

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5147012号
(P5147012)

(45) 発行日 平成25年2月20日(2013.2.20)

(24) 登録日 平成24年12月7日(2012.12.7)

(51) Int.Cl. F I
G 1 O L 25/78 (2013.01) G 1 O L 11/02
G 1 O L 15/04 (2013.01) G 1 O L 15/04 3 O O B

請求項の数 9 (全 24 頁)

(21) 出願番号	特願2008-214097 (P2008-214097)	(73) 特許権者	000004226
(22) 出願日	平成20年8月22日 (2008.8.22)		日本電信電話株式会社
(65) 公開番号	特開2010-49086 (P2010-49086A)		東京都千代田区大手町二丁目3番1号
(43) 公開日	平成22年3月4日 (2010.3.4)	(73) 特許権者	504132272
審査請求日	平成23年6月27日 (2011.6.27)		国立大学法人京都大学
			京都府京都市左京区吉田本町36番地1
		(74) 代理人	100121706
			弁理士 中尾 直樹
		(74) 代理人	100128705
			弁理士 中村 幸雄
		(74) 代理人	100147773
			弁理士 義村 宗洋
		(74) 代理人	100066153
			弁理士 草野 卓

最終頁に続く

(54) 【発明の名称】 目的信号区間推定装置、目的信号区間推定方法、目的信号区間推定プログラム及び記録媒体

(57) 【特許請求の範囲】

【請求項1】

目的信号区間を推定する目的信号区間推定装置であって、
 複数のセンサで観測された各信号をそれぞれ所定の時間区間であるフレーム毎に切り出す信号切出部と、

上記信号切出部で切り出された各フレームの信号を周波数領域に変換し、時間周波数ビン毎の周波数領域信号を各センサについて生成する周波数領域変換部と、

上記信号切出部で切り出された各フレームの信号の基本周波数をそれぞれ推定する基本周波数推定部と、

上記基本周波数又はその各倍音成分に最も近い前記時間周波数ビンを含む所定の時間軸方向及び所定の周波数軸方向の範囲内の1以上の前記時間周波数ビンを含む有限の時間周波数区間であるグリッドを上記フレーム毎に1以上ずつ特定し、各グリッドに属する各時間周波数ビンの上記周波数領域信号を抽出する時間周波数領域分割部と、

上記センサに含まれる特定の基準センサに対応する上記時間周波数領域分割部で抽出された上記周波数領域信号を基準として、少なくとも当該基準センサ以外の上記センサに対応する上記時間周波数領域分割部で抽出された各周波数領域信号を正規化し、上記センサで観測された信号の到来方向に対応する正規化信号値を時間周波数ビン毎に生成する正規化部と、

上記正規化信号値の上記グリッド毎の偏在性を示す偏在値を求め、当該グリッド毎の偏在値を用い、上記正規化信号値の上記フレーム毎の偏在性を示す偏在性指標値を算出する

10

20

偏在性指標値算出部と、

上記偏在性指標値を指標とし、各フレームが上記目的信号区間に対応するか否かを判定する判定部と、

を有することを特徴とする目的信号区間推定装置。

【請求項 2】

請求項 1 の目的信号区間推定装置であって、

上記偏在性指標値算出部は、

上記正規化信号値を量子化し、量子化された正規化信号値の頻度を上記グリッド毎に求め、上記グリッド毎のヒストグラムを生成するヒストグラム生成部と、

上記グリッド毎のヒストグラムを用い、当該ヒストグラムの分布の偏りを示す偏在値を上記グリッド毎に算出する偏在性計算部と、

同一のフレームに対応する各グリッドの上記偏在値を平均し、その平均値を当該フレームの上記偏在性指標値として算出する平均部と、

を有することを特徴とする目的信号区間推定装置。

【請求項 3】

請求項 2 の目的信号区間推定装置であって、

上記偏在性計算部は、

上記グリッド毎のヒストグラムを用い、上記量子化された各正規化信号値にそれぞれ対応する値を確率変数とする確率密度関数を上記グリッド毎に求める確率密度関数生成部と

、
上記確率密度関数のエントロピーに対して単調増加する関数値、又は、当該エントロピーに対して単調減少する関数値を、上記偏在値として求める偏在値算出部と、

を有することを特徴とする目的信号区間推定装置。

【請求項 4】

請求項 1 から 3 の何れかに記載の目的信号区間推定装置であって、

上記判定部は、

各フレームの上記偏在性指標値又はそれらの写像と、所定の閾値と、の大小を比較し、各フレームが上記目的信号区間であるか否かを判定する、

ことを特徴とする目的信号区間推定装置。

【請求項 5】

請求項 1 から 3 の何れかに記載の目的信号区間推定装置であって、

上記判定部は、

判定対象のフレームの上記偏在性指標値と、非目的信号区間のフレームの上記偏在性指標値と、の比である除算値又は当該除算値の写像が、所定の閾値以上であった場合に上記判定対象のフレームが上記目的信号区間であると判定するか、当該所定の閾値を超える場合に上記判定対象のフレームが上記目的信号区間であると判定する閾値判定部と、を有する、

ことを特徴とする目的信号区間推定装置。

【請求項 6】

請求項 1 から 3 の何れかに記載の目的信号区間推定装置であって、

上記判定部は、

事前学習されたフレームの上記偏在性指標値と当該フレームが上記目的信号区間であるか否かの判定結果との関係を用いたパターン認識により、上記偏在性指標値算出部で算出された上記偏在性指標値に対応するフレームが上記目的信号区間であるか否かを判定する、

ことを特徴とする目的信号区間推定装置。

【請求項 7】

目的信号区間を推定する目的信号区間推定装置の目的信号区間推定方法であって、

信号切出部が、複数のセンサで観測された各信号をそれぞれ所定の時間区間であるフレーム毎に切り出すステップと、

10

20

30

40

50

周波数領域変換部が、上記信号切出部で切り出された各フレームの信号を周波数領域に変換し、時間周波数ビン毎の周波数領域信号を各センサについて生成するステップと、

基本周波数推定部が、上記信号切出部で切り出された各フレームの信号の基本周波数をそれぞれ推定するステップと、

時間周波数領域分割部が、上記基本周波数又はその各倍音成分に最も近い前記時間周波数ビンを含む所定の時間軸方向及び所定の周波数軸方向の範囲内の1以上の前記時間周波数ビンを含む有限の時間周波数区間であるグリッドを上記フレーム毎に1以上ずつ特定し、各グリッドに属する各時間周波数ビンの上記周波数領域信号を抽出するステップと、

正規化部が、上記センサに含まれる特定の基準センサに対応する上記時間周波数領域分割部で抽出された上記周波数領域信号を基準として、少なくとも当該基準センサ以外の上記センサに対応する上記時間周波数領域分割部で抽出された各周波数領域信号を正規化し、上記センサで観測された信号の到来方向に対応する正規化信号値を時間周波数ビン毎に生成するステップと、

偏在性指標値算出部が、上記正規化信号値の上記グリッド毎の偏在性を示す偏在値を求め、当該グリッド毎の偏在値を用い、上記正規化信号値の上記フレーム毎の偏在性を示す偏在性指標値を算出するステップと、

判定部が、上記偏在性指標値を指標とし、各フレームが上記目的信号区間に対応するか否かを判定するステップと、

を有することを特徴とする目的信号区間推定方法。

【請求項8】

請求項1から6の何れかに記載の目的信号区間推定装置としてコンピュータを機能させるための目的信号区間推定プログラム。

【請求項9】

請求項8に記載の目的信号区間推定プログラムを格納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、信号処理技術に関し、特に、雑音が含まれる観測信号から目的信号が存在する区間を推定する技術に関する。

【背景技術】

【0002】

音声信号や音楽信号などの目的信号を処理対象とする符号化、雑音信号の抑圧、残響除去、自動音声認識などの音響信号処理技術では、複数種類の信号が含まれる入力音響信号から目的信号が存在する区間を推定する必要がある。この目的信号区間推定の精度はその後の信号処理性能に大きく影響する。

【0003】

環境雑音下での目的信号区間推定を行う際に複数のマイクロフォンが利用可能な場合には、目的音響信号の区間を推定するために信号の到達時間差の情報を利用できる。従来、目的信号の到来方向を既知として到来方向の信号を強調することで目的信号区間推定を行う手法（非特許文献1）や、零交差数などの音響特徴に対する閾値を目的信号の推定到来方向の信頼度によって決定する方法（非特許文献2）や、空間スペクトルのピークの有無で音声の有無を推定する手法（非特許文献3）や、推定信号到来方向が時間的に一定となる区間を音声の存在する区間とする手法（非特許文献4）などがある。しかしながら、これらの方法で十分な精度を得るためには、目的信号の到来方向が既知であったり、周辺環境が静かであったりする必要がある。

【0004】

また、複数のマイクロフォンを利用し、各マイクロフォンの信号に対してそれぞれ目的信号区間を推定した後、各マイクロフォンに対応する推定結果を比較して最終的な目的信号区間推定結果を得る手法が存在する（非特許文献5）。しかしながら、この方法では、

10

20

30

40

50

複数のマイクロフォンを利用することによる空間情報（目的信号の到来方向の情報）を十分利用できていなかった。

【0005】

その一方で、複数の音響信号が同時にあらゆる方向・あらゆる周波数帯域において到来するような環境（例えば街頭や駅・空港のような日常環境）で信号の到達時間差を用いて十分な目的信号区間推定精度を得るための手法として、一定範囲の時間周波数区間で推定した信号の到達時間差がある一定の値に偏る度合い（偏在性）を利用する手法（非特許文献6）がある。

【非特許文献1】 Alvarez, A., Gomez, P., Nieto, V., Martinez, R., and Rodellar, V., "Application of a first-order differential microphone for efficient voice activity detection in a car platform", Proceedings of Interspeech, 2669-2672, 2005.

10

【非特許文献2】 田中貴雅, 傳田遊亀, 中山雅人, 西浦敬信, "Weighted CSP法と音声特徴量に基づくハンズフリー発話区間検出の検討", 日本音響学会2006年度春期全国大会講演論文集, 1-P-3, pp. 149-150, Mar. 2006.

【非特許文献3】 山本潔, 浅野太, 吉村隆, 本村陽一, 麻生英樹, 原功, 市村直幸, 緒方淳, 北脇信彦, "音響情報と画像情報の統合による発話区間検出・分離システムの評価," 日本音響学会秋季研究発表会講演論文集, 3-6-10, P121-122, 2003.

【非特許文献4】 藤本雅清, 有木康雄, 堂下修司, "マルチモーダルインタラクションによるニュース映像中の人物認識," 日本音響学会誌, Vol. 62, No. 3, P182-192, 2006.

20

【非特許文献5】 荒木章子, 藤本雅清, 石塚健太郎, 澤田宏, 牧野昭二, 「音声区間検出と方向情報を用いた会議音声話者識別システムとその評価」, 日本音響学会春季研究発表会, pp. 1-4, 2008.

