

音場再現技術の基本原理と展開

東京大学大学院 情報理工学系研究科 システム情報学専攻
(ex-日本電信電話株式会社 NTTメディアインテリジェンス研究所)

小山 翔一

物理現象の記録と再生

音響再生方式の変遷

あたかも同じ空間を共有しているかのような音響再生システムとは？

Holodeck (Star Trek)



2200~?

空間
解像
度

Monophonic



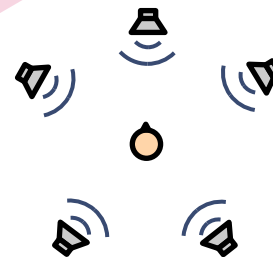
1857

Stereophonic



1931

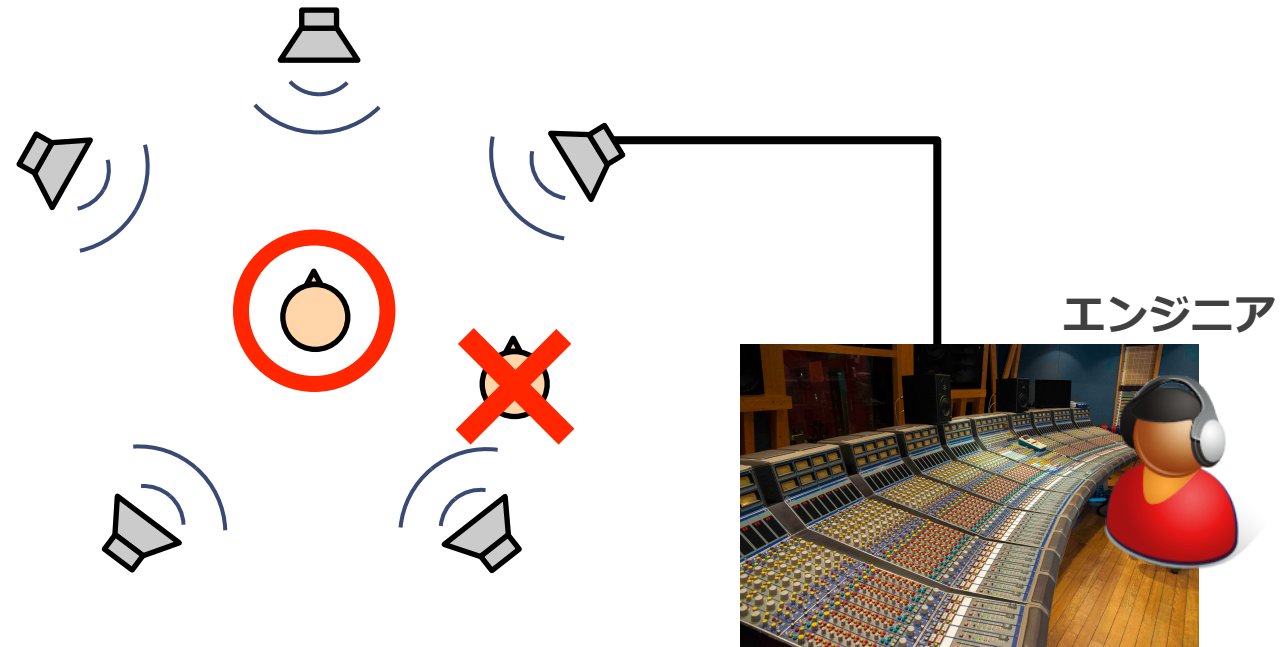
5.1ch Surround



1994

従来の音響再生技術における課題

- ステレオ、サラウンド5.1chなど

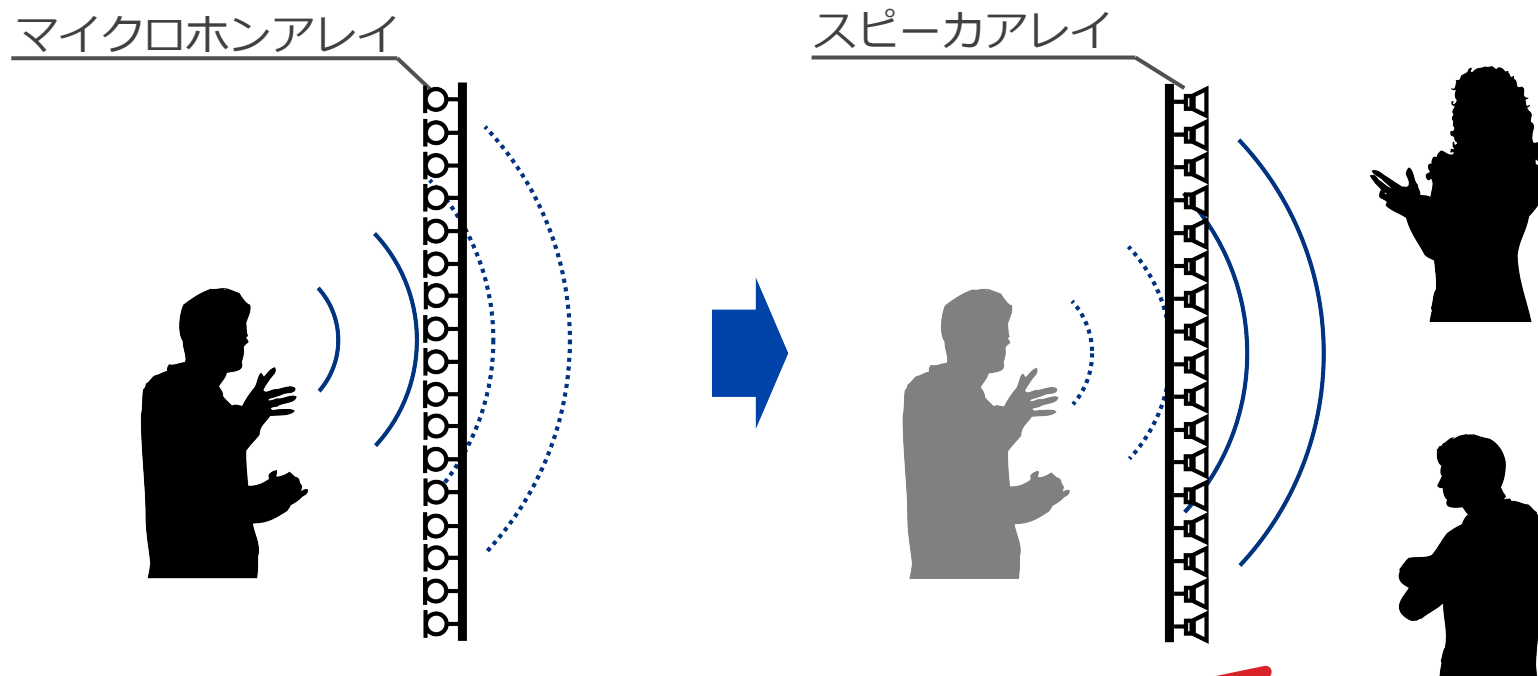


欠点

- ① 聴くことができる位置がスピーカを中心限定される
(スイートスポット)
- ② あくまで人工的な音の表現であり、音をデザインする人が必要

音場再現による高臨場音響再生

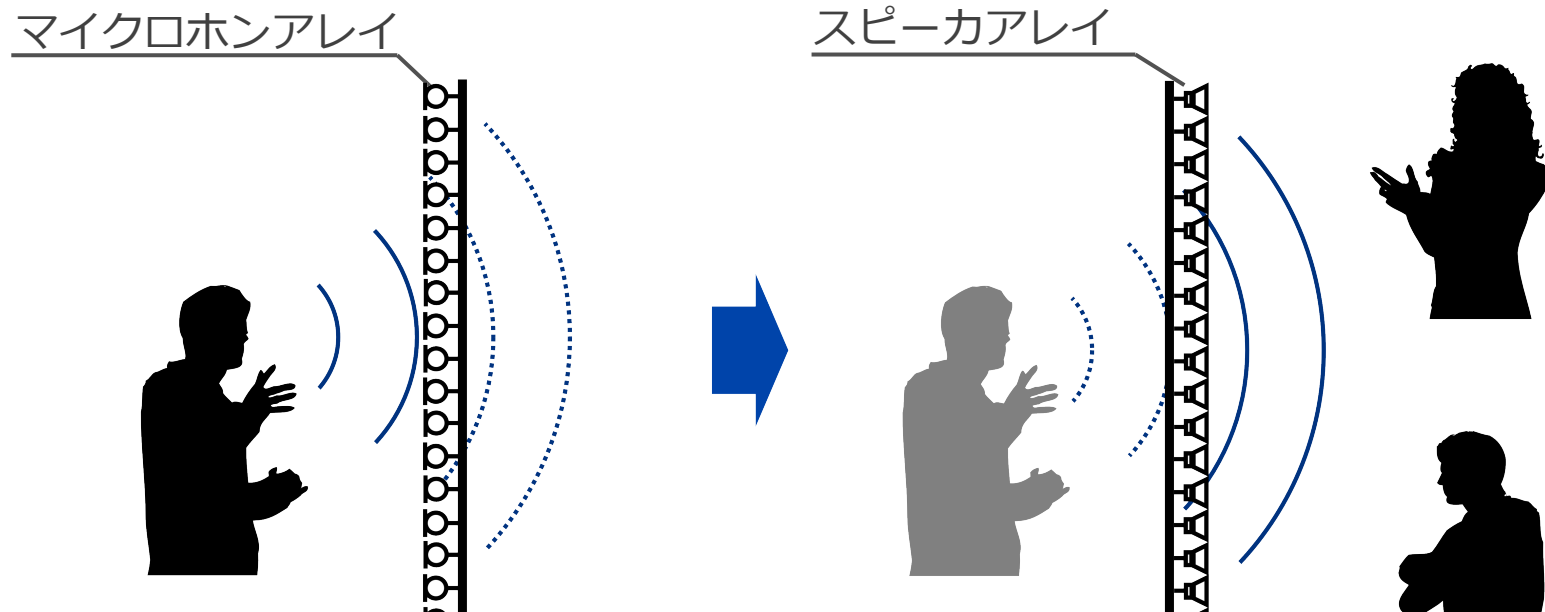
音場再現：空間的な音場そのものの再現



- 広い受聴領域が実現可能であり、複数の受聴者がいる場合でも適用可能
- 音源の距離感も含めた再現が可能
- リアルタイムの収音・再生システムへ適用が可能

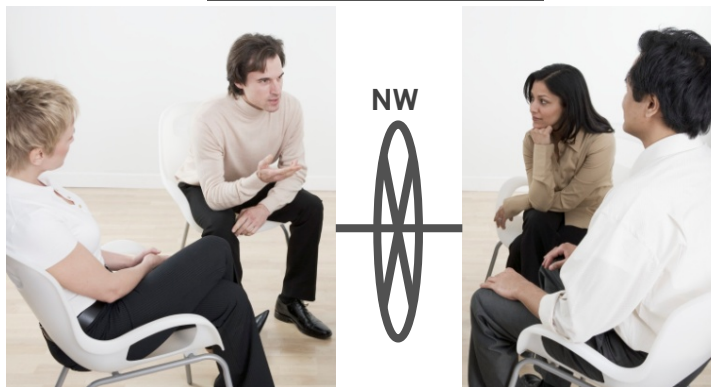
音場再現による高臨場音響再生

音場再現：空間的な音場そのものの再現



アプリケーション例

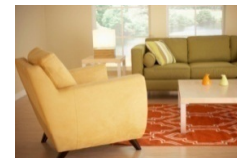
テレプレゼンス



コンテンツ配信



家庭



劇場



アウトライン

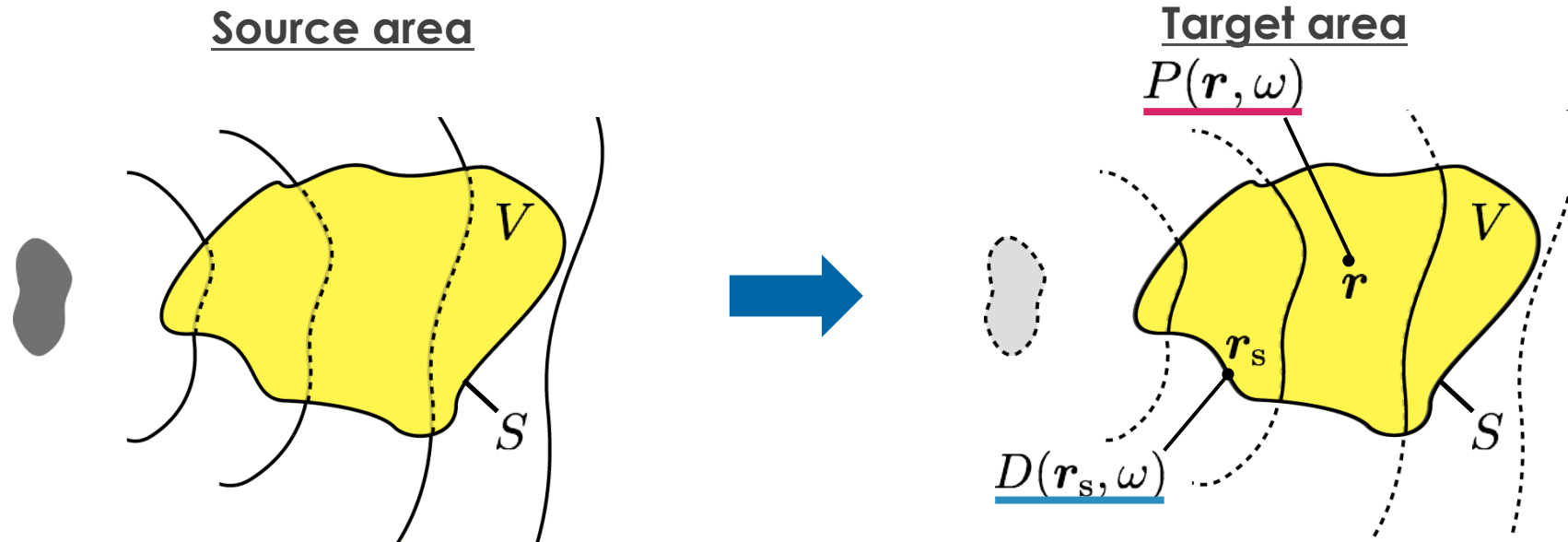
- 音場再現技術の基本原理
 - 音場再現問題とその従来手法
 - 音場収音・再現のための波面再構成フィルタ
 - リアルタイム音場伝送システムの実装例
- 超解像型音場再現に向けて
 - 空間エイリアシング問題回避の必要性
 - 音源位置事前情報を用いた音場収音・再現



