

近傍周波数ビンの連結による周波数領域 ICA の提案

Frequency Domain ICA Connecting of Adjacent Frequency Bins

大田 健紘[†] 野口 忠繁[†] レアンドロ・ディ・ペルシア[‡] 柳田 益造[†]

Kenko OTA[†] Tadashige NOGUCHI[†] Leandro DI PERSIA[‡] and Masuzo YANAGIDA[†]

[†] 同志社大学工学部

[†] Faculty of Engineering, Doshisha University

[‡] エントレリオス国立大学

[‡] Universidad Nacional de Entre Ríos

[†] {etf1704, dtg0731}@mail4.doshisha.ac.jp, [‡] ldpersia@ciudad.com.ar, [†] myanagid@mail.doshisha.ac.jp

あらまし

本稿では、周波数領域 ICA を行う際に問題となるパーミュテーション問題を回避し、高精度に分離を行う手法を提案している。従来の周波数領域 ICA では各周波数ビン毎に分離処理を行っており、周波数毎では最適な分離が行えるが、パーミュテーションの問題が発生する。以前に、受信信号の周波数成分の一時的な統合によって構成される一本の長いベクトルを分離するパーミュテーションフリーICAを開発した。しかし、パーミュテーションフリーICAによって得られる分離行列は、全周波数帯域で共通の指向性のパターンを持つ。そのため、この手法はパーミュテーション問題を避けることができるが、各周波数ビンで最適な分離行列を探すことはできない。そこで本稿では、混合行列を元の音源信号に分離する際に、分離対象とする周波数ビンと近傍の周波数ビンを一つのグループとして扱う Multi-Bin ICA を提案する。提案法では、パーミュテーションの問題を抑えながら、周波数毎に適切な分離行列を求めることができる。Multi-Bin ICA の動作を評価するために、従来の周波数領域 ICA や Permutation-Free ICA とのセグメンタル SNR を比較する。

Abstract

This paper proposes a method for solving “the permutation problem”. Conventional frequency-domain ICA yields the optimal separation for each frequency bin, but it suffers from the permutation problem. Permutation-free ICA as a separation scheme by obtaining the separation matrix for a long vector consisting of temporal changes of all frequency components of the received signal has been developed for avoiding the permutation problem. The permutation-free ICA, however, only yields a common separation matrix for all frequency bins. So, the separation matrix obtained in the permutation-free

ICA has a common directivity pattern for all frequency bins, though the method can avoid the permutation problem. Proposed in this paper is a scheme of multi-bin ICA, that deconvolves mixed signals into original source signals by shifting piecewise integration of a set of frequency bins consisting of the frequency bin in concern and neighboring frequency bins. The proposed method can yield nearly optimal directivity in the form of the separation matrix for each frequency bin avoiding the permutation problem. Performance of the proposed multi-bin ICA is compared with that of conventional frequency-domain ICA and that of permutation-free ICA, employing segmental SNR as an evaluation index.

1 はじめに

現在の音声認識システムは、マイクロホンに接近した位置からの音声に対してはある程度高い認識精度を有する。一方で、マイクロホン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、認識精度は著しく低下する[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前処理を行い、認識しやすい信号にすることが有効と考えられる。その方法の一つとして、ブラインド音源分離 (BSS: Blind Source Separation)がある。BSSとは、複数の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、観測信号のみから各音源信号へ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の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[3]、これらを統合した方法[6]、さらには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 DiPersia らは、これまでに全周波数ビンを連結して 1 本のベクトルと考えて分離を行う Permutation-Free ICA (PF-ICA) を提案した[9][10]。この手法では、分離された信号の周波数ビン間の入れ替わりが生じない利点があるが、分離行列の周波数特性を考慮することができないという欠点があった。本研究では、対象とする周波数ビンの前後数本の周波数ビンを連結することによって、当該周波数ビンに関する分離行列を安定して求める Multi-Bin ICA (MB-ICA) を提案する。提案法は、近隣の周波数ビンを同時処理することにより分離行列の平滑化を行いパーミュテーション問題を抑えながら、分離行列の周波数特性を考慮できるため、高精度な分離が行えると期待できる。

