

パーミュテーションフリーICAの動作解析

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

[†] エントレリオス国立大学 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, [‡] (dtg0731,etf1704)@mail4.doshisha.ac.jp myanagid@mail.doshisha.ac.jp

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けるパーミュテーション問題を回避したパーミュテーションフリーICAの動作解析を行っている。これまで、パーミュテーション問題の解決法には、各周波数の指向特性、分離信号の周波数間の相関、音声信号の調波性に基ついた方法などが提案されているが、完全には解決されていない。これに対してパーミュテーションフリーICAは、全周波数ビンをもつベクトルとして分離処理を行っているため、パーミュテーション問題は発生しない。このパーミュテーションフリーICAは、短時間の音声や非定常雑音にも効果があることがわかった。しかし、パーミュテーションフリーICAでは、全周波数ビンで同じ指向性を持つ分離行列になってしまう、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る。本研究では、その問題への対策を検討している。

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

Performance of Permutation-Free ICA

Leandro Di Persia[†] Tadashige NOGUCHI[‡] 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, [‡] (dtg0731,etf1704)@mail4.d0.shisha.ac.jp myanagid@mail.doshisha.ac.jp

Abstract Performance evaluation of Permutation-Free ICA for blind source separation is presented. Permutation-Free ICA was proposed as a method that could avoid so-called "Permutation problem" that appears in frequency-domain ICA from the necessity to align components of each source. Up to now, several approaches for the permutation problem have been proposed, for example those based on directivity patterns of each frequency bin, those on correlation between adjacent frequency bins and those based on harmonic structure of source sounds. These methods, however, do not give the perfect solution to the permutation problem. On the other hand, Permutation-Free ICA carries out separation on a set of long vectors consisting of connected temporal changes of frequency components of the received signal. Investigating the performance of Permutation-Free ICA, it appears that the method is effective for cases where short-time speech or nonstationary noise are involved. The separation matrix obtained in the Permutation-Free ICA, however, has a common directivity pattern for all frequency. To make Permutation-Free ICA effective for real situations, some means that can find the optimum separation matrix for each frequency bin referring to the results of Permutation-Free ICA.

Keyword Blind Source Separation, Independent Component Analysis, Permutation

1. はじめに

近年、ブラインド音源分離 (BSS: Blind Source Separation) が種々の応用を目的として検討されている。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らないでも、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。最近では、BSSの手法の一つである独立成

分分析 (ICA: Independent Component Analysis) 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[1]。しかし、残響環境下では畳み込み混合となってしまう、このような場合については周波数領域ICAが用いられる[2,3]。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、周波数ビン毎の混合として

ICAを適用するものである。

しかし、周波数領域ICAを周波数ビン毎に適用した場合には、周波数ビン毎の分離チャンネルを揃える操作、つまり入れ違いがあれば、それを正しく並び替える操作の必要がある。これはパーミュテーション問題と呼ばれるが、これ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[4]、分離信号の各周波数ビンでの相関を用いる方法[1]、これらを統合した方法[5]、さらに音声の調波構造を用いることによりそれを改良した方法[6]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、我々は各周波数ビンで分離処理を行うのではなく、全周波数ビンを連結して1つのベクトルとして分離処理を行うパーミュテーションフリーICAを提案した[7]。

本報告では、まず、パーミュテーションフリーICAがどのような状況で有効であるのかを調査し、短時間の音声や非定常雑音に効果があることを示す。パーミュテーションフリーICAでは、分離行列が全周波数で同じ指向性を持つてしまうが、実環境では、反射面によって周波数特性が異なるなどの理由で、周波数毎に伝達特性が異なるため、パーミュテーションフリーICAの分離行列は最適ではないという問題がある。

そこで、パーミュテーションフリーICAの長所と短所を把握し、従来の手法にパーミュテーションフリーICAの長所を組み合わせることによって、パーミュテーション問題を解決し、分離精度の向上を図る手法を検討する。

2. 周波数領域ICAの問題

2.1 ブラインド音源分離 (BSS)

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。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

