

近傍周波数一括処理 ICA によるブラインド音源分離の 後処理としての Wiener フィルタの効果

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

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

[‡] エントレリオス国立大学 アルゼンチン パラナー市

E-mail: [†] (dtg0731,etf1704,myanagid@mail.doshisha.ac.jp [‡] ldpersia@ciudad.com.ar,

あらまし 従来の周波数領域 ICA では、周波数ビン毎に分離を行っており、周波数ビン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、分離実行後に音源について並べ替えを行う必要がある。この問題を避けるために、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)が提案されている。この手法では周波数ビンの入れ替わりが生じない利点があるものの、分離行列を周波数ビン毎に考慮することができないという欠点がある。本報告では、ICA の後処理として時間周波数領域で Wiener フィルタを適用することによってこの欠点を克服し、周波数毎の特性を考慮できるようにしている。Wiener フィルタの適用による効果はセグメンタル SNR・PESQ の2つの評価指標で評価している。

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問題, Wiener フィルタ

Effects of Wiener Filter on Blind Source Separation by processing on sequentially connected bins

Tadashige NOGUCHI[†] Kenko OTA[†] Leandro Di Persia[‡] and Masuzo YANAGIDA[†]

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

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

E-mail: [†] (dtg0731,etf1704,myanagid@mail.doshisha.ac.jp [‡] ldpersia@ciudad.com.ar

Abstract

Permutation-Free ICA is a method that can avoid so-called "Permutation problem" that appears in conventional frequency-domain ICA due to indeterminicity among source identity in each frequency bin. Permutation-Free ICA carries out separation on a set of long vectors consisting of connected temporal changes of frequency components of the received signal. The separation matrix obtained in Permutation-Free ICA, however, has common directivity for all frequency bins. Applying Wiener Filter on a resultant signal of Blind Source Separation, frequency characteristics of the separation matrix are thought to be taken into account. In this paper, effects of applying Wiener Filter after Blind Source Separation are evaluated using segmental SNR and PESQ as evaluation indexes.

Keyword Blind Source Separation, Independent Component Analysis, Permutation Problem, Wiener Filter

1. はじめに

近年、ブライント音源分離 (BSS: Blind Source Separation) についての研究が盛んに行われている。ブライント音源分離は、音源信号や混合過程に関する情報が未知であるという状況において、マイクロホンで観測した複数の信号を分離して音源信号を推定する技術である。独立成分分析 (ICA: Independent Component Analysis) は、BSS を実現する手法の一つであり、瞬時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、実環境では壁や床などによる反射の影響を受けるため、混合過程が瞬時混合でなく畳み込み混合になってしまう。そこで、一般には信号を時間領域から周波数領域に変換して、分離処理を実行する周波数領域 ICA によって、残響に対処している [2][3]。

しかし、BSS に周波数領域 ICA を適用した場合、周波数ビン毎に分離を行うことになり、周波数ビン間で分離した信号の入れ替わりが発生する可能性がある。これがパーミュテーション問題と呼ばれるものである。そこで、各チャンネルの周波数ビンの入れ替わりを正しく並び替える必要があり、これまでに様々な解決策が提案されている [2][4]。また、パーミュテーション問題を回避する手法としては Kim *et al.* [5] によるものがある。この手法は、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、パーミュテーション問題を回避している。

Di Persia らは、パーミュテーション問題を避けるために、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) を提案している [6][7]。しかし、この手法では求まる分離行列がすべての周波数ビンについて共通のもの 1 つだけになるので、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、PF-ICA の後処理として、時間周波数領域で周波数ビン毎、あるいは時間フレーム毎に Wiener フィルタを適用することによって、PF-ICA では考慮できなかった周波数特性をいくらか救済する手法を提案する。

