実世界情報システムプロジェクト
～ネオサイバネティックス研究グループ～
マイクロフォンアレイ計測

眞溪　歩
新領域創成科学研究科　複雑理工学専攻

1. 実音響環境での音響信号センシング

実音響環境において，対象話者の発話を聞き取るためには,大きく 3 つの問題として,雑音,残響，対象話者以外の発話が存在する．小さな会議室，大きなホール，大勢が机を並べる事務オフィス，自動車の中，パーティ会場など，これら 3 つの問題の程度は異なる.一方,人間は,さまざまな聴覚機構(2 つの耳と蝸牛でのアクティブセンシングなど)，視覚機構(リップモーションの認識など)，運動機構(音源の方向への姿勢制御など)，心理機構(発話内容の予測など),脳内計算処理(クロスモダリティの統合など)を利用し，これら 3 つの問題に立ち向かい，聞き取りを行なっている．もちろん聞き取れないという状況も発生するが，人間はかなり柔軟に対処できる．

機械による音声認識では，先の 3 つの問題の影響は極めて深刻である．S/N 比の低下や残響の増加に応じて，音声認識性能は著しく低化する．非常に限定された場面では，音響環境自体を改善することも可能であるし，話者ごとに接話マイクロフォンをつけるなども考えられる．しかし，ハンズフリーがキーワードとなるように，音声認識が期待される場面はさまざまであり，音声のセンシングに工夫が必要になる．センシングにおいては，3 つの問題に独立に対処するシステムを開発し，それらを直列に接続しても大きな効果は期待できない．この理由には,3 つの問題自体が独立ではないことが挙げられる.

このような問題に対し，マイクロフォンアレイの利用が考えられてきた．基本的に，マイクロフォンアレイは実音響環境に設置され，3 つの問題を含む音声を測定する．収集された音声は，到来方向の推定や独立成分分析などの枠組みで信号処理される．本研究では，実音響環境にしばしば存在する環境音楽にシステム同定用の信号を重畳させ，適応フィルタを設計する方法を提案する.

2. 実音響環境のシステム同定

実音響環境をシステムとして知るために，まず実音響環境のシステム同定を行なった．対象とした部屋は，サイズが 5m(W)x8m(L)x3m(H)程度あり，空調機器，コンピュータ・サーバなどの OA 機器から到来する雑音，隣接する部屋や近接する道路から到来する都市雑音が存在する一般的なオフィス環境であった．この部屋の中心付近に 2m 程度の間隔でスピーカとマイクロフォンを設置した.システム同定には M 系列信号の相関特性を利用することとした．しかし，一般的な音響機材では，帯域制限のために値域が 1 と–1 の M 系列信号を音声信号化できない．そこで，M 系列信号を離散フーリエ変換し，音声帯域以上の高周波成分に零づめし，逆フーリエ変換した信号を出力した．この信号を図 1 上段に示す．この処理はフーリエ級数の部

分和を取り出すことに相当し，生成される信号はもとの M 系列信号のリップルつき近似となり帯域制限される．また，この M 系列信号の部分和は，システム同定に要求される良好なデルタ状の自己相関特性と，異なる M 系列間の良好な無相関特性を保存していた．

図 1 下段に，この部屋の音響システムのインパルス応答を示す．ただし，伝播に伴うむだ時間は表示していない．横軸はデータポイント表しており，このときのサンプリング周波数は 12kHz であった．300 ポイント(25ms)程度までは比較的大きい残響が確認される．なお，スピーカから出力する M 系列信号の近似波形の音圧を上げると，有限振幅音響の領域に入り，非線形波形ひずみが観察された．