2 従来の周波数領域 ICA

周波数領域 ICA は受音信号 $x(t)$ の第 k フレームの短時間フーリエ変換により得られた観測信号群ベクトル $X(f,k)$ に対して、周波数ビン f において学習された分離行列 $W(f,k)$ を用いて分離を行う。分離行列を得る方法には様々なものがあるが、本研究では JADE を用いる[11]。ここで、 $Y(f,k)$ を分離された信号群のスペクトルを表すとすると分離過程は、

$$Y(f,k) = W(f,k) X(f,k) \quad (1)$$

と表される。ここで、 k はフレーム番号を表す。ICA は音源信号の独立性を最大にするという基準により分離を行うので、 $W(f,k)$ の各行が入れ替わったとしても、 $Y(f,k)$ の独立性が最大になる可能性がある。従ってこの時、 $Y(f,k)$ のある周波数ビンにおいて、本来対応するはずの音源が別の音源のものに入れ替わっている可能性がある。パーミュテーション問題とは、周波数ビンによって入れ替わった分離信号の音源番号を整列させる問題である。

従来の周波数領域 ICA では分離後に必ずパーミュテーション問題を解決しなければならないが、これまで完全には解決されておらず、本質的に並び替えに失敗する可能性がある。

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、方向推定結果に基づいて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[3]。

そこで、各周波数ビンでの相関を用いた方法[3]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[3]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で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び替えに失敗するという問題を常に含んでいる。

4 Permutation-Free ICA (PF-ICA)

パーミュテーション問題を解決するために、DiPersia らは周波数ビンの入れ替わりが発生しない

PF-ICAを提案した[9][10]. この手法は、時間-周波数領域における受信信号の各周波数ビンのフレーム時系列を f_1, f_2, \dots, f_n の順に連結して、全周波数ビンを一本のベクトルとして扱うものであった. この処理を図1に示す. 図1の左上の図は時間周波数領域上での各周波数ビンを表しており、それらを右下の図のように各周波数ビンを連結し、マイクロホン毎に1本のベクトルを形成する.

そして、このベクトルに対して分離を実行し、分離後は各周波数ビンに分割を行う. この手法では、全ての周波数ビンを一括して処理するため、各周波数ビンにおいて音源番号を並べ替える必要がなくなる. つまり、パーミュテーション問題を回避したことになる. しかし、この手法では全ての周波数において同じ分離行列を用いて分離を行うことになるため、壁からの反射などによって生じる周波数毎の特性を考慮することができず、必ずしも最適な分離結果が得られるとは限らない. なぜなら、PF-ICAでは(マイクロホン数) \times (マイクロホン数)のサイズの分離行列を1つ推定するだけであるからである.

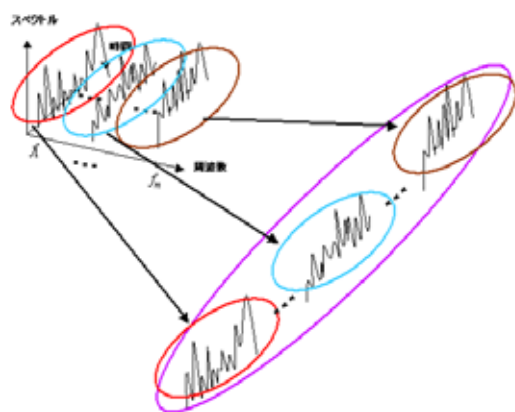


図1. 1本のマイクロホンへの混合信号の短時間スペクトルの時系列を各周波数ビン毎の時系列とみて全周波数ビンの時系列を連結したベクトル

5 Multi-Bin ICA (MB-ICA)

本稿では、パーミュテーション問題を抑えつつ、周波数特性を考慮できるMB-ICAを提案する. 提案法は、PF-ICAと同様に、周波数ビンを連結してベクトルとして扱い、分離を実行する. しかし、PF-ICAのように全周波数ビンを連結して一本のベクトルとして扱うのではなく、分離対象の周波数ビンの前後 k 本ずつの周波数ビンを連結し、単一周波数ビンのベクトルの $2k+1$ 倍の長さのベクトルを作成する. ただし、前後に k 本ずつの周波数ビンを取ることができない場合はその限りではない. そして、分離対象の周波数ビンは1本ずつシフト

する. この処理を図2に示す. 図2の上方の図は時間周波数領域上での各周波数ビンを表しており、そして下方の図は左から順に周波数 $f_1, f_2, f_3, \dots, f_{n-2}, f_{n-1}, f_n$ に対して生成されたベクトルを表している.

