

# **Boundary Integral Approach to Sound Field Transform and Reproduction**

(音場の変換と再現への境界値積分アプローチ)

**Shoichi Koyama**

小山 翔一



## Abstract

Two engineering problems are investigated in this thesis: sound field reproduction and sound source localization. For sound field reproduction problem, methods that directly gives driving signals of secondary sources when a sound pressure distribution on a boundary of a target region is only known are proposed. These methods make it possible for signal conversion of sound field recording and reproduction by using commonly used microphones and loudspeakers. Specifically, when the array configurations are planar, linear, cylindrical, spherical, or circular, these methods are based on a transform filter called *wave field reconstruction (WFR) filter*. It is shown that these methods are computationally efficient and provides equivalent reproduction accuracy to conventional methods based on numerical algorithms. Additionally, a method, which enables higher reproduction accuracy when an approximate location of a reproducing sound source can be given as a prior information, is proposed. As an actual implementation, by applying the WFR filter to linear arrays of microphones and loudspeakers, a real-time sound field transmission system is proven feasible. For sound source localization problem, a parameter estimation method called *weighted integral method* is applied to a partial differential equation that constraints a sound source location and observation values. Then, an exact algebraic equation, which includes a sound source location as unknown variables and finite integrals of observation values, is derived. By using an algorithm based on this equation, locations of the sound sources is given as direct solutions of the algebraic equations and observation values can be obtained as spatial or spatio-temporal discrete Fourier transform of the sensor signals; this means that sound source locations are obtained very accurately at fine temporal resolution.



## 概要

本論文では、音場再現と音源位置推定という二つの問題を扱う。音場再現問題においては、再現対象領域の境界面上の音圧のみが既知の場合に、境界面上に配置された二次音源の駆動信号を得るための手法を提案する。これは従来困難であった、一般的なマイクロホン・スピーカ素子を用いて、音場の收音・再現のための信号変換を可能にするものである。特に、アレイ配置が平面、直線、円筒、球、円状の場合に、波面再構成フィルタ (wave field reconstruction filter: WFR filter) と呼ぶ変換フィルタを用いる手法を提案し、従来の数値解析的なアルゴリズムに基づく手法などに比べて、高速かつ効率的な手法であることを示す。さらに、再現対象のおおよその音源位置を事前情報として与えられる場合には、より高精度な再現が可能な手法についても示す。実際のシステムとして、直線状アレイのための WFR フィルタを適用することで、音場のリアルタイム伝送が実現可能であることを示す。音源位置推定問題においては、音源位置と観測値とを拘束する偏微分方程式に対し、荷重積分法と呼ばれるパラメータ推定手法を適用することで、未知数である音源位置と、観測値の有限区間での積分値とを含むような代数方程式を導く。これにより、音源位置が代数方程式の直接解として得られ、また観測値はセンサ信号の空間または時空間の離散フーリエ変換として得られることから、高精度かつ高時間分解能な音源位置推定が可能となる。



# Contents

|                  |   |           |
|------------------|---|-----------|
| <b>Chapter 1</b> | <b>Introduction</b>   | <b>1</b>  |
| 1.1              | Recording and reproducing physical phenomena . . . . .  | 1         |
| 1.2              | Spatial audio reproduction . . . . .  | 2         |
| 1.3              | Sound source localization . . . . .   | 3         |
| 1.4              | Outline . . . . .   | 4         |
| <b>Chapter 2</b> | <b>Mathematical Representaion of Sound Field</b>  | <b>6</b>  |
| 2.1              | Wave equation and Helmholtz equation . . . . .  | 6         |
| 2.2              | Kirchhoff-Helmholtz integral equation . . . . .   | 7         |
| 2.3              | Dirichlet and Neumann Green Functions . . . . .   | 9         |
| 2.4              | Fourier acoustics . . . . .   | 11        |
| 2.5              | Location constraint partial differential equation for single point source . . .                           | 12        |
| 2.6              | Special functions and selected identities . . . . .   | 14        |
| <b>Chapter 3</b> | <b>Sound Field Reproduction and its Previous Studies</b>  | <b>15</b> |
| 3.1              | Sound Field Reproduction Problem . . . . .  | 15        |
| 3.2              | Current sound field reproduction methods . . . . .  | 17        |
| 3.3              | Signal conversion for real-time sound field transmission . . . . .  | 22        |
| <b>Chapter 4</b> | <b>Inverse Wave Propagator in Sound Field Reproduction</b>  | <b>24</b> |
| 4.1              | Shifting reconstruction position based on inverse wave propagator . . . . .                               | 25        |
| 4.2              | Filter representation for shifting reconstruction line . . . . .  | 29        |
| 4.3              | Experiments . . . . .   | 30        |
| 4.4              | Conclusions . . . . .   | 34        |
| <b>Chapter 5</b> | <b>Wave Field Reconstruction Filter for Planar or Linear Arrays of Micro-<br/>phones and Loudspeakers</b> | <b>35</b> |
| 5.1              | WFR equation for planar distributions of receivers and secondary sources .                                | 36        |
| 5.2              | WFR equation for linear distributions of receivers and secondary sources . .                              | 38        |
| 5.3              | Implementation of WFR filter as FIR filter . . . . .  | 41        |
| 5.4              | Sound field shifting and rotating by parametric control of WFR filter . . . .                             | 51        |

|                   |  |            |
|-------------------|--|------------|
| 5.5               | WFR filter for directional secondary sources . . . . .   | 57         |
| 5.6               | Reproduction in reverberant environment by using circular harmonics beam-forming . . . . .     | 60         |
| 5.7               | Conclusions . . . . .  | 67         |
| <b>Chapter 6</b>  | <b>Wave Field Reconstruction Filter for Cylindrical Arrays of Microphones and Loudspeakers</b> | <b>68</b>  |
| 6.1               | WFR equation for cylindrical distributions of receivers and secondary sources                  | 69         |
| 6.2               | WFR equation for circular distributions of receivers and secondary sources .                   | 74         |
| 6.3               | Implementation of WFR filter for multiple linear arrays . . . . .                              | 79         |
| 6.4               | Experiments . . . . .  | 82         |
| 6.5               | Conclusions . . . . .  | 89         |
| <b>Chapter 7</b>  | <b>Wave Field Reconstruction Filter for Spherical Arrays of Microphones and Loudspeakers</b>   | <b>90</b>  |
| 7.1               | WFR equation for spherical distributions of receivers and secondary sources                    | 90         |
| 7.2               | WFR equation for circular distributions of receivers and secondary sources .                   | 96         |
| 7.3               | Implementation of WFR filter for circular arrays . . . . .                                     | 102        |
| 7.4               | Experiments . . . . .  | 104        |
| 7.5               | Conclusions . . . . .  | 108        |
| <b>Chapter 8</b>  | <b>Source Location Informed Sound-Pressure-to-Driving-Signal Conversion</b>                    | <b>110</b> |
| 8.1               | Sparse sound field representation . . . . .  | 111        |
| 8.2               | Problem formulation . . . . .  | 113        |
| 8.3               | MAP Estimation of Driving Signals of Secondary Sources . . . . .                               | 114        |
| 8.4               | Experiments . . . . .  | 118        |
| 8.5               | Conclusions . . . . .  | 120        |
| <b>Chapter 9</b>  | <b>Practical Implementations of Sound Field Transmission Systems</b>                           | <b>122</b> |
| 9.1               | System specifications . . . . .  | 122        |
| 9.2               | Measurement of reproduced sound field . . . . .  | 126        |
| 9.3               | Subjective listening test of sound localization . . . . .                                      | 130        |
| 9.4               | Demonstration system including 3D images . . . . .   | 134        |
| 9.5               | Conclusions . . . . .  | 135        |
| <b>Chapter 10</b> | <b>Direct Algebraic Method for Sound Source Localization</b>                                   | <b>136</b> |
| 10.1              | Weighted finite integral of LC-PDE in space . . . . .  | 136        |
| 10.2              | Weighted finite integral of LC-PDE in space and time . . . . .                                 | 143        |
| 10.3              | Experiments . . . . .  | 149        |
| 10.4              | Conclusions . . . . .  | 149        |

---

|                   |   |            |
|-------------------|---|------------|
| <b>Chapter 11</b> | <b>Conclusions</b>                              | <b>154</b> |
|                   | <b>Acknowledgement (in Japanese)</b>            | <b>156</b> |
|                   | <b>Bibliography</b>                             | <b>158</b> |
|                   | <b>List of Publications</b>                     | <b>166</b> |
|                   | <b>List of Awards</b>                           | <b>172</b> |
| <b>Appendix A</b> | <b>Definitions of Fourier transforms</b>        | <b>173</b> |
| A.1               | Fourier transform . . . . .                     | 173        |
| A.2               | Fourier series . . . . .                        | 175        |
| A.3               | Spherical harmonic transform . . . . .          | 175        |
| A.4               | Fourier transforms of Green functions . . . . . | 176        |



# Chapter 1

## Introduction

### 1.1 Recording and reproducing physical phenomena

Humans perceive everything around them by sensing physical phenomena. The surrounding environment has massive amounts of information, but we can extract necessary elements through cognitive processes. However, we do not discard subconscious information that is not extracted through cognitive processes, and this information affects our consciousness. This subconscious information actually gives us a sense of reality.

Physical phenomena cannot occur more than once at a particular time and place in nature. Many attempts have been made to record and reproduce physical phenomena in various fields. Technology for this has been developed mainly for visual and auditory information starting from the invention of the photography and phonograph. Today, three-dimensional imaging achieved by the stereoscopic principle are becoming common as 3D movies. Systems to reproduce physical phenomena have been investigated to reproduce not only the visual and auditory senses but also the other senses in virtual reality. Ivan E. Sutherland [1] envisioned the “ultimate display”, which serves as many senses as possible and constructs a virtual world. David Zeltzer [2] classified virtual reality systems by using three elements: autonomy (A), interaction (I), and presence (P). These elements are called the AIP cube, and a system fulfilling these requirements is considered to be an optimal virtual reality system. Many literature, especially science fiction stories, depict worlds in which these systems exist, so developing such systems appears to be a technological ideal.

#### 1.1.1 Acoustic phenomena and its perception

Many attempts have also been made to record and reproduce acoustic phenomena. Generally, microphones are used to receive sound waves, and these signals are converted to digital information and stored in a recording medium. For reproduction, this information is reconverted to analog signals and retransformed to sound waves by using loudspeakers. However, conven-

tional recording and reproducing methods result in the loss of a large amount of information when four-dimensional information of sound waves in the spatio-temporal domain is converted to one-dimensional information of an acoustic signal in the temporal domain or when combining one-dimensional information. This is because the information of acoustic signals in the temporal structure is sufficient to preserve elements of verbal information for communication. However, it is difficult to perceive a total sense of reality from such compressed information. To fulfill the presence requirement for auditory displays, four-dimensional information, i.e., spatio-temporal information, must be presented accurately.

Humans can actually perceive spatial elements of acoustical phenomena. Acoustic phenomena propagate as sound waves in the air. Humans sense these waves through the ears and recognize acoustical environments. In nature, a sound wave has four-dimensional information in the spatio-temporal domain. Humans use only the information obtained through two sensors, i.e., the ears, but through highly advanced cognitive processing. For example, the directions of the arrived sound waves are recognized through the interaural time difference (ITD) and interaural intensity difference (IID) [3]. Microscopic cues, such as the head-related transfer function (HRTF), which indicates the transfer characteristics from a sound source to both ears, play important roles in the recognition of the elevation angle of sound sources. By integrating spatial and temporal features of sounds, the decomposition and extraction of acoustic signals, such as focusing on a specific sound, is possible. This reintegration process is based on Gestalt psychology, and is called *auditory scene analysis* [4]. Overall, humans can process acoustical tasks in real-time and extract necessary elements from massive amounts of acoustic phenomena information, something that is very difficult with computers of our time.

We deal with two engineering problems: physical sound field reproduction and sound source localization. These two engineering challenges can provide solutions for numerous applications.

## 1.2 Spatial audio reproduction

The most widely used system today for spatial audio reproduction is the stereo loudspeaker setup. Usually, two loudspeakers are placed in front on the left and right of the listener. In recent years, multi-channel loudspeaker systems such as 5.1 surround [5] become widely used. A new 22.2 ch sound system is proposed by NHK (Japan Broadcasting Corporation) [6]. In these systems, amplitude and/or phase differences of the loudspeaker driving signals are controlled to present spatial sound images based on the perceptual phenomenon of summing localization [7]. Various techniques for controlling the perceived location of the spatial sound image, which are referred to as panning law, are proposed such as Vector Base Amplitude Panning (VBAP) [8]. A major drawback of these systems is that the optimal listening position called *sweet spot* is limited to the center of the loudspeakers or its neighboring region. Therefore, it is difficult to serve multiple listeners appropriate spatial impression. Moreover, it is difficult that driving signals of the loudspeakers are

directly calculated from the received signals of the microphones by using the panning law; therefore, these systems are usually limited to playing back artificial sounds designed by recording engineers.

Binaural audio systems aim to present simulated signals at ear entrances of the listeners. Typically, the binaural signals are recorded by using microphones mounted at the ear entrances of a listener or a dummy head, or artificially generated by filtering the source signals with HRTFs. Headphones are commonly used for playback to directly stimulate the ears with binaural signals. When loudspeakers are used, crosstalk cancellation techniques can be applied so that the signals at the ears of the listeners corresponds to the binaural signals. These techniques are called transaural [9]. It is important that the movement of the listeners are tracked and the binaural signals are accordingly controlled. Obviously, it is not a trivial task to accurately control the binaural signals of the multiple listeners. Furthermore, differences in HRTFs between individuals are a difficult issue for consistent and accurate presentation of binaural signals.

We deal with a physical sound field reproduction problem. Sound field reproduction methods aim to control a physical sound field in a large area. It can be expected that multiple listeners can perceive and share the same sound space. Obviously, a large number of loudspeakers are necessary for controlling a large area with higher spatial resolution. This is a defect in these methods that the other methods do not exhibit. However, the feasibility of such systems will gradually increase due to the evolution of devices and improvements in computational performance.

### 1.3 Sound source localization

Sound source localization methods can be classified into two categories in the context of array signal processing [10]: time difference of arrival (TDOA) and direction of arrival (DOA). The TDOA-based methods are based on a time difference of sound wave between multiple sensors [11]. The time-difference can be estimated by using cross correlation between the signals obtained by sensors. The sound source location can be obtained by using the estimated time-difference when the locations of the sensors are known. The DOA-based methods use signals obtained by relatively closely spaced sensors. Delay-and-sum beamformer, a well-known DOA-based method, creates sensitivity towards a particular direction by adding appropriately delayed multiple sensor signals [10]. Subspace methods such as MUSIC [12] and ESPRIT [13] are based on dividing into signal and noise subspaces by using eigenvalue expansion of the spatial correlation matrix of the sensor signals. Several attempts have been made to combine these methods with statistical approach in order to achieve highly robust and accurate location estimate [14–17]. Blind source separation methods such as independent component analysis (ICA) [18] aims to separate mixed-signals under the condition that the mixing process is unknown.

Most of current source localization methods use statistical properties of sensor signals such as spatial correlation. Therefore, long observation time is required for accurate estimation, but this

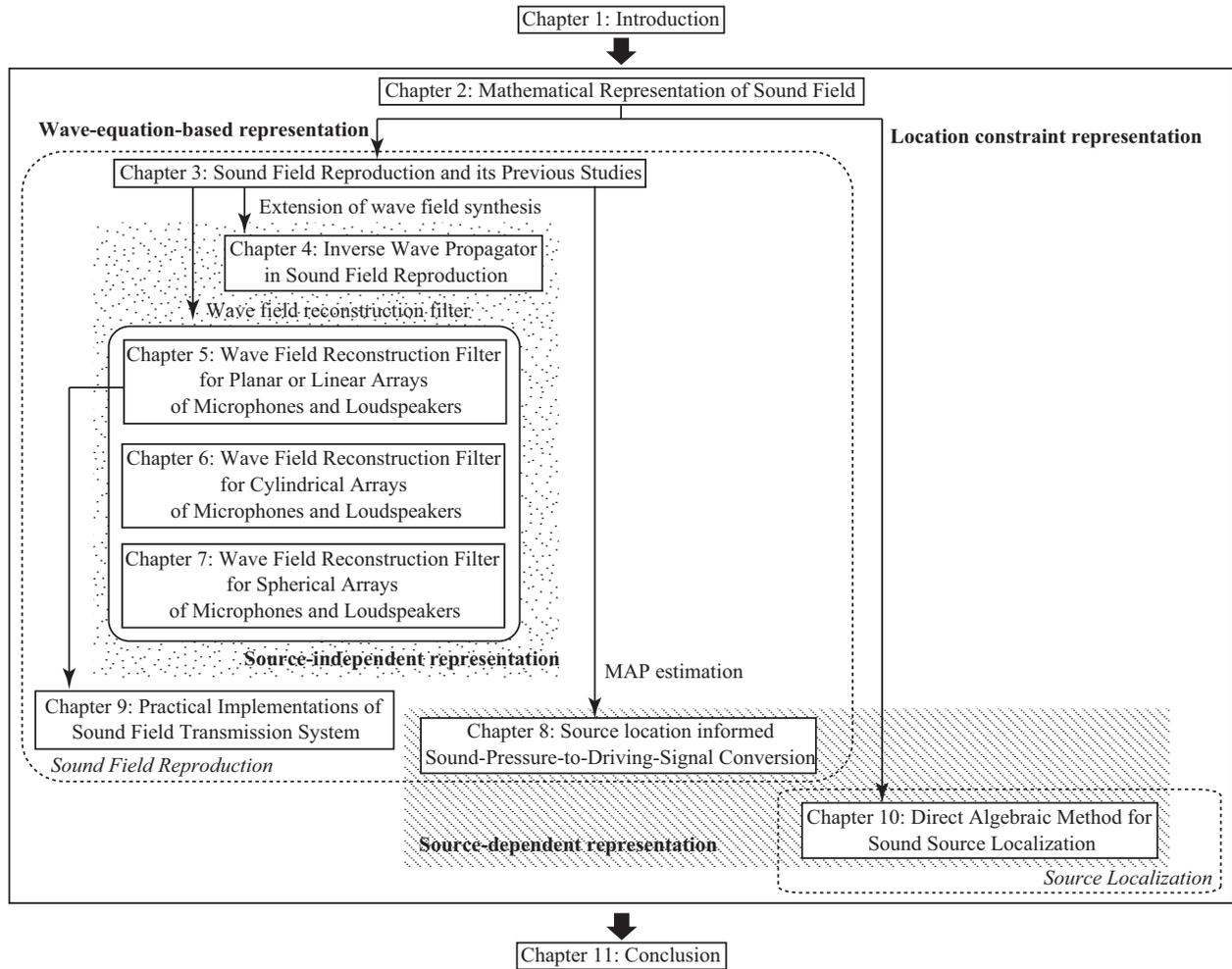


Fig. 1.1. Outline

causes error when the sound source moves fast. Furthermore, scanning of directional sensitivity in the DOA-based methods is computationally expensive, especially when direction and distance of the sound source needs to be estimated [19].

A sound source localization method based on a partial differential equation that constraints the source location and observation values is proposed. By applying a parameter estimation method called the weighted integral method [20], an exact algebraic equation for source localization, which includes finite integral of observation values, is derived; therefore, source location is obtained very accurately at fine temporal resolution by using the proposed algorithm.

## 1.4 Outline

The organization of this thesis is given in Fig. 1.1. In Chapter 2, mathematical representations of a sound field is presented for preliminaries. In Chapter 3, a sound field reproduction problem

and its previous studies are introduced. Problems on the current sound field reproduction methods are posed and are dealt with in the subsequent Chapters. For reproducing virtual sound sources in front of a loudspeakers, wave field synthesis (WFS), one of the sound field reproduction methods, is extended using inverse wave propagator, and this is given in Chapter 4. In Chapters 5, 6, and 7, wave field reconstruction (WFR) filtering method that directly converts received signals of a microphone array into driving signals of a loudspeaker array is proposed. The WFR filters are derived for planar, linear, cylindrical, spherical, and circular arrays of microphones and loudspeakers. In Chapter 8, a sound field reproduction method, which enables higher reproduction accuracy when an approximate location of a reproducing sound source can be given as a prior information, is proposed. Practical implementation using the WFR filter for linear arrays is presented in Chapter 9. The sound source localization problem is dealt with in Chapter 10. The proposed method is derived by applying the weighted integral method to the location constraint partial differential equation (LC-PDE). Finally, Chapter 11 concludes this thesis.

## Chapter 2

# Mathematical Representaion of Sound Field

Several equations that describe physical properties of sound waves are introduced. These equations play important role in this thesis.

### 2.1 Wave equation and Helmholtz equation

When  $p(\mathbf{r}, t)$  denotes an infinitesimal variation of acoustic pressure at  $\mathbf{r} = (x, y, z)$ ,  $p(\mathbf{r}, t)$  satisfies the acoustic wave equation for a homogeneous fluid with no viscosity as [21–23]:

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0, \quad (2.1)$$

where  $c$  is the speed of sound and  $\nabla^2$  is described as:

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

in the Cartesian coordinates.

The wave equation in the frequency domain is obtained by computing the Fourier transform of Eq. (2.1) with respect to  $t$  as [21–23]:

$$\nabla^2 P(\mathbf{r}, \omega) + k^2 P(\mathbf{r}, \omega) = 0, \quad (2.2)$$

where  $k = \omega/c$  is the wave number and  $P(\mathbf{r}, \omega)$  is described as:

$$P(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} p(\mathbf{r}, t) e^{j\omega t} dt. \quad (2.3)$$

Equation (2.2) is referred to as the homogeneous Helmholtz equation.

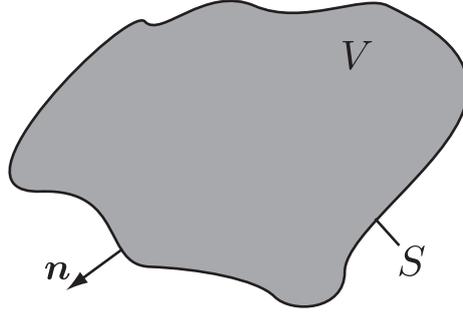


Fig. 2.1. Green's theorem is applied to volume  $V$  with bounding surface  $S$  and outward normal  $n$

## 2.2 Kirchhoff-Helmholtz integral equation

Let  $V$  be a three-dimensional volume bounded by a surface  $S$  as shown in Fig. 2.1. A position  $\mathbf{r}$  is inside this volume. When two functions  $\Psi(\mathbf{r})$  and  $\Phi(\mathbf{r})$  are assumed and are finite and continuous along with their first and second partial derivatives, the Green's theorem is applied as:

$$\iiint_V (\Phi \nabla^2 \Psi - \Psi \nabla^2 \Phi) dV = \iint_S \left( \Phi \frac{\partial \Psi}{\partial \mathbf{n}} - \Psi \frac{\partial \Phi}{\partial \mathbf{n}} \right) dS, \quad (2.4)$$

where  $\partial/\partial \mathbf{n}$  is the derivative with respect to the outward normal.

These two functions are assumed to satisfy the homogeneous Helmholtz equation Eq. (2.2) on  $S$  and in  $V$  as:

$$\nabla^2 \Phi + k^2 \Phi = 0 \quad (2.5)$$

$$\nabla^2 \Psi + k^2 \Psi = 0. \quad (2.6)$$

The integrand on the left hand side of Eq. (2.4) becomes zero as:

$$(\Phi \nabla^2 \Psi - \Psi \nabla^2 \Phi) = \Phi(-k^2 \Psi) - \Psi(-k^2 \Phi) = 0. \quad (2.7)$$

Therefore, the right hand side of Eq. (2.4) must be zero as:

$$\iint_S \left( \Phi \frac{\partial \Psi}{\partial \mathbf{n}} - \Psi \frac{\partial \Phi}{\partial \mathbf{n}} \right) dS = 0. \quad (2.8)$$

If the functions  $\Psi$  and  $\Phi$  have no singularities within or on the surface  $S$ , Eq. (2.8) must be satisfied.

Now assume the case that  $\Psi$  has a monopole singularity within  $V$  while  $\Phi$  is continuous as before. The singularity is located at  $\mathbf{r} = \mathbf{r}'$  and  $\Psi(\mathbf{r})$  is a solution of the inhomogeneous Helmholtz equation,

$$\nabla^2 \Psi + k^2 \Psi = -\delta(\mathbf{r} - \mathbf{r}'). \quad (2.9)$$

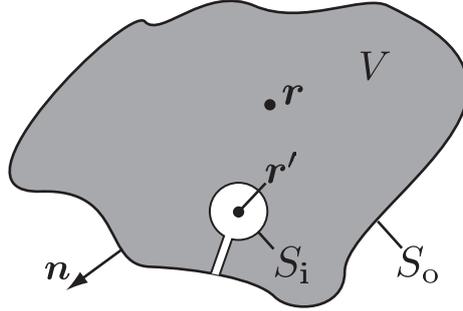


Fig. 2.2. Monopole singularity is located at  $\mathbf{r}' \in V$ .

The general solution of Eq. (2.9) is represented as the sum of the particular solution  $\Psi_p$  and homogeneous solution  $\Psi_h$  as:

$$\Psi = \Psi_p + \Psi_h. \quad (2.10)$$

For simplicity, we set  $\Psi_h$  as zero; therefore,  $\Psi$  becomes the three-dimensional free-field Green function as:

$$\Psi = \Psi_p = G(\mathbf{r}|\mathbf{r}') = \frac{e^{jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \quad (2.11)$$

As shown in Fig. 2.2, we modify the volume so that the singular point is excluded from  $V$  by surrounding the point by a vanishingly small sphere with surface  $S_i$ . The outer surface is denoted as  $S_o$  and  $S = S_o + S_i$ .

By applying Eq. (2.8),

$$\iint \left( \Phi \frac{\partial G(\mathbf{r}|\mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}|\mathbf{r}') \frac{\partial \Phi}{\partial \mathbf{n}} \right) dS_o + \lim_{\epsilon \rightarrow 0} \iint \left( \Phi \frac{\partial G(\mathbf{r}|\mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}|\mathbf{r}') \frac{\partial \Phi}{\partial \mathbf{n}} \right) dS_i = 0, \quad (2.12)$$

where  $\epsilon$  is the radius of  $S_i$ . Because the second integral of the above equation becomes  $\Phi(\mathbf{r}')$  (cf. [21]), the Kirchhoff-Helmholtz integral equation is derived as:

$$\Phi(\mathbf{r}') = \iint_{S_o} \left( G(\mathbf{r}|\mathbf{r}') \frac{\partial \Phi}{\partial \mathbf{n}} - \Phi \frac{\partial G(\mathbf{r}|\mathbf{r}')}{\partial \mathbf{n}} \right) dS_o. \quad (2.13)$$

The full statement of the Kirchhoff-Helmholtz integral equation is described by replacing  $\Psi$  into sound pressure  $P(\mathbf{r}, \omega)$  as:

$$\alpha P(\mathbf{r}', \omega) = \iint_{S_o} \left( G(\mathbf{r}|\mathbf{r}') \frac{\partial P(\mathbf{r}, \omega)}{\partial \mathbf{n}} - P(\mathbf{r}, \omega) \frac{\partial G(\mathbf{r}|\mathbf{r}')}{\partial \mathbf{n}} \right) dS_o, \quad (2.14)$$

where

$$\alpha = \begin{cases} 1 & \mathbf{r}' \in V \text{ and } \mathbf{r}' \notin S_o \\ 1/2 & \mathbf{r}' \in S_o \\ 0 & \mathbf{r}' \notin V \text{ and } \mathbf{r}' \notin S_o. \end{cases} \quad (2.15)$$

Refer to [21] for the derivation in the case of  $\mathbf{r}' \notin V$ . Even if the homogeneous solution  $\Psi_h$  in Eq. (2.10) is not zero, Eq. (2.14) still holds by replacing  $G(\mathbf{r}|\mathbf{r}')$  in Eq. (2.14) with  $G(\mathbf{r}|\mathbf{r}') + \Psi_h$ . The Kirchhoff-Helmholtz integral equation gives sound field inside  $V$  when the sources are outside  $V$  and sound pressures and its normal derivatives, i.e., velocity, on the surface of  $V$  are known.

## 2.3 Dirichlet and Neumann Green Functions

The Kirchhoff-Helmholtz integral equation, Eq. (2.14), require a knowledge of both the pressure and velocity on the surface in order to compute the sound field inside the surface even though either the pressure or velocity is inherently sufficient. This overspecification of the boundary conditions can be avoided by use of a special forms of the Green function.

Now consider a non-zero homogeneous solution  $\Psi_h$  of Eq. (2.9). Because  $\Psi_h$  will be generally a function of both  $\mathbf{r}$  and  $\mathbf{r}'$ , we define

$$g_N(\mathbf{r}|\mathbf{r}') \equiv \Psi_h, \quad (2.16)$$

and  $g_N$  satisfies Eq. (2.8). Therefore, the sum of  $g_N$  and the free-field Green function,  $G_N \equiv G + g_N$ , is still a solution to Eq. (2.14). The prescribed boundary condition for the Neumann problem is

$$\frac{\partial G_N}{\partial \mathbf{n}} = 0 \quad (2.17)$$

everywhere on  $S_o$ . This choice eliminates one of the terms in Eq. (2.14) as:

$$\alpha P(\mathbf{r}', \omega) = \iint_{S_o} G_N(\mathbf{r}|\mathbf{r}') \frac{\partial P(\mathbf{r}, \omega)}{\partial \mathbf{n}} dS_o. \quad (2.18)$$

Note that this boundary condition is applied only to  $G_N$ , not to  $P(\mathbf{r}, \omega)$ .  $G_N$  is called the Neumann Green function. When  $G_N$  is known, the normal derivative of pressure on  $S_o$  is only required to determine a sound field inside  $V$ . Practically, it is difficult to determine  $g_N$  especially when the surface does not lie on the contours of one of the separable coordinate systems.

The Dirichlet Green function,  $G_D \equiv G + g_D$ , is a second choice of Green function, which satisfies the Dirichlet boundary condition on  $S_o$  as:

$$G + g_D = 0. \quad (2.19)$$

Therefore, Eq. (2.14) becomes

$$\alpha P(\mathbf{r}', \omega) = - \iint_{S_o} P(\mathbf{r}, \omega) \frac{\partial G_N(\mathbf{r}|\mathbf{r}')}{\partial \mathbf{n}} dS_o. \quad (2.20)$$

When  $G_D$  is known, the pressure on  $S_o$  is only required to determine sound field inside  $V$ .

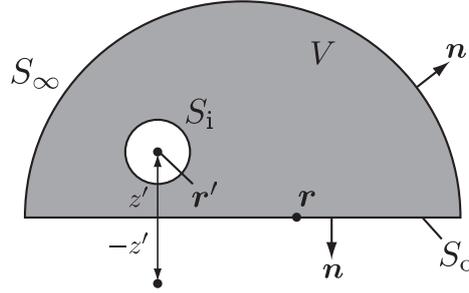


Fig. 2.3. Surface  $S$  is spread out to infinite plane.

### 2.3.1 Example with plane

As an example, the Neumann and Dirichlet Green functions for planar surface are derived. The surface is modified as shown in Fig. 2.3 so that the total surface becomes  $S = S_o + S_i + S_\infty$ . Because we let the radius of the hemispherical surface goes infinite, the surface integral on  $S_\infty$  becomes zero based on the Sommerfeld radiation condition [21, 23] and  $S_o$  becomes an infinite plane. In the Cartesian coordinates, the free-field Green function is described as:

$$G(\mathbf{r}|\mathbf{r}') = \frac{e^{jkR}}{4\pi R}, \quad (2.21)$$

where

$$R = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}. \quad (2.22)$$

Since  $\partial/\partial\mathbf{n} = -\partial/\partial z$  on  $S_o$ , Eq. (2.14) becomes

$$\alpha P(\mathbf{r}', \omega) = \frac{1}{4\pi} \iint_{S_o} \left( P(x, y, 0, \omega) \frac{\partial}{\partial y} \frac{e^{jkR}}{R} - \frac{e^{jkR}}{R} \frac{\partial}{\partial y} P(x, y, 0, \omega) \right) dS_o. \quad (2.23)$$

In order to construct the Neumann Green function, an image source located at  $(x', y', -z')$  with a distance of

$$R_{\text{im}} = \sqrt{(x - x')^2 + (y - y')^2 + (z + z')^2} \quad (2.24)$$

is introduced. Because this image source, given by  $e^{jkR_{\text{im}}}/R_{\text{im}}$ , is finite inside  $V$ , this is a solution to the homogeneous Helmholtz equation as required for  $g_N$ . The Neumann Green function is obtained by adding this to the free-field Green function as:

$$G_N(x, y, z|x', y', z') = \frac{1}{4\pi} \left( \frac{e^{jkR}}{R} + \frac{e^{jkR_{\text{im}}}}{R_{\text{im}}} \right). \quad (2.25)$$

When  $y = 0$ , since  $R = R_{\text{im}}$ , the normal derivative of  $G_N$  on  $S_o$  is calculated as:

$$\left. \frac{\partial G_N}{\partial z} \right|_{z=0} = 0. \quad (2.26)$$

Therefore, the boundary condition is indeed satisfied by  $G_N$ . Furthermore,  $G_N$  on  $S_o$  becomes simply twice the value of the free-field Green function as:

$$G_N|_{z=0} = \frac{1}{2\pi} \frac{e^{jkR}}{R}. \quad (2.27)$$

By using  $G_N$  for  $G$  in Eq. (2.23), the following equation is derived:

$$P(\mathbf{r}', \omega) = -\frac{1}{2\pi} \iint_{S_o} \frac{e^{jkR}}{R} \frac{\partial}{\partial z} P(x, y, 0, \omega) dS_o. \quad (2.28)$$

This equation is referred to as the Rayleigh integral of the first kind. Noted that  $\alpha$  in Eq. (2.23) becomes  $\alpha = 1$  in Eq. (2.28) even when  $\mathbf{r}'$  is not inside  $V$  (cf. [21]).

The Dirichlet Green function for the planar surface is also obtained by using the image source as:

$$G_D(x, y, z|x', y', z') = \frac{1}{4\pi} \left( \frac{e^{jkR}}{R} - \frac{e^{jkR_{im}}}{R_{im}} \right). \quad (2.29)$$

When  $y = 0$ ,  $G_D$  on  $S_o$  becomes zero as:

$$G_D|_{z=0} = 0. \quad (2.30)$$

The boundary condition is indeed satisfied by  $G_D$ . The normal derivative of  $G_D$  on  $S_o$  becomes twice of the normal derivative of  $G$  as:

$$\left. \frac{\partial G_D}{\partial z} \right|_{z=0} = 2 \frac{\partial}{\partial z} \frac{e^{jkR}}{R}. \quad (2.31)$$

By using  $G_D$  for  $G$  in Eq. (2.23), the following equation is derived:

$$P(\mathbf{r}', \omega) = \frac{1}{2\pi} \iint_{S_o} P(x, y, 0, \omega) \frac{\partial}{\partial z} \frac{e^{jkR}}{R} dS_o. \quad (2.32)$$

This equation is referred to as the Rayleigh integral of the second kind.

## 2.4 Fourier acoustics

Basics of Fourier acoustics are shortly introduced in the Cartesian coordinates [21, 24]. Similar relations can be derived in the cylindrical and spherical (polar) coordinates [21].

Let us consider a general unknown, steady state pressure distribution  $P(x, y, z, \omega)$  in a source-free half space,  $z > 0$ . This pressure can be uniquely and completely expressed by a sum of plane waves. The amplitudes and phases of each plane wave are denoted as  $\tilde{P}(k_x, k_y, \omega)$ , where  $k_x$  and  $k_y$  are wave numbers in the  $x$  and  $y$  directions, respectively. The wave number in the  $z$  direction is determined as:

$$k_z = \pm \sqrt{k^2 - k_x^2 - k_y^2}. \quad (2.33)$$

Note that the plane waves turns into evanescent waves under the condition of  $k_x^2 + k_y^2 > k^2$ . Therefore, because of the infinite extent and continuity of  $k_x$  and  $k_y$ ,  $P(x, y, z, \omega)$  is described as:

$$P(x, y, z, \omega) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \tilde{P}(k_x, k_y, \omega) e^{j(k_x x + k_y y + k_z z)}. \quad (2.34)$$

The constant  $1/4\pi^2$  is introduced in order to relate with the definition of Fourier transforms Eq. (A.10). Under the assumption that the sources are confined to  $z \leq 0$ , positive  $k_z$  values only hold. By substituting  $z = 0$  into Eq. (2.34), we have

$$P(x, y, 0, \omega) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \tilde{P}(k_x, k_y) e^{j(k_x x + k_y y)}, \quad (2.35)$$

which represents that the inverse Fourier transform of  $\tilde{P}(k_x, k_y, \omega)$  with respect to  $k_x$  and  $k_y$  equals to the pressure on the plane at  $z = 0$ ,  $P(x, y, 0, \omega)$ .

Based on the above discussion, once  $\tilde{P}(k_x, k_y)$  is known, which can be computed from the pressure on the plane at  $z = 0$ , then the pressure field in  $z \geq 0$  can be extrapolated by using Eq. (2.34). The general expression of the extrapolation from the plane at  $z = z'$  to a plane at  $z = z$  is represented as:

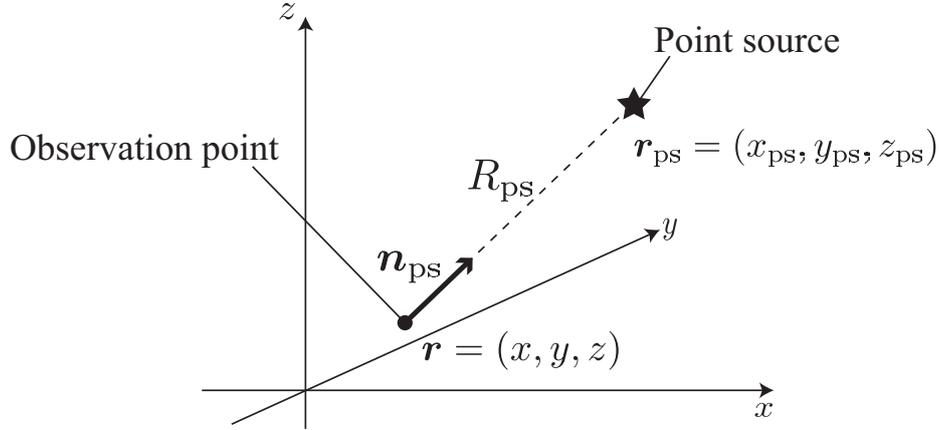
$$\tilde{P}(k_x, k_y, z) = \tilde{P}(k_x, k_y, z') e^{jk_z(z-z')} = \mathcal{F}_x \mathcal{F}_y [P(x, y, z')] e^{jk_z(z-z')}, \quad (2.36)$$

where  $z \geq z'$ . This relation also forms the foundation of the nearfield acoustical holography (NAH) [21].

## 2.5 Location constraint partial differential equation for single point source

The theoretical development so far is based on the homogeneous wave equation, Eq. (2.1), that constraints only on the property of wave propagation; therefore, constrains on sound sources are not directly imposed. A partial differential equation (PDE) that constraints directly on source locations may lead to various advanced algorithms that exploit or estimate source locations. For example, it may be difficult to develop an efficient algorithm for source localization based on Eq. (2.1) and subsequent equations because these equations does not directly include parameters of a source location. We introduce the location constraint partial differential equation (LC-PDE) that uniquely describes the relation between a wave field created by a single point source and observation values [25]. In Chapter 10, a source localization method is proposed based on the LC-PDE. It may be possible to apply the LC-PDE to various applications other than the source localization.

As shown in Fig. 2.4,  $\mathbf{r} = (x, y, z)$  and  $\mathbf{r}_{ps} = (x_{ps}, y_{ps}, z_{ps})$  denote the position vectors of the observation point and single point source, respectively. When a temporal waveform of the



**Fig. 2.4.** Relation between observation point and sound source location in LC-PDE.

sound source is denoted as  $s(t)$  and the sound speed is denoted as  $c$ , an observation value (sound pressure) at  $\mathbf{r}$  can be described as:

$$p(\mathbf{r}, t) = \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|} s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right), \quad (2.37)$$

where

$$|\mathbf{r} - \mathbf{r}_{ps}| = \sqrt{(x - x_{ps})^2 + (y - y_{ps})^2 + (z - z_{ps})^2}.$$

When the unit vector towards  $\mathbf{r}_{ps}$  from  $\mathbf{r}$  is denoted as  $\mathbf{n}_{ps}$ , the following equations can be derived:

$$\nabla |\mathbf{r} - \mathbf{r}_{ps}| = \frac{\mathbf{r} - \mathbf{r}_{ps}}{|\mathbf{r} - \mathbf{r}_{ps}|} \equiv -\mathbf{n}_{ps} \quad (2.38)$$

$$\nabla \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|} = -\frac{\mathbf{r} - \mathbf{r}_{ps}}{|\mathbf{r} - \mathbf{r}_{ps}|^3} = \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|^2} \mathbf{n}_{ps}. \quad (2.39)$$

Therefore, the spatial derivative of  $p(\mathbf{r}, t)$  is derived as:

$$\begin{aligned} \nabla p(\mathbf{r}, t) &= \nabla \left\{ \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|} s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right) \right\} \\ &= \nabla \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|} s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right) + \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|} \nabla s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right) \\ &= \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|^2} s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right) \mathbf{n}_{ps} + \frac{1}{c|\mathbf{r} - \mathbf{r}_{ps}|} \frac{\partial}{\partial t} s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right) \mathbf{n}_{ps}. \end{aligned} \quad (2.40)$$

The temporal derivative of  $p(\mathbf{r}, t)$  is described as:

$$\frac{\partial}{\partial t} p(\mathbf{r}, t) = \frac{1}{|\mathbf{r} - \mathbf{r}_{ps}|} \frac{\partial}{\partial t} s\left(t - \frac{|\mathbf{r} - \mathbf{r}_{ps}|}{c}\right). \quad (2.41)$$

Substituting Eqs. (2.37) and (2.41) into (2.40), the LC-PDE is derived by eliminating  $s(t)$  as:

$$\nabla p(\mathbf{r}, t) = \left\{ \frac{1}{R_{\text{ps}}} p(\mathbf{r}, t) + \frac{1}{c} \frac{\partial}{\partial t} p(\mathbf{r}, t) \right\} \mathbf{n}_{\text{ps}}, \quad (2.42)$$

where  $R_{\text{ps}} \equiv |\mathbf{r} - \mathbf{r}_{\text{ps}}|$  is the distance between the sound source location and the observation point.

## 2.6 Special functions and selected identities

Unless otherwise noted, the functions listed in Table 2.1 represent each special function. Several selected identities of these special functions are also described as follows.

**Table. 2.1.** Table of special functions

|  |                              |
|--|------------------------------|
| $n$ -th order Bessel function of first kind                      | $J_n(x)$                     |
| $n$ -th order Hankel function of first and second kind           | $H_n^{(1)}(x), H_n^{(2)}(x)$ |
| $n$ -th order spherical Bessel function                          | $j_n(x)$                     |
| $n$ -th order spherical Hankel function of first and second kind | $h_n^{(1)}(x), h_n^{(2)}(x)$ |
| Associated Legendre function of degree $n$ and order $m$         | $P_n^m(x)$                   |
| Spherical harmonic function of degree $n$ and order $m$          | $Y_n^m(\theta, \phi)$        |

The Hankel functions for large arguments ( $x \gg 0$ ) are approximated as:

$$H_n^{(1)}(x) \simeq \sqrt{\frac{2}{\pi x}} e^{j(x - (2n+1)\frac{\pi}{4})} \quad (2.43)$$

and

$$H_n^{(2)}(x) \simeq \sqrt{\frac{2}{\pi x}} e^{-j(x - (2n+1)\frac{\pi}{4})}. \quad (2.44)$$

The equations including Bessel functions and their derivatives are described as:

$$J_n(x)H_n^{(1)'}(x) - J_n'(x)H_n^{(1)}(x) = \frac{2j}{\pi x} \quad (2.45)$$

and

$$j_n(x)h_n^{(x)'}(x) - j_n'(x)h_n^{(1)}(x) = \frac{j}{x^2}. \quad (2.46)$$

The relation between Bessel and spherical Bessel functions is described as:

$$j_n(x)P_n^m(0) = \frac{j^m}{2j^n} \int_0^\pi J_m(x \sin \theta)P_n^m(\cos \theta) \sin \theta d\theta. \quad (2.47)$$

This equation is important to relate  $j_n(\cdot)$  with  $J_n(\cdot)$ .

## Chapter 3

# Sound Field Reproduction and its Previous Studies

Physical reproduction of a sound field aims at coinciding a sound field with a desired one by using multiple loudspeakers aligned on a surface of a target region. Sound field reproduction methods make it possible to calculate driving signals of the loudspeakers. Because these methods enables to reproduce the sound field in a large area, they have possibilities for numerous applications such as immersive audio systems.

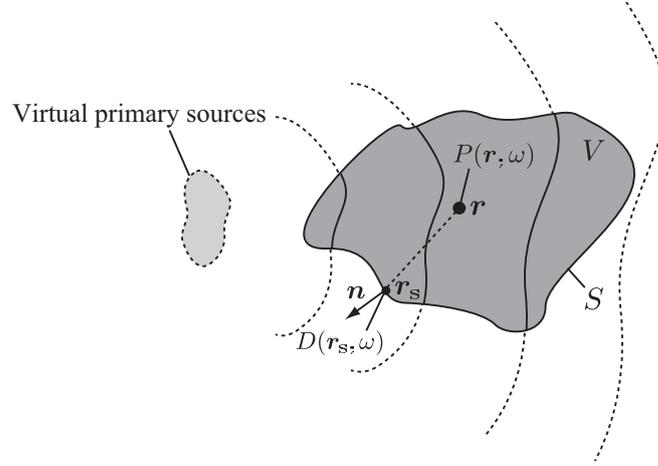
General formulation of sound field reproduction inside a closed surface is derived by Fazi, *et al* [26]. We introduce this formulation based on the single layer potential. Then, current sound field reproduction methods, wave field synthesis (WFS), higher order Ambisonics (HOA), and least squares (LS) method for sound pressure control, are derived. Finally, we point out problems of these current methods, which are dealt with in the following chapters.

### 3.1 Sound Field Reproduction Problem

As shown in Fig. 3.1, when  $P(\mathbf{r}, \omega)$  denotes sound pressure at  $\mathbf{r} \in V$  of the temporal frequency  $\omega$ , sound field reproduction problem is obtaining driving signals,  $D(\mathbf{r}_s, \omega)$  ( $\mathbf{r}_s \in S$ ), of secondary sources arranged on  $S$ , i.e., the boundary surface of  $V$ , so that  $P(\mathbf{r}, \omega)$  coincides with a desired sound field,  $P_{\text{des}}(\mathbf{r}, \omega)$ . It is assumed that  $P(\mathbf{r}, \omega)$  satisfies the homogeneous Helmholtz equation, Eq. (2.2), in  $V$ . When secondary sources are assumed to be continuously distributed on  $S$  and the transfer function between  $\mathbf{r}_s$  and  $\mathbf{r}$  is denoted as  $G(\mathbf{r} - \mathbf{r}_s, \omega)$ ,  $P(\mathbf{r}, \omega)$  can be described as:

$$P(\mathbf{r}, \omega) = \iint_S D(\mathbf{r}_s, \omega) G(\mathbf{r} - \mathbf{r}_s, \omega) dS. \quad (3.1)$$

Therefore, it is necessary to obtain  $D(\mathbf{r}_s, \omega)$  so that  $P(\mathbf{r}, \omega)$  coincide with  $P_{\text{des}}(\mathbf{r}, \omega)$ .



**Fig. 3.1.** Sound field inside  $V$  is reproduced by using secondary sources on  $S$ .

### 3.1.1 Single layer potential and jump relation

When each secondary source is assumed to be monopole, which means that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  is equivalent to three-dimensional free-field Green function, analytical formulation of  $D(\mathbf{r}_s, \omega)$  can be obtained. The three-dimensional free-field Green function introduced in Eq (2.11) is redefined as:

$$G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega) = \frac{e^{jk|\mathbf{r} - \mathbf{r}_s|}}{4\pi|\mathbf{r} - \mathbf{r}_s|}, \quad (3.2)$$

where  $k = \omega/c$  is the wave number and  $c$  is the sound speed. Therefore, Eq. (3.1) is described as:

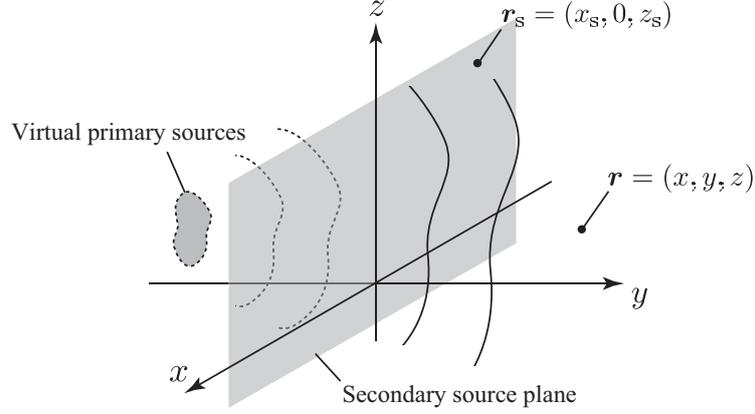
$$P(\mathbf{r}, \omega) = \iint_S D(\mathbf{r}_s, \omega) \frac{e^{jk|\mathbf{r} - \mathbf{r}_s|}}{4\pi|\mathbf{r} - \mathbf{r}_s|} dS. \quad (3.3)$$

Eq. (3.3) is known as the single layer potential [26, 27] or simple source formulation [21]. In the context of the single layer potential,  $D(\mathbf{r}_s, \omega)$  is referred to as the density of the potential. When  $P_{\text{int}}(\mathbf{r}_s, \omega)$  and  $P_{\text{ext}}(\mathbf{r}_s, \omega)$  respectively denotes interior and exterior sound fields of  $V$ , the single layer potential satisfies the homogeneous Helmholtz equation, Eq. (2.2), in both the regions.

When the density  $D(\mathbf{r}_s, \omega)$  is continuous, sound pressures on  $S$  of the internal and external regions,  $P_{\text{int}}(\mathbf{r}_s, \omega)$  and  $P_{\text{ext}}(\mathbf{r}_s, \omega)$ , are equivalent as:

$$P_{\text{int}}(\mathbf{r}_s, \omega) = P_{\text{ext}}(\mathbf{r}_s, \omega). \quad (3.4)$$

However, the normal derivatives of the sound pressures on  $S$ ,  $\nabla_{\mathbf{n}} P_{\text{int}}(\mathbf{r}_s, \omega)$  and  $\nabla_{\mathbf{n}} P_{\text{ext}}(\mathbf{r}_s, \omega)$ , are generally different. This difference originates from the presence of the layer of the secondary sources on  $S$ .



**Fig. 3.2.** Geometry of secondary sources for WFS in three-dimensions.

The jump relation of the single layer potential gives the relation between the normal derivatives of sound pressures and the density of the potential as [28]:

$$D(\mathbf{r}_s, \omega) = \nabla_{\mathbf{n}} P_{\text{int}}(\mathbf{r}_s, \omega) - \nabla_{\mathbf{n}} P_{\text{ext}}(\mathbf{r}_s, \omega). \quad (3.5)$$

When  $P_{\text{int}}(\mathbf{r}_s, \omega)$  is chosen to be equal to the desired sound field,  $P_{\text{des}}(\mathbf{r}, \omega)$ , the driving signals of the secondary sources can be obtained by using Eq. (3.5). Fazi, *et al.* [26] reformulated the current sound field reproduction methods, wave field synthesis (WFS) and higher order Ambisonics (HOA), based on Eq. (3.5).

## 3.2 Current sound field reproduction methods

We introduce three representative sound field reproduction methods, wave field synthesis (WFS), higher order ambisonics (HOA), and least squares (LS) method for sound pressure control.

### 3.2.1 Wave field synthesis

Wave field synthesis (WFS) [29] is a sound field reproduction method based on the Kirchhoff-Helmholtz integral that is derived in Section 2.2. We reformulate the Kirchhoff-Helmholtz integral Eq. (2.14) in this context as:

$$P(\mathbf{r}, \omega) = \iint_S \left( G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega) \frac{\partial P(\mathbf{r}_s, \omega)}{\partial \mathbf{n}} - P(\mathbf{r}_s, \omega) \frac{\partial G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega)}{\partial \mathbf{n}} \right) dS, \quad (3.6)$$

where  $\mathbf{r} \in V$ . This equation means that the sound field inside  $V$  can be controlled by using continuously distributed monopole and dipole secondary sources on  $S$  when the sound pressure and its normal derivative on  $S$  is known. However, it is difficult to employ dipole sources in practical implementations. In general, Eq. (3.6) is simplified to the Rayleigh integral of the first

kind that is derived in Section 2.3.1. As shown in Fig. 3.2,  $S$  is spread out as an infinite plane. In this context, Eq. (2.28) is reformulated as:

$$P(\mathbf{r}, \omega) = -2 \iint_{-\infty}^{\infty} \frac{\partial P(\mathbf{r}_s, \omega)}{\partial y_s} G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega) dx_s dz_s, \quad (3.7)$$

where the abbreviated notation  $\partial/\partial y_s$  means the directional gradient in the direction of  $y$  at  $\mathbf{r}_s$ . The dipole source term in Eq. (3.6) is eliminated in Eq. (3.7); therefore, monopole secondary sources distributed on the  $x$ - $z$ -plane at  $y = 0$  are necessary to reproduce the sound field in the half space of  $y > 0$ . Based on Eq. (3.7), the driving signals of the secondary sources can be obtained as:

$$D_{\text{WFS}}(\mathbf{r}_s, \omega) = -2 \frac{\partial P(\mathbf{r}_s, \omega)}{\partial y_s}. \quad (3.8)$$

This equation indicates that the sound pressure gradient in the direction of  $y$  of the desired sound field is necessary to obtain the driving signals of the secondary sources.

When considering reproduction on the  $x$ - $y$ -plane at  $z = 0$  only, linear secondary source distribution along the  $x$ -axis is approximately applicable. The conventional formulation of the driving signals for WFS in two-dimensions is based on the stationary phase approximation of Eq. (3.7) [30–32]. However, this formulation requires monopole assumption of primary sources. Spors, *et al.* [33] formulated WFS in two-dimensions based on the Rayleigh integral of the first kind in two-dimensions as [21]:

$$P(\mathbf{r}, \omega) = -2 \int_{-\infty}^{\infty} \frac{\partial P(\mathbf{r}_s, \omega)}{\partial y_s} G_{2D}(\mathbf{r} - \mathbf{r}_s, \omega) dx_s, \quad (3.9)$$

where  $G_{2D}(\mathbf{r} - \mathbf{r}_s, \omega)$  is the two-dimensional Green function defined as

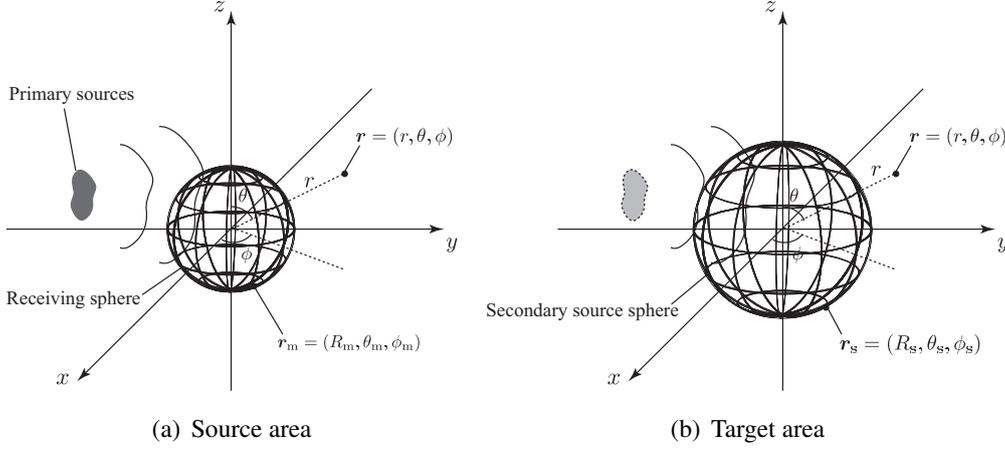
$$G_{2D}(\mathbf{r} - \mathbf{r}_s, \omega) = \frac{j}{4} H_0^{(1)}(k|\mathbf{r} - \mathbf{r}_s|). \quad (3.10)$$

For complete reproduction on the  $x$ - $y$ -plane at  $z = 0$ , each secondary source must exhibit line source characteristics, Eq. (3.10). However, enclosed loudspeakers, which are generally used as secondary sources, have approximately monopole characteristics,  $G_{3D}(\cdot)$ , especially in low frequencies. Therefore, Eq. (3.9) is approximated under the assumption of the Hankel function for large arguments, Eq. (2.43), as:

$$G_{2D}(\mathbf{r} - \mathbf{r}_s, \omega) \simeq \sqrt{\frac{2\pi j|\mathbf{r} - \mathbf{r}_s|}{k}} G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega). \quad (3.11)$$

Therefore, the driving signal for linear secondary sources can be described as:

$$\begin{aligned} D_{\text{WFS}}(\mathbf{r}_s, \omega) &= -2 \sqrt{\frac{2\pi j|\mathbf{r} - \mathbf{r}_s|}{k}} \frac{\partial P(\mathbf{r}_s, \omega)}{\partial y_s} \\ &\simeq -2 \sqrt{\frac{2\pi j\zeta}{k}} \frac{\partial P(\mathbf{r}_s, \omega)}{\partial y_s}. \end{aligned} \quad (3.12)$$



**Fig. 3.3.** Geometries of receivers and secondary sources for HOA.

This equation indicates that the amplitude can only be corrected for one listener position by the constant factor  $\zeta$ . This approximation that derives from mismatch of secondary source characteristics results in artifacts, such as faster amplitude decay than desired.

Ahrens and Spors [34, 35] proposed an alternative formulation for planar or linear secondary source distribution referred to as spectral division method (SDM). The SDM is the method to derive analytical form of driving signals based on inversion of spatio-temporal frequency spectrum.

### 3.2.2 Higher order Ambisonics

Ambisonics is the method proposed by Gerzon [36] and is extended to Higher Order Ambisonics (HOA) [37–40]. Typically, HOA includes encoding and decoding stages that correspond to recording and reproduction of sound field. As shown in Fig. 3.3, it is assumed that receivers and secondary sources are arranged on spheres, i.e., receiving and secondary source spheres. The receivers are assumed to be mounted on a rigid spherical baffle. The receiving and secondary source spheres are placed in the source and target areas, respectively. The position vectors on the receiving and secondary source spheres are respectively denoted as  $\mathbf{r}_m = (R_m, \theta_m, \phi_m)$  and  $\mathbf{r}_s = (R_s, \theta_s, \phi_s)$ , and  $R_m$  and  $R_s$  are the radii of the spheres.

In the encoding stage, spherical harmonic coefficients of the sound field in the source area,  $A_n^m(\omega)$ , is estimated from sound pressure distribution on the receiving sphere,  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$ , as [38]:

$$A_n^m(\omega) = -jk^2 R_m^2 h_n'(kR_m) \int_0^{2\pi} \int_0^\pi P_{\text{rcv}}(\mathbf{r}_m, \omega) Y_n^m(\theta_m, \phi_m)^* \sin \theta_m d\theta_m d\phi_m, \quad (3.13)$$

where  $n$  and  $m$  represent the order of spherical harmonics. By introducing the assumption of the rigid spherical baffle in the source area, Eq. (3.13) can be calculated, preventing numerical instabilities [41, 42].

In the decoding stage, the driving signals of secondary sources are calculated by using  $A_n^m(\omega)$  estimated in the encoding stage. Loudspeakers are discretely arranged on the secondary source sphere and their positions are denoted as  $\mathbf{r}_{s,l} = (R_s, \theta_{s,l}, \phi_{s,l})$ , where  $l$  ( $1 \leq l \leq L$ ) is the index of the loudspeakers. When the driving signal of the  $l$ -th loudspeaker is denoted as  $D_l(\omega)$ , a mode-matching equation that relates synthesized spherical harmonic spectrum to  $A_n^m(\omega)$  is described as [38]:

$$\sum_{l=1}^L D_l(\omega) Y_n^m(\theta_{s,l}, \phi_{s,l})^* = \frac{A_n^m(\omega)}{jkh_n(kR_s)}. \quad (3.14)$$

Each secondary source is assumed to be monopole. When the maximum order of spherical harmonics is  $N$ , the total number of mode to be matched  $(n, m)$  is  $(N + 1)^2$ .

Eq. (3.14) can be written in matrix notation for all spherical harmonics up to maximum order  $N$  as:

$$\Psi \mathbf{D}(\omega) = \Xi(\omega), \quad (3.15)$$

where  $\mathbf{D}(\omega)$  is the  $L$ -size vector of the driving signals  $D_l(\omega)$ , each element of  $(N + 1)^2 \times L$ -size matrix  $\Psi$  is represented as

$$\Psi_{\nu l} = Y_n^m(\theta_{s,l}, \phi_{s,l})^* \quad (3.16)$$

for  $\nu = n^2 + n + m + 1$ , and  $\Xi(\omega)$  is the  $(N + 1)^2$ -size vector and each element is described as

$$\Xi_{\nu}(\omega) = \frac{A_n^m(\omega)}{jkh_n(kR_s)}. \quad (3.17)$$

When the number of modes to be matched is less than the number of loudspeakers, i.e.,  $(N + 1)^2 < L$ , Eq. (3.15) is solved as minimum-norm solution as:

$$\hat{\mathbf{D}}^{\text{HOA}}(\omega) = \Psi^H (\Psi \Psi^H)^{-1} \Xi(\omega), \quad (3.18)$$

where  $^H$  denotes the conjugate transpose. If the computation of Eq. (3.18) is ill-conditioned, a regularized solution is generally used as:

$$\hat{\mathbf{D}}^{\text{HOA}}(\omega) = \Psi^H (\Psi \Psi^H + \beta \mathbf{I})^{-1} \Xi(\omega), \quad (3.19)$$

where  $\beta$  denotes a pre-selected regularization parameter and  $\mathbf{I}$  is an  $(N + 1)^2 \times (N + 1)^2$  unit matrix.

Various encoding and decoding methods are formulated analytically or numerically [37–40,43].

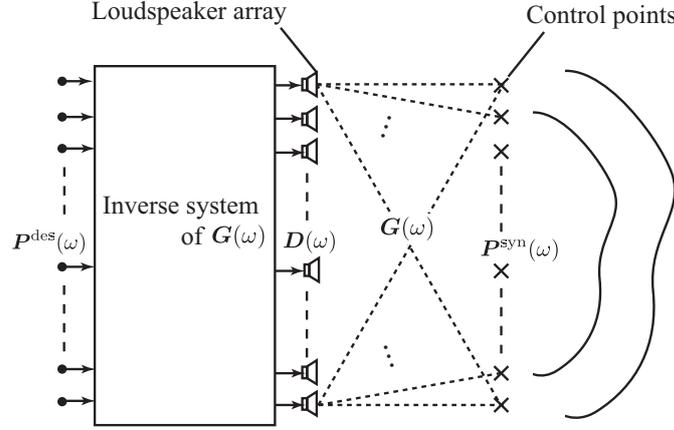


Fig. 3.4. Geometry of loudspeakers and control points for LS method.

### 3.2.3 Least squares method for sound pressure control

Methods based on least squares (LS) algorithm aim to control sound pressures at discrete points in the target area [44–49]. These methods can be derived as the extension of multi-point control of sound pressure by using inverse filtering [50]. Related works can be found in the context of room compensation for WFS [51–53].

