Spatial Acoustic Vector Based Sound Field Reproduction

Huanyu Zuo

Master of Engineering, Beijing Institute of Technology, China
Bachelor of Engineering, Beijing Institute of Technology, China

May 2021

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University

School of Engineering
College of Engineering and Computer Science
The Australian National University

Declaration

The contents of this thesis are the results of original research and have not been submitted for a higher degree to any other university or institution. Much of this work has either been published or submitted for publications as journal papers and conference proceedings. Following is a list of these papers.

Journal Publications

- **H. Zuo**, P. N. Samarasinghe, T. D. Abhayapala, and G. Dickins, "Spatial sound intensity vectors in spherical harmonic domain," *The Journal of the Acoustical Society of America (JASA)*, vol. 145, no. 2, pp. EL149–EL155, 2019.

- **H. Zuo**, P. N. Samarasinghe, T. D. Abhayapala, and G. Dickins, "Erratum: Spatial sound intensity vectors in spherical harmonic domain [j. acoust. soc. am. 145 (2), el149–el155 (2019)]," *The Journal of the Acoustical Society of America (JASA)*, vol. 146, no. 1, pp. 164–165, 2019.

- **H. Zuo**, P. N. Samarasinghe, and T. D. Abhayapala, "Intensity based spatial soundfield reproduction using an irregular loudspeaker array," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 1356–1369, 2020.

- **H. Zuo**, T. D. Abhayapala, and P. N. Samarasinghe, "Particle velocity assisted three dimensional sound field reproduction using a modal-domain approach," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 2119–2133, 2020.

- **H. Zuo**, L. I. Birnie, P. N. Samarasinghe, T. D. Abhayapala, and V. Tourbabin, "Particle velocity based mixed-source sound field translation for binaural reproduction," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 2021 (in peer review).

Conference Proceedings

- **H. Zuo**, P. N. Samarasinghe, and T. D. Abhayapala, "Exterior-interior 3d sound field separation using a planar array of differential microphones," in *Proc. 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 216–220, Tokyo, Japan, September, 2018.
- **H. Zuo**, P. N. Samarasinghe, and T. D. Abhayapala, "Intensity based sound-field reproduction over multiple sweet spots using an irregular loudspeaker array," in *Proc. 28th European Signal Processing Conference (EUSIPCO)*, pp. 486–490, Amsterdam, Netherlands, January, 2021.
- **H. Zuo**, T. D. Abhayapala, and P. N. Samarasinghe, "3D multizone sound-field reproduction in a reverberant environment using intensity matching method," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 416–420, Toronto, Canada, June, 2021.

The research work presented in this thesis has been performed jointly with Prof. Thushara D. Abhayapala and Dr. Prasanga N. Samarasinghe. Approximately 80% of this work is my own.

Huanyu Zuo
School of Engineering
The Australian National University
Canberra ACT 2601
May 2021

Acknowledgments

Without the support of the many faces in my life, this work would not have been possible. I would like to acknowledge and thank each of the following.

- First and foremost, my supervisors, Prof. Thushara D. Abhayapala and Dr. Prasanga N. Samarasinghe, for their professional guidance, consistent encouragement, and valuable friendship. They provided me with knowledge and experience not only in research, but also in many other aspects of life.
- Dr. Hanchi Chen for his valuable help with my research at the beginning of my PhD as an associate supervisor.
- Dr. Glenn Dickins for his helpful ideas and collaboration on some of this work.
- Dr. Vladimir Tourbabin and Mr. Lachlan I. Birnie for their collaboration on some of this work.
- The China Scholarship Council and The Australian National University, for the PhD scholarship and funding.
- My friends and colleagues in the Audio & Acoustic Signal Processing Group, specially Aimee, Fei, Lachlan, June, and Yonggang for their true friendship.
- My parents for their overwhelming love, unconditional support, and constant encouragement.
- Finally, my girlfriend Yuyi, for her love and emotional support.

Abstract

Spatial sound field reproduction aims to recreate an immersive sound field over a spatial region. The existing sound pressure based approaches to spatial sound field reproduction focus on the accurate approximation of original sound pressure over space, which ignores the perceptual accuracy of the reproduced sound field. The acoustic vectors of particle velocity and sound intensity appear to be closely linked with human perception of sound localization in literature. Therefore, in this thesis, we explore the spatial distributions of the acoustic vectors, and seek to develop algorithms to perceptually reproduce the original sound field over a continuous spatial region based on the vectors.

A theory of spatial acoustic vectors is first developed, where the spatial distributions of particle velocity and sound intensity are derived from sound pressure. To extract the desired sound pressure from a mixed sound field environment, a 3D sound field separation technique is also formulated. Based on this theory, a series of reproduction techniques are proposed to improve the perceptual performance.

The outcomes resulting from this theory are: (i) derivation of a particle velocity assisted 3D sound field reproduction technique which allows for non-uniform loudspeaker geometry with a limited number of loudspeakers, (ii) design of particle velocity based mixed-source sound field translation technique for binaural reproduction that can provide sound field translation with good perceptual experience over a large space, (iii) derivation of an intensity matching technique that can reproduce the desired sound field in a spherical region by controlling the sound intensity on the surface of the region, and (iv) two intensity based multizone sound field reproduction algorithms that can reproduce the desired sound field over multiple spatial zones. Finally, these techniques are evaluated by comparing to the conventional approaches through numerical simulations and real-world experiments.

List of Acronyms

CMOS	Comparison mean opinion score
DirAC	Directional audio coding
DOA	Direction of arrival
HOA	Higher order Ambisonics
HRTF	Head related transfer function
ILD	Interaural level difference
IM/MM	Intensity matching/Mode matching
IRLS	Iteratively reweighted least squares
ITD	Interaural time difference
LASSO	Least absolute shrinkage and selection operator
MUSHRA	Multiple stimulus with hidden reference and anchor
NNLS	Non-negative least squares
P-CFS	Pressure based closed-form solution
P-SS	Pressure based sparse solution
PM(M)	Pressure matching (method)
PVM	Pressure and velocity matching
SNR	Signal to noise ratio
TWS	True wireless stereo
V-LSS	Velocity based least squares solution
V-SS	Velocity based sparse solution
VAM	Velocity assisted method
VBAP	Vector base amplitude panning
VR/AR	Virtual reality/augmented reality
WFS	Wave field synthesis
WGN	White Gaussian noise

Notations and Symbols

$\lceil \cdot \rceil$	Ceiling operator
$(\cdot)^*$	Complex conjugate
$[\cdot]^T$	Transpose of a matrix
$[\cdot]^H$	Complex conjugate transpose of a matrix
$\lvert \cdot \rvert$	Absolute value
$\lvert\lvert \cdot \rvert\rvert_1$	ℓ_1 norm
$\lvert\lvert \cdot \rvert\rvert_2$	ℓ_2 norm
$\nabla^2(\cdot)$	Laplacian operator
$\vec{\nabla}$	Gradient
$\boldsymbol{A}^\dagger$	Matrix psuedoinverse: $\boldsymbol{A}^\dagger = [\boldsymbol{A}^H \boldsymbol{A}]^{-1} \boldsymbol{A}^H$
$\boldsymbol{x} \cdot \boldsymbol{y}$	Dot product between two vectors
$\boldsymbol{x} \times \boldsymbol{y}$	Cross product between two vectors
$\mathrm{Re}\{\cdot\}$	Real part
$\mathrm{Im}\{\cdot\}$	Imaginary part
$\delta\{\cdot\}$	Dirac delta function
$\delta_{nm}\{\cdot\}$	Kronecker delta function
i	$\sqrt{-1}$

Contents

Declaration **i**

Acknowledgements **iii**

Abstract **v**

List of Acronyms **vii**

Notations and Symbols **ix**

List of Figures **xvii**

List of Tables **xxvii**

1 Introduction **1**

1.1 Motivation and Scope 1

1.2 Problem Description and Proposed Solution 7

1.3 Thesis Overview and Outline 8

2 Background Theory: Spatial Sound Field Reproduction Using Spherical Harmonic Synthesis **13**

2.1 Spherical Harmonic Decomposition of a Sound Field 13

2.1.1 Sound field expansion 14

2.1.2 Properties of the expansion 18

2.1.3 Soundfield coefficient estimation from measurements 22

2.2 Spherical Harmonic Synthesis of Mode Matching 23

2.2.1 Single zone solution 24

2.2.2 Multizone solution 26
2.3 Summary 27

3 Theory of Spatial Acoustic Vectors 29
3.1 Desired Sound Field Extraction 30
3.1.1 Introduction 30
3.1.2 Problem formulation 31
3.1.3 Sound field recording using a planar array of differential microphones. 32
3.1.4 Sound field separation 34
3.1.5 Simulation results 37
3.2 Spatial Particle Velocity Vector 40
3.2.1 Introduction 40
3.2.2 Problem formulation 41
3.2.3 Velocity coefficients 42
3.2.4 Velocity truncation error 43
3.3 Spatial Sound Intensity Vector 45
3.3.1 Introduction 45
3.3.2 Problem formulation 46
3.3.3 Intensity coefficients 47
3.3.4 Intensity truncation error 49
3.4 Summary and Contributions 51
3.5 Related Publications 52
3.6 Appendix 53
3.6.1 Proof of Theorem 1 53
3.6.2 Proof of Theorem 2 55

4 Particle Velocity Assisted Spatial Sound Field Reproduction 57
4.1 Introduction 57
4.2 Problem Definition 59
4.3 Velocity Assisted Reproduction Algorithm 60
4.3.1 Point-to-region transfer function 61
4.3.2 Velocity assisted optimization 62

4.3.3 Reproduction error 65
4.4 Simulations 65
4.4.1 Simulation setup 67
4.4.2 Determination of the weighting coefficient 68
4.4.3 Simulation results in free field 70
4.4.4 Simulation results in reverberant rooms 75
4.5 Experimental Verification 82
4.6 Summary and Contributions 86
4.7 Related Publications 87
4.8 Appendix 88
4.8.1 Proof of equation (4.7) 88
4.8.2 Proof of equation (4.18) 89

5 Particle Velocity Based Sound Field Translation for Binaural Reproduction 91
5.1 Introduction 92
5.2 Problem Formulation 94
5.3 Particle Velocity Based Mixed-Source Expansion 96
5.3.1 Mixed-source model 96
5.3.2 Particle velocity based expansion 97
5.4 Sound Field Translation and Synthesis for Binaural Reproduction . 100
5.5 Simulation Study 102
5.5.1 Simulation setup and criteria 102
5.5.2 Simulation results 104
5.5.3 Discussion 113
5.6 Experimental Verification 114
5.6.1 Experimental methodology 115
5.6.2 Experimental results 116
5.7 Summary and Contributions 118
5.8 Related Publications 119

6 Spatial Sound Field Reproduction Based on Sound Intensity 121
6.1 Introduction 122

6.2 Problem Formulation 124
6.3 Loudspeaker Weights Design 124
6.4 Simulations 129
6.4.1 Simulation setup and criteria 129
6.4.2 Performance comparison for the intensity matching and the pressure matching 132
6.4.3 Performance comparison for the intensity matching and the HOA max-r_E decoding 138
6.4.4 Robustness analysis to microphone noise 142
6.5 Perceptual Localization Experiments 144
6.5.1 Experimental setup and criteria 146
6.5.2 Experimental results 147
6.6 Summary and Contributions 149
6.7 Related Publications 150

7 Intensity Based Sound Field Reproduction over Multiple Spatial Zones 151
7.1 Introduction 151
7.2 Sound Field Reproduction over Multiple Sweet Spots 153
7.2.1 Problem formulation 153
7.2.2 Multiple sweet spots reproduction algorithm 153
7.3 Spatial Multizone Reproduction in a Reverberant Room 156
7.3.1 Problem Formulation 156
7.3.2 Multizone reproduction algorithm for a reverberant room . . 157
7.4 Evaluations 160
7.4.1 Multiple sweet spots reproduction 161
7.4.2 Multizone reproduction 166
7.5 Summary and Contributions 169
7.6 Related Publications 170

8 Conclusions and Future Work 171
8.1 Conclusions 171
8.2 Future Work 174

Bibliography **177**

List of Figures

1.1 Layout of a 7.1 surround sound system [1]. 2

1.2 The signal flow of a binaural synthesis system for headphones [2]. . 5

1.3 Thesis overview. Blocks in orange and blue correspond to the chapters answering sub-question (i) and (ii), respectively. 9

2.1 Definition of the spherical coordinate system throughout this thesis. 14

2.2 Plots of the spherical Bessel function $j_n(x)$ for n from 0 to 10. . . . 19

3.1 Example of planar arrays of differential microphones 32

3.2 Interior and exterior sound field reconstruction on a spherical surface with radius of 0.08 m. The frequency of the sources is 850 Hz. (a) Original interior sound field, (b) original exterior sound field, (c) reconstructed interior sound field without noise, (d) reconstructed exterior sound field without noise, (e) reconstructed interior sound field with SNR of 40 dB, (f) reconstructed exterior sound field with SNR of 40 dB. 38

3.3 Reconstruction error of interior and exterior sound field against different radii of spherical surfaces. The frequency of the sources is 850 Hz. 39

3.4 The normalized truncation error with respect to velocity truncation order for (a) velocity in the θ direction and (b) velocity in the ϕ direction with various pressure truncation orders N. 45

3.5 The relationship between normalized truncation error and intensity truncation order P_θ for various pressure truncation orders N. 50

3.6 Sound intensity on a sphere with radius of 0.05 m, generated by a plane wave from $(3\pi/4, 5\pi/6)$, with frequency 600 Hz. (a-c) Sound intensity in r, θ and ϕ direction, separately, calculated using the proposed theory, (d-f) sound intensity in r, θ and ϕ direction, separately, obtained from point by point measurement. 51

4.1 Geometry of the sound field reproduction system in this paper. An array of loudspeakers is located outside χ with driving signals $w_\ell(k)$ and acoustic transfer function $H_\ell(\boldsymbol{x}, k)$ from the ℓth loudspeaker to a point $\boldsymbol{x}$ within the region. 59

4.2 The impact of the weighting coefficient on the reproduction system for various frequencies in (a) the free field and (b) the reverberant environment (the large room) in the 22 channel loudspeaker layout. The sub-figures on the left are the enlarged curves in the rectangular boxes correspondingly. 69

4.3 Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for the 22 channel system. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by the pressure matching method (PMM); (c) Reproduced sound field controlled by the velocity-assisted method (VAM); (d) Difference field controlled by PMM; (e) Difference field controlled by VAM. 71

4.4 Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for the 25 channel system. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by PMM; (c) Reproduced sound field controlled by VAM; (d) Difference field controlled by PMM; (e) Difference field controlled by VAM. 72

4.5 3D reproduction error with respect to the virtual source positions controlled by (a, c) PMM and (b, d) VAM for (a, b) the 22 channel system and (c, d) the 25 channel system. The source frequency is 1200 Hz. 73

4.6 3D reproduction error with respect to the change of frequency controlled by both methods for (a) the 22 channel system and (b) the 25 channel system. 74

4.7 The intensity direction error on plane $z = 0$ controlled by (a, c) PMM and (b, d) VAM for (a, b) the 22 channel system and (c, d) the 25 channel system. The source frequency is 1200 Hz. 76

4.8 Reconstructed sound field on the observation plane controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz. 77

4.9 Difference between the desired sound field and the reconstructed sound field controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz. 78

4.10 3D reproduction error with respect to the virtual source positions controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz. 79

4.11 3D reproduction error with respect to the change of frequency controlled by both methods in (a) the large room and (b) the small room in case of the 22 channel system. 80

4.12 The intensity direction error on plane $z = 0$ controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz. 81

4.13 Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for source #1. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by PMM; (c) Reproduced sound field controlled by VAM; (d) Difference field controlled by PMM; (e) Difference field controlled by VAM. 83

4.14 Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for source #2. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by PMM; (c) Reproduced sound field controlled by VAM; (d) Difference field controlled by PMM; (e) Difference field controlled by VAM. 84

4.15 3D reproduction error with respect to the change of frequency controlled by both methods for (a) source #1 and (b) source #2. . . . 85

4.16 The intensity direction error on plane $z = 0$ controlled by (a, c) PMM and (b, d) VAM for (a, b) source #1 and (c, d) source #2. The source frequency is 1200 Hz. 86

5.1 The true pressure field and recorded pressure field at 1500Hz in the single source scenario. The black circle denotes the receiver region. (a) the true pressure field; (b) the recorded pressure field. 105

5.2 The reconstructed pressure field and the error field controlled by the closed-form solution (P-CFS) and the sparse solution (P-SS) using the sound pressure based expansion in the single source scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by P-CFS ; (b) Error field controlled by P-CFS; (c) Reconstructed pressure field controlled by P-SS; (d) Error field controlled by P-SS. 106

5.3 The reconstructed pressure field and the error field controlled by the least squares solution (V-LSS) and the sparse solution (V-SS) using the particle velocity based expansion in the single source scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by V-LSS ; (b) Error field controlled by V-LSS; (c) Reconstructed pressure field controlled by V-SS; (d) Error field controlled by V-SS. 107

5.4 The true pressure field and recorded pressure field at 1500Hz in the multiple sources scenario. The black circle denotes the receiver region. (a) the true pressure field; (b) the recorded pressure field. . 108

5.5 The reconstructed pressure field and the error field controlled by the closed-form solution (P-CFS) and the sparse solution (P-SS) using the sound pressure based expansion in the multiple sources scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by P-CFS ; (b) Error field controlled by P-CFS; (c) Reconstructed pressure field controlled by P-SS; (d) Error field controlled by P-SS. 109

5.6 The reconstructed pressure field and the error field controlled by the least squares solution (V-LSS) and the sparse solution (V-SS) using the particle velocity based expansion in the multiple sources scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by V-LSS ; (b) Error field controlled by V-LSS; (c) Reconstructed pressure field controlled by V-SS; (d) Error field controlled by V-SS. 110

5.7 The directional error ζ_V and the magnitude of $\boldsymbol{r}_V$ with respect to frequency at the translated position of (0.2, 0.3, 0) m in the single source scenario. (a) Directional error ζ_V; (b) Magnitude of $\boldsymbol{r}_V$. . . . 111

5.8 The directional error ζ_E and the magnitude of $\boldsymbol{r}_E$ with respect to frequency at the translated position of (0.2, 0.3, 0) m in the single source scenario. (a) Directional error ζ_E; (b) Magnitude of $\boldsymbol{r}_E$. . . . 111

5.9 The directional error ζ_V and the magnitude of $\boldsymbol{r}_V$ as a function of translation distance along the positive y-axis in the single source scenario. The frequency of the source is 300 Hz. (a) Directional error ζ_V; (b) Magnitude of $\boldsymbol{r}_V$. 112

5.10 The directional error ζ_E and the magnitude of $\boldsymbol{r}_E$ as a function of translation distance along the positive y-axis in the single source scenario. The frequency of the source is 1500 Hz. (a) Directional error ζ_E; (b) Magnitude of $\boldsymbol{r}_E$. 112

5.11 Results of the listening experiment scores for the source localization test and the basic audio quality test. Each box bounds the 25th and 75th percentile of the scores with the central red line indicating the median value, and the whiskers are the outward extension of the 25th and 75th percentile by 1.5 times of the interquartile range. The v-shaped notches represent the 95% confidence intervals. The symbols + denote the outliers and the symbols $*$ denote the mean values. (a) Source localization test; (b) Basic audio quality test. . . 117

6.1 Geometry of soundfield reproduction system in this paper. An irregular array of loudspeakers is located outside the spherical region. 123

6.2 2D sketch of (a) the 5 channel system layout and (b) the 22 channel system layout. 130

6.3 Desired sound intensity in (a) r direction, (b) θ direction, and (c) ϕ direction induced by a plane wave with frequency of 900 Hz. 131

6.4 Reconstructed sound intensity in (a, d) r direction, (b, e) θ direction, and (c, f) ϕ direction controlled by (a-c) the proposed method, and (d-f) the pressure matching method in case of the 5 channel system. The source frequency is 900 Hz. 132

6.5 Reconstructed sound intensity in (a, d) r direction, (b, e) θ direction, and (c, f) ϕ direction controlled by (a-c) the proposed method, and (d-f) the pressure matching method in case of the 22 channel system. The source frequency is 900 Hz. 133

6.6 Relative error with respect to the radius for (a) the 5 channel system and (b) the 22 channel system controlled by the intensity matching (IM) and the pressure matching (PM). The source frequency is 900 Hz. 134

6.7 Relative error on the surface of the target region with respect to the incident directions for (a) the 5 channel system and (b) the 22 channel system controlled by IM and PM. Symbol ■ denotes the loudspeakers placed on the plane $z = 0$. The source frequency is 900 Hz. 135

6.8 Relative error on the surface of the target region with respect to the frequency for (a) the 5 channel system and (b) the 22 channel system controlled by IM and PM. 135
6.9 The intensity direction error of (a,c) the 5 channel system and (b,d) the 22 channel system on plane $z = 0$ controlled by (a-b) IM and (c-d) PM. Black circles denote the target region. The source frequency is 900 Hz. 137
6.10 Reconstructed sound intensity in (a, d) r direction, (b, e) θ direction, and (c, f) ϕ direction controlled by the HOA max-r_E decoding in case of (a-c) the 5 channel system, and (d-f) the 22 channel system. The source frequency is 900 Hz. 139
6.11 Relative error with respect to the radius for (a) the 5 channel system and (b) the 22 channel system controlled by the intensity matching (IM) and the HOA max-r_E decoding (rE). The source frequency is 900 Hz. 141
6.12 Relative error on the surface of the target region with respect to the incident directions for (a) the 5 channel system and (b) the 22 channel system controlled by IM and rE. Symbol ■ denotes the loudspeakers placed on the plane $z = 0$. The source frequency is 900 Hz. 141
6.13 The intensity direction error of (a) the 5 channel system and (b) the 22 channel system on plane $z = 0$ controlled by the HOA max-r_E decoding. Black circles denote the target region. The source frequency is 900 Hz. 142
6.14 Averaged relative error across the entire target region with respect to SNR for (a) the 5 channel system and (b) the 22 channel system. The source frequency is 900 Hz. 143
6.15 Relative error with respect to the radius for (a) the 5 channel system and (b) the 22 channel system under the noiseless and noisy conditions. The source frequency is 900 Hz. 144
6.16 The intensity direction error of (a) the 5 channel system and (b) the 22 channels when noise exists. The source frequency is 900 Hz. . . . 145
6.17 Experimental setup for the 5 channel system. 145

6.18 Results of the perceptual localization experiments in the case of the 5 channel system controlled by the intensity matching (red lines with ◯) and the pressure matching (blue lines with □) for desired sound source (a) R1, (b) R2, and (c) R3. The symbols + denote the source positions and the symbols × denote the loudspeakers on the plane $z = 0$ of the 5 channel system. 148
6.19 CMOS scores for the three desired sources with 95% confidence intervals. 148

7.1 Geometry of soundfield reproduction system in this paper. The sweet spots are arbitrarily chosen within the spherical region of radius R. An irregular array of loudspeakers is located outside the spherical region. 154
7.2 Geometry of the multizone reproduction system in this work. The loudspeaker array surrounds the zones. 157
7.3 3D sketch of the 8-channel loudspeaker array. The blue squares denote the loudspeakers. 161
7.4 Polar plots of DO and CR at sweet spots using the intensity matching (IM) and the pressure and velocity matching (PVM) for the virtual source moving on the horizontal plane ((a) sweet spot #1, (c) #2 and (e) #3) and the vertical plane ((b) sweet spot #1, (d) #2 and (f) #3). The source frequency is 600 Hz. 163
7.5 The intensity direction error on plane $z = 0$ controlled by (a) the intensity matching and (b) the pressure and velocity matching. Black circles denote human head zones around the sweet spots. The source frequency is 600 Hz. 164
7.6 The intensity direction error with respect to frequency. 165
7.7 CMOS scores for the three sweet spots with 95% confidence intervals. 165
7.8 Relative error with respect to the radius for (a) zone #1 and (b) zone #2 controlled by the intensity matching (IM) and the mode matching (MM). The frequency of the sources is 900 Hz. 167
7.9 Relative error on the surface of the target region with respect to the frequency for (a) zone #1 and (b) zone #2 controlled by IM and MM. 167

7.10 The intensity direction error on plane $z = 0$ for (a, c) zone #1 and (b, d) zone #2 controlled by (a-b) IM and (c-d) MM. Black circles denote the target zones. The frequency of the sources is 900 Hz. . . 168

List of Tables

4.1 Loudspeaker positions of the 22 channel system 66

6.1 Loudspeaker positions of the 5 channel system 129
6.2 The locations of three desired sound sources 145
6.3 CMOS scale 146

7.1 Loudspeaker locations of the 8-channel array. 160
7.2 Average absolute perceived direction error. 164

Chapter 1

Introduction

1.1 Motivation and Scope

In our daily life, sound is all around us, moving freely in any direction. For example, when watching a live show of a concert, you can hear the sound of instruments from different directions and can feel the voice change as the vocalist moves on the stage. The recreation of such a lifelike audio experience is a challenge for a spatial sound field reproduction system, whereas the ability of this recreation is vital to many commercial applications such as home entertainment systems (e.g., 5.1/7.1 systems), modern cinemas (e.g., Dolby Atmos and DTS:X), and Virtual Reality/Augmented Reality (VR/AR). In this thesis, we focus on this challenge and find ways to create an immersive sound field over a spatial region so that the listener inside the region can experience a realistic but virtual replication of the original sound.

Sound field reproduction is achieved by controlling the placement of a set of loudspeakers usually put on the boundary that encloses the spatial region of interest and deriving the signals emitted from the loudspeakers [2]. As an example, a general layout of a 7.1 surround sound system is shown in Fig. 1.1.

Until now, there have been various studies of sound field reproduction and they have evolved in two directions: physically motivated techniques and perceptually motivated techniques [3]. Physically motivated techniques aim to reproduce an accurate physical approximation of a sound field. A notable example is Wave

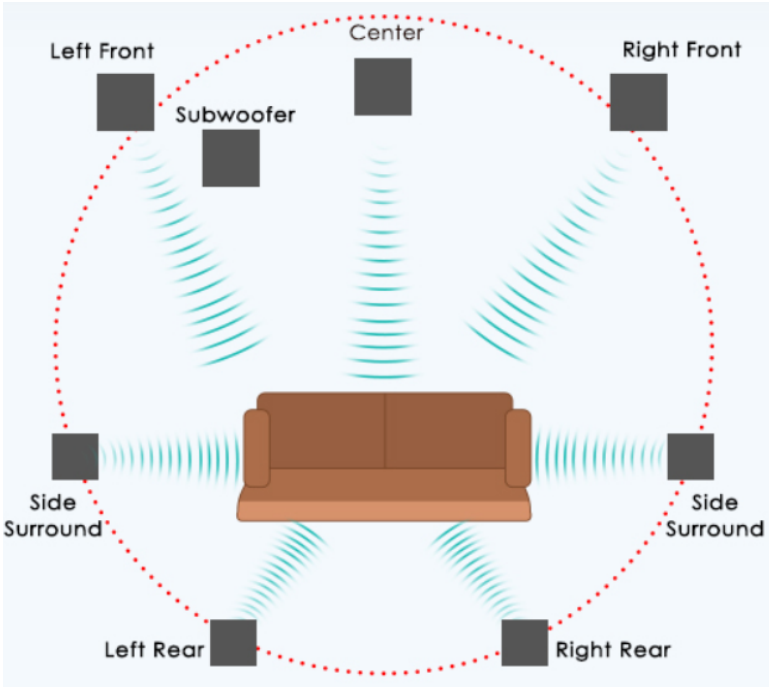


Figure 1.1: Layout of a 7.1 surround sound system [1].

Field Synthesis (WFS) [4–6], which synthesizes the desired sound field in a target region based on the Huygens–Fresnel principle by individually driven loudspeakers. Theoretically, it requires a continuous distribution of monopole and normally oriented dipole secondary sources on the boundary of the region [7]. An approximation to the continuous distribution is implemented by an array of equally placed loudspeakers, and the reproduction artifacts due to the approximation is analyzed in [8]. To facilitate the implementation of WFS in a real reproduction environment, an adaptive WFS algorithm is proposed in [6]. Two dimensional WFS using linear and planar arrays is thoroughly studied in [9–12]. To enhance the synthesis accuracy of azimuthal localization around the listener's position, a series of local WFS methods are presented and compared with each other [13]. A practical implementation is called 2.5D WFS, which controls 2D sound fields using point sources serving as the secondary sources instead of line sources [14, 15]. However, there is a limited number of work implementing WFS in 3D space due to the fact that a significant number of loudspeakers are required for broadband reproduction over a large region [16–18].

Ambisonics [19], which is another physically motivated approach to soundfield reproduction, is based on the first order spherical harmonic analysis of the soundfield to be reproduced. To achieve a large listening area, high order spherical harmonics have been introduced in Higher Order Ambisonics (HOA) [20–24], which is also widely used in multi-zone soundfield reproduction systems [25–27]. Array geometry is an important factor for HOA. Spherical arrays were extensively studied in literature [20, 28–30], where there is a requirement to place loudspeakers uniformly on a sphere that surrounds the target region so that the desired sound field within the region can be reconstructed perfectly. To facilitate the implementation of spherical arrays, multiple circular arrays have been proposed in HOA [31–33]. Chen *et al.* developed a method to reproduce 3D exterior sound field with a planar arrays of dipole (or first-order) loudspeakers [33]. However, the performance of HOA deteriorates when the geometry of loudspeaker array becomes neither spherical nor circular due to the poorly conditioned matrix inversion. An improved least squares method with a weighted penalty function was developed for such irregular loudspeaker layouts [34]. This method requires the regularization parameter to be dependent on each loudspeaker arrangement. In [35], the authors proposed a panning method, named all-round Ambisonic panning, which can recreate the sound with stable loudness using arbitraty loudspeaker arrangements. Ueno *et al.* proposed a weighted mode matching method, which is flexible with various loudspeaker geometries, to avoid a relatively large effort devoted to matching the insignificant modes [36]. This method was also exploited for multizone reproduction [37, 38]. Sparse methods based HOA was investigated in [39, 40], which results in a significantly reduced number of the required microphones for the measuring process. In addition to the problem in terms of loudspeaker placement, there is another limitation for HOA. For a given order of a reproduction system, which is determined by the maximum frequency of the desired sound and the radius of the target region, HOA requires sufficient loudspeakers to match all the spherical harmonics to the given order in order to avoid spatial aliasing [20].

Sound field reproduction can also be implemented by controlling sound pressure at a set of discrete sampling points, which is basically based on the least squares method [41, 42]. To improve the performance of the reproduction system in a large area, other optimization techniques are exploited, such as the wave field

reconstruction filter [43], the singular value decomposition based optimization [44], and the least absolute shrinkage and selection operator (LASSO) [45, 46].

In addition to their own shortcomings, the aforementioned physically motivated techniques have a common limitation. They derive the loudspeaker driving signals usually by minimizing the error between the desired sound pressure and the reconstructed sound pressure without considering the perceptual performance. Recently, the perceptual performance of these techniques has been evaluated [47, 48]. A detailed comparison of these techniques on the perceptual performance was presented in [49]. It shows that the original sound image can be perceptually reconstructed in some applications of these techniques, even for the off-center positions [50, 51]. However, the performance largely depends on the loudspeaker arrangement. For an irregular loudspeaker array that cannot perfectly reconstruct the desired sound pressure, these techniques with the goal of optimally reconstructing sound pressure may lead to perception deterioration.

The perceptually motivated techniques are more appropriate for such cases, which aim to render the perceptually relevant aspects of the original sound. For example, there are binaural techniques [52], which provide a convincing experience over two channels. Binaural synthesis is based on the knowledge of the acoustic transfer paths between the source and the two ears. These paths are characterized by head related transfer function (HRTF) in the frequency domain [3]. HRTF is not only determined by the relative position of the source and the ear, but it also varies between people [53]. Therefore, individualized HRTF is normally used for binaural synthesis. Also, dynamic and environmental cues should be considered to achieve high fidelity [2]. The signal flow of a binaural synthesis system for headphones is shown in Fig. 1.2. A representative binaural synthesis method is the magnitude least square, which designs the binaural rendering filters based on Ambisonic signals using magnitude-only optimization [54–56]. The binaural synthesis has been widely applied in personal entertainment, such as commercial true wireless stereo (TWS) headphones and VR/AR products. There have also been some works on multi-channel perceptually motivated systems. One notable example is Vector Base Amplitude Panning (VBAP) [57–59], which is designed to recreate the sound image of a virtual source over flexible loudspeaker layouts. VBAP is based on a triplet-wise panning law and the virtual source cannot be positioned outside

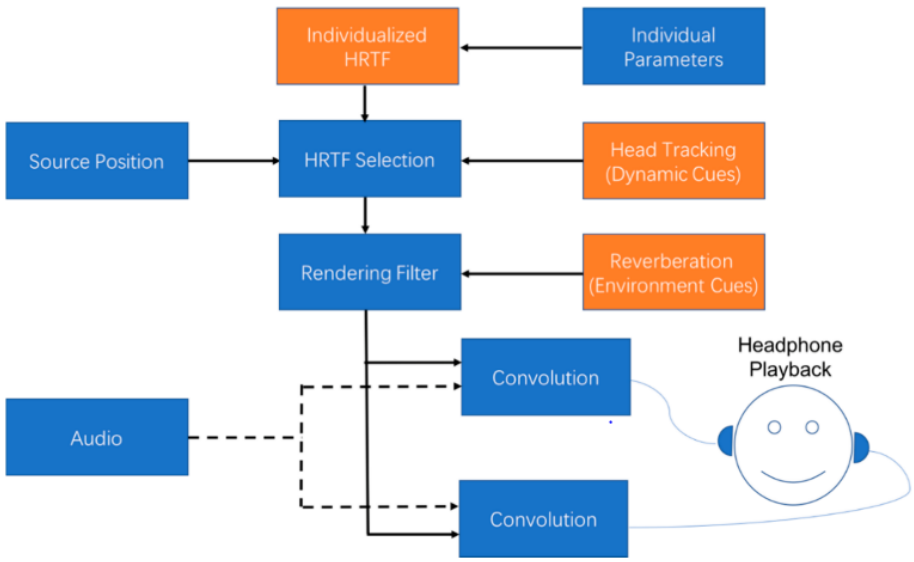


Figure 1.2: The signal flow of a binaural synthesis system for headphones [2].

a specific region, which decreases its applicability. Another example is directional audio coding (DirAC) based on spatial impulse response rendering [60, 61]. It records the impulse response of a room and analyzes the diffuseness and direction of arrival of sound from the recordings. During the synthesis step, DirAC assigns the analyzed direction and diffuseness level for each time-frequency bin to reconstruct the original perception. An improved DirAC was presented in the spherical harmonic domain by combining the perceptual effectiveness of DirAC with HOA signals [62]. Recently, Sena *et al.* also developed a new multi-channel perceptual soundfield reconstruction system [63], which captures the directional cues by a designed microphone array, and reconstructs them using a loudspeaker array in the same configuration as the microphone array. However, the reproduction quality of these methods for the off-center positions is not as good as that for the central listening position.

Acoustic vectors, such as particle velocity vector and sound intensity vector, are important acoustic quantities that contain energy and directivity information of a sound field. These two acoustic vectors also imply the information of interaural time difference (ITD) and interaural level difference (ILD), which are the two important human localization cues. In [64], Gerzon first developed particle

velocity and sound intensity theories of sound localization for reproducing psychoacoustically optimum sound, and showed the particle velocity based r_V vector and sound intensity based r_E vector satisfy low frequency (below 700 Hz) ITD localization and high frequency (above 500 Hz) ILD localization, respectively. The r_E vector, a measure of superimposed intensities, was then extended to predict perceptual sweet area and describe the changes in the timbre [50, 65]. Recently, these acoustic vectors (i.e., particle velocity and sound intensity) have been widely controlled in soundfield reproduction systems to improve the performance of perceptual localization. In addition to good performance for regularly or evenly placed loudspeakers [66–69], the particle velocity or sound intensity based methods also perform well in irregular loudspeaker arrangements [70, 71]. A sound signal conversion method between different loudspeaker systems is proposed in [72], which reproduces the spatial impression of the original sound by maintaining the pressure and direction of sound (controlled by particle velocity). A recent particle velocity based sound field reproduction technique was proposed to achieve good localization performance over a large region [70, 71], however, it requires a prior knowledge of discrete particle velocity distributions on a sphere. For the sound intensity based sound field techniques, they have been extensively investigated with subjective experiments in [50, 73, 74]. However all the previous works based on sound intensity only focus on a single reproduction position, and therefore perception degrades when the listener is moved from this exact reproduction position.

To reproduce the sound field over a large continuous region, the prerequisite is to capture the accurate distributions of the sound field over a continuous space. The spatial sound field is usually captured by a microphone array with a 3D geometry, such as EigenMike (a spherical microphone array) [75]. The estimation of spatial sound field using spherical arrays has been extensively analyzed in [76, 77]. To facilitate the real-world implementations, a special planar microphone array that can record 3D spatial sound field is also developed [78]. All these works aim to capture the spatial pressure distribution of a sound field. As for the spatial distributions of acoustic vectors, they can be approximated by measuring the vectors at multiple sampling points over space. For a large region, it requires a large number of sampling points to ensure the measurement accuracy. The dominating method of measuring the vectors is based on the combination of two pressure microphones

and it can mainly be divided into two categories. One is to use the cross spectrum of the pressure measurements [79–84] and another is to approximate particle velocity using two closely placed microphones [85–87]. In [88], authors studied the finite difference approximation error of the latter. With the development of hardware, a micro-machined particle velocity transducer called Microflown has become available [89]. Witold *et al.* determined the spatial distribution of sound intensity using the Microflown coupled with a condenser microphone [90]. Then, an intensity probe based on the Microflown in combination with a small pressure microphone is available for intensity measurement [91, 92]. This intensity probe was compared to the method with the combination of two pressure microphones in [93]. The intensity probe is more difficult to be calibrated but it can eliminate phase mismatch in the measurement system of two pressure microphones. These methods can accurately measure the vectors at a point, however, it is time-consuming and costly to obtain the spatial distributions of the vectors by point-by-point measurement using either of the methods.

Based on the above discussions, we pose the following research problem to address in this thesis:

How to recreate an immersive experience for listeners over single or multiple spatial regions using practical reproduction systems based on spatial acoustic vectors?

1.2 Problem Description and Proposed Solution

We breakdown this research problem into two further sub-questions:

(i) How to accurately achieve desired acoustic vector distributions over space from a mixed sound field environment using a practical microphone array?

We start with a mixed sound field environment over space, where the desired and undesired sound field co-exist. We develop an algorithm to extract the desired sound field from the mixed sound field using a practical microphone array. The desired sound field is characterized by spherical harmonic coefficients of sound pressure. The control objective for spatial acoustic vector based sound field reproduction should be the acoustic vectors over the entire target region rather than

multiple sampling points within the region. Therefore, we derive the closed form representations of the acoustic vectors over space from spherical harmonic coefficients of sound pressure. This means that we can obtain the continuous acoustic vector distributions over space from a measurement of a microphone array. There are infinite closed form coefficients for the acoustic vectors representations. The truncation theorems are also proposed for practical implementations.

(ii) How to perceptually reproduce the desired sound field over a continuous region by exploiting the spatial acoustic vector distributions?

Once we have the desired acoustic vector distributions over space, we can design the reproduction system to optimize the distributions over the target region. We develop a series of reproduction techniques to improve the perceptual performance, all of which are based on the theory of spatial acoustic vectors. The spatial acoustic vectors include spatial particle velocity vector and spatial sound intensity vector. Therefore, the developed techniques fall into two categories. One is the particle velocity based technique. We first propose a particle velocity assisted 3D sound field reproduction technique, which can provide accurate sound field reproduction over a large region with a limited number of loudspeakers. We then develop a new algorithm to recreate the sense of spatial direction in binaural reproduction by combining the spatial particle velocity vector with the virtual mixed-source model and HRTF synthesis. Another is the sound intensity based technique. We overcome the limitation of single sweet spot in the previous works of sound intensity based reproduction and propose a reproduction technique that can reconstruct the original sound image over a continuous spatial region. We also extend this technique to multizone reproduction.

1.3 Thesis Overview and Outline

This thesis provides some useful tools for spatial sound field analysis, and various spatial sound field reproduction techniques based on the tools. The flow diagram of the thesis overview is shown in Fig. 1.3 where the blocks represent core chapters. Blocks in orange and blue correspond to the chapters answering sub-question (i) and (ii), respectively. Chapter 3 produces two useful tools, i.e., sound field separation and spatial acoustic vectors including particle velocity vector and sound

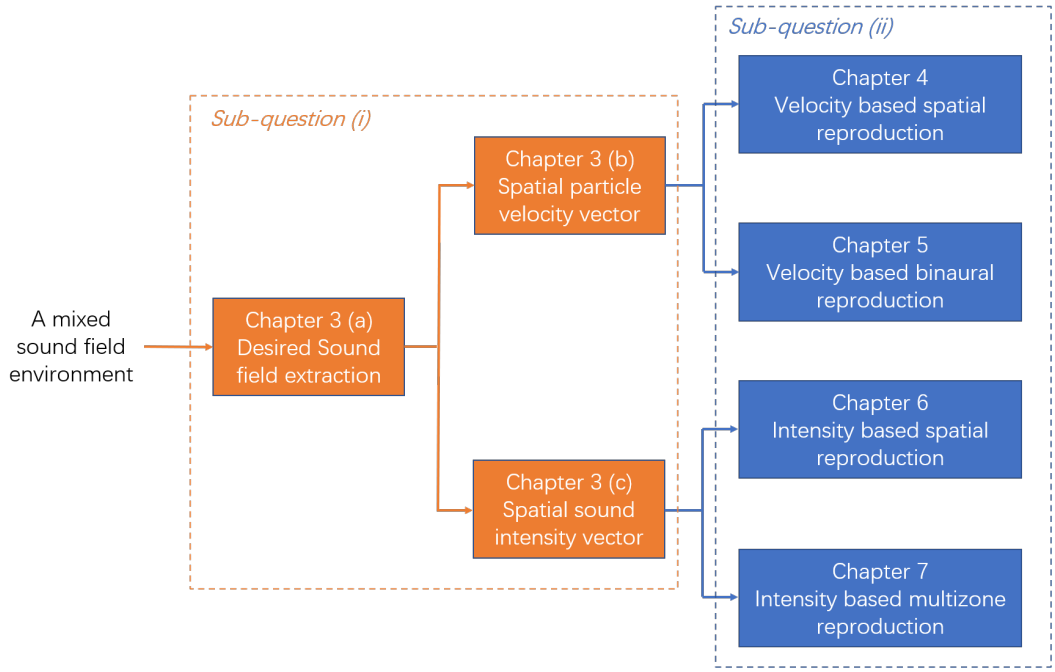


Figure 1.3: Thesis overview. Blocks in orange and blue correspond to the chapters answering sub-question (i) and (ii), respectively.

intensity vector. These tools can be used as standalone applications as well as an assist to the spatial sound field reproduction algorithms introduced in the following chapters of this thesis (see the arrow paths in Fig. 1.3). The contributions to spatial sound field reproduction in this thesis are distributed into four chapters. Chapter 4 and 5 produce two advanced sound field reproduction algorithms based on particle velocity, whereas Chapter 6 and 7 provide solutions to spatial sound field reproduction using sound intensity vectors.

The structure of the thesis is as follows:

Chapter 2: Background Theory

This chapter briefly reviews the theory of spherical harmonic analysis for spatial sound, which are used throughout this thesis. We also outline the concept of the conventional mode matching method for spatial sound field reproduction, which is based on the spherical harmonic analysis. This method is also the benchmark when we evaluate the proposed methods in this thesis.

Chapter 3: Theory of Spatial Acoustic Vectors

In this chapter, we propose a series of tools for spatial sound field reproduction. First, we develop a spatial sound field separation method to extract the desired sound field from a complex sound environment. This method can achieve 3D sound field separation using a planar array of differential microphones. The separated spatial sound field by this method is described by pressure coefficients in the spherical harmonic domain. From pressure coefficients we then theoretically derive the representations of spatial particle velocity vector and spatial sound intensity vector. We conclude that the spatial particle velocity/sound intensity vector can be characterized by the their corresponding coefficients, which are directly linked with pressure coefficients. We also indicate a truncation order for these representations.

Chapter 4: Particle Velocity Assisted Spatial Sound Field Reproduction

This chapter aims to reproduce the desired sound field in the target region by exploiting the theory of spatial particle velocity. To achieve this goal, we build a weighted cost function to optimize the distributions of sound pressure and particle velocity on the boundary of the target region. The proposed method allows for non-uniform loudspeaker geometry with a limited number of loudspeakers. Compared to the benchmark of mode matching method, this method can provide more accurate sound field reproduction with a wider frequency range using a non-uniform loudspeaker array. The proposed method can also reconstruct the original sound direction in various environments. We consider both free fields and reverberant environments when modeling the reproduction system. The reverberation can compensate for the non-uniformity of the loudspeaker array. We implement this method using the impulse response measurements of a real-world room to show its superiority.

Chapter 5: Particle Velocity Based Sound Field Translation for Binaural Reproduction

Chapter 5 investigates the sound field translation technique for binaural reproduction to recreate the virtual sense of spatial direction using headphones. We develop a new mixed-source expansion, which includes a mix of near-field and far-

field virtual sources, based on particle velocity. We represent the mixed sources using velocity coefficients and calculate the driving signals of the mixed source distribution by minimizing the difference between desired velocity coefficients and reconstructed velocity coefficients. We provide two solutions to this problem. One is the least square solution, which distributes energy throughout all virtual sources and inherits the spatial artifacts caused by the truncated measurement. Another is the sparse solution, which leads to better perceptual experience by modeling fewer virtual sources from propagation directions that are similar to the original sound. We introduce two localization metrics to evaluate the solutions. A MUSHRA experiment is also implemented to verify the performance of the proposed method on source localization and audio quality.

Chapter 6: Spatial Sound Field Reproduction Based on Sound Intensity
In this chapter, we explore the possibility of applying spatial sound intensity theory to the sound field reproduction system. We propose an intensity matching technique to optimally reproduce sound intensity over a continuous spatial region using an irregular loudspeaker array. Inspired by Kirchhoff-Helmholtz integral equation, we derive the loudspeaker driving signals and reproduce the sound intensity inside a spherical region by controlling the sound intensity distribution on the surface of the region. We compare the proposed method with the benchmark of mode matching method and HOA max-r_E decoding that optimizes sound intensity at a single position with two different irregular loudspeaker layouts (i.e., a 5 channel system and a 22 channel system) through numerical simulations. We also assess the impact of microphone noise on reproduction performance. The 5 channel system is then built for perceptual experiments.

Chapter 7: Intensity Based Sound Field Reproduction over Multiple Spatial Zones
This chapter offers two algorithms for intensity based multizone sound field reproduction. First, in Section 7.2, we reproduce the desired sound field by controlling the sound intensity at multiple sweet spots to recreate the original sound image over multiple different locations simultaneously. The sweet spots can be arbitrarily chosen within the target region. Then, in Section 7.3, we extend the method in Chapter

6 to multiple spatial zones, concurrently taking room acoustics into consideration. We integrate the sound intensity distributions of all the target zones to formulate the cost function. Both methods are evaluated through simulations/experiments with a spatially non-uniform loudspeaker arrangement.

Chapter 8: Conclusion and Future Work

This chapter provides a summary of the results drawn from this thesis and possible directions for future work.

Chapter 2

Background Theory: Spatial Sound Field Reproduction Using Spherical Harmonic Synthesis

This chapter provides a brief overview of the background knowledge concerning spatial sound field reproduction in the spherical harmonic domain. We first review the spherical harmonic expansion of a sound field. The properties of the expansion and the estimation of spherical harmonic coefficients are also discussed. We then review the conventional spatial sound synthesis method using the spherical harmonic expansion (i.e., the mode matching method) for the cases of a single spatial zone and multiple spatial zones.

2.1 Spherical Harmonic Decomposition of a Sound Field

Spherical harmonic decomposition is a powerful tool that can be used to describe the wave propagation over 3D space. The essential principle of the decomposition is to express the wave field as a weighted sum of a set of orthogonal basis functions (i.e., spherical harmonics), which are solutions of the Helmholtz wave equation in spherical coordinates. In acoustics, spherical harmonic decomposition is an important technique for sound field analysis and synthesis, and it has been used to

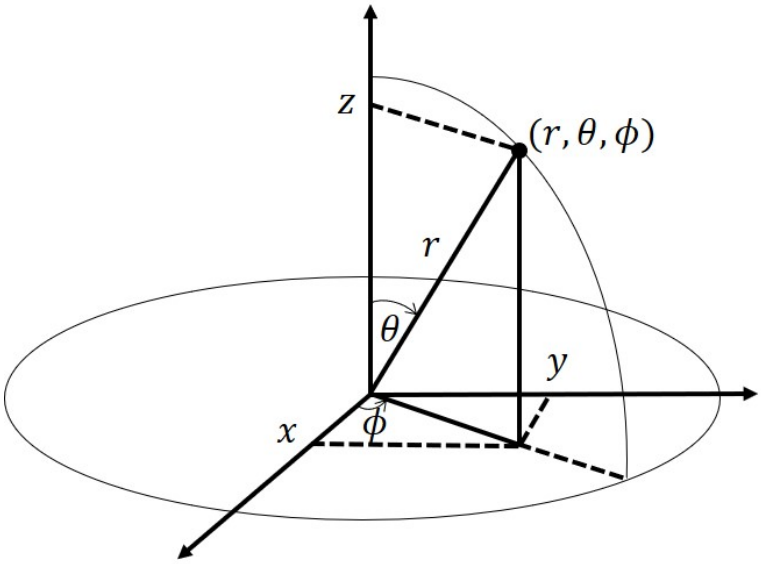


Figure 2.1: Definition of the spherical coordinate system throughout this thesis.

solve various acoustic problems such as spatial sound field reproduction [20–24, 28, 30], spatial active noise control [94–97], source localization [98–101], room acoustic modelling [102–104], beamforming [98, 105, 106], and dereverberation [107, 108]. Since spherical harmonic decomposition reveals the underlying characteristics of a sound field, it naturally provides more insight into the sound field and allows for more accurate analysis and synthesis of the sound field, compared to the structure of multiple sampling points over space. In this thesis, we make an extensive use of the spherical harmonic decomposition and develop the algorithms and models in the spherical harmonic domain.

We briefly introduce the theory and techniques related to spherical harmonic decomposition in this section.

2.1.1 Sound field expansion

A sound field is usually characterized by sound pressure. In this subsection, we give the representations of sound pressure in the spherical harmonic domain.

Coordinate system

The mathematical expressions of the representations vary for different definitions of the coordinate systems. Hence, we start with defining the spherical coordinate system used throughout this thesis, which is given in Fig. 2.1. The spherical coordinates of a point in space can be represented by (r, θ, ϕ), where radius r is the Euclidean distance from the origin to the point, elevation θ is the angle between the zenith direction and the line segment from the origin to the point, and azimuth ϕ is the angle between positive x and the projection of the line segment on x-y plane.

The spherical coordinates are related to the Cartesian coordinates through

$$x = r \sin\theta \cos\phi, \tag{2.1}$$

$$y = r \sin\theta \sin\phi, \tag{2.2}$$

$$x = r \cos\theta, \tag{2.3}$$

so that $r = \sqrt{x^2 + y^2 + z^2}$, $\theta = \tan^{-1}(\sqrt{x^2 + y^2}/z)$, and $\phi = \tan^{-1}(y/x)$. To define a unique set of coordinates in the spherical coordinate system, we impose the following constraints on the ranges of r, θ, and ϕ

$$r \in [0, \infty), \tag{2.4}$$

$$\theta \in [0, 180^\circ], \tag{2.5}$$

$$\phi \in [0, 360^\circ]. \tag{2.6}$$

Interior and exterior field solutions

Consider the propagation of sound waves in a homogeneous isotropic medium assumed to be an inviscid fluid, with stationary initial conditions. Sound pressure $p(\boldsymbol{x}, \omega)$ at an arbitrary point $\boldsymbol{x} = (r, \theta, \phi)$ follows the reduced wave equation or

Helmholtz equation [109]

$$\nabla^2 p(\boldsymbol{x}, \omega) + k^2 p(\boldsymbol{x}, \omega) = 0, \tag{2.7}$$

where ∇^2 denotes the Laplacian operator, expressed in spherical coordinates as

$$\nabla^2(\cdot) = \frac{1}{r^2}\frac{\partial}{\partial r}[r^2\frac{\partial}{\partial r}(\cdot)] + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}[\sin\theta\frac{\partial}{\partial \theta}(\cdot)] + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2}(\cdot), \tag{2.8}$$

and k denotes wavenumber, which is related to the angular frequency ω and wavelength λ by

$$k = \frac{\omega}{c} = \frac{2\pi f}{c} = \frac{2\pi}{\lambda}, \tag{2.9}$$

where f is the frequency and c denotes the speed of sound.

According to the relative positions of sound sources, the representations of a sound field can be divided into two cases: the interior field representation and the exterior field representation, both of which are the solutions to (2.7). The interior field is an incoming sound field caused by sources positioned completely outside the region of interest. The solution for the interior field is

$$p(\boldsymbol{x}, k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{2.10}$$

where $\alpha_{nm}(k)$ are the spherical harmonic coefficients of the interior field, $j_n(\cdot)$ is the nth order spherical Bessel function of the first kind, and $Y_{nm}(\theta, \phi)$ is the spherical harmonic of order n and degree m, defined by

$$Y_{nm}(\theta, \phi) \triangleq \underbrace{\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}}_{A_{nm}} P_{nm}(\cos\theta) e^{im\phi}, \tag{2.11}$$

where $P_{nm}(\cos\theta)$ is the associated Legendre function. Therefore, the sound field at an arbitrary point within a homogeneous interior sound field can be expressed by (2.10).

Conversely, the exterior field is an outgoing sound field caused by sources that are positioned within a limited enclosed area where the region of interest is the

space enclosing the source area. The solution for the exterior field is

$$p(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_{nm}(k) h_n(kr) Y_{nm}(\theta, \phi), \tag{2.12}$$

where $h_n(\cdot)$ is the nth order spherical Hankel function of the first kind. Similarly, (2.12) represents the sound field at any arbitrary point within a homogeneous exterior sound field.

Note that the sound field coefficients $\alpha_{nm}(k)$ and $\beta_{nm}(k)$ are independent of the observation point, therefore, they can be used to characterize the sound field over an entire continuous space for a given frequency k.

Spherical wave and plane wave models

There are two basic models of acoustic sources that are widely used when we study acoustic problems, and they are spherical wave model (also called point source model) and plane wave model. We here present the spherical harmonic expansions of the sound fields due to the two types of sources.

We first consider the spherical wave model. The sound field due to a source located at $\boldsymbol{x}_s = (r_s, \theta_s, \phi_s)$ can be described by the Green's function [110]

$$g(\boldsymbol{x}|\boldsymbol{x}_s, k) = \frac{e^{ik\|\boldsymbol{x}-\boldsymbol{x}_s\|_2}}{4\pi\|\boldsymbol{x}-\boldsymbol{x}_s\|_2}, \tag{2.13}$$

where $||\cdot||_2$ denotes the Euclidean norm. The expression (2.13) can be decomposed in terms of spherical harmonic functions as [111]

$$\frac{e^{ik\|\boldsymbol{x}-\boldsymbol{x}_s\|_2}}{4\pi\|\boldsymbol{x}-\boldsymbol{x}_s\|_2} = \begin{cases} ik \sum\limits_{n=0}^{\infty} \sum\limits_{m=-n}^{n} h_n(kr_s) Y_{nm}^*(\theta_s, \phi_s) j_n(kr) Y_{nm}(\theta, \phi), & r_s > r \\ ik \sum\limits_{n=0}^{\infty} \sum\limits_{m=-n}^{n} j_n(kr_s) Y_{nm}^*(\theta_s, \phi_s) h_n(kr) Y_{nm}(\theta, \phi), & r_s < r, \end{cases} \tag{2.14}$$

where $*$ denotes complex conjugate. We should note that (2.14) does not define the sound field at the source location $\boldsymbol{x}_s$, thus representing two separate homogeneous sound fields inside and outside the virtual sphere of radius r_s. The upper case of

(2.14) is an interior field, which complies with the interior field solution (2.10) with

$$\alpha_{nm}(k) = ikh_n(kr_s)Y^*_{nm}(\theta_s, \phi_s). \tag{2.15}$$

In contrast, the lower case is an exterior field, which satisfies the exterior field solution (2.12) with

$$\beta_{nm}(k) = ikj_n(kr_s)Y^*_{nm}(\theta_s, \phi_s). \tag{2.16}$$

Another model we discuss here is the plane wave model, which can be interpreted as point sources at infinity ($r_s \to \infty$) with appropriate normalization [112]. Substituting $r_s = \infty$ into (2.13) and performing the normalization, the sound field due to a plane wave yields

$$p(\boldsymbol{x}, k) = e^{-ik\widehat{\boldsymbol{y}}_i \cdot \boldsymbol{x}}, \tag{2.17}$$

where $\widehat{\boldsymbol{y}}_i = (\theta_i, \phi_i)$ is the incident direction of the plane wave. The spherical harmonic decomposition of (2.17) is [113]

$$e^{ik\widehat{\boldsymbol{y}}_i \dot{\boldsymbol{x}}_s} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi(-i)^n Y^*_{nm}(\theta_i, \phi_i) j_n(kr) Y_{nm}(\theta, \phi), \tag{2.18}$$

which also satisfies the interior field solution (2.10) with

$$\alpha_{nm}(k) = 4\pi(-i)^n Y^*_{nm}(\theta_i, \phi_i). \tag{2.19}$$

The spherical wave model and plane wave model can be used to describe the near-field propagation and far-field propagation, respectively. Therefore, (2.14) and (2.18) can accurately represent any arbitrary source-free sound field due to one or more sources lying in the near-field or far-field.

2.1.2 Properties of the expansion

In this thesis, we deal with various of properties of the expansion, and present them subsequently.

