独立成分分析に基づくブラインド音源分離
Audio Source Separation based on Independent Component Analysis

牧野 昭二† 荒木 章子† 向井 良† 澤田 宏†
Shoji MAKINO Shoko ARAKI Ryo MUKAI Hiroshi SAWADA

† NTT コミュニケーション科学基礎研究所
〒 619–0237 京都府相楽郡精華町光台 2–4
{maki, shoko, ryo, sawada}@cslab.kecl.ntt.co.jp

あらまし

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．独立成分分析に基づくブラインド音源分離は，ある人が話している声と別の人の声，背景に流れる音楽，雑音等，それぞれの音は互いに統計的に独立であるという仮定により，複数のマイクで観測した信号を互いに独立な信号に分離すれば，それぞれの元の音を復元できる，という原理に基づいている．この手法は，音源や混合系の情報を原理的に必要としない，いわゆるブラインドな分離が可能である．本稿では，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りやすく論じる [1].

abstract

This paper introduces the blind source separation (BSS) of convolutive mixtures of acoustic signals, especially speech. A statistical and computational technique, called independent component analysis (ICA), is examined. By achieving nonlinear decorrelation, nonstationary decorrelation, or nonwhite decorrelation, we can find source signals only from observed mixed signals. Particular attention is paid to the physical interpretation of BSS from the acoustical signal processing point of view. Frequency-domain BSS is shown to be equivalent to two sets of frequency domain adaptive microphone arrays, i.e., adaptive beamformers (ABFs). Although BSS can reduce reverberant sounds to some extent in the same way as ABF, it mainly removes the sounds from the jammer direction. This is why BSS has difficulties with long reverberation in the real world. If sources are not “independent,” the dependence results in bias noise when obtaining the correct unmixing filter coefficients. Therefore, the performance of BSS is limited by that of ABF. Although BSS is upper bounded by ABF, BSS has a strong advantage over ABF. BSS can be regarded as an intelligent version of ABF in the sense that it can adapt without any information on the array manifold or the target direction, and sources can be simultaneously active in BSS.

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，一人が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，目的の人以外の声，背景に流れる音楽，周囲騒音，残響のような邪魔な音があると認識率は急激に低下し，そのような状況ではコンピュータに私たちの話した声を認識させることはできない．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．一方，人間は騒々しいカクテルパーティでも，目的とする音声を聞き取り会話することができる．これは，カクテルパーティ効果として，良く知られている人間特有の能力である．

複数人の遠隔発話を認識する場合には，残響と混合が問題となり，このような状況でも聞きたい音声をコンピューターで認識するために，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．具体的には，会議室などの実環境において，複数話者が同時に発話することが想定される状況下での音源分離技術の実現を目標としている．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，分離出力が互いに独立となる

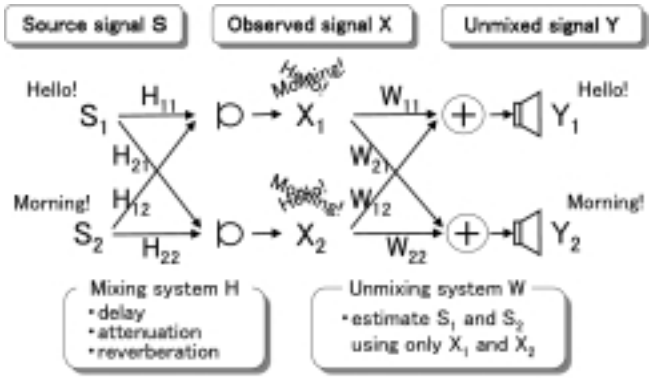

図 1: ブラインド音源分離システム

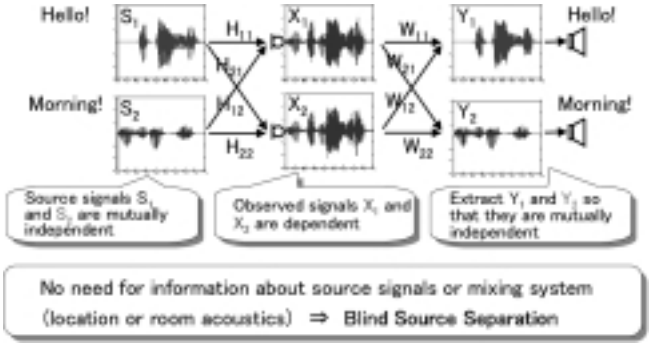

図 2: ブラインド音源分離の課題

ようなフィルタを求める手法である．この手法は，音源位置の知識や目的音 (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等の仮定も用いない，完全なブラインド分離が可能である．

