

高詳細音響属性の抽出による 音空間編集システム構築に向けて

Toward an editable audio-space system
using high-resolution sound properties

岡本 拓磨^{*†}, 岩谷 幸雄^{*‡}, 鈴木陽一^{*‡}

* 東北大学 電気通信研究所

† 東北大学 大学院工学研究科

‡ 東北大学 大学院情報科学研究科学

研究背景：3次元音空間の提示と集音

■ モノラル・ステレオ再生を超える3次元音空間提示技術

■ 3次元音空間提示技術への関心・技術革新

- ✳ (5.1 ch)サラウンドシステム・ホームシアター音響
- ✳ ヘッドホンやスピーカを用いた聴覚ディスプレイによる3次元音空間提示



高臨場感な音空間を創出可能に

■ 近未来のセンサネットワーク

■ 多数のマイク空間を丸ごと所に配置が可能

- ✳ 音空間丸ごとの集音が可能



単に集音するだけではなく…

- ・音源位置情報, ・原音情報, ・放射指向特性, ・反射音成分

といった音空間の属性値の高精度解析が可能に

音空間を構成する5つの属性

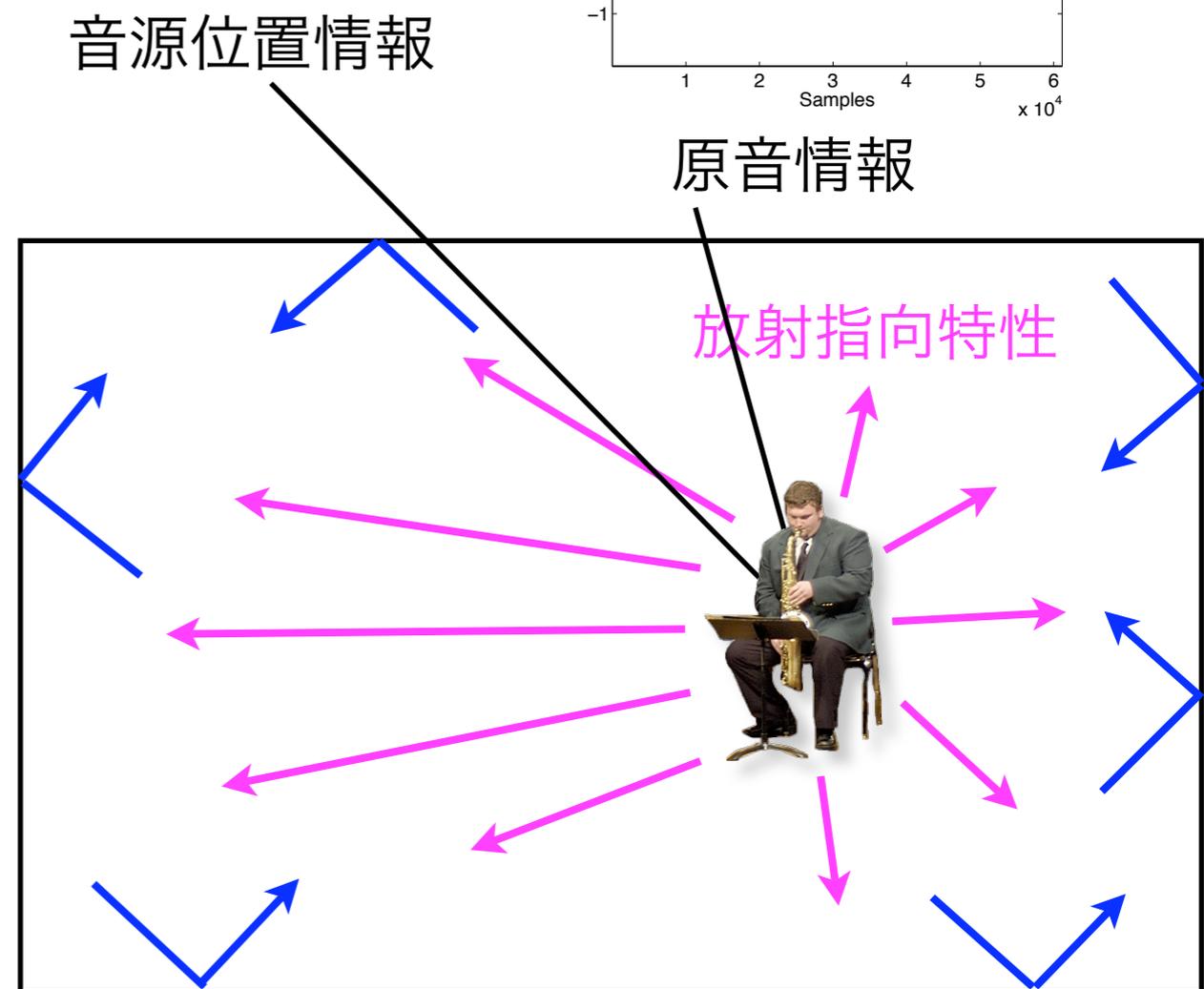
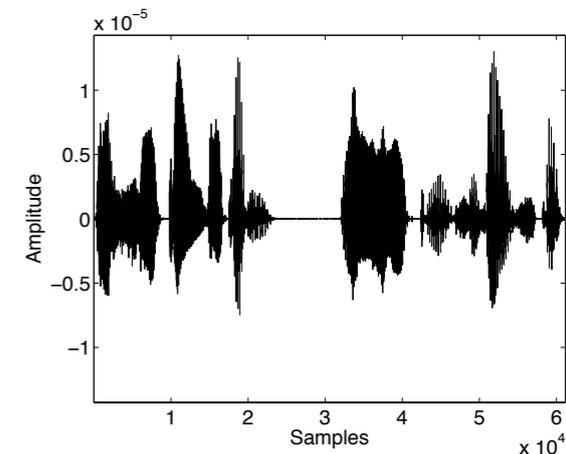
■ 室内音空間を構成する5つの属性

■ 音源の属性

1. 音源位置情報
2. 原音情報
3. 放射指向特性
…包囲による伝達特性の違い
(例：話者や楽器の向き)

■ 空間 (=室内)の属性

4. 反射音情報
5. 後部残響特性
…反射や散乱等を繰り返し
方向性を失った波面の特性



➡ 各音源ずつ存在, 極めて複雑な室内音場

エディタブルオーディオシステム

■ 既存の収音再生型音空間提示システム (音場再現)

■ 波面合成法 Fletcher *et al.* 1934

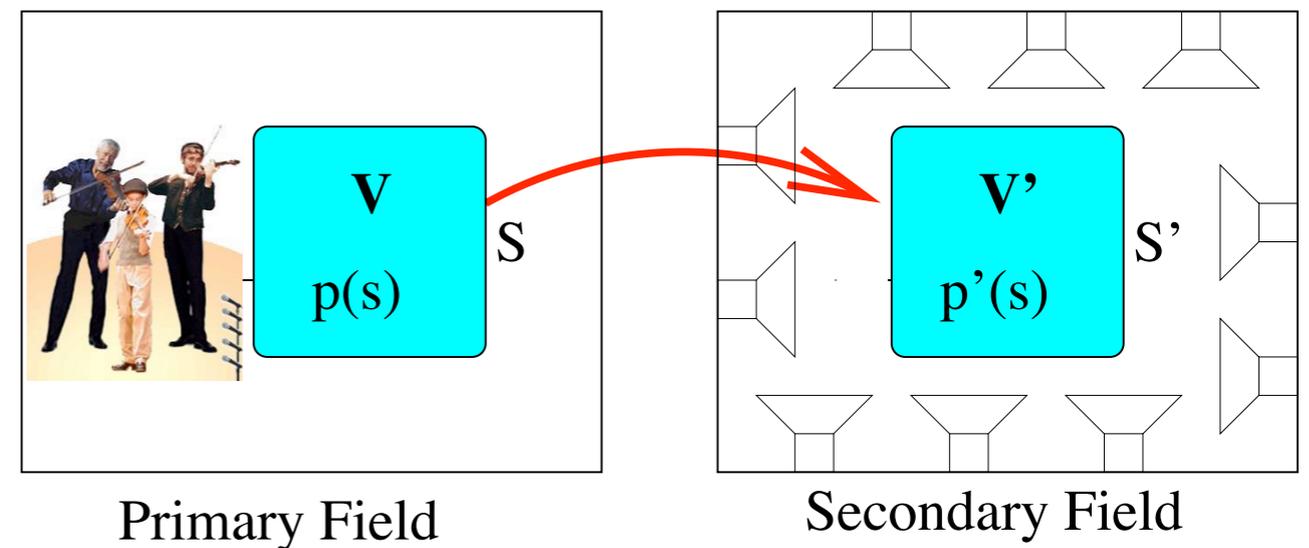
■ 境界音場制御 S. Ise 1997

✳ 閉領域内の境界の音情報を制御



➡ 領域内の音場をまるごと再現可能

➡ 音属性を分解することは困難



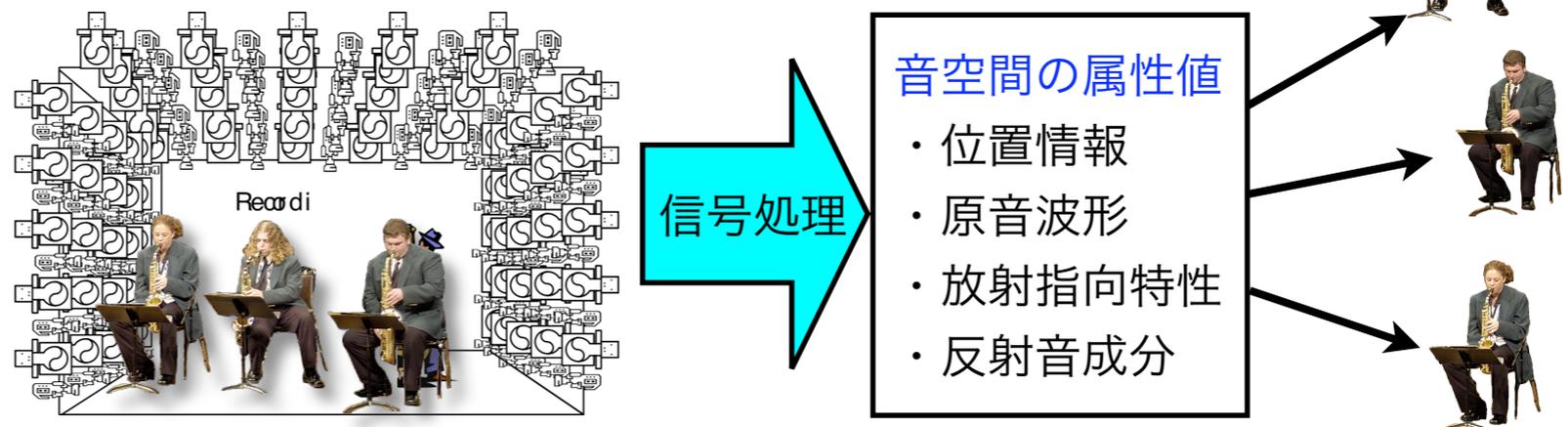
閉領域の音場再現

■ エディタブルオーディオシステムの提案

■ 音空間の属性値を収音情報から高精細に抽出



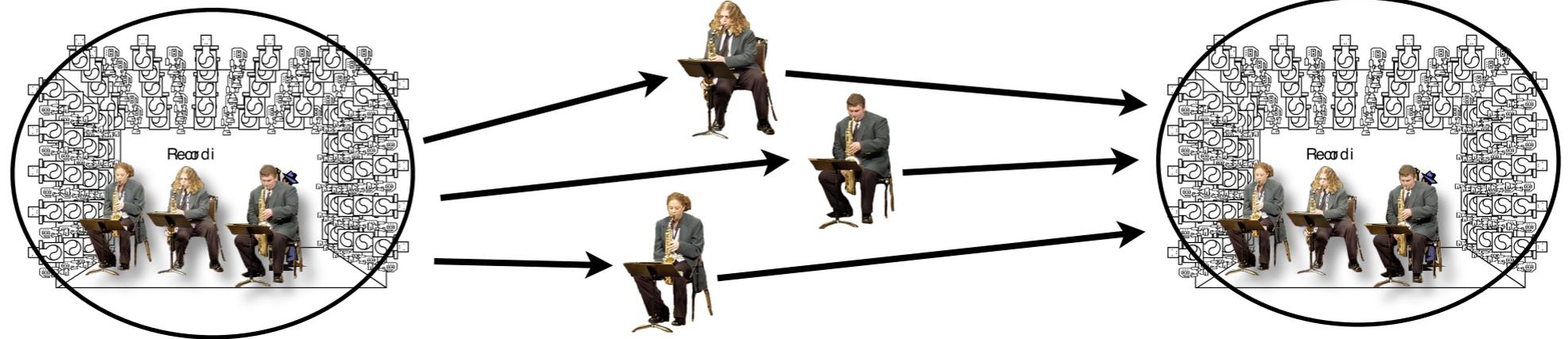
■ 属性値ごとに自由に提示



エディタブルオーディオシステム (2)

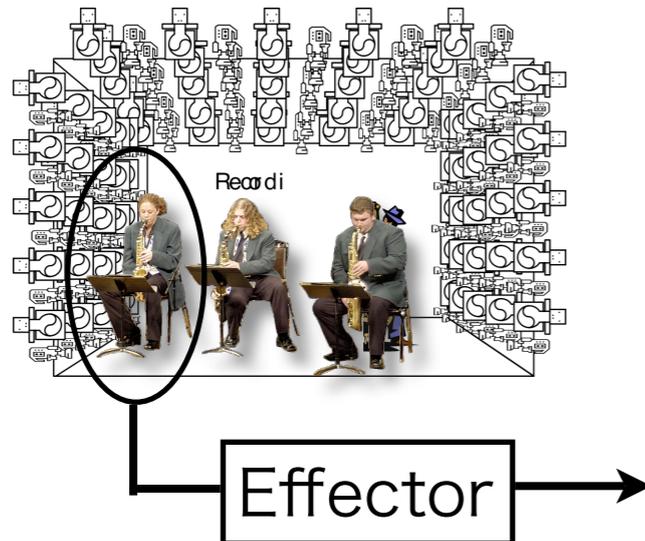
■ エディタブルオーディオで実現できる音空間

■ 忠実な再現

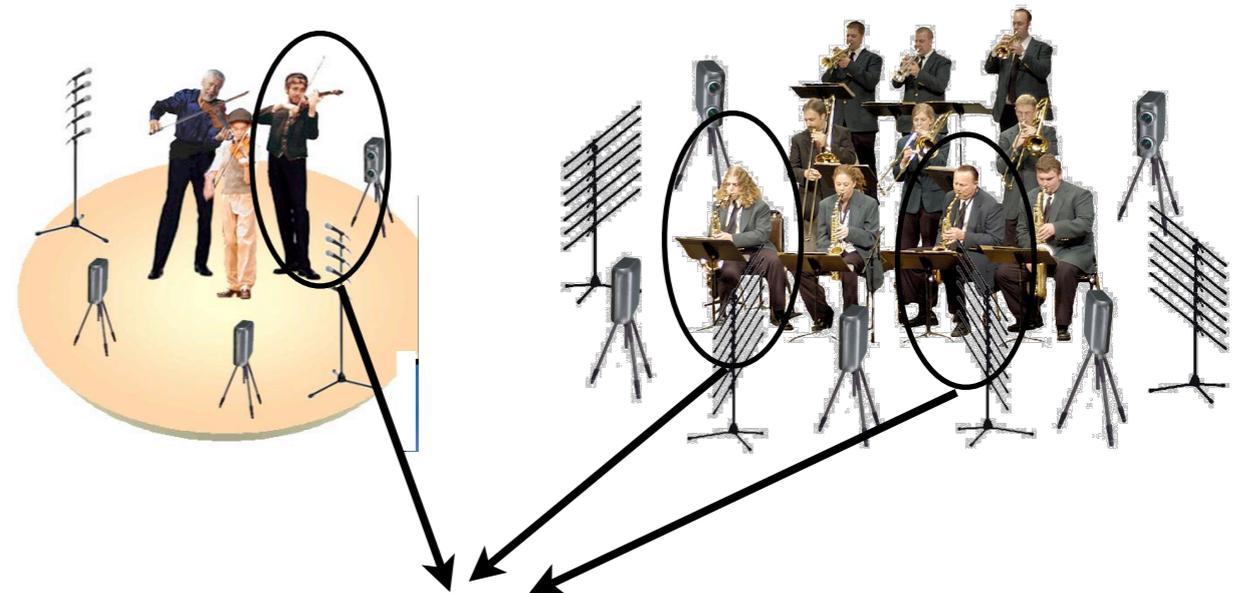


のみならず...

