

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4234746号  
(P4234746)

(45) 発行日 平成21年3月4日(2009.3.4)

(24) 登録日 平成20年12月19日(2008.12.19)

(51) Int.Cl.		F I			
<b>H04R</b>	<b>3/00</b>	<b>(2006.01)</b>	H04R	3/00	320
<b>G10L</b>	<b>21/02</b>	<b>(2006.01)</b>	G10L	21/02	201D
			G10L	21/02	203Z

請求項の数 12 (全 34 頁)

(21) 出願番号	特願2006-259343 (P2006-259343)	(73) 特許権者	000003078
(22) 出願日	平成18年9月25日 (2006.9.25)		株式会社東芝
(65) 公開番号	特開2008-79255 (P2008-79255A)		東京都港区芝浦一丁目1番1号
(43) 公開日	平成20年4月3日 (2008.4.3)	(74) 代理人	100075812
審査請求日	平成18年9月25日 (2006.9.25)		弁理士 吉武 賢次
		(74) 代理人	100088889
			弁理士 橘谷 英俊
		(74) 代理人	100082991
			弁理士 佐藤 泰和
		(74) 代理人	100096921
			弁理士 吉元 弘
		(74) 代理人	100103263
			弁理士 川崎 康

最終頁に続く

(54) 【発明の名称】 音響信号処理装置、音響信号処理方法及び音響信号処理プログラム

(57) 【特許請求の範囲】

【請求項1】

それぞれ異なる位置に配置された複数のマイクによって得られる複数の音響信号を入力する音響信号入力手段と、

前記各音響信号を複数の周波数成分にそれぞれ分解し、前記各周波数成分毎に、信号レベル及び位相が対応付けられた周波数分解情報を前記音響信号毎に生成する周波数分解手段と、

前記複数の周波数分解情報のうち、所望の2つの前記周波数分解情報を比較することにより、対応する前記周波数成分毎に、前記2つの周波数分解情報間の位相差をそれぞれ算出する位相差算出手段と、

前記周波数成分を第1の軸とし、かつ前記位相差を第2の軸とする、2次元座標系上に配置された点群からなる2次元データを生成する2次元データ化手段と、

前記点群に対してそれぞれハフ変換を行って、ハフ投票空間に、前記点群それぞれに対応する複数の軌跡を生成し、前記複数の軌跡が通過する前記ハフ投票空間における位置に投票値を加算する際、前記2つの周波数分解情報がそれぞれ示す第1及び第2の前記信号レベルのレベル差に基づいて、前記投票値を変化させながら加算する投票手段と、

前記投票値が最大となる位置を探索することにより、前記2次元データの中から、探索された位置に対応する図形であって、かつ前記周波数成分及び前記位相差の比例関係を示し、前記各音響信号の音源方向を推定するための前記図形を検出する図形検出手段と

を備える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、

前記音源方向の角度を検出する際の分解能が、前記音源方向の角度を検出可能な範囲にわたって略同一になるように、前記周波数成分及び前記位相差の比例関係を示す前記図形を検出するための分解能を変化させる

ことを特徴とする請求項 1 に記載の音響信号処理装置。

## 【請求項 3】

前記位相差算出手段によって比較された前記 2 つの周波数分解情報を単位として、異なる前記単位毎に前記検出された図形を用いて、音源の候補に関する複数の音源候補情報を生成し、前記生成された前記複数の音源候補情報を対応付ける図形照合手段

をさらに備えることを特徴とする請求項 1 に記載の音響信号処理装置。

## 【請求項 4】

前記対応付けられた前記複数の音源候補情報に基づいて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

をさらに備えることを特徴とする請求項 3 に記載の音響信号処理装置。

## 【請求項 5】

それぞれ異なる位置に配置された複数のマイクによって得られる複数の音響信号を入力する音響信号入力ステップと、

前記各音響信号を複数の周波数成分にそれぞれ分解し、前記各周波数成分毎に、信号レベル及び位相が対応付けられた周波数分解情報を前記音響信号毎に生成する周波数分解ステップと、

前記複数の周波数分解情報のうち、所望の 2 つの前記周波数分解情報を比較することにより、対応する前記周波数成分毎に、前記 2 つの周波数分解情報間の位相差をそれぞれ算出する位相差算出ステップと、

前記周波数成分を第 1 の軸とし、かつ前記位相差を第 2 の軸とする、2 次元座標系上に配置された点群からなる 2 次元データを生成する 2 次元データ化ステップと

前記点群に対してそれぞれハフ変換を行って、ハフ投票空間に、前記点群それぞれに対応する複数の軌跡を生成し、前記複数の軌跡が通過する前記ハフ投票空間における位置に投票値を加算する際、前記 2 つの周波数分解情報がそれぞれ示す第 1 及び第 2 の前記信号レベルのレベル差に基づいて、前記投票値を変化させながら加算する投票ステップと、

前記投票値が最大となる位置を探索することにより、前記 2 次元データの中から、探索された位置に対応する図形であって、かつ前記周波数成分及び前記位相差の比例関係を示し、前記各音響信号の音源方向を推定するための前記図形を検出する図形検出ステップとを備えることを特徴とする音響信号処理方法。

## 【請求項 6】

前記図形検出ステップは、

前記音源方向の角度を検出する際の分解能が、前記音源方向の角度を検出可能な範囲にわたって略同一になるように、前記周波数成分及び前記位相差の比例関係を示す前記図形を検出するための分解能を変化させる

ことを特徴とする請求項 5 に記載の音響信号処理方法。

## 【請求項 7】

前記位相差算出手段によって比較された前記 2 つの周波数分解情報を単位として、異なる前記単位毎に前記検出された図形を用いて、音源の候補に関する複数の音源候補情報を生成し、前記生成された前記複数の音源候補情報を対応付ける図形照合ステップ

をさらに備えることを特徴とする請求項 5 に記載の音響信号処理方法。

## 【請求項 8】

前記対応付けられた前記複数の音源候補情報に基づいて音源情報を生成する音源情報生成ステップと、

10

20

30

40

50

前記音源情報を出力する出力ステップと  
をさらに備えることを特徴とする請求項 7 に記載の音響信号処理方法。

【請求項 9】

コンピュータに、

それぞれ異なる位置に配置された複数のマイクによって得られる複数の音響信号を入力する音響信号入力ステップと、

前記各音響信号を複数の周波数成分にそれぞれ分解し、前記各周波数成分毎に、信号レベル及び位相が対応付けられた周波数分解情報を前記音響信号毎に生成する周波数分解ステップと、

前記複数の周波数分解情報のうち、所望の 2 つの前記周波数分解情報を比較することにより、対応する前記周波数成分毎に、前記 2 つの周波数分解情報間の位相差をそれぞれ算出する位相差算出ステップと、

前記周波数成分を第 1 の軸とし、かつ前記位相差を第 2 の軸とする、2 次元座標系上に配置された点群からなる 2 次元データを生成する 2 次元データ化ステップと

、

前記点群に対してそれぞれハフ変換を行って、ハフ投票空間に、前記点群それぞれに対応する複数の軌跡を生成し、前記複数の軌跡が通過する前記ハフ投票空間における位置に投票値を加算する際、前記 2 つの周波数分解情報がそれぞれ示す第 1 及び第 2 の前記信号レベルのレベル差に基づいて、前記投票値を変化させながら加算する投票ステップと、

前記投票値が最大となる位置を探索することにより、前記 2 次元データの中から、探索された位置に対応する図形であって、かつ前記周波数成分及び前記位相差の比例関係を示し、前記各音響信号の音源方向を推定するための前記図形を検出する図形検出ステップと  
を実行させるための音響信号処理プログラム。

【請求項 10】

前記図形検出ステップは、

前記音源方向の角度を検出する際の分解能が、前記音源方向の角度を検出可能な範囲にわたって略同一になるように、前記周波数成分及び前記位相差の比例関係を示す前記図形を検出するための分解能を変化させる

ことを特徴とする請求項 9 に記載の音響信号処理プログラム。

【請求項 11】

前記位相差算出手段によって比較された前記 2 つの周波数分解情報を単位として、異なる前記単位毎に前記検出された図形を用いて、音源の候補に関する複数の音源候補情報を生成し、前記生成された前記複数の音源候補情報に対応付ける図形照合ステップ

をさらに前記コンピュータに実行させることを特徴とする請求項 9 に記載の音響信号処理プログラム。

【請求項 12】

前記対応付けられた前記複数の音源候補情報に基づいて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

をさらに前記コンピュータに実行させることを特徴とする請求項 11 に記載の音響信号処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は音響信号を処理する装置に関し、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を推定可能な装置に係る。

【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

10

20

30

40

50

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によれば、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を推定する方法が示されている。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。

10

【0004】

また、例えば、下記非特許文献2によれば、一对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が提案されている。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から成る周波数構造）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。

20

30

【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、従来技術は、（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を抱えており、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していない。

40

【0006】

本発明は上記の問題点に鑑みて成されたものであり、その目的とするところは、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、音響信号処理方法及び音響信号処理プログラムを提供することである。

【課題を解決するための手段】

【0007】

本発明の一態様による音響信号処理装置は、

50

それぞれ異なる位置に配置された複数のマイクによって得られる複数の音響信号を入力する音響信号入力手段と、

前記各音響信号を複数の周波数成分にそれぞれ分解し、前記各周波数成分毎に、信号レベル及び位相が対応付けられた周波数分解情報を前記音響信号毎に生成する周波数分解手段と、

前記複数の周波数分解情報のうち、所望の2つの前記周波数分解情報を比較することにより、対応する前記周波数成分毎に、前記2つの周波数分解情報間の位相差をそれぞれ算出する位相差算出手段と、

前記周波数成分を第1の軸とし、かつ前記位相差数を第2の軸とする、2次元座標系上に配置された点群からなる2次元データを生成する2次元データ化手段と、

前記点群に対してそれぞれハフ変換を行って、ハフ投票空間に、前記点群それぞれに対応する複数の軌跡を生成し、前記複数の軌跡が通過する前記ハフ投票空間における位置に投票値を加算する際、前記2つの周波数分解情報がそれぞれ示す第1及び第2の前記信号レベルのレベル差に基づいて、前記投票値を変化させながら加算する投票手段と、

