

RAP-MUSICを用いた 反射音に頑健な音源位置推定

○岡本 拓磨 西村 竜一 岩谷 幸雄

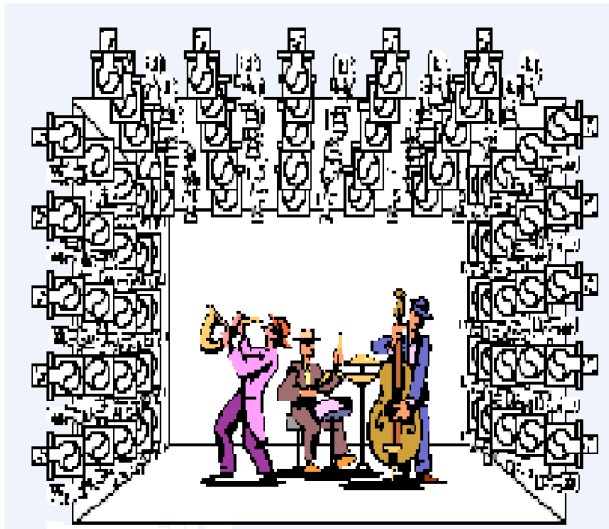
東北大学 電気通信研究所 / 大学院情報科学研究科

発表の概要

- 研究の背景
- 包囲型マイクロホンアレイの特性
- 空間平均化法の適応
- RAP-MUSICの考案
- まとめ

研究背景

- ある空間をそっくりそのまま保存し，別の部屋で再現するシステムの構築
 - FIR³ (FIeld R ecording, R ecognition and R eproduction)



時空間を越えた音空間の再現

FIR³ プロトタイプ

空間 A



包囲型マイクロホンアレイ (157 ch)

B&K Type 4591 (157本)

※実物は後ほどの研究施設紹介にて

空間 B



直接音

聴覚ディスプレイにより
仮想的に提示

耳介開放型音場再生デバイス LADOMi

位置センサ INTERSENSE IS-900

1次反射音



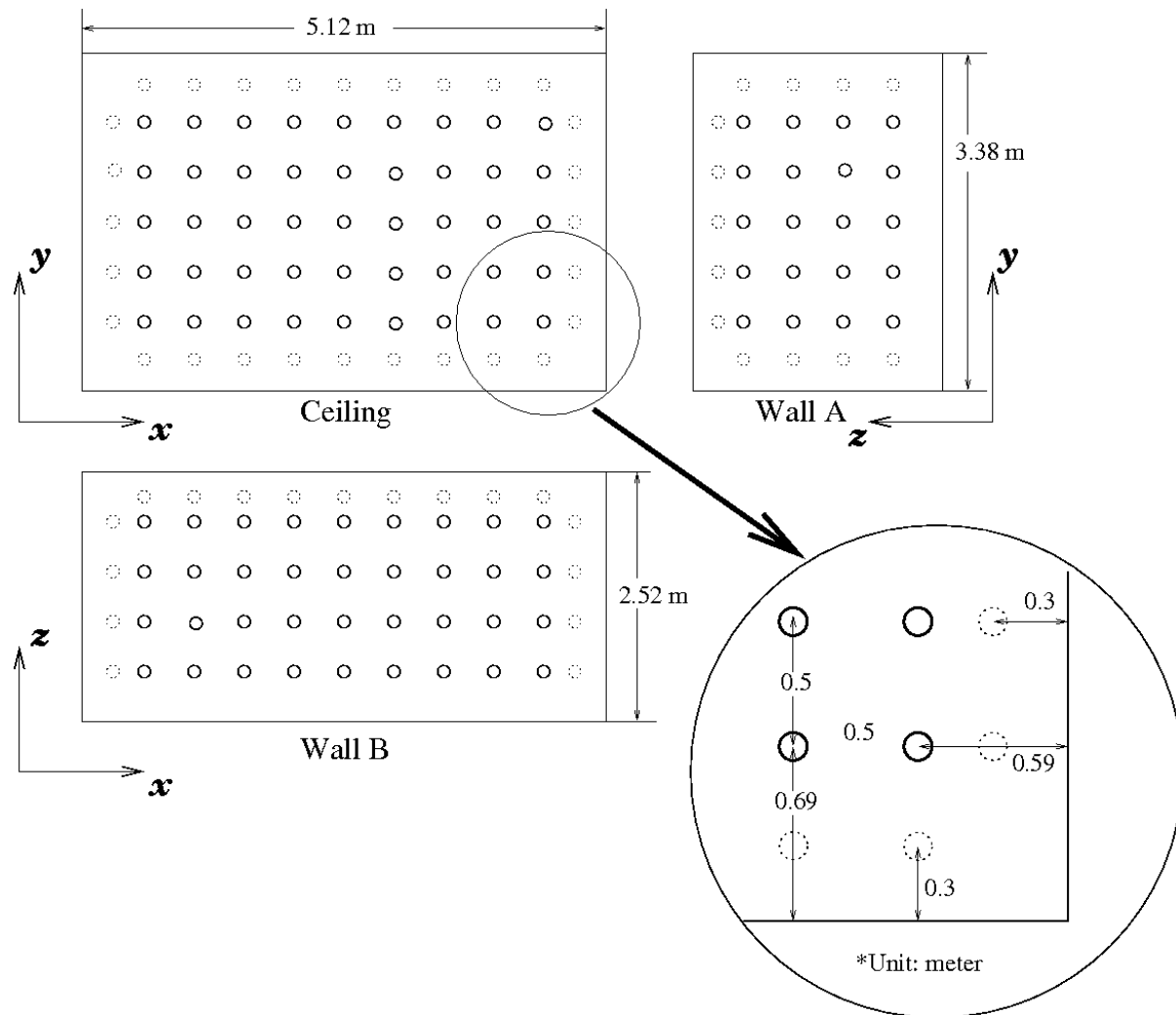
包囲型スピーカアレイ (157 ch)

