

包囲型マイクロホンアレイを用いた 高精度線形予測ブラインド残響除去に関する検討

岡本 拓磨, 岩谷 幸雄, 鈴木 陽一

東北大学 大学院情報科学研究科 / 電気通信研究所

はじめに

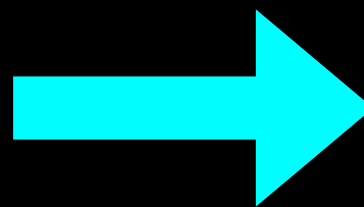
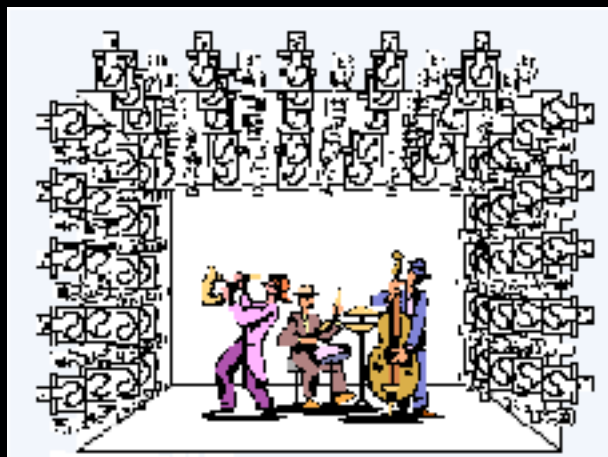
究極の目標

・音空間情報の厳密な保存・解析・再現

保存：空間全体の音情報を収録

解析：どこでどの音が鳴っているかを収録情報より解析

再現：保存した音空間情報をそっくりそのまま別の空間で再現



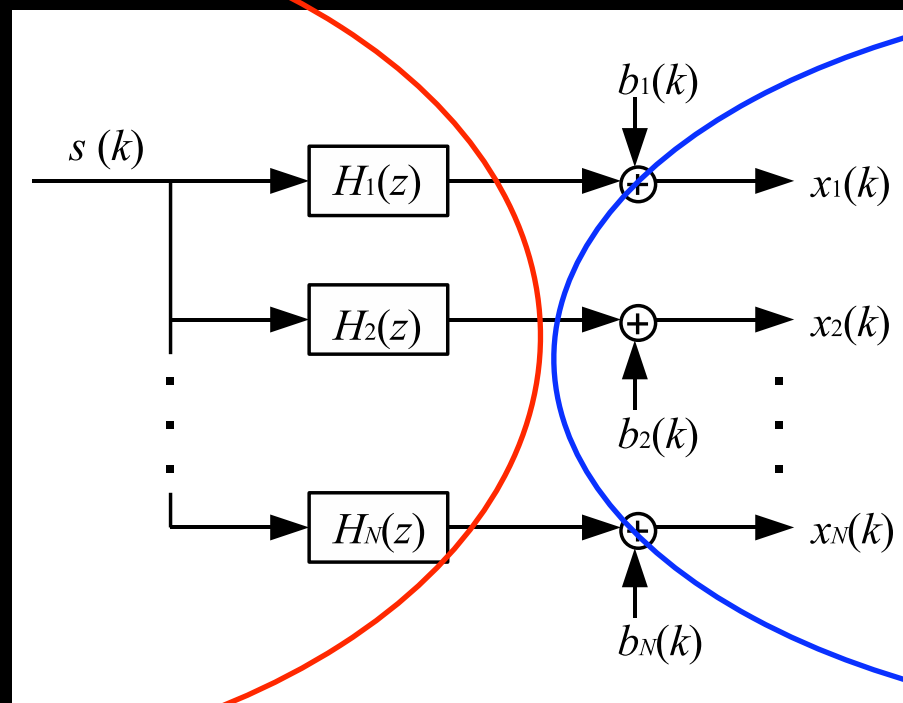
時空間を超えた音空間情報の再現

本研究の目的

包囲型マイクロホンアレイを用いた

室内音空間情報の高精度推定

- ・ 音源位置情報
- ・ 直接音情報
- ・ 反射音情報
- ・ 音源放射特性



SIMOモデル



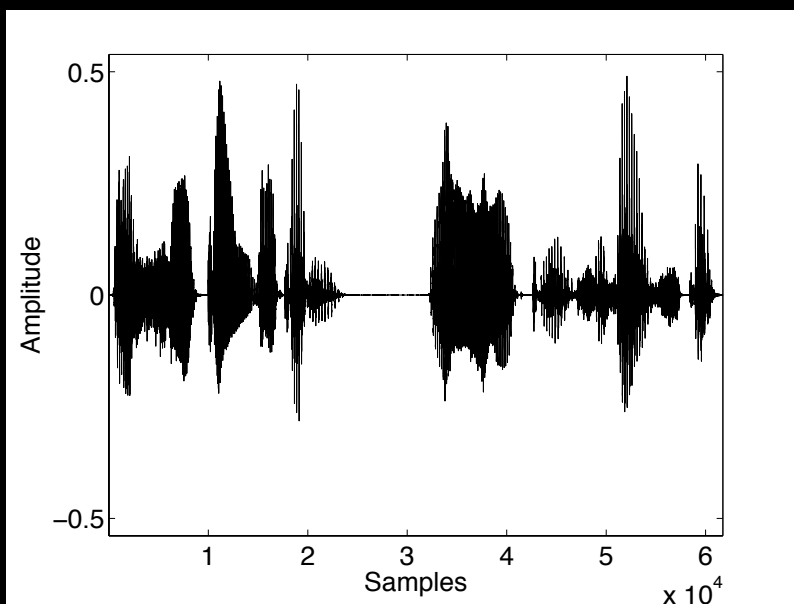
包囲型マイクロホンアレイ
157 ch

入力信号とマイクロホン位置情報のみから
室内音源特性を推定

多チャネル信号を用いた残響除去

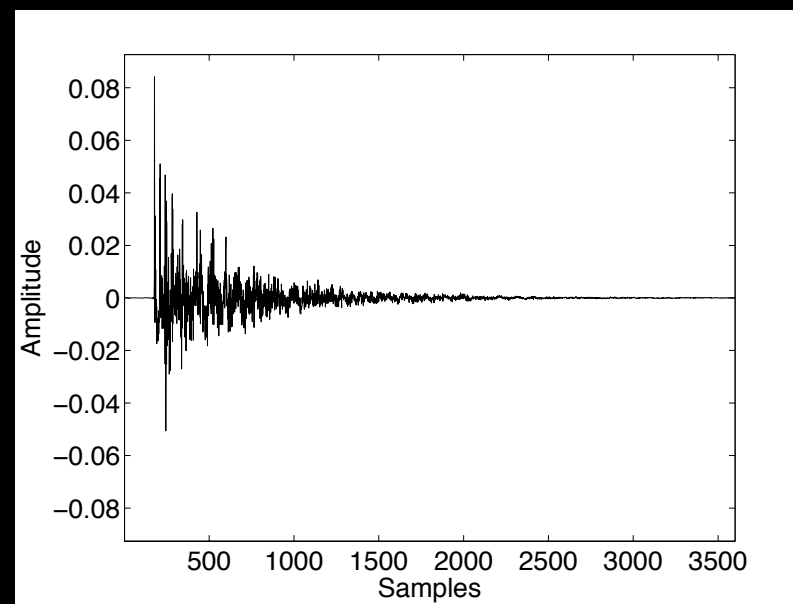
各マイクロホンでの信号

$$s(n) * h_i(n) = x_i(n)$$



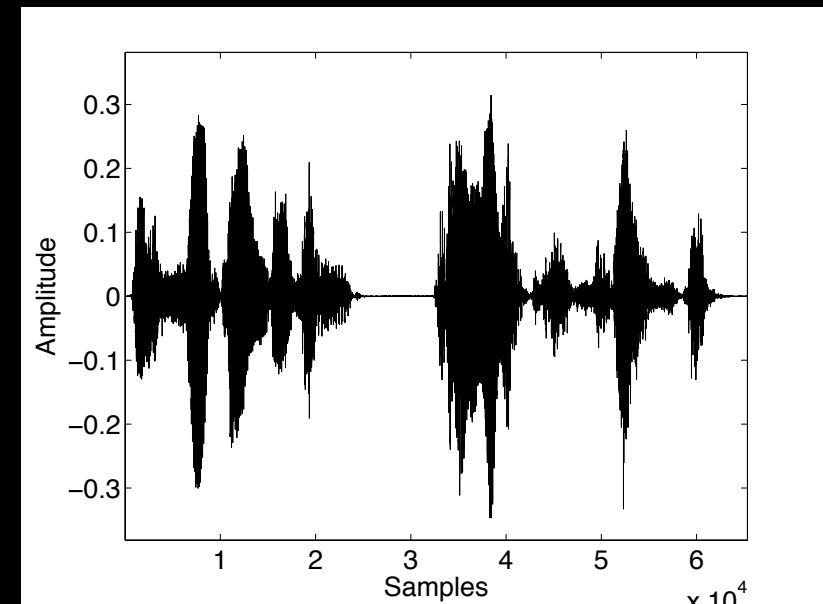
音源信号

*



残響(=インパルス応答)

