

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4496186号
(P4496186)

(45) 発行日 平成22年7月7日(2010.7.7)

(24) 登録日 平成22年4月16日(2010.4.16)

(51) Int. Cl. F I
G 1 0 L 21/02 (2006.01)
 G 1 0 L 21/02 2 0 1 D
 G 1 0 L 21/02 2 0 2 B
 G 1 0 L 21/02 2 0 3 Z

請求項の数 8 (全 29 頁)

(21) 出願番号 特願2006-241861 (P2006-241861)
 (22) 出願日 平成18年9月6日(2006.9.6)
 (65) 公開番号 特開2007-219479 (P2007-219479A)
 (43) 公開日 平成19年8月30日(2007.8.30)
 審査請求日 平成19年3月1日(2007.3.1)
 (31) 優先権主張番号 特願2006-14419 (P2006-14419)
 (32) 優先日 平成18年1月23日(2006.1.23)
 (33) 優先権主張国 日本国(JP)

(73) 特許権者 000001199
 株式会社神戸製鋼所
 兵庫県神戸市中央区脇浜町二丁目10番2
 6号
 (73) 特許権者 504143441
 国立大学法人 奈良先端科学技術大学院大
 学
 奈良県生駒市高山町8916-5
 (74) 代理人 100084135
 弁理士 本庄 武男
 (72) 発明者 稗方 孝之
 兵庫県神戸市西区高塚台1丁目5番5号
 株式会社神戸製鋼所 神戸総合技術研究所
 内

最終頁に続く

(54) 【発明の名称】 音源分離装置、音源分離プログラム及び音源分離方法

(57) 【特許請求の範囲】

【請求項1】

所定の音響空間に複数の音源と複数の音声入力手段とが存在する状態で該音声入力手段各々を通じて入力される前記音源各々からの音源信号が重畳された複数の混合音声信号から、1以上の前記音源信号を分離した分離信号を生成する音源分離装置であって、

独立成分分析法に基づくブラインド音源分離方式の音源分離処理により複数の前記混合音声信号から1以上の前記音源信号に対応するS I M O信号を分離生成する第1の音源分離手段と、

前記第1の音源分離手段で分離生成された前記S I M O信号の全部若しくは一部である複数の特定信号について、複数の区分された周波数成分ごとに予め設定された重み係数により該周波数成分ごとの信号レベルを重み付けして補正し、その補正後の信号について前記周波数成分ごとに選択処理若しくは合成処理を行う所定の間中処理を行うことにより中間処理後信号を得る中間処理実行手段と、

前記中間処理実行手段により得られた複数の前記中間処理後信号に、又は該中間処理後信号及び前記第1の音源分離手段で分離生成された前記S I M O信号の一部の信号に、バイナリーマスキング処理を施すことにより得られる信号を前記音源信号に対応する前記分離信号とする第2の音源分離手段と、

前記重み係数を所定の操作入力に従って設定する中間処理パラメータ設定手段と、
 を具備してなることを特徴とする音源分離装置。

【請求項2】

10

20

前記中間処理実行手段が、前記補正後の信号の中から前記周波数成分ごとに信号レベルが最大のものを選択する処理を行うものである請求項1に記載の音源分離装置。

【請求項3】

前記第1の音源分離手段が、

時間領域における複数の前記混合音声信号に短時間離散フーリエ変換処理を施して周波数領域における複数の混合音声信号へ変換する短時間離散フーリエ変換手段と、

前記周波数領域における複数の混合音声信号に対し所定の分離行列に基づく分離処理を施すことにより前記混合音声信号ごとに前記音源信号のいずれかに対応した第1の分離信号を生成するFDICA音源分離手段と、

前記周波数領域における複数の混合音声信号各々から当該混合音声信号に基づき前記FDICA音源分離手段により分離された前記第1の分離信号を除く残りの前記第1の分離信号を減算した第2の分離信号を生成する減算手段と、

前記第1の分離信号及び前記第2の分離信号に基づく逐次計算により前記FDICA音源分離手段における前記分離行列を計算する分離行列計算手段と、を具備する周波数領域SIMO独立成分分析法に基づくブラインド音源分離方式の音源分離手段である請求項1又は2のいずれかに記載の音源分離装置。

10

【請求項4】

前記第1の音源分離手段が、周波数領域独立成分分析法と逆射影法との連結手法に基づくブラインド音源分離方式の音源分離処理を行うものである請求項1～3のいずれかに記載の音源分離装置。

20

【請求項5】

前記第1の音源分離手段が、時系列に入力される前記混合音声信号が予め定められた周期で区分された区間信号ごとに、該区間信号に対し所定の分離行列に基づく分離処理を順次実行して前記SIMO信号を生成するとともに、該分離処理により生成した前記区間信号の時間帯に対応する全ての時間帯の前記SIMO信号に基づいて以降に用いる前記分離行列を求める逐次計算を行うものであって、該逐次計算の回数を前記予め定められた周期の時間内で実行可能な回数に制限してなる請求項1～4のいずれかに記載の音源分離装置。

【請求項6】

前記第1の音源分離手段が、時系列に入力される前記混合音声信号が予め定められた周期で区分された区間信号ごとに、該区間信号に対し所定の分離行列に基づく分離処理を順次実行して前記SIMO信号を生成するとともに、該分離処理により生成した前記区間信号の時間帯のうちの先頭側の一部の時間帯に対応する前記SIMO信号に基づいて以降に用いる前記分離行列を求める逐次計算を前記予め定められた周期の時間内で実行してなる請求項1～5のいずれかに記載の音源分離装置。

30

【請求項7】

所定の音響空間に複数の音源と複数の音声入力手段とが存在する状態で該音声入力手段各々を通じて入力される前記音源各々からの音源信号が重畳された複数の混合音声信号から、1以上の前記音源信号を分離した分離信号を生成する音源分離処理をコンピュータに実行させるための音源分離プログラムであって、

40

独立成分分析法に基づくブラインド音源分離方式の音源分離処理により複数の前記混合音声信号から1以上の前記音源信号に対応するSIMO信号を分離生成する第1の音源分離工程と、

前記第1の音源分離工程で分離生成された前記SIMO信号の全部若しくは一部である複数の特定信号について、複数の区分された周波数成分ごとに予め設定された重み係数により該周波数成分ごとの信号レベルを重み付けして補正し、その補正後の信号について前記周波数成分ごとに選択処理若しくは合成処理を行う所定の中間処理を行うことにより中間処理後信号を得る中間処理実行工程と、

前記中間処理実行工程により得られた複数の前記中間処理後信号に、又は該中間処理後信号及び前記第1の音源分離工程により分離生成された前記SIMO信号の一部の信号に

50

、バイナリーマスキング処理を施すことにより得られる信号を前記音源信号に対応する前記分離信号とする第2の音源分離工程と、

前記重み係数を所定の操作入力に従って設定する中間処理パラメータ設定工程と、
をコンピュータに実行させるための音源分離プログラム。

【請求項8】

所定の音響空間に複数の音源と複数の音声入力手段とが存在する状態で該音声入力手段各々を通じて入力される前記音源各々からの音源信号が重畳された複数の混合音声信号から、1以上の前記音源信号を分離した分離信号を生成する音源分離方法であって、

独立成分分析法に基づくブラインド音源分離方式の音源分離処理により複数の前記混合音声信号から1以上の前記音源信号に対応するS I M O信号を分離生成する第1の音源分離工程と、

前記第1の音源分離工程で分離生成された前記S I M O信号の全部若しくは一部である複数の特定信号について、複数の区分された周波数成分ごとに予め設定された重み係数により該周波数成分ごとの信号レベルを重み付けして補正し、その補正後の信号について前記周波数成分ごとに選択処理若しくは合成処理を行う所定の中間処理を行うことにより中間処理後信号を得る中間処理実行工程と、

前記中間処理実行工程により得られた複数の前記中間処理後信号に、又は該中間処理後信号及び前記第1の音源分離工程により分離生成された前記S I M O信号の一部の信号に、バイナリーマスキング処理を施すことにより得られる信号を前記音源信号に対応する前記分離信号とする第2の音源分離工程と、

前記重み係数を所定の操作入力に従って設定する中間処理パラメータ設定工程と、
を有してなることを特徴とする音源分離方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、所定の音響空間に複数の音源と複数の音声入力手段とが存在する状態で、その音声入力手段各々を通じて入力される前記音源各々からの個別音声信号が重畳された複数の混合音声信号から、1以上の前記個別音声信号を同定（分離）する音源分離装置、音源分離プログラム及び音源分離方法に関するものである。

【背景技術】

【0002】

所定の音響空間に複数の音源と複数のマイク（音声入力手段）とが存在する場合、その複数のマイクごとに、複数の音源各々からの個別音声信号（以下、音源信号という）が重畳された音声信号（以下、混合音声信号という）が取得される。このようにして取得（入力）された複数の前記混合音声信号のみに基づいて、前記音源信号各々を同定（分離）する音源分離処理の方式は、ブラインド音源分離方式（Blind Source Separation方式、以下、B S S方式という）と呼ばれる。

さらに、B S S方式の音源分離処理の1つに、独立成分分析法（Independent Component Analysis、以下、I C A法という）に基づくB S S方式の音源分離処理がある。このI C A法に基づくB S S方式は、複数のマイクを通じて入力される複数の前記混合音声信号（時系列の音声信号）において、前記音源信号どうしが統計的に独立であることを利用して所定の逆混合行列を最適化し、入力された複数の前記混合音声信号に対して最適化された逆混合行列によるフィルタ処理を施すことによって前記音源信号の同定（音源分離）を行う処理方式である。このようなI C A法に基づくB S S方式の音源分離処理は、例えば、非特許文献1や非特許文献2、及び非特許文献6や非特許文献7等に詳説されている。

一方、音源分離処理としては、パイノーラル信号処理（分解）による音源分離処理も知られている。これは、人間の聴覚モデルに基づいて複数の入力音声信号に時変のゲイン調節を施して音源分離を行うものであり、比較的低い演算負荷で実現できる音源分離処理である。これについては、例えば、非特許文献3や非特許文献4等に詳説されている。

【非特許文献1】猿渡洋「アレー信号処理を用いたブラインド音源分離の基礎」電子情報

10

20

30

40

50

通信学会技術報告、vol.EA2001-7、pp.49-56、April 2001.

【非特許文献2】高谷智哉他「SIMOモデルに基づくICAを用いた高忠実度なブラインド音源分離」電子情報通信学会技術報告、vol.US2002-87、EA2002-108、January 2003.

【非特許文献3】R.F.Lyon, "A computational model of binaural localization and separation," In Proc. ICASSP, 1983.

【非特許文献4】M. Bodden, "Modeling human sound-source localization and the cocktail-party-effect," Acta Acoustica, vol.1, pp.43--55, 1993.

【非特許文献5】N.Murata and S. Ikeda. A on-line algorithm for blind source separation on speech signals. In Proceedings of NOLTA'98, pp. 923-926,1998

【非特許文献6】梶田、小林、武田、板倉、「ヒューマンスピーチライク雑音に含まれる音声的特徴の分析」、日本音響学会誌、53巻5号、pp.337-345 (1997)

【非特許文献7】鷓飼訓史他、「周波数領域ICAと時間領域ICAを統合したSIMOモデル信号のブラインド抽出法の評価」、電子情報通信学会技術報告、vol.EA2004-23, p.37-42, June 2004

【発明の開示】

【発明が解決しようとする課題】

【0003】

しかしながら、前記音源信号（個別音声信号）の独立性に着目したICA法に基づくBSS方式による音源分離処理は、これを実環境で用いた場合、音声信号の伝達特性や背景ノイズ等の影響により、統計量を高精度で推定できず（即ち、前記逆混合行列が十分に最適化されず）、十分な音源分離性能（前記音源信号の同定性能）が得られない場合があるという問題点があった。

また、パイノール信号処理による音源分離処理は、処理が簡易で演算負荷が低い反面、音源の位置に対しての頑健性が悪い等、一般に音源分離性能は劣るという問題点があった。

一方、音源分離処理は、その適用対象によっては、分離後の音声信号に特定の音源以外の他の音源からの音声信号が極力含まれていないこと（音源分離性能が高いこと）が特に重視される場合や、また、分離後の音声信号の音質が良い（スペクトル歪みが小さい）ことが特に重視される場合がある。しかしながら、従来の音源分離装置は、そのように重視される目的に応じた音源分離を行うことができないという問題点もあった。

従って、本発明は上記事情に鑑みてなされたものであり、その目的とするところは、ノイズの影響がある等の多様な環境下においても高い音源分離性能が得られるとともに、重視される目的（音源分離性能又は音質）に応じた音源分離処理が可能な音源分離装置、音源分離プログラム及び音源分離方法を提供することにある。

【課題を解決するための手段】

【0004】

上記目的を達成するために本発明は、所定の音響空間に複数の音源と複数の音声入力手段（マイクロホン）とが存在する状態でその音声入力手段各々を通じて入力される前記音源各々からの音源信号が重畳された複数の混合音声信号から、1以上の前記音源信号を分離（抽出）した分離信号を生成するものであり、以下の各工程を実行する手段を備える音源分離装置であること、又は以下の各工程をコンピュータに実行させるプログラムであること、或いは以下の（1）～（4）の各工程を有する音源分離方法であることを特徴とするものである。

（1）独立成分分析法に基づくブラインド音源分離方式の音源分離処理により、複数の前記混合音声信号から1以上の前記音源信号に対応するSIMO（single-input multiple-output）信号を分離生成（抽出）する工程。以下、この工程を第1の音源分離工程といい、この工程で実行される処理を第1の音源分離処理という。

（2）第1の音源分離工程により分離生成された前記SIMO信号の全部若しくは一部である複数の信号（以下、特定信号という）について、複数の区分された周波数成分ごとに予め設定された重み係数により該周波数成分ごとの信号レベルを重み付けして補正し、そ

10

20

30

40

50

の補正後の信号について前記周波数成分ごとに選択処理若しくは合成処理を行う所定の処理（以下、中間処理という）を行うことによりその中間処理が施された信号（以下、中間処理後信号という）を得る工程。以下、この工程を中間処理実行工程という。

