

sinc() Research Center for Signals, Systems and Computational Intelligence (fich.uml.edu.ar/sinc)  
Kenko Ohta, Tadashige Noguchi, Kohei Yasui, L. Di Persia & Masuzo Yanagida; "Frequency Domain ICA Connecting of Adjacent Frequency Bins"  
IEICE Transactions on Information and Systems (Japanese Edition), Vol. J91-D, No. 1, pp. 130-135, 2008.

THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

**IEICE** **電子情報通信学会**  
**D** **論文誌** VOL.J91-D NO.1 JANUARY  
情報・システム 2008

情報・システムソサイエティ

社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

## 研究速報

## 近傍周波数ビンの連結による周波数領域 ICA の提案

大田 健紘<sup>†,††</sup> 野口 忠繁<sup>†</sup>  
 安井 耕平<sup>†</sup> レアンドロ デイ ベルシア<sup>†††</sup>  
 柳田 益造<sup>†</sup> (正員)

Frequency Domain ICA Connecting of Adjacent Frequency Bins  
 Kenko OTA<sup>†,††</sup>, Tadashige NOGUCHI<sup>†</sup>, Kohei YASUI<sup>†</sup>,  
 Leandro DI PERSIA<sup>†††</sup>, *Nonmembers*,  
 and Masuzo YANAGIDA<sup>†</sup>, *Member*

<sup>†</sup>同志社大学, 京田辺市  
 Doshisha University, Kyotanabe-shi, 610-0321 Japan  
<sup>††</sup>リール中央学院, France  
 École Centrale de Lille, Villeneuve d'Ascq, 59651  
<sup>†††</sup>エントレリオス国立大学, Argentina  
 Universidad Nacional de Entre Rios, Parana, 3100

あらまし 本論文では、周波数領域 ICA を行う際に問題となるパーミューテーション問題を回避し、高精度に分離を行う Multi-Bin ICA を提案する。Multi-Bin ICA は混合行列をもと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つ一つのグループとして扱うことにより、パーミューテーションの問題を抑えながら、周波数ごとに適切な分離行列を求めることができる。

キーワード 独立成分分析, パーミューテーション問題, ブラインド音源分離, 周波数領域

## 1. まえがき

近年、複数の信号が複数のマイクロホンへ入力されたときに、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、ブラインド音源分離 (BSS: Blind Source Separation) に研究の興味が集まっている。BSS の解法の一つとして、独立成分分析 (ICA: Independent Component Analysis) [1], [2] に基づく手法が盛んに研究されており、瞬時 (畳込みでない) 混合に対しては十分な分離性能が得られている。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、周波数ビンごとの瞬時混合として 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、各チャネルの周波数ビンの入れ違いを正しく並べ換える (パーミューテーション問題を解決する) 必要がある。パーミュー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 [4]、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 [5]、更には音声の調波構造を用いることにより文献 [5] の改善を行った方法 [6]、及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 [7] などが提案されている。

しかし、いまだパーミューテーション問題の完全な解決には至っていない。そこで 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 [8], [9]。この手法では、分離された信号の周波数ビン間の入れ換わりが生じない利点があるが、混合系 (音源-マイクロホン間) 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。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ーテーション問題を抑えながら、混合系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。

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。

パーミューテーションを回避する手法として関連する研究は、Kim et al. [10] によるものがある。Kim et al. は、分離行列を学習するためのコスト関数を多変量関数に拡張することにより、パーミューテーションを回避している。また近隣の周波数ビンを同時処理する手法としては、Robledo-Arnuncio et al. [11] によるものがある。Robledo-Arnuncio et al. による研究は、分離行列を学習するためのデータが少ない場合 (発話時間が短い場合) に分離行列の学習精度が低下するのを回避するために近隣周波数ビンの同時処理を行っており、パーミューテーションに関しては [6] を用いて対処している。