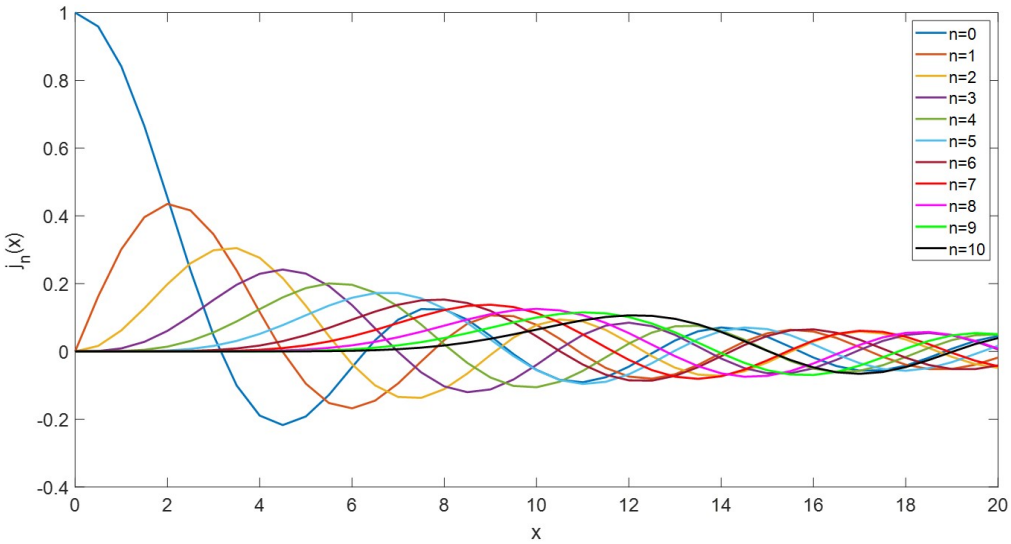


Figure 2.2: Plots of the spherical Bessel function $j_n(x)$ for n from 0 to 10.

Spherical Bessel functions

The spherical Bessel function can be expressed in terms of trigonometric functions

$$j_n(x) = (-x)^n(\frac{1}{x}\frac{d}{dx})^n(\frac{\sin x}{x}). \tag{2.20}$$

Figure 2.2 shows the plots of $j_n(x)$ for n from 0 to 10. From Fig. 2.2, it shows that the spherical Bessel functions $j_n(x)$ inherit a high pass behavior in terms of x (i.e., given a value of x, higher order $j_n(x)$ remains close to zero). In other words, for a given kr, $j_n(kr)$ is non-zero only up to an order of N, therefore, the infinite summation expressions that involve spherical Bessel functions can be appropriately truncated. A rule of thumb used to determine the truncation order is $N = \lceil kr \rceil$ with a error of around 4% [20], where $\lceil \cdot \rceil$ denotes rounding up to the nearest integer. An alternative rule is $N = \lceil ker/2 \rceil$ with improved accuracy [114].

According to this property of the spherical Bessel functions, the interior field of (2.10) can be truncated to $N_i = \lceil kR \rceil$ or $N_i = \lceil keR/2 \rceil$, assuming the target region is a spherical region with radius of R. For the exterior field of (2.12) there exist no spherical Bessel functions, however, the exterior fields can be given in terms of one or more points sources as mentioned in the lower case of (2.14), which includes the

spherical Bessel functions. Hence, the exterior field of (2.12) can also be truncated to $N_e = \lceil kR_s \rceil$ or $N_e = \lceil keR_s/2 \rceil$, where R_s denotes the radius of the minimum sphere enclosing all the sources.

There also exist a set of recurrence relations for spherical Bessel functions. In particular,

$$\frac{2n+1}{x} j_n(x) = j_{n-1}(x) + j_{n+1}(x), \tag{2.21}$$

and

$$j_n'(x) = j_{n-1}(x) - \frac{n+1}{x} j_n(x), \tag{2.22}$$

where $(\cdot)'$ denotes the corresponding first derivative term.

Spherical harmonics

Spherical harmonics are a set of orthonormal basis functions, therefore, we have the orthonormality of spherical harmonics

$$\int_0^{2\pi} d\phi \int_0^{\pi} Y_{nm}(\theta,\phi) Y_{n'm'}^*(\theta,\phi) \sin\theta d\theta = \delta_{nn'}\delta_{mm'}, \tag{2.23}$$

where $\delta_{nn'}$ is the Kronecker delta function, expressed as

$$\delta_{nn'} = \begin{cases} 1, & n = n' \\ 0, & n \neq n'. \end{cases} \tag{2.24}$$

Another important property for spherical harmonics is about the integral of product of three different spherical harmonics over a sphere [115]

$$\begin{aligned} &\int_0^{2\pi} d\phi \int_0^{\pi} Y_{n_1m_1}(\theta,\phi) Y_{n_2m_2}(\theta,\phi) Y_{n_3m_3}(\theta,\phi) \sin\theta d\theta \\ &= \sqrt{\frac{(2n_1+1)(2n_2+1)(2n_3+1)}{4\pi}} \begin{pmatrix} n_1 & n_2 & n_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} n_1 & n_2 & n_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \end{aligned} \tag{2.25}$$

where $\begin{pmatrix} \cdot \end{pmatrix}$ in (2.25) denotes Wigner-3j symbols [116]. In addition, the spherical harmonics have the following symmetry [109]

$$Y_{n(-m)}(\theta,\phi) = (-1)^m Y_{nm}^*(\theta,\phi). \tag{2.26}$$

The asscociated Legendre functions, one of the components of spherical harmonics, also have an important recurrence relation that we extensively use in this thesis

$$(1-x^2)\frac{dP_{nm}(x)}{dx} = (n+1)xP_{nm}(x) - (n-m+1)P_{(n+1)m}(x). \tag{2.27}$$

Addition theorems

The addition theorems are used to transform a sound field representation with respect to a given origin O in space into a similar representation with respect to a different origin O_1 in space. Here we introduce the addition theorems for the spherical Bessel functions and Hankel functions, which are related to this thesis.

Let $\boldsymbol{x}_0 = (r_0, \theta_0, \phi_0)$ and $\boldsymbol{x}_1 = (r_1, \theta_1, \phi_1)$ be the positions of a general point in space with respect to O, and O_1, respectively. Let $\boldsymbol{x}_2 = (r_2, \theta_2, \phi_2)$ be the position of O_1 with respect to O, so that $\boldsymbol{x}_0 = \boldsymbol{x}_1 + \boldsymbol{x}_2$. The addition theorem for spherical Bessel functions is then given by [117]

$$j_n(kr_0)Y_{nm}(\theta_0, \phi_0) = \sum_{v=0}^{\infty}\sum_{\mu=-v}^{v} \widehat{S}_{nv}^{m\mu}(k, \boldsymbol{x}_2) j_v(kr_1) Y_{v\mu}(\theta_1, \phi_1) \tag{2.28}$$

where v and μ denote higher order indices and

$$\widehat{S}_{nv}^{m\mu}(k, \boldsymbol{x}_2) = 4\pi i^{v-n} \sum_{l=0}^{n+v} i^l (-1)^{2m-\mu} j_l(kr_2) Y^*_{l(\mu-m)}(\theta_2, \phi_2) W_1 W_2 \xi \tag{2.29}$$

with

$$W_1 = \begin{pmatrix} n & v & l \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad W_2 = \begin{pmatrix} n & v & l \\ m & -\mu & (\mu - m) \end{pmatrix} \tag{2.30}$$

denoting Wigner-3j symbols and $\xi = \sqrt{(2n+1)(2v+1)(2l+1)/4\pi}$.

Similarly, the addition theorem for spherical Hankel functions is [117]

$$h_n(kr_0)Y_{nm}(\theta_0,\phi_0) = \begin{cases} \sum_{v=0}^{\infty}\sum_{\mu=-v}^{v} S_{nv}^{m\mu}(k,\boldsymbol{x}_2)j_v(kr_1)Y_{v\mu}(\theta_1,\phi_1) & r_1 \le r_2 \\ \\ \sum_{v=0}^{\infty}\sum_{\mu=-v}^{v} \widehat{S}_{nv}^{m\mu}(k,\boldsymbol{x}_2)h_v(kr_1)Y_{v\mu}(\theta_1,\phi_1) & r_1 \ge r_2 \end{cases} \tag{2.31}$$

where

$$S_{nv}^{m\mu}(k,\boldsymbol{x}_2) = 4\pi i^{v-n}\sum_{l=0}^{n+v} i^l(-1)^{2m-\mu}h_l(kr_2)Y_{l(\mu-m)}^*(\theta_2,\phi_2)W_1W_2\xi. \tag{2.32}$$

Note that $\widehat{S}_{nv}^{m\mu}(\cdot)$ and $S_{nv}^{m\mu}(\cdot)$ are also known as re-expansion operators, which denote translatory re-expansion of a sound field with respect to different coordinate systems.

2.1.3 Soundfield coefficient estimation from measurements

As mentioned earlier, a complete set of spherical harmonic coefficients can fully describe a sound field. To achieve a real-world sound field in such a way, it requires to extract the coefficients from microphone recordings. In this subsection, we briefly introduce the estimation of the coefficients using a spherical microphone array [76].

Multiplying both sides of (2.10) by $Y_{n'm'}^*(\theta,\phi)$ and integrating them over a sphere, we have

$$\begin{aligned} &\int_0^{2\pi}\int_0^{\pi} p(\boldsymbol{x},k)Y_{n'm'}^*(\theta,\phi)\sin\theta d\theta d\phi \\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\alpha_{nm}(k)j_n(kr)\int_0^{2\pi}\int_0^{\pi} Y_{nm}(\theta,\phi)Y_{n'm'}^*(\theta,\phi)\sin\theta d\theta d\phi. \end{aligned} \tag{2.33}$$

Due to the property of orthonormality (2.23), (2.33) can be reduced to

$$\alpha_{nm}(k) = \frac{1}{j_n(kr)}\int_0^{2\pi}\int_0^{\pi} p(\boldsymbol{x},k)Y_{nm}^*(\theta,\phi)\sin\theta d\theta d\phi. \tag{2.34}$$

To facilitate the real-world implementation, we approximate the integral in (2.34)

by a finite summation

$$\alpha_{nm}(k) = \frac{1}{j_n(kR_M)} \sum_{q=1}^{Q} \omega_q p(\boldsymbol{x}_q, k) Y_{nm}^*(\theta_q, \phi_q), \tag{2.35}$$

where Q is the number of microphones, R_M is the radius of the spherical microphone array, $p(\boldsymbol{x}_q, k)$ is the qth microphone measurement, and $\omega_q, q = 1, ..., Q$ are the corresponding microphone weights. Given a truncation order of N, it requires at least $(N+1)^2$ microphones on a sphere to avoid losing information, and their corresponding weights are determined by the spatial sampling scheme in order to ensure

$$\sum_{q=1}^{Q} \omega_q Y_{nm}(\theta_q, \phi_q) Y_{n'm'}^*(\theta_q, \phi_q) \approx \delta_{nn'} \delta_{mm'}. \tag{2.36}$$

An example of the spatial sampling scheme is given in [118].

We note that (2.35) is developed for a open spherical microphone array. For a rigid spherical array, $j_n(\cdot)$ in (2.35) need to be replaced with $b_n(\cdot)$, where

$$b_n(\cdot) = j_n(\cdot) - \frac{j_n'(\cdot)}{h_n'(\cdot)} h_n(\cdot). \tag{2.37}$$

2.2 Spherical Harmonic Synthesis of Mode Matching

The previous section introduces the spherical harmonic decomposition of a sound field, which can be used to describe the sound field over a continuous spatial region. Therefore, it has an advantage in solving the problem of spatial sound field reproduction. In this section, we review the spherical harmonic decomposition based spatial sound field reproduction technique of mode matching (also called pressure matching in the modal domain), including a solution to single zone reproduction and a solution to multizone reproduction. This technique is also the benchmark for the proposed reproduction techniques in the following chapters.

2.2.1 Single zone solution

We first briefly outline the concept of the mode matching method for a single continuous spatial region [20]. In the spherical harmonic domain, the sound pressure at any point $\boldsymbol{x}$ in a spherical region, due to a desired source outside the region, can be obtained from (2.10)

$$p_d(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{2.38}$$

where $\alpha_{nm}(k)$ denote the spherical harmonic coefficients of the desired sound field. The reproduced sound field is generated by a set of N_L loudspeakers located at $\boldsymbol{x}_\ell = (r_\ell, \theta_\ell, \phi_\ell)$ outside the spherical region, with $\ell = 1, ..., N_L$. We assume that the loudspeakers used for reproduction are point sources. According to the point source model of (2.14), the reproduced sound field due to the ℓth loudspeaker can be expressed as

$$g(\boldsymbol{x}|\boldsymbol{x}_\ell, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_{nm}^{(\ell)}(k) j_n(kr) Y_{nm}(\theta, \phi). \tag{2.39}$$

with

$$\beta_{nm}^{(\ell)}(k) = ikh_n(kr_\ell) Y_{nm}^*(\theta_\ell, \phi_\ell), \tag{2.40}$$

where $\beta_{nm}^{(\ell)}(k)$ denote the spherical harmonic coefficients of the sound field due to the ℓth loudspeaker. The reproduced sound pressure due to the loudspeaker array can be written as

$$p_a(\boldsymbol{x}, k) = \sum_{\ell=0}^{N_L} w_\ell(k) g(\boldsymbol{x}|\boldsymbol{x}_\ell, k), \tag{2.41}$$

where $w_\ell(k)$ is the driving signal applied to the ℓth loudspeaker.

Substituting (2.39) into (2.41), and then equating (2.41) with (2.38), we have

$$\sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{\ell=0}^{N_L} w_\ell(k) \beta_{nm}^{(\ell)}(k) j_n(kr) Y_{nm}(\theta, \phi). \tag{2.42}$$

Multiplying both sides in (2.42) by $Y_{n'm'}^*(\theta, \phi)$ and integrating them over the unit

sphere with respect to θ and ϕ, gives

$$\alpha_{nm}(k) = \sum_{\ell=0}^{N_L} w_\ell(k)\beta_{nm}^{(\ell)}(k). \tag{2.43}$$

Rewriting (2.43) in matrix form as

$$\boldsymbol{\alpha} = \boldsymbol{\beta}\boldsymbol{G}, \tag{2.44}$$

where $\boldsymbol{\alpha} = [\alpha_{00}, \alpha_{1-1}, ..., \alpha_{NN}]^T$, $\boldsymbol{G} = [w_1, ..., w_{N_L}]^T$, and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{00}^{(1)} & \cdots & \beta_{00}^{(N_L)} \\ \beta_{1-1}^{(1)} & \cdots & \beta_{1-1}^{(N_L)} \\ \vdots & \ddots & \vdots \\ \beta_{NN}^{(1)} & \cdots & \beta_{NN}^{(N_L)} \end{bmatrix}. \tag{2.45}$$

For notations convenience, we omit the dependence on k. The number of loudspeakers N_L specifies how the linear system (2.44) can be solved. There are three cases of interest. When the system is over-determined (i.e.,$(N+1)^2 > N_L$), there will be no exact solution in general. It becomes a least square problem

$$\min_{\boldsymbol{G}} ||\boldsymbol{\beta}\boldsymbol{G} - \boldsymbol{\alpha}||_2^2. \tag{2.46}$$

If $\boldsymbol{\beta}$ is a square non-singular matrix, a unique solution to (2.44) is given by $\boldsymbol{G} = \boldsymbol{\beta}^{-1}\boldsymbol{\alpha}$. Finally, for an under-determined system (i.e.,$(N+1)^2 < N_L$), there may either be no solution or an infinite number of solutions. For the latter case, we can find the weights to satisfy

$$\min_{\boldsymbol{G}} ||\boldsymbol{G}||_2^2 \ \text{ s.t. } \ \boldsymbol{\alpha} = \boldsymbol{\beta}\boldsymbol{G}. \tag{2.47}$$

Tikhonov regularization can be introduced here to improve the robustness of the solution. The solutions of both the problems (2.46) and (2.47) are well studied in [35, 47, 119].

2.2.2 Multizone solution

The essential idea of the multizone solution is to derive an equivalent global sound field that consists of a number of individual multizone sound fields by the addition theorems [25]. Assume that there are B non-overlapping spatial zones, and the origin of the bth zone is denoted as $\mathcal{O}_b$ and located at $\boldsymbol{x}_o^{(b)}$ with respect to the global origin $\mathcal{O}$. The desired sound field of the bth zone can be described by spherical harmonic coefficients $\alpha_{nm}^{(b)}(k)$, such that the sound pressure at an arbitrary observation point $\boldsymbol{x}_b = (r_b, \theta_b, \phi_b)$ with respect to $\mathcal{O}_b$ within the bth zone can be expressed as

$$p^{(b)}(\boldsymbol{x}_b, k) = \sum_{n=0}^{N_b} \sum_{m=-n}^{n} \alpha_{nm}^{(b)}(k) j_n(kr_b) Y_{nm}(\theta_b, \phi_b), \tag{2.48}$$

where $N_b = \lceil keR_b \rceil$, and R_b is the radius of the bth zone.

The sound pressure at the same point with respect to $\mathcal{O}$ (i.e., $\boldsymbol{x}_{bo} = \boldsymbol{x}_o^{(b)} + \boldsymbol{x}_b = (r_{bo}, \theta_{bo}, \phi_{bo})$) can be written as

$$p^{(g)}(\boldsymbol{x}_{bo}, k) = \sum_{n=0}^{N_g} \sum_{m=-n}^{n} \beta_{nm}^{(g)}(k) j_n(kr_{bo}) Y_{nm}(\theta_{bo}, \phi_{bo}), \tag{2.49}$$

where $N_g = \lceil keR_g \rceil$, R_g is the radius of the smallest sphere that encloses all spatial zones of interest, and $\beta_{nm}^{(g)}(k)$ are the global pressure coefficients. The coefficients $\alpha_{nm}^{(b)}(k)$ and $\beta_{nm}^{(g)}(k)$ are related by the addition theorem of (2.28)

$$\alpha_{v\mu}^{(b)}(k) = \sum_{v=0}^{\infty} \sum_{\mu=-v}^{v} \widehat{S}_{nv}^{m\mu}(\boldsymbol{x}_o^{(b)}) \beta_{nm}^{(g)}(k). \tag{2.50}$$

Our objective is to achieve the global pressure coefficients that can reproduce all sound field zones. We write (2.50) for $b = 1, ..., B$ and express them in matrix form as

$$\boldsymbol{A} = \widehat{\boldsymbol{S}}\boldsymbol{C}, \tag{2.51}$$

where $\boldsymbol{C}$=$[\beta_{00}^{(g)}, \beta_{1-1}^{(g)}, ..., \beta_{N_g N_g}^{(g)}]^T$, $\boldsymbol{A}$=$[\alpha_{00}^{(1)}, \alpha_{1-1}^{(1)}, ..., \alpha_{N_1 N_1}^{(1)}, ..., \alpha_{00}^{(B)}, \alpha_{1-1}^{(B)}, ..., \alpha_{N_B N_B}^{(B)}]^T$,

and

$$\widehat{\boldsymbol{S}} = \begin{bmatrix} \widehat{S}_{00}^{00}(\boldsymbol{x}_o^{(1)}) & \widehat{S}_{10}^{-10}(\boldsymbol{x}_o^{(1)}) & \cdots & \widehat{S}_{N_g0}^{N_g0}(\boldsymbol{x}_o^{(1)}) \\ \widehat{S}_{01}^{0-1}(\boldsymbol{x}_o^{(1)}) & \widehat{S}_{11}^{-1-1}(\boldsymbol{x}_o^{(1)}) & \cdots & \widehat{S}_{N_g1}^{N_g-1}(\boldsymbol{x}_o^{(1)}) \\ \vdots & \ddots & \vdots & \\ \widehat{S}_{0N_1}^{0N_1}(\boldsymbol{x}_o^{(1)}) & \widehat{S}_{1N_1}^{-1N_1}(\boldsymbol{x}_o^{(1)}) & \cdots & \widehat{S}_{N_gN_1}^{N_gN_1}(\boldsymbol{x}_o^{(1)}) \\ \vdots & \ddots & \vdots & \\ \widehat{S}_{00}^{00}(\boldsymbol{x}_o^{(B)}) & \widehat{S}_{10}^{-10}(\boldsymbol{x}_o^{(B)}) & \cdots & \widehat{S}_{N_g0}^{N_g0}(\boldsymbol{x}_o^{(B)}) \\ \widehat{S}_{01}^{0-1}(\boldsymbol{x}_o^{(B)}) & \widehat{S}_{11}^{-1-1}(\boldsymbol{x}_o^{(B)}) & \cdots & \widehat{S}_{N_g1}^{N_g-1}(\boldsymbol{x}_o^{(B)}) \\ \vdots & \ddots & \vdots & \\ \widehat{S}_{0N_B}^{0N_B}(\boldsymbol{x}_o^{(B)}) & \widehat{S}_{1N_B}^{-1N_B}(\boldsymbol{x}_o^{(B)}) & \cdots & \widehat{S}_{N_gN_B}^{N_gN_B}(\boldsymbol{x}_o^{(B)}) \end{bmatrix}. \tag{2.52}$$

The global sound field coefficients can be obtained by solving (2.51) using the least squares method

$$\boldsymbol{C} = \widehat{\boldsymbol{S}}^{\dagger}\boldsymbol{A}, \tag{2.53}$$

where $\widehat{\boldsymbol{S}}^{\dagger} = [\widehat{\boldsymbol{S}}^H\widehat{\boldsymbol{S}}]^{-1}\widehat{\boldsymbol{S}}^H$ is the Moore–Penrose (Pseudo) inverse of $\widehat{\boldsymbol{S}}$. Note that the (2.53) involves a matrix inversion of $\widehat{\boldsymbol{S}}^H\widehat{\boldsymbol{S}}$, which is dependent of the zone positions, therefore, each zone needs to be carefully positioned to avoid the poorly conditioned inverse problem [25].

Once we obtain the global sound field coefficients, the multizone reproduction is reduced to single zone reproduction in Section 2.2.1.

2.3 Summary

This chapter offered an introduction on the theory of spherical harmonic analysis that the rest of thesis is based on, including sound field expansion, important properties related to the expansion, and spherical harmonic coefficients estimation. The spherical harmonic synthesis of sound fields was also discussed, where the conventional mode matching method was outlined for both single zone reproduction and multizone reproduction. This method is compared to the proposed methods in the Chapters 4, 5, and 7 as the benchmark.

Chapter 3

Theory of Spatial Acoustic Vectors

In this chapter, we introduce the theory of spatial acoustic vectors, which is the foundation of the spatial sound field reproduction systems in the following chapters of this thesis. The theory is derived in terms of the pressure distribution of a sound field. In most cases, the sound field radiated by a desired sound source (exterior or outgoing sound field) may be mixed with the reverberations and the sound fields incident from other sources that may be present (interior or incident sound field). To achieve the spatial pressure distribution radiated by the desired sound source, we first propose a 3D sound field separation method that can extract the desired outgoing sound field in a mixed sound field environment using a planar microphone array that is more practical than the traditional spherical array. The desired sound field is given by sound pressure in spherical harmonic domain. We then derive the closed form representations of the spatial acoustic vectors, including spatial particle velocity vector and spatial sound intensity vector, from sound pressure in spherical harmonic domain. We also give the truncated expressions of the vectors for practical implementations.

3.1 Desired Sound Field Extraction

An accurate capture of desired sound field is the prerequisite for a sound field reproduction system. In this section, we introduce the desired sound field extraction from a mixed sound field.

3.1.1 Introduction

In a complex 3D acoustic environment, there usually exist multiple sources and even reverberations. The mixed sound field can be divided into interior sound field and exterior sound field [109] on the basis of the location of receivers. It is a challenging task to separate and extract the desired sound field from a mixed measurement in such an environment. However, in nature, the ability to separate these sound fields will benefit a plethora of applications including noise cancellation [120], selective sound field reproduction, room equalization and scattering measurement/correction. The representation of the interior or exterior sound field has been studied in literature [76, 121]. Efren *et al.* proposed a separation method based on the measurement of particle velocity in two layers or the pressure and the velocity in a single layer [122, 123]. The method introduced in [124], is called equivalent source method which separates wave components by measuring pressure on arbitrarily shaped surfaces in the spatial domain. All these works do not involve whole 3D sound field separation. Weinreich *et al.* determined the interior and exterior components independently using a spherical measurement array with a single rotating microphone [125]. In [126], authors introduced a method to extract the exterior field from a mixed sound field for 2D height-invariant sound propagation and then they extended the method to 3D sound field by using a sparse array of higher order spherical microphones [127].

Array geometry is an important factor when we consider sound field recording. Spherical arrays were extensively studied in literature [20,28,29]. Gupta *et al.* used a set of multiple circular arrays to reproduce 3D sound fields [128], however the arrays were placed on different planes. Chen *et al.* developed a method to capture 3D spatial sound field with 2D planar microphone arrays [78] with two possible configurations of arrays. One was to use a hybrid of omni-directional microphones

and first order microphones and another was to use omni-directional pairs. The first order microphones were utilized to capture 3D spatial exterior sound field in [129]. All these geometries can help to record sound fields. However, none of these arrays address the problem of sound field separation.

In the following subsections, we propose a solution to separate 3D interior and exterior sound field using a planar array of differential microphones, which can be used to extract the desired sound field for spatial sound field reproduction systems. We then demonstrate the algorithm exhibits a good performance through simulations.

3.1.2 Problem formulation

We consider a region χ between two concentric spherical surfaces with radii r_{i} and r_{e}, where $r_{\mathrm{e}} < r_{\mathrm{i}}$. We choose the origin in such a way that all sound sources are located inside the sphere of radius r_{e} or outside the sphere of radius r_{i}. In this work, we assume that the desired sound sources are located inside the sphere of radius r_{e} and the other sources are located outside the sphere of radius r_{i}. According to (2.10) and (2.12), the 3D sound field at any point $\boldsymbol{x}$ in χ due to the sources outside the sphere of radius r_{i} is given by

$$P_{\mathrm{i}}(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{3.1}$$

and the sound field at the same point $\boldsymbol{x}$ due to the sources inside the sphere of radius r_{e} is given by

$$P_{\mathrm{e}}(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_{nm}(k) h_n(kr) Y_{nm}(\theta, \phi), \tag{3.2}$$

where P_{i} and P_{e} are known as the interior and exterior sound fields, respectively, and

$$Y_{nm}(\theta, \phi) = \mathcal{P}_{n|m|}(\cos\theta) E_m(\phi), \tag{3.3}$$

where $\mathcal{P}_{n|m|}(\cos\theta) = [(2n+1)(n-|m|)!/2\pi(n+|m|)!]^{1/2} P_{n|m|}(\cos\theta) e^{im\phi}$ and $E_m(\phi) = (1/(2\pi)^{1/2}) e^{im\phi}$ are the normalized associated Legendre functions and normalized

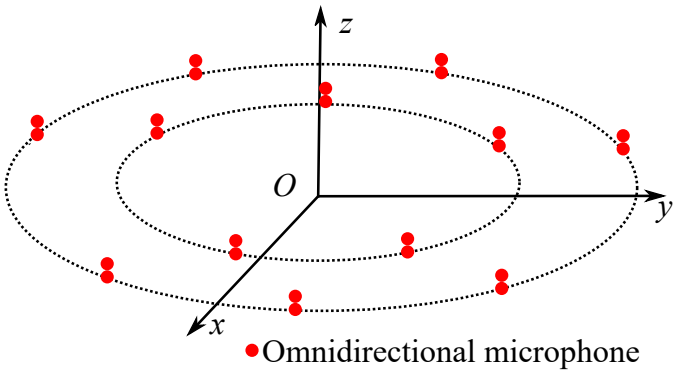


Figure 3.1: Example of planar arrays of differential microphones

exponential functions respectively.

The total sound field at $\boldsymbol{x}$ in χ would follow the superposition principle

$$P_{\mathrm{t}}(\boldsymbol{x}, k) = P_{\mathrm{i}}(\boldsymbol{x}, k) + P_{\mathrm{e}}(\boldsymbol{x}, k). \tag{3.4}$$

Given the measured total sound pressure $P_{\mathrm{t}}(\boldsymbol{x}, k)$, our objective is to obtain the interior sound field $P_{\mathrm{i}}(\boldsymbol{x}, k)$ and exterior sound field $P_{\mathrm{e}}(\boldsymbol{x}, k)$ separated.

3.1.3 Sound field recording using a planar array of differential microphones.

Array geometry for a mixed sound field

The planar array of differential microphones is placed on the x-y plane and consists of circles of closely placed omni-directional microphone pairs (see Fig. 3.1). The desired sources should be surrounded by the array, and the other sources should be located outside the array. Each microphone pair is used in two different ways: one of the omni-directional microphones in the pair is used to calculate even[1] coefficients, while the difference between the output signals of two omni-directional microphones placed perpendicular to the x-y plane are used to calculate odd coefficients [78]. The number of microphones on the qth circular array is $N_q = 2\lceil keR_q/2 \rceil + 1$, which depends on the radius of array R_q and the wave number k (see [78] for details).

[1]Even and odd coefficients are defined when the value of $n + |m|$ is even and odd, respectively.

Output representations of the array

Consider an arbitrary point on the circle in Fig. 3.1 given by $(R_q, \pi/2, \phi)$, and the output of an omni-directional microphone at this point is given by

$$\begin{aligned} P_{\mathrm{t}}(R_q, \pi/2, \phi, k) = &\sum_{n=0}^{N_{\mathrm{i}}} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kR_q) \mathcal{P}_{n|m|}(0) E_m(\phi) \\ &+ \sum_{n=0}^{N_{\mathrm{e}}} \sum_{m=-n}^{n} \beta_{nm}(k) h_n(kR_q) \mathcal{P}_{n|m|}(0) E_m(\phi), \end{aligned} \tag{3.5}$$

where $N_{\mathrm{i}} = \lceil ker_{\mathrm{i}}/2 \rceil$ and $N_{\mathrm{e}} = \lceil ker_{\mathrm{e}}/2 \rceil$ are truncation limits of the sound field orders [103, 114].

Then the output of one microphone pair, which serves as a differential microphone, on the circle at $(R_q, \pi/2, \phi)$ is given by

$$\begin{aligned} P_{\mathrm{t}}^{(\mathrm{d})}(R_q, \pi/2, \phi, k) &= \left. \frac{\partial P_{\mathrm{t}}(r, \theta, \phi, k)}{\partial \theta} \right|_{r=R_q, \theta=\frac{\pi}{2}} \\ &= -\sum_{n=0}^{N_{\mathrm{i}}} \sum_{m=-n}^{n} d_{nm} \alpha_{nm}(k) j_n(kR_q) \mathcal{P}_{(n-1)|m|}(0) E_m(\phi) \\ &\quad - \sum_{n=0}^{N_{\mathrm{e}}} \sum_{m=-n}^{n} d_{nm} \beta_{nm}(k) h_n(kR_q) \mathcal{P}_{(n-1)|m|}(0) E_m(\phi), \end{aligned} \tag{3.6}$$

where $d_{nm} = [(2n+1)(n^2 - m^2)/(2n-1)]^{1/2}$, and the superscript 'd' here means differential.

Sound field recording

Due to the orthogonality of exponential functions, multiplying both sides of (3.5) by $E_m(-\phi)$ and integrating with respect to ϕ, we can get

$$\begin{aligned} X_m(R_q, k) &= \int_0^{2\pi} P_{\mathrm{t}}(R_q, \pi/2, \phi, k) E_m(-\phi) d\phi \\ &= \sum_{n=|m|}^{N_{\mathrm{i}}} \alpha_{nm}(k) j_n(kR_q) \mathcal{P}_{n|m|}(0) + \sum_{n=|m|}^{N_{\mathrm{e}}} \beta_{nm}(k) h_n(kR_q) \mathcal{P}_{n|m|}(0). \end{aligned} \tag{3.7}$$

Note that (3.7) only includes the even mode harmonics, because $\mathcal{P}_{n|m|}(0) = 0$ for odd values of $n + |m|$ [130].

Similarly, by multiplying both sides of (3.6) by $E_m(-\phi)$ and integrating with respect to ϕ, we obtain

$$\begin{aligned} X_m{}^{(\mathrm{d})}(R_q, k) &= \int_0^{2\pi} P_{\mathrm{t}}^{(\mathrm{d})}(R_q, \pi/2, \phi, k) E_m(-\phi) d\phi \\ &= -\sum_{n=|m|}^{N_{\mathrm{i}}} d_{nm}\alpha_{nm}(k) j_n(kR_q)\mathcal{P}_{(n-1)|m|}(0) - \sum_{n=|m|}^{N_{\mathrm{e}}} d_{nm}\beta_{nm}(k) h_n(kR_q)\mathcal{P}_{(n-1)|m|}(0). \end{aligned} \tag{3.8}$$

Note that only the odd mode harmonics are present in (3.8), since $\mathcal{P}_{(n-1)|m|}(0) = 0$ for even values of $n + |m|$ [130].

However, in practice, we only have a finite set of microphone pairs and thus we approximate the integration in (3.7) and (3.8) by summations

$$X_m(R_q, k) \approx \frac{2\pi}{N_q} \sum_{s=1}^{N_q} P_{\mathrm{t}}(R_q, \pi/2, \phi_s, k) E_m(-\phi_s), \tag{3.9}$$

$$X_m^{(\mathrm{d})}(R_q, k) \approx \frac{2\pi}{N_q} \sum_{s=1}^{N_q} P_{\mathrm{t}}^{(\mathrm{d})}(R_q, \pi/2, \phi_s, k) E_m(-\phi_s), \tag{3.10}$$

where ϕ_s is the azimuth angle of the location of the sth microphone on the qth array.

3.1.4 Sound field separation

Even coefficients separation

Let there be a total of Q circles placed at different radii but all on the x-y plane. Rewriting (3.7) in matrix form

$$\boldsymbol{X_m(k)} = \boldsymbol{C_m^{(\mathrm{e})}(k) U_m(k)}, \tag{3.11}$$

where $\boldsymbol{X_m(k)} = [X_m(R_1,k), X_m(R_2,k), ..., X_m(R_Q,k)]$, $\boldsymbol{C_m^{(e)}(k)} = [\boldsymbol{\alpha_m^{(e)}}\boldsymbol{\beta_m^{(e)}}]$ with $\boldsymbol{\alpha_m^{(e)}} =$

$$\begin{cases} [\alpha_{mm}, \alpha_{(m+2)m}, ..., \alpha_{N_\mathrm{i}m}], & \text{if } (m+N_\mathrm{i}) \text{ is even} \\ [\alpha_{mm}, \alpha_{(m+2)m}, ..., \alpha_{(N_\mathrm{i}-1)m}], & \text{otherwise,} \end{cases} \tag{3.12}$$

and $\boldsymbol{\beta_m^{(e)}} =$

$$\begin{cases} [\beta_{mm}, \beta_{(m+2)m}, ..., \beta_{N_\mathrm{e}m}], & \text{if } (m+N_\mathrm{e}) \text{ is even} \\ [\beta_{mm}, \beta_{(m+2)m}, ..., \beta_{(N_\mathrm{e}-1)m}], & \text{otherwise,} \end{cases} \tag{3.13}$$

and

$$\boldsymbol{U_m(k)} = \begin{bmatrix} U^{(\mathrm{j})}_{m|m|}(kR_1) & \cdots & U^{(\mathrm{j})}_{m|m|}(kR_Q) \\ \vdots & \ddots & \vdots \\ U^{(\mathrm{j})}_{N_\mathrm{i}|m|}(kR_1) & \cdots & U^{(\mathrm{j})}_{N_\mathrm{i}|m|}(kR_Q) \\ U^{(\mathrm{h})}_{m|m|}(kR_1) & \cdots & U^{(\mathrm{h})}_{m|m|}(kR_Q) \\ \vdots & \ddots & \vdots \\ U^{(\mathrm{h})}_{N_\mathrm{e}|m|}(kR_1) & \cdots & U^{(\mathrm{h})}_{N_\mathrm{e}|m|}(kR_Q) \end{bmatrix} \tag{3.14}$$

with $U^{(\mathrm{j})}_{n|m|}(kR_q) = j_n(kR_q)\mathcal{P}_{n|m|}(0)$ and $U^{(\mathrm{h})}_{n|m|}(kR_q) = h_n(kR_q)\mathcal{P}_{n|m|}(0)$ for the case when both $N_\mathrm{i}+m$ and $N_\mathrm{e}+m$ are even. If $N_\mathrm{i}+m$ is odd, we replace N_i in (3.14) by $N_\mathrm{i}-1$, and similarly we replace N_e in (3.14) by $N_\mathrm{e}-1$, if $N_\mathrm{e}+m$ is odd. We can estimate the even coefficients of interior and exterior sound field from (3.11), as

$$\boldsymbol{C_m^{(e)}(k)} = \boldsymbol{X_m(k)}\boldsymbol{U_m^{\dagger}(k)}, \tag{3.15}$$

where $\boldsymbol{U_m^{\dagger}}$ is the pseudo inverse of $\boldsymbol{U_m}$.

Odd coefficients separation

Similarly, rewriting (3.8) in the matrix form

$$\boldsymbol{X_m^{(d)}(k)} = \boldsymbol{C_m^{(o)}(k)}\boldsymbol{V_m(k)}, \tag{3.16}$$

where $\boldsymbol{X}_{\boldsymbol{m}}^{(\mathrm{d})}(\boldsymbol{k}) = [X_m^{(\mathrm{d})}(R_1, k), X_m^{(\mathrm{d})}(R_2, k), ..., X_m^{(\mathrm{d})}(R_Q,\ k)]$, $\boldsymbol{C}_{\boldsymbol{m}}^{(\mathrm{o})}(\boldsymbol{k}) = [\boldsymbol{\alpha}_{\boldsymbol{m}}^{(\mathrm{o})}\ \boldsymbol{\beta}_{\boldsymbol{m}}^{(\mathrm{o})}]$ with $\boldsymbol{\alpha}_{\boldsymbol{m}}^{(\mathrm{o})} =$

$$\begin{cases} [\alpha_{(m+1)m}, \alpha_{(m+3)m}, ..., \alpha_{(N_\mathrm{i}-1)m}], & \text{if } (m+N_\mathrm{i}) \text{ is even} \\ [\alpha_{(m+1)m}, \alpha_{(m+3)m}, ..., \alpha_{N_\mathrm{i}m}], & \text{otherwise,} \end{cases} \tag{3.17}$$

and $\boldsymbol{\beta}_{\boldsymbol{m}}^{(\mathrm{o})} =$

$$\begin{cases} [\beta_{(m+1)m}, \beta_{(m+3)m}, ..., \beta_{(N_\mathrm{e}-1)m}], & \text{if } (m+N_\mathrm{e}) \text{ is even} \\ [\beta_{(m+1)m}, \beta_{(m+3)m}, ..., \beta_{N_\mathrm{e}m}], & \text{otherwise,} \end{cases} \tag{3.18}$$

and

$$\boldsymbol{V}_{\boldsymbol{m}}(\boldsymbol{k}) = \begin{bmatrix} V_{(m+1)|m|}^{(\mathrm{j})}(kR_1) & \cdots & V_{(m+1)|m|}^{(\mathrm{j})}(kR_Q) \\ \vdots & \ddots & \vdots \\ V_{(N_\mathrm{i}-1)|m|}^{(\mathrm{j})}(kR_1) & \cdots & V_{(N_\mathrm{i}-1)|m|}^{(\mathrm{j})}(kR_Q) \\ V_{(m+1)|m|}^{(\mathrm{h})}(kR_1) & \cdots & V_{(m+1)|m|}^{(\mathrm{h})}(kR_Q) \\ \vdots & \ddots & \vdots \\ V_{(N_\mathrm{e}-1)|m|}^{(\mathrm{h})}(kR_1) & \cdots & V_{(N_\mathrm{e}-1)|m|}^{(\mathrm{h})}(kR_Q) \end{bmatrix} \tag{3.19}$$

with

$V_{n|m|}^{(\mathrm{j})}(kR_q) = -d_{nm} j_n(kR_q)\mathcal{P}_{(n-1)|m|}(0)$ and $V_{n|m|}^{(\mathrm{h})}(kR_q) = -d_{nm} h_n(kR_q)\mathcal{P}_{(n-1)|m|}(0)$ for the case when both $N_\mathrm{i}+m$ and $N_\mathrm{e}+m$ are even. If $N_\mathrm{i}+m$ is odd, we replace $N_\mathrm{i}-1$ in (3.19) by N_i, and similarly we replace $N_\mathrm{e}-1$ in (3.19) by N_e, if $N_\mathrm{e}+m$ is odd. We can estimate the odd coefficients of interior and exterior sound field from (3.16), as

$$\boldsymbol{C}_{\boldsymbol{m}}^{(\mathrm{o})}(\boldsymbol{k}) = \boldsymbol{X}_{\boldsymbol{m}}^{(\mathrm{d})}(\boldsymbol{k})\boldsymbol{V}_{\boldsymbol{m}}^{\dagger}(\boldsymbol{k}), \tag{3.20}$$

where $\boldsymbol{U}_{\boldsymbol{m}}^{\dagger}$ is the pseudo inverse of $\boldsymbol{U}_{\boldsymbol{m}}$. Note that, to avoid an under-determined system of (3.14) and (3.19) , no less than $2\lceil N_\mathrm{i}/2\rceil$ circles of microphone pairs are required.

We also note that there are zeros in the spherical Bessel functions at some radius and frequency values, which leads to the loss of the information for certain modes and the ill-conditioning for matrix inversion. Therefore, the radius of each circular array should be carefully determined such that the spherical Bessel zeros

for the target frequency band are avoided [78].

Thus the complete set of interior and exterior sound field coefficients are derived and separated by solving (3.15) and (3.20). Then we can use the coefficients to reconstruct corresponding sound field.

3.1.5 Simulation results

Sound field separation in a reverberant room

We consider the case separating the desired sound field and undesired sound field in a reverberant room with $(4\times5\times3)$ m dimensions. We used the image source method [131] to simulate reverberant conditions. The reflection coefficient of each wall is $(0.9, 0.8, 0.7, 0.5, 0.7, 0.9)$ and the number of image depths is 2 in this simulation. The point source producing the desired sound field is located at $(0.01, 0.01, 0.01)$ m and there is another undesired point source located at $(-1, 1, 1)$ m, both with $f = 850$ Hz, as the origin is located at the centre of the room. In this case, we define the region χ is between the sphere with radii $r_{\mathrm{i}} = 0.12$ m and the sphere with radii $r_{\mathrm{e}} = 0.06$ m. Sound propagation speed $c = 340$ m/s, thus the maximum order of the sound field is $N_{\mathrm{i}} = 3$. It requires at least 4 circular arrays to obtain the interior (undesired) sound field and exterior (desired) sound field separated. The radii of the arrays are set to be 0.12, 0.1, 0.08, and 0.06 m. Thus the numbers of microphone pairs on each array are 7, 7, 5, and 5, respectively. In this simulation, we record and separate the interior and exterior sound fields, and then reconstruct them on a spherical surface with radius of 0.08 m. Fig. 3.2(a) and Fig. 3.2(b) show the original interior and exterior sound field, respectively. Figure 3.2(c) and Fig. 3.2(d) are the reconstructed interior and exterior sound field, respectively. To evaluate the robustness against microphone thermal noise, we add a White Gaussian Noise (WGN) with signal to noise ratio (SNR) of 40 dB at each microphone. The reconstructed results are shown in Fig. 3.2(e) and Fig. 3.2(f). We observe that both reconstructed interior and exterior sound fields are similar to the original ones, with or without noise. Note that the closer the desired point source is to the array, the worse the performance achieves.

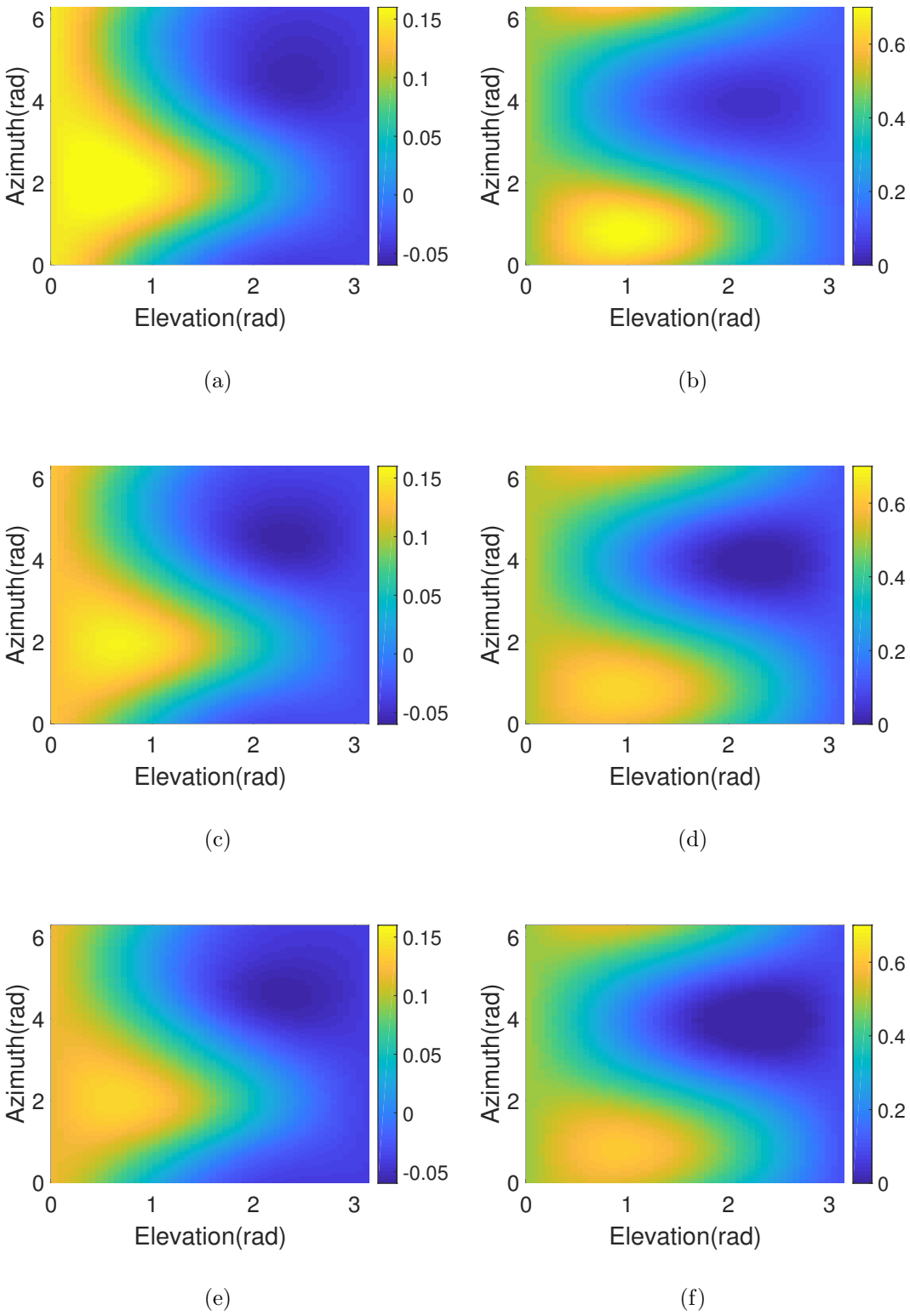


Figure 3.2: Interior and exterior sound field reconstruction on a spherical surface with radius of 0.08 m. The frequency of the sources is 850 Hz. (a) Original interior sound field, (b) original exterior sound field, (c) reconstructed interior sound field without noise, (d) reconstructed exterior sound field without noise, (e) reconstructed interior sound field with SNR of 40 dB, (f) reconstructed exterior sound field with SNR of 40 dB.

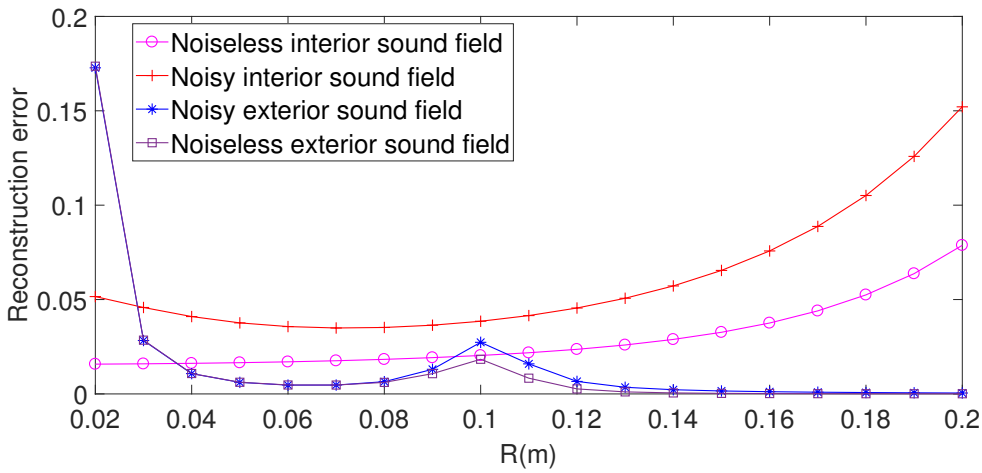


Figure 3.3: Reconstruction error of interior and exterior sound field against different radii of spherical surfaces. The frequency of the sources is 850 Hz.

Reconstruction accuracy

To examine the reconstruction accuracy of interior and exterior sound field on different spherical surfaces, we calculate the reconstruction error choosing radius of the surface from 0.02 m to 0.2 m with SNR of 100 dB (noiseless) and 40 dB, respectively. The reconstruction error is measured by

$$\epsilon(k) = \frac{\sum_{\forall \boldsymbol{x}} \left| P(\boldsymbol{x}, k) - \widehat{P}(\boldsymbol{x}, k) \right|^2}{\sum_{\forall \boldsymbol{x}} \left| P(\boldsymbol{x}, k) \right|^2} \tag{3.21}$$

where $\widehat{P}(\boldsymbol{x}, k)$ is reconstructed sound pressure at point $\boldsymbol{x}$ on the spherical surface. Fig. 3.3 depicts the error for different radii. Observe that the error of interior sound field is no more than 0.06 when the radius is below 0.12 m. Beyond this value, the error increases because the order of active spherical harmonics also increases. Differently, the error of exterior sound field is large on the sphere close to the source and the error becomes less than 0.04 as the radius increases. Note that there is a peak at $R = 0.1$ m, which is caused by truncation error of exterior sound field.

3.2 Spatial Particle Velocity Vector

In the previous section, we extracted the desired sound field, described by spherical harmonic coefficients of sound pressure, from a mixed sound field using a practical microphone array. To achieve the desired particle velocity distribution over space from the desire sound pressure, we next are going to derive the representations of spatial particle velocity vector from spherical harmonic coefficients of sound pressure in this section.

3.2.1 Introduction

Particle velocity is one of fundamental acoustic quantities that can be widely used in different acoustic scenarios, such as direction of arrival (DOA) estimations [132, 133], sound field reconstruction [69, 71], sound field separation [123], near field acoustic holography [134], etc. However, all these applications are based on particle velocity at a single point or multiple discrete points over space. Currently, the dominating and economical method of measuring particle velocity over space is based on the combination of two closely placed pressure microphones or a particle velocity sensor [135] (e.g., the Microflown [89]), which requires performing a measurement in the volume of interest point by point. While this process can be done automatically by an industrial robot, the design and implementation of it with high accuracy is comparably time-consuming and costly. Therefore, if we can develop a representation for particle velocity over a continuous spatial region, it could be a powerful tool that can facilitate the measurement of particle velocity over space. In such instances, spherical harmonics can play a role because they cover the entire space together with the radial functions and they are also spatial basis functions that can be used to describe a variety of acoustics based functions [102, 136] in the three-dimensional space. In the following subsections, we represent particle velocity in terms of sound pressure, and then decompose it in spherical harmonic domain. We also derive closed form velocity coefficients and introduce truncated expressions to enable ease of implementation.

3.2.2 Problem formulation

Consider a source-free spherical region with radius R. The sound pressure at $\boldsymbol{x} = (r, \theta, \phi)$ in the spherical region, resulting from sources outside the region, is given by [109]

$$P(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(kr) Y_{nm}(\theta, \phi). \tag{3.22}$$

The particle velocity vector $\boldsymbol{V}(\boldsymbol{x}, k)$ is related to the sound pressure at $\boldsymbol{x}$ by the Euler's equation [109],

$$\boldsymbol{V}(\boldsymbol{x}, k) = \frac{i}{k\rho_0 c} \vec{\nabla} P(\boldsymbol{x}, k), \tag{3.23}$$

where ρ_0 is the medium density, $\boldsymbol{V}(\boldsymbol{x}, k) = V_r(\boldsymbol{x}, k)\widehat{r} + V_\theta(\boldsymbol{x}, k)\widehat{\theta} + V_\phi(\boldsymbol{x}, k)\widehat{\phi}$, and

$$\vec{\nabla} \equiv \frac{\partial}{\partial r}\widehat{r} + \frac{1}{r}\frac{\partial}{\partial \theta}\widehat{\theta} + \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\widehat{\phi}, \tag{3.24}$$

where $\widehat{r}$, $\widehat{\theta}$, and $\widehat{\phi}$ are the unit vectors in the r, θ, and ϕ directions, respectively. All of the above velocity components are defined on a sphere and therefore, can be decomposed in terms of spherical harmonic functions[2]

$$V_\Psi(\boldsymbol{x}, k) = \sum_{p=0}^{\infty} \sum_{q=-p}^{p} X_{pq}^{(\Psi)}(k, r) Y_{pq}(\theta, \phi); \qquad \Psi = \{r, \theta, \phi\}, \tag{3.25}$$

where $X_{pq}^{(\Psi)}(k, r)$ denote velocity coefficients in the Ψ direction.

Our objective is to derive complete sets of closed form velocity coefficients $X_{pq}^{(\Psi)}(k, r)$ related to each r, θ, ϕ component of the particle velocity vector.

[2]Spherical harmonics are orthogonal spatial basis functions which can be used to decompose any arbitrary function on the sphere.

3.2.3 Velocity coefficients

Substituting (3.22) into (3.23), we have

$$V_r(\boldsymbol{x}, k) = \frac{i}{k\rho_0 c} \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n'(kr) Y_{nm}(\theta, \phi), \tag{3.26}$$

$$V_\theta(\boldsymbol{x}, k) = \frac{i}{k\rho_0 c} \sum_{n=0}^{N} \sum_{m=-n}^{n} A_{nm}\alpha_{nm}(k) \frac{j_n(kr)}{r} P_{nm}'(\cos\theta) e^{im\phi}, \tag{3.27}$$

$$V_\phi(\boldsymbol{x}, k) = -\frac{1}{k\rho_0 c} \sum_{n=0}^{N} \sum_{m=-n}^{n} m A_{nm}\alpha_{nm}(k) \frac{j_n(kr)}{r} \frac{P_{nm}(\cos\theta)}{\sin\theta} e^{im\phi}. \tag{3.28}$$

where $j_n'(kr)$ and $P_{nm}'(\cos\theta)$ denote the derivative of $j_n(kr)$ in terms of r and $P_{nm}(\cos\theta)$ in terms of θ, respectively, expressed as

$$j_n'(kr) = \frac{nkj_{n-1}(kr) - (n+1)kj_{n+1}(kr)}{2n+1}, \tag{3.29}$$

and

$$P_{nm}'(\cos\theta) = (n-m+1)\frac{P_{(n+1)m}(\cos\theta)}{\sin\theta} - (n+1)\frac{\cos\theta P_{nm}(\cos\theta)}{\sin\theta}. \tag{3.30}$$

To achieve velocity coefficients $X_{pq}^{(\Psi)}(k, r)$, the problem is now reduced to decompose (3.26), (3.27), and (3.28) in the form of (3.25). The velocity coefficients in the r direction can be obtained from (3.26) directly as

$$X_{pq}^{(r)}(k, r) = \frac{i}{k\rho_0 c} \alpha_{pq}(k) j_p'(kr). \tag{3.31}$$

Derivations for the velocity coefficients in the θ and ϕ direction are given in Appendix 3.6.1. We introduce the results in the following theorem.

Theorem 1: The velocity coefficients in the θ direction $X_{pq}^{(\theta)}(k, r)$ and ϕ direction $X_{pq}^{(\phi)}(k, r)$ can be expressed, respectively, as

$$X_{pq}^{(\theta)}(k, r) = \frac{2\pi i}{k\rho_0 c} H_{pq} \sum_{n=|q|}^{N} H_{nq}\alpha_{nq}(k) \frac{j_n(kr)}{r} G_1, \tag{3.32}$$

$$X_{pq}^{(\phi)}(k,r) = -\frac{2\pi q}{k\rho_0 c} H_{pq} \sum_{n=|q|}^{N} H_{nq}\alpha_{nq}(k)\frac{j_n(kr)}{r} G_2, \tag{3.33}$$

where

$$H_{pq} = \frac{(-1)^{\frac{q+|q|}{2}}}{2^{|q|}(|q|)!}\sqrt{\frac{(2p+1)}{4\pi}\frac{(p+|q|)!}{(p-|q|)!}}, \tag{3.34}$$

$$\begin{aligned} G_1 = {} & (n+|q|+1)\mathcal{G}(a, b+\frac{1}{2}-\delta_{n+|q|+1}; -\mu_1-\delta_{n+|q|+1}, -\mu_2; \nu_1+\delta_{n+|q|}, \nu_2; \xi_1, \xi_2) \\ & - (n+1)\mathcal{G}(a, b+\frac{1}{2}; -\mu_1, -\mu_2; \nu_1, \nu_2; \xi_1, \xi_2), \end{aligned} \tag{3.35}$$

and

$$\begin{aligned} G_2 &= \mathcal{G}(a, b; -\mu_1, -\mu_2; \nu_1, \nu_2; \xi_1, \xi_2) \\ &= [(-1)^{2b+1}+1]\sum_{j_1=0}^{\mu_1}\sum_{j_2=0}^{\mu_2}\frac{(-\mu_1)_{j_1}(\nu_1)_{j_1}}{(\xi_1)_{j_1}j_1!}\frac{(-\mu_2)_{j_2}(\nu_2)_{j_2}}{(\xi_2)_{j_2}j_2!} \times \frac{B(j_1+j_2+a, b)}{2}, \end{aligned} \tag{3.36}$$

with

$$\delta_M = \begin{cases} 1, & \text{if } M \text{ is even} \\ 0, & \text{if } M \text{ is odd,} \end{cases} \tag{3.37}$$

$$(a)_j = \begin{cases} 1, & \text{if } j = 0 \\ a(a+1)...(a+j-1), & \text{if } j = 1, 2, ..., \end{cases} \tag{3.38}$$

and $a = (2|q|+1)/2$, $b = (3-\delta_{n+|q|}-\delta_{p+|q|})/2$, $\mu_1 = -(1-n+|q|-\delta_{n+|q|})/2$, $\mu_2 = -(1-p+|q|-\delta_{p+|q|})/2$, $\nu_1 = (2+n+|q|-\delta_{n+|q|})/2$, $\nu_2 = (2+p+|q|-\delta_{p+|q|})/2$, $\xi_1 = \xi_2 = |q|+1$, and $B(\cdot)$ denotes the beta function. We note that both $X_{pq}^{(\theta)}(k,r)$ and $X_{pq}^{(\phi)}(k,r)$ are zero for any q whose absolute value is more than N. From the above derivations we show that given the spatial pressure field coefficients, spatial velocity coefficients can be directly calculated.