=



入力信号



残響除去

先行研究

- ・ ビームフォーミング

- ・ マイクロホン, 音源位置が既知
- ・ 残響抑圧精度がマイクロホン間隔に依存

- ・ ブラインド残響除去

- ・ マイクロホン, 音源位置は未知でOK!
- ・ 残響抑圧精度がマイクロホン間隔に依存しない

- ・ ブラインドシステム同定→MINTによる手法

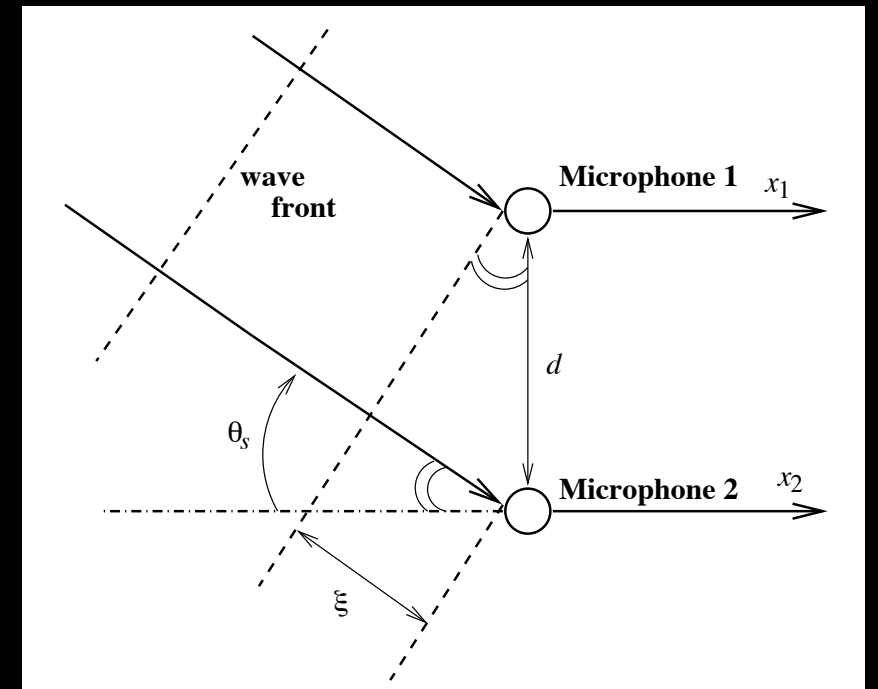
- ・ 入力信号のみからインパルス応答を推定
- ・ 推定したインパルス応答の逆フィルタをMINTにより計算
- ・ 逆フィルタにより残響を完全に除去

※1989 M. Miyoshi et.al

$$H_1(z)G_1(z) + H_2(z)G_2(z) = 1$$

- ・ 問題点

- ・ インパルスのタップ長がわからないと精度が悪い



先行研究 (2)

- ブラインドデコンボリューションによる残響除去
- Semi-blind-MINT (2001 K. Furuya et. al)
 - メリット
 - 入力信号から直接逆フィルタを算出
 - インパルス応答長が未知の場合でもそれより長いタップとして推定すればOK!
 - 背景雑音にもロバスト
 - デメリット
 - 入力信号が白色雑音を過程しているため、音声などでは抑圧性能劣化

線形予測を用いたブラインド残響除去

2005 M. Delcroix et. al

$$s(n) = \sum_{k=1}^N a_k s(n-k) + e(n)$$

音源信号をARモデルで表現

入力信号から音源信号のAR係数と予測残差を推定

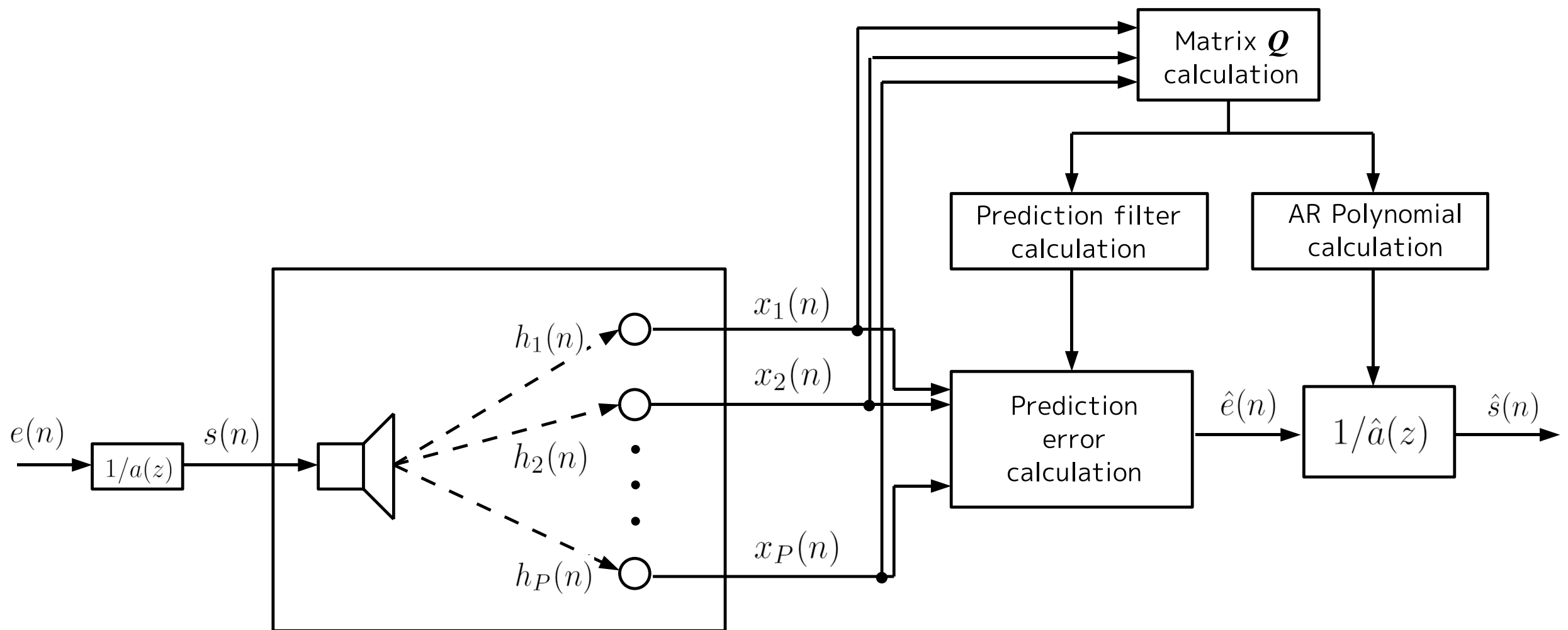
$$\mathbf{Q} = \left(E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\} \right)^+ E\{\mathbf{x}(n-1)\mathbf{x}^T(n)\}$$

$$\hat{e}(n) = x_1(n) - \mathbf{x}^T(n-1)\mathbf{w}$$

$$\mathbf{C} = \begin{pmatrix} a_1 & 1 & 0 & \cdots & 0 \\ a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 1 \\ a_N & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

$$f_c(\mathbf{Q}, \lambda) = f_c(\mathbf{C}, \lambda)$$

線形予測を用いたブラインド残響除去 (2)



Linear-predictive Multi-channel Equalization (LIME)

- ・ メリット
 - ・ インパルス応答長が未知の場合でも精度よく残響を除去
- ・ デメリット
 - ・ 背景雑音に弱い

ブラインド残響除去の課題

1. インパルス応答長が未知である
2. インパルス応答長が大きい (数千タップ)
3. 背景雑音の混入

上記3点をクリアした残響除去手法は
未だ存在しない

上記2点をクリアしているLIMEアルゴリズムを用いて
包囲型マイクロホンアレイでの残響除去を試みる

包囲型マイクロホンアレイを用いた 線形予測ブラインド残響除去

2007 M. Delcroix et. alの検討では…

- ・ 8 kHzサンプリング
- ・ 2～6 ch
- ・ 残響時間0.48 s (=3840taps)

今回の検討事項

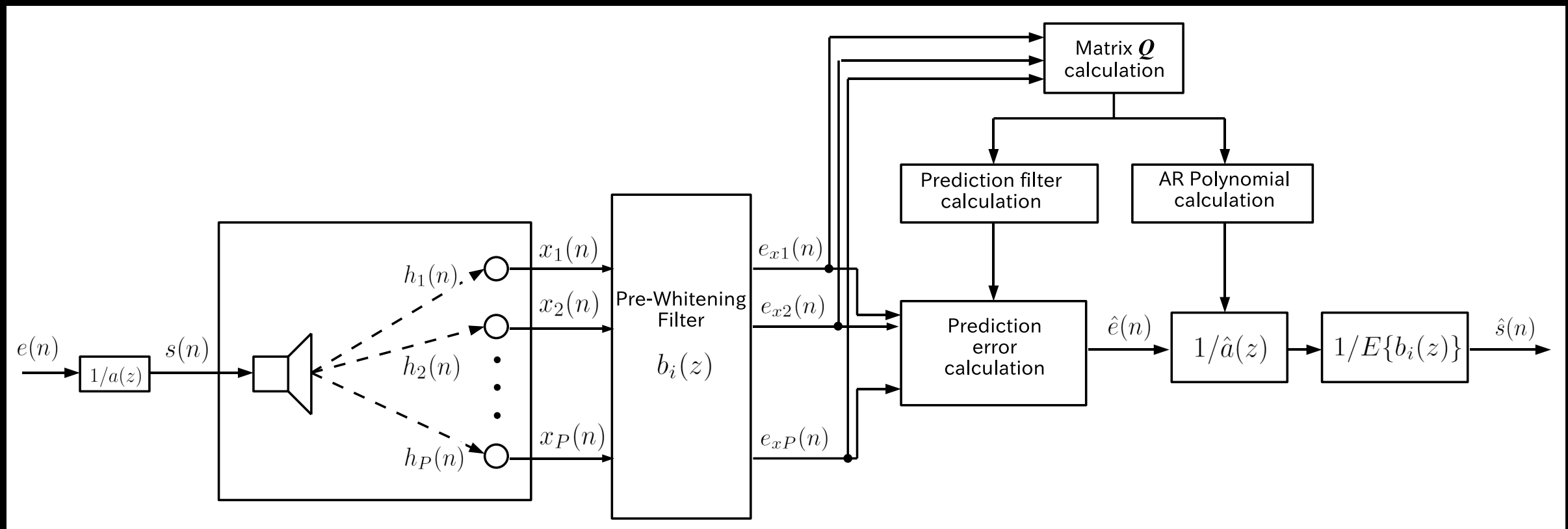
- ・ 高サンプリングレートでの実現 (24 kHz)
- ・ 4～157 chにおける性能評価
- ・ 放射特性を持つ音源に対する残響除去の効果