音場再現技術の基本原理

音場再現問題とは

対象領域 V の境界面 S 上に配置した二次音源(=スピーカアレイ)を用いて、
所望の音場を V 内で再構成するための駆動信号 $D(\mathbf{r}_s, \omega)$ を求める



Kirchhoff-Helmholtz積分方程式：境界面 S から V 内の音場を決める物理式

$$\underbrace{P(\mathbf{r}, \omega)}_{\text{内部音場}} = \iint_S \left(\underbrace{G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega)}_{\substack{\text{3次元自由空間} \\ \text{Green関数}}} \underbrace{\frac{\partial P(\mathbf{r}_s, \omega)}{\partial n}}_{\text{音圧勾配}} - \underbrace{P(\mathbf{r}_s, \omega)}_{\text{音圧}} \underbrace{\frac{\partial G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega)}{\partial n}}_{\text{音圧勾配}} \right) dS$$

- ➡ 境界面 S 上のモノポール・ダイポール音源で音圧と音圧勾配を制御
- ➡ 実デバイスを用いて実現可能な形に変形することが必要！

音場再現手法の比較

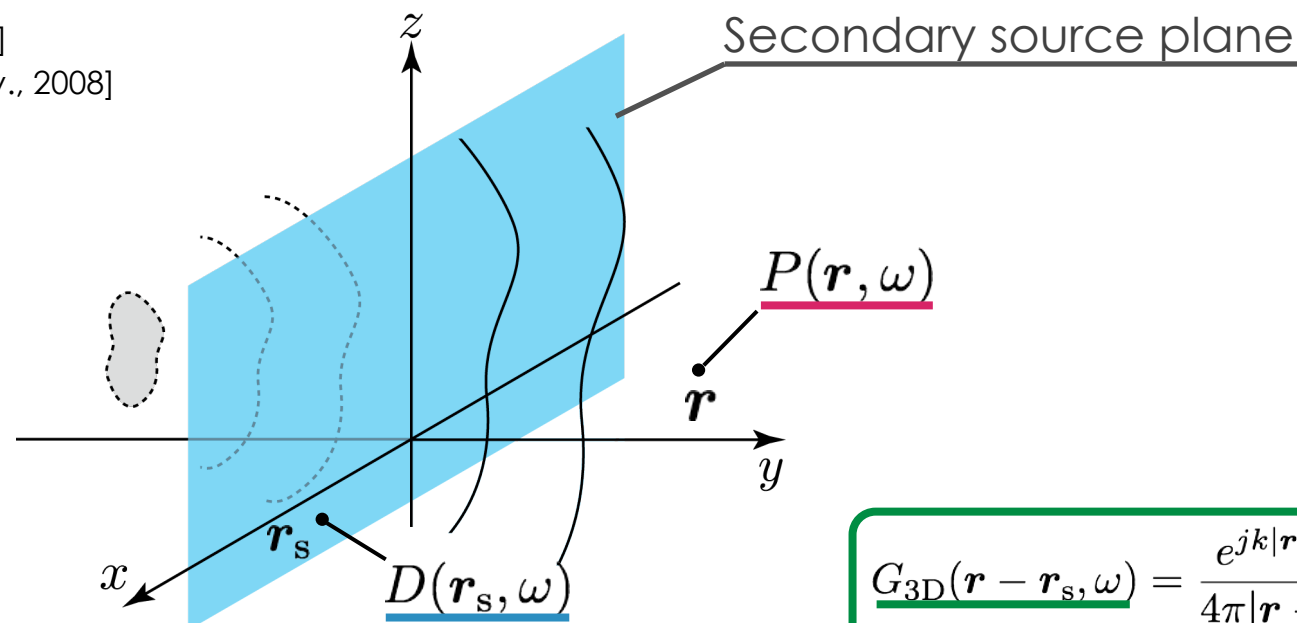
	アレイ配置	概要
Wave Field Synthesis (WFS)	平面／直線	Kirchhoff-Helmholtz積分／Rayleigh積分に基づくモデルベース手法
Higher Order Ambisonics (HOA)	球	球面調和関数展開に基づくエンコーディング／デコーディング
逆フィルタに基づく手法	任意	最小二乗法などに基づく多点音圧制御
波面再構成(WFR)フィルタ	平面／直線／円筒／球／円	空間スペクトル上での直接的な信号変換による収音・再現

➡ ここでは平面／直線状アレイに適用可能な手法にフォーカス

Wave Field Synthesis (WFS)

[Berkhout, et al., JASA, 1993]

[Spors, et al., AES 124th Conv., 2008]



第1種Rayleigh積分に基づく所望音場

$$\underline{P_{\text{des}}(\mathbf{r}, \omega)} = -2 \iint_{-\infty}^{\infty} \frac{\partial \underline{P(\mathbf{r}_s, \omega)}}{\partial y_s} \underbrace{G_{3D}(\mathbf{r} - \mathbf{r}_s, \omega)}_{\substack{\text{3次元自由空間} \\ \text{Green関数}}} dx_s dz_s$$

二次音源によって合成される音場

$$\underline{P_{\text{syn}}(\mathbf{r}, \omega)} = \iint_{-\infty}^{\infty} \underline{D(\mathbf{r}_s, \omega)} \underbrace{G(\mathbf{r} - \mathbf{r}_s, \omega)}_{\text{伝達関数}} dx_s dz_s$$

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

各二次音源が点音源と近似
($\underline{G} \simeq \underline{G_{3D}}$)すると,

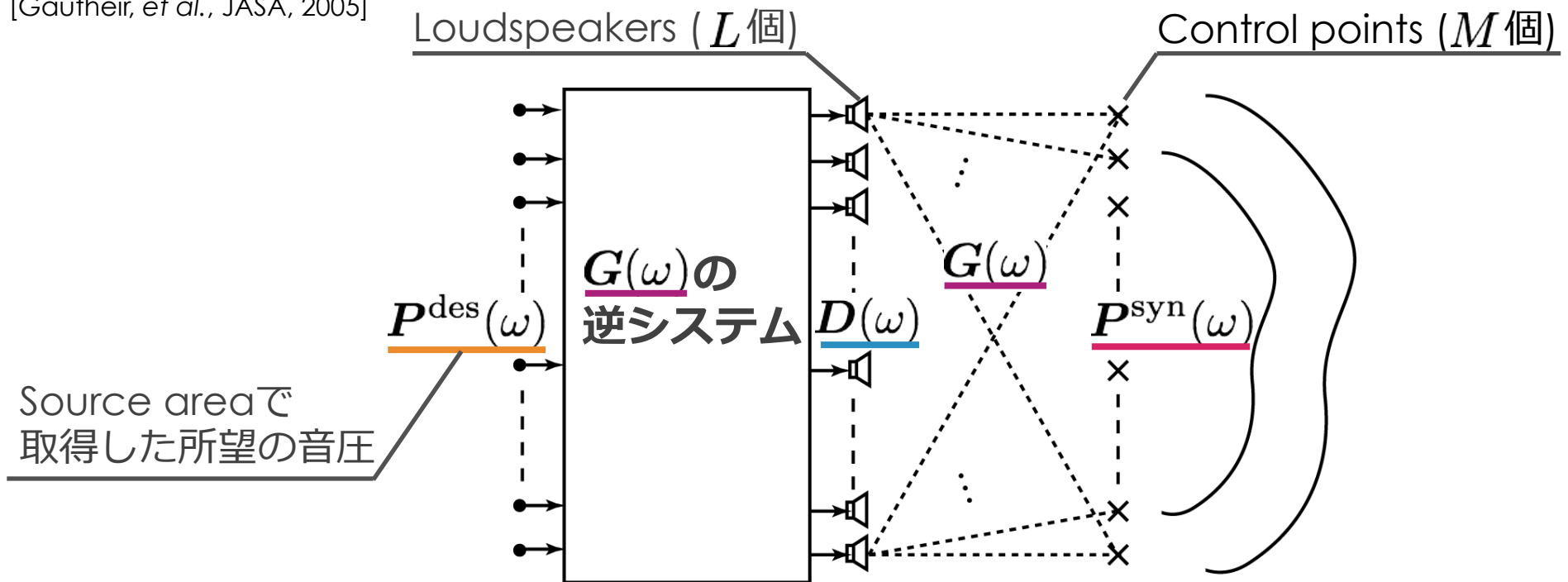
$$\underline{D(\mathbf{r}_s, \omega)} = -2 \frac{\partial \underline{P(\mathbf{r}_s, \omega)}}{\partial y_s}$$

➡ 音圧勾配と一致

所望音場における境界面上の音圧勾配が必要となるため、シミュレートした音場の合成に用いられる場合が多い

逆フィルタに基づく手法

[Gauthier, et al., JASA, 2005]



制御点上で所望の音圧と一致するような逆システムを用いる

➡ 最小二乗法 (LS法) による $G(\omega) \in \mathbb{C}^{L \times M}$ の逆フィルタの設計

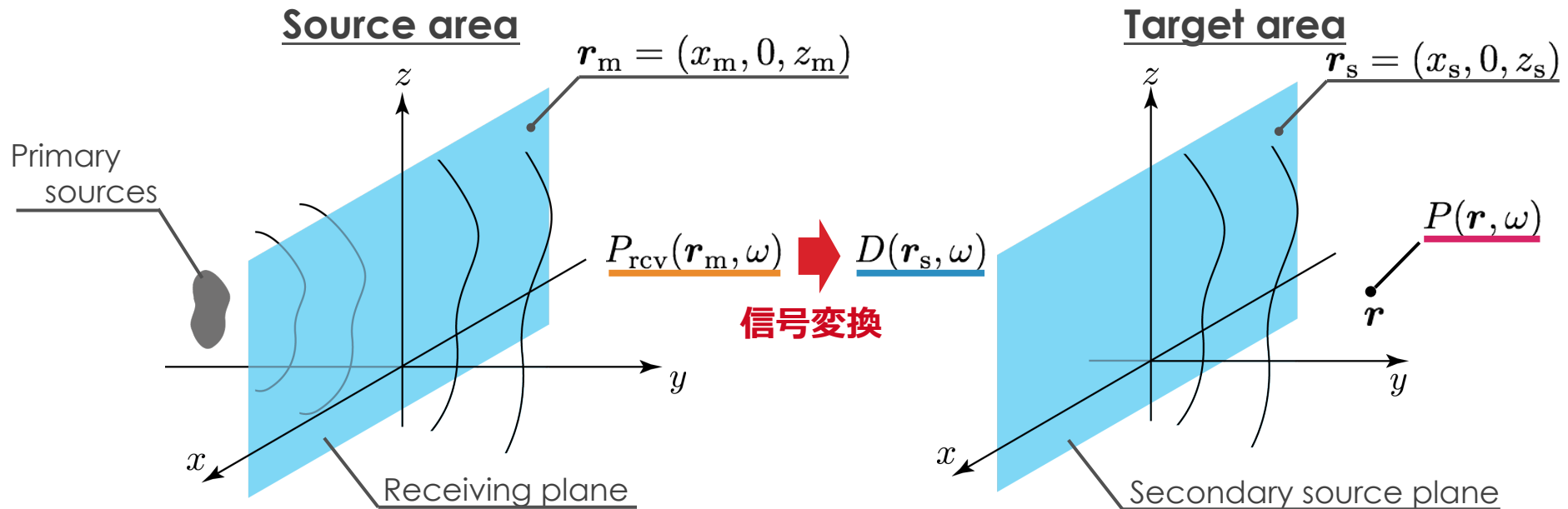
$$\hat{D}(\omega) = (\underline{G^H(\omega)} \underline{G(\omega)} + \beta \underline{I})^{-1} \underline{G^H(\omega)} \underline{P^{\text{des}}(\omega)}$$

