

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3906230号
(P3906230)

(45) 発行日 平成19年4月18日(2007.4.18)

(24) 登録日 平成19年1月19日(2007.1.19)

(51) Int. Cl.	F I		
HO4R 1/40 (2006.01)	HO4R 1/40	310	
G1OL 15/28 (2006.01)	G1OL 15/28	400	
G1OL 21/02 (2006.01)	G1OL 21/02	201D	
HO4S 7/00 (2006.01)	HO4S 7/00	Z	

請求項の数 19 (全 31 頁)

<p>(21) 出願番号 特願2005-69824 (P2005-69824)</p> <p>(22) 出願日 平成17年3月11日(2005.3.11)</p> <p>(65) 公開番号 特開2006-254226 (P2006-254226A)</p> <p>(43) 公開日 平成18年9月21日(2006.9.21)</p> <p>審査請求日 平成17年3月25日(2005.3.25)</p> <p>(出願人による申告)平成16年度独立行政法人新エネルギー・産業技術総合開発機構「次世代ロボット実用化プロジェクト(プロトタイプ開発支援事業)人とのインタラクションにおける高性能聴覚機能ロボットの研究開発」委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p>	<p>(73) 特許権者 000003078 株式会社東芝 東京都港区芝浦一丁目1番1号</p> <p>(74) 代理人 100058479 弁理士 鈴江 武彦</p> <p>(74) 代理人 100091351 弁理士 河野 哲</p> <p>(74) 代理人 100088683 弁理士 中村 誠</p> <p>(74) 代理人 100108855 弁理士 蔵田 昌俊</p> <p>(74) 代理人 100075672 弁理士 峰 隆司</p> <p>(74) 代理人 100109830 弁理士 福原 淑弘</p>
--	---

最終頁に続く

(54) 【発明の名称】音響信号処理装置、音響信号処理方法、音響信号処理プログラム、及び音響信号処理プログラムを記録したコンピュータ読み取り可能な記録媒体

(57) 【特許請求の範囲】

【請求項1】

空間的に同一でない2以上の地点で捉えられた複数の音響信号を入力する音響信号入力手段と、

前記複数の音響信号の各々を分解し、周波数毎の位相値を表す複数の周波数分解データセットを求める周波数分解手段と、

前記複数の周波数分解データセットの異なる組において、周波数毎の位相差値を算出する位相差算出手段と、

前記組のそれぞれについて、周波数のスカラ倍を第1の軸とし、位相差値のスカラ倍を第2の軸とする2次元座標系上の座標値を有する点群を表す2次元データを生成する2次元データ化手段と、

前記2次元データ化手段により生成された座標値を有する点を、直線ハフ変換により投票空間に投票することにより得票分布を生成し、前記生成された得票分布から、得票が所定閾値以上となる極大位置を上位所定数まで検出することにより、同一音源に由来する周波数と位相差との間の比例関係を反映した直線を前記2次元データから検出する図形検出手段と、

前記直線のそれぞれを前記音響信号から抽出された音源と成し、前記音響信号の発生源に相当する音源の数、各音源の空間的な存在範囲、前記各音源の発した音声の時間的な存在期間、音源毎に分離された分離音声、各音源が発した音声の記号的内容の少なくともいずれかの情報と、各音源の発した音声の成分構成の情報とを含む、前記抽出された各音源

10

20

に関する音源情報を生成する音源情報生成手段と、

前記音源情報を出力する出力手段とを具備する音響信号処理装置であって、

前記音源情報生成手段は、前記図形検出手段により検出された直線について、周波数毎に、前記座標値と前記直線との距離を算出し、該距離に基づいて前記直線に対応する音源が発する音の周波数成分を推定する音源成分推定手段を具備したことを特徴とする音響信号処理装置。

【請求項 2】

前記図形検出手段は、前記投票空間に固定の値を投票し、前記 2 次元座標系において各周波数の点の多くを通る直線を検出する請求項 1 記載の音響信号処理装置。

【請求項 3】

前記周波数分解手段は、前記周波数毎の位相値のみならず周波数毎のパワー値をも計算し、前記図形検出手段は、前記パワー値に基づく数値を投票し、前記 2 次元座標系において各周波数のパワーの大きい点を多く通る直線を検出する請求項 1 記載の音響信号処理装置。

【請求項 4】

前記図形検出手段は、前記得票分布から所定閾値以上の得票を獲得した極大位置を検出するに際して、前記 2 次元座標系上の特定の位置を通過する直線に対応した前記投票空間上の位置についてのみ前記極大位置を求める請求項 1 記載の音響信号処理装置。

【請求項 5】

前記図形検出手段は、前記得票分布から所定閾値以上の得票を獲得した極大位置を検出するに際して、前記直線と傾きを同じくし、該傾きに応じて算出される一定の距離だけ離れた平行な直線群の、各直線に対応した前記投票空間上の得票の合計値を算出し、該合計値が所定閾値以上となる極大位置を求める請求項 1 記載の音響信号処理装置。

【請求項 6】

前記音源情報生成手段は、前記図形検出手段により検出された直線の傾きに基づいて、前記音源の空間的な存在範囲を、前記音響信号を捉えた 2 つの地点間を結ぶ線分に対する角度として算出する方向推定手段を具備する請求項 1 記載の音響信号処理装置。

【請求項 7】

前記音源情報生成手段は、前記図形検出手段により検出された直線について、周波数毎に、前記座標値と前記直線との距離を算出し、該距離に基づいて前記直線に対応する音源が発する音の周波数成分を推定する音源成分推定手段と、

推定された前記音の周波数成分から当該音源が発する音響信号データを合成する分離音声抽出手段とを具備する請求項 1 記載の音響信号処理装置。

【請求項 8】

前記音源成分推定手段は、前記座標値の前記直線に対する距離が所定閾値内である周波数を前記直線に対応する音源の発する音の周波数成分とする請求項 1 記載の音響信号処理装置。

【請求項 9】

前記音源成分推定手段は、前記座標値の前記直線に対する距離が所定閾値内である周波数を前記直線に対応する音源の発する音の周波数成分の候補とし、同一周波数成分について、最も近い方の直線に帰属させる請求項 1 記載の音響信号処理装置。

【請求項 10】

前記周波数分解手段は、前記周波数毎の位相値のみならず周波数毎のパワー値をも計算し、前記音源成分推定手段は、前記座標値の前記直線に対する距離の増加に応じて単調減少する非負係数を算出し、前記周波数のパワーに前記非負係数を乗じた値を前記直線に対応する音源の発する音の当該周波数成分のパワー値とする請求項 1 に記載の音響信号処理装置。

【請求項 11】

前記音源情報生成手段は、前記図形検出手段により検出された直線の傾きに基づいて、前記音源の空間的な存在範囲を、前記音響信号を捉えた 2 つの地点間を結ぶ線分に対する

10

20

30

40

50

角度として算出する方向推定手段と、

前記角度に基づいて音源方向に関する追従範囲を設定し、該追従範囲内に存在する音源からの音声のみを通過させることで当該音源が発する音の音響信号のデータを抽出する適応アレイ処理手段とを具備する請求項 1 記載の音響信号処理装置。

【請求項 1 2】

装置の動作に関する設定情報の確認と変更を利用者が行うためのユーザインタフェース手段を具備する請求項 1 記載の音響信号処理装置。

【請求項 1 3】

装置の動作に関する設定情報の保存と読み込みを利用者が行うためのユーザインタフェース手段を具備する請求項 1 記載の音響信号処理装置。

10

【請求項 1 4】

前記 2 次元データ、又は前記図形を利用者に呈示するためのユーザインタフェース手段を具備する請求項 1 記載の音響信号処理装置。

【請求項 1 5】

前記音源情報を利用者に呈示するためのユーザインタフェース手段を具備する請求項 1 記載の音響信号処理装置。

【請求項 1 6】

前記図形検出手段は、前記 2 次元データセットの時系列からなる 3 次元データセットから前記図形を検出する請求項 1 記載の音響信号処理装置。

【請求項 1 7】

空間的に同一でない 2 以上の地点で捉えられた複数の音響信号を入力する音響信号入力ステップと、

20

前記複数の音響信号の各々を分解し、周波数毎の位相値を表す複数の周波数分解データセットを求める周波数分解ステップと、

前記複数の周波数分解データセットの異なる組において、周波数毎の位相差値を算出する位相差算出ステップと、

前記組のそれぞれについて、周波数のスカラ倍を第 1 の軸とし、位相差値のスカラ倍を第 2 の軸とする 2 次元座標系上の座標値を有する点群を表す 2 次元データを生成する 2 次元データ化ステップと、

前記 2 次元データ化ステップにより生成された座標値を有する点を、直線ハフ変換により投票空間に投票することにより得票分布を生成し、前記生成された得票分布から、得票が所定閾値以上となる極大位置を上位所定数まで検出することにより、同一音源に由来する周波数と位相差との間の比例関係を反映した直線を前記 2 次元データから検出する図形検出ステップと、

30

前記直線のそれぞれを前記音響信号から抽出された音源と成し、前記音響信号の発生源に相当する音源の数、各音源の空間的な存在範囲、前記各音源の発した音声の時間的な存在期間、音源毎に分離された分離音声、各音源が発した音声の記号的内容の少なくともいずれかの情報と、各音源の発した音声の成分構成の情報とを含む、前記抽出された各音源に関する音源情報を生成する音源情報生成ステップと、

前記音源情報を出力する出力ステップとを具備する音響信号処理方法であって、

40

前記音源情報生成ステップは、前記図形検出ステップにより検出された直線について、周波数毎に、前記座標値と前記直線との距離を算出し、該距離に基づいて前記直線に対応する音源が発する音の周波数成分を推定する音源成分推定ステップを具備したことを特徴とする音響信号処理方法。

【請求項 1 8】

空間的に同一でない 2 以上の地点で捉えられた複数の音響信号を入力する音響信号入力手順と、

前記複数の音響信号の各々を分解し、周波数毎の位相値を表す複数の周波数分解データセットを求める周波数分解手順と、

前記複数の周波数分解データセットの異なる組において、周波数毎の位相差値を算出す

50

る位相差算出手順と、

前記組のそれぞれについて、周波数のスカラ倍を第1の軸とし、位相差値のスカラ倍を第2の軸とする2次元座標系上の座標値を有する点群を表す2次元データを生成する2次元データ化手順と、

前記2次元データ化手順により生成された座標値を有する点を、直線ハフ変換により投票空間に投票することにより得票分布を生成し、前記生成された得票分布から、得票が所定閾値以上となる極大位置を上位所定数まで検出することにより、同一音源に由来する周波数と位相差との間の比例関係を反映した直線を前記2次元データから検出する図形検出手順と、

前記直線のそれぞれを前記音響信号から抽出された音源と成し、前記音響信号の発生源に相当する音源の数、各音源の空間的な存在範囲、前記各音源の発した音声の時間的な存在期間、音源毎に分離された分離音声、各音源が発した音声の記号的内容の少なくともいずれかの情報と、各音源の発した音声の成分構成の情報とを含む、前記抽出された各音源に関する音源情報を生成する音源情報生成手順と、

前記音源情報を出力する出力手順とをコンピュータに実行させるための音響信号処理プログラムであって、

前記音源情報生成手順は、前記図形検出手順により検出された直線について、周波数毎に、前記座標値と前記直線との距離を算出し、該距離に基づいて前記直線に対応する音源が発する音の周波数成分を推定する音源成分推定手順を具備したことを特徴とする音響信号処理プログラム。

【請求項19】

請求項18に記載の音響信号処理プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は音響信号処理に関し、特に、媒質中を伝播してきた音波の発信源の数、各発信源の方向、各発信源から到来した音波の周波数成分などの推定に関する。

【背景技術】

【0002】

近年、ロボット用の聴覚研究の分野で、雑音環境下で複数の目的音源音の数とその方向を推定し（音源定位）、各音源音を分離抽出する（音源分離）方式が提案されている。

【0003】

例えば、下記非特許文献1には、背景雑音のある環境でN個の音源音をM個のマイクロホンで観測し、各マイクロホン出力を短時間フーリエ変換（FFT）処理したデータから空間相関行列を生成し、これを固有値分解して値の大きい主要な固有値を求めることで、この主要な固有値の数として音源の数Nを推定する方法が記載されている。これは、音源音のように方向性のある信号は主要な固有値にマッピングされ、方向性のない背景雑音は全ての固有値にマッピングされる性質を利用したものである。主要な固有値に対応する固有ベクトルは音源からの信号が張る信号部分空間の基底ベクトルとなり、残りの固有値に対応する固有ベクトルは背景雑音信号が張る雑音部分空間の基底ベクトルとなる。この雑音部分空間の基底ベクトルを利用してMUSIC法を適用することで、各音源の位置ベクトルを探索することができ、探索の結果得られた方向に指向性を与えられたビームフォーマで当該音源からの音声を抽出することができる。しかしながら、音源数Nがマイクロホン数Mと同数であるときは、雑音部分空間を定義できず、また、音源数NがMを越えるときは検出できない音源が存在することになる。したがって、推定可能な音源の数はマイクロホン数M以上となることはない。この方法は、音源について特に大きな制約はなく、数学的にもきれいな方法であるが、多数の音源を扱うためには、それを越える数のマイクロホンが必要になるという制限がある。

【0004】

10

20

30

40

50

また、下記非特許文献2には、一対のマイクロホンを用いて音源定位と音源分離を行う方法について記載されている。この方法は、人間の声のように管（調音器官）を通して発生される音声に特有の調波構造（基本周波数とその高調波とからなる周波数構造）に着目し、マイクロホンで捉えた音声信号をフーリエ変換したデータから基本周波数の異なる調波構造を検出することで、検出された調波構造の数を発話者の数とし、調波構造毎の両耳間位相差（IPD）と両耳間強度差（IID）とを用いてその方向を確信度付きで推定し、調波構造それ自体によって各音源音を推定する。この方法は、フーリエ変換データから複数の調波構造を検出することで、マイクロホン数以上の音源を処理することができる。しかしながら、音源数と方向と音源音の推定の基礎を調波構造に立脚して行うため、扱うことのできる音源は人間の声のような調波構造を持つものに限定され、さまざまな音に対応できるものではない。

