ブラインドな処理が可能な音源分離技術

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ます．ここでは，音源や音の混ざり方（音源位置，部屋の残響など）の情報を原理的に必要としない，すなわちブラインドな処理が可能な，独立成分分析に基づく音源分離技術を紹介します．

牧野 昭二（まきの しょうじ）/荒木 章子（あらき しょうこ）
向井 良（むかい りょう）/澤田 宏（さわだ ひろし）
NTTコミュニケーション科学基礎研究所

聞き分け可能なコンピュータをつくる

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ます．しかしその一方で，さまざまな背景音，例えば周囲の人の声，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します．私たち人間が無意識に行っている「聞きたい音を聞き分ける」能力がコンピュータには欠けているのです．

NTTコミュニケーション科学基礎研究所では，「コミュニケーション環境理解」というテーマを掲げ，人間のように物を見たり，音や声を聞いたり，人と話をすることができるコンピュータを実現することにより，生活を豊かにすることを目指して研究を進めています．特に，音声による高度で自然な人との情報交換機能を持つコンピュータの実現を目指し，音声認識システムの適用領域をこれまでの接話・単一話者の場合から，さまざまな実環境コミュニケーションシーンで有効なマイクから離れた発話・複数人話者での対話へ広げるための研究を推進しています．

マイク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雑音との混合や残響の影響が問題となります．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ります．これが音源分離の目標です．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す．

音源分離技術の確立を目指して，NTTコミュニケーション科学基礎研究所では，独立成分分析（Independent Component Analysis: ICA）を用いる分離手法の研究を行っています．例えば，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す．同様に，実環境におけるほとんどすべての音源は互いに独立であるとみなすことができます．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・復元します．

独立成分分析に基づくブラインド音源分離

独立成分分析に基づくブラインド音源分離（Blind Source Separation: BSS）の原理を図１を用いて説明します．音源信号は互いに独立であると仮定します．この仮定は，実環境の音源信号については通常成り立ちます．２音を分離する場合，混合音声を収音するマイクを２本用います．２本のマイクで収音された観測信号には相関があります．この相関のある観測信号を入力として分離フィルタを推定し，互いに独立な出力を分離・抽出することが目標です．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します．この操作により，音源信号の推定値が分離出力に得られます．音源位置の知識や雑音区間の切り出し，さらに音の混ざり方の情報を原理的に必要としません．そのため，ブラインド音源分離と呼ばれています．

独立成分分析

独立成分分析は統計的な手法で，

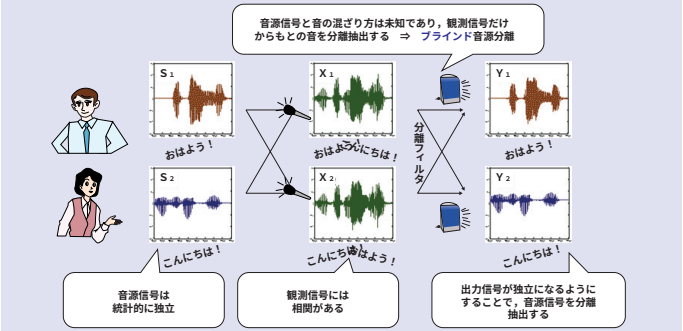

図1　独立成分分析（ICA）に基づくブラインド音源分離（BSS）の原理

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，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ます．

独立成分分析の理論には，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ます．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」として扱われます．簡単にいえば，独立成分分析は観測信号のみから，分離信号が互いに独立となるような線形な分離フィルタと分離信号の両方を推定する手法です．1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ます．

独立の概念

「独立」という概念は「無相関」の概念より強い概念です．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きます．簡単にいえば，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す．

独立成分分析には3つの理論があります．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*1の最大化，の3つに基づきます．面白いことに，上記3つの解は同一です．

(1)　相互情報量の最小化

独立成分分析の1つ目の理論は，相互情報量の最小化に基づきます．相互情報量は，2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す．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になります．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ます．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します．

(2)　非ガウス性の最大化

独立成分分析の2つ目の理論は，非ガウス性の最大化に基づきます．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*2に近づきます．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くなります．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ます．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した．これに対して，独立成分分析の理

＊1　ゆう度：観測された事象がある確率分布から生起するということが，どのくらいもっともらしいかを表す量．
＊2　ガウス分布：確率変数の分布曲線がガウス関数であるような分布．正規分布．

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す．

音声信号は，尖った確率密度関数を持っています．すなわち，ガウス分布に比べて０である確率が高い分布です．音声信号の一例とその確率密度関数を図２，３に示します．

⑶　ゆう度の最大化

独立成分分析の３つ目の理論は，ゆう度の最大化に基づきます．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，一般的にはゆう度を最大化するような未知の確率分布を求めます．

独立成分分析では，混合系も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の係数を求めます．

学習則

分離フィルタは，まず初期状態にある分離フィルタを用いて分離音を求めます．次に，分離フィルタを変化させ，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フィルタを求めます．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音を生成します．この操作は，勾配法＊３により実現できます．

ブラインド音源分離のメカニズム

独立成分分析に基づくブラインド音源分離は統計的，あるいは数学的な手法であり，物理的な動作のメカニズムはよく分かっていませんでした．単に分離信号を互いに独立にしているだけです．では，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でしょうか？