（３）中間処理実行工程により得られた複数の前記中間処理後信号に、又はその中間処理後信号及び前記第１の音源分離工程により分離生成された前記ＳＩＭＯ信号の一部の信号に、バイナリーマスキング処理を施すことにより得られる信号を前記音源信号に対応する前記分離信号とする工程。以下、この工程を第２の音源分離工程といい、この工程で実行される処理を第２の音源分離処理という。

（４）前記重み係数を所定の操作入力に従って設定する中間処理パラメータ設定工程。

本発明に係る音源分離装置（又は音源分離方法）は、２段階の音源分離処理（前記第１の音源分離処理及び前記第２の音源分離処理）を行う。その結果、後述するように、ノイズの影響がある等の多様な音響環境の下においても、高い音源分離性能が得られることがわかった。また、前記中間処理の内容により、音源分離性能が特に高まるような音源分離処理を実現したり、或いは分離後の音声信号の音質が特に高まるような音源分離処理を実現したりすることができる。

特に、前記重み係数を所定の操作入力に従って設定する前記中間処理パラメータ設定工程（手段）により、重視する目的に応じた音源分離処理が行われるように調節することがより容易となる。

【０００５】

より具体的な前記中間処理としては、前記補正後の信号の中から前記周波数成分ごとに信号レベルが最大のものを選択する処理を行うものが考えられる。

このような構成によれば、前記重み係数を調節することにより、音源分離性能が特に高まるような音源分離処理を実現したり、或いは分離後の音声信号の音質が特に高まるような音源分離処理を実現したりすることができる。

【０００６】

また、前記第１の音源分離処理としては、周波数領域ＳＩＭＯ独立成分分析法に基づくブラインド音源分離方式の音源分離処理や、周波数領域独立成分分析法と逆射影法との連結手法に基づくブラインド音源分離方式の音源分離処理を行うことが考えられる。

なお、前記周波数領域ＳＩＭＯ独立成分分析法に基づくブラインド音源分離方式の音源分離処理とは、後述するように、次の（１－１）～（１－４）に示す各処理を実行する処理である。

（１－１）時間領域における複数の前記混合音声信号に短時間離散フーリエ変換処理を施して、周波数領域における複数の混合音声信号へ変換する短時間離散フーリエ変換処理。

（１－２）前記周波数領域における複数の混合音声信号に対し、所定の分離行列に基づく分離処理を施すことにより前記混合音声信号ごとに前記音源信号のいずれかに対応した分離信号（第１の分離信号）を生成するＦＤＩＣＡ音源分離処理。

（１－３）前記周波数領域における複数の混合音声信号各々から、当該混合音声信号に基づいて前記ＦＤＩＣＡ音源分離処理により分離された分離信号（前記第１の分離信号）を除く残りのものを減算した分離信号（第２の分離信号）を生成する減算処理。

（１－４）前記第１の分離信号及び前記第２の分離信号に基づく所定の評価関数を用いた逐次計算を行うことによつて、前記ＦＤＩＣＡ音源分離処理で用いる前記分離行列を計算する分離行列計算処理。

前記周波数領域ＳＩＭＯ独立成分分析法に基づくブラインド音源分離方式の音源分離処理は、時間領域の混合音声信号を時間領域のまま処理する時間領域ＳＩＭＯ独立成分分析法に基づくブラインド音源分離方式の音源分離処理（前記非特許文献２等を参照）に比べ、処理負荷を大幅に軽減できる。

【０００７】

ところで、一般に、前記ＩＣＡ法に基づくＢＳＳ方式による音源分離処理は、十分な音源分離性能を得るためには、分離処理（フィルタ処理）に用いる分離行列を求めるための逐次計算（学習計算）の回数が増えるので演算負荷が高くなる。その逐次計算（学習計算

10

20

30

40

50

)は、製品組み込み用として実用的なプロセッサで実行された場合、入力される混合音声信号の時間長に対して数倍の時間を要し、リアルタイム処理には適さない。また、前記逐次計算(学習計算)の回数を制限することは、音響環境の変化(音源の移動や音源の追加・変更等)が大きい場合に、十分な音源分離性能が得られなくなるという結果を招く。

一方、前記バイナリーマスキング処理は、製品組み込み用として実用的なプロセッサによってリアルタイム処理が可能であるとともに、音響環境が変化しても比較的安定した音源分離性能を発揮するが、前記分離行列の学習が十分なされた前記ICA法に基づくBSS方式による音源分離処理に比べて音源分離性能はるかに劣る。

しかしながら、以上示した本発明に係る音源分離処理によれば、以下のような構成により、音源分離性能を確保しつつリアルタイム処理が可能となる。

10

例えば、第1の音源分離処理における分離行列の逐次計算回数を制限することが考えられる。

即ち、前記第1の音源分離処理(前記第1の音源分離手段の処理)において、時系列で入力される前記混合音声信号が予め定められた周期で区分された区間信号それぞれに対し、所定の分離行列に基づく分離処理を順次実行することによって前記SIMO信号を生成するとともに、その分離処理により生成した前記区間信号の時間帯に対応する全ての時間帯の前記SIMO信号に基づいて、その後(以降に)用いる前記分離行列を求める逐次計算(学習計算)を行うものであって、その逐次計算の回数を前記予め定められた周期の時間内で実行可能な回数に制限することが考えられる。

このように、前記第1の音源分離処理(第1段階目の前記ICA法に基づくBSS方式による音源分離処理)において、前記分離行列を求める逐次計算(学習計算)の回数を、リアルタイム処理が可能な範囲に制限すると学習が不十分となるため、得られる前記SIMO信号は、十分な音源分離(同定)がなされた信号にならないことが多い。

20

しかしながら、それによって得られた前記SIMO信号に基づく前記中間処理によって得られる信号に、リアルタイム処理が可能な第2段階目の前記バイナリーマスキング処理をさらに施すことによって音源分離性能が向上するので、音源分離性能を確保しつつリアルタイム処理が可能となる。

【0008】

また、前記第1の音源分離処理における前記分離行列の逐次計算に用いるSIMO信号のサンプル数を減らすことも考えられる。

30

即ち、前記第1の音源分離処理(前記第1の音源分離手段の処理)において、時系列に入力される前記混合音声信号が予め定められた周期で区分された区間信号ごとに、その区間信号に対し所定の分離行列に基づく分離処理を順次実行して前記SIMO信号を生成するとともに、その分離処理により生成した前記区間信号の時間帯のうちの先頭側の一部の時間帯に対応する前記SIMO信号に基づいて、以降に用いる前記分離行列を求める逐次計算を前記予め定められた周期の時間内で実行することが考えられる。

このように、前記第1の音源分離処理(前記ICA法に基づくBSS方式による音源分離処理)において、前記分離行列を求める逐次計算(学習計算)に用いる前記SIMO信号を先頭側の一部の時間帯の信号に限定することにより、十分な回数の前記逐次計算(学習)を行ってもリアルタイム処理が可能にはなる(前記予め定められた周期の時間内で十分な学習が可能となる)が、学習に用いるサンプル数が少ないため、やはり得られる前記SIMO信号は、音源が十分に分離(同定)された信号にならないことが多い。しかしながら、本発明に係る音源分離処理装置(又は音源分離方法)は、それによって得られた前記SIMO信号にリアルタイム処理が可能な第2段階目の前記バイナリーマスキング処理をさらに施す。これにより、音源分離性能が向上し、高い音源分離性能を確保しつつリアルタイム処理が可能となる。

40

【発明の効果】

【0009】

本発明によれば、前記独立成分分析法に基づくブラインド音源分離方式の音源分離処理に、比較的簡易な前記バイナリーマスキング処理による音源分離処理を加えた2段階処理

50

を行うことにより、ノイズの影響がある等の多様な環境下においても高い音源分離性能が得られる。

さらに、本発明では、前記独立成分分析法に基づくブラインド音源分離方式の音源分離処理によって得られる前記S I M O信号に基づく前記中間処理を実行し、その中間処理後の信号について前記バイナリーマスキング処理を施す。これにより、前記中間処理の内容に応じて、音源分離性能が特に高まるような音源分離処理を実現したり、或いは分離後の音声信号の音質が特に高まるような音源分離処理を実現したりすることができる。その結果、重視する目的（音源分離性能又は音質）に応じて柔軟に対応できる音源分離処理が可能となる。

また、前記第1の音源分離処理として、前記周波数領域S I M O独立成分分析法に基づくブラインド音源分離方式の音源分離処理や、前記周波数領域独立成分分析法と前記逆射影法との連結手法に基づく前記ブラインド音源分離方式の音源分離処理を行うことにより、前記時間領域S I M O独立成分分析法に基づくブラインド音源分離方式の音源分離処理に比べ、処理負荷を大幅に軽減できる。

また、前記第1の音源分離処理における前記分離行列の逐次計算回数を制限したり、その逐次計算に用いる前記S I M O信号のサンプル数を減らすことにより、音源分離性能を確保しつつリアルタイム処理が可能となる。

【発明を実施するための最良の形態】

【0010】

以下添付図面を参照しながら、本発明の実施の形態について説明し、本発明の理解に供する。尚、以下の実施の形態は、本発明を具体化した一例であって、本発明の技術的範囲を限定する性格のものではない。

ここに、図1は本発明の実施形態に係る音源分離装置Xの概略構成を表すブロック図、図2は本発明の第1実施例に係る音源分離装置X1の概略構成を表すブロック図、図3はT D I C A法に基づくB S S方式の音源分離処理を行う従来の音源分離装置Z1の概略構成を表すブロック図、図4はT D - S I M O - I C A法に基づく音源分離処理を行う従来の音源分離装置Z2の概略構成を表すブロック図、図5はF D I C A法に基づく音源分離処理を行う従来の音源分離装置Z3の概略構成を表すブロック図、図6はF D - S I M O - I C A法に基づく音源分離処理を行う音源分離装置Z4の概略構成を表すブロック図、図7はF D I C A - P B法に基づく音源分離処理を行う従来の音源分離装置Z5の概略構成を表すブロック図、図8はバイナリーマスキング処理を説明するための図、図9はS I M O信号にバイナリーマスキング処理を施す前後の信号における周波数成分ごとの信号レベル分布の第1例（音源信号各々の周波数成分に重複がない場合）を模式的に表した図、図10はS I M O信号にバイナリーマスキング処理を施す前後の信号における周波数成分ごとの信号レベル分布の第2例（音源信号各々の周波数成分に重複がある場合）を模式的に表した図、図11はS I M O信号にバイナリーマスキング処理を施す前後の信号における周波数成分ごとの信号レベル分布の第3例（目的音源信号のレベルが比較的小さい場合）を模式的に表した図、図12は音源分離装置X1におけるS I M O信号に対する音源分離処理の第1例の内容を模式的に表した図、図13は音源分離装置X1におけるS I M O信号に対する音源分離処理の第2例の内容を模式的に表した図、図14は音源分離装置X1を用いた音源分離性能評価の実験条件を表す図、図15は従来の音源分離装置と本発明に係る音源分離装置との各々により所定の実験条件の下で音源分離を行ったときの音源分離性能及び音質の評価値を表すグラフ、図16は音源分離装置Xにおける分離行列計算の第1例を説明するためのタイムチャート、図17は音源分離装置Xにおける分離行列計算の第2例を説明するためのタイムチャート、図18は音源分離装置X1におけるS I M O信号に対する音源分離処理の第3例の内容を模式的に表した図である。

【0011】

まず、本発明の実施形態について説明する前に、図3～図7に示すブロック図を参照しつつ、各種のI C A法に基づくブラインド音源分離方式（I C A法に基づくB S S方式）の音源分離装置について説明する。

なお、以下に示す音源分離処理或いはその処理を行う装置等は、いずれも所定の音響空間に複数の音源と複数のマイクロホン（音声入力手段）とが存在する状態で、そのマイクロホン各々を通じて入力される前記音源各々からの個別の音声信号（以下、音源信号という）が重畳された複数の混合音声信号から、1以上の音源信号を分離（同定）した分離信号を生成する音源分離処理或いはその処理を行う装置等に関するものである。

【0012】

図3は、ICA法的一种である時間領域独立成分分析法（time-domain independent component analysis法、以下、TDICA法という）に基づくBSS方式の音源分離処理を行う従来の音源分離装置Z1の概略構成を表すブロック図である。なお、本処理の詳細は、非特許文献1や非特許文献2等に示されている。

10

前記音源分離装置Zは、分離フィルタ処理部11により、2つの音源1、2からの音源信号 $S_1(t)$ 、 $S_2(t)$ （音源ごとの音声信号）を2つのマイクロホン（音声入力手段）111、112で入力した2チャンネル（マイクロホンの数）の混合音声信号 $x_1(t)$ 、 $x_2(t)$ について、分離行列 $W(z)$ によりフィルタ処理を施すことによって音源分離を行う。

図3には、2つの音源1、2からの音源信号 $S_1(t)$ 、 $S_2(t)$ （個別音声信号）を2つの前記マイクロホン111、112で入力した2チャンネル（マイクロホンの数）の前記混合音声信号 $x_1(t)$ 、 $x_2(t)$ に基づいて音源分離を行う例について示しているが、2チャンネル以上であっても同様である。前記ICA法に基づくBSS方式による音源分離の場合、（入力される混合音声信号のチャンネル数 n （即ち、マイクロホンの数））（音源の数 m ）であればよい。

20

複数の前記マイクロホン111、112各々で集音された前記混合音声信号 $x_1(t)$ 、 $x_2(t)$ それぞれには、複数音源からの前記音源信号が重畳されている。以下、前記混合音声信号 $x_1(t)$ 、 $x_2(t)$ を総称して $x(t)$ と表す。この混合音声信号 $x(t)$ は前記音源信号 $S(t)$ の時間的空間的な畳み込み信号として表現され、次の(1)式のように表される。

【数1】

$$\mathbf{x}(t) = \mathbf{A}(z) \cdot \mathbf{s}(t) \quad \dots(1)$$

前記TDICA法による音源分離の理論は、前記音源信号 $S(t)$ のそれぞれの音源同士が統計的に独立であることを利用すると、 $x(t)$ がわかれば $S(t)$ を推測することができ、従って、音源を分離することができるという発想に基づく理論である。