本稿では ICA を広く捉える，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 3 つの手法を統一的に論じる [2]．現実の世界でブラインド音源分離のアプリケーションは多い [3]．しかし，分離性能はまだ十分とは言えない [4], [5]．

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 [6], [7], [8]，従来の音響信号処理技術との関係を解明した．

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) と呼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．2 マイクの ABF の支配的な動作は妨害音に 1 つの死角を向ける動作である．これより，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．我々は統計的な手法と，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を目指している．

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りやすく論じる [1]．

2 ブラインド音源分離

ブラインド音源分離 (Blind Source Separation: BSS) 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 [9], [10]．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

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

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

分離システムは，Q タップの分離フィルタ $w_{ij}(k)$ を推定し，分離信号

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

を得る．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 $N=M=2$ (図 1) 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．観測信号 x_1, x_2 には相関がある．分離システム 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．この操作により，音源信号 s_1, s_2 が分離出力 y_1, y_2 に得られる．音源位置の知識や目的音 (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呼ばれている (図 2).

分離システム w_{ij} には，スケーリングとパーミュテーションの任意性が残る．なお，残響除去やデコンボリューションは行なわない [11].

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呼ばれ，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} が定数である場合，すなわち，遅延や残響がなく，例えば，ミキサーを使って音をミキシングしたような場合には，瞬時混合の問題と呼ばれる．

実際，画像, CDMA 無線信号, functional MRI や EEG などの医療信号などのアプリケーションは，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 フィルタで表わし，分離フィルタを時間領域で推定する [12], [13], [14]．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 [15], [16].

2.6 畳込み混合に対する時間領域手法

畳込み混合を時間領域で分離するためには，分離システム w_{ij} は FIR フィルタでも IIR フィルタでも構わない．しかし，非最小位相系を実現するために，通常，FIR フィルタが用いられる [12].

畳込み混合の時間領域 BSS には，Multichannel Blind Deconvolution と Convolutive BSS の 2 つを明確に区別する必要がある [14].

Multichannel Blind Deconvolution は，分離出力が互いに，そして，自分自身に対しても独立になるように動作する．源信号は，互いに，そして，自分自身に対しても独立であると仮定する．すなわち，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．これに対して，Convolutive BSS は，分離出力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なわない．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，分離出力に不適切な作用が発生し，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 [14], [17].

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 [12], [17]．この問題は取扱うタップ長 P が大きい場合に特に顕著である．

時間領域 BSS の利点は，パーミュテーションの問題，すなわち，分離信号の出現順の任意性の問題を回避できる点にある (2.8 節参照).

2.7 畳込み混合に対する周波数領域手法

Smaragdis[15] は，非線形関数を複素数に拡張し，周波数領域 BSS を提案した．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 [15], [16], [18], [19].

(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，$\mathbf{S}(\omega, m) = [S_1(\omega, m), S_2(\omega, m)]^T$ は音源信号ベクトル，$\mathbf{X}(\omega, m) = [X_1(\omega, m), X_2(\omega, m)]^T$ 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．

分離システムは各周波数 ω で

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

と表わされる．ここで，$\mathbf{Y}(\omega, m) = [Y_1(\omega, m), Y_2(\omega, m)]^T$ 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なわれる．ここでは，短時間フーリエ変換のフレーム長 T と分離フィルタ長 Q は，等しいものとする．

本章以降では，特に断らない限り，畳込み混合の問題を周波数領域で論じる．ディジタル信号処理は，周

波数領域と時間領域で本質的には変わらず，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 bin が個別に取り扱われるというパーミュテーションの問題がある．その結果，分離信号の各周波数成分はそれぞれの周波数 bin で別々の順番で現れる．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 では分離出力が逆にな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出力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 [20], [21].

周波数領域 BSS 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 bin で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．

周波数領域 BSS の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 [22], [23].