## 2. 従来の周波数領域 ICA

周波数領域 ICA は受信信号の第  $k$  フレームの短時間フーリエ変換により得られた  $M$  本のマイクロホンによる観測信号群ベクトル  $\mathbf{X}(f_n, k) = [X_1(f_n, k), X_2(f_n, k), \dots, X_M(f_n, k)]^T$  に対して、周波数ビン  $f_n$  において学習された分離行列  $W(f_n)$  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 [12] を用いている。ここで  $P$  個の音源信号群ベクトルを  $\mathbf{S}(f_n, k) = [S_1(f_n, k), S_2(f_n, k), \dots, S_P(f_n, k)]^T$  とすると、音源信号の混合過程は

$$X(f_n, k) = H(f_n)S(f_n, k)$$

と表される。ただし、 $H(f_n)$  は混合系の周波数特性である。また、 $Y(f_n, k) = [Y_1(f_n, k), Y_2(f_n, k), \dots, Y_P(f_n, k)]^T$  を  $P$  個に分離された音源信号群のスペクトルとすると分離過程は、

$$Y(f_n, k) = W(f_n)X(f_n, k)$$

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より分離を行うので、 $W(f_n)$  の各行が入れ換わったとしても、 $Y(f_n, k)$  の独立性が最大になることに変わりはない。したがってこのとき、 $Y(f_n, k)$  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換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換わった分離信号の音源番号を整理させる問題である。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、本質的に並べ換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べ換えを行う方法がある [4]。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ことが指摘されている [5]。

一方、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も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 [5]。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長い場合に不安定になることが報告されている [6]。残響が長い場合に不安定になる原因として、低周波数域での方向推定の困難さが挙げられている [6]。

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 [6]。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もっており、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換えに失敗するという問題を依然として含んでいる。

## 4. パーミュテーション問題を回避する周波数領域 ICA

### 4.1 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 らは周波数ビンの入れ換わりが発生しない PF-ICA を提案した [8], [9]。この手法は、時間-周波数領域における受信信号のマイクロホン  $m$  における周波数ビン  $f_n$  のフレーム時系列

$$X_m(f_n) = [X_m(f_n, 1), X_m(f_n, 2), \dots, X_m(f_n, T)]$$

を全周波数ビンについて連結し、

$$X_m = [X_m(f_1), X_m(f_2), \dots, X_m(f_N)]^T$$

とする。ただし  $N$  はフレーム長の半分 +1 であり、 $T$  はフレーム数である。この処理を図 1 に示す。図 1 の上の図は時間-周波数領域上での各周波数ビンを表しており、それらを下の図のように全周波数ビンを連結し、マイクロホンごとに 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する。

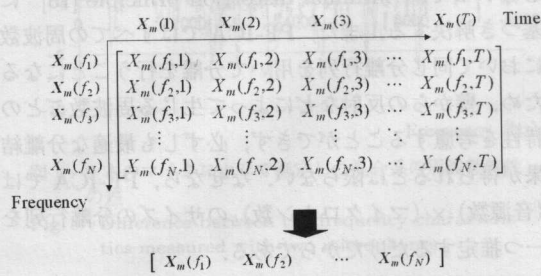


図 1 PF-ICA のベクトル構成  
 Fig. 1 Vector structure of PF-ICA.



$$Y = WX$$

ただし、 $X = [X_1, X_2, \dots, X_M]^T$  であり、 $Y$ 、 $W$  及び  $X$  それぞれの行列のサイズは、 $P \times TN$ 、 $P \times M$  及び  $M \times TN$  である。JADE を用いて分離行列  $W$  を得るには、まず  $X$  の共分散  $R_x$  ( $M \times M$  の行列) を求め、白色化行列  $B$  ( $P \times M$  の行列) を得る。次に白色化した系列  $Z = BX$  の尖度  $K$  ( $P^2 \times P^2$  の行列) を求め、 $K$  の固有値・固有ベクトルの組のうち固有値の大きい方から  $P$  個の組  $\{\lambda_r, D_r | 1 \leq r \leq P\}$  を選ぶ。ただし、 $\lambda_r$  は固有値、 $D_r$  は固有ベクトル ( $P^2 \times 1$ ) を  $P$  個の成分ずつ列に並べた  $P \times P$  の行列を表す。そして、独立性の基準として白色化系列  $Z$  ( $P \times TN$ ) を行列  $V$  ( $P \times P$ ) により分離した  $A = VZ$  の4次のクロススキュムラント