2. 従来の周波数領域 ICA

周波数領域 ICA は M 個のマイクロホンでの観測信号 $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ 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のベクトル $\mathbf{X}(f, k) = [X_1(f, k), X_2(f, k), \dots, X_M(f, k)]^T$ に対して、周波数ビン f において学習された分離行列 $\mathbf{W}(f)$ 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 (Joint Approximate Diagonaliz-

ation of Eigenmatrices) を用いる [8]。ここで、 $\mathbf{Y}(f, k)$ を分離された信号のベクトルとすると分離過程は、

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

と表される。

周波数領域 ICA では音源信号が互いに独立であるという仮定に基づき、各周波数ビン毎に分離後の信号がそれぞれ独立となるように分離行列 $\mathbf{W}(f)$ を最適化する。そのため、 $\mathbf{W}(f)$ の行が入れ替わっても、 $\mathbf{Y}(f, k)$ の各行の独立性は保たれるので、ある周波数 f_n と f_{n+1} において、 $\mathbf{Y}(f_n, k)$ と $\mathbf{Y}(f_{n+1}, k)$ の各行が必ずしも同じ音源に対応しているとは限らず、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ず、これに対して多様な手法が提案されている [2][4]。しかし、この問題は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. Permutation Free ICA (PF-ICA)

Di Persia らは、パーミュテーション問題を避けるために、周波数ビンの入れ替わりが本質的に発生しない PF-ICA を提案した [6][7]。この手法では、分離行列を生成する際に発生するパーミュテーション問題を避けるため、分離行列生成過程の前後でパーミュテーションの発生を避ける処理を行っている。具体的には、時間-周波数領域における観測信号の各周波数ビン $f_1, f_2, \dots, f_n, \dots, f_N$ 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い、分離の際に使用する観測行列 \mathbf{X}_{PF} を以下のように作成する。

$$\mathbf{X}_{PF} = [\mathbf{X}_{1PF}, \mathbf{X}_{2PF}, \dots, \mathbf{X}_{mPF}, \dots, \mathbf{X}_{MPF}]^T \quad (2)$$

$$\begin{aligned} \text{ただし, } \mathbf{X}_{mPF} &= [\mathbf{X}_m(f_1), \mathbf{X}_m(f_2), \dots, \mathbf{X}_m(f_N)]^T \\ \mathbf{X}_m(f_n) &= [X_m(f_n, 1), X_m(f_n, 2), \dots, X_m(f_n, K)] \end{aligned}$$

ここで、 K は全フレーム数、 $\mathbf{X}_m(f_n)$ は m 番目のマイクロホンで観測した周波数ビン f_n のフレーム時系列、 \mathbf{X}_{mPF} は $\mathbf{X}_m(f_1), \mathbf{X}_m(f_2), \dots, \mathbf{X}_m(f_N)$ を連結したベクトルを表している。

この処理を図 1 に示す。図 1 の左図は、時間-周波数領域上での各周波数ビンを表しており、各周波数ビンを連結し、右のようにマイクロホン毎に 1 本のベクトルを生成する。

そして、 M 本のベクトルに対して分離を実行し、次のように分離信号 \mathbf{Y}_{PF} を求める。

$$\mathbf{Y}_{PF} = \mathbf{W} \mathbf{X}_{PF} \quad (3)$$

但し、 \mathbf{W} は PF-ICA で用いられる分離行列であり、全ての周波数で同じ \mathbf{W} を用いる。このようにして求められた l 番目の音源ベクトル $\mathbf{Y}_{lPF} = [Y_{lPF}(1), Y_{lPF}(2), \dots, Y_{lPF}(NK)]$ を各周波数ビン $\mathbf{Y}_l(f_1), \mathbf{Y}_l(f_2), \dots, \mathbf{Y}_l(f_n), \dots, \mathbf{Y}_l(f_N)$ に分割し、信号を時間-周波数領域に戻す。

$$\begin{aligned} \mathbf{Y}_l(f_1) &= [Y_{lPF}(1), Y_{lPF}(2), \dots, Y_{lPF}(K)] \\ \mathbf{Y}_l(f_2) &= [Y_{lPF}(K+1), Y_{lPF}(K+2), \dots, Y_{lPF}(2K)] \\ &\vdots \\ \mathbf{Y}_l(f_n) &= [Y_{lPF}((n-1)K+1), Y_{lPF}((n-1)K+2), \dots, Y_{lPF}(nK)] \\ &\vdots \\ \mathbf{Y}_l(f_N) &= [Y_{lPF}((N-1)K+1), Y_{lPF}((N-1)K+2), \dots, Y_{lPF}(NK)] \end{aligned} \quad (4)$$

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 \mathbf{W} 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(音源数)×(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。

