「独立成分分析とその応用特集号」 解　説

畳込み混合のブラインド音源分離

牧野　昭二*・荒木　章子*・向井　良*・澤田　宏*

1. まえがき

コンピュータによる音声認識技術は年々進歩しており，静かな環境で接話マイクに向かって丁寧に話した言葉であれば，かなり高い精度で認識できるようになっている．しかしその一方で，様々な背景音，たとえば周囲の人の声，音楽，騒音，さらには残響などがある環境では，認識性能は急激に低下する．私たちが普段それほど意識せずに行っている「聞きたい音を聞き分ける」という能力がコンピュータには欠けているのである．

マイクロホンから離れた複数人の発話を認識する場合には，目的音声とその他の妨害音との混合や，残響の影響が問題となる．このような状況でも聞きたい音声をコンピュータで認識するためには，たくさんの音の中から聞きたい音を分離抽出することが必要となる．これが音源分離の目標である．この音源分離技術は，多様な音が存在する中で音声認識システムへ適切な入力を与えるための重要な要素技術である．

たくさんの音の中から聞きたい音を聞き分ける音源分離技術として，近年，独立成分分析（Independent Component Analysis: ICA）に基づくブラインド音源分離（Blind Source Separation: BSS）が脚光を浴びている．これは，複数音源が統計的に互いに独立であるという仮定のみを用い，分離信号が互いに独立となるようなフィルタを求める手法である．たとえば，2人の人が同時に話しても，それぞれの音声は互いに独立である．同様に，実環境におけるほとんど全ての音源は互いに独立であるとみなすことができる．独立成分分析に基づく分離手法は，聞きたい音とそれ以外の音，すなわち雑音との間の独立性に着目し，音源に関する事前情報を用いずに，すなわちブラインドな処理で，収音した混合信号から聞きたい音声を分離・復元する．この手法は，音源位置の知識や妨害音区間の切り出しを原理的に必要とせず，音源信号の調波構造などの仮定も用いない．

本稿ではICAを広くとらえる，すなわち，高次統計量に基づく非線形相関除去手法だけでなく2次統計量に基づく非定常相関除去手法，非白色相関除去手法も含めてこれらの三つの手法を統一的に論じる [1–3].

統計的処理であるICAは，物理的，音響的にはある種のブラックボックスであり，その中で何が行われているのか，何がどこまで分離できるのかがあまり分かっていなかった．我々はこれまでの研究により，統計的手法であるICAを音響信号処理的な観点から分析して物理的意味付けを与え [4]，従来の音響信号処理技術との関係を解明した [5].

そして，ICAに基づくブラインド音源分離が，適応ビームフォーマ（Adaptive Beamformer: ABF）とよばれるマイクロホンアレーと同じ動作原理を実現していることを明らかにした [6]．2マイクのABFの支配的な動作は妨害音に一つの死角を向ける動作である．これにより，ブラインド音源分離の性能改善の糸口が明らかとなった．統計的な手法と，音響信号処理的手法との長所を上手く関連づけることで，新しい分離技術を得ることが可能となったのである．

ここでは，独立成分分析とは何か，ブラインド音源分離とは何か，どのようにして分離が達成されるのか，分離のメカニズムはどのようなものか，などについて，できるだけ直感的に分かりやすく論じる．

2. ブラインド音源分離

ブラインド音源分離は，観測された混合音声 $x_j(n)$ のみを用いて，音源信号 $s_i(n)$ を推定する手法である．いくつかのマイクロホンで収音した混合音声や，複数センサで収録した脳波，無線基地局の複数のアンテナに到達する複数の無線信号の分離などが代表的な応用例である．

2.1 音声信号の混合モデル

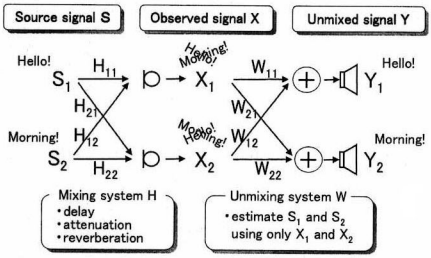
音声信号を分離する場合，音源信号が異なる時間差とレベル差で混合するように，いくつかのマイクロホンを空間の異なる位置に配置する．音源信号が音声であり混合系が部屋である実環境では，マイクロホンで収音された混合音声は残響の影響を受ける．したがって，M 個のマイクロホンで収音された N 個の混合音声は

$$x_j(n)=\sum_{i=1}^{N}\sum_{p=1}^{P}h_{ji}(p)s_i(n-p+1)\,(j=1,\cdots,M)\quad(1)$$

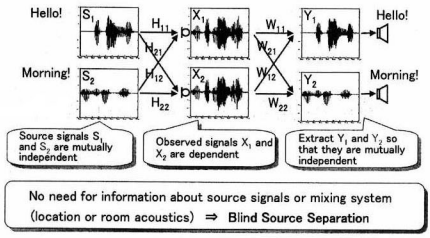
とモデル化できる．ここで，s_i は音源 i からの音源信号，x_j はマイクロホン j で収音された混合音声，h_{ji} は

* NTT コミュニケーション科学基礎研究所

Key Words: blind source separation, independent component analysis, unsupervised adaptive filtering, adaptive beamformer, microphone array.