As shown in Fig. 3.4, the control points are arranged in the target area. The number of loudspeakers and control points are denoted as  $L$  and  $M$ , respectively. Synthesized and desired sound pressures at the control points in the temporal-frequency domain are respectively denoted as

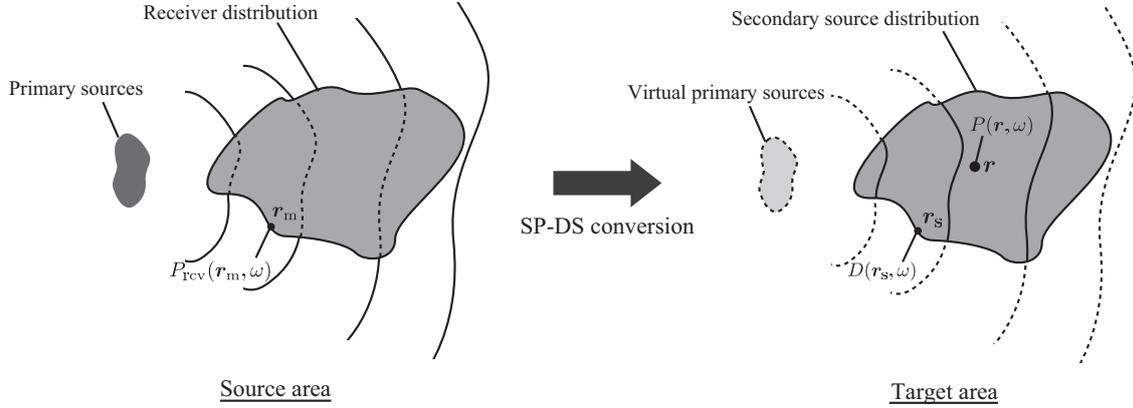
$$\begin{aligned} \mathbf{P}^{\text{syn}}(\omega) &= [P_1^{\text{syn}}(\omega), \dots, P_M^{\text{syn}}(\omega)]^T \\ \mathbf{P}^{\text{des}}(\omega) &= [P_1^{\text{des}}(\omega), \dots, P_M^{\text{des}}(\omega)]^T. \end{aligned}$$

The transfer function matrix is denoted as  $\mathbf{G}(\omega) \in \mathbb{C}^{L \times M}$ , which has the transfer function between each loudspeaker and control point in each element. These elements are given by measuring or modeling transfer functions. When the driving signals of loudspeakers are denoted as the  $L$ -size vector  $\mathbf{D}(\omega)$ , the synthesized sound pressures are described as:

$$\mathbf{P}^{\text{syn}}(\omega) = \mathbf{G}(\omega)\mathbf{D}(\omega). \quad (3.20)$$

The objective of the LS method is to solve the following minimization problem at temporal frequency  $\omega$ :

$$\begin{aligned} \hat{\mathbf{D}}^{\text{LS}}(\omega) &= \arg \min_{\mathbf{D}(\omega)} \|\mathbf{P}^{\text{syn}}(\omega) - \mathbf{P}^{\text{des}}(\omega)\|_2^2 \\ &= \arg \min_{\mathbf{D}(\omega)} \|\mathbf{G}(\omega)\mathbf{D}(\omega) - \mathbf{P}^{\text{des}}(\omega)\|_2^2. \end{aligned} \quad (3.21)$$



**Fig. 3.5.** Direct conversion of  $P_{rcv}(\mathbf{r}_m, \omega)$  into  $D(\mathbf{r}_s, \omega)$  for reproducing sound field.

When the number of control points is equal to or more than the number of loudspeakers,  $M \geq L$ , Eq. (3.21) is solved as least squares solution as

$$\hat{\mathbf{D}}^{\text{LS}}(\omega) = \mathbf{G}^\dagger(\omega) \mathbf{P}^{\text{des}}(\omega), \quad (3.22)$$

where  $\mathbf{G}^\dagger(\omega) = (\mathbf{G}^H(\omega) \mathbf{G}(\omega))^{-1} \mathbf{G}^H(\omega)$ . If the computation of Eq. (3.22) is ill-conditioned, which can be evaluated by the condition number defined as the ratio of the largest and smallest singular value of  $\mathbf{G}(\omega)$  [44, 54], the power-regularizing term is generally added to the cost function, Eq. (3.21). This method is referred to as the regularized LS method, and the driving signal based on this method is described as [47]:

$$\hat{\mathbf{D}}^{\text{LSreg}}(\omega) = (\mathbf{G}^H(\omega) \mathbf{G}(\omega) + \beta \mathbf{I})^{-1} \mathbf{G}^H(\omega) \mathbf{P}^{\text{des}}(\omega), \quad (3.23)$$

where  $\beta > 0$  denotes a pre-selected regularization parameter, and  $\mathbf{I}$  is an  $M \times M$  unit matrix.

When the geometry of the target area is a half space as in WFS, the control points can be aligned on the boundary of the target area, i.e., plane or line [48]. If the geometry is a closed region and the control points are aligned on the boundary surface, the inverse of the transfer function matrix cannot be uniquely determined due to forbidden frequencies [21, 55]. Ise [56] proposed to utilize both sound pressure and sound pressure gradient as controlled variables, which is referred to as the boundary surface control principle.

An alternative cost function introducing an  $\ell_1$ -regularization term to Eq. (3.21) is proposed by Lilis, *et al* [57].

### 3.3 Signal conversion for real-time sound field transmission

The above-described sound field reproduction methods make it possible to calculate the driving signals of the secondary sources (loudspeakers),  $D(\mathbf{r}_s, \omega)$ , for reproducing the desired sound field,  $P_{\text{des}}(\mathbf{r}, \omega)$ . In order to achieve spatial audio recording and reproducing systems, for example,

sound field transmission systems, it is preferable that  $D(\mathbf{r}_s, \omega)$  can be calculated with a sound pressure distribution on the boundary surface of the target area,  $P_{rcv}(\mathbf{r}_m, \omega)$ , that is obtained by using receivers (microphones) in the source area (Fig. 3.5). This is because any parameters of the desired sound field other than the sound pressures, for example, source positions, directions, and original signals, are unknown and difficult to obtain. Therefore, methods for directly transforming the sound pressures into the driving signals are necessary. We call this type of transformation sound-pressure-to-driving-signal (SP-DS) conversion.

The driving signals of WFS is calculated as Eq. (3.8) or (3.12). These equations include a planar or linear distribution of sound pressure gradient. Because it is difficult to obtain them by using ordinary microphones, WFS cannot be used for SP-DS conversion. HOA can be regarded as SP-DS conversion methods through encoding and decoding processes, Eqs. (3.13) and (3.18). However, these two stages are not suitable for real-time transmission systems from the viewpoint of computational complexity. The LS-based methods can be applied to SP-DS conversion by using Eq. (3.22) or (3.23). However, these extended methods from the multi-point control technique causes several difficulties in designing and applying LS-based filters as discussed in Chapter 5.

In the following five chapters, we propose five SP-DS conversion methods. In Chapter 4, a method for reproducing virtual sound sources in front of the planar or linear loudspeaker array is proposed. The proposed method is derived by introducing inverse wave propagator in WFS and can be used as an SP-DS conversion. In Chapter 5, 6, and 7, SP-DS conversion methods for planar, linear, cylindrical, spherical, and circular arrays of microphones and loudspeakers are proposed. Transform filters defined as wave field reconstruction (WFR) filters, which converts the received signals of the microphones array into the driving signals of the loudspeaker array, are analytically derived based on physical properties of acoustic waves. The WFR filters are derived in the spatial frequency spectrum domain and the signal conversions are represented as spatial convolution forms. Therefore, the WFR filters have many advantages in terms of filter design, filter size, computational cost, and filter stability. Although methods for direct SP-DS conversion are important, approximate locations of reproducing sound sources may be available, for example, by using the method presented in Chapter 10. Other types of sensors are also useful to detect locations of reproducing sound sources, such as talkers and musicians. In Chapter 8, an SP-DS conversion method for exploiting the prior information of the reproducing sound source locations is proposed. A practical implementation of sound field transmission system is presented in Chapter 9.

## Chapter 4

# Inverse Wave Propagator in Sound Field Reproduction

For future applications of sound field reproduction, it would be useful to create sound images in front of loudspeaker arrays to synchronize with 3D visual images. However, general sound field reproduction methods have been formulated for reconstructing a sound field in which sound sources exist behind a loudspeaker array. We propose a method for reconstructing a sound field in which sound sources exist in front of a planar or linear loudspeaker array.

Recent studies have shown that it is possible to reproduce point sources in front of a loudspeaker array by applying the time reversal principle [58–63]. Such sources are referred to as focused sources. The focused-source method requires source positions and spherical wave approximation of sound sources to be reproduced to calculate the driving signals of loudspeakers. Therefore, this method cannot be applied to SP-DS conversion.

The proposed method can be applied to SP-DS conversion, and is based on the combination of sound field reproduction using a planar or linear loudspeaker array and inverse wave propagator for a planar or linear microphone array. Virtual sound sources are recreated in front of a loudspeaker array when the desired sound field is given as an inversely obtained sound field in the opposite direction to wave propagation. Such inverse problems of wave propagation are called acoustical holography [21, 64] or inverse wavefield extrapolation [65, 66] and used for analysis of sound fields. We derive the combined formulation of sound field reproduction and inverse wave propagator, which is described as the transform from planar or linear sound pressure distribution to the driving signals of planar or linear secondary sources. The filter representation to displace the sound field to be reproduced towards the listener in front of a loudspeaker array is derived and determined by setting only a displacement parameter for shifting. The proposed method makes it possible to place virtual sound sources between the listener and loudspeaker array.

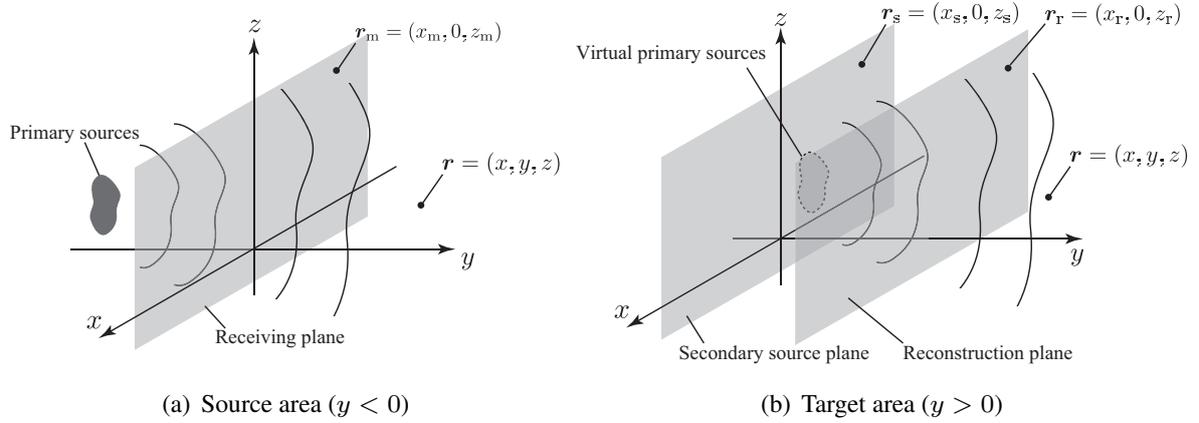


Fig. 4.1. Shifting reconstruction plane.

## 4.1 Shifting reconstruction position based on inverse wave propagator

To reproduce primary sources in front of secondary sources, we assume that the sound field captured at the receiving plane is reconstructed at an arbitrary shifted plane in front of secondary sources, as shown in Fig. 4.1. This plane is defined as the reconstruction plane. When the shift distance is larger than that between primary sources and the receiving plane, primary sources are virtually recreated in front of secondary sources. It is assumed that sound pressure distribution is obtained at the receiving plane by using continuously distributed omni-directional receivers. The resulting formulation is derived as the transformation from the sound pressure distribution at the receiving plane into the driving signals in order to shift the reconstruction plane forward, i.e., SP-DS conversion form. The transformation is determined by shifting distance only. Therefore, the sound pressure distribution at the receiving plane is only required to obtain the driving signals of secondary sources. On the contrary, the conventional focused-source method requires decomposed parameters of the desired sound field [59–62].

To shift the reconstruction plane forward, the sound field at the secondary source plane is obtained from the sound field at the reconstruction plane that coincides with the sound field at the receiving plane. To obtain information on the sound field in the opposite direction of wave propagation, the analytical solution of inverse wave propagator in a similar manner to holography [64, 67, 68] is adopted. This is because the sound field between virtual primary and secondary sources is ensured by the wave equation while the back region of virtual primary sources is inverted.

### 4.1.1 Inverse wave propagator for shifting reconstruction plane

The position vectors on the receiving, secondary source, and reconstruction plane are denoted as  $\mathbf{r}_m = (x_m, 0, z_m)$ ,  $\mathbf{r}_s = (x_s, 0, z_s)$ , and  $\mathbf{r}_r = (x_r, 0, z_r)$ , respectively. The sound pressures at  $\mathbf{r}_s$  and  $\mathbf{r}_r$  in the target area are denoted as  $P(\mathbf{r}_s, \omega)$  and  $P(\mathbf{r}_r, \omega)$ , respectively. Here,  $\omega$  is the temporal frequency. The sound pressure at  $\mathbf{r}_m$  in the source area is denoted as  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$ . Propagation directly from the secondary source plane  $\mathbf{r}_s$  to the reconstruction plane  $\mathbf{r}_r$  is described as the Rayleigh integral of the second kind, Eq. (2.32), derived in Section 2.3.1. In this context, Eq. (2.32) is reformulated as:

$$P(\mathbf{r}_r, \omega) = \iint_{-\infty}^{\infty} P(\mathbf{r}_s, \omega) \mathcal{K}(\mathbf{r}_r, \mathbf{r}_s) dx_s dz_s, \quad (4.1)$$

where

$$\mathcal{K}(\mathbf{r}_r, \mathbf{r}_s) = \frac{\partial}{\partial y_s} \left\{ \frac{e^{jk|\mathbf{r}_r - \mathbf{r}_s|}}{2\pi|\mathbf{r}_r - \mathbf{r}_s|} \right\}. \quad (4.2)$$

The back propagation of Eq. (4.1) can be derived by using the inverse of Eq. (4.2), which is called inverse wave propagator [64]. Now, we assume that inverse wave propagator  $\mathcal{K}^{-1}(\mathbf{r}_s, \mathbf{r}_r)$  can be approximated as the homogeneous component only in place of the sum of the homogeneous and inhomogeneous components:

$$\mathcal{K}^{-1}(\mathbf{r}_s, \mathbf{r}_r) \simeq \frac{\partial}{\partial y_r} \left\{ -\frac{e^{-jk|\mathbf{r}_s - \mathbf{r}_r|}}{2\pi|\mathbf{r}_s - \mathbf{r}_r|} \right\}. \quad (4.3)$$

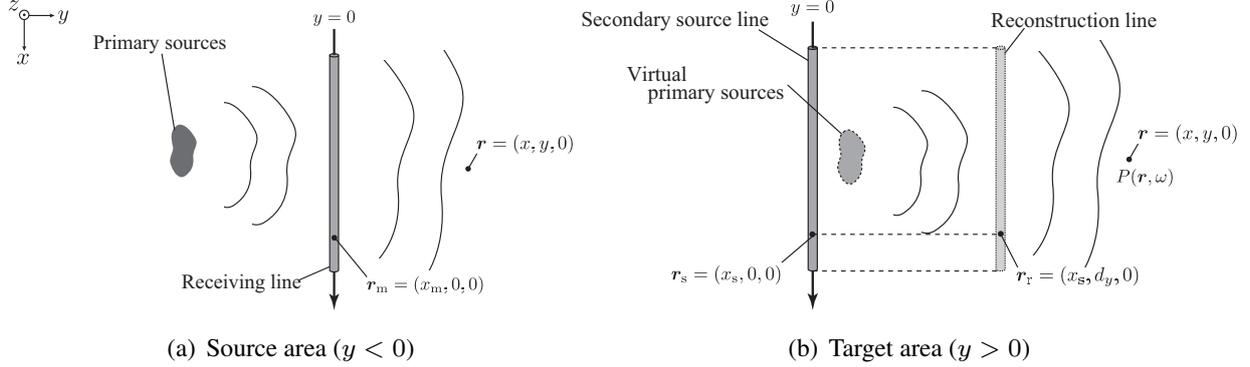
Therefore, the inversion formula of Eq. (4.1) is approximated as

$$\begin{aligned} P(\mathbf{r}_s, \omega) &\simeq \iint_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) \frac{\partial}{\partial y_r} \left\{ -\frac{e^{-jk|\mathbf{r}_s - \mathbf{r}_r|}}{2\pi|\mathbf{r}_s - \mathbf{r}_r|} \right\} dx_r dz_r \\ &= - \iint_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) (y_s - y_r) \frac{jk|\mathbf{r}_s - \mathbf{r}_r| + 1}{2\pi|\mathbf{r}_s - \mathbf{r}_r|^3} e^{-jk|\mathbf{r}_s - \mathbf{r}_r|} dx_r dz_r. \end{aligned} \quad (4.4)$$

A sound field in front of the secondary sources must satisfy the homogeneous equation to reproduce virtual sound sources in this region, preserving causality. Therefore, the approximation introduced in Eq. (4.4) is required for deriving the stable solution of  $\mathcal{K}^{-1}(\mathbf{r}_s, \mathbf{r}_r)$ . A similar approximation for the focused-source method can be found in Ref. [63].

The driving signal of the secondary sources for shifting the reconstruction plane,  $D(\mathbf{r}_s, \omega)$ , is derived based on WFS. By substituting Eq. (4.4) into Eq. (3.8),  $D(\mathbf{r}_s, \omega)$  is described as:

$$\begin{aligned} D(\mathbf{r}_s, \omega) &= -2 \frac{\partial}{\partial y_s} \left[ - \iint_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) (y_s - y_r) \frac{jk|\mathbf{r}_s - \mathbf{r}_r| - 1}{2\pi|\mathbf{r}_s - \mathbf{r}_r|^3} e^{-jk|\mathbf{r}_s - \mathbf{r}_r|} dx_r dz_r \right]_{y_s=0} \\ &= \iint_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) \left\{ \frac{jk|\mathbf{r}_s - \mathbf{r}_r| + 1}{\pi|\mathbf{r}_s - \mathbf{r}_r|^3} + y_r^2 \frac{k^2|\mathbf{r}_s - \mathbf{r}_r|^2 - 3jk|\mathbf{r}_s - \mathbf{r}_r| - 3}{\pi|\mathbf{r}_s - \mathbf{r}_r|^5} \right\} \\ &\quad e^{-jk|\mathbf{r}_s - \mathbf{r}_r|} dx_r dz_r. \end{aligned} \quad (4.5)$$



**Fig. 4.2.** Shifting reconstruction line.

The sound pressure distribution at the reconstruction plane is given by the sound pressure distribution at the receiving plane as  $P(\mathbf{r}_r, \omega) = P_{\text{rcv}}(\mathbf{r}_m, \omega)$ . Equation (4.5) is calculated only by setting the position of the reconstruction plane. Therefore, the reconstruction plane can be arbitrarily shifted in front of secondary source plane. Note that the distribution of the sound pressure gradient at the receiving plane is replaced by the distribution of sound pressure. This means that omni-directional receivers can be applied at the receiving plane and Eq. (4.5) can be applied as an SP-DS conversion.

When the driving signals of secondary sources are calculated based on Eq. (4.5), the sound pressure distribution at the receiving plane is reconstructed at the shifted reconstruction plane in the target area. The sound field between the secondary source and reconstruction planes follows Eq. (4.1), except inhomogeneous components. Therefore, primary sources are virtually recreated in front of secondary source plane when the shift distance is larger than the distance between primary sources and the receiving plane. However, the sound field between virtual primary sources and the secondary source plane is inverted because secondary sources emit a sound wave travels towards virtual primary sources.

### 4.1.2 Inverse wave propagator for shifting reconstruction line

In linear secondary source distribution, the receiving and reconstruction planes are replaced with the receiving and reconstruction lines, respectively. We assume the two-dimensional inverse propagation of sound pressure distribution from the reconstruction line  $\mathbf{r}_r$  to the secondary source line  $\mathbf{r}_s$ . This is derived in a similar manner as the three-dimensional case by neglecting inhomogeneous components:

$$\begin{aligned}
 P(\mathbf{r}_s, \omega) &\simeq \int_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) \frac{\partial}{\partial y_r} \left\{ -\frac{j}{2} H_0^{(2)}(k|\mathbf{r}_s - \mathbf{r}_r|) \right\} dx_r \\
 &= - \int_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) \frac{jk}{2} \frac{y_s - y_r}{|\mathbf{r}_s - \mathbf{r}_r|} H_1^{(2)}(k|\mathbf{r}_s - \mathbf{r}_r|) dx_r
 \end{aligned} \tag{4.6}$$

The driving signal of secondary sources for shifting the reconstruction line,  $D(\mathbf{r}_s, \omega)$ , is derived based on WFS for linear secondary sources. By substituting Eq. (4.6) into Eq. (3.12),  $D(\mathbf{r}_s, \omega)$  is described as:

$$\begin{aligned}
D(\mathbf{r}_s, \omega) &= -2\sqrt{\frac{2\pi j\zeta}{k}} \frac{\partial}{\partial y_s} \left[ - \int_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) \frac{jk}{2} \frac{y_s - y_r}{|\mathbf{r}_s - \mathbf{r}_r|} H_1^{(2)}(k|\mathbf{r}_s - \mathbf{r}_r|) dx_r \right]_{y_s=0} \\
&= -\sqrt{\frac{2\pi k\zeta}{j}} \int_{-\infty}^{\infty} P(\mathbf{r}_r, \omega) \\
&\quad \left\{ \frac{|\mathbf{r}_s - \mathbf{r}_r|^2 - 2y_r^2}{|\mathbf{r}_s - \mathbf{r}_r|^3} H_1^{(2)}(k|\mathbf{r}_s - \mathbf{r}_r|) + \frac{y_r^2}{|\mathbf{r}_s - \mathbf{r}_r|^2} k H_0^{(2)}(k|\mathbf{r}_s - \mathbf{r}_r|) \right\} dx_r
\end{aligned} \tag{4.7}$$

The sound pressure distribution at the reconstruction line is given by the sound pressure distribution at the receiving line. Equation (4.7) is calculated only by setting the position of the reconstruction line, the same as with Eq. (4.5). Therefore, the reconstruction line can be arbitrarily shifted in front of the secondary source line. In the linear case, it is assumed that primary sources only exist on the  $x$ - $y$ -plane at  $z = 0$ .

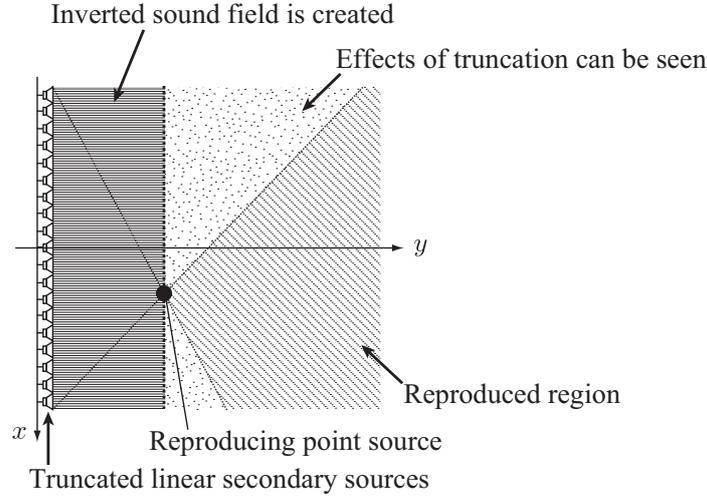
### 4.1.3 Comparison With Focused-Source Method

The driving signals of secondary sources for focused sources are presented in Ref. [62]. We compared the formulations of the proposed and focused-source methods in two-dimensional reproduction. In the focused-source method, the positions of the focused sources  $\mathbf{r}_{fs} = (x_{fs}, y_{fs})$  and the original signals of primary sources  $S(\omega)$  must be given, and the primary sources are assumed as line sources located at the focal points. In this context, the driving signals of secondary sources for focused sources are described as:

$$D_{FS}(\mathbf{r}_s, \omega) = \sqrt{2\pi\zeta} jk \frac{y_{fs}}{|\mathbf{r}_s - \mathbf{r}_{fs}|} H_1^{(2)}(k|\mathbf{r}_s - \mathbf{r}_{fs}|). \tag{4.8}$$

This equation is suitable when decomposed parameters of primary sources are known. On the other hand, it is possible to apply Eq. (4.7) when the sound pressure distribution at the receiving line is known while parameters of primary sources are unknown, i.e., SP-DS conversion.

The sound field reproduced by focused sources are only correct in their front region [62]. The back region of focused sources is inverted because secondary sources emit a sound wave that travels towards the focal points. This limitation coincides with that of the proposed method in which the back region of virtual primary sources is inverted.



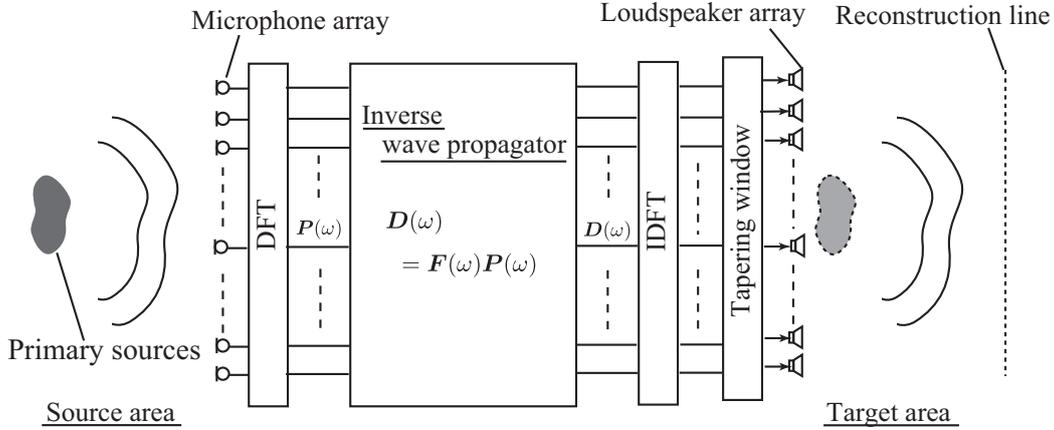
**Fig. 4.3.** Limited reproduced region by using truncated secondary sources when virtual point source is recreated in front of secondary sources.

## 4.2 Filter representation for shifting reconstruction line

We describe the filter representation applied as SP-DS conversion using linear arrays. Practically, spatially discrete microphones and loudspeakers are used as receiving and secondary source lines, respectively. The discretization leads to artifacts of spatial aliasing [33, 69]. The spatial truncation of the distributions of the receivers and secondary sources must be introduced. The truncation of secondary sources leads to limitations of the listening area, as shown in Fig. 4.3 [62]. Additionally, truncation leads to diffraction from the ends of the secondary source distribution. This artifacts can be reduced by applying a tapering window to the driving signals [30, 33]. The truncation of receiving line leads to a smaller reproduced region when the distance between the virtual point source and secondary sources is smaller than that between the primary point source and receivers.

A block diagram of the proposed method is depicted in Fig. 4.4. The sound pressure distribution in the source area is obtained by using omni-directional microphone array. A loudspeaker array is arranged in the target area, and each loudspeaker is assumed to be omni-directional. The numbers of microphones and loudspeakers are denoted as  $M$  and  $L$ , respectively. Each position of the microphone and loudspeaker is denoted as  $\mathbf{r}_{m,i} = (x_{m,i}, 0, 0)$  ( $1 \leq i \leq M$ ) and  $\mathbf{r}_{s,j} = (x_{s,j}, 0, 0)$  ( $1 \leq j \leq L$ ), respectively. The position of the reconstruction line is denoted as  $\mathbf{r}_{r,i} = (x_{r,i}, d_y, 0)$ , where  $d_y$  is the shift distance. The received signals of the microphone array and the driving signals of the loudspeaker array in the temporal frequency domain are respectively denoted as

$$\begin{aligned} \mathbf{P}(\omega) &= [P_1(\omega), \dots, P_M(\omega)]^T \\ \mathbf{D}(\omega) &= [D_1(\omega), \dots, D_L(\omega)]^T. \end{aligned}$$



**Fig. 4.4.** Block diagram of inverse wave propagator for shifting reconstruction line.

According to Eq. (4.7),  $\mathbf{D}(\omega)$  can be calculated as

$$\mathbf{D}(\omega) = \mathbf{F}(\omega)\mathbf{P}(\omega), \quad (4.9)$$

where  $\mathbf{F}(\omega)$  is  $M \times L$ -size matrix and  $F_{ij}(\omega)$ ,  $(i, j)$ -th element of  $\mathbf{F}(\omega)$ , is described as:

$$F_{ij}(\omega) = -\sqrt{\frac{2\pi k\zeta}{j}} \left\{ \frac{|\mathbf{r}_{s,j} - \mathbf{r}_{r,i}|^2 - 2d_y^2}{|\mathbf{r}_{s,j} - \mathbf{r}_{r,i}|^3} H_1^{(2)}(k|\mathbf{r}_{s,j} - \mathbf{r}_{r,i}|) + \frac{d_y^2}{|\mathbf{r}_{s,j} - \mathbf{r}_{r,i}|^2} kH_0^{(2)}(k|\mathbf{r}_{s,j} - \mathbf{r}_{r,i}|) \right\}, \quad (4.10)$$

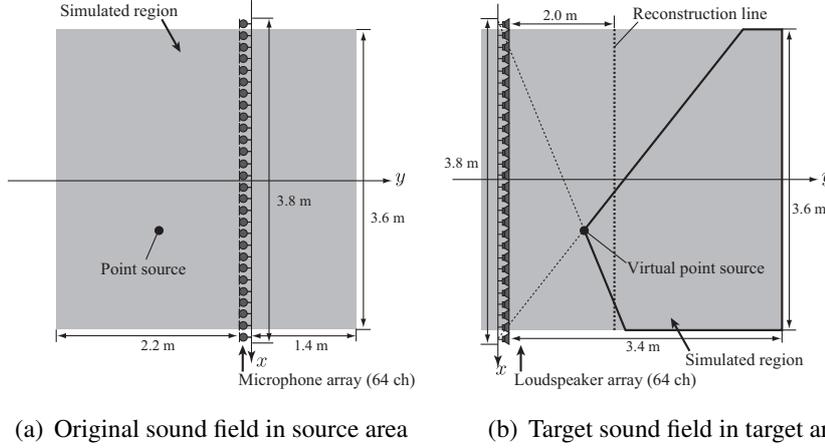
where

$$|\mathbf{r}_{s,j} - \mathbf{r}_{r,i}| = \sqrt{(x_{s,j} - x_{r,i})^2 + d_y^2}. \quad (4.11)$$

Therefore, the filter of shifting the reconstruction line is determined simply by setting  $d_y$ . A sufficient length of the frame for convergence of the filter in the time domain is needed.

### 4.3 Experiments

We conducted numerical simulations of reproducing virtual point sources to compare the proposed and focused-source methods under the free-field assumption. Figure 4.5 shows the simulation set up. Identical linear microphone and loudspeaker arrays were located at  $y = 0$ , with 64 channels in each array. The directivity of the array elements were assumed to be omni-directional. The array elements were equally spaced 6.0 cm apart, so the array lengths were 3.84 m. Therefore, the spatial Nyquist frequency was about 2.8 kHz. The point sources as primary sources located at  $y < 0$ , i.e., the source area, were observed with a linear microphone array at  $y = 0$  to obtain the driving



**Fig. 4.5.** Numerical simulation setup. Linear microphone and loudspeaker arrays were located at  $x$ -axis, with 64 channels in each array. The array elements were equally spaced 6 cm apart.

signals of the proposed method. The Tukey window function was applied as a window whose sides tapered by 10% [21]. The driving signals of the focused-source method was calculated based on Eq. (4.8) and the same tapering window as the proposed method was applied. In the focused-source method, the locations and original signals of the primary sources to be reproduced were given *a priori*. The original sound field created by the primary sources in the source area was reproduced at  $y > 0$ , i.e., the target area, with the linear loudspeaker array at  $y = 0$ . The original and target sound fields were simulated in  $3.6 \times 3.6$  m regions (shaded regions in Fig. 4.5) at every 1.5 cm. The reconstruction line was shifted 2 m forward, i.e.,  $d_y = 2$  m in Eq. (4.10), so that the target sound field corresponded with the original sound field in the reproduced region (bounded region in Fig. 4.5). The amplitude of the reproduced signals was normalized at the center of the simulated region. The sampling frequency was 48 kHz.

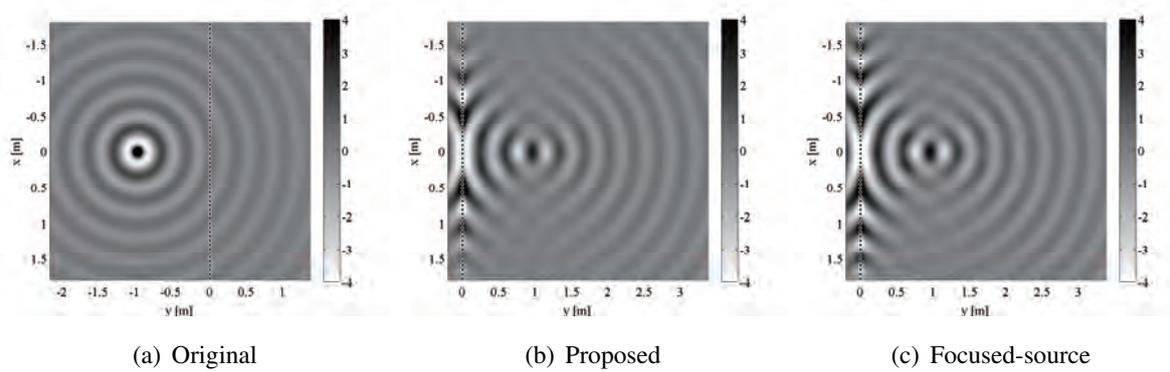
The reproduced sound fields were evaluated by using the time averaged squared error at every simulated discrete position  $(x_i, y_j)$  defined as

$$\text{Err}(x_i, y_j) = 10 \log_{10} \frac{\sum_k |p_{\text{rep}}(x_i, y_j, t_k) - p_{\text{org}}(x_i, y_j, t_k)|^2}{\sum_k |p_{\text{org}}(x_i, y_j, t_k)|^2}, \quad (4.12)$$

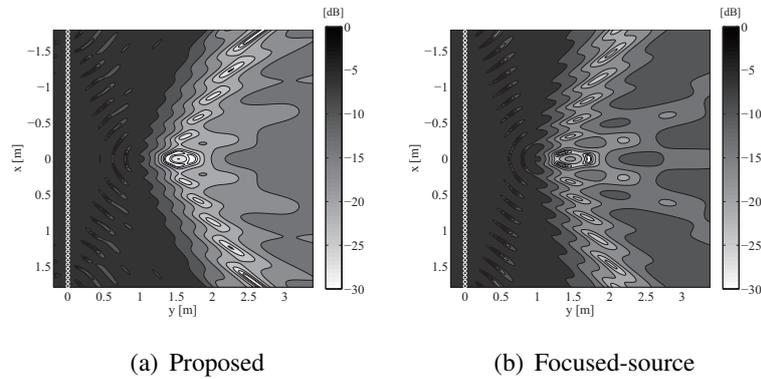
where  $p_{\text{rep}}(x_i, y_j, t_k)$  and  $p_{\text{org}}(x_i, y_j, t_k)$  are the reproduced and original sound pressure distributions in the time domain, respectively, and  $t_k$  denotes the discrete time. The total number of time samples were set at 480 sample, i.e., 10 ms. As the ratio of the original sound pressure distribution to the error of reproduced sound pressure distribution, the signal-to-distortion ratio (SDR) is defined as:

$$\text{SDR} = 10 \log_{10} \frac{\sum_i \sum_j \sum_k |p_{\text{org}}(x_i, y_j, t_k)|^2}{\sum_i \sum_j \sum_k |p_{\text{rep}}(x_i, y_j, t_k) - p_{\text{org}}(x_i, y_j, t_k)|^2}. \quad (4.13)$$

The SDR was calculated in the reproduced region in which the sound field can be reproduced in principle as shown in Fig. 4.5. The SDR was not calculated within 0.1 m of the primary sources



**Fig. 4.6.** Simulation results of original and reproduced sound pressure distribution by using proposed and focused-source methods when source signal was 1-kHz sinusoidal wave.

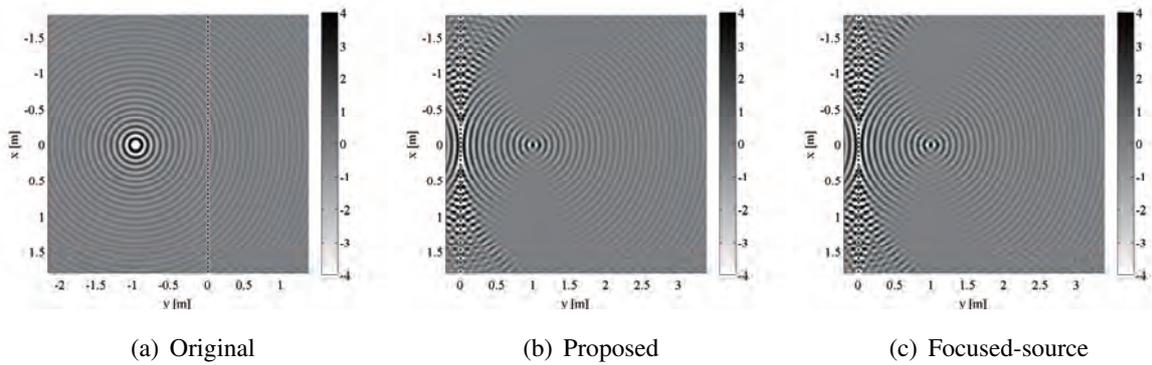


**Fig. 4.7.** Time averaged squared error of sound pressure distribution when source signal was 1-kHz sinusoidal wave.

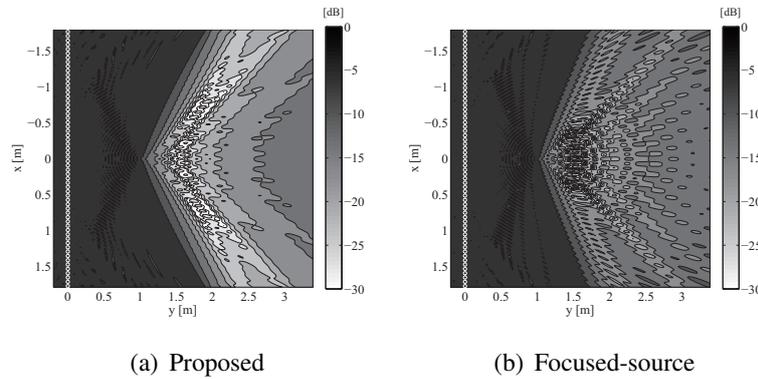
because the amplitude of the original sound pressure is quite high in this region and the SDR greatly decreases. The SDR is used for evaluating accuracy of spatial reconstruction of sound field. However, amplitude errors exist in principle when linear arrays are used. Therefore, the SDR was used for investigating the differences between the proposed and focused-source methods.

Figure 4.6 shows the simulation results when the point source was located at  $(0.0, -1.0)$  m; therefore, the virtual point source was located at  $(0.0, 1.0)$  m. The source signal was a 1-kHz sinusoidal wave. Figure 4.6 (a) shows the original sound pressure distribution in the source area. Figures 4.6 (b) and (c) show the reproduced sound pressure distribution by using the proposed and focused-source methods, respectively. While the accurately reproduced region was restricted, the point source was virtually recreated in the target area by using the two methods. Figure 4.7 shows the time averaged squared error of sound pressure distribution. The SDRs of the proposed and focused-source methods were 7.9 and 7.8 dB, respectively. The error distributions were slightly different in the two methods, but the reproducing accuracies were generally the same.

Figures 4.8 and 4.9 show the simulation results when the source signal was a 4-kHz sinusoidal



**Fig. 4.8.** Simulation results of original and reproduced sound pressure distribution by using proposed and focused-source methods when source signal was 4-kHz sinusoidal wave.



**Fig. 4.9.** Time averaged squared error of sound pressure distribution when source signal was 4-kHz sinusoidal wave.

wave. The SDRs of the proposed and focused-source methods were 8.4 and 9.0 dB, respectively. Similar to the focused-source method [62], the closer the region of the virtual point source, the less effect from spatial aliasing in the proposed method. Figure 4.10 shows the relation between SDRs and frequency of the source signal. The SDRs were almost the same in both methods up to 5.0 kHz, but the SDR of the proposed method decreased faster than that of the focused-source method above 5.0 kHz. However, when the microphone array consisted of 128 channels and the interval of the elements were 3 cm while the loudspeaker array consisted of 64 channels, the SDRs of both methods were almost the same above 5.0 kHz. Therefore, the near-field of the virtual primary sources was hardly affected by spatial aliasing in the proposed method, but the total reproduction accuracy was affected by the alignment of the microphone array. The accurate reproduction in the simulated area was still limited below 5.6 kHz by spatial aliasing even when the number of microphones was increased. A smaller interval length of loudspeakers is needed to reconstruct sound fields higher frequencies to avoid spatial aliasing.

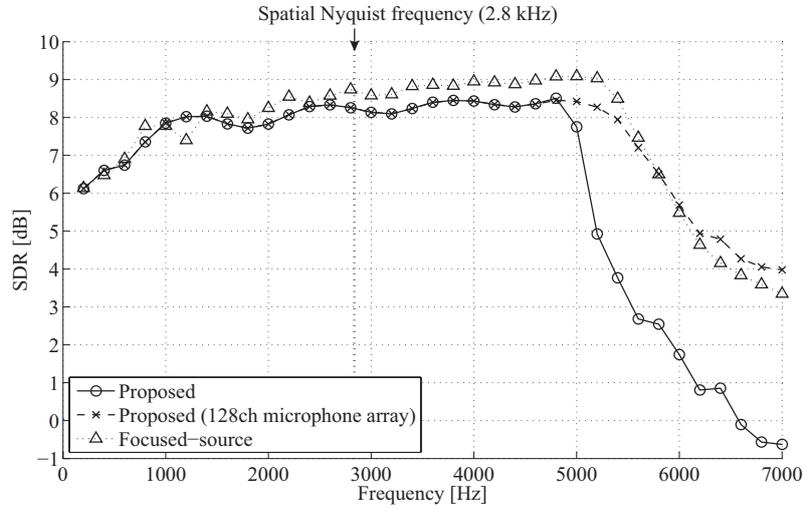


Fig. 4.10. Relation between SDR and frequency of source signal.

## 4.4 Conclusions

A SP-DS conversion method for recreating virtual sound sources in front of a loudspeaker array was proposed. The conventional focused-source method is suitable when decomposed parameters of the sound field to be reproduced is known. However, it cannot be applied when the planar or linear sound pressure distribution is only known. By combining the sound field reproduction method using the planar or linear secondary source distribution and inverse wave propagator for planar or linear sound pressure distribution in the proposed method, we derived a filter representation to shift the sound field to be reproduced towards the listeners while secondary sources are fixed. The filter is determined only by setting a displacement parameter and can be applied as SP-DS conversion. Numerical simulations were conducted to compare the proposed and focused-source methods. While the two methods have different parameters for calculating driving signals of loudspeakers, their reproduction accuracies were almost the same.

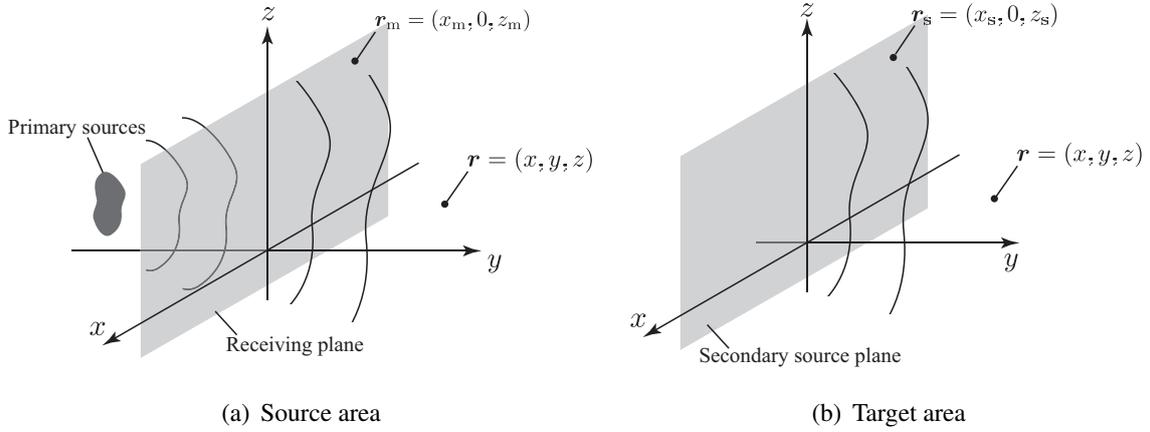
## Chapter 5

# Wave Field Reconstruction Filter for Planar or Linear Arrays of Microphones and Loudspeakers

We propose an SP-DS conversion method for planar or linear arrays of microphones and loudspeakers. For a simple procedure of designing and applying a transform filter, we introduce concepts of continuous distributions of receivers and secondary sources, i.e., microphones and loudspeakers, and physical properties of the wave propagation. These concepts enable to derive a continuous transform equation that relates sound pressure distribution and driving signals for reproducing sound fields in the spatio-temporal frequency domain. We analytically derive this transform equation by simultaneously solving the sound field synthesized by secondary sources and the desired sound field defined as the Rayleigh integral of the first kind, which is defined as the *wave field reconstruction equation* (WFR equation). By discretization, the transform filter that converts the received signals of the microphone array into driving signals of the loudspeaker array is uniquely determined. We define this transform filter as a *wave field reconstruction filter* (WFR filter). Therefore, signals obtained using only a planar or linear omni-directional microphone array are all that is required to calculate driving signals of a planar or linear loudspeaker array. Since the WFR filter is represented as a spatial convolution in addition to its analytical derivation, it has many advantages in terms of filter design, filter size, computational cost, and filter stability compared to the conventional LS method.

Additionally, it is possible to arbitrarily control the corresponding position of the received and reconstructed sound fields. Parametric control of the WFR filter allows to shift and rotate the reconstructed sound field.

Although the method described above assumes that each secondary source has monopole characteristics, it is possible to extend to the case that each secondary source has directivity. The



**Fig. 5.1.** Sound pressure distribution at receiving plane is obtained in source area. In target area, sound field is reproduced by using planar distribution of secondary sources.

difference of the directivity of the secondary sources leads to smaller reproduced region. We consider the model of the directivity of the secondary sources in two approaches and derive the WFR filter considering directivity.

Finally, we consider a problem specific to linear alignments of the array elements. When the linear omni-directional microphone array is assumed, reverberation of the reproduced sound field increases and direct-to-reverberant energy ratio (DRR) decreases. Circular harmonics beamforming technique is applied as a preprocessing of the WFR filter in order to match the DRR to the original sound field.

## 5.1 WFR equation for planar distributions of receivers and secondary sources

We derive the WFR equation that relates sound pressure distribution obtained by receivers to driving signals of secondary sources in the Cartesian coordinates. As shown in Fig. 5.1, continuous and infinite planar distributions of receivers and secondary sources are arranged on the  $x$ - $z$ -plane at  $y = 0$  in the source and target areas, respectively. These planar distributions are assumed to be acoustically transparent. The sound field in the half-space of  $y > 0$  must be reproduced to coincide with the sound field created in the half-space of  $y < 0$ . The position vectors on the receiving and secondary source plane are denoted as  $\mathbf{r}_m = (x_m, 0, z_m)$  and  $\mathbf{r}_s = (x_s, 0, z_s)$ , respectively. The sound pressure at  $\mathbf{r}_m$  and driving signal of the secondary source at  $\mathbf{r}_s$  of the temporal frequency  $\omega$  are denoted as  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  and  $D(\mathbf{r}_s, \omega)$ , respectively.

### 5.1.1 Synthesized sound field

The synthesized sound field created by secondary sources in the target area is denoted as  $P_{\text{syn}}(\mathbf{r}, \omega)$ . When the transfer function from  $\mathbf{r}_s$  to  $\mathbf{r}$  is denoted as  $G(\mathbf{r} - \mathbf{r}_s, \omega)$ ,  $P_{\text{syn}}(\mathbf{r}, \omega)$  is described as

$$P_{\text{syn}}(\mathbf{r}, \omega) = \iint_{-\infty}^{\infty} D(\mathbf{r}_s, \omega) G(\mathbf{r} - \mathbf{r}_s, \omega) dx_s dz_s. \quad (5.1)$$

Equation (5.1) can be regarded as the convolution of  $D(\cdot)$  and  $G(\cdot)$  with respect to  $x$  and  $z$  [34]. Therefore, the spatial Fourier transform of Eq. (5.1) with respect to  $x$  and  $z$  is derived as

$$\tilde{P}_{\text{syn}}(k_x, y, k_z, \omega) = \tilde{D}(k_x, k_z, \omega) \cdot \tilde{G}(k_x, y, k_z, \omega), \quad (5.2)$$

where  $k_x$  and  $k_z$  are the spatial frequency. The spatial Fourier transform is defined as Eq. (A.9) in Appendix. The variables in the spectral domain, i.e., spatio-temporal frequency domain or angular spectrum domain, is represented by a tilde.

### 5.1.2 Derivation of WFR equation for planar receivers

We introduce the Rayleigh integral of the first kind in three dimensions, Eq. (2.28), as representation of the desired sound field. This equation represents the sound field in a half space of a planar boundary based on the planar distribution of the sound pressure gradient on the boundary as follows [21]:

$$P_{\text{des}}(\mathbf{r}, \omega) = -2 \iint_{-\infty}^{\infty} \frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m} G_{3D}(\mathbf{r} - \mathbf{r}_m, \omega) dx_m dz_m, \quad (5.3)$$

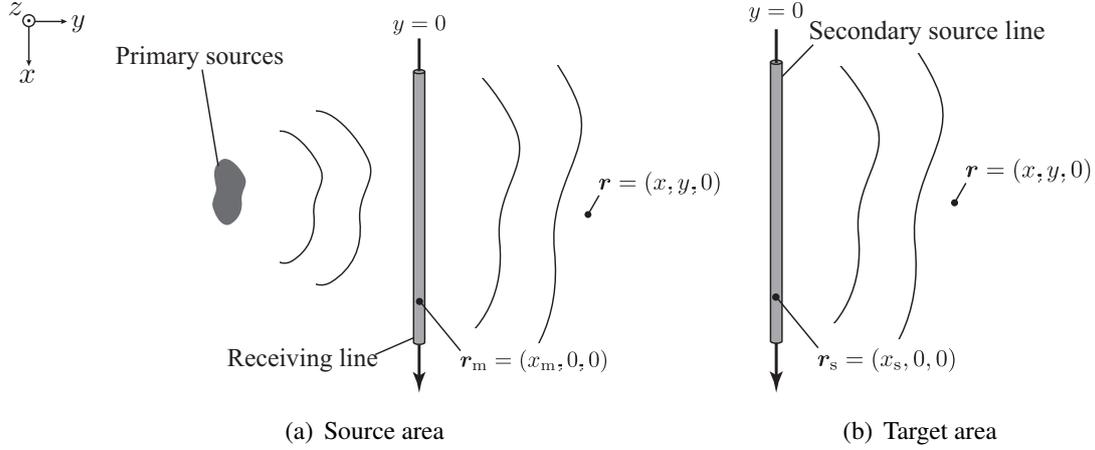
Here,  $P_{\text{des}}(\cdot)$  is the desired sound pressure, and  $G_{3D}(\cdot)$  is the free-field Green function in three dimensions defined as Eq. (3.2).

Now, we consider the spatial Fourier transform of Eq. (5.3) with respect to  $x$  and  $z$ . When it is assumed that the primary sources are only in the region of  $y < 0$ , the distribution of the sound pressure gradient can be estimated from the spatial frequency spectrum on the receiving plane as:

$$\begin{aligned} \frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m} &= \frac{\partial}{\partial y} \left[ \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) e^{j(k_x x + k_y y + k_z z)} dk_x dk_z \right]_{y=0} \\ &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} j k_y \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) e^{j(k_x x + k_z z)} dk_x dk_z \\ &= \mathcal{F}_x^{-1} \mathcal{F}_z^{-1} \left[ j k_y \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) \right], \end{aligned} \quad (5.4)$$

where

$$k_y = \sqrt{k^2 - k_x^2 - k_z^2}. \quad (5.5)$$



**Fig. 5.2.** Sound pressure distribution at receiving line is obtained in source area. In target area, sound field is reproduced by using linear distribution of secondary sources.

Therefore, the spatial Fourier transform of Eq. (5.3) with respect to  $x$  and  $z$  is represented as

$$\tilde{P}_{\text{des}}(k_x, y, k_z, \omega) = -2jk_y \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) \cdot \tilde{G}_{3\text{D}}(k_x, y, k_z, \omega) \quad (5.6)$$

This equation represents the desired sound field in the spatio-temporal frequency domain.

The synthesized and desired sound fields are represented as Eqs. (5.2) and (5.6), respectively, in the spatio-temporal frequency domain. When these equations are simultaneously solved, i.e.,  $\tilde{P}_{\text{syn}}(\cdot) = \tilde{P}_{\text{des}}(\cdot)$ , the WFR equation that relates the spatial frequency spectrum of the received signals with that of the driving signals is derived.

$$\tilde{D}(k_x, k_z, \omega) = -2jk_y \frac{\tilde{G}_{3\text{D}}(k_x, y, k_z, \omega)}{\tilde{G}(k_x, y, k_z, \omega)} \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) \quad (5.7)$$

When each secondary source is assumed to be monopole for simplicity, i.e.,  $G(\cdot) \simeq G_{3\text{D}}(\cdot)$ , Eq. (5.7) can be approximated as

$$\tilde{D}(k_x, k_z, \omega) = -2jk_y \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega). \quad (5.8)$$

Only the sound pressure distribution on the receiving plane is required to calculate Eq. (5.8).

## 5.2 WFR equation for linear distributions of receivers and secondary sources

Because the sound field in the horizontal ear-plane is the most important from a perception viewpoint [3], we discuss reproduction of only a sound field at a constant height. For reproduction on the  $x$ - $y$ -plane at  $z = 0$  only, the linear receiver and secondary source distributions at the  $x$ -axis

are approximately applicable (Fig. 5.2). The WFR equation for the linear distributions are derived in a similar way as the planar case.

### 5.2.1 Synthesized sound field

Because the secondary sources are assumed to be continuously distributed along the  $x$ -axis in the target area, the synthesized sound field is described as:

$$P_{\text{syn}}(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} D(\mathbf{r}_s, \omega) G(\mathbf{r} - \mathbf{r}_s, \omega) dx_s. \quad (5.9)$$

Here, only the  $x$ - $y$ -plane at  $z = 0$  is focused on. The spatial Fourier transform with respect to  $x$  of Eq. (5.9) is represented as

$$\tilde{P}_{\text{syn}}(k_x, y, 0, \omega) = \tilde{D}(k_x, \omega) \cdot \tilde{G}(k_x, y, 0, \omega). \quad (5.10)$$

As in the planar case,  $G(\cdot)$  in Eq. (5.9) is the transfer function between  $\mathbf{r}$  and  $\mathbf{r}_s$  in three dimensions because dimensionality does not depend on the array configuration. Since only the  $x$ - $y$ -plane at  $z = 0$  is focused on, the  $z$ -coordinate in Eq. (5.10) is set as 0.

### 5.2.2 Derivation of WFR equation for linear receivers

Now, we assume that the desired sound field follows the Rayleigh integral of the first kind in two dimensions. This equation represents the two-dimensional sound field in the half-plane of a linear boundary based on the linear distribution of sound pressure gradient on the boundary as follows:

$$P_{\text{des}}(\mathbf{r}, \omega) = -2 \int_{-\infty}^{\infty} \frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m} G_{2D}(\mathbf{r} - \mathbf{r}_m, \omega) dx_m. \quad (5.11)$$

Here,  $G_{2D}(\cdot)$  is the free-field Green function in two dimensions defined as Eq. (3.10). Although the sound fields in the source and target areas are in the three dimensions, the desired sound field is assumed to be two dimensional. This assumption means that captured and reproduced sound fields are invariant with regard to changes in the  $z$ -axis in the source and target areas, respectively. Obviously, this assumption is difficult to meet in practical situations; therefore, it leads to some artifacts as discussed further below.

The spatial Fourier transform of Eq. (5.11) with respect to  $x$  can be derived by using an estimation of the sound pressure gradient in the spatio-temporal frequency domain. This estimation can be derived in a similar way as the planar case:

$$\frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m} = \mathcal{F}_x^{-1} \left[ jk_\rho \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega) \right], \quad (5.12)$$

where

$$k_\rho = \sqrt{k^2 - k_x^2}. \quad (5.13)$$

As mentioned before, it is assumed that the desired sound field is invariant with regard to changes in the  $z$ -axis, i.e., two-dimensions. Therefore, Eq. (5.12) is only valid when the primary sources are on the  $x$ - $y$ -plane at  $z = 0$  only in three-dimensional space. Based on the convolution theorem, the spatial Fourier transform of Eq. (5.11) with respect to  $x$  is represented as

$$\tilde{P}_{\text{des}}(k_x, y, 0, \omega) = -2jk_\rho \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega) \cdot \tilde{G}_{2\text{D}}(k_x, y, \omega). \quad (5.14)$$

This equation represents the desired sound field on the  $x$ - $y$ -plane at  $z = 0$  in the spatio-temporal frequency domain.

When the equations of the synthesized and desired sound fields, Eqs. (5.10) and (5.14), are simultaneously solved, i.e.,  $\tilde{P}_{\text{syn}}(\cdot) = \tilde{P}_{\text{des}}(\cdot)$ , the WFR equation for linear distributions is derived as:

$$\tilde{D}(k_x, \omega) = -2jk_\rho \frac{G_{2\text{D}}(k_x, y, 0, \omega)}{G(k_x, y, 0, \omega)} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega). \quad (5.15)$$

The three-dimensional synthesized and two-dimensional desired sound fields are combined in this derivation [34]. As in the planar case, it is assumed that each secondary source can be approximated as monopole,  $G(\cdot) \simeq G_{3\text{D}}(\cdot)$ , for simplicity as:

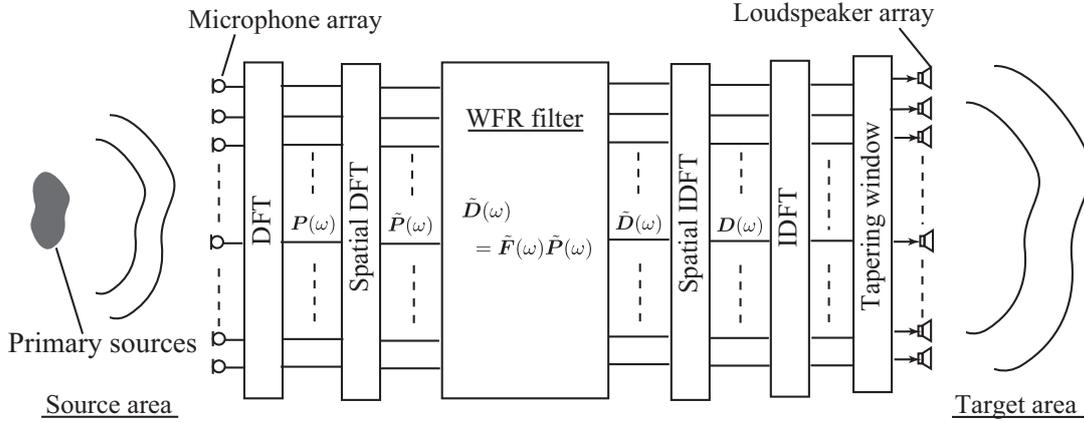
$$\tilde{D}(k_x, \omega) = -4j \frac{e^{jk_\rho y}}{H_0^{(1)}(k_\rho y)} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega). \quad (5.16)$$

The analytical solutions for the spatial Fourier transform of Eqs. (3.2) and (3.10) with respect to  $x$ , Eqs. (A.22) and (A.24) in the Appendix, are used to derive Eq. (5.16).

Equation (5.16) depends on  $y$  in the target area. This means that Eqs. (5.10) and (5.14) can be equivalent only on a line parallel to the  $x$ -axis. Therefore, the reference line  $y = y_{\text{ref}}$  must be set, and it leads to faster amplitude decay than desired. This artifact comes from the mismatch between the two-dimensional assumption of the desired sound field and the three-dimensional synthesized sound field. If  $G(\cdot)$  in Eq. (5.10) has the same characteristic as a line source, i.e., it is invariant with regards to changes in the  $z$ -axis, this artifact may not appear. Because  $G(\cdot)$  is assumed to have monopole characteristic in order to derive Eq. (5.16), the synthesized sound field propagates axisymmetrically with a central axis on the secondary source line.

Similar to the planar case, only the sound pressure distribution along the  $x$ -axis is needed to calculate Eq. (5.16). As mentioned before, the desired sound field is assumed to be two-dimensional in the source area, which only requires the primary sources to be on the  $x$ - $y$ -plane at  $z = 0$ . Therefore, all primary sources along the axisymmetric position with a central axis on the receiving line are projected on a two-dimensional plane.

For computational simplicity, a simplified form of Eq. (5.16) is derived. The Hankel function can be approximated for large arguments as Eq. (2.43). Therefore, Eq. (5.16) can be approximated



**Fig. 5.3.** Block diagram of WFR filtering for linear arrays. Received signals of microphone array are transformed into driving signals of loudspeaker array by using WFR filter.

for  $|k_\rho y_{\text{ref}}| \gg 0$  as

$$\begin{aligned} \tilde{D}(k_x, \omega) &\simeq -4j \frac{e^{-jk_\rho y_{\text{ref}}}}{\sqrt{\frac{2}{\pi k_\rho y_{\text{ref}}} e^{-j(k_\rho y_{\text{ref}} - \frac{\pi}{4})}}} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega) \\ &= -2\sqrt{2\pi j k_\rho y_{\text{ref}}} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega) \end{aligned} \quad (5.17)$$

This equation is simpler than Eq. (5.16) to calculate numerically.

If the planar receiver and linear secondary source distributions are combined, i.e., Fig. 5.1 (a) and 5.2 (b), the desired sound field in the source area is regarded as three-dimensional. Since the distribution of the sound pressure gradient can be estimated based on Eq. (5.4), the primary sources are not required to be only on the  $x$ - $y$ -plane at  $z = 0$ . The WFR equation for the planar receiver and linear secondary source distributions is derived by replacing  $k_\rho$  with  $k_y$  in Eq. (5.16) or (5.17). In the target area, the sound field in the source area on the  $x$ - $y$ -plane at the arbitrarily chosen constant  $z$  is extracted and reproduced.

## 5.3 Implementation of WFR filter as FIR filter

### 5.3.1 WFR filter for linear arrays of microphones and loudspeakers

In a practical implementation, in order to use microphone and loudspeaker arrays, the receiver and secondary source distributions must be discretized and truncated. The transformation of the received signals into the driving signals is also processed as a digital filter, i.e., the WFR filter as a transform filter. A block diagram of the proposed method using linear arrays is depicted in Fig. 5.3 when the WFR filter is designed as a finite impulse response (FIR) filter. The sound pressure distribution in the source area is obtained by using an equally spaced omni-directional microphone array. A loudspeaker array is arranged at the coinciding position with the microphone

array in the target area, and each loudspeaker is assumed to be omni-directional. The numbers of microphones and loudspeakers are the same, and are denoted as  $M$ . The received signals of the microphone array and the driving signals of the loudspeaker array in the temporal frequency domain are respectively denoted as

$$\begin{aligned}\mathbf{P}(\omega) &= [P_1(\omega), \dots, P_M(\omega)]^T \\ \mathbf{D}(\omega) &= [D_1(\omega), \dots, D_M(\omega)]^T.\end{aligned}$$

The spatial discrete Fourier transform (DFT) is achieved with the proper size of zero-padding, and the number of samples for the DFT is denoted as  $N_{\text{FT}}$ . The spatial frequency spectrum of  $\mathbf{P}(\omega)$  and  $\mathbf{D}(\omega)$  is respectively denoted as

$$\begin{aligned}\tilde{\mathbf{P}}(\omega) &= [\tilde{P}_1(\omega), \dots, \tilde{P}_{N_{\text{FT}}}(\omega)]^T \\ \tilde{\mathbf{D}}(\omega) &= [\tilde{D}_1(\omega), \dots, \tilde{D}_{N_{\text{FT}}}(\omega)]^T.\end{aligned}$$

According to Eq. (5.16),  $\tilde{D}_i(\omega)$ , the  $i$ -th element of  $\tilde{\mathbf{D}}(\omega)$ , can be calculated as

$$\tilde{D}_i(\omega) = \tilde{F}_i(\omega) \tilde{P}_i(\omega), \quad (5.18)$$

where

$$\tilde{F}_i(\omega) = -4j \frac{e^{j\sqrt{k^2 - k_{x,i}^2} y_{\text{ref}}}}{H_0^{(1)}\left(\sqrt{k^2 - k_{x,i}^2} y_{\text{ref}}\right)}. \quad (5.19)$$

Here,  $k_{x,i}$  denotes the spatial frequency in the  $i$ -th bin. Therefore, the matrix representation of transformation can be described as

$$\tilde{\mathbf{D}}(\omega) = \tilde{\mathbf{F}}(\omega) \tilde{\mathbf{P}}(\omega), \quad (5.20)$$

where

$$\tilde{\mathbf{F}}(\omega) = \text{diag}\left(\tilde{F}_1(\omega), \dots, \tilde{F}_{N_{\text{FT}}}(\omega)\right). \quad (5.21)$$

Equation (5.21) is the WFR filter in the spatio-temporal frequency domain. The WFR filter in the spatial domain, which is denoted by  $\mathbf{F}(\omega)$ , needs to be trimmed to a length  $K$  that is sufficient for reproduction accuracy because this filter is designed as an FIR filter. Although it may be possible to design it as an infinite impulse response (IIR) filter, FIR filters are easier to design than IIR filters, and furthermore, maintaining the stability of IIR filters is not entirely straightforward. The longer  $K$  is, the more precise the WFR filter becomes. The spatial DFT size,  $N_{\text{FT}}$ , must be longer or equal to  $M + K - 1$ . This is because Eq. (5.20) can be considered as the convolution of the

$M$ -size vector  $\mathbf{P}(\omega)$  and the  $K$ -size vector  $\text{diag}(\mathbf{F}(\omega))$  in the spatial domain. The driving signal  $\mathbf{D}(\omega)$  is sorted from the spatial inverse DFT (IDFT) of  $\tilde{\mathbf{D}}(\omega)$ .