時間領域でのBSSを考えると、式(1)のような畳み込

み混合となるが、 $h_{mn}(t)$ 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、混合過程は各成分ごとに

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

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、単なる積の形になるため、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、畳み込み混合の場合は学習規則が複雑で、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学習された分離行列 $W(f, k)$ を用いて分離を行う。ここで、 $Y(f, k)$ 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。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 $Y_n(f, k)$ が必ずしも同じ音源に対応しているとはかぎらない。従って、 $W(f, k)$ の行の並びを $Y_n(f, k)$ と $Y_n(f, k)$ 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 周波数領域ICAとビームフォーマの関係

図1は、 $M=2$ の場合のマイクロホンアレイによるビームフォーミングを表している。複数の観測信号 $x_m(t)$ の位相差を利用して、ある方向 θ からの周波数 f の指向特性は

$$D(f, \theta) = \sum_m^M A_m \exp \left(-j \left(2\pi f \frac{d \sin \theta}{c} - \delta_m \right) \right) \quad (5)$$

となる．ここで， A_m は振幅， d はマイク間の距離， δ_m はマイク M_1 の波形から見たマイク M_m の位相差， c は音速を表している．

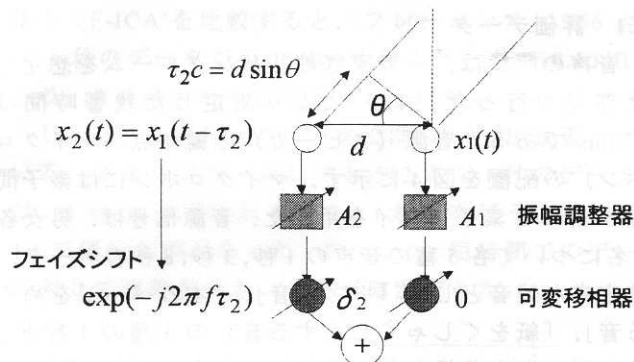


図 1. マイクロホンアレイによるビームフォーミング

周波数領域 ICA における分離行列 $W(f)$ は，マイクの本数 $M=2$ ，音源数 $N=2$ の場合は，

$$W(f) = \begin{bmatrix} W_{11}(f) & W_{12}(f) \\ W_{21}(f) & W_{22}(f) \end{bmatrix} \quad (6)$$

のように表すことができ，また，周波数領域で考えているので $W(f)$ の要素 $W_{nm}(f)$ は複素数になるため，各成分は振幅と位相に分けて

$$W_{nm}(f) = |W_{nm}| \exp(-j\angle W_{nm}(f)) \quad (7)$$

のように表すことができる．式(5)の A_m と式(7)の $|W_{nm}|$ は振幅，式(5)の δ_m と式(7)の $\exp(-j\angle W_{nm}(f))$ は位相を表していることより，分離という処理は，対象音源以外の音源方向に死角を持ったビームフォーマによる処理と等価であると考えられる．

2.4 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方法は，これまでいくつか考えられている．一つは，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である[4]．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ないで，方向推定精度が低くなってしまう．

他の解決法として，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[1]．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．

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[5]．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，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．

パーミュテーション問題の難しさの原因のひとつに，低周波数域での方向推定の困難さがある．この問題を解決するために，音声の調波構造を利用する方法が提案されている[6]．有声音に対しては，調波構造を用いることで，より確実にパーミュテーション問題を解決することができる．

しかし，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. パーミュテーションフリーICA

3.1 従来のICAの処理

従来の周波数領域 ICA は，周波数ビンごとの時間波形を持ってきて，分離操作を行う．図2の上部は，受信信号に対して分析窓を移動しながら短時間フーリエ変換(STFT)を行い，周波数スペクトルの時間変化を各周波数ビン毎の時間波形として見た様子を表している．従来の周波数領域 ICA の手法では，各周波数ビン（図2の上部の1つの丸で囲まれた部分）ごとに分離処理を行う．しかし，このように周波数ビンごとに処理を行えば，分離行列を求めた際に，行が入れ替わってしまい，パーミュテーション問題が発生する．そのため，分離行列を生成した後に，2.4 で説明したような方法でパーミュテーション問題の解決を図る．そして，各周波数ごとに振幅への不確定性の対処を行い，各音源に対する各分析フレーム時間の周波数スペクトルを得て，短時間フーリエ変換を行い，時間領域に戻す．これが従来の周波数領域 ICA の手法であり，この処理の様子を図3の右部に示す．