FOSTEX FE38E (157本)

收音側で必要な技術

- 聴覚ディスプレイによる直接音の提示
 - 直接音の音源位置推定
 - 包囲型マイクロホンアレイによる音源位置推定
 - マイクロホン配置についての検討
 - 室内における音源位置推定
 - 直接音と波形が近い反射音による性能の低下
 - 反射音に頑健な音源位置推定手法の提案
 - 直接音の波形取り出し (=直接音の分離)
- スピーカアレイによる反射音の提示
 - 1次反射音の分離

包囲型マイクロホンアレイ



- 天井 45 ch
- 壁A 20 ch × 2
- 壁B 36 ch × 2

157 ch

マイクロホン間隔 50 cm

包囲型マイクロホンアレイの性質

- アレイ長が大きい
 - 到来信号を球面波として取り扱う
 - 到来方向ではなく、3次元位置推定が可能
- マイクロホン間隔
 - 50 cmのため、遅延和などの空間的なフーリエ変換に基づく手法では340 Hz以上でサイドローブ発生
 - 空間相関行列の固有空間構造を用いたMUSIC法による音源位置推定

MUSIC法 (Multiple Signal Classification)

各マイクロホンにおける入力信号

$$\mathbf{x}(\omega, t) = [x_1(\omega, t), x_2(\omega, t), \dots, x_n(\omega, t)]^T$$

空間相関行列

$$\mathbf{R} = \text{E}[\mathbf{x}(\omega, t)\mathbf{x}^H(\omega, t)]$$

導出



直交補空間の基底ベクトル

$$\mathbf{V}_n$$

各マイクロホン i と探査点 (x, y, z) 間の伝達関数

$$a_i(\omega, x, y, z) = \frac{1}{r_i} e^{j\omega \frac{r_i}{c}} \quad (1)$$

ステアリングベクトル

$$\mathbf{a} = [a_1, a_2, \dots, a_n]^T$$

MUSIC法

$$P(x, y, z) = \frac{1}{\|\mathbf{V}_n^H \mathbf{a}(x, y, z)\|^2} \quad (2)$$

$\mathbf{V}_n$ は音源位置の伝達関数と直交

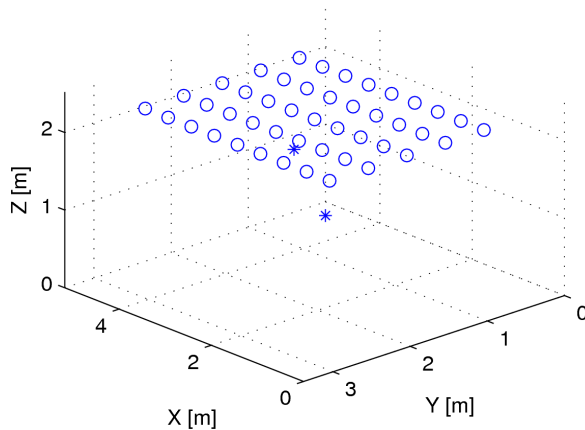


$\mathbf{a}$ が音源の位置に来た時だけ

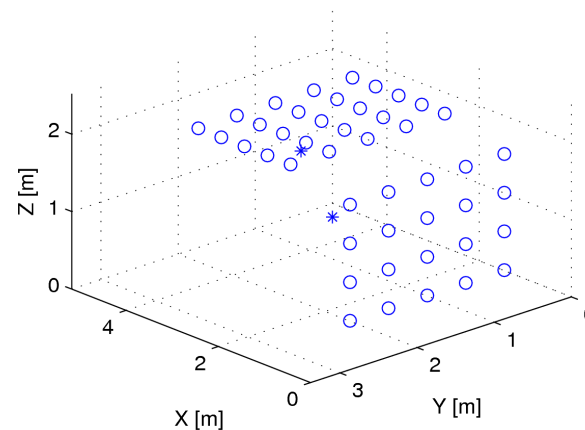
P は鋭いピークを形成する

包囲型マイクロホンアレイの性質

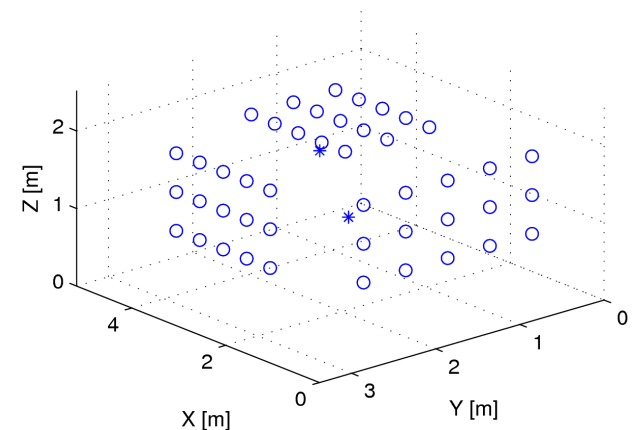
- マイクロホン配置が3次元配置
 - 配置の違いによる音源位置推定精度評価実験
 - 面の取り方の違いによる性能の違いを検討