The elements of the WFR filter, Eq. (5.19), are determined only by setting the position of the reference line  $y_{\text{ref}}$ . The Hankel function can be calculated numerically, e.g., as shown in Ref. [70], or the discrete form of Eq. (5.17) can be used as the WFR filter without the Hankel function.

The discretization and truncation of the receiver and secondary source distributions lead to several artifacts. The properties of these artifacts are the same as those of other methods using a planar or linear loudspeaker array [33, 34]. Because of discretization, spatial aliasing errors occur above the spatial Nyquist frequency that is defined as

$$f_{\text{nyq}} = \frac{c}{2\Delta x} \quad (5.22)$$

where  $\Delta x$  denotes the interval of loudspeakers. The truncation of the secondary source distribution causes unnecessary reflections from the edge. Furthermore, the steep truncation of the received signals causes severe error on the edge of the driving signals. Therefore, a tapering window is applied to reduce the truncation errors [30].

### 5.3.2 Property comparison between WFR and LS-based filters

In this context, methods based on the LS algorithm introduced in Section 3.2.3 can be applied in order that sound pressures at discrete control points aligned in front of the loudspeaker array correspond to sound pressures obtained by the microphone array in the source area. Based on Eq. (3.22) and (3.23), the transform filter obtained by the LS and regularized LS methods can be respectively described as:

$$\mathbf{F}^{\text{LS}}(\omega) = (\mathbf{G}^H(\omega)\mathbf{G}(\omega))^{-1}\mathbf{G}^H(\omega) \quad (5.23)$$

and

$$\mathbf{F}^{\text{LSreg}}(\omega) = (\mathbf{G}^H(\omega)\mathbf{G}(\omega) + \beta\mathbf{I})^{-1}\mathbf{G}^H(\omega). \quad (5.24)$$

The number of microphones, control points, and loudspeakers is assumed to be the same, i.e.,  $M = L$ . As is clear from the derivations so far, the concept of the LS and regularized LS methods differs greatly from that of the proposed method. The LS methods are based on controlling sound pressure at discrete points by using an inverse filter. Therefore, the physical properties such as alignments of loudspeakers and characteristics of wave propagation are included in the known transfer function matrix  $\mathbf{G}(\omega)$ , and its inverse is numerically solved in the least square error sense. In contrast, the proposed method is based on the continuous WFR equation that is analytically derived based on the physical equation of wave propagation. These facts give rise to many differences between the properties of these transform filters. We remark on the several differences in the properties. Note that only the spatial aspects of the transform filters are discussed.

*a) Filter design:*

In the LS and regularized LS methods, it is possible to use either a modeled or measured transfer function matrix for  $\mathbf{G}(\omega)$ . However, because the inverse of  $\mathbf{G}(\omega)$  becomes unstable in many cases, heuristic processes for stabilization, such as adjustment of the regularization parameter  $\beta$ , are required. Several systematic methods for determining  $\beta$  have been proposed such as the L-curve [71] and generalized cross validation methods [72].

The proposed WFR filter is analytically determined; therefore, it has an advantage in that no heuristic processes for designing the filter are needed.

*b) Filter size:*

The size of the LS-based filters,  $\mathbf{F}^{\text{LS}}(\omega)$  and  $\mathbf{F}^{\text{LSreg}}(\omega)$ , is  $M^2$ . Because  $M$  is assumed to be a very large number, e.g.,  $M \geq 32$ ,  $\mathbf{F}^{\text{LS}}(\omega)$  and  $\mathbf{F}^{\text{LSreg}}(\omega)$  generally become very large matrices in order to avoid the spatial aliasing artifacts. To make them smaller, it is necessary to reduce the number of either loudspeakers or control points; however, this leads to severe errors.

The size of the proposed WFR filter,  $\mathbf{F}(\omega)$ , is  $K$ , and is flexible because it is designed as an FIR filter, which is a matter of trimming. The longer  $K$  is, the more precise the WFR filter becomes. Making  $K$  smaller reduces the spatial resolution of  $\mathbf{F}(\omega)$ . However, the energy of  $\mathbf{F}(\omega)$  is concentrated near the origin. Therefore, the size  $K \ll M^2$  is usually sufficient for obtaining  $\mathbf{D}(\omega)$ , which leads to computational efficiency. As a result, the size of  $\tilde{\mathbf{F}}(\omega)$ ,  $N_{\text{FT}} \geq M + K - 1$ , can also be a much smaller number than the size of  $\mathbf{F}^{\text{LS}}(\omega)$  and  $\mathbf{F}^{\text{LSreg}}(\omega)$ ,  $M^2$ .

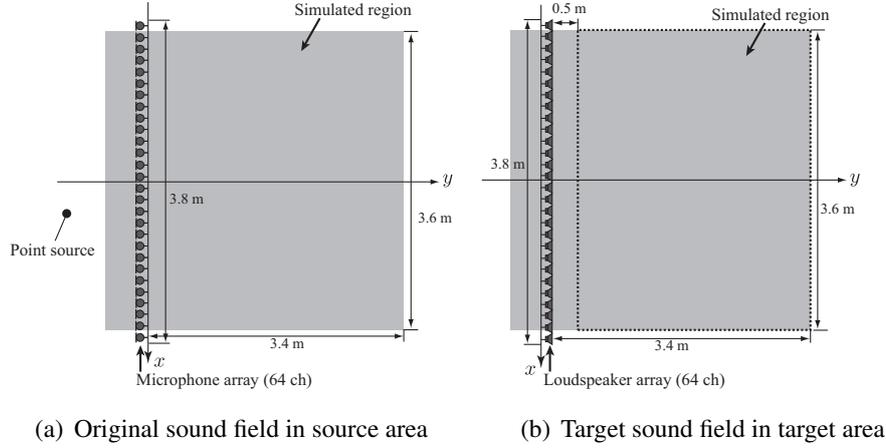
*c) Computational cost:*

The computational cost for calculating the driving signals in the temporal frequency domain by using the LS methods is represented as  $O(M^2)$  because the multiplication of the  $M^2$ -size matrices is included.

In the proposed method, the computational cost of Eq. (5.20) is  $O(N_{\text{FT}})$  because that indicates the multiplication of each element of  $N_{\text{FT}}$ -size vectors in the spatio-temporal frequency domain. The computational cost of the spatial DFT is  $O(N_{\text{FT}} \log N_{\text{FT}})$  when the fast Fourier transform (FFT) algorithm is applied. As a result, the computational cost of the proposed method in the temporal frequency domain is represented as  $O(N_{\text{FT}} \log N_{\text{FT}})$ . In most cases, the computational cost of the proposed method is much smaller than that of the LS methods.

*d) Filter stability:*

The stability of  $\mathbf{F}^{\text{LS}}(\omega)$  can be evaluated as the condition number of  $\mathbf{G}(\omega)$ , which depends on the linear independence of each transfer function [54]. Therefore, with the constraints that the alignments of the loudspeakers and control points are linear,  $\mathbf{F}^{\text{LS}}(\omega)$  becomes very unstable. Although it is necessary to apply  $\mathbf{F}^{\text{LSreg}}(\omega)$  and to adjust  $\beta$  to stabilize  $\mathbf{F}^{\text{LSreg}}(\omega)$ , it is difficult with such a large matrix. Making  $\beta$  larger increases reproduction



**Fig. 5.4.** Numerical simulation setup. Linear microphone and loudspeaker arrays were located at  $x$ -axis, with 64 channels in each array. The array elements were equally spaced 6 cm apart.

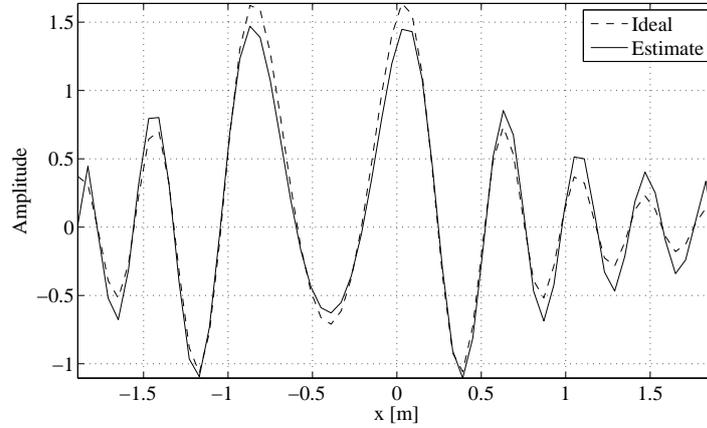
errors.

Because the proposed WFR filter is a uniquely determined vector defined in the orthogonal space, it is much more stable than that of the LS methods. Additionally, the more stable the transform filter is, the shorter the filter length that can be achieved in the time domain.

These advantages of the proposed method derive from the fact that the WFR filter is represented as the convolution form in the spatial domain and is analytically determined. These advantages are closely related to the advantages of acoustical holography based on the angular spectrum representation in the context of sound field analysis [21, 24]. The differences described above are investigated and quantified in the following section.

### 5.3.3 Experiments

Numerical simulations of reproducing point sources were performed by using linear arrays of microphones and loudspeakers under the free-field assumption. We compare the proposed WFR filtering method with the LS and regularized LS methods. Figure 5.4 shows the simulation setup. Identical linear microphone and loudspeaker arrays were located at  $y = 0$ , with 64 channels in each array. These positions were assumed for simplicity because the WFR filter is derived under the assumption that the microphones and loudspeakers are set at the coinciding positions. If their positions are not coincident, the original and reproduced sound fields would be distorted or misaligned. For the loudspeaker intervals to be different from the microphone ones, the spatial DFT or IDFT must be calculated accordingly. The directivity of the array elements was omni-directional. The array elements were equally spaced 6 cm apart, so the array lengths were 3.8 m. Therefore, the spatial Nyquist frequency defined in Eq. (5.22) was about 2.8 kHz. The point sources as primary sources located at  $y < 0$ , i.e., the source area, were observed with the microphone array.



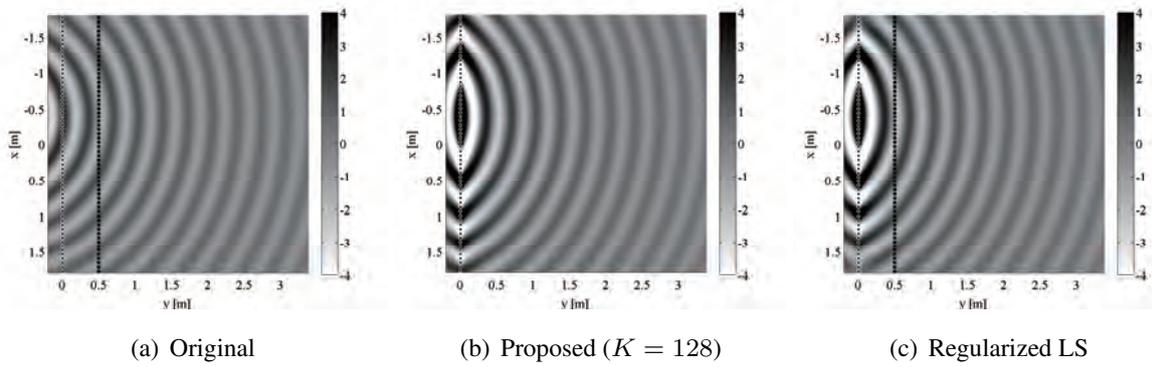
**Fig. 5.5.** Estimated sound pressure gradient by using proposed method when source signal was 1-kHz sinusoidal wave.

The observed signals were transformed into the driving signals of the loudspeaker array by using each transform filter. In the WFR filtering method, Eq. (5.21) was used as the transform filter and  $y_{\text{ref}}$  was set at 1.0 m. The Tukey window function was applied as a tapering window whose sides tapered by 10%. In the LS and regularized LS methods, Eqs. (5.23) and (5.24) were used as the transform filters, respectively. Since the control line  $y_{\text{cl}}$  must be set in front of the loudspeaker array, the location of the microphone array was shifted to  $y = y_{\text{cl}}$  to match the target sound field with that of the WFR filtering method. The control line  $y_{\text{cl}}$  was set at 0.5 m. The regularization parameter,  $\beta$  in Eq. (5.24), was determined by using the L-curve method [71] at each temporal frequency; therefore,  $\beta$  was different at each temporal frequency. The original and target sound fields were simulated in  $3.6 \times 3.6$  m regions (shaded regions in Fig. 5.4) at every 1.5 cm. The reproduced sound pressure distributions in the target area were normalized at the center of the simulated region. The sampling frequency was 48 kHz.

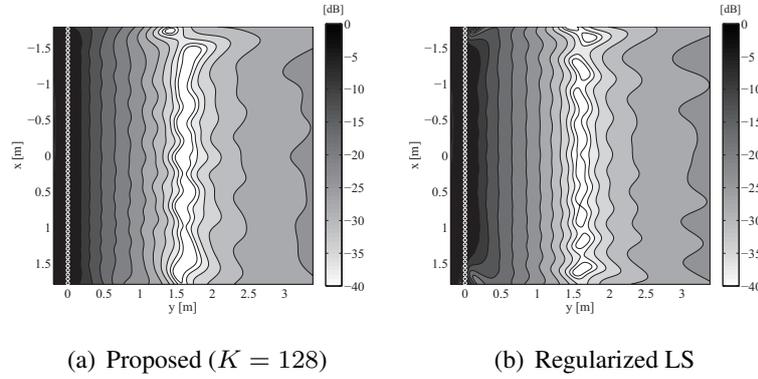
The reproduced sound fields were evaluated by using the time averaged squared error at every simulated discrete position  $(x_i, y_j)$  defined as Eq. (4.12). As the ratio of the original sound pressure distribution to the error of reproduced sound pressure distribution, the signal-to-distortion ratio (SDR) is defined as Eq. (4.13). The SDR was calculated in the region bounded by the dashed line in Fig. 5.4 (b).

First, we demonstrate simulation result of estimated sound pressure gradient based on Eq. (5.12). Figure 5.5 shows the ideal and estimated distributions of sound pressure gradient on the  $x$ -axis when the point source was located at  $(-0.4, -1.0, 0.0)$  m and the source signal was a 1-kHz sinusoidal wave. The filter size was set at  $K = 128$ . In this simulation, the tapering window was not applied. As expected, the estimate accuracy at the neighboring region of the origin was higher than that at the edge of the microphone array due to finite array length.

Figure 5.6 shows the simulation results when the point source was located at  $(-0.4, -1.0, 0.0)$  m.



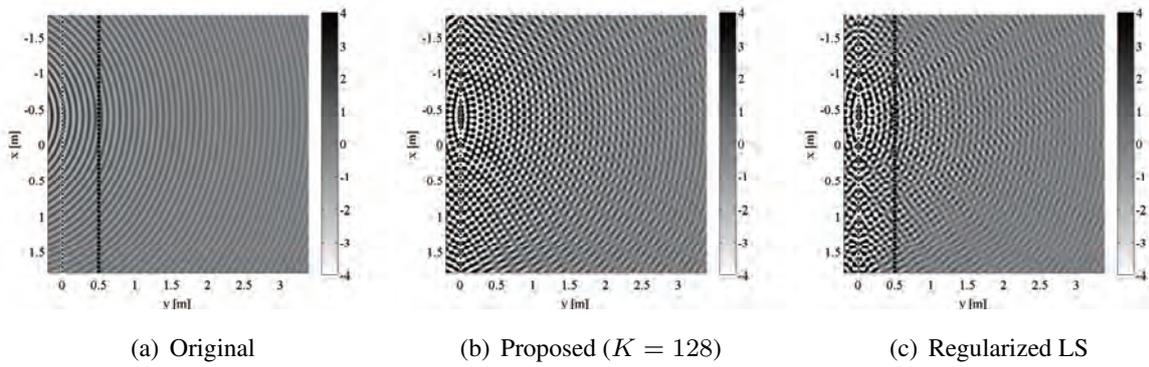
**Fig. 5.6.** Simulation results of original and reproduced sound pressure distribution by using proposed and regularized LS methods when source signal was 1-kHz sinusoidal wave.



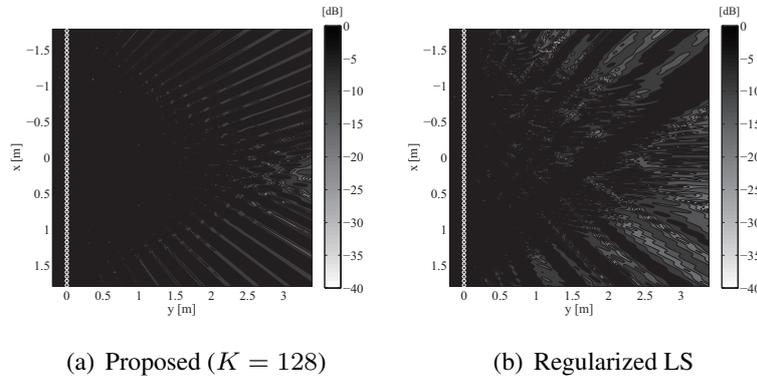
**Fig. 5.7.** Time averaged squared error of sound pressure distribution when source signal was 1-kHz sinusoidal wave.

The source signal was a 1-kHz sinusoidal wave. Figure 5.6 (a) shows the original sound pressure distribution in the source area. Figure 5.6 (b) and (c) show the reproduced sound pressure distribution using the proposed and regularized LS methods, respectively. The filter size of the proposed method was set at  $K = 128$ . Figure 5.7 shows the time-averaged squared error of sound pressure distribution. In both methods, the reproduction accuracies were distinctly high along the line parallel to the loudspeaker array. Those in the region off the line are low because of the faster amplitude decay. The SDRs of both the proposed and regularized LS methods were 19.8 dB. Almost the same performance as the regularized LS method was achieved with the proposed method.

The results when the source signal was a 4-kHz sinusoidal wave are shown in Figs. 5.8 and 5.9. Because the frequency of the source signal was above the spatial Nyquist frequency, severe errors can be seen in both methods. In the proposed method, it appears that the errors derive from the replication of the spatial frequency spectra. On the contrary, in the regularized LS method, the synthesized sound pressure distribution is more complicated. The SDRs of the proposed and regularized LS methods were 6.6 and 5.7 dB, respectively.



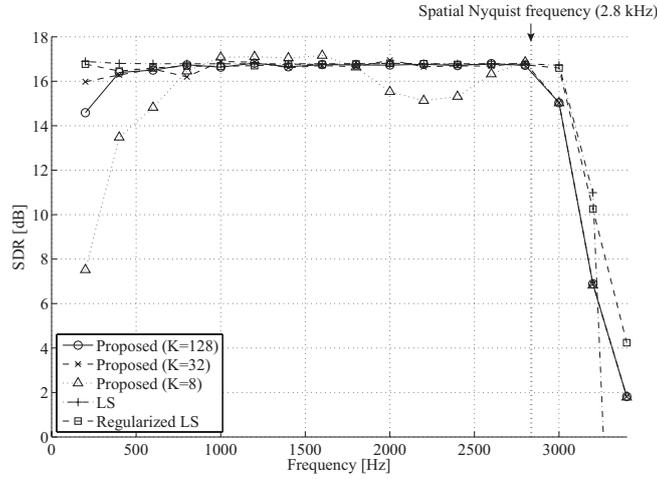
**Fig. 5.8.** Simulation results of original and reproduced sound pressure distribution by using proposed and regularized LS methods when source signal was 4-kHz sinusoidal wave.



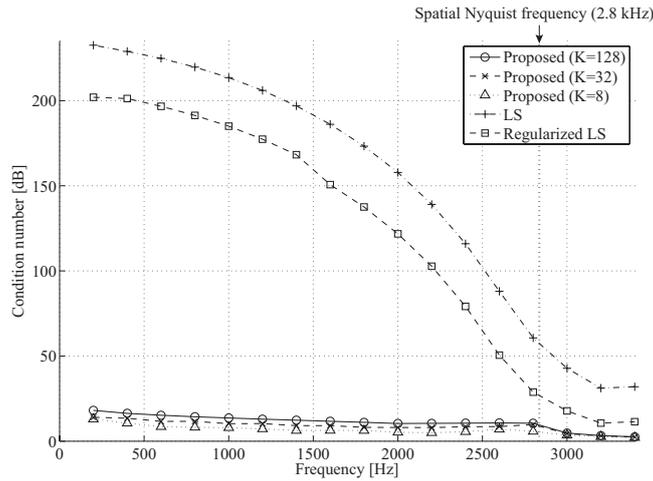
**Fig. 5.9.** Time averaged squared error of sound pressure distribution when source signal was 4-kHz sinusoidal wave.

Figure 5.10 plots the relation between SDRs and the frequency of the source signal. The position of the point source was  $(-0.4, -1.0, 0.0)$  m. The results of the proposed method are shown for three filter sizes,  $K = 128, 32,$  and  $8$ . The SDRs of the proposed method using the shorter filter size ( $K = 8$ ) were lower than those of the LS and regularized LS methods, especially at low frequencies. By using the proposed method with  $K = 128$ , the SDRs were almost the same as the LS methods above 400 Hz. The longer the  $K$ , the higher the reproduction accuracy becomes at low frequency. The SDRs of all the methods were very low above the spatial Nyquist frequency. The reproduction accuracies of the proposed method were almost the same as those of the LS methods at most frequencies.

The relation between the condition number in dB and the frequency of the source signal is shown in Fig. 5.11. Although the condition number of the proposed method was not originally defined, we calculated it as  $20 \log_{10}(\text{cond}(\tilde{\mathbf{F}}(\omega)))$ , where  $\text{cond}(\cdot)$  indicates the operation of the ratio of the largest and smallest eigenvalue. The condition numbers of the LS and regularized LS methods were calculated as  $20 \log_{10}(\text{cond}(\mathbf{G}(\omega)))$  and  $20 \log_{10}(\text{cond}((\mathbf{G}^H(\omega)\mathbf{G}(\omega) +$



**Fig. 5.10.** Relation between SDR and frequency of source signal.



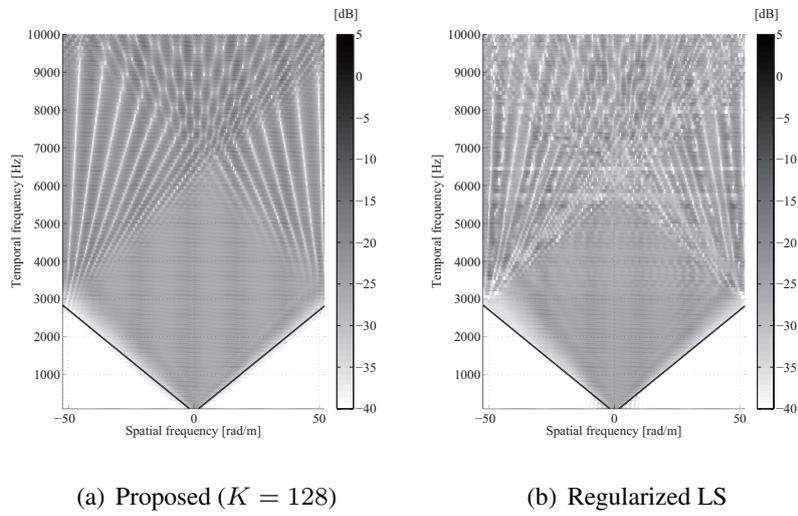
**Fig. 5.11.** Relation between condition number in dB and frequency of source signal.

$\beta \mathbf{I})^{-1} \mathbf{G}^H(\omega)$ ), respectively. A smaller condition number means that the filter is more stable. Note that these values do not depend on input signals. The condition numbers of the LS method were distinctly higher than those of the proposed method. Although the condition numbers of the regularized LS method were lower than those of the LS method, they were still higher than those of the proposed method. Almost the same performance as the LS and regularized LS methods was achieved by using the proposed method, whereas the filter stability of the proposed method was higher than that of these LS methods. These results are summarized in Table 5.1. Obviously, the filter size and computational cost of the proposed method were much smaller than those of the LS methods.

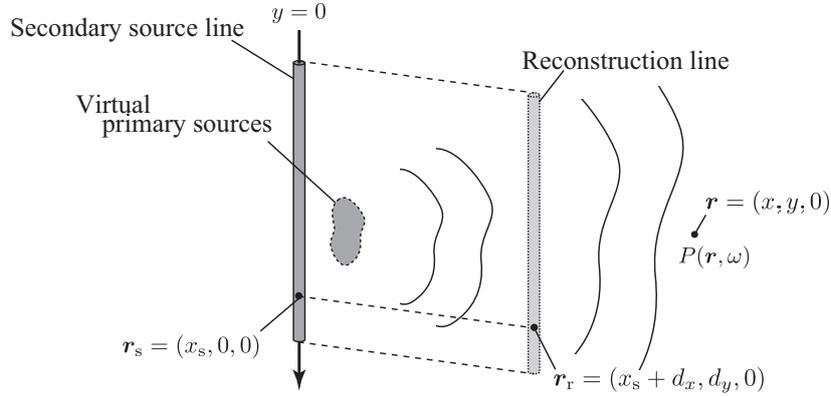
Figures 5.12 (a) and (b) show the amplitude of the driving signals of the proposed and regularized LS methods in the spatio-temporal frequency domain, respectively. The horizontal and ver-

**Table 5.1.** Summarized results of experiments.

|                | Filter size  | Computational cost  |                       | Filter stability<br>(Maximum value of<br>condition number) | Averaged SDR<br>below 2.8 kHz |
|----------------|--------------|---|-----------------------|--|-------------------------------|
| Proposed       | $K = 128$    | $O(N_{\text{FT}} \log N_{\text{FT}})$<br>$(N_{\text{FT}} \geq M + K - 1)$ | $N_{\text{FT}} = 256$ | 18.1 dB  | 16.5 dB                       |
|                | $K = 32$     |   | $N_{\text{FT}} = 128$ | 14.0 dB  | 16.6 dB                       |
|                | $K = 8$      |   | $N_{\text{FT}} = 128$ | 13.0 dB  | 15.5 dB                       |
| LS             | $M^2 = 4096$ | $O(M^2)$  | $M = 64$              | 233 dB   | 16.8 dB                       |
| Regularized LS |              |   |                       | 202 dB   | 16.7 dB                       |

**Fig. 5.12.** Amplitude of driving signals in spatio-temporal frequency domain when primary source is point source.

tical axes indicate the spatial frequency  $k_x$  rad/m and temporal frequency  $\omega/2\pi$  Hz, respectively. The amplitudes were calculated as  $20 \log_{10} |\tilde{\mathbf{D}}(\omega)|$  with spatial DFT. The size of the spatial DFT was 256. These amplitudes were normalized at 2 kHz. The bold line indicates the line of  $k_x = k$ ; therefore, the region below this line represents the evanescent wave components. Although it seems that the amplitude of the regularized LS method is slightly concentrated compared to that of the proposed method in the region below the spatial Nyquist frequency, this is because the loudspeaker array position of the regularized LS method was closer to the virtual primary source than that of the proposed method. In the region above the spatial Nyquist frequency, the aliasing noise of the proposed method seems to be a simple replication of the spatial frequency spectra. On the contrary, that of the regularized LS method was very different from the simple replication of the spatial frequency spectra. This difference can be considered as an effect of the off-diagonal components of the transform filter  $\mathbf{F}^{\text{LSreg}}(\omega)$ .



**Fig. 5.13.** Reproduction of arbitrary-shifted sound field. Reconstruction line is shifted from secondary source line in target area.

## 5.4 Sound field shifting and rotating by parametric control of WFR filter

By using the WFR filter derived in Sections 5.1 and 5.2, the reconstructed sound field corresponds to the received sound field at the position of the loudspeaker array. In future application of sound field reproduction, it would be useful to be able to control the corresponding position of these sound fields, for instance, to adjust them to the visual images or add spatial effects to them. Parametric control of the WFR filter allows to shift and rotate the reconstructed sound field. Therefore, it is possible to arbitrarily control the corresponding position of the received and reconstructed sound fields. The shift of the reconstructed sound field can be regarded as an extension of the method presented in Chapter 4.

The use of linear distributions of receivers and secondary sources is only considered because the extension to the planar case is straightforward.

### 5.4.1 WFR equation for reproducing arbitrary-shifted sound field

By using the driving signals calculated by Eq. (5.16), the sound field obtained on the receiving line is reconstructed except for the amplitude errors on the secondary source line. To shift the reproduced sound field in the target area, the sound field obtained on the receiving line is reconstructed on an arbitrary shifted line, as shown in Fig. 5.13. This line is defined as a reconstruction line. The position of the reconstruction line is denoted as  $\mathbf{r}_r = (x_s + d_x, d_y, 0)$ , where  $d_x$  and  $d_y$  are the shift distances in the directions of  $x$  and  $y$ , respectively. It is necessary to derive the WFR equation for shifting the reconstruction line to  $\mathbf{r}_r$ .

The sound pressure on the secondary source line can be estimated from that on the reconstruct-

tion line by using the phase-shift of the spatio-temporal frequency as:

$$\tilde{P}_{\text{des}}(k_x, 0, 0, \omega) = \tilde{P}_{\text{rec}}(k_x, d_y, 0, \omega)e^{j(k_x d_x + k_\rho d_y)}, \quad (5.25)$$

where  $\tilde{P}_{\text{des}}(k_x, 0, 0, \omega)$  and  $\tilde{P}_{\text{rec}}(k_x, d_y, 0, \omega)$  are the respective spatio-temporal frequency spectra at  $\mathbf{r}_s$  and  $\mathbf{r}_r$ . Estimation in the direction of  $y$  can be explained by the principle of near-field acoustical holography in two-dimensions [21].

Because the spatio-temporal frequency spectrum at the reconstruction line corresponds to that at the receiving line,  $\tilde{P}_{\text{rcv}}(\cdot)$ , the desired sound field can be derived by using Eqs. (5.14) and (5.25) as:

$$\tilde{P}_{\text{des}}(k_x, y, 0, \omega) = -2jk_\rho e^{j(k_x d_x + k_\rho d_y)} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega) \cdot \tilde{G}_{2D}(k_x, y, \omega). \quad (5.26)$$

Therefore, the WFR equation for shifting the reproduced sound field can be derived by solving Eqs. (5.10) and (5.26) as:

$$\tilde{D}(k_x, \omega) = -2jk_\rho e^{j(k_x d_x + k_\rho d_y)} \frac{G_{2D}(k_x, y, 0, \omega)}{G(k_x, y, 0, \omega)} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega) \quad (5.27)$$

$$= -4j \frac{e^{jk_\rho y}}{H_0^{(1)}(k_\rho y)} e^{j(k_x d_x + k_\rho d_y)} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega). \quad (5.28)$$

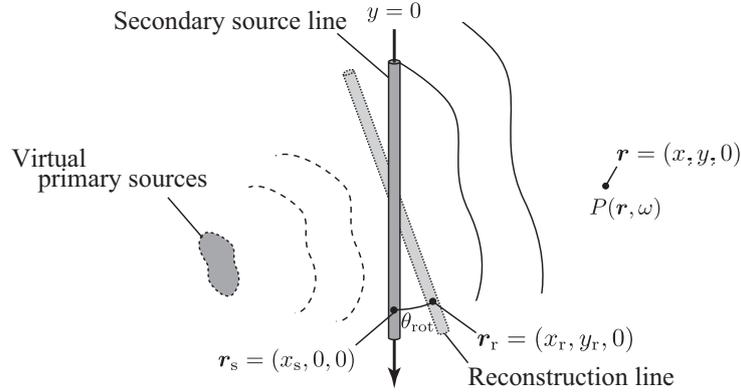
This equation means the phase-shift of the WFR filter corresponds to shifting the position of the reproduced sound field in the target area. Because evanescent components causes instability of the WFR filter as discussed in Chapter 4, it is necessary to apply a tapering window that reduces components of  $|k_x| > k$  to Eq. (5.28).

As a general acoustic-holographic objective, such as imaging of sound pressure distribution, Eq. (5.25) is only valid when the distance between the receiving line and primary sources is smaller than  $d_y$ . However, in this method, when  $d_y$  is larger than the distance between the receiving line and primary sources, the primary sources are virtually reconstructed in front of the secondary source line as in Chapter 4. The region between the virtual primary sources and the secondary source line is not ensured because the secondary source emit a sound wave that travels towards the reconstruction line.

#### 5.4.2 WFR equation for reproducing arbitrary-rotated sound field

To rotate the reproduced sound field in the target area, the captured sound field on the receiving line is reconstructed on an arbitrary-rotated line around the  $z$ -axis, as shown in Fig. 5.14. Again, this line is defined as a reconstruction line, and the position of the reconstruction line is denoted as  $\mathbf{r}_r = (x_r, y_r, 0)$ . When rotation angle is denoted as  $\theta_{\text{rot}}$ ,  $x_r$  and  $y_r$  are represented as:

$$\begin{aligned} x_r &= x \cos \theta_{\text{rot}} \\ y_r &= x \sin \theta_{\text{rot}}. \end{aligned}$$



**Fig. 5.14.** Reproduction of arbitrary-rotated sound field. Reconstruction line is rotated from secondary source line in target area.

It is necessary to derive the WFR equation for rotating the reconstruction line to  $\mathbf{r}_r$ .

When  $P_{\text{rec}}(\mathbf{r}_{\text{rec}}, \omega)$  denotes the sound pressure on the reconstruction line,  $P_{\text{rec}}(\mathbf{r}_{\text{rec}}, \omega)$  is obtained by modulating the sound pressure on the  $x$ -axis of the desired sound field as:

$$P_{\text{rec}}(x_r, y_r, 0, \omega) = P_{\text{des}}(x, 0, 0, \omega) \cdot e^{-jk_{x,\text{rot}}x}. \quad (5.29)$$

Here, the modulating frequency  $k_{x,\text{rot}}$  is determined so that the spatial frequency spectrum on the  $x$ -axis,  $k_x$ , shifts to the rotated one as:

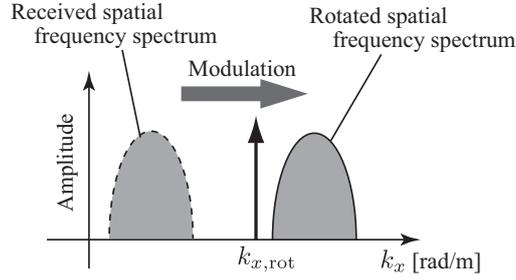
$$k_{x,\text{rot}} = k \sin \theta_{\text{rot}}. \quad (5.30)$$

By using Eq. (5.29), the spatial frequency spectrum on the  $x$ -axis of the desired sound field can be described as:

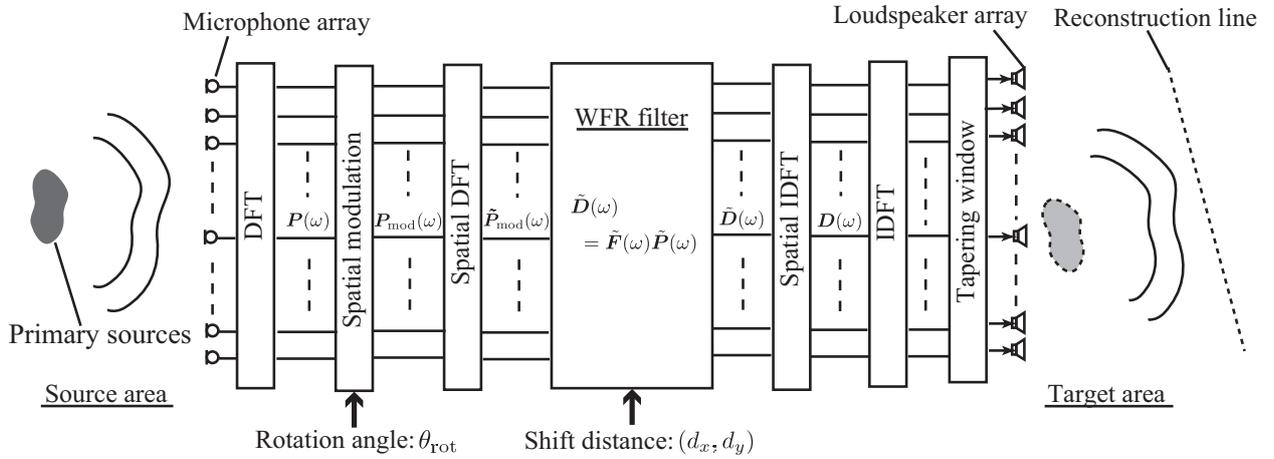
$$\begin{aligned} \tilde{P}_{\text{des}}(k_x, 0, 0, \omega) &= \int_{-\infty}^{\infty} P_{\text{des}}(x, 0, 0, \omega) e^{-jk_x x} dx \\ &= \int_{-\infty}^{\infty} P_{\text{rec}}(x_r, y_r, 0, \omega) e^{-j(k_x - k_{x,\text{rot}})x} dx \\ &= \tilde{P}_{\text{rec}}(k_x - k_{x,\text{rot}}, 0, \omega). \end{aligned} \quad (5.31)$$

Because the spatial frequency spectrum on the receiving line,  $\tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega)$ , corresponds to that on the reconstruction line, the desired sound field can be described by using  $\tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega)$  as:

$$\begin{aligned} \tilde{P}_{\text{des}}(k_x, y, 0, \omega) &= -2jk_{\rho} \tilde{P}_{\text{des}}(k_x, 0, 0, \omega) \cdot \tilde{G}_{2\text{D}}(k_x, y, \omega) \\ &= -2jk_{\rho} \tilde{P}_{\text{rcv}}(k_x - k_{x,\text{rot}}, 0, 0, \omega) \cdot \tilde{G}_{2\text{D}}(k_x, y, \omega) \end{aligned} \quad (5.32)$$



**Fig. 5.15.** Modulating spatial frequency spectrum for rotating reproduced sound field.



**Fig. 5.16.** Block diagram of parametric control of WFR filter for linear arrays.

By using Eqs. (5.32) and (5.10), the driving signals for rotating the sound field can be derived as:

$$\tilde{D}(k_x, \omega) = -2jk_\rho \frac{G_{2D}(k_x, y, 0, \omega)}{G(k_x, y, 0, \omega)} \tilde{P}_{rcv}(k_x - k_{x,rot}, 0, 0, \omega) \quad (5.33)$$

$$= -4j \frac{e^{-jk_\rho y}}{H_0^{(2)}(k_\rho y)} \tilde{P}_{rcv}(k_x - k_{x,rot}, 0, 0, \omega). \quad (5.34)$$

Therefore, the sound pressure on the receiving line is modulated by  $k_{x,rot}$ , and then converted to the driving signals. As shown in Fig. 5.15, for example, a plane wave component in the direction  $\theta_{pw}$  can be rotated to that in the direction  $\theta_{pw} + \theta_{rot}$  by the spatial modulation, and then the rotated plane wave is reproduced in the target area.

### 5.4.3 Practical implementation

Practical implementation of sound field shifting and rotating can be simply extended from that presented in Section 5.3. For rotation, the received signals of the microphone array is spatially modulated according to Eq. (5.29) as:

$$P_{mod}(\omega) = P(\omega) \cdot e^{-jk \sin \theta_{rot}}, \quad (5.35)$$

where the  $M$ -size vector  $\mathbf{P}_{\text{mod}}(\omega)$  denotes the modulated signals. The modulating frequency is set according to the rotation angle  $\theta_{\text{rot}}$ . The spatial frequency spectrum of the modulated signal,  $\tilde{\mathbf{P}}_{\text{mod}}(\omega)$ , is used as the input of the WFR filter as:

$$\tilde{\mathbf{D}}(\omega) = \tilde{\mathbf{F}}(\omega)\tilde{\mathbf{P}}_{\text{mod}}(\omega), \quad (5.36)$$

where the  $i$ -th element of the diagonal component of  $\tilde{\mathbf{F}}(\omega)$  is described as:

$$\tilde{F}_i(\omega) = -4j \frac{e^{j\sqrt{k^2 - k_{x,i}^2} y_{\text{ref}}}}{H_0^{(1)}\left(\sqrt{k^2 - k_{x,i}^2} y_{\text{ref}}\right)} e^{j(k_{x,i} d_x + \sqrt{k^2 - k_{x,i}^2} d_y)}. \quad (5.37)$$

The shift distances,  $d_x$  and  $d_y$ , are set as the phase-shift of the WFR filter in the spatio-temporal frequency domain.

When the LS-based method presented in Section 5.3.2 is applied to sound field shifting and rotating, the control line must be shifted and rotated accordingly. Therefore, recalculation of the inverse transfer function matrix is necessary when the rotation angle or shift distance is changed. Obviously, this recalculation is computationally complicated. On the other hand, in the proposed method, the recalculation is simply achieved and the position of the reproduced sound field is parametrically controlled by changing modulating frequency and shifting phase.

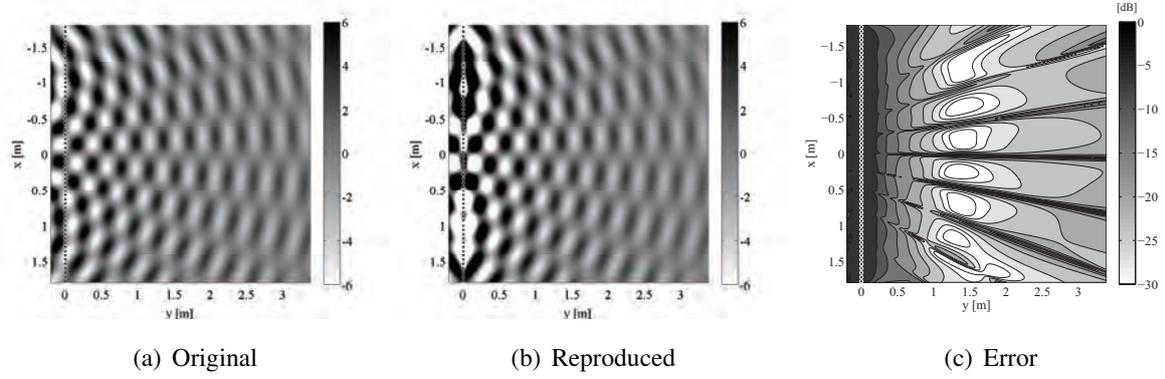
## 5.4.4 Experiments

Numerical simulations were performed under the same conditions as Section 5.3.3. As shown in Fig. 5.4, identical linear microphone and loudspeaker arrays were located at  $y = 0$ , with 64 channels in each array. The filter size of the proposed method was set at  $K = 128$ . The reference line,  $y_{\text{ref}}$ , was set at 1.0 m. The Tukey window function was applied as a tapering window whose sides tapered by 10%.

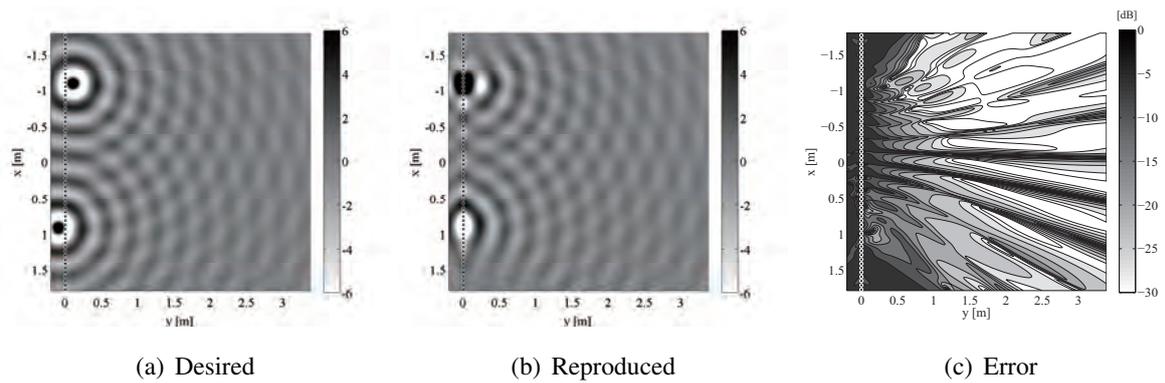
Figures 5.17, 5.18, and 5.19 show the simulation results when the two point sources were located at  $(-1.0, -0.9, 0.0)$  m and  $(1.0, -1.1, 0.0)$  m. The source signal was 1-kHz sinusoidal wave. Figure 5.17 shows the original and reproduced sound pressure and error distributions when  $(d_x, d_y)$  and  $\theta_{\text{rot}}$  were set as  $(0.0, 0.0)$  m and 0 deg, respectively. Therefore, the reconstructed sound field corresponds with the received sound field at the position of the loudspeaker array as in Section 5.3.3.

Figure 5.18 shows the desired and reproduced sound pressure and error distributions when  $(d_x, d_y)$  and  $\theta_{\text{rot}}$  were set as  $(-0.1, 1.0)$  m and 0 deg, respectively. The reproduced sound field was correctly shifted from the original one. As in Chapter 4, the point source at  $(-1.0, -0.9, 0.0)$  m was virtually reproduced in front of the loudspeaker array. The similar limitation of the reproduced sound field as Fig. 4.3 can be seen.

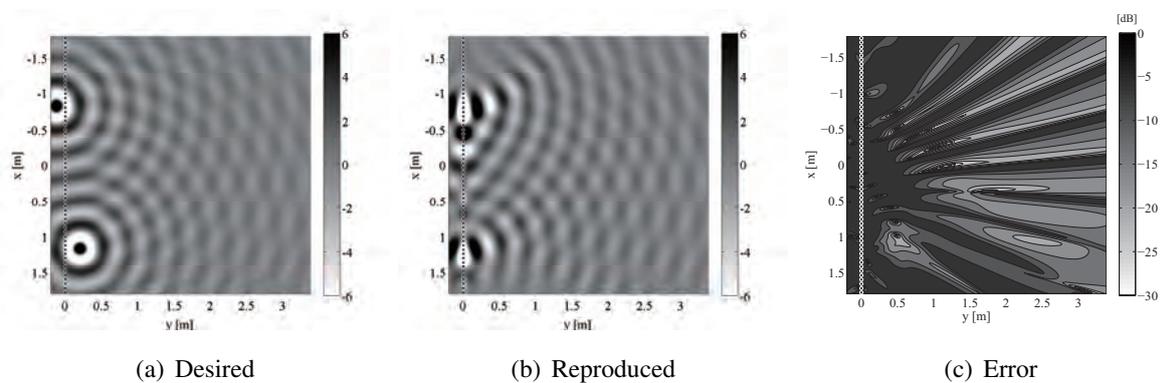
Figure 5.19 shows the desired and reproduced sound pressure and error distributions when  $(d_x, d_y)$  and  $\theta_{\text{rot}}$  were set as  $(-0.1, 1.0)$  m and 15 deg, respectively. The reproduced sound field



**Fig. 5.17.** Original and reproduced sound pressure, and error distribution when two sound sources were located at  $(-1.0, -0.9, 0.0)$  m and  $(1.0, -1.1, 0.0)$  m.



**Fig. 5.18.** Desired and reproduced sound pressure, and error distribution when  $(d_x, d_y)$  was set as  $(-0.1, 1.0)$  m.



**Fig. 5.19.** Desired and reproduced sound pressure, and error distribution when  $(d_x, d_y)$  and  $\theta_{rot}$  were set as  $(-0.1, 1.0)$  m and 15 deg, respectively.

was correctly rotated from the original one. The point source at (1.0, -1.1, 0.0) m was virtually reproduced in front of the loudspeaker array in this rotation case.

These results indicate that the position of the reconstructed sound field can be arbitrarily controlled. It may be highly advantage that the corresponding positions are simply set as shifting and rotating parameters.

## 5.5 WFR filter for directional secondary sources

In Sections 5.1 and 5.2, the WFR equation is derived under the assumption that each secondary source have monopole characteristics. For deriving the WFR equation that includes the directivity of secondary sources, we describe two strategies, multipole expansion and direct measurement. In both approaches, an assumption must be introduced to derive the WFR equation in the spatio-temporal frequency domain, which is that all directivities of secondary sources are the same. This is because the synthesized sound field is described by using a convolution form.

The use of linear distributions of receivers and secondary sources is focused on. A similar formulation is applicable to the case of planar distributions.

### 5.5.1 Multipole expansion approach for WFR equation including directivity of secondary sources

Arbitrary directivity of a secondary source can be described using multipole expansion as [21, 73]:

$$G(\mathbf{r}, \omega) = \sum_p \sum_q \sum_s d_{pqs} \partial_x^p \partial_y^q \partial_z^s \left( \frac{e^{jkr}}{r} \right), \quad (5.38)$$

where  $r = |\mathbf{r}|$ ,  $d_{pqs}$  denotes the coefficient of each pole, and  $p, q, s$  denote the order of multipoles in the direction of  $x, y, z$ , respectively. The spatial Fourier transform of Eq. (5.38) with respect to  $x$  is calculated as:

$$\begin{aligned} \tilde{G}(k_x, y, z, \omega) &= \mathcal{F}_x \left\{ \sum_p \sum_q \sum_s d_{pqs} \partial_x^p \partial_y^q \partial_z^s \left( \frac{e^{jkr}}{r} \right) \right\} \\ &= \sum_p \sum_q \sum_s d_{pqs} (jk_x)^p \partial_y^q \partial_z^s \left( \mathcal{F}_x \left\{ \frac{e^{jkr}}{r} \right\} \right) \\ &= \pi \sum_p \sum_q \sum_s d_{pqs} j^{(p+1)} k_x^p \partial_y^q \partial_z^s \left( H_0^{(1)} \left( k_\rho \sqrt{y^2 + z^2} \right) \right) \end{aligned} \quad (5.39)$$

Therefore, the WFR equation that includes the directivity of secondary sources can be derived by using Eq. (5.15) as:

$$\tilde{D}(k_x, \omega) = \frac{e^{jk_\rho y}}{\pi \sum_p \sum_q \sum_s d_{pqs} j^{(p+1)} k_x^p \partial_y^q \partial_z^s H_0^{(1)}(k_\rho y)} \tilde{P}_{\text{rcv}}(k_x, \omega) \quad (5.40)$$

This is a general formulation of the WFR filter including the directivity of secondary sources. The coefficients of multipole expansion,  $d_{pqS}$ , must be known in order to calculate Eq. (5.40), which is obtained by prior measurements of loudspeaker characteristics. It is possible to calculate  $d_{pqS}$  by fitting to Eq. (5.38) with a linear regression method or analyzing it with a spherical harmonic transform. Because the coefficients of multipole expansion and those of the spherical harmonic transform are different, the conversion of mutual coefficients is required [74, 75].

As an example, we describe the WFR filter when Eq. (5.38) is ignored above the first order, i.e., the sum of the monopole and dipole components. In this approximate case,  $\tilde{G}(k_x, y, 0, \omega)$  can be represented as:

$$\tilde{G}(k_x, y, 0, \omega) \simeq -\pi j \left\{ d_{000} H_0^{(1)}(k_\rho y) - d_{010} k_\rho H_1^{(1)}(k_\rho y) \right\}. \quad (5.41)$$

By using Eq. (5.41), the WFR equation, Eq. (5.40), is approximated as:

$$\tilde{D}(k_x, \omega) = \frac{j}{\pi} \frac{e^{jk_\rho y}}{d_{000} H_0^{(1)}(k_\rho y) - d_{010} k_\rho H_1^{(1)}(k_\rho y)} \tilde{P}_{\text{rcv}}(k_x, \omega). \quad (5.42)$$

For computational simplicity, a simplified form of Eq. (5.42) can be described by using an approximation of the Hankel function in a large argument, Eq. (2.43), as:

$$\tilde{D}(k_x, \omega) = \sqrt{\frac{jk_\rho y}{2\pi}} \frac{1}{d_{000} - d_{010} j k_\rho} \tilde{P}_{\text{rcv}}(k_x, \omega). \quad (5.43)$$

Even when a higher order of Eq. (5.38) is available, the WFR equation can be derived in a similar manner.

Similar to the monopole approximation case, Eqs. (5.42) and (5.43) depends on  $y$ ; therefore, the reference line  $y_{\text{ref}}$  must be set.

### 5.5.2 Direct measurement approach for WFR equation including directivity of secondary sources

A more straightforward way is to directly measure  $G(\mathbf{r}, \omega)$  on a line parallel to the secondary source line, i.e., the reference line. A similar approach in the spectral division method is proposed by Ahrens and Spors [76]. When the spatial Fourier transform of the measured transfer function on  $y = y_{\text{ref}}$  is denoted as  $\tilde{G}_{\text{msr}}(k_x, y_{\text{ref}}, 0, \omega)$ , the WFR equation is described as:

$$\tilde{D}(k_x, \omega) = \frac{e^{jk_\rho y_{\text{ref}}}}{\tilde{G}_{\text{msr}}(k_x, y_{\text{ref}}, 0, \omega)} \tilde{P}_{\text{rcv}}(k_x, \omega). \quad (5.44)$$

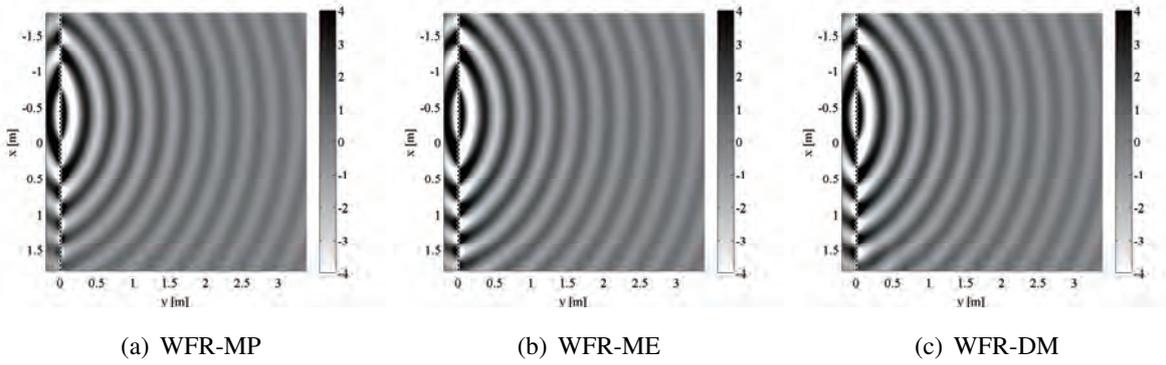


Fig. 5.20. Reproduced sound pressure distribution.

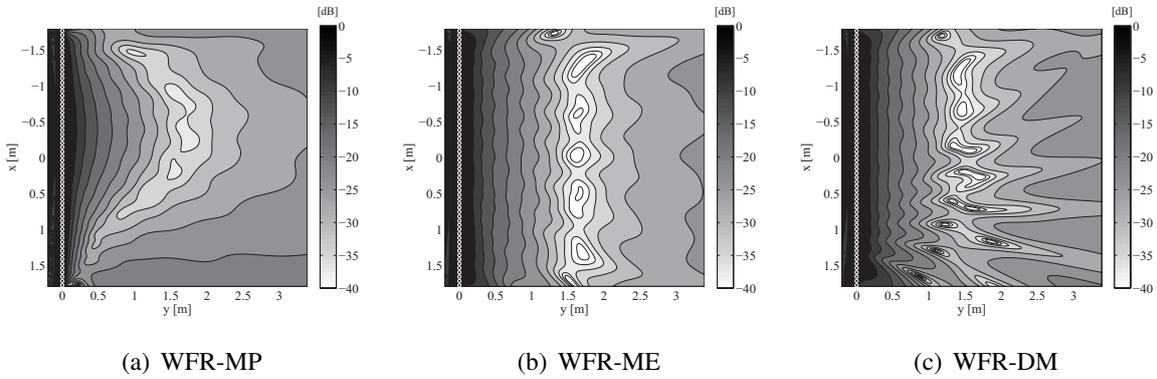


Fig. 5.21. Error distribution.

Practically,  $G(\mathbf{r}, \omega)$  is measured as impulse responses to the linear microphone array on  $y = y_{\text{ref}}$  and  $\tilde{G}_{\text{msr}}(k_x, y_{\text{ref}}, 0, \omega)$  is calculated using the spatial DFT.

### 5.5.3 Experiments

Numerical simulations were performed under the same condition as Section 5.3.3. Three methods, the WFR filtering under the monopole assumption, multipole expansion approach, and direct measurement approach, were compared when the loudspeakers have directivities. These three methods were denoted as WFR-MP, WFR-ME, and WFR-DM, respectively. The WFR filters of the WFR-MP, WFR-ME, and WFR-DM were designed based on Eq. (5.16), (5.42), and (5.44), respectively. As shown in Fig. 5.4, linear microphone and loudspeaker arrays were located at  $y = 0$ , with 64 channels in each array. The interval of the array elements was 6 cm. The directivities of the microphones were assumed to be monopole. Those of the loudspeakers were assumed that  $d_{000} = d_{010} = 0.5$  and the other coefficients were zero in Eq. (5.38). The filter size of the three methods was set at  $K = 128$ . The reference line,  $y_{\text{ref}}$ , was set at 1.0 m. The Tukey window function was applied as a tapering window whose sides tapered by 10%.

Fig. 5.20 and 5.21 show the simulation results when the point source was located at  $(-0.4, -1.0, 0.0)$  m. The source signal was 1-kHz sinusoidal wave. Fig. 5.20 shows the reproduced sound pressure distribution by using the three methods. Fig. 5.21 shows the error distribution calculated based on Eq. (4.12). The reproduction accuracy of the WFR-ME and WFR-DM was high along the reference line. In WFR-DM, the reproduction accuracy in the region of  $x > 0.5$  was slightly disturbed due to the finite length of the directivity measurement. On the other hand, the high reproduction accuracy region of the WFR-MP did not correspond to the reference line. Because the directivities of actual loudspeakers are different depending on frequencies, this fact means that the sound field region that was accurately reproduced using the WFR-MP was smaller than that reproduced using the WFR-ME and WFR-DM.

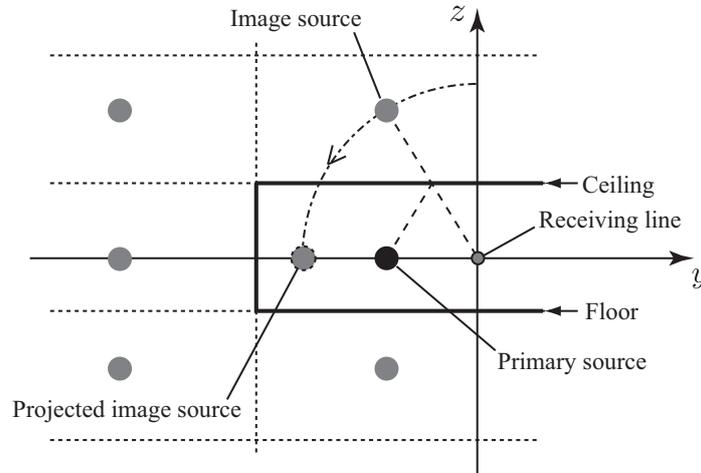
## 5.6 Reproduction in reverberant environment by using circular harmonics beamforming

In practical applications, linear array configuration of microphones and loudspeakers is simple to implement; however, sound sources at different heights cannot be properly reproduced because each microphone in the linear array is assumed to be omni-directional. When sound sources are located at approximately the same height in the recording room, listeners can localize the directions of the sources based on the precedence effect [3] even though the reflections synthesized by the loudspeaker array are different from the original ones. However, in a reverberant environment, this artifact leads to an increase in the reverberation time and a decrease in the direct-to-reverberant energy ratio (DRR), which will be discussed in the following section. Because a different DRR from the original produces a different perceived distance [77], it is important to match the DRR to the original DRR at the time the direct sound wave is properly reproduced. Although several methods have been proposed to improve reproduction accuracy in a reverberant reproduction room [47, 51, 52], this type of artifact originating from reverberation in the recording room has never been discussed in detail.

We propose a method combining circular harmonics (CH) beamforming [78, 79] with the WFR filter as a preprocessing step. A cylindrical microphone array is used instead of a linear microphone array. Each circular array produces a beamformed signal in which the reflections from the ceiling and floor are reduced. The signal is then used as the input signal of the WFR filter for linear arrays; therefore, this preprocess enables the extraction of a two-dimensional sound field in the recording room.

### 5.6.1 Limitation of WFR filter for linear arrays in reverberant environment

Although the linear array configurations are simple to implement, some artifacts appear as discussed in Section 5.2. The major artifact is derived from the two-dimensional assumption of the

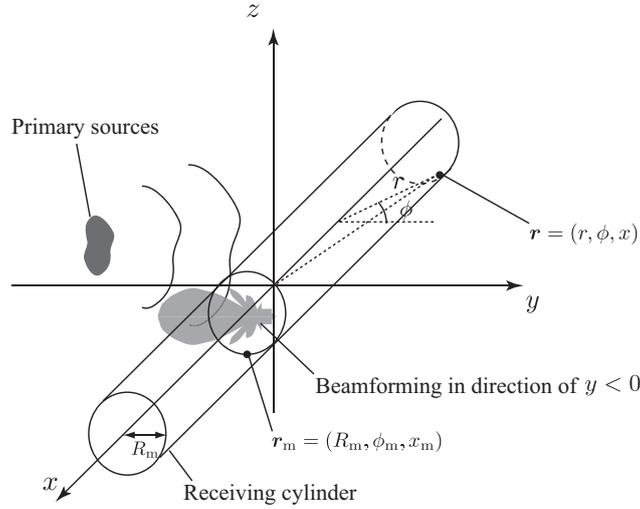


**Fig. 5.22.** All primary sources at axisymmetric positions with central axis on receiving line are equivalent to being projected onto  $x$ - $y$ -plane at  $z = 0$  in source area.

desired sound field, where the primary sources cannot be properly reproduced unless they are on the  $x$ - $y$ -plane at  $z = 0$ . Even though sound sources, such as talkers, can be assumed to be approximately at the same height in practical situations, for example, in telecommunication systems, the image sources are produced by reverberation [80] at different heights in the reverberant source area. As shown in Fig. 5.22, all the image sources at axisymmetric positions relative to the central axis on the receiving line are equivalent to being projected onto the  $x$ - $y$ -plane at  $z = 0$  because the receivers are assumed to have an omni-directional characteristic. This artifact increases reverberation time and reduces the DRR in the target area compared with the sound field generated directly by the primary sources, i.e., original sound field. The WFR filter can properly reproduce the direct sound wave from the primary sources; therefore, listeners can localize the direction of the primary sources based on the precedence effect [3]. However, the difference in DRR may lead to a different perceived distance to a reproduced sound source because the DRR contains the information necessary for sound source distance judgement in human auditory perception [77]. Even if the transform filter is designed using another method, for example, the LS method presented in Section 3.2.3, the same artifact may appear as long as the linear microphone array in the source area is assumed to be omni-directional.

### 5.6.2 DRR matching by combining with CH beamforming

Initial investigation and simulations suggest that the DRR of the reproduced sound field using the WFR filter is comparable to that of the original sound field if the reflections from the floor, ceiling and wall in the  $y > 0$  direction are reduced in the source area. Therefore, it is desirable to extract the sound wave from  $y < 0$  on the  $x$ - $y$ -plane at  $z = 0$  in reverberant environments. To achieve this, we combine a beamforming technique with the WFR filtering. We investigate CH



**Fig. 5.23.** Beamforming as preprocess of WFR filtering by using cylindrical microphone array.

beamforming [78, 79] because it exhibit higher performance at low frequencies compared to the conventional delay-and-sum (DS) beamforming [10]. As shown in Fig. 5.23, the CH beamforming is applied to the received signals of circular arrays at each  $x$  of the cylindrical microphone array as preprocess of WFR filtering.