30

ここで、当該音源分離処理に用いる前記分離行列を $W(z)$ とすれば、前記分離信号（即ち、同定信号） $y(t)$ は、次の(2)式で表される。

【数2】

$$\mathbf{y}(t) = \mathbf{W}(z) \cdot \mathbf{x}(t) \quad \dots(2)$$

ここで、 $W(z)$ は、出力 $y(t)$ から逐次計算により求められる。また、前記分離信号は、チャンネルの数だけ得られる。

なお、音源合成処理はこの $W(z)$ に関する情報により、逆演算処理に相当する行列を形成し、これを用いて逆演算を行えばよい。

40

このようなICA法に基づくBSS方式による音源分離を行うことにより、例えば、人の歌声とギター等の楽器の音とが混合した複数チャンネル分の混合音声信号から、歌声の音源信号と楽器の音源信号とが分離（同定）される。

ここで、(2)式は、次の(3)式のように書き換えて表現できる。

【数 3】

$$\mathbf{y}(t) = \sum_{n=0}^{D-1} \mathbf{w}(n)\mathbf{x}(t-n) \quad \dots(3)$$

但し、 D は分離フィルタ $\mathbf{w}(n)$ のタップ数。

そして、(3)式における分離フィルタ(分離行列) $W(n)$ は、次の(4)式により逐次計算される。即ち、前回(j)の出力 $y(t)$ を(4)式に逐次適用することより、今回($j+1$)の $W(n)$ を求める。

【数 4】

$$\mathbf{w}^{[j+1]}(n) = \mathbf{w}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \left\{ \text{off} - \text{diag} \left\langle \varphi(\mathbf{y}^{[j]}(t)) \mathbf{y}^{[j]}(t-n+d)^T \right\rangle_t \right\} \cdot \mathbf{w}^{[j]}(d) \quad \dots(4)$$

但し、 α は更新係数、 $[j]$ は更新回数、 $\langle \dots \rangle_t$ は時間平均を表す。

off-diag X は行列 X の全ての対角要素をゼロに置き換える演算処理を表す。

$\varphi(\dots)$ はシグモイド関数等を要素にもつ適当な非線形ベクトル関数を表す。

10

【0013】

次に、図4に示すブロック図を用いて、TDICA法の一種である時間領域SIMO独立成分分析法(Time-Domain single-input multiple-output ICA法、以下、TD-SIMO-ICA法という)に基づく音源分離処理を行う従来の音源分離装置2の構成について説明する。なお、図4は、2チャンネル(マイクロホンの数)の混合音声信号 $x_1(t)$ 、 $x_2(t)$ に基づいて音源分離を行う例について示しているが、3チャンネル以上であっても同様であり、その詳細は非特許文献2等に示されている。

20

TD-SIMO-ICA法による音源分離の特徴は、図4中に示すFidelity Controller 12により、マイクロホン入力信号である各混合音声信号 $x_i(t)$ から、前記分離フィルタ処理部11による音源分離処理(TDICA法に基づく音源分離処理)によって分離(同定)された分離信号(同定信号)を減算し、その減算により得られる信号成分の統計的独立性も評価することによって分離フィルタ $W(Z)$ の更新(逐次計算)を行う点である。ここで、混合音声信号 $x_i(t)$ 各々から減算する分離信号(同定信号)は、各々異なる1つの分離信号(当該混合音声信号に基づく音源分離処理により得られた分離信号)を除く残り全ての分離信号である。これにより、チャンネル(マイクロホン)ごとに2つの分離信号(同定信号)が得られることになり、また、音源信号 $S_i(t)$ ごとに2つの分離信号が得られることになる。図4の例では、分離信号 $y_{11}(t)$ と $y_{12}(t)$ 、分離信号 $y_{22}(t)$ と $y_{21}(t)$ が、各々同一の音源信号に対応する分離信号(同定信号)である。なお、分離信号 y の添字(数字)において、前段の数字は音源の識別番号を、後段の数字はマイクロホン(即ち、チャンネル)の識別番号を表している(以下同様)。

30

このように、ある音響空間に複数の音源と複数の音声入力手段(マイクロホン)とが存在する状態で、その音声入力手段各々を通じて入力される音源各々からの音源信号(個別音声信号)が重畳された複数の混合音声信号から、1以上の音源信号を分離(同定)した場合に、音源信号ごとに得られる複数の分離信号(同定信号)群をSIMO(single-input multiple-output)信号という。図4の例では、分離信号 $y_{11}(t)$ と $y_{12}(t)$ の組合せ、分離信号 $y_{22}(t)$ と $y_{21}(t)$ の組合せの各々がSIMO信号である。

40

ここで、分離フィルタ(分離行列) $W(Z)$ を表現し直した $W(n)$ の更新式は、次の(5)式で表される。

【数5】

$$\begin{aligned} \mathbf{w}_{ICA1}^{[j+1]}(n) = & \mathbf{w}_{ICA1}^{[j]}(n) - \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \varphi \left(\mathbf{y}_{ICA1}^{[j]}(t) \right) \mathbf{y}_{ICA1}^{[j]}(t-n+d)^T \right\rangle_t \right\} \cdot \mathbf{w}_{ICA1}^{[j]}(d) \\ & + \alpha \sum_{d=0}^{D-1} \left\{ \text{off-diag} \left\langle \varphi \left(x \left(t - \frac{D}{2} \right) - \sum_{l=1}^{L-1} \mathbf{y}_{ICA1}^{[j]}(t) \right) \cdot \left(x \left(t - \frac{D}{2} - n + d \right) - \sum_{l=1}^{L-1} \mathbf{y}_{ICA1}^{[j]}(t-n+d) \right)^T \right\rangle_t \right\} \\ & \cdot \left(\mathbf{I} \delta \left(d - \frac{D}{2} \right) - \sum_{l=1}^{L-1} \mathbf{w}_{ICA1}^{[j]}(d) \right) \end{aligned} \quad \dots(5)$$

但し、 α は更新係数、 $[j]$ は更新回数、 $\langle \dots \rangle_t$ は時間平均を表す。

off-diag \mathbf{X} は行列 \mathbf{X} の全ての対角要素をゼロに置き換える演算処理を表す。

$\varphi(\dots)$ はシグモイド関数等を要素にもつ適当な非線形ベクトル関数を表す。

Wおよびyの添字「ICA1」はSIMO-ICA部内の1(エル)番目のICA成分を表す。

10

この(5)式は、前述の(4)式に対して第3項目が加わったものであり、この第3項は、Fidelity Controller 1 2により生成される信号の成分の独立性を評価している部分である。

【0014】

次に、図5に示すブロック図を用いて、ICA法の一種であるFDICA法(Frequency-Domain ICA)に基づく音源分離処理を行う従来の音源分離装置Z3について説明する。

FDICA法では、まず、入力された混合音声信号 $x(t)$ について、ST-DFT処理部13によって所定の周期ごとに区分された信号であるフレーム毎に短時間離散フーリエ変換(Short Time Discrete Fourier Transform、以下、ST-DFT処理という)を行い、観測信号の短時間分析を行う。そして、そのST-DFT処理後の各チャンネルの信号(各周波数成分の信号)について、分離フィルタ処理部11fにより分離行列 $W(f)$ に基づく分離フィルタ処理を施すことによって音源分離(音源信号の同定)を行う。ここで f を周波数ビン、 m を分析フレーム番号とすると、分離信号(同定信号) $y(f, m)$ は、次の(6)式のように表すことができる。

20

【数6】

$$\mathbf{Y}(f, m) = \mathbf{W}(f) \cdot \mathbf{X}(f, m) \quad \dots(6)$$

ここで、分離フィルタ $W(f)$ の更新式は、例えば次の(7)式のように表すことができる。

30

【数7】

$$\mathbf{W}_{(ICA)}^{[i+1]}(f) = \mathbf{W}_{(ICA)}^{[i]}(f) - \eta(f) \left[\text{off-diag} \left\{ \varphi \left(\mathbf{Y}_{(ICA)}^{[i]}(f, m) \right) \mathbf{Y}_{(ICA)}^{[i]}(f, m)^H \right\}_m \right] \mathbf{W}_{(ICA)}^{[i]}(f) \quad \dots(7)$$

但し、 $\eta(f)$ は更新係数、 i は更新回数、 $\langle \dots \rangle$ は時間平均、 H はエルミート転置を表す。

off-diag \mathbf{X} は行列 \mathbf{X} の全ての対角要素をゼロに置き換える演算処理を表す。

$\varphi(\dots)$ はシグモイド関数等を要素にもつ適当な非線形ベクトル関数を表す。

40

このFDICA法によれば、音源分離処理が各狭帯域における瞬時混合問題として取り扱われ、比較的簡単かつ安定に分離フィルタ(分離行列) $W(f)$ を更新することができる。

【0015】

次に、図6に示すブロック図を用いて、FDICA法の一種である周波数領域SIMO独立成分分析法(Frequency-Domain single-input multiple-output ICA法、以下、FD-SIMO-ICA法という)に基づく音源分離処理を行う音源分離装置Z4について説明する。

FD-SIMO-ICA法では、前述のTD-SIMO-ICA法(図4)と同様に、

50

Fidelity Controller 1 2 により、各混合音声信号 $x_i(t)$ に対して S T - D F T 処理を施した信号各々から、F D I C A 法 (図 5) に基づく音源分離処理によって分離 (同定) された分離信号 (同定信号) を減算し、その減算により得られる信号成分の統計的独立性も評価することによって分離フィルタ $W(f)$ の更新 (逐次計算) を行うものである。

この F D - S I M O - I C A 法に基づく音源分離装置 Z 4 では、時間領域における複数の前記混合音声信号 $x_1(t)$ 、 $x_2(t)$ に対して前記 S T - D F T 処理部 1 3 により短時間離散フーリエ変換処理を施して、周波数領域における複数の混合音声信号 $x_1(f)$ 、 $x_2(f)$ へ変換する (短時間離散フーリエ変換手段の一例) 。

次に、変換後の周波数領域における複数の混合音声信号 $x_1(f)$ 、 $x_2(f)$ に対し、前記分離フィルタ処理部 1 1 f によって所定の分離行列 $W(f)$ に基づく分離処理 (フィルタ処理) を施すことにより、前記混合音声信号ごとに前記音源信号 $S_1(t)$ 、 $S_2(t)$ のいずれかに対応した第 1 の分離信号 $y_{11}(f)$ 、 $y_{22}(f)$ を生成する (F D I C A 音源分離手段の一例) 。

さらに、前記周波数領域における複数の混合音声信号 $x_1(f)$ 、 $x_2(f)$ 各々から当該混合音声信号に基づき前記分離フィルタ処理部 1 1 f により分離された前記第 1 の分離信号 ($x_1(f)$ に基づき分離された $y_{11}(f)$ 、 $x_2(f)$ に基づき分離された $y_{22}(f)$) を除く残りの前記第 1 の分離信号を、前記 Fidelity Controller 1 2 (減算手段の一例) により減算した第 2 の分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ を生成する。

一方、不図示の分離行列計算部により、前記第 1 の分離信号 $y_{11}(f)$ 、 $x_2(f)$ 及び前記第 2 の分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ の両方に基づく逐次計算を行い、前記分離フィルタ処理部 1 1 f (F D I C A 音源分離手段) で用いられる前記分離行列 $W(f)$ を計算する (分離行列計算手段の一例) 。

これにより、チャンネル (マイクロホン) ごとに 2 つの分離信号 (同定信号) が得られることになり、また、音源信号 $S_i(t)$ ごとに 2 つ以上の分離信号 (S I M O 信号) が得られることになる。図 6 の例では、分離信号 $y_{11}(f)$ と $y_{12}(f)$ の組合せ、及び分離信号 $y_{22}(f)$ と $y_{21}(f)$ の組合せの各々が S I M O 信号である。

ここで、前記分離行列計算部は、前記第 1 の分離信号及び前記第 2 の分離信号に基づいて、次の (8) 式で表される分離フィルタ (分離行列) $W(f)$ の更新式により前記分離行列 $W(f)$ を計算する。

【数 8】

$$\begin{aligned} \mathbf{W}_{(ICA)}^{[i+1]}(f) = & \mathbf{W}_{(ICA)}^{[i]}(f) - \eta(f) \left[\text{off-diag} \left\{ \phi \left(\mathbf{Y}_{(ICA)}^{[i]}(f, m) \mathbf{Y}_{(ICA)}^{[i]}(f, m)^H \right) \right\}_m \right] \mathbf{W}_{(ICA)}^{[i]}(f) \\ & - \text{off-diag} \left\{ \phi \left(\mathbf{X}(f, m) - \sum_{l=1}^{L-1} \mathbf{Y}_{(ICA)}^{[l]}(f, m) \left(\mathbf{X}(f, m) - \sum_{l=1}^{L-1} \mathbf{Y}_{(ICA)}^{[l]}(f, m) \right)^H \right) \right\}_m \\ & \left(\mathbf{I} - \sum_{l=1}^{L-1} \mathbf{W}_{(ICA)}^{[l]}(f) \right) \end{aligned} \quad \dots(8)$$

但し、 $\eta(f)$ は更新係数、 i は更新回数、 $\langle \dots \rangle$ は時間平均、 H はエルミート転置を表す。

off-diag \mathbf{X} は行列 \mathbf{X} の全ての対角要素をゼロに置き換える演算処理を表す。

$\phi(\dots)$ はシグモイド関数等を要素にもつ適当な非線形ベクトル関数を表す。

【0016】

次に、図 7 に示すブロック図を用いて、F D I C A 法の一つである周波数領域独立成分分析法と逆射影法との連結手法 (Frequency-Domain ICA & Projection back法、以下、F D I C A - P B 法という) に基づく音源分離処理を行う従来の音源分離装置 Z 5 について説明する。なお、P D I C A - P B 法の詳細については、特許文献 5 等に示されている。

F D I C A - P B 法では、各混合音声信号 $x_i(t)$ から前述の F D I C A 法に基づく音源分離処理 (図 5) により得られた分離信号 (同定信号) $y_i(f)$ 各々について、逆行列

演算部 14 によって分離行列 $W(f)$ の逆行列 $W^{-1}(f)$ の演算処理を施すことにより、最終的な分離信号（音源信号の同定信号）を得るものである。ここで、逆行列 $W^{-1}(f)$ による処理対象の信号のうち、各分離信号 $y_i(f)$ 以外の残りの信号成分は、0（ゼロ）入力として設定する。