$$c(V) = \sum_{i,j,l=1 \dots P} |\text{Cum}(A_i, A_i^*, A_j, A_l^*)|^2$$

を定義し、 $c(V)$  を最小化する  $V$  を見つける。ただし  $(\cdot)^*$  は共役を表し、 $A$  のインデックスは音源信号の ID に対応している。JADE では、直接  $c(V)$  の最小化を行わず、同時対角化により最適な  $V$  を推定する。具体的には、 $\mathcal{N} = \{N_{\nabla} = \lambda_{\nabla} D_{\nabla} | 1 \leq \nabla \leq P\}$  で表される  $P$  個の  $P \times P$  の行列の組を定義し、

$$c(V, \mathcal{N}) = \sum_{\nabla=1 \dots P} |\text{diag}(V^H N_{\nabla} V)|^2$$

を最小化する  $V$  を Givens 変換により求める。ただし、 $|\text{diag}(\cdot)|$  は行列の対角成分のノルムを表し、 $(\cdot)^H$  は複素共役転置を表す。そして分離行列  $W$  は  $W = VB$  により求める。

PF-ICA では、すべ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換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、振幅の不確定性は解決しないので、“Minimal distortion principle [13]” に基づき解決する。また、PF-ICA ではすべ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よるため、壁からの反射などによって生じる周波数ごと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 (音源数)  $\times$  (マイクロホン数) のサイズの分離行列を一つ推定するだけだからである。

#### 4.2 Multi-Bin ICA (MB-ICA)

本論文では、パーミュテーション問題を抑えつつ、

混合系の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1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。周波数ビン  $f_n$  のフレーム時系列  $X_m(f_n)$  を連結し、分離対象のベクトルを再定義する。

$$X_m(f_1) = [X_m(f_1), X_m(f_2), \dots, X_m(f_{k+1})]^T$$

$$X_m(f_2) = [X_m(f_1), X_m(f_2), \dots, X_m(f_{k+2})]^T$$

...

$$X_m(f_{k+1}) = [X_m(f_1), \dots, X_m(f_{2k+1})]^T$$

...

$$X_m(f_N) = [X_m(f_{N-k-1}), \dots, X_m(f_N)]^T$$

ただし、 $X_m(f_{k+1})$  は分離対象の周波数が  $f_{k+1}$  のベクトルを表す。この処理を図2に示す。図2の上の図は時間-周波数領域上での各周波数ビンを表しており、そして下の図は上から順に周波数  $f_1, f_2, \dots, f_N$  に対して生成されたベクトルを表している。そして、このベクトルごとに分離を行い、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。

$$Y(f_n) = W(f_n)X(f_n)$$

ただし、 $X(f_n) = [X_1(f_n), X_2(f_n), \dots, X_M(f_n)]^T$

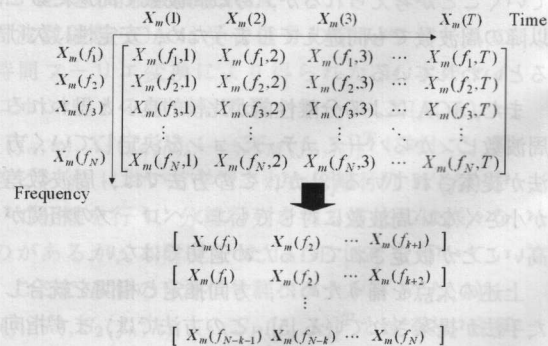


図2 MB-ICA のベクトル構成  
Fig. 2 Vector structure of MB-ICA.

提案法を連続的に接続し、隣接する周波数ビン間の周波数特性の差を小さくする。また、周波数特性の差を小さくするために、隣接する周波数ビン間の周波数特性の差を小さくする。また、周波数特性の差を小さくするために、隣接する周波数ビン間の周波数特性の差を小さくする。

