Sound Field Recording and Reproduction and Its Extension to Super-resolution

Shoichi Koyama

Graduate school of information science and technology,

The University of Tokyo

About me

➢ **Shoichi Koyama, Ph.D.**

– 2007 B.E. and 2009 M.S. degrees from UTokyo
  - *Source localization with weighted integral method*

– 2009.04 – 2014.03 NTT Media Intelligence Lab.
  - *Sound field recording and reproduction*

NTT

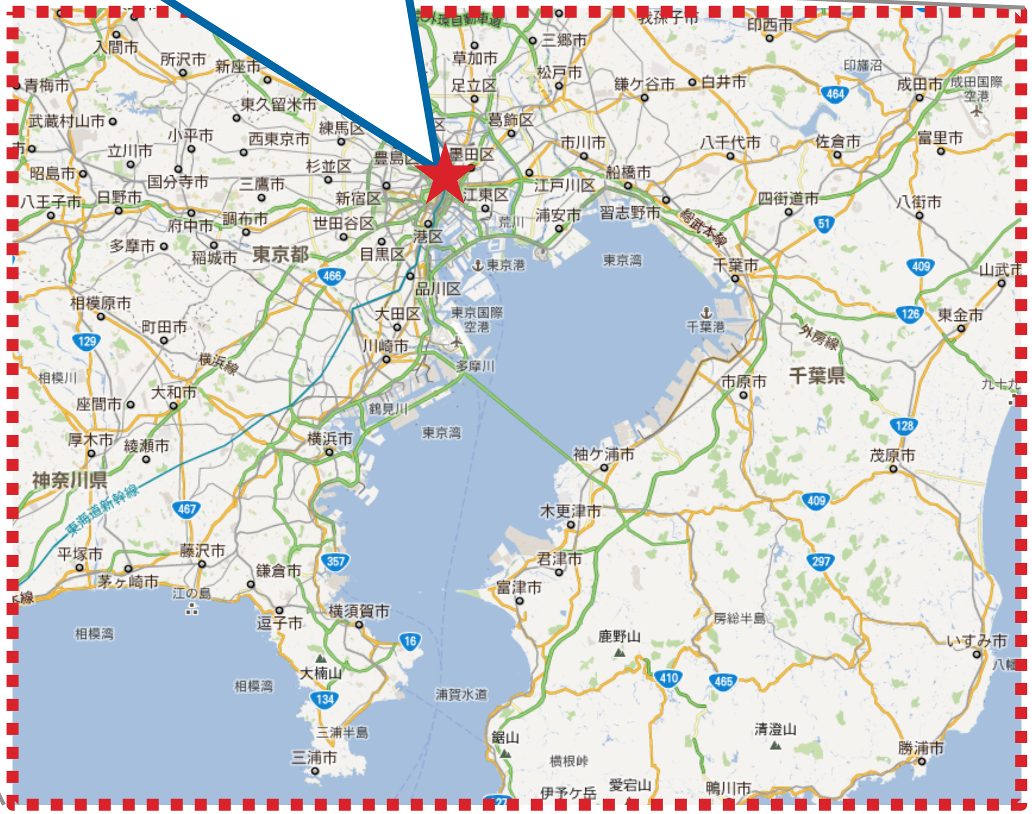
– 2014.01 Ph.D. (Inf. Sci.&Tech.) from UTokyo

– 2014.04 – Assistant Prof. (Research Associate) at UTokyo
  - *Super-resolution in sound field recording and reproduction*

東京大学
THE UNIVERSITY OF TOKYO

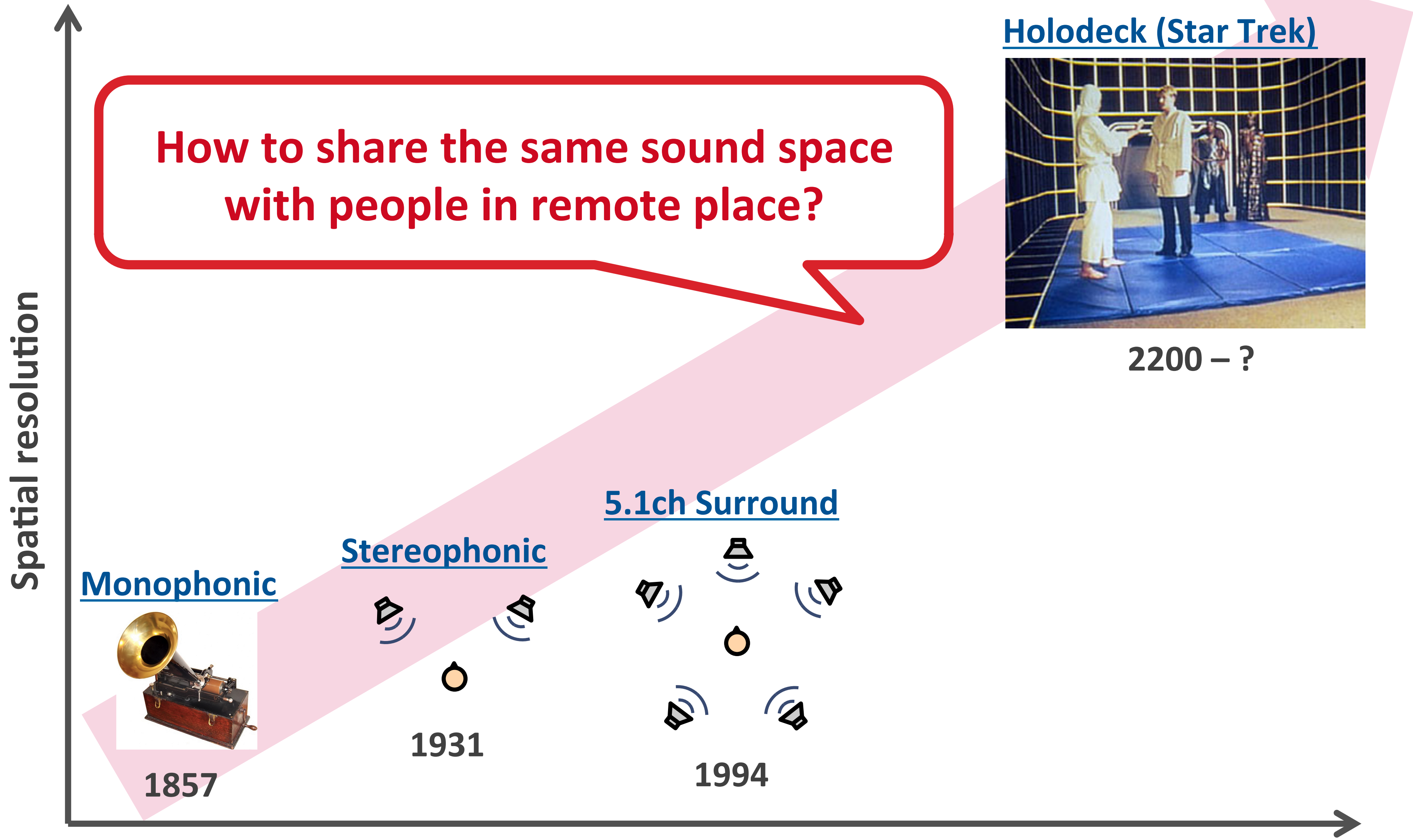
System #1 Lab., The University of Tokyo

- **Our lab has just started on Apr. 2014**
  - Staffs: Prof. Hiroshi Saruwatari & me
  - 6 students
  - Audio, speech, and music signal processing



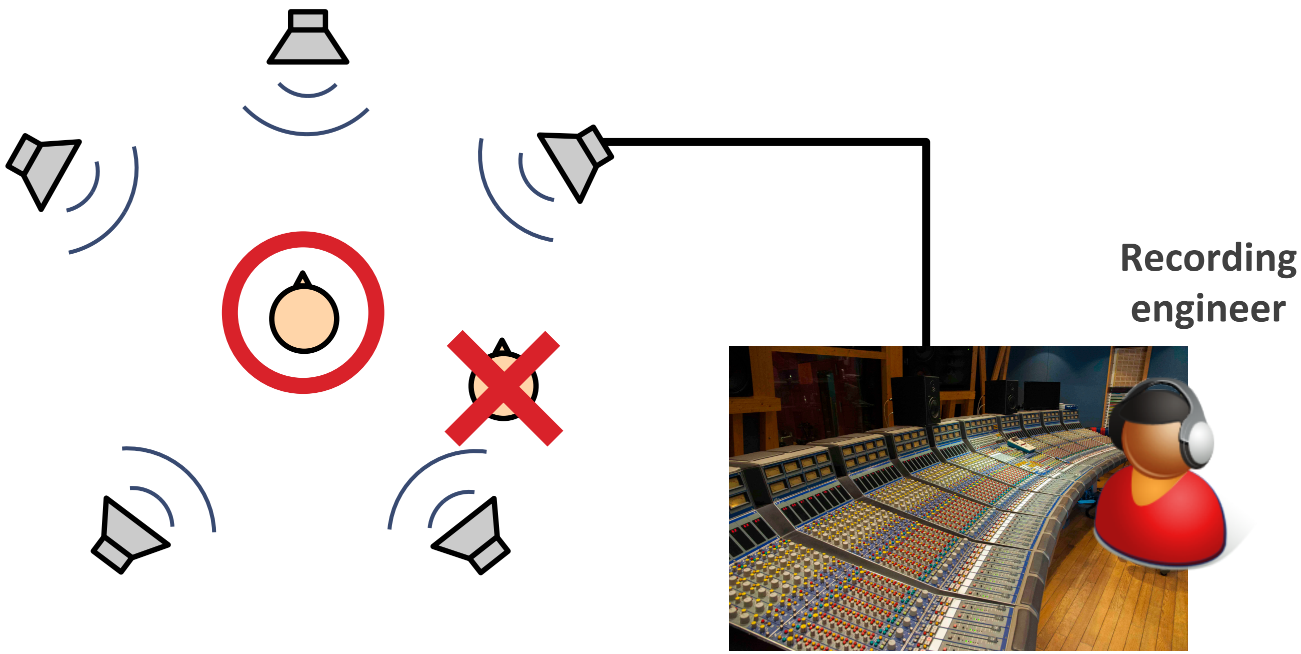
Recording and Reproduction

History of audio reproduction system



Conventional audio systems

- Stereo, 5.1ch surround, etc…



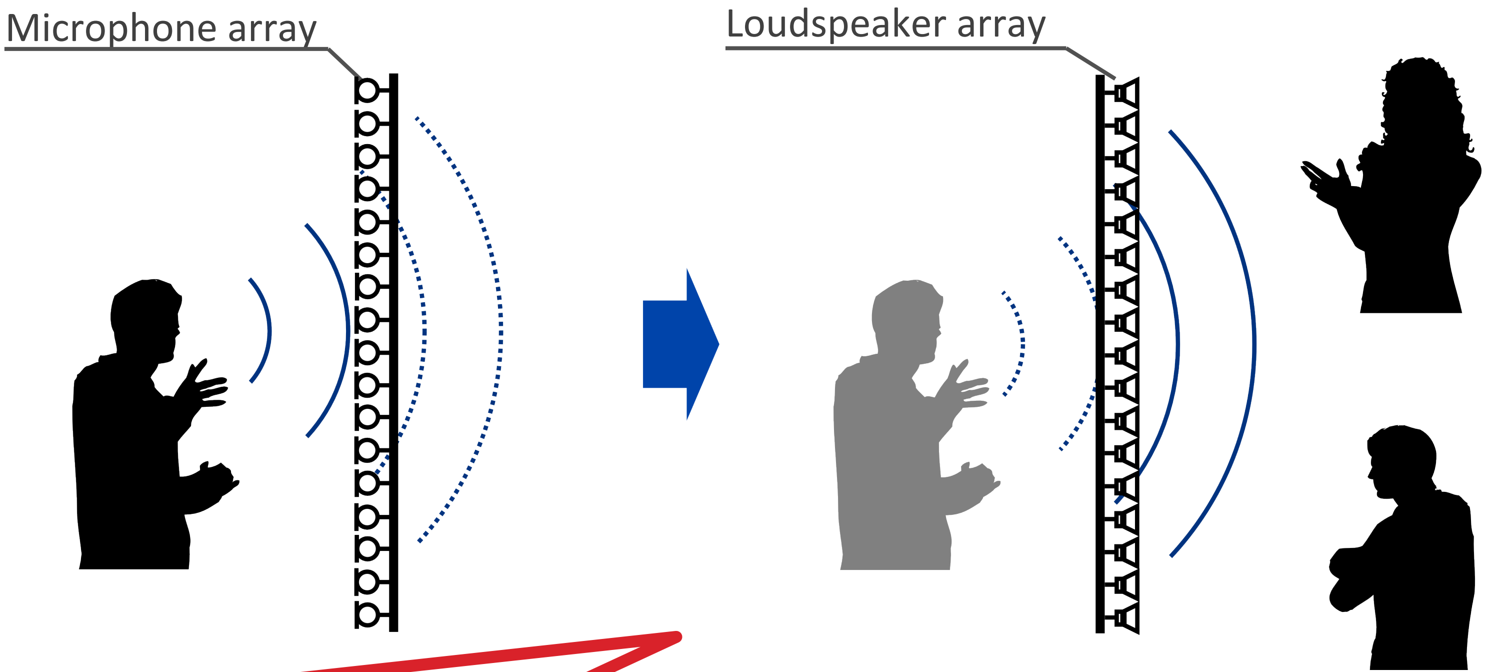
Drawbacks

☹ Listening position is limited around center of loudspeakers (sweet spot)

☹ Reproduced sound is artificially designed by recording engineers

Sound field reproduction for audio system

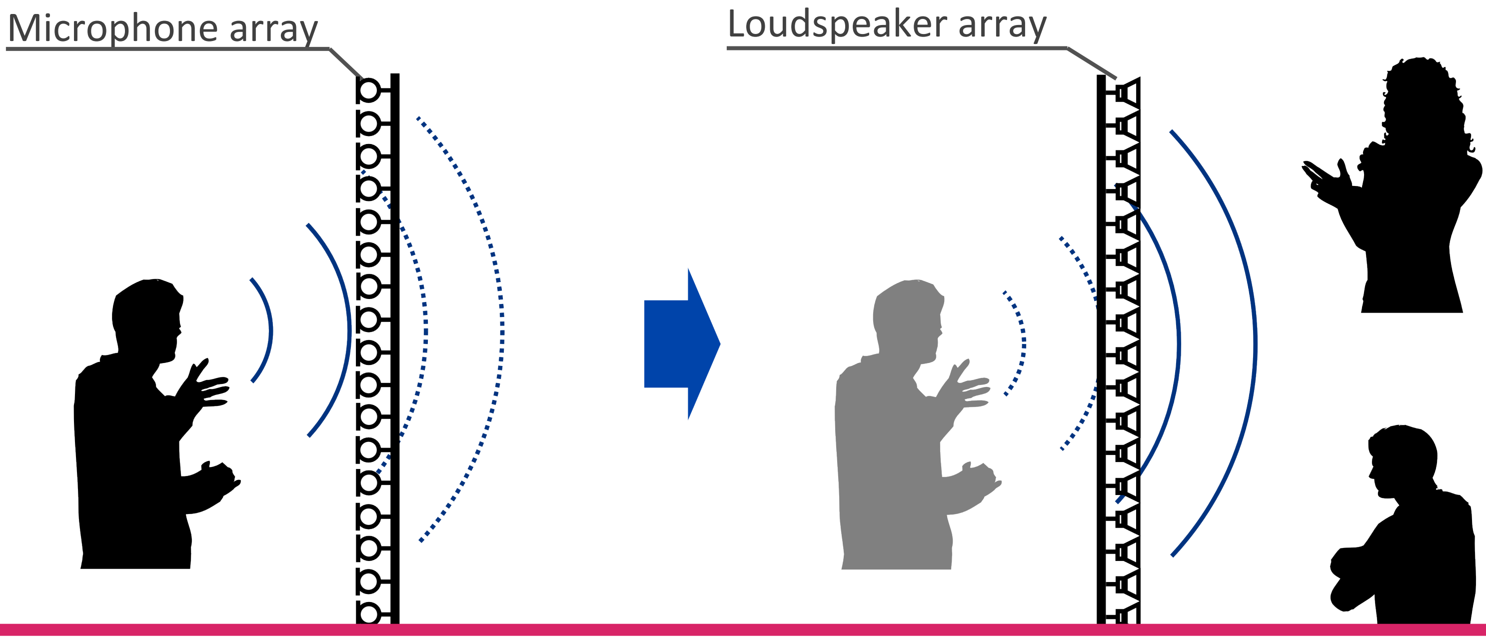
Sound field recording and reproduction



- ❑ Large listening area can be achieved
- ❑ Listeners can perceive source distance
- ❑ Real-time recording and reproduction can be achieved without recording engineers

Sound field reproduction for audio system

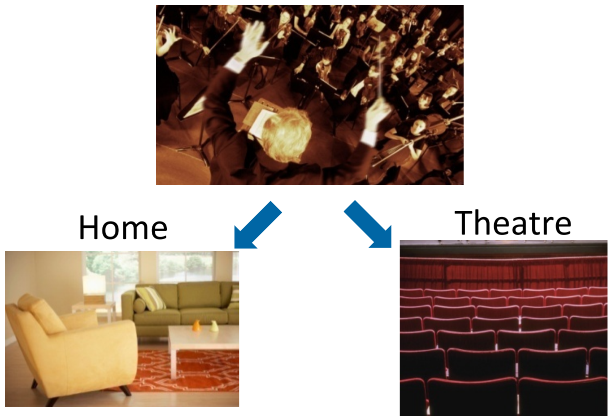
Sound field recording and reproduction



Applications

Telecommunication system

Live broadcasting



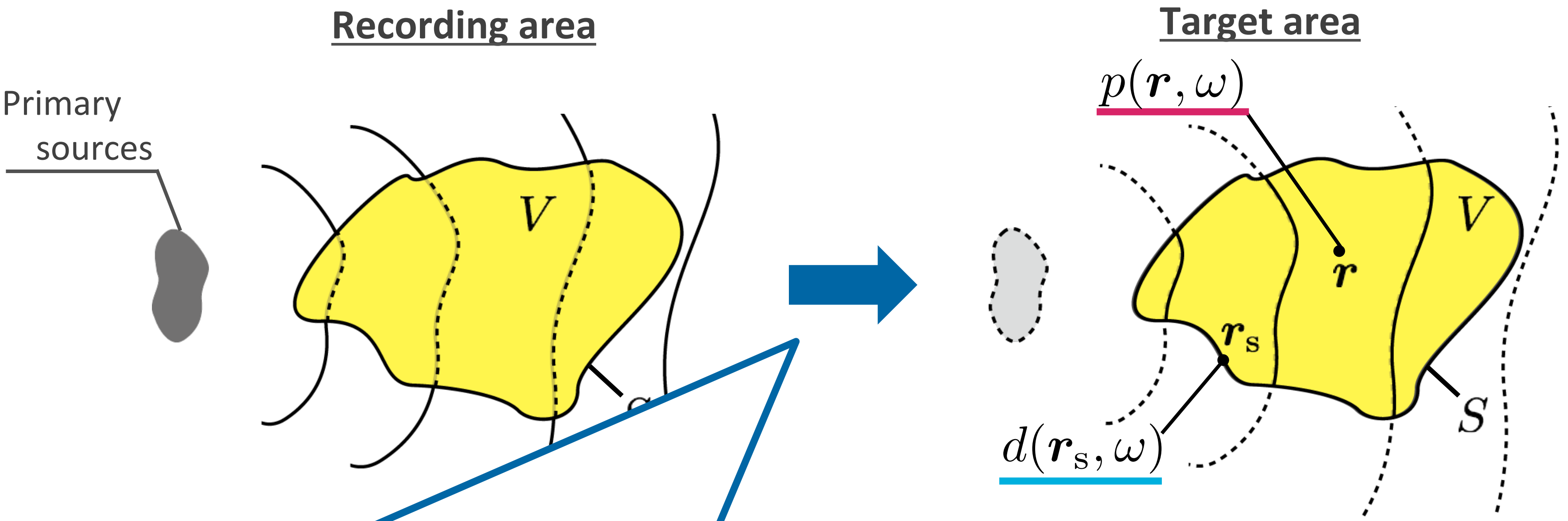
Outline

- **Signal Conversion for Sound Field Recording and Reproduction**
  - Signal conversion from signals received by microphone array into driving signals of loudspeakers for reconstruction of sound field
  - Fast and stable signal conversion method for sound field recording and reproduction is proposed
- **Super-resolution in Sound Field Recording and Reproduction**
  - Improve reproduction accuracy at frequencies above spatial Nyquist frequency determined by intervals of array elements
  - Two approaches of super-resolution are introduced