【非特許文献6】 Juan E. Rubio, Kentaro Ishizuka, Hiroshi Sawada, Shoko Araki, Tomohiro Nakatani, and Masakiyo Fujimoto, "Two-Microphone Voice Activity Detection Based on the Homogeneity of the Direction of Arrival Estimates," Proceedings of the 32nd International Conference on Acoustics, Speech, and Signal Processing, Vol. 4, pp. 385-388, 2007.

【発明の開示】

30

【発明が解決しようとする課題】

【0006】

しかし、非特許文献6の方法では、全ての時間周波数区間での偏在性を計算するため多大な計算量を要し、また方向性の雑音がある場合にはそれも目的信号として検出してしまうという問題点がある。

【0007】

本発明はこのような点に鑑みてなされたものであり、雑音環境下であって、なおかつ、目的信号の到来方向を正確に知ることが出来ない状況において、少ない計算量で精度よく目的信号区間を推定することが可能な技術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】

40

【0008】

本発明では上記課題を解決するために、まず、信号抽出部が、複数のセンサで観測された各信号をそれぞれ所定の時間区間であるフレーム毎に切り出し、周波数領域変換部が、信号切出部で切り出された各フレームの信号を周波数領域に変換し、時間周波数ビン毎の周波数領域信号を各センサについて生成する。また、基本周波数推定部が、信号切出部で切り出された各フレームの信号の基本周波数をそれぞれ推定し、時間周波数領域分割部が、基本周波数又はその各倍音成分をそれぞれ含む有限の時間周波数区間であるグリッドをフレーム毎に1以上ずつ特定し、各グリッドに属する各時間周波数ビンの周波数領域信号を抽出する。その後、正規化部が、センサに含まれる特定の基準センサに対応する時間周波数領域分割部で抽出された周波数領域信号を基準として、少なくとも当該基準センサ以

50

外のセンサに対応する各周波数領域信号を正規化し、センサで観測された信号の到来方向に対応する時間周波数領域分割部で抽出された正規化信号値を時間周波数ビン毎に生成する。そして、偏在性指標値算出部が、正規化信号値のグリッド毎の偏在性を示す偏在値を求め、当該グリッド毎の偏在値を用い、正規化信号値のフレーム毎の偏在性を示す偏在性指標値を算出する。

【0009】

ここで、本発明の正規化部が生成する正規化信号値は信号の到来方向に対応する値となる。通常、環境雑音は多様な方向からセンサに到来するのに対し、目的信号は或る方向のみからセンサに到来するという性質（性質1）を持つ。そのため、目的信号が存在しない時間周波数ビンの正規化信号値は広く分布する（偏在性が低い）のに対し、目的信号が存在する時間周波数ビンの正規化信号値は目的信号の到来方向に対応する値の付近に偏って分布する（偏在性が高い）。また、同一の目的信号の基本周波数又はその倍音成分（基本周波数の整数倍の周波数成分）はそれぞれ時間周波数領域で狭く分布するのに対し、雑音のパワーは時間周波数領域に広く分布する（性質2）。本発明では、これらの性質を利用し、基本周波数又はその倍音成分をそれぞれ含む有限の時間周波数区間であるグリッド毎の偏在性を示す偏在値を求め、当該グリッド毎の偏在値を用い、正規化信号値のフレーム毎の偏在性を示す偏在性指標値を算出する。これにより、目的信号区間を精度よく推定することができる。また、本発明では、各グリッドに属する各時間周波数ビンのみについて正規化信号値を求め、各グリッドのみについて偏在値を求めるため、全ての時間周波数区間で正規化信号値や偏在値を求める場合に比べ演算量を削減できる。なお、このように正規化信号値の偏在性を指標とする場合、目的信号の到来方向を正確に知る必要はない。よって、本発明では、目的信号の正確な到来方向を推定できない場合であっても、適切に目的信号区間を推定することができる。

【0010】

また、本発明において好ましくは、偏在性指標値算出部は、正規化信号値を量子化し、量子化された正規化信号値の頻度をグリッド毎に求め、グリッド毎のヒストグラムを生成するヒストグラム生成部と、グリッド毎のヒストグラムを用い、当該ヒストグラムの分布の偏りを示す偏在値をグリッド毎に算出する偏在性計算部と、同一のフレームに対応する各グリッドの偏在値を平均し、その平均値を当該フレームの偏在性指標値として算出する平均部と、を有する。

【0011】

このように、同一のフレームに対応する各グリッドの偏在値を平均し、その平均値を当該フレームの偏在性指標値として算出することにより、パワーや到来方向が時間周波数領域に広く分布する雑音成分の影響を低減させ、目的信号区間の推定精度を向上させることができる。

【0012】

また、本発明において好ましくは、ヒストグラム生成部は、重み係数を用い、量子化された正規化信号値の頻度を重み付けし、当該重み付けされた頻度を用いてグリッド毎のヒストグラムを生成する。この重み係数を適宜設定することにより、各環境に最適な目的信号区間推定方法が構築できる。

【0013】

また、本発明において好ましくは、偏在性計算部は、グリッド毎のヒストグラムを用い、量子化された各正規化信号値にそれぞれ対応する値を確率変数とする確率密度関数をグリッド毎に求める確率密度関数生成部と、確率密度関数のエントロピーに対して単調増加する関数値、又は、当該エントロピーに対して単調減少する関数値を、偏在値として求める偏在値算出部とを有する。このように偏在値を求めることにより、目的信号が存在する区間で小さな値をとり、目的信号が存在しない区間で大きな値をとる偏在性指標値、又は、目的信号が存在する区間で大きな値をとり、目的信号が存在しない区間で小さな値をとる偏在性指標値を生成することが可能となる。

【0014】

また、本発明において好ましくは、正規化部は、基準センサに対応する周波数領域信号の位相及び／又は振幅を基準とし、少なくとも当該基準センサ以外のセンサに対応する各周波数領域信号の位相及び／又は振幅を正規化し、当該正規化値又はその写像である正規化信号値を生成する。また、この場合に好ましくは、正規化信号値は、周波数成分が正規化され、周波数依存性が排除された値である。正規化信号値の周波数依存性が排除されていない場合、目的信号の時間周波数ビンにおける正規化信号値は、信号の到来方向と周波数とに依存した値となる。一方、正規化信号値の周波数依存性が排除されていた場合、目的信号の時間周波数ビンにおける正規化信号値は、信号の到来方向のみに依存した値となる。すなわち、同じ目的信号に対応する正規化信号値であったとしても、周波数依存性が排除された正規化信号値のほうが、周波数依存性が排除されていない正規化信号値よりも偏在性が高い。その結果、目的信号に起因する正規化信号値の偏在性がより明確に表れた偏在性指標値を得ることができ、偏在性指標値を指標として行われる目的信号区間の推定精度が向上する。

10

【0015】

また、本発明において判定部は、例えば、各フレームの偏在性指標値又はそれらの写像と、所定の閾値と、の大小を比較し、各フレームが目的信号区間であるか否かを判定する。また、判定部が、判定対象のフレームの偏在性指標値と、を非目的信号区間のフレームの偏在性指標値との比である除算値又は当該除算値の写像が、所定の閾値以上であった場合に判定対象のフレームが目的信号区間であると判定するか、当該所定の閾値を超える場合に判定対象のフレームが目的信号区間であると判定してもよい。また、例えば、判定部が、事前学習されたフレームの偏在性指標値と当該フレームが目的信号区間であるか否かの判定結果との関係を用いたパターン認識により、偏在性指標値算出部で算出された偏在性指標値に対応するフレームが目的信号区間であるか否かを判定してもよい。

20

【発明の効果】

【0016】

以上のように本発明では、雑音環境下であって、なおかつ、目的信号の到来方向を正確に知ることが出来ない状況において、少ない計算量で精度よく目的信号区間を推定することが可能となる。

【発明を実施するための最良の形態】

【0017】

以下、本発明を実施するための最良の形態を図面を参照して説明する。

図1は、本形態の目的信号区間推定装置10の全体構成を例示したブロック図である。また、図2は、図1の偏在性指標値算出部16の詳細構成を例示したブロック図である。また、図3は、図1の判定部17の詳細構成を例示したブロック図である。

30

【0018】

<構成>

図1に例示するように、本形態の目的信号区間推定装置10は、信号切出部11と周波数領域変換部12と基本周波数推定部13と時間周波数領域分割部14と正規化部15と偏在性指標値算出部16と判定部17と制御部18と記憶部19とを具備し、 $S(S-2)$ 個のセンサ20-1～ S で観測され、サンプリング部30でサンプリングされた信号が入力され、目的信号区間の分析結果を出力する装置である。また、図2(a)に例示するように、この例の偏在性指標値算出部16は、ヒストグラム生成部16aと確率密度関数計算部16bとエントロピー計算部16cと平均部16dを具備する。また、図2(b)に例示するように、この例の判定部17は、相対値計算部17aと尤度比算出部17bと閾値判定部17cとを具備する。

40

【0019】

なお、目的信号区間推定装置10は、例えば、CPU (central processing unit) , RAM (random access memory) , ROM (read only memory) 等から構成される公知のコンピュータに所定のプログラムを実行させることによって構成されるものである。

【0020】

50

< 処理 >

次に、本形態の目的信号区間推定方法について説明する。

本形態の目的信号区間推定方法では、複数のセンサ 20 - 1 ~ S (S 2) で観測された各信号を時間周波数分析し、特定の基準センサを基準とした正規化信号値を求め、所定の時間周波数区間であるグリッド内における正規化信号値の偏在性に基づいて、目的信号の有無を検出して出力する。なお、本形態では、複数のセンサ 20 - 1 ~ S としてマイクロフォンを利用し、それらで観測された各音響信号を用い、音声信号や音楽信号などの目的信号の有無を検出して出力する場合を例示する。また、以下では明記しないが、目的信号区間推定装置 10 は、制御部 18 の制御に基づいて各演算処理を実行し、各演算処理の過程で得られたデータは記憶部 19 に逐次格納され、それ以降の各演算処理に利用される。

10

【 0 0 2 1 】

図 4 は、本形態の目的信号区間推定方法を説明するためのフローチャートである。また、図 5 は、ステップ S 7 の詳細を説明するためのフローチャートであり、図 6 は、ステップ S 8 の詳細を説明するためのフローチャートである。以下、これらのフローチャートに沿って本形態の目的信号区間推定方法を説明していく。

【 0 0 2 2 】

まず、S (S 2) 個のセンサ 20 - 1 ~ S でそれぞれ観測された各信号がサンプリング部 30 に入力される。これらの信号には音声信号や音楽信号等目的信号の他に環境雑音信号も含まれる。サンプリング部 30 は、各信号を所定 (例えば 16,000Hz) のサンプリング周波数 f_s でサンプリングし、これにより各センサ 20 - 1 ~ S に対応する時間領域の信号 $x(1,t), \dots, x(S,t)$ を抽出する (ステップ S 1) 。なお、t は t 番目のサンプリング点を示す。

