

DOCUMENT 1/2
DOCUMENT NUMBER
@: unavailable

DETAIL

JAPANESE

LEGAL STATUS

- 1. [JP.2008-026625.A](#)
- 2. [JP.2007-248975.A](#)

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2008-026625
(43)Date of publication of application : 07. 02. 2008

(51) Int. Cl. *G10L 21/02 (2006. 01)*

(21)Application number : **2006-199420** (71)Applicant : **DOSHISHA
UNIV NACIONAL DE ENTRE RIOS
UNIV NACIONAL DEL LITORAL**
(22)Date of filing : **21. 07. 2006** (72)Inventor : **DI PERSIA LEANDRO EZEQUIEL
MILONE DIEGO
YANAGIDA MASUZO**

(54) **MULTI-BIN INDEPENDENT COMPONENT ANALYSIS AND BLIND SOUND SOURCE SEPARATION DEVICE USING THE SAME**

(57)Abstract:
PROBLEM TO BE SOLVED: To propose a blind sound source separation technique which substantially eliminates the need for considering a permutation problem and can perform nearly optimum separation relating to each frequency characteristic.
SOLUTION: In the case where the sounds output from a plurality of sound sources are received in a plurality of sound receivers and a desired sound is separated, the signals of a time region when the sounds are received in the sound receivers are subjected to Fourier transformation for every analysis frame and the signals of the time frequency zone are obtained. The time variation of the spectral component for every frequency bin is therefrom extracted from the spectra in the total analysis frame, and the time variation with respect to the spectral components of a plurality of the frequency bins adjoining the certain frequency bin is linked. The linking signals for the respective sound receivers are separated to the signal in the time region in the frequency bin for each sound source.



sinc(i) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
L. Di Persia, D. H. Milone & Masuzo Yanagida; "Multi-bin Independent Component Analysis and Blind Sound Source Separation Device Using the Same"
Japan Patent Office, No. 2008-02662, 2008.

BACK NEXT
MENU SEARCH
NUMBER LIST HELP

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-26625

(P2008-26625A)

(43) 公開日 平成20年2月7日(2008. 2. 7)

(51) Int. Cl.

G10L 21/02 (2006.01)

F I

G10L 21/02 203Z

G10L 21/02 201B

テーマコード(参考)

審査請求 未請求 請求項の数 4 O L (全 15 頁)

(21) 出願番号 特願2006-199420(P2006-199420)

(22) 出願日 平成18年7月21日(2006. 7. 21)

(71) 出願人 503027931

学校法人同志社

京都府京都市上京区今出川通烏丸東入玄武町601

(71) 出願人 506092444

ユニベルシダー・ナショナル・デ・エントレ・リオス

アルゼンチン国エントレリオス州コンセプション・デル・ウルグアイ市エバ・ペロン通り24

(71) 出願人 506092466

ユニベルシダー・ナショナル・デル・リトラール

アルゼンチン国サンタフェ州サンタフェ市ペレグリーニ大通り2750番地

最終頁に続く

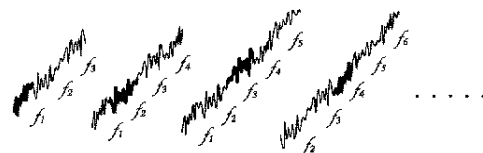
(54) 【発明の名称】 マルチピン独立成分分析およびそれを用いたブラインド音源分離装置

(57) 【要約】

【課題】パーミュテーション問題をほとんど考慮する必要がなく、しかも、各周波数特性についてはほぼ最適な分離を行うことのできるブラインド音源分離手法を提案する。

【解決手段】複数の音源から出力された音を複数の受音装置で受音して目的の音を分離する場合、まず、受音装置で受音した時間領域の信号を、分析フレーム毎にフーリエ変換して、時間周波数領域の信号を得る。そして、これから、全分析フレームにおけるスペクトルから、各周波数ビン毎のスペクトル成分の時間変化を抽出し、ある周波数ビンに隣接する複数の周波数ビンのスペクトル成分に対する時間変化を連結する。そして、各受音装置に対する連結信号を、各音源に対する当該周波数ビンにおける時間領域の信号に分離する。