Pre-Whitening フィルタを用いた LIMEアルゴリズムの高サンプリング化

有色信号の場合、サンプリング周波数を上げると予測残差やAR係数が計算できなくなる

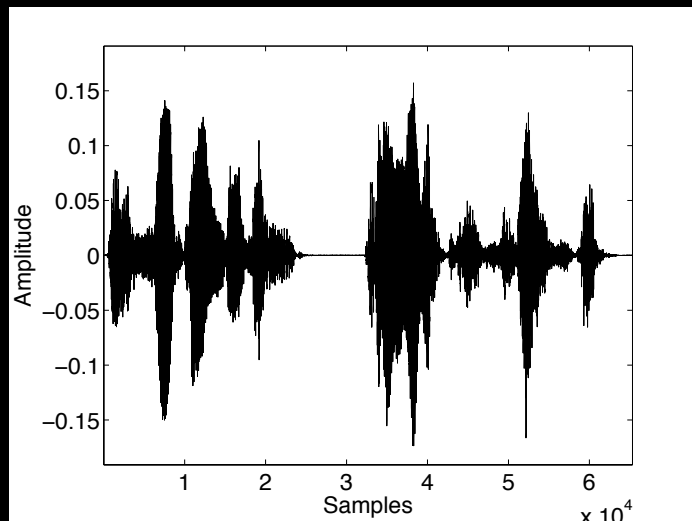
各入力信号のAR係数の平均のフィルタを各信号に畳み込み、白色化

高サンプリングでもしっかりと動作可能

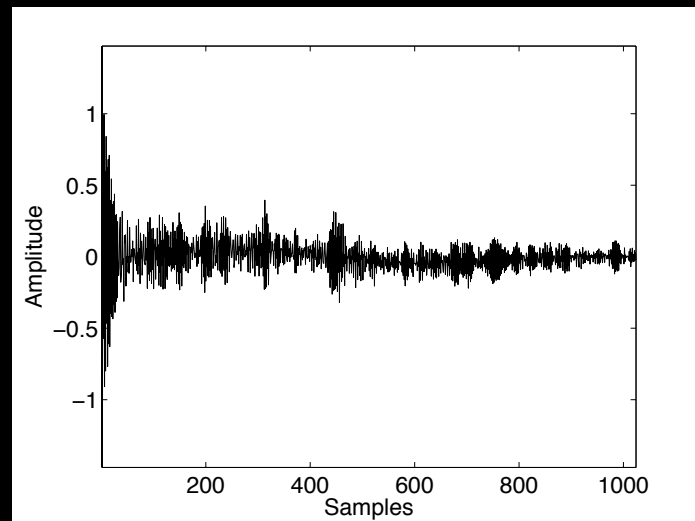


Pre-Whitening LIME

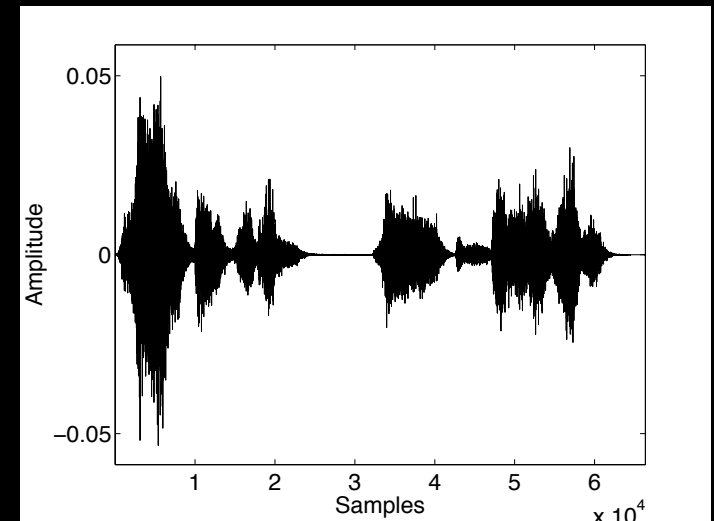
Pre-Whitening フィルタを用いた LIMEアルゴリズムの高サンプリング化 (2)



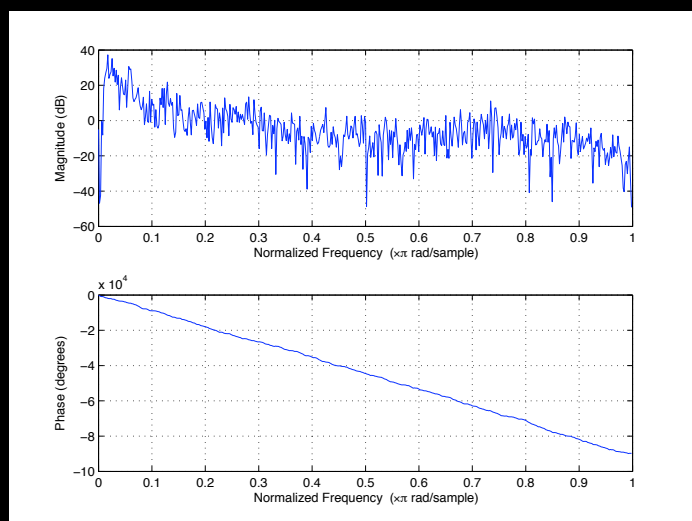
$$x_i(n)$$



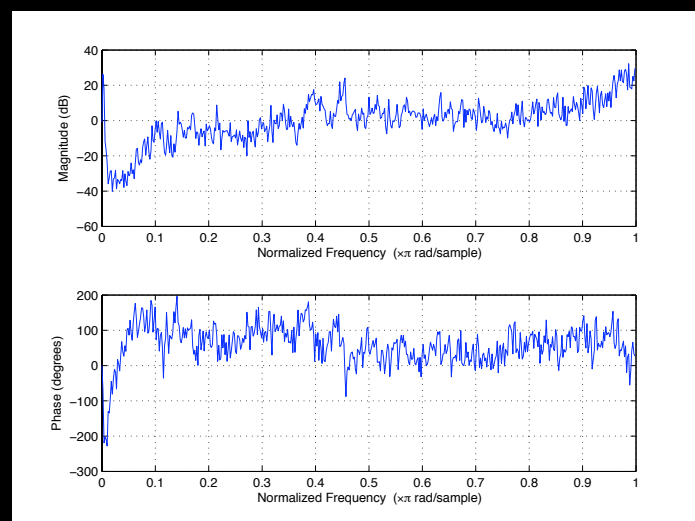
$$E\{b_i(z)\}$$



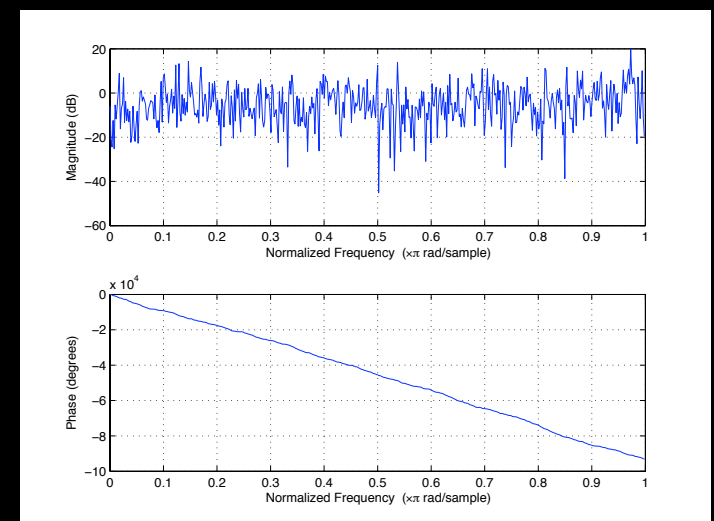
$$e_i(n)$$



$$X_i(f)$$

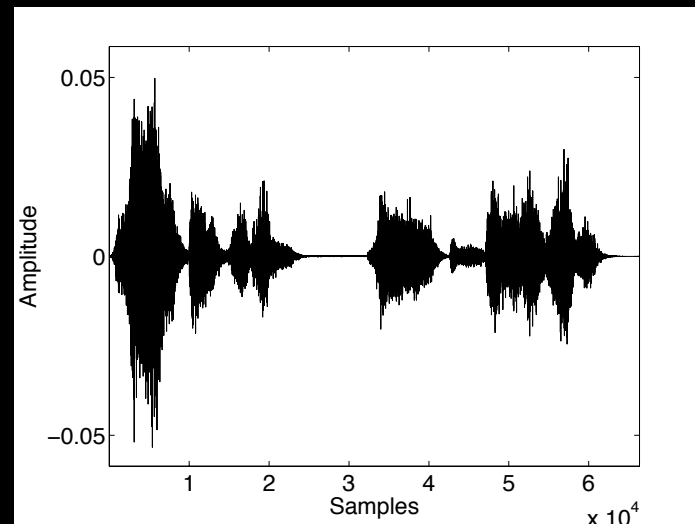


$$E\{B_i(f)\}$$



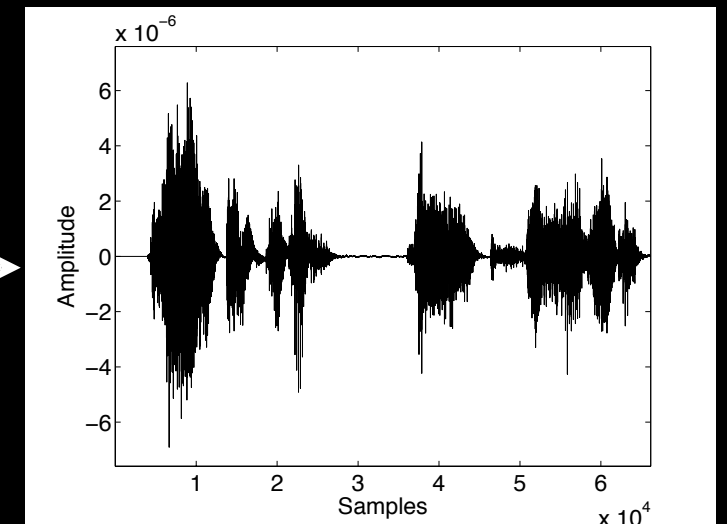
$$e_i(f)$$

Pre-Whitening フィルタを用いた LIMEアルゴリズムの高サンプリング化 (3)