パターン1 (1面)



パターン2 (2面)



パターン3 (3面)

※各パターン共に45 chを使用

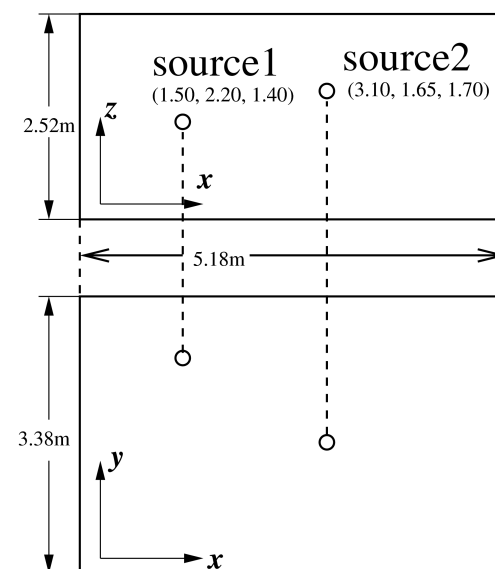
MUSIC法による音源位置推定実験

• 実験条件

- 音源1 男性音声 (a/i/kya/ku) 63.7 dB Lceq
- 音源2 女性音声 (i/chi/yu/u) 50.4 dB Lceq

(親密度別単語理解度試験用音源データベース NTTアドバンステクノロジー株式会社)

- サンプリング周波数 48000 Hz
- 信号の長さ 1 s
- 窓関数 ハミング窓 (512 ポイント)
- 解析周波数帯域 500 ~ 8000 Hz



音源位置

$$k = \frac{\text{各探査位置における } P}{\text{正しい音源位置における } P} \quad (3)$$

※音源位置で1, その他の位置では0に近い値であるほど, Pは鋭いピークを持つ

実験風景

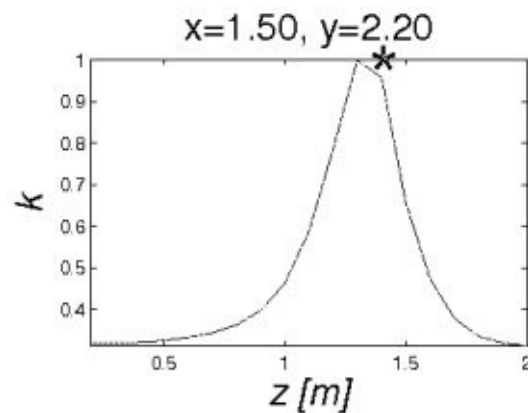


残響時間 0.15 s

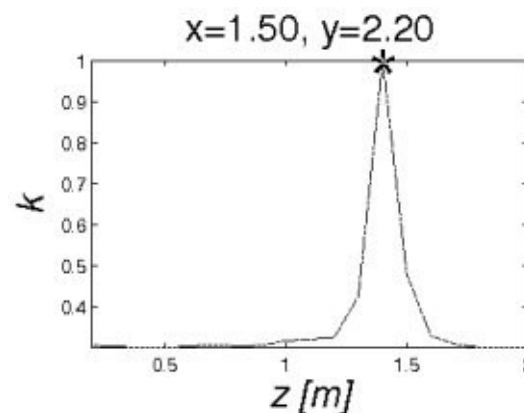
32面体点音源スピーカ (直径 7 cm)を使用

MUSIC法による音源位置推定結果

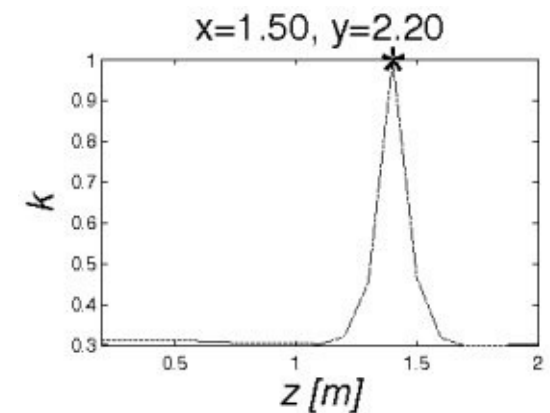
- (x, y) 座標を固定して z 座標で捜査した探査結果 (音源1)



パターン1 (1面)



パターン2 (2面)



パターン3 (3面)

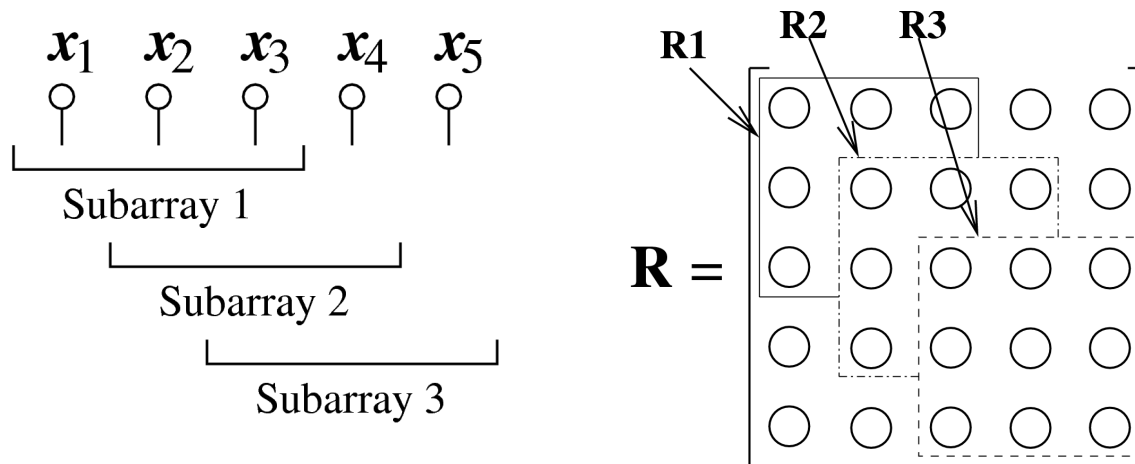
- 1面においては面と垂直方向の推定精度が悪い
- 2面以上では正しい推定が可能

