.1 Experimental conditions and evaluation criteria

To illustrate the properties of the proposed method in an actual environment, the frequency domain and spatial domain descriptions of the synthesized wavefront are used for experiments. The configuration of the acoustic system and the measurement system are shown in Fig. 37. The secondary sources are conducted via 25ch linear-array loudspeakers ($N = 25$), and the loudspeakers are live-sound SM-1A04S loudspeakers. To observe and to visualize the synthesized wavefront, I use the wavefront measurement system [51]. The measurement system uses 10 microphones for the measurment, and the microphones are audio-technica ATM14a omnidirectional microphones. The width and height of the observation area are 3.0 m on each axis. To verify the wide applicability of the proposed method in the spatial domain, I calculate the direct wavefront property of the synthesized secondary wavefront

$$\mathrm{SWF}(x_O, y_O, \omega) = \hat{\boldsymbol{Z}}(x_O, y_O, \omega)\boldsymbol{Z}^{-}_{\mathrm{opt}}(\omega)\boldsymbol{W}(\omega) * H(\omega) \tag{51}$$

where x_O and y_O denote the coordinates of the observation point, including the locations of the control points and monitoring sensors, $H(\omega)$ denotes the window function for removal of the wavefront reflected from the room wall surface, $*$ is the convolution operator, and $\hat{\boldsymbol{Z}}(x_O, y_O, \omega)$ denotes the spatial impedance vector which can be written as

$$\hat{\boldsymbol{Z}}(x, y, \omega) = \left[\hat{Z}_1(x, y, \omega), \cdots, \hat{Z}_N(x, y, \omega)\right], \tag{52}$$

where $\hat{Z}_n(x, y, \omega)$ denotes the spatial impedance between the nth secondary source and the point geometry (x, y) in measurement environment. In this dissertation, I designate the frequency property of the Hanning window as $H(\omega)$. To evaluate the accuracy of the reproduction at the control points, I compared the normalized quadratic reproduction error of the proposed method with that of the MCWS at the control points using [49, 50]

$$E_{LS}(\omega) = \sum_{x,y} \frac{|\mathrm{SWF}(x, y, \omega) - \mathrm{PWF}(x, y, \omega)|^2}{|\mathrm{SWF}(x, y, \omega)|^2}, \tag{53}$$

where PWF(x, y, ω) denotes the calculated primary point source wavefront at the evaluation point (x, y).

5.5.2 Experimental results

Figure 38 shows the normalized quadratic reproduction error $E_{LS}(\omega)$ of the proposed method with the reference wavefront of monopole source at position $(x_P, y_P, z_P) = (0.0, -1.0, 1.2)$ [m] at the control points. Figures 39(a) and 39(b) respectively show the wave fields synthesized by multi-point controlled wavefront synthesis (MCWS) [51] and the proposed method in the measurement room for a monopole source, where the radiated signal frequency is 1600 Hz. The evaluated wavefront frequency of 1600 Hz is the upper limit of major cues for sound source localization. As can be seen in Fig. 39(a), the reproduction error of MCWS is large because the commonly used inverse filter cannot guarantee the correct wavefront outside the control points. In contrast, in Figs. 39(a) and 39(b), the reproduction error of the proposed method is smaller than that of MCWS, and is generally smallest in the vicinity of the control points. By using WFS, the proposed method overcomes the compensation limitation of auditory distance and azimuth perception outside the control points.

The results of measurement revealed that the proposed method balances the above goals and has wide applicability in a spatial domain with high accuracy of reproduction in actual environment.

5.6. Summary of this Chapter

In this chapter, I proposed a new method for balancing the listening area and reproduction accuracy with absolute accuracy by using an inverse filter of the room acoustics: the null space of the generalized inverse matrix given by a compensation filter of the wave field outside the control points. To develop an expression for the compensation filter, I used the loudspeaker driving function of WFS instead of the filter used in conventional studies. By using WFS, the proposed method overcomes the compensation limitation of auditory distance and azimuth percep-

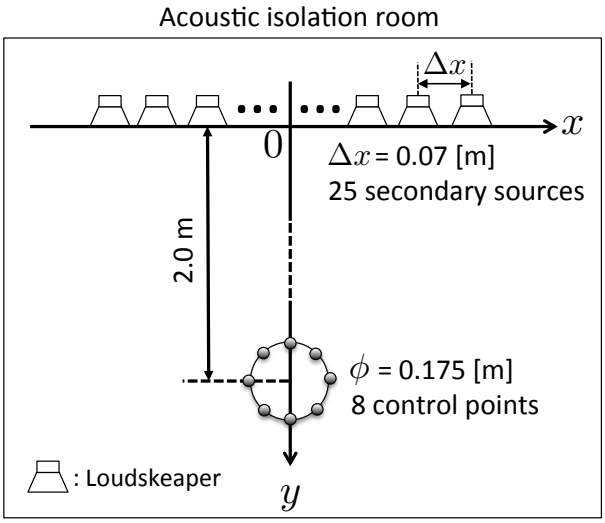


Figure 37. Configurations of loudspeaker array and control points.

tion outside the control points. The results of computer simulations revealed that the proposed method balances the above goals and has wide applicability in a spatial domain with high accuracy of reproduction both under free-field conditions and in a simulation model with room reflection. Next, I evaluated our proposed method through the wavefront measurement. We measured the wavefront generated by our proposed method to evaluate its efficiencies. The measurement results showed that our proposed method can synthesize the primary wavefront with a high degree of accuracy of the reproduction at the control-points. Finally, these results revealed that the proposed method balances the competing goals of reproduction accuracy and a wide listening area, and has wide applicability in a spatial domain with a high accuracy of reproduction in the actual environment.

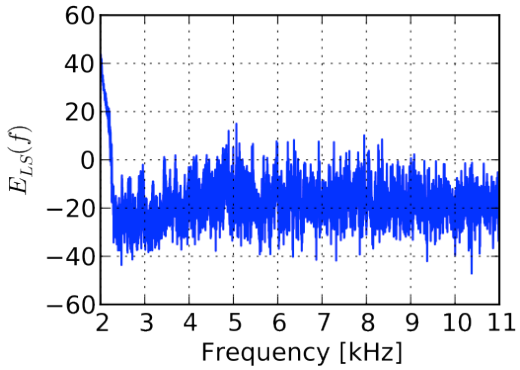


Figure 38. Normalized quadratic reproduction error $E_{LS}(\omega)$ of proposed method with reference wavefront of monopole source at the control points.

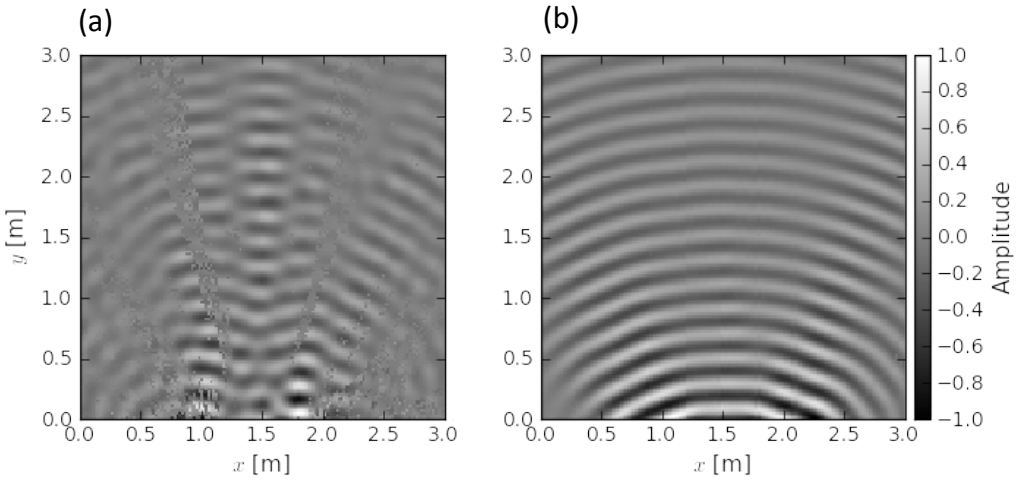


Figure 39. Measured direct wavefront synthesized at 1600 Hz in acoustic isolation room by (a) MCWS and (b) proposed method.

CHAPTER 6

HYBRID WAVE FIELD SYNTHESIS WITH EFFORT VARIATION REGULARIZATION BASED ON GENERALIZED SINGULAR VALUE DECOMPOSITION

6.1. Introduction

In the previous chapter, HWFS was proposed as a new integration method combining multi-point sound field reproduction and WFS. Although the theory of conventional sound field reproduction methods, such as adaptive wave field synthesis (AWFS) [52] and multi-point controlled wavefront synthesis with effort variation regularization (MCWS-EVR) [53], has been well studied, a comparative study on HWFS with these conventional methods has not yet been carried out.

This chapter begins with a review and discussion of related studies on sound field reproduction methods without HWFS. Next, the mathematical relationship and differences between the conventional methods and HWFS are given with a new interpretation based on generalized singular value decomposition (GSVD) [34, 35, 36, 37]. Furthermore, a method that overcomes the degradation of reproduction accuracy of the conventional methods is presented.

6.2. Generalized singular value decomposition

Generalized singular value decomposition (GSVD) was first introduced by Van Loan [34] and has become one of the essential matrix decompositions in recent studies. GSVD is a useful way to explain and solve several least squares problems. In this section, I use the extended GSVD without loss of generality of the dimensions of the decomposed matrix, which was described in [35, 36, 37], to apply GSVD to several inverse problems in sound field reproduction.

Let us consider the joint decomposition of two matrices $\boldsymbol{Z} \in \mathbb{C}^{(M \times N)}$ and $\boldsymbol{D} \in$

$\mathbb{C}^{(P\times N)}$ for any arbitrary natural numbers M, P and N. The GSVD of $\boldsymbol{Z}$ and $\boldsymbol{D}$ is the joint factorization $\left[\boldsymbol{Z}^{\mathrm{T}}, \boldsymbol{D}^{\mathrm{T}}\right]^{\mathrm{T}}$, and the results of decomposition are expressed as

$$\underset{(M\times N)}{\boldsymbol{Z}} = \underset{(M\times M)}{\boldsymbol{U}}\,\underset{(M\times N)}{\boldsymbol{\Sigma}_1}\,\underset{(N\times N)}{\boldsymbol{R}}\,\underset{(N\times N)}{\boldsymbol{Q}^{\mathrm{H}}}, \tag{54}$$

$$\underset{(P\times N)}{\boldsymbol{D}} = \underset{(P\times P)}{\boldsymbol{V}}\,\underset{(P\times N)}{\boldsymbol{\Sigma}_2}\,\underset{(N\times N)}{\boldsymbol{R}}\,\underset{(N\times N)}{\boldsymbol{Q}^{\mathrm{H}}}, \tag{55}$$

where $\boldsymbol{U} \in \mathbb{C}^{(M\times M)}$, $\boldsymbol{V} \in \mathbb{C}^{(P\times P)}$ and $\boldsymbol{Q} \in \mathbb{C}^{(N\times N)}$ are unitary matrices, $\boldsymbol{R} \in \mathbb{C}^{(N\times N)}$ is a nonsingular upper triangular matrix and Σ_1 and Σ_2 are one of the following structures.

1. If $M \geq N$ and $P \geq N$,

$$\boldsymbol{\Sigma}_1 = \begin{array}{c} \\ N \\ M-N \end{array}\!\!\begin{array}{c} \begin{array}{c} N \end{array} \\ \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{0} \end{bmatrix} \end{array}, \quad \boldsymbol{\Sigma}_2 = \begin{array}{c} \\ N \\ P-N \end{array}\!\!\begin{array}{c} \begin{array}{c} N \end{array} \\ \begin{bmatrix} \boldsymbol{S} \\ \boldsymbol{0} \end{bmatrix} \end{array}. \tag{56}$$

2. If $M \geq N$ and $P < N$,

$$\boldsymbol{\Sigma}_1 = \begin{array}{c} \\ N-P \\ P \\ M-N \end{array}\!\!\begin{array}{c} \begin{array}{cc} N-P & P \end{array} \\ \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \end{array}, \quad \boldsymbol{\Sigma}_2 = \begin{array}{c} \\ P \end{array}\!\!\begin{array}{c} \begin{array}{cc} N-P & P \end{array} \\ \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S} \end{bmatrix} \end{array}. \tag{57}$$

3. If $M < N$ and $P \geq N$,

$$\boldsymbol{\Sigma}_1 = \begin{array}{c} \\ M \end{array}\!\!\begin{array}{c} \begin{array}{cc} M & N-M \end{array} \\ \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \end{array}, \quad \boldsymbol{\Sigma}_2 = \begin{array}{c} \\ M \\ N-M \\ P-N \end{array}\!\!\begin{array}{c} \begin{array}{cc} M & N-M \end{array} \\ \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \end{array}. \tag{58}$$

4. If $M < N$, $P < N$,

$$\boldsymbol{\Sigma}_1 = \begin{array}{c} \\ N-P \\ M-(N-P) \end{array} \begin{array}{c} \begin{array}{ccc} N-P & M-(N-P) & N-M \end{array} \\ \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \end{array}, \tag{59}$$

$$\boldsymbol{\Sigma}_2 = \begin{array}{c} \\ M-(N-P) \\ N-M \end{array} \begin{array}{c} \begin{array}{ccc} N-P & M-(N-P) & N-M \end{array} \\ \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \end{array}, \tag{60}$$

where $\boldsymbol{C}$ and $\boldsymbol{S}$ are matrices whose diagonal entries α_i in $\boldsymbol{C}$ and β_i in $\boldsymbol{S}$ are nonnegative real values, which are collectively called *generalized singular value pairs*, and the ratios $\sigma_i = \alpha_i/\beta_i$ are called *generalized singular values*, which are analogous to singular values.

In addition, Van Loan's letter form of GSVD is expressed as

$$\underset{(M\times N)}{\boldsymbol{Z}} = \boldsymbol{U}\,\boldsymbol{\Sigma}_1\,\boldsymbol{X}^{-1}, \tag{61}$$

$$\underset{(P\times N)}{\boldsymbol{D}} = \boldsymbol{V}\,\boldsymbol{\Sigma}_2\,\boldsymbol{X}^{-1}. \tag{62}$$

The GSVD form given in Eqs. (54) and (55) can be converted to Van Loan's form by taking the nonsingular matrix $\boldsymbol{X}$ as

$$\underset{(N\times N)}{\boldsymbol{X}} = \boldsymbol{Q}\boldsymbol{R}^{-1}. \tag{63}$$

6.3. Least squares problem with equality-constrained and unconstrained weighting approximation

In several sound field reproduction methods, to reproduce the primary sound field at the secondary sound field, the multiple-input/output inverse theorem (MINT) is often used to solve the inverse problem of a spatial impedance matrix given by the room transfer functions between the secondary source loudspeakers and the control-point microphones. To ensure the stable controllability of the secondary sound field at the control-point microphones, the MINT-based method must satisfy the condition that the number of control-points M is less than the number of

secondary sources N. Linear problems under such conditions are called underdetermined problems. Solving this type of underdetermined problem requires the use of a regularization for which there is no unique solution; thus, the problem is singular and additional assumptions have to be made to reconstruct a unique meaningful solution.

