一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SP2012-93(2012-12)

特徴量領域音源分離のためのクロススペクトル抑圧

安藤　厚志†　　丹羽　健太††　北岡　教英†　　武田　一哉†

† 名古屋大学大学院情報科学研究科
〒 464-8601 愛知県名古屋市千種区不老町
†† 日本電信電話/NTT メディアインテリジェンス研究所
E-mail: †atsushi.ando@g.sp.m.is.nagoya-u.ac.jp, ††niwa.kenta@lab.ntt.co.jp,
†††{kitaoka,kazuya.takeda}@nagoya-u.ac.jp

あらまし　複数人同時発話の音声認識には，音源分離技術が不可欠である．しかし従来の音源分離技術，特にブラインド音源分離技術は分離フィルタの更新学習を行うため，計算コストが大きいことが課題であった．そこで我々は，音声認識の特徴量領域で音源分離を行うことで，更新学習が必要な分離フィルタの数を減少させ，計算量の削減を図ろうと考えた．そのためには，特徴量領域で音源と観測信号との間に線形性が成立する必要がある．本稿では，音源と観測信号との間に線形性を成立させるための，クロススペクトルの抑圧法を提案する．提案法では，複数のマイクロホンで観測したパワースペクトルを平均化することで，クロススペクトルの抑圧を図る．提案法の結果，クロススペクトルの抑圧が確認され，提案法を用いてクロススペクトルを抑圧することで音源分離における分離信号のケプストラム歪が改善されることが示された．

キーワード　音源分離，音声認識，フィルタバンク出力，クロススペクトル

Reduction of cross spectrum for feature-domain sound source separation

Atsushi ANDO†, Kenta NIWA††, Norihide KITAOKA†, and Kazuya TAKEDA†

† Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601, Japan
†† Nippon Telegraph and Telephone Corporation/ NTT Media Intelligence Laboratories
E-mail: †atsushi.ando@g.sp.m.is.nagoya-u.ac.jp, ††niwa.kenta@lab.ntt.co.jp,
†††{kitaoka,kazuya.takeda}@nagoya-u.ac.jp

Abstract　Speech source separation is utilized for recognition of simultaneous speech. Conventional source separation methods, especially blind source separation, have a huge computational cost because they require iterative learning steps to estimate separation filters. We therefore try to separate sounds in the feature domain, and the features are then used as inputs for speech recognition, in order to reduce the number of estimated separation filters. For this purpose, linearity between sources and recorded signals is needed in the domain. In this paper, we propose a cross spectrum reduction method between sources to approximate linearity. We prove that taking the average of the power spectra over multiple microphones can reduce the cross spectrum. Experimental results showed that the proposed method could reduce the cross spectrum, and that cepstrum distortions of separated signals were also improved.

Key words　source separation, speech recognitions, filterbank outputs, cross spectrum

1. はじめに

現在，スマートフォンにおける音声入力型情報検索システムを始めとして，音声認識技術を利用したサービスが増加している．しかし現在の技術レベルでは，マイクのそばで一人の話者が発話した音声でなければ，高精度な音声認識は困難である．複数人が同時に話す音声を認識することが可能となれば，議会やミーティングの議事録の自動作成や，家族の団らんから興味のある情報を抽出するなど，音声認識技術の応用の幅が広がると考えられる．本研究の目的は，複数人が同時に発話してい

る環境下において，個々の発話を高精度に音声認識することである．

上記の目的を達成するための従来法は，音源分離技術を用いて個々の発話信号を推定し，個々の音声信号を認識するアプローチが代表的である．中でも，複数のマイクロホンから構成されるマイクロホンアレイにより観測した信号を用いて空間的な指向性を形成するビームフォーミング [1] が盛んに研究されてきた．音源の位置が既知である場合には，遅延和法 [2] や最小分散法 [3] などの様々な逆フィルタの生成法が知られている．一方で，音源の位置が未知である場合のフィルタ推定問題はブラインド音源分離 (Blind source separation : BSS) [4–6] と呼ばれ，周波数領域独立成分分析 (Frequency domain independent component analysis : FDICA) [6–8] に基づく手法が有効な逆フィルタ推定方法の一つとして知られている．我々の最終的な目標としても，音源位置を事前に知らなくても，複数の発話を同時に音声認識したいと考えている．しかしこの手法は，各周波数ビン毎にフィルタをブラインドで更新して推定するために，計算コストが大きいことが課題であった．

