(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2005-227511
(P2005-227511A)
(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	301E	5D015
G10L 11/02	G10L	3/00	511	
G10L 15/02	G10L	3/00	513B	
G10L 15/04	G10L	9/08	301A	
G10L 15/28				

審査請求 未請求　請求項の数 15　ＯＬ　（全 22 頁）　最終頁に続く

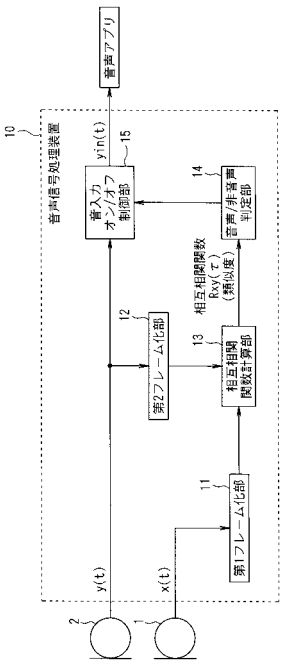
(21)出願番号　特願2004-35618(P2004-35618)
(22)出願日　平成16年2月12日(2004.2.12)

(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
(72)発明者　赤坂　貴志
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 DD02　DD03　EE04

(54)【発明の名称】対象音検出方法、音信号処理装置、音声認識装置及びプログラム

(57)【要約】
【課題】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を可能にする。
【解決手段】音声信号処理装置は、発話音又は雑音のいずれか一方を受音するように配置されている単一指向性マイク１と、発話音及び雑音を受音するように配置されている無指向性マイク２と、単一指向性マイク１に入力された音声信号ｘ（ｔ）と無指向性マイク２に入力された音声信号ｙ（ｔ）との相互相関関数Ｒｘｙ（τ）を算出するためのフレーム化部１１，１２及び相互相関関数計算部１３と、相互相関関数計算部１３が算出した相互相関関数Ｒｘｙ（τ）に基づいて、発話音の発話区間を検出する音声／非音声判定部１４とを備える。
【選択図】図３



【特許請求の範囲】
【請求項１】
無指向性マイクで検出対象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又は前記雑音のいずれか一方を受音し、前記無指向性マイクに入力された声信号と単一指向性マイクに入力された音信号とを比較し、その比較結果に基づいて、前記検出対象音を検出することを特徴とする対象音検出方法。
【請求項２】
前記無指向性マイクに入力された音信号と単一指向性マイクに入力された音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、前記検出対象音を検出することを特徴とする請求項１記載の対象音検出方法。
【請求項３】
前記無指向性マイクに入力された音信号のパワースペクトルと、単一指向性マイクに入力された音信号のパワースペクトルとを比較して、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１又は２記載の対象音検出方法。
【請求項４】
前記無指向性マイクに入力された音信号と単一指向性マイクに入力された音信号との比較により得た相関度、及び前記無指向性マイクに入力された音信号のパワースペクトルと、単一指向性マイクに入力された音信号のパワースペクトルと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１乃至３のいずれか１に記載の対象音検出方法。
【請求項５】
検出対象音及び雑音を受音するように配置されている無指向性マイクと、
前記検出対象音又は前記雑音のいずれか一方を受音するように配置されている単一指向性マイクと、
前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較する比較手段と、
前記比較手段の比較結果に基づいて、前記検出対象音を検出する対象音検出手段と、
を備えることを特徴とする音信号処理装置。
【請求項６】
前記比較手段は、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記対象音検出手段は、前記比較手段が算出した相関度と所定の第１しきい値とを比較して、前記検出対象音を検出することを特徴とする請求項５記載の音信号処理装置。
【請求項７】
前記比較手段は、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、
前記対象音検出手段は、前記パワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値とを比較して、前記検出対象音を検出することを特徴とする請求項５又は６記載の音信号処理装置。
【請求項８】
前記比較手段は、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、
前記比較手段は、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記対象音検出手段は、前記比較手段が算出した相関度と所定の第１しきい値との比較結果と、前記比較手段のパワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値との比較結果とに基づいて、

前記検出対象音を検出することを特徴とする請求項５乃至７のいずれか１に記載の音信号処理装置。

【請求項９】

前記無指向性マイクに入力された音信号及び単一指向性マイクに入力された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、

前記比較手段は、前記フレーム化手段から出力されるフレーム単位で、前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較し、前記対象音検出手段は、前記比較手段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５乃至８のいずれか１に記載の音信号処理装置。

【請求項１０】

発話音及び雑音を受音するように配置されている無指向性マイクと、

前記発話音又は前記雑音のいずれか一方を受音するように配置されている単一指向性マイクと、

前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較する比較手段と、

前記比較手段の比較結果に基づいて、前記発話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検出した前記発話音の発話区間について、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１１】

前記比較手段は、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記発話区間検出手段は、前記比較手段が算出した相関度と所定の第１しきい値とを比較して、前記発話音の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】

前記比較手段は、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、

前記発話区間検出手段は、前記パワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値とを比較して、前記発話音の発話区間を検出することを特徴とする請求項１０又は１１記載の音声認識装置。

【請求項１３】

前記比較手段は、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、

前記比較手段は、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記発話区間検出手段は、前記比較手段が算出した相関度と所定の第１しきい値との比較結果と、前記比較手段のパワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値との比較結果とに基づいて、前記発話音の発話区間を検出することを特徴とする請求項１０乃至１２のいずれか１に記載の音声認識装置。

【請求項１４】

前記無指向性マイクに入力された音信号及び単一指向性マイクに入力された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、

前記比較手段は、前記フレーム化手段から出力されるフレーム単位で、前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較し、前記発話

区間検出手段は、前記比較手段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音の発話区間を検出し、前記音声認識処理手段は、前記発話区間検出手段が検出したフレーム単位の前記発話音の発話区間について、音声認識処理を行うことを特徴とする請求項１０乃至１３のいずれか１に記載の音声認識装置。

【請求項１５】

無指向性マイクで受音した検出対象音及び雑音の音信号と単一指向性マイクで受音した前記検出対象音又は前記雑音のいずれか一方の音信号とを比較し、その比較結果に基づいて、前記検出対象音を検出する処理を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、入力音中から検出対象音を検出する対象音検出方法及びこれを実現するプログラム、入力された音信号を処理する音信号処理装置、並びに入力された発話音について音声認識処理を行う音声認識装置に関する。

【背景技術】

【０００２】

音声は、人間の用いる種々の通信の形態の中でも最も根源的なものであると同時に、他のどの情報送出方法よりも高速度に情報を送り出すことのできる優れた通信手段である。このようなことから、音声は、古くから現在に至るまで人間の通信手段の根幹を担っている。

音声認識技術は、そのような音声を認識するための技術である。音声認識とは、その音声に含まれる情報の中で、最も基本的な意味内容に関する情報、つまり音韻情報をコンピュータなどにより抽出し、その抽出内容を判定することである。近年では、計算機プロセッサ技術の飛躍的な発達と、インターネットに代表される高度な情報ネットワークの構築により、様々な分野においてマン・マシンインタフェースとしての音声認識技術の適用が試みられている。

現在の音声認識システムの認識性能は、確率・統計的手法により格段に向上しており、理想的な環境下での音声や接話マイクロホンで収録された近距離音声などでは、非常に高い認識率が得られるようになっている。