In recent studies, various regularization techniques have been applied to underdetermined problems in sound field reproduction. The regularization techniques that have been used in recent approaches have been able to obtain an approximate form of the least-squares minimization with a quadratic equality constraint (LSE) [54]. The LSE problem is defined as follows:

$$\boldsymbol{Q}^{(\mathrm{D})}(\omega) = \min_{\boldsymbol{Z}(\omega)\boldsymbol{Q}(\omega)=\boldsymbol{W}(\omega)} \|\boldsymbol{D}(\omega)\boldsymbol{Q}(\omega) - \boldsymbol{H}(\omega)\|_{\mathrm{F}}\,, \tag{64}$$

where the dimensions of the matrices depends on those of each symbol defined in the previous section. The approximated LSE problem can be considered as the minimization problem of the quadratic cost function $J_\mu(\boldsymbol{Q}(\omega))$ for the column vector formed from driving functions of the secondary sources $\boldsymbol{Q}(\omega)$. $J_\mu(\boldsymbol{Q}(\omega))$ is given by

$$J_\mu(\boldsymbol{Q}(\omega)) = \boldsymbol{E}_M^{\mathrm{H}}(\omega)\boldsymbol{E}_M(\omega) + \mu\boldsymbol{E}_C^{\mathrm{H}}(\omega)\boldsymbol{E}_C(\omega), \tag{65}$$

$$\boldsymbol{E}_M(\omega) = \boldsymbol{Z}(\omega)\boldsymbol{Q}(\omega) - \boldsymbol{W}(\omega), \tag{66}$$

$$\boldsymbol{E}_C(\omega) = \boldsymbol{D}(\omega)\boldsymbol{Q}(\omega) - \boldsymbol{H}(\omega), \tag{67}$$

where the positive scalar value μ is a penalty term, $\boldsymbol{D}(\omega)$ is the smoothing matrix of the sharp edges of the exact solution [55] and $\boldsymbol{H}(\omega)$ is a vector related to the a priori known solution. From Eqs. (64) and (65), the approximated LSE problem can be written in matrix form as

$$\begin{aligned}\boldsymbol{Q}_\mu^{(\mathrm{D})}(\omega) &= \min_{\boldsymbol{Q}(\omega)} \left\| J_\mu(\boldsymbol{Q}(\omega)) \right\|_{\mathrm{F}} \\ &= \min_{\boldsymbol{Q}(\omega)} \left\| \begin{bmatrix} \boldsymbol{Z}(\omega) \\ \mu\boldsymbol{D}(\omega) \end{bmatrix} \boldsymbol{Q}(\omega) - \begin{bmatrix} \boldsymbol{W}(\omega) \\ \mu\boldsymbol{H}(\omega) \end{bmatrix} \right\|_{\mathrm{F}}.\end{aligned} \tag{68}$$

The GSVD of $[\boldsymbol{Z}(\omega)^{\mathrm{T}}\boldsymbol{D}(\omega)^{\mathrm{T}}]^{\mathrm{T}}$ sheds light on the approximate solution of these problems. In addition, in the LSE problem of sound field reproduction, the number

of rows and columns of $\boldsymbol{H}(\omega)$ and $\boldsymbol{D}(\omega)$, P, is often set to N. Therefore, the exact solution of Eq. (64) is given in [56] as

$$\boldsymbol{Q}^{(\mathrm{D})}(\omega) = \sum_{i=1}^{M} \frac{\boldsymbol{u}_i^{\mathrm{H}} \boldsymbol{W}(\omega)}{\alpha_i} \boldsymbol{x}_i + \sum_{i=M+1}^{N} \frac{\boldsymbol{v}_i^{\mathrm{H}} \boldsymbol{H}(\omega)}{\beta_i} \boldsymbol{x}_i, \tag{69}$$

and the approximate weighting solution of Eq. (64) is

$$\boldsymbol{Q}_{\mu}^{(\mathrm{D})}(\omega) = \sum_{i=1}^{M} \frac{\alpha_i \boldsymbol{u}_i^{\mathrm{H}} \boldsymbol{W}(\omega)}{\alpha_i^2 + \mu^2 \beta_i^2} \boldsymbol{x}_i + \sum_{i=1}^{M} \frac{\mu^2 \beta_i \boldsymbol{v}_i^{\mathrm{H}} \boldsymbol{H}(\omega)}{\alpha_i^2 + \mu^2 \beta_i^2} \boldsymbol{x}_i + \sum_{i=M+1}^{N} \frac{\boldsymbol{v}_i^{\mathrm{H}} \boldsymbol{H}(\omega)}{\beta_i} \boldsymbol{x}_i, \tag{70}$$

where $\boldsymbol{U}(\omega) = [\boldsymbol{u}_1(\omega), \ldots, \boldsymbol{u}_M(\omega)]$, $\boldsymbol{V}(\omega) = [\boldsymbol{v}_1(\omega), \ldots, \boldsymbol{v}_P(\omega)]$ and $\boldsymbol{X}(\omega) = [\boldsymbol{x}_1(\omega), \ldots, \boldsymbol{x}_N(\omega)]$. Since

$$\boldsymbol{Q}^{(\mathrm{D})}(\omega) - \boldsymbol{Q}_{\mu}^{(\mathrm{D})}(\omega) = \sum_{i=1}^{M} \frac{\boldsymbol{u}_i^{\mathrm{H}} \boldsymbol{W}(\omega)}{\alpha_i} \boldsymbol{x}_i - \sum_{i=1}^{M} \frac{\alpha_i \boldsymbol{u}_i^{\mathrm{H}} \boldsymbol{W}(\omega)}{\alpha_i^2 + \mu^2 \beta_i^2} \boldsymbol{x}_i - \sum_{i=1}^{M} \frac{\mu^2 \beta_i \boldsymbol{v}_i^{\mathrm{H}} \boldsymbol{H}(\omega)}{\alpha_i^2 + \mu^2 \beta_i^2} \boldsymbol{x}_i, \tag{71}$$

it follows that $\boldsymbol{Q}_{\mu}^{(\mathrm{D})}(\omega) \to \boldsymbol{Q}^{(\mathrm{D})}(\omega)$ as $\mu \to 0$. From Eq. (70), the solution $\boldsymbol{Q}_{\mu}^{(\mathrm{D})}(\omega)$ can be rewritten in matrix form as

$$\underset{(N\times M)}{\boldsymbol{Q}_{\mu}^{(\mathrm{D})}(\omega)} = \underset{(N\times M)}{\boldsymbol{X}(\omega)\boldsymbol{\Xi}_1(\omega)} \boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega) + \underset{(N\times P)}{\boldsymbol{X}(\omega)\boldsymbol{\Xi}_2(\omega)} \boldsymbol{V}^{\mathrm{H}}(\omega)\boldsymbol{H}(\omega), \tag{72}$$

where $\boldsymbol{\Xi}_1(\omega)$ and $\boldsymbol{\Xi}_2(\omega)$ are the inverse matrices of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively, which are given by

$$\boldsymbol{\Sigma}_1 = \begin{matrix} \\ M \end{matrix} \begin{matrix} \begin{matrix} M & N-M \end{matrix} \\ \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \end{matrix}, \quad \boldsymbol{\Sigma}_2 = \begin{matrix} \begin{matrix} M \\ N-M \end{matrix} \\ \end{matrix} \begin{matrix} \begin{matrix} M & N-M \end{matrix} \\ \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \end{matrix}. \tag{73}$$

Hence, $\boldsymbol{\Xi}_1(\omega)$ and $\boldsymbol{\Xi}_2(\omega)$ are given by

$$\underset{(N\times M)}{\boldsymbol{\Xi}_1(\omega)} = \left[\begin{array}{ccc} \xi_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \xi_M \\ \hline & \boldsymbol{0} & \end{array}\right], \tag{74}$$

$$\underset{(N\times P)}{\boldsymbol{\Xi}_2(\omega)} = \left[\begin{array}{ccc|c} \zeta_1 & & & \\ & \ddots & & \boldsymbol{0} \\ & & \zeta_M & \\ \hline & \boldsymbol{0} & & \boldsymbol{I} \end{array}\right], \tag{75}$$

where the diagonal entries of $\boldsymbol{\Xi}_1$ and $\boldsymbol{\Xi}_2$ are expressed as

$$\xi_i = \frac{\alpha_i}{\alpha_i^2 + \mu^2\beta_i^2} = \frac{\sigma_i}{\sigma_i^2 + \mu^2}\frac{1}{\beta_i}, \tag{76}$$

$$\zeta_i = \frac{\mu^2\beta_i}{\alpha_i^2 + \mu^2\beta_i^2} = \frac{\mu^2}{\sigma_i^2 + \mu^2}\frac{1}{\beta_i}. \tag{77}$$

6.4. Effort regularization based on GSVD

Effort regularization is the most common form of Tikhonov regularization used in inverse filtering techniques in sound field reproduction. In the context of sound field control, it is used to circumvent any ill-conditioning of the problem and is conceptually connected to the leaky adaptation of a feedforward system. In sound reproduction, it has been shown that effort regularization can affect the lifetime of the control filters.

Using this method, the minimization of the reproduction errors is accomplished using a cost function that is a special case of Eq. (65) with $\boldsymbol{D}(\omega) = \boldsymbol{I}$, $\boldsymbol{H}(\omega) = \boldsymbol{0}$ and $N = P$,

$$J_\mu(\boldsymbol{Q}(\omega)) = \boldsymbol{E}_M^{\mathrm{H}}(\omega)\boldsymbol{E}_M(\omega) + \boldsymbol{Q}^{\mathrm{H}}(\omega)\boldsymbol{Q}(\omega).$$

Minimization of this cost function is guaranteed to provide a solution even in the case of an underdetermined or poorly conditioned system. The optimum secondary source driving functions obtained by this regularization are

$$\boldsymbol{Q}_\mu^{(\mathrm{ER})}(\omega) = (\boldsymbol{Z}^{\mathrm{H}}(\omega)\boldsymbol{Z}(\omega) + \mu\boldsymbol{I})^{-1}\boldsymbol{Z}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega). \tag{78}$$

In addition, from the decomposition of the smoothing matrix $\boldsymbol{D}$ given by Eq. (62), we can derive the nonsingular matrix $\boldsymbol{X}(\omega)$ of GSVD as

$$\boldsymbol{D}(\omega) \;\; = \boldsymbol{V}(\omega)\boldsymbol{S}(\omega)\boldsymbol{X}^{-1}(\omega) \Leftrightarrow \boldsymbol{X}(\omega) \;\; = \boldsymbol{V}(\omega)\boldsymbol{S}(\omega). \tag{79}$$

From Eq. (70) and this equation, we can rewrite the optimum secondary source driving functions obtained by effort regularization using GSVD as

$$\boldsymbol{Q}_\mu^{(\mathrm{ER})}(\omega) = \boldsymbol{V}(\omega)\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_1(\omega)\boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega), \tag{80}$$

where the diagonal entries of $\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_1(\omega)$ are given by

$$\beta_i \xi_i = \frac{\sigma_i}{\sigma_i^2 + \mu^2}. \tag{81}$$

This method can be considered as a method based on MCWS with effort regularization (MCWS-ER). One of the sound field reproduction systems based on MCWS-ER is boundary surface control (BoSC) [57].

6.5. Adaptive wave field synthesis based on GSVD

As an example of a recently proposed method, adaptive wave field synthesis (AWFS) is a sound field control method that combines MINT and wave field synthesis. AWFS includes a regularization term that quadratically penalizes deviations from the a-priori known WFS solution. It is based on minimizing the reproduction errors and penalizing deviations from the a-priori WFS solution $\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$ by minimizing a cost function of the form

$$\begin{aligned} J_\mu(\boldsymbol{Q}(\omega)) = {} & \boldsymbol{E}_M^{\mathrm{H}}(\omega)\boldsymbol{E}_M(\omega) \\ & + \mu(\boldsymbol{Q}(\omega) - \boldsymbol{Q}^{(\mathrm{WFS})}(\omega))^{\mathrm{H}}(\boldsymbol{Q}(\omega) - \boldsymbol{Q}^{(\mathrm{WFS})}(\omega)). \end{aligned} \tag{82}$$

Thus, the optimum secondary source driving functions for this regularization are

$$\begin{aligned} \boldsymbol{Q}_\mu^{(\mathrm{AWFS})}(\omega) = {} & \left(\boldsymbol{Z}^{\mathrm{H}}(\omega)\boldsymbol{Z}(\omega) + \mu\boldsymbol{I}\right)^{-1} \\ & (\boldsymbol{Z}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega) + \mu\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)). \end{aligned} \tag{83}$$

In the related literature on AWFS, the secondary loudspeaker array has a circular form and Eq. (83) has been interpreted as independent radiation mode penalization with singular values of a spatial impedance matrix $\boldsymbol{Z}(\omega)$ that has been decomposed by SVD. However, the interpretation does not give a sufficient description of the synthesized wavefront that is generated by summing the penalized independent radiation modes of WFS and MCWS over the signal-space and null-space. These spaces are spanned by the singular vectors that originate from the variation in the regularization parameter μ with an arbitrary constitution of the secondary loudspeaker array. Therefore, in this dissertation, a new interpretation is given

for the fundamental relationship between the sums of the wavefronts in WFS and MCWS based on the theory of AWFS using GSVD.

Taking the AWFS solution given by Eq. (83) and using the GSVD of the spatial impedance matrix $\boldsymbol{Z}(\omega)$ and smoothing matrix $\boldsymbol{D}(\omega) = \boldsymbol{I}$, Eq. (83) can be expanded as

$$\begin{aligned}\boldsymbol{Q}_{\mu}^{(\mathrm{AWFS})}(\omega) = \boldsymbol{V}(\omega)\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_1(\omega)\boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega) \\ + \boldsymbol{V}(\omega)\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_2(\omega)\boldsymbol{V}^{\mathrm{H}}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega),\end{aligned} \tag{84}$$

where the diagonal entries of $\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_2(\omega)$ are given by

$$\beta_i\zeta_i = \frac{\mu^2}{\sigma_i^2 + \mu^2}. \tag{85}$$

6.6. Hybrid wave field synthesis based on GSVD

On the other hand, from Eq. (43), the secondary-wavefront-synthesized filter in HWFS, $\boldsymbol{Q}^{(\mathrm{HWFS})}(\omega)$, can be rewritten as

$$\begin{aligned}\boldsymbol{Q}^{(\mathrm{HWFS})}(\omega) &= \boldsymbol{V}(\omega)\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{V}_{\mathrm{null}}^{\mathrm{H}}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)\boldsymbol{W}_{\mathrm{ortho}}(\omega)\boldsymbol{U}(\omega) \end{bmatrix}\boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega) \\ &= \boldsymbol{V}(\omega)\begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega) + \boldsymbol{V}(\omega)\left[\begin{array}{c|c} \boldsymbol{0}_{\mathrm{null}} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{I}_{\mathrm{null}} \end{array}\right]\boldsymbol{V}^{\mathrm{H}}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega),\end{aligned} \tag{86}$$

where $\boldsymbol{I}_{\mathrm{null}}$ is an $(N - R_T \times N - R_T)$ unit matrix and $\boldsymbol{0}_{\mathrm{null}}$ is an $(R_T \times R_T)$ zero matrix, where R_T is the number of efficient ranks of the truncated spatial impedance matrix $\boldsymbol{Z}_{R_T}(\omega)$. From Eqs. (64) and (86), HWFS can be redefined as an LSE problem with truncated GSVD as follows:

$$\boldsymbol{Q}^{(\mathrm{HWFS})}(\omega) = \min_{\boldsymbol{Z}_{R_T}(\omega)\boldsymbol{Q}(\omega)=\boldsymbol{W}(\omega)} \left\|\boldsymbol{Q}(\omega) - \boldsymbol{Q}^{(\mathrm{WFS})}(\omega)\right\|_{\mathrm{F}}. \tag{87}$$

6.7. Effort variation regularization based on GSVD

In image processing methods, the general form of Tikhonov regularization is employed, where the penalty term $(\boldsymbol{D}(\omega)\boldsymbol{Q}(\omega))^{\mathrm{H}}(\boldsymbol{D}(\omega)\boldsymbol{Q}(\omega))$ of Eq. (65) is based on

a modification of the smoothing matrix $\boldsymbol{D}(\omega)$. In sound field reproduction, recent work has demonstrated the use of effort variation regularization in the problem of global sound equalization [59]. This regularization has helped realize the successful reproduction of a traveling plane wave inside a room by penalizing the square of the norm of variation in the secondary source driving signals.

In a recent study [53], effort variation regularization was based on a cost function of the form

$$\begin{aligned} J_{\mu}(\boldsymbol{Q}(\omega)) = \boldsymbol{E}_{M}^{\mathrm{H}}(\omega)\boldsymbol{E}_{M}(\omega) \\ + \mu(\boldsymbol{D}^{(l)}(\omega)\boldsymbol{Q}(\omega))^{\mathrm{H}}(\boldsymbol{D}^{(l)}(\omega)\boldsymbol{Q}(\omega)), \end{aligned} \tag{88}$$

in which the smoothing matrix $\boldsymbol{D}^{(l)}(\omega)$ represents an lth-order complex differential matrix. The elements of the complex differential operators can be expressed as

$$D_{n,m}^{(1)}(\omega) = \begin{cases} -1 & (\text{if } n = m), \\ d_{n,m}(\omega) & (\text{if } m - n = 1), \\ 0 & (\text{otherwise}), \end{cases} \tag{89}$$

and

$$D_{n,m}^{(2)}(\omega) = \begin{cases} -2 & (\text{if } n = m), \\ d_{n,m}(\omega) & (\text{if } |n - m| = 1), \\ 0 & (\text{otherwise}), \end{cases} \tag{90}$$

for the first-order and second-order differential operators, where

$$\begin{aligned} d_{n,m}(\omega) &= e^{j(\arg(G_n(\omega)) - \arg(G_m(\omega)))} \\ &= e^{-jk(r_{PS_n} - r_{PS_m})} \end{aligned} \tag{91}$$

and $G_n(\omega)$ is the impedance from the primary source to the nth secondary source, and r_{PS_n} and r_{PS_m} are the distances between the primary source and the nth and mth secondary sources, respectively. The resulting first-order and second-order

differential matrices are $N \times N$ square matrices of the form

$$\boldsymbol{D}^{(1)}(\omega) = \begin{bmatrix} -1 & d_{1,2}(\omega) & 0 & \cdots & 0 \\ 0 & -1 & d_{2,3}(\omega) & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & -1 & d_{N-1,N}(\omega) \\ 0 & & \cdots & 0 & -1 \end{bmatrix}, \tag{92}$$

$$\boldsymbol{D}^{(2)}(\omega) = \begin{bmatrix} -2 & d_{1,2}(\omega) & 0 & \cdots & 0 \\ d_{2,1}(\omega) & -2 & d_{2,3}(\omega) & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & d_{N-1,N-2}(\omega) & -2 & d_{N-1,N}(\omega) \\ 0 & & \cdots & d_{N,N-1}(\omega) & -2 \end{bmatrix}. \tag{93}$$

where $d_{n,n\pm1}(\omega)$ is the a-priori phase difference between neighboring reproduction sources n and $n \pm 1$. It can be seen that the minimization of $\|\boldsymbol{D}^{(l)}(\omega)\boldsymbol{Q}(\omega)\|^2$ in the cost function given by Eq. (88) tends to align the complex strengths of neighboring secondary sources n and $n \pm 1$ according to the predefined phase differentials $k(r_{PS_n} - r_{PS_m})$. Assuming that the secondary sources are numbered in ascending order and placed equidistantly,

$$\boldsymbol{Q}_{\mu}^{(\mathrm{EVR})}(\omega) = \left[\mu \boldsymbol{D}^{(l)\mathrm{H}}(\omega)\boldsymbol{D}^{(l)}(\omega) + \boldsymbol{Z}^{\mathrm{H}}(\omega)\boldsymbol{Z}(\omega)\right]^{-1} \boldsymbol{Z}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega). \tag{94}$$

In recent studies, although the solution based on generalized Tikhonov regularization was given, the expanded solution based on GSVD has not been given. Therefore, from Eq. (94), the optimum secondary source driving functions are derived by effort regularization with GSVD and the solution can be rewritten as

$$\boldsymbol{Q}_{\mu}^{(\mathrm{EVR})}(\omega) = \boldsymbol{D}^{(l)^{-1}}(\omega)\boldsymbol{V}(\omega)\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_1(\omega)\boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega). \tag{95}$$

We call this method MCWS with effort variation regularization (MCWS-EVR). From Eq. (95), effort variation regularization can be used to interpret the theory based on MCWS with an anti-phase having an a-priori known phase differential smoothing matrix $\boldsymbol{D}^{(l)^{-1}}(\omega)$. The smoothing term of MCWS-EVR includes an a

priori condition not only for the phase condition of the wavefront generated by the primary wave field but also for the sound pressures at the secondary sources in the primary sound field. The relationship between the optimal reproduced sound pressure at the secondary sources and the synthesized wavefront in front of the secondary sources in reverberant rooms has been investigated [60]. In ref. [60], it was shown that the sound pressures at the secondary sources with an anti-phase having a primary sound field is related to the synthesized sound field. In addition, the anti-phase processing of the first-order and second-order differential matrices in the smoothing term $\boldsymbol{D}^{(l)}(\omega)$ reduces the reverberation modes in the actual environment and significantly increases the area of sweet-spots compared with those for MCWS and MCWS-ER in the actual environment [53].