第1図 ブラインド音源分離システム



第2図 ブラインド音源分離の課題

音源 i からマイクロホン j への P タップのインパルス応答である．本稿では，非ガウス，非定常，非白色，ゼロ平均である音声信号を音源とする場合について論じる．

2.2 分離モデル

分離フィルタとは，Q タップのフィルタ $w_{ij}(q)$ により，分離信号

$$y_i(n)=\sum_{j=1}^{M}\sum_{q=1}^{Q}w_{ij}(q)x_j(n-q+1)\,(i=1,\cdots,N) \quad (2)$$

を推定するものである．分離フィルタは，分離信号が統計的に互いに独立になるように求める．本稿では，一般性を損なうことなく，2入力2出力の問題，つまり $N=M=2$（第1図）の場合を取扱う．

2.3 音源分離の課題

音源信号 s_1, s_2 は互いに独立であると仮定する．この仮定は，実環境の音源信号については，通常，成り立つ．混合音声を収音するマイクロホンを2本用いると，観測信号 x_1, x_2 には相関がある．この相関のある観測信号を入力として分離フィルタ w_{ij} を推定し，互いに独立な出力 y_1, y_2 を分離・抽出することが目標である．ここで，独立成分分析を用いて出力を互いに独立とする分離フィルタを逐次的に学習する．この操作により，音源信号 s_1, s_2 の推定値が分離信号 y_1, y_2 として得られる．音源位置の知識や妨害音区間の切り出し，さらに，混合系 h_{ji} の情報を原理的に必要としない．そのため，ブラインド音源分離とよばれている（第2図）．

2.4 瞬時混合と畳込み混合

2.4.1 畳込み混合

部屋の中で音を分離する場合には，混合系 h_{ji} は P が数千タップに及ぶFIRフィルタになる．この問題は畳込み混合の問題とよばれ，大変難しい問題であり，比較的新しい課題である．

2.4.2 瞬時混合

これに対して，混合系 h_{ji} で P が1である場合，すなわち，遅延や残響がなく，たとえば，ミキサーを使って音をミキシングしたような場合には，瞬時混合の問題とよばれる．

実際，画像，functional MRI や EEG などの医療信号などのアプリケーションでは，ほとんどが瞬時混合の問題である．瞬時混合の問題は畳込み混合の問題より簡単であり，検討も多数なされ，良い成果も多数得られている．

2.5 時間領域手法と周波数領域手法

畳込み混合の問題を解くために，いくつかの方法が提案されている．時間領域手法は，混合系のインパルス応答 h_{ji} を FIR フィルタで表し，分離フィルタを時間領域で推定する [7–9]．また，周波数領域手法は，時間領域の畳込み混合を，周波数領域の複数の瞬時混合に変換して解く [10–14]．

2.6 畳込み混合に対する時間領域手法

畳込み混合の時間領域BSSには，Multichannel Blind Deconvolution と Convolutive BSS の二つを明確に区別する必要がある [7]．

まず，Multichannel Blind Deconvolution では，源信号はチャネル間ばかりでなくチャネル内でも独立であると仮定する．そして，分離信号がチャネル間で互いに，かつ，チャネル内でも独立になるようにデコンボリューションする．これに対して，Convolutive BSS は，分離信号を互いに独立にするのみであり，デコンボリューションは行わない．音声信号には自己相関があるため，音声信号を分離する場合には Convolutive BSS が適している．もし音声信号に Multichannel Blind Deconvolution を適用すれば，スペクトルの等価，周波数特性の平坦化，白色化が起こってしまう．したがって，元の音声の周波数特性を保つために，プレフィルタリングやポストフィルタリングなどを施す必要が生じる [7,8]．

時間領域BSSの欠点は，演算量が多いこと，収束が遅いこと，初期値の影響を強く受けることである [8,15]．この問題は取扱うタップ長 P が大きい場合にとくに顕著である．

2.7 畳込み混合に対する周波数領域手法

周波数領域BSSでは，時間領域の畳込み混合を，周波数領域の複数の瞬時混合に変換して解く [10–14]．

(1)式に T-ポイント短時間フーリエ変換を施せば，周波数領域の時間系列信号が得られる．

$$\mathbf{X}(\omega,m)=\mathbf{H}(\omega)\mathbf{S}(\omega,m) \tag{3}$$

ここで，ω は周波数，m は時間，$\mathbf{S}(\omega,m)=[S_1(\omega,m), S_2(\omega,m)]^{\mathrm{T}}$ は音源信号ベクトル，$\mathbf{X}(\omega,m)=[X_1(\omega,m), X_2(\omega,m)]^{\mathrm{T}}$ は観測信号ベクトルである．(2×2) 混合行列 $\mathbf{H}(\omega)$ は逆行列を持ち，$H_{ji}(\omega)\neq 0$ と仮定する．さらに $\mathbf{H}(\omega)$ は時間 m に依らないと仮定する．

分離フィルタは各周波数 ω で

$$\mathbf{Y}(\omega,m)=\mathbf{W}(\omega)\mathbf{X}(\omega,m) \tag{4}$$

と表わされる．ここで，$\mathbf{Y}(\omega,m)=[Y_1(\omega,m),Y_2(\omega,m)]^{\mathrm{T}}$ は分離信号ベクトル，$\mathbf{W}(\omega)$ は周波数 ω における (2×2) 分離行列である．分離信号 $Y_1(\omega,m)$, $Y_2(\omega,m)$ が互いに独立になるように，分離行列 $\mathbf{W}(\omega)$ を求める．この計算は各周波数でそれぞれ行われる．ここでは，短時間フーリエ変換のフレーム長 T と分離フィルタのタップ長 Q は等しいものとする．

