

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4851447号  
(P4851447)

(45) 発行日 平成24年1月11日(2012.1.11)

(24) 登録日 平成23年10月28日(2011.10.28)

(51) Int.Cl. F I  
**G 1 O L 11/04 (2006.01)** G 1 O L 11/04  
**G 1 O L 11/00 (2006.01)** G 1 O L 11/00 4 O 2 H  
 G 1 O L 11/00 1 O 1 D  
 G 1 O L 11/00 1 O 1 A

請求項の数 8 (全 16 頁)

(21) 出願番号	特願2007-520082 (P2007-520082)	(73) 特許権者	500430877 株式会社エイ・ジー・アイ 東京都品川区北品川5丁目13-15-1 01
(86) (22) 出願日	平成18年6月2日(2006.6.2)	(73) 特許権者	504119583 光吉 俊二 千葉県市川市真間5-14-20
(86) 国際出願番号	PCT/JP2006/311123	(74) 代理人	100072718 弁理士 古谷 史旺
(87) 国際公開番号	W02006/132159	(74) 代理人	100116001 弁理士 森 俊秀
(87) 国際公開日	平成18年12月14日(2006.12.14)	(72) 発明者	光吉 俊二 東京都品川区北品川5丁目13-15-1 01
審査請求日	平成20年9月9日(2008.9.9)		
(31) 優先権主張番号	特願2005-169414 (P2005-169414)		
(32) 優先日	平成17年6月9日(2005.6.9)		
(33) 優先権主張国	日本国(JP)		
(31) 優先権主張番号	特願2005-181581 (P2005-181581)		
(32) 優先日	平成17年6月22日(2005.6.22)		
(33) 優先権主張国	日本国(JP)		

最終頁に続く

(54) 【発明の名称】 ピッチ周波数を検出する音声解析装置、音声解析方法、および音声解析プログラム

(57) 【特許請求の範囲】

【請求項 1】

被験者の音声信号を取り込む音声取得部と、  
 前記音声信号を周波数スペクトルに変換する周波数変換部と、  
 前記周波数スペクトルを周波数軸上でずらしながら自己相関波形を求める自己相関部と

、  
前記自己相関波形における複数の極値の出現順番と前記複数の極値の出現位置を示す  
らし周波数である出現周波数との分布を回帰分析し、回帰直線の傾きに基づいてピッチ  
周波数を求めるピッチ検出部と、

を備えたことを特徴とする音声解析装置。

【請求項 2】

請求項 1 に記載の音声解析装置において、  
 前記自己相関部は、前記周波数スペクトルを前記周波数軸上で離散的にずらしながら、  
 前記自己相関波形の離散データを求め、  
 前記ピッチ検出部は、前記自己相関波形の前記離散データを補間して、前記極値の前記  
出現周波数を求める

ことを特徴とする音声解析装置。

【請求項 3】

請求項 1 または請求項 2 に記載の音声解析装置において、  
 前記ピッチ検出部は、(前記出現順番, 前記出現周波数) の母集団から、前記自己相関

波形のレベル変動の小さな標本を除き、残った母集団について前記回帰分析を行い、前記回帰直線の傾きに基づいて前記ピッチ周波数を求める

ことを特徴とする音声解析装置。

【請求項 4】

請求項 1 ないし請求項 3 のいずれか 1 項に記載の音声解析装置において、

前記ピッチ検出部は、

前記自己相関波形を曲線近似することにより、前記自己相関波形に含まれる『フォルマントに依存する成分』を抽出する抽出部と、

前記自己相関波形から前記成分を除去することにより、フォルマントの影響を軽減した自己相関波形を求める減算部とを備え、

前記フォルマントの影響を軽減した前記自己相関波形に基づいて、前記ピッチ周波数を求める

ことを特徴とする音声解析装置。

【請求項 5】

請求項 1 ないし請求項 4 のいずれか 1 項に記載の音声解析装置において、

少なくとも『ピッチ周波数』と『感情状態』との対応関係を記憶する対応記憶部と、

前記ピッチ検出部で検出された前記ピッチ周波数を前記対応関係に照会して、前記被験者の感情状態を推定する感情推定部とを備えた

ことを特徴とする感情検出用の音声解析装置。

【請求項 6】

請求項 1 に記載の音声解析装置において、

前記ピッチ検出部は、『前記回帰直線に対する（前記出現順番、前記出現周波数）の分散度合い』および『前記回帰直線と原点とのずれ』の少なくとも一方を、前記ピッチ周波数の不規則性として求め、

