(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2005-227512
(P2005-227512A)
(43)公開日　平成17年8月25日(2005.8.25)

(51)Int.Cl.[7]	F I			テーマコード（参考）
G10L 15/20	G10L	3/02	301Z	5D015
G10L 11/02	H04R	3/00	320	5D020
G10L 15/00	G10L	3/00	511	
G10L 15/02	G10L	3/02	301E	
G10L 15/04	G10L	3/00	513B	

審査請求 未請求　請求項の数 12　OL　（全 26 頁）　最終頁に続く

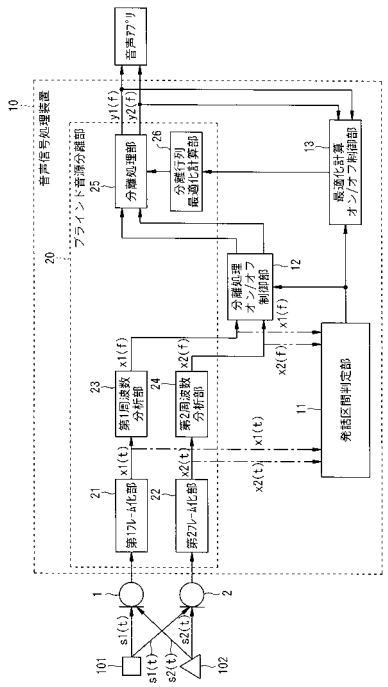
(21)出願番号　特願2004-35619(P2004-35619)
(22)出願日　平成16年2月12日(2004.2.12)

(71)出願人　000010076
ヤマハ発動機株式会社
静岡県磐田市新貝２５００番地
(74)代理人　100066980
弁理士　森　哲也
(74)代理人　100075579
弁理士　内藤　嘉昭
(74)代理人　100103850
弁理士　崔　秀▲てつ▼
(72)発明者　有宗　伸泰
静岡県磐田市新貝２５００番地　ヤマハ発動機株式会社内

Ｆターム(参考)　5D015 AA06　BB02　CC17　DD02　EE05
FF07　KK02
5D020 BB07

(54)【発明の名称】音信号処理方法及びその装置、音声認識装置並びにプログラム

(57)【要約】
【課題】ブラインド音源分離をリアルタイムで行うことを可能にする。
【解決手段】音声信号処理装置１０は、話者音源１０１からの音声と雑音源１０２からの音との混合音が入力される第１及び第２マイク１，２と、話者音源１０１から出力された発話区間を検出する発話区間判定部１１と、発話区間判定部１１が発話区間を検出した場合、第１及び第２マイク１，２に入力された音声信号ｘ１（ｔ），ｘ２（ｔ）を用いて、分離行列を最適化する分離処理オン／オフ制御部１２、最適化計算オン／オフ制御部１３及び分離行列最適化計算部２６と、分離行列最適化計算部２６が最適化した分離行列を用いて、混合音から話者音源１０１からの音と雑音源１０２からの音とを分離する分離処理部２５とを備える。
【選択図】図２



【特許請求の範囲】
【請求項１】
検出対象音源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該最適化した分離行列を用いて、前記混合音から検出対象音源からの音と雑音源からの音とを分離するブラインド音源分離（ＢＢＳ:BlindSource Separation）を行う音信号処理方法において、
前記検出対象音源からの音の検出の有無により前記分離行列の最適化の実行を切換可能にするとともに、前記検出対象音源からの音を検出した場合、前記分離行列の最適化を行い、前記検出対象音源からの音を検出できない場合、前記分離行列の最適化を行わないことを特徴とする音信号処理方法。
【請求項２】
前記検出対象音源からの音が所定長以上の音の場合、前記分離行列の最適化を行い、前記検出対象音源からの音が所定長未満の音の場合、前記分離行列の最適化を行わないことを特徴とする請求項１記載の音信号処理方法。
【請求項３】
前記ブラインド音源分離では、無指向性マイクに前記混合音が入力され、単一指向性マイクに前記検出対象音源からの音又は前記雑音源からの音のいずれか一方が入力され、前記無指向性マイク及び単一指向性マイクに入力された音の音信号を用いて前記分離行列の最適化を行い、かつ当該最適化した分離行列を用いて前記混合音から検出対象音源からの音と雑音源からの音とを分離しており、
前記無指向性マイクに入力された混合音の音信号と前記単一指向性マイクに入力された音の音信号とを比較し、その比較結果に基づいて、前記検出対象音源からの音を検出することを特徴とする請求項１又は２記載の音信号処理方法。
【請求項４】
検出対象音源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該最適化した分離行列を用いて、前記混合音から検出対象音源からの音と雑音源からの音とを分離するブラインド音源分離（ＢＢＳ:BlindSource Separation）を行う音信号処理装置において、
前記混合音が入力される第１マイクと、
前記検出対象音源からの音と雑音源からの音とのうちの少なくとも一方が入力される第２マイクと、
前記検出対象音源からの音を検出する対象音検出手段と、
前記対象音検出手段が検出対象音を検出した場合、前記第１及び第２マイクに入力された音の音信号を用いて前記分離行列を最適化する分離行列最適化手段と、
前記分離行列最適化手段が最適化した分離行列を用いて、前記第１マイクに入力された混合音から検出対象音源からの音と雑音源からの音とを分離する分離手段と、
を備えることを特徴とする音信号処理装置。
【請求項５】
前記分離行列最適化手段は、前記検出対象音検出手段が検出した検出対象音が所定長以上の音の場合、前記分離行列の最適化を行うことを特徴とする請求項４記載の音信号処理装置。
【請求項６】
前記第１マイクは、前記混合音が入力されるように配置された無指向性マイクであり、第２のマイクは、前記検出対象音源からの音と雑音源からの音とのうちのいずれか一方が入力されるように配置された単一指向性マイクであり、
前記対象音検出手段は、前記第１マイクに入力された混合音の音信号と前記第２マイクに入力された音の音信号とを比較して、その比較結果に基づいて、前記検出対象音を検出することを特徴とする請求項４又は５記載の音信号処理装置。
【請求項７】
前記第１マイクに入力された混合音の音信号及び第２マイクに入力された検出対象音源

からの音と雑音源からの音とのうちのいずれか一方の音の音信号を時分割してフレーム化するフレーム化手段を備えており、
前記対象音検出手段は、前記フレーム化手段から出力されるフレーム単位で、前記第１マイクに入力された混合音の音信号と、前記第２マイクに入力された音の音信号とを比較し、その比較結果に基づいて、前記フレーム化手段から出力されるフレーム単位で前記検出対象音を検出することを特徴とする請求項６記載の音信号処理装置。
【請求項８】
発話源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該最適化した分離行列を用いて、前記混合音から発話源からの音と雑音源からの音とを、ブラインド音源分離（ＢＢＳ:BlindSource Separation）により分離し、その分離した発話源からの音について音声認識処理を行う音声認識装置において、
前記混合音が入力される第１マイクと、
前記発話源からの音と雑音源からの音とのうちの少なくとも一方が入力される第２マイクと、
前記発話源からの音の発話区間を検出する発話区間検出手段と、
前記発話区間検出手段が発話区間を検出した場合、前記第１及び第２マイクに入力された音信号を用いて前記分離行列を最適化する分離行列最適化手段と、
前記分離行列最適化手段が最適化した分離行列を用いて、前記第１マイクに入力された混合音から発話源からの音と雑音源からの音とを分離する分離手段と、
前記分離手段が分離した発話源からの音について、音声認識処理を行う音声認識処理手段と、
を備えることを特徴とする音声認識装置。
【請求項９】
前記分離行列最適化手段は、前記発話区間検出手段が検出した発話区間が所定長以上の場合、前記分離行列の最適化を行うことを特徴とする請求項８記載の音声認識装置。
【請求項１０】
前記第１マイクは、前記混合音が入力されるように配置された無指向性マイクであり、第２のマイクは、前記発話源からの音と雑音源からの音とのうちのいずれか一方が入力されるように配置された単一指向性マイクであり、
前記発話区間検出手段は、前記第１マイクに入力された混合音の音信号と前記第２マイクに入力された音の音信号とを比較して、その比較結果に基づいて、前記発話区間を検出することを特徴とする請求項８又は９記載の音声認識装置。
【請求項１１】
前記第１に入力された混合音の音信号及び第２マイクに入力された発話源からの音と雑音源からの音とのうちのいずれか一方の音の音信号を時分割してフレーム化するフレーム化手段を備えており、
前記発話区間検出手段は、前記フレーム化手段から出力されるフレーム単位で、前記第１マイクに入力された混合音の音信号と、前記第２マイクに入力された音の音信号とを比較し、その比較結果に基づいて、前記フレーム化手段から出力されるフレーム単位で前記発話区間を検出することを特徴とする請求項１０記載の音声認識装置。
【請求項１２】
検出対象音源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該分離行列を用いて、前記混合音から検出対象音源からの音と雑音源からの音とを分離するブラインド音源分離（ＢＢＳ:BlindSource Separation）をコンピュータに実現させるプログラムにおいて、
前記検出対象音源からの音の検出の有無により前記分離行列の最適化の実行を切換可能にするとともに、前記検出対象音源からの音を検出した場合、前記分離行列の最適化を行い、前記検出対象音源からの音を検出できない場合、前記分離行列の最適化を行わないようにコンピュータに実行させることを特徴とするプログラム。
【発明の詳細な説明】

【技術分野】
【０００１】
本発明は、音信号処理方法、音信号処理装置、音声認識装置及びプログラムに関し、特に混合音から検出対象音を分離して取り出すブラインド音源分離（ＢＢＳ:BlindSource Separation）が適合される音信号処理方法、音信号処理装置、音声認識装置及びプログラムに関する。
【背景技術】
【０００２】
ブラインド音源分離（ＢＢＳ:Blind Source Separation）では、複数チャンネルに入力された混合音を用いて、独立成分分析（ＩＣＡ:IndependentComponent Analysis）の技術により、分離行列を最適化（学習）する。これにより、分離行列が目的とする音を分離する最適解に近づく。そして、ブラインド音源分離では、そのように最適化した分離行列を用いて、混合音から目的の音を分離して取り出している。ここで、混合音として、話者音源（発話源）からの音（発話）と雑音源からの音が混ざり合った音が挙げられ、このような場合、分離目的の音は、話者音源からの音（発話）になる。
【発明の開示】
【発明が解決しようとする課題】
【０００３】
リアルタイムでブラインド音源分離をする場合、混合音で分離行列を最適化しつつ、混合音から目的の音を分離するような態様となる。このようにリアルタイムでブラインド音源分離を実現する場合には、混合音中に分離目的の音が断続的又は不規則に含まれるようになる。
しかし、従来のブラインド音源分離のシステムは、オフライン処理によりブラインド音源分離をすることを前提としている。すなわち、従来のブラインド音源分離のシステムは、分離目的の音が連続して入力される場合を前提とし、その前提の下、分離行列を最適化しつつ、その最適化した分離行列で目的の音を分離するように構成されている。このようなことから、従来のシステムでリアルタイムでブラインド音源分離を行うと、混合音に分離目的の音が断続的又は不規則に含まれる結果、分離目的の音の特定が困難になることから、分離行列を最適化できなくなる。この結果、目的の音を高精度で分離できなくなる。
【０００４】
このように、従来のシステムは、オフライン処理でブラインド音源分離を行う必要があった。このような結果、従来のシステムは、実用性に欠けたものとなっていた。
さらに、従来のシステムでは、分離行列の最適化処理中に分離目的としない他の音が長時間継続して入力されてしまうと、当該他の音で分離行列を最適化してしまう。この場合、分離行列が間違った局所最適解に落ち込んでしまう。このように分離行列が間違った局所最適解に落ち込んでしまうと、その後、分離目的の音を入力しても、分離行列が最適化しなくなってしまう。
本発明は、前記問題に鑑みてなされたものであり、ブラインド音源分離をリアルタイムで行うことを可能にする音信号処理方法、音信号処理装置、音声認識装置及びプログラムの提供を目的とする。
【課題を解決するための手段】
【０００５】
請求項１記載の音信号処理方法は、検出対象音源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該最適化した分離行列を用いて、前記混合音から検出対象音源からの音と雑音源からの音とを分離するブラインド音源分離（ＢＢＳ:BlindSource Separation）を行う音信号処理方法である。この音信号処理方法は、前記検出対象音源からの音の検出の有無により前記分離行列の最適化の実行を切換可能にするとともに、前記検出対象音源からの音を検出した場合、前記分離行列の最適化を行い、前記検出対象音源からの音を検出できない場合、前記分離行列の最適化を行わないことを特徴とする。なお、検出対象音源からの音には、人間が発する発話音の他、物体が発する音も含まれ