これにより、音源信号 $S_i(t)$ 各々に対応したチャンネル数分（複数）の分離信号（同定信号）である S I M O 信号が得られる。図 7 において、分離信号 $y_{11}(f)$ と $y_{12}(f)$ 、分離信号 $y_{21}(f)$ と $y_{22}(f)$ が、各々同一の音源信号に対応する分離信号（同定信号）であり、各逆行列 $W^{-1}(f)$ による処理後の信号である分離信号 $y_{11}(f)$ と $y_{12}(f)$ の組合せ、分離信号 $y_{21}(f)$ と $y_{22}(f)$ の組合せの各々が S I M O 信号である。

【 0 0 1 7 】

以下、図 1 に示すブロック図を用いて、本発明の実施形態に係る音源分離装置 X について説明する。

音源分離装置 X は、ある音響空間に複数の音源 1、2 と複数のマイクロホン 1 1 1、1 1 2（音声入力手段）とが存在する状態で、そのマイクロホン 1 1 1、1 1 2 各々を通じて入力される音源 1、2 各々からの音源信号（個別の音声信号）が重畳された複数の混合音声信号 $X_i(t)$ から、1 以上の音源信号（個別音声信号）を分離（同定）した分離信号（同定信号） y を生成するものである。

そして、音源分離装置 X の特徴は、以下の（1）～（3）の構成要素を備える点にある。

（1）複数の混合音声信号 $X_i(t)$ から独立成分分析（ICA）法に基づくブラインド音源分離（BSS）方式の音源分離処理により、1 以上の音源信号 $S_i(t)$ を分離（同定）した S I M O 信号（1 つの音源信号に対応する複数の分離信号）を分離生成する S I M O - I C A 処理部 10（第 1 の音源分離手段の一例）。

（2）S I M O - I C A 処理部 10 により生成された S I M O 信号のうちの一部である複数の信号について、複数に区分された周波数成分ごとに選択処理若しくは合成処理を行うことを含む所定の間接処理を行い、この中間処理により得られる中間処理後信号 $y_{d1}(f)$ 、 $y_{d2}(f)$ を出力する 2 つの中間処理実行部 4 1、4 2（中間処理実行手段の一例）。ここで、周波数成分の区分は、例えば、予め定められた周波数幅での均等な区分とすることが考えられる。

なお、図 1 に例示した中間処理実行部 4 1、4 2 各々は、4 つの分離信号からなる S I M O 信号のうち、3 つの分離信号（特定信号の一例）に基づいて前記中間処理を行い、それぞれ 1 つの中間処理後信号 $y_{d1}(f)$ 、 $y_{d2}(f)$ を出力するものである。

（3）中間処理実行部 4 1、4 2 により得られた（出力された）前記中間処理後信号 $y_{d1}(f)$ 、 $y_{d2}(f)$ と、S I M O - I C A 処理部 10 により分離生成された S I M O 信号の一部の信号との各々を入力信号とし、その入力信号にバイナリマスキング処理を施して得られる信号を、1 以上の音源信号について分離（同定）した分離信号として生成する 2 つのバイノーラル信号処理部 2 1、2 2（第 2 の音源分離手段の一例）。

なお、S I M O - I C A 処理部 10 が音源分離処理を行う工程が、第 1 の音源分離工程の一例であり、中間処理実行部 4 1、4 2 が前記中間処理を行う工程が、中間処理実行工程の一例であり、さらに、バイノーラル信号処理部 2 1、2 2 がバイナリマスキング処理を行う工程が、第 2 の音源分離工程の一例である。

【 0 0 1 8 】

図 2 に示す例では、一方のバイノーラル信号処理部 2 1 に入力される S I M O 信号は、これに対応する中間処理実行部 4 1 が中間処理の対象としていない S I M O 信号である。同様に、他方のバイノーラル信号処理部 2 2 に入力される S I M O 信号も、これに対応する中間処理実行部 4 2 が中間処理の対象としていない S I M O 信号である。但し、図 2 に示す例はあくまで一例であり、前記中間処理実行部 4 1、4 2 が、前記バイノーラル信号処理部 2 1、2 2 に入力される前記 S I M O 信号（図 2 における $y_{11}(f)$ や $y_{22}(f)$ など）を前記中間処理の対象として入力する構成も考えられる。

ここで、S I M O - I C A 処理部 10（第 1 の音源分離手段）としては、図 4 に示した

10

20

30

40

50

T D - S I M O - I C A 法に基づく音源分離処理を行う前記音源分離装置 Z 2 や、図 6 に示した F D - S I M O - I C A 法に基づく音源分離処理を行う F D - S I M O - I C A 法に基づく音源分離処理を行う前記音源分離装置 Z 4、或いは図 7 に示した F D I C A - P B 法に基づく音源分離処理を行う前記音源分離装置 Z 5 等を採用することが考えられる。

但し、S I M O - I C A 処理部 1 0 として、前記 T D - S I M O - I C A 法に基づく前記音源分離装置 Z 2 を採用する場合や、F D - S I M O - I C A 法若しくは F D I C A - P B 法に基づく音源分離処理後の信号が I D F T 処理（逆離散フーリエ変換処理）により時間領域の信号に変換されている場合には、その S I M O - I C A 処理部 1 0（音源分離装置 Z 2 等）により得られる分離信号（同定信号）について、バイナリーマスキング処理を施す前に、離散フーリエ変換処理（D F T 処理）を施す手段を設ける。これにより、前記

10

パイノール信号処理部 2 1、2 2 や中間処理実行部 4 1、4 2 への入力信号を、時間領域の離散信号から周波数領域の離散信号へ変換する。

さらに、図 1 には示されていないが、音源分離装置 X は、前記パイノール信号処理部 2 1 の出力信号（周波数領域の分離信号）を時間領域の信号に変換する（逆離散フーリエ変換処理を施す）I D F T 処理部も備えている。

【 0 0 1 9 】

また、図 1 には、チャンネル数（マイクロホンの数）の分だけ生成される S I M O 信号各々について、バイナリーマスキング処理による音源分離処理を施す構成例を示しているが、一部の音源信号の分離（同定）を行うことを目的とする場合は、一部のチャンネルに対応する S I M O 信号（或いは、一部のマイクロホン若しくは一部の復号音声信号 $x_i(t)$ ）に対応する S I M O 信号ともいえる）についてのみ、バイナリーマスキング処理を施す構成も考えられる。

20

さらに、図 1 には、チャンネル数が 2 つ（マイクロホンの数が 2 つ）である例について示したが、（入力される混合音声信号のチャンネル数 n （即ち、マイクロホンの数））（音源の数 m ）であれば、3 チャンネル以上であっても同様の構成により実現できる。

ここで、各構成要素 1 0、2 1、2 2、4 1、4 2 は、それぞれ D S P（Digital Signal Processor）又は C P U 及びその周辺装置（R O M、R A M 等）と、その D S P 若しくは C P U により実行されるプログラムとにより構成されたものや、或いは、1 つの C P U 及びその周辺装置を有するコンピュータにより、各構成要素 1 0、2 1、2 2、4 1、4 2 が行う処理に対応するプログラムモジュールを実行するよう構成されたもの等が考えられる。また、所定のコンピュータに各構成要素 1 0、2 1、2 2、4 1、4 2 の処理を実行させる音源分離プログラムとして提供することも考えられる。

30

【 0 0 2 0 】

一方、前記パイノール信号処理部 2 1、2 2 における信号分離処理は、前述したように、人間の聴覚モデルに基づいて前記混合音声信号に時変のゲイン調節を施して音源分離を行うものであり、例えば、非特許文献 3 や非特許文献 4 等に詳説されている。

図 8 は、パイノール信号処理の考え方を起源とする信号処理の一例であって、比較的処理がシンプルなバイナリーマスキング処理を説明するための図である。

バイナリーマスキング処理を実行する装置やプログラムは、複数の入力信号（本発明においては S I M O 信号を構成する複数の音声信号）の比較処理を行う比較部 3 1 と、その比較部 3 1 による比較処理の結果に基づいて入力信号にゲイン調節を施して信号分離（音源分離）を行う分離部 3 2 とを有している。

40

バイナリーマスキング処理では、まず、前記比較部 3 1 において、入力信号（本発明においては S I M O 信号）各々について周波数成分ごとの信号レベル（振幅）分布 A L、A R を検出し、同じ周波数成分における信号レベルの大小関係を判別する。

図 8 において、B L、B R は、入力信号各々における周波数成分ごとの信号レベル分布と、その信号レベルごとに他方の対応する信号レベルに対する大小関係（ $>$ 、 $<$ ）を表した図である。図中、「 $>$ 」印は、前記比較部 3 1 による判別の結果、他方の対応する信号レベルよりも当該信号の信号レベルの方が大きかったことを表し、「 $<$ 」印は同じく当該信号レベルの方が小さかったことを表している。

50

次に、前記分離部 3 2 により、前記比較部 3 1 による信号比較の結果（大小判別の結果）に基づいて、入力信号各々にゲイン乗算（ゲイン調節）を施すことにより分離信号（同定信号）を生成する。この分離部 3 2 における最も簡単な処理の例としては、入力信号について、周波数成分ごとに、信号レベルが最も大きいと判別された入力信号の周波数成分にゲイン 1 を乗算し、その他の入力信号全てと同じ周波数成分にゲイン 0（ゼロ）を乗算すること等が考えられる。

これにより、入力信号と同数の分離信号（同定信号） C_L 、 C_R が得られる。この分離信号 C_L 、 C_R のうち、一方は、入力信号（前記 $SIMO-ICA$ 処理部 1 0 による分離信号（同定信号））の同定の対象となった音源信号に相当するものとなり、他方は入力信号に混在するノイズ（同定対象の音源信号以外の音源信号）に相当するものとなる。従って、前記 $SIMO-ICA$ 処理部 1 0 と前記バイノーラル信号処理部 2 1、2 2 とによる 2 段処理（直列的処理）によって、ノイズの影響がある等の多様な環境下においても高い音源分離性能が得られる。

なお、図 8 には、2 つの入力信号に基づくバイナリーマスキング処理の例を示すが、3 つ以上の入力信号に基づく処理であっても同様である。

例えば、まず、複数チャンネル分の入力信号各々について、複数に区分された周波数成分ごとに信号レベルを比較し、最大のものにゲイン 1 を乗算するとともに、その他のものにゲイン 0 を乗算し、その乗算により得られた信号を全てのチャンネルについて加算する。そして、この加算により得られる周波数成分ごとの信号を、全ての周波数成分について算出し、それらを組合せた信号を出力信号とすればよい。これにより、3 チャンネル分以上の入力信号に対しても、図 8 に示したのと同様にバイナリーマスキング処理を行うことができる。

【0021】

（第 1 実施例）

前記音源分離装置 X における前記 $SIMO-ICA$ 処理部 1 0 として、図 6 に示した $FD-SIMO-ICA$ 法に基づく音源分離処理を行う $FD-SIMO-ICA$ 法に基づく音源分離処理を行う前記音源分離装置 Z 4、或いは図 7 に示した $FDICA-PB$ 法に基づく音源分離処理を行う前記音源分離装置 Z 5 を採用したものを、以下、第 1 実施例とする。なお、図 2 は、そのような本発明の第 1 実施例に係る音源分離装置 X 1 の概略構成を表すブロック図であり、前記音源分離装置 X における前記 $SIMO-ICA$ 処理部 1 0 として、図 6 に示した $FD-SIMO-ICA$ 法に基づく音源分離処理を行う前記音源分離装置 Z 4 を採用した場合の例を示している。

この音源分離装置 X 1 の構成により、畳み込み演算が必要なために演算負荷が高い $TD-SIMO-ICA$ 法に基づく音源分離処理（図 4）を採用した構成よりも、比較的演算負荷が抑えられる。

また、本第 1 実施例に係る音源分離装置 X 1 では、前記 $SIMO-ICA$ 処理部 1 0 で用いる分離行列 $W(f)$ の初期値は、予め定められた値が設定される。

また、音源分離装置 X 1 のバイノーラル信号処理部 2 1、2 2 は、バイナリーマスキング処理を行う。

【0022】

図 2 に示す音源分離装置 X 1 では、前記 $SIMO-ICA$ 処理部 1 0 により、2 つの入力チャンネル（マイクロホン）ごとに 2 つの分離信号、即ち、合計 4 つの分離信号が得られ、この 4 つの分離信号が $SIMO$ 信号である。

また、一方の中間処理実行部 4 1 は、 $SIMO$ 信号の一部である分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ （特定信号の一例）を入力し、これらの信号に基づいて前記中間処理を実行する。同様に、他方の中間処理実行部 4 2 は、 $SIMO$ 信号の一部である分離信号 $y_{11}(f)$ 、 $y_{12}(f)$ 、 $y_{21}(f)$ （特定信号の一例）を入力し、これらの信号に基づいて前記中間処理を実行する。中間処理の具体的内容については後述する。

また、一方のバイノーラル信号処理部 2 1 は、これに対応する中間処理実行部 4 1 により出力される前記中間処理後信号 $y_{d1}(f)$ と、その中間処理実行部 4 1 が中間処理の対象

10

20

30

40

50

としていない分離信号 $y_{11}(f)$ (SIMO 信号の一部) とを入力し、その入力信号に対してバイナリマスキング処理を行い、最終的な分離信号 $Y_{11}(f)$ 及び $Y_{12}(f)$ を出力する。また、これら周波数領域の分離信号 $Y_{11}(f)$ 及び $Y_{12}(f)$ は、逆離散フーリエ変換処理を実行する IDFT 処理部 15 によって時間小域の分離信号 $y_{11}(t)$ 、 $y_{12}(t)$ に変換される。

同様に、他方のバイノーラル信号処理部 22 は、これに対応する中間処理実行部 42 により出力される前記中間処理後信号 $y_{d2}(f)$ と、その中間処理実行部 42 が中間処理の対象としていない分離信号 $y_{22}(f)$ (SIMO 信号の一部) とを入力し、その入力信号に対してバイナリマスキング処理を行い、最終的な分離信号 $Y_{21}(f)$ 及び $Y_{22}(f)$ を出力する。また、これら周波数領域の分離信号 $Y_{21}(f)$ 及び $Y_{22}(f)$ は、前記 IDFT 処理部 15 によって時間小域の分離信号 $y_{21}(t)$ 、 $y_{22}(t)$ に変換される。

