

信号処理論特論 第5回 (10/25)

情報理工学系研究科システム情報学専攻

猿渡 洋 (資料作成協力: 小山翔一 先生)

hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

講義予定

- 9/27: 第1回 統計的音声音響信号処理概論
- 10/04: 第2回 非負値行列因子分解
- 10/11: 第3回 独立因子分析(ICA, IVA, ILRMA)
- 10/18: 第4回 独立因子分析(続き)
- 10/25: 第5回 音場再現・スパース最適化
- 11/01: 第6回 エンハンスメント・高次統計量解析
- 11/15: 第7回 【レポート課題1】
- 11/22: 第8回 話者認識
- 11/29: 休講
- 12/06: 第9回 音声合成・変換1
- 12/13: 休講
- 12/20: 第10回 音声合成・変換2
- 01/10: 第11回 【レポート課題2】

} 入れ替えました

講義資料と成績評価

■ 講義資料

- <http://www.sp.ipc.i.u-tokyo.ac.jp/~saruwatari/>

(システム情報第一研究室からたどれるようになってます)

■ 成績評価

- 出席点
- レポート点(2回の提出が必須)

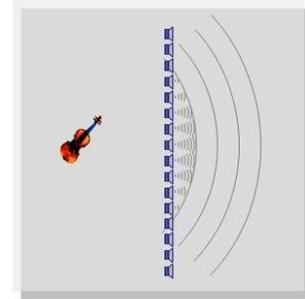
音場再現

高臨場音響再生技術の動向

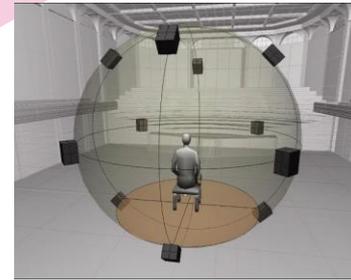
受聴領域の大きさ

広い受聴領域と高い空間解像度を目指した物理的な音場再現へ

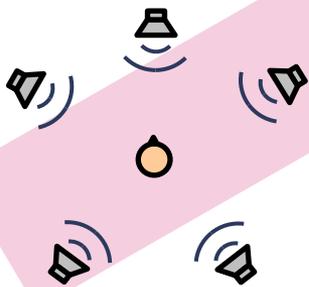
Wave Field Synthesis



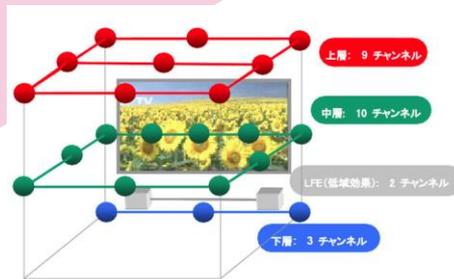
Higher Order Ambisonics



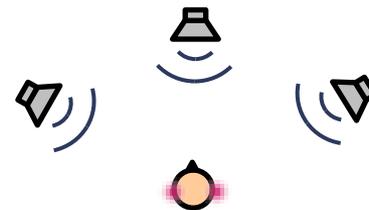
5.1ch サラウンド



22.2ch サラウンド



トランスオーラル



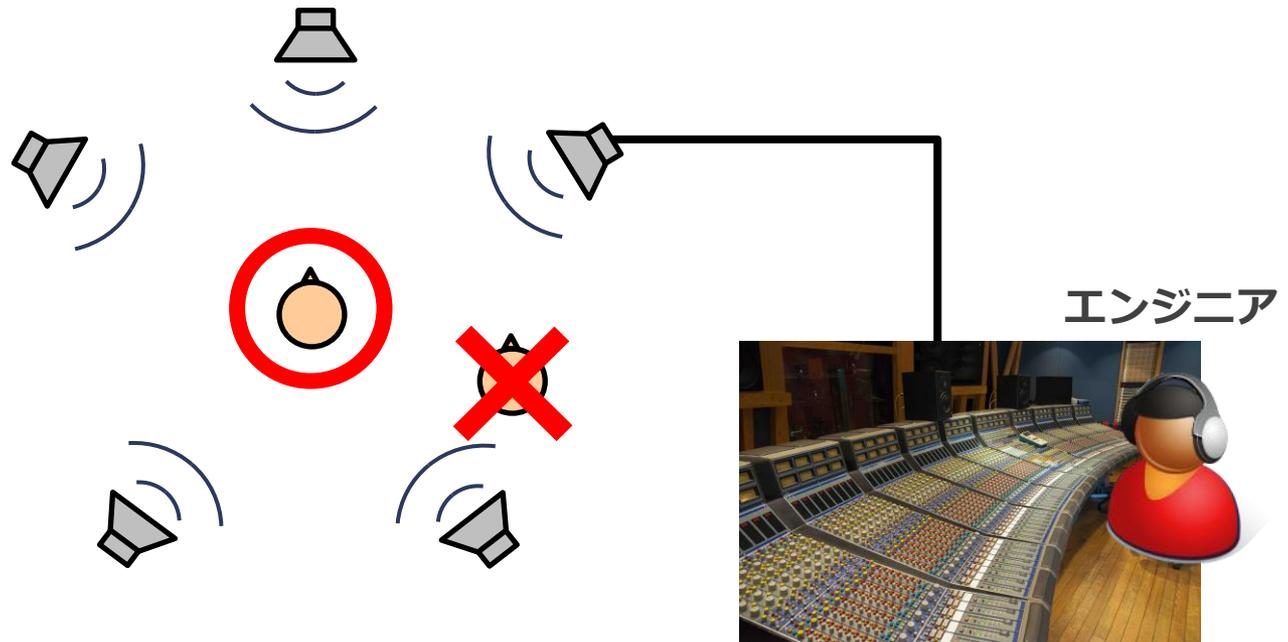
バイノーラル



空間解像度

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

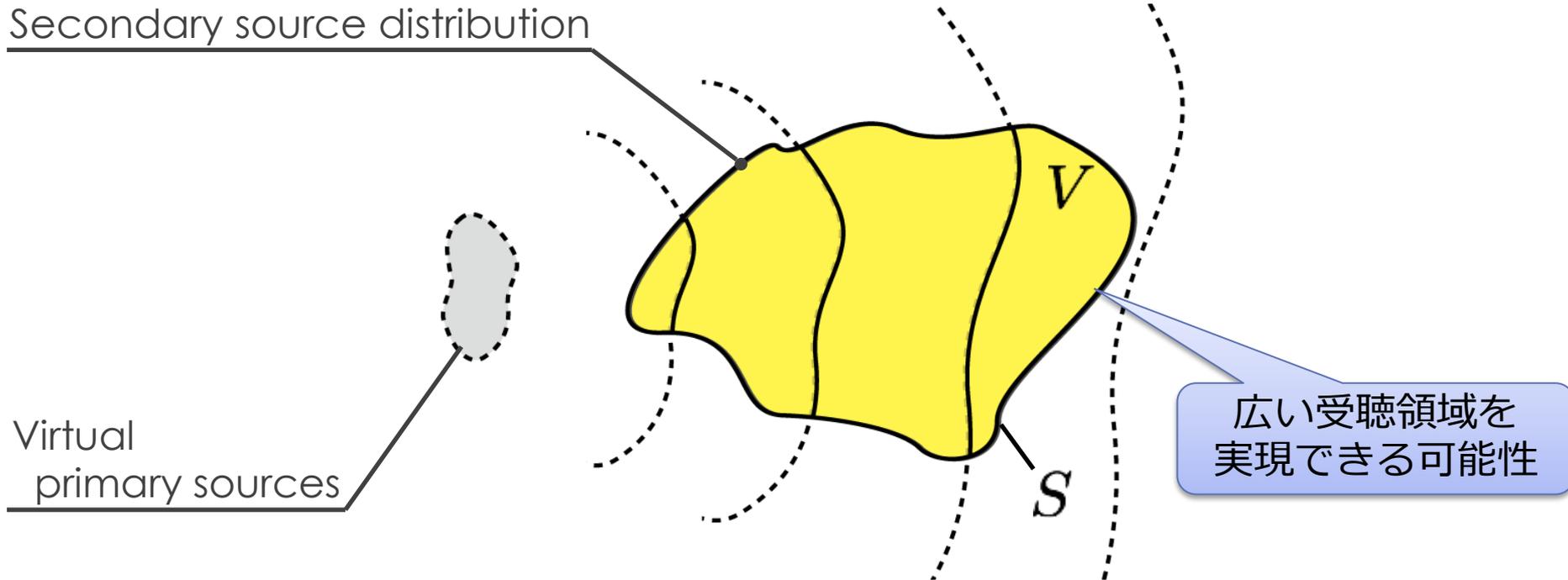


欠点

- 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)
- 音をデザインする人が必要 (あくまで人工的な音の表現)

音場再現による高臨場音響再生

音場そのものを物理的に再現



- 対象領域 V 内の音場を，境界面 S 上に配置した二次音源（=スピーカ）を用いて，所望の音場と一致させる

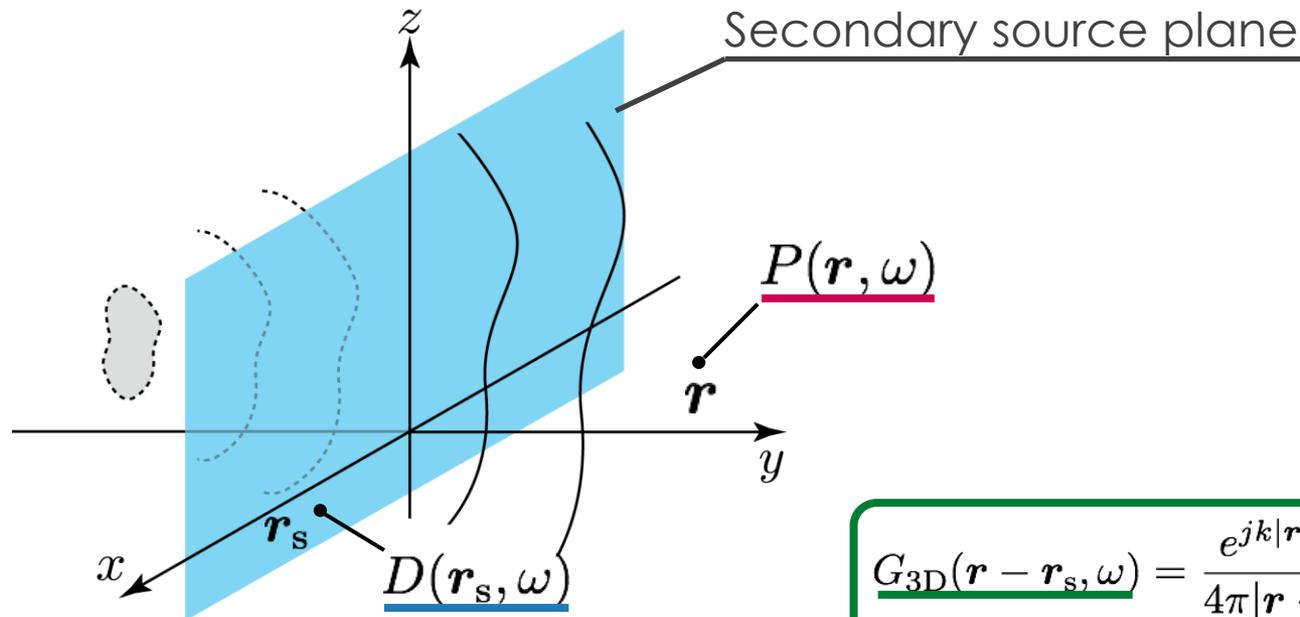
➡ 音場再現問題

音場再現手法の比較

	アレイ配置	概要	収録音場の再現
Wave Field Synthesis (WFS)	平面／直線	Kirchhoff-Helmholtz積分／Rayleigh積分に基づくスピーカ駆動信号	×
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 (e.g., 境界音場制御)	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) フィルタ法	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	○

Wave Field Synthesis (WFS)

[Berkhout+ JASA 1993]
[Spors+ AES Conv 2008]



第1種Rayleigh積分に基づく所望音場

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

二次音源によって合成される音場

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

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

各二次音源が点音源と近似
($G \simeq G_{3D}$)すると,

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

➡ 音圧勾配と一致

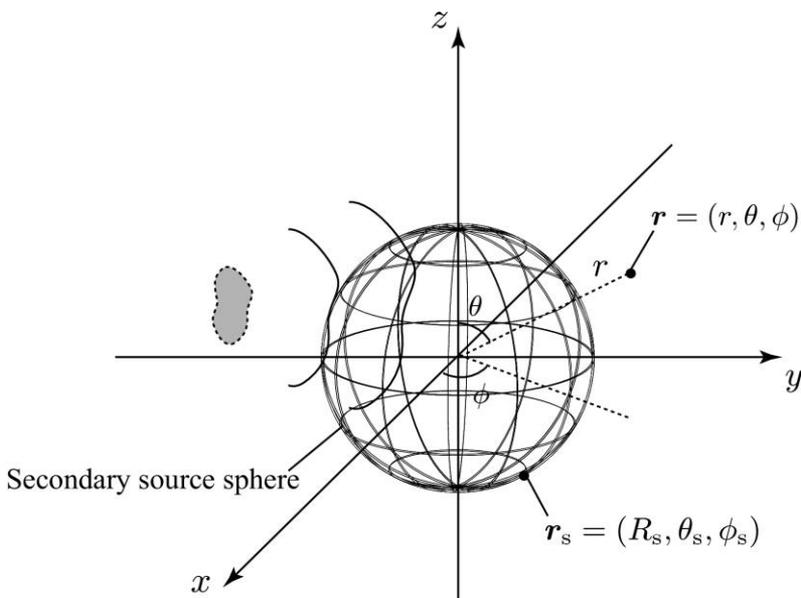
境界面上の音圧勾配を二次音源の駆動信号とすればよい

Higher Order Ambisonics (HOA)

[Daniel AES Conf 2003]

[Poletti JAES 2005]

- スピーカ中心を原点とする球面調和スペクトル領域で合成音場が所望音場と一致するように制御



$$\underbrace{\sum_{l=1}^L \frac{D_l(\omega) Y_n^m(\theta_{s,l}, \phi_{s,l})^*}{\text{球面調和関数}}}_{\text{合成音場の球面調和スペクトル}} = \underbrace{\frac{A_n^m(\omega)}{jkh_n(kR_s)}}_{\text{所望音場の球面調和スペクトル}}$$

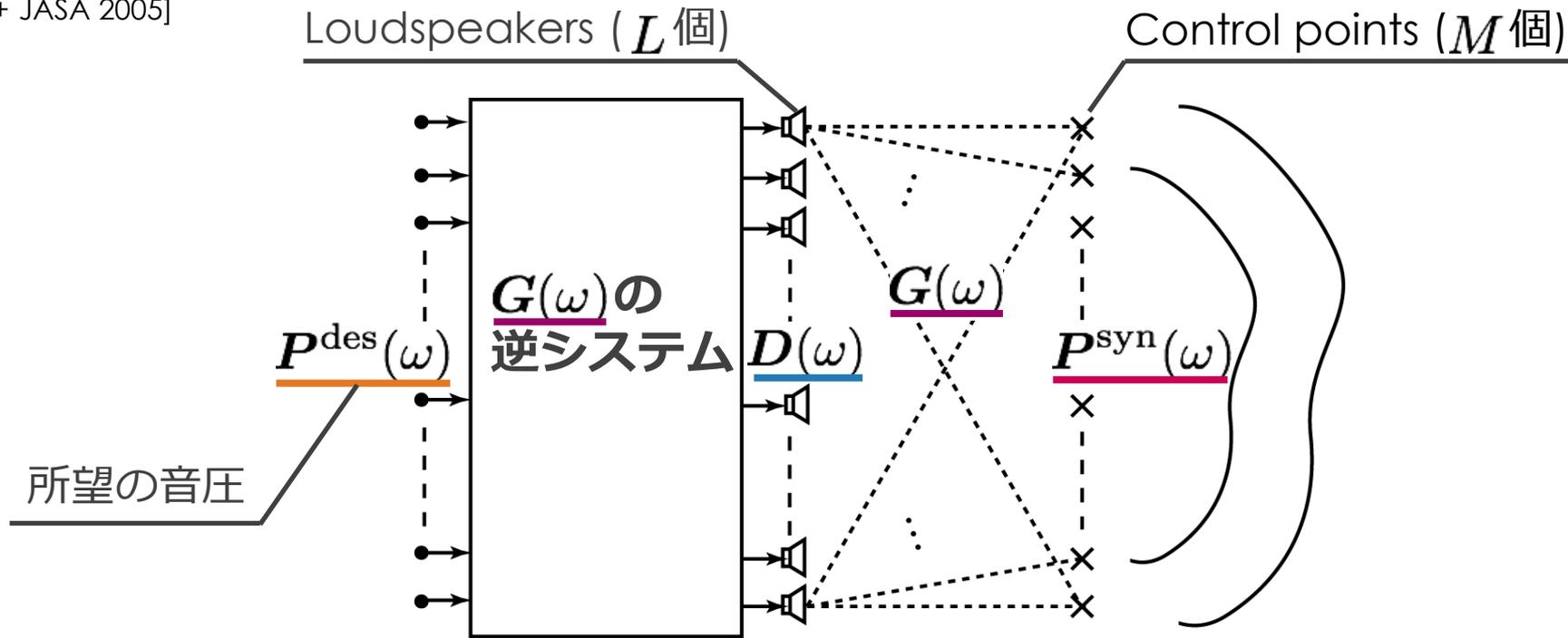
➡ 球面調和関数を要素にもつ行列の（一般化）逆行列を用いて駆動信号を得る

- 收音側では球状マイクロフォンアレイを用いて所望の $A_n^m(\omega)$ を取得することになるが、アレイ半径で決まる**禁止周波数**への対処が必要

球状アレイを用いることで全方位の音場を再現可能

逆フィルタに基づく手法

[Gauthier+ 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

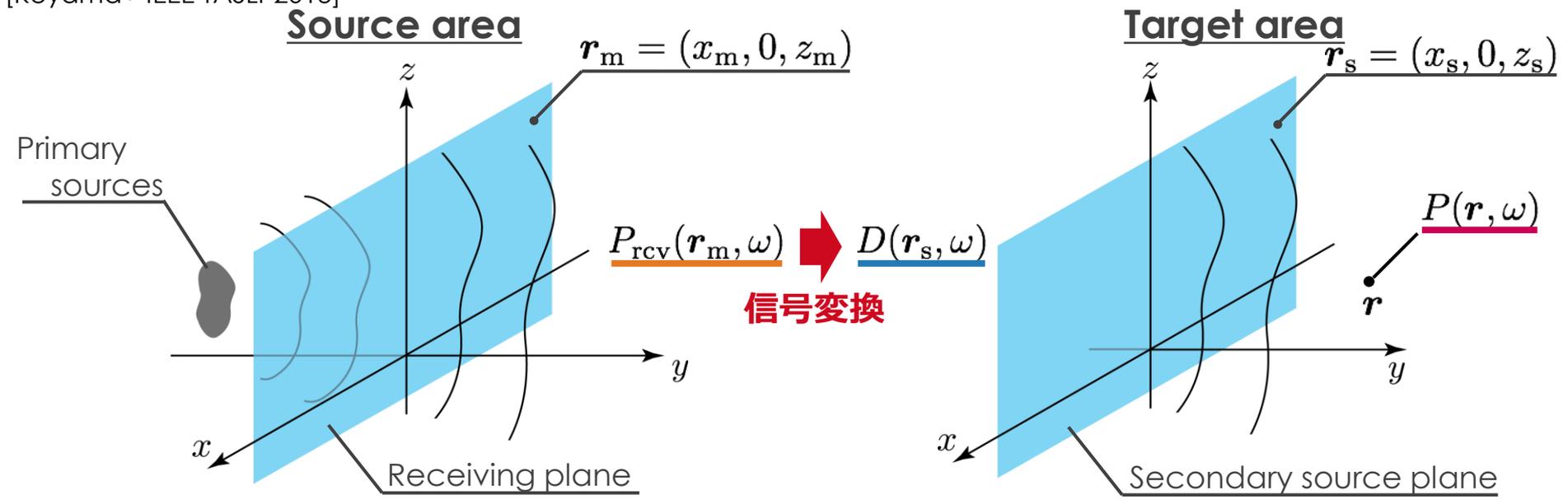
$$\hat{D}(\omega) = (\underline{G^H(\omega)G(\omega)} + \beta I)^{-1} \underline{G^H(\omega)P^{des}(\omega)}$$

