雑音除去　事後確率最大化基準　独立成分分析

Technology Reports

マイク・アレイを用いた高速雑音除去技術

移動端末に入力される音声に対する高速雑音除去手法として，独立成分分析に基づくマイク・アレイを用いた手法を提案する．提案手法の有効性を実際のモバイル環境を再現した評価実験により確認した．

張　志鵬　栄藤　稔

1. まえがき

音声雑音除去技術は，移動端末における通話品質の向上および音声認識・音声翻訳などの音声ユーザインタフェースの重要な技術である．背景雑音を含む信号から音声を強調する技術研究が広く行われている．従来から雑音スペクトルサブトラクション[1]と呼ばれるマイクを1つ用いて背景雑音を抑圧する技術が知られている．この技術は有声区間と無声区間を区別し，無声区間の入力信号を背景雑音とし雑音スペクトルを生成する．次に音声と雑音が混在した有声区間の入力信号から，雑音スペクトルを減算することで雑音を抑圧している．したがって，雑音が定常信号である場合は，良好な雑音除去性能を示し，かつ実現が容易であるため現在広く用いられている．しかし，にぎやかなレストランや車の行き交う雑踏の中では背景雑音が非定常信号となり，良好な雑音除去性能が得られない．このため，複数のマイクを用いた手法（マイク・アレイ）の研究も行われている．マイク・アレイは音源からの信号が各マイクに到達するときに生じる位相差などの空間情報を利用し雑音を抑圧するため，信号の定常性を前提とする1 マイクの手法に比べて雑音抑圧性能が優れている．このマイク・アレイに基づく手法にはビームフォーミング法[2]と独立成分分析（ICA： Independent Component Analysis）を用いたブラインド音源分離（BSS : Blind Source Separation）法[3]がある．2つの手法の比較を**表1**に示す．

移動端末には搭載スペース，計算能力の制限があるため数多くのマイクを搭載するのは現時点では難しいが，マイクの数を限定した小規模なマイク・アレイの実用化は可能となっている．ビームフォーミング法は歴史が長く，すでに，適応型ビームフォーミングと非線形信号処理を組合せ，ハンズフリー通話用のPDA端末が商品化されている．しかし，原理的にビームフォーミング法は，抽出したい音源が他の音源と異なる位置にあることを前提とするため，音源の位置が変動する場合や，雑音と目的音源が同じ方向にある場合は性能が低下する場合がある．

「異なる空間にある音源を分離する」というビームフォーミング法に対して，ICAに基づくBSS法は「独立した統計的性質をもつ音源を分離する」という信号の独立性を利用するため，原理的に位置情報を必要としない．すなわち，雑音と同じ方向でも目的音源を抽出することができる利点があり，より広い適応範囲を有している．しかし一方で，ICAは

表1　マイク・アレイに基づく手法の比較

	ビームフォーミング法	ICAに基づくBSS法
メリット	すでに商用化され，実績がある	移動音源への自動追随が可能
デメリット	移動音源への自動追随に性能が低下	分離行列の推定に計算がかかる 商用化には，ハードウェアとマイクにかかるコストが問題

統計的性質の逐次学習（正確には非ガウス性[*1]の最大化）を必要とする．この最適化問題の解釈には，非線形であり反復を伴う逐次処理が必要であり，そのためICAは実時間処理には不向きであった．演算処理の高速化を実現することにより，リアルタイムの実環境下での分離性能を大幅に向上させたモジュールも開発されたが，専用ハードウェアと指向性マイクが必要なため実現コストが商用化の障害となり，課題となっている．

本稿では，考案した少ないマイク数で高品質な音声信号を得るため新しいICAに基づくBSS法を提案する．本提案手法は，ユーザの口元から移動端末のマイクまでの伝達関数[*2]のパラメータが所定の範囲に収まることを利用し，このパラメータの事後確率最大化（MAP：Maximum A Posterior probability）を用いたものである．これは，パラメータを推定する手法の一種で，音声データに基づきパラメータの事後確率を最大にするようにパラメータを推定する方法である．また，本手法は従来のICAに比べ収束が早く，より高品質な音声を抽出できる．