10

【非特許文献1】浅野太、“音を分ける”、計測と制御、第43巻、第4号、pp. 325 - 330、2004年4月号

【非特許文献2】中臺一博ほか、“視聴覚情報の階層的統合による実時間アクティブ人物追跡”、人工知能学会AIチャレンジ研究会、SIG-Challenge-0113-5、pp. 35 - 42、June 2001

【発明の開示】

【発明が解決しようとする課題】

【0005】

以上で述べたように、(1)音源に制約を設けない場合は、音源数はマイクロホン数以上にはできない、(2)音源数をマイクロホン数以上にする場合、音源に例えば調波構造を仮定するなどの制約がある、という二律背反の問題あり、音源を制約せずにマイクロホン数以上の音源を扱うことのできる方式は確立されていない。

20

【0006】

本発明は上記の問題点に鑑みてなされたものであり、音源への制約をより緩和し、かつ、マイクロホン数以上の音源を扱うことのできる音源定位と音源分離のための音響信号処理装置、音響信号処理方法、音響信号処理プログラム、及び音響信号処理プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

【課題を解決するための手段】

【0007】

本発明の一観点に係る音響信号処理装置は、空間的に同一でない2以上の地点で捉えられた複数の音響信号を入力する音響信号入力手段と、前記複数の音響信号の各々を分解し、周波数毎の位相値を表す複数の周波数分解データセットを求める周波数分解手段と、前記複数の周波数分解データセットの異なる組において、周波数毎の位相差値を算出する位相差算出手段と、前記組のそれぞれについて、周波数の関数を第1の軸とし、前記位相差算出手段により算出された位相差値の関数を第2の軸とする2次元座標系上の座標値を有する点群を表す2次元データを生成する2次元データ化手段と、同一音源に由来する周波数と位相差との間の比例関係を反映した図形を前記2次元データから検出する図形検出手段と、前記音響信号の発生源に相当する音源の数、各音源の空間的な存在範囲、各音源が発した音声の時間的な存在期間、各音源が発した音声の成分構成、音源毎に分離された分離音声、各音源が発した音声の記号的内容の少なくともいずれかを含み、区別された音源に関する音源情報を前記図形に基づいて生成する音源情報生成手段と、前記音源情報を出力する出力手段とを具備する音響信号処理装置である。

30

40

【発明の効果】

【0008】

本発明によれば、音源への制約をより緩和し、かつ、マイクロホン数以上の音源を扱うことのできる音源定位と音源分離のための音響信号処理装置、音響信号処理方法、音響信号処理プログラム、及び音響信号処理プログラムを記録したコンピュータ読み取り可能な記録媒体を提供できる。

【発明を実施するための最良の形態】

50

【 0 0 0 9 】

以下、本発明に係る音響信号処理装置の実施形態を図面に従って説明する。

【 0 0 1 0 】

図1は本発明の一実施形態に係る音響信号処理装置の機能ブロック図である。この音響信号処理装置は、マイク1aと、マイク1bと、音響信号入力部2と、周波数分解部3と、2次元データ化部4と、図形検出部5と、音源情報生成部6と、出力部7と、ユーザインタフェース部8とを具備する。

【 0 0 1 1 】

[周波数成分毎の位相差に基づく音源推定の基本概念]

マイク1aとマイク1bは、空気などの媒質中に所定の距離をあけて配置された2つのマイクロホンであり、異なる2地点での媒質振動(音波)をそれぞれ電気信号(音響信号)に変換するための手段である。以後、マイク1aとマイク1bとをひとまとめに扱う場合、これをマイクロホン対と呼ぶことにする。

10

【 0 0 1 2 】

音響信号入力部2は、マイク1aとマイク1bによる2つの音響信号を所定のサンプリング周期 F_r で定期的にA/D変換することで、マイク1aとマイク1bによる2つの音響信号のデジタル化された振幅データを時系列的に生成する手段である。

【 0 0 1 3 】

マイクロホン間距離に比べて十分遠い場所に音源が位置していることを仮定すると、図2(a)に示すように、音源100から発してマイクロホン対に到達する音波の波面101はほぼ平面となる。マイク1aとマイク1bとを用いることにより異なる2地点でこの平面波を観測すると、マイク1aとマイク1bとを結ぶ線分102(これをベースラインと呼ぶ)に対する音源100の方向Rに応じて、マイクロホン対で変換される音響信号に所定の到達時間差 T が観測されるはずである。なお、音源が十分遠いとき、この到達時間差 T が0になるのは、音源100がベースライン102に垂直な平面上に存在するときであり、この方向をマイクロホン対の正面方向と定義する。

20

【 0 0 1 4 】

参考文献1「鈴木薫ほか、「視聴覚連携によるホームロボットの「呼べば来る」機能の実現」、第4回計測自動制御学会システムインテグレーション部門講演会(SI2003)講演論文集、2F4-5、2003」には、一方の振幅データのどの部分が他方の振幅データのどの部分と類似しているかをパターン照合によって探索することで、2つの音響信号(図2(b)の103と104)の間の到達時間差 T を導き出す方法について記載されている。しかしながら、この方法は強い音源が1つしか存在しない場合には有効であるが、強い背景雑音や複数の音源が存在する場合、複数方向からの強い音の混在した波形上に類似部分が明瞭に現れず、パターン照合に失敗することがある。

30

【 0 0 1 5 】

そこで、本発明では入力された振幅データを周波数成分毎の位相差に分解して解析する。複数の音源が存在する場合、各音源の周波数成分について、2つのデータ間でその音源方向に応じた位相差が観測される。そこで、もし音源についての強い制約を仮定することなく周波数成分毎の位相差を同方向のグループに分けることができれば、より幅広い種類の音源について、幾つの音源が存在し、その各々がどちらの方向にあり、それぞれが主に、どのように特徴的な周波数成分の音波を発しているかを把握できるはずである。この理屈自体は非常に単純明快であるが、実際のデータを分析する際には幾つかの克服すべき課題が存在している。その課題とともに、このグループ分けを行なうための機能ブロック(周波数分解部3、2次元データ化部4、図形検出部5)について引き続き説明する。

40

【 0 0 1 6 】

[周波数分解部3]

さて、振幅データを周波数成分に分解する一般的な手法として高速フーリエ変換(FFT)がある。代表的なアルゴリズムとしては、Cooley-Turkey DFTアルゴリズムなどが知られている。

50

【 0 0 1 7 】

周波数分解部 3 は、図 3 に示すように、音響信号入力部 2 による振幅データ 1 1 0 について、連続する N 個の振幅データをフレーム (T 番目のフレーム 1 1 1) として抜き出して高速フーリエ変換を行うとともに、この抜き出し位置をフレームシフト量 1 1 3 ずつずらしながら繰り返す (T + 1 番目のフレーム 1 1 2) 。

【 0 0 1 8 】

フレームを構成する振幅データは、図 4 (a) に示すように窓掛け 1 2 0 を施された後、高速フーリエ変換 1 2 1 がなされる。この結果、入力されたフレームの短時間フーリエ変換データが実部バッファ R [N] と虚部バッファ I [N] (1 2 2) に生成される。なお、窓掛け関数 (Hamming 窓掛けあるいは Hanning 窓掛け) 1 2 4 の一例を

10

【 0 0 1 9 】

ここで生成される短時間フーリエ変換データは、当該フレームの振幅データを N / 2 個の周波数成分に分解したデータであり、 k 番目の周波数成分 f k についてバッファ 1 2 2 内の実部 R [k] と虚部 I [k] の数値が、図 4 (c) に示すように複素座標系 1 2 3 上の点 P k を表す。この P k の原点 O からの距離の 2 乗が当該周波数成分のパワー P o (f k) であり、 P k の実部軸からの符号付き回転角度 { : - > [ラジアン] } が当該周波数成分の位相 P h (f k) である。

【 0 0 2 0 】

サンプリング周波数が F r [H z]、フレーム長が N [サンプル] のとき、 k は 0 から (N / 2) - 1 までの整数値をとり、 k = 0 が 0 [H z] (直流)、 k = (N / 2) - 1 が F r / 2 [H z] (最も高い周波数成分) を表し、その間を周波数分解能 f = (F r / 2) ÷ ((N / 2) - 1) [H z] で等分したものが各 k における周波数となり、 f k = k · f で表される。

20

【 0 0 2 1 】

なお、前述したように、周波数分解部 3 はこの処理を所定の間隔 (フレームシフト量 F s) を空けて連続的に行うことで、入力振幅データの周波数毎のパワー値と位相値とからなる周波数分解データセットを時系列的に生成する。

【 0 0 2 2 】

(2 次元データ化部 4 と図形検出部 5)

30

図 5 に示すように、2 次元データ化部 4 は位相差算出部 3 0 1 と座標値決定部 3 0 2 とを具備する。図形検出部 5 は投票部 3 0 3 と直線検出部 3 0 4 とを具備する。

【 0 0 2 3 】

[位相差算出部 3 0 1]

位相差算出部 3 0 1 は、周波数分解部 3 により得られた同時期の 2 つの周波数分解データセット a と b とを比較して、同じ周波数成分毎に両者の位相値の差を計算して得た a b 間位相差データを生成する手段である。例えば図 6 に示すように、ある周波数成分 f k の位相差 P h (f k) は、マイク 1 a における位相値 P h 1 (f k) とマイク 1 b における位相値 P h 2 (f k) との差を計算し、その値が { P h (f k) : - < P h (f k) } に収まるように、2 の剰余系として算定する。

40

【 0 0 2 4 】

[座標値決定部 3 0 2]

座標値決定部 3 0 2 は、位相差算出部 3 0 1 により得られた位相差データを元に、各周波数成分に両者の位相値の差を計算して得た位相差データを所定の 2 次元の X Y 座標系上の点として扱うための座標値を決定する手段である。ある周波数成分 f k の位相差 P h (f k) に対応する X 座標値 x (f k) と Y 座標値 y (f k) は、図 7 に示す式によって決定される。X 座標値は位相差 P h (f k)、Y 座標値は周波数成分番号 k である。

【 0 0 2 5 】

[同一時間差に対する位相差の周波数比例性]

位相差算出部 3 0 1 によって、図 6 に示したように算出される周波数成分毎の位相差は

50

、同一音源（同一方向）に由来するものどうしが同じ到達時間差を表しているはずである。FFTによって得られたある周波数の位相値及びマイクロホン間の位相差はその周波数の周期を2として算出された値である。ここで、同じ時間差であっても、周波数が2倍になれば位相差も2倍となるような比例関係の存在に着目する。これを図8に示す。図8(a)に示すように、同一時間Tについて、周波数 f_k [Hz]の波130は1/2周期、すなわちだけの位相区間を含むが、2倍の周波数 $2f_k$ [Hz]の波131は1周期、すなわち2の位相区間を含む。これは、位相差についても同様である。すなわち、同一時間差Tについて、位相差は周波数に比例して大きくなる。このような位相差と周波数との間の比例関係を図8(b)に示す。同一音源から発せられてTを共通にする各周波数成分の位相差を図7に示した座標値計算により2次元座標系上にプロットすると、各周波数成分の位相差を表す座標点132が直線133の上に並ぶことがわかる。Tが大きいほど、すなわち音源までの距離がマイクロホン間で異なるほど、この直線の傾きは大きくなる。

10

【0026】

[位相差の循環性]

但し、マイクロホン間の位相差が図8(b)に示したように全域で周波数に比例するのは、解析対象となる最低周波数から最高周波数まで通して真の位相差が \pm を逸脱しない場合に限られる。この条件はTが、最高周波数（サンプリング周波数の半分） $F_r/2$ [Hz]の1/2周期分の時間、すなわち $1/F_r$ [秒]以上としないことである。もし、Tが $1/F_r$ 以上となる場合には、次に述べるように位相差が循環性を持つ値としてしか得られないことを考慮しなければならない。

20

【0027】

周波数成分毎の位相値は図4に示した回転角度の値として2の幅（本実施形態ではからの間の2の幅）でしか得ることができない。これは、その周波数成分における実際の位相差がマイクロホン間で1周期以上であっても、周波数分解結果として得られる位相値からそれを知ることができないことを意味する。そのため、本実施形態では位相差を図6に示したようにからの間で得るようにしている。しかし、Tに起因する真の位相差は、ここで求められた位相差の値に2を加えたり差し引いたり、あるいはさらに4や6を加えたり差し引いたりした値である可能性がある。これを模式的に示すと図9のようになる。図9において、周波数 f_k の位相差 $Ph(f_k)$ が黒丸140で表すように+であるとき、1つ高い周波数 f_k+1 の位相差は白丸141で表すように+を超えている。しかしながら、計算された位相差 $Ph(f_k+1)$ は、本来の位相差から2を差し引いた、黒丸142で表すように-よりもやや大きい値となる。図示しないが、その3倍の周波数でも同様の値を示すことになるが、これは実際の位相差から4を差し引いた値である。このように位相差は周波数が高くなるにつれて2の剰余系として-からの間で循環する。この例のように、Tが大きくなると、ある周波数 f_k+1 から上では、白丸で表した真の位相差が黒丸で示したように反対側に循環してしまう。

30

【0028】

[複数音源存在時の位相差]

一方、複数の音源から音波が発せられている場合、周波数と位相差のプロット図は図10に模式的に示すような様相となる。この図は2つの音源がマイクロホン対に対して異なる方向に存在している場合を示したものであり、図10(a)は2つの音源音が互いに同じ周波数成分を含んでいない場合であり、図10(b)は一部の周波数成分が双方に含まれている場合である。図10(a)では、各周波数成分の位相差はTを共通にする直線のいずれかに乗っており、傾きの小さい直線150では5点、傾きの大きい直線151（循環した直線152を含む）では6点が直線上に配置されている。図10(b)では、双方に含まれる2つの周波数成分153と154では波が混ざって位相差が正しく出ないため、いずれの直線にも乗らず、特に傾きの小さい直線155では3点しか直線上に乗っていない。