20

【 0 0 2 3 】

サンプリング部 30 で抽出された各時間領域の信号 $x(1,t), \dots, x(S,t)$ は、目的信号区間推定装置 10 の信号切出部 11 に入力される。信号切出部 11 は、入力された各信号 $x(1,t), \dots, x(S,t)$ を所定の時間区間であるフレーム毎に切り出し、各センサ 20 - 1 ~ S についての各フレーム i (i はフレームインデックスを示す) の信号 $x'(1,i,n), \dots, x'(S,i,n)$ を抽出する (ステップ S 2) 。なお、n はフレーム i における n 番目のサンプル点を表す。具体的には、信号切出部 11 は、例えば、入力された各信号 $x(1,t), \dots, x(S,t)$ に対し、それぞれ所定の窓関数を例えば時間軸方向に 16ms ずつ移動 (シフト) させながら乗じ、これにより、例えば 32ms の時間長の信号 $x'(1,i,n), \dots, x'(S,i,n)$ を切り出す。より具体的には、例えば、サンプリング周波数が 16,000Hz の場合、信号切出部 11 は、入力された各信号 $x(1,t), \dots, x(S,t)$ に対し、それぞれ例えば式 (1) のハニング窓を 256 サンプル点 (16,000Hz \times 16ms) ずつ移動 (シフト) させながら乗じ、センサ 20 - 1 ~ S 毎に、512 サンプル点 (16,000Hz \times 32ms) の離散信号を 1 フレーム分の信号として切り出す。ここで、L は切り出される 1 フレーム分の信号のサンプル点数 (フレーム長 : 上述の例では L=512) を表す。

30

【 0 0 2 4 】

【 数 1 】

$$w(n) = 0.5 - 0.5 \cdot \cos \frac{2\pi n}{L-1} \quad \dots (1)$$

40

信号切出部 11 は、以上のように切り出した各センサ 20 - 1 ~ S についての各フレーム i の信号 $x'(1,i,n), \dots, x'(S,i,n)$ を出力し、これらは周波数領域変換部 12 に入力される。

【 0 0 2 5 】

周波数領域変換部 12 では、各センサ 20 - 1 ~ S についての各フレーム i の信号 $x'(1,i,n), \dots, x'(S,i,n)$ を周波数領域に変換し、時間周波数ビン (i,k) 毎の周波数領域信号 (周波数領域スペクトル) $X(1,i,k), \dots, X(S,i,k)$ を各センサ 20 - 1 ~ S について生成

50

する（ステップS3）。離散フーリエ変換によってこの変換を行う場合、周波数領域変換部12は、以下の式(2)のように周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ を算出する。

【0026】

【数2】

$$X(s,i,k) = \sum_{n=0}^{M-1} x'(s,i,n) \cdot \exp\left(-j \frac{2\pi}{M} \cdot k \cdot n\right) \quad \dots (2)$$

ここで、 j は虚数単位を示し、 s ($s \in \{1, \dots, S\}$)は各センサ20-1~Sの番号を示す。また、 k ($k=0, \dots, M-1$)は周波数インデックスであり、サンプリング周波数 f_s を M 等分した離散点を表す。 M はフレーム長 L 以上の自然数であり、例えば、 $M=512$ とする。周波数領域変換部12は、以上のような変換によって得られた周波数領域信号（周波数スペクトル） $X(1,i,k), \dots, X(S,i,k)$ を出力する。

10

【0027】

また、信号切出部11で切り出された各センサ20-1~Sについての各フレーム i の信号 $x'(1,i,n), \dots, x'(S,i,n)$ は、基本周波数推定部13にも入力される。基本周波数推定部13は、各フレーム i の時間領域の信号 $x'(1,i,n), \dots, x'(S,i,n)$ を用い、各センサ s 及びフレーム i 毎の基本周波数 $F_0(1,i), \dots, F_0(S,i)$ をそれぞれ推定する（ステップS4）。この推定には、例えば以下の自己相関法（例えば、「Quatieri, T. F., "Discrete-time Speech Signal Processing principles and practice," Prentice-Hall, 2002; pp. 504-505」参照）を用いる。この場合、基本周波数推定部13は、まず、以下のように各センサ s 及びフレーム i 毎の自己相関関数の係数 $c(s,i,n)$ を $n=1, \dots, L$ について求める。

20

【0028】

【数3】

$$c(s,i,n) = \frac{1}{L} \sum_{\tau=1}^{L-1-n} x'(s,i,n) \cdot x'(s,i,n+\tau) \quad \dots (3)$$

なお、周波数領域変換部12が離散フーリエ変換によって周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ を算出する場合、基本周波数推定部13は、周波数領域変換部12の出力である各周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ の絶対値を2乗して逆フーリエ変換し、各自自己相関関数の係数 $c(s,i,n)$ を求めることもできる。

30

【0029】

次に、基本周波数推定部13は、各センサ s 及びフレーム i 毎に、 n の一定の探索範囲、例えば、 $32 \leq n \leq 320$ （サンプリング周波数 $f_s=16,000\text{Hz}$ の場合の50Hzから500Hzの周波数範囲に該当）の範囲内において、自己相関関数の係数 $c(s,i,n)$ が最大となる n を検出する。その結果得られた n は、入力信号 $x'(1,i,n), \dots, x'(S,i,n)$ の当該探索範囲において最も支配的な周期成分の周期長に対応し、入力信号 $x'(1,i,n), \dots, x'(S,i,n)$ がそれぞれ単一の完全な周期信号（例えば、正弦波）の場合にはその周期長に対応する。基本周波数推定部13は、サンプリング周波数 f_s を各センサ s 及びフレーム i 毎に得られた n で割ることで、各センサ s 及びフレーム i 毎の基本周波数 $F_0(1,i), \dots, F_0(S,i)$ を生成し、これら

40

【0030】

次に、時間周波数領域分割部14に、各基本周波数 $F_0(1,i), \dots, F_0(S,i)$ と、周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ とが入力される。時間周波数領域分割部14は、入力された各基本周波数 $F_0(1,i), \dots, F_0(S,i)$ 又はその各倍音成分（基本周波数の整数倍の周波数成分）をそれぞれ含む有限の時間周波数区間であるグリッドを各センサ s 及びフレーム i 毎に1以上ずつ特定し、各グリッドに属する各時間周波数ピンの周波数領域信号 $X_{GRID_1}(1,i,k), \dots, X_{GRID_G}(S,i,k)$ を抽出して出力する（ステップS5）。なお、センサ s のフレーム

50

i の各グリッドは、基本周波数 $F_0(s, i)$ 又はその各倍音成分の近傍にある一定の時間周波数区間であり、例えば、基本周波数 $F_0(s, i)$ に最も近い時間周波数ビンから所定の時間周波数範囲にある時間周波数区間、及び、基本周波数 $F_0(s, i)$ の各倍音成分にそれぞれ最も近い時間周波数ビンから所定の時間周波数範囲にある時間周波数区間である。例えば、基本周波数 $F_0(s, i)$ に最も近い周波数ビン及び基本周波数 $F_0(s, i)$ の各倍音成分にそれぞれ最も近い周波数ビン k' と表現すると($f=f_s \cdot k'/M$ が基本周波数 $F_0(s, i)$ 又はその各倍音成分に最も近くなる)、センサ s のフレーム i の各グリッドに属する各時間周波数ビンの周波数領域信号 $XGRID_g(s, i, k)$ ($g=1, \dots, G$) は以下のように表現できる。

$$XGRID_g(s, i, k) = \{X(s, i+P, k'+Q)\} \quad \dots (4)$$

【 0 0 3 1 】

【 数 4 】

$$\forall P \in \left\{ -\frac{A-1}{2}, \dots, +\frac{A-1}{2} \right\} \quad \dots (5)$$

$$\forall Q \in \left\{ -\frac{B-1}{2}, \dots, +\frac{B-1}{2} \right\} \quad \dots (6)$$

ここで、 A は時間方向の幅を示し、 B は周波数方向の幅を示し、 $\{ \cdot \}$ は \cdot を要素とする集合を意味する。 A 、 B には例えば $A=9$ 、 $B=5$ (時間幅160ms、周波数幅156.25Hz)を用いる。また、 G は基本周波数及びその各倍音成分の合算数を示す。 G は例えば定数である。

【 0 0 3 2 】

[好ましいグリッド幅の設定方法]

なお、前述したように、本発明では、正規化信号値のグリッド毎の偏在性を示す偏在値を求め、当該グリッド毎の偏在値を用い、正規化信号値のフレーム毎の偏在性を示す偏在性指標値を算出し、そのフレームが目的信号区間であるか否かの判定を行う。ここで、グリッドの時間周波数区間が広すぎるとグリッド内での正規化信号値の偏在性が平坦化され、偏在性から目的信号区間であるか非目的信号区間であるかを判定することが困難となる。逆にグリッドの時間周波数区間が狭すぎると、サンプル数が少ないため全てのグリッドでの正規化信号値の偏在性が高くなり、偏在性から目的信号区間であるか非目的信号区間であるかを判定することが困難となる。よって、グリッド幅はこのような問題が生じない範囲で設定する必要がある。以下に好ましいグリッド幅の設定方法について説明する。

【 0 0 3 3 】

《 式 (5) の A について 》

信号が音声信号の場合、およそ音声信号の定常性が仮定できる50~300 msの時間長に対応する A を決定すればよい。すなわち、フレームシフトの幅を SF msとすると、 $50/SF \sim 300/SF$ の間の整数値を A とすればよい。また、話者の発話速度 SR syllables/sec (1秒あたりに発話される音節数)が事前に分かるならば、 $(1000/SR)/SF$ 近傍の(例えば最も近い)整数値を A としてもよい(例えば、 $SR=7$ syllables/sec, $SF=16$ msならば、 $(1000/SR)/SF=(1000/7)/16=8.93$ なので、 $A=9$ とする、など)。また、対象信号が音楽信号ならば、音楽のリズム(音声の SR に対応)から同様に A を求める値を用いることが望ましい。

【 0 0 3 4 】

《 式 (5) の B について 》

基本的に、窓関数 $w(n)$ のメインローブ幅から得られる幅を用いるとよい。例えば、窓関数 $w(n)$ の離散フーリエ変換値を $W(k)$ とし、 $1 < k < M/2$ の範囲で $20 \log_{10}(W(k)/W(0)) > -60$ dBを満たす最大の周波数ビン k を cf とし、 $cf \cdot 2 + 1$ 近傍の(例えば最も近い)整数値を B とする。この値はサンプリング周波数 f_s 、分析フレーム長 L 、離散フーリエ変換の周波数ビンの総数 M に応じて変化する(例えば、サンプリング周波数 8kHz、窓関数の幅が256サンプル点、 $M=256$ ならば、 $cf=2$ となり、 $B=5$ となる)。

【 0 0 3 5 】

10

20

30

40

50

また、基本周波数推定部 1 3 で推定された基本周波数が $F_0(s, i)$ Hz であった場合には、一つのグリッドに 2 つ以上の音声信号の調波成分が入らないように、例えば $B=2 \cdot F_0(s, i) / (f_s/M)+1$ により定め、これが上記のメインローブ幅から求まる幅より大きくなる場合は、上記のメインローブ幅から求める値を採用することとしてもよい。例えば、サンプリング周波数 8kHz、窓関数の幅が 256 サンプル点、 $M=256$ のとき、 $F_0(s, i)=50$ Hz なら $B=2 \cdot 50 \cdot (8000/256)+1=4.2$ となるので、例えば $B=4$ とする。一方、 $F_0(s, i)=200$ Hz なら $B=2 \cdot 200 \cdot (8000/256)+1=13.8$ となるが、上記のメインローブ幅から求める値 $B=5$ よりも大きくなるため、 $B=5$ を採用する。これは、メインローブ幅の中でのみ音声信号の到来方向が偏在することによる。これらは目的信号が音楽信号である場合も同様である（「好ましいグリッド幅の設定方法」の説明終わり）。