【発明の開示】

【発明が解決しようとする課題】

【０００３】

実環境下の音声認識は、学習データと観測データとの間の環境や発話内容のミスマッチ等により、その認識率が劣化する。また、受音系となる接話マイクヘッドセットの装着によりユーザが受ける負担や不快感は大きく、音声認識システム実用化の大きな障害のひとつになっている。

また、Ｓ／Ｎ比の低下や背景雑音、室内残響の影響などにより認識が困難な遠隔音声に関し、複数の遠隔マイクロホンを用いた音声認識手法の研究が多くなされている。その代表的なものとして、マイクロホンアレーを用いる手法が挙げられる。この手法では、音源位置検出処理、目的音強調処理、雑音抑制処理、の３つの空間的な信号処理を行なうことができる。このような手法により遠隔音声の音声認識が盛んに研究されている。

【０００４】

しかし、この手法は、正確な話者方向同定処理のために複数のマイクロホンを一定間隔にて固定配置する必要があり、小型化、携帯化が困難であるため、様々な環境・状況下での音声入力への応用が難しく、用途が限定されるという問題がある。

本発明は、前述の問題に鑑みてなされたものであり、マイクロホン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を可能にする対象音検出方法、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記載の対象音検出方法は、無指向性マイクで検出対象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又は前記雑音のいずれか一方を受音し、前記無指向性マイクに入力された声信号と単一指向性マイクに入力された音信号とを比較し、その比較結果に基づいて、前記検出対象音を検出することを特徴とする。
また、請求項２記載の対象音検出方法は、請求項１記載の対象音検出方法において、前記無指向性マイクに入力された音信号と単一指向性マイクに入力された音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、前記検出対象音を検出することを特徴とする。
【０００６】
また、請求項３記載の対象音検出方法は、請求項１又は２記載の対象音検出方法において、前記無指向性マイクに入力された音信号のパワースペクトルと、単一指向性マイクに入力された音信号のパワースペクトルとを比較して、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。
また、請求項４記載の対象音検出方法は、請求項１乃至３のいずれか１に記載の対象音検出方法において、前記無指向性マイクに入力された音信号と単一指向性マイクに入力された音信号との比較により得た相関度、及び前記無指向性マイクに入力された音信号のパワースペクトルと単一指向性マイクに入力された音信号のパワースペクトルと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。
【０００７】
また、請求項５記載の音信号処理装置は、検出対象音及び雑音を受音するように配置されている無指向性マイクと、前記検出対象音又は前記雑音のいずれか一方を受音するように配置されている単一指向性マイクと、前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較する比較手段と、前記比較手段の比較結果に基づいて、前記検出対象音を検出する対象音検出手段と、を備えることを特徴とする。
【０００８】
また、請求項６記載の音信号処理装置は、請求項５記載の音信号処理装置において、前記比較手段が、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記対象音検出手段が、前記比較手段が算出した相関度と所定の第１しきい値とを比較して、前記検出対象音を検出することを特徴とする。
また、請求項７記載の音信号処理装置は、請求項５又は６記載の音信号処理装置において、前記比較手段が、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、前記対象音検出手段が、前記パワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値とを比較して、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８記載の音信号処理装置は、請求項５乃至７のいずれか１に記載の音信号処理装置において、前記比較手段が、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、前記比較手段が、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記対象音検出手段は、前記比較手段が算出した相関度と所定の第１しきい値との比較結果と、前記比較手段のパワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値との比較結果とに基づいて、前記検出対象音を検出することを特徴とする。
【００１０】
また、請求項９記載の音信号処理装置は、請求項５乃至８のいずれか１に記載の音信号

処理装置において、前記無指向性マイクに入力された音信号及び単一指向性マイクに入力された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、前記比較手段が、前記フレーム化手段から出力されるフレーム単位で、前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較し、前記対象音検出手段は、前記比較手段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。

【００１１】

また、請求項１０記載の音声認識装置は、発話音及び雑音を受音するように配置されている無指向性マイクと、前記発話音又は前記雑音のいずれか一方を受音するように配置されている単一指向性マイクと、前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較する比較手段と、前記比較手段の比較結果に基づいて、前記発話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検出した前記発話音の発話区間について、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。

【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、前記比較手段が、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記発話区間検出手段が、前記比較手段が算出した相関度と所定の第１しきい値とを比較して、前記発話音の発話区間を検出することを特徴とする。

【００１３】

また、請求項１２記載の音声認識装置は、請求項１０又は１１記載の音声認識装置において、前記比較手段が、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、前記発話区間検出手段が、前記パワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値とを比較して、前記発話音の発話区間を検出することを特徴とする。

【００１４】

また、請求項１３記載の音声認識装置は、請求項１０乃至１２のいずれか１に記載の音声認識装置において、前記比較手段が、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各音信号のパワースペクトルをそれぞれ算出するパワースペクトル算出手段と、前記パワースペクトル算出手段が算出した２つのパワースペクトルの比を算出するパワースペクトル比算出手段とを備えており、前記比較手段が、前記無指向性マイクに入力された音信号と前記単一指向性マイクに入力された音信号との相関度を算出し、前記発話区間検出手段は、前記比較手段が算出した相関度と所定の第１しきい値との比較結果と、前記比較手段のパワースペクトル比算出手段が算出した前記パワースペクトルの比と所定の第２しきい値との比較結果とに基づいて、前記発話音の発話区間を検出することを特徴とする。

【００１５】

また、請求項１４記載の音声認識装置は、請求項１０乃至１３のいずれか１に記載の音声認識装置において、前記無指向性マイクに入力された音信号及び単一指向性マイクに入力された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、前記比較手段が、前記フレーム化手段から出力されるフレーム単位で、前記無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較し、前記発話区間検出手段が、前記比較手段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音の発話区間を検出し、前記音声認識処理手段は、前記発話区間検出手段が検出したフレーム単位の前記発話音の発話区間について、音声認識処理を行うことを特徴とする。

【００１６】

また、請求項１５記載のプログラムは、無指向性マイクで受音した検出対象音及び雑音の音信号と単一指向性マイクで受音した前記検出対象音又は前記雑音のいずれか一方の音信号とを比較し、その比較結果に基づいて、前記検出対象音を検出する処理をコンピュータに実行させることを特徴とする。

無指向性マイクで検出対象音及び雑音を受音し、単一指向性マイクで検出対象音又は雑音のいずれか一方を受音するようにした場合において、無指向性マイク及び単一指向性マイクが配置されている環境で雑音だけを発したときと検出対象音及び雑音を発したときとで、無指向性マイク及び単一指向性マイクに入力された音信号が異なってくる。なお、検出対象音には、人間が発する発話音の他、物体が発する音も含まれる。

【００１７】

そこで、請求項１、５、１５記載の発明では、無指向性マイクに入力された音信号と単一指向性マイクに入力された音信号とを比較することで、検出対象音又は発話音を検出している。また、請求項１０記載の発明では、検出対象音が人間が発する発話音であり、検出対象音の検出として、発話音の音声区間の検出を行っている。

また、請求項２、６、１１記載の発明では、無指向性マイクに入力された音信号と単一指向性マイクに入力された音信号との相関度により、検出対象音の検出又は発話音の発話区間の検出を行っている。