40

50

【 0 0 2 9 】

音源の数と方向を推定する問題は、このようなプロット図上で、図示したような直線を発見することに帰着できる。また、音源毎の周波数成分を推定する問題は、検出された直線に近い位置に配置された周波数成分を選別することに帰着できる。本実施形態において、2次元データ化部4が出力する2次元データは、周波数分解部3による周波数分解データセットの2つを使って周波数と位相差の関数として決定される点群、もしくはそれら点群を2次元座標系上に配置(プロット)した画像とする。なお、この2次元データは時間軸を含まない2軸によって定義され、故に、2次元データの時系列としての3次元データが定義できる。図形検出部5はこの2次元データ(もしくはその時系列たる3次元データ)として与えられる点群配置から直線状の配置を図形として検出するものとする。

10

【 0 0 3 0 】

[投票部 3 0 3]

投票部303は、座標値決定部302によって(x, y)座標を与えられた各周波数成分に対して、後述するように直線ハフ変換を適用し、その軌跡をハフ投票空間に所定の方法で投票する手段である。ハフ変換については、参考文献2「岡崎彰夫、“はじめての画像処理”、工業調査会、2000年10月20日発行」の第100頁～第102頁に解説されているが、ここでもう一度説明する。

【 0 0 3 1 】

[直線ハフ変換]

図11に模式的に示すように、2次元座標上の点p(x, y)を通り得る直線は160、161、162に例示するように無数に存在するが、原点Oから各直線に下ろした垂線163のX軸からの傾きをとし、この垂線163の長さをとして表現すると、1つの直線についてとは一意に決まる。ある点(x, y)を通る直線の取り得るとの組は、座標系上で(x, y)の値に固有の軌跡164($= x \cos + y \sin$)を描くことが知られている。このような、(x, y)座標値から、(x, y)を通り得る直線の(,)の軌跡への変換を直線ハフ変換という。なお、直線が左に傾いているときは正值、垂直のときは0、右に傾いているときは負値であるとし、また、の定義域は{ : < }を逸脱することはない。

20

【 0 0 3 2 】

ハフ曲線はXY座標系上の各点について独立に求めることができるが、図12に示すように、例えば3点p1、p2、p3を共通に通る直線170は、p1、p2、p3に対応した軌跡171、172、173が交差する点174の座標(0, 0)で定められる直線として求めることができる。多くの点を通る直線であればあるほど、その直線を表すとの位置を多くの軌跡が通過する。このように、ハフ変換は点群から直線を検出する用途に向いている。

30

【 0 0 3 3 】

[ハフ投票]

点群から直線を検出するため、ハフ投票という工学的な手法が使われる。これはとを座標軸とする2次元のハフ投票空間に各軌跡の通過するとの組を投票することで、ハフ投票空間の得票の大きい位置に多数の軌跡の通過するとの組、すなわち直線の存在を示唆させるようにする手法である。一般的には、まずとについての必要な探索範囲分の大きさを持つ2次元の配列(ハフ投票空間)を用意して0で初期化しておく。次いで、点毎の軌跡をハフ変換によって求め、この軌跡が通過する配列上の値を1だけ加算する。これをハフ投票という。全ての点についてその軌跡を投票し終わると、得票0の位置(軌跡が1つも通過しなかった)には直線が存在せず、得票1の位置(軌跡が1つだけ通過した)には1つの点を通る直線が、得票2の位置(軌跡が2つだけ通過した)には2つの点を通る直線が、得票nの位置(軌跡がn個だけ通過した)にはn個の点を通る直線がそれぞれ存在することがわかる。ハフ投票空間の分解能を無限大にできれば、上述した通り、軌跡の通過する点のみが、そこを通過する軌跡の数だけの得票を得ることになるが、実際のハフ投票空間はとについて適当な分解能で量子化されているため、複数の軌跡

40

50

が交差する位置の周辺にも高い得票分布が生じる。そのため、ハフ投票空間の得票分布から極大値を持つ位置を探すことで、軌跡の交差する位置をより正確に求める必要がある。

【 0 0 3 4 】

投票部 3 0 3 は、次の投票条件を全て満たす周波数成分についてハフ投票を行う。この条件により、所定の周波数帯で所定閾値以上のパワーを持つ周波数成分のみが投票されることになる。

【 0 0 3 5 】

すなわち、投票条件 1 は、周波数が所定範囲にあるもの（低域カットと高域カット）とする。また、投票条件 2 は、当該周波数成分 f_k のパワー $P(f_k)$ が所定閾値以上のもの、とする。

10

【 0 0 3 6 】

投票条件 1 は、一般に暗騒音が乗っている低域をカットしたり、FFT の精度の落ちる高域をカットしたりする目的で使われる。この低域カットと高域カットの範囲は運用に合わせて調整可能である。最も広く周波数帯域を使う場合、低域カットは直流成分のみ、高域カットは最大周波数のみとする設定が適している。

【 0 0 3 7 】

暗騒音程度の非常に弱い周波数成分では FFT 結果の信頼性が高くないと考えられる。投票条件 2 は、このような信頼性の低い周波数成分をパワーで閾値処理することで投票に参加させないようにする目的で使われる。マイク 1 a におけるパワー値 $P_{o1}(f_k)$ 、マイク 1 b におけるパワー値 $P_{o2}(f_k)$ とすると、このとき評価されるパワー $P(f_k)$ の決め方には次の 3 つが考えられる。なお、いずれの条件を使用するかは運用に合わせて設定可能である。

20

【 0 0 3 8 】

（平均値）： $P_{o1}(f_k)$ と $P_{o2}(f_k)$ の平均値とする。両方のパワーがともに適度に強いことを必要とする条件である。

【 0 0 3 9 】

（最小値）： $P_{o1}(f_k)$ と $P_{o2}(f_k)$ の小さい方とする。両方のパワーが最低でも閾値以上あることを必要とする条件である。

【 0 0 4 0 】

（最大値）： $P_{o1}(f_k)$ と $P_{o2}(f_k)$ の大きい方とする。一方が閾値未満でも他方が十分強ければ投票するという条件である。

30

【 0 0 4 1 】

また、投票部 3 0 3 は、投票に際して次の 2 つの加算方式を行うことが可能である。

【 0 0 4 2 】

すなわち、加算方式 1 では、軌跡の通過位置に所定の固定値（例えば 1）を加算する。加算方式 2 では、軌跡の通過位置に当該周波数成分 f_k のパワー $P(f_k)$ の関数値を加算する。

【 0 0 4 3 】

加算方式 1 は、ハフ変換による直線検出問題で一般的によく用いられている方式であり、通過する点の多さに比例して得票に順位がつくため、多くの周波数成分を含む直線（すなわち音源）を優先的に検出するのに適している。このとき、直線に含まれる周波数成分について調波構造（含まれる周波数が等間隔であること）の制限がないので、人間の音声に限らずより幅広い種類の音源を検出することができる。

40

【 0 0 4 4 】

また、加算方式 2 は、通過する点が少なくても、パワーの大きい周波数成分を含んでいれば上位の極大値を得ることのできる方式であり、周波数成分が少なくてもパワーの大きい有力な成分を持つ直線（すなわち音源）を検出するのに適している。加算方式 2 におけるパワー $P(f_k)$ の関数値は $G(P(f_k))$ として計算される。図 1 3 は、 $P(f_k)$ を $P_{o1}(f_k)$ と $P_{o2}(f_k)$ の平均値とした場合の $G(P(f_k))$ の計算式を示したものである。この他にも上述した投票条件 2 と同様、 $P_{o1}(f_k)$ と $P_{o2}(f_k)$

50

k) の最小値や最大値として $P(f_k)$ を計算することも可能であり、投票条件 2 とは別に運用に合わせて設定可能である。中間パラメータ V の値は $P(f_k)$ の対数值 $\log_{10}(P(f_k))$ に所定のオフセットを足した値として計算される。そして V が正であるときは $V + 1$ の値を、 V がゼロ以下であるときには 1 を、関数 $G(P(f_k))$ の値とする。このように最低でも 1 を投票することで、パワーの大きい周波数成分を含む直線(音源)が上位に浮上するだけでなく、多数の周波数成分を含む直線(音源)も上位に浮上するという加算方式 1 の多数決的な性質を併せ持たせることができる。投票部 303 は、設定によって加算方式 1 と加算方式 2 のいずれを行うことも可能であるが、特に後者を用いることで、周波数成分の少ない音源も同時に検出することが可能になり、さらに幅広い種類の音源を検出できるようになる。

10

【0045】

[複数 FFT 結果をまとめて投票]

さらに、投票部 303 は、1 回の FFT 毎に投票を行うことも可能だが、一般的に連続する m 回 ($m \geq 1$) の時系列的な FFT 結果についてまとめて投票を行なうこととする。長期的には音源の周波数成分は変動するものであるが、このようにすることで、周波数成分の安定している適度に短期間の複数次刻の FFT 結果から得られるより多くのデータを用いて、より信頼性の高いハフ投票結果を得ることができるようになる。なお、この m は運用に合わせてパラメータとして設定可能とする。

【0046】

[直線検出部 304]

直線検出部 304 は、投票部 303 によって生成されたハフ投票空間上の得票分布を解析して有力な直線を検出する手段である。但し、このとき、図 9 で述べた位相差の循環性など、本問題に特有の事情を考慮することで、より高精度な直線検出を実現する。

20

【0047】

図 14 に、室内雑音環境下で 1 人の人物がマイクロホン対の正面約 20 度左より発話した実際の音声を用いて処理したときの周波数成分のパワースペクトル、連続する 5 回分 (前述の $m = 5$) の FFT 結果から得た周波数成分毎の位相差プロット図、同じ 5 回分の FFT 結果から得たハフ投票結果 (得票分布) を示す。ここまでの処理は音響信号入力部 2 から投票部 303 までの一連の機能ブロックで実行される。

【0048】

マイクロホン対で取得された振幅データは、周波数分解部 3 によって周波数成分毎のパワー値と位相値のデータに変換される。図 14 において、180 と 181 は、横軸を時間として、周波数成分毎のパワー値の対数を輝度表示 (黒いほど大きい) したものである。縦の 1 ラインが 1 回の FFT 結果に対応し、これを時間経過 (右向き) に沿ってグラフ化した図である。上段 180 がマイク 1a、下段 181 がマイク 1b からの信号を処理した結果であり、多数の周波数成分が検出されている。この周波数分解結果を受けて、位相差算出部 301 により周波数成分毎の位相差が求められ、座標値決定部 302 によりその (x, y) 座標値が算出される。図 14 において、182 はある時刻 183 から連続 5 回分の FFT によって得た位相差をプロットした図である。この図で原点から左に傾いた直線 184 に沿う点群分布が認められるが、その分布は直線 184 上にきれいに乗っているわけではなく、またこの直線 184 から離れた多数の点が存在している。投票部 303 により、このような分布を示している各点がハフ投票空間に投票されて得票分布 185 を形成する。なお、図の 185 は加算方式 2 を用いて生成された得票分布である。

30

40

【0049】

[$\theta = 0$ の制約]

ところで、マイク 1a とマイク 1b の信号が音響信号入力部 2 によって同相で A/D 変換される場合、検出されるべき直線は必ず $\theta = 0$ 、すなわち XY 座標系の原点を通る。したがって、音源の推定問題は、ハフ投票空間上で $\theta = 0$ となる θ 軸上の得票分布 $S(\theta, 0)$ から極大値を探索する問題に帰着する。図 14 で例示したデータに対して θ 軸上で極大値を探索した結果を図 15 に示す。

50

【0050】

図15において、得票分布190は図13における得票分布185と同一のものである。棒グラフ192は、軸191上の得票分布 $S(x, 0)$ を $H(x)$ として抜き出したものである。この得票分布 $H(x)$ には幾つか極大箇所(突出部)が存在している。直線検出部304は、得票分布 $H(x)$ に対して、(1)ある位置について左右に自身と同点のものが続く限り探索したときに、最後に自身より低得票のものだけが現れた箇所を残す。この結果、得票分布 $H(x)$ 上の極大部が抽出されるが、この極大部には平坦な頂を持つものが含まれるので、そこでは極大値が連続する。そこで直線検出部304は、(2)細線化処理によって極大部の中央位置だけを極大位置193として残す。そして最後に、(3)得票が所定閾値以上となる極大位置のみを直線として検出する。このようにすることで十分な得票を得た直線の傾きを正確に割り出すことができる。図の例では、上記(2)において検出された極大位置194、195、196のうち、極大位置194が平坦な極大部から細線化処理によって残された中央位置(偶数連続時は右が優先)である。また、唯一196だけが閾値以上の得票を得て検出された直線となる。この極大位置196によって与えられる傾き θ ($\theta = 0$)とにより、直線(基準直線)197は定義される。なお、細線化処理のアルゴリズムは、ハフ変換の説明で参照した参考文献2の第89頁~第92頁に記載される「田村の方法」を1次元化して使うことが可能である。直線検出部304は、このようにして1乃至複数の極大位置(所定閾値以上の得票を得た中央位置)を検出すると、その得票の多い順に順位を付けて各極大位置の傾き θ と θ の値を出力する。

10

【0051】

[位相差循環を考慮した直線群の定義]

ところで、図15で例示した直線197は、傾き $\theta = 0$ なる極大位置196によって定義されたXY座標原点を通る直線である。しかし、実際には位相差の循環性によって、図15の直線197が傾き $\theta = 199$ だけ平行移動してX軸上の反対側から循環してくる直線198もまた197と同じ到達時間差を示す直線である。この直線198のように直線197を延長してXの値域からはみ出した部分が反対側から循環的に現れる直線を、直線197の「循環延長線」、基準となった直線197を「基準直線」とそれぞれ呼ぶことにする。もし、基準直線197がさらに傾いておれば、循環延長線はさらに数を増すことになる。ここで係数 a を0以上の整数とすると、到達時間差を同じくする直線は全て傾き $\theta = 0$ で定義される基準直線197を傾き $\theta = a$ ずつ平行移動させた直線群(傾き $\theta = 0, a$)となる。さらに、起点となる傾き $\theta = 0$ の制約をはずして傾き $\theta = 0$ として一般化すると、直線群は傾き $\theta = 0, a + 0$ として記述できることになる。このとき、傾き θ は直線の傾き θ の関数 $f(\theta)$ として図16に示す式で定義される符号付きの値である。

