

## 近傍周波数ビンの連結による周波数領域 ICA の検討

安井 耕平<sup>†</sup> 野口 忠繁<sup>†</sup> 大田 健紘<sup>†</sup> レアンドロ・ディ・ペルシア<sup>‡</sup> 柳田 益造<sup>†</sup>

<sup>†</sup>同志社大学工学部 〒610-0321 京都府京田辺市多々羅都谷 1-3

<sup>‡</sup>エントリオス国立大学 アルゼンチン パラナー市

E-mail: <sup>†</sup>{btd0102,dtg0731,etf1704}@mail4.doshisha.ac.jp myanagida@mail.doshisha.ac.jp,

<sup>‡</sup>ldpersia@ciudad.com.ar

あらまし 周波数領域独立成分分析に基づくブラインド音源分離において、近傍周波数ビンを連結することにより分離対象となる周波数ビンに関する分離行列を安定して求める周波数領域 ICA の検討を行っている。Di Persia らの提案したパーミュテーションフリーICA では、全周波数ビンで分離行列が共通になってしまい、周波数ビン毎に最適な分離行列を生成することになっていないという問題があった。そこで、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、周波数特性の方向依存性を考慮し、かつパーミュテーションも抑えることのできる ICA として Multi-Bin ICA を提案している。動作解析には反射による周波数特性がどの方向も同一に近い場合と、強制的に周波数特性が方向によって異なるようにした部屋で行った。前者ではパーミュテーションフリーICA と Multi-Bin ICA に差はなかったが、後者において Multi-Bin ICA の有効性を確認できた。

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

## Source Separation by Frequency-Domain ICA on Locally Connected Frequency Bins

Kohei YASUI<sup>†</sup> Tadashige NOGUCHI<sup>†</sup> Kenko OHTA<sup>†</sup> Leandro Di Persia<sup>‡</sup> Masuzo YANAGIDA<sup>†</sup>

<sup>†</sup> Faculty of Engineering, Doshisha University 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, 610-0321 Japan

<sup>‡</sup> Universidad Nacional de Entre Ríos, Paraná, Argentina

E-mail: <sup>†</sup>{btd0102,dtg0731,etf1704}@mail4.doshisha.ac.jp myanagida@mail.doshisha.ac.jp,

<sup>‡</sup>ldpersia@ciudad.com.ar

**Abstract** Proposed is a processing scheme of Blind Source Separation(BSS) employing frequency-domain Independent Component Analysis(f-ICA) to a vector consisting of time series of adjacent frequency components, contrasting with Permutation-Free ICA(PF-ICA) applying f-ICA to a vector consisting of all the frequency components. The proposed method is named Multi-Bin ICA(MB-ICA). It can treat directionality of frequency characteristics of the sound field, while PF-ICA proposed by Di Persia cannot treat directionality of frequency characteristics. Its performance is confirmed by comparing with those of ordinary ICA and PF-ICA for room environments of almost equal directionality and of forcedly asymmetrical directionality in the room characteristics.

**Keyword** Blind Source Separation, Independent Component Analysis, Permutation

### 1. はじめに

従来の周波数領域 ICA[1]では周波数ビン毎に分離を行っており、周波数ビン間で分離した信号の入れ替わりが発生する。これはパーミュテーション問題と呼ばれ、これまでに完全な解決法は見出されていない。Di Persia らは、これまでに全周波数ビンを連結して1本のベクトルと考慮して分離を行う Permutation-Free ICA (PF-ICA) を提案した[2][3]。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って分離行列の平滑化を行って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、最適ではないが大きな間違いのない分離が行えると期待できる。

本報告では、2つの収録環境を使って MB-ICA の評価を行った。1つ目は一般家庭のリビングルームを想定したものである。しかし、部屋に設置したものは TV

と音声収録装置のみであり、4面の壁は同一の建材であるので、反射による周波数特性が方向によってそれほど変わらない可能性がある。MB-ICAは周波数特性が方向によって違う環境での効果を狙ったものなので、この音場環境では、提案法の正常な評価ができないと考えられる。

そこで、2つ目の収録環境として、音場環境に変化を与えるために吸音材やよく反響する素材を部屋に置いた環境での評価を行った。これにより強制的に反射による周波数特性の方向依存性を作り出した。