正則化パラメータ

実環境の収音場を再現する場合にも適用可能であるが
フィルタの安定性が問題となる場合が多い

波面再構成(WFR)フィルタ

マイクロホンアレイ信号の直接変換により実環境で收音した音場を再現



第1種Rayleigh積分に基づく所望音場

$$P_{\text{des}}(\mathbf{r}, \omega) = -2 \iint_{-\infty}^{\infty} \frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m} \underbrace{G_{3D}(\mathbf{r} - \mathbf{r}_m, \omega)}_{\substack{\text{3次元自由空間} \\ \text{Green関数}}} dx_m dz_m$$

二次音源によって合成される音場

$$P_{\text{syn}}(\mathbf{r}, \omega) = \iint_{-\infty}^{\infty} \underbrace{D(\mathbf{r}_s, \omega)}_{\text{伝達関数}} \underbrace{G(\mathbf{r} - \mathbf{r}_s, \omega)}_{\text{伝達関数}} dx_s dz_s$$

各二次音源が点音源と近似
($G \simeq G_{3D}$)すると,

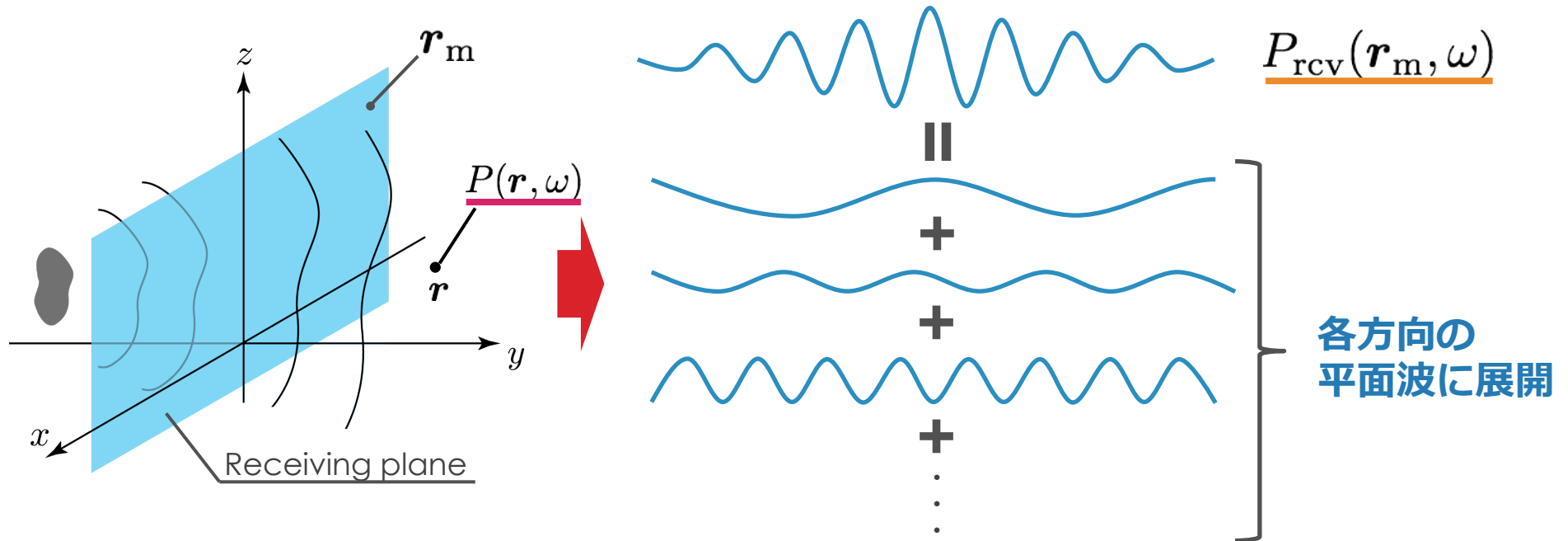
$$D(\mathbf{r}_s, \omega) = -2 \frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m}$$

☹ $P_{\text{rcv}}(\mathbf{r}_m, \omega)$ のみが既知

受音平面上の音圧分布から、音圧勾配分布の推定が必要

音場の平面波展開

受信平面上の音圧分布を平面波の和として表現



各平面波について y 方向の伝播が一意に決まることから音圧勾配を推定できる

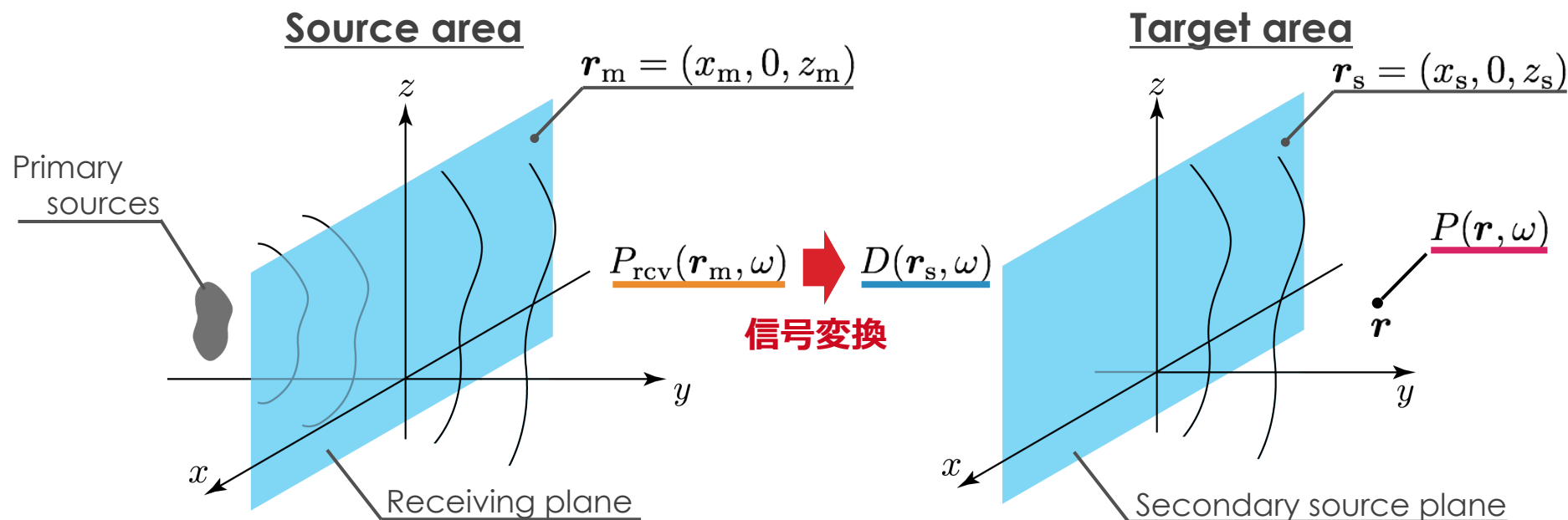
➡ 平面波展開は音場を空間スペクトル領域で表現することと等価

$$\underline{P(\mathbf{r}, \omega)} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) e^{j(k_x x + \textcircled{k_y} y + k_z z)}$$

受信平面上の
空間スペクトル

$$k_y = \pm \sqrt{k^2 - k_x^2 - k_z^2}$$

波面再構成フィルタ



第1種Rayleigh積分に基づく所望音場の空間スペクトル [Koyama, et al., IEEE Trans. ASLP, 2013]

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

二次音源によって合成される音場の空間スペクトル

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

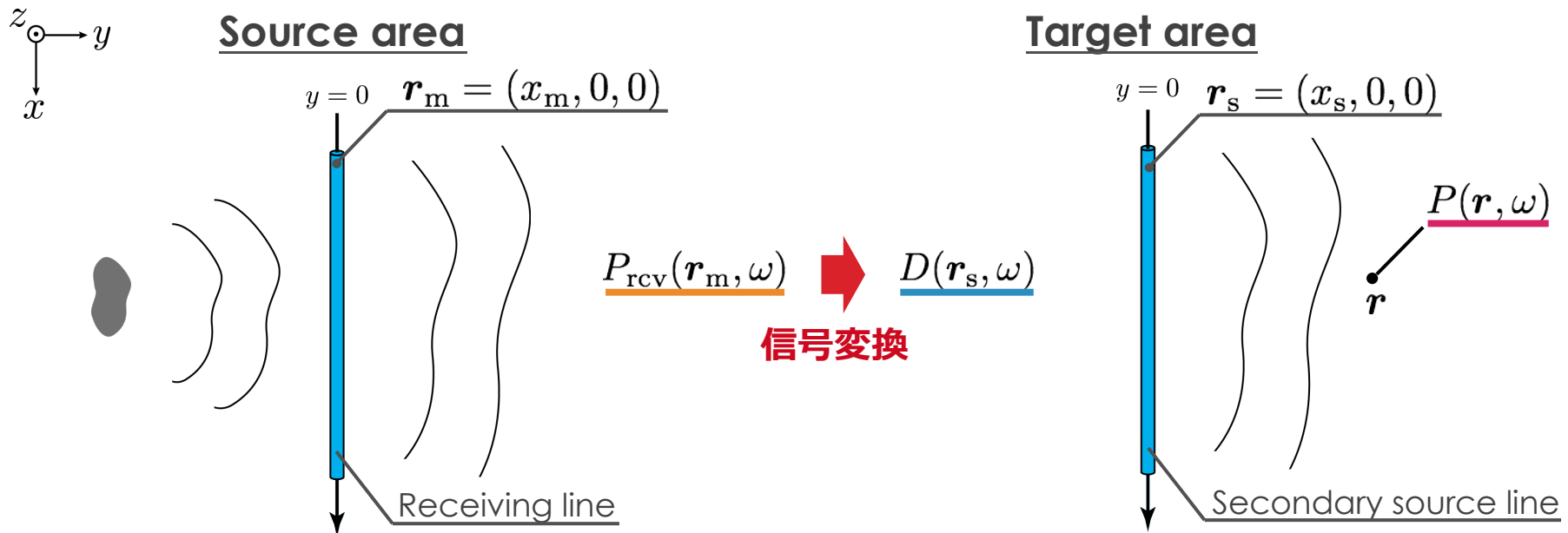
$$\Rightarrow \tilde{D}(k_x, k_z, \omega) = -2jk_y \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega)$$

