周波数領域ICAの後処理としてWienerフィルタを用いた短時間混合音声の分離

半田晶寛 †　　レアンドロ・ディ・ペルシア ‡　　大田健紘 †　　柳田益造 †
† 同志社大学工学部　　‡ エントレリオス国立大学 (アルゼンチン)

あらまし 残響環境下で，周波数領域 ICA によるブラインド音源分離を行った際，持続時間が比較的長い混合音声に対してはある程度の分離精度を得ることができるが，1 秒前後の短時間混合音声に対しては十分な分離精度をあげるに至っていない．主な原因は各周波数ビンでのデータ量不足と考えられる．そこで本稿では 1 秒前後の短時間混合音声における短時間フーリエ変換をする際の最適な窓長とシフト幅の調査を行い，データ量不足の影響を軽減させ，さらに周波数領域 ICA の後処理として各周波数ビンで Wiener フィルタを適用させることで，分離精度の向上を図った．

Separation of mixed speech signals of short duration using Wiener filter as postprocessing for Frequency-Domain ICA

Akihiro HANDA†　Leandro Di Persia‡　Kenko OHTA†　Masuzo YANAGIDA†
†Faculty of Engineering, Doshisha University　　‡Universidad Nacional de Entre Rios, Argentina

Abstract Frequency-domain ICA is effective for separating mixed speech signals of long duration but it is not the case for signals of short duration in environments having ordinary reverberation time. The main reason would be lack of data in each frequency bin. The optimal window size and shifting interval for separating short speech are investigated, and Wiener filter is adopted in each frequency bin as post-processing of frequency-domain ICA.

1　はじめに

近年情報技術の進歩に伴い，家電操作においても多機能化が見られ，操作の複雑化が問題となっている．その解決策として音声によるコマンド操作が考えられる．しかし現在の音声認識システムは，マイクに接近した位置からの音声に対してはそれなりに高い認識精度を有する一方で，マイク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て認識精度は著しく低下してしまう [1]．実環境で音声認識を行うためには音声認識システムに入力される音声に対して何らか処理を行う必要がある．その方法の一つがブラインド音源分離（BSS: Blind Source Separation）である．BSS とは，複数のマイクロホンへ線形に混合された信号が入力された時に，音源信号や混合過程を知ることなく，観測信号のみから音源信号を推定し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．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 を適用することができる．しかし，この手法を用いても高残響下での分離や 1 秒前後の短時間音声による混合音声の分離に関しては十分な分離精度を得ることができない [3]．そこで本稿では，高残響下で 1 秒前後の音声を目的音とした混合音声に対する周波数領域 ICA の処理条件を検討し，その後処理として Wiener フィルタを用いて分離精度の向上を試みた．

2　研究の目的

ICA を用いた BSS の手法は，線形混合に対しては有効な手法であるといえるが，実環境の音は壁などからの反射音が畳み込まれて混合されるために，畳み込みを考慮した時間領域 BSS や，周波数領域に変換して瞬時混合の ICA を適用する周波数領域 BSS など，残響に対応するための手法が提案されている [4]．しかし実環境下（高残響下）や対象データが短時間信号である場合には，十分

な効果を挙げるには至っていない．そこで本研究では実環境下で 1 秒前後の短時間混合音声（2 音源）に対して，周波数領域 ICA による BSS の後処理としての Wiener フィルタの有効性を検証することを目的とした．

混合音声の収録は一般家庭のリビングルームを想定して設計された部屋で行い，観測点（マイクロホン）は固定し，音源位置（スピーカ）と音源の種類（男女計 6 名）を変えて収録した混合音声を用いて分離を行った．周波数領域 ICA による分離では，JADE（Joint Approximate Diagonalization of Eigenmatrices）法 [5] の結果を初期値に用いた FastICA[6] を適用し，それらの結果を Wiener フィルタ適応させた周波数領域 BSS の分離精度を検証する．認識対象の音声にはテレビの操作コマンドを用いる．

3 混合過程と従来の周波数領域 BSS

3.1 混合過程

マイクロホン数を M，音源数を N とし（本稿では $M = N = 2$ の場合を考える），短時間フーリエ変換を用いることにより，複数の音源信号の畳み込み混合モデルは次のように表される．

$$\boldsymbol{X}(f,t) = \boldsymbol{A}(f)\boldsymbol{S}(f,t)$$

ここで，$\boldsymbol{X}(f,t) = (X_1(f,t),\ldots,X_M(f,t))^T$ は観測ベクトル，$\boldsymbol{S}(f,t) = (S_1(f,t),\ldots,S_N(f,t))^T$ は音源信号ベクトル，$\boldsymbol{A}(f)$ は混合行列である. また f は周波数，t はフレーム番号（時刻に相当）である.