3 独立成分分析

独立成分分析 (Independent Component Analysis: ICA) は統計的手法で，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て来た [24], [25], [26], [27], [28], [29], [30], [31]．ICA は信号処理，ニューラルネット，統計理論，情報理論，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．この手法はブラインド音源分離に盛んに使われており，音，画像，CDMA 無線信号， functional MRI や EEG などの医療信号などを始めとして，たくさんの応用例がある．

ICA の理論には，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 の両方を推定する問題である．分離行列 $\mathbf{W}(\omega)$ は，それぞれの分離信号がそれぞれの音源信号の情報をできるだけ多く含むように決定される．ひと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，相関は 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．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，

$$E[\Phi(y_1)y_2] = E[\Phi(y_1)]E[y_2] = 0 \qquad (5)$$

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良いか？

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 3 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，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．相互情報量は常に否負であり，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と言える．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分布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分布は，元の信号の確率密度分布より，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分布としてガウス分布を仮定することが多かった．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分布として非ガウスの分布を仮定することに注意されたい．

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，尖った確率密度分布を有する．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分布 (pdf) の一例を (図 3) に示す．

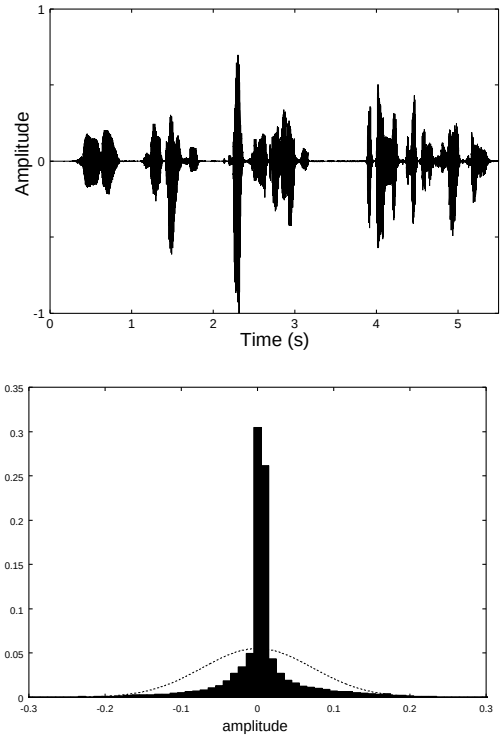

図 3: 音声信号の一例とその確率密度分布 (pdf)．点線はガウス分布．

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，統計推定の基本原理であり，独立成分分析においても大変良く使われる手法である．最も確からしい分離信号を得るように，最ゆう推定のパラメータを求める．

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．入出力間の相互情報量の最大化は，分離信号のエントロピー最大化と等価であり，インフォマックスは最ゆう推定法と等価である．

3.5 3 つの解は同一

面白いことに，上記 3 つの解は同一である [32]．分離信号 y_1, y_2 間の相互情報量 $I(y_1, y_2)$ は

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

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 (pdf) $p(y)$ を用いて

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

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，3 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 (pdf) 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 [12], [33], [34], [35] を参照されたい．

3.6 学習則

分離を達成するためには， (4) 式の分離行列 $\mathbf{W}(\omega)$ を変化させ，分離信号の確率密度分布がどのように変化するかを観測すれば良い．分離信号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．この操作は勾配法を用いて達成できる．

Bell は簡潔な勾配法を導いた [36]．甘利はナチュラルグラジェントを提案し，安定性と収束速度を改善した [37]．このアルゴリズムは通常の相関除去を非線形相関除去に拡張したものであり，実際，非線形相関除去アルゴリズムの一種である．

以降，源信号 (音声信号) の確率密度関数 (pdf) 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 (pdf) 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 (2×2) の行列である．

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

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，分離出力 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 型アルゴリズムと呼ばれ， $c_1 = \langle \Phi(Y_1)Y_1 \rangle$, $c_2 = \langle \Phi(Y_2)Y_2 \rangle$ の場合には，Nonholonomic 型アルゴリズムと呼ばれる．

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つようにすることによって，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 [38].