6.8. Summary of sound field reproduction methods based on GSVD

As mentioned in the previous sections, the conventional and proposed sound field reproduction methods can be generalized by GSVD. In this section, a summary of these sound field reproduction methods based on GSVD is given. Table 5 shows the parameter relationships for each sound field reproduction method based on GSVD.

In MCWS-ER, when $\mu \to \infty$, the wavefront-synthesized filter $\boldsymbol{Q}_{\mu}^{(\mathrm{ER})}(\omega)$ reduces the contributions of the singular values of the signal-space in $\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_1(\omega)$ and when $\mu = \infty$, the synthesized wavefront disappears completely because $\boldsymbol{\Sigma}_2(\omega)$ $\boldsymbol{\Xi}_1(\omega) = \boldsymbol{0}$. This fact raises issues of sensitivity regarding the determination of the regularization parameter μ under conditions that do not require excessive control of the sound field, such as under the free-field condition. Therefore, more sophisticated approaches are needed in MCWS-ER to obtain an approximate solution under the condition of $\mu \to \infty$. In addition, in MCWS-EVR, the same problem of the determination of μ as that in MCWS-ER occurs because if $\mu = 0$, then $\boldsymbol{\Sigma}_2(\omega)\boldsymbol{\Xi}_1(\omega) = \boldsymbol{0}$ in Eq. (95). Therefore, compensation theory is required when applying MCWS-ER under the condition of $\mu = 0$.

Table 5. Summary of parameter relationships for each sound field reproduction method based on GSVD

Name of method	$\boldsymbol{Z}(\omega)$	$\boldsymbol{D}(\omega)$	$\boldsymbol{H}(\omega)$	RP[1]	$\boldsymbol{Q}(\omega)$ (RP = 0)	$\boldsymbol{Q}(\omega)$ (RP = ∞)
MCWS-ER	$\boldsymbol{Z}(\omega)$	$\boldsymbol{I}$	$\boldsymbol{0}$	μ	$\boldsymbol{Q}^{(\mathrm{D})}(\omega)$	$\boldsymbol{0}$
MCWS-EVR	$\boldsymbol{Z}(\omega)$	$\boldsymbol{D}^{(l)}(\omega)$	$\boldsymbol{0}$	μ	$\boldsymbol{Q}^{(\mathrm{D})}(\omega)$	$\boldsymbol{0}$
AWFS	$\boldsymbol{Z}(\omega)$	$\boldsymbol{I}$	$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$	μ	$\boldsymbol{Q}^{(\mathrm{D})}(\omega)$	$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$
HWFS	$\boldsymbol{Z}_{R_T}(\omega)$	$\boldsymbol{I}$	$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$	ϵ	$\boldsymbol{Q}^{(\mathrm{D})}(\omega)$	$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$
HWFS-EVR	$\boldsymbol{Z}_{R_T}(\omega)$	$\boldsymbol{D}^{(l)}(\omega)$	$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$	ϵ	$\boldsymbol{Q}^{(\mathrm{D})}(\omega)$	$\boldsymbol{Q}^{(\mathrm{WFS})}(\omega)$

The first term of Eq. (86) is the driving function of MCWS and the second term corresponds to the embedding filter that compensates the sound field outside the control points designed by WFS. By carefully observing the newly transcribed wavefront-synthesized filter of HWFS in Eq. (86), it can be seen that the wavefront-synthesized filter of HWFS is identical to that of AWFS under the extreme conditions of the regularization parameter of RP = 0 and RP = ∞. This suggests that HWFS is different from AWFS in terms of regularization. Moreover, the accurate mathematical aspect of summing the first and second terms of AWFS in Eq. (84) in comparison with that of Eq. (86) reveals that the embedding part of AWFS affects the part used for sweet-spot control by adjusting the regularization term μ. In contrast, these terms in HWFS are normal to each other. Therefore, in AWFS, when $0 < \mu < \infty$, the sound pressures at the control points are affected by the addition of the sound pressure that is generated by WFS.

In addition, when $\mu \to 0$, the wavefront-synthesized filter of AWFS, $\boldsymbol{Q}_{\mu}^{(\mathrm{AWFS})}(\omega)$, reduces the contribution of the WFS driving function $Q^{(\mathrm{WFS})}$ in the second term of Eq. (84), and when $\mu = 0$, the synthesized wavefront is precisely generated by MCWS-ER. In contrast, when $\mu \to \infty$, the filter reduces the contribution of MCWS-ER, which is given by the first term of Eq. (84), and when $\mu = \infty$, the synthesized wavefront is precisely generated by WFS. These two situations are the limiting cases of AWFS, and the same holds for HWFS (in practice, we use truncated SVD (TSVD) [47, 48] to obtain the driving function of the secondary sources in HWFS in order to maintain the orthogonality of the wavefronts gener-

[1]RP : regularization parameter

ated by the two methods at the control-points). The limiting cases of AWFS and HWFS appear to act as an image (synthesized wavefront) metamorphosis between MCWS and WFS [58].

Finally, the formulation of HWFS is refined by the introduction of a regularization parameter using GSVD to incorporate its improved points over the conventional methods discussed in the previous sections. Considering the characteristics of (a) the sound field metamorphosis between WFS and MCWS in HWFS and (b) the improvement of the sweet-spot characteristics by the anti-phase smoothing of the secondary source driving function using the primary wave field in MCWS-EVR, a new balancing method for achieving both these advantages of HWFS and MCWS-EVR is proposed. The proposed method involving the integration of HWFS and MCWS-EVR is named *hybrid wave field synthesis with effort variation regularization* (HWFS-EVR). The algorithm of the proposed method is described in detail below.

From Eqs. (87) and (95), the mixed problem of HWFS and MCWS-EVR in the LSE problem has the form

$$\boldsymbol{Q}^{(\mathrm{D})}(\omega) = \min_{\mathbf{Z}_{R_T}(\omega)\boldsymbol{Q}(\omega)=\boldsymbol{W}(\omega)} \left\| \boldsymbol{D}^{(l)}(\omega)\boldsymbol{Q}(\omega) - \boldsymbol{D}^{(l)}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega) \right\|_{\mathrm{F}}. \tag{96}$$

From Eq. (86), Eq. (96) can be rewritten after GSVD as

$$\begin{aligned} \boldsymbol{Q}^{(\mathrm{D})}(\omega) = {} & \boldsymbol{D}^{(l)^{-1}}(\omega)\boldsymbol{V}(\omega) \begin{bmatrix} \boldsymbol{\Lambda}(\omega) \\ \mathbf{0} \end{bmatrix} \boldsymbol{U}^{\mathrm{H}}(\omega)\boldsymbol{W}(\omega) \\ & + \boldsymbol{V}(\omega) \left[\begin{array}{c|c} \mathbf{0}_{\mathrm{null}} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{I}_{\mathrm{null}} \end{array}\right] \boldsymbol{V}^{\mathrm{H}}(\omega)\boldsymbol{Q}^{(\mathrm{WFS})}(\omega), \end{aligned} \tag{97}$$

where the first term of Eq. (97) is the same as the driving function of MCWS-EVR and the second term of Eq. (97) is the same as the second term of the HWFS, which is the sound field control function outside the control-points in null-space. Thus, the secondary source driving function, which is the sum of the first and second terms of Eq. (97), includes the advantages of HWFS and MCWS-EVR.

6.9. Computer-based simulation and discussion

6.9.1 Simulation results in free field

To illustrate the properties of the conventional and proposed methods, the frequency domain and spatial domain descriptions of the synthesized wavefront are used for numerical simulations. The wavefront calculation conditions and evaluation criteria are the same as those in Fig. 32, Chapter 5.

Figures 40(a) to 40(d) respectively show the wavefronts synthesized at 1600 Hz under the free-field condition by (a) HWFS, (b) MCWS-EVR, (c) AWFS and (d) the proposed method for a monopole source at position $(x_P, y_P, z_P) = (0.0, -1.0, 1.3)$ [m]. The evaluated wavefront frequency of 1600 Hz is the upper limit of the major cues for sound source localization. Figures 41(a) to 41(d) show the values of $E_S(x, y, \omega)$ corresponding to Figs. 40(a) to 40(d), respectively. In Fig. 41, the dark areas represent high reproduction accuracy with a small error. In addition, the solid red line indicates the contour $E_S(x, y, \omega) = -6.0$ dB, which was suggested to be an acceptable reproduction region limit for comparison purposes as mentioned in the previous chapter. In these calculations, the regularization parameter μ and the differential order of $\boldsymbol{D}(\omega)$ are set equal to 10^{-2} and 2, respectively. As can be seen in Figs. 40(a) and 41(a), the reproduction error of HWFS is smaller than those of the other methods and is generally the smallest in the vicinity of the control points because the null-space-based embedding filter design obtained by WFS does not affect the wavefront inside the control points. In addition, the results of MCWS-EVR shown in Figs. 40(b) and 41(b) indicate that it significantly increases the effective region of the sweet-spots, despite the inefficiency of null-space-embedded filtering methods such as HWFS and AWFS. In contrast, in Figs. 40(c), 40(d), 41(c) and 41(d), the reproduction errors of AWFS and the proposed method are smaller than those of HWFS and MCWS-EVR. However, the reproduction error of the sweet-spots was maintained throughout the observation area in this calculation for all methods.

In Figs. 42(a) to 42(d) and Figs. 43(a) to 43(d), the values of $E_{LS}(\omega)$ for (a) WFS and HWFS, (b) MCWS-EVR, (c) AWFS and (d) the proposed method at the

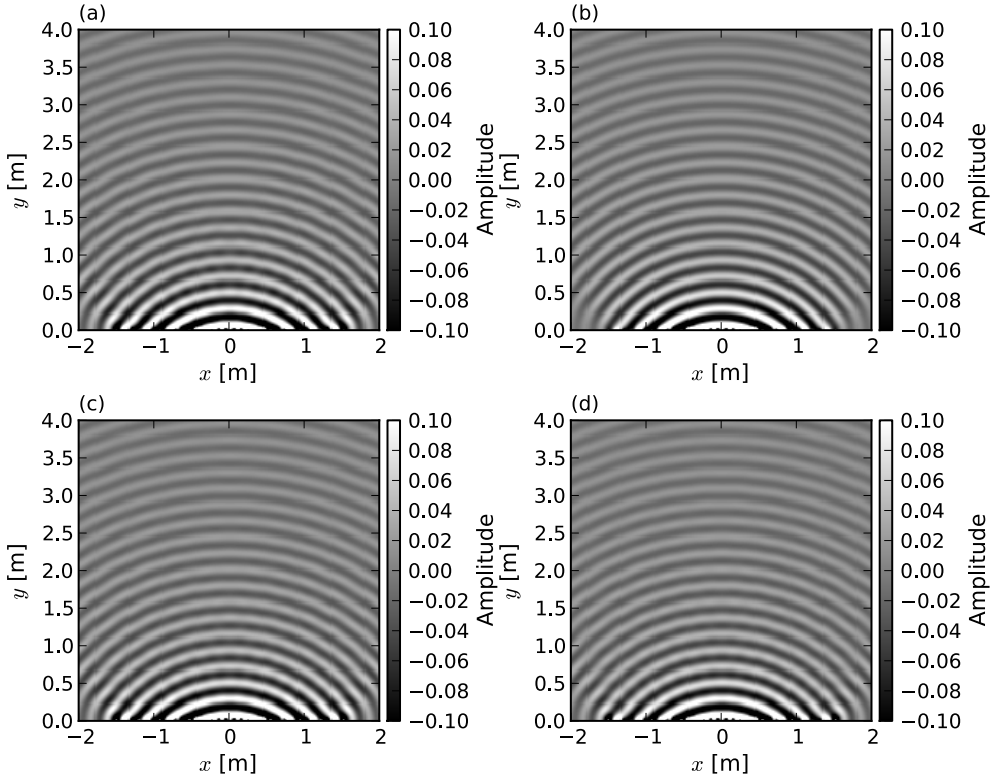


Figure 40. Wavefronts synthesized at 1600 Hz in free-field obtained by (a) HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

control points and monitoring sensors are respectively shown. The frequency of 1700 Hz is indicated in these figures, corresponding to the WFS spatial aliasing frequency. In Fig. 42(a), $E_{LS}(\omega)$ for HWFS at the control points is extremely small compared with that for the other methods throughout the range of frequencies; however, the reproduction accuracy at the control points for the other methods shown in Figs. 42(b)-42(d) maintains sufficient accuracy of $E_{LS}(\omega) \leq 10^{-2}$ when $\mu = 10^{-2}$ below 1700 Hz. In addition, in Figs. 42(b) to 42(d), with increasing regularization parameter μ, $E_{LS}(\omega)$ also increases and a difference can be observed between the regularization limitation for AWFS, MCWS-EVR and the proposed method. In MCWS-EVR, when $\mu \to \infty$, the synthesized wavefront disappears

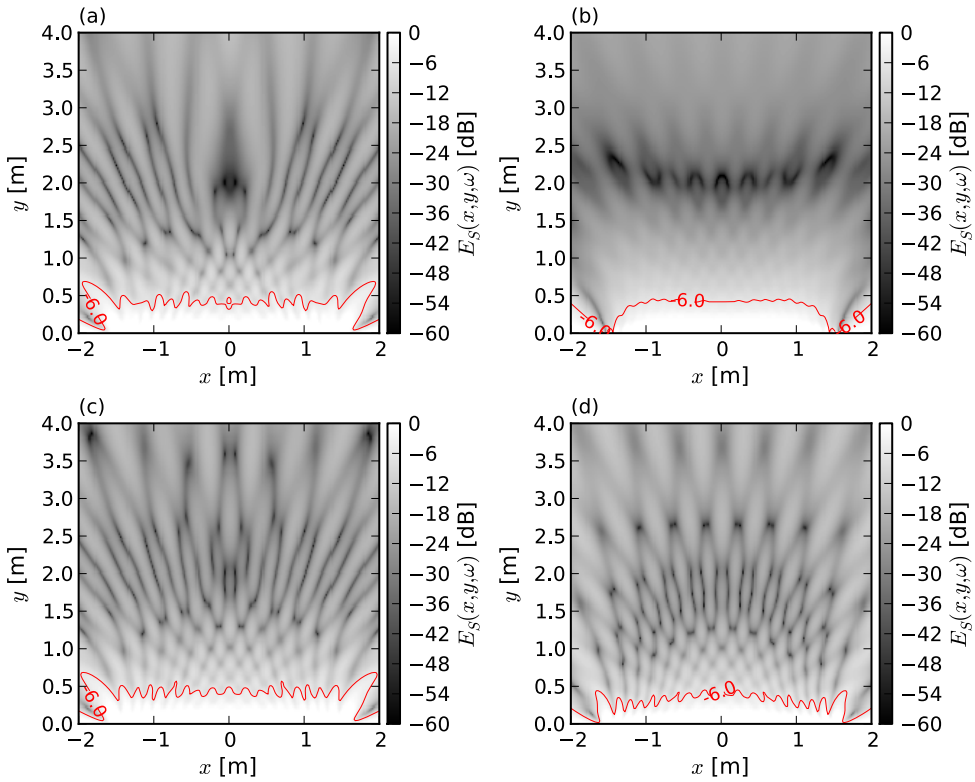


Figure 41. $E_S(x, y, \omega)$ values at 1600 Hz in free-field obtained by (a) HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

completely and $E_{LS}(\omega)$ becomes equal to 1 throughout the frequency range. In contrast, the limitation of $E_{LS}(\omega)$ for HWFS and the proposed method are limited to that for WFS. In Figs. 43(c) and 43(d), the $E_{LS}(\omega)$ values for AWFS and the proposed method at the monitoring sensors behave in a similar manner throughout the range of frequencies, and are almost equivalent to or smaller than that for WFS. Thus, it is possible to compensate for the reproduction accuracy outside the control points using the proposed method. However, similarly to the case of HWFS, in the proposed method, the coloration problem of WFS mentioned in the previous chapter will also occur outside the control points because the tendency for the error to increase in the proposed method is similar to that of WFS above