10

【 0 0 3 6 】

次に、時間周波数領域分割部 1 4 から出力された各グリッドに属する各時間周波数ピンの周波数領域信号 $XGRID_1(1, i, k), \dots, XGRID_G(S, i, k)$ が正規化部 1 5 に入力される。正規化部 1 5 は、特定の基準センサ $s_B \{1, \dots, S\}$ に対応する時間周波数領域分割部 1 4 で抽出された周波数領域信号 $XGRID_1(s_B, i, k), \dots, XGRID_G(s_B, i, k)$ を基準として、少なくとも当該基準センサ s_B 以外のセンサ $s (s_B)$ に対応する時間周波数領域分割部 1 4 で抽出された各周波数領域信号 $XGRID_1(s, i, k), \dots, XGRID_G(s, i, k)$ を正規化し、センサで観測された信号の到来方向に対応する正規化信号値 $ZGRID_1(i, k), \dots, ZGRID_G(i, k)$ を時間周波数ピン (i, k) 毎に生成する（ステップ S 5）。なお、このような各正規化信号値 $ZGRID_1(i, k), \dots, ZGRID_G(i, k)$ は、目的信号が存在する時間周波数ピン (i, k) において、目的信号の到来方向に対応する値に偏ったものとなる。以下に正規化部 1 5 が生成する正規化信号値 $ZGRID_g(i, k) (g=1, \dots, G)$ の例を示す。

20

【 0 0 3 7 】

[正規化信号値 $ZGRID_g(i, k)$ の例]

本形態では正規化信号値 $ZGRID_g(i, k)$ の一例として、 $S=2$ とし、基準センサ 2 0 - 1 に対応する周波数領域信号 $XGRID_g(1, i, k)$ と、他方のセンサ 2 0 - 2 に対応する周波数領域信号 $XGRID_g(2, i, k)$ とから信号到来方向を推定し、その信号到来方向推定値を正規化信号値 $ZGRID_g(i, k)$ とする（正規化信号値 $ZGRID_g(i, k)$ の例 1）。この例では、正規化部 1 5 は、以下の式（7）（8）によって算出された信号到来方向 $\tau_g(i, k)$ を正規化信号値 $ZGRID_g(i, k)$ とする。なお、 v は音速（約 340m/秒）を示し、 d はセンサ間距離（m）を示し、 f は周波数ピン k に対応する離散周波数 $f=f_s \cdot k/M$ を示し、 $\arg(\cdot)$ は \cdot の位相（偏角）を示す。また、 $\tau_g(i, k)$ は信号源から各センサ 2 0 - 1, 2 までの信号到達時間差を示し、 $\theta_g(i, k)$ は信号到来方向推定値を示す。また、式（8）によって算出される信号到来方向 $\theta_g(i, k)$ は、センサ 2 0 - 1, 2 を結ぶ線分の中点を通り、その線分と直交する方向を 0 radian とした角度（radian）である。なお、このように算出された正規化信号値 $ZGRID_g(i, k)$ は、周波数成分 f が正規化され、周波数依存性が排除された値となる。

30

【 0 0 3 8 】

【 数 5 】

$$\tau_g(i, k) = \frac{\arg(XGRID_g(2, i, k) / XGRID_g(1, i, k))}{2 \cdot \pi \cdot f} \quad \dots (7)$$

40

$$\theta_g(i, k) = \arcsin\left(\frac{v \cdot \tau_g(i, k)}{d}\right) \quad \dots (8)$$

また、前述の式（7）で算出された信号到達時間差 $\tau_g(i, k)$ を正規化信号値 $ZGRID_g(i, k)$ としてもよい（正規化信号値 $ZGRID_g(i, k)$ の例 2）。なお、このように算出された正規化信号値 $ZGRID_g(i, k)$ も周波数成分 f が正規化され、周波数依存性が排除された値となる。

【 0 0 3 9 】

また、周波数領域信号 $XGRID_g(2, i, k)$ の位相に対する周波数領域信号 $XGRID_g(1, i, k)$ の位

50

相差 $\arg(XGRID_g(2, i, k)/XGRID_g(1, i, k))$ を正規化信号値 $ZGRID_g(i, k)$ としてもよいし(正規化信号値 $ZGRID_g(i, k)$ の例3)、周波数領域信号 $XGRID_g(2, i, k)$ の位相と周波数領域信号 $XGRID_g(1, i, k)$ の位相との差 $\arg(XGRID_g(2, i, k)) - \arg(XGRID_g(1, i, k))$ を正規化信号値 $ZGRID_g(i, k)$ としてもよい(正規化信号値 $ZGRID_g(i, k)$ の例4)。さらに、周波数領域信号 $XGRID_g(2, i, k)$ 振幅に対する周波数領域信号 $XGRID_g(1, i, k)$ の振幅の比 $|XGRID_g(2, i, k)| / |XGRID_g(1, i, k)|$ を正規化信号値 $ZGRID_g(i, k)$ としてもよいし(正規化信号値 $ZGRID_g(i, k)$ の例5)、周波数領域信号 $XGRID_g(1, i, k)$ のパワーに対する周波数領域信号 $XGRID_g(2, i, k)$ のパワーの比 $|XGRID_g(2, i, k)|^2 / |XGRID_g(1, i, k)|^2$ を正規化信号値 $ZGRID_g(i, k)$ としてもよい(正規化 $ZGRID_g(i, k)$ の例6)。何れの場合も、目的信号が存在する時間周波数ビン (i, k) においてのみ、目的信号の到来方向に対応する値に偏った値を取るため、正規化信号値 $ZGRID_g(i, k)$ の偏在性を指標として目的信号が存在するか否かを判定できる。

10

【0040】

また、上記ではセンサの数が2つの場合を例示したが、センサの数が3以上の場合は、例えば以下のように、目的信号の到来方位角推定値 $\theta_g(i, k)$ と仰角推定値 $\phi_g(i, k)$ とを求め、それら2つの値を時間周波数ビン (i, k) に対する正規化信号値 $ZGRID_g(i, k)$ としてもよい(正規化信号値 $ZGRID_g(i, k)$ の例7)。

【0041】

まず、各センサ $20-s$ ($s=1, \dots, S$)の空間中の座標ベクトルを $d_s = [x \text{ 座標}, y \text{ 座標}, z \text{ 座標}]$ とする。また、 J ($J(1, \dots, S)$)番目のセンサ $20-J$ を基準センサとし、基準センサ $20-J$ と各センサ $20-s$ との距離ベクトル D を以下の式(9)のように設定

20

【0042】

$$D = [d_1 - d_J, d_2 - d_J, \dots, d_S - d_J]^T \quad \dots (9)$$

また、基準センサ $20-J$ と各センサ $20-s$ との信号到達時間差 $\tau_g(s, i, k)$ を以下の式(10)で求め、それらを要素とした信号到達時間差ベクトル $\tau_g'(i, k)$ を以下の式(11)のように求める。

【0043】

【数6】

$$\tau_g(s, i, k) = \frac{\arg(XGRID_g(s, i, k)/XGRID_g(J, i, k))}{2 \cdot \pi \cdot f} \quad \dots (10)$$

30

$$\tau_g'(i, k) = [\tau_g(1, i, k), \tau_g(2, i, k), \dots, \tau_g(S, i, k)]^T \quad \dots (11)$$

上述の式(9)~(11)には以下の式(12)の関係が成り立ち、以下の式(12)から目的信号の到来方位角推定値 $\theta_g(i, k)$ と仰角推定値 $\phi_g(i, k)$ とを求める。なお、式(12)における D^{-1} はムーア・ペンローズ型一般化逆行列などの一般化逆行列である。また、目的信号の到来方位角とは $x-y$ 平面上の目的信号の到来方向を意味し、目的信号の仰角とは $x-z$ 平面上の目的信号の到来方向を意味する。また、 y 軸線方向が0radianである。

【0044】

$$D^{-1} \cdot \tau_g'(i, k) = [\cos \theta_g(i, k) \cos \phi_g(i, k), \sin \theta_g(i, k) \cos \phi_g(i, k), \sin \theta_g(i, k) \sin \phi_g(i, k)]^T \quad \dots (12)$$

40

また、上述の正規化信号値 $Z(i, k)$ の例1~7で例示した正規化信号値 $ZGRID_g(i, k)$ を組み合わせ、時間周波数ビン (i, k) 毎に2以上の正規化信号値 $ZGRID_g(i, k)$ 算出する構成であってもよい(正規化信号値 $ZGRID_g(i, k)$ の例8)。例えば、位相差 $\arg(XGRID_g(2, i, k)/XGRID_g(1, i, k))$ と振幅の比 $|XGRID_g(2, i, k)| / |XGRID_g(1, i, k)|$ との組を時間周波数ビン (i, k) の正規化信号値 $ZGRID_g(i, k)$ としてもよい。また、例えば $S=3$ とし、位相差 $\arg(XGRID_g(2, i, k)/XGRID_g(1, i, k))$ と振幅の比 $|XGRID_g(3, i, k)| / |XGRID_g(1, i, k)|$ との組を時間周波数ビン (i, k) の正規化信号値 $ZGRID_g(i, k)$ としてもよい。また、上述のように生成した値の写像を正規化信号値 $ZGRID_g(i, k)$ としてもよい([正規化信号値 $ZGRID_g(i, k)$ の例] の説

50

明終わり)。

【0045】

以上のようにステップS6では、正規化部15が上述のような正規化信号値 $ZGRID_1(i,k), \dots, ZGRID_G(i,k)$ を生成し、出力する。

【0046】

正規化部15から出力された各正規化信号値 $ZGRID_1(i,k), \dots, ZGRID_G(i,k)$ は、偏在性指標値算出部16に入力される。偏在性指標値算出部16は、各正規化信号値 $ZGRID_1(i,k), \dots, ZGRID_G(i,k)$ のグリッド毎の偏在性を示す偏在値 $H_1(i,k), \dots, H_G(i,k)$ を求め、当該グリッド毎の偏在値 $H_1(i,k), \dots, H_G(i,k)$ を用い、正規化信号値のフレームi毎の偏在性を示す偏在性指標値 $H(i)$ を算出する(ステップS7)。以下にステップS7の詳細を例示する。

10

【0047】

[ステップS7の例]

この例の場合、まず、偏在性指標値算出部16のヒストグラム生成部16a(図2)が、入力された各正規化信号値 $ZGRID_1(i,k), \dots, ZGRID_G(i,k)$ をそれぞれC個の値 $Z(c)$ ($c=1, \dots, C$)に量子化し、量子化された正規化信号値 $Z(c)$ の頻度 $bin_1(i,k,c), \dots, bin_G(i,k,c)$ ($c=1, \dots, C$)をグリッド毎に求め、グリッド毎のヒストグラムを生成する(ステップS71)。例えば、正規化信号値 $ZGRID_g(i,k)$ が信号到来方向 $\theta_g(i,k)$ であり、 $C=32$ であった場合、各正規化信号値 $ZGRID_g(i,k)$ は以下のようなC個の正規化信号値 $Z(c)$ に量子化される。