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（Adaptive Beamformer: ABF）と等価です（図４）．マイクが２本の場合，適応ビームフォーマは雑音方向に適応的な空間的死角を１つ形成し，目的音を抽出します．ブラインド音源分離も適応ビームフォーマと同様に，雑音方向に適応的な死角を１つ形成し，目的音を抽出します[1]．

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ましょう．図５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す．図５(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，図５(c)，(d)は適応ビーム

＊３　勾配法：あるコスト関数を最小（最大）化するパラメータを求める逐次的手法の１つ．各時点でのコスト関数の傾き（勾配）の絶対値が最大である方向へパラメータを少しずつ変化させます．

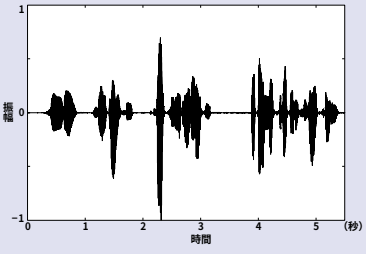

図２　音声信号の一例

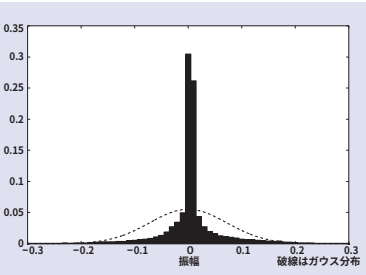

図３　音声信号の確率密度関数

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す．残響時間0 msの場合には，図5(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ともに，鋭く深い空間的指向性パターンが得られています．これに対して，残響時間300 msの場合には，図5(b)，(d)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ともに，幅が広く底の浅い空間的指向性パターンが得られています．

ブラインド音源分離も適応ビームフォーマも，雑音に適応的な指向特性の死角を形成する，すなわち雑音の方向に適応的な空間的ノッチ（notch）をつくって感度を下げ，目的音を取り出すメカニズムであることが理解できます．

ブラインド音源分離は適応ビームフォーマの高機能版

適応ビームフォーマもブラインド音

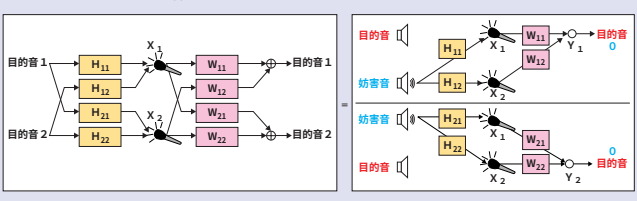

図4　ブラインド音源分離（BSS）と適応ビームフォーマ（ABF）の等価性

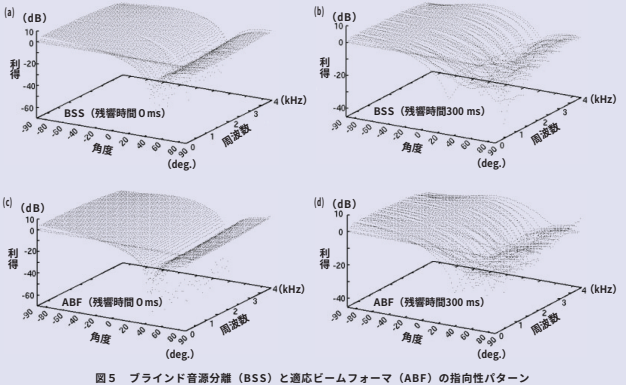

図5　ブラインド音源分離（BSS）と適応ビームフォーマ（ABF）の指向性パターン

源分離も，雑音方向に適応的な空間的死角を形成して目的音を取り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ません．

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，マイクの位置や音源の位置情報などは不要です．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，目的音がなく雑音のみが鳴っている時間を検出して，そのとき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います．これに対して，ブラインド音源分離では，分離出力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雑音のみが鳴っている時間の検出が不要です．

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，雑音のみが鳴っている時間の検出誤りに非常に影響されます．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ありません．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す．このように考えれば，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いえるでしょう．

動き回る３人の話者のリアルタイム分離にも成功

NTTコミュニケーション科学基礎研究所では，これまでの研究により統計的手法である独立成分分析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，従来の音響信号処理技術との関係を解明しました(1)．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りました．統計的な手法と，音響信号処理的手法との長所をうまく関連づけることで，新しい分離技術を得る努力を重ねています．

さらに，実際の部屋に音源分離システムを構築し，３人の話者が同時に発声した声をその場で録音，バッチ処理により分離フィルタを計算，分離音を再生することに成功しました（図６）．さらに，動き回る３人の話者をリアルタイムで追跡しながら分離・再生する実時間音源分離にも成功しています．

図６　３人の話者を用いた音源分離

今後の展開

今後は音声認識システムとの統合を進め，これまで接話・単一話者に限定されていた音声認識システムの適用領域をマイクから離れた複数人話者での対話へ広げ，ヒューマノイドロボットの耳や会議の議事録を自動編集するシステムなどへの応用を目指します．

■参考文献
(1) 牧野・向井・荒木・片桐：“混じりあった声を解く―遠隔発話の認識を目指して，”NTT R&D，Vol.50，No.12，pp.937-944，2001.

（左から）牧野　昭二/ 荒木　章子/
向井　　良/ 澤田　　宏

人間に学び，人間を超えたヒューマノイドロボットを実現することが目標です．

◆問い合わせ先
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
TEL　046-240-5210
FAX　046-270-2359
E-mail　maki@cslab.kecl.ntt.co.jp