この2つの環境で各アルゴリズムを評価することにより、各アルゴリズムの長所と短所を確認し、さらなる分離精度の向上を図る手法を検討した。

## 2. 従来の周波数領域ICA

### 2.1. ブラインド音源分離

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、

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

と表される。ここで  $h_{mn}(t)$  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  $S_n$  の信号波形であり、\* は畳み込み積分を表す。式(1)において、観測可能な要素は  $x_m(t)$  だけで、 $h_{mn}(t)$  と  $s_n(t)$  は未知である。BSSは、この状況において  $s_n(t)$  を推定する問題である。しかし、時間領域のBSSでは、反射を伴う混合に対しては、遅延時間に相当する長さの混合フィルタのデコンボリューションを行う必要があるため、残響が長くなるとこの問題は非常に困難になる。

### 2.2. 周波数領域ICA

時間領域で一般に受音点での信号は、式(1)のような畳み込み混合の和となるが、 $h_{mn}(t)$  の逆特性を求めるのは複雑な問題である。そこで、フーリエ変換を行い、周波数領域で考えると、式(1)で表される混合過程は成分毎に

$$X_m(f) = \sum_{n=1}^N H_{mn}(f) S_n(f) \quad (2)$$

と表され、これをまとめると

$$X(f) = H(f) S(f) \quad (3)$$

のようになる。ここで  $f$  は周波数、 $X_m(f)$ 、 $H_{mn}(f)$ 、 $S_n(f)$  はそれぞれ  $x_m(t)$ 、 $h_{mn}(t)$ 、 $s_n(t)$  のフーリエ変換であり、 $X(f) = [X_1(f), \dots, X_M(f)]^T$ 、 $S(f) = [S_1(f), \dots, S_N(f)]^T$ 、 $H(f) = [H_{mn}(f)]$  を表している。時間領域での畳み込みは周波数領域では、周波数毎の積の形になるため、問題を簡単化できる。

ICAはBSSを解く一般的な手法であり、混合されている各音源波形の独立性を仮定して分離を行うものである。時間領域ICAは、残響環境下では収束性が悪いなどの欠点があるが、周波数領域ICAは、時間領域での畳み込みを周波数ごとの積すなわち瞬時混合の形に置き換えて解くことができるため、実環境で有望な手法として盛んに用いられている。

周波数領域ICAは  $x_m(t)$  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  $X(f, k)$  に対して、各周波数ビンにおいて求められた分離行列  $W(f, k)$  を用いて分離を行う。ここで、 $k$  はフレーム番号を表す。 $Y(f, k)$  を分離信号ベクトルとすると分離過程は、

$$Y(f, k) = W(f, k) X(f, k) \quad (4)$$

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、 $W(f, k)$  の行が入れ替わっても、 $Y(f, k)$  の独立性は保たれるので、任意の周波数  $f_p$  と  $f_q$  において、 $Y_n(f, k)$  と  $Y_m(f, k)$  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、 $W(f, k)$  の行の並びを  $Y_n(f, k)$  と  $Y_m(f, k)$  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。

### 2.3. 従来のパーミュテーション問題の解決法

パーミュテーション問題の解決法は、これまでいくつか考えられている。一つは、同じ音源である場合には近傍の周波数での周波数特性のエンベロープは高い相関を持つという性質を利用する方法がある[4]。しかし、この方法では、ある周波数での誤りが以降の推定を失敗させるため、安定性に欠ける。

他の解決法として、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べ替えを行う方法である[5]。しかし、この方法では低周波数域で指向性が鈍くなり、死角が顕著に形成されるので、方向推定精度が低くなってしまう。

そこで、方向推定と相関の方法を統合し、それぞれの欠点を補う手法が提案されている[6]。この方法は、各周波数ビンで指向特性を求めた後、確実に推定できる周波数ビンで指向特性を求めた後、その確実に推定できる周波数から順に並べ方を決めていく。そして、未解決の周波数に対しては、パーミュテーションが解決されている周波数と周波数差が小さく、相関の和が最大になるように決定していく。しかし、この方法でも間違いを完全に排除することは難しく、分離行列が周波数によって大きく異なる場合は困難であり、残響の強い場合には低周波数域で不安定になってしまう。

これらのどの方法でも本質的に並べ替えに失敗する可能性を含んでいる。

### 3. Permutation-Free ICA(PF-ICA)

パーミュテーション問題を解決するために、Di Persia らは周波数ピンの入れ替わりが本質的に発生しない PF-ICA を提案した[2][3].

