

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-47831

(P2009-47831A)

(43) 公開日 平成21年3月5日(2009.3.5)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/02 (2006.01)	G 1 0 L 15/02 3 0 0 K	5 D 0 1 5
G 1 0 L 11/00 (2006.01)	G 1 0 L 11/00 1 0 1 Z	
G 1 0 L 11/04 (2006.01)	G 1 0 L 11/04	

審査請求 未請求 請求項の数 9 O L (全 13 頁)

(21) 出願番号	特願2007-212739 (P2007-212739)	(71) 出願人	000003078
(22) 出願日	平成19年8月17日 (2007.8.17)		株式会社東芝
			東京都港区芝浦一丁目1番1号
		(74) 代理人	100089118
			弁理士 酒井 宏明
		(72) 発明者	益子 貴史
			東京都港区芝浦一丁目1番1号 株式会社東芝内
		Fターム(参考)	5D015 CC12

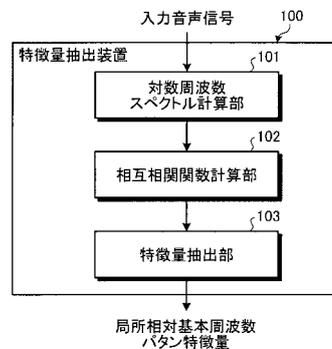
(54) 【発明の名称】 特徴量抽出装置、プログラムおよび特徴量抽出方法

(57) 【要約】

【課題】ピッチ抽出やピッチ周期の範囲指定を必要とせずに基本周波数パターン情報を得ることができるとともに、背景雑音の影響を受けにくくすることができる特徴量抽出装置を提供する。

【解決手段】フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算手段101と、時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算手段102と、前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出手段103と、を備える。

【選択図】 図2



【特許請求の範囲】

【請求項 1】

フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算手段と、

時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算手段と、

前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出手段と、

を備えることを特徴とする特徴量抽出装置。

10

【請求項 2】

前記スペクトル計算手段が計算する前記対数周波数スペクトルは、スペクトル包絡成分を除いた残差成分の対数周波数スペクトルである、

ことを特徴とする請求項 1 記載の特徴量抽出装置。

【請求項 3】

前記スペクトル計算手段は、前記対数周波数スペクトルに対して振幅の正規化を行なう、

ことを特徴とする請求項 1 または 2 記載の特徴量抽出装置。

【請求項 4】

時刻毎に計算された前記相互相関関数の列から時刻毎に当該時刻の相互相関関数と当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の相互相関関数との相互相関関数を再帰的に繰り返し計算する再帰計算手段をさらに備え、

20

前記特徴量抽出手段は、前記再帰計算手段により再帰的に繰り返し計算された前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する、

ことを特徴とする請求項 1 ないし 3 のいずれか一記載の特徴量抽出装置。

【請求項 5】

前記相互相関関数の次元を時刻毎に圧縮する次元圧縮手段をさらに備え、

前記特徴量抽出手段は、前記次元圧縮手段による次元圧縮後の前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する、

ことを特徴とする請求項 1 ないし 4 のいずれか一記載の特徴量抽出装置。

30

【請求項 6】

前記相互相関関数から時刻毎に近似関数を求める近似関数計算手段をさらに備え、

前記特徴量抽出手段は、前記近似関数計算手段で求められた近似関数をフレームにおける局所相対基本周波数パターン特徴量として抽出する、

ことを特徴とする請求項 1 ないし 4 のいずれか一記載の特徴量抽出装置。

【請求項 7】

前記近似関数上の相互相関関数値の列およびそれらの統計量を近似関数の信頼度として求める信頼度計算手段をさらに備え、

前記特徴量抽出手段は、前記信頼度計算手段で求められた信頼度をフレームにおける局所相対基本周波数パターン特徴量として抽出する、

ことを特徴とする請求項 6 記載の特徴量抽出装置。

40

【請求項 8】

フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算機能と、

時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算機能と、

前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出機能と、

をコンピュータに実行させることを特徴とするプログラム。

50

【請求項 9】

フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算工程と、

時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算工程と、

前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出工程と、

を含むことを特徴とする特徴量抽出方法。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、特徴量抽出装置、プログラムおよび特徴量抽出方法に関する。

【背景技術】

【0002】

音声の韻律情報の要素の1つに、アクセントやイントネーション、有声/無声についての情報を獲得する基本周波数パターン情報がある。このような基本周波数パターン情報は、音声認識装置や音声区間検出装置、ピッチ抽出装置あるいは話者認識装置などで使用される。このような基本周波数パターン情報を得ようとする場合、非特許文献1に示されている手法などを用いて、ピッチ抽出を行う必要がある。