反射音に頑健な音源位置推定

- 空間平均化法 (T. Shan ら 1985)

- アレイをいくつかのサブアレイに分割し，サブアレイの空間相関行列を平均化したものを用いてMUSICを適用

↑
フルランクまで回復



$$\bar{\mathbf{R}} = \sum_{n=1}^N \frac{1}{N} \mathbf{R}_n \quad (4)$$

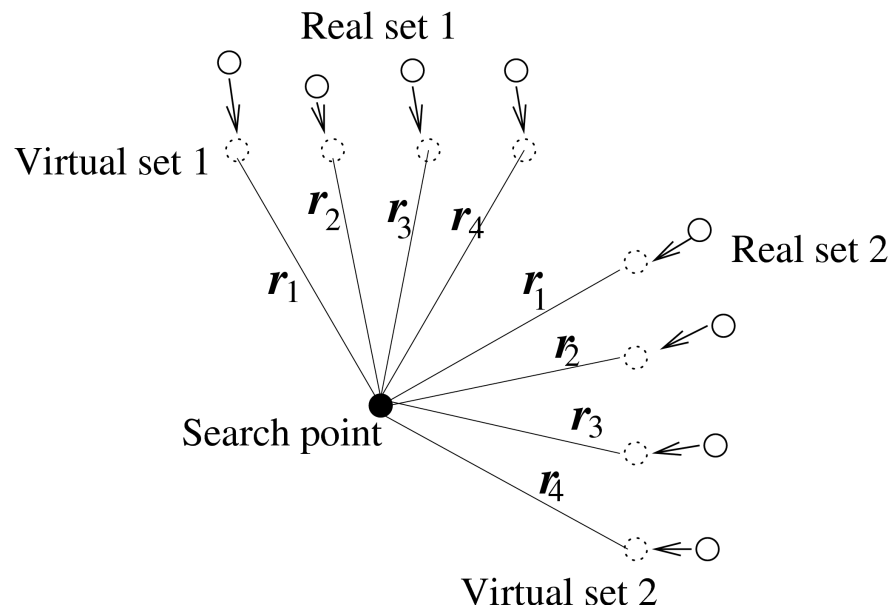
- サブアレイの形が全て同じであれば，方向推定の際には，どのサブアレイのステアリングベクトルも全て等しいことを利用
- 不規則形状アレイ，球面波仮定の音源位置推定にはそのまま適応できない

包囲型マイクロホンアレイへの 空間平均化法の適用

- 不規則形状アレイへの適応

- 補完処理によってサブアレイの信号を仮想的に作成

(B. Friedlander ら 1992)



補完処理 = 振幅 & 位相補正

仮想チャネルにおける信号

$$x'_i(t) = \frac{r_i}{r'_i} x_i(t) e^{-j\omega(\frac{r_i - r'_i}{c})} \quad (5)$$

仮想チャネルより構成されるサブアレイを用いて
空間相関行列を平均化する

包囲型マイクロホンアレイへの 空間平均化法の適用

- 3次元位置推定への適用

- 各探査点と各マイクロホンの距離は探査点が変化することによって変化する
 - 探査点ごとに仮想チャンネルを計算し直し、空間平均化法を適用する

$$x'_i(t) = \frac{r_i}{r'_i} x_i(t) e^{-j\omega(\frac{r_i - r'_i}{c})} \quad (5)$$

- 相関行列はフルランクまで回復しない
- 相関行列の直接音が形成する固有値に反射音の要素が入り込まないことが重要

入力信号の定義

- ある点に音源 $S(t)$ ，別の点に干渉波 $C(t)$

- マイクロホン i と音源および干渉波との距離，到来時間

- r_{Ci} , r_{Si} $\tau_{Si} = r_{Si}/c$, $\tau_{Ci} = r_{Ci}/c$

- マイクロホン i での入力信号

- $$x_i(t) = \frac{1}{r_{Si}} S(t) e^{-j\omega\tau_{Si}} + \frac{1}{r_{Ci}} C(t) e^{-j\omega\tau_{Ci}} \quad (6)$$

- 音源と仮想チャネルまでの距離，到来時間

- r'_S , $\tau'_{Si} = r'_{Si}/c$

- 仮想チャネル i での入力信号

- $$\begin{aligned} x'_i(t) &= \frac{r_{Si}}{r'_{Si}} x_i(t) e^{-j\omega(\tau'_{Si} - \tau_{Si})} \\ &= \frac{1}{r'_{Si}} S(t) e^{-j\omega\tau'_{Si}} + \frac{r_{Si}}{r'_{Si} r_{Ci}} C(t) e^{-j\omega(\tau_{Ci} - \tau_{Si} + \tau'_{Si})} \end{aligned} \quad (7)$$