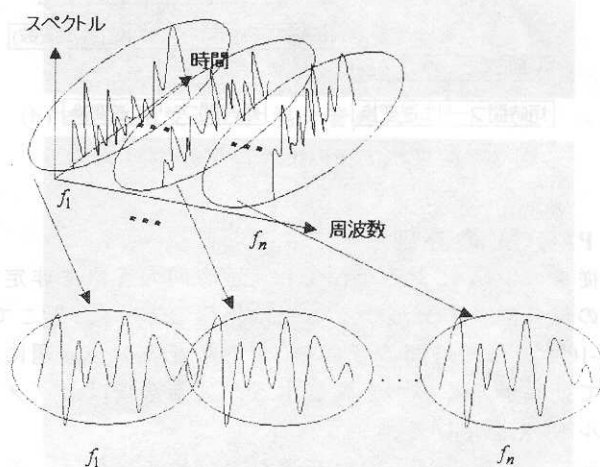


図 2 従来の周波数領域 ICA と PF-ICA の対応関係

3.2 パーミュテーションフリーICA (PF-ICA)

我々は本質的にパーミュテーションが発生しない周波数領域 BSS を提案した[7]. パーミュテーション問題は、分離行列を生成する際に発生するため、その前にパーミュテーションが起らないように処理を行う。

受信信号に対して STFT を行った後、従来法では各周波数ビン毎に独立に分離が行われるが、PF-ICA では全周波数ビンに対して一括して分離を行う。この方法では、全周波数について共通の指向特性を持つビームフォーマがあると仮定している。これは、2.3 で説明したように、全周波数について共通の大きさと遅延を持つ遅延和ビームフォーマを求めることと等価である。つまり、全周波数ビンにわたって共通のビームフォーマとして最適なものを見つけることになる。そこで、PF-ICA では、図 2 の下部のように、分割されている周波数ビンを連結して、1 つの長いベクトルとして、受信信号を一括して扱ってビームフォーマを求める。そのようにすれば、単一のビームフォーマが得られ、分離行列の行が周波数ビンによって入れ替わることはない。分離処理後は、各周波数ビンへの分割を行い、短時間フーリエ逆変換を行い、時間領域に戻す。この処理の様子を図 3 の左部に示す。従来法では、振幅に関する不確実性があったが、この方法では全体の振幅を相対的にどうするかが決まらないが、音源ごとに求まったあとで各音源について必要があれば調整すればよいだけで、個別の振幅については不確実性は生じない。

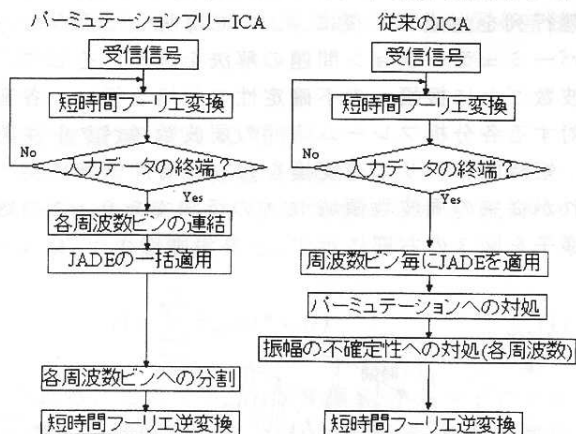


図 3 従来法と PF-ICA のアルゴリズム

4. PF-ICA の評価

従来の ICA による手法では、短時間の音声や非定常音の混合信号を分離することは困難であった。そこで、PF-ICA では短時間の音声や、非定常音の混合問題に対して効果があるのかを評価する。評価基準はセグメンタル SNR を用いる。