そして、このベクトル毎に分離を行い、その分離結果の中央の周波数ビンに対応する部分を取り出し、分離信号を時間-周波数領域に戻す.

対象とする周波数ビン毎に分離を実行すると、従来法のように音源番号の並べ替えが必要となるが、本手法では、分離行列の隣接周波数ビン間での差は平滑化されるので、困難は生じにくい. 平滑化の程度は連結する周波数ビン数 k によって制御できる.

ここで、従来法、PF-ICA、MB-ICAの3手法のアルゴリズムを図3に対比的に示す.

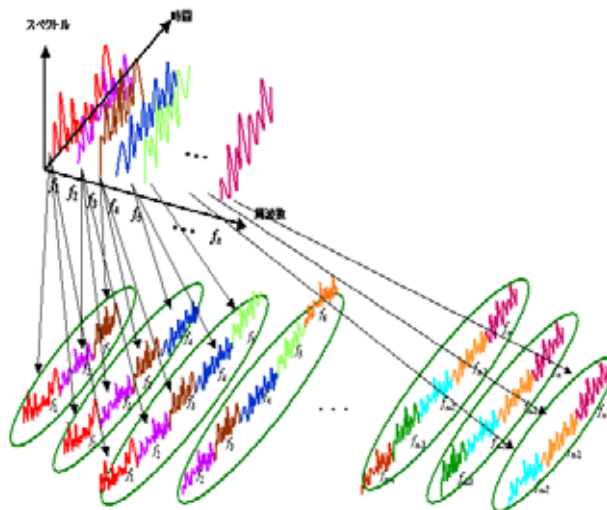


図2. 隣接周波数ビンを連結したベクトル



図3. 各手法のアルゴリズムの比較

6 評価

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。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。TSP(Time Stretched Pulse)により測定した残響時間は約370msである。音源データは、男性と女性それぞれ10発話で100通りの組み合わせである。ここでは、音源位置としてP2とP3を用いて音声の収録を行った。図3のそれぞれのアルゴリズムによる分離結果をセグメンタルSNRによって比較する(MB-ICAでは、前後に連結する周波数ビンをもとにする)。セグメンタルSNRの定義を式(2)に示す。

$$SNR(k) = \frac{\int |S(f,k)|^2 df}{\int |S(f,k) - \hat{S}(f,k)|^2 df} \quad (2)$$

ここで $S(f,k)$, $\hat{S}(f,k)$ はそれぞれ音源信号 $s(t)$ と分離信号 $\hat{s}(t)$ のフーリエ変換である。

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.[3]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。

図5にセグメンタルSNRを示す。ただし、音響分析条件は、サンプリングレートが16ksamples/sec、量子化ビット数が16bits、フレームサイズが512点、フレームシフトが256点である。

図5の”☆”は提案法により有意に認識率が向上したことを表しており、”△”は有意差なしを表しており、”x”は認識率が有意に低下したことを表している。従来法では女性のSNRの改善が0.7dBであり、男性の2.5dBよりも分離精度が悪かったが、PF-ICAでは男性・女性共に2.2~2.3dBとほぼ同じ改善程度であった。また、MB-ICAにおいては、男性の場合では前後4,5本程度の周波数ビンを連結した場合(周波数では300Hzの幅)に3.5dBと最もS/Nが良くなった。しかし、女性の場合では、連結する周波数ビンの本数を増やすほど分離精度が良くなっており、PF-ICAのSNRに近づいている。

次に連結する周波数ビン数を固定して、フレーム長を変化させた場合の分離性能を調べた。図6に男性の結果、図7に女性の結果を示す。図6,7をみると、連結する周波数ビンの数が少ない場合は、フレーム長が短い方がSNRが高くなっており、連結本数が増えるにしたがって、フレーム長が長い方がSNRが高くなる傾向があることがわかる。