空間スペクトル領域での定式化に基づく信号変換により、
実環境で收音した音場を再現

直線状アレイのためのWFRフィルタ

- 聴覚上、水平角方向の定位が仰角方向に比べて重要
- 平面状アレイの構成は実装上困難

直線状分布
のための定式化



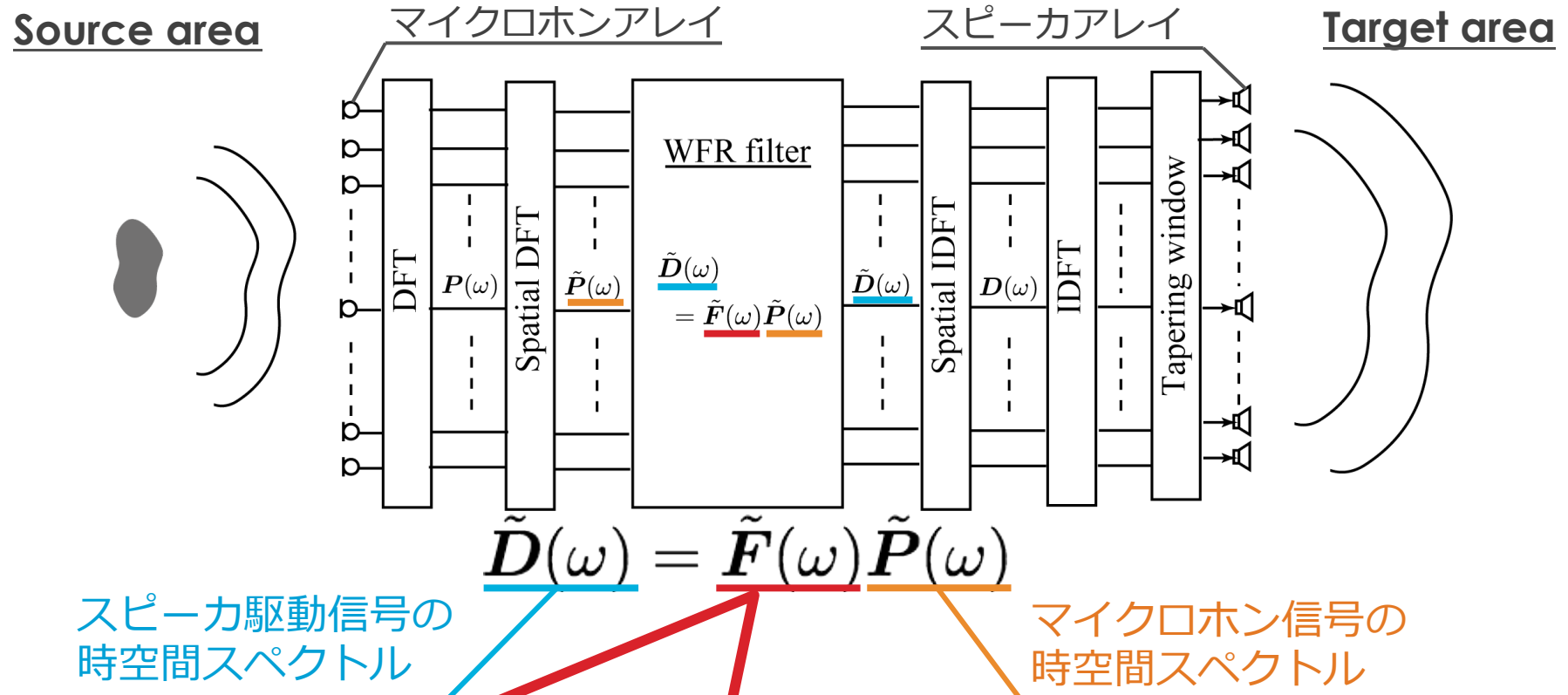
$$\tilde{D}(k_x, \omega) = -4j \frac{e^{jk_\rho y_{\text{ref}}}}{H_0^{(1)}(k_\rho y_{\text{ref}})} \tilde{P}_{\text{rcv}}(k_x, 0, 0, \omega)$$

(y_{ref} : 受聴直線=振幅が所望音場と一致する直線)

2次元音場の近似により、振幅減衰が所望音場に比べて大きい

直線状アレイのためのWFRフィルタ

実装上は時空間の2次元FIRフィルタ畳み込み：**WFRフィルタ**



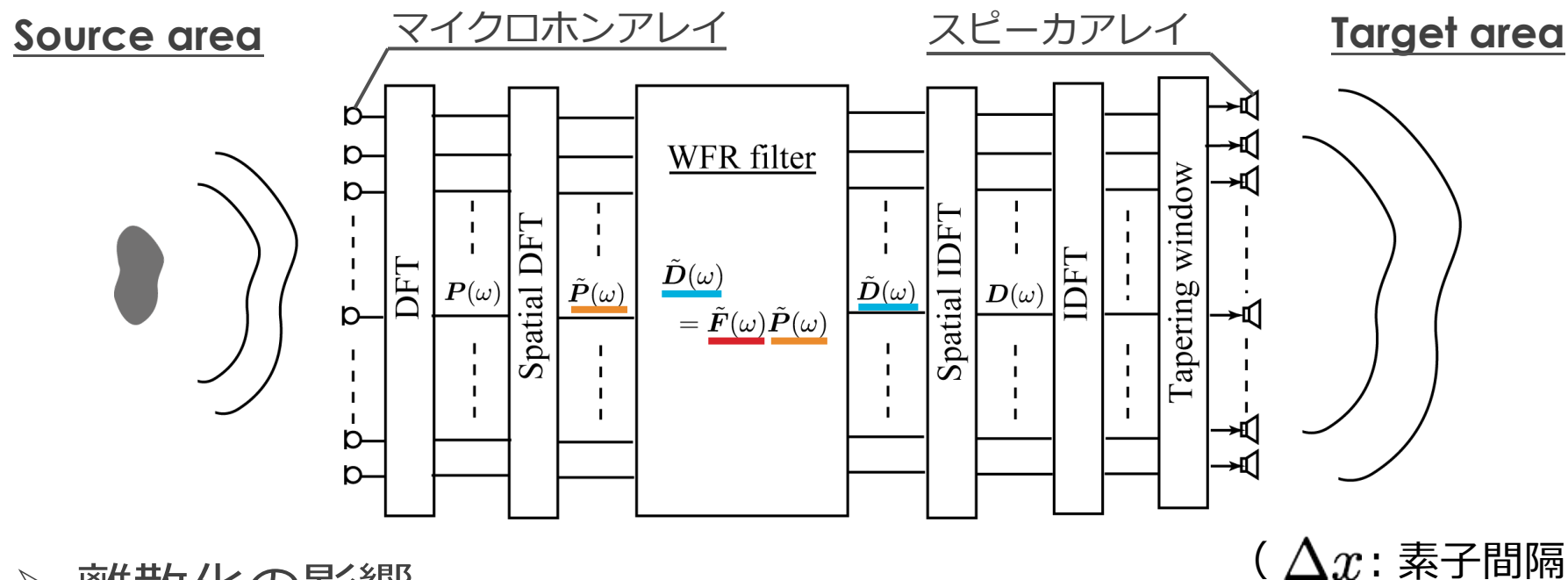
WFRフィルタ：各要素が以下で決まる対角行列

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

リアルタイムでの信号変換による音場収音・再現も実現可能

離散化・有限長での打ち切りによる影響

実装上は時空間の2次元FIRフィルタ畳み込みとなる：波面再構成（WFR）フィルタ



➤ 離散化の影響

- $|k_{x,i}| > \pi/\Delta x$ の成分は空間エイリアシングの影響を受ける

➡ $c/2\Delta x$ Hz以上の帯域で音色に変化

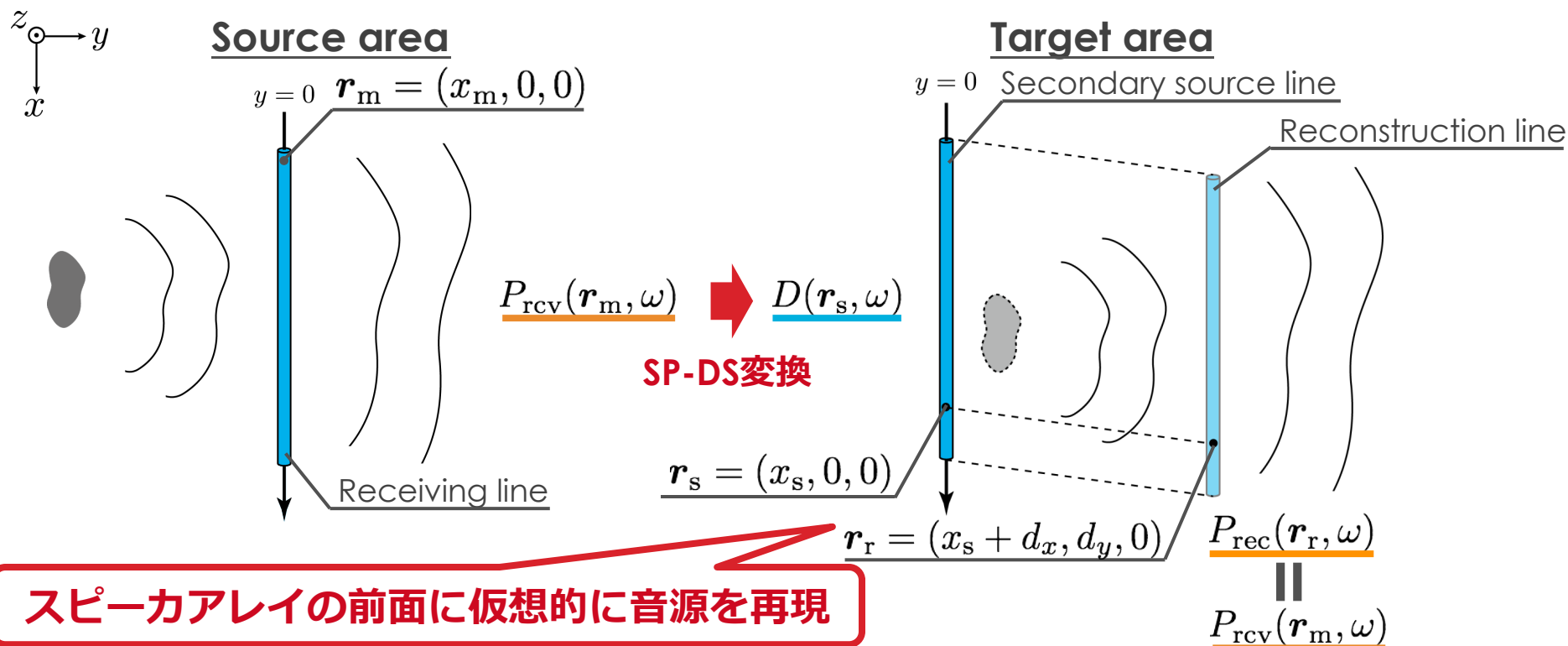
➤ 有限長での打ち切りの影響

- アレイ両端からの不要な反射波が発生

➡ 両端で滑らかに減衰するような窓掛けにより抑制

音場再現位置の操作

観測音圧分布をシフトした位置で再現



Reconstruction line上の音圧分布からSecondary source line上の音圧分布を推定

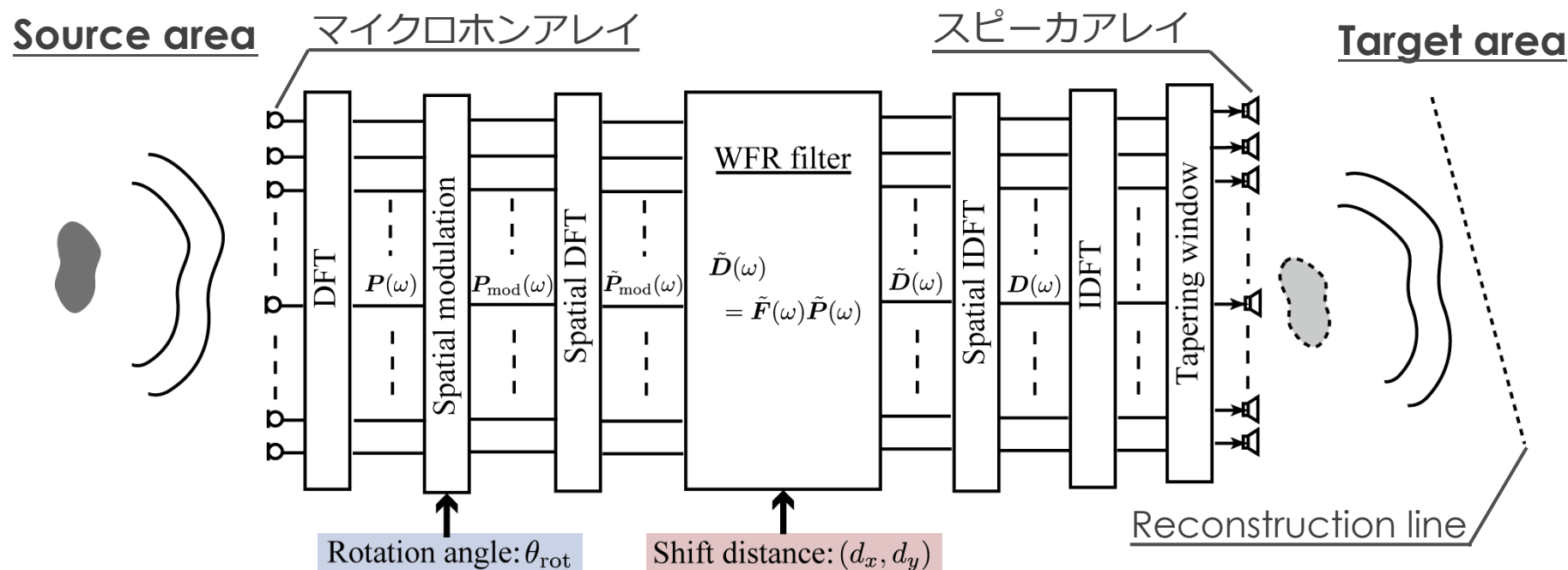
➡ 空間スペクトルの位相シフトとして定式化できる

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

近距離場音響ホログラフィの原理に基づく

再現位置操作のためのWFRフィルタ

空間周波数変調・位相シフトによる再現音場の回転とシフト



変調

$$P_{\text{mod}}(\omega) = P(\omega) \cdot e^{-jk \sin \theta_{\text{rot}}} \quad \Rightarrow \quad \theta_{\text{rot}} \text{ だけ回転}$$