16

空間相関行列の検討

- 平均化を適用しない場合の相関行列

$$\begin{aligned} R(p, q) &= \mathbb{E}[x_p(t)x_q(t)^*] \\ &= \frac{1}{r_{Sp}r_{Sq}} P_S e^{-j\omega(\tau_{Sp}-\tau_{Sq})} && \left. \begin{array}{l} \\ \end{array} \right\} \text{直接音} \\ &\quad + \frac{1}{r_{Sp}r_{Cq}} \mathbb{E}[S(t)C(t)^*] e^{-j\omega(\tau_{Sp}-\tau_{Cq})} \\ &\quad + \frac{1}{r_{Cp}r_{Sq}} \mathbb{E}[C(t)S(t)^*] e^{-j\omega(\tau_{Cp}-\tau_{Sq})} && \left. \begin{array}{l} \\ \end{array} \right\} \text{反射音} \\ &\quad + \frac{1}{r_{Cp}r_{Cq}} P_C e^{-j\omega(\tau_{Cp}-\tau_{Cq})} \end{aligned} \tag{8}$$

$$P_S = \mathbb{E}[S(t)S(t)^*]$$

$$P_C = \mathbb{E}[C(t)C(t)^*]$$

空間相関行列の検討

- 簡略化のため, $r'_{Si} \approx r_{Si}$ として考える

- 第2項から第4項は多重化されない

$$\begin{aligned}\bar{R}'(p, q) \approx & \frac{1}{r_{Sp}r_{Sq}} P_S e^{-j\omega(\tau_{Sp}-\tau_{Sq})} & \left. \vphantom{\frac{1}{r_{Sp}r_{Sq}}} \right\} & \text{直接音} \\ & + \frac{1}{N} \frac{1}{r_{Sp}r_{Cq}} E[S(t)C(t)^*] e^{-j\omega(\tau_{Sp}-\tau_{Cq})} \\ & + \frac{1}{N} \frac{1}{r_{Cp}r_{Sq}} E[C(t)S(t)^*] e^{-j\omega(\tau_{Cp}-\tau_{Sq})} & \left. \vphantom{\frac{1}{N}} \right\} & \text{反射音} \\ & + \frac{1}{N} \frac{1}{r_{Cp}r_{Cq}} P_C e^{-j\omega(\tau_{Cp}-\tau_{Cq})}\end{aligned}\tag{11}$$



干渉波の影響を低減

RAP-MUSIC

- 3次元位置推定における空間平均化法

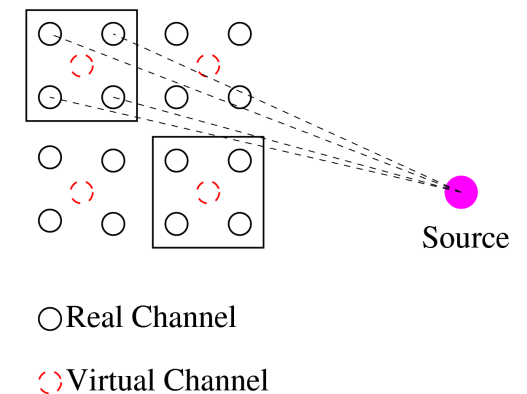
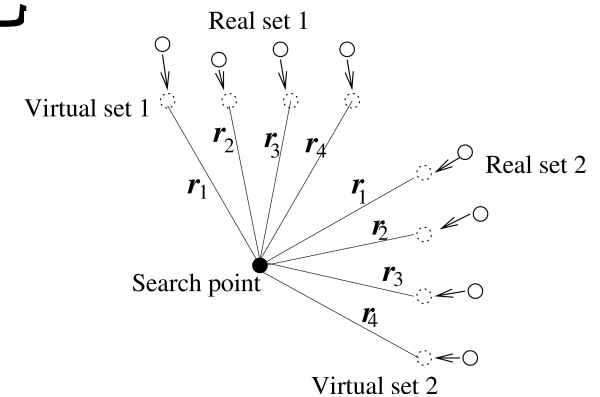
- 探査点ごとに遅延量を変化
- 各仮想チャネルにおいて、探査点からの信号は同相化



サブアレイを全て同じ位置にする

- RAP-MUSIC (Re-arrangement And Pre-smoothing for MUSIC)

- 相関行列を平均化するのではなく、仮想チャネルの信号を最初から平均化
- MUSICを行う前に仮想チャネルのサブアレイに遅延和を適用



RAP-MUSIC

- N 個のチャネルで同相化された信号の和 $\bar{x}'_i(t)$

$$\begin{aligned}\bar{x}'_i(t) &= \sum_{n=1}^N \frac{1}{N} x'_i(t) \\ &= \frac{1}{r'_{Si}} S(t) e^{-j\omega\tau'_{Si}} + \sum_{n=1}^N \frac{1}{N} \frac{r_{Si}}{r'_{Si} r_{Cn}} C(t) e^{-j\omega(\tau_{Cn} - \tau_{Si} + \tau'_{Si})}\end{aligned}\quad (12)$$

第2項は全て位相が異なるため、 $1/N$ になると仮定



$$\bar{x}'_i(t) \approx \frac{1}{r'_{Si}} S(t) e^{-j\omega\tau'_{Si}} + \frac{1}{N} \frac{r_{Si}}{r'_{Si} r_{Ci}} C(t) e^{-j\omega(\tau_{Ci} - \tau_{Si} + \tau'_{Si})}\quad (13)$$