少なくとも『ピッチ周波数』および『ピッチ周波数の不規則性』と、『感情状態』との対応関係を記憶する対応記憶部と、

前記ピッチ検出部で求めた『ピッチ周波数』および『ピッチ周波数の不規則性』を前記対応関係に照会して、前記被験者の感情状態を推定する感情推定部とを備えた

ことを特徴とする感情検出用の音声解析装置。

【請求項 7】

被験者の音声信号を取り込むステップと、

前記音声信号を周波数スペクトルに変換するステップと、

前記周波数スペクトルを周波数軸上でずらしながら自己相関波形を求めるステップと、

前記自己相関波形における複数の極値の出現順番と前記複数の極値の出現位置を示すずらし周波数数量である出現周波数との分布を回帰分析し、回帰直線の傾きに基づいてピッチ周波数を求めるステップと、

を備えたことを特徴とする音声解析方法。

【請求項 8】

コンピュータを、請求項 1 ないし請求項 6 のいずれか 1 項に記載の音声解析装置として機能させるための音声解析プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声のピッチ周波数を検出する音声解析の技術に関する。

また、本発明は、音声のピッチ周波数から感情を推定する感情検出の技術に関する。

【背景技術】

【0002】

従来、被験者の音声信号を分析して、被験者の感情を推定する技術が開示されている。

例えば、特許文献 1 には、歌唱音声の基本周波数を求め、歌い終わりにおける基本周波数の上下変化から、歌唱者の感情を推定する技術が提案されている。

10

20

30

40

50

【特許文献1】特開平10-187178公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

ところで、楽器音では、基本周波数が明瞭に現れるため、基本周波数を検出することが容易である。

しかしながら、一般的な音声では、しわがれ声や震えた声などを含むため、基本周波数が揺らぐ。また、倍音の構成成分が不規則になる。そのため、この種の音声から、基本周波数を確実に検出する有効な方法が確立していない。

そこで、本発明の目的は、音声の周波数を正確かつ確実に検出する技術を提供することである。

また、本発明の別の目的は、音声処理に基づく新しい感情推定の技術を提供することである。

【課題を解決するための手段】

【0004】

《1》 本発明の音声解析装置は、音声取得部、周波数変換部、自己相関部、およびピッチ検出部を備える。

音声取得部は、被験者の音声信号を取り込む。

周波数変換部は、音声信号を周波数スペクトルに変換する。

自己相関部は、周波数スペクトルを周波数軸上でずらしながら自己相関波形を求める。

ピッチ検出部は、自己相関波形における複数の極値の出現順番と複数の極値の出現位置を示すずらし周波数である出現周波数との分布を回帰分析し、回帰直線の傾きに基づいてピッチ周波数を求める。

《2》 なお好ましくは、自己相関部は、周波数スペクトルを周波数軸上で離散的にずらしながら、自己相関波形の離散データを求める。ピッチ検出部は、この自己相関波形の離散データを補間し、その補間ラインから極値の出現周波数を求める。

《3》 なお好ましくは、ピッチ検出部は、複数求めた（出現順番，出現周波数）の母集団から、自己相関波形のレベル変動の小さな標本を除く。ピッチ検出部は、このようにして残った母集団について回帰分析を行い、得られた回帰直線の傾きに基づいてピッチ周波数を求める。

《4》 また好ましくは、ピッチ検出部は、抽出部および減算部を備える。

抽出部は、自己相関波形を曲線近似することによって、自己相関波形に含まれる『フォルマントに依存する成分』を抽出する。

減算部は、自己相関波形からこの成分を除去することにより、フォルマントの影響を軽減した自己相関波形を求める。

この構成により、ピッチ検出部は、フォルマントの影響を軽減した自己相関波形に基づいて、ピッチ周波数を求めることが可能になる。

《5》 なお好ましくは、上述した音声解析装置に、対応記憶部、感情推定部を備える。

対応記憶部は、少なくとも『ピッチ周波数』と『感情状態』との対応関係を記憶する。

感情推定部は、ピッチ検出部で検出されたピッチ周波数を対応関係に照会して、被験者の感情状態を推定する。

《6》 なお好ましくは、上記《1》の音声解析装置において、ピッチ検出部は、『回帰直線に対する（出現順番，出現周波数）の分散度合い』および『回帰直線と原点とのずれ』の少なくとも一方を、ピッチ周波数の不規則性として求める。この音声解析装置に、対応記憶部、感情推定部を備える。