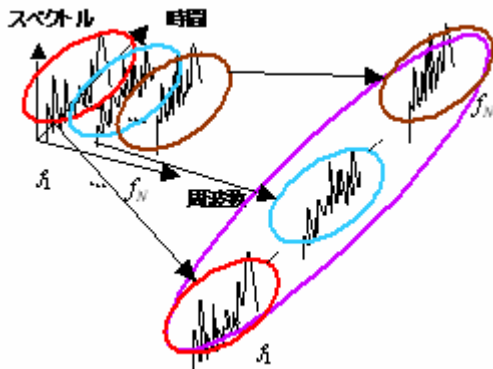


図1. 1本のマイクロホンの観測信号について、時間周波数領域の信号から PF-ICA の分離で用いるベクトルへの変換

4. Wiener フィルタ [9]

雑音の中にある信号をできるだけ正確に選び出すフィルタ特性 H を考える(図2)。入力信号のパワースペクトル $P_s(f, k)$ と存在している雑音のパワースペクトル $P_n(f, k)$ は既知のものとする。ここで信号自体から生ずる誤差 $E_s(f, k)$ と雑音 $n(t)$ のスペクトル $N(f, k)$ により発生する

誤差 $E_n(f, k)$ をスペクトル表示すると以下のようになる。

$$\begin{aligned} E_s(f, k) &= H(f, k)S(f, k) - S(f, k) \\ &= (H(f, k) - 1)S(f, k) \end{aligned} \quad (5)$$

$$E_n(f, k) = N(f, k)H(f, k) \quad (6)$$

これらのパワースペクトルを $P_{Es}(f, k)$, $P_{En}(f, k)$ とすると、入力信号 $s(t)$ と出力信号 $\hat{s}(t)$ の誤差 $e(t)$ の2乗平均は、Parseval の等式からパワースペクトル密度を全周波数で積分したものに比例する。

$$\begin{aligned} \overline{e^2(t)} &= \int_{-\infty}^{\infty} [\{H(f, k) - 1\} \{H^*(f, k) - 1\} P_s(f, k) \\ &\quad + H(f, k)H^*(f, k) P_n(f, k)] df \end{aligned} \quad (7)$$

この値を最小にする $H(f, k)$ を考えると、 \int 内を H で微分して0とおくことにより、

$$H(f, k) = \frac{P_s(f, k)}{P_s(f, k) + P_n(f, k)} \quad (8)$$

となる。これを Wiener フィルタという。

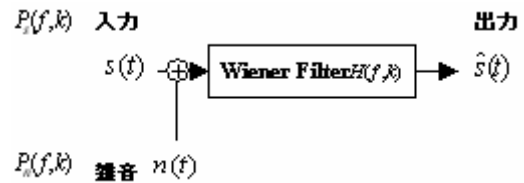


図2. Wiener フィルタ

5. 提案法

ここで、PF-ICA の後処理として、時間-周波数領域で Wiener フィルタを適用することを提案する。このフィルタを適用することによって、PF-ICA の周波数特性を考慮できなかった欠点を補う。ここで簡単化のため、音源・マイクロホンの数をそれぞれ2とする。

Wiener フィルタを用いる場合、本来は目的信号と雑音信号のパワースペクトルが既知でなくてはならないが、BSSにおいては、目的・雑音信号に関する事前情報は一切未知である。そこで、PF-ICA によって得られた分離信号 $\mathbf{Y}_1, \mathbf{Y}_2$ は完全ではないがある程度目的・雑音信号に近いと考えて、分離信号 $\mathbf{Y}_1, \mathbf{Y}_2$ のパワースペクトル $P_{Y_1}(f, k)$ と $P_{Y_2}(f, k)$ を目的・雑音信号のパワースペクトル $P_s(f, k)$, $P_n(f, k)$ として用いて、Wiener フィルタを

$$H_{W_1}(f, k) = \frac{P_{Y_1}(f, k)}{P_{Y_1}(f, k) + CP_{Y_2}(f, k)} \quad (9)$$

$$H_{W_2}(f, k) = \frac{P_{Y_2}(f, k)}{P_{Y_2}(f, k) + CP_{Y_1}(f, k)} \quad (10)$$