20

30

【0052】

図16において、基準直線200は傾き $\theta = 0$ で定義される。このとき、基準直線200が右に傾いているので定義に従い傾き θ は負値であるが、図ではその絶対値として扱う。図16における直線201は基準直線200の循環延長線であり、点RにおいてX軸と交差している。また、基準直線200と循環延長線201の間隔は補助線202で示す通りであり、補助線202は点Oにおいて基準直線200と垂直に交差し、点Uにおいて循環延長線201と垂直に交差している。このとき、基準直線が右に傾いているので定義に従い傾き θ も負値であるが、図ではその絶対値として扱う。図16における $\triangle OQP$ は辺OQの長さが 2 となる直角三角形であり、これと合同な三角形が $\triangle RTS$ である。故に辺RTの長さも 2 であり、 $\triangle OUR$ の斜辺ORの長さが 2 であることがわかる。このとき、傾き θ は辺OUの長さであるから、 $\cos \theta = 2 / 2 = 1$ となる。そして、傾き θ の符号を考慮すると図の計算式が導き出される。

40

【0053】

[位相差循環を考慮した極大位置検出]

位相差の循環性から、音源を表す直線は1つではなく基準直線と循環延長線とからなる直線群として扱われるべきであることを述べた。このことは得票分布から極大位置を検出する際にも考慮されなければならない。通常、位相差の循環が起きないか、起きても小規

50

模で収まるマイクロホン対の正面付近のみで音源を検出する場合に限れば、 $\theta = 0$ （あるいは $\theta = 0$ ）上の得票値（すなわち基準直線の得票値）のみで極大位置を探索する上述の方法は性能的に十分であるばかりか、探索時間の短縮と精度の向上に効果がある。しかし、より広い範囲に存在する音源を検出しようとする場合には、ある θ について ずつ離れた数箇所の得票値を合計して極大位置を探索する必要がある。この違いを以下で説明する。

【0054】

図17に、室内雑音環境下で2人の人物がマイクロホン対の正面約20度左と約45度右から同時に発話した実際の音声を用いて処理したときの周波数成分のパワースペクトル、5回分（ $m = 5$ ）のFFT結果から得た周波数成分毎の位相差プロット図、同じ5回分のFFT結果から得たハフ投票結果（得票分布）を示す。

10

【0055】

マイクロホン対で取得された振幅データは、周波数分解部3によって周波数成分毎のパワー値と位相値のデータに変換される。図17において、210と211は、縦軸を周波数、横軸を時間として、周波数成分毎のパワー値の対数を輝度表示（黒いほど大きい）したものである。縦の1ラインが1回のFFT結果に対応し、これを時間経過（右向き）に沿ってグラフ化した図である。上段210がマイク1a、下段211がマイク1bからの信号を処理した結果であり、多数の周波数成分が検出されている。この周波数分解結果を受けて、位相差算出部301により周波数成分毎の位相差が求められ、座標値決定部302によりその（ x, y ）座標値が算出される。プロット図212は、ある時刻213から連続5回分のFFTによって得た位相差をプロットしものである。このプロット図212において原点から左に傾いた基準直線214に沿う点群分布と右に傾いた基準直線215に沿う点群分布が認められる。投票部303により、このような分布を示している各点がハフ投票空間に投票されて得票分布216を形成する。なお、得票分布216は、加算方式2を用いて生成されたものである。

20

【0056】

図18は θ 軸上の得票値のみで極大位置を探索した結果を示した図である。図18における得票分布220は図17における得票分布216と同一のものである。棒グラフ222は、 θ 軸221上の得票分布 $S(\theta, 0)$ を $H(\theta)$ として抜き出して棒グラフにしたものである。この得票分布 $H(\theta)$ には幾つか極大箇所（突出部）が存在しているが、総じて θ の絶対値が大きくなるほど得票が少なくなることがわかる。この得票分布 $H(\theta)$ からは、極大位置グラフ223に示すように4つの極大位置224、225、226、227が検出される。このうち、唯一、極大位置227だけが閾値以上の得票を得る。これにより1つの直線群（基準直線228と循環延長線229）が検出される。この直線群はマイクロホン対の正面約20度左からの音源を検出したものであるが、マイクロホン対の正面約45度右からの音源は検出できていない。原点を通る基準直線ではその角度が大きいほど X の値域を超えるまでに少ない周波数帯しか通過できないため、基準直線が通過する周波数帯の広さは θ によって異なる（不公平がある）。そして、 $\theta = 0$ の制約は、この不公平な条件で基準直線だけの得票を競わせることになるため、角度の大きい直線ほど得票で不利になるのである。これが約45度右からの音源を検出できなかった理由である。

30

40

【0057】

一方、図19は ずつ離れた数箇所の得票値を合計して極大位置を探索した結果を示した図である。図中の240は、図17における得票分布216上に、原点を通る直線を ずつ平行移動させたときの θ の位置を破線242～249で表示したものである。このとき、 θ 軸241と破線242～245、及び θ 軸241と破線246～249はそれぞれ θ （ θ ）の自然数倍で等間隔に離れている。なお、直線が X の値域を越えずにプロット図の天井まで抜けることが確実な $\theta = 0$ には破線が存在しない。

【0058】

ある $\theta = 0$ の得票 $H(\theta = 0)$ は、 $\theta = 0$ の位置で縦に見たときの θ 軸241上の得票と破線242～249上の得票の合計値、すなわち $H(\theta = 0) = \{ S(\theta = 0, a$ （

50

0) }として計算される。この操作は $\theta = 0$ となる基準直線とその循環延長線の得票を合計することに相当する。この得票分布 $H(\theta)$ を棒グラフにしたものが図中の 250 である。図 18 の 222 と異なり、この分布では θ の絶対値が大きくなっても得票が少なくなっていない。これは、得票計算に循環延長線を加えたことで全ての θ について同じ周波数帯を使うことができるようになったからである。この得票分布 250 からは図中 251 に示す 10 個の極大位置が検出される。このうち、極大位置 252 と 253 が閾値以上の得票を得て、マイクロホン対の正面約 20 度左からの音声を検出した直線群 (極大位置 253 に対応する基準直線 254 と循環延長線 255) と、マイクロホン対の正面約 45 度右からの音声を検出した直線群 (極大位置 252 に対応する基準直線 256 と循環延長線 257 と 258) の 2 つが検出される。このように θ ずつ離れた箇所の得票値を合計して極大位置を探索することで、角度の小さい直線から角度の大きい直線まで安定に検出できるようになる。

10

【0059】

[非同相の場合を考慮した極大位置検出：一般化]

さて、マイク 1 a とマイク 1 b の信号が音響信号入力部 2 によって同相で A/D 変換されない場合、検出されるべき直線は $\theta = 0$ 、すなわち XY 座標原点を通らない。この場合は $\theta = 0$ の制約をはずして極大位置を探索する必要がある。

【0060】

$\theta = 0$ の制約をはずした基準直線を一般化して (θ_0, θ_0) と記述すると、その直線群 (基準直線と循環延長線) は $(\theta_0, a \cos(\theta - \theta_0) + \theta_0)$ と記述できる。ここで (θ_0) は θ_0 によって決まる循環延長線の平行移動量である。音源がある方向から来るときに、それに対応した θ_0 における直線群は最も有力なものが 1 つ存在するだけである。その直線群は、様々に θ_0 を変えたときの直線群の得票 $\{S(\theta_0, a \cos(\theta - \theta_0) + \theta_0)\}$ が最大となる θ_0 の値 $\theta_{0 \max}$ を使って $(\theta_0, a \cos(\theta - \theta_0) + \theta_{0 \max})$ で与えられる。そこで、各 θ_0 における得票 $H(\theta_0)$ をそれぞれの θ_0 における最大得票値 $\{S(\theta_0, a \cos(\theta - \theta_0) + \theta_{0 \max})\}$ とすることで、 $\theta = 0$ の制約時と同じ極大位置検出アルゴリズムを適用した直線検出を行なうことができるようになる。

20

【0061】

なお、このようにして検出された直線群の数が音源の数である。

【0062】

[音源情報生成部 6]

図 20 に示すように、音源情報生成部 6 は、方向推定部 311 と、音源成分推定部 312 と、音源音再合成部 313 と、時系列追跡部 314 と、継続時間評価部 315 と、同相化部 316 と、適応アレイ処理部 317 と、音声認識部 318 とを具備する。

30

【0063】

[方向推定部 311]

方向推定部 311 は、以上で述べた直線検出部 304 による直線検出結果、すなわち直線群毎の θ 値を受けて、各直線群に対応した音源の存在範囲を計算する手段である。このとき、検出された直線群の数が音源の数 (全候補) となる。マイクロホン対のベースラインに対して音源までの距離が十分遠い場合、音源の存在範囲はマイクロホン対のベースラインに対してある角度を持った円錐面となる。これを図 21 を参照して説明する。

40

【0064】

マイク 1 a とマイク 1 b の到達時間差 T は $\pm T_{\max}$ の範囲で変化し得る。図 21 (a) のように、正面から入射する場合、 T は 0 となり、音源の方位角 θ は正面を基準にした場合 0° となる。また、図 21 (b) のように音声が右真横、すなわちマイク 1 b 方向から入射する場合、 T は $+ T_{\max}$ に等しく、音源の方位角 θ は正面を基準にして右回りを正として $+90^\circ$ となる。同様に、図 21 (c) のように音声が左真横、すなわちマイク 1 a 方向から入射する場合、 T は $- T_{\max}$ に等しく、方位角 θ は -90° となる。このように、 T を音が右から入射するとき正、左から入射するとき負となるように定義する。

50

【0065】

以上を踏まえて図21(d)のような一般的な条件を考える。マイク1aの位置をA、マイク1bの位置をBとし、音声線分PA方向から入射すると仮定すると、PABは頂点Pが直角となる直角三角形となる。このとき、マイク間中心O、線分OCをマイクロホン対の正面方向として、OC方向を方位角 0° とした左回りを正にとる角度を方位角と定義する。QOBはPABの相似形となるので、方位角の絶対値はOBQ、すなわちABPに等しく、符号はTの符号に一致する。また、ABPはPAとABの比の \sin^{-1} として計算可能である。このとき、線分PAの長さをこれに相当するTで表すと、線分ABの長さは T_{max} に相当する。したがって、符号も含めて、方位角は $=\sin^{-1}(T/T_{max})$ として計算することができる。そして、音源の存在範囲は点Oを頂点、ベースラインABを軸として、 $(90 -)^\circ$ 開いた円錐面 260 として推定される。音源はこの円錐面 260 上のどこかにある。

10

【0066】

図22に示すように、 T_{max} はマイク間距離 $L[m]$ を音速 $V_s[m/sec]$ で割った値である。このとき、音速 V_s は気温 $t[]$ の関数として近似できることが知られている。今、直線検出部304によって直線270がハフの傾きで検出されているとする。この直線270は右に傾いているのでは負値である。 $y = k$ (周波数 f_k)のとき、直線270で示される位相差 Ph は k との関数として $k \cdot \tan(-)$ で求めることができる。このとき $T[sec]$ は、位相差 $Ph(, k)$ の2に対する割合を、周波数 f_k の1周期 $(1/f_k)[sec]$ に乗じた時間となる。が符号付きの量なので、 T も符号付きの量となる。すなわち、図21(d)で音が右から入射する(位相差 Ph が正值となる)とき、は負値となる。また、図21(d)で音が左から入射する(位相差 Ph が負値となる)とき、は正值となる。そのために、の符号を反転させている。なお、実際の計算においては、 $k = 1$ (直流成分 $k = 0$ のすぐ上の周波数)で計算を行えば良い。

20

【0067】

[音源成分推定部312]

音源成分推定部312は、座標値決定部302により与えられた周波数成分毎の (x, y) 座標値と、直線検出部304により検出された直線との距離を評価することで、直線近傍に位置する点(すなわち周波数成分)を当該直線(すなわち音源)の周波数成分として検出し、この検出結果に基づいて音源毎の周波数成分を推定するための手段である。

30

【0068】

[距離閾値方式による検出]

図23に複数の音源が存在するときの音源成分推定の原理を模式的に示す。図23(a)は図9に示したものと同一周波数と位相差のプロット図であり、2つの音源がマイクロホン対に対して異なる方向に存在している場合を示している。図23(a)の280は1つの直線群をなし、図23(a)の281と282は別の直線群をなす。図23(a)の黒丸は周波数成分毎の位相差位置を表している。

【0069】

直線群280に対応する音源音を構成する周波数成分は、図23(b)に示すように、直線280から左右にそれぞれ水平距離283だけ離れた直線284と直線285に挟まれる領域286内に位置する周波数成分(図の黒丸)として検出される。ある周波数成分がある直線の成分として検出されることを、周波数成分が直線に帰属する(あるいは属する)ということにする。

40

【0070】

同様に、直線群(281、282)に対応する音源音を構成する周波数成分は、図23(c)に示すように、直線281と直線282から左右にそれぞれ水平距離283だけ離れた直線に挟まれる領域287と288内に位置する周波数成分(図の黒丸)として検出される。

【0071】

50

なお、このとき周波数成分 289 と原点（直流成分）の 2 点は、領域 286 と領域 288 の両方に含まれるので、両音源の成分として二重に検出される（多重帰属）。このように、周波数成分と直線との水平距離を閾値処理して、直線群（音源）毎に閾値内に存在する周波数成分を選択し、そのパワーと位相をそのまま当該音源音の成分とする方式を「距離閾値方式」と呼ぶことにする。