20

【0048】

$$Z(1) \quad (-1/2 \leq ZGRID_g(i,k) < -7/16)$$

$$Z(2) \quad (-7/16 \leq ZGRID_g(i,k) < -3/16)$$

...

$$Z(C) \quad (7/16 \leq ZGRID_g(i,k) < 1/2)$$

なお、前述の式(7)で算出された信号到達時間差 $\tau_g(i,k)$ を正規化信号値 $ZGRID_g(i,k)$ とした場合には、ヒストグラム生成部16aは、例えば $| \tau_g(i,k) | \cdot (d/c) \times ($ (は正の定数)の単位で正規化信号値 $ZGRID_g(i,k)$ をC個に量子化する。

【0049】

そして、ヒストグラム生成部16aは、時間周波数ビン (i,k) 毎に正規化信号値 $ZGRID_g(i,k)$ がいずれの正規化信号値 $Z(c)$ に対応するかを判断し、その頻度をグリッド毎にカウントし、グリッド毎のヒストグラムを生成する。この際、ヒストグラム生成部16aが、或る重み係数を用い、量子化された正規化信号値 $Z(c)$ の頻度を重み付けし、当該重み付けされた頻度を用いてグリッド毎のヒストグラムを生成してもよい。例えば、ヒストグラム生成部16aが、頻度のカウントの際、対応する時間周波数ビン (i,k) の重み係数 $W(i,k)$ で頻度を重み付けしてもよい。より具体的には、例えば、時間周波数ビン $(1,2)$ の正規化信号値 $ZGRID_g(i,k)$ を量子化した値が $Z(5)$ であった場合、 $Z(5)$ に対する頻度として $W(1,2)$ をカウントする。すなわち、正規化信号値 $ZGRID_g(i,k)$ に対する、量子化された正規化信号値 $Z(c)$ の頻度 $bin_g(i,k,c)$ ($c=1, \dots, C$)は、例えば、以下の式(13)のようにカウントされてもよい。

30

40

【0050】

$$bin_g(i,k,c) = W(i+k, k+Q) \quad \text{if } ZGRID_g(i+k, k+Q) \in Z(c) \quad \dots (13)$$

【0051】

《重み係数 $W(i,k)$ の例》

以下に重み係数 $W(i,k)$ の例を示す。重み係数 $W(i,k)$ の例として、例えば、以下の式(14)のように、全センサについての周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ のパワーを合算し、それを全センサ・全周波数についての周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ のパワー総和で正規化した値を例示できる(重み係数 $W(i,k)$ の例1)。

【0052】

【数 7】

$$W(i,k) = \frac{\sum_{s=1}^S |X(s,i,k)|^2}{\sum_{k=0}^M \sum_{s=1}^S |X(s,i,k)|^2} \quad \dots (14)$$

また、重み係数 $W(i,k)$ として、例えば、以下の式(15)のように、全センサについての周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ の振幅の絶対値を合算し、それを全センサ・全周波数についての周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ の振幅の絶対値の総和で正規化した値を用いてもよい(重み係数 $W(i,k)$ の例2)。

10

【0053】

【数 8】

$$W(i,k) = \frac{\sum_{s=1}^S |X(s,i,k)|}{\sum_{k=0}^M \sum_{s=1}^S |X(s,i,k)|} \quad \dots (15)$$

また、式(14)(15)のような正規化を行わないで重み係数 $W(i,k)$ を求めてもよい(重み係数 $W(i,k)$ の例3)。この場合であっても、雑音環境によっては十分に目的信号区間推定が可能な場合もある。例えば、以下の式(11)(12)のように重み係数 $W(i,k)$ を求めてもよい。

20

【0054】

【数 9】

$$W(i,k) = \sum_{s=1}^S |X(s,i,k)|^2 \quad \dots (16)$$

$$W(i,k) = \sum_{s=1}^S |X(s,i,k)| \quad \dots (17)$$

また、全センサについての周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ の振幅の絶対値やパワーを合算するのではなく、一部のセンサについての周波数領域信号 $X(1,i,k), \dots, X(S,i,k)$ の振幅の絶対値やパワーを合算したり、以下の式(18)(19)のように1個のセンサ20-Jの周波数領域信号 $X(J,i,k)$ の振幅の絶対値やパワーを重み係数 $W(i,k)$ としたりしてもよい(重み係数 $W(i,k)$ の例4)。なおこの場合には、できるだけ信号源に近い(出来れば最も近い)センサ20-Jの周波数領域信号 $X(J,i,k)$ を用いることが望ましい。信号源に近いセンサ20-Jほど、遅延や畳み込みの影響が少なく、適切な重み係数 $W(i,k)$ を算出できるからである。

30

【0055】

$$W(i,k) = |X(J,i,k)| \quad \dots (18)$$

$$W(i,k) = |X(J,i,k)|^2 \quad \dots (19)$$

なお、重み係数 $W(i,k)$ を1などの固定値としてもよい。また、雑音環境や目的信号の状況に応じ、重み係数 $W(i,k)$ を1などの固定値とする場合と、重み係数 $W(i,k)$ の例1~4のように重み係数 $W(i,k)$ を逐次算出する場合と、を切り替え制御可能な構成としてもよい(《重み係数 $W(i,k)$ の例》の説明終わり)。

40

【0056】

図11は、以上のように生成したヒストグラムを、横軸を量子化された正規化信号値(信号到来方向)Z(c)とし、縦軸を正規化された重み付け後の頻度 $\text{bin}_g(i,k,c)$ として表示した例である。ここで、図11(a)は、目的信号が存在する時間周波数ビンを含むグリッドについて作成されたヒストグラムであり、図11(b)は、目的信号が存在せず、雑音信号のみが存在する時間周波数ビンを含むグリッドについて作成されたヒストグラムの例示である。なお、これらの例において重み係数 $W(i,k)$ は1である。

【0057】

50

図 1 1 (a) (b) の対比から分かるように、目的信号が存在する時間周波数ビンを含むグリッドのヒストグラム (図 1 1 (a)) は、正規化信号値 $Z(c)$ が特定の値に偏った分布をみせる (偏在性が高い) のに対し、目的信号が存在せず、雑音信号のみが存在する時間周波数ビンを含むグリッドのヒストグラム (図 1 1 (b)) は、幅広く分布する形状となることが分かる。

【 0 0 5 8 】

ヒストグラム生成部 1 6 a は、以上のように生成したグリッド毎のヒストグラムを特定するための $\text{bin}_g(i, k, c)$ ($c=1, \dots, C$) を出力し、 $\text{bin}_g(i, k, c)$ は確率密度関数計算部 1 6 b に入力される。

【 0 0 5 9 】

確率密度関数計算部 1 6 b は、 $\text{bin}_g(i, k, c)$ を用い、以下の式 (2 0) のようにヒストグラムを確率密度関数 $P_g(i, k, c)$ とみなし、量子化された各正規化信号値にそれぞれ対応する値 $c=1, \dots, C$ を確率変数とする確率密度関数 $P_g(i, k, c)$ を算出し、出力する (ステップ S 7 2) 。

【 0 0 6 0 】

【 数 1 0 】

$$P_g(i, k, c) = \frac{\text{bin}_g(i, k, c)}{\sum_{c=1}^C \text{bin}_g(i, k, c)} \quad \dots (20)$$

10

20

出力されたグリッド毎の各確率密度関数 $P_g(i, k, c)$ は、エントロピー計算部 1 6 c に入力され、エントロピー計算部 1 6 c は、以下の式 (2 1) のようにグリッド毎のエントロピー $H_g(i, k)$ を求め、各グリッドの偏在値として出力する。

【 0 0 6 1 】

【 数 1 1 】

$$H_g(i, k) = -\sum_{c=1}^C P_g(i, k, c) \cdot \log_2(P_g(i, k, c)) \quad \dots (21)$$

このように算出したエントロピー $H_g(i, k)$ は、正規化信号値 $Z(c)$ のヒストグラムが特定の値に偏った分布をみせる場合には低い値となり、幅広く分布する場合には高い値となり、ヒストグラムの分布の偏りを示す。すなわち、図 1 1 (a) のように、目的信号が存在する時間周波数ビンを含むグリッドのヒストグラムは、正規化信号値 $Z(c)$ が特定の値に偏るため、エントロピー $H_g(i, k)$ は小さくなる。なお、このエントロピーの大きさを反転させるため、エントロピー計算部 1 6 c がさらに以下の計算を行い、その演算結果を各グリッドの偏在値として出力としてもよい。

【 0 0 6 2 】

【 数 1 2 】

$$\bar{H}_g(i, k) = (1 - H_g(i, k)) / \log_2(C) \quad \dots (21)$$

この演算結果は、エントロピー $H_g(i, k)$ の大きさを反転させたものであり、目的信号が存在するグリッドで大きい値をとり、それ以外のグリッドで小さい値をとり、ヒストグラムの分布の偏りを示す。以下では、エントロピー計算部 1 6 c から出力される各グリッドの偏在値を式 (2 1) の演算結果も含めて $H_g(i, k)$ と表現する。

【 0 0 6 3 】

エントロピー計算部 1 6 c から出力された各グリッドの偏在値 $H_g(i, k)$ は、平均部 1 6 d に入力される。平均部 1 6 d は、同一のフレーム i に対応する各グリッドの偏在値 $H_g(i, k)$ を平均し、その平均値を当該フレーム i の偏在性指標値 $H(i)$ として算出する (ステップ S 7 4) 。すなわち、平均部 1 6 d は、同一のフレーム i に対応する各グリッドの偏在値 $H_g(i, k)$ を $g=1, \dots, G$ について合算し、 G で割ることにより当該フレーム i の偏在性指標値 $H(i)$ を算出する。

30

40

50

【 0 0 6 4 】

なお、ここではヒストグラムの偏りを示す指標としてエントロピーを用い、それを偏在性指標値 $H(i)$ としているが、その他の正規化信号値 $ZGRID_g(i,k)$ の偏在性を示す指標を在性指標値 $H(i)$ としてもよい。以下に他の偏在性指標値 $H(i)$ を例示する。

【 0 0 6 5 】

《 偏在性指標値 $H(i)$ の変形例 》

例えば、図 2 の偏在性指標値算出部 1 6 の代わりに、図 7 の偏在性指標値算出部 1 6 を用いてもよい (偏在性指標値 $H(i)$ の変形例 1)。この例では分散を偏在性指標値 $H(i)$ として用いる。この場合、まず、偏在性指標値算出部 1 6 の平均値算出部 1 6 e に、正規化信号値 $ZGRID_g(i,k)$ が入力される。平均値算出部 1 6 e は、以下の式 (2 2) のように、時間周波数ビン (i,k) 毎に各正規化信号値 $ZGRID_g(i,k)$ を重み係数 $W(i,k)$ で重み付けし、グリッド毎に重み付け後の平均値 $E_g(i,k)$ を求めて出力する。なお、 μ はグリッド毎の正規化信号値 $ZGRID_g(i,k)$ の要素数である。