SIGNAL CONVERSION FOR SOUND FIELD RECORDING AND REPRODUCTION

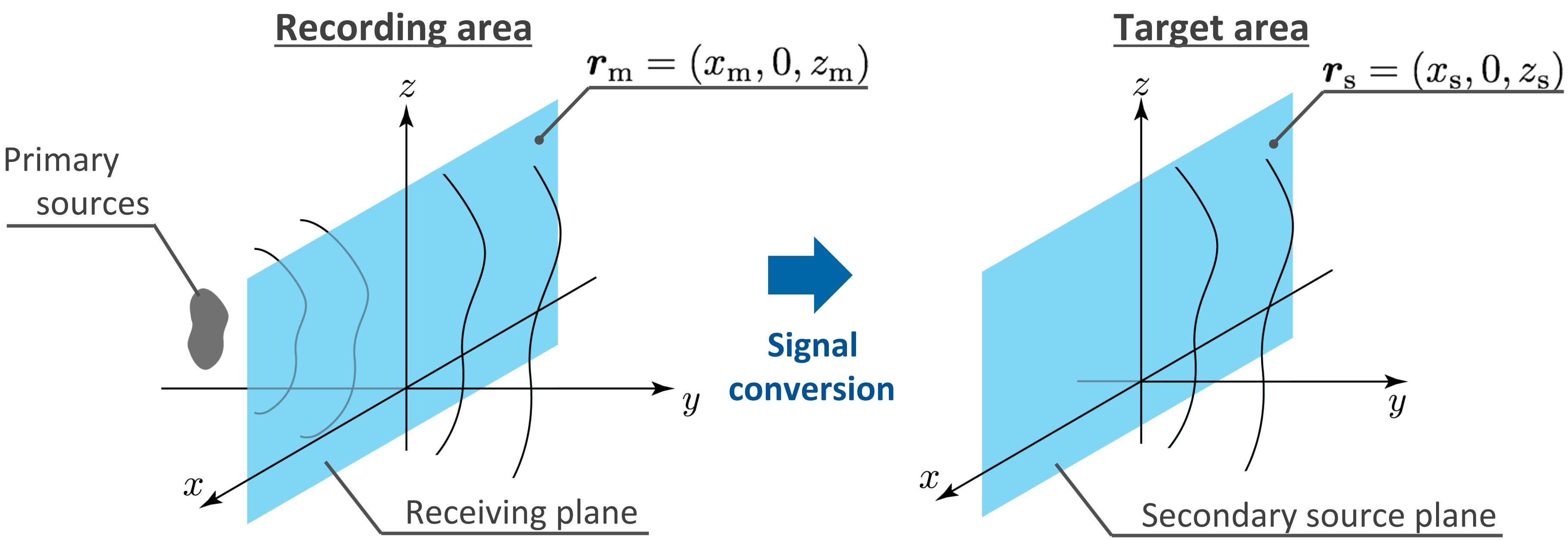
Sound Field Recording and Reproduction

Obtain driving signals $d(\boldsymbol{r}_\mathrm{s}, \omega)$ of secondary sources (= loudspeakers) arranged on S to reconstruct desired sound field inside V



- ✧ Inherently, sound pressure and its gradient on S is required to obtain $d(\boldsymbol{r}_\mathrm{s}, \omega)$, but sound pressure is usually only known
- ✧ Our objective: Fast and stable signal conversion for sound field recording and reproduction with ordinary acoustic sensors and transducers

Recording and Reproduction with Planar Arrays



Desired sound field based on Rayleigh integral

$$p_{\mathrm{des}}(\boldsymbol{r},\omega) = -2 \iint_{-\infty}^{\infty} \frac{\partial p(\boldsymbol{r}_{\mathrm{m}},\omega)}{\partial y_{\mathrm{m}}} G(\boldsymbol{r}|\boldsymbol{r}_{\mathrm{m}},\omega)\, dx_{\mathrm{m}} dz_{\mathrm{m}}$$

Green's function

Synthesized sound field by secondary sources

$$p_{\mathrm{syn}}(\boldsymbol{r},\omega) = \iint_{-\infty}^{\infty} d(\boldsymbol{r}_{\mathrm{s}},\omega) H(\boldsymbol{r}-\boldsymbol{r}_{\mathrm{s}},\omega)\, dx_{\mathrm{s}} dz_{\mathrm{s}}$$

Transfer function

Approximate secondary sources as point source ($H \simeq G$)

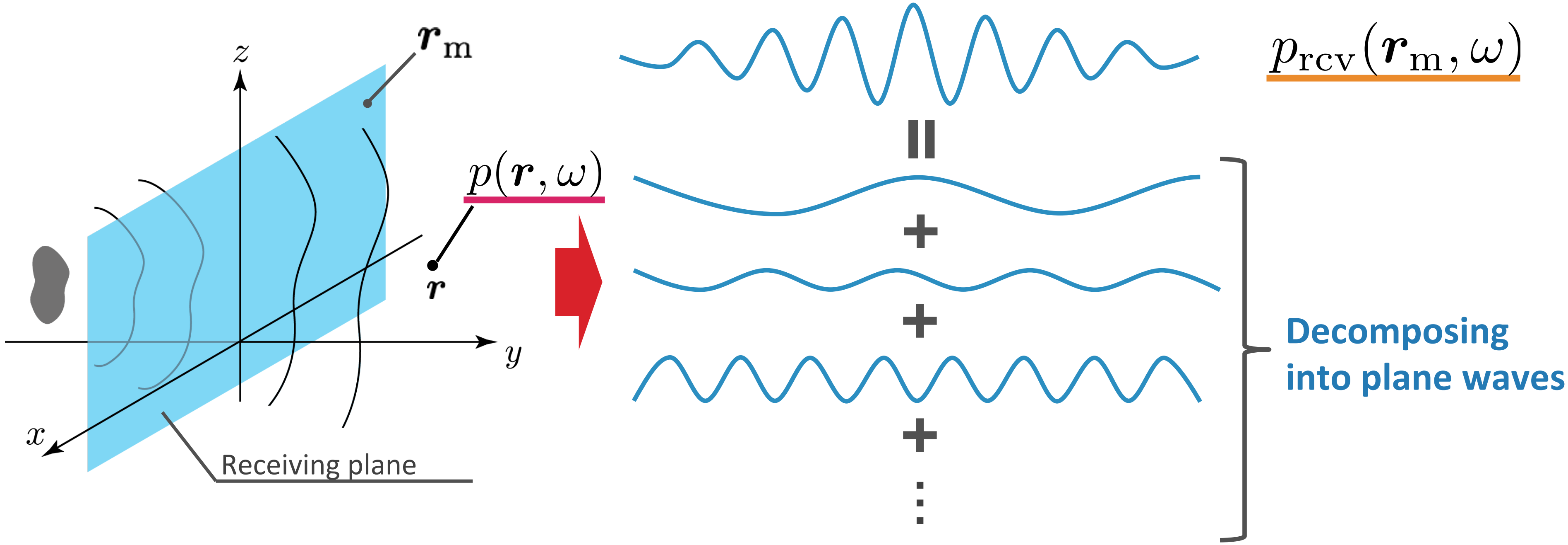
$$d(\boldsymbol{r}_{\mathrm{s}},\omega) = -2 \frac{\partial p(\boldsymbol{r}_{\mathrm{m}},\omega)}{\partial y_{\mathrm{m}}}$$

☹ $p(\boldsymbol{r}_{\mathrm{m}},\omega)$ **is only known**

Distribution of sound pressure gradient needs to be estimated

Formulation in Spatial Frequency Domain

Sound pressure distribution is represented as a sum of plane waves



Since each plane wave uniquely determines propagation in y direction, sound pressure gradient can be estimated

➡ **Plane wave decomposition corresponds to spatial Fourier transform**

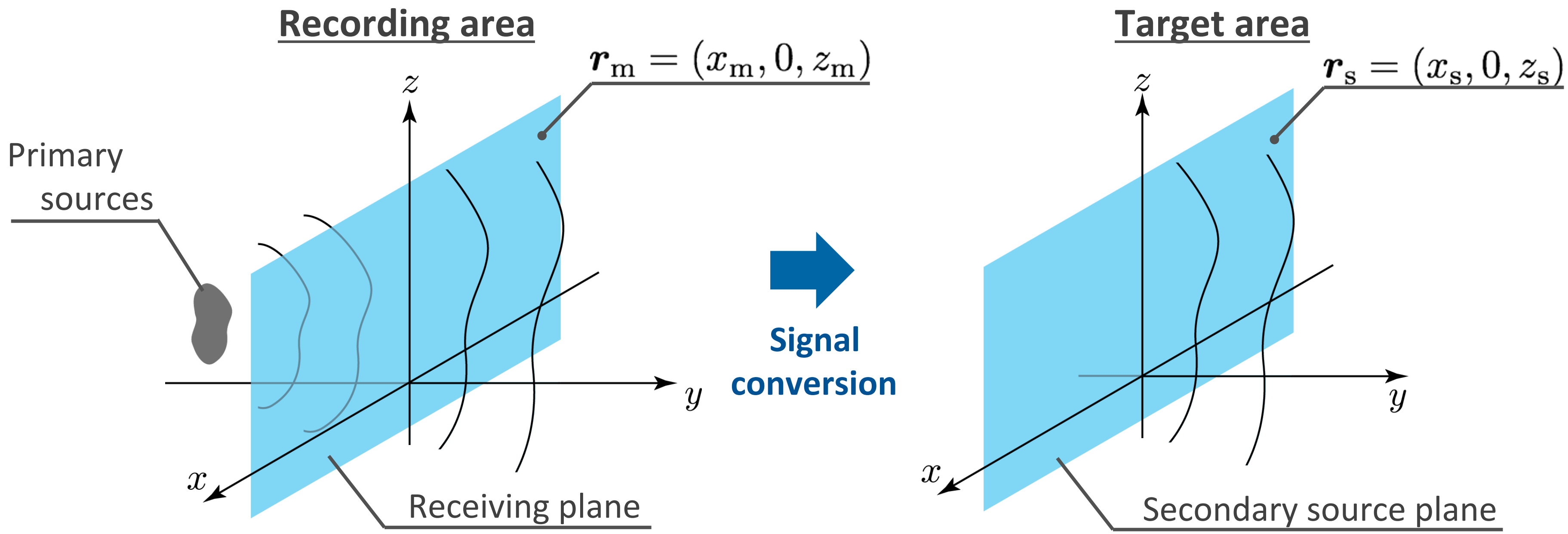
$$p(\boldsymbol{r}, \omega) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z P_{\mathrm{rcv}}(k_x, 0, k_z, \omega) e^{j(k_x x + k_y y + k_z z)}$$

Spatial frequency spectrum on receiving plane

$\left(k_y = \pm\sqrt{k^2 - k_x^2 - k_z^2} \right)$

Signal Conversion in Spatial Freq Domain

[Koyama+ IEEE TASLP 2013]



Desired sound field based on Rayleigh integral

$$p_{\rm des}(\boldsymbol{r},\omega) = -2\iint_{-\infty}^{\infty} \frac{\partial p(\boldsymbol{r}_{\rm m},\omega)}{\partial y_{\rm m}} G(\boldsymbol{r}|\boldsymbol{r}_{\rm m},\omega) dx_{\rm m} dz_{\rm m}$$

Green's function

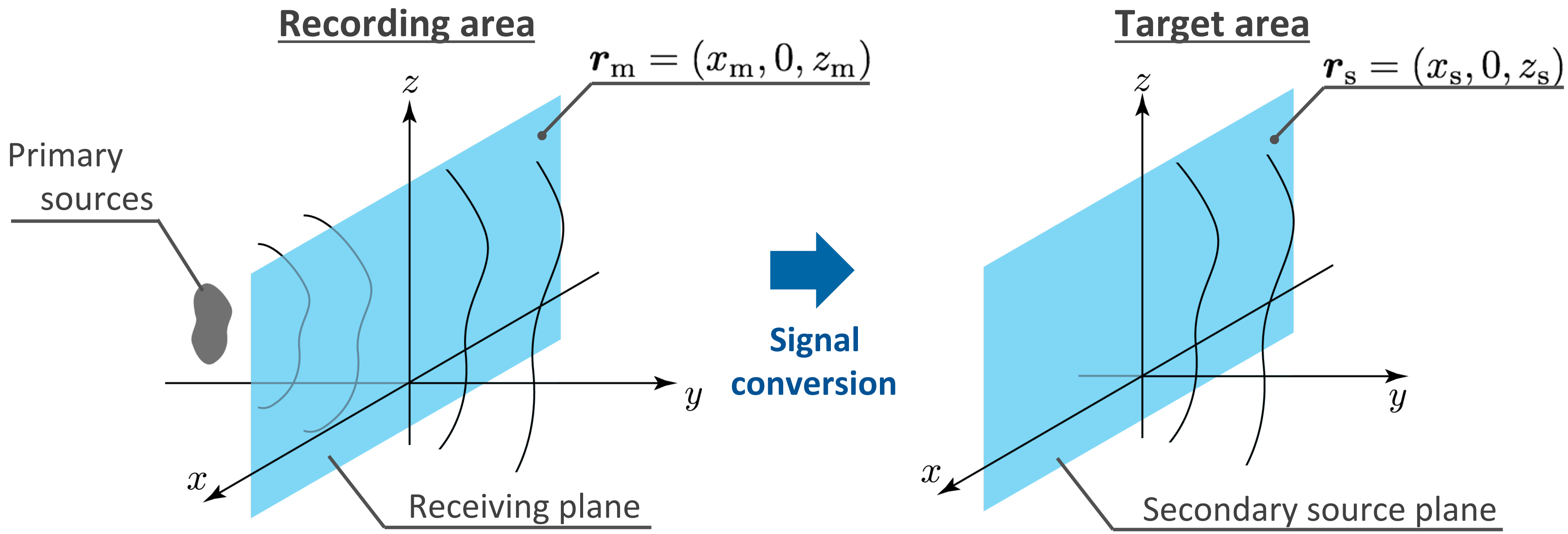
Synthesized sound field by secondary sources

$$p_{\rm syn}(\boldsymbol{r},\omega) = \iint_{-\infty}^{\infty} d(\boldsymbol{r}_{\rm s},\omega) H(\boldsymbol{r}-\boldsymbol{r}_{\rm s},\omega) dx_{\rm s} dz_{\rm s}$$

Transfer function

Signal Conversion in Spatial Freq Domain

[Koyama+ IEEE TASLP 2013]



Desired sound field based on Rayleigh integral in spatial freq domain

$$P_{\mathrm{des}}(k_x, y, k_z, \omega) = -2jk_y P_{\mathrm{rcv}}(k_x, 0, k_z, \omega) G(k_x, y, k_z, \omega)$$

Green's function

Synthesized sound field by secondary sources in spatial freq domain

$$P_{\mathrm{syn}}(k_x, y, k_z, \omega) = D(k_x, k_z, \omega) H(k_x, y, k_z, \omega)$$

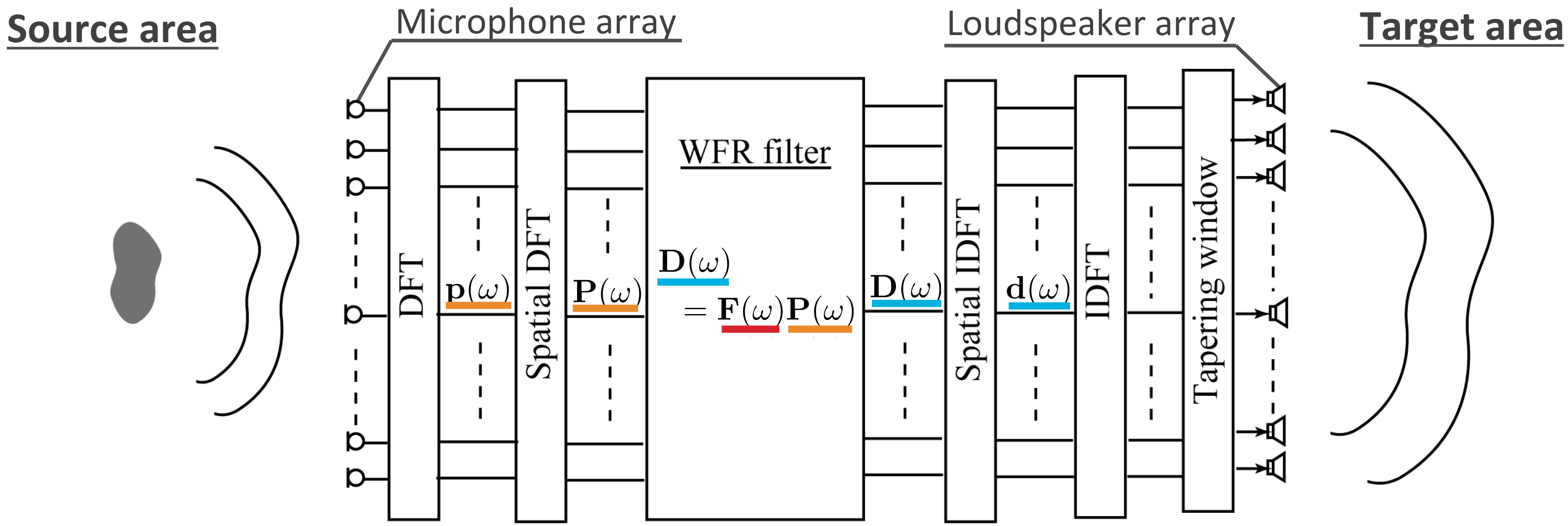
Transfer function

$$D(k_x, k_z, \omega) = -2jk_y P_{\mathrm{rcv}}(k_x, 0, k_z, \omega)$$

$d(\boldsymbol{r}_{\mathrm{s}}, \omega)$ **can be obtained only from** $p_{\mathrm{rcv}}(\boldsymbol{r}_{\mathrm{m}}, \omega)$ **!**

Wave Field Reconstruction (WFR) Filtering

Wave Field Reconstruction (WFR) Filter: Driving signal can be obtained by applying spatio-temporal filter [Koyama+ IEEE TASLP2013]



- Linear arrays of microphones and loudspeakers: 2D Convolution of WFR filter $\mathbf{F}(\omega)$

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

Driving signals in spatial freq domain

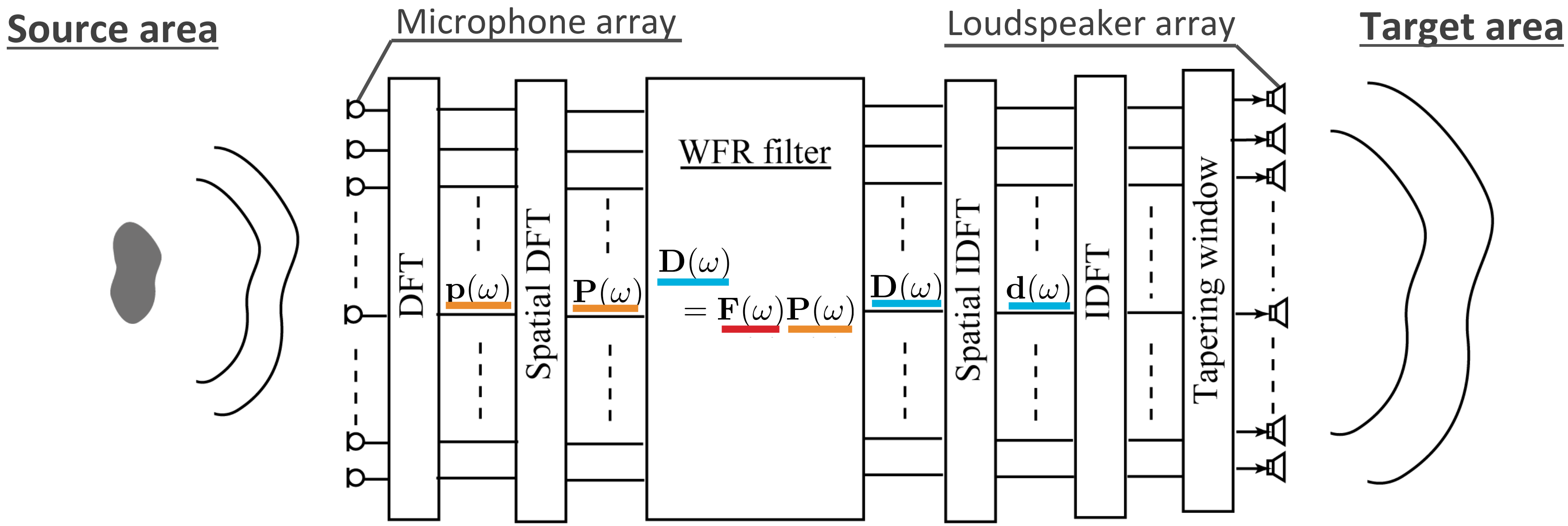
Diagonal matrix consisting of

$$F_i(\omega) = -4j \frac{\exp\left(\sqrt{k^2 - k_{x.i}^2}\, y_{\mathrm{ref}}\right)}{H_0^{(1)}\left(\sqrt{k^2 - k_{x.i}^2}\, y_{\mathrm{ref}}\right)}$$