【選択図】図5



【特許請求の範囲】**【請求項1】**

複数の音源から出力された音を複数の受音装置で受音し、目的の音を抽出するブラインド音源分離方法において、
受音装置によって受音した信号をフレームシフトさせながらフーリエ変換し、時間周波数領域の信号を得るステップと、
全分析フレームの周波数スペクトルから、各周波数ビン毎のスペクトル成分の時間変化を抽出するステップと、
ある周波数ビンに隣接する一定数の周波数ビンの成分に対する時間変化波形を連結して連結波形を得るステップと、
複数の受音装置についての当該連結波形を、各音源に対する周波数ビンの時間信号に分離するステップと、
当該分離された結果を各音源信号の時間周波数領域表現とみて、周波数スペクトルのフレーム毎の時間変化を得るステップと、
各音源についてのフレーム毎の時間変化をフーリエ逆変換することによって、時間領域の信号を得るステップと、
各音源についてのフレーム毎の時間波形を連結して全時間波形を得るステップを備えたことを特徴とするブラインド音源分離方法。

【請求項2】

複数の音源から出力された音を複数の受音装置で受音し、目的の音を抽出するブラインド音源分離方法において、
受音装置によって受音した信号をフレームシフトさせながらフーリエ変換し、時間周波数領域の信号を得るステップと、
全分析フレームの周波数スペクトルから、各周波数ビン毎のスペクトル成分の時間変化を抽出するステップと、
全周波数ビンの成分に対する時間変化波形を連結し、当該連結された連結波形を各音源における時間周波数領域の信号に一括分離するステップと、
ある周波数ビンに隣接する一定数の周波数ビンの成分に対する時間変化波形を連結し、当該連結された連結波形を、各音源に対する周波数ビンの時間信号に分離するステップと、
当該分離された信号を、前記一括分離された信号をもとに周波数ビン毎に並び替えるステップと、
当該並び替えられた結果の信号を各音源信号の時間周波数領域表現とみて、周波数スペクトルのフレーム毎の時間変化を得るステップと、
各音源についてのフレーム毎の時間変化をフーリエ逆変換することによって、時間領域の信号を得るステップと、
各音源についてのフレーム毎の時間波形を連結して全時間波形を得るステップを備えたことを特徴とするブラインド音源分離方法。

【請求項3】

複数の音源から出力された音を複数の受音装置で受音し、目的の音を抽出するブラインド音源分離装置において、
受音装置によって受音した信号をフレームシフトさせながらフーリエ変換し、時間周波数領域の信号を得る周波数領域変換部と、
全分析フレームの周波数スペクトルから、各周波数ビン毎のスペクトル成分の時間変化を読み取る第一の読取部と、
ある周波数ビンに隣接する一定数の周波数ビンの成分に対する時間変化波形を連結して連結波形を得る連結部と、
複数の受音装置についての当該連結波形を、各音源に対する周波数ビンの時間信号に分離する分離部と、
当該分離された結果を各音源信号の時間周波数領域表現とみて、周波数スペクトルのフレ

ーム毎の時間変化を読み取る第二の読取部と、
各音源についてのフレーム毎の時間変化をフーリエ逆変換することによって、時間領域の信号を得る時間領域変換部と、
各音源についてのフレーム毎の時間波形を連結して全時間波形を得て出力する出力部とを備えたことを特徴とするブラインド音源分離装置。

【請求項4】

複数の音源から出力された音を複数の受信装置で受信し、目的の音を抽出するブラインド音源分離装置において、
受信装置によって受信した信号をフレームシフトさせながらフーリエ変換し、時間周波数領域の信号を得る周波数領域変換部と、
全分析フレームの周波数スペクトルから、各周波数ビン毎のスペクトル成分の時間変化を読み取る第一の読取部と、
各周波数ビンを全周波数標本点について連結し、当該連結された全周波数ビンの信号を各音源における時間周波数領域の信号に一括分離するとともに、ある周波数ビンに隣接する一定数の周波数ビンの成分に対する時間変化波形を連結して連結波形を得る連結部と、
複数の受信装置についての当該連結波形を各音源に対する周波数ビンの時間信号に分離する分離部と、
該連結された連結波形を、各音源に対する周波数ビンの時間信号に分離するステップと、
当該分離された信号を、前記一括分離された信号をもとに周波数ビン毎に並び替える並び替え部と、
当該並び替えられた分離の信号を各音源信号の時間周波数領域表現とみて、周波数スペクトルのフレーム毎の時間変化を読み取る第二の読取部と、
各音源についてのフレーム毎の時間変化をフーリエ逆変換することによって、時間領域の信号を得る時間領域変換部と、
各音源についてのフレーム毎の時間波形を連結して全時間波形を得て出力する出力部とを備えたことを特徴とするブラインド音源分離装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、独立成分分析を用いた音源分離方法においてパーミュテーション問題をほとんど考慮する必要のない手法に関するものである。