【００１８】

また、請求項３、７、１２記載の発明では、無指向性マイクに入力された音信号のパワースペクトルと単一指向性マイクに入力された音信号のパワースペクトルとを比較することで、検出対象音の検出又は発話音の発話区間の検出を行っている。

また、請求項４、８、１３記載の発明では、無指向性マイクに入力された音信号と単一指向性マイクに入力された音信号との相関度と、無指向性マイクに入力された音信号のパワースペクトルと単一指向性マイクに入力された音信号のパワースペクトルとの比較結果とに基づいて、検出対象音の検出又は発話音の発話区間の検出を行っている。

また、請求項９、１４記載の発明では、前記無指向性マイクに入力された音信号及び単一指向性マイクに入力された音信号を時分割してフレーム化し、フレーム単位でその後の処理を行う。

【発明の効果】

【００１９】

本発明によれば、無指向性マイクで検出対象音（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（又は発話音）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、検出対象音（又は発話音の音声区間）を検出することができる。これにより、マイクロホン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】

【００２０】

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。

第１マイク１は単一指向性マイクであり、第２マイク２は無指向性マイクであり、第１及び第２マイク１，２は例えば装着型マイクである。第１及び第２マイク１，２は、図２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をその指向方向が音源（ユーザ）の位置に対して反対側となるように配置している。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２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２に示す一点鎖線は、第２マイク２の指向特性を示す。

このように第１及び第２マイク１，２を配置すると、雑音源からの音は、第１及び第２

マイク１，２で受音でき、音源（ユーザ）からの音は第２マイク２だけが受音できるようになる。

【００２１】

図３は、第１の実施形態の音声信号処理装置１０の構成を示す。

図３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１１，１２、相互相関関数計算部１３、音声／非音声判定部１４並びに音入力オン／オフ制御部１５を備えている。

第１及び第２マイク１，２から入力された２ｃｈの音声信号はそれぞれ、第１及び第２フレーム化部１１，１２に入力される。また、第２マイク２から入力された音声信号は、音入力オン／オフ制御部１５に入力される。ここで、第１マイク１に入力された音声信号をｘ（ｔ）とし、第２マイク２に入力された音声信号をｙ（ｔ）とする。

【００２２】

第１フレーム化部１１では、第１マイク１から入力された音声信号ｘ（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ｘ（ｔ）を相互相関関数計算部１３に出力する。また、第２フレームか部１２では、第２マイク２から入力される音声信号ｙ（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ｙ（ｔ）を相互相関関数計算部１３に出力する。ここで、第１及び第２フレーム化部１１，１２は、入力されてくる音声信号ｘ（ｔ），ｙ（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。

【００２３】

相互相関関数計算部１３は、第１フレーム化部１１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ｘ（ｔ）と、第２マイク２に入力された音声信号ｙ（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１３は、下記（１）式により、相互相関関数Ｒｘｙ（τ）を算出する。

【００２４】

【数１】

$$R_{xy}(\tau)=\frac{1}{T}\int_{-T/2}^{T/2}x(t)y(t-\tau)dt \quad \cdots(1)$$

【００２５】

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Ｒｘｙ（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Ｒｘｙ（τ）を得ることができ、精度よく相互相関関数Ｒｘｙ（τ）を得ることができる。

【００２６】

このように算出された相互相関関数Ｒｘｙ（τ）はフレーム単位で各音声信号ｘ（ｔ），ｙ（ｔ）の波形形状の類似度を示す値となる。具体的には、相関関係を求める２つの音声信号ｘ（ｔ），ｙ（ｔ）が似ているほど、相互相関関数Ｒｘｙ（τ）は大きい値となり、相関関係を求める２つの音声信号ｘ（ｔ），ｙ（ｔ）が異なっているほど、相互相関関数Ｒｘｙ（τ）は０に近くなる。相互相関関数計算部１３は、このような相互相関関数Ｒｘｙ（τ）を音声／非音声判定部１４に出力する。

【００２７】
音声／非音声判定部１４は、相互相関関数Ｒｘｙ（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。
前述したように、音源（ユーザ）と雑音源に対して図２のように第１及び第２マイク１，２を配置することで、雑音源からの音を第１及び第２マイク１，２で受音し、音源（ユーザ）からの音を第２マイク２だけで受音している。
【００２８】
一方、相互相関関数Ｒｘｙ（τ）は、前述したように、相関関係を求める２つの音声信号ｘ（ｔ），ｙ（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ｘ（ｔ），ｙ（ｔ）が異なっているほど０に近くなる。
このようなことから、雑音源からの音だけを第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Ｒｘｙ（τ）は大きい値になる。一方、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Ｒｘｙ（τ）は０に向かって減少する。
【００２９】
このように、音源（ユーザ）から発話があった場合には相互相関関数Ｒｘｙ（τ）は０に向かって減少することから、音声／非音声判定部１４は、相互相関関数Ｒｘｙ（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１４は、相互相関関数Ｒｘｙ（τ）が判定用しきい値δｒ１未満の場合（Ｒｘｙ（τ）＜δｒ１）、音声区間と判定し、それ以外の場合（Ｒｘｙ（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１４は、このような判定をフレーム単位で行う。音声／非音声判定部１４は、その判定結果をフレーム単位で音入力オン／オフ制御部１５に出力する。
【００３０】
音入力オン／オフ制御部１５には、第２マイク２からの音声信号ｙ（ｔ）が入力されており、音入力オン／オフ制御部１５は、音声／非音声判定部１４の判定結果に基づいて、第２マイク２からの音声信号ｙ（ｔ）の後段への出力のオンとオフとを切り換える。具体的には、音声／非音声判定部１４が音声区間と判定した場合、音入力オン／オフ制御部１５は、オン制御として当該音声区間に対応する音声信号ｙ（ｔ）の区間を後段に出力して、音声／非音声判定部１４が非音声区間と判定した場合、音入力オン／オフ制御部１５は、オフ制御として当該非音声区間に対応する音声信号ｙ（ｔ）の区間を後段に出力しないようにする。
【００３１】
以上のように音声信号処理装置１０が構成されている。この音声信号処理装置１０における一連の動作は次のようになる。
先ず、第１及び第２フレーム化部１１，１２が、第１及び第２マイク１，２から入力された２ｃｈの音声信号ｘ（ｔ），ｙ（ｔ）をそれぞれフレーム化し、フレーム単位で音声信号ｘ（ｔ），ｙ（ｔ）を相互相関関数計算部１３に出力する。
【００３２】
相互相関関数計算部１３では、第１及び第２フレーム化部１１，１２それぞれから出力されるフレーム単位の音声信号ｘ（ｔ），ｙ（ｔ）について相互相関関数Ｒｘｙ（τ）を算出して、算出した相互相関関数Ｒｘｙ（τ）を音声／非音声判定部１４に出力する。
音声／非音声判定部１４では、相互相関関数Ｒｘｙ（τ）と判定用しきい値δｒ１とを比較し、相互相関関数Ｒｘｙ（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１４は、その判定結果を音入力オン／

オフ制御部１５に出力する。

【００３３】

音入力オン／オフ制御部１５は、音声／非音声判定部１４が音声区間と判定した場合、オン制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力して、音声／非音声判定部１４が非音声区間と判定した場合、オフ制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力しないようにする。このとき、音入力オン／オフ制御部１５から出力される音声信号ｙ（ｔ）は、音源（ユーザ）からの音と雑音源からの音とからなる音声信号となる。