Received signals in spatial freq domain

Wave Field Reconstruction (WFR) Filtering

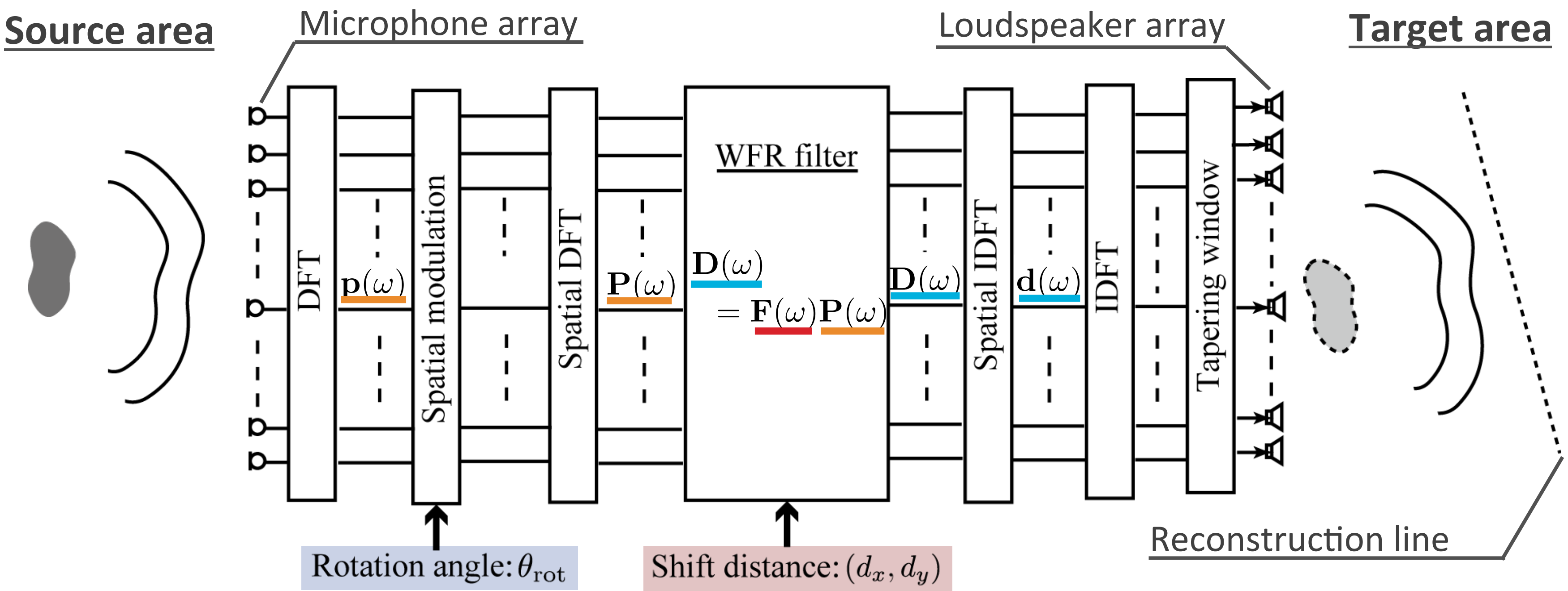
Wave Field Reconstruction (WFR) Filter: Driving signal can be obtained by convolution of spatio-temporal filter [Koyama+ IEEE TASLP2013]



- Linear arrays of microphones and loudspeakers: 2D Convolution of WFR filter $\mathbf{F}(\omega)$
  - Filter for signal conversion is stable compared to that designed by numerical approach (e.g., Least squares)
  - Fast computation of signal conversion can be achieve by using FFT algorithm

Wave Field Reconstruction (WFR) Filtering

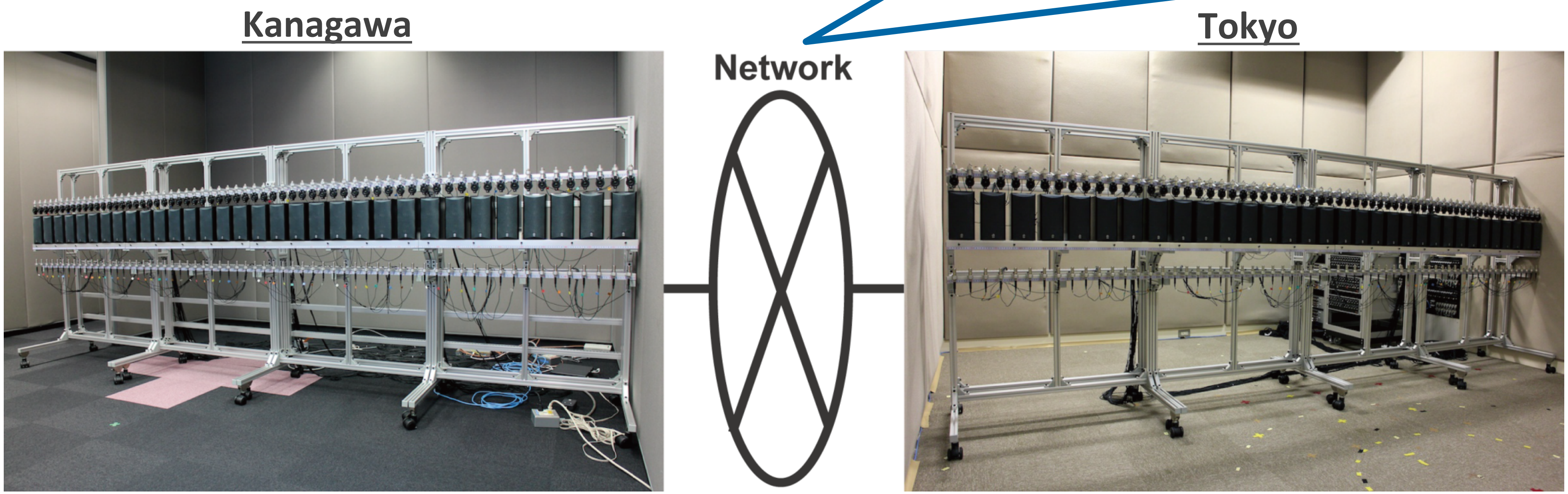
Position of reproduced sound field can be controlled by spatial phase shift and modulation



- ✧ **Reconstruction line can be shifted by spatial phase shift and rotated by spatial modulation**
- ✧ **It is possible to recreate virtual sound sources in front of loudspeaker array**

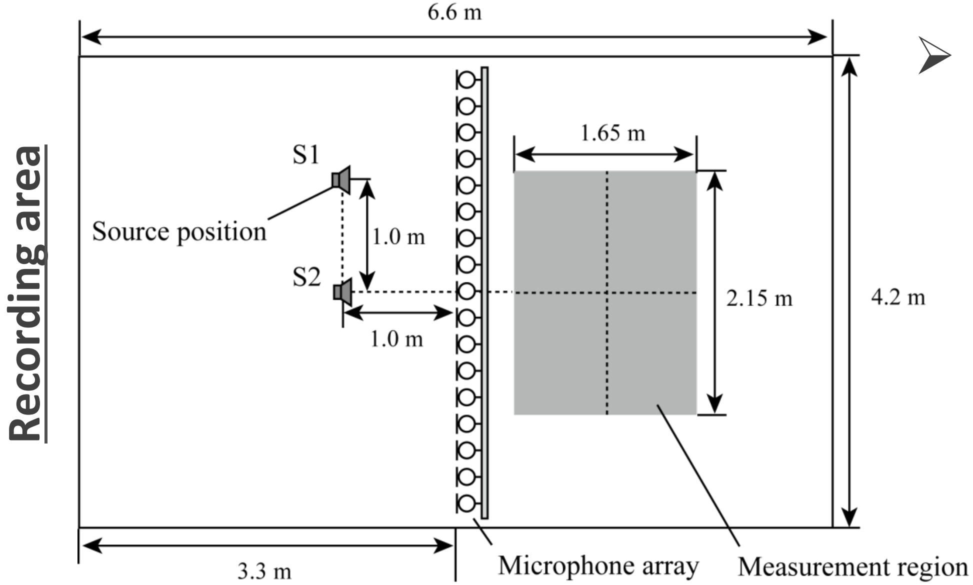
Practical Implementation

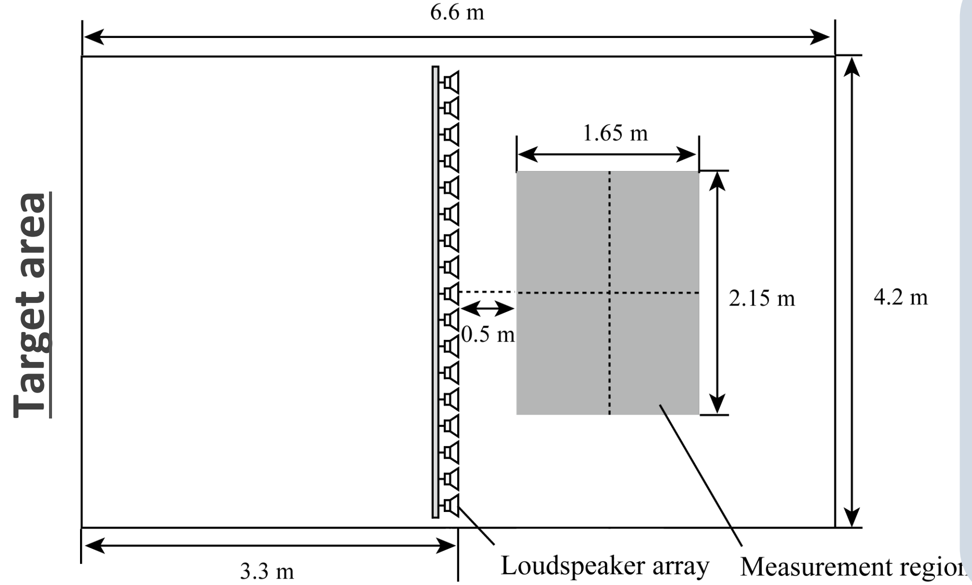
Real-time sound field transmission system by NTT [Koyama+ IEICE Trans 2014]



- Loudspeakers (for high freq.): 64, 6cm intervals
- Loudspeakers (for low freq.): 32, 12cm intervals
- Microphones: 64, 6cm intervals
- Array size: 3.84 m
- Sampling freq.: 48 kHz, Delay: 152 ms

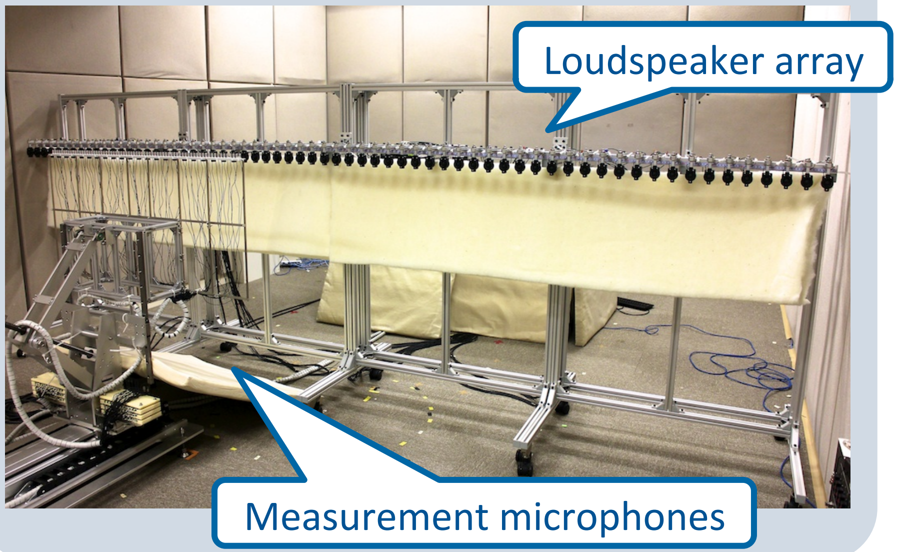
Measurement experiments





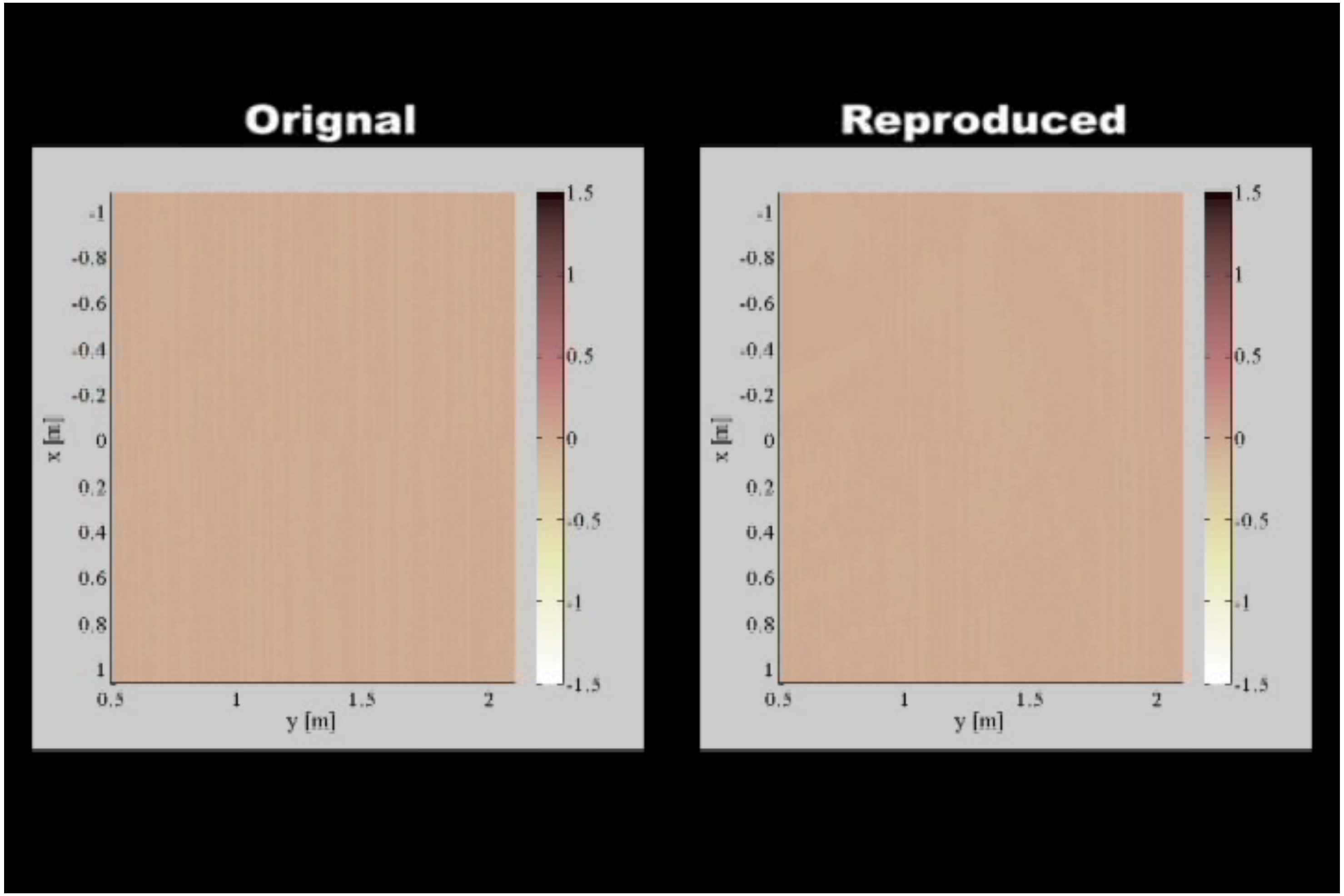
- **Visualize reproduced sound field**
  - Impulse responses were measured on planar region of 2.15 x 1.65 m at 1.5 cm intervals (144x108 points)
  - Reverberation time (T_{60}): 167ms
  - Source: Loudspeaker
  - Array of smaller loudspeakers was only used

Impulse response measurement



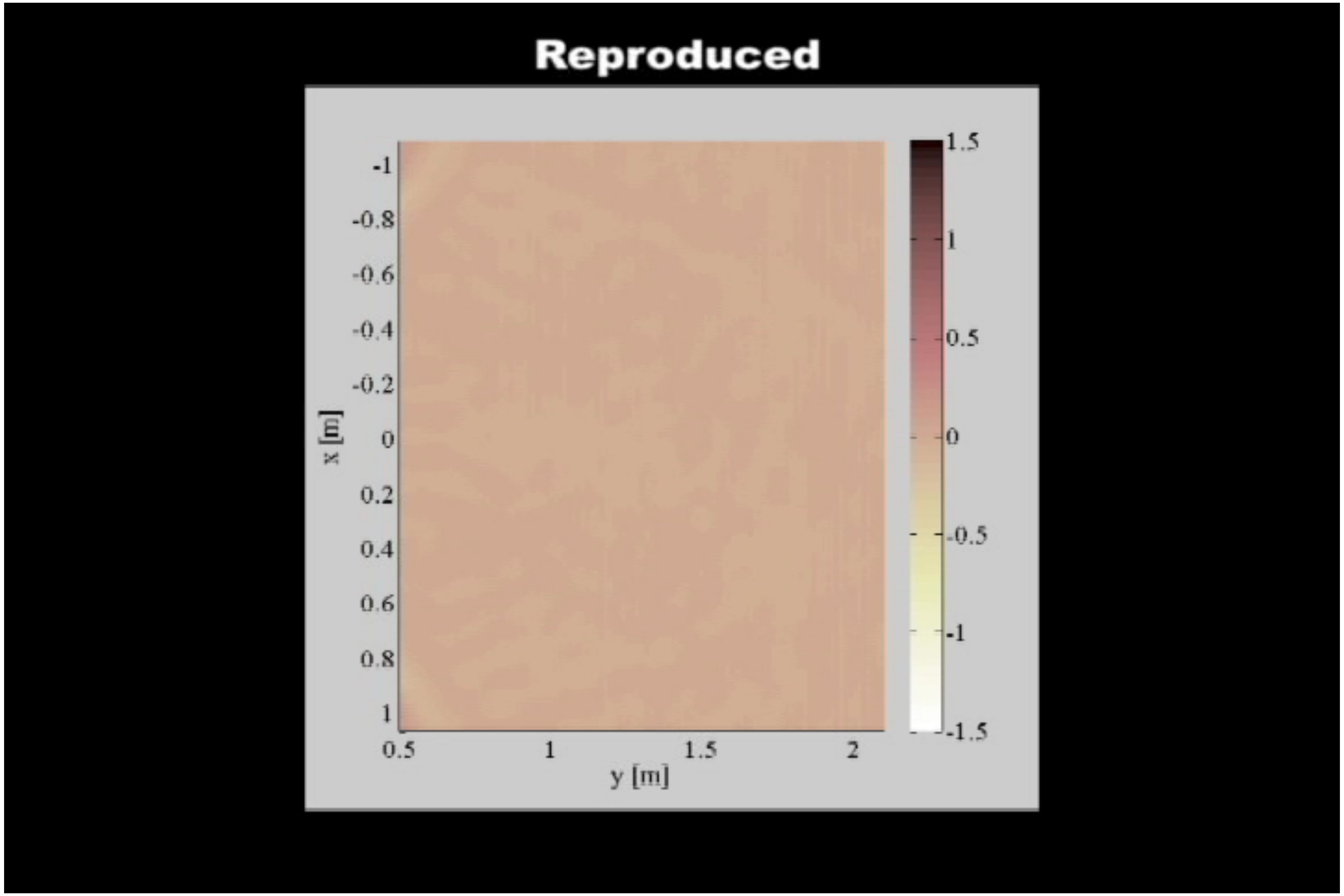
Measurement of Reproduced Sound Field

– Source signal: Low-passed pulse (0 – 2.6kHz)
– Source: Loudspeaker, Position: (-1.0, -1.0, 0.0) m

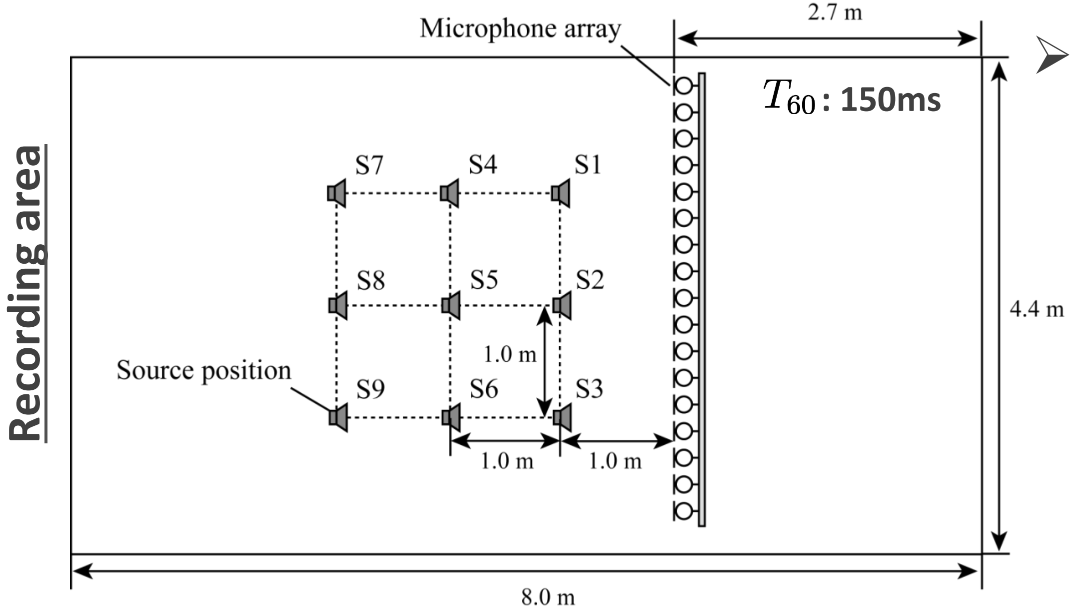


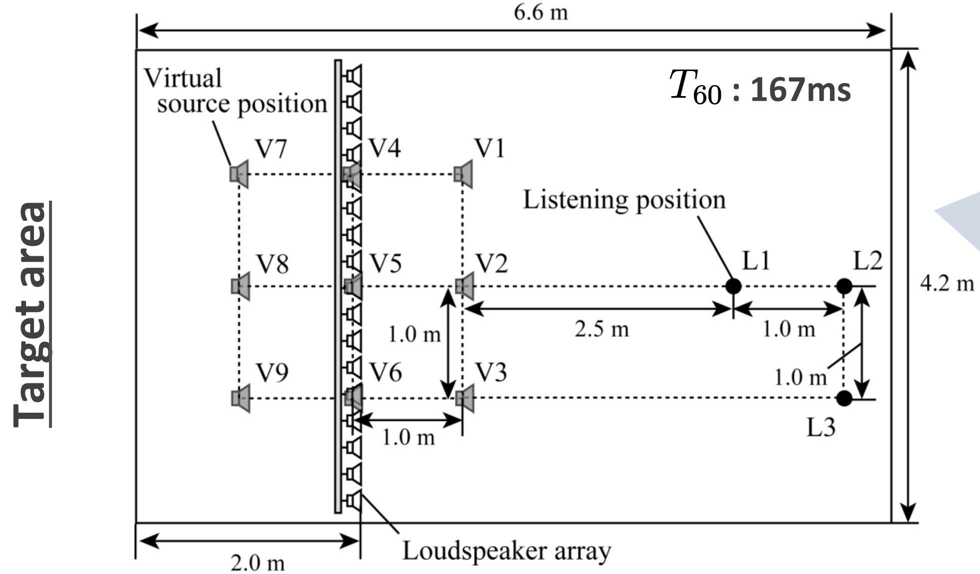
Measurement of Reproduced Sound Field

– Source signal: Low-passed pulse (0 – 2.6kHz)
– Source: Loudspeaker, Position: (-1.0, -1.0, 0.0) m, 2.0 m forward shift