RAP-MUSIC

- RAP-MUSICの空間相関行列

$$\begin{aligned}\bar{R}'(p, q) &= \text{E}[x'_p(t)x'_q(t)^*] \\ &= \frac{1}{r_{Sp}r_{Sq}} P_S e^{-j\omega(\tau_{Sp}-\tau_{Sq})} \quad \left. \vphantom{\frac{1}{r_{Sp}r_{Sq}} P_S e^{-j\omega(\tau_{Sp}-\tau_{Sq})}} \right\} \text{直接音} \\ &\quad + \frac{1}{N} \frac{1}{r_{Sp}r_{Cq}} \text{E}[S(t)C(t)^*] e^{-j\omega(\tau_{Sp}-\tau_{Cq})} \\ &\quad + \frac{1}{N} \frac{1}{r_{Cp}r_{Sq}} \text{E}[C(t)S(t)^*] e^{-j\omega(\tau_{Cp}-\tau_{Sq})} \quad \left. \vphantom{\frac{1}{N} \frac{1}{r_{Cp}r_{Sq}} \text{E}[C(t)S(t)^*] e^{-j\omega(\tau_{Cp}-\tau_{Sq})}} \right\} \text{反射音} \\ &\quad + \frac{1}{N^2} \frac{1}{r_{Cp}r_{Cq}} P_C e^{-j\omega(\tau_{Cp}-\tau_{Cq})} \quad \left. \vphantom{\frac{1}{N^2} \frac{1}{r_{Cp}r_{Cq}} P_C e^{-j\omega(\tau_{Cp}-\tau_{Cq})}} \right\} \text{反射音}\end{aligned} \quad (14)$$

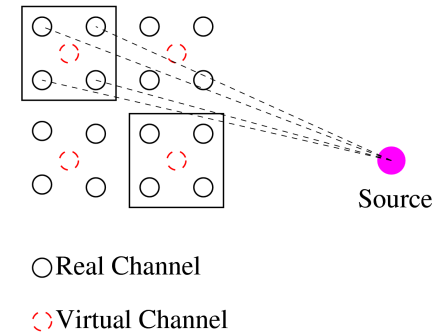


空間平均化法よりも干渉波の影響を低減

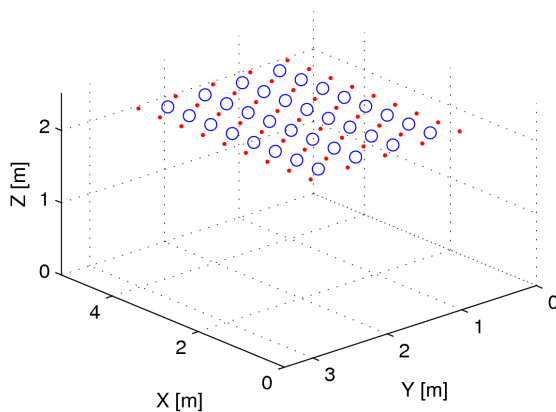
実験による性能評価

- 実験条件は先ほど同様

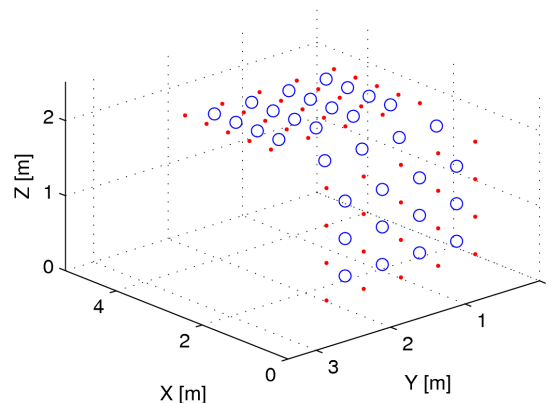
- 空間平均化法, RAP-MUSICの仮想チャネルは4つずつの各マイクロホンの重心位置に設定, 4つで多重化



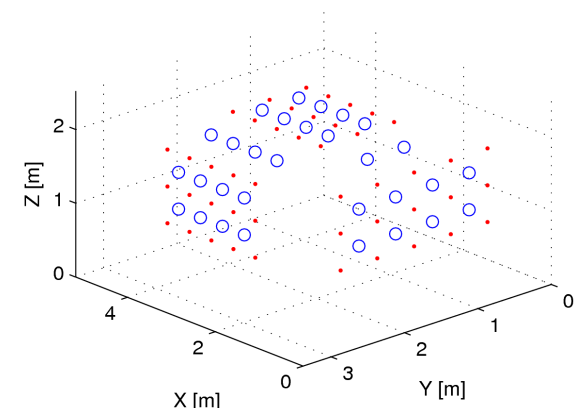
- 各パターンについて, MUSICのみ, 空間平均化法, RAP-MUSICの性能を比較する



パターン1 (32 ch)



