Sound Field Analysis and Synthesis: Theoretical Advances and Applications to Spatial Audio Reproduction

小山 翔一 / Shoichi Koyama

東京大学 大学院情報理工学系研究科 講師

Lecturer, Graduate School of Information Science and Technology, The University of Tokyo

About me

➢ Shoichi Koyama, Ph.D.

- 2009: Master of Inf. Sci. Tech., UTokyo
- 2009 – 2014: NTT Media Intelligence Labs
- 2014: Ph.D. (Inf. Sci. Tech.), UTokyo
- 2014 – 2018: Research Associate, UTokyo
- 2016 – 2018: Visiting researcher, Paris Diderot Univ.
- 2018 – : Lecturer, UTokyo
- 2020 – : Visiting Associate Prof., Tohoku Univ.

Sound field analysis/synthesis and its applications

Room acoustic analysis

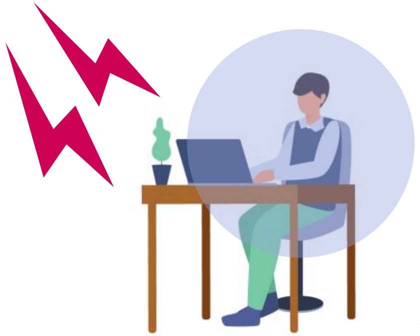
VR/AR audio

Signal enhancement

Basic Technologies of Sound Field Analysis and Synthesis

Active noise control

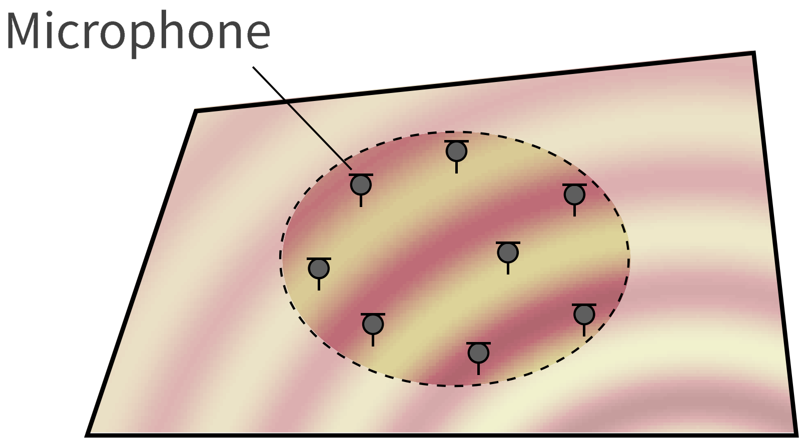
Local-field recording and reproduction



Visualization/auralization

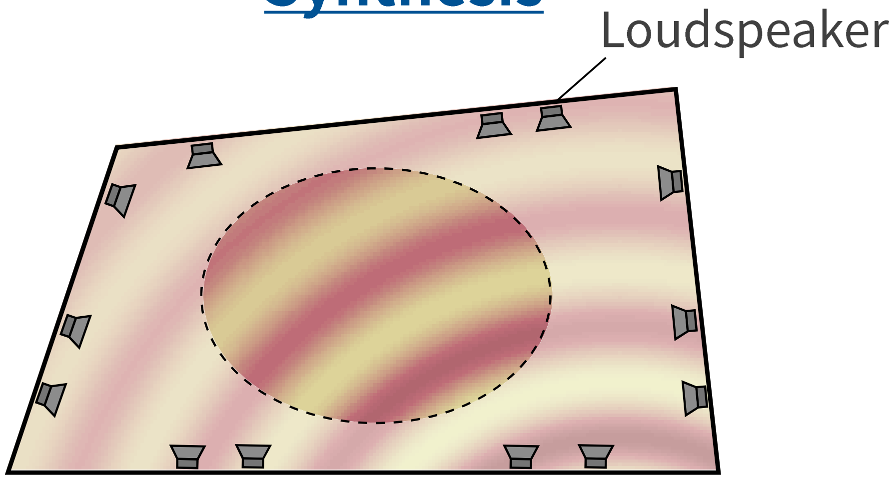
What is sound field analysis/synthesis?

Analysis



Estimating sound field inside target region using multiple mics

Synthesis

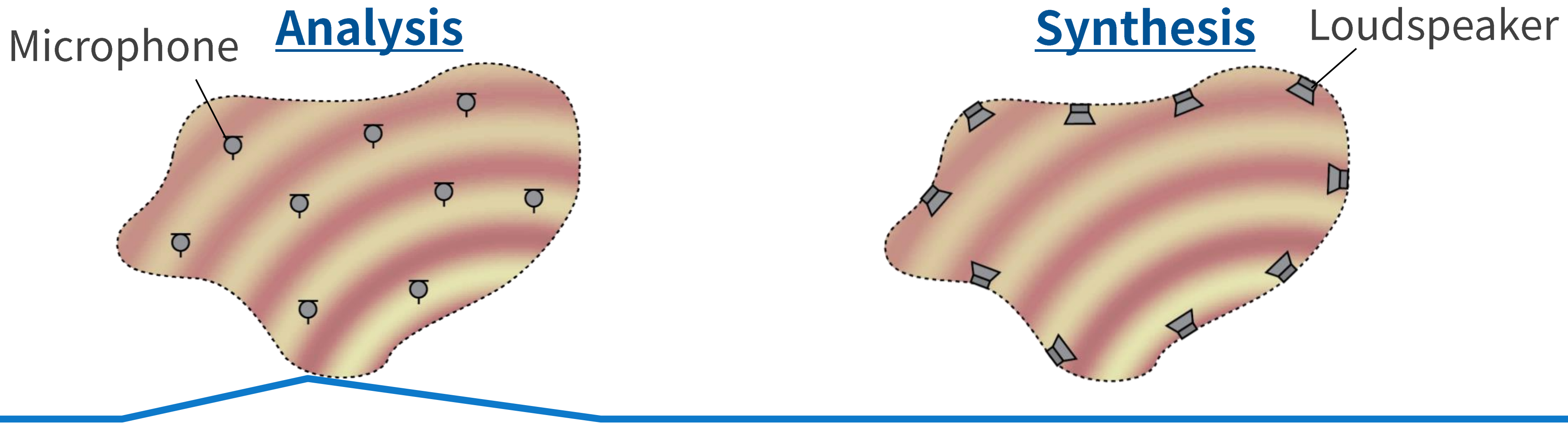


Synthesizing desired sound field inside target region using multiple loudspeakers

Wavefield-informed signal processing and machine learning for sound field analysis and synthesis

Our work on basic technologies

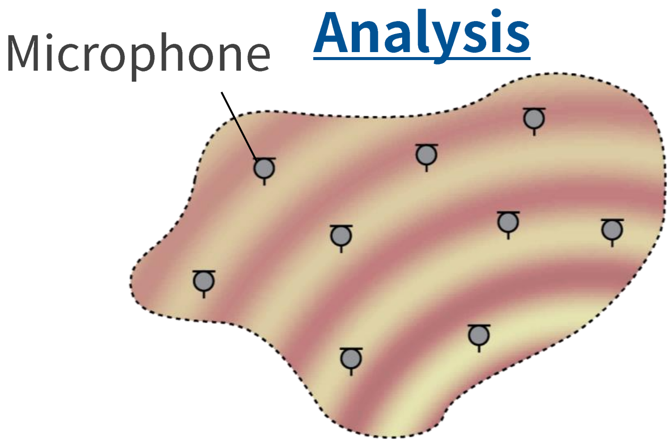
Wavefield-informed signal processing and machine learning

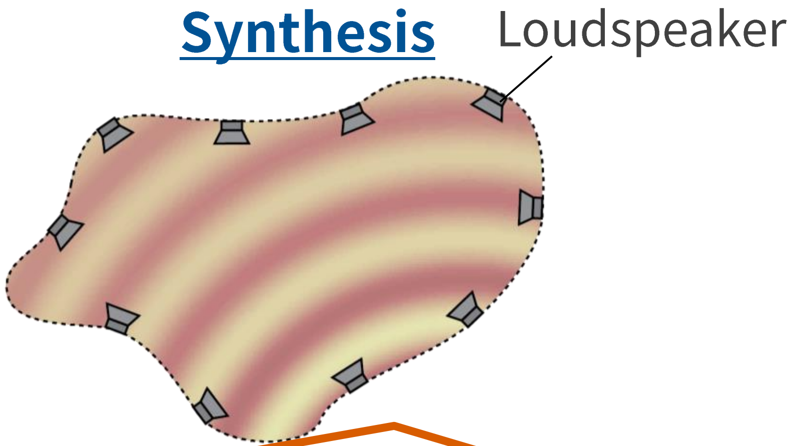


- Kernel interpolation with constraint of Helmholtz eq
  [Ueno+ IEEE SPL 2018, IEEE TSP 2021]
- Sparsity-based super-resolution
  [Murata+ IEEE TSP 2018, Koyama+ JASA 2018, IEEE JSTSP 2019]
- Analysis based on Reciprocity Gap Functional
  [Takida+ Signal Process 2019]

Our work on basic technologies

Wavefield-informed signal processing and machine learning





- Weighted pressure and mode matching for sound field control
  [Ueno+ IEEE/ACM TASLP 2019, Koyama+ JAES 2022]
- Optimization of source and sensor placement
  [Koyama+ IEEE/ACM TASLP 2020, Nishida+ IEEE TSP 2022]
- Amplitude matching for multizone control
  [Koyama+ IEEE ICASSP 2021, Abe+ IEEE/ACM TASLP (under review)]

Enhancing flexibility and scalability to make the range of applications broader

PRELIMINARIES

Governing equations in acoustic field

Sound propagation is governed by wave equation in time domian and Helmholtz equation in frequency domain

- Sound pressure u at position $\boldsymbol{r} \in \mathbb{R}^3$
  - Wave equation for time t

$$\nabla^2 u(\boldsymbol{r}, t) - \frac{1}{c^2} \frac{\partial^2 u(\boldsymbol{r}, t)}{\partial t^2} = 0$$

Fourier transform w.r.t. time

  - Helmholtz equation for wave number $k = \omega / c$

$$(\nabla^2 + k^2) u(\boldsymbol{r}, k) = 0$$

Hereafter, all the formulations are in frequency domain

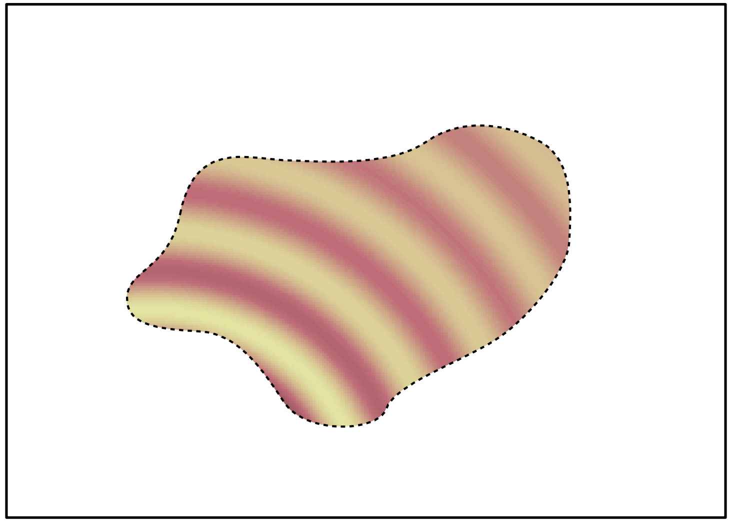
Representations of acoustic field

- Two important acoustic-field representations
  - Boundary-integral representations
    - Describing sound propagation from boundary surface to its interior/exterior region
    - Sound field representation without explicit source parameters
  - Wavefunction expansions
    - Sound field is represented by superposition of wavefunctions, i.e., elementary solutions of Helmholtz equation
    - Complete set of wavefunctions fairly approximates any solutions of homogeneous Helmholtz equation

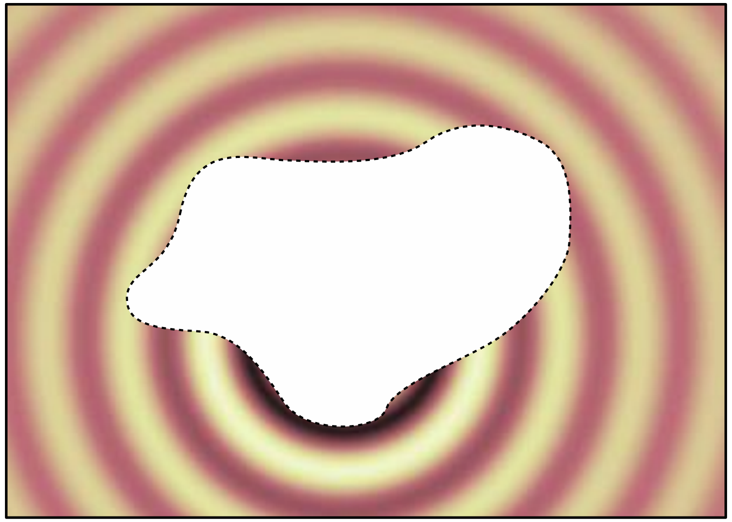
Most of sound field estimation/control methods are based on these two representations

Boundary-integral representation

- **Boundary integral equations** for Helmholtz equation allow predicting interior/exterior sound field from boundary values
  - Kirchhoff–Helmholtz integral
  - Single/double layer potential



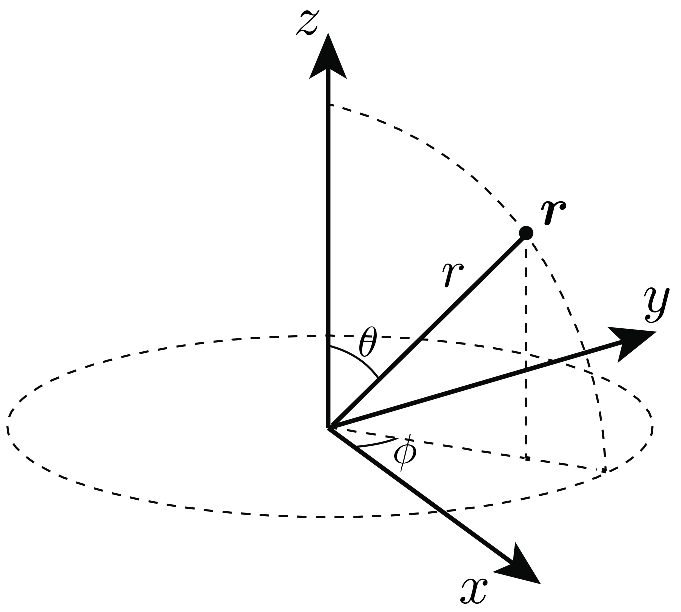
Interior problem



Exterior problem

Wavefunction expansion

- Representing solutions of (homogeneous) Helmholtz equation by complete set of eigenfunctions
- Two representative wavefunction expansions
  - Plane wave expansion
    - Equivalent to general solution in Cartesian coordinate
  - Spherical wavefunction expansion
    - Equivalent to general solution in spherical coordinate



Plane wave expansion

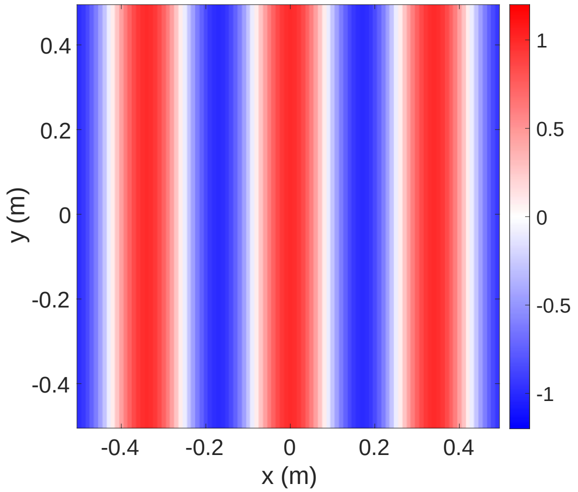
- Plane wave expansion

Expansion coefficient

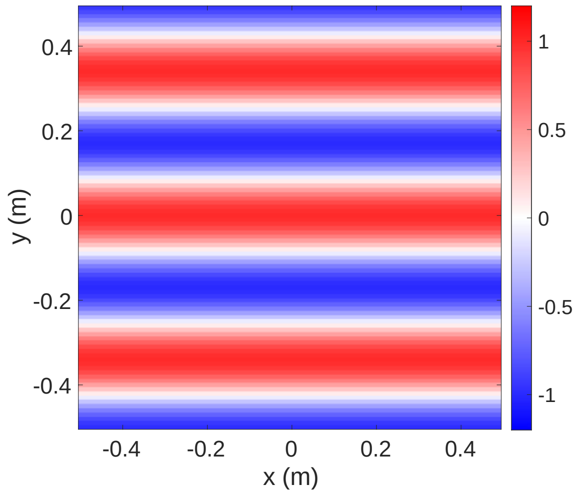
$$u(\boldsymbol{r}) = \int_{\boldsymbol{x}\in\mathbb{S}_2} \tilde{u}(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}k\boldsymbol{x}\cdot\boldsymbol{r}} \mathrm{d}\chi$$