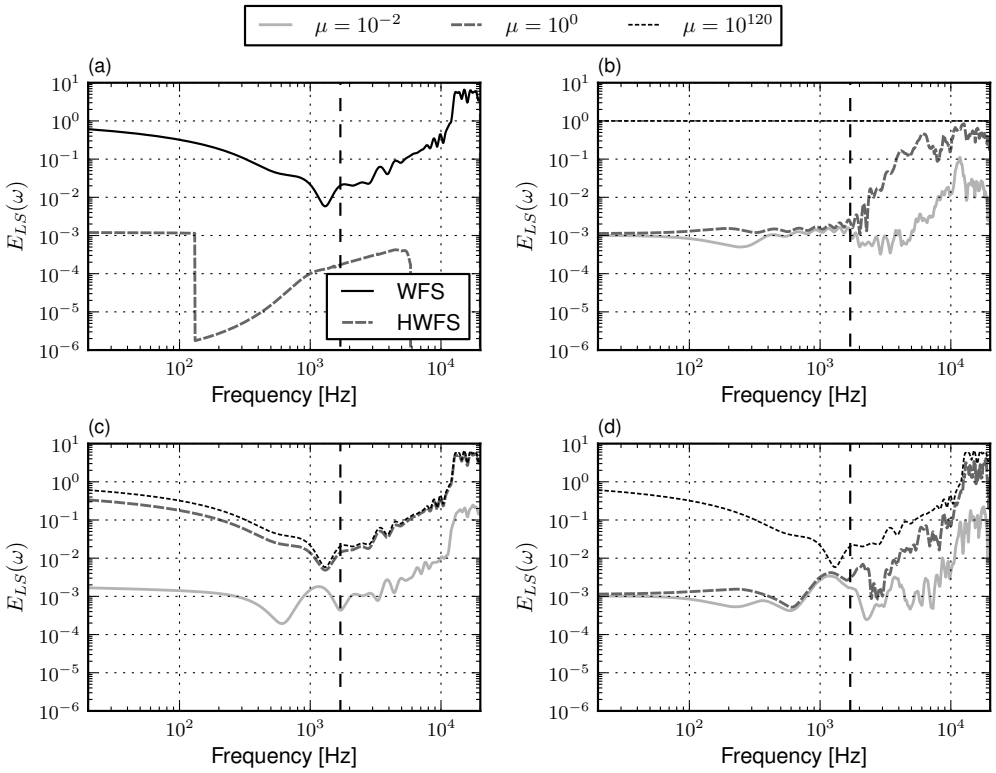


Figure 42. $E_{LS}(\omega)$ values in free-field at control-points for (a) WFS and HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

the spatial aliasing frequency. In addition, the discontinuity of $E_{LS}(\omega)$ due to the truncation of singular values in SVD is shown in the results of MCWS and the proposed method in Figs. 42(a) and 43(a).

From these findings, the proposed method is shown to be feasible.

6.9.2 Simulation results in reverberant room

Figure 44 shows the wavefronts synthesized at 1600 Hz under room-like conditions by (a) HWFS, (b) MCWS-EVR, (c) AWFS and (d) the proposed method for a monopole source at position $(x_P, y_P, z_P) = (0.0, -1.0, 1.3)$ [m]. The wavefront frequency of 1600 Hz used for evaluation is the upper limit of the major cues for sound source localization. Figures 45(a) to 45(d) show the values of $E_S(x, y, \omega)$

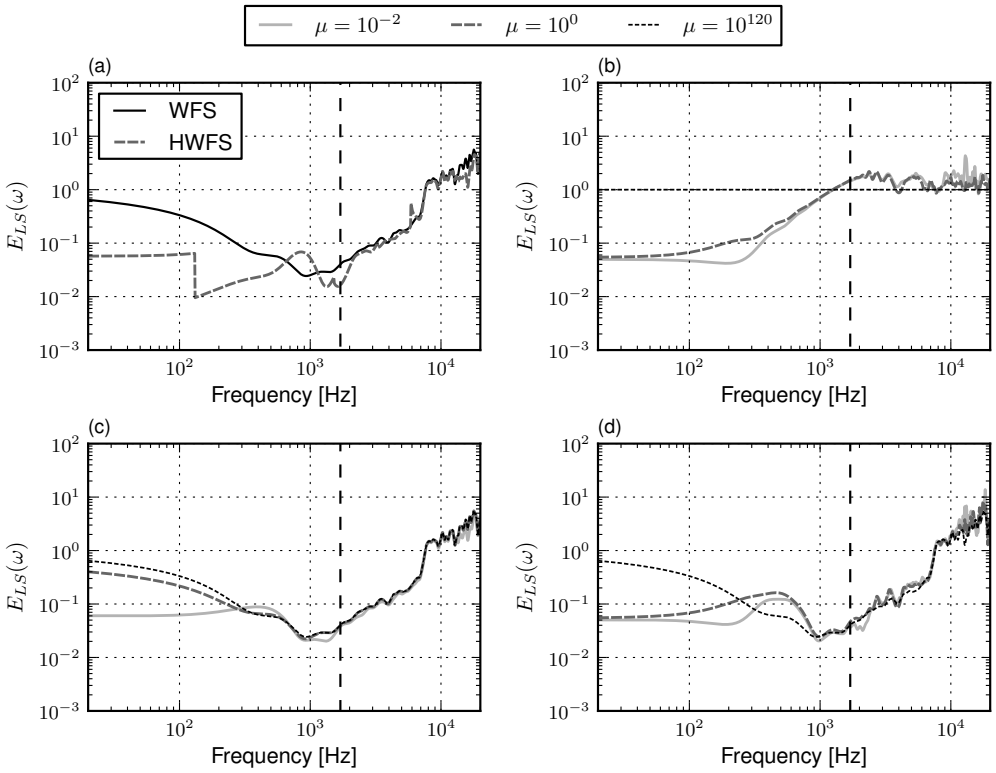


Figure 43. $E_{LS}(\omega)$ values in free field at monitoring sensors for (a) WFS and HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

corresponding to Figs. 44(a) to 44(d), respectively. In these calculation results, the regularization parameter μ and the differential order of $\boldsymbol{D}(\omega)$ are set equal to 10^{-2} and 2, respectively. In Figs. 44(a) to 44(d), it can seen that the reflected waves in the room disturb the synthesized wavefront. As shown in Figs. 45(a) to 45(d), since interference by the room-reflected waves occurs, the acceptable reproduction regions are narrower than those in the free-field simulation; however, AWFS, MCWS-EVR and the proposed method maintain relatively large regions compared with HWFS. The gap in the acceptable region for HWFS is caused by the discontinuity of the complex strengths of the neighbor secondary sources, whereas the gap is suppressed in the result for the proposed method shown in Fig. 45(d) even when the TSVD form given by Eq. (97) is used.

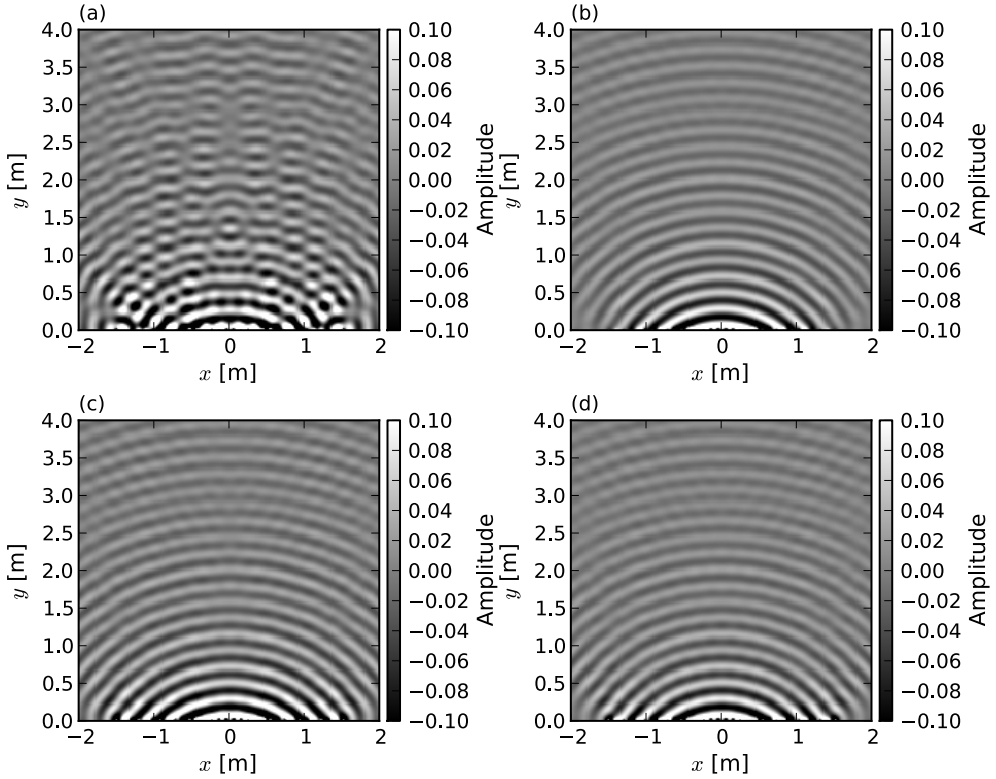


Figure 44. Wavefronts synthesized at 1600 Hz in room by (a) HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

Figures 46(a) to 46(d) and Figs. 47(a) to 47(d) show the $E_{LS}(\omega)$ values for (a) WFS and MCWS, (b) MCWS-EVR, (c) AWFS and (d) the proposed method at the control points and monitoring sensors. As shown in Fig. 46(d), the proposed method also maintains the reproduction accuracy at the control points. However, in Fig. 47(a), $E_{LS}(\omega)$ for HWFS increases outside the control points and fluctuates significantly between the results for MCWS and WFS. Thus, as the room reverberation increases, the reproduction errors outside the control points for the proposed method probably approach those for MCWS.

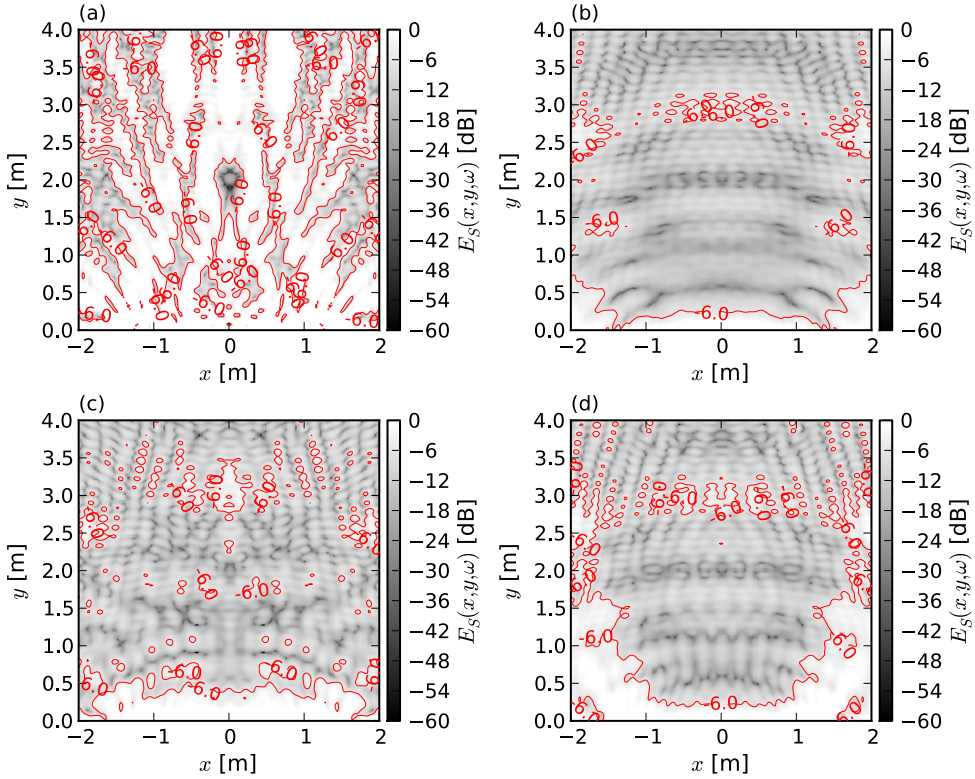


Figure 45. $E_S(x, y, \omega)$ values at 1600 Hz in free field obtained by (a) HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

6.10. Summary of this chapter

In this chapter, we first reviewed and discussed the related literature on sound field reproduction methods such as MCWS-ER, AWFS, MCWS-EVR and HWFS. Next, a new interpretation of the mathematical relationships and differences between MCWS-ER, AWFS, MCWS-EVR and HWFS was given by applying GSVD, and a new method was proposed for achieving the sound field metamorphosis between WFS and MCWS in AWFS and for improving the sweet-spot characteristics by anti-phase smoothing of the secondary source driving function using the primary wave field in MCWS-EVR within HWFS. The results of computer simulations revealed that the proposed method balances the above goals and has wide

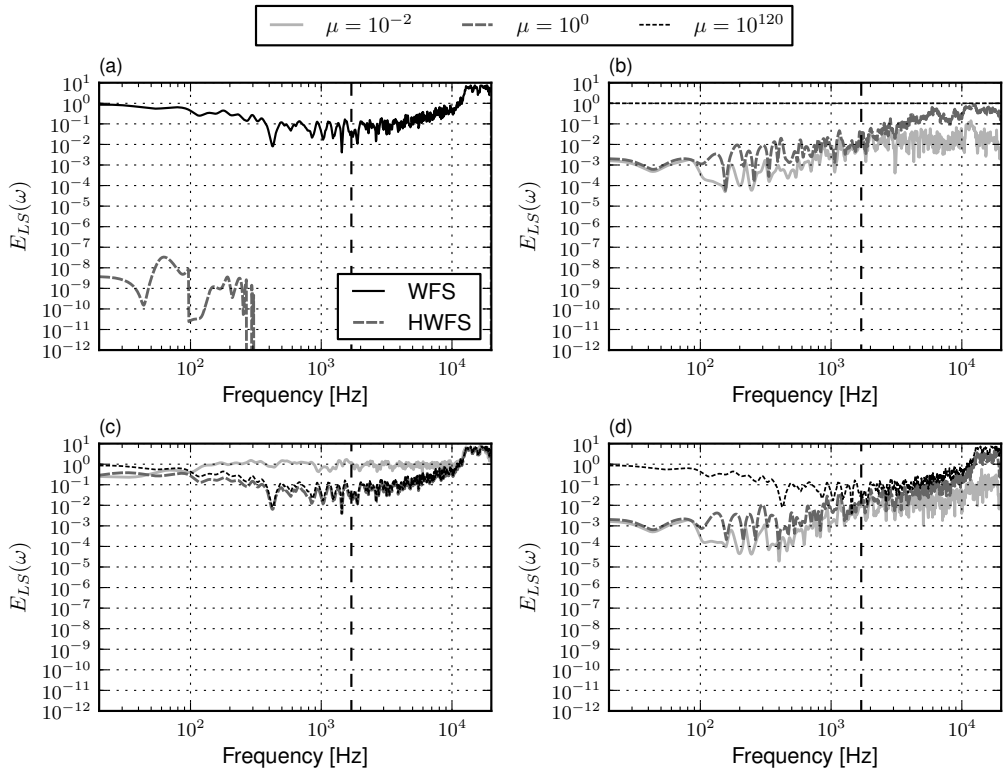


Figure 46. $E_{LS}(\omega)$ values in room at control-points for (a) WFS and HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

applicability in a spatial domain with high accuracy of reproduction in numerical calculations both under free-field conditions and in a model simulating a room with room reflection.