なお、バイノーラル信号処理部 21、22 は、必ずしも 2 チャンネル分の信号分離処理を行うものに限らず、3 チャンネル分以上のバイナリマスキング処理を行うものを採用することも考えられる。

【0023】

次に、図 9 ~ 図 11 を参照しつつ、SIMO-ICA 処理部 10 により得られる SIMO 信号をバイノーラル信号処理部 21 or 22 への入力信号とする場合における、バイノーラル信号処理部 21 or 22 への入力信号の組合せと、バイノーラル信号処理部 21 or 22 による信号分離性能及び分離信号の音質との関係について説明する。ここで、図 9 ~ 図 11 は、SIMO 信号にバイナリマスキング処理を施す前後の信号における周波数成分ごとの信号レベル (振幅) の分布の一例 (第 1 例 ~ 第 3 例) をバークラフにより模式的に表したものである。なお、バイノーラル処理部 21 or 22 は、バイナリマスキング処理を行うものであるとする。

また、以下に示す例では、一方のマイクロホン 111 に近い方の音源 1 の音声信号 $S_1(t)$ を、最終的に分離信号として得たい信号であるものとし、その音源信号 $S_1(t)$ 及びその音を、目的音源信号及び目的音と称する。そして、その他の音源 2 の音声信号 $S_2(t)$ 及びその音を、非目的音源信号及び非目的音と称する。

ところで、4 つの分離信号 $y_{11}(f)$ 、 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ からなる SIMO 信号を 2 入力のバイナリマスキング処理の入力信号とする場合、バイナリマスキング処理への入力信号の組合せは 6 パターン考えられる。その中で、主として目的音源信号 $S_1(t)$ に対応する分離信号 $y_{11}(f)$ を含む組合せは 3 パターン考えられるが、SIMO-ICA 法に基づく音源分離処理の性質上、 $y_{11}(f)$ 及び $y_{22}(f)$ との組合せと、 $y_{11}(f)$ 及び $y_{21}(f)$ との組合せとは、定性的には同じ傾向の性質を有する。よって、図 9 ~ 図 11 は、 $y_{11}(f)$ 及び $y_{12}(f)$ との組合せと、 $y_{11}(f)$ 及び $y_{22}(f)$ との組合せとの各々について、バイナリマスキング処理を行う場合の例について示している。

【0024】

また、図 9 は音源信号各々の周波数成分に重複がない場合、図 10 は同周波数成分に重複がある場合の例を表す。一方、図 11 は、音源信号各々の周波数成分に重複がなく、かつ、目的音源信号 $S_1(t)$ の信号レベルが非目的音源信号 $S_2(t)$ の信号レベルに対して相対的に低い (振幅が小さい) 場合の例を表す。

さらに、図 9 (a)、図 10 (a) 及び図 11 (a) は、バイノーラル信号処理部 21 or 22 への入力信号を、分離信号 $y_{11}(f)$ と $y_{12}(f)$ の組合せ (SIMO 信号) とした場合 (以下、「パターン a」という) の例を表す。

一方、図 9 (b)、図 10 (b) 及び図 11 (b) は、バイノーラル信号処理部 21 or 22 への入力信号を、分離信号 $y_{11}(f)$ と $y_{22}(f)$ の組合せとした場合 (以下、「パターン b」という) の例を表す。

また、図 9 ~ 図 11 において、目的音源信号 $S_1(t)$ の周波数成分に対応する部分のバークラフは網掛け模様により、非目的音源信号 $S_2(t)$ の周波数成分に対応する部分のバークラフは斜線模様により各々表している。

【0025】

図9及び図10に示すように、バイノーラル信号処理部21 or 22への入力信号には、その同定の対象となった音源信号の成分が支配的ではあるものの、それ以外に、ノイズとして他の音源信号の成分も若干混在している。

このようなノイズを含む入力信号(分離信号)に対してバイナリーマスキング処理を施した場合、図9(a)、(b)の出力信号のレベル分布(右側のバーグラフ)に示すように、音源信号各々の周波数成分に重複がない場合には、入力信号の組合せにかかわらず、前記第1音源信号と前記第2音源信号とが良好に分離された分離信号($Y_{11}(f)$ と $Y_{12}(f)$)、及び $Y_{11}(f)$ と $Y_{22}(f)$)が得られる。

このように各音源信号の周波数成分に重複がない場合、バイノーラル信号処理部21 or 22への両入力信号各々において、同定の対象となった音源信号の周波数成分における信号レベルが高く、その他の音源信号の周波数成分における信号レベルが低くなるというレベル差が明確となり、周波数成分ごとの信号レベルに応じて信号分離を行うバイナリーマスキング処理によって信号が確実に分離されやすい。その結果、入力信号の組合せにかかわらず、高い分離性能が得られる。

【0026】

しかしながら、一般に、現実の音響空間(音環境)では、同定対象とする目的音源信号とその他の非目的音源信号との間で、周波数成分(周波数帯域)が全く重ならないということはほとんどなく、複数の音源信号間で多少なりとも周波数成分が重複する。

ここで、音源信号各々の周波数成分に重複がある場合であっても、図10(a)の出力信号 $Y_{11}(f)$ 、 $Y_{12}(f)$ のレベル分布(右側のバーグラフ)に示すように、前記「パターンa」では、音源信号各々で重複する周波数成分において若干のノイズ信号(同定対象以外の音源信号の成分)が残るものの、その他の周波数成分においてはノイズ信号が確実に分離される。

この図10(a)に示す「パターンa」では、バイノーラル信号処理部21 or 22への両入力信号は、同一の音源信号を、各々異なるマイクロホンにより収録された音声信号に基づいて分離(同定)した信号であり、それらの信号レベルは、同定対象の音源からマイクロホンまでの距離に応じたレベル差を有している。そのため、バイナリーマスキング処理において、そのレベル差により信号が確実に分離されやすい。このことが、「パターンa」では、音源信号各々の周波数成分に重複があっても高い分離性能が得られる理由として考えられる。

さらに、図10(a)に示す「パターンa」では、両入力信号ともに同一の音源信号(目的音源信号 $S_1(t)$)の成分が支配的である(即ち、混在する他の音源信号の成分のレベルは小さい)ため、比較的信号レベルの低い同定対象外の音源信号の成分(ノイズ成分)が信号分離に悪影響を及ぼしにくいことも、高い分離性能が得られる理由の1つと考えられる。

【0027】

一方、音源信号各々の周波数成分に重複がある場合、図10(b)に示すように、前記「パターンb」では、音源信号各々で重複する周波数成分において、出力信号(分離信号) $Y_{11}(f)$ において本来出力されるべき信号成分(同定対象の音源信号の成分)が欠損するという不都合な現象が生じる(図10(b)における破線で囲んだ部分)。

このような欠損は、その周波数成分について、同定対象の目的音源信号 $S_1(t)$ のマイクロホン112への入力レベルよりも、非目的音源信号 $S_2(t)$ のマイクロホン112へのレベルの方が高いために生じる現象である。このような欠損が生じると音質が悪化する。

従って、一般的には、前記「パターンa」を採用すれば、良好な分離性能が得られる場合が多いと言える。

【0028】

しかしながら、実際の音響環境では、各音源信号の信号レベルは変化し、状況によっては、図11に示すように、目的音源信号 $S_1(t)$ の信号レベルが非目的音源信号 $S_2(t)$ の信号レベルに対して相対的に低くなることもある。

10

20

30

40

50

このような場合、S I M O - I C A 処理部 1 0 で十分な音源分離がなされなかった結果、マイクロホン 1 1 1 に対応する分離信号 $y_{11}(f)$ 及び $y_{12}(f)$ に残留する非目的音源信号 $S_2(t)$ の成分が相対的に大きくなる。このため、図 1 1 (a) に示す「パターン a」を採用すると、図 1 1 (a) において矢印で示すように、目的音源信号 $S_1(t)$ に対応するものとして出力される分離信号 $Y_{11}(f)$ に、非目的音源信号 $S_1(t)$ の成分が残存するという不都合な減少が生じてしまう。この現象が発生すると、音源分離性能が悪化する。

これに対し、図 1 1 (b) に示す「パターン b」を採用すると、具体的な信号レベルにもよるが、出力信号 $Y_{11}(f)$ に図 1 1 (a) の矢印で示したような非目的音源信号 $S_1(t)$ の成分が残存することを回避できる可能性が高い。

【 0 0 2 9 】

次に、図 1 2 及び図 1 3 を参照しつつ、音源分離装置 X 1 により音源分離処理を行った場合の効果について説明する。

図 1 2 は、音源分離装置 X 1 における S I M O 信号に対する音源分離処理の第 1 例の内容 (S I M O 信号及びバイナリーマスキング処理後の信号についての周波数成分ごとの信号レベル分布を含む) を模式的に表した図である。なお、図 1 2 には、バイノーラル信号処理部 2 1 及びこれに対応する中間処理実行部 4 1 のみをピックアップして表記している。

図 1 2 に示す例では、中間処理実行部 4 1 は、まず、3 つの分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ (特定信号の一例) を、所定の周波数幅で均等に区分された周波数成分ごとに、その周波数成分の信号に所定の重み係数 a_1 、 a_2 、 a_3 を乗算することによって信号レベルを補正 (即ち、重み付けにより補正) し、さらに、その補正後の信号の中から、前記周波数成分ごとに信号レベルが最大のものを選択する中間処理を行う。この中間処理を、 $Max [a_1 \cdot y_{12}(f), a_2 \cdot y_{21}(f), a_3 \cdot y_{22}(f)]$ と表すものとする。

さらに、中間処理実行部 4 1 は、この中間処理により得られた中間処理後信号 $y_{d1}(f)$ (周波数成分ごとに信号レベルが最大のものが組み合わされた信号) をバイノーラル信号処理部 2 1 へ出力する。ここで、 $a_2 = 0$ かつ $a_1 > a_3$ である。例えば、 $a_1 = 1$ 、 $a_3 = 0.5$ である。なお、 $a_2 = 0$ であるため、分離信号 $y_{21}(f)$ の周波数分布の表記は省略している。また、図 1 2 に示す S I M O 信号は、図 1 0 に示した S I M O 信号と同じである。

このように、 $a_1 > a_3$ となるように重み付け補正をした後の信号のうち、周波数成分ごとに信号レベルが最大のものをバイナリーマスキング処理の入力信号とすることにより、音源分離装置 X 1 は、以下のように動作する。

即ち、分離信号 $y_{12}(f)$ が、分離信号 $y_{22}(f)$ に対して $a_1 \cdot y_{12}(f) > a_3 \cdot y_{22}(f)$ となる範囲の信号レベルで出力されている周波数成分については、バイノーラル信号処理部 2 1 には分離信号 $y_{11}(f)$ と分離信号 $y_{12}(f)$ とが入力され、図 9 (a) や図 1 0 (a) に示したような良好な信号分離状況が得られると考えられる。

一方、分離信号 $y_{12}(f)$ が、分離信号 $y_{22}(f)$ に対して $a_1 \cdot y_{12}(f) < a_3 \cdot y_{22}(f)$ となる範囲の信号レベルまで低下している周波数成分については、バイノーラル信号処理部 2 1 には、分離信号 $y_{11}(f)$ と、分離信号 $y_{22}(f)$ が (a_3) 倍に減縮補正された信号とが入力され、図 9 (a) や図 1 1 (b) に示したような良好な信号分離状況が得られると考えられる。

【 0 0 3 0 】

図 1 3 は、音源分離装置 X 1 における S I M O 信号に対する音源分離処理の第 2 例の内容 (S I M O 信号及びバイナリーマスキング処理後の信号についての周波数成分ごとの信号レベル分布を含む) を模式的に表した図である。

図 1 3 に示す例も、図 1 2 に示した例と同様に、中間処理実行部 4 1 は、まず、3 つの分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ (特定信号の一例) を、所定の周波数幅で均等に区分された周波数成分ごとに、その周波数成分の信号に所定の重み係数 a_1 、 a_2 、 a_3 を乗算することによって信号レベルを補正 (即ち、重み付けして補正) し、さらに、その補正後の信号の中から、前記周波数成分ごとに信号レベルが最大のものを選択する中間処

10

20

30

40

50

理（図中、 $\text{Max} [a_1 \cdot y_{12}(f), a_2 \cdot y_{21}(f), a_3 \cdot y_{22}(f)]$ と表記）を行う。さらに、中間処理実行部 4 1 は、この中間処理により得られた中間処理後信号 $y_{d1}(f)$ （周波数成分ごとに信号レベルが最大のもので組み合わせられた信号）をバイノーラル信号処理部 2 1 へ出力する。例えば、 $1 \quad a_1 > a_2 > a_3 \quad 0$ である。

同様に、中間処理実行部 4 2 は、まず、3つの分離信号 $y_{11}(f)$ 、 $y_{12}(f)$ 、 $y_{21}(f)$ （特定信号の一例）を、所定の周波数幅で均等に区分された周波数成分ごとに、その周波数成分の信号に所定の重み係数 b_1 、 b_2 、 b_3 を乗算することによって信号レベルを補正し、さらに、その補正後の信号の中から、前記周波数成分ごとに信号レベルが最大のものを選択する中間処理（図中、 $\text{Max} [b_1 \cdot y_{11}(f), b_2 \cdot y_{12}(f), b_3 \cdot y_{21}(f)]$ と表記）を行う。さらに、中間処理実行部 4 2 は、この中間処理により得られた中間処理後信号 $y_{d2}(f)$ （周波数成分ごとに信号レベルが最大のもので組み合わせられた信号）をバイノーラル信号処理部 2 2 へ出力する。例えば、 $1 \quad b_1 > b_2 > b_3 \quad 0$ である。なお、図 1 3 に示す S I M O 信号は、図 1 0 に示した S I M O 信号と同じである。