3.2.4 Velocity truncation error

The representation (3.25) has an infinite number of orthogonal modes. To facilitate the implementation, we can truncate this decomposition to a finite number at a cost of a truncation error. For the particle velocity in the θ direction and ϕ direction,

the infinite summations can be truncated to Q_θ and Q_ϕ, respectively, as

$$\widehat{V}_\theta(\boldsymbol{x},k) = \sum_{p=0}^{Q_\theta}\sum_{q=-p}^{p} X_{pq}^{(\theta)}(k,r)Y_{pq}(\theta,\phi), \tag{3.39}$$

$$\widehat{V}_\phi(\boldsymbol{x},k) = \sum_{p=0}^{Q_\phi}\sum_{q=-p}^{p} X_{pq}^{(\phi)}(k,r)Y_{pq}(\theta,\phi), \tag{3.40}$$

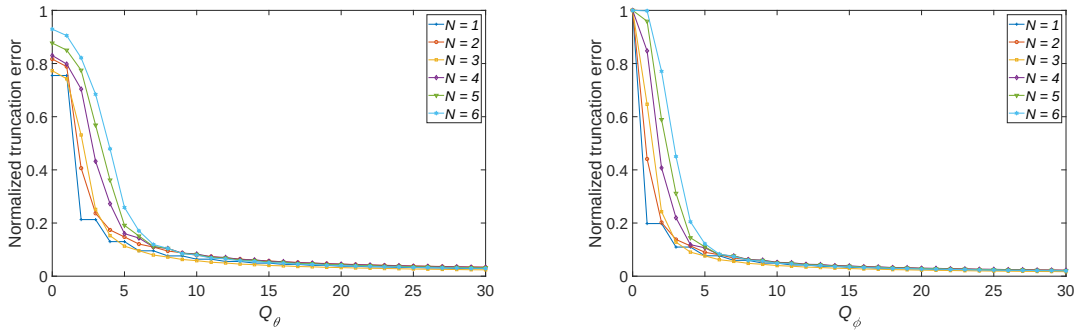
where $\widehat{V}_\theta(\boldsymbol{x},k)$ and $\widehat{V}_\phi(\boldsymbol{x},k)$ denote truncated particle velocity in the θ and ϕ direction, respectively. However, for the particle velocity in the r direction, the representation with finite modes can be written directly by rearranging (3.26) as

$$V_r(\boldsymbol{x},k) = \sum_{p=0}^{Q_r}\sum_{q=-p}^{p} X_{pq}^{(r)}(k,r)Y_{pq}(\theta,\phi), \tag{3.41}$$

where $Q_r = N$. To investigate the relationship between the truncation error and the truncation order, we define the normalized truncation error as

$$\epsilon_{\widetilde{\Psi}}(k) = \frac{\sum\limits_{\forall \boldsymbol{x}} \left| V_{\widetilde{\Psi}}(\boldsymbol{x},k) - \widehat{V}_{\widetilde{\Psi}}(\boldsymbol{x},k)\right|^2}{\sum\limits_{\forall \boldsymbol{x}} \left| V_{\widetilde{\Psi}}(\boldsymbol{x},k)\right|^2}; \qquad \widetilde{\Psi} = \{\theta,\phi\}. \tag{3.42}$$

where $V_{\widetilde{\Psi}}(\boldsymbol{x},k)$ is non-truncated particle velocity, which can be obtained from (3.27) and (3.28), and $\widehat{V}_{\widetilde{\Psi}}(\boldsymbol{x},k)$ is truncated particle velocity, which is measured by (3.39) and (3.40). Given a pressure truncation order, N, which are determined by the radius of the region of interest and the frequency as indicated in Section 3.2.2, the normalized truncation error with respect to the velocity truncation order is shown in Fig. 3.4. We observe that the normalized truncation error decreases as the velocity truncation order increases, and falls to a small value rapidly for all the cases, which means the lower modes contain the majority of the energy and therefore we can truncate the infinite summations to a particular order with a tolerant error. Note that the higher the truncation order is, the less the truncation error becomes but the greater computational complexity the system has.



(a) (b)

Figure 3.4: The normalized truncation error with respect to velocity truncation order for (a) velocity in the θ direction and (b) velocity in the ϕ direction with various pressure truncation orders N.

3.3 Spatial Sound Intensity Vector

Similar to Section 3.2, in this section, we derive the representations of spatial sound intensity vector from the spherical harmonic coefficients of sound pressure.

3.3.1 Introduction

Sound intensity is another fundamental quantity in acoustics that is defined as the power carried by sound waves per unit area in a direction perpendicular to that area. It indicates the rate of energy flow and also gives a measure of direction of energy flow. Therefore, sound intensity has both magnitude and direction components, and that is why it is also referred to as "intensity vector". Sound intensity is useful for localization of sources, reproduction of sound fields, measurement of sound power, measurement of transmission loss, identification of transmission paths, etc. [137] A practical example for sound intensity based applications is seen in high speed trains, where noise sources are measured and identified to create a quiet environment for passengers [138]. The reason why sound intensity is useful for a range of applications as mentioned above is because it plays an important role in directional psychoacoustics. As shown by Gerzon, the human ability to localize

is related to the ratio of the sound intensity vector gain to the total energy [64]. Using this relationship, Arteaga *et al.* propose a novel method for sound reproduction using intensity matching [73, 74]. In addition, intensity is an effective tool to estimate DOA because there is no need to compute a spatial cost function by directly computing the direction of energy flow [99, 101].

However, similar to particle velocity, all the applications mentioned above are based on sound intensity at a single point or several points, but these applications can be extended and facilitated by using spatial sound intensity vectors. For example, sound intensity based reproduction can be realized over space instead of at a single point using spatial intensity matching so that the original sound can be reproduced over a large region for more listeners. Besides, spatial sound intensity vectors facilitates sound intensity measurement over space that is necessary in most of aforementioned applications. In the following subsections, we formulate spatial sound intensity vectors in spherical harmonic domain such that the vectors contain energy and directivity information over continuous spatial regions. The closed form expressions of spatial sound intensity are given and finite modes of spherical harmonics in each expression are also indicated.

3.3.2 Problem formulation

Given sound pressure $P(\boldsymbol{x}, k)$ and the particle velocity vector in spherical coordinates $\boldsymbol{V}(\boldsymbol{x}, k)$=[$V_r(\boldsymbol{x}, k)$,$V_\theta(\boldsymbol{x}, k)$, $V_\phi(\boldsymbol{x}, k)$] at $\boldsymbol{x}$ in a homogeneous medium in space. The intensity relationship for the steady state field is defined as [109]

$$\boldsymbol{I}(\boldsymbol{x}, k) = P^*(\boldsymbol{x}, k)\boldsymbol{V}(\boldsymbol{x}, k), \tag{3.43}$$

where $\boldsymbol{I}(\boldsymbol{x}, k) = [I_r(\boldsymbol{x}, k), I_\theta(\boldsymbol{x}, k), I_\phi(\boldsymbol{x}, k)]$ is sound intensity vector in spherical coordinates. We note that the sound intensity vector $\boldsymbol{I}(\boldsymbol{x}, k)$ is a complex-valued quantity, which contains the active intensity (the real part of $\boldsymbol{I}(\boldsymbol{x}, k)$) and the reactive intensity (the imaginary part of $\boldsymbol{I}(\boldsymbol{x}, k)$).

Similar to particle velocity in (3.25), sound intensity can also be represented as

a linear combination of spherical harmonics as

$$I_\Psi(\boldsymbol{x},k) = \sum_{p=0}^{\infty}\sum_{q=-p}^{p} S_{pq}^{(\Psi)}(k,r)Y_{pq}(\theta,\phi); \qquad \Psi = \{r,\theta,\phi\}, \tag{3.44}$$

where $S_{pq}^{(\Psi)}(k,r)$ are intensity coefficients in Ψ direction.

The aim is to derive complete sets of closed form intensity coefficients $S_{pq}^{(\Psi)}(k,r)$ related to each r, θ, ϕ component of the sound intensity vector.

3.3.3 Intensity coefficients

Substituting (3.26), (3.27), and (3.28), separately, with (3.22) into (3.43), we can get

$$I_r(\boldsymbol{x},k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} R_{nmn'm'}(k,r)Y_{nm}^*(\theta,\phi)Y_{n'm'}(\theta,\phi), \tag{3.45}$$

$$I_\theta(\boldsymbol{x},k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} A_{n'm'}\frac{T_{nmn'm'}(k,r)}{r}Y_{nm}^*(\theta,\phi)P'_{n'm'}(\cos\theta)e^{im'\phi}, \tag{3.46}$$

$$I_\phi(\boldsymbol{x},k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} im'A_{n'm'}\frac{T_{nmn'm'}(k,r)}{r}Y_{nm}^*(\theta,\phi)\frac{P_{nm}(\cos\theta)}{\sin\theta}e^{im'\phi}, \tag{3.47}$$

where $R_{nmn'm'}(k,r) = i\alpha_{nm}^*(k)\alpha_{n'm'}(k)j_n(kr)j'_{n'}(kr)/(k\rho_0 c)$ and $T_{nmn'm'}(k,r) = i\alpha_{nm}^*(k)\alpha_{n'm'}(k)j_n(kr)j_{n'}(kr)/(k\rho_0 c)$.

We now represent (3.45), (3.46), (3.47) in the form of (3.44), and we have the following theorem. The proof of the theorem is given in Appendix 3.6.2.

Theorem 2: The intensity coefficients in the r direction $S_{pq}^{(r)}(k,r)$, θ direction $S_{pq}^{(\theta)}(k,r)$, and ϕ direction $S_{pq}^{(\phi)}(k,r)$ can be expressed, respectively, as

$$S_{pq}^{(r)}(k,r) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} (-1)^{m+q}C_{nn'p}W_1W_2R_{nmn'm'}(k,r), \tag{3.48}$$

$$S_{pq}^{(\theta)}(k,r) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} A_{nm}A_{n'm'}A_{pq}\mathcal{P}_{nmn'm'pq}^{(\theta)}\mathcal{E}_{mm'q}\frac{T_{nmn'm'}(k,r)}{r}, \quad (3.49)$$

$$S_{pq}^{(\phi)}(k,r) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} im'A_{nm}A_{n'm'}A_{pq}\mathcal{P}_{nmn'm'pq}^{(\phi)}\mathcal{E}_{mm'q}\frac{T_{nmn'm'}(k,r)}{r}, \quad (3.50)$$

where $C_{nn'p} = \sqrt{(2n+1)(2n'+1)(2p+1)/4\pi}$, $\mathcal{E}_{mm'q} = 2\pi$ when $m'-m-q=0$, otherwise $\mathcal{E}_{mm'q} = 0$, $W_1 = \begin{pmatrix} n & n' & p \\ 0 & 0 & 0 \end{pmatrix}$ and $W_2 = \begin{pmatrix} n & n' & p \\ -m & m' & -q \end{pmatrix}$ denoting Wigner 3-j symbols, $\mathcal{P}_{nmn'm'pq}^{(\theta)} = (n'-m'+1)\mathcal{J}_1 - (n'+1)\mathcal{J}_2$ with

$$\begin{aligned}\mathcal{J}_1 =& H(n,m)H(n'+1,m')H(p,q)\mathcal{J}(\frac{m+m'+q+1}{2}, \frac{4-\delta_{m+n}-\delta_{m'+n'+1}-\delta_{p+q}}{2};\\ &\frac{1+m-n-\delta_{m+n}}{2}, \frac{m'-n'-\delta_{m'+n'+1}}{2}, \frac{1+q-p-\delta_{p+q}}{2}; \frac{2+m+n}{2}\\ &-\frac{\delta_{m+n}}{2}, \frac{3+m'+n'-\delta_{m'+n'+1}}{2}, \frac{2+p+q-\delta_{p+q}}{2}; m+1, m'+1, q+1),\end{aligned} \quad (3.51)$$

$$\begin{aligned}\mathcal{J}_2 =& H(n,m)H(n',m')H(p,q)\mathcal{J}(\frac{m+m'+q+1}{2}, \frac{5-\delta_{m+n}-\delta_{m'+n'}-\delta_{p+q}}{2};\\ &\frac{1+m-n-\delta_{m+n}}{2}, \frac{1+m'-n'-\delta_{m'+n'}}{2}, \frac{1+q-p-\delta_{p+q}}{2}; \frac{2+m+n}{2}\\ &-\frac{\delta_{m+n}}{2}, \frac{2+m'+n'-\delta_{m'+n'}}{2}, \frac{2+p+q-\delta_{p+q}}{2}; m+1, m'+1, q+1),\end{aligned} \quad (3.52)$$

where $H(n,m) = (-1)^m(n+m)!/[2^m m!(n-m)!]$, and

$$\begin{aligned}\mathcal{J}(\alpha,\beta;-n_1,-n_2,-n_3;a_1,a_2,a_3;c_1,c_2,c_3) =& \sum_{j_1=0}^{n_1}\sum_{j_2=0}^{n_2}\sum_{j_3=0}^{n_3}\frac{(-n_1)_{j_1}(a_1)_{j_1}}{(c_1)_{j_1}j_1!}\\ \times& \frac{(-n_2)_{j_2}(a_2)_{j_2}}{(c_2)_{j_2}j_2!}\frac{(-n_3)_{j_3}(a_3)_{j_3}}{(c_3)_{j_3}j_3!}\frac{((-1)^{2\beta+1}+1)B(j_1+j_2+j_3+\alpha,\beta)}{2},\end{aligned} \quad (3.53)$$

$$\begin{aligned}\mathcal{P}^{(\phi)}_{nmn'm'pq} =& H(n,m)H(n',m')H(p,q)\mathcal{J}(\frac{m+m'+q+1}{2}, \frac{4-\delta_{m+n}-\delta_{m'+n'}-\delta_{p+q}}{2};\\ & \frac{1+m-n-\delta_{m+n}}{2}, \frac{1+m'-n'-\delta_{m'+n'}}{2}, \frac{1+q-p-\delta_{p+q}}{2}; \frac{2+m+n}{2}\\ & -\frac{\delta_{m+n}}{2}, \frac{2+m'+n'-\delta_{m'+n'}}{2}, \frac{2+p+q-\delta_{p+q}}{2}; m+1, m'+1, q+1).\end{aligned} \tag{3.54}$$

We note that the intensity coefficients are directly determined by the pressure coefficients $\alpha_{nm}(k)$ together with the radial functions (i.e., the spherical Bessel function), and all the rest terms in the expressions are either constants or indices-only dependent variables.

3.3.4 Intensity truncation error

Likewise, it would be beneficial to truncate the intensity expression (3.44), expressed as

$$\widehat{I}_{\Psi}(\boldsymbol{x},k) = \sum_{p=0}^{P_{\Psi}} \sum_{q=-p}^{p} S^{(\Psi)}_{pq}(k,r) Y_{pq}(\theta,\phi), \tag{3.55}$$

where P_{Ψ} denotes the truncation order for intensity in the Ψ direction. Given the pressure truncation order of N, the order of sound intensity expressions in r direction is $P_r = 2N$ because of the selection rule of Wigner 3-j symbols that W_1 and W_2 equal zero when $p > n + n'$. However, it is quite different and hard to identify the active modes directly for sound intensity expressions in θ and ϕ direction. In order to show the active modes, intuitively, we calculate the normalized truncation error in terms of the truncation order of intensity in the θ direction P_θ for five different pressure truncation orders N determined by different radii of the region of interest[3], which is shown in Fig. 3.5. Note that the normalized truncation error is measured by

$$\epsilon_\theta(k) = \frac{\sum_{\forall \boldsymbol{x}} \left| I_\theta(\boldsymbol{x},k) - \widehat{I}_\theta(\boldsymbol{x},k) \right|^2}{\sum_{\forall \boldsymbol{x}} \left| I_\theta(\boldsymbol{x},k) \right|^2}, \tag{3.56}$$

[3] $S^{(\theta)}_{pq}(k,r)$ and $S^{(\phi)}_{pq}(k,r)$ have similar expressions, and here we take the θ as an example to show the truncation error.

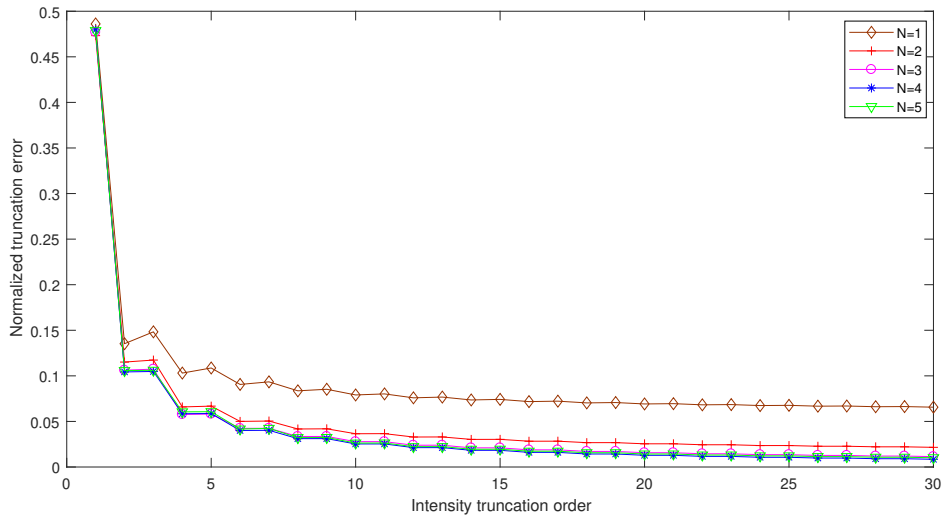


Figure 3.5: The relationship between normalized truncation error and intensity truncation order P_θ for various pressure truncation orders N.

where $\widehat{I}_\theta(\boldsymbol{x}, k)$ is reconstructed sound intensity at point $\boldsymbol{x}$ using (3.44) with given truncation orders. Observe that the intensity truncation error becomes less and falls to an acceptable value as intensity truncation order increases no matter what N is. Also, as N grows, it has less influence on intensity truncation error.

As mentioned at the beginning of this section, the theory of spatial sound intensity vectors in spherical harmonic domain can be largely useful in many applications including intensity measurement over continuous spatial regions, sound field reproduction, DOA estimation, etc. We here provide simulation results for the application of spatial intensity measurement, using pressure coefficients of a sound field (spherical harmonic domain). Figure 3.6 shows sound intensity on a sphere obtained from the proposed theory using (3.55) (note that the spherical pressure coefficients were obtained using a spherical microphone array) against point by point measurement (note that point by point intensity was simulated using the theoretical expression for pressure due to a plane wave and the relationship in (3.43)). We observe that reconstructed sound intensity vectors by using the proposed theory are similar to actual point by point measurements. The simulation results also indicate that the truncation leads to little error for an appropriate truncation order.

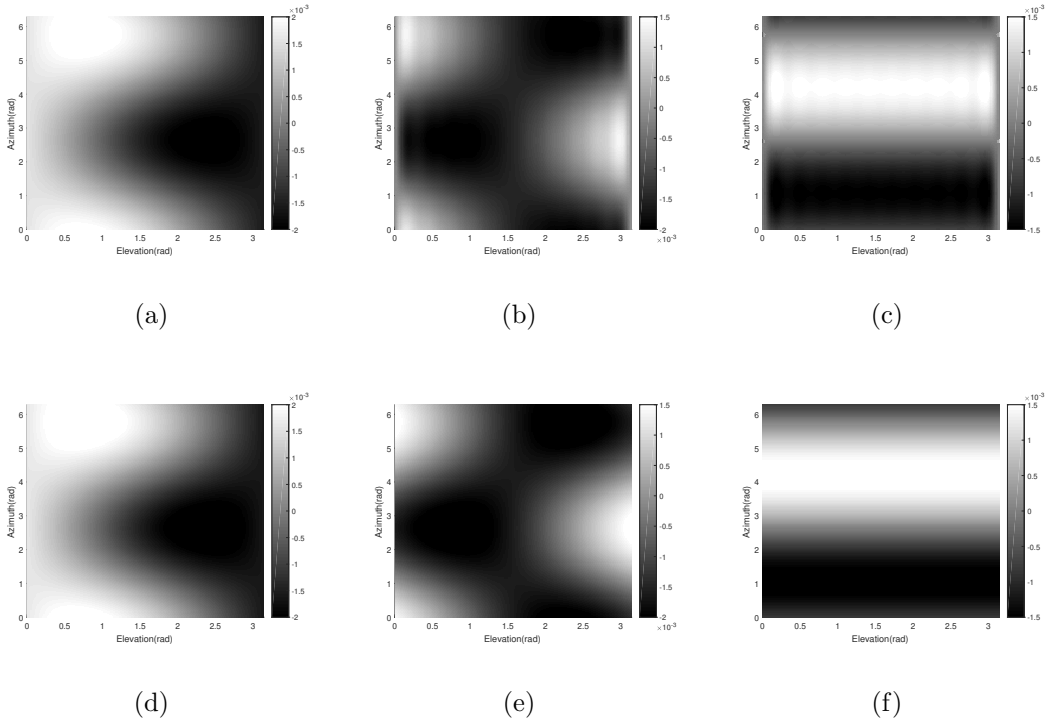


Figure 3.6: Sound intensity on a sphere with radius of 0.05 m, generated by a plane wave from $(3\pi/4, 5\pi/6)$, with frequency 600 Hz. (a-c) Sound intensity in r, θ and ϕ direction, separately, calculated using the proposed theory, (d-f) sound intensity in r, θ and ϕ direction, separately, obtained from point by point measurement.

3.4 Summary and Contributions

In this chapter, we have presented three useful techniques that are closely related to the spatial sound field systems in the following chapters of this thesis. To the end, we list the major contributions of this chapter below:

- We proposed a method of separating 3D interior and exterior sound field using a planar array of differential microphones, which can be used to extract the desired sound pressure in a mixed sound field environment. The proposed method was proved to be effective in a noisy reverberant room.

- We formulated continuous particle velocity expressions over space as a function of spherical harmonic coefficients of sound pressure. The velocity trun-

cation error was also analyzed for the implementation.

- We formulated the theory of spatial sound intensity vectors in spherical harmonic domain that is applicable to a variety of acoustic scenarios. The complete sets of closed form intensity coefficients were derived, and finite modes of spherical harmonics were suggested for practical implementations.

3.5 Related Publications

Much of this chapter's work has been published in the following journal papers and conference proceedings.

- **H. Zuo**, T. D. Abhayapala, and P. N. Samarasinghe, "Particle velocity assisted three dimensional sound field reproduction using a modal-domain approach," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 2119-2133, 2020.

- **H. Zuo**, P. N. Samarasinghe, T. D. Abhayapala, and G. Dickins, "Spatial sound intensity vectors in spherical harmonic domain," *The Journal of the Acoustical Society of America (JASA)*, vol. 145, no. 2, pp. EL149-EL155, 2019.

- **H. Zuo**, P. N. Samarasinghe, T. D. Abhayapala, and G. Dickins, "Erratum: Spatial sound intensity vectors in spherical harmonic domain [j. acoust. soc. am. 145 (2), el149–el155 (2019)]," *The Journal of the Acoustical Society of America (JASA)*, vol. 146, no. 1, pp. 164-165, 2019.

- **H. Zuo**, P. N. Samarasinghe, and T. D. Abhayapala, "Exterior-interior 3d sound field separation using a planar array of differential microphones," in *Proc. 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 216–220, Tokyo, Japan, September, 2018.

3.6 Appendix

3.6.1 Proof of Theorem 1

By multiplying both sides of (3.25) by $Y^*_{p'q'}(\theta,\phi)$ and integrating them with respect to θ and ϕ, for the particle velocity in the θ direction, we have

$$\begin{aligned}\int_0^{2\pi}\int_0^{\pi} V_\theta(\boldsymbol{x},k)Y^*_{p'q'}(\theta,\phi)\sin\theta d\theta d\phi \\ = \int_0^{2\pi}\int_0^{\pi}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} X^{(\theta)}_{pq}(k,r)Y_{pq}(\theta,\phi)Y^*_{p'q'}(\theta,\phi)\sin\theta d\theta d\phi,\end{aligned} \tag{3.57}$$

Due to the orthogonality of the spherical harmonics, (3.57) can be simplified as

$$X^{(\theta)}_{pq} = \int_0^{2\pi}\int_0^{\pi} V_\theta(\boldsymbol{x},k)Y^*_{pq}(\theta,\phi)\sin\theta d\theta d\phi \tag{3.58}$$

Substituting (3.27) into (3.58), we have

$$\begin{aligned}X^{(\theta)}_{pq} = \frac{i}{k\rho_0 c}\int_0^{2\pi}\int_0^{\pi}\sum_{n=0}^{N}\sum_{m=-n}^{n} A_{nm}A_{pq}\alpha_{nm}(k)\frac{j_n(kr)}{r} \\ \times P'_{nm}(\cos\theta)P_{pq}(\cos\theta)e^{i(m-q)\phi}\sin\theta d\theta d\phi\end{aligned} \tag{3.59}$$

Replacing $P'_{nm}(\cos\theta)$ in (3.59) with (3.30) and letting $\cos\theta$ be t, it reduces to calculate the following two integrals,

$$\mathcal{P}^{(\theta)}_{nmpq} = \int_{-1}^{1}[(n-m+1)\frac{P_{(n+1)m}(t)P_{pq}(t)}{\sqrt{1-t^2}} - (n+1)\frac{tP_{nm}(t)P_{pq}(t)}{\sqrt{1-t^2}}]dt, \tag{3.60}$$

and

$$\mathcal{E}_{mq} = \int_0^{\pi} e^{i(m-q)\phi}d\phi = \begin{cases} 2\pi, & \text{if } m=q \\ 0, & \text{otherwise.}\end{cases} \tag{3.61}$$

In order to calculate the integral of (3.60), we begin with the Euler integral,

$$\int_0^1 (1-t)^{a-1}t^{b-1}dt = B(a,b) \quad (\mathrm{Re}(a)>0, \mathrm{Re}(b)>0), \tag{3.62}$$

where Re(·) denotes the real part. From (3.62), we can obtain

$$\int_{-1}^{1}(1-t^2)^{a-1}t^{2b-1}dt = \frac{((-1)^{2b+1}+1)B(a,b)}{2} \qquad (\mathrm{Re}(a)>0, \mathrm{Re}(b)>0). \quad (3.63)$$

Into (3.63) we add the product of two hypergeometric polynomials defined by

$$F(-\mu,\nu;\xi;x) = \sum_{j=0}^{\mu}\frac{(-\mu)_j(\nu)_j}{(\xi)_j j!}x^j. \quad (3.64)$$

to obtain

$$\begin{aligned}&\int_{-1}^{1}(1-t^2)^{\overline{a}-1}t^{2\overline{b}-1}F(-\overline{\mu}_1,\overline{\nu}_1;\overline{\xi}_1;1-t^2)F(-\overline{\mu}_2,\overline{\nu}_2;\overline{\xi}_2;1-t^2)dt\\ &=\mathcal{G}(\overline{a},\overline{b};-\overline{\mu}_1,-\overline{\mu}_2;\overline{\nu}_1,\overline{\nu}_2,\overline{\xi}_1,\overline{\xi}_2) \quad (\mathrm{Re}(\overline{a})>0, \mathrm{Re}(\overline{b})>0),\end{aligned} \quad (3.65)$$

The associated Legendre function can be represented by the hypergeometric polynomials, which is given as [139]

$$\begin{aligned}P_{nm}(t) = D_{nm}[&\delta_{|m|+n}(1-t^2)^{|m|/2}F(\frac{|m|-n}{2},\frac{1+|m|}{2}+\frac{n}{2};|m|+1;1-t^2)\\ &+\delta_{|m|+n+1}(1-t^2)^{|m|/2}tF(\frac{1+|m|-n}{2},\frac{2+|m|+n}{2};|m|+1;1-t^2)],\end{aligned} \quad (3.66)$$

where

$$D_{nm} = \begin{cases}\frac{(-1)^{|m|}(n+|m|)!}{2^{|m|}|m|!(n-|m|)!}, & \text{when } m\geq 0\\ \frac{1}{2^{|m|}|m|!}, & \text{when } m<0.\end{cases} \quad (3.67)$$

Substituting (3.66) into (3.60) together with (3.65), we have

$$\begin{aligned}\mathcal{P}^{(\theta)}_{nmpq} = (n-m+1)D_{(n+1)m}D_{pq}\mathcal{G}(a,b+\frac{1}{2}-\delta_{n+|m|+1};-\mu_1-\delta_{n+|m|+1},-\mu_2;\\ \nu_1+\delta_{n+|m|},\nu_2;\xi_1,\xi_2)-(n+1)D_{nm}D_{pq}\mathcal{G}(a,b+\frac{1}{2};-\mu_1,-\mu_2;\nu_1,\nu_2;\xi_1,\xi_2).\end{aligned} \quad (3.68)$$

Substituting (3.68) and (3.61) into (3.59) completes the proof of (3.32).

Similarly, for the particle velocity in the ϕ direction, we have

$$X_{pq}^{(\phi)} = -\frac{1}{k\rho_0 c}\sum_{n=0}^{N}\sum_{m=-n}^{n} mA_{nm}A_{pq}\alpha_{nm}(k)\frac{j_n(kr)}{r}\mathcal{P}_{nmpq}^{(\phi)}\mathcal{E}_{mq}, \tag{3.69}$$

where

$$\begin{aligned}\mathcal{P}_{nmpq}^{(\phi)} &= \int_{-1}^{1}\frac{P_{nm}(t)P_{pq}(t)}{\sqrt{1-t^2}}dt \\ &= D_{nm}D_{pq}\mathcal{G}(a,b;-\mu_1,-\mu_2;\nu_1,\nu_2;\xi_1,\xi_2).\end{aligned} \tag{3.70}$$

Substituting (3.70) and (3.61) into (3.69) completes the proof of (3.33).

3.6.2 Proof of Theorem 2

Multiplying both sides of (3.44) by $Y_{pq}^*(\theta,\phi)$ and integrating with respect to θ and ϕ,

$$S_{pq}^{(\Psi)}(k,r) = \int_0^{2\pi}\int_0^{\pi} I_\Psi(\boldsymbol{x},k)Y_{pq}^*(\theta,\phi)\sin\theta d\theta d\phi; \qquad \Psi = \{r,\theta,\phi\}. \tag{3.71}$$

In the r direction, substituting (3.45) into (3.71),

$$\begin{aligned}&S_{pq}^{(r)}(k,r) \\ &= \int_0^{2\pi}\int_0^{\pi}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\sum_{n'=0}^{\infty}\sum_{m'=-n'}^{n'} R_{nmn'm'}(k,r)Y_{nm}^*(\theta,\phi)Y_{n'm'}(\theta,\phi)Y_{pq}^*(\theta,\phi)\sin\theta d\theta d\phi.\end{aligned} \tag{3.72}$$

The integral of products of three spherical harmonics is given by [140],

$$\begin{aligned}&\int_0^{2\pi}\int_0^{\pi} Y_{l_1m_1}(\theta,\phi)Y_{l_2m_2}(\theta,\phi)Y_{l_3m_3}(\theta,\phi)\sin\theta d\theta d\phi \\ &= \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}}\begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3\end{pmatrix},\end{aligned} \tag{3.73}$$

as well as $Y_{nm}^*(\theta,\phi) = (-1)^m Y_{n(-m)}(\theta,\phi)$. Substituting (3.73) into (3.72) completes the proof of (3.48).

For the θ direction, we substitute (3.46) into (3.71), (3.49) can be achieved with

$$\mathcal{P}^{(\theta)}_{nmn'm'pq} = \int_{-1}^{1} P_{nm}(\cos\theta) P'_{n'm'}(\cos\theta) P_{pq}(\cos\theta) d\cos\theta, \tag{3.74}$$

$$\mathcal{E}_{mm'q} = \int_{0}^{2\pi} e^{-im\phi} e^{im'\phi} e^{-iq\phi} d\phi = \begin{cases} 2\pi, & \text{if } m' - m - q = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.75}$$

To solve the integral of (3.74), we begin with the transformed Euler integral, adding the product of three hypergeometric polynomials, defined by

$$\begin{aligned} &\int_{-1}^{1} (1-t^2)^{\alpha-1} t^{2\beta-1} F(-n_1, a_1; c_1; 1-t^2) F(-n_2, a_2; c_2; 1-t^2) F(-n_3, a_3; c_3; 1-t^2) dt = \\ &\sum_{j_1=0}^{n_1} \sum_{j_2=0}^{n_2} \sum_{j_3=0}^{n_3} \frac{(-n_1)_{j_1}(a_1)_{j_1}}{(c_1)_{j_1} j_1!} \frac{(-n_2)_{j_2}(a_2)_{j_2}}{(c_2)_{j_2} j_2!} \frac{(-n_3)_{j_3}(a_3)_{j_3}}{(c_3)_{j_3} j_3!} \frac{((-1)^{2\beta+1}+1) B(j_1+j_2+j_3+\alpha, \beta)}{2}, \end{aligned} \tag{3.76}$$

and then the relation between the associated Legendre function and the hypergeometric function can be given as [139]

$$\begin{aligned} P_{nm}(t) =& H(n,m)[\delta_{m+n}(1-t^2)^{m/2} F(\frac{m-n}{2}, \frac{1+m+n}{2}; m+1; 1-t^2) \\ &+ \delta_{m+n+1}(1-t^2)^{m/2} t F(\frac{1+m-n}{2}, \frac{2+m+n}{2}; m+1; 1-t^2)], \end{aligned} \tag{3.77}$$

Substituting (3.77) into (3.74) together with (3.76) completes the proof.

Similarly, for the ϕ direction, by substituting (3.47) into (3.71) we have

$$\mathcal{P}^{(\phi)}_{nmn'm'pq} = \int_{-1}^{1} \frac{P_{nm}(\cos\theta) P'_{n'm'}(\cos\theta) P_{pq}(\cos\theta)}{\sqrt{1-\cos^2\theta}} d\cos\theta, \tag{3.78}$$

Substituting (3.77) into (3.78) together with (3.76) completes the proof.

Chapter 4

Particle Velocity Assisted Spatial Sound Field Reproduction

In literature, particle velocity has been introduced to improve performance of spatial sound field reproduction systems. This chapter develops a practical and accurate 3D sound field reproduction system with the aid of the theory of spatial particle velocity vector introduced in Chapter 3. The sound field within a target region is controlled by a weighted cost function we built to optimize the continuous particle velocity, as well as sound pressure, on the boundary of the region. In contrast to the conventional spatial sound field reproduction methods in the spherical harmonic domain, the proposed method allows for non-uniform loudspeaker geometry with a limited number of loudspeakers, thus providing a flexible array arrangement. The performance of the proposed method is evaluated through numerical simulations in both a free field and a reverberant room. Finally, we prove the proposed method in an objective experiment with real-world measurements of room impulse response.

4.1 Introduction

As we reviewed in Chapter 1, there have been various studies on spatial sound field reproduction, such as Ambisonics [19], HOA [20–24], and WFS [4, 5, 7, 15], all of which aim to reconstruct the original sound pressure. However, there are limita-

tions for these methods to implement in real-world applications. WFS requires a continuous distribution of monopole and normally oriented dipole secondary sources on the boundary of the region. Although the continuous distribution can be approximated by an array of equally placed loudspeakers, a significant number of loudspeakers are required for broadband reproduction over a large region. Similarly, for a given order of a reproduction system, HOA requires sufficient loudspeakers to match all the spherical harmonics to the given order in order to avoid spatial aliasing [20]. The performance of HOA also deteriorates when the geometry of loudspeaker array is non-uniform.

In addition to sound pressure, particle velocity is another acoustic quantity describing a sound field, which has been incorporated in active control to enlarge the zones of quiet [141,142]. Particle velocity can also play a significant role in the sound field reproduction systems. Shin *et al.* proposed a particle velocity controlled sound field reproduction technique to simplify the regularization of the inverse problem for a non-uniformly spaced loudspeaker array [70, 71]. The work only considers particle velocity without sound pressure. A joint optimization, based on both sound pressure and particle velocity, was presented for broadband multi-zone sound field reproduction in [68,69], where how the particle velocity impacts on the reproduction performance was explicitly analyzed. This work is restricted to reproduce the desired sound field in a 2D plane. Besides, all of these previous particle velocity based sound field reproduction methods focus on controlling the particle velocity at multiple discrete points on the boundary of the target region. To guarantee the accuracy of reproduced sound field over a large region, it requires a large quantity of control points, which implies that a practical system implementation requires to measure the particle velocity point by point for all the control points. The process of this measurement is time-consuming and costly. In this chapter, we incorporate the theory of continuous particle velocity over space in Chapter 3 into HOA to investigate a flexible scheme for 3D sound field reproduction that can relax the limitations of loudspeaker number and placement.

The remainder of this chapter is structured as follows. Section 4.2 contains problem definition and the objective of the work. In Section 4.3, the algorithm of velocity assisted 3D sound field reproduction is explicitly investigated for free fields and reverberant rooms. Section 4.4 validates the proposed method by comparing

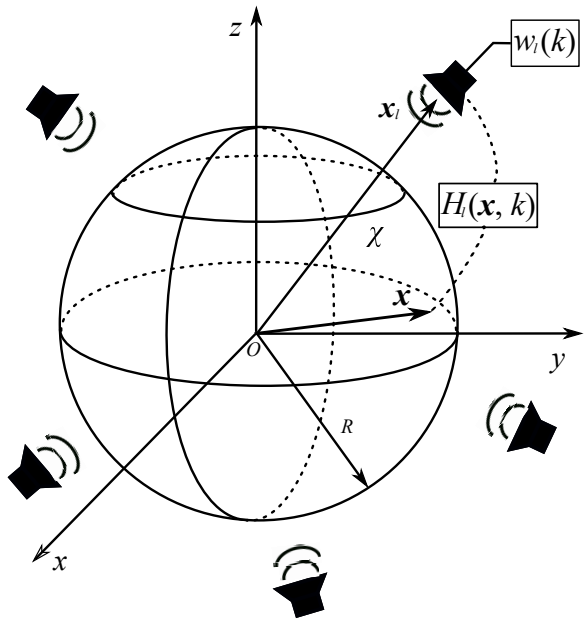


Figure 4.1: Geometry of the sound field reproduction system in this paper. An array of loudspeakers is located outside χ with driving signals $w_\ell(k)$ and acoustic transfer function $H_\ell(\boldsymbol{x}, k)$ from the ℓth loudspeaker to a point $\boldsymbol{x}$ within the region.

it with the conventional HOA (i.e., the mode matching) in the aforementioned two environments. In Section 4.5, we validate the algorithm with an objective experiment using impulse response measurements of a real-world room.

4.2 Problem Definition

Our objective is to reproduce the pressure $P_d(\boldsymbol{x}, k)$ of a desired sound field at any point $\boldsymbol{x}$ within a source free spherical region χ of radius R as shown in Fig. 4.1. The desired sound field can be a plane wave, a point source or a field measured in a real-world scenario. Let there be an array of loudspeakers outside of χ, with the ℓth loudspeaker located at $\boldsymbol{x}_\ell = (r_\ell, \theta_\ell, \phi_\ell)$ with respect to the origin O. The sound pressure at $\boldsymbol{x}$ due to the ℓth loudspeaker can be written as

$$P_\ell(\boldsymbol{x}, k) = w_\ell H_\ell(\boldsymbol{x}, k), \tag{4.1}$$

where w_ℓ is the driving signal applying to the ℓth loudspeaker, and $H_\ell(\boldsymbol{x}, k)$ is the acoustic transfer function between the ℓth loudspeaker to $\boldsymbol{x}$ for arbitrary environments. Assume there is a total of L loudspeakers, therefore, the total reproduced sound pressure at $\boldsymbol{x}$ is given by

$$P_a(\boldsymbol{x}, k) = \sum_{\ell=1}^{L} w_\ell H_\ell(\boldsymbol{x}, k). \tag{4.2}$$

The design task of sound field reproduction is to find the loudspeaker driving signals that can reproduce the desired sound field optimally. The popular approach to solving this problem is to minimize the least squares error between the desired sound field and the reproduced sound field within the region χ, i.e.,

$$\min_{\boldsymbol{x}\in\chi} |P_d(\boldsymbol{x}, k) - P_a(\boldsymbol{x}, k)|^2. \tag{4.3}$$

This can be formulated as mode matching [20] by expressing $P_d(\boldsymbol{x}, k)$ and $P_a(\boldsymbol{x}, k)$ in terms of spherical harmonics similar to (3.22) in the spherical harmonic domain. Although a 3D sound field can be accurately reproduced by this method, it requires at least $(N+1)^2$ loudspeakers uniformly distributed on a sphere. The performance of mode matching degrades when there is only a limited number of loudspeakers or the loudspeaker geometry is irregular. It is the constraint we wish to relax through the introduction of spatial particle velocity vectors, which we present next.

4.3 Velocity Assisted Reproduction Algorithm

The velocity-assisted sound field reproduction algorithm is presented in this section. We first model the point-to-region transfer function between each loudspeaker to the target region for both a free field and a reverberant environment in Section 4.3.1. The driving signals of loudspeakers are then designed by jointly optimizing the sound pressure and the particle velocity in Section 4.3.2. Finally, 3D sound field reproduction error is given in Section 4.3.3.

4.3.1 Point-to-region transfer function

Acoustic transfer function plays an important role in sound field reproduction because it reflects how much sound from a source is observed by a receiver. The acoustic transfer function between the ℓth loudspeaker to $\boldsymbol{x}$, $H_\ell(\boldsymbol{x}, k)$, can be expressed by the modal decomposition as

$$H_\ell(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_{nm}^{(\ell)}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{4.4}$$

where $\beta_{nm}^{(\ell)}(k)$ are the sound field coefficients of the acoustic transfer function. We also refer to (4.4) as the point-to-region transfer function, because the acoustic transfer function between the ℓth loudspeaker to any point in the region χ is readily available if $\beta_{nm}^{(\ell)}(k)$ are known. The coefficients $\beta_{nm}^{(\ell)}(k)$ can also represent the directional characteristics of the loudspeaker [143], therefore, (4.4) can be the acoustic transfer function due to not only omni-directional sources but also directional loudspeakers. For simplicity, in the following two sections, we assume all loudspeakers to be omni-directional. Note that in practice the aforementioned point-region transfer function coefficients related to a given loudspeaker can be extracted from recordings of higher order microphones such as an EigenMike [75], by feeding a sweep signal into the loudspeaker.

Free field

In a free field, the point-to-region transfer function $H_\ell^{\text{dir}}(\boldsymbol{x}, k)$ only contains the direct path component, which is determined by the Green's function [109],

$$H_\ell^{\text{dir}}(\boldsymbol{x}, k) = \frac{e^{ik||\boldsymbol{x}-\boldsymbol{x}_\ell||_2}}{4\pi||\boldsymbol{x}-\boldsymbol{x}_\ell||_2} = \sum_{n=0}^{N} \sum_{m=-n}^{n} ikh_n(kr_\ell) Y_{nm}^*(\theta_\ell, \phi_\ell) j_n(kr) Y_{nm}(\theta, \phi). \tag{4.5}$$

In this case, the coefficients of the point-to-region transfer function due to the ℓth loudspeaker can be expressed as

$$\beta_{nm}^{(\ell)}(k) = ikh_n(kr_\ell) Y_{nm}^*(\theta_\ell, \phi_\ell). \tag{4.6}$$

Reverberant environment

We model the reverberant environment as a shoebox room, and derive the point-to-region transfer function using the spherical harmonics based generalized image source method [143]. The point-to-region transfer function in the reverberant environment due to the ℓth loudspeaker $H_\ell^{\text{rvb}}(\boldsymbol{x}, k)$ can be expressed as

$$H_\ell^{\text{rvb}}(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{ik}{\sqrt{4\pi}} \alpha_{nm}^{00}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{4.7}$$

where $\alpha_{nm}^{00}(k)$ denote the coupling coefficients. The exact expression of $\alpha_{nm}^{00}(k)$, as well as the proof of (4.7), is given in Appendix 4.8.1. Therefore, in the reverberant environment, the expression of the point-to-region transfer function coefficients due to the ℓth loudspeaker is

$$\beta_{nm}^{(\ell)}(k) = \frac{ik}{\sqrt{4\pi}} \alpha_{nm}^{00}(k), \tag{4.8}$$

where the location information of the ℓth loudspeaker is incorporated in $\alpha_{nm}^{00}(k)$. Note that (4.8) holds only for the simulated shoebox room. For a real-world room with arbitrary geometries, one can estimate the coefficients from impulse response measurements, which we present in Section 4.5.

4.3.2 Velocity assisted optimization

The proposed method is inspired by Kirchhoff-Helmholtz integral equation. We borrow the idea from the equation that the sound pressure is completely determined within a volume free of sources, if sound pressure and particle velocity are controlled in all points on its surface. We therefore propose to reproduce the sound field within a spherical region (free of sources) by controlling the continuous distributions of sound pressure and particle velocity on its surface in the spherical harmonic domain, which is different from controlling multiple discrete points on the surface [68]. The 3D sound field reproduction problem is then reduced to optimizing the sound pressure and particle velocity on the surface of the region χ. From (3.22), the

desired sound pressure can be given by

$$P_d(\boldsymbol{x},k) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \underbrace{\alpha_{nm}^{(d)}(k) j_n(kr)}_{\widehat{\alpha}_{nm}^{(d)}(k,r)} Y_{nm}(\theta,\phi), \tag{4.9}$$

where $\alpha_{nm}^{(d)}(k)$ are the pressure coefficients of the desired sound field. Replacing $\alpha_{nm}(k)$ with $\alpha_{nm}^{(d)}(k)$ in (3.31), (3.32), and (3.33) and substituting them into (3.25), we have the desired particle velocity as

$$V_{\Psi}^{(d)}(\boldsymbol{x},k) = \sum_{p=0}^{Q_\Psi}\sum_{q=-p}^{p} X_{pq}^{(\Psi,d)}(k,r) Y_{pq}(\theta,\phi). \tag{4.10}$$

Similarly, by replacing $\alpha_{nm}(k)$ with $\beta_{nm}^{(\ell)}(k)$, the reproduced sound pressure and particle velocity due to the loudspeaker array can be written, respectively, as

$$P_a(\boldsymbol{x},k) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \underbrace{\sum_{\ell=1}^{L} w_\ell \beta_{nm}^{(\ell)}(k) j_n(kr)}_{\widehat{\beta}_{nm}^{(a)}(k,r)} Y_{nm}(\theta,\phi), \tag{4.11}$$

$$V_{\Psi}^{(a)}(\boldsymbol{x},k) = \sum_{p=0}^{Q_\Psi}\sum_{q=-p}^{p} \underbrace{\sum_{\ell=1}^{L} w_\ell X_{pq}^{(\Psi,\ell)}(k,r)}_{X_{pq}^{(\Psi,a)}(k,r)} Y_{pq}(\theta,\phi). \tag{4.12}$$

Note that $\widehat{\alpha}_{nm}^{(d)}(k,r)/X_{pq}^{(\Psi,d)}(k,r)$ and $\widehat{\beta}_{nm}^{(a)}(k,r)/X_{pq}^{(\Psi,a)}(k,r)$ are the desired and reproduced pressure/particle velocity coefficients on the surface with the radius of r, respectively, which represent the continuous sound pressure/particle velocity on the surface. Using these coefficients, a weighted least squares optimization criterion, including both sound pressure and particle velocity on the surface of χ, can be formulated

$$\min_{\boldsymbol{W}}\{\tau^2||\boldsymbol{\beta}_A(k,R)\boldsymbol{W}(k) - \boldsymbol{\alpha}_D(k,R)||_2^2 + (1-\tau)^2||\boldsymbol{X}_A(k,R)\boldsymbol{W}(k) - \boldsymbol{X}_D(k,R)||_2^2\}, \tag{4.13}$$

where $\tau \in [0,1]$ is the weighting coefficient to adjust the relative weight of sound pressure and particle velocity,
$\boldsymbol{\alpha}_D(k,R) = [\alpha_{00}^{(d)}(k)j_0(kR), \alpha_{1(-1)}^{(d)}(k)j_1(kR), ..., \alpha_{NN}^{(d)}(k)j_N(kR)]^T$ is a $(N+1)^2$ long vector,
$\boldsymbol{X}_D(k,R) = [X_{00}^{(r,d)}(k,R), ..., X_{Q_rQ_r}^{(r,d)}(k,R), X_{00}^{(\theta,d)}(k,R), ..., X_{Q_\theta Q_\theta}^{(\theta,d)}(k,R), X_{00}^{(\phi,d)}(k,R), ..., X_{Q_\phi Q_\phi}^{(\phi,d)}(k,R)]^T$ is a $[(Q_r+1)^2+(Q_\theta+1)^2+(Q_\phi+1)^2]$ long vector containing all three components of the particle velocity vector, $\boldsymbol{W}(k) = [w_1(k), w_2(k), ..., w_L(k))]^T$ is a L long vector, $\boldsymbol{\beta}_A(k,R)$ is a $(N+1)^2$ by L matrix given by

$$\boldsymbol{\beta}_A(k,R) = \begin{bmatrix} \beta_{00}^{(1)} j_0(kR) & \cdots & \beta_{00}^{(L)} j_0(kR) \\ \beta_{1(-1)}^{(1)} j_1(kR) & \cdots & \beta_{1(-1)}^{(L)} j_1(kR) \\ \vdots & \ddots & \vdots \\ \beta_{NN}^{(1)} j_N(kR) & \cdots & \beta_{NN}^{(L)} j_N(kR) \end{bmatrix}, \tag{4.14}$$

and $\boldsymbol{X}_A(k,R) = [\boldsymbol{X}_r(k,R)^T, \boldsymbol{X}_\theta(k,R)^T, \boldsymbol{X}_\phi(k,R)^T]^T$ is a $[(Q_r+1)^2+(Q_\theta+1)^2+(Q_\phi+1)^2]$ by L matrix with

$$\boldsymbol{X}_\Psi(k,R) = \begin{bmatrix} X_{00}^{(\Psi,1)}(k,R) & \cdots & X_{00}^{(\Psi,L)}(k,R) \\ X_{1(-1)}^{(\Psi,1)}(k,R) & \cdots & X_{1(-1)}^{(\Psi,L)}(k,R) \\ \vdots & \ddots & \vdots \\ X_{Q_\Psi Q_\Psi}^{(\Psi,1)}(k,R) & \cdots & X_{Q_\Psi Q_\Psi}^{(\Psi,L)}(k,R) \end{bmatrix}. \tag{4.15}$$

The optimization problem (4.13) can be rearranged using stacked matrices as

$$\min_{\boldsymbol{W}} ||\widehat{\boldsymbol{X}}_A(k,R)\boldsymbol{W}(k) - \widehat{\boldsymbol{X}}_D(k,R)||_2^2, \tag{4.16}$$

where $\widehat{\boldsymbol{X}}_A(k,R) = [\tau\boldsymbol{\beta}_A(k,R)^T, (1-\tau)\boldsymbol{X}_A(k,R)^T]^T$, and $\widehat{\boldsymbol{X}}_D(k,R) = [\tau\boldsymbol{\alpha}_D(k,R)^T, (1-\tau)\boldsymbol{X}_D(k,R)^T]^T$. This problem can be solved using a Moore-Penrose inverse with Tikhonov regularization [119]. Although the regularization can improve the condition number of the $\widehat{\boldsymbol{X}}_A$ matrix, we still need to pay close attention to the conditioning of the matrix. The conditioning of $\widehat{\boldsymbol{X}}_A$ is determined primarily by the loudspeaker arrangement. Specifically, the loudspeaker locations should be maximally distributed over space in some sense to avoid the ill-conditioning [20].

We introduce the weighting coefficient τ to make the optimization criterion ad-

justable for different scenarios. Note that the pressure coefficients in (4.16) are weighted by the Bessel functions and therefore $\tau = 1$ corresponds to the weighted mode matching by the Bessel functions rather than the conventional pressure matching in the spherical harmonic domain. The Bessel functions in the optimization criterion can be replaced with other functions such as the Gaussian-weighted Bessel functions [38], and the proposed velocity-assisted method can also work.

4.3.3 Reproduction error

We define the normalized reproduction error over the spherical target region as

$$\epsilon(k) = \sqrt{\frac{\int_0^R \int_0^{2\pi} \int_0^{\pi} |P_d(\boldsymbol{x}, k) - P_a(\boldsymbol{x}, k)|^2 d\Omega}{\int_0^R \int_0^{2\pi} \int_0^{\pi} |P_d(\boldsymbol{x}, k)|^2 d\Omega}}, \tag{4.17}$$

where $d\Omega = r^2 \sin\theta d\theta d\phi dr$. Substituting (4.9) and (4.11) into (4.17), it can be simplified as

$$\epsilon(k) = \sqrt{\frac{\sum_{n=0}^{N} \sum_{m=-n}^{n} |\alpha_{nm}^{(d)}(k) - \sum_{\ell=1}^{L} w_l \beta_{nm}^{(\ell)}(k)|^2 \mathcal{J}_n(k, R)}{\sum_{n=0}^{N} \sum_{m=-n}^{n} |\alpha_{nm}^{(d)}(k)|^2 \mathcal{J}_n(k, R)}}, \tag{4.18}$$

where

$$\mathcal{J}_n(k, R) = \begin{cases} [2kR - \sin(2kR)]/4k^3, & \text{if } n = 0 \\ R^3[j_n^2(kR) - j_{n-1}(kR) j_{n+1}(kR)]/2, & \text{if } n > 0. \end{cases} \tag{4.19}$$

Equation (4.18) is a closed-form expression for 3D sound field reproduction error, and the proof of (4.18) is given in Appendix 4.8.2.

4.4 Simulations

In this section, we illustrate the performance of the proposed method in various simulated environments using two different loudspeaker geometries. First, the analysis of determining the weighting coefficient for the reproduction system is given. Using this weighting coefficient, the system is implemented and compared with the

Table 4.1: Loudspeaker positions of the 22 channel system

Loudspeaker No.	r [m]	θ [deg]	ϕ [deg]
1	0.97	0	0
2	1.55	51.3	0
3	1.55	51.3	45
4	1.55	51.3	90
5	1.55	51.3	135
6	1.55	51.3	180
7	1.55	51.3	225
8	1.55	51.3	270
9	1.55	51.3	315
10	1.21	90	0
11	1.21	90	22.5
12	1.21	90	45
13	1.21	90	90
14	1.21	90	135
15	1.21	90	180
16	1.21	90	225
17	1.21	90	270
18	1.21	90	315
19	1.21	90	337.5
20	1.34	115	0
21	1.34	115	45
22	1.34	115	315

conventional pressure matching in the spherical harmonic domain (or mode matching) [20, 21], in terms of the reproduced sound pressure, virtual source position, frequency, and reproduced sound direction.

4.4.1 Simulation setup

In this simulation example, the radius of the region χ is set to $R = 0.18$ m. To evaluate the proposed method in both regular and irregular loudspeaker arrangements, we simulate two different 3D loudspeaker geometries. One is the regular array with 25 loudspeakers uniformly placed on the surface of a 1 m sphere at locations determined by [118]. Another loudspeaker array is based on the cylindrical layout of 22.2 channel system proposed by NHK [144] without the two low-frequency channels, where most loudspeakers are non-uniformly placed in the upper hemisphere. The loudspeaker position of the NHK array is given in Table 4.1. The loudspeakers are all modeled as ideal point sources. Note that both the loudspeaker arrangements have an upper frequency limit for the target region, beyond which they are not capable of accurately reproducing the sound field. It usually requires more loudspeakers to support such a scenario. For the reverberant environments, we simulate a large room of size $(5, 6, 5)$ m and a small room of size $(2.5, 3, 2.4)$ m, with the same reflection coefficients of $\boldsymbol{d} = [0.8, 0.7, 0.8, 0.7, 0.78, 0.81]$. To show the rooms in terms of reverberant time, we also calculate the T_{60} of the rooms using the Sabine formula [145]. The T_{60} is 0.35 s for the large room and it is 0.17 s for the small room. We note that, in the following simulations, we truncate the image depth of the rooms to $\boldsymbol{R}_{\text{depth}} = (4, 4, 4)$ when calculating the reverberation in the rooms for simplicity. A virtual point source with frequency of 1200 Hz at $(2, \pi/9, 2\pi/3)$ is the desired sound source throughout the simulations, except the evaluation in terms of the virtual source position and frequency. Sound speed is $c = 343$ m/s and air density is $\rho_0 = 1.29$ kg/m^3. Therefore, the pressure truncation order is $N = 6$. The velocity truncation order is set to $Q_\theta = Q_\phi = 2N$.

The reproduction error over the whole spherical target region is calculate by (4.18) throughout the paper. To show the difference between the reproduced sound field and the desired sound field intuitively, we define the mismatch between them

at any point $\boldsymbol{x}$ as

$$\varepsilon(\boldsymbol{x}, k) = \frac{|P_d(\boldsymbol{x}, k) - P_a(\boldsymbol{x}, k)|^2}{|P_d(\boldsymbol{x}, k)|^2} \times 100(\%). \tag{4.20}$$

From the perspective of realistic perception of source location, to reproduce the original sound, it requires to ensure the reproduction of the direction of the sound. A convenient measure for the direction of energy propagation in an arbitrary sound field is provided by the active sound intensity, which is defined for steady state fields as [109]

$$\boldsymbol{I}(\boldsymbol{x}, k) = \frac{1}{2}\mathrm{Re}\{P^*(\boldsymbol{x}, k)\boldsymbol{V}(\boldsymbol{x}, k)\}. \tag{4.21}$$

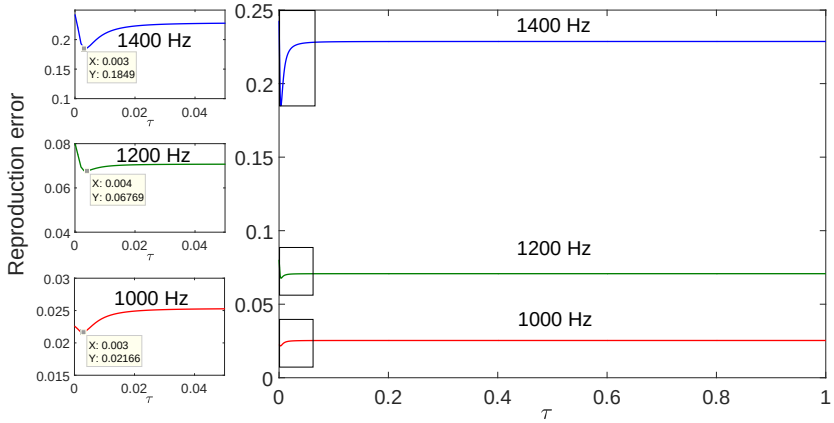
Note that the active sound intensity is a vector containing the information of direction, and it has been shown as a good predictor of localization perception [50, 71, 146]. To show the difference in angles between two vectors, we define the intensity direction error ζ at $\boldsymbol{x}$ as

$$\zeta(\boldsymbol{x}, k) = \cos^{-1}[\widehat{\boldsymbol{I}}_a(\boldsymbol{x}, k) \cdot \widehat{\boldsymbol{I}}_d(\boldsymbol{x}, k)]/\pi \times 100\%, \tag{4.22}$$

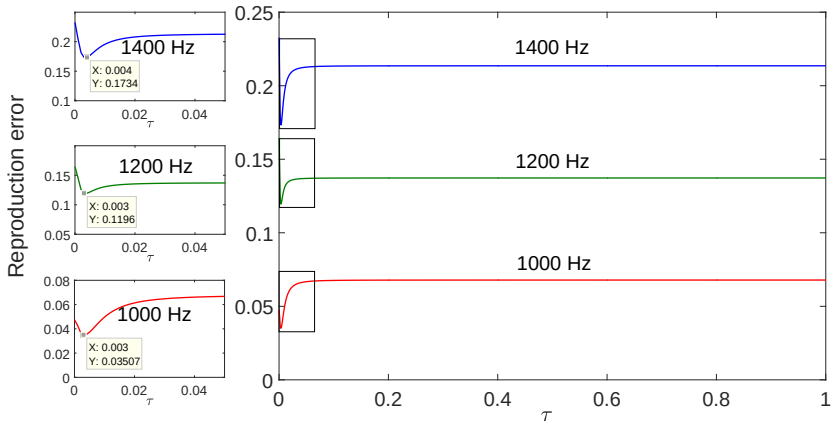
where $\widehat{\boldsymbol{I}}_a(\boldsymbol{x}, k) = \boldsymbol{I}_a(\boldsymbol{x}, k)/|\boldsymbol{I}_a(\boldsymbol{x}, k)|$ and $\widehat{\boldsymbol{I}}_d(\boldsymbol{x}, k) = \boldsymbol{I}_d(\boldsymbol{x}, k)/|\boldsymbol{I}_d(\boldsymbol{x}, k)|$ are the unit active intensity vector of the reproduced sound field and the unit active intensity vector of the desired sound field, respectively.