20

【0003】

また、特許文献1では、ある時刻(フレーム)tの音声の予測残差の自己相関関数と別の時刻(フレーム)sの音声の予測残差の自己相関関数との相互相関関数をピッチ周波数差分特徴量とすることにより、ピッチ抽出誤りの影響を低減し複数のピッチ周波数の候補を考慮したピッチ周波数差分情報を得る方法が提案されている。

【0004】

【非特許文献1】古井貞熙、「デジタル音声処理」、東海大学出版会、pp. 57 - 59 (1985)

【特許文献1】特許第2940835号公報

【発明の開示】

30

【発明が解決しようとする課題】

【0005】

しかしながら、特許文献1に記載されている方法によれば、音声の予測残差に基づいているため、背景雑音の影響により特徴量が劣化しやすいという問題がある。また、予測残差の自己相関関数にはピッチ周期の整数倍の位置に複数のピークが現れるが、整数倍の位置のピークを用いると差分値も整数倍になるため、正しいピッチ周波数差分情報を求めるためには相互相関関数を求める予測残差自己相関関数の範囲を正しいピッチ周期の付近に限定する必要があり、そのためには事前にピッチ周期を求めたり、話者の声の高さに応じてピッチ周期の範囲を適切に定める必要がある。

【0006】

40

本発明は、上記に鑑みてなされたものであって、ピッチ抽出やピッチ周期の範囲指定を必要とせずに基本周波数パターン情報を得ることができるとともに、背景雑音の影響を受けにくくすることができる特徴量抽出装置、プログラムおよび特徴量抽出方法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】

上述した課題を解決し、目的を達成するために、本発明の特徴量抽出装置は、フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算手段と、時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間

50

幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算手段と、前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出手段と、を備える。

【 0 0 0 8 】

また、本発明のプログラムは、フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算機能と、時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算機能と、前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出機能と、をコンピュータに実行させる。

10

【 0 0 0 9 】

また、本発明の特徴量抽出方法は、フレーム毎に入力音声信号から対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算するスペクトル計算工程と、時刻毎に計算された前記対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する関数計算工程と、前記相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する特徴量抽出工程と、を含む。

【 発明の効果 】

20

【 0 0 1 0 】

本発明によれば、対数周波数スペクトルの相互相関関数に基づいて局所相対基本周波数パターン特徴量を求めることにより、基本周波数の変動による対数周波数スペクトルのピーク（調波成分）のシフト量はどのピーク（調波成分）に対しても同じになり、相互相関関数のラグ 0 付近のピークの変動は基本周波数の変動に対応することになるので、ピッチ抽出やピッチ周期の範囲指定を必要とせずに基本周波数パターン情報を得ることができるとともに、背景雑音の影響を受けにくくすることができる、という効果を奏する。

【 発明を実施するための最良の形態 】

【 0 0 1 1 】

以下に添付図面を参照して、この発明にかかる特徴量抽出装置、プログラムおよび特徴量抽出方法の最良な実施の形態を詳細に説明する。

30

【 0 0 1 2 】

[第 1 の実施の形態]

本発明の第 1 の実施の形態を図 1 ないし図 6 に基づいて説明する。本実施の形態は、音声認識装置に備えられる特徴量抽出装置への適用例である。

【 0 0 1 3 】

図 1 は、本発明の第 1 の実施の形態にかかる音声認識装置 1 のハードウェア構成を示すブロック図である。本実施の形態の音声認識装置 1 は、概略的には、人間の音声コンピュータで自動的に認識する音声認識処理を行なうものである。

【 0 0 1 4 】

40

図 1 に示すように、音声認識装置 1 は、例えばパーソナルコンピュータであり、コンピュータの主要部であって各部を集中的に制御する CPU (Central Processing Unit) 2 を備えている。この CPU 2 には、BIOS などを記憶した読み専用メモリである ROM (Read Only Memory) 3 と、各種データを書換え可能に記憶する RAM (Random Access Memory) 4 とがバス 5 で接続されている。

【 0 0 1 5 】

さらにバス 5 には、各種のプログラム等を格納する HDD (Hard Disk Drive) 6 と、配布されたプログラムであるコンピュータソフトウェアを読み取るための機構として CD (Compact Disc) - ROM 7 を読み取る CD - ROM ドライブ 8 と、音声認識装置 1 とネットワーク 9 との通信を司る通信制御装置 10 と、各種操作指示を行うキーボードやマウ

50

スなどの入力装置 11 と、各種情報を表示する CRT (Cathode Ray Tube)、LCD (Liquid Crystal Display) などの表示装置 12 とが、図示しない I/O を介して接続されている。