このような第 2 例においても、前記第 1 例（図 1 2 参照）で説明したのと同様の作用効果を奏する。

【 0 0 3 1 】

図 1 8 は、音源分離装置 X 1 における S I M O 信号に対する音源分離処理の第 3 例の内容（S I M O 信号及びバイナリーマスキング処理後の信号についての周波数成分ごとの信号レベル分布を含む）を模式的に表した図である。

図 1 8 に示す第 3 例は、図 1 3 に示した前記第 2 例に対して前記中間処理実行部 4 1、4 2 が実行する処理と、前記バイノーラル信号処理部 2 1、2 2 が実行する処理とが若干異なるが、全体として前記第 2 例（図 1 3 参照）と実質的に同じ処理を実行する音源分離装置 X 1 を表す。

即ち、図 1 8 に示す第 3 例においては、前記中間処理実行部 4 1 は、まず、4つの分離信号 $y_{11}(f)$ 、 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ （特定信号の一例）を、所定の周波数幅で均等に区分された周波数成分ごとに、その周波数成分の信号に所定の重み係数（ 1 、 a_1 、 a_2 、 a_3 ）を乗算することによって信号レベルを補正（即ち、重み付けして補正）し、さらに、その補正後の信号の中から、前記周波数成分ごとに信号レベルが最大のものを選択する中間処理（図中、 $\text{Max} [y_{11}, a_1 \cdot y_{12}(f), a_2 \cdot y_{21}(f), a_3 \cdot y_{22}(f)]$ と表記）を行う。さらに、前記中間処理実行部 4 1 は、この中間処理により得られた前記中間処理後信号 $y_{d1}(f)$ （周波数成分ごとに信号レベルが最大のもので組み合わせられた信号）を前記バイノーラル信号処理部 2 1 へ出力する。例えば、 $1 \quad a_1 > a_2 > a_3 \quad 0$ である。

同様に、前記中間処理実行部 4 2 は、まず、4つの分離信号 $y_{11}(f)$ 、 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ （特定信号の一例）を、所定の周波数幅で均等に区分された周波数成分ごとに、その周波数成分の信号に所定の重み係数（ b_1 、 b_2 、 b_3 、 1 ）を乗算することによって信号レベルを補正し、さらに、その補正後の信号の中から、前記周波数成分ごとに信号レベルが最大のものを選択する中間処理（図中、 $\text{Max} [b_1 \cdot y_{11}(f), b_2 \cdot y_{12}(f), b_3 \cdot y_{21}(f), y_{22}(f)]$ と表記）を行う。さらに、前記中間処理実行部 4 2 は、この中間処理により得られた前記中間処理後信号 $y_{d2}(f)$ （周波数成分ごとに信号レベルが最大のもので組み合わせられた信号）をバイノーラル信号処理部 2 2 へ出力する。例えば、 $1 \quad b_1 > b_2 > b_3 \quad 0$ である。なお、図 1 8 に示す S I M O 信号は、図 1 0 に示した S I M O 信号と同じである。

【 0 0 3 2 】

ここで、この第 3 例における前記バイノーラル信号処理部 2 1 は、これに入力される信号（分離信号 $y_{11}(f)$ 及び前記中間処理後信号 $y_{d1}(f)$ ）について、周波数成分ごとに以下の処理を実行する。

即ち、前記バイノーラル信号処理部 2 1 は、周波数成分ごとに、前記中間処理後信号 $y_{d1}(f)$ の信号レベルが前記分離信号 $y_{11}(f)$ の信号レベルと等しい場合（同じ信号である場合）にはその中間処理後信号 $y_{d1}(f)$ 又は分離信号 $y_{11}(f)$ の成分を出力信号 $Y_{11}(f)$

10

20

30

40

50

の信号成分として採用し、そうでない場合には予め定められた定数値（ここでは、0 値）を出力信号 $Y_{11}(f)$ の信号成分として採用する。

同様に、この第 3 例における前記バイノーラル信号処理部 2 2 は、これに入力される信号（分離信号 $y_{22}(f)$ 及び前記中間処理後信号 $y_{d2}(f)$ ）について、周波数成分ごとに、前記分離信号 $y_{22}(f)$ の信号レベルと前記中間処理後信号 $y_{d2}(f)$ の信号レベルとが等しい場合（同じ信号である場合）には、その分離信号 $y_{22}(f)$ 又はその中間処理後信号 $y_{d2}(f)$ の成分を出力信号 $Y_{22}(f)$ の信号成分として採用し、そうでない場合には予め定められた定数値（ここでは、0 値）を出力信号 $Y_{22}(f)$ の信号成分として採用する。

ここで、前記バイノーラル信号処理部 2 1 は、一般的なバイナリマスキング処理を実行する場合、周波数成分ごとに、前記分離信号 $y_{11}(f)$ の信号レベルが前記中間処理後信号 $y_{d1}(f)$ の信号レベル以上である場合（ $y_{11}(f) \geq y_{d1}(f)$ ）には、その分離信号 $y_{11}(f)$ の成分を出力信号 $Y_{11}(f)$ の信号成分として採用し、そうでない場合には予め定められた定数値（ここでは、0 値）を出力信号 $Y_{11}(f)$ の信号成分として採用する。

しかしながら、前記中間処理実行部 4 1 において、バイナリマスキング処理の対象となる（重み係数「1」が乗算される）前記分離信号 $y_{11}(f)$ と、重み係数 $a_1 \sim a_3$ が乗算されるその他の前記分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ とについて、周波数成分ごとにレベルが最大のものが選択された信号が前記中間処理後信号 $y_{d1}(f)$ とされる。このため、前述したように、前記バイノーラル信号処理部 2 1 が、「 $y_{11}(f) = y_{d1}(f)$ 」である場合に、前記分離信号 $y_{11}(f)$ 又は前記中間処理後信号 $y_{d1}(f)$ の成分を出力信号 $Y_{11}(f)$ の信号成分として採用するものであっても、そのバイノーラル信号処理部 2 1 は、実質的に一般的なバイナリマスキング処理を実行するものと実質的に同じ（等価）である。このことは、前記バイノーラル信号処理部 2 2 についても同様である。

ここで、一般的なバイナリマスキング処理とは、「 $y_{11}(f) \geq y_{d1}(f)$ 」であるか否かにより、出力信号 $Y_{11}(f)$ の信号成分として前記分離信号 $y_{11}(f)$ 又は前記中間処理後信号 $y_{d1}(f)$ の成分を採用するか、定数値（0 値）を採用するかを切り替える処理である。

従って、図 1 8 に示すこれら中間処理実行部 4 1、4 2 及びバイノーラル信号処理部 2 1、2 2 も、本発明に係る音源分離装置を構成する前記中間処理実行手段及び前記第 2 の音源分離手段の実施形態の一例である。

以上に示した第 3 例においても、前記第 1 例（図 1 2 参照）で説明したのと同様の作用効果を奏する。

【0033】

次に、前記音源分離装置 X 1 を用いた音源分離性能評価の実験結果について説明する。

図 1 4 は、音源分離装置 X 1 を用いた音源分離性能評価の実験条件を説明するための図である。

図 1 4 に示すように、音源分離性能評価の実験は、4.8 m（幅）× 5.0 m（奥行き）の広さの居室内において、所定の 2 箇所に存在する 2 名の話者を音源とし、その音源（話者）各々からの音声信号（話者の声）を相互に反対方向に向けた 2 つのマイクロホン 1 1 1、1 1 2 で入力し、入力された 2 チャンネルの混合音声信号から、各話者の音声信号（音源信号）を分離する性能を評価する実験条件である。ここで、音源となる話者は、2 名の男性及び 2 名の女性（計 4 名）の中から選択した 2 名の順列である 1 2 通りの条件下で実験を行い（同じ 2 名の話者を音源とする場合でも、その 2 名の配置が入れ替われば異なる条件であるとした）、音源分離性能評価は各組合せの下での評価値の平均値により行った。

また、いずれの実験条件においても、残響時間は 200 ms、音源（話者）から最寄りのマイクロホンまでの距離は 1.0 m とし、2 つのマイクロホン 1 1 1、1 1 2 は、5.8 cm の間隔を隔てて配置した。なお、マイクロホンの機種は、SONY 社製の ECM-DS70P である。

ここで、上方から見て、相互に反対方向に向けられた両マイクロホン 1 1 1、1 1 2 の向きに対する直角方向を基準方向 R 0 としたときに、この基準方向 R 0 と一方の音源 S 1

10

20

30

40

50

(話者)から両マイクロホン111、112の中間点Oに向かう方向R1とのなす角度を1とする。また、前記基準方向R0と他方の音源S2(話者)から前記中間点Oに向かう方向R2とのなす角度2とする。このとき、1と2との組合せを、3パターンの条件(1、2) = (-40°、30°)、(-40°、10°)、(-10°、10°)となるように設定(機器配置)し、その各々の条件下で実験を行った。

【0034】

図15(a)、(b)は、従来の音源分離装置と本発明に係る音源分離装置との各々により、前述の実験条件の下で音源分離を行ったときの音源分離性能及び分離後の音声の音質の評価結果を表すグラフである。

ここで、図15(a)に示す音源分離性能の評価値(グラフの縦軸)としては、NR R (Noise Reduction Rate)を用いた。このNR Rは、雑音除去の程度を表す指標であり、単位は(dB)である。NR Rの定義は、例えば非特許文献2の(21)式等に示されている。このNR R値が大きいほど音源分離性能が高いといえる。

また、図15(b)に示す音質の評価値(グラフの縦軸)としては、CD(Cepstral distortion)を用いた。このCDは、音質の程度を表す指標であり単位は(dB)である。このCDは、音声信号のスペクトル歪みを表し、分離対象となる元の音源信号と、その音源信号を混合音声信号から分離した分離信号とのスペクトル包絡の距離を表す。CD値が小さいほど音質が良いといえる。なお、図15(b)に示す音質評価の結果は、(1、2) = (-40°、30°)である場合のもののみである。

【0035】

また、各バーグラフに対応する図中の表記P1~P6は、以下の場合の処理結果を表す。

P1(BM)と表記しているものは、バイナリーマスキング処理を行った場合の結果を表す。

P2(ICA)と表記しているものは、図6に示したFD-SIMO-ICA法に基づく音源分離処理を行った場合の結果を表す。

P3(ICA+BM)と表記しているものは、図6に示したFD-SIMO-ICA法に基づく音源分離処理(音源分離処理装置Z4)により得られたSIMO信号にバイナリーマスキング処理を施した場合の結果を表す。即ち、図9~図11に示した構成により音源分離処理を行った結果に相当する。

P4~P6(SIMO-ICA+SIMO-BM)と表記しているものは、図2に示した音源分離処理装置X1により音源分離処理を行った結果を表す。ここで、P4は、補正係数 $[a_1, a_2, a_3] = [1.0, 0, 0]$ である場合、P5は補正係数 $[a_1, a_2, a_3] = [1, 0, 0.1]$ である場合、P6は補正係数 $[a_1, a_2, a_3] = [1.0, 0, 0.7]$ である場合を表す。以下、P4、P5、P6の各補正係数の条件を、補正パターンP4、補正パターンP5、補正パターンP6という。

【0036】

図15に示すグラフから、バイナリーマスキング処理やICA法に基づくBSS方式の音源分離処理を単独で行う場合(P1、P2)や、これにより得られるSIMO信号にバイナリーマスキング処理を施す場合(P3)よりも、ICA法に基づくBSS方式の音源分離処理で得られるSIMO信号に基づいて前記中間処理を行い、その中間処理後の信号を用いてバイナリーマスキング処理を施す音源分離処理である本発明に係る音源分離処理(P4~P6)の方が、NR R値が大きく、音源分離性能が優れていることがわかる。

同様に、P1~P3の音源分離処理よりも、本発明に係る音源分離処理(P4~P6)の方が、CD値が小さく、分離後の音声信号が高音質であることがわかる。

また、本発明に係る音源分離処理(P4~P6)の中では、補正パターンP4及びP5に設定した場合、音源分離性能向上と音質性能向上とのバランスがとれている。これは、図10及び図11を用いて説明した不都合な現象の発生が少ないため、音源分離性能及び音質性能が高まったものと考えられる。

一方、補正パターンP6では、補正パターンP4及びP5よりも、さらに高い音源分離

10

20

30

40

50

性能が得られている（NRR値が高い）反面、音質性能が若干犠牲になっている（CD値がやや高い）。これは、補正パターンP4及びP5の場合よりも、図11を用いて説明した不都合な現象の発生頻度がより抑えられたことにより、音源分離性能がさらに向上する一方、図10を用いて説明した不都合な現象の発生頻度が若干増え、その結果、音質性能がやや犠牲になっているものと考えられる。

【0037】

以上に示したように、音源分離装置X1では、中間処理実行部41、42における前記中間処理に用いられるパラメータ（重み係数 $a_1 \sim a_3$ 、 $b_1 \sim b_3$ ）を調節するだけで、重視する目的（音源分離性能又は音質性能）に応じた音源分離処理が可能となる。

従って、音源分離装置X1が、調節ツマミや、数値入力用操作キー等の操作入力部（中間処理パラメータ設定手段の一例）を備え、さらに、前記中間処理実行部41、42が、その操作入力部を通じて入力された情報に従って、中間処理実行部41、42（中間処理実行手段の一例）における前記中間処理に用いられるパラメータ（ここでは、重み係数 $a_1 \sim a_3$ 、 $b_1 \sim b_3$ ）を設定（調節）する機能を備えたものであれば、重視する目的に応じた装置の調整が容易となる。

例えば、音源分離装置X1が、ロボットやカーナビゲーションシステム等で用いられる音声認識装置に適用される場合、雑音除去を優先するために、NRR値が高くなる方向に、重み係数 $a_1 \sim a_3$ 、 $b_1 \sim b_3$ を設定すればよい。

一方、音源分離装置X1が、携帯電話機やハンズフリー電話機等の音声通話装置に適用される場合、音質が良くなるように、CD値が高くなる方向に、重み係数 $a_1 \sim a_3$ 、 $b_1 \sim b_3$ を設定すればよい。

より具体的には、重み係数 a_1 、 b_1 の値に対する重み係数 a_2 、 a_3 、 b_2 、 b_3 の値の比が、より大きくなるように設定すれば、音源分離性能を重視するという目的に沿い、その比が、より小さくなるように設定すれば、音質性能を重視するという目的に沿うことになる。