4.4.2 Determination of the weighting coefficient

Before evaluating the overall performance of the proposed method, we first investigate how the weighting coefficient τ impacts the reproduction system. As shown in Section 4.3.2, the weighting coefficient controls the relative weight of sound pressure and particle velocity. Note that the larger the weighting coefficient is, the less weight it assigns to the particle velocity. We calculate the 3D reproduction error with respect to τ for various frequencies in both the free field and the reverberant environment (the large room) using the 22 channel loudspeaker array, which is shown in Fig. 4.2. The curves in the rectangular boxes are enlarged and shown on the left correspondingly. We observe that the reproduction system has minimum error when τ is 0.003 approximately for all the evaluated frequencies in both the



(a)



(b)

Figure 4.2: The impact of the weighting coefficient on the reproduction system for various frequencies in (a) the free field and (b) the reverberant environment (the large room) in the 22 channel loudspeaker layout. The sub-figures on the left are the enlarged curves in the rectangular boxes correspondingly.

free field and the reverberant environment. For the values more than 0.1, τ has little influence on the reproduction system, which means more weight should be assigned to the particle velocity to make it take effect because of the significant difference between the values of sound pressure and particle velocity. This result

also means the velocity-assisted method with an appropriate weighting coefficient has better performance than the weighted mode matching by the Bessel functions (i.e., $\tau = 1$). We also notice that the reproduction system has more error when the frequency increases, because the system order goes up. The optimization criterion with $\tau = 0.003$ also works well for the 25 channel spherical array and the small room, therefore, we use the weighting coefficient $\tau = 0.003$ for the reproduction system in the following analysis. Note that this value is optimal only for the simulation scenarios in this paper. One should recalculate the optimal weighting coefficient if the scenario totally changes because the difference between the values of pressure coefficients and velocity coefficients may also change, which may result in a different optimal value.

4.4.3 Simulation results in free field

The simulations in this section study the overall performance of the proposed method in free field for the two loudspeaker arrangements.

Reproduced sound pressure

We now evaluate the performance of the system on the reproduction of sound pressure. The observation plane is the plane of $z = 0$. Figure 4.3 and Fig. 4.4 show the desired sound field and the reconstructed sound fields controlled by the pressure matching method and the velocity-assisted method in both loudspeaker layouts. The black circle denotes the target region. As expected, the velocity-assisted method provides a better sound field reproduction than the pressure matching method within the target region. The mismatches between the desired sound field and the reconstructed sound field for both methods are also shown in Fig. 4.3 and Fig. 4.4, which are calculated by (4.20). We observe that the pressure matching method fails to reproduce the desired sound field for the whole target region in the 22 channel system. When the loudspeaker array changes to the 25 channel regular spherical array, the performance of the pressure matching method has a significant improvement, however, it is still not as good as the velocity-assisted method. In this case, it requires at least $(N + 1)^2 = 49$ uniformly distributed loudspeakers to accurately reproduce the desired sound field in the target region for the pressure

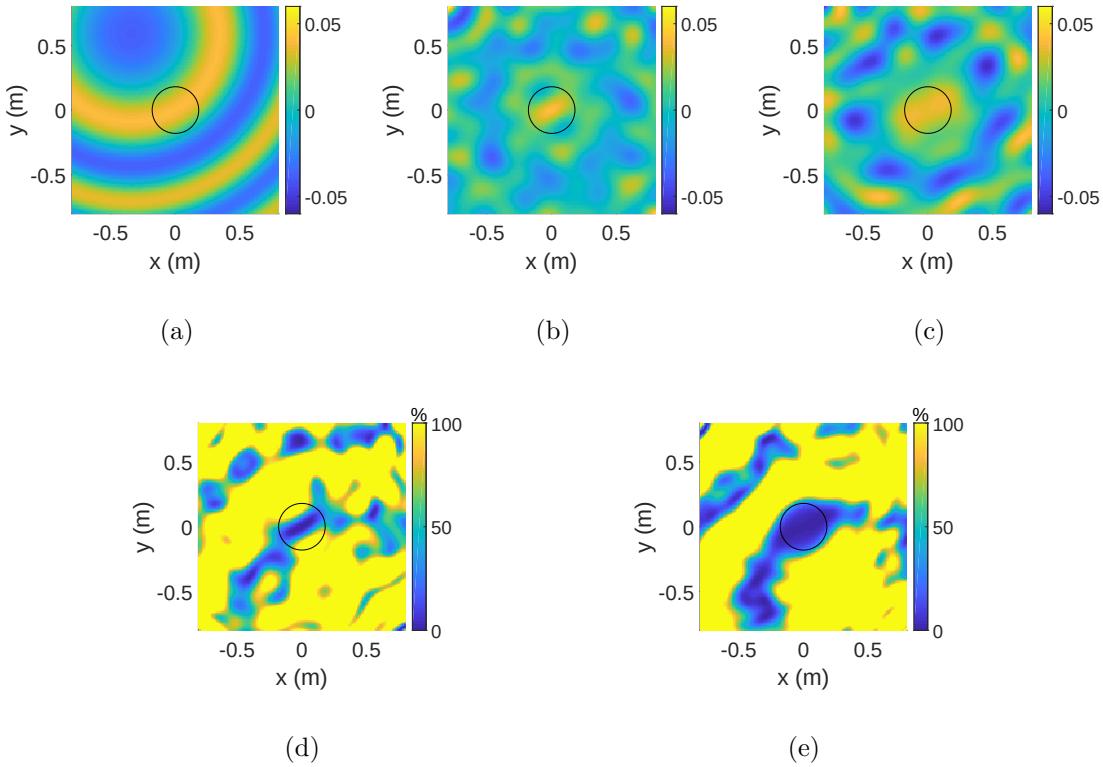


Figure 4.3: Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for the 22 channel system. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by the pressure matching method (PMM); (c) Reproduced sound field controlled by the velocity-assisted method (VAM); (d) Difference field controlled by PMM; (e) Difference field controlled by VAM.

matching method, whereas there are only 22 non-uniformly placed loudspeakers or 25 regular loudspeakers available. However, the velocity-assisted method reproduces the desired sound field in the target region with much less error than the pressure matching method verifying that the proposed method can overcome this limitation.

Virtual source position

To investigate the overall performance of the proposed method with respect to the virtual source position, we keep the distance (2 m) from the virtual source

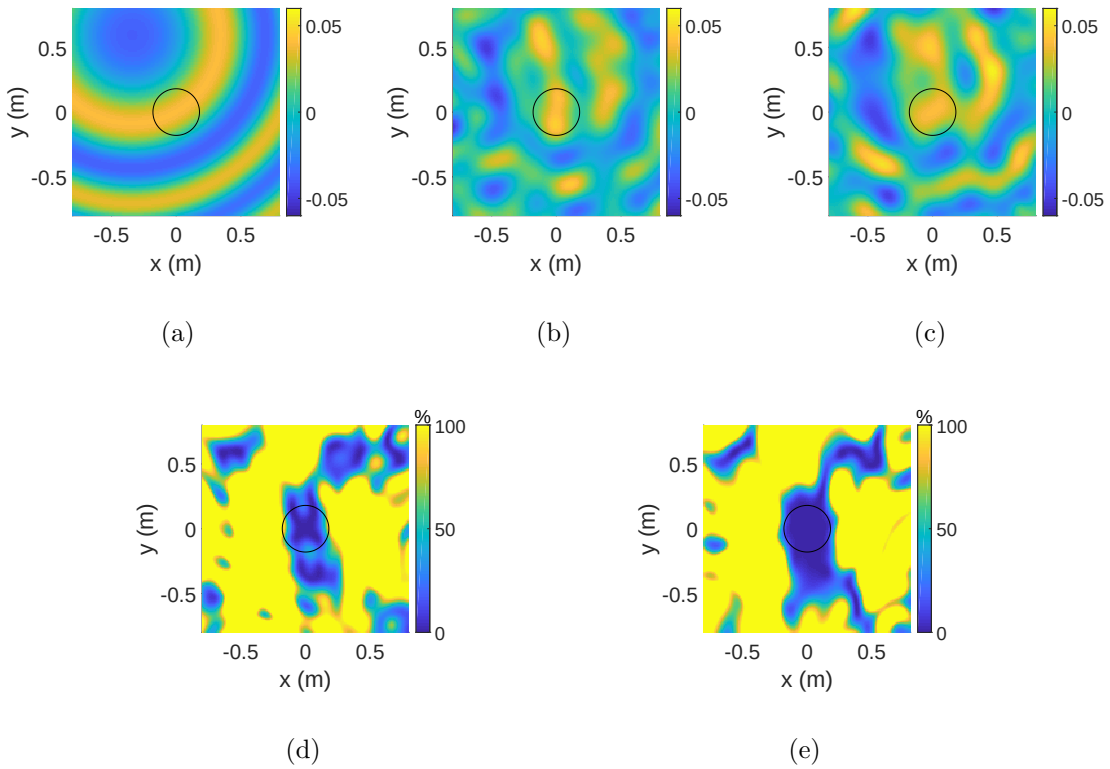
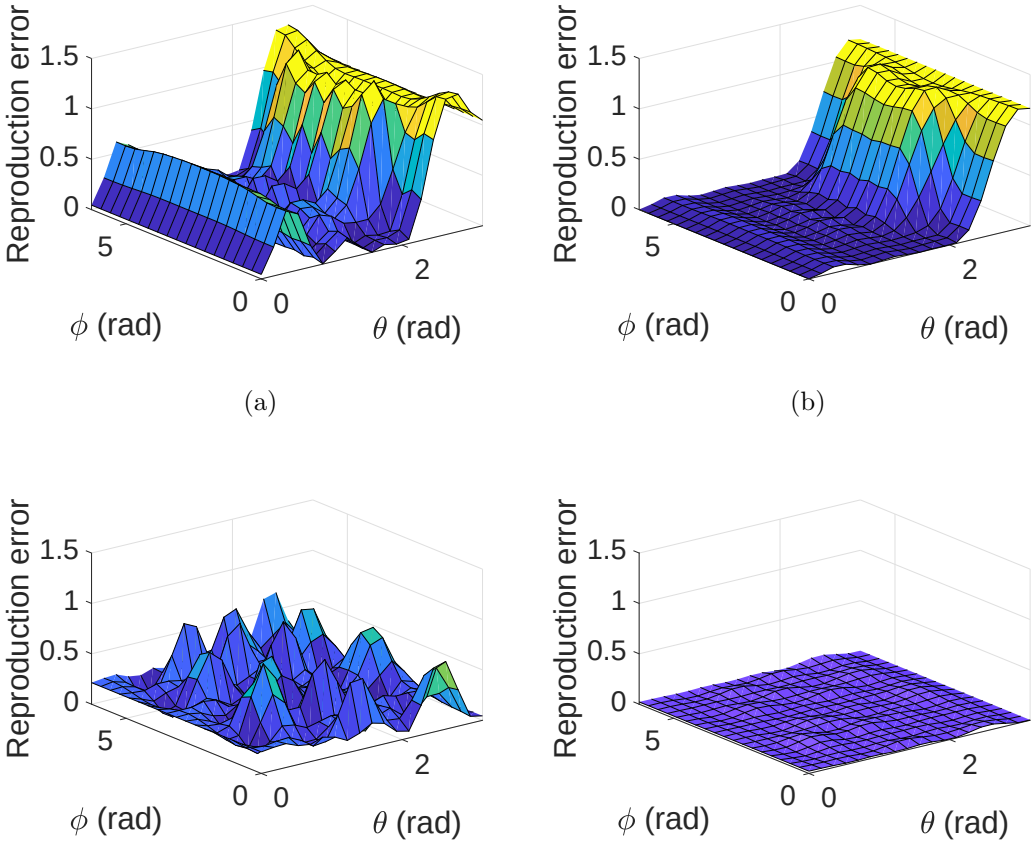


Figure 4.4: Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for the 25 channel system. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by PMM; (c) Reproduced sound field controlled by VAM; (d) Difference field controlled by PMM; (e) Difference field controlled by VAM.

to the origin fixed and calculate the 3D reproduction error in terms of θ and ϕ. The three-dimensional surface plots showing the results are given in Fig. 4.5. By comparing the error plots for the two loudspeaker arrays, we notice that the change of the reproduction error is regular with the virtual source position change for the 25 channel regular loudspeaker array, whereas for the 22 channel loudspeaker array the error fluctuates dramatically as the virtual source position changes due to the non-uniformly distributed loudspeakers. For the 22 channel irregular loudspeaker array, the velocity-assisted method performs better than the pressure matching method when the virtual source is located in the upper hemisphere ($0 \leq \theta \leq \pi/2$), however, the performance of both methods degrades for the virtual sources from



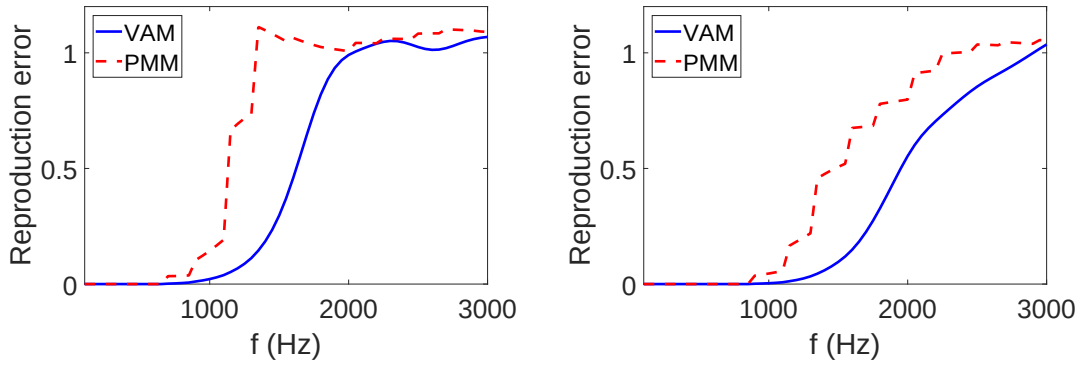
(a) (b) (c) (d)

Figure 4.5: 3D reproduction error with respect to the virtual source positions controlled by (a, c) PMM and (b, d) VAM for (a, b) the 22 channel system and (c, d) the 25 channel system. The source frequency is 1200 Hz.

the lower hemisphere ($\pi/2 < \theta \leq \pi$) because there are only 3 out of 22 loudspeakers located below x-y plane. In contrast, the 3D reproduction error is less than 0.1 for all the virtual source positions using the 25 channel regular loudspeaker array, which is much better than the pressure matching method.

Frequency

To analyze the broadband performance of the proposed method, we also calculate the 3D reproduction error with respect to the change of frequency from 100 Hz to



(a) (b)

Figure 4.6: 3D reproduction error with respect to the change of frequency controlled by both methods for (a) the 22 channel system and (b) the 25 channel system.

3000 Hz, which is shown in Fig. 4.6. From Fig. 4.6, we observe that the error remains very low for both methods in the 22 channel system at frequencies less than 650 Hz approximately, where the system order N is smaller than 3. The 22 loudspeakers are enough to accurately reproduce the desired sound field in the target region for both methods. As the frequency increases (N also increases), the 22 loudspeakers are not enough any more and the error increases dramatically for the pressure matching method. However, the error builds up more gradually for the velocity-assisted method. For example, the error due to the pressure matching method is more than 1, whereas the error due to the velocity-assisted method is only around 0.3 when the frequency increases to 1500 Hz. Therefore, given a loudspeaker array and a target region, the velocity-assisted method can reproduce the desired sound field with a wider frequency range than the pressure matching method. We note that, when the frequency continues to increase to 3000 Hz, the velocity-assisted method loses this advantage and its error also increases to around 1 because of the extreme scarcity of loudspeakers. The overall trend of error curve in the 25 channel system is same as that in the 22 channel system for both methods, however, the former is lower than the latter for all evaluated frequencies due to the increase of the number of loudspeakers. Besides, the frequency range in which the velocity-assisted method outperforms the pressure matching is around 1-2 kHz in

Fig. 4.6(a), whereas the frequency range is around 1-3 kHz in Fig. 4.6(b), which indicates that the proposed method has advantages over a wider frequency range as the number of loudspeakers increases. This point also infers the limitation of the proposed method that it requires a large number of loudspeakers to reproduce the sound with very high frequency in a large area.

Intensity direction

As we mentioned in Section 4.4.1, it is clear that if we are concerned to create a realistic perception of the original sound, it requires to ensure the reproduction of the direction of travel of the sound. Therefore, we calculate the intensity direction error on plane $z = 0$ using (4.22), which is shown in Fig. 4.7. In the 22 channel system, the intensity direction error of the velocity-assisted method is less than 20% within the region of interest, whereas it is around 40% for the pressure matching method. When the loudspeaker array becomes regular, the intensity direction error of the pressure matching method reduces, however, it is still larger than the velocity-assisted method. This analysis shows that the velocity-assisted method can reproduce the direction of original sound inside the region of interest in free field.

From the above analysis, we conclude that the velocity-assisted method has better overall performance than the pressure matching method in free field, especially for a non-uniformly distributed loudspeaker array with a limited number of loudspeakers.

4.4.4 Simulation results in reverberant rooms

The following simulations investigate the overall performance of the proposed method in the two different rooms. Here we only consider the 22 channel non-uniformly distributed loudspeaker array.

Reproduced sound pressure

We also first evaluate the reproduction of sound pressure within the region of interest, and the results in the different sizes of rooms are given in Fig. 4.8. The

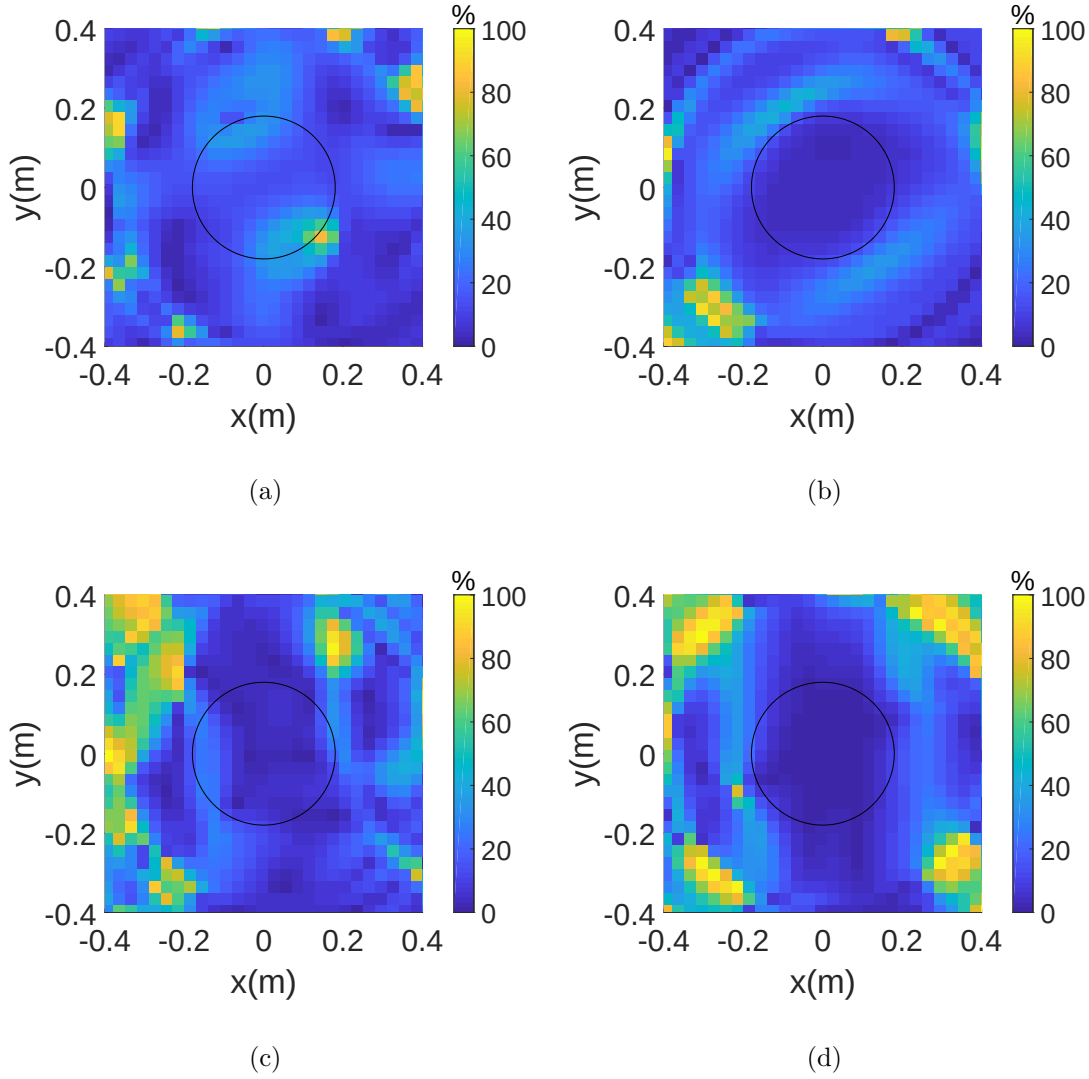


Figure 4.7: The intensity direction error on plane $z = 0$ controlled by (a, c) PMM and (b, d) VAM for (a, b) the 22 channel system and (c, d) the 25 channel system. The source frequency is 1200 Hz.

difference between the desired sound field and the reconstructed sound field is shown in Fig. 4.9. It shows that the velocity-assisted method provides more accurate pressure reproduction in both the reverberant rooms. By comparing with the results in the free field, we observe that for this particular virtual source the performance in the reverberant environment is a little worse than that in the free

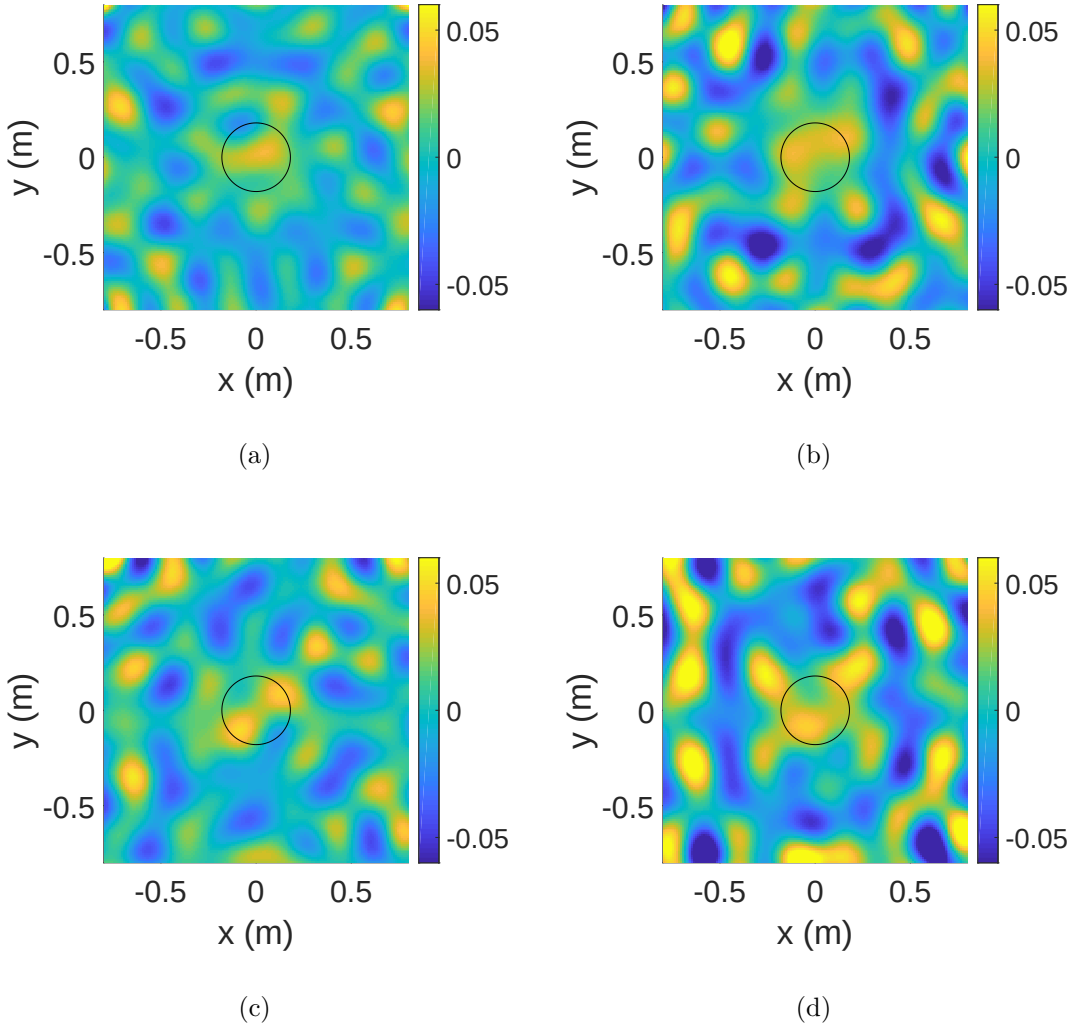


Figure 4.8: Reconstructed sound field on the observation plane controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz.

field. However, this result is not consistent for all the virtual source positions. The reproduction performance with respect to the position of the virtual source in the reverberant rooms is discussed next.

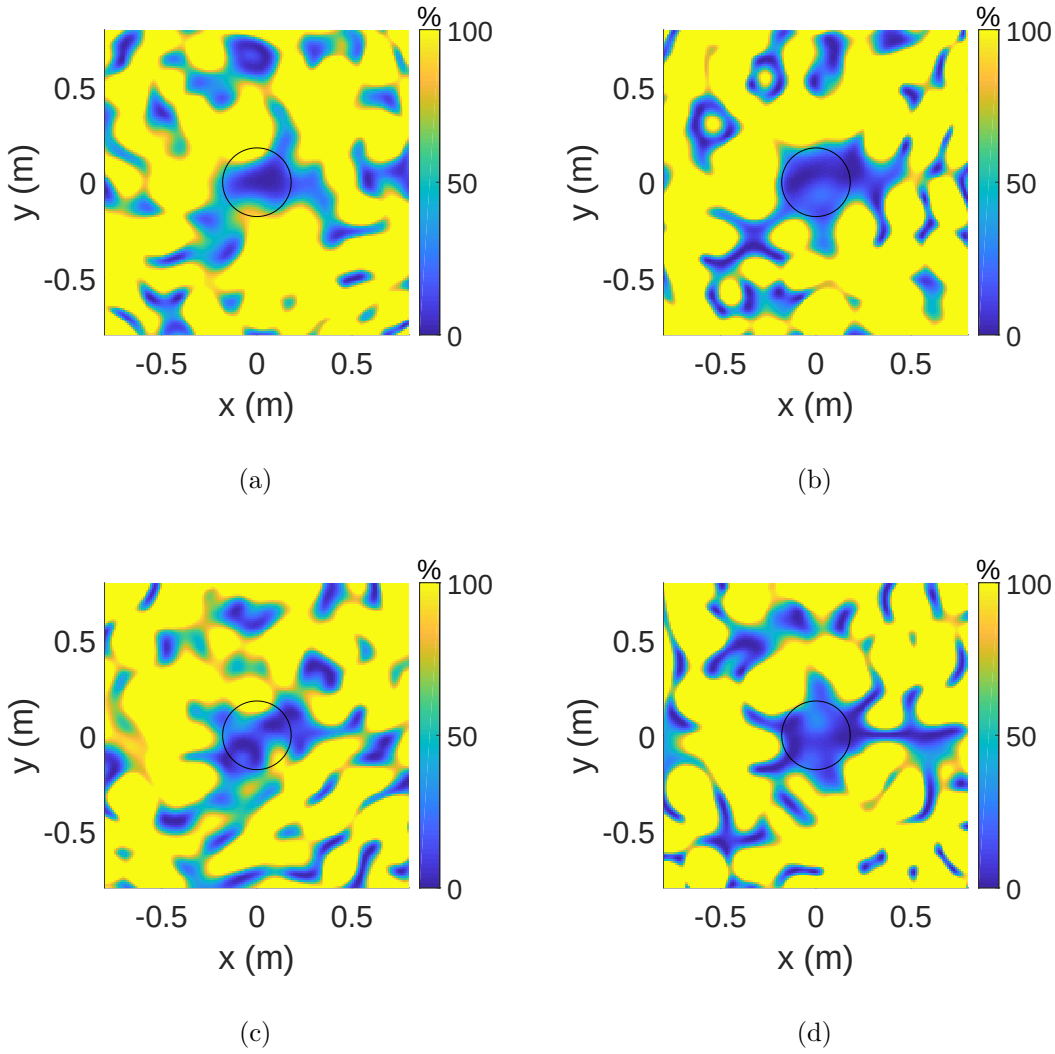


Figure 4.9: Difference between the desired sound field and the reconstructed sound field controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz.

Virtual source position

The same virtual source positions as in Section 4.4.3 are evaluated for the reverberant environments. The results are given in Fig. 4.10. By comparing the results in the free field and the reverberant environment, we notice that the performance in the free field is parallel to the performance in the reverberant environment for

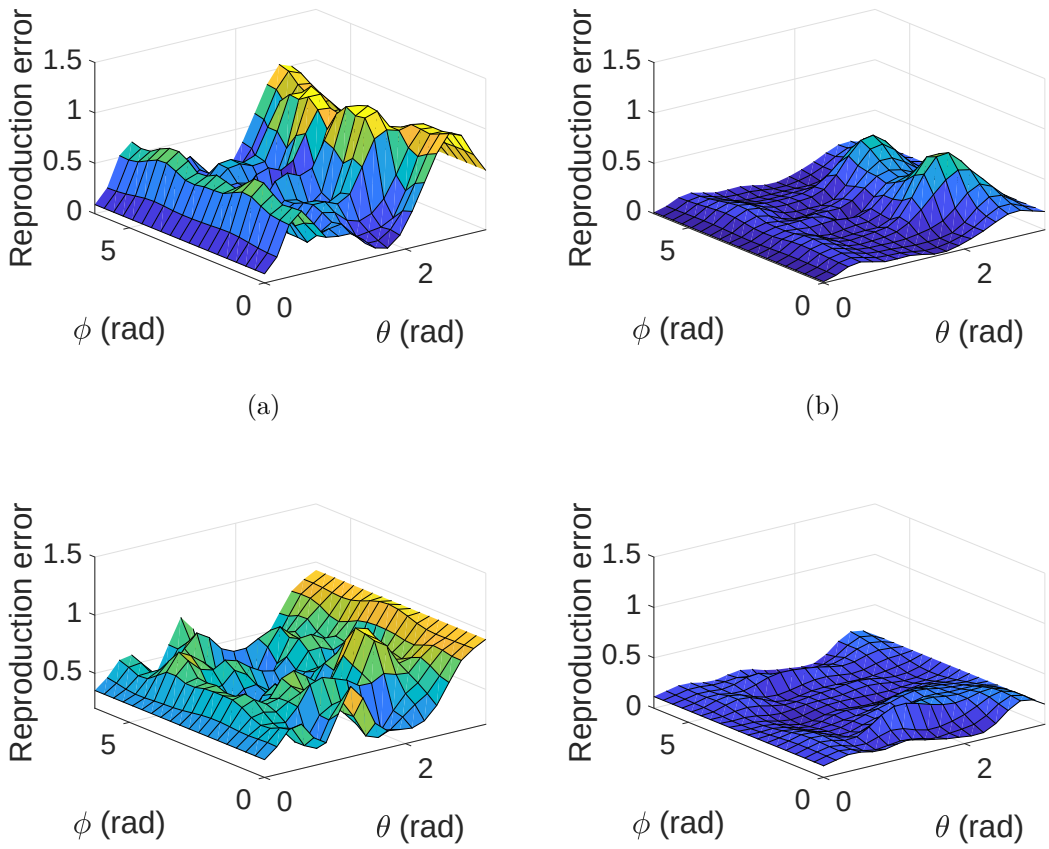
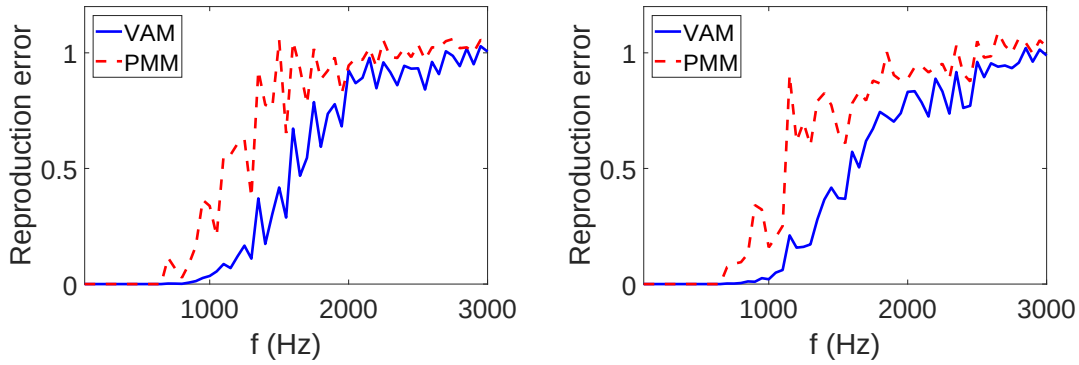


Figure 4.10: 3D reproduction error with respect to the virtual source positions controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz.

the virtual sources from the upper hemisphere, where there are more loudspeakers. However, for the virtual sources from the lower hemisphere where there are few loudspeakers, the sound pressure reproduction in the reverberant environment has less error than that in the free field. This is because the reflections (image sources) in the reverberant room make up for the insufficiency of loudspeakers in the lower hemisphere in this case. Compared with the pressure matching method, the 3D reproduction error of the velocity-assisted method is less than 0.5 for all the virtual source positions in the reverberant environment, which is much better than



(a) (b)

Figure 4.11: 3D reproduction error with respect to the change of frequency controlled by both methods in (a) the large room and (b) the small room in case of the 22 channel system.

the pressure matching method. We also note that the velocity-assisted method has more improvement from the remedy due to the reflections in the reverberant room than the pressure matching method. This is because not only sound pressure but also particle velocity is exploited for the reflections that are incorporated in the optimization criterion of the velocity-assisted method.

Frequency

Fig. 4.11 shows the results of the two methods with respect to the change of frequency in the two different sizes of rooms. The performance in the large room is similar to that in the small room for both methods. Similar to the results in the free field, the error starts to increase at the frequency of 650 Hz, and reaches 1 at the frequency of 2000 Hz. Therefore, the overall trend of error curve in the free field is same as that in the reverberant environment for both methods, however, the former is more smooth than the latter due to the more complicated environment in a reverberant room.

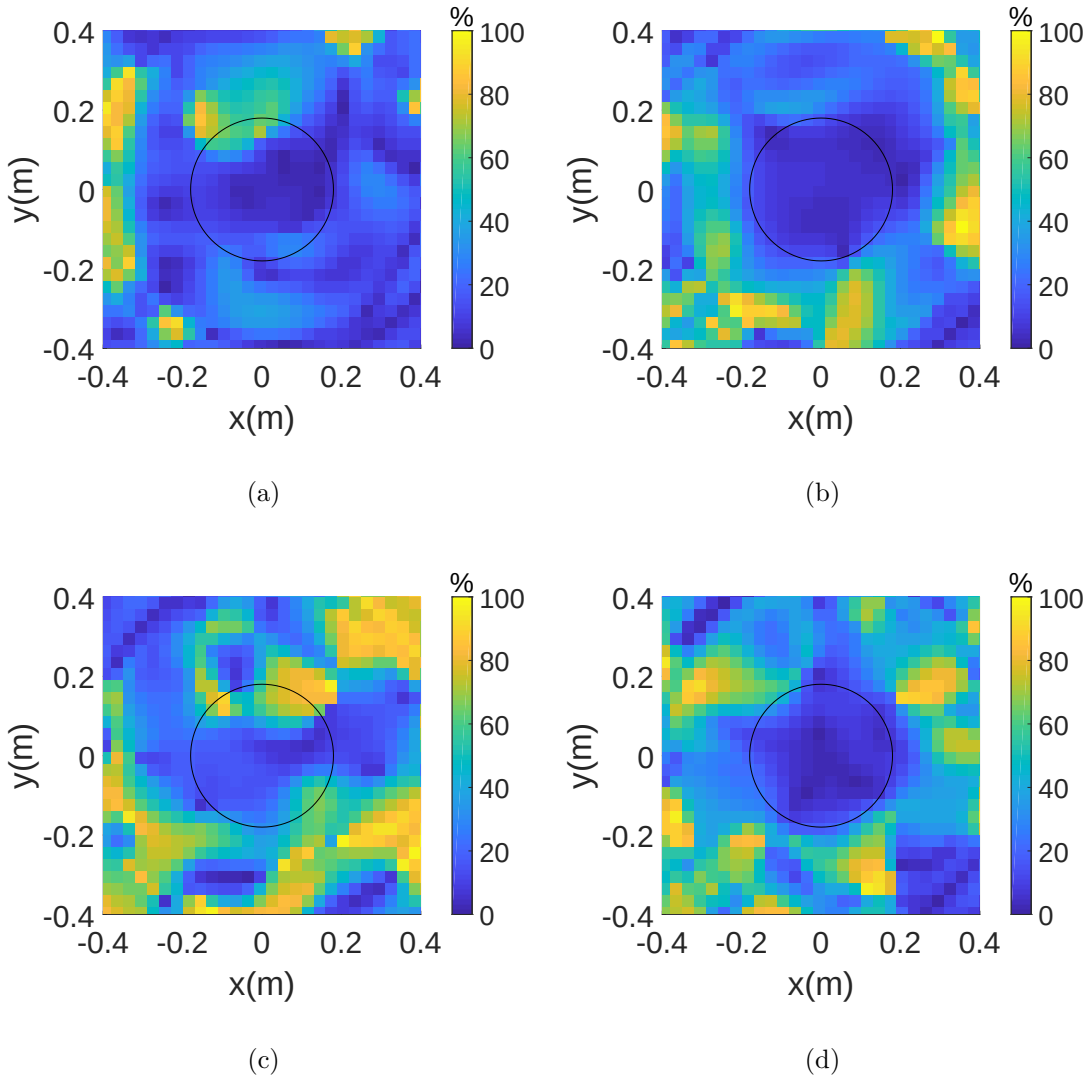


Figure 4.12: The intensity direction error on plane $z = 0$ controlled by (a, c) PMM and (b, d) VAM in (a, b) the large room and (c, d) the small room in case of the 22 channel system. The source frequency is 1200 Hz.

Intensity direction

The intensity direction error is also calculated for both methods in the reverberant rooms, which is given in Fig. 4.12. From Fig. 4.12, we show that the pressure matching method cannot guarantee accurate reproduction of sound direction inside the whole target region in the reverberant environment, whereas the intensity

direction error of the velocity-assisted method is less than 20% for the whole target region. Although the performance of the pressure matching method in the large room is slightly better than that in the small room, it is at the nearly same level for the performance of the velocity-assisted method in the two rooms. Therefore, we prove that the velocity-assisted method can reconstruct the original sound direction in various sizes of rooms. According to the above analysis, we conclude that the velocity-assisted method performs better than the pressure matching method in the reverberant environment for all the assessed aspects. The size of the reverberant room has little effect on the performance of the reproduction system.

The evaluation in this section has shown that the proposed method is better than the pressure matching method for overall performance in both free fields and reverberant environments. However, we should note that the computational complexity of the proposed method is higher than that of the pressure matching method due to the introduction of particle velocity.

4.5 Experimental Verification

In this section, we use the impulse response measurements in a real-world room to validate the proposed method. The impulse responses were measured in an office space at the Australian National University with dimensions of $(3.54, 4.06, 2.70)$ m. For simplicity, we consider a circular loudspeaker array with radius of 1 m, consisting of six non-uniformly placed loudspeakers. We use such a circular loudspeaker array as it has a significantly reduced implementation complexity compared to the loudspeaker geometries in Section 4.4.1. However, the implementation steps for the proposed method remain unchanged. We set the center of the circular array as the origin, and the loudspeakers are positioned on the horizontal plane ($\theta = 90°$) with $\phi = 48°$, $92°$, $148°$, $184°$, $256°$, and $336°$, respectively. An EigenMike [75] located at the origin is used for sound field recording, where the EigenMike is a 32 channel rigid spherical microphone array with radius of 0.042 m. A 2 second linear sweep signal is fed into each loudspeaker and the resulting sound field is recorded by the EigenMike. The impulse response between each loudspeaker and each EigenMike sensor can be obtained from the recordings, therefore we have 32×6 impulse responses measurements. The coefficients of the point-to-region transfer function

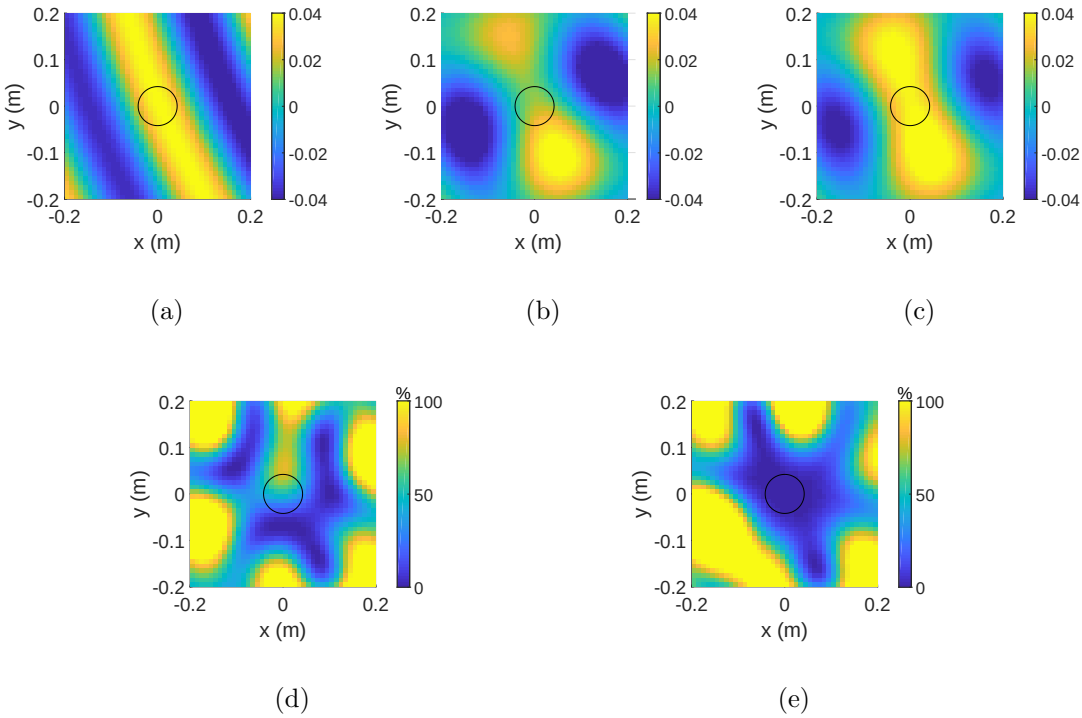


Figure 4.13: Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for source #1. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by PMM; (c) Reproduced sound field controlled by VAM; (d) Difference field controlled by PMM; (e) Difference field controlled by VAM.

between each loudspeaker and the receiver region (i.e., the EigenMike area) are estimated from the corresponding 32 channel measurements by

$$\widehat{\beta}_{nm}^{(\ell)}(k) = \frac{1}{b_n(kR_E)} \sum_{q_M=1}^{Q_M} \widehat{H}_\ell(k, \boldsymbol{x}_{q_M}) Y_{nm}^*(\theta_{q_M}, \phi_{q_M}), \tag{4.23}$$

where $\widehat{H}_\ell(k, \boldsymbol{x}_{q_M})$ is the measurement between the ℓth loudspeaker and the q_M^{th} EigenMike sensor located at $\boldsymbol{x}_{q_M} = (r_{q_M}, \theta_{q_M}, \phi_{q_M})$, $Q_M = 32$ is the number of EigenMike sensors, $R_E = 0.042$ m is the radius of the EigenMike, and

$$b_n(kR_E) = j_n(kR_E) - \frac{j_n'(kR_E)}{h_n'(kR_E)} h_n(kR_E). \tag{4.24}$$

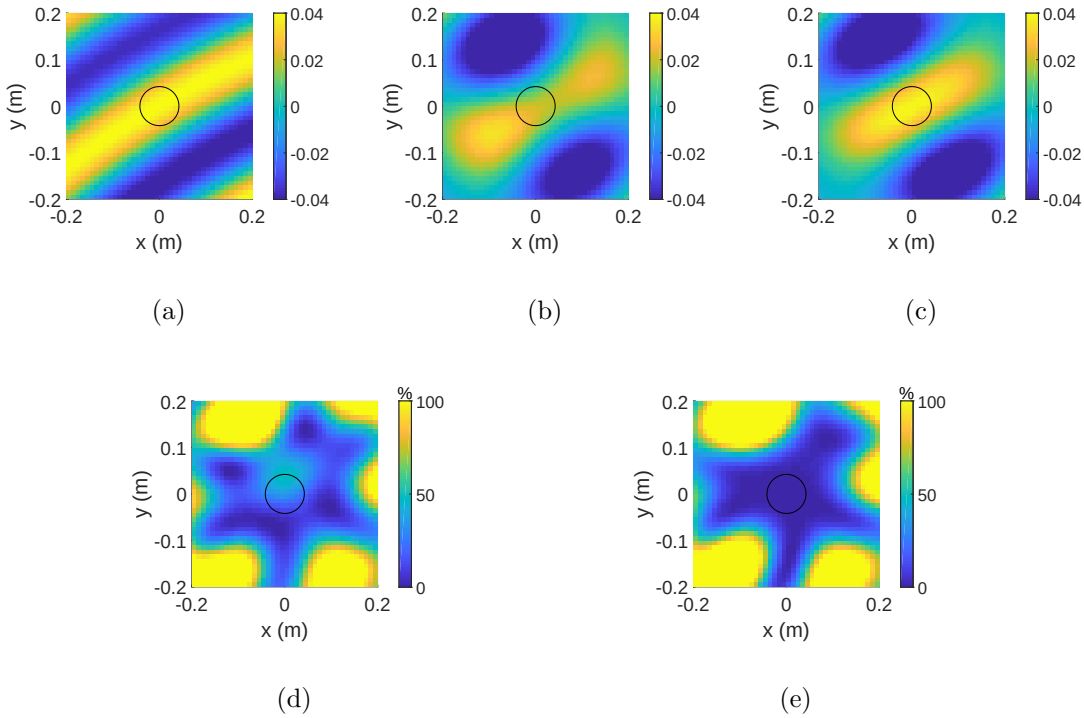
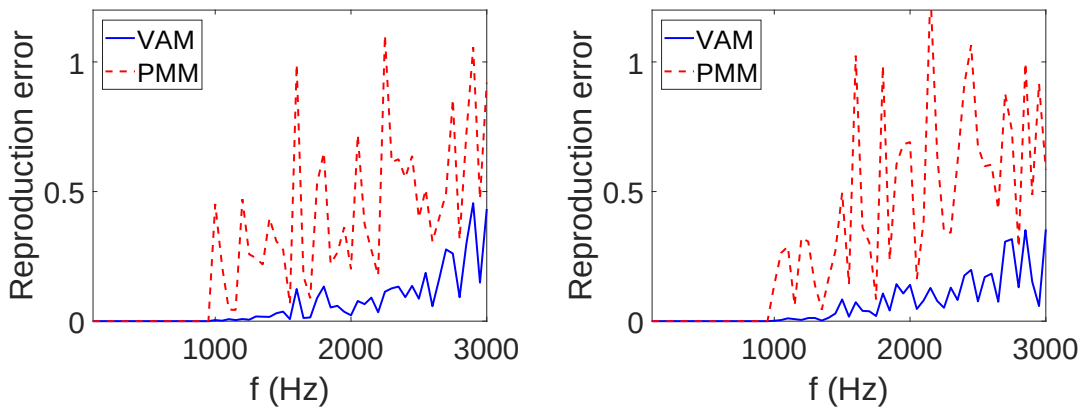


Figure 4.14: Desired sound field, reconstructed sound field, and the difference between them on the observation plane controlled by both methods for source #2. The source frequency is 1200 Hz. (a) Desired sound field; (b) Reproduced sound field controlled by PMM; (c) Reproduced sound field controlled by VAM; (d) Difference field controlled by PMM; (e) Difference field controlled by VAM.

We assume the desired sound fields are given, and they are produced by a virtual point source located at $(2, \pi/2, \pi/8)$ (source #1) and $(2, \pi/2, 5\pi/3)$ (source #2), respectively, with frequency of 1200 Hz. We calculate the loudspeaker weights using (4.16) by replacing $\{\beta_{nm}^{(\ell)}(k)\}$ with $\{\widehat{\beta}_{nm}^{(\ell)}(k)\}$, and plot the sound fields reproduced by the circular array controlled by the velocity-assisted method. The plots including the results of both methods are given in Fig. 4.13 and Fig. 4.14, where the difference between the reconstructed sound fields and the desired sound fields for both methods are also shown. Note that in this case the target region is a spherical region with radius of 0.042 m, which is same as the size of the EigenMike. With the observation of the error fields between the reproduced and desired pressure fields, the velocity-assisted method can provide accurate pressure reproduction inside the



(a) (b)

Figure 4.15: 3D reproduction error with respect to the change of frequency controlled by both methods for (a) source #1 and (b) source #2.

target region, whereas the pressure matching method has around 50% error for both virtual sources.

The EigenMike can record spatial sound field up to 4th order, therefore, the $\{\widehat{\beta}_{nm}^{(\ell)}(k)\}$ obtained from the measurements is also up to 4th order. Given the radius of 0.042 m, the frequency limit is 3800 Hz approximately. We process the measurements frequency bin by frequency bin, ranging from 100 Hz to 3000 Hz. The results showing the reproduction error with respect to different frequency bins for both methods are given in Fig. 4.15. The six channel loudspeaker array can provide accurate pressure reproduction for the first order system (i.e., when the frequency is less than 950 Hz) using both methods. When the frequency goes up, the pressure matching method results in a large error for most of the frequency bins, however, the error of the velocity-assisted method remains low for all the evaluated frequency bins. As shown in Fig. 4.16, the intensity direction error of the velocity-assisted method within the target region is also less than that of the pressure matching method, which is consistent with the simulation results in Section 4.4. Consequently, we conclude that the proposed method also has good performance for a real-world environment.

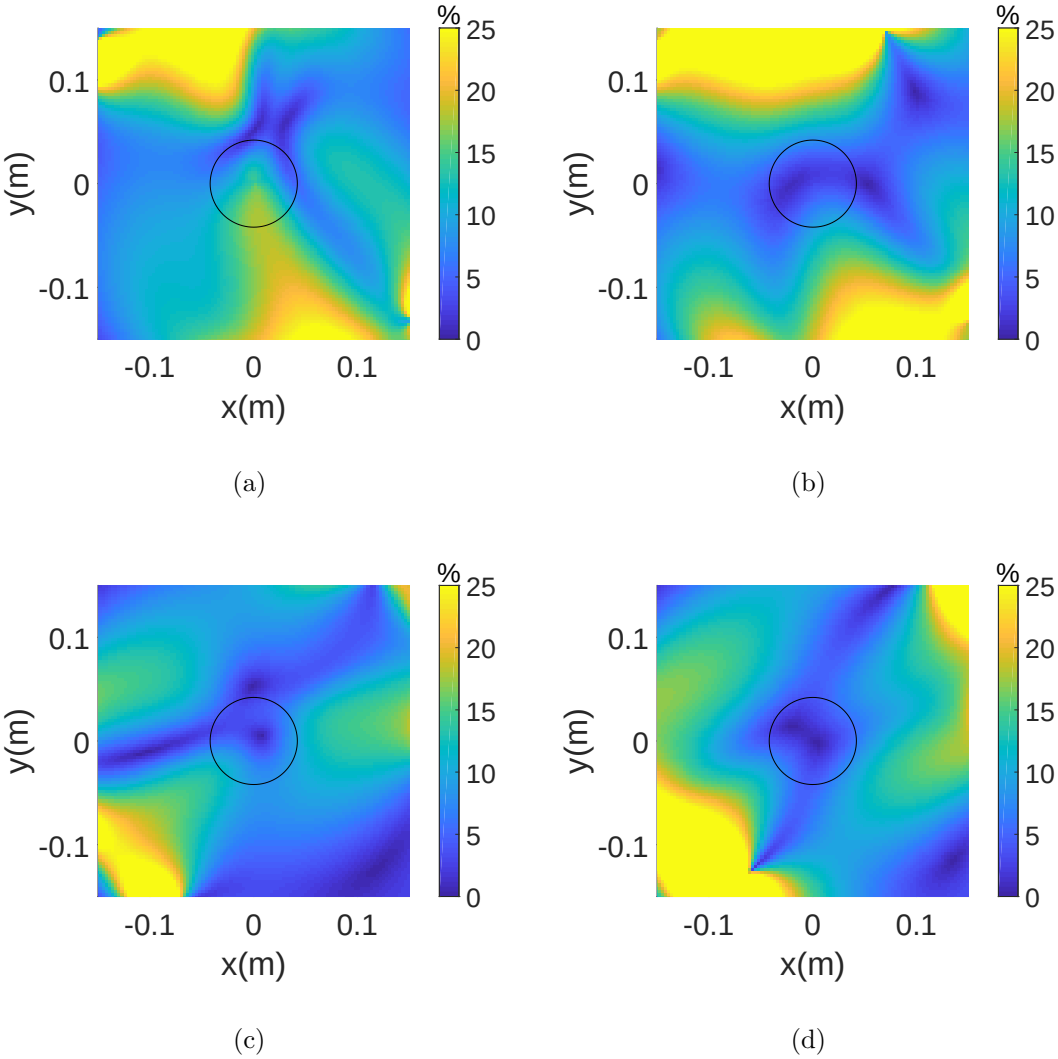


Figure 4.16: The intensity direction error on plane $z = 0$ controlled by (a, c) PMM and (b, d) VAM for (a, b) source #1 and (c, d) source #2. The source frequency is 1200 Hz.

4.6 Summary and Contributions

In this chapter, we have proposed a 3D sound field reproduction method based on particle velocity in the spherical harmonic domain which is appropriate for non-uniform loudspeaker geometries. The chapter offers the following major contributions:

- We built an error cost function and designed loudspeaker driving signals by controlling the continuous particle velocity and sound pressure on the surrounding contour.

- We evaluated the proposed method in both a free field and a reverberant room by comparing it with the conventional pressure matching in the spherical harmonic domain. We demonstrated that the proposed method can provide more accurate sound field reproduction with a wider frequency range in the target region using a limited number of loudspeakers, as verified by simulation results. Nevertheless, we should note that there still exist an upper reproduction frequency limit for the proposed method with the given loudspeaker array. It usually requires more loudspeakers to increase this frequency limit.

- We proved that the proposed method can reconstruct the original direction of the sound in various environments. The simulations also revealed that the reflections in the reverberant room can compensate for the non-uniformity of the loudspeaker array and thus improve the reproduction performance.

- We carried out an objective experiment with the impulse response measurements of a real-world room to validate the proposed method. The results suggested that the proposed method can also provide good performance in a real-world environment.

4.7 Related Publications

Much of this chapter's work has been published in the following journal paper.

- **H. Zuo**, T. D. Abhayapala, and P. N. Samarasinghe, "Particle velocity assisted three dimensional sound field reproduction using a modal-domain approach," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 2119-2133, 2020.