る。

【０００６】

また、請求項２記載の音信号処理方法は、請求項１記載の音信号処理方法において、前記検出対象音源からの音が所定長以上の音の場合、前記分離行列の最適化を行い、前記検出対象音源からの音が所定長未満の音の場合、前記分離行列の最適化を行わないことを特徴とする。

また、請求項３記載の音信号処理方法は、請求項１又は２記載の音信号処理方法において、前記ブラインド音源分離では、無指向性マイクに前記混合音が入力され、単一指向性マイクに前記検出対象音源からの音又は前記雑音源からの音のいずれか一方が入力され、前記無指向性マイク及び単一指向性マイクに入力された音の音信号を用いて前記分離行列の最適化を行い、かつ当該最適化した分離行列を用いて前記混合音から検出対象音源からの音と雑音源からの音とを分離しており、前記無指向性マイクに入力された混合音の音信号と前記単一指向性マイクに入力された音の音信号とを比較し、その比較結果に基づいて、前記検出対象音源からの音を検出することを特徴とする。

【０００７】

また、請求項４記載の音信号処理装置は、検出対象音源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該最適化した分離行列を用いて、前記混合音から検出対象音源からの音と雑音源からの音とを分離するブラインド音源分離（ＢＢＳ:BlindSource Separation）を行う音信号処理装置である。この音信号処理装置は、前記混合音が入力される第１マイクと、前記検出対象音源からの音と雑音源からの音とのうちの少なくとも一方が入力される第２マイクと、前記検出対象音源からの音を検出する対象音検出手段と、前記対象音検出手段が検出対象音を検出した場合、前記第１及び第２マイクに入力された音の音信号を用いて前記分離行列を最適化する分離行列最適化手段と、前記分離行列最適化手段が最適化した分離行列を用いて、前記第１マイクに入力された混合音から検出対象音源からの音と雑音源からの音とを分離する分離手段と、を備えることを特徴とする。

【０００８】

また、請求項５記載の音信号処理装置は、請求項４記載の音信号処理装置において、前記分離行列最適化手段が、前記検出対象音検出手段が検出した検出対象音が所定長以上の音の場合、前記分離行列の最適化を行うことを特徴とする。

また、請求項６記載の音信号処理装置は、請求項４又は５記載の音信号処理装置において、前記第１マイクが、前記混合音が入力されるように配置された無指向性マイクであり、第２のマイクが、前記検出対象音源からの音と雑音源からの音とのうちのいずれか一方が入力されるように配置された単一指向性マイクであり、前記対象音検出手段が、前記第１マイクに入力された混合音の音信号と前記第２マイクに入力された音の音信号とを比較して、その比較結果に基づいて、前記検出対象音を検出することを特徴とする。

【０００９】

また、請求項７記載の音信号処理装置は、請求項６記載の音信号処理装置において、前記第１マイクに入力された混合音の音信号及び第２マイクに入力された検出対象音源からの音と雑音源からの音とのうちのいずれか一方の音の音信号を時分割してフレーム化するフレーム化手段を備えており、前記対象音検出手段が、前記フレーム化手段から出力されるフレーム単位で、前記第１マイクに入力された混合音の音信号と、前記第２マイクに入力された音の音信号とを比較し、その比較結果に基づいて、前記フレーム化手段から出力されるフレーム単位で前記検出対象音を検出することを特徴とする。

【００１０】

また、請求項８記載の音声認識装置は、発話源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該最適化した分離行列を用いて、前記混合音から発話源からの音と雑音源からの音とを、ブラインド音源分離（ＢＢＳ:BlindSource Separation）により分離し、その分離した発話源からの音について音声認識処理を行う音声認識装置である。この音声認識装置は、前記混合音が入力される第１マイクと、前記発話源から

の音と雑音源からの音とのうちの少なくとも一方が入力される第２マイクと、前記発話源からの音の発話区間を検出する発話区間検出手段と、前記発話区間検出手段が発話区間を検出した場合、前記第１及び第２マイクに入力された音信号を用いて前記分離行列を最適化する分離行列最適化手段と、前記分離行列最適化手段が最適化した分離行列を用いて、前記第１マイクに入力された混合音から発話源からの音と雑音源からの音とを分離する分離手段と、前記分離手段が分離した発話源からの音について、音声認識処理を行う音声認識処理手段と、を備える。

【００１１】

また、請求項９記載の音声認識装置は、請求項８記載の音声認識装置において、前記分離行列最適化手段が、前記発話区間検出手段が検出した発話区間が所定長以上の場合、前記分離行列の最適化を行うことを特徴とする。

また、請求項１０記載の音声認識装置は、請求項８又は９記載の音声認識装置において、前記第１マイクが、前記混合音が入力されるように配置された無指向性マイクであり、第２のマイクが、前記発話源からの音と雑音源からの音とのうちのいずれか一方が入力されるように配置された単一指向性マイクであり、前記発話区間検出手段が、前記第１マイクに入力された混合音の音信号と前記第２マイクに入力された音の音信号とを比較して、その比較結果に基づいて、前記発話区間を検出することを特徴とする。

【００１２】

また、請求項１１記載の音声認識装置は、請求項１０記載の音声認識装置において、前記第１に入力された混合音の音信号及び第２マイクに入力された発話源からの音と雑音源からの音とのうちのいずれか一方の音の音信号を時分割してフレーム化するフレーム化手段を備えており、前記発話区間検出手段が、前記フレーム化手段から出力されるフレーム単位で、前記第１マイクに入力された混合音の音信号と、前記第２マイクに入力された音の音信号とを比較し、その比較結果に基づいて、前記フレーム化手段から出力されるフレーム単位で前記発話区間を検出することを特徴とする。

【００１３】

また、請求項１２記載のプログラムは、検出対象音源からの音と雑音源からの音との混合音により分離行列を最適化していき、当該分離行列を用いて、前記混合音から検出対象音源からの音と雑音源からの音とを分離するブラインド音源分離（ＢＢＳ:BlindSource Seperation）をコンピュータに実現させるプログラムである。このプログラムは、前記検出対象音源からの音の検出の有無により前記分離行列の最適化の実行を切換可能にするとともに、前記検出対象音源からの音を検出した場合、前記分離行列の最適化を行い、前記検出対象音源からの音を検出できない場合、前記分離行列の最適化を行わないようにコンピュータに実行させることを特徴とする。

【発明の効果】

【００１４】

本発明によれば、発話源からの音を検出した場合、分離行列の最適化を行い、前記発話源からの音を検出できない場合、分離行列の最適化を行わないので、断続的又は不規則にシステムに入力される発話源からの音に対してのみ分離行列の最適化を行うことができる。これにより、リアルタイムでブラインド音源分離を行うことができる。

また、請求項２、５及び９記載の発明によれば、検出対象音源からの音又は発話源からの音が所定長以上の場合、分離行列の最適化を行うようにすることで、検出対象音源からの音又は発話源からの音に対して最適解の分離行列を得ることができる。

【００１５】

また、請求項３、６及び１０記載の発明によれば、無指向性マイクで検出対象音源からの音又は発話音及び雑音を受音し、単一指向性マイクで前記検出対象音源からの音（発話音）又は前記雑音のいずれか一方を受音するように、無指向性マイク及び単一指向性マイクを配置する限り、前記検出対象音源からの音（発話源からの音）を検出することができる。これにより、マイクの取り付け位置の変化等による環境の変化、話者の移動や姿勢の変化等による音源の移動に対してロバストな受音系の構築が可能になる。

【発明を実施するための最良の形態】
【００１６】
本発明を実施するための最良の形態（以下、実施形態という。）を図面を参照しながら詳細に説明する。
第１の実施形態は、図１に示すように、第１及び第２マイク１，２に入力された音声信号を処理する音声信号処理装置１０である。
図２は音声信号処理装置１０の構成を示す。
図２に示すように、音声信号処理装置１０は、第１及び第２フレーム化部２１，２２、第１及び第２周波数分析部２３，２４、分離処理部２５、分離行列最適化計算部２６、発話区間判定部１１、分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３を備えている。
【００１７】
なお、第１及び第２フレーム化部２１，２２、第１及び第２周波数分析部２３，２４、分離処理部２５及び分離行列最適化計算部２６は、ブラインド音源分離（ＢＢＳ:BlindSource Separation）を実現するブラインド音源分離部２０を構成している。すなわち、ブラインド音源分離部２０は、このような構成を備えることで、複数チャンネルに入力された混合音により、独立成分分析（ＩＣＡ:IndependentComponent Analysis）の技術を用いて分離行列を最適化する一方、当該最適化した分離行列を用いて、前記混合音から話者音源（発話源）からの音（発話音）と雑音源からの音（雑音）とを分離する音声信号処理を実現する。
【００１８】
このような音声信号処理装置１０の構成において、第１及び第２マイク１，２から入力された２ｃｈの音声信号ｘ１（ｔ），ｘ２（ｔ）はそれぞれ、第１及び第２フレーム化部２１，２２に入力される。
ここで、音声信号ｘ１（ｔ），ｘ２（ｔ）は、話者音源（発話源）１０１が発した音ｓ１（ｔ）と雑音源１０２が発した音ｓ２（ｔ）とが混ざり合った混合音信号である。雑音ｓ２（ｔ）としては、話者音源の周囲の音、話者音源以外の他の者の音声等が挙げられる。
【００１９】
第１フレーム化部２１では、第１マイク１から入力された音声信号ｘ１（ｔ）を時分割でフレーム化（或いはフレーム分割）して、複数フレームにした音声信号ｘ１（ｔ）を第１周波数分析部２３に出力する。第２フレーム化部２２では、第２マイク２から入力される音声信号ｘ２（ｔ）を時分割でフレーム化（或いはフレーム分割）して、複数フレームにした音声信号ｘ２（ｔ）を第２周波数分析部２４に出力する。ここでは、第１及び第２フレーム化部２１，２２は、入力されてくる音声信号ｘ１（ｔ），ｘ２（ｔ）を所定時間間隔でサンプリングしていき、所定のサンプル数を１フレームとして次々にフレーム化していく。
【００２０】
第１及び第２周波数分析部２３，２４はそれぞれ、フレーム単位で音声信号ｘ１（ｔ），ｘ２（ｔ）をＦＦＴ（Fast Fourier Transform）により周波数分析して、観測信号（Observedsignals）ｘ１（ｆ），ｘ２（ｆ）を生成し、その観測信号ｘ１（ｆ），ｘ２（ｆ）を分離処理オン／オフ制御部１２に出力する。
なお、観測信号ｘ１（ｆ），ｘ２とは、当該ブラインド音源分離（ＢＢＳ:Blind Source Separation）の技術において、混合音の分離を行う分離行列に入力される信号のことをいう。
【００２１】
分離処理オン／オフ制御部１２は、発話区間判定部１１からの発話区間判定結果（制御信号）に基づいて、第１及び第２周波数分析部２３，２４それぞれからの観測信号ｘ１（ｆ），ｘ２（ｆ）を後段の分離処理部２５に出力する。
発話区間判定部１１は、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ）