7 考察

7.1 男女混合音に対する従来法との比較

Multi-Bin ICAでは条件を適切に設定すれば、従来法と比較して標本点数が多いため高精度の分離ができると考えられる。また、従来法による分離では前後の周波数との関連性がないため分離結果の並びがバラバラになりやすいが、提案法では周波数ビンオーバーラップさせながらベクトルを1本ずつ形成しているため、隣接の周波数ベクトルで共通部分が多く、近隣周波数の分離行列も似

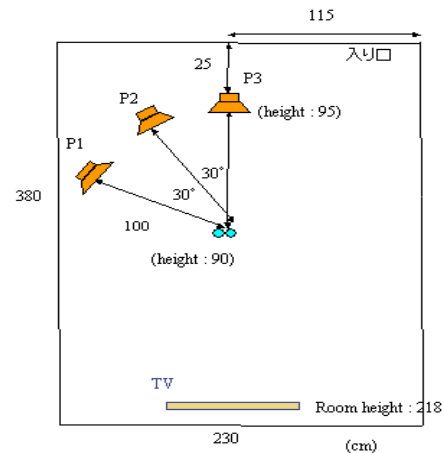


図4. 収録環境

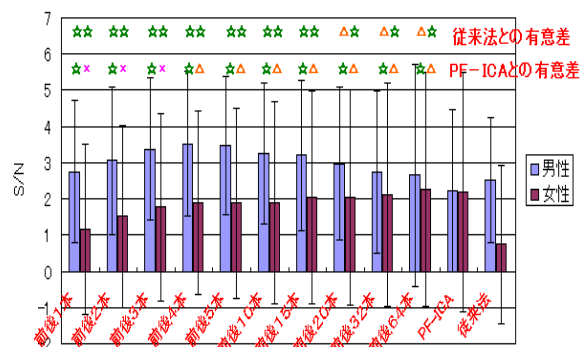


図5. フレーム長とフレームシフトを固定した場合のSNRの比較

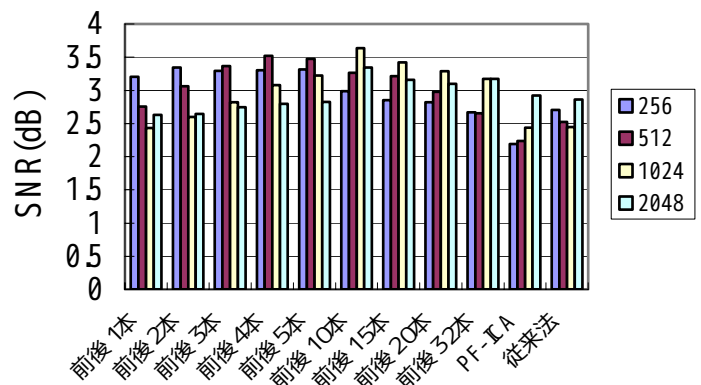


図6. 連結する周波数ビン数を固定した場合のSNRの比較(目的音は男性)

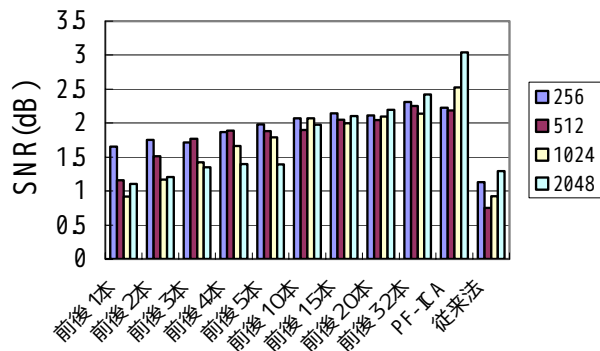


図7 連結する周波数ビン数を固定した場合のSNRの比較 (目的音は女性)

たような特性になる. 今回は PF-ICA を前処理とした分離行列の並べ替えを行ったが, 隣接周波数ビンの関連性が強くなるため, 隣接フレーム間で似た並べ方をすることで従来法よりも良くなったと考えられる.