Sound Localization Listening Test

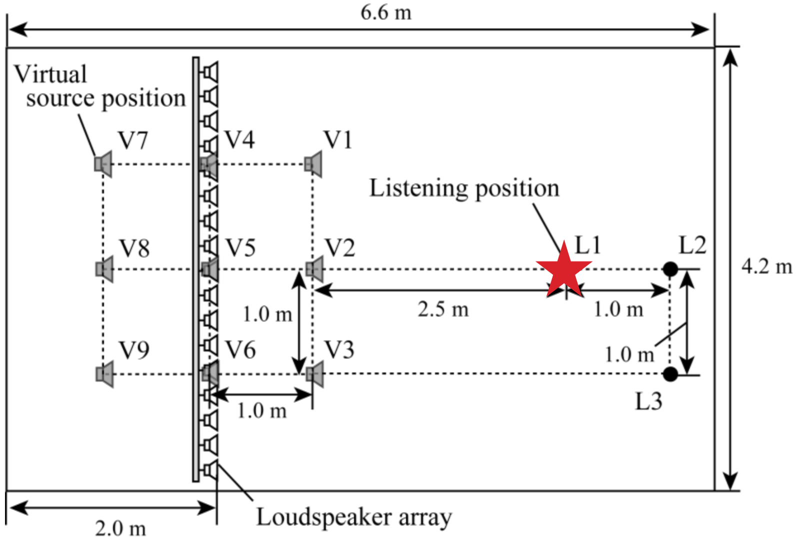




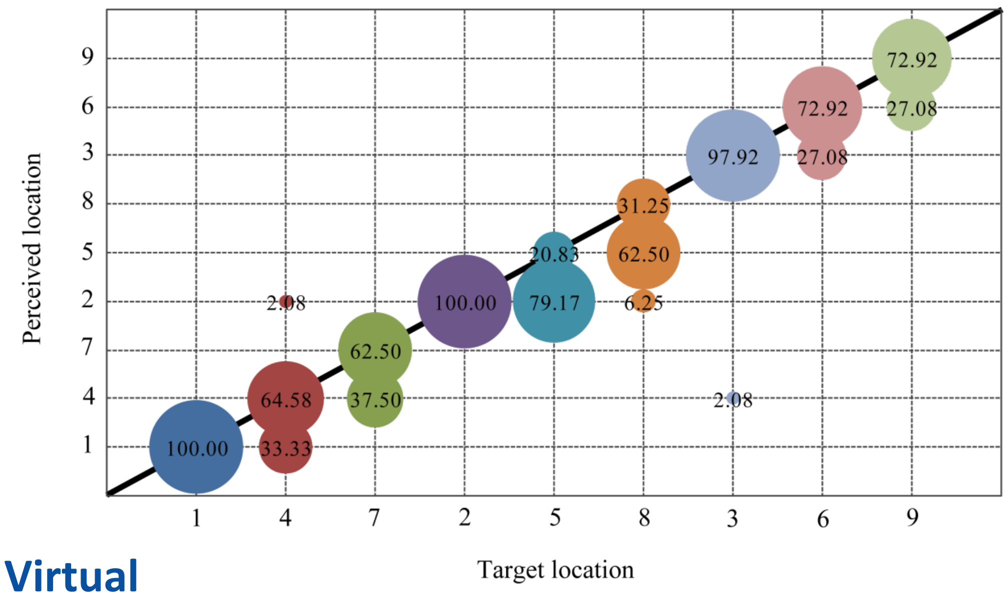
- **Sound localization test**
  - 6 listeners answered sound location from 9 positions
  - Compare localization accuracy between real and virtual sound sources
  - Source signal: Speech (4 sec)
  - Each sound source position was randomly selected at 144 times in total

Sound localization listening test

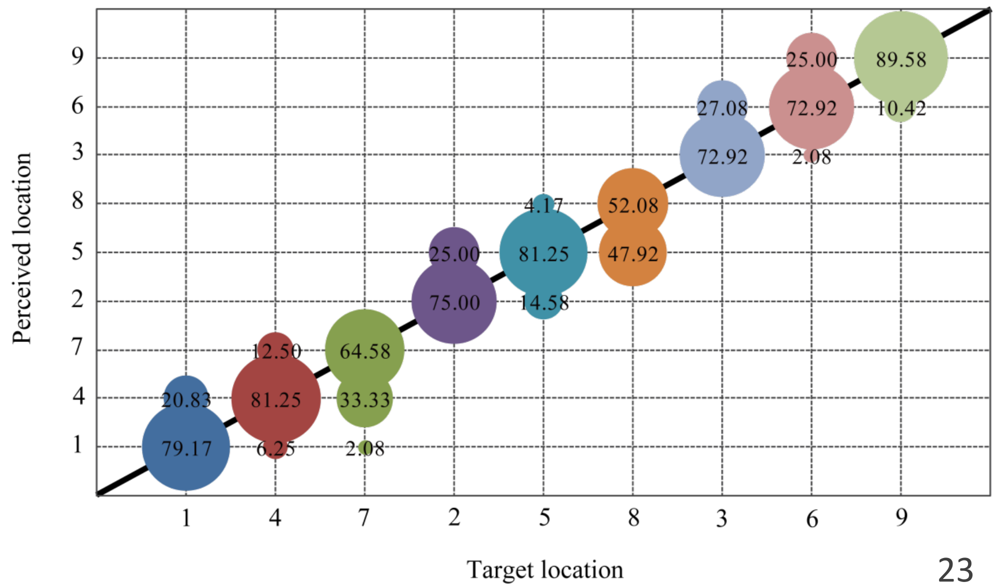
Listening Test



Real



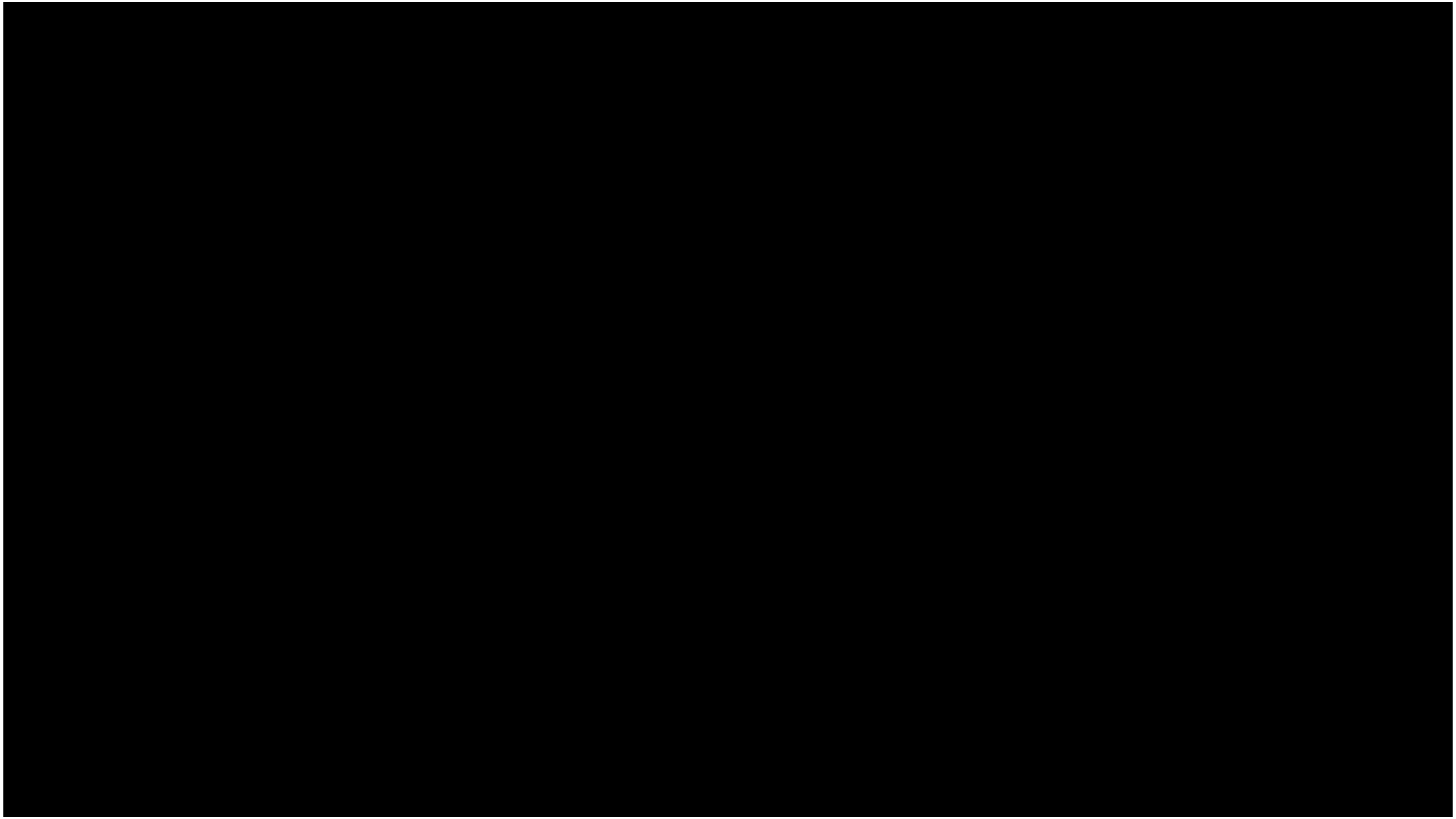
Virtual



- **Result of perceived location**
  - Almost the same localization accuracy
  - Accuracy rate of virtual sources at V1 - V3 was slightly low
  - Low accuracy rate of real sound sources in line with front-facing position (V5 & V8)

Demonstration System with 3D image

DigInfo TV: http://www.diginfo.tv/v/13-0016-r-en.php



Signal Conversion for Recording and Reproduction

- ✧ Fast and stable signal conversion for sound field recording and reproduction
- ✧ WFR filter was analytically derived in spatio-temporal freq domain
- ✧ Signal conversion by spatio-temporal convolution of WFR filter
- ✧ Real-time sound field transmission system and its evaluation

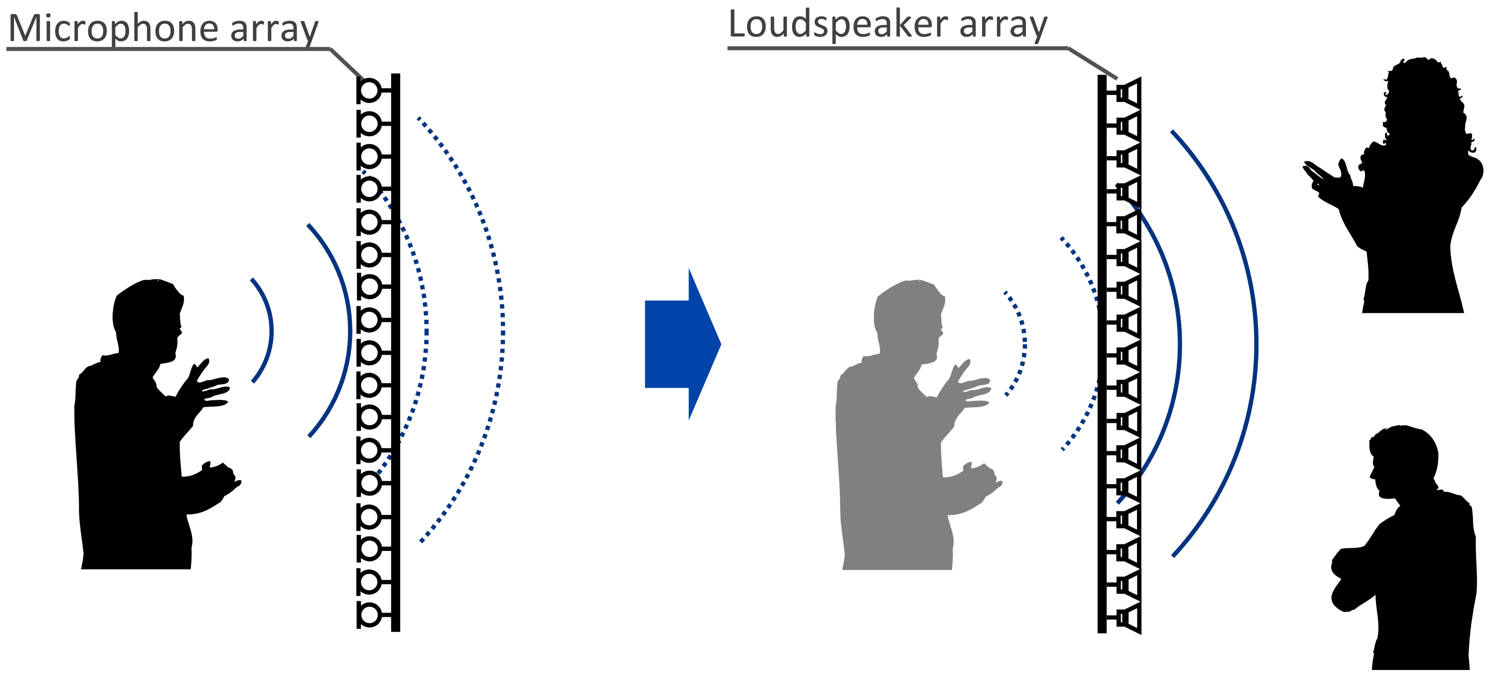
Related publications

- S. Koyama, *et al*. "Wave field reconstruction filtering in cylindrical harmonic domain for with-height recording and reproduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1546-1557, 2014.
- S. Koyama, *et al*. "Real-time sound field transmission system by using wave field reconstruction filter and its evaluation," *IEICE Trans. Fund. Electron. Comm. Comput. Sci.*, vol. E97-A, no. 9, pp. 1840-1848, 2014.
- S. Koyama, *et al*. "Analytical Approach to Wave Field Reconstruction Filtering in Spatio-temporal Frequency Domain," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 4, 2013.
- S. Koyama, *et al*. "Reproducing Virtual Sound Sources in Front of a Loudspeaker Array Using Inverse Wave Propagator," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 6, 2012.

SUPER-RESOLUTION IN SOUND FIELD RECORDING AND REPRODUCTION

Sound field reproduction for audio system

Super-resolution in sound field recording and reproduction

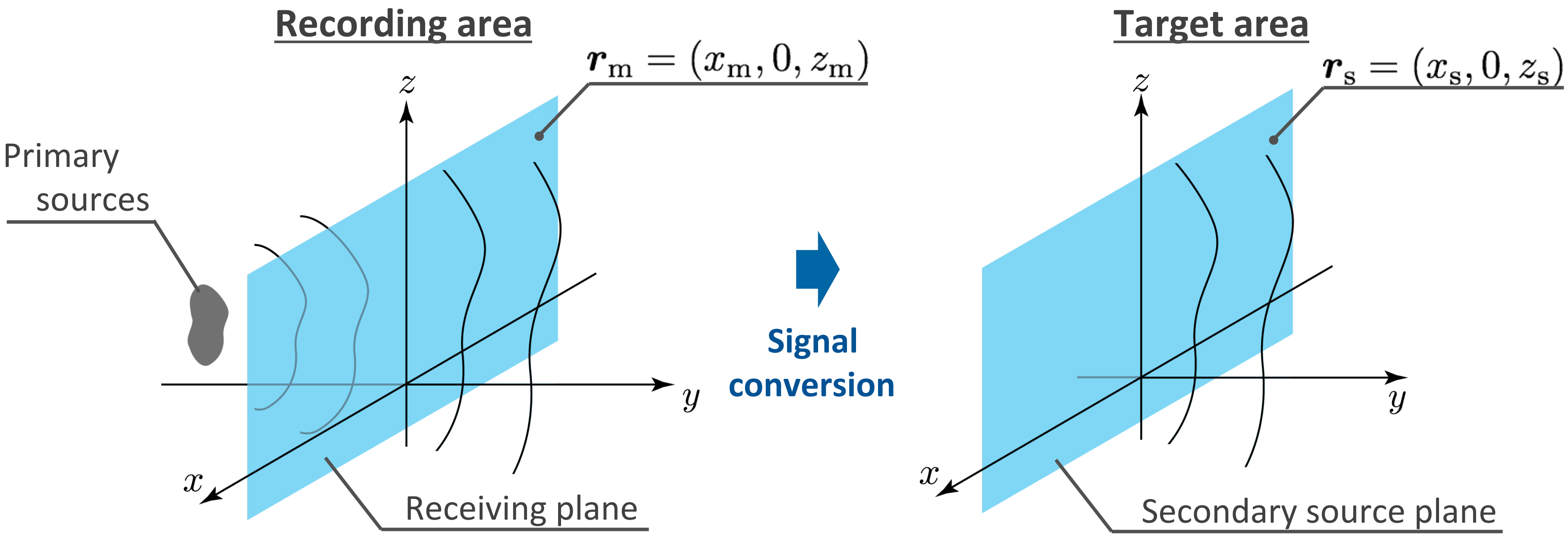


Improve reproduction accuracy when less microphones than loudspeakers

- # of microphones > # of loudspeakers [Ahrens+ AES Conv. 2010]
  - Higher reproduction accuracy within local region of target area
- # of microphones < # of loudspeakers [Koyama+ WASPAA 2013], [Koyama+ ICASSP 2014]
  - Higher reproduction accuracy of sources in local region of recording area

Wave field reconstruction (WFR) filtering method

[Koyama+ IEEE TASLP 2013]



$$p(\boldsymbol{r}_{\mathrm{m}}, \omega) = \sum_n b_n(\omega) \phi_n(\boldsymbol{r}_{\mathrm{m}}, \omega) \quad \Rightarrow \quad d(\boldsymbol{r}_{\mathrm{s}}, \omega) = \sum_n c_n(\omega) \phi_n(\boldsymbol{r}_{\mathrm{s}}, \omega)$$

- $p(\boldsymbol{r}_{\mathrm{m}}, \omega)$: Received signals
- $d(\boldsymbol{r}_{\mathrm{s}}, \omega)$: Driving signals
- ϕ_n: Plane wave

Each plane wave determines entire sound field

Spatial aliasing artifacts due to plane wave decomposition ➡ Significant error at high freq. even when microphone < loudspeaker

Sound field decomposition for super-resolution

Plane wave decomposition

$$p(\boldsymbol{r},\omega) = \sum_n b_n(\omega)\phi_n(\boldsymbol{r},\omega)$$

Plane wave

- Computationally efficient
- Larger effect of spatial aliasing

Decompose into source signal, location, transfer function, etc…

$$p(\boldsymbol{r},\omega) = \sum_k s_k(\omega)h_k(\boldsymbol{r},\omega)$$

Transfer function

- Smaller effect of spatial aliasing
- Accurate estimation in real environment is very difficult

Intermediate representation of sound field is required for super-resolution in recording and reproduction

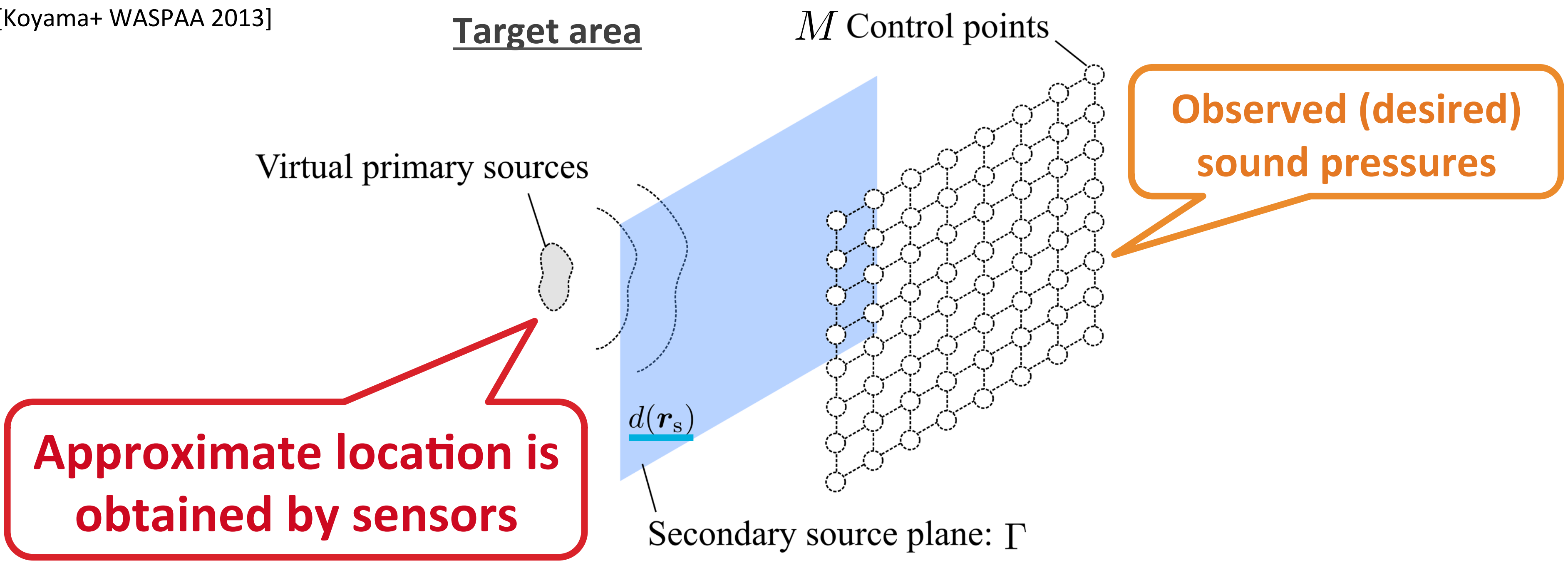
Two Approaches to Super-resolution

- **Source-Location-Informed Sound Field Recording and Reproduction** [Koyama+ WASPAA 2013]
  - Assume that approximate locations of sound sources in recording area are known
  - Driving signals of loudspeakers are obtained by using prior information on source locations