【0016】

RAM 4 は、各種データを書換え可能に記憶する性質を有していることから、CPU 2 の作業エリアとして機能してバッファ等の役割を果たす。

【0017】

図 1 に示す CD-ROM 7 は、この発明の記憶媒体を実施するものであり、OS (Operating System) や各種のプログラムが記憶されている。CPU 2 は、CD-ROM 7 に記憶されているプログラムを CD-ROM ドライブ 8 で読み取り、HDD 6 にインストールする。

10

【0018】

なお、記憶媒体としては、CD-ROM 7 のみならず、DVD などの各種の光ディスク、各種光磁気ディスク、フレキシブルディスクなどの各種磁気ディスク等、半導体メモリ等の各種方式のメディアを用いることができる。また、通信制御装置 10 を介してインターネットなどのネットワーク 9 からプログラムをダウンロードし、HDD 6 にインストールするようにしてもよい。この場合に、送信側のサーバでプログラムを記憶している記憶装置も、この発明の記憶媒体である。なお、プログラムは、所定の OS (Operating System) 上で動作するものであってもよいし、その場合に後述の各種処理の一部の実行を OS に肩代わりさせるものであってもよいし、所定のアプリケーションソフトや OS などを構成

20

【0019】

このシステム全体の動作を制御する CPU 2 は、このシステムの主記憶として使用される HDD 6 上にロードされたプログラムに基づいて各種処理を実行する。

【0020】

次に、音声認識装置 1 の HDD 6 にインストールされている各種のプログラムが CPU 2 に実行させる機能のうち、本実施の形態の音声認識装置 1 が備える特長的な機能について説明する。

【0021】

図 2 は、音声認識装置 1 が備える特徴量抽出装置 100 の機能構成を示すブロック図である。図 2 に示すように、音声認識装置 1 は、プログラムに従うことにより、局所相対基本周波数パターン特徴量を抽出する特徴量抽出装置 100 を備える。この局所相対基本周波数パターン特徴量は、音声認識処理に用いる音声の韻律情報の要素の 1 つであって、アクセントやイントネーション、有声/無声についての情報を獲得可能な基本周波数パターン情報である。

30

【0022】

本実施の形態の特徴量抽出装置 100 は、図 2 に示すように、対数周波数スペクトル計算部 101 と、相互相関関数計算部 102 と、特徴量抽出部 103 とを備えている。対数周波数スペクトル計算部 101 は、スペクトル計算手段として機能するものであって、入力音声信号から所定の間隔の時刻 (フレーム) 毎に対数周波数軸上で等間隔に求められた周波数成分からなる対数周波数スペクトルを計算する。相互相関関数計算部 102 は、関数計算手段として機能するものであって、対数周波数スペクトル計算部 101 で時刻毎に計算された対数周波数スペクトルの列から時刻毎に当該時刻の対数周波数スペクトルと当該時刻の前後の一定の時間幅に含まれる一つまたは複数の時刻の対数周波数スペクトルとの相互相関関数を計算する。特徴量抽出部 103 は、特徴量抽出手段として機能するものであって、相互相関関数計算部 102 で計算された相互相関関数の組をフレームにおける局所相対基本周波数パターン特徴量として抽出する。以下において、対数周波数スペクトル計算部 101、相互相関関数計算部 102 および特徴量抽出部 103 について、詳述する。

40

【0023】

50

まず、対数周波数スペクトル計算部 101 について説明する。対数周波数スペクトル計算部 101 は、まず、入力音声信号から 1 フレーム（例えば 10 ms）毎に対数周波数軸上で等間隔となる周波数点上で求められた周波数成分からなる対数周波数スペクトル $S_t(w)$ を求める。ここで、 t はフレーム番号、 w ($0 < w < W$) は周波数点番号を表している。この対数周波数スペクトル $S_t(w)$ は、具体的には対数周波数軸上で等間隔となる周波数点に基づくフーリエ変換やウェーブレット変換、また直線周波数軸上で等間隔となる周波数点に基づくフーリエ変換により求められた直線周波数スペクトルからの周波数軸変換などにより求められる。

【0024】

なお、対数周波数スペクトルは、振幅の正規化を行った対数周波数スペクトルでも良い。振幅の正規化は、具体的には対数周波数スペクトルの振幅の平均を一定値（例えば 0）にする方法、分散を一定値（例えば 1）にする方法、最小値と最大値を一定値（例えば 0 と 1）にする方法、あるいは対数周波数スペクトルを求める音声波形の振幅の分散を一定値（例えば 1）にする方法などにより行われる。

10

【0025】

また、対数周波数スペクトルは、スペクトル包絡成分を除いた残差成分の対数周波数スペクトルでも良い。この残差成分の対数周波数スペクトルは、線形予測分析などにより得られる残差信号から求めても良いし、ケプストラムの高次成分のフーリエ変換から求めても良い。さらに、この残差成分の対数周波数スペクトルに対して振幅の正規化を行っても良い。