前記投票値が最大となる位置を探索することにより、前記2次元データの中から、探索された位置に対応する図形であって、かつ前記周波数成分及び前記位相差の比例関係を示し、前記各音響信号の音源方向を推定するための前記図形を検出する図形検出手段とを備える。

【発明の効果】

【0008】

本発明によれば、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。また、算出される音源方向をより精度良く算出することができる。

【発明を実施するための最良の形態】

【0009】

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。

【0010】

(全体構成)

図1に本発明の1実施形態に係る音響信号処理装置の機能ブロック構成を示す。本実施形態装置は、3つ以上のn個のマイク1a~1c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、図形照合部6と、音源情報生成部7と、出力部8と、ユーザインタフェース部9とを有している。

【0011】

n個のマイク1a~1cは相異なる2つを対とする異なる組み合わせによる2つ以上のm組の対を成している。マイク1a~1cと音響信号入力部2を經由して入力されたnチャンネルの振幅データは、周波数解析部3によってそれぞれ周波数分解情報に変換される。2次元データ化部4は対を成す2つの周波数分解情報から周波数毎の位相差を求める。求められた周波数毎の位相差は2次元の(x, y)座標値を与えられて2次元データ化される。この2次元データは時系列に並べると時間軸を加えた3次元データとなる。図形検出部5は生成されたXY平面上の2次元データ、あるいは時間軸Tを加えたXYT空間中の3次元データを解析して所定の図形を検出する。この検出はm組の対に対してそれぞれ実行される。また、検出された図形はそれぞれが音源の存在を示唆する候補情報である。図形照合部6は検出された図形の情報を処理して、組の異なる音源候補のうち同一音源に由来するものを推定して対応付ける。音源情報生成部7は対応付けられた音源候補の情報を処理して、音源の数、各音源の空間的な存在範囲、各音源を発した音声の時間的な存在期間、各音源音声の成分構成、音源毎の分離音声、各音源音声の記号的内容を含む音源情報を生成し、出力部8はこれらの情報を出力し、ユーザインタフェース部9は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、及び各種情報や各種中間派生データの利用者への呈示を実行する。

10

20

30

40

50

## 【0012】

この音響信号処理装置は、少数の強い周波数成分を発する音源か、弱くても多数の周波数成分を発する音源であれば、人の声に限らず様々な音源を背景雑音の中から検出可能であり、またマイク数を越えた数の音源の検出が可能である。

## 【0013】

このとき、一对のマイクから音源候補として音源数の推定・音源方向の推定を行い、その結果を複数の対について照合・統合することで音源の方向だけでなくその空間位置を推定することを可能にする。また、1つの音源について複数のマイク対の中から適切なものを選ぶことで、単一のマイク対では悪条件となる方向の音源に対して好条件のマイク対によるデータから音源音声を品質良く抽出・認識することを可能にする。

10

## 【0014】

(周波数成分毎の位相差から音源を推定するという基本コンセプト)

マイク1a~1cは、空気などの媒質中に所定の距離をあけて配置されたn個のマイクロホンであり、異なるn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。n個のマイクは異なる2つを対として、異なる組み合わせによる2つ以上のm組のマイクロホン対(あるいはマイク対)を成す。

## 【0015】

音響信号入力部2は、マイク1a~1cによるnチャンネルの音響信号を所定のサンプリング周期Frで定期的にA/D変換することで、nチャンネル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。

20

## 【0016】

音源がマイクロホン間距離に比べて十分遠いと仮定するならば、図2に示すように、音源100を発してマイクロホン対に到達する音波の波面101はほぼ平面となっている。例えばマイク1aとマイク1bを用いて異なる2地点でこの平面波を観測すると、両マイクロホンを結ぶ線分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、両マイクロホン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。

## 【0017】

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、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の103と104)の間の到達時間差  $T$  を導き出す方法が開示されている。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

## 【0018】

そこで、本実施形態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。このようにすることで、複数の音源が存在しても、各音源に特有の周波数成分については、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ため、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な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

## 【0019】

(周波数分解部3)

50

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。

【 0 0 2 0 】

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110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111）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113ずつずらしながら繰り返す（T+1番目のフレーム112）。

【 0 0 2 1 】

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（図中120）を施された後、高速フーリエ変換（図中121）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]（図中122）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）を図中の124に示す。

【 0 0 2 2 】

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となり、k番目の周波数成分fkについてバッファ122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123上の点Pkを表すようになっている。このPkの原点0からの距離の2乗が当該周波数成分のパワーPo(fk)であり、Pkの実部軸からの符号付き回転角度 {  $\theta$  : -  $\pi$  <  $\theta$  <  $\pi$  } [ラジアン]が当該周波数成分の位相Ph(fk)となる。

【 0 0 2 3 】

サンプリング周波数がFr[Hz]、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[Hz]（直流）、k=(N/2) - 1がFr/2[Hz]（最も高い周波数成分）を表し、その間を周波数分解能  $f = (Fr/2) \div ((N/2) - 1)$  [Hz]で等分したものが各kにおける周波数となり、 $f_k = k \cdot f$ で表される。

【 0 0 2 4 】

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Fs）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成る周波数分解データを時系列的に生成する。

【 0 0 2 5 】

（2次元データ化部4と図形検出部5）

図5に示すように、2次元データ化部4は位相差算出部301と座標値決定部302とから成っており、図形検出部5は投票部303と直線検出部304とから成っている。

【 0 0 2 6 】

（位相差算出部301）

位相差算出部301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ab間位相差データを生成する手段である。図6に示すように、ある周波数成分fkの位相差 Ph(fk)は、マイク1aにおける位相値Ph1(fk)とマイク1bにおける位相値Ph2(fk)の差を計算し、その値が{  $\Delta \theta$  : -  $\pi$  <  $\Delta \theta$  <  $\pi$  }に収まるように2πの剰余系として算定する。

【 0 0 2 7 】

（座標値決定部302）

座標値決定部302は、位相差算出部301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XY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fkの位相差 Ph(fk)に対応するX座標値x(fk)とY座標値y(fk)は、図7に示す式によって決定される。X座標値は位相差Ph(fk)、Y座標値は周波数成分番号kである。

【 0 0 2 8 】

（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301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、同

10

20

30

40

50

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。このとき、FFTによって得られたある周波数の位相値及び両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なので、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倍となる比例関係にある。これを示すと図8のようになる。図8(a)に例示するように、同一時間Tに対して、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では1周期、すなわち2の位相区間を含む。位相差についても同様であり、同一時間差Tに対する位相差は周波数に比例して大きくなる。これを模式的に示したものが図8(b)である。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の上に並ぶことを示している。

10

Tが大きいほど、すなわち音源までの距離が両マイク間で異なるほど、この直線の傾きは大きくなる。

## 【0029】

（位相差の循環性）

但し、両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 $\pm\pi$ 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]以上とにならないことである。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

## 【0030】

手に入れることのできる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両マイクロホン間で1周期以上開いていても、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。図中、周波数 $f_k$ の位相差 $\text{Ph}(f_k)$ が図中の黒丸140で表すように+であるとき、1つ高い周波数 $f_{k+1}$ の位相差は図中の白丸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 $\text{Ph}(f_{k+1})$ は、本来の位相差から2を差し引いた、図中の黒丸142で表すように-よりやや大きい値となる。さらに、図示はしていないが、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

## 【0031】

（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、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ってこず、特に傾きの小さい直線155では3点しか直線上に配置されていない。

40

## 【0032】

本実施形態における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

50



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。そこで、本実施形態装置における2次元データ化部4の出力する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。

#### 【0033】

（投票部303）

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100～102ページに解説されているが、ここでもう一度説明する。

#### 【0034】

（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 $p(x, y)$ を通り得る直線は図中の160、161、162に例示するごとく無数に存在するが、原点0から各直線に下ろした垂線163のX軸からの傾きを、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り、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、 $(x, y)$ 座標値からそこを通り得る直線の( $,$ )の軌跡への変換を直線ハフ変換と云う。なお、直線が左に傾いているときは正值、垂直のとき0、右に傾いているとき負値であるとし、また、の定義域は{ : < }を逸脱することはない。

#### 【0035】

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 $p_1, p_2, p_3$ を共通に通る直線170は、 $p_1, p_2, p_3$ 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。

#### 【0036】

（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。これをハフ投票と云う。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。

#### 【0037】

10

20

30

40

50

投票部303は、次の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。

(投票条件1) 周波数が所定範囲にあるもの(低域カットと高域カット)

(投票条件2) 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

【0038】

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。

10

【0039】

暗騒音程度の非常に弱い周波数成分ではFFT結果の信頼性が低いと考えられる。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1aにおけるパワー値 $Po1(f_k)$ 、マイク1bにおけるパワー値 $Po2(f_k)$ 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。

(平均値) :  $Po1(f_k)$ と $Po2(f_k)$ の平均値

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。

(最小値) :  $Po1(f_k)$ と $Po2(f_k)$ の小さい方

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。

20

(最大値) :  $Po1(f_k)$ と $Po2(f_k)$ の大きい方

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。

【0040】

また、投票部303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。

(加算方式1) 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) 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。

【0041】

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。

30

【0042】

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13は、 $P(f_k)$ を $Po1(f_k)$ と $Po2(f_k)$ 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、 $Po1(f_k)$ と $Po2(f_k)$ 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 $P(f_k)$ の対数値 $\log_{10}(P(f_k))$ 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。

40

【0043】

(本実施の形態による音源定位(音源方向推定)処理)

50

ところで、音源方向推定処理を行う際、任意のフレームでマッピングした周波数 - 位相差空間に対してハフ変換を行い、投票空間に投票を行う段階において、例えば投票する値を一定値（最大値又は平均値）として投票を行うと、マイクロホン間における音声データの音量レベル差が大きい場合には、音源方向算出を誤るといった問題が生じる。