Plane wave function

  - $\boldsymbol{x}$: Unit vector of arrival direction $(\boldsymbol{x} := -\boldsymbol{k}/k)$
  - $\int_{\boldsymbol{x}\in\mathbb{S}_2} \mathrm{d}\chi$: Integral over unit sphere



$\boldsymbol{x} = [1, 0, 0]^\mathsf{T}$



$\boldsymbol{x} = [0, 1, 0]^\mathsf{T}$

Spherical wavefunction expansion

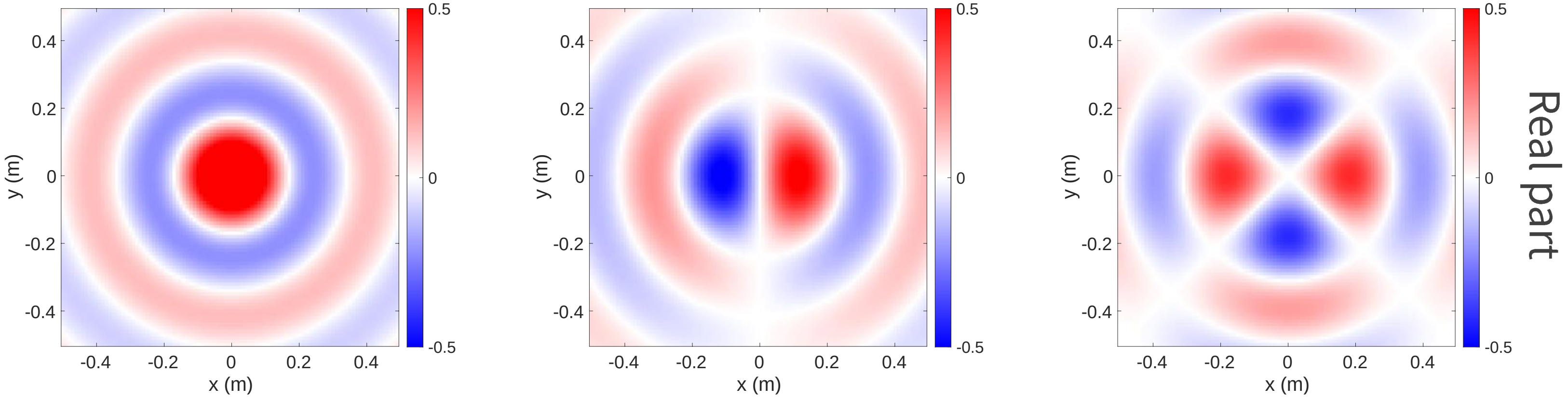
➢ Spherical wavefunction expansion for interior problem

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \mathring{u}_{\nu,\mu} \sqrt{4\pi} j_\nu(k\|\boldsymbol{r}\|) Y_{\nu,\mu}(\boldsymbol{r}/\|\boldsymbol{r}\|)$$

Spherical wavefunction: $\sqrt{4\pi} j_\nu(k\|\boldsymbol{r}\|) Y_{\nu,\mu}(\boldsymbol{r}/\|\boldsymbol{r}\|)$

Expansion coefficient: $\mathring{u}_{\nu,\mu}$

– $j_\nu(\cdot)$: ν th-order spherical Bessel function
– $Y_{\nu,\mu}(\cdot)$: Spherical harmonic function of order ν and degree μ



$\nu = 0,\ \mu = 0$ $\nu = 1,\ \mu = 1$ $\nu = 2,\ \mu = 2$

Spherical wavefunction expansion

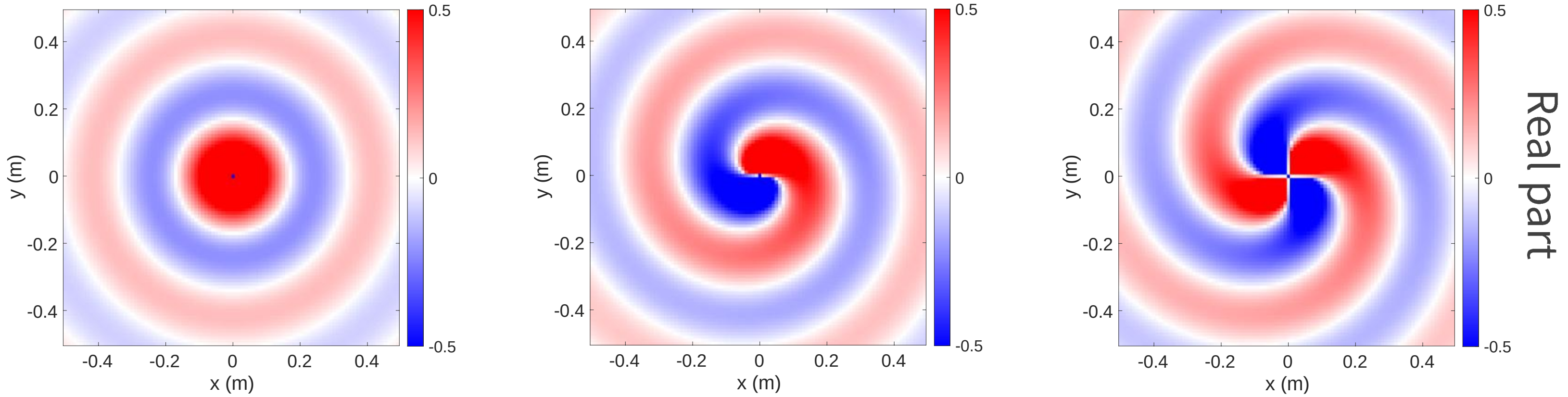
➢ Spherical wavefunction expansion for exterior problem

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \mathring{u}_{\nu,\mu} \sqrt{4\pi} h_\nu(k\|\boldsymbol{r}\|) Y_{\nu,\mu}(\boldsymbol{r}/\|\boldsymbol{r}\|)$$

Spherical wavefunction

Expansion coefficient

- $h_\nu(\cdot)$: ν th-order spherical Hankel function of 1st kind
- $Y_{\nu,\mu}(\cdot)$: Spherical harmonic function of order ν and degree μ



$\nu = 0,\ \mu = 0$ $\nu = 1,\ \mu = 1$ $\nu = 2,\ \mu = 2$

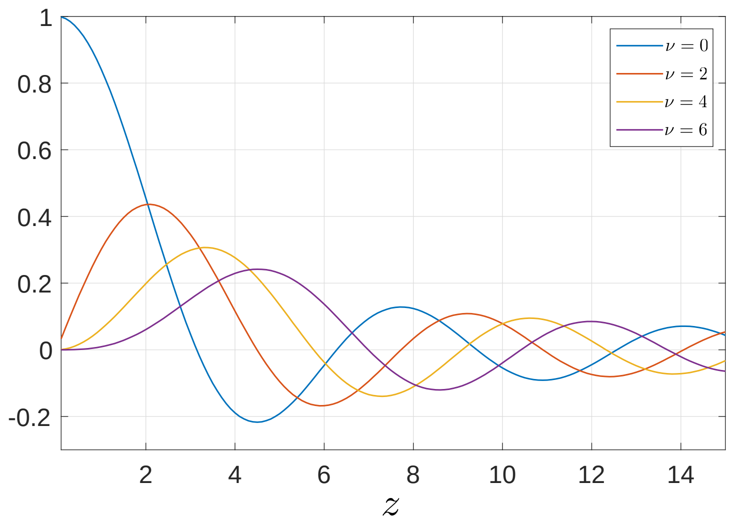
($h_\nu(\cdot)$ has singularity at origin)

Spherical wavefunction expansion

- Spherical Bessel function

$$j_\nu(z) = \sqrt{\frac{\pi}{2z}} J_{\nu+1/2}(z)$$

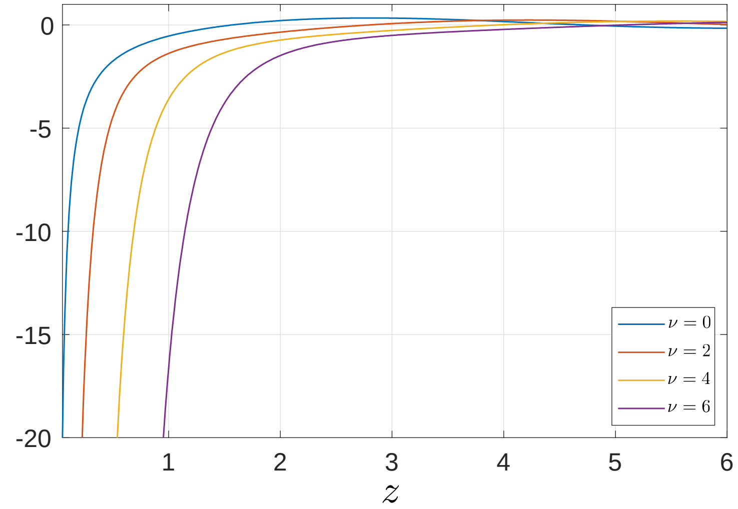
Bessel function



- Spherical Neumann function

$$n_\nu(z) = \sqrt{\frac{\pi}{2z}} N_{\nu+1/2}(z)$$

Neumann function



- Spherical Hankel function of 1st kind

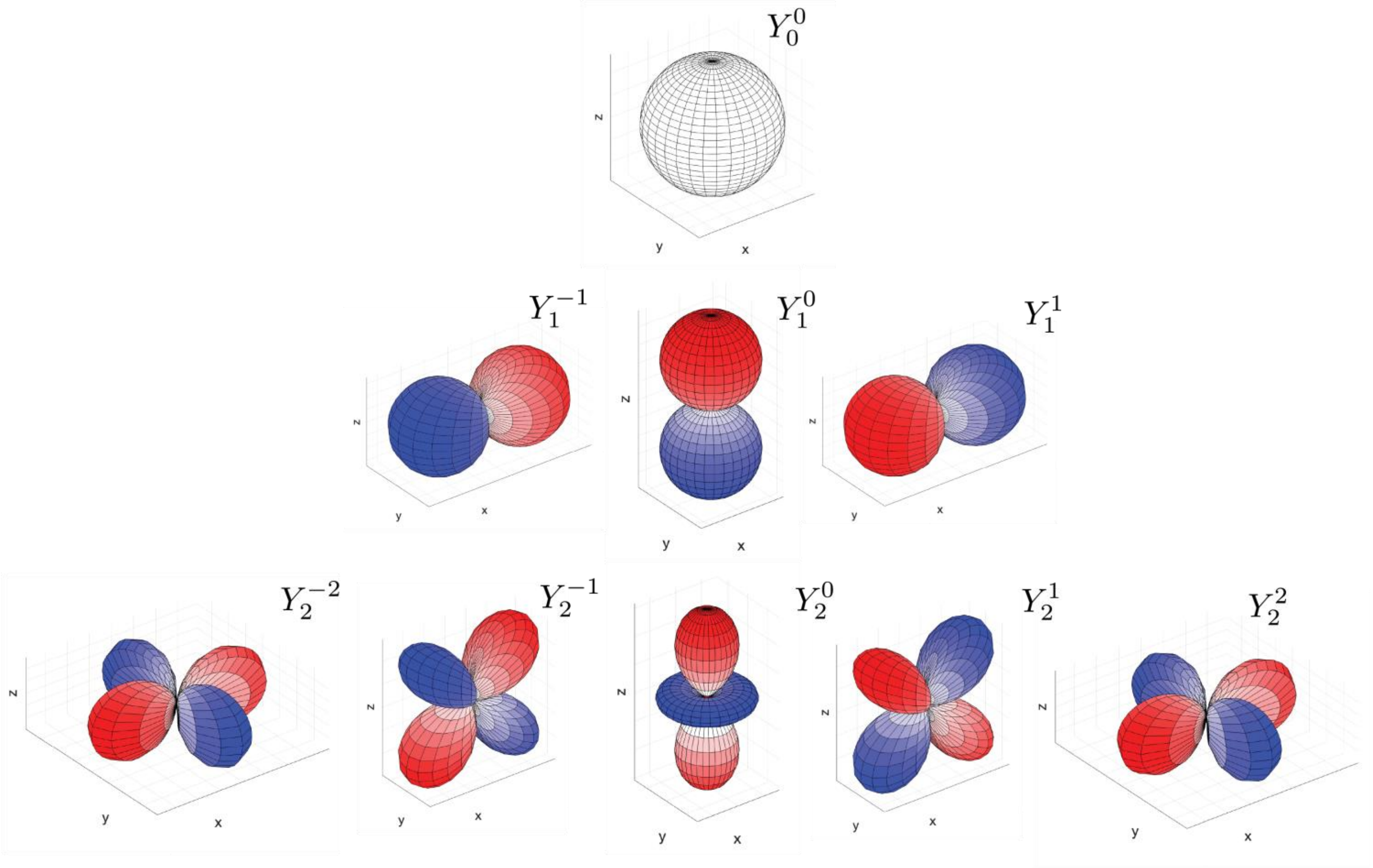
$$h_\nu(z) = j_\nu(z) + \mathrm{j} n_\nu(z)$$

Spherical wavefunction expansion

- Spherical harmonic function

Associated Legendre function

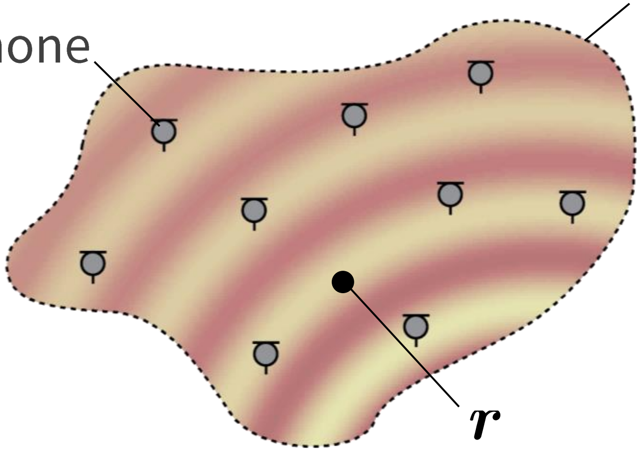
$$Y_{\nu,\mu}(\theta,\phi) = \sqrt{\frac{(2\nu+1)}{4\pi}\frac{(\nu-\mu)!}{(\nu+\mu)!}} P_\nu^\mu(\cos\theta) \mathrm{e}^{\mathrm{j}\mu\phi}$$



SOUND FIELD ANALYSIS

Sound field estimation

Formulation of sound field estimation problem



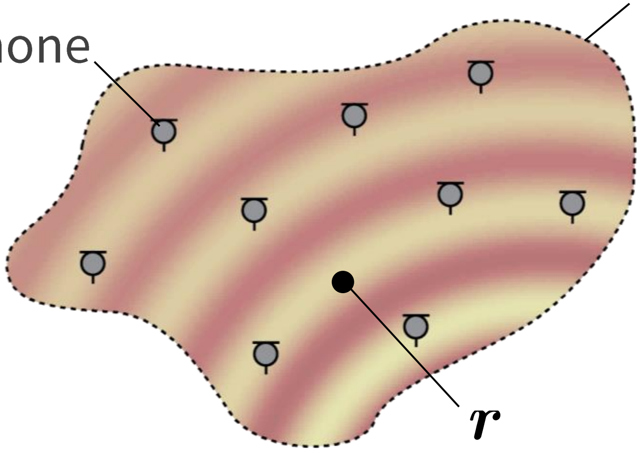
(P1)

Estimate pressure distribution $u(\boldsymbol{r})$ $(\boldsymbol{r} \in \Omega)$ with observations $\{s_m\}_{m=1}^{M}$ at discrete set of M mics $\{\boldsymbol{r}_m\}_{m=1}^{M}$

➡ Ω : Source-free and simply-connected interior region

Sound field estimation

Formulation of sound field estimation problem



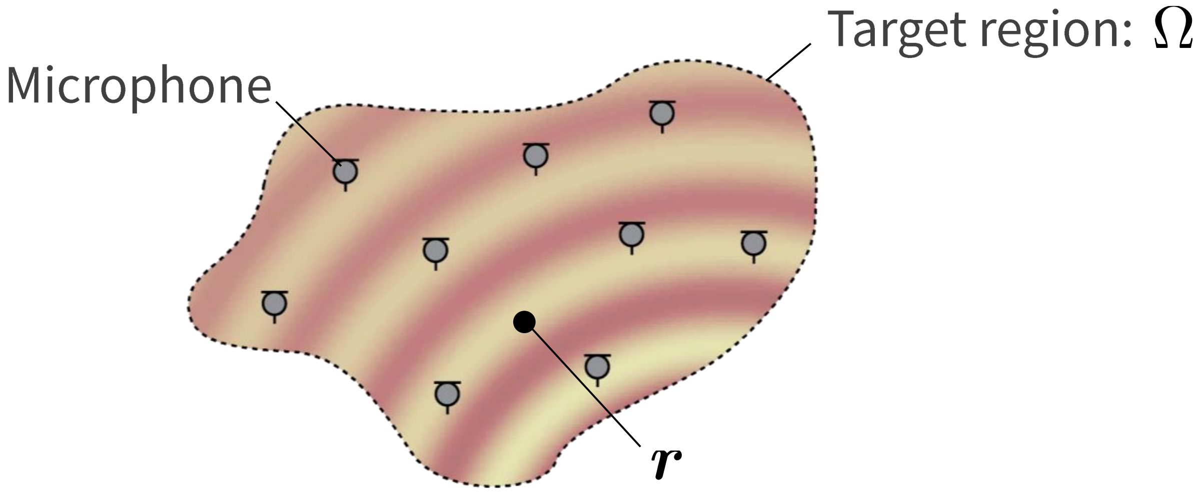
(P2)