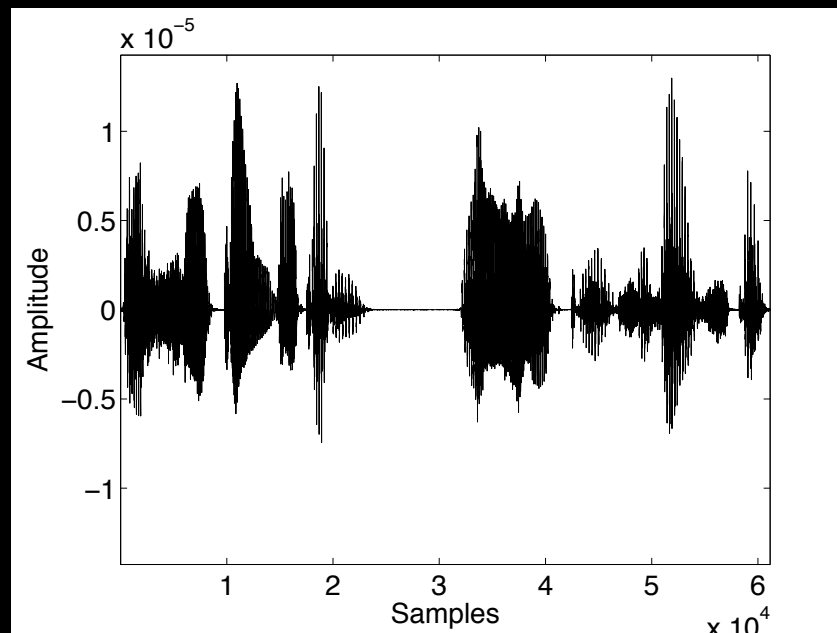


$e_i(n)$

LIME



$e_i(\hat{n})$



$(\hat{n})$

$$1/E\{b(z)\}$$

Pre-Whiteningフィルタでデコンボリューションすれば
回復信号が得られる

計算条件

計算条件

音源信号：24kHzサンプリング男性音声 (2 s)

インパルス応答：包囲型マイクロホンアレイ部屋での実測値

(48 kHzサンプリングを24 kHzにダウンサンプリング)

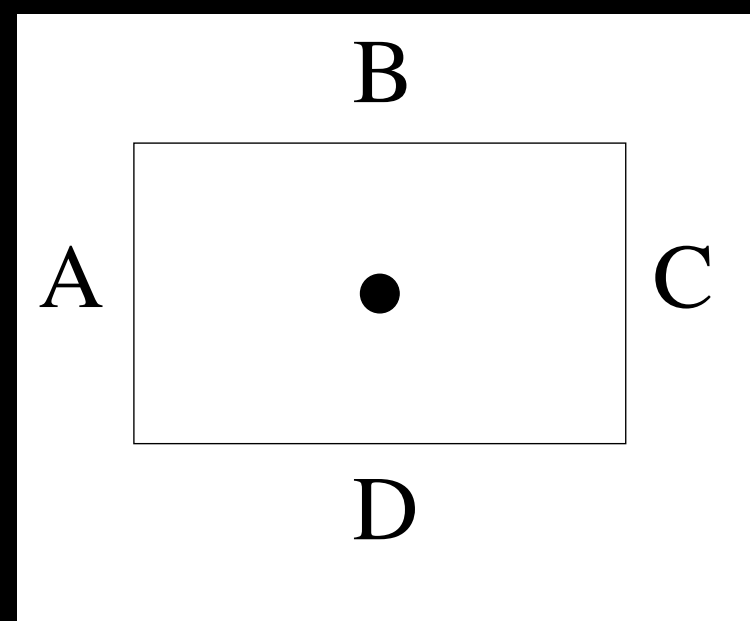
スピーカ：正12面体点音源スピーカ Solid Acoustics

残響時間：0.15 s = 3600 taps

部屋の大きさ：5.18 m x 3.38 m x 2.52 m

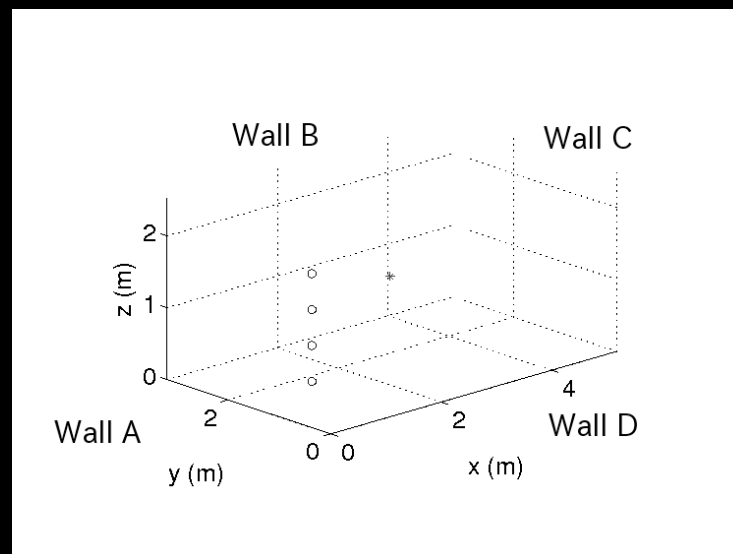


点音源スピーカ

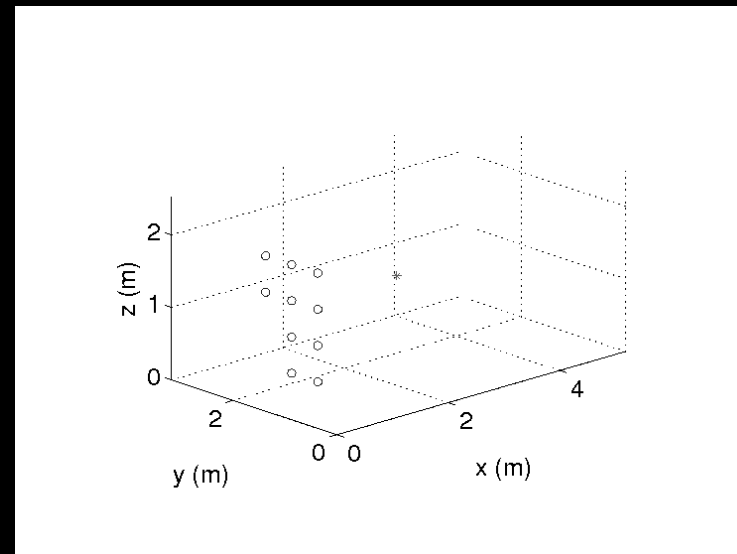


音源位置 (部屋を中心)

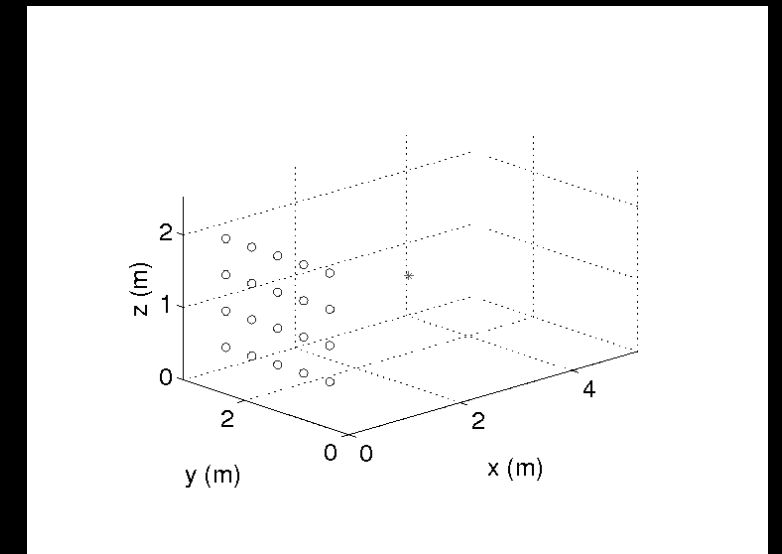
チャネル数の検討



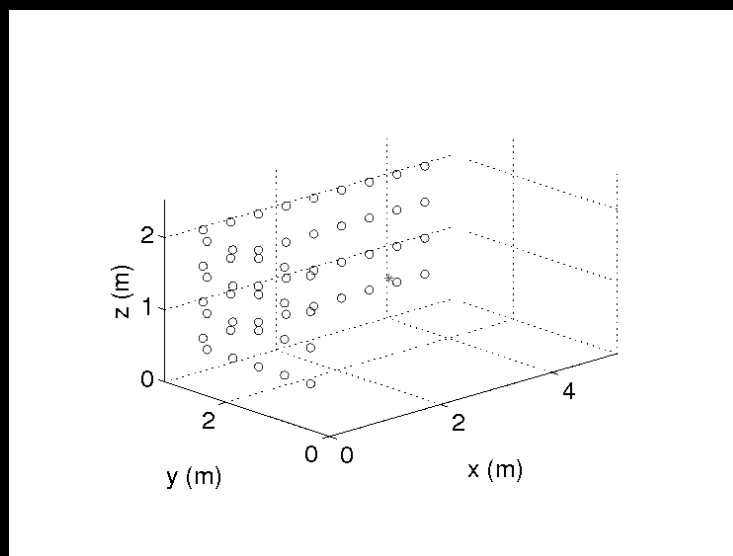
4 ch (1面)



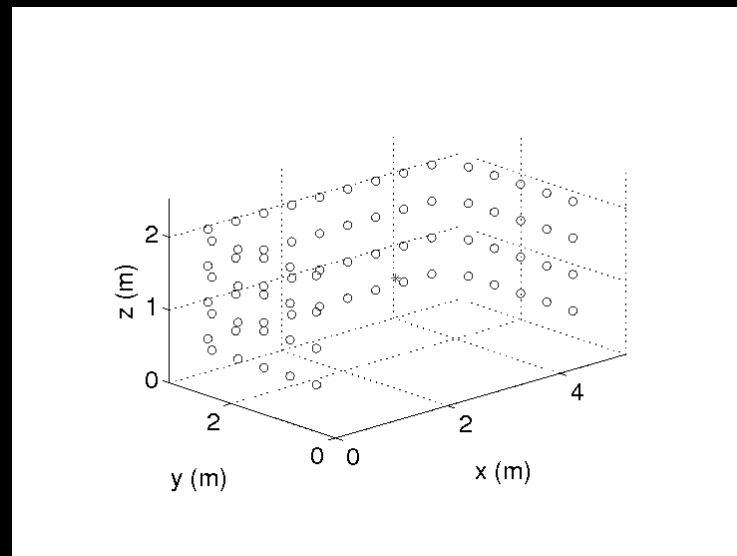
10 ch (1面)