，ｘ２（ｔ）に基づいて、当該音声信号ｘ１（ｔ），ｘ２（ｔ）に含まれている発話音声の区間（発話区間）を判定するように構成されている。例えば、発話区間判定部１１は、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）、具体的には第１及び第２フレーム化部１１，１２から出力されたフレーム単位の音声信号ｘ１（ｔ），ｘ２（ｔ）又は第１及び第２周波数分析部２３，２４から出力された信号ｘ１（ｆ），ｘ２（ｆ）に基づいて、当該フレーム単位で発話区間の判定を行う。具体的には、発話区間判定部１１は、所定長（所定時間）以上の発話区間を検出したときに、発話区間を検出した旨の信号を判定結果（制御信号）として、分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。なお、発話区間判定部１１の具体的な構造については、後述する第２乃至第４の実施形態として説明する。

【００２２】

これにより、分離処理オン／オフ制御部１２は、発話区間判定部１１から発話区間を検出した結果が入力された場合、分離処理部２５のオン制御として、第１及び第２周波数分析部２３，２４それぞれからの観測信号ｘ１（ｆ），ｘ２（ｆ）を分離処理部２５に出力する。また、分離処理オン／オフ制御部１２は、発話区間判定部１１が発話区間を検出していない場合、分離処理部２５のオフ制御として、第１及び第２周波数分析部２３，２４それぞれからの観測信号ｘ１（ｆ），ｘ２（ｆ）を分離処理部２５に出力しない。このとき、分離処理オン／オフ制御部１２から分離処理部２５への観測信号ｘ１（ｆ），ｘ２（ｆ）の出力のオン及びオフは、発話区間判定部１１が発話区間を検出したフレームに対応するフレームを単位として行う。

【００２３】

分離処理部２５は、分離行列最適化計算部２６により最適化された分離行列により、観測信号ｘ１（ｆ），ｘ２（ｆ）から分離信号ｙ１（ｆ），ｙ２（ｆ）を分離抽出する。そして、分離処理部２５は、音声信号ｓ１（ｔ），ｓ２（ｔ）とされる分離信号ｙ１（ｆ），ｙ２（ｆ）を後段に出力する。

分離行列最適化計算部２６は、分離処理部２５が得た分離信号ｙ１（ｆ），ｙ２（ｆ）が入力されており、この分離信号ｙ１（ｆ），ｙ２（ｆ）に基づく分離行列の最適化処理として、最適解の分離行列を得る。そして、分離行列最適化計算部２６は、その最適化した分離行列を分離処理部２５に出力する。すなわち、分離処理部２５は、当該分離処理部２５が得る分離信号ｙ１（ｆ），ｙ２（ｆ）を用いて分離行列最適化計算部２６で最適化された分離行列を用いて、それ以降に当該分離処理部２５に入力される観測信号ｘ１（ｆ），ｘ２（ｆ）から分離信号ｙ１（ｆ），ｙ２（ｆ）を分離抽出しているのである。

【００２４】

一方、分離行列最適化計算部２６は、最適化計算オン／オフ制御部１３によりオン及びオフ制御がなされる。具体的には、最適化計算オン／オフ制御部１３は、発話区間判定部１１から発話区間を検出した結果が入力された場合、分離行列最適化計算部２６をオン制御しており、分離行列最適化計算部２６はこのオン制御により、分離処理部２５が出力した分離信号ｙ１（ｆ），ｙ２（ｆ）に基づいて、分離行列の最適化処理を実施する。また、最適化計算オン／オフ制御部１３は、発話区間判定部１１が発話区間を検出していない場合、分離行列最適化計算部２６をオフ制御しており、分離行列最適化計算部２６はこのオフ制御により、分離行列の最適化処理を停止（一時的に停止）する。

【００２５】

以上のように音声信号処理装置１０が構成されている。

次に図３を用いて、第１及び第２マイク１，２から入力された２ｃｈの音声信号（混合音信号）ｘ１（ｔ），ｘ２（ｔ）に対する処理に沿って、音声信号処理装置１０の一連の動作を説明する。なお、ここでの動作は、分離行列を最適化（学習）する際の動作になる。

第１及び第２マイク１，２からの音声信号ｘ１（ｔ），ｘ２（ｔ）は、第１及び第２フレーム化部２１，２２に入力される。

第１及び第２フレーム化部２１，２２は、各音声信号ｘ１（ｔ），ｘ２（ｔ）をフレー

ム化（或いはフレーム分割）して、複数フレームにした音声信号ｘ１（ｔ），ｘ２（ｔ）を第１及び第２周波数分析部２３，２４に出力する（ステップＳ１）。

【００２６】

第１及び第２周波数分析部２３，２４では、フレーム単位で、音声信号ｘ１（ｔ），ｘ２（ｔ）から観測信号ｘ１（ｆ），ｘ２（ｆ）を生成し、その観測信号ｘ１（ｆ），ｘ２（ｆ）を分離処理オン／オフ制御部１２に出力する（ステップＳ２）。

一方、発話区間判定部１１は、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）中の発話区間の判定をフレーム単位で行い（ステップＳ３）、発話区間（発話フレーム）を検出する（ステップＳ４）。そして、発話区間判定部１１は、発話区間を検出した場合、当該発話区間が最短発話長以上か否かを判定する（ステップＳ５）。ここで、発話区間判定部１１は、発話区間が最短発話長以上の場合、発話区間を検出した旨の判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。また、発話区間判定部１１は、発話区間を検出できなかった場合、又は発話区間は検出できたが、その発話区間が最短発話長未満であった場合、発話区間を検出できなかったとして、その旨の判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。

【００２７】

分離行列最適化計算部２６は、分離処理部２５から分離行列を読み出す（ステップＳ６）。そして、分離行列最適化計算部２６は、その読み出した分離行列の最適化計算を行う（ステップＳ７）。具体的には次のような処理により分離行列の最適化計算を行う。

分離処理オン／オフ制御部１２では、発話区間判定部１１が発話区間を検出した場合、第１及び第２周波数分析部２３，２４それぞれからの観測信号ｘ１（ｆ），ｘ２（ｆ）を後段の分離処理部２５に出力する。そして、分離処理部２５は、最新の分離行列により観測信号ｘ１（ｆ），ｘ２（ｆ）から分離信号ｙ１（ｆ），ｙ２（ｆ）を得る。

【００２８】

その一方で、最適化計算オン／オフ制御部１３は、発話区間判定部１１が発話区間を検出した場合、分離行列最適化計算部２６をオン制御する。分離行列最適化計算部２６は、オン制御により、分離処理部２５が得た分離信号ｙ１（ｆ），ｙ２（ｆ）を取り込み、この分離信号ｙ１（ｆ），ｙ２（ｆ）に基づいて前記読み出した分離行列を最適化する。

このように分離行列最適化計算部２６で分離行列の最適化計算を行う。そして、分離行列最適化計算部２６は、その最適化した分離行列を分離処理部２５に出力し、分離処理２５は、その分離行列を保存する（ステップＳ８）。

【００２９】

そして、分離処理部２５は、このように最適化された最新の分離行列を用いて、観測信号ｘ１（ｆ），ｘ２（ｆ）から分離信号ｙ１（ｆ），ｙ２（ｆ）を得る（ステップＳ９）。

このように音声信号処理装置１０は、分離処理部２５で得た分離信号ｙ１（ｆ），ｙ２（ｆ）を例えば音声アプリケーションに出力する。

音声アプリケーションは、例えば音声を認識して各種処理を行うアプリケーションである。例えば、音声アプリケーションとしては、音声認識システム、放送システム、携帯電話及びトランシーバが挙げられる。このような音声アプリケーションは、話者音源（発話源）１０１が発した音声信号ｓ１（ｔ）である分離信号ｙ１（ｆ）に基づいて、音声を認識して、所定の処理を行う。

【００３０】

次に第１の実施形態における効果を説明する。

前述したように、音声信号処理装置１０は、発話区間を検出した場合にのみ、分離行列の最適化計算を行っている。これにより、分離目的の音である発話音源からの音が音声信号処理装置１０に断続的又は不規則に入力されてくる場合でも、音声信号処理装置１０は、分離行列を最適化することができる。これにより、分離目的の音である発話音源からの音が音声信号処理装置１０に断続的又は不規則に入力されてくる場合でも、音声信号処理

装置１０は、目的の音である発話音源からの音を高精度で分離できるようになる。このように、音声信号処理装置１０は、リアルタイムでブラインド音源分離を実現できるようになり、実用性に優れたものとなる。

【００３１】

また、このように発話区間を検出した場合にのみ分離行列の最適化計算を行うようにすることで、分離目的外の音が入力されても分離行列の最適化計算が行われないので、そのような分離目的外の音により分離行列が間違った局所最適解に落ち込んでしまうようなことを防止できる。

また、前述したように、音声信号処理装置１０は、発話区間が最短発話長以上の場合に限って、分離行列の最適化計算を行っている。一般的には、ブラインド音源分離のシステムに入力される音（学習対象の音）がある一定以上の長さがあると、分離行列の最適化は良好となる。このようなことから、音声信号処理装置１０は、発話区間が最短発話長以上の場合に限って分離行列の最適化計算を行うようにすることで、分離目的の音に最適解の分離行列を得ることができるようになる。なお、音声信号処理装置１０が分離した音声を音声認識システム（音声アプリケーション）が利用するとした場合、前記一定以上の長さとは、例えばコマンド最短長さや、１発話最短長さとなる。

【００３２】

そして、このように音声信号処理装置１０では、高精度で目的の音声を分離できるので、このように音声信号処理装置１０が分離した音声を利用することにより、音声認識システムでは、高認識率、低誤認識率の音声認識が可能になり、また、携帯電話やトランシーバでは、信頼性の高いハンズフリー半二重通信が可能になり、放送システムでは、通信システムの送信電力低減が可能になる。

【００３３】

また、前述したように、発話区間判定部１１が発話区間を検出した場合には、分離処理オン／オフ制御部１２が第１及び第２周波数分析部２３，２４それぞれからの観測信号ｘ１（ｆ），ｘ２（ｆ）を後段の分離処理部２５に出力する一方で、最適化計算オン／オフ制御部１３が分離行列最適化計算部２６をオン制御して、分離行列最適化計算部２６に分離行列の最適化処理を実施させている。

【００３４】

よって、発話区間判定部１１が発話区間を検出した場合にのみ、分離行列の最適化計算を行うのであれば、分離処理オン／オフ制御部１２と最適化計算オン／オフ制御部１３とのいずれか一方を備えるだけでよいといえる。しかし、分離処理オン／オフ制御部１２や最適化計算オン／オフ制御部１３の応答性を考慮して、これら両方をシステムに備えることで、それら構成要素の特性に対するロバスト性を上げて、分離行列の最適化処理を行うことができるようになる。