Estimate expansion coefficients around $\boldsymbol{r}_0$, i.e., $\mathring{u}_{\nu,\mu}(\boldsymbol{r}_0)$, up to order N with observations $\{s_m\}_{m=1}^{M}$

➡ Ω : Source-free and simply-connected interior region

Sound field estimation

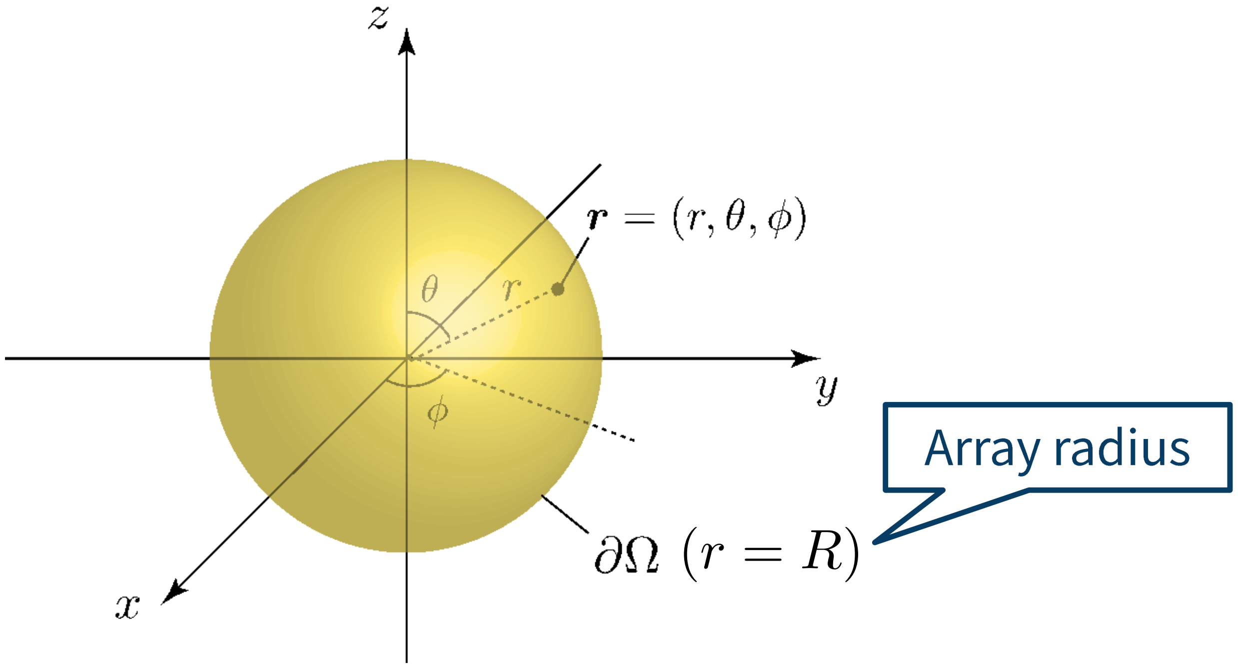
Two major categories of sound field estimation methods



- Integral-equation-based method
  - Based on discretization of boundary integral equation
- Least-squares-based method
  - Based on minimization of square error

Integral-equation-based method for spherical mic array

- Simplify the problem by setting Ω to sphere of radius R
- Spherical array is typically used for spatial audio recording
  - Goal is to estimate expansion coefficients $\mathring{u}_{\nu,\mu}(\boldsymbol{r}_0)$ around array center $\boldsymbol{r}_0$ from observations $\{s_m\}_{m=1}^{M}$ on $\partial\Omega$ (P2)



Integral-equation-based method for spherical mic array

[Poletti 2005]

- Spherical harmonic coefficients on $\partial\Omega$ is obtained by

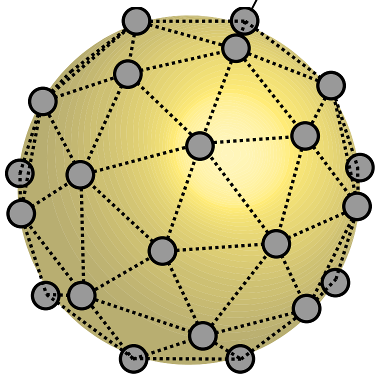
$$U_{\nu,\mu}(R) = \int_0^{2\pi}\int_0^{\pi} u(R,\theta,\phi) Y_\nu^\mu(\theta,\phi)^* \sin\theta \mathrm{d}\theta \mathrm{d}\phi$$

⬇ Discretization by M microphone positions on $\partial\Omega$

Observation s_m

$$U_{\nu,\mu}(R) = \sum_m \gamma_m u(R,\theta_m,\phi_m) Y_\nu^\mu(\theta_m,\phi_m)^*$$

Weight



- Expansion coefficients $\mathring{u}_{\nu,\mu}$ are estimated by

$$\hat{\mathring{u}}_{\nu,\mu} = \frac{1}{\sqrt{4\pi} j_\nu(kR)} U_{\nu,\mu}(R)$$

➡ Incomputable when $j_\nu(kR) = 0$! (forbidden frequency problem)

How to avoid forbidden frequency problem?

- Several established techniques for avoiding forbidden frequency problem
  1. Mics mounted on rigid spherical baffle
  2. Array of directional mics (e.g., unidirectional mics)
  3. Two (or more) layers of spherical mic array

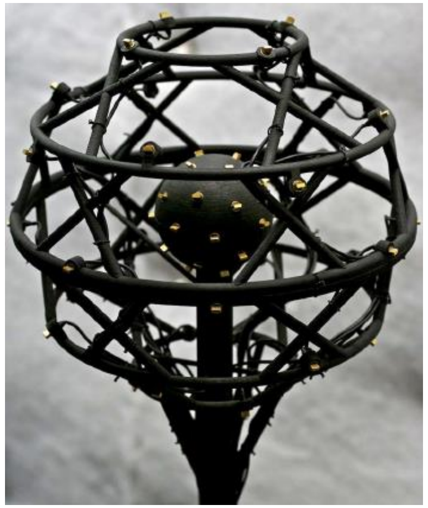
1.

mh acoustics
em32 Eigenmike®

2.

Core Sound
OctoMic™

3.



[Jin+ IEEE/ACM TASLP 2014]

Estimation by rigid spherical mic array

[Poletti 2005]

- Sound field scattered by rigid spherical baffle of radius R

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty}\sum_{\mu=-\nu}^{\nu} \mathring{u}_{\nu,\mu}\sqrt{4\pi}\left[j_\nu(kr) - \frac{j'_\nu(kR)}{h'_\nu(kR)}h_\nu(kr)\right] Y_\nu^\mu(\theta,\phi)$$

- Expansion coefficients are estimated by

$$\hat{\mathring{u}}_{\nu,\mu} = \frac{1}{\sqrt{4\pi}\left[j_\nu(kR) - \frac{j'_\nu(kR)}{h'_\nu(kR)}h_\nu(kR)\right]}U_{\nu,\mu}(R)$$

$$= -\frac{\mathrm{j}k^2R^2}{\sqrt{4\pi}}h'_\nu(kR)U_{\nu,\mu}(R)$$

Does not include spherical Bessel function in denominator

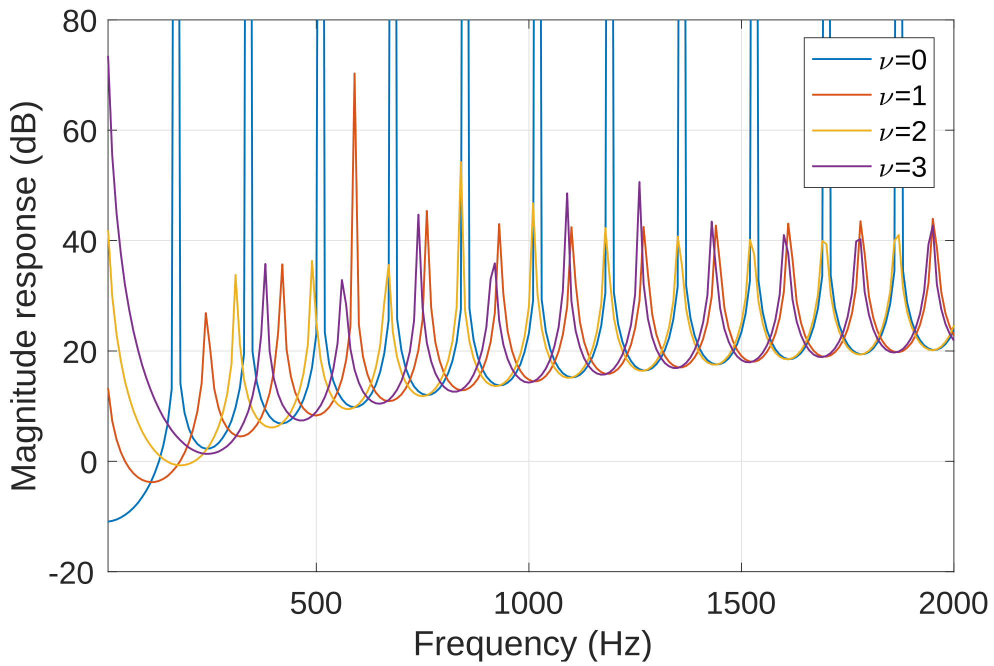
(Wronskian relation: $j_\nu(z)h'_\nu(z) - j'_\nu(z)h_\nu(z) = \frac{\mathrm{j}}{z^2}$)

➡ Much more robust than open spherical mic array

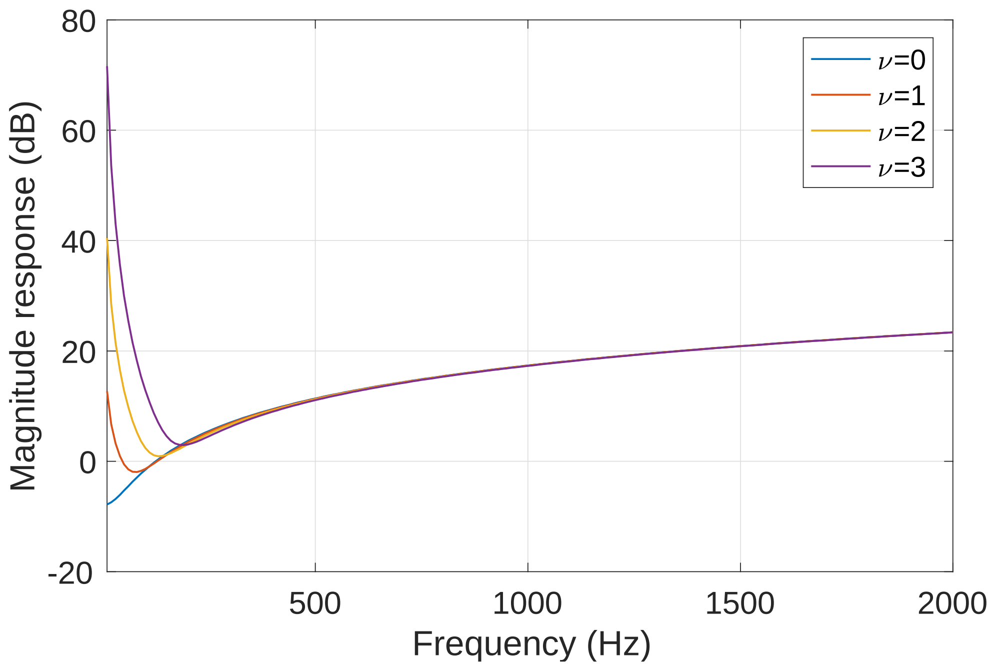
Estimation by rigid spherical mic array

➢ Comparison of array response

Open array

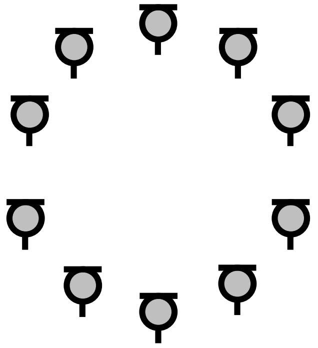


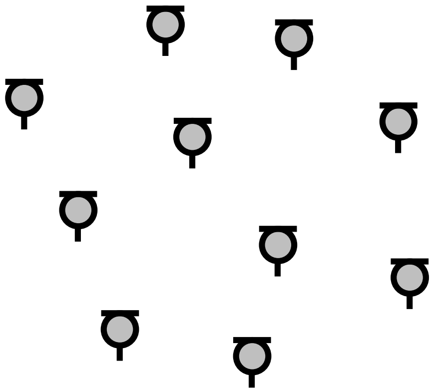
Rigid array



Least-squares-based method

- Limitation of integral-equation-based method
  - Simple array geometry (e.g., sphere, plane)
  - Simple microphone directivity (e.g., omnidirectional, unidirectional)





More flexible method is required

Least-squares-based method

- Least-squares-based sound field estimation
  - Applicable to arbitrary array geometry and microphone directivity
  - Based on decomposition of sound field into basis functions

- 4 steps in least-squares-based method for P1
  1. **Decomposition of sound field**
  2. **Formulation of observation model**
  3. **Formulation of optimization problem**
  4. **Derivation of optimal solution**

Least-squares-based method

- **Decomposition of sound field**

$$u(\boldsymbol{r}) \approx \sum_{n=1}^{N} a_n \psi_n(\boldsymbol{r})$$

Expansion coefficient: a_n

Basis function: $\psi_n(\boldsymbol{r})$

  – Examples of basis function
    - Spherical wavefunction

$$u(\boldsymbol{r}) \approx \sum_{\nu=0}^{N} \sum_{\mu=-\nu}^{\nu} \mathring{u}_{\nu,\mu} \sqrt{4\pi} j_\nu(k\|\boldsymbol{r}\|) Y_{\nu,\mu}(\boldsymbol{r}/\|\boldsymbol{r}\|)$$

    - Plane wave function

$$u(\boldsymbol{r}) \approx \sum_{n=1}^{N} \tilde{u}_n \mathrm{e}^{-\mathrm{j}k\boldsymbol{x}_n \cdot \boldsymbol{r}}$$

Least-squares-based method

- **Formulation of observation model**
  - Decomposition of sound field

$$u(\boldsymbol{r}) \approx \sum_{n=1}^{N} a_n \psi_n(\boldsymbol{r})$$

Expansion coefficient: a_n

Basis function: $\psi_n(\boldsymbol{r})$

  - Observation by m th microphone (superposition principle)

$$s_m = \sum_{n=1}^{N} a_n c_{m,n} + \epsilon_m$$

Sensor noise: ϵ_m

Response to ψ_n: $c_{m,n}$

➡ $c_{m,n}$ is determined by microphone's position and directivity [Laborie+ 2003]

Least-squares-based method

- **Formulation of observation model**
  - Observation by m th microphone

$$s_m = \sum_{n=1}^{N} a_n c_{m,n} + \epsilon_m$$

  - Matrix-vector representation

$$\boldsymbol{s} = \boldsymbol{C}\boldsymbol{a} + \boldsymbol{\epsilon}$$

- **Formulation of optimization problem**

$$\underset{\boldsymbol{a}}{\text{minimize}}\, \mathcal{J}(\boldsymbol{a}) = \underbrace{\|\boldsymbol{C}\boldsymbol{a} - \boldsymbol{s}\|_2^2}_{\text{Loss term}} + \underbrace{\lambda \|\boldsymbol{a}\|_p^p}_{\text{Regularization term}}$$

Least-squares-based method

- **Formulation of optimization problem**

$$\underset{\boldsymbol{a}}{\mathrm{minimize}}\, \mathcal{J}(\boldsymbol{a}) = \|\boldsymbol{C}\boldsymbol{a} - \boldsymbol{s}\|_2^2 + \lambda \|\boldsymbol{a}\|_p^p$$

- **Derivation of optimal solution**
  - For $p = 2$

$$\hat{\boldsymbol{a}} = \boldsymbol{C}^{\mathsf{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathsf{H}} + \lambda \boldsymbol{I})^{-1}\boldsymbol{s}$$

  - Estimated sound field

$$\hat{u}(\boldsymbol{r}) = \sum_{n=1}^{N} \hat{a}_n \psi_n(\boldsymbol{r})$$

Infinite-dimensional extension

- Limitation of finite-dimensional decomposition of sound field
  - Necessity of parameter setting in an empirical manner
    - Number of basis functions
    - Position of expansion center for spherical wavefunction
    - Direction of xyz-axes for plane wave function