【背景技術】

【0002】

現在の音声認識システムは、マイクロホン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、マイクロホン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率は著しく低下してしまう（非特許文献1）。実環境で音声認識を行うためには音声認識システムそのものの認識能力の向上とともに、それに入力される音声に対して何らかの処理を行って認識しやすい形にしておくことが有効であると考えられる。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。一般に、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、

【0003】

【数1】

$$x_m(t) = \sum_{n=1}^N h_{mn}(t) * s_n(t)$$

$h_{mn}(t)$: 音源 s_n からマイクロホン M_m へのインパルス応答

$s_n(t)$: 音源 s_n の信号

* : 畳み込み積分
(コンボリューション)

と表される。数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はこのような状況において、 $s_n(t)$ を推定する問題である。しかし、時間領域のBSSでは、計算量が膨大になり、現実的な処理時間で分離フィルタを計算するのが困難であることが指摘されている。

【0004】

そこで、一般に、このフィルタを直接求めずその周波数特性を求める周波数領域BSSを用いることが多い。マイクロホン数 M 、音源数 N の混合モデルを求めるために、数1の両辺をフーリエ変換すると、 $X_m(f) = \sum_n H_{mn}(f) S_n(f)$ と表される。ここで $X_m(f)$ 、 $H_{mn}(f)$ 、 $S_n(f)$ は、それぞれ $x_m(t)$ 、 $h_{mn}(t)$ 、 $s_n(t)$ のフーリエ変換である。そして、 $m=1 \sim M$ についてまとめると、周波数領域BSSの混合モデルは、

【0005】

【数2】

$$X(f) = H(f)S(f)$$

【0006】

と表される。ただし、 $X(f) = (X_1(f), \dots, X_M(f))^T$ は観測ベクトル、 $S(f) = (S_1(f), \dots, S_N(f))^T$ は音源信号ベクトル、そして、 $H(f)$ は周波数 f を変数とする M 行 N 列の混合行列である。

【0007】

BSSを解く手法として、各音源信号の独立性を仮定した独立成分分析(ICA)が一般的に用いられる。ICAには時間領域で処理する方法と周波数領域で処理する方法とがある。周波数領域ICAは、畳み込みを瞬時混合の問題に置き換えて解くことができるため、現在、実環境で使える手法として盛んに用いられている。

【0008】

周波数領域ICAは、 $x_m(t)$ の標準化列に対する短時間フーリエ変換により得られた観測信号ベクトル $X_m(f)$ の離散表現に対して、各周波数ビンにおいて学習された $W(f)$ を用いて分離を行う。ここで、 $Y(f) = (Y_1(f), \dots, Y_N(f))^T$ を分離された信号のベクトル表現とすると分離過程は、連続領域で

【0009】

【数3】

$$Y(f) = W(f)X(f)$$

と表される。

【0010】

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、 $W(f)$ の

行が入れ替わったとしても $Y(f)$ の独立性は保たれるので、任意の周波数 f_1 と f_2 において、 $Y_n(f_1)$ と $Y(f_2)$ が必ずしも同じ音源に対応しているとは限らない。従って、 $W(f)$ の行の並びを $Y_n(f_1)$ と $Y_n(f_2)$ が同じ音源に属するように並び替える「パーミュテーション問題」を解決する必要がある。

【0011】

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、方向推定結果を基に並び替えを行う方法がある（非特許文献5）。しかし、すべての周波数において理想的なビームあるいは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と指摘されている。

【0012】

そこで、パーミュテーション問題への対処法として各周波数ビンでの相関を用いた方法（非特許文献2）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持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い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（非特許文献2）が、この方法では、離れた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。

【0013】

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（非特許文献6）。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、未定の周波数ビン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ビンとの相関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強い場合に不安定になることが報告されている（非特許文献7）。

【0014】

残響が強い場合に不安定になる原因として、低域での方向推定の困難さが挙げられている（非特許文献7）。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（非特許文献7）。有声音は基本周波数の整数倍のところで当該周波数ビンの成分の時間波形の相関が高くなる性質を持っており、調波構造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。

【非特許文献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.

【非特許文献2】T.W.Lee：“Independent Component Analysis”，Kluwer，1998.