■ 特定の属性のみの編集・提示



■ 他イベントとの融合



多様性に富んだ臨場感創出音空間提示システムの実現

エディタブルオーディオシステムの実現に向けて

■ 室内音源特性情報のモデル化

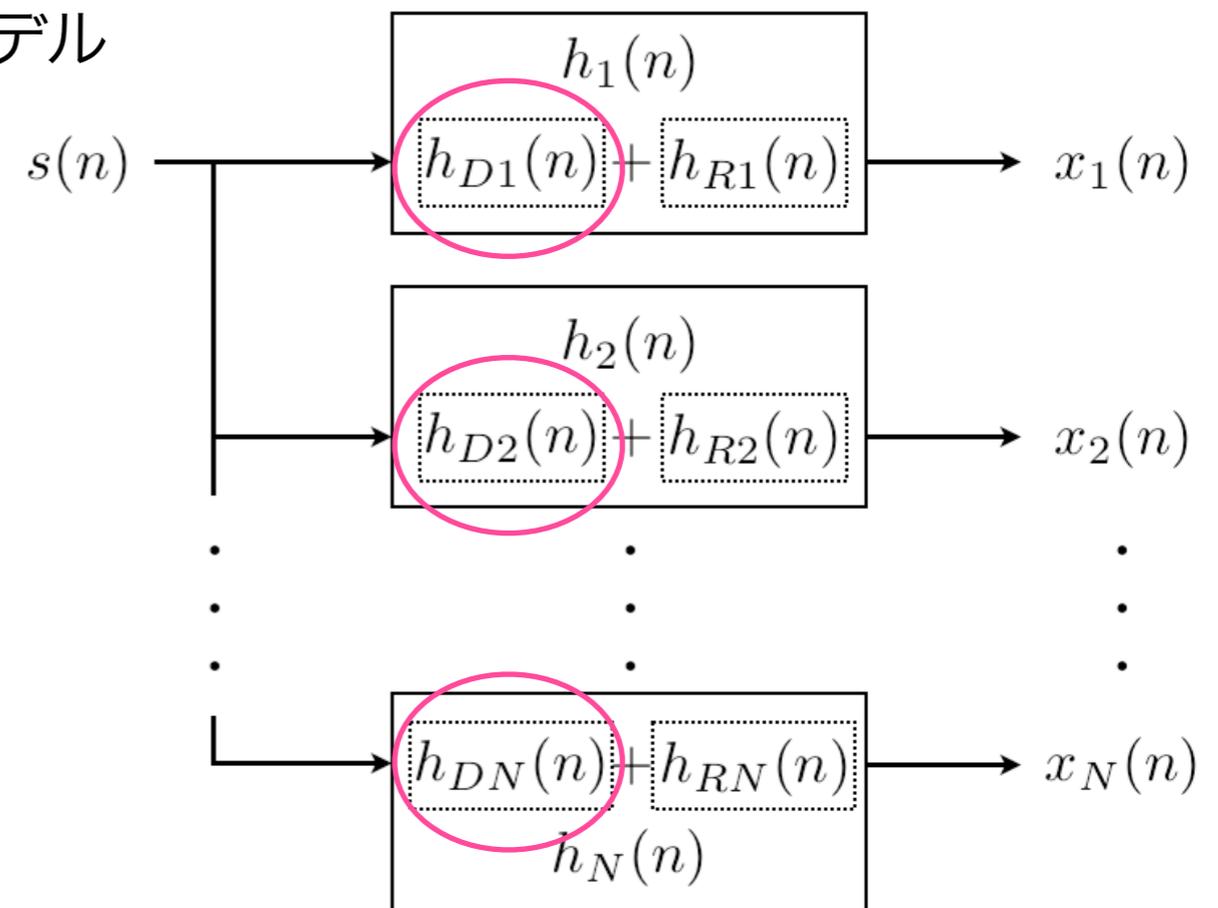
- 室内に音源が1つであると仮定
- Single-Input Multiple-Output(SIMO)モデル

- ✳ 音源波形信号： $s(n)$
- ✳ 各観測点での観測信号： $x_i(n)$

- 音源の方向による放射指向特性を伝達関数として定義 (オリジナル)

- ✳ 放射指向特性： $H_{Di}(z)$
 - ✳ 室内伝達特性： $H_{Ri}(z)$
- } $H_i(z)$

- まずは観測信号 $x_i(n)$ のみから位置情報 (x, y, z) , 原音 $s(n)$, 放射指向特性 $H_{Di}(z)$ を抽出



SIMOモデル(放射指向特性あり)

空間全体を保存するために構築したシステム

■ 包囲型マイクロホンアレイ

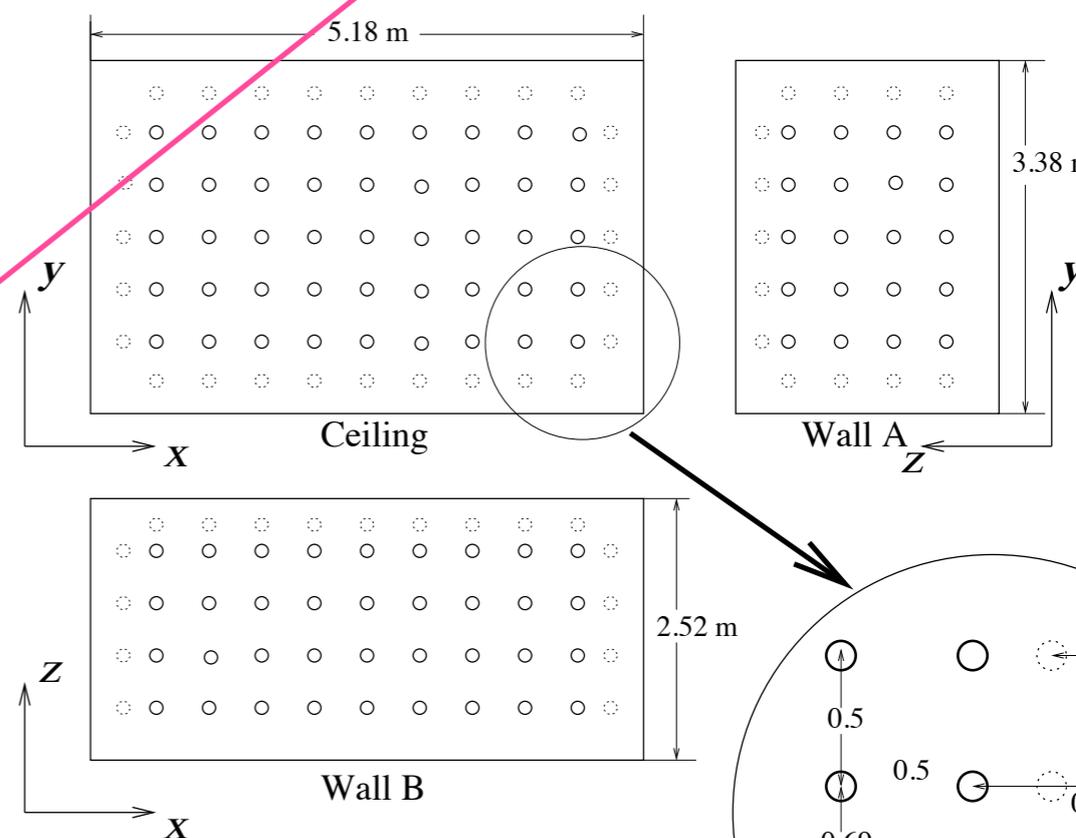
- 天井と壁4面にマイクロホンを格子状に配置
- 壁面から30 cm, 50 cm間隔にて格子状に配置
 - ※ マイクロホン：B & K Type 4951 × 157 ch
 - ※ マイクロホンアンプ：B & K Type 2694 × 10



B & K Type 4951



包囲型マイクロホンアレイ



*Unit: meter

■ 48 kHzサンプリング157 chの同時収録

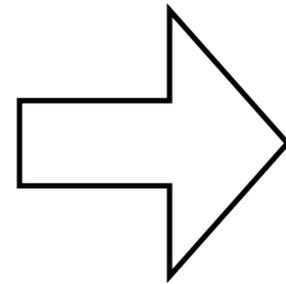
➡音源特性を余すところなく収録可能

包囲型マイクロホンアレイを用いた 音源位置情報と原音情報の抽出

■ 音源位置情報の抽出と課題

■ 室内での音源位置推定

- ✳ 反射音により精度が劣化
- ➡ 音源位置推定では未検討課題



あらかじめ反射音成分を低減する
信号処理を施した後、位置推定を
行う方法RAP-MUSIC (Re-Arrangement
and Pre-smoothing for Multiple Signal Classification)
を提案 T. Okamoto *et al.* 2007

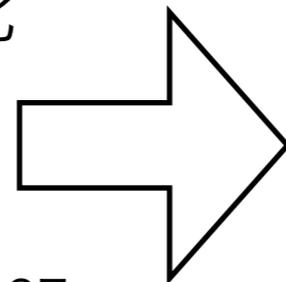


反射音に頑健な位置推定が可能

■ 原音抽出(=残響除去)と課題

■ 直接逆特性を求めるブラインド残響除去

- ✳ 広帯域では残響除去精度の劣化
- ➡ 広帯域での残響除去が課題



残響除去処理の前に入力信号を白
色化するWhite-LIMEを提案

岡本ら 2009



広帯域での残響除去が可能

従来方法：LIME M. Delcroix *et al.* 2007

Linear-predictive Multi-input Equalization



これらを用いて音源の放射指向特性の抽出を試みる

研究の目的：音源の放射指向特性の抽出

■ 音源の放射指向特性

- 実環境の音源は振動パターンや音源自体に大きさ
- 点音源の球面波の伝達特性とは異なる
- 音源からの距離が等しくても方位によって伝達特性が異なる

音源を忠実に再現するためには放射指向特性の抽出は必須

■ 大半の研究：人の顔やスピーカの正面方向の音を収録

- 伝達関数を平面波や球面波と仮定 → 放射指向特性は考慮なし

■ 放射指向特性推定に関する先行研究

K. Nakadai *et al.* 2005

- 音源位置と方向の同時推定
 - 正面は音圧レベルが強い、後方は弱いといった大雑把な仮定の推定
 - 厳密な推定ではない

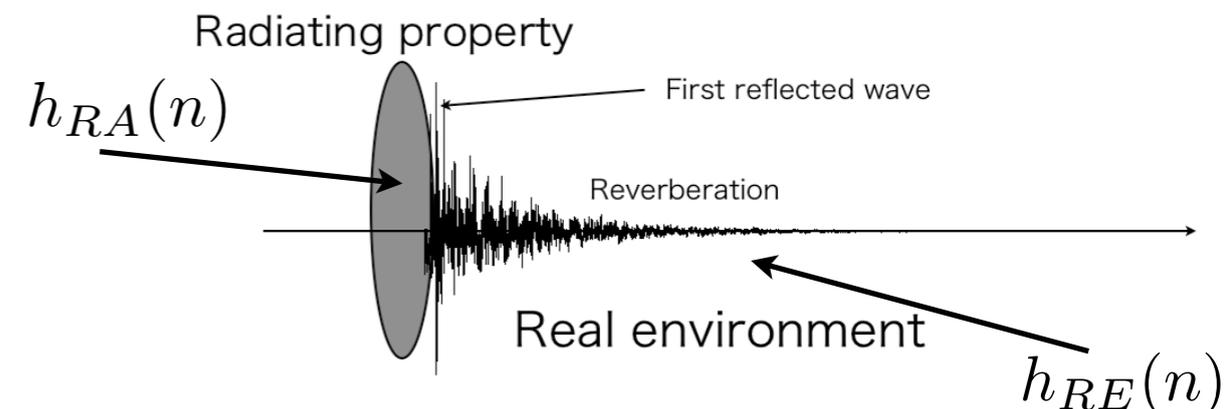
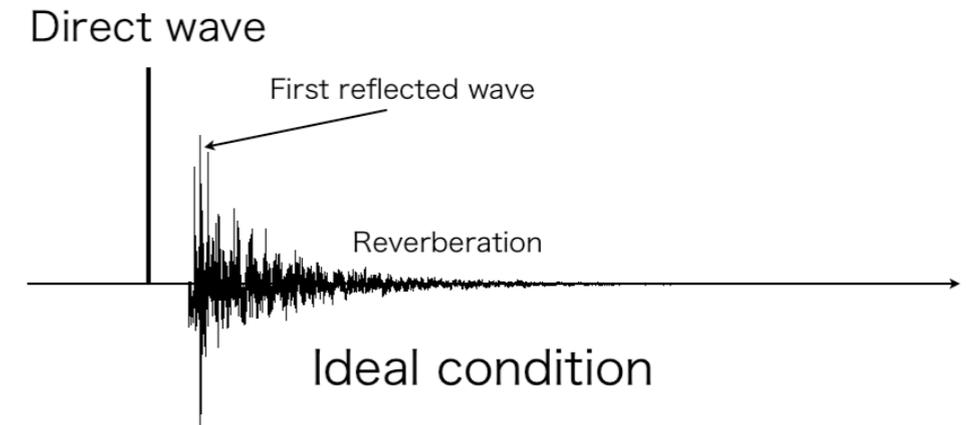
放射指向特性抽出は未開拓の問題

放射指向特性と室内残響の分離

放射指向特性と室内残響の関係

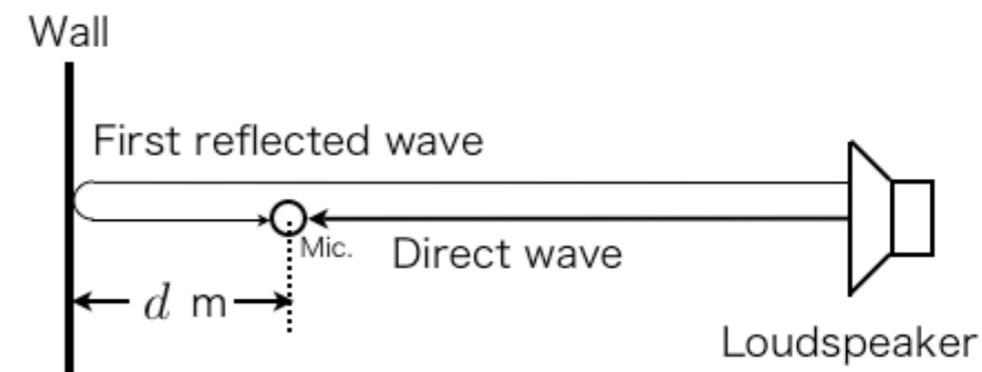
- 点音源 … 直接音はインパルス応答の最初の応答
- 放射指向特性 … ある一定時間の応答
 - ✱ 放射指向特性： $h_{RA}(n)$
 - ✱ 室内残響特性： $h_{RE}(n)$
- 観測点でのインパルス応答： $h(n)$

$$h(n) = h_{RA}(n) + h_{RE}(n)$$



放射指向特性と室内残響特性の分離

- 直接音と初期反射音到来までの時間差を利用
- $t = 2d/c$ までは直接音(=放射指向特性)のみの応答
- 初期応答から $t = 2d/c$ までの応答を切り出し