【0044】

これは、どちらのマイクで取得した音声レベルがどのくらい大きいのかという情報が反映されていないことが原因である。つまり上述の加算方式2を使用することにより周波数毎に投票値は異なるが、周波数が同じであればどの角度に対しても同じ投票値が投票されてしまう為、結果的に音声レベル差に関しての情報が音源方向推定処理の結果に対して反映されないということである。

10

【0045】

これに対して、本実施の形態では、IID（Interaural Intensity Defference）を導入し、音源方向を推定する。例えば2つのマイクロホンa、bからなるマイクアレイにおいて音源方向を推定するために、位相差 - 周波数空間の点をハフ変換を用いて投票空間へ投票するとき、その点を通る直線の傾きである値によって、投票する値を変化させる。

【0046】

この変化のパラメータとして、2つのマイクロホンa、bそれぞれで得られた音量レベル値を用いる。例えば、マイクロホンaがbより音量レベル値が大きければ、a寄りの方向を示す傾き値の時は投票値を大きく、b寄りの方向を示す傾き値のときは投票値を小さくすることで、ハフ変換による直線検出にIIDの要素を導入することができ、精度良く音源方向を推定することができる。

20

【0047】

なお、周波数 - 位相差空間における直線の傾きを示す値は、音源方向に対応し、当該直線の傾きを示す値に対して所定の演算処理を施すことにより、音源方向を算出することができる。

【0048】

ここで図14(a)を用いて、投票値を一定値として投票を行う場合における、ハフ変換を用いた音源定位の処理手順を説明する。

【0049】

まず、マイクアレイを構成する2つのマイク（マイクa、マイクbとする）に入力された音声波形データそれぞれに対してFFT処理を行い、各周波数における強度値（すなわち音量レベル値を示す信号レベル） $I_a(\omega_i)$ 、 $I_b(\omega_i)$ を得る。

30

【0050】

次に、任意の周波数 $\omega_i$ において、その周波数におけるマイクaとマイクbの強度値の平均値

【数1】

$$\frac{I_a(\omega_i) + I_b(\omega_i)}{2}$$

を算出し、これをハフ投票値 $V(\omega_i)$ とする。または、その周波数におけるマイクaとマイクbの強度値の最大値 $\max(I_a(\omega_i), I_b(\omega_i))$ を算出し、これをハフ投票値 $V(\omega_i)$ とする。

40

【0051】

これ以降は、ハフ変換による直線検出処理を、周波数 - 位相差空間に対して適用する。このとき投票値としては $V(\omega_i)$ を用いる。

【0052】

すなわち、周波数 $\omega_i$ と、当該周波数 $\omega_i$ におけるマイクaとマイクbの位相差値 $\phi(\omega_i)$ （FFT処理により算出済み）を基に、周波数 - 位相差空間内に一点が定まる。この周波数 - 位相差空間で定められた点を通る直線のうち、傾き $\theta$ が $-60^\circ$ から $60^\circ$ の範囲（ $2^\circ$ 刻み）における、61本の直線についてそれぞれ原点との距離 $r$ を算出し、 $r$  - 空間における61個の点 $(\theta, r)$ に対して、投票値 $V(\omega_i)$ を積算していく。なお、 $r$  - 空間内の各点

50

の投票値の初期値は0である。また、距離 を算出する場合には、予め求めておいた 値のテーブルから引用しても良い。

【 0 0 5 3 】

そして、全ての周波数  $\omega_i$  において、 $(\omega_i, \theta(\omega_i))$  から  $(\omega, \theta)$  へのハフ変換と、  
- 空間への投票 (投票値は  $V(\omega_i)$ ) とを行う。その後、音声入力後に関しては、A/D変換時の同時性が専用ボードによって保証されていることから、求める直線は必ず原点を通る ( $\theta = 0$ 、直流成分での位相差は0)。このため、 $\theta = 0$ の部分の得票値 ( $\theta$  軸上の値) を抜き出して  $\theta$  値に対する得票値列を作成する。ただし、位相差に関しては循環性 ( $\theta = \theta_0 + 2k\pi, k=0, \pm 1, \pm 2, \dots$ ) がある為、同じ  $\theta_0$  となる直線があれば、抜き出した得票値列に積算しておく。

10

【 0 0 5 4 】

この得票値列を用いて、最も投票値の高い点  $(\omega, \theta)$  を表す周波数 - 位相差空間の直線を、音源から到来した音声の周波数とマイクaとマイクbの位相差の関係を表す直線として求める。この関係が音源の方向を示している。また音源が2個以上あると考えられる場合は、第2位以下の投票値の高い点  $(\omega, \theta)$  を求めて、それぞれ対応する音源の方向を求める。

【 0 0 5 5 】

ところで、図14(b)に示すように、本実施の形態の場合には、投票値は周波数 - 位相差空間の直線の傾き によって変化する。ゆえに上記の計算式を用いて、 $\theta_a = -60^\circ$ ,  $\theta_b = 60^\circ$  であるとして、

20

【数2】

$$V(\omega_i, \theta) = \frac{I_b(\omega_i) - I_a(\omega_i)}{\theta_b - \theta_a} \cdot \theta + I_a(\omega_i) \quad \dots (1)$$

で投票値を求めることとなる。ただし、 $\theta_a = -60^\circ$   $\theta_b = 60^\circ$  ( $2^\circ$  刻み) とする。

【 0 0 5 6 】

なお、 $(\omega_i, \theta(\omega_i))$  から  $(\omega, \theta)$  へのハフ変換に関しては、上述の場合と同じ行程を踏む。投票時には、 $\theta$  - 空間における61個の点  $(\omega, \theta)$  に対して、投票値  $V(\omega_i, \theta)$  を積算していく。なお、 $\theta$  - 空間内の各点の初期値は0であるとする。ここで、 $V(\omega_i, \theta)$  は各  $\theta$  値に応じた値となるため、その都度計算することとなる。この場合、マイクbよりマイクaの方が大きい ( $I_a(\omega) > I_b(\omega)$ ) ため、マイクa側端が最も高く ( $I_a(\omega)$ )、徐々に投票値が下がっていき、マイクb側端では最低値となる  $I_b(\omega)$  を投票する。

30

【 0 0 5 7 】

ところで、図15に示すように、2つのマイクロホンによってマイクアレイを構成した場合に、2つのマイクロホンを結んだ直線 (ベースラインと呼ぶ)  $BL$  と直交する方向 (図中  $0^\circ$  方向) 近傍における方向角度検出分解能と、ベースライン  $BL$  方向近傍における方向角度検出分解能とが異なる。このため、音源位置によって角度の精度が異なる問題や、複数のマイクアレイを利用して音源定位を行う場合にも、その不均等性から最終的な精度に大きな影響を与えるという問題が生じる。

40

【 0 0 5 8 】

これに対して、本実施の形態では、最終的に算出される音源方向値  $\theta_{direct}$  の分解能を均等にするように、ハフ変換を行う際の  $\theta_{hough}$  値 (周波数 - 位相差空間における直線の傾き) の分解能を不均等にする。この  $\theta_{hough}$  と  $\theta_{direct}$  の間の関係は以下の通りに記述される。

【数3】

$$\theta_{direct} = \sin^{-1} \left( \frac{V}{d_{a-b}} \cdot \frac{1}{(f_s/2)} \cdot (-\tan \theta_{hough}) \cdot \frac{R_\omega}{R_{\Delta\phi}} \right) \quad \dots (2)$$

【 0 0 5 9 】

50

ただし、音速をV、マイクロホンa、b間の距離をd<sub>a-b</sub>、周波数をω<sub>i</sub>とし、括弧内の値が[-1,1]のときのみ考慮するものとする。また、f<sub>s</sub>は音声取得時のサンプリング周波数、R<sub>ω</sub>、R<sub>Δφ</sub>は位相差 - 周波数平面上のR<sub>ω</sub>、R<sub>Δφ</sub>の範囲（無次元化した後の範囲）とする。

【0060】

これをθ<sub>hough</sub>について逆展開して得られる以下の式、  
【数4】

$$\theta_{hough} = \tan^{-1} \left( -\frac{d_{a-b}}{V} \cdot (f_s/2) \cdot \sin \theta_{direc} \cdot \frac{R_{\Delta\phi}}{R_{\omega}} \right) \dots (3)$$

10

を用いて、θ<sub>direc</sub>が等間隔の場合に算出されるθ<sub>hough</sub>値を求め、これをハフ変換を行う際に用いる。これによって、投票によって最も得票が得られたθ<sub>hough</sub>値を用いて直線を決定した後に、式(3)を用いて算出される音源方向値θ<sub>direc</sub>が、等間隔に算出されるようになる。

【0061】

ここで図15(a)に、θ<sub>hough</sub>値の分解能が均等である場合を示す。この場合では、θ<sub>hough</sub>の範囲を-60° ~ 60°（2°刻み）と設定して計算している。この場合、正面方向を0°、右側を正、左側を負として、音源方向をθ<sub>direc</sub>で表すと、

【数5】

20

$$\theta_{direc} = \sin^{-1} \left( \frac{\left( \frac{\Delta\phi}{2\pi\omega_i} \right)}{\left( \frac{d_{a-b}}{V} \right)} \right), \frac{\Delta\phi}{2\pi} \cdot R_{\Delta\phi} = (-\tan \theta_{hough}) \cdot \frac{\omega_i}{(f_s/2)} \cdot R_{\omega} + 2k\pi \dots (4)$$

となる。ただし、f<sub>s</sub>は音声取得時のサンプリング周波数、R<sub>ω</sub>、R<sub>Δφ</sub>は位相差 - 周波数平面上のR<sub>ω</sub>、R<sub>Δφ</sub>の範囲（無次元化した後の範囲）とする（図15(c)参照）。

ここでk = 0とすると、

【数6】

30

$$\theta_{direc} = \sin^{-1} \left( \frac{V}{d_{a-b}} \cdot \frac{1}{(f_s/2)} \cdot (-\tan \theta_{hough}) \cdot \frac{R_{\omega}}{R_{\Delta\phi}} \right) \dots (5)$$