3.2 周波数領域 BSS

周波数領域 ICA による BSS では短時間フーリエ変換を用いることにより得られた観測信号ベクトル $\boldsymbol{X}(f,t)$ を各周波数ビンで分離行列 $\boldsymbol{W}(f)$ を更新させて分離を行う. ここで $\boldsymbol{Y}(f,t) = (Y_1(f,t),\ldots,Y_N(f,t))^T$ を分離信号ベクトルとすると分離過程は，次のように表される.

$$\boldsymbol{Y}(f,t) = \boldsymbol{W}(f)\boldsymbol{X}(f,t)$$

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，各周波数ビン内で観測信号がそれぞれ独立となるように分離行列 $\boldsymbol{W}(f)$ を最適化する.

4 学習アルゴリズム

音声認識システムに入力する音声の前処理として十分な役割を果たすために収束の速さを考慮して FastICA による学習アルゴリズムを適用する．さらに，解の収束性を高めるために，JADE 法によって得た分離行列 $\boldsymbol{W}(f)$ を FastICA の初期値として採用する．

4.1 FastICA

FastICA はの Hyvärinen らによって提案された不動点法を用いた ICA のアルゴリズムである [6,7]．FastICA は次にあげる二つの性質において，勾配法に基づくアルゴリズムよりも優れた点をもつ．第一にこのアルゴリズムの収束は 3 次的であり，非常に速い．第二に，勾配法に基づくアルゴリズムと対照的に，学習係数や他のパラメータを必要としない安定したアルゴリズムである．

4.1.1 独立性の定義

本稿では独立性を negentropy の最大化によって定義した FastICA を適用している．独立性の定義としてしばしば使われる尖度を用いないのは，尖度がはずれ値に対して過敏ということ [6] から，尖度は非ガウス性の頑健な尺度ではないと考えられるためである．negentropy は以下のように定義される．

$$J(\boldsymbol{y}) = H(\boldsymbol{y_{Gauss}}) - H(\boldsymbol{y}) \tag{1}$$

ここで，ベクトル $\boldsymbol{y}$ は復元信号を表す時間関数としての確率変数，$\boldsymbol{y_{Gauss}}$ は $\boldsymbol{y}$ と同じ共分散行列となる標準正規分布に従う確率変数である．$H(\boldsymbol{y})$ は $\boldsymbol{y}$ の entropy を表し，$\boldsymbol{y}$=$\boldsymbol{y_{Gauss}}$ の時に最大となる．negentropy は非負であり，復元信号 $\boldsymbol{y}$ が正規分布に従う時に最小の 0 となる．つまり，negentropy を最大化するということは，復元信号 $\boldsymbol{y}$ の非正規性を最大化することである．中心極限定理の逆の考え方から，非正規性の最大化は独立性の最大化につながる．ただし，negentropy の推定は困難であるので，非 2 次の非線形関数 G を用いて式 (1) を次のように近似する．ここで復元信号ベクトル $\boldsymbol{y}$ の 1 要素を求めることを考える．

$$J(y) = k[E\{G(y)\} - E\{G(y_{Gauss})\}]^2 \tag{2}$$

ここで，y，y_{Gauss} は平均 0，分散 1 に正規化されているスカラー量とする．また k は正の定数である．この近似が非常に正確だとはいえない場合でも，それは常に非負であり，変数 y が正規分布になるときには 0 になるという意味で無矛盾となる．ただし非線形関数の選び方として，あまり速く増加しない関数 G を選ぶ必要がある．そうすることで，より頑健な推定量が得られる．

4.1.2 アルゴリズムの導出

$y{=}\boldsymbol{w^T z}$ ($\boldsymbol{z}$ は白色化された観測ベクトルである) を negentropy の近似式 (式 (2)) に適用し，極大を得ることを考える．$E\{G(\boldsymbol{w^T z})\}$ の極大を与える $\boldsymbol{w}$ は，Lagrange の未定係数法によれば，制約条件 $\|\boldsymbol{w}\|^2{=}1$ のもとでの最適化として，評価関数

$$F = E\{G(\boldsymbol{w}^T\boldsymbol{z})\} + \lambda(\|\boldsymbol{w}\|^2 - 1) \tag{3}$$