10

【 0 0 6 6 】

【 数 1 3 】

$$E_g(i,k) = \frac{1}{\mu} \sum W(i,k) \cdot ZGRID_g(i,k) \quad \dots (22)$$

偏在性指標値算出部 1 6 の分散計算部 1 6 f には、平均値 $E_g(i,k)$ と、各正規化信号値 $ZGRID_g(i,k)$ とが入力され、以下の式 (2 3) のように分散 $H_g(i,k)$ を計算し、それを各グリッドの偏在値 $H_g(i,k)$ として出力する。

20

【 0 0 6 7 】

【 数 1 4 】

$$H_g(i,k) = \frac{1}{\mu - 1} \sum (W(i,k) \cdot ZGRID_g(i,k) - E_g(i,k))^2 \quad \dots (23)$$

分散計算部 1 6 f から出力された各グリッドの偏在値 $H_g(i,k)$ は、平均部 1 6 d に入力される。平均部 1 6 d は、同一のフレーム i に対応する各グリッドの偏在値 $H_g(i,k)$ を平均し、その平均値を当該フレーム i の偏在性指標値 $H(i)$ として算出する。

【 0 0 6 8 】

30

また、図 2 の偏在性指標値算出部 1 6 の代わりに、図 8 の偏在性指標値算出部 1 6 を用いてもよい (偏在性指標値 $H(i)$ の変形例 2)。この例では尖度を偏在性指標値 $H(i)$ として用いる。

【 0 0 6 9 】

この場合、まず、偏在性指標値算出部 1 6 の平均値算出部 1 6 e に、正規化信号値 $ZGRID_g(i,k)$ が入力される。平均値算出部 1 6 e は、式 (2 2) のように、時間周波数ビン (i,k) 毎に各正規化信号値 $ZGRID_g(i,k)$ を重み係数 $W(i,k)$ で重み付けし、重み付け後の平均値 $E_g(i,k)$ を求めて出力する。また、偏在性指標値算出部 1 6 の分散計算部 1 6 g には、平均値 $E_g(i,k)$ と、各正規化信号値 $ZGRID_g(i,k)$ とが入力され、式 (2 3) と同様に分散 $H_g(i,k)$ を計算し出力する。

40

【 0 0 7 0 】

さらに尖度計算部 1 6 h に、分散 $H_g(i,k)$ と、平均値 $E_g(i,k)$ と、各正規化信号値 $ZGRID_g(i,k)$ とが入力され、尖度計算部 1 6 h は、例えば以下の式 (2 4) によって尖度 $H_g(i,k)$ を求めて出力する。

【 0 0 7 1 】

【 数 1 5 】

$$H_g(i,k) = \sum \left\{ \left(E_g(i,k) - W(i,k) \cdot ZGRID_g(i,k) \right)^4 / \mu \cdot \sigma_g(i,k)^2 \right\} \quad \dots (24)$$

尖度計算部 1 6 h から出力された各グリッドの偏在値 $H_g(i,k)$ は、平均部 1 6 d に入力

50

される。平均部 16d は、同一のフレーム i に対応する各グリッドの偏在値 $H_g(i, k)$ を平均し、その平均値を当該フレーム i の偏在性指標値 $H(i)$ として算出する。

【0072】

また、標準偏差等その他正規化信号値 $ZGRID_g(i, k)$ の偏在性を示す統計量を各グリッドの偏在値 $H_g(i, k)$ とし、それをフレーム i 毎に平均して偏在性指標値 $H(i)$ としてもよい。

【0073】

さらに、時間周波数ビン (i, k) 毎に 2 種類以上の正規化信号値 $ZGRID_g(i, k)$ (例えば位相差と振幅比) が生成されている場合には、当該 2 種類以上の正規化信号値 $ZGRID_g(i, k)$ の偏在性をそれぞれ示す 2 以上の偏在性指標値 $H(i)$ を算出してもよいし、当該 2 種類以上の正規化信号値 $ZGRID_g(i, k)$ を要素とするベクトルの偏在性を示す偏在性指標値 $H(i)$ を算出してもよいが、2 以上の偏在性指標値 $H(i)$ を算出する場合と 1 種類の偏在性指標値 $H(i)$ を算出する場合とでは、後述する判定部 17 での処理内容が相違する(《偏在性指標値 $H(i)$ の変形例》[ステップ S7 の例]の説明終わり)。

【0074】

上述のように偏在性指標値算出部 16 から出力された正規偏在性指標値 $H(i)$ は、判定部 17 に入力され、判定部 17 は、偏在性指標値 $H(i)$ を指標とし、各フレームが目的信号区間であるか否かを判定する(ステップ S8)。

【0075】

本形態の判定部 17 は、判定対象のフレームの偏在性指標値と、非目的信号区間のフレームの偏在性指標値と、の比である除算値又は当該除算値の写像が、所定の閾値以上であった場合に判定対象のフレームが目的信号区間であると判定するか、当該所定の閾値を超える場合に判定対象のフレームが目的信号区間であると判定する。

【0076】

[ステップ S8 の詳細]

図 3 及び図 6 に示す例では、まず、判定部 17 の相対値算出部 17a (図 3) は、式(21)によって算出されて偏在性指標値算出部 16 から出力された偏在性指標値のうち、判定対象のフレームの偏在性指標値 $H'(i)$ とし、目的信号が存在しないと推定される非目的信号区間のフレームの偏在性指標値 $H''(i)$ とし、それらの比である除算値 $\gamma(i)$ を以下のように算出して出力する(ステップ S81)。なお、目的信号が存在しないと推定される非目的信号区間の一例、例えば、 $i=1, \dots, 20$ などの初期区間である。

【0077】

$$\gamma(i) = H'(i) / H''(i) \quad \dots (25)$$

次に、尤度比算出部 17b に除算値 $\gamma(i)$ が入力され、尤度比算出部 17b は、以下の式(26)に従って尤度比 $\Lambda(i)$ を算出して出力する(ステップ S82)。なお、式(26)の対数は自然対数である。また、この尤度比の計算式は、例えば、Shon, J, Kim, N.-S., and Sung, W., "A Statistical Model-based Voice Activity Detection," IEEE Signal Processing Letters, Vol. 6, No. 1, pp.1-3, 1999. 等に関示されている。

【0078】

【数 16】

$$\log \Lambda(i) = \gamma(i) - \log \gamma(i) - 1 \quad \dots (26)$$

次に、閾値判定部 17c に尤度比 $\Lambda(i)$ が入力され、閾値判定部 17c は尤度比 $\Lambda(i)$ と所定の閾値 th とを比較し、尤度比 $\Lambda(i)$ に対応するフレーム i が目的信号区間であるか否か、すなわち、フレーム i が目的信号区間であるか否かを判定し、その判定結果を出力する(ステップ S83)。具体的には、閾値判定部 17c は、例えば、尤度比 $\Lambda(i)$ が所定の閾値 th より大きい場合(「閾値 th 以上の場合」としてもよい)、目的信号がフレーム i に含まれるとして 1 を出力し(ステップ S84)、尤度比 $\Lambda(i)$ が所定の閾値 th より小さい場合(「閾値 th 以下の場合」としてもよい)、目的信号がフレーム i に含まれないとして 0 を出力する(ステップ S85)。なお、閾値 th は、尤度比 $\Lambda(i)$ の時間長平均(複数のフレーム i に対する平均)や分散などの統計量を用いて設定されてもよいし、 $th=0.2$ など

10

20

30

40

50

の固定値を事前に設定しておいてもよい。分散などの統計量を用いて閾値 th を設定する場合の一例としては、目的信号が存在しないと推定されるフレームを判定対象のフレームとして尤度比 (i) を求め、それらの平均値から所定のマージンを設けた値を閾値 th とする方法がある。

【0079】

なお、偏在性指標値 $H(i)$ を指標として目的信号区間を判定する方法はこれに限定されない。前述のように偏在性指標値 $H(i)$ の大きさは、各フレーム i が目的信号区間であるか否かによって変化する値である。偏在性指標値 $H(i)$ の大きさを評価し、その評価結果を各フレーム i が目的信号区間であるか否かの判定結果に対応付ける方法であれば、どのような方法を用いてもよい。以下に目的信号区間判定方法の変形例を示す。

10

【0080】

[目的信号区間判定方法の変形例]

例えば、図3の判定部17の代わりに、図9の判定部17を用いてもよい(目的信号区間判定方法の変形例1)。この変形例の場合、第1値算出部17aに前述の判定対象のフレームの偏在性指標値 $H'(i)$ と、目的信号が存在しないと推定される非目的信号区間のフレームの偏在性指標値 (i) が入力され、それらの比である除算値 (i) を上述の式(25)のように算出して出力する。次に、閾値判定部17dに除算値 (i) が入力され、閾値判定部17dは、フレーム i 毎に除算値 (i) と閾値 th とを比較し、除算値 (i) が閾値 th よりも大きい場合(「閾値 th 以上の場合」としてもよい)、除算値 (i) に対応するフレーム i が目的信号区間に対応し、そうでなければ除算値 (i) に対応するフレーム i が非目的信号区間であると判定し、その判定結果(1 or 0)を出力する。また、除算値 (i) を用いる代わりに、偏在性指標値 $H'(i)$ から偏在性指標値 (i) を減算した減算値を用い、その減算値に対して上記と同様な閾値処理を行って、目的信号区間であるか否かを推定してもよい。

20

【0081】

また、例えば、図3の判定部17の代わりに、図10(a)の判定部17を用いてもよい(目的信号区間判定方法の変形例2)。この変形例の場合、判定部17の閾値判定部17iに、式(21)によって算出されて偏在性指標値算出部16から出力された偏在性指標値 $H(i)$ が入力され、閾値判定部17iは、フレーム i 毎に除算値 (i) と閾値 th とを比較し、除算値 (i) が閾値 th よりも大きい場合(「閾値 th 以上の場合」としてもよい)、除算値 (i) に対応するフレーム i が目的信号区間に対応し、そうでなければ除算値 (i) に対応するフレーム i が非目的信号区間に対応すると判定し、その判定結果(1 or 0)を出力する。なお、閾値 th は、例えば、閾値算出部17hが入力された偏在性指標値 $H(i)$ のフレーム i 毎の平均値等の統計量をもとに動的に設定される。また、閾値 th は固定値であってもよい。

30

【0082】

なお、上述した以外の偏在性指標値 $H(i)$ を用い、上述のように目的信号区間を判定してもよい。この場合の閾値判定は偏在性指標値 $H(i)$ の特性による。すなわち、偏在性が高いほど値が大きくなる偏在性指標値 $H(i)$ を用いる際には、偏在性指標値 $H(i)$ 又はその写像が所定の閾値を超えた場合(又は「以上の場合」)に目的信号区間であると判定し、偏在性指標値 $H(i)$ 又はその写像が所定の閾値を未満場合(又は「以下の場合」)に目的信号区間でないと判定する。一方、偏在性が低いほど値が大きくなる偏在性指標値 $H(i)$ を用いる際には、偏在性指標値 $H(i)$ 又はその写像が所定の閾値を超えた場合(又は「以上の場合」)に目的信号区間でないと判定し、偏在性指標値 $H(i)$ 又はその写像が所定の閾値を未満場合(又は「以下の場合」)に目的信号区間であると判定する。