→ 放射指向特性を抽出

原音情報を用いたインパルス応答の抽出

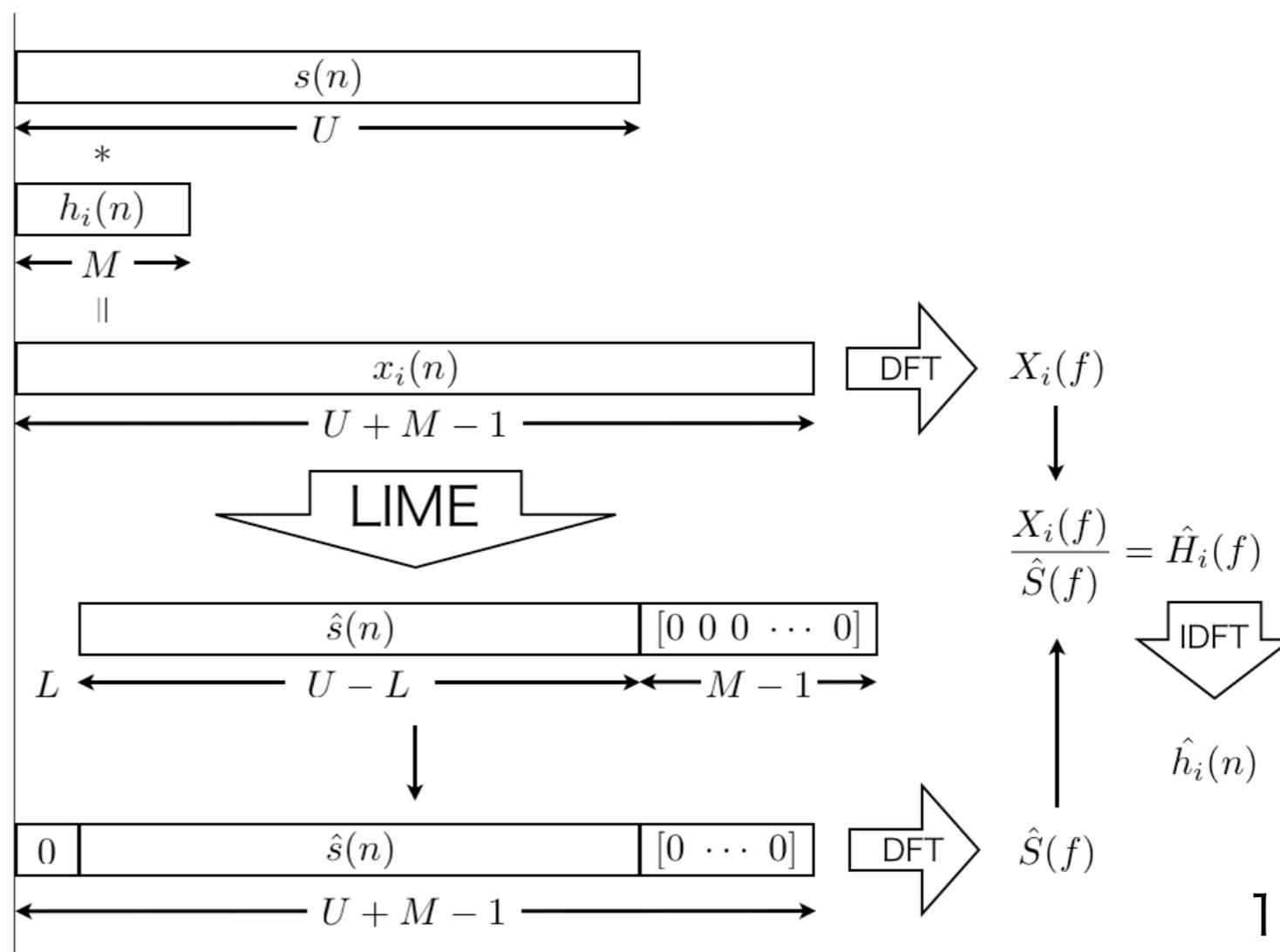
■ インパルス応答(Impulse Response:IR)の抽出

- ブラインドシステム同定は困難
- White-LIMEを考案 → 原音情報を抽出可能
- 入力(原音)と出力(観測信号)から伝達関数(室内インパルス応答)を抽出

- 抽出した原音信号 : $\hat{S}(f)$
- 観測信号 : $X_i(f)$
- 抽出するIR : $H_i(f)$

$$\hat{H}_i(f) = \frac{X_i(f)}{\hat{S}(f)}$$

- 逆畳み込みにより抽出可能



放射指向特性の抽出

■ インパルス応答の切り出し

- 抽出したインパルスの初期反射までの最短時間を放射指向特性として切り出し
- 観測点 \leftrightarrow 壁面：30 cm
- 44.1 kHzサンプリング：
 $44100 \times 2 \times 0.3 / 340 \rightarrow 78$ タップ
- 初期応答から78タップを切り出し

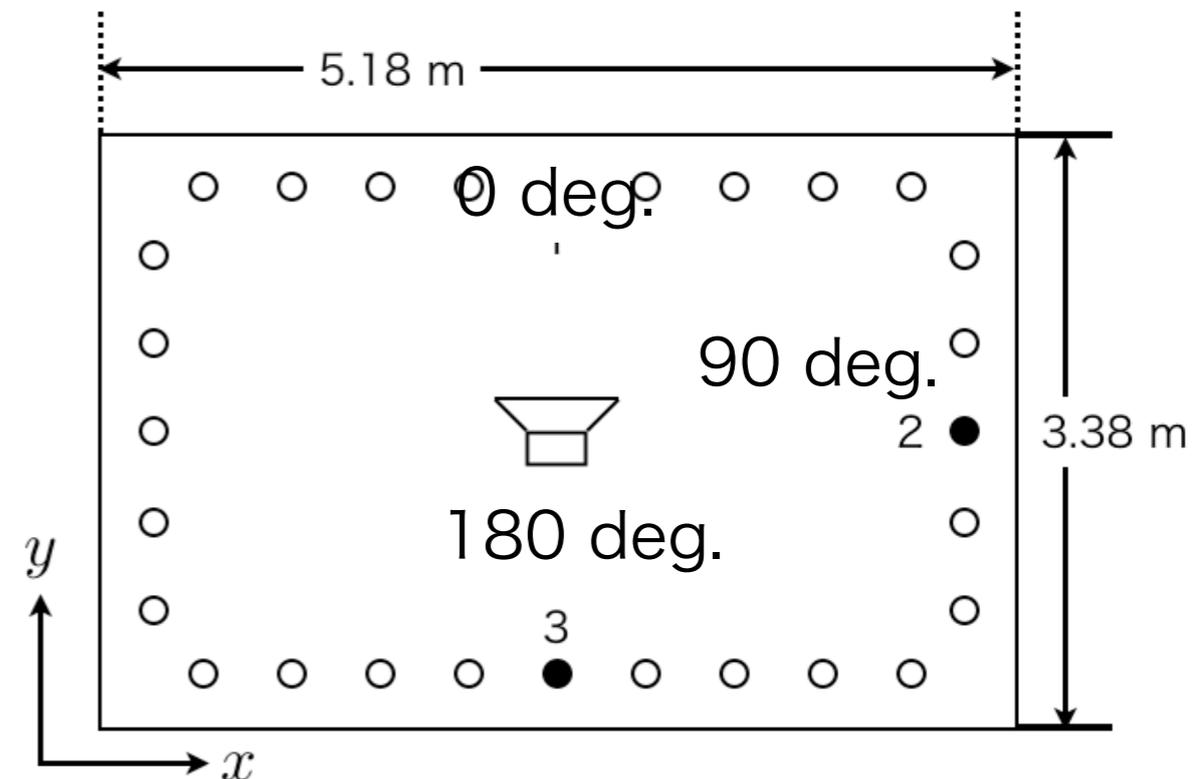


Micropure AP5001

■ 抽出精度の性能評価

- スピーカ正面, 真横, 背面での応答を無響室での応答と比較
- 方向ごとの周波数特性を1/3オクターブバンド解析した結果を比較
- 距離減衰は音源位置情報から補正

$$h_{Di}(n) = r_i h'_{Di}(n)$$

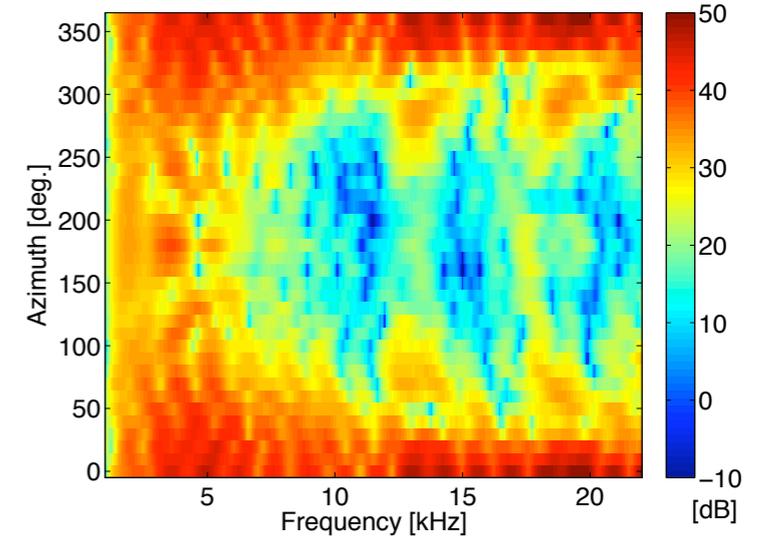
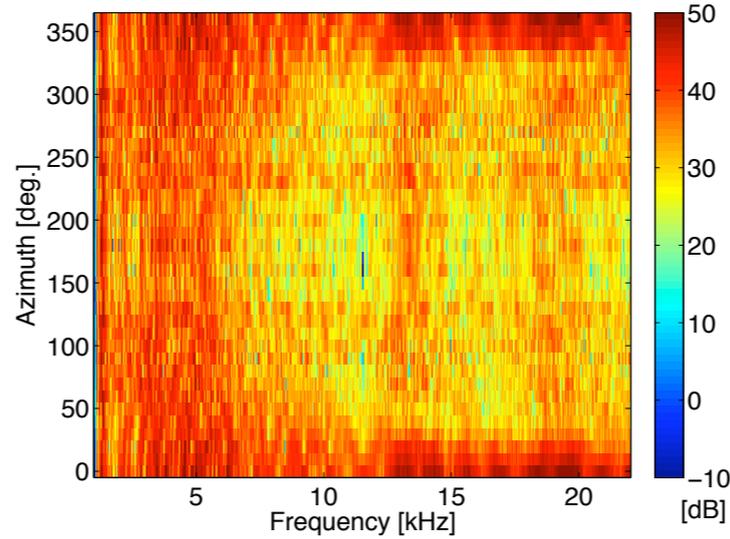
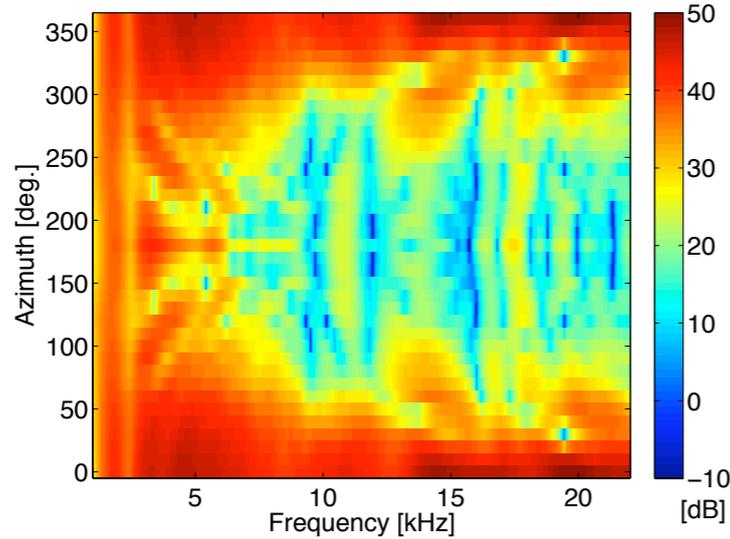


マイクロホン高さ：1.0 m

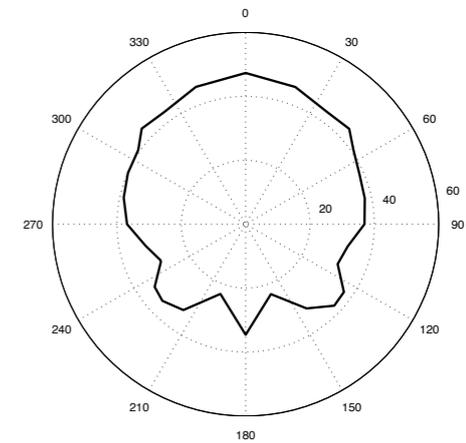
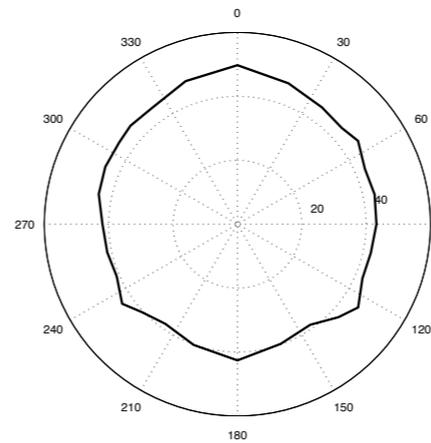
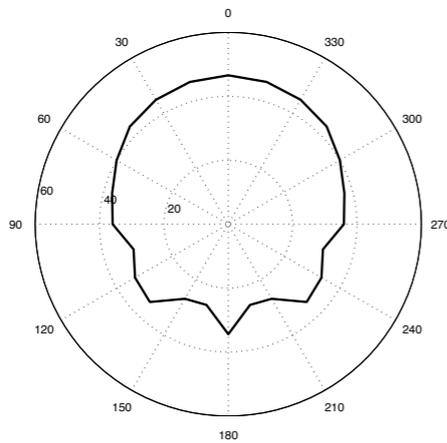
スピーカ高さ：1.1 m

結果比較(全帯域, 2, 8 kHzの音圧レベル)

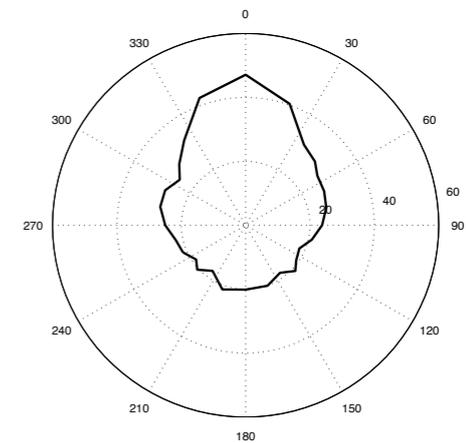
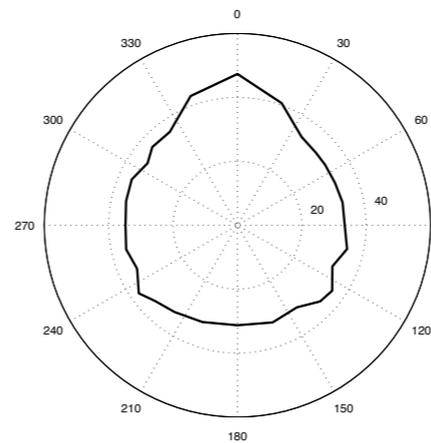
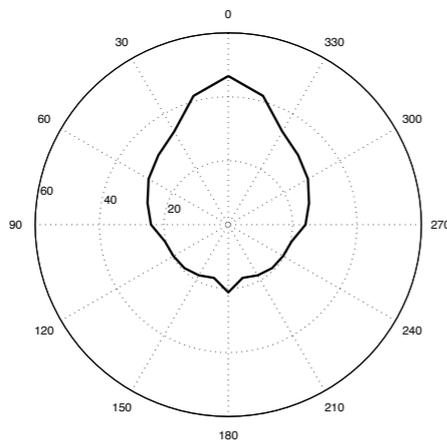
全帯域
全方向



2 kHz



8 kHz



無響室測定の特徴

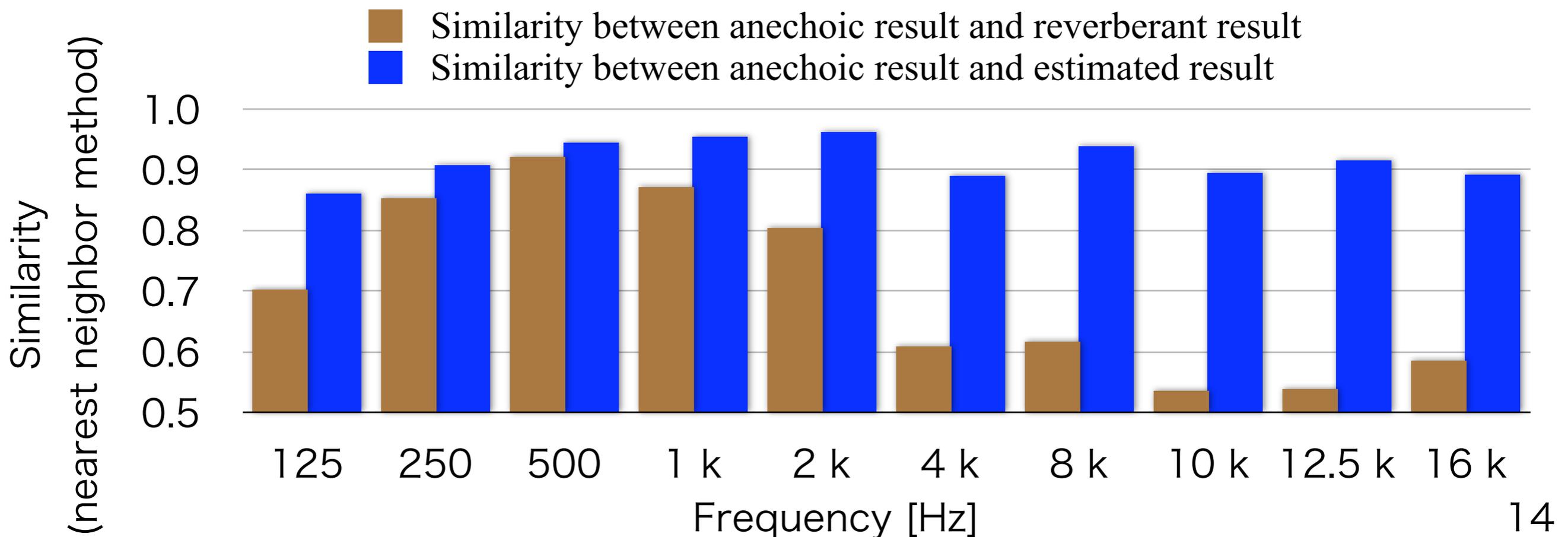
室内インパルスの特徴

抽出した放射指向特性 13

パターン照合における類似度を用いた性能評価

- パターンベクトルを用いた類似度(最短距離法)の比較
- パターンベクトル $P(f) = (x_{\theta_1}(f), x_{\theta_2}(f), \dots, x_{\theta_N}(f))$