位相シフト

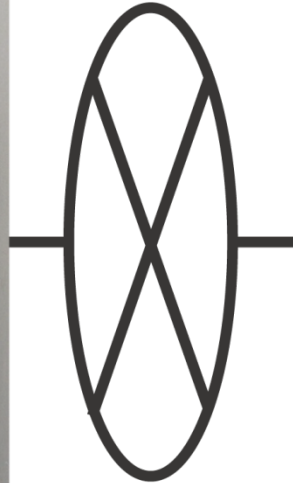
$$\tilde{F}_i(\omega) = -4j \frac{e^{j\sqrt{k^2 - k_{x,i}^2} y_{\text{ref}}}}{H_0^{(1)}(\sqrt{k^2 - k_{x,i}^2} y_{\text{ref}})} e^{j(k_{x,i} d_x + \sqrt{k^2 - k_{x,i}^2} d_y)} \quad \Rightarrow \quad (d_x, d_y) \text{ だけシフト}$$

リアルタイム音場伝送システム

横須賀



Network

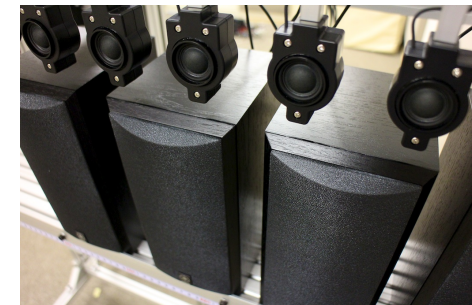


武蔵野



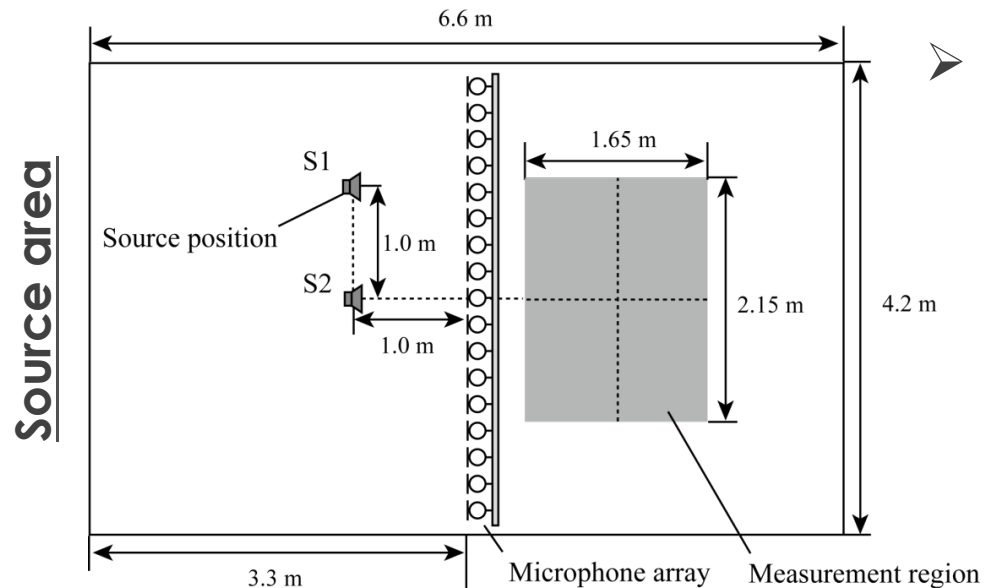
➤ IPネットワークを用いたリアルタイム音場伝送システム

- 直線状マイクロホン・スピーカアレイ+WFRフィルタによる信号変換
- マイクロホン：64ch, 6cm間隔
- スピーカ（小, 高域用）：64ch, 6cm間隔
- スピーカ（大, 低域用）：32ch, 12cm間隔
- アレイ長：3.84m
- サンプリング周波数：48kHz
- 遅延時間：信号処理 21ms+バッファ 131ms=152ms



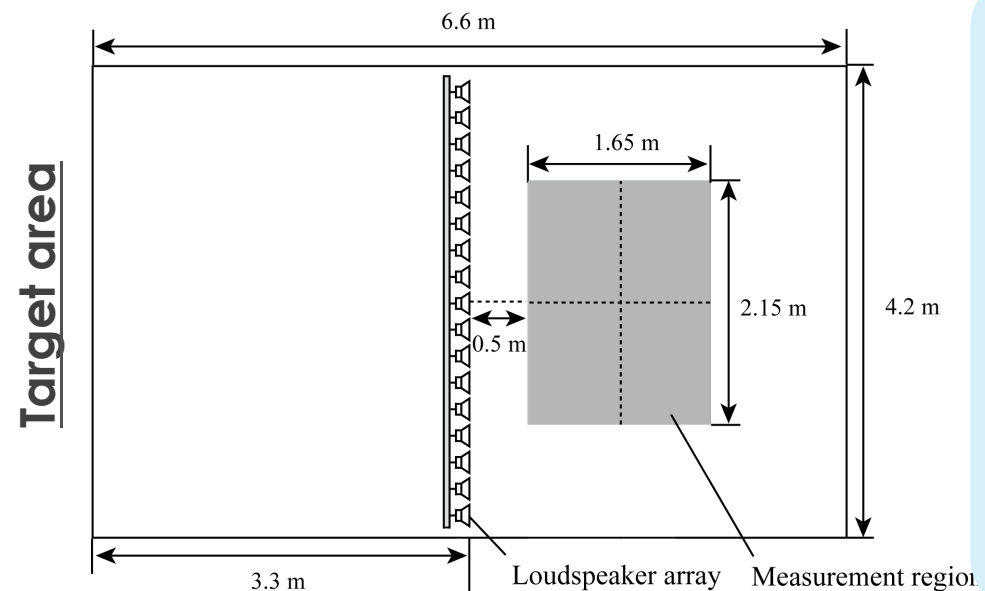
※片方向の伝送・再現のみ

実環境での測定実験



➤ 再現音場の可視化・原音場との比較

- 平面状測定領域 2.15x1.65mにおいて、インパルス応答を1.5cm間隔で計測（144x108点）
- 測定環境の残響時間 (T_{60}) : 167ms
- 音源には一般的なスピーカを利用
- スピーカアレイは高域用のみ使用
- 原音場も同様にして測定

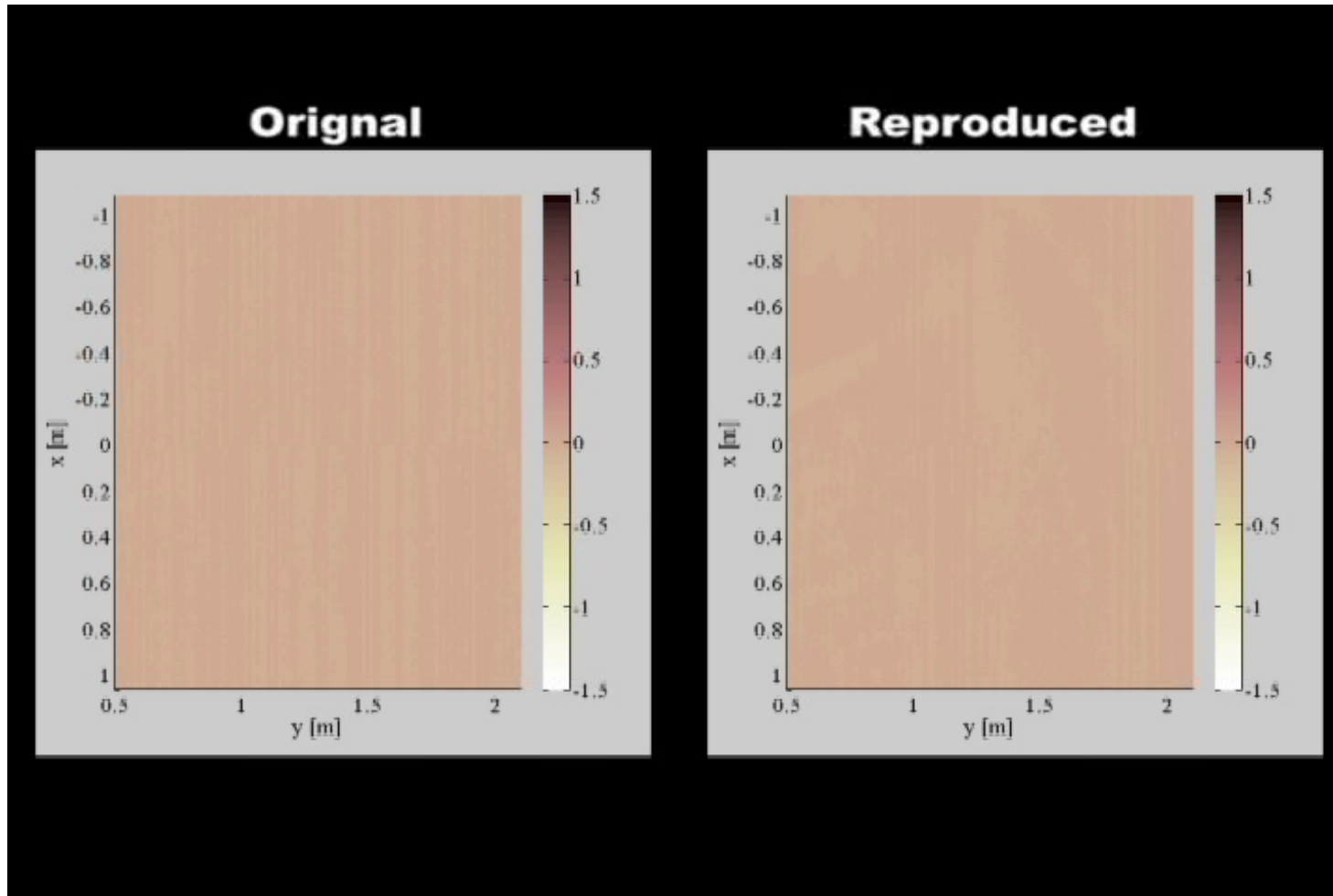


測定実験の様子



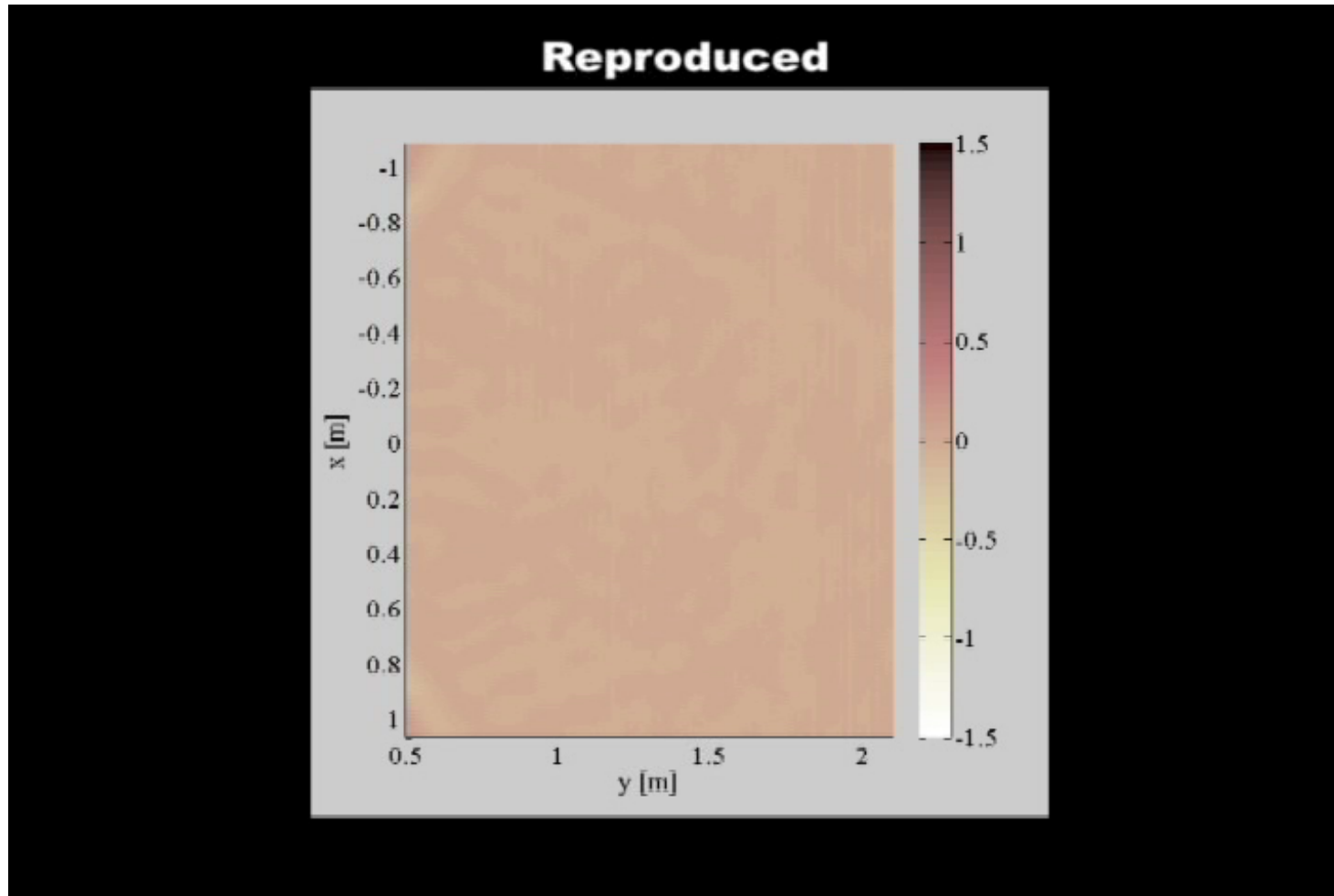
再現音場の可視化①

- 音源信号：帯域制限パルス信号（～2.6kHz）
- 音源：スピーカ，位置 (-1.0, -1.0, 0.0) m



再現音場の可視化②

- 音源信号：帯域制限パルス信号（ $\sim 2.6\text{kHz}$ ）
- 音源：スピーカ，位置 $(0.0, -1.0, 0.0)$ m，シフト 2.0m （y方向のみ）