我々は，複数話者の音声発話から低演算量で音声認識するために，周波数領域でビームフォーミングにより分離した信号から音声認識するのではなく，特徴量領域で音源を分離できないかと考えた．音声認識の特徴量 (e.g. MFCC) は，音声信号の振幅スペクトルやパワースペクトルを周波数方向に重みづけ加算したフィルタバンク出力と呼ばれる値に基づく．このとき，パワースペクトル領域では音源と観測信号との間に線形性が成立しないため，線形フィルタを用いた音源分離は不可能である．これは，パワースペクトルを求める過程で行う短時間分析において異なる音源間で相関が生じるためであり，この異なる音源間の相関はクロススペクトルと呼ばれる．ここで，もし観測系の制御 [9] や信号処理によりクロススペクトルが除去できるならば，パワースペクトル領域で音源混合過程の線形性が成り立つ観測信号を得られると考えられる．広帯域に渡ってそれを達成できた場合，パワースペクトルを周波数方向に重み付け加算したフィルタバンク出力領域でも同様に，音源の混合過程に線形性が成り立つことになるだろう．このとき，フィルタバンクの次元数分の分離フィルタを用意することで分離処理が可能となるため，複数話者の音声認識の特徴量を低演算量で推定できるのではないかと考えた．本稿では，パワースペクトル領域での音源混合過程に線形性が成り立つためのクロススペクトルの削減法について議論を進める．

本論文の構成は以下の通りである．2 節では，どの領域での音源分離を試みるかについて述べる．3 節では，パワースペクトル領域における混合モデルの線形性と，クロススペクトル抑圧の先行研究を述べる．4 節では，クロススペクトル削減による混合モデルの線形近似法として，多観測信号平均化を提案する．5 節では，提案法によりクロススペクトル抑圧を行った音源混合信号に対し，パワースペクトル領域とフィルタバンク出力領域での音源分離実験を行い，提案法の分離性能への影響を調査する．

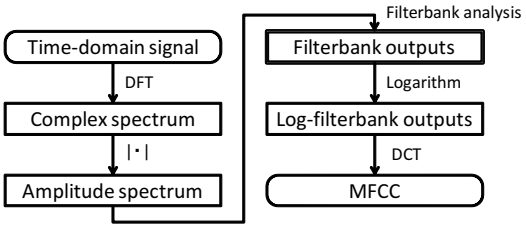

図 1　MFCC とフィルタバンク出力の関係.

2.　音源分離を行う領域

前節で述べたように，音声認識で用いる特徴量に近い領域で音源分離を行うことで，推定しなければならない分離フィルタの数が減り，音源分離の低演算量化が達成できると考えられる．

分離を行う領域は，1)MFCC にできるだけ近い領域であること，2) 音源と観測信号との関係性が明瞭であること，の二点を満たすことが望ましい．上記を踏まえ，本研究ではパワースペクトルフィルタバンク出力領域での音源分離を試みる．フィルタバンク出力は，MFCC の抽出過程において得られる信号である (図 1)．なお，MFCC は振幅スペクトルに対し求めるのが一般的であるが，パワースペクトルから求めた MFCC を用いて音声認識を行う先行研究 [11, 12] も存在し，いずれの方法を用いて MFCC を抽出する場合でも，認識率に大きな差異は表れていない．定式化における簡潔さのため，本稿ではパワースペクトルフィルタバンク出力を用いる．

ただし，パワースペクトルフィルタバンク出力領域では音源と観測信号との間に線形性が成立しないため，音源の分離は困難である．これは，パワースペクトル領域においてクロススペクトルと呼ばれる音源間の相関に関する項が生じるためである．

3.　パワースペクトル領域における線形性

本節では，パワースペクトル領域における混合モデルとクロススペクトルとの関係，クロススペクトル除去の先行研究について述べる．

3.1　パワースペクトル領域の混合モデル

M 個のマイクロホン，K 個の音源が存在する場合を考える．周波数領域における m 番目のマイクロホンでの観測信号を $X_m(\omega, t)$，k 番目の音源を $S_k(\omega, t)$ とする．ω は周波数のインデックスである．音源は互いに無相関な時系列とする．m 番目のマイクと k 番目の音源間の伝達関数の周波数特性を $A_{m,k}(\omega)$ とする．周波数領域での観測信号と音源との関係は以下のように表される．

$$X_m(\omega, t) = \sum_{k=1}^{K} A_{m,k}(\omega) S_k(\omega, t) \qquad (1)$$

このとき，音源と観測信号間で線形性が成立していることから，$A_{m,k}(\omega)$ の逆特性をもつ線形フィルタを推定することで音源の分離が可能となる．

一方で，観測信号のパワースペクトル $|X_m(\omega, t)|^2$ と音源のパワースペクトル $|S_k(\omega, t)|^2$ の関係は以下の通りとなる．

$$|X_m(\omega,t)|^2 = \left|\sum_{k=1}^{K} A_{m,k}(\omega)S_k(\omega,t)\right|^2 = \sum_{k=1}^{K}|A_{m,k}(\omega)|^2|S_k(\omega,t)|^2 + C_m(\omega,t), \quad (2)$$