となり、θ<sub>hough</sub>とθ<sub>direc</sub>の関係式が求められる。これを逆に展開すると、

【数7】

$$\theta_{hough} = \tan^{-1} \left( -\frac{d_{a-b}}{V} \cdot (f_s/2) \cdot \sin \theta_{direc} \cdot \frac{R_{\Delta\phi}}{R_{\omega}} \right) \dots (6)$$

40

となる。

【0062】

これから、-90° ~ 90°（2°刻み）でθ<sub>hough</sub>を求めると、図15(b)に示すように、不均等な刻み幅のθ<sub>hough</sub>値列が得られる。すなわち、θ<sub>hough</sub>値の絶対値が大きい範囲における分解能が高くなる。

【0063】

このθ<sub>hough</sub>値を周波数 - 位相差空間における直線の傾きとして用いてθ<sub>hough</sub>を求め、投票を行い、最も投票値が高い点に対して抽出直線として出力する。これによって、θ<sub>hough</sub>値を、音源方向を示すθ<sub>direc</sub>値へ変換することで、均等に分割された解像度のθ<sub>direc</sub>値が得られる（図15(b)）。このθ<sub>hough</sub>値からθ<sub>direc</sub>値へ変換は、図形照合部6によ

50

って行われる。なお、図16に、`hough`と`direc`の関係を示す。

【0064】

(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( $m-1$ )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時刻の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。

【0065】

(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。

【0066】

図17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。

【0067】

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。図中の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。図中の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。

【0068】

( $\theta=0$ 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 $\theta$ 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17で例示したデータに対して $\theta$ 軸上で極大値を探索した結果を図18に示す。

【0069】

図中の190は図17における得票分布185と同一のものである。 $\theta$ 軸191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が図中の192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。直線検出部304は、得票分布 $H(\theta)$ 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 $H(\theta)$ 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図中193に示すように極大部の中央位置だけを極大位置として残す。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

10

20

30

40

50

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、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  $\theta = 0$  で定義される直線が図中の197である。なお、細線化処理のアルゴリズムは、ハフ変換の説明で紹介した参考文献2の89~92ページに記載される「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  $\theta$  の値を出力する。

【0070】

(位相差循環を考慮した直線群の定義)

ところで、図18で例示した直線197は、 $(0, 0)$ 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18の直線197が  $\theta = 0$  (図中の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  $(0, a)$  で定義される基準直線197を  $\theta = 0$  ずつ平行移動させた直線群  $(0, a)$  となる。さらに、起点となる  $\theta = 0$  について  $\theta = 0$  の制約をはずして  $\theta = 0$  として一般化すると、直線群は  $(0, a + 0)$  として記述できることになる。このとき、 $\theta$  は直線の傾きの関数  $\theta(\theta)$  として図19に示す式で定義される符号付きの値である。

【0071】

図中の200は  $(0, 0)$  で定義される基準直線である。このとき、基準直線が右に傾いているので定義に従い  $\theta$  は負値であるが、図ではその絶対値として扱う。図中の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  $\theta = 0$  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  $\theta$  も負値であるが、図ではその絶対値として扱う。図中の  $\triangle OQP$  は辺OQの長さが  $\theta$  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  $\theta$  であり、 $\triangle OUR$  の斜辺ORの長さが  $2\theta$  であることがわかる。このとき、 $\theta$  は辺OUの長さであるから、 $\theta = 2\theta \cos \theta$  となる。そして、 $\theta$  と  $\theta$  の符号を考慮すると図の計算式が導き出される。

【0072】

(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から成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  $\theta$  について  $\theta$  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。

【0073】

図20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。

【0074】

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

10

20

30

40

50

ワー値と位相値のデータに変換される。図中の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。図中の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た図である。この図で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、図の216は加算方式2を用いて生成された得票分布である。

10

【 0 0 7 5 】

図21は 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。図中の220は図20における得票分布216と同一のものである。 軸221上の得票分布S( , 0)をH( )として抜き出して棒グラフにしたものが図中の222である。この得票分布H( )には幾つか極大箇所（突出部）が存在しているが、総じて 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( )からは図中223に示す4つの極大位置224、225、226、227が検出される。このうち、唯一227だけが閾値以上の得票を得て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度左からの音声を検出したものであるが、マイクロホン対の正面約45度右からの音声を検出できていない。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 によって異なる（不公平がある）。そして、 = 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度右からの音声を検出できなかった理由である。

20

【 0 0 7 6 】

一方、図22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20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 の位置を破線242～249で表示したものである。このとき、 軸241と破線242～245、及び 軸241と破線246～249はそれぞれ ( )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 = 0には破線が存在しない。

30

【 0 0 7 7 】

ある 0の得票H( 0)は、 = 0の位置で縦に見たときの 軸241上の得票と破線242～249上の得票の合計値、すなわちH( 0) = {S( 0, a ( 0))}として計算される。この操作は = 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( )を棒グラフにしたものが図中の250である。図21の222と異なり、この分布では 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 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 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。

40

【 0 0 7 8 】

（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変換されない場合、検出されるべき直線は = 0、すなわちXY座標原点を通らない。この場合は = 0の制

50



約をはずして極大位置を探索する必要がある。

【0079】

$\theta = 0$ の制約をはずした基準直線を一般化して $(\theta, \theta)$ と記述すると、その直線群(基準直線と循環延長線)は $(\theta, a(\theta) + \theta)$ と記述できる。ここで $(\theta)$ は $\theta$ 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 $\theta$ 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 $\theta$ を変えたときの直線群の得票 $\{S(\theta, a(\theta) + \theta)\}$ が最大となる $\theta$ の値 $\theta_{max}$ を使って $(\theta, a(\theta) + \theta_{max})$ で与えられる。そこで、各 $\theta$ における得票 $H(\theta)$ をそれぞれの $\theta$ における最大得票値 $\{S(\theta, a(\theta) + \theta_{max})\}$ 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。

10

【0080】

(図形照合部6)

なお、検出された直線群はマイクロホン対毎に独立に推定された各時刻における音源の候補である。このとき、同一の音源を発生した音声は複数のマイクロホン対で同時刻にそれぞれ直線群として検出される。したがって、複数のマイクロホン対で同一の音源に由来する直線群を対応付けることができれば、より信頼性の高い音源の情報を得ることができるはずである。図形照合部6はそのための対応付けを行なう手段である。このとき、図形照合部6で直線群毎に編集された情報を音源候補情報と呼ぶことにする。

20

【0081】

図23に示すように、図形照合部6は、方向推定部311と、音源成分推定部312と、時系列追跡部313と、継続時間評価部314と、音源成分照合部315とから成っている。

【0082】

(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 $\theta$ 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候補数となる。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24で説明する。

30

【0083】

マイク1aとマイク1bの到達時間差 $T$ は $\pm T_{max}$ の範囲で変化し得る。図中(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 $\theta$ は正面を基準にした場合 $0^\circ$ となる。また、図中(b)のように音声が右真横、すなわちマイク1b方向から入射する場合、 $T$ は $+T_{max}$ に等しく、音源の方位角 $\theta$ 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、図中(c)のように音声が左真横、すなわちマイク1a方向から入射する場合、 $T$ は $-T_{max}$ に等しく、方位角 $\theta$ は $-90^\circ$ 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。

【0084】

以上を踏まえて図中(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、音声が線分PA方向から入射すると仮定すると、 $\triangle PAB$ 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 $0^\circ$ とした左回りを正にとる角度を方位角 $\theta$ と定義する。 $\triangle QOB$ は $\triangle PAB$ の相似形となるので、方位角 $\theta$ の絶対値は $\angle OBQ$ 、すなわち $\angle ABP$ に等しく、符号は $T$ の符号に一致する。また、 $\angle ABP$ 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 $\theta = \sin^{-1}(T/T_{max})$ 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、 $(90 - \theta)^\circ$ 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。

40

【0085】

図25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

50

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  $\theta$  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  $\theta$  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  $\phi$  は $k$  と  $\theta$  の関数として $k \cdot \tan(\theta)$  で求めることができる。このとき  $T$  [sec]は、位相差  $\phi$  ( $\phi$ ,  $k$ )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。  $\phi$  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24 ( d ) で音が右から入射する(位相差  $\phi$ が正值となる)とき、  $T$  は負値となる。また、図24 ( d ) で音が左から入射する(位相差  $\phi$ が負値となる)とき、  $T$  は正值となる。そのために、  $T$  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。

10

## 【 0 0 8 6 】

(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)を当該直線群(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。

## 【 0 0 8 7 】

(距離閾値方式による検出)

図26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。図中 ( 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 20 に対して異なる方向に存在している場合を示している。図中 ( a ) の280は1つの直線群を、図中 ( a ) の281と282は別の直線群を成している。図中 ( a ) の黒丸は周波数成分毎の位相差位置を表している。

## 【 0 0 8 8 】

直線群(280)に対応する音源音を構成する周波数成分は、図中 ( 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)と云うことにする。

## 【 0 0 8 9 】

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、図中 ( 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。

30

## 【 0 0 9 0 】

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。

## 【 0 0 9 1 】

(最近傍方式による検出)

図27は、図26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、図中 ( 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。

40

50

## 【0092】

(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成していた。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乗じることで、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。

## 【0093】

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乗じた値を、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28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。

## 【0094】

(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26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。

## 【0095】

(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(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

(1) N等分(距離閾値方式と最近傍方式に適用可能)

(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)

(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。

(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。

## 【0096】

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

10

20

30

40

50

ワー保存オプションを選択することが可能である。

【0097】

(時系列追跡部313)

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  $\theta$  は方向推定部311により計算される音源方向  $\theta$  と1対1に対応しているため、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  $\theta$ ) 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  $\theta$  (あるいは  $\theta$ ) 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。

10

【0098】

時系列追跡部313は、このように図形検出周期毎に求められる  $\theta$  を時間軸上で連続なグループに分けることで、  $\theta$  の時間軸上の軌跡を求める手段である。図29を用いてグループ分けの方法を説明する。