【００３５】

また、前述したように、音声信号ｘ１（ｔ），ｘ２（ｔ）を第１及びイ第２フレーム化部２１，２２でフレーム化したものを、分離処理オン／オフ制御部１２及び分離処理部２５に出力している。このようにすることで、結果的に、音声信号処理装置１０から出力される分離信号ｙ１（ｆ）である音声信号ｓ１（ｔ）もフレーム化されているものとなり、これにより、音声信号処理装置１０から出力される音声信号ｓ１（ｔ）を利用する音声アプリケーションでは、解りやすいフレーム化された音声信号ｓ１（ｔ）で処理をすることができるようになる。

【００３６】

ここで、図４を用いて効果を説明する。

図４中（Ａ）は、オフラインによりブラインド音源分離を行う場合を示し（従来の手法）、図４中（Ｂ）及び（Ｃ）は、リアルタイムでブラインド音源分離を行う場合を示す。

従来の手法をそのまま適用して、リアルタイムでブラインド音源分離をしてしまうと、図４中（Ｂ）に従来法として示すように、システムに雑音のみが入力されている場合でも、その雑音により分離行列を最適化してしまう。この場合、分離行列が劣化してしまう。

この結果、最適化された分離行列では、目的とする信号（音声信号）を分離することができなくなる（結果不明となる）。

【００３７】

一方、本発明を適用して、リアルタイムでブラインド音源分離をした場合、図４中（Ｂ）に本発明法として示すように、システムに雑音のみが入力されているときには、分離行列の最適化は実施されず、システムに雑音と目的とする信号（音声信号）とが入力されたときに、分離行列の最適化は実施される。この結果、最適化された分離行列により、雑音とともに入力されてきた目的とする信号（音声信号）を精度よく分離することができる。

【００３８】

また、従来の手法をそのまま適用して、リアルタイムでブラインド音源分離をした場合、システムに雑音のみ又は雑音と分離目的外の信号とが混じり合い、長時間入力されると、図４中（Ｃ）に従来法として示すように、分離行列が間違った局所最適解に落ち込んでしまう。この結果、目的とする信号（音声信号）を分離することができなくなる（結果不明となる）。

しかし、本発明を適用した場合には、図４中（Ｃ）に本発明法として示すように、システムに雑音と目的とする信号（音声信号）とが入力されたときに分離行列の最適化を実施するので、そのように分離行列が間違った局所最適解に落ち込んでしまうことを防止できる。

【００３９】

次に第２の実施形態を説明する。

この第２の実施形態は、発話区間判定部１１を具体的な構成とした音声信号処理装置１０であり、発話区間判定部１１が、第１及び第２マイク１，２で受音した音声信号ｘ１（ｔ），ｘ２（ｔ）の相関度により発話区間を検出するように構成されている。

図５は、その第２の実施形態における発話区間判定部１１の構成を示し、図６は、発話区間判定部１１の構成に対応する第１及び第２マイクの配置を示す。

【００４０】

この第２の実施形態では、第１マイク１として単一指向性マイクを使用し、第２マイク２として無指向性マイクを使用している。そして、第１及び第２マイク１，２は、図６に示すように、第１及び第２マイク１，２をできるだけ近づけて配置するとともに、単一指向性マイクである第１マイク１をその指向方向が発話音源（ユーザ）の位置に対して反対側となるように配置する。また、第１マイク１の指向方向に、雑音源が存在している。なお、図６に示す点線は、雑音源を基準にした第１マイク１の指向特性を示し、図６に示す一点鎖線は、第２マイク２の指向特性を示す。

【００４１】

このように第１及び第２マイク１，２を配置すると、雑音源からの音ｓ２（ｔ）は、第１及び第２マイク１，２で受音でき、発話音源（ユーザ）からの音ｓ１（ｔ）は第２マイク２だけが受音できるようになる。

このように配置した第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）はそれぞれ、前述したように、第１及び第２フレーム化部２１，２２に入力される。そして、前述したように、第１フレーム化部２１では、第１マイク１から入力された音声信号ｘ１（ｔ）をフレーム化（或いはフレーム分割）し、また、第２フレームか部２２では、第２マイク２から入力される音声信号ｘ２（ｔ）をフレーム化（或いはフレーム分割）する。そして、このように各フレーム化部２１，２２で複数フレームにされた音声信号ｘ１（ｔ），ｘ２（ｔ）は発話区間判定部１１に入力される。

【００４２】

発話区間判定部１１は、図５に示すように、相互相関関数計算部３１及び音声／非音声判定部４１を備えている。このような発話区間判定部１１において、各フレーム化部２１，２２で複数フレームにされた音声信号ｘ１（ｔ），ｘ２（ｔ）が相互相関関数計算部３１に入力される。

相互相関関数計算部３１は、第１フレーム化部２１から出力されるフレームと、第２フ

レーム化部１２から出力されるフレームとを比較する。すなわち、第１マイク１に入力された音声信号ｘ１（ｔ）と、第２マイク２に入力された音声信号ｘ２（ｔ）とをフレーム単位で比較する。その比較結果として、相互相関関数計算部３１は、下記（１）式により、相互相関関数Ｒ（τ）を算出する。

【００４３】

【数１】

$$R(\tau)=\frac{1}{T}\int_{-T/2}^{T/2} x1(t)x2(t-\tau)dt \quad \cdots (1)$$

【００４４】

ここで、τは第１マイク１と第２マイク２との間の距離によって決まる遅延時間である。また、Ｔはフレーム長である。

前述したように第１及び第２マイク１，２をできるだけ近づけて配置している場合には、遅延時間τを近似的に０とおくことができる。しかし、後述するような本発明の効果を満たす限り、第１マイク１と第２マイク２とを離して配置することは可能であり、この場合、遅延時間τを適切に与える必要がある。すなわち例えば、第１マイク１と第２マイク２との間の距離を１０ｃｍにしている場合には、その１０ｃｍ相当分の遅延時間τを与えて、相互相関関数Ｒ（τ）を算出する。このようにすれば、第１マイク１と第２マイク２との間の距離を考慮して、相互相関関数Ｒ（τ）を得ることができ、精度よく相互相関関数Ｒ（τ）を得ることができる。

【００４５】

このように算出された相互相関関数Ｒ（τ）は、相関関係を求める２つの音声信号ｘ１（ｔ），ｘ２（ｔ）が似ているほど、大きい値となり、相関関係を求める２つの音声信号ｘ１（ｔ），ｘ２（ｔ）が異なっているほど、０に近くなる。相互相関関数計算部３１は、このような相互相関関数Ｒ（τ）を音声／非音声判定部３２に出力する。

音声／非音声判定部３２は、相互相関関数Ｒ（τ）に基づいて音声区間（発話区間）と非音声区間（非発話区間）とを判定する。具体的には、次のように音声区間と非音声区間とを判定する。

【００４６】

前述したように、発話音源（ユーザ）と雑音源に対して図６のように第１及び第２マイク１，２を配置することで、雑音源からの音ｓ２（ｔ）を第１及び第２マイク１，２で受音し、発話音源（ユーザ）からの音ｓ１（ｔ）を第２マイク２だけで受音している。

一方、相互相関関数Ｒ（τ）は、前述したように、相関関係を求める２つの音声信号ｘ１（ｔ），ｘ２（ｔ）が似ているほど大きい値となり、相関関係を求める２つの音声信号ｘ１（ｔ），ｘ２（ｔ）が異なっているほど０に近くなる。

【００４７】

このようなことから、雑音源からの音ｓ２（ｔ）だけを第１及び第２マイク１，２で受音している場合には、同じ音声信号が第１及び第２マイク１，２に入力されているので、すなわち、第１及び第２マイク１，２の入力音声信号のＳ／Ｎ比が同程度になるので、相互相関関数Ｒ（τ）は大きい値になる。一方、発話音源（ユーザ）から発話があった場合には、その発話を第２マイク２だけが受音するので、第１及び第２マイク１，２それぞれに異なる音声信号が入力されるようになり、すなわち第２マイク２の入力音声信号のＳ／Ｎ比の方が大きくなるので、相互相関関数Ｒ（τ）は０に向かって減少する。

【００４８】

このように、発話音源（ユーザ）から発話があった場合には相互相関関数Ｒ（τ）は０に向かって減少することから、音声／非音声判定部３２は、相互相関関数Ｒ（τ）と判定用しきい値（類似度を示すしきい値）δｒ１とを比較して、音声区間を判定する。すなわち、音声／非音声判定部３２は、相互相関関数Ｒ（τ）が判定用しきい値δｒ１未満の場合（Ｒ（τ）＜δｒ１）、音声区間と判定し、それ以外の場合（Ｒ（τ）≧δｒ１）、非

音声区間と判定する。ここで、判定用しきい値δｒ１は例えば実験により得る。そして、音声／非音声判定部３２は、このような判定をフレーム単位で行う。発話区間判定部１１は、このように音声／非音声判定部３２で得た音声区間（発話区間）の判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。

【００４９】

以上のように、発話区間判定部１１では、相互相関関数計算部１３が、第１及び第２フレーム化部２１，２２それぞれから出力されるフレーム単位で相互相関関数Ｒ（τ）を算出して、算出した相互相関関数Ｒ（τ）を音声／非音声判定部３２に出力する。音声／非音声判定部３２では、相互相関関数Ｒ（τ）と判定用しきい値δｒ１とを比較し、相互相関関数Ｒ（τ）に対応するフレームが音声区間のものか、非音声区間のものかを判定する。そして、音声／非音声判定部３２は、その判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。

【００５０】

そして、分離処理オン／オフ制御部１２は、前述したように、発話区間判定部１１からの発話区間の判定結果に基づいて、分離処理部２５への観測信号ｘ１（ｆ），ｘ２（ｆ）の出力をオン及びオフ制御する。また、最適化計算オン／オフ制御部１３は、前述したように、発話区間判定部１１からの発話区間の判定結果に基づいて、分離行列最適化計算部２６のオン及びオフを制御する。

【００５１】

なお、第１及び第２マイク１，２の配置については、前記図６に示した態様に限定されるものではない。例えば、発話音源（ユーザ）からの音を第１及び第２マイク１，２で受音し、雑音源からの音を第１マイク１だけで受音するように、第１及び第２マイク１，２を配置してもよい。具体的には、第１マイク１に無指向性マイクを用い、第２マイク２に単一指向性マイクを用いる。そして、図７に示すように、第１及び第２マイク１，２をできるだけ近づけて配置するとともに、単一指向性マイクである第２マイク２を、その指向方向が発話音源（ユーザ）に向かい、かつその指向方向外に雑音源が位置されるように、配置する。なお、図７に示す点線は、第１マイク１の指向特定を示し、図７に示す一点鎖線は、発話音源（ユーザ）を基準にした第２マイク２の指向特性を示す。