【００３４】

このように、音声信号処理装置１０は、第２マイク２への入力音中の発話区間（音声区間）を検出することができる。

例えば、第１マイク１，２と音声アプリケーションとの間にこのような音声信号処理装置１０を備えることで、音声アプリケーションは、確実に発話区間についての処理を行うことができる。ここで、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。例えば、音声アプリケーションが音声認識システムであるとすれば、音声認識システムは、音声信号処理装置１０が出力する発話区間の音声信号に基づいて音声認識できるようになる。

【００３５】

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、無指向性マイクである第２のマイク２で発話音及び雑音を受音し、単一指向性マイクである第１マイク１で雑音を受音し、第１マイク１で受音した雑音の音声信号と第２マイク２で受音した発話音及び雑音からなる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、発話音の発話区間を特定している。

【００３６】

これにより、第２のマイク２で発話音及び雑音を受音し、かつ第１マイク１で雑音を受音するように第１及び第２のマイク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受音系を構築することができる。

また、精度よく発話区間を検出することができる。そして、このように音声信号処理装置１０が検出した発話区間の音声信号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。

【００３７】

次に第２の実施形態を説明する。

この第２の実施形態も前述の第１の実施形態と同様、前記図１に示したように２つのマイク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。そして、音声信号処理装置１０の構成は、前述の第１の実施形態と同様、図３に示したような構成になる。しかし、第２の実施形態では、第１及び第２マイク１，２の配置を前述の第１の実施形態における配置と異ならせている。

前述の第１の実施形態では、第１マイク１に単一指向性マイクを用い、第２マイク２に無指向性マイクを用い、前記図２に示したように、雑音源からの音を第１及び第２マイク１，２で受音し、音源（ユーザ）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を配置している。

【００３８】

一方、この第２の実施形態では、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している。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４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４に示す点線は、第１マイク１の指向特定を示し、図４に示す一点鎖線は、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。

【００３９】

このように第１及び第２マイク１，２を配置した場合、前述の第１の実施形態と比較し、特に相互相関関数計算部１３で算出される相互相関関数Ｒｘｙ（τ）が異なる傾向を示すようになる。

すなわち、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２マイク２だけが受音しているので、雑音源からの音だけを第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、このとき相互相関関数Ｒｘｙ（τ）は０に近い値になる。一方、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に入力されるようになり、これにより、相互相関関数Ｒｘｙ（τ）は大きい値になる。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。

【００４０】

このように、音源（ユーザ）から発話があった場合には、相互相関関数Ｒｘｙ（τ）が大きくなり、第２の実施形態で得る相互相関関数Ｒｘｙ（τ）は、前述の第１の実施形態とは反対の傾向を示すようになる。

このようなことから、第２の実施形態では、音声／非音声判定部１４は、相互相関関数Ｒｘｙ（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Ｒｘｙ（τ）が判定用しきい値δｒ２より大きい場合（Ｒｘｙ（τ）＞δｒ２）、音声区間と判定し、それ以外の場合（Ｒｘｙ（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。

【００４１】

そして、前述の第１の実施形態と同様に、音声／非音声判定部１４が音声区間と判定した場合、音入力オン／オフ制御部１５は、オン制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力して、音声／非音声判定部３４が非音声区間と判定した場合、音入力オン／オフ制御部１５は、オフ制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力しないようにする。このとき、音入力オン／オフ制御部１５から出力される音声信号ｙ（ｔ）は、音源（ユーザ）からの音のみからなる音声信号となる。

このように、第２の実施形態の音声信号処理装置１０は、第２マイク２への入力音中の発話区間（音声区間）を検出することができる。

【００４２】

次に第２の実施形態における効果を説明する。

前述したように、無指向性マイクである第１のマイク１で発話音及び雑音を受音し、単一指向性マイクである第２マイク２で発話音を受音し、第１マイク１で受音した発話音及び雑音からなる音声信号と第２マイク２で受音した発話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、発話音の発話区間を特定している。

【００４３】

これにより、第１のマイク１で発話音及び雑音を受音し、かつ第２マイク２で発話音を受音するように第１及び第２のマイク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受音系を構築することができる。

また、精度よく発話区間を検出することができる。そして、このように音声信号処理装置１０が検出した発話区間の音声信号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。

【００４４】
次に第３の実施形態を説明する。
この第３の実施形態も前述の第１の実施形態と同様、前記図１に示したように２つのマイク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。
前述の第１及び第２の実施形態では、相互相関関数計算部１３により音声信号ｘ（ｔ），ｙ（ｔ）の相互相関関数Ｒｘｙ（τ）を算出し、この相互相関関数Ｒｘｙ（τ）に基づいて音声区間と非音声区間とを判定している。
【００４５】
これに対して、第３の実施形態の音声信号処理装置１０は、音声信号ｘ（ｔ），ｙ（ｔ）それぞれのパワースペクトルを算出し、このパワースペクトルに基づいて音声区間と非音声区間とを判定するように構成されている。図５は、その第３の実施形態の音声信号処理装置１０の構成を示す。
図５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１１，１２、第１及び第２パワースペクトラム計算部２１，２２、パワー比計算部２３、音声／非音声判定部２４並びに音入力オン／オフ制御部１５を備えている。各部の処理内容は次のようになる。
【００４６】
なお、第１及び第２フレーム化部１１，１２及び音入力オン／オフ制御部１５については、前述の第１の実施形態のものと同様な処理を行うので、その説明を省略する。また、前述の第１の実施形態と同様に、第１マイク１は単一指向性マイクであり、第２マイク２は無指向性マイクである。さらに、第１及び第２マイク１，２の配置についても、前記図２に示したような配置にしている。これにより、雑音源からの音を第１及び第２マイク１，２で受音し、音源（ユーザ）からの音を第２マイク２だけで受音している。
【００４７】
第１及び第２フレーム化部１１，１２それぞれで複数フレームにされた音声信号ｘ（ｔ），ｙ（ｔ）は第１及び第２パワースペクトラム計算部２１，２２に入力される。
第１パワースペクトラム計算部２１は、フレーム単位で音声信号ｘ（ｔ）の第１パワースペクトル値Ｐｘ（ω）を算出し、その算出した第１パワースペクトル値Ｐｘ（ω）をパワー比計算部２３に出力する。また、第２パワースペクトラム計算部２２は、フレーム単位で音声信号ｙ（ｔ）の第２パワースペクトル値Ｐｙ（ω）を算出し、その算出した第２パワースペクトル値Ｐｙ（ω）をパワー比計算部２３に出力する。
パワー比計算部２３は、下記（２）式により、第１パワースペクトラム計算部２１からの第１パワースペクトル値Ｐｘ（ω）と、第２パワースペクトラム計算部２２からの第２パワースペクトル値Ｐｙ（ω）との比（以下、パワー比という。）Ｐｘｙ（ω）を算出する。
【００４８】
【数２】

$$P_{xy}(\omega)=G_{xy}\frac{P_x(\omega)}{P_y(\omega)} \qquad \cdots (2)$$

【００４９】
ここで、Ｇｘｙ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。このように算出されたパワー比Ｐｘｙ（ω）はフレーム単位で各音声信号ｘ（ｔ），ｙ（ｔ）の波形形状の類似度を示す値となる。パワー比計算部２３は、このようなパワー比Ｐｘｙ（ω）を音声／非音声判定部２４に出力する。
音声／非音声判定部２４は、パワー比Ｐｘｙ（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、音源（ユーザ）と雑音源に対して前記図２のように第１及び第２マイク１，２を配置することで、雑音源からの音を第１及び第２マイク１，２で受音し、音源