受信信号に対して短時間フーリエ変換を行った後、従来法では周波数ビン毎に独立に分離が行われるが、PF-ICA では、時間-周波数領域における受信信号の各周波数ピンを、順序はマイク間で統一がとれてさえいればどうでも構わないが、特に理由がなければ  $f_1, f_2, \dots, f_n$  の順に連結して、全周波数ピンを一本のベクトルとして扱うものであった。この様子を図1に示す。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、全ての周波数ピ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共通の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特性の方向による違いを考慮することができず、不均一な音場ではいい分離結果が得られない。

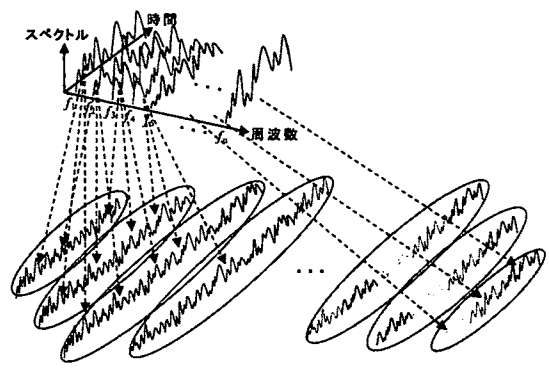


図2. 隣接周波数ピンを連結したベクトル

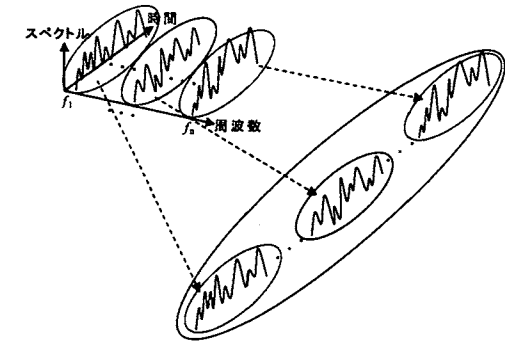


図1. 1本のマイクへの混合信号の短時間スペクトルの時系列を各周波数ビン毎の時系列とみて全周波数ビンの時系列を連結したベクトル

### 4. 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、周波数特性の方向による違いを考慮できる MB-ICA を提案する。提案法は、PF-ICA と同様に、周波数ピ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 のように全周波数ピンを連結して一本のベクトルとして扱うのではなく、局所的に、具体的には対象とする周波数ビンの前後  $k$  本の周波数ピンを連結し、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。対象とする周波数ビンは1本ずつシフトする。この様子を図2に示す。

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分離結果を取り出し、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、パーミュテーション誤りはほぼ生じない。平滑化の程度は  $k$  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。

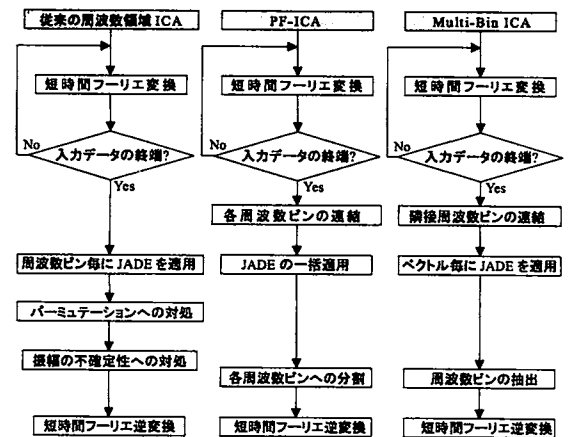


図3. 各手法のアルゴリズムの比較

## 5. MultiBin ICA の評価

### 5.1. 音場環境

#### (a) 収録環境 1

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。今回使用した配置は P1-P2, P1-P3, P1-P5, P1-P8, P3-P10, P8-P10, P2-P9,