(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  $\theta$  値と  $\theta$  値(直線検出部304による)と、この直線群に対応した音源方向を表す  $\theta$  値(方向推定部311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312による)と、それらが取得された時刻とから成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。

20

(2) ある時刻 $T$ において、新しく得られた  $\theta$  のそれぞれ(以後  $n$ とし、図中では黒丸303と黒丸304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 $Kd$ (図中の矩形301と302)の直線群データ $Ld$ (図中の矩形内に配置された黒丸)を参照し、その  $\theta$  値と  $n$ の差(図中の305と306)が所定角度閾値  $\theta_{th}$  内にあり、かつその取得時刻の差(図中の307と308)が所定時間閾値  $t_{th}$  内にある $Ld$ を持つ軌跡データを検出する。この結果、黒丸303については軌跡データ301が検出されたが、黒丸304については最も近い軌跡データ302も上記条件を満たさなかったとする。

30

(3) 黒丸303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  $\theta$  値と  $\theta$  値と周波数成分と現時刻 $T$ とを当該軌跡 $Kd$ 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 $Te$ 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 $Ts$ 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 $Te$ 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303は軌跡データ301に追加される。

40

(4) 黒丸304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 $Ts$ と終了時刻 $Te$ を共に現時刻 $T$ とし、 $n$ とそれに対応した  $\theta$  値と  $\theta$  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 $Ln$ 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304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。

(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 $Te$ から)現時刻 $T$ までに前記所定時間  $t_{th}$  を経過したものがあれば

50

、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315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302がこれに該当する。

【0099】

(継続時間評価部314)

継続時間評価部314は、時系列追跡部313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超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  $\theta$  と  $\phi$  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314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。

【0100】

(音源成分照合部315)

音源成分照合部315は、時系列追跡部313と継続時間評価部314を経て異なるマイクロホン対に対してそれぞれ得られた音源ストリーム情報を、同一の音源に由来するもの同士で対応付けて音源候補対応情報を生成する。同一時刻において同一音源を発した音声は、その周波数成分において類似しているはずである。そこで、音源成分推定部312によって推定された直線群毎の各時刻の音源成分に基づいて、音源ストリーム間で同時刻の周波数成分のパターンを照合して類似度を計算し、所定閾値以上で最大の類似度を獲得した周波数成分のパターンを持つ音源ストリーム同士を対応付ける。このとき、音源ストリームの全域でパターン照合することも可能であるが、照合される音源ストリームが同時に存在する期間の数時刻の周波数成分パターンを照合して、その合計類似度、あるいは平均類似度が所定閾値以上で最大となるものを探索するようにすれば効率的である。この照合されるべき数時刻は、照合される両ストリームのパワーが共に所定閾値以上となる時刻とすることで、照合の信頼性はより向上することが期待できる。

【0101】

なお、図形照合部6の各機能ブロックは、図23に図示しない結線によって必要に応じて相互に情報のやりとりが可能であるものとする。

【0102】

(音源情報生成部7)

図30に示すように、音源情報生成部7は、音源存在範囲推定部401と、対選択部402と、同相化部403と、適応アレイ処理部404と、音声認識部405とから成っている。音源情報生成部7は、図形照合部6によって対応付けられた音源候補情報から、音源に関するより精密で信頼できる情報を生成する手段である。

【0103】

(音源存在範囲推定部401)

音源存在範囲推定部401は、図形照合部6により生成された音源候補対応情報に基づいて、音源の空間的存在範囲を算出する手段である。その算出方式には次の2つがあり、パラメータによって切り替え可能である。

(算出方式1) 同一音源に由来するとして対応付けられた音源ストリーム情報のそれぞれが示す音源方向を、それぞれの音源ストリームを検出したマイクロホン対の中点を頂点とする円錐面(図21のd)とみなし、対応付けられた全ての音源ストリームからそれぞれ得られる円錐面の交差する曲線もしくは点の所定近傍を音源の空間的な存在範囲として計算する。

(算出方式2) 同一音源に由来するとして対応付けられた音源ストリーム情報のそれぞれが示す音源方向を、最小二乗誤差で全て満たす空間中の点を算出し、音源の空間的な存在範囲として計算する。このとき、装置の原点を中心とする同心球面上の離散的な点について、各マイクロホン対への角度を計算したテーブルを用意することで、該角度と前記音

10

20

30

40

50

源方向との誤差の2乗和が最小となる点を前記テーブル上で探索する。

【0104】

(対選択部402)

対選択部402は、図形照合部6により生成された音源候補対応情報に基づいて、音源音声の分離抽出に最も適した対を選択する手段である。その選択方式には次の2つがあり、パラメータによって切り替え可能である。

(選択方式1)同一音源に由来するとして対応付けられた音源ストリーム情報のそれぞれが示す音源方向を比較して、最も正面寄りの音源ストリームを検出したマイクロホン対を選択する。この結果、最も正面から音源音声を捉えたマイクロホン対が音源音声の抽出に利用される。

10

(選択方式2)同一音源に由来するとして対応付けられた音源ストリーム情報のそれぞれが示す音源方向をそれぞれの音源ストリームを検出したマイクロホン対の中点を頂点とする円錐面(図24のd)とみなし、該円錐面から他の音源が最も遠い音源ストリームを検出したマイクロホン対を選択する。この結果、他の音源の影響の最も少ないマイクロホン対が音源音声の抽出に利用される。

【0105】

(同相化部403)

同相化部403は、対選択部402によって選択された音源ストリーム情報から、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  $\text{mid} = (\max + \min) / 2$  を計算して幅  $w = \max - \text{mid}$  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  $\text{mid}$  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。

20

【0106】

あるいは、方向推定部311による各時刻の音源方向を  $\text{mid}$  として、2つの周波数分解データ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。

【0107】

(適応アレイ処理部404)

適応アレイ処理部404は、抽出・同相化された2つの周波数分解データaとbの時系列データを、正面  $0^\circ$  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(周波数成分の時系列データ)を高精度に分離抽出する。なお、適応アレイ処理には、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2004，VOL.59，NO.9，2004」に開示されるように、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つ用いることで、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用いることができる。

30

【0108】

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、本実施形態装置では、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。

40

【0109】

また、このとき、事前に2つの周波数分解データ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。

50

## 【0110】

(音声認識部405)

音声認識部405は、適応アレイ処理部404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。

## 【0111】

(出力部8)

出力部8は、図形照合部6による音源候補情報として、図形検出部5による直線群の数として得られる音源候補の数、方向推定部311により推定された前記音響信号の発生源たる音源候補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312により推定された前記音源候補を発生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)、時系列追跡部313と継続時間評価部314による雑音源を除く音源候補(音源ストリーム)の数、時系列追跡部313と継続時間評価部314による前記音源候補(音源ストリーム)を発生した音声の時間的な存在期間、の少なくとも1つを含む情報、

もしくは、音源情報生成部7による音源情報として、図形照合部6による対応の付いた直線群(音源ストリーム)の数として得られる音源の数、音源存在範囲推定部401により推定された前記音響信号の発生源たる音源のより精密な空間的な存在範囲(円錐面の交差範囲やテーブル引きされた座標値)、対選択部402と正面化部403と適応アレイ部404による前記音源毎の分離音声(振幅値の時系列データ)、音声認識部405による前記音源音声の記号的な内容、の少なくとも1つを含む情報、を出力する手段である。

## 【0112】

(ユーザインタフェース部9)

ユーザインタフェース部9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17や図19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23や図24に示した(6)直線群に帰属する周波数成分の表示、図26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へ呈示したり、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。このようにすることで、利用者が本実施形態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実施形態装置を利用したりすることが可能になる。

## 【0113】

(処理の流れ図)

また、図31に本実施形態装置における処理の流れを示す。本実施形態装置における処理は、初期設定処理ステップS1と、音響信号入力処理ステップS2と、周波数分解処理ステップS3と、2次元データ化処理ステップS4と、図形検出処理ステップS5と、図形照合処理ステップS6と、音源情報生成処理ステップS7と、出力処理ステップS8と、終了判断処理ステップS9と、確認判断処理ステップS10と、情報呈示・設定受理処理ステップS11と、終了処理ステップS12とから成る。

## 【0114】

初期設定処理ステップS1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。

## 【0115】

音響信号入力処理ステップS2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。

## 【0116】

周波数分解処理ステップS3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

10

20

30

40

50

テップであり、前記音響信号入力処理ステップS2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。

【0117】

2次元データ化処理ステップS4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S3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XY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。

【0118】

図形検出処理ステップS5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S4による2次元データから所定の図形を検出する。

10

【0119】

図形照合処理ステップS6は、上述した図形照合部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S5により検出された図形を音源候補とし、音源候補を異なるマイク対の間で対応付けることで、同一音源に対する複数のマイク対による図形の情報（音源候補対応情報）を統合する。

【0120】

音源情報生成処理ステップS7は、上述した音源情報生成部7における処理を実行する処理ステップであり、前記図形照合処理ステップS6により統合された同一音源に対する複数のマイク対による図形の情報（音源候補対応情報）に基づいて、前記音響信号の発生源たる音源の数、各音源のより精密な空間的な存在範囲、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の内容、の少なくとも1つを含む音源情報を生成する。

20

【0121】

出力処理ステップS8は、上述した出力部8における処理を実行する処理ステップであり、前記図形照合処理ステップS6により生成された音源候補情報や、前記音源情報生成処理ステップS7により生成された音源情報を出力する。

【0122】

終了判断処理ステップS9は、上述したユーザインタフェース部9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、終了命令が有る場合には終了処理ステップS12へ（左分岐）、無い場合には確認判断処理ステップS10へ（上分岐）と処理の流れを制御する。

30

【0123】

確認判断処理ステップS10は、上述したユーザインタフェース部9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、確認命令が有る場合には情報呈示・設定受理処理ステップS11へ（左分岐）、無い場合には音響信号処理ステップS2（上分岐）と処理の流れを制御する。

【0124】

情報呈示・設定受理処理ステップS11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9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。

40

【0125】

終了処理ステップS12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9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

50



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。

【0126】

(効果)