本章以降では，とくに断らない限り，畳込み混合の問題を周波数領域で論じる．ここで議論される周波数領域における議論は，時間領域における畳込み混合の問題に対しても，本質的に成り立つ．

2.8　スケーリングとパーミュテーション

周波数領域 BSS には，各周波数が個別に取り扱われるために生じるパーミュテーションの問題がある．これは，分離信号の各周波数成分がそれぞれの周波数で別々の順番で現れるという問題である．分離信号の各周波数成分が揃わなければ全体としての分離は達成できないため，周波数領域 BSS でパーミュテーションは非常に重要な問題となる．これに対して，時間領域 BSS では分離信号の順番が入れ替わるだけの問題である．

周波数領域 BSS のパーミュテーションの解法として，音源の方向情報と分離信号の相関を利用した方法が提案され，ほぼ完璧なパーミュテーション解決ができるようになった [14].

ブラインド音源分離では，スケーリングの問題も大きな問題である．分離信号の各周波数成分はそれぞれの周波数において別々のゲインで現れる．各周波数におけるスケーリングの任意性は，分離信号の畳込みの任意性，すなわち，フィルタリングの任意性となって現れる．このことは，独立な信号をフィルタリングしたものもまた独立であるという事実を反映している．

スケーリングの解法として，Minimal Distortion Principle に基づく方法が提案され，マイクロホン位置における音源信号まで回復できるようになった [9].

3.　独立成分分析

独立成分分析は統計的な手法で，もともとニューラルネットや無線通信の分野で提案され [16]，統計理論，情報理論をベースに発展し，さらに，さまざまなアプリケーション領域において，近年脚光を浴びている．

ICA の理論には，信号どうしの統計的独立性という，統計理論においてもっとも一般的な特徴が利用されている．ブラインド音源分離の問題において，音源信号は「独立成分」として扱われる．簡単に言えば，ICA は，観測信号 x_j のみから，分離信号 y_i が互いに独立となるような線形な分離行列 $\mathbf{W}(\omega)$ と分離信号 y_i の両方を推定する手法である．一つの成分は他の成分に何の情報も与えず，分離信号が互いに独立になったとき，音源信号が抽出される．

3.1　独立の概念

「独立」という概念は「無相関」の概念より強い．すなわち，相関が2次の統計量に基づくものであるのに対して，独立は高次の統計量に基づく．簡単に言えば，独立とは，片方の信号がもう一方の信号に関する情報を持っていないということである．独立な成分は，非線形相関除去，非定常相関除去，非白色相関除去により求めることができる．

もし，分離行列 $\mathbf{W}(\omega)$ が正しく，分離信号 y_1, y_2 が独立で，ゼロ平均であり，非線形関数 $\Phi(\cdot)$ が奇関数で，$\Phi(y_1)$ がゼロ平均であれば，

$$E[\Phi(y_1)y_2]=E[\Phi(y_1)]E[y_2]=0 \tag{5}$$

が成り立つ．非線形相関除去の手法では，(5) 式を満足するような分離行列 $\mathbf{W}(\omega)$ を求めていく．では，非線形関数 $\Phi(\cdot)$ はどのように選べばよいか？

この疑問には，独立成分分析のいくつかの理論から答えることができる．これらの理論のうちのどれを用いても，適切な非線形関数 $\Phi(\cdot)$ を選ぶことができる．これらは，相互情報量の最小化，非ガウス性の最大化，ゆう度の最大化，の三つに基づく．

なお，非定常相関除去と非白色相関除去の手法については，**4** 章を参照されたい．

3.2　相互情報量の最小化

独立成分分析の一つ目の理論は，相互情報量の最小化に基づく．相互情報量は，二つの信号間の統計的独立性を測るための，情報理論に基づく自然な規範である．相互情報量は常に非負であり，統計的に独立なときにのみ0になる．したがって，分離信号間の相互情報量を最小化することによって，独立な音源信号成分を推定しようとすることは自然なことといえる．分離信号間の相互情報量の最小化は，分離信号間の独立性の最大化を意味する．

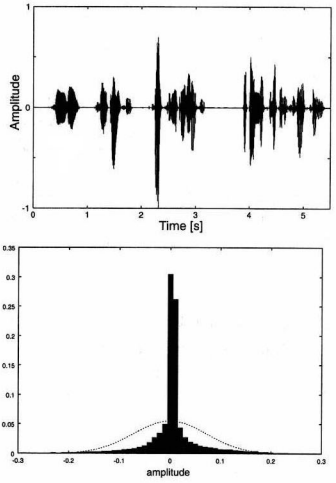
3.3　非ガウス性の最大化

独立成分分析の二つ目の理論は，非ガウス性の最大化に基づく．統計理論の中心極限定理によれば，独立な成

分の和の確率密度関数はガウス分布に近づく．独立な成分が混合した信号の確率密度関数は，元の信号の確率密度関数より，ガウス分布に近い．したがって，分離信号の非ガウス性を最大化することによって，独立な成分，すなわち元の音源信号を，分離・抽出することができる．従来の多くの統計理論においては，音源信号の確率密度関数としてガウス分布を仮定することが普通であった．これに対して，独立成分分析の理論においては，音源信号の確率密度関数として非ガウスの分布を仮定することがポイントである．