ただし，

$$C_m(\omega,t) = 2\sum_{k=1}^{K}\sum_{k'\neq k}^{K}|A_{m,k}(\omega)S_k(\omega,t)||A_{m,k'}(\omega)S_{k'}(\omega,t)| \cdot\cos(\angle A_{m,k}(\omega)S_k(\omega,t) + \angle A_{m,k'}(\omega)S_{k'}(\omega,t)) \quad (3)$$

は，クロススペクトルと呼ばれる，異なる二つの音源信号の相関項の総和である．また，$\angle$ は位相を表す．式 (2) から，パワースペクトル領域では音源と観測信号の間に線形性が成立しないため，線形フィルタによる音源分離は不可能である．しかし，クロススペクトル $C_m(\omega,t) \approx 0$ が成り立つならば，パワースペクトル領域での音源分離が可能となると考えられる．

3.2 クロススペクトル抑圧の先行研究

クロススペクトルの抑圧法として，北岡らは時間方向スムージング [10] を提案した．これは，音源ごとの位相差がフレーム間で無相関であり，かつ隣接するフレームでは音源がほぼ定常と仮定し，隣接する T フレームでパワースペクトルの平均化を行うことでクロススペクトルを抑圧する手法である．この手法はスペクトルサブトラクション法の性能向上を目的として考案され，音声認識実験においてその有効性が確かめられている．

この手法は中心極限定理に基づいている．音源が互いに無相関であるとき，クロススペクトルの平均は 0 である．このとき，クロススペクトルの標準偏差を σ とすると，中心極限定理から，クロススペクトルの T フレーム平均 $\overline{C_m(\omega,t)} = \{C_m(\omega,t)+C_m(\omega,t+1)+\cdots+C_m(\omega,T-1)\}/T$ は，平均 0，標準偏差 $\sigma/\sqrt{T}$ となる．つまり，時間方向に平均化したクロススペクトルは，通常のクロススペクトルに比べ瞬時値の大きさが減少しているとみなすことができる．

しかし北岡らの手法では，クロススペクトルを限りなく 0 に近づけるのは困難である．クロススペクトルの時間平均の標準偏差は $1/\sqrt{T}$ の割合で削減することができるが，T を大きくすると音源の定常性の仮定が崩れてしまうためである．

4. 提 案 手 法

4.1 提案手法の概要

北岡らの手法は，位相差が時間的に無相関であるクロススペクトルを平均化することで，その分散を減少させるという手法であった．我々はこの考えを拡張し，複数のマイクロホンで得たパワースペクトルの平均化，すなわちパワースペクトルの空間方向の平均化によりクロススペクトルを削減する手法を提案する．

式 (3) において，位相差を $\theta(\omega,t,k,k',m)$ とする．このとき，複素数の偏角の加法性から，

$$\theta(\omega,t,k,k',m) = \angle A_{m,k}(\omega)S_k(\omega,t) + \angle A_{m,k'}(\omega)S_{k'}(\omega,t) = \angle A_{m,k}(\omega) + \angle S_k(\omega,t) + \angle A_{m,k'}(\omega) + \angle S_{k'}(\omega,t) \quad (4)$$

と表せる．すなわち，クロススペクトルにおける位相差は，伝達関数 $A_{m,k}(\omega)$ に関する項と音源 $S_k(\omega)$ に関する項の二つにより決定することが分かる．従って，ある音源から異なる二つのマイク m, m' への伝達関数の位相差が無相関であると仮定すると，$\theta(\omega,t,k,k',m)$ と $\theta(\omega,t,k,k',m')$ もまた無相関となると考えられる．この仮定は，空間中にマイクがランダムに配置されている，すなわち分散マイクロホンアレイのようなマイクロホン配置であるときに妥当である．