【非特許文献3】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, Aussiaos, France, Jan., 1999.

【非特許文献4】A.Hyvarinen, J.Karhunen, and E.Oja：“Independent Component Analysis”，John Wiley, New York, 2001.

【非特許文献5】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.

【非特許文献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問題の解法”，音響学会講演論文集, pp. 541-542, Sep, 2002.

【非特許文献7】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.

【発明の開示】

【発明が解決しようとする課題】

【0015】

しかし、これらの方法によっても本質的に並べ替えに失敗する可能性があり、パーミュテーション問題は解決されない。

【0016】

さらに、周波数ビン毎に音声を分離する場合、分離される周波数ビンの標本点数が多ければ多いほど、その周波数特性に応じた最適な分離を行うことができるが、音声分離を行う際には、この分離の最適性を確保することが必要とされる。

【0017】

そこで、本発明は上記課題に着目してなされたもので、パーミュテーション問題をほとんど考慮する必要がなく、しかも、各周波数特性についてほぼ最適な分離を行うことのできるブラインド音源分離手法を提案することを目的とする。

【課題を解決するための手段】

【0018】

すなわち、本発明は上記課題を解決するために、複数の音源から出力された音を複数の受信装置で受信して目的の音を分離する場合、まず、受信装置で受信した時間領域の信号を、分析フレーム毎にフーリエ変換して、時間周波数領域の信号を得る。そして、これから、各周波数ビン毎のスペクトル成分の時間変化を抽出し、ある周波数ビンに隣接する複数の周波数ビンのスペクトル成分に対する時間変化を連結する。そして、各受信装置に対する連結信号を各音源に対する当該周波数ビンにおける時間領域の信号に分離するようにしたものである。

【0019】

すなわち、従来の周波数領域BSSでは、受信した時間領域の信号を、図3に示すように分析フレーム毎にフーリエ変換し、これを図4に示すように、各周波数ビン毎のスペクトル成分の時間変化として抽出して、各周波数ビン毎の分離行列を用いて分離していたが、この手法では、各周波数ビン毎に異なる分離行列を用いるため、各分離された信号がいずれの音源の信号なのかを判別することができない。このため、組み合わせに自由度ができてしまい、いわゆるパーミュテーション問題を生じてしまっていた。さらに、学習された分離行列を用いて分離を行う場合、標本点の数が多ければ多いほど学習能力が高くなり、分離精度がよくなるが、従来の方法であれば分析フレームの数だけしか標本点を取ることができないため、分離の精度がよくなる。加えて、従来の手法では、周波数ビン毎に異なる分離行列を用いていたため、分離された信号の振幅に不確実性を生じてしまい、これを修正しなければならなくなる。これに対して、本発明によれば、各周波数ビン毎のスペクトル成分の時間変化を複数連結した後、一括して分離行列を用いて分離するようにしたので、組み合わせの自由度を少なくしてパーミュテーション問題をほとんど解決することができ、さらには、連結された周波数ビンに分析フレーム数を乗じた標本点の数に基づいて分離処理を行うので、分離精度もよくなる。

【0020】

また、このような発明において、全周波数ビンを連結して一括分離するとともに、各受信装置について隣接周波数ビンのスペクトル成分の時間変化を部分的に連結して各音源毎の信号に分離する。そして、一括分離された結果を参照して、各音源毎に分離された時間領域の信号を並び替える。

【0021】

このようにすれば、全周波数ビンについて連結して一括分離の結果を参照することで、パーミュテーション問題をほとんど考慮する必要がなくなり、しかも、複数の周波数ビン毎にその周波数特性を考慮して分離された信号を並び替えることで、パーミュテーション問題と分離の最適化の妥協点を見つけることができる。

【発明の効果】

【0022】

本発明によれば、複数の周波数ビンを連結して分離処理を行うので、パーミュテーション問題における組み合わせの自由度を激減させることができ、また、連結された周波数ビンに分析フレーム数を乗じた数の標本点に基づいて分離処理を行うので分離精度もよくな

る。

【発明を実施するための最良の形態】

【0023】

以下、本発明の一実施の形態について図面を参照して説明する。図1は、周波数領域BS手法を用いた装置の機能ブロック図を示したものである。以下では、音源として2つの音源S1と音源S2を用い、集音装置として2つのマイクロホンM1、M2を設けた場合を例とする。ここで、音源の種類としては、人間の声などのような可聴周波数領域の音波を出力する音源を用いる。また、マイクロホンの数は音源の数以上であればよく、3個以上のマイクロホンを設けるようにしてもよい。