➡ **Translation and rotation invariant estimation by infinite-dimensional representation**

- 4 steps in infinite-dimensional extension
  1. **Infinite-dimensional representation of sound field**
  2. **Formulation of observation model**
  3. **Formulation of optimization problem**
  4. **Derivation of optimal solution**

➢ Infinite-dimensional representation of sound field

- Expansion by spherical wavefunction

$$u(\boldsymbol{r}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \mathring{u}_{\nu,\mu} \sqrt{4\pi} j_\nu(k\|\boldsymbol{r}\|) Y_{\nu,\mu}(\boldsymbol{r}/\|\boldsymbol{r}\|)$$

- Expansion by plane wave function

$$u(\boldsymbol{r}) = \int_{\boldsymbol{x}\in\mathbb{S}_2} \tilde{u}(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}k\boldsymbol{x}\cdot\boldsymbol{r}} \mathrm{d}\chi$$

➡ Hilbert space for representing sound fields can be defined as

$$\mathscr{H} = \left\{ u(\boldsymbol{r}) = \sum_{\nu,\mu} \mathring{u}_{\nu,\mu} \sqrt{4\pi} j_\nu(k\|\boldsymbol{r}\|) Y_{\nu,\mu}(\boldsymbol{r}/\|\boldsymbol{r}\|) \;\middle|\; \|u\|_{\mathscr{H}} = \left(\sum_{\nu,\mu} |\mathring{u}_{\nu,\mu}|^2 \right)^{\frac{1}{2}} < \infty \right\}$$

$$= \left\{ u(\boldsymbol{r}) = \int_{\mathbb{S}_2} \tilde{u}(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}k\boldsymbol{x}\cdot\boldsymbol{r}} \mathrm{d}\chi \;\middle|\; \|u\|_{\mathscr{H}} = \left(\int_{\mathbb{S}_2} |\tilde{u}(\boldsymbol{x})|^2 \mathrm{d}\chi \right)^{\frac{1}{2}} < \infty \right\}$$

Infinite-dimensional extension

- **Infinite-dimensional representation of sound field**
  - Representation capability of $\mathscr{H}$
    - Any solution of Helmholtz equation in Ω can be approximated arbitrarily by function in $\mathscr{H}$ in sense of uniform convergence on compact sets

[Ueno+ 2021]

➡ Sufficient representation capability without any parameter

Infinite-dimensional extension

- **Formulation of observation model**
  - Observation by m th microphone

$$s_m = \mathcal{F}_m u + \epsilon_m$$

Linear functional of response

  - $\mathcal{F}_m$: determined by microphone's position and directivity

$$\mathcal{F}_m u = \int_{\boldsymbol{x} \in \mathbb{S}_2} \tilde{u}(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j} k \boldsymbol{x} \cdot \boldsymbol{r}_m} \gamma_m(\boldsymbol{x}) \mathrm{d}\chi$$

Position: $\boldsymbol{r}_m$

Directivity: γ_m

Infinite-dimensional extension

- **Formulation of optimization problem**
  - Regularized least squares in infinite-dimensional Hilbert space $\mathscr{H}$

$$\underset{u \in \mathscr{H}}{\text{minimize}}\, \mathcal{J}(u) = \sum_{m=1}^{M} \frac{1}{\sigma_m^2} |\mathcal{F}_m u - s_m|^2 + \lambda \|u\|_{\mathscr{H}}^2$$

Reformulation using inner product

$$\underset{u \in \mathscr{H}}{\text{minimize}}\, \mathcal{J}(u) = \sum_{m=1}^{M} \frac{1}{\sigma_m^2} |\langle v_m, u \rangle_{\mathscr{H}} - s_m|^2 + \lambda \|u\|_{\mathscr{H}}^2$$

  - v_m : Determined by microphone's position and directivity

$$v_m(\boldsymbol{r}) = \frac{1}{4\pi} \int_{\boldsymbol{x} \in \mathbb{S}_2} \gamma_m(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}k\boldsymbol{x}\cdot(\boldsymbol{r}-\boldsymbol{r}_m)} \mathrm{d}\chi$$

Position: $\boldsymbol{r}_m$

Directivity: $\gamma_m(\boldsymbol{x})$

Infinite-dimensional extension

➢ Derivation of optimal solution

– Optimization problem

$$\underset{u \in \mathscr{H}}{\text{minimize}} \, \mathcal{J}(u) = \sum_{m=1}^{M} \frac{1}{\sigma_m^2} |\langle v_m, u \rangle_{\mathscr{H}} - s_m|^2 + \lambda \|u\|_{\mathscr{H}}^2$$

– Optimal solution

$$\hat{u}_m(\boldsymbol{r}) = \sum_{m=1}^{M} \hat{\alpha}_m v_m(\boldsymbol{r})$$

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{K} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{s}$$

$$\boldsymbol{K} := \begin{bmatrix} \langle v_1, v_1 \rangle_{\mathscr{H}} & \cdots & \langle v_1, v_M \rangle_{\mathscr{H}} \\ \vdots & \ddots & \vdots \\ \langle v_M, v_1 \rangle_{\mathscr{H}} & \cdots & \langle v_M, v_M \rangle_{\mathscr{H}} \end{bmatrix}$$

$$\boldsymbol{\Sigma} := \text{diag}\left(\sigma_1^2, \ldots, \sigma_M^2\right)$$

- No necessity to set parameters for finite-dimensional decomposition
- Optimal solution can be obtained in closed form

Interpretation as kernel ridge regression

- In case of pressure microphones (= interpolation problem of sound field)
  - Estimated sound field

$$\hat{u}_m(\boldsymbol{r}) = \sum_{m=1}^{M} \hat{\alpha}_m \kappa(\boldsymbol{r}, \boldsymbol{r}_m)$$

$$\kappa(\boldsymbol{r}, \boldsymbol{r}_m) = j_0(k\|\boldsymbol{r} - \boldsymbol{r}_m\|)$$: Kernel function

  - Coefficients

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{K} + \lambda \boldsymbol{\Sigma})^{-1} \boldsymbol{s}$$

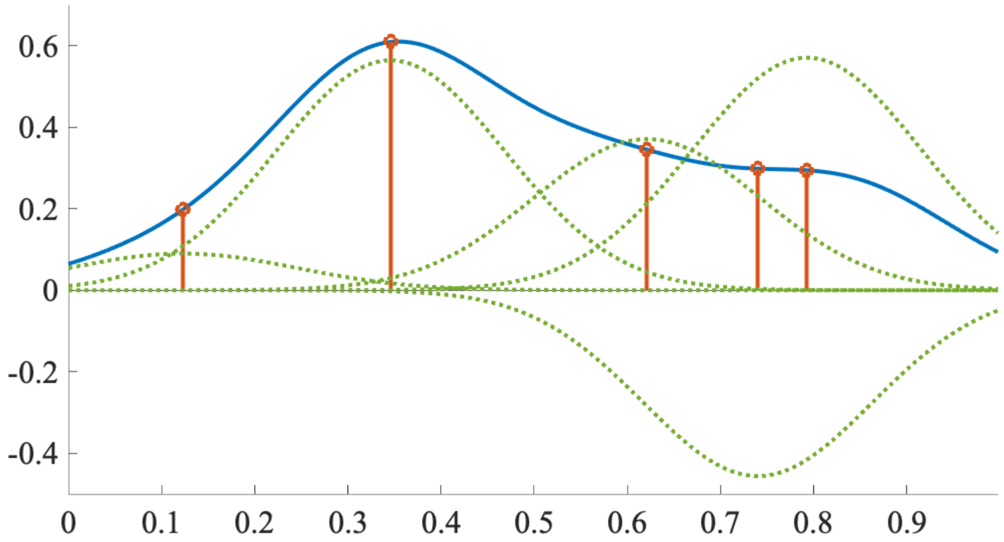
$$\boldsymbol{K} := \begin{bmatrix} \kappa(\boldsymbol{r}_1, \boldsymbol{r}_1) & \cdots & \kappa(\boldsymbol{r}_1, \boldsymbol{r}_M) \\ \vdots & \ddots & \vdots \\ \kappa(\boldsymbol{r}_M, \boldsymbol{r}_1) & \cdots & \kappa(\boldsymbol{r}_M, \boldsymbol{r}_M) \end{bmatrix}$$: Gram matrix

Equivalent to kernel ridge regression with constraint of Helmholtz equation

Interpretation as kernel ridge regression

- Interpolation is achieved by linear combination of kernel functions at data points
  - Typically-used kernel function in machine learning is Gaussian kernel

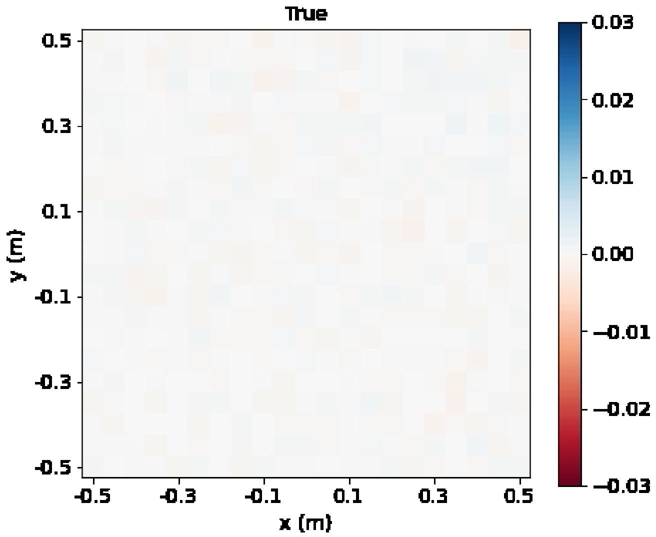
$$\kappa(\boldsymbol{r}_1, \boldsymbol{r}_2) = \exp\left(-\frac{\|\boldsymbol{r}_1 - \boldsymbol{r}_2\|^2}{\sigma^2}\right)$$

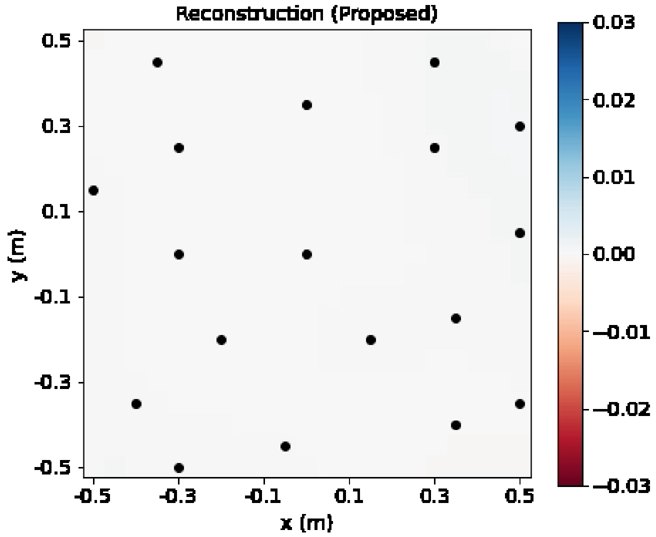


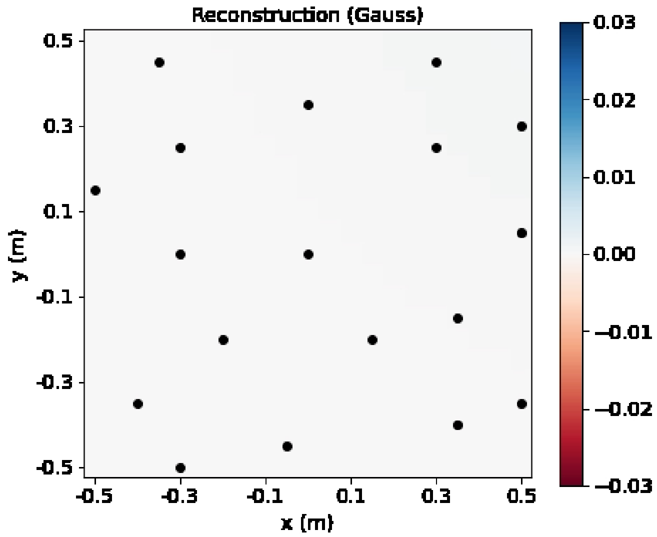
Experimental example

- Experimental results using real data using MeshRIR data set [Koyama+ 2021]
  - Reconstructing pulse signal from single loudspeaker w/ 18 mic

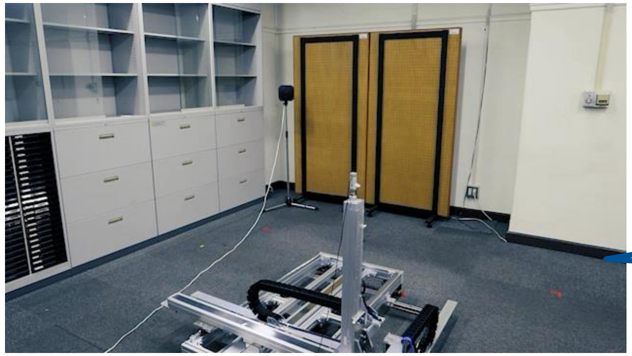
True | **Proposed** | **Gaussian kernel**







(Black dots indicate mic positions)



Impulse response measurement system

Related work

- **Infinite-dimensional harmonic analysis** [Ueno+ 2018]
  - Estimation of expansion coefficient at arbitrary expansion center (P2)
  - No truncation in expansion of sound field or in translation of expansion coefficient

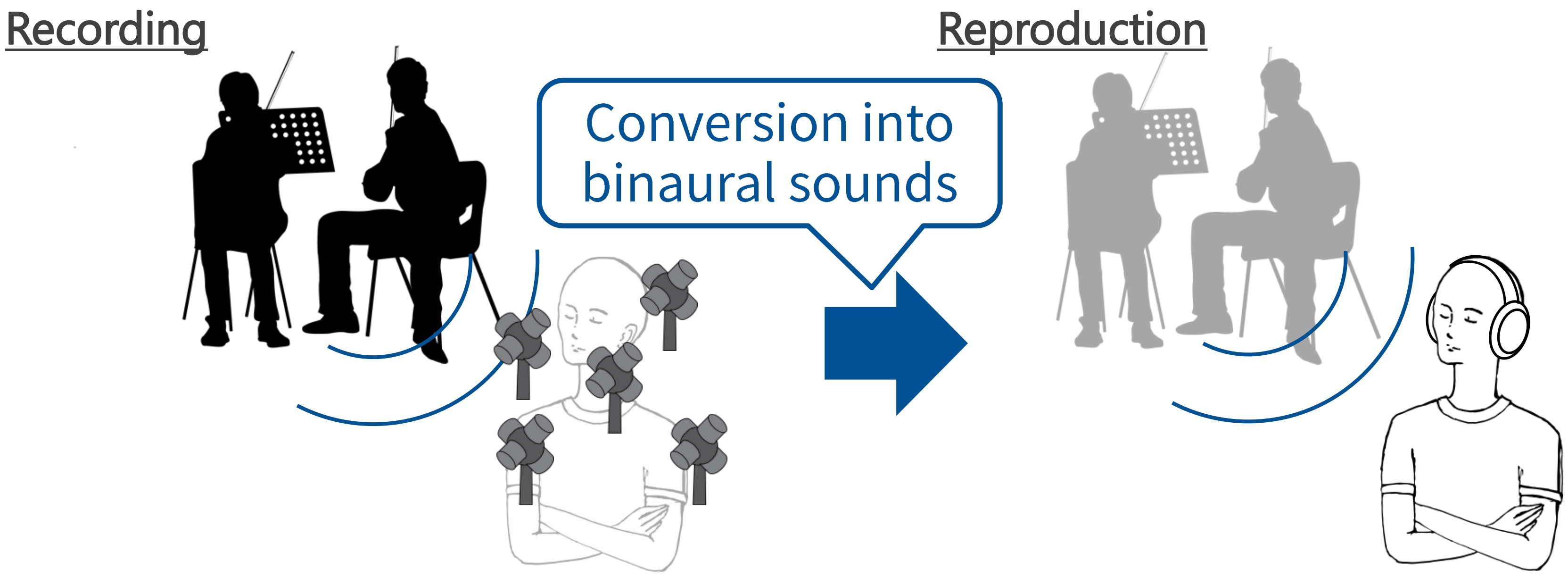
- **Estimation exploiting prior information on source direction** [Ueno+ 2021]
  - Based on directional weighting for norm of sound field
  - Enhancing estimation accuracy based on prior information

- **Learning-based approach** [Horiuchi+ 2021]
  - Modeling by weighted sum of multiple kernel functions
  - Multiple kernel learning to adapt parameters of kernel functions to environment