（ユーザ）からの音を第２マイク２だけで受音している。

【００５０】

これにより、雑音源からの音だけを第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、このときに第１及び２パワースペクトラム計算部２１，２２で算出される第１及び第２パワースペクトル値Ｐｘ（ω），Ｐｙ（ω）は同程度になる。一方、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（ω）よりも第２パワースペクトル値Ｐｙ（ω）の方が大きくなる。このとき、パワー比計算部２３が算出するパワー比Ｐｘｙ（ω）は小さくなる。

【００５１】

なお、このとき、雑音源や音源（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（ω），Ｐｙ（ω）が特に変化する。

このように、音源（ユーザ）から発話があった場合にはパワー比Ｐｘｙ（ω）は小さくなることから、音声／非音声判定部２４は、パワー比Ｐｘｙ（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。

【００５２】

ここで、第１及び第２パワースペクトラム計算部２１，２２では、パワースペクトル値Ｐｘ（ω），Ｐｙ（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Ｐｘｙ（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（ω），Ｐｙ（ω）について各周波数で得ているパワー比Ｐｘｙ（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。

【００５３】

なお、判定対象としてパワースペクトル値Ｐｘ（ω），Ｐｙ（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、音源（ユーザ）の特性を示す特定の周波数帯のパワー比Ｐｘｙ（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Ｐｘｙ（ω）の平均値と判定用しきい値δｐ１とを比較したり、又は音源（ユーザ）の特性を示す特定の周波数帯のパワー比Ｐｘｙ（ω）と雑音源の特性を示す特定の周波数帯のパワー比Ｐｘｙ（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。

【００５４】

そして、音声／非音声判定部２４は、パワー比Ｐｘｙ（ω）が判定用しきい値δｐ１未満の場合（Ｐｘｙ（ω）＜δｐ１）、音声区間と判定し、それ以外の場合（Ｐｘｙ（ω）≧δｐ１）、非音声区間と判定する。ここで、音声／非音声判定部２４は、このような判定をフレーム単位で行う。そして、音声／非音声判定部２４は、その判定結果を音入力オン／オフ制御部１５に出力する。

【００５５】

以上のように第３の実施形態の音声信号処理装置１０が構成されている。この音声信号処理装置１０における一連の動作は次のようになる。

先ず、第１及び第２フレーム化部１１，１２が、第１及び第２マイク１，２から入力された２ｃｈの音声信号ｘ（ｔ），ｙ（ｔ）をそれぞれフレーム化し、フレーム単位で音声信号ｘ（ｔ），ｙ（ｔ）を第１及び第２パワースペクトラム計算部２１，２２に出力する。

【００５６】

パワースペクトラム計算部２１，２２ではそれぞれ、第１及び第２フレーム化部１１，１２それぞれから出力されるフレーム単位の音声信号ｘ（ｔ），ｙ（ｔ）について第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２３に出力する。

パワー比計算部２３では、パワースペクトラム計算部２１，２２それぞれから出力される第１及び第２パワースペクトル値Ｐｘ（ω），Ｐｙ（ω）について、フレーム単位でパワー比Ｐｘｙ（ω）を算出して、算出したパワー比Ｐｘｙ（ω）を音声／非音声判定部２４に出力する。

【００５７】

音声／非音声判定部２４では、パワー比Ｐｘｙ（ω）と判定用しきい値δｐ１とを比較し、パワー比Ｐｘｙ（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２４は、その判定結果を音入力オン／オフ制御部１５に出力する。

音入力オン／オフ制御部１５では、第２マイク２からの音声信号ｙ（ｔ）の後段への出力のオンとオフとを切り換える。具体的には、音声／非音声判定部２４が音声区間と判定した場合、音入力オン／オフ制御部１５は、オン制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力して、音声／非音声判定部２４が非音声区間と判定した場合、音入力オン／オフ制御部１５は、オフ制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力しないようにする。このとき、音入力オン／オフ制御部１５から出力される音声信号ｙ（ｔ）は、音源（ユーザ）からの音と雑音源からの音とからなる音声信号となる。

このように、第３の実施形態の音声信号処理装置１０は、第２マイク２への入力音中の発話区間（音声区間）を検出することができる。

【００５８】

次に第３の実施形態における効果を説明する。

前述したように、無指向性マイクである第２のマイク２で発話音及び雑音を受音し、単一指向性マイクである第１マイク１で雑音を受音し、第２マイク２で受音した発話音及び雑音からなる音声信号のパワースペクトルと、第１マイク１で受音した雑音の音声信号のパワースペクトルとを比較して、その比較結果に基づいて、前記発話音の発話区間を特定している。

【００５９】

これにより、第２のマイク２で発話音及び雑音を受音し、かつ第１マイク１で雑音を受音するように第１及び第２のマイク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受音系を構築することができる。

また、精度よく発話区間を検出することができる。そして、このように音声信号処理装置１０が検出した発話区間の音声信号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。

【００６０】

次に第４の実施形態を説明する。

この第４の実施形態も前述の第３の実施形態と同様、音声信号ｘ（ｔ），ｙ（ｔ）それぞれのパワースペクトルを算出し、このパワースペクトルに基づいて音声区間と非音声区間とを判定する音声信号処理装置１０である。音声信号処理装置１０の構成は、前述の第３の実施形態と同様、前記図５に示したような構成になる。そして、第４の実施形態では、第１及び第２マイク１，２の配置を前述の第３の実施形態における配置と異ならせている。

【００６１】

前述の第３の実施形態では、前述の第１の実施形態と同様、第１マイク１に単一指向性マイクを用い、第２マイク２に無指向性マイクを用い、前記図２に示したように、雑音源からの音を第１及び第２マイク１，２で受音し、音源（ユーザ）からの音を第２マイク２だけで受音するように、第１及び第２マイク１，２を配置している。

一方、この第４の実施形態では、前述の第２の実施形態と同様、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だけで受音している。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、前記図４に示したように、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。

【００６２】

このように第１及び第２マイク１，２を配置した場合、パワー比計算部２３で算出されるパワー比Ｐｘｙ（ω）は次のような傾向を示す。

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しているので、雑音源からの音だけを第１マイク１で受音している場合には、第１マイク１の受音感度の方が大きくなるので、第１パワースペクトル値Ｐｘ（ω）が大きくなり、これにより、パワー比Ｐｘｙ（ω）が大きくなる。一方、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、これにより、第２パワースペクトル値Ｐｙ（ω）も大きくなり、パワー比Ｐｘｙ（ω）が小さくなる。