上記を用いて式 (3) を整理すると，

$$C_m(\omega,t) = 2\sum_{k=1}^{K}\sum_{k'\neq k}^{K}\alpha_{k,k'}^{(m)}(\omega)|S_k(\omega,t)||S_{k'}(\omega,t)| \cdot\cos\theta(\omega,t,k,k',m), \quad (5)$$

ここで，$\alpha_{k,k'}^{(m)}(\omega) = |A_{m,k}(\omega)||A_{m,k'}(\omega)|$ とした．式 (5) を M 本のマイクに対し平均化すると，

$$\frac{1}{M}\sum_{m=1}^{M}C_m(\omega,t) = 2\sum_{k=1}^{K}\sum_{k'\neq k}^{K}|S_k(\omega,t)||S_{k'}(\omega,t)| \cdot\frac{1}{M}\sum_{m=1}^{M}\alpha_{k,k'}^{(m)}(\omega)\cos\theta(\omega,t,k,k',m) \quad (6)$$

となる．式 (6) において m に関する部分のみを考えると，$\alpha_{k,k'}^{(m)}(\omega)\cos\theta(\omega,t,k,k',m)$ の m に対する平均化は，$\alpha_{k,k'}^{(m)}(\omega)$ が m の変化に対し一定である場合，平均ゼロ，分散は $1/\sqrt{M}$ に減少する (実際には $\alpha_{k,k'}^{(m)}(\omega)$ は m により変化するため，分散は $1/\sqrt{M}$ よりも大きな値をとる)．以上から，式 (2) で表されるパワースペクトルを各マイクロホン m で平均化することにより，平均化されたクロススペクトルはゼロ平均かつ極めて小さい分散をとる．したがって，クロススペクトルを無視することができ，音源と観測信号の間に近似的に線形性が成立すると考えられる．

4.2 相互相関項低減の検証実験

提案法によりどの程度相互相関項が減少するかを調査する実験を行った．

4.2.1 実 験 条 件

本実験は全てシミュレーション上で行われるものとする．実験の流れは以下の通りである．まず，音源 n とマイクロホン m の間の伝達関数 $A_{mn}(\omega)$ をランダムに作成する．各音源に対し伝達関数を畳み込み観測信号を作成し，そのパワースペクトルを求め，各マイクロホンで観測したパワースペクトルを得る．それらを全てのマイクで平均化し，平均化信号に含まれるクロススペクトルの大きさを評価する．

本実験において，音源の総数 $N = 2$ とし，音源として互いに無相関なインパルスのペア，白色雑音のペア (各 1 秒)，音声のペア (男女各一発話，各 5 秒程度) の 3 種類を用いた．平均化を行うマイクロホンの数 M は 1 から 100 まで 3 刻みで変化

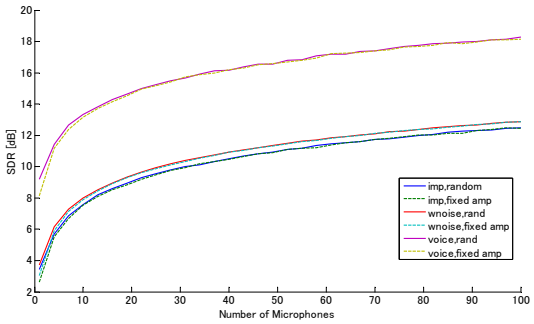

図 2　パワースペクトル平均化を行うマイクロホンの数と信号対歪み比との関係.

させ，それぞれの点でのクロススペクトルの大きさを評価した．また伝達関数の作成法として，1) 平均 0，分散 1 の複素正規分布から抽出，2) 複素平面において振幅が全て 1，位相がランダムである分布から抽出，の二通りを用いた．サンプリングレートは 16kHz，フレーム長 16ms，シフト幅 8ms とし，窓関数はハニング窓とした．

評価尺度として，信号対歪み比 (Signal-to-distortion ratio : SDR) を用いる．ただし本実験においては，信号を各音源に起因するパワースペクトル，歪みをクロススペクトルとして扱う．

$$SDR_k = 10\log_{10}\frac{\sum_\omega\sum_t\sum_m\sum_k|A_{m,k}(\omega,t)S_k(\omega,t)|^2}{\sum_\omega\sum_t|\sum_m\{|X_m(\omega,t)|^2-\sum_k|A_{m,k}(\omega)S_k(\omega,t)|^2\}|} \tag{7}$$