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を複数の遅延時間に対して成り立つようにすることによって，問題を解くことができる．

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

これが,「非白色相関除去」の手法である [39], [40]. これらの 2 つが,「2 次統計量に基づく手法 (Second Order Statistics: SOS)」である．

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

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

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 [11]. 2 次統計量 (SOS) に基づく手法を用いた音源分離の試みは [4], [41] でもなされている．

簡単に言えば，各周波数において， W_{ij} の未知数が 4 つであるのに対して，$\Phi(Y_i) = Y_i$ の場合には $Y_1Y_2 = Y_2Y_1$ であるから， (9), (10) 式には 3 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．

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

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

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

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．

$$\begin{aligned} &\arg\min_{W(\omega)} \sum_k ||\mathrm{diag}\{\mathbf{W}(\omega)\mathbf{R}_X(\omega, k)\mathbf{W}^H(\omega)\} \\ &\qquad -\mathbf{W}(\omega)\mathbf{R}_X(\omega, k)\mathbf{W}^H(\omega)||^2 \\ &s.t., \ \sum_k ||\mathrm{diag}\{\mathbf{W}(\omega)\mathbf{R}_X(\omega, k)\mathbf{W}^H(\omega)\}||^2 \neq 0 \end{aligned} \tag{19}$$

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

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

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 (Higher Order Statistics: 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，4 つの未知数 W_{ij} に対して，(9), (10) の中に式が 4 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 [15], [16]. 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甘利によって提案された [42]．ナチュラルグラジェントを用いれば，最適な分離行列 $\mathbf{W}(\omega)$ は次のような逐次勾配法によって

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

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わす．

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 [15].

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 [43], [44].

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，数学的は手法であり，動作のメカニズムは良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？

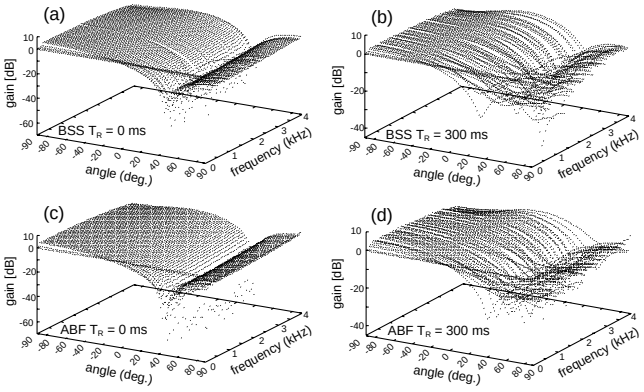

図 4: 指向性パターン (a) obtained by BSS (T_{R}=0 ms), (b) obtained by BSS (T_{R}=300 ms), (c) obtained by ABF (T_{R}=0 ms), and (d) obtained by ABF (T_{R}=300 ms).

ブラインド音源分離の動作のメカニズムは，2 組の適応ビームフォーマ (ABF) である [45]．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空間的死角を 1 つ適応的に形成し，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，妨害音方向に死角を 1 つ適応的に形成し，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．図 4 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システムの指向性パターンである．図 4 において， (a), (b) はブラインド音源分離によって得られた分離システム $\mathbf{W}$ の指向性パターン， (c), (d) は適応ビームフォーマによって得られた分離システム $\mathbf{W}$ の指向性パターンである．残響時間 $T_{\mathrm{R}}=0$ の場合には，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 [図 4(a), (c)]．これに対して，残響時間 $T_{\mathrm{R}}=300$ ms の場合には，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 [図 4(b), (d)].

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 [46]．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，分離システムを求める際にバイアスノイズとなる．そのため，ブラインド音源分離の性能にとって適応ビームフォーマの性能が上限となる．

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位置情報などは不要である．ブラインド音源分離では，目的音の方向の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では，目的音が無く妨害音のみが鳴っている時間を検出して，その時だけパワー最小化の規範により適応動作を行なう．これに対して，ブラインド音源分離では，目的音と妨害音の相関を除去するクロスパワー最小化の規範により適応動作を行なう．適応ビームフォーマにおける出力誤差に対する 2 乗誤差最小規範とブラインド音源分離における分離出力間の相関除去規範は等価である．誤差の最小化は相関のゼロサーチと等価である．

ブラインド音源分離の性能は適応ビームフォーマの性能を超えることはできない．しかし，ブラインド音源分離には適応ビームフォーマよりも優れている点がある．適応ビームフォーマにおいて，1 チャネルのパワー最小化規範は，クロストークあるいはダブルトークに非常に弱い．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幾何学的情報と目的音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っても過言ではない．

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，互いに独立な分離出力を取り出す，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．

参考文献

[1] S. Makino, “Blind Source Separation of Convolutive Mixtures of Speech,” in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds., Springer, Berlin, Jan. 2003.