- **Sparse Sound Field Representation for Super-resolution in Recording and Reproduction** [Koyama+ ICASSP 2014]
  - Signals received by microphone array are decomposed into monopole source and plane wave components
  - Sparse signal decomposition algorithm is applied under the assumption of sparse distribution of monopole source components

Source-Location-Informed Recording and Reproduction

[Koyama+ WASPAA 2013]



Constraint on driving signals

$$\boldsymbol{p}_{\mathrm{des}} = \int_{\Gamma} d(\boldsymbol{r}_{\mathrm{s}}) \boldsymbol{G}(\boldsymbol{r}_{\mathrm{s}}) d\Gamma + \boldsymbol{\epsilon}$$

Desired pressures

Transfer function

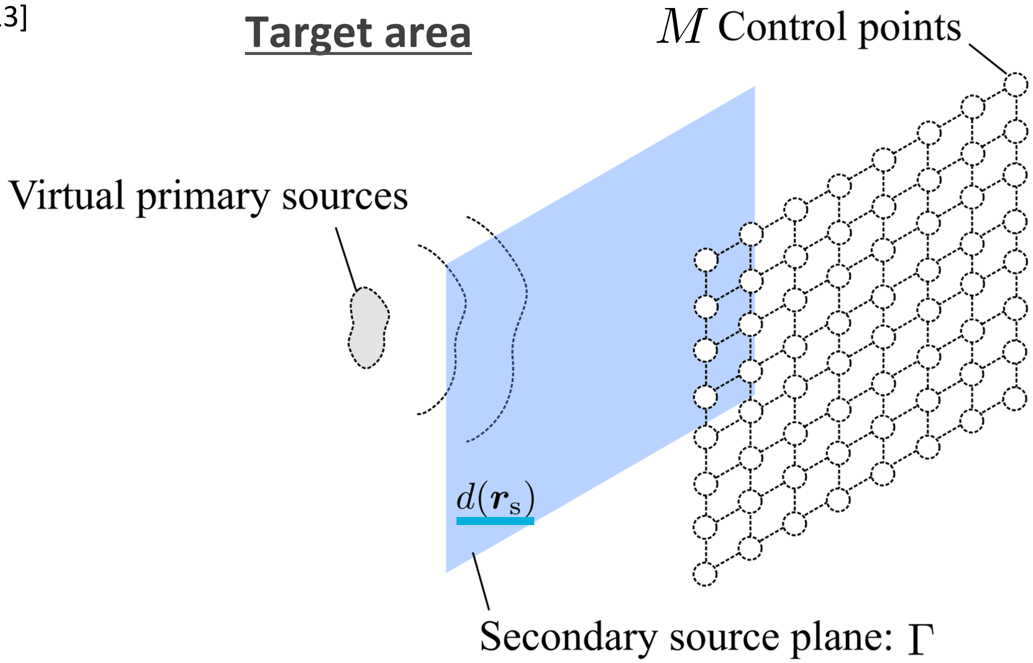
Linear combination of spatial basis functions

$$\hat{d}(\boldsymbol{r}_{\mathrm{s}}) = \sum_{n=1}^{N} c_n \phi_n(\boldsymbol{r}_{\mathrm{s}}) = \boldsymbol{\phi}^T(\boldsymbol{r}_{\mathrm{s}}) \boldsymbol{c}$$

Optimize $\boldsymbol{\phi}(\boldsymbol{r}_{\mathrm{s}})$ and $\boldsymbol{c}$ by using prior information on source locations

MAP Estimation of Driving Signals

[Koyama+ WASPAA 2013]



Maximum *a posteriori* (MAP) estimation

$$(\hat{\boldsymbol{c}}, \hat{\boldsymbol{\phi}}(\boldsymbol{r}_\mathrm{s})) = \arg \max_{\boldsymbol{c}, \boldsymbol{\phi}(\boldsymbol{r}_\mathrm{s})} \pi \left(d(\boldsymbol{r}_\mathrm{s}) = \boldsymbol{\phi}^T(\boldsymbol{r}_\mathrm{s})\boldsymbol{c} | \boldsymbol{p}_\mathrm{des} \right)$$

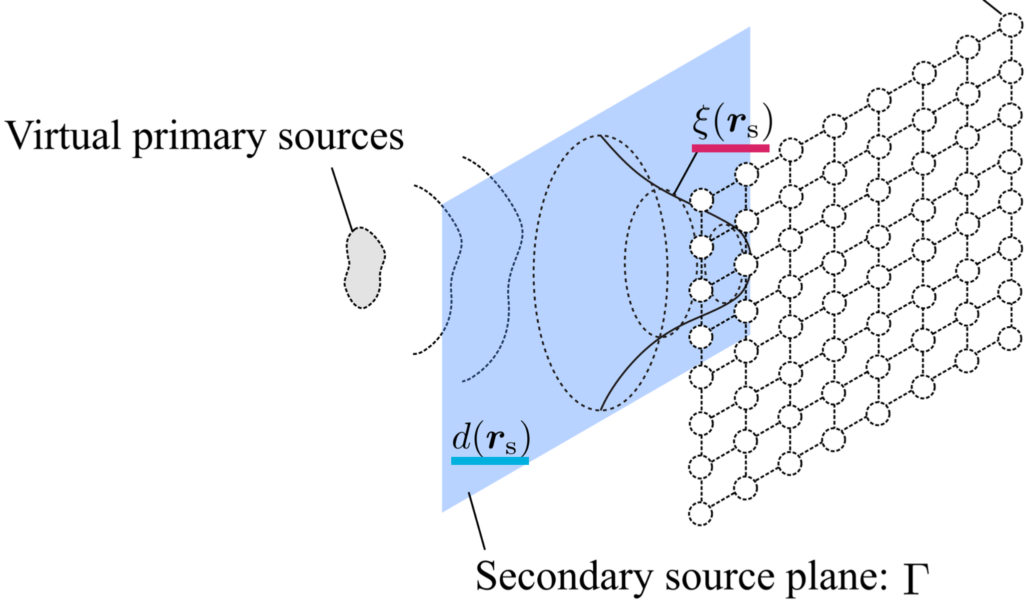
Bayes' rule ➡

$$\pi \left(d(\boldsymbol{r}_\mathrm{s}) | \boldsymbol{p}_\mathrm{des} \right) \propto \pi \left(\boldsymbol{p}_\mathrm{des} | d(\boldsymbol{r}_\mathrm{s}) \right) \pi \left(d(\boldsymbol{r}_\mathrm{s}) \right)$$

Likelihood function   Prior distribution

MAP Estimation of Driving Signals

[Koyama+ WASPAA 2013]



Maximum *a posteriori* (MAP) estimation

$$(\hat{\boldsymbol{c}}, \hat{\boldsymbol{\phi}}(\boldsymbol{r}_\mathrm{s})) = \arg \max_{\boldsymbol{c}, \boldsymbol{\phi}(\boldsymbol{r}_\mathrm{s})} \pi \left(d(\boldsymbol{r}_\mathrm{s}) = \boldsymbol{\phi}^T(\boldsymbol{r}_\mathrm{s})\boldsymbol{c} | \boldsymbol{p}_\mathrm{des} \right)$$

Bayes' rule

$$\pi \left(d(\boldsymbol{r}_\mathrm{s}) | \boldsymbol{p}_\mathrm{des} \right) \propto \pi \left(\boldsymbol{p}_\mathrm{des} | d(\boldsymbol{r}_\mathrm{s}) \right) \pi \left(d(\boldsymbol{r}_\mathrm{s}) \right)$$

Likelihood function   Prior distribution

Use amplitude distribution of $d(\boldsymbol{r}_\mathrm{s})$ obtained from prior source locations

MAP Estimation of Driving Signals

- **Assume spatial basis functions are M orthogonal functions, which satisfies the following relation of singular value decomposition**

$$\xi(\boldsymbol{r}_{\mathrm{s}})\boldsymbol{G}^{H}(\boldsymbol{r}_{\mathrm{s}}) = \boldsymbol{\phi}^{T}(\boldsymbol{r}_{\mathrm{s}})\boldsymbol{\Lambda}\boldsymbol{U}^{H}$$

- **Optimal spatial basis functions and their coefficients**

$$\hat{\boldsymbol{\phi}}(\boldsymbol{r}_{\mathrm{s}}) = \xi(\boldsymbol{r}_{\mathrm{s}})\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^{T}\boldsymbol{G}^{*}(\boldsymbol{r}_{\mathrm{s}})$$

$$\hat{\boldsymbol{c}} = \left(\alpha\boldsymbol{I} + \boldsymbol{\Lambda}^{2}\right)^{-1}\boldsymbol{\Lambda}\boldsymbol{U}^{H}\boldsymbol{p}_{\mathrm{des}}$$

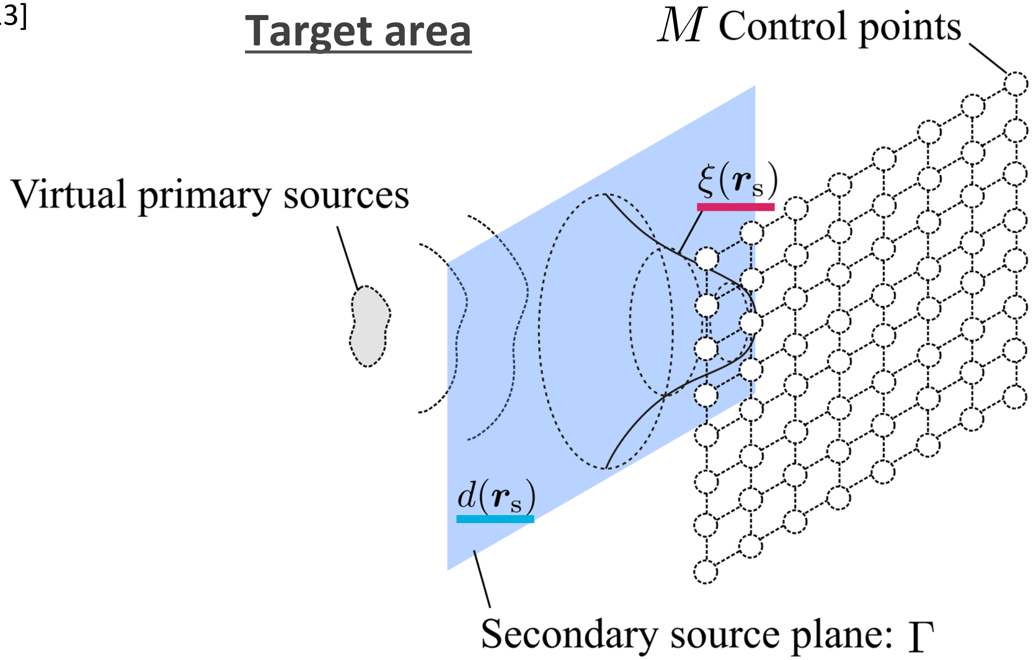
(α : regularization parameter)

- **Driving signals obtained by MAP estimation**

$$\hat{d}(\boldsymbol{r}_{\mathrm{s}}) = \xi(\boldsymbol{r}_{\mathrm{s}})\boldsymbol{G}^{H}(\boldsymbol{r}_{\mathrm{s}})\boldsymbol{U}(\alpha\boldsymbol{I} + \boldsymbol{\Lambda}^{2})^{-1}\boldsymbol{U}^{H}\boldsymbol{p}_{\mathrm{des}}$$

Predicted Amplitude Distribution of Driving Signals

[Koyama+ WASPAA 2013]



Assume spherical wave propagation from prior source locations

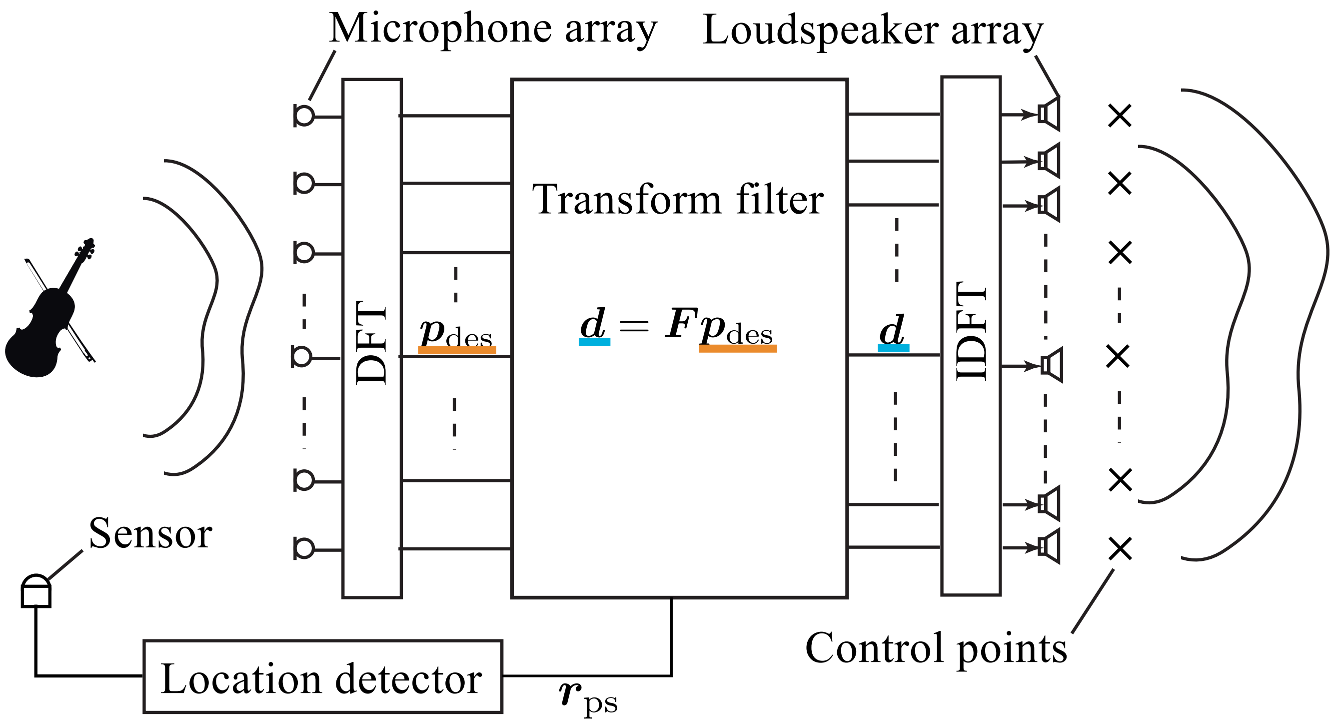
$$\bar{d}(\boldsymbol{r}_{\mathrm{s}}) = 2\frac{\partial p_{\mathrm{des}}(\boldsymbol{r}_{\mathrm{s}})}{\partial y_{\mathrm{s}}}$$

$$\simeq (y - y_{\mathrm{ps}})\frac{jk|\boldsymbol{r}_{\mathrm{s}} - \boldsymbol{r}_{\mathrm{ps}}| - 1}{2\pi|\boldsymbol{r}_{\mathrm{s}} - \boldsymbol{r}_{\mathrm{ps}}|}e^{jk|\boldsymbol{r}_{\mathrm{s}} - \boldsymbol{r}_{\mathrm{ps}}|}$$

Normalization

$$\xi(\boldsymbol{r}_{\mathrm{s}}) = |\bar{d}(\boldsymbol{r}_{\mathrm{s}})| \Big/ \int_{\boldsymbol{r}_{\mathrm{s}} \in \Gamma} |\bar{d}(\boldsymbol{r}_{\mathrm{s}})|^2 d\Gamma$$

Algorithm



Discretize secondary source distribution

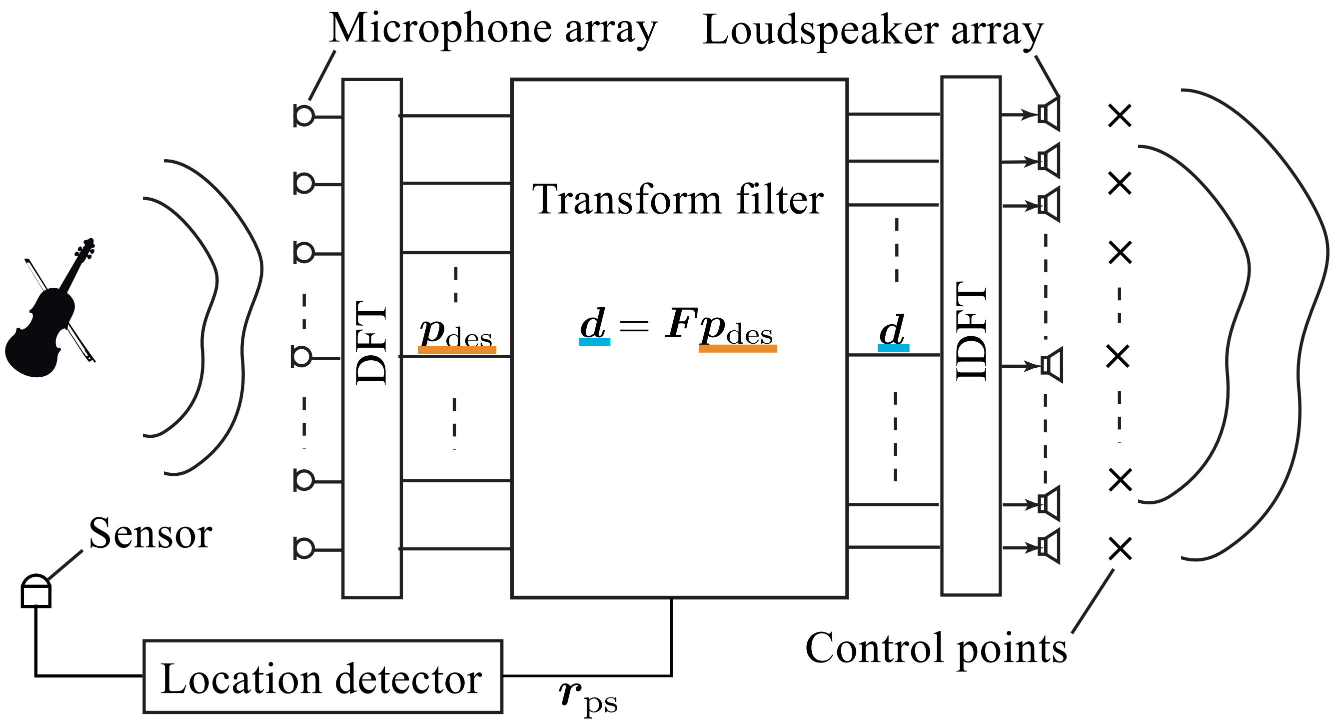
$$\boldsymbol{d} = [d(\boldsymbol{r}_{\text{s},1}), d(\boldsymbol{r}_{\text{s},2}) \cdots, d(\boldsymbol{r}_{\text{s},L})]^T$$

$$\boldsymbol{H} = [\boldsymbol{G}(\boldsymbol{r}_{\text{s},1}), \boldsymbol{G}(\boldsymbol{r}_{\text{s},2}) \cdots, \boldsymbol{G}(\boldsymbol{r}_{\text{s},L})]$$

Amplitude distribution for prior

$$\boldsymbol{\xi} = \text{diag}\left([\xi(\boldsymbol{r}_{\text{s},1}), \xi(\boldsymbol{r}_{\text{s},2}) \cdots, \xi(\boldsymbol{r}_{\text{s},L})]\right)$$