シミュレーションでは音源および雑音信号は、それぞれ個別に収録し、SNR の調整を行うことなく、受音

点で個別に収録した受信信号をそのまま加算して混合音を作る。

4.1 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 370ms 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。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 名について各 5 種の音声の 1 秒、3 秒、8 秒のデータ、非定常な雑音として「ドアの音」、「本のページをめくる音」、「紙をくしゃくしゃする音」の 3 種の 1 秒と 3 秒のデータを収録した。組み合わせは、女性 + 男性 (5 × 5 × 3)、女性 + 非定常雑音 (5 × 3 × 2)、男性 + 非定常雑音 (5 × 3 × 2) の計 135 個の混合信号を調査対象とした。

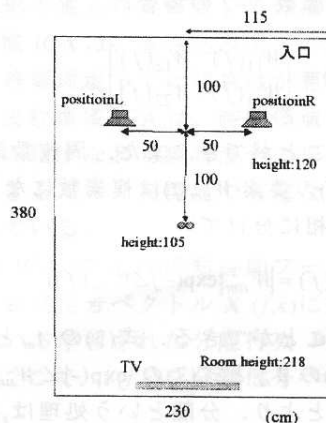


図 4 データ収録環境

4.2 評価条件

図 3 で示した 2 種類のアルゴリズム間で比較を行う。従来アルゴリズムは、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 Ikeda et al.[1]による方法である。

評価指標については、以下のように定義した SNR を用いる。

$$SNR = \frac{\int |S(\omega)|^2 d\omega}{\int |S(\omega) - \hat{S}(\omega)|^2 d\omega} \quad (8)$$

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 $\hat{s}(t)$ のフーリエ変換である。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	256.512 点
分析窓	ハミング窓

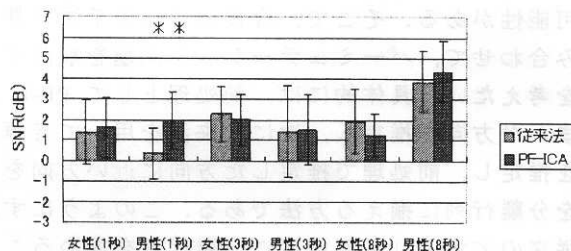
*TSP の 20dB 減衰時間を 3 倍して求めた

4.3 評価結果

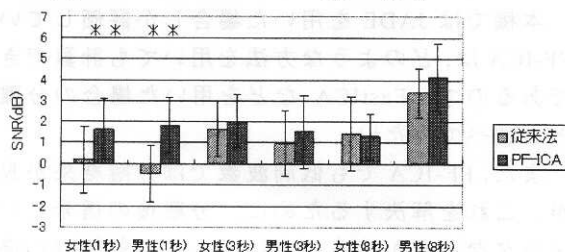
4.3.1 SNR による動作比較

まず、男性+女性の場合の SNR を図 5 に示す。従来法と PF-ICA を比較すると、フレームシフト 256 点では 3 秒のデータで -0.08dB、8 秒のデータで -0.10dB であり、2つの手法で有意差は見られない。

しかし、1 秒の音声データでは、0.93dB 改善されている。また、フレームシフト 512 点で 1 秒のデータでは、1.86dB 改善されており、t 検定を行うと優位水準 1% で優位な差があった。これより、短時間でのデータに対して改善が大きいことが確認できた。