【0038】

また、以上に示した実施例では、中間処理実行部41、42により、 $\text{Max} [a_1 \cdot y_{12}(f), a_2 \cdot y_{21}(f), a_3 \cdot y_{22}(f)]$ 或いは、 $\text{Max} [b_1 \cdot y_{11}(f), b_2 \cdot y_{12}(f), b_3 \cdot y_{21}(f)]$ という中間処理を行う例を示した。

しかしながら、前記中間処理は、これに限るものではない。

中間処理実行部41、42により実行される前記中間処理としては、以下のような例も考えられる。

即ち、まず、中間処理実行部41が、3つの分離信号 $y_{12}(f)$ 、 $y_{21}(f)$ 、 $y_{22}(f)$ （特定信号の一例）を、所定の周波数幅で均等に区分された周波数成分ごとに、その周波数成分の信号に所定の重み係数 a_1 、 a_2 、 a_3 を乗算することによって信号レベルを補正（即ち、重み付けして補正）する。さらに、その補正後の信号を、前記周波数成分ごとに合成（加算）する。即ち、 $a_1 \cdot y_{12}(f) + a_2 \cdot y_{21}(f) + a_3 \cdot y_{22}(f)$ という中間処理を行う。

さらに、中間処理実行部41は、この中間処理により得られた中間処理後信号 $y_{d1}(f)$ （周波数成分ごとに重み付け補正がなされた信号を合成した）をバイノーラル信号処理部21へ出力する。

このような中間処理を採用しても、前述した実施例と同様の作用効果が得られる。もちろん、このような2種類の中間処理に限られず、他の中間処理を採用することも考えられる。また、チャンネル数を、3チャンネル以上に拡張した構成も考えられる。

【0039】

前述したように、ICA法に基づくBSS方式による音源分離処理は、音源分離性能を向上させるためには多大な演算が必要となり、リアルタイム処理には適さない。

一方、バイノーラル信号処理による音源分離は、一般に演算量は少なくリアルタイム処理に適している反面、音源分離性能はICA法に基づくBSS方式による音源分離処理に比べて劣る。

10

20

30

40

50

これに対し、SIMO-ICA処理部10が、例えば以下に示す要領で分離行列 $W(f)$ の学習を行うように構成すれば、音源信号の分離性能を確保しつつリアルタイム処理が可能な音源分離処理装置を実現できる。

【0040】

次に、図16及び図17に示すタイムチャートを用いて、分離行列 $W(f)$ の学習に用いられる混合音声信号と、その学習により得られる分離行列 $W(f)$ を用いて音源分離処理が施される混合音声信号との対応関係の第1例(図16)及び第2例(図17)について説明する。

ここで、図16は、分離行列 $W(f)$ の計算と、音源分離処理との各々に用いる混合音声信号の区別の第1例をタイムチャートの的に表したものである。

10

この第1例は、SIMO-ICA処理部10の音源分離処理において、逐次入力される混合音声信号を、所定時間長(例えば3秒)分のフレーム信号(以下、Frameという)ごとにその全てを用いて学習計算を行う。その一方で、SIMO-ICA処理部10の音源分離処理における分離行列の逐次計算回数を制限するものである。また、図1に示す例では、SIMO-ICA処理部10は、分離行列の学習計算と、その分離行列に基づくフィルタ処理(行列演算)により分離信号を生成(同定)する処理とを、異なるFrameを用いて実行する。

図16に示すように、SIMO-ICA処理部10は、時刻 $T_i \sim T_{i+1}$ の期間(周期： $T_{i+1} - T_i$)に入力された前記混合音声信号全てに相当するFrame(i)を用いて分離行列の計算(学習)を行い、それにより求められた分離行列を用いて時刻 $(T_{i+1} + T_d) \sim (T_{i+2} + T_d)$ の期間に入力された前記混合音声信号全てに相当するFrame(i+1)'について分離処理(フィルタ処理)を実行する。ここで、 T_d は1つのFrameを用いた分離行列の学習に要する時間である。即ち、ある1期間の混合音声信号に基づき計算された分離行列を用いて、Frame時間長+学習時間だけずれた次の1期間の混合音声信号の分離処理(同定処理)を行う。このとき、ある1期間のFrame(i)を用いて計算(学習)された分離行列を、次の1期間のFrame(i+1)'を用いて分離行列を計算(逐次計算)する際の初期値(初期分離行列)として用いる。さらに、SIMO-ICA処理部10は、分離行列の逐次計算(学習計算)の繰り返し回数を、1フレーム分の時間長(周期)の範囲内の時間 T_d で実行可能な回数に制限する。

20

【0041】

以上示したように、図16(第1例)に示すタイムチャートに従って分離行列の計算を行うSIMO-ICA処理部10は、時系列に入力される前記混合音声信号を予め定められた周期で区分されたFrame(区間信号の一例)ごとに、そのFrameに対し所定の分離行列に基づく分離処理を順次実行して前記SIMO信号を生成するものであり、また、その分離処理により生成した全ての時間帯(Frame(区間信号)の時間帯に対応する全ての時間帯)の前記SIMO信号に基づいて、以降に用いる前記分離行列を求めるための逐次計算(学習計算)を行うものである。

30

このように、1つのFrameの時間長以内に、1つのFrame全体に基づく分離行列の学習計算を完了させることができれば、全ての混合音声信号を学習計算に反映させながらリアルタイムでの音源分離処理が可能となる。

40

しかしながら、学習計算を複数のプロセッサで分担して並行処理した場合でも、1つのFrameの時間範囲内($T_i \sim T_{i+1}$)に、十分な音源分離性能を確保できるだけの十分な学習計算(逐次計算処理)を常には完了させられないことも考え得る。

そこで、この第1例におけるSIMO-ICA処理部10は、分離行列の逐次計算の回数を、Frame(区間信号)の時間(予め定められた周期)の範囲に収まる時間 T_d で実行可能な回数に制限する。これにより、学習計算の収束が早まり、リアルタイム処理が可能となる。

【0042】

一方、図17に示す第2例は、逐次入力される混合音声信号を所定時間長(例えば3秒)分のフレーム信号(Frame)ごとに、そのフレーム信号の先頭側の一部を用いて学習計

50

算を行う例、即ち、分離行列の逐次計算に用いる混合音声信号のサンプル数を通常よりも減らす（間引く）例である。

これにより、学習計算の演算量が抑えられるので、より短周期で分離行列の学習を行うことが可能となる。

図17も、図16と同様に、分離行列 $W(f)$ の計算と、音源分離処理との各々に用いる混合音声信号の区分の第2例をタイムチャートの表したものである。

また、図17に示す第2例も、分離行列の学習計算と、その分離行列に基づくフィルタ処理（行列演算）により分離信号を生成（同定）する処理とを、異なるFrameを用いて実行する例である。

この第2例では、図17に示すように、時刻 $T_i \sim T_{i+1}$ の期間（周期： $T_{i+1} - T_i$ ）に入力された前記混合音声信号（Frame）であるFrame(i)のうち、先頭側の一部（例えば、先頭から所定時間分）の信号（以下、Sub-Frame(i)という）を用いて分離行列の計算（学習）を行い、それにより求められた分離行列を用いて時刻 $T_{i+1} \sim T_{i+2}$ の期間に入力された前記混合音声信号全てに相当するFrame(i+1)について分離処理（フィルタ処理）を実行する。即ち、ある1期間の混合音声信号の先頭側の一部に基づき計算された分離行列を用いて次の1期間の混合音声信号の分離処理（同定処理）を行う。このとき、ある1期間のFrame(i)の先頭側の一部を用いて計算（学習）された分離行列を、次の1期間のFrame(i+1)を用いて分離行列を計算（逐次計算）する際の初期値（初期分離行列）として用いる。これにより、逐次計算（学習）の収束が早まり好適である。

【0043】

以上示したように、図17（第2例）に示すタイムチャートに従って分離行列の計算を行うSIMO-ICA処理部10も、時系列に入力される前記混合音声信号を予め定められた周期で区分されたFrame（区間信号の一例）ごとに、そのFrameに対し所定の分離行列に基づく分離処理を順次実行して前記SIMO信号を生成するものであり、また、その分離処理により生成した全ての時間帯（Frame（区間信号）の時間帯に対応する全ての時間帯）の前記SIMO信号に基づいて、以降に用いる前記分離行列を求める逐次計算（学習計算）を行うものである。

さらに、この第2例に対応するSIMO-ICA処理部10は、分離行列を求める学習計算に用いる混合音声信号を、フレーム信号ごとにその先頭側の一部の時間帯の信号に限定する。これにより、より短周期での学習計算が可能となり、その結果、リアルタイム処理が可能となる。

【産業上の利用可能性】

【0044】

本発明は、音源分離装置への利用が可能である。

【図面の簡単な説明】

【0045】

【図1】本発明の実施形態に係る音源分離装置Xの概略構成を表すブロック図。

【図2】本発明の第1実施例に係る音源分離装置X1の概略構成を表すブロック図。

【図3】TDICA法に基づくBSS方式の音源分離処理を行う従来の音源分離装置Z1の概略構成を表すブロック図。

【図4】TD-SIMO-ICA法に基づく音源分離処理を行う従来の音源分離装置Z2の概略構成を表すブロック図。

【図5】FDICA法に基づく音源分離処理を行う従来の音源分離装置Z3の概略構成を表すブロック図。

【図6】FD-SIMO-ICA法に基づく音源分離処理を行う音源分離装置Z4の概略構成を表すブロック図。

【図7】FDICA-PB法に基づく音源分離処理を行う従来の音源分離装置Z5の概略構成を表すブロック図。

【図8】バイナリマスキング処理を説明するための図。

【図9】SIMO信号にバイナリマスキング処理を施す前後の信号における周波数成分

10

20

30

40

50

ごとの信号レベル分布の第 1 例（音源信号各々の周波数成分に重複がない場合）を模式的に表した図。

【図 1 0】S I M O 信号にバイナリーマスキング処理を施す前後の信号における周波数成分ごとの信号レベル分布の第 2 例（音源信号各々の周波数成分に重複がある場合）を模式的に表した図。

【図 1 1】S I M O 信号にバイナリーマスキング処理を施す前後の信号における周波数成分ごとの信号レベル分布の第 3 例（目的音源信号のレベルが比較的小さい場合）を模式的に表した図。