20

【0026】

なお、対数周波数スペクトルを求める際、周波数成分を求める範囲を、例えば音声のエネルギーが相対的に大きい 200 Hz から 1600 Hz までとすることにより、背景雑音による影響を受けにくい対数周波数スペクトルが得られる。

【0027】

次に、相互相関関数計算部 102 について説明する。相互相関関数計算部 102 は、各フレーム t において、当該フレームの対数周波数スペクトル $S_t(w)$ とその前後の一定の時間幅（近傍 N ）に含まれるフレーム $t + n$ の対数周波数スペクトル $S_{t+n}(w)$ との相互相関関数 $C_t(\tau, n)$ を計算する。 n は対数周波数軸上でのずれの大きさ（ラグ）を表し、その値は $-(W-1)$ から $(W-1)$ までに含まれる一定の整数値の集合 L により与えられる。相互相関関数 $C_t(\tau, n)$ は、下記に示す数式 (1) により計算される。

30

【数 1】

$$C_t(\tau, n) = \frac{1}{W - |n|} \sum_i S_t(i) S_{t+\tau}(i+n) \quad \dots(1)$$

ただし、 $S_t(w) = 0$ ($w < 0, w \geq W$) とする。

【0028】

なお、数式 (1) の右辺の項 $1 / (W - |n|)$ は、ラグの絶対値の増加に伴う相互相関関数の計算に用いられる周波数成分の数の減少を補正する項であり、必ずしも必要ではない。また、 $C_t(\tau, n) = -C_t(\tau, -n)$ の関係を用いることにより、数式 (1) の計算量を削減することが可能である。

40

【0029】

特徴量抽出部 103 は、上述のようにして求められた相互相関関数の組 $C_t(\tau, n)$ ($\tau \in N, n \in L$) をフレーム t における局所相対基本周波数パターン特徴量として抽出する。

【0030】

ここで、対数周波数スペクトルおよび相互相関関数の例を図 3 から図 6 に示す。

50

【 0 0 3 1 】

図 3 は、クリーン音声の有声音区間に含まれる 5 フレームの対数周波数スペクトルを示すグラフである。図 3 の横軸は周波数点番号、縦軸はフレーム番号である。図 3 中の対数周波数スペクトルは 2 0 0 H z から 1 6 0 0 H z までの周波数帯域から対数周波数軸上で等間隔となる 2 5 6 点の周波数成分からなっており、振幅を平均が 0、分散が 1 となるように正規化している。

【 0 0 3 2 】

図 4 は、対数周波数スペクトルの相互相関関数を示すグラフである。図 4 は、図 3 のフレーム 7 7 を基準フレームとして求めた対数周波数スペクトルを表している。図 4 の横軸はラグ、縦軸の数字は基準フレームと相互相関関数を求めたフレームとのフレーム番号の差分を表している。例えば、差分 - 2 はフレーム 7 7 とフレーム 7 5 との相互相関関数である。ただし、差分 0 は自己相関関数に等しい。各フレームの枠の縦軸は - 1 から 1 までの相互相関関数の値を示しており、枠中央の水平の点線は 0 を表している。

10

【 0 0 3 3 】

すなわち、図 4 の相互相関関数の組が、近傍 $N = \{ - 2 , - 1 , 0 , 1 , 2 \}$ としたときのフレーム 7 7 における局所相対基本周波数パターン特徴量となる。

【 0 0 3 4 】

図 3 に示す対数周波数スペクトルには 4 つから 5 つのピークが現れており、それぞれ基本周波数の整数倍の位置にある調波成分に対応している。この対数周波数スペクトルのピークは、フレーム番号が大きくなるにしたがって右にシフトしているが、これは基本周波数の上昇に対応している。また、図 4 においては、ラグ 0 付近のピークもフレーム番号が大きくなるにしたがって右にシフトしているが、これは対数周波数スペクトルのピークのシフトに対応している。つまり、相互相関関数のラグ 0 付近のピークの変動は、基本周波数の変動に対応している。

20

【 0 0 3 5 】

ここで、図 3 のグラフによれば、基本周波数の変動による対数周波数スペクトルのピーク（調波成分）のシフト量はどのピークに対しても同様であることが分かる。すなわち、どのピーク（調波成分）に対しても同じシフト量となる。