4.2.2　実験結果

平均化を行うマイクロホン数と SDR の関係を図 2 に示す．

図 2 から，平均化を行うマイクロホン数が増加することで SDR が対数的に増加することが分かる．これは，中心極限定理において平均化する個数 M に対し標準偏差が $1/\sqrt{M}$ となることと一致しており，平均化したクロススペクトルの標準偏差が減少したことを表すと考えられる．また，平均化を行う前の SDR は音源により異なるものの，平均化により向上する SDR の大きさは音源に殆ど依存せず，$M=10$ で約 5dB，$M=100$ で約 10dB 向上することが分かる．実際には，マイクロホンを 100 本用いて収録を行うのは現実的ではなく，またクロススペクトルがゼロ近似できるほど抑圧出来ているかは分からないが，先行研究である時間方向平滑化と組み合わせることで，少数のマイクでも充分なクロススペクトル抑圧ができる可能性がある．

5.　音源分離可能性の検証

5.1　分離問題の再定式化

始めに，パワースペクトル領域での音源分離問題を考える．(2) 式を行列形式で書くと，以下の通りとなる．

$$\mathbf{X}(\omega,t)=\mathbf{A}(\omega)\mathbf{S}(\omega,t)+\mathbf{C}(\omega,t) \tag{8}$$

ただし，$\mathbf{X}(\omega,t)=[|X_1(\omega,t)|^2,\cdots,|X_M(\omega,t)|^2]^{\mathrm{T}}$ は観測信号ベクトル，$\mathbf{S}(\omega,t)=[|S_1(\omega,t)|^2,\cdots,|S_K(\omega,t)|^2]^{\mathrm{T}}$ は音源ベクトル，$\mathbf{C}(\omega,t)=[|C_1(\omega,t)|^2,\cdots,|C_M(\omega,t)|^2]^{\mathrm{T}}$ はクロススペクトルのベクトルを表し，$\mathbf{A}(\omega)$ の要素 (m,k) は $|A_{m,k}(\omega)|^2$ である．式 (8) の両辺からクロススペクトルを引き，$\mathbf{A}(\omega)$ の逆行列 $\mathbf{A}(\omega)^{-1}$ を左から作用させることで，推定音源 $\mathbf{Y}(\omega,t)=[|Y_1(\omega,t)|^2,\cdots,|Y_K(\omega,t)|^2]^{\mathrm{T}}$ を得る．

$$\mathbf{Y}(\omega,t)=\mathbf{A}(\omega)^{-1}\mathbf{X}(\omega,t)-\mathbf{A}(\omega)^{-1}\mathbf{C}(\omega,t) \tag{9}$$

提案法の空間平均化によりクロススペクトル $\mathbf{C}(\omega,t)=0$ となれば，$\mathbf{Y}(\omega,t)=\mathbf{S}(\omega,t)$ となり完全な分離が保証される．

次に，パワースペクトルフィルタバンク出力領域での音源分離問題を考える．式 (2) にフィルタバンクを作用させると，観測信号のフィルタバンク出力 $F_{X_m}(n,t)$ は以下のように表される．

$$\begin{aligned}F_{X_m}(n,t) &= \sum_\omega Z_n(\omega)|X_m(\omega,t)|^2 \\ &= \sum_\omega Z_n(\omega)\left\{\sum_{k=1}^{K}|A_{m,k}(\omega)|^2|S_k(\omega,t)|^2 + C_m(\omega,t)\right\} \qquad (10)\\ &= \sum_{k=1}^{K}|A_{m,k}(n)|^2F_{S_k}(n,t)+F_{C_m}(n,t) \qquad (11)\end{aligned}$$

ここで，$F_{S_k}(n,t)$ は音源のフィルタバンク出力，$F_{C_m}(n,t)$ はクロススペクトルのフィルタバンク出力を表す．n はフィルタバンクの次元，$Z_n(\omega)$ はフィルタバンク係数を表し，音声認識には一般に図 3 のようなメルスケールのフィルタバンクが用いられる．また，式 (10) から式 (11) への変形には，あるフィルタバンクに含まれる周波数帯域では伝達関数の振幅が等しいという仮定を用いている．これは，近接する周波数帯域では伝達関数に類似性が現れる [6] という事実に基づく．

式 (11) をパワースペクトルの混合モデルと同様に変形することで，パワースペクトルフィルタバンク領域の推定音源 $\mathbf{F_Y}(n,t)$ が得られる．

$$\mathbf{F_Y}(n,t)=\mathbf{A}(n)^{-1}\mathbf{F_X}(n,t)-\mathbf{A}(n)^{-1}\mathbf{F_C}(n,t) \tag{12}$$

パワースペクトル領域における混合と同様に，クロススペクトルが 0 である場合には $\mathbf{F_Y}(n,t)=\mathbf{F_S}(n,t)$ となることが予想される．

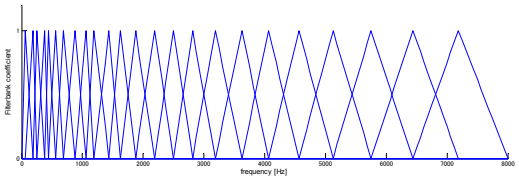

図 3　フィルタバンク係数の例.