【0024】

図1の機能ブロックについて説明すると、まず、受信装置1はマイクロホンやマイクロホンアレイなどの装置によって構成される。そして、図2に示す2つのビームを形成し、音源と雑音方向を推定する。このビームフォーマーを求めるために、推定音源の独立性を表す表か指標を用いる。

【0025】

この受信装置1によって受信された信号は、まず、時間領域の信号として記憶部に格納される。そして、この信号を周波数領域変換部2にて時間周波数領域の信号に変換し、同様に記憶部に格納していく。時間領域の信号を時間周波数領域の信号に変換する場合、分析フレームをオーバーラップさせて少しずつずらしながら各分析フレーム内で短時間フーリエ変換(STFT)する。ここでは分析フレームの個数を m 個としている。これにより、図3や図4に示すような時間周波数領域における離散的な信号が得られる。この領域のうち時間軸上においては、 m 個の分析フレームに基づく離散的な信号が存在し、また、周波数軸上においては、 n 個の離散的な信号が存在する。なお、図3と図4は同じ信号空間を示しており、図3は分析フレーム毎の周波数スペクトルとして、また、図4は各周波数ビンの成分の時間変化として示している。

【0026】

第一の読取部3は、この記憶部に格納されている時間周波数領域内の信号から、各周波数ビン毎におけるスペクトル成分の時間変化を読み取る。すなわち、分析フレーム毎に図3のように並べられた離散的な信号を、今度は、図4に示すように周波数ビン毎の信号として読み取る。

【0027】

連結部4では、この図5に示すように、各周波数ビン毎のスペクトルの時間変化を強制的に連結した信号を得る。この連結に際しては、ある周波数ビンに隣接する予め定められた個数の周波数ビンを連結する。このとき、ある周波数 f_k のスペクトル成分における $t=t_1 \sim t_m$ の時間変化を表す信号の前に、周波数 f_{k-1}, f_{k-2} のスペクトル成分における $t=t_1 \sim t_m$ の時間変化を表す信号を低周波数側から順に連結し、また、周波数 f_k のスペクトル成分の時間波形の後に、周波数 f_{k+1}, f_{k+2} のスペクトル成分の波形を順に連結する。そして、この連結された周波数ビンのスペクトル成分の時間は計を後述する分離部5によって分離処理し、その中心となる周波数 f_k におけるスペクトル成分の時間変化を各音源の独立性が最大となるようにして求める。また、同様にして、全ての周波数ビンのスペクトル成分の時間変化を連結し、一つの連結された信号とする。そして、次の分離部5を用いて一括した分離処理を行う。

【0028】

分離部5は、連結された周波数ビンの時間信号を、各音源の時間信号に分離する。この分離手法としては、JADEやFAST ICAなどの手法が存在するが、JADEを用いた場合について説明する。

【0029】

JADEとは、Cardosoらにより提案された手法で、4次クロススキュムラントを対角化する手法である。4次クロススキュムラントは次のように定義される。

【0030】

【数4】

$$\begin{aligned} & \text{cum}(x_i, x_j, x_k, x_l) \\ &= E(x_i, x_j, x_k, x_l) - E(x_i, x_j)E(x_k, x_l) \\ & \quad - E(x_i, x_k)E(x_j, x_l) - E(x_i, x_l)E(x_j, x_k) \end{aligned}$$

【0031】

観測信号 $x(t)$ を平均0として無相関化したものを $z(t)$ とし、音源信号 $s(t)$ と無相関化された観測信号 $z(t)$ はある直交行列 $U = (u_1, \dots, u_N)$ により、

【0032】

【数5】

$$z = Us$$

という関係で結ばれている。独立性の仮定から、

【0033】

【数6】

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

【0034】

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

【0035】

【数7】

$$C_{i,j}(M) = \left(\sum_{k,l=1}^N \text{cum}(z_i, z_j, z_k, z_l) m_{kl} \right)$$

【0036】

を考える。 s のインデックスが全て同じ時だけがゼロでない値をもつから、 $\kappa_i = \text{cum}(s_i, s_i, s_i, s_i)$ として数7を変形すると