【００５２】

そして、このように第１及び第２マイク１，２を配置した場合には、相互相関関数計算部３１及び音声／非音声判定部３２は次のような計算を行う。

発話音源（ユーザ）からの音ｓ１（ｔ）を第１及び第２マイク１，２で受音し、雑音源からの音ｓ２（ｔ）を第１マイク１だけが受音しているので、雑音源からの音ｓ２（ｔ）だけを第１マイク１で受音している場合には、第１及び第２マイク１，２それぞれに異なる音声信号が入力されるようになり、相互相関関数Ｒ（τ）は０に近い値になる。一方、発話音源（ユーザ）から発話があった場合には、その発話を第１及び第２マイク１，２で受音するので、ほぼ同じ音声信号が第１及び第２マイク１，２に入力される。このとき、相互相関関数Ｒ（τ）は大きい値になる。そして、このとき第２マイク２の入力音声信号のＳ／Ｎ比は高くなり、第１マイク１の入力音声信号のＳ／Ｎ比は、第２マイク２ほどではないが、高くなる。

【００５３】

このように、相互相関関数計算部３１は、発話音源（ユーザ）から発話があった場合には、大きい相互相関関数Ｒ（τ）を得る。

このようなことから、音声／非音声判定部３２は、相互相関関数Ｒ（τ）と判定用しきい値（類似度を示すしきい値）δｒ２とを比較して、相互相関関数Ｒ（τ）が判定用しきい値δｒ２より大きい場合（Ｒ（τ）＞δｒ２）、音声区間と判定し、それ以外の場合（Ｒ（τ）≦δｒ２）、非音声区間と判定する。ここで、判定用しきい値δｒ２は例えば実験により得る。そして、音声／非音声判定部３２は、その判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。

【００５４】

次に第２の実施形態における効果を説明する。

先ず、第２の実施形態では、前述した第１の実施形態と同様な効果を得ることができる。

さらに、第２の実施形態では、無指向性マイクに発話音源からの音及び雑音源からの音からなる混合音が入力され、単一指向性マイクに発話音源からの音又は雑音源からの音のいずれか一方が入力され、無指向性マイクに入力された混合音の音声信号と単一指向性マイクに入力された発話音源からの音又は雑音源からの音のいずれか一方の音の音声信号との比較により相関度を得て、その相関度に基づいて、発話区間を検出している。

【００５５】

これにより、無指向性マイクに発話音源からの音及び雑音源からの音からなる混合音が入力され、単一指向性マイクに発話音源からの音又は雑音源からの音のいずれか一方が入力されるように、無指向性マイク及び単一指向性マイク（第１及び第２マイク１，２）を配置する限り、マイクの取り付け位置の変化等による環境の変化、話者の移動や姿勢の変化等による音源の移動に対してロバストな発話区間検出システムを構築することができる。

そして、このように精度よく発話区間を検出することができるので、分離行列を劣化させることなく、最適化することができるようになる。

【００５６】

次に第３の実施形態を説明する。

この第３の実施形態は、発話区間判定部１１を具体的な構成とした音声信号処理装置１０であり、発話区間判定部１１が、第１及び第２マイク１，２で受音した音声信号ｘ１（ｔ），ｘ２（ｔ）のパワースペクトルに基づいて発話区間を検出するように構成されている。

図８は、その第３の実施形態における発話区間判定部１１の構成を示す。

また、前述の第２の実施形態と同様、第１マイク１として単一指向性マイクを使用し、第２マイク２として無指向性マイクを使用している。そして、第１及び第２マイク１，２の配置についても、前記図６に示したような配置にしている。これにより、雑音源からの音ｓ２（ｔ）を第１及び第２マイク１，２で受音し、発話音源（ユーザ）からの音ｓ１（ｔ）を第２マイク２だけで受音するようにしている。

【００５７】

このように配置した第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）はそれぞれ、前述したように、第１及び第２フレーム化部２１，２２に入力される。そして、前述したように、第１フレーム化部２１では、第１マイク１から入力された音声信号ｘ１（ｔ）をフレーム化（或いはフレーム分割）し、また、第２フレームか部２２では、第２マイク２から入力される音声信号ｘ２（ｔ）をフレーム化（或いはフレーム分割）する。そして、このように各フレーム化部２１，２２で複数フレームにされた音声信号ｘ１（ｔ），ｘ２（ｔ）は発話区間判定部１１に入力される。

【００５８】

発話区間判定部１１は、図８に示すように、パワースペクトラム計算部４１、パワー比計算部４２及び音声／非音声判定部４３を備えている。このような発話区間判定部１１において、各フレーム化部２１，２２で複数フレームにされた音声信号ｘ１（ｔ），ｘ２（ｔ）がパワースペクトラム計算部４１に入力される。

パワースペクトラム計算部４１は、フレーム単位で音声信号ｘ１（ｔ），ｘ２（ｔ）の第１及び第２パワースペクトル値$Px_1(\omega)$，$Px_2(\omega)$を算出し、その算出した第１及び第２パワースペクトル値$Px_1(\omega)$，$Px_2(\omega)$をパワー比計算部４２に出力する。

パワー比計算部４２は、下記（２）式により、パワースペクトラム計算部４１からの第１パワースペクトル値$Px_1(\omega)$と第２パワースペクトル値$Px_2(\omega)$との比（以下、パワー比という。）Ｐ（ω）を算出する。

【００５９】

【数２】

$$P(\omega)=G\frac{P_x(\omega)}{P_y(\omega)} \quad \cdots(2)$$

【００６０】

ここで、Ｇは、第１及び第２マイク１，２の感度によって決まる補正係数である。

パワー比計算部４２は、このようなパワー比Ｐ（ω）を音声／非音声判定部４３に出力する。

音声／非音声判定部４３は、パワー比Ｐ（ω）に基づいて音声区間と非音声区間とを判定する。具体的には、次のように音声区間と非音声区間とを判定する。

前述したように、発話音源（ユーザ）と雑音源に対して前記図６のように第１及び第２マイク１，２を配置することで、雑音源からの音ｓ２（ｔ）を第１及び第２マイク１，２で受音し、話者音源（ユーザ）からの音ｓ１（ｔ）を第２マイク２だけで受音している。

【００６１】

これにより、雑音源からの音ｓ２（ｔ）だけを第１及び第２マイク１，２で受音している場合には、同じ音声信号が第１及び第２マイク１，２に入力されているので、すなわち第１及び第２マイク１，２の受音感度が同程度であるので、このときにパワースペクトラム計算部４１で算出される第１及び第２パワースペクトル値Px_1（ω），Px_2（ω）は同程度になる。一方、発話音源（ユーザ）から発話があった場合には、その発話を第２マイク２だけが受音するので、すなわち第２マイク２の受音感度の方が大きくなるので、このときに第１パワースペクトル値Px_1（ω）よりも第２パワースペクトル値Px_2（ω）の方が大きくなる。このとき、パワー比計算部４２が算出するパワー比Ｐ（ω）は小さくなる。

【００６２】

なお、このとき、雑音源や発話音源（ユーザ）の特性に応じて、所定の周波数域のパワースペクトル値Px_1（ω），Px_2（ω）が特に変化する。

このように、発話音源（ユーザ）から発話があった場合にはパワー比Ｐ（ω）は小さくなることから、音声／非音声判定部４３は、パワー比Ｐ（ω）と判定用しきい値（類似度を示すしきい値）δｐ１とを比較して、音声区間を判定する。

【００６３】

ここで、パワースペクトラム計算部４１では、パワースペクトル値Px_1（ω），Px_2（ω）を所定の周波数域を対象として得ている。よって、パワー比Ｐ（ω）は、各周波数帯について得ることができる。

このようなことから、パワースペクトル値Px_1（ω），Px_2（ω）について各周波数で得ているパワー比Ｐ（ω）の総和平均値を算出し、判定では、その総和平均値と判定用しきい値δｐ１とを比較する。ここで、判定用しきい値δｐ１は例えば実験により得る。

【００６４】

なお、判定対象としてパワースペクトル値Px_1（ω），Px_2（ω）の全周波数域の総和平均値を用いることに限定されるものではない。例えば、発話音源（ユーザ）の特性を示す特定の周波数帯のパワー比Ｐ（ω）の総和平均値と判定用しきい値δｐ１とを比較したり、雑音源の特性を示す特定の周波数帯のパワー比Ｐ（ω）の平均値と判定用しきい値δｐ１とを比較したり、又は発話音源（ユーザ）の特性を示す特定の周波数帯のパワー比Ｐ（ω）と雑音源の特性を示す特定の周波数帯のパワー比Ｐ（ω）との平均値と判定用しきい値δｐ１とを比較したりしてもよい。この場合、それに応じて、判定用しきい値δｐ１を設定する。

【００６５】

そして、音声／非音声判定部４３は、パワー比Ｐ（ω）が判定用しきい値δｐ１未満の場合（Ｐ（ω）＜δｐ１）、音声区間と判定し、それ以外の場合（Ｐ（ω）≧δｐ１）、

非音声区間と判定する。ここで、音声／非音声判定部４３は、このような判定をフレーム単位で行う。そして、発話区間判定部１１は、このように音声／非音声判定部４３で得た音声区間（発話区間）の判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。

【００６６】

以上のように、発話区間判定部１１では、パワースペクトラム計算部４１が第１及び第２フレーム化部２１，２２それぞれから出力されるフレーム単位で第１及び第２パワースペクトル値Ｐｘ（ω），Ｐｙ（ω）を算出して、算出した第１及び第２パワースペクトル値Ｐｘ（ω），Ｐｙ（ω）をパワー比計算部４２に出力する。パワー比計算部４２では、パワースペクトラム計算部４１から出力される第１及び第２パワースペクトル値$Ｐｘ_{1}$（ω），$Ｐｘ_{2}$（ω）について、フレーム単位でパワー比Ｐ（ω）を算出して、算出したパワー比Ｐ（ω）を音声／非音声判定部４３に出力する。

【００６７】

音声／非音声判定部４３では、パワー比Ｐ（ω）と判定用しきい値δｐ１とを比較し、パワー比Ｐ（ω）に対応するフレームが音声区間のものか、非音声区間のものかを判定する。そして、音声／非音声判定部４３は、その判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。

そして、分離処理オン／オフ制御部１２は、前述したように、発話区間判定部１１からの発話区間の判定結果に基づいて、分離処理部２５への観測信号ｘ１（ｆ），ｘ２（ｆ）の出力をオン及びオフ制御する。また、最適化計算オン／オフ制御部１３は、前述したように、発話区間判定部１１からの発話区間の判定結果に基づいて、分離行列最適化計算部２６のオン及びオフを制御する。

【００６８】

このように、第３の実施形態として、発話区間判定部１１を構成することにより、前述した第１の実施形態に加えて、第２の実施形態と同様な効果を得ることができる。すなわち、無指向性マイクに発話音源からの音及び雑音源からの音からなる混合音が入力され、単一指向性マイクに発話音源からの音又は雑音源からの音のいずれか一方が入力されるように、無指向性マイク及び単一指向性マイク（第１及び第２マイク１，２）を配置する限り、マイクの取り付け位置の変化等による環境の変化、話者の移動や姿勢の変化等による音源の移動に対してロバストな発話区間検出システムを構築することができる。そして、このように精度よく発話区間を検出することができるので、分離行列を劣化させることなく、最適化することができるようになる。

【００６９】

次に第４の実施形態を説明する。

この第４の実施形態は、発話区間判定部１１を具体的な構成とした音声信号処理装置１０であり、第１及び第２マイク１，２で受音した音声信号ｘ１（ｔ），ｘ２（ｔ）のクロススペクトルに基づいて発話区間を検出するように構成されている。

なお、第１及び第２マイク１，２で受音した音声信号ｘ１（ｔ），ｘ２（ｔ）のクロススペクトルに基づいて発話区間を検出する技術については、例えば、多々良潔による「複数マイクロホンを用いた音声認識用耐雑音受音系の研究」（名古屋大学大学院工学研究科，修士論文，２００３年３月）で開示されている。