■ 類似度
$$S(P_1(f), P_2(f)) = 1.0 - \frac{|P_1(f) - P_2(f)|}{|P_1(f)|}$$



まとめと今後の課題

■ まとめ

- エディタブルオーディオシステムの提案
- 包囲型マイクロホンアレイによる音源特性情報の抽出
 - ✳ 反射音の存在する環境における音源位置情報の抽出
 - ✳ 広帯域残響除去の実現による原音情報の高精度抽出
 - ✳ 音源位置情報, 原音情報を用いた室内インパルス応答の抽出
 - ✳ 壁面とマイクロホンとの距離を規範とした放射指向特性の抽出

■ 今後の課題

- ロバストな環境における原音情報の抽出
 - ✳ White-Limeの背景雑音に対するせい弱性問題の解決
- 放射指向特性の低域成分の抽出
 - ✳ マイクロホンと壁面との距離を変化させた測定による検討

包囲型マイクロホンアレイを用いた 音源位置情報と原音情報の抽出

■ 反射音に頑健な高精度音源位置情報抽出法の提案

T. Okamoto *et al.* 2007

- 従来法である空間平均化法を3次元位置推定に適用
- Delay-and-Sumにより反射音を低減後, MUSICにより高精度な音源位置推定を行うRAP-MUSICを提案
 - ➔ 従来法よりも高精度な音源位置情報抽出が可能

■ 広帯域ブラインド残響除去法の提案 岡本ら 2009

- スパース性のある観測信号にPre-Whiteningフィルタを導入し, 相関行列の条件数を緩和するWhite-LIMEを提案
 - ✳ スパース性のある観測信号に対しても高精度な残響除去が可能
 - ✳ 従来よりも広範囲な音信号に対する残響除去が可能
 - ➔ 高品質な原音の抽出が可能



これらを用いて音源の放射指向特性の抽出を試みる

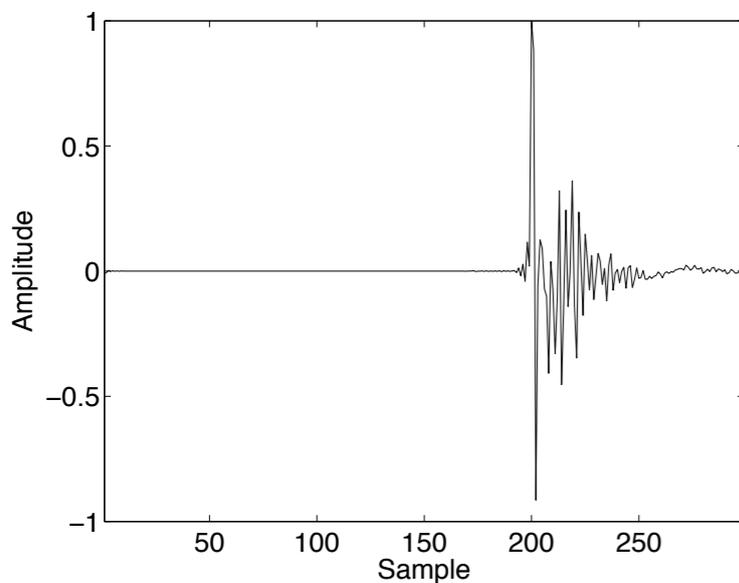
放射指向特性の測定

■ 放射指向特性のインパルス応答測定

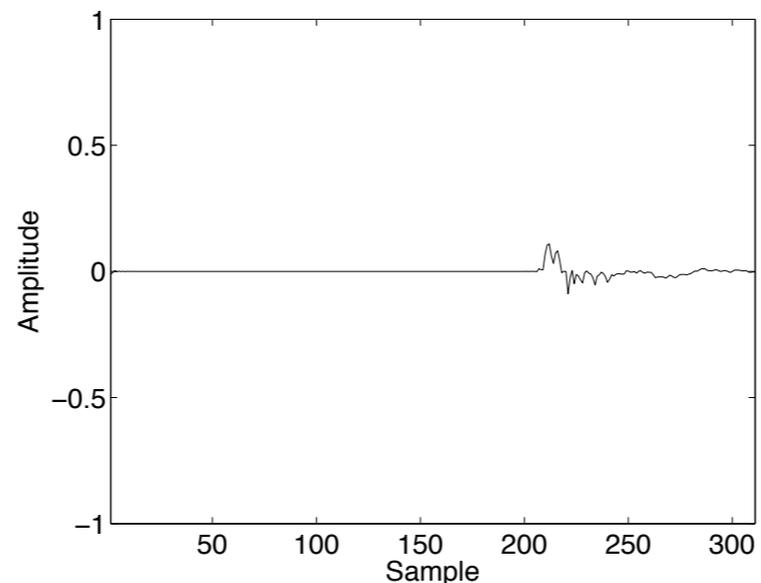
- 1ウェイスピーカの放射指向特性を測定
- スピーカ正面を0 deg. : 15 deg.間隔 ~ 180 deg.
- スピーカとの距離 : 1.5 m



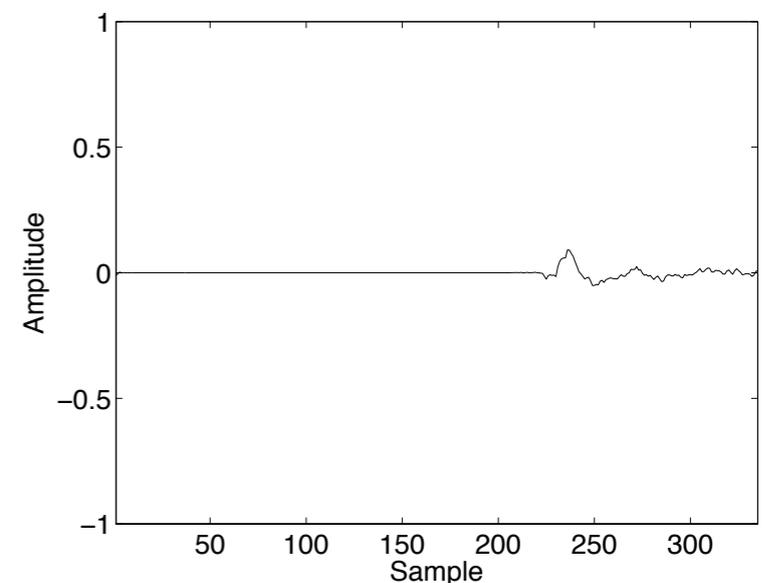
Micropure AP5001



0 deg.



90 deg.



180 deg.

測定したインパルス応答

インパルス応答の抽出シミュレーション

■ 原音情報を用いたインパルス応答の抽出

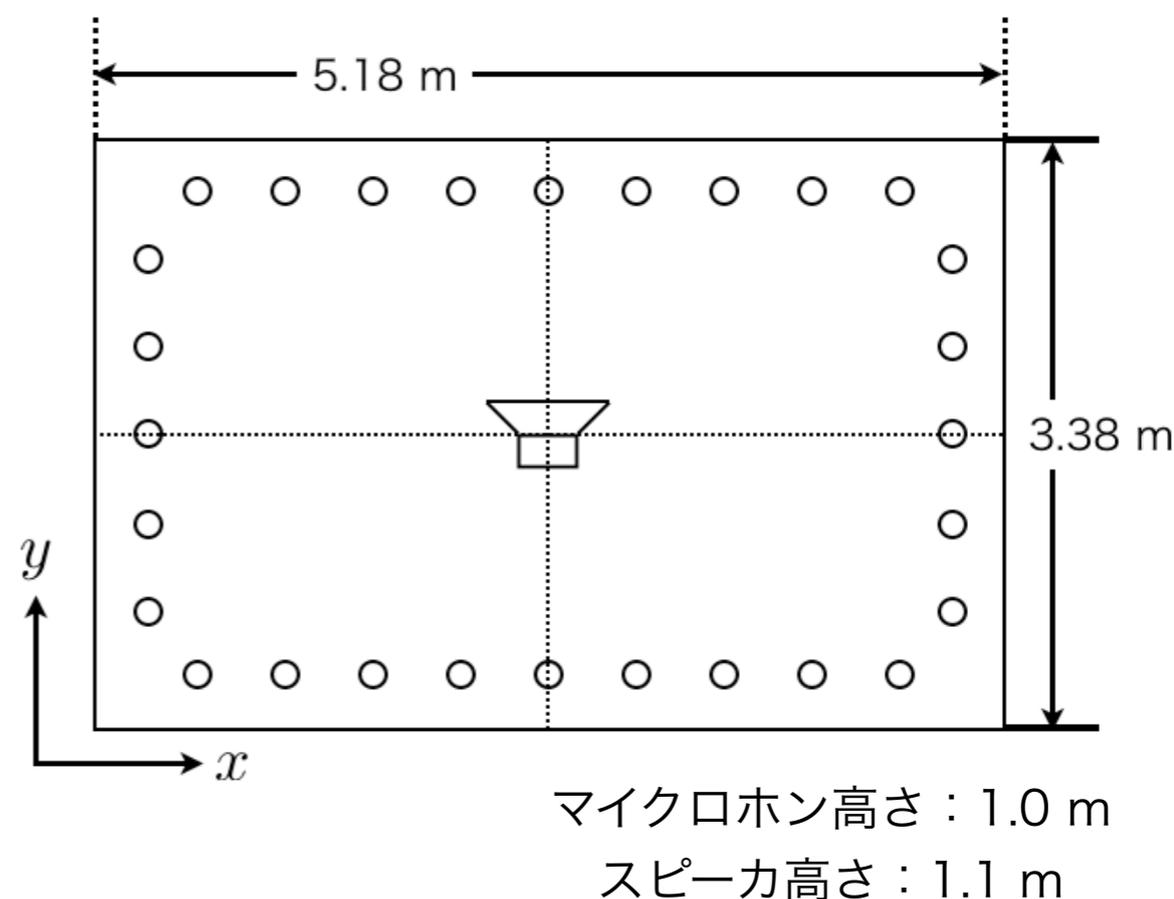
- 観測信号28 chから原音情報(44.1 kHzサンプリング音楽信号 2.6 s)を抽出
- 抽出した原音と観測信号の逆畳み込みからインパルス応答(6615タップ)を抽出

■ 抽出精度の性能評価

- Signal to Distortion Ratio (SDR)

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

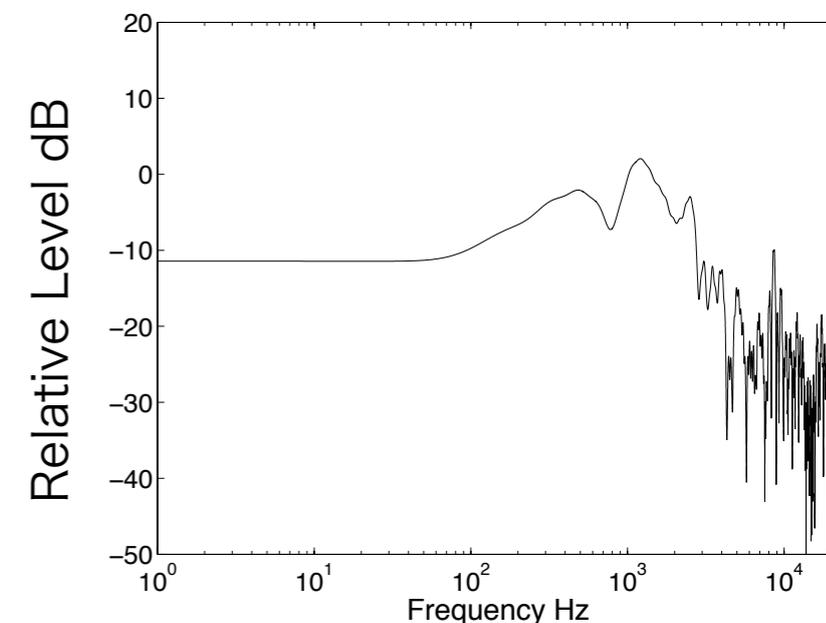
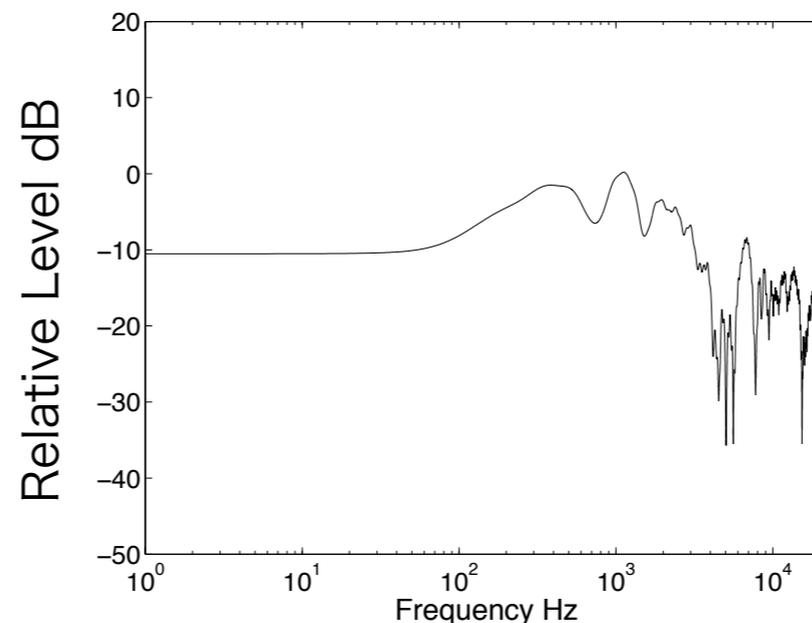
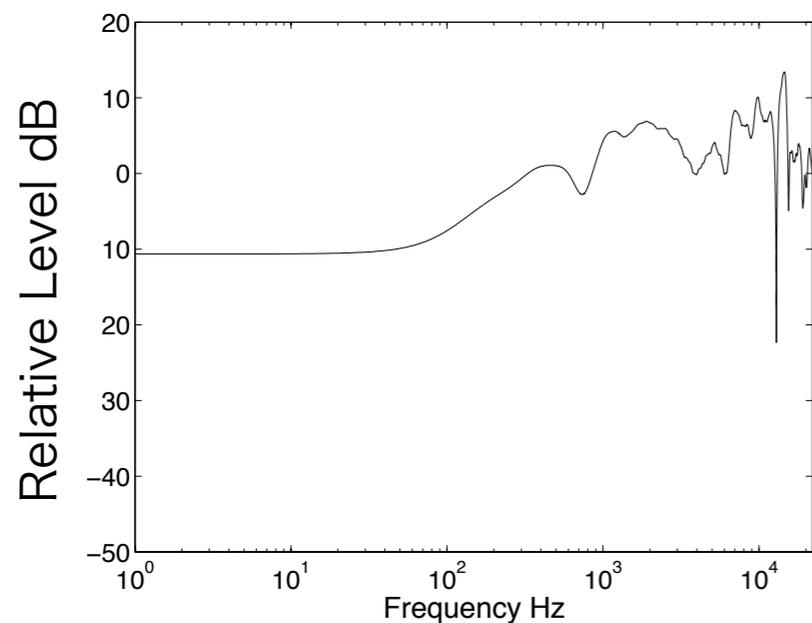
- 抽出された原音のSDR : 57.3 dB
- 抽出したインパルスのSDR : 62.6 dB
(28ch 平均値)



