

ICAにおけるパーミュテーション問題の解決法の提案

レアンドロ・ディ・ペルシア[†] 大田 健紘[‡] 柳田 益造[‡]

[†] エントレリオス大学バイオ工学部 アルゼンチン パラナー市
[‡] 同志社大学工学部 〒610-0321 京都府京田辺市多々羅都谷 1-3

E-mail: [†] ldpersia@ciudad.com.ar, [‡] etf1704@mail4.doshisha.ac.jp myanagid@mail.doshisha.ac.jp

あらまし 本稿では、周波数領域独立成分分析に基づく手法によりブラインド音源分離を行った際に問題となる周波数ビン間で信号の並び順を揃えるパーミュテーション問題の解決法を提案している。これまでに、パーミュテーション問題の解決法としては、各周波数の指向特性を基に解決する方法や、分離信号の周波数間での相関に基づいた方法、さらには音声信号の調波性に基づいた方法などが提案されているが、完全な解決には至っていない。そこで、筆者らはパーミュテーション問題を回避する手法を提案している。従来の周波数領域 ICA は、各周波数ビンに時間-周波数の行列として扱い分離処理を行っていたため、パーミュテーション問題と対峙する必要があった。しかし、提案法では、各周波数ビンを連結し1つのベクトルとして扱い分離処理を行っている。これにより、全周波数ビンが同時に入れ替わる可能性はあるが、各周波数ビンが他のチャンネルに入れ替わることはない。

キーワード ブラインド音源分離, 独立成分分析, パーミュテーション

A Method for Solving the Permutation Problem in ICA

Leandro Di Persia[†] Kenko OHTA[‡] and Masuzo YANAGIDA[‡]

[†] Universidad Nacional de Entre Ríos, Paraná, Argentina

[‡] Faculty of Engineering, Doshisha University 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, 610-0321 Japan

E-mail: [†] ldpersia@ciudad.com.ar, [‡] etf1704@mail4.doshisha.ac.jp myanagid@mail.doshisha.ac.jp

Abstract Proposed is a method for solving the permutation problem in frequency-domain independent component analysis (ICA). Up to now, several approaches for the permutation problem are proposed, for example those based on directivity pattern of each frequency bin, the correlation between frequency bins and harmonic structure. These methods, however, do not give the complete solution. So, it is necessary to develop a method which solves or does not need to solve the permutation problem. The traditional frequency-domain ICA executes separation using a matrix consisting of the short time Fourier transform of the received signal. The proposed method, however, carries out separation on a long vector constructed as connected frequency bins of the received signal. Though, there is a possibility that whole the channel of the separated signal might be exchanged with the other channel, there is no possibility that a single channel or a group of channels are permuted with other channels.

Keyword Blind Source Separation, Independent Component Analysis, Permutation

1. はじめに

現在の音声認識システムは、マイクに接近した位置からの音声に対してはそれなりに高い認識精度を有する一方で、マイクから離れた位置からの音声に対しては、周囲の雑音や部屋の残響の影響を受けて認識精度は著しく低下してしまう[1]。実環境で音声認識を行うためには音声認識システムそのものの認識能力の向上とともに、それに入力される音声に対して何らかの処理を行って認識しやすい形にしておくことが有効である

と考えられる。その方法の一つがブラインド音源分離 (BSS: Blind Source Separation) である。BSSとは、線形に混合された信号が複数のマイクロホンへ入力された時に、音源信号や混合過程を知ることなく、観測信号のみから音源信号を推定し分離する技術である。近年、独立成分分析 (ICA: Independent Component Analysis) に基づく手法が盛んに研究されており、瞬時(畳み込みでない)混合の問題に対しては十分な分離性能が得られている[2]。また、残響に対処する方法と

しては周波数領域 ICA がある[3,4]. これは信号を短時間フーリエ変換により周波数領域に変換することによって, 周波数ビン毎の瞬時混合として ICA を適用するものである.

しかし, BSS に周波数領域 ICA を適用した場合に, 各チャンネルの周波数ビンの入れ違いを正しく並び替えるパーミュテーション問題を解決する必要がある. パーミュテーション問題に対する解決策としてはこれまでに, 各周波数ビンでの指向特性を求め, 雑音の方向を推定する方法[5], 分離信号の各周波数ビンでの相関を用いる方法[2], これらを統合した方法[6], さらに音声の調波構造を用いることにより文献[6]の改善を行った方法[7], および近距離場モデルを用いて音源の位置情報を推定する方法[8]などが提案されている. しかし, いまだパーミュテーション問題の完全な解決には至っていない.