正則化パラメータ

逆フィルタの設計さえできれば任意のレイ形状に適用可能

波面再構成 (WFR) フィルタ法

[Koyama+ IEEE TASLP2013]



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,

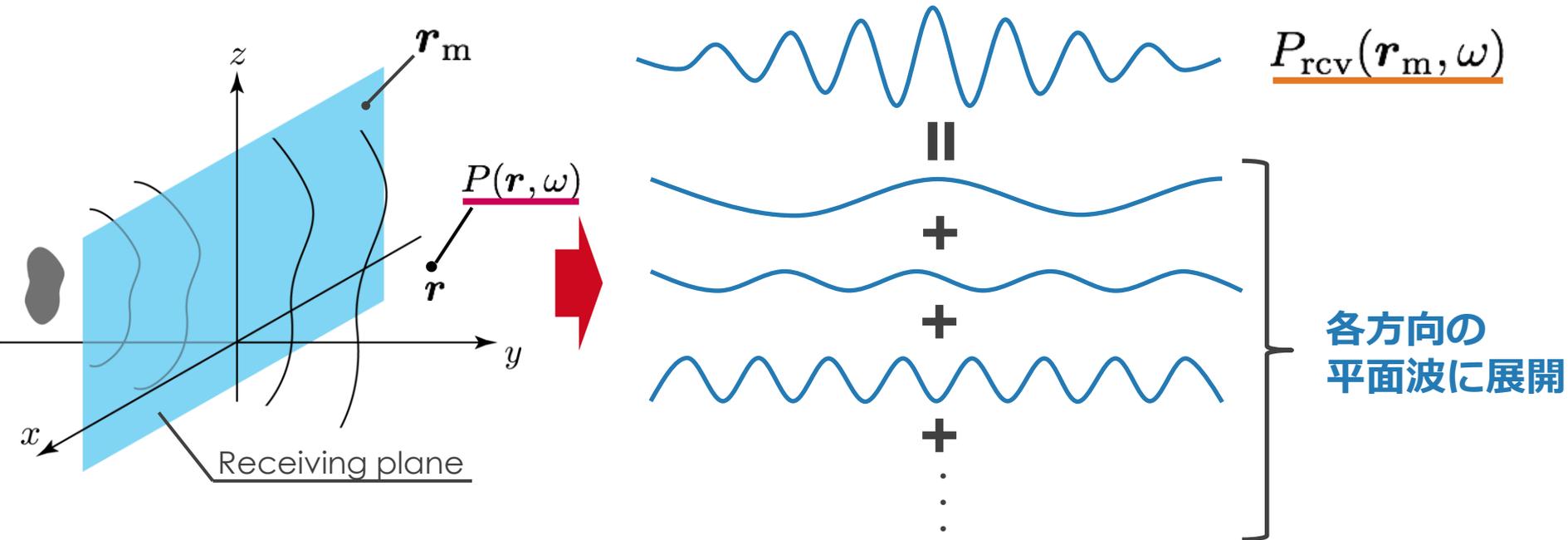
$$D(\mathbf{r}_s, \omega) = -2 \frac{\partial P_{\text{rcv}}(\mathbf{r}_m, \omega)}{\partial y_m}$$

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

収録音場の再現には、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

受音平面上的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

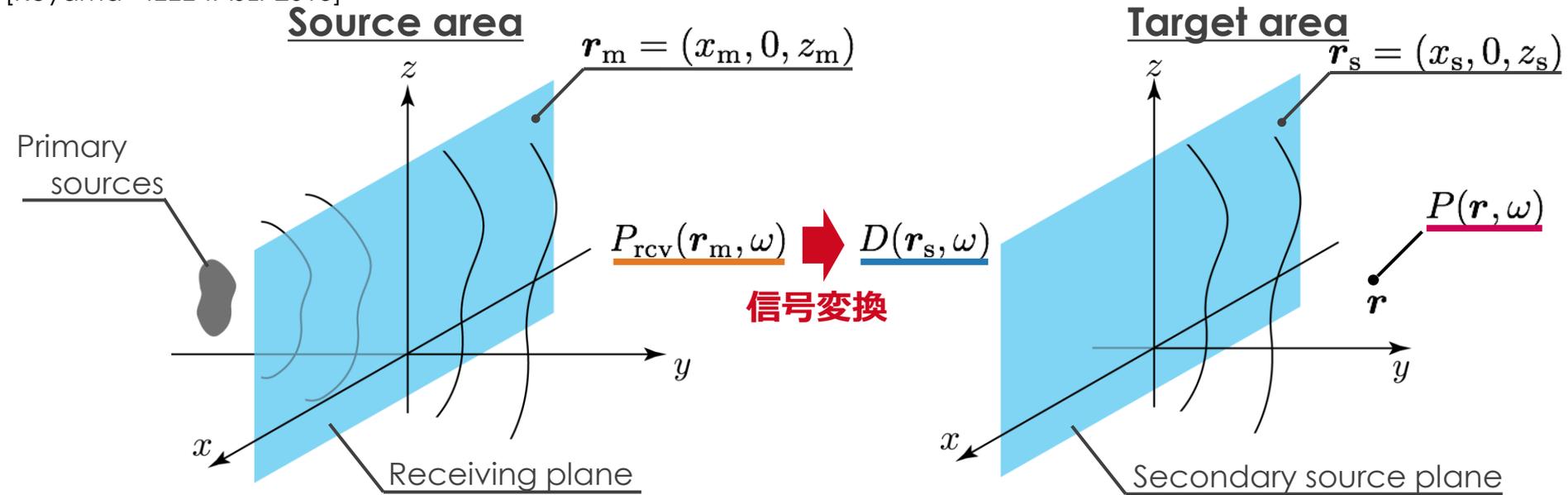
$$P(\mathbf{r}, \omega) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{P}_{\text{rcv}}(k_x, 0, k_z, \omega) e^{j(k_x x + k_y y + k_z z)}$$

受音平面上的
空間スペクトル

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

波面再構成 (WFR) フィルタ法

[Koyama+ IEEE TASLP2013]



第1種Rayleigh積分に基づく所望音場の空間スペクトル

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

二次音源によって合成される音場の空間スペクトル

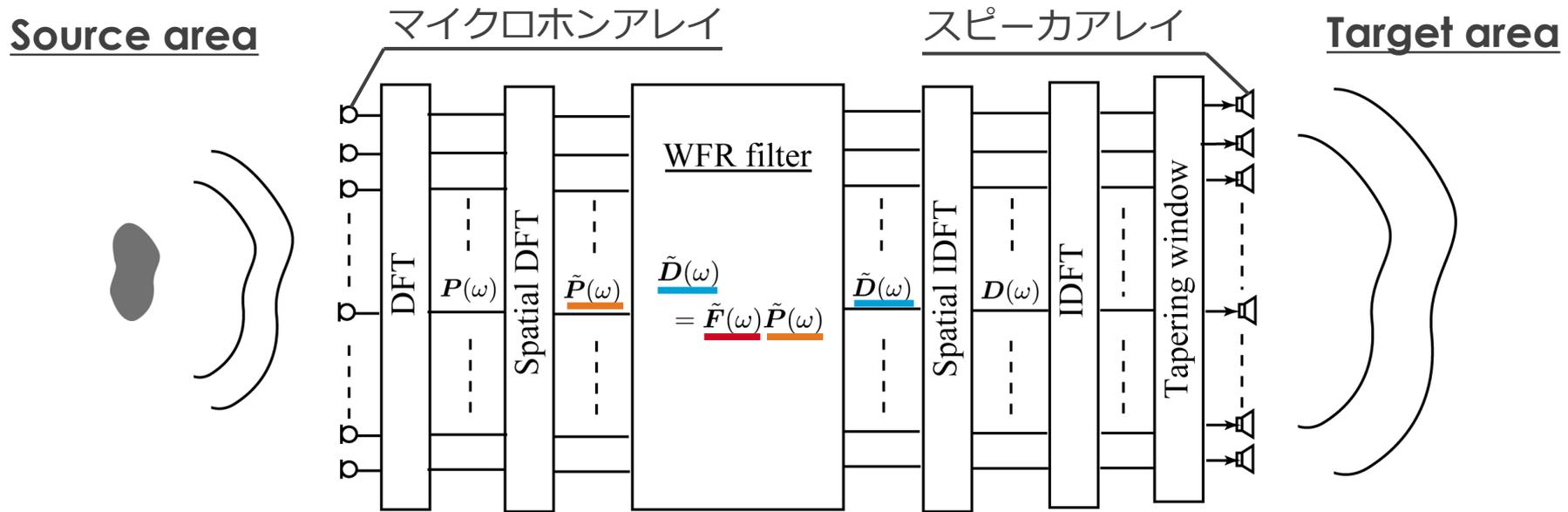
$$\tilde{P}_{\text{syn}}(k_x, y, k_z, \omega) = \tilde{D}(k_x, k_z, \omega) \cdot \tilde{G}(k_x, y, k_z, \omega)$$

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

空間スペクトル領域での定式化による信号変換

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：波面再構成 (WFR) フィルタ



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

スピーカ駆動信号の時空間スペクトル

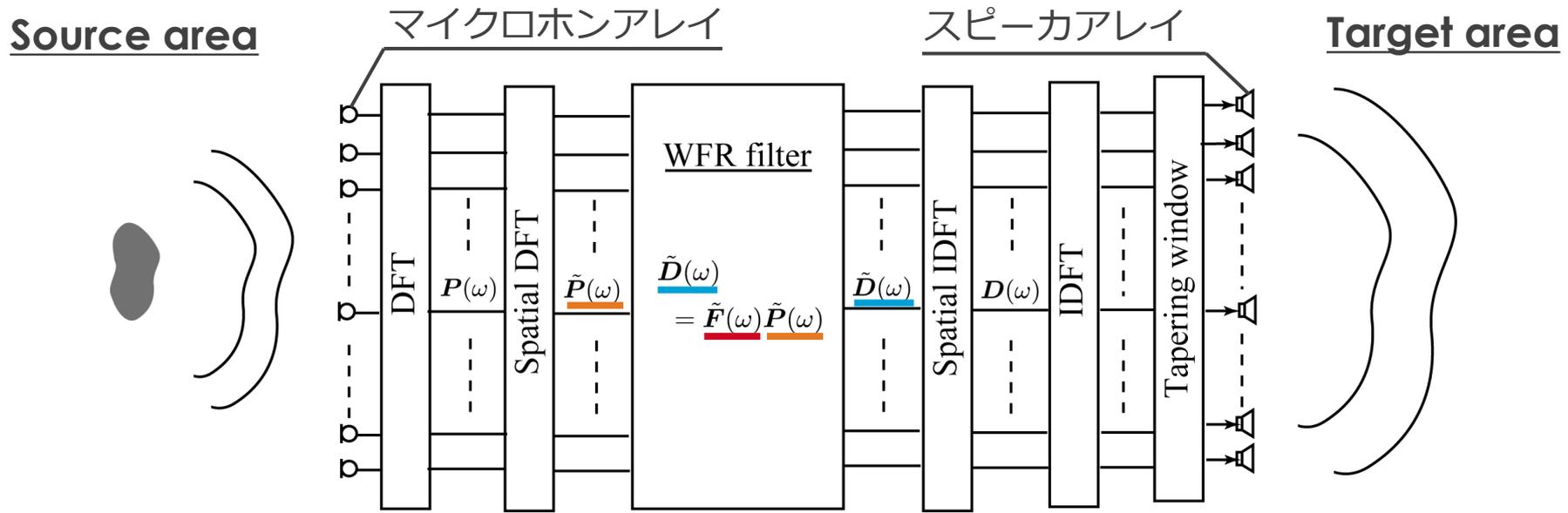
マイクロホン信号の時空間スペクトル

WFRフィルタ：各要素が以下で決まる対角行列

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

離散化・有限長での打ち切りの影響

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**



(Δx : 素子間隔)