音声信号をはじめとして，多くの実世界の信号は，スーパーガウシアン分布を有する．スーパーガウシアン分布は，尖った確率密度関数を有する．すなわち，ガウス分布に比べて0である確率が高い．

音声信号の確率密度関数の一例を第3図に示す．



第3図 音声信号の一例とその確率密度関数．点線はガウス分布

3.4 ゆう度の最大化

独立成分分析の三つ目の理論は，ゆう度の最大化に基づく．最ゆう推定は，学習理論で用いられることの多い統計的手法であり，ゆう度を最大化するような未知の確率分布を求める．独立成分分析では，混合系も音源信号の確率分布も未知であるため，音源信号の確率分布を仮定したうえで，観測信号のゆう度が最大になる分離フィルタを求める．

最ゆう推定は，ニューラルネットの規範である情報量最大化原理（infomax）に密接に関係している．インフォマックスは，非線形出力を有するニューラルネットの出力エントロピーあるいは情報フローの最大化に基づいている．入力 x_i と出力 y_i の間の相互情報量を最大化するのである．

3.5 三つの解は同一

面白いことに，上記三つの解は同一である [17]．分離信号 y_1, y_2 間の相互情報量 $I(y_1,y_2)$ は

$$I(y_1,y_2)=\sum_{i=1}^{2}H(y_i)-H(y_1,y_2) \tag{6}$$

と表される．ここで，$H(y_i)$ は分離信号 y_i の marginal エントロピー，$H(y_1,y_2)$ は joint エントロピーである．分離信号 y のエントロピー $H(y)$ は y の確率密度関数 $p(y)$ を用いて

$$H(y)=E[\log\frac{1}{p(y)}]=\sum p(y)\log\frac{1}{p(y)} \tag{7}$$

と表される．

3.2 節で説明した相互情報量 $I(y_1,y_2)$ の最小化は，(6) 式第一項の最小化あるいは第二項の最大化によって達成される．ガウス信号は第一項を最大化するため，**3.3** 節で説明した非ガウス性の最大化により，(6) 式第一項の最小化が達成される．一方，**3.4** 節で説明した分離信号の joint エントロピー最大化により，(6) 式第二項の最大化が達成される．以上のように，三つの手法は等価である．これらの理論から，非線形関数は，音源信号の確率密度関数の対数を微分したものに対応させることが良いことが明らかにされている．これらの理論の詳細は，参考文献 [15,18–20] を参照されたい．

3.6 学習則

まず，初期状態にある分離行列 $\mathbf{W}(\omega)$ を用いて (4) 式の操作により分離信号 Y_i, Y_j を求める．つぎに，分離行列 $\mathbf{W}(\omega)$ を変化させ，分離信号 Y_i, Y_j 間の相互情報量を最小化する，非ガウス性を最大化する，あるいは，ゆう度を最大化する分離行列 $\mathbf{W}(\omega)$ を求める．この更新を繰り返す，いわゆる学習を経て，システムは互いに独立な分離信号を生成する．この操作は，勾配法により実現できる [21]．

以降，源信号 (音声信号) の確率密度関数は既知である，すなわち，音源信号のスーパーガウシアン分布は既知であると仮定する．さらに，非線形関数は適切に与えられている，すなわち，音源信号の確率密度関数の対数を微分したものに対応していると仮定する．つまり，非線形関数には音声信号に適切な $\tanh(\cdot)$ を用いる．

4. ブラインド音源分離のメカニズム

本稿では，音響信号処理の観点から，ブラインド音源分離のメカニズムを直感的に分かりやすく論じる．独立成分分析に基づくブラインド音源分離のフレームワークで，どのようにして分離が達成できるのか?

最も簡潔な答えは，「$\mathbf{R}_Y$ を対角化することによって分離する」である．$\mathbf{R}_Y$ は次の (2×2) の行列である．

$$\mathbf{R}_Y = \begin{bmatrix} \langle \Phi(Y_1)Y_1 \rangle & \langle \Phi(Y_1)Y_2 \rangle \\ \langle \Phi(Y_2)Y_1 \rangle & \langle \Phi(Y_2)Y_2 \rangle \end{bmatrix} \tag{8}$$

ここで，関数 $\Phi(\cdot)$ は非線形関数，$\langle\cdot\rangle$ は統計量を取り出すための平均操作である．非対角成分を最小化し，同時に，対角成分を適切な値に保つことにより，分離が達成できる．

行列 $\mathbf{R}_Y$ の成分は Y_i, Y_j 間の相互情報量に対応する．収束後には，Y_1, Y_2 間の相互情報量を表わす非対角成分は最小化され，0 に近づく．

$$\langle \Phi(Y_1)Y_2 \rangle = 0, \quad \langle \Phi(Y_2)Y_1 \rangle = 0 \tag{9}$$

これと同時に，分離信号 Y_1 および Y_2 の大きさを表わす対角成分は適切な値に保たれる．

$$\langle \Phi(Y_1)Y_1 \rangle = c_1, \quad \langle \Phi(Y_2)Y_2 \rangle = c_2 \tag{10}$$

解を収束させるために，逐次修正式を用いる．

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta \Delta \mathbf{W}_i, \tag{11}$$

$$\Delta \mathbf{W}_i = \begin{bmatrix} c_1 - \langle \Phi(Y_1)Y_1 \rangle & \langle \Phi(Y_1)Y_2 \rangle \\ \langle \Phi(Y_2)Y_1 \rangle & c_2 - \langle \Phi(Y_2)Y_2 \rangle \end{bmatrix} \mathbf{W}_i \tag{12}$$