そこで, 本稿ではパーミュテーション問題を回避できる周波数領域 BSS を提案し, 提案法を用いることによって分離性能が向上することを確認する.

2. ブラインド音源分離

BSS とは, 複数のマイクロホンへ線形に混合された信号が入力された時に, 音源信号や混合過程を知ることなく, 観測信号のみから音源信号を推定し分離する技術である. これを定式化すると, マイクロホン数を M , 音源数を N とすると, マイクロホン M_m での観測信号は,

$$x_m(t) = \sum_{n=1}^N h_{mn}(t) * s_n(t) \quad (1)$$

で表される. ただし, h_{mn} は音源 S_n からマイクロホン M_m へのインパルス応答, $s_n(t)$ は音源 S_n の信号であり, $*$ は畳み込み演算を表す. 式(1)において, 観測可能な要素は x_m だけで, h_{mn} と s_n は未知である. BSS はこの状況において, h_{mn} と s_n を推定する問題である. しかし, 時間領域の BSS では, 式(1)で表される実環境での残響が伴う混合に対して, 残響時間に適した長さのフィルタ係数を推定する必要がある. 残響が長くなるとこの問題は非常に困難になる.

そこで, このフィルタを直接求めずにその周波数特性を求める周波数領域 BSS を考える. マイクロホン数 M , 音源数 N の混合モデルを求めるために, 式(1)の両辺のフーリエ変換を行うと, $X_m(f) = \sum_n H_{mn}(f) S_n(f)$ と表される. そして, $m=1 \sim M$ についてまとめると, 周波数領域 BSS の混合モデルは

$$X(f) = H(f)S(f) \quad (2)$$

と表される. ただし, $X(f) = (X_1(f), \dots, X_M(f))^T$ は観測ベク

トル, $S(f) = (S_1(f), \dots, S_N(f))^T$ は音源信号ベクトルそして, $H(f)$ は混合行列である. f は周波数である.

BSS を解く手法として, 混合信号中の各要素の独立性に基づいた ICA が一般的に用いられる. ICA には時間領域の方法と周波数領域の方法がある. 先に述べたように, 周波数領域 ICA は, 短時間の残響を瞬時混合の問題に置き換えて解くことができるため, 現在, 実環境で頑健な手法として盛んに用いられている.

周波数領域 ICA は, x_m の短時間フーリエ変換により得られた観測信号ベクトル $X(f)$ に対して, 各周波数ビンにおいて学習された $W(f)$ を用いて分離を行う. ここで, $Y(f) = (Y_1(f), \dots, Y_N(f))^T$ を分離信号ベクトルとすると分離過程は,

$$Y(f) = W(f)X(f) \quad (3)$$

と表される.

ICA は音源信号の独立性を最大にするという基準に基づいて分離を行っており, $W(f)$ の行が入れ替わったとしても $Y(f)$ の独立性は保たれるので, 任意の周波数 f_1 と f_2 において, $Y_n(f_1)$ と $Y_n(f_2)$ が必ずしも同じ音源に対応しているとは限らない. 従って, $W(f)$ の行の並びを $Y_n(f_1)$ と $Y_n(f_2)$ が同じ音源に属するように並べ替える「パーミュテーション問題」を解決する必要がある.

3. 従来のパーミュテーション問題解決法

パーミュテーション問題に対する従来の解決手法として, 各周波数ビンでの指向特性を調べ, 雑音源の方向を推定し, 方向推定結果を基に並び替えを行う方法がある[5]. しかし, すべての周波数において理想的な死角が形成できるわけではなく, 特に低周波数における方向推定性能が低いと指摘されている.

そこで, 各周波数ビンでの相関を用いた方法[2]が提案されている. 各周波数ビンのエンベロープは, 同じ音源については, 特に近傍の周波数で高い相関を持つと考えるのが妥当である. そのため, 周波数差が比較的小さい範囲内で相関の和が最大になるように決定していくことが考えられるが, ある周波数で間違えると以降の周波数でも間違えてしまうため, 安定性に欠けるといわれている. また, ICA による分離性能が比較的良好と思われる周波数ビンからパーミュテーションを決定していく方法が提案されている[2]が, この方法では, 周波数差が小さくない周波数に対してもエンベロープの相関が高いことが仮定されているため適切ではない.