4.8 Appendix

4.8.1 Proof of equation (4.7)

We denote the size of the shoebox room as (L_x, L_y, L_z) for length, width and height. The room origin O_{room} is set at the left-front-bottom corner of the room, and the origin O is located at $\boldsymbol{x}_{qo} = (x_{qo}, y_{qo}, z_{qo}) = (L_x/2, L_y/2, L_z/2)$ with respect to O_{room}. Therefore, the ℓth loudspeaker is located at $\boldsymbol{x}_{\ell o} = \boldsymbol{x}_\ell + \boldsymbol{x}_{qo}$ with respect to O_{room}. Reverberant characteristics are modeled with the reflection coefficients of the wall surface, denoted as $\boldsymbol{d} = (d_{x1}, d_{x2}, d_{y1}, d_{y2}, d_{z1}, d_{z2})$. According to the spherical harmonics based generalized image source method [143], the point-to-region transfer function in the reverberant environment is given by

$$H_\ell^{\text{rvb}}(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{v=0}^{V} \sum_{u=-v}^{v} \gamma_{vu}^{(\ell)}(k) \alpha_{nm}^{vu}(k) j_n(kr) Y_{nm}(\theta, \phi), \tag{4.25}$$

where $V = \lceil keR_s/2 \rceil$ is the source region truncation order, R_s is the radius of the source region, $\gamma_{vu}^{(\ell)}(k) = ikj_v(kR_s)Y_{vu}^*(\theta_s, \phi_s)$ are the outgoing sound field coefficients due to the ℓth loudspeaker at $\boldsymbol{x}_s = (R_s, \theta_s, \phi_s)$ with respect to the origin of the source region O_s, and $\alpha_{nm}^{vu}(k)$ are coupling coefficients, defined by [147]

$$\begin{aligned} \alpha_{nm}^{vu}(k) = &\sum_{\boldsymbol{p}=0}^{1} \sum_{\boldsymbol{r}=-\infty}^{\infty} d_{x1}^{|r_1-p_1|} d_{x2}^{|r_1|} d_{y1}^{|r_2-p_2|} d_{y2}^{|r_2|} d_{z1}^{|r_3-p_3|} d_{z2}^{|r_3|} \\ &\times (-1)^{(p_2+p_3)u+p_3 v} S_{vn}^{((-1)^{p_1+p_2}u)m}(\boldsymbol{x}_{\boldsymbol{p}} + \boldsymbol{x}_{\boldsymbol{r}}), \end{aligned} \tag{4.26}$$

where

$$S_{vn}^{um}(\boldsymbol{y}) = 4\pi i^{n-v} \sum_{l=0}^{v+n} i^l (-1)^{2u-m} h_l(kr_y) Y_{l(m-u)}^*(\theta_y, \phi_y) W_1 W_2 \xi, \tag{4.27}$$

$W_1 = \begin{pmatrix} v & n & l \\ 0 & 0 & 0 \end{pmatrix}$, $W_2 = \begin{pmatrix} v & n & l \\ u & -m & m-u \end{pmatrix}$, $\xi = \sqrt{(2v+1)(2n+1)(2l+1)/4\pi}$, $\boldsymbol{x}_{\boldsymbol{p}} = (x_{qo} - x_{\ell o} + 2p_1 x_{\ell o}, y_{qo} - y_{\ell o} + 2p_2 y_{\ell o}, z_{qo} - z_{\ell o} + 2p_3 z_{\ell o})$, $\boldsymbol{x}_{\boldsymbol{r}} = (2r_1 L_x, 2r_2 L_y, 2r_3 L_z)$, $\sum_{\boldsymbol{p}=0}^{1}$ for $\boldsymbol{p} = (p_1, p_2, p_3)$ and $\sum_{\boldsymbol{r}=-\infty}^{\infty}$ for $\boldsymbol{r} = (r_1, r_2, r_3)$ both represent triplet sum-

mations for indexing room images over three dimensional space. The infinite image depth $\boldsymbol{r}$ is truncated to $\boldsymbol{R}_{\text{depth}}$ for simplicity.

We assume that each loudspeaker constitute a source region and the location of the source coincides with the origin the source region (e.g., $\boldsymbol{x}_s = (0, 0, 0)$ for the ℓth loudspeaker). Therefore, we have $V = 0$, and (4.25) can be rewritten as (4.7), which completes the proof.

4.8.2 Proof of equation (4.18)

We begin with the following integral

$$I_a(k) = \int_0^R \int_0^{2\pi} \int_0^{\pi} |P_d(\boldsymbol{x}, k) - P_a(\boldsymbol{x}, k)|^2 d\Omega \tag{4.28}$$

Substituting (4.9) and (4.11) into (4.28), we have

$$\begin{aligned} I_a(k) = &\sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{n'=0}^{N} \sum_{m'=-n'}^{n'} [\alpha_{nm}^{(d)}(k) - \sum_{\ell=1}^{L} w_l \beta_{nm}^{(\ell)}(k)][\alpha_{n'm'}^{(d)}(k) - \sum_{\ell=1}^{L} w_l \beta_{n'm'}^{(\ell)}(k)]^* \\ &\times \int_0^R j_n(kr) j_{n'}(kr) r^2 dr \int_0^{2\pi} \int_0^{\pi} Y_{nm}(\theta, \phi) Y_{n'm'}(\theta, \phi) \sin\theta d\theta d\phi. \end{aligned} \tag{4.29}$$

According to the orthogonality of the spherical harmonics, it becomes

$$I_a(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} |\alpha_{nm}^{(d)}(k) - \sum_{\ell=1}^{L} w_\ell \beta_{nm}^{(\ell)}(k)|^2 int_0^R j_n^2(kr) r^2 dr. \tag{4.30}$$

Also, the integral in term of r is given by [148]

$$\int_0^R j_n^2(kr) r^2 dr = \mathcal{J}_n(k, R). \tag{4.31}$$

Therefore, we have

$$I_a(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} |\alpha_{nm}^{(d)}(k) - \sum_{\ell=1}^{L} w_l \beta_{nm}^{(\ell)}(k)|^2 \mathcal{J}_n(k, R). \tag{4.32}$$

Similarly,

$$\begin{aligned} I_d(k) &= \int_0^R \int_0^{2\pi} \int_0^{\pi} |P_d(\boldsymbol{x}, k)|^2 d\Omega \\ &= \sum_{n=0}^{N} \sum_{m=-n}^{n} |\alpha_{nm}^{(d)}(k)|^2 \mathcal{J}_n(k, R). \end{aligned} \tag{4.33}$$

Substituting (4.32) and (4.33) into (4.17) completes the proof.

Chapter 5

Particle Velocity Based Sound Field Translation for Binaural Reproduction

Binaural reproduction can offer high quality stereophonic sound audio, however, the reproductions are usually spatially confined to the perspective of each recording microphone. With the rise of virtual reality, there is a demand for sound field translation techniques that can allow users to interact and move throughout the reproduction. In this chapter, we propose a particle velocity based mixed-source sound field translation method for binaural reproduction. We develop a new mixed-source expansion based on the theory of spatial particle velocity vector. The driving signals of the virtual mixed sources are estimated by constructing cost functions to optimize the spatial particle velocity vectors. Compared to the state-of-the-art method, sound pressure based mixed-source expansion, we show through numerical simulations that the proposed particle velocity based mixed-source expansion has better reconstruction performance in sparse solutions, allowing for sound field translation with better perceptual experience over a larger space. We also perceptually validate the proposed method through a Multiple Stimulus with Hidden Reference and Anchor (MUSHRA) experiment for the single source scenario. The experimental results support the better perceptual experience of the proposed method in the single source scenario.

5.1 Introduction

Binaural reproduction is to recreate the real-life experience of hearing sound, which is achieved by binaural recording. However, the perspective is usually fixed to the recording microphone position. To obtain an interactive experience, it requires to record and reproduce the real-world experiences that enables a listener to move about an acoustic space with a sustained perceptual immersion experience. In such a case, sound field translation can play an important role. A typical example of its applications is in virtual-reality reproductions of real-world experiences, where the listener is allowed to virtually explore by moving their body/head and they can experience a spatially-accurate perception that is same as in the real-world space through sound field translation [149].

There have been various studies of sound field translation that allows for a listener to interact and move within a sound field reproduction, and they can be divided into two main categories: interpolation based techniques [150] and extrapolation based techniques [151]. The interpolation based techniques aim to interpolate the sound field to listeners during reproduction using a grid of higher order microphones over space [152]. For example, a simple interpolation based technique is shown in [153], where the room impulse response is interpolated to enable real-time auralizations with head. The interpolation based techniques usually suffer from significant localization error [152, 154] and comb-filtering spectral distortions [155], in particular when the sound source is nearer to one microphone than to another. Although the comb-filtering spectral distortions can be avoided using a parametric Ambisonic room impulse response interpolation system [156], it requires extra efforts to compensate for varying interference. Also, the interpolation assumes multiple distributed microphones or even arrays of microphones, which is a significant hardware limitation in many AR/VR applications [77, 150]. To enhance the performance of perceptual source localization for displaced listeners, a modified method is proposed, where the directional components of the microphone nearest to the listener are emphasized [157]. This method also interpolates the sound field to listeners and is subject to the fundamental limitation. Methods that alleviate the fundamental limitation have been investigated, however, they usually require additional source localization and separation of direct

sound field components [158, 159]. On the other hand, the extrapolation based techniques spherically extend the translation from the recording microphone to the nearest source, which can be achieved by Ambisonics [160, 161], discrete virtual sources [162], plane-wave expansion [163–165], and equivalent source method [166]. A comparison between different extrapolation based techniques is given in [167]. To improve the performance of translation and source localization, especially for off-center positions, optimized decoding methods, such as max-r_E decoding [50] and near-field-compensated decoding [168], have been proposed. However, most of the extrapolated based techniques are limited by the order of the recordings, which is determined by both the upper frequency band and the microphone radius [20, 22]. Consequently, for a low-order recording, the range of translation is significantly limited. Attempting to move beyond this limited range, even after extrapolation, results in spectral distortions [169–171], degraded source localization [167,172], and a poor perceptual listening experience.

Parametric models have been shown to provide an efficient way to describe sound scenes [173, 174], which can be used in binaural reproduction. To overcome the limitation of translation range discussed above, a parametric model is built to describe first-order Ambisonics recordings for translation with known information of listener's position and source distances [175]. A similar approach described in [176], which extends to support arbitrary listener movement, however with multiple source directions and distances. In [177], the authors improve the spatial localization accuracy by additional spatial information obtained from multi-perspective Ambisonics recordings. Higher order Ambisonics signals are used in [178] that translates the sound field based on multi-directional decomposition that estimates source distances. Another parametric decomposition from a higher order Ambisonics signal has been proposed in [179], where it offers excellent localization performance even for strong spatial displacements. However, it also requires a prior knowledge of source distance information and an additional analysis stage of direction of arrival estimation. The sound field translation with very few measurements in a room is investigated in [180]. Recently, an alternative secondary source model for extrapolated virtual reproduction has been developed, where it sparsely expands the recording into the equivalent virtual environment using a mix of near-field and far-field virtual sources [181]. It has been shown that this method

can relax the limitation and offer an improved perceptual experience through perceptual listening tests [113]. However, this method expands the sound field based on sound pressure, which is not directly linked with human perception.

As we mentioned in Chapter 4, particle velocity, which is an acoustic quantity containing the direction information of the sound field, has been widely controlled in sound field reproduction systems [64,68,69,182]. Particle velocity also reflects interaural time difference (ITD), which is one of human localization cues. Therefore, particle velocity is an important predictor of perceived source directions, especially for low frequencies (below 700 Hz) [146]. Furthermore, the particle velocity controlled sound field shows an accurate reproduction of sound intensity in the reproduction region [70, 71], where sound intensity is another good predictor of localization perception, however, it is most appropriate at high frequencies (above 500 Hz) [73, 74]. In this chapter, we propose a new mixed-source expansion based on particle velocity for sound field translation and synthesis. We will show that sparsely expanding the sound field in term of particle velocity offers more immersive perceptual experience at the translated positions, especially for source localization.

The remainder of this chapter is structured as follows. First the sound field translation problem is formulated in Section 5.2. In Section 5.3, we develop the particle velocity based expansion by exploiting the theory of spatial particle velocity vector in the mixed-source model. Section 5.4 translates the sound field and synthesizes binaural signals for translated listeners. In Section 5.5, we introduce two localization metrics, in addition to the relative pressure error, as evaluation criteria. With these criteria, we demonstrate the reproduction accuracy of the proposed method by comparing it with the state-of-the-art method, sound pressure based mixed-source expansion, through numerical simulations. The perceptual validation through a MUSHRA experiement is presented in Section 5.6.

5.2 Problem Formulation

Consider a listener within an acoustic environment, where the listening space of the environment is centered by the origin $\boldsymbol{o}$. The listener is positioned at $\boldsymbol{d} = (r, \theta, \phi)$, where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ in spherical coordinates, with respect to the origin. Let there be a total of K sound sources outside the listening space, with the κth

source located at $\boldsymbol{b}_\kappa$. The free-field binaural sound perceived by the listener can be synthesized by filtering the source signals with the listener's head-related transfer functions (HRTFs), expressed as

$$P_{\mathrm{L,R}}(\boldsymbol{d}, k) = \sum_{\kappa=1}^{K} H_{\mathrm{L,R}}(k, \boldsymbol{b}_\kappa; \boldsymbol{d}) s_\kappa(k), \tag{5.1}$$

where $P_{\mathrm{L,R}}(\boldsymbol{d}, k)$ is the sound pressure at the listener's left and right ear, $H_{\mathrm{L,R}}(k, \boldsymbol{b}_\kappa; \boldsymbol{d})$ are the HRTFs between the κth source and the listener's ears, $s_\kappa(k)$ is the sound signal of the κth source.

We aim to record and reproduce the auditory experience of (5.1) for every possible listening position. The acoustic environment can be characterized using a spherical harmonic decomposition of the sound pressure, expressed as

$$P(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(k|\boldsymbol{x}|) Y_{nm}(\widehat{\boldsymbol{x}}), \tag{5.2}$$

where $\boldsymbol{x}$ is an arbitrary position in the space, $|\cdot| \equiv r$, $\widehat{\cdot} \equiv (\theta, \phi)$, $\alpha_{nm}(k)$ are the pressure coefficients which completely describe the source-free acoustic environment centered about $\boldsymbol{o}$. In practice, the pressure coefficients describing the acoustic environment $\{\alpha_{nm}(k)\}$ can be estimated from recordings of a higher order microphone or a microphone array [76]. We now consider a Nth order receiver centered at $\boldsymbol{o}$, such as a spherical array [75] or planar array [78], that we use to estimate the acoustic environment for $\{\alpha_{nm}(k)\}$ up to order N. However, the limited order N introduces a spatial reproduction constraint that the sound field can only be accurately reproduced within the receiver region (i.e., $|\boldsymbol{x}| \leq R$), where R is the radius of the receiver region that is related by $N = \lceil kR \rceil$. This constraint restricts the listener to move within the receiver region (i.e., $|\boldsymbol{d}| \leq R$). If the listener attempts to move beyond the receiver region, they would experience spectral distortions and a loss in perceptual immersion.

The objective of this work is to perceptually enhance the performance of translation to $|\boldsymbol{d}| > R$ for binaural reproduction from a mode limited measurement so that a listener can move over a large space in virtual reconstruction with a sustained perceptual experience, through the proposed particle velocity based mixed-source

expansion.

5.3 Particle Velocity Based Mixed-Source Expansion

In this section, we first introduce the mixed-source model in Section 5.3.1. The particle velocity based mixed-source expansion is developed in Section 5.3.2.

5.3.1 Mixed-source model

Different from the plane wave model, which represents a sound field with an equivalent superposition of virtual plane wave sources, the mixed-source model expands the sound field in terms of a mixture of near-field and far-field virtual sources [181]. This mixed-source model can overcome the difficulties of the plane wave expansion in synthesizing the near-field sources.

The mixed-source is defined as

$$P(\boldsymbol{x}, k; \boldsymbol{y}) = |\boldsymbol{y}| e^{-ik|\boldsymbol{y}|} \frac{e^{ik|\boldsymbol{y}-\boldsymbol{x}|}}{4\pi|\boldsymbol{y}-\boldsymbol{x}|}, \tag{5.3}$$

where $\boldsymbol{y}$ denotes the position of the virtual mixed source. In the spherical harmonic domain, it can be expressed as

$$P(\boldsymbol{x}, k; \boldsymbol{y}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \underbrace{ik|\boldsymbol{y}| e^{-ik|\boldsymbol{y}|} h_n(k|\boldsymbol{y}|) Y_{nm}^*(\widehat{\boldsymbol{y}})}_{\beta_{nm}(k)} j_n(k|\boldsymbol{x}|) Y_{nm}(\widehat{\boldsymbol{x}}), \tag{5.4}$$

where $\beta_{nm}(k)$ are the pressure coefficients of the mixed source. We note that the mixed source has the property of [20]

$$\lim_{|\boldsymbol{y}| \to +\infty} P(\boldsymbol{x}, k; \boldsymbol{y}) = \frac{e^{-ik\widehat{\boldsymbol{y}} \cdot \boldsymbol{x}}}{4\pi}, \tag{5.5}$$

which shows that the plane-wave expansion can be modeled using the mixed-source expansion.

We construct a virtual equivalent sound field using two concentric spheres of virtual mixed sources. Therefore, the sound field can be expressed, using the mixed-source model, as [181]

$$\begin{aligned} P(\boldsymbol{x},k) = &\int \psi(k, R_N\widehat{\boldsymbol{y}};\boldsymbol{o}) \underbrace{R_N e^{-ikR_N} \frac{e^{ik||R_N\widehat{\boldsymbol{y}}-\boldsymbol{x}||}}{4\pi||R_N\widehat{\boldsymbol{y}}-\boldsymbol{x}||}}_{P(\boldsymbol{x},k;R_N\widehat{\boldsymbol{y}})} d\widehat{\boldsymbol{y}} \\ &+ \int \psi(k, R_F\widehat{\boldsymbol{y}};\boldsymbol{o}) \underbrace{R_F e^{-ikR_F} \frac{e^{ik||R_F\widehat{\boldsymbol{y}}-\boldsymbol{x}||}}{4\pi||R_F\widehat{\boldsymbol{y}}-\boldsymbol{x}||}}_{P(\boldsymbol{x},k;R_F\widehat{\boldsymbol{y}})} d\widehat{\boldsymbol{y}}, \end{aligned} \tag{5.6}$$

where R_N is the radius of the virtual sphere placed in the near-field, R_F is the radius of the virtual sphere placed in the far-field, $\psi(k, R_N\widehat{\boldsymbol{y}};\boldsymbol{o})$ and $\psi(k, R_F\widehat{\boldsymbol{y}};\boldsymbol{o})$ are the driving functions of the mixed-source distributions as observed at $\boldsymbol{o}$, and $P(\boldsymbol{x},k;R_N\widehat{\boldsymbol{y}})$ and $P(\boldsymbol{x},k;R_F\widehat{\boldsymbol{y}})$ are the mixed sources on the two spheres. Note that the mixed-source model expands the sound field with respect to $\boldsymbol{o}$, which allows us to observe the source distribution at $\boldsymbol{o}$ and estimate the sound at any translated position $\boldsymbol{x}$.

5.3.2 Particle velocity based expansion

According to the results in Chapter 3, we have the continuous particle velocity in the spherical harmonic domain, shown as

$$V_\Psi(\boldsymbol{x},k) = \sum_{p=0}^{Q_\Psi} \sum_{q=-p}^{p} X_{pq}^{(\Psi)}(k,|\boldsymbol{x}|) Y_{pq}(\widehat{\boldsymbol{x}}); \qquad \Psi \in \{r,\theta,\phi\}, \tag{5.7}$$

where Q_Ψ is the particle velocity truncation order in the Ψ direction, and $X_{pq}^{(\Psi)}(k,|\boldsymbol{x}|)$ denote the velocity coefficients in the Ψ direction.

Similar to (3.23), for the mixed source, we have

$$\boldsymbol{V}(\boldsymbol{x},k;R_X\widehat{\boldsymbol{y}}) = \frac{i}{k\rho_0 c}\vec{\nabla}P(\boldsymbol{x},k;R_X\widehat{\boldsymbol{y}}), \tag{5.8}$$

where $R_X \in \{R_N, R_F\}$, therefore, taking the gradient operation on both sides of

(5.6), we can obtain the particle velocity based mixed-source expansion

$$V_\Psi(\boldsymbol{x},k) = \int \psi(k, R_N\widehat{\boldsymbol{y}};\boldsymbol{o})V_\Psi(\boldsymbol{x},k;R_N\widehat{\boldsymbol{y}})d\widehat{\boldsymbol{y}} + \int \psi(k, R_F\widehat{\boldsymbol{y}};\boldsymbol{o})V_\Psi(\boldsymbol{x},k;R_F\widehat{\boldsymbol{y}})d\widehat{\boldsymbol{y}}, \tag{5.9}$$

where $V_\Psi(\boldsymbol{x},k;R_N\widehat{\boldsymbol{y}})$ and $V_\Psi(\boldsymbol{x},k;R_F\widehat{\boldsymbol{y}})$ represent the particle velocity of the virtual sources in the near-field and the far-field, respectively, in the Ψ direction.

In practice, the particle velocity based mixed-source expansion (5.9) can be approximated by a finite set of L known virtual sources for each virtual sphere, therefore, (5.9) can be expressed as

$$\begin{aligned} V_\Psi(\boldsymbol{x},k) = &\sum_{\ell=1}^{L} w_\ell \psi(k, R_N\widehat{\boldsymbol{y}}_\ell;\boldsymbol{o})V_\Psi(\boldsymbol{x},k;R_N\widehat{\boldsymbol{y}}_\ell) \\ &+ \sum_{\ell=1}^{L} w_\ell \psi(k, R_F\widehat{\boldsymbol{y}}_\ell;\boldsymbol{o})V_\Psi(\boldsymbol{x},k;R_F\widehat{\boldsymbol{y}}_\ell), \end{aligned} \tag{5.10}$$

where w_ℓ are the the sampling weights for the source distribution, and $\widehat{\boldsymbol{y}}_\ell$ denotes the position of the ℓth virtual source on each sphere.

Note that the parameters of the mixed-source expansion need to be carefully selected in the implementations. The near-field radius of the virtual sphere can be selected based on the desired maximum translation distance, because the listener cannot move beyond the near-field sphere. As for the far-field sphere, it can be placed anywhere so long as it is in the far-field. Besides, the selection of the number of the virtual sources per sphere is a trade-off between angular resolution and computation cost. The more virtual sources there are, the more opportunity they have to match the true sound direction of arrival, however, the higher computation cost the system has.

By replacing $\alpha_{nm}(k)$ with $\beta_{nm}^{(\ell)}$ (i.e., the pressure coefficients of the ℓth virtual source) in (3.31), (3.32), and (3.33), and then substituting them into (5.7), the particle velocity due to the ℓth virtual source can be written as

$$V_\Psi(\boldsymbol{x},k;R_X\widehat{\boldsymbol{y}}_\ell) = \sum_{p=0}^{Q_\Psi}\sum_{q=-p}^{p} X_{pq}^{(\Psi,\ell)}(k,|\boldsymbol{x}|;R_X\widehat{\boldsymbol{y}}_\ell)Y_{pq}(\widehat{\boldsymbol{x}}), \tag{5.11}$$

where $X_{pq}^{(\Psi,\ell)}(k,|\boldsymbol{x}|)$ are the velocity coefficients of the ℓth virtual source. Substituting (5.7) and (5.11) into (5.10), (5.10) can be simplified as

$$X_{pq}^{(\Psi)}(k,|\boldsymbol{x}|) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \widehat{\boldsymbol{y}}_\ell; \boldsymbol{o}) X_{pq}^{(\Psi,\ell)}(k,|\boldsymbol{x}|; R_\ell \widehat{\boldsymbol{y}}_\ell), \tag{5.12}$$

where $R_\ell = R_N$ for $\ell \in [1, L]$, and for $\ell \in [L+1, 2L]$ we have $R_\ell = R_F$, and $w_\ell = w_{\ell-L}$.

We denote (5.12), in matrix form, as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\psi}, \tag{5.13}$$

where $\boldsymbol{X} = [\boldsymbol{X}_r^T, \boldsymbol{X}_\theta^T, \boldsymbol{X}_\phi^T]^T$ is a $[(Q_r+1)^2 + (Q_\theta+1)^2 + (Q_\phi+1)^2]$ long vector containing all three components of the particle velocity vector with $\boldsymbol{X}_\Psi = [X_{00}^{(\Psi)}, X_{1(-1)}^{(\Psi)}, ..., X_{Q_\Psi Q_\Psi}^{(\Psi)}]^T$, $\boldsymbol{\psi} = [\psi_1, ..., \psi_{\mathcal{L}}]^T$ is a $\mathcal{L} = 2L$ long vector with $\psi_{\mathcal{L}} = \psi(k, R_{\mathcal{L}} \widehat{\boldsymbol{y}}_{\mathcal{L}}; \boldsymbol{o})$, and $\boldsymbol{A} = [\boldsymbol{A}_r^T, \boldsymbol{A}_\theta^T, \boldsymbol{A}_\phi^T]^T$ is the $[(Q_r+1)^2 + (Q_\theta+1)^2 + (Q_\phi+1)^2]$ by $\mathcal{L}$ expansion matrix with

$$\boldsymbol{A}_\Psi = \begin{bmatrix} w_1 X_{00}^{(\Psi,1)} & \cdots & w_{\mathcal{L}} X_{00}^{(\Psi,\mathcal{L})} \\ w_1 X_{1(-1)}^{(\Psi,1)} & \cdots & w_{\mathcal{L}} X_{1(-1)}^{(\Psi,\mathcal{L})} \\ \vdots & \ddots & \vdots \\ w_1 X_{Q_\Psi Q_\Psi}^{(\Psi,1)} & \cdots & w_{\mathcal{L}} X_{Q_\Psi Q_\Psi}^{(\Psi,\mathcal{L})} \end{bmatrix}. \tag{5.14}$$

Note that the dependence is omitted here for notational simplicity. The expansion is now reduced to calculate the driving function of the mixed source distribution $\psi(k, \boldsymbol{y}; \boldsymbol{o})$ for a given measured sound field. We introduce two solutions here.

Least squares solution

The least squares method is to estimate the driving function $\boldsymbol{\psi}$ so as to minimize the difference between $\boldsymbol{X}$ and $\boldsymbol{A}\boldsymbol{\psi}$, and it can be formulated as

$$\boldsymbol{\psi} = \arg\min_{\boldsymbol{\psi}} ||\boldsymbol{A}\boldsymbol{\psi} - \boldsymbol{X}||_2^2. \tag{5.15}$$

This problem can be solved using a Moore-Penrose inverse with Tikhonov regularization [119].

Sparse solution

It is shown that modeling fewer virtual sources from propagation directions that are similar to the original sound will lead to better perceptual experience [181]. Therefore, we also construct a sparse constrained objective function, expressed as

$$\boldsymbol{\psi} = \arg\min_{\boldsymbol{\psi}} ||\boldsymbol{A}\boldsymbol{\psi} - \boldsymbol{X}||_2^2 + \lambda||\boldsymbol{\psi}||_1, \tag{5.16}$$

where λ is the parameter controlling the strength of the sparsity constraint for the driving function $\boldsymbol{\psi}$. This sparse linear regression problem can be solved using LASSO [45]. Other optimization techniques such as the iteratively reweighted least squares (IRLS) [183] can also be applied to solving this problem. We direct readers to [184] for more details about compressive sensing. Note that the applicability of the above solutions will depend on the condition number of matrix $\boldsymbol{A}$, which determined by the distribution of the virtual sources. It shows that a well-conditioned $\boldsymbol{A}$ matrix should result from a geometry in which the virtual sources are maximally distributed over space in literature [20].

5.4 Sound Field Translation and Synthesis for Binaural Reproduction

In this section, we translate the sound field in the spherical harmonic domain, and then synthesize the left and right ear signals at the translated position.

Solving (5.15) or (5.16) gives us the reproduced virtual-reality sound field, expressed as

$$\widetilde{P}(\boldsymbol{x}, k) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \widehat{\boldsymbol{y}}_\ell; \boldsymbol{o}) R_\ell e^{-ikR_\ell} \frac{e^{ik||R_\ell \widehat{\boldsymbol{y}}_\ell - \boldsymbol{x}||}}{4\pi||R_\ell \widehat{\boldsymbol{y}}_\ell - \boldsymbol{x}||}. \tag{5.17}$$

With respect to the translated position $\boldsymbol{d}$, the virtual sources are located at

$\boldsymbol{z}_\ell = R_\ell \widehat{\boldsymbol{y}}_\ell - \boldsymbol{d}$, therefore, the translated sound field can be written as

$$\widetilde{P}(\boldsymbol{x}, k; \boldsymbol{d}) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \widehat{\boldsymbol{y}}_\ell; \boldsymbol{o}) R_\ell e^{-ikR_\ell} \frac{e^{ik||\boldsymbol{z}_\ell - \boldsymbol{x}||}}{4\pi||\boldsymbol{z}_\ell - \boldsymbol{x}||}. \tag{5.18}$$

We decompose (5.18) in the spherical harmonic domain about the position $\boldsymbol{d}$, and the pressure coefficients of the translated sound field due to the virtual sources at $\boldsymbol{z}_\ell$ can expressed as

$$\gamma_{vu}(k; \boldsymbol{d}) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \widehat{\boldsymbol{y}}_\ell; \boldsymbol{o}) ikR_\ell e^{-ikR_\ell} h_v(k|\boldsymbol{z}_\ell|) Y_{vu}^*(\widehat{\boldsymbol{z}}_\ell), \tag{5.19}$$

where v and u index the translated harmonics, centered at $\boldsymbol{d}$. The left and right ear signals can be synthesized, in the spherical harmonic domain, directly from the translated pressure coefficients $\gamma_{vu}(k; \boldsymbol{d})$ and the coefficients of the listener's HRTFs, $H_{\mathrm{L,R}}^{vu}(k)$, with [185, 186]

$$\widetilde{P}_{\mathrm{L,R}}(k; \boldsymbol{d}) = \sum_{v=0}^{\infty} \sum_{u=-v}^{v} \gamma_{vu}(k; \boldsymbol{d}) H_{\mathrm{L,R}}^{vu}(k). \tag{5.20}$$

Note that $H_{\mathrm{L,R}}^{vu}(k)$ is obtained from the spherical harmonic decomposition of the HRTF $H_{\mathrm{L,R}}(k, \boldsymbol{y}; \boldsymbol{d})$.

Alternatively, the left and right ear signals can be obtained by applying the mixed-source driving functions to the HRTFs based on the listener's translated position, given as

$$\widetilde{P}_{\mathrm{L,R}}(k; \boldsymbol{d}) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \widehat{\boldsymbol{y}}_\ell; \boldsymbol{o}) H_{\mathrm{L,R}}(k, \boldsymbol{y}_\ell; \boldsymbol{d}). \tag{5.21}$$

We direct the reader to [55, 187] for other options of rendering the ear signals at the translated positions.

5.5 Simulation Study

In this section, we firstly examine the reproduced sound pressure to assess the proposed particle velocity based mixed-source expansion in two simulated acoustic environment, including a single source and multiple sources, respectively. Secondly, we also evaluate the proposed method in terms of perception of sound localization using two localization metrics for the single source scenario. The simulated environment and the evaluation criteria are explained in Section 5.5.1. The simulation results and discussions are given in Section 5.5.2 and Section 5.5.3, respectively.

5.5.1 Simulation setup and criteria

In this simulation, we simulate two acoustic scenarios. We firstly consider an acoustic environment that contains a single free-field point source with location of $\boldsymbol{b}_1 = (1, \pi/3, \pi/4)$. We then consider a scenario where multiple sources exist, which is constructed by adding two more point sources with randomly selected locations of $\boldsymbol{b}_2 = (0.5, 3\pi/4, 4\pi/3)$ and $\boldsymbol{b}_3 = (1.8, \pi/2, \pi/2)$, respectively, in the former acoustic environment. The true sound pressure at an arbitrary position $\boldsymbol{x}$ due to the jth point source is

$$P_j(\boldsymbol{x}, k) = \frac{e^{ik||\boldsymbol{b}_j - \boldsymbol{x}||}}{4\pi||\boldsymbol{b}_j - \boldsymbol{x}||}, \qquad j = \{1, 2, 3\}. \tag{5.22}$$

The total sound pressure due to multiple sources follows the principle of linear superposition. We assume that the receiver can record the sound field up to the 4th order (i.e., $N = 4$), and the radius of the receiver region is $R = 0.042$ m.[1] The pressure coefficients describing the acoustic environment $\{\alpha_{nm}(k)\}$ can be extracted from the receiver's recording. Here, for simplicity, we use the theoretical pressure coefficients up to the Nth order, expressed as

[1]We simulate the radius of the EigenMike, which is a popular commercial microphone array [75].

$$\alpha_{nm}(k) = \begin{cases} ikh_n(k|\boldsymbol{b}_1|)Y^*_{nm}(\widehat{\boldsymbol{b}}_1), \quad n \in [0, N], & \text{single source scenario,} \\ \\ \sum_{j=1}^{3} ikh_n(k|\boldsymbol{b}_j|)Y^*_{nm}(\widehat{\boldsymbol{b}}_j), \quad n \in [0, N], & \text{multiple sources scenario.} \end{cases} \tag{5.23}$$

The mixed-source model consists of two sets of $L = 36$ virtual sources distributed on the two spheres, both with the positions arranged on Fliege nodes [188]. The first set is placed in near-field at $R_N = 2$ m, and the second set is placed in far-field at $R_F = 20$ m.

Throughout the simulations, the frequency of the sources is $f = 1500$ Hz except the evaluation in terms of the frequency. Sound speed is $c = 343$ m/s and air density is $\rho_0 = 1.29$ kg/m^3. The truncation orders of the velocity coefficients in the r, θ, ϕ direction are $Q_r = N$, $Q_\theta = 2N$, and $Q_\phi = 2N$, respectively. For the sparse solution, we use a 500 iterations LASSO.

As the first objective performance measure for the reproduction method, we define the mismatch between the reproduced sound field and the true sound field at any point $\boldsymbol{x}$ as

$$\varepsilon(\boldsymbol{x}, k) = \frac{|P(\boldsymbol{x}, k) - \widetilde{P}(\boldsymbol{x}, k)|^2}{|P(\boldsymbol{x}, k)|^2} \times 100(\%). \tag{5.24}$$

A good perception of sound localization requires accurate reproduction of sound direction. To evaluate the performance of the reproduction of sound direction at the translated position $\boldsymbol{d}$, two localization metrics are introduced [167]. The first one is related to the velocity vector at the translated position, which is given by

$$\boldsymbol{r}_V(k; \boldsymbol{d}) = \mathrm{Re}\left(\frac{\sum_{\ell=1}^{\mathcal{L}} \mathcal{U}_\ell(k) \widehat{\boldsymbol{z}}_\ell}{\sum_{\ell=1}^{\mathcal{L}} \mathcal{U}_\ell(k)} \right), \tag{5.25}$$

where $\mathcal{U}_\ell(k)$ are the effective source gains (accounting for point-source radiation), expressed as

$$\mathcal{U}_\ell(k) = \frac{e^{ik(|\boldsymbol{z}_\ell| - R_\ell)}}{|\boldsymbol{z}_\ell|/R_\ell} \psi(k, R_\ell \widehat{\boldsymbol{y}}_\ell; \boldsymbol{o}). \tag{5.26}$$

$\boldsymbol{r}_V(k)$ is used to predict localization at low frequencies (below 700 Hz). For high

frequencies (above 500 Hz), the second metric related to the intensity vector is given by

$$\boldsymbol{r}_E(k;\boldsymbol{d}) = \frac{\sum_{\ell=1}^{\mathcal{L}} |\mathcal{U}_\ell(k)|^2 \widehat{\boldsymbol{z}}_\ell}{\sum_{\ell=1}^{\mathcal{L}} |\mathcal{U}_\ell(k)|^2}. \tag{5.27}$$

The directions of $\boldsymbol{r}_V(k;\boldsymbol{d})$ and $\boldsymbol{r}_E(k;\boldsymbol{d})$ indicate the expected localization direction, and their magnitudes, $|\boldsymbol{r}_V(k;\boldsymbol{d})|$ and $|\boldsymbol{r}_E(k;\boldsymbol{d})|$, indicate the quality of localization [167]. For the point source at $\boldsymbol{b}$, the directional error $\zeta_{V,E}(k;\boldsymbol{d})$ at the translated position $\boldsymbol{d}$ is defined as[2]

$$\zeta_{V,E}(k;\boldsymbol{d}) = \left| \frac{\boldsymbol{r}_{V,E}(k;\boldsymbol{d})}{|\boldsymbol{r}_{V,E}(k;\boldsymbol{d})|} - \frac{\boldsymbol{b}-\boldsymbol{d}}{|\boldsymbol{b}-\boldsymbol{d}|} \right|. \tag{5.28}$$

Ideally, $\boldsymbol{r}_V(k;\boldsymbol{d})$ and $\boldsymbol{r}_E(k;\boldsymbol{d})$ should be unit vectors that point to the direction of the point source $\boldsymbol{b}$ (i.e., $|\boldsymbol{r}_V(k;\boldsymbol{d})| = |\boldsymbol{r}_E(k;\boldsymbol{d})| = 1$, and $\zeta_{V,E}(k;\boldsymbol{d}) = 0$).

We also implement the two corresponding solutions (i.e., the closed-form solution and the sparse solution) of the state-of-the-art technique, the sound pressure based mixed-source expansion [113, 181], for comparison.

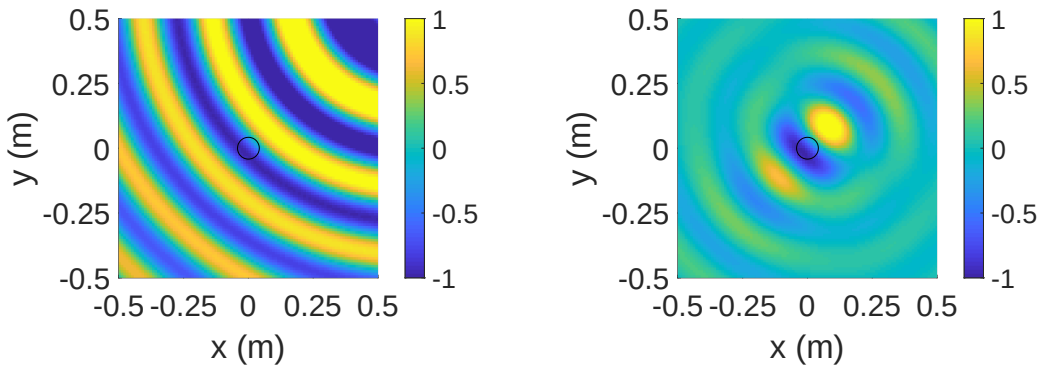
5.5.2 Simulation results

In this subsection, we first examine the accuracy of the proposed particle velocity based mixed-source expansion to reconstruct the pressure field in the two scenarios, and then we evaluate the localization performance in terms of the frequency and translated position using the metrics in Section 5.5.1 in the single source scenario.

Reproduced pressure field

In terms of the single source scenario, Figure 5.1 shows the true pressure field calculated by (5.22), and the recorded pressure field calculated using the theoretical pressure coefficients in (5.23). The black circle denotes the receiver region. The receiver can record the sound field within the receiver region. However, truncation error degrades the accuracy of the pressure field beyond the receiver region, which is the constraint in Section 5.2 that we wish to relax through the proposed method.

[2]The direction error is related to the direction difference in angles between the two unit vectors in (5.28) by $\arccos(1-\zeta^2/2)$.



(a) (b)

Figure 5.1: The true pressure field and recorded pressure field at 1500Hz in the single source scenario. The black circle denotes the receiver region. (a) the true pressure field; (b) the recorded pressure field.

To compare the proposed method with the sound pressure based expansion, we first show the results of the sound pressure based expansion for the closed-form solution and the sparse solution in Fig. 5.2. We observe that the closed-form solution can guarantee the reproduction accuracy within the receiver region. However, for a position outside the receiver region, the reconstructed pressure field is similar (but not exact) to the recorded pressure field, which is not good enough for translation. The sparse solution improves the performance and reconstructs the pressure field accurately around the receiver region, which allows for the listener to move to the position exterior to the receiver region. Figure 5.3 shows the reconstructed pressure field and the error field controlled by the least squares solution and the sparse solution using the particle velocity based expansion. We notice that the least squares solution of the particle velocity based expansion has the same performance as the closed-form solution of the sound pressure based expansion. However, for the sparse solution of the particle velocity based expansion, it provides a larger area with smaller error around the receiver region compared with the sparse solution of the sound pressure based expansion, which allows for the listener to translate further away from the original receiver region.

As for the multiple sources scenario, the true pressure field and recorded pres-

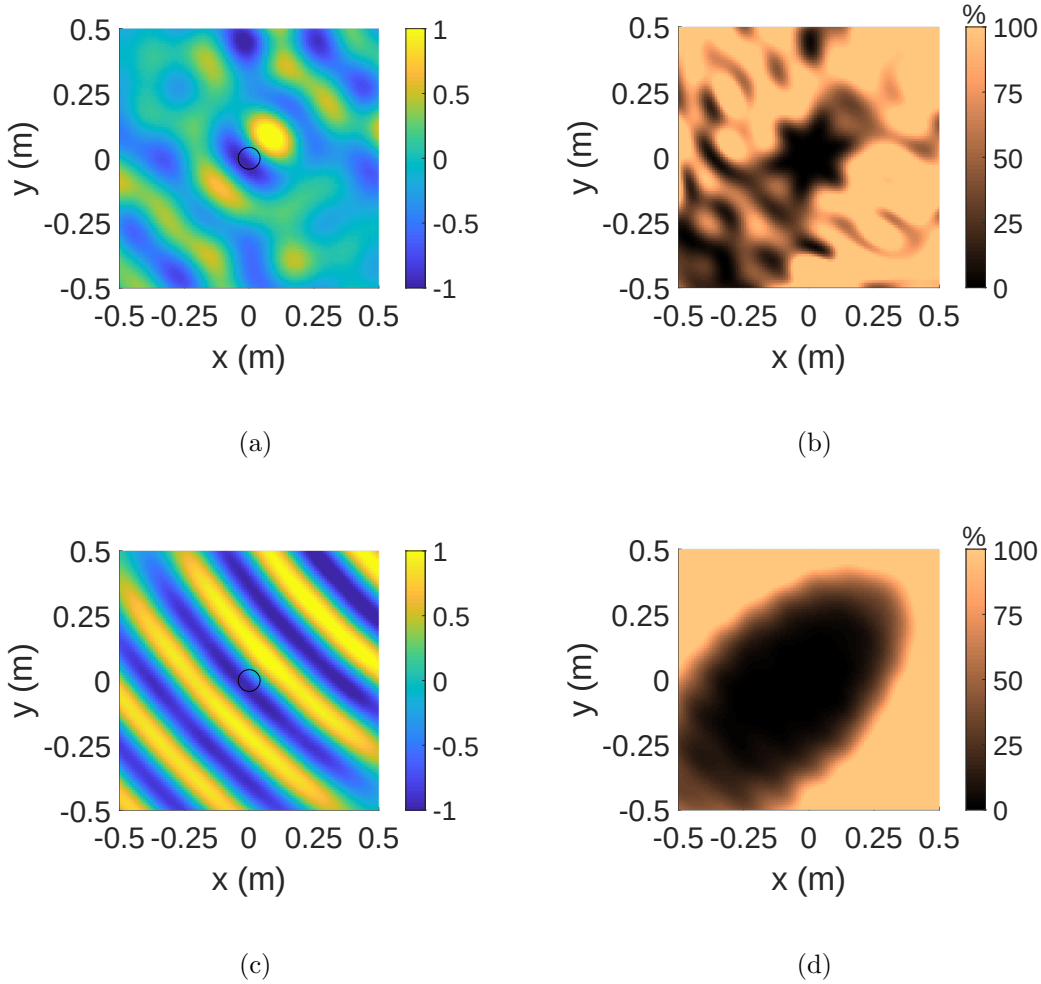


Figure 5.2: The reconstructed pressure field and the error field controlled by the closed-form solution (P-CFS) and the sparse solution (P-SS) using the sound pressure based expansion in the single source scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by P-CFS ; (b) Error field controlled by P-CFS; (c) Reconstructed pressure field controlled by P-SS; (d) Error field controlled by P-SS.

sure field are given in Fig. 5.4. The difference between them also increases gradually as the observation point moves away from the target region due to the truncation error. Similar to the evaluation for the single source scenario, we first show the reconstructed pressure fields controlled by the closed-form solution and the sparse solution based on the sound pressure based expansion in the multiple sources sce-

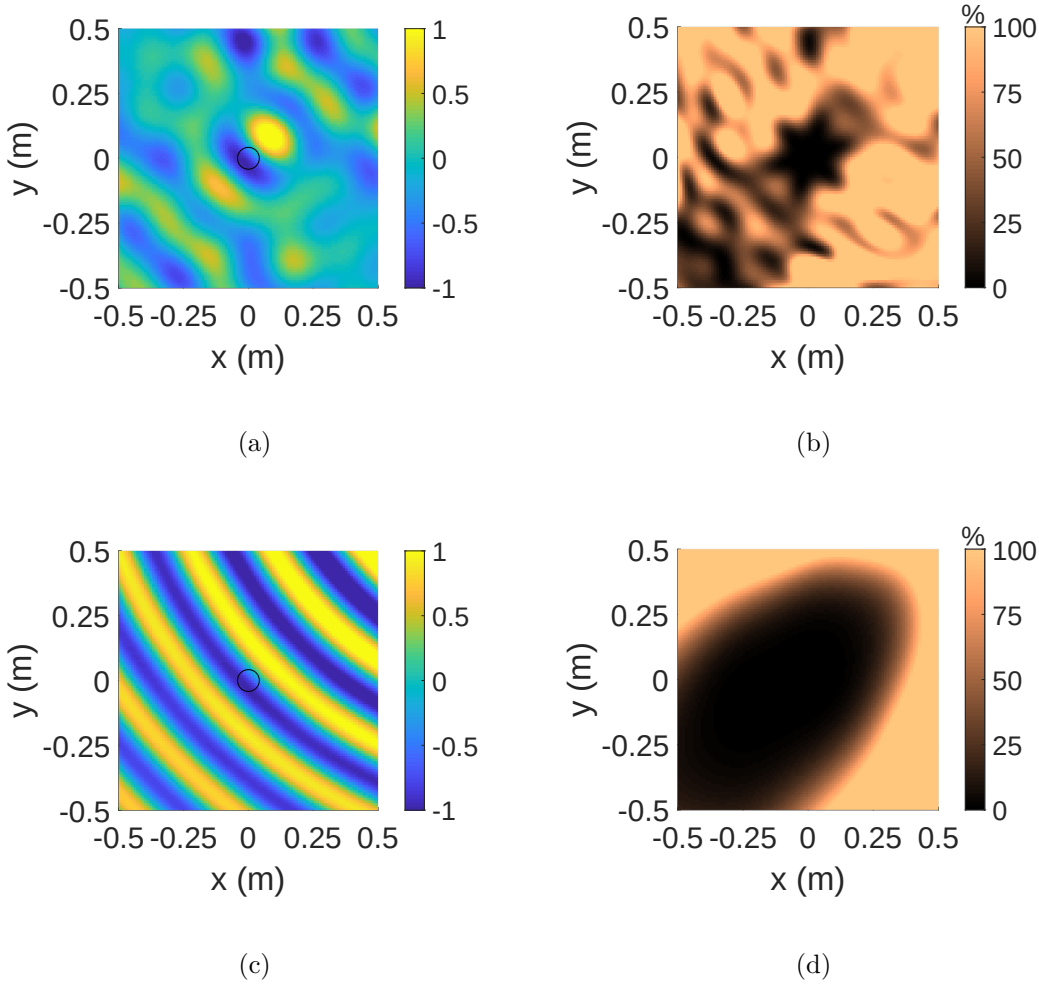
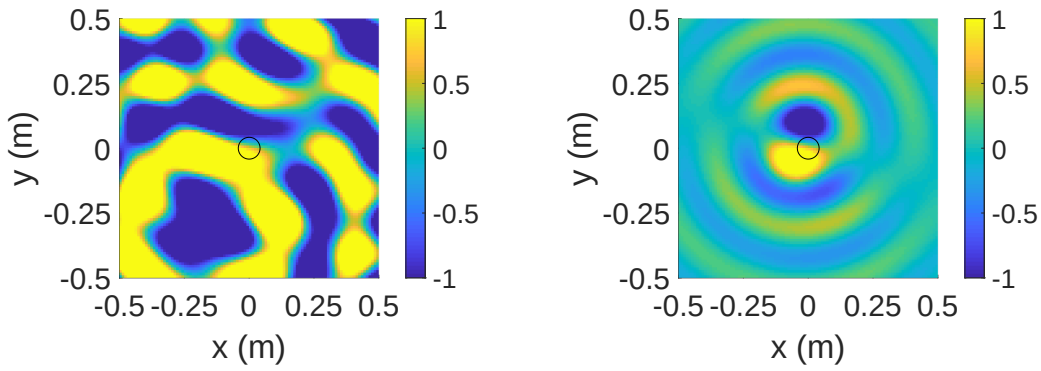


Figure 5.3: The reconstructed pressure field and the error field controlled by the least squares solution (V-LSS) and the sparse solution (V-SS) using the particle velocity based expansion in the single source scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by V-LSS ; (b) Error field controlled by V-LSS; (c) Reconstructed pressure field controlled by V-SS; (d) Error field controlled by V-SS.

nario. The results with the error fields are shown in Fig. 5.5. It shows that the closed-form solution still can only guarantee accurate reproduction around the target region, whereas the sparse solution can provide greater translation area. However, this area is much smaller than that in the single source scenario. Figure 5.6 shows the results of the least squares solution and the sparse solution based



(a) (b)

Figure 5.4: The true pressure field and recorded pressure field at 1500Hz in the multiple sources scenario. The black circle denotes the receiver region. (a) the true pressure field; (b) the recorded pressure field.

on the particle velocity based expansion. Similarly, the result of the least squares solution under the particle velocity based expansion is identical to the result of the closed-form solution under the sound pressure based expansion. Although the sparse solution of particle velocity based expansion in the multiple sources scenario can provide smaller translation area than that in the single source scenario, it still has an improved performance compared to the sparse solution of the sound pressure based expansion in the multiple sources scenario.

Frequency

To evaluate the broadband performance of the proposed method on direction reproduction at a translated position, we calculate the directional error $\zeta_{V,E}$ and the magnitude of $\boldsymbol{r}_{V,E}$ (i.e., $|\boldsymbol{r}_{V,E}|$) in the single source scenario using the two localization metrics introduced in Section 5.5.1, with respect to the change of frequency. The translated position is $(0.2, 0.3, 0)$ m in Cartesian coordinates. Figure 5.7 shows the results of ζ_V and $|\boldsymbol{r}_V|$. As expected, the sparse solution of the particle velocity based expansion is better than all the other solutions, especially for low frequencies. The results of ζ_E and $|\boldsymbol{r}_E|$ are given in Fig. 5.8. We observe that the the sparse solution of the particle velocity based expansion has the least directional er-

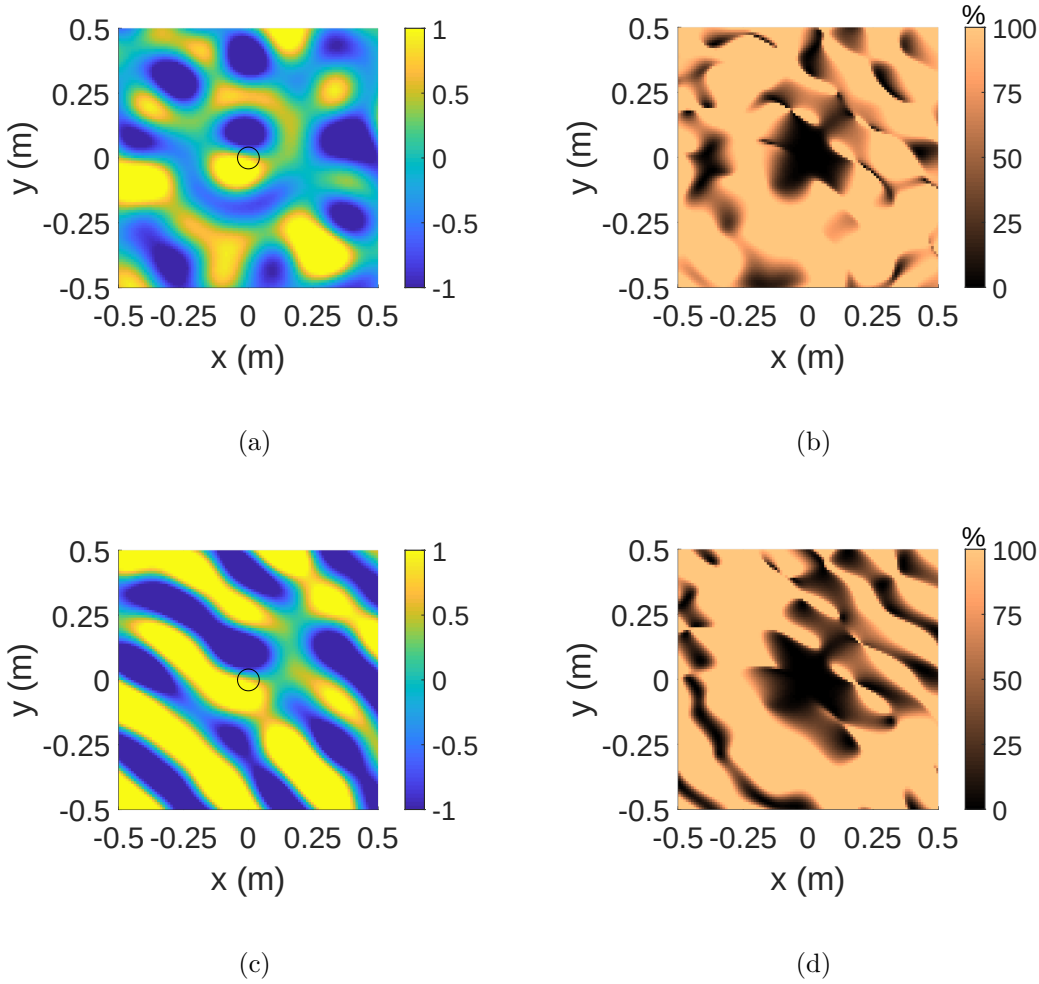


Figure 5.5: The reconstructed pressure field and the error field controlled by the closed-form solution (P-CFS) and the sparse solution (P-SS) using the sound pressure based expansion in the multiple sources scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by P-CFS ; (b) Error field controlled by P-CFS; (c) Reconstructed pressure field controlled by P-SS; (d) Error field controlled by P-SS.

ror, and the $|\boldsymbol{r}_E|$ for this solution is closest to 1 among all the evaluated methods. Compared to curves for the other method, the curve for the sparse solution of the particle velocity based expansion has little fluctuation in terms of frequency, which means the the sparse solution of the particle velocity based expansion can provide an uniform performance for broadband signals. Also, the least squares solution of

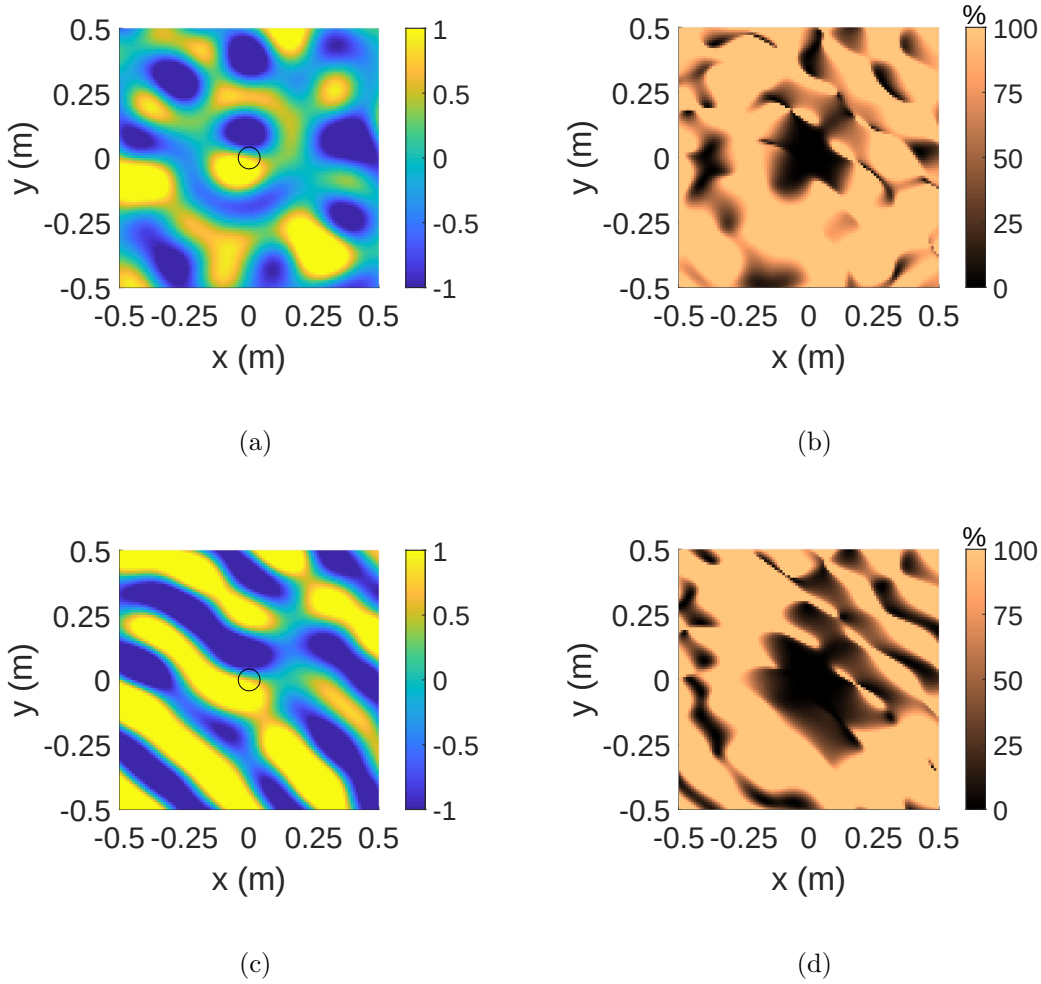
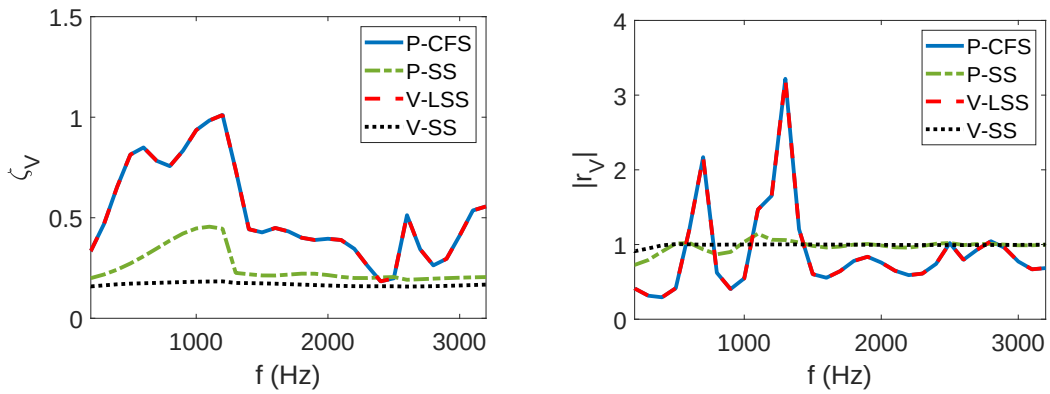


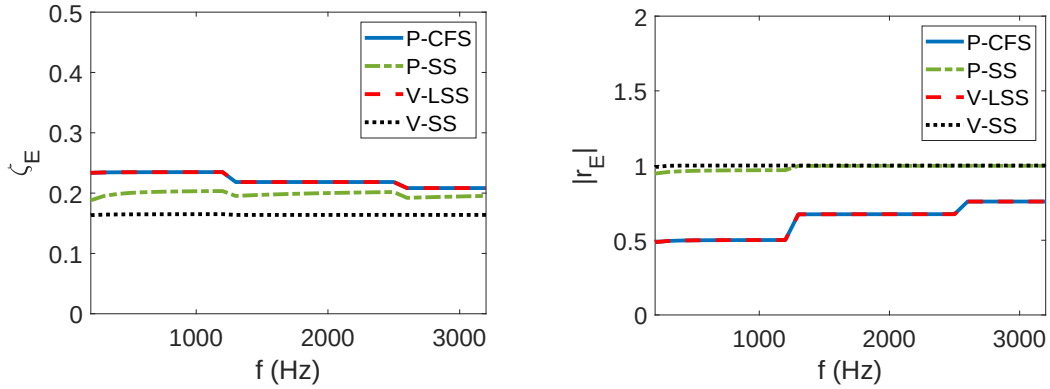
Figure 5.6: The reconstructed pressure field and the error field controlled by the least squares solution (V-LSS) and the sparse solution (V-SS) using the particle velocity based expansion in the multiple sources scenario. The source frequency is 1500 Hz. (a) Reconstructed pressure field controlled by V-LSS ; (b) Error field controlled by V-LSS; (c) Reconstructed pressure field controlled by V-SS; (d) Error field controlled by V-SS.

the particle velocity based expansion is the same as the closed-form solution of the sound pressure based expansion for the performance in terms of frequency.