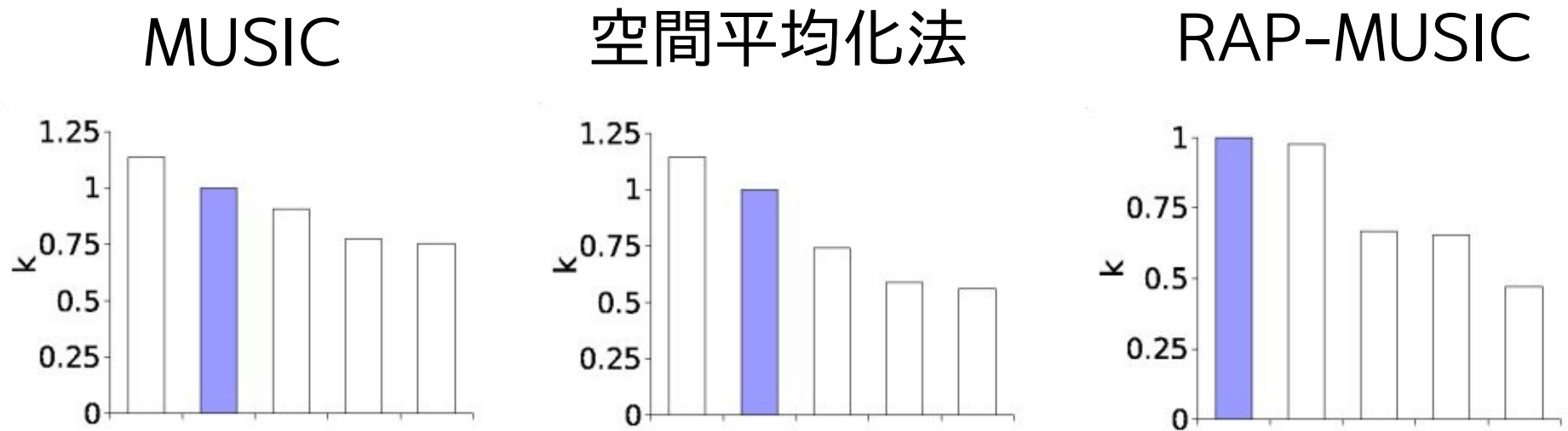
パターン2 (32 ch)



パターン3 (30 ch)

実験結果

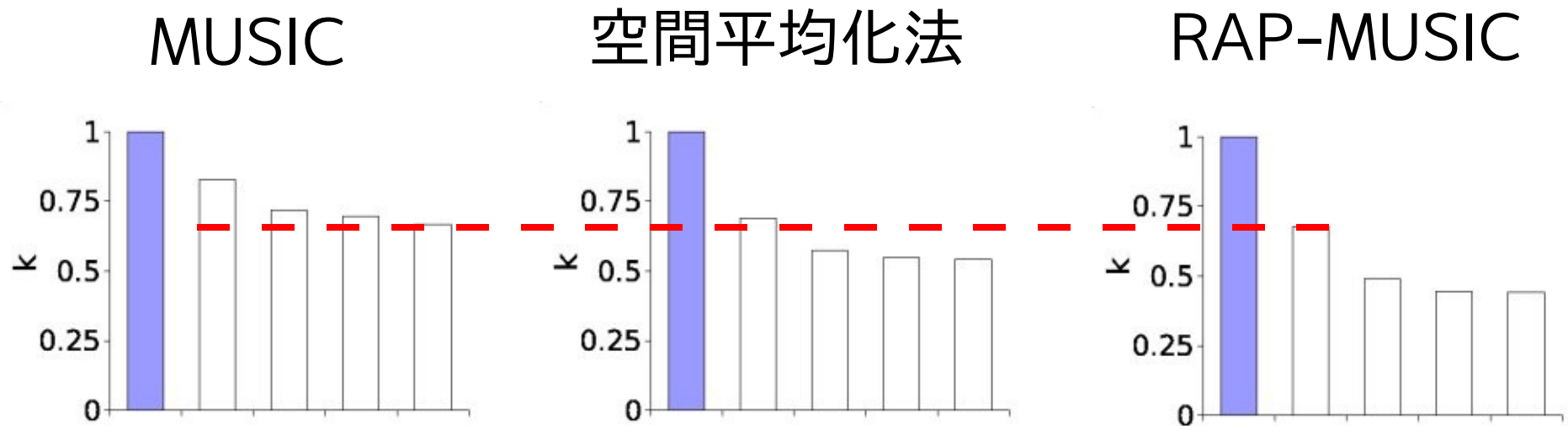
- 音源1における性能の比較 (パターン1)



RAP-MUSICにより推定精度改善

実験結果

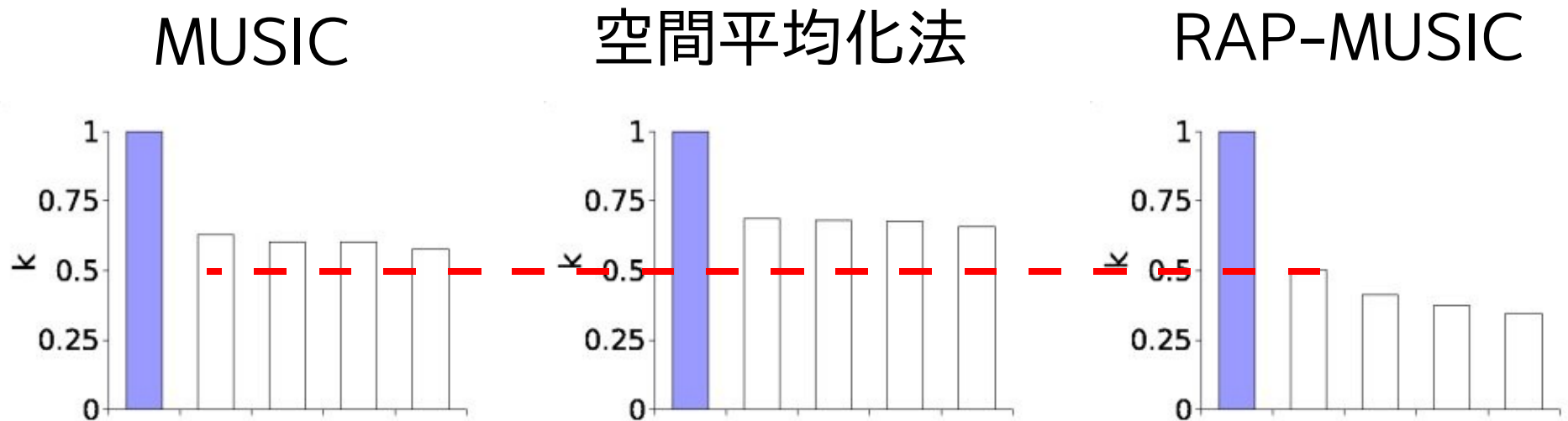
- 音源1における性能の比較 (パターン2)