P4-P6, P5-7, P7-10 の 10 通りである。マイクロホンには素子間隔 5cm の 2 素子アレイを用いた。

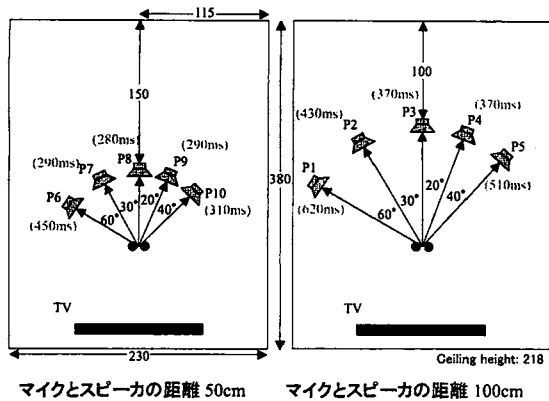


図 4. 収録環境 1

図 4 の()内の数値はその配置における残響時間。残響時間の計測は Time-Stretched Pulse の 20dB 減衰を 3 倍して求めた。

(b) 収録環境 2

また、MB-ICA は反射などによる周波数特性が方向によって異なる環境での効果を期待したものであるが、図 4 の収録環境では壁面材料が同一であるので反射による周波数特性が方向によってそれほど変わらない可能性があった。そこで、PF-ICA と MB-ICA の性能比較を明確にするために、強制的に反射による周波数特性が方向によって異なるような収録環境を図 4 と同じ部屋を使い作った。具体的には吸音材やベニヤ板、大きな楽器などを設置し音声の収録を行った。音源（スピーカー）と観測点（マイクロホン）の配置を図 5 に示す。マイクロホンには素子間隔 5cm の 2 素子アレイを用いた。

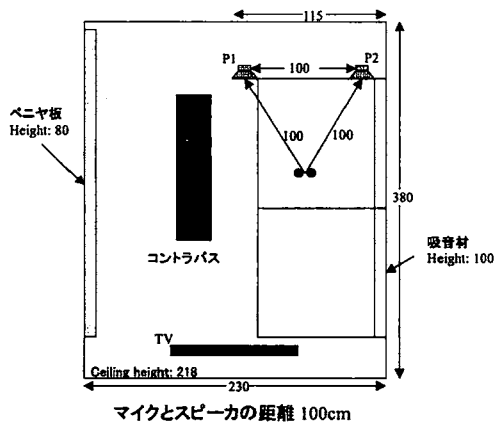


図 5. 収録環境 2

5.2. 音源信号

音源信号は、男性 1 名、女性 2 名について各 10 種の音声のデータ(1~2 秒)である。分離に使う混合音は、各配置から各音声を個別に録音し、SNR の調整を行うことなく収録した受信信号をそのまま加算して作った。組み合わせは、男性+女性 1(10×10×2×1)、男性×女性 2(10×10×2×1)で、一つの配置に対し計 200 個の混合信号を調査対象とした。

5.3. 評価条件

図 3 で示した 3 種類のアルゴリズム間による分離結果の比較を行う。MB-ICA では、前後に連結する周波数ビンを 1,2,4,8,16,32,64,128 本とした。従来法のアルゴリズムとして、JADE により周波数特性 ICA を行い、相関を用いてパーミュテーション問題の解決を行う Ikeda et al.[4]による方法を用いた。

評価指標については、以下の式(5)で定義した SNR を用いた。

$$SNR = \frac{\int |S(\omega)|^2 d\omega}{\int |S(\omega) - \hat{S}(\omega)|^2 d\omega} \quad (5)$$

ここで  $S(\omega)$ 、 $\hat{S}(\omega)$  はそれぞれ音源信号  $\hat{s}(t)$  と分離信号  $\hat{s}(t)$  のフーリエ変換。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	16000 samples/sec
量子化ビット数	16 bits
フレームサイズ	1024 点
フレームシフト	512 点
分析窓	ハミング窓

5.4. 評価結果

収録環境 1 における各配置の SNR を図 6, 図 7, 図 8 に、今回の収録で使った全配置の SNR の平均を図 9 に示す。スピーカーとマイクロホンの距離の違いで図を 3 つに分けた。図 6 のスピーカーの配置が P8-10 の組合せのとき以外 PF-ICA, MB-ICA のいずれも SNR が向上した。PF-ICA と MB-ICA では配置によっては SNR の向上があったが、全ての配置において SNR が向上したわけではなかった。