【0037】

【数8】

$$C(M) = \sum_{i=1}^N (\kappa_i w_i^T M w_i) w_i w_i^T$$

となる。ここで

【0038】

【数9】

$$D(M) = \text{diag}(\kappa_1 w_1^T M w_1, \dots, \kappa_N w_N^T M w_N)$$

とすると、

【0039】

【数10】

$$C(M) = WD(M)W^T$$

直交行列であるWで対角行列を挟んでいるので、C(M)は対称行列になっている。このC(M)を対角化するW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は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。

【0040】

周波数上でICAを行う場合は、各周波数ビンに対して分離行列W(f)を求める必要があり、通常はここで、パーミュテーションが問題となる。しかし、提案法において一括分離を行う場合は、分離行列を求めるのは一度でよいので、パーミュテーションは問題とならない。

【0041】

分割部6は、分離部5によって分離された信号を周波数ビン毎のスペクトルの時間変化に戻す。すなわち、この処理においては、連結された信号を個々の周波数ビンの分離信号に分割し、そのうち中心となる周波数ビンの信号波形を抽出して記憶部に格納する。また、同様に、連結された信号波形に対して、それを個々の周波数ビンの時間信号に分割し、記憶部に格納する。

【0042】

並び替え部7は、この一括分離された信号をもとに、必要ならば個々に分離された信号の入れ替え処理を行う。すなわち、一括分離された信号のうち、分析フレームにおける分離信号がどちらの音源に対応しているかを読み取り、これに類似するように個々に分離された信号を読み出して音源毎に並び替えていく。そして、この並び替えられた信号を、時間周波数領域に並べていく。

【0043】

第二の読取部8は、この時間周波数領域に並べられた信号を周波数スペクトルの分析フレーム毎の変化として読み取り、これを時間領域変換部9にて分析フレーム毎にフーリエ逆変換することによって時間領域の信号に戻す。そして、この時間信号に戻された信号を、分析フレーム順に連結し、各音源毎に分離された信号として出力部10から出力する。

【0044】

次に、このように構成された周波数領域BSS手法を用いた処理のフローチャートについて図6を用いて説明する。

【0045】

まず、受信装置1で受信された信号を(ステップS1)、時間領域の波形として記憶部に記憶する。そして、周波数領域変換部2により、分析フレーム毎に短時間フーリエ変換を行うとともに(ステップS2)、すべての分析フレームについて短時間フーリエ変換を行うまでこの処理を続ける(ステップS3)。このとき、図3に示すように、時間周波数領域内において分析フレーム分(m個)の離散的な周波数スペクトルが並べられる。次に、この並べられた信号を、図4に示すように、各周波数ビン毎のスペクトル成分の時間変化として読み出し(ステップS4)、各周波数ビンでのスペクトル成分の時間変化を全周波数ビンについて連結する(ステップS5)。その後、受信装置毎に連結された周波数成分の時間信号をJADEを用いて一括して各音源毎に分離し(ステップS6)、その分離された時間周波数領域での信号を各周波数ビンへの信号へ分割する(ステップS7)。

【0046】

また、対象とする周波数ビンの左右に隣接する所定の個数(例えば、2個)の周波数ビンを連結し(ステップS8)、5個の連結された周波数ビンの信号を対象としてJADEを用いて分離する(ステップS9)。このとき、帯域の両側の周波数ビンにおける周波数ビンについては、左右に規定の同数の周波数ビンを連結することができないため、図5に示すように、利用可能な周波数ビンのみを連結するか、もしくは、図7に示すように、規定より小

さい数で左右同一の周波数ビンだけを用いて分離処理を行う。この段階での分離結果は、対象とする周波数ビンに対するものと考えて、当該部分のみを各音源の周波数ビンに関する時間波形とみなす(ステップS10)。そして、この抽出された分離信号が、一括分離された信号においてどちらの音源に属しているのかを判定し、この判定結果に基づいて分離信号を各音源に並び替えていく(ステップS11)。そして、分離された時間周波数領域の信号を分析フレーム毎の周波数スペクトルの時間経過として読み出し、分析フレーム単位でフーリエ逆変換することによって時間領域の信号に戻す(ステップS12)。そして、このように分離された音声を連結して出力する(ステップS13)。

【0047】

上述のように本実施の形態によれば、各周波数に隣接する周波数ビンを連結して分離するようにしたので、隣接周波数における標本点数を用いることによって実効的に時間軸上の標本点数(分析フレーム数)を多くした状態で分離処理することができ、周波数特性に応じた精度良い分離処理を行うことができるようになる。また、このとき、全周波数ビンを一括分離した結果を参照し、各分離された信号がいずれの音源に属するのかを振り分けるようにしたので、分離精度の維持を図りながらパーミュテーション問題を解決することができるようになる。