5.2　パワースペクトル領域での音源分離実験

パワースペクトル領域での音源分離の際にクロススペクトルがどう影響するかを調査するため，パワースペクトル領域での音源分離実験を行った．

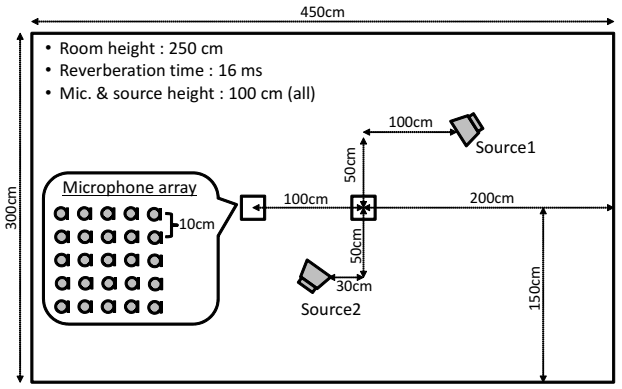

図 4　音源とマイクロホンの配置

5.2.1　実 験 条 件

本実験において，音源 n からマイクロホン m までの伝達関数は既知とする．このとき，クロススペクトルが 0 であれば，伝達関数の二乗の逆特性を観測信号に対し掛けることで完全な音源分離が可能となる．本実験では，観測信号 (クロススペクトルが大きく存在するパワースペクトル)，提案法により空間平均化 (クロススペクトルが抑圧されたパワースペクトル)，クロススペクトルなし (音源ごとのパワースペクトルの和により作成) の 3 つの入力に対し音源分離を行い，クロススペクトルが音源分離にどのように影響するかを検証する．

インパルス応答の生成には，鏡像法を用いた．このときの音源およびマイクロホンの配置を図 4 に示す．音源として，実験 1 で用いた音声ペア (男女 1 発話，各 5 秒程度) を用いた．また，サンプリングレートは 16kHz，フレーム幅 16ms，シフト幅 8ms とした．

また，音源分離時に分離信号が負の値を取る場合があった．分離信号は音源のパワースペクトルに相当することから，負の値を取ることは不適切である．したがって，分離信号が負の値を取る場合，観測信号で置換を行った．

$$Y_k(\omega,t) = \begin{cases} Y_k(\omega,t) & \text{if } Y_k(\omega,t) \geqq 0 \\ X_1(\omega,t) & \text{if } Y_k(\omega,t) < 0 \end{cases} \tag{13}$$

これは，分離信号において極端に不連続な値が現れるのを防ぐためである．

評価尺度として，スペクトル歪 (SD) 及びケプストラム歪 (CD) を用いた．音源 k に対するスペクトル歪 SD_k 及びケプストラム歪 CD_k は以下のように表される．

$$SD_k = \frac{1}{T}\sum_{t=1}^{T} 10\log_{10}\left(\sum_{\omega}\sqrt{\left(|Y_k(\omega,t)|^2 - |S_k(\omega,t)|^2\right)^2}\right) \tag{14}$$

$$CD_k = \frac{1}{T}\sum_{t=1}^{T} \frac{10}{\ln 10}\left(\sqrt{\sum_{p=1}^{P} 2\left(c_{Y_k}(p,t) - c_{S_k}(p,t)\right)^2}\right) \tag{15}$$

$c_{Y_k}(p,t), c_{S_k}(p,t)$ は分離信号のケプストラム係数，音源のケプストラム係数を表す．また，$P = 12$ とした．

表 1　パワースペクトル領域音源分離における観測信号及び分離信号のスペクトル歪とケプストラム歪．

		SD [dB]	CD [dB]
分離前		8.3	47.9
分離後	観測信号	3.3	36.5
	空間平均化 (提案法)	2.0	33.4
	クロススペクトルなし	-137.9	0.0