Assume that microphones are spaced equally and mounted on a sufficiently long and rigid cylindrical baffle with a radius of  $R_m$  (Fig. 5.23). The position vector of each microphone is denoted as  $\mathbf{r}_{m,ij} = (R_m, \phi_{m,j}, x_{m,i})$  in cylindrical coordinates that corresponds to  $\mathbf{r}_{m,ij} = (x_{m,i}, R_m \cos \phi_{m,j}, R_m \sin \phi_{m,j})$  in the Cartesian coordinates. Here,  $i$  ( $1 \leq i \leq M_x$ ) and  $j$  ( $1 \leq j \leq M_\phi$ ) are the indexes of each microphone in the directions of  $x$  and  $\phi$ , respectively.

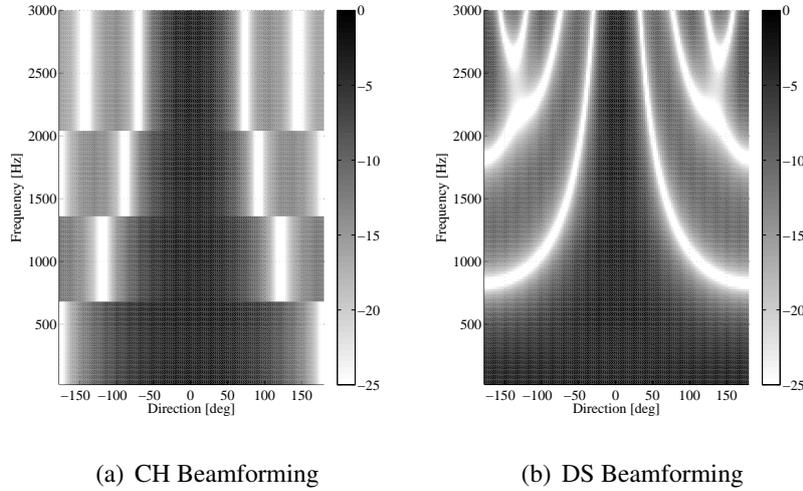
The CH beamforming is derived based on plane wave decomposition of the captured sound field in the CH domain [78]. When the discrete received and beamformed signals are denoted as  $P_{rcv,ij}(\omega)$  and  $P_{BF,i}(\omega)$ , respectively,  $P_{BF,i}(\omega)$  can be obtained as:

$$P_{BF,i}(\omega) = \frac{1}{M_\phi} \sum_{j=1}^{M_\phi} W_j(\omega) P_{rcv,ij}(\omega) \quad (5.45)$$

with

$$W_j(\omega) = \sum_{n=-N}^N \frac{(-j)^{-n} \pi k R_m}{2j} H_n^{(1)'}(k R_m) e^{jn(\phi_{m,j} - \phi_0)}, \quad (5.46)$$

where  $\phi_0$  is the desired direction of beamforming and is set as  $\pi$ ,  $N$  denotes the maximum order of circular harmonics. To minimize sampling and truncation errors,  $N$  for each  $\omega$  should follow  $N \leq \lceil k R_m \rceil$  and  $N \leq \lceil M/2 \rceil$ , where  $\lceil \cdot \rceil$  represents the ceiling function. As an example, directivity



**Fig. 5.24.** Comparison of directivity pattern

pattern of the CH beamforming and DS beamforming are shown in Fig. 5.24 when the number of microphone is 8 and the radius of the circular array is 8.0 cm.

The beamformed signal  $P_{\text{BF},i}(\omega)$  is used as the input of the WFR filter for linear arrays. While each output of the CH beamforming enhances the sound wave from  $\phi_0 = \pi$ , it has an omnidirectional characteristics in the direction of  $x$ ; therefore, direct sound waves from primary sources are properly reproduced.

Fig. 5.25 shows a block diagram of a system using the CH beamforming with the WFR filter for linear arrays. The omni-directional microphones are aligned on the sufficiently long and rigid cylindrical baffle. The sound wave from the direction of  $y < 0$  on the  $x$ - $y$ -plane at  $z = 0$  is enhanced by using the CH beamforming. The beamformed signals are transformed into the spatio-temporal frequency domain, where the WFR filter is applied in order to obtain driving signals of the linear loudspeaker array.

### 5.6.3 Experiments

Numerical simulations were conducted to evaluate the performance of the proposed method combining the CH beamforming. For comparison, the methods combining the DS beamforming and without beamforming using linear microphone array were also explored and tested. These three methods are denoted as WFR-CHBF, WFR-DSBF, and WFR-LINE, respectively. As shown in Fig. 5.26, both the source and target areas were assumed to be (3.84, 7.0, 3.0) m with a reflection coefficient of 0.83. The number of microphones in the cylindrical array,  $M_x \times M_\phi$ , was  $64 \times 3$ . The microphones were spaced 6.0 cm apart along the length of the cylinder, and the radius of the cylinder was 1.8 cm. The center of the cylindrical array was located at (0.0, 0.0, -0.5) m, and its center axis was located at  $(y, z) = (0.0, -0.5)$  m. In the WFR-LINE, the linear microphone

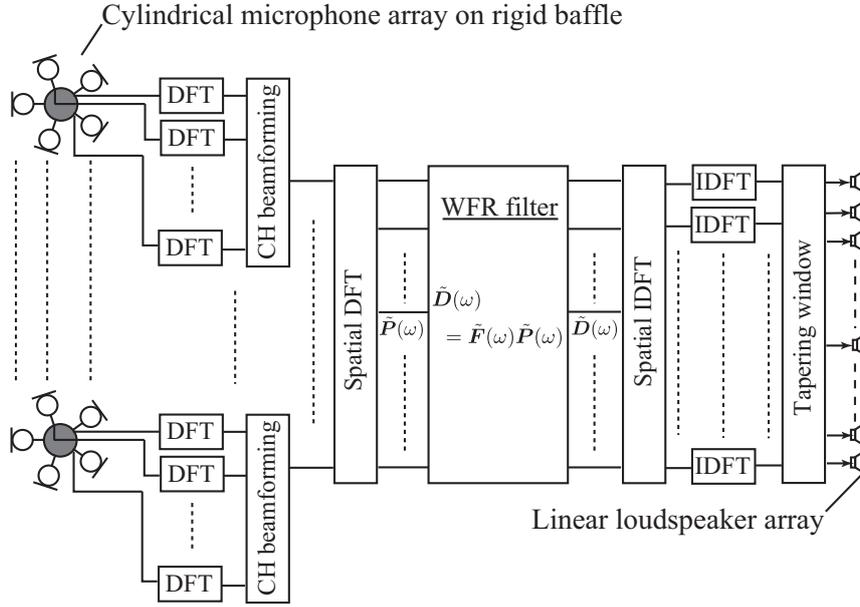


Fig. 5.25. Block diagram of WFR filter with CH beamforming.

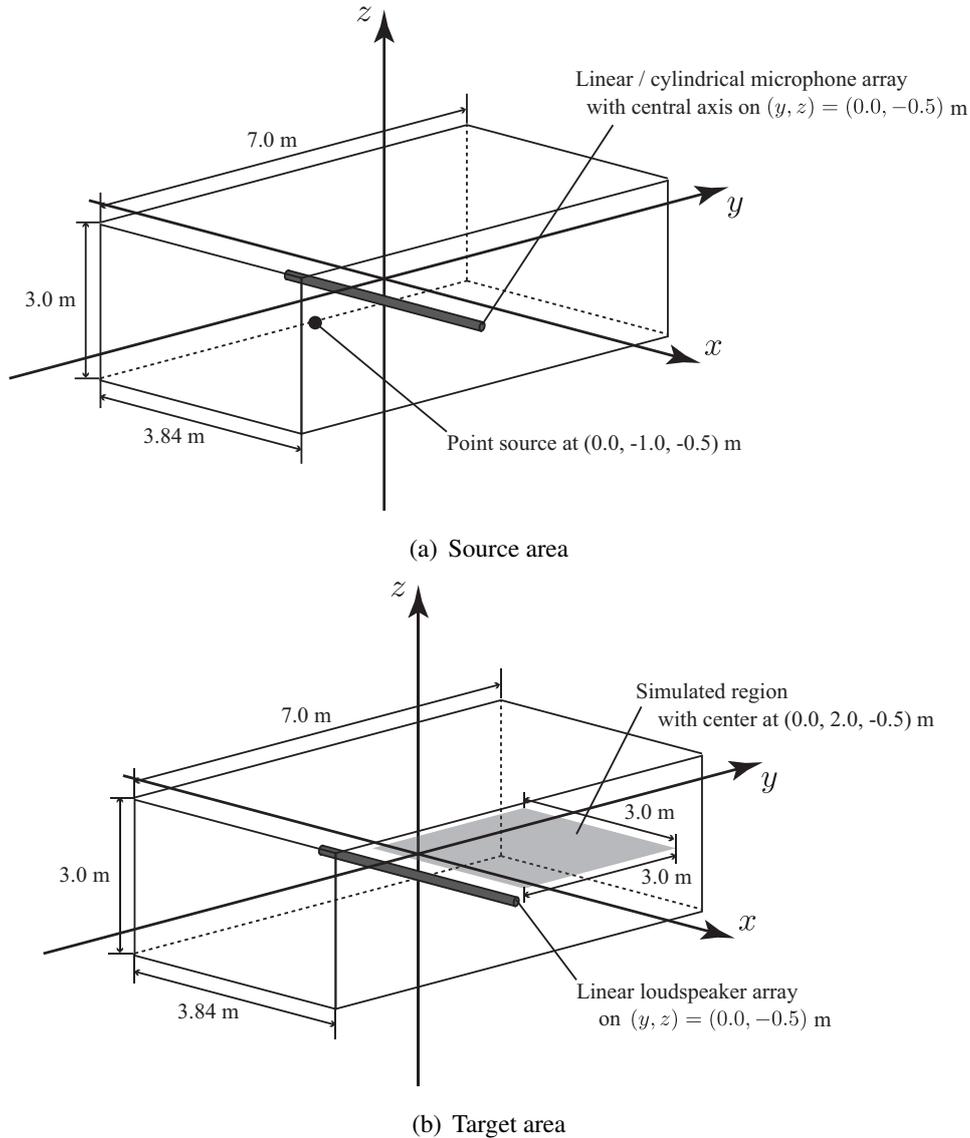
array with 64 channels was located along  $(y, z) = (0.0, -0.5)$  m. The linear loudspeaker array consisted of 64 loudspeakers spaced 6.0 cm apart in the target area. Similarly, its center axis was located at  $(y, z) = (0.0, -0.5)$  m. The directivity of both the microphones and loudspeakers was assumed to be omni-directional. The primary source was assumed to be a point source located at  $(0.0, -1.0, -0.5)$  m. The simulated results were sampled within a  $3.0 \times 3.0$  m area centered at  $(0.0, 2.0, -0.5)$  m with intervals of 0.3 m. The sampling frequency was 48kHz, and the frequency band of interest was chosen to be between 20 Hz to 3 kHz to eliminate the effects of spatial aliasing. The image source method [80] was used to simulate a reverberant environment.

For simplicity, the beamformed signals were generated using the beam pattern of the beamformer. This is because generating signals scattered on the baffled cylinder is very time-consuming due to the large number of elements and the extensive equations that need to be computed. It is verified that this alternative method provided a low-error approximation for all the investigated array sizes. An infinite-length baffled cylindrical microphone array was used to derive the directivity pattern for the WFR-CHBF, and unbaffled, i.e., acoustically transparent, cylindrical microphone array was used for WFR-DSBF.

The reproduced sound field was evaluated using methods and procedures intended to quantify the performance of the proposed method. The results were compared with the original sound field and the three methods. The DRR is calculated as [81]:

$$\text{DRR} = 10 \log_{10} \left( \frac{\sum_{k=0}^{k_d} h^2(x_i, y_j, t_k)}{\sum_{k=k_d+1}^{\infty} h^2(x_i, y_j, t_k)} \right), \quad (5.47)$$

where  $h(x_i, y_j, t_k)$  is the reproduced impulse response from the primary source to the position in



**Fig. 5.26.** Experimental setup

the simulated area,  $(x_i, y_j, -0.5)$ , and samples from  $t_0$  to  $t_{k_d}$  are assumed to represent only the direct-path propagation. Here,  $t_{k_d}$  was set as 16 ms. The value of the DRR changes according to the distance between the measured point and the source; the DRR decreases as the distance increases. Therefore, the DRR plots were generated by taking the average of the DRR values for each  $y_j$  in the simulated region, which is shown in Fig. 5.27. The plot shows that the WFR-DSBF improved the DRR reproduction compared with the WFR-LINE. Additionally, the WFR-CHBF was much closer to the original DRR compared with the WFR-DSBF.

Table 5.2 lists the average reverberation times,  $T_{60}$  [82, 83], in the simulated region. Although the reproduced reverberation time was improved for both beamforming techniques, the WFR-CHBF showed better performance. However, the difference between the original sound field and

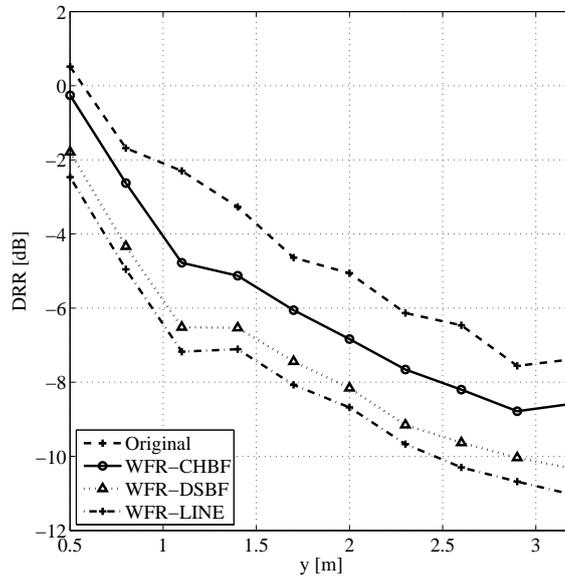


Fig. 5.27. Comparison of DRR averaged at each  $y$ .

Table. 5.2. Comparison of averaged reverberation time

|          | Original | WFR-CHBF | WFR-DSBF | WFR-LINE |
|----------|----------|----------|----------|----------|
| $T_{60}$ | 372 ms   | 473 ms   | 492 ms   | 508 ms   |

the sound field reproduced using the WFR-CHBF was still apparent. This is because the late reverberations were not sufficiently reduced, whereas the early reflections were reduced by the CH beamforming. Therefore, the spatial impression of the sound field reproduction using the WFR-CHBF may be different from that of the original one.

## 5.7 Conclusions

We proposed an SP-DS conversion method for planar or linear arrays of microphones and loudspeakers. The transform filter that converts the received signals of the microphone array into the driving signals of the loudspeaker array, the WFR filter, was analytically derived based on the continuous WFR equation in the spatio-temporal frequency domain. Numerical simulations show that the WFR filter can achieve the same performance as that obtained using the LS method. However, the properties of the proposed WFR filter have many advantages in filter design, filter size, computational cost, and filter stability compared to those of the LS-based transform filter.

It is possible to arbitrarily control the corresponding position of the received and reconstructed sound fields. Phase-shift of the WFR filter corresponds to shifting the reproduced sound field in the horizontal and depth directions. Spatial modulation of the received sound pressure distribution corresponds to rotating the reproduced sound field and the rotation angle is arbitrarily set as a modulating frequency. These parametric control of the reproduced sound has advantages in computational efficiency.

The WFR filter is initially derived under the assumption that secondary sources have monopole characteristics; however, it is possible to include the directivity of the secondary source. Two approaches, multipole expansion and direct measurement, are applied in order to model the directivity. The reproduction accuracy of the sound field can be increased by using the WFR filter including the directivity of the secondary source.

In linear array configuration, the major artifact of the WFR filter is an increase in reverberation time and decrease in the DRR. This artifact may appear as long as a linear configuration of the omni-directional microphone array is assumed even when other SP-DS conversion method is applied. We proposed a method combining the CH beamforming with the WFR filter as preprocessing step. Numerical simulation results indicated that the DRR reproduced using the proposed method was much closer to the original DRR compared with that reproduced by the WFR filter without beamforming. Therefore, it can be considered that perceived distance is improved by using the proposed method.

## Chapter 6

# Wave Field Reconstruction Filter for Cylindrical Arrays of Microphones and Loudspeakers

In order to achieve immersive audio systems, it is important to reproduce a sound field including height (with-height sound field). However, SP-DS conversion for three-dimensional sound field reproduction requires a large number of microphones and loudspeakers as well as a large quantity of signal transmission. From a perceptual point of view, the sound field at a horizontal angle is more important than that at an elevation angle [3]. Therefore, it is more efficient to reproduce a sound field at a lower resolution in elevation direction as opposed to higher resolution in horizontal direction for with-height reproduction.

In the previous chapter, we proposed an SP-DS conversion method for planar or linear microphone and loudspeaker arrays based on the WFR filter in the spatio-temporal frequency domain. It may seem that using the WFR filter for planar arrays with a lower vertical resolution may be an immediate solution. However, a large vertical interval between the array elements leads to severe reproduction error due to spatial aliasing.

In the HOA presented in Section 3.2.2, lower resolution at elevation angle than horizontal angle can be achieved by using lower order expansion coefficients. This type of technique is called the mixed order approach [37, 43]. An alternative method using multiple circular loudspeaker arrays was proposed by Gupta and Abhayapala [84]. In the HOA, the reproduction accuracy on the neighboring region at the center of the spherical array is preserved at a broad frequency band and its width is determined by the maximum available order of the spherical harmonic coefficients [85]. Because the maximum order of the captured sound field is largely depends on the radius and number of samples of the spherical microphone array [42], it is difficult to achieve large listening area in practical implementations as SP-DS conversion.

We propose an SP-DS conversion method using horizontally-arranged cylindrical array configurations, which is based on WFR filtering in the cylindrical harmonic domain. As in the planar and linear cases in Chapter 5, continuous distributions of receivers and secondary sources are assumed and the continuous WFR equation that relates a received sound pressure distribution with driving signals for reproducing sound fields is analytically derived in the cylindrical harmonic domain. The WFR filter that converts the received signals of microphones into driving signals of loudspeakers is uniquely determined by discretizing the WFR equation. In the case of the cylindrical arrays, the WFR filter is stabilized by introducing a model in which microphones are mounted on a rigid cylindrical baffle. The SP-DS conversion is achieved by two-dimensional spatial convolution of the WFR filter and the received signals. With the use of the cylindrical configurations, the reproduction accuracy on the central axis of the cylindrical array is expected to be preserved even when a small number of horizontal linear arrays is used. Therefore, the resulting property may be similar to the WFS at horizontal direction and HOA at elevation direction.

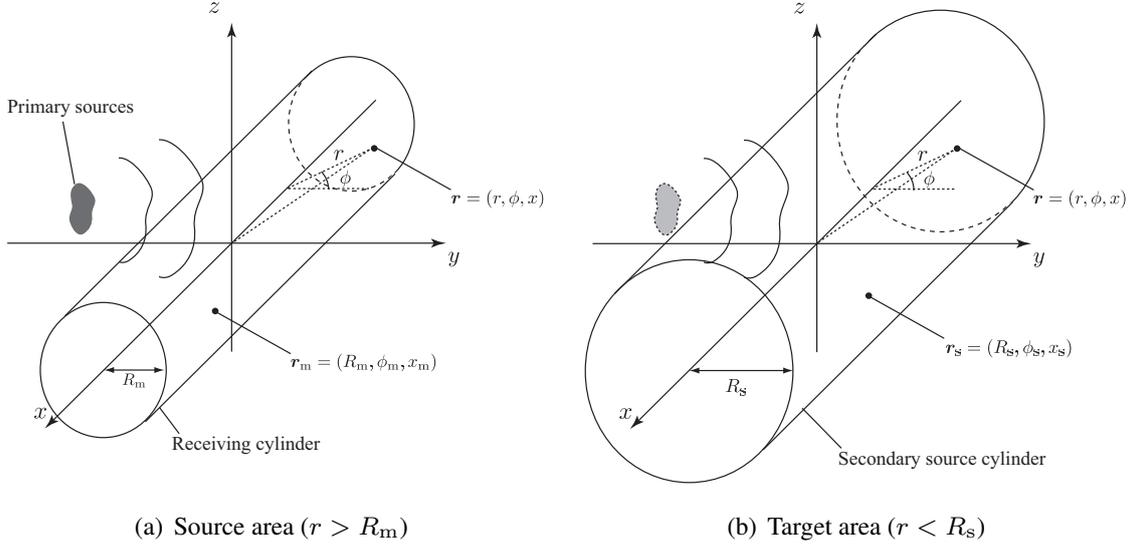
In addition, we also propose an SP-DS conversion method for circular arrays of microphones and loudspeakers based on two-dimensional representation of the WFR filtering method for cylindrical arrays.

## 6.1 WFR equation for cylindrical distributions of receivers and secondary sources

The WFR equation that relates sound pressure distribution obtained by receivers to driving signals of secondary sources is derived in the cylindrical coordinates. As shown in Fig. 6.1, continuous and infinitely-long cylindrical distributions of receivers and secondary sources are respectively arranged in the source and target areas with the central axes on the  $x$ -axis. These cylindrical distributions are assumed to be acoustically transparent (open cylinder). The radii of the receiving and secondary source cylinders are  $R_m$  and  $R_s$ , respectively. It is assumed that  $R_m \leq R_s$  and the sound field inside the secondary source cylinder is reproduced to coincide with that inside the receiving cylinder. The position vector is denoted as  $\mathbf{r} = (r, \phi, x)$  in the cylindrical coordinates, which corresponds to the Cartesian coordinates as follows.

$$\begin{aligned} x &= x \\ y &= r \cos \phi \\ z &= r \sin \phi \end{aligned} \tag{6.1}$$

The position vectors on the receiving and secondary source cylinders are denoted as  $\mathbf{r}_m = (R_m, \phi_m, x_m)$  and  $\mathbf{r}_s = (R_s, \phi_s, x_s)$ , respectively. The received sound pressure at  $\mathbf{r}_m$  and driving signal of the secondary source at  $\mathbf{r}_s$  of the temporal frequency  $\omega$  are denoted as  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  and  $D(\mathbf{r}_s, \omega)$ , respectively. Therefore, the WFR equation is needed which relates  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  with



**Fig. 6.1.** Cylindrical distributions of receivers and secondary sources.

$D(\mathbf{r}_s, \omega)$ .

### 6.1.1 Synthesized sound field

The synthesized sound field created by secondary sources in the target area is denoted as  $P_{\text{syn}}(\mathbf{r}, \omega)$ . When the transfer function from  $\mathbf{r}_s$  to  $\mathbf{r}$  is denoted as  $G(\mathbf{r} - \mathbf{r}_s, \omega)$ ,  $P_{\text{syn}}(\mathbf{r}, \omega)$  is described as:

$$P_{\text{syn}}(\mathbf{r}, \omega) = \int_0^{2\pi} \int_{-\infty}^{\infty} R_s D(\mathbf{r}_s, \omega) G(\mathbf{r} - \mathbf{r}_s, \omega) dx_s d\phi_s. \quad (6.2)$$

Here,  $dS = R_s dx_s d\phi_s$  is used. Equation (6.2) can be regarded as the convolution of  $D(\cdot)$  and  $G(\cdot)$  with respect to  $x$  and  $\phi$ . Therefore, the spatial Fourier transform of Eq. (6.2) with respect to  $x$  and  $\phi$ , i.e., cylindrical Fourier transform, is derived as:

$$\tilde{P}_{\text{syn},n}(r, k_x, \omega) = 2\pi R_s \tilde{D}_n(k_x, \omega) \tilde{G}_n(r - R_s, k_x, \omega), \quad (6.3)$$

where  $k_x$  is the spatial frequency and  $n$  is the order. The variables in the spectral domain, i.e., cylindrical harmonic domain or helical wave spectrum domain [21], is represented by a tilde. When the desired sound field, which is denoted as  $P_{\text{des}}(\mathbf{r}, \omega)$ , is known,  $D(\mathbf{r}_s, \omega)$  can be calculated by using Eq. (6.3) because  $P_{\text{syn}}(\cdot)$  needs to equate to  $P_{\text{des}}(\cdot)$ . However, in this context, the sound pressure distribution on the receiving cylinder is only known. The desired sound field  $P_{\text{des}}(\mathbf{r}, \omega)$  needs to be estimated from  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$ .

### 6.1.2 Derivation of WFR equation for cylindrical receivers

When the receiving cylinder is assumed to be acoustically transparent, the desired sound field,  $P_{\text{des}}(\cdot)$ , corresponds to the internal sound field of the receiving cylinder. Therefore, based on the Fourier acoustics [21], the desired sound field can be described in the cylindrical harmonic domain as:

$$\tilde{P}_{\text{des},n}(r, k_x, \omega) = \tilde{P}_{\text{rcv},n}(R_m, k_x, \omega) \frac{J_n(k_r r)}{J_n(k_r R_m)} \quad (6.4)$$

where

$$k_r = \sqrt{k^2 - k_x^2}. \quad (6.5)$$

When the synthesized and desired sound fields, Eqs. (6.3) and (6.4), are simultaneously solved in the cylindrical harmonic domain, the WFR equation is derived as:

$$\tilde{D}_n(k_x, \omega) = \frac{1}{2\pi R_s \tilde{G}_n(r - R_s, k_x, \omega)} \frac{J_n(k_r r)}{J_n(k_r R_m)} \tilde{P}_{\text{rcv},n}(R_m, k_x, \omega) \quad (6.6)$$

For simplicity, we assume that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  can be approximated as monopole, Eq. (3.2). Because the representation in the cylindrical harmonic domain of Eq. (3.2),  $\tilde{G}_{3D,n}(r - R_s, k_x, \omega)$ , is analytically derived as Eq. (A.27) in the Appendix, the WFR equation for monopole secondary source is derived by substituting Eq. (A.27) into Eq. (6.6) as:

$$\tilde{D}_n(k_x, \omega) = \frac{2}{\pi j R_s H_n^{(1)}(k_r R_s) J_n(k_r R_m)} \tilde{P}_{\text{rcv},n}(R_m, k_x, \omega). \quad (6.7)$$

Only the sound pressure distribution on the receiving cylinder is required to calculate Eq. (6.7). However, the denominator of Eq. (6.7) includes  $J_n(k_r R_m)$ , which is a real valued function that have many zeros. This fact means that  $\tilde{D}_n(k_x, \omega)$  diverges at the zeros of  $J_n(k_r R_m)$ . This instability originates from the non-uniqueness at the forbidden frequencies [55].

### 6.1.3 Derivation of WFR equation for cylindrical receivers with rigid baffle

In order to avoid the instability of Eq. (6.7), we introduce a cylindrical baffle for the receivers in the source area. The reflection characteristics on the surface of the baffle is assumed to be known. Several attempts to use a rigid baffle for avoiding computational instabilities can be found in the context of plane wave decomposition by using spherical or circular microphone array [38, 41, 42, 79, 86–88].

It is assumed that the acoustically rigid and infinitely-long cylindrical baffle (rigid cylinder) of radius  $R_b$  ( $\leq R_m$ ) is set along the central axis on the  $x$ -axis in the source area. In this case, the

effect of the rigid cylinder needs to be removed from the received sound field to obtain the desired sound field. The sound field in the source area,  $P_{\text{rcv}}(\mathbf{r}, \omega)$ , can be represented as the summation of the incident and scattering sound fields,  $P_{\text{inc}}(\mathbf{r}, \omega)$  and  $P_{\text{sct}}(\mathbf{r}, \omega)$ , as:

$$P_{\text{rcv}}(\mathbf{r}, \omega) = P_{\text{inc}}(\mathbf{r}, \omega) + P_{\text{sct}}(\mathbf{r}, \omega). \quad (6.8)$$

The sound pressure gradient on the surface of the rigid baffle becomes zero as:

$$\left. \frac{\partial}{\partial r} \{P_{\text{inc}}(\mathbf{r}, \omega) + P_{\text{sct}}(\mathbf{r}, \omega)\} \right|_{r=R_b} = 0. \quad (6.9)$$

Here,  $P_{\text{inc}}(\cdot)$  and  $P_{\text{sct}}(\cdot)$  can be represented in the cylindrical harmonic domain as [21]:

$$P_{\text{inc}}(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \check{P}_{\text{inc},n}(k_x, \omega) J_n(k_r r) e^{jn\phi} e^{jk_x x} dk_x, \quad (6.10)$$

and

$$P_{\text{sct}}(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \check{P}_{\text{sct},n}(k_x, \omega) H_n^{(1)}(k_r r) e^{jn\phi} e^{jk_x x} dk_x, \quad (6.11)$$

where the variables excluding radial component of the cylindrical harmonic coefficients are represented by breve, i.e.,  $\check{P}_{\text{inc},n}(r, k_x, \omega) = \check{P}_{\text{inc},n}(k_x, \omega) J_n(k_r r)$ . By substituting Eqs. (6.10) and (6.11) into Eq. (6.9),

$$\begin{aligned} & \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \check{P}_{\text{inc},n}(k_x, \omega) k_r J'_n(k_r R_b) e^{jn\phi} e^{jk_x x} dk_x \\ & + \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \check{P}_{\text{sct},n}(k_x, \omega) k_r H_n^{(1)'}(k_r R_b) e^{jn\phi} e^{jk_x x} dk_x = 0. \end{aligned} \quad (6.12)$$

Therefore, the following relation between the incident and scattering sound fields can be derived:

$$\check{P}_{\text{sct},n}(k_x, \omega) = -\frac{J'_n(k_r R_b)}{H_n^{(1)'}(k_r R_b)} \check{P}_{\text{inc},n}(k_x, \omega). \quad (6.13)$$

From Eqs. (6.8), (6.10), (6.11), and (6.13), the sound field in the source area,  $P_{\text{rcv}}(\mathbf{r}, \omega)$ , can be described by using  $\check{P}_{\text{inc},n}(k_x, \omega)$  as:

$$P_{\text{rcv}}(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left( J_n(k_r r) - \frac{J'_n(k_r R_b)}{H_n^{(1)'}(k_r R_b)} H_n^{(1)}(k_r r) \right) \check{P}_{\text{inc},n}(k_x, \omega) e^{jn\phi} e^{jk_x x} dk_x. \quad (6.14)$$

Therefore,  $\check{P}_{\text{inc},n}(k_x, \omega)$  is represented by using  $\check{P}_{\text{rcv},n}(r, k_x, \omega)$  as

$$\check{P}_{\text{inc},n}(k_x, \omega) = \frac{H_n^{(1)'}(k_r R_b)}{J_n(k_r r) H_n^{(1)'}(k_r R_b) - J'_n(k_r R_b) H_n^{(1)}(k_r r)} \check{P}_{\text{rcv},n}(r, k_x, \omega). \quad (6.15)$$

The desired sound field,  $P_{\text{des}}(\cdot)$ , corresponds to the incident sound field,  $P_{\text{inc}}(\cdot)$ . Therefore,  $P_{\text{des}}(\cdot)$  can be described by using the cylindrical harmonic coefficients on the receiving cylinder,  $\tilde{P}_{\text{rcv},n}(R_m, k_x, \omega)$ , from Eq. (6.15) as:

$$\begin{aligned} \tilde{P}_{\text{des},n}(r, k_x, \omega) &= \check{P}_{\text{inc},n}(k_x, \omega) J_n(k_r r) \\ &= \frac{J_n(k_r r) H_n^{(1)'}(k_r R_b)}{J_n(k_r R_m) H_n^{(1)'}(k_r R_b) - J_n'(k_r R_b) H_n^{(1)}(k_r R_m)} \tilde{P}_{\text{rcv},n}(R_m, k_x, \omega). \end{aligned} \quad (6.16)$$

Similar to the open cylinder case, the WFR equation can be derived by simultaneously solving the synthesized and desired sound fields, Eqs. (6.3) and (6.16), as:

$$\begin{aligned} \tilde{D}_n(k_x, \omega) &= \frac{1}{2\pi R_s \tilde{G}_n(r - R_s, k_x, \omega)} \\ &\quad \cdot \frac{J_n(k_r R_m) H_n^{(1)'}(k_r R_b)}{J_n(k_r R_m) H_n^{(1)'}(k_r R_b) - J_n'(k_r R_b) H_n^{(1)}(k_r R_m)} \tilde{P}_{\text{rcv},n}(R_m, \omega) \\ &= \frac{2H_n^{(1)'}(k_r R_b)}{\pi j R_s H_n^{(1)}(k_r R_s) \left( J_n(k_r R_m) H_n^{(1)'}(k_r R_b) - J_n'(k_r R_b) H_n^{(1)}(k_r R_m) \right)} \\ &\quad \cdot \tilde{P}_{\text{rcv},n}(R_m, \omega). \end{aligned} \quad (6.17)$$

Again, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. When the radii of the receivers and rigid baffle are the same, i.e.,  $R_b = R_m$ , Eq. (6.17) can be simplified as:

$$\tilde{D}_n(k_x, \omega) = -\frac{k_r R_m H_n^{(1)'}(k_r R_m)}{R_s H_n^{(1)}(k_r R_s)} \tilde{P}_{\text{rcv},n}(R_m, \omega). \quad (6.18)$$

Here, Eq. (2.45) is used to derive Eq. (6.18). The Bessel function of the denominator of Eq. (6.7) is eliminated in Eq. (6.18). Therefore, Eq. (6.18) is much more stable than Eq. (6.17). The driving signals can be stably obtained from the sound pressure distribution on the receiving cylinder in the rigid cylinder case.

#### 6.1.4 Derivation of WFR equation for cylindrical receivers with soft baffle

As another example of a baffle, we consider an acoustically soft baffle. Practically, such a baffle is difficult to implement. It is assumed that the acoustically soft and infinitely-long cylindrical baffle of radius  $R_b$  ( $\leq R_m$ ) is set along the central axis on the  $x$ -axis in the source area. Similar to the rigid baffle case, the sound field in the source area can be represented as the summation of the incident and scattering sound fields,  $P_{\text{inc}}(\mathbf{r}, \omega)$  and  $P_{\text{sct}}(\mathbf{r}, \omega)$ , as Eq. (6.8). The sound pressure on the surface of the soft baffle becomes zero as:

$$\{P_{\text{inc}}(\mathbf{r}, \omega) + P_{\text{sct}}(\mathbf{r}, \omega)\}|_{r=R_b} = 0 \quad (6.19)$$

By substituting Eqs. (6.10) and (6.11) into Eq. (6.19),

$$\begin{aligned} & \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \check{P}_{\text{inc},n}(k_x, \omega) J_n(k_r R_b) e^{jn\phi} e^{jk_x x} dk_x \\ & + \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \check{P}_{\text{sct},n}(k_x, \omega) H_n^{(1)}(k_r R_b) e^{jn\phi} e^{jk_x x} dk_x = 0 \end{aligned} \quad (6.20)$$

Therefore, the following relation between the incident and scattering sound fields can be derived:

$$\check{P}_{\text{sct},n}(k_x, \omega) = -\frac{J_n(k_r R_b)}{H_n^{(1)}(k_r R_b)} \check{P}_{\text{inc},n}(k_x, \omega) \quad (6.21)$$

From Eqs. (6.8), (6.10), (6.11), and (6.21), the sound field in the source area,  $P_{\text{rcv}}(\mathbf{r}, \omega)$ , can be described by using  $\check{P}_{\text{inc},n}(k_x, \omega)$  as:

$$P_{\text{rcv}}(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left( J_n(k_r r) - \frac{J_n(k_r R_b)}{H_n^{(1)}(k_r R_b)} H_n^{(1)}(k_r r) \right) \check{P}_{\text{inc},n}(k_x, \omega) e^{jn\phi} e^{jk_x x} dk_x \quad (6.22)$$

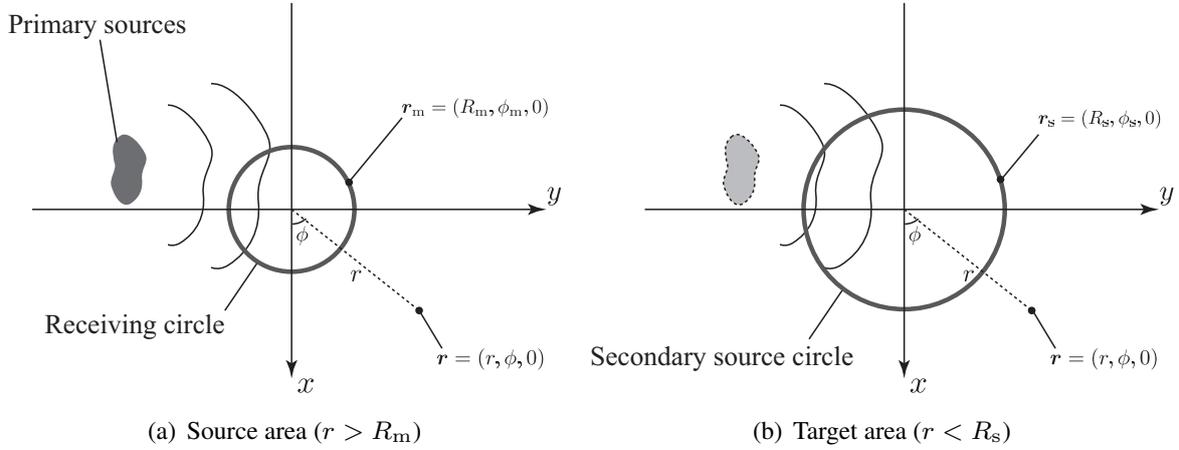
In a similar manner to the rigid cylindrical baffle case, the WFR equation can be derived as

$$\begin{aligned} \tilde{D}_n(k_x, \omega) = & \frac{2H_n^{(1)}(k_r R_b)}{\pi j R_s H_n^{(1)}(k_r R_s) \left( J_n(k_r R_m) H_n^{(1)}(k_r R_b) - J_n(k_r R_b) H_n^{(1)}(k_r R_m) \right)} \\ & \cdot \check{P}_{\text{rcv},n}(R_m, k_x, \omega). \end{aligned} \quad (6.23)$$

Again, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. The denominator of Eq. (6.23) may not also exhibit zeros when  $R_b$  is smaller than  $R_m$ .

## 6.2 WFR equation for circular distributions of receivers and secondary sources

The WFR equation for circular distributions of receivers and secondary sources can be obtained based on the WFR equation for cylindrical distributions presented in the previous section. In this section, we derive the WFR equation for circular distributions in the cylindrical coordinates. As shown in Fig. 6.2, continuous circular distributions of receivers and secondary sources are arranged in the source and target areas, respectively, with the center of the circle on the origin. These circular distributions are assumed to be acoustically transparent (open circle). The radii of the receiving and secondary source circles are  $R_m$  and  $R_s$ , respectively. It is assumed that  $R_m \leq R_s$  and the sound field inside the secondary source circle is reproduced to coincide with that inside the receiving circle. The position vector is denoted as  $\mathbf{r} = (r, \phi, z)$  in the cylindrical



**Fig. 6.2.** Circular distributions of receivers and secondary sources.

coordinates. Note that the correspondence relation with the Cartesian coordinates is different from that in Section 6.1.

$$\begin{aligned}
 x &= r \cos \phi \\
 y &= r \sin \phi \\
 z &= z
 \end{aligned} \tag{6.24}$$

The position vectors on the receiving and secondary source circles are denoted as  $\mathbf{r}_m = (R_m, \phi_m, 0)$  and  $\mathbf{r}_s = (R_s, \phi_s, 0)$ , respectively. The sound pressure at  $\mathbf{r}_m$  and driving signal of the secondary source at  $\mathbf{r}_s$  of the temporal frequency  $\omega$  are denoted as  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  and  $D(\mathbf{r}_s, \omega)$ , respectively.

### 6.2.1 Synthesized sound field

The synthesized sound field created by secondary sources in the target area is denoted as  $P_{\text{syn}}(\mathbf{r}, \omega)$ . When the transfer function from  $\mathbf{r}_s$  to  $\mathbf{r}$  is denoted as  $G(\mathbf{r} - \mathbf{r}_s, \omega)$ ,  $P_{\text{syn}}(\mathbf{r}, \omega)$  is described as

$$P_{\text{syn}}(\mathbf{r}, \omega) = \int_0^{2\pi} R_s D(\mathbf{r}_s, \omega) G(\mathbf{r} - \mathbf{r}_s, \omega) d\phi_s \tag{6.25}$$

Here,  $dS = R_s d\phi_s$  is used. Equation (6.25) can be regarded as the convolution of  $D(\cdot)$  and  $G(\cdot)$  with respect to  $\phi$ . Therefore, the spatial Fourier transform of Eq. (6.25) with respect to  $\phi$  is derived as

$$\tilde{P}_{\text{syn},n}(r, 0, \omega) = 2\pi R_s \tilde{D}_n(\omega) \tilde{G}_n(r - R_s, \omega), \tag{6.26}$$

where  $n$  is the order. The variables in the spectral domain, i.e., circular harmonic domain, is represented by a tilde.

### 6.2.2 Derivation of WFR equation for circular receivers

The three-dimensional desired sound field needs to be estimated from the sound pressure distribution on the receiving circle. Therefore, we assumed that the sound field in the source area is invariant with regard to changes in the  $z$ -axis. Under this assumption, the desired sound field can be represented as an extrapolation of the sound field in the circular harmonics domain as:

$$P_{\text{des}}(r, \phi, 0, \omega) = \sum_{n=-\infty}^{\infty} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega) \frac{J_n(kr)}{J_n(kR_m)} e^{jn\phi}. \quad (6.27)$$

This assumption is only valid when the primary sound sources are only on the  $x$ - $y$ -plane at  $z = 0$ .

Because the synthesized and desired sound fields are to be matched, the WFR equation can be derived by solving Eqs. (6.26) and (6.27) as

$$\tilde{D}_n(\omega) = \frac{1}{2\pi R_s \tilde{G}_n(r - R_s, \omega)} \frac{J_n(kr)}{J_n(kR_m)} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega) \quad (6.28)$$

For simplicity, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics, Eq. (3.2). The cylindrical harmonic coefficients of  $G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega)$  is described as Eq. (A.28) in the Appendix. Substituting Eq. (A.28) into Eq. (6.28), the WFR equation for monopole secondary sources is derived as

$$\tilde{D}_n(\omega) = -4j \frac{J_n(kr)}{R_s J_n(kR_m)} \left( \int_{-\infty}^{\infty} H_n^{(1)}(k_r R_s) J_n(k_r r) dk_z \right)^{-1} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega), \quad (6.29)$$

where

$$k_r = \sqrt{k^2 - k_z^2}. \quad (6.30)$$

The integral in Eq. (6.29) must be calculated numerically. Because  $r$  is included in Eq. (6.29), the reference radius,  $r = R_{\text{ref}}$ , must be set. The amplitude of the synthesized sound field correspond to that of the desired one only on the reference radius.

When it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has line source characteristics, i.e., two-dimensional Green function Eq. (3.10), the WFR equation becomes simpler. The circular harmonic coefficients of  $G_{2D}(\mathbf{r} - \mathbf{r}_s, \omega)$  is described as Eq. (A.30). Substituting Eq. (A.30) into Eq. (6.28), the WFR equation for line source characteristics is derived as

$$\tilde{D}_n(\omega) = \frac{2}{\pi j R_s H_n^{(1)}(kR_s) J_n(kR_m)} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega). \quad (6.31)$$

However, the characteristics of actual loudspeakers is closer to monopole source than line source. The difference between monopole and line sources can be described as Eq. (3.11). Therefore, the

synthesized sound field can be coincide with the desired one only at the origin when monopole sources are approximately employed as an alternative to line sources. In this approximate case, it is necessary to compensate frequency characteristics of the driving signals by multiplying  $\sqrt{2\pi j R_s/k}$ .

Similar to the cylindrical distributions case, the denominator of Eqs. (6.29) and (6.31) includes  $J_n(kR_m)$ . Therefore,  $\tilde{D}_n(\omega)$  diverges at the zeros of  $J_n(kR_m)$ .

### 6.2.3 Derivation of the WFR equation for circular receivers with rigid cylindrical baffle

Again, we introduce a rigid cylindrical baffle in the source area to avoid instability of Eqs. (6.29) and (6.31). As in Eq. (6.14), when the rigid and infinitely-long cylindrical baffle of radius  $R_b$  is set along the central axis on the  $z$ -axis in the source area, the sound field in the source area,  $P_{rcv}(\cdot)$ , is represented in the cylindrical harmonic domain as:

$$P_{rcv}(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left( J_n(k_r r) - \frac{J'_n(k_r R_b)}{H_n^{(1)'}(k_r R_b)} H_n^{(1)}(k_r r) \right) \check{P}_{inc,n}(k_z, \omega) e^{jn\phi} e^{jk_z z} dk_z, \quad (6.32)$$

where  $\check{P}_{inc,n}(\cdot)$  denotes the cylindrical harmonic coefficients excluding radial components of the incident sound field. Therefore,  $\check{P}_{inc,n}(\cdot)$  can be described by using  $\tilde{P}_{rcv,n}(r, 0, \omega)$  when the desired sound field is assumed to be invariant to changing in  $z$ -axis:

$$\check{P}_{inc,n}(0, \omega) = \frac{H_n^{(1)'}(kR_b)}{J_n(kr)H_n^{(1)'}(kR_b) - J'_n(kR_b)H_n^{(1)}(kr)} \tilde{P}_{rcv,n}(r, 0, \omega) \quad (6.33)$$

The desired sound field,  $P_{des}(\cdot)$ , correspond to the incident sound field,  $P_{inc}(\cdot)$ . Therefore,  $P_{des}(\cdot)$  can be described by using the circular harmonic coefficients on the receiving circle,  $\tilde{P}_{rcv,n}(R_m, 0, \omega)$ , from Eq. (6.33) as

$$\tilde{P}_{des,n}(r, 0, \omega) = \frac{J_n(kr)H_n^{(1)'}(kR_b)}{J_n(kR_m)H_n^{(1)'}(kR_b) - J'_n(kR_b)H_n^{(1)}(kR_m)} \tilde{P}_{rcv,n}(R_m, 0, \omega) \quad (6.34)$$

By solving Eqs. (6.26) and (6.34), the WFR equation can be derived as

$$\tilde{D}_n(\omega) = \frac{J_n(kr)H_n^{(1)'}(kR_b)}{2\pi^3 j R_s \left( J_n(kR_m)H_n^{(1)'}(kR_b) - J'_n(kR_b)H_n^{(1)}(kR_m) \right)} \cdot \left( \int_{-\infty}^{\infty} H_n^{(1)}(k_r R_s) J_n(k_r r) dk_x \right)^{-1} \tilde{P}_{rcv,n}(R_m, 0, \omega). \quad (6.35)$$

Again, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. Because  $r$  is included in Eq. (6.17), the reference radius,  $r = R_{ref}$ , must be set. When the radii of the receivers and rigid

baffle are the same, i.e.,  $R_b = R_m$ , Eq. (6.35) can be simplified as:

$$\tilde{D}_n(\omega) = -\frac{kR_m}{4\pi^2 R_s} J_n(kR_{\text{ref}}) H_n^{(1)'}(kR_m) \cdot \left( \int_{-\infty}^{\infty} J_n(k_r R_{\text{ref}}) H_n^{(1)}(k_r R_s) dk_x \right)^{-1} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega). \quad (6.36)$$

Here, Eq (2.45) is used.

When it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has line source characteristics, i.e., two-dimensional Green function, the WFR equation becomes simpler. The line source characteristics is represented as Eq. (A.30). Therefore, the WFR equation for line source characteristics is derived as

$$\tilde{D}_n(\omega) = -\frac{kR_m H_n^{(1)'}(kR_m)}{4\pi^2 R_s H_n^{(1)}(kR_s)} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega) \quad (6.37)$$

Similar to the open receiving circle case, it is necessary to compensate frequency characteristics of the driving signals by multiplying  $\sqrt{2\pi j R_s/k}$  when monopole sources are approximately employed as an alternative to line sources.

## 6.2.4 Derivation of the WFR equation for circular receivers with soft cylindrical baffle

As another example of a baffle, an acoustically soft baffle is considered. As in Eq. (6.22), when the soft and infinitely-long cylindrical baffle is set along the central axis on the  $z$ -axis in the source area, the sound field in the source area,  $P_{\text{rcv}}(\cdot)$ , is represented in the cylindrical harmonic domain as:

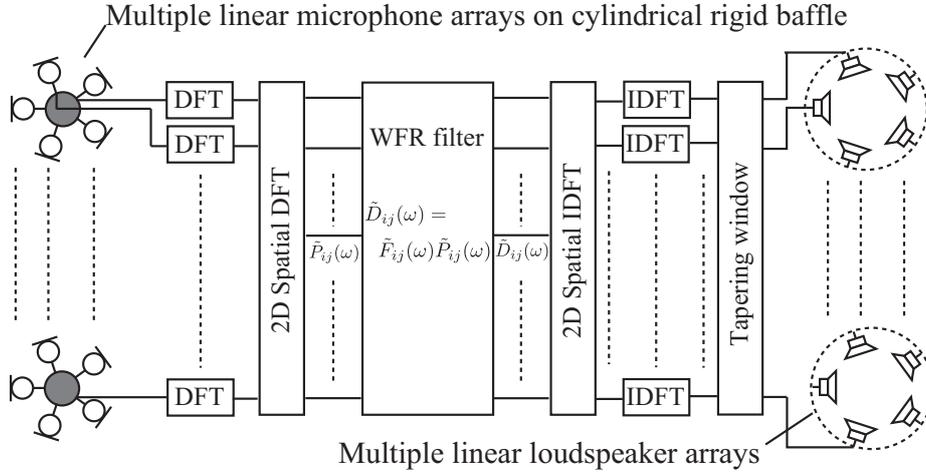
$$P_{\text{rcv}}(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left( J_n(k_r r) - \frac{J_n(k_r R_b)}{H_n^{(1)}(k_r R_b)} H_n^{(1)}(k_r r) \right) \check{P}_{\text{inc},n}(k_z, \omega) e^{jn\phi} e^{jk_z z} dk_z, \quad (6.38)$$

where  $\check{P}_{\text{inc},n}(\cdot)$  denotes the cylindrical harmonic coefficients excluding radial components of the incident sound field.

In a similar manner to the rigid cylindrical baffle case, the WFR equation can be derived as

$$\tilde{D}_n(\omega) = \frac{J_n(kr) H_n^{(1)}(kR_b)}{2\pi^3 j R_s \left( J_n(kR_m) H_n^{(1)}(kR_b) - J_n(kR_b) H_n^{(1)}(kR_m) \right)} \cdot \left( \int_{-\infty}^{\infty} H_n^{(1)}(k_r R_s) J_n(k_r r) dk_x \right)^{-1} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega) \quad (6.39)$$

Again, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. Because  $r$  is included in Eq. (6.17), the reference radius,  $r = R_{\text{ref}}$ , must be set. Similar to the cylindrical receivers with the



**Fig. 6.3.** Block diagram of WFR filter for multiple linear arrays.

soft baffle case,  $R_b$  must be smaller than  $R_m$  in order to avoid that the denominator of Eq. (6.39) becomes zero.

When it is assumed that  $G(\mathbf{r}-\mathbf{r}_s, \omega)$  has line source characteristics, the WFR equation becomes simpler. The line source characteristics is represented as Eq. (A.30) in the Appendix. Therefore, the WFR equation for line source characteristics is derived as:

$$\tilde{D}_n(\omega) = \frac{H_n^{(1)}(kR_b)}{2\pi^3 j R_s H_n^{(1)}(kR_s) \left( J_n(kR_m) H_n^{(1)}(kR_b) - J_n(kR_b) H_n^{(1)}(kR_m) \right)} \tilde{P}_{\text{rcv},n}(R_m, 0, \omega). \quad (6.40)$$

Similar to the open receiving circle case, it is necessary to compensate frequency characteristics of the driving signals by multiplying  $\sqrt{2\pi j R_s/k}$  when monopole sources are approximately employed as an alternative to line sources.

### 6.3 Implementation of WFR filter for multiple linear arrays

When the WFR filter is designed based on the WFR equation using cylindrical microphone and loudspeaker arrays, it may be possible to reproduce sound field at lower resolution at an elevation angle than that at a horizontal angle. The WFR filter for cylindrical arrays with rigid baffle is described based on the WFR equation in the cylindrical harmonic domain.

In a practical implementation, the receiver and secondary source distributions must be discretized and truncated. A block diagram of the proposed method using cylindrical arrays of microphones and loudspeakers is depicted in Fig. 6.3. We employed the sampling scheme that the respective  $\phi$  and  $x$  are equally discretized in order to apply 2D-FFT algorithm to cylindrical DFT. The interval and number of the array elements at the angular positions ( $\phi$ ) are respectively

denoted as  $\Delta\phi$  and  $M_\phi$ , and those at the axial positions ( $x$ ) are respectively denoted as  $\Delta x$  and  $M_x$ . The received signals of the microphone array and the driving signals of the loudspeaker array in the temporal frequency domain are respectively denoted as:

$$\mathbf{P}(\omega) = \begin{bmatrix} P_{11}(\omega) & P_{12}(\omega) & \cdots & P_{1M_x}(\omega) \\ P_{21}(\omega) & P_{22}(\omega) & & P_{2M_x}(\omega) \\ \vdots & & \ddots & \vdots \\ P_{M_\phi 1}(\omega) & P_{M_\phi 2}(\omega) & \cdots & P_{M_\phi M_x}(\omega) \end{bmatrix}$$

and

$$\mathbf{D}(\omega) = \begin{bmatrix} D_{11}(\omega) & D_{12}(\omega) & \cdots & D_{1M_x}(\omega) \\ D_{21}(\omega) & D_{22}(\omega) & & D_{2M_x}(\omega) \\ \vdots & & \ddots & \vdots \\ D_{M_\phi 1}(\omega) & D_{M_\phi 2}(\omega) & \cdots & D_{M_\phi M_x}(\omega) \end{bmatrix}.$$

When the DFT sizes with respect to  $\phi$  and  $x$  are respectively denoted as  $N_{\text{FT},\phi}$  and  $N_{\text{FT},x}$ , the cylindrical harmonic coefficients of the received signals and those of the driving signals are represented as  $N_{\text{FT},\phi} \times N_{\text{FT},x}$ -size matrices as:

$$\tilde{\mathbf{P}}(\omega) = \begin{bmatrix} \tilde{P}_{11}(\omega) & \tilde{P}_{12}(\omega) & \cdots & \tilde{P}_{1N_{\text{FT},x}}(\omega) \\ \tilde{P}_{21}(\omega) & \tilde{P}_{22}(\omega) & & \tilde{P}_{2N_{\text{FT},x}}(\omega) \\ \vdots & & \ddots & \vdots \\ \tilde{P}_{N_{\text{FT},\phi}1}(\omega) & \tilde{P}_{N_{\text{FT},\phi}2}(\omega) & \cdots & \tilde{P}_{N_{\text{FT},\phi}N_{\text{FT},x}}(\omega) \end{bmatrix}$$

and

$$\tilde{\mathbf{D}}(\omega) = \begin{bmatrix} \tilde{D}_{11}(\omega) & \tilde{D}_{12}(\omega) & \cdots & \tilde{D}_{1N_{\text{FT},x}}(\omega) \\ \tilde{D}_{21}(\omega) & \tilde{D}_{22}(\omega) & & \tilde{D}_{2N_{\text{FT},x}}(\omega) \\ \vdots & & \ddots & \vdots \\ \tilde{D}_{N_{\text{FT},\phi}1}(\omega) & \tilde{D}_{N_{\text{FT},\phi}2}(\omega) & \cdots & \tilde{D}_{N_{\text{FT},\phi}N_{\text{FT},x}}(\omega) \end{bmatrix}.$$

According to Eq. (6.18),  $\tilde{D}_{ij}(\omega)$ , the  $(i, j)$ -th element of  $\tilde{\mathbf{D}}(\omega)$ , can be calculated as

$$\tilde{D}_{ij}(\omega) = \tilde{F}_{ij}(\omega)\tilde{P}_{ij}(\omega), \quad (6.41)$$

where

$$\tilde{F}_{ij}(\omega) = -\sqrt{k^2 - k_{x,j}^2} \frac{R_m H_{n_i}^{(1)'}(\sqrt{k^2 - k_{x,j}^2} R_m)}{R_s H_{n_i}^{(1)}(\sqrt{k^2 - k_{x,j}^2} R_s)}. \quad (6.42)$$

Here,  $n_i$  and  $k_{x,j}$  denotes the order and spatial frequency in the  $(i, j)$ -th bin. Equation (6.42) is the WFR filter for cylindrical arrays with rigid baffle in the cylindrical harmonic domain. Therefore, the WFR filter can be designed as a two-dimensional FIR filter in the spatial domain. When the WFR filter in the spatial domain is denoted as a matrix  $\mathbf{F}(\omega)$ , the size of  $\mathbf{F}(\omega)$  is trimmed so that reproduction accuracy is sufficient. This filter size is denoted as  $K_\phi \times K_x$ . Equation (6.41) is regarded as the convolution of  $\mathbf{F}(\omega)$  and  $\mathbf{P}(\omega)$ . Note that the convolution with respect to  $\phi$  is circular and that with respect to  $x$  is linear. Therefore,  $N_{\text{FT},\phi}$  is equal to  $M_\phi$ , and  $N_{\text{FT},x}$  is larger or equal to  $M_x + K_x - 1$ . The proper size of zero-padding with respect to  $x$  is necessary at the cylindrical DFT.

The spatial aliasing artifact originating from the discretization of  $x$  arises when the received sound field includes the components of  $|k_x| > \pi/\Delta x$ . The discretization of  $\phi$  limits the maximum order of  $n$ , which is denoted as  $N$  and determined as  $N = \lfloor M_\phi/2 \rfloor$ . Here,  $\lfloor \cdot \rfloor$  is the floor function. The upper bound of  $N$  leads to limited reproduction region in the neighboring region of the central axis of the cylindrical array. This property of the proposed method is similar to that of sound field reproduction methods using spherical or circular array that the reproduction region is limited to the neighboring region of the center of the array [37–40].

The evanescent wave components of the received sound field are represented as:

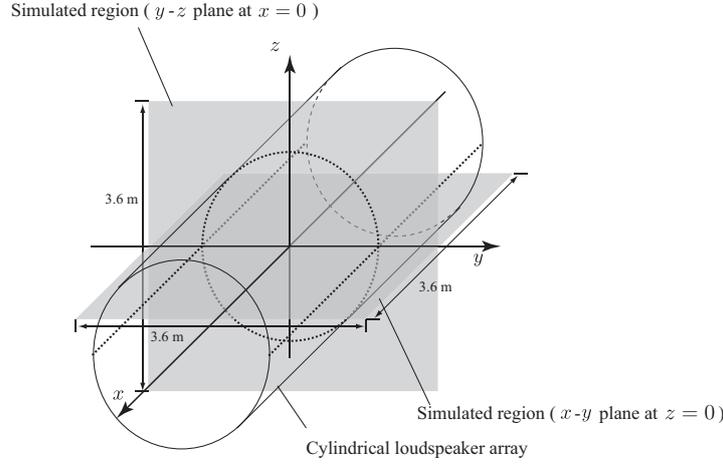
$$|n| > |kR_m| \quad \text{and} \quad |k_x| > |k|. \quad (6.43)$$

When the radius of the microphone array is smaller than that of the loudspeaker array, i.e.,  $R_m < R_s$ , the WFR filter becomes significantly unstable because inverse propagation of the received sound field is included as in Chapter 4. Therefore, a filter in the cylindrical harmonic domain to reduce the evanescent wave components is necessary. We employ a simple rectangular window function in the cylindrical harmonic domain,  $\tilde{W}_{ij}(\omega)$ , that reduces the components of Eq. (6.43) as:

$$\tilde{W}_{ij}(\omega) = \begin{cases} 1 & |n_i| \leq |kR_m| \text{ and } |k_{x,j}| \leq |k| \\ 0 & \text{otherwise} \end{cases} \quad (6.44)$$

This window function is combined with the WFR filter as  $\tilde{W}_{ij}(\omega) \times \tilde{F}_{ij}(\omega)$ . It should be noted that this multiplication of Eq. (6.44) also limits the maximum available order.

The finite array length also produces artifacts as in the planar or linear array case (c.f. Chapter 5). Therefore, a spatial tapering window at axial positions is necessary. The rigid cylindrical baffle needs to be sufficiently long compared to the microphone array length because the WFR filter Eq. (6.42) is derived under the assumption of the infinitely-long cylindrical baffle.



**Fig. 6.4.** Target sound field in target area.

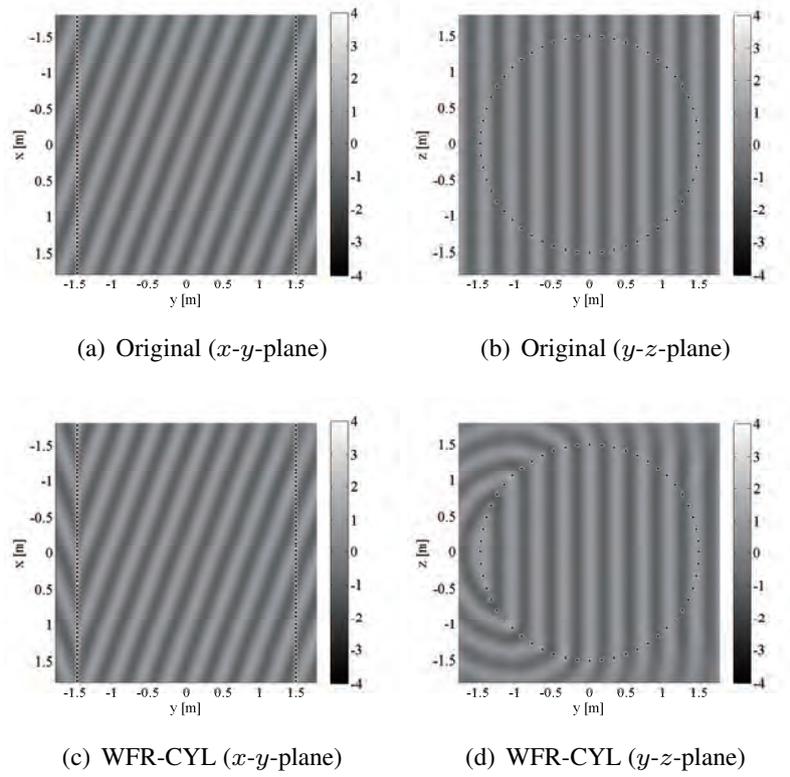
## 6.4 Experiments

Numerical simulations were conducted under the free-field condition. First, simulation results of the proposed method in the case of large  $M_\phi$  and  $M_x$  are shown to demonstrate the properties of the proposed method. Second, the results when  $M_\phi$  was smaller than  $M_x$  are shown as a more practical situation for with-height reproduction. The proposed WFR filtering method for cylindrical arrays is compared with the WFR filtering method for planar arrays presented in Chapter 5. These two methods are denoted as WFR-CYL and WFR-PLN, respectively.

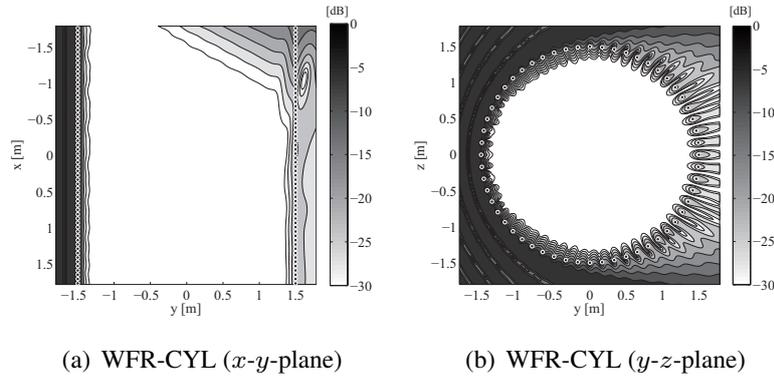
### 6.4.1 Simulations using large number of array elements

The number of elements in the cylindrical microphone and loudspeaker arrays was set as  $M_\phi = 56$  and  $M_x = 128$ . The length of the WFR filter was set as  $K_\phi = 56$  and  $K_x = 128$ , and  $N_{FT,\phi}$  and  $N_{FT,x}$  were 56 and 256, respectively. Both the radii of the microphone and loudspeaker arrays,  $R_m$  and  $R_s$ , were set as 1.5 m. The interval of the array elements at the axial positions was 6 cm; therefore, the length of the cylindrical arrays was 7.68 m. The directivity of the array elements was assumed to be omni-directional. The Tukey window function was applied as a spatial window whose sides tapered by 10%. The received signals of the microphone array was generated in the cylindrical harmonic domain because simulation of scattering sound fields in the spatial domain is time consuming; therefore, the rigid cylindrical baffle was simulated as infinitely long. As shown in Fig. 6.4, the sound pressure distribution was simulated in  $3.6 \times 3.6$  m regions on the  $x$ - $y$ -plane at  $z = 0$  and  $y$ - $z$ -plane at  $x = 0$  at intervals of 1.5 cm. The amplitudes of the reproduced sound fields were normalized at the origin. The sampling frequency was 48 kHz.