【実施例】

【0048】

以下に、上記実施の形態を用いたシステムの評価結果について説明する。

【0049】

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8に示す。観測点は固定し、音源の配置は図中の $\theta = -30^\circ, 0^\circ$ (時計回りを正)となるようにし、それぞれの配置をL, M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1名ずつで、女性(配置:M)による10フレーズ、男性(配置:L)による10フレーズの組み合わせで、合計100個の混合音声について考える。女性および男性の音声を個別に収録し、S/Nの調整を行うことなくそのまま加算した。

【0050】

2. 評価指標

S/Nを用いて分離信号を評価する。S/Nは以下のように定義した。

【0051】

【数11】

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

$S(\omega)$ と $\hat{S}(\omega)$ は音源信号 $s(t)$ と分離信号 $\hat{s}(t)$ をそれぞれ周波数領域に変換したもの

【0052】

3. 評価対象

比較対象のアルゴリズムは、JADE法によって周波数領域ICAを行い、パーミュテーションに関しては、周波数ビン間の相関を用いるIkeda et al. (非特許文献2)の手法である。

【0053】

4. 評価結果

評価は、比較対象のアルゴリズムでパーミュテーション問題の解決ができない例に対して、提案法では解決できていることを、スペクトログラムをスペクトル上での比較によって行う。音響分析条件を表1に示す。

【0054】

【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	512 点
フレームシフト	256 点
分析窓	ハミング窓

【0055】

図9の縦軸は分離処理後のS/N(信号対雑音比)で、値の大きい方が分離がいいことを示し、横軸は処理方式を表す。各方式に、左から女性、男性、平均の結果が示されている。左側は通常の周波数領域ICAで、パーミュテーション誤りがいくつかあるためにS/Nがそれほどよくない。その右は一括分離による方法で、各周波数ビンについての最適性が失われていることによってS/Nは男性ではむしろ下がっている。その右が本手法で、同時に処理する左右の周波数ビンの片側の数nを示す。図は512点FFTを用い、nを1,2,3,4,5,10,15,20としたときのS/Nとその平均である。男性については隣接高調波までの間隔が狭く、同時処理する周波数ビンの数nは多くなくてもいい(n=4~5で最適)が、女性は基本周波数が高いため、nを多くしないと分離性能が上がらない。

【図面の簡単な説明】

【0056】

【図1】本発明の一実施の形態におけるパーミュテーションフリー・ブラインド音源分離装置の機能ブロック図

【図2】同形態において推定すべき指向特性を示す図

【図3】同形態における分析フレーム区間毎にSTFTした値を並べた図

【図4】図3における信号を各周波数ビン毎にみた図

【図5】同形態における隣接する周波数ビンを連結した状態を示す図

【図6】同形態における処理のフローチャートを示す図

【図7】同形態における隣接する周波数ビンを連結した状態を示す図

【図8】本発明の実施例におけるデータ収録環境を示す図

【図9】同実施例における全分離結果のセグメンタルSNRの平均値の比較を示す図

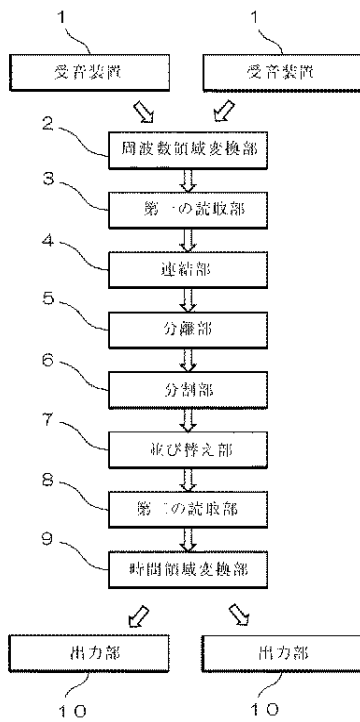
【符号の説明】

【0057】

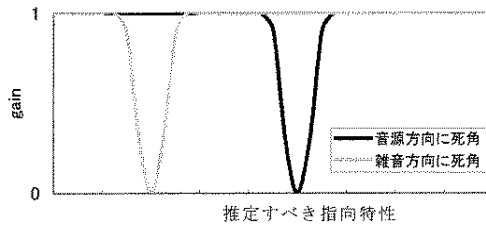
- 1・・・受音装置
- 2・・・周波数領域変換部
- 3・・・第一の読取部
- 4・・・連結部
- 5・・・分離部
- 6・・・分割部
- 7・・・並び替え部
- 8・・・第二の読取部
- 9・・・時間領域変換部