【 0 0 3 6 】

このように本実施の形態によれば、対数周波数スペクトルの相互相関関数に基づいて局所相対基本周波数パターン特徴量を求めることにより、基本周波数の変動による対数周波数スペクトルのピーク（調波成分）のシフト量はどのピーク（調波成分）に対しても同じになり、相互相関関数のラグ 0 付近のピークの変動は基本周波数の変動に対応することになるので、ピッチ抽出やピッチ周期の範囲指定を必要とせず基本周波数パターン情報を得ることができる。すなわち、特定の調波成分を選択して用いる必要はなく、あらかじめ基本周波数を求めたり話者の基本周波数の範囲を指定したりすることなく局所相対基本周波数パターン特徴量を求めることが可能となる。

30

【 0 0 3 7 】

また、図 5 は図 3 で用いている音声に 1 0 d B の白色雑音を加えた音声から求められた対数周波数スペクトル、図 6 は図 5 の対数周波数スペクトルから求めた相互相関関数を表している。図 5 を図 3 と比較すると、特に周波数の低い帯域においてよく似た対数周波数スペクトルが得られていることがわかる。これは 2 0 0 H z から 1 6 0 0 H z 付近が比較的音声のエネルギーが大きい帯域であるためである。また、図 6 においては、ラグ 0 付近のピークは図 4 と同様に变化しており、図 4 とよく似た局所相対基本周波数パターン特徴量が得られていることが分かる。

40

【 0 0 3 8 】

このように本実施の形態によれば、背景雑音の影響を受けにくくすることができるので、雑音の影響をあまり受けずに安定した局所相対基本周波数パターン特徴量を求めることが可能となる。

【 0 0 3 9 】

50

[第 2 の実施の形態]

次に、本発明の第 2 の実施の形態を図 7 に基づいて説明する。なお、前述した第 1 の実施の形態と同じ部分は同じ符号で示し説明も省略する。

【 0 0 4 0 】

図 7 は、本発明の第 2 の実施形態にかかる特徴量抽出装置 1 0 0 の機能構成を示すブロック図である。図 7 に示すように、本実施の形態の特徴量抽出装置 1 0 0 は、相互相関関数計算部 1 0 2 で時刻毎に計算された相互相関関数から時刻毎に相互相関関数を再帰的に計算する相互相関関数再起計算部 1 0 4 を備えている点で、第 1 の実施の形態とは異なるものである。

【 0 0 4 1 】

相互相関関数再起計算部 1 0 4 は、再帰計算手段として機能するものであって、 $C_t^{(i)}(\tau, n) = C_t^{(i-1)}(\tau, n)$ として、各フレーム t において、当該フレームの相互相関関数の組 $C_t^{(i-1)}(\tau, n) (N, n-L)$ とその前後の一定の時間幅 (近傍 N) に含まれるフレーム $t + \tau$ の相互相関関数の組 $C_{t+\tau}^{(i-1)}(\tau, n) (N, n-L)$ との相互相関関数 $C_t^{(i)}(\tau, n)$ を、下記に示す数式 (2) のように、再帰的に計算する。

【 数 2 】

$$C_t^{(i)}(\tau, n) = \sum_u \sum_j C_t^{(i-1)}(u, j) C_{t+\tau}^{(i-1)}(u - \tau, j + n) \quad (i \geq 2) \quad \dots(2)$$

【 0 0 4 2 】

なお、数式 (1) と同様に、数式 (2) の右辺に計算に用いられる相互相関関数値の数による変動を補正する項 ($1 / (W - |n|)$) を加えても良い。また、対数周波数スペクトルと同様に、相互相関関数 $C_t^{(i-1)}(\tau, n)$ の振幅に対する正規化を行っても良い。

【 0 0 4 3 】

特徴量抽出部 1 0 3 は、このようにして求められた相互相関関数の組 $C_t^{(i)}(\tau, n) (N, n-L)$ をフレーム t における局所相対基本周波数パタン特徴量として抽出する。

【 0 0 4 4 】

このように本実施の形態によれば、当該フレーム以外のフレーム同士の相互相関も考慮することにより、当該フレームと他のフレームとの相互相関のみを考慮する場合よりも安定した局所相対基本周波数パタン特徴量を求めることが可能となる。

【 0 0 4 5 】

[第 3 の実施の形態]

次に、本発明の第 3 の実施の形態を図 8 ないし図 1 0 に基づいて説明する。なお、前述した第 1 の実施の形態と同じ部分は同じ符号で示し説明も省略する。

【 0 0 4 6 】

図 8 は、本発明の第 3 の実施形態にかかる特徴量抽出装置 1 0 0 の機能構成を示すブロック図である。図 8 に示すように、本実施の形態の特徴量抽出装置 1 0 0 は、相互相関関数計算部 1 0 2 で時刻毎に計算された相互相関関数を時刻毎に次元圧縮する次元圧縮部 1 0 5 を備えている点で、第 1 の実施の形態とは異なるものである。