40

【0083】

また、時間周波数ピン (i, k) 毎に2種類以上の正規化信号値 $ZGRID_g(i, k)$ が生成され、各グリッドに属する2種類以上の正規化信号値 $ZGRID_g(i, k)$ を要素とするベクトルの偏在性を示す偏在性指標値 $H(i)$ がフレーム i 毎に算出されている場合であっても、判定部17は、上述と同様に目的信号区間であるか否かの判定を行うことができる。

50

【 0 0 8 4 】

一方、時間周波数ビン (i, k) 毎に2種類以上の正規化信号値 $ZGRID_g(i, k)$ が生成され、各グリッドに属する2種類以上の正規化信号値 $ZGRID_g(i, k)$ の偏在性をそれぞれ示す2以上の偏在性指標値 $H(i)$ がフレーム i 毎に算出されている場合、判定部17は、例えば、フレーム i 毎の2以上の偏在性指標値 $H(i)$ に重み付けを行い、当該重み付け後の偏在性指標値を指標とし、各フレーム i が目的信号区間であるか否かを判定する。具体的には、例えば、2以上の偏在性指標値 $H(i)$ の重み付け和が所定の閾値を超えるか否かによって、フレーム i が目的信号区間であるか否かを判定する。

【 0 0 8 5 】

また、上述のように偏在性指標値 $H(i)$ 又はその写像と、所定の閾値との大小を比較して目的信号区間であるか否かを判定する代わりに、事前学習されたフレームの偏在性指標値と当該フレームが目的信号区間であるか否かの判定結果との関係を用いたパターン認識により、偏在性指標値算出部で算出された偏在性指標値に対応するフレームが目的信号区間に対応するか否かを判定してもよい。この場合、例えば、図10(b)の判定部17のように、パラメータ学習部17hに、フレームの音響特徴量(偏在性指標値や (i) など)と当該フレームが目的信号区間であるか否かの判定結果との組からなる学習サンプルを入力し、パラメータ学習部17hでパターン認識学習を行い、モデルパラメータを求める。そして、パターン認識部17iにこのパラメータと判定対象の偏在性指標値 $H(i)$ とを入力し、パターン認識によって偏在性指標値 $H(i)$ に対応するフレーム i が目的信号区間のものであるか否かを判定する。なお、パターン認識技術には、例えば、公知のサポートベクターマシン(津田宏治, “サポートベクターマシンとは何か”, 電子情報通信学会誌, 2000: 460~466頁)や、隠れマルコフモデル(北研二, 中村哲, 永田昌明, “音声言語処理”, 森出版株式会社, 1996: 57~90頁)等を利用する。

【 0 0 8 6 】

判定部17が、目的信号区間であるか否かの判定結果を出力するのではなく、前述の尤度比 (i) そのもの、または、それを確率値に変換した $(i)/(1+(i))$ などを出力してもよい。

【 0 0 8 7 】

< 実験結果 >

本形態の効果を示すための実験結果を示す。この実験では、センサとして2本のマイクロフォンを用い、音声信号と雑音信号が混在する音響信号を観測し、その音響信号を本形態の信号区間推定方法によって分析し、音声信号区間を検出する実施例を示す。なお、この実験では、正規化信号値 $Z(i, k)$ として信号到来方向推定値を用い、偏在性指標値 $H(i)$ として式(21)によって算出されて偏在性指標値算出部16から出力された偏在性指標値を用いて目的信号区間の推定を行った。

【 0 0 8 8 】

使用した音響信号データは大学の研究室内で学生がポスターを用いて自身の研究の発表を行っている発話を収録した信号で、サンプリング周波数16kHz、量子化ビット数16ビットで離散サンプリングされたものを用いた。収録には2本のマイクロフォンを用いており、2本のマイクロフォンは同一直線上に4cmの間隔で配置されている。図12(a)に収録された音響信号を示す。なお、図12(a)における横軸は時間であり、縦軸は音響信号の振幅である。この音響信号には、冒頭の部分にポスター発表が行われている研究室のドアを開閉する音(方向性雑音)が混入している。この音響信号に対し、本形態による信号区間推定方法を、1フレームの時間長を32ms(512サンプル点)とし、16ms(256サンプル点)毎にフレーム始点を移動させて適用する。図12(b)は、これによって各フレームで推定された偏在性指標値 $H(i)$ (音響特徴量)を示すグラフである。なお、図12(a)における横軸は時間であり、縦軸は偏在性指標値 $H(i)$ の振幅である。また、非特許文献6記載の方法で求まる偏在性指標値 $H(i)$ を図12(c)に示す。

【 0 0 8 9 】

図示のように本形態による目的信号区間推定方法により出力される偏在性指標値 $H(i)$ が

、音声信号の存在区間Bでは高い値を示し、それ以外の区間では小さい値を示すことがわかる。また、図12(b)と(c)とを比較すると、図12(a)のデータに含まれるドアの開閉音のみが収録されている区間Aにおいて、本形態による手法では偏在性指標値H(i)の値が小さいままなのに対し(図12(b))、非特許文献6記載の方法では音声信号区間と同程度の高い値が見られる(図12(c))。このことから、本形態によって得られる偏在性指標値H(i)が、調波性を持たない、ドアの開閉音のような方向性雑音に対し頑健であることがわかる。加えて、本形態では基本周波数とその倍音成分の近傍における到達時間差の偏在性のみを利用しているため、全時間周波数帯域で到達時間差の偏在性を計算する非特許文献6記載の方法よりも高速に偏在性指標値H(i)を計算できる。この実験においては、本形態の方法は非特許文献6記載の方法の9%の計算時間で音響特徴の計算を行

10

うことができた。なお、最悪の計算量は、
(サンプリング周波数 / 基本周波数推定の探索範囲の下限周波数) / 離散フーリエ変換の点数

により計算することができる。本形態に例示したサンプリング周波数(16kHz)と離散フーリエ変換の点数(512点)、及び基本周波数推定の探索範囲の下限周波数(50Hz)の場合は、 $(16000/50)/512$ =約31.25%の計算量で済む。実際は、推定される基本周波数の値は50~500Hzに分布するため、上記の実験により示されたように、最悪の計算量よりも少ない時間で計算が可能となる。

【0090】

以上のことから、本形態により、方向性雑音の影響を受けにくく、高速に目的音響信号区間を検出することが可能になることがわかる。なお、非特許文献6記載の方法と比較して、本形態では観測信号の基本周波数を推定する必要があるが、本形態記載の基本周波数推定方法は高速での実行が可能であることから、全体の計算量には影響を与えない。

20

【0091】

なお、本発明は上述の実施の形態に限定されるものではない。例えば、本形態のステップS4では、基本周波数推定部13が全てのセンサに対応する基本周波数を推定し、それらをそれ以降の各センサに対応する処理にそれぞれ利用することとした。しかし、ステップS4において、基本周波数推定部13が一部のセンサ(例えば1つのセンサ)に対応する基本周波数のみを推定し、それをそれ以降の全てのセンサに対応する処理に利用してもよい。

30

【0092】

上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されてもよい。また、信号区間推定装置10がサンプリング部30を包含する構成であってもよいし、信号区間推定装置10の機能を複数のコンピュータで分散処理する構成であってもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

【0093】

また、上述の構成をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。

40

【0094】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

【0095】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

50

【 0 0 9 6 】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆる A S P (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

10

【 0 0 9 7 】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい

【 産業上の利用可能性 】

20

【 0 0 9 8 】

本発明の利用分野としては、例えば、音声信号や音楽信号などの目的信号が雑音信号とともに観測される環境においてなされる、目的信号の符号化、雑音信号の抑圧、残響除去、自動音声認識などの音響信号処理分野を例示できる。もちろん、音響信号以外の信号処理に本発明を適用してもかまわない。

【 図面の簡単な説明 】

【 0 0 9 9 】

【 図 1 】 図 1 は、本形態の目的信号区間推定装置の全体構成を例示したブロック図である。

。

【 図 2 】 図 2 は、図 1 の偏在性指標値算出部 1 6 の詳細構成を例示したブロック図である

30

。

【 図 3 】 図 3 は、図 1 の判定部 1 7 の詳細構成を例示したブロック図である。

【 図 4 】 図 4 は、本形態の目的信号区間推定方法を説明するためのフローチャートである

。

【 図 5 】 図 5 は、ステップ S 7 の詳細を説明するためのフローチャートである。

【 図 6 】 図 6 は、ステップ S 8 の詳細を説明するためのフローチャートである。

【 図 7 】 図 7 は、偏在性指標値算出部の構成例を示すブロック図である。

【 図 8 】 図 8 は、偏在性指標値算出部の構成例を示すブロック図である。

【 図 9 】 図 9 は、判定部の構成例を示すブロック図である。

【 図 1 0 】 図 1 0 (a) (b) は、判定部の構成例を示すブロック図である。

40

【 図 1 1 】 図 1 1 (a) (b) は、本形態の処理によって得られたヒストグラムを示すグラフである。

【 図 1 2 】 図 1 2 (a) は、実験において収録された音響信号を示すグラフである。図 1 2 (b) は、実験において各フレームで推定された偏在性指標値 $H(i)$ (音響特徴量)を示すグラフである。図 1 2 (c) は、非特許文献 6 記載の方法で求まる偏在性指標値 $H(i)$ を示すグラフである。