1 0 . . . 出力部

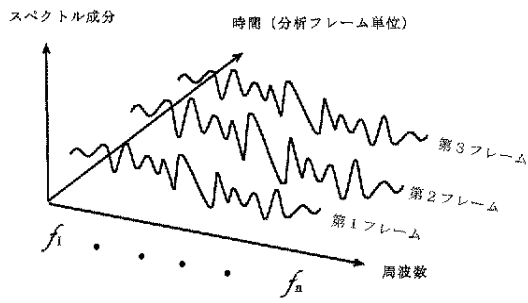
【図1】



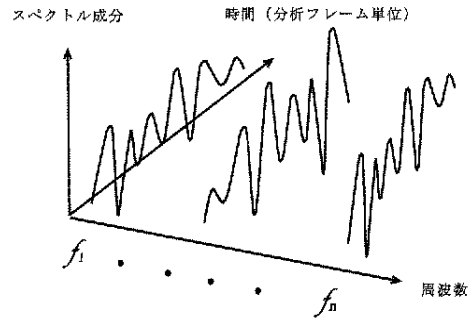
【図2】



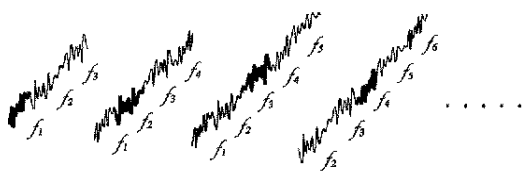
【図3】



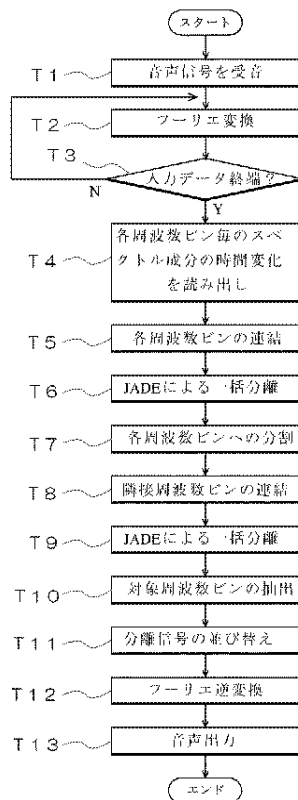
【図4】



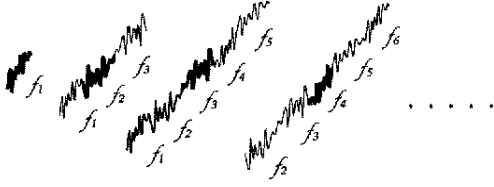
【図5】



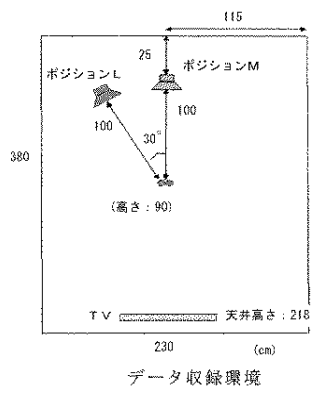
【図6】



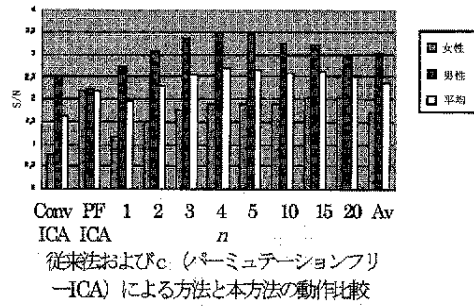
【図7】



【図8】



【図9】



- (74)代理人 100111349
弁理士 久留 徹
- (72)発明者 レアンドロ・エセキエル・ディ・ペルシア
アルゼンチン国エントレリオス州パラナー市ディアマンテ通り320番地9号室
- (72)発明者 ディエゴ・ミローネ
アルゼンチン国サンタフェ州サンタフェ市リバダビア通り2631番地2号室
- (72)発明者 柳田 益造
京都府京田辺市多々羅都谷1-3 同志社大学工学部内