【００６３】

このように、音源（ユーザ）から発話があった場合には、パワー比Ｐｘｙ（ω）が小さくなる傾向を示すようになる。

このようなことから、第４の実施形態では、音声／非音声判定部２４は、パワー比Ｐｘｙ（ω）と判定用しきい値（類似度を示すしきい値）δｐ２とを比較して、パワー比Ｐｘｙ（ω）が判定用しきい値δｐ２未満の場合（Ｐｘｙ（ω）＜δｐ２）、音声区間と判定し、それ以外の場合（Ｐｘｙ（ω）≧δｐ２）、非音声区間と判定する。ここで、判定用しきい値δｐ２は例えば実験により得る。

【００６４】

そして、前述の第３の実施形態と同様に、音入力オン／オフ制御部１５は、音声／非音声判定部２４が音声区間と判定した場合、オン制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力して、音声／非音声判定部２４が非音声区間と判定した場合、オフ制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力しないようにする。このとき、音入力オン／オフ制御部１５から出力される音声信号ｙ（ｔ）は、音源（ユーザ）からの音のみからなる音声信号となる。

【００６５】

このように、第４の実施形態の音声信号処理装置１０は、第２マイク２への入力音中の発話区間（音声区間）を検出することができる。

次に第４の実施形態における効果を説明する。

前述したように、無指向性マイクである第１のマイク１で発話音及び雑音を受音し、単一指向性マイクである第２マイク２で発話音を受音し、第１マイク１で受音した発話音及び雑音からなる音声信号のパワースペクトルと、第２マイク２で受音した発話音の音声信号のパワースペクトルとを比較して、その比較結果に基づいて、前記発話音の発話区間を特定している。

【００６６】

これにより、第１のマイク１で発話音及び雑音を受音し、かつ第２マイク２で発話音を受音するように第１及び第２のマイク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受音系を構築することができる。

また、精度よく発話区間を検出することができる。そして、このように音声信号処理装置１０が検出した発話区間の音声信号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。

【００６７】
次に第５の実施形態を説明する。
前述の第１及び第２の実施形態では、相互相関関数計算部１３により音声信号ｘ（ｔ），ｙ（ｔ）の相互相関関数Ｒｘｙ（τ）を算出し、この相互相関関数Ｒｘｙ（τ）に基づいて音声区間と非音声区間とを判定し、また、前述の第３及び第４の実施形態では、音声信号ｘ（ｔ），ｙ（ｔ）それぞれのパワースペクトルＰｘ（ω），Ｐｙ（ω）を算出し、このパワースペクトルＰｘ（ω），Ｐｙ（ω）（具体的にはパワー比Ｐｘｙ（ω））に基づいて音声区間と非音声区間とを判定している。第５の実施形態では、第１の実施形態（第２の実施形態）の処理と、第３の実施形態（第４の実施形態）の処理とを組み合わせた処理により、音声区間と非音声区間とを判定している。すなわち、第５の実施形態では、音声信号ｘ（ｔ），ｙ（ｔ）の相互相関関数Ｒｘｙ（τ）を算出するとともに、音声信号ｘ（ｔ），ｙ（ｔ）それぞれのパワースペクトルＰｘ（ω），Ｐｙ（ω）を算出し、相互相関関数Ｒｘｙ（τ）とパワースペクトルＰｘ（ω），Ｐｙ（ω）（具体的にはパワー比Ｐｘｙ（ω））との両面から音声区間と非音声区間とを判定している。図６は、それを実現する音声信号処理装置１０の構成を示す。
【００６８】
図６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１１，１２、相互相関関数計算部１３、音声／非音声判定部１４、音入力オン／オフ制御部１５、第１及び第２パワースペクトラム計算部２１，２２、パワー比計算部２３、並びに音声／非音声判定部３１を備えている。
このような構成において、第１及び第２フレーム化部１１，１２、相互相関関数計算部１３、音声／非音声判定部１４、音入力オン／オフ制御部１５、第１及び第２パワースペクトラム計算部２１，２２、並びにパワー比計算部２３は、前述の第１及び第２の実施形態と同様な処理を行う。
【００６９】
すなわち、第１及び第２フレーム化部１１，１２は、第１及び第２マイク１，２から入力された２ｃｈの音声信号ｘ（ｔ），ｙ（ｔ）をそれぞれフレーム化し、フレーム単位で音声信号ｘ（ｔ），ｙ（ｔ）を相互相関関数計算部１３に出力する。また、第１及び第２フレーム化部１１，１２はそれぞれ、フレーム単位で各音声信号ｘ（ｔ），ｙ（ｔ）を第１及び第２パワースペクトラム計算部２１，２２それぞれに出力する。
【００７０】
相互相関関数計算部１３は、第１及び第２フレーム化部１１，１２それぞれから出力されるフレーム単位の音声信号ｘ（ｔ），ｙ（ｔ）について相互相関関数Ｒｘｙ（τ）を算出して、算出した相互相関関数Ｒｘｙ（τ）を音声／非音声判定部３１に出力する。
また、第１及び第２パワースペクトラム計算部２１，２２は、第１及び第２フレーム化部１１，１２それぞれから出力されるフレーム単位の音声信号ｘ（ｔ），ｙ（ｔ）について第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２３に出力する。パワー比計算部２３は、パワースペクトラム計算部２１，２２それぞれから出力される第１及び第２パワースペクトル値Ｐｘ（ω），Ｐｙ（ω）について、フレーム単位でパワー比Ｐｘｙ（ω）を算出して、算出したパワー比Ｐｘｙ（ω）を音声／非音声判定部３１に出力する。
【００７１】
音声／非音声判定部３１では、次のような判定処理を行う。この音声／非音声判定部３１で行う処理が第５の実施形態において特有の処理になる。ここで、前記図２に示したように、雑音源からの音を第１及び第２マイク１，２で受音し、音源（ユーザ）からの音を第２マイク２だけで受音するように、第１及び第２マイク１，２を配置した場合（第１又は第３の実施形態の場合）と、前記図４に示したように、雑音源からの音を第１マイク１だけで受音し、音源（ユーザ）からの音を第１及び第２マイク２で受音するように、第１及び第２マイク１，２を配置した場合（第２又は第４の実施形態の場合）とを分けして説

明する。

【００７２】

先ず、前記図２に示したように、雑音源からの音を第１及び第２マイク１，２で受音し、音源（ユーザ）からの音を第２マイク２だけで受音するように、第１及び第２マイク１，２を配置した場合について説明する。

この場合、音声／非音声判定部３１は、音源（ユーザ）から発話があった場合に相互相関関数Ｒｘｙ（τ）が０に向かって減少することから、相互相関関数Ｒｘｙ（τ）が判定用しきい値δｒ１未満の場合（Ｒｘｙ（τ）＜δｒ１）、音声区間とし、それ以外の場合（Ｒｘｙ（τ）≧δｒ１）、非音声区間とする第１判定結果を得る。また、音声／非音声判定部３１は、音源（ユーザ）から発話があった場合にはパワー比Ｐｘｙ（ω）が小さくなることから、パワー比Ｐｘｙ（ω）が判定用しきい値δｐ１未満の場合（Ｐｘｙ（ω）＜δｐ１）、音声区間とし、それ以外の場合（Ｐｘｙ（ω）≧δｐ１）、非音声区間とする第２判定結果を得る。