Application to binaural reproduction

Binaural reproduction from mic array recordings for VR audio

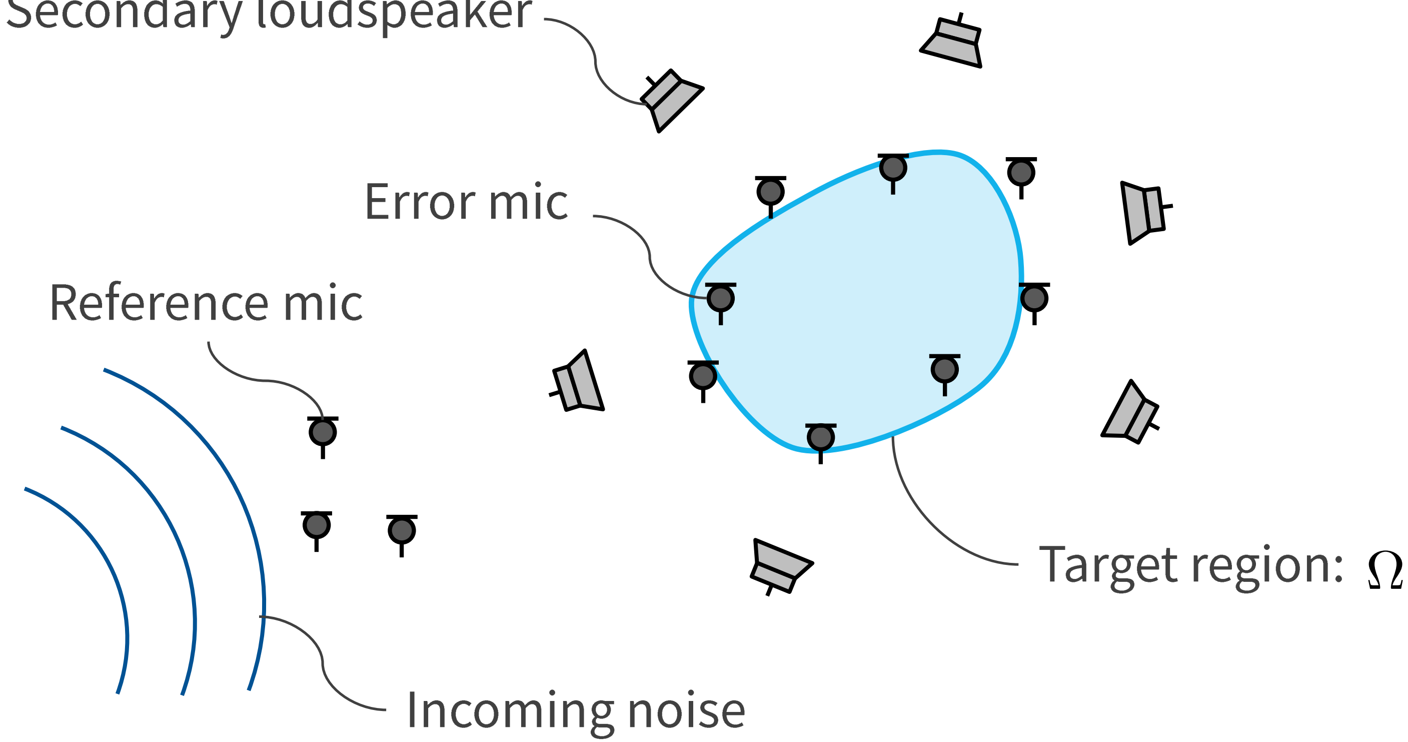
[Iijima+ 2021]



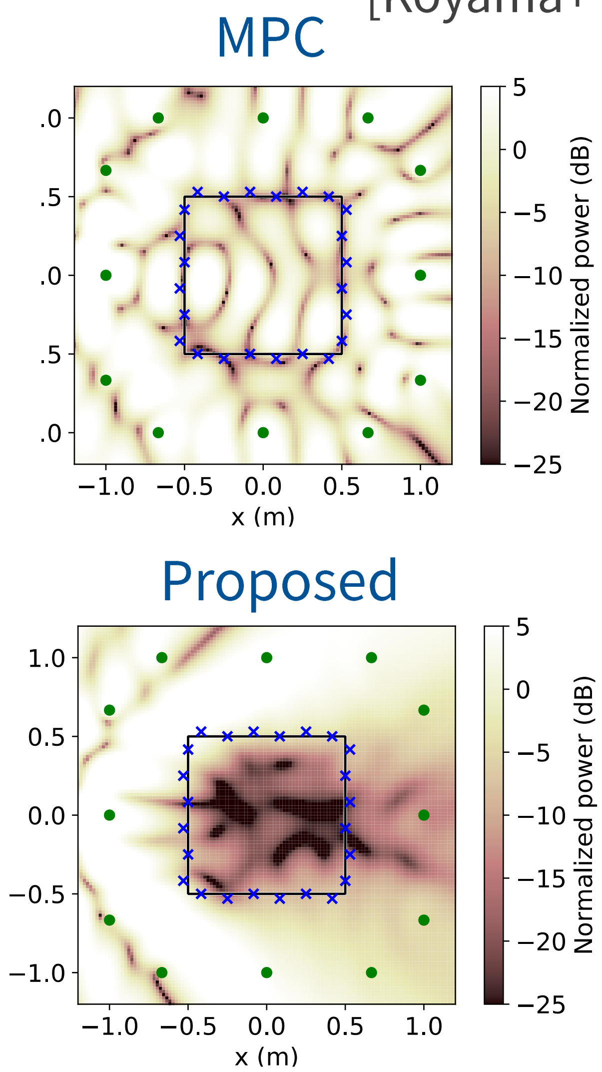
- Binaural reproduction from recordings of multiple small arrays instead of single spherical array
- Broad listening area by using flexible and scalable recording system
- Demo available on YouTube https://youtu.be/tsGIITmQiug

Application to spatial active noise control

Suppression noise over spatial target region by using multiple loudspeakers



[Koyama+ 2021]

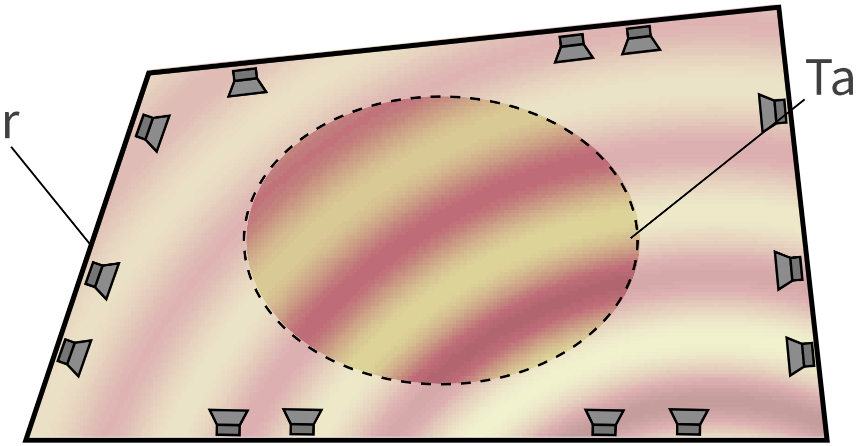


Regional noise reduction by predicting sound field based on kernel interpolation

SOUND FIELD SYNTHESIS

Sound field synthesis

Synthesizing desired pressure field w/ multiple loudspeakers



- Two major categories of sound field synthesis:
  - **Analytical approach** based on boundary integral equation:
    - Fast and stable computation, but array geometry must be simple
  - **Numerical approach** based on minimization of squared error:
    - Flexible array geometry, but computational cost is relatively high

Problem formulation

Goal: Synthesizing desired sound field $u_{\mathrm{des}}(\boldsymbol{r}, \omega)$ inside Ω with L secondary sources (loudspeakers)

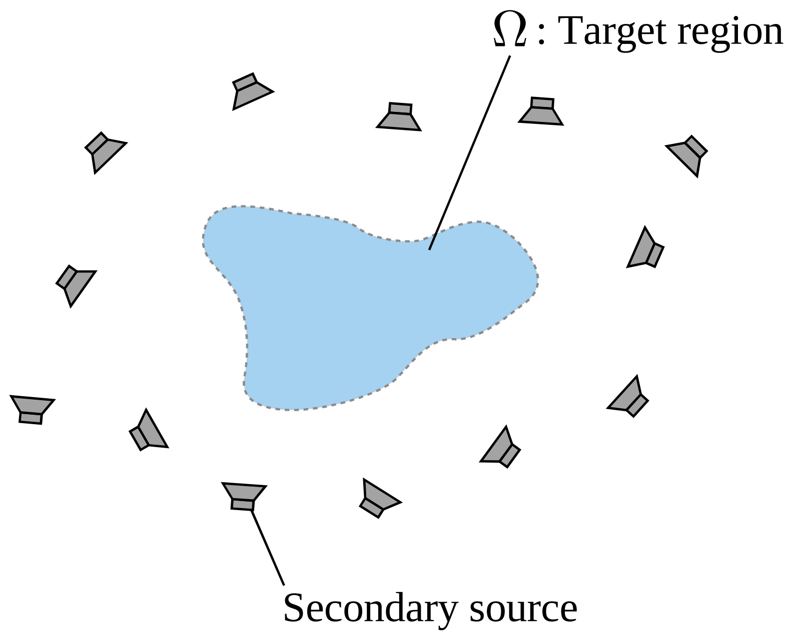
- Optimization problem to be solved

$$\underset{\{d_l\}_{l=1}^{L}}{\mathrm{minimize}}\, J := \int_{\Omega} \left| \sum_{l=1}^{L} d_l g_l(\boldsymbol{r}) - u_{\mathrm{des}}(\boldsymbol{r}) \right|^2 \mathrm{d}\boldsymbol{r}$$

Synthesized sound field

- d_l : Driving signal of lth secondary sources
- $g_l(\boldsymbol{r})$: Transfer function of lth secondary source

➡ Difficult to solve owing to regional integration



Pressure matching

- Discretize target region Ω into $N\ (\geq L)$ control points
- Optimization problem for pressure matching becomes simple least-squares problem

Driving signal vector

Desired pressure vector

$$\underset{\boldsymbol{d}\in\mathbb{C}^L}{\text{minimize}} \left\|\boldsymbol{G}\boldsymbol{d} - \boldsymbol{u}^{\text{des}}\right\|^2 + \eta\|\boldsymbol{d}\|^2$$

Transfer function matrix

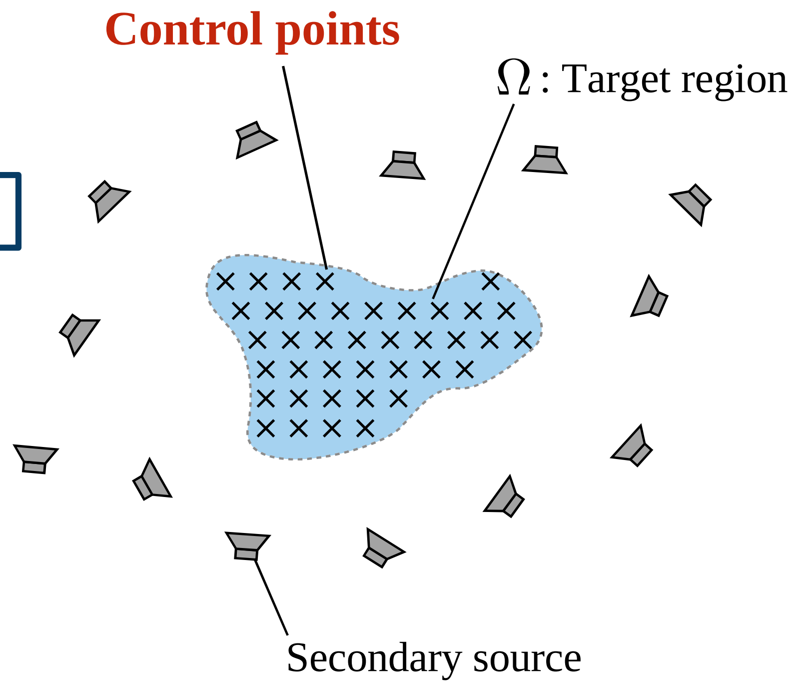
Regularization term

➡ Closed-form solution is obtained as

$$\boldsymbol{d} = \left(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} + \eta\boldsymbol{I}\right)^{-1}\boldsymbol{G}^{\mathsf{H}}\boldsymbol{u}^{\text{des}}$$

☺ Simple implementation

☹ Fine discretization of Ω is necessary



Weighted pressure matching

How can we take region between control points into consideration?

– Pressure matching is simple for implementation, but there is no guarantee that sound field in region between control points is accurately synthesized

– Mode Matching/ Weighted Mode Matching [Poletti 2005, Ueno+ 2019] can be used to synthesize continuous sound field based on expansion representation, but sometimes implementation is costly

Our idea: Incorporating sound field interpolation technique into pressure matching

- Continuous sound field estimated from measurements at control points is synthesized
- Resulting algorithm is still simple for implementation

Weighted pressure matching

Pressure matching for continuous region based on kernel interpolation of sound field

- Transfer functions $\{g_l(\boldsymbol{r})\}_{l=1}^{L}$ and desired sound field $u_{\mathrm{des}}(\boldsymbol{r})$ are estimated from those at control points:

$$\hat{g}_l(\boldsymbol{r}) = \boldsymbol{\kappa}_l(\boldsymbol{r})^{\mathsf{T}} \left(\boldsymbol{K}_l + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{g}_l := \boldsymbol{z}_l(\boldsymbol{r})^{\mathsf{T}} \boldsymbol{g}_l$$

$$\hat{u}_{\mathrm{des}}(\boldsymbol{r}) = \boldsymbol{\kappa}^{\mathrm{des}}(\boldsymbol{r})^{\mathsf{T}} \left(\boldsymbol{K}^{\mathrm{des}} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{u}^{\mathrm{des}} := \boldsymbol{z}^{\mathrm{des}}(\boldsymbol{r})^{\mathsf{T}} \boldsymbol{u}^{\mathrm{des}}$$

$\boldsymbol{g}_l$: lth column vector of $\boldsymbol{G}$
$\boldsymbol{z}_l$: Interpolation filter for lth secondary source
$\boldsymbol{z}^{\mathrm{des}}$: Interpolation filter for desired sound field

Weighted pressure matching

Pressure matching for continuous region based on kernel interpolation of sound field

- Original cost function is approximated as

$$
\begin{aligned}
J &\approx \int_{\Omega} \left| \sum_{l=1}^{L} d_l \hat{g}_l(\boldsymbol{r}) - \hat{u}_{\mathrm{des}}(\boldsymbol{r}) \right|^2 \mathrm{d}\boldsymbol{r} \\
&= \boldsymbol{d}^{\mathsf{H}} \boldsymbol{W}_{gg} \boldsymbol{d} - \boldsymbol{d}^{\mathsf{H}} \boldsymbol{W}_{gu} \boldsymbol{u}^{\mathrm{des}} + C
\end{aligned}
$$

$$
\boldsymbol{W}_{gg} = \int_{\Omega} \hat{\boldsymbol{g}}(\boldsymbol{r})^* \hat{\boldsymbol{g}}(\boldsymbol{r})^{\mathsf{T}} \mathrm{d}\boldsymbol{r}
$$

$$
\boldsymbol{W}_{gu} = \int_{\Omega} \hat{\boldsymbol{g}}(\boldsymbol{r})^* \boldsymbol{z}^{\mathrm{des}}(\boldsymbol{r})^{\mathsf{T}} \mathrm{d}\boldsymbol{r}
$$

Term unrelated to optimization

Weighted pressure matching

Pressure matching for continuous region based on kernel interpolation of sound field

- Optimal driving signal is obtained by solving

$$\underset{\boldsymbol{d}\in\mathbb{C}^L}{\text{minimize}}\, \boldsymbol{d}^{\mathsf{H}}\boldsymbol{W}_{gg}\boldsymbol{d} - \boldsymbol{d}^{\mathsf{H}}\boldsymbol{W}_{gu}\boldsymbol{u}^{\text{des}} + \eta\|\boldsymbol{d}\|^2$$

$$\Rightarrow \quad \hat{\boldsymbol{d}} = \left(\boldsymbol{W}_{gg} + \eta\boldsymbol{I}\right)^{-1}\boldsymbol{W}_{gu}\boldsymbol{u}^{\text{des}}$$

Driving signals can still be obtained in closed form with W_{gg} and W_{gu} computed in advance

Weighted pressure matching

Pressure matching for continuous region based on kernel interpolation of sound field

- When using same kernel function,

$$\boldsymbol{z}_l(\boldsymbol{r})^{\mathsf{T}} = \boldsymbol{z}^{\mathrm{des}}(\boldsymbol{r})^{\mathsf{T}} = \boldsymbol{\kappa}(\boldsymbol{r})^{\mathsf{T}} \left(\boldsymbol{K} + \lambda \boldsymbol{I}\right)^{-1} := \boldsymbol{z}(\boldsymbol{r})^{\mathsf{T}}$$

$$\hat{\boldsymbol{d}} = \underset{\boldsymbol{d} \in \mathbb{C}^L}{\arg\min} \left(\boldsymbol{G}\boldsymbol{d} - \boldsymbol{u}^{\mathrm{des}}\right)^{\mathsf{H}} \boldsymbol{W} \left(\boldsymbol{G}\boldsymbol{d} - \boldsymbol{u}^{\mathrm{des}}\right)$$

$$= \left(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{G} + \eta \boldsymbol{I}\right)^{-1} \boldsymbol{G}^{\mathsf{H}}\boldsymbol{W}\boldsymbol{u}^{\mathsf{des}}$$