の偏微分を 0 とおいて, すなわち $\frac{\partial F}{\partial \boldsymbol{w}} = \boldsymbol{0}$ として，

$$E\{\boldsymbol{z}g(\boldsymbol{w}^T\boldsymbol{z})\} + \beta\boldsymbol{w} = \boldsymbol{0} \tag{4}$$

を満たす $\boldsymbol{w}$ として求まる．ここで，関数 g は関数 G の導関数，$\beta = 2\lambda$ である．この方程式を Newton 法で解くことを考える．式 (4) の左辺を $\boldsymbol{f}$ とおくと

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}} = E\{\boldsymbol{z}\boldsymbol{z^T}g'(\boldsymbol{w}^T\boldsymbol{z})\} + \beta\boldsymbol{I} \tag{5}$$

と求まる. ただし，g' は g の導関数である．データは白色化されているので,

$$E\{\boldsymbol{z}\boldsymbol{z}^T g'(\boldsymbol{w}^T\boldsymbol{z})\} \approx E\{g\prime(\boldsymbol{w}^T\boldsymbol{z})\}\boldsymbol{I} \tag{6}$$

が可能となる．こうすることで $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}}$ は対角行列になり簡単に逆行列が求まる．そこで以下の近似 Newton 法を得たことになる．

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - [E\{\boldsymbol{z}g(\boldsymbol{w}^T\boldsymbol{z})\} + \beta\boldsymbol{w}]/[E\{g'(\boldsymbol{w}^T\boldsymbol{z})\} + \beta] \tag{7}$$

さらに両辺に $E\{g\prime(\boldsymbol{w}^T\boldsymbol{z})\} + \beta$ を乗じ，$\boldsymbol{w}$ のノルムを 1 に正規化することによって以下の式を得る．

$$\boldsymbol{w} \leftarrow E\{\boldsymbol{z}g(\boldsymbol{w}^T\boldsymbol{z})\} - E\{g'(\boldsymbol{w}^T\boldsymbol{z})\}\boldsymbol{w} \tag{8}$$

ただし，本稿では，周波数領域で ICA を適用しているためデータは複素数値となり，式 (8) は以下のようになる．

$$\begin{aligned}\boldsymbol{w} \leftarrow\ & E\{\boldsymbol{z}(\boldsymbol{w}^H\boldsymbol{z})^* g(|\boldsymbol{w}^H\boldsymbol{z}|^2)\} \\ & -E\{g(|\boldsymbol{w}^H\boldsymbol{z}|^2) + |\boldsymbol{w}^H\boldsymbol{z}|^2 g\prime(|\boldsymbol{w}^H\boldsymbol{z}|^2)\}\boldsymbol{w}\end{aligned} \tag{9}$$

*は複素共役，H は共役転置行列を表す．

このアルゴリズムを用いた場合の収束の精度は，分離ベクトル $\boldsymbol{w}$ の初期値に大きく依存する．よって，初期値の選び方が重要となる．

4.1.3 逐次的直交化 (Deflation 法)

これまで 1 成分の推定を考えてきた．しかし複数の成分を推定するためには，異なった成分が同じ最大値に収束するのを防ぐために各繰り返し後の出力値 $\boldsymbol{w}_1^T\boldsymbol{z},\ldots,$ $\boldsymbol{w}_N^T\boldsymbol{z}$ が無相関となるようにしなければならない．これはつまり $\boldsymbol{w}_1^T,\ldots,$ $\boldsymbol{w}_N^T$ の直交化を要請していることになる．本研究では Gram-Schmidt の正規直交化法を用いて各成分を一つずつ求める Deflation 法 [6] を適用する．手順は次のとおりである．

1. $\boldsymbol{w}_p^T \leftarrow \boldsymbol{w}_p^T - \sum_{j=1}^{p-1}(\boldsymbol{w}_p^T\boldsymbol{w}_j)\boldsymbol{w}_j$

2. $\boldsymbol{w}_p \leftarrow \dfrac{\boldsymbol{w}_p}{\|\boldsymbol{w}_p\|}$

ここで p は音源 ID であり，$1 < p \leq N$ である．この処理を繰り返し行うことによって推定ベクトル $\boldsymbol{y}$ の無相関化と分離ベクトル $\boldsymbol{w}$ の直交化を行う．