のように設計する．ここで，分離結果の一方を目的音，他方を雑音としてフィルタを設計し，次に目的音と雑音を入れ替えてフィルタを設計する．BSSによる分離が不十分な場合，分離処理によって推定した信号 Y_1 と Y_2 において，雑音信号の中にも目的信号が含まれているため，過度の減算を行ってしまう場合がある．これを避けるために，式(9)(10)のフィルタを設計する際に，次のような分離信号の相関 C を用いている．

$$C(f) = \frac{|Y_1(f, k) \cdot Y_2(f, k)|}{|Y_1(f, k)| |Y_2(f, k)|} \quad (11)$$

PF-ICA の後処理として Wiener フィルタを適用する提案法全体の処理の流れを図3に示す．

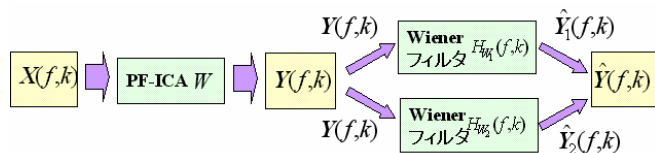


図3. 提案法全体の処理の流れ

式(9)(10)では Wiener フィルタを適用する際に，周波数毎にフィルタを設計している．しかし，一般的にアレイ処理では低周波領域での分離精度が悪いため，周波数毎にフィルタを設計するよりも，低周波領域・高周波領域をそれぞれ別の方法でフィルタを設計する方が効果的であると考えられる．そこで，低周波領域では，式(9)(10)を用いて設計した Wiener フィルタをそのまま適用せずに，高周波領域で求めたフィルタを利用して，次のように低周波領域のフィルタを設計する．ある周波数 f_H 以上のフィルタ係数 $H_{W_j}(f_n)$ を利用して，ある周波数 f_L 以下のフィルタ $H_{W_j}(f_L)$ を次のように設計する．

$$H_{W_j}(f_L, k) = \frac{1}{N - H + 1} \sum_{n=H}^N H_{W_j}(f_n, k) \quad j=1,2 \quad (12)$$

ここで， f_{L+1} 以上の周波数については，式(9)(10)によって周波数毎に設計したフィルタ $H_{W_j}(f_n)$ を用いる．

6. 提案法の評価

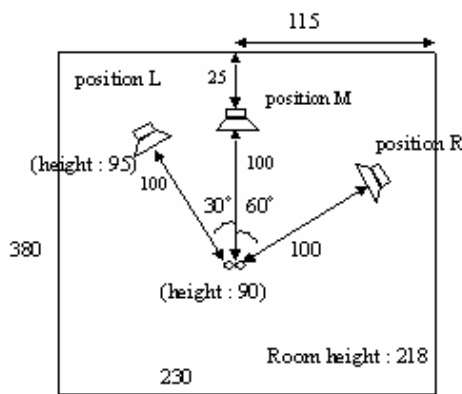
音声の収録は，一般家庭のリビングルームを想定した部屋(幅 230cm×奥行き 380cm×高さ 218cm)で行った．Time Stretched Pulse を用いて測定した部屋の残響時間は 370ms である．音源(スピーカ)と観測点(マイクロホン)の配置を図4に示す．観測点は固定し，音源の配置は観測点を中心に -30° ， 0° ， 60° (時計回りを正)となるようにし，それぞれの配置を L, M, R とする．また，マイクロホンの間隔は 5cm とした．今回使用した音源データを表1，音響分析条件を表2に示す．

表1 音源データ

音源信号	男女各3名・5種類
雑音信号	男女各3名・2種類
混合音声の組み合わせ	300通り
配置の組み合わせ	3種類(M-L, M-R, L-R)
平均データ長	1.25秒

表2 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	512点
分析窓	ハミング窓



(単位: cm)

図4. 収録環境

6.1 評価指標

評価指標には，セグメンタル SNR (Signal to Noise Ratio) と PESQ (Perceptual Evaluation of Speech Quality) を用いた [10]．

セグメンタル SNR は，等間隔の周波数軸上で目的信号に対する雑音信号のパワーの比を表しており，式(13)のように定義されている．

$$SNR(k) = \frac{\int |S(f,k)|^2 df}{\int |S(f,k) - Z(f,k)|^2 df} \quad (13)$$

ここで $S(f,k), Z(f,k)$ はそれぞれ音源信号 $s(t)$ と評価対象信号 $z(t)$ のフーリエ変換である。評価対象は、 X, Y, \hat{Y} である。