また, PF-ICA と比較すると, 各周波数のベクトルごとに分離を実行しているため, 男性の場合では分精度が向上したと考えられる. しかし, 女性の場合では, 男性の場合ほど, 分離精度の向上が見られなかった. これは女性の声では, 基本周波数が高いことが原因となっていると考えられる. 基本周波数が高いと, 母音などで高調波間の周波数間隔が広くなり, MB-ICA では, 近隣周波数の情報を有効に使用することができなかったと考えられる. つまり, 複数の高調波にまたがる程度に周波数ビンを連結するにはかなりの数の周波数ビンを連結する必要があるが, そうすると, 処理対象の周波数帯域が広がってしまい, PF-ICA と同様の欠点が生じるため, それほど改善できないことになると考えられる.

7.2 同性の声の混合の場合の動作

次にスピーカのセッティングを P1 と P2 の位置に変更して, 男性同士, 女性同士の混合音声を生離した場合でも同様の傾向が現れるかを調査した. つまり, 男性同士の混合であれば, 前後にある一定数の周波数ビンを連結したところで SNR のピークがあり, 一方女性同士の混合であれば, 前後に周波数ビンを連結するに従い SNR が改善するかを調べた.

音声データは, 男性 5 人, 女性 5 人による各 10 発話を用いた. そして, 男性同士 5 組, 女性同士 5 組, 男性と女性 5 組のデータを振幅を調整することなく計算機上で加算した. 分離に用いたフレーム長は 1024 点, フレームシフトは 512 点である.

まず, 従来法との比較をスペクトログラムでもって見る.

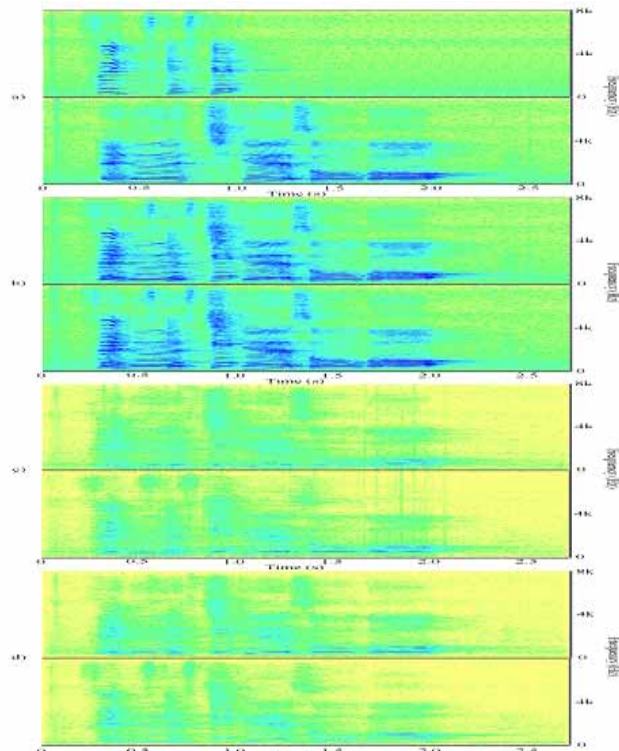


図8 スペクトログラムの比較 a)混合前の各話者の音声信号, b)混合後の信号, c)従来法による分離結果, d)MB-ICA による分離結果

図8にスペクトログラムの一例を示す. 図8のc)の従来法による分離結果の SNR の平均は 5.54dB であった. 一方, d)の提案法による分離結果では SNR の平均は 6.46dB となった. スペクトログラム上で比較する限りでは, 高周波数部分で若干の改善があることが確認される程度である. そこで, 図9にこの場合の各フレームでの SNR を比較したグラフを示す.

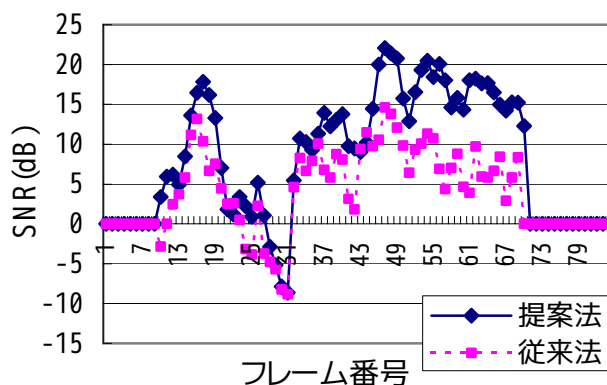


図9 各フレームでのセグメンタル SNR の変動の比較

フレームごとに見ると, 特定の音声についての結果であるがパーミュテーションを高精度に解決し各周波数の特性を考慮した結果 SNR が改善し

ている様子がよくわかる. 図9では, 最大で 14.1dB, SNR が改善した.