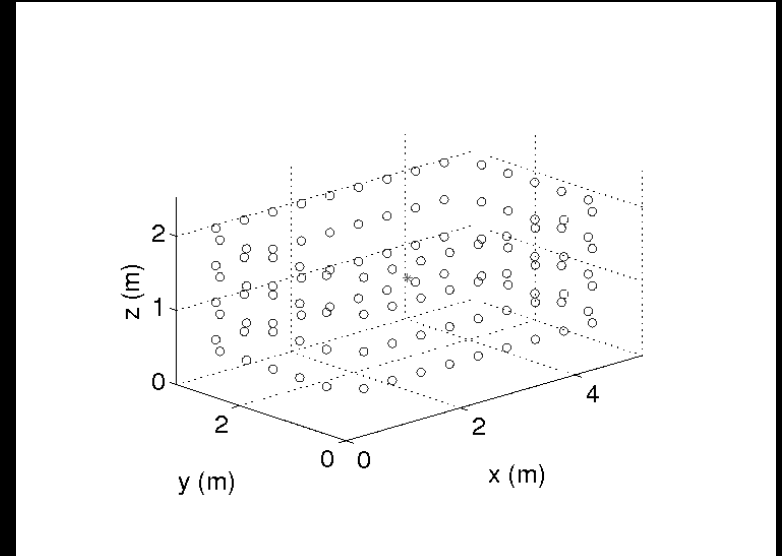
20 ch (1面)



56 ch (2面)



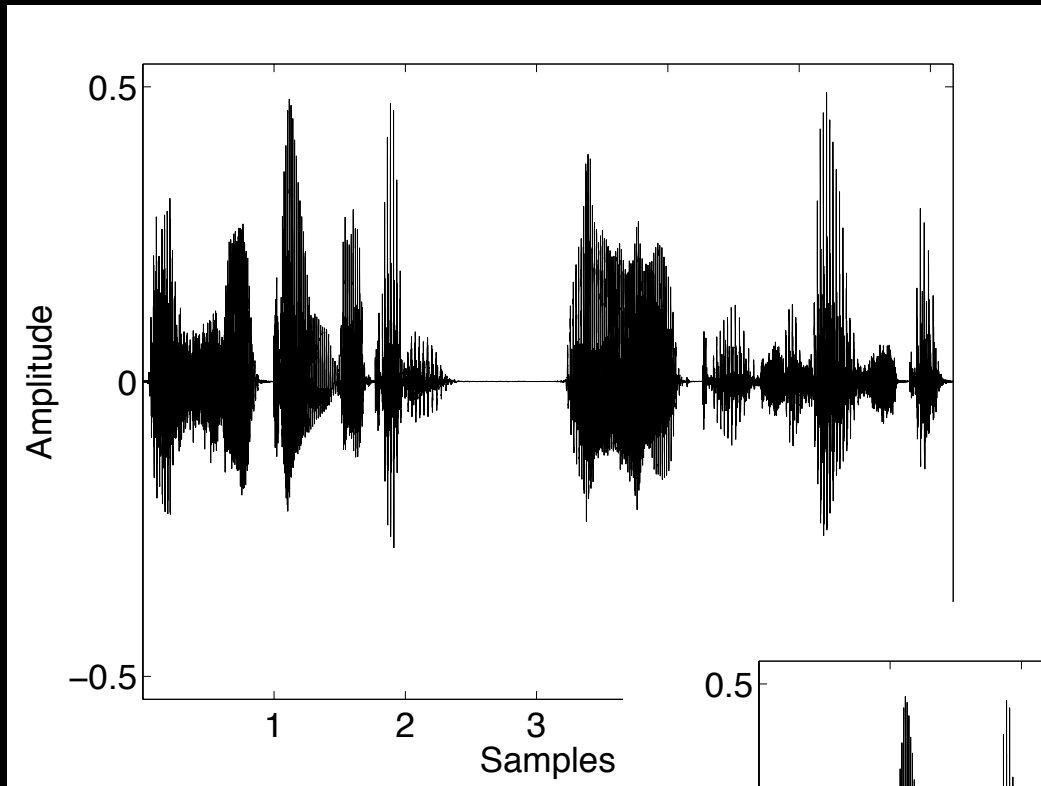
76 ch (3面)



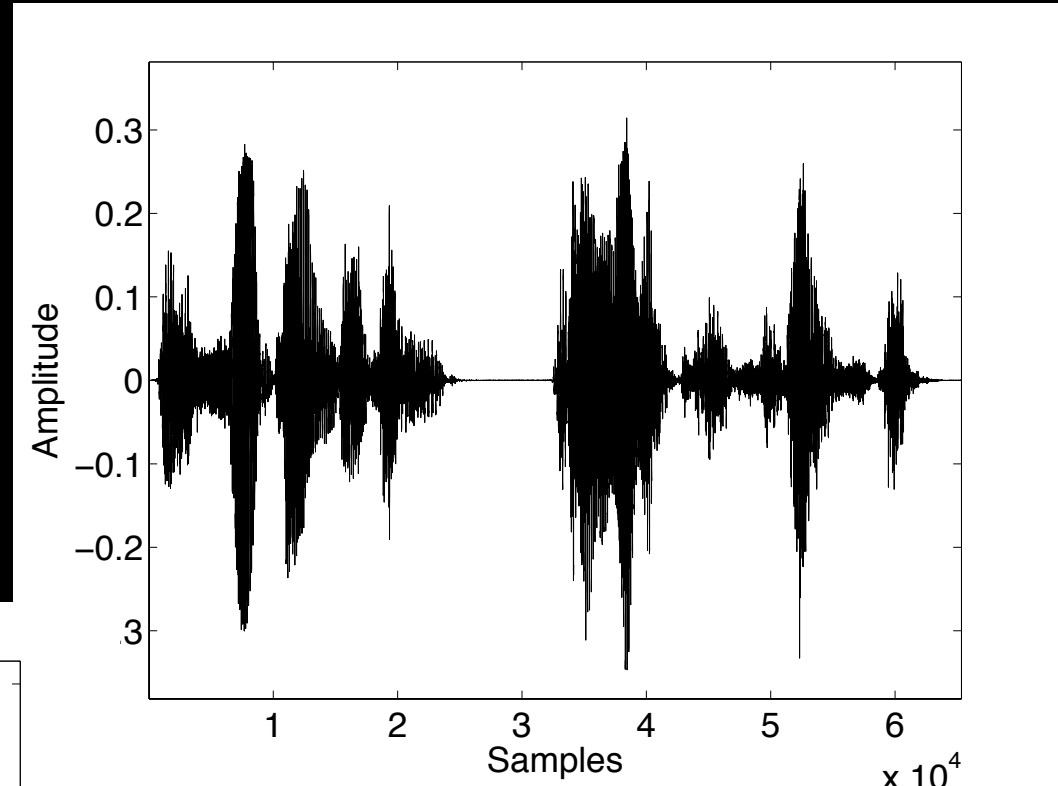
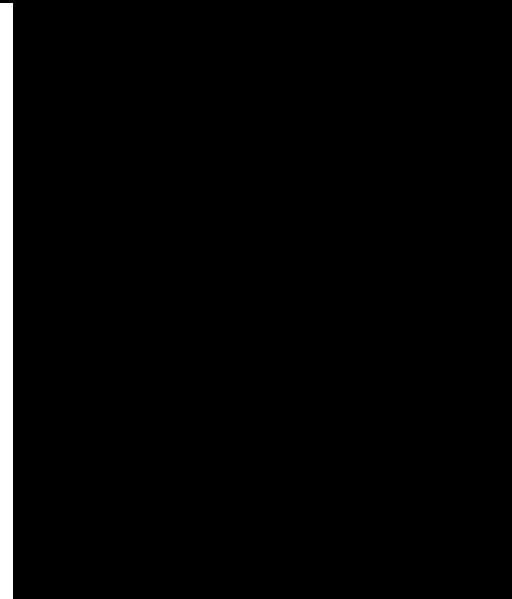
112 ch (4面)

157 ch (5面)

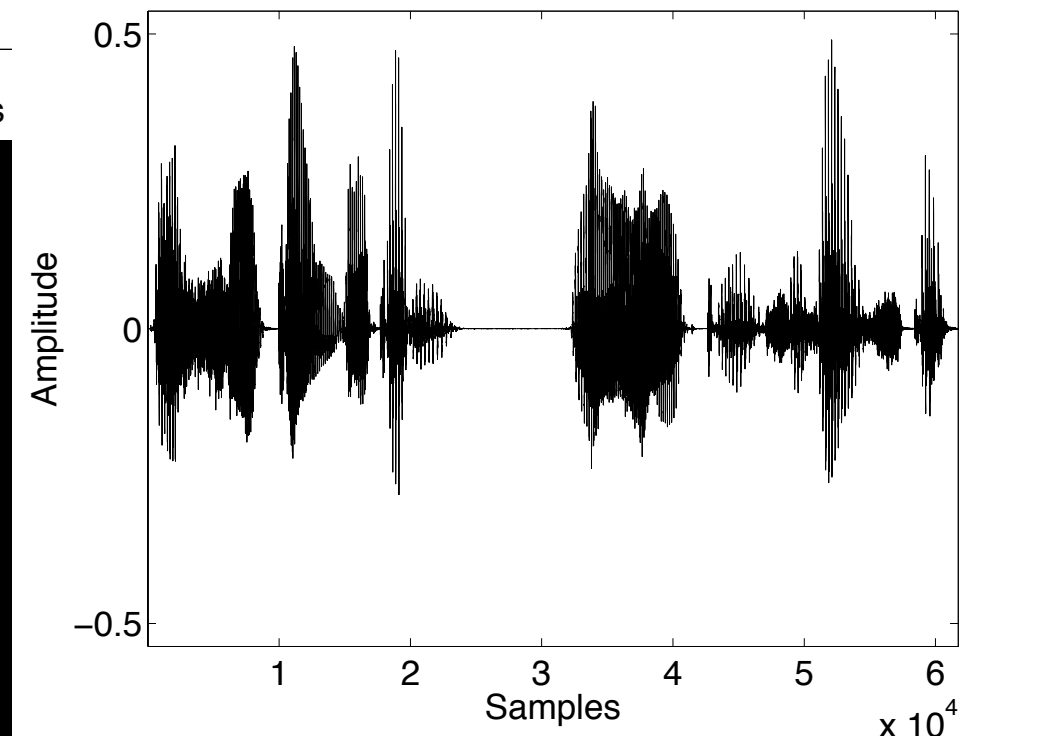
チャンネル数の検討結果 (1)