以下にICAを用いた雑音除去，伝達関数を活用したICA，評価実験について述べる．

2. ICAを用いた雑音除去

前章にて，目的信号を分離抽出する手法の1つとしてICAに基づくBSS法があることを述べた．これは，複数の線形混合された信号を，元の信号や混合過程についての知識を全く用いることなしに推定する手法である．BSS法には，信号時系列をそのまま扱う実空間領域のICAに基づくものと，信号を周波数領域へ変換し各周波数において分離行列を求めるICAに基づくものの2種類がある．提案手法では，伝達関数の扱いの平易さから周波数領域のICAを用いる．また，環境が発声（目的音源）と雑音（干渉音源）という2つの音源からなると仮定し，目標とするシステムを単純化する．これによりマイク数を2つとし，計算の複雑さと実現コストの低減を狙う．

2.1 実環境での混合信号（観測信号）モデル

移動端末2マイクICAシステムにおける混合信号・分離モデルを**図1**に示す．図1に示すようにsは音源信号を表す．s_1は目的音源からのユーザの発声である．s_2は干渉音源からの雑音を表す．2つのマイクで観測される信号y_1とy_2は音源信号が伝達経路を経て収録されたものである．信号源からマイクまでの伝達関数をAとすると，$y=As$で信号源，伝達関数と観測データの線形関係を表す．ただしAは混合行列で，Aの各要素が目的および干渉それぞれの音源から2つのマイクまでの伝達状況を表す．ここで干渉音源と目的音源は統計的に独立であると仮定する．

2.2 分離信号のモデル

BSS法では，信号の独立性を利用し，混合信号yから元の信号源を復元する．分離行列Wを求め，分離信号zを以下の式(1)で求める．

$$z=Wy \tag{1}$$

信号源からマイクまでのすべての伝達関数が分かれば，$W=A^{-1}$でWを計算することにより，zを音源信号に復元できるが，そもそも伝達関数Aの詳細は事前には未知である．また音源が移動する場合，伝達関数が変わるため何らかの手段で分離行列Wの変動を追随しなければならない．そこで信号源に関する独立性

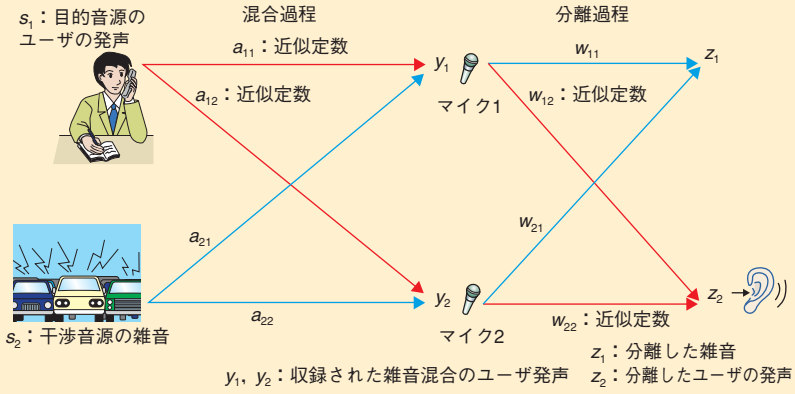

図1　移動端末の2マイクICAシステムにおける混合信号・分離モデル

＊1　**非ガウス性**：確率分布においてガウス分布を示さない確率変数の性質．独立した信号において最大となる．
＊2　**伝達関数**：ある伝達システムの信号伝達に関する出力信号のラプラス変換と，入力信号のラプラス変換の比．

の仮定のもとで，独立性が最大となるよう分離行列Wを推定し，元の信号源を再現するのがBSS法である．

2.3 分離行列の推定

分離行列の推定によく利用される尤度最大化基準[*3]に基づく手法[4][5]について解説する．$p(y)$は観測信号yの確率分布関数を表す場合，Wを変数とする尤度は$p\{y(t)/W\}$で示される．実際には尤度の対数のほうが計算上の便利からより用いられる．対数尤度最大基準での推定は以下の式(2)で表す．

$$\hat{W}=\arg\max\sum_{t=1}^{T}\log p\{y(t)/W\} \quad (2)$$

一般的にこの式より$\hat{W}$を解析的に解くことはできないため，式(3)のように逐次適応法による繰返しの計算が必要になる．

$$W_{i+1}=W_i+\eta\Delta W \quad (3)$$

勾配法[*4]により

$$\Delta W=\frac{\partial \log p\{y(t)/W\}}{\partial W}=(I-E[\phi(y)y^T])W \quad (4)$$