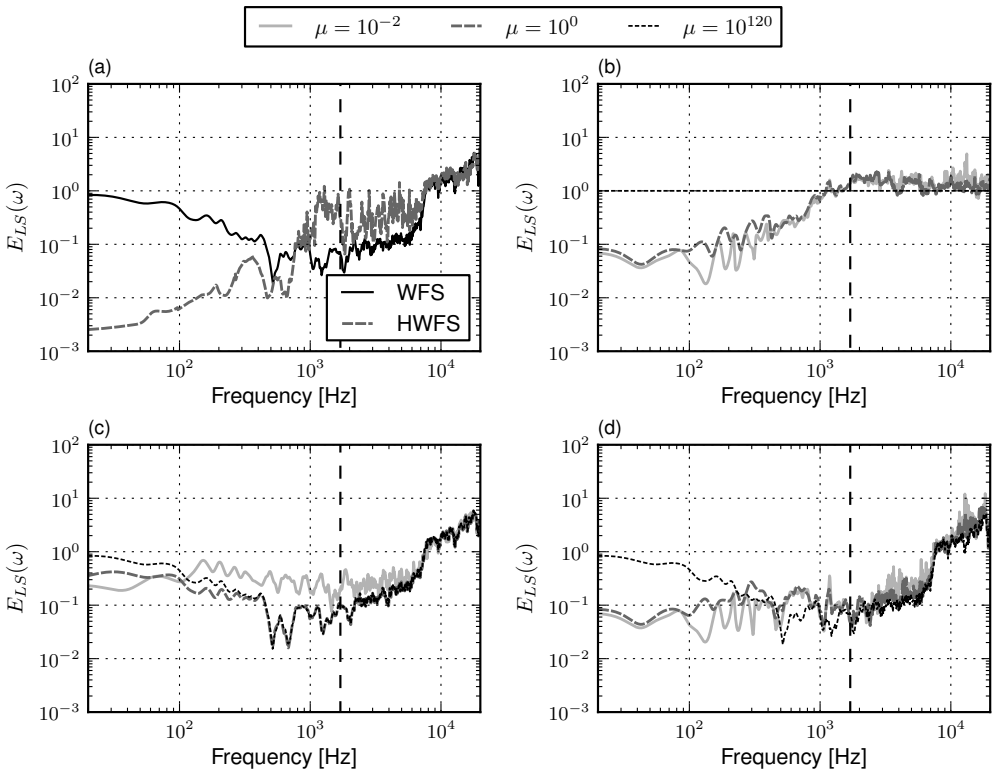


Figure 47. $E_{LS}(\omega)$ values in room at monitoring sensors for (a) WFS and HWFS, (b) MCWS-EVR, (c) AWFS and (d) proposed method.

CHAPTER 7

INTERACTIVE CONTROLLER FOR AUDIO OBJECT LOCALIZATION BASED ON SPATIAL REPRESENTATIVE VECTOR OPERATION

7.1. Introduction

In applying WFS to commercially available sound contents, the localization information of each sound source is required as a cue for the primary sources generated in WFS. This means that a method of analyzing and decomposing the sound sources is indispensable for the total system of multichannel sound reproduction. Therefore, in this Chapter, we propose a new audio coding framework enabling the localization operation of audio objects by the temporal quantization of spatial information and introduce an audio object controller developed for interactive sound field reproduction using audio object localization based on spatial representative vector operations.

7.2. System configuration

In this section, we give an overview of the proposed system. Figure 48 shows the configuration of the proposed system, which consists of three functions: a batch processing encoder, a real-time processing decoder and an interactive audio object localization controller.

First, the stereo mixed signal is inputted into the encoder, and the input signal is analyzed and decomposed to the audio objects contained within, and the encoder outputs are stored as a storage file by a dedicated format container file. The interactive controller reads the storage file and displays the relative positions of the audio objects. Using the touchscreen panel, the listener can control the relative positions via the graphical user interface (GUI). The real-time decoder

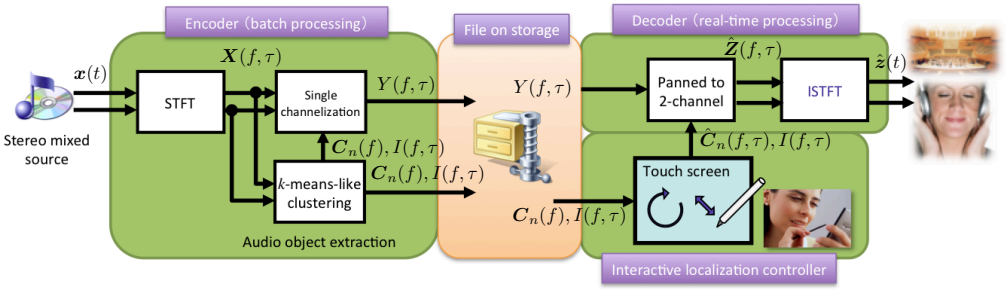


Figure 48. Configuration of proposed system.

interprets each listener operation and the operation is reflected immediately in the reproduced sound field. The mathematical principles of the proposed system are described in the following section.

7.3. Mathematical principles of proposed system

7.3.1 Encoder

We assume that the composition of musical instruments sound does not vary significantly in conventional audio signals. On the basis of this, the proposed method quantizes spatial information in the time-frequency domain. Hence, we store spatial localization information in a number of time-invariant basis vectors and extract audio objects efficiently by clustering the mixture of sounds using the basis vectors. Figure 49 shows the quantization error, which is the distance between the M-channel time-series complex vector $\boldsymbol{X}(f,\tau) = [X_1(f,\tau), \cdots, X_M(f,\tau)]^{\mathrm{T}}$, which consists of input signals, and the complex decoded signal vector $\boldsymbol{Z}(f,\tau) = [Z_1(f,\tau), \cdots, Z_M(f,\tau)]^{\mathrm{T}}$. Here we focus on the decoded signals that minimize the quantization error E, which signals are expressed in terms of the orthogonal projection of $\boldsymbol{X}(f,\tau)$ and supplementary information $\boldsymbol{H}(f,\tau)$, which includes quantized spatial information, as follows:

$$\boldsymbol{Z}(f,\tau) = \boldsymbol{H}^{\mathrm{H}}(f,\tau)\boldsymbol{X}(f,\tau)\frac{\boldsymbol{H}(f,\tau)}{\|\boldsymbol{H}(f,\tau)\|}, \tag{98}$$

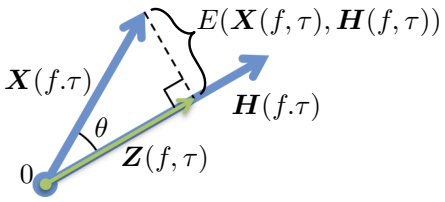


Figure 49. Relations between the input signal $\boldsymbol{X}$, the decoded signal $\boldsymbol{Z}$, the quantization spatial basis vector $\boldsymbol{H}$ and the quantization error E.

where superscript H denotes the complex conjugate transposition of a matrix, $\|\cdot\|$ denotes the Euclidean norm and $\boldsymbol{H}(f,\tau)$ is described as

$$\boldsymbol{H}(f,\tau) = \boldsymbol{C}_{I(f,\tau)}(f), \tag{99}$$

$$\boldsymbol{C}_n(f) = [C_{(n,1)}(f),\ldots,C_{(n,M)}(f)]^{\mathrm{T}} \quad (n=1,\ldots,N), \tag{100}$$

where $\boldsymbol{C}_n(f)$ is the nth centroid and the complex basis vector derived from the clustering described below, $I(f,\tau)$ denotes the nth index of the centroid that minimizes the quantization error between $\boldsymbol{X}(f,\tau)$ and $\boldsymbol{Z}(f,\tau)$ at every time-frequency grid in all channels, and N denotes the number of centroids. The basis vector used in the orthogonal projection is determined by the cosine-distance weighted k-means [38] to minimize the quantization error. We can obtain the resultant single-channel encoded signal $Y(f,\tau)$ using this basis vector.

To minimize the quantization error between input signals and decoded signals, we formulate the magnitude of the compression error. First, we calculate the nth basis vector $\boldsymbol{C}_n(f)$ and the decoded signal vector $\boldsymbol{Z}(f,\tau)$. The quantization error $E(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau))$ can be expressed in terms of the cosine-distance as

$$\begin{aligned} E(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau)) &= \|\boldsymbol{X}(f,\tau)\|\sin(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau)) \\ &= \|\boldsymbol{X}(f,\tau)\|\sqrt{1-\cos^2(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau))}, \end{aligned} \tag{101}$$

where $\cos(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau))$ is the cosine-distance between $\boldsymbol{X}(f,\tau)$ and $\boldsymbol{H}(f,\tau)$, as

$$\cos(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau)) = \frac{|\boldsymbol{X}^{\mathrm{H}}(f,\tau)\boldsymbol{H}(f,\tau)|}{\|\boldsymbol{X}(f,\tau)\|\|\boldsymbol{H}(f,\tau)\|}. \tag{102}$$

We optimize the basis vector $\boldsymbol{H}(f,\tau)$ so that the total error of the signals, $E_{\text{total},n}(f)$, is minimized. The total error is given by

$$E_{\text{total},n}(f) = \sum_{\tau\in\Theta_n}\Big(E(\boldsymbol{X}(f,\tau),\boldsymbol{H}(f,\tau))\Big)^2, \tag{103}$$

where Θ_n is the nth class of the cluster. Optimization of the basis vector $\boldsymbol{H}(f,\tau)$ minimizing the error is equivalent to the k-means clustering problem for the cosine-distance as follows.

[STEP 1] Prototype centroids are generated as

$$\boldsymbol{C}_n^{[k]}(f) = \left[C_{(n,1)}^{[k]}(f),\ldots,C_{(n,M)}^{[k]}(f)\right]^{\mathrm{T}} \quad (n=1,\ldots,N), \tag{104}$$

where k is the number of iterations used to update the centroid.

[STEP 2] Each input signal $\boldsymbol{X}(f,\tau)$ is assigned to the nth class Θ_n based on the error between the input signal $\boldsymbol{X}(f,\tau)$ and the centroid vector $\boldsymbol{C}_n(f)$ as follows:

$$I^{[k]}(f,\tau) = \underset{n}{\operatorname{argmin}}\, E(\boldsymbol{X}(f,\tau),\boldsymbol{C}_n^{[k]}(f))^2, \tag{105}$$

$$\Theta_n = \{\tau : I^{[k]}(f,\tau) = n\}, \tag{106}$$

where $\operatorname{argmin}_n \cdot$ denotes the minimization function, $\{\tau\}$ denotes the class that corresponds to a set of τ, and $I^{[k]}(f,\tau)$ is the index function of the kth iteration. The class to which $\boldsymbol{X}(f,\tau)$ belongs is determined by $I^{[k]}(f,\tau)$.

[STEP 3] The optimal basis vector is extracted to minimize the error, as

$$\begin{aligned}
\boldsymbol{C}_n^{[k+1]}(f) &= \underset{\boldsymbol{C}_n^{[k]}(f)}{\operatorname{argmin}}\, E_{\text{total},n}(f) \\
&= \underset{\boldsymbol{C}_n^{[k]}(f)}{\operatorname{argmin}} \sum_{\tau\in\Theta_n}\Big(E(\boldsymbol{X}(f,\tau),\boldsymbol{C}_n^{[k]}(f))\Big)^2 \\
&= \underset{\boldsymbol{C}_n^{[k]}(f)}{\operatorname{argmin}} \sum_{\tau\in\Theta_n}\|\boldsymbol{X}(f,\tau)\|^2\Big(1-\cos^2(\boldsymbol{X}(f,\tau),\boldsymbol{C}_n^{[k]}(f))\Big) \\
&= \underset{\boldsymbol{C}_n^{[k]}(f)}{\operatorname{argmin}} \sum_{\tau\in\Theta_n}\|\boldsymbol{X}(f,\tau)\|^2\left(1-\frac{|\boldsymbol{X}^{\mathrm{H}}(f,\tau)\boldsymbol{C}_n^{[k]}(f)|^2}{\|\boldsymbol{X}(f,\tau)\|^2\|\boldsymbol{C}_n^{[k]}(f)\|^2}\right) \\
&= \underset{\boldsymbol{C}_n^{[k]}(f)}{\operatorname{argmin}} \sum_{\tau\in\Theta_n}-|\boldsymbol{X}^{\mathrm{H}}(f,\tau)\boldsymbol{C}_n^{[k]}(f)|^2 \\
&= \underset{\boldsymbol{C}_n^{[k]}(f)}{\operatorname{argmax}}\boldsymbol{C}_n^{[k]^{\mathrm{H}}}(f)\Big(\sum_{\tau\in\Theta_n}\boldsymbol{X}(f,\tau)\boldsymbol{X}^{\mathrm{H}}(f,\tau)\Big)\boldsymbol{C}_n^{[k]}(f).
\end{aligned} \tag{107}$$

Owing to the constraint $\|\boldsymbol{C}_n(f)\| = 1$, the maximization problem on the right-hand side of (107) is equivalent to finding the maximum eigenvalue of $\sum_{\tau\in\Theta_n} \boldsymbol{X}(f,\tau)\boldsymbol{X}^{\mathrm{H}}(f,\tau)$. Therefore, the basis vector $\boldsymbol{H}_i(f)$ is derived as the maximum eigenvector of $\sum_{t\in\Theta_n} \boldsymbol{X}(f,\tau)\boldsymbol{X}^{\mathrm{H}}(f,\tau)$.

[STEP 4] If the centroid vector does not change from that obtained by the previous iteration, the optimal vectors are determined to be the basis vector $\boldsymbol{C}_n(f)$. If the centroid vector changes, the algorithm returns to **[STEP 2]** with $k = k + 1$.

From (98), (100) and the constraint $\|\boldsymbol{C}_n(f)\| = 1$, the single-channel encoded signal $Y(f,\tau)$ is obtained as follows:

$$\begin{aligned}\boldsymbol{Z}(f,\tau) =& \boldsymbol{H}^{\mathrm{H}}(f,\tau)\boldsymbol{X}(f,\tau)\frac{\boldsymbol{H}(f,\tau)}{\|\boldsymbol{H}(f,\tau)\|}\\ =& \left\{\boldsymbol{C}^{\mathrm{H}}_{I(f,\tau)}(f)\boldsymbol{X}(f,\tau)\right\}\boldsymbol{C}_{I(f,\tau)}(f)\\ =& Y(f,\tau)\boldsymbol{C}_{I(f,\tau)}(f), \qquad (108)\\ Y(f,\tau) =& \boldsymbol{C}^{\mathrm{H}}_{I(f,\tau)}(f)\boldsymbol{X}(f,\tau). \qquad (109)\end{aligned}$$

7.3.2 Decoder

In the decoding process, multichannel decoded signals $\boldsymbol{Z}(f,\tau)$ are produced by the single-channel compressed signal $Y(f,\tau)$ and the basis vector $\boldsymbol{C}_{I(f,\tau)}(f)$ as follows:

$$\boldsymbol{Z}(f,\tau) = Y(f,\tau)\boldsymbol{C}_{I(f,\tau)}(f), \qquad (110)$$

and the nth clustered audio object signal is described as follows:

$$\boldsymbol{Z}_n(f,\tau) = \begin{cases} Y(f,\tau)\boldsymbol{C}_{I(f,\tau)}(f) & (I(f,\tau) = n),\\ 0 & (otherwise).\end{cases}$$

In (110), the quantization vector $\boldsymbol{C}_{I(f,\tau)}(f)$ represents the clustered audio object and its angle expresses the spatial image of the audio object. Therefore, we refer to the quantization vector as the *spatial representative vector* (SRV) hereafter. Finally, the decoder converts $\boldsymbol{Z}(f,\tau)$ into the time-domain expression $z(t)$ by an inverse short-time Fourier transformation (ISTFT) and outputs it to a pair of headphones or earphones. The time-domain signal $z(t)$ is described as

$$z(t) = \mathrm{ISTFT}(\boldsymbol{Z}(f,\tau)). \qquad (111)$$

7.3.3 Localization operation of audio objects

In this section, we describe signal processing for the localization operation of audio objects by the rotation and expansion/contraction of the SRV. The SRV is operated in the following manner:

$$\hat{\boldsymbol{C}}_n(f,\tau) = G_n(f,\tau)\boldsymbol{R}_n(f,\tau)\boldsymbol{C}_n(f), \tag{112}$$

where $\hat{\boldsymbol{C}}_n(f,\tau)$ is the operated SRV, and $G_n(f,\tau)$ and $\boldsymbol{R}_n(f,\tau)$ are filters that control the perceptual distance (gain) and the azimuth of the perceptual sound image of the nth audio object, respectively. $R_n(f,\tau)$ is obtained as the following rotation matrix:

$$\boldsymbol{R}_n(f,\tau) = \begin{bmatrix} \cos(\vartheta_n(f,\tau)) & -\sin(\vartheta_n(f,\tau)) \\ \sin(\vartheta_n(f,\tau)) & \cos(\vartheta_n(f,\tau)) \end{bmatrix}, \tag{113}$$

where $\vartheta_n(f,\tau)$ denotes the angle of rotation for the nth SRV. By filtering the SRV with $G_n(f,\tau)$ and $\boldsymbol{R}_n(f,\tau)$, we can control the localization of the corresponding audio object of interest on the same horizontal plane as that of the ears of the listener. The operated signal is obtained by restoring the single-channel signal using the SRV as

$$\hat{\boldsymbol{Z}}(f,\tau) = Y(f,\tau)\hat{\boldsymbol{C}}_{I(f,\tau)}(f,\tau). \tag{114}$$

7.4. Evaluation of audio object operations

7.4.1 Experimental conditions

In this section, we assess the effectiveness of localization and the sound quality of an audio object after performing individual operations to verify the operation of the interactive controller via a subjective evaluation. In this experiment, we use three audio signals recorded by a professional musician and mixed using a real mixing console. The front direction corresponds to 0°, and the clockwise direction is denoted by a positive angle. The audio objects that are mixed are located at {-45°, 0°, 45°} in the real space. The reference signal in the subjective test is the unprocessed original signal. The number of SRVs is three, which is the same as

the number of audio objects and the SRVs are set to three directions {-45°, 0°, 45° }, which are the same as the directions of the audio objects. The signal length is 5 s, fast Fourier transform (FFT) points is 8192, the number of shift points is 4096, the sampling rate is 44.1 kHz, the number of quantization bits is 16 bits and the maximum number of updates of the centroid is 50. The values of $G_n(\tau)$ are set to {0.1, 1, 2, 5}. The values of $\vartheta_n(\tau)$ are set to {0, 22.5, 45}. The test subjects are five adult males with normal audibility. We use a pair-comparison method with the reconstructed signal operated by (112) to evaluate the effectiveness of audio object localization. The test subjects evaluate how the perceptual gain of the audio object and the localization of each reconstructed signal changed by selecting one of the three phrases {became large, did not change, became small} and {rotated left, did not change, rotated right}, respectively. We use the mean opinion score [61] (MOS) method to evaluate the sound quality using the following five grades: {5: excellent, 4: good, 3: fair, 2: poor, 1: very poor}.