4.2 JADE 法

JADE は Cardoso らによって提案された手法で，4 次クロスキュムラントを対角化する手法 [5,8] である．4 次クロスキュムラントは次のように定義される．

$$\begin{aligned}\mathrm{cum}(x_i, x_j, x_k, x_l) &= E(x_i, x_j, x_k, x_l) \\ -E(x_i, x_j)E(x_k, x_l) &- E(x_i, x_k)E(x_j, x_l) \\ &-E(x_i, x_l)E(x_j, x_k)\end{aligned} \tag{10}$$

観測信号 $\boldsymbol{x}$ が平均 0 で無相関化 ($\boldsymbol{z}$ とする) されている時，音源信号 $\boldsymbol{s}$ と無相関化された観測信号 z はある直交行列 $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ により,

$$\boldsymbol{z} = \boldsymbol{U}\boldsymbol{s} \tag{11}$$

という関係で結ばれている．独立性の仮定から

$$\mathrm{cum}(s_i, s_j, s_k, s_l) = \begin{cases}\kappa_i, & i = j = k = l \\ 0, & \text{otherwise}\end{cases} \tag{12}$$

となる．ここで，行列 $\boldsymbol{M}{=}(m_{ij})$ により縮約された kurtosis の行列の第 i, j 要素は

$$C_{i,j}(\boldsymbol{M}) = (\sum_{k,l=1}^{N}\mathrm{cum}(z_i, z_j, z_k, z_l)\, m_{kl}) \tag{13}$$

を考える．S のインデックスが全て同じ時だけが問題であるから，$\kappa_i = cum(s_i, s_i, s_i, s_i)$ として式 (13) を変形すると

$$C(M) = \sum_{i=1}^{N}(\kappa_i\boldsymbol{u}_i^T\boldsymbol{M}\boldsymbol{u}_i)\boldsymbol{u}_i\boldsymbol{u}_i^T \tag{14}$$

となる．ここで

$$\boldsymbol{D}(\boldsymbol{M}) = \mathrm{diag}(\kappa_1\boldsymbol{u}_1^T\boldsymbol{M}\boldsymbol{u}_1, \ldots, \kappa_N\boldsymbol{u}_N^T\boldsymbol{M}\boldsymbol{u}_N) \tag{15}$$

とすると，

$$C(\boldsymbol{M}) = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{M})\boldsymbol{U}^T \tag{16}$$

直交行列である $\boldsymbol{U}$ で対角行列を挟んでいるので，$C(\boldsymbol{M})$ は対称行列になっている．この $C(\boldsymbol{M})$ を対角化する U を探すのが JADE である．行列の対角化には Jacobi 法を用いることができる．Jacobi 法は理想的には 2 次収束することが知られているので，JADE も収束の速いアルゴリズムといえる．ただし，JADE は 4 次元配列を用いるので，信号数があまり多くない時には問題がないが，信号数が多い時には計算上の問題がある．

5　提案手法

上述したアルゴリズムにより得られた分離結果を用いた後処理として Wiener フィルタを適用させせることで分離精度の向上を考える．

5.1　Wiener フィルタ [9]

雑音中の信号をできるだけ正確に選び出すフィルタの特性 $H_W(\omega)$ を考える (図1)．今，入力信号 $x(t)$ のパワースペクトル $P_x(\omega)$ と存在している雑音 $n(t)$ のパワースペクトル $P_n(\omega)$ は既知のものとする．ここで対象信号に対する誤差と雑音に対する推定誤差をスペクトル表示すると，

$$E_x(\omega) = H_W(\omega)X(\omega) - X(\omega) = X(\omega)[H_W(\omega) - 1] \tag{17}$$

$$E_n(\omega) = N(\omega)H_W(\omega) \tag{18}$$

これらのパワースペクトルを，$P_{E_x}(\omega)$，$P_{E_n}(\omega)$ とすると，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一致する

$$\overline{e^2}(t) = \int_{-\infty}^{\infty} [\{H_W(\omega) - 1\}\{H_W^*(\omega) - 1\} \cdot P_x(\omega) + H_W(\omega)H_W^*(\omega) \cdot P_n(\omega)]d\omega \tag{19}$$