上述の欠点を補うために, 方向推定と相関を統合した手法が提案されている[6]. この方法では, まず指向特性を各周波数ビンに対して求め, 確実に推定できる周波数に対してパーミュテーション問題を解決する.

次に、未定の周波数に対して周波数差が小さく、かつパーミュテーションが決定されている周波数との相関の和が最大になるように決定していく。しかし、この方法でも残響が強い場合に不安定になることが報告されている[7]。

残響が強い場合に不安定になる原因として、低域での方向推定の困難さが挙げられている[7]。そして、この問題を解決するために、調波構造を用いた方法が提案されている[7]。音声は基本周波数の整数倍のところ相関が高くなる性質を持っており、調波性を用いることにより、より確実にパーミュテーション問題を解決することが可能となる。

しかし、これらの方法は本質的に並べ替えに失敗するという問題を常に含んでいる。

4. パーミュテーションフリーICA

4.1 提案法概要

そこで、我々は本質的にパーミュテーション問題が起きない周波数領域 BSS を提案する。提案法は 4 ステップからなり、パーミュテーション問題回避の基本的なアイデアは最初のステップにある。

従来の周波数領域 ICA では、受信信号に対して短時間フーリエ変換(STFT)が行われ、各周波数ビンに対して分離が行われるが、提案法では全周波数ビンに対して最適な指向特性を持つビームフォーマーがあると仮定している。これは、適切な大きさと遅延を持つ遅延和ビームフォーマーを求めることと等価である。そして図 1 に示す 2 つのビームフォーマーを求め、音源と雑音両方の推定を行う。

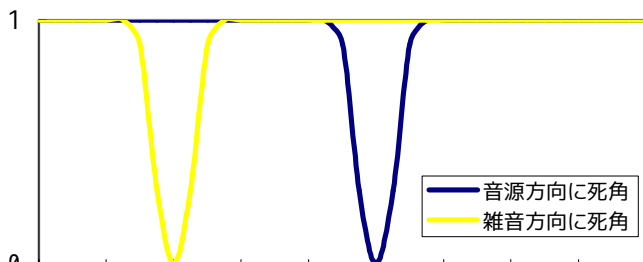


図 1 推定すべき指向特性

このビームフォーマーを求めるために、図 2 の関係を用いる。図 2 の上部は、受信信号が STFT により 2 次元の時間-周波数平面にマッピングされる様子を示しており、従来法では各周波数に対して分離処理を行うことになっていた。一方、提案法では図 2 の下部に示したように、各周波数ビンを連結し、全体で一つの大きなベクトルとして扱ってビームフォーマーを求めている。さらに、各音源の独立性を仮定しているので、時間方向の順序は重要ではなく、図 3 に示すように各時刻における全周波数の値をベクトルの成分として連

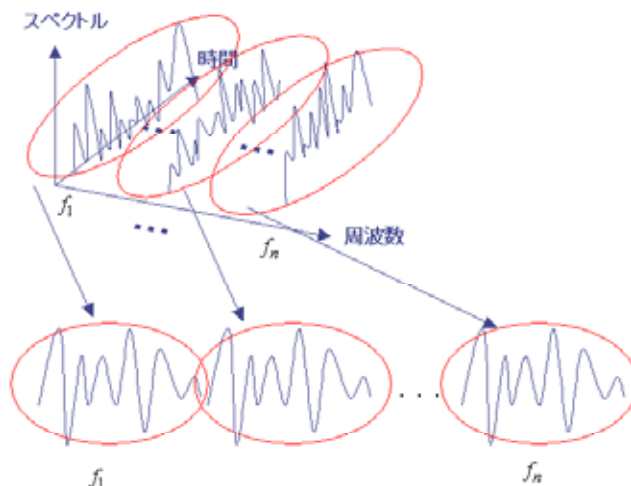


図 2 従来の周波数領域 ICA と提案法の対応関係

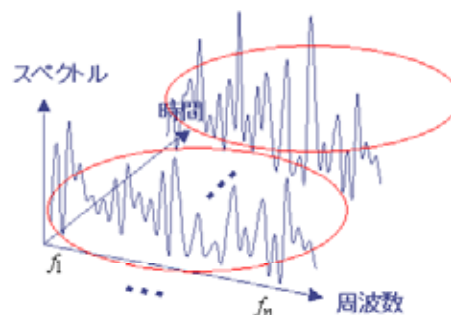


図 3 時間と周波数を入れ替えた例

結した場合も同じ結果になる。(このことは、シミュレーションによっても確認している。)