7.4.2 Results of subjective evaluation

Figures 50 and 51 show the correct answer rate and the perceived sound quality of the audio objects after localization control by the proposed method. First, as the gain parameter $G_n(\tau)$ increases, the perceptual audio gain changes. In addition, it is possible to adjust the gain of individual audio objects corresponding to each SRV while maintaining sufficient sound quality except for the case of $G_n(\tau) = 5$.

Next, as the rotation parameter $\vartheta_n(\tau)$ changes, the perceptual azimuth of the audio objects changes. In addition, the rotation operation maintains higher sound quality than the gain operation. The relations among the gains of the audio objects are not severely changed by the rotation operation, therefore the rotation operation can suppress the occurrence of spectrum distortion due to this operation by perceptual masking effects. However, the gain operation destroys these relations; thus, it appears that the gain operation has a harmful effect on the sound quality.

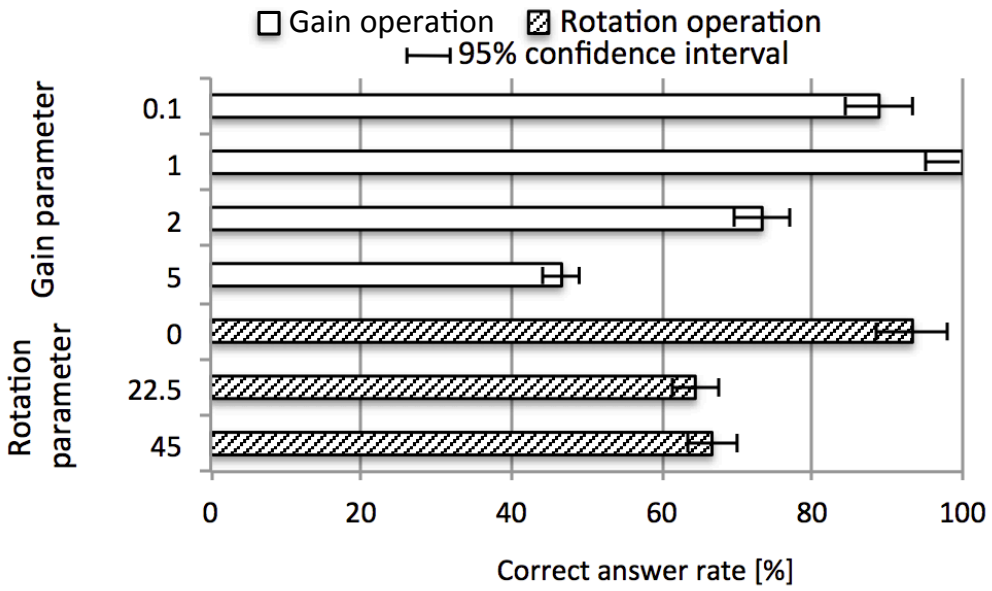


Figure 50. Subjective evaluation results for operability of gain and rotation operations.

7.5. Development of interactive controller for audio object localization

In this section, we introduce the proposed interactive controller for audio object localization using an SRV operation. Figure 52 shows an overview of the controller. This controller is equipped with a capacitive touchscreen panel, and the listener can intuitively operate every audio object displayed on the touchscreen panel with a touch pen.

Figure 53 shows the geometry of two audio objects and their manipulative variables on the touchscreen panel. There is a subject icon at the center of the touchpanel screen; the front direction of the subject corresponds to 0°, and the clockwise direction is denoted by a positive angle. The default position of a displayed audio object in polar coordinates is described as follows:

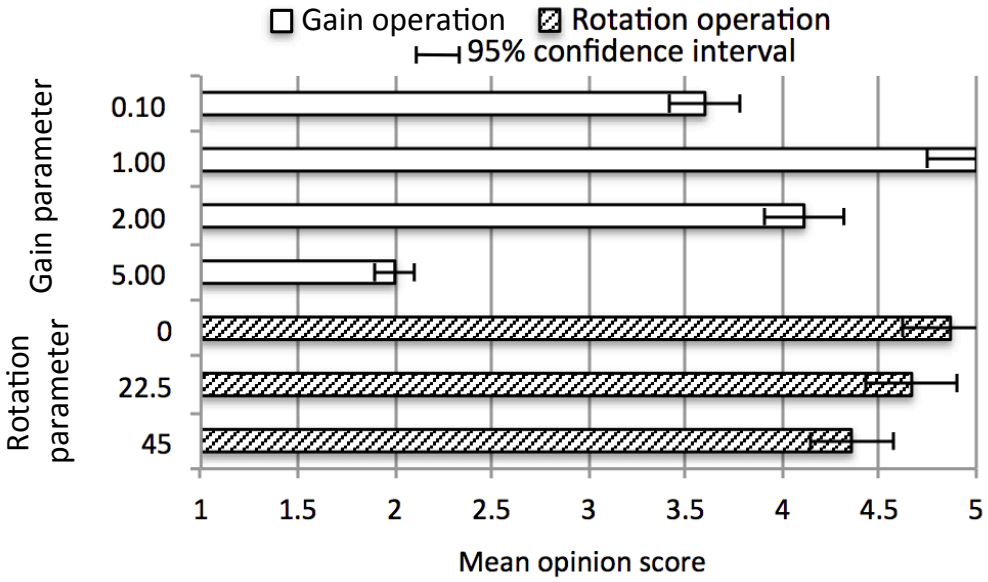


Figure 51. Subjective evaluation results for sound quality of gain and rotation operations.

$$r_{n_0} = 1, \tag{115}$$

$$\varphi_{n_0} = -2\left\{\left\langle \arg(\boldsymbol{C}_{\mathrm{amp},n}(f))\right\rangle_f - \frac{\pi}{4}\right\}, \tag{116}$$

where $\langle\cdot\rangle_f$ denotes the function that calculates the average degree over frequency f, and $\boldsymbol{C}_{\mathrm{amp},n}(f)$ and the $\arg(\cdot)$ operator are described as

$$\boldsymbol{C}_{\mathrm{amp},n}(f) = [|C_{(n,1)}(f)|, \ldots, |C_{(n,M)}(f)|]^{\mathrm{T}} \quad (n = 1, \ldots, N), \tag{117}$$

$$\arg(\boldsymbol{C}_{\mathrm{amp},n}(f)) = \arctan\left(\frac{C_{(n,L)}(f)}{C_{(n,R)}(f)}\right). \tag{118}$$

In (116), $\pi/4$ is subtracted from the argument of the channel power ratio of the audio object $\boldsymbol{C}_{\mathrm{amp},n}(f)$ so that $\varphi_{n_0} = 0$ when the channel power ratio is 1, that is, $\langle \arg(\boldsymbol{C}_{\mathrm{amp},n}(f))\rangle_f = \pi/4$. The result is then doubled to display the audio object on the touchscreen panel as an instrument icon. This transformation provides display positions consistent with the audio objects. From (112) and Fig. 53, the

Figure 52. Prototype of real-time interactive controller.

relationship between the manipulative variables on the display and the operation on the audio objects is described as follows:

$$G_n(\tau) = \frac{1}{\{r_{n_0} + r_n(\tau)\}^2} = \frac{1}{\{1 + r_n(\tau)\}^2}, \tag{119}$$

$$\boldsymbol{R}_n(\tau) = \begin{bmatrix} \cos(\frac{\vartheta_n(\tau)}{2}) & -\sin(\frac{\vartheta_n(\tau)}{2}) \\ \sin(\frac{\vartheta_n(\tau)}{2}) & \cos(\frac{\vartheta_n(\tau)}{2}) \end{bmatrix}. \tag{120}$$

The total time required before getting the operation results is 40 to 50 ms, including the delays of block buffer processing. This total time is equivalent to the shift length of STFT (the length being 2048 samples in this implementation), the audio-event handling and the antialiasing digital filtering of the DA converter in the 48 kHz sampling frequency.

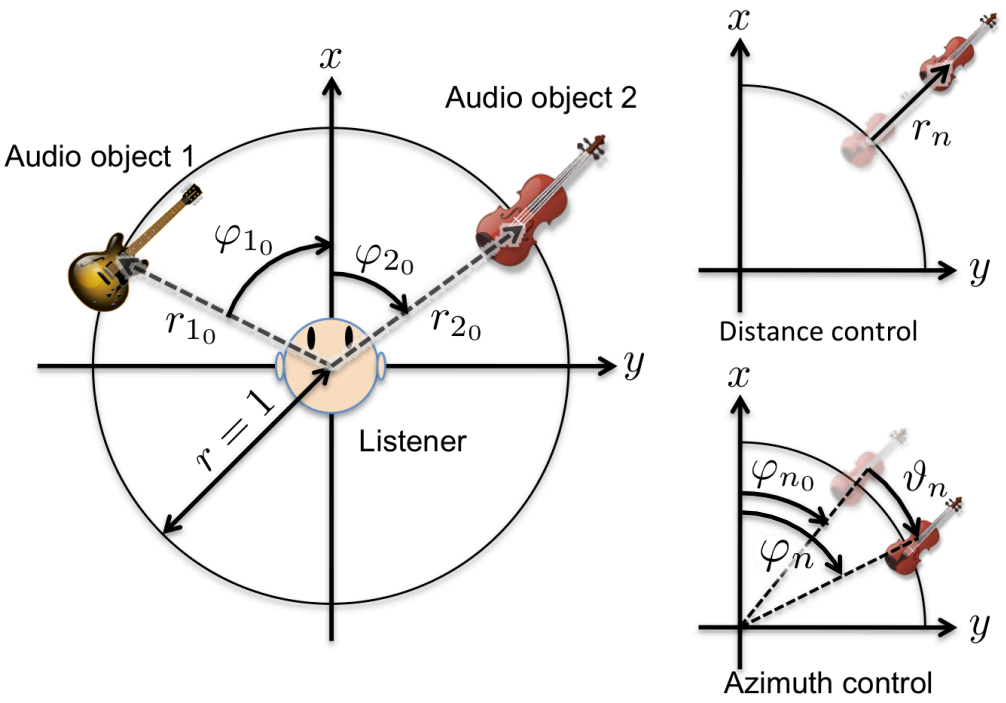


Figure 53. Geometry of audio objects and their manipulative variables on the touchscreen panel.

7.6. Evaluation of interactive controller

7.6.1 Experimental conditions

In this section, we assess the sound quality, the operability and the GUI usability of localization of an audio object after performing a single operation of the interactive controller via a subjective evaluation. In this experiment, we use three audio signals recorded by a professional musician and mixed using a real mixing console, locating at {-45°, 0°, 45°} in the real space. The number of SRVs is three, which is the same as the number of audio objects and the SRVs are initialized to three directions {-45°, 0°, 45°}. There are 8192 FFT points, the number of shift points is 4096, the sampling rate is 48 kHz, there are 16 quantization bits, and the maximum number of updates of the centroid is 50. The user-operable range of

$G_n(\tau)$ is set to {1/1.5–1.5 (±1.76 dB), 1/2–2.0 (±3.01 dB), 1/3–3.0 (±4.77 dB)} and the range of $\varphi_n(\tau)$ is set to $\{-\pi/2–\pi/2\}$. The test subjects are 18 adult males and females with normal audibility. We use the MOS to evaluate the sound quality, the operability and the GUI usability of localization by the interactive controller; the following five levels are requested: {5: excellent, 4: good, 3: fair, 2: poor, 1: very poor}.

7.6.2 Results of subjective evaluation

Figure 54 shows the subjective evaluation results for sound quality, the operability and the GUI usability of localization of the audio objects after performing signal operations by the interactive controller. In Fig. 54 (a), as the range of gain parameters increases, the perceptual audio quality decreases, whereas the rotation operation maintains the higher sound quality. Figures 54 (b) and 54 (c) show that the qualities on the sound localization and the GUI qualify almost as good and satisfactory to the users because these scores are around 4 regardless of the type of operations. These results clarify that this interactive controller enables the listener to change the gain and the localization of audio objects without sound degradation, apart from an excessive gain operation. With the development of this system, various operations based on this interactive controller and decoder enable a seamless control of audio objects. The usability of the developed system is demonstrated at the following URL:

`http://spalab.naist.jp/aocdemo.html`.

7.7. Summary of this Chapter

In this chapter, a new interactive controller for audio object localization based on spatially representative vector operations was proposed. First, I introduced the configuration of the proposed system. Second, I described the mathematical principles of our system in detail, proving that the spatially representative vectors

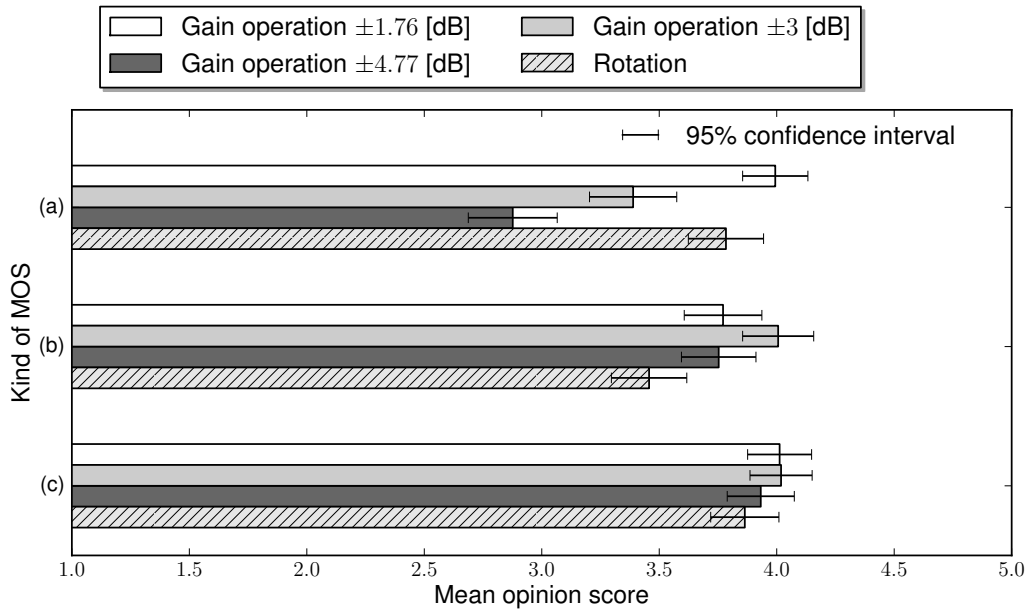


Figure 54. Subjective evaluation results using MOS to verify the gain and rotation operations of the interactive controller; (a) results of the sound quality after each operation; (b) results of the operability of each operation; (c) results of the GUI usability.

are closely related to the centroids of cosine-distance weighted *k*-means, and I described the audio object localization control method using the derived vectors.

Next, I assessed the effectiveness of localization and the sound quality of an audio object after performing single operations to verify the operation of the interactive controller via a subjective evaluation. The results of the experiments showed that the interactive controller enables the listener to change the localization of audio objects without sound degradation if the gain operation is not extreme.

Chapter 8

Epilogue

8.1. Dissertation summary

In this dissertation, we extended the current multichannel sound field coding and multichannel sound field reproduction frameworks with more flexibility of user experiments.

In Chapter 2, WFS, MCWS and DMCWS were described theoretically as a basic frameworks of the sound field reproduction.

Next, the optimal control-point geometry and the behavior of the synthesized secondary wavefront in DMCWS were investigated, and the effectiveness of DMCWS through the comparison with WFS was evaluated in Chapter 3.

Since the wavefront has the frequency characteristics of actual audio applications, I derived the spatial spectrum characteristics from the impulse responses at each observation point and measure the wavefront of DMCWS using the constructed wavefront measurement system in an actual environment in Chapter 4. From these evaluations and measurement results , it has been shown that the accuracy of the synthesized secondary wavefront is related to the control-point coordinates and the wavefront measurement results of DMCWS in an actual environment. Numerical wavefront calculations clarified that the optimum y-coordinate of the directly aligned control point is y_C = 0.6–0.7 m. Also numerically obtained WFS and DMCWS wavefronts compared using this range of y_C that DMCWS has a larger listening area with fewer amplitude and phase errors than WFS, whereas they have a similar attenuation error.