【００７３】

そして、音声／非音声判定部３１は、前記第１及び第２判定結果に基づいて、音声区間の最終的な判定結果を得る。例えば、音声／非音声判定部３１は、第１及び第２判定結果が共に音声区間である判定結果となった場合、最終的な判定結果を音声区間とする。または、音声／非音声判定部３１は、第１判定結果又は第２判定結果の少なくとも一方が音声区間である判定結果となった場合、最終的な判定結果を音声区間とする。そして、音声／非音声判定部３１は、それ以外の場合、最終的な判定結果を非音声区間とする。

【００７４】

一方、前記図４に示したように、雑音源からの音を第１マイク１だけで受音し、音源（ユーザ）からの音を第１及び第２マイク２で受音するように、第１及び第２マイク１，２を配置した場合には、次のような判定を行う。

音声／非音声判定部３１は、音源（ユーザ）から発話があった場合に相互相関関数Ｒｘｙ（τ）が大きくなることから、相互相関関数Ｒｘｙ（τ）が判定用しきい値δｒ２より大きい場合（Ｒｘｙ（τ）≧δｒ２）、音声区間とし、それ以外の場合（Ｒｘｙ（τ）＜δｒ２）、非音声区間とする第１判定結果を得る。また、音声／非音声判定部３１は、音源（ユーザ）からの音の出力（発話）があった場合にはパワー比Ｐｘｙ（ω）が小さくなることから、パワー比Ｐｘｙ（ω）が判定用しきい値δｐ２未満の場合（Ｐｘｙ（ω）＜δｐ２）、音声区間とし、それ以外の場合（Ｐｘｙ（ω）≧δｐ２）、非音声区間とする第２判定結果を得る。

【００７５】

そして、音声／非音声判定部３１は、前記第１及び第２判定結果に基づいて、音声区間の最終的な判定結果を得る。例えば、音声／非音声判定部３１は、第１及び第２判定結果が共に音声区間である判定結果となった場合、最終的な判定結果を音声区間とする。または、音声／非音声判定部３１は、第１判定結果又は第２判定結果の少なくとも一方が音声区間である判定結果となった場合、最終的な判定結果を音声区間とする。そして、音声／非音声判定部３１は、それ以外の場合、最終的な判定結果を非音声区間とする。

【００７６】

以上のようにして音声／非音声判定部３１は、音声区間又は非音声区間を判定している。そして、音声／非音声判定部３１は、その判定結果を音入力オン／オフ制御部１５に出力する。

音入力オン／オフ制御部１５は、音声／非音声判定部３１が音声区間と判定した場合、オン制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力して、音声／非音声判定部３１が非音声区間と判定した場合、オフ制御として第２マイク２からの音声信号ｙ（ｔ）を後段に出力しないようにする。このとき、第１及び第２のマイク１，２の配置が前記図２に示した配置であれば、音入力オン／オフ制御部１５から出力される音声信号ｙ（ｔ）は、音源（ユーザ）からの音と雑音源からの音とからなる音声信号となる。また、第１及び第２のマイク１，２の配置が前記図４に示した配置であれば、音入力オン／オフ制

御部１５から出力される音声信号ｙ（ｔ）は、音源（ユーザ）からの音のみからなる音声信号となる。

【００７７】

次に第５の実施形態における効果を説明する。

前述したように、無指向性マイクで発話音及び雑音を受音し、単一指向性マイクで発話音又は雑音のいずれか一方を受音し、無指向性マイクで受音した発話音及び雑音からなる音声信号と単一指向性マイクで受音した発話音又は雑音のいずれか一方の音声信号の比較により相関度を得ている。その一方で、無指向性マイクで受音した発話音及び雑音からなる音声信号のパワースペクトルと、単一指向性マイクで受音した発話音又は雑音のいずれか一方の音声信号のパワースペクトルとを比較して、その比較結果としてパワー比を得ている。そして、前記相関度とパワー比との両方に基づいて、最終的に前記発話音の発話区間を特定している。

【００７８】

このように、相関度とパワー比との両方に基づいて最終的に前記発話音の発話区間を特定することで、その特定を精度よく行うことができる。そして、このように音声信号処理装置１０が検出した発話区間の音声信号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。

【００７９】

また、前述の第１乃至第４の実施形態と同様に、無指向性マイクで発話音及び雑音を受音し、かつ単一指向性マイクで発話音又は雑音のいずれか一方を受音するように第１及び第２のマイク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受音系を構築することができる。

【００８０】

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１１，１２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１１，１２に入力するようにする。また、前述の実施形態では、第２マイク２に入力された音声信号ｘ１（ｔ）を、音入力オン／オフ制御部１５に入力しているが、第２マイク２に入力され、フレーム化した音声信号ｘ１（ｔ）を音入力オン／オフ制御部１５に入力する。これらの仕様を、例えば前述の第１の実施形態の音声信号処理装置１０の構成に適用すると、図７に示すような構成になる。

【００８１】

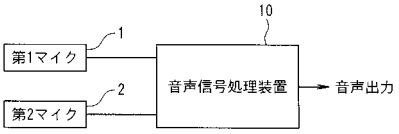
この図７に示すように、第１及び第２マイク１，２から入力された音声信号ｘ１（ｔ），ｘ２（ｔ）をそれぞれ、第１及び第２ＡＤ変換部４１，４２でＡＤ変換した後、第１及び第２フレーム化部１１，１２に入力する。また、第２ＡＤ変換部３２でＡＤ変換された信号は、第２フレーム化部１２でフレーム化されてから音入力オン／オフ制御部１５に入力される。ここで、第１及び第２ＡＤ変換部４１，４２でＡＤ変換されたデータ形式は、例えば１１０２５Ｈｚ、１６ｂｉｔ、リニアＰＣＭである。また、第１及び第２フレーム化部１１，１２でフレーム化された信号のフレーム長は、例えば５１２サンプルフレーム長である。

【００８２】

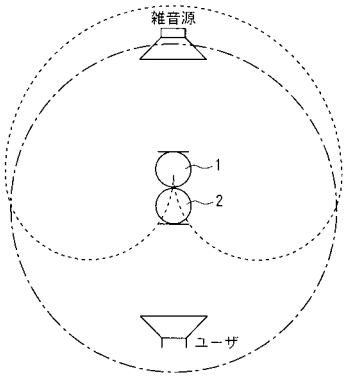
例えば、音声信号ｘ２（ｔ）を第２フレーム化部１２でフレーム化してから音入力オン／オフ制御部１５に出力することで、結果的に、音声信号処理装置１０から出力される音声信号ｘ２（ｔ）もフレーム化されているものとなり、これにより、音声信号処理装置１０から出力される音声信号ｘ２（ｔ）を利用する音声アプリケーションでは、解りやすいフレーム化された音声信号ｘ２（ｔ）で処理をすることができるようになる。

【００８３】
　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、相互相関関数計算部１３又はパワースペクトラム計算部２１，２２及びパワー比計算部２３は、無指向性マイクに入力された音信号と、単一指向性マイクに入力された音信号とを比較する比較手段を実現しており、音声／非音声判定部１４，２４，３１は、比較手段の比較結果に基づいて、検出対象音を検出する検出対象音検出手段又は発話音の発話区間を検出する発話区間検出手段を実現している。
【００８４】
　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。
【００８５】
　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、無指向性マイクで受音した発話音及び雑音の音声信号と単一指向性マイクで受音した前記発話音又は前記雑音のいずれか一方の音声信号とを比較し、その比較結果に基づいて、前記発話音の発話区間を検出する処理を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】
【００８６】
【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】本発明の第１の実施形態におけるマイクの配置を示す図である。
【図３】本発明の第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４】本発明の第２の実施形態におけるマイクの配置を示す図である。
【図５】本発明の第３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６】本発明の第５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７】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】
【００８７】
１，２　マイク
１０　音声信号処理装置
１１，１２　フレーム化部
１３　相互相関関数計算部
１４，２４，３１　音声／非音声判定部
１５　音入力オン／オフ制御部
２１，２２　パワースペクトラム計算部
２３　パワー比計算部