対応記憶部は、少なくとも『ピッチ周波数』および『ピッチ周波数の不規則性』と、『感情状態』との対応関係を記憶する。

感情推定部は、ピッチ検出部で求めた『ピッチ周波数』および『ピッチ周波数の不規則性』を対応関係に照会して、被験者の感情状態を推定する。

《7》 本発明の音声解析方法は、次のステップを有する。

10

20

30

40

50

(ステップ1) 被験者の音声信号を取り込むステップ

(ステップ2) 音声信号を周波数スペクトルに変換するステップ

(ステップ3) 周波数スペクトルを周波数軸上ですらしながら自己相関波形を求めるステップ

(ステップ4) 自己相関波形における複数の極値の出現順番と複数の極値の出現位置を示すずらし周波数である出現周波数との分布を回帰分析し、回帰直線の傾きに基づいてピッチ周波数を求めるステップ

《8》本発明の音声解析プログラムは、コンピュータを、上記《1》～《6》のいずれか1項に記載の音声解析装置として機能させるためのプログラムである。

【発明の効果】

【0005】

[1] 本発明では、音声信号を周波数スペクトルに一旦変換する。この周波数スペクトルには、基本周波数の揺らぎや倍音成分の不規則性がノイズ分として含まれる。そのため、この周波数スペクトルから基本周波数を読み取ることは困難である。

そこで、本発明は、この周波数スペクトルを周波数軸上ですらしながら自己相関波形を求める。この自己相関波形では、周期性の低いスペクトルノイズが抑制される。その結果、自己相関波形には、周期性の強い倍音成分が山となって周期的に現れる。

本発明では、この低ノイズ化された自己相関波形における複数の極値の出現順番と複数の極値の出現位置を示すずらし周波数である出現周波数との分布を回帰分析し、回帰直線の傾きに基づいてピッチ周波数を正確に求める。

このように得られたピッチ周波数は、基本周波数に類似する場合もあるが、自己相関波形の最大ピークや1番目のピークから求めるわけではないため、必ずしも基本周波数とは一致しない。むしろ、山と山(または谷と谷)の間隔から求めることにより、基本周波数の不明瞭な音声からも安定かつ正確にピッチ周波数を求めることが可能となる。

また、音声によっては、自己相関波形に周期的に現れるローカルな山と山(または谷と谷)の間隔が不等間隔になる場合もある。このとき、どこか1箇所の間隔だけを参照してピッチ周波数を決定しては、正確なピッチ周波数を求めることができない。そこで、自己相関波形の山または谷である極値の(出現順番, 出現周波数)を複数求めることが好ましい。これら(出現順番, 出現周波数)を回帰直線で近似することによって、不等間隔の変動を均したピッチ周波数を求めることが可能になる。

このようなピッチ周波数の求め方により、極めて微弱な発話音声からでもピッチ周波数を正確に求めることが可能になる。その結果、ピッチ周波数の分析が困難な音声についても、感情推定の成功率を高めることが可能になる。

[2] また、本発明においては、周波数スペクトルを周波数軸上で離散的にずらしながら、自己相関波形の離散データを求めることが好ましい。このような離散的な処理により、演算回数を軽減し、処理時間の短縮を図ることができる。しかし、離散的にずらす周波数を大きくすると、自己相関波形の分解能が低くなり、ピッチ周波数の検出精度が低下する。そこで、自己相関波形の離散データを補間して、極値の出現周波数を精密に求めることにより、離散データの分解能よりも細かい精度でピッチ周波数を求めることが可能になる。

[3] なお、自己相関波形のレベル変動が小さい箇所は、なだらかな山(または谷)となるため、山や谷の出現周波数を正確に求めることが困難となる。そこで、上記のように求めた(出現順番, 出現周波数)の母集団から、自己相関波形のレベル変動の小さな標本を除くことが好ましい。このようにして限定した母集団について回帰分析を行うことにより、ピッチ周波数を一段と安定かつ正確に求めることが可能になる。

[4] 音声の周波数成分には、時間的に移動する特定のピークが現れる。このピークをフォルマントと言う。自己相関波形にも、波形の山谷とは別に、このフォルマントを反映した成分が現れる。そこで、自己相関波形の揺らぎにフィッティングする程度の曲線で近似する。この曲線は、自己相関波形に含まれる『フォルマントに依存する成分』であると推定できる。この成分を、自己相関波形から除くことによって、フォルマントの影響を軽