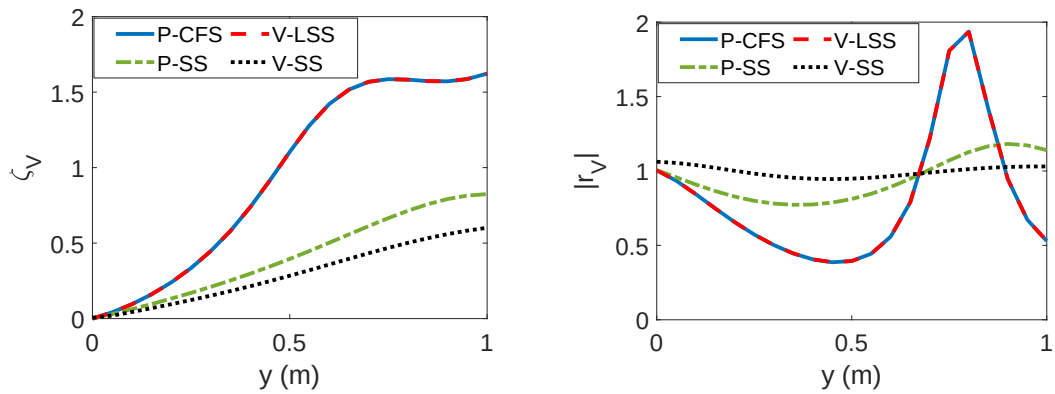
(a) (b)

Figure 5.7: The directional error ζ_V and the magnitude of $\boldsymbol{r}_V$ with respect to frequency at the translated position of (0.2, 0.3, 0) m in the single source scenario. (a) Directional error ζ_V; (b) Magnitude of $\boldsymbol{r}_V$.



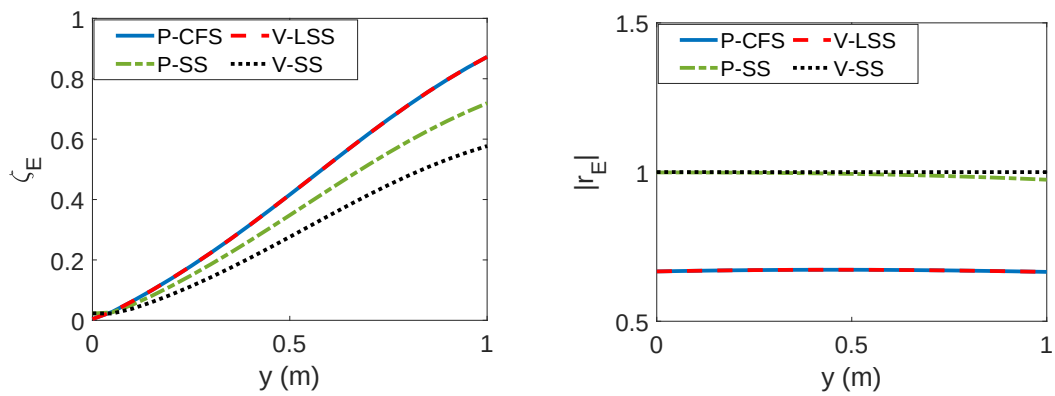
(a) (b)

Figure 5.8: The directional error ζ_E and the magnitude of $\boldsymbol{r}_E$ with respect to frequency at the translated position of (0.2, 0.3, 0) m in the single source scenario. (a) Directional error ζ_E; (b) Magnitude of $\boldsymbol{r}_E$.



(a) (b)

Figure 5.9: The directional error ζ_V and the magnitude of $\boldsymbol{r}_V$ as a function of translation distance along the positive y-axis in the single source scenario. The frequency of the source is 300 Hz. (a) Directional error ζ_V; (b) Magnitude of $\boldsymbol{r}_V$.



(a) (b)

Figure 5.10: The directional error ζ_E and the magnitude of $\boldsymbol{r}_E$ as a function of translation distance along the positive y-axis in the single source scenario. The frequency of the source is 1500 Hz. (a) Directional error ζ_E; (b) Magnitude of $\boldsymbol{r}_E$.

Translated position

We also evaluate the performance of each method when the translated position moves away from the receiver region in the single source scenario. Figures 5.9 and

5.10 show the results of ζ_V and $|\boldsymbol{r}_V|$ at 300 Hz, and the results of ζ_E and $|\boldsymbol{r}_E|$ at 1500 Hz, respectively, as a function of translation distance along the positive y-axis. Note that we plot Fig. 5.9 and Fig. 5.10 at different frequencies because $\boldsymbol{r}_V$ is a good predictor for low frequencies whereas $\boldsymbol{r}_E$ is a good predictor for high frequencies, as mentioned in Section 5.5.1. From Figs. 5.9(a) and 5.10(a), we show that all the methods have little error at the origin where the receiver is positioned, however, the performance worsens as the translation distance increases for all the methods. Comparing to the sparse solution of the sound pressure based expansion, the sparse solution of the particle velocity based expansion has less directional error and this advantage becomes more significant for a further translated position, which leads to a better direction reproduction, especially for the translated position far away from the receiver region. The results of Figs. 5.9(b) and 5.10(b) also indicate that the sparse solution of the particle velocity based expansion can provide a better quality of localization for the listener than the sparse solution of the sound pressure based expansion. Once again, we notice that the least squares solution and the closed-form solution have same performance, and both are worse than the the sparse solutions of the particle velocity based expansion and the sound pressure based expansion.

From the above analysis, we conclude that the least squares solution of the particle velocity based expansion is identical to the closed-form solution of the sound pressure based expansion, whereas the sparse solution of the particle velocity based expansion has better overall performance than the sparse solution of the sound pressure based expansion.

5.5.3 Discussion

Based on the above simulation results, we give the following comments:

- The particle velocity based expansion is derived directly from the sound pressure based expansion by the gradient calculation. Therefore, the least squares solution of the particle velocity based expansion is mathematically parallel to the closed-form solution of the sound pressure based expansion, both of which aim to reconstruct the original truncated recording by distributing energy throughout all virtual sources and inherit the spatial artifacts caused

by the truncated measurement. This leads to poor reproduction outside the receiver region.

- Sparse solutions can improve the performance of reproduction outside the receiver region due to the fact that most sound fields can be reproduced accurately by a single or a few virtual sources, and thus exhibit sparsity in space. Therefore, provided that the desired sound is sparse in space, the sparse solutions can be used to relax the restriction we mentioned in Section 5.2. We should note that the region of accurate translation in the multiple sources scenario becomes smaller than that in the single source scenario for the sparse solution. Also, the sparse solution is less applicable to highly reverberant fields where the sparsity does not hold.

- Particle velocity contains the direction information of a sound field. For the sparse solutions, we can achieve more accurate sound field reproduction by controlling particle velocity than sound pressure. Furthermore, particle velocity reflects ITD, therefore, velocity vector is directly related to the localization predictor for human perception at low frequencies. Particle velocity is also one of the quantities that determine sound intensity vector (the localization predictor at high frequencies), which reflects another human localization cue of interaural level difference (ILD). Therefore, the sparse solution of particle velocity based expansion is expected to provide an enhanced perceptual experience for the listener.

We examine the perceptual advantages of the proposed sparse solution of particle velocity based expansion by experiments in the next section.

5.6 Experimental Verification

In this section, we evaluate the perceptual performance of the proposed method against the state-of-the-art method in a listening experiment. The experimental methodology is introduced in detail in Section 5.6.1, and then we present the statistical results in Section 5.6.2.

5.6.1 Experimental methodology

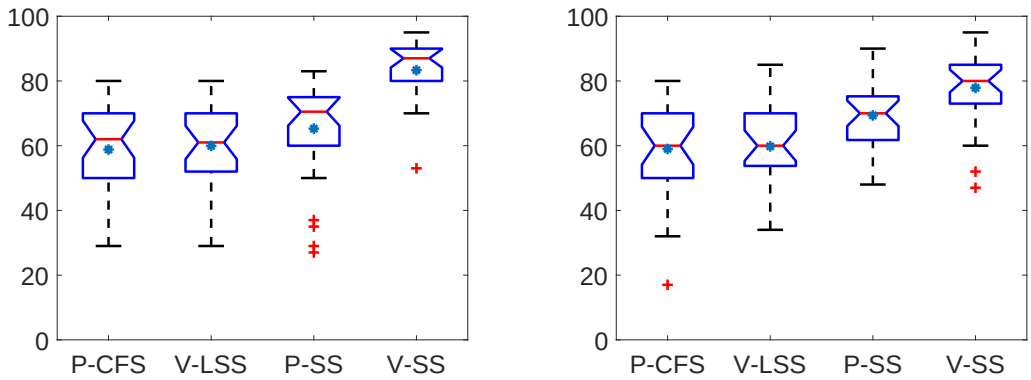
To generate the binaural test signals, we construct a virtual experimental environment, where there is a 4th order receiver located at the origin and a free-field point source located at $(1, 0, 0)$ m, with respect to the origin. The receiver region is a spherical region with a radius of 0.042 m. We assume the microphone array works ideally, and the theoretical pressure coefficients (5.23) can be extracted from the ideal microphone recording. The parameters of the mixed-source model is the same as those in Section 5.5.1. We reproduce the sound image for the listener away from the recording area by extrapolation using the proposed method. The listener faces positive x-axis, and the listener's head is aligned vertically with the point source on the x-y plane. We used the HRTFs of the FABIAN head and torso simulator [189] from the HUTUBS dataset [190, 191]. The source signal is a clip of music with 10 s duration and its spectrum energy mainly distributes in the frequency band below 8 kHz. We process the signals at a frame size of 4096 with 50% overlap and a 16 kHz sampling frequency. We carry out a MUSHRA [192] experiment to compare four translation methods, therefore, for a translated position there are a total of six binaural test signals in this experiment:

- *Reference / Hidden reference*: The ground truth, which is obtained directly from filtering the theoretical point source signal with the HRTFs.
- *Anchor*: Signals of the truncated recording at the origin, which are simulated using the pressure coefficients up to order N in (5.23). No translation is processed, but filtered by the HRTFs.
- *P-CFS*: Signals reconstructed using the closed-form solution of the sound pressure based expansion, and rendered by the HRTFs.
- *V-LSS*: Signals reconstructed using the least squares solution of the particle velocity based expansion, and rendered by the HRTFs.
- *P-SS*: Signals reconstructed using the sparse solution of the sound pressure based expansion, and rendered by the HRTFs.
- *V-SS*: Signals reconstructed using the sparse solution of the particle velocity based expansion, and rendered by the HRTFs.

We generate the above binaural test signals at six different translated positions outside the receiver region using (5.21). All the translated positions are on the x-y plane, and they are $(0, 1, 0)$ m, $(0, -1, 0)$ m, $(0.5, 0.8, 0)$ m, $(0.5, -0.8, 0)$ m, $(1, 1, 0)$ m, and $(1, -1, 0)$ m, respectively. 31 subjects with normal hearing are invited to participate in this listening experiment, and a random set out of the six sets of test signals is selected for each hearing subject. The test signals are played on a computer and the hearing subjects need to wear headphones to listen to the test signals. The experiment includes two tests, and they are the source localization test and the basic audio quality test, respectively. The order in which the test signals are played is also random for each test. For the source localization test, the hearing subjects are asked to score all the test signals against the reference for the perceived direction of the sound source, whereas the hearing subjects need to score on the spectral distortions and other audible processing artifacts with respect to the reference for the basic audio quality test. The score ranges from 0 to 100, and the reference is preseted to 100. We disqualify the hearing subjects who score the hidden reference below 90, and their results are removed from the final data set of scores.

5.6.2 Experimental results

After removing the scores of the disqualified subjects, we collect the scores of 30 subjects in total for the source localization test, and 29 scores for the basic audio quality test. Box plots showing the results for both tests are given in Fig. 5.11. On each box, the central short red line indicates the median value, and the edges of the box are the 25th and 75th percentile of the scores. The extremes of the whiskers correspond to $q_3 + 1.5(q_3 - q_1)$ and $q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentile of the scores, respectively. If a score goes beyond the extremes of the whiskers, it is considered an outlier, indicated by the symbol $+$ in the plots. The symbol $*$ denotes the mean value, and the v-shaped notches refer to the 95% confidence interval. The endpoints of the notches are calculated by $q_2 - 1.57(q_3 - q_1)/\sqrt{N_s}$ and $q_2 + 1.57(q_3 - q_1)/\sqrt{N_s}$, where q_2 is the median value and N_s is the number of the scores. The intervals can be used to compare the median values. If two intervals do not overlap, their true medians are different



(a) (b)

Figure 5.11: Results of the listening experiment scores for the source localization test and the basic audio quality test. Each box bounds the 25th and 75th percentile of the scores with the central red line indicating the median value, and the whiskers are the outward extension of the 25th and 75th percentile by 1.5 times of the interquartile range. The v-shaped notches represent the 95% confidence intervals. The symbols + denote the outliers and the symbols $*$ denote the mean values. (a) Source localization test; (b) Basic audio quality test.

with 95% confidence.

From Fig. 5.11(a), we observe that, for the source localization, the closed-form solution of the sound pressure based expansion and the least squares solution of the particle velocity based mixed-source solution have similar mean values and median values, and their confidence intervals have nearly 100% overlapping. The sparse solutions of both expansions have a higher mean values and median values. However, the sparse solution of the particle velocity based expansion has a more significant improvement in score and more centralized score distribution compared to the sparse solution of the sound pressure based expansion. Similar results are shown for the basic audio quality test in Fig. 5.11(b), where no significant difference is found between the closed-form solution and the least squares solution. For the sparse solutions, the particle velocity based expansion is still observed to perform better than the sound pressure based expansion, however, both of which outperform the closed-form solution and the least squares solution. We also notice that for the sparse solution of the particle velocity based expansion the scores in the source

localization test are relatively higher than those in the basic audio quality test.

According to the above experimental results, we conclude that in the perceptual criteria of source localization and audio quality the sparse solutions have better performance than the closed-formed solution and the least squares solution, both of which have similar performance. In the sparse solutions, the particle velocity based expansion shows an improvement against the sound pressure based expansion, especially for the source localization. The experimental results are consistent with the discussions in Section 5.5.3.

5.7 Summary and Contributions

In this chapter, we have proposed a new mixed-source expansion, which exploits particle velocity to enhance the performance of sound field translation for binaural reproduction. We provide an itemized list of our contributions:

- We built two different cost functions that optimize spatial particle velocity vectors. The solutions to the cost functions were introduced.

- The corresponding two solutions for the state-of-the-art expansion based on sound pressure were also implemented for comparison. The simulation results showed that the least squares solution of the particle velocity based expansion is parallel to the closed-form solution of the sound pressure based expansion, whereas the sparse solution of the particle velocity based expansion can reproduce the sound field over a larger area with less error, and has more accurate direction reproduction at the translated positions than the sparse solution of the sound pressure based expansion.

- We conducted MUSHRA experiments to validate the simulation results. The experimental results proved that for the sparse solution the particle velocity based expansion can provide greater perception accuracy at the translated positions on both source localization and audio quality than the sound pressure based expansion, and the improvement for source localization is particularly significant.

5.8 Related Publications

Much of this chapter's work has been submitted for publication in the following journal paper.

- **H. Zuo**, L. I. Birnie, P. N. Samarasinghe, T. D. Abhayapala, and V. Tourbabin, "Particle velocity based mixed-source sound field translation for binaural reproduction," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 2021 (in peer review).

Chapter 6

Spatial Sound Field Reproduction Based on Sound Intensity

In addition to particle velocity, sound intensity can also be controlled to create a high level of realism to humans in soundfield reproduction systems due to the fact that sound intensity is closely linked with human perception of sound location. In this chapter, we present an intensity matching technique to optimally reproduce sound intensity over a continuous spatial region using an irregular loudspeaker array. This avoids several known limitations in the previous works on intensity based soundfield reproduction, such as a single sweet spot for the listener and a regular loudspeaker geometry that is difficult to implement in real-world applications. In contrast to the previous works, the new technique uses a cost function we built to optimize sound intensity over space by exploiting the spatial sound intensity distributions we proposed in Section 3.3. The spatial sound intensity distribution is represented by spherical harmonic coefficients of sound pressure, which are widely used to describe a spatial soundfield. Compared to the conventional spatial soundfield reproduction method of pressure matching in the spherical harmonic domain (i.e., the mode matching) and the HOA max-r_E decoding method optimizing sound intensity at a single position, we show that the intensity matching technique has better overall performance with two different irregular loudspeaker layouts through simulations. The impact of microphone noise on reproduction performance is also assessed. Finally, we carry out perceptual localization experiments to validate the

proposed method.

6.1 Introduction

Since Gerzon developed particle velocity and sound intensity theories of sound localization for reproducing psychoacoustically optimum sound [64], acoustic quantities containing directivity information, such as particle velocity and sound intensity, have been controlled in soundfield reproduction systems to improve the performance of perceptual localization. In addition to good performance for regularly or evenly placed loudspeakers [66–69], the particle velocity or sound intensity based methods also perform well in irregular loudspeaker arrangements [70–74]. In Chapter 4 and Chapter 5, we have also shown the proposed continuous particle velocity based soundfield reproduction techniques can offer good localization performance, which in particular proves that particle velocity is an important predictor of perceived source directions. This predictor is based on interaural phase cues and is most appropriate at frequencies below 700 Hz [64]. The sound intensity (i.e., the product of the sound pressure and particle velocity) is also a good predictor of localization perception [146]. It was exploited in the max-r_E decoding method [73], which is to reproduce the energy and acoustic intensity of the desired soundfield, and psychoacoustically to create the impression of the desired sound, especially for high frequencies (above 500 Hz). By extending [73], the localization performance was improved by means of HOA [74]. However, these works are all restricted to a center position, and therefore perception degrades when the listener is moved from this exact reproduction position. To overcome this limitation, an Ambisonics based energy vector model was proposed [51], where it shows that the method can provide accurate localization at off-center positions through perceptual experiments. In addition to accurate localization, the improved max-r_E model, proposed by Frank *et al.* [50], can provide less coloration (i.e., changes in the timbre) at off-center positions on the horizontal plane. In this chapter, we aim to reproduce the original sound image over a continuous spatial region by exploiting spatial intensity distributions.

We have formulated spatial sound intensity in the spherical harmonic domain in Section 3.3, which means we can represent sound intensity that contains both

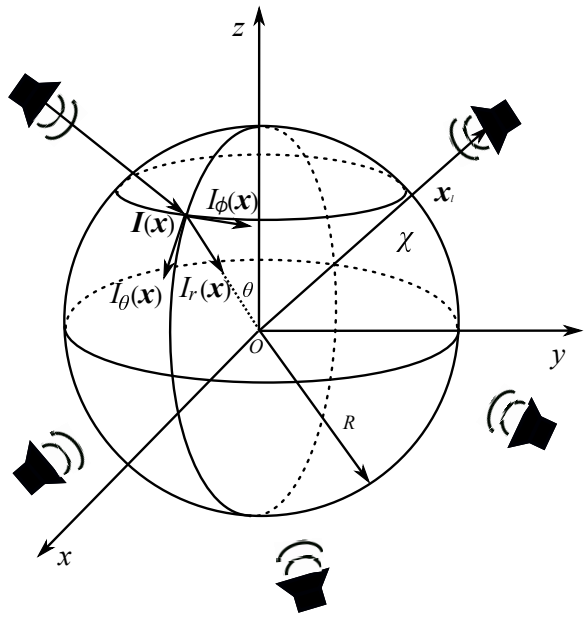


Figure 6.1: Geometry of soundfield reproduction system in this paper. An irregular array of loudspeakers is located outside the spherical region.

energy and directivity information over continuous spatial regions. Therefore, similar to pressure matching in the spherical harmonic domain, intensity matching can also be realized over continuous space instead of at a single point so that the original sound can be perceptually reproduced over a large region for more listeners. In this work, we reproduce the sound intensity inside a continuous spatial region by matching sound intensity on the surrounding contour of the region using an irregular loudspeaker array.

The remainder of this chapter is structured as follows. First we formulate the reproduction problem in Section 6.2. In Section 6.3, the proposed intensity matching method is introduced in detail. Section 6.4 demonstrates the accuracy of the proposed method by comparing it with the pressure matching and the HOA max-r_E decoding through numerical simulations. Perceptual listening test results are presented in Section 6.5.

6.2 Problem Formulation

Consider a spherical target region χ of radius R as seen in Fig. 6.1. Let there be an irregular array of loudspeakers outside of χ, with the ℓth loudspeaker located at $\boldsymbol{x}_\ell$ with respect to the origin O. The free field assumption is made and diffraction/scattering is assumed to be negligible.

The complex acoustic intensity at any point $\boldsymbol{x}$ in χ can be represented by [193]

$$\boldsymbol{I}(\boldsymbol{x}, k) = P^*(\boldsymbol{x}, k)\boldsymbol{V}(\boldsymbol{x}, k), \tag{6.1}$$

where
$\boldsymbol{I}(\boldsymbol{x}, k) = [I_r(\boldsymbol{x}, k), I_\theta(\boldsymbol{x}, k), I_\phi(\boldsymbol{x}, k)]$ and $\boldsymbol{V}(\boldsymbol{x}, k) = [V_r(\boldsymbol{x}, k), V_\theta(\boldsymbol{x}, k), V_\phi(\boldsymbol{x}, k)]$ are sound intensity and particle velocity vector, respectively, $P(\boldsymbol{x}, k)$ is sound pressure. The real part of $\boldsymbol{I}(\boldsymbol{x}, k)$ is often referred to as the active sound intensity, which represents the propagating sound energy and shows the direction of propagation at the point in space. The imaginary part of it, on the other hand, is referred to as the reactive sound intensity, which represents the non-propagating sound energy [194].

A soundfield in a spherical region is usually characterized by spherical harmonic coefficients of sound pressure (which can be extracted by using higher order microphones such as an EigenMike) instead of the spatial distribution of sound intensity[1]. Given spherical harmonic coefficients of sound pressure for the desired soundfield $\{\alpha_{nm}(k)\}$ in the spherical region χ, and an appropriate irregular loudspeaker array geometry, our objective is to calculate the spatial sound intensity (6.1) in χ and find the loudspeaker driving signals that can reproduce the desired sound intensity over the continuous spatial region within χ.

6.3 Loudspeaker Weights Design

We have formulated spatial sound intensity vectors in the spherical harmonic domain in Section 3.3. We have shown that the spatial distribution of sound intensity

[1]The sound intensity can be estimated by measuring the particle velocity over the volume of interest point by point, which is time-consuming [93].

can be expressed directly from the spherical harmonic coefficients of sound pressure.

Considering the spherical harmonic coefficients of sound pressure for the desired soundfield $\{\alpha_{nm}(k)\}$ are given, we can use the results in Section 3.3 to calculate the desired sound intensity at any point in χ. According to the results, the components of desired sound intensity $\boldsymbol{I}^d(\boldsymbol{x},k) = [I_r^d(\boldsymbol{x},k), I_\theta^d(\boldsymbol{x},k), I_\phi^d(\boldsymbol{x},k)]$ at any arbitrary $\boldsymbol{x}$ within χ can be decomposed, respectively, as

$$I_r^d(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p} S_{pq}^{(r,d)}(k,r)Y_{pq}(\theta,\phi), \tag{6.2}$$

$$I_\theta^d(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p} S_{pq}^{(\theta,d)}(k,r)Y_{pq}(\theta,\phi), \tag{6.3}$$

$$I_\phi^d(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p} S_{pq}^{(\phi,d)}(k,r)Y_{pq}(\theta,\phi), \tag{6.4}$$

where $S_{pq}^{(r,d)}(k,r)$, $S_{pq}^{(\theta,d)}(k,r)$, and $S_{pq}^{(\phi,d)}(k,r)$ denote spherical harmonic coefficients of desired sound intensity in r, θ, and ϕ direction obtained by substituting $\alpha_{nm}(k)$ into (3.48), (3.49), and (3.50), respectively, $P = 2N$ is the truncation order for intensity expressions. Note that we select the same truncation order for all the components of sound intensity for simplicity in this work, at the cost of relatively more error on the components in θ and ϕ direction compared to r direction.

The loudspeakers are modeled as point sources, and the spherical harmonic coefficients of sound pressure due to the ℓth loudspeaker in free field is given in (4.6). By replacing $\alpha_{nm}(k)$ with $\beta_{nm}^{(\ell)}(k)$ in (6.2) and (3.48), we obtain sound intensity in r direction due to the ℓth loudspeaker at $\boldsymbol{x}$ as

$$I_r^{(\ell)}(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p} S_{pq}^{(r,\ell)}(k,r)Y_{pq}(\theta,\phi), \tag{6.5}$$

where $S_{pq}^{(r,\ell)}(k,r)$ are spherical harmonic coefficients of sound intensity in r direction due to the ℓth loudspeaker.

In order to drive loudspeakers, we apply a weight to each loudspeaker. There-

fore, the total reproduced sound intensity in r direction at $\boldsymbol{x}$ can be written as

$$I_r^a(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p}\underbrace{\sum_{\ell=0}^{N_L}|w_\ell(k)|^2 S_{pq}^{(r,\ell)}(k,r)}_{S_{pq}^{(r,a)}(k,r)}\, Y_{pq}(\theta,\phi), \tag{6.6}$$

where N_L is the number of loudspeakers. We note that here we assume each loudspeaker is an independent source, and only time-average quantities (i.e., sound intensity) from them add up at the listeners' ears based on the fact that statistically out-of-phase sound signal summation occurs at the ears due to head diffraction [146, 195].

Similar to (6.6), the reproduced sound intensity in θ and ϕ directions at $\boldsymbol{x}$ can be expressed, respectively, as

$$I_\theta^a(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p}\underbrace{\sum_{\ell=0}^{N_L}|w_\ell(k)|^2 S_{pq}^{(\theta,\ell)}(k,r)}_{S_{pq}^{(\theta,a)}(k,r)}\, Y_{pq}(\theta,\phi), \tag{6.7}$$

and

$$I_\phi^a(\boldsymbol{x},k) = \sum_{p=0}^{P}\sum_{q=-p}^{p}\underbrace{\sum_{\ell=0}^{N_L}|w_\ell(k)|^2 S_{pq}^{(\phi,\ell)}(k,r)}_{S_{pq}^{(\phi,a)}(k,r)}\, Y_{pq}(\theta,\phi). \tag{6.8}$$

The reproduction problem is now reduced to calculate the loudspeaker driving signals/weights that can reconstruct the desired sound intensity over the target region. Inspired by Kirchhoff-Helmholtz integral equation, which shows that the soundfield inside a region can be controlled by evoking the sound pressure and the gradient of sound pressure on the surrounding contour of the region, we reproduce the sound intensity inside a spherical region by controlling the sound intensity (i.e., the product of sound pressure and its gradient) on the surface of the region[2]. Therefore, in this work, the problem is translated into minimizing the difference

[2]While sound intensity matching on the surface of the target region is inspired by the Kirchhoff-Helmholtz principle, a formal proof showing that sound intensity matching on the boundary implies intensity matching within the volume is not currently available, and is a subject of future work.

between the desired sound intensity and the reproduced sound intensity on the surface of the target region. The difference for sound intensity in r direction can be given as

$$E_r = \int_0^{2\pi} \int_{-1}^{1} \left| I_r^d(R,\theta,\phi,k) - I_r^a(R,\theta,\phi,k) \right|^2 d(\cos\theta) d\phi. \tag{6.9}$$

Substitute (6.2) and (6.6) into (6.9), we have

$$\begin{aligned} E_r = & \sum_{p=0}^{P} \sum_{q=-p}^{p} \sum_{p'=0}^{P} \sum_{q'=-p'}^{p'} [S_{pq}^{(r,d)}(k,R) - S_{pq}^{(r,a)}(k,R)]^* [S_{p'q'}^{(r,d)}(k,R) - S_{p'q'}^{(r,a)}(k,R)] \\ & \times \int_0^{2\pi} \int_{-1}^{1} Y_{pq}(\theta,\phi)^* Y_{p'q'}(\theta,\phi) d(\cos\theta) d\phi. \end{aligned} \tag{6.10}$$

Equation (6.10) can be simplified, by using the orthogonality of spherical harmonics, as

$$E_r = \sum_{p=0}^{P} \sum_{q=-p}^{p} \left| S_{pq}^{(r,d)}(k,R) - S_{pq}^{(r,a)}(k,R) \right|^2. \tag{6.11}$$

Similarly, the difference for sound intensity in θ and ϕ direction are given, respectively, as

$$E_\theta = \sum_{p=0}^{P} \sum_{q=-p}^{p} \left| S_{pq}^{(\theta,d)}(k,R) - S_{pq}^{(\theta,a)}(k,R) \right|^2, \tag{6.12}$$

and

$$E_\phi = \sum_{p=0}^{P} \sum_{q=-p}^{p} \left| S_{pq}^{(\phi,d)}(k,R) - S_{pq}^{(\phi,a)}(k,R) \right|^2. \tag{6.13}$$

A cost function containing all the components is given by

$$f = c_r E_r + c_\theta E_\theta + c_\phi E_\phi, \tag{6.14}$$

where c_r, c_θ, and c_ϕ are weighting coefficients. Different values of the weighting coefficients can be selected if different weights on the components are preferred. Also, one or two of the weighting coefficients can be set to zero (e.g., if one only focuses on the radial part of sound intensity, c_θ and c_ϕ can be set to zero). To

minimize the cost function f, the problem can be formulated as

$$\min_{\boldsymbol{W}} ||\boldsymbol{S}_A\boldsymbol{W} - \boldsymbol{S}_D||_2^2, \quad \text{s.t.}\boldsymbol{W} \geq 0, \tag{6.15}$$

where $\boldsymbol{S}_D = [\boldsymbol{S}^{(r,d)}(k)^T, \boldsymbol{S}^{(\theta,d)}(k)^T, \boldsymbol{S}^{(\phi,d)}(k)^T]^T$ is a $6(Q+1)^2$ long vector with

$$\boldsymbol{S}^{(\Psi,d)}(k) = \sqrt{c_\Psi}[\text{Re}\{S_{00}^{(\Psi,d)}(k)\}, \text{Im}\{S_{00}^{(\Psi,d)}(k)\}, ..., \text{Re}\{S_{QQ}^{(\Psi,d)}(k)\}, \text{Im}\{S_{QQ}^{(\Psi,d)}(k)\}]^T, \tag{6.16}$$

where $\text{Im}\{\cdot\}$ denotes the imaginary part, $\boldsymbol{W} = [|w_1(k)|^2, |w_2(k)|^2, ..., |w_{N_L}(k)|^2]^T$ is a N_L long vector and $\boldsymbol{S}_A = [\boldsymbol{S}^{(r,a)}(k)^T, \boldsymbol{S}^{(\theta,a)}(k)^T, \boldsymbol{S}^{(\phi,a)}(k)^T]^T$ is a $6(Q+1)^2$ by N_L matrix with

$$\boldsymbol{S}^{(\Psi,a)}(k) = \sqrt{c_\Psi} \begin{bmatrix} \text{Re}\{S_{00}^{(\Psi,1)}(k)\} & \cdots & \text{Re}\{S_{00}^{(\Psi,N_L)}(k)\} \\ \text{Im}\{S_{00}^{(\Psi,1)}(k)\} & \cdots & \text{Im}\{S_{00}^{(\Psi,N_L)}(k)\} \\ \vdots & \ddots & \vdots \\ \text{Re}\{S_{QQ}^{(\Psi,1)}(k)\} & \cdots & \text{Re}\{S_{QQ}^{(\Psi,N_L)}(k)\} \\ \text{Im}\{S_{QQ}^{(\Psi,1)}(k)\} & \cdots & \text{Im}\{S_{QQ}^{(\Psi,N_L)}(k)\} \end{bmatrix}. \tag{6.17}$$

$\boldsymbol{W} \geq 0$ means that each component of the vector $\boldsymbol{W}$ should be non-negative. Note that we deal with the real part and imaginary part of sound intensity separately as $\boldsymbol{W}$ is a real vector.

In mathematical optimization, this problem is known as non-negative least squares (NNLS). A widely used algorithm for solving this problem is the active-set method proposed by Lawson and Hanson in 1974 [196]. Other algorithms such as Landweber's gradient descent method [197] and coordinate-wise optimization method [198] can also be applied to solving this problem. In this work, we solve (6.15) using Lawson and Hanson's method in [196]. We note that the weights obtained from the proposed method is real. To make the system causal and stable, a minimum-phase filter can be achieved from the real weight using a Hilbert transformer [199].

Table 6.1: Loudspeaker positions of the 5 channel system

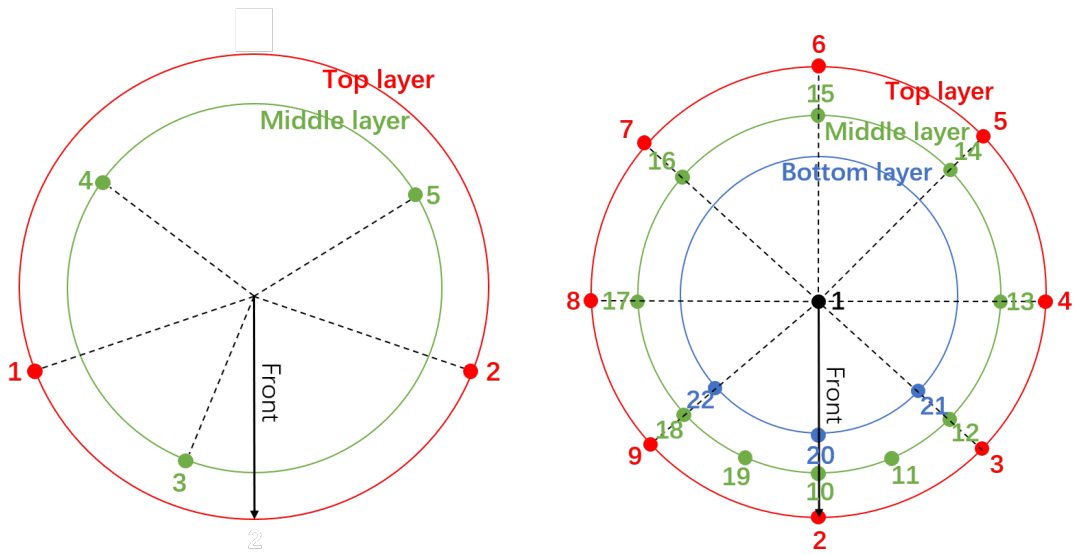
Loudspeaker No.	r [m]	θ [deg]	ϕ [deg]
1	1	58.3	288
2	1	58.3	72
3	1	90	342
4	1	90	234
5	1	90	126

6.4 Simulations

The reproduction performance of the intensity matching, the pressure matching and the HOA max-r_E decoding is now evaluated for a simulated free field environment, where the simulation setup and criteria are explained in Section 6.4.1. Afterward, a comparison for the intensity matching and the pressure matching is conducted in Section 6.4.2, and a comparison for the intensity matching and the HOA max-r_E decoding is shown in Section 6.4.3. Finally, it is investigated in Section 6.4.4 how inherent noise of the microphones, which are required to measure the soundfield in practice, affects the reproduction performance.

6.4.1 Simulation setup and criteria

The target region is a spherical region χ with radius $R = 0.15$ m. To evaluate the overall performance of the proposed method, we simulate two different 3D loudspeaker layouts. One is composed of 5 separate loudspeakers selected from a spherical array that exists in our lab (also for the perceptual experiments in Section 6.5) and their positions are listed in Table 6.1, which is used to evaluate the performance of the proposed method when there are only a small number of loudspeakers available. Note that the loudspeaker number of this layout is too small to cover most of the 3D space, therefore, it may lead to poor reproduction performance for some incident directions of the desired sound. Another loudspeaker layout is based on the cylindrical layout of 22.2 channel system, which is same as the layout we used in Chapter 4. The loudspeaker positions of the 22 channel system is given in Table 4.1. A 2D sketch for both loudspeaker layouts is given in Fig. 6.2. The 22-channel system . We assume that the desired soundfield is a



(a) (b)

Figure 6.2: 2D sketch of (a) the 5 channel system layout and (b) the 22 channel system layout.

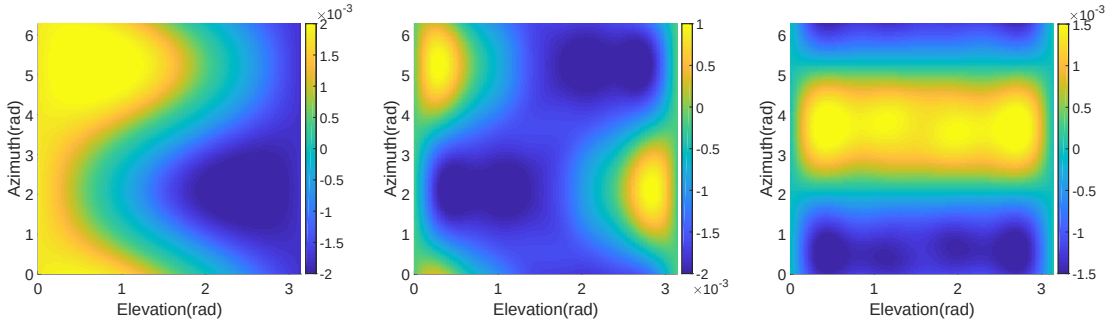
monochromatic plane wave. The spherical harmonic coefficients of sound pressure for the desired soundfield can be expressed by (2.19)

$$\alpha_{nm} = 4\pi(-i)^n Y^*_{nm}(\theta_d, \phi_d), \tag{6.18}$$

where (θ_d, ϕ_d) is the incident direction of the plane wave. Sound speed c is 343 m/s and air density is 1.29 kg/m^3. We set weighting coefficients $c_r = c_\theta = c_\phi = 1$. As discussed in Section 6.2, the active sound intensity represents the propagating sound energy and shows the propagation energy, therefore, we only evaluate the real part of the complex acoustic intensity in the following simulations.

As the first objective performance measure for the reproduction methods, the relative error ϵ of sound intensity is evaluated

$$\epsilon(k) = 10\log_{10}\left(\frac{\sum_{\forall\widehat{\boldsymbol{x}}}\left|\mathrm{Re}\{I^d_\Psi(\widehat{\boldsymbol{x}},k) - I^a_\Psi(\widehat{\boldsymbol{x}},k)\}\right|^2}{\sum_{\forall\widehat{\boldsymbol{x}}}\left|\mathrm{Re}\{I^d_\Psi(\widehat{\boldsymbol{x}},k)\}\right|^2}\right)\ (\mathrm{dB}), \tag{6.19}$$



(a) (b) (c)

Figure 6.3: Desired sound intensity in (a) r direction, (b) θ direction, and (c) ϕ direction induced by a plane wave with frequency of 900 Hz.

where $\widehat{\boldsymbol{x}}$ denotes the evaluated point. To reproduce a soundfield with a realistic perception of the sound, it requires to ensure the reproduction of the direction of the sound. Therefore, the intensity direction error η is also defined to show the difference in angles between two vectors

$$\eta(\widehat{\boldsymbol{x}}, k) = \cos^{-1}(\mathrm{DOT})/\pi \times 100(\%), \tag{6.20}$$

where

$$\mathrm{DOT} = \frac{\mathrm{Re}\{\boldsymbol{I}^a(\widehat{\boldsymbol{x}}, k)\}}{\parallel \mathrm{Re}\{\boldsymbol{I}^a(\widehat{\boldsymbol{x}}, k)\} \parallel_2} \cdot \frac{\mathrm{Re}\{\boldsymbol{I}^d(\widehat{\boldsymbol{x}}, k)\}}{\parallel \mathrm{Re}\{\boldsymbol{I}^d(\widehat{\boldsymbol{x}}, k)\} \parallel_2}, \tag{6.21}$$

with the desired intensity vector $\boldsymbol{I}^d(\widehat{\boldsymbol{x}}, k) = [I_r^d(\widehat{\boldsymbol{x}}, k), I_\theta^d(\widehat{\boldsymbol{x}}, k), I_\phi^d(\widehat{\boldsymbol{x}}, k)]$ and the reproduced intensity vector $\boldsymbol{I}^a(\widehat{\boldsymbol{x}}, k) = [I_r^a(\widehat{\boldsymbol{x}}, k), I_\theta^a(\widehat{\boldsymbol{x}}, k), I_\phi^a(\widehat{\boldsymbol{x}}, k)]$, respectively, at $\widehat{\boldsymbol{x}}$.

The following simulations study the performance of the proposed method, in terms of the reconstructed sound intensity, incident direction of the desired sound, frequency and reproduced sound direction. A plane wave with frequency of 900 Hz from $(2\pi/9, 5\pi/3)$ is the desired soundfield throughout the simulations, except the evaluation in terms of the incident direction and frequency.

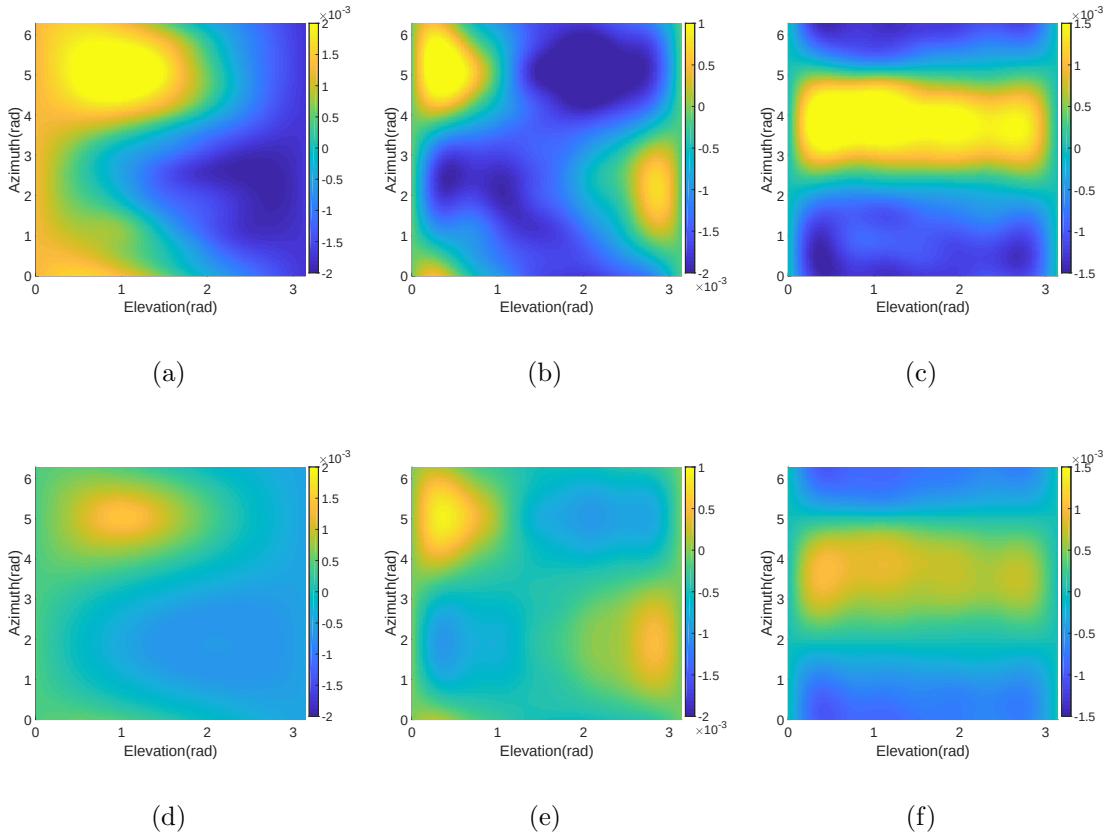


Figure 6.4: Reconstructed sound intensity in (a, d) r direction, (b, e) θ direction, and (c, f) ϕ direction controlled by (a-c) the proposed method, and (d-f) the pressure matching method in case of the 5 channel system. The source frequency is 900 Hz.

6.4.2 Performance comparison for the intensity matching and the pressure matching

Reproduced sound intensity

We first evaluate the reproduction of sound intensity on the surface of the target region. Figures 6.3(a), 6.3(b), and 6.3(c) show the desired sound intensity in r, θ, and ϕ direction due to the plane wave, respectively. Figure 6.4 and Fig. 6.5 show the reconstructed sound intensity controlled by the intensity matching and the pressure matching using the 5 and 22 channel systems, respectively. Observe

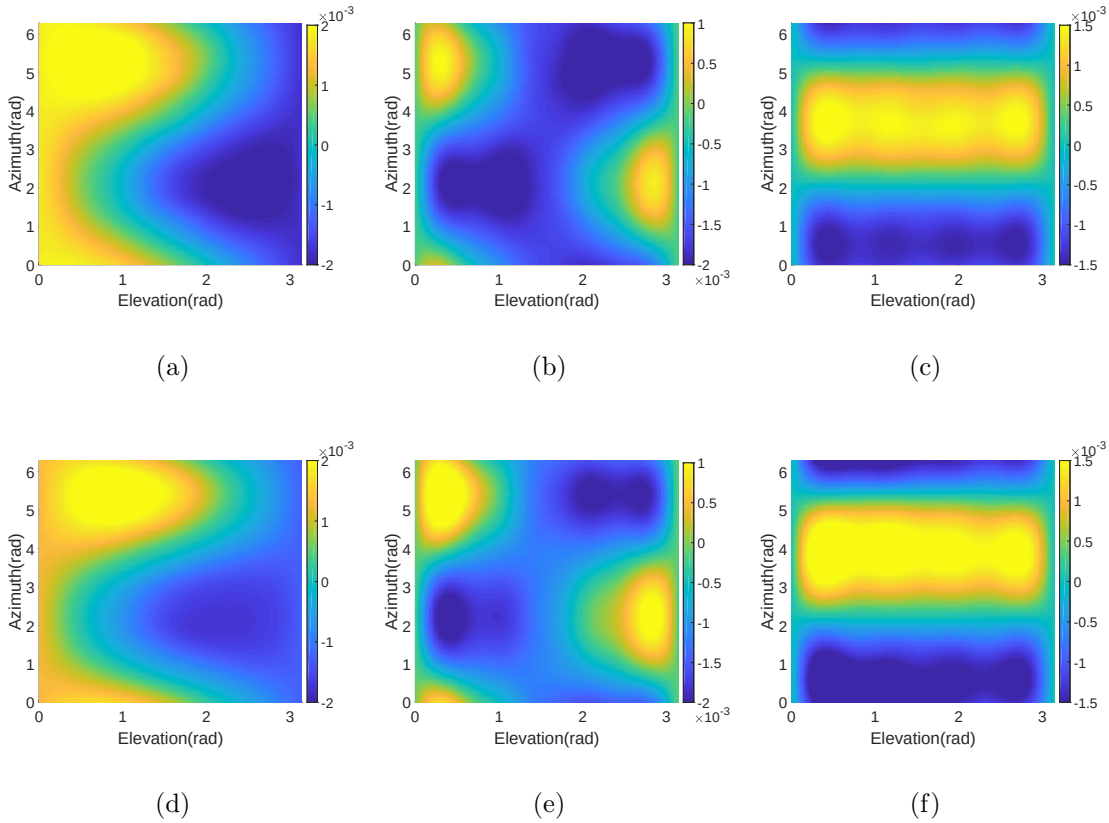
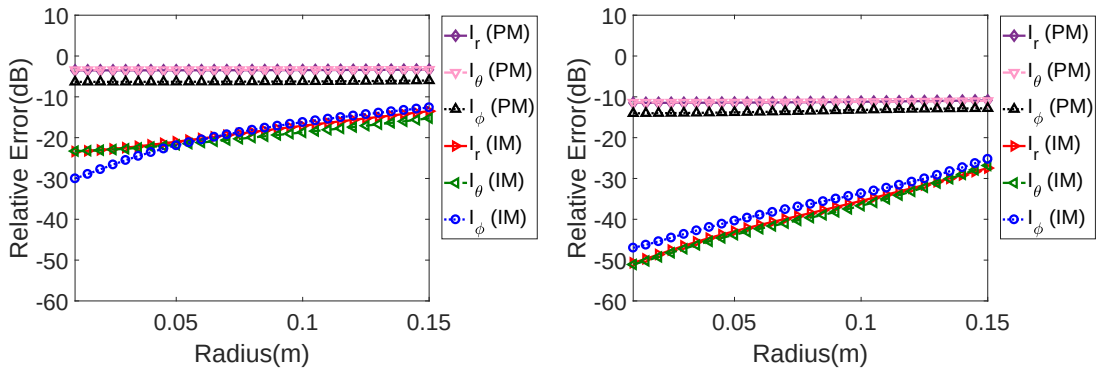


Figure 6.5: Reconstructed sound intensity in (a, d) r direction, (b, e) θ direction, and (c, f) ϕ direction controlled by (a-c) the proposed method, and (d-f) the pressure matching method in case of the 22 channel system. The source frequency is 900 Hz.

that, for the 5 channel system, the proposed method has much better performance than the pressure matching on the surface of the target region. As the number of channels increases, both of the methods perform better. However, the reconstructed sound intensity using the proposed method is still closer to the desired intensity than the pressure matching. The relative error on different spherical surfaces within the target region, which is calculated by (6.19), is also analyzed. The results are given in Fig. 6.6. For the 5 channel system, the relative error of the pressure matching for all the components of sound intensity is around -5 dB in the target region, which is much larger than that of the intensity matching. That is because the matrix $\boldsymbol{\beta}$ in (2.44) for the 5 channel system is ill-conditioned. However, in



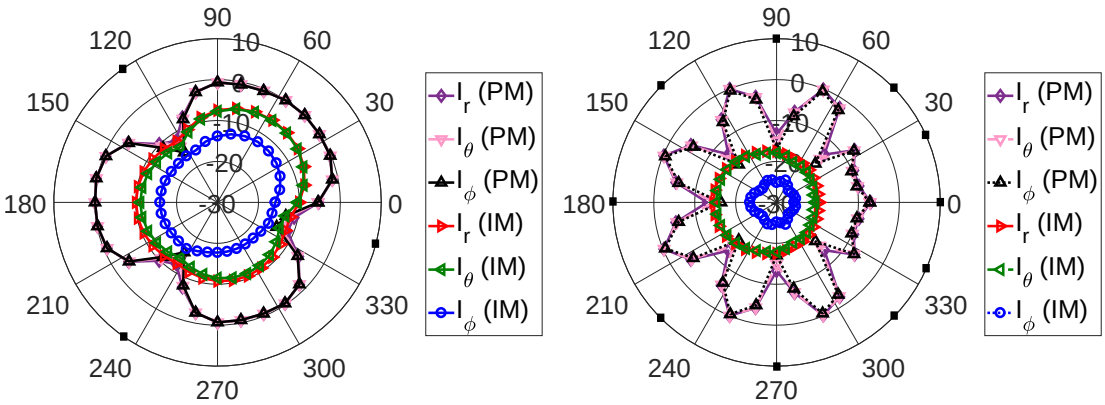
(a) (b)

Figure 6.6: Relative error with respect to the radius for (a) the 5 channel system and (b) the 22 channel system controlled by the intensity matching (IM) and the pressure matching (PM). The source frequency is 900 Hz.

this case, the intensity matching can provide a relatively accurate sound intensity reproduction with the error no more than -10 dB within the target region. When more channels are available, the pressure matching performs better and the relative error for the whole region is around -10 dB. Meanwhile, the performance of the intensity matching is also improved, and the relative error is below -25 dB within the target region. We note that, for both the 5 and 22 channel systems, the performance degrades as the radius increases because the active modes of spherical harmonics also increase when the radius increases. In this work, the system truncation limit is determined by the radius of the target region and the signal frequency, therefore, the error on the surrounding contour is larger than that on the spherical surface inside the region due to the property of spherical Bessel function.

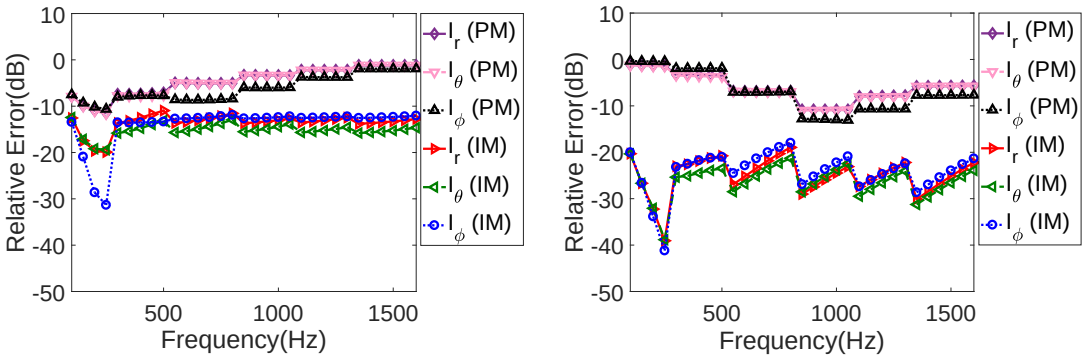
Incident direction of the desired sound

To investigate the influence on the performance of the two methods for different incident directions of the desired soundfield, we consider all the incident directions on the plane $z = 0$ (i.e., $\theta_d = \pi/2$, and ϕ_d is from 0 to 2π). We choose the plane with the maximum error ($R = 0.15$ m) as the observation sphere. Polar plots showing the relative error are given in Fig. 6.7, where symbol ■ denotes the loudspeakers



(a) (b)

Figure 6.7: Relative error on the surface of the target region with respect to the incident directions for (a) the 5 channel system and (b) the 22 channel system controlled by IM and PM. Symbol ■ denotes the loudspeakers placed on the plane $z = 0$. The source frequency is 900 Hz.



(a) (b)

Figure 6.8: Relative error on the surface of the target region with respect to the frequency for (a) the 5 channel system and (b) the 22 channel system controlled by IM and PM.

placed on the plane $z = 0$. From Fig. 6.7, it shows that the intensity matching and the pressure matching have similar performance when the incident direction is coincident with the location of one of the loudspeakers. Except for this case,

the intensity matching performs better than the pressure matching at any other evaluated incident directions for both the loudspeaker layouts. An improvement of about 10 dB is obtained when the number of loudspeakers is increased from 5 to 22. We note that it has relatively better performance for the directions where there are more loudspeakers for both the methods. For example, in Fig. 6.7(b), there are more loudspeakers located at the positions with ϕ_d from 0° to 60° and from 300° to 360°, therefore, there is less error for the plane waves from these two direction intervals on plane $z = 0$. An irregular loudspeaker array cannot well cover all the incident directions, which leads to relatively poor performance for the directions at which there are few loudspeakers. Moreover, the relative error of the pressure matching decreases sharply when the incident direction changes to be coincident with a loudspeaker location. For the intensity matching, it changes more smoothly, which can avoid the sudden deterioration of reproduction performance when the desired incident direction changes. Note that the patterns shown in Fig. 6.7 are dependent of the loudspeaker layouts.

Frequency

Figure 6.8 shows the relative error with respect to the change of frequency. The intensity matching is still better than the pressure matching within the frequency range evaluated for both the loudspeaker layouts. For the pressure matching, the system is either over-determined or under-determined in the loudspeaker arrangements discussed in this work, which causes more error on the reproduction of sound intensity. However, the intensity matching optimizes the sound intensity directly using a cost function, which shows a better performance. Within a particular truncation limit N, similar to the increase of the radius, the relative error also increases as the signal frequency increases because of the property of spherical Bessel function. When the frequency increases to a value that makes the truncation limit become $N + 1$, the relative error decreases compared to the error at the last frequency value, which makes the curves for the intensity matching sawtooth-shaped in Fig. 6.8. Also, the 22 channel system provides much less reproduction error in a larger frequency range than the 5 channel system, which indicates that we can achieve a small reproduction error in a large frequency range by increasing the

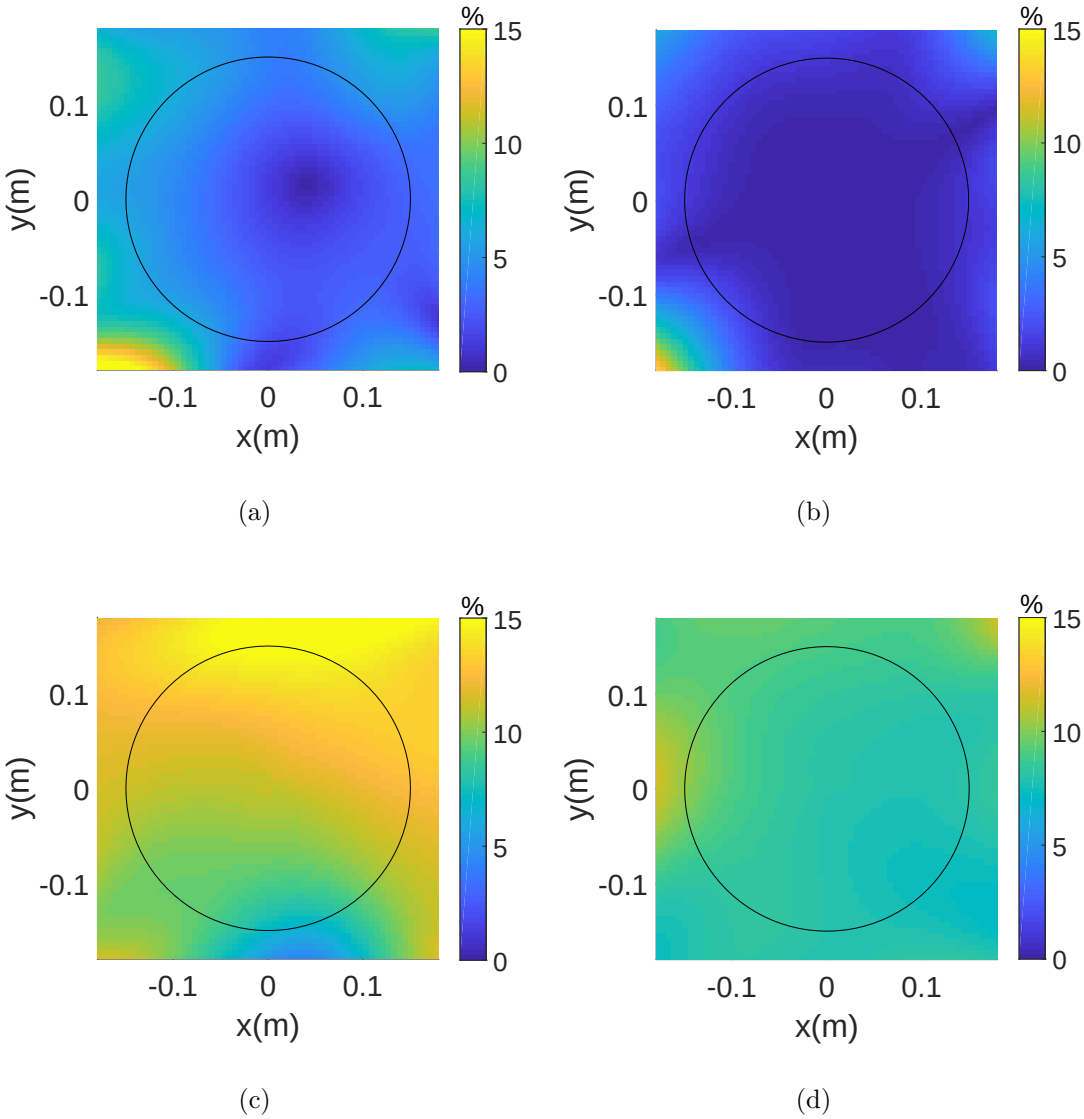


Figure 6.9: The intensity direction error of (a,c) the 5 channel system and (b,d) the 22 channel system on plane $z = 0$ controlled by (a-b) IM and (c-d) PM. Black circles denote the target region. The source frequency is 900 Hz.

number of loudspeakers.