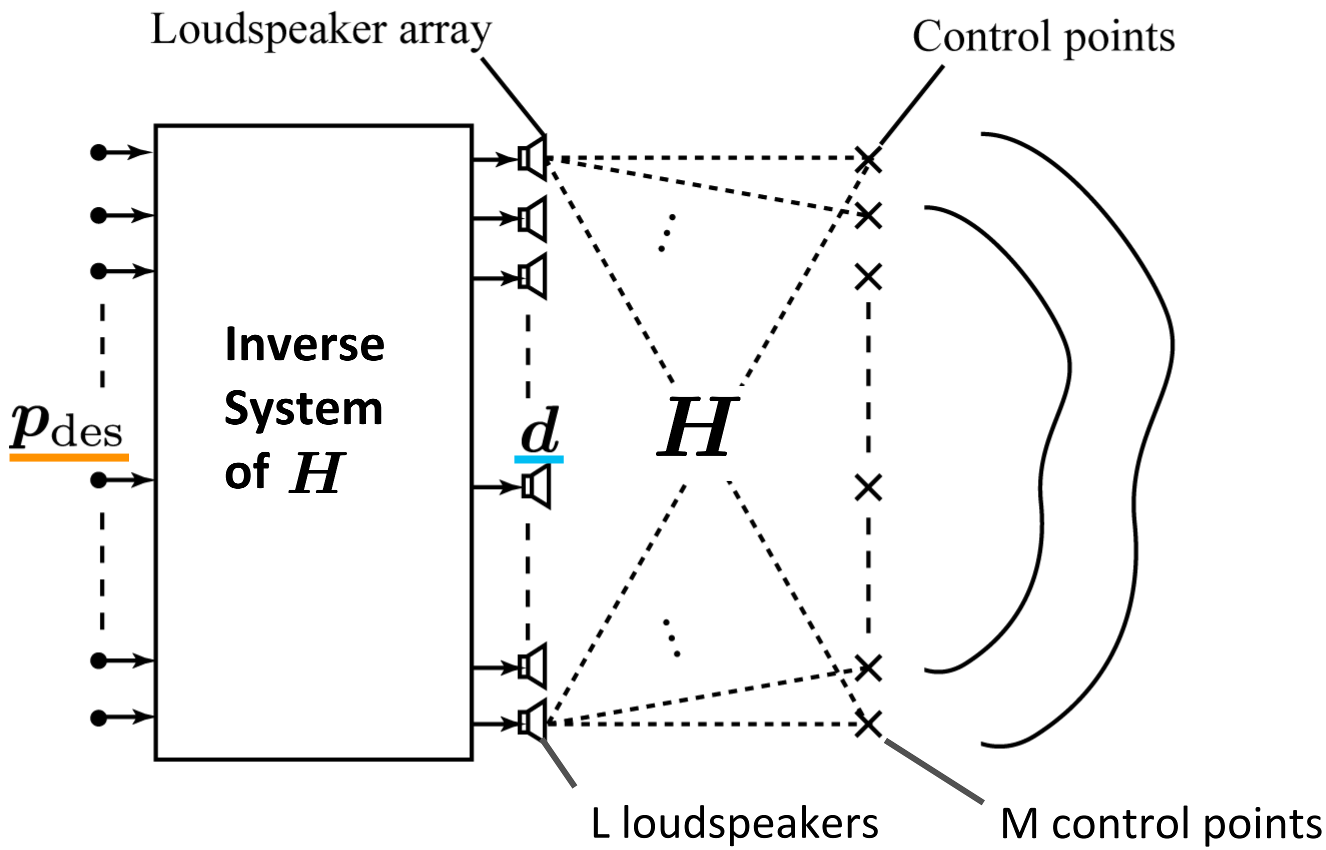
Algorithm



1. Detect source location ($\boldsymbol{r}_{\mathrm{ps}}$)
2. Calculate amplitude distribution $\boldsymbol{\xi}$
3. Calculate $\boldsymbol{R} \in \mathbb{C}^{M \times M}$ as $\boldsymbol{R} = \boldsymbol{H}\boldsymbol{\xi}\boldsymbol{H}^H$
4. Eigenvalue decomposition of $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}^2\boldsymbol{U}^H$
5. Obtain transform filter as $\boldsymbol{F} = \boldsymbol{\xi}\boldsymbol{H}^H\boldsymbol{U}\left(\alpha\boldsymbol{I} + \boldsymbol{\Lambda}^2\right)^{-1}\boldsymbol{U}^H$

Relation to Sound Pressure Control method

[Gautheir+ JASA 2005]



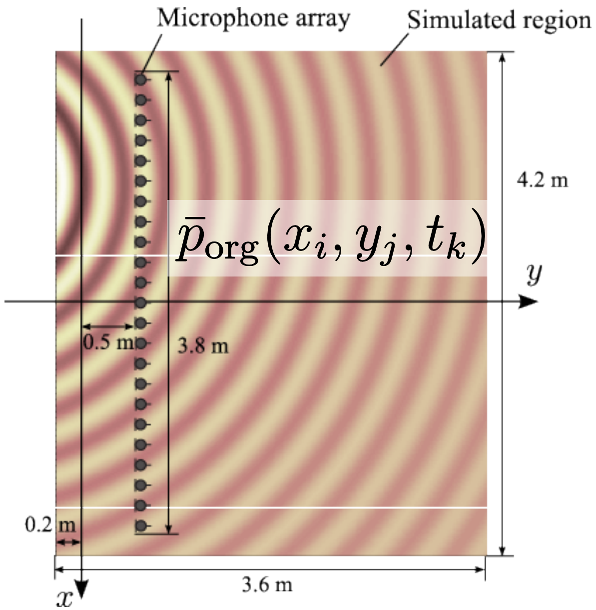
Inverse filter based on minimum-norm solution

$$\hat{\boldsymbol{d}} = \boldsymbol{H}^{H}(\boldsymbol{H}\boldsymbol{H}^{H} + \beta \boldsymbol{I})^{-1}\boldsymbol{p}_{\mathrm{des}}$$

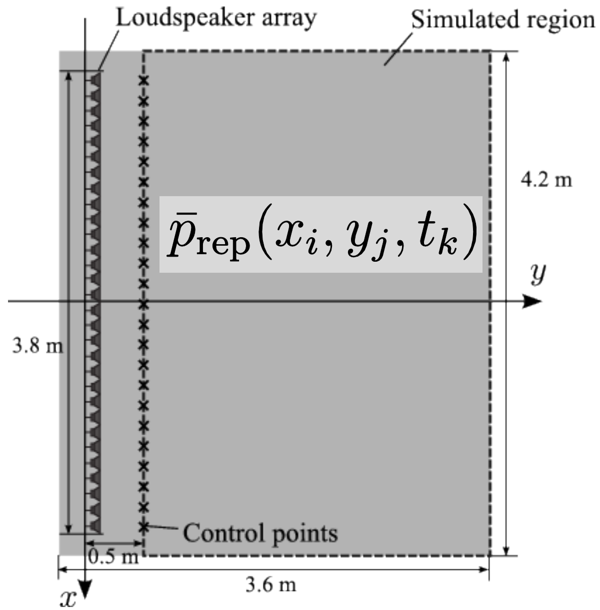
Correspond to the case that $\xi(\boldsymbol{r}_{\mathrm{s}})$ is uniform distribution!

Simulation Experiments

Original sound field



Reproduced sound field

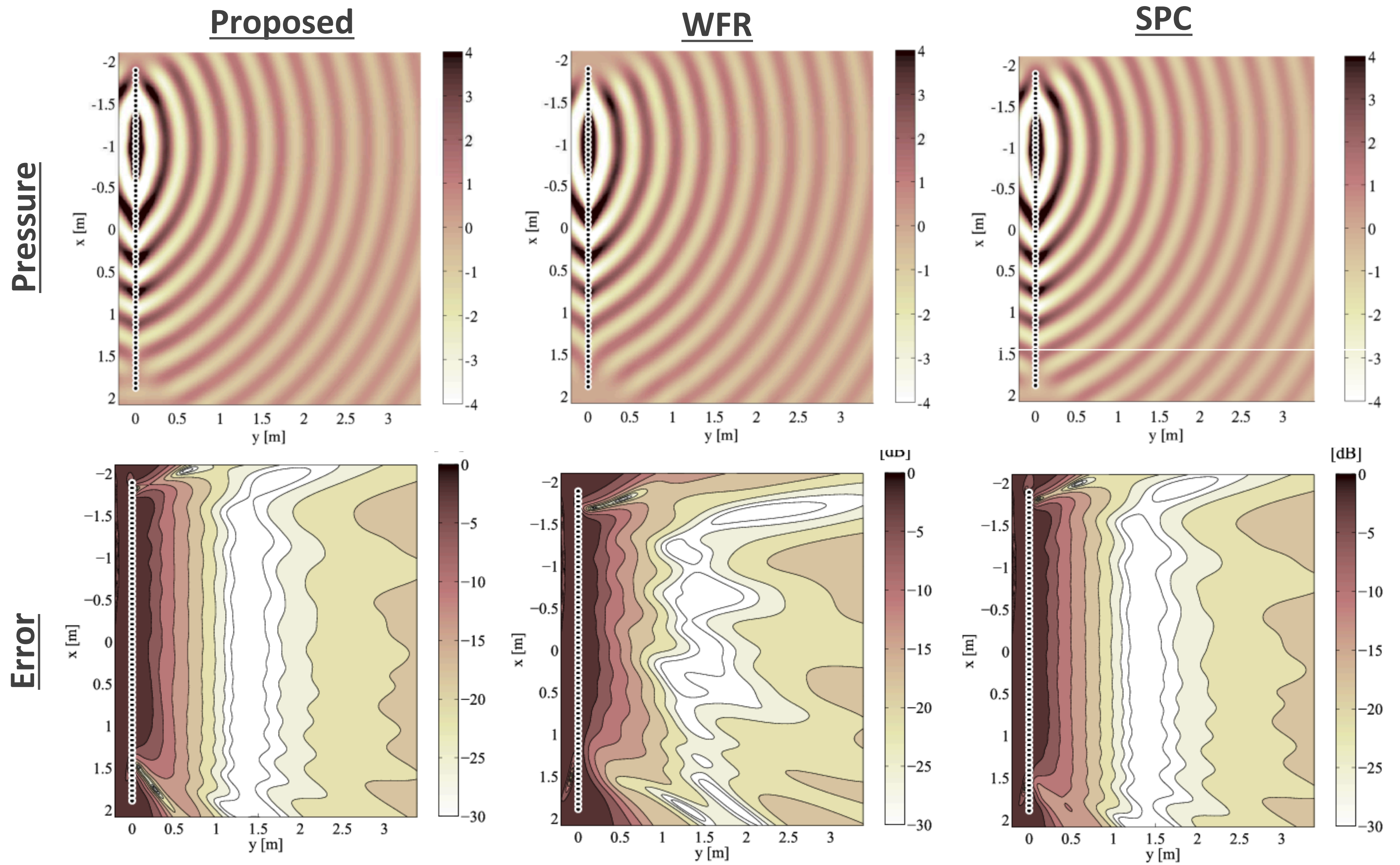


- Proposed method (**Proposed**), WFR filtering method (**WFR**), and sound pressure control method (**SPC**) methods were compared
- 32 microphones (12 cm intervals) and 64 loudspeakers (6cm intervals)
- Signal-to-distortion ratio (SDR) for evaluation

$$\mathrm{SDR} = 10\log_{10}\frac{\sum_i\sum_j\sum_k |\bar{p}_{\mathrm{org}}(x_i, y_j, t_k)|^2}{\sum_i\sum_j\sum_k |\bar{p}_{\mathrm{rep}}(x_i, y_j, t_k) - \bar{p}_{\mathrm{org}}(x_i, y_j, t_k)|^2}$$

Reproduced Sound Pressure Distribution (1.0 kHz)

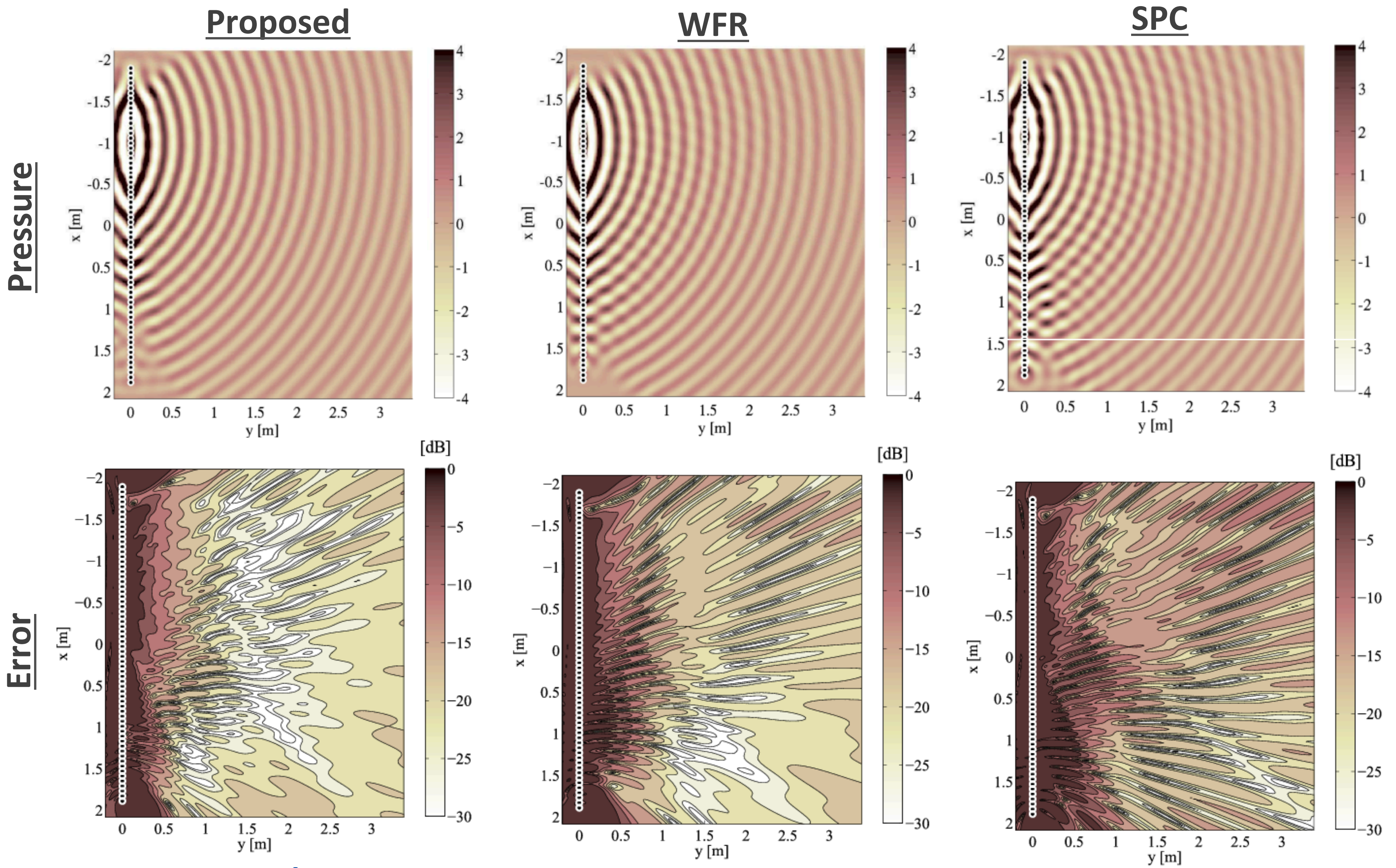
- Source location: (-1.0, -1.0, 0.0) m, Exact prior source location was given



17.9 dB   16.4 dB   17.8 dB

Reproduced Sound Pressure Distribution (1.7 kHz)

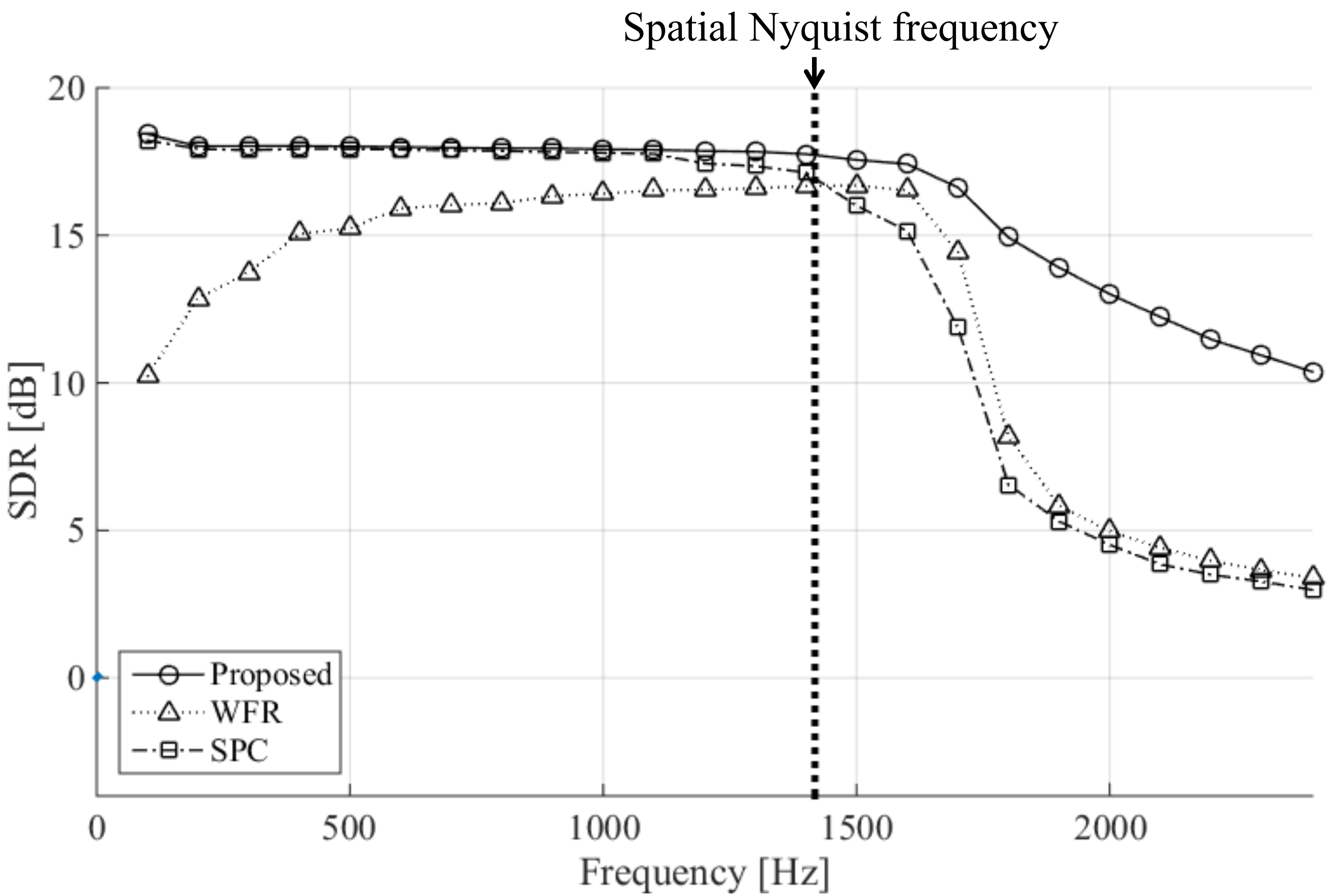
- Source location: (-1.0, -1.0, 0.0) m, Exact prior source location was given



16.6 dB | **14.5 dB** | **11.9 dB**

Frequency vs. SDR

- Source location: (-1.0, -1.0, 0.0) m, Exact prior source location was given



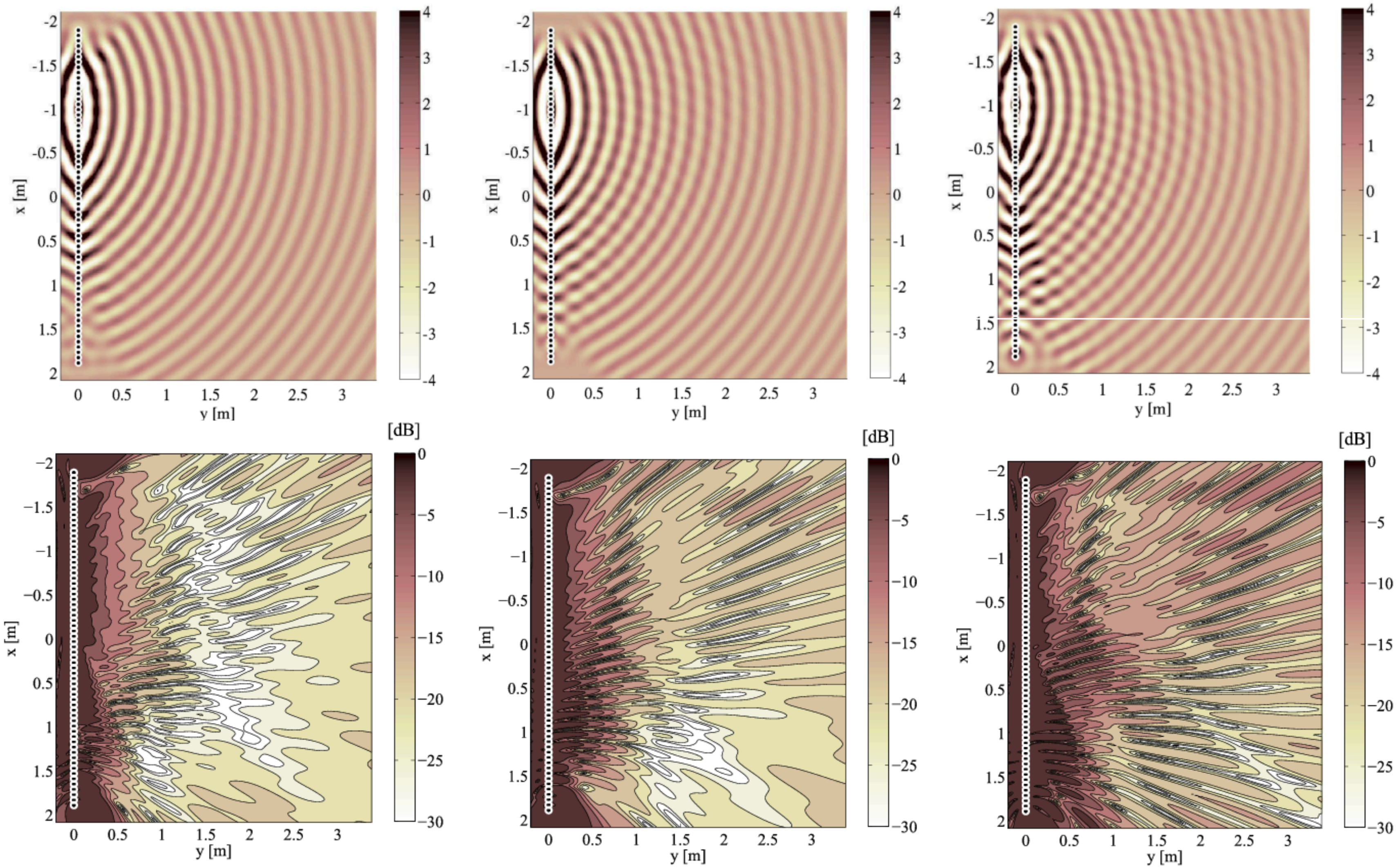
SDRs above spatial Nyquist frequency were improved