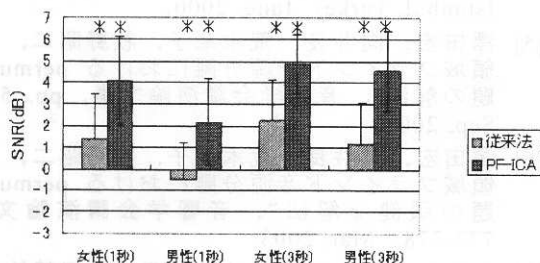
(a) フレームシフト 256 点の場合



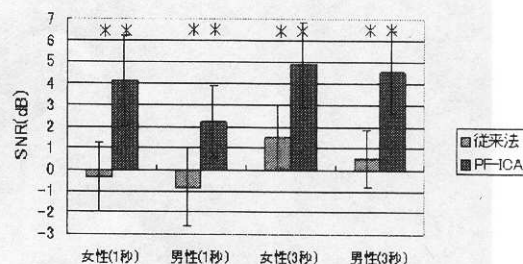
(b) フレームシフト 512 点の場合

図 5 音声+音声の混合に対する動作比較

次に、女性+非定常雑音、男性+非定常雑音の場合の SNR を図 6 に示す。フレームシフト 256 点の場合、1 秒では 2.68dB、3 秒では 3.00dB の改善が見られ、フレームシフト 512 点の場合、1 秒では 3.75dB、3 秒では 3.70dB の改善が見られた。t 検定を行うと 1 秒と 3 秒で共に優位水準 1% で優位な差があった。これより、音声と非定常雑音が混合された場合、PF-ICA を用いると従来法と比較して SNR が改善されることが確認できた。



(a) フレームシフト 256 点の場合

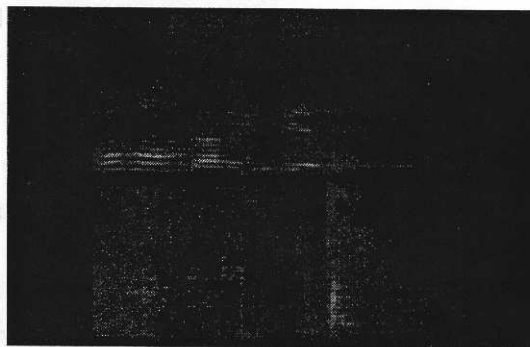


(b) フレームシフト 512 点の場合

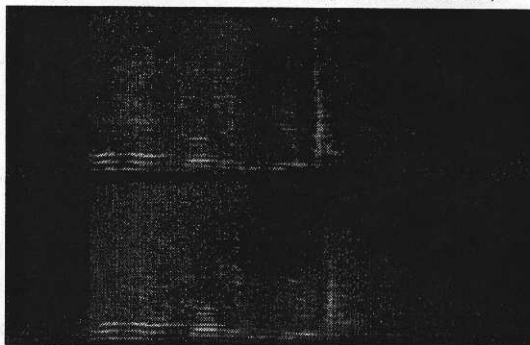
図 6 音声+雑音の混合に対する動作比較

4.3.3 スペクトログラムによる動作比較

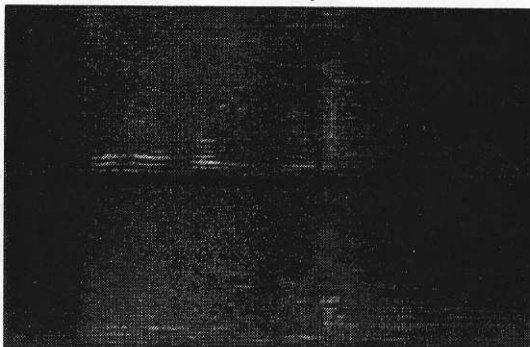
次に、どのように分離が行われているかをスペクトログラムで確認する。ここでは、「女性の音声」に「本のページをめくる音」が加わった場合の例を図 7 に示す。図 7(a)の上が音声で、下が雑音である。図 7(b)は受音点で受音した混合音であり、図 7(c)は従来法、(d)は PF-ICA の分離結果である。



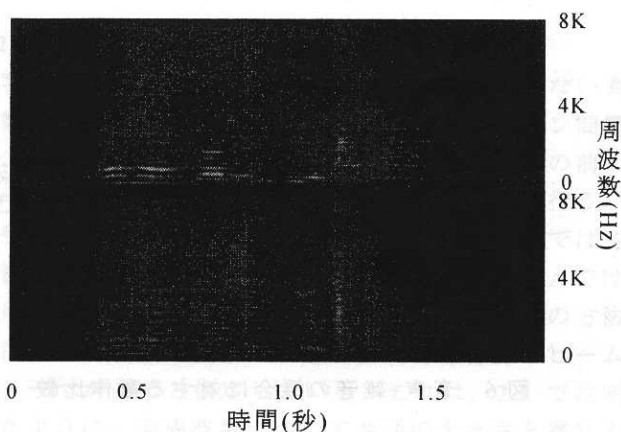
(a) 元の音データ(上は音声, 下は雑音)