原音声



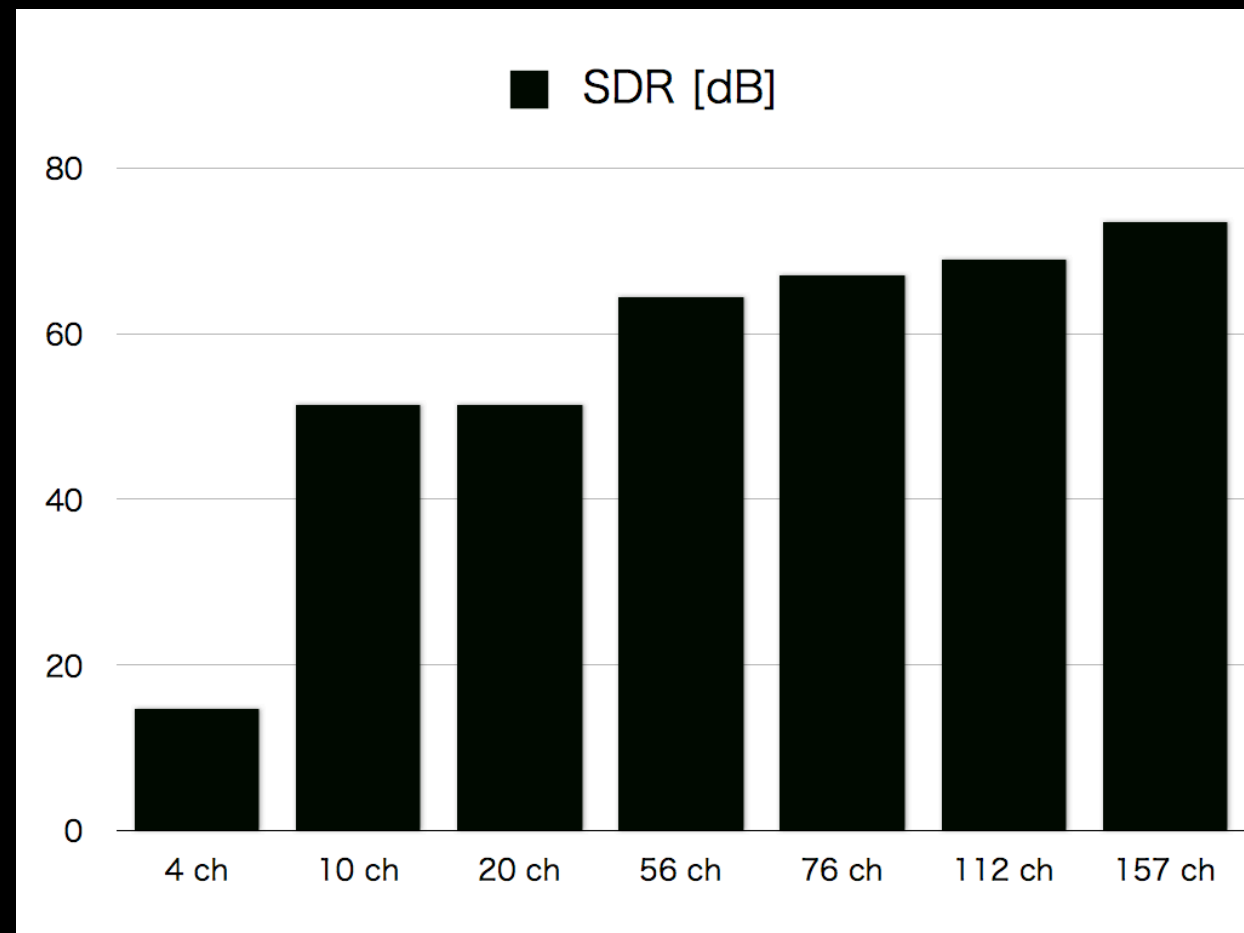
マイクロホン1での信号



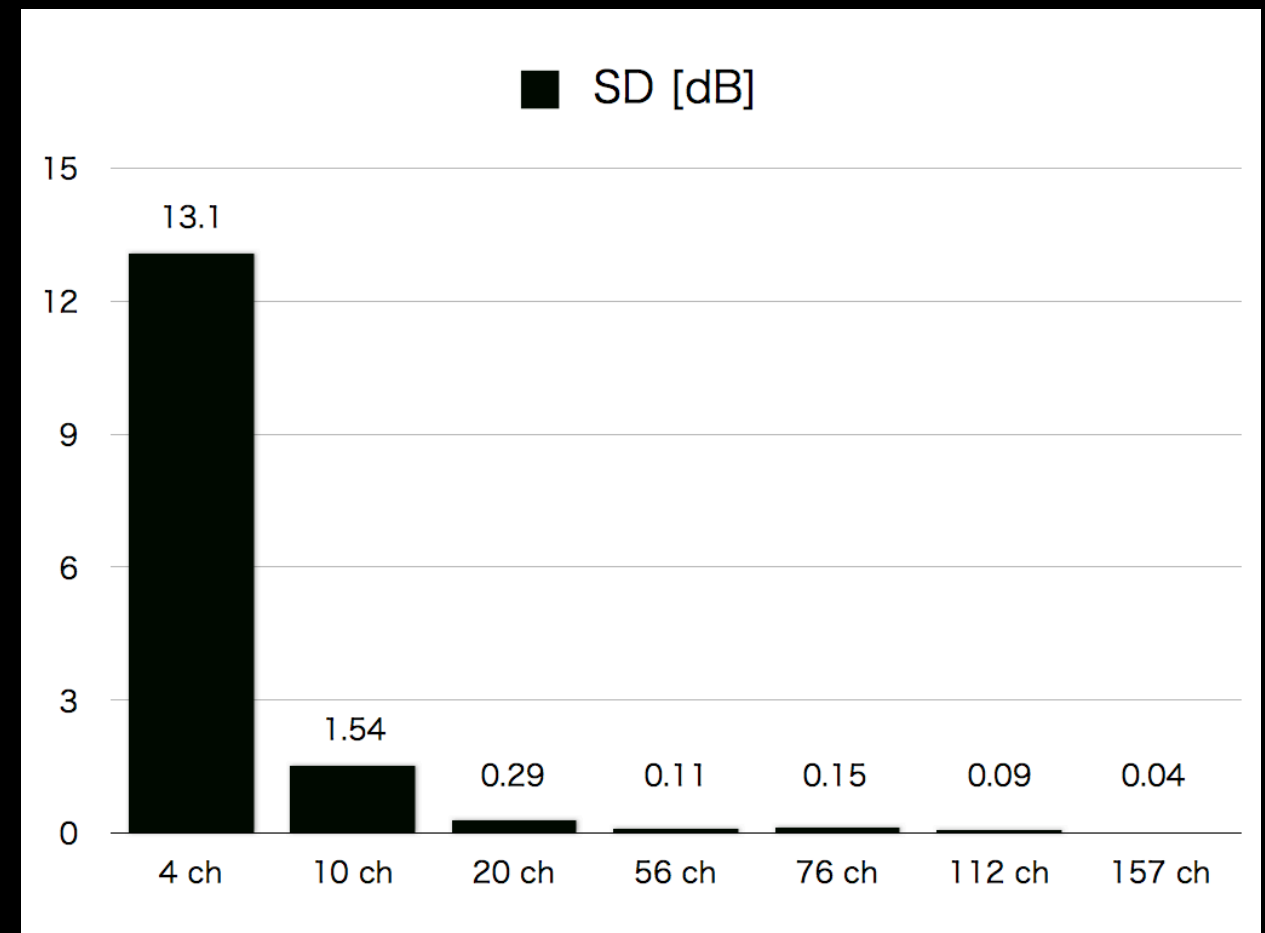
回復信号

チャンネル数の検討結果 (2)

Signal to Distortion Ratio (SDR)



Spectral Distortion (SD)



$$\text{SDR} = 10\log_{10} \left(\frac{\sum |s(n)|^2}{\sum |s(n) - \hat{s}(n)|^2} \right)$$

$$\text{SD} = \sqrt{\frac{1}{F} \sum_{f=0}^{F-1} \left(20\log|P(f)| - 20\log|\hat{P}(f)| \right)^2}$$

チャンネルを増やすと若干ではあるが性能が上がる 17

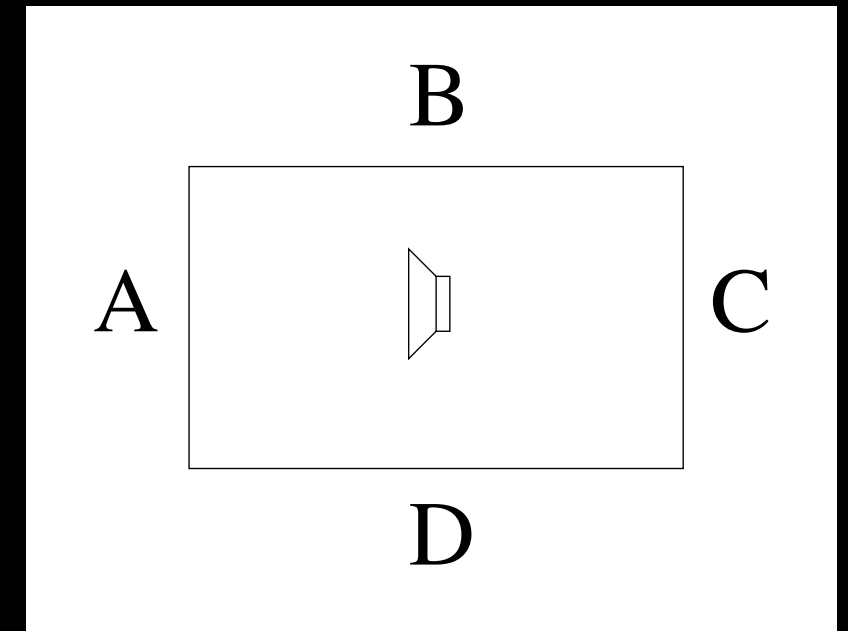
放射特性を持つ音源に対する残響除去 および放射特性推定に関する検討

実環境の音

→ 方向ごとに放射特性を持つ

A…直接音+反射音

C…回折音+反射音



ここでの検討

チャンネル数を一定にした上で…

- ・ Cのような位置の信号から回折音から残響除去はできるのか
- ・ 2面以上の壁を使った場合

さらには…

- ・ Cでの推定結果から残響だけを取り除き

放射特性推定の可能性!!