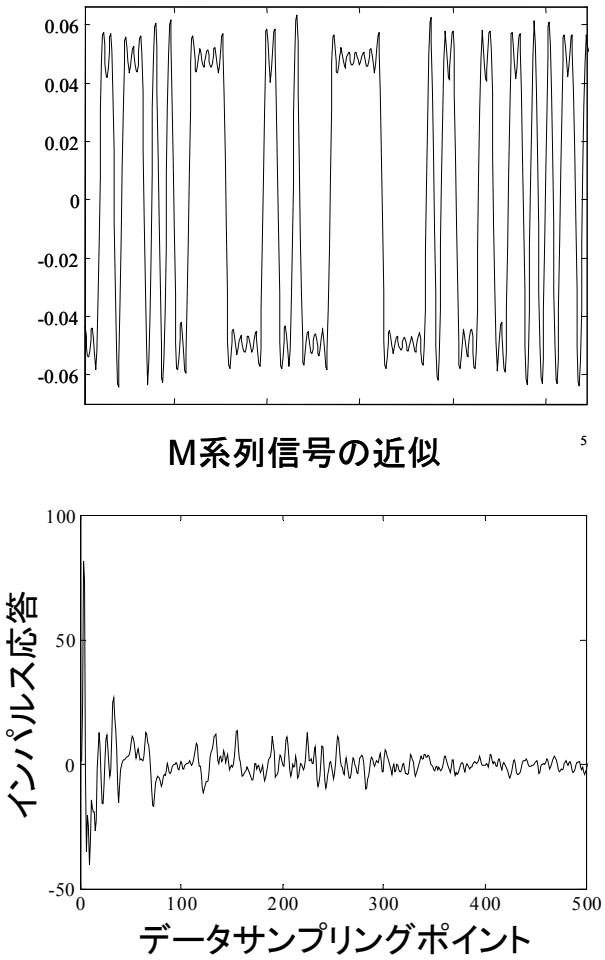

図 1．実音響空間のシステム同定

3．マイクロフォンアレイの利用

図 2．マイクロフォンアレイ信号処理

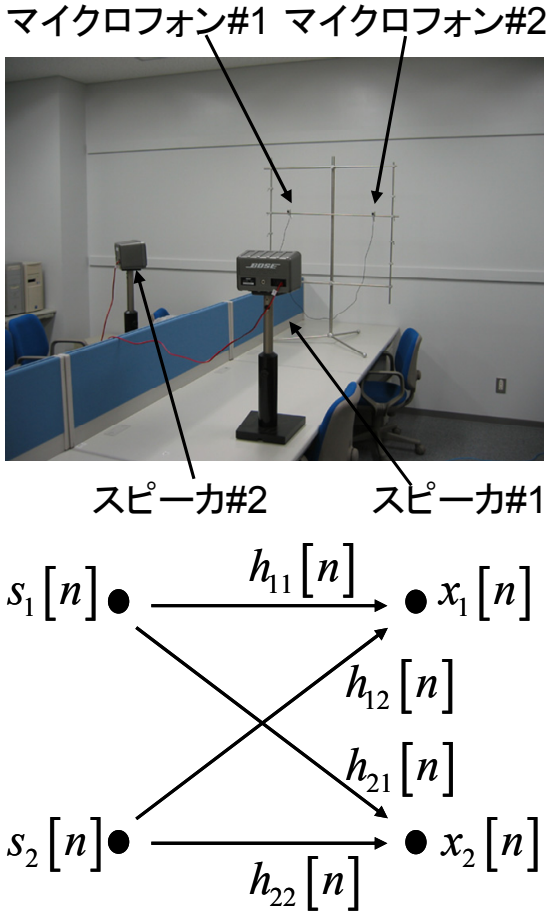

図 2 上段に示すように，室内に 2 つのスピーカと 2 つのマイクロフォンを設置する．スピーカ#1，#2 はそれぞれ音声 $s_1[n], s_2[n]$ を出力し，マイクロフォン#1, #2 はこれらの混合音声をそれぞれ $x_1[n], x_2[n]$ として測定する．いま，図 2 下段に示すように，スピーカ $\#p, (p=1,2)$ からマイクロフォン $\#m, (m=1,2)$ への伝達となるインパルス応答を $h_{mp}[n]$ とすると，

$$\begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} = \begin{bmatrix} h_{11}[n] & h_{12}[n] \\ h_{21}[n] & h_{22}[n] \end{bmatrix} \otimes \begin{bmatrix} s_1[n] \\ s_2[n] \end{bmatrix} \tag{1}$$

となる．ここで，⊗はたたみ込みを表す記号である．式(1)を z 変換 $X(z) = \sum_{n \in \mathbb{Z}} x[n] z^{-n}$ すると，

$$\begin{bmatrix} X_1(z) \\ X_2(z) \end{bmatrix} = \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix} \begin{bmatrix} S_1(z) \\ S_2(z) \end{bmatrix} \qquad (2)$$
$$\mathbf{X}(z) = \mathbf{H}(z)\mathbf{S}(z)$$