(b) マイクで受音した混合音のスペクトログラム



(c) 従来法による分離結果のスペクトログラム



(d)PF-ICA による分離結果のスペクトログラム
図 7 音声データのスペクトログラム

図 7(c),(d)を見ると、共に上の方が調波構造の縞が強いことが確認できるので、上の図が音声の推定結果であることがわかる。そこで、図 7(c),(d)の下図を比較すると、図 7(c)の方が調波構造が強く残っているのがわかる。図 7(a)の下図の非定常雑音を見ると、調波構造はないので、図 7(d)の方が分離結果がよくなっていることがわかる。

5. 考察

PF-ICA は、短い音声に対して効果があることがわかった。これは、短い音声に対しては、従来法では 1 箇所の分離行列の行の入れ替えが全体に及ぼす影響が大きいためであると考えられる。また、音声短い場合には、従来法は相関を用いたパーミュテーションの解決がうまくいかなかったのではないかと考えられる。

また、音声に非定常雑音が加わった場合にも PF-ICA の効果があることがわかった。これは、非定常な音源が雑音として混入した場合、従来法ではパーミュテーションの問題が解決できないことが大きな原因となっていると考えられる。

目的音と雑音がともに音声の組み合わせの場合には、3 秒あるいは 8 秒のように 1 秒より十分長いデータでは、従来法と PF-ICA で SNR に大きな違いはなかった。従来法では完全にパーミュテーション問題を解決できていないにもかかわらず、同じ SNR になるということは、PF-ICA にも欠点があるということになる。この原因は、PF-ICA の指向特性にあると考えられる。実環境を考えた場合、壁からの反射に周波数特性があることを考えると、各周波数に対して精密に死角の方向を推定したほうが好ましいが、PF-ICA では全周波数で同じ方向に死角を持たせることになるため、従来法と比較して SNR が改善されない周波数があると考えられる。

6. 今後の課題

PF-ICA では、全周波数について共通の方向に指向特性を持つような分離行列を生成した。これによって、ある程度方向制御ができてはいるはずである。しかし、実環境では、一般には、各周波数ごとに反射特性が異なるため、周波数によって異なる方向に指向特性を持たせるのがよいはずである。従来法では、周波数特性を考慮したような指向特性を得ることができ、ほとんどの場合は正確な方向に指向特性を持つことができるかもしれない。しかし、分離行列の行が入れ替わり、目的音の方向ではなく、雑音の方向に指向特性が向く可能性がある。そこで、今後は 2 つの手法の長所を組み合わせ、パーミュテーション問題を解決することを考えたい。具体的には、前処理として PF-ICA で大まかな方向を推定し、次に従来法を用いて音源の方向を推定し、前処理で推定した方向に近い方向を持つ行を分離行列に揃える方法である。このようにすれば、従来の方法よりも高精度に分離行列を求めることができると考えられる。

本稿では JADE を用いた場合しか評価していない。PF-ICA は、どのような方法を用いても計算できるはずであるので、FastICA などを用いた場合の分離結果なども調べてみたい。

また、PF-ICA でも低周波数では分離結果が良くないが、これを解決するために、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を検討したい。

文 献

- [1] T.W.Lee: "Independent Component Analysis", Kluwer, 1998.
- [2] 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, Aussios, France, Jan., 1999.
- [3] A.Hyvärinen, J.Karhunen, and E.Oja: "Independent Component Analysis", John Wiley, New York, 2001.
- [4] 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.
- [5] 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.
- [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 問題の頑健な解法", 音響学会講演論文集, pp. 777-778, Mar, 2003.
- [7] 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.