PESQ は、ITU-T (Telecommunication Standardization Sector of the International Telecommunication Union) で勧告されている規格であり、信号に対して人間の知覚特性を考慮した処理を施しており、被験者が音を聞いて評価する主観音質評価の1つである、MOS (Mean Opinion Score) と高い相関がある。ここで、MOS 値は 1(悪い) 5(良い)の値をとり、PESQ 値は -0.5 4.5 の値をとる。PESQ を用いて評価することは、客観音質評価によって、人が知覚した時の主観評価に近い結果が得られるということである。ここで、PESQ では、以下の処理を行うことによって、人間の知覚特性を考慮している。

- ・ 高域・低域の周波数分解能の違いに対応するために、間隔が非線形な周波数軸(バークスケール)に変換する
- ・ スペクトル強度のスケールを人間が感じる音の強さ(ラウドネススケール)の感度に変換する

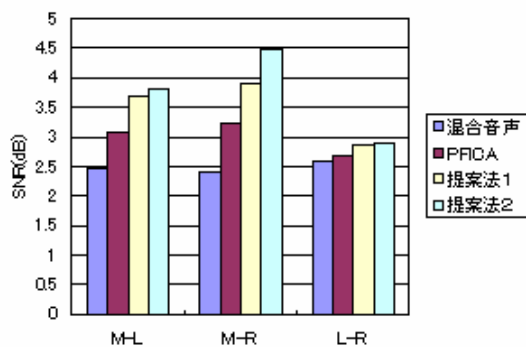


図 5. SNR による比較

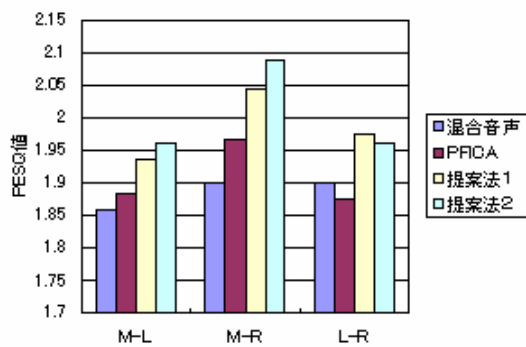


図 6. PESQ による比較

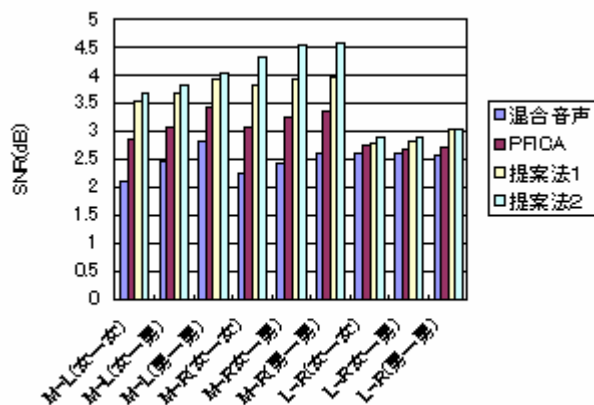


図 7. SNR による比較(男女別組み合わせ)

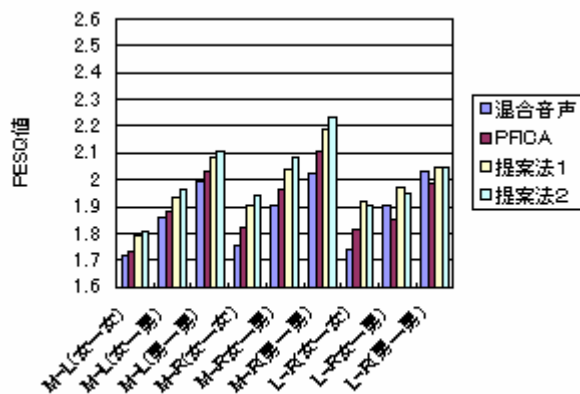


図 8. PESQ による比較(男女別組み合わせ)

6.2 評価する手法

PF-ICA の後処理として、式(9)(10)を用いて周波数毎に Wiener フィルタを設計する方法を提案法 1 とし、低周波領域に式(12)を用いて Wiener フィルタを設計する方法を提案法 2 とする。ここで、 $f_L=2000\text{Hz}$, $f_H=4000\text{Hz}$ とし、 2000Hz 以下の周波数を低周波領域、 4000Hz 以上の周波数を高周波領域として Wiener フィルタを適用した。