【0072】

[最近傍方式による検出]

図 24 は、図 23 において多重帰属する周波数成分 289 について、どちらか最も近い方の直線群にのみ属させるようにした結果を示した図である。直線 280 と直線 282 に対する周波数成分 289 の水平距離を比較した結果、周波数成分 289 は直線 282 に最も近いことが判明する。このとき、周波数成分 289 は直線 282 近傍の領域 288 に入っている。よって、周波数成分 289 は、図 24 (b) に示すように直線群 (281、282) に属する成分として検出される。このように、周波数成分毎に水平距離で最も近い直線（音源）を選択し、その水平距離が所定閾値内にある場合に当該周波数成分のパワーと位相をそのまま当該音源音の成分とする方式を「最近傍方式」と呼ぶことにする。なお、直流成分（原点）は特別扱いとして両方の直線群（音源）に帰属させるものとする。

10

【0073】

[距離係数方式による検出]

上記 2 つの方式は、直線群を構成する直線に対して所定の水平距離閾値内に存在する周波数成分だけを選択し、そのパワーと位相をそのままにして直線群に対応する音源音の周波数成分としていた。一方、次に述べる「距離係数方式」は、周波数成分と直線との水平距離 d の増大に応じて単調減少する非負係数 を計算し、これを当該周波数成分のパワーに乘じることで、直線から水平距離で遠い成分ほど弱いパワーで音源音に寄与するようにした方式である。

20

【0074】

このとき、水平距離による閾値処理を行う必要はなく、ある直線群に対する各周波数成分の水平距離（直線群の中の最も近い直線との水平距離） d を求め、その水平距離 d に基づいて定められる係数 を当該周波数成分のパワーに乘じた値を、当該直線群における当該周波数成分のパワーとする。水平距離 d の増大に応じて単調減少する非負係数 の計算式は任意であるが、一例として図 25 に示すシグモイド（S 字曲線）関数 $= \exp(- (B \cdot d)^C)$ が挙げられる。このとき図に例示したように、 B を正の数値（図では 1.5）、 C を 1 より大きい数値（図では 2.0）とすると、 $d = 0$ のとき $= 1$ 、 d のとき 0 となる。非負係数 の減少の度合いが急峻、すなわち B が大きいと直線群からはずれた成分が排除され易くなるので、音源方向に対する指向性が鋭くなり、逆に非負係数 の減少の度合いが緩慢、すなわち B が小さいと指向性が鈍くなる。

30

【0075】

[複数 FFT 結果の扱い]

既に述べたように、投票部 303 は 1 回の FFT 毎に投票を行うことも、連続する m 回 ($m - 1$) の FFT 結果をまとめて投票することも可能である。したがって、ハフ投票結果を処理する直線検出部 304 以降の機能ブロックは 1 回のハフ変換が実行される期間を単位として動作する。このとき、 $m = 2$ でハフ投票が行われるときは、複数時刻の FFT 結果がそれぞれの音源音を構成する成分として分類されることになり、時刻の異なる同一周波数成分が別々の音源音に帰属させられることも起こり得る。これを扱うために、 m の値に関わらず、座標値決定部 302 によって、各周波数成分（すなわち、図 24 に例示した黒丸）にはそれが取得されたフレームの開始時刻が取得時刻の情報として付与されており、どの時刻のどの周波数成分がどの音源に帰属するかを参照可能にする。すなわち、音源音がその周波数成分の時系列データとして分離抽出される。

40

【0076】

[パワー保存オプション]

なお、上述した各方式において、複数（ N 個）の直線群（音源）に属する周波数成分（

50

最近傍方式では直流成分のみ、距離係数方式では全周波数成分が該当)では、各音源に配分される同一時刻の当該周波数成分のパワーを、その合計が配分前の当該時刻のパワー値 $P_o(f, k)$ に等しくなるように正規化してN分割することも可能である。このようにすることで、同一時刻の周波数成分毎に音源全体での合計パワーを入力と同じに保つことができる。これを「パワー保存オプション」と呼ぶことにする。配分の仕方には次の2つの考え方がある。

【0077】

すなわち、(1) N等分(距離閾値方式と最近傍方式に適用可能)と、(2)各直線群との距離に応じた配分(距離閾値方式と距離係数方式に適用可能)である。

【0078】

(1)はN等分することで自動的に正規化が達成される配分方法であり、距離に関係なく配分を決める距離閾値方式と最近傍方式に適用可能である。

【0079】

(2)は距離係数方式と同様にして係数を決めた後、さらにそれらの合計が1になるように正規化することでパワーの合計を保存する配分方法であり、原点以外で多重帰属の発生する距離閾値方式と距離係数方式に適用可能である。

【0080】

なお、音源成分推定部312は、設定によって距離閾値方式と最近傍方式と距離係数方式のいずれを行うことも可能である。また、距離閾値方式と最近傍方式において上述したパワー保存オプションを選択することが可能である。

【0081】

[音源音再合成部313]

音源音再合成部313は、各音源音を構成する同一取得時刻の周波数成分を逆FFT処理することによって、当該時刻を開始時刻とするフレーム区間の当該音源音(振幅データ)を再合成する。図3に図示したように、1つのフレームは次のフレームとフレームシフト量だけの時間差をおいて重複している。このように複数のフレームで重複している区間では、重複する全てのフレームの振幅データを平均して最終的な振幅データと成すことができる。このような処理によって、音源音をその振幅データとして分離抽出することが可能になる。

【0082】

[時系列追跡部314]

上述した通り、投票部303によるハフ投票毎に直線検出部304により直線群が求められる。ハフ投票は連続するm回(m-1)のFFT結果についてまとめて行われる。この結果、直線群はmフレーム分の時間を周期(これを「図形検出周期」と呼ぶことにする)として時系列的に求められることになる。また、直線群の は方向推定部305により計算される音源方向 と1対1に対応しているので、音源が静止していても移動していても、安定な音源に対応している(あるいは)の時間軸上の軌跡は連続しているはずである。一方、直線検出部304により検出された直線群の中には、閾値の設定具合によって背景雑音に対応する直線群(これを「雑音直線群」と呼ぶことにする)が含まれていることがある。しかしながら、このような雑音直線群の(あるいは)の時間軸上の軌跡は連続していないか、連続していても短いことが期待できる。

【0083】

時系列追跡部314は、このように図形検出周期毎に求められる を時間軸上で連続なグループに分けることで、 の時間軸上の軌跡を求める手段である。図26を用いてグループ分けの方法を説明する。

【0084】

(1)軌跡データバッファを用意する。軌跡データバッファは軌跡データの配列である。1つの軌跡データKdは、その開始時刻Tsと、終了時刻Teと、当該軌跡を構成する直線群データLdの配列(直線群リスト)と、ラベル番号Lnとを保持することができる。1つの直線群データLdは、当該軌跡を構成する1つの直線群の 値と 値(直線検出

10

20

30

40

50

部 3 0 4 による)と、この直線群に対応した音源方向を表す 値 (方向推定部 3 1 1 による)と、この直線群に対応した周波数成分 (音源成分推定部 3 1 2 による)と、それらが取得された時刻とからなる一群のデータである。なお、軌跡データバッファは最初空である。また、ラベル番号を発行するためのパラメータとして新規ラベル番号を用意し、初期値を 0 に設定する。

【 0 0 8 5 】

(2) ある時刻 T において、新しく得られた のそれぞれ (以後 n とし、図中では黒丸 3 0 3 と黒丸 3 0 4 で示される 2 つが得られたものとする) について、軌跡データバッファに保持されている軌跡データ K d (図中の矩形 3 0 1 と 3 0 2) の直線群データ L d (図中の矩形内に配置された黒丸) を参照し、その 値と n の差 (図中の 3 0 5 と 3 0 6) が所定角度閾値 内にあり、かつその取得時刻の差 (図中の 3 0 7 と 3 0 8) が所定時間閾値 t 内にある L d を持つ軌跡データを検出する。この結果、黒丸 3 0 3 については軌跡データ 3 0 1 が検出されたが、黒丸 3 0 4 については最も近い軌跡データ 3 0 2 も上記条件を満たさなかったとする。

【 0 0 8 6 】

(3) 黒丸 3 0 3 のように、もし、(2) の条件を満たす軌跡データが見つかった場合は、 n はこの軌跡と同一の軌跡を成すものとして、この n とそれに対応した 値と 値と周波数成分と現時刻 T とを当該軌跡 K d の新たな直線群データとして直線群リストに追加し、現時刻 T を当該軌跡の新たな終了時刻 T e とする。このとき、複数の軌跡が見つかった場合には、それら全てが同一の軌跡を成すものとして、最も若いラベル番号を持つ軌跡データに統合して、残りを軌跡データバッファから削除する。統合された軌跡データの開始時刻 T s は統合前の各軌跡データの中で最も早い開始時刻であり、終了時刻 T e は統合前の各軌跡データの中で最も遅い終了時刻であり、直線群リストは統合前の各軌跡データの直線群リストの和集合である。この結果、黒丸 3 0 3 は軌跡データ 3 0 1 に追加される。

【 0 0 8 7 】

(4) 黒丸 3 0 4 のように、もし、(2) の条件を満たす軌跡データが見つからなかった場合は、新規の軌跡の始まりとし、軌跡データバッファの空き部分に新しい軌跡データを作成し、開始時刻 T s と終了時刻 T e を共に現時刻 T とし、 n とそれに対応した 値と 値と周波数成分と現時刻 T とを直線群リストの最初の直線群データとし、新規ラベル番号の値をこの軌跡のラベル番号 L n として与え、新規ラベル番号を 1 だけ増加させる。なお、新規ラベル番号が所定の最大値に達したときは、新規ラベル番号を 0 に戻す。この結果、黒丸 3 0 4 は新たな軌跡データとして軌跡データバッファに登録される。

【 0 0 8 8 】

(5) もし、軌跡データバッファに保持されている軌跡データで、最後に更新されてから (すなわちその終了時刻 T e から) 現時刻 T までに前記所定時間 t を経過したものがあれば、追加すべき新たな n の見つからなかった、すなわち追跡の満了した軌跡として、この軌跡データを次段の継続時間評価部 3 1 5 に出力した後、当該軌跡データを軌跡データバッファから削除する。図の例では軌跡データ 3 0 2 がこれに該当する。

【 0 0 8 9 】

[継続時間評価部 3 1 5]

継続時間評価部 3 1 5 は、時系列追跡部 3 1 4 により出力された追跡の満了した軌跡データの開始時刻と終了時刻から当該軌跡の継続時間を計算し、この継続時間が所定閾値を越えるものを音源音に基づく軌跡データと認定し、それ以外を雑音に基づく軌跡データと認定する。音源音に基づく軌跡データを音源ストリーム情報と呼ぶことにする。音源ストリーム情報には、当該音源音の開始時刻 T s 、終了時刻 T e 、当該音源方向を表す と の時系列的な軌跡データが含まれる。なお、図形検出部 5 による直線群の数が音源の数を与えるが、そこには雑音源も含まれている。継続時間評価部 3 1 5 による音源ストリーム情報の数は、雑音に基づくものを除いた信頼できる音源の数を与えてくれる。

【 0 0 9 0 】

10

20

30

40

50

[同相化部 3 1 6]

同相化部 3 1 6 は、時系列追跡部 3 1 4 による音源ストリーム情報を参照することで、当該ストリームの音源方向 の時間推移を得て、 の最大値 max と最小値 min から中間値 $mid = (max + min) / 2$ を計算して幅 $w = max - mid$ を求める。そして、当該音源ストリーム情報の元となった 2 つの周波数分解データセット a と b の時系列データを、当該ストリームの開始時刻 T_s より所定時間遡った時刻から終了時刻 T_e より所定時間経過した時刻まで抽出して、中間値 mid で逆算される到達時間差をキャンセルするように補正することで同相化する。

【 0 0 9 1 】

あるいは、方向推定部 3 1 1 による各時刻の音源方向 を mid として、2 つの周波数分解データセット a と b の時系列データを常時同相化することもできる。音源ストリーム情報を参照するか、各時刻の を参照するかは動作モードで決定され、この動作モードはパラメータとして設定・変更可能である。

10

【 0 0 9 2 】

[適応アレイ処理部 3 1 7]

適応アレイ処理部 3 1 7 は、抽出・同相化された 2 つの周波数分解データセット a と b の時系列データを、正面 0° に中心指向性に向け、 $\pm w$ に所定のマージンを加えた値を追従範囲とする適応アレイ処理に掛けることで、当該ストリームの音源音の周波数成分の時系列データを高精度に分離抽出する。この処理は方法こそ異なるが、周波数成分の時系列データを分離抽出する点において音源成分推定部 3 1 2 と同様の働きをする。それ故、音源音再合成部 3 1 3 は、適応アレイ処理部 3 1 7 による音源音の周波数成分の時系列データからも、その音源音の振幅データを再合成することができる。

20

【 0 0 9 3 】

なお、適応アレイ処理としては、参考文献 3 「天田皇ほか “音声認識のためのマイクロホンアレイ技術”，東芝レビュー 2 0 0 4，VOL. 5 9，NO. 9，2 0 0 4」に記載のように、それ自体がビームフォーミングの構成方法として知られている「Griffith-Jim 型一般化サイドローブキャンセラ」を主副 2 つに用いるなど、設定された指向性範囲内の音声を明瞭に分離抽出する方法を適用することができる。

【 0 0 9 4 】

通常、適応アレイ処理を用いる場合、事前に追従範囲を設定し、その方向からの音声のみを待ち受ける使い方をするため、全方位からの音声を待ち受けるためには追従範囲を異ならせた多数の適応アレイを用意する必要があった。一方、本実施形態では、実際に音源の数とその方向を求めたうえで、音源数に応じた数の適応アレイだけを稼働させることができ、その追従範囲も音源の方向に応じた所定の狭い範囲に設定することができるので、音声を効率良くかつ品質良く分離抽出できる。