【 0 0 4 7 】

次元圧縮部 1 0 5 は、次元圧縮手段として機能するものであって、各フレーム t において、相互相関関数計算部 1 0 2 で計算された相互相関関数 $C_t(\tau, n) (n-L)$ の次元数を離散コサイン変換や主成分分析などを用いて圧縮する。

【 0 0 4 8 】

ここで、図 9 は図 4 に示される相互相関関数からラグの範囲が - 3 0 から 3 0 までの部分を取り出したものである。このとき、相互相関関数 $C_t(\tau, n) (-30 \leq n \leq 30)$

10

20

30

40

50

の次元数は 61 となっている。

【0049】

一方、図10は図9に示される相互相関関数をそれぞれ5次元の離散コサイン変換係数で近似したものである。図10から、次元圧縮を行っても元の相互相関関数とほぼ同等のパターンが得られていることが分かる。

【0050】

特徴量抽出部103は、このようにして得られた次元圧縮後の相互相関関数の組を局所相対基本周波数パターン特徴量として抽出する。

【0051】

このように本実施の形態によれば、少ない次元数で効率よく表現された局所相対基本周波数パターン特徴量を求めることが可能となる。

10

【0052】

なお、本実施の形態の特徴量抽出装置100においては、相互相関関数計算部102で時刻毎に計算された相互相関関数を、次元圧縮部105によって時刻毎に次元圧縮するようにしたが、これに限るものではない。例えば、第2の実施の形態で説明したように、相互相関関数計算部102で時刻毎に計算された相互相関関数から相互相関関数再起計算部104で時刻毎に相互相関関数を再帰的に計算した後、次元圧縮部105によって時刻毎に次元圧縮するようにしても良い。

【0053】

[第4の実施の形態]

20

次に、本発明の第4の実施の形態を図11および図12に基づいて説明する。なお、前述した第1の実施の形態と同じ部分は同じ符号で示し説明も省略する。

【0054】

図11は、本発明の第3の実施形態にかかる特徴量抽出装置100の機能構成を示すブロック図である。図11に示すように、本実施の形態の特徴量抽出装置100は、相互相関関数計算部102で時刻毎に計算された相互相関関数から時刻毎に基本周波数パターン近似関数を求める近似関数計算部106と、相互相関関数計算部102で時刻毎に計算された相互相関関数と近似関数計算部106で時刻毎に計算された基本周波数パターン近似関数から時刻毎に前記基本周波数パターン近似関数の信頼度を計算する信頼度計算部107を備えている点で、第1の実施の形態とは異なるものである。

30

【0055】

近似関数計算部106は、近似関数計算手段として機能するものであって、各フレーム t において、相互相関関数計算部102で計算された相互相関関数の組 $C_t(\tau, n)$ ($N, n = 1, \dots, L$) から局所相対基本周波数パターン近似関数 $F_t(\tau)$ を求める。この近似関数 $F_t(\tau)$ は、例えば最小二乗誤差基準を用いる場合には、下記に示す数式(3)に示される誤差 E_t を最小化することにより求められる。

【数3】

$$E_t = \sum_{\tau \in N(t)} \sum_{n \in L} C_t(\tau, n) \{F_t(\tau) - n\}^2 \quad \dots(3)$$

40

【0056】

信頼度計算部107は、信頼度計算手段として機能するものであって、各フレーム t において、相互相関関数計算部102で計算された相互相関関数の組 $C_t(\tau, n)$ ($N, n = 1, \dots, L$) と、近似関数計算部106で計算された局所相対基本周波数パターン近似関数 $F_t(\tau)$ から、近似関数 $F_t(\tau)$ の信頼度を求める。この信頼度は、近似関数 $F_t(\tau)$ 上の相互相関関数の値の組 $C_t(\tau, F_t(\tau))$ (N) や、これらの平均、分散、最大値などの統計量により与えられる。

【0057】

特徴量抽出部103は、このようにして求められた局所相対基本周波数パターン近似関数

50

$F_t(\quad)$ およびその信頼度を、フレーム t における局所相対基本周波数パターン特徴量として抽出する。

【0058】

ここで、図12は無声区間における相互相関関数の例を示すグラフである。図12に示されるように、無声区間では基本周波数が存在しないため、ラグ0の自己相関関数を除いて、相互相関関数には明確なピークが存在しない。しかし、数式(3)によれば、このような場合でも近似関数を得ることができる。

【0059】

また、図12に示されるように、基本周波数が存在しない場合には、相互相関関数の値が全体的に小さいため、局所相対基本周波数パターン近似関数上の相互相関関数の値も小さくなる。逆に、図4に示されるように基本周波数が存在し相互相関関数に明確なピークが存在する場合には、局所相対基本周波数パターン近似関数上の相互相関関数の値は大きくなる。つまり、局所相対基本周波数パターン近似関数上の相互相関関数の値は、基本周波数の存在の確からしさを表している。