$\mathbf{R}_Y$ が対角化されたときには，$\Delta\mathbf{W}$ は 0 に収束する．

$c_1 = c_2 = 1$ の場合には，Holonomic 型アルゴリズムとよばれ，$c_1 = \langle \Phi(Y_1)Y_1 \rangle$, $c_2 = \langle \Phi(Y_2)Y_2 \rangle$ の場合には，Nonholonomic 型アルゴリズムとよばれる．

4.1 2 次統計量に基づく手法と高次統計量に基づく手法

もし $\Phi(Y_1) = Y_1$ の場合には，非対角項は単純な相互相関除去になる．

$$\langle \Phi(Y_1)Y_2 \rangle = \langle Y_1 Y_2 \rangle = 0 \tag{13}$$

この条件だけでは独立にはならない．しかし，音源信号が非定常信号である場合には，このような式を複数の時間ブロックに対して成り立たせる共通の分離行列 $\mathbf{W}(\omega)$ を求めることによって，問題を解くことができる．これが「非定常相関除去」の手法である [22,23].

また，音源信号が非白色信号である場合には，時間遅れを伴なった相互相関除去

$$\langle \Phi(Y_1)Y_2 \rangle = \langle Y_1(m)Y_2(m+\tau_i) \rangle = 0 \tag{14}$$

を複数の遅延時間 τ_i に対して成り立たせる共通の分離行列 $\mathbf{W}(\omega)$ を求めることによって，問題を解くことができる．これが「非白色相関除去」の手法である [24]. これらの二つは「2 次統計量に基づく手法 (Second Order Statistics: SOS)」とよばれる．

一方，たとえば $\Phi(Y_1) = \tanh(Y_1)$ の場合には，非対角項は

$$\langle \Phi(Y_1)Y_2 \rangle = \langle \tanh(Y_1)Y_2 \rangle = 0 \tag{15}$$

となる．$\tanh(Y_1)$ をテイラー展開すれば，(15) 式は

$$\langle (Y_1 - \frac{Y_1^3}{3} + \frac{2Y_1^5}{15} - \frac{17Y_1^7}{315} \cdots)\, Y_2 \rangle = 0 \tag{16}$$

となり，2 次ばかりでなく高次の相関除去，あるいは，「非線形相関除去」が現れる．(16) 式を満たすような解を求めることによって，問題を解くことができる．これが高次統計量に基づく手法 (Higher Order Statistics: HOS) である．

4.2 2 次統計量に基づく手法

2 次統計量 (SOS) に基づく手法は，2 次統計量と音源信号の非定常/非白色性を利用している．すなわち，追加情報として音源信号の非定常/非白色性を用いたクロストークの最小化である．Weinstein らは，音源信号の非定常性を用いれば分離行列 $\mathbf{W}(\omega)$ を求めることが可能であることを指摘し，非定常相関除去に基づく手法を提案した [23].

簡単に言えば，各周波数において，W_{ij} の未知数が 4 つであるので対して，$\Phi(Y_i) = Y_i$ の場合には $Y_1Y_2 = Y_2Y_1$ であるから，(9), (10) 式には三つの式しか存在しない．すなわち，連立方程式は不定であり，そのため，連立方程式は解けない．

しかし，音源信号が非定常である場合には，それぞれの時間ブロックで 2 次統計量が異なる．同様に，音源信号が非白色である場合には，それぞれの遅延時間に対して 2 次統計量が異なる．その結果，より多くの式が利用可能となり，連立方程式を解くことが可能となる．

非定常相関除去に基づく手法では，音源信号 $S_1(\omega,m)$, $S_2(\omega,m)$ はゼロ平均，無相関であると仮定する．分離信号 $Y_1(\omega,m)$, $Y_2(\omega,m)$ が互いに独立となる分離行列 $\mathbf{W}(\omega)$ を求めるために，すべての時間ブロック k に対して共分散行列 $\mathbf{R}_Y(\omega,k)$ を同時対角化する $\mathbf{W}(\omega)$ を求める．

$$\begin{aligned} \mathbf{R}_Y(\omega,k) &= \mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^{\mathrm{H}}(\omega) \\ &= \mathbf{W}(\omega)\mathbf{H}(\omega)\mathbf{\Lambda}_s(\omega,k)\mathbf{H}^{\mathrm{H}}(\omega)\mathbf{W}^{\mathrm{H}}(\omega) \\ &= \mathbf{\Lambda}_c(\omega,k) \end{aligned} \tag{17}$$

ここで $^{\mathrm{H}}$ は共役転置を表わし，$\mathbf{R}_X$ は $\mathbf{X}(\omega)$ の共分散行列

$$\mathbf{R}_X(\omega,k) = \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{X}(\omega, Mk+m)\mathbf{X}^{\mathrm{H}}(\omega, Mk+m) \tag{18}$$

である．$\mathbf{\Lambda}_s(\omega,k)$ は音源信号の共分散行列であり，時間ブロック k に対して異なる対角行列である．$\mathbf{\Lambda}_c(\omega,k)$ は任意の対角行列である．

$\mathbf{R}_Y(\omega,k)$ の対角化は最小 2 乗法で解ける．

$$\arg\min_{W(\omega)} \sum_k ||\mathrm{diag}\{\mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^{\mathrm{H}}(\omega)\}$$

$$-\mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^{\mathrm{H}}(\omega)||^2$$

$$s.t.,\ \sum_k ||\mathrm{diag}\{\mathbf{W}(\omega)\mathbf{R}_X(\omega,k)\mathbf{W}^{\mathrm{H}}(\omega)\}||^2 \neq 0 \quad (19)$$