【 符号の説明 】

【 0 1 0 0 】

1 0 信号区間推定装置

【図1】

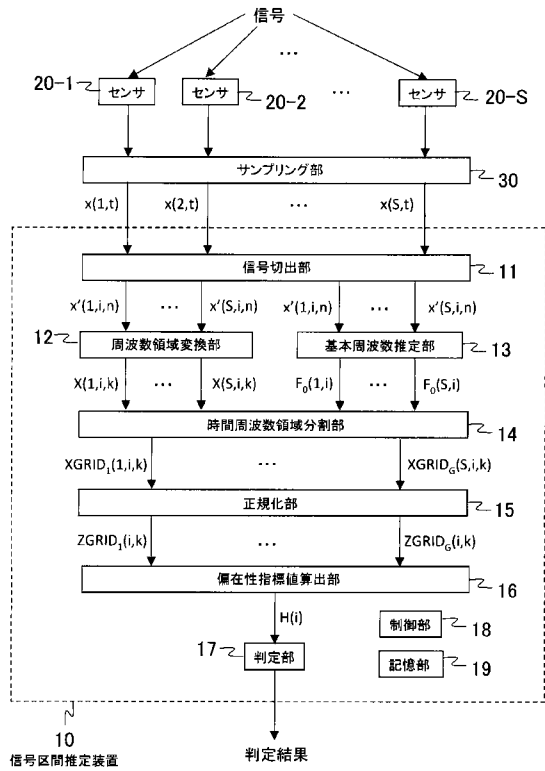


図1

【図2】

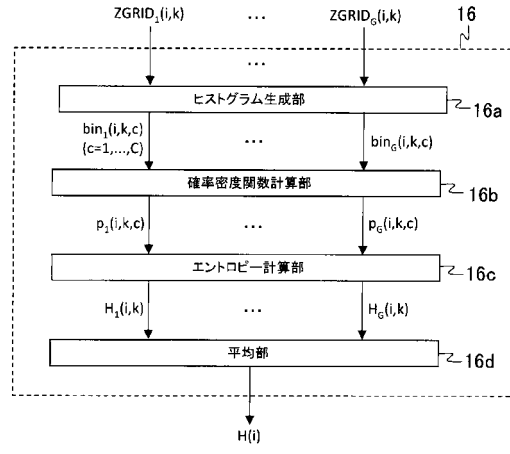


図2

【図3】

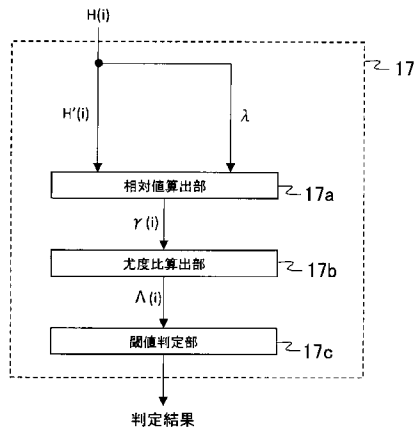


図3

【図4】

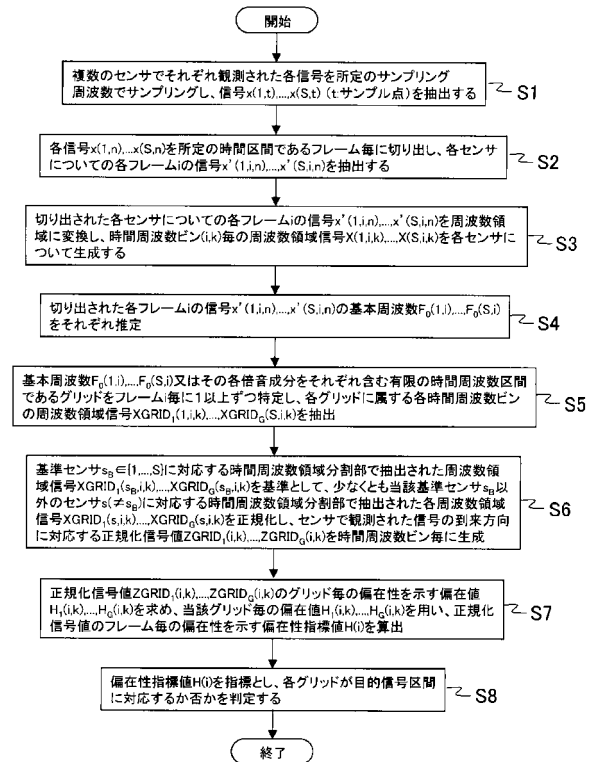


図4

【 図 5 】

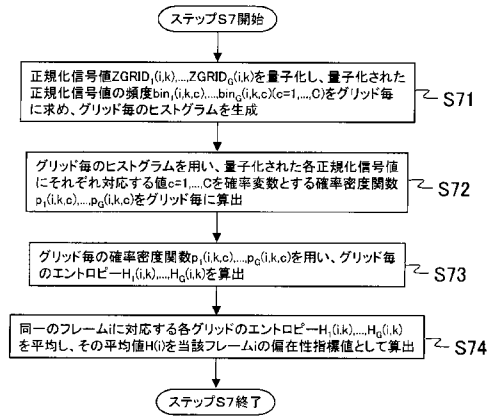


図5

【 図 6 】

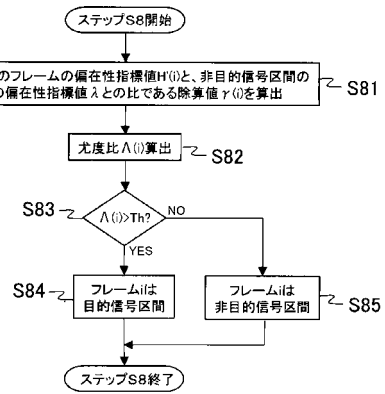


図6

【 図 7 】

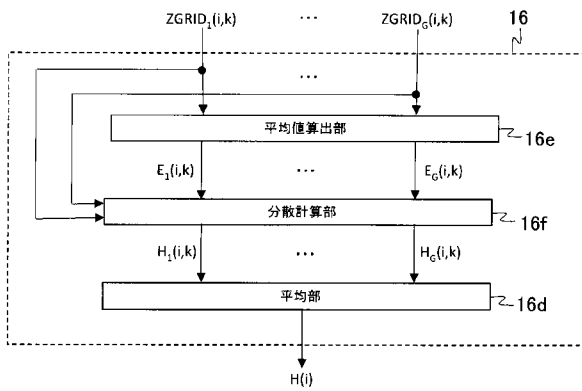


図7

【 図 8 】

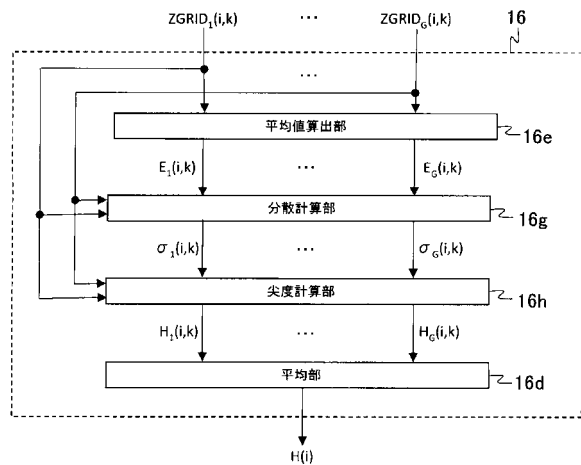


図8

【 図 9 】

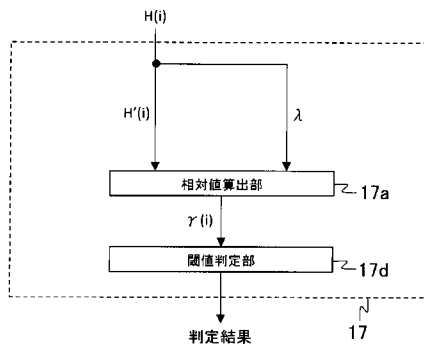


図9

【 図 10 】

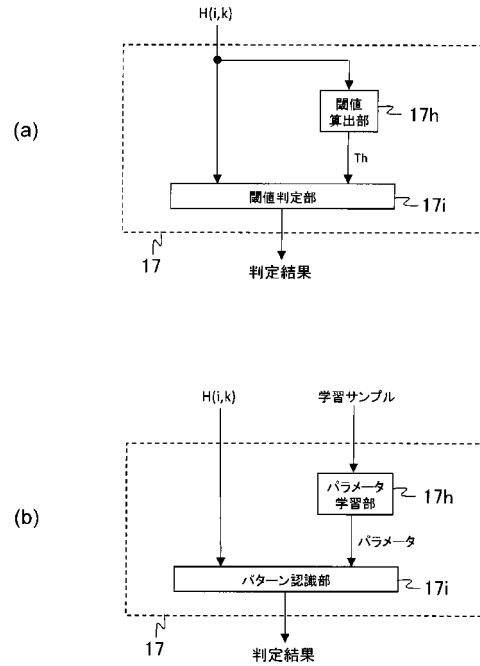


図10

【 図 11 】

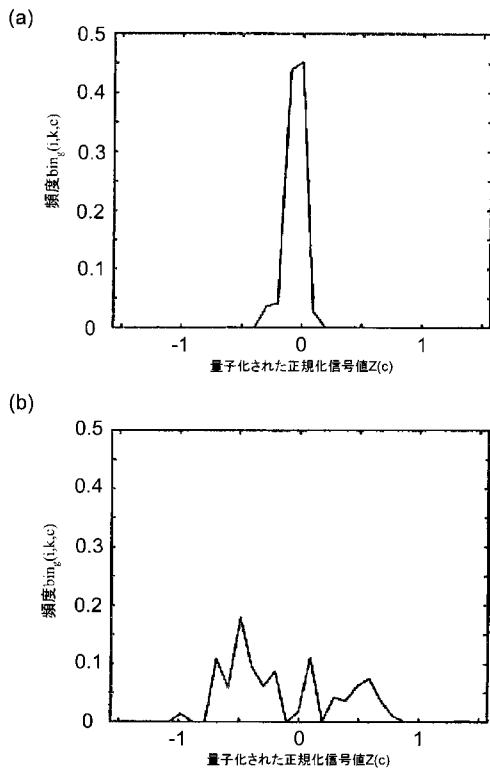


図11

【 図 12 】

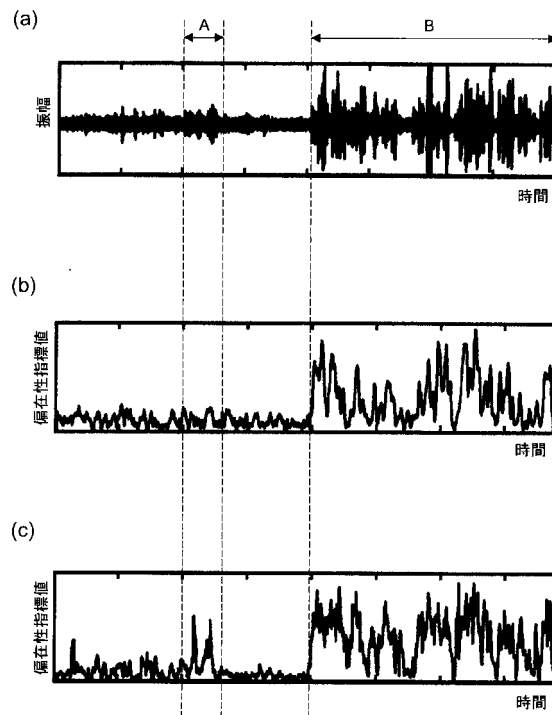


図12

フロントページの続き

- (72)発明者 石塚 健太郎
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 荒木 章子
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 河原 達也
京都府京都市左京区吉田本町 国立大学法人京都大学大学院情報学研究科内

審査官 安田 勇太

- (56)参考文献 石塚 健太郎 Kentaro ISHIZUKA, 信号到来方向の推定値の偏りを用いた耐雑音音声区間検出法, 日本音響学会 2007年 秋季研究発表会講演論文集CD-ROM [CD-ROM], 日本, 日本音響学会, 2007年 9月21日
- Juan E. Rubio, Kentaro Ishizuka, Hiroshi Sawada, Shoko Araki, "Two-Microphone Voice Activity Detection Based on the Homogeneity of the Direction of Arrival Estimates," , Proceedings of the 32nd International Conference on Acoustics, Speech, and Signal Processing, , 2007年, Vol. 4, , pp. 385-388

(58)調査した分野(Int.Cl., DB名)

G10L 11/02

G10L 15/04

G10L 21/02 JSTPlus/JMEDPlus/JST7580(JDreamI
I)