放射特性を推定結果内に残すことは可能か

計算条件

計算条件

音源信号：24kHzサンプリング男性音声 (2 s)

インパルス応答：包囲型マイクロホンアレイ部屋での実測値

(48 kHzサンプリングを24 kHzにダウンサンプリング)

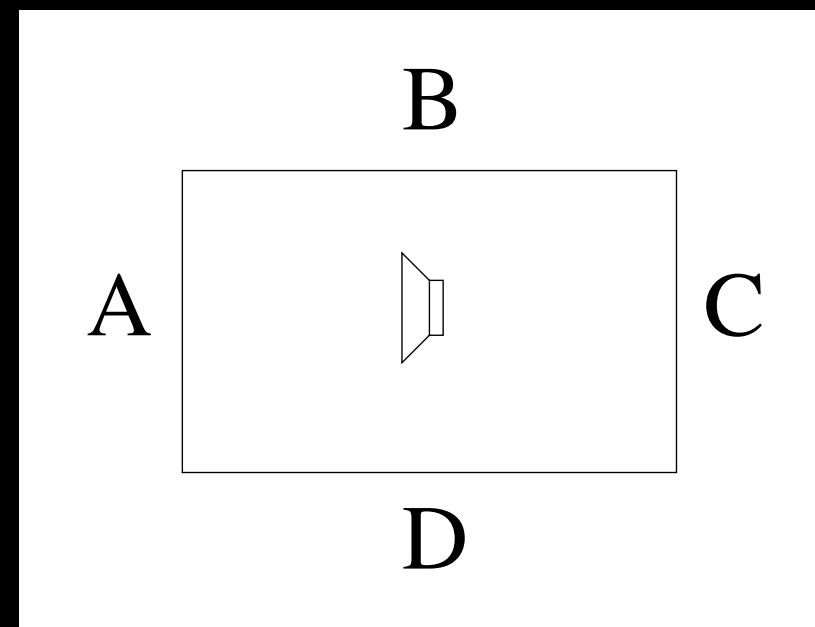
スピーカ：1ウェイスピーカ (マイクロピュア AP5001)

残響時間：0.15 s = 3600 taps

部屋の大きさ：5.18 m x 3.38 m x 2.52 m

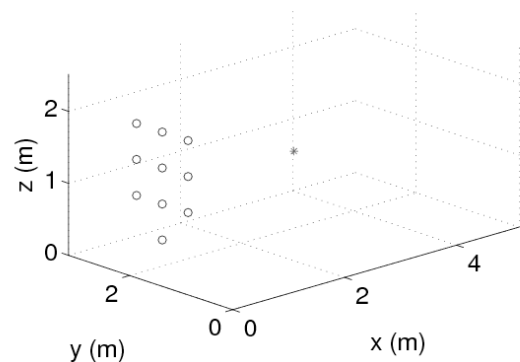


点音源スピーカ

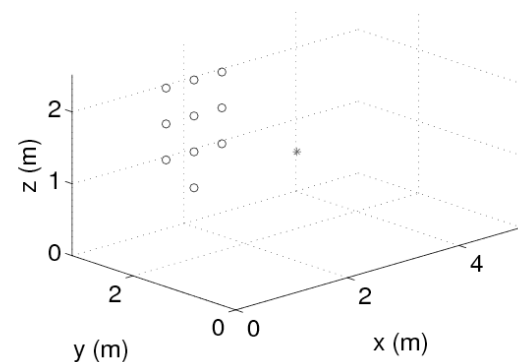


音源位置 (部屋を中心)

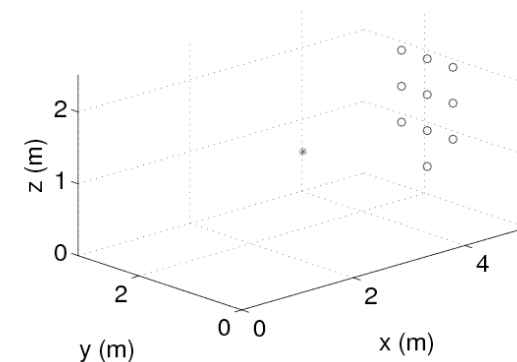
マイクホン配置の検討 (10 ch)



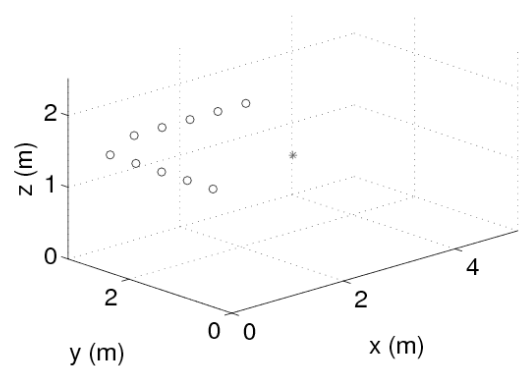
パターン1 (前面)



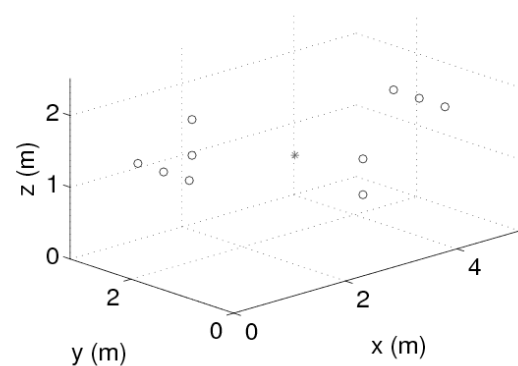
パターン2 (横面)



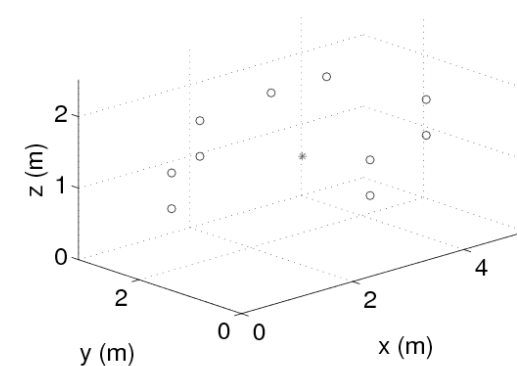
パターン3 (背面)



パターン4 (2面)

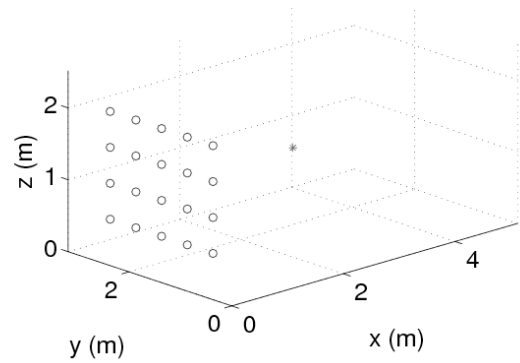


パターン5 (4面)

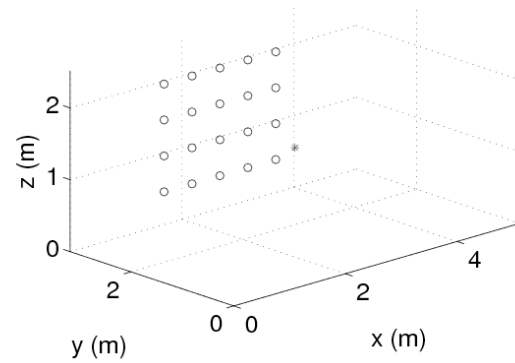


パターン6 (5面) 20

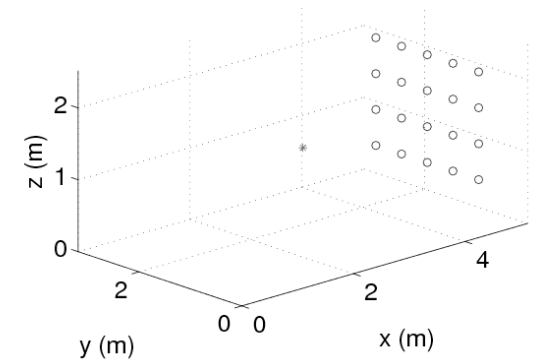
マイクホン配置の検討 (20 ch)



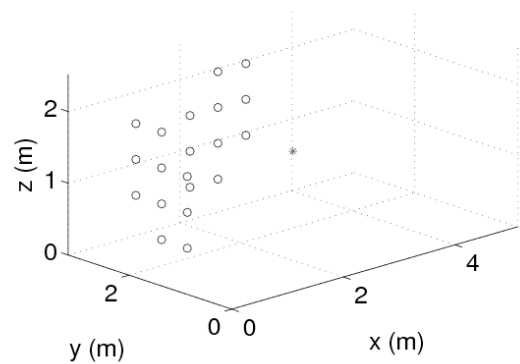
パターン1 (前面)