音場再現技術の基本原理についてのまとめ

➤ 音場再現問題とその従来手法

- 平面／直線状アレイのための従来手法
(WFS, 多点制御法)
- WFSでは所望音場における音圧勾配の取得が必要

➤ 音場収音・再現のためのWFRフィルタ

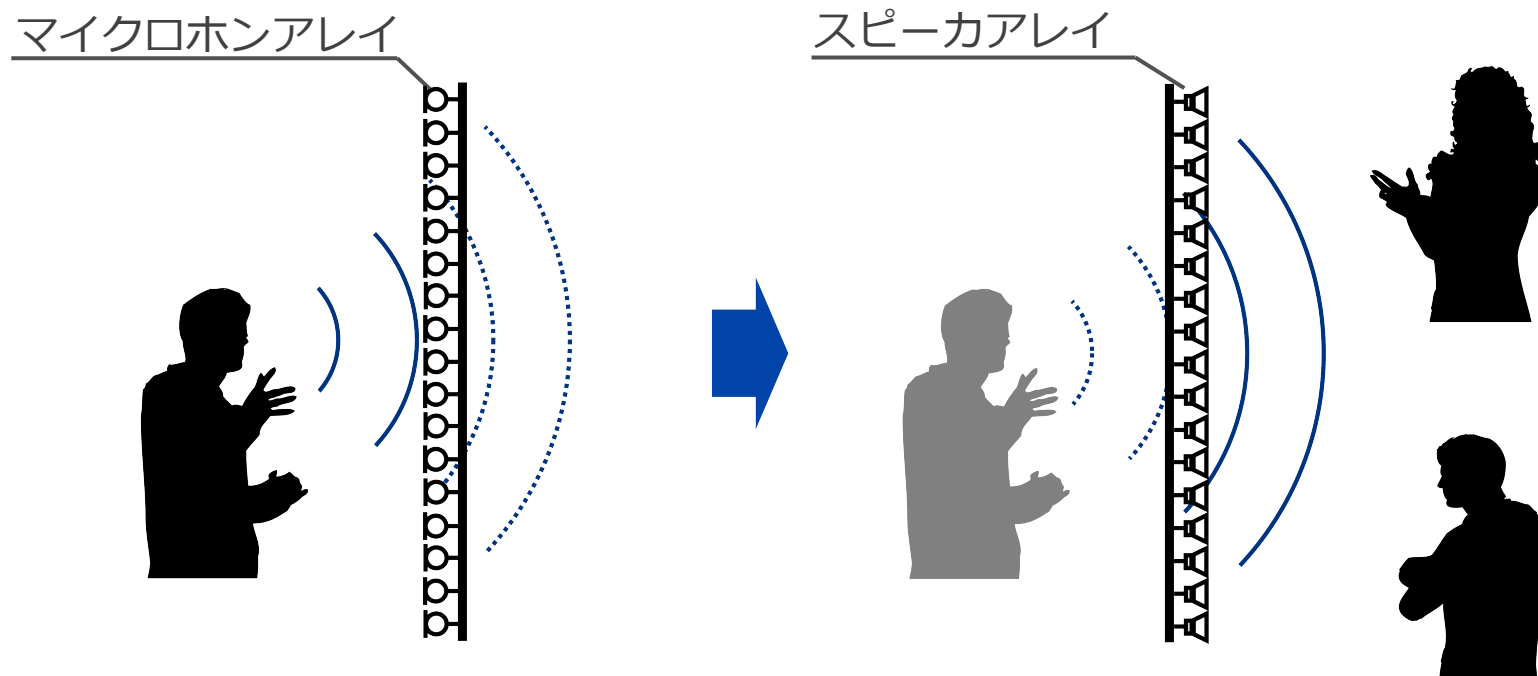
- 実環境で収音した音場を再現するため、音圧から音圧勾配を推定
- 変調／位相シフトによる再現位置の操作
- リアルタイム音場伝送システムと再現音場の可視化



超解像型音場再現に向けて

超解像型音場再現とは

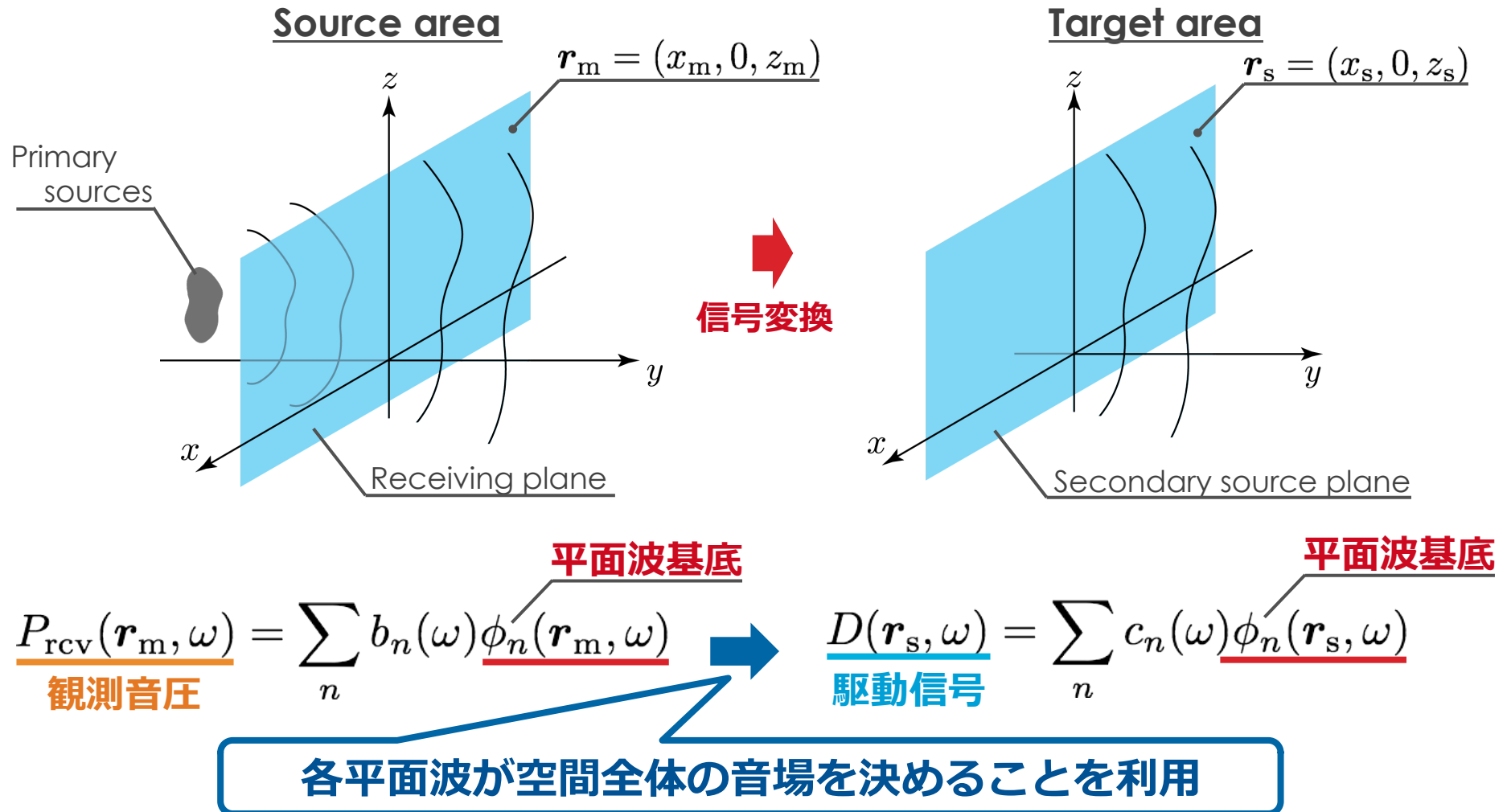
マイクロホン・スピーカ数が理想よりも少ない場合の再現精度を向上



- 空間エイリアシングの問題を回避 ⇒ 音色・定位感の劣化を改善
- マイクロホン数 > スピーカ数 [Ahrens, et al., AES Conv., 2010]
 - 局所的な再生領域で再現精度を向上
- マイクロホン数 < スピーカ数 [Koyama, et al., WASPAA, 2013], [Koyama, et al., ICASSP, 2014]
 - 局所的な収音領域の再現精度を向上

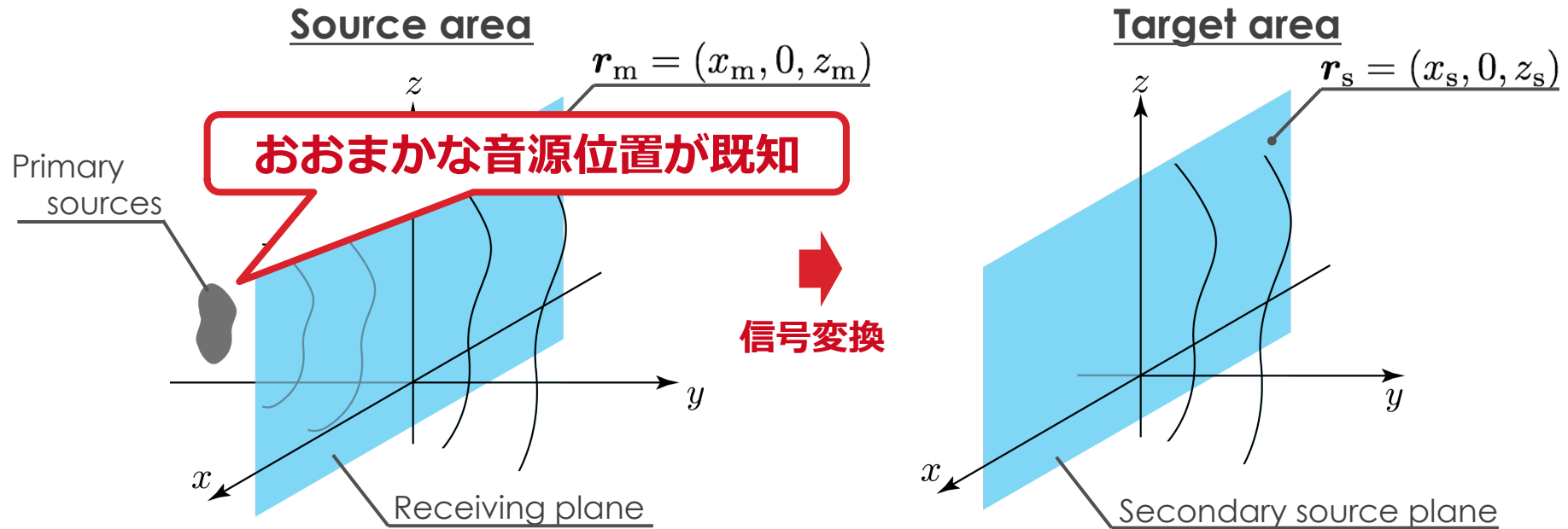
音源位置の事前情報を利用した超解像型再現

平面状アレイのためのWFRフィルタ



平面波に展開しているため、空間エイリアシングの問題が顕著
➡ マイク数 < スピーカ数の場合も高域での再現精度が劣化

音源位置事前情報を利用した基底展開



$$\underbrace{P_{\text{rcv}}(\mathbf{r}_m, \omega)}_{\text{観測音圧}} = \sum_n b_n(\omega) \phi_n(\mathbf{r}_m, \omega) \quad \rightarrow \quad \underbrace{D(\mathbf{r}_s, \omega)}_{\text{駆動信号}} = \sum_n c_n(\omega) \phi_n(\mathbf{r}_s, \omega)$$