We evaluated the reproduced sound fields by using the time averaged squared error at every



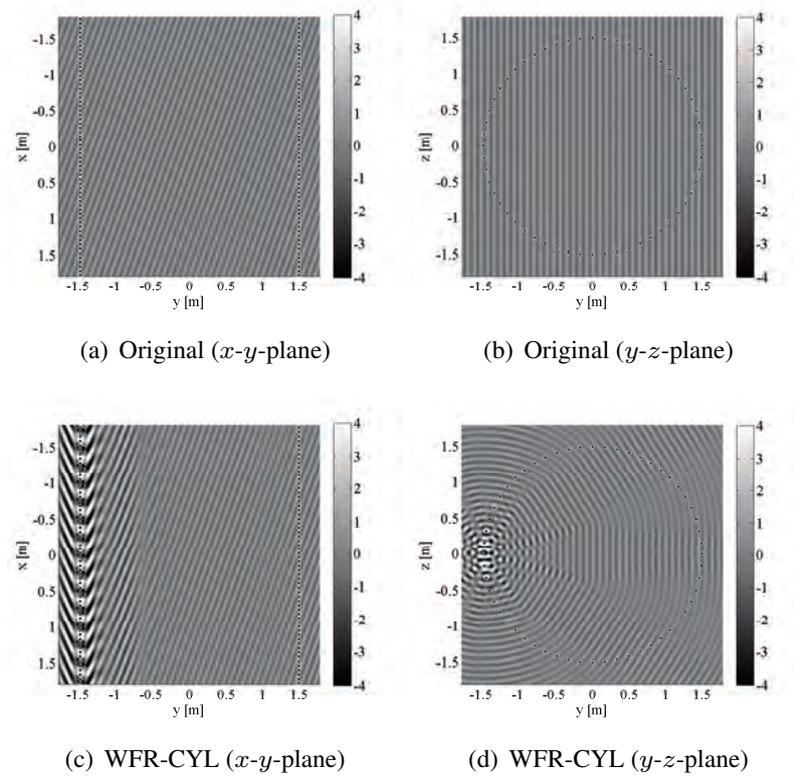
**Fig. 6.5.** Original and reproduced sound pressure distribution when source signal was 1-kHz sinusoidal wave



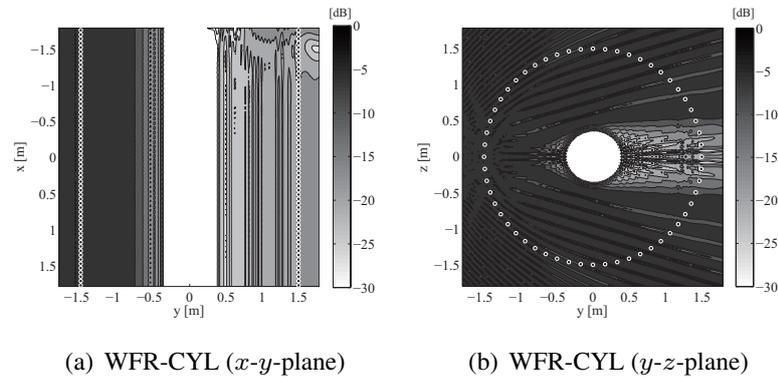
**Fig. 6.6.** Time averaged squared error of sound pressure distribution when source signal was 1-kHz sinusoidal wave.

simulated discrete position  $(x_i, y_j)$  defined as Eq. (4.12). The total number of time samples was set at 480 samples, i.e., 10 ms.

Figure 6.5 shows the simulation results of the original and reproduced sound pressure distributions when the primary source was a single plane wave and its arrival direction was  $(\theta_{pw}, \phi_{pw}) = (20 \text{ deg}, 0 \text{ deg})$ . These angles were defined so that the unit directional vector of the

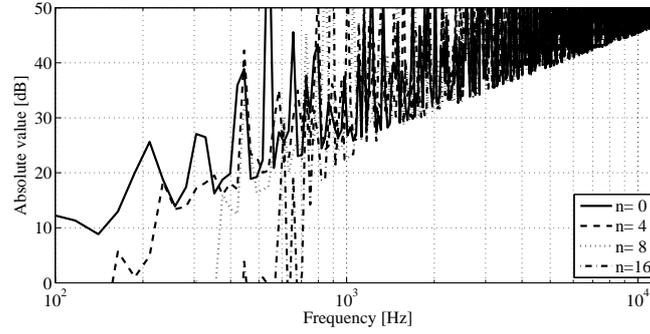


**Fig. 6.7.** Original and reproduced sound pressure distribution when source signal was 4-kHz sinusoidal wave

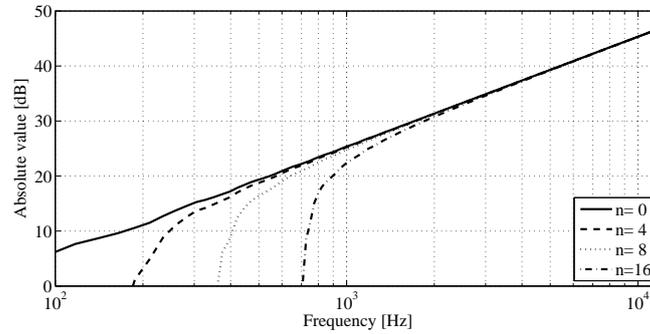


**Fig. 6.8.** Time averaged squared error of sound pressure distribution when source signal was 4-kHz sinusoidal wave.

plane wave becomes  $(\cos \theta_{pw}, \phi_{pw}, \sin \theta_{pw})$  in the cylindrical coordinates. The source signal was 1-kHz sinusoidal wave. Figure 6.6 shows the time-averaged squared error distributions of the reproduced sound field. Because this experimental condition does not arise the spatial aliasing errors, the reproduction accuracy inside the cylindrical loudspeaker array was distinctly high. Small error on the region of  $x < -1.0$  m and  $y > -0.5$  m at  $z = 0$  was caused by truncation of



(a) Open cylinder



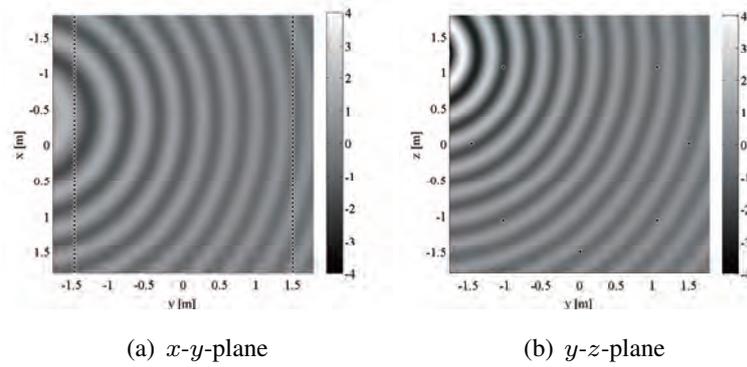
(b) Rigid cylinder

**Fig. 6.9.** Amplitude of WFR filter for open and rigid cylinders in cylindrical harmonic domain

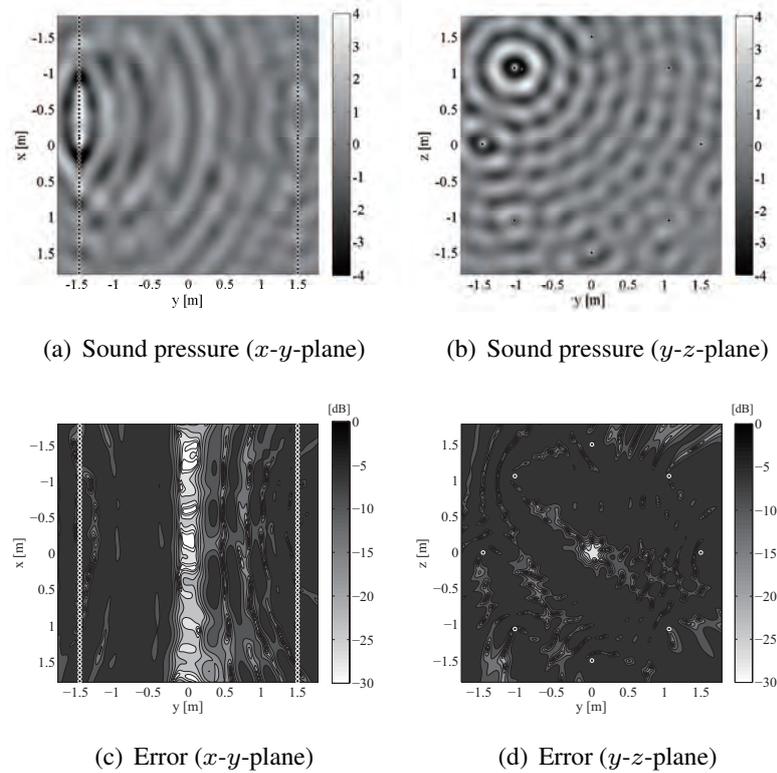
the array length.

Figures 6.5 and 6.8 show the simulation results when the source signal was a 4-kHz sinusoidal wave. The other conditions were the same as those of Figs. 6.5 and 6.6. Because the maximum order  $N$  was limited to 28, the high reproduction accuracy region was limited to the neighboring region of the central axis of the cylindrical array. The higher frequency of the source signal is, the smaller the radius of the high reproduction accuracy region becomes. Therefore, the reproduction accuracy along the central axis of the cylindrical array was preserved at broad frequency band. When the components of  $|k_x| > \pi/\Delta x$  are included in the cylindrical harmonic coefficients of the received signals, the reproduction accuracy on this region cannot be maintained.

Figures 6.9 (a) and (b) respectively plot the amplitude of the WFR filter for open and rigid cylinder, i.e., the filter designed based on Eqs. (6.7) and (6.18), in the cylindrical harmonic domain. The amplitudes for  $k_x = 0$  and  $n = 0, 4, 8, 16$  are only shown. Note that these filters do not depend on the input signals. As shown in Fig. 6.9 (a), the WFR filter for open cylinder diverged at many frequencies. This instability originates from the Bessel function of the denominator of Eq. (6.7). The WFR filter for rigid cylinder was much more stable compared to that for open cylinder.



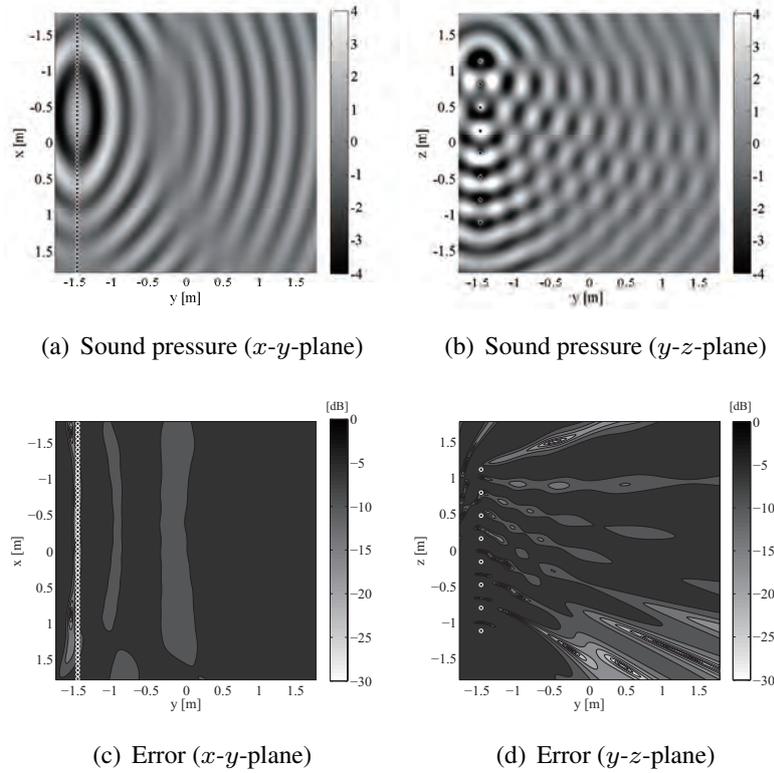
**Fig. 6.10.** Original sound pressure distribution for 1-kHz sinusoidal wave case. Point source was located at (2.5 m, 150 deg, -4.0 m).



**Fig. 6.11.** Simulation results of sound pressure and time-averaged squared error distribution of WFR-CYL for 1-kHz sinusoidal wave case.

### 6.4.2 Simulations using multiple linear arrays

The simulation results in the previous section suggest that the reproduction accuracy in the neighboring region of the central axis of the cylindrical array is preserved at broad frequency band.

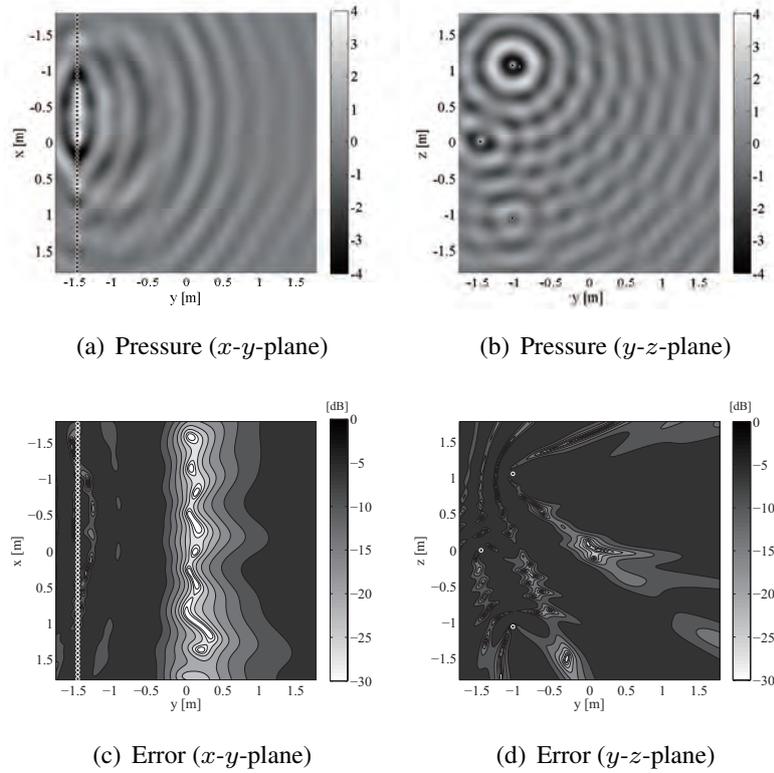


**Fig. 6.12.** Simulation results of sound pressure and time-averaged squared error distribution of WFR-PLN for 1-kHz sinusoidal wave case.

Therefore, a similar effect is expected to arise even when the number of the array elements at the angular positions is small. We compared the WFR-CYL and WFR-PLN when the number of array elements at elevation angle is small. This situation may be more practical and efficient for with-height reproduction. The WFR-CYL and WFR-PLN are compared under the condition that the number of the array elements are the same; therefore, the array configuration and the filter for SP-DS conversion were the only differences in these two methods.

In the proposed WFR-CYL, the number of array elements was set as  $M_\phi = 8$  and  $M_x = 64$ . The filter length of the WFR-CYL was set as  $K_\phi = 8$  and  $K_x = 64$ , and  $N_{\text{FT},\phi}$  and  $N_{\text{FT},x}$  were 8 and 128, respectively. The radii of the microphone and loudspeaker arrays,  $R_m$  and  $R_s$ , were set as 0.22 and 1.5 m, respectively. The interval of the array elements at the axial positions was 6 cm; therefore, the length of the cylindrical arrays was 3.84 m.

In the WFR-PLN, the number of array elements at the vertical and horizontal positions ( $M_z$  and  $M_x$ ) was set as the same number as the WFR-CYL, i.e.,  $M_z = 8$  and  $M_x = 64$ . The interval of the array elements were 32 cm at the positions of  $z$  and 6 cm at the positions of  $x$ ; therefore, the arrays were  $2.56 \times 3.84$  m in size. The planar arrays were set on the  $x$ - $z$ -plane at  $y = -1.5$  m. The WFR filter for the WFR-PLN becomes a two-dimensional spatial filter and its size was set as  $8 \times 64$  taps. The filter was applied in the spatial frequency domain; therefore, the spatial DFT size



**Fig. 6.13.** Reproduced sound pressure and time-averaged squared error distributions

was  $16 \times 128$ . A two-dimensional Tukey window whose sides tapered by 10% was applied as a spatial tapering window.

Figure 6.10 show the original sound pressure distribution when a point source as a primary source was located at  $(r, \phi, x) = (2.5 \text{ m}, 150 \text{ deg}, -4.0 \text{ m})$  in the source area. Figs. 6.11 and 6.12 show the reproduced sound pressure and time-averaged squared error distributions. The source signal was a 1-kHz sinusoidal wave. As shown in Fig. 6.11, the reproduction accuracy of the proposed WFR-CYL was distinctly high in the neighboring region along the central axis of the cylindrical loudspeaker array. In contrast, the reproduction accuracy of the WFR-PLN was low in the entire region. This reproduction error comes from the larger intervals of the microphone and loudspeaker arrays in the positions of  $z$ , i.e., spatial aliasing. Even though it is possible to reduce this artifact by bringing these intervals closer, the sound field captured by the microphone array becomes smaller when the number of elements at the positions of  $z$  is small.

### 6.4.3 Truncation of loudspeaker array at angular positions

In the WFR-CYL, the loudspeakers in the direction where primary sources do not exist, which were the loudspeakers at  $-90 \leq \phi \leq 90 \text{ deg}$  in the experimental condition of the previous section, can be removed for approximate reproduction. Figure 6.13 shows the reproduced sound pressure

and time-averaged squared error distributions when the linear loudspeaker arrays at  $\phi = 135, 180,$  and  $-135$  deg were only used. The number of microphones was  $8 \times 64$ . The reproduction accuracy on the neighboring region along the  $x$ -axis was still preserved. The effect of this removing largely depends on the position of the primary sources. However, this approximation strategy may be useful when the positions of the primary sources are limited.

## 6.5 Conclusions

We proposed an SP-DS conversion method using cylindrical array configurations of microphones and loudspeakers. The proposed method is based on the WFR filter that is analytically derived in the cylindrical harmonic domain. The WFR filter is stabilized by introducing a model in which microphones are mounted on a cylindrical baffle. The proposed method has advantages that reproduction accuracy in the neighboring region along the central axis of the cylindrical array is preserved at a broad frequency band and the width of that region depends on the maximum order of the cylindrical harmonic coefficients. This method is also extended to the SP-DS conversion method using circular arrays of microphones and loudspeakers. Numerical simulations were conducted to compare the WFR filtering method for cylindrical arrays and that for planar arrays. The reproduction accuracy of the proposed WFR filtering for cylindrical arrays was distinctly high on the neighboring region along the central axis of the cylinder. In contrast, that of the WFR filtering for planar arrays was low in the entire region due to spatial aliasing. Additionally, the loudspeakers in the direction where the primary sources do not exist can be removed for approximate reproduction. Therefore, the proposed method is efficient for reproducing with-height sound field when lower resolution at an elevation angle is sufficient.

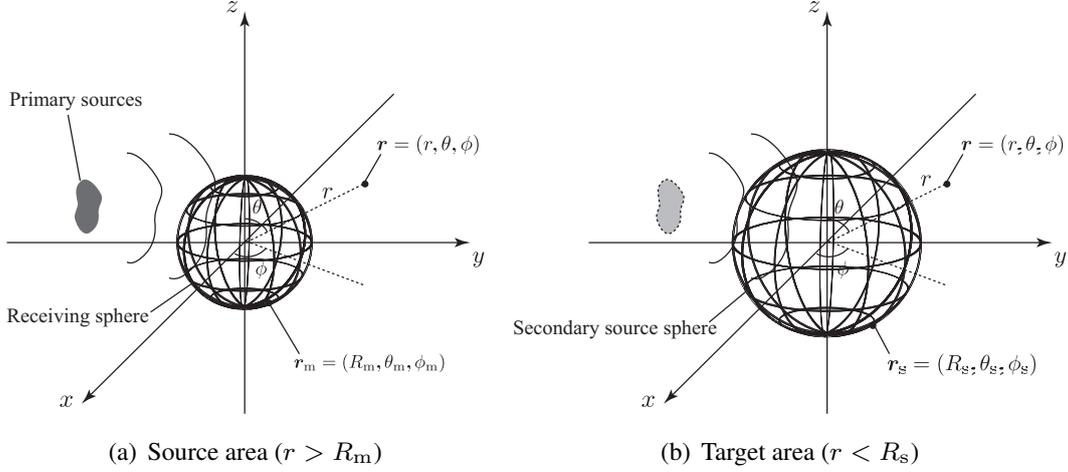
## Chapter 7

# Wave Field Reconstruction Filter for Spherical Arrays of Microphones and Loudspeakers

By using spherical or circular arrays, it is possible to reproduce a sound field that listeners are surrounded by sounds. Similar concept as the WFR filtering method for planar, linear, and cylindrical arrays is applied to spherical and circular array configurations for SP-DS conversion. The HOA presented in Chapter 3 can be applied as an SP-DS conversion for the spherical arrays. However, the two stages of the signal conversion, encoding and decoding processes, may not be efficient for real-time recording and reproducing systems even though such signal conversion scheme have scalability of the reproduction systems. The WFR filtering method for spherical arrays makes it possible to convert signals in a single convolution of the WFR filter and is essentially equivalent to the HOA. In circular array configurations, because spherical harmonic coefficients of the desired sound field are unknown, the HOA cannot be directly applied. The WFR filter for circular arrays can be applied as an SP-DS conversion and is derived under the assumption that the captured sound field is two-dimensional.

### 7.1 WFR equation for spherical distributions of receivers and secondary sources

The WFR equation that relates sound pressure distribution obtained by receivers to driving signals of secondary sources is derived in the spherical coordinates. As shown in Fig. 7.1, continuous spherical distributions of receivers and secondary sources are arranged in the source and target areas, respectively, with the center at the origin. These spherical distributions are assumed to be acoustically transparent (open spheres). The radii of the receiving and secondary source spheres



**Fig. 7.1.** Spherical distributions of receivers and secondary sources.

are  $R_m$  and  $R_s$ , respectively. It is assumed that  $R_m \leq R_s$  and the sound field inside the secondary source sphere is reproduced to coincide with that inside the receiving sphere. The position vector is denoted as  $\mathbf{r} = (r, \theta, \phi)$  in the spherical coordinates, which corresponds to the Cartesian coordinates as follows:

$$\begin{aligned} x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta. \end{aligned} \quad (7.1)$$

The position vectors on the receiving and secondary source sphere are denoted as  $\mathbf{r}_m = (R_m, \theta_m, \phi_m)$  and  $\mathbf{r}_s = (R_s, \theta_s, \phi_s)$ , respectively. The sound pressure at  $\mathbf{r}_m$  and driving signal of the secondary source at  $\mathbf{r}_s$  of the temporal frequency  $\omega$  are denoted as  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  and  $D(\mathbf{r}_s, \omega)$ , respectively.

### 7.1.1 Synthesized sound field

The synthesized sound field created by secondary sources in the target area is denoted as  $P_{\text{syn}}(\mathbf{r}, \omega)$ . When the transfer function from  $\mathbf{r}_s$  to  $\mathbf{r}$  is denoted as  $G(\mathbf{r} - \mathbf{r}_s, \omega)$ ,  $P_{\text{syn}}(\mathbf{r}, \omega)$  is described by using spherical convolution form as [35, 86, 89]:

$$P_{\text{syn}}(\mathbf{r}, \omega) = \int_{u \in SO(3)} D(u^{-1}\mathbf{r}_s, \omega) G(\mathbf{r} - u\boldsymbol{\eta}, \omega) R_s^2 du, \quad (7.2)$$

where  $u$  is a rotation operation, and  $\boldsymbol{\eta} = (0, 0, R_s)$  denotes the north pole of the spherical surface in the Cartesian coordinates. Therefore, the spherical convolution theorem, Eq. (A.21) in the

Appendix, can be applied as:

$$\tilde{P}_{\text{syn},n}^m(r, \omega) = 2\pi R_s^2 \sqrt{\frac{4\pi}{2n+1}} \tilde{D}_n^m(\omega) \tilde{G}_n^0(r - R_s, \omega), \quad (7.3)$$

where  $n$  and  $m$  are the orders. The variables in the spectral domain, i.e., spherical harmonic domain [21], is represented by a tilde. When the desired sound field, which is denoted as  $P_{\text{des}}(\mathbf{r}, \omega)$ , is known,  $D(\mathbf{r}_s, \omega)$  can be calculated by using Eq. (7.3) because  $P_{\text{syn}}(\cdot)$  needs to equate to  $P_{\text{des}}(\cdot)$ . However, in this context, the sound pressure distribution on the receiving sphere is only known. The desired sound field  $P_{\text{des}}(\mathbf{r}, \omega)$  needs to be estimated from  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$ .

### 7.1.2 Derivation of WFR equation for spherical receivers

When the receiving sphere is assumed to be acoustically transparent (open sphere), the desired sound field,  $P_{\text{des}}(\cdot)$ , correspond to the internal sound field of the receiving sphere. Therefore, based on the Fourier acoustics [21], the desired sound field can be described in the spherical harmonic domain as:

$$\tilde{P}_{\text{des},n}^m(r, \omega) = \tilde{P}_{\text{rcv},n}^m(R_m, \omega) \frac{j_n(kr)}{j_n(kR_m)}. \quad (7.4)$$

Because the synthesized and desired sound fields are to be matched, the WFR equation can be derived by simultaneously solving Eqs. (7.3) and (7.4) as

$$\tilde{D}_n^m(\omega) = \sqrt{\frac{2n+1}{4\pi}} \frac{1}{2\pi R_s^2 \tilde{G}_n^0(r - R_s, \omega)} \frac{j_n(kr)}{j_n(kR_m)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega). \quad (7.5)$$

For simplicity, we assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  can be approximated as monopole, Eq. (3.2). The spherical harmonic transform of Eq. (3.2),  $\tilde{G}_{3D,n}^0(r - R_s, \omega)$ , is analytically derived as Eq. (A.33) in the Appendix. Substituting Eq. (A.33) into Eq. (7.5), the WFR equation is derived as

$$\tilde{D}_n^m(\omega) = \frac{1}{2\pi j k R_s^2 h_n^{(1)}(kR_s) j_n(kR_m)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega). \quad (7.6)$$

Only the sound pressure distribution on the receiving sphere is required to calculate Eq. (7.6). However, the denominator of Eq. (7.6) includes  $j_n(kR_m)$ , which is a real valued function that have many zeros. This fact means that  $\tilde{D}_n^m(\omega)$  diverges at the zeros of  $j_n(kR_m)$ . This instability originates from the non-uniqueness at the forbidden frequencies [55].

### 7.1.3 Derivation of WFR equation for spherical receivers with rigid baffle

In order to avoid the instability of Eq. (7.6), we introduce a spherical baffle in the source area as in the cylindrical case. The reflection characteristics on the surface of the baffle is assumed to be known.

It is assumed that the acoustically rigid spherical baffle (rigid sphere) of radius  $R_b$  ( $\leq R_m$ ) is set with the center at the origin in the source area. In this case, the effect of the rigid sphere needs to be removed from the received sound field to obtain the desired sound field. The sound field in the source area,  $P_{\text{rcv}}(\mathbf{r}, \omega)$ , can be represented as the summation of the incident and scattering sound fields,  $P_{\text{inc}}(\mathbf{r}, \omega)$  and  $P_{\text{sct}}(\mathbf{r}, \omega)$ , as:

$$P_{\text{rcv}}(\mathbf{r}, \omega) = P_{\text{inc}}(\mathbf{r}, \omega) + P_{\text{sct}}(\mathbf{r}, \omega). \quad (7.7)$$

The sound pressure gradient on the surface of the rigid baffle becomes zero as:

$$\left. \frac{\partial}{\partial r} \{P_{\text{inc}}(\mathbf{r}, \omega) + P_{\text{sct}}(\mathbf{r}, \omega)\} \right|_{r=R_b} = 0 \quad (7.8)$$

Here,  $P_{\text{inc}}(\cdot)$  and  $P_{\text{sct}}(\cdot)$  can be represented in the spherical harmonic domain as [21]:

$$P_{\text{inc}}(\mathbf{r}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \check{P}_{\text{inc},n}^m(\omega) j_n(kr) Y_n^m(\theta, \phi) \quad (7.9)$$

and

$$P_{\text{sct}}(\mathbf{r}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \check{P}_{\text{sct},n}^m(\omega) h_n^{(1)}(kr) Y_n^m(\theta, \phi). \quad (7.10)$$

The variables excluding radial component of the spherical harmonic coefficients are represented by a breve, i.e.,  $\check{P}_{\text{inc},n}(r, k_x, \omega) = \check{P}_{\text{inc},n}(k_x, \omega) j_n(kr)$ . By substituting Eqs. (7.9) and (7.10) into Eq. (7.8),

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \check{P}_{\text{inc},n}^m(\omega) k j_n'(kR_b) Y_n^m(\theta, \phi) + \sum_{n=0}^{\infty} \sum_{m=-n}^n \check{P}_{\text{sct},n}^m(\omega) k h_n^{(1)'}(kR_b) Y_n^m(\theta, \phi) = 0 \quad (7.11)$$

Therefore, the following relation between the incident and scattering sound fields can be derived:

$$\check{P}_{\text{sct},n}^m(\omega) = -\frac{j_n'(kR_b)}{h_n^{(1)'}(kR_b)} \check{P}_{\text{inc},n}^m(\omega) \quad (7.12)$$

From Eqs. (7.7), (7.9), (7.10), and (7.12), the sound field in the source area,  $P_{\text{rcv}}(\mathbf{r}, \omega)$ , can be described by using  $\check{P}_{\text{inc},n}^m(\omega)$  as:

$$P_{\text{rcv}}(\mathbf{r}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left( j_n(kr) - \frac{j_n'(kR_b)}{h_n^{(1)'}(kR_b)} h_n^{(1)}(kr) \right) \check{P}_{\text{inc},n}^m(\omega) Y_n^m(\theta, \phi). \quad (7.13)$$

Therefore,  $\check{P}_{\text{inc},n}^m(\omega)$  is represented by using  $\check{P}_{\text{rcv},n}^m(r, \omega)$  as

$$\check{P}_{\text{inc},n}^m(\omega) = \frac{h_n^{(1)'}(kR_b)}{j_n(kr) h_n^{(1)'}(kR_b) - j_n'(kR_b) h_n^{(1)}(kr)} \check{P}_{\text{rcv},n}^m(r, \omega) \quad (7.14)$$

The desired sound field,  $\tilde{P}_{\text{des}}(\cdot)$ , corresponds to the incident sound field,  $P_{\text{inc}}(\cdot)$ . Therefore,  $P_{\text{des}}(\cdot)$  can be described by using the spherical harmonic coefficients on the receiving sphere,  $\tilde{P}_{\text{rcv},n}^m(R_m, \omega)$ , from Eq. (7.14) as:

$$\begin{aligned}\tilde{P}_{\text{des},n}^m(r, \omega) &= \check{P}_{\text{inc},n}^m(\omega) j_n(kr) \\ &= \frac{j_n(kr) h_n^{(1)'}(kR_b)}{j_n(kR_m) h_n^{(1)'}(kR_b) - j_n'(kR_b) h_n^{(1)}(kR_m)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega)\end{aligned}\quad (7.15)$$

Similar to the open sphere case, the WFR equation can be derived by simultaneously solving Eqs. (7.3) and (7.15), as:

$$\begin{aligned}\tilde{D}_n^m(\omega) &= \frac{1}{2\pi j k R_s^2 h_n^{(1)}(kR_s) j_n(kR_m)} \\ &\quad \cdot \frac{j_n(kR_m) h_n^{(1)'}(kR_b)}{j_n(kR_m) h_n^{(1)'}(kR_b) - j_n'(kR_b) h_n^{(1)}(kR_m)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega) \\ &= \frac{h_n^{(1)'}(kR_b)}{2\pi j k R_s^2 h_n^{(1)}(kR_s) \left( j_n(kR_m) h_n^{(1)'}(kR_b) - j_n'(kR_b) h_n^{(1)}(kR_m) \right)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega)\end{aligned}\quad (7.16)$$

Again, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. When the radii of the receivers and rigid baffle are the same, i.e.,  $R_b = R_m$ , Eq. (7.16) can be simplified as:

$$\begin{aligned}\tilde{D}_n^m(\omega) &= \frac{h_n^{(1)'}(kR_m)}{2\pi j k R_s^2 h_n^{(1)}(kR_s) \left( j_n(kR_m) h_n^{(1)'}(kR_m) - j_n'(kR_m) h_n^{(1)}(kR_m) \right)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega) \\ &= -\frac{kR_m^2 h_n^{(1)'}(kR_m)}{2\pi R_s^2 h_n^{(1)}(kR_s)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega)\end{aligned}\quad (7.17)$$

Here, Eq. (2.46) is used to derive Eq. (7.17). The spherical Bessel function of the denominator of Eq. (7.6) is eliminated in Eq. (7.17). Therefore, Eq. (7.17) is much more stable than Eq. (7.6). The driving signals can be stably obtained from the sound pressure distribution on the receiving sphere in the rigid sphere case.

#### 7.1.4 Derivation of WFR equation for spherical receivers with soft baffle

As an another example of a baffle, we consider a acoustically soft baffle. Practically, such a baffle is difficult to implement. It is assumed that the acoustically soft spherical baffle of radius  $R_b$  ( $\leq R_m$ ) is set with the center at the origin in the source area. Similar to the rigid baffle case, the sound field in the source area can be represented as the summation of the incident and scattering sound fields,  $P_{\text{inc}}(\mathbf{r}, \omega)$  and  $P_{\text{sct}}(\mathbf{r}, \omega)$ , as Eq. (7.7). The sound pressure on the surface of the soft

baffle becomes zero as:

$$\{P_{\text{inc}}(\mathbf{r}, \omega) + P_{\text{sct}}(\mathbf{r}, \omega)\}|_{r=R_b} = 0 \quad (7.18)$$

Here,  $P_{\text{inc}}(\cdot)$  and  $P_{\text{ext}}(\cdot)$  are described in the spherical harmonic domain as Eqs. (7.9) and (7.10). By substituting Eqs. (7.9) and (7.10) into Eq. (7.18),

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \check{P}_{\text{inc},n}^m(\omega) j_n(kR_b) Y_n^m(\theta, \phi) + \sum_{n=0}^{\infty} \sum_{m=-n}^n \check{P}_{\text{sct},n}^m(\omega) h_n^{(1)}(kR_b) Y_n^m(\theta, \phi) = 0 \quad (7.19)$$

Therefore, the following relation between the incident and scattering sound fields can be derived:

$$\check{P}_{\text{sct},n}^m(\omega) = -\frac{j_n(kR_b)}{h_n^{(1)}(kR_b)} \check{P}_{\text{inc},n}^m(\omega). \quad (7.20)$$

From Eqs. (7.7), (7.9), (7.10), and (7.20), the sound field in the source area,  $P_{\text{rcv}}(\mathbf{r}, \omega)$ , can be described by using  $\check{P}_{\text{inc},n}^m(\omega)$  as:

$$P_{\text{rcv}}(\mathbf{r}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left( j_n(kr) - \frac{j_n(kR_b)}{h_n^{(1)}(kR_b)} h_n^{(1)}(kr) \right) \check{P}_{\text{inc},n}^m(\omega). \quad (7.21)$$

Therefore,  $\check{P}_{\text{inc},n}^m(\omega)$  is represented by using  $\tilde{P}_{\text{rcv},n}^m(r, \omega)$  as

$$\check{P}_{\text{inc},n}^m(\omega) = \frac{h_n^{(1)}(kR_b)}{j_n(kr)h_n^{(1)}(kR_b) - j_n(kR_b)h_n^{(1)}(kr)} \tilde{P}_{\text{rcv},n}^m(r, \omega). \quad (7.22)$$

In a similar manner to the rigid spherical baffle case, the WFR equation can be derived as

$$\tilde{D}_n^m(\omega) = \frac{h_n^{(1)}(kR_b)}{2\pi j k R_s^2 h_n^{(1)}(kR_s) \left( j_n(kR_m)h_n^{(1)}(kR_b) - j_n(kR_b)h_n^{(1)}(kR_m) \right)} \tilde{P}_{\text{rcv},n}^m(R_m, \omega) \quad (7.23)$$

Again, it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. The denominator of Eq. (7.23) may not also exhibit zeros when  $R_b$  is smaller than  $R_m$ .

## 7.1.5 Comparison with HOA

The WFR filter for spherical arrays of microphones and loudspeakers with rigid baffle is designed based on Eq. (7.17). In the HOA, the encoding and decoding stages are necessary to convert the received signals of the microphones into the driving signals of the loudspeakers. The WFR filtering method make it possible to convert these signals in a single convolution of the WFR filter

and is essentially equivalent to the HOA. However, it is difficult to calculate spherical convolution of the WFR filter and the received signals of the microphones in an efficient way compared to the other array configuration cases because the spherical harmonic transform includes the associated Legendre transform with respect to  $\theta$ . Furthermore, the required numbers of the microphones and loudspeakers and the computational efficiency of the spherical harmonic transform depend on the sampling manner of the spherical surface. For example, the equiangle sampling requires  $4(N + 1)^2$  samples when  $N$  denotes the maximum order of  $n$  and the FFT algorithm can be used with respect to  $\phi$  [42]. The nearly uniform sampling requires  $(N + 1)^2$  samples at least and the FFT algorithm cannot be used. Although several algorithms of fast spherical harmonic transform are proposed, these are not as efficient as the FFT algorithm [90].

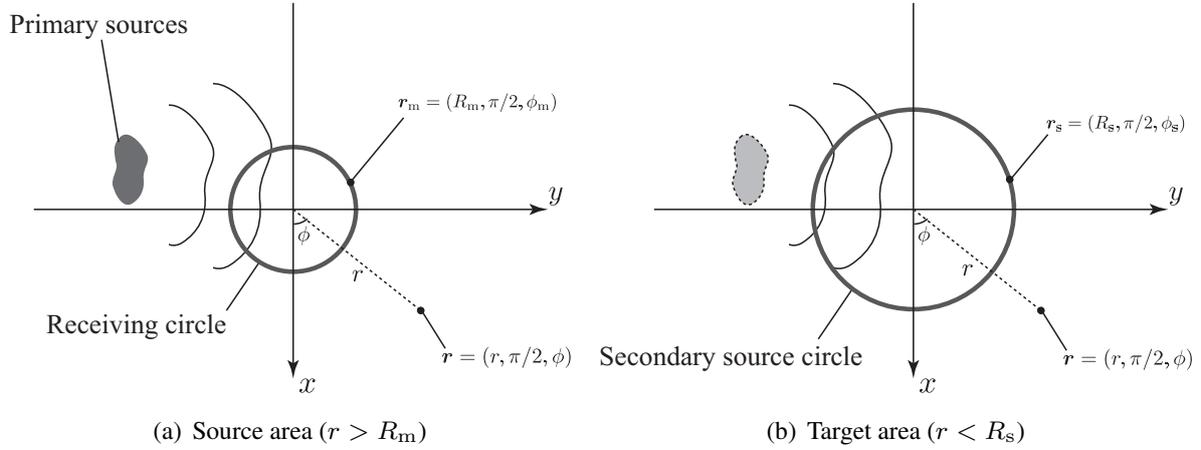
## 7.2 WFR equation for circular distributions of receivers and secondary sources

We consider the WFR equation for circular distributions of receivers and secondary sources based on the WFR equation for spherical distributions. Even though several decoding methods of the HOA for circular loudspeaker array are proposed, most of them assumes that the spherical harmonic coefficients of the desired sound field is known [39,43,91]. When the array configuration of the microphones is circular, it is difficult to obtain the spherical harmonic coefficients. We derive the WFR equation for circular distributions under the assumption that the captured sound field is two-dimensional. Unlike the WFR equation for circular distributions derived in Section 6.2, we consider them in the spherical coordinates.

As shown in Fig. 7.2, continuous circular distributions of receivers and secondary sources are arranged in the source and target areas, respectively, with the center of the circle at the origin. These circular distributions are assumed to be acoustically transparent (open circles). The radii of the receiving and secondary source circle are  $R_m$  and  $R_s$ , respectively. It is assumed that  $R_m \leq R_s$  and the sound field inside the secondary source circle is reproduced to coincide with that inside the receiving circle. The position vectors on the receiving and secondary source circle are denoted as  $\mathbf{r}_m = (R_m, \pi/2, \phi_m)$  and  $\mathbf{r}_s = (R_s, \pi/2, \phi_s)$ , respectively. The sound pressure at  $\mathbf{r}_m$  and driving signal of the secondary source at  $\mathbf{r}_s$  of the temporal frequency  $\omega$  are denoted as  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  and  $D(\mathbf{r}_s, \omega)$ , respectively.

### 7.2.1 Synthesized sound field

The synthesized sound field created by secondary sources in the target area is denoted as  $P_{\text{syn}}(\mathbf{r}, \omega)$ . When the transfer function from  $\mathbf{r}_s$  to  $\mathbf{r}$  is denoted as  $G(\mathbf{r} - \mathbf{r}_s, \omega)$ ,  $P_{\text{syn}}(\mathbf{r}, \omega)$  is



**Fig. 7.2.** Circular distributions of receivers and secondary sources.

described as

$$P_{\text{syn}}(\mathbf{r}, \omega) = \int_0^{2\pi} R_s D(\mathbf{r}_s, \omega) G(\mathbf{r} - \mathbf{r}_s, \omega) d\phi_s. \quad (7.24)$$

Here,  $dS = R_s d\phi_s$  is used. Equation (7.24) can be regarded as the convolution of  $D(\cdot)$  and  $G(\cdot)$  with respect to  $\phi$ . Therefore, the spatial Fourier transform of Eq. (7.24) with respect to  $\phi$  is derived as

$$\tilde{P}_{\text{syn},m} \left( r, \frac{\pi}{2}, \omega \right) = 2\pi R_s \tilde{D}_m(\omega) \tilde{G}_m \left( r - R_s, \frac{\pi}{2}, \omega \right), \quad (7.25)$$

where  $m$  is the order. The variables in the spectral domain, i.e., circular harmonic domain, is represented by a tilde.

### 7.2.2 Derivation of WFR equation for circular receivers

The three-dimensional desired sound field needs to be estimated from the sound pressure distribution on the receiving circle. When it is assumed that the receiving circle is acoustically transparent, the desired sound field corresponds to the sound field in the source area.

$$\begin{aligned} \tilde{P}_{\text{des},n}^m(r, \omega) &= \check{P}_{\text{des},n}^m(\omega) j_n(kr) \\ &= \tilde{P}_{\text{rcv},n}^m(r, \omega) \end{aligned} \quad (7.26)$$

We assume that the desired sound field is invariant to changes in the  $z$ -axis; therefore, the spherical harmonics coefficients of the desired sound field is described as

$$\begin{aligned}
 \tilde{P}_{\text{des},n}^m(r, \omega) &= \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta P_{\text{des}}(r, \theta, \phi, \omega) Y_n^m(\theta, \phi)^* \\
 &= \int_0^\pi \check{P}_{\text{des},m}\left(\frac{\pi}{2}, \omega\right) J_m(kr \sin \theta) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \theta) \sin \theta d\theta \\
 &= \check{P}_{\text{des},m}\left(\frac{\pi}{2}, \omega\right) \frac{2j^n}{j^m} j_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0). \tag{7.27}
 \end{aligned}$$

Here, Eq. (2.47) is used. Therefore, the relation between the spherical and circular harmonic coefficients of the desired sound field can be derived as follows:

$$\check{P}_{\text{des},n}^m(\omega) = \check{P}_{\text{des},m}\left(\frac{\pi}{2}, \omega\right) \frac{2j^n}{j^m} \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0). \tag{7.28}$$

The circular harmonic coefficients on the receiving circle is described as:

$$\begin{aligned}
 \tilde{P}_{\text{rcv},m}\left(R_m, \frac{\pi}{2}, \omega\right) &= \sum_{n=|m|}^{\infty} \tilde{P}_{\text{rcv},n}^m(R_m, \omega) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) \\
 &= \sum_{n=|m|}^{\infty} \check{P}_{\text{des},n}^m(\omega) j_n(kR_m) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) \\
 &= \sum_{n=|m|}^{\infty} \check{P}_{\text{des},m}(\omega) \frac{2j^n}{j^m} j_n(kR_m) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2. \tag{7.29}
 \end{aligned}$$

Therefore, the circular harmonic coefficients of the desired sound field can be described by using those on the receiving circle as

$$\begin{aligned}
 \tilde{P}_{\text{des},m}\left(r, \frac{\pi}{2}, \omega\right) &= \check{P}_{\text{des},m}\left(\frac{\pi}{2}, \omega\right) J_m(kr) \\
 &= \left( \sum_{n=|m|}^{\infty} \frac{2j^n}{j^m} j_n(kR_m) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\
 &\quad \cdot J_m(kr) \tilde{P}_{\text{rcv},m}\left(R_m, \frac{\pi}{2}, \omega\right) \tag{7.30}
 \end{aligned}$$

This equation is only valid when the primary sources are only on the  $x$ - $y$ -plane at  $z = 0$ .

Because the synthesized and desired sound fields are to be matched, the WFR equation can be described by solving Eqs. (7.25) and (7.30) as

$$\begin{aligned}
 \tilde{D}_m(\omega) &= \frac{1}{2\pi R_s \tilde{G}_m(r - R_s, \pi/2, \omega)} \left( \sum_{n=|m|}^{\infty} \frac{2j^n}{j^m} j_n(kR_m) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\
 &\quad \cdot J_m(kr) \tilde{P}_{\text{rcv},m}\left(R_m, \frac{\pi}{2}, \omega\right) \tag{7.31}
 \end{aligned}$$

For simplicity, we assume that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. The representation in the circular harmonic domain of  $G(r - R_s, \pi/2, \phi, \omega)$  is derived as Eq. (A.35) in the Appendix. Substituting Eq. (A.35) into Eq. (7.31), the WFR equation for monopole secondary sources is derived as:

$$\begin{aligned} \tilde{D}_m(\omega) = & \frac{1}{2\pi R_s} \left( \sum_{n=|m|}^{\infty} \frac{2j^n}{j^m} j_n(kR_m) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\ & \cdot \left( \sum_{n=|m|}^{\infty} jk j_n(kr) h_n^{(1)}(kR_s) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} J_m(kr) \tilde{P}_{rcv,m}(R_m, \omega) \end{aligned} \quad (7.32)$$

Because  $r$  is included in Eq. (7.32), the reference radius,  $r = R_{\text{ref}}$ , must be set. The amplitude of the synthesized sound field correspond to that of the desired one only on the reference radius.

When it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has line source characteristics, i.e., two-dimensional Green function, the WFR equation becomes simpler. The line source characteristics is described as Eq. (A.36) in the Appendix. Substituting Eq. (A.36) into Eq. (7.31), the WFR equation for line sources as secondary sources is derived as:

$$\tilde{D}_m(\omega) = \frac{2}{\pi j R_s H_m^{(1)}(kR_s)} \left( \sum_{n=|m|}^{\infty} \frac{2j^n}{j^m} j_n(kR_m) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \tilde{P}_{rcv,m}(R_m, \omega). \quad (7.33)$$

However, the characteristics of actual loudspeakers is closer to monopole source than line source. Based on Eq. (3.11), it is necessary to compensate frequency characteristics of the driving signals by multiplying  $\sqrt{2\pi j R_s/k}$  when monopole sources are approximately employed as an alternative to line sources. Similar to the spherical array configuration case, Eqs. (7.32) and (7.33) may be unstable due to non-uniqueness at forbidden frequencies [55]. This fact is demonstrated in the experiments.

### 7.2.3 Derivation of WFR equation for circular receivers with rigid spherical baffle

Again, we introduce a rigid spherical baffle in the source area to avoid the instability of Eqs. (7.32) and (7.33). When the rigid spherical baffle is set with the center on the origin in the source area, the spherical harmonic coefficients excluding radial components of the incident sound field can be described by using those of the sound field in the source area as Eq. (7.14). As in Eq. (7.29), when it is assumed that the incident sound field is invariant in changes to  $z$ -axis, the spherical harmonic

coefficients of the incident sound field is described as

$$\tilde{P}_{\text{inc},n}^m(r,\omega) = \check{P}_{\text{inc},m}\left(\frac{\pi}{2},\omega\right) \frac{2j^n}{j^m} j_n(kr) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0). \quad (7.34)$$

Therefore, the relation between the spherical and circular harmonic coefficients of the incident sound field can be described as follows.

$$\check{P}_{\text{inc},n}^m(\omega) = \check{P}_{\text{inc},m}\left(\frac{\pi}{2},\omega\right) \frac{2j^n}{j^m} \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0). \quad (7.35)$$

The circular harmonic coefficients on the receiving circle is described as

$$\begin{aligned} & \tilde{P}_{\text{rcv},m}\left(R_m, \frac{\pi}{2}, \omega\right) \\ &= \sum_{n=|m|}^{\infty} \tilde{P}_{\text{rcv},n}^m\left(R_m, \frac{\pi}{2}, \omega\right) \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) \\ &= \sum_{n=|m|}^{\infty} \check{P}_{\text{inc},n}^m\left(\frac{\pi}{2}, \omega\right) \frac{j_n(kR_m)h_n^{(1)'}(kR_b) - j_n'(kR_b)h_n^{(1)}(kR_m)}{h_n^{(1)'}(kR_b)} \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) \\ &= \sum_{n=|m|}^{\infty} \check{P}_{\text{inc},m}\left(\frac{\pi}{2}, \omega\right) \frac{j_n(kR_m)h_n^{(1)'}(kR_b) - j_n'(kR_b)h_n^{(1)}(kR_m)}{h_n^{(1)'}(kR_b)} \\ & \quad \cdot \frac{2j^n}{j^m} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2. \end{aligned} \quad (7.36)$$

Because the desired sound field corresponds to the incident sound field, the circular harmonic coefficients of the desired sound field can be described by using those on the receiving circle as

$$\begin{aligned} & \tilde{P}_{\text{des},m}\left(r, \frac{\pi}{2}, \omega\right) \\ &= \check{P}_{\text{inc},m}\left(\frac{\pi}{2}, \omega\right) J_m(kr) \\ &= \left( \sum_{n=|m|}^{\infty} \frac{j_n(kR_m)h_n^{(1)'}(kR_b) - j_n'(kR_b)h_n^{(1)}(kR_m)}{h_n^{(1)'}(kR_b)} \frac{2j^n}{j^m} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\ & \quad \cdot J_m(kr) \tilde{P}_{\text{rcv},m}\left(R_m, \frac{\pi}{2}, \omega\right). \end{aligned} \quad (7.37)$$

By solving Eqs. (7.25) and (7.37), the WFR equation can be derived as

$$\begin{aligned} \tilde{D}_m(\omega) = & \frac{1}{2\pi R_s} \left( \sum_{n=|m|}^{\infty} \frac{j_n(kR_m)h_n^{(1)'}(kR_b) - j_n'(kR_b)h_n^{(1)}(kR_m)}{h_n^{(1)'}(kR_b)} \frac{2j^n}{j^m} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\ & \cdot \left( \sum_{n=|m|}^{\infty} jk j_n(kr)h_n^{(1)}(kR_s) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} J_m(kr) \tilde{P}_{rcv,m} \left( R_m, \frac{\pi}{2}, \omega \right). \end{aligned} \quad (7.38)$$

Again, we assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole characteristics. Because  $r$  is included in Eq. (7.38), the reference radius,  $r = R_{ref}$ , must be set. When the radii of the receivers and rigid baffle are the same, i.e.,  $R_b = R_m$ , Eq. (7.38) can be simplified as:

$$\begin{aligned} \tilde{D}_m(\omega) = & -\frac{j^m R_m^2 J_m(kR_{ref})}{4\pi R_s} \left( \sum_{n=|m|}^{\infty} \frac{j^n}{h_n^{(1)'}(kR_m)} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\ & \cdot \left( \sum_{n=|m|}^{\infty} j_n(kR_{ref})h_n^{(1)}(kR_s) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \tilde{P}_{rcv,m} \left( R_m, \frac{\pi}{2}, \omega \right). \end{aligned} \quad (7.39)$$

Here, Eq. (2.46) is used.

When it is assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has line source characteristics, the WFR equation becomes simpler. The line source characteristics is represented as Eq. (A.36) in the Appendix. Therefore, the WFR equation for line source characteristics is derived as

$$\begin{aligned} \tilde{D}_m(\omega) = & -\frac{j^m k R_m^2}{\pi R_s H_m^{(1)}(kR_s)} \left( \sum_{n=|m|}^{\infty} \frac{j^n}{h_n^{(1)'}(kR_m)} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\ & \tilde{P}_{rcv,m} \left( R_m, \frac{\pi}{2}, \omega \right) \end{aligned} \quad (7.40)$$

Similar to the open receiving circle case, it is necessary to compensate frequency characteristics of the driving signals by multiplying  $\sqrt{2\pi j R_s / k}$  when monopole sources are approximately employed as an alternative to line sources.

#### 7.2.4 Derivation of WFR equation for circular receivers with soft spherical baffle

As another example of a baffle, we consider a acoustically soft baffle. As in Eq. (7.22), when the soft spherical baffle is set with the center at the origin in the source area, the spherical harmonic

coefficients excluding radial components of the incident sound field can be described by using those of the sound field in the source area as Eq. (7.22). We assumed that the incident sound field is invariant in changes to the  $z$ -axis.

In a similar manner to the rigid spherical baffle case, the WFR equation can be derived as

$$\begin{aligned} \tilde{D}_m(\omega) = & \frac{1}{2\pi R_s} \left( \sum_{n=|m|}^{\infty} \frac{j_n(kR_m)h_n^{(1)}(kR_b) - j_n(kR_b)h_n^{(1)}(kR_m)}{h_n^{(1)}(kR_b)} \frac{2j^n}{j^m} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right) \\ & \cdot \left( \sum_{n=|m|}^{\infty} jk j_n(kr) h_n^{(1)}(kR_s) \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^{m2}(0) \right)^{-1} J_m(kr) \tilde{P}_{rcv,m}(R_m, \omega) \end{aligned} \quad (7.41)$$

Again, we assumed that  $G(\mathbf{r} - \mathbf{r}_s, \omega)$  has monopole source characteristics. Because  $r$  is included in Eq. (7.41), the reference radius,  $r = R_{\text{ref}}$ , must be set.

The WFR equation for line source characteristics is derived as

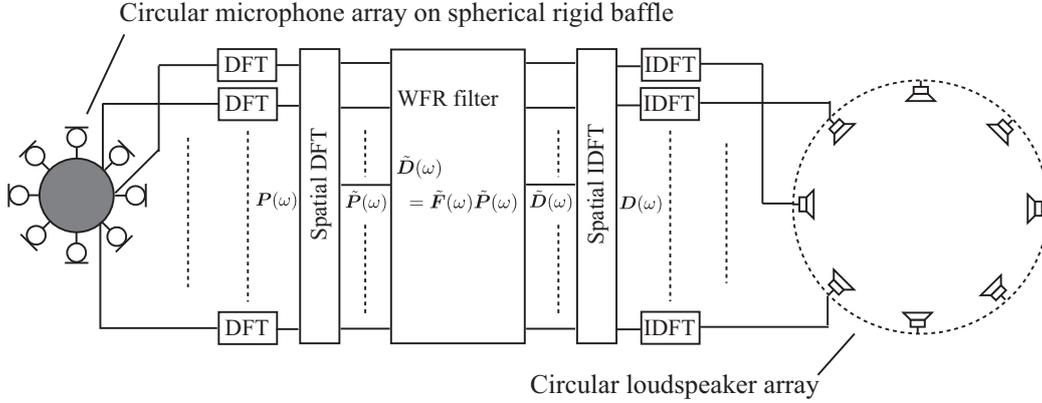
$$\begin{aligned} \tilde{D}_m(\omega) = & \frac{2}{\pi j R_s H_m^{(1)}(kR_s)} \\ & \cdot \left( \sum_{n=|m|}^{\infty} \frac{j_n(kR_m)h_n^{(1)}(kR_b) - j_n(kR_b)h_n^{(1)}(kR_m)}{h_n^{(1)}(kR_b)} \frac{2j^n}{j^m} \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right)^{-1} \\ & \tilde{P}_{rcv,m}\left(R_m, \frac{\pi}{2}, \omega\right) \end{aligned} \quad (7.42)$$

Similar to the open receiving circle case, it is necessary to compensate frequency characteristics of the driving signals by multiplying  $\sqrt{2\pi j R_s/k}$  when monopole sources are approximately employed as an alternative to line sources.

### 7.3 Implementation of WFR filter for circular arrays

In practical implementation, the spherical or circular distributions of the receivers and secondary sources must be discretized. Such a system can provide sound field that listeners are surrounded by sounds. We describe the WFR filter for circular arrays with rigid spherical baffle based on the WFR equation in the spherical harmonic domain.

A block diagram of the system using circular arrays of microphones and loudspeakers with rigid spherical baffle and the WFR filter as a two-dimensional FIR filter is shown in Fig. 7.3. The sound pressure distribution in the source area is obtained by using an equiangularly spaced omni-directional microphone array. A loudspeaker array is arranged in the target area, and each loudspeaker is assumed to be omni-directional. The numbers of microphones and loudspeakers



**Fig. 7.3.** Block diagram of WFR filter for circular arrays.

are assumed to be the same, and denoted as  $M_\phi$ . The microphone array is mounted on a great circle of the rigid spherical baffle. The received signals of the microphone array and the driving signals of the loudspeaker array in the temporal frequency domain are respectively denoted as:

$$\begin{aligned} \mathbf{P} &= [P_1(\omega), \dots, P_{M_\phi}(\omega)]^T \\ \mathbf{D} &= [D_1(\omega), \dots, D_{M_\phi}(\omega)]^T. \end{aligned}$$

When the number of the samples of the spatial DFT is denoted as  $N_{\text{FT}}$ , the circular harmonic spectrum of  $\mathbf{P}(\omega)$  and  $\mathbf{D}(\omega)$  is respectively denoted as:

$$\begin{aligned} \tilde{\mathbf{P}} &= [\tilde{P}_1(\omega), \dots, \tilde{P}_{N_{\text{FT}}}(\omega)]^T \\ \tilde{\mathbf{D}} &= [\tilde{D}_1(\omega), \dots, \tilde{D}_{N_{\text{FT}}}(\omega)]^T, \end{aligned}$$

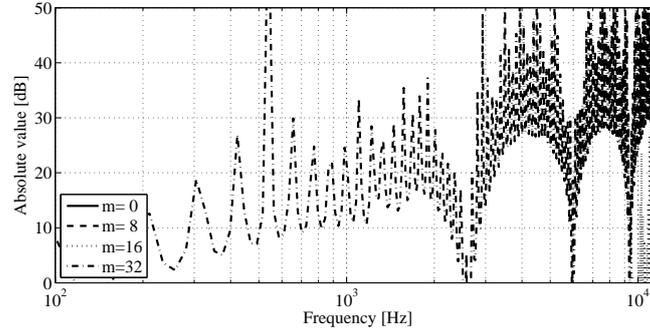
In the circular case, the convolution of the WFR filter becomes circular and a zero-padding is not required in the spatial DFT; therefore,  $N_{\text{FT}}$  is normally set as  $M_\phi$ . According to Eq. (7.39),  $\tilde{D}_i(\omega)$ , the  $i$ -th element of  $\tilde{\mathbf{D}}(\omega)$ , can be calculated as

$$\tilde{D}_i(\omega) = \tilde{F}_i(\omega)\tilde{P}_i(\omega), \quad (7.43)$$

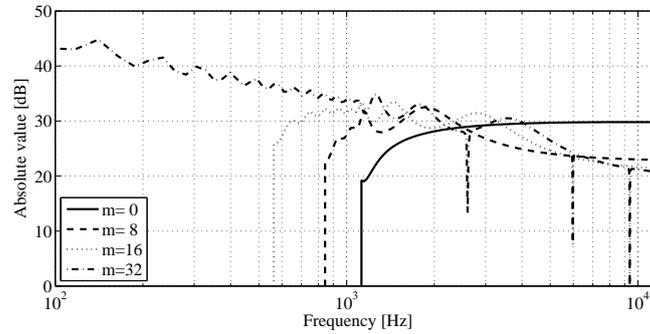
where

$$\begin{aligned} \tilde{F}_i(\omega) &= -\frac{j^{m_i} R_m^2 J_{m_i}(kR_{\text{ref}})}{4\pi R_s} \left( \sum_{n=|m_i|}^{\infty} \frac{j^n}{h_n^{(1)'}(kR_m)} \frac{2n+1}{4\pi} \frac{(n-m_i)!}{(n+m_i)!} P_n^{m_i}(0)^2 \right)^{-1} \\ &\quad \cdot \left( \sum_{n=|m_i|}^{\infty} j_n(kR_{\text{ref}}) h_n^{(1)}(kR_s) \frac{2n+1}{4\pi} \frac{(n-m_i)!}{(n+m_i)!} P_n^{m_i}(0)^2 \right)^{-1}. \end{aligned} \quad (7.44)$$

Here,  $m_i$  denotes the order in the  $i$ -th bin. Therefore, the WFR filter can be designed as an FIR filter. The driving signals can be obtained by convolution of the WFR filter and the received signals.



(a) Open sphere



(b) Rigid sphere

**Fig. 7.4.** Amplitude of WFR filter for open and rigid spheres in circular harmonic domain

The discretization of  $\phi$  limits the maximum order of  $m$ , which is denoted as  $M$  and determined as  $M = \lfloor M_\phi \rfloor$ . Here,  $\lfloor \cdot \rfloor$  is the floor function. The upper bound of  $M$  leads to limited reproduction region to the neighboring region of the center of the circular array.

The evanescent wave components of the received sound field are represented as:

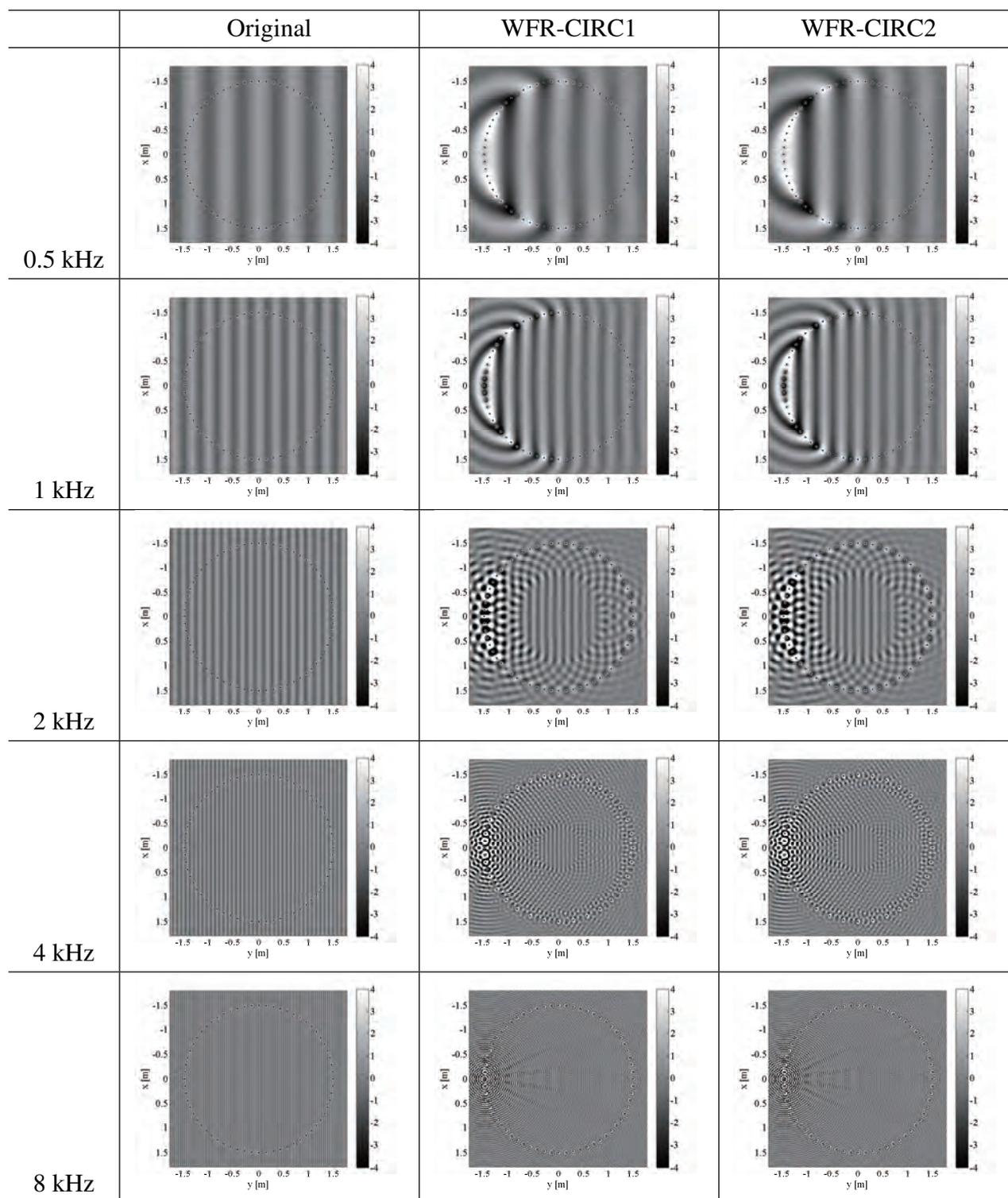
$$|m| > |kR_m|. \quad (7.45)$$

When the radius of the microphone array is smaller than that of the loudspeaker array, i.e.,  $R_m \ll R_s$ , the WFR filter becomes significantly unstable because inverse propagation of the received sound field is included as in Chapter 4. Therefore, it is necessary to apply a filter that reduces the components of  $|m_i| > |kR_m|$ . We employ a simple rectangular window function in the circular harmonic domain. The condition number of the WFR filter can be also used as an alternative criteria of the cutoff order to Eq. (7.45).