【００７０】

以下、このような開示技術を適用して構成した第４の実施形態における発話区間判定部１１の構成を説明する。図９は、その第４の実施形態における発話区間判定部１１の構成を示す。

図９に示すように、発話区間判定部１１は、クロススペクトル計算部５１、位相抽出処理部５２、位相unwrap処理部５３及び主計算部６０を備えている。また、主計算部６０は、周波数帯域分割部６１、第１乃至第Ｎ傾き計算部６２_{1}～６２_{N}、ヒストグラム等計算部６３及び音声／非音声判定部６４を備えている。なお、第１及び第２マイク１，２の配置については、当該第１及び第２マイク１，２に発話音源（ユーザ）からの音ｓ１（ｔ）

が入力されるように配置されている。

【００７１】

このような発話区間判定部１１において、各周波数分析部２３，２４から出力された信号ｘ１（ｆ），ｘ２（ｆ）がクロススペクトル計算部５１に入力される。

ここで、例えば、第１マイク１と第２マイク２といった複数のマイクで受音した音声信号を時間軸上でみた場合、受音した音声信号間に位相差が生じる。これは、音源から各マイク１，２までの距離の違いにより、音源から各マイク１，２までの音声信号の到達時間に差が生じた結果である。

【００７２】

ここで、第１マイク１と第２マイク２とにより受音した音声信号間の遅延時間を計測し、その計測した遅延時間に基づいて位相を同相化し、その後、第１マイク１と第２マイクとでそれぞれ受音した音声信号を加算して同期加算音声を得る場合を考える。例えば、M. Omologo, P.Svaizerらの文献「“Acousticevent localization using a crosspower-spectruum phase based technique”, Proc.ICASSP94, pp.274-276, (1994)」に、そのように同期加算音声を得る技術が記載されている。

【００７３】

ここで、２つのマイク１，２で受音した音声信号ｘ１（ｔ），ｘ２（ｔ）をフーリエ変換して得られる周波数関数をＸ１（ω），Ｘ２（ω）とする。ここで、ｘ２（ｔ）は、下記（３）式のようにｘ１（ｔ）の時間移動波形であると仮定する。

$$x2(t) = x1(t - t_0) \quad \cdots (3)$$

このように仮定した場合、周波数関数Ｘ１（ω）と周波数関数Ｘ２（ω）との関係は下記（４）式のようになる。

$$X2(\omega) = e^{-j\omega t_0} X1(\omega) \quad \cdots (4)$$

そして、この周波数関数Ｘ１（ω）と周波数関数Ｘ２（ω）とからクロススペクトルG_{12}（ω）が下記（５）式として得られる。

$$G_{12}(\omega) = X1(\omega) X2^{*}(\omega) = X1(\omega) e^{j\omega t_0} X1^{*}(\omega) = |X1|^2 e^{j\omega t_0} \quad \cdots (5)$$

【００７４】

ここで、クロススペクトルG_{12}（ω）の指数項はスペクトル領域のチャネル間の時間遅れに対応する。したがって、周波数関数Ｘ２に遅延項$e^{j\omega t_0}$をかけたＸ２（ω）$e^{j\omega t_0}$は、周波数関数Ｘ１と同相化され、これにより、Ｘ１（ω）＋Ｘ２（ω）$e^{j\omega t_0}$の逆フーリエ変換をチャネル同期加算音声として扱うことができるようになる。

【００７５】

クロススペクトル計算部５１では、このようなクロススペクトルG_{12}（ω）を得る。そのため、第１周波数分析部２３は、第１フレーム化部２１からの音声信号をフーリエ変換して前記周波数関数Ｘ１（ω）を算出して、その周波数関数Ｘ１（ω）（ｘ１（ｆ））をクロススペクトル計算部５１に出力する。また、第２周波数分析部２４は、第２フレーム化部２２からの音声信号をフーリエ変換して周波数関数Ｘ２（ω）を算出して、その周波数関数Ｘ２（ω）（ｘ２（ｆ））をクロススペクトル計算部５１に出力する。ここで、第１及び第２周波数分析部２３，２４は、フレーム毎に音声信号をフーリエ変換する。

クロススペクトル計算部５１は、第１及び第２周波数分析部２３，２４からの周波数関数Ｘ１（ω）、Ｘ２（ω）である前記信号ｘ１（ｆ），ｘ２（ｆ）に基づいて、前記（５）式によりクロススペクトルG_{12}（ω）を算出する。

【００７６】

なお、図１０は、１フレームについての音声信号のクロススペクトルの位相を示す。ここで、図１０中（Ａ）は自動車内で発した音声について得たクロススペクトルの位相であり、図１０中（Ｂ）はオフィススペース内で発した音声について得たクロススペクトルの位相であり、図１０中（Ｃ）は防音室内で発した音声について得たクロススペクトルの位相であり、図１０中（Ｄ）は歩道（屋外）で発した音声について得たクロススペクトルの位相である。この図１０に示すように、フレーム内で（すなわち局所的に）、音源と第１

マイク１までの距離と音源と第２マイク２までの距離との差に対応して、クロススペクトルの位相が周波数に対してほぼ一定の傾きを示すことがわかる。すなわち、音源と第１マイク１までの距離と音源と第２マイク２までの距離との差に対応して、クロススペクトルの位相成分が一定の傾きを有している。

【００７７】

そして、第１及び第２マイク１，２で受音した音声信号のＳ／Ｎ比が高ければ、そのように傾きが一定となる傾向は顕著になる。よって、第１及び第２マイク１，２により音声（発話）を受音した場合のその音声信号はＳ／Ｎ比が高くなり、この場合、明らかに一定の傾きを示すものになる。

クロススペクトル計算部５１は、このような特性を有するクロススペクトルG_{12}（ω）を位相抽出部５２に出力する。

【００７８】

位相抽出部５２では、クロススペクトル計算部５１からのクロススペクトルG_{12}（ω）から位相を抽出（検出）して、その抽出結果を位相unwrap処理部５３に出力する。

位相unwrap処理部５３では、位相抽出部５２の位相抽出結果に基づいて、クロススペクトルG_{12}（ω）をunwrap処理して、主計算部６０の周波数帯域分割部６１に出力する。

周波数帯域分割部６１は、帯域分割（セグメント分割）した位相を第１乃至第Ｎ傾き計算部62_1～62_Nそれぞれに出力する。

【００７９】

ここで、音声の入力されていない非音声区間フレームと音声が入力されている音声区間フレームとで、クロススペクトルの位相成分に大きな違いがある。すなわち、音声区間フレームでは、前述したようにクロススペクトルの位相が周波数に対してほぼ一定の傾きを示すが、非音声区間フレームでは、そのようにはならない。ここで、図１１を用いて説明する。

【００８０】

図１１はクロススペクトルの位相を示しており、図１１中（Ａ）は、音声区間フレームのクロススペクトルの位相であり、図１１中（Ｂ）は、非音声区間フレームのクロススペクトルの位相である。

この図１１中（Ａ）と図１１中（Ｂ）との比較からもかわるように、非音声区間フレームでは、クロススペクトルの位相は、周波数に対して特定のトレンドをもたない。すなわち、周波数に対してクロススペクトルの位相が一定の傾きを持つ結果とはならない。これは、ノイズの位相がランダムだからである。

【００８１】

これに対して、音声区間フレームでは、周波数に対してクロススペクトルの位相が一定の傾きをもつようになる。そして、この傾きは、音源から各マイク１，２までの距離の差に対応した大きさになる。

このように、音声の入力されていない非音声区間フレームと音声が入力されている音声区間フレームとでは、クロススペクトルの位相成分に大きな違いがある。

【００８２】

このようなことから、位相の回転が生じた場合にも正確にトレンドを追随するために、周波数帯域分割部６１により、位相成分を小さな周波数セグメントに分割（或いは帯域分割）し、後段の第１乃至第Ｎ傾き計算部62_1～62_Nで、最小２乗法を適用することでセグメント毎に傾きを計算している。この第１乃至第Ｎ傾き計算部62_1～62_Nはそれぞれ、算出した傾きをヒストグラム等計算部６３に出力する。

【００８３】

ここで、最小２乗法によりセグメント毎に傾きを求める手法は、公知の技術であり、例えば、『「信号処理」「画像処理」のための入門工学社』（高井信勝著，工学社，２０００）にその技術が記載されている。

ヒストグラム等計算部６３は、第１乃至第Ｎ傾き計算部62_1～62_Nが算出した前記傾きについて、ヒストグラムを得る。

【００８４】
図１２は、ヒストグラム等計算部６３が得たヒストグラムであり、セグメント毎に得た傾きについてのヒストグラムを示す。すなわち、この図１２は、位相の傾きの分布を示し、全セグメントに対する、各傾きのセグメント数の割合、すなわち頻度を縦軸にとっている。ここで、図１２中（Ａ）は、音声区間フレームについてのヒストグラムを示し、図１２中（Ｂ）は、非音声区間フレームについてのヒストグラムを示す。
【００８５】
図１２中（Ａ）と図１２中（Ｂ）との比較からもかわるように、音声区間フレームでは、ヒストグラムに明らかにピーク値があり、すなわち傾きがごく狭い範囲に局在しており、これにより、ある傾きについて頻度が高くなっている。すなわち、帯域毎のそれぞれの傾きが特定の傾きに集中する傾向が強くなっている。一方、非音声区間フレームでは、ヒストグラムが平滑となり、傾きが広い範囲にわたって分布している。
【００８６】
このヒストグラム等計算部６３は、このようなヒストグラム化して得た頻度を音声／非音声判定部６４に出力する。なお、このヒストグラム等計算部６３の処理については後で具体例を説明する。
音声／非音声判定部６４は、ヒストグラム等計算部６３からの前記頻度に基づいて、音声区間と非音声区間とを判定する。例えば、前記頻度の平均値周辺の所定の範囲に含まれる傾きの出現頻度が所定のしきい値以上の場合、音声区間と判定し、頻度が所定のしきい値未満の場合、非音声区間と判定する。
なお、ここでは、前段の処理がフレーム単位の処理となっているので、当該フレームが、音声区間フレーム又は非音声区間フレームのいずれかであるかを判定する。音声／非音声判定部６４は、その判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。
【００８７】
次にヒストグラム等計算部６３の具体的な構成を説明する。図１３は、その構成例を示す。
ヒストグラム等計算部６３は、第１乃至第Ｎ傾き計算部62_1～62_Nが算出した前記傾きのうちから頻度が高い（最頻度の）傾きを算出する構成として、第１スイッチ６３Ｓ１、第２スイッチ６３Ｓ２及び最頻値計算部６３Ｃを備えている。これにより、第１スイッチ６３Ｓ１を一定時間オン（閉）にして、第１乃至第Ｎ傾き計算部62_1～62_Nが算出した一定時間の前記傾きのデータ（或いはデータベース）６３Ｄ１を作成する。このとき、第２スイッチ６３Ｓ２については、オフ（開）にしておく。そして、データ６３Ｄ１を作成したら、第２スイッチ６３Ｓ２をオン（閉）にして、そのデータ６３Ｄ１を最頻値計算部６３Ｃに出力する。
【００８８】
最頻値計算部６３Ｃでは、データ６３Ｄ１から前記図１２に示すような前記傾きについてのヒストグラムを作成して、そのヒストグラム中の最頻度の傾き（以下、最頻傾きという。）τ０を算出する。
なお、最頻度の傾きを算出するようにしてもよいが、平均値の傾きτ０を算出したり、或いは最頻度の傾きと傾きの平均値とを組み合わせた傾きτ０を算出するようにしてもよい。これにより、各帯域の傾きが特定の傾きに集中する傾向が強くなったとき、当該特定の傾きの値そのもの或いはそれに近い傾きの値を得ることができる。なお、本実施の形態では、最頻値計算部６３Ｃが最頻傾きτ０を算出しているものとする。
【００８９】
そして、最頻値計算部６３Ｃは、算出した最頻傾きτ０を前記音声／非音声判定部６４に出力する。ここで、最頻傾きτ０をデータ６３Ｄ２として前記音声／非音声判定部６４に出力する。
音声／非音声判定部３４では、ヒストグラム等計算部６３からの最頻傾きτ０に基づいて、音声区間と非音声区間とを判定する。