【図 1 2】音源分離装置 X 1 における S I M O 信号に対する音源分離処理の第 1 例の内容を模式的に表した図。

10

【図 1 3】音源分離装置 X 1 における S I M O 信号に対する音源分離処理の第 2 例の内容を模式的に表した図。

【図 1 4】音源分離装置 X 1 を用いた音源分離性能評価の実験条件を表す図。

【図 1 5】従来の音源分離装置と本発明に係る音源分離装置との各々により所定の実験条件の下で音源分離を行ったときの音源分離性能及び音質の評価値を表すグラフ。

【図 1 6】音源分離装置 X における分離行列計算の第 1 例を説明するためのタイムチャート。

【図 1 7】音源分離装置 X における分離行列計算の第 2 例を説明するためのタイムチャート。

【図 1 8】音源分離装置 X 1 における S I M O 信号に対する音源分離処理の第 3 例の内容を模式的に表した図。

20

【符号の説明】

【 0 0 4 6 】

X ... 本発明の実施形態に係る音源分離装置

X 1 ... 本発明の第 1 実施例に係る音源分離装置

1、2 ... 音源

1 0 ... S I M O - I C A 処理部

1 1、1 1 f ... 分離フィルタ処理部

1 2 ... Fidelity Controller

1 3 ... S T - D F T 処理部

1 4 ... 逆行列演算部

1 5 ... I D F T 処理部

2 1、2 2 ... バイノーラル信号処理部

3 1 ... バイナリーマスキング処理における比較部

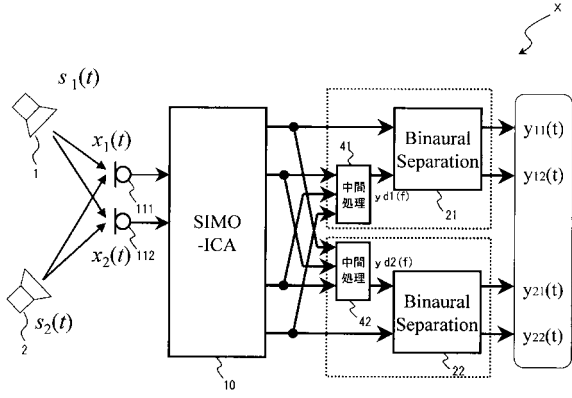
3 2 ... バイナリーマスキング処理における分離部

4 1、4 2 ... 中間処理実行部

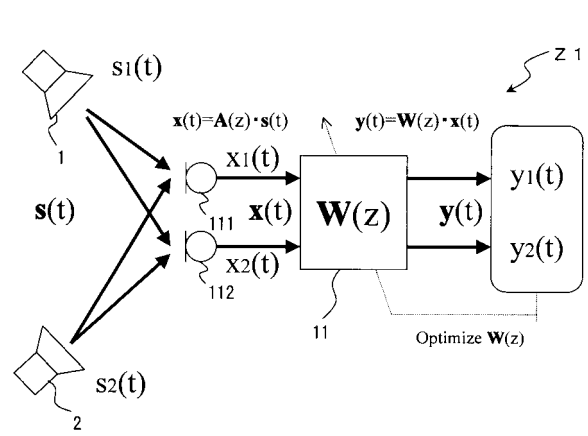
1 1 1、1 1 2 ... マイクロホン

30

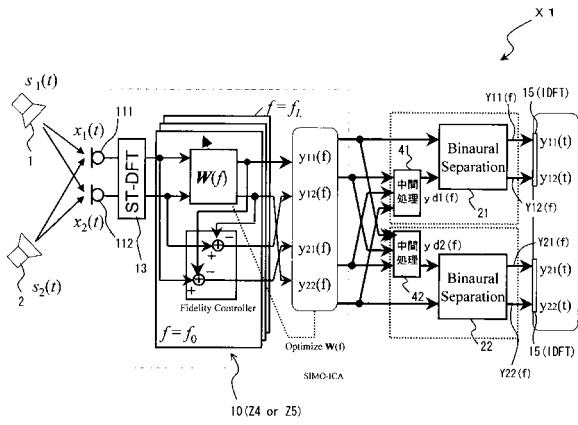
【図1】



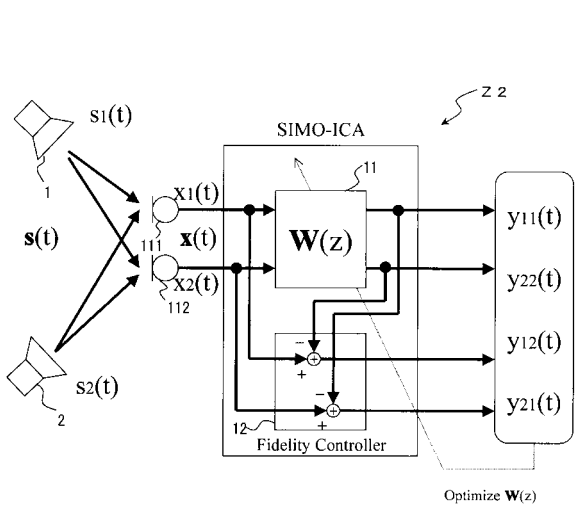
【図3】



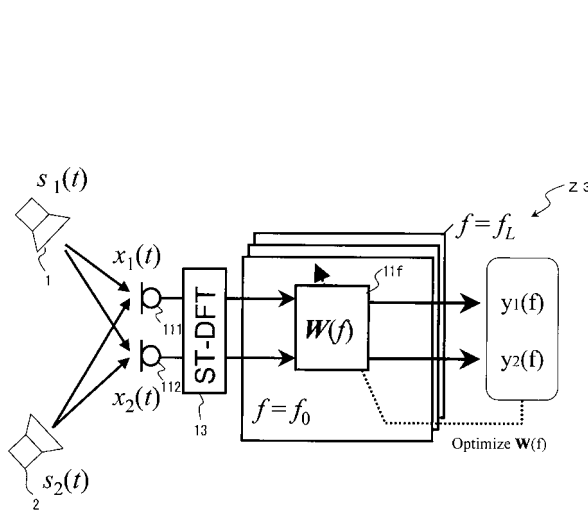
【図2】



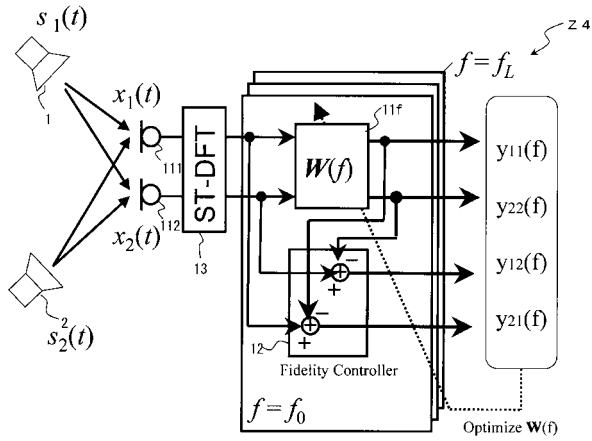
【図4】



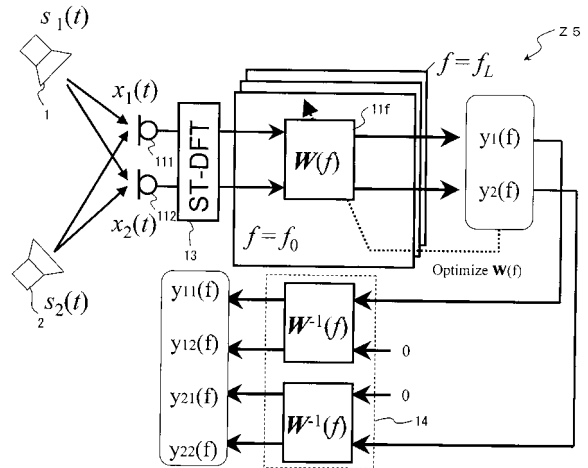
【図5】



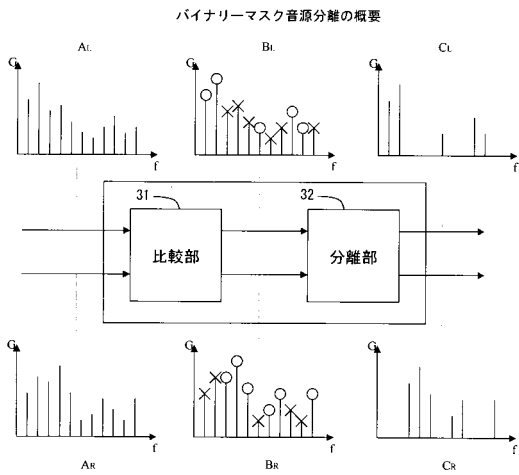
【図6】



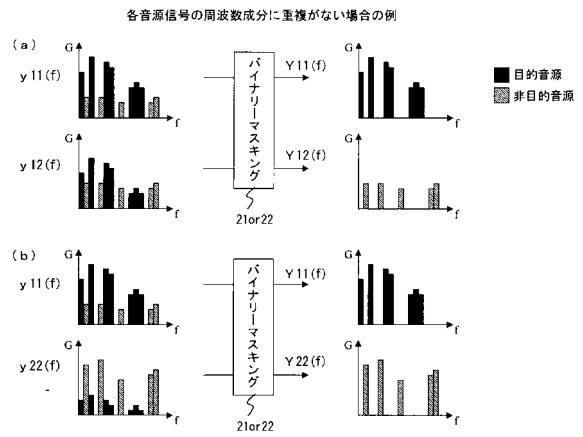
【図7】



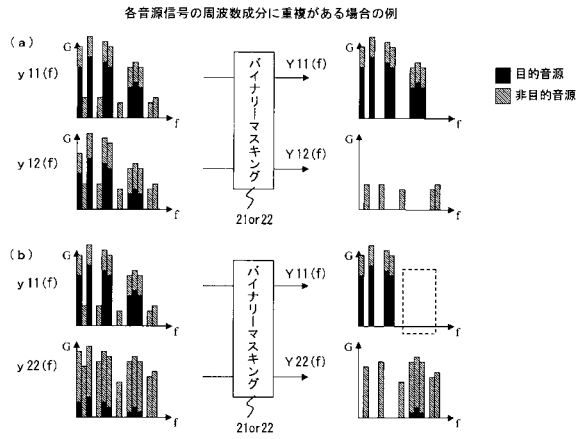
【図8】



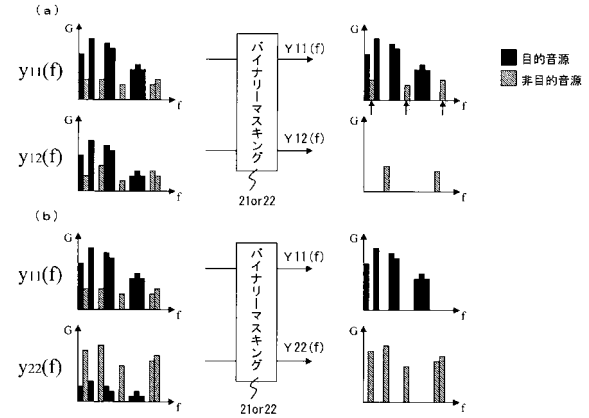
【図9】



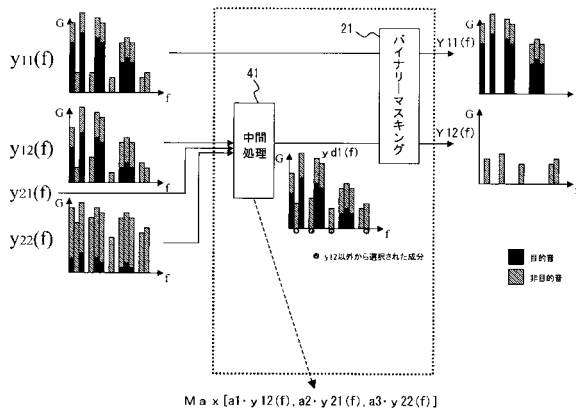
【図10】



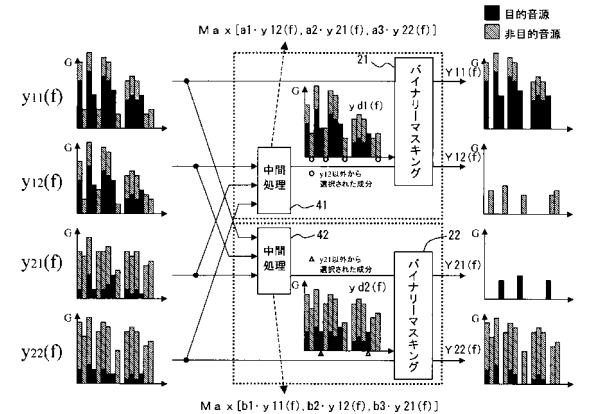
【図11】



【図12】

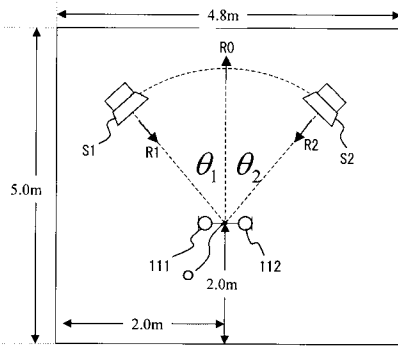


【図13】

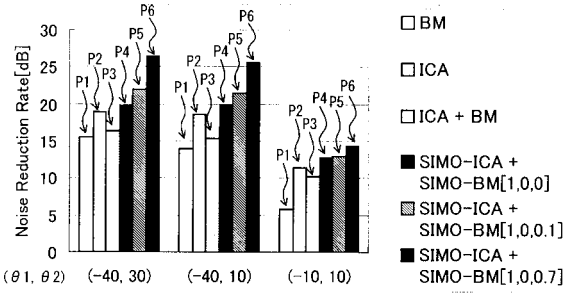


【 図 1 4 】

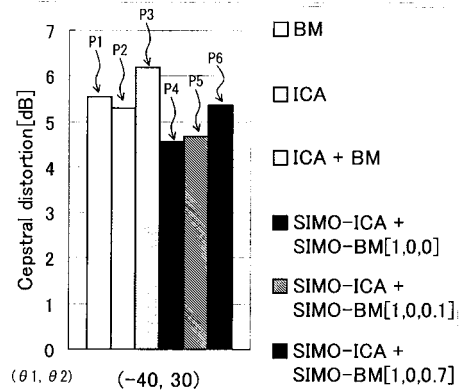
残響時間	200ms
音源信号	話者2名(男女各2名から12通りを選択)
音源の高さ	1.0m
マイク間距離、マイク種	5.8cm (Sony ECM-DS70P)
マイク-話者間距離	1.0m
角度(θ_1, θ_2)	(-40°, 30°), (-40°, 10°), (-10°, 10°)



【 図 1 5 】

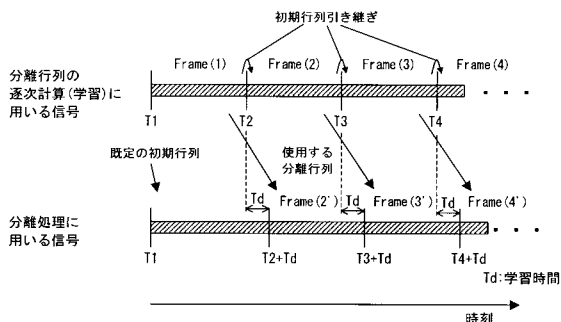


(a)分離性能評価結果

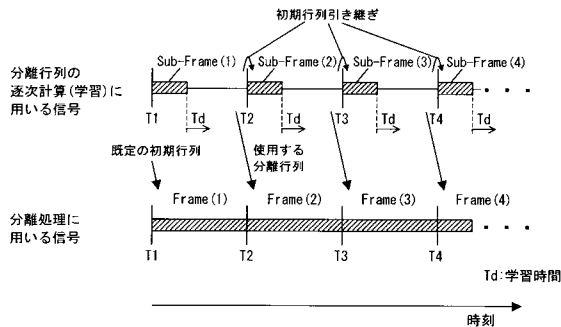


(b)音質評価結果

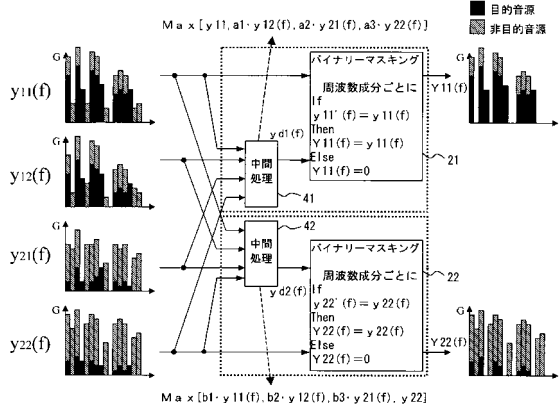
【 図 1 6 】



【 図 1 7 】



【 図 1 8 】



フロントページの続き

- (72)発明者 森田 孝司
兵庫県神戸市西区高塚台1丁目5番5号 株式会社神戸製鋼所 神戸総合技術研究所内
- (72)発明者 猿渡 洋
奈良県生駒市高山町8916-5
- (72)発明者 森 康充
奈良県生駒市高山町8916-5

審査官 山下 剛史

- (56)参考文献 特開2005-031169(JP,A)
特開2005-091560(JP,A)
特開2001-051689(JP,A)
国際公開第2005/024788(WO,A1)
森康充他, "SIMOモデルに基づくICAとバイナリ・マスク処理を組み合わせたブラインド音源分離", 電子情報通信学会技術研究報告, Vol.104, No.497, EA2004-115(2004-12), pp.71-76
森康充他, "SIMO-ICAとバイナリマスク処理を組み合わせた2段型リアルタイムブラインド音源分離", 人工知能学会研究会資料, SIG-Challenge-0522-4(2005-10), pp.23-28

(58)調査した分野(Int.Cl., DB名)

G10L 15/20, 21/02
C i N i i