10

20

30

40

50

減した自己相関波形を求めることができる。このような処理を施した自己相関波形は、フォルマントによる乱れが少なくなる。そのため、ピッチ周波数をより正確かつ確実に求めることが可能になる。

〔5〕 このように得られるピッチ周波数は、声の高さや声質などの特徴を表すパラメータであり、発話時の感情によっても敏感に変化する。そのため、このピッチ周波数を感情推定の材料とすることにより、基本周波数の検出困難な音声においても確実に感情推定を行うことが可能になる。

〔6〕 さらに、周期的な山と山（または谷と谷）の間隔の不規則性を新たな音声特徴として検出することが好ましい。例えば、回帰直線に対する（出現順番，出現周波数）の分散度合いを統計的に求める。また例えば、回帰直線と原点とのずれを求める。

このように求めた不規則性は、音声の集音環境の善し悪しを示すと共に、声の微妙な変化を表すものである。そこで、このピッチ周波数の不規則性を感情推定の材料に加えることにより、推定可能な感情の種類を増やしたり、微妙な感情の推定成功率を高めることが可能になる。

なお、本発明における上述した目的およびそれ以外の目的は、以下の説明と添付図面とにおいて具体的に示される。

【図面の簡単な説明】

【0006】

【図1】感情検出装置（音声解析装置を含む）11のブロック図である。

【図2】感情検出装置11の動作を説明する流れ図である。

【図3】音声信号の処理過程を説明する図である。

【図4】自己相関波形の補間処理を説明する図である。

【図5】回帰直線とピッチ周波数との関係を説明する図である。

【発明を実施するための最良の形態】

【0007】

〔実施形態の構成〕

図1は、感情検出装置（音声解析装置を含む）11のブロック図である。

図1において、感情検出装置11は、下記の構成を備える。

【0008】

- (1) マイク12・・・被験者の音声を音声信号に変換する。
- (2) 音声取得部13・・・音声信号を取り込む。
- (3) 周波数変換部14・・・取り込まれた音声信号を周波数変換し、音声信号の周波数スペクトルを求める。
- (4) 自己相関部15・・・周波数スペクトルについて周波数軸上で自己相関を求め、周波数軸上に周期的に現れる周波数成分を自己相関波形として求める。
- (5) ピッチ検出部16・・・自己相関波形の山と山（または谷と谷）の周波数間隔を、ピッチ周波数として求める。
- (6) 対応記憶部17・・・ピッチ周波数や分散などの判断材料と、被験者の感情状態との対応関係を記憶する。この対応関係は、ピッチ周波数や分散などの実験データと、被験者の申告する感情状態（怒り、喜び、緊張、または悲しみなど）とを対応付けることにより作成できる。この対応関係の記述方式としては、対応テーブルや判断ロジックやニューラルネットなどが好ましい。
- (7) 感情推定部18・・・ピッチ検出部16で求めたピッチ周波数を、対応記憶部17の対応関係に照会して、対応する感情状態を決定する。決定された感情状態は、推定感情として出力される。

【0009】

なお、上述した構成13～18については、その一部または全部をハードウェア的に構成してもよい。また、コンピュータにおいて感情検出プログラム（音声解析プログラムを含む）を実行することにより、構成13～18の一部または全部をソフトウェア的に実現してもよい。

10

20

30

40

50

## 【 0 0 1 0 】

[ 感情検出装置 1 1 の動作説明 ]

図 2 は、感情検出装置 1 1 の動作を説明する流れ図である。

以下、図 2 に示すステップ番号に沿って、具体的な動作を説明する。

## 【 0 0 1 1 】

ステップ S 1 : 周波数変換部 1 4 は、音声取得部 1 3 から F F T (Fast Fourier Transform) 演算に必要な区間の音声信号を切り出す (図 3 [ A ] 参照)。このとき、切り出し区間の両端の影響を軽減するよう、切り出し区間に対してコサイン窓などの窓関数を施す。

## 【 0 0 1 2 】

ステップ S 2 : 周波数変換部 1 4 は、窓関数で加工した音声信号に対して F F T 演算を施し、周波数スペクトルを求める (図 3 [ B ] 参照)。