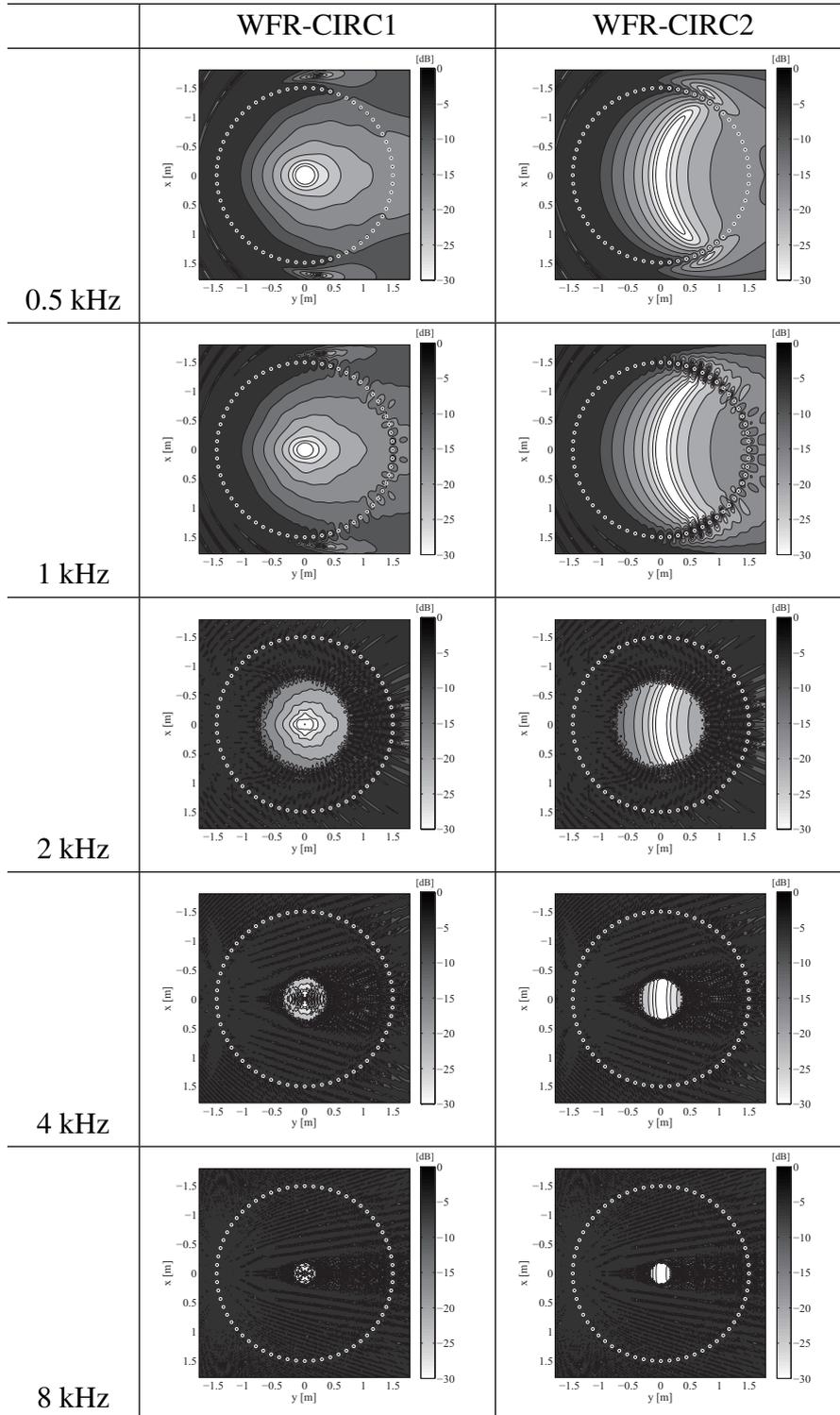
## 7.4 Experiments

Numerical simulations were conducted under the free-field condition. The proposed two kinds of methods using circular arrays, which were based on Eq. (7.39) and (7.40), were compared. These

Table 7.1. Original and reproduced sound pressure distribution.



**Table 7.2.** Time-averaged squared error distribution.



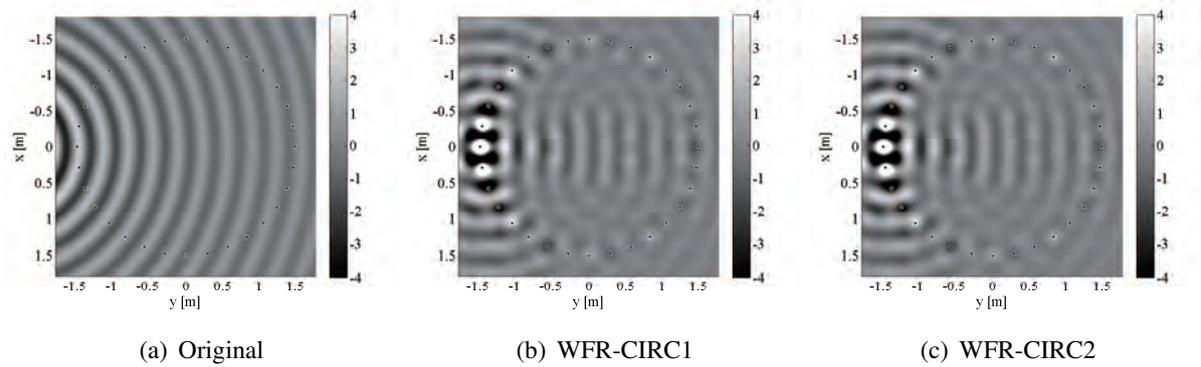
two methods are denoted as WFR-CIRC1 and WFR-CIRC2, respectively.

First, simulation results of two methods when the maximum order  $M$  is large are shown. The number of elements in the circular microphone and loudspeaker arrays was 64. The microphone array was mounted on a rigid spherical baffle. The radius of the arrays,  $R_m$  and  $R_s$ , was 1.5 m. The directivity of the array elements was assumed to be omni-directional. In the WFR-CIRC1, the reference radius  $R_{ref}$  was set at 0.5 cm. The sound pressure distribution was simulated in the  $3.6 \times 3.6$  m regions on the  $x$ - $y$ -plane at  $z = 0$  at intervals of 1.5 cm. The amplitudes were normalized at the origin.

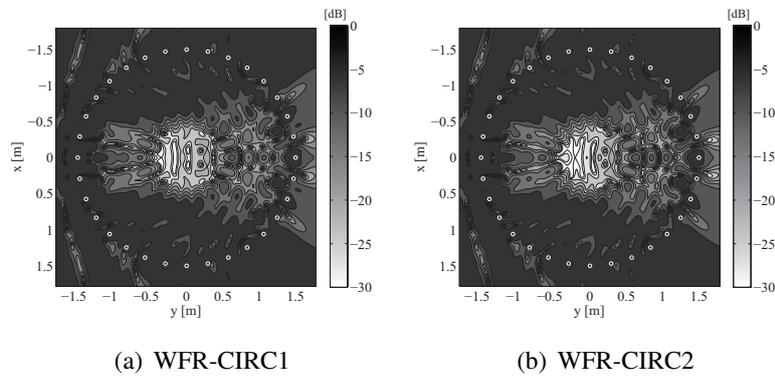
Table 7.1 shows the original and reproduced sound pressure distributions when the primary source was a plane wave in direction of 0 deg and the source signal was a sinusoidal wave of 0.5, 1, 2, 4, and 8 kHz. Table 7.2 is the time-averaged squared error distributions. Because the maximum order  $M$  is 32, the reproduction accuracy of 0.5 and 1 kHz was high in the whole region of the circular array in the both methods. In WFR-CIRC1, the reproduction accuracy was distinctly high on the neighboring region of the reference radius. On the other hand, in the WFR-CIRC2, the reproduction accuracy was high on a curve line due to the faster amplitude decay of a monopole source than that of a line source. Therefore, the high reproduction accuracy region of the WFR-CIRC2 may be limited only at the origin when there are multiple primary sources, whereas that of the WFR-CIRC1 is maintained on the reference radius. In 2, 4, and 8 kHz, the high reproduction accuracy region becomes smaller as the frequency becomes higher. This is because the maximum order  $M$  was limited.

Figure 7.4 (a) and (b) plots the amplitude of the WFR filter for open and rigid spherical baffle, i.e., the filter designed based on Eqs. (7.32) and (7.39), in the circular harmonic domain, respectively. The amplitudes for  $m = 0, 8, 16, 32$  are only shown. Note that these filters do not depend on the input signals. As shown in these figures, the WFR filter for open sphere diverged at many frequencies. This instability originates from the spherical Bessel function of the denominator of Eq. (7.32). The WFR filter for rigid sphere was much more stable than that for open sphere.

The small radius of the microphone array may be more practical. In the next experimental setup, the numbers of microphones and loudspeakers were 32. The radius of the arrays,  $R_m$  and  $R_s$ , were 0.25 and 1.5 m, respectively. A point source as a primary source was located at  $(r, \theta, \phi) = (2.5 \text{ m}, 90 \text{ deg}, -90 \text{ deg})$  in the source area. The source signal was a 1-kHz sinusoidal wave. Figure 7.5 is the original and reproduced sound pressure distributions, and Fig. 7.6 is the time-averaged squared error distributions. The maximum order  $M$  was 9, which was determined so that the condition number of the WFR filter was below 40 dB. Therefore, the high reproduction accuracy region was limited on the neighboring region of the center in the both methods.



**Fig. 7.5.** Original and reproduced sound pressure distribution when source signal was 1-kHz sinusoidal wave



**Fig. 7.6.** Time-averaged error distribution when source signal was 1-kHz sinusoidal wave

## 7.5 Conclusions

We proposed an SP-DS conversion method for spherical and circular array configurations of microphones and loudspeakers. Although the proposed WFR filtering method for spherical arrays is essentially equivalent to the HOA, the signal conversion is achieved in a single convolution of the WFR filter, whereas the HOA includes two stages of encoding and decoding. However, spatio-temporal convolution for the spherical array configurations is not as efficient as that for the other array configurations because the FFT algorithm can be applied under the limited conditions.

In circular array configurations, the WFR filtering method can be applied as an SP-DS conversion whereas the HOA cannot be directly applied. We derived the WFR filter for circular arrays under the assumption that the captured sound field is invariant with changes in the vertical direction. Numerical simulations were conducted to compare the two kinds of the WFR filter for circular arrays. The reproduction accuracy of the proposed WFR filtering methods was distinctly high in the neighboring region at the center of the circle at a broad frequency band. The regions

that the amplitude of the reproduced sound field coincide with that of the desired one were different in the two WFR filtering methods.

## Chapter 8

# Source Location Informed Sound-Pressure-to-Driving-Signal Conversion

From Chapter 4 to 7, we addressed SP-DS conversion methods for sound field reproduction. Especially in the WFR filtering methods, because the captured sound pressure distribution is transformed into the spatial frequency domain and the WFR filter that is analytically derived based on the property of wave propagation is applied, it is possible to instantly obtain driving signals of loudspeakers from received signals of microphones.

The SP-DS conversion methods that we presented so far do not take into account *a priori* knowledge of approximate positions of sound sources to be reproduced, *primary sources*, for SP-DS conversion. However, approximate positions of primary sources can be obtained by using, for example, the source localization method presented in Chapter 10. To exploit the prior information of the primary source locations, we propose an SP-DS conversion method based on the maximum *a posteriori* (MAP) estimation [92]. Spatial basis functions and their coefficients representing the driving signals of the loudspeakers are optimized based on the prior information. Therefore, reproduction accuracy is expected to be improved in the frequency bands above the spatial Nyquist frequency compared to the current SP-DS conversion methods, especially when more loudspeakers than microphones are available. This is because the driving signals of the loudspeakers are represented by using a few number of optimized basis functions and more precise interpolation can be achieved. The proposed method is closely related to the method used in near-field acoustical holography (NAH) based on the MAP estimation [93]. In this method, a predefined aperture function is introduced as a prior distribution for focusing on a specific location. We formulated the prior distribution of the driving signals of the loudspeakers based on the Rayleigh integral. We focus on the SP-DS conversion problem when the array configuration of the microphones and

loudspeakers are planar or linear.

## 8.1 Sparse sound field representation

From Chapter 5 to 7, we started from the homogeneous wave equation, or Helmholtz equation, and SP-DS conversion is achieved by decomposing a sound field into orthonormal basis functions which depend on coordinate systems. For example, in the WFR filtering method for linear arrays (Chapter 5), the received signals of microphones  $P_{\text{rcv}}(\mathbf{r}_m, \omega)$  and the driving signals of loudspeakers  $D(\mathbf{r}_s, \omega)$  are represented by using the spatial Fourier basis functions as:

$$P_{\text{rcv}}(\mathbf{r}_m, \omega) = \sum_n b_n(\omega) \phi_n(\mathbf{r}_m, \omega) \quad (8.1)$$

$$D(\mathbf{r}_s, \omega) = \sum_n c_n(\omega) \phi_n(\mathbf{r}_s, \omega), \quad (8.2)$$

where  $\phi_n(\cdot)$  is the predefined Fourier basis functions,  $b_n(\omega)$  and  $c_n(\omega)$  are their coefficients, and  $n$  is their index. This is because the homogeneous wave equation only constrains the property of the sound propagation without directly dealing with sound sources. On the other hand, the LC-PDE introduced in Section 2.5 constrains a source location and observation values, which is derived by assuming spherical wave propagation from a single point source.

Representing an observed sound pressure distribution in the Fourier basis functions makes it possible for stable and efficient signal conversions. However, in this procedure, artifacts originating from spatial aliasing notably occur, which depends on intervals of microphones and loudspeakers. As an example, assume a system in which loudspeakers are densely arranged compared to microphones. Even though this system has potential to reproduce a sound field in the frequency bands up to the spatial Nyquist frequency determined by the loudspeaker intervals, it can be hardly expected to reproduce it in the frequency bands above the spatial Nyquist frequency determined by the microphone intervals by using the WFR filtering method. This fact may be predictable from Fig. 5.12 (a) in Chapter 5. Even when a sound field generated by a single point source is observed using a linear microphone array, the energy of the driving signals of the loudspeakers spread over spatial frequency domain; therefore, it is difficult to avoid folding noise. However, if the driving signals of the loudspeakers can be represented using a small number of basis functions, it can be expected to reproduce the sound field in the frequency bands above the spatial Nyquist frequency determined by the microphone intervals because the signals may be interpolated more precisely. This can be referred to as a super-resolution of sound field reproduction.

This kind of sparse signal representation has been attracted in recent years, especially in *compressed sensing* literature [94–96]. A number of algorithms for solving underdetermined linear system of equations with regularization promoting sparsity of a solution vector have been proposed [97]. When the number of nonzero entries of a solution vector is used as an optimization criteria, this optimization problem requires exhaustive search. Therefore, computationally feasi-

ble algorithms that can successfully obtain a sparse solution have been investigated. Several algorithms that train basis functions from observed signals simultaneously solving sparse solutions have also been proposed, such as Method of Optimal Directions (MOD) [98] and K-SVD [99]. This kind of algorithms is referred to as *dictionary learning* [100].

In the field of sound field recording and reproduction, sparse representation of a sound field may be applied not only for the above-mentioned super-resolution of sound field reproduction, but also signal compression, echo cancellation, and so on. The most reasonable approach for sparse representation of the sound field may be focusing on sound sources to be reproduced. It can be expected that basis functions depending on primary sources provides sparse basis expansion of  $P_{rcv}(\mathbf{r}_m, \omega)$  and  $D(\mathbf{r}_s, \omega)$ .

Now, we introduce the inhomogeneous Helmholtz equation as an approach for dealing with sound sources directly. When a source distribution in an enclosed three-dimensional space  $\Omega$  is denoted as  $Q(\mathbf{r}, \omega)$ , a sound pressure at  $\mathbf{r}$  of frequency  $\omega$ ,  $P(\mathbf{r}, \omega)$ , can be described as:

$$(\nabla^2 + k^2) P(\mathbf{r}, \omega) = \begin{cases} -Q(\mathbf{r}, \omega), & \mathbf{r} \in \Omega \\ 0, & \mathbf{r} \notin \Omega \end{cases}. \quad (8.3)$$

It is assumed that any sound sources does not exist in the region of  $\mathbf{r} \notin \Omega$ . The solution of Eq. (8.3) can be represented as a sum of homogeneous and inhomogeneous components,  $P_h(\mathbf{r}, \omega)$  and  $P_i(\mathbf{r}, \omega)$ . The inhomogeneous component  $P_i(\mathbf{r}, \omega)$  can be described as a convolution of  $Q(\mathbf{r}, \omega)$  and the free-field Green function  $G(\mathbf{r}|\mathbf{r}')$  [21]. Therefore, the solution of Eq. (8.3) becomes

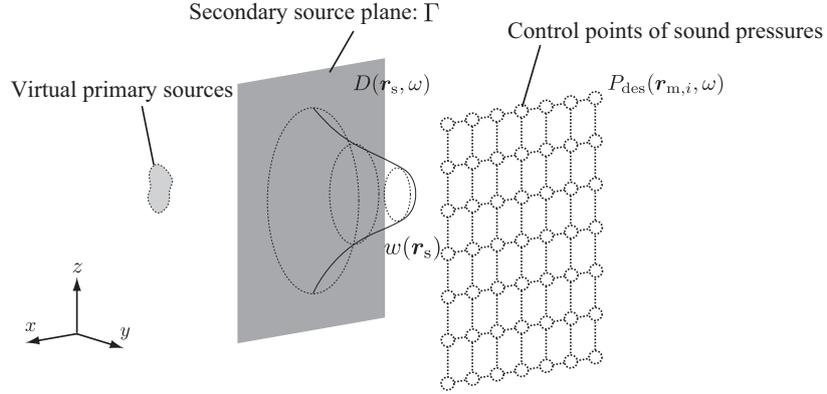
$$\begin{aligned} P(\mathbf{r}, \omega) &= P_i(\mathbf{r}, \omega) + P_h(\mathbf{r}, \omega) \\ &= \int_{\mathbf{r}' \in \Omega} Q(\mathbf{r}', \omega) G(\mathbf{r}|\mathbf{r}') d\mathbf{r}' + P_h(\mathbf{r}, \omega). \end{aligned} \quad (8.4)$$

This can be confirmed by substituting Eq. (8.4) into Eq. (8.3) as:

$$\begin{aligned} (\nabla^2 + k^2) \left\{ \int_{\mathbf{r}' \in \Omega} Q(\mathbf{r}', \omega) G(\mathbf{r}|\mathbf{r}') d\mathbf{r}' + P_h(\mathbf{r}, \omega) \right\} &= - \int_{\mathbf{r}' \in \Omega} \delta(\mathbf{r} - \mathbf{r}') Q(\mathbf{r}', \omega) d\mathbf{r}' \\ &= \begin{cases} -Q(\mathbf{r}, \omega), & \mathbf{r} \in \Omega \\ 0, & \mathbf{r} \notin \Omega \end{cases}. \end{aligned} \quad (8.5)$$

The source distribution  $Q(\mathbf{r}, \omega)$  is determined by locations of monopole sources and signal intensity of each source. Therefore,  $Q(\mathbf{r}, \omega)$  can be estimated from the observed signals by using source localization methods such as the method presented in Chapter 10. Other sensor data such as images may also useful to detect the source locations.

The dominant components of the received signals lie in  $P_i(\mathbf{r}, \omega)$  when the sound sources are located adjacent to the receivers. Additionally,  $Q(\mathbf{r}, \omega)$  ( $\mathbf{r} \in \Omega$ ) may become sparse because it can be considered that the sound sources exist only at a few locations in  $\Omega$ . Therefore, if the basis functions reflecting the primary source locations can be obtained, the sparse representation



**Fig. 8.1.** Geometry of secondary source plane and control points of sound pressures in target area.

of the received sound field may be achieved. The proposed method in this chapter optimizes the basis function based on the prior information of  $Q(\mathbf{r}, \omega)$ , which is formulated by using the MAP estimation framework.

Another way to directly deal with the sound sources is the LC-PDE. One of the advantages of the LC-PDE is that parameters of the source locations are included in itself. Because to obtain  $Q(\mathbf{r}, \omega)$  in Eq. (8.4) is not a trivial task, it may be possible to offer a more efficient algorithm for SP-DS conversion with the LC-PDE; therefore, a future work may be to develop the SP-DS conversion based on the LC-PDE. One of the advantages of the LC-PDE is presented in Chapter 10 as an application to the source localization.

## 8.2 Problem formulation

As shown in Fig. 8.1, we assume that the sound field is reproduced using a continuous distribution of secondary sources (loudspeakers) on the plane  $\Gamma$  in the target area. Desired sound pressures are obtained using  $M$  microphones in the source area. The driving signals of secondary sources of the temporal frequency  $\omega$  are denoted as  $D(\mathbf{r}_s, \omega)$ , where  $\mathbf{r}_s$  is the position on  $\Gamma$ . The desired sound pressures are denoted as

$$\mathbf{P}^{\text{des}}(\omega) = [P_{\text{des}}(\mathbf{r}_{m,1}, \omega), P_{\text{des}}(\mathbf{r}_{m,2}, \omega), \dots, P_{\text{des}}(\mathbf{r}_{m,M}, \omega)]^T,$$

where  $\mathbf{r}_{m,i}$  ( $i = 1, \dots, M$ ) is the positions that the sound pressures are controlled to correspond with the desired ones, i.e., control points. The control points are also distributed on a plane because the planar microphone array in the source area is assumed. Additionally, an enclosed distribution of the control points causes non-uniqueness of the driving signals [55] as presented in Chapter 6 and 7. The driving signals  $D(\mathbf{r}_s, \omega)$  need to satisfy the following equation:

$$\mathbf{P}^{\text{des}}(\omega) = \int_{\Gamma} D(\mathbf{r}_s, \omega) \mathbf{G}(\mathbf{r}_s, \omega) d\Gamma + \boldsymbol{\varepsilon}, \quad (8.6)$$

where

$$\mathbf{G}(\mathbf{r}_s, \omega) = [G(\mathbf{r}_{m,1} - \mathbf{r}_s, \omega), G(\mathbf{r}_{m,2} - \mathbf{r}_s, \omega), \dots, G(\mathbf{r}_{m,M} - \mathbf{r}_s, \omega)]^T$$

is the  $M$ -size vector with the elements of the transfer functions between  $\mathbf{r}_{m,i}$  and  $\mathbf{r}_s$ , and  $\boldsymbol{\varepsilon}$  is the  $M$ -size vector with the elements of the errors at each control point. For simplicity,  $G(\mathbf{r}_{m,i} - \mathbf{r}_s, \omega)$  is assumed to be monopole as:

$$G(\mathbf{r}_{m,i} - \mathbf{r}_s, \omega) = \frac{e^{jk|\mathbf{r}_{m,i} - \mathbf{r}_s|}}{4\pi|\mathbf{r}_{m,i} - \mathbf{r}_s|}. \quad (8.7)$$

The goal is to derive optimal  $D(\mathbf{r}_s, \omega)$  that satisfy Eq. (8.6).

We assume that an estimate of  $D(\mathbf{r}_s, \omega)$  is represented as the following form:

$$\hat{D}(\mathbf{r}_s) = \sum_{n=1}^N c_n \phi_n(\mathbf{r}_s) = \boldsymbol{\phi}^T(\mathbf{r}_s) \mathbf{c}, \quad (8.8)$$

where

$$\mathbf{c} = [c_1, \dots, c_N]^T$$

are the coefficients that depend on  $\mathbf{P}^{\text{des}}$  and

$$\boldsymbol{\phi}(\mathbf{r}_s) = [\phi_1(\mathbf{r}_s), \dots, \phi_N(\mathbf{r}_s)]^T$$

is the spatial basis functions that are independent from  $\mathbf{P}^{\text{des}}$ . For simplicity,  $\omega$  is omitted from here on. In the WFR filtering method presented in Chapter 5, the Fourier basis functions that are equivalent to plane waves are used as  $\boldsymbol{\phi}(\mathbf{r}_s)$  for fast computation of SP-DS conversion with the FFT algorithm, and  $\mathbf{c}$  is obtained by applying the WFR filter to  $\mathbf{P}^{\text{des}}$  that is analytically derived based on the wave propagation properties. The problem considered here is how  $\boldsymbol{\phi}(\mathbf{r}_s)$  and  $\mathbf{c}$  are optimized when the prior information of  $D(\mathbf{r}_s)$  is available.

## 8.3 MAP Estimation of Driving Signals of Secondary Sources

### 8.3.1 Objective function and its solution

The above-mentioned problem can be formulated based on the MAP estimation framework [92]. It can be considered that the optimal  $\boldsymbol{\phi}(\mathbf{r}_s)$  and  $\mathbf{c}$  maximize the posterior probability distribution given  $\mathbf{P}^{\text{des}}$  as:

$$\left( \hat{\mathbf{c}}, \hat{\boldsymbol{\phi}}(\mathbf{r}_s) \right) = \arg \max_{\mathbf{c}, \boldsymbol{\phi}(\mathbf{r}_s)} \pi \left( D(\mathbf{r}_s) = \boldsymbol{\phi}^T(\mathbf{r}_s) \mathbf{c} \mid \mathbf{P}^{\text{des}} \right) \quad (8.9)$$

Based on the Bayes' rule, the right-hand side of Eq. (8.9) becomes

$$\pi \left( D(\mathbf{r}_s) \mid \mathbf{P}^{\text{des}} \right) \propto \pi \left( \mathbf{P}^{\text{des}} \mid D(\mathbf{r}_s) \right) \pi \left( D(\mathbf{r}_s) \right) \quad (8.10)$$

where the likelihood function  $\pi(\mathbf{P}^{\text{des}}|D(\mathbf{r}_s))$  represents the probability of  $\mathbf{P}^{\text{des}}$  given  $D(\mathbf{r}_s)$ , and  $\pi(D(\mathbf{r}_s))$  is the prior probability distribution of  $D(\mathbf{r}_s)$ .

From Eq. (8.6),  $\pi(\mathbf{P}^{\text{des}}|D(\mathbf{r}_s))$  corresponds with the probability distribution of  $\varepsilon$ . Therefore, we assume that  $\pi(\mathbf{P}^{\text{des}}|D(\mathbf{r}_s))$  follows a complex Gaussian distribution  $\mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I})$  as:

$$\pi(\mathbf{P}^{\text{des}}|D(\mathbf{r}_s)) \propto \exp\left(\frac{1}{\sigma_\varepsilon^2} \left\| \mathbf{P}^{\text{des}} - \int_{\Gamma} D(\mathbf{r}_s) \mathbf{G}(\mathbf{r}_s) d\Gamma \right\|_2^2\right), \quad (8.11)$$

where  $\|\cdot\|_2^2$  represents the squared  $\ell_2$ -norm.

The prior distribution  $\pi(D(\mathbf{r}_s))$  needs to reflect some sort of *a priori* information of  $D(\mathbf{r}_s)$ . We assumed that the source locations in the source area can be obtained from the received signals of the microphones or other sensor data. Under this assumption, approximate knowledge about the amplitude distribution of  $D(\mathbf{r}_s)$  may be available because the locations of the virtual primary sources are approximately known. To formulate this idea, the spatial covariance function of  $D(\mathbf{r}_s)$  is described as:

$$E[D(\mathbf{r}_s)D(\mathbf{r}'_s)^*] = \sigma_d^2 w(\mathbf{r}_s) \delta(\mathbf{r}_s - \mathbf{r}'_s) \quad (8.12)$$

It is assumed that the driving signals are spatially uncorrelated. Although it may be possible to include spatial correlations of the driving signals as prior information, this assumption leads to a simple algorithm of SP-DS conversion. The function  $w(\mathbf{r}_s)$  represents the distribution of the squared amplitude of  $D(\mathbf{r}_s)$  with normalization as  $\int_{\Gamma} w(\mathbf{r}) d\Gamma = 1$ , and  $\sigma_d^2$  reflects its mean value. The probability distribution of  $D(\mathbf{r}_s)$  is assumed as a complex Gaussian distribution  $\mathcal{N}(0, \sigma_d^2 w(\mathbf{r}_s))$ :

$$\begin{aligned} \pi(D(\mathbf{r}_s)) &\propto \exp\left(-\frac{1}{\sigma_d^2} \|D(\mathbf{r}_s)\|_w^2\right) \\ &= \exp\left(-\frac{1}{\sigma_d^2} \|\boldsymbol{\phi}^T \mathbf{c}\|_w^2\right), \end{aligned} \quad (8.13)$$

where

$$\|D(\mathbf{r}_s)\|_w^2 \equiv \int_{\Gamma} |D(\mathbf{r}_s)|^2 w^{-1}(\mathbf{r}_s) d\Gamma. \quad (8.14)$$

The details on determining  $w(\mathbf{r}_s)$  are described in Section 8.3.2.

Now, the objective function for Eq. (8.9),  $J(\mathbf{c}, \boldsymbol{\phi}(\mathbf{r}_s))$ , can be formulated. The negative of the logarithm of  $\pi(D(\mathbf{r}_s)|\mathbf{P}^{\text{des}})$  is used for  $J(\mathbf{c}, \boldsymbol{\phi}(\mathbf{r}_s))$  as:

$$\begin{aligned} J(\mathbf{c}, \boldsymbol{\phi}(\mathbf{r}_s)) &= -\log \pi(D(\mathbf{r}_s)|\mathbf{P}^{\text{des}}) \\ &= -\log \pi(\mathbf{P}^{\text{des}}|D(\mathbf{r}_s)) - \log \pi(D(\mathbf{r}_s)) \\ &= \text{const.} + \frac{1}{\sigma_d^2} \|\boldsymbol{\phi}^T \mathbf{c}\|_w^2 + \frac{1}{\sigma_\varepsilon^2} \left\| \mathbf{P}^{\text{des}} - \int_{\Gamma} \mathbf{G}(\mathbf{r}_s) \boldsymbol{\phi}^T(\mathbf{r}_s) \mathbf{c} d\Gamma \right\|_2^2. \end{aligned} \quad (8.15)$$

Note that the argument  $\mathbf{r}_s$  is omitted for simplicity. Therefore,  $\phi$  and  $\mathbf{c}$  that minimize  $J(\mathbf{c}, \phi)$  should be optimal. The differential of Eq. (8.15) with respect to  $\mathbf{c}^*$  and  $\phi^*$  becomes

$$\frac{\partial J(\mathbf{c}, \phi)}{\partial \mathbf{c}^*} = \frac{1}{\sigma_d^2} \int \phi^* \phi^T \mathbf{c} w^{-1} d\Gamma + \frac{1}{\sigma_\varepsilon^2} \left( \iint \phi^* \mathbf{G}^H \mathbf{G} \phi^T \mathbf{c} d\Gamma^2 - \int \phi^* \mathbf{G}^H \mathbf{P}^{\text{des}} d\Gamma \right) \quad (8.16)$$

and

$$\frac{\partial J(\mathbf{c}, \phi)}{\partial \phi^*} = \mathbf{c}^H \left( \frac{1}{\sigma_d^2} \int \phi^T \mathbf{c} w^{-1} d\Gamma + \frac{1}{\sigma_\varepsilon^2} \left( \iint \mathbf{G}^H \mathbf{G} \phi^T \mathbf{c} d\Gamma^2 - \int \mathbf{G}^H \mathbf{P}^{\text{des}} d\Gamma \right) \right), \quad (8.17)$$

respectively. By setting  $\partial J(\mathbf{c}, \phi)/\partial \mathbf{c}^* = 0$  and  $\partial J(\mathbf{c}, \phi)/\partial \phi^* = 0$ , the following equations are respectively derived:

$$\left( \alpha \int \phi^* \phi^T w^{-1} d\Gamma + \iint \phi^* \mathbf{G}^H \mathbf{G} \phi^T d\Gamma^2 \right) \mathbf{c} = \int \phi^* \mathbf{G}^H \mathbf{P}^{\text{des}} d\Gamma \quad (8.18)$$

and

$$\mathbf{c}^H \left( \alpha \int \phi^T w^{-1} d\Gamma + \iint \mathbf{G}^H \mathbf{G} \phi^T d\Gamma^2 \right) \mathbf{c} = \mathbf{c}^H \int \mathbf{G}^H \mathbf{P}^{\text{des}} d\Gamma, \quad (8.19)$$

where  $\alpha \equiv \sigma_\varepsilon^2/\sigma_d^2$  is defined.

To derive  $\mathbf{c}$  and  $\phi$  that satisfy Eqs. (8.18) and (8.19), we introduce a matrix composed by descretizing  $\mathbf{r}_s$  of  $w(\mathbf{r}_s)\mathbf{G}^H(\mathbf{r}_s)$ . A row vector of the singular value decomposition of this matrix is described as:

$$w(\mathbf{r}_s)\mathbf{G}^H(\mathbf{r}_s) = \boldsymbol{\psi}^T(\mathbf{r}_s)\boldsymbol{\Lambda}\mathbf{U}^H, \quad (8.20)$$

where

$$\int \boldsymbol{\psi}(\mathbf{r}_s)\boldsymbol{\psi}^H(\mathbf{r}_s)d\Gamma = \mathbf{I},$$

$$\mathbf{U}\mathbf{U}^H = \mathbf{I},$$

and  $\boldsymbol{\Lambda}$  is a diagonal matrix with non-negative elements. The optimal basis function in Eq. (8.9) corresponds with the eigen-function of Eq. (8.20), i.e.,  $\hat{\phi}(\mathbf{r}_s) \equiv \boldsymbol{\psi}(\mathbf{r}_s)$ . This fact is proved by substituting Eq. (8.20) into Eqs. (8.18) and (8.19) by a similar procedure presented in Ref. [93]. Therefore, the optimal  $\mathbf{c}$  and  $\phi$  are derived as:

$$\hat{\mathbf{c}} = (\alpha\mathbf{I} + \boldsymbol{\Lambda}^2)^{-1} \boldsymbol{\Lambda}\mathbf{U}^H \mathbf{P}^{\text{des}} \quad (8.21)$$

and

$$\hat{\phi}(\mathbf{r}_s) = w(\mathbf{r}_s)\boldsymbol{\Lambda}^{-1}\mathbf{U}^T \mathbf{G}^*(\mathbf{r}_s). \quad (8.22)$$

Finally, the driving signals are obtained using Eqs. (8.8), (8.21) and (8.22) as:

$$\hat{D}(\mathbf{r}_s) = w(\mathbf{r}_s) \mathbf{G}^H(\mathbf{r}_s) \mathbf{U} (\alpha \mathbf{I} + \mathbf{\Lambda}^2)^{-1} \mathbf{U}^H \mathbf{P}^{\text{des}} \quad (8.23)$$

The constant value  $\alpha$  can be regarded as a regularization parameter.

### 8.3.2 Prior based on primary source locations

The remaining problem is how to choose a proper  $w(\mathbf{r}_s)$  in Eq. (8.12). As mentioned previously,  $w(\mathbf{r}_s)$  is related to the distribution of the squared amplitude of  $D(\mathbf{r}_s)$ . Under the assumption that the source locations in the source area are available, the Rayleigh integral of the first kind in three dimensions suggests a shape of  $w(\mathbf{r}_s)$ :

$$P_{\text{des}}(\mathbf{r}) = 2 \int_{\Gamma} \frac{\partial P_{\text{des}}(\mathbf{r}_s)}{\partial y_s} G(\mathbf{r} - \mathbf{r}_s) d\Gamma, \quad (8.24)$$

where  $P_{\text{des}}(\mathbf{r})$  is the desired sound pressure distribution at position  $\mathbf{r}$ . As in WFS presented in Section 3.2.1, comparing Eq. (8.24) with Eq. (8.6),  $D(\mathbf{r}_s)$  should equal the sound pressure gradient of the desired sound field on  $\Gamma$ . We assume that the primary sources has monopole characteristics. When the approximate location of virtual primary source is denoted as  $\mathbf{r}_{\text{ps}}$ , the squared amplitude of  $D(\mathbf{r}_s)$ , is approximated as:

$$\begin{aligned} \check{w}(\mathbf{r}_s) &= \left| 2 \frac{\partial P_{\text{des}}(\mathbf{r}_s)}{\partial y_s} \right|^2 \\ &= \left| (y_s - y_{\text{ps}}) \frac{jk|\mathbf{r}_s - \mathbf{r}_{\text{ps}}| - 1}{2\pi|\mathbf{r}_s - \mathbf{r}_{\text{ps}}|^2} e^{jk|\mathbf{r}_s - \mathbf{r}_{\text{ps}}|} \right|^2. \end{aligned} \quad (8.25)$$

Additionally, we applied a spatial tapering window to  $\check{w}(\mathbf{r}_s)$  to reduce the truncation error, as in WFS. Therefore,  $w(\mathbf{r}_s)$  is obtained by normalizing  $\check{w}(\mathbf{r}_s)$  as

$$w(\mathbf{r}_s) = \check{w}(\mathbf{r}_s) / \int_{\Gamma} \check{w}(\mathbf{r}_s) d\Gamma.$$

When there are multiple primary sources, the relative amplitude of each source is required to calculate Eq. (8.25).

### 8.3.3 Practical algorithm of proposed method

The practical algorithm of the proposed method is described as follows:

1. Detect source location ( $\mathbf{r}_{\text{ps}}$ ) with microphones or other sensors in the source area.
2. Calculate  $w(\mathbf{r}_s)$  based on Eq. (8.25).
3. Compute the spatial correlations as

$$R_{ij} = \int_{\Gamma} G(\mathbf{r}_i - \mathbf{r}_s) G^*(\mathbf{r}_j - \mathbf{r}_s) w(\mathbf{r}_s) d\Gamma. \quad (8.26)$$

4. Compute eigenvalue decomposition as

$$\mathbf{R} = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^H, \quad (8.27)$$

where  $\mathbf{R}$  is the  $M \times M$  matrix with elements  $R_{ij}$ .

5. Obtain transform filter for SP-DS conversion based on Eq. (8.23).

In step 2, because the loudspeaker locations, i.e., secondary source distribution, are discretized in practice, the computation is performed by matrix multiplications. The regularization parameter  $\alpha$  in Eq. (8.23) can be obtained using the L-curve [71] or generalized cross validation method [72].

The proposed method can be straightforwardly extended to the linear array configuration case. In the linear case,  $w(\mathbf{r}_s)$  should be calculated based on the Rayleigh integral of the first kind in two dimensions as in WFS for linear loudspeaker array.

### 8.3.4 Comparison with sound pressure control using minimum norm solution

For comparison, we reformulate the sound pressure control method presented in Section 3.2.3 in this context. When the position of the loudspeakers are discretized as  $\mathbf{r}_{s,j}$  ( $j = 1, \dots, L$  ( $L > M$ )) and  $\mathbf{H}$  denotes the  $L \times M$ -size matrix with the element of  $G(\mathbf{r}_{m,i} - \mathbf{r}_{s,j})$ , the  $L$ -size vector of driving signals  $\mathbf{D}^{\text{MN}}$  can be obtained as a minimum norm solution:

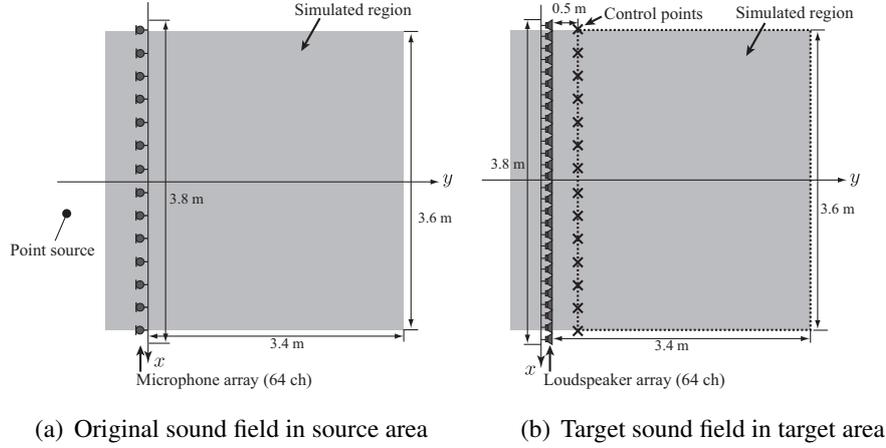
$$\hat{\mathbf{D}}^{\text{MN}} = \mathbf{H}^H (\mathbf{H} \mathbf{H}^H + \beta \mathbf{I})^{-1} \mathbf{P}^{\text{des}}, \quad (8.28)$$

where  $\beta$  is a regularization parameter. It should be noted that Eq. (8.28) corresponds with the case in which  $w(\mathbf{r}_s)$  is uniform in Eq. (8.23). Furthermore, Eq. (8.23) is regularized on the projected space of  $\mathbf{U}$  that depends on  $w(\mathbf{r}_s)$ .

## 8.4 Experiments

Numerical simulations were conducted under the free-field assumption to compare the three methods; the proposed method based on MAP estimation, WFR filtering method presented in Section 5.3, and sound pressure control method presented in Section 8.3.4. These three methods are respectively denoted as MAP, WFR, and MN.

As shown in Fig. 8.2, the loudspeakers were aligned along the  $x$ -axis with the center at the origin in the target area. The control points were aligned along the line  $y = 0.5$  m at  $z = 0$  with the center on the  $y$ -axis. The numbers of the loudspeakers and control points (microphones) were 64 and 32, respectively. The interval of the loudspeakers was 6 cm and that of the control points was 12 cm; therefore, both the array lengths of the loudspeakers in the target area and microphones in the source area were 3.84 m. The directivity of the array elements was assumed to be omnidirectional. A single point source as a primary source was located in the recording area and its



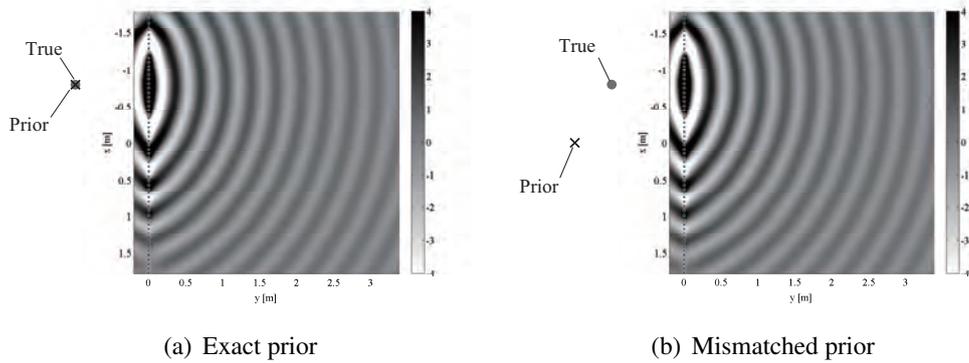
**Fig. 8.2.** Numerical simulation setup. Linear microphone and loudspeaker arrays were located at  $x$ -axis.

corresponding position of the virtual point source was at  $(-0.8, -1.0, 0.0)$  m. With the linear loudspeaker array, the amplitude decay of the virtual primary source becomes faster than desired even though this artifact may not be critical to listeners' perception. Therefore, we eliminated this effect for feasible evaluation by modifying the amplitude decay of the primary point source in the original sound field from  $1/r$  to  $1/r\sqrt{r}$ , where  $r$  is the distance from the source location. The sound pressure distributions were simulated in a  $3.6 \times 3.6$  m region (shaded region in Fig. 8.2) at intervals of 1.5 cm on the  $x$ - $y$ -plane at  $z = 0$ . The amplitudes were normalized at the center of the simulated region. The sampling frequency was 48 kHz. The general reproduction error was evaluated using the SDR defined in Eq. (4.13). The SDR was calculated in the region of  $y \geq 0.5$  m.

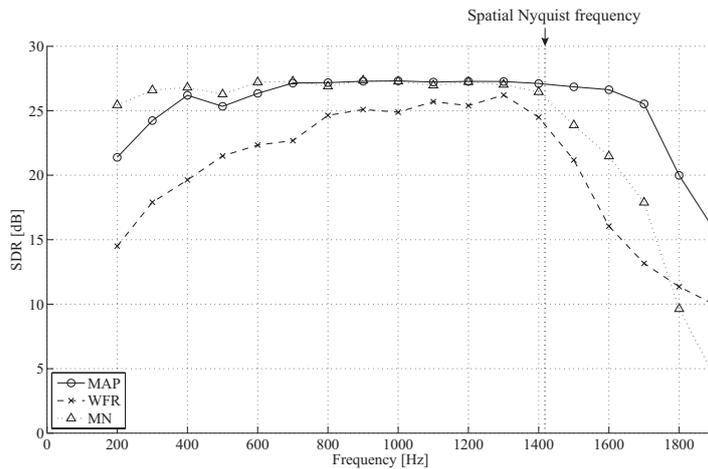
A Tukey window function was applied to the WFR and  $\tilde{w}(\mathbf{r}_s)$  in Eq. (8.25) of the MAP as a spatial window whose sides tapered by 10%. The regularization parameters of the MAP and MN were determined using the L-curve method. In the WFR, sinc interpolation is used for upsampling of the received sound pressures, and the method presented in Section 5.4.1 was used for extrapolation of the sound field between the loudspeaker array and control points.

Figure 8.3(a) shows the simulation result when the source signal was a 1-kHz sinusoidal wave. In the MAP, it was assumed that the exact position of the virtual point source was known. The SDRs of the MAP, WFR, and MN were 27.3, 24.9, and 27.3 dB, respectively. Fig. 8.3(b) shows the simulation results when the prior information of the virtual point source position was set as  $(0.0, -1.5, 0.0)$  m, which is different from the true position. Despite the mismatched prior distribution, the SDR of the MAP was 27.4 dB. Because the spatial basis functions of the MAP is derived as a complete set of eigen-functions, the mismatch between the prior and ideal distributions does not have a major negative impact.

Figure 8.4 plots the relation between SDRs and the frequency of the source signal. The exact position of the virtual point source was set as prior in the MAP. The spatial Nyquist frequency



**Fig. 8.3.** Reproduced sound pressure distributions by using MAP when source signal was 1-kHz sinusoidal wave. Prior mismatch does not have major negative impact on reproduction accuracy.

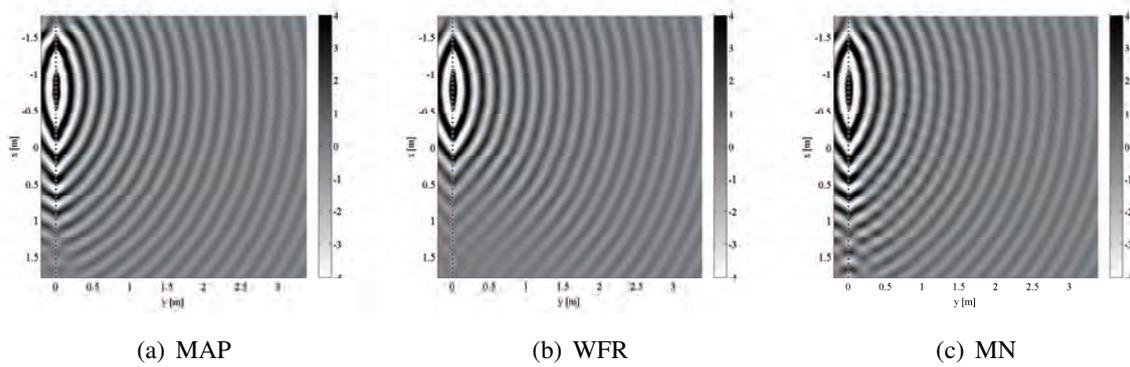


**Fig. 8.4.** Relation between SDR and frequency of source signal.

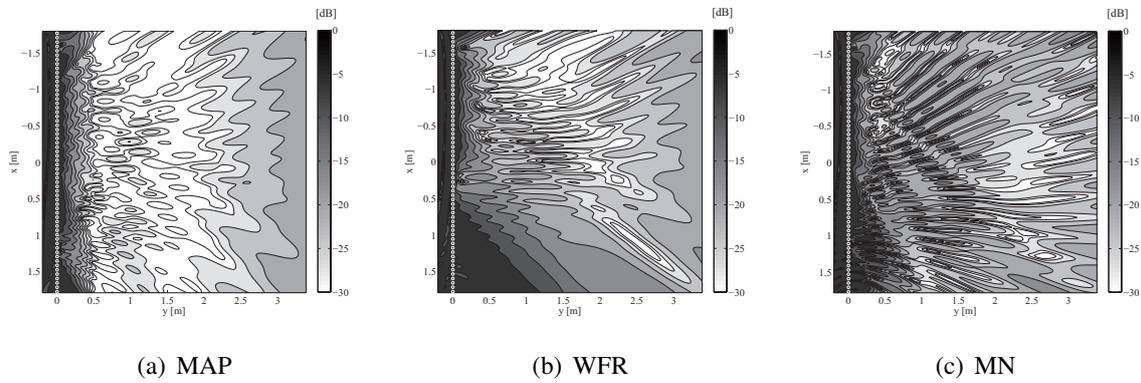
determined by the interval of the microphones is indicated by the dashed line. The SDRs of the WFR were smaller than the others, especially at low frequencies. This error originated from the error of the sinc interpolation. The SDRs of the MAP at a higher frequency than the spatial Nyquist frequency were maintained compared to the MN and WFR. Therefore, the proposed MAP can be considered as being more robust against the spatial aliasing artifacts. The reproduced sound pressure and error distributions when the source signal was 1.7 kHz sinusoidal wave are shown in Figs. 8.5 and 8.6, respectively.

## 8.5 Conclusions

We proposed SP-DS conversion method based on the MAP estimation. The method is derived from optimizing the spatial basis function and its coefficients of the driving signals of the loud-



**Fig. 8.5.** Reproduced sound pressure distributions when source signal was 1.7 kHz sinusoidal wave.



**Fig. 8.6.** Error distributions when source signal was 1.7 kHz sinusoidal wave.

speakers when the distribution of the squared amplitude of the driving signals is available as the prior. Numerical simulations were conducted to compare the proposed method with the WFR filtering method and sound pressure control method based on the minimum norm solution. The reproduction accuracy of the proposed method was better than that of the other methods, especially at frequencies above the spatial Nyquist frequency. It is assumed that the prior is obtained beforehand by using the approximate knowledge of the source locations in the source area, but the mismatch did not have a major negative impact on reproduction accuracy.

## Chapter 9

# Practical Implementations of Sound Field Transmission Systems

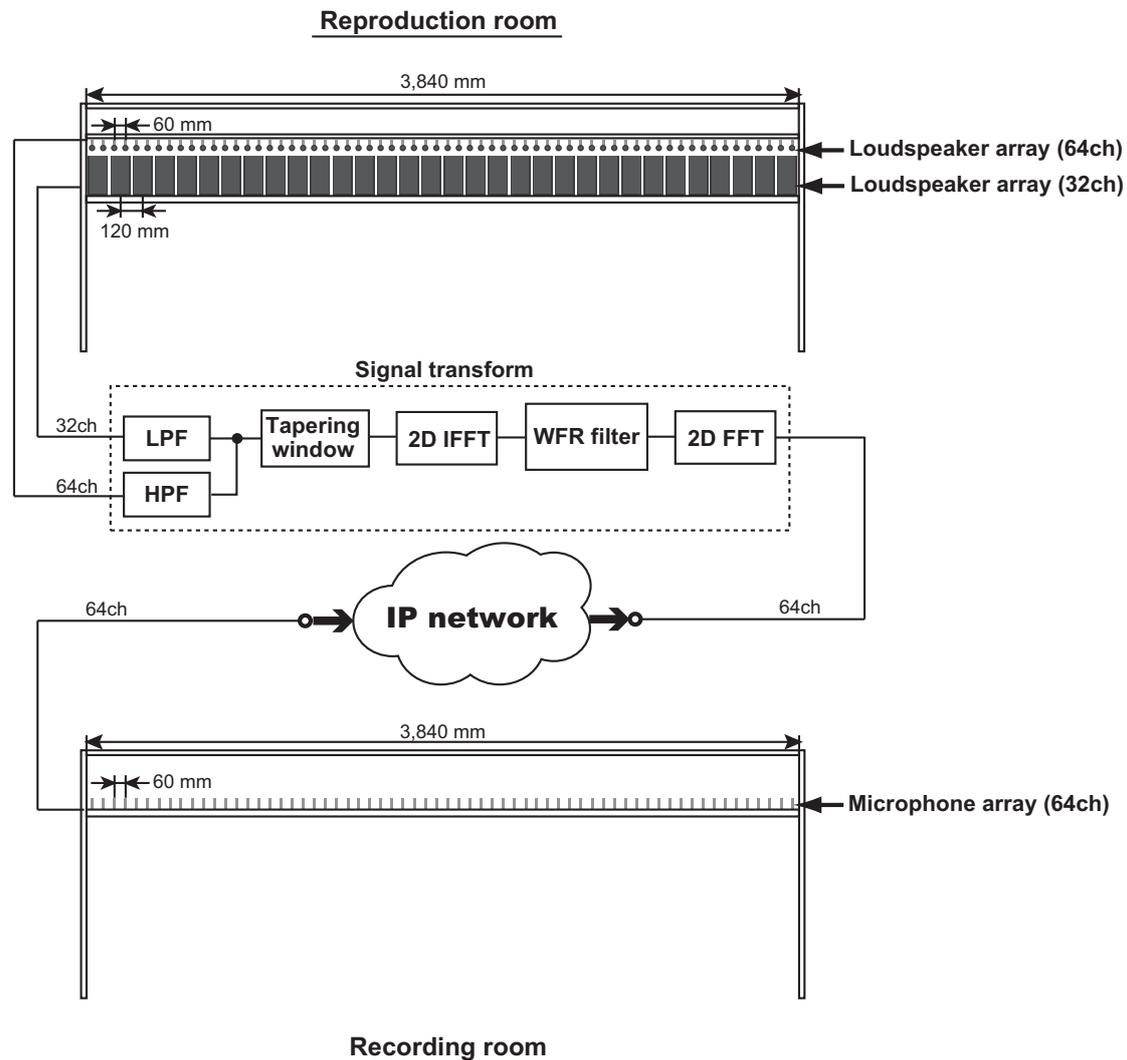
Sound field transmission systems by using linear arrays of microphones and loudspeakers have the potential for many applications such as live broadcasting and teleconference systems because a large listening area and simple implementation compared to systems using planar arrays can be achieved.

As discussed in Chapter 5, because WFS and the related analytical methods for linear loudspeaker arrays cannot be directly applied as an SP-DS conversion [31, 101, 102], they are not suitable for real-time sound field transmission systems. Methods based on numerical algorithm have difficulties in designing and applying the transform filters even though they can be applied to SP-DS conversion in principle. Therefore, the WFR filtering method presented in Section 5.3 is a suitable way to implement real-time sound field transmission systems.

We developed a real-time sound field transmission system over an IP network. The sound field received by the linear microphone array at the far-end is transmitted and reconstructed using the linear loudspeaker array and the WFR filter in real-time at the near-end. We conducted measurement experiments and subjective listening tests of this system. A demonstration system including 3D images is also introduced.

### 9.1 System specifications

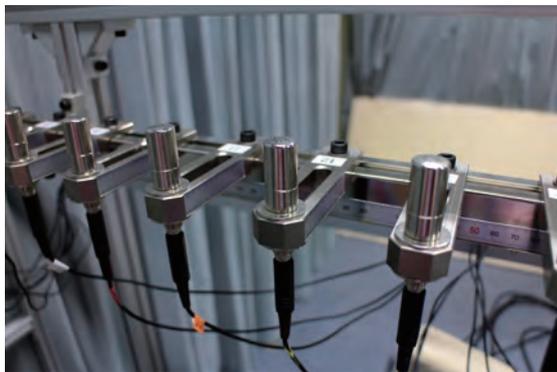
A block diagram of the system is depicted in Fig. 9.1. Linear microphone and loudspeaker arrays were set in the recording and reproduction rooms, respectively. Photos of the array elements are shown in Fig. 9.2. The microphone array was made up of 64 omni-directional microphones, and each microphone was equally spaced 6 cm apart. Primo EM172 were used. The loudspeaker array was composed of two kinds of loudspeakers. One was for higher frequency band; there were 64 of these loudspeakers equally spaced 60 mm apart. The other was for lower frequency band,



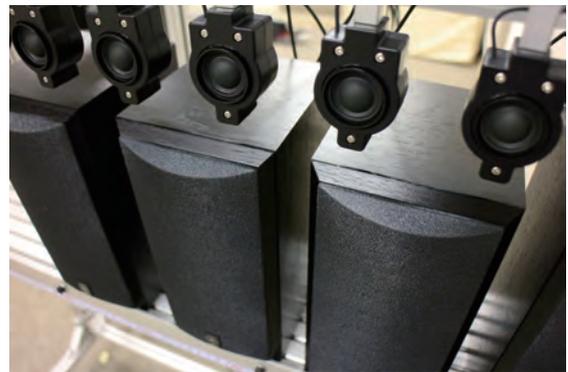
**Fig. 9.1.** Block diagram of real-time sound field transmission system.

and there were 32 loudspeakers equally spaced 120 mm apart. Therefore, the spatial Nyquist frequency was about 2.8 kHz. We used Foster 411222 loudspeakers for higher frequency band and Yamaha NS-B210 loudspeakers for the lower frequency band. Figure 9.3 shows measured directivity of the loudspeakers for higher frequency band. We measured impulse responses of one of the loudspeakers by using time-stretched pulse (TSP) signals [103] at every 5.0 deg in an anechoic room. Differences of the signal power between 0 deg and  $\pm 90$  deg were within 1.0 dB below 1.3 kHz; therefore, the directivity of the loudspeakers can be assumed as omni-directional below 1.3 kHz. The microphone and loudspeaker arrays were 3.84 m long. A MOTU 24 I/O interface was used for A/D and D/A converters.

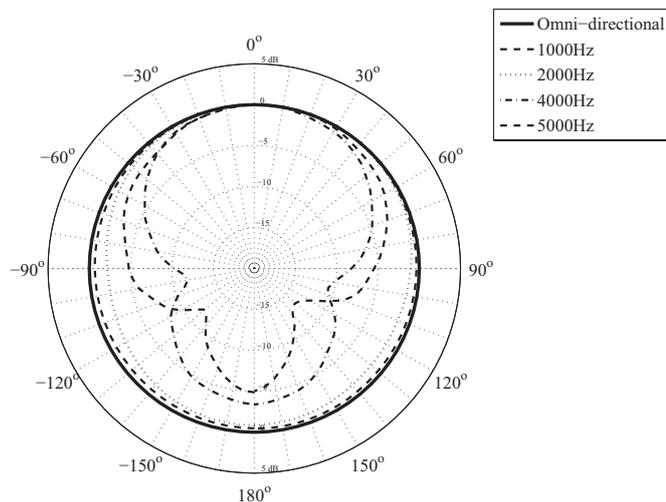
The signals obtained by the microphone array in the recording room were transmitted to the reproduction room by using UDP (User Datagram Protocol) over an exclusive IP network called GEMnet2 [104]. The signals were encoded as 16-bit PCM (Pulse Code Modulation) data, and the



(a) Microphone unit



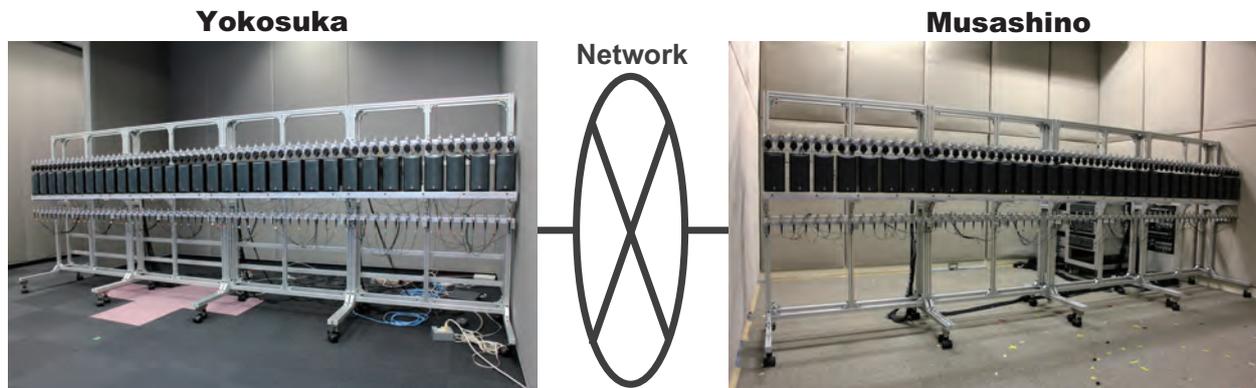
(b) Loudspeaker unit

**Fig. 9.2.** Photos of microphones and loudspeakers.**Fig. 9.3.** Measured directivity of loudspeaker.

sampling frequency was 48 kHz. Therefore, the transmission rate was about 50 Mbps.

The signals of the microphone array were transformed into driving signals of the loudspeaker array by using the WFR filtering method. The WFR filter is designed as a 2D FIR filter by using Eq. (5.37). Therefore, conversion was achieved as 2D convolution of the WFR filter and the received signals. The length of the WFR filter was 1024 taps in the time domain and 64 taps in the space domain. Therefore, a 2D FFT of  $2048 \times 128$  was applied for 2D linear convolution. The Tukey window function was applied as a tapering window whose sides tapered by 10%. The converted signals were divided into two frequency bands, and the cut-off frequency was 0.6 kHz. These signals were used for the driving signals of the two loudspeaker arrays. The signals of the 32 channels were created by thinning the signals of the 64 channels.

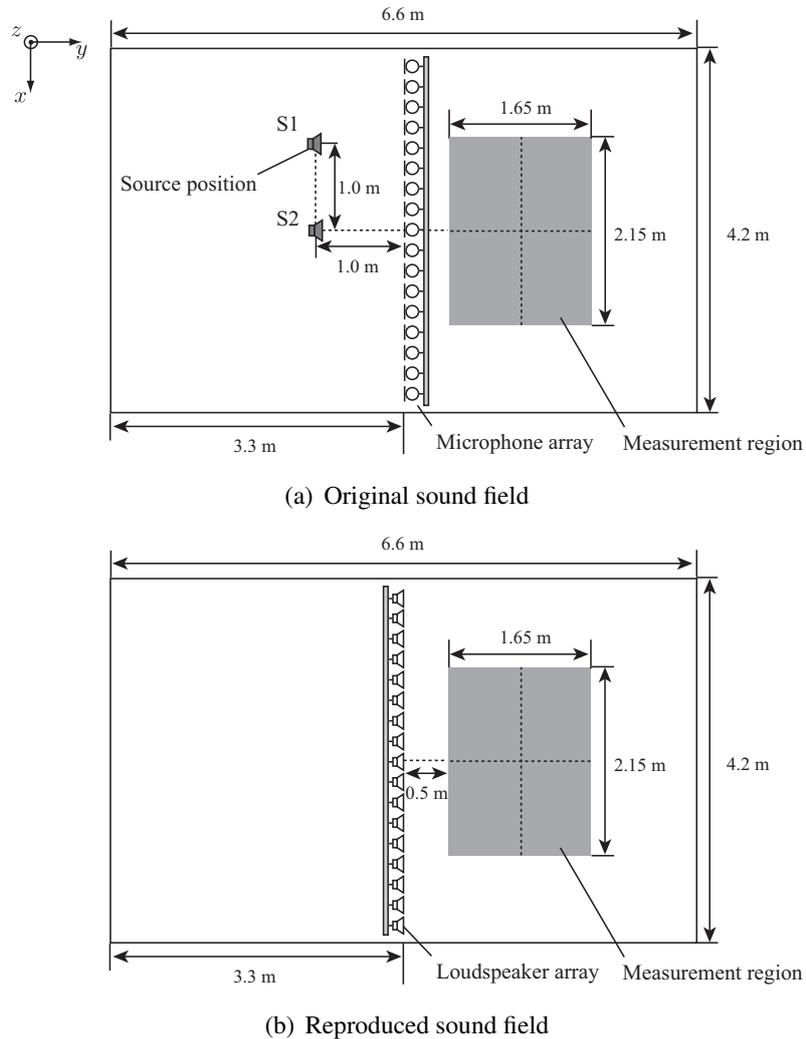
The delay times of the WFR filtering and data buffering were about 21 and 131 ms, respectively.