前述の非特許文献2による方法は、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云える。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。

【0127】

また、非特許文献1による方法は、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。

10

【0128】

本実施の形態によれば、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。また、算出される音源方向をより精度良く算出することが出来る。

【0129】

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、異なる実施形態にわたる構成要素を適宜組み合わせてもよい。

20

【図面の簡単な説明】

【0130】

【図1】本発明の実施の形態による機能ブロック構成図。

【図2】音源方向と到達時間差の関係図。

【図3】フレームとフレームシフト量の関係図。

【図4】FFT処理と短時間フーリエ変換データの説明図。

【図5】2次元データ化部と図形検出部の内部構成図。

【図6】位相差算出の説明図。

【図7】座標値計算の説明図。

30

【図8】同一時間に対する周波数と位相の比例関係図。

【図9】位相差の循環性説明図。

【図10】複数音源存在時のプロット図。

【図11】直線ハフ変換の説明図。

【図12】複数の点を通る直線があれば軌跡が1点で交わることの説明図。

【図13】投票される平均パワーの関数値の説明図。

【図14】IIDを導入したハフ投票値の利用。

【図15】ハフ投票で投票する値と結果として出る実際の方向値の分布。

【図16】 $\theta_{\text{hough}}$ と $\theta_{\text{direc}}$ の関係式のグラフ。

【図17】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、ハフ投票結果を示した図。

40

【図18】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した図。

【図19】との関係図。

【図20】同時発話時の周波数成分、位相差プロット図、ハフ投票結果を示した図。

【図21】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。

【図22】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。

【図23】図形照合部の内部構成図。

【図24】方向推定の説明図。

【図25】とTの関係図。

50

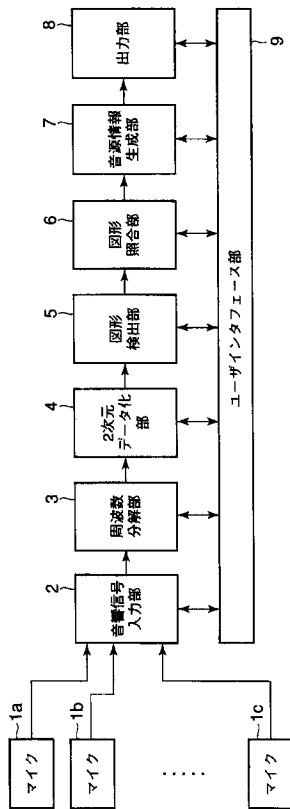
- 【図26】複数音源存在時の音源成分推定（距離閾値方式）の説明図。
- 【図27】最近傍方式の説明図。
- 【図28】 の計算式の例とそのグラフを示した図。
- 【図29】 の時間軸上の追跡を説明した図。
- 【図30】音源乗法生成部の内部構成図。
- 【図31】処理の流れを示した図。

【符号の説明】

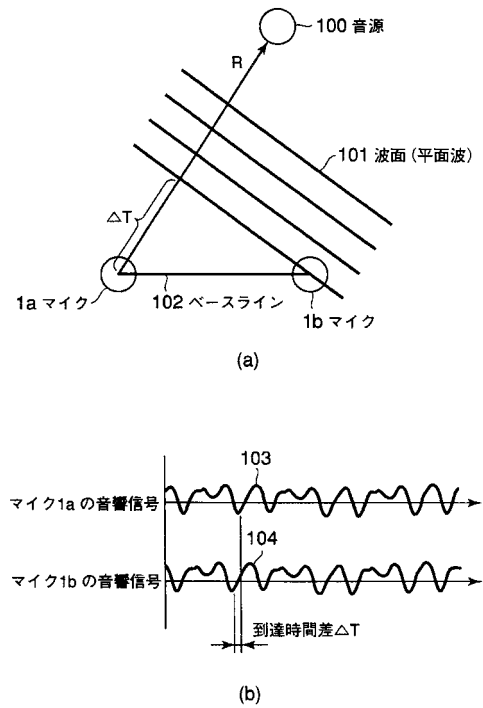
【0131】

- 1a・・・マイク
- 1b・・・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
- 8・・・ユーザインタフェース部

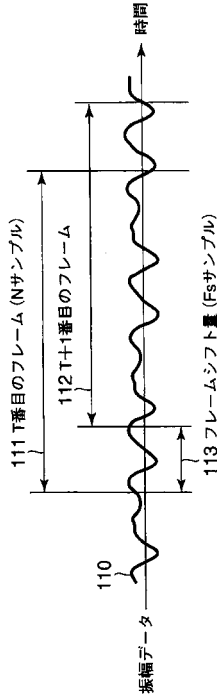
【図1】



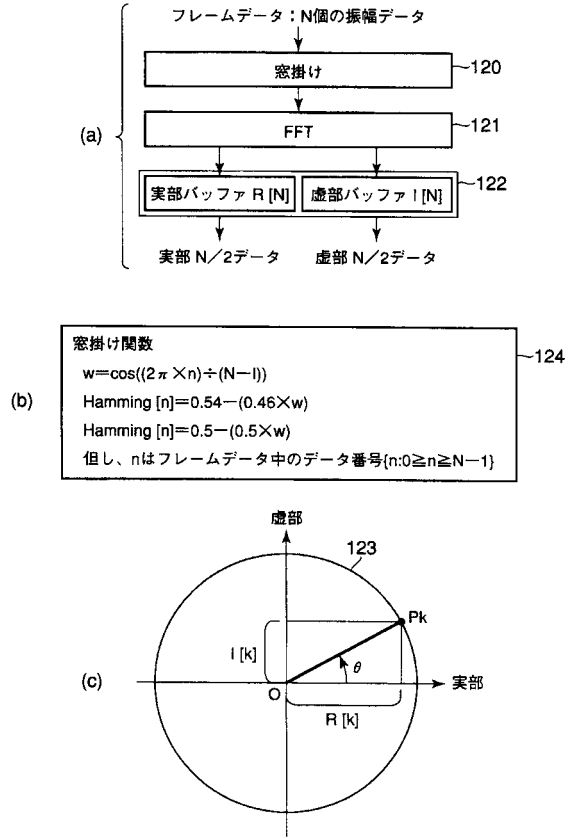
【図2】



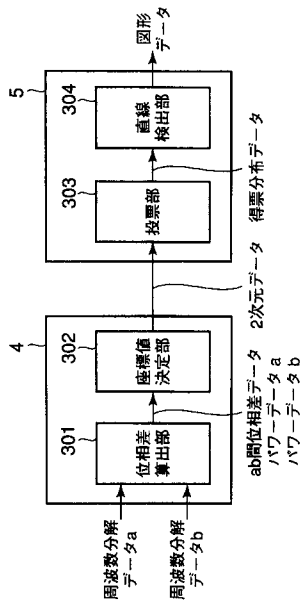
【図3】



【図4】



【図5】



【図6】

```

位相差
ΔPh(k) = Ph2(k) - Ph1(k);
while(1)
  if(ΔPh(k) ≤ -π) {ΔPh(k) = ΔPh(k) + 2π; continue;}
  break;
}
while(1)
  if(ΔPh(k) > π) {ΔPh(k) = ΔPh(k) - 2π; continue;}
  break;
}
但し、
Ph1(k)はマイク1aの周波数成分における位相値
Ph2(k)はマイク1bの周波数成分における位相値
ΔPh(k)の種値は、{ΔPh(k); -π < ΔPh(k) ≤ π}

```

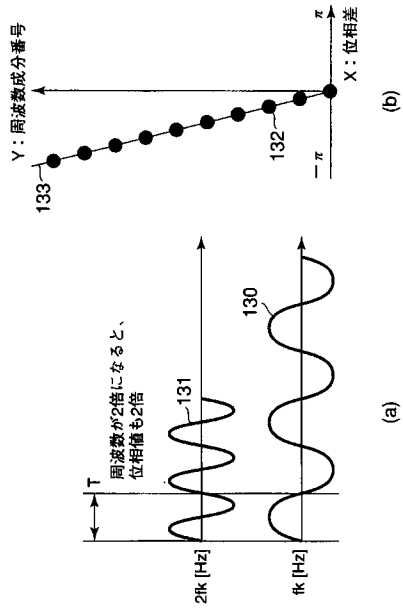
【図7】

```

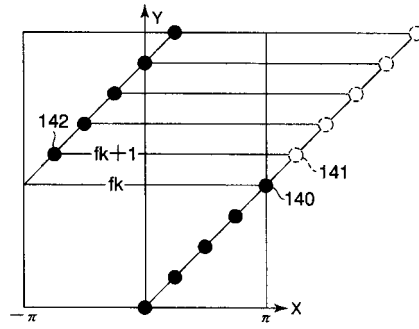
座標差
x(k) = ΔPh(k)
y(k) = k