提案法では、窓長に関する制約を仮定していないので、異なる窓長に対して動作し、適切な窓長選択に関する問題もなくなる。

4.2 処理手順

まず、 N 個の音源信号を M 個のマイクロホンで受信した信号に対してSTFTを行い、各周波数ビンを連結し M 個の長いベクトルを作成する(ただし本稿では $M=2$, $N=2$ とする)。次に、JADE[9]を用いて各ベクトルに対して分離処理を行う。JADEとは、Cardosoらにより提案された手法で、4次クロスコキュムラントを対角化する手法である。4次クロスコキュムラントは次のように定義される。

$$\begin{aligned}
 \text{cum}(x_i, x_j, x_k, x_l) &= E(x_i, x_j, x_k, x_l) - E(x_i, x_j)E(x_k, x_l) \\
 &\quad - E(x_i, x_k)E(x_j, x_l) - E(x_i, x_l)E(x_j, x_k)
 \end{aligned} \tag{4}$$

観測信号 x が平均0で無相関化したものを z とし、音源信

号 s と無相関化された観測信号 z はある直交行列 $U = (\mathbf{u}_1, \dots, \mathbf{u}_N)$ により,

$$\mathbf{z} = \mathbf{U}\mathbf{s} \quad (5)$$

という関係で結ばれている。独立性の仮定から

$$\text{cum}(s_i, s_j, s_k, s_l) = \begin{cases} \kappa_i, & i = j = k = l \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

となる。ここで、 $N \times N$ 行列 $\mathbf{M}=(m_{ij})$ により縮約された kurtosisの行列の第 i, j 要素は

$$C_{i,j}(\mathbf{M}) = \left(\sum_{k,l=1}^N \text{cum}(z_i, z_j, z_k, z_l) m_{kl} \right) \quad (7)$$

を考える。 s のインデックスが全て同じ時だけが問題であるから、 $\kappa_i = \text{cum}(s_i, s_i, s_i, s_i)$ として式(7)を変形すると

$$\mathbf{C}(\mathbf{M}) = \sum_{i=1}^N (\kappa_i \mathbf{w}_i^T \mathbf{M} \mathbf{w}_i) \mathbf{w}_i \mathbf{w}_i^T \quad (8)$$

となる。ここで

$$\mathbf{D}(\mathbf{M}) = \text{diag}(\kappa_1 \mathbf{w}_1^T \mathbf{M} \mathbf{w}_1, \dots, \kappa_N \mathbf{w}_N^T \mathbf{M} \mathbf{w}_N) \quad (9)$$

とすると,

$$\mathbf{C}(\mathbf{M}) = \mathbf{W} \mathbf{D}(\mathbf{M}) \mathbf{W}^T \quad (10)$$

直交行列である \mathbf{W} で対角行列を挟んでいるので、 $\mathbf{C}(\mathbf{M})$ は対称行列になっている。この $\mathbf{C}(\mathbf{M})$ を対角化する \mathbf{W} を探すのがJADEである。行列の対角化にはJacobi法を用いることができる。Jacobi法は理想的には2次元収束することが知られているので、JADEは収束の速いアルゴリズムといえる。ただし、JADEは4次元配列を用いるので、信号数があまり多くない時には問題がないが、信号数が多い時には計算上の問題がある。

上記では時間領域ICAでの説明を行ったが、周波数領域ICAにもアルゴリズムを変更することなく拡張できる。周波数上で行う場合は、各周波数ビンに対して分離行列 $\mathbf{W}(f)$ を求める必要があり、通常はここで、パーミュテーションが問題となる。しかし、提案法では、分離行列を求めるのは1度でよいので、パーミュテーションが問題にならない。