➤ 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

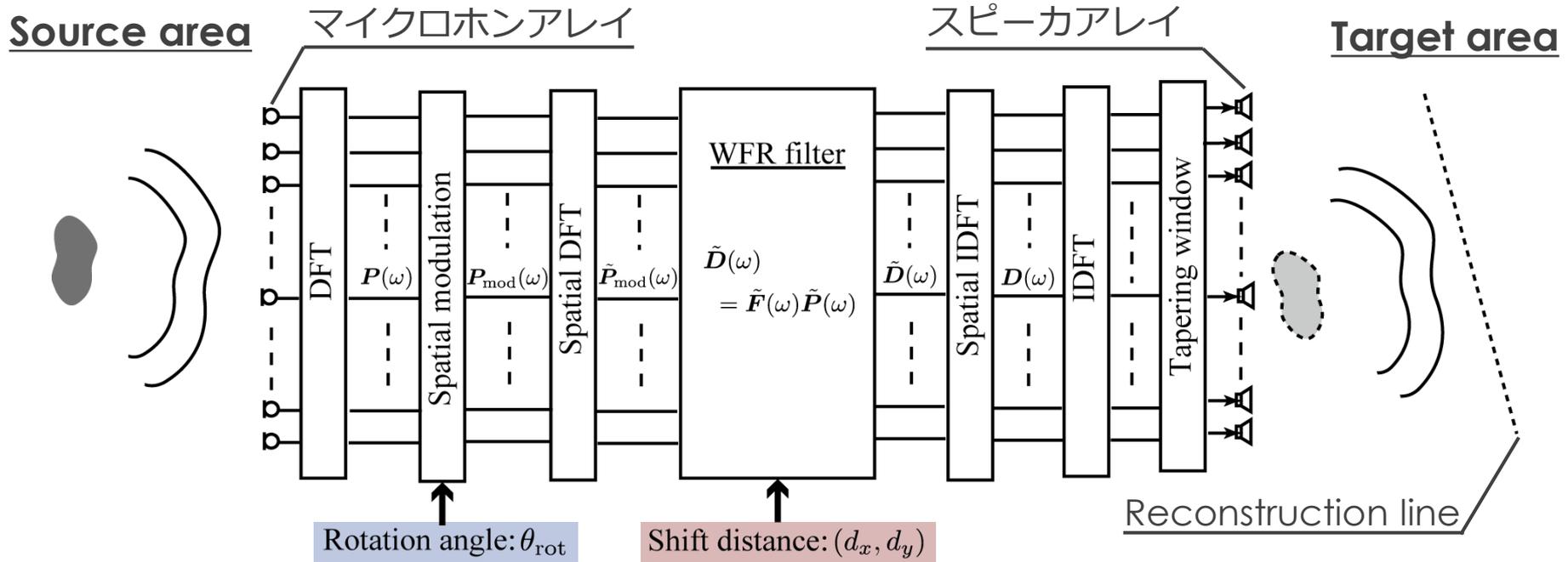
➤ 有限長での打ち切りの影響

- 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

直線状アレイのための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



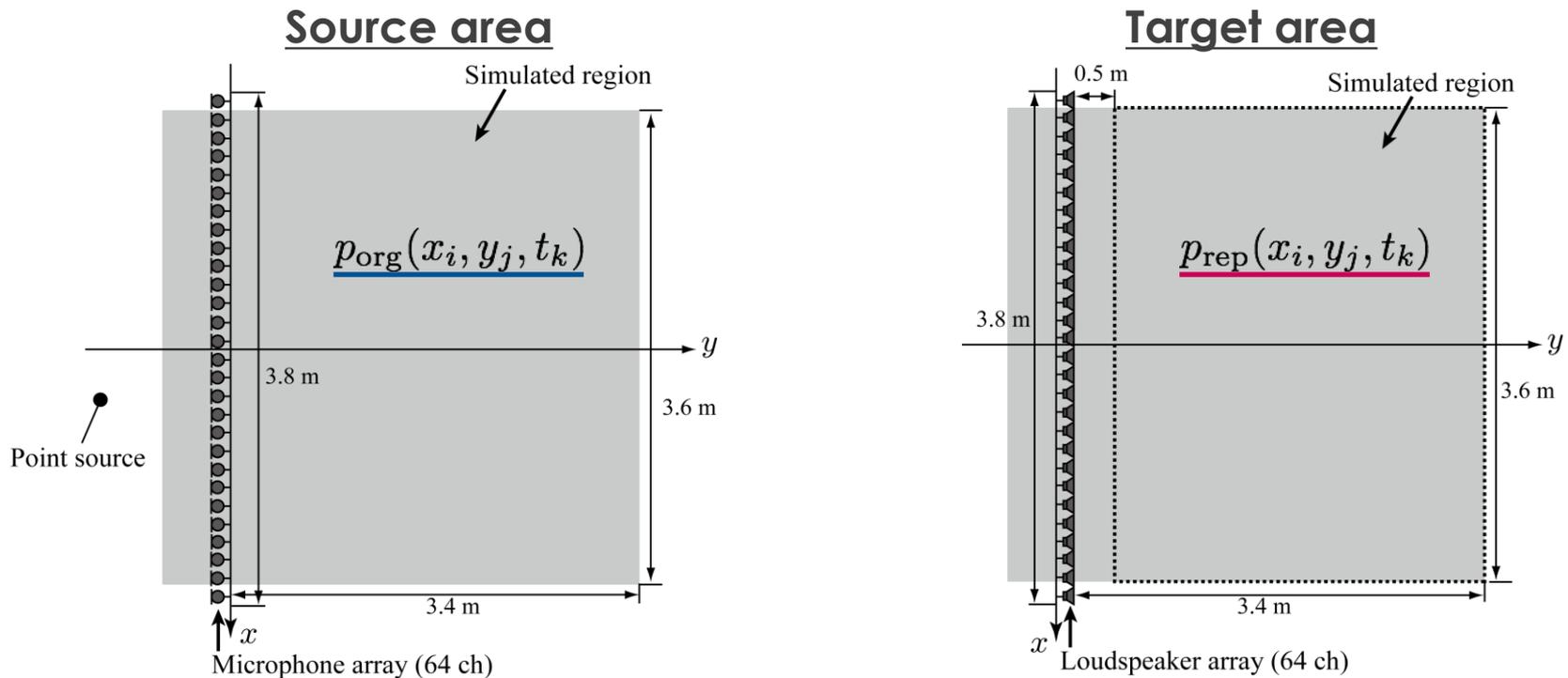
変調

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

位相シフト

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

シミュレーション実験



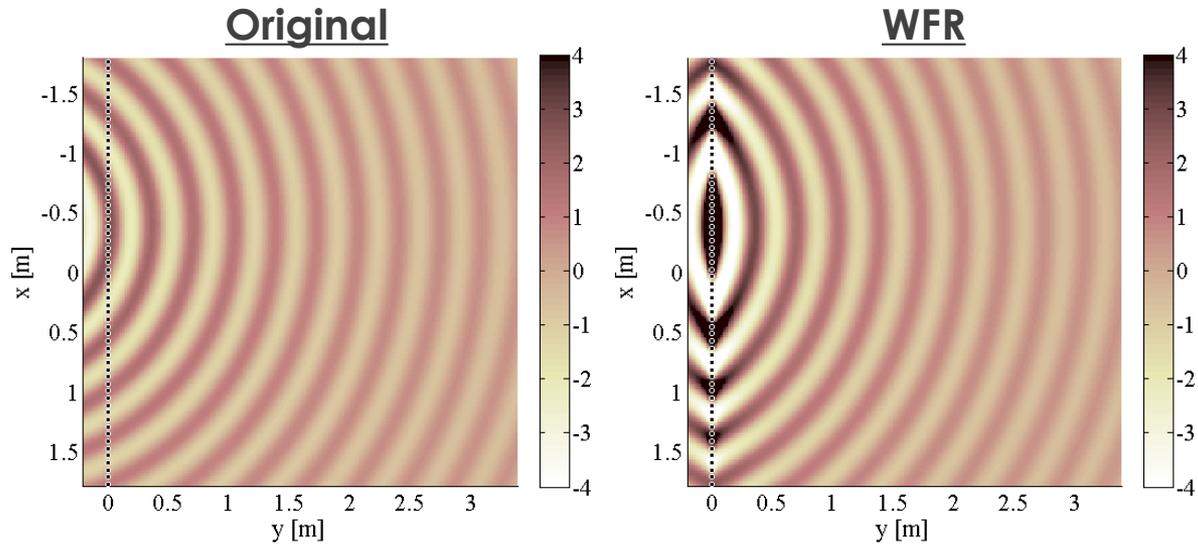
- マイクロホン・スピーカ素子数：64，素子間隔：6 cm
- 3.6 x 3.6mの領域で1.5cm間隔の音圧分布を計算（中心で正規化）
- WFRフィルタ長（空間方向）：128，受聴直線位置：1 m
- 誤差分布：

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

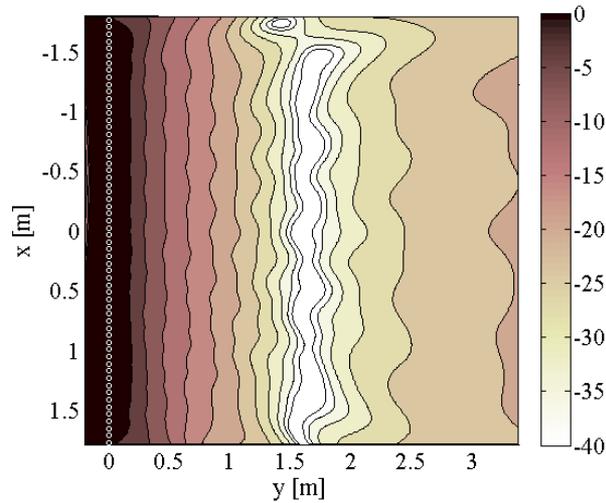
音圧分布の比較 (1 kHz)

位置 $(-0.4, -1.0, 0.0)$ m の点音源, 1 kHz正弦波

音圧分布



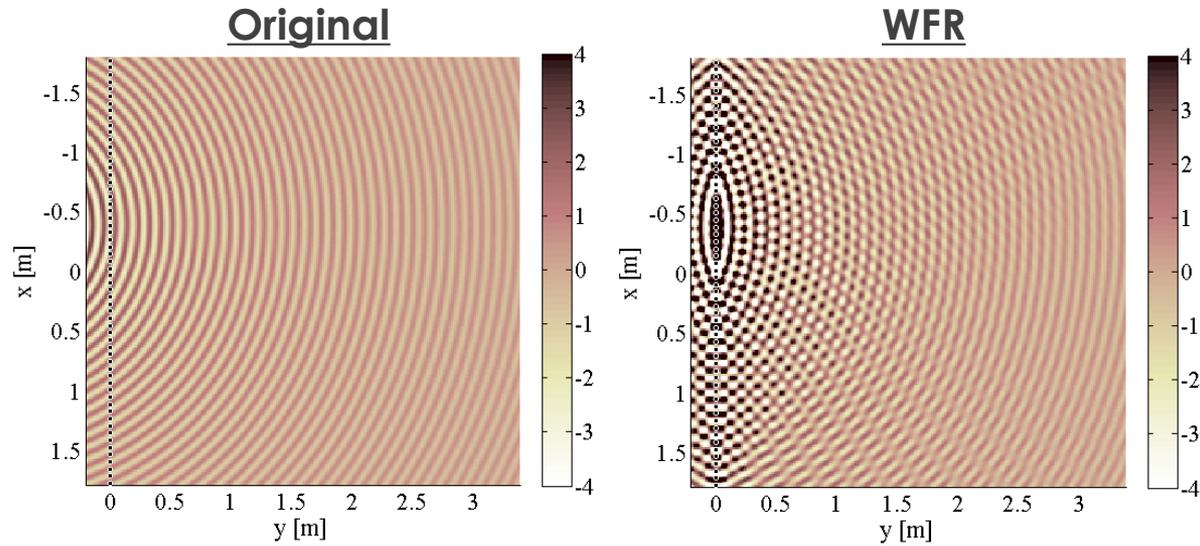
誤差分布



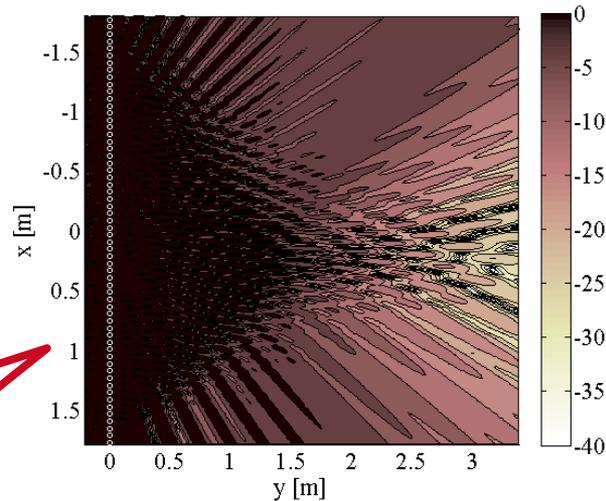
音圧分布の比較 (3.6kHz)

位置 (-0.4, -1.0, .0.0)m の点音源, 3.6 kHz正弦波

音圧分布



誤差分布

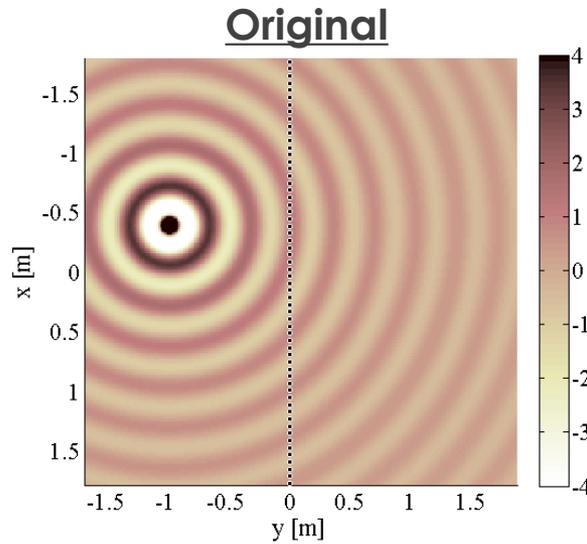


高い周波数帯域では
空間エイリアシングによる誤差が発生

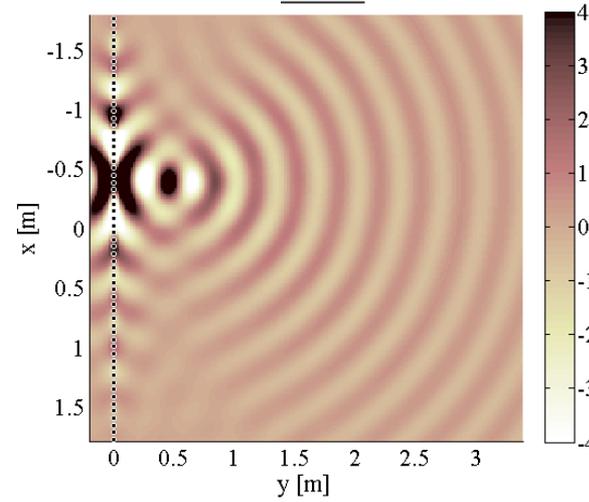
シミュレーション実験 (1kHz)

位置 $(-0.4, -1.0, .0.0)$ m の点音源, 1 kHz正弦波, シフト長 2m (y方向のみ)

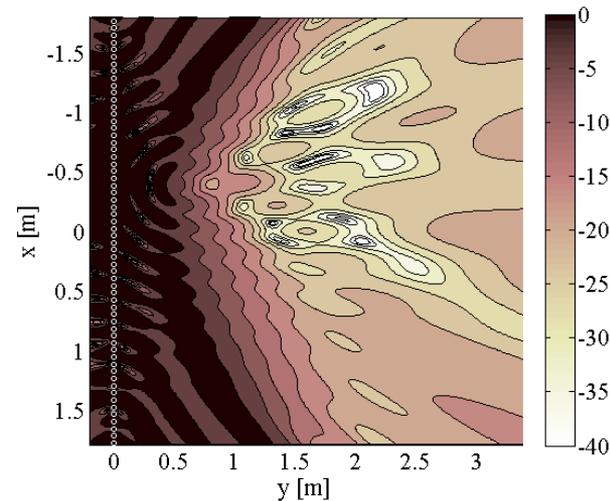
音圧分布



WFR



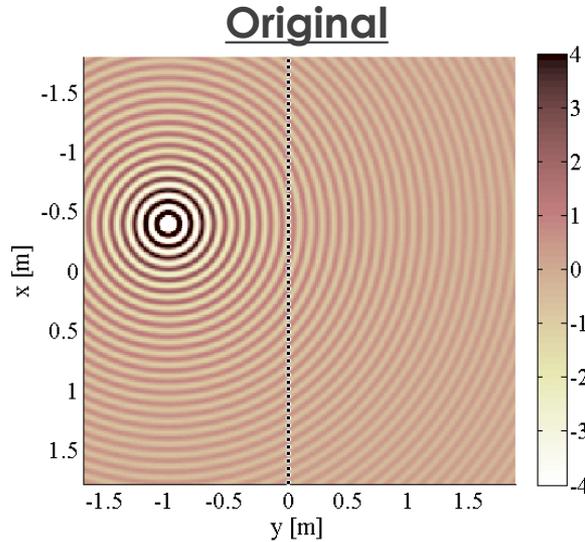
誤差分布



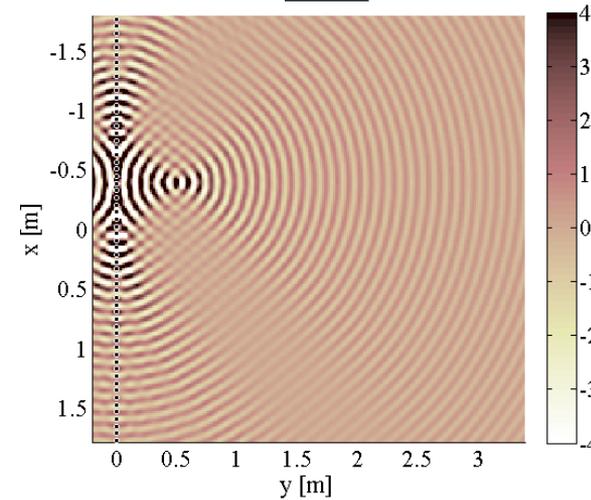
シミュレーション実験 (3.6kHz)

位置 (-0.4, -1.0, .0.0)m の点音源, 3.6 kHz正弦波, シフト長 2m (y方向のみ)