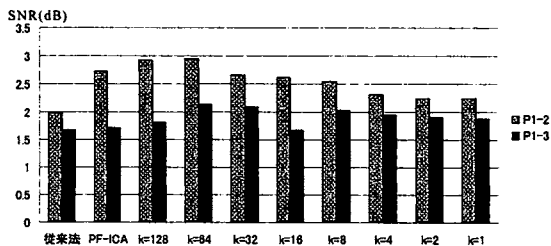


図 6. P1-2 と P1-3 の SNR の平均

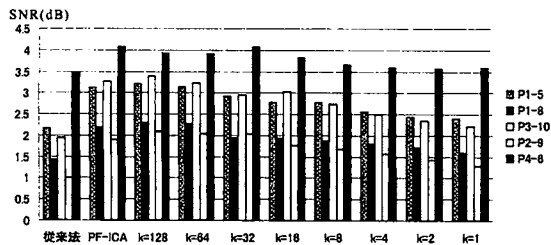


図 7. P1-5, P1-8, P3-10, P2-9, P4-6 の SNR の平均

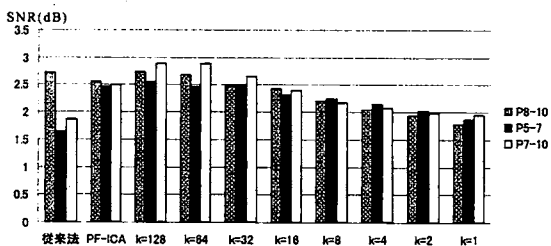


図 8. P8-10, P5-7, P7-10 の SNR の平均

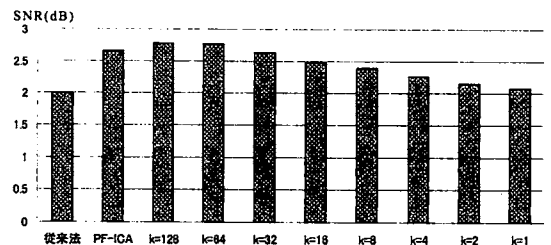


図 9. 全配置における SNR の平均

全ての配置の SNR の平均は従来法が 1.99dB, PF-ICA が 2.65dB, MB-ICA は 2.78dB ( $k=128$ ) であった。従来法と PF-ICA, 従来法と MB-ICA では両側  $t$  検定による有意水準 5% で差があったが, PF-ICA と MB-ICA では有意差がなかった。

また, 収録環境 2 における女性 1+男性, 女性 2+男性, 前 2 つの SNR の平均をそれぞれ図 10, 図 11, 図 12 に示した。いずれも従来法よりも PF-ICA の SNR が

低くなっている。また条件によっては Multi-Bin ICA の SNR に改善が見られた。

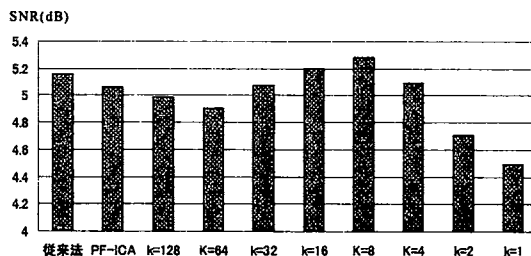


図 10. 女性 1 + 男性の SNR の平均

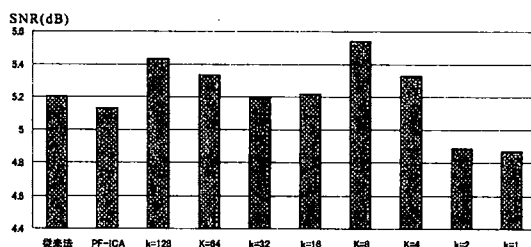


図 11. 女性 2 + 男性の SNR の平均

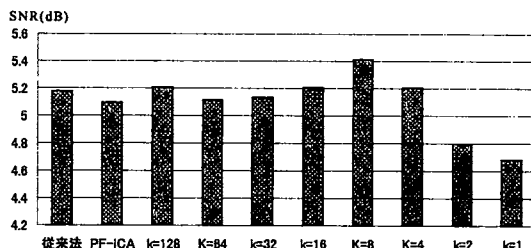


図 12. 女性 1+男性と女性 2+男性の SNR の平均

図 12 で示した平均での従来法の SNR は 5.18dB, PF-ICA の SNR は 5.09dB, Multi-Bin ICA の最高の SNR は 5.41dB であった。両側  $t$  検定による有意水準 5% で従来法と PF-ICA では有意差がなかったが, 従来法と MB-ICA では有意差があった。また PF-ICA と MB-ICA でも有意水準 5% で差があった。