ただし，Iは単位行列，$E[X]$はXの期待値，$\phi(y)$はyの確率分布関数の微分を表す．最終的にWの更新式は以下になる．

$$W_{i+1}=W_i+\eta\{I-E[\phi(y)y^T]\}W_i \quad (5)$$

このWを基に各周波数領域で独立成分分離が達成され，周波数の分離信号を時間領域に再変換することにより，時間領域での独立信号を得る．

3. 伝達関数を活用したICA

従来の尤度最大化基準に基づく推定法は，局所最適値が複数あり，繰返しが必要となる一般的な非線形最適化の手法の1つであり，何らかの工夫を施さない限り，結果は初期値に依存し反復数も不定かつ多くなる可能性がある．特にこれらの課題は，雑音が非定常な場合，追従性の顕著な問題となって現れる．この問題を解決するために，ユーザの口元からマイクまでの伝達関数パラメータがある範囲に収まるという事前知識を利用し，高速かつ安定な最適化手法を提案する．

提案手法のフローを**図2**に示す．提案手法の大きな特徴は初期化と反復推定の二段階の計算を行うところにある．まず，最初の段階ではユーザの発話位置とマイクの位置関係から伝達関数パラメータの初期値を求める．次に求めた初期値を利用し，分離行列を推定する．通話環境においてはユーザの口元はマイクとの相対位置が一定の範囲以内に収まると考えられる．このため，ユーザの口元からマイクまでの伝達関数パラメータを利用する．一方で，背景雑音については未知パラメータとなるが発話最初の雑音区間のデータを利用し，推定できる．事前にパラメータの分布が，すべてではなくとも一部でも既知であれば，パラメータ推定は，尤度最大推定ではなくMAP推定で定式化でき，より少ない繰返しの回数で分離行列の推定を精度よく安定して推定できることが期待される．ユーザの口元はマイクとの相対位置が一定の範囲以内に収まるため，ここでは混合行列の一部（a_{11}とa_{12}）をほぼ一定とみなすことができ，分離行列の対応部分（w_{12}とw_{22}）も同様にほぼ一定とみなすことができる．

3.1 初期値の推定

まず，ユーザの口元からマイクまでの周波数応答を測定する．この測定から以下の手順で分離行列の目的音源の伝達に関する要素の初期値を推定する．

① w_{11}の初期値の設定

ユーザが発声してない区間を検出する．ユーザの発声してない区間の観測データ（y_1とy_2）と以下の式(6)を用いて，出力z_1のエネルギーが最小化となるようにw_{11}を求める．

$$w_{11}=\arg\min\ R[z_1,\bar{z}_1] \quad (6)$$

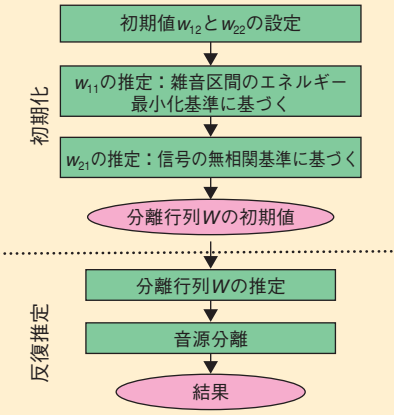

図2　処理手順

＊3 **尤度最大化基準**：ある確率論的モデルを仮定しているときに，その観測データが得られる確率が最大となる基準．

＊4 **勾配法**：関数の勾配（ベクトル場の微分）を利用して，数値最適化を図る手法の一種である．

ただし，$R[x, y]$はxとyの相関を表す[5]．

w_{11}を式(5)，(6)により以下のように推定する．

$$w_{11}=a_1(R_{22}-R_{21})/(R_{11}-R_{12}) \quad (7)$$

ここで，

$R_{ij}=R[y_i, y_j]$．

② w_{21}の初期値の設定

z_1とz_2が無相関の条件に基づく．

$$R[z_1, z_2]=0 \quad (8)$$

この基準でw_{21}の初期値を求める．

$$w_{21}=-\frac{w_{22}w_{11}E(y_1y_2)+w_{12}w_{22}E(y_2y_2)}{w_{11}E(y_1y_1)+w_{12}E(y_1y_2)} \quad (9)$$