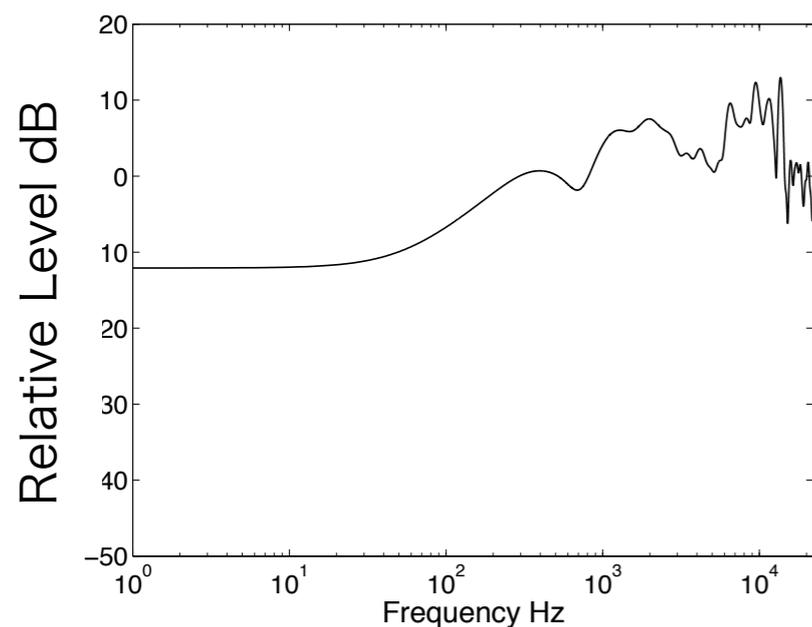
抽出した原音からインパルス応答を高精度に抽出可能

結果比較(0, 90, 180 deg.の周波数特性)

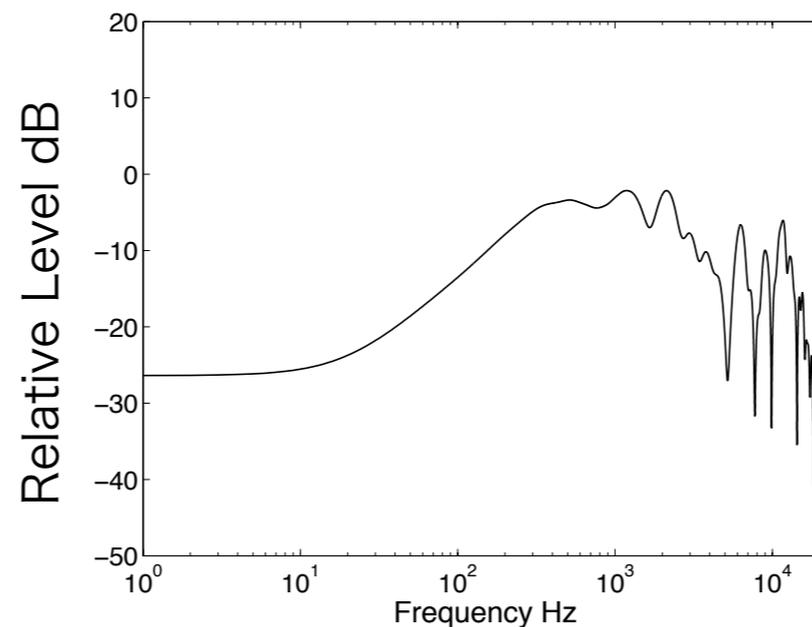
無響室における測定結果の周波数応答 (対数スケール)



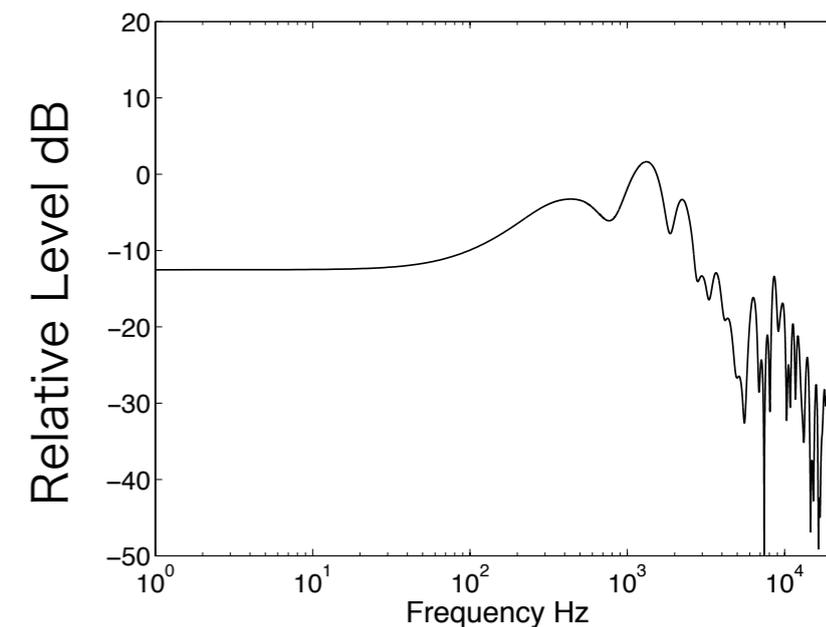
抽出した結果の周波数応答 (対数スケール)



0 deg.



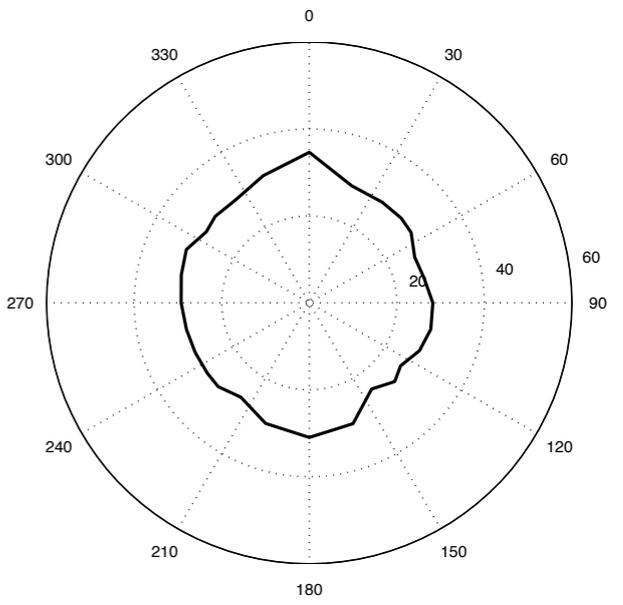
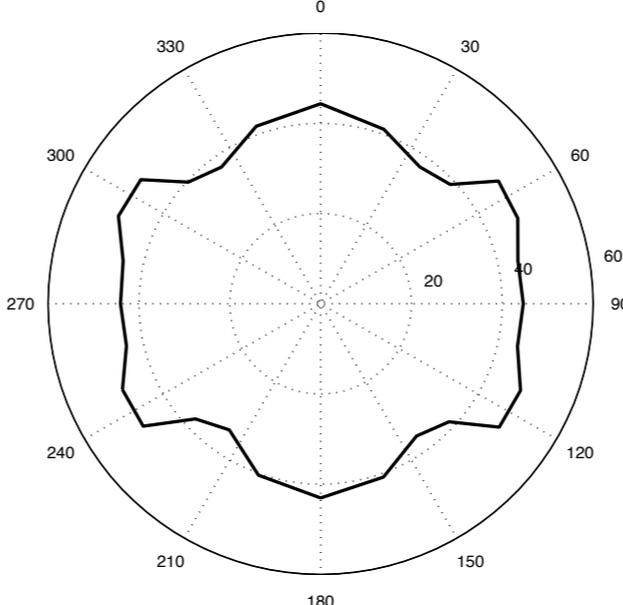
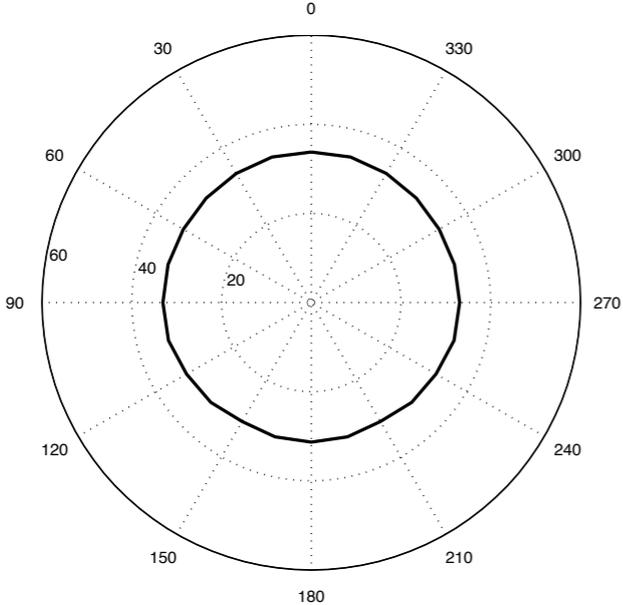
90 deg.



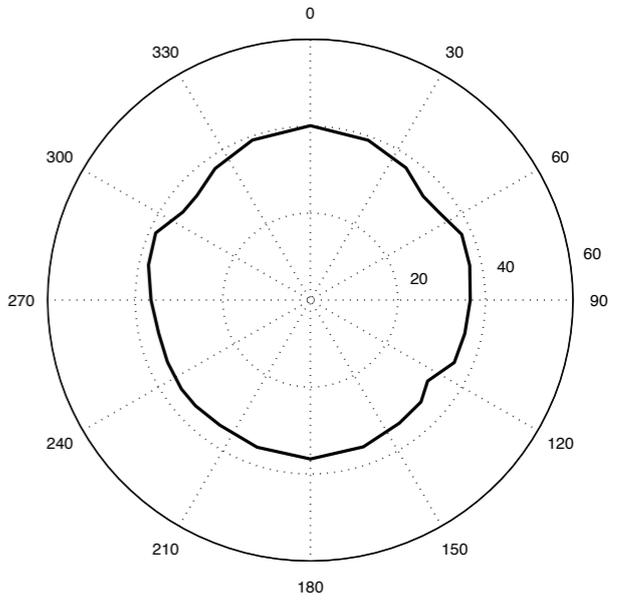
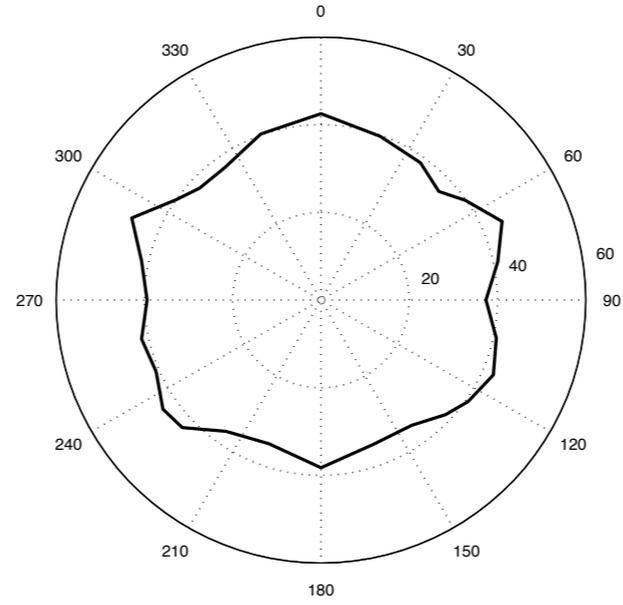
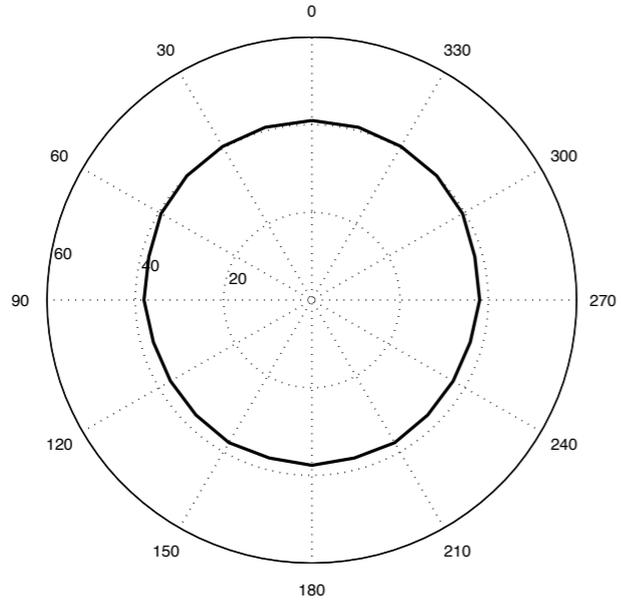
180 deg.

結果比較(125, 500 Hzの音圧レベル)

125 Hz



500 Hz



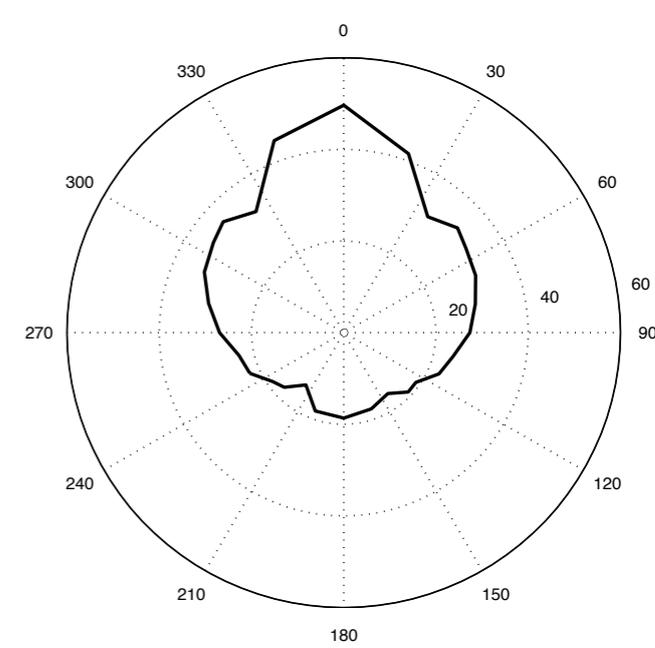
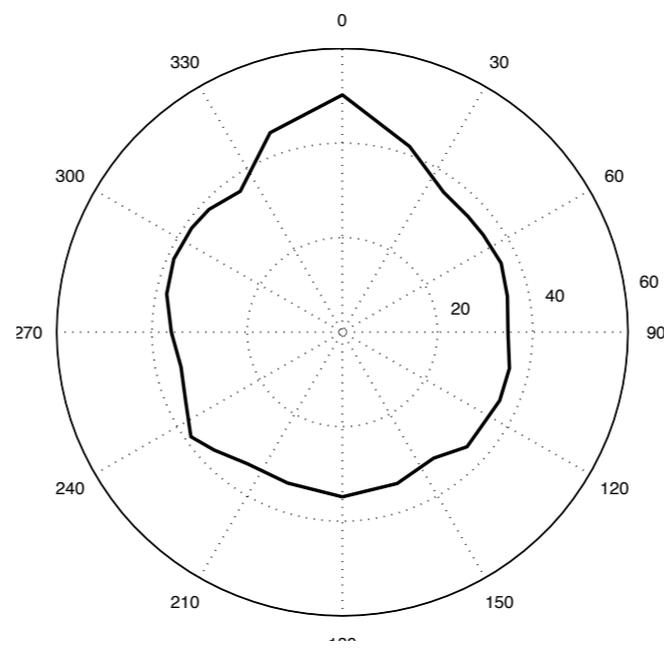
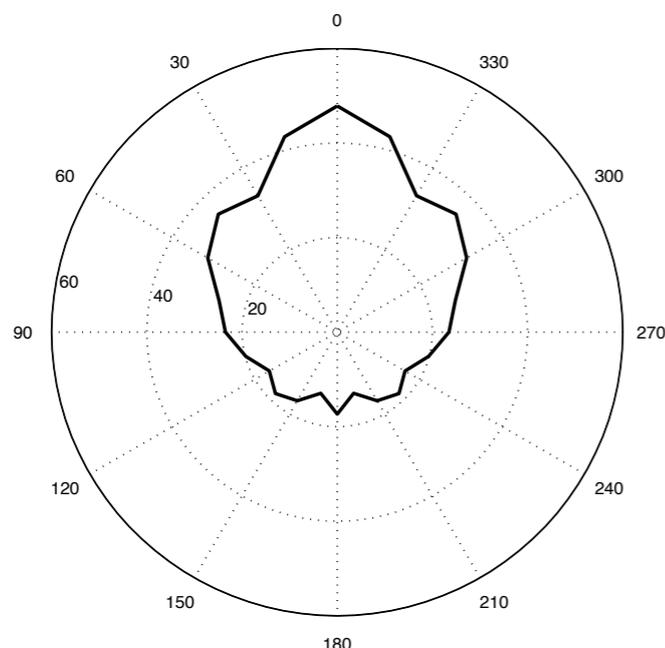
無響室測定の特徴