Reproduced Sound Pressure Distribution (1.7 kHz)

- Source location: (-1.0, -1.0, 0.0) m, **Prior source location: (-0.9, -1.2, 0.0) m**

Proposed | **WFR** | **SPC**

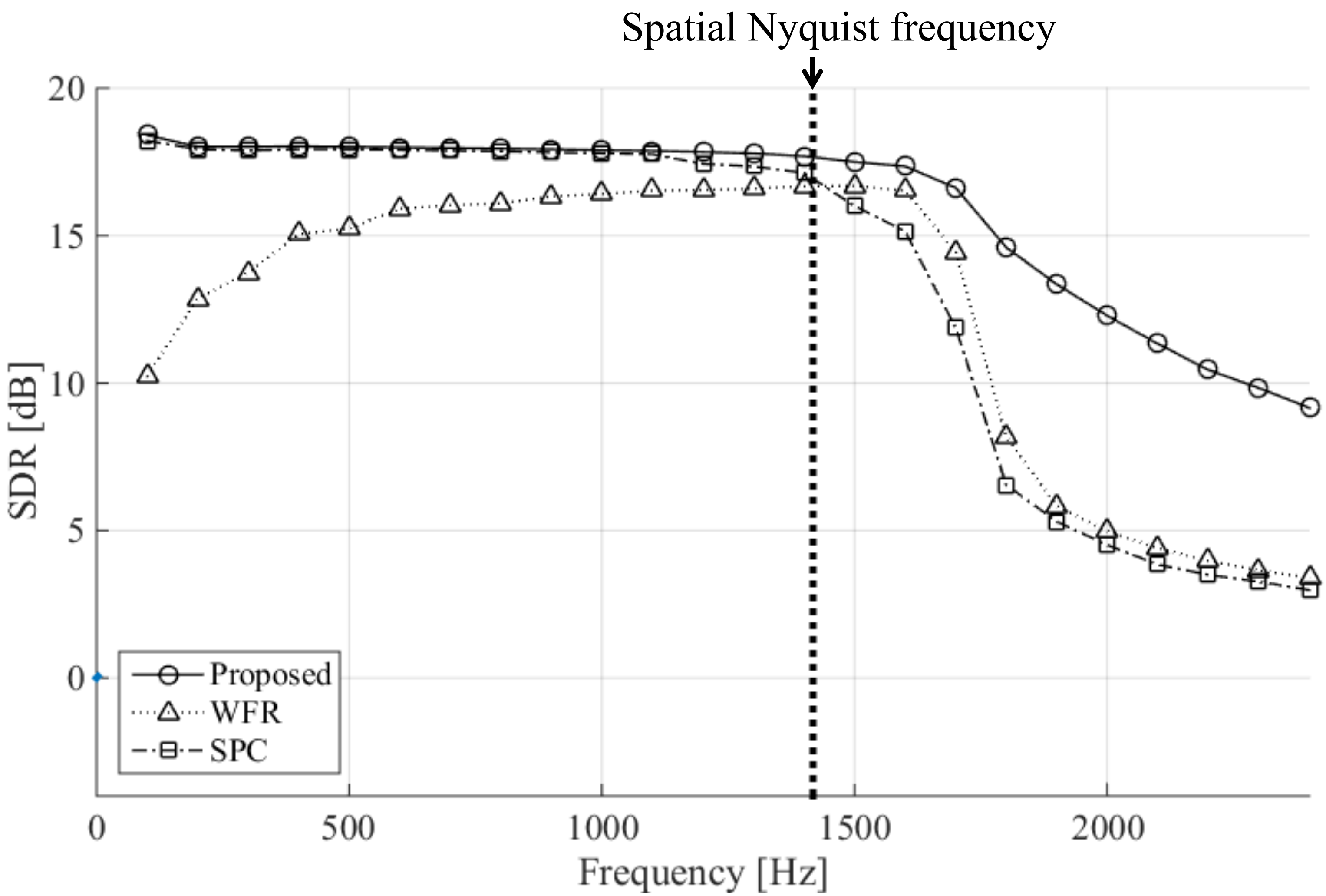
Pressure / **Error**



16.6 dB | **14.5 dB** | **11.9 dB**

Frequency vs. SDR

- Source location: (-1.0, -1.0, 0.0) m, **Prior source location: (-0.9, -1.2, 0.0) m**



Proposed method is robust against mismatch of prior source locations

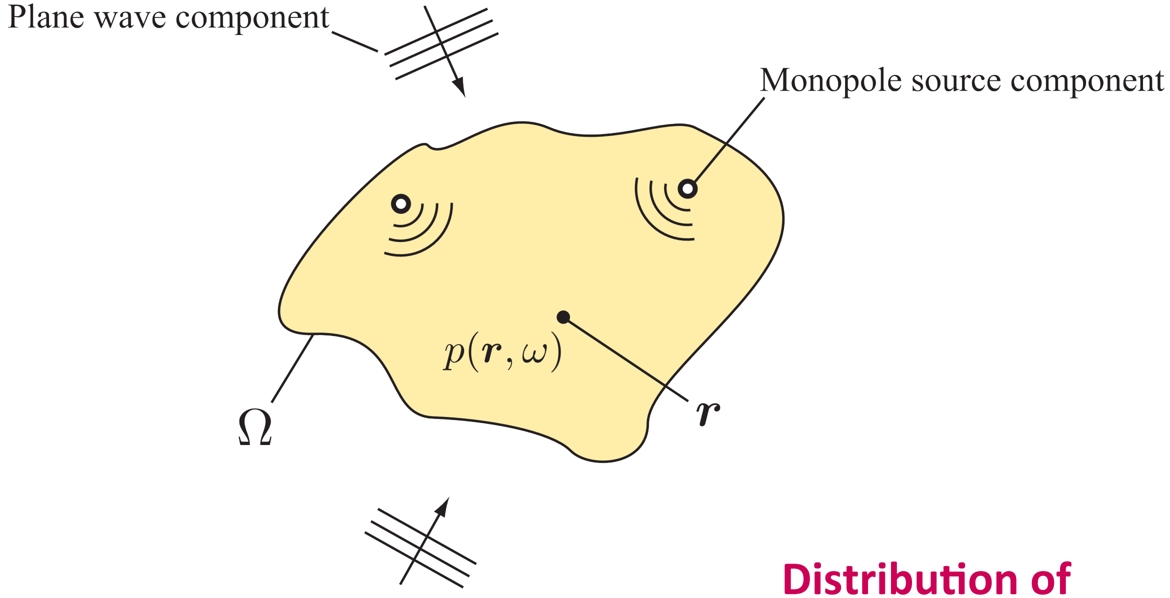
Two Approaches to Super-resolution

- **Source-Location-Informed Sound Field Recording and Reproduction** [Koyama+ WASPAA 2013]
  - Assume that approximate locations of sound sources in recording area are known
  - Driving signals of loudspeakers are obtained by using prior information of source locations

- **Sparse Sound Field Representation for Super-resolution in Recording and Reproduction** [Koyama+ ICASSP 2014]
  - Signals received by microphone array are decomposed into monopole source and plane wave components
  - Sparse signal decomposition algorithm is applied under the assumption of sparse distribution of monopole source components

Generative model of sound field

[Koyama+ ICASSP 2014]



- **Inhomogeneous and homogeneous Helmholtz eq.**

Distribution of monopole components

$$\left(\nabla^2 + k^2\right) p(\boldsymbol{r},\omega) = \begin{cases} -Q(\boldsymbol{r},\omega), & \boldsymbol{r} \in \Omega \\ 0, & \boldsymbol{r} \notin \Omega \end{cases}$$

Inhomogeneous + homogeneous terms

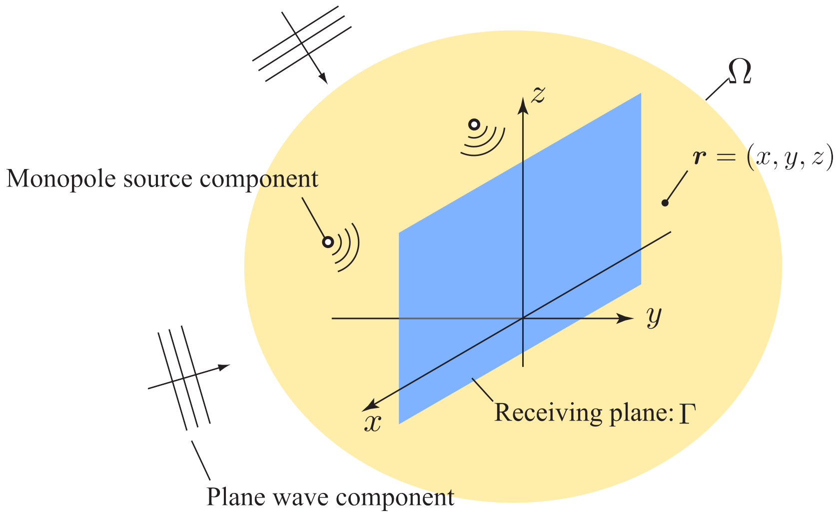
$$p(\boldsymbol{r},\omega) = p_{\mathrm{i}}(\boldsymbol{r}) + p_{\mathrm{h}}(\boldsymbol{r})$$

$$= \int_{\boldsymbol{r}' \in \Omega} Q(\boldsymbol{r}') G(\boldsymbol{r}|\boldsymbol{r}') d\boldsymbol{r}' + p_{\mathrm{h}}(\boldsymbol{r})$$

Green's function

Generative model of sound field

- **Observe sound pressure distribution on Γ**



- **Conversion into driving signals**

$$d(\boldsymbol{r}) = \frac{\partial}{\partial y} p(\boldsymbol{r}) \Big|_{y=0}$$

$$= \int_{\boldsymbol{r}' \in \Omega} Q(\boldsymbol{r}') \frac{\partial G(\boldsymbol{r}|\boldsymbol{r}')}{\partial y} \Big|_{y=0} d\boldsymbol{r}' + \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z j k_y P_{\mathrm{h}}(k_x, k_z) e^{j(k_x x + k_z z)}$$

Synthesize monopole sources [Spors+ AES Conv. 2008]

Applying WFR filtering method [Koyama+ IEEE TASLP 2013]

Decomposition into two components may lead to higher reproduction accuracy above spatial Nyquist freq.

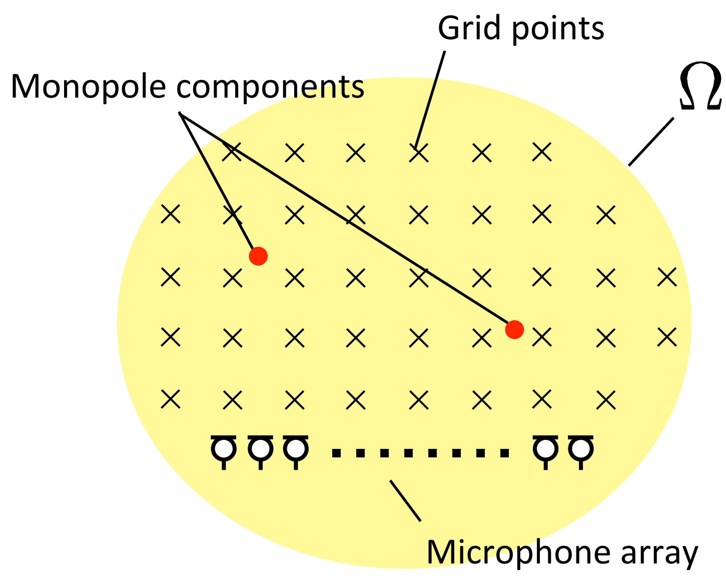
Sparse signal decomposition

➤ **Sparse representation of sound field**

$$p(\boldsymbol{r}) = \int_{\boldsymbol{r}' \in \Omega} Q(\boldsymbol{r}') G(\boldsymbol{r}|\boldsymbol{r}') d\boldsymbol{r}' + p_{\mathrm{h}}(\boldsymbol{r})$$

Discretization

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{h}$$



N : # of grid points within Ω, M : # of microphones, $N \gg M$

$\mathbf{y} \in \mathbb{C}^M$: Observed signal,

$\mathbf{x} \in \mathbb{C}^N$: Distribution of monopole components,

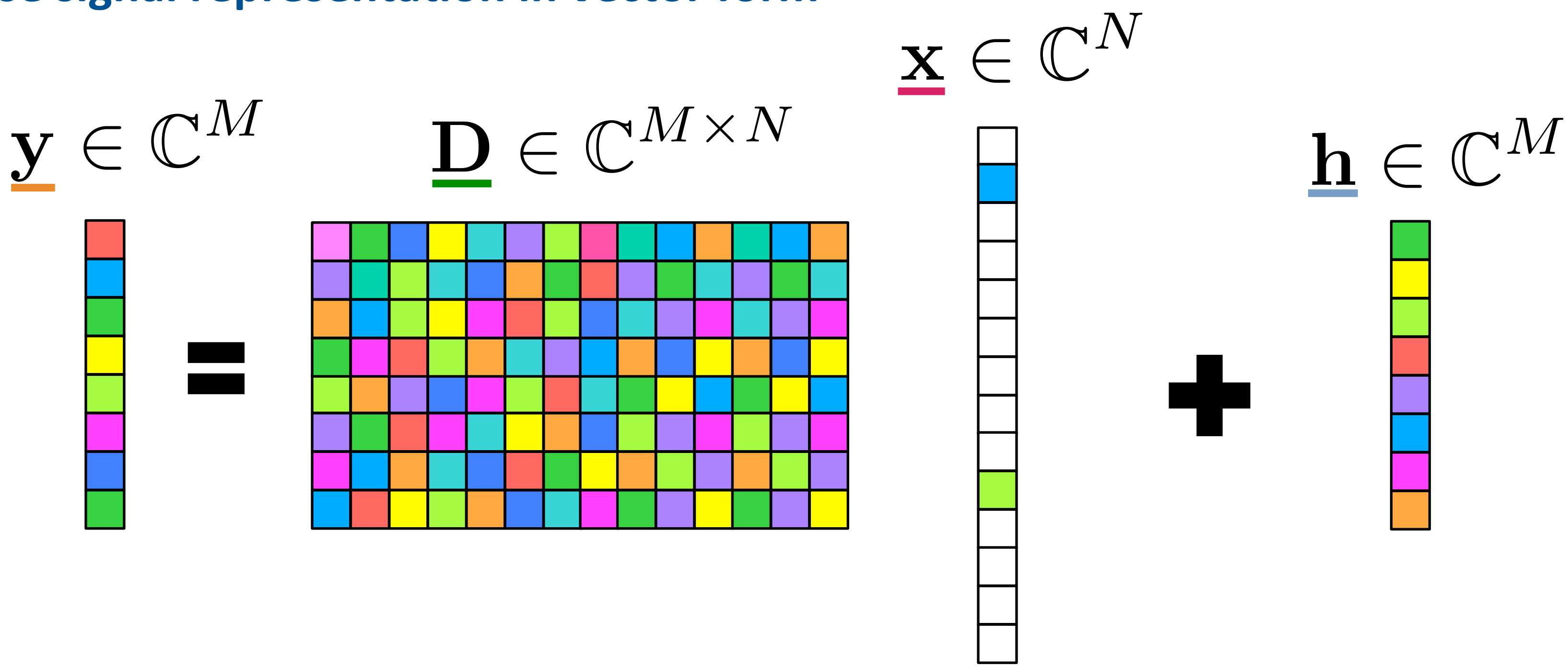
$\mathbf{D} \in \mathbb{C}^{M \times N}$: Dictionary matrix of Green's functions

$\mathbf{h} \in \mathbb{C}^M$: Ambient components

A few elements of $\mathbf{x}$ may have non-zero components

Sparse signal decomposition

- Sparse signal representation in vector form



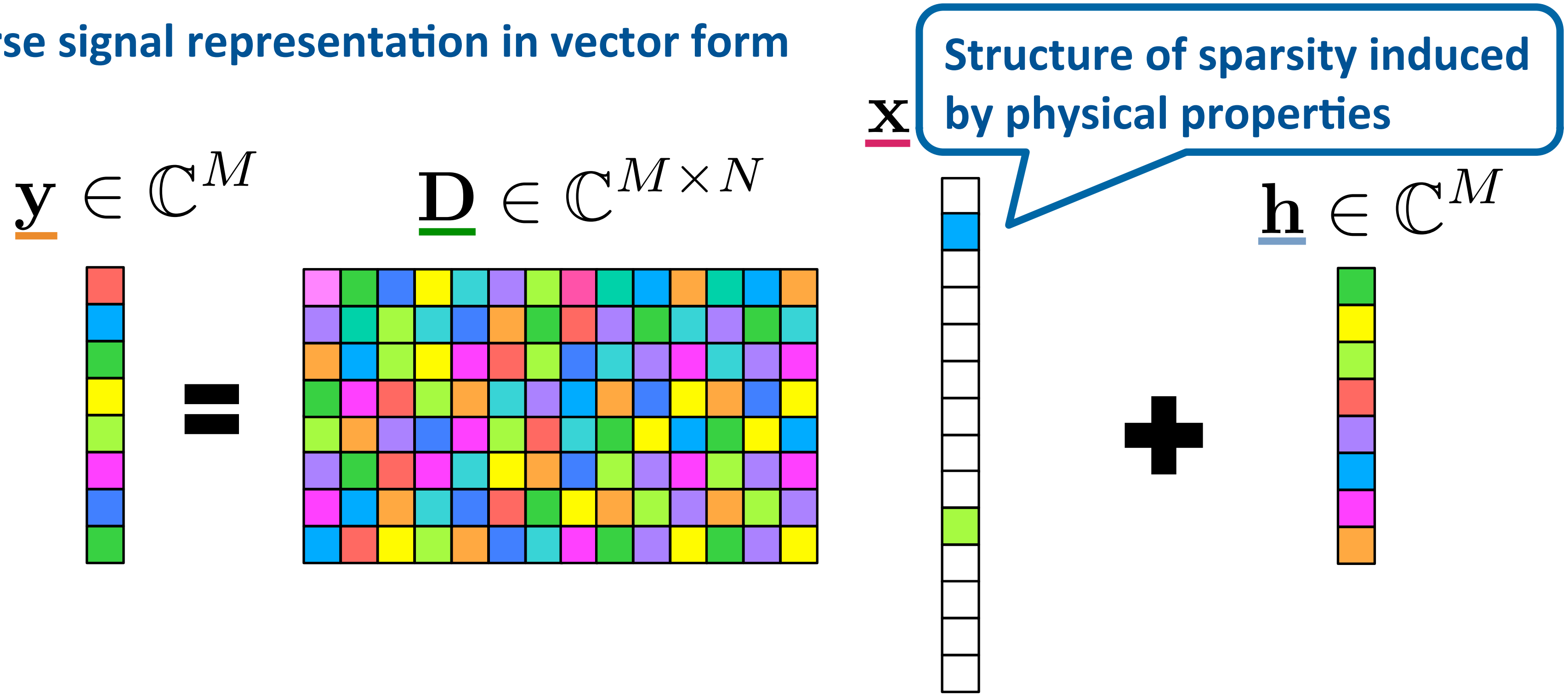
- Signal decomposition based on sparsity of $\mathbf{x}$

$$\text{minimize } \|\mathbf{x}\|_p^p \quad (p \le 1)$$
$$\text{subject to } \mathbf{y} = \mathbf{D}\mathbf{x}$$

Minimize ℓ_p -norm of $\mathbf{x}$

Structured sparsity based on physical properties

➢ **Sparse signal representation in vector form**



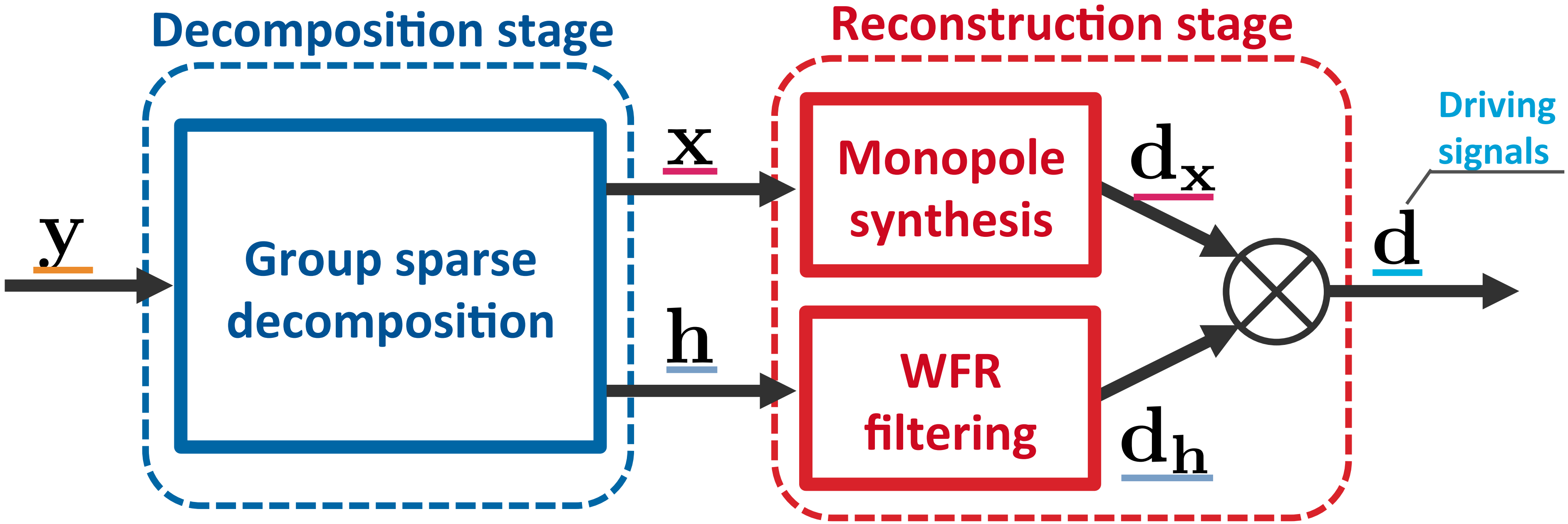
➢ **Group sparse signal models for robust decomposition**