であり、分離対象の周波数ビンの前後に  $k$  本の周波数ビンを連結した場合  $\mathbf{Y}(f_n)$ ,  $\mathbf{W}(f_n)$  及び  $\mathbf{X}(f_n)$  それぞれの行列サイズは、 $P \times T(2k + 1)$ ,  $P \times M$  及び  $M \times T(2k + 1)$  である。対象とする周波数ビンごとに PF-ICA と同様 JADE を適用し分離を実行すると、従来法のように音源番号の並べ換えが必要となるが、本手法では、オーバーラップさせながら形成した分離対象ベクトルを用いて分離行列を学習するので、隣接する分離対象周波数ビン間での分離行列は大きな変化をすることなく、滑らかに変化すると考えられるため、その困難は生じにくい。連結周波数ビン数  $k$  を変えることにより、隣接する分離対象周波数ビン間での分離行列の変化の程度を制御できる。提案法では、各周波数ビンでの分離行列の学習は独立に行っており、ある周波数  $f_k$  の分離に用いた分離行列を、異なる周波数  $f_{k+1}$  の分離行列の学習の初期値に用いることは行っていない。

5. 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（残響時間<sup>(注1)</sup>：約 370 ms）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 3、図 4 に示す。マイクロホン間隔は 5 cm である。図 3 の環境は方向により混合系の周波数特性の違いが小さい環境であり、図 4 の環境は方向により混合系の周波数特性が大きく異なる環境を想定している。

環境 2 において方向により混合系の周波数特性が大きく異なることを確認するために、それぞれの環境で周波数特性の測定を行った。環境 1 に関しては、スピーカを P3 の位置に設置し、環境 2 に関しては、スピーカをマイクロホンから垂直の方向に 100cm 離れた位置にスピーカを設置した。もし方向により混合系の周波数特性が大きく異なる場合は、2 本のマイクロ

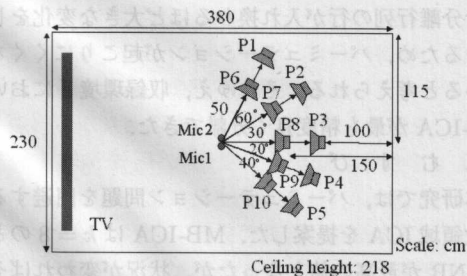


図 3 収録環境 1 (方向により混合系の周波数特性の違いが小さい)

Fig. 3 Environment 1. (Frequency characteristics of directionality are almost same)

ホンで測定した混合系の周波数特性の違いが大きくなると考えられる。図 5 に環境 1 及び環境 2 で測定した混合系の周波数特性の 2 本のマイクロホン間での差を示す。混合系の周波数特性の差は以下の式により計算した。

$$\Delta \hat{H}(f_n) = 10 \log \left( \frac{|\hat{H}_1(f_n)|}{|\hat{H}_2(f_n)|} \right) \quad (1)$$

ただし、 $\hat{H}_1(f_n)$  はマイクロホン 1,  $\hat{H}_2(f_n)$  はマイクロホン 2 で測定した混合系の周波数特性を表す。

図 5 より環境 2 において混合系の周波数特性が方向により大きく異なることが分かる。ただし、位相特性に関してはマイクロホン間での信号の到達時間差により周波数依存性をもつため、環境 1 のように方向によ

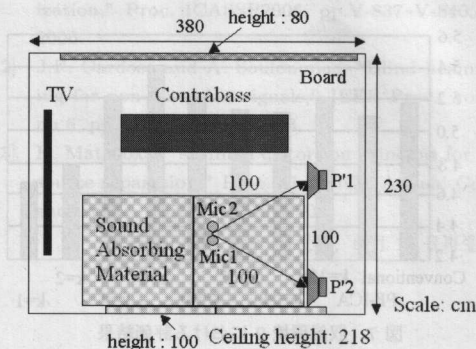


図 4 収録環境 2 (方向により混合系の周波数特性が異なる)

Fig. 4 Environment 2. (Frequency characteristics of directionality are forced to be different by putting some objects)

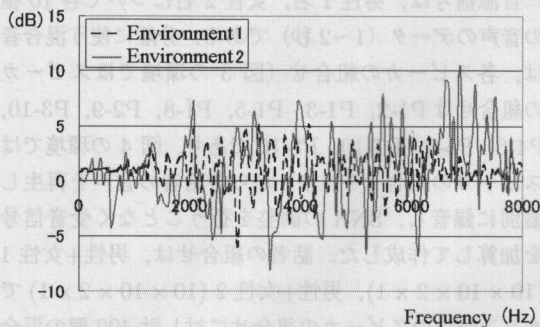


図 5 2 本のマイクロホンで測定した混合系の周波数特性の差

Fig. 5 Difference between the frequency characteristics measured with two microphones.