となる．簡単のためにスピーカ(音源)2 つ，マイクロフォン 2 つの例を取り挙げているが，サイズの拡張は容易であり，雑音も音源として取り込むことができる．ただし，計測系に付随する雑音は加法的になる．いま，インパルス応答の z 変換である伝達関数 $\mathbf{H}(z)$ は既知であるとすると，シンプルな雑音，残響，音声分離は

$$\mathbf{H}^{-1}(z) = \frac{1}{\det\{\mathbf{H}(z)\}} \begin{bmatrix} H_{22}(z) & -H_{12}(z) \\ -H_{21}(z) & H_{11}(z) \end{bmatrix} = \frac{1}{\det\{\mathbf{H}(z)\}} \tilde{\mathbf{H}}(z) \qquad (3)$$

を求めることに他ならない．式(3)において，$1/\det\{\mathbf{H}(z)\}$ は $S_1(z), S_2(z)$ のどちらにも共通の影響を与えているため，雑音・音声分離は余因子行列 $\tilde{\mathbf{H}}(z)$ のみによって行なわれていることがわかる．残る，残響除去と周波数特性補正は余因子行列 $\tilde{\mathbf{H}}(z)$ と $1/\det\{\mathbf{H}(z)\}$ で分担している．ここで，$\det\{\mathbf{H}(z)\}$ の零点がすべて z 平面の単位円内に存在しなければ，すなわち $\det\{\mathbf{H}(z)\}$ が最小位相特性でなければ，$1/\det\{\mathbf{H}(z)\}$ は因果的で安定にはならない．最小位相特性は，実音響環境ではまったく期待できないため，このような条件設定では音声分離よりむしろ残響除去の方が困難である．さらに，もとの音声 $S_p(z)$ は伝達関数 $\mathbf{H}(z)$ の要素として $H_{mp}(z)$ なるシステムを経験して混合音声 $X_m(z)$ となるが，式(3)による処理過程で，もう一度余因子行列 $\tilde{\mathbf{H}}(z)$ の要素として $H_{mp}(z)$ なるシステムを経験することになる．つまり，余因子行列 $\tilde{\mathbf{H}}(z)$ だけを作用させた状態では，音源分離は可能であっても，周波数特性はさらに劣化し，残響もさらに長くなっている．伝達関数 $\mathbf{H}(z)$ の行列のサイズが大きくなれば，この影響はより深刻になる．

さて，$1/\det\{\mathbf{H}(z)\}$ の安定化には，ある程度の処理の遅れを許容し，非因果的な部分を吸収することになる．

遅れを許容するなら，ウィナーフィルタ $\mathbf{F}(z)$ を，

$$\hat{\mathbf{S}}(z) = \mathbf{F}(z)\mathbf{X}(z)$$
$$\hat{\mathbf{S}}(z) \simeq \begin{bmatrix} z^{-p} & 0 \\ 0 & z^{-q} \end{bmatrix} \mathbf{S}(z) \qquad (4)$$

となるように設計できる．なお，近似は最小 2 乗近似の意味である．ここで，ウィナーフィルタ $\mathbf{F}(z)$ は，たたみ込みをテープリッツ行列を用いた行列積で表現することによって求める．この場合では，まず伝達関数 $\mathbf{H}(z)$ の各要素の係数列を個々に逐次シフトして並べたテープリッツ行列を作成し，つぎにこれらの行列をあらためて小行列とする行列を作成し，最後にこの行列の擬似逆行列を求める．定性的には，ウィナーフィルタ $\mathbf{F}(z)$ の音源分離・残響除去性能は高い．

ウィナーフィルタ $\mathbf{F}(z)$ は高性能ではあるが，実音響環境での利用を考えると，伝達関数 $\mathbf{H}(z)$ が既知であるというところに疑問が残る．実践的場面では，スピーカは話者に変り，話者からマイクロフォンへの音響システムは未知である．このような条件設定での音源分離は，