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

$$H_W(\omega) = \frac{P_n(\omega)}{P_x(\omega) + P_n(\omega)} \tag{20}$$

となる．これを Wiener フィルタ特性という．

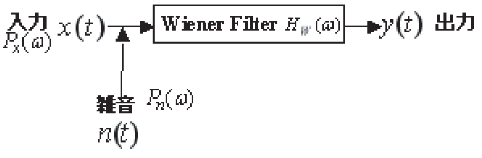

図 1: Wiener フィルタ

5.2　周波数領域 ICA と Wiener フィルタの組合せ

前節であげたように，Wiener フィルタを用いる場合，音源と雑音 (妨害音) のパワースペクトルが既知でなくてはならないが，BSS においてはどちらの情報も事前情報として与えられていない．そこで周波数領域 ICA によって分離された信号を音源信号と雑音信号として Wiener フィルタを適用する．このため ICA により分離された信号は，ある程度の分離精度を保っていることが必要となる．そのため，分離には JADE 法により分離行列 $\boldsymbol{W}(f)$ を求めて，それを初期値として設定し，FastICA を適用している．Wiener フィルタを後処理として用いた分離アルゴリズムを図 2 に示す．

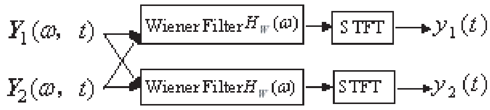

図 2: Wiener フィルタを適用したアルゴリズム

ここで図中の $\boldsymbol{Y}_1$,$\boldsymbol{Y}_2$ は周波数領域 ICA によって得られた分離結果である．Wiener フィルタを適用する際一方を目的信号，他方を雑音信号とするが，本研究で用いた周波数領域 ICA アルゴリズムではどちらが目的信号かを識別することはできない．そのため，ICA による分離信号の一方を目的信号，他方を雑音信号とし見なして Wiener フィルタを適用し，次に目的信号と雑音信号を入れ替えてフィルタを適用する．図 2 中の $\boldsymbol{Y}_1$,$\boldsymbol{Y}_2$ に適用する Filter 特性を次に示す．分離信号 1，2 のパワースペクトルをそれぞれ $P_{Y_1}(\omega)$, $P_{Y_2}(\omega)$ とすると，式 (20) より図 2 中の $\boldsymbol{Y}_1$,$\boldsymbol{Y}_2$ に適用する Wiener フィルタ特性は

$$H_1(\omega) = \frac{P_{Y_1}(\omega)}{P_{Y_1(\omega)} + cP_{Y_2}(\omega)} \tag{21}$$

$$H_2(\omega) = \frac{P_{Y_2}(\omega)}{P_{Y_2(\omega)} + cP_{Y_1}(\omega)} \tag{22}$$

となる．ここで，式中の c は妨害音の過度の減算を避けるための定数である．

6　動作評価

6.1　混合音声の収録

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 310ms である．音源（スピーカ）と観測点（マイクロホン）の配置を図 3 に示す．観測点は固定し，音源の配置は図中の θ_1=-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．混合音声の組み合わせは M+L，M+R，L+R とする．マイクロホンには素子間隔 5cm の 2 素子アレイを用いた．音源信号は，男女各

3 名ずつの 30 通りの組み合わせで，目的音声を 5 フレーズ妨害音声を 2 フレーズとし M+L，M+R，L+R の各配置それぞれ 300 通りの混合音声をについて考える．また，この音声データの長さは平均約 1.2 秒，サンプリング速度は 16Ksamples/sec で，音源の SN 比は，受音点で 0dB と設定する．

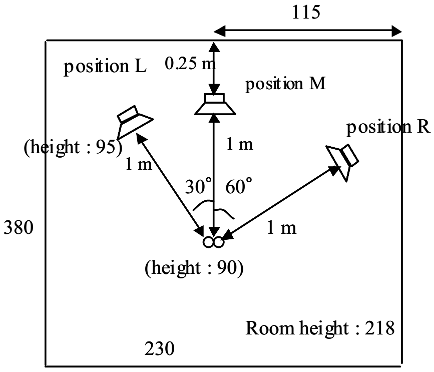

図 3: 収録環境

6.2 評価法

S/N を用いて分離信号を評価する．SN は以下のように定義した．

$$SNR = 10\log_{10}\frac{\int |S(\omega)|^2 d\omega}{\int |S(\omega) - Y(\omega)|^2 d\omega} \tag{23}$$