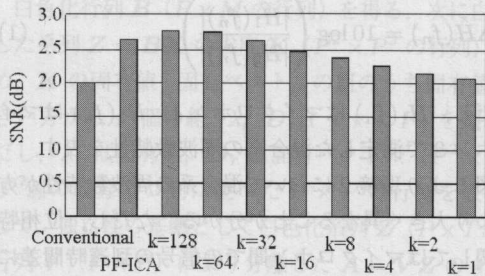
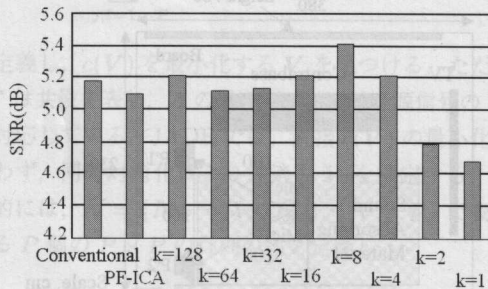
(注1)：残響時間は Time-Stretched Pulse の 20 dB 減衰を 3 倍して求めた。



表 1 音響分析条件

Table 1 Conditions of acoustic analysis.

Sampling Rate	16000 samples/s
Frame Size	1024
Frame Shift	512
Window	Hamming

図 6 収録環境 1 における評価結果  
Fig. 6 SNR in the environment 1.図 7 収録環境 2 における評価結果  
Fig. 7 SNR in the environment 2.

り混合系の周波数特性の違いが小さい環境であっても位相特性の違いも小さいとはいえない。

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、各スピーカの組合せ (図 3 の環境ではスピーカの組合せは P1-2, P1-3, P1-5, P1-8, P2-9, P3-10, P4-6, P5-7, P7-10, P8-10 であり、図 4 の環境ではスピーカの組合せ P'1-2) から各話者の音声を再生し個別に録音し、SNR の調整を行うことなく受信信号を加算して作成した。話者の組合せは、男性+女性 1 ( $10 \times 10 \times 2 \times 1$ )、男性+女性 2 ( $10 \times 10 \times 2 \times 1$ ) であり、一つのスピーカの組合せに対し計 400 個の混合信号を調査対象とした。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 により周波数領域 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 Murata et al. [3] による方法を用い

た。評価指標には以下に定義するセグメンタル SNR を用いた。

$$SNR_i(t) = \frac{\sum_{n=0}^{N-1} |S_i(f_n, t)|^2}{\sum_{n=0}^{N-1} |S_i(f_n, t) - \hat{Y}_i(f_n, t)|^2}$$

表 1 に音響分析条件を示す。

図 6, 図 7 は図 3, 図 4 における SNR による評価結果を示している。

## 6. 考 察

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は最も良い、若しくは少なくとも PF-ICA より良い結果になるはずである。しかし、実際に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。それゆえ一般に、収録環境 1 のように混合系の周波数特性の違いが方向により小さい環境の場合は、PF-ICA の分離精度が良くなる。しかし、実際の環境では混合系の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 の評価結果では PF-ICA の SNR が従来法よりも低くなっている。また 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 では混合系の周波数特性が方向によって異なっていることから、混合系の周波数特性の方向による違いを扱えない PF-ICA では分離精度が落ちる。一方で、MB-ICA は全周波数ビンを一括して処理するのではなく、ある一定の周波数帯域ごとに処理を行っているため、混合系の周波数特性の方向による違いに対処することができる。更に、分離対象ベクトルをオーバーラップさせながら形成し、このベクトルから分離行列を学習するため、隣接する分離対象ベクトル間で学習される分離行列の違いが小さくなる。つまり、隣接する周波数ビン間で分離行列の行が入れ換わるほど大きな変化をしなくなるため、パーミュテーションが起こりにくくなっていると考えられる。それゆえ、収録環境 2 において MB-ICA が最も精度良く分離できた。

## 7. む す び

本研究では、パーミュテーション問題を回避する周波数領域 ICA を提案した。MB-ICA 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。収録環境 1 の結果は  $k=128$  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混合系の周波数特性によって提案