7.3 動作の連結ビン数依存性

次に, 分離結果をこれまでと同様に SNR を用いて評価した. 結果の連結ビン数依存性を図 10, 11 に示す. ただし, 前後に連結する周波数ビンの数を 1, 2, 4, 8, 16, 32, 64, 128 本とした.

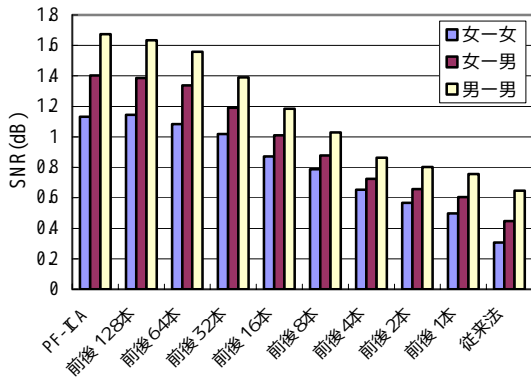


図 10 P1側のデータ(凡例の左側)を目的音とした場合の SNR の比較

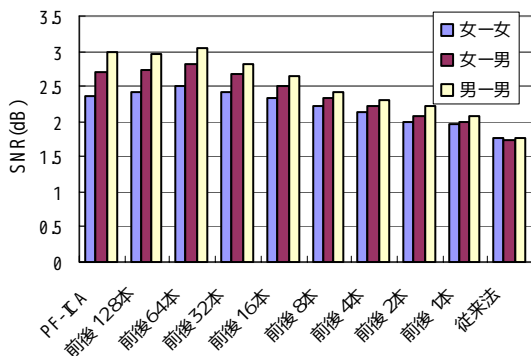


図 11 P2側のデータ(凡例の右側)を目的音とした場合の SNR の比較

図 10, 11 より, 従来法の SNR が最も低く, そして前後に周波数ビンを連結していくごとに SNR が改善している. 最終的に 64 本または 128 本の周波数ビンを前後に連結したところで, PF-ICA と同等もしくは若干の改善が見られる. しかし, 図 5 と比較すると異なる傾向を示している.

つまり, MB-ICA は, 局所的に見ると各話者の持つ基本周波数とその組み合わせに依存した性能を持つと考えられる. 事実, 図 5 に示した評価と同一の話者を用いた場合は図 5 の SNR の変化と同様の傾向を示した. 男性話者の特定の発話の基本周波数を調べたところ, 平均で約 100Hz 程度であった. 女性話者の基本周波数は平均で約 300Hz であった. このことより, 混合される話者の基本周

波数の差が大きい場合に提案法が有効になると考えられる. 一方, 大局的に見ると図 10 と図 11 のグラフのように, 周波数ビンを連結するほど, PF-ICA に近づき, 若干の性能の向上が見られると考えられる.

以上のことから, 今後課題として, 各話者の平均的な基本周波数を調査し, 話者の組み合わせによる性能の違いを詳しく評価する必要がある. さらにスピーカの配置の影響も調べる必要があると考えられる.

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ー創成事業の援助を受けた.

参考文献

- [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.Hyvarinen, 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, pp.3140-3143, 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] レアンドロ・ディ・ペルシア, 大田健紘, 柳田益造, "ICA におけるパーミュテーション問題の解決法の提案", 信学技報 Vol.105 No.686, pp. 53-58, Mar, 2006.
- [10] レアンドロ・ディ・ペルシア, 野口忠繁, 大田健紘, 柳田益造, "パーミュテーションフリー ICA の動作解析", 信学技報 Vol.106 No.78, pp.1-6, May, 2006.
- [11] J.F.Cardoso, and A.Souloumiac: "Blind beamforming for non Gaussian signals", IEEE Proceeding-F, vol. 140, no 6, pp.362-370, Dec 1993.