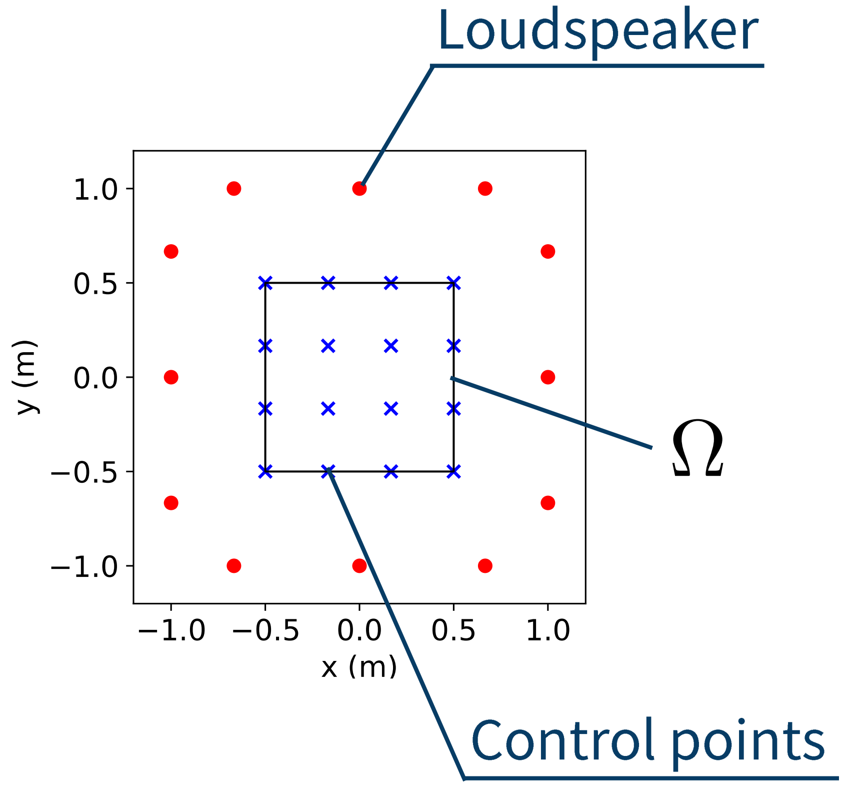
$$\boldsymbol{W} = \int_{\Omega} \boldsymbol{z}(\boldsymbol{r})^{*} \boldsymbol{z}(\boldsymbol{r})^{\mathsf{T}} \mathrm{d}\boldsymbol{r}$$

- **Simple implementation as pressure matching**
- **Equivalent to pressure matching when setting $\boldsymbol{W} = \boldsymbol{I}$**

- Setting
  - 2D free field
  - Target region Ω : square of 1.0 m x 1.0 m
  - 12 loudspeakers along square of 2.0 m x 2.0 m
  - 16 control points regularly placed over Ω
  - Desired field: plane wave (direction π/4 rad)
  - Methods:
    - Pressure matching **(PM)**
    - Weighted pressure matching **(WPM uniform/directional)**
  - Evaluation measure:

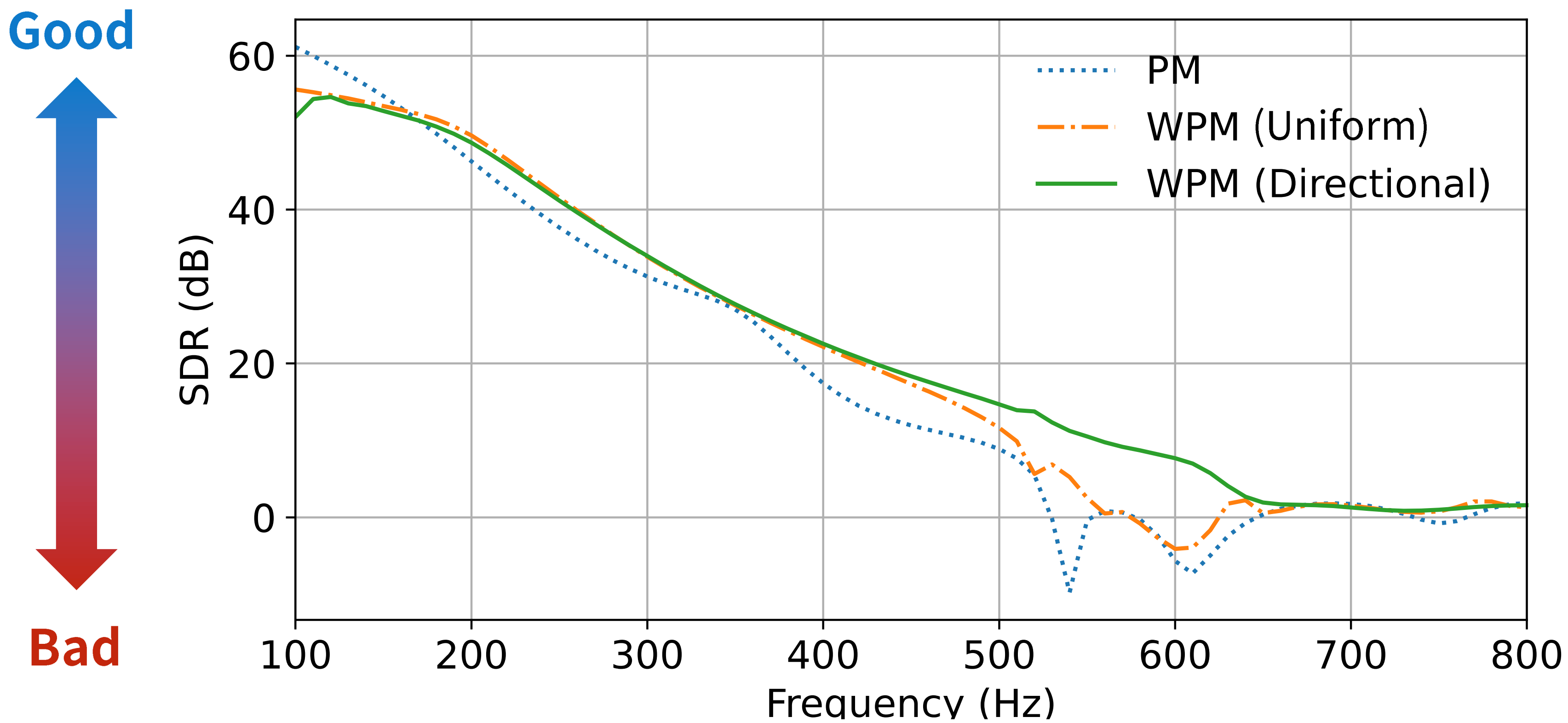
$$\mathrm{SDR}(\omega) = \frac{\int_\Omega |u_{\mathrm{des}}(\boldsymbol{r},\omega)|^2 \mathrm{d}\boldsymbol{r}}{\int_\Omega |u_{\mathrm{syn}}(\boldsymbol{r},\omega) - u_{\mathrm{des}}(\boldsymbol{r},\omega)|^2 \mathrm{d}\boldsymbol{r}}$$

Synthesized sound field



Result: Frequency vs. SDR

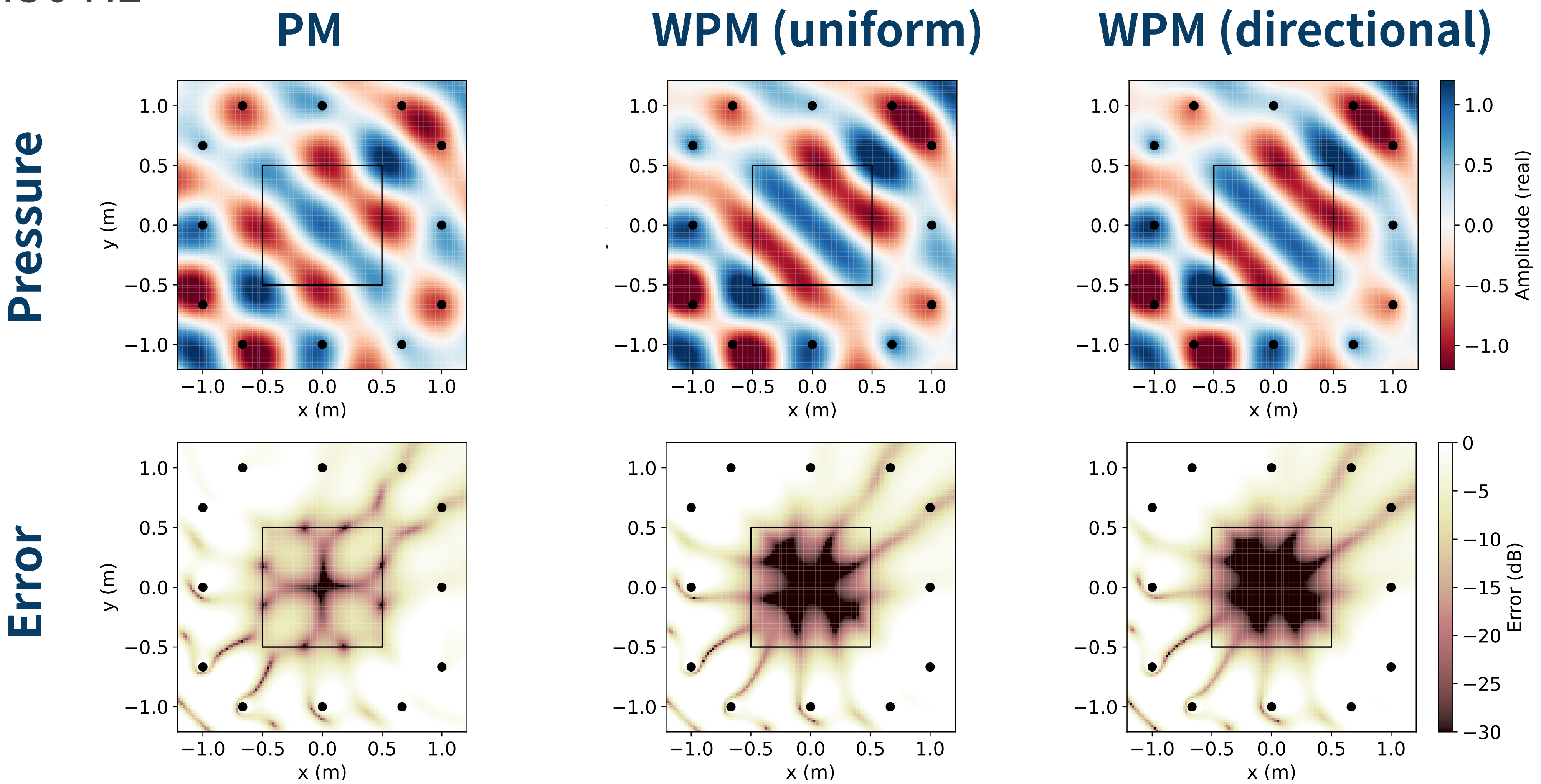
➤ SDR between 100–800Hz



WPM outperformed PM particularly at high frequencies

Result: Pressure and error distribution

➢ At 450 Hz



SDR: 11.9 dB **SDR: 17.3 dB** **SDR: 18.3 dB**

Comparison with weighted mode matching

- Weighted mode matching [Ueno+ 2019]
  - Solve weighted least-squares problem for expansion coefficients of spherical wave function

Expansion coefs of $\boldsymbol{G}$

$$\text{minimize} \left(\mathring{\boldsymbol{G}}\boldsymbol{d} - \mathring{\boldsymbol{u}}^{\text{des}}\right)^{\mathsf{H}} \boldsymbol{W}_{\text{MM}} \left(\mathring{\boldsymbol{G}}\boldsymbol{d} - \mathring{\boldsymbol{u}}^{\text{des}}\right) + \eta \|\boldsymbol{d}\|^2$$

Expansion coefs of $\boldsymbol{u}^{\text{des}}$

$$\boldsymbol{W}_{\text{MM}} = \int_{\Omega} \boldsymbol{\psi}(\boldsymbol{r})^* \boldsymbol{\psi}(\boldsymbol{r})^{\mathsf{T}} \mathrm{d}\boldsymbol{r}$$

Spherical wavefunctions

  - Closed-form solution

$$\hat{\boldsymbol{d}} = \left(\mathring{\boldsymbol{G}}^{\mathsf{H}} \boldsymbol{W}_{\text{MM}} \mathring{\boldsymbol{G}} + \eta \boldsymbol{I}\right)^{-1} \mathring{\boldsymbol{G}}^{\mathsf{H}} \boldsymbol{W}_{\text{MM}} \mathring{\boldsymbol{u}}^{\text{des}}$$

➡ **Equivalent to weighted pressure matching when expansion coefficients are estimated by kernel ridge regression**

[Koyama+ 2022 (in press)]

How to avoid spatial aliasing artifacts

- Owing to discrete placement of secondary sources (and control points), spatial aliasing artifacts are unavoidable in sound field synthesis methods
- Significant decrease in synthesis accuracy at high frequencies:
  - Degradation of sound localization
  - Coloration of source signals
- Optimal source (/sensor) placement [Koyama+ 2020, Kimura+ 2021] is one of the solutions, but still has limitation

Our idea: Synthesizing amplitude distribution leaving phase distribution arbitrary at high frequencies

- Interaural level difference (ILD) is dominant cue for horizontal sound localiz ation above 1500 Hz, compared with interaural time difference (ITD)
- Amplitude response should be accurately synthesized as much as possible, rather than phase response, to alleviate coloration effects

➡ **Applying amplitude matching for high frequencies**

Amplitude matching

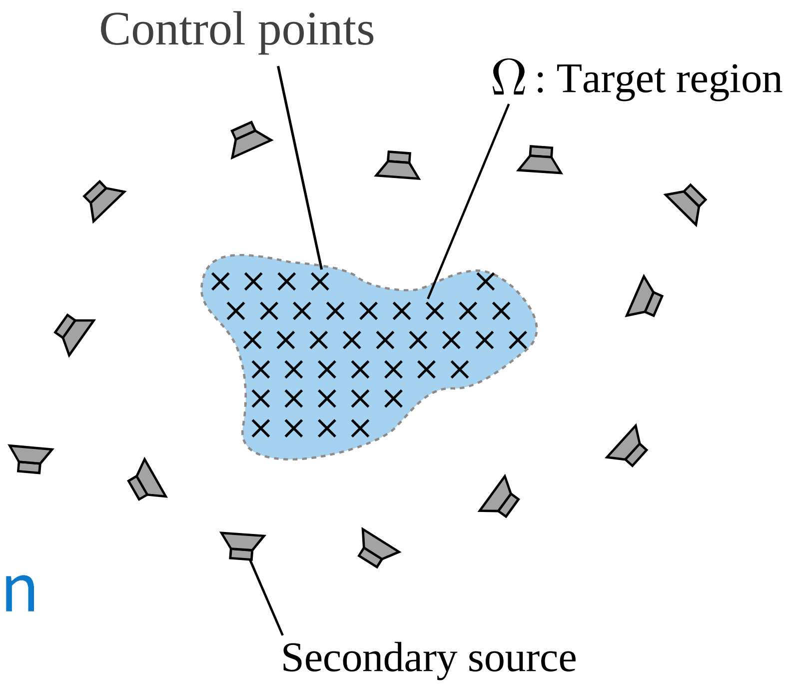
- Synthesizing desired amplitude at control points [Koyama+ 2021, Abe+ (under review)]
  - By leaving phase arbitrary, number of parameters to be control can be reduced
  - First proposed for multizone sound field control for personal audio
- Optimization problem of amplitude matching