```

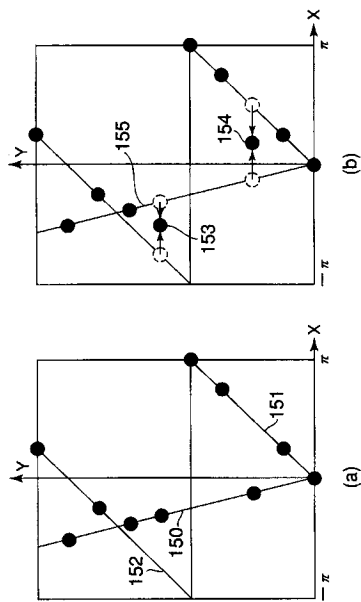
【 図 8 】



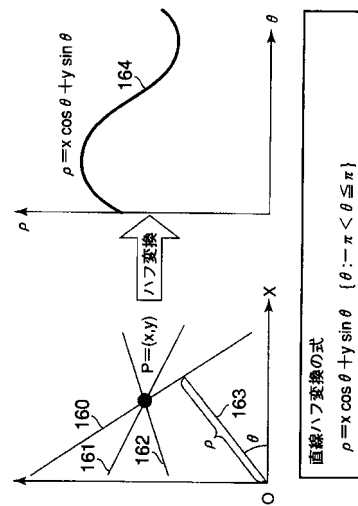
【 図 9 】



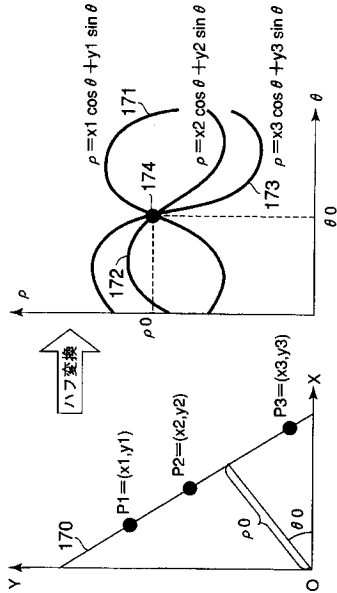
【 図 10 】



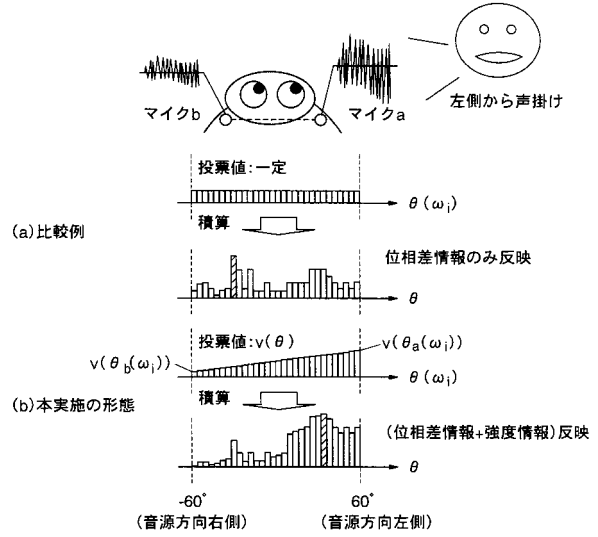
【 図 11 】



【 図 1 2 】



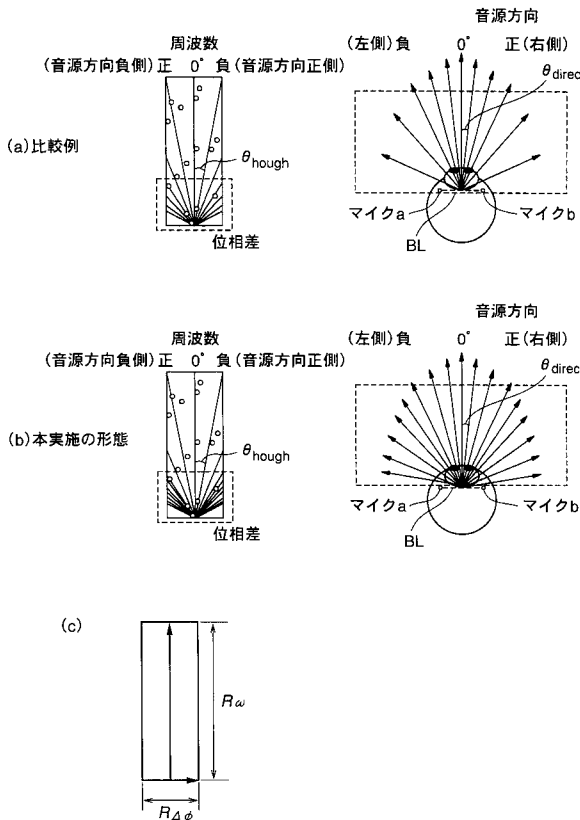
【 図 1 4 】



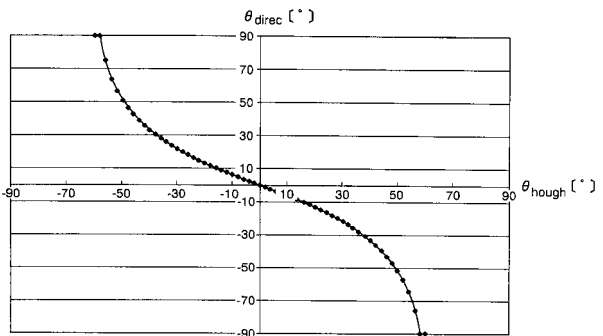
【 図 1 3 】

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

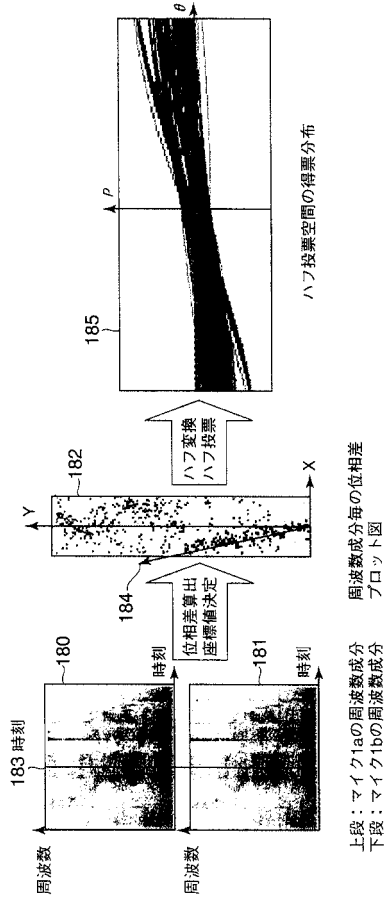
【 図 1 5 】



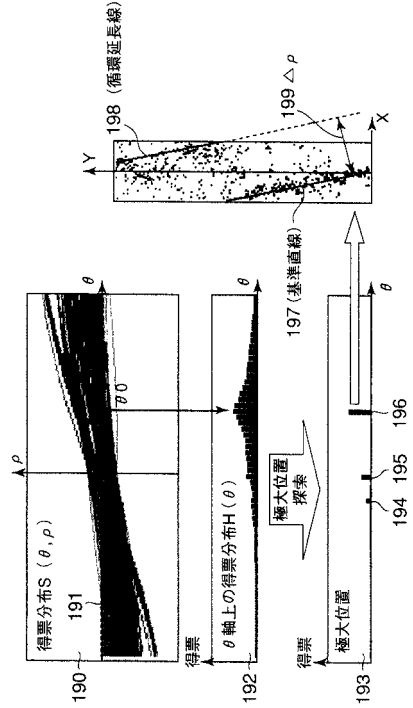
【 図 1 6 】



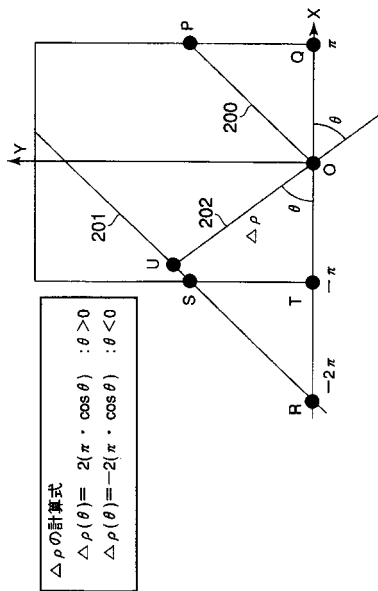
【図 17】



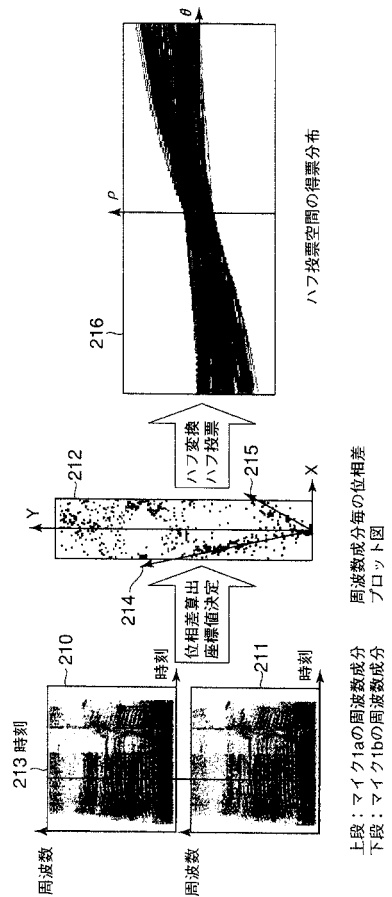
【図 18】



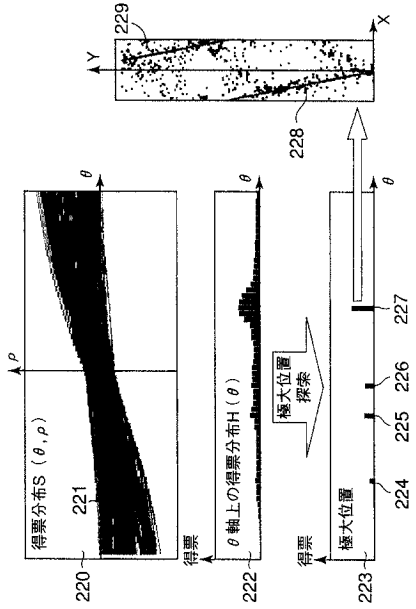
【図 19】



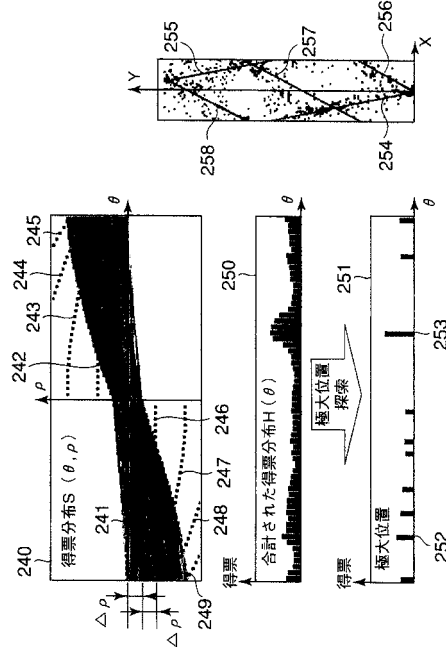
【図 20】



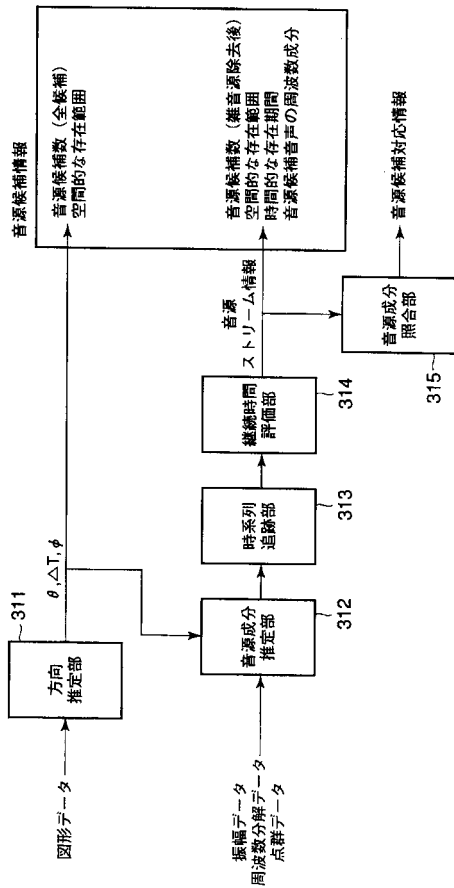
【図 2 1】