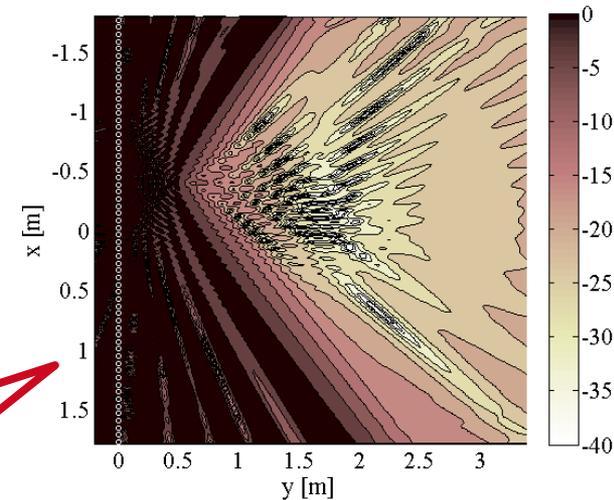
音圧分布



WFR



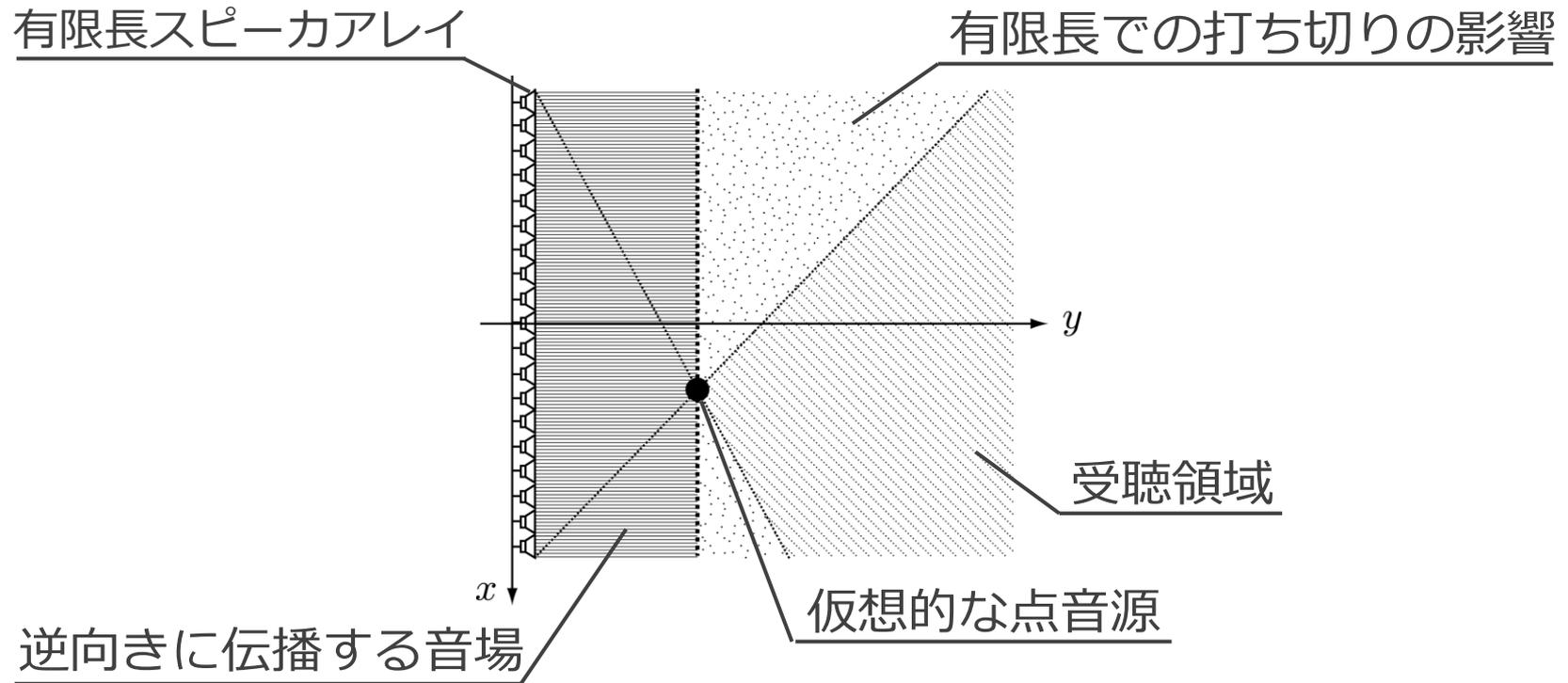
誤差分布



仮想音源位置近傍では
空間エイリアシングの影響が小さい

焦点型音源の再現における制約

スピーカアレイ前面への仮想音源の再現は、焦点を形成することに対応



- 受音直線上の空間スペクトルにエバネッセント波成分が含まれると、信号変換は極めて不安定になる

➡ $|k_{x,i}| > k$ の成分を抑制するような空間周波数領域上の窓掛けが必要

エバネッセント波成分 = 音源近傍で急峻に減衰する成分

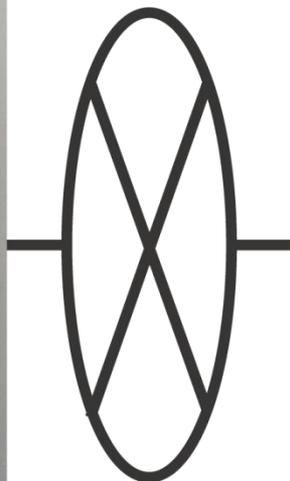
リアルタイム音場伝送システム@NTT研究所

横須賀

武蔵野



Network



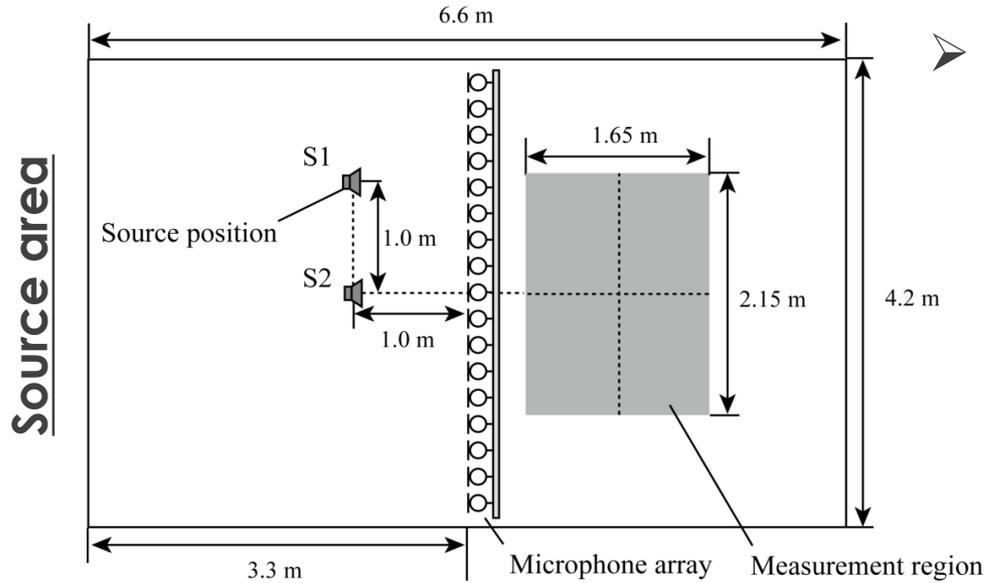
▶ IPネットワークを用いたリアルタイム音場伝送システム

- 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

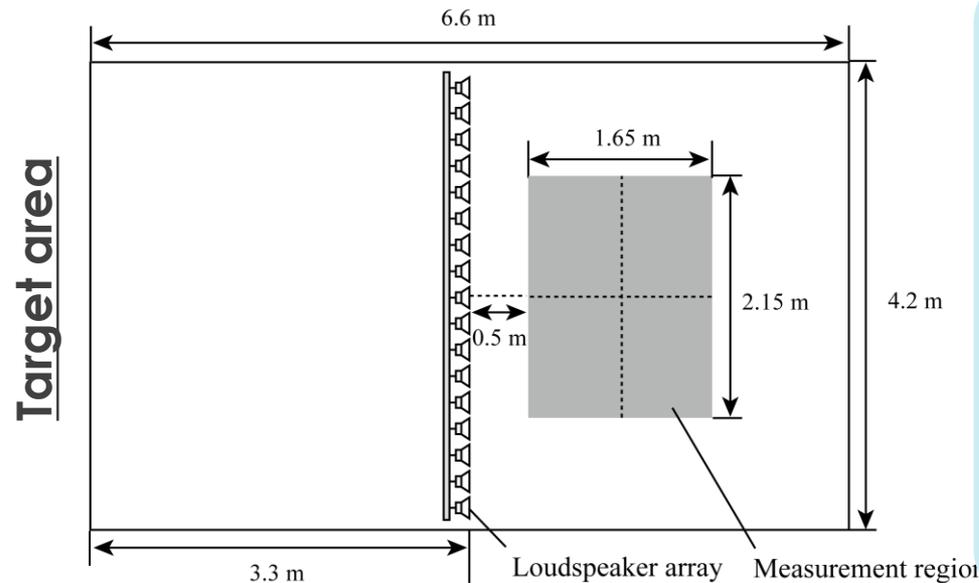


※片方向の伝送・再現のみ

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

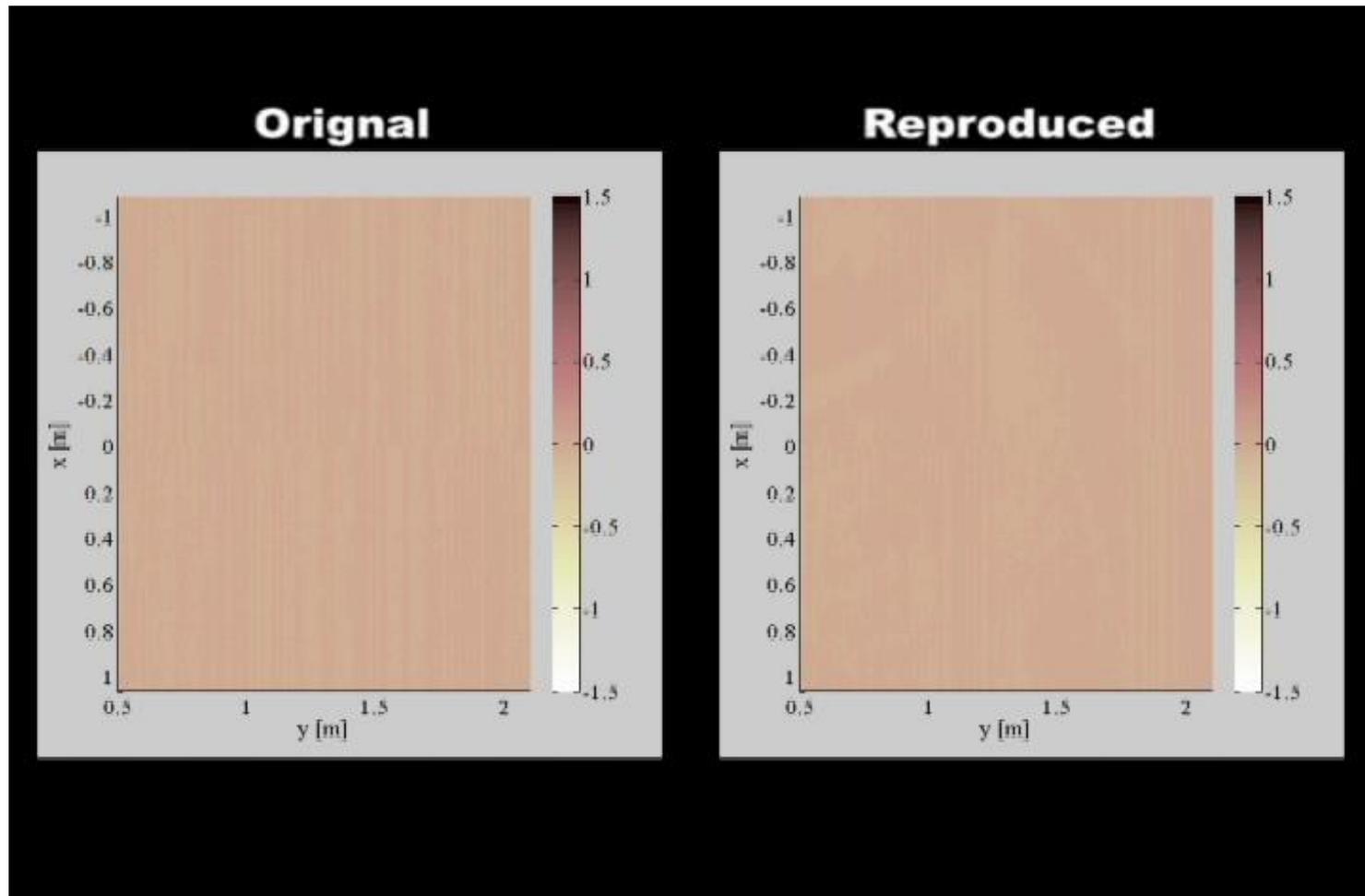


測定実験の様子



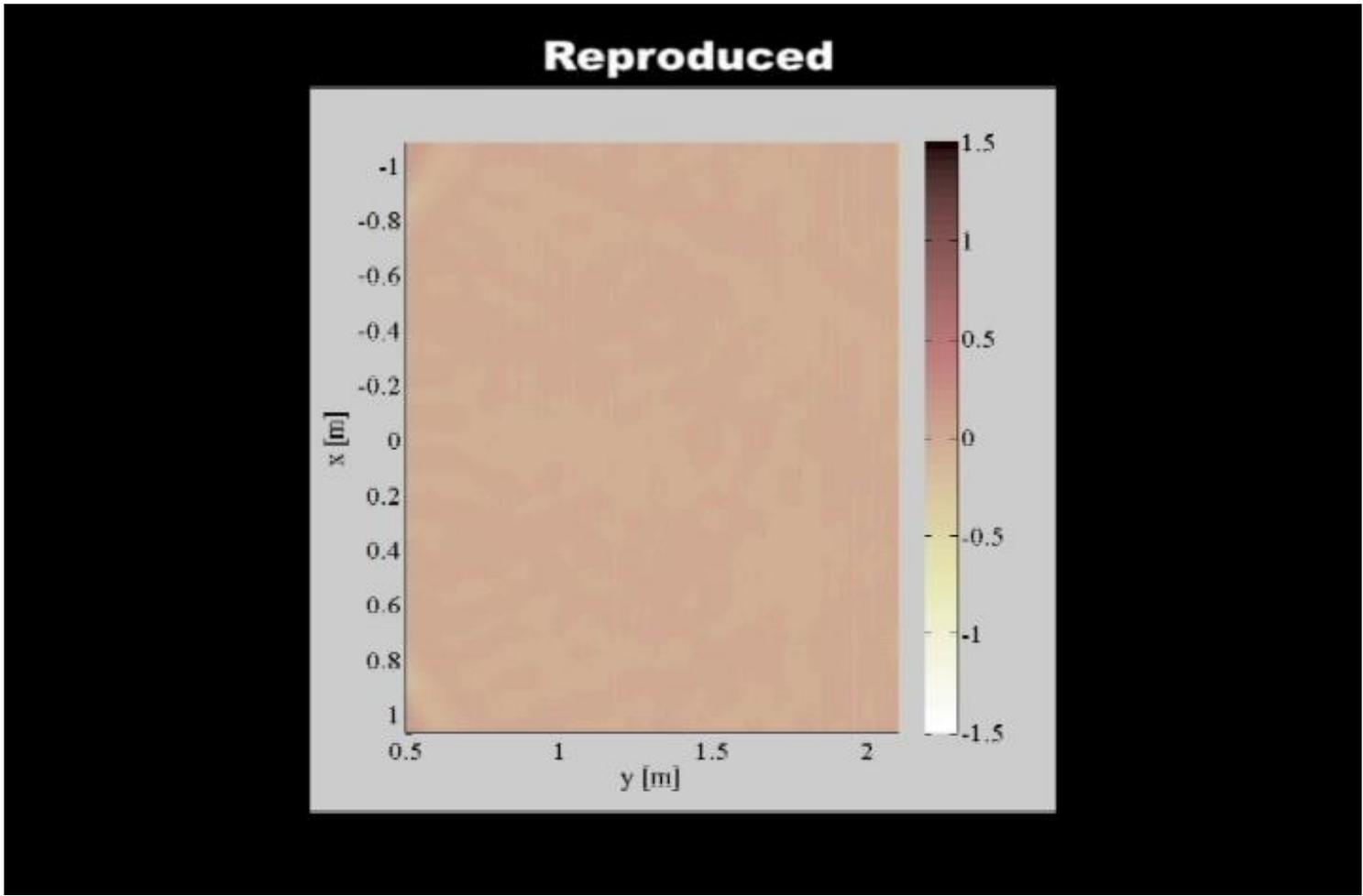
再現音場の可視化①

- 音源信号：帯域制限パルス信号 ($\sim 2.6\text{kHz}$)
- 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)



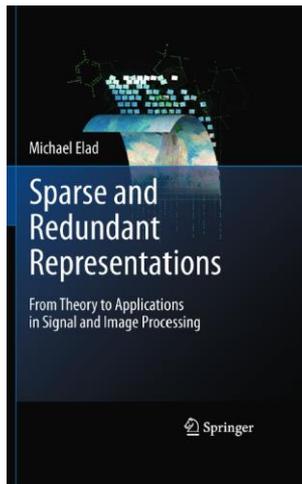
関連文献

- A. J. Berkhout, *et al.* “Acoustic control by wave field synthesis,” *JASA*, 1993.
- S. Spors, *et al.* “The theory of wave field synthesis revisited,” *Proc. 124th Conv. AES*, 2008.
- J. Daniel, “Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format,” *Proc. 23rd Conf. AES*, 2003.
- M. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” *J. AES*, 2005.
- P. A. Gauthier and A. Berry, “Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain,” *JASA*, 2005.
- 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.* “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.

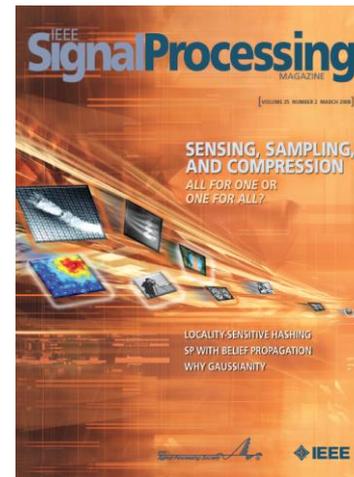
スパース信号表現

スパース信号表現とは

- 信号を、ある基底の下で多くの成分の大きさがゼロ、またはゼロとみなせるほど小さくなる（**スパース**になる）ように表現すること。
- **圧縮センシング（Compressed sensing）**
 - スパースに表現可能な信号を少数の計測データから再構成を行う。
 - Donoho, Candés, Taoら（2003年頃）



M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing," Springer, 2010.



IEEE Signal Processing Magazine, Volume 25, Issue 2, 2008.