## 6. 考察

収録環境 1 での評価結果では PF-ICA と MB-ICA の分離精度に大きな差はなかった。これは, 予想した通り図 4 の収録環境では反射による周波数特性が方向によってあまり変わらなかったからではないかと考えられる。そのため PF-ICA の結果と MB-ICA での  $k$  の値が大きいきの SNR に差がなく,  $k$  の値が小さくなるにつれ SNR が悪くなったと考えられる。図 6~図 8 を

比べると各配置において多少の傾向の違いはあるが、ほとんどの配置で図9と同じような傾向になっている。つまり図4の環境ではPF-ICAが一番安定して分離できるということが分かった。

本来パーミュテーション問題を完全に解決できれば従来法の分離精度が一番よくなるはずである。しかし、実際は完全にパーミュテーション問題を解決することができない。一般に、図4の収録環境1のように周波数特性が方向によって変わらない環境の場合は、PF-ICAの分離精度がよくなる。

しかし、実際の環境では周波数特性が方向によって変わらないことは少ない。収録環境2の評価結果をみると図12でPF-ICAのSNRが従来法よりも低くなっている。ゆえに、収録環境2では周波数特性が方向によって異なり、分離行列が周波数依存をもつこと予想される。

図10～図12でMB-ICAは $k=8$ のときにSNRが最大で、従来法とPF-ICAよりもSNRが高くなっている。これは収録環境2では周波数特性が方向によって大きく違っていたため周波数特性が方向によって変わらなないと考えるPF-ICAでは分離精度が落ち、また従来法による分離では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、MB-ICAでは周波数ピンをオーバーラップさせながらベクトルを1本ずつ形成していくため、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たような特性になる。ゆえに、MB-ICAでは周波数特性を考慮することができ、パーミュテーション誤りを上手く抑えられたということである。

またMB-ICAで $k=8$ より大きい値のSNRの値が下がったのは上で述べたように周波数特性が方向によって大きく異なったためであり、 $k=8$ よりも小さい値では逆にパーミュテーション誤りが起こりやすくなりSNRが下がったと考えられる。MB-ICAは条件を適切に設定すれば、従来法やPF-ICAと比較して高精度の分離ができると考えられる。

## 7. 今後の課題

今回の図5の収録環境2ではMB-ICAの有効性が確認できた。しかし、図10と図11を比べるとSNRが一番高くなる $k$ の値は同じであるがそれより大きい $k$ についてのSNRの傾向は違った。このことから、本方法の動作は混合信号の性質に依存しているかもしれない。そこで、今後はデータ数を増強し評価を行う必要がある。また、妨害音として音声だけではなく、定常雑音や非定常雑音が妨害音となった場合等でも評価を行う予定である。また、今回は従来法として相関を用いてパーミュテーション問題に対処したが、分離行列の方

向性を考慮する方法を比較対象とすることも検討する。

今回はフレーム長を1024点で固定したが、512点や256点等と変更したときの動作評価、分離後の信号に対してフィルタなどをかけることによって分離精度の高精度化なども検討したい。

MB-ICAで今回は $k=8$ のときにSNRが最高の値となったが、状況が変わればその値が最適とは限らない。実際収録環境1の結果では $k=128$ のときにSNRが最高となっている。このことから部屋の反射などによる周波数特性によってMB-ICAの最適な $k$ の値が変わることである。この最適な $k$ の値をどのようにして決めるかがMB-ICAの今後の課題である。

また、MB-ICAで周波数ピンに対して重み付けを行い、分離対象とする周波数ピンに大きな重みを付け、離れるにしたがって小さな重みをつける。そうすることにより $k$ の値を大きくしてもMB-ICAより周波数特性を考慮でき、かつパーミュテーションも起こりにくくなるのでMB-ICAよりも高精度な分離ができると考えられる。

今回はMB-ICAの分離行列の並べえに、PF-ICAを前処理とした並べ替えを行ったが、相関によるパーミュテーション問題の解決を行う方法など、従来法と組み合わせることも検討したい。

本研究の一部は、同志社大学学術フロンティア推進事業、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] A.Hyvärinen, "Independent Component Analysis", John Wiley, New York, 2001.
- [2] レアンドロ・ディ・ペルシア, "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.
- [3] レアンドロ・ディ・ペルシア, "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.
- [4] T.W.Lee: "Independent Component Analysis", Kluwer, 1998.
- [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ト音源分離における permutation 問題の解法", 音響学会講演論文集, pp. 541-542, Sep, 2002.