RAP-MUSICにより推定精度改善

実験結果

- 音源1における性能の比較 (パターン3)



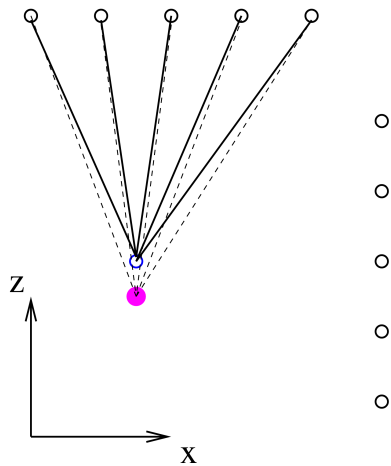
RAP-MUSICにより推定精度改善

まとめ

- 包囲型マイクロホンアレイを用いた音源位置推定に関する検討
 - 2次元配置よりも3次元配置の方が高精度に推定できることを示した
 - 反射音の効果を低減する空間平均化法を3次元位置推定に適用した
 - 空間平均化法を発展させたRAP-MUSICを考案し、従来法よりも高精度な音源位置推定を可能とした

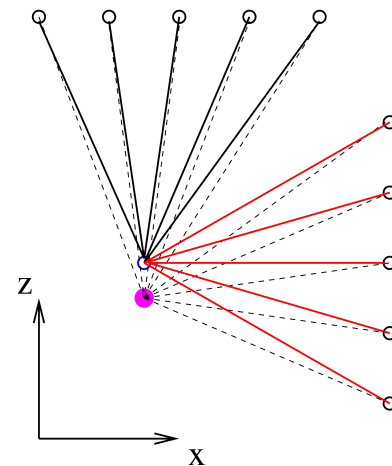
マイクロホン配置に関する考察

- 2面や3面では精度よく推定可能
- 1面では音源のz方向の推定精度だけが低い
 - ステアリングベクトル $\mathbf{a}(x, y, z)$ についての検討



1面

$$\mathbf{a}(x, y, z + dz) \simeq c\mathbf{a}(x, y, z)$$



2面以上

$$\mathbf{a}(x, y, z + dz) \neq \mathbf{a}(x, y, z)$$

空間相関行列の検討

- 平均化を適用する場合のサブアレイの相関行列

$$\begin{aligned} R'(p, q) &= E[x'_p(t)x'_q(t)^*] \\ &= \frac{1}{r'_{Sp}r_{Sq}} P_S e^{-j\omega(\tau'_{Sp}-\tau'_{Sq})} \\ &\quad + \frac{r_{Sq}}{r'_{Sp}r'_{Sq}r_{Cq}} E[S(t)C(t)^*] e^{-j\omega(\tau'_{Sp}-\tau_{Cq}+\tau_{Sq}-\tau'_{Sq})} \\ &\quad + \frac{r_{Sp}}{r'_{Sp}r'_{Sq}r_{Cq}} E[C(t)S(t)^*] e^{-j\omega(\tau_{Cp}-\tau_{Sp}+\tau'_{Sp}-\tau'_{Sq})} \\ &\quad + \frac{r_{Sp}r_{Sq}}{r'_{Sp}r_{Cp}r'_{Sq}r_{Cq}} P_C e^{-j\omega(\tau_{Cp}-\tau_{Sp}+\tau'_{Sp}-\tau_{Cq}+\tau_{Sq}-\tau'_{Sq})} \end{aligned} \tag{9}$$

空間相関行列の検討

- $\bar{R} = \sum_{n=1}^N \frac{1}{N} R_n$ によって平均化された相関行列

$$\begin{aligned}\bar{R}'(p, q) &= \sum_{n=1}^N \frac{1}{N} R(p, q) \\ &= \frac{1}{r'_{Sp} r_{Sq}} P_S e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \\ &\quad + \sum_{n=1}^N \frac{1}{N} \frac{r_{Sq}}{r'_{Sp} r'_{Sq} r_{Cn}} E[S(t)C(t)^*] e^{-j\omega(\tau'_{Sp} - \tau_{Cn} + \tau_{Sq} - \tau'_{Sq})} \\ &\quad + \sum_{n=1}^N \frac{1}{N} \frac{r_{Sp}}{r'_{Sp} r'_{Sq} r_{Cn}} E[C(t)S(t)^*] e^{-j\omega(\tau_{Cn} - \tau_{Sp} + \tau'_{Sp} - \tau'_{Sq})} \\ &\quad + \sum_{n=1}^N \frac{1}{N} \frac{r_{Sp} r_{Sq}}{r'_{Sp} r_{Cn} r'_{Sq} r_{Cn}} P_C e^{-j\omega(\tau_{Cn} - \tau_{Sp} + \tau'_{Sp} - \tau_{Cn} + \tau_{Sq} - \tau'_{Sq})}\end{aligned}$$

(10)