3.2 分離行列の推定

一般的にパラメータに関する事前知識が分かる場合はMAP基準で推定を行うほうが有効と考えられる．分離行列の事後確率$p(W/y)$はWの事前確率$p(W)$と尤度$p(y/W)$の積で表す．

$$p(W/y)=p(W)p(y/W) \quad (10)$$

この式から分かるようにWに関する事前情報がなく，$p(W)$は一様分布と仮定した場合はMAP基準と尤度最大化基準は同じになる．Wに関する事前確率$p(W)$が分かれば，より正確な推定はできる．MAP基準に基づくWの推定式は以下になる．

$$\hat{W}=\arg\max\sum_{t=1}^{T}\log\left[p(W)p\{y(t)/W\}\right] \quad (11)$$

ここではWに関する事前確率$p(W)$を正規分布と仮定し，密度関数$p(W)$は以下のようになる．期待値μは3.1節で求めたWの初期値とする．分散値σ^2は分離行列の事前推定値の変動を表す．

$$p(W)=\frac{1}{\sigma\sqrt{2\pi}}\exp\{\frac{-(W-\mu)^2}{2\sigma^2}\} \quad (12)$$

MAP基準に基づいて，勾配法によりWを推定する場合は，

$$\Delta W=\frac{\partial\log[\,p(W)\,p\{y(t)/W\}\,]}{\partial W}=\frac{\partial\log p(W)}{\partial W}+\frac{\partial\log[\,p\{y(t)/W\}\,]}{\partial W} \quad (13)$$

式(12)の第二項は尤度最大推定と同じ，第一項は以下のようになる．

$$\frac{\partial\log p(W)}{\partial W}=(W-\mu)/\sigma^2 \quad (14)$$

このため，

$$\Delta W=\eta\{I-\phi(y)y^T\}W+(W-\mu)/\sigma^2 \quad (15)$$

更新式は，

$$W_{i+1}=W_i+\eta\{I-\phi(y)y^T\}W_i+(W_i-\mu)/\sigma^2 \quad (16)$$

以上で推定した分離行列を用いて分離を行い，目的信号を抽出する．

4. 評価実験

4.1 評価用データ

音声認識で提案する手法を評価する．評価に用いたのは連続数字認識である．1人の女性話者からの連続した数字30個を評価に利用した．サンプリングレートは16kHzである．混合行列

$$A=\begin{bmatrix}2, 3\\4, 1\end{bmatrix}$$

を用いて空港の雑音と無雑音の音声を混合し（周波数領域）雑音下の音声データを作成した．次の実験では分離行列の一部（w_{12}：3.0，w_{22}：2.0）は既知と仮定する．

4.2 音声認識実験概要

音声認識にはUniversity of Cambridgeより一般に公開されている隠れマルコフモデル（HMM：Hidden Markov Model）[*5]による音声認識ソフトウェア[6]を用いた．このソフトウェアは，MFCC（Mel-Frequency Cepstrum Coefficient）[*6]と呼ばれる周波数特徴と正規化パワー[*7]からなる12次元の特徴ベクトルを特徴量に用いている．HMMのパラメータに有限個の状態および各状態における出力の確率分布関数がある．音声認識では複数の正規ガウス分布の混合で各状態における出力の確率関数を表現する．この実験におけるHMMのパラメータの状態数は5，各状態の正規ガウス分布の混合数はすべて4とした．

4.3 評価結果

従来と提案のICAに基づくBSS法により抽出した音声信号のサンプルを**図3**に示す．提案手法による音声信号は従来手法に比べ雑音成分がより抑圧されていることが確認できる．また実利用での有効性を確認す