なお、周波数スペクトルについては、一般的な対数演算によるレベル抑圧処理を施すと、負値が発生するため、後述する自己相関演算が複雑かつ困難になる。そこで、周波数スペクトルについては、対数演算のレベル抑圧処理ではなく、ルート演算などの正の値が得られるレベル抑圧処理を施しておくことが好ましい。

また、周波数スペクトルのレベル変化を強調する場合には、周波数スペクトルの値を 4 乗演算するなどの強調処理を施してもよい。

## 【 0 0 1 3 】

ステップ S 3 : 周波数スペクトルには、楽器音と言えば倍音に相当するスペクトルが周期的に現れる。しかし、発話音声の周波数スペクトルは、図 3 [ B ] に示すように複雑な成分を含むため、このままでは周期的なスペクトルを明確に区別することが難しい。そこで、自己相関部 1 5 は、この周波数スペクトルを周波数軸方向に所定幅ずつずらしながら自己相関値を順次求める。この演算により得られる自己相関値の離散データを、ずらし周波数ごとにプロットすることによって自己相関波形が得られる (図 3 [ C ] 参照)。

## 【 0 0 1 4 】

なお、周波数スペクトルには、音声帯域以外の不要な成分 (直流成分や極端に低域の成分) が含まれる。これらの不要な成分は、自己相関の演算を狂わせる。そこで、自己相関の演算に先立って、周波数変換部 1 4 は、周波数スペクトルからこれらの不要な成分を抑制または除去しておくことが好ましい。

例えば、周波数スペクトルから、直流成分 (例えば 6 0 ヘルツ以下など) をカットしておくことが好ましい。

また例えば、所定の下限レベル (例えば周波数スペクトルの平均レベル) を設定して周波数スペクトルの足切り (下限リミット) を行い、微小な周波数成分をノイズとしてカットしておくことが好ましい。

このような処理により、自己相関演算において生じる波形乱れを未然に防ぐことができる。

## 【 0 0 1 5 】

ステップ S 4 : 自己相関波形は、図 4 に示すように離散データである。そこで、ピッチ検出部 1 6 は、離散データを補間することにより、複数の山および / または谷について出現周波数を求める。例えば、ここでの補間方法としては、山や谷の付近の離散データについて、直線補間や曲線関数で補間する方法が簡便で好ましい。なお、離散データの間隔が十分に狭い場合は、離散データの補間処理を省略することも可能である。このようにして、(出現順番, 出現周波数) の標本データを複数求める。

## 【 0 0 1 6 】

なお、自己相関波形のレベル変動が小さい箇所は、なだらかな山 (または谷) となるため、この山や谷の出現周波数を正確に求めることが難しい。そのため、不正確な出現周波数をそのまま標本として含めると、後から検出するピッチ周波数の精度が下がる。そこで、上記のように求めた (出現順番, 出現周波数) の母集団から、自己相関波形のレベル変動の小さな標本データを判定する。このように判定された標本データを母集団から取り除

10

20

30

40

50

くことにより、ピッチ周波数の分析に適した母集団を得る。

【 0 0 1 7 】

ステップ S 5 : ピッチ検出部 1 6 は、ステップ S 4 で求めた母集団から標本データをそれぞれ取り出して、出現周波数を出現順番ごとに並べる。このとき、自己相関波形のレベル変動が小さいために取り除かれた出現順番については欠番となる。

ピッチ検出部 1 6 は、このように標本データを並べた座標空間において回帰分析を実施し、回帰直線の傾きを求める。この傾きに基づいて、出現周波数の揺らぎを排除したピッチ周波数を求めることができる。

【 0 0 1 8 】

なお、回帰分析を実施する際に、ピッチ検出部 1 6 は、回帰直線に対する出現周波数の分散を統計的に求め、ピッチ周波数の分散とする。

また、回帰直線と原点とのずれ（例えば、回帰直線の切片）を求め、このずれが、予め定められた許容限界よりも大きい場合、ピッチ周波数の検出に適さない音声区間（騒音など）であると判定してもよい。この場合、その音声区間を除いて、残りの音声区間についてピッチ周波数を検出することが好ましい。

【 0 0 1 9 】

ステップ S 6 : 感情推定部 1 8 は、ステップ S 5 で求めた（ピッチ周波数，分散）のデータを、対応記憶部 1 7 の対応関係に照会して、対応する感情状態（怒り、喜び、緊張、または悲しみなど）を決定する。

【 0 0 2 0 】

[ 本実施形態の効果など ]