【0060】

このように本実施の形態によれば、局所相対基本周波数パターン近似関数を求めることにより、本来基本周波数が存在しない無声区間においても局所相対基本周波数パターン特徴量を得ることが可能となる。さらに、局所相対基本周波数パターン近似関数の信頼度も求めることにより、基本周波数の存在の確からしさを含む局所相対基本周波数パターン特徴量を得ることが可能となる。

【0061】

なお、本実施の形態の特徴量抽出装置100においては、相互相関関数計算部102で時刻毎に計算された相互相関関数から、近似関数計算部106によって時刻毎に基本周波数パターン近似関数を求め、相互相関関数計算部102で時刻毎に計算された相互相関関数と近似関数計算部106で時刻毎に計算された基本周波数パターン近似関数から時刻毎に前記基本周波数パターン近似関数の信頼度を計算するようにしたが、これに限るものではない。例えば、第2の実施の形態で説明したように、相互相関関数計算部102で時刻毎に計算された相互相関関数から相互相関関数再起計算部104で時刻毎に相互相関関数を再帰的に計算した後、近似関数計算部106によって時刻毎に基本周波数パターン近似関数を求めるようにしても良い。

【0062】

なお、本発明は上述した各実施の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、上記各実施の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、実施の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、異なる実施の形態にわたる構成要素を適宜組み合わせてもよい。

【0063】

また、各実施の形態においては、音声認識装置に備えられる特徴量抽出装置への適用例を示したが、これに限るものではなく、基本周波数パターン情報を必要とする音声区間検出装置、ピッチ抽出装置あるいは話者認識装置などに備えられる特徴量抽出装置へ適用しても良い。

【図面の簡単な説明】

【0064】

【図1】本発明の第1の実施の形態にかかる音声認識装置のハードウェア構成を示すブロック図である。

【図2】特徴量抽出装置の機能構成を示すブロック図である。

【図3】クリーン音声の有声音区間に含まれる5フレームの対数周波数スペクトルを示すグラフである。

【図4】対数周波数スペクトルの相互相関関数を示すグラフである。

【図5】雑音を加えた音声から求められた対数周波数スペクトルを示すグラフである。

10

20

30

40

50

【図6】図5の対数周波数スペクトルの相互相関関数を示すグラフである。

【図7】本発明の第2の実施形態にかかる特徴量抽出装置の機能構成を示すブロック図である。

【図8】本発明の第3の実施形態にかかる特徴量抽出装置の機能構成を示すブロック図である。

【図9】対数周波数スペクトルの相互相関関数を部分的に示すグラフである。

【図10】図9の相互相関関数を近似した結果を示すグラフである。

【図11】本発明の第3の実施形態にかかる特徴量抽出装置の機能構成を示すブロック図である。

【図12】無声区間における相互相関関数の例を示すグラフである。

10

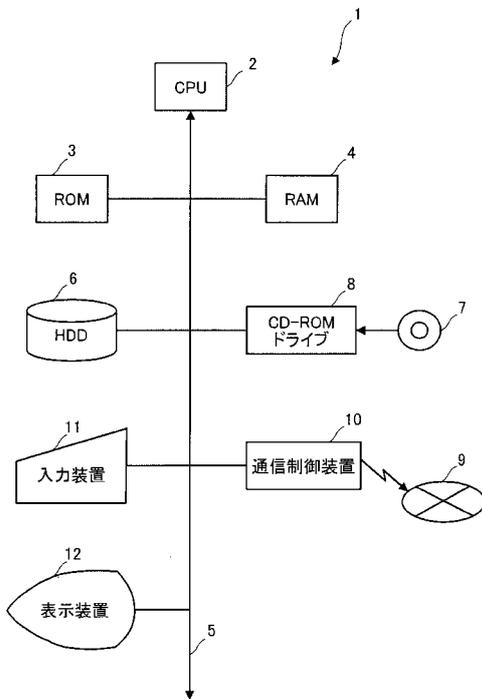
【符号の説明】

【0065】

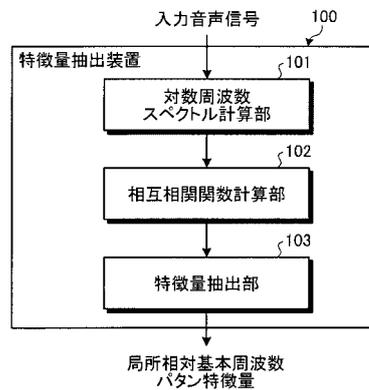
- 100 特徴量抽出装置
- 101 スペクトル計算手段
- 102 関数計算手段
- 103 特徴量抽出手段
- 104 再帰計算手段
- 105 次元圧縮手段
- 106 近似関数計算手段
- 107 信頼度計算手段