ここで $S(\omega)$，$Y(\omega)$ は音源信号 $s(t)$ と分離信号 $y(t)$ をそれぞれ周波数領域に変換したものである．周波数領域で考えたのは，マイクロホンを 2 つ使用しているため，各音源情報はそれぞれ 2 通り存在する．これらをもとに分離信号が得られるため，音源信号と分離信号の同期を取るのが困難となるためである．

6.3 窓長とシフト間隔

観測信号を STFT(Short Time Fourier Transform) により周波数領域に変換する際の窓長とシフト間隔について最適なものを選ぶことを考える．一般的に畳み込まれた信号の FFT を実行する際，残響時間より十分長い窓長が必要となる．しかし，残響時間のみを考慮して窓長を長くしすぎると各周波数ビンにおけるデータ数が減少し，音源信号の独立性の仮定が成立しなくなることがわかっている [3]．よってサンプリング速度 8k samples / sec，データ長 3 秒程度の混合信号に対する最適な窓長として 1024 点，シフト間隔 512 が推奨されている [10]．これらを確認するため，本研究で扱った平均データ長約 1.2 秒の短時間混合音声に対して最適な窓長の探索を試みた．窓長を 64 点〜1024 点，シフト間隔を窓長の 1/2 の長さにした結果を図 4 に示す．

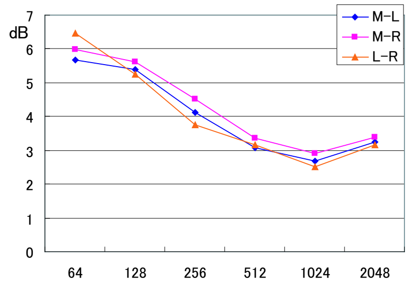

図 4: 各配置の分離結果．横軸は窓長，縦軸は S/N．

平均データ長が約 1.2 秒であることから短い窓長を使用した場合でも各周波数ビンにおけるデータ数不足となるため，各窓長においてシフト間隔を変えて実験を行った．結果を図 5 に示す．値は 3 つの配置の平均である．

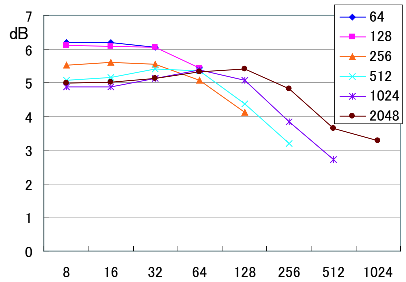

図 5: 各窓長におけるシフト間隔と SN の比較．横軸はシフト間隔，縦軸は S/N．

図 4 より，短時間音声においては，さらに窓長を短くする必要があるのがわかる．図 5 より窓長とシフト間隔を変えることによって，同一アルゴリズムを用いても 1dB の違いが出ている．また，データ不足の対策として用いた小さいシフト間隔においては窓長 256 点以上のものに対して悪くなる傾向がある．これは重複データが多くなるため相関が大きくなるからであると考えられる．しかし 64 点と 128 点においては，シフト間隔を小さくしても各フレームに占める重複幅は小さいので，長い窓長の場合より相関が小さくなると考えられる．以上のことから，短時間混合音声を分離するには，推奨される窓長とシフト間隔では不十分であり，これらをさらに小さくする必要があることがわかった．

6.4 従来法と提案法による分離精度の比較

6．3 で得られた分析窓長とシフト間隔を用いて従来法である JADE+FastICA による分離信号と，後処理として Wiener フィルタを適用した分離信号の比較を行った．図 3 における 2 音源の各配置 M+L，M+R，L+R において目的信号とする音源の位置をそれぞれ M，M，L とする．各配置において，女性+女性，男性+男性，女

性+男性（女性が目的信号），男性+女性（男性が目的信号），男女混合の 5 種類の混合音声に分別し，各種類の混合音声に対して従来法と提案法により分離し，それぞれ平均した結果を比較したものを，図 6，7，8 に示す．また，従来法と提案法において，すべての混合音声の分離結果の平均を，配置別に比較したものを図 9 に示す．

7 考察

図 6〜8 より各配置における，すべての混合信号に対して提案法によって，平均約 0.4dB の分離精度の向上が見られた．符号検定の結果，すべての種類の混合音声に対して従来法（JADE + FastICA）と提案法（従来法+Wiener フィルタ）の分離結果に危険率 0.01 以下で有意差が確認されている．また，図 9 より各配置における全混合信号の平均を比較しても分離精度が向上していることがわかる．以上のことから，提案手法では音声においては混合の種類（ここでは男女の差）や配置に関係なく同程度の精度を保持できることがわかる．