次に、振幅の不確定性を解決する。提案法によりパーミュテーション問題は解決されるが、振幅に対する不確定性が残る。振幅の不確定性は、文献[2]により提案されている方法により解決する。分離行列 $\mathbf{W}(f)$ を得

た後に混合行列 $\mathbf{H}(f)$ の近似として分離行列の逆行列を求める。

$$\hat{\mathbf{H}} = \mathbf{W}^{-1} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \quad (11)$$

次に、JADEによる分離の結果,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{W}\mathbf{x} \quad (12)$$

を得る。式(11), (12)を用いて振幅の不確定性を解決した結果,

$$\hat{\mathbf{s}} = \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = (\delta_{ij} \hat{\mathbf{H}}) \mathbf{y} = \begin{bmatrix} h_{11} y_1 \\ h_{22} y_2 \end{bmatrix} \quad (13)$$

を得る。ただし、 δ_{ij} はデルタ関数で

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

と定義する。

5. 提案法の評価

5.1 評価データ

音声の収録は一般家庭のリビングルームを想定した部屋で行った。TSPにより測定した残響時間は310ms^(*)である。音源(スピーカ)と観測点(マイクロホン)の配置を図4に示す。観測点は固定し、音源の配置は図中の $\theta = -30^\circ, 0^\circ$ (時計回りを正)となるようにし、それぞれの配置をL, Mとする。マイクロホンには素子間隔5cmの2素子アレイを用いた。音源信号は、男女各1名ずつで、女性(配置:M)による10フレーズ、男性(配置:L)による10フレーズの組み合わせで、合計100個の混合音声について考える。女性および男性の音声を個別に収録し、S/Nの調整を行うことなくそのまま加算した。

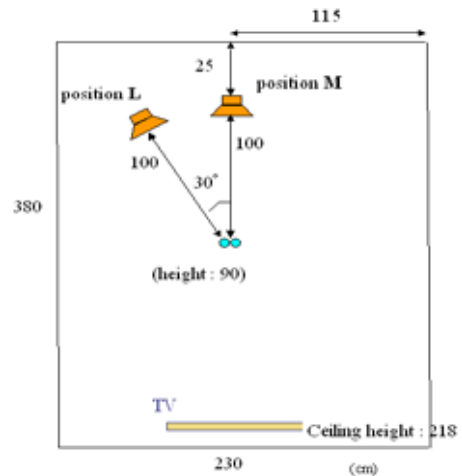


図4 データ収録環境

^{*}白色雑音の15dB減衰時間を4倍したが、直接音の減衰部分を見ていた可能性があるため、実際はこれより長い

5.2 評価指標

S/N を用いて分離信号を評価する。S/N は以下のよう
に定義した。

$$SNR = \frac{\int |S(\omega)|^2 d\omega}{\int |S(\omega) - \hat{S}(\omega)|^2 d\omega} \quad (15)$$

ここで $S(\omega)$, $\hat{s}(\omega)$ は音源信号 $s(t)$ と分離信号 $\hat{s}(t)$ をそれぞれ周波数領域に変換したものである。周波数領域で考えたのは、マイクロホンを2つ使用しているため、各音源情報はそれぞれ2通り存在する。これらをもとに分離信号が得られるため、音源信号と分離信号の同期を取るのが困難となるためである

5.3 比較対象

比較対象のアルゴリズムは、JADE により周波数領域 ICA を行い、Ikeda et al.[2]による提案法である、周波数ビン間での相関を用いた手法によりパーミュテーション問題を解決する。

5.4 評価結果

評価はまず、比較対象のアルゴリズムでパーミュテーション問題の解決ができない例に対して、提案法では解決できていることをスペクトログラムを比較することにより行う。音響分析条件を表 1 に示す。

表 1 音響分析条件

サンプリングレート	16ksamples/sec
量子化ビット数	16bits
フレームサイズ	1024 点
フレームシフト	128 点
分析窓	ハミング窓

図 5 は、a)混合前の女性音声、b)混合前の男性音声、c)従来 ICA による分離信号 1(女性)、d)従来 ICA による分離信号 2(男性)、e)提案法による分離信号 1(女性)そして f)提案法による分離信号 2(男性)のスペクトログラムを示している。図 5 の c)と d)において、特に 4kHz 以上の領域でパーミュテーションの解決に失敗していることがわかる。しかし、図 5 の e)と f)ではパーミュテーションの問題が起きていないことが確認できる。