例：画像のスパース表現

▶ ウェーブレット基底での画像の表現



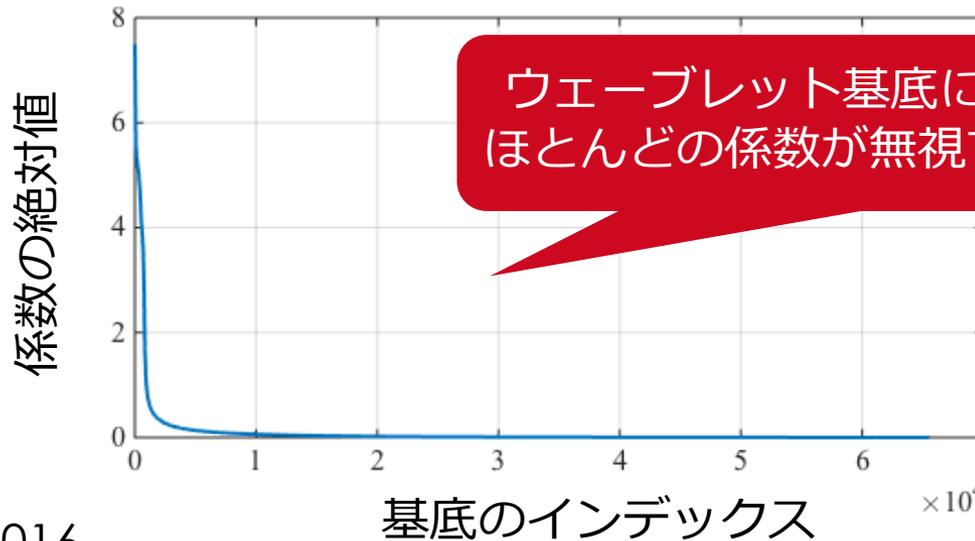
原画像



上位10%



上位3%



スパース信号表現とは

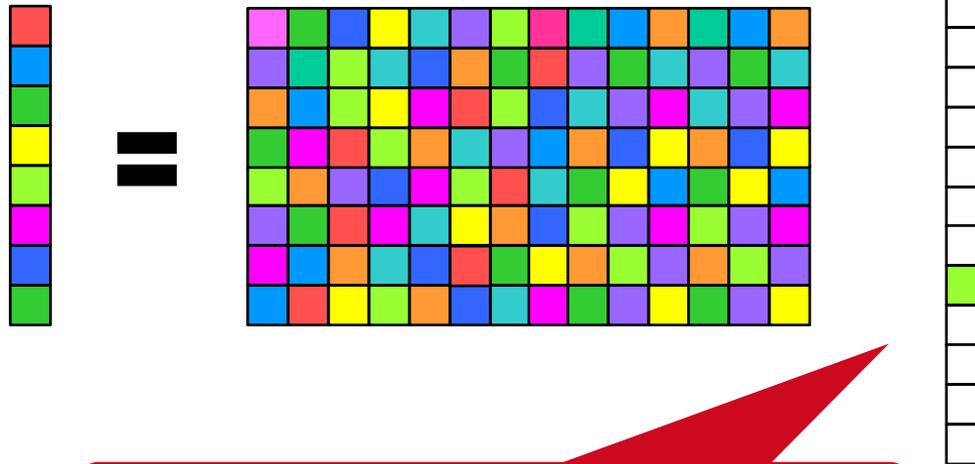
- 行列形式では、線形方程式として書ける。

各列が既知の基底関数に対応

観測信号 : y

基底の行列 (辞書行列) : D

基底の係数 : x



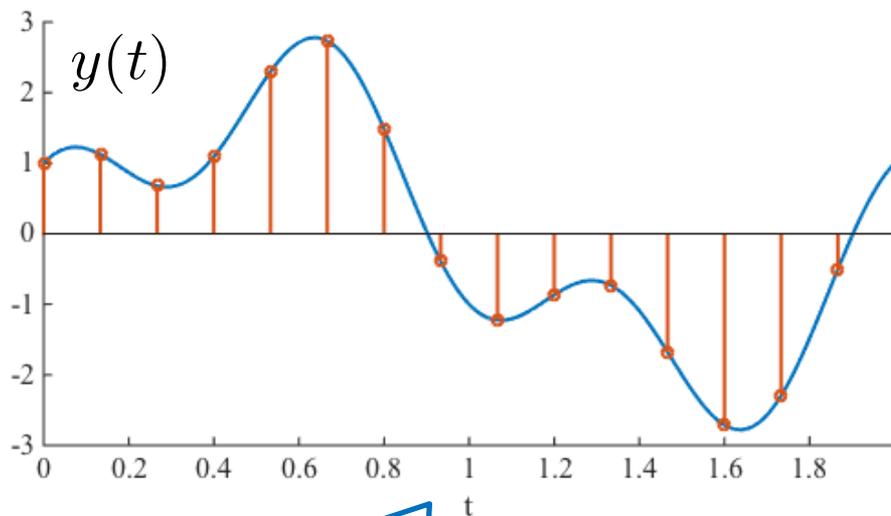
非ゼロの成分が少数 = スパース

x がスパースになるとなぜ嬉しいのか？
どうやってスパースな x を求めるのか？

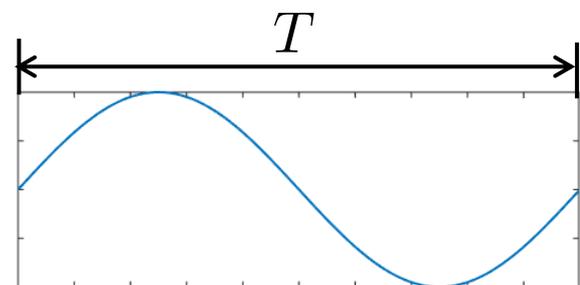
サンプリング定理 再考

➤ 周期 T の信号 $y(t)$ のフーリエ級数展開

$$y(t) = a_0 + \sum_{k=1}^K \left[a_k \cos \left(\frac{2\pi kt}{T} \right) + b_k \sin \left(\frac{2\pi kt}{T} \right) \right]$$



線形和

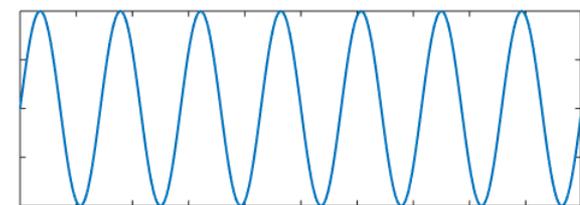


波数1

+

⋮

+



波数K

各フーリエ係数 ($a_0, a_1, b_1, \dots, a_K, b_K$)
がわかれば、信号 $y(t)$ を再構成可能

サンプリング定理 再考

- 等間隔サンプリングにより、フーリエ係数間の独立な一次方程式を得る

$$y(n\Delta t) = \underline{a_0} + \sum_{k=1}^K \left[\underline{a_k} \cos \left(\frac{2\pi kn\Delta t}{T} \right) + \underline{b_k} \sin \left(\frac{2\pi kn\Delta t}{T} \right) \right]$$

- 求めたいフーリエ係数は**2K+1**個であることから、1周期の間に**2K+1**点以上のサンプリングによって、信号の完全再構成が可能となる。

方程式の数 \geq 未知数の数

サンプリング定理 再考

➤ 行列形式 ($y = Dx$) で表現

2K+1 サンプルの観測値

$$\begin{bmatrix} y(\Delta t) \\ y(2\Delta t) \\ \vdots \\ y((2K+1)\Delta t) \end{bmatrix} =$$

2K+1 個の未知数

$$\begin{bmatrix} 1 & \cos\left(\frac{2\pi\Delta t}{T}\right) & \sin\left(\frac{2\pi\Delta t}{T}\right) & \cos\left(\frac{2\pi 2\Delta t}{T}\right) & \cdots & \sin\left(\frac{2\pi K\Delta t}{T}\right) \\ 1 & \cos\left(\frac{2\pi 2\Delta t}{T}\right) & \sin\left(\frac{2\pi 2\Delta t}{T}\right) & \cos\left(\frac{2\pi 2 \cdot 2\Delta t}{T}\right) & \cdots & \sin\left(\frac{2\pi K \cdot 2\Delta t}{T}\right) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(\frac{2\pi(2K+1)\Delta t}{T}\right) & \sin\left(\frac{2\pi(2K+1)\Delta t}{T}\right) & \cos\left(\frac{2\pi 2(2K+1)\Delta t}{T}\right) & \cdots & \sin\left(\frac{2\pi K(2K+1)\Delta t}{T}\right) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ \vdots \\ b_K \end{bmatrix}$$

(2K+1) × (2K+1) の行列

辞書行列の逆行列が存在することから、信号の完全再構成が可能

サンプリング定理 再考

- ▶ 周期 T の間に $2K + 1$ 個のサンプルが必要なことから、サンプリング周波数 F_s は以下のように書ける。

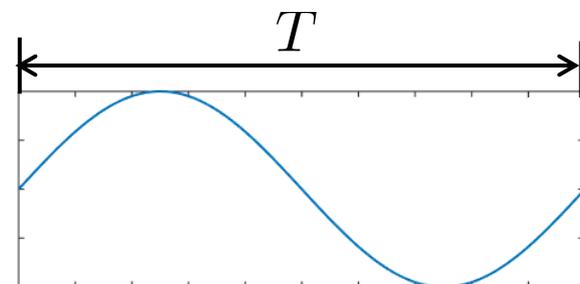
$$F_s > \frac{2K + 1}{T}$$

- ▶ 波数の最大値 K に対応する周波数は、

$$f_{\max} = \frac{K}{T}$$

- ▶ サンプリング周波数と周波数の最大値との関係が以下のように得られる。

$$F_s > \frac{2K + 1}{T} = 2f_{\max} + \frac{1}{T}$$
$$\xrightarrow{T \rightarrow \infty} 2f_{\max}$$

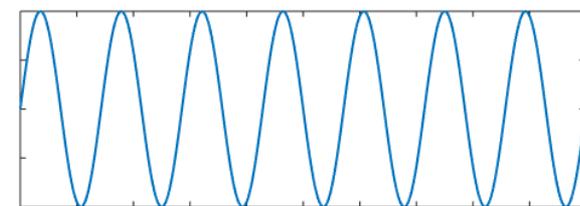


波数1

+

⋮

+



波数K

サンプリング定理 再考

- ▶ フーリエ係数に関する解の一意性の条件としてサンプリング定理を再考した。

$$\underline{y} = \underline{D} \underline{x}$$

The diagram shows the equation $y = Dx$. The variable y is underlined in orange and has a label '観測信号' (Observed signal) pointing to it from the bottom left. The matrix D is underlined in green and has a label 'フーリエ基底' (Fourier basis) pointing to it from the bottom right. The vector x is underlined in pink and has a label 'フーリエ係数' (Fourier coefficients) pointing to it from the top right.

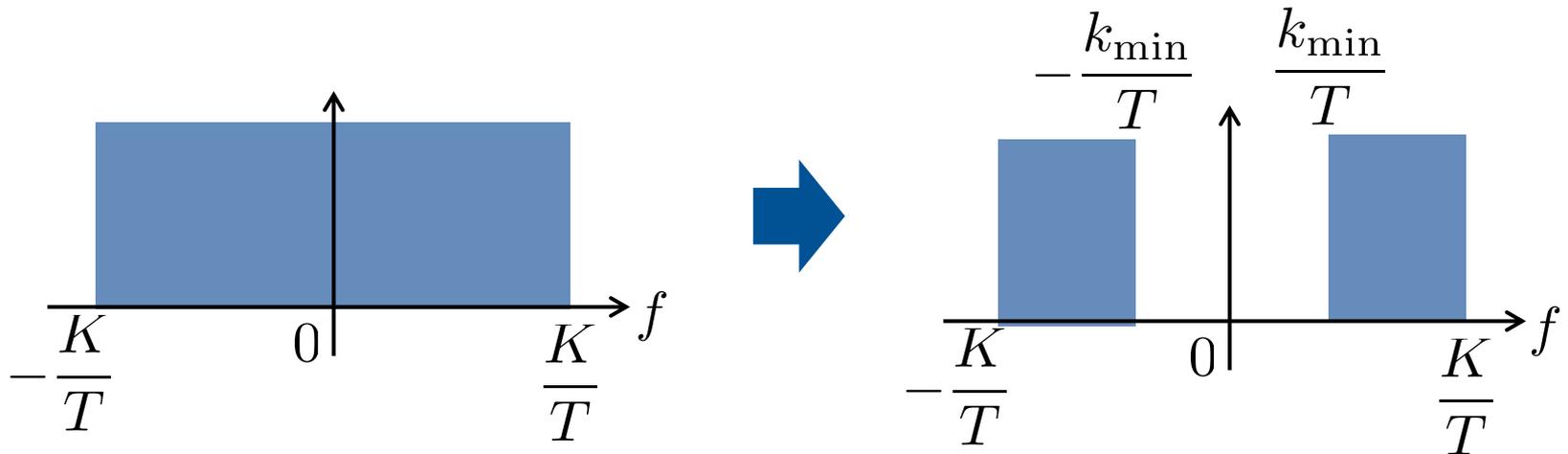
- ▶ 一般化して考えると、
 - 等間隔サンプリングでなくても、フーリエ係数と同じ数の独立なサンプル値さえ得られればよい。
 - 辞書行列は、ウェーブレットのような別の基底でも良いし、直交基底でなくても良い。

スパース性を利用した信号復元

- ▶ サンプリング定理の結果よりも、少ないサンプル数で信号を復元可能な例

$$y(n\Delta t) = \sum_{k=k_{\min}}^K \left[a_k \cos \left(\frac{2\pi kn\Delta t}{T} \right) + b_k \sin \left(\frac{2\pi kn\Delta t}{T} \right) \right]$$

k の最小値が k_{\min} として制約されている場合



$2(K - k_{\min} + 1)$ 個のサンプルで信号復元が可能

スパース性を利用した信号復元

➤ 行列形式で考えると、

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

両辺に \mathbf{D}_s^T を掛ける

$$\mathbf{y}_s = \mathbf{D}_s^T \mathbf{y}$$

ゼロとなる係数が存在する場合、
より少ないサンプル数で信号復元が可能

圧縮センシング

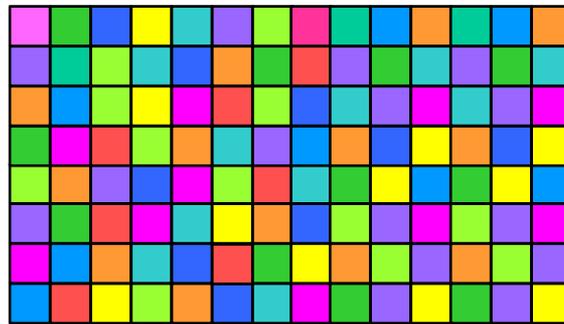
➤ ゼロとなる要素が未知の場合 ➡ 圧縮センシングの問題

観測信号 : y



=

辞書行列 : D



係数 : x



- 未知数の数 > 方程式の数の劣決定問題
- ゼロの成分が多いことはわかっている(場所は不明)

- 信号がある領域で多くのゼロ成分を持つものは多い

- どのようなアルゴリズムでスパースな解を求め、少ないサンプルから信号復元を行うか
- どのような条件において、信号復元の性能が保証されるか

信号復元アルゴリズム

▶ ベクトル $\mathbf{x} \in \mathbb{R}^N$ のノルム

$$\|\mathbf{x}\|_p = \left(\sum_{n=1}^N |x_n|^p \right)^{1/p} = \sqrt[p]{|x_1|^p + \cdots + |x_N|^p}$$

▶ 以下のコスト関数を最小化する \mathbf{x} が、スパースな解として最良

$$\text{minimize } \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{D}\mathbf{x}$$

非ゼロの要素の個数

非ゼロ要素に関するあらゆる数・位置の組み合わせが存在し、
解を求めることは計算量的に困難（NP困難な問題）

信号復元アルゴリズム

➤ 現実的な解法のいくつかを紹介

➤ 貪欲法

- 最良のサポートを順に付け加えることでサポートを構成
- Orthogonal Matching Pursuit (OMP)など

➤ 凸緩和法

- コスト関数を凸関数として近似する
- Basis Pursuit, 近接勾配法 (proximal gradient method) など

➤ 補助関数法

- コスト関数に対する補助関数を設計し, 交互に最適化
- FOCal Underdetermined System Solver (FOCUSS)など

貪欲法 (Greedy algorithms)

- 観測ベクトル \underline{y} を最も良く近似する列ベクトルを順にサポートに付け加える

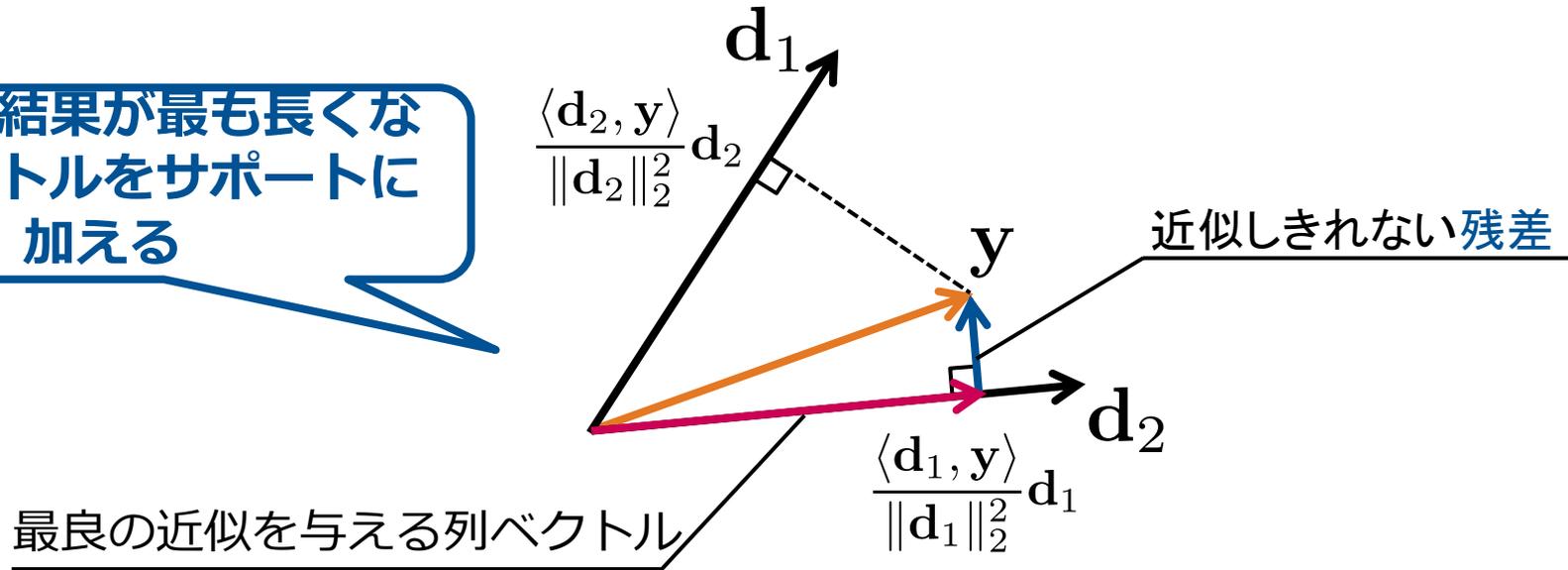
$\underline{y} = \underline{D} \underline{x} = x_1 \mathbf{d}_1 + x_2 \mathbf{d}_2 + \cdots + x_N \mathbf{d}_N$

- 辞書行列の各列 $\{\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N\}$ の中から、 \underline{y} を最も良く近似するものを探索
- 最良の候補を \underline{x} のサポート S の要素として加えて行く

貪欲法 (Greedy algorithms)

- 単一の列ベクトルによる近似を繰り返す

射影した結果が最も長くなる列ベクトルをサポートに加える



最良近似を与える射影 $\hat{x}_n = \arg \min_{x_n} \|\mathbf{y} - x_n \mathbf{d}_n\|_2 = \frac{\langle \mathbf{d}_2, \mathbf{y} \rangle}{\|\mathbf{d}_2\|_2} \mathbf{d}_2$

近似誤差 $\epsilon_n = \min_{x_n} \|\mathbf{y} - x_n \mathbf{d}_n\|_2 = \|\mathbf{y}\|_2 - \frac{\langle \mathbf{d}_n, \mathbf{y} \rangle^2}{\|\mathbf{d}_n\|_2^2}$

Orthogonal Matching Pursuit (OMP)

- **Initialization:** Initialize $k=0$, and set

$$\mathbf{x}^0 = \mathbf{0}, \mathbf{r}^0 = \mathbf{y} - \mathbf{D}\mathbf{x}^0, S^0 = \emptyset$$

- **Main Iteration:** Increment k by 1 and perform the following steps

- **Sweep:** $\epsilon_n = \min_{x_n} \|\mathbf{r}^{k-1} - x_n \mathbf{d}_n\|_2^2 = \|\mathbf{r}^{k-1}\|_2^2 - \frac{\langle \mathbf{d}_n, \mathbf{r}^{k-1} \rangle^2}{\|\mathbf{d}_n\|_2^2}$

- **Update support:** $\hat{n} = \arg \min_{n \notin S^{k-1}} \epsilon_n, S^k = S^{k-1} \cup \{\hat{n}\}$

- **Update provisional solution:** $\hat{\mathbf{x}}^k = \arg \min_{\mathbf{x}_{S^k}} \|\mathbf{y} - \mathbf{D}_{S^k} \mathbf{x}_{S^k}\|_2^2$

- **Update residual:** $\mathbf{r}^k = \mathbf{y} - \mathbf{D}_{S^k} \hat{\mathbf{x}}^k$

- **Stopping rule:** Stop if $\|\mathbf{r}^k\|_2 < \epsilon$ holds.