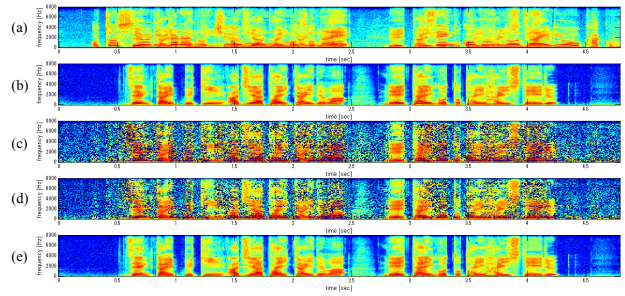

図 5　パワースペクトルの時間-周波数分布．それぞれ，(a) 観測信号，(b) 音源，(c) 観測信号をそのまま用いた時の分離信号，(d) マイク平均化を行った時の分離信号 (c) クロススペクトルが存在しない時の分離信号．また，濃い青色の部分は分離信号において負の値を取った領域を表す．

5.2.2　実 験 結 果

観測信号及び分離信号のスペクトル歪とケプストラム歪を表 1 に，パワースペクトルの時間-周波数分布の例を図 5 に示す．

図 5(c), (e) から，クロススペクトルが存在しない信号を分離すると音源とほぼ同じ信号が得られるのに対し，観測信号をそのまま分離すると音源に雑音が乗ったような信号が得られることが分かる．このことから，クロススペクトルは分離性能を低下させること，クロススペクトルが全時間，全周波数でゼロならば音源の完全な分離が行えることがいえ，式 (9) が正しいことが分かる．また (c) と (d) を比べると，空間平均化を行うことでより音源に近い分離音が得られている．これは平均化によるクロススペクトルの減少が原因であると考えられる．以上の結果は，表 1 からも得られる．すなわち，クロススペクトルが減少するにつれ，スペクトル歪とケプストラム歪が小さくなるといえる．

また分離信号の負値に注目すると，片方の音源のパワーが支配的である時間や周波数では負値が現れにくいことが分かった．このことから，負値が現れる部分は両方の音源のパワーが同程度であると考えられる．この特性を考慮し，置換する値を変えたり別の分離モデルを用いて音源分離を行うべきだといえる．

5.3　パワースペクトルフィルタバンク領域での音源分離実験

前項と同様に，パワースペクトルフィルタバンク出力領域での音源分離におけるクロススペクトルの影響，また 5.1 節における仮定 (近接する周波数帯域では伝達関数のパワーが類似) の妥当性の二点を検証する実験を行った．

5.3.1　実 験 条 件

前項と同様に伝達関数既知とし，パワースペクトルフィルタバンク出力領域において逆特性を求め音源分離を行った．インパルス応答の生成法，音源配置，使用音源及びフレーム長など

表 2　パワースペクトルフィルタバンク出力領域における観測信号及び分離信号のケプストラム歪.

		CD [dB]
分離前		47.9
分離後	観測信号	41.8
	空間平均化 (提案法)	36.1
	クロススペクトルなし	35.4

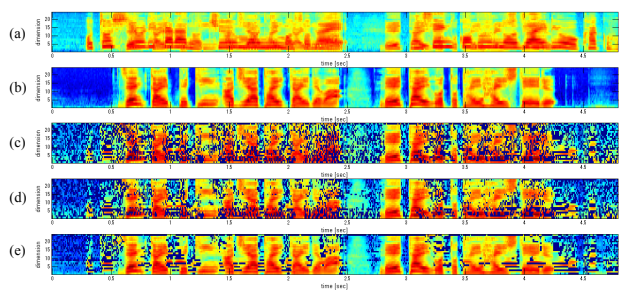

図 6　パワースペクトルフィルタバンク出力の時間-次元分布．それぞれ，(a) 観測信号，(b) 音源，(c) 観測信号をそのまま用いた時の分離信号，(d) マイク平均化を行った時の分離信号 (c) クロススペクトルが存在しない時の分離信号．また，濃い青色の部分は分離信号において負の値を取った領域を表す.

は前項と同様とした．フィルタバンク数は 24 個とし，メルスケールで等間隔に配置した．また，分離信号に負値が現れたときには，前項と同様に観測信号で置換を行った．評価尺度は前項のケプストラム歪を用いた.