Our wavefront measurement system and our algorithm using impulse responses measured in an acoustically isolated room clarified that the measured wavefront is affected by the mural-reflected wave and the frequency-amplitude characteristics of the secondary source loudspeaker. In addition, DMCWS can be performed above the spatial aliasing frequency of WFS according to the results of numerical calculations and measurement. From these findings, we can conclude the advan-

tageousness of DMCWS compared with WFS.

In Chapter 5, to mitigate the trade-off problem between the accuracy and reproducible region caused by the practical location of the control-points of the reproduced sound field in both of conventional methods, we proposed a new method of balancing the listening area and reproduction accuracy with absolute accuracy using an inverse filter of the room acoustics: the null space of the generalized inverse matrix given by a compensation filter of the wave field outside the control points. The efficacy of the proposed method was ascertained by objective evaluation through computer-based simulations and wavefront-measurement. The results of computer simulations revealed that the proposed method balances the above goals and has wide applicability in a spatial domain with high accuracy of reproduction both under free-field conditions and in a simulation model with room reflection. Next, we measured the wavefront generated by our proposed method to evaluate the efficiencies. From the measurement results showed that our proposed method can synthesize the primary wavefront with the accuracy of the reproduction at the control-points. Finally, these results revealed that the proposed method balances the competing goals and has wide applicability in a spatial domain with high accuracy of reproduction under the actual environment then numerical calculations.

In Chapter 6, I first reviewed and discussed of the related literature of active sound field reproduction methods such as MCWS-ER, AWFS, MCWS-EVR and HWFS. Next, a new interpretation of the mathematical relationship and differences between MCWS-ER, AWFS, MCWS-EVR and HWFS was given by applying GSVD, and a new method was proposed for archiving the sound field metamorphosis between WFS and MCWS in AWFS for improving the sweet-spot characteristics by the anti-phase smoothing of the secondary source driving function using the primary wave field in MCWS-EVR within the HWFS. The results of computer simulations revealed that the proposed method balances the above goals and has wide applicability in a spatial domain with high accuracy of reproduction in numerical calculations both under free-field conditions and in a model simulating a room with room reflection.

In Chapter 7, a new audio coding framework enabling the localization opera-

tion of audio objects (e.g., vocal, guitar, drums) by the temporal quantization of spatial information was proposed, and the audio object controller developed for interactive sound field reproduction using audio object localization based on spatial representative vector operations was introduced. First, I described the mathematical principles of our system in detail, proving that the spatially representative vectors are closely related to the centroids of cosine-distance weighted k-means, and we described the audio object localization control method using the derived vectors. In addition, the proposed audio object controller allows a listener to operate an initial sound field and build a new virtual sound field by performing sound field operations on a stereo mixed source. Next, I assessed the effectiveness of localization and the sound quality of an audio object after performing single operations to verify the operation of the interactive controller via a subjective evaluation. The results of the experiments clarified that the interactive controller enables the listener to change the localization of audio objects without sound degradation if the gain operation is not extreme.

8.2. Future work

In the dissertation, I have improved the reproducible area and the reproduced accuracy for multichannel sound field reproduction systems, and the proposed algorithm has archived enough performance for those systems. However, the following problems are still opened.

First, in the proposed method, the fluctuation of room transfer functions degrades reproduced sound in which the coefficients of the inverse filter are fixed. In order to resolve this problem, an iterative adaptive algorithm for an inverse filter performed in SVD framework is needed [62, 63, 64, 65].

Next, for evaluating the reproduced sound field in actual environment, the measurement system must deal comprehensively on principal factors of the reality of the primary sound field not only direct wavefront but also reflecting wavefront, reverberation and the auditory perception of sound images which arise in the actual reproduced sound field. In this dissertation, I proposed the wavefront

measurement system and the measurement theory. However, the measurement theory dealt with only a factor of direct wavefront. Therefore, more sophisticated approaches are needed to evaluate the synthesized secondary wave field.

Finally, in applying WFS to commercially available sound contents, localization information of each sound source is required as a cue of the primary sources generated in WFS. This means that a method for analyzing and decomposing the sound sources is indispensable in the total system of multichannel sound reproduction. In this dissertation, I proposed the audio object coding method which provides the direction of the localization of the audio object. However, the required sound localization information in the phase of sound field reproduction not only the direction of the audio object but also distance of the audio object. Therefore, the analyzing method to estimate the localization distance of the audio object which is included in the commercially available sound contents are required.

In the future, it is needed that the analysis in terms of another multichannel signal processing and its integration methods for our proposed method.

References

[1] A. Blumlein, "Improvements in and relating to sound transmission, sound-recording and sound reproduction systems," Patent 394 325, 1931.

[2] M. A. H. Clark, G. F. Dutton, P.B. Vanderlyn, "The 'stereosonic' recording and reproducing system: A two-channel systems for domestic tape records," *J. Audio Eng. Soc*, vol. 6, no. 2, pp. 102–117, 1958.

[3] P. S. Gaskel, P. A. Ratliff, "Quadraphony: developments in matrix h decoding," in *Tech. Rep.* BBC Research Department, 1977.

[4] Rec. ITU-R BS.775-1, *Multichannel Stereophonic Sound System with and without Accompanying Pictures*.

[5] Rec. ITU-R BS.1116-1, *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*.

[6] W. E. Garity, J. N. A. Hawkins, "Fantasound," *Journal of the Society of Motion Picture Engineers*, vol. 37, no. 2, pp. 127–146, 1941.

[7] W. B. Snow, *Auditory Perspective*. Bell Laboratories Record, 1934, vol. 12.

[8] ——, "Basic principle of stereophonic sound," *Journal of SMPTE*, vol. 61, pp. 567–589, 1953.

[9] A. S. Harma, "Coding principles for virtual acoustic openings," in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, 2002.

[10] S. T. Guijarro, J. A. Beracoechea, F. J. C. Quiros, I. P. Garcia, "Coding strategies and quality measure for multichannel audio," in *Audio Engineering Society Convention 116*, 2004.

[11] S. Miyabe, K. Masatoki, H. Saruwatari, K. Shikano, and T. Nomura, "Temporal quantization of spatial information using directional clustering for multichannel audio coding," in *Proc. of WASPAA 09*, 2009, p. 261??264.

[12] Christiaan Huygens, *Traité de la Lumière*, 1678.

[13] T. C. E. Corteel, "Sound scene creation and manipulation using wave field synthesis," 2004.

[14] A. J. Berkhout, "A holographic approach to acoustic control," *J. Audio Eng. Soc.*, vol. 36, no. 12, pp. 977–995, 1988.

[15] A. J. Berkhout, D. de Vries, P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, 1993.

[16] E. N. G. Verheijen, "Sound reproduction by wave field synthesis," Ph.D. dissertation, Deflt University of Technology, 1997.

[17] W. de Bruijn, "Application of wave field synthesis in videoconferencing," Ph.D. dissertation, Delft University of Technology, 2004.

[18] M. R. Schroeder, B. S. Atal, "Computer simulation of sound transmission in rooms," *Proceedings of the IEEE*, vol. 51, pp. 536–537, 1963.

[19] M. R. Schroeder, "Digital simulation of sound transmission in reverberant spaces," *J. Acoust. Soc. Am.*, vol. 47, no. 2, pp. 424–431, 1970.

[20] M. Miyoshi, Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. on ASSP*, vol. 36, no. 2, pp. 683–705, 1988.

[21] N. Takashima, T. Sugano, S. Uto, H. Harada, "The basic study of the sound field control using multi-channel adaptive digital filters," *IEICE Technical Report*, vol. EA94, no. 38, pp. 1–8, 1994.

[22] T. Ohno, H. Hokari, S. Shimada, I. Oohira, "A study on cylindrical wave field synthesis with ultrasmall loudspeaker array," *IEICE Technical Report*, vol. EA2006, no. 56, pp. 7–12, 2006.

[23] M. Kolundzija, C. Faller, M. Vetterli, "Sound field reconstruction: An improved approach for wave field synthesis," in *Audio Engineering Society Convention 126*, 2009.

[24] N. Kamado, H. Hokari, S. Shoji, H. Saruwatari, K. Shikano, "Sound field reproduction by wavefront synthesis using directly aligned multi point control," in *Proc. 40th AES Conference*, 2010.

[25] M. V. M. Kolundzija, C. Faller, "Reproducing sound fields using mimo acoustic channel inversion," *J. Audio Eng. Soc*, vol. 59, no. 10, pp. 721–734, 2011.

[26] S. Miyabe, M. Shimada, T. Takatani, H. Saruwatari, K. Shikano, "Multichannel inverse filtering with loudspeaker selection and enhancement for robust sound field reproduction," in *Proc. IWAENC 2006*, 2006.

[27] Y. Yuyama, S. Miyabe, H. Saruwatari, K. Shikano, "Hybrid structure of inverse filtering and DOA-parameterized wavefront synthesis," in *Proc. ICASSP 2008*, 2008, pp. 401–404.

[28] S. B. T. Sporer, J. Plogsties, "Carrouso - an european approach to 3d-audio," in *Audio Engineering Society Convention 110*, 2001.

[29] *Information technology – MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC).* ISO/IEC FCD 23003-2, 2010.

[30] J. Herre, S. Disch, "New concepts in parametric stereo coding in mpeg-4," in *Proc. ICME*, 2007, pp. 2–4.

[31] L. Villemoes, J. Herre, J. Breebaart, G. Hotho, S. Disch, H. Purnhagen, K. Kjörling, "Mpeg-surround: the forthcoming iso standard for spatial audio coding," in *Proc. AES*, 2006.

[32] C. Faller, F. Baumgarte, "Binaural cue coding - part ii: schemes and applications," *IEEE Trans. Speech Audio Process*, vol. 11, pp. 520–531, 2003.

[33] S. Tanoue, H. Hokari, S. Shimada, "The mono sound localization on the wave field synthesis with spatial transfer functions," *IEICE Technical Report*, vol. EA2007, no. 93, pp. 43–48, 2007.

[34] C. F. Van Loan, "Generalizing the singular value decomposition," *SIAM Journal on Numerical Analysis*, vol. 13, no. 1, pp. 76–83, 1976.

[35] C. Paige, "Computing the generalized singular value decomposition," *SIAM Journal on Scientific Computing*, vol. 7, pp. 1126–1146, 1986.

[36] Z. Bai, J. W. Demmel, "Computing the generalized singular value decomposition," *SIAM Journal on Scientific Computing*, vol. 14, pp. 1464–1486, 1993.

[37] Z. Bai, H. Zha, "A new preprocessing algorithm for the computation of the generalized singular value decomposition," *SIAM Journal on Scientific Computing*, vol. 14, pp. 1007–1012, 1993.

[38] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symp. on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[39] D. de Vries, "Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics," *J. Audio Eng. Soc.*, vol. 44, no. 12, pp. 1120–1131, 1996.

[40] E. W. Start, "Direct sound enhancement by wave field synthesis," Ph.D. dissertation, Delft University of Technology, 1997.

[41] E. Corteel, "Caractérisation et extensions de la wave field synthesis en conditions réelles d'écoute," Ph.D. dissertation, Université de Paris VI, Paris, France, 2004.

[42] D. J. K. F. L. Wightman, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.*, vol. 91, no. 3, 1992.

[43] H. Timo, "Two-way acoustic window using wave field synthesis," Master's thesis, Helsinki University of Technology, 2007.

[44] H. Wittek, "Perception of spatially synthesized sound fields - literature review about wfs," http://www.hauptmikrofon.de/HW/Wittek_WFS_LitReview.pdf, 2003.

[45] D. D. Rife, J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–443, 1989.

[46] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[47] R. J. Hanson, "A numerical method for solving fredholm integral equations of the first kind using singular values," *SIAM Journal on Numerical Analysis*, vol. 8, pp. 616–622, 1971.

[48] J. M. Varah, "On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems," *SIAM Journal on Numerical Analysis*, vol. 10, no. 2, pp. 257–267, 1973.

[49] O. Kirkeby, P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, 1993.

[50] P. A. Gauthier, A. Berry, W. Woszczyk, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *J. Acoust. Soc. Am.*, vol. 117, no. 2, pp. 662–678, 2006.

[51] N. Kamado, H. Hokari, S.Shoji, H. Saruwatari, K. Shikano, "Sound field reproduction by wavefront synthesis using directly aligned multi point control," *IEICE Trans. Fundamentals.*, vol. E94-A, no. 3, 2011.

[52] P. A. Gauthier, A. Berry, "Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory," *J. Acoust. Soc. Am.*, vol. 119, pp. 2721–2737, 2006.

[53] N. Stefanakis, F. Jacobsen, J. Sarris, "Effort variation regularization in sound field reproduction," *J. Acoust. Soc. Am.*, vol. 128, pp. 740–750, 2010.

[54] G. Golub, C. V. Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, 1996.

[55] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.

[56] C. V. Loan, "On the method of weighting for equality-constrained least-squares problems," *SIAM Journal on Numerical Analysis*, vol. 22, no. 5, pp. 851–864, 1985.

[57] S. Ise, "A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems," *Acustica*, vol. 85, pp. 78–87, 1999.

[58] G. Wolberg, "Image morphing: A survey," *Visual Computer*, vol. 14, pp. 360–372, 1998.

[59] N. Stefanakis, J. Sarris, F. Jacobsen, "Regularization in global sound equalization based on effort variation," *J. Acoust. Soc. Am.*, vol. 126, pp. 666–675, 2009.

[60] A. O. Santillán, "Spatially extended sound equalization in rectangular rooms," *J. Acoust. Soc. Am.*, vol. 110, pp. 1989–1997, 2001.

[61] *Telephone transmission quality, Methods for subjective determination of transmission quality*. ITU-T Recommendation P.800 Annex B, 1996.

[62] P. A. Gauthier, A. Berry, "Adaptive wave field synthesis for sound field reproduction: Theory, experiments, and future perspectives," *J. Audio Eng. Soc.*, vol. 55, no. 12, 2007.

[63] T. Yosuke, U. Shigefumi, S. Hiroshi, S. Kiyohiro, "On-line relaxation algorithm applicable to acoustic fluctuation for inverse filter in multichannel sound reproduction system," *IEICE Trans. Fundamentals*, vol. E88-A, no. 7, pp. 1747–1756, 2005.

[64] P. A. Gauthier, A. Berry, "Adaptive wave field synthesis for broadband active sound file reproduction: Signal processing," *J. Audio Eng. Soc.*, vol. 123, no. 4, 2008.

[65] ——, "Adaptive wave field synthesis for broadband active sound file reproduction: Experimental results," *J. Audio Eng. Soc.*, vol. 123, no. 4, 2008.

[66] N. Bleistein, *Mathematical methods for wave phenomena*. Academic Press Inc., Orlando, 1984.

Appendix

A. Derivation of attenuation law of WFS

In this section, we introduce to estimate the radiation profiles of the synthesized wavefront emitted by an infinite line with the driving function obtained by WFS [41]. Estimation in the frequency domain which is given by the stationary phase approximation [66] gives the attenuation law of synthesized wavefront.