凸緩和法 (Convex relaxation)

- ℓ_0 ノルムは離散最適化となり困難であるため、連続関数で近似する

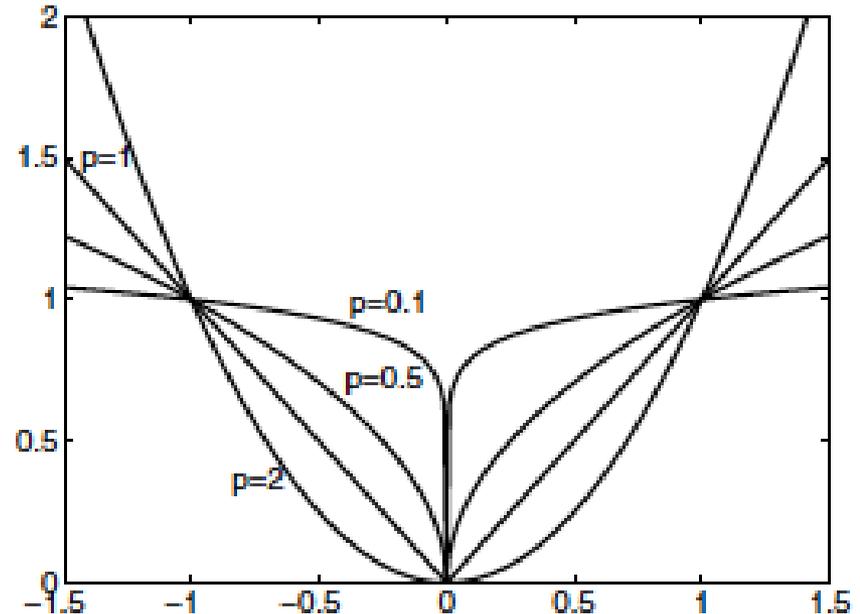
ℓ_0 ノルム

$$\|\mathbf{x}\|_0 = \lim_{p \rightarrow 0} \sum_{n=1}^N |x_n|^p$$

ℓ_p ノルム

$$\|\mathbf{x}\|_p = \left(\sum_{n=1}^N |x_n|^p \right)^{1/p}$$

$|x|^p$ のグラフ



M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing," Springer, 2010.

凸緩和法 (Convex relaxation)

- 凸関数となれば最適化が容易となる → $p \geq 1$
- 凸関数かつ最適解がスパースになることが望ましい

ℓ_p ノルム

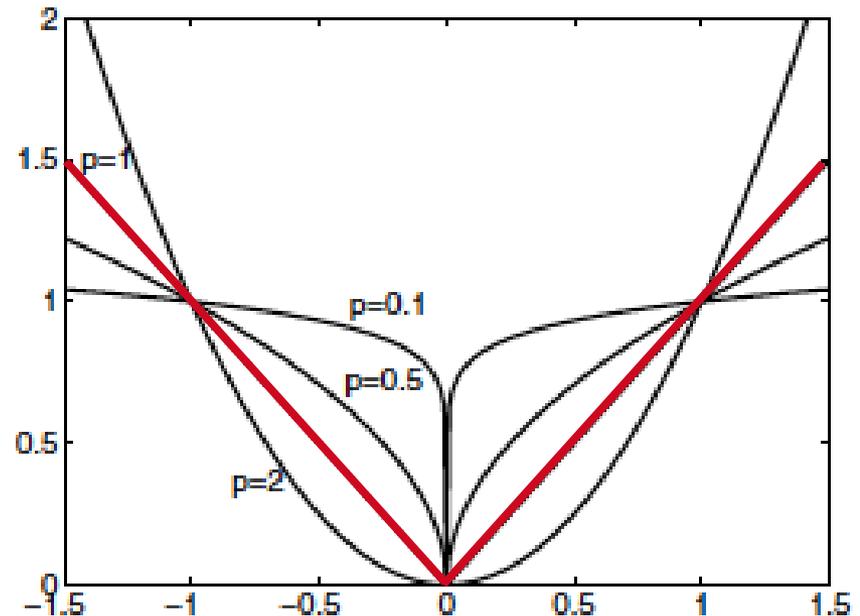
$$\|\mathbf{x}\|_p = \left(\sum_{n=1}^N |x_n|^p \right)^{1/p}$$

↓ $p = 1$

ℓ_1 ノルム

$$\|\mathbf{x}\|_1 = \sum_{n=1}^N |x_n|$$

$|x|^p$ のグラフ

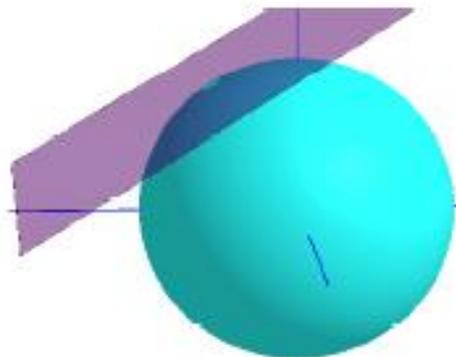


M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing," Springer, 2010.

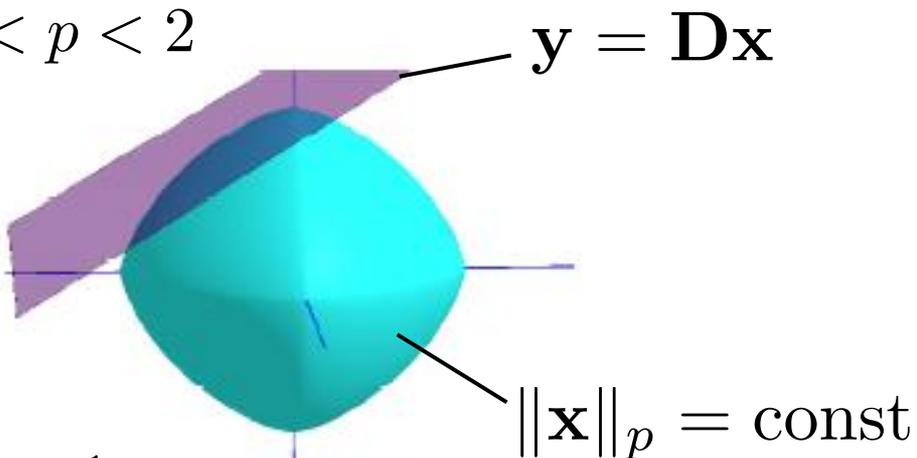
ノルムと解との関係

$$\text{minimize } \|\mathbf{x}\|_p \text{ subject to } \mathbf{y} = \mathbf{D}\mathbf{x}$$

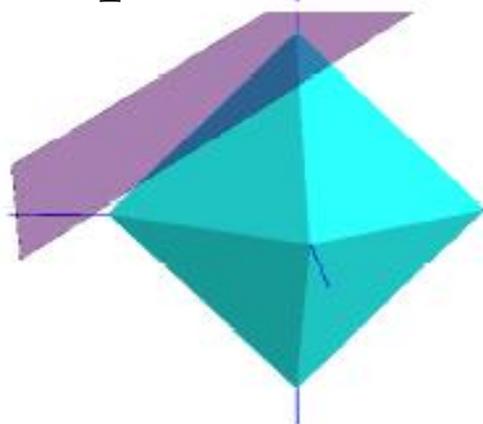
$p = 2$



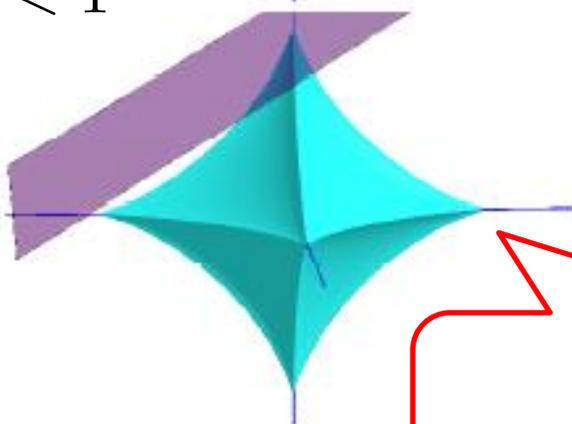
$1 < p < 2$



$p = 1$



$p < 1$



$p \leq 1$ の場合に
解がスパースになりやすい

M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing," Springer, 2010.

Basis Pursuit

➤ minimize $\|\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{D}\mathbf{x}$



➤ 線形計画問題へ書き換える

$$\mathbf{x} = \mathbf{u} - \mathbf{v} \quad \begin{cases} \mathbf{u} : \mathbf{X} \text{の正の成分はそのまま、負の成分をゼロ} \\ \mathbf{v} : \mathbf{X} \text{の正の成分はゼロ、負の成分をその絶対値} \end{cases}$$

$$\mathbf{z} = [\mathbf{u}^T, \mathbf{v}^T]^T \in \mathbb{R}^{2N}$$

全ての要素が1のベクトル

$$\Rightarrow \|\mathbf{x}\|_1 = \mathbf{1}^T (\mathbf{u} + \mathbf{v}), \quad \mathbf{D}\mathbf{x} = \mathbf{D}(\mathbf{u} - \mathbf{v}) = [\mathbf{D}, -\mathbf{D}] \mathbf{z}$$

$$\Rightarrow \text{minimize } \mathbf{1}^T \mathbf{z} \text{ subject to } \mathbf{y} = [\mathbf{D}, -\mathbf{D}] \mathbf{z} \text{ and } \mathbf{z} \succeq \mathbf{0}$$

線形計画問題を解くためのアルゴリズムが使える

近接勾配法

- L1ノルムを罰則項とした以下の問題を解く

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

- 第1項（損失項）を $\mathbf{x} = \mathbf{x}^t$ まわりで線形近似

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \\ &\simeq f(\mathbf{x}^t) + \langle \nabla f(\mathbf{x}^t), \mathbf{x} - \mathbf{x}^t \rangle + \frac{1}{2\eta} \|\mathbf{x} - \mathbf{x}^t\|_2^2 \end{aligned}$$



$$\underset{\mathbf{x}}{\text{minimize}} \lambda \|\mathbf{x}\|_1 + \frac{1}{2\eta} \left\| \mathbf{x} - \left(\mathbf{x}^t - \eta \nabla f(\mathbf{x}^t) \right) \right\|_2^2$$

prox作用素を利用できる形に式変形

近接勾配法

➤ L1ノルムに関するprox作用素を利用

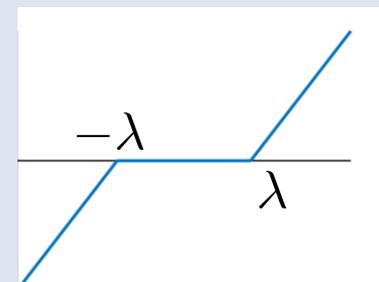
$$\text{prox}_{\lambda}^{\ell_1}(\mathbf{y}) = \arg \min_{\mathbf{w}} \left(\frac{1}{2} \|\mathbf{y} - \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1 \right)$$

上記のL1ノルムに関するprox作用素は、以下のように解析的に書ける

$$\left[\text{prox}_{\lambda}^{\ell_1}(\mathbf{y}) \right] = \begin{cases} y_m + \lambda, & y_m < -\lambda \\ 0, & -\lambda \leq y_m \leq \lambda \\ y_m - \lambda, & y_m > \lambda \end{cases}$$

yの各要素

ソフト閾値関数



※証明は富岡「スパース性に基づく機械学習」などを参照

近接勾配法

- 先程の式にprox作用素を適用すると，更新式が得られる

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

線形近似

➡
$$\underset{\mathbf{x}}{\text{minimize}} \lambda \|\mathbf{x}\|_1 + \frac{1}{2\eta} \|\mathbf{x} - (\mathbf{x}^t - \eta \nabla f(\mathbf{x}^t))\|_2^2$$

prox作用素

➡
$$\mathbf{x}^{t+1} = \text{prox}_{\lambda\eta}^{\ell_1} (\mathbf{x}^t - \eta \nabla f(\mathbf{x}^t))$$

ソフト閾値処理を反復的に適用することで最適解が得られる

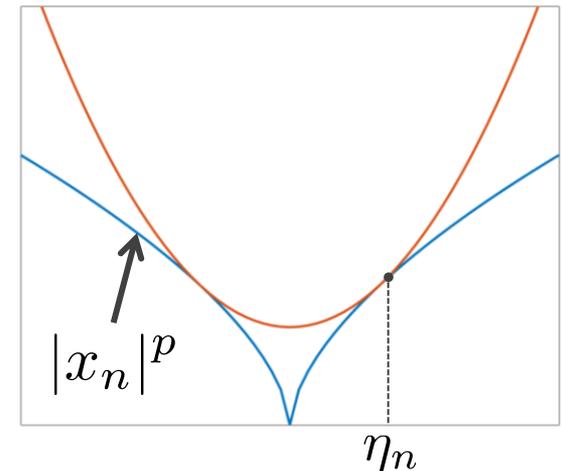
1 反復あたりの計算量が非常に小さい

補助関数法

- 以下のような最適化問題に書き換える

$$\begin{cases} \underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + J(\mathbf{x}) \\ J(\mathbf{x}) = \|\mathbf{x}\|_p^p = \sum_n |x_n|^p \end{cases}$$

- $p < 1$ の場合はコスト関数が非凸



➡ **最適化しやすい上限関数を定義して反復的に最小化**

$$J(\mathbf{x}) = \sum_n |x_n|^p$$

$$\leq \frac{1}{2} \sum_n \{ p\eta_n^{p-2} x_n^2 + (2-p)\eta_n^p \}$$

$$\equiv J^+(\mathbf{x}, \boldsymbol{\eta})$$

接放物線を上限関数とする

($x_n = \eta_n$ のときに等号成立)

FOCUSS

➤ 以下を繰り返す

- \mathbf{x} の更新

$$\begin{aligned}\mathbf{x}^t &= \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda J^+(\mathbf{x}, \boldsymbol{\eta}^t) \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_n \underbrace{p\eta_n^{p-2} x_n^2}_{\text{重み付きリッジ正則化項}}\end{aligned}$$

- $\boldsymbol{\eta}$ の更新

$$\eta_n^{t+1} = |x_n^t| \quad \text{として補助変数を更新}$$

アルゴリズムとしては繰り返し重み付き縮小 (iteratively reweighted shrinkage) 法としても知られる

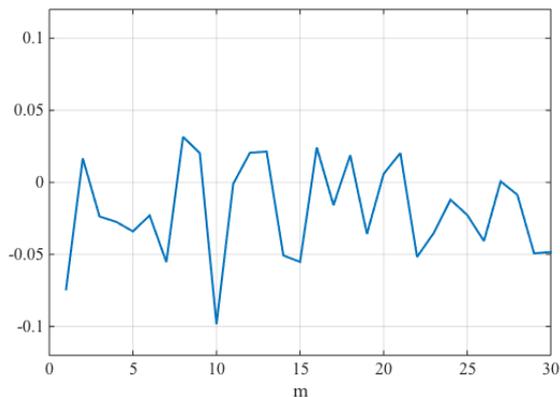
局所解に落ちる可能性はあるものの、
 p を1より小さい値にしても適用可能

Basis Pursuitを用いた数値例

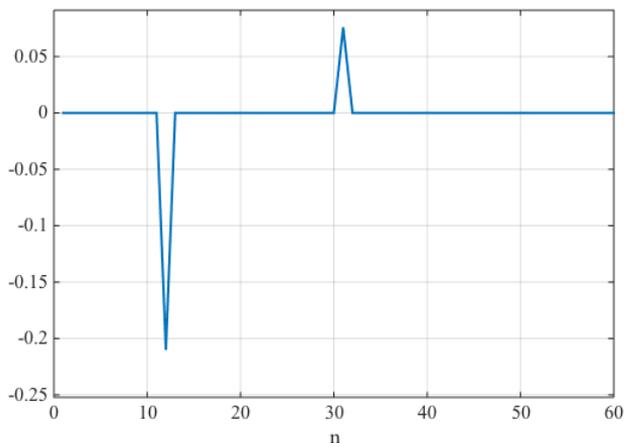
➤ $\mathbf{y} \in \mathbb{R}^{30}$ $\mathbf{x} \in \mathbb{R}^{60}$ $\mathbf{D} \in \mathbb{R}^{30 \times 60}$ の場合の実験結果

各要素を乱数で生成

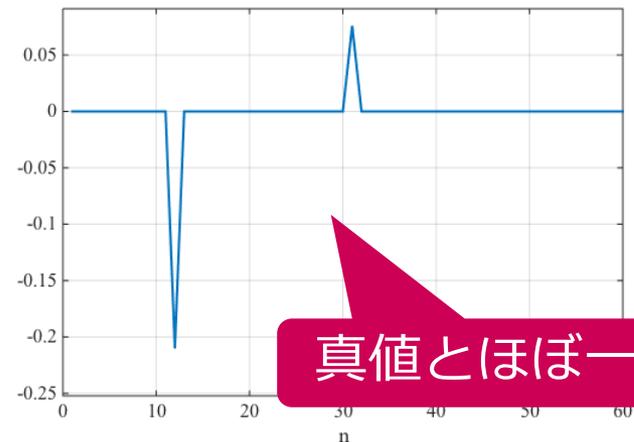
観測信号 \mathbf{y}



真値 \mathbf{x}

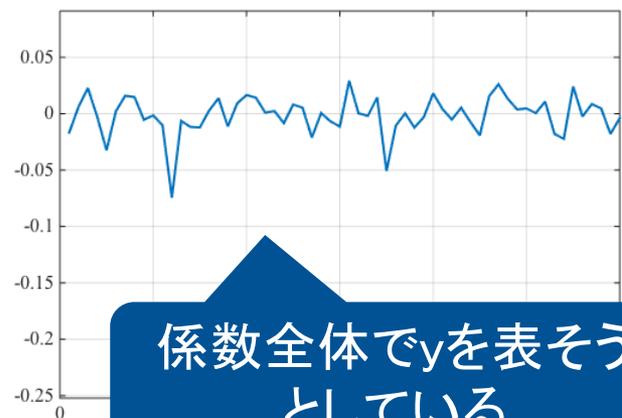


Basis Pursuit (p=1)