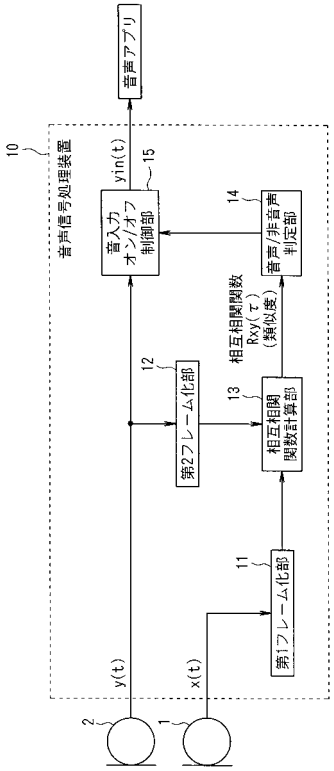
【図１】


10
1
第1マイク
2
第2マイク
音声信号処理装置
音声出力

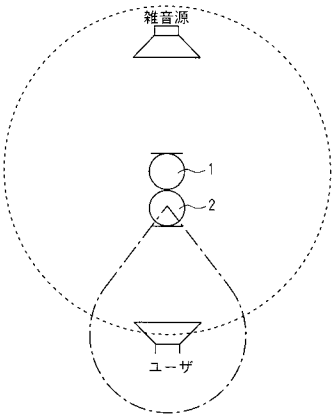
【図２】


雑音源
1
2
ユーザ

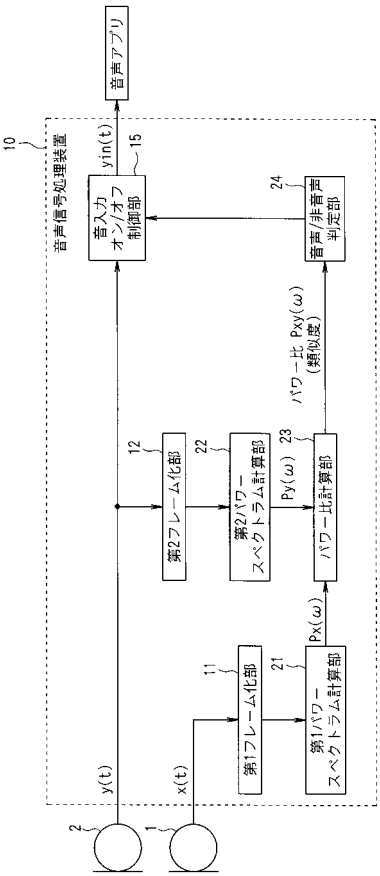
【図３】


音声アプリ
10
音声信号処理装置
yin(t)
15
音入力
オン/オフ
制御部
14
音声/非音声
判定部
相互相関関数
Rxy(τ)
(類似度)
12
第2フレーム化部
13
相互相関
関数計算部
11
第1フレーム化部
y(t)
x(t)
2
1

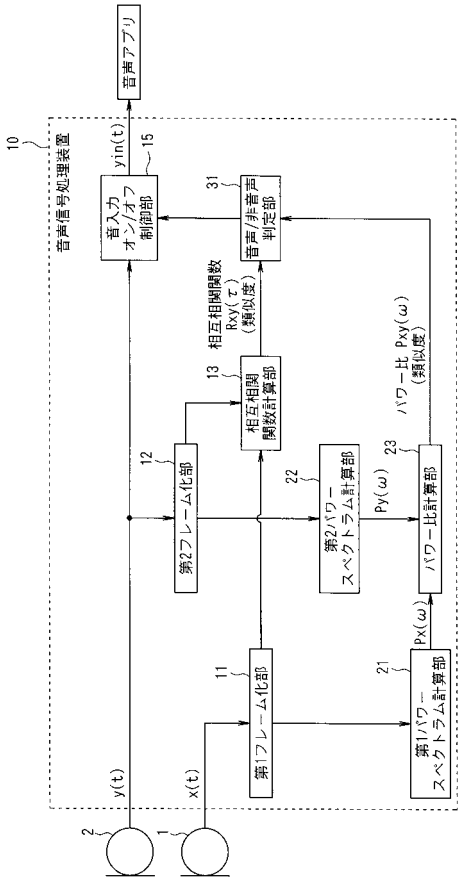
【図４】


雑音源
1
2
ユーザ

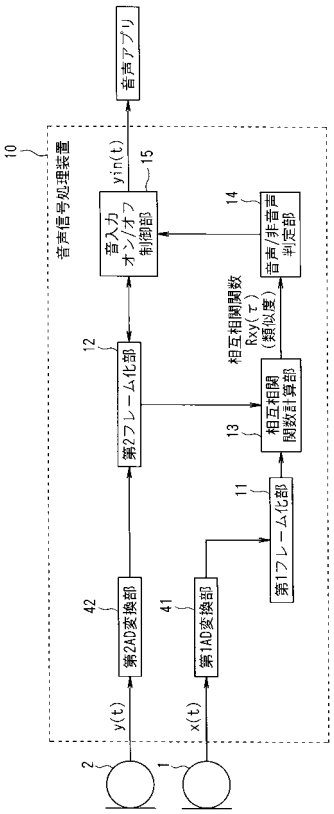
【図５】


音声アプリ
10
音声信号処理装置
yin(t)
15
音入力
オン/オフ
制御部
24
音声/非音声
判定部
パワー比 Pxy(ω)
(類似度)
12
第2フレーム化部
22
第2パワー
スペクトラム計算部
Py(ω)
23
パワー比計算部
Px(ω)
11
第1フレーム化部
21
第1パワー
スペクトラム計算部
y(t)
x(t)
2
1

【図６】


音声アプリ
10
音声信号処理装置
yin(t)
15
音入力
オン/オフ
制御部
31
音声/非音声
判定部
相互相関関数
Rxy(τ)
(類似度)
13
相互相関
関数計算部
12
第2フレーム化部
22
第2パワー
スペクトラム計算部
Py(ω)
23
パワー比 Pxy(ω)
(類似度)
パワー比計算部
Px(ω)
11
第1フレーム化部
21
第1パワー
スペクトラム計算部
y(t)
x(t)
2
1

【図７】


音声アプリ
10
音声信号処理装置
yin(t)
15
音入力
オン/オフ
制御部
14
音声/非音声
判定部
相互相関関数
Rxy(τ)
(類似度)
12
第2フレーム化部
13
相互相関
関数計算部
11
第1フレーム化部
42
第2AD変換部
41
第1AD変換部
y(t)
x(t)
2
1

フロントページの続き

(51)Int.Cl.[7]	FI	テーマコード（参考）
G10L 21/02		