音源位置の事前情報を利用し、
駆動信号を表現する基底を最適化

➡ 空間ナイキスト周波数以上での再現精度が改善

問題の定式化

Target area

Secondary source plane: Γ

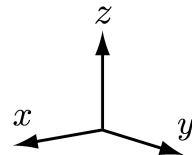
Control points of sound pressures

Virtual primary sources

$D(\mathbf{r}_s)$

$P_{\text{des}}(\mathbf{r}_{m,i})$

おおまかな位置を
センサ等で取得



音圧の観測値
(所望の音圧)

駆動信号が満たす制約式

$$\underline{P^{\text{des}}}(\omega) = \int_{\Gamma} \underline{D(\mathbf{r}_s, \omega)} \underline{G(\mathbf{r}_s, \omega)} d\Gamma + \epsilon$$

伝達関数 誤差

空間的な基底による表現

音源位置の事前情報に基づいて
 $\phi(\mathbf{r}_s)$ と c を最適化

$$\underline{\hat{D}(\mathbf{r}_s, \omega)} = \sum_{n=1}^N c_n(\omega) \underline{\phi_n(\mathbf{r}_s, \omega)} = \underline{\phi^T(\mathbf{r}_s, \omega)} c(\omega)$$

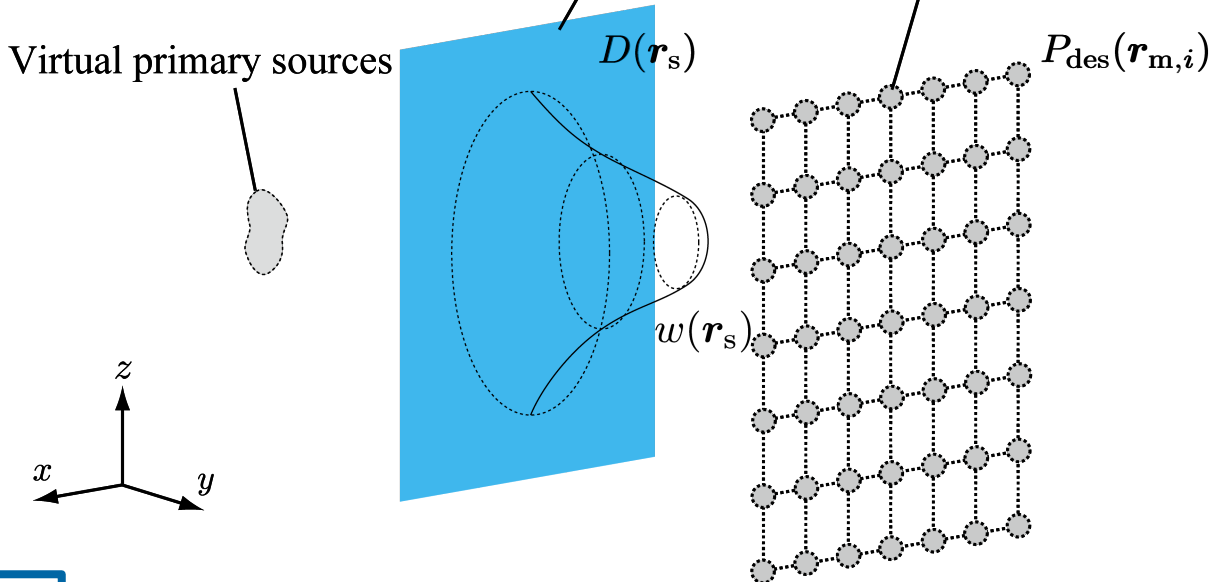
MAP推定に基づく最適化

Target area

Secondary source plane: Γ

Control points of sound pressures

Virtual primary sources



事後確率最大化

$$\left(\hat{c}, \hat{\phi}(\mathbf{r}_s) \right) = \arg \max_{\mathbf{c}, \phi(\mathbf{r}_s)} \pi \left(D(\mathbf{r}_s) = \phi^T(\mathbf{r}_s) \mathbf{c} | \mathbf{P}^{\text{des}} \right)$$

Bayes' rule

$$\pi \left(D(\mathbf{r}_s) | \mathbf{P}^{\text{des}} \right) \propto \underbrace{\pi \left(\mathbf{P}^{\text{des}} | D(\mathbf{r}_s) \right)}_{\text{尤度関数}} \underbrace{\pi \left(D(\mathbf{r}_s) \right)}_{\text{事前分布}}$$

駆動信号 $D(\mathbf{r}_s)$ のおおよその振幅分布 $w(\mathbf{r}_s)$ を事前分布として与える

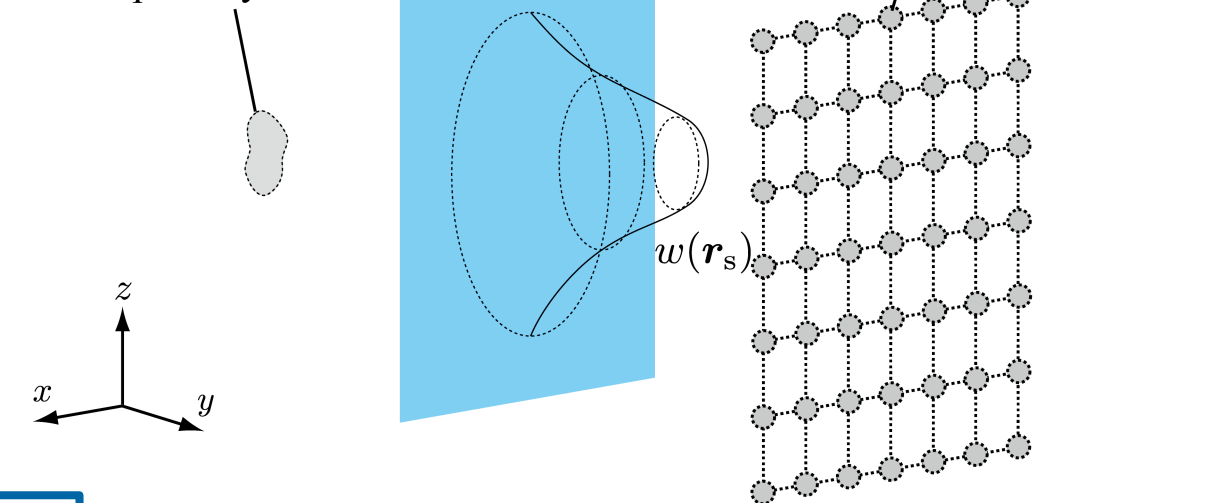
MAP推定に基づく最適化

Target area

Secondary source plane: Γ

Control points of sound pressures

Virtual primary sources



事前分布の定式化

共分散関数

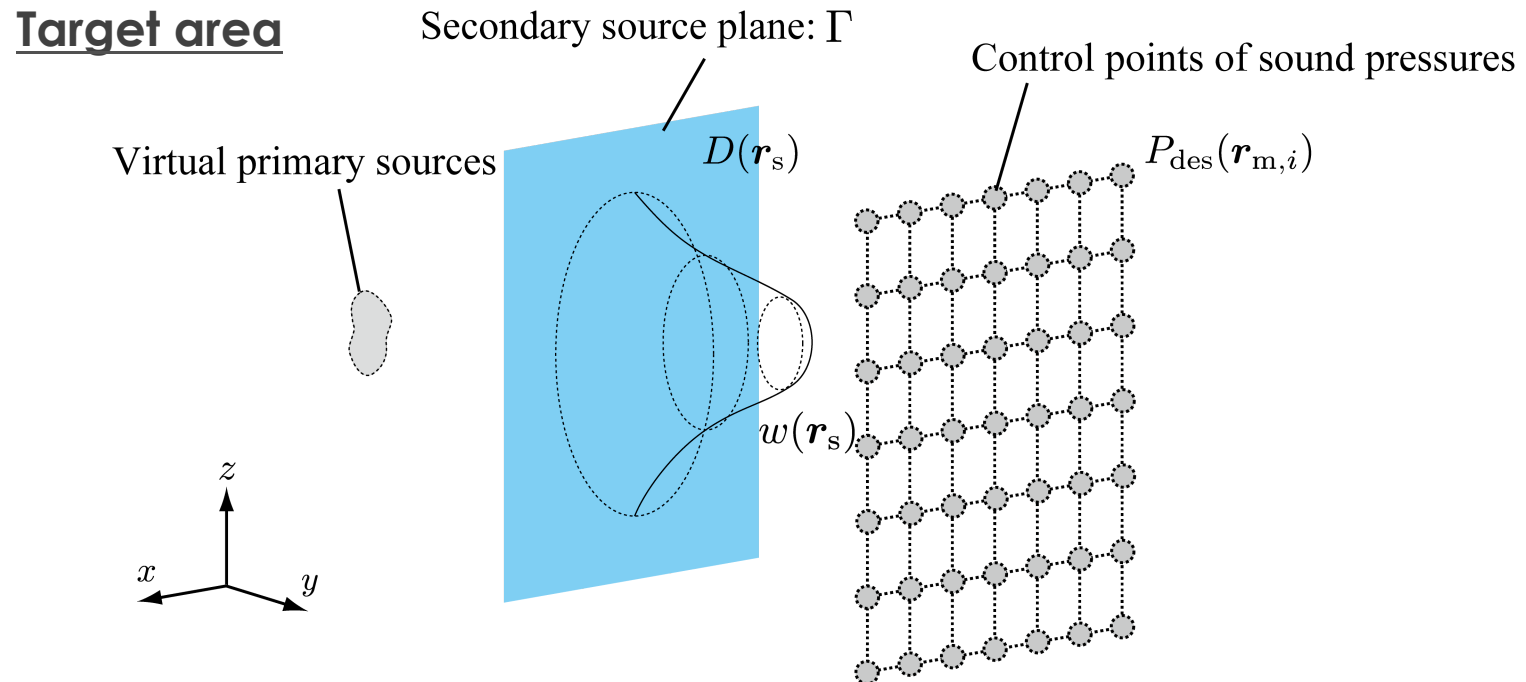
$$E[D(\mathbf{r}_s)D(\mathbf{r}'_s)^*] = \sigma_d^2 \underline{w(\mathbf{r}_s)} \delta(\mathbf{r}_s - \mathbf{r}'_s)$$

正規化した振幅分布

事前分布

$$\underline{\pi(D(\mathbf{r}_s))} \propto \exp\left(-\frac{1}{\sigma_d^2} \int_{\Gamma} |D(\mathbf{r}_s)|^2 \underline{w^{-1}(\mathbf{r}_s)} d\Gamma\right)$$

振幅分布の与え方



WFSに基づけば, $D(\mathbf{r}_s)$ は Γ 上の音圧勾配にほぼ一致するはず

[Spors, et al, AES Conv., 2008]

➡ 音源位置 $\mathbf{r}_{\text{ps}}$ からの球面波伝播を仮定すると, $D(\mathbf{r}_s)$ の振幅分布は

$$\begin{aligned} \check{w}(\mathbf{r}_s) &= \left| 2 \frac{\partial P_{\text{des}}(\mathbf{r}_s)}{\partial y_s} \right|^2 \\ &\simeq \left| (y_s - y_{\text{ps}}) \frac{jk|\mathbf{r}_s - \mathbf{r}_{\text{ps}}| - 1}{2\pi|\mathbf{r}_s - \mathbf{r}_{\text{ps}}|} e^{jk|\mathbf{r}_s - \mathbf{r}_{\text{ps}}|} \right|^2 \end{aligned}$$

正規化 ➡ $w(\mathbf{r}_s) = \check{w}(\mathbf{r}_s) / \int_{\Gamma} \check{w}(\mathbf{r}_s) d\Gamma$

MAP推定に基づく最適化

以下の特異値分解の関係が成り立つ

$$\underline{w(\mathbf{r}_s)} \mathbf{G}^H(\mathbf{r}_s) = \phi^T(\mathbf{r}_s) \mathbf{\Lambda} \mathbf{U}^H$$



基底 $\hat{\phi}(\mathbf{r}_s) = \underline{w(\mathbf{r}_s)} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{G}^*(\mathbf{r}_s)$