室内インパルスの特徴

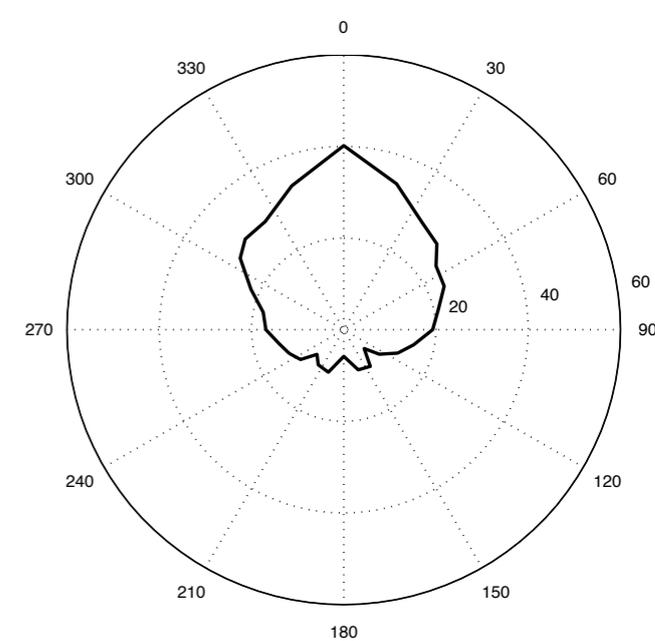
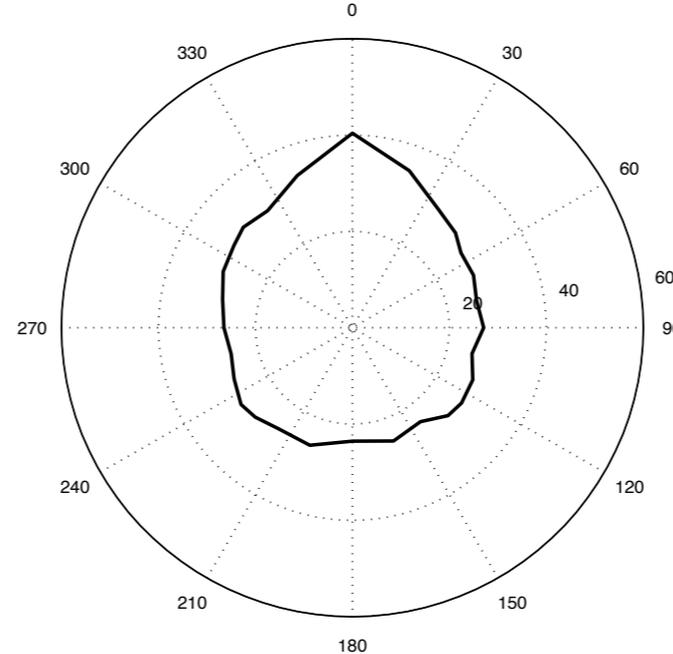
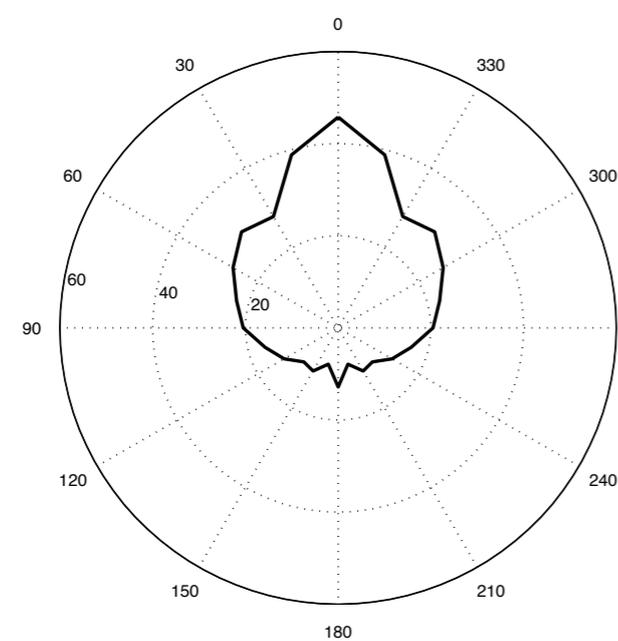
抽出した放射指向特性

結果比較(10, 16 kHzの音圧レベル)

10 kHz



16 kHz



無響室測定の特徴

室内インパルスの特徴

抽出した放射指向特性

評価結果について

■ 測定結果と抽出結果の比較

■ 正面(0 deg.), 真横(90 deg.), 背面(180 deg.)の結果

- ✳ 真横の～300 Hz以外では測定結果と同様の傾向

■ 方向ごとの1/3オクターブバンド解析結果

- ✳ 500 Hz以上では放射指向特性を抽出可能

■ パターン照合の結果

- ✳ 室内インパルス応答そのものは(波長が短い)高周波帯域では反射音の影響により精度が悪い
- ✳ 抽出したものは高周波数でも精度よく放射指向特性を模擬

■ 結果

- インパルス応答の初期反射までの切り出しにより放射指向特性を抽出可能

音空間を構成する5つの属性

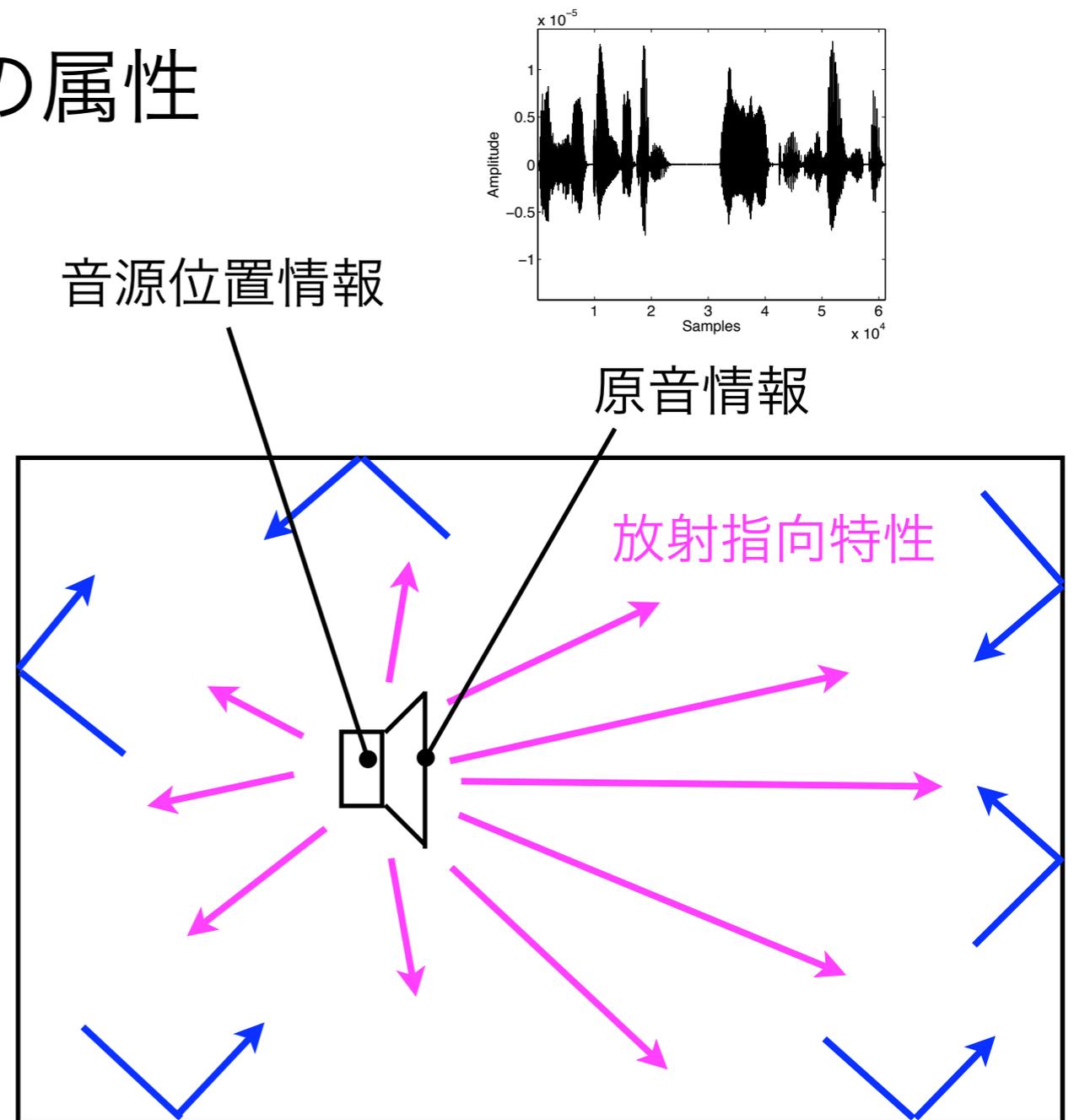
■ 室内音空間を構成する5つの属性

■ 音源の属性

1. 音源位置情報
2. 原音情報
3. 放射指向特性
…包囲による伝達特性の違い
(例：話者やスピーカの向き)

■ 空間 (=室内)の属性

4. 反射音情報
5. 後部残響特性
…反射や散乱等を繰り返し
方向性を失った波面の特性

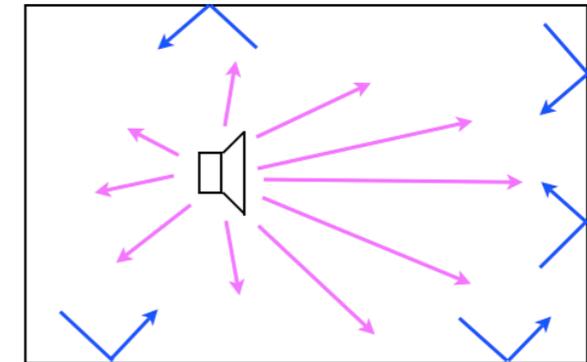


➡ 各音源ずつ存在, 極めて複雑な室内音場

エディタブルオーディオシステムの構築

■ 従来の音空間提示システム

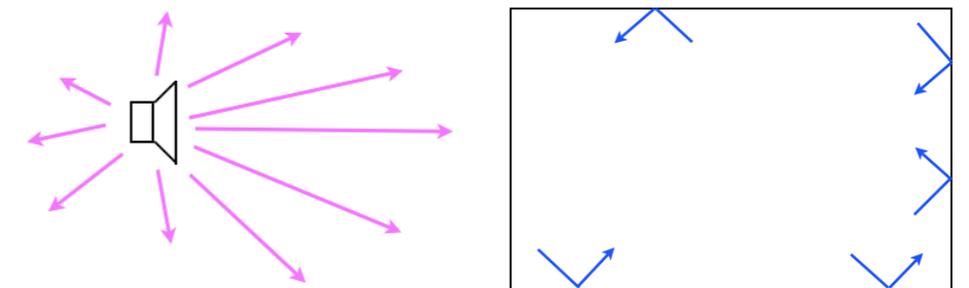
- 実空間全体の音空間の提示 (音場再現)
 - ➔ 属性ごとへの分解や提示は不可能



音空間全体

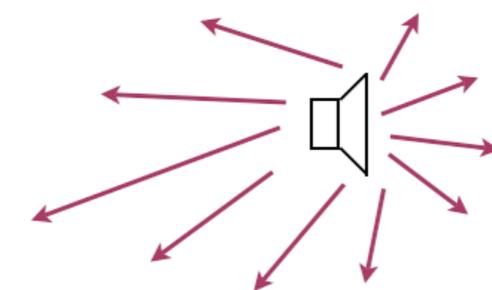
■ 次世代音空間提示システム

- 音空間から各属性ごとの完全な抽出
- +
- 各属性ごとのレンダリング技術の確立
 - ✳ 音場全体の提示
 - ✳ 特定の属性ごとの選択的な提示
 - ✳ 他イベントの属性の入れ替えや編集も可能