Element-wise absolute value

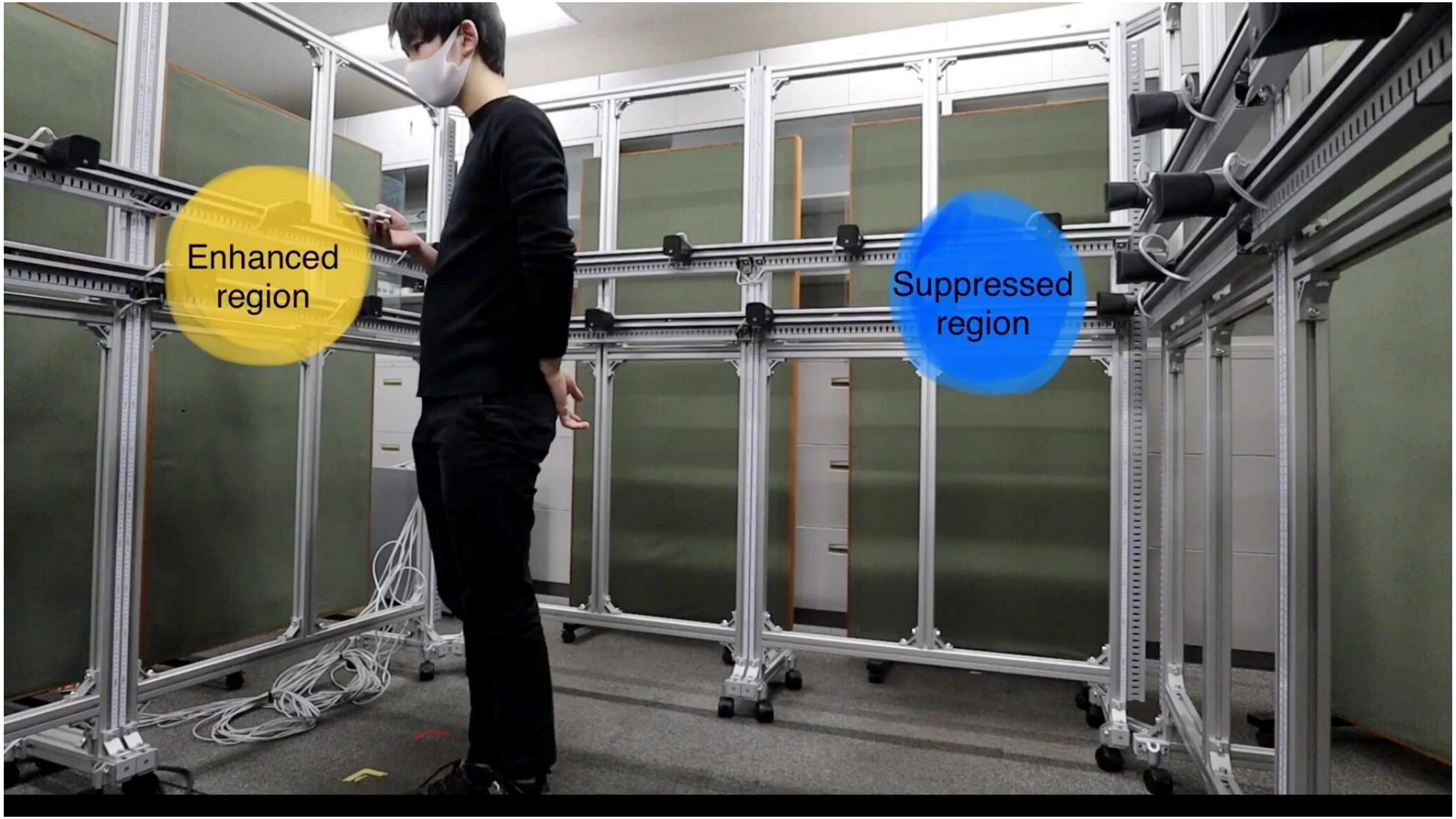
$$\underset{\boldsymbol{d}\in\mathbb{C}^L}{\text{minimize}} \left\| |\boldsymbol{G}\boldsymbol{d}| - |\boldsymbol{u}^{\text{des}}| \right\|^2 + \lambda \|\boldsymbol{d}\|^2$$

Desired amplitude

No closed form solution, but majorization minimization (MM) algorithm or alternating direction method of multipliers (ADMM) can be applied



Amplitude matching



Full version: https://youtu.be/MZKZofGI_q0

Proposed method for perceptual quality enhancement

- Combination of pressure and amplitude matching [Kimura+ (in prep)]

$$\underset{\boldsymbol{d}\in\mathbb{C}^L}{\text{minimize}}\, J(\boldsymbol{d}) := (1-\beta)\|\boldsymbol{G}\boldsymbol{d}-\boldsymbol{u}^{\mathrm{des}}\|_2^2 + \beta\||\boldsymbol{G}\boldsymbol{d}|-|\boldsymbol{u}^{\mathrm{des}}|\|_2^2 + \lambda\|\boldsymbol{d}\|_2^2$$

  – β is determined so that $\beta=0$ for low frequencies and $\beta=1$ for high frequencies
  – For example, β can be defined as sigmoid function

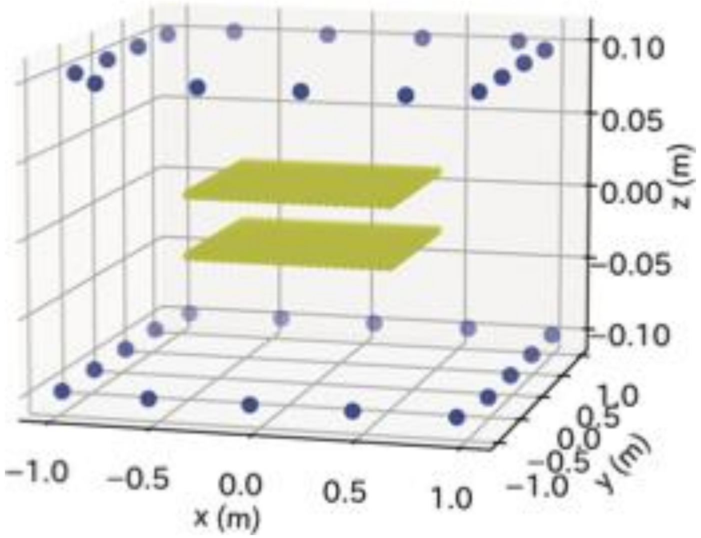
$$\beta(\omega) = \frac{1}{1+\mathrm{e}^{-\frac{\sigma}{2\pi}(\omega-\omega_{\mathrm{T}})}}$$

Transition frequency

➡ Can still be solved by MM algorithm or ADMM

Numerical experiments

- Setting
  - 3D free field
  - Target region Ω: Cuboid of 1.0 m x 1.0 m x 0.4 m
  - 32 loudspeakers on borders of squares of 2.0 m x 2.0 m at z=±0.02 m
  - 1152 control points regularly placed over Ω
  - Desired sound field: point source at (2.0 m, 0.0 m, 0.0 m)
  - Proposed method and pressure matching (PM) are compared



Numerical experiments

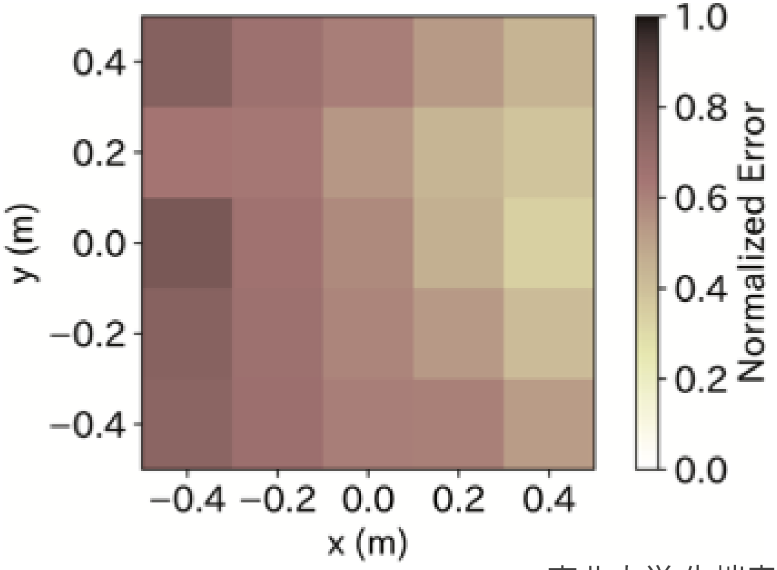
- Evaluation of ILD
  - Binaural signals in the synthesized sound field were calculated by using transfer functions from loudspeakers to a listener obtained by Mesh2HRTF [Ziegelwanger+ 2015]
  - Evaluation measure was normalized error of ILD:

$$\mathrm{NE}(\boldsymbol{r}_{\mathrm{H}}) = \frac{\sum_{\phi_{\mathrm{H}}} |\mathrm{ILD}_{\mathrm{syn}}(\boldsymbol{r}_{\mathrm{H}}, \phi_{\mathrm{H}}) - \mathrm{ILD}_{\mathrm{true}}(\boldsymbol{r}_{\mathrm{H}}, \phi_{\mathrm{H}})|}{\sum_{\phi_{\mathrm{H}}} |\mathrm{ILD}_{\mathrm{true}}(\boldsymbol{r}_{\mathrm{H}}, \phi_{\mathrm{H}})|}$$

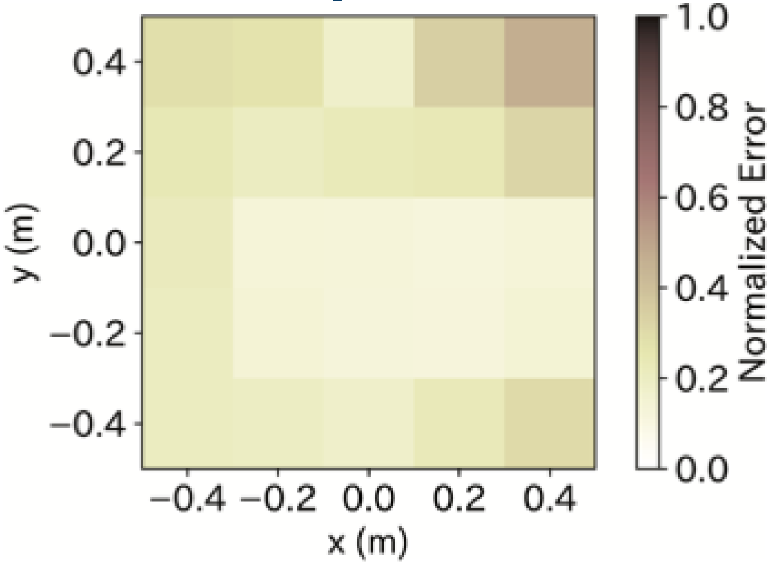
Position and direction of listener's head

  - Distribution of NE

PM



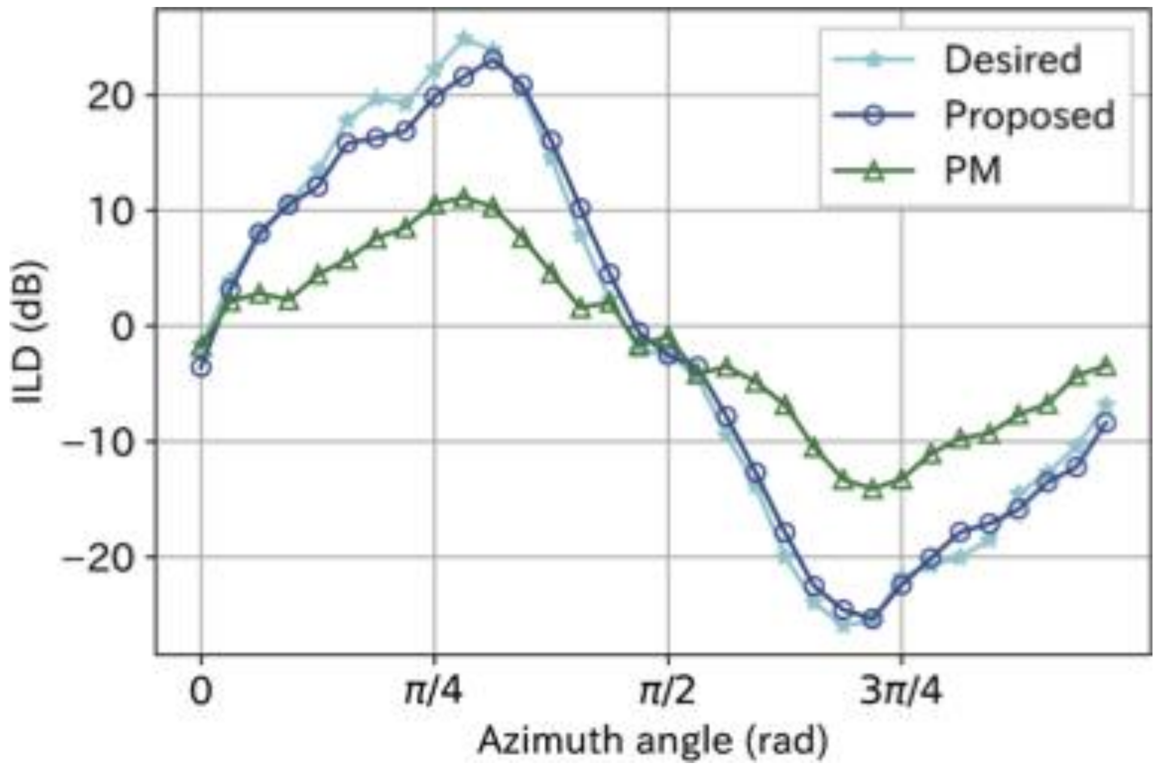
Proposed



Numerical experiments

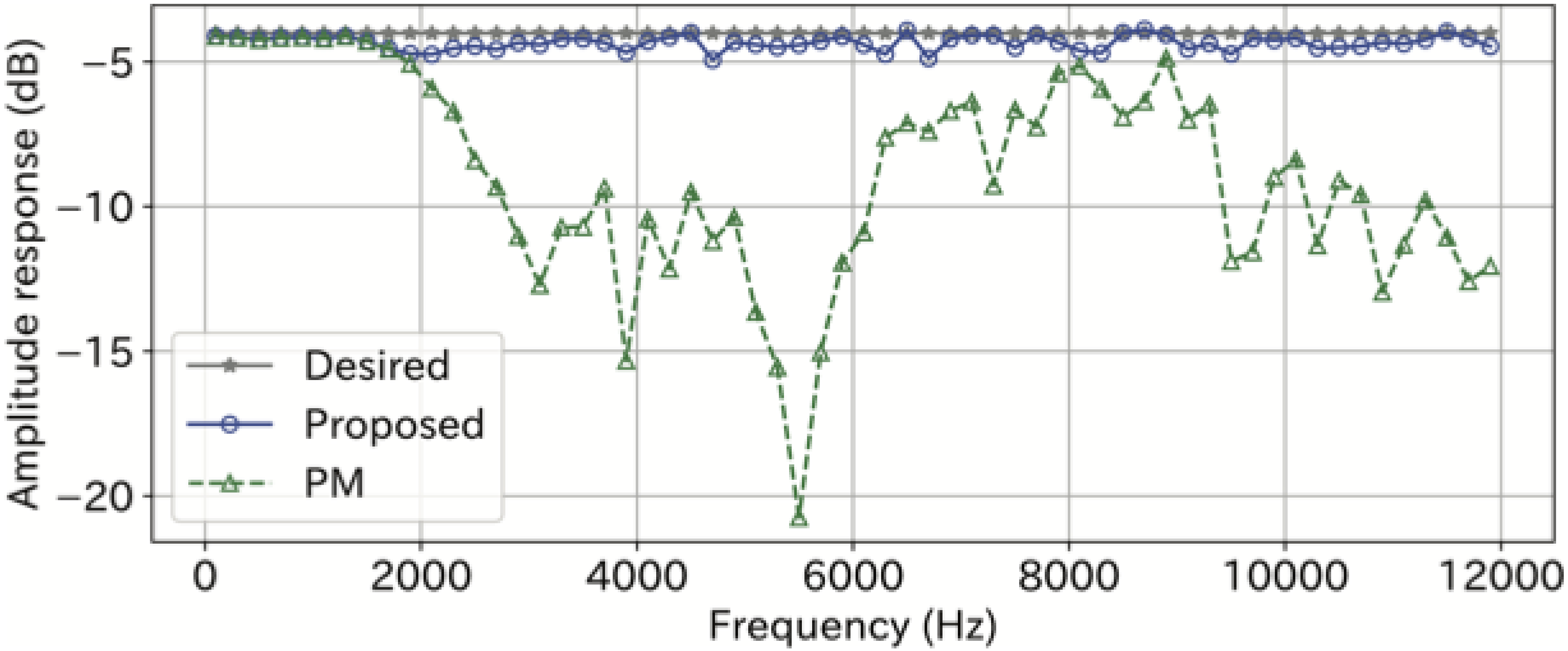
➢ Evaluation of ILD

– ILD with respect to direction of lisener's head at (0.1 m, 0.0 m)