ここで，$||\mathbf{x}||$ はフロベニウスノルム，$\mathrm{diag}\mathbf{A}$ は行列 $\mathbf{A}$ の対角成分である．解は勾配法を用いて求めることが出来る．

非白色相関除去に基づく手法では，$\mathbf{R}_X$ は

$$\mathbf{R}_X(\omega,\tau_i)=\frac{1}{M}\sum_{m=0}^{M-1}\mathbf{X}(\omega,m)\mathbf{X}^{\mathrm{H}}(\omega,m+\tau_i) \quad (20)$$

と定義される．すべての遅延時間 τ_i に対して共分散行列 $\mathbf{R}_Y(\omega,\tau_i)$ を同時対角化する $\mathbf{W}(\omega)$ を求める．

4.3 高次統計量に基づく手法

高次統計量に基づく手法（HOS）は，音源信号の非ガウス性を利用している．もっと簡単に言えば，それぞれの周波数で，四つの未知数 W_{ij} に対して，(9), (10) の中に式が四つある．その結果，連立方程式は解ける．分離行列 $\mathbf{W}(\omega)$ を求めるために，Kullback-Leibler divergence の最小化に基づくアルゴリズムが提案されている [10,11]．安定で収束速度の速いアルゴリズムとして，ナチュラルグラジェントに基づくアルゴリズムが Amari らによって提案された [25]．ナチュラルグラジェントを用いれば，最適な分離行列 $\mathbf{W}(\omega)$ は逐次勾配法によって

$$\mathbf{W}_{i+1}(\omega)=\mathbf{W}_i(\omega)+\eta\left[\mathrm{diag}\left(\langle\Phi(\mathbf{Y})\mathbf{Y}^{\mathrm{H}}\rangle\right)-\langle\Phi(\mathbf{Y})\mathbf{Y}^{\mathrm{H}}\rangle\right]\mathbf{W}_i(\omega) \quad (21)$$

と表される．ここで，$\mathbf{Y}=\mathbf{Y}(\omega,m)$，$\langle\cdot\rangle$ は平均操作，i は繰り返しの i-番目，η はステップサイズを表す．

さらに，複素数の信号に対する非線形関数 $\Phi(\cdot)$ を

$$\Phi(\mathbf{Y})=\tanh(\mathbf{Y}^{(\mathrm{R})})+j\ \tanh(\mathbf{Y}^{(\mathrm{I})}) \quad (22)$$

と表わす．ここで，$\mathbf{Y}^{(\mathrm{R})}$, $\mathbf{Y}^{(\mathrm{I})}$ は，それぞれ $\mathbf{Y}$ の実数部と虚数部である [10]．

複素信号に対する非線形関数として，極座標表示に基づく

$$\Phi(\mathbf{Y})=\tanh(\mathrm{abs}(\mathbf{Y}))\ e^{j\arg(\mathbf{Y})} \quad (23)$$

が，直交座標表示に基づく非線形関数 (22) より適していることが，理論的，実験的に示されている [26]．

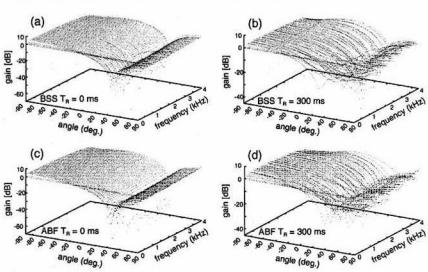
5. ブラインド音源分離の音響信号処理からの解釈

独立成分分析 (ICA) に基づくブラインド音源分離 (BSS) は統計的，あるいは，数学的な手法であり，動作のメカニズムはよく分っていなかった．単に分離信号 Y_1, Y_2 を互いに独立にしているだけである．それでは，独立成分分析に基づくブラインド音源分離の音響信号処理的解釈は何であろうか？

ブラインド音源分離の動作のメカニズムは，古くからよく知られている2組の適応ビームフォーマ（ABF）と等価である [6]．マイクロホンが2本の場合，適応ビームフォーマは妨害音方向に適応的な空間的死角を一つ形成し，目的音を抽出する．ブラインド音源分離も適応ビームフォーマと同様に，妨害音方向に適応的な死角を一つ形成し，目的音を抽出する．

ブラインド音源分離と適応ビームフォーマの動作の様子を比較してみよう．第4図はブラインド音源分離と適応ビームフォーマによって得られた分離フィルタの指向性パターンである．第4図 (a), (b) はブラインド音源分離によって得られた分離フィルタの指向性パターン，第4図 (c), (d) は適応ビームフォーマによって得られた分離フィルタの指向性パターンである．残響時間 $T_{\mathrm{R}}=0$ の場合には，第4図 (a), (c) に示すようにブラインド音源分離と適応ビームフォーマ共に，鋭く深い空間的指向性パターンが得られている．これに対して，残響時間 $T_{\mathrm{R}}=300$[ms] の場合には，第4図 (b), (d) に示すようブラインド音源分離と適応ビームフォーマ共に，幅が広く底の浅い空間的指向性パターンが得られている．