【００９０】
なお、先の説明では、音声／非音声判定部３４がヒストグラム等計算部６３からの前記頻度に基づいて音声区間と非音声区間とを判定する場合について説明した。ここでは、音声／非音声判定部６４は、ヒストグラム等計算部６３からの最頻傾きτ０と第１乃至第Ｎ傾き計算部62_1～62_Nが算出した前記傾きτｉに基づいて、音声区間と非音声区間とを判定しており、これに対応して、音声／非音声判定部６４に、第１乃至第Ｎ傾き計算部62_1～62_Nが算出した前記傾きが入力されるようになっている。
【００９１】
すなわち、音声／非音声判定部６４は、第１乃至第Ｎ傾き計算部62_1～62_Nが算出した前記傾きτｉと最頻傾きτ０とを下記（６）式により比較する。
$$|\tau i-\tau 0|<\delta_\tau \quad \cdots (6)$$
ここで、δ_τは判定用のしきい値（傾きしきい値）である。
音声／非音声判定部３４は、この（６）式の条件が満たされていることが所定の割合を超えた場合（ＹＥＳ）、音声区間と判定し、そうでない場合（ＮＯ）、非音声区間と判定する。そして、音声／非音声判定部６４は、その判定結果を分離処理オン／オフ制御部１２及び最適化計算オン／オフ制御部１３に出力する。
【００９２】
次に第４の実施形態における効果を説明する。
先ず、第４の実施形態では、前述した第１の実施形態と同様な効果を得ることができる。
さらに、第４の実施形態では、第１及び第２マイク１，２に入力された音声信号間のクロススペクトルの位相を検出し、その検出したクロススペクトルの位相の周波数に対する傾きに基づいて、当該複数のマイクロホンが受音した音声信号中の発話区間を検出している。すなわち、音声が入力（発話入力）されていない音声信号と音声が入力（発話入力）されている音声信号とをクロススペクトルでみた場合に、そのクロススペクトルの位相成分に大きな違いがあることを利用して、当該複数のマイクロホンが受音した音声信号中の発話区間を検出している。具体的には、クロススペクトルの位相を帯域分割（セグメント分割）し、帯域毎（セグメント毎）の位相の傾きからヒストグラムを生成し、そのヒストグラムから頻度（具体的には最頻値）を得て、その頻度に基づいて、発話区間を検出している。これにより、精度よく発話区間を検出することができる。そして、このように精度よく発話区間を検出することができるので、分離行列を劣化させることなく、最適化することができるようになる。
【００９３】
なお、前述の実施形態では、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）を、直接第１及び第２フレーム化部２１，２２にそれぞれ入力しているが、具体的には、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）を、ＡＤ（アナログ／デジタル）変換した後、第１及び第２フレーム化部２１，２２に入力するようにする。これを、図２に示した実施形態の音声信号処理装置１０の構成に適用すると、図１４に示すような構成になる。
【００９４】
この図１４に示すように、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）をそれぞれ、第１及び第２ＡＤ変換部７１，７２でＡＤ変換した後、第１及び第２フレーム化部２１，２２に入力する。
ここで、第１及び第２ＡＤ変換部７１，７２でＡＤ変換されたデータ形式は、例えば１０２５Ｈｚ、１６ｂｉｔ、リニアＰＣＭである。また、第１及び第２フレーム化部２１，２２でフレーム化された信号のフレーム長は、例えば５１２サンプルフレーム長である。
【００９５】
また、前述の実施形態では、検出対象音が人間が発する発話音である場合を説明したが、検出対象音は、人間以外の物体が発する音でもよい。

また、前述の実施形態の説明において、発話区間判定部１１は、検出対象音源からの音を検出する対象音件手段又は発話源からの音の発話区間を検出する発話区間検出手段を実現しており、分離処理オン／オフ制御部１２、最適化計算オン／オフ制御部１３及び分離行列最適化計算部２６は、前記対象音検出手段又は発話区間検出手段が検出対象音源からの音又は発話区間を検出した場合、第１及び第２マイクに入力された音信号を用いて分離行列を最適化する分離行列最適化手段を実現しており、分離処理部２５は、前記分離行列最適化手段が最適化した分離行列を用いて、混合音から検出対象音源の音又は発話源からの音と雑音源からの音とを分離する分離手段を実現している。

【００９６】

また、前述の実施形態の音声信号処理装置１０を音声認識装置に適用することができる。この場合、音声認識装置は、前述したような音声信号処理装置１０の構成に加えて、音声信号処理装置１０が検出した発話区間の音声信号について音声認識処理をする音声認識処理手段を備える。

ここで、音声認識技術としては、例えば、旭化成株式会社が提供する音声認識技術「ＶＯＲＥＲＯ」（商標）（http://www.asahi-kasei.co.jp/vorero/jp/vorero/feature.html参照）等があり、このような音声認識技術の用いた音声認識装置に適用することもできる。

【００９７】

また、前述の実施形態の音声信号処理装置１０をコンピュータで実現することができる。そして、前述したような音声信号処理装置１０の処理内容をコンピュータが所定のプログラムにより実現する。この場合、プログラムは、検出対象音源からの音の検出の有無により分離行列の最適化の実行を切換可能にするとともに、検出対象音源からの音を検出した場合、分離行列の最適化を行い、検出対象音源からの音を検出できない場合、分離行列の最適化を行わないようにコンピュータに実行させるプログラムになる。

【図面の簡単な説明】

【００９８】

【図１】本発明の実施形態の音声信号処理装置を含むシステム全体の構成を示すブロック図である。

【図２】前記第１の実施形態の音声信号処理装置の構成を示すブロック図である。

【図３】前記第１の実施形態の音声信号処理装置の一連の動作順序を示すフローチャートである。

【図４】前記第１の実施形態における効果の説明に使用した図である。

【図５】本発明の第２の実施形態における発話区間判定部の構成を示すブロック図である。

【図６】前記第２の実施形態におけるマイクの配置を示す図である。

【図７】前記第２の実施形態におけるマイクの他の配置を示す図である。

【図８】本発明の第３の実施形態における発話区間判定部の構成を示すブロック図である。

【図９】本発明の第４の実施形態における発話区間判定部の構成を示すブロック図である。

【図１０】各環境のクロススペクトルの位相を示す特性図である。

【図１１】クロススペクトルの位相を示す特性図であり、（Ａ）は、音声区間フレームのクロススペクトルの位相を示す特性図であり、（Ｂ）は、非音声区間フレームのクロススペクトルの位相を示す特性図である。

【図１２】クロススペクトルの位相に基づいて得たヒストグラムを示す特性図であり、（Ａ）は、音声区間フレームのヒストグラムを示す特性図であり、（Ｂ）は、非音声区間フレームのヒストグラムを示す特性図である。

【図１３】前記第４の実施形態におけるヒストグラム等計算部などの構成を示すブロック図である。

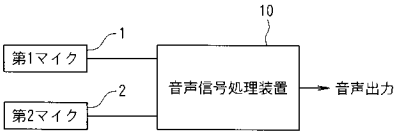
【図１４】前記第１の実施形態の他の構成例を示すブロック図である。

【符号の説明】

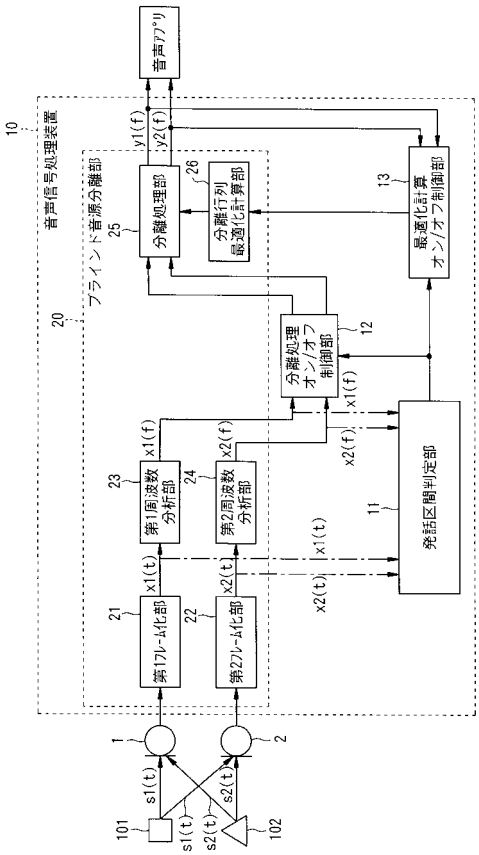
【００９９】

１，２　マイク
１０　音声信号処理装置
１１　発話区間判定部
１２　分離処理オン／オフ制御部
１３　最適化計算オン／オフ制御部
２０　ブラインド音源分離部
２１，２２　フレーム化部
２３，２４　周波数分析部
２５　分離処理部
２６　分離行列最適化計算部
１０１　話者音源
１０２　雑音源