次に、パーミュテーション問題が解決できたことにより S/N がどの程度向上するのかを、セグメンタル SNR を用いて評価する。

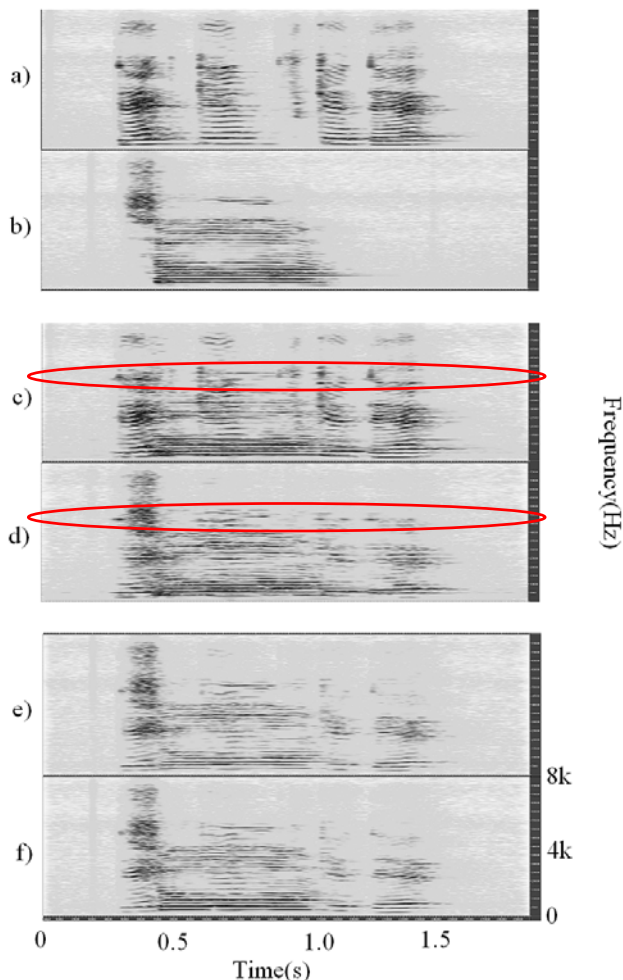


図 5 パーミュテーション問題解決の例

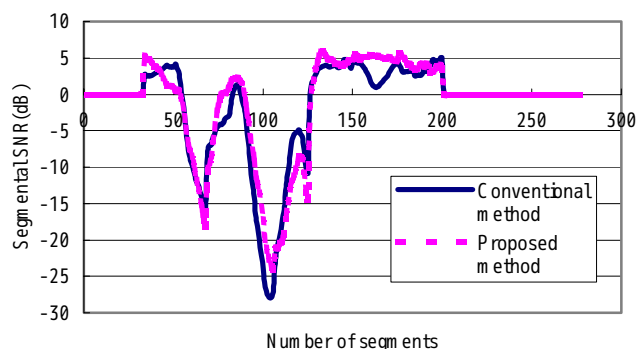


図 6 女性音声のセグメンタル SNR の比較

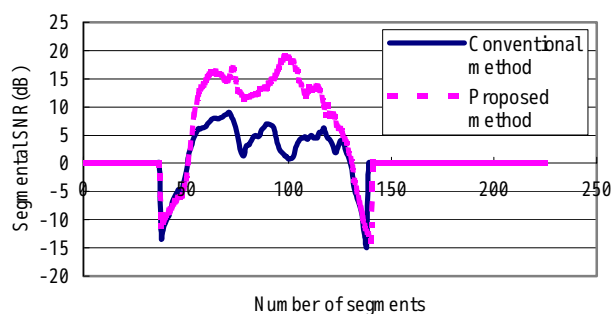


図 7 男性音声のセグメンタル SNR による比較

図 6 と 7 を比較すると、女性音声側のセグメンタル SNR の平均値で約 0.65dB 改善し、男性音声側のセグメンタル SNR の平均値では提案法により約 5.9dB の改善が確認できた。これにより、パーミュテーション問題を回避することによって解決することが分離性能の向上につながるということがわかる。