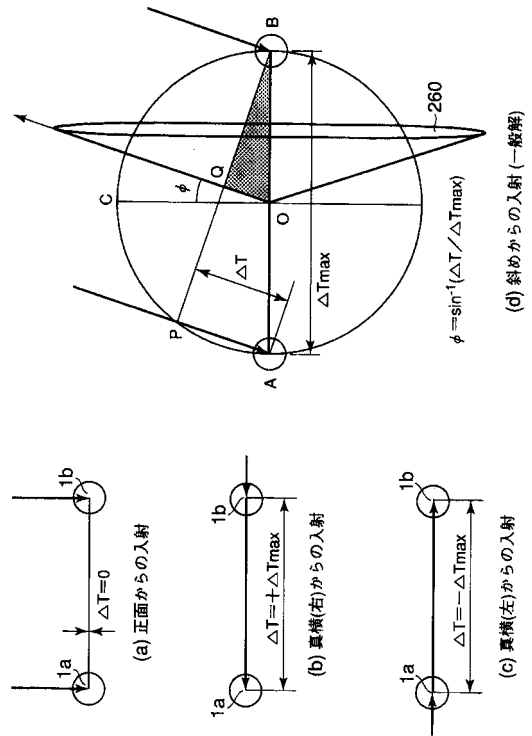
【図 2 2】



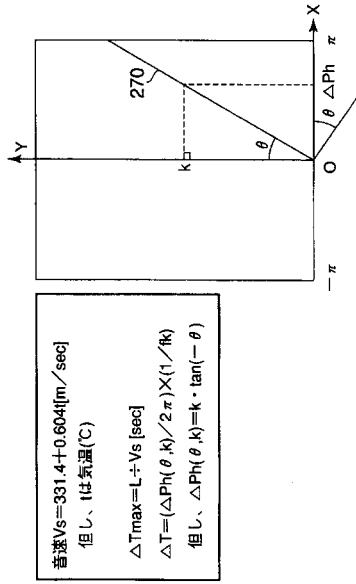
【図 2 3】



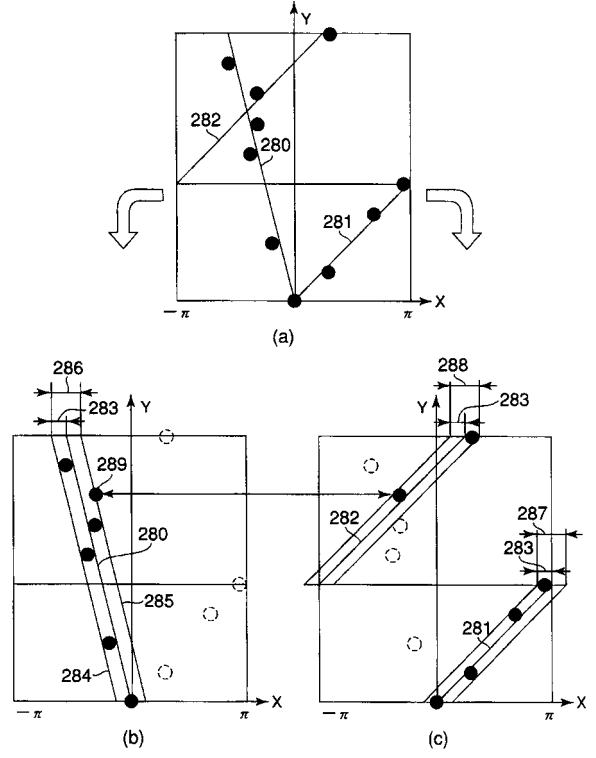
【図 2 4】



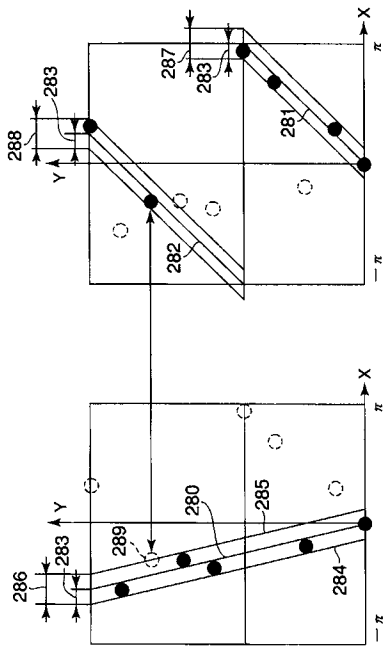
【 図 2 5 】



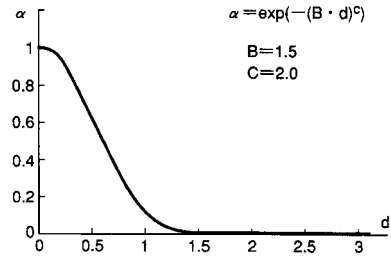
【 図 2 6 】



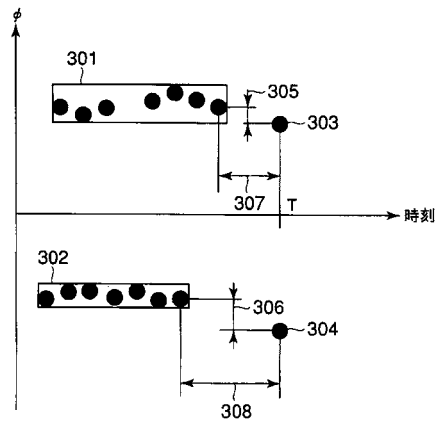
【 図 2 7 】



【 図 2 8 】

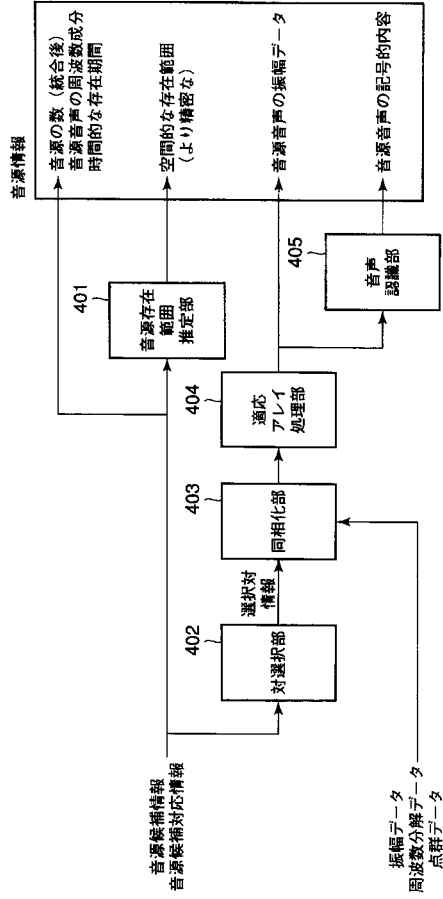


【 図 2 9 】

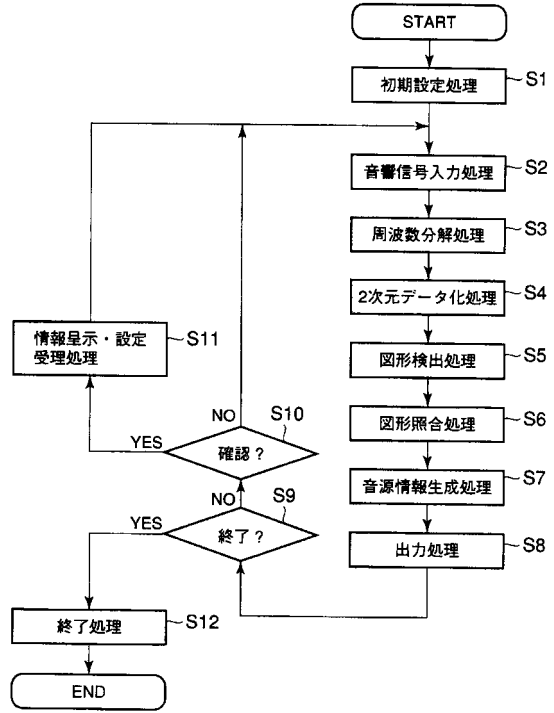




【図30】



【図31】



---

フロントページの続き

- (72)発明者 古賀敏之  
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 鈴木 薫  
東京都港区芝浦一丁目1番1号 株式会社東芝内

審査官 益戸 宏

- (56)参考文献 特開2006-254226(JP,A)  
特開2003-337164(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04R 3/00 - 3/14  
H04R 1/20 - 1/40  
G10L 15/28  
G10L 21/02  
H04S 1/00 - 7/00