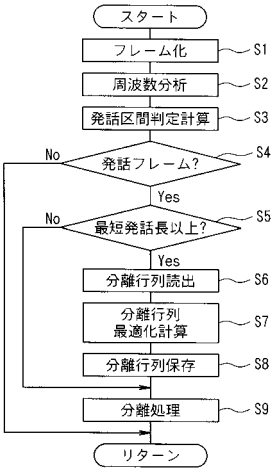
【図１】



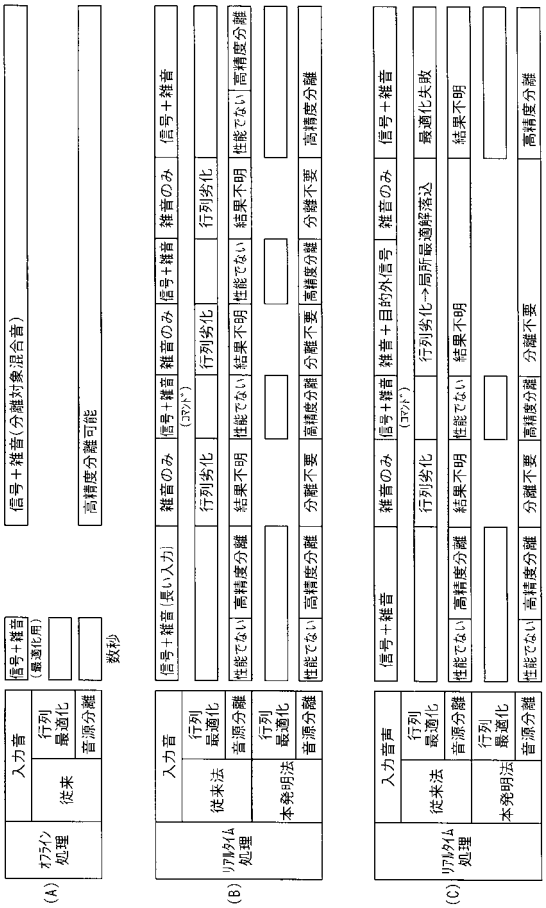
【図２】



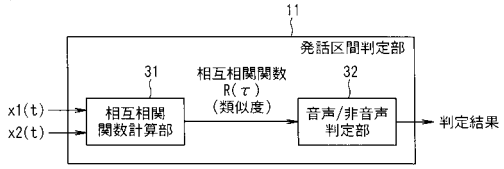
【図３】



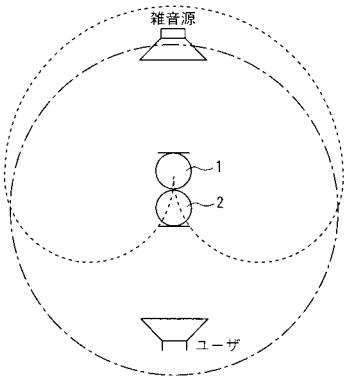
【図４】



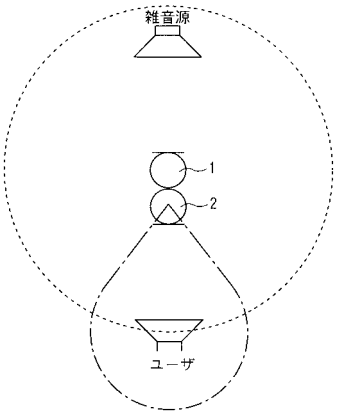
【図５】



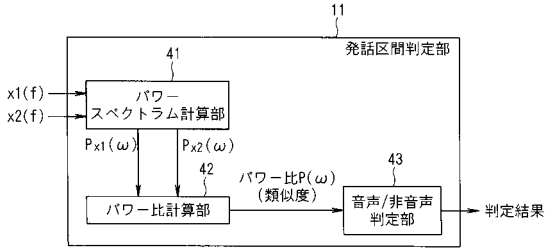
【図６】



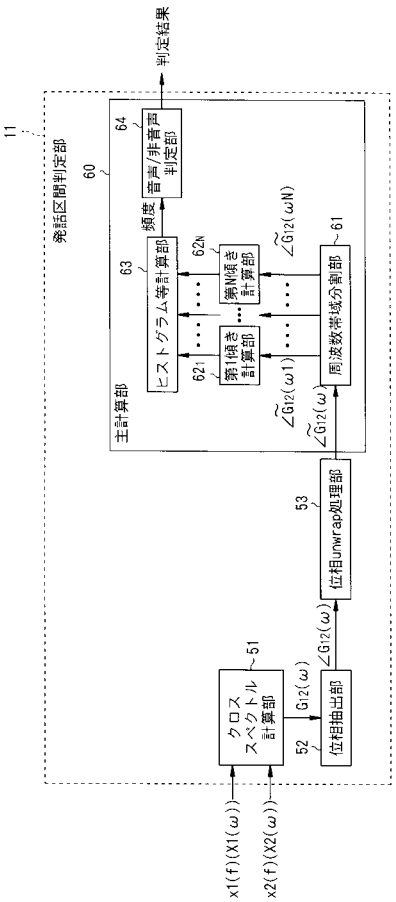
【図７】



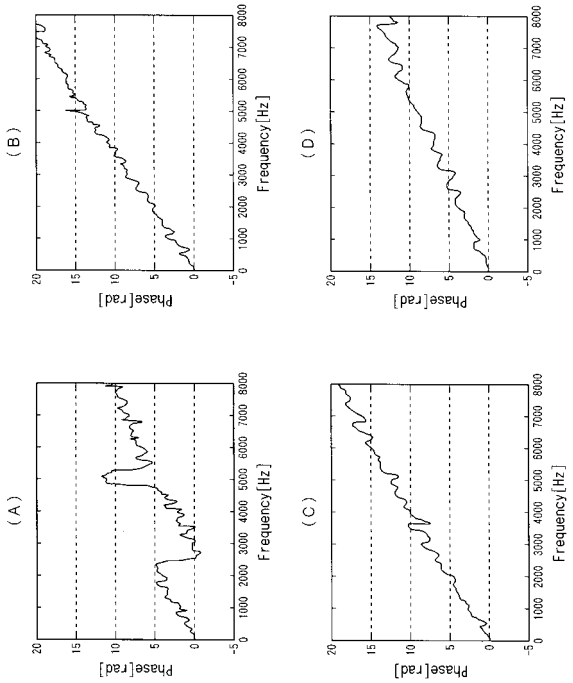
【図８】



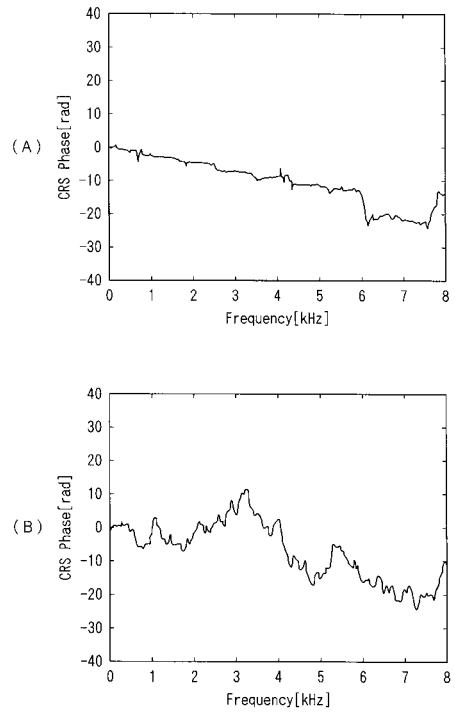
【図９】



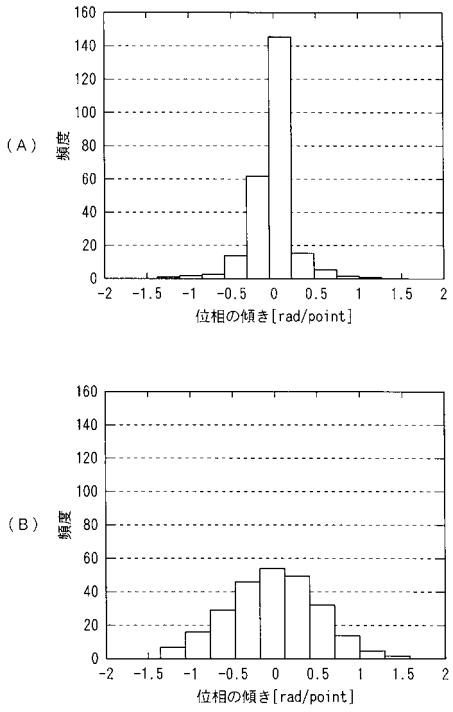
【図１０】



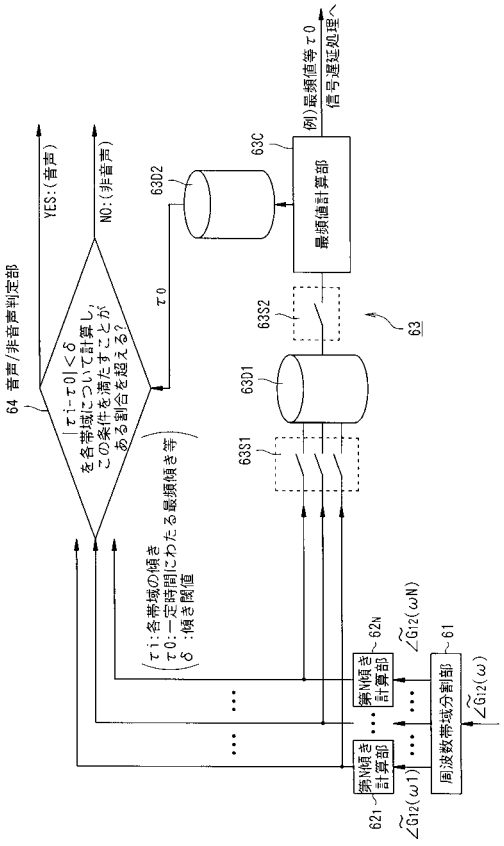
【図１１】



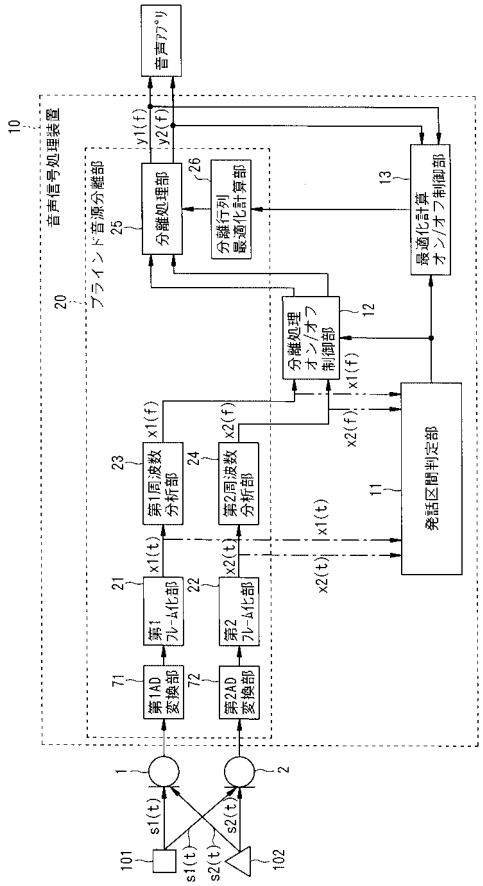
【図１２】



【図１３】


64 音声/非音声判定部
YES:(音声)
NO:(非音声)
|τi-τ0|<δ
を各帯域について計算し，
この条件を満たすことが
ある割合を超える?
τi:各帯域の傾き
τ0:一定時間にわたる最頻傾き等
δ :傾き閾値
τ0
6302
63C
例)最頻値等τ0
信号遅延処理へ
最頻値計算部
63S2
63
6301
6331
第Nの傾き
計算部
62N
第1の傾き
計算部
621
∠G12(ω1)
∠G12(ωN)
周波数帯域分割部
61
∠G12(ω)

【図１４】


10
音声信号処理装置
20
ブラインド音源分離部
音声ｱﾌﾟﾘ
y1(f)
y2(f)
25
分離処理部
26
分離行列
最適化計算部
13
最適化計算
オン/オフ制御部
12
分離処理
オン/オフ
制御部
x1(f)
x2(f)
23
第1周波数
分析部
24
第2周波数
分析部
11
発話区間判定部
21
第1
ﾌﾚｰﾑ化部
22
第2
ﾌﾚｰﾑ化部
71
第1AD
変換部
72
第2AD
変換部
x1(t)
x2(t)
1
2
101
102
s1(t)
s2(t)

フロントページの続き

(51)Int.Cl.[7]		FI			テーマコード（参考）
Ｇ１０Ｌ	15/28	Ｇ１０Ｌ	3/02	３０１Ｆ	
Ｇ１０Ｌ	21/02	Ｇ１０Ｌ	3/00	５５１Ａ	
Ｈ０４Ｒ	3/00	Ｇ１０Ｌ	9/08	３０１Ａ	