さらに図 8 に分離結果 100 個についてのセグメンタル SNR の平均値の比較を示す。

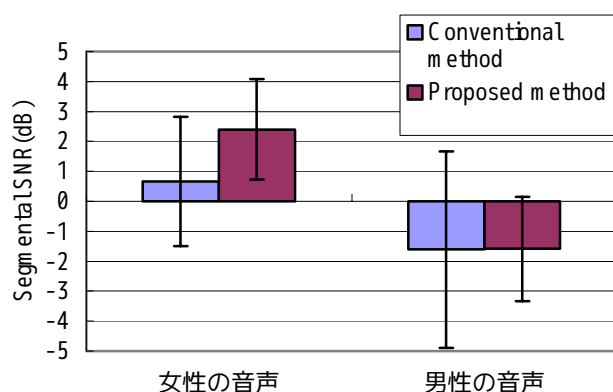


図 8 全分離結果のセグメンタル SNR の平均値の比較

図 8 より、SNR はデータにより大きくばらついているが、女性については約 1.7dB の改善、男性については平均値の改善についてはほぼ 0 であるが、標準偏差が小さくなっていることがわかる。図 8 からパーミュテーション回避による S/N の改善が確認できた。

6. まとめ

周波数領域 ICA を行う際に問題となっていたパーミュテーション問題を回避することによって解決する手法を提案した。従来法では、パーミュテーションをうまく解決できなかった信号に対して、提案法では解決できることを確認した。また、セグメンタル SNR で評価した結果、パーミュテーション問題の解決によりセグメンタル SNR を改善できることが確認できた。

7. 今後の課題

今後、我々は提案法の更なる分離精度向上のために、後処理として Wiener フィルタなどのフィルタリング手法を採用することを考えている。

また、提案法の属性として、異なるフレーム長に対しても動作すると述べた。フレーム長 N とサンプリングレート f_s と周波数分解能 Δf の間には、

$$\Delta f = \frac{f_s}{N} \quad (16)$$

の関係があるため、フレーム長を変更することは周波数分解能を変更することと等価である。そのため、異なるフレーム長の窓による STFT と JADE による分離処理を反復して行うことは、異なる周波数分解能で計算したビームフォーマーを複数求めることになる。提案したアルゴリズムでは、フレーム長 1024 点の Hamming 窓による STFT と JADE による分離を 1 度行っているだけなので、全周波数ビンに対して同一の指向特性を持つビームフォーマーを 1 つ求めているだけであり、分離性能は十分であるとはいえない。それゆえ、複数のフレーム長の窓を用いた多重解像度解析により、分離性能を向上することが期待される。

謝辞 本研究の一部は同志社大学学術フロンティア事業、ならびに文科省知的クラスタ創成事業の援助を受けた。

文 献

- [1] 中村哲: "実音響環境に頑健な音声認識を目指して", 電子情報通信学会技術報告, SP2002-12, pp.31-36, 2002.
- [2] T.W.Lee: "Independent Component Analysis", Kluwer, 1998.
- [3] S.Ikeda, and N.Murata: "A method of ICA in time-frequency domain", Proc WS on Independent Component Analysis and Blind Signal Separation (ICA' 99), pp.365-371, Aussiais, France, Jan., 1999.
- [4] A.Hyvärinen, J.Karhunen, and E.Oja: "Independent Component Analysis", John Wiley, New York, 2001.
- [5] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura: "Evaluation of blind signal separation method using directivity pattern under reverberant conditions", Proc. ICASSP2000, pp3140-3143, Istanbul, Turkey, June, 2000.
- [6] 澤田宏, 向井良, 荒木章子, 牧野昭二, "周波数領域ブラインド音源分離における permutation 問題の解法", 音響学会講演論文集, pp. 541-542, Sep, 2002.
- [7] 澤田宏, 向井良, 荒木章子, 牧野昭二, "周波数領域ブラインド音源分離における permutation 問題の頑健な解法", 音響学会講演論文集, pp. 777-778, Mar, 2003.
- [8] 向井良, 澤田宏, 荒木章子, 牧野昭二, "周波数領域 BSS における近距離場モデルを用いたパーミュテーションの解法", 音響学会講演論文集, pp. 589-590, Sep, 2003.
- [9] J.F.Cardoso, and A.Souloumiac: "Blind beamforming for non Gaussian signals", IEEE Proceeding-F, vol. 140, no 6, pp.362-370, Dec 1993.