Intensity direction

The discussion above has shown that all the components of sound intensity can be reproduced with limited error using the intensity matching. However, it is clear that if we are concerned to create a realistic perception of the original sound, it requires to ensure the reproduction of the direction of travel of the sound. Therefore, we calculate the intensity direction error on plane $z = 0$ using (6.20), which is shown in Fig. 6.9. The black circles denote the target region on plane $z = 0$. For the 5 channel system, the intensity direction error of the pressure matching is more than 10% for the whole region, whereas it is only approximately 5% within the target region for the intensity matching. When the number of channels increases to 22, the intensity direction error of the pressure matching is decreased to less than 10%, and the performance of the intensity matching also becomes better, with the error around 2%. It also shows that the intensity direction error inside the target region is less than that on the surrounding contour of the region for the intensity matching, which is consistent with the result of the relative error with respect to the radius.

6.4.3 Performance comparison for the intensity matching and the HOA max-r_E decoding

We also compare the proposed method with the HOA max-r_E decoding [73, 74], which optimizes the energy and acoustic intensity at the origin. In the HOA max-r_E decoding, the driving signals fed to loudspeakers are given as

$$\boldsymbol{G} = \boldsymbol{D}\boldsymbol{\alpha}, \tag{6.22}$$

where $\boldsymbol{D}$ is the decoding matrix. The max-r_E decoding also assumes incoherent superposition of loudspeaker signals, and the loudspeakers are assumed as points sources. Therefore, the signal energy at the origin is given as

$$E = \sum_{i=1}^{N_L} \frac{|w_i|^2}{(4\pi d_i)^2}, \tag{6.23}$$

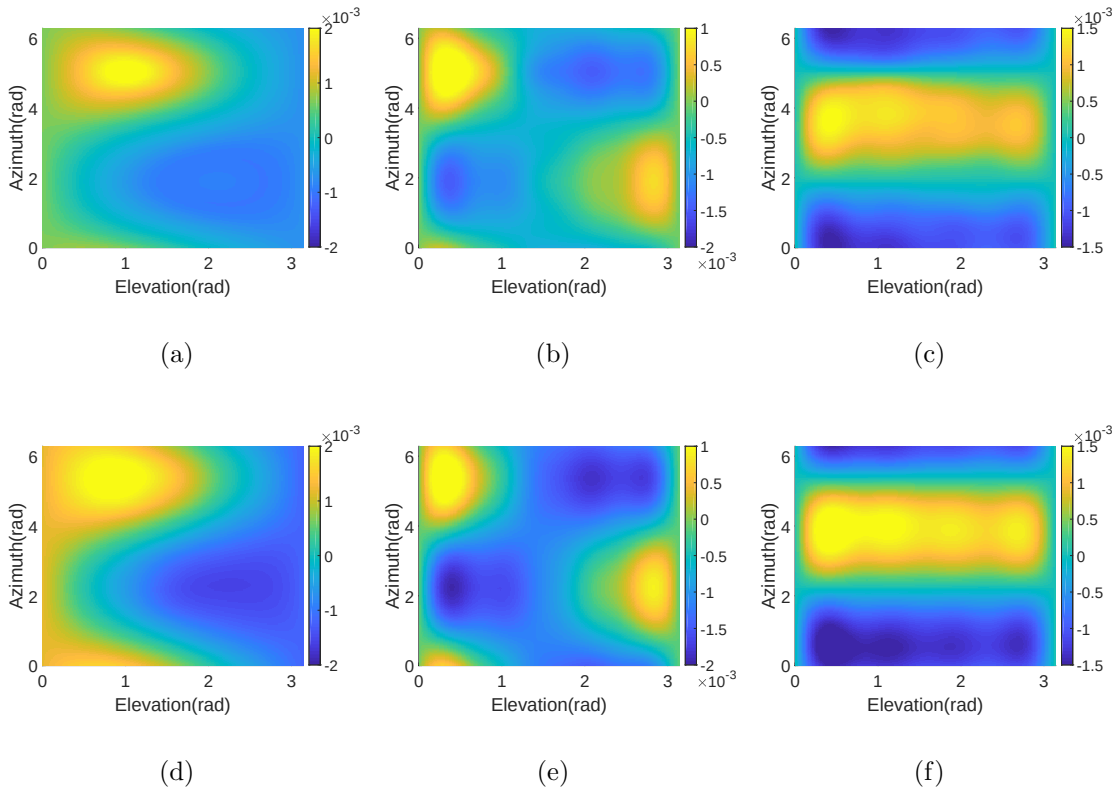


Figure 6.10: Reconstructed sound intensity in (a, d) r direction, (b, e) θ direction, and (c, f) ϕ direction controlled by the HOA max-r_E decoding in case of (a-c) the 5 channel system, and (d-f) the 22 channel system. The source frequency is 900 Hz.

where d_i is the distance from the ith loudspeaker to the origin. The normalized sound intensity at the origin can be written as

$$\begin{aligned} \boldsymbol{I} &= \frac{1}{E}\sum_{i=1}^{N_L}\frac{|w_i|^2}{(4\pi d_i)^2}\widehat{\boldsymbol{u}}_i \\ &= r_E\widehat{\boldsymbol{v}}_E, \end{aligned} \tag{6.24}$$

where $\widehat{\boldsymbol{u}}_i$ is the unit vector corresponding to the position of the ith loudspeaker, $\widehat{\boldsymbol{v}}_E$ is the unit vector representing the direction of the reproduced sound. This decoding is to maximize the value of r_E with the condition that the direction of the reproduced sound is the direction of the desired sound. The cost function in [73]

is used, without the energy and in-phase components,

$$f = (1 - \boldsymbol{I} \cdot \widehat{\boldsymbol{v}}_j)^2 + (||\boldsymbol{I} \times \widehat{\boldsymbol{v}}_j||)^2, \tag{6.25}$$

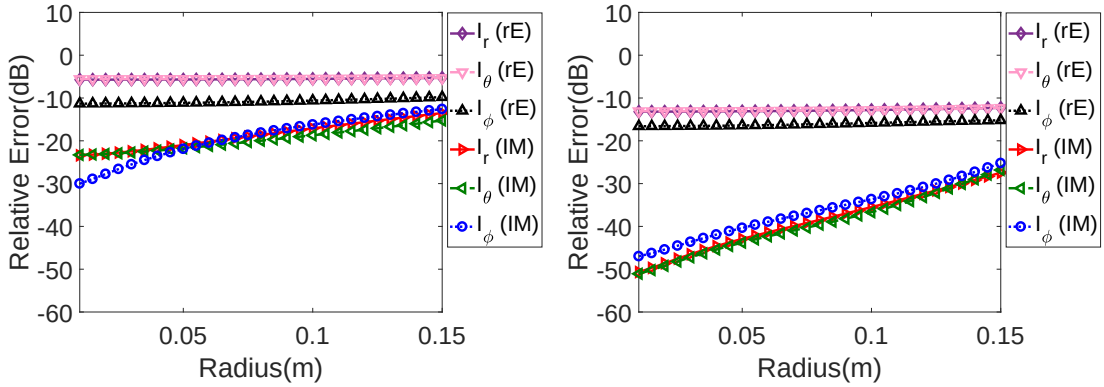
where $\widehat{\boldsymbol{v}}_j$ is the unit vector representing the direction of the desired sound. We note that (6.25) is independent of frequency. This non-linear problem is solved by simulated annealing [73], which is a probabilistic technique, inspired from annealing in metallurgy, for approximating the global optimum of a given function.

Reproduced sound intensity

We also first evaluate the reproduction of sound intensity on the surface of the region. Fig. 6.10 shows the reconstructed sound intensity controlled by the HOA *max*-r_E decoding for the 5 channel loudspeaker layout and the 22 channel loudspeaker layout, respectively. It shows that the intensity matching has better performance than the HOA *max*-r_E decoding for both loudspeaker geometries, by comparing the results with the reconstructed sound intensity controlled by the intensity matching. The relative error of sound intensity on different spherical surfaces within the target region is given in Fig. 6.11. We notice that the performance of the HOA *max*-r_E decoding is also worse than that of the intensity matching, although it is slightly better than the pressure matching.

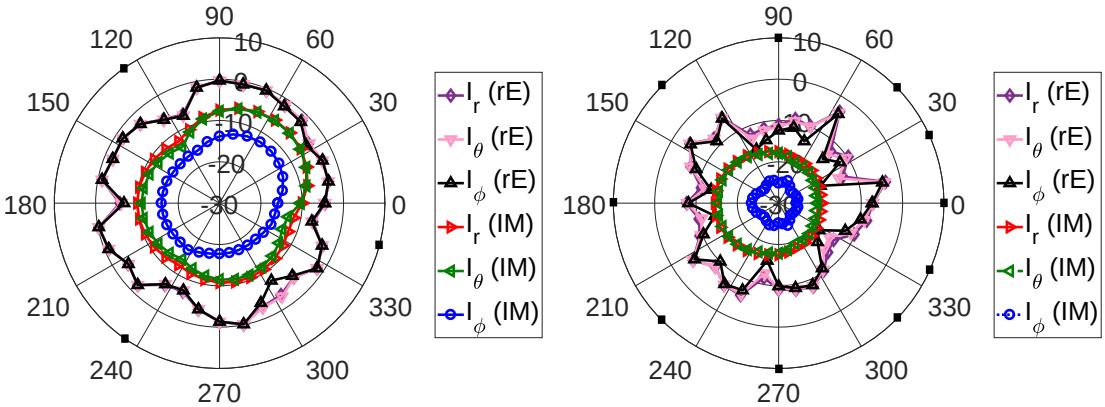
Incident direction of the desired sound

The same incident directions as in Section 6.4.2 are evaluated for the HOA *max*-r_E decoding, also compared with the intensity matching. The results are given in Fig. 6.12. In the 5 channel system, the intensity matching is approximately 10 dB better than the HOA *max*-r_E decoding for the evaluated incident directions. When the number of loudspeaker increases to 22, the performance of the HOA *max*-r_E decoding also becomes better, however, it is not as good as the intensity matching. Different from the pressure matching, the HOA *max*-r_E decoding does not have a significant improvement for the incident directions that are coincident with the location of one of the loudspeakers compared to other incident directions. The HOA *max*-r_E decoding has better performance than the pressure matching for most of



(a) (b)

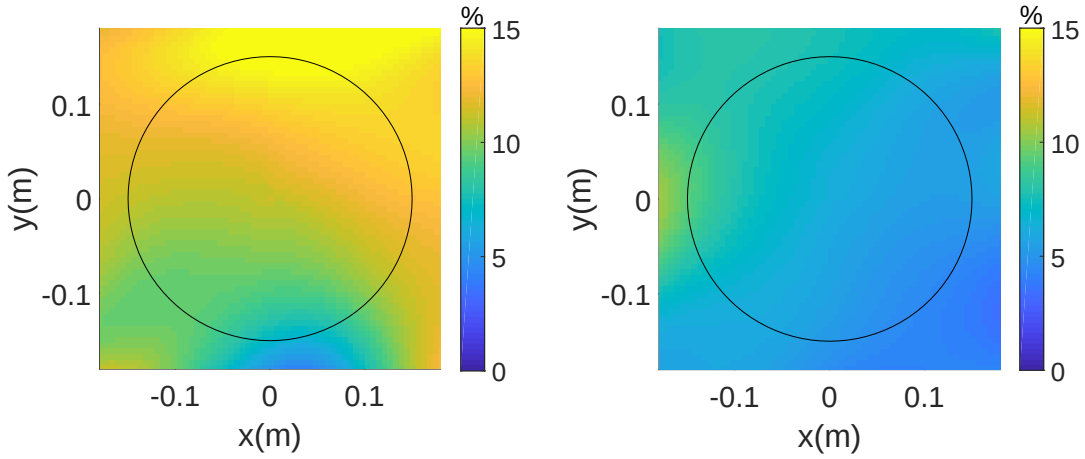
Figure 6.11: Relative error with respect to the radius for (a) the 5 channel system and (b) the 22 channel system controlled by the intensity matching (IM) and the HOA max-r_E decoding (rE). The source frequency is 900 Hz.



(a) (b)

Figure 6.12: Relative error on the surface of the target region with respect to the incident directions for (a) the 5 channel system and (b) the 22 channel system controlled by IM and rE. Symbol ■ denotes the loudspeakers placed on the plane $z = 0$. The source frequency is 900 Hz.

the non-coincident directions. In terms of the smoothness of the error curves, the intensity matching also performs better than the HOA max-r_E decoding.



(a) (b)

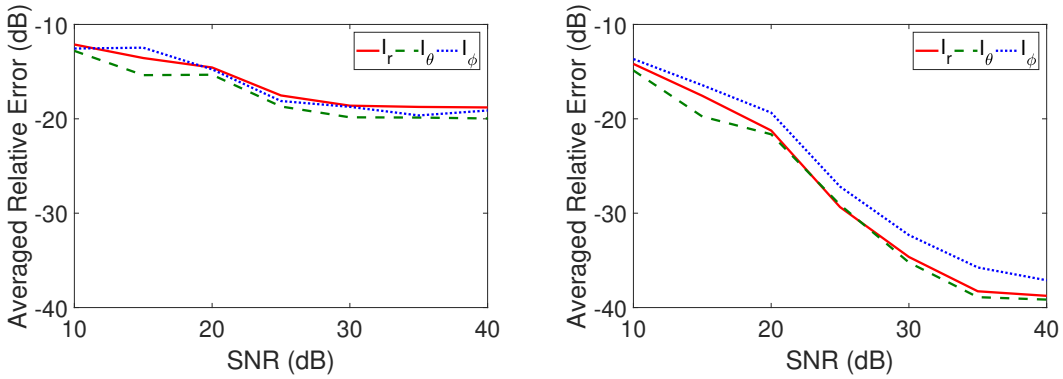
Figure 6.13: The intensity direction error of (a) the 5 channel system and (b) the 22 channel system on plane $z = 0$ controlled by the HOA max-r_E decoding. Black circles denote the target region. The source frequency is 900 Hz.

Intensity direction

Intensity direction error of the HOA max-r_E decoding is also calculated for both loudspeaker layouts, which is shown in Fig. 6.13. We compare the results with the intensity direction error of the intensity matching and the pressure matching correspondingly. For both loudspeaker geometries, the intensity matching is much better than the HOA max-r_E decoding. It is important to note that the performance of the HOA max-r_E decoding is similar to the performance of the pressure matching for the 5 channel system. However, the HOA max-r_E decoding is better than the pressure matching for the 22 channel system.

6.4.4 Robustness analysis to microphone noise

All above simulations are based on theoretical spherical harmonic coefficients of sound pressure of the desired soundfield. In practice, however, we calculate the coefficients based on sound pressure measured by a microphone array, which may introduce noise. To simulate this process, a spherical microphone array is consid-



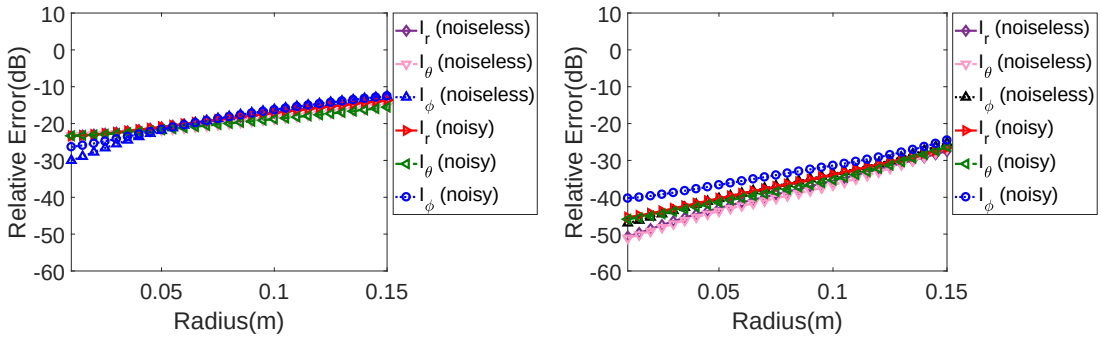
(a) (b)

Figure 6.14: Averaged relative error across the entire target region with respect to SNR for (a) the 5 channel system and (b) the 22 channel system. The source frequency is 900 Hz.

ered and a WGN with different SNR is added at each microphone. The truncation limit is $N = 4$ for this case. Theoretically, a fourth order system should only require $(N+1)^2 = 25$ microphones on the surface of the target region. More accurate results are obtained, however, if more microphones are used. Therefore, we use a total of $(N+2)^2 = 36$ microphones placed on the surface of the 0.15 m sphere at location determined by [118]. The noisy pressure coefficients can be calculated by

$$\widehat{\alpha}_{nm}(k) = \frac{1}{j_n(kR)} \sum_{n_M=1}^{N_M} \widehat{P}(r_M, \theta_M, \phi_M, k) Y^*_{nm}(\theta_M, \phi_M), \tag{6.26}$$

where $\widehat{P}(r_M, \theta_M, \phi_M, k)$ is the sound pressure measured at the Mth microphone, and $N_M = 36$ is the number of microphones. We first assess the proposed method by varying the SNR. The relative error is averaged across the entire target region and the results for both loudspeaker layouts are shown in Fig. 6.14. We note that the averaged relative error is less than -10 dB for both loudspeaker layouts even though the SNR is as low as 10 dB, and the average relative error decreases as the SNR increases. To compare the performance on different observation sphere for the noiseless and noisy conditions, we calculate the relative error by adding a WGN with SNR of 30 dB to the microphones. The results are given in Fig. 6.15. For the



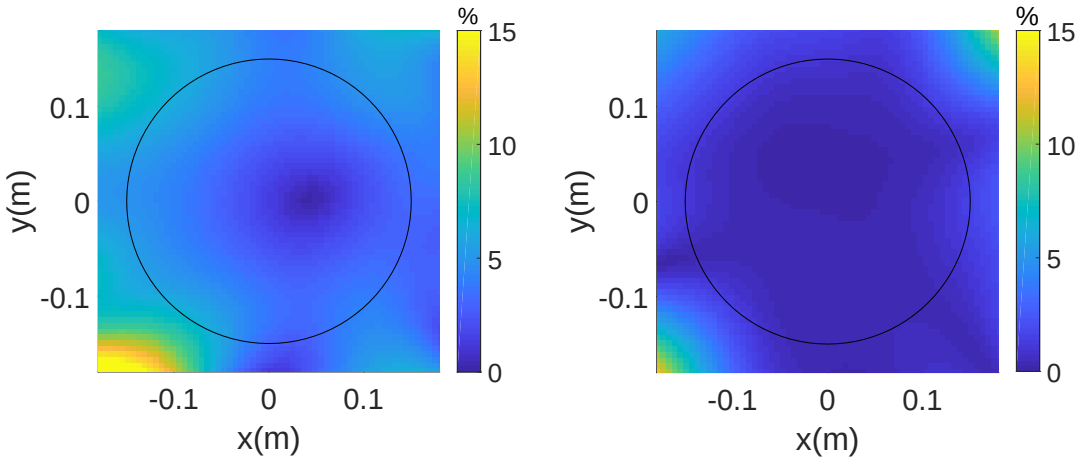
(a) (b)

Figure 6.15: Relative error with respect to the radius for (a) the 5 channel system and (b) the 22 channel system under the noiseless and noisy conditions. The source frequency is 900 Hz.

5 channel system, the curves for the noiseless and noisy cases are nearly coincided with each other, which means 30 dB WGN has little influence on it. For the 22 channel system, there is 5 dB deterioration near the center of the region for all the components of sound intensity when noise is introduced, however, the deterioration becomes less as the radius of the observation sphere increases. We also calculate the intensity direction error, which is shown in Fig. 6.16. By comparing them with Fig. 6.9(a) and Fig. 6.9(b), we observe that the intensity direction error does not change a lot when the noise with SNR of 30 dB exists. These results indicate that microphone noise is not a severe problem for the proposed method in practice. Note that the proposed method does not work any more if the energy of the noise is high enough (e.g., 0 dB) to ruin the pressure measurements.

6.5 Perceptual Localization Experiments

The 5 channel system is built for perceptual experiments in this section. Section 6.4 has shown the HOA max-r_E decoding has similar performance to the pressure matching in the 5 channel system, therefore, we only assess the error between the actual directions and the perceived directions of sound sources as rendered by the intensity matching and the pressure matching in the experiments.



(a) (b)

Figure 6.16: The intensity direction error of (a) the 5 channel system and (b) the 22 channels when noise exists. The source frequency is 900 Hz.

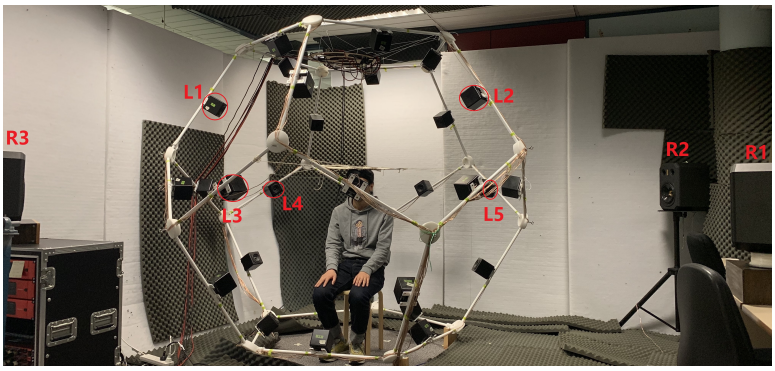


Figure 6.17: Experimental setup for the 5 channel system.

Table 6.2: The locations of three desired sound sources

Label	r [m]	θ [deg]	ϕ [deg]
R1	3	90	60
R2	3	90	126
R3	3	90	306

Table 6.3: CMOS scale

The direction of A compared to that of B is	Score
much (more than 50°) closer to the reference	3
(more than 30°) closer to the reference	2
slightly (more than 10°) closer to the reference	1
(less than or equal to 10°) about the same	0
slightly (more than 10°) further to the reference	-1
(more than 30°) further to the reference	-2
much (more than 50°) further to the reference	-3

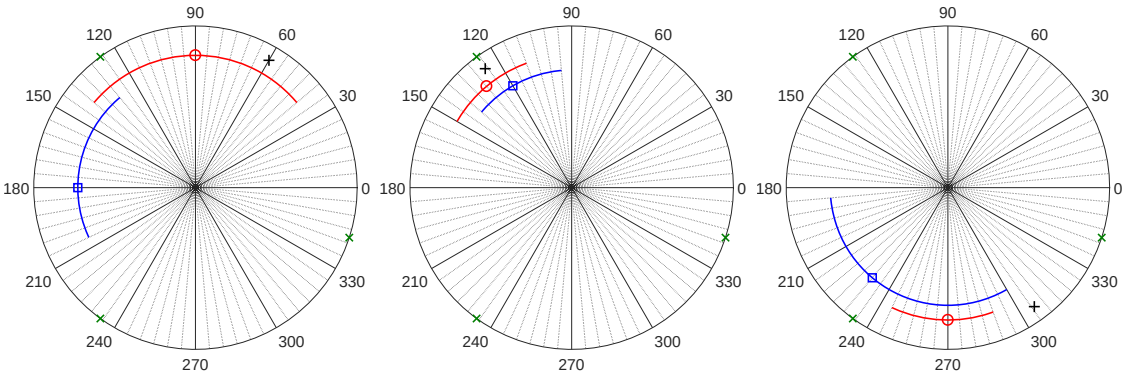
6.5.1 Experimental setup and criteria

The 5 channel system in Table 6.1 is built from a spherical loudspeaker array as shown in Fig 6.17, where only the five loudspeakers (circled and marked from L1 to L5) are activated and all the other loudspeakers are turned off. These five loudspeakers are used to play back stimuli generated by the intensity matching method and the pressure matching method. The center of the spherical array is set as the origin, and the direction where the listening subject faces is the positive x axis. The target region χ is a spherical region centered by the origin with radius of $R = 0.15$ m, which is large enough to fit a human head. There are another three loudspeakers (marked from R1 to R3) located outside of the spherical array, whose position is given in Table 6.2. They serve as the desired sound sources/reference sources. All the loudspeakers are calibrated to produce the same sound pressure level within an error of ± 1 dB. Sound absorbing materials are used in walls, floor and ceiling to reduce reflection. Diffraction/scattering is assumed to be negligible. This experiment is composed of three blind listening tests. For each test, the test samples consist of A, B and a reference, where the reference is the actual sound signal from the reference sources, and A and B are reproduced by the 5 channel system using the intensity matching method and the pressure matching method, respectively. The hearing subjects do not have the knowledge of which method produces A, B, or the reference, and the hearing subjects are instructed to put their heads at the center and face the positive x direction. During the period when the test samples are presented, the hearing subjects are allowed to slightly move their heads from the center. In each test, the hearing subjects are required to

listen to all of the three test samples, and are given 10 s to record their perceived directions for each test sample from 72 angles ((θ,ϕ)=(90°, 5° $\times$ $(n-1)$), where $n = 1, ..., 72$) on a provided answer sheet after listening to the samples. Note that all the reference sources here are placed on the horizontal plane and therefore the hearing subjects are only required to record the azimuth angles to facilitate the implementation of the experiment. Each test sample is played twice, and there is a 2 s of pause between every two samples. A clean speech of 5 s duration is used as a source signal and the sample rate adopted is 48 kHz for signal inputs to all of the loudspeakers. The reference is used not only to compare with A and B, but also to identify an outlier within a data set (i.e., if the reference direction that a hearing subject perceives greatly deviates from its actual direction, the result from this subject is removed from the data set). To straightforwardly show the preferences (between A and B) of the hearing subjects, we compare the perceived directions of A and B with the perceived direction of the reference, respectively, and evaluate the performance using Comparison Mean Opinion Score (CMOS). The score has 7 levels, which is shown in Table 6.3. The positive scores indicate that A is preferred, whereas the negative scores indicate that B is preferred. For the positive scores, the higher the CMOS score is, the better the intensity matching is versus the pressure matching, and vice versa.

6.5.2 Experimental results

17 subjects with normal hearing are recruited to participate in the listening tests. A valid data set containing 15 subjects is obtained after the outliers are removed. Figure 6.18 shows 10-90 percentile ranks that are described by lines for different desired sound sources. The symbols (◯ and □) on the lines indicate the 50 percentile rank or median value. The red lines with ◯ represent the subjective results of the intensity matching method, whereas the blue lines with □ represent the subjective results of the pressure matching method. The percentile rank is used when there is a relatively small number of subjects because it does not consider any predefined distribution of the sample data set, which is regarded as a more representative expression [71]. As shown in Fig. 6.18(a), the test result controlled by the intensity matching shows much closer localization performance in terms of median value,



(a) (b) (c)

Figure 6.18: Results of the perceptual localization experiments in the case of the 5 channel system controlled by the intensity matching (red lines with ◯) and the pressure matching (blue lines with □) for desired sound source (a) R1, (b) R2, and (c) R3. The symbols + denote the source positions and the symbols × denote the loudspeakers on the plane $z = 0$ of the 5 channel system.

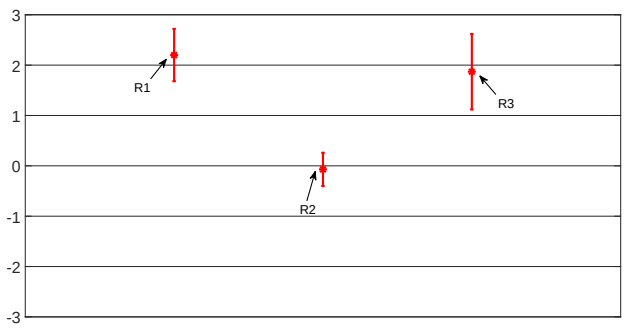


Figure 6.19: CMOS scores for the three desired sources with 95% confidence intervals.

although with slightly larger variations indicated by the angular range of the curve than the pressure matching when the desired sound source is located at position R1. In the case of desired source R2 in Fig. 6.18(b), which is coincident with the location of one of the loudspeakers in the 5 channel system, both the intensity matching and the pressure matching have good localization performance, which is consistent with the simulation results in Section 6.4.2. For desired source R3 in

Fig. 6.18(c), the intensity matching provides more accurate localization performance with smaller variations than the pressure matching. Results in Fig. 6.18 also indicate that the proposed method can reproduce a strong direction perception of the desired sound in the case of there only being a small number of loudspeakers available.

We also calculate the CMOS scores for each desired source according to Table 6.3. The CMOS scores with 95% confidence intervals are given in Fig. 6.19. A 95% confidence interval is a range of values that you can be 95% certain contains the true mean of the population. The average score for desired source R2 is -0.07, which means the direction reproduced by the intensity matching is about the same compared to that reproduced by the pressure matching. The average scores for desired sources R1 and R3 are 2.2 and 1.87, respectively, which means the intensity matching can reproduce a better original sound image compared to the pressure matching when the direction of desired sound is not coincident with the location of one of the loudspeakers for reproduction.

6.6 Summary and Contributions

In this chapter, we have studied the intensity matching technique for spatial soundfield reproduction using an irregular loudspeaker array. The proposed method has an edge over the previous reproduction methods in the following aspects: i) a continuous listening area over space, ii) flexibility of the spatial loudspeaker arrangement, iii) a good human perception of sound localization. We provide an itemized list of our contributions:

- We used the theory of spatial sound intensity vector and reproduced the soundfield inside the target region by controlling the sound intensity vectors on the surrounding contour of the region.

- The intensity matching, as well as the pressure matching and the HOA max-r_E decoding, has been implemented in two loudspeaker arrangements, respectively, consisting of 5 and 22 non-uniformly placed loudspeakers in numerical simulations. The simulation results showed that the proposed method can

reproduce all the components of sound intensity and the direction of the desired sound inside the target region with less error than the pressure matching method and the HOA max-r_E decoding method. It was also proved that the proposed method has good robustness to microphone noise.

- We conducted the perceptual localization experiments with the 5 channel system. The results suggested that the proposed method can reproduce a better original sound image compared to the pressure matching with a small number of loudspeakers.

6.7 Related Publications

Much of this chapter's work has been published in the following journal paper.

- **H. Zuo**, P. N. Samarasinghe, and T. D. Abhayapala, "Intensity based spatial soundfield reproduction using an irregular loudspeaker array," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 28, pp. 1356–1369, 2020.

Chapter 7

Intensity Based Sound Field Reproduction over Multiple Spatial Zones

In the previous chapter, we proposed a spatial sound field reproduction technique based on intensity matching. However, this technique only focuses on a single reproduction zone. This chapter addresses this limitation and extends intensity matching to multizone sound field reproduction. We first propose a method to optimally match the desired sound intensity at multiple sweet spots (i.e., multiple points) in space. We then propose to reproduce sound field over multiple spatial zones by matching the spherical harmonic coefficients of sound intensity within the zones. Both the proposed methods are evaluated by multiple simulations/experiments using a non-uniform loudspeaker array.

7.1 Introduction

Spatial multizone soundfield reproduction aims to reproduce desired soundfields over multiple spatial regions, which has various applications such as simultaneous entertainment systems in cars and personal audio systems in shared office spaces. To achieve such personal sound zones, the multizone reproduction was firstly formulated as creating a bright zone and a dark zone by maximizing the ratio of energy

in the two zones, which is known as the acoustic contrast control method [200–202]. Since then, different approaches based on the acoustic contrast control have been proposed, including the pressure matching approach [46, 203–205], and the modal-domain approach [26, 27, 206, 207]. Besides of creating such two kinds of sound zones (i.e., the bright zone and dark zone), the generation of two zones of silence was investigated in [208]. In [25], the authors recreated two active sound zones by translating the local soundfields to an equivalent global soundfield and then reconstructing the equivalent soundfield using mode matching approach. However, these works are based on sound pressure optimization, which is not directly linked with human perception. This may lead to perception degradation when there are only a limited number of non-uniformly distributed loudspeakers [71].

As we reviewed in Section 6.1, sound intensity has been controlled to improve reproduction performance in the perspective of perception, and the intensity based reproduction methods can provide good localization perception in spatially non-uniform loudspeaker arrangements [50,73,74]. However, these intensity based works are all restricted to a single reproduction position, and therefore perception degrades when listeners are moved from this exact reproduction position. In Chapter 6, we have proposed an intensity matching technique to optimally reproduce sound intensity over a continuous spatial region using a non-uniformly placed loudspeaker array, which allows listeners to freely move over the region with good perception.

In this chapter, we extend the intensity matching technique to multiple sweet spots/spatial zones. We first consider a practical reproduction system that is modeled to reproduce the desired soundfield at multiple sweet spots by driving an irregular loudspeaker array (e.g., in cinemas, we only need to create a good sound perception at seat positions rather than the whole region, and it is easier to implement if the loudspeaker array can be irregular). It is important to note that these sweet spots can be arbitrarily located in a predefined spatial region. We optimize the sound intensity at the sweet spots and compare this technique with the pressure and velocity matching through numerical simulations and perceptual experiments. We then consider a spatial multizone reproduction system in a reverberant room. The aim is to psychoacoustically create the impression of the desired sound in each spatial zone by controlling sound intensity within the zone. We derive sound intensity expressions in a reverberant room and formulate the multizone reproduction

as an optimization problem based on sound intensity. We compare the proposed method with the conventional multizone reproduction method of mode matching in a spatially non-uniform loudspeaker arrangement through numerical simulations.

The remainder of this chapter is organised as follows. Section 7.2 introduces the problem and develops an intensity matching technique for sound field reproduction over multiple sweet spots. The problem is redefined for multiple spatial zones in a reverberant room in Section 7.3 and a new intensity matching technique is proposed by optimizing the spherical harmonic coefficients of sound intensity. Both the techniques are scrutinised under different acoustic environments and the evaluation results are presented in Section 7.4.

7.2 Sound Field Reproduction over Multiple Sweet Spots

7.2.1 Problem formulation

Consider a set of distinct sweet spots, located at $\boldsymbol{x}_b = (r_b, \theta_b, \phi_b)$ for the bth sweet spot, within a spherical region χ of radius R as seen in Fig. 7.1. Note that the sweet spots can be arbitrarily chosen within χ. Let there be an irregular array of loudspeakers with the ℓth loudspeaker located at $\boldsymbol{x}_\ell = (r_\ell, \theta_\ell, \phi_\ell)$ outside of χ. The free field assumption is made and scattering is assumed to be negligible.

The complex acoustic intensity at any point $\boldsymbol{x} = (r, \theta, \phi)$ in χ can be represented by (6.1). Given spherical harmonic coefficients of sound pressure for the desired soundfield $\{\alpha_{nm}(k)\}$ in the spherical region χ, the number of arbitrary sweet spots N_P from χ, and an appropriate irregular loudspeaker array geometry, our objective is to calculate the spatial sound intensity (6.1) in χ and find the loudspeaker driving signals that reproduce the desired sound intensity at the sweet spots in the region.

7.2.2 Multiple sweet spots reproduction algorithm

This subsection introduces the algorithm for multiple sweet spots reproduction. We model each sweet spot as a point and optimize sound intensity at the target points.

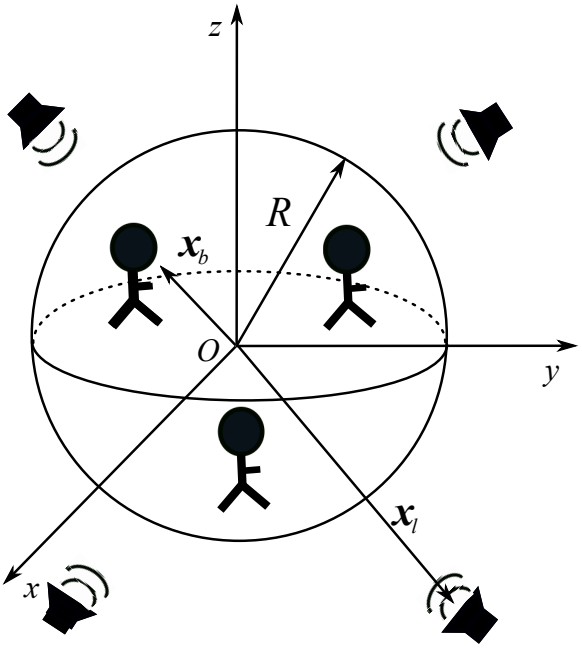


Figure 7.1: Geometry of soundfield reproduction system in this paper. The sweet spots are arbitrarily chosen within the spherical region of radius R. An irregular array of loudspeakers is located outside the spherical region.

Representations of desired and reproduced sound intensity

According to Section 6.3 in Chapter 6, the components of desired sound intensity $\boldsymbol{I}^d(\boldsymbol{x},k) = [I_r^d(\boldsymbol{x},k), I_\theta^d(\boldsymbol{x},k), I_\phi^d(\boldsymbol{x},k)]$ at any arbitrary $\boldsymbol{x}$ within χ can be written as (6.2), (6.3), and (6.4), respectively. Assuming that the loudspeakers in the reproduction system are point sources, sound intensity in r direction due to the ℓth loudspeaker at $\boldsymbol{x}$ can also be expressed by (6.5).

In order to drive loudspeakers, we apply a weight to each loudspeaker. Therefore, the reproduced sound intensity in r direction at $\boldsymbol{x}$ can be written as

$$I_r^a(\boldsymbol{x},k) = \sum_{\ell=0}^{N_L} |w_\ell(k)|^2 I_r^{(\ell)}(\boldsymbol{x},k), \tag{7.1}$$

where $w_\ell(k)$ is the weight applying to the ℓth loudspeaker and N_L is the number of loudspeakers. Here we also assume incoherent superposition of loudspeaker signals, similar to max-r_E decoding method in [73] and [74], which is used to reproduce the

energy and acoustic intensity of the desired soundfield, and psychoacoustically to create the impression of the desired sound, especially for high frequencies (above 500 Hz).

Similar to (7.1), the reproduced sound intensity in θ and ϕ directions at $\boldsymbol{x}$ can be expressed, respectively, as

$$I_\theta^a(\boldsymbol{x}, k) = \sum_{\ell=0}^{N_L} |w_\ell(k)|^2 I_\theta^{(\ell)}(\boldsymbol{x}, k), \tag{7.2}$$

$$I_\phi^a(\boldsymbol{x}, k) = \sum_{\ell=0}^{N_L} |w_\ell(k)|^2 I_\phi^{(\ell)}(\boldsymbol{x}, k). \tag{7.3}$$

Therefore, the reproduced intensity vector is $\boldsymbol{I}^a(\boldsymbol{x}, k) = [I_r^a(\boldsymbol{x}, k), I_\theta^a(\boldsymbol{x}, k), I_\phi^a(\boldsymbol{x}, k)]$.

Loudspeaker weights design

Given the desired soundfield, the reproduction problem is now reduced to calculate the loudspeaker driving signals/weights that can reconstruct the original sound intensity at all sweet spots. Pressure based least squares method [209] is a common approach to calculate loudspeaker weights in a soundfield reproduction system. However, for irregular loudspeaker arrays, this method may lead to errors, which is detrimental to the perception of the location of the source. In order to overcome this limitation, an alternative version of the least squares, the intensity matching technique, is given, which is based on the quantity of sound intensity closely linked with human perception of sound localization [64].

We equate the desired sound intensity to the reproduced sound intensity at the sweet spots to design the loudspeaker weights, i.e.,

$$\boldsymbol{I}^d(\boldsymbol{x}_b, k) = \boldsymbol{I}^a(\boldsymbol{x}_b, k), \qquad b = 1, 2, ..., N_P, \tag{7.4}$$

where $\boldsymbol{I}^d(\boldsymbol{x}_b, k)$=$[I_r^d(\boldsymbol{x}_b, k), I_\theta^d(\boldsymbol{x}_b, k), I_\phi^d(\boldsymbol{x}_b, k)]$ and $\boldsymbol{I}^a(\boldsymbol{x}_b, k) = [I_r^a(\boldsymbol{x}_b, k), I_\theta^a(\boldsymbol{x}_b, k), I_\phi^a(\boldsymbol{x}_b, k)]$. This can be expressed in matrix form as

$$\boldsymbol{I}_D = \boldsymbol{I}_A \boldsymbol{W}, \tag{7.5}$$

where $\boldsymbol{I}_D = [\boldsymbol{I}_r^d(k)^T, \boldsymbol{I}_\theta^d(k)^T, \boldsymbol{I}_\phi^d(k)^T]^T$ is a $6N_P$ long vector with

$$\boldsymbol{I}_\Psi^d(k) = [\mathrm{Re}\{I_\Psi^d(\boldsymbol{x}_1, k)\}, \mathrm{Im}\{I_\Psi^d(\boldsymbol{x}_1, k)\}, ..., \mathrm{Re}\{I_\Psi^d(\boldsymbol{x}_{N_P}, k)\}, \mathrm{Im}\{I_\Psi^d(\boldsymbol{x}_{N_P}, k)\}]^T, \tag{7.6}$$

where $\boldsymbol{W} = [|w_1(k)|^2, |w_2(k)|^2, ..., |w_{N_L}(k)|^2]^T$ is a N_L long vector and $\boldsymbol{I}_A = [\boldsymbol{I}_r^a(k)^T, \boldsymbol{I}_\theta^a(k)^T, \boldsymbol{I}_\phi^a(k)^T]^T$ is a $6N_P$ by N_L matrix with

$$\boldsymbol{I}_\Psi^a(k) = \begin{bmatrix} \mathrm{Re}\{I_\Psi^{(1)}(\boldsymbol{x}_1, k)\} & \cdots & \mathrm{Re}\{I_\Psi^{(N_L)}(\boldsymbol{x}_1, k)\} \\ \mathrm{Im}\{I_\Psi^{(1)}(\boldsymbol{x}_1, k)\} & \cdots & \mathrm{Im}\{I_\Psi^{(N_L)}(\boldsymbol{x}_1, k)\} \\ \vdots & \ddots & \vdots \\ \mathrm{Re}\{I_\Psi^{(1)}(\boldsymbol{x}_{N_P}, k)\} & \cdots & \mathrm{Re}\{I_\Psi^{(N_L)}(\boldsymbol{x}_{N_P}, k)\} \\ \mathrm{Im}\{I_\Psi^{(1)}(\boldsymbol{x}_{N_P}, k)\} & \cdots & \mathrm{Im}\{I_\Psi^{(N_L)}(\boldsymbol{x}_{N_P}, k)\} \end{bmatrix}. \tag{7.7}$$

To minimize the difference between the desired sound intensity and the reproduced sound intensity, the problem is formulated as

$$\min_{\boldsymbol{W}} ||\boldsymbol{I}_A \boldsymbol{W} - \boldsymbol{I}_D||_2^2, \quad \text{s.t.} \boldsymbol{W} \geq 0. \tag{7.8}$$

A technique for solving this problem is known as NNLS [196].

7.3 Spatial Multizone Reproduction in a Reverberant Room

7.3.1 Problem Formulation

We assume that there are a set of distinct 3D spherical zones in a reverberant room and corresponding desired spatial soundfields. As shown in Fig. 7.2, the radius and the origin of the bth spherical zone are denoted as R_b and O_b, respectively, where O_b is located at the spherical coordinate $(r_o^{(b)}, \theta_o^{(b)}, \phi_o^{(b)})$ with respect to the global origin O. Any arbitrary observation point within this b^{th} spherical zone is denoted as $\boldsymbol{x}_b = (r_b, \theta_b, \phi_b)$ with respect to the corresponding local origin O_b. The loudspeakers are placed outside of a sphere with radius of $R_{\max}$ from O in the room, where $R_{\max}$ is the maximum value of $r_o^{(b)} + r_b$ for all the zones (i.e., the loudspeaker

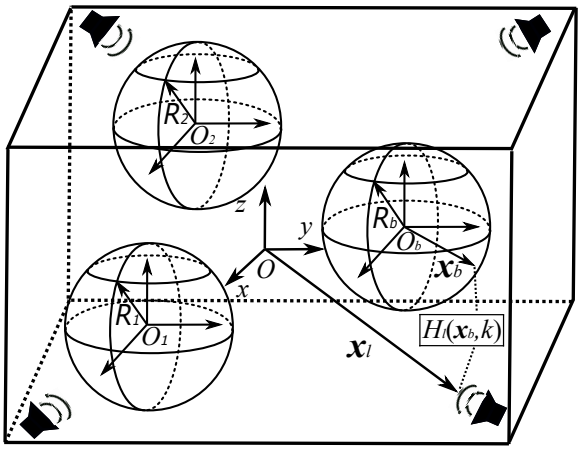


Figure 7.2: Geometry of the multizone reproduction system in this work. The loudspeaker array surrounds the zones.

array encompasses all the spherical zones). The location of the ℓth loudspeaker is denoted as $\boldsymbol{x}_\ell = (r_\ell, \theta_\ell, \phi_\ell)$ with respect to O.

Given the desired spatial soundfield $\{\alpha_{nm}^{(b)}(k)\}$, where $\alpha_{nm}^{(b)}(k)$ are desired pressure coefficients, for the bth spherical zone, the components of desired sound intensity $\boldsymbol{I}^d(\boldsymbol{x}_b, k) = [I_r^d(\boldsymbol{x}_b, k), I_\theta^d(\boldsymbol{x}_b, k), I_\phi^d(\boldsymbol{x}_b, k)]$ at any arbitrary $\boldsymbol{x}_b$ within the zone can be written, according to (3.55), as

$$I_\Psi^d(\boldsymbol{x}_b, k) = \sum_{p=0}^{P} \sum_{q=-p}^{p} S_{pq}^{(\Psi,d)}(k, r_b) Y_{pq}(\theta_b, \phi_b). \tag{7.9}$$

The desired sound intensity distributions of all the spatial zones are available in the form of (7.9). Our objective is to design the driving functions for a loudspeaker array that will simultaneously reproduce all of those intensity distributions in multiple zones.

7.3.2 Multizone reproduction algorithm for a reverberant room

This subsection considers multizone sound field reproduction in a reverberant room. We describe the sound intensity distributions using the theory of spatial sound intensity vectors and optimize the distribution in the target zones.

Representations of reproduced sound intensity

In a reverberant room, reflections should be taken into consideration in addition to the direct path. We first present the point-to-region transfer function in this case, which incorporates both the direct path and reflections.

With respect to O_b, the acoustic transfer function of the ℓth loudspeaker to $\boldsymbol{x}_b$, $H_\ell(\boldsymbol{x}_b, k)$, can be expressed by the spherical harmonic decomposition as

$$H_\ell(\boldsymbol{x}_b, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_{nm}^{(\ell,b)}(k) j_n(kr_b) Y_{nm}(\theta_b, \phi_b), \tag{7.10}$$

where $\beta_{nm}^{(\ell,b)}(k)$ are the sound field coefficients of the acoustic transfer function of the ℓth loudspeaker to the bth zone. Note that the position of the ℓth loudspeaker is $\boldsymbol{x}_\ell - \boldsymbol{O}_b$ with respect to O_b. We also call (7.10) point-to-region transfer function, because the acoustic transfer function of the ℓth loudspeaker to any point in the bth zone is readily available if $\beta_{nm}^{(\ell,b)}(k)$ are known.

We simulate $\beta_{nm}^{(\ell,b)}(k)$ using the spherical harmonics based generalized image source method. We assume all the loudspeakers as omni-directional point sources, and model the reverberant environment as a shoebox room with the size denoted by (L_x, L_y, L_z) for length, width and height. Reverberant characteristics are modeled with the reflection coefficients of the wall surface, denoted as $\boldsymbol{d} = (d_{x1}, d_{x2}, d_{y1}, d_{y2}, d_{z1}, d_{z2})$. The infinite image depth is truncated to $\boldsymbol{R}_{\text{depth}}$ for simplicity. Every single loudspeaker in the loudspeaker array is regarded as a separate source region. According to the results in Section 4.3.1 of Chapter 4, the coefficients of the point-to-region transfer function of the ℓth loudspeaker to the bth zone in the reverberant environment can be written as

$$\beta_{nm}^{(\ell,b)}(k) = \frac{ik}{\sqrt{4\pi}} \alpha_{nm}^{00}(k), \tag{7.11}$$

where $\alpha_{nm}^{00}(k)$ are the coupling coefficients. The location information of the ℓth loudspeaker and the bth zone, as well as the room parameters, is incorporated in $\alpha_{nm}^{00}(k)$. We should note that we assume the point-to-region transfer functions are known at all times in this work, however, the transfer functions may change over time in practice due to the changes in room acoustic conditions, which may degrade

the performance of the proposed method. Thus future advancements in real time learning of room environments will be beneficial in the realization of this work.

The sound intensity for the bth zone due to the ℓth loudspeaker can be written, by replacing $\alpha_{nm}^{(b)}(k)$ with $\beta_{nm}^{(\ell,b)}(k)$ in $S_{pq}^{(\Psi,d)}(k,r_b)$ of (7.9), as

$$I_{\Psi}^{(\ell)}(\boldsymbol{x}_b,k)=\sum_{p=0}^{P}\sum_{q=-p}^{p}S_{pq}^{(\Psi,\ell)}(k,r_b)Y_{pq}(\theta_b,\phi_b), \tag{7.12}$$

where $S_{pq}^{(\Psi,\ell)}(k,r_b)$ are intensity coefficients in Ψ direction.

Applying a frequency-dependent weight $w_\ell(k)$ to the ℓth loudspeaker, the total sound intensity at $\boldsymbol{x}_b$ due to the loudspeaker array is given by

$$I_{\Psi}^{a}(\boldsymbol{x}_b,k)=\sum_{\ell=0}^{N_L}|w_\ell(k)|^2 I_{\Psi}^{(\ell)}(\boldsymbol{x}_b,k), \tag{7.13}$$

where N_L is the number of loudspeakers. Note that here we still assume incoherent superposition of loudspeaker signals.

Loudspeaker weights design

The multizone reproduction is now reduced to calculate the loudspeaker weights that can reconstruct the desired sound intensity for multiple spatial zones. To accurately reconstruct the desired sound intensity within a bounded zone, it requires that the desired intensity coefficients on the surface of the zone (i.e., $S_{pq}^{(\Psi,d)}(k,R_b)$) are accurately reconstructed. Therefore, the problem can be formulated mathematically as

$$\min_{\boldsymbol{W}}\{\sum_{b=1}^{N_P}\tau_b^2||\boldsymbol{S}_A(k,R_b)\boldsymbol{W}(k)-\boldsymbol{S}_D(k,R_b)||_2^2\},\text{s.t.}\boldsymbol{W}\geq 0, \tag{7.14}$$

where N_P is the number of the spatial zones, and τ_b is the weighting coefficient to adjust the relative weight of different zones, $\boldsymbol{S}_D(k,R_b)$=$[\boldsymbol{S}^{(r,d)}(k,R_b)^T,\boldsymbol{S}^{(\theta,d)}(k,R_b)^T,\boldsymbol{S}^{(\phi,d)}(k,R_b)^T]^T$ is a $6(Q+1)^2$ long vector with

$$\boldsymbol{S}^{(\Psi,d)}(k,R_b)=[\text{Re}\{S_{00}^{(\Psi,d)}\},\text{Im}\{S_{00}^{(\Psi,d)}\},...,\text{Re}\{S_{QQ}^{(\Psi,d)}\},\text{Im}\{S_{QQ}^{(\Psi,d)}\}]^T, \tag{7.15}$$

Table 7.1: Loudspeaker locations of the 8-channel array.

Number	r [m]	θ [deg]	ϕ [deg]
1	1	58.3	288
2	1	58.3	216
3	1	58.3	72
4	1	90	18
5	1	90	126
6	1	121.7	324
7	1	121.7	180
8	1	148.3	72

$\boldsymbol{W}(k) = [|w_1(k)|^2, |w_2(k)|^2, ..., |w_{N_L}(k)|^2]^T$ is a N_L long vector, and $\boldsymbol{S}_A(k, R_b)$= $[\boldsymbol{S}^{(r,a)}(k, R_b)^T, \boldsymbol{S}^{(\theta,a)}(k, R_b)^T, \boldsymbol{S}^{(\phi,a)}(k, R_b)^T]^T$ is a $6(Q+1)^2$ by N_L matrix with

$$\boldsymbol{S}^{(\Psi,a)}(k, R_b) = \begin{bmatrix} \mathrm{Re}\{S_{00}^{(\Psi,1)}\} & \cdots & \mathrm{Re}\{S_{00}^{(\Psi,N_L)}\} \\ \mathrm{Im}\{S_{00}^{(\Psi,1)}\} & \cdots & \mathrm{Im}\{S_{00}^{(\Psi,N_L)}\} \\ \vdots & \ddots & \vdots \\ \mathrm{Re}\{S_{QQ}^{(\Psi,1)}\} & \cdots & \mathrm{Re}\{S_{QQ}^{(\Psi,N_L)}\} \\ \mathrm{Im}\{S_{QQ}^{(\Psi,1)}\} & \cdots & \mathrm{Im}\{S_{QQ}^{(\Psi,N_L)}\} \end{bmatrix}. \tag{7.16}$$

The optimization problem (7.14) can be rearranged as

$$\min_{\boldsymbol{W}} ||\widehat{\boldsymbol{S}}_A \boldsymbol{W} - \widehat{\boldsymbol{S}}_D||_2^2, \quad \text{s.t.} \boldsymbol{W} \geq 0, \tag{7.17}$$

where $\widehat{\boldsymbol{S}}_A = [\tau_1 \boldsymbol{S}_A(k, R_1)^T, ..., \tau_{N_P} \boldsymbol{S}_A(k, R_{N_P})^T]^T$, and $\widehat{\boldsymbol{S}}_D = [\tau_1 \boldsymbol{S}_D(k, R_1)^T, ..., \tau_{N_P} \boldsymbol{S}_D(k, R_{N_P})^T]^T$. This problem can be solved using NNLS [196]. Note that (7.17) can provide optimal sound intensity reproduction in each zone given an appropriate loudspeaker array, which allows for a non-uniform array arrangement.

7.4 Evaluations

This section contains the evaluations and analysis of the performance of the two algorithms.

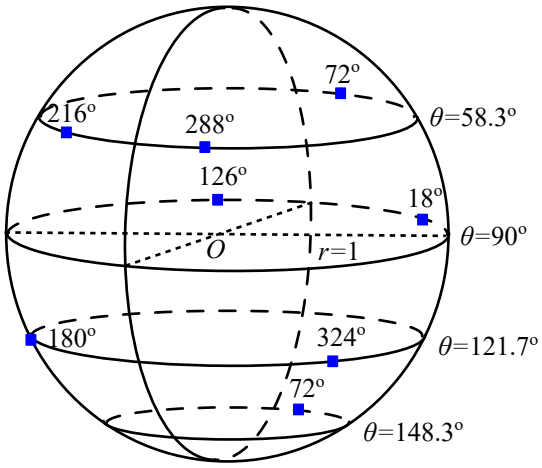


Figure 7.3: 3D sketch of the 8-channel loudspeaker array. The blue squares denote the loudspeakers.

7.4.1 Multiple sweet spots reproduction

Experimental setup and criteria

We consider a spherical region χ with radius $R = 0.4$ m. The sweet spots can be selected at random within the region. In this case, we select sweet spot #1, #2 and #3 located at (0.2 m, $\pi/2$, 0), (0.25 m, $\pi/2$, $\pi/2$) and (0.2 m, $\pi/2$, π), respectively. Eight channels are selected from the spherical array in our lab (shown in Fig. 6.17) to build the irregular loudspeaker array for reproduction in this work. The locations of the loudspeakers are given in Table 7.1, and a 3D sketch of the loudspeaker array is shown in Fig. 7.3. The desired soundfield is produced by a virtual point source which is 2 m away from the origin. Sound speed c is 343 m/s and air density is 1.29 kg/m^3. To reproduce a soundfield with a realistic perception of the sound, it requires to ensure the reproduction of the direction of the sound. Therefore, we define two direction related quantitie as follows

$$\mathrm{DO} = \frac{\mathrm{Re}\{\boldsymbol{I}^a(\boldsymbol{x},k)\}}{\parallel \mathrm{Re}\{\boldsymbol{I}^a(\boldsymbol{x},k)\} \parallel_2} \cdot \frac{\mathrm{Re}\{\boldsymbol{I}^d(\boldsymbol{x},k)\}}{\parallel \mathrm{Re}\{\boldsymbol{I}^d(\boldsymbol{x},k)\} \parallel_2}, \tag{7.18}$$

$$\mathrm{CR} = \frac{\mathrm{Re}\{\boldsymbol{I}^a(\boldsymbol{x},k)\}}{\parallel \mathrm{Re}\{\boldsymbol{I}^a(\boldsymbol{x},k)\} \parallel_2} \times \frac{\mathrm{Re}\{\boldsymbol{I}^d(\boldsymbol{x},k)\}}{\parallel \mathrm{Re}\{\boldsymbol{I}^d(\boldsymbol{x},k)\} \parallel_2}, \tag{7.19}$$

where $\cdot$ is dot product of two vectors and $\times$ is cross product of two vectors. DO and CR describe the difference between the reproduced direction and the desired direction. Ideally, if the direction of the sound at $\boldsymbol{x}$ is reconstructed perfectly, we have DO = 1 and CR = 0. Note that only the real part of complex acoustic intensity, which represents the propagation direction of sound energy, is considered here for evaluation. The intensity direction error η is also used, which can be obtained from either of DO and CR. We here represent it in terms of DO as

$$\eta = \cos^{-1}(\mathrm{DO})/\pi \times 100(\%). \tag{7.20}$$

The pressure and velocity matching is implemented by the least squares method.