30

【 0 0 9 5 】

また、このとき、事前に 2 つの周波数分解データセット a と b の時系列データを同相化することで、適応アレイ処理における追従範囲を正面付近にのみ設定するだけで、あらゆる方向の音を処理できるようになる。

【 0 0 9 6 】

[音声認識部 3 1 8]

音声認識部 3 1 8 は、音源成分推定部 3 1 2 もしくは適応アレイ処理部 3 1 7 により抽出された音源音の周波数成分の時系列データを解析照合することで、当該ストリームの記号的な内容、すなわち、言語的な意味や音源の種別や話者の別を表す記号（列）を抽出する。

40

【 0 0 9 7 】

なお、方向推定部 3 1 1 から音声認識部 3 1 8 までの各機能ブロックは、必要に応じて図 2 0 に図示しない結線によって情報のやりとりが可能であるものとする。

【 0 0 9 8 】

[出力部 7]

50

出力部 7 は、音源情報生成部 6 による音源情報として、図形検出部 5 による直線群の数として得られる音源の数、方向推定部 3 1 1 により推定される、音響信号の発生源たる各音源の空間的な存在範囲（円錐面を決定させる角度）、音源成分推定部 3 1 2 により推定される、各音源が発した音声の成分構成（周波数成分毎のパワーと位相の時系列データ）、音源音再合成部 3 1 3 により合成される、音源毎に分離された分離音声（振幅値の時系列データ）、時系列追跡部 3 1 4 と継続時間評価部 3 1 5 とに基づいて決定される、雑音源を除く音源の数、時系列追跡部 3 1 4 と継続時間評価部 3 1 5 とにより決定される、各音源が発した音声の時間的な存在期間、正面化部 3 1 6 と適応アレイ部 3 1 7 とにより求められる、音源毎の分離音声（振幅値の時系列データ）、音声認識部 3 1 8 により求められる、各音源音声の記号的内容、の少なくとも 1 つを含む情報を出力する手段である。

10

【 0 0 9 9 】

[ユーザインタフェース部 8]

ユーザインタフェース部 8 は、上述した音響信号処理に必要な各種設定内容の利用者への呈示、利用者からの設定入力受理、設定内容の外部記憶装置への保存と外部記憶装置からの読み出しを実行したり、図 1 7 や図 1 9 に示した（ 1 ）マイク毎の周波数成分の表示、（ 2 ）位相差（あるいは時間差）プロット図の表示（すなわち 2 次元データの表示）、（ 3 ）各種得票分布の表示、（ 4 ）極大位置の表示、（ 5 ）プロット図上の直線群の表示、図 2 3 や図 2 4 に示した（ 6 ）直線群に帰属する周波数成分の表示、図 2 6 に示した（ 7 ）軌跡データの表示、のように各種処理結果や中間結果を可視化して利用者に呈示したり、所望のデータを利用者に選択させてより詳細に可視化するための手段である。このよ

20

【 0 1 0 0 】

[処理の流れ図]

図 2 7 は本実施形態に係る音響信号処理装置が実行する処理の流れを示すフローチャートである。この処理は、初期設定処理ステップ S 1 と、音響信号入力処理ステップ S 2 と、周波数分解処理ステップ S 3 と、2 次元データ化処理ステップ S 4 と、図形検出処理ステップ S 5 と、音源情報生成処理ステップ S 6 と、出力処理ステップ S 7 と、終了判断処理ステップ S 8 と、確認判断処理ステップ S 9 と、情報呈示・設定受理処理ステップ S 1 0 と、終了処理ステップ S 1 1 とを有する。

30

【 0 1 0 1 】

初期設定処理ステップ S 1 は、上述したユーザインタフェース部 8 における処理の一部を実行する処理ステップであり、音響信号処理に必要な各種設定内容を外部記憶装置から読み出して、装置を所定の設定状態に初期化する。

【 0 1 0 2 】

音響信号入力処理ステップ S 2 は、上述した音響信号入力部 2 における処理を実行する処理ステップであり、空間的に同一でない 2 つの位置で捉えられた 2 つの音響信号を入力する。

【 0 1 0 3 】

周波数分解処理ステップ S 3 は、上述した周波数分解部 3 における処理を実行する処理ステップであり、前記音響信号入力処理ステップ S 2 による入力音響信号のそれぞれを周波数分解して、周波数毎の少なくとも位相値（と必要ならパワー値）を算出する。

40

【 0 1 0 4 】

2 次元データ化処理ステップ S 4 は、上述した 2 次元データ化部 4 における処理を実行する処理ステップであり、前記周波数分解処理ステップ S 3 により算出された各入力音響信号の周波数毎の位相値を比較して、両者の周波数毎の位相差値を算出し、該周波数毎の位相差値を、周波数の関数を Y 軸、位相差値の関数を X 軸とする X Y 座標系上の点として、各周波数とその位相差により一意に決定される（ x , y ）座標値に変換する。

【 0 1 0 5 】

50

図形検出処理ステップ S 5 は、上述した図形検出部 5 における処理を実行する処理ステップであり、前記 2 次元データ化処理ステップ S 4 による 2 次元データから所定の図形を検出する。

【 0 1 0 6 】

音源情報生成処理ステップ S 6 は、上述した音源情報生成部 6 における処理を実行する処理ステップであり、前記図形検出処理ステップ S 5 により検出された図形の情報に基づいて、前記音響信号の発生源たる音源の数、各音源の空間的な存在範囲、前記各音源を発生した音声の成分構成、前記音源毎の分離音声、前記各音源を発生した音声の時間的な存在期間、前記各音源を発生した音声の記号的内容、の少なくとも 1 つを含む音源情報を生成する。

10

【 0 1 0 7 】

出力処理ステップ S 7 は、上述した出力部 7 における処理を実行する処理ステップであり、前記音源情報生成処理ステップ S 6 により生成された音源情報を出力する。

【 0 1 0 8 】

終了判断処理ステップ S 8 は、上述したユーザインタフェース部 8 における処理の一部を実行する処理ステップであり、利用者からの終了命令の有無を検査して、終了命令がある場合には終了処理ステップ S 11 へ（左分岐）、無い場合には確認判断処理ステップ S 9 へ（上分岐）と処理の流れを制御する。

【 0 1 0 9 】

確認判断処理ステップ S 9 は、上述したユーザインタフェース部 8 における処理の一部を実行する処理ステップであり、利用者からの確認命令の有無を検査して、確認命令がある場合には情報呈示・設定受理処理ステップ S 10 へ（左分岐）、無い場合には音響信号処理ステップ S 2（上分岐）と処理の流れを制御する。

20

【 0 1 1 0 】

情報呈示・設定受理処理ステップ S 10 は、利用者からの確認命令を受けて実行される、上述したユーザインタフェース部 8 における処理の一部を実行する処理ステップであり、音響信号処理に必要な各種設定内容の利用者への呈示、利用者からの設定入力受理、保存命令による設定内容の外部記憶装置への保存、読み出し命令による設定内容の外部記憶装置からの読み出しを実行したり、各種処理結果や中間結果を可視化して利用者に呈示したり、所望のデータを利用者に選択させてより詳細に可視化することで、利用者が音響信号処理の動作を確認したり、所望の動作を行ない得るように調整したり、以後調整済みの状態で処理を継続したりすることを可能にする。

30

【 0 1 1 1 】

終了処理ステップ S 11 は、利用者からの終了命令を受けて実行される、上述したユーザインタフェース部 8 における処理の一部を実行する処理ステップであり、音響信号処理に必要な各種設定内容の外部記憶装置への保存を自動実行する。

【 0 1 1 2 】

[変形例] ここで、上述した実施形態の変形例を説明する。

【 0 1 1 3 】

[垂直線の検出]

40

2 次元データ化部 4 は、その座標値決定部 302 によって、図 7 に示すように X 座標値を位相差 $P_h(f, k)$ 、Y 座標値を周波数成分番号 k として点群を生成した。このとき、X 座標値を位相差 $P_h(f, k)$ からさらに計算される到達時間差の周波数毎の推定値 $T(f, k) = (P_h(f, k) / 2) \times (1 / f, k)$ とすることも可能である。位相差の代わりに到達時間差を使うと、同一の到達時間差を持つ、すなわち同一音源に由来する点は垂直な直線上に並ぶことになる。

【 0 1 1 4 】

このとき、周波数が高くなるほど $P_h(f, k)$ で表現可能な時間差 $T(f, k)$ は小さくなる。図 28 (a) に模式的に示すように、周波数 f, k の波 290 の 1 周期が表す時間を T とすると、2 倍の周波数 $2f, k$ の波 291 の 1 周期が表すことのできる時間は $T /$

50

2と半分になってしまう。このとき、図28(a)のようにX軸を時間差とすると、その範囲は $\pm T_{max}$ であり、これを超えて時間差が観測されることはない。ところが、 T_{max} が $1/2$ 周期(すなわち)以下となる限界周波数292以下の低い周波数では、位相差 $Ph(fk)$ から到達時間差 $T(fk)$ が一意に求められるが、限界周波数292を超えた高い周波数では、算出された $T(fk)$ は理論上可能な T_{max} よりも小さく、図28(b)に示すように直線293と294に挟まれる範囲しか表現できない。これは上述した位相差循環の問題と同じ問題である。

【0115】

そこで、この位相差循環の問題を解決するために、限界周波数292を超える周波数域については、図29に模式的に示すように、座標値決定部302が1つの $Ph(fk)$ について2、4、6などを足したり引いたりした位相差に対応する T の位置にも冗長な点を、 $\pm T_{max}$ の範囲内で生成して2次元データとする。生成された点群が図中の黒丸であり、限界周波数292を超えた周波数域では1つの周波数について複数の黒丸がプロットされている。

10

【0116】

このようにすることで、1つの位相差値に対して1乃至複数の点として生成される2次元データから、投票部303と直線検出部304から有力な垂直線(図中の295)をハフ投票によって検出することが可能になる。このとき、垂直線はハフ投票空間上で $=0$ となる直線なので、垂直線の検出問題はハフ投票後の得票分布で $=0$ となる軸上の極大位置で所定閾値以上の得票を得るものを検出することで解くことができる。ここで検出された極大位置の値が垂直線とX軸の交点、すなわち到達時間差 T の推定値を与えてくれる。なお、投票に際しては投票部303の説明に記載した投票条件と加算方式をそのまま用いることが可能である。また、音源に対応した直線は直線群ではなく単一の垂直線である。

20

【0117】

この極大位置を求める問題は、上述の冗長な点群のX座標値を投票した1次元の得票分布(Y軸方向へ射影投票した周辺分布)上の極大位置で所定閾値以上の得票を得るものを検出することで解くこともできる。このように、位相差の代わりに到達時間差をX軸に用いることで、異なる方向に存在する音源を表す証拠が全て同じ傾きの(すなわち垂直な)直線に写されるので、ハフ変換によらずとも周辺分布によって簡便に検出可能になる。

30

【0118】

垂直線を求めることで得られる音源方向の情報はではなくとして得られる到達時間差 T である。よって、方向推定部311はを介在させることなく T から直ちに音源方向を算出可能となる。

【0119】

このように、2次元データ化部4による2次元データは1種類に限らず、図形検出部5による図形の検出法も1つとは限らない。なお、図29に例示した到達時間差を使った点群のプロット図と検出された垂直線もユーザインタフェース部8による利用者への情報呈示対象である。

【0120】

[複数システムの並列実装]

また、以上の例はマイクロホンを2つ備えた最も単純な構成で説明したものであるが、図30に示すように、マイクロホンを $N(N \geq 3)$ 個備え、最大 $M(1 \leq M \leq N C_2)$ 個のマイクロホン対を構成することも可能である。

40

【0121】

図中の11~13は N 個のマイクロホンである。図中の20は N 個のマイクロホンによる N 個の音響信号を入力する手段であり、図中の21は入力された N 個の音響信号をそれぞれ周波数分解する手段である。図中の22は N 個の音響信号のうちの2つからなる $M(1 \leq M \leq N C_2)$ 組の対の各々について2次元データを生成する手段であり、図中の23は生成された M 組の2次元データからそれぞれ所定の図形を検出する手段である。図中の

50

24は検出されたM組の図形の情報のそれぞれから音源の情報を生成する手段であり、図中の25は生成された音源の情報を出力する手段である。図中の26は各対を構成するマイクロホンの情報を含む各種設定値の利用者への呈示、利用者からの設定入力受理、外部記憶装置への設定値の保存、外部記憶装置からの設定値の読み出し、及び各種処理結果の利用者への呈示を実行する手段である。各マイクロホン対における処理は上述した実施形態と同様であり、そのような処理が複数のマイクロホン対について並列的に実行される。

【0122】

このようにすることで、1つのマイクロホン対では方向への得て不得手があっても、複数の対でカバーすることで正しい音源情報を取りこぼす危険を軽減することが可能になる。

10

【0123】

[汎用コンピュータを使った実施：プログラム]

また、本発明は、図31に示すように本発明に係る音響信号処理機能を実現するためのプログラムを実行可能な汎用コンピュータとして実施することも可能である。図中の31~33はN個のマイクロホンである。図中の40はN個のマイクロホンによるN個の音響信号を入力するA/D変換手段であり、図中の41は入力されたN個の音響信号を処理するためのプログラム命令を実行するCPUである。図中の42~47はコンピュータを構成する標準的なデバイスであり、それぞれRAM42、ROM43、HDD44、マウス/キーボード45、ディスプレイ46、LAN47である。また、図中の50~52は外部から記憶メディアを介してプログラムやデータをコンピュータに供給するためのドライ
20
ブ類であり、それぞれCDROM50、FDD51、CF/SDカード52である。図中の48は音響信号を出力するためのD/A変換手段であり、その出力にスピーカ49が繋がっている。このコンピュータ装置は、図27に示した処理ステップを実行するための音響信号処理プログラムをHDD44に記憶し、これをRAM42に読み出してCPU41で実行することで音響信号処理装置として機能する。また、外部記憶装置としてのHDD44、操作入力を受け付けるマウス/キーボード45、情報呈示手段としてのディスプレイ46とスピーカ49を使うことで、上述したユーザインタフェース部8の機能を実現する。また、音響信号処理によって得られた音源情報をRAM42やROM43やHDD44に保存出力したり、LAN47を介して通信出力する。