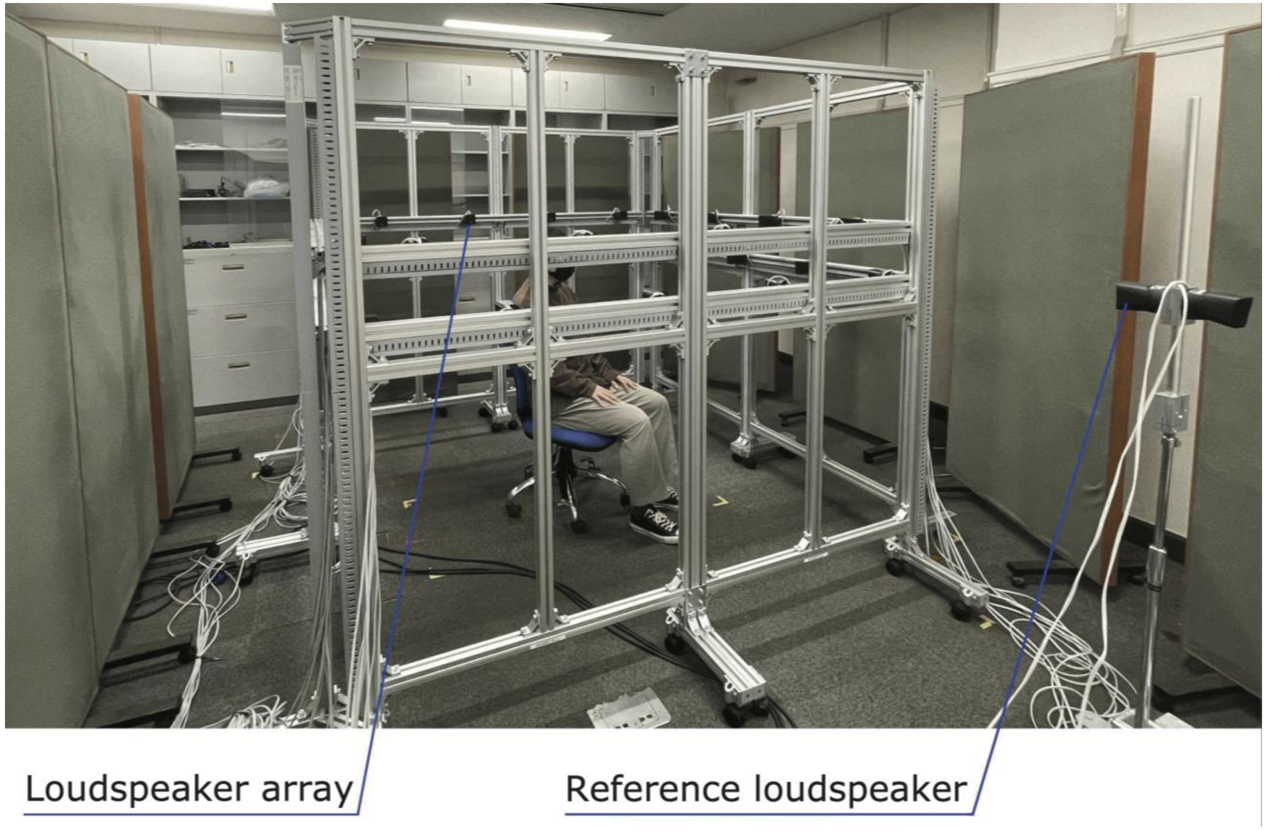
Numerical experiments

- Evaluation of amplitude response
  - At origin



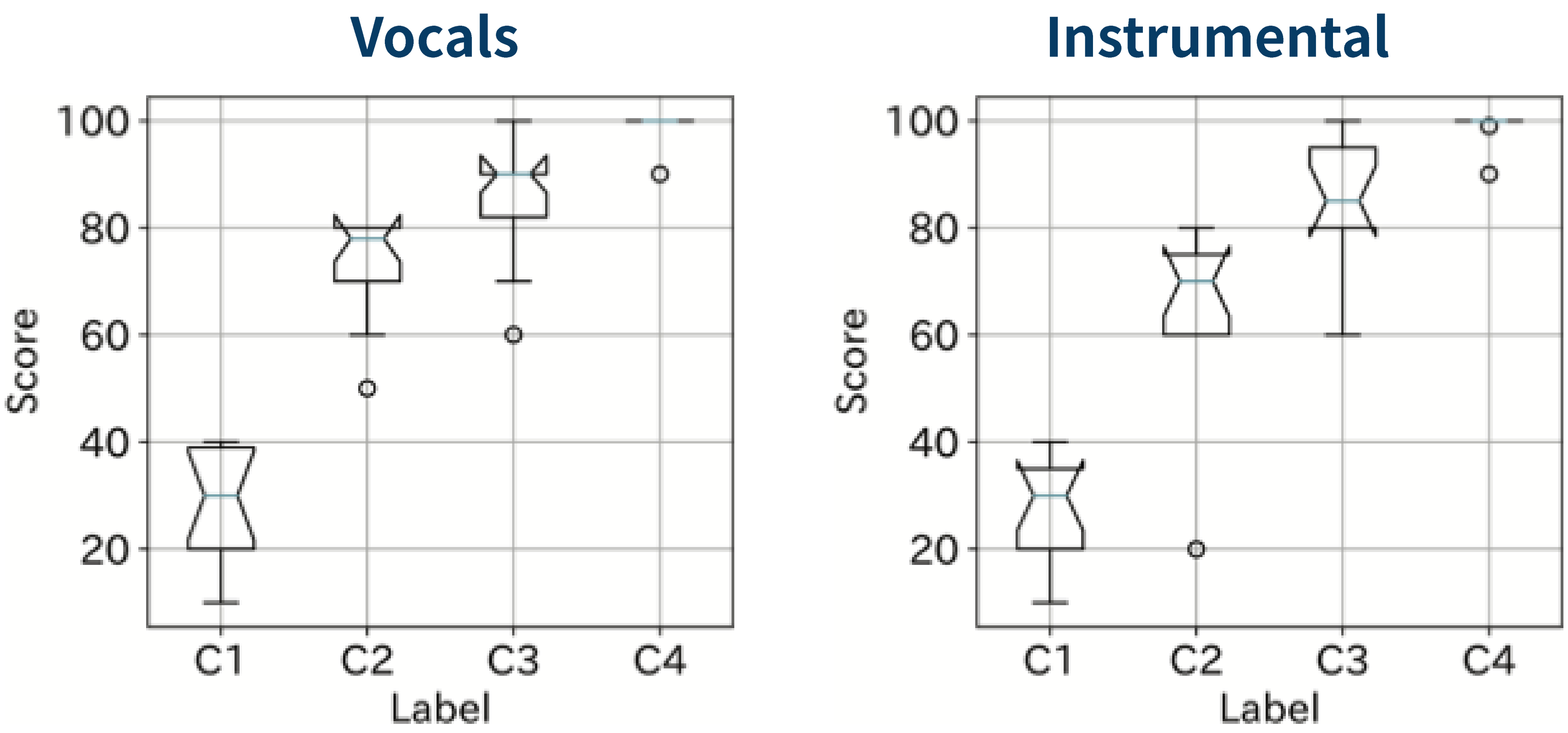
Listening experiments

- Evaluation by MUSHRA
  - Desired sound field: point source at (2.0 m, 0.5 m, 0.0 m)
  - Reverberation time (T_{60}): 0.19 s
  - 14 male subjects in 20-30s
  - Listening at center of target region
  - Test signals:
    - **Reference:** Source signal from reference loudspeaker
    - **C1/Hidden anchor:** lowpass-filtered source signal up to 3.5 kHz
    - **C2/PM:** Synthesized sound by PM
    - **C3/Proposed:** Synthesized sound by Proposed
    - **C4/Hidden reference:** Same as reference



Listening experiments

➢ Results of two source signals (Vocals/Instrumental)



C1/Hidden anchor
C2/PM
C3/Proposed
C4/Hidden reference

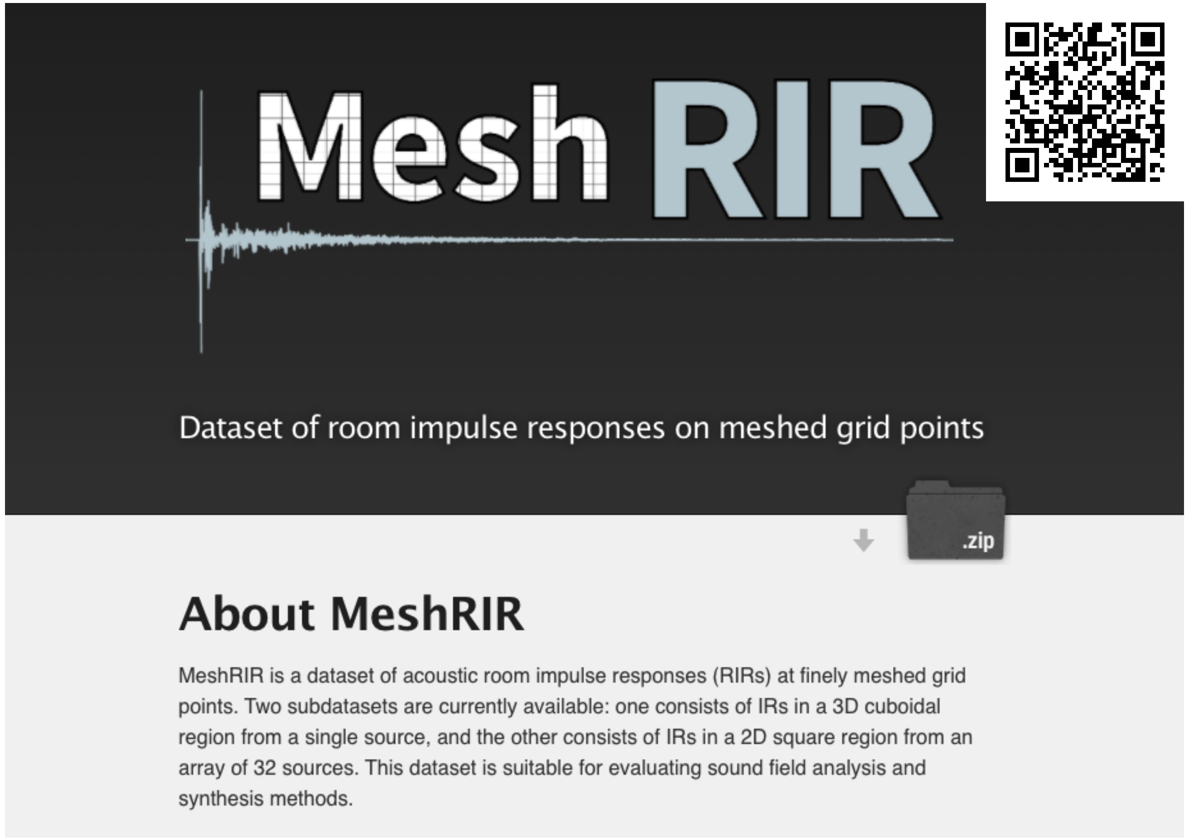
Synthesized sound by Proposed is perceptually close to reference sound compared to PM

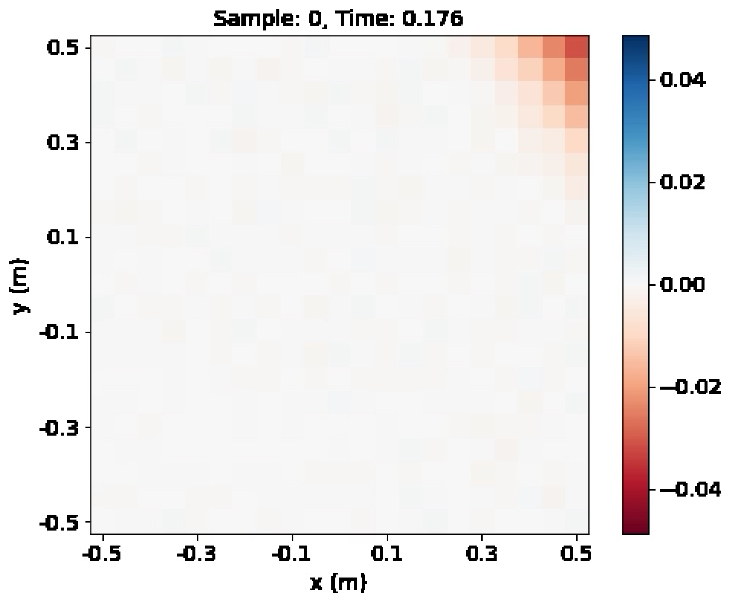
Conclusion

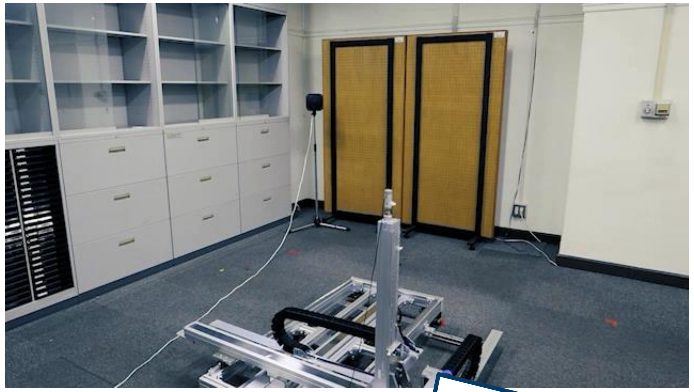
- Recent advances in sound field analysis and synthesis
  - **Sound field analysis:**
    - Overview of sound field estimation methods
    - Infinite-dimensional extension of least-squares-based sound field estimation
  - **Sound field synthesis:**
    - Weighted pressure matching
    - Combination of amplitude matching for perceptual quality enhancement

Dataset of room impulse responses (RIRs)

- Released RIR dataset on meshed grid points with example codes
  - https://sh01k.github.io/MeshRIR/







RIR measurement system

References

1. N. Ueno, S. Koyama, and H. Saruwatari, "Sound Field Recording Using Distributed Microphones Based on Harmonic Analysis of Infinite Order," IEEE SPL, 25(1), pp. 135-139, 2018.
2. N. Ueno, S. Koyama, and H. Saruwatari, "Directionally Weighted Wave Field Estimation Exploiting Prior Information on Source Direction," IEEE Trans. SP, 69, pp. 2383-2395, 2021.
3. N. Murata, S. Koyama, N. Takamune, and H. Saruwatari, "Sparse Representation Using Multidimensional Mixed-Norm Penalty With Application to Sound Field Decomposition," IEEE Trans. SP, 66(12), pp. 3327-3338, 2018.
4. S. Koyama, N. Murata, H. Saruwatari, "Sparse Sound Field Decomposition for Super-resolution in Recording and Reproduction," J. Acoust. Soc. Am., 143(6), pp. 3780-3895, 2018.
5. S. Koyama and L. Daudet, "Sparse Representation of a Spatial Sound Field in a Reverberant Environment," IEEE J. STSP, 13(1), 2019.
6. S. Koyama, "Sparsity-based sound field reconstruction," Acoust. Sci. & Tech., 41(1), pp. 269-275, 2020.
7. Y. Takida, S. Koyama, N. Ueno, H. Saruwatari, "Reciprocity gap functional in spherical harmonic domain for gridless sound field decomposition," Signal Processing, Elsevier, 169, 2020.
8. R. Horiuchi, S. Koyama, J. G. C. Ribeiro, N. Ueno, and H. Saruwatari, "Kernel Learning for Sound Field Estimation With L1 and L2 Regularizations," in Proc. IEEE WASPAA, 2021.
9. T. Nishida, N. Ueno, S. Koyama, and H. Saruwatari, "Region-restricted Sensor Placement Based on Gaussian Process for Sound Field Estimation," IEEE Trans. SP, 70, pp. 1718-1733, 2022.
10. J. G. C. Ribeiro, N. Ueno, S. Koyama, and H. Saruwatari, "Region-to-region Kernel Interpolation of Acoustic Transfer Functions Constrained by Physical Properties," IEEE/ACM Trans. ASLP, 30, pp. 2944-2954, 2022.
11. N. Iijima, S. Koyama, and H. Saruwatari, "Binaural Rendering from Microphone Array Signals of Arbitrary Geometry," J. Acoust. Soc. Am., 150(4), pp. 2479-2491, 2021.

References

12. H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, “Feedforward Spatial Active Noise Control Based on Kernel Interpolation of Sound Field,” in Proc. IEEE ICASSP, 2019.
13. S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari, “Spatial Active Noise Control Based on Kernel Interpolation of Sound Field,” IEEE/ACM Trans. ASLP, 29, pp. 3052-3063, 2021.
14. N. Ueno, S. Koyama, and H. Saruwatari, “Three-Dimensional Sound Field Reproduction Based on Weighted Mode-Matching Method,” IEEE/ACM Trans. ASLP, 27(12), pp. 1852-1867, 2019.
15. K. Kimura, S. Koyama, N. Ueno, and H. Saruwatari, “Mean-Square-Error-Based Secondary Source Placement in Sound Field Synthesis With Prior Information on Desired Field,” in Proc. IEEE WASPAA, 2021.
16. S. Koyama, G. Chardon, and L. Daudet, “Optimizing Source and Sensor Placement for Sound Field Control: An Overview,” IEEE/ACM Trans. ASLP, 28, pp. 686-714, 2020.
17. S. Koyama, T. Amakasu, N. Ueno, and H. Saruwatari, “Amplitude Matching: Majorization-Minimization Algorithm for Sound Field Control Only With Amplitude Constraint,” in Proc. IEEE ICASSP, 2021.
18. T. Abe, S. Koyama, N. Ueno, and H. Saruwatari, “Amplitude Matching for Multizone Sound Field Control,” IEEE/ACM Trans. ASLP, under review.
19. S. Koyama and K. Arikawa, “Weighted Pressure Matching Based on Kernel Interpolation for Sound Field Reproduction,” in Proc. ICA, 2022.
20. S. Koyama, K. Kimura, and N. Ueno, “Weighted Pressure and Mode Matching for Sound Field Reproduction: Theoretical and Experimental Comparisons,” J. Audio Eng. Soc., 2022. (in press)
21. S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, “MeshRIR: A Dataset of Room Impulse Responses on Meshed Grid Points for Evaluating Sound Field Analysis and Synthesis Methods,” in Proc. IEEE WASPAA, 2021.