6.3 評価結果

混合音声と混合音声に PF-ICA, 提案法 1, 提案法 2 それぞれを適用した時の分離信号のセグメンタル SNR, PESQ の各配置における結果の平均をそれぞれ図 5,6 に示す。また、男女別の組み合わせについての結果の平均を図 7,8 に示す。

図 5 の結果より L-R の配置では、SNR の改善がほとんどないが、M-L・M-R の配置では提案法によって SNR・PESQ が改善されていることがわかる。M-R の配置で、PF-ICA と提案法 1 の SNR を比較すると 0.6dB 程度、提案法 2 と比較すると 1.3dB 程度改善されている。さら

に、図 7 から同じ配置で男女の組み合わせによる SNR の改善の傾向はほとんど同じであった。

また、図 5,6 の結果より M-L, M-R の配置では、PESQ と SNR は同じような傾向になっていることがわかる。PF-ICA と提案法の PESQ 値を比較すると、わずか 0.06 0.15 程度であるが、有意水準 5% での t 検定の結果、有意に改善していることがわかった。さらに、図 8 よりどのような男女の組み合わせでも、PF-ICA と提案法において PESQ 値が改善していることがわかる。

7. 考察

図 5 8 より、PF-ICA の後処理に Wiener フィルタを適用すると、分離精度が向上することがわかった。さらに、信号と雑音の比のみで評価する SNR だけでなく、人間の知覚特性を考慮して評価を行う PESQ 値も t 検定で有意な差があり、わずかだが向上していると言える。

また、男女の組み合わせによる比較では、SNR は男女の組み合わせに関係なく同じような傾向になっているのに対して、PESQ では男-男の組み合わせが聞こえやすいという結果になった。

さらに、提案法 1 のように周波数毎に Wiener フィルタを設計するよりも、提案法 2 のように高周波領域のフィルタを利用して低周波領域のフィルタを設計した方が、フィルタの効果が高くなった。これより、Wiener フィルタを適用する際に、大きく高周波領域と低周波領域に分けてフィルタ係数の設計するだけで十分効果的であることがわかる。

8. 今後の課題

低周波領域で Wiener フィルタを適用する際に、高周波領域でのフィルタ係数を用いると効果的であることがわかった。本稿では 4000Hz 以上の高周波領域のフィルタについての平均を 2000Hz 以下の低周波領域のフィルタとしただけであった。そこで、境界周波数をどのように設定するかなど、低周波領域での Wiener フィルタの設計方法をさらに検討する必要がある。さらに、本稿は周波数毎に時間方向に Wiener フィルタを適用しただけであったが、フレーム毎に周波数方向にも Wiener フィルタを適用することも検討したい。

また、Wiener フィルタを適用する際に入力信号・雑音信号を PF-ICA の分離結果と仮定したため、PF-ICA の分離精度が悪いと Wiener フィルタの効果が小さいと考えられる。PF-ICA の分離精度を向上させれば、Wiener フィルタ設計に利用する信号は本来の入力・雑音信号に近づき、Wiener フィルタの効果も大き

くなるはずである。PF-ICA では全周波数を一括で処理しているが、音声信号は低周波領域のパワーが強いため、分離行列がほとんど低周波数の情報によって決定されるため、今後は PF-ICA の前処理として、ハイパスフィルタを適用するなどの処理を行い、PF-ICA 自体の分離精度を向上させる手法を検討したい。

謝辞 本研究の一部は同志社大学学術フロントティア事業「人間と生物の賢さの解明とその応用」の援助を受けた。

参考文献

- [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] 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.
- [5] T.Kim, H.Attias, S-Y.Lee, and T-W.Lee:
"Blind source separation exploiting higher-order frequency dependencies", IEEE Transactions on Speech and Audio Processing Vol.15 No.1, 2007
- [6] 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.
- [7] 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.
- [8] J.F.Cardoso, and A.Souloumiac:
"Blind beamforming for non Gaussian signals", IEEE Proceeding-F, vol. 140, no 6, pp.362-370, Dec, 1993.
- [9] 谷口慶二: "信号処理の基礎", 共立出版株式会社, pp.43-44.
- [10] "Perceptual Evaluation of Speech Quality, an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs ", ITU-T Recommendation P. 862, Feb,2001