真値とほぼ一致

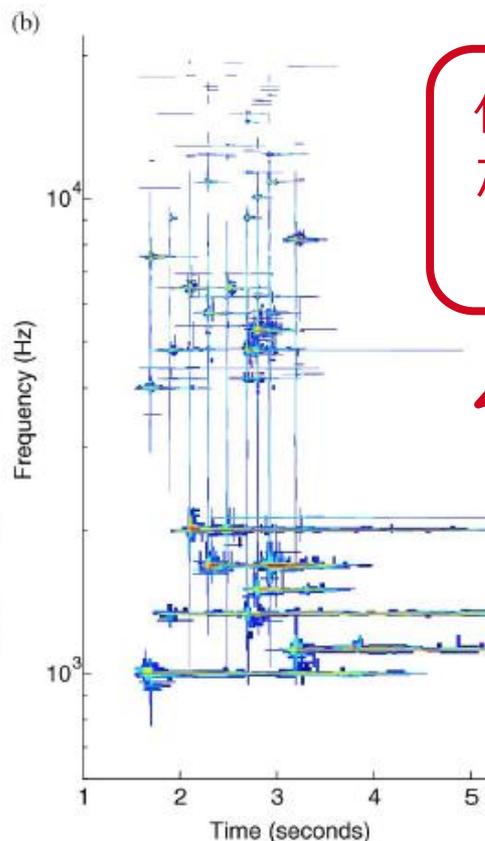
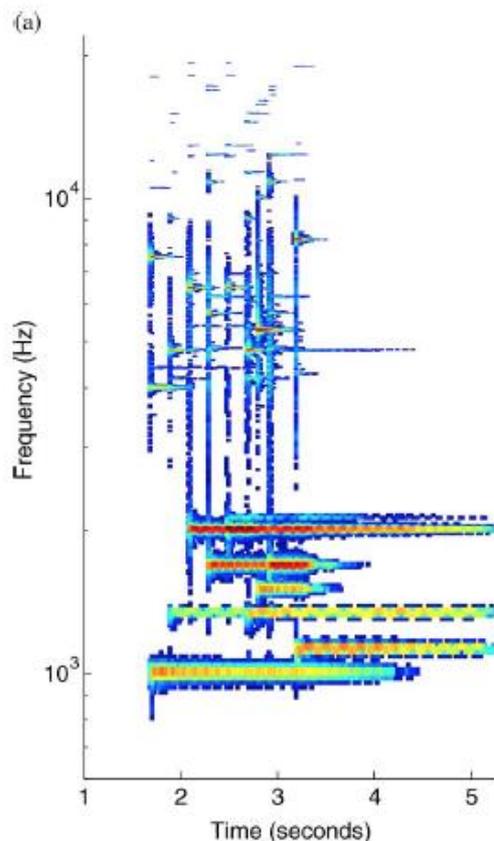
最小ノルム解 (p=2)



係数全体でyを表そう
としている

音声・音響信号処理への応用①

➤ オーディオ符号化 [Plumbley+ Proc. IEEE 2010]



係数のほとんどがゼロとなるため、より低ビットレートでデータ圧縮が可能となる

一般的なMDCT基底

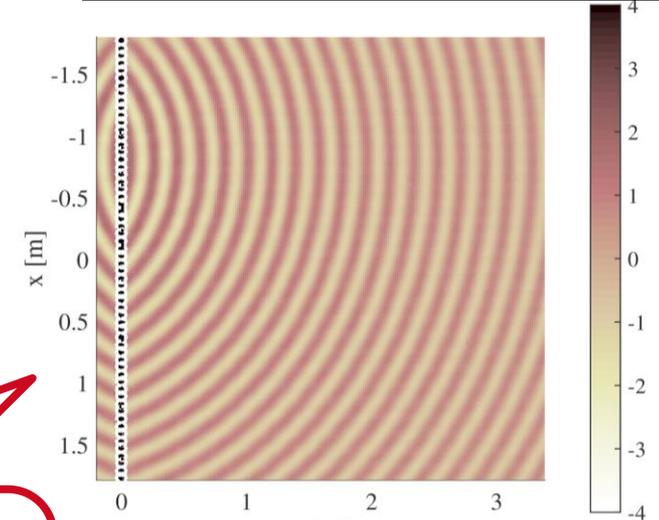
窓長を変えた8つのMDCT基底の組み合わせ

音声・音響信号処理への応用②

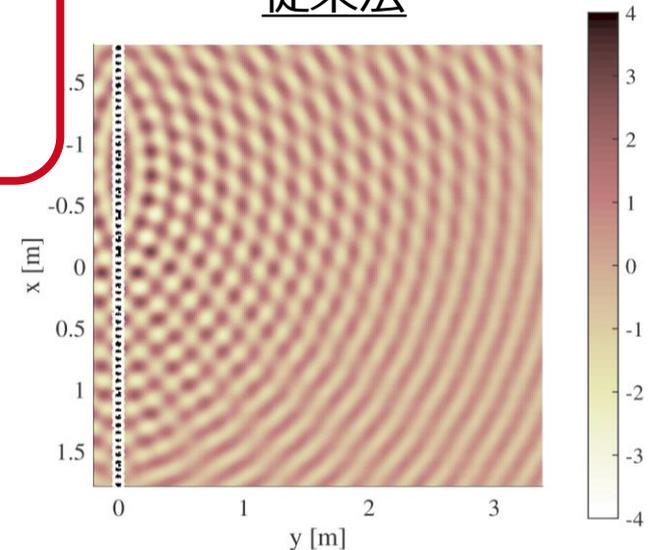
音場收音・再現の超解像化 [Koyama+ ICASSP 2014]

- 音場をモノポール音源成分と平面波成分の和としてモデル化
- モノポール音源成分が空間的にスパースに分布していることを仮定
- 各成分に分解した上でスピーカ駆動信号に変換し、空間エイリアシングの誤差を低減

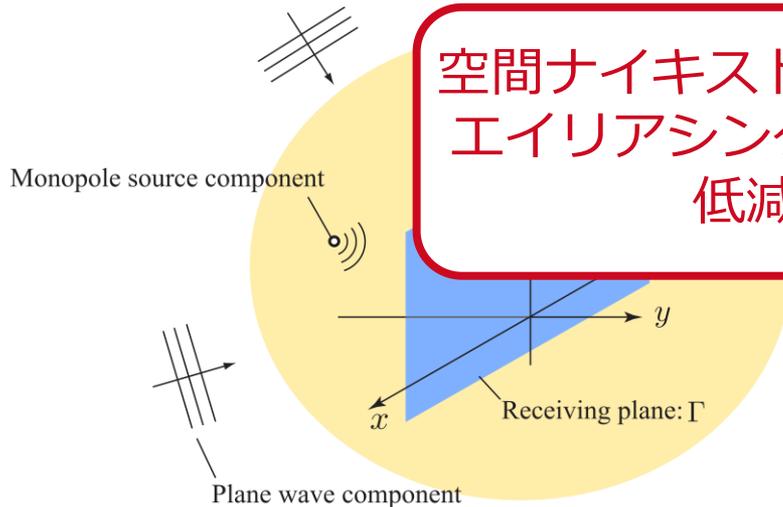
スパース分解に基づく方法



従来法



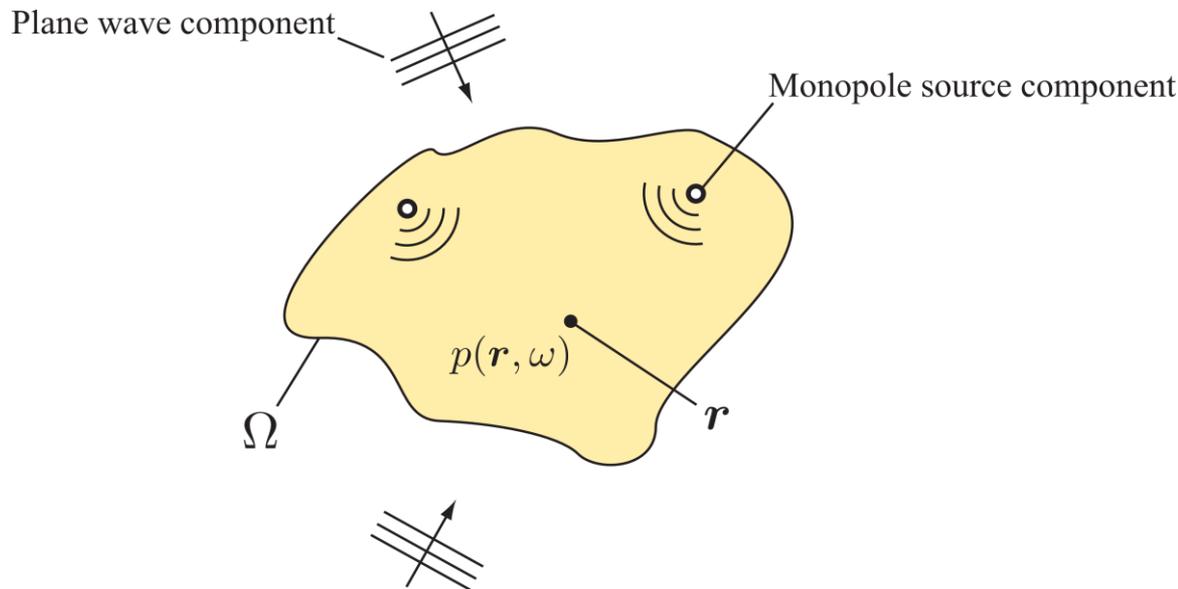
空間ナイキスト周波数以上で、エイリアシングによる誤差を低減可能



スパース信号表現に基づく 音場再現の超解像化

音場の生成モデル

[Koyama+ ICASSP 2014]



➤ 非斉次・斉次ヘルムホルツ方程式

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

モノポール音源成分の分布

非斉次項 + 斉次項



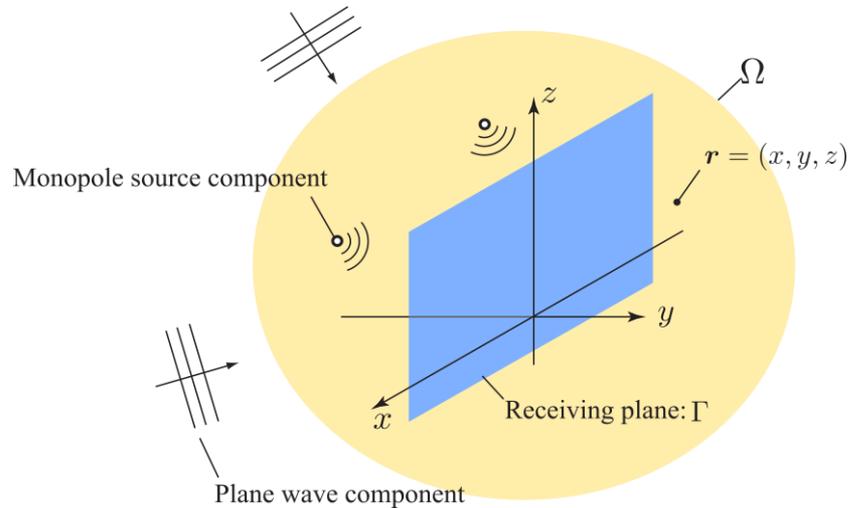
$$\underline{p(\mathbf{r}, \omega)} = p_i(\mathbf{r}) + p_h(\mathbf{r})$$

$$= \int_{\mathbf{r}' \in \Omega} \underline{Q(\mathbf{r}')} \underline{G(\mathbf{r}|\mathbf{r}')} d\mathbf{r}' + \underline{p_h(\mathbf{r})}$$

グリーン関数

音場の生成モデル

▶ 受音平面 Γ 上での音圧分布の取得



▶ 駆動信号への変換

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

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

モノポール音源の合成 [Spors+ AES Conv 2008]

WFRフィルタ法の適用 [Koyama+ IEEE TASLP 2013]

二つの成分に音場を分解できればマイクロフォン間隔で決まる空間ナイキスト以上での再現が可能となる

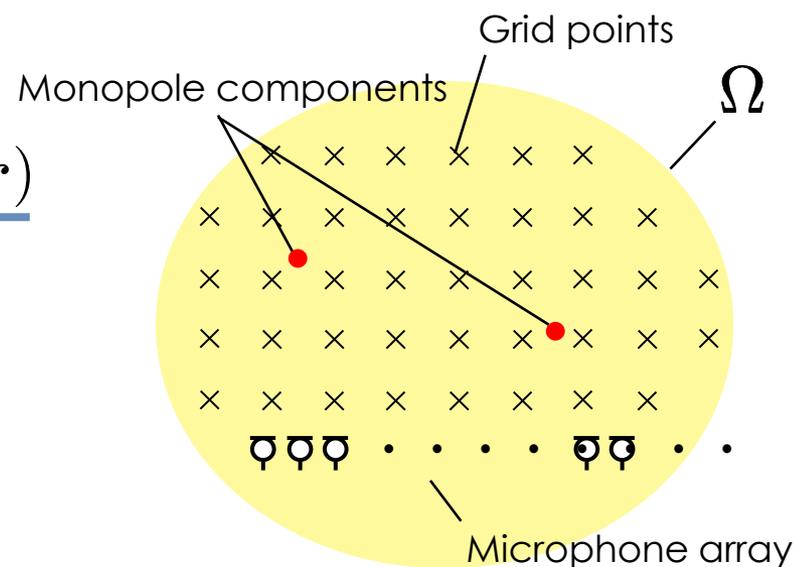
スパース音場分解

音場のスパース表現

$$\underline{p}(\mathbf{r}) = \int_{\mathbf{r}' \in \Omega} \underline{Q}(\mathbf{r}') \underline{G}(\mathbf{r}|\mathbf{r}') d\mathbf{r}' + \underline{p}_h(\mathbf{r})$$

↓ 離散化

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



N : # of grid points within Ω , M : # of microphones, $N \gg M$
 $\underline{y} \in \mathbb{C}^M$: Observed signal,
 $\underline{x} \in \mathbb{C}^N$: Distribution of monopole components,
 $\underline{D} \in \mathbb{C}^{M \times N}$: Dictionary matrix of Green's functions
 $\underline{h} \in \mathbb{C}^M$: Ambient components

\mathbf{x} は少数の要素にのみ非ゼロ成分を持つ

スパース音場分解

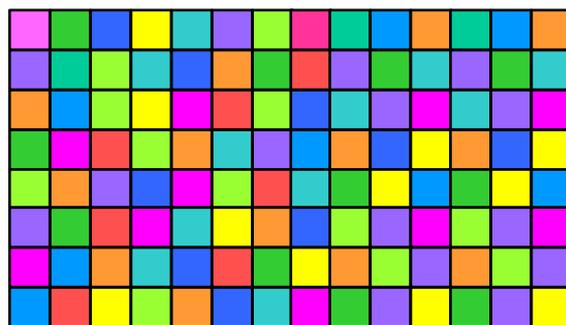
➤ 行列形式での表現

$$\underline{\mathbf{y}} \in \mathbb{C}^M$$



=

$$\underline{\mathbf{D}} \in \mathbb{C}^{M \times N}$$



$$\underline{\mathbf{x}} \in \mathbb{C}^N$$



$$\underline{\mathbf{h}} \in \mathbb{C}^M$$



+

➤ x のスパース性に基づく信号分解

$$\text{minimize } \|\underline{\mathbf{x}}\|_p^p \quad (p \leq 1)$$

$$\text{subject to } \underline{\mathbf{y}} = \underline{\mathbf{D}}\underline{\mathbf{x}}$$

x の ℓ_p -norm を最小化

スパース音場分解

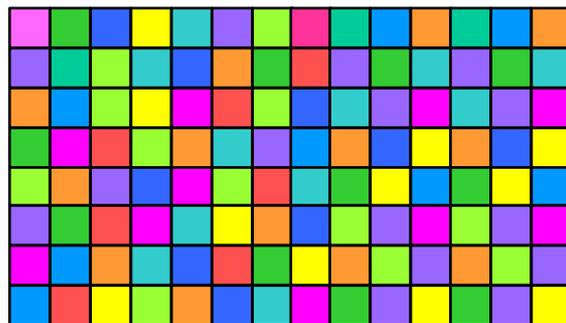
➤ 行列形式での表現

$$\underline{\mathbf{y}} \in \mathbb{C}^M$$



=

$$\underline{\mathbf{D}} \in \mathbb{C}^{M \times N}$$



音場の物理的な特性に基づく
スパース性の構造を考慮

$$\underline{\mathbf{h}} \in \mathbb{C}^M$$



+

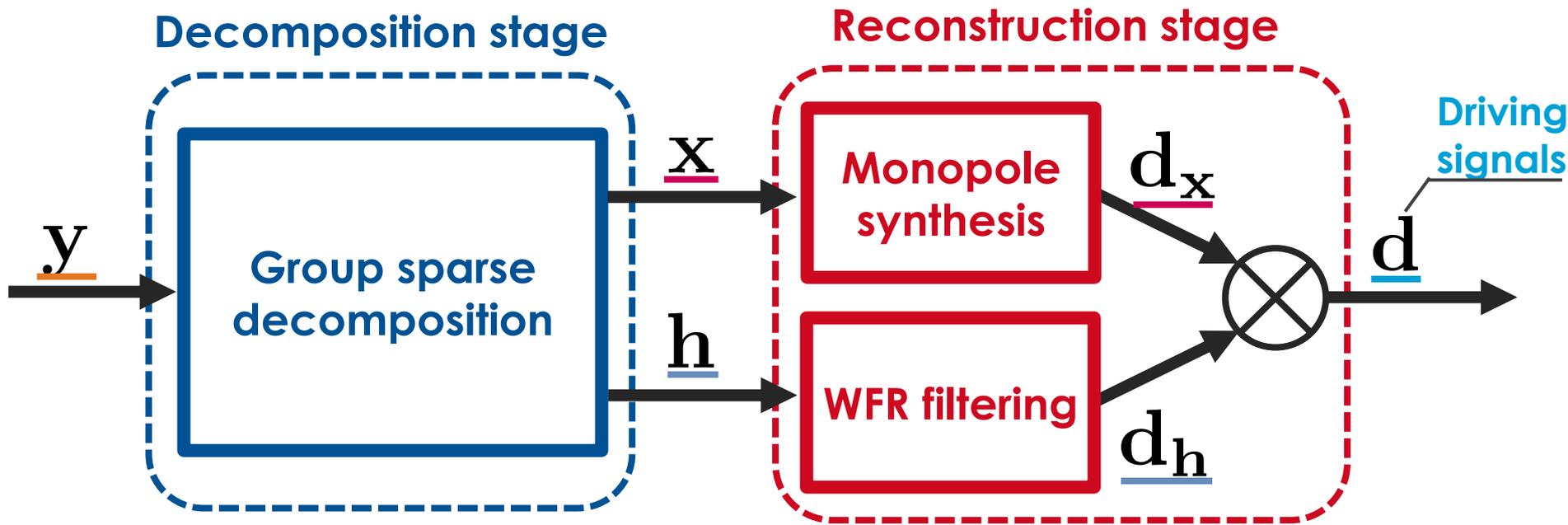


➤ グループスパース性に基づくロバスト化

- 複数時間フレーム：音源位置が不変
- 時間周波数：音源信号は広帯域
- 多重極音源：音源の指向特性

➤ FOCUSSの拡張アルゴリズムなどを適用可能 [Koyama+ ICASSP 2015]