[2] J. F. Cardoso, “The three easy routes to independent component analysis; contrasts and geometry,” in *Proc. ICA*, Dec. 2001, pp. 1–6.

[3] T. W. Lee, A. J. Bell, and R. Orglmeister, “Blind source separation of real world signals,” *Neural Networks*, vol. 4, pp. 2129–2134, 1997.

[4] M. Z. Ikram and D. R. Morgan, “Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment,” in *Proc. ICASSP*, June 2000, pp. 1041–1044.

[5] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, “Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech,” in *Proc. ICASSP*, May 2001, vol. 5, pp. 2737–2740.

[6] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, “Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers,” in *Proc. Eurospeech*, Sept. 2001, pp. 2595–2598.

[7] R. Mukai, S. Araki, and S. Makino, “Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment,” in *Proc. Eurospeech*, Sept. 2001, pp. 2599–2602.

[8] R. Mukai, S. Araki, H. Sawada, and S. Makino, “Separation and dereverberation performance evaluation of frequency domain blind source separation,” *Acoust. Sci. & Tech.*, accepted.

[9] S. C. Douglas, “Blind separation of acoustic signals,” in *Microphone Arrays: Techniques and Applications*, M. Brandstein and D. B. Ward, Eds., pp. 355–380, Springer, Berlin, 2001.

[10] K. Torkkola, “Blind separation of delayed and convolved sources,” in *Unsupervised Adaptive Filtering, Vol. I*, S. Haykin, Ed., pp. 321–375, John Wiley & Sons, 2000.

[11] E. Weinstein, M. Feder, and A. V. Oppenheim, “Multi-channel signal separation by decorrelation,” *IEEE Trans. Speech Audio Processing*, vol. 1, no. 4, pp. 405–413, Oct. 1993.

[12] T. W. Lee, *Independent Component Analysis - Theory and Applications*, Kluwer, 1998.

[13] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, “Real world blind separation of convolved non-stationary signals,” in *Proc. ICA*, Jan. 1999, pp. 347–352.

[14] X. Sun and S. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," in *Proc. ICA*, Dec. 2001, pp. 59–64.

[15] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[16] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. ICA*, Jan. 1999, pp. 365–370.

[17] 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," in *Proc. NNSP*, Sept. 2002, pp. 445–454.

[18] J. Anemüeller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA*, June 2000, pp. 215–220.

[19] 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," in *Proc. ICASSP*, May 2001, vol. 5, pp. 2729–2732.

[20] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust approach to the permutation problem of frequency-domain blind source separation," in *Proc. ICASSP*, Apr. 2003, pp. 381–384.

[21] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," in *Proc. ICA*, Apr. 2003, pp. 505–510.

[22] K. Matsuoka and S. Nakashima, " Minimal distortion principle for blind source separation," in *Proc. ICA*, Dec. 2001, pp. 722–727.

[23] K. Matsuoka and S. Nakashima, " A robust algorithm for blind separation of convolutive mixture of sources," in *Proc. ICA*, Apr. 2003, pp. 927–932.

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

[25] C. Jutten and J. Herault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[26] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources, part II: problems statement," *Signal Processing*, vol. 24, pp. 11–20, 1991.

[27] E. Sorouchyari, "Blind separation of sources, part III: stability analysis," *Signal Processing*, vol. 24, pp. 21–29, 1991.

[28] A. Cichocki and L. Moszczynski, "A new learning algorithm for blind separation of sources," *Electronics Letters*, vol. 28, no. 21, pp. 1986–1987, 1992.

[29] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[30] P. Comon, "Independent component analysis - a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[31] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Trans. Circuits and Systems*, vol. 43, no. 11, pp. 894–906, 1996.

[32] 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, Mar. 2000.

[33] A. Hyvärinen, H. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[34] S. Haykin, *Unsupervised Adaptive Filtering*, John Wiley & Sons, 2000.

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

[36] 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.

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

[38] 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.

[39] 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.

[40] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[41] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[42] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, 1998.

[43] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," in *Proc. ICASSP*, May 2002, vol. 1, pp. 1001–1004.

[44] 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.

[45] 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*, accepted.

[46] 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, Mar. 2003.