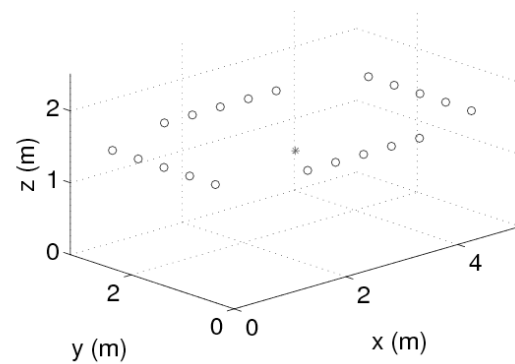
パターン2 (横面)



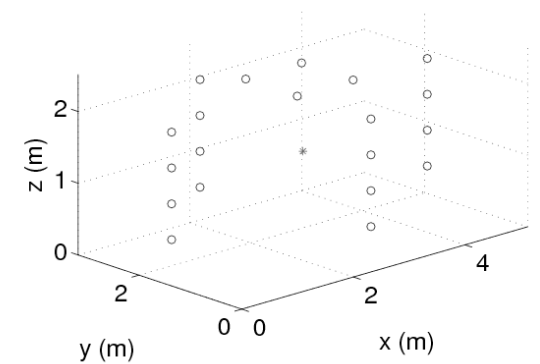
パターン3 (背面)



パターン4 (2面)

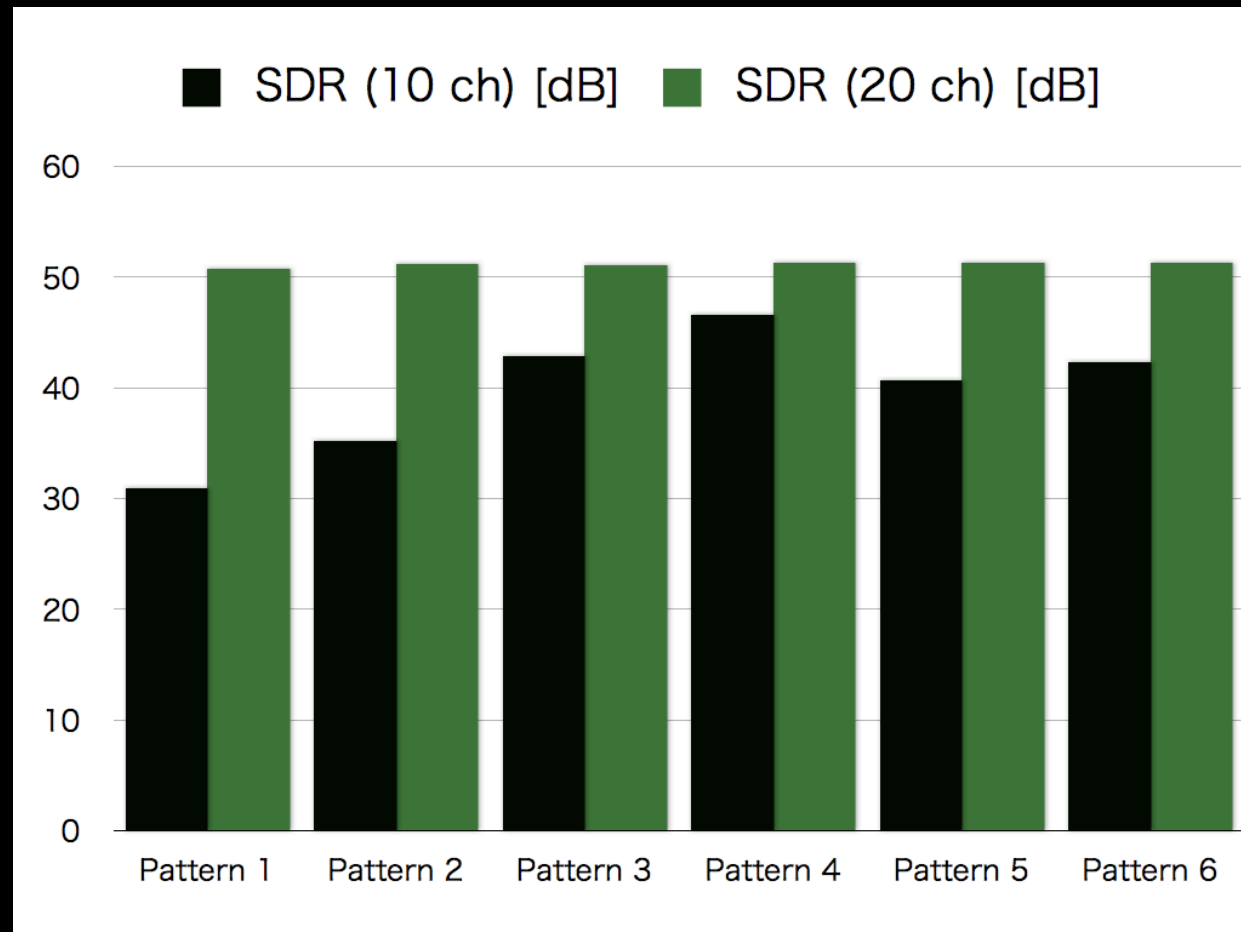


パターン5 (4面)

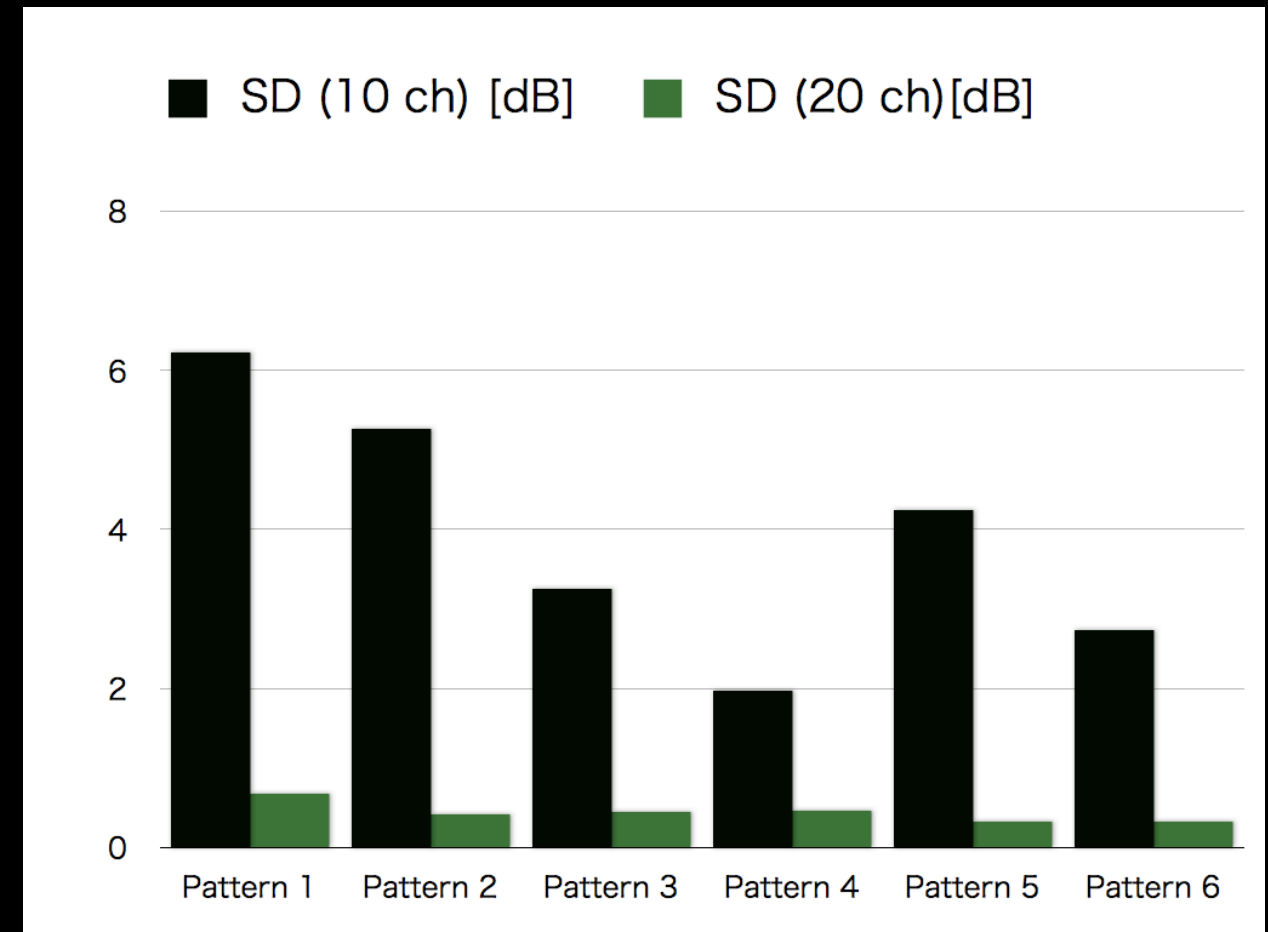


パターン6 (5面) 21

マイクロホン配置の検討結果



Signal to Distortion Ratio (SDR)



Spectral Distortion (SD)

マイクロホン配置の検討結果 (2)

- ・ 10チャンネルだと場所によって多少の差は出るが、
基本的にはどのパターンも良好に残響除去している
- ・ パターン2の結果からもわかる通り、
回折音と反射音のチャンネルからでも残響除去が可能
- ・ 放射特性成分を残響除去した信号に
残すことはできなかった
- ・ 放射特性成分も残響成分同様各チャンネルによって
異なるためであると考えられる

まとめ

- Pre-Whitening フィルタを用いてLIMEアルゴリズムの
高サンプリング化を行った
- チャンネル数を増やすことにより,
性能が飽和することなく多チャンネルの効果を確認した
- 放射特性のある音源の場合においても,
精度よく残響除去ができることを確認した

今後の課題

- ・ 背景雑音にロバストな残響除去アルゴリズムの検討
- ・ 包囲型マイクロホンアレイを用いた放射特性の推定