20

【0124】

30

[記録媒体]

また、本発明は図32に示すようにコンピュータ読み取り可能な記録媒体として実施することも可能である。図中の61は本発明に係る音響信号処理プログラムを記録したCD-ROMやCFやSDカードやフロッピー（登録商標）ディスクなどで実現される記録媒体である。この記録媒体61をテレビやコンピュータなどの電子装置62や電子装置63やロボット64に挿入することで当該プログラムを実行可能としたり、あるいはプログラムを供給された電子装置63から通信によって別の電子装置65やロボット64に当該プログラムを供給することで、電子装置65やロボット64上で当該プログラムを実行可能とする。

【0125】

40

[温度センサによる音速の補正]

また、本発明は装置に外気温を測定するための温度センサを備え、該温度センサによって計測された気温データに基づいて図22における音速Vsを補正して、正確なTmaxを求めるように実施することも可能である。

【0126】

あるいは、本発明は装置に所定の間隔を空けて配置された音波の発信手段と受信手段とを備え、該発信手段を発した音波が該受信手段に到達するまでの時刻を計測手段で測ることで、直接的に音速Vsを計算・補正して、正確なTmaxを求めるように実施することも可能である。

【0127】

50

[の等間隔化のための の不等間隔化]

また、本発明は直線群の傾きを得るためにハフ変換を実行する際に を例えば 1° 刻みというように量子化を行うが、このように を等間隔に刻むと推定可能な音源方向 の値が不等間隔に量子化されてしまう。そこで、本発明は を等間隔とするように の量子化を行うことで、音源方向の推定精度に粗密が生じにくいように実施することも可能である。

【 0 1 2 8 】

上記非特許文献 2 に記載の方法は、周波数分解データから調波構造を構成する基本周波数成分とその高調波成分を検出することで、音源の数、方向、成分の推定を行っている。調波構造を仮定することから、この方法は人間の声に特化したものであるといえる。しかし、実際の環境では、ドアの開閉音など、調波構造を持たない音源も多く存在するため、この方法ではそのような音源音を扱うことができない。

10

【 0 1 2 9 】

また、上記非特許文献 1 に記載の方法は、特定のモデルに縛られないが、2つのマイクロホンを使う限り、扱うことのできる音源は1つに限られてしまう。

【 0 1 3 0 】

一方、本発明の実施形態によれば、ハフ変換を使って周波数成分毎の位相差を音源毎のグループに分けることで、2つのマイクロホンを使いながら2つ以上の音源を定位し、かつ分離する機能を実現することができる。このとき、調波構造のような限定的なモデルを使用しないので、より広範な性質の音源に適用することができる。

20

【 0 1 3 1 】

本発明の実施形態が奏する他の作用効果を纏めると以下の通りである。

【 0 1 3 2 】

・ハフ投票に際して周波数成分の多い音源やパワーの強い音源の検出に適した投票方法を使うことで、幅広い種類の音源を安定に検出することができる。

【 0 1 3 3 】

・直線検出に際して $= 0$ の制約や位相差循環の考慮を行うことで、効率良くかつ精度良く音源を検出することができる。

【 0 1 3 4 】

・直線検出結果を用いて、音響信号の発生源たる音源の空間的な存在範囲、音源を発生した音源音の時間的な存在期間、音源音の成分構成、音源音の分離音声、音源音の記号的内容を含む有益な音源情報を求めることができる。

30

【 0 1 3 5 】

・各音源音の周波数成分を推定する際に、単純に直線近傍の成分を選択したり、ある成分がどの直線に帰属するかを判定したり、各直線と成分の距離に応じた係数掛けを行うことで、簡便な方法で音源音を個々に分離することができる。

【 0 1 3 6 】

・各音源の方向を予め知ることで、適応アレイ処理の指向性範囲を適応的に設定して、より高精度に音源音を分離することができる。

【 0 1 3 7 】

・各音源音を高精度に分離して認識することで、音源音の記号的内容を判別することができる。

40

【 0 1 3 8 】

・利用者が本装置の働きを確認したり、所望の動作を行ない得るように調整したり、以後調整済みの状態で本装置を利用したりすることが可能になる。

【 0 1 3 9 】

なお、本発明は上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実

50

施形態にわたる構成要素を適宜組み合わせてもよい。

【図面の簡単な説明】

【0140】

【図1】本発明の一実施形態に係る音響信号処理装置の機能ブロック図

【図2】音源方向と、音響信号において観察される到達時間差とを示す図

【図3】フレームとフレームシフト量との関係を示す図

【図4】FFT処理の手順および短時間フーリエ変換データを示す図

【図5】2次元データ化部および図形検出部のそれぞれの内部構成を示す機能ブロック図

【図6】位相差算出の手順を示す図

【図7】座標値計算の手順を示す図

【図8】同一時間について周波数と位相との間の比例関係、および同一時間差について位相差と周波数との間の比例関係を示す図

【図9】位相差の循環性を説明するための図

【図10】複数の音源が存在する場合の周波数と位相差のプロット図

【図11】直線ハフ変換について説明するための図

【図12】ハフ変換により点群から直線を検出することについて説明するための図

【図13】投票される平均パワーの関数（計算式）を示す図

【図14】実際の音声から生成された周波数成分、位相差プロット図、ハフ投票結果を示す図

【図15】実際のハフ投票結果から求められた極大位置と直線を示す図

【図16】 と の関係を示す図

【図17】2人の人物の同時発話時の周波数成分、位相差プロット図、ハフ投票結果を示す図

【図18】 軸上の得票値のみで極大位置を探索した結果を示す図

【図19】 ずつ離れた数箇所の得票値を合計して極大位置を探索した結果を示す図

【図20】音源情報生成部の内部構成を示す機能ブロック図

【図21】方向推定を説明するための図

【図22】 と Tとの関係を示す図

【図23】複数音源存在時の音源成分推定（距離閾値方式）について説明するための図

【図24】最近傍方式について説明するための図

【図25】係数 の計算式の例とそのグラフを示す図

【図26】 の時間軸上の追跡を説明した図

【図27】音響信号処理装置が実行する処理の流れを示すフローチャート

【図28】周波数と表現可能な時間差との関係を示す図

【図29】冗長点を生成した場合における時間差のプロット図

【図30】N個のマイクロホンを具備する変形実施形態に係る音響信号処理装置の機能ブロック構成図

【図31】本発明に係る音響信号処理機能を汎用コンピュータを用いて実現する実施形態に係る機能ブロック図

【図32】本発明に係る音響信号処理機能を実現するためのプログラムを記録した記録媒体による実施形態を示した図

【符号の説明】

【0141】

1 a , 1 b ... マイク ;

2 ... 音響信号入力部 ;

3 ... 周波数分解部 ;

4 ... 2次元データ化部 ;

5 ... 図形検出部 ;

6 ... 音源情報生成部 ;

7 ... 出力部 ;

10

20

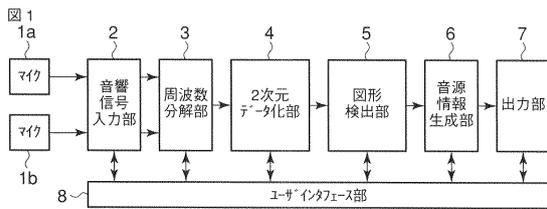
30

40

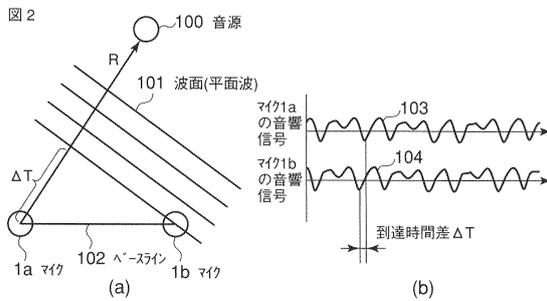
50

8 ...ユーザインタフェース部

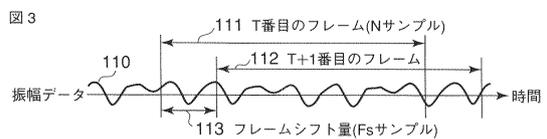
【図1】



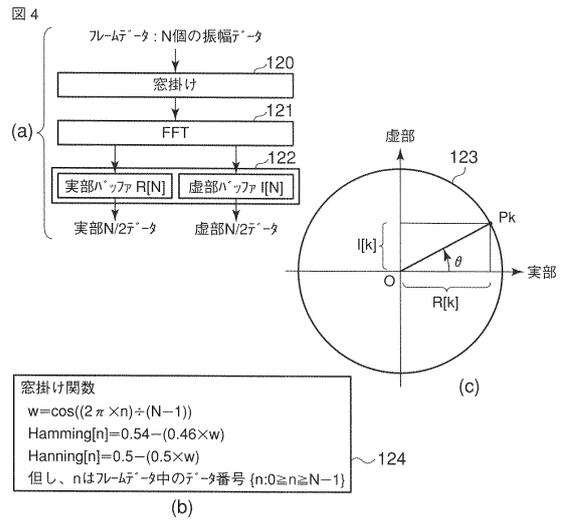
【図2】



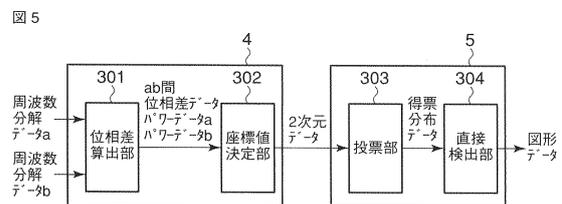
【図3】



【図4】



【図5】



【 図 6 】

図 6

位相差

```

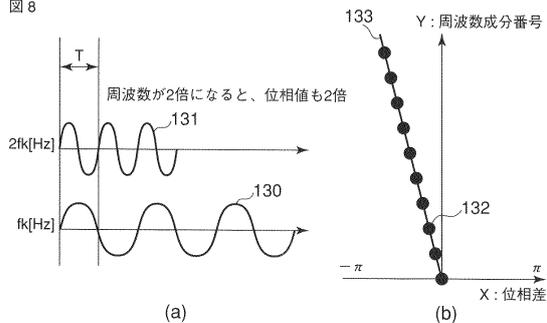
 $\Delta Ph(fk) = Ph2(fk) - Ph1(fk);$ 
while(1){
    if( $\Delta Ph(fk) \leq -\pi$ ){  $\Delta Ph(fk) = \Delta Ph(fk) + 2\pi$ ; continue;}
    break;
}
while(1){
    if( $\Delta Ph(fk) > \pi$ ){  $\Delta Ph(fk) = \Delta Ph(fk) - 2\pi$ ; continue;}
    break;
}

```

但し、
 Ph1(fk)はマイク1aの周波数成分fkにおける位相値
 Ph2(fk)はマイク1bの周波数成分fkにおける位相値
 $\Delta Ph(fk)$ の値域は、 $\{\Delta Ph(fk): -\pi < \Delta Ph(fk) \leq \pi\}$

【 図 8 】

図 8



【 図 7 】

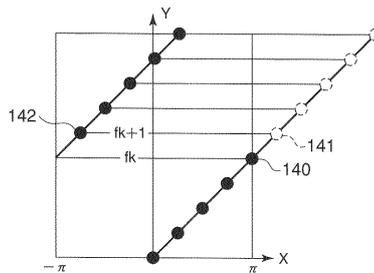
図 7

座標値

$x(fk) = \Delta Ph(fk)$
 $y(fk) = k$

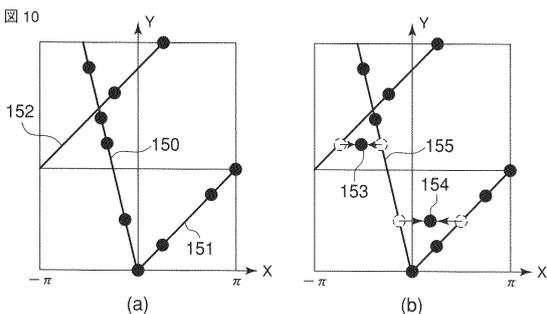
【 図 9 】

図 9



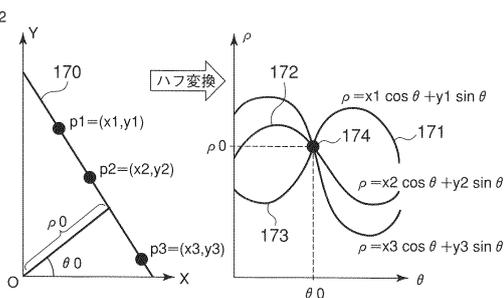
【 図 10 】

図 10



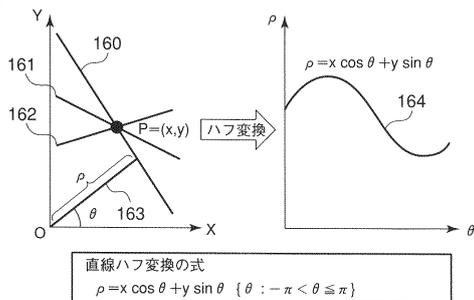
【 図 12 】

図 12



【 図 11 】

図 11



【 図 13 】

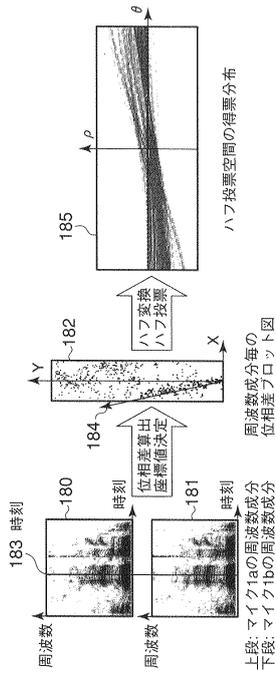
図 13

パワーの関数

$G(P(fk)) = V + 1 : V > 0$
 $G(P(fk)) = 1 : V \leq 0$
 但し、
 $V = \log_{10}(P(fk)) + \alpha$
 $P(fk) = (Po2(fk) + Po1(fk)) / 2$