According to Eq. (1) and Fig. 5, the synthesized wavefront emitted by the WFS driving function of the ideal of infinite size linear distribution of monopole sources (x_S, y_S) arranged along the x axis ($y_S = 0$) is given as

$$SWF(x_O, y_O, \omega) = S_P(\omega)\sqrt{\frac{\mathrm{j}k}{2\pi}}C(y_R, y_P)\int_{-\infty}^{\infty}\frac{e^{-\mathrm{j}kr_{PS}}}{\sqrt{r_{PS}}}\cos\theta_{PS}\frac{e^{-\mathrm{j}kr_{SO}}}{r_{SO}}dx_S, \tag{121}$$

where r_{PS} is the spatial distance between the primary source and the secondary point source, r_{SO} is the spatial distance between the secondary point source and the observation point and θ_{PS} is the angle between the y-axis and the line connecting the primary souce and the secondary point source. According to Eq. (121), the integral term can be written as

$$\begin{aligned} I(\omega) &= \int_{\infty}^{\infty}\frac{e^{-\mathrm{j}kr_{PS}}}{\sqrt{r_{PS}}}\cos\theta_{PS}\frac{e^{-\mathrm{j}kr_{SO}}}{r_{SO}}dx_S \\ &= \int_{-\infty}^{\infty} f(x_S)e^{-\mathrm{j}k\chi(x_S)}dx_S, \end{aligned} \tag{122}$$

where the functions $f(x_S)$ and $\chi(x_S)$ are given as

$$f(x_S) = \frac{1}{r_{SO}\sqrt{r_{PS}}}\cos\theta_{PS}, \tag{123}$$

$$\chi(x_S) = r_{SO} + r_{PS}. \tag{124}$$

The functions $f(x_S)$ and $\chi(x_S)$ are called envelope and phase function, respectively. $\chi(x_S)$ gives the derivatives of 1st and 2nd order as

$$\begin{aligned} \chi'(x_S) &= \frac{x_S - x_P}{r_{PS}} - \frac{x_O - x_S}{r_{SO}} \\ &= \sin\theta_{PS} - \sin\theta_{SO}, \end{aligned} \tag{125}$$

$$\begin{aligned}\chi''(x_S) &= \frac{1}{r_{PS}} - \frac{(x_S - x_P)^2}{r_{PS}^3} + \frac{1}{r_{SO}} - \frac{(x_O - x_S)^2}{r_{SO}^3} \\ &= \frac{1}{r_{PS}} - \frac{\sin^2\theta_{SO}}{r_{PS}} + \frac{1}{r_{SO}} - \frac{\sin^2\theta_{PS}}{r_{SO}} \\ &= \frac{\cos^2\theta_{SO}}{r_{PS}} + \frac{\cos^2\theta_{PS}}{r_{SO}},\end{aligned} \tag{126}$$

where θ_{SO} is the angle between the y-axis and the line connecting the secondary point source and the observation point. The stationary phase approximation gives the estimation of this type of integral. Consider a phase point $x_S = x_{S0}$ for the phase function $\chi(x_S)$. We assume that the phase point x_{S0} satisfy the condition,

$$\begin{cases} \chi'(x_{S0}) & = 0 \\ \chi''(x_{S0}) & \neq 0 \end{cases}, \tag{127}$$

and the phase function admit an extremum only at x_{S0}. It is a maximum if $\chi''(x_{S0}) < 0$, if a minimum $\chi''(x_{S0}) > 0$. On either side of $x = x_{S0}$, the phase function is strictly monotonic and thus the term $e^{-\mathrm{j}k\chi(x_S)}$ is oscillatory. The point $x = x_{S0}$ is then called the stationary phase point. In this case, the stationary phase point x_{S0} is unique and given by

$$\theta_{SO} = \theta_{PS} = \theta_0. \tag{128}$$

Then, $f'(x_S)$ and $f''(x_S)$ which is the derivatives of 1st and 2nd order of $f(x_S)$ are continuous on $]-\infty \sim \infty[$, $f(x)$ is monotone around x_{S0}, and $f(x_{S0}) \neq 0$, the Taylor expansion around x_{S0} of expression under the integral gives:

$$f(x_S)e^{-\mathrm{j}k\chi(x_S)} \approx f(x_{S0})e^{-\mathrm{j}k\left[\chi(x_{S0}) + \frac{(x_S - x_{S0})^2}{2}\chi''(x_{S0})\right]}. \tag{129}$$

Therefore, we obtain the approximation of $I(\omega)$ as

$$\begin{aligned}\hat{I}(\omega) &= \int_{-\infty}^{\infty} f(x_{S0})e^{-\mathrm{j}k\left[\chi(x_0) + \frac{(x_S - x_{S0})^2}{2}\chi''(x_0)\right]}dx_S \\ &= f(x_{S0})e^{-\mathrm{j}k\chi(x_{S0})}\int_{-\infty}^{\infty} e^{-\mathrm{j}k\frac{(x_S - x_{S0})^2}{2}\chi''(x_{S0})}dx_S.\end{aligned} \tag{130}$$

In making the change of variable

$$\kappa = (x_S, x_{S0})\sqrt{\frac{k\chi''(x_{S0})}{2}}, \tag{131}$$

and provided that $\chi(x_{S0})'' \neq 0$, Equation (130) can be written as

$$\hat{I}(\omega) = f(x_{S0}) \sqrt{\frac{2\pi}{k|\chi''(x_{S0})|}} e^{-\mathrm{j}k\chi(x_{S0})} \int_{-\infty}^{\infty} e^{-\mathrm{j}\mathrm{sign}(k\chi''(x_{S0}))\kappa^2} d\kappa, \tag{132}$$

where sign denotes the signum function. Note that

$$\int_{-\infty}^{\infty} e^{\pm\mathrm{j}\kappa^2} d\kappa = \sqrt{\pi} e^{\pm\frac{\pi}{4}}, \tag{133}$$

the solution of the integral is given as

$$\hat{I}(\omega) = f(x_{S0}) \sqrt{\frac{2\pi}{k|\chi''(x_{S0})|}} e^{-\mathrm{j}\left[k\chi(x_{S0}) + \mathrm{sign}(\chi''(x_{S0}))\frac{\pi}{4}\right]} \tag{134}$$

From Eqs. (126) and (128), $\chi''(x_{S0})$ is given as

$$\chi''(x_{S0}) = \cos^2\theta_0 \left(\frac{1}{r_{PS0}} + \frac{1}{r_{SO0}}\right), \tag{135}$$

where r_{PS0} and r_{SO0} are r_{PS} and r_{SO} in the stationary phase point, respectively. Finally, the pressure of the synthesized wavefront $SWF(x_O, y_O, \omega)$ can be rewritten as

$$\begin{aligned} SWF(x_O, y_O, \omega) &= S_P(\omega) \sqrt{\frac{jk}{2\pi}} C(y_R, y_P) \hat{I}(\omega) \\ &= S_P(\omega) C(y_R, y_P) \sqrt{\frac{r_{PS0} - r_{SO0}}{r_{SO0}}} \frac{e^{-\mathrm{j}k(r_{PS0} + r_{SO0})}}{r_{PS0} + r_{SO0}}. \end{aligned} \tag{136}$$

From this equation, the pressure law in front of the primary source can be written as

$$SWF(y_O, \omega) = S_P(\omega) \sqrt{\frac{|y_R|}{|y_O|}} \sqrt{\frac{|y_O| + |y_P|}{|y_R| + |y_P|}} \frac{e^{-\mathrm{j}k|y_O - y_P|}}{|y_O - y_P|}. \tag{137}$$

Therefore, we obtain the attenuation law denoted in Eq. (4).

Acknowledgements

This dissertation is a summary of two years and a half studies carried out at Graduate School of Information Science, Nara Institute in Technology, Japan.

I would like to express my sincere thanks to Professor Kiyohiro Shikano in Nara Institute of Science and Technology, my dissertation adviser, for his valuable guidance and constant encouragement.

I also would like to express my appreciation Professor Kazushi Ikeda of Nara Institute of Science and Technology, a member of the thesis committee, for his valuable comments to the thesis.

I also would like to express my appreciation Emeritus Professor Shoji Shimada and Technical Staff Haruhide Hokari in Nagaoka University of Technology for their continuous teaching and essential advice on both technical and non-technical issues.

I would especially like to express my deeply grateful and sincere acknowledgement to Associate Professor Hiroshi Saruwatari in Nara Institute of Science and Technology for his continuous teaching and essential advice on both technical and non-technical issues. This work could not have been accomplished his well-directed advice, beneficial suggestion, and fruitful discussion with him. I could learn a lot of things as a researcher from his attitude toward a study. I have always been happy to conduct researches with him.

I would also like to thank Associate Professor Tomoki Toda and Assistant Professor Hiromichi Kawanami of Nara Institute of Science and Technology for their instructive comments.

The former part of this work could never have been completed without the collaboration of many researchers. I would like to express my gratitude to Dr. Shigeki Miyabe, who is currently an Assistant Professor of Tsukuba University for his valuable suggestions and shrewd advice on techniques of sound field reproduction. Also, I would like to acknowledge to Dr. Yu Takahashi, who is currently a researcher in Yamaha Co., Ltd., Dr. Yamato Ohtani, who is currently a researcher in Toshiba Co., Ltd., and Dr. Keigo Nakamura, who is currently an engineer in Rakuten Co., Ltd., for their instructive advice on real-time implemen-

tation of my proposed audio object controller.

The later part of this work also could not have been achieved without collaborations of many researchers. I especially thank Mr. Ko Mizuno and Mr. Isao Kakuhari, who are researchers in Panasonic Corporation, for their beneficial and valuable comments on sound field reproduction and its driving devices.

A lot of staff and my research group members had supported me to carry out experiments and write this dissertation at Nara Institute of Science and Technology; I especially would like to express my appreciation for their valuable discussion on technical issues of speech signal processing and digital signal processing, and their arrangement of the comfortable computer environment in our laboratory. I also wish to represent my deeply gratitude to Mrs. Toshie Nobori, who is the secretary in our laboratory, for her kind help and support in all aspects of my research.

I appreciate to study together with all of students who had been in our research group at Nara Institute of Science and Technology. I thank Mr Shota Takeuchi, Mr. Rafael Torres, Mr. Yusuke Kisaki and Mr. Keigo Kubo, who are Ph.D. candidates in Nara Institute of Science and Technology, for their useful discussions on this work.

Finally, I am deeply grateful to all members of my family Aki Kamado and Ryo Kamado for their support they gave me during all these years.

List of Publications

Journal Papers

1. Noriyoshi Kamado, Haruhide Hokari, Shoji Shimada, Hiroshi Saruwatari, Kiyohiro Shikano, "Sound Field Reproduction by Wavefront Synthesis Using Directly Aligned Multi Point Control," IEICE Trans. Fundamentals, vol.E94-A, no.3, pp.907–920, 2011.

2. Noriyoshi Kamado, Shigeki Miyabe, Hiroyuki Nawata, Hiroshi Saruwatari, Kiyohiro Shikano, "Development of interactive controller for audio object localization based on spatial representative vector operation," Journal of Signal Processing, 2012. (in Japanese, accepted)

Conference Proceedings

1. Noriyoshi Kamado, Haruhide Hokari, Shoji Shimada, Hiroshi Saruwatari, Kiyohiro Shikano, "Comparative Study on Directly-Aligned Multi-Point Controlled Wavefront Synthesis and Wave Field Synthesis," APSIPA Annual Summit and Conference, pp. 242–245, 2009.

2. Noriyoshi Kamado, Haruhide Hokari, Shoji Shimada, Hiroshi Saruwatari, Kiyohiro Shikano, "Sound Field Reproduction by Wavefront Synthesis Using Directly Aligned Multi Point Control," Proc. of AES conference, 2010.

3. Noriyoshi Kamado, Hiroyuki Nawata, Hiroshi Saruwatari, Kiyohiro Shikano, Toshiyuki Nomura, "Interactive controller for audio object localization based on spatial representative vector operation," International Workshop on Acoustic Echo and Noise Control (IWAENC2010), Tel Aviv, Israel, Aug. 2010.

4. Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Robust sound field reproduction integrating multi-point sound field control and wave field synthesis," Proc. of the 36th International Conference on Acoustics, Speech and Signal Processing (ICASSP2011), pp.441–444, 2011.

5. Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Experimental evaluation of hybrid 3D sound field reproduction integrating multi-point sound field

control and wave field synthesis," Proc. of the 3D Systems and Applications (3DSA2011), pp.213–216, 2011.

6. S. Suzuki, S. Miyabe, N. Kamado, H. Saruwatari, K. Shikano, and T. Nomura, " Audio object individual operation and its application to earphone leakage noise reduction," Proceedings of International Symposium on Communications, Control and Signal Processing (ISCCSP2010), Th. 5.6, Feb. 2010.

7. Hiroyuki Nawata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Automatic music thumbnailing using localization information of audio object," APSIPA Annual Summit and Conference 2010 (APSIPA2010), pp.915–918, 2010.

8. Hiroyuki Nawata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Automatic musical thumbnailing based on audio object localization and its evaluation," Proc. of the 36th International Conference on Acoustics, Speech and Signal Processing (ICASSP2011), pp.41–44, 2011.

9. Masayuki Hirata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Spatial depth perception of focusing source by wave field synthesis for 3D sound reproduction," Proc. of the 3D Systems and Applications (3DSA2011), pp.137–140, 2011.

Domestic Conference

1. Noriyoshi Kamado, Haruhide Hokari, Shoji Shimada, Hiroshi Saruwatari, Kiyohiro Shikano, "Comparison of Wavefront Synthesized by Directly-Aligned Multi-Point Controlled Wavefront Synthesis and Wave Field Synthesis," Technical report of IEICE. EA, vol. 109, no. 136, EA2009-40, pp. 19-24, 2009.

2. Noriyoshi Kamado, Hiroyuki Nawata, Hiroshi Saruwatari, Kiyohiro Shikano, "Development of real-time audio localization control system," IEICE Technical Report, vol.110, no. 54, pp.13–18, 2010.

3. Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Sound reproduction based on multi-channel inverse filtering and WFS," IEICE Technical Report, vol. 110, no. 285, pp. 55–60, 2010.

4. Shota Suzuki, Shigeki Miyabe, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, Toshiyuki Nomura, "Audio object individual operation method and its application to earphone leakage noise reduction", IEICE Technical Report EA, EA2009-95, pp.13–18, 2009.

5. Hiroyuki Nawata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Automatic music thumbnailing using localization information of audio object", IEICE Technical Report EA, EA2010-29, pp.41–46, 2010.

6. Masayuki Hirata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Spatial depth perception control of focusing source by wave field synthesis," IEICE Technical Report EA, vol.110, no.471, pp.17–22, 2011.

7. Masayuki Hirata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Sound localization at several listening positions using wave field synthesis with sound separation and azimuth estimation," IEICE Technical Report EA, 2012.

8. Noriyoshi Kamado, Hiroyuki Nawata, Hiroshi Saruwatari and Kiyohiro Shikano, Interactive Controller for Audio Object Localization and Automatic Thumbnail Music Generator, International Society of Music Information Retrieval Conference (ISMIR) 2010, http://ismir2010.ismir.net/proceedings/late-breaking-demo-05.pdf

9. Noriyoshi Kamado, Haruhide Hokari, Shoji Shimada, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of Directly-aligned multi-point controlled wavefront synthesis in actual environment," Proc. of 2009 Autumn Meeting of Acoustical Society of Japan, 1-4-6, pp. 1407–1410, September 2009.

10. Noriyoshi Kamado, Haruhide Hokari, Shoji Shimada, Hiroshi Saruwatari, Kiyohiro Shikano, "Direct measurement of actual synthesized wavefront spectrogram using wavefront," Proc. of 12th Young Researcher's Joint meeting of Acoustical Society of Japan at Kansai, p.26, Dec 2009.

11. Noriyoshi Kamado, Shigeki Miyabe, Shota Suzuki, Hiroyuki Nawata, Hiroshi Saruwatari, Kiyohiro Shikano, " Development of audio object controller," Proc. of 2010 Spring Meeting of Acoustical Society of Japan, 3-P-28, pp. 931–932, 2010.

12. Noriyoshi Kamado, Hiroyuki Nawata, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of audio object controller based on directional clustering of spatial informa-

tion," Proc. of 2010 Autumn Meeting of Acoustical Society of Japan, 1-10-22, pp. 1403–1406, Sept. 2010.

13. Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation and measurement of secondary wavefront with hybrid wave field synthesis, " Proc. of 2011 Spring Meeting of Acoustical Society of Japa, 3-9-4, pp. 733-736, March 2011.

14. Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Compensation method of phase-gap problem of hybrid wave field synthesis ," Proc. of 2011 Autumn Meeting of Acoustical Society of Japan, 2-4-9, pp.657–660, 2011.

15. Shota Suzuki, Shigeki Miyabe, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, Toshiyuki Nomura, "Earphone leakage noise reduction method based on Audio object individual operation method", Proc. of 12th Young Researcher's Joint meeting of Acoustical Society of Japan at Kansai, p. 56, 2009.

16. Shota Suzuki, Shigeki Miyabe, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, Toshiyuki Nomura, "Evaluation of earphone leakage noise reduction method based on Audio object individual operation method", Proc. of 2010 Spring Meeting of Acoustical Society of Japan, 3-5-7, 2010.

17. Hiroyuki Nawata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Evaluation of Automatic Music Thumbnail Generator based on Audio Object Localization Estimator," Proc. of 2010 Autumn Meeting of Acoustical Society of Japan, 1-10-23, pp. 1407–1410, 2010.

18. Hiroyuki Nawata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Music Thumbnails Generation Method based on Time-spatial Information Analysis, " Proc. of 2011 Spring Meeting of Acoustical Society of Japan, 3-1-13, pp. 1063–1066, 2011.

19. Masayuki Hirata, Noriyoshi Kamado, Hiroshi Saruwatari, Kiyohiro Shikano, "Depth Control of Sound Images using Wavefront Synthesis," Proc. of 2011 Autumn Meeting of Acoustical Society of Japan, 2-4-8, pp.653-656, September 2011.

20. Yuji Onuma, Noriyoshi Kamado, Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano, "Implementation of Real-time Blind Spatial Subtraction Array on Kinect,",

Proc. of 3rd Young Researcher's Meeting of Advanced Language Information Forum, p.3, 2011.

21. Yuji Onuma, Noriyoshi Kamado, Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano, "Implementation and Evaluation of Blind Spatial Subtraction Array on Kinect," Proc. of 34th SIG-Challenge 2011, 2011.

Article

1. Noriyoshi Kamado, Yuji Onuma, Hiroshi Saruwatari, Kiyohiro Shikano, "Microphone Array Signal Processing using Kinect", INTERFACE , CQ publishing, Tokyo, pp.112–116, January, 2012.

Awards

1. Student Award on Convention of Acoustical Society of Japan, March, 2011.

2. Adapted the Exploratory IT Human Resources Project (MITOH Project) by Information-technology Promotion Agency, Japan (IPA), 2009 .

Open Source Software

1. Open RAF
   An implementation of real-time multichannel audio signal processing library
   http://openraf.sourceforge.net/