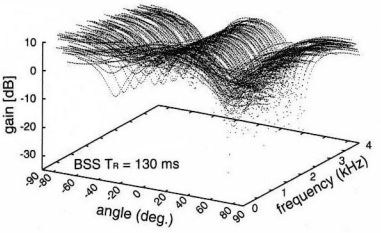
第5図は三つの音源を3本のマイクロホンを用いて分離した場合の分離フィルタの指向性パターンである．二つの妨害音方向に適応的な空間的死角を形成し，中央の目的音を抽出している様子が見て取れる．



第4図 指向性パターン (a) obtained by BSS (T_{R}=0 ms), (b) obtained by BSS (T_{R}=300 ms), (c) obtained by ABF (T_{R}=0 ms), (d) obtained by ABF (T_{R}=300 ms)

ブラインド音源分離も適応ビームフォーマも，妨害音に指向特性の死角を形成する．すなわち，妨害音の方向に空間的なノッチを作って感度を下げ，目的音を取り出すメカニズムであることが理解できる．

適応ビームフォーマもブラインド音源分離も，妨害音方向に空間的死角を形成して目的音を取出すメカニズムであるため，残響がある場合の分離性能の低下は避けられない [4]．この解釈により，残響が長い実際の室内においてブラインド音源分離の性能が悪い理由が直感的に理



第5図　3音源分離時の指向性パターン BSS (T_R=130 ms)

解できる．もし，音源信号の独立性の仮定が成り立たない場合は，分離フィルタを求める際にバイアスノイズが発生する．そのため，正確に拘束条件を与えられた適応ビームフォーマの性能が，ブラインド音源分離の性能の上限となる．

しかしながら，適応ビームフォーマと違って，ブラインド音源分離には，マイクロホンの位置や音源の情報などは不要である．適応ビームフォーマでは，目的音の位置情報を拘束条件としながら，目的音が無く妨害音のみが鳴っている時間を検出して，その時だけ出力誤差に対する2乗誤差最小化の規範により適応動作を行う．これに対して，ブラインド音源分離では，分離信号間の相関除去の規範により適応動作を行うため，目的音の位置情報や妨害音のみが鳴っている時間の検出が不要である．

適応ビームフォーマにおいて，出力誤差に対する2乗誤差最小化の規範は，妨害音のみが鳴っている時間の検出誤りに非常に影響される．これに対して，ブラインド音源分離では，音源信号は同時に鳴っていても全く問題ない．さらに，適応ビームフォーマではマイクロホンアレーのマイク配置に関する幾何学的情報と目的音方向の情報が必要である．以上のように考えれば，ブラインド音源分離は適応ビームフォーマの高機能版と言える．

なお，本稿では，二つの音源を2本のマイクロホンで分離する場合を例に説明してきたが，最近では，動き回る3人の話者を3本のマイクロホンでリアルタイムに追跡しながら実時間で分離したり [14]，六つの音源を8本のマイクロホンで分離したり [27]，四つの音源を2本のマイクロホンで分離する [28] 検討も行われている．

6. あとがき

音響信号，特に音声信号の畳込み混合を対象としたブラインド音源分離について述べた．非線形相関除去，非定常相関除去，非白色相関除去を用いることにより，観測した混合音声のみを使って，音源信号を分離・抽出することができる．統計的手法である独立成分分析を音響信号処理の観点から論じた．

ブラインド音源分離は2組の適応ビームフォーマと同じ動作原理を有している．畳込み混合のブラインド音源分離は，互いに独立な分離信号を取り出す，あるいはもっと簡単に言えば，クロストークを最小化する複数の適応ビームフォーマとして解釈できる．

本稿が，独立成分分析に基づくブラインド音源分離「開眼」の一助となれば幸いである．

謝　辞

日頃ご討論頂く猿渡洋博士に謝意を表する．

(2004年3月22日受付)

参考文献

[1] S. Makino: Blind source separation of convolutive mixtures of speech; *Adaptive Signal Processing: Applications to Real-World Problems* (J. Benesty and Y. Huang, Eds.), Springer (2003)

[2] S. Makino, S. Araki, R. Mukai and H. Sawada: Audio source separation based on independent component analysis; *Proc. ISCAS*, pp. V-668–V-671 (2004)

[3] J. F. Cardoso: The three easy routes to independent component analysis; contrasts and geometry; *Proc. ICA*, pp. 1–6 (2001)

[4] S. Araki, R. Mukai, S. Makino, T. Nishikawa and H. Saruwatari: The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech; *IEEE Trans. Speech Audio Processing*, Vol. 11, No. 2, pp. 109–116 (2003)

[5] R. Mukai, S. Araki, H. Sawada and S. Makino: Evaluation of separation and dereverberation performance in frequency domain blind source separation; *J. Acoust. Sci. & Tech.*, Vol. 25, No. 2, pp. 119–126 (2004)

[6] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa and H. Saruwatari: Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures; *EURASIP Journal on Applied Signal Processing*, Vol. 2003, No. 11, pp. 1157–1166 (2003)

[7] X. Sun and S. Douglas: A natural gradient convolutive blind source separation algorithm for speech mixtures; *Proc. ICA*, pp. 59–64 (2001)

[8] R. Aichner, S. Araki, S. Makino, T. Nishikawa and H. Saruwatari: Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming; *Proc. NNSP*, pp. 445–454 (2002)

[9] K. Matsuoka and S. Nakashima: Minimal distortion principle for blind source separation; *Proc. ICA*, pp. 722–727 (2001)