**Fig. 9.4.** A photo of sound field transmission system over IP network.

Therefore, the total delay time was about 152 ms. The rate of packet loss was 0.0% when the received data were observed for 600 sec.

Using this system makes it possible to transmit a sound field of about 4 m in real time. We also tested the system between two locations as far apart as about 55 km away (Fig. 9.4). In this case, the delay time of the IP network was about 1 ms. It was possible to transmit a sound field to any place that the network was connected.



**Fig. 9.5.** Measurement experiment setup.

## 9.2 Measurement of reproduced sound field

We conducted measurement experiments to compare the original and reproduced sound fields. The setup of the experiments is shown in Fig. 9.5. In these experiments, we only used the loudspeaker arrays of 64 channels as shown in Fig. 9.6 and did not use the IP network. The origin was set at the center of the room. Ordinary enclosed loudspeakers were used as primary sources and located at S1 and S2. Reverberation time ( $T_{60}$ ) of the experiment room was about 167 ms.

We measured impulse responses from each loudspeaker to grid points in a planar measurement region. These impulse responses were used to obtain sound pressure distribution of the original and reproduced sound fields by post-processing. The impulse responses were measured by using TSP signals [103] in  $2.2 \times 1.6$  m regions (shaded regions in Fig. 9.5) at intervals of 1.5 cm, i.e.,



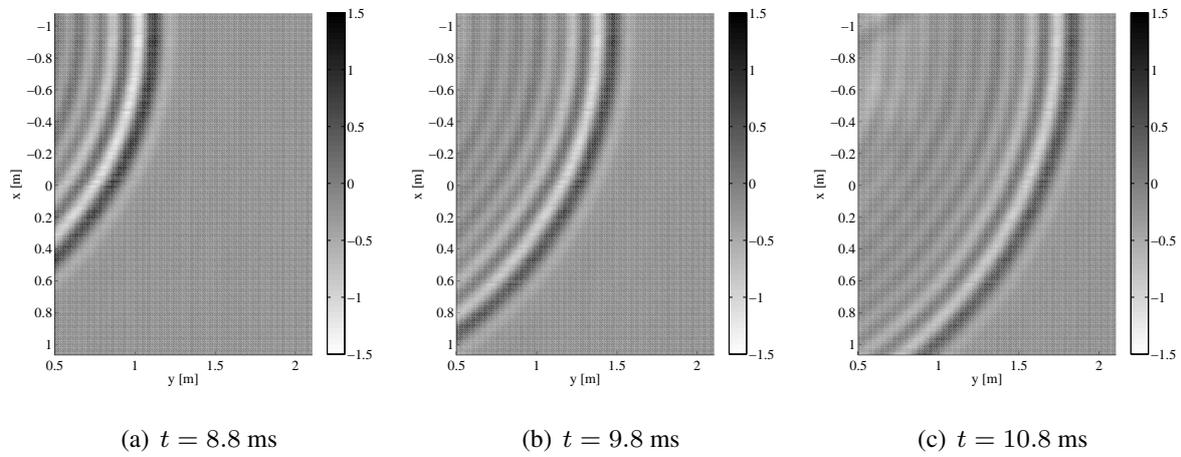
**Fig. 9.6.** Photo of measurement experiment.

144×108 points. The measurements were conducted in 4 divided regions by using automated measurement microphones as shown in Fig. 9.6. The signals of the linear microphone array were also obtained from measured impulse responses. The amplitudes and phases of the original and reproduced signals were normalized at the center of the measurement region, (0.0, 1.3) m.

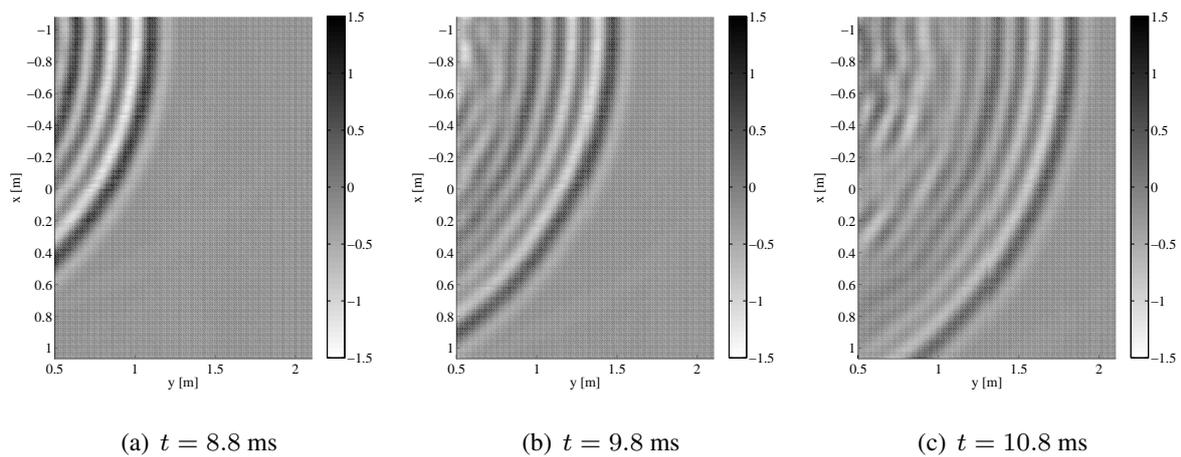
Figures 9.7 and 9.8 show the instantaneous sound pressure distributions of the original and reproduced sound fields in the measurement region, respectively. Here,  $t$  indicates the time. The primary source was located at S1 and the source signal was a pulse signal that was band-limited below 2.6 kHz in order to avoid spatial aliasing artifacts. It can be seen that the direct sound wave was propagating from S1 on the measurement region. This direct sound wave was correctly reproduced by using the WFR filtering method as shown in Fig. 9.8. The disturbed waves that occur after the reproduced direct sound wave are thought to be reflections from the frame structure of the array. When the source signal includes frequency spectrum components above 2.8 kHz, the reproduction accuracy of the direct sound wave cannot be maintained. Because the array configurations are linear, sound waves from vertical directions were not able to be reproduced.

Figure 9.9 plots the measured space-time signals of the original and reproduced sound fields along the line at  $y = 1.3$  m. The source position and signal were the same as those of the previous results. It can be seen that the reflections from side walls were also reproduced, for example, the pulse signal arrived at about  $t = 16$  ms. As discussed in Section 5.6, the reproduced reflections from the ceiling and floor seem to be different from the original ones.

Figure 9.10 shows the reproduced sound pressure distribution when the source position was S2 and the shift parameter of the WFR filter,  $(d_x, d_y)$  in Eq. (5.37), was set as (0.0, 2.0) m. The reproduced sound wave was focused at (0.0, 1.0) m as shown in Fig. 9.10 (b); therefore, a virtual sound source was created at this position as expected. When the listeners are in front of the virtual sound source, i.e.,  $y > 1.0$  m, it is expected that the listeners can localize the source at

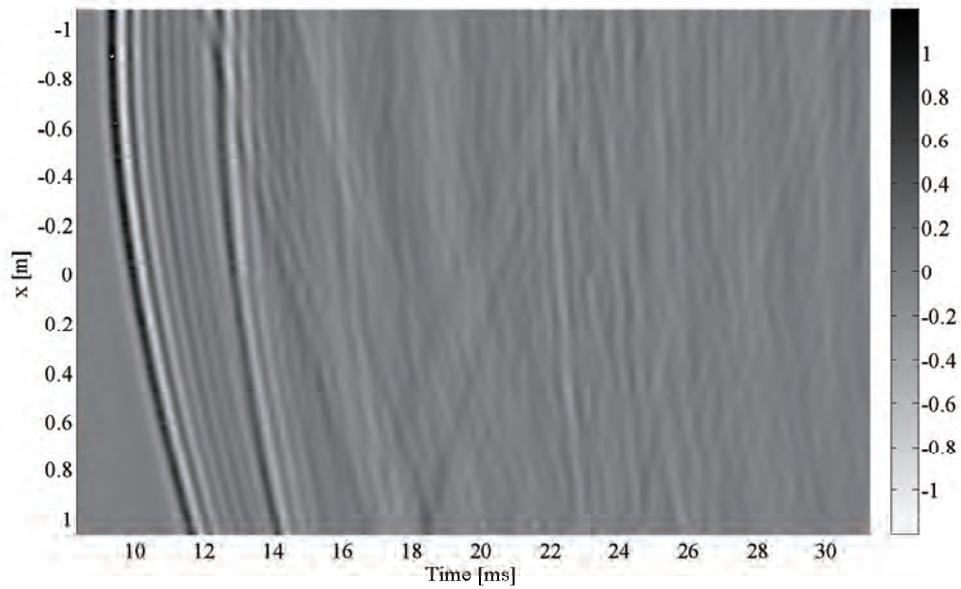


**Fig. 9.7.** Measured sound pressure distribution of original sound field.

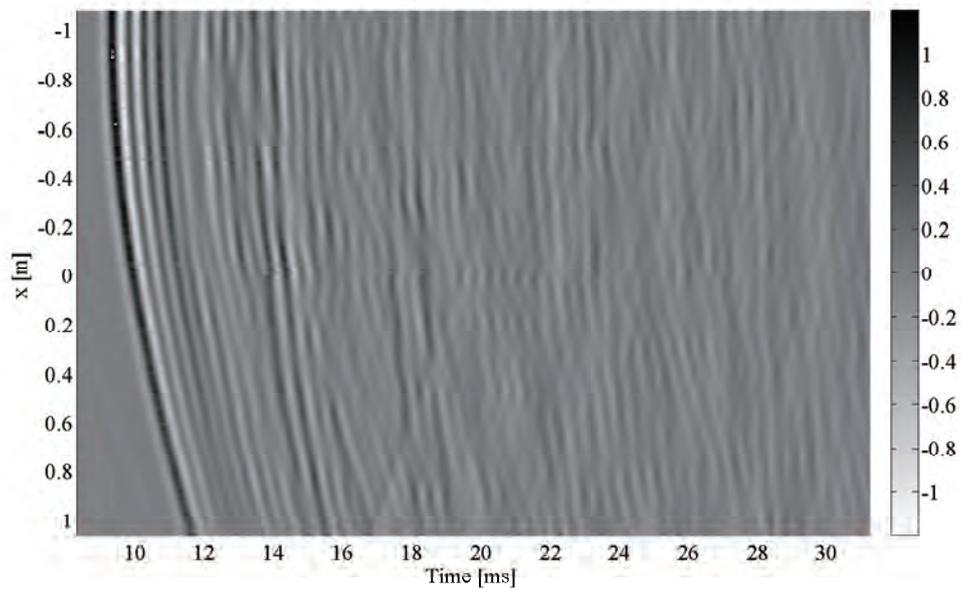


**Fig. 9.8.** Measured sound pressure distribution of reproduced sound field.

this position. It should be noted that finite array length approximation also makes the reproduced region smaller as discussed in Chapter 4.

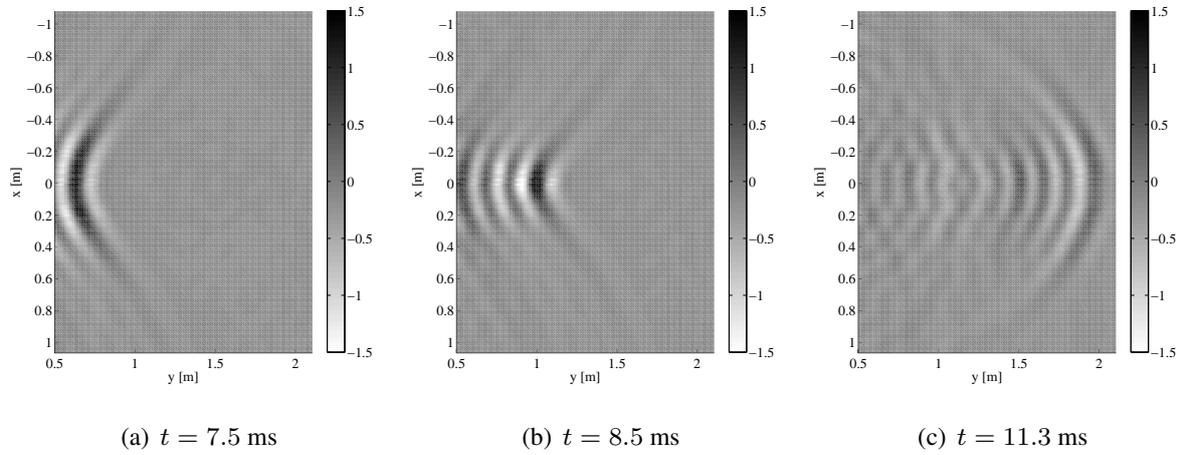


(a) Original



(b) Reproduced

**Fig. 9.9.** Measured space-time signal at  $y = 1.3$  m.



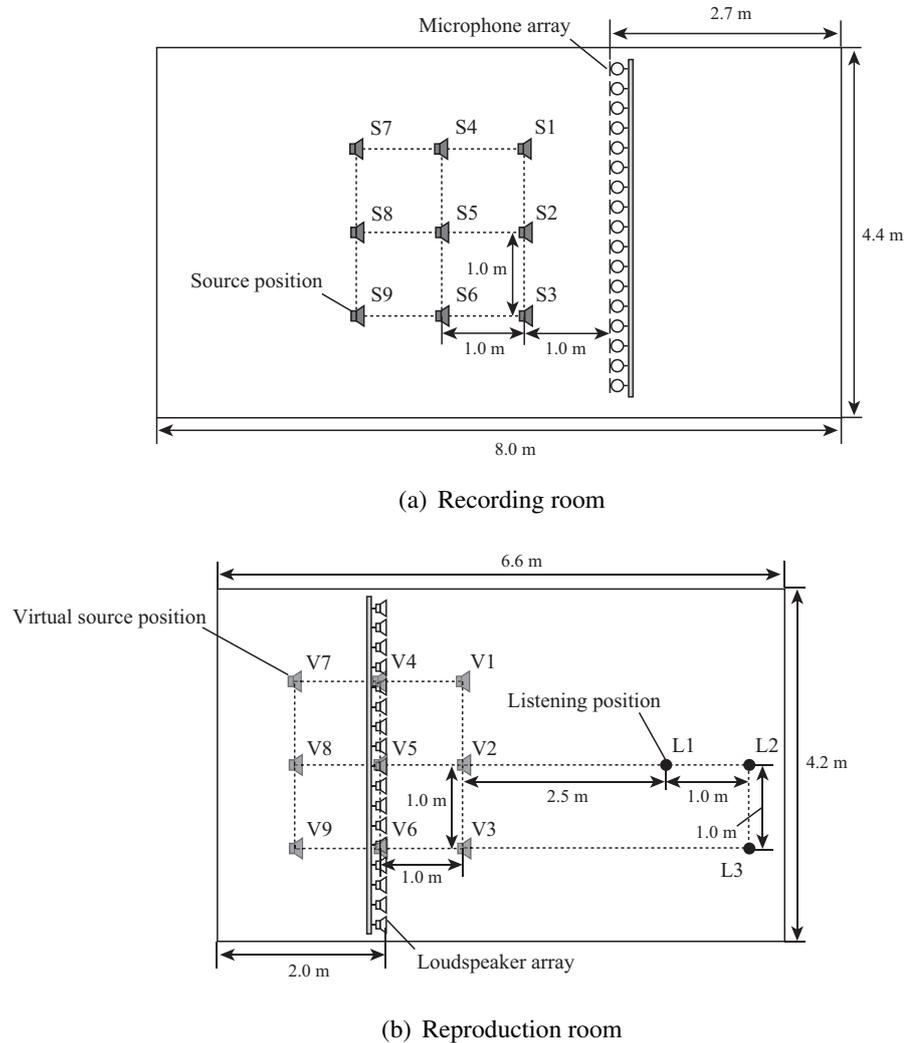
**Fig. 9.10.** Measured sound pressure distribution of reproduced sound field.

### 9.3 Subjective listening test of sound localization

To perceptually evaluate the proposed system, a sound localization listening test was conducted. Figure 9.11 shows the subjective listening test setup. The system described in the previous section was used for the experiments. The loudspeaker array was set about 1.2 m above the floor. The ordinary loudspeakers as primary sources were set at nine positions, S1 – S9, in the recording room. The shift parameter of the WFR filter,  $(d_x, d_y)$  in Eq. (5.37), was set as (0.0, 2.0) m. Therefore, reproduced sound sources were located at V1 – V9 in the reproduction room. As references, real loudspeakers of the same type used in the recording room were also set at V1 – V9. A listener was positioned at L1 – L3. There were 6 listeners between the ages of 25 and 40. A listener was able to move his or her head in the chair and see the positions signs at V1 – V9 (Fig. 9.12). The room sizes of the recording and reproduction rooms were  $4.4 \times 8.0 \times 3.0$  m and  $4.2 \times 6.6 \times 3.1$  m, respectively. The reverberation times ( $T_{60}$ ) were about 150 and 167 ms, respectively. The background noise level of the reproduction room was about 27.9 dBA.

The speech signals of female and male utterances were used as source signals. The source signals were recorded and played at a 48-kHz sampling rate. The duration of each source signal was 4 sec and its interval was 2 sec. The signals of real and virtual sound sources were played randomly 144 times at each listening position; therefore, the signal of each source was played 8 times. The order of the listening positions was L1, L2, and L3, and the listeners had a break when the listening position was changed. The listeners chose the perceived positions of the sound source from V1 – V9.

Figures 9.13, 9.14, and 9.15 shows the average rate of listeners noting each sound source location. The horizontal and vertical axes are target and perceived locations, respectively. Therefore, the accuracy rate is shown on the diagonal line. The sound source locations on the axes are sorted



**Fig. 9.11.** Subjective listening test setup.

vertically. The results of L1 (Fig. 9.13) show that the accuracies of real and virtual source localization were generally the same. The accuracy rate of the virtual sound sources at V1 – V3 was relatively low compared to that of the real sound sources. It seems that the real sound sources at short distances can be easily localized compared to virtual ones. Because it is difficult to distinguish the source locations in line with the front-facing position of the listener, i.e., V2, V3, and V8 [3], the accuracy rate of V5 and V8 for the real sound sources was distinctly low, and that for the virtual sound sources was higher. This is because the faster amplitude decay due to the linear array approximation, which leads to larger amplitude differences in distances, provides a clue to distinguish them. The results of L2 (Fig. 9.14) were of similar tendency to those of L1.

The results of L3 (Fig. 9.15) were also of similar tendency to those of L1. The accuracy rate of real sound sources at V6 and V9 was distinctly low because these sources were in line with the front-facing position of the listener. Because L3 was almost at the boundary of the reproduction



**Fig. 9.12.** Reproduction room for subjective listening test.

area of V1 for the virtual sound source as in Fig. 4.3, its accuracy rate was lower than that of the real one. These results indicated that the proposed system provides sound localization accuracy for virtual sound sources comparable to that for real sound sources in a large listening area.

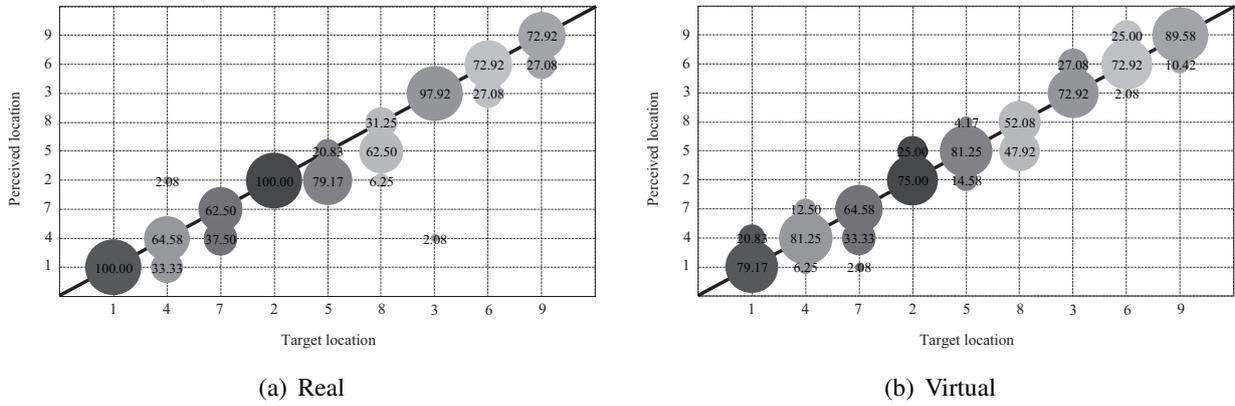


Fig. 9.13. Results of subjective listening test at L1 (%)

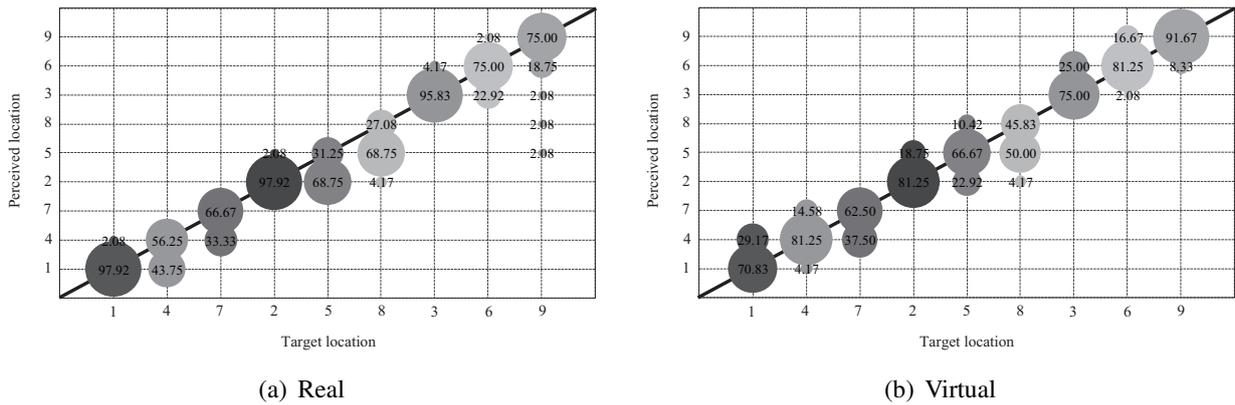


Fig. 9.14. Results of subjective listening test at L2 (%)

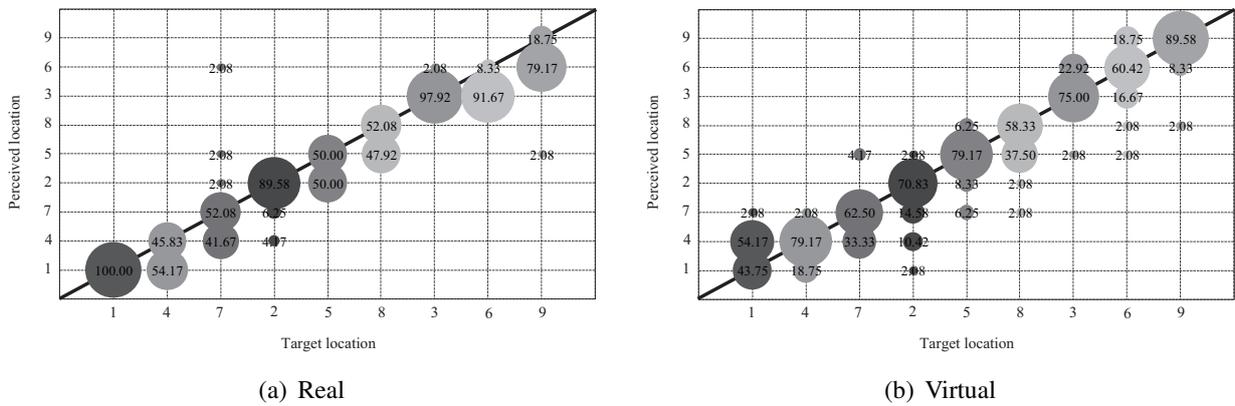
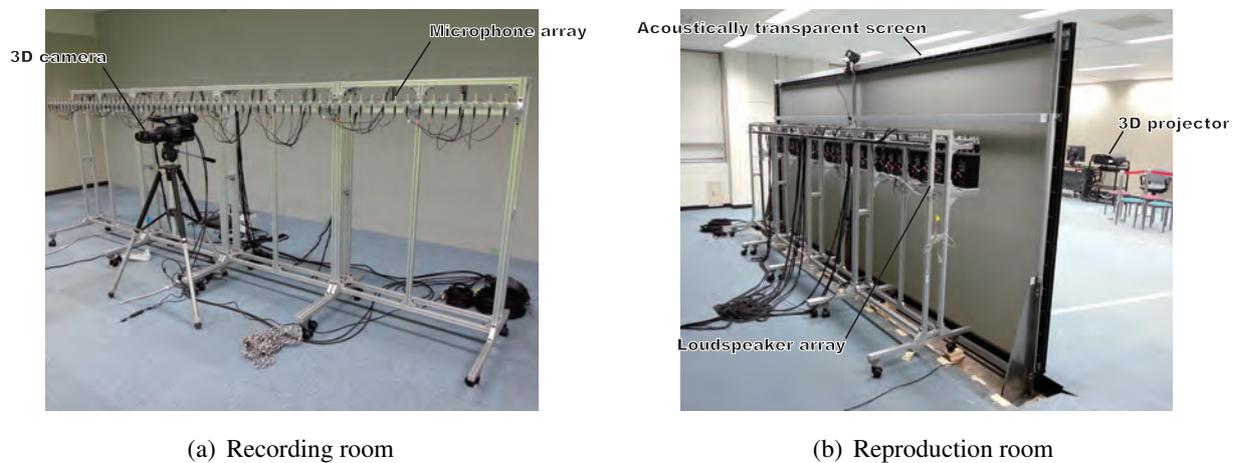
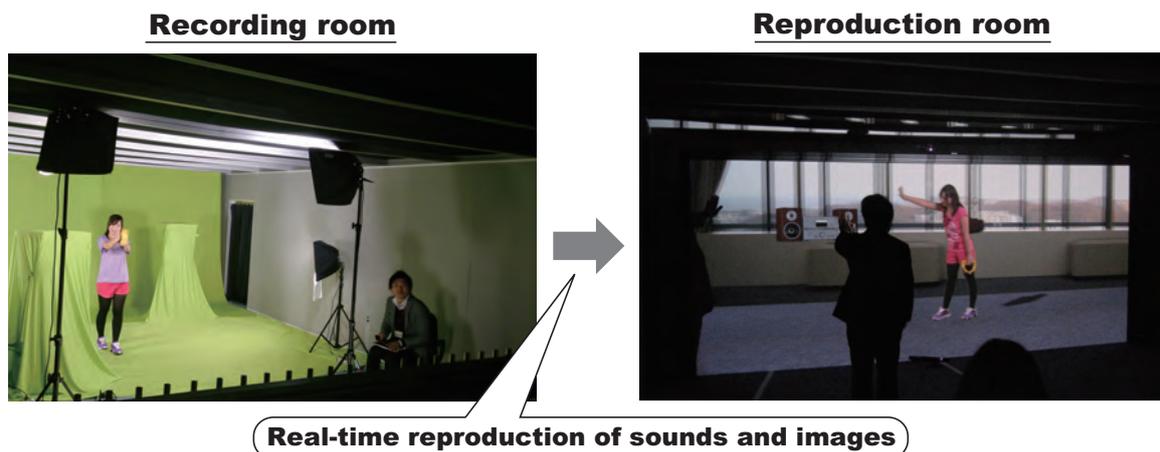


Fig. 9.15. Results of subjective listening test at L3 (%)



**Fig. 9.16.** System with 3D images.



**Fig. 9.17.** Photo of system with 3D images at NTT R&D forum 2013.

## 9.4 Demonstration system including 3D images

We developed a system combining the sound system described in Section 9.1 with 3D images. The system was composed of the linear arrays of microphones and loudspeakers, 3D camera and projector, and acoustically transparent screen (Fig 9.16). In order to generate images corresponding to a viewer's position, the position was detected by a sensor and the images were adaptively synthesized in real-time [105]. Therefore, sounds and images according to the viewer's position can be correctly presented by using this system. This system was exhibited at NTT R&D forum 2013 (Fig. 9.17).

## 9.5 Conclusions

A novel real-time sound field transmission system by using the WFR filter is developed. The system is composed of linear arrays of microphones and loudspeakers. The received signals transmitted over an IP network are transformed to the driving signals of the loudspeaker array by using the WFR filter for linear arrays. The measurement experiments were conducted and their results indicated that the sound field can be reproduced by using the actual system. The result of the subjective listening tests showed that this system can provide a sound localization accuracy for virtual sound sources comparable to that for real sound sources in a large listening area. We also introduced the demonstration system including 3D images.

## Chapter 10

# Direct Algebraic Method for Sound Source Localization

We deal with an estimation problem of directions and distances of sound sources in the three-dimensional space. The LC-PDE derived in Chapter 2 provides a necessary and sufficient description of sound field generated by a single point source at direction  $\mathbf{n}_{ps}$  and distance  $R_{ps}$ . In previous studies, spatial differential values of sound pressures in the LC-PDE are directly obtained and used for the estimation [25, 106, 107]. This approach has advantages in computational efficiency and temporal resolution; however, direct measurement of spatial differential values has several difficulties in robustness against measurement noise. To remove differentials, we apply a parameter estimation method referred to as weighted integral method [20] to the LC-PDE. The LC-PDE is integrated in a finite observation region with complex sinusoidal weighting functions. By using a well-known class of window function to eliminate the integral boundary terms, an exact finite set of algebraic equations is obtained including  $\mathbf{n}_{ps}$  and  $R_{ps}$  as the unknown variables and two-dimensional DFT components of the sound field as measurements. Recently, a similar method is developed for a circular microphone array [108].

### 10.1 Weighted finite integral of LC-PDE in space

We assume that the LC-PDE, Eq. (2.42), is satisfied on a two-dimensional finite observation region on the  $x$ - $y$ -plane,  $\Gamma$ . An observation value, i.e., sound pressure, at a position  $\mathbf{r} = (x, y, 0)$  is denoted as  $f(\mathbf{r}, t)$  ( $\mathbf{r} \in \Gamma$ ). Here,  $t$  is the time. By using an arbitrary spatial weighting function  $w(\mathbf{r})$ , the following identity is satisfied:

$$\begin{aligned} \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \partial_t f(\mathbf{r}, t) \right) \mathbf{n} &= \mathbf{0} \quad \forall \mathbf{r} \in \Gamma \\ \Leftrightarrow \iint_{\Gamma} \left\{ \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \partial_t f(\mathbf{r}, t) \right) \mathbf{n} \right\} w(\mathbf{r}) d\mathbf{r} &= \mathbf{0} \quad \forall w(\mathbf{r}), \end{aligned} \quad (10.1)$$

where the orientation vector  $\mathbf{n}_{ps}$  and distance  $R_{ps}$  of the sound source are replaced by  $\mathbf{n}$  and  $R$  respectively, for notational simplicity. We assume that the observation region  $\Gamma$  is sufficiently small compared to the sound source distance  $R$  so that  $\mathbf{n}$  and  $R$  are constant on  $\Gamma$ . Under this assumption, the right-hand side of Eq. (10.1) is calculated as:

$$\begin{aligned}
 & \iint_{\Gamma} w(\mathbf{r}) \left\{ \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \partial_t f(\mathbf{r}, t) \right) \mathbf{n} \right\} d\mathbf{r} \\
 &= \left[ \begin{array}{l} \iint_{\Gamma} \partial_x (w(\mathbf{r}) f(\mathbf{r}, t)) d\mathbf{r} - \iint_{\Gamma} f(\mathbf{r}, t) \partial_x w(\mathbf{r}) d\mathbf{r} \\ \iint_{\Gamma} \partial_y (w(\mathbf{r}) f(\mathbf{r}, t)) d\mathbf{r} - \iint_{\Gamma} f(\mathbf{r}, t) \partial_y w(\mathbf{r}) d\mathbf{r} \\ \partial_z \iint_{\Gamma} w(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r} \end{array} \right] \\
 &\quad - \frac{\mathbf{n}}{R} \iint_{\Gamma} w(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r} - \frac{\mathbf{n}}{c} \partial_t \iint_{\Gamma} w(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r} \\
 &= \left[ \begin{array}{l} \oint w(\mathbf{r}) f(\mathbf{r}, t) dy - g^{[x]}(t) \\ - \oint w(\mathbf{r}) f(\mathbf{r}, t) dx - g^{[y]}(t) \\ \partial_z g(t) \end{array} \right] - \frac{\mathbf{n}}{R} g(t) - \frac{\mathbf{n}}{c} \partial_t g(t) \\
 &= \mathbf{0}, \tag{10.2}
 \end{aligned}$$

where  $g(t)$ ,  $g^{[x]}(t)$ , and  $g^{[y]}(t)$  are defined as:

$$g(t) \equiv \iint_{\Gamma} w(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r} \tag{10.3}$$

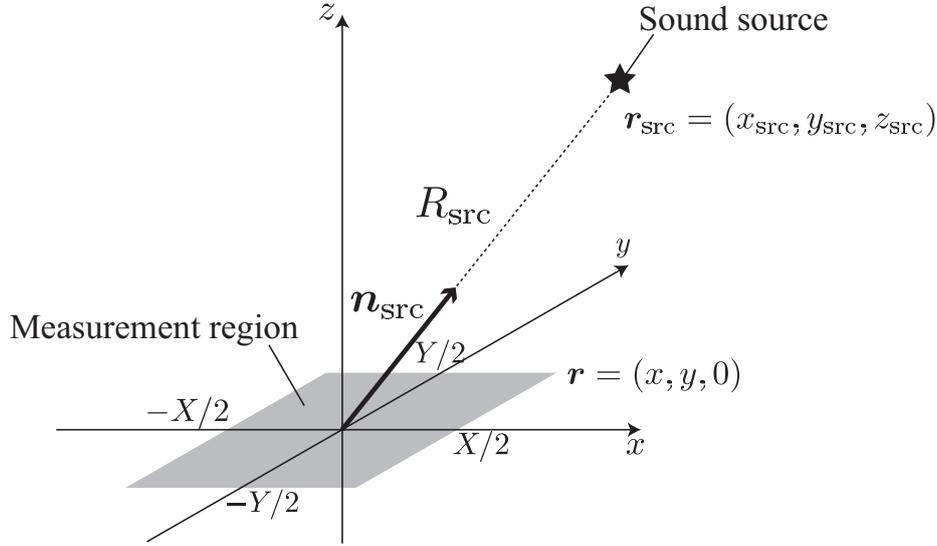
$$g^{[x]}(t) \equiv \iint_{\Gamma} \partial_x w(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r} \tag{10.4}$$

$$g^{[y]}(t) \equiv \iint_{\Gamma} \partial_y w(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r}. \tag{10.5}$$

The variables  $g(t)$ ,  $g^{[x]}(t)$ , and  $g^{[y]}(t)$  are the finite integral of  $f(\mathbf{r}, t)$  on  $\Gamma$  multiplying the spatial weighting function  $w(\mathbf{r})$  and its differential with respect to  $x$  and  $y$ ; therefore, these variables can be assumed as observed values. It is difficult to handle the weighted integral of boundary values,  $\oint w(\mathbf{r}) f(\mathbf{r}, t) dy$  and  $\oint w(\mathbf{r}) f(\mathbf{r}, t) dx$ , as observed values. However, these terms can be eliminated by two kinds of procedure:

1. Select a weighting function so that these boundary values equal to zero
2. Eliminate by combining multiple weighting functions

When these terms are eliminated, the unknown real variables,  $\mathbf{n}$  and  $R$ , can be algebraically derived.



**Fig. 10.1.** LC-PDE for rectangular measurement region.

### 10.1.1 Algebraic equation of source localization for rectangular observation region

#### Derivation by using window function

Now, let us assume the observation region is a rectangular region  $[-X/2, X/2] \times [-Y/2, Y/2]$  as shown in Fig. 10.1. We use a set of complex sinusoidal signals with a window function as the weighting function  $w(\mathbf{r})$ :

$$w(\mathbf{r}) = q(\mathbf{r})e^{-j(ux+vy)}, \quad (10.6)$$

where  $u = 2\pi n/X$  and  $v = 2\pi m/Y$  are the discrete spatial frequencies, and  $n$  and  $m$  are non-negative integers. Here,  $q(\mathbf{r})$  is a two-dimensional window function whose boundary value equals to zero as:

$$q(\pm X/2, \pm Y/2) = 0. \quad (10.7)$$

Therefore, the weighted integral of the observation values can be described as:

$$g_{uv}(t) = \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) q(\mathbf{r}) e^{-j(ux+vy)} dx \quad (10.8)$$

$$g_{uv}^{[x]}(t) = \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) \{(\partial_x - ju)q(\mathbf{r})\} e^{-j(ux+vy)} dx \quad (10.9)$$

$$g_{uv}^{[y]}(t) = \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) \{(\partial_y - jv)q(\mathbf{r})\} e^{-j(ux+vy)} dy. \quad (10.10)$$

The weighted integral of the boundary values are eliminated by the window function as:

$$\begin{bmatrix} -g_{uv}^{[x]}(t) \\ -g_{uv}^{[y]}(t) \\ \partial_z g_{uv}(t) \end{bmatrix} - \frac{\mathbf{n}}{R} g_{uv}(t) - \frac{\mathbf{n}}{c} \partial_t g_{uv}(t) = \mathbf{0}. \quad (10.11)$$

Because it is difficult to obtain  $\partial_z g_{uv}(t)$ , we describe the equation to be solved as:

$$\begin{bmatrix} -g_{uv}^{[x]}(t) \\ -g_{uv}^{[y]}(t) \end{bmatrix} = g_{uv}(t) \begin{bmatrix} n_x/R \\ n_y/R \end{bmatrix} + \frac{\partial_t g_{uv}(t)}{c} \begin{bmatrix} n_x \\ n_y \end{bmatrix}. \quad (10.12)$$

This equation includes four equations combining real and imaginary parts; therefore, the four unknown real values,  $n_x/R$ ,  $n_y/R$ ,  $n_x$ , and  $n_y$ , can be uniquely solved. Because  $\partial_z g_{uv}(t)$  is unknown, it is difficult to determine whether  $n_z = \pm\sqrt{1 - n_x^2 - n_y^2}$ , i.e.  $z$ -coordinate of the sound source location, is positive or negative. Eq. (10.12) can be solved by using  $g_{uv}(t)$ ,  $g_{uv}^{[x]}(t)$ ,  $g_{uv}^{[y]}(t)$ , and  $\partial_t g_{uv}(t)$  that can be obtained from an instantaneous observation value. In addition, Eqs. (10.8), (10.9), and (10.10) are close to two-dimensional Fourier transform of  $f(\mathbf{r}, \omega)$  with the window function. Therefore, these values can be calculated by using the 2D DFT with compensation in practical implementations.

### Application of Hann window

As an specific example of a window function  $q(\mathbf{r})$ , we consider the Hann window described as:

$$q(x, y) = (1 + \cos \Delta u x)(1 + \cos \Delta v y) \quad (10.13)$$

$$\partial_x q(x, y) = -\Delta u (1 + \cos \Delta v y) \sin \Delta u x \quad (10.14)$$

$$\partial_y q(x, y) = -\Delta v (1 + \cos \Delta u x) \sin \Delta v y, \quad (10.15)$$

where  $\Delta u \equiv 2\pi/X$  and  $\Delta v \equiv 2\pi/Y$ . These equations can be rewritten by using exponential as:

$$q(x, y) = \frac{1}{4} (2 + e^{j\Delta u x} + e^{-j\Delta u x}) (2 + e^{j\Delta v y} + e^{-j\Delta v y}) \quad (10.16)$$

$$\partial_x q(x, y) = -\frac{\Delta u}{4j} (2 + e^{j\Delta v y} + e^{-j\Delta v y}) (e^{j\Delta u x} - e^{-j\Delta u x}) \quad (10.17)$$

$$\partial_y q(x, y) = -\frac{\Delta v}{4j} (2 + e^{j\Delta u x} + e^{-j\Delta u x}) (e^{j\Delta v y} - e^{-j\Delta v y}) \quad (10.18)$$

The weighted integral of the observation value without the window function is defined as:

$$h_{nm} \equiv \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}) e^{-j(\Delta u n x + \Delta v m y)} dx dy. \quad (10.19)$$

Hereafter, the temporal variable  $t$  is omitted for simplicity. By substituting Eqs. (10.16), (10.17), and (10.18) into Eqs. (10.8), (10.9), and (10.10),  $g_{nm}$ ,  $g_{nm}^{[x]}$ , and  $g_{nm}^{[y]}$ , i.e., discrete notations of

$g_{uv}$ ,  $g_{uv}^{[x]}$ , and  $g_{uv}^{[y]}$ , can be described by using  $h_{nm}$  as:

$$\begin{aligned}
g_{nm} &= \frac{1}{4} \iint f(\mathbf{r})(2 + e^{j\Delta ux} + e^{-j\Delta ux})(2 + e^{j\Delta vy} + e^{-j\Delta vy})e^{-j(\Delta unx + \Delta vmy)} dx dy \\
&= h_{nm} + \frac{1}{2}(h_{n-1m} + h_{n+1m} + h_{nm-1} + h_{nm+1}) \\
&\quad + \frac{1}{4}(h_{n-1m-1} + h_{n+1m-1} + h_{n-1m+1} + h_{n+1m+1})
\end{aligned} \tag{10.20}$$

$$\begin{aligned}
g_{nm}^{[x]} &= -\frac{\Delta u}{4j} \iint f(\mathbf{r})(2 + e^{j\Delta vy} + e^{-j\Delta vy}) \\
&\quad \cdot (e^{j\Delta ux} - e^{-j\Delta ux})e^{-j(\Delta unx + \Delta vmy)} dx dy - j u g_{nm} \\
&= -\frac{\Delta u}{4j} (2h_{n-1m} + h_{n-1m-1} + h_{n-1m+1} \\
&\quad - 2h_{n+1m} - h_{n+1m-1} - h_{n+1m+1}) - j u g_{nm}
\end{aligned} \tag{10.21}$$

$$\begin{aligned}
g_{nm}^{[y]} &= -\frac{\Delta v}{4j} \iint f(\mathbf{r})(2 + e^{j\Delta ux} + e^{-j\Delta ux}) \\
&\quad \cdot (e^{j\Delta vy} - e^{-j\Delta vy})e^{-j(\Delta unx + \Delta vmy)} dx dy - j v g_{nm} \\
&= -\frac{\Delta v}{4j} (2h_{nm-1} + h_{n-1m-1} + h_{n+1m-1} \\
&\quad - 2h_{nm+1} - h_{n-1m+1} - h_{n+1m+1}) - j v g_{nm}
\end{aligned} \tag{10.22}$$

Therefore, the weighted integrals with the window function,  $g_{nm}$ ,  $g_{nm}^{[x]}$ , and  $g_{nm}^{[y]}$  can be obtained by using the weighted integrals without the window function,  $h_{nm}$ . Adjacent  $3 \times 3$  frequencies of  $h_{nm}$  are only required to calculate  $g_{nm}$ ,  $g_{nm}^{[x]}$ , and  $g_{nm}^{[y]}$ . These frequencies should be chosen from around a peak of  $h_{nm}$  because these components include a lot of information on the source location.

This solution by using the Hann window is equivalent to simultaneously solve Eq. (10.2) by using adjacent  $3 \times 3$  frequencies for eliminating the weighted integral of the boundary values. When the Blackman window is used as  $q(\mathbf{r})$ ,  $g_{nm}$ ,  $g_{nm}^{[x]}$ , and  $g_{nm}^{[y]}$  can be calculated by using adjacent  $5 \times 5$  frequencies of  $h_{nm}$ .

### 10.1.2 Algorithm for estimating sound source locations

A sound source localization algorithm based on the method presented in Section 10.1.1 is summarized in Algorithm 1. A peak of  $h_{nm}$  by a single sound source can be assumed to be locally-concentrated; therefore, multiple sound source locations can be estimated by calculating  $\mathbf{n}$  and  $R$  from each peak of  $h_{nm}$ . This algorithm is computationally efficient because the 2D DFT of the observation sensor data and algebra operations are only required to estimate sound source

---

**Algorithm 1** Estimate  $n_x$ ,  $n_y$ , and  $R$  of multiple sound sources

---

$f(x, y) \leftarrow$  observation sensor data on a rectangular region  
 $\partial_t f(x, y) \leftarrow$  numerical differential of  $f(x, y)$  in time  
 $h_{nm} \leftarrow$  2D DFT of  $f(x, y)$   
 $\partial_t h_{nm} \leftarrow$  2D DFT of  $\partial_t f(x, y)$   
**for all** peak of  $h_{nm}$  **do**  
    compute  $g_{nm}$ ,  $g_{nm}^{[x]}$ , and  $g_{nm}^{[y]}$  by using adjacent  $3 \times 3$  frequencies of  $h_{nm}$   
    compute  $\partial_t g_{nm}$  by using adjacent  $3 \times 3$  frequencies of  $\partial_t h_{nm}$   
     $n_x, n_y$ , and  $R \leftarrow$  solve Eq. (10.12)  
**end for**

---

locations. In addition, instantaneous observation sensor data is only required, and sound source locations can be obtained every time sample. Therefore, this algorithm is useful, especially when sound sources move fast.

Because  $h_{nm}$  is the weighted finite integral of the observation values as in Eq. (10.19), a compensation for the 2D DFT is required. When the 2D DFT of  $f(x, y)$  is denoted as  $h_{nm}^D$ , the compensation is achieved by half sample phase shift as [20]:

$$h_{nm} \simeq h_{nm}^D e^{-j(u\Delta x + v\Delta y)/2}. \quad (10.23)$$

As shown in Algorithm 1, it is necessary to obtain the temporal differential of  $f(x, y)$  and  $\partial_t f(x, y)$ . A simple way to obtain  $\partial_t f(x, y)$  is to use a finite difference approximation. However, Eq. (10.12) includes the weighted integrals of both  $f(x, y)$  and  $\partial_t f(x, y)$ . When  $\partial_t f(x, y)$  is calculated by using a finite difference approximation, an inconsistency between  $f(x, y)$  and  $\partial_t f(x, y)$  causes an estimation error [109]. Therefore,  $f(x, y)$  and  $\partial_t f(x, y)$  are obtained by using smoothing and differential filters, respectively. These filters are derived by minimizing inconsistency between them. When the filter length is 5 taps, the filters can be numerically derived as:

$$(0.0354596, 0.248928, 0.431224, 0.248928, 0.0354596) \quad (10.24)$$

$$(-0.108237, -0.283525, 0.000000, 0.283525, 0.108237) \quad (10.25)$$

The variables  $f(x, y)$  and  $\partial_t f(x, y)$  are obtained by applying these filters to observation sensor data.

In practice, Eq. (10.12) can be analytically solved as:

$$n_x = \frac{\Im\{g_{nm}\}\Re\{g_{nm}^{[x]}\} - \Re\{g_{nm}\}\Im\{g_{nm}^{[x]}\}}{\Re\{g_{nm}\}\Im\{\partial_t g_{nm}/c\} - \Re\{\partial_t g_{nm}/c\}\Im\{g_{nm}\}} \quad (10.26)$$

$$n_y = \frac{\Im\{g_{nm}\}\Re\{g_{nm}^{[y]}\} - \Re\{g_{nm}\}\Im\{g_{nm}^{[y]}\}}{\Re\{g_{nm}\}\Im\{\partial_t g_{nm}/c\} - \Re\{\partial_t g_{nm}/c\}\Im\{g_{nm}\}} \quad (10.27)$$

The distance  $R$  can be obtained in two ways as:

$$R = \frac{\Im\{g_{nm}\}\Re\{g_{nm}^{[x]}\} - \Re\{g_{nm}\}\Im\{g_{nm}^{[x]}\}}{\Re\{g_{nm}^{[x]}\}\Im\{\partial_t g_{nm}/c\} - \Re\{\partial_t g_{nm}/c\}\Im\{g_{nm}^{[x]}\}} \quad (10.28)$$

and

$$R = \frac{\Im\{g_{nm}\}\Re\{g_{nm}^{[y]}\} - \Re\{g_{nm}\}\Im\{g_{nm}^{[y]}\}}{\Re\{g_{nm}^{[y]}\}\Im\{\partial_t g_{nm}/c\} - \Re\{\partial_t g_{nm}/c\}\Im\{g_{nm}^{[y]}\}}. \quad (10.29)$$

Therefore,  $n$  and  $R$  can be obtained as direct solutions.

## 10.2 Weighted finite integral of LC-PDE in space and time

In the previous section, the weighted integral method is applied to LC-PDE only in the spatial domain. Therefore, the resulting equation for source localization includes temporal differentials of observed signals. We apply the weighted integral method in both the spatial and temporal domains; therefore, the temporal differentials may be replaced by a finite integral of observed signals.

We assume that the LC-PDE is satisfied on a two-dimensional observation region on the  $x$ - $y$ -plane,  $\Gamma$ , and an observation time interval,  $T$ . An observation value at a position  $\mathbf{r} = (x, y, 0)$  and time  $t$  is denoted as  $f(\mathbf{r}, t)$  ( $\mathbf{r} \in \Gamma$ ,  $t \in T$ ). By using an arbitrary spatio-temporal weighting function  $W(\mathbf{r}, t)$ , the following identity is satisfied:

$$\begin{aligned} \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right) \mathbf{n} &= 0 \quad \forall t \in T, \forall \mathbf{r} \in \Gamma \\ \Leftrightarrow \int_T dt \iint_{\Gamma} d\mathbf{r} \left\{ \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right) \mathbf{n} \right\} W(\mathbf{r}, t) &= 0 \quad \forall W(\mathbf{r}, t) \end{aligned} \quad (10.30)$$

We assume that the weighting function  $W(\mathbf{r}, t)$  can be separated as  $W(\mathbf{r}, t) = w_T(t)w_{\Gamma}(\mathbf{r})$ . In a similar way to the spatial weighted integral case, the right-hand side of Eq. (10.30) is calculated as:

$$\begin{aligned} & \int_T \iint_{\Gamma} w_{\Gamma}(\mathbf{r}) w_T(t) \left\{ \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right) \mathbf{n} \right\} dt d\mathbf{r} \\ &= \int_T w_T(t) \left[ \begin{aligned} & \iint_{\Gamma} \partial_x (w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t)) d\mathbf{r} - \iint_{\Gamma} f(\mathbf{r}, t) \partial_x w_{\Gamma}(\mathbf{r}) d\mathbf{r} \\ & \iint_{\Gamma} \partial_y (w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t)) d\mathbf{r} - \iint_{\Gamma} f(\mathbf{r}, t) \partial_y w_{\Gamma}(\mathbf{r}) d\mathbf{r} \\ & \partial_z \iint_{\Gamma} w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) d\mathbf{r} \end{aligned} \right] dt \\ & \quad - \frac{\mathbf{n}}{R} \int_T \iint_{\Gamma} w_T(t) w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt d\mathbf{r} \\ & \quad - \frac{\mathbf{n}}{c} \iint_{\Gamma} w_{\Gamma}(\mathbf{r}) \left\{ [w_T(t) f(\mathbf{r}, t)]_T - \int_T \partial_t w_T(t) f(\mathbf{r}, t) dt \right\} d\mathbf{r} \\ &= \left[ \begin{aligned} & \int_T \oint w_T(t) w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt dy - G^{[x]} \\ & - \int_T \oint w_T(t) w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt dx - G^{[y]} \\ & \partial_z G \end{aligned} \right] \\ & \quad - \frac{\mathbf{n}}{R} G - \frac{\mathbf{n}}{c} \left\{ \iint_{\Gamma} w_{\Gamma}(\mathbf{r}) [w_T(t) f(\mathbf{r}, t)]_T d\mathbf{r} - G^{[t]} \right\} \\ &= \mathbf{0} \end{aligned} \quad (10.31)$$

where  $G$ ,  $G^{[x]}$ ,  $G^{[y]}$ , and  $G^{[t]}$  are defined as:

$$G \equiv \int_T \iint_{\Gamma} w_T(t) w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt d\mathbf{r} \quad (10.32)$$

$$G^{[x]} \equiv \int_T \iint_{\Gamma} w_T(t) \partial_x w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt d\mathbf{r} \quad (10.33)$$

$$G^{[y]} \equiv \int_T \iint_{\Gamma} w_T(t) \partial_y w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt d\mathbf{r} \quad (10.34)$$

$$G^{[t]} \equiv \int_T \iint_{\Gamma} \partial_t w_T(t) w_{\Gamma}(\mathbf{r}) f(\mathbf{r}, t) dt d\mathbf{r} \quad (10.35)$$

Again,  $\mathbf{n}$  and  $R$  are assumed to be constant on  $\Gamma$  during  $T$ . The variables  $G$ ,  $G^{[x]}$ ,  $G^{[y]}$ , and  $G^{[t]}$  are the finite integral on  $\Gamma$  and  $T$  multiplying the spatio-temporal weighting function  $w_T(t)w_{\Gamma}(\mathbf{r})$  and its differential with respect to  $x$ ,  $y$ , and  $t$ ; therefore, these variables can be assumed as observed values. It is difficult to handle the weighted integral of boundary values,  $\int_T \oint w_T(t)w_{\Gamma}(\mathbf{r})f(\mathbf{r}, t)dt dy$ ,  $\int_T \oint w_T(t)w_{\Gamma}(\mathbf{r})f(\mathbf{r}, t)dt dx$ , and  $\iint_{\Gamma} w_{\Gamma}(\mathbf{r})[w_T(t)f(\mathbf{r}, t)]_T d\mathbf{r}$ , as observed values. These terms can be eliminated in a similar manner to the spatial weighted integral case, and the unknown variables,  $\mathbf{n}$  and  $R$  can be algebraically derived.

### 10.2.1 Algebraic equation of source localization for rectangular measurement region

#### Derivation by using window function

Let us assume the observation time interval is  $[-T/2, T/2]$  and the observation region is a rectangular region  $[-X/2, X/2] \times [-Y/2, Y/2]$ . We use a set of complex sinusoidal signals with a window function as the weighting functions  $w_T(t)$  and  $w_{\Gamma}(\mathbf{r})$ :

$$w_T(t) = q_T(t) e^{-j\omega t} \quad (10.36)$$

$$w_{\Gamma}(\mathbf{r}) = q_{\Gamma}(\mathbf{r}) e^{-j(ux+vy)}, \quad (10.37)$$

where  $\omega = 2\pi n/T$  is the discrete temporal frequency,  $u = 2\pi m/X$ , and  $v = 2\pi l/Y$  are the discrete spatial frequencies, and  $n$ ,  $m$ , and  $l$  are non-negative integers. Here,  $q_T(t)$  and  $q_{\Gamma}(\mathbf{r})$  are one-dimensional and two-dimensional window functions, respectively, whose boundary values equals to zero as:

$$q_T(\pm T/2) = 0 \quad (10.38)$$

$$q_{\Gamma}(\pm X/2, \pm Y/2) = 0. \quad (10.39)$$

Therefore, the weighted integral of the observation values can be described as:

$$G_{\omega uv} = \int_{-T/2}^{T/2} \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) q_T(t) q_\Gamma(\mathbf{r}) e^{-j(\omega t + ux + vy)} dt d\mathbf{r} \quad (10.40)$$

$$G_{\omega uv}^{[t]} = \int_{-T/2}^{T/2} \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) \{(\partial_t - j\omega) q_T(t)\} q_\Gamma(\mathbf{r}) e^{-j(\omega t + ux + vy)} dt d\mathbf{r} \quad (10.41)$$

$$G_{\omega uv}^{[x]} = \int_{-T/2}^{T/2} \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) q_T(t) \{(\partial_x - ju) q_\Gamma(\mathbf{r})\} e^{-j(\omega t + ux + vy)} dt d\mathbf{r} \quad (10.42)$$

$$G_{\omega uv}^{[y]} = \int_{-T/2}^{T/2} \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) q_T(t) \{(\partial_y - jv) q_\Gamma(\mathbf{r})\} e^{-j(\omega t + ux + vy)} dt d\mathbf{r} \quad (10.43)$$

The weighted integral of the boundary values are eliminated by the window function as:

$$\begin{bmatrix} -G_{\omega uv}^{[x]} \\ -G_{\omega uv}^{[y]} \\ \partial_z G_{\omega uv} \end{bmatrix} - \frac{\mathbf{n}}{R} G_{\omega uv} - \frac{\mathbf{n}}{c} G_{\omega uv}^{[t]} = \mathbf{0} \quad (10.44)$$

Because it is difficult to obtain  $\partial_z G_{\omega uv}$ , we describe the equation to be solved as:

$$\begin{bmatrix} -G_{\omega uv}^{[x]} \\ -G_{\omega uv}^{[y]} \end{bmatrix} = G_{\omega uv} \begin{bmatrix} n_x/R \\ n_y/R \end{bmatrix} + \frac{G_{\omega uv}^{[t]}}{c} \begin{bmatrix} n_x \\ n_y \end{bmatrix} \quad (10.45)$$

This equation includes four equations combining real and imaginary parts; therefore, the four unknown real values,  $n_x/R$ ,  $n_y/R$ ,  $n_x$ , and  $n_y$  can be uniquely solved. Because  $\partial_z G_{\omega uv}$  is unknown, it is difficult to determine whether  $n_z = \pm \sqrt{1 - n_x^2 - n_y^2}$ , i.e.,  $z$ -coordinate of the sound source location, is positive or negative. Eq. (10.45) can be solved by using  $G_{\omega uc}$ ,  $G_{\omega uv}^{[t]}$ ,  $G_{\omega uv}^{[x]}$ , and  $G_{\omega uv}^{[y]}$  that can be obtained from a finite observation value. In addition, Eqs. (10.40), (10.41), (10.42), and (10.43) are close to three-dimensional Fourier transform of  $f(\mathbf{r}, t)$  with the window function. Therefore, these values can be calculated by using the 3D DFT with compensation in practical implementations.

### Application of Hann window

As an specific example of window functions  $q_T(t)$  and  $q_\Gamma(\mathbf{r})$ , we consider the 1D and 2D Hann windows as:

$$q_T(t) = 1 + \cos \Delta\omega t \quad (10.46)$$

$$q_\Gamma(x, y) = (1 + \cos \Delta ux)(1 + \cos \Delta vy) \quad (10.47)$$

$$\partial_t q_T(t) = -\Delta\omega \sin \Delta\omega t \quad (10.48)$$

$$\partial_x q_\Gamma(x, y) = -\Delta u(1 + \cos \Delta vy) \sin \Delta ux \quad (10.49)$$

$$\partial_y q_\Gamma(x, y) = -\Delta v(1 + \cos \Delta ux) \sin \Delta vy \quad (10.50)$$

where  $\Delta\omega \equiv 2\pi/T$ ,  $\Delta u \equiv 2\pi/X$ , and  $\Delta v \equiv 2\pi/Y$ . These equations can be rewritten by using exponential as:

$$q_T(t) = 1 + \frac{1}{2}(e^{j\Delta\omega t} + e^{-j\Delta\omega t}) \quad (10.51)$$

$$q_\Gamma(x, y) = \frac{1}{4}(2 + e^{j\Delta ux} + e^{-j\Delta ux})(2 + e^{j\Delta vy} + e^{-j\Delta vy}) \quad (10.52)$$

$$\partial_t q_T(t) = j\frac{\Delta\omega}{2}(e^{j\Delta\omega t} - e^{-j\Delta\omega t}) \quad (10.53)$$

$$\partial_x q_\Gamma(x, y) = j\frac{\Delta u}{4}(2 + e^{j\Delta vy} + e^{-j\Delta vy})(e^{j\Delta ux} - e^{-j\Delta ux}) \quad (10.54)$$

$$\partial_y q_\Gamma(x, y) = j\frac{\Delta v}{4}(2 + e^{j\Delta ux} + e^{-j\Delta ux})(e^{j\Delta vy} - e^{-j\Delta vy}). \quad (10.55)$$

The weighted integral of the observation value without a window function is defined as:

$$H_{nml} = \int_{-T/2}^{T/2} \int_{-X/2}^{X/2} \int_{-Y/2}^{Y/2} f(\mathbf{r}, t) e^{-j(\Delta\omega nt + \Delta umx + \Delta vly)} dt dx dy \quad (10.56)$$

By substituting Eqs. (10.51), (10.52), (10.53), (10.54), and (10.55) into Eqs. (10.40), (10.41), (10.42), and (10.43),  $G_{nml}$ ,  $G_{nml}^{[t]}$ ,  $G_{nml}^{[x]}$ , and  $G_{nml}^{[y]}$ , i.e., discrete notations of  $G_{\omega uv}$ ,  $G_{\omega uv}^{[t]}$ ,  $G_{\omega uv}^{[x]}$ , and  $G_{\omega uv}^{[y]}$ , can be described by using  $H_{nml}$  as:

$$\begin{aligned} G_{nml} = & H_{nml} + \frac{1}{2}(H_{n-1ml} + H_{n+1ml} \\ & + H_{nm-1l} + H_{nm+1l} + H_{nml-1} + H_{nml+1}) \\ & + \frac{1}{4}(H_{nm-1l-1} + H_{nm+1l-1} + H_{nm-1l+1} + H_{nm+1l+1} \\ & + H_{n-1m-1l} + H_{n-1m+1l} + H_{n-1ml-1} + H_{n-1ml+1} \\ & + H_{n+1m-1l} + H_{n+1m+1l} + H_{n+1ml-1} + H_{n+1ml+1}) \\ & + \frac{1}{8}(H_{n-1m-1l-1} + H_{n-1m+1l-1} + H_{n-1m-1l+1} + H_{n-1m+1l+1} \\ & + H_{n+1m-1l-1} + H_{n+1m+1l-1} + H_{n+1m-1l+1} + H_{n+1m+1l+1}) \end{aligned} \quad (10.57)$$

$$\begin{aligned}
 G_{nml}^{[t]} = j \frac{\Delta\omega}{2} & \left\{ H_{n-1ml} - H_{n+1ml} \right. \\
 & + \frac{1}{2} (H_{n-1m-1l} + H_{n-1m+1l} + H_{n-1ml-1} + H_{n-1ml+1} \\
 & \quad - H_{n+1m-1l} - H_{n+1m+1l} - H_{n+1ml-1} - H_{n+1ml+1}) \\
 & + \frac{1}{4} (H_{n-1m-1l-1} + H_{n-1m+1l-1} + H_{n-1m-1l+1} + H_{n-1m+1l+1} \\
 & \quad - H_{n+1m-1l-1} - H_{n+1m+1l-1} - H_{n+1m-1l+1} - H_{n+1m+1l+1}) \left. \right\} \\
 & - j\omega G_{nml} \quad (10.58)
 \end{aligned}$$

$$\begin{aligned}
 G_{nml}^{[x]} = j \frac{\Delta u}{2} & \left\{ H_{nm-1l} - H_{nm+1l} \right. \\
 & + \frac{1}{2} (H_{n-1m-1l} + H_{n+1m-1l} + H_{nm-1l-1} + H_{nm-1l+1} \\
 & \quad - H_{n-1m+1l} - H_{n+1m+1l} - H_{nm+1l-1} - H_{nm+1l+1}) \\
 & + \frac{1}{4} (H_{n-1m-1l-1} + H_{n+1m-1l-1} + H_{n-1m-1l+1} + H_{n+1m-1l+1} \\
 & \quad - H_{n-1m+1l-1} - H_{n+1m+1l-1} - H_{n-1m+1l+1} - H_{n+1m+1l+1}) \left. \right\} \\
 & - ju G_{nml} \quad (10.59)
 \end{aligned}$$

$$\begin{aligned}
 G_{nml}^{[y]} = j \frac{\Delta v}{2} & \left\{ H_{nml-1} - H_{nml+1} \right. \\
 & + \frac{1}{2} (H_{n-1ml-1} + H_{n+1ml-1} + H_{nm-1l-1} + H_{nm+1l-1} \\
 & \quad - H_{n-1ml+1} - H_{n+1ml+1} - H_{nm-1l+1} - H_{nm+1l+1}) \\
 & + \frac{1}{4} (H_{n-1m-1l-1} + H_{n+1m-1l-1} + H_{n-1m+1l-1} + H_{n+1m+1l-1} \\
 & \quad - H_{n-1m-1l+1} - H_{n+1m-1l+1} - H_{n-1m+1l+1} - H_{n+1m+1l+1}) \left. \right\} \\
 & - jv G_{nml} \quad (10.60)
 \end{aligned}$$

Therefore, the weighted integrals with the window functions,  $G_{nml}$ ,  $G_{nml}^{[t]}$ ,  $G_{nml}^{[x]}$ , and  $G_{nml}^{[y]}$ , can be obtained by using the weighted integrals without the window function,  $H_{nml}$ , in a similar manner to the spatial weighted integral case. Adjacent  $3 \times 3 \times 3$  frequencies of  $H_{nml}$  are only required to calculate  $G_{nml}$ ,  $G_{nml}^{[t]}$ ,  $G_{nml}^{[x]}$ , and  $G_{nml}^{[y]}$ . These frequencies should be chosen from around a peak of  $H_{nml}$  because these components include a lot of information.