信号変換のブロック図



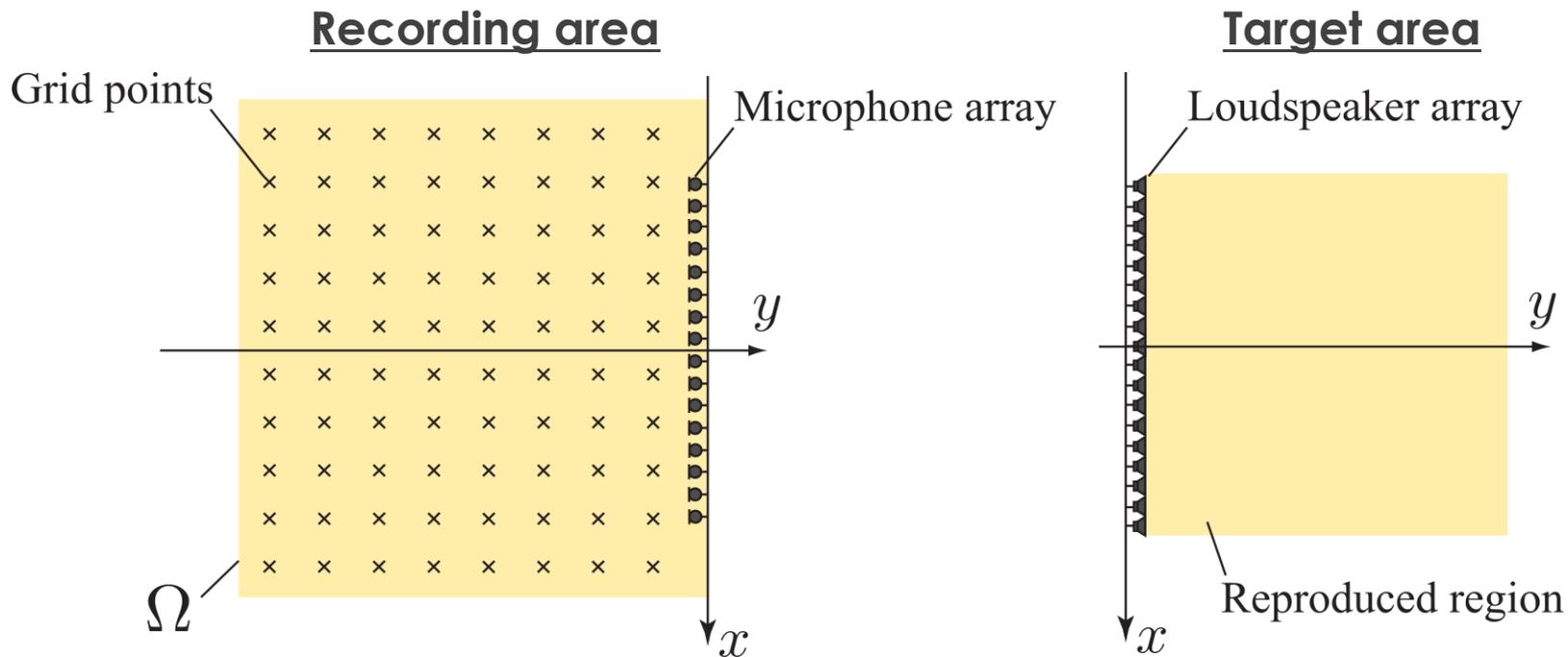
➤ Decomposition stage

- \underline{y} のグループスパース信号分解

➤ Reconstruction stage

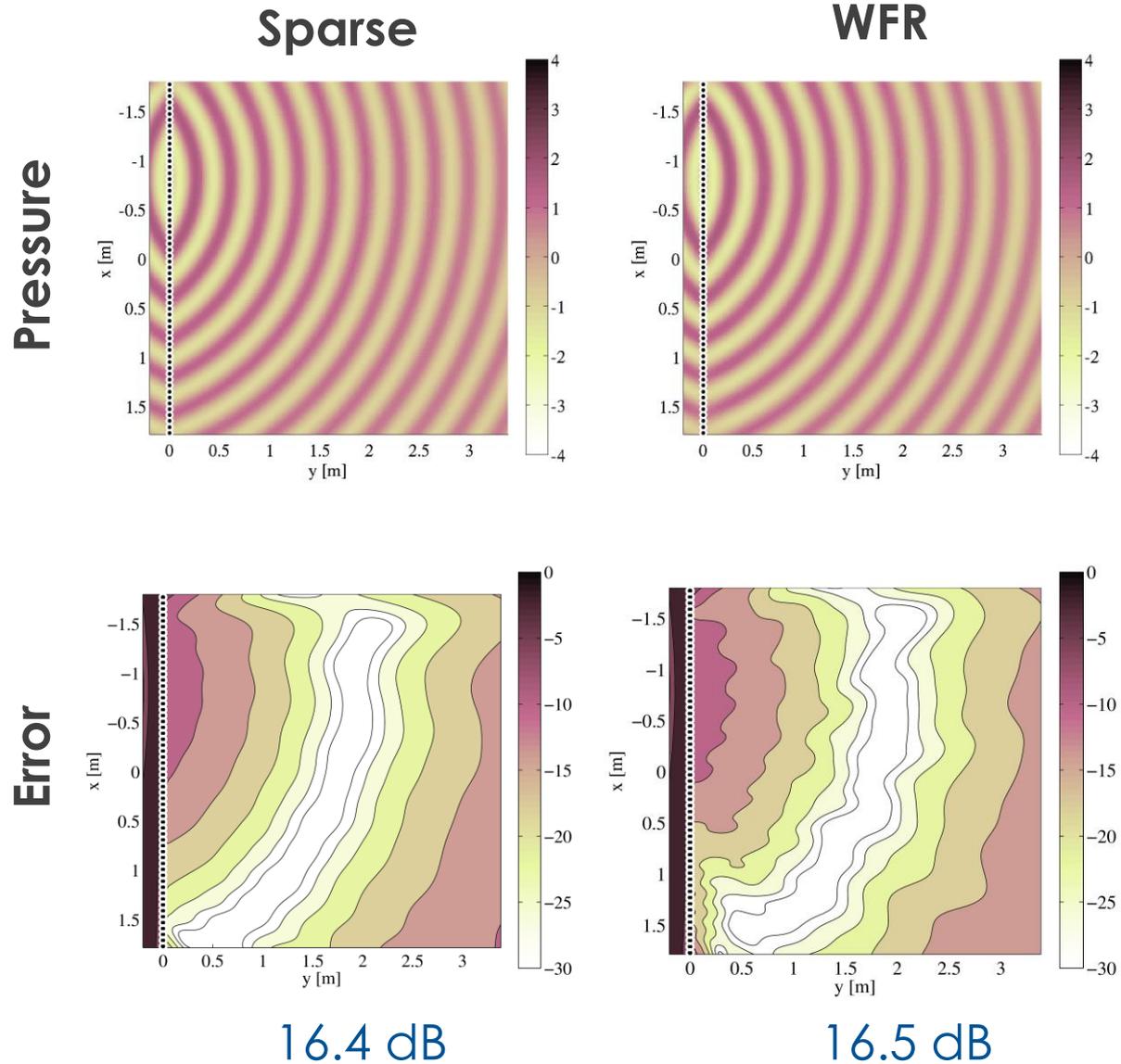
- \underline{x} と \underline{h} をそれぞれ駆動信号 \underline{d}_x と \underline{d}_h とに変換し、それらの和 \underline{d} をスピーカ出力とする

シミュレーション実験

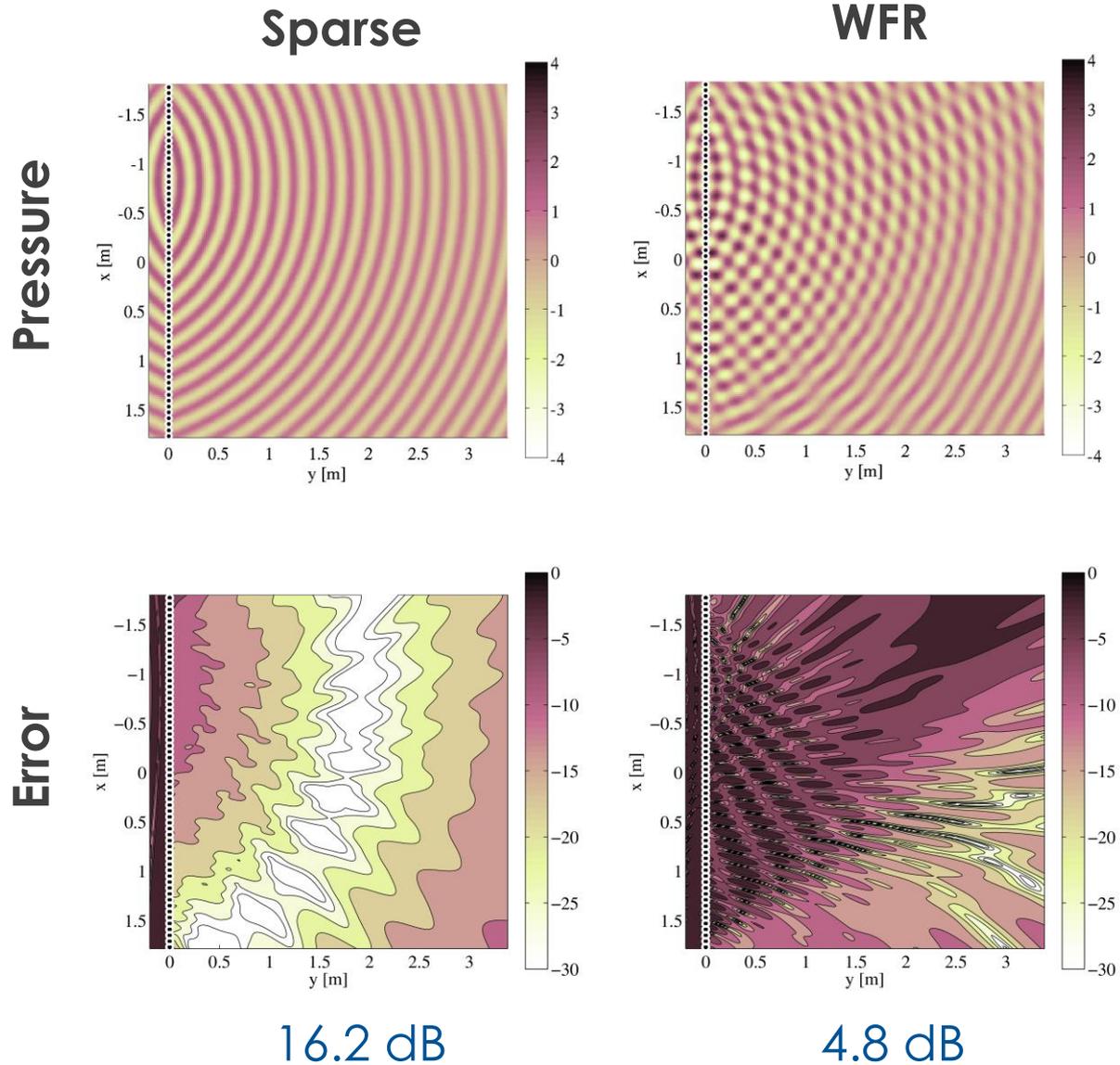


- スパース音場分解に基づく手法 (**Sparse**), WFRフィルタ法 (**WFR**)との比較
- マイク数 : 32 (12 cm間隔) , スピーカ数 : 64 (6 cm間隔)
- Ω : 4.0x4.0 mの矩形領域, 格子点: (10cm, 20cm) 間隔
- 音源位置: (-0.82, -0.86, 0.0) m

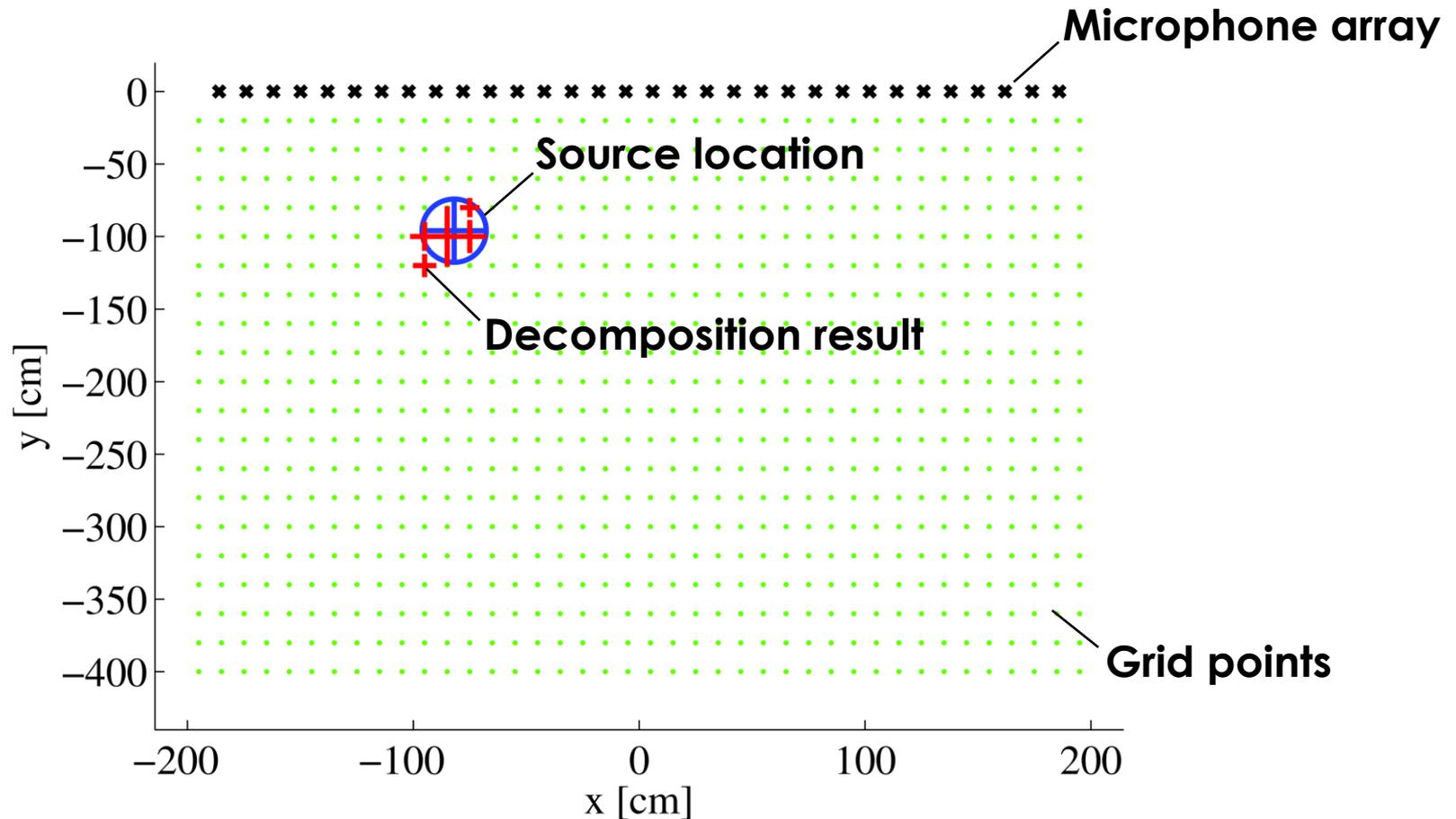
再現音圧分布 (1.0 kHz)



再現音圧分布 (1.8 kHz)



信号分解の結果 (1.8 kHz)



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

関連文献

- M. Elad, “Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing,” *Springer*, 2010.
- 樺島, “圧縮センシングの数理的側面～Compressed Sensing / Compressive Sensing～,” セミナー資料, 2013.
- 富岡, “スパース性に基づく機械学習,” 講談社, 2015.
- M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies, “Sparse Representations in Audio and Music: From Coding to Source Separation,” *Proc. IEEE*, vol. 98, no. 6, 2010.
- E. Ravelli, G. Richard, and L. Daudet, “Union of MDCT Bases for Audio Coding,” *IEEE Trans. ASLP*, vol. 16, no. 8, 2008.
- S. Koyama, S. Shimauchi, and H. Ohmuro, “Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts,” *Proc. IEEE ICASSP*, 2014.