5.3.2　実 験 結 果

分離信号及び観測信号のケプストラム歪を表 2 に，パワースペクトルフィルタバンク出力の時間-次元分布を図 6 に示す．図 6(e) から，パワースペクトル領域では，クロススペクトルのない信号を用いることで音源ごとに完全に分離できていたのに対し，パワースペクトルフィルタバンク出力領域では完全な分離が行えなくなることがわかる．また表 2 と前項の表 1 の全体の傾向を比較すると，パワースペクトルフィルタバンク出力領域ではパワースペクトル領域に比べ分離信号のケプストラム距離が大きくなっている．これらの結果は，フィルタバンク分析を行うときに置いた近接する周波数間で伝達関数の振幅が等しいという仮定が完全には成立しないことに起因すると考えられる．特に高次元においては，フィルタバンク分析を行う周波数帯域が極めて広い (1000Hz 程度) ため，仮定が成り立たない可能性が高い．この問題に対しては，例えば伝達特性が既知である場合，周波数間で伝達特性の振幅がフラットになるように補正を行うなどの処理を行うことでフィルタバンク出力領域においても正確な分離が可能となる可能性がある.

前項及び本項の結果から，パワースペクトルフィルタバンク出力領域での正確な音源分離のためには，1) クロススペクトルのさらなる抑圧，2) 周波数間の伝達特性の類似性の向上　の二点が必要であることが分かった．前者は時間方向平滑化との組み合わせや拡散センシング [13] の考え方を利用した伝達関数の制御を，後者は近接する周波数間での伝達関数補正を行うことで対処しようと考えている.

6. Conclusion

本稿は，特徴量領域での音源分離を行うための，クロススペクトル抑圧法の提案を行った．特徴量領域における音源分離の利点としては，分離の高速化が挙げられる．提案法では，中心極限定理の考えを用い，マイクロホンごとに得たパワースペクトルを平均化することでクロススペクトルの抑圧を図った．提案法の結果，クロススペクトルの瞬時的な大きさは平均化するマイクロホン数の対数に反比例するという結果を得た．また提案法により得た信号を音源分離にかけた場合，観測信号をそのまま音源分離にかけるよりも分離性能が良くなるという結果を得た．今後の課題として，時間方向平滑化と組み合わせたクロススペクトルのさらなる抑圧や，フィルタバンク出力領域での分離をより正確に行うための伝達関数の補正などが挙げられる.

謝辞　本研究は，科学技術振興事業団の戦略的基礎研究推進事業 CREST により行われた.

文　　献

[1] 浅野太, “音のアレイ信号処理,” コロナ社, 2011.
[2] H. L. Van Trees, “Optimum array processing,” Wiley, 2002.
[3] D. H. Johnson and D. E. Dudgeon, “Array signal processing,” Prentice Hall, Englewood Cliffs, NJ, 1993.
[4] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines, “A blind source separation technique based on second order statistics,” *IEEE Trans. Signal Processing*, vol. 45, pp.434–444, 1997.
[5] S. Araki, H. Sawada, R. Mukai and S. Makino, “Underdetermined blind source separation for arbitrarily arranged multiple sensors,” *Signal Proc.*, vol. 87, pp. 1833–1847, 2007.
[6] M. S. Pedersen *et al.*, “A survey of convolutive blind source separation methods,” *Springer Handbook on Speech,* Nov., 2007.
[7] A. Hyvärinen, “ Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Trans. on Neural Networks*, vol. 10, No. 3, pp.626–634, 1999.
[8] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain”, *Neurocomput.*, vol. 22, pp.21–34, 1998.
[9] K. Niwa, S. Sakauchi, K. Furuya, M. Okamoto and Y. Haneda, “Diffused sensing for sharp directivity microphone array,” *ICASSP 2012*, pp. 225–228, Mar., 2012.
[10] 北岡教英, 赤堀一郎, 中川聖一. ”スペクトルサブトラクションと時間方向スムージングを用いた雑音環境下音声認識”, 電子情報通信学会論文誌 (D-II), Vol. J83–D–II No. 2, pp. 500–509, Feb., 2000.
[11] 西村義隆，篠崎隆宏，岩野公司，古井貞熙，“周波数帯域ごとの重みつき尤度を用いた雑音に頑健な音声認識,” 電子情報通信学会技術報告，vol. 103, pp. 19–24, Dec., 2003.
[12] 滝口哲也，有木康雄，“Kernel PCA を用いた残響下におけるロバスト特徴量抽出の検討,” 情報処理学会論文誌, Vol. 47, No. 6, pp. 1767–1773, 2006.
[13] 丹羽健太，日岡裕輔，荒木章子，古家賢一，羽田陽一，“最大 SN 比法への拡散センシングの適用,” 日本音響学会講演論文集, pp. 761–762, Sep., 2012.