[10] P. Smaragdis: Blind separation of convolved mixtures in the frequency domain; *Neurocomputing*, Vol. 22, pp. 21–34 (1998)

[11] S. Ikeda and N. Murata: A method of ICA in time-frequency domain; in *Proc. ICA*, pp. 365–370 (1999)

[12] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura: Evaluation of blind signal separation

method using directivity pattern under reverberant conditions; *Proc. ICASSP*, pp. 3140–3143 (2000)

[13] F. Asano, S. Ikeda, M. Ogawa, H. Asoh and N. Kitawaki: A combined approach of array processing and independent component analysis for blind separation of acoustic signals; *Proc. ICASSP*, pp. 2729–2732 (2001)

[14] H. Sawada, R. Mukai, S. Araki and S. Makino: Convolutive blind source separation for more than two sources in the frequency domain; *Proc. ICASSP*, pp. III-885–III-888 (2004)

[15] T. W. Lee: *Independent Component Analysis - Theory and Applications*, Kluwer (1998)

[16] J. Herault and C. Jutten: Space or time adaptive signal processing by neural network models; *Neural Networks for Computing: AIP Conference Proceedings 151* (J. S. Denker, Ed.), American Institute of Physics (1986)

[17] T. W. Lee, M. Girolami, A. J. Bell and T. J. Sejnowski: A unifying information-theoretic framework for independent component analysis; *Computers and Mathematics with Applications*, Vol. 31, No. 11, pp. 1–12 (2000)

[18] A. Hyvärinen, J. Karhunen and E. Oja: *Independent Component Analysis*, John Wiley & Sons (2001)

[19] S. Haykin: *Unsupervised Adaptive Filtering*, John Wiley & Sons (2000)

[20] A. Cichocki and S. Amari: *Adaptive Blind Signal and Image Processing*, John Wiley & Sons (2002)

[21] A. J. Bell and T. J. Sejnowski: An information-maximization approach to blind separation and blind deconvolution; *Neural Computation*, Vol. 7, No. 6, pp. 1129–1159 (1995)

[22] K. Matsuoka, M. Ohya and M. Kawamoto: A neural net for blind separation of nonstationary signals; *Neural Networks*, Vol. 8, No. 3, pp. 411–419 (1995)

[23] E. Weinstein, M. Feder and A. V. Oppenheim: Multichannel signal separation by decorrelation; *IEEE Trans. Speech Audio Processing*, Vol. 1, No. 4, pp. 405–413 (1993)

[24] L. Molgedey and H. G. Schuster: Separation of a mixure of independent signals using time delayed correlations; *Physical Review Letters*, Vol. 72, No. 23, pp. 3634–3636 (1994)

[25] S. Amari, A. Cichocki and H. Yang: A new learning algorithm for blind source separation; *Advances in Neural Information Processing Systems 8*, pp. 757–763, MIT Press (1996)

[26] H. Sawada, R. Mukai, S. Araki and S. Makino: Polar coordinate based nonlinear function for frequency-domain blind source separation; *IEICE Trans. Fundamentals*, Vol. E86-A, No. 3, pp. 590–596, Mar. (2003)

[27] R. Mukai, H. Sawada, S. Araki and S. Makino: Near-field frequency domain blind source separation for convolutive mixtures; *Proc. ICASSP*, pp. IV-49–IV-52 (2004)

[28] S. Araki, S. Makino, A. Blin, R. Mukai and H. Sawada: Underdetermined blind separation for speech in real environments with sparseness and ICA; *Proc. ICASSP*, pp. III-881–III-884 (2004)

著者略歴

牧野 昭二

1956 年 6 月 4 日生．1981 年東北大学大学院工学研究科機械工学専攻修士課程修了．同年日本電信電話公社 (現 NTT) 入社．2003 年 NTT コミュニケーション科学基礎研究所メディア情報研究部長となり現在に至る．音響エコーキャンセラ，ブラインド音源分離などの音響信号処理の研究に従事．1993 年工学博士 (東北大学)．IEEE Fellow．日本音響学会，電子情報通信学会会員．

荒木 章子

1975 年 4 月 11 日生．2000 年東京大学大学院工学系研究科計数工学専攻修士課程修了．同年日本電信電話株式会社 NTT コミュニケーション科学基礎研究所，現在に至る．音響信号処理の研究，とくに音源分離の研究に従事．2001 年日本音響学会粟屋潔学術奨励賞，2003 年 IWAENC Best Paper Award，2004 年電気通信普及財団テレコムシステム技術賞受賞．IEEE，日本音響学会会員．

向井 良

1968 年 3 月 19 日生．1992 年東京大学大学院理学系研究科修士課程修了．同年 NTT 入社．交換ノード用プロセッサ，分散処理システムの研究開発に従事，2000 年より，音響信号処理，音源分離の研究に従事し，現在に至る．IEEE，ACM，日本音響学会，電子情報通信学会，情報処理学会などの会員．

澤田 宏

1968 年 10 月 31 日生．1991 年京都大学工学部情報工学科卒業．1993 年同修士課程修了．同年日本電信電話株式会社 (NTT) 入社．以来，同社コミュニケーション科学基礎研究所にて，VLSI 向け CAD およびアーキテクチャの研究に従事．2000 年より，信号処理，特に独立成分分析を用いたブラインド音源分離の研究に従事．2001 年京都大学博士 (情報学)．2000 年 IEEE Circuits and Systems Society Best Paper Award．電子情報通信学会，日本音響学会，IEEE 会員．