20

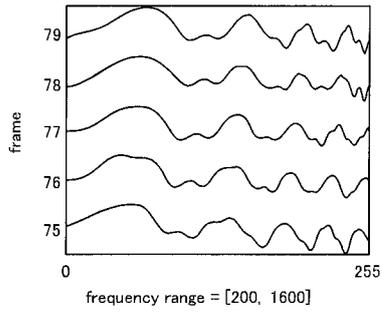
【図1】



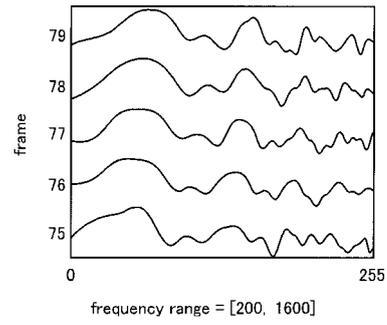
【図2】



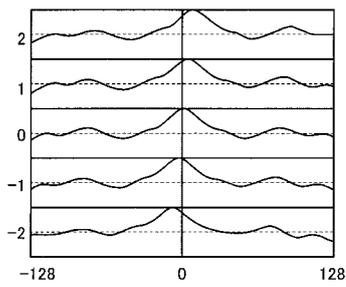
【 図 3 】



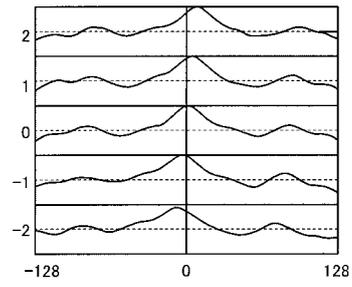
【 図 5 】



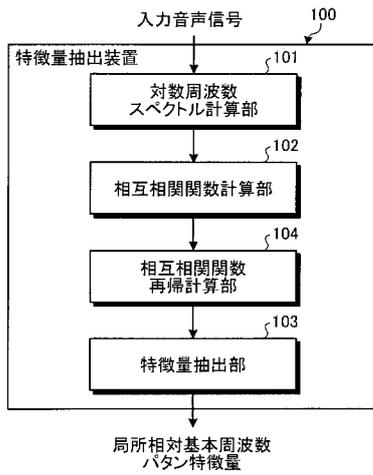
【 図 4 】



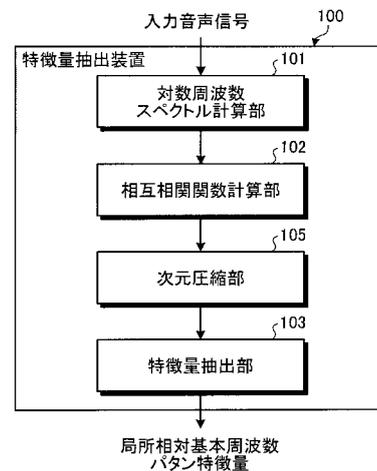
【 図 6 】



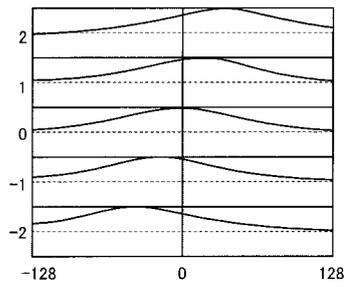
【 図 7 】



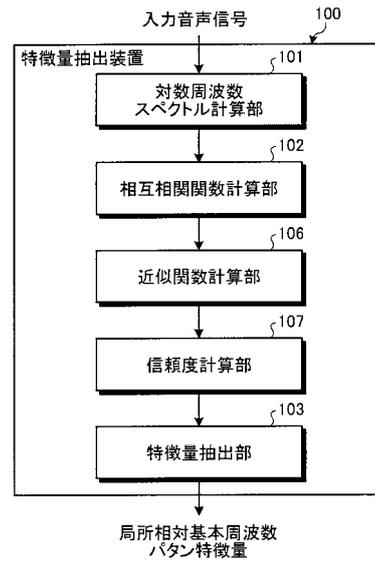
【 図 8 】



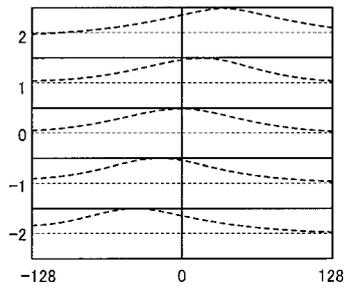
【 図 9 】



【 図 1 1 】



【 図 1 0 】



【 図 1 2 】