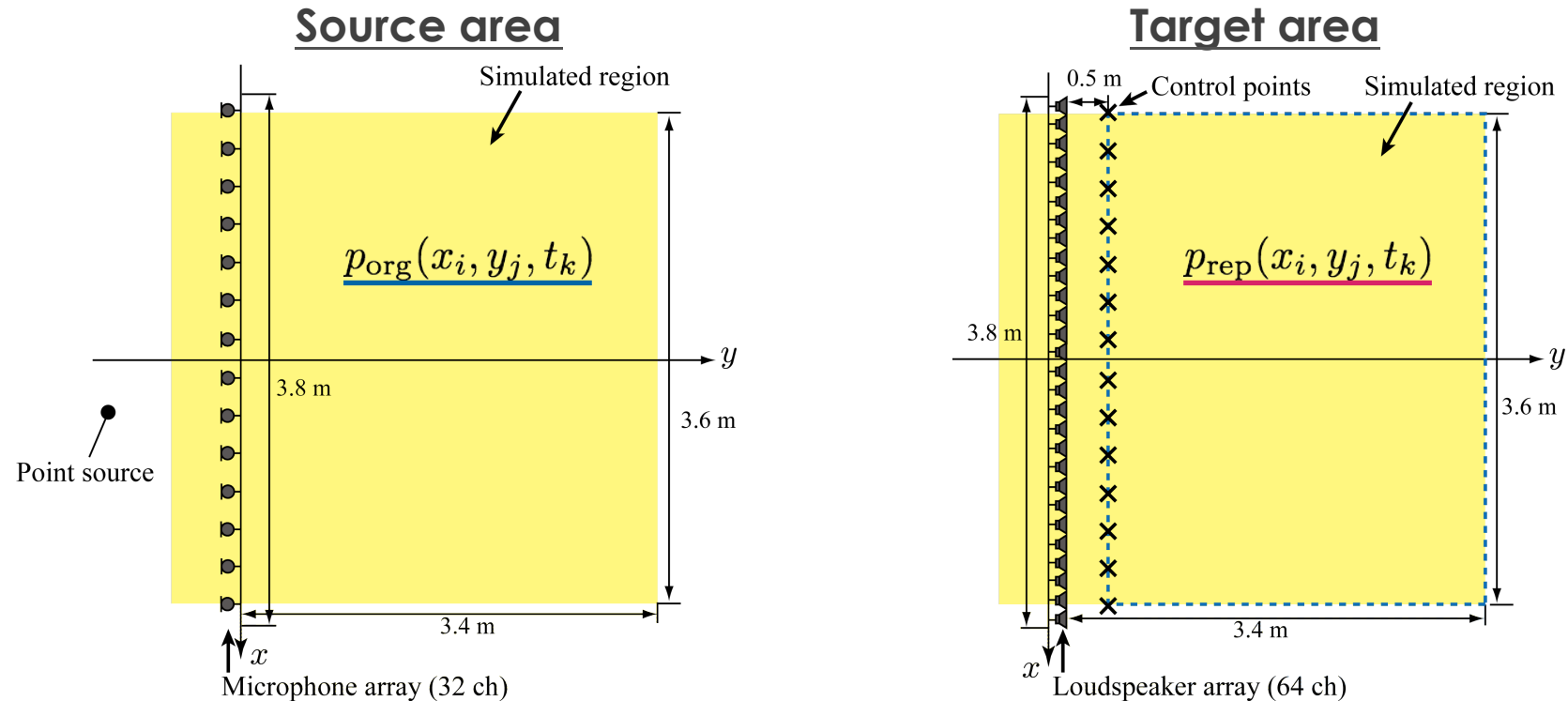
係数 $\hat{\mathbf{c}} = (\alpha \mathbf{I} + \mathbf{\Lambda}^2)^{-1} \mathbf{\Lambda} \mathbf{U}^H \underline{\mathbf{P}^{\text{des}}}$

➡ $\underline{\hat{D}(\mathbf{r}_s)} = \phi^T(\mathbf{r}_s) \mathbf{c}$ より

$$\underline{\hat{D}(\mathbf{r}_s)} = \underline{w(\mathbf{r}_s)} \mathbf{G}^H(\mathbf{r}_s) \mathbf{U} (\alpha \mathbf{I} + \mathbf{\Lambda}^2)^{-1} \mathbf{U}^H \underline{\mathbf{P}^{\text{des}}}$$

(α : 正則化パラメータ)

シミュレーション実験

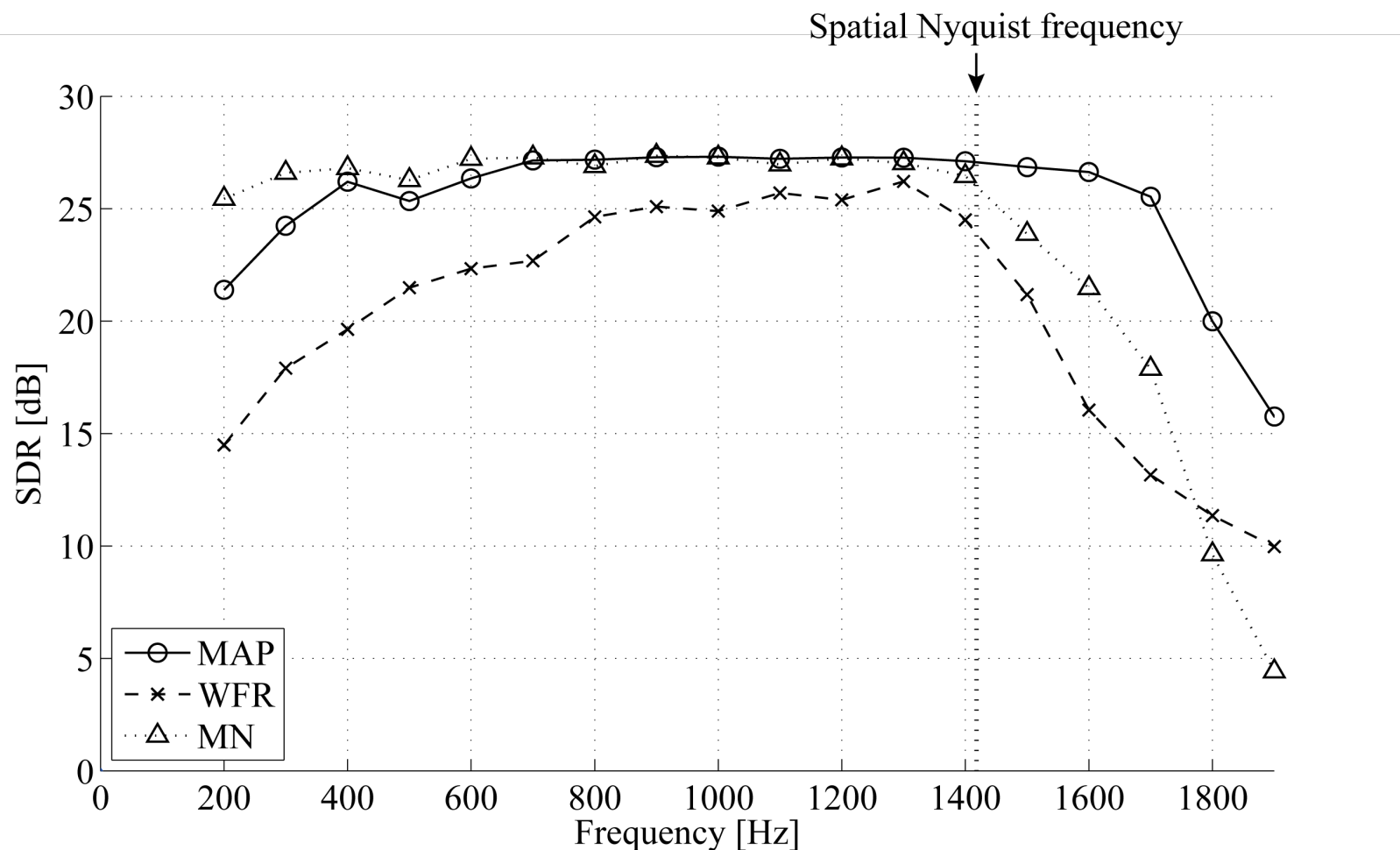


- 提案法 (**MAP**) , WFRフィルタに基づく手法 (**WFR**) , 最小ノルム解に基づく手法 (**MN**) を比較
- Signal-to-distortion-ratio (SDR)による評価

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

Frequency vs. SDR

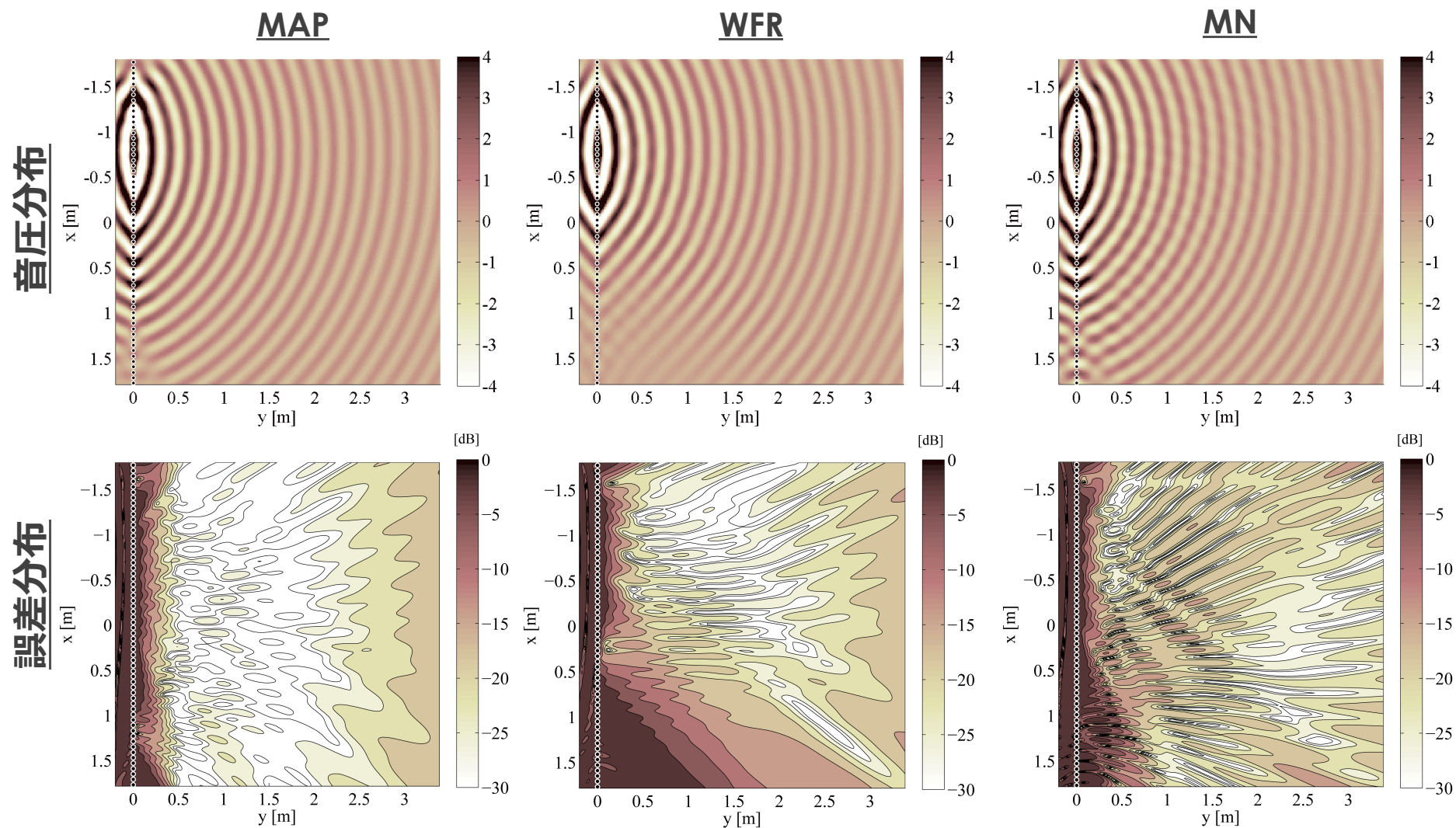
位置 (-0.8, -1.0, 0.0)m の点音源



提案法では空間ナイキスト周波数以上で再現精度が保たれる

音圧分布の比較

位置 $(-0.8, -1.0, 0.0)$ m の点音源, 1.7 kHz正弦波



まとめ

- 音場再現技術の基本原理
 - 音場再現問題とその従来手法
 - 音場収音・再現のためのWFRフィルタ
 - リアルタイム音場伝送システムの実装例
- 超解像型音場再現に向けて
 - 空間エイリアシングによる音質・定位感の劣化を改善
 - 音源位置事前情報を用いた超解像型音場再現

今後の展開

➤ 超解像型音場再現

- 今回紹介した手法はあくまで一例
- 収音側の音場を適応的に高精度化するアルゴリズムや、マイクロホン数>スピーカ数の場合にも適用可能な手法が必要

➤ 聴覚的な観点からの最適化

- 現状は物理的な指標による客観評価
- 臨場感を評価することは難しい。。
- 聴覚的に空間解像度を最適化？