属性ごとの抽出と提示

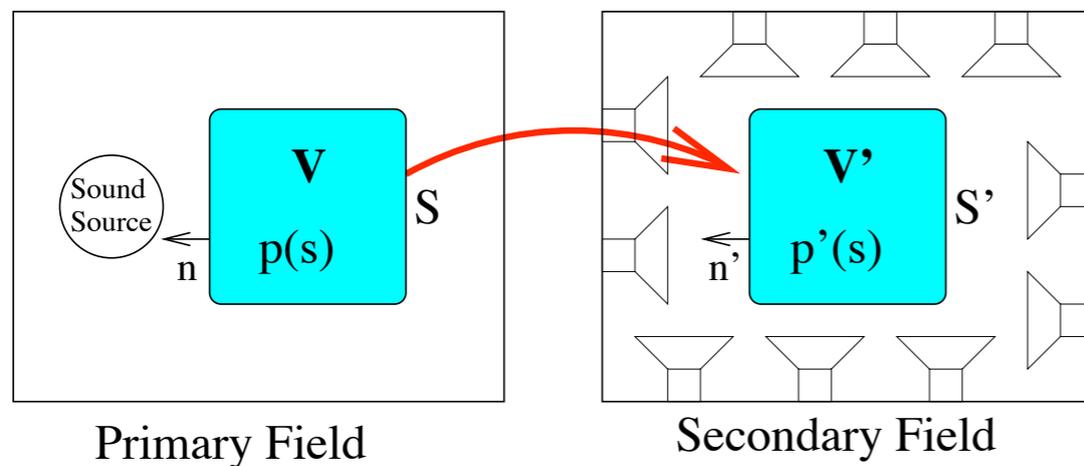
- ➔ これまでにはない新たな提示方法
- ➔ 多様性に富んだ臨場感通信技術へ



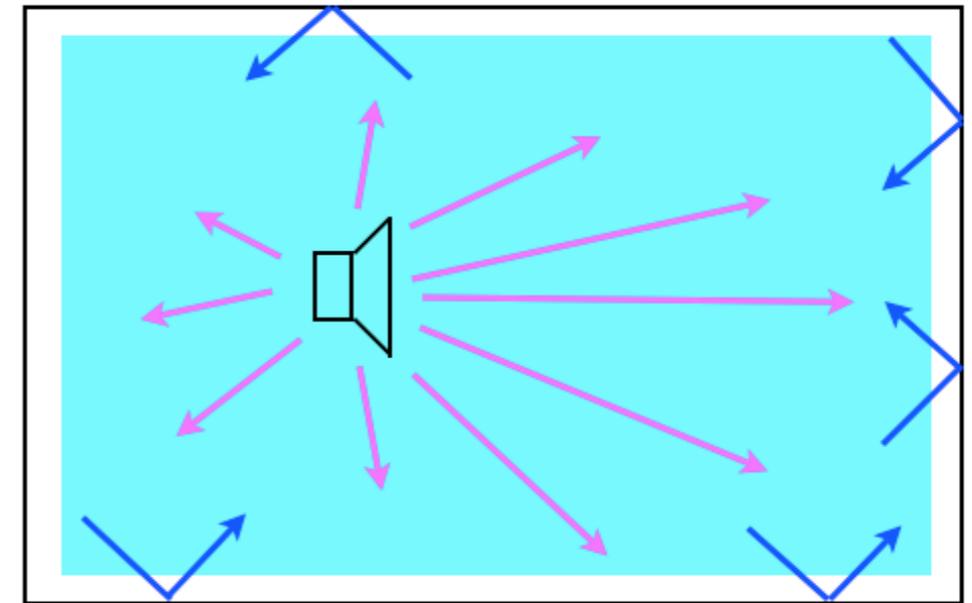
属性の入れ替え・編集

音空間全体の記録と音源特性の高精度抽出

■ 音空間全体の記録



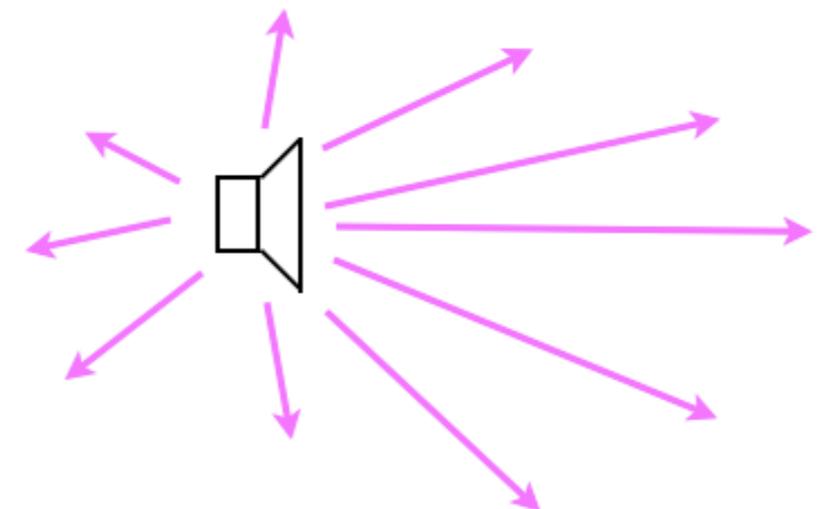
従来法：音場内の数点における收音



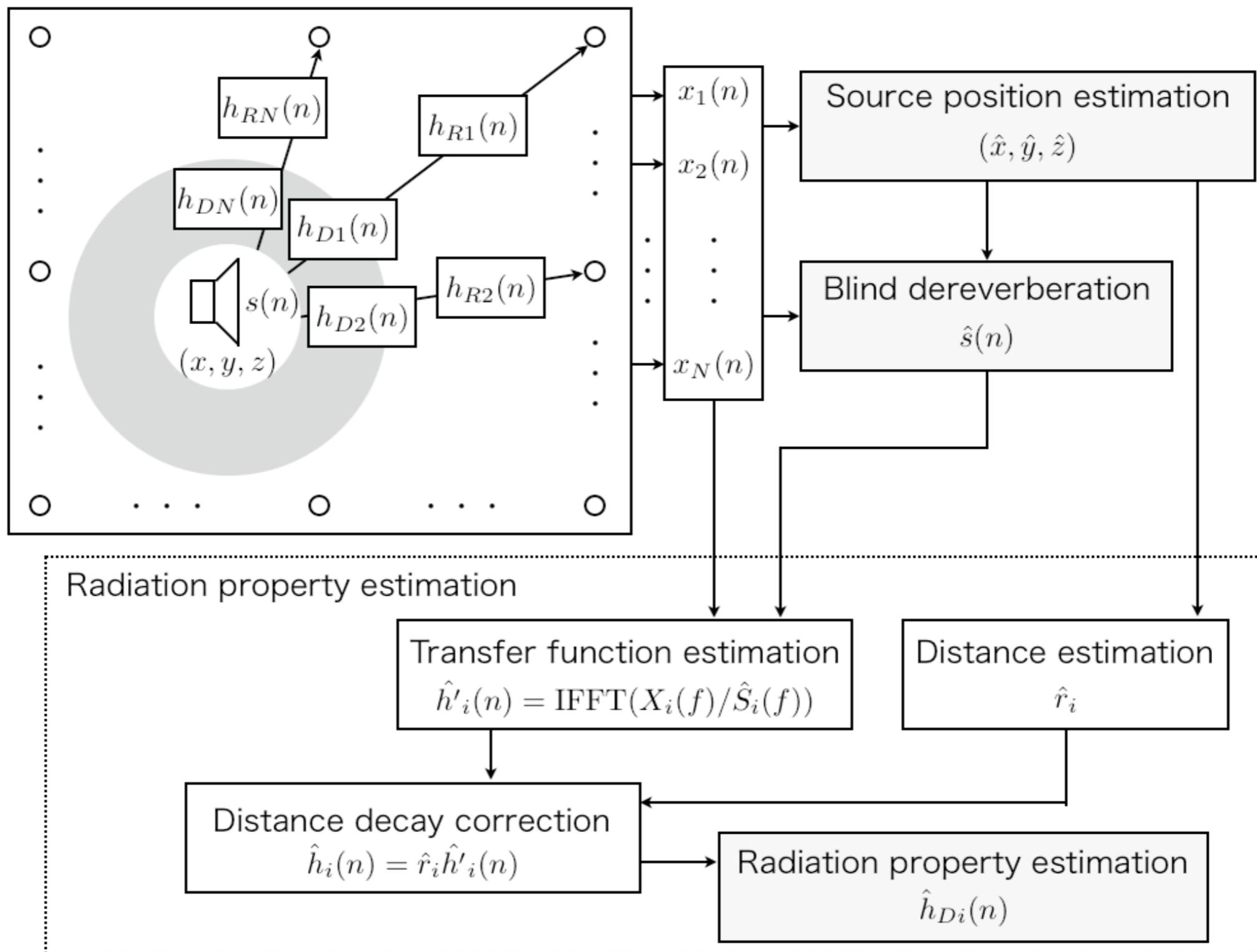
音空間全体の記録：音源を囲む

■ 放射指向特性を含む音源特性の抽出

1. 音源位置情報
2. 原音情報
3. 放射指向特性
 - ✳ 実環境の音源は点音源ではない
 - ✳ 従来法では抽出不可能



音源特性の属性値抽出の流れ



音源の放射指向特性抽出と課題

■ 放射指向特性抽出と課題

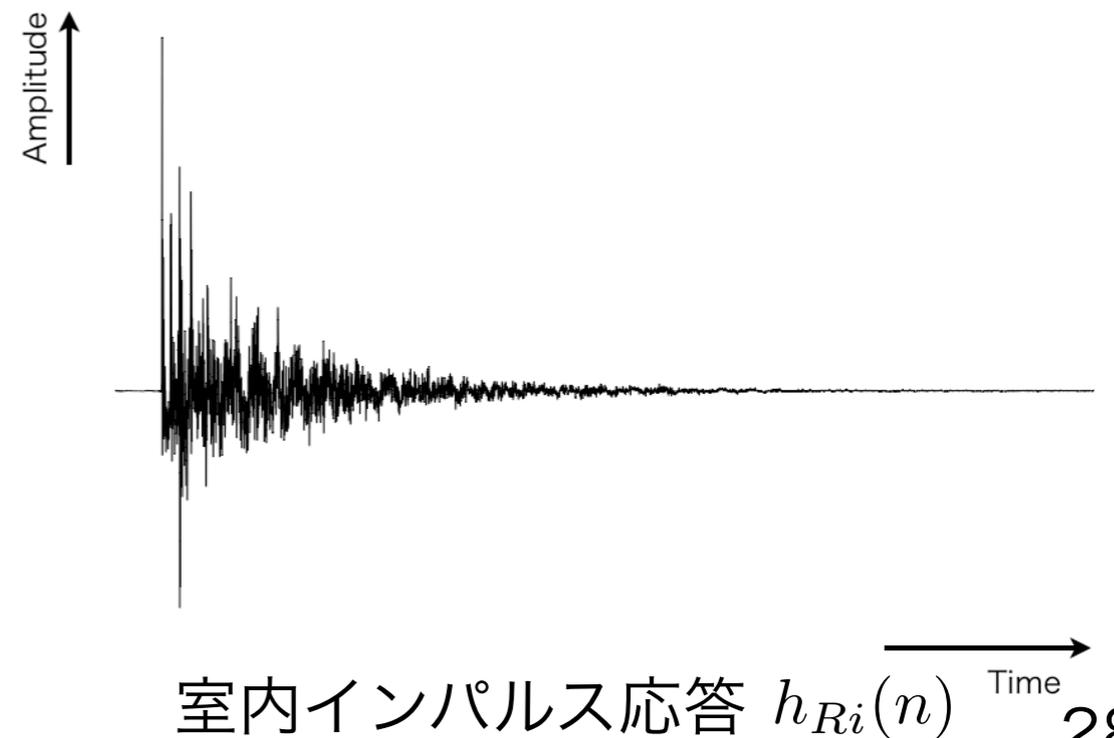
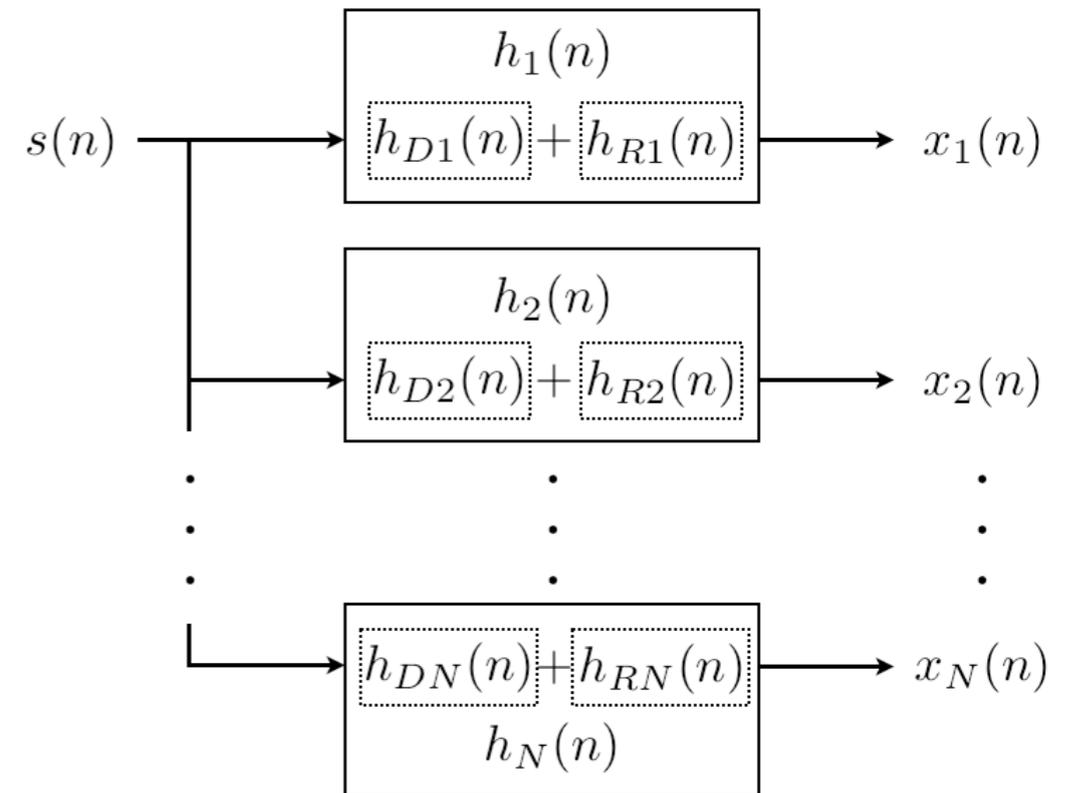
- 放射指向特性と室内伝達特性の分離
- 室内インパルス応答は非常に長い
 - ✳ 観測信号から直接 $h_i(n)$ 抽出は困難
 - ➔ 最初に位置情報と原音を抽出
 - ➔ 放射指向特性の分離と抽出