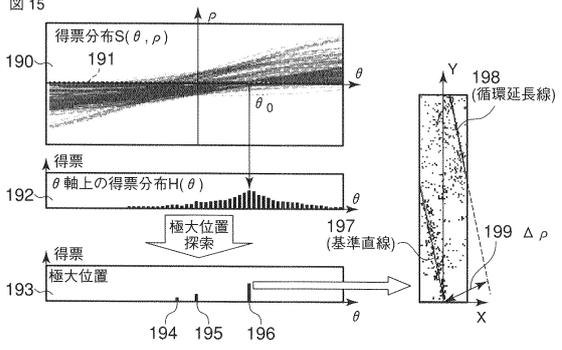
【 図 14 】

図 14



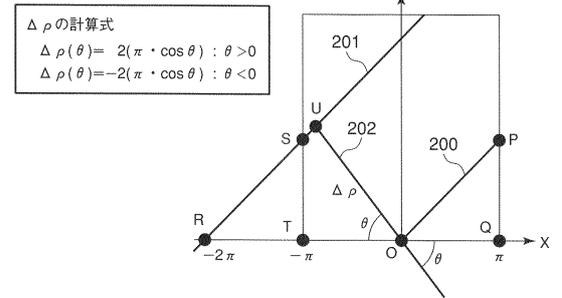
【 図 15 】

図 15



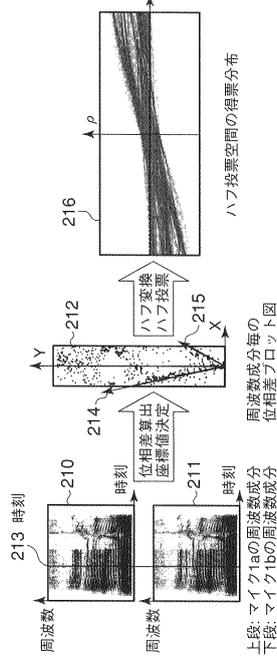
【 図 16 】

図 16



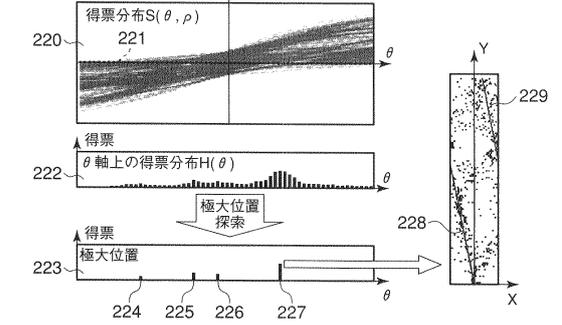
【 図 17 】

図 17



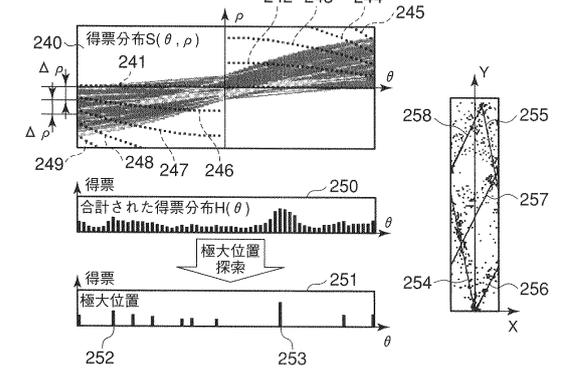
【 図 18 】

図 18

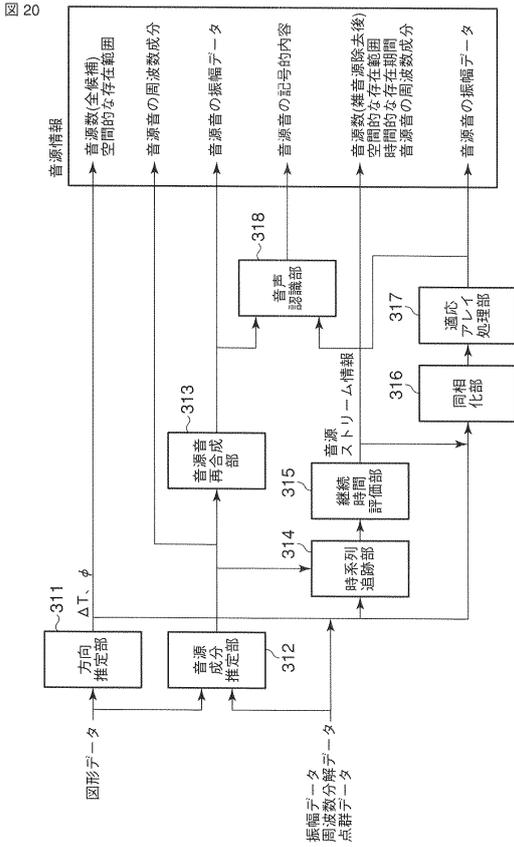


【 図 19 】

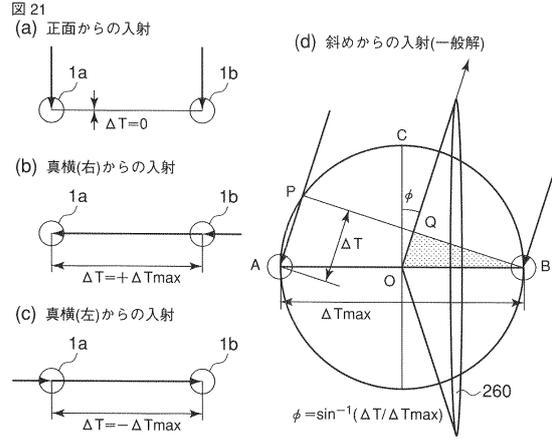
図 19



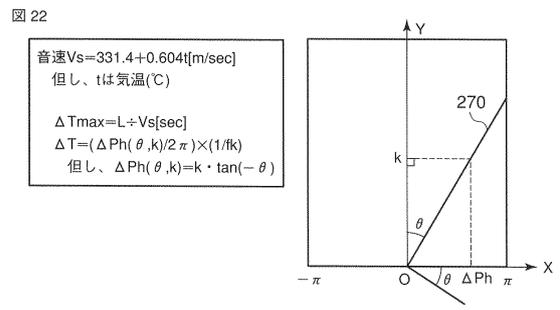
【 図 2 0 】



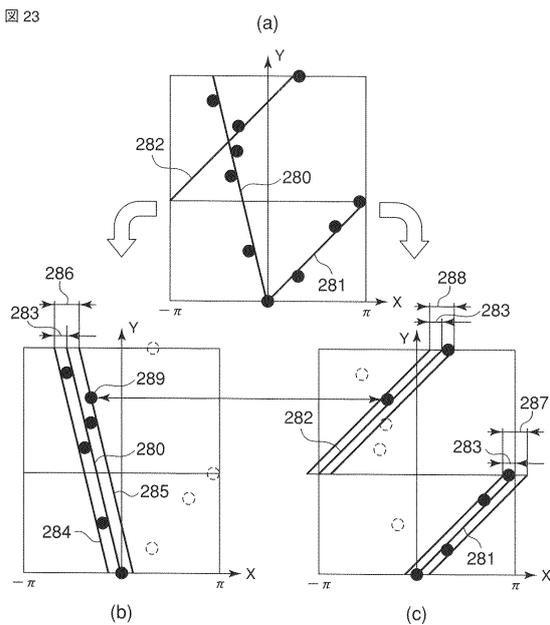
【 図 2 1 】



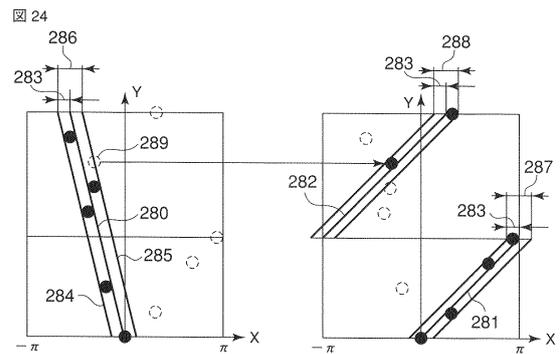
【 図 2 2 】



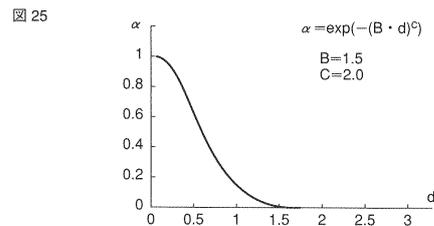
【 図 2 3 】



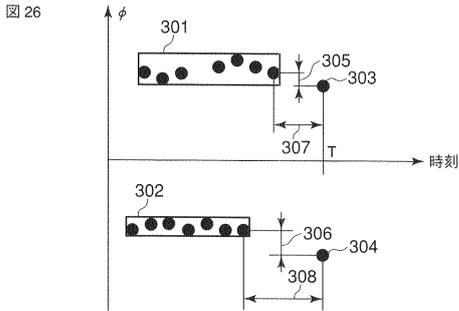
【 図 2 4 】



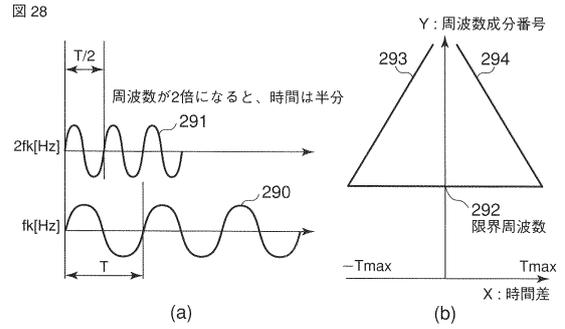
【 図 2 5 】



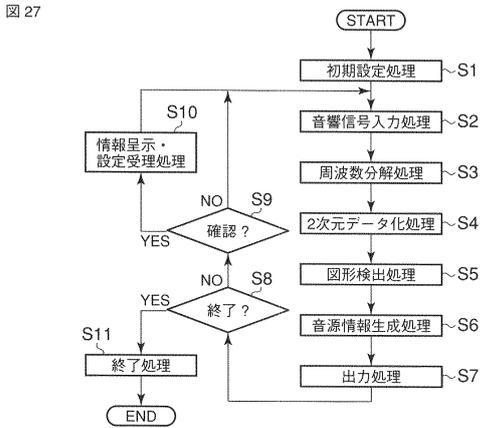
【 図 2 6 】



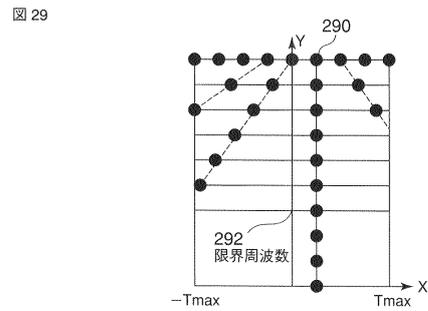
【 図 2 8 】



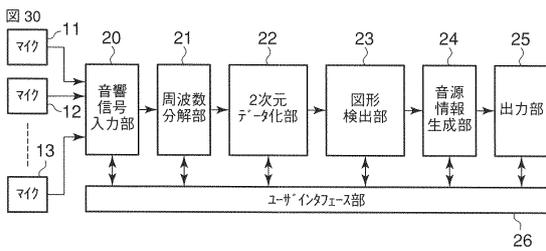
【 図 2 7 】



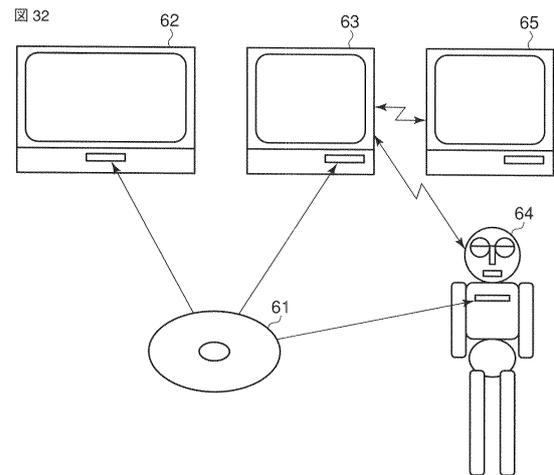
【 図 2 9 】



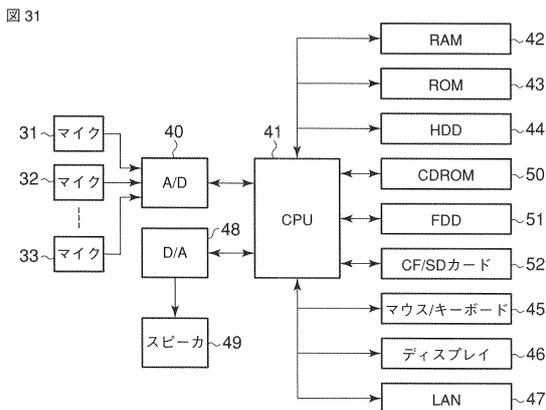
【 図 3 0 】



【 図 3 2 】



【 図 3 1 】



フロントページの続き

- (74)代理人 100084618
弁理士 村松 貞男
- (74)代理人 100092196
弁理士 橋本 良郎
- (72)発明者 鈴木 薫
神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内
- (72)発明者 古賀 敏之
神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

審査官 井出 和水

- (56)参考文献 特開2003-337164(JP,A)
特開平11-096374(JP,A)
特開2003-269937(JP,A)
特開2004-289762(JP,A)
特開2003-111183(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04R	1/20	-	H04R	1/40
H04S	1/00	-	H04S	7/00
G10L	15/28			
G10L	21/02			
G06T	7/00			
G06T	7/20	-	G06T	7/60