＊5 **隠れマルコフモデル（HMM）**：不確定な時系列のデータをモデル化するための統計的手法．

＊6 **MFCC**：人間の聴覚を模した音声の特徴量係数の系列．

＊7 **正規化パワー**：音声信号のパワーを対数領域に正規化した値．

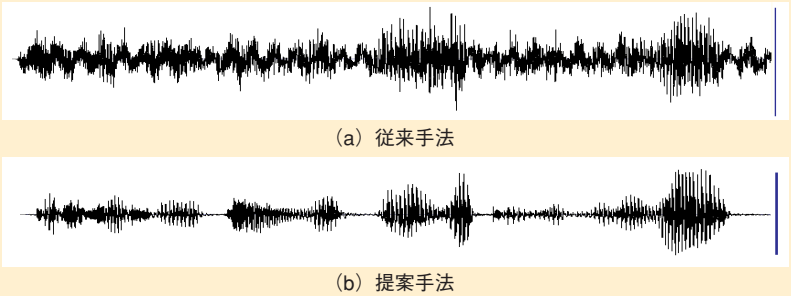

図3　音声信号サンプル：従来と提案のICAに基づくBSS法

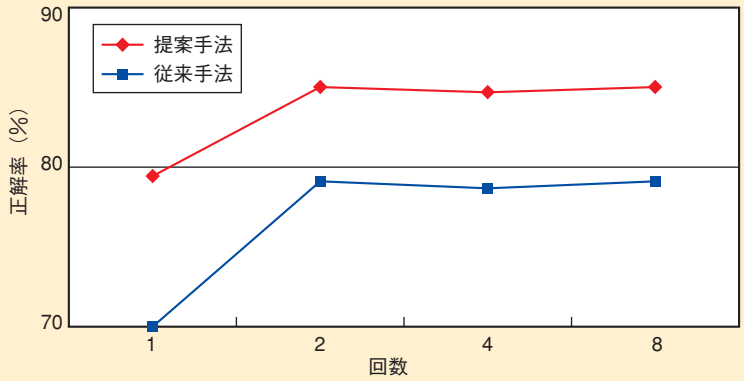

図4　評価結果

るため音声認識の前処理としての評価実験を行った．音源分離による目的音源を抽出し音声認識により評価を行った．提案手法と従来手法の評価結果（正解率）を**図4**に示す．横軸は分離行列の推定に必要な繰返しの回数を示す．提案手法では一回の推定で従来手法の複数回の推定結果とほぼ同じ性能になる．提案手法により従来手法に比べて79%から84%に認識率が向上することを確認した．

5. あとがき

考案したマイク・アレイを用いた雑音除去技術について述べた．本技術は，ICAという汎用性の高い信号統計処理に，実際の通話環境から得られる音源の位置情報を利用した最適化手法を導入した2マイク・アレイによる雑音除去技術である．実験では通話環境での利用シーンを再現し，ユーザの口元からマイクまでの伝達関数パラメータを計測・利用することにより，少ない計算量で従来手法より音声認識性能を向上できることを確認した．

マイク・アレイを用いた本雑音除去技術は，今後の音声コミュニケーションの幅を広げる技術として，また音声認識・翻訳サービスへの基本技術として期待される．

文　献

[1] S. F. Boll: “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Transactions on ASSP, No. 2, pp. 113–120, 1979.

[2] 金田　豊：“音響システムとディジタル処理，”1995.

[3] 西川 剛樹，荒木 章子，牧野 昭二，猿渡 洋：“帯域分割型ICAを用いたBlind Source Separationにおける帯域分割数の最適化，”日本音響学会2001年春季研究発表会，2001.

[4] T-W. Lee, et al.: “Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources,” Neural Computation, Vol. 11, pp. 417–441, 1999.

[5] A. Bell, T. Sejnowski: “An Information Maximization Approach to Blind Separation and Blind Deconvolution,” Neural Computation, 7 : 1129–1159, 1995.

[6] M. J. F. Gales, P.C. Woodland: “Recent advances in large vocabulary continuous speech recognition: An HTK perspective,” ICASSP, 2006.