■ 原音抽出(=残響除去)と課題

- 直接逆特性を求めるブラインド残響除去
 - ✳ 広帯域では残響除去精度の劣化
 - ➔ 広帯域での残響除去が課題

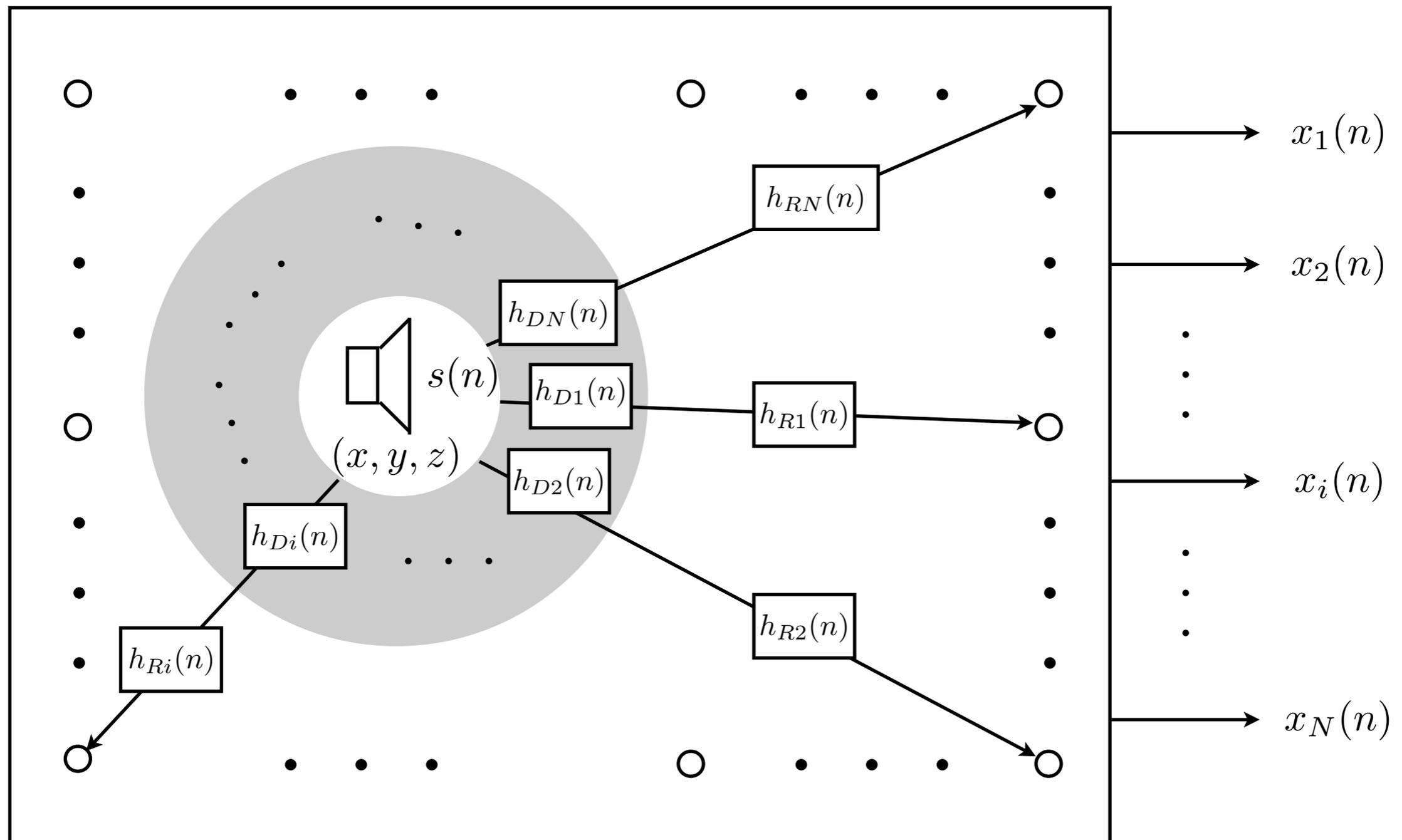
■ 音源位置情報の抽出と課題

- 室内での音源位置推定
 - ✳ 反射音により精度が劣化
 - ➔ 音源位置推定では未検討課題

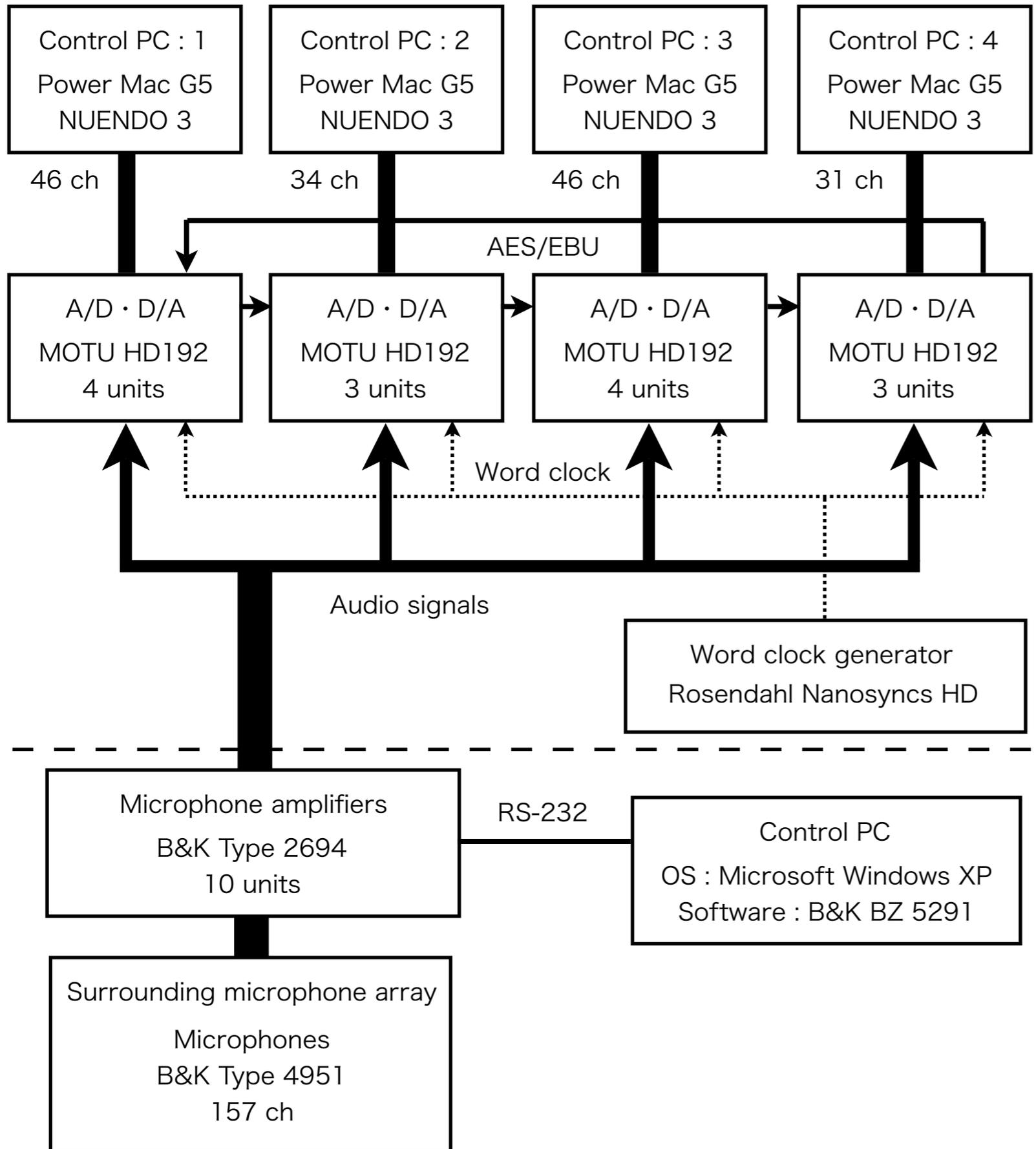


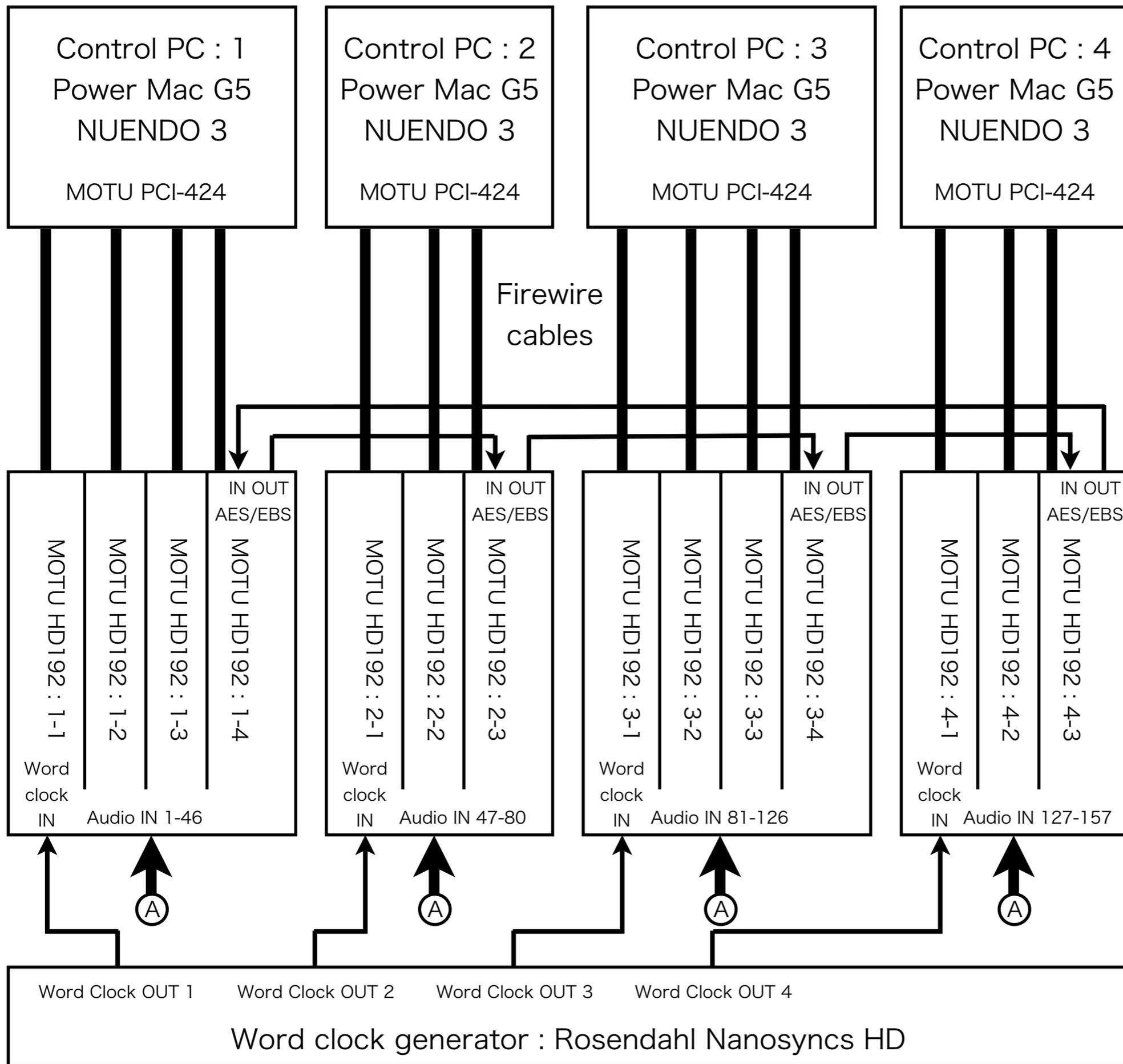
音源の放射指向特性抽出

■ 放射指向特性を考慮したSIMOモデル



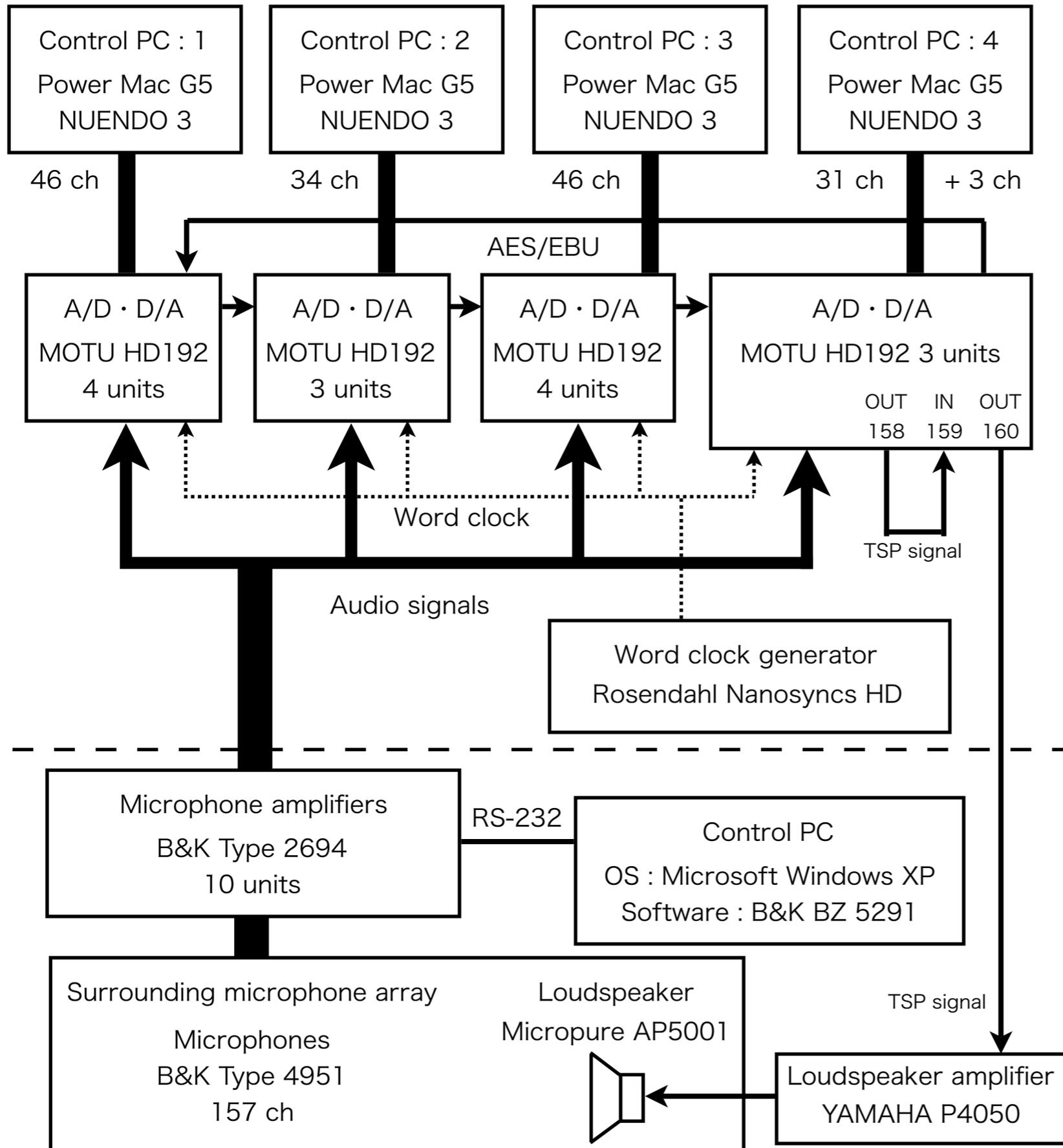
157 ch synchronized recording system



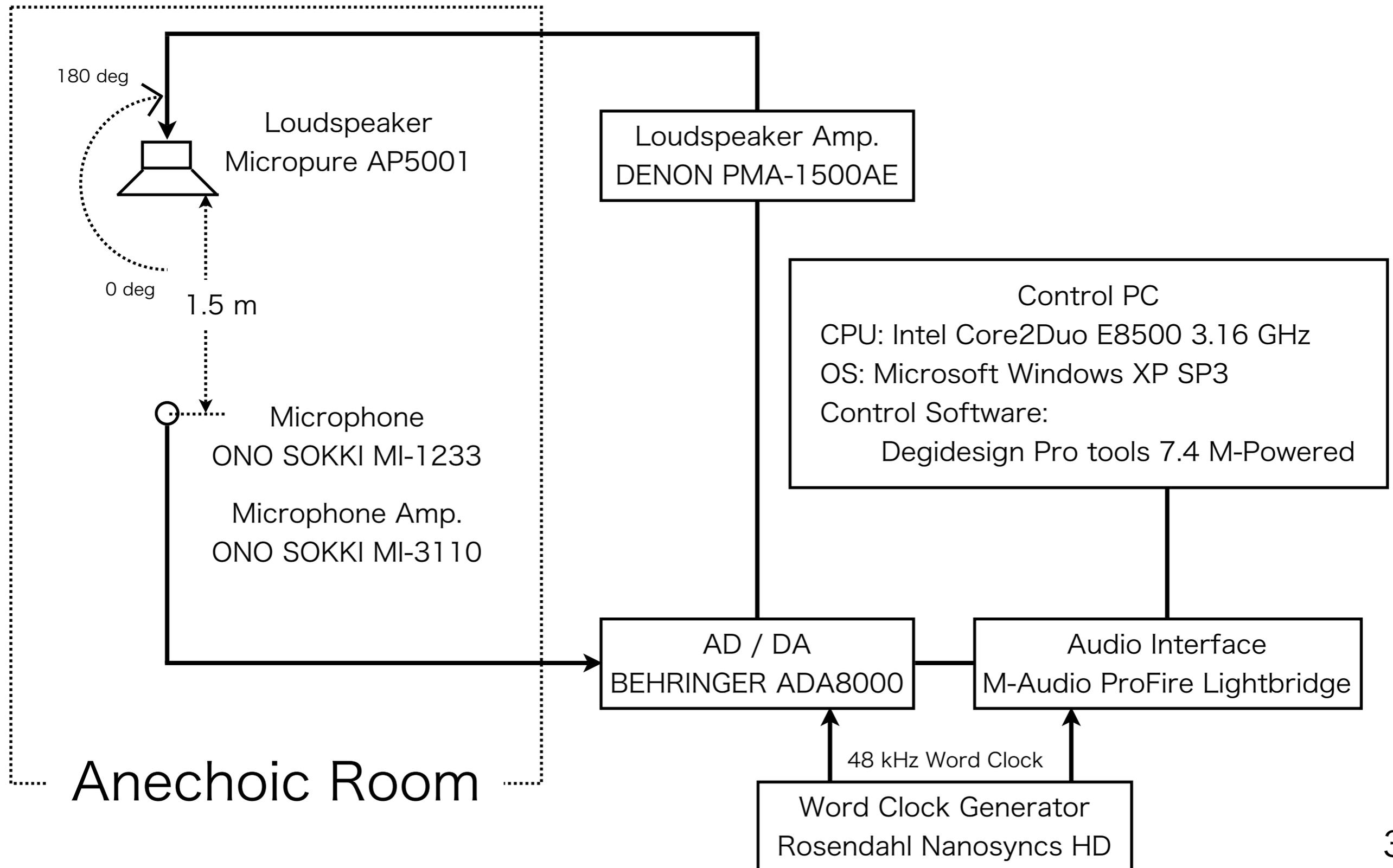


Ⓐ : Audio signals

157 ch synchronized recording system

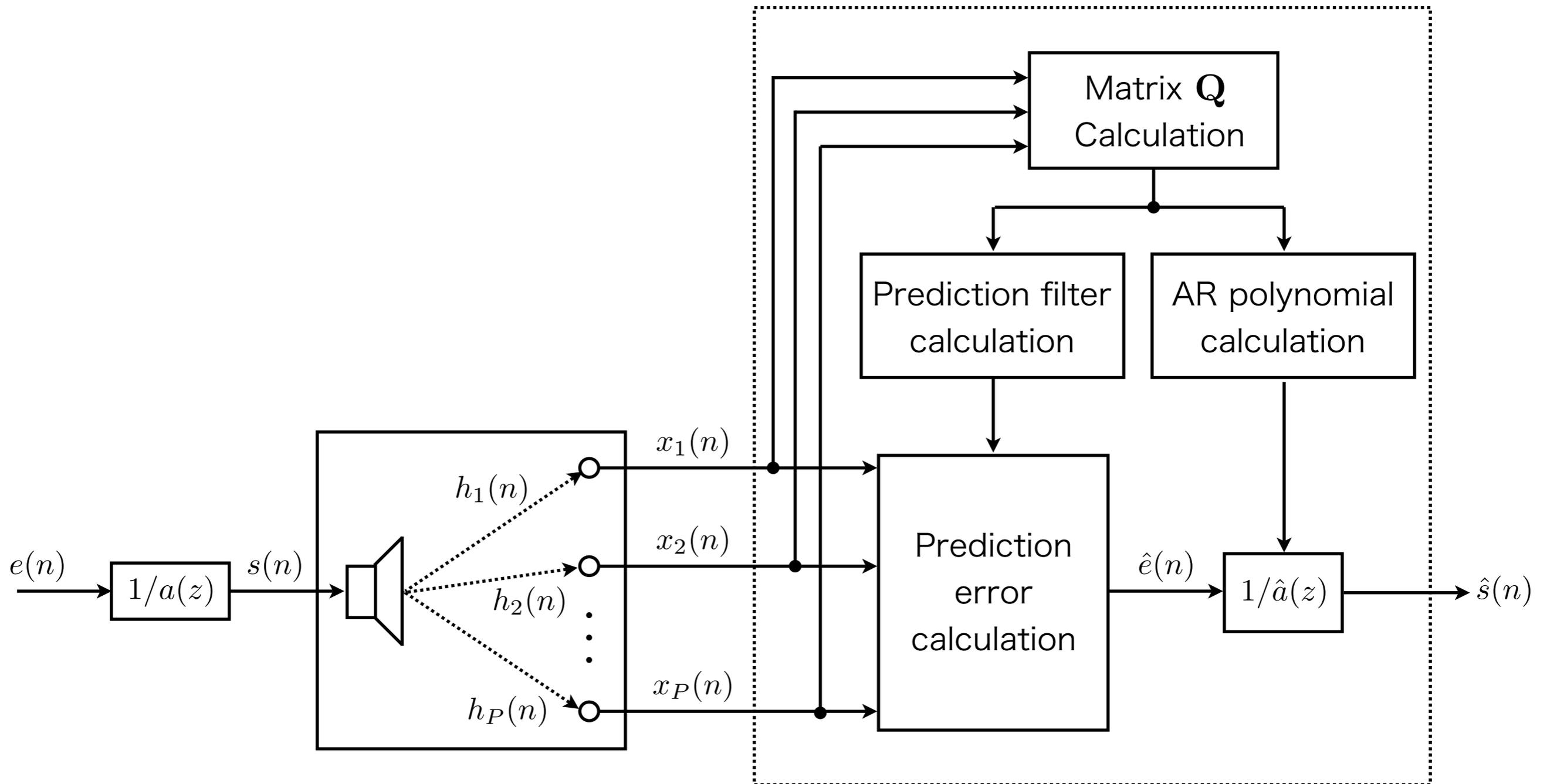


無響室における放射指向特性の測定系



LIMEアルゴリズム

M. Delcroix *et al.* 2007



Linear-predictive Multi-input Equalization (LIME)

Wihte-LIME

岡本ら 2009