8 まとめと今後の課題

ICA による分離が困難な残響下での短時間混合音声に対して，周波数領域 ICA の後処理として Wiener フィルタを用いれば，ほぼ確実に分離精度の向上を得られることがわかった．ただし，飛躍的な向上は見られない．今後の課題として，分離された結果に残響除去のアルゴリズムを施し，認識率の向上を図ることを考える．

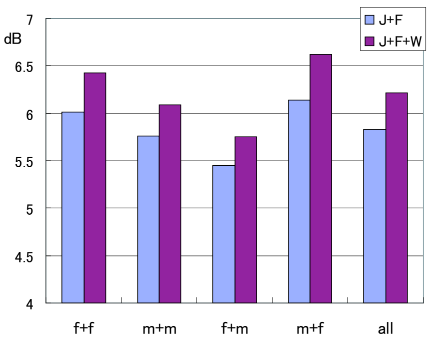

図 6: M+L での比較．横軸は男女の組合せ, 縦軸は S/N.

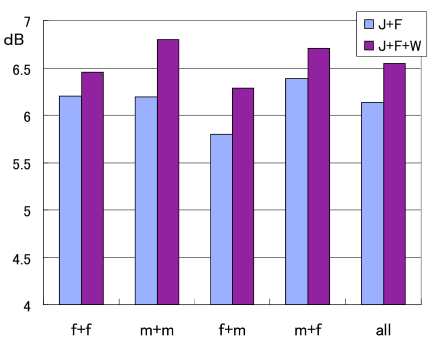

図 7: M+R での比較．横軸は男女の組合せ, 縦軸は S/N.

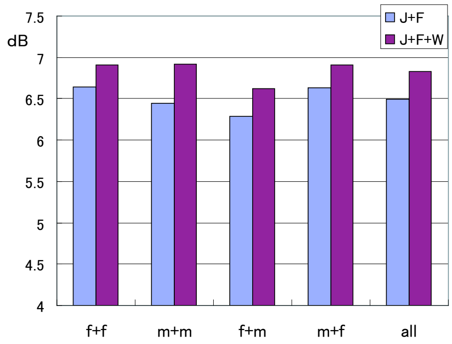

図 8: L+R での比較．横軸は男女の組合せ, 縦軸は S/N.

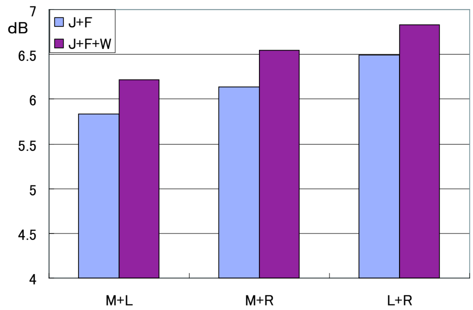

図 9: 音源の配置による比較．横軸は音源の配置, 縦軸は S/N.

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 ", Kluewer,1998.

[3] S. Araki, etc:" Fundamental Limitation of Frequency dmain Blind Source Separation for Convolved Mixture of Speech ", ICASSP2001, pp. 2737–2740, 2001.

[4] S.Ikeda, and N.Murata:" A method of ICA in time-frequency domain ", Proc WS on Independent Component Analysis ond Blind Signal Separation (ICA'99), pp.365-371, Aussios, France, Jan., 1999.

[5] J.F.Cardoso, and A.Souloumiac:" Blind beamforming for non Gaussian signals ", IEEE Proceeding-F, vol. 140, no 6, pp.362-370, Dec 1993.

[6] A.Hyvärinen, J.Karhunen, and E.Oja:" Independent Component Analysis ", John Wiley, New York, 2001.

[7] R.Prasad, H.Saruwatari, A.Lee, and K.Shikano:" A Fixed-point ICA algorithm for convoluted speech signal separation ", 4th Intern.Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003), pp.579-584, Nara, Japan, April 2003.

[8] S. Amari, etc:" 独立成分分析〜多変量解析の新しい方法〜 ", サイエンス社 pp55-62, 2002.

[9] K. Taniguchi，etc:" 信号処理の基礎 ", 共立出版, pp43-44, 2001.

[10] H. Saruwatari:" 音声・音響信号を対象としたブラインド音源分離 ", 電子情報通信学会 DSP 研究会, pp. 397 − 404, 2002.