Experimental results

We first evaluate the reproduction methods for different positions of the virtual source. The source frequency is 600 Hz. Polar plots showing the results are given in Fig. 7.4. Among them, Fig. 7.4(a), Fig. 7.4(c) and Fig. 7.4(e) are the results for virtual source moving on the horizontal plane (the elevation is $\pi/2$, and the azimuth changes from 0 to 2π). Figure 7.4(b), Fig. 7.4(d) and Fig. 7.4(f) are the results moving on the vertical plane (the azimuth is 0 and the elevation changes from 0 to π for the upper half circle, and the azimuth is π and the elevation changes from π to 0 for the lower half circle). We observe that for all the evaluated positions of the virtual source, the intensity matching performs much better than the pressure and velocity matching at all sweet spots. We note that an irregular loudspeaker array cannot well cover all the incident directions, which leads to relatively poorer performance for the directions at which there are fewer loudspeakers.

The radius of human head is about 0.1 m, therefore, we define a sphere with radius of 0.1 m, centred by the sweet spot, as a human head zone. To evaluate the reproduction performance in human head zones, we calculate the intensity error on plane $z = 0$ for both the intensity matching and the pressure and velocity matching, which is given in Fig. 7.5. The black circles denote the human head zones on plane $z = 0$. The virtual source is located at (2 m, $\pi/2$, $2\pi/3$). It shows that the intensity direction error controlled by the intensity matching is less than that controlled by the pressure and velocity matching in human head zones around

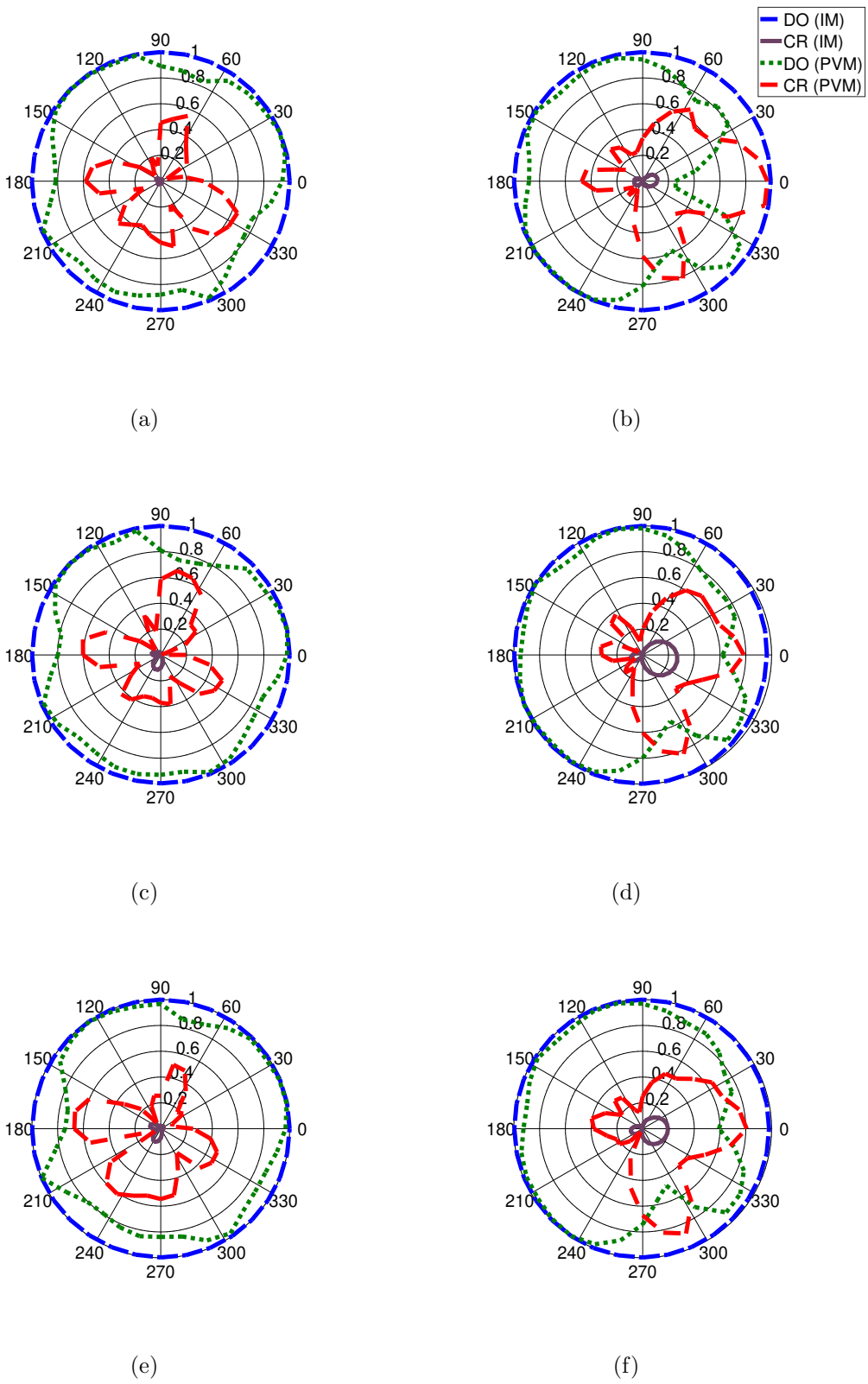
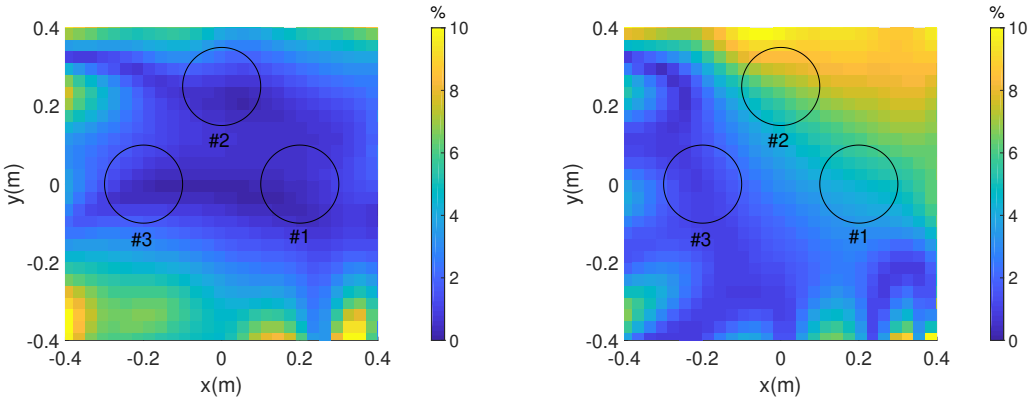


Figure 7.4: Polar plots of DO and CR at sweet spots using the intensity matching (IM) and the pressure and velocity matching (PVM) for the virtual source moving on the horizontal plane ((a) sweet spot #1, (c) #2 and (e) #3) and the vertical plane ((b) sweet spot #1, (d) #2 and (f) #3). The source frequency is 600 Hz.



(a) (b)

Figure 7.5: The intensity direction error on plane $z = 0$ controlled by (a) the intensity matching and (b) the pressure and velocity matching. Black circles denote human head zones around the sweet spots. The source frequency is 600 Hz.

Table 7.2: Average absolute perceived direction error.

Method	Sweet spot #1	Sweet spot #2	Sweet spot #3
IM	41.3°	31.7°	39.0°
PVM	68.0°	69.7°	82.7°

the sweet spots. To examine the reproduction performance with respect to the change of frequencies, we also calculate the intensity direction error at the sweet spots for the frequency range from 50 to 3000 Hz, which mostly covers the frequency range of human voice. The comparison between the intensity matching and the pressure and velocity matching is shown in Fig. 7.6. The intensity direction error of intensity matching is no more than 3% within the frequency range at all the sweet spots. Compared with the intensity matching, the pressure and velocity matching has relatively larger error and the error noticeably fluctuates with frequency.

A perceptual listening test is carried out to validate the theory. The loudspeaker layout in Table 7.1 is built. The test samples consist of A, B and a reference, where the reference is the original sound signal produced by a loudspeaker located at (2 m, $\pi/2$, $2\pi/3$), and A and B are reproduced by the eights channels using the intensity matching and the pressure and velocity matching, respectively. The

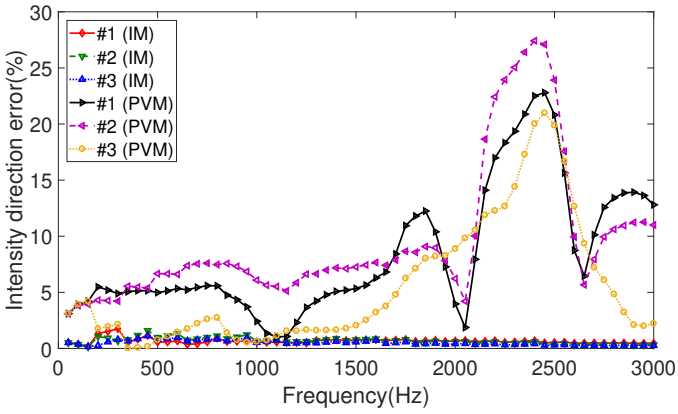


Figure 7.6: The intensity direction error with respect to frequency.

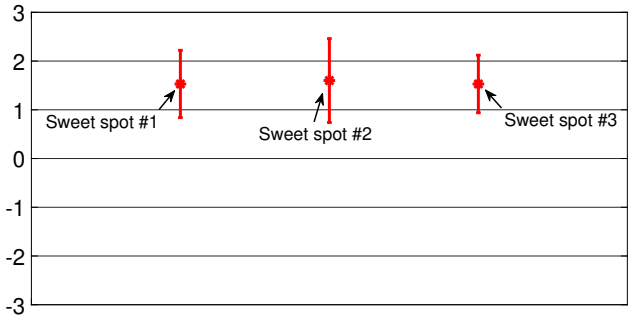


Figure 7.7: CMOS scores for the three sweet spots with 95% confidence intervals.

hearing subjects do not have the knowledge of which method produces A, B or the reference. The hearing subjects are required to listen to A, B and the reference at each sweet spot and record their perceived directions from 72 angles ((θ,ϕ)=(90°, $5° \times (n-1)$), where $n = 1, ..., 72$) on a provided answer sheet. During the test, the hearing subjects face positive y direction and keep their heads at the center of each sweet spot. A clean speech of 5 s duration is used as a source signal and the sample rate adopted is 48 kHz for signal inputs to all of the loudspeakers. The reference is used not only to compare with A and B, but also to identify an outlier within a data set (i.e., if the reference direction that a hearing subject perceives greatly deviates from its actual direction, the result from this subject is removed from the data set). A valid data set containing 15 hearing subjects is obtained after two outliers are removed. We calculate the average absolute perceived direction difference in

angles between A/B and the reference, which is given in Table 7.2. It shows that the absolute error of the proposed method is less than that of the pressure and velocity matching for all the sweet spots. We also analyze the relative perceived direction error of A and B using Comparison Mean Opinion Score (CMOS), which has 7 levels shown in Table 6.3. The higher the CMOS score is, the better the intensity matching is versus the pressure and velocity matching. The test results with 95% confidence intervals are given in Fig. 7.7. The average scores at the three sweet spots are 1.53, 1.60 and 1.53, respectively, which means for this experimental setup the intensity matching can reproduce a better original sound image at all the sweet spots compared to the pressure and velocity matching.

7.4.2 Multizone reproduction

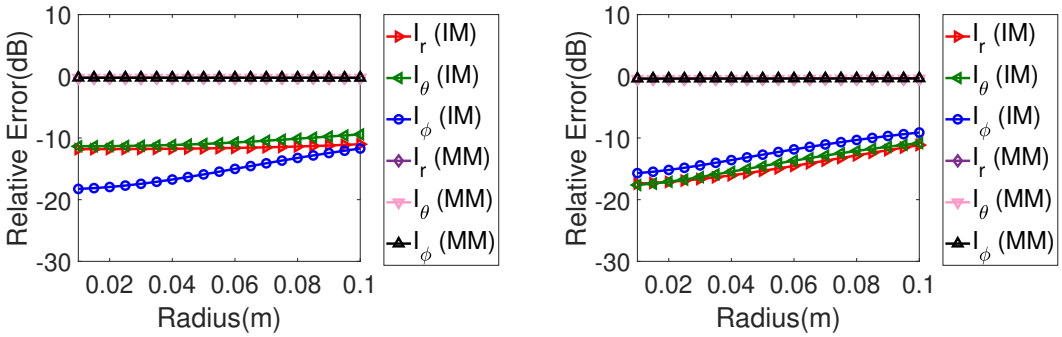
Simulation setup and criteria

In this simulation example, we simulate a rectangular room of size $(5, 6, 3)$ m, with the reflection coefficients of $\boldsymbol{d} = (0.6, 0.7, 0.6, 0.7, 0.6, 0.7)$. The image depth is $\boldsymbol{R}_{\text{depth}} = (3, 3, 3)$. The loudspeaker array consists of eight loudspeakers non-uniformly distributed on a sphere with radius of 1 m, as shown in Fig. 7.3. There are two 3D spherical zones (zone #1 and zone #2) with radius $R_1 = R_2 = 0.1$ m inside the loudspeaker array, and their origins are located at $(0.3, 0, 0)$ m and $(-0.3, 0, 0)$ m, respectively. The desired soundfield is a plane wave coming from $(\pi/3, 4\pi/3)$ with frequency $f = 900$ Hz for both zones. We treat the two zones equally, i.e., $\tau_1 = \tau_2 = 1$. Sound speed is $c = 343$ m/s and air density is $\rho_0 = 1.29$ kg/m^3. The mode matching method [25] is also implemented for comparison.

We have shown that accurate reconstruction of sound intensity guarantees a good localization perception within the target region by perceptual experiments. Therefore, we here use the intensity relative error of (6.19) and the intensity direction error of (6.20) in Chapter 6 to evaluate this algorithm.

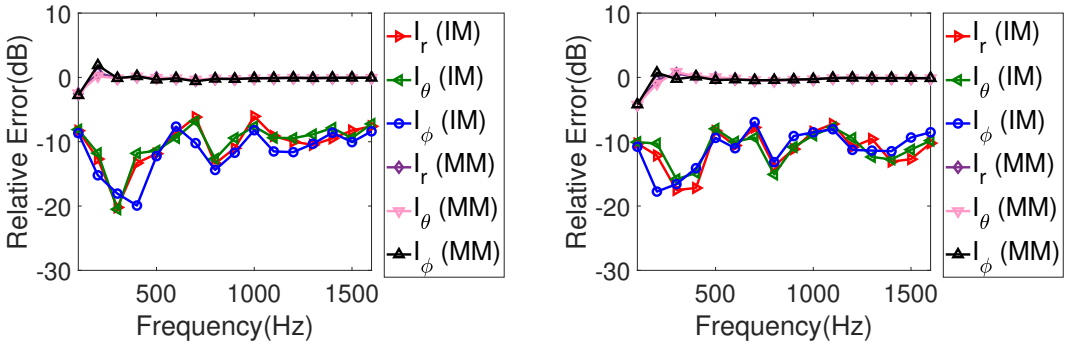
Simulation results

We first evaluate the reproduction of sound intensity within the two target zones for both the intensity matching method and the mode matching method. The results



(a) (b)

Figure 7.8: Relative error with respect to the radius for (a) zone #1 and (b) zone #2 controlled by the intensity matching (IM) and the mode matching (MM). The frequency of the sources is 900 Hz.



(a) (b)

Figure 7.9: Relative error on the surface of the target region with respect to the frequency for (a) zone #1 and (b) zone #2 controlled by IM and MM.

showing the relative error of sound intensity in the r, θ, and ϕ directions are given in Fig. 7.8. The relative error controlled by the mode matching is around 0 dB for all the components of sound intensity vector within the two target zones. For the intensity matching, the relative error lies between -20 dB and -10 dB within the target zones, which is much better than the mode matching. We note that the relative error gradually increases for the intensity matching, whereas it remains for

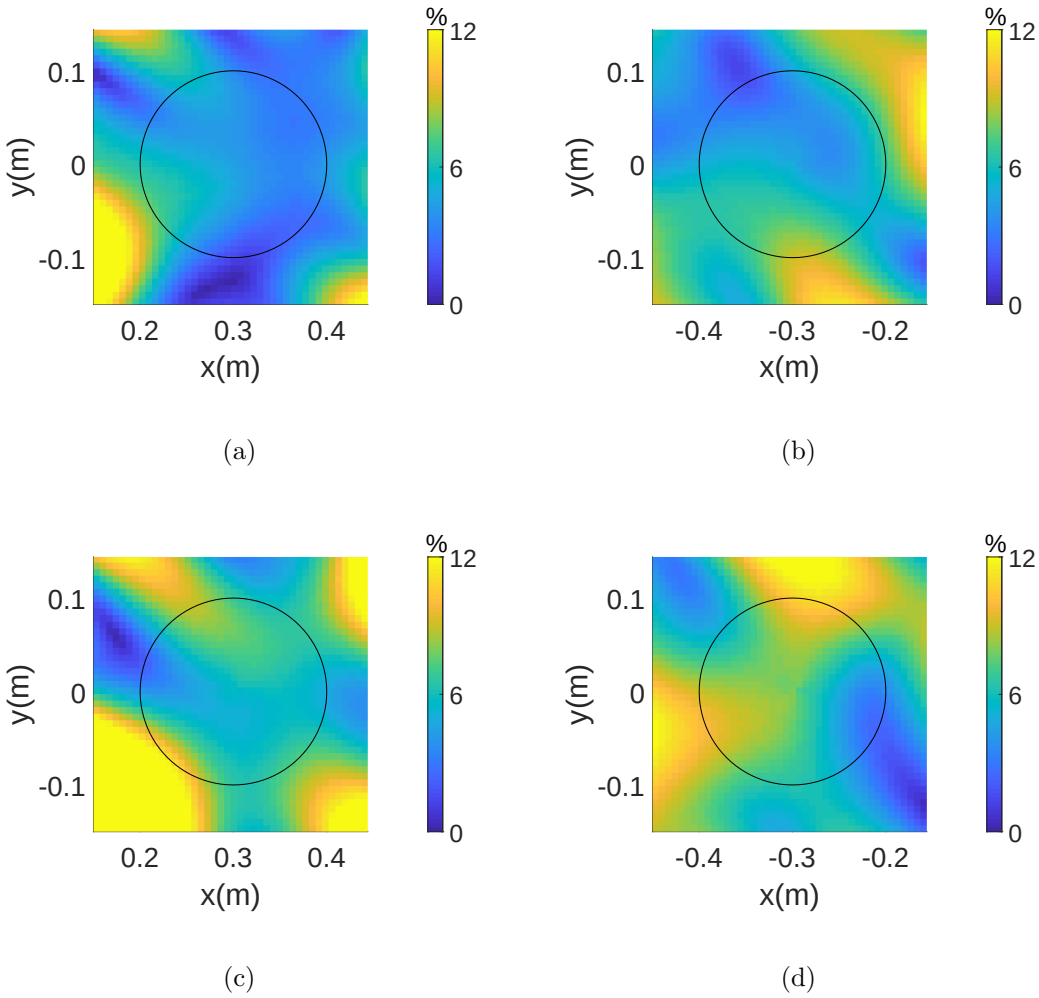


Figure 7.10: The intensity direction error on plane $z = 0$ for (a, c) zone #1 and (b, d) zone #2 controlled by (a-b) IM and (c-d) MM. Black circles denote the target zones. The frequency of the sources is 900 Hz.

the mode matching, as the radius goes up.

To examine the performance of the proposed method with respect to the changes of frequencies, we calculate the relative error by varying the frequency between 100 Hz and 1600 Hz. The observation sphere is the surface of each target zone (i.e., the spherical surface with the maximum error). The results are shown in Fig. 7.9. We observe that the proposed method has around 10 dB improvement compared with the mode matching for the evaluated frequency range in both zones. The different

components of intensity vectors have similar curve fluctuation. This analysis proves that the proposed method can work for broadband multizone soundfield reproduction. Note that the big fluctuations along the frequency for the proposed method are due to the non-linear optimization. Also, similar to the method proposed in Chapter 6, it requires more loudspeakers to guarantee the accurate reproduction in a larger frequency range in the given target region.

The discussion above has shown that all the components of sound intensity can be reproduced with a limited error within the two target zones using the intensity matching method. However, it is clear that if we are concerned to create a realistic perception of the original sound, it requires to ensure the reproduction of the sound direction. Therefore, we calculate the intensity direction error on plane $z = 0$ using (6.20). The results of the intensity matching are given in Fig. 7.10(a) and Fig. 7.10(b). The black circles denote the target zones in the figures. It shows that the intensity matching has less direction error within both the zones by comparing with the results of the mode matching in Fig. 7.10(c) and Fig. 7.10(d).

7.5 Summary and Contributions

In this chapter, we have developed two sound intensity based approaches to multizone soundfield reproduction, which are the extensions of single zone sound intensity reproduction in Chapter 6. Key to these methods is also the theory of spatial sound intensity vectors, which provides the representations of continuous sound intensity for multiple regions. We provide an itemized list of our contributions:

- We have studied a new intensity matching technique used for sound intensity reproduction at multiple sweet spots for irregular loudspeaker arrays to achieve a distributed listening area.

- We compared the intensity matching technique with the pressure and velocity matching method. We demonstrated in the numerical simulations that the intensity matching is better than the pressure and velocity matching for overall reproduction performance. The perceptual experiment results also showed that intensity matching can reproduce the direction of the sound at all sweet spots with less error than the pressure and velocity matching.

- A new intensity based reproduction method for multiple spatial zones in a reverberant room was proposed. We demonstrated, in the simulation example with a spatially non-uniform loudspeaker arrangement, that the proposed method can reproduce the desired sound intensity (both magnitude and direction) within the target zones with a smaller error than the conventional multizone reproduction method of mode matching, which may lead to better perception for listeners.

7.6 Related Publications

Much of this chapter's work has been published in the following conference proceedings.

- **H. Zuo**, P. N. Samarasinghe, and T. D. Abhayapala, "Intensity based sound-field reproduction over multiple sweet spots using an irregular loudspeaker array," in *Proc. 28th European Signal Processing Conference (EUSIPCO)*, pp.486–490, Amsterdam, Netherlands, January, 2021.

- **H. Zuo**, T. D. Abhayapala, and P. N. Samarasinghe, "3D multizone sound-field reproduction in a reverberant environment using intensity matching method," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 416–420, Toronto, Canada, June, 2021.

Chapter 8

Conclusions and Future Work

In this chapter, we state the general conclusions drawn from this thesis. We also outline some future research directions arising from this work.

8.1 Conclusions

This thesis focused on spatial sound field reproduction based on the acoustic vectors of particle velocity and sound intensity. We first developed the spatial distributions of the acoustic vectors from sound pressure in the spherical harmonic domain. Based on the spatial particle velocity vector, we developed a spatial sound field reproduction algorithm applicable for various environments, and proposed a sound field translation technique over a virtual space for binaural reproduction. Based on the spatial sound intensity vector, we designed a sound field reproduction system over a continuous spatial region with an irregular loudspeaker array, and explored multizone reproduction methods.

In the literature, spatial sound field reproduction has been addressed using different approaches, as reviewed in Chapter 1. However, most of these approaches aim to reproduce the physical approximation of sound pressure, which is not directly linked with human perception. Meanwhile, the previous sound field reproduction works based on the acoustic vectors have the limitation of a single sweet spot. To overcome the above constraints, in this thesis, we addressed the two sub-questions posed at the beginning of the thesis.

We solved the problem of *'(i) How to accurately achieve desired acoustic vector distributions over space from a mixed sound field environment using a practical microphone array?'* by theoretically deriving the spatial distributions of the acoustic vectors from a mixed sound field recording in Chapter 3.

We addressed this problem in two steps. We first proposed to extract the pressure coefficients of the desired sound field from a mixed sound field recording using a planar array of differential microphones. The proposed method can provide accurate sound field separation and is robust to the noise. We then derived the spatial distributions of the acoustic vectors from sound pressure in spherical harmonic domain. The spatial distributions of the acoustic vectors can be obtained from the spherical harmonic coefficients of the acoustic vectors, which are directly determined by the pressure coefficients. We showed the infinite number of orthogonal modes of the derived representations can be truncated to a finite number with a small truncation error in order to enable ease of implementation.

We solved the problem of *'(ii) How to perceptually reproduce the desired sound field over a continuous region by exploiting the spatial acoustic vector distributions?'* using the distributions of the spatial particle velocity in Chapter 4 and Chapter 5.

In Chapter 4, we formulated a weighted cost function to optimize the distribution of particle velocity, as well as sound pressure on the boundary of the target region. We analyzed the point-to-region transfer functions that are capable of describing various environments including free fields and reverberant rooms. The performance of the proposed method was validated in various environments (i.e., a free field and two reverberant rooms with different reverberant times) using both uniform and non-uniform loudspeaker arrays through numerical simulations. We also carried out an objective experiment in a six channel non-uniform loudspeaker arrangement using the impulse response measurements in a real-world room. The proposed velocity assisted method can achieve better performance on sound field and sound direction reproduction comparing to the mode matching method, especially for the non-uniform loudspeaker arrays with a limited number of loudspeakers. In Chapter 5, we incorporated the spatial particle velocity vector into the mixed-source model, resulting in the particle velocity based mixed-source ex-

pansion. We offered two solutions (i.e., the least squares solution and the sparse solution) to calculating the driving signals of the mixed-source distribution, and filtered the driving signals by HRTFs to generate the ear signals for binaural reproduction. The sparse solution can translate the sound field over a large region with a small error in the simulation examples, and provide great perception accuracy at the translated positions on both source localization and audio quality in the MUSHRA experiment.

We solved the problem of *'(ii) How to perceptually reproduce the desired sound field over a continuous region by exploiting the spatial acoustic vector distributions?'* by exploiting the distributions of the spatial sound intensity in Chapter 6 and Chapter 7.

We formulated the sound field reproduction problem using spatial sound intensity in Chapter 6. We extended intensity based sound field reproduction at a single sweet spot to a continuous spatial region. We reconstructed the sound field by controlling the intensity distributions on the spherical surface of the target region. The proposed method was implemented using two non-uniform loudspeaker geometries (i.e., a 5 channel array and a 22 channel array), and assessed by comparing to the mode matching and HOA max-r_E decoding through numerical simulations and perceptual localization experiments. Using the proposed method, we can achieve a good perceptual localization in a large and continuous area, even for a loudspeaker array with a limited number of loudspeakers. The proposed method is also robust to microphone noise. We then proposed two algorithms for multizone sound field reproduction based on sound intensity in Chapter 7. The intensity based multizone sound field reproduction can be implemented by optimizing the intensity distributions either at multiple positions over space or on the surface of the target zones. We evaluated the reproduction performance of the two algorithms using a eight channel non-uniform loudspeaker array. The former algorithm can reproduce the original sound direction at multiple different locations in the numerical simulations and perceptual experiments, and the latter algorithm can reproduce the desired sound intensity within all the target zones in a reverberant room in the simulation examples.

All of the above conclusions demonstrate that the spatial acoustic vectors of particle velocity and sound intensity are connected to spatial perception, and the

proposed algorithms are promising in solving the problem of spatial sound field reproduction from the perspective of perception. It also shows that the proposed algorithms are applicable for various reproduction scenarios in this thesis. Therefore, we believe that the outcomes of this research will underpin the future development of spatial sound field reproduction in industry and can be applied across a number of applications such as home entertainments, modern cinemas and VR/AR.

8.2 Future Work

Based on the foregoing discussion, we list the following future research directions to improve the current state of the solution.

Perceptual validation of the intensity based multizone reproduction in reverberant environments

In Chapter 7, we developed an intensity matching technique for 3D multizone sound field reproduction in a reverberant room, and showed the proposed method can reconstruct the original sound intensity within the target zones through numerical simulations. We also verified that the intensity matching method can provide good localization performance in free fields by perceptual localization experiments. However, we have not carried out perceptual experiments to validate the intensity matching method for multiple zones in a reverberant room. The acoustic environment in a reverberant room is more complicated than that in a free field. In such a environment, the perceptual performance may or may not be as good as that in a free field. Therefore, how the reverberation affects the perceptual performance needs further study. The results of such a study would benefit the deployment of the intensity matching method in real-world applications.

Sound quality improvement

In this thesis, we mainly focused on the direction of the reproduced sound when we built the cost functions and defined the evaluation metrics. However, sound quality which represents the accuracy, fidelity, or intelligibility of an audio output, is also important in a sound reproduction system. Although we have showed that the proposed particle velocity based method can offer better sound quality than the

conventional method in Chapter 5, the reproduced sound is obviously different from the original sound, and therefore there is still a plenty of room for improvement. A possible approach to address this problem is to find sound quality related quantities, and incorporate these quantities into the cost functions. Meanwhile, it is also necessary to design the objective metrics for sound quality evaluation.

Broadband reproduction using both particle velocity and sound intensity

As we discussed throughout this thesis, the velocity theory is based on ITD, which is most appropriate at frequencies below 700 Hz, whereas the intensity theory is based on ILD and most suitable for high frequencies (above 500 Hz). All the proposed methods in this thesis exploit either the former or the latter. Therefore, it may be possible to integrate the two theories into one algorithm where particle velocity is controlled for low frequency and sound intensity is controlled for high frequency, which would lead to improved broadband reproduction. The challenge to this problem is the hybrid design for the transition frequency band.

Bibliography

[1] "5.1 vs. 7.1: Battle of the surround sound systems," https://techspirited.com/5-1vs7-1-battle-of-surround-sound-systems, accessed: 2021-04-12.

[2] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Applied Sciences*, vol. 7, no. 5, pp. 532, 2017.

[3] H. Hacihabiboglu, E. De Sena, Z. Cvetkovic, J. Johnston, and J. O. Smith III, "Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics," *IEEE Signal Processing Magazine*, vol. 34, no. 3, pp. 36–54, 2017.

[4] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, Jan. 1993.

[5] A. J. Berkhout, "A holographic approach to acoustic control," *Journal of the Audio Engineering Society*, vol. 36, no. 12, pp. 977–995, Dec. 1988.

[6] P. A. Gauthier and A. Berry, "Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2721–2737, Apr. 2006.

[7] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *Audio Engineering Society Convention 124*. AES, 2008, pp. 17–20.

[8] R. Rabenstein and S. Spors, "Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis," in *Audio Engineering Society Convention 120*. AES, 2006.

[9] M. M. Boone, E. N. G. Verheijen, and P. F. Van Tol, "Spatial sound-field reproduction by wave-field synthesis," *Journal of the Audio Engineering Society*, vol. 43, no. 12, pp. 1003–1012, Dec. 1995.

[10] D. De Vries and M. M. Boone, "Wave field synthesis and analysis using array technology," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1999, pp. 15–18.

[11] M. M. Boone, "Multi-actuator panels (maps) as loudspeaker arrays for wave field synthesis," *Journal of the Audio Engineering Society*, vol. 52, no. 7/8, pp. 712–723, Jul. 2004.

[12] 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, Feb. 2010.

[13] F. Winter, *Local sound field synthesis*, Ph.D. thesis, University of Rostock, Germany, 2019.

[14] S. Spors and J. Ahrens, "Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis," in *Audio Engineering Society Convention 128*. AES, 2010.

[15] G. Firtha, P. Fiala, F. Schultz, and S. Spors, "Improved referencing schemes for 2.5 d wave field synthesis driving functions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 1117–1127, May 2017.

[16] N. Epain and E. Friot, "Active control of sound inside a sphere via control of the acoustic pressure at the boundary surface," *Journal of Sound and Vibration*, vol. 299, no. 3, pp. 587–604, Jan. 2007.

[17] M. Naoe, T. Kimura, Y. Yamakata, and M. Katsumoto, "Performance evaluation of 3d sound field reproduction system using a few loudspeakers and wave field synthesis," in *Second International Symposium on Universal Communication*, 2008, pp. 36–41.

[18] F. Fazi, P. Nelson, J. E. Christensen, and J. Seo, "Surround system based on three-dimensional sound field reconstruction," in *Audio Engineering Society Convention 125*. AES, 2008.

[19] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, Nov. 1985.

[20] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on speech and audio processing*, vol. 9, no. 6, pp. 697–707, 2001.

[21] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, Jan. 2005.

[22] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.

[23] J. Ahrens and S. Spors, "Analytical driving functions for higher order ambisonics," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 373–376.

[24] Z. Li, R. Duraiswami, and N. A. Gumerov, "Capture and recreation of higher order 3d sound fields via reciprocity," in *International Conference Audio Display*, 2004, pp. 6–9.

[25] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1711–1720, Dec. 2010.

[26] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2343–2355, Dec. 2015.

[27] W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi, "Analysis and control of multi-zone sound field reproduction using modal-domain approach," *The Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. 2134–2144, Sept. 2016.

[28] 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, pp. 107–116, 2009.

[29] M. F. Zha, C. C. Bao, and M. S. Jia, "3d multizone soundfield reproduction in the reverberant room using a spherical loudspeaker array," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2015, pp. 23–26.

[30] M. Poletti, F. M. Fazi, and P. A. Nelson, "Sound-field reproduction systems using fixed-directivity loudspeakers," *The Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3590–3601, Apr. 2010.

[31] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1149–1159, Oct. 2010.

[32] W. Zhang and T. D. Abhayapala, "Three dimensional sound field reproduction using multiple circular loudspeaker arrays: functional analysis guided approach," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 7, pp. 1184–1194, May 2014.

[33] P. Chen, P. N. Samarasinghe, and T. D. Abhayapala, "3d exterior soundfield reproduction using a planar loudspeaker array," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 471–475.

[34] M. Poletti, "Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts," *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, pp. 598–610, Jul. 2007.

[35] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 807–820, 2012.

[36] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field reproduction with exterior cancellation using analytical weighting of harmonic coefficients," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 466–470.

[37] J. Zhang, W. Zhang, T. D. Abhayapala, and L. Zhang, "2.5 d multizone reproduction using weighted mode matching: Performance analysis and experimental validation," *The Journal of the Acoustical Society of America*, vol. 147, no. 3, pp. 1404–1417, Feb. 2020.

[38] N. Ueno, S. Koyama, and H. Saruwatari, "Three-dimensional sound field reproduction based on weighted mode-matching method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 1852–1867, Aug. 2019.

[39] W. Jin, "Adaptive reverberation cancelation for multizone soundfield reproduction using sparse methods," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 509–513.

[40] P. Setiawan and W. Jin, "Compressing higher order ambisonics of a multizone soundfield," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 466–470.

[41] F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson, "Generation of private sound with a circular loudspeaker array and the weighted pressure matching method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1579–1591, May 2017.

[42] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint,"

The Journal of the Acoustical Society of America, vol. 135, no. 2, pp. 734–741, Feb. 2014.

[43] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. 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, Nov. 2012.

[44] Q. Zhu, P. Coleman, X. Qiu, M. Wu, J. Yang, and I. Burnett, "Robust personal audio geometry optimization in the svd-based modal domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 610–620, Dec. 2018.

[45] 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, Apr. 2010.

[46] N. Radmanesh and I. S. Burnett, "Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 378–387, Nov. 2012.

[47] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*, Springer Nature, 2019.

[48] H. Wierstorf, *Perceptual assessment of sound field synthesis*, Ph.D. thesis, Technische Universitaet Berlin, Germany, 2014.

[49] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial sound with loudspeakers and its perception: A review of the current state," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 1920–1938, 2013.

[50] M. Frank, *Phantom sources using multiple loudspeakers in the horizontal plane*, Ph.D. thesis, University of Music and Performing Arts Graz, Austria, 2013.

[51] P. Stitt, *Ambisonics and higher-order ambisonics for off-centre listeners: Evaluation of perceived and predicted image direction*, Ph.D. thesis, Queen's University Belfast, UK, 2015.

[52] H. Møller, "Fundamentals of binaural technology," *Applied acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[53] B. Xie, *Head-related transfer function and virtual auditory display*, J. Ross Publishing, 2013.

[54] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[55] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, 2018, vol. 44, pp. 339–342.

[56] D. Sun, "Generation and perception of three-dimensional sound fields using higher order ambisonics," 2013.

[57] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[58] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources i: stereophonic panning," *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 739–752, 2001.

[59] V. Pulkki, "Localization of amplitude-panned virtual sources ii: Two-and three-dimensional panning," *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 753–767, 2001.

[60] J. Merimaa and V. Pulkki, "Spatial impulse response rendering i: Analysis and synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 12, pp. 1115–1127, 2005.

[61] V. Pulkki and J. Merimaa, "Spatial impulse response rendering ii: Reproduction of diffuse sound and listening tests," *Journal of the Audio Engineering Society*, vol. 54, no. 1/2, pp. 3–20, 2006.

[62] A. Politis, *Microphone array processing for parametric spatial audio techniques*, Ph.D. thesis, Aalto University, Finland, 2016.

[63] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Analysis and design of multichannel systems for perceptual sound field reconstruction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1653–1665, 2013.

[64] M. A. Gerzon, "Optimal reproduction matrices for multispeaker stereo," in *Audio Engineering Society Convention 91*. AES, 1991.

[65] E. Kurz and M. Frank, "Prediction of the listening area based on the energy vector," *Proc. ICSA, Graz*, 2017.

[66] H. Hacihabiboğlu and Z. Cvetković, "Panoramic recording and reproduction of multichannel audio using a circular microphone array," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 117–120.

[67] A. Ando and K. Hamasaki, "Sound intensity-based three-dimensional panning," in *Audio Engineering Society Convention 126*. AES, 2009.

[68] M. Buerger, R. Maas, H. W. Löllmann, and W. Kellermann, "Multizone sound field synthesis based on the joint optimization of the sound pressure and particle velocity vector on closed contours," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015, pp. 1–5.

[69] M. Buerger, C. Hofmann, and W. Kellermann, "Broadband multizone sound rendering by jointly optimizing the sound pressure and particle velocity," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. 1477–1490, Feb. 2018.

[70] M. Shin, F. M. Fazi, P. A. Nelson, and J. Seo, "Control of velocity for sound field reproduction," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control-Engineering and Perception.* AES, 2013.

[71] M. Shin, P. A. Nelson, F. M. Fazi, and J. Seo, "Velocity controlled sound field reproduction by non-uniformly spaced loudspeakers," *Journal of Sound and Vibration*, vol. 370, pp. 444–464, Feb. 2016.

[72] A. Ando, "Conversion of multichannel sound signal maintaining physical properties of sound in reproduced sound field," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1467–1475, 2010.

[73] D. Arteaga, "An ambisonics decoder for irregular 3-d loudspeaker arrays," in *Audio Engineering Society Convention 134*. AES, 2013.

[74] D. Scaini and D. Arteaga, "Decoding of higher order ambisonics to irregular periphonic loudspeaker arrays," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*. AES, 2014.

[75] MH Acoustics, "Em32 eigenmike microphone array release notes (v17. 0)," *25 Summit Ave, Summit, NJ 07901, USA*, 2013.

[76] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, vol. 2, pp. II–1949.

[77] P. N. Samarasinghe, T. D. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 647–658, 2014.

[78] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015.

[79] F. J. Fahy, "Measurement of acoustic intensity using the cross-spectral density of two microphone signals," *The Journal of the Acoustical Society of America*, vol. 62, no. 4, pp. 1057–1059, 1977.

[80] A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," *the Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1362–1370, 1977.

[81] H. Bodén and M. Åbom, "Influence of errors on the two-microphone method for measuring acoustic properties in ducts," *The Journal of the Acoustical Society of America*, vol. 79, no. 2, pp. 541–549, 1986.

[82] J. Y. Chung, "Cross-spectral method of measuring acoustic intensity without error caused by instrument phase mismatch," *The Journal of the Acoustical Society of America*, vol. 64, no. 6, pp. 1613–1616, 1978.

[83] M. P. Waser, "Introduction to the two-microphone cross-spectral method of determining sound intensity," *Noise Control Engineering Journal*, vol. 2, no. 3, pp. 76–85, 1984.

[84] J. Y. Chung and J. Pope, "Practical measurement of acoustic intensity-the two-microphone cross-spectral method," in *Inter-Noise 78: Designing for Noise Control*, 1978, pp. 893–900.

[85] J. C. Pascal and C. Carles, "Systematic measurement errors with two microphone sound intensity meters," *Journal of Sound and Vibration*, vol. 83, no. 1, pp. 53–65, 1982.

[86] F. Jacobsen, "A note on finite difference estimation of acoustic particle velocity," *Journal of Sound and Vibration*, vol. 256, no. 5, pp. 849–859, 2002.

[87] H. Li, Y. Li, Z. Y. Geng, and L. J. Wang, "Identification of rotor system noise source based on sound intensity measurement," in *Advanced Materials Research*. Trans Tech Publication Ltd., 2012, vol. 479, pp. 1169–1173.

[88] J. K. Thompson and D. R. Tree, "Finite difference approximation errors in acoustic intensity measurements," *Journal of Sound and Vibration*, vol. 75, no. 2, pp. 229–238, 1981.

[89] H. E. De Bree, "The microflown: An acoustic particle velocity sensor," *Acoustics Australia*, vol. 31, no. 3, pp. 91–94, 2003.

[90] W. Mickiewicz, M. Jablonski, and M. Pyla, "Automatized system for 3d sound intensity field measurement," in *16th International Conference on Methods and Models in Automation and Robotics (MMAR)*. IEEE, 2011, pp. 200–203.

[91] F. Jacobsen and H. De Bree, "Measurement of sound intensity: pu probes versus pp probes," *Proceedings of NOVEM*, 2005.

[92] R. Raangs, W. F. Druyvesteyn, and H. E. De Bree, "A low-cost intensity probe," *Journal of the Audio Engineering Society*, vol. 51, no. 5, pp. 344–357, 2003.

[93] F. Jacobsen and H. de Bree, "A comparison of two different sound intensity measurement principles," *The Journal of the Acoustical Society of America*, vol. 118, no. 3, pp. 1510–1517, 2005.

[94] F. Ma, W. Zhang, and T. D. Abhayapala, "Active control of outgoing noise fields in rooms," *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1589–1599, 2018.

[95] F. Ma, W. Zhang, and T. D. Abhayapala, "Active control of outgoing broadband noise fields in rooms," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 529–539, 2019.

[96] J. A. Zhang, N. Murata, Y. Maeno, P. N. Samarasinghe, T. D. Abhayapala, and Y. Mitsufuji, "Coherence-based performance analysis on noise reduction in multichannel active noise control systems," *The Journal of the Acoustical Society of America*, vol. 148, no. 3, pp. 1519–1528, 2020.

[97] H. Sun, T. D. Abhayapala, and P. N. Samarasinghe, "A realistic multiple circular array system for active noise control over 3d space," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 3041–3052, 2020.

[98] L. Kumar and R. M. Hegde, "Near-field acoustic source localization and beamforming in spherical harmonics domain," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3351–3361, 2016.

[99] A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 178–192, 2016.

[100] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2009, pp. 221–224.

[101] S. Hafezi, A. H. Moore, and P. A. Naylor, "Augmented intensity vectors for direction of arrival estimation in the spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 1956–1968, 2017.

[102] P. N. Samarasinghe, T. D. Abhayapala, and H. Chen, "Estimating the direct-to-reverberant energy ratio using a spherical harmonics-based spatial correlation model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 310–319, 2016.

[103] P. N. Samarasinghe, T. D. Abhayapala, M. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 12, pp. 2217–2227, 2015.

[104] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 261–270, 2012.

[105] S. Delikaris-Manias, J. Vilkamo, and V. Pulkki, "Signal-dependent spatial filtering based on weighted-orthogonal beamformers in the spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1511–1523, 2016.

[106] S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 361–371, 2010.

[107] A. H. Moore and P. A. Naylor, "Linear prediction based dereverberation for spherical microphone arrays," in *IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016, pp. 1–5.

[108] Y. Yamamoto and Y. Haneda, "Spherical microphone array post-filtering for reverberation suppression using isotropic beamformings," in *IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016, pp. 1–5.

[109] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Elsevier, 1999.

[110] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, vol. 348, Springer, 2007.

[111] P. N. Samarasinghe, *Modal based solutions for the acquisition and rendering of large spatial soundfields*, Ph.D. thesis, The Australian National University, Australia, 2015.

[112] T. D. Abhayapala, *Modal analysis and synthesis of broadband nearfield beamforming arrays*, Ph.D. thesis, The Australian National University, Australia, 1999.

[113] L. I. Birnie, T. D. Abhayapala, V. Tourbabin, and P. N. Samarasinghe, "Mixed source sound field translation for virtual binaural application with perceptual validation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021.

[114] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2542–2556, 2007.

[115] A. R. Edmonds, *Angular momentum in quantum mechanics*, Princeton university press, 2016.

[116] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST handbook of mathematical functions*, Cambridge university press, 2010.

[117] P. A. Martin, *Multiple scattering: interaction of time-harmonic waves with N obstacles*, Number 107. Cambridge University Press, 2006.

[118] N. J. A. Sloane, R. H. Hardin, and W. D. Smith, *Spherical Codes*, [Online]. Available: http://neilsloane.com/packings/.

[119] C. F. Van Loan and G. H. Golub, *Matrix computations*, Johns Hopkins University Press, 1983.

[120] F. Ma, W. Zhang, and T. D. Abhayapala, "Reference signal generation for broadband anc systems in reverberant rooms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 216–220.

[121] M. A. Poletti, T. D. Abhayapala, and P. N. Samarasinghe, "Interior and exterior sound field control using two dimensional higher-order variable-directivity sources," *The Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 3814–3823, 2012.

[122] E. Fernandez-Grande and F. Jacobsen, "Sound field separation with a double layer velocity transducer array (l)," *The Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 5–8, 2011.

[123] E. Fernandez-Grande, F. Jacobsen, and Q. Leclere, "Sound field separation with sound pressure and particle velocity measurements," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3818–3825, 2012.

[124] C. X. Bi, X. Z. Chen, and J. Chen, "Sound field separation technique based on equivalent source method and its application in nearfield acoustic holography," *The Journal of the Acoustical Society of America*, vol. 123, no. 3, pp. 1472–1478, 2008.

[125] G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," *The Journal of the Acoustical Society of America*, vol. 68, no. 2, pp. 404–411, 1980.

[126] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Extraction of exterior field from a mixed sound field for 2d height-invariant sound propagation," in *IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016, pp. 1–5.

[127] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *Hands-free Speech Communications and Microphone Arrays (HSCMA)*. IEEE, 2017, pp. 151–155.

[128] A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1149–1159, 2011.

[129] L. Birnie, P. N. Samarasinghe, and T. D. Abhayapala, "3d exterior soundfield capture using pressure and gradient microphone array on 2d plane," in *25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1225–1229.

[130] T. D. Abhayapala and A. Gupta, "Higher order differential-integral microphone arrays," *The Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. EL227–EL233, 2010.

[131] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[132] M. Hawkes and A. Nehorai, "Acoustic vector-sensor beamforming and capon direction estimation," *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2291–2304, 1998.

[133] K. T. Wong and M. D. Zoltowski, "Root-music-based azimuth-elevation angle-of-arrival estimation with uniformly spaced but arbitrarily oriented velocity hydrophones," *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3250–3260, 1999.

[134] F. Jacobsen and Y. Liu, "Near field acoustic holography with particle velocity transducers," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3139–3144, 2005.

[135] A. Henning, S. Kröber, and L. Koop, "Acoustic particle velocity measurements: a cross comparison between modern sensor technologies," *Measurement Science and Technology*, vol. 24, no. 8, pp. 085303, 2013.

[136] P. N. Samarasinghe, T. D. Abhayapala, and W. Kellermann, "Acoustic reciprocity: An extension to spherical harmonics domain," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. EL337–EL343, 2017.

[137] F. J. Fahy and V. Salmon, "Sound intensity," 1990.

[138] R. Fan, Z. Su, G. Meng, and C. He, "Application of sound intensity and partial coherence to identify interior noise sources on the high speed train," *Mechanical Systems and Signal Processing*, vol. 46, no. 2, pp. 481–493, 2014.

[139] T. P. Higgins and Z. Kopal, "Volume integrals of the products of spherical harmonics and their application to viscous dissipation phenomena in fluids," *Astrophysics and Space Science*, vol. 2, no. 3, pp. 352–369, 1968.

[140] D. Sébilleau, "On the computation of the integrated products of three spherical harmonics," *Journal of Physics A: Mathematical and General*, vol. 31, no. 34, pp. 7157, 1998.

[141] S. J. Elliott and J. Garcia-Bonito, "Active cancellation of pressure and pressure gradient in a diffuse sound field," *Journal of Sound and Vibration*, vol. 186, no. 4, pp. 696–704, 1995.

[142] M. E. Johnson and S. J. Elliott, "Active control of sound radiation using volume velocity cancellation," *The Journal of the Acoustical Society of America*, vol. 98, no. 4, pp. 2174–2186, 1995.

[143] P. N. Samarasinghe, T. D. Abhayapala, Y. Lu, H. Chen, and G. Dickins, "Spherical harmonics based generalized image source method for simulating room acoustics," *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1381–1391, Aug. 2018.

[144] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 multichannel sound system and its application," in *Audio Engineering Society Convention 118*. AES, 2005.

[145] E. A. Lehmann, A. M. Johansson, and S. Nordholm, "Reverberation-time prediction method for room impulse responses simulated with the image-source model," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2007, pp. 159–162.

[146] M. A. Gerzon, "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*. AES, 1992.

[147] L. I. Birnie, T. D. Abhayapala, and P. N. Samarasinghe, "Reflection assisted sound source localization through a harmonic domain music framework," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Nov. 2019.

[148] J. K. Bloomfield, S. H. P. Face, and Z. Moss, "Indefinite integrals of spherical bessel functions," *arXiv preprint arXiv:1703.06428*, 2017.

[149] J. G. Tylka and E. Y. Choueiri, "Models for evaluating navigational techniques for higher-order ambisonics," in *Proceedings of Meetings on Acoustics 173EAA*, 2017, vol. 30, p. 050009.

[150] J. G. Tylka and E. Y. Choueiri, "Fundamentals of a parametric method for virtual navigation within an array of ambisonics microphones," *Journal of the Audio Engineering Society*, vol. 68, no. 3, pp. 120–137, 2020.

[151] J. G. Tylka and E. Y. Choueiri, "Performance of linear extrapolation methods for virtual sound field navigation," *Journal of the Audio Engineering Society*, vol. 68, no. 3, pp. 138–156, 2020.

[152] N. Mariette and B. Katz, "Sounddelta—large scale, multi-user audio augmented reality," in *Proceedings of the EAA Symposium on Auralization*, 2009, pp. 15–17.

[153] A. Southern, J. Wells, and D. Murphy, "Rendering walk-through auralisations using wave-based acoustical models," in *17th European Signal Processing Conference*. IEEE, 2009, pp. 715–719.

[154] N. Mariette, B. F. G. Katz, K. Boussetta, and O. Guillerminet, "Sounddelta: a study of audio augmented reality using wifi-distributed ambisonic cell rendering," in *Audio Engineering Society Convention 128*. AES, 2010.

[155] J. G. Tylka and E. Choueiri, "Soundfield navigation using an array of higher-order ambisonics microphones," in *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*. AES, 2016.

[156] K. Müller and F. Zotter, "Auralization based on multi-perspective ambisonic room impulse responses," *Acta Acustica*, vol. 4, no. 6, pp. 25, 2020.

[157] E. Patricio, A. Ruminski, A. Kuklasinski, L. Januszkiewicz, and T. Zernicki, "Toward six degrees of freedom audio recording and playback using multiple ambisonics sound fields," in *Audio Engineering Society Convention 146*. AES, 2019.

[158] Y. Wang and K. Chen, "Translations of spherical harmonics expansion coefficients for a sound field using plane wave expansions," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3474–3478, 2018.

[159] O. Thiergart, G. Del Galdo, M. Taseska, and E. A. Habets, "Geometry-based spatial sound acquisition using distributed microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2583–2594, 2013.

[160] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3d ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. AES, 2003.

[161] D. Menzies and M. Al-Akaidi, "Ambisonic synthesis of complex sources," *Journal of the Audio Engineering Society*, vol. 55, no. 10, pp. 864–876, 2007.

[162] T. Pihlajamaki and V. Pulkki, "Synthesis of complex sound scenes with transformation of recorded spatial sound in virtual reality," *Journal of the Audio Engineering Society*, vol. 63, no. 7/8, pp. 542–551, 2015.

[163] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2005, pp. 150–153.

[164] D. Menzies and M. Al-Akaidi, "Nearfield binaural synthesis and ambisonics," *The Journal of the Acoustical Society of America*, vol. 121, no. 3, pp. 1559–1563, 2007.

[165] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head-movements," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control-Engineering and Perception*. AES, 2013.

[166] E. Fernandez-Grande, "Sound field reconstruction using a spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 139, no. 3, pp. 1168–1178, 2016.

[167] J. G. Tylka and E. Choueiri, "Comparison of techniques for binaural navigation of higher-order ambisonic soundfields," in *Audio Engineering Society Convention 139*. AES, 2015.

[168] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *Audio*

Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction. AES, 2003.

[169] N. Hahn and S. Spors, "Modal bandwidth reduction in data-based binaural synthesis including translatory head-movements," in *The annual German Conference on Acoustics (DAGA)*, 2015, pp. 1122–1125.

[170] N. Hahn and S. Spors, "Physical properties of modal beamforming in the context of data-based sound reproduction," in *Audio Engineering Society Convention 139.* AES, 2015.

[171] A. Kuntz and R. Rabenstein, "Limitations in the extrapolation of wave fields from circular measurements," in *15th European Signal Processing Conference.* IEEE, 2007, pp. 2331–2335.

[172] F. Winter, F. Schultz, and S. Spors, "Localization properties of data-based binaural synthesis including translatory head-movements," in *Forum Acusticum*, 2014, vol. 31.

[173] K. Kowalczyk, O. Thiergart, M. Taseska, G. Del Galdo, V. Pulkki, and E. A. P. Habets, "Parametric spatial sound processing: A flexible and efficient solution to sound scene acquisition, modification, and reproduction," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 31–42, 2015.

[174] T. Laitinen, M. V.and Pihlajamäki, C. Erkut, and V. Pulkki, "Parametric time-frequency representation of spatial sound in virtual worlds," *ACM Transactions on Applied Perception (TAP)*, vol. 9, no. 2, pp. 1–20, 2012.

[175] A. Plinge, S. J. Schlecht, O. Thiergart, T. Robotham, O. Rummukainen, and E. A. P. Habets, "Six-degrees-of-freedom binaural audio reproduction of first-order ambisonics with distance information," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality.* AES, 2018.

[176] E. Stein and M. M. Goodwin, "Ambisonics depth extensions for six degrees of freedom," in *Audio Engineering Society Conference: 2019 AES International Conference on Headphone Technology.* AES, 2019.

[177] M. Blochberger and F. Zotter, "Particle-filter tracking of sounds for frequency-independent 3d audio rendering from distributed b-format recordings," *Acta Acustica*, vol. 5, pp. 20, 2021.

[178] Andrew A. and Bastiaan K., "Ambisonics soundfield navigation using directional decomposition and path distance estimation," in *International Conference on Spatial Audio*, 2017.

[179] M. Kentgens, A. Behler, and P. Jax, "Translation of a higher order ambisonics sound scene based on parametric decomposition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 151–155.

[180] S. Werner, F. Klein, A. Neidhardt, U. Sloma, C. Schneiderwind, and K. Brandenburg, "Creation of auditory augmented reality using a position-dynamic binaural synthesis system—technical components, psychoacoustic needs, and perceptual evaluation," *Applied Sciences*, vol. 11, no. 3, pp. 1150, 2021.

[181] L. Birnie, T. D. Abhayapala, P. N. Samarasinghe, and V. Tourbabin, "Sound field translation methods for binaural reproduction," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, pp. 140–144.

[182] S. Wang, R. Hu, S. Chen, X. Wang, B. Peng, Y. Yang, and W. Tu, "Sound physical property matching between non central listening point and central listening point for nhk 22.2 system reproduction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 436–440.

[183] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 3869–3872.

[184] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[185] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized hrtf fitting using spherical harmonics," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 257–260.

[186] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347–2357, 2010.

[187] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 972–983, 2014.

[188] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999.

[189] A. Lindau, T. Hohn, and S. Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments," in *Audio Engineering Society Convention 122*. AES, 2007.

[190] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated hrtfs including 3d head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.

[191] B. Fabian, D. Manoj, P. Robert, W. Jan Joschka, S. Fabian, V. Daniel, G. Peter, and W. Stefan, "The hutubs head-related transfer function (hrtf) database," [Online]. Available: http://dx.doi.org/10.14279/depositonce-8487, 2019.

[192] "Itu-r bs. 1534-3: Method for the subjective assessment of intermediate quality level of audio systems," ITU Radiocommunication Assembly, Technical Report, October 2015.

[193] J. A. Mann III, J. Tichy, and A. J. Romano, "Instantaneous and time-averaged energy transfer in acoustic fields," *The Journal of the Acoustical Society of America*, vol. 82, no. 1, pp. 17–30, 1987.

[194] H. Chen, T. D. Abhayapala, P. N. Samarasinghe, and W. Zhang, "Direct-to-reverberant energy ratio estimation using a first-order microphone," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 226–237, 2016.

[195] J. Daniel, J. B. Rault, and J. D. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in *Audio Engineering Society Convention 105*. AES, 1998.

[196] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, SIAM, 1995.

[197] B. Johansson, T. Elfving, V. Kozlov, Y. Censor, P. E. Forssén, and G. Granlund, "The application of an oblique-projected landweber method to a model of supervised learning," *Mathematical and computer modelling*, vol. 43, no. 7-8, pp. 892–909, 2006.

[198] V. Franc, V. Hlaváč, and M. Navara, "Sequential coordinate-wise algorithm for the non-negative least squares problem," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2005, pp. 407–414.

[199] M. Hayes, J. Lim, and A. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 672–680, 1980.

[200] J. W. Choi and Y. H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1695–1700, 2002.

[201] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2123–2133, 2012.

[202] P. Coleman, P. Jackson, M. Olik, M. Møller, M. Olsen, and J. A. Pederson, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1929–1940, 2014.

[203] M. Poletti, "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125*. AES, 2008.

[204] J. H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *The Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4518–4525, 2012.

[205] M. F. S. Gálvez, S. J Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1869–1878, 2015.

[206] J. Donley, C. Ritz, and W. B. Kleijn, "Multizone soundfield reproduction with privacy-and quality-based speech masking filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1041–1055, 2018.

[207] Z. Han, M. Wu, Q. Zhu, and J. Yang, "Two-dimensional multizone sound field reproduction using a wave-domain method," *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. EL185–EL190, 2018.

[208] M. A. Poletti and F. M. Fazi, "An approach to generating two zones of silence with application to personal sound systems," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 598–605, 2015.

[209] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2992–3000, 1993.