This solution by using the Hann window is equivalent to simultaneously solve Eq. (10.31) by using adjacent  $3 \times 3 \times 3$  frequencies for eliminating the weighted integral of the boundary values.

## 10.2.2 Algorithm for estimating sound source locations

---

**Algorithm 2** Estimate  $n_x$ ,  $n_y$ , and  $R$  of multiple sound sources

---

$f(x, y, t) \leftarrow$  observation sensor data of time interval  $T$  on a rectangular region

$H_{nml} \leftarrow$  3D DFT of  $f(x, y, t)$

**for all** peak of  $H_{nml}$  **do**

    compute  $G_{nml}$ ,  $G_{nml}^{[t]}$ ,  $G_{nml}^{[x]}$ , and  $G_{nml}^{[y]}$  by using adjacent  $3 \times 3 \times 3$  frequencies of  $H_{nml}$

$n_x$ ,  $n_y$ , and  $R \leftarrow$  solve Eq. (10.45)

**end for**

---

A sound source localization algorithm based on the method presented in Section 10.2.1 is summarized in Algorithm 2. Similar to the spatial weighted integral case, a peak of  $H_{nml}$  by a single sound source can be assumed to be locally-concentrated. In the spatio-temporal weighted integral case, the sparseness of speech signals in the time-frequency domain [110] can also be utilized. Therefore, multiple sound source locations can be estimated by calculating  $\mathbf{n}$  and  $R$  from each peak of  $H_{nml}$ . This algorithm is also computationally efficient because the 3D DFT of the observation sensor data and algebraic operations are only required.

A compensation by half sample phase shift can also be applied to the spatio-temporal weighted finite integral case. When the 3D DFT of  $f(x, y, t)$  is denoted as  $H_{nml}^D$ , the compensation is achieved as:

$$H_{nml} \simeq H_{nml}^D e^{-j(\omega\Delta t + u\Delta x + v\Delta y)} \quad (10.61)$$

Analytical solution of Eq. (10.45) can also be used in practice as:

$$n_x = \frac{\Im\{G_{nml}\}\Re\{G_{nml}^{[x]}\} - \Re\{G_{nml}\}\Im\{G_{nml}^{[x]}\}}{\Re\{G_{nml}\}\Im\{G_{nml}^{[t]}/c\} - \Re\{G_{nml}^{[t]}/c\}\Im\{G_{nml}\}} \quad (10.62)$$

$$n_y = \frac{\Im\{G_{nml}\}\Re\{G_{nml}^{[y]}\} - \Re\{G_{nml}\}\Im\{G_{nml}^{[y]}\}}{\Re\{G_{nml}\}\Im\{G_{nml}^{[t]}/c\} - \Re\{G_{nml}^{[t]}/c\}\Im\{G_{nml}\}} \quad (10.63)$$

The distance  $R$  can be obtained in two ways as:

$$R = \frac{\Im\{G_{nml}\}\Re\{G_{nml}^{[x]}\} - \Re\{G_{nml}\}\Im\{G_{nml}^{[x]}\}}{\Re\{G_{nml}^{[x]}\}\Im\{G_{nml}^{[t]}/c\} - \Re\{G_{nml}^{[t]}/c\}\Im\{G_{nml}^{[x]}\}} \quad (10.64)$$

and

$$R = \frac{\Im\{G_{nml}\}\Re\{G_{nml}^{[y]}\} - \Re\{G_{nml}\}\Im\{G_{nml}^{[y]}\}}{\Re\{G_{nml}^{[y]}\}\Im\{G_{nml}^{[t]}/c\} - \Re\{G_{nml}^{[t]}/c\}\Im\{G_{nml}^{[y]}\}}. \quad (10.65)$$

Similar to the spatial weighted integral case,  $\mathbf{n}$  and  $R$  can be obtained as direct solutions.

## 10.3 Experiments

Numerical simulations of sound source localization were conducted by using Algorithm 2 under the free-field assumption. Omni-directional microphones were aligned on a rectangular observation region. The size of the observation region was  $0.1 \times 0.1$  m and the number of microphones was  $8 \times 8$ . The sampling frequency was 48 kHz. The observed duration was 256 samples, i.e., about 5.3 ms. The sound source was assumed as a single point source and the source signal was a single frequency sinusoidal wave.

Figure 10.2 shows the estimated location of the sound source when the source position was (1.0 m, 45 deg, 80 deg) in the polar coordinates. The frequency of the sound source was changed from 100 Hz to 10 kHz by 100 Hz. Figure 10.2 (a) is the estimated direction and the true direction is indicated by the large circle with the cross mark. The direction of the sound source was accurately estimated at all the frequencies. Figures 10.2 (b) and (c) are the estimated distance by using Eqs. (10.64) and (10.65), respectively. The true distance is indicated by the dashed line. The distance of the sound source was also accurately estimated at all the frequencies.

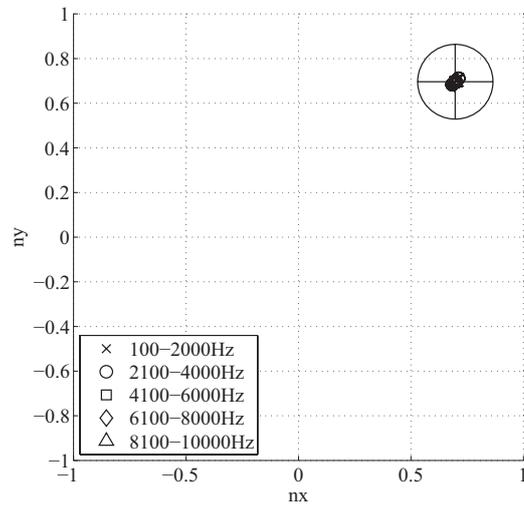
Figure 10.3 shows the estimated location of the sound source when the source position was (1.0 m, 70 deg, 60 deg). Again, the direction of the sound source was accurately estimated at all the frequencies (Fig. 10.3 (a)). The estimated distance have large errors as shown in Figs. 10.3 (b) and (c). This is because the spatial frequency spectrum was concentrated on low frequency components due to the small zenith angle of the sound source position and Eq. (10.64) and (10.65) became ill-conditioned.

Figure 10.4 shows the estimated location of the sound source when the source position was (1.0 m, 70 deg, 60 deg) and white Gaussian noise was added to each microphone so that signal-to-noise ratio (SNR) was 0 dB. Although the estimated direction of the sound source was fluctuated at low frequencies, it was accurately estimated at above 2 kHz as shown in Fig. 10.4 (a). Severe effect of the noise can be seen in the estimated distance (Figs. 10.4 (b) and (c)).

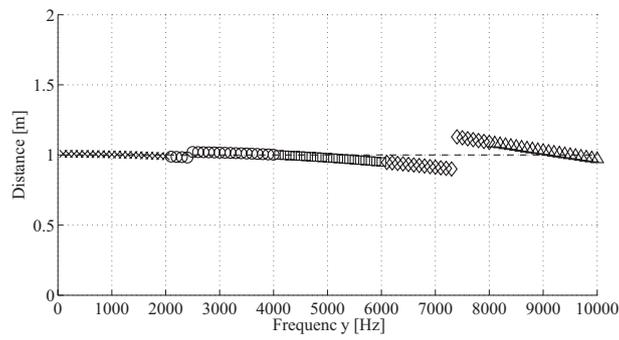
## 10.4 Conclusions

We proposed a sound source localization method by applying the weighted integral method to the LC-PDE in the spatial and spatio-temporal domain. We derived an exact algebraic equation including a sound source location as unknown variables and finite integrals of observation values; therefore, estimate values of sound source locations are algebraically obtained. Additionally, the observation values can be obtained as two or three dimensional DFT of sensor array signals when complex sinusoidal signals are used as the weighting functions. Sound source locations can be estimated at fine temporal resolution by using this method. The experimental results indicated that directions of sound sources were able to be estimated in high accuracy. The estimation accuracy

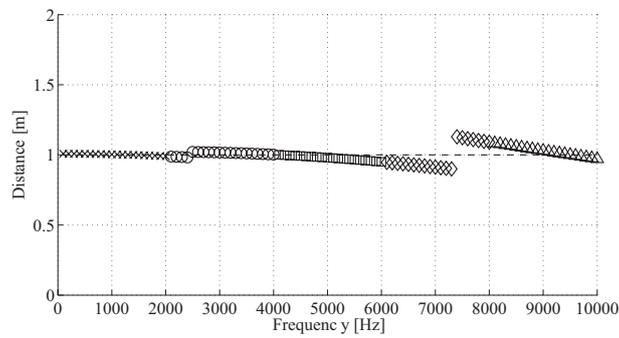
of sound source distances was limited in some conditions. However, when multiple sensor arrays are used, sound source distance may be accurately estimated by using the estimated directions with a triangulation method.



(a) Estimated direction

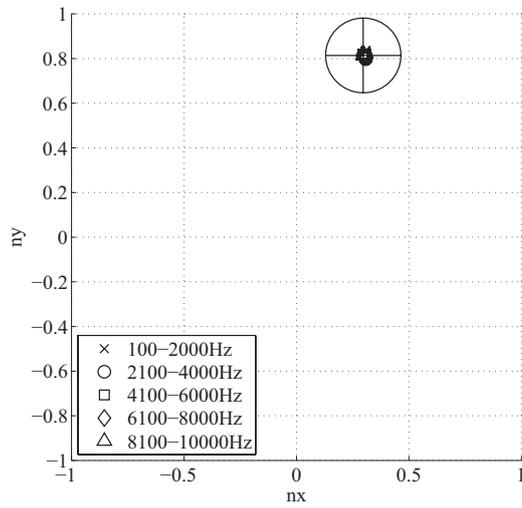


(b) Estimated distance ( $n_x$ )

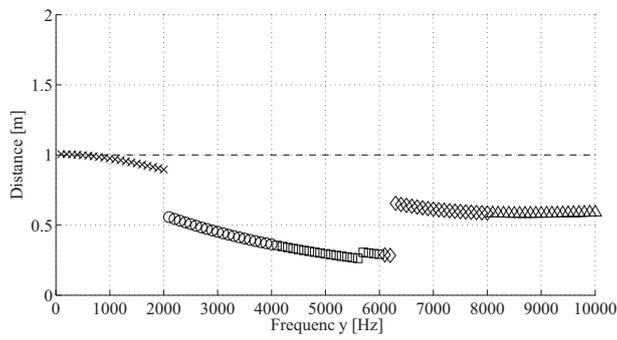


(c) Estimated distance ( $n_y$ )

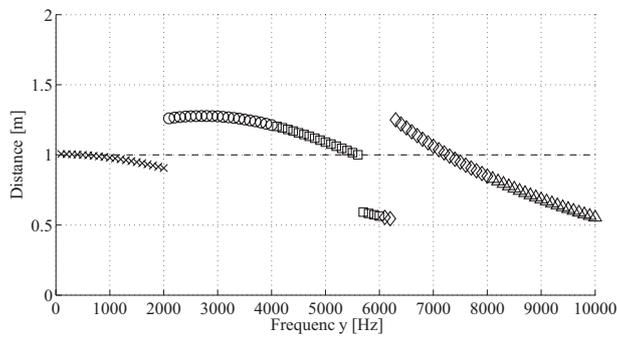
**Fig. 10.2.** Estimated location when source position was (1.0 m, 45 deg, 80 deg).



(a) Estimated direction

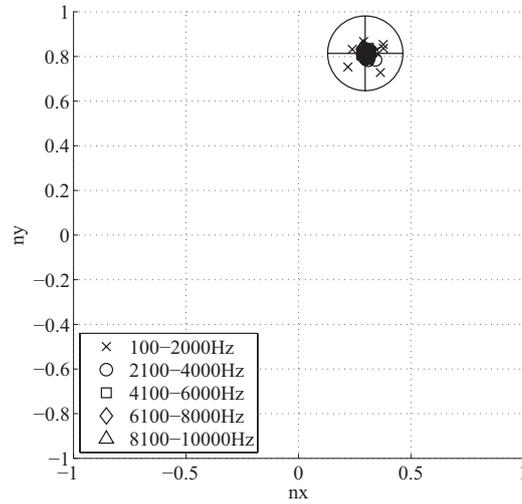


(b) Estimated distance ( $n_x$ )

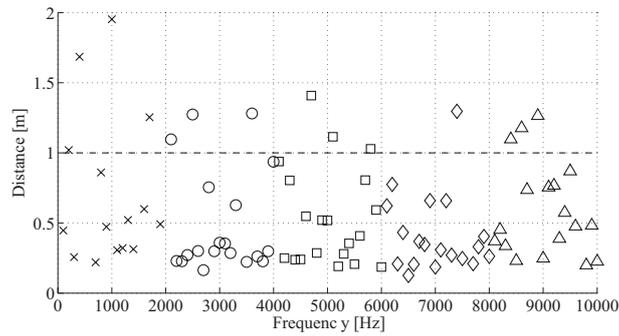
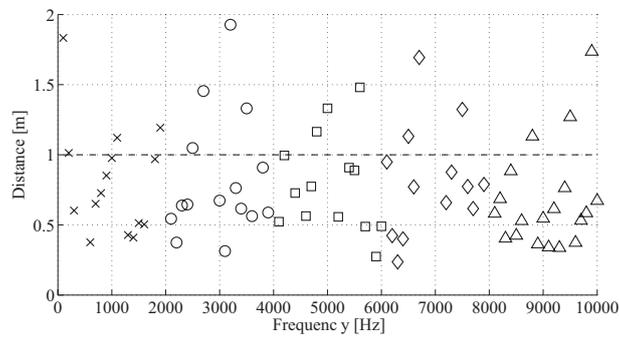


(c) Estimated distance ( $n_y$ )

**Fig. 10.3.** Estimated location when source position was (1.0 m, 70 deg, 60 deg).



(a) Estimated direction

(b) Estimated distance ( $n_x$ )(c) Estimated distance ( $n_y$ )

**Fig. 10.4.** Estimated location when source position was (1.0 m, 70 deg, 60 deg) and white Gaussian noise was added so that SNR was 0 dB.

# Chapter 11

## Conclusions

This thesis addressed two engineering problems: physical sound field reproduction and sound source localization.

From Chapter 3 to 9, we dealt with a sound field reproduction problem. The current sound field reproduction methods were introduced in Chapter 3 and we pointed out that these methods are difficult to apply as an SP-DS conversion. Efficient SP-DS conversion algorithms are highly important for sound field recording and reproducing systems. We proposed several methods for SP-DS conversion from Chapter 4 to 8.

In Chapter 4, we proposed a sound field reproduction method using planar or linear loudspeaker array for reproducing virtual sound sources in front of the loudspeaker array. This method was derived as an extension of WFS by combining with inverse wave propagator and can be applied as SP-DS conversion. The reproduction accuracy of the proposed method was comparable to the conventional focused-source method.

In Chapter 5, 6, and 7, we proposed WFR filtering methods for efficient SP-DS conversion. The WFR filter for planar or linear arrays of microphones and loudspeakers was derived based on the continuous WFR equation in the spatio-temporal frequency domain. The proposed WFR filter has many advantages in filter design, filter size, computational cost, and filter stability compared to the conventional LS-based transform filter. Additionally, the corresponding position of the received and reconstructed sound fields can be controlled by parametric control of the WFR filter.

The WFR filtering method was also derived in the cylindrical coordinates. The proposed WFR filter for cylindrical array configurations is useful when the multiple linear arrays of microphones and loudspeakers are applied for with-height reproduction. The reproduction accuracy of the proposed WFR filter for cylindrical arrays was distinctly high in the neighboring region along the central axis of the cylinder.

Similar concept to the WFR filtering method for planar, linear, or cylindrical arrays was applied to spherical and circular array configurations. Although the resulting WFR filter for spherical arrays was essentially equivalent to the HOA, the driving signals can be obtained in a single convolution of the WFR filter. The WFR filter for circular arrays can be applied as an SP-DS

---

conversion whereas the HOA cannot be directly applied. The WFR filter for circular arrays is derived under the assumption that the captured sound field is invariant with changes in the vertical direction.

In Chapter 8, we proposed an SP-DS conversion method based on the maximum *a posteriori* (MAP) estimation which exploits the prior information of the primary source locations. Spatial basis functions and their coefficients representing the driving signals of the secondary sources are optimised based on the prior information. The reproduction accuracy of the proposed method was better than that of the other methods, especially at frequencies above the spatial Nyquist frequency.

A practical implementation using the WFR filter for linear arrays was presented in Chapter 9. We developed a real-time sound field transmission system over an exclusive IP network. The measurement experiments were conducted and their results indicated that the sound field can be reproduced by using the actual system. The subjective listening test results indicated that this system can provide a sound localization accuracy for virtual sound sources comparable to that for real sound sources in a large listening area. The demonstration system including 3D images was also introduced.

A sound source localization method was proposed in Chapter 10, which was derived by applying the weighted integral method to the LC-PDE in the spatial and spatio-temporal domains. The resulting algorithm is based on the exact algebraic equation including a source location as unknown variables and finite integral of observation values. The observation values can be obtained as two or three dimensional DFT of the sensor array signals. Therefore, sound source locations can be efficiently and accurately estimated at fine temporal resolution. Numerical simulation results indicated that direction of a sound source can be accurately estimated, whereas the estimation accuracy of the sound source distance was limited in some conditions. The proposed method may be useful when sound sources move fast or appropriate initial values for statistical algorithms is required.

# Acknowledgement (in Japanese)

本論文は、筆者が東京大学大学院 情報理工学系研究科 システム情報学専攻 安藤研究室に修士課程として在学した2年間、および日本電信電話株式会社 (NTT) サイバースペース研究所 (現・メディアインテリジェンス研究所) に勤めた4年間の成果をまとめたものです。

私が安藤研究室で研究をはじめたのは、工学部計数工学科の卒業論文のときでした。それから修士課程を修了するまでの約2年半、安藤繁教授には研究に関してあらゆる面でご指導をいただきました。基礎的な理論展開からシミュレーション、実装まで、研究の方法論を幅広くご教授頂いただけでなく、物事の捉え方や、研究者としての姿勢など、安藤先生から受けた影響は多大なものであったと確信しています。本論文の第10章にあたる内容は、私が修士課程で取り組んだテーマであり、安藤先生のご指導があつてのものです。さらに、本論文の学位審査の主査も務めていただき、また、時には安藤研究室のメンバーの皆さんに現在の研究内容を紹介させていただく機会も設けていただき、大変有益なご助言をいただきました。安藤先生には、一研究者としての大きな指針を示していただいたと感じており、深く感謝いたします。

お忙しいところ、予備審査会において有益なご助言やご指導を頂きました。石川正俊教授、苗村健教授、奈良高明准教授、国立情報学研究所の小野順貴准教授に感謝いたします。様々な視点から頂いたご助言は、論文をより充実させるにあたって、大変有益なものでした。奈良先生には、安藤研究室在籍時にも、毎週の検討会等を通して多くのご指導を頂きました。

修士課程在籍時には、安藤研究室の栗原徹助教にも大変お世話になりました。研究に関する議論だけでなく、実験環境の整備などの面でもお力添えをいただき、大変感謝しています。

NTTに入社後は、音声言語メディア処理プロジェクト 音環境情報処理グループにおいて、音響信号処理の研究に従事することとなりました。本論文の第3～9章の内容がその成果にあたります。研究内容について特に多くのご指導をいただいたのが、現・大分大学の古家賢一教授です。古家先生がNTTに在籍されていた際には、本当に長い時間研究の議論に付き合ってくださいました。また、当時のグループリーダーであり、現・電気通信大学の羽田陽一教授にも、研究の議論から、論文の細かい内容に至るまでご指導をいただきました。羽田先生も独自の研究スタイルを確立されており、やはり一つの指針として、大きな影響を受けたと感じています。現・NTT ファシリティーズの日和崎祐介博士には、私が入社して最初の指導者を務めていただきました。特にプレゼンテーションや英語に関する細やかな指導をいただき、企業研究者としてのあり方をご教授いただきました。実システムを用いた実験では、金澤港氏、山西健一氏にお手伝いをいただき、研究を加速化する一助となりました。実験では膨大なチャンネル数のシステムを扱うことが多く、大変な作業をこなしていただき、両氏なくしては実験が進

められなかったと思います。

NTT に海外インターンシップ生として来ていた、ブリティッシュ・コロンビア大学の Timothy Lee 氏には、本論文の第 5 章の一部の実験を行っていただきました。当時はまだ学部 3 年生だったにも関わらず、粘り強く取り組んでいただき、大きな成果を得ることができました。

東北大学の鈴木陽一教授には、共同研究をさせていただいており、本論文の第 6 章にあたる内容の検討をはじめのきっかけを与えていただきました。また、鈴木研究室のトレビーニョ・ホルヘ氏とは研究の議論をさせていただく機会もいただき、大きな刺激を受けたと感じています。

最後になりますが、これまで遠くから支えて頂いた両親に、深く感謝したいと思います。

# Bibliography

- [1] I. E. Sutherland, "The ultimate display," *Proceedings of IFIP*, vol. 65, no. 2, pp. 506–508, 1965.
- [2] D. Zeltzer, "Autonomy, interaction, and presence," *Presence: Teleoperators and Virtual Environment*, vol. 1, no. 1, pp. 127–132, 1992.
- [3] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1996.
- [4] A. S. Bregman, *Auditory Scene Analysis: Perceptual Organization of Sound*. Bradford Books, 1990.
- [5] ITU-R, "Multichannel stereophonic sound system with and without accompanying picture," *ITU-R Recommend. BS-775-1*, 1994.
- [6] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando, "A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)," *SMPTE Motion Imaging Journal*, vol. 117, no. 3, pp. 40–49, 2008.
- [7] F. Rumsey, *Spatial Audio*. Oxford: Focal Press, 2001.
- [8] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of Audio Engineering Society*, vol. 45, pp. 456–466, 1997.
- [9] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *Proceedings of the IEEE*, vol. 51, no. 3, pp. 536–537, 1963.
- [10] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. P T R Prentice-Hall, 1993.
- [11] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.
- [12] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, 1986.
- [13] A. Paulraj and R. R. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [14] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4,

- pp. 1560–1577, 1988.
- [15] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood, and Cramer-Rao Bound,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–714, 1989.
- [16] P. Stoica and K. Sharman, “Maximum likelihood methods for direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1132–1143, 1990.
- [17] M. I. Miller and D. R. Fhurmman, “Maximum-Likelihood Narrow-Band Direction Finding and the EM Algorithm,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 9, pp. 1560–1577, 1990.
- [18] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, Inc, 2001.
- [19] F. Asano, H. Asoh, and T. Matsui, “Sound source localization and separation in near field,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, no. 11, pp. 2286–2294, 2000.
- [20] S. Ando and T. Nara, “An exact direct method of sinusoidal parameter estimation derived from finite fourier intergral of differential equation,” *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3317–3329, 2009.
- [21] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London: Academic Press, 1999.
- [22] A. Sommerfeld, *Lectures on theoretical Physics: Optics*. Academic Press, 1954.
- [23] M. Born and E. Wolf, *Principles of Optics*. Pergamon Press, 1975.
- [24] O. K. Ersoy, *Diffraction, Fourier Optics and Imaging*. John Wiley & Sons, 2006.
- [25] S. Ando, H. Shinoda, K. Ogawa, and S. Mitsuyama, “A three-dimensional sound localization sensor system based on the spatio-temporal gradient method,” *Journal of Society of Instrument and Control Engineers*, vol. 29, no. 5, pp. 520–528, 1993 (in Japanese).
- [26] F. Fazi, P. Nelson, and R. Potthast, “Analogies and differences between three methods for sound field reproduction,” in *Proceedings of Ambisonics Symposium*, (Graz), June 25-27 2009.
- [27] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, 1998.
- [28] D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory*. John Wiley & Sons, Inc., 1983.
- [29] A. J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” *Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.
- [30] E. Verheijen, *Sound Field Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997.
- [31] D. de Vries, *Wave Field Synthesis*. AES Monograph, Audio Engineering Society, 2009.

- [32] S. Spors, H. Teutsch, A. Kuntz, and R. Rabenstein, "Sound field synthesis," in *Audio Signal Processing For Next-Generation Multimedia Communication Systems* (Y. Huang and J. Benesty, eds.), ch. 12, Kluwer Academic Publishers, 2004.
- [33] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *Proceedings of 124th AES Convention*, (Amsterdam, The Netherlands), Oct. 2008.
- [34] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2038–2050, 2010.
- [35] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Springer-Verlag, 2012.
- [36] M. A. Gerzon, "Periphony: With-height sound field reproduction," *Journal of the Audio Engineering Society*, vol. 21, pp. 2–10, Jan. 1973.
- [37] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format," in *Proceedings of 23rd AES Conference*, (Copenhagen, Denmark), May 2003.
- [38] M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 56, no. 11, pp. 1004–1025, 2005.
- [39] J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 988–999, 2008.
- [40] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, 2009.
- [41] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the sound field," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1781–1784, 2002.
- [42] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.
- [43] C. Travis, "A new mixed-order scheme for ambisonic signals," in *Proceedings of Ambisonics Symposium*, (Graz), 2009.
- [44] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2992–3000, 1993.
- [45] P. A. Nelson, "Active control of acoustic fields and the reproduction of sound," *Journal of Sound and Vibration*, vol. 177, no. 4, pp. 447–477, 1993.
- [46] O. Kirkeby, P. A. Nelson, F. O. Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *Journal of the Acoustical Society of America*, vol. 100, no. 3, 1996.
- [47] P. A. Gauthier and A. Berry, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *Journal of the Acoustical Society of Amer-*

- ica, vol. 117, no. 2, pp. 662–678, 2005.
- [48] N. Kamado, H. Hokari, S. Shimada, H. Saruwatari, and K. Shikano, “Sound field reproduction by wavefront synthesis using directly aligned multi point control,” in *Proceedings of 40th AES Conference*, (Tokyo), Oct. 2010.
- [49] M. Kolundžija, C. Faller, and M. Vitterli, “Reproducing sound fields using MIMO acoustic channel inversion,” *Journal of the Audio Engineering Society*, vol. 59, no. 10, pp. 721–734, 2011.
- [50] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.
- [51] J. J. López, A. González, and L. Fuster, “Room compensation in wave field synthesis by means of multichannel inversion,” in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz), Oct. 2005.
- [52] E. Corteel, “Equalization in an extended area using multichannel inversion and wave field synthesis,” *Journal of the Audio Engineering Society*, vol. 54, pp. 1140–1161, 2006.
- [53] S. Spors, H. Buchner, and R. Rabenstein, “Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering,” *Journal of the Acoustical Society of America*, vol. 54, no. 12, pp. 354–369, 2007.
- [54] F. Asano, Y. Suzuki, and T. Sone, “Sound equalization using derivative constraints,” *Acta Acustica united with Acustica*, vol. 82, no. 2, pp. 311–320, 1996.
- [55] F. Fazi and P. A. Nelson, “Nonuniqueness of the solution of the sound field reproduction problem with boundary pressure control,” *Acta Acustica united with Acustica*, vol. 98, pp. 1–14, 2012.
- [56] S. Ise, “A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems,” *Acta Acustica united with Acustica*, vol. 85, no. 1, pp. 78–87, 1999.
- [57] G. N. Lilis, D. Angelosante, and G. B. Giannakis, “Sound field reproduction using the Lasso,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, 2010.
- [58] S. Yon *et al.*, “Sound focusing in rooms: The time-reversal approach,” *Journal of the Acoustical Society of America*, vol. 113, no. 3, pp. 1533–1543, 2003.
- [59] M. Boone, E. Verheijen, and G. Jansen, “Virtual reality by sound reproduction based on wave field synthesis,” in *Proceedings of 100th AES Convention*, 1996.
- [60] T. Caulkins, E. Corteel, and O. Warusfel, “Wave field synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room,” in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, 2003.
- [61] S. Komiyama, Y. Nakayama, K. Ono, and S. Koizumi, “A loudspeaker array to control sound image distance,” *Acoustical Science and Technology*, vol. 24, no. 5, pp. 242–249,

- 2003.
- [62] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and perceptual properties of focused sources in wave field synthesis," in *Proceedings of 127th AES Convention*, 2009.
- [63] S. Spors and J. Ahrens, "Reproduction of focused sources by the spectral division method," in *Proceedings of the 4th IEEE-EURASIP International Symposium on Communications, Control and Signals Processing*, (Limassol), 2010.
- [64] E. Lalor, "Inverse wave propagator," *Journal of Mathematical Physics*, vol. 9, no. 12, pp. 2001–2006, 1968.
- [65] C. P. A. Wapenaar, G. L. Peeis, V. Budejicky, and A. J. Berkhout, "Inverse extrapolation of primary seismic waves," *GEOPHYSICS*, vol. 54, no. 7, pp. 853–863, 1989.
- [66] E. Hulusebos, D. de Vries, and E. Bourdillat, "Improved microphone array configuration for auralization of sound fields by wave field synthesis," *Journal of the Audio Engineering Society*, vol. 50, no. 10, pp. 779–790, 2002.
- [67] M. Nieto-Vesperinas, *Scattering and Diffraction in Physical Optics*. John Wiley & Sons, Inc, 1991.
- [68] J. R. Shewell and E. Wolf, "Inverse diffraction and a new reciprocity theorem," *Journal of Optical Society of America*, vol. 58, no. 12, 1968.
- [69] S. Spors and R. Rabenstein, "Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis," in *Proceedings of 120th AES Convention*, (Paris, France), May 2006.
- [70] D. E. Amos, "A portable package for Bessel functions of a complex argument and non-negative order," *ACM Transactions on Mathematical Software*, vol. 12, no. 3, pp. 265–273, 1986.
- [71] P. C. Hansen and D. P. O’Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993.
- [72] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2 ed., 2009.
- [73] P. A. Nelson and S. J. Elliott, *Active Control of Sound*. Academic Press, 1992.
- [74] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum*. World Scientific, 1988.
- [75] Y. Haneda, K. Furuya, and H. Itou, "Design of multipole loudspeaker array based on spherical harmonic expansion," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, (Prague), 2011.
- [76] J. Ahrens and S. Spors, "An analytical approach to 2.5D sound field reproduction employing linear distributions of non-omni-directional loudspeakers," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 105–108, 2010.

- 
- [77] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica United with Acustica*, vol. 91, pp. 409–420, 2005.
- [78] E. Tiana-Roig, F. Jacobsen, and E. F. Grande, "Beamforming with a circular microphone array for localization of environmental noise sources," *Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3535–3542, 2010.
- [79] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wave-field Decomposition*. Springer, 2007.
- [80] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [81] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*. Springer, 2010.
- [82] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, 2008.
- [83] H. Kuttruff, *Room Acoustics*. Elsevier Applied Science, 1991.
- [84] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, 2011.
- [85] S. Spors and J. Ahrens, "A comparison of wave field synthesis and higher order ambisonics with respect to physical properties and spatial sampling," in *Proceedings of 125th AES Convention*, (San Francisco, USA), Oct. 2008.
- [86] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *Journal of the Acoustical Society of America*, vol. 116, pp. 2149–2157, Oct. 2004.
- [87] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 2–16, 2010.
- [88] J. Ahonen, G. D. Galdo, F. Kuech, and V. Pulkki, "Directional analysis with microphone array mounted on rigid cylinder for directional audio coding," *J. Audio Eng. Soc.*, vol. 60, no. 5, pp. 311–324, 2012.
- [89] J. R. Driscoll and D. M. Healy Jr., "Computing fourier transforms and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, pp. 202–250, 1994.
- [90] M. J. Mohlemkamp, "A fast transform for spherical harmonics," *Journal of Fourier Analysis and Applications*, vol. 5, no. 2-3, pp. 159–184, 1999.
- [91] M. A. Poletti, T. Betlehem, and T. D. Abhayapala, "Analysis of 2d sound reproduction with fixed-directivity loudspeakers," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 377–380, 2012.
- [92] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*. New York: Springer-Verlag, 2010.

- [93] J. Antoni, "A bayesian approach to sound source reconstruction: Optimal basis, regularization, and focusing," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 2873–2890, 2012.
- [94] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [95] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–120, 2007.
- [96] E. J. Candès, "Compressive sampling," in *International Congress of Mathematicians*, vol. 3, pp. 1433–1452, 2006.
- [97] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.
- [98] K. Engan, S. O. Aase, and J. H. Hakon-Husoy, "Method of optimal directions for frame design," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 2443–2446, 1999.
- [99] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [100] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.
- [101] U. Horbach and A. Karamustafaoglu, "Practical implementation of a data-based wave field reproduction system," in *Proceedings of 108th AES Convention*, (Paris), Feb. 2000.
- [102] S. Brix, T. Sporer, and J. Plogsties, "Carrouso - an European approach to 3D-audio," in *Proceedings of 110th AES Convention*, (Amsterdam), May 2001.
- [103] Y. Suzuki, F. Asano, H. Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1119–1123, 1995.
- [104] M. Shimomura, A. Masuda, H. Uose, N. Inoue, and N. Nakayama, "GEMnet2 R&D testbed network," *NTT Technical Review*, vol. 11, no. 1, 2013.
- [105] H. Takada, M. Date, S. Koyama, S. Ozawa, S. Mieda, and A. Kojima, "High-reality space composition using stably-positioned imaging and acoustic wave field synthesis," in *Proceedings of European Conference on Visual Perception (ECVP 2013)*, Aug. 2013.
- [106] S. Ando and N. Ono, "Partial-differential equation (PDE)-based theory of sound source localization," in *Proceedings of 4th Joint Meeting ASA and ASJ*, (Honolulu), 2006.
- [107] S. Ando, N. Ono, and Y. Fujita, "Parital differential equation-based algorithm of sound source localization with finest granularity in both time and frequency," in *Proceedings of International Conference on Networked Seinsing Systems (INSS)*, pp. 229–234, 2007.
- [108] S. Ando, T. Nara, and T. Levi, "Partial differential equation-based localization of a monopole source from a circular array," *Journal of Acoustical Society of America*, vol. 134, no. 4, pp. 2799–2813, 2013.
- [109] S. Ando, "Consistent gradient operators," *IEEE Transactions on Pattern Analysis and Ma-*

- chine Intelligence*, vol. 22, no. 3, pp. 252–265, 2000.
- [110] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, 2004.
- [111] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. Academic Press, 2007.

# List of Publications

## Journal papers and articles

- [J1] Shoichi Koyama, Toru Kurihara, and Shigeru Ando, “A Theory and Experiment of Instantaneous Wave Source Localization from a Wave Distribution on a Small Region,” *IEEE Transactions on Sensors and Micromachines*, vol. 129, no. 10, pp. 350-356, 2009 (in Japanese).
- [J2] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Reproducing virtual sound sources in front of a loudspeaker array using inverse wave propagator,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1746-1758, 2012.
- [J3] Shoichi Koyama, “Mathematical formulation of sound field reproduction. –Mathematics in Wave Field Synthesis and Higher Order Ambisonics–,” *Journal of the Acoustical Society of Japan*, vol. 68, no. 11, 2012 (in Japanese).
- [J4] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Analytical Approach to Wave Field Reconstruction Filtering in Spatio-Temporal Frequency Domain,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 4, pp. 685-696, 2013.
- [J5] Yoichi Haneda, Ken’ichi Furuya, Shoichi Koyama, Kenta Niwa, Kazunori Kobayashi, “Sound field simulation for circular array based on spatial circular convolution,” *Acoustical Science and Technology*, 2013 (in press).
- [J6] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, Yoichi Haneda, and Yôiti Suzuki, “Wave Field Reconstruction Filtering in Cylindrical Harmonic Domain for With-Height Reproduction,” *IEEE Transactions on Audio, Speech, and Language Processing* (submitted).

## International conferences

- [C1] Shigeru Ando, Shoichi Koyama, Nobutaka Ono, Naoki Ikeuchi. “Silicon Cochlear Microphone “Fishbone” with Complex Filtering and Hilbert Transform Capability for Instantaneous Signal Detection and Analysis,” in *Proceedings of Transducers '07*, pp. 1295-1298, 2007.
- [C2] Shoichi Koyama, Shigeru Ando, Nobutaka Ono, and Naoki Ikeuchi, “On-Sensor Abnormal Signal Detection and Wireless Transfer Using “Fishbone” Acoustic Sensor,” in *Proceedings of the 24th Sensor Symposium*, pp. 51-54, 2007.
- [C3] Shoichi Koyama, Toru Kurihara, and Shigeru Ando, “Instantaneous Wave Source Localization Based on Weighted Integral of Partial Differential Equation,” in *Proceedings of the 25th Sensor Symposium*, pp. 667-670, 2008.
- [C4] Shoichi Koyama, Toru Kurihara, and Shigeru Ando, “A Direct Algebraic Method of Instantaneous Wave Source Localization,” in *Proceedings of SICE Annual Conference 2008*, pp. 532-537, 2008.
- [C5] Shoichi Koyama and Risa Suzuki, “Scape Synthesis: Generation of a Novel Acoustical Space Based on Auditory Perceptual Features,” in *Proceedings of the 2009 International Computer Music Conference (ICMC 2009)*, pp. 533-536, 2009.
- [C6] Shoichi Koyama, Yusuke Hiwasaki, Ken'ichi Furuya, and Yoichi Haneda, “Inverse Wave Propagation in Wave Field Synthesis,” in *Proceedings of 40th AES International Conference*, 2010.
- [C7] Shoichi Koyama, Yusuke Hiwasaki, Ken'ichi Furuya, and Yoichi Haneda, “Inverse Wave Propagation for Reproducing Virtual Sources in Front of Loudspeaker Array,” in *Proceedings of European Signal Processing Conference (EUSIPCO)*, pp. 1322-1326, 2011.
- [C8] Shoichi Koyama, Ken'ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Design of Transform Filter for Sound Field Reproduction Using Microphone Array and Loudspeaker Array,” in *Proceedings of Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, pp. 5-8, Oct. 2011.
- [C9] Shoichi Koyama, Ken'ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Sound Field

- Recording and Reproduction Using Transform Filter Designed in Spatio-Temporal Frequency Domain,” in *Proceedings of 131st AES Convention*, 2011.
- [C10] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Design of Transform Filter For Reproducing Arbitrary Shifted Sound Field Using Phase-Shift of Spatio-Temporal Frequency,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 381-384, 2012.
- [C11] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Sound Field Reproduction Method in Spatio-Temporal Frequency Domain Considering Directivity of Loudspeakers,” in *Proceedings of 132nd AES Convention*, 2012.
- [C12] Satoru Emura, Shoichi Koyama, Ken’ichi Furuya, and Yoichi Haneda, “Posterior Residual Echo Canceling and its Complexity Reduction in the Wave Domain,” in *Proceedings of International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.
- [C13] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, Yoichi Haneda, and Yôiti Suzuki, “Sound field reproduction using multiple linear arrays based on wave field reconstruction filtering in helical wave spectrum domain,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 271-275, Mar., 2013.
- [C14] Shoichi Koyama, Timothy Lee, Ken’ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “Improvement using circular harmonics beamforming on reverberation problem of wave field reconstruction filtering,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 276-280, Mar., 2013.
- [C15] Hideaki Takada, Munekazu Date, Shoichi Koyama, Shiro Ozawa, Satoshi Mieda, Akira Kojima, “High-reality space composition using stably-positioned imaging and acoustic wave field synthesis,” in *Proceedings of European Conference on Visual Perception (ECVP 2013)*, Aug., 2013.
- [C16] Shoichi Koyama, Ken’ichi Furuya, Hisashi Uematsu, Yusuke Hiwasaki, and Yoichi Haneda, “Real-time sound field transmission system by using wave field reconstruction filter and its subjective listening test,” in *Proceedings of 52nd AES International Conference*, 2013.
- [C17] Shoichi Koyama, Ken’ichi Furuya, Yusuke Hiwasaki, and Yoichi Haneda, “MAP Estimation of Driving Signals of Loudspeakers for Sound Field Reproduction From Pressure Mea-

surements,” in *Proceedings of Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013.

## Domestic conferences (in Japanese)

- [D1] 小山翔一, 安藤 繁, 小野 順貴, 池内 直樹, “Fishbone 音響センサを用いるオンセンサ異常音抽出と無線伝送,” 計測自動制御学会 第 24 回センシングフォーラム資料, pp. 377-382, 2007.
- [D2] 小山翔一, 栗原 徹, 安藤 繁, “空間積分型音源恒等式による瞬時音源定位,” 日本音響学会 春季研究発表会講演論文集, 3-6-3, pp. 711-714, 2008.
- [D3] 小山翔一, 栗原 徹, 安藤 繁, “偏微分方程式の空間荷重積分による瞬時音源定位,” 日本音響学会 秋季研究発表会講演論文集, 2-8-20, pp. 679-682, 2008.
- [D4] 小山翔一, 栗原 徹, 安藤 繁. “偏微分方程式の空間荷重積分による波源定位の直接解法: 理論とシミュレーション,” 応用音響/信号処理研究会, 信学技報, vol. 108, no. 68, EA2008-12, pp. 37-42, 2008.
- [D5] 小山翔一, 栗原 徹, 安藤 繁, “偏微分方程式の空間荷重積分による波源定位の直接解法,” 計測自動制御学会 第 25 回センシングフォーラム資料, pp. 25-29, 2009.
- [D6] 小山翔一, 栗原 徹, 安藤 繁, “荷重積分法による瞬時波源定位と残響の時空間分解シミュレーション,” 電気学会全国大会講演論文集 [3], pp. 206, 2009.
- [D7] 阪内 澄宇, 古家 賢一, 小山翔一, 羽田 陽一, “高臨場遠隔通信のためのサウンドウォールシステム,” 音響学会 春季研究発表会講演論文集, 3-P-30, pp. 935-936, 2010.
- [D8] 小山翔一, 日和崎 祐介, 古家 賢一, 羽田 陽一, “角度スペクトル微分による音圧勾配取得に基づく波面合成法,” 音響学会 秋季研究発表会講演論文集, 1-10-9, pp. 555-556, 2010.
- [D9] 小山翔一, 日和崎 祐介, 古家 賢一, 羽田 陽一, “逆伝播演算による波面合成位置の操作,” 音響学会 秋季研究発表会講演論文集, 3-Q-6, pp. 751-752, 2010.
- [D10] 小山翔一, 日和崎 祐介, 古家 賢一, 羽田 陽一, “角度スペクトル微分を用いた波面合成法の実音場評価,” 音響学会 春季研究発表会講演論文集, 3-9-6, pp. 741-742, 2011.

- [D11] 小山翔一, 古家 賢一, 日和崎 祐介, 羽田 陽一, “音場収音・再現のための時空間周波数領域信号変換法,” 音響学会秋季研究発表会講演論文集, 2-4-1, pp. 635-636, 2011.
- [D12] 小山翔一, 古家 賢一, 日和崎 祐介, 羽田 陽一, “時空間スペクトルの位相シフトによる音場再現位置の制御,” 音響学会秋季研究発表会講演論文集, 2-4-2, pp. 637-638, 2011.
- [D13] 小山翔一, 古家 賢一, 日和崎 祐介, 羽田 陽一, “スピーカ指向性を考慮した時空間周波数領域での音場再現フィルタ設計手法,” 音響学会春季研究発表会講演論文集, 2-Q-6, pp. 913-914, 2012.
- [D14] 小山翔一, 古家 賢一, 日和崎 祐介, 羽田 陽一, 鈴木 陽一, “円筒状マイクロホン・スピーカアレイのための波面再構成フィルタ,” 音響学会秋季研究発表会講演論文集, 1-9-13, pp. 605-608, 2012.
- [D15] 江村 暁, 小山翔一, 古家 賢一, 羽田 陽一, “波数領域における残留エコー消去法,” 音響学会秋季研究発表会講演論文集, 3-4-4, pp. 667-668, 2012.
- [D16] 小山翔一, 古家 賢一, 日和崎 祐介, 羽田 陽一, “球・円状マイクロホン・スピーカアレイのための波面再構成フィルタ,” 音響学会春季研究発表会講演論文集, 1-10-5, pp. 679-682, 2013.
- [D17] 小山翔一, Timothy Lee, 古家 賢一, 日和崎 祐介, 羽田 陽一, “円調和ビームフォーミングによる波面再構成フィルタの残響環境下性能改善,” 音響学会春季研究発表会講演論文集, 1-P-19, pp. 861-862, 2013.
- [D18] トレビーニョ ホルヘ, 小山翔一, 坂本 修一, 鈴木 陽一, “Scalable encoding of sound field recordings made with cylindrical microphone arrays,” 音響学会春季研究発表会講演論文集, 2-Q-36, pp. 957-960, 2013.
- [D19] 高田 英明, 伊達 宗和, 小山翔一, 小澤 史朗, 三枝 知史, 小島 明, “3D 映像音響システムによる遠隔地間の空間共有,” 3次元画像コンファレンス, 2013.
- [D20] 羽田 陽一, 古家 賢一, 小山翔一, 丹羽 健太, “球面調和関数展開による音圧内挿を用いた超接話球面マイクロホンアレイ,” 音響学会秋季研究発表会講演論文集, 1-1-15, pp. 611-612, 2013.
- [D21] 小山翔一, 古家 賢一, 日和崎 祐介, 羽田 陽一, “音源位置事前情報を用いた音圧-駆動信号変換に基づく音場再現,” 音響学会秋季研究発表会講演論文集, 3-1-2, pp. 679-682,

2013.

- [D22] トレビーニョ ホルへ, 小山翔一, 坂本 修一, 鈴木 陽一, “3D cylindrical Ambisonics: Encoding sound field information using the cylindrical harmonic functions,” 音響学会秋季研究発表会講演論文集, 1-Q-16, pp. 723-726, 2013.

# List of Awards

1. Young Researcher Paper Award on Measurement Division of the Society of Instrument and Control Engineers (SICE), 2009.  
計測自動制御学会, 計測部門 研究技術奨励賞, 2009.
2. Best Young Researcher Paper Award on Sensors and Micromachines Society of the Institute of Electrical Engineers of Japan (IEEJ), 2010.  
電気学会, センサ・マイクロマシン部門 奨励論文賞, 2010.
3. Awaya Prize Young Researcher Award from Acoustical Society of Japan (ASJ), 2011.  
日本音響学会, 栗屋 潔学術奨励賞, 2011.

## Appendix A

# Definitions of Fourier transforms

Several definitions employed in this thesis are described in the following.

### A.1 Fourier transform

The Fourier transform,  $\tilde{F}(k_x)$ , of a function  $F(x)$  with respect to  $x$  is defined as

$$\begin{aligned}\tilde{F}(k_x) &= \mathcal{F}_x [F(x)] \\ &= \int_{-\infty}^{\infty} F(x) e^{-jk_x x} dx,\end{aligned}\tag{A.1}$$

where  $\mathcal{F}_x$  represent the Fourier transform operator. The inverse transform corresponding to Eq. (A.1),  $\mathcal{F}_x^{-1}$ , is

$$\begin{aligned}F(x) &= \mathcal{F}_x^{-1} [\tilde{F}(k_x)] \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{F}(k_x) e^{jk_x x} dk_x.\end{aligned}\tag{A.2}$$

When the spatial Fourier transform pair is defined by Eqs. (A.1) and (A.2), the temporal Fourier transform pair is described as:

$$\begin{aligned}F(\omega) &= \mathcal{F}_t [f(t)] \\ &= \int_{-\infty}^{\infty} f(t) e^{j\omega t} dt\end{aligned}\tag{A.3}$$

and

$$\begin{aligned}f(t) &= \mathcal{F}_t^{-1} [F(\omega)] \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{-j\omega t} d\omega.\end{aligned}\tag{A.4}$$

Note that the sign of the exponential term is reversed. This is for retaining the meaning of a plane wave given by  $\exp(j(k_x x + k_y y + k_z z - \omega t))$  [21]. Thus the Fourier transform pair with respect

to  $x$  and  $t$  is described as:

$$\begin{aligned}\tilde{F}(k_x, \omega) &= \mathcal{F}_x \mathcal{F}_t [f(x, t)] \\ &= \iint_{-\infty}^{\infty} f(x, t) e^{-jk_x x} e^{j\omega t} dx dt\end{aligned}\quad (\text{A.5})$$

and

$$\begin{aligned}f(x, t) &= \mathcal{F}_x^{-1} \mathcal{F}_t^{-1} [\tilde{F}(k_x, \omega)] \\ &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{F}(k_x, \omega) e^{jk_x x} e^{-j\omega t} dk_x d\omega.\end{aligned}\quad (\text{A.6})$$

One of the important theorems regarding the Fourier transform is the convolution theorem as:

$$\begin{aligned}\mathcal{F}_x [F(x) * G(x)] &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} F(x - x') G(x') dx' \right] e^{-jk_x x} dx \\ &= \tilde{F}(k_x) \tilde{G}(k_x),\end{aligned}\quad (\text{A.7})$$

where “\*” denotes convolution:

$$F(x) * G(x) = \int_{-\infty}^{\infty} F(x - x') G(x') dx'. \quad (\text{A.8})$$

The two-dimensional Fourier transform pair of a two-dimensional function  $F(x, z)$  is defined as:

$$\begin{aligned}\tilde{F}(k_x, k_z) &= \mathcal{F}_x \mathcal{F}_z [f(x, z)] \\ &= \iint_{-\infty}^{\infty} F(x, z) e^{-j(k_x x + k_z z)} dx dz\end{aligned}\quad (\text{A.9})$$

and

$$\begin{aligned}F(x, z) &= \mathcal{F}_x^{-1} \mathcal{F}_z^{-1} [\tilde{F}(k_x, k_z)] \\ &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{F}(k_x, k_z) e^{j(k_x x + k_z z)} dk_x dk_z.\end{aligned}\quad (\text{A.10})$$

The two-dimensional convolution theorem is

$$\begin{aligned}\mathcal{F}_x \mathcal{F}_z [F(x, z) ** G(x, z)] \\ &= \iint_{-\infty}^{\infty} \left[ \iint_{-\infty}^{\infty} F(x - x', z - z') G(x', z') dx' dz' \right] e^{-j(k_x x + k_z z)} dx dz \\ &= \tilde{F}(k_x, k_z) \tilde{G}(k_x, k_z),\end{aligned}\quad (\text{A.11})$$

where “\*\*” denotes two-dimensional convolution:

$$F(x, z) ** G(x, z) = \iint_{-\infty}^{\infty} F(x - x', z - z') G(x', z') dx' dz'. \quad (\text{A.12})$$

## A.2 Fourier series

When circular (polar) coordinates are given by  $r$  and  $\phi$ , a function  $F(r, \phi)$  can be represented in a Fourier series in the  $\phi$  coordinate as:

$$\begin{aligned} F(r, \phi) &= \mathcal{F}_\phi^{-1} [\tilde{F}_n(r)] \\ &= \sum_{n=-\infty}^{\infty} \tilde{F}_n(r) e^{jn\phi}, \end{aligned} \quad (\text{A.13})$$

where the coefficient functions,  $\tilde{F}_n(r)$ , are given by

$$\begin{aligned} \tilde{F}_n(r) &= \mathcal{F}_\phi [F(r, \phi)] \\ &= \frac{1}{2\pi} \int_0^{2\pi} F(r, \phi) e^{-jn\phi} d\phi. \end{aligned} \quad (\text{A.14})$$

These coefficients are referred to as circular harmonic coefficients.

The convolution relationship for Fourier series is derived as

$$\begin{aligned} \mathcal{F}_\phi [F(\phi) * G(\phi)] &= \frac{1}{2\pi} \int_0^{2\pi} \left[ \int_0^{2\pi} F(\phi - \phi') G(\phi') d\phi' \right] e^{-jn\phi} d\phi \\ &= 2\pi \tilde{F}_n \tilde{G}_n, \end{aligned} \quad (\text{A.15})$$

where “\*” denotes circular convolution as:

$$F(\phi) * G(\phi) = \int_0^{2\pi} F(\phi - \phi') G(\phi') d\phi'. \quad (\text{A.16})$$

## A.3 Spherical harmonic transform

When spherical (polar) coordinates are given by  $r$ ,  $\theta$ , and  $\phi$ , the spherical harmonic function  $Y_n^m(\theta, \phi)$  is defined as [21]:

$$Y_n^m(\theta, \phi) \equiv \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \theta) e^{jm\phi}, \quad (\text{A.17})$$

where  $P_n^m(\cdot)$  is the associated Legendre functions. The spherical harmonic function  $Y_n^m(\theta, \phi)$  is orthonormal.

It is possible that any arbitrary function on a sphere,  $F(r, \theta, \phi)$ , can be expanded in terms of the spherical harmonic functions as:

$$\begin{aligned} F(r, \theta, \phi) &= \mathcal{S}^{-1} [\tilde{F}_n^m(r)] \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \tilde{F}_n^m(r) Y_n^m(\theta, \phi), \end{aligned} \quad (\text{A.18})$$

where the coefficient functions,  $\tilde{F}_n^m(r)$ , are given by

$$\begin{aligned}\tilde{F}_n^m(r) &= \mathcal{S}[F(r, \theta, \phi)] \\ &= \int_0^{2\pi} d\phi \int_0^\pi F(r, \theta, \phi) Y_n^m(\theta, \phi)^* \sin \theta d\theta.\end{aligned}\quad (\text{A.19})$$

These coefficients are referred to as spherical harmonic coefficients.

Spherical convolution “ $*_{\text{sph}}$ ” of two functions on a sphere,  $F(\gamma)$  and  $G(\gamma)$ , is defined as [89]:

$$F(\gamma) *_{\text{sph}} G(\gamma) = \int_{u \in SO(3)} F(u^{-1}\gamma) G(u\eta) r^2 du, \quad (\text{A.20})$$

where  $u \in SO(3)$  is a rotational operation,  $\eta = (0, 0, r)$  denotes the north pole, and  $\gamma = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)$  in the Cartesian coordinates. The convolution relationship for spherical harmonic coefficients is described as:

$$\mathcal{S}[F(\gamma) *_{\text{sph}} G(\gamma)] = 2\pi r^2 \sqrt{\frac{4\pi}{2n+1}} \tilde{F}_n^m \tilde{G}_n^0. \quad (\text{A.21})$$

This equation assumes that  $G(\gamma)$  is invariant with respect to rotation around the coordinate origin.

## A.4 Fourier transforms of Green functions

The three- and two-dimensional Green functions in the spectral domain are analytically derived in the Cartesian, cylindrical, and spherical coordinates. Because these equations are equivalent to monopole and line source characteristics, respectively, they are used to derive the WFR equations in each coordinate.

### A.4.1 Derivation in Cartesian coordinates

The Fourier transform of the three-dimensional Green function, Eq. (3.2), with respect to  $x$  is derived as [111] (3.876-1), (3.876-2):

$$\begin{aligned}\tilde{G}_{3D}(k_x, y, z, \omega) &= \mathcal{F}_x [G_{3D}(x, y, z, \omega)] \\ &= \int_{-\infty}^{\infty} \frac{e^{jk\sqrt{x^2+y^2+z^2}}}{4\pi\sqrt{x^2+y^2+z^2}} e^{jk_x x} dx \\ &= \frac{j}{4} H_0^{(1)} \left( \sqrt{k^2 - k_x^2} \sqrt{y^2 + z^2} \right).\end{aligned}\quad (\text{A.22})$$

The Fourier transform of Eq. (3.2) with respect to  $x$  and  $z$ , i.e., the Fourier transform of  $\tilde{G}(k_x, y, z, \omega)$  with respect to  $z$ , is derived as [111] (6.677-3), (6.677-4):

$$\begin{aligned}\tilde{G}_{3D}(k_x, y, k_z, \omega) &= \mathcal{F}_z \left[ \tilde{G}_{3D}(k_x, y, z, \omega) \right] \\ &= \int_{-\infty}^{\infty} \frac{j}{4} H_0^{(1)} \left( \sqrt{k^2 - k_x^2} \sqrt{y^2 + z^2} \right) e^{jk_z z} dz \\ &= \frac{j}{2} \frac{e^{j\sqrt{k^2 - k_x^2 - k_z^2} y}}{\sqrt{k^2 - k_x^2 - k_z^2}}\end{aligned}\quad (\text{A.23})$$

In a similar manner, the Fourier transform of the two-dimensional Green function, Eq. (3.10), is derived as:

$$\begin{aligned}\tilde{G}_{2D}(k_x, y, \omega) &= \mathcal{F}_x [G_{2D}(x, y, \omega)] \\ &= \int_{-\infty}^{\infty} \frac{j}{4} H_0^{(1)} \left( k \sqrt{x^2 + y^2} \right) e^{jk_x x} \\ &= \frac{j}{2} \frac{e^{j\sqrt{k^2 - k_x^2} y}}{\sqrt{k^2 - k_x^2}}.\end{aligned}\quad (\text{A.24})$$

### A.4.2 Derivation in cylindrical coordinates

Based on the addition theorem of the Hankel function [27],  $\tilde{G}_{3D}(r - r', \phi - \phi', k_z, \omega)$  can be described as:

$$\begin{aligned}\tilde{G}_{3D}(r - r', \phi - \phi', k_z, \omega) &= \mathcal{F}_z [G_{3D}(r - r', \phi - \phi', z, \omega)] \\ &= \frac{j}{4} H_0^{(1)}(\sqrt{k^2 - k_z^2} |\mathbf{r} - \mathbf{r}'|) \\ &= \sum_{n=-\infty}^{\infty} \frac{j}{4} H_n^{(1)} \left( \sqrt{k^2 - k_z^2} r' \right) J_n \left( \sqrt{k^2 - k_z^2} r \right) e^{jn(\phi - \phi')},\end{aligned}\quad (\text{A.25})$$

where  $\mathbf{r} = (r, \phi, 0)$ , and  $\mathbf{r}' = (r', \phi', 0)$ . It is assumed that  $r < r'$ . Therefore,

$$\begin{aligned}G_{3D}(r - r', \phi - \phi', z - z', \omega) \\ = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{j}{4} H_n^{(1)} \left( \sqrt{k^2 - k_z^2} r' \right) J_n \left( \sqrt{k^2 - k_z^2} r \right) e^{jn(\phi - \phi')} e^{jk_z(z - z')} dk_z\end{aligned}\quad (\text{A.26})$$

The cylindrical harmonic coefficients of  $G_{3D}(r - r', \phi, z, \omega)$  is derived as:

$$\begin{aligned}\tilde{G}_{3D,n}(r - r', k_z, \omega) &= \mathcal{F}_\phi \left[ \tilde{G}_{3D}(r - r', \phi, k_z, \omega) \right] \\ &= \frac{j}{4} H_n^{(1)} \left( \sqrt{k^2 - k_z^2} r' \right) J_n \left( \sqrt{k^2 - k_z^2} r \right).\end{aligned}\quad (\text{A.27})$$

When  $z$  is set as  $z = 0$ ,  $\tilde{G}_{3D,n}(r - r', 0, \omega)$  can be described by using Eq. (A.26) as:

$$\tilde{G}_{3D,n}(r - r', 0, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{j}{4} H_n^{(1)}\left(\sqrt{k^2 - k_z^2} r'\right) J_n\left(\sqrt{k^2 - k_z^2} r\right) dk_z \quad (\text{A.28})$$

This equation is used to derive the WFR equation for two-dimensional reproduction.

In a similar manner, based of the addition theorem of the Hankel function,  $G_{2D}(r - r', \phi - \phi', \omega)$  can be described as:

$$\begin{aligned} G_{2D}(r - r', \phi - \phi', \omega) &= \frac{j}{4} H_0^{(1)}(k|\mathbf{r} - \mathbf{r}'|) \\ &= \sum_{n=-\infty}^{\infty} \frac{j}{4} H_n^{(1)}(kr') J_n(kr) e^{jn(\phi - \phi')}. \end{aligned} \quad (\text{A.29})$$

Therefore, the circular harmonic coefficients of  $G_{2D}(r - r', \phi, \omega)$  is derived as:

$$\begin{aligned} G_{2D,n}(r - r', \omega) &= \mathcal{F}_\phi [G_{2D}(r - r', \phi, \omega)] \\ &= \frac{j}{4} H_n^{(1)}(kr') J_n(kr). \end{aligned} \quad (\text{A.30})$$

### A.4.3 Derivation in spherical coordinates

Based on the addition theorem of the spherical Hankel function,  $G_{3D}(r - r', \theta - \theta', \phi - \phi', \omega)$  can be described as:

$$G_{3D}(r - r', \theta - \theta', \phi - \phi', \omega) = jk \sum_{n=0}^{\infty} j_n(r) h_n(kr') \sum_{m=-n}^n Y_n^m(\theta', \phi')^* Y_n^m(\theta, \phi). \quad (\text{A.31})$$

Here, it is assumed that  $r < r'$ . Therefore, the spherical harmonic coefficients of  $G_{3D}(r - r', \theta, \phi, \omega)$  is derived as:

$$\begin{aligned} \tilde{G}_{3D,n}^m(r - r', \omega) &= \mathcal{S} [G_{3D}(r - r', \theta, \phi, \omega)] \\ &= jk j_n(r) h_n(kr') Y_n^m(0, 0)^*. \end{aligned} \quad (\text{A.32})$$

When  $m$  is set as  $m = 0$ , Eq. (A.32) is described as:

$$\begin{aligned} \tilde{G}_{3D,n}^0(r - r', \omega) &= jk j_n(kr) h_n^{(1)}(kr') Y_n^0(0, 0)^* \\ &= \sqrt{\frac{2n+1}{4\pi}} jk j_n(kr) h_n^{(1)}(kr'). \end{aligned} \quad (\text{A.33})$$

This equation is important when the WFR equation is described as spherical convolution form.

When  $\theta$  is set as  $\theta = \pi/2$ ,  $G_{3D}(r - r', \pi/2, \phi, \omega)$  can be described as:

$$\begin{aligned}
G_{3D}\left(r - r', \frac{\pi}{2}, \phi, \omega\right) &= \sum_{n=0}^{\infty} \sum_{m=-n}^n jk j_n(kr) h_n^{(1)}(kr') Y_n^m\left(\frac{\pi}{2}, 0\right)^* Y_n^m\left(\frac{\pi}{2}, \phi\right) \\
&= \sum_{m=-\infty}^{\infty} \left\{ \sum_{n=|m|}^{\infty} jk j_n(kr) h_n^{(1)}(kr') \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2 \right\} e^{jm\phi}
\end{aligned} \tag{A.34}$$

Therefore, circular harmonic coefficients of  $G_{3D}(r - r', \pi/2, \phi, \omega)$  is derived as:

$$\begin{aligned}
\tilde{G}_{3D,m}\left(r - r', \frac{\pi}{2}, \omega\right) &= \mathcal{F}_{\phi} \left[ G_{3D}\left(r - r', \frac{\pi}{2}, \phi, \omega\right) \right] \\
&= \sum_{n=|m|}^{\infty} jk j_n(kr) h_n^{(1)}(kr') \frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} P_n^m(0)^2.
\end{aligned} \tag{A.35}$$

This equation is used to derive the WFR equation for two-dimensional reproduction.

By using Eq. (A.30), the circular harmonic coefficients of  $G_{2D}(r - r', \phi, \omega)$  is derived as:

$$\begin{aligned}
G_{2D,m}(r - r', \omega) &= \mathcal{F}_{\phi} [G_{2D}(r - r', \phi, \omega)] \\
&= \frac{j}{4} H_m^{(1)}(kr') J_m(kr).
\end{aligned} \tag{A.36}$$