ブラインド音源分離と呼ばれる．さて，伝達関数$\mathbf{H}(z)$は因果的で安定と考えられるので，zを複素平面の単位円上にとる離散時間フーリエ変換$\mathbf{H}[\Omega]=\left[\sum_{n\in\mathbb{Z}} h_{mp}[n]\exp(-\mathrm{j}\Omega n)\right]$に置き換えても情報の損失はない．ここで，$\Omega\in\mathbb{R}[-\pi,\pi)$は規格化角周波数である．また，より実践的に，混合音声$\mathbf{x}[n]$をある時間幅Nで切り出し，伝達関数$\mathbf{H}(z)$を離散フーリエ変換$\mathbf{H}[k]=\left[\sum_{n=0}^{N-1} h_{mp}[n]\exp(-\mathrm{j}2\pi kn/N)\right]$で代替する．ここで，$k$は$k$番目の離散周波数である．利用したい性質は，式(1)と式(2)のように，変換を介してたたみ込みが積に変化する性質であり，循環性の問題はあるものの，この性質は離散フーリエ変換にも継承されている．この状況設定では，Minimum Variance Beamformer (MVB)，Multiple Signal Classification (MUSIC)，Independent Component Analysis (ICA)などが利用できる．ただし，ICA 以外は完全にブラインド音源分離とは呼べず，ステアリングベクトルが必要となる．ステアリングベクトルとは，マイクロフォンアレイに対し角θをなす方向の離散周波数kの位相応答を，各マイクロフォンに対して並べた列ベクトル$\mathbf{d}[\theta,k]$である．MVBによるθ方向の音源の推定は，

$$\frac{\mathbf{d}^{\mathrm{H}}[\theta,k]\left\langle\mathbf{X}[k]\mathbf{X}^{\mathrm{H}}[k]\right\rangle^{-1}}{\mathbf{d}^{\mathrm{H}}[\theta,k]\left\langle\mathbf{X}[k]\mathbf{X}^{\mathrm{H}}[k]\right\rangle^{-1}\mathbf{d}[\theta,k]}\mathbf{X}[k] \tag{5}$$

となる．ここで$\langle\bullet\rangle$は周波数平均である．一方，MUSIC によるθ方向の音源の推定は，

$$\frac{\mathbf{d}^{\mathrm{H}}[\theta,k]\mathbf{d}[\theta,k]}{\mathbf{d}^{\mathrm{H}}[\theta,k]\mathbf{V}_{\mathrm{N}}\mathbf{V}_{\mathrm{N}}^{\mathrm{H}}\mathbf{d}[\theta,k]} \tag{6}$$

となる．ここで，$\mathbf{V}_{\mathrm{N}}$は$\left\langle\mathbf{X}[k]\mathbf{X}^{\mathrm{H}}[k]\right\rangle$の相対的に小さい固有値に対応する固有ベクトルだけを列として並べた縦長の行列である．MVB(5)，MUSIC(6)ともに，若干形は異なるものの，離散フーリエ変換$\mathbf{X}[k]$のステアリングベクトル$\mathbf{d}[\theta,k]$への計量行列つき射影ということができる．一方，ステアリングベクトルも必要としない完全なブラインド設定の ICA には, MVB(5)や MUSIC(6)のように陽には記述できないものの，infomax や fastICA などさまざまなアルゴリズムが存在する．残響が存在しない場合，わざわざ周波数領域に変換する必要もなく，時間領域での ICA の音源分離性能は著しい．ただし，もとの音声が互いに無相関を通り越して互いに独立である必要がある．現実には，ICA は独立の必要条件のいくかを手がかりにするだけで，もとの音声の独立性が崩れることが，直ちに音源分離の破綻を意味するわけではない．一方，ICA には分離された個々の信号の順番とスケールに不確定性が存在する．残響が存在する場合，MVB(5)や MUSIC(6)のように周波数領域での ICA の利用となるが，切り出された各時間帯での分離結果と，異なる時間帯での分離結果との対応が問題となる．また，MVB，MUSIC，ICA ともに，積極的な残響除去対策は講じられていない．また，ここで紹介した手法以外に，ヒューリスティックな手法も多く提案されている．

このように，マイクロフォンアレイによる音源分離手法は，問題設定が異なるため，一概に比較できない．また，実音響環境の問題設定は，完全にブラインドでもなく，かといってシステムが既知ともいえない．実音響環境にしばしば存在する環境音楽にここで紹介したような M 系列信号を重畳させれば，システムは部分的に既知となる．このことを利用すれば，ウィナーフィルタを部分的に更新していくことができる．