- Multiple time frames
- Temporal frequencies
- Image sources and multipole components

➢ **Decomposition algorithm extending M-FOCUSS**

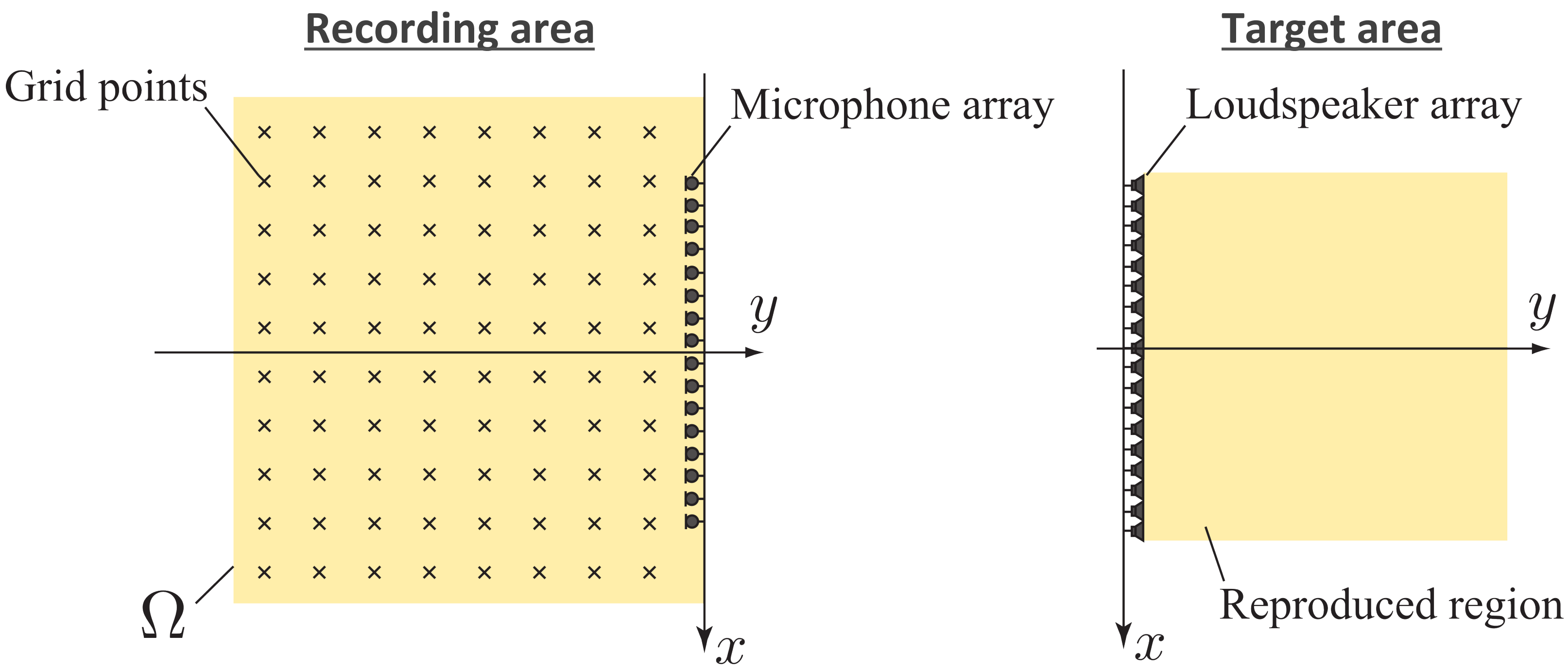
[Koyama+ ICASSP 2015 (to appear)]

Block diagram of signal conversion



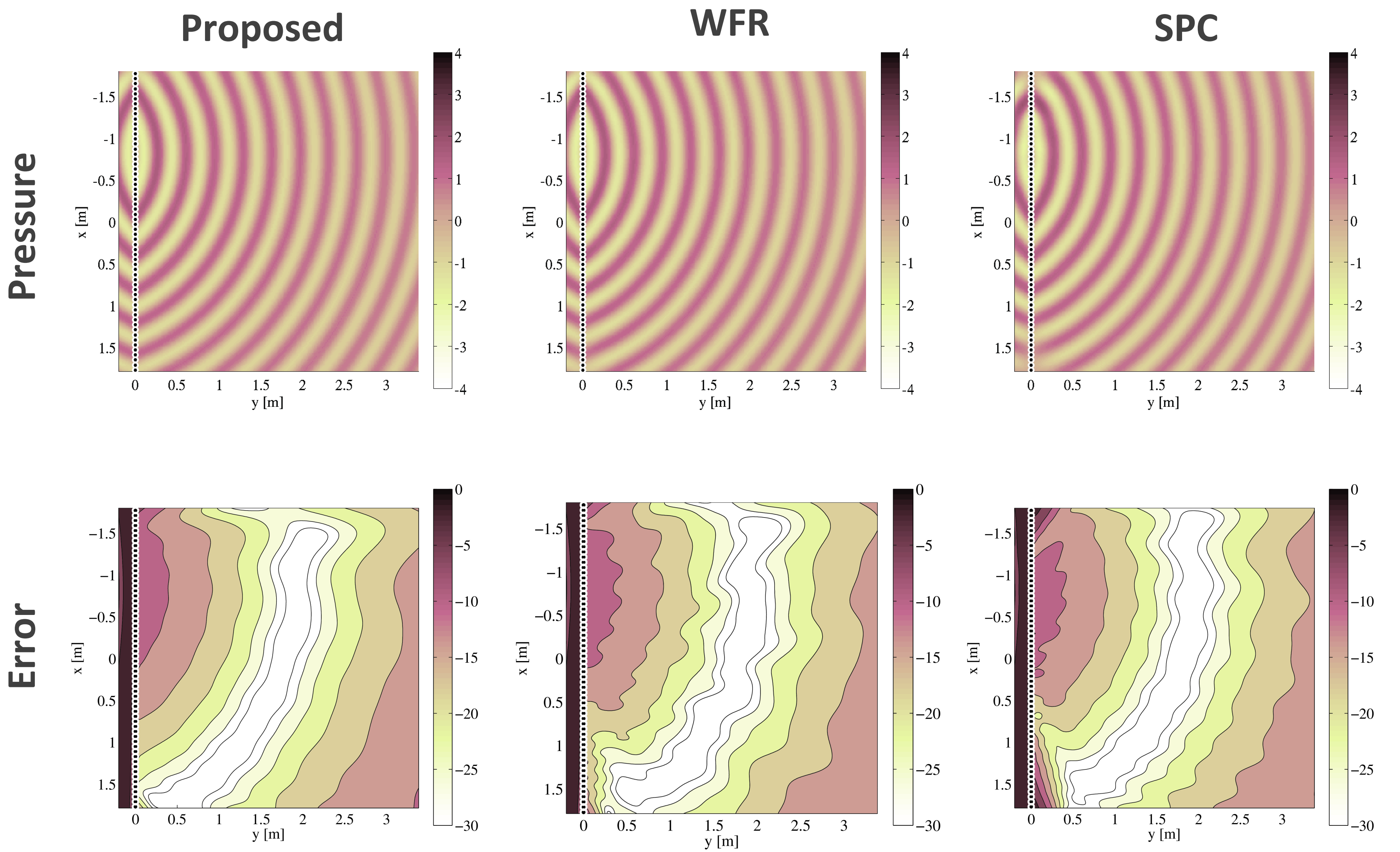
- **Decomposition stage**
  - Group sparse decomposition of $\mathbf{y}$
- **Reconstruction stage**
  - $\mathbf{x}$ and $\mathbf{h}$ are respectively converted into driving signals
  - $\mathbf{d}$ is obtained as sum of $\mathbf{d_x}$ and $\mathbf{d_h}$

Simulation Experiment



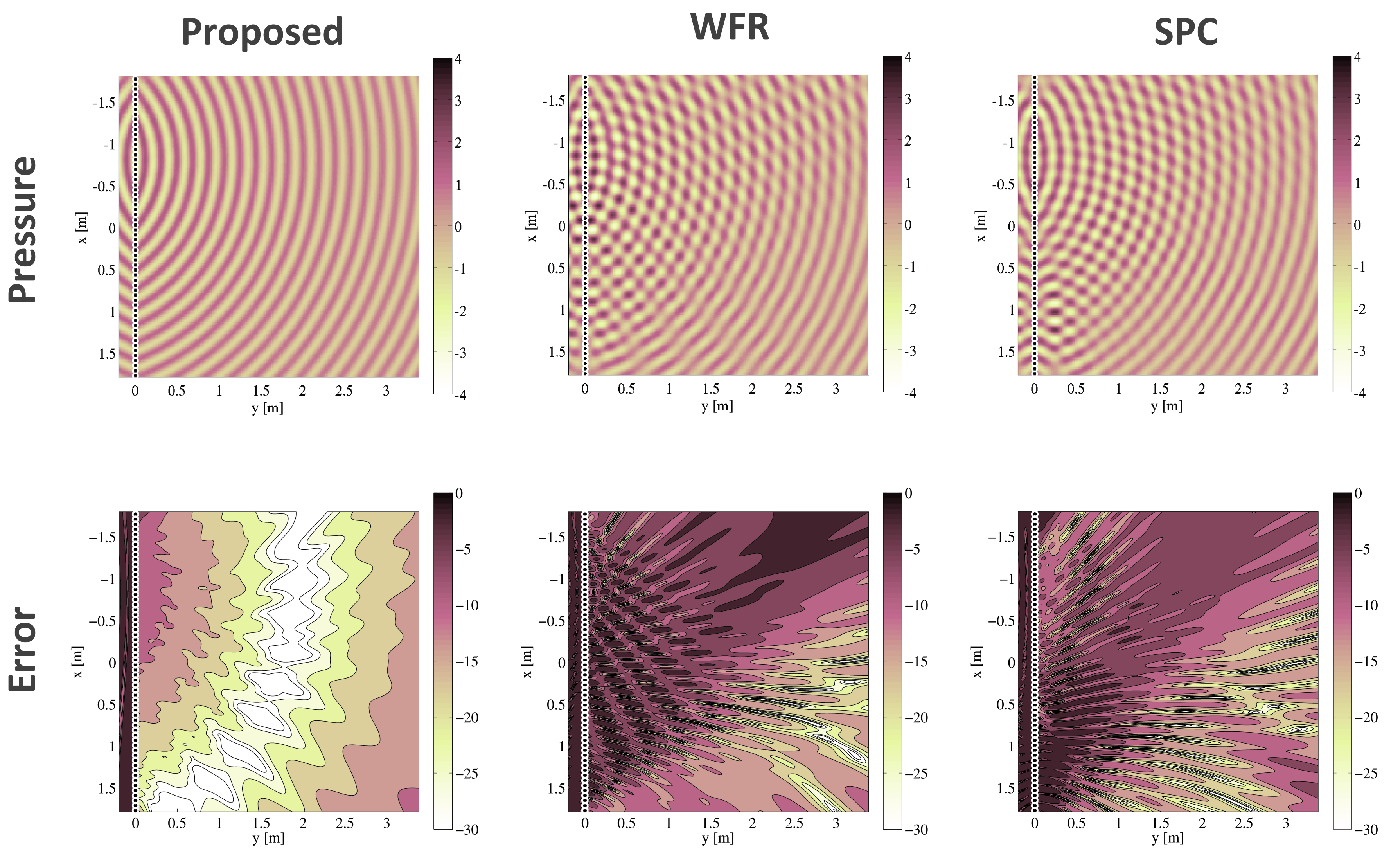
- Proposed method (**Proposed**), WFR filtering method (**WFR**), Sound pressure control method (**SPC**) were compared
- 32 microphones (12 cm intervals) and 64 loudspeakers (6 cm intervals)
- Ω: Rectangular region of 4.0x4.0 m, Grid points: (10cm, 20cm) intervals
- Source location: (-0.82, -0.86, 0.0) m

Reproduced sound pressure distribution (1.0 kHz)



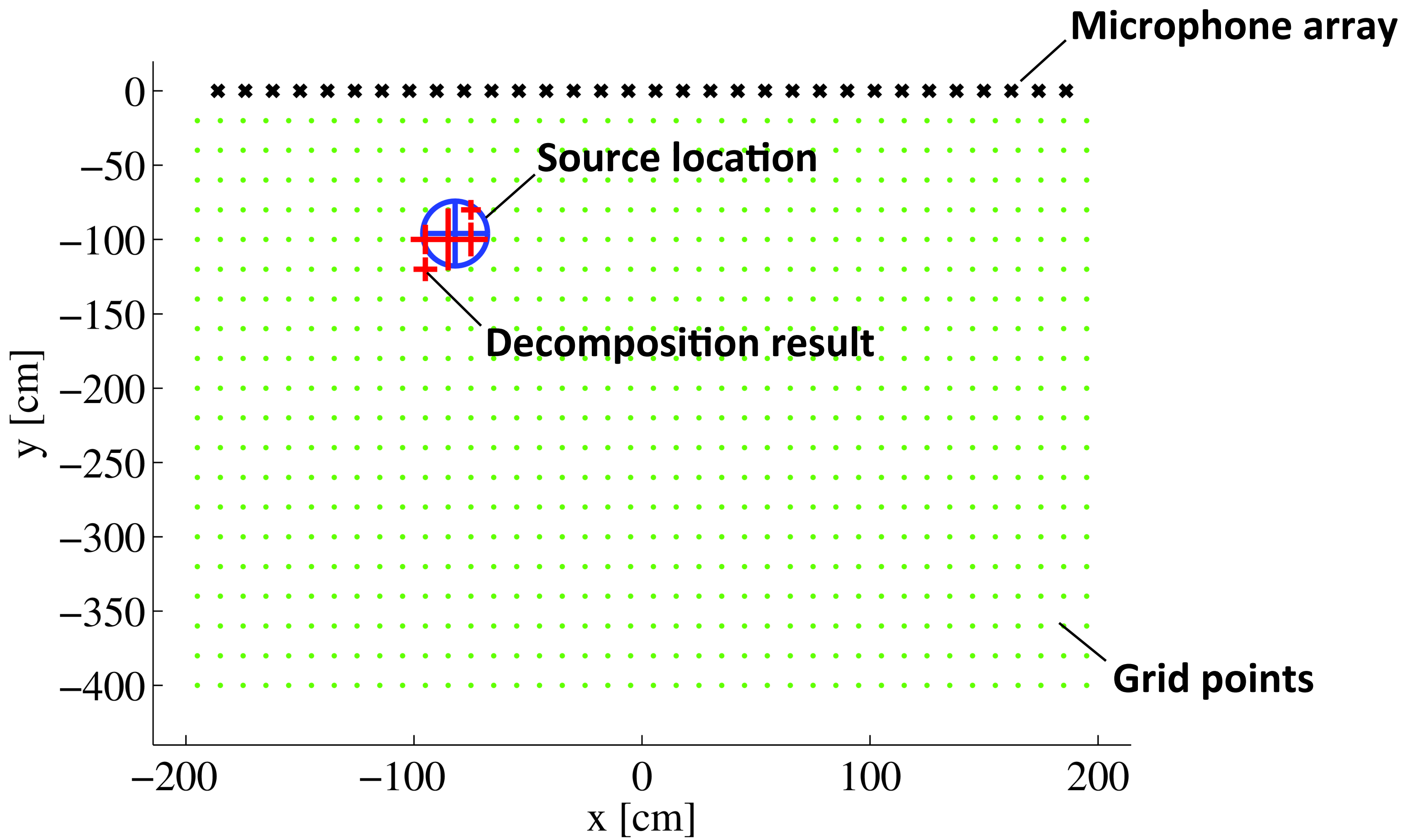
16.4 dB | 16.5 dB | 16.5 dB

Reproduced sound pressure distribution (1.8 kHz)



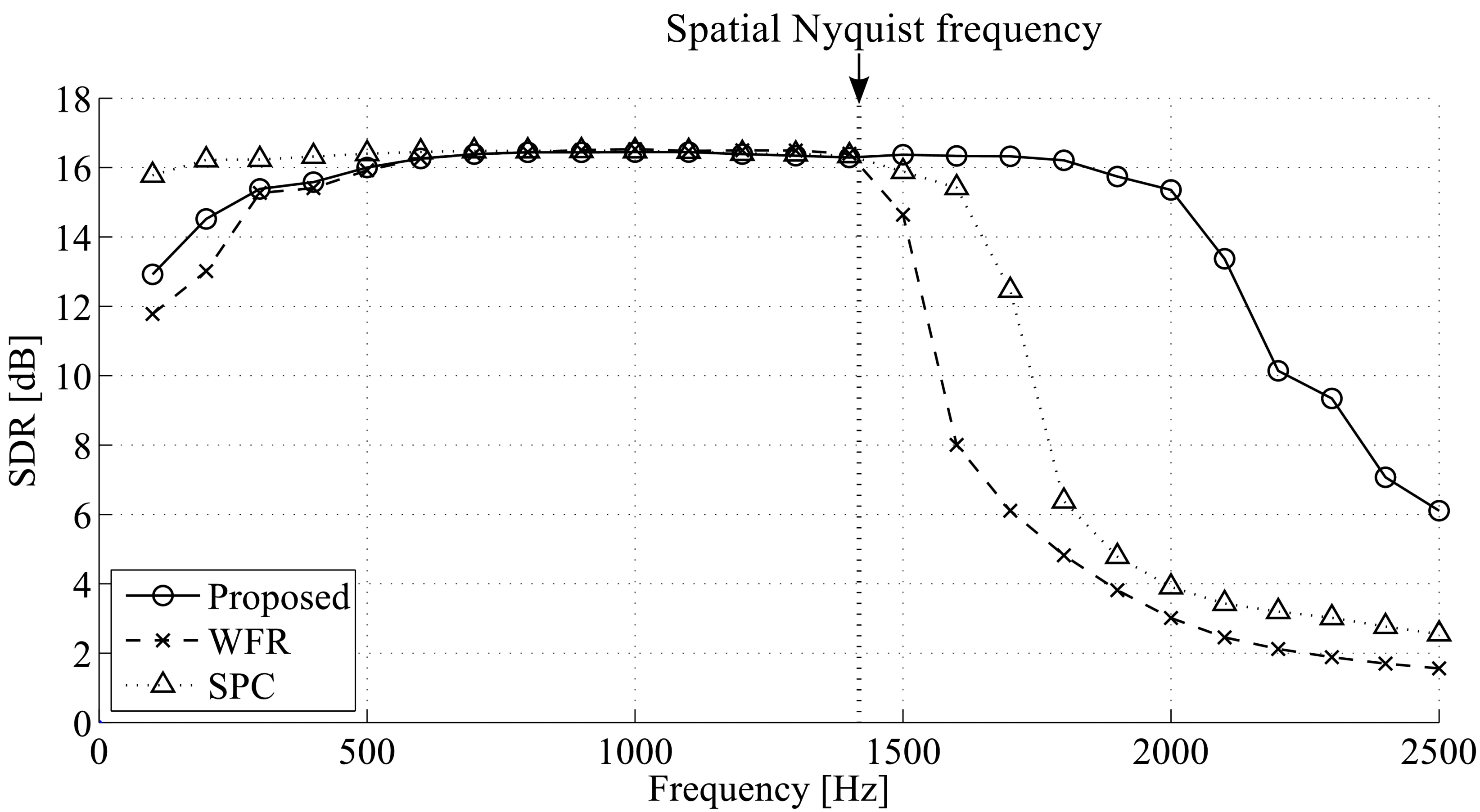
16.2 dB | 4.8 dB | 6.4 dB

Decomposition result (1.8 kHz)



A few dictionaries around source location is used for representing observed array signals

Frequency vs. SDR



SDRs above spatial Nyquist frequency were improved

Super-resolution in Sound Field Recording and Reproduction

- ✧ Two approaches to super-resolution in sound field recording and reproduction
- ✧ Source-location-informed recording and reproduction use prior information on source locations
- ✧ Sparse sound field representation based on generative model as a sum of monopole source and plane wave components
- ✧ Reproduction accuracy above spatial Nyquist frequency was improved in both methods

Related publications

- S. Koyama, *et al*. "Structured sparse signal models and decomposition algorithm for super-resolution in sound field recording and reproduction," *Proc. IEEE ICASSP*, 2015 (to appear).
- S. Koyama, *et al*. "Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts," *Proc. IEEE ICASSP*, pp. 4476-4480, 2014.
- S. Koyama, *et al*. "MAP estimation of driving signals of loudspeakers for sound field reproduction from pressure measurements," *Proc. IEEE WASPAA*, 2013.

Conclusion

- **Signal Conversion for Sound Field Recording and Reproduction**
  - Wave field reconstruction filtering method for fast and stable signal conversion for recording and reproduction
  - Real-time sound field transmission system

- **Super-resolution in Sound Field Recording and Reproduction**
  - Source-location-informed sound field recording and reproduction
  - Super-resolution based on sparse sound field representation

Thank you for your attention!