まず、図 5 [ A ] [ B ] を用いて、本実施形態と、従来技術との違いについて説明する。

本実施形態のピッチ周波数は、自己相関波形の山と山（または谷と谷）の間隔に相当し、図 5 [ A ] [ B ] では、回帰直線の傾きに対応する。一方、従来の基本周波数は、図 5 [ A ] [ B ] に示す一番目の山の出現周波数に相当する。

【 0 0 2 1 】

図 5 [ A ] では、回帰直線が原点近傍を通過し、その分散が小さい。この場合、自己相関波形には、山がほぼ等間隔に規則正しく現れる。したがって、従来技術でも、基本周波数を明瞭に検出できるケースである。

【 0 0 2 2 】

一方、図 5 [ B ] は、回帰直線が原点から大きく外れ、分散が大きい。この場合、自己相関波形の山は不等間隔に現れる。したがって、基本周波数が不明瞭な音声であり、基本周波数を特定することが困難となる。従来技術では、一番目の山の出現周波数から求めるため、このようなケースにおいては、間違った基本周波数を求めてしまう。

【 0 0 2 3 】

本発明では、このようなケースでは、山の出現周波数から求めた回帰直線が原点近傍を通るか否か、ピッチ周波数の分散が小さいか否かなどによって、ピッチ周波数の信頼性を判断することができる。したがって、本実施形態では、図 5 [ B ] の音声信号については、ピッチ周波数の信頼性が低いと判断して感情推定の材料から除くことが可能になる。そのことにより、信頼性の高いピッチ周波数のみを使用することが可能になり、感情推定の成功率を一段と高めることが可能になる。

【 0 0 2 4 】

なお、図 5 [ B ] のようなケースにおいては、傾きの程度を広義のピッチ周波数として求めることが可能である。この広義のピッチ周波数を感情推定の材料とすることも好ましい。さらに、『分散度合い』および/または『回帰直線と原点とのずれ』をピッチ周波数の不規則性として求めることも可能である。このように求めた不規則性を、感情推定の材料とすることも好ましい。もちろん、このように求めた広義のピッチ周波数およびその不規則性を、感情推定の材料とすることも好ましい。これらの処理では、狭義のピッチ周波数に限らず、音声周波数の特徴や変化を総合的に反映した感情推定が可能になる。

10

20

30

40

50

## 【0025】

また、本実施形態では、自己相関波形の離散データを補間して、ローカルな山と山（または谷と谷）の間隔を求める。したがって、一段と高い分解能でピッチ周波数を求めることが可能になる。その結果、ピッチ周波数の変化をより細かく検出することが可能になり、より精細な感情推定が可能になる。

## 【0026】

さらに、本実施形態では、ピッチ周波数の分散度合い（分散や標準偏差など）も、感情推定の判断材料に加える。このピッチ周波数の分散度合いは、音声信号の不安定さや不協和音の度合いなどの独特な情報を示すものであり、発話者の自信の無さや緊張度合いなどの感情を検出するのに適している。また、この緊張度合いなどからうそ特有の感情を検出するうそ発見器を実現することなどが可能になる。

10

## 【0027】

## [実施形態の補足事項]

なお、上述した実施形態では、自己相関波形からそのまま山や谷の出現周波数を求めている。しかしながら、本発明はこれに限定されるものではない。

## 【0028】

例えば、音声信号の周波数成分には、時間的に移動する特定のピーク（フォルマント）が現れる。自己相関波形にも、ピッチ周波数とは別に、このフォルマントを反映した成分が現れる。そこで、自己相関波形を、山谷の細かな変動にフィッティングしない程度の曲線関数で近似することで、自己相関波形に含まれる『フォルマントに依存する成分』を推定することが好ましい。このように推定した成分（近似曲線）を、自己相関波形から減算することによって、フォルマントの影響を軽減した自己相関波形を求めることができる。このような処理を施すことにより、自己相関波形からフォルマントによる乱れ波形を除くことが可能になり、ピッチ周波数をより正確かつ確実に求めることが可能になる。

20

## 【0029】

また例えば、特殊な音声信号では、自己相関波形の山と山の間に小さな山が出現する。この小さな山を、自己相関波形の山と誤認識すると、ハーフピッチの周波数を求めてしまうことになる。この場合、自己相関波形の山の高さを比較して、小さな山については波形の谷と見なすことが好ましい。この処理により、正確なピッチ周波数を求めることが可能になる。

30

## 【0030】

また例えば、自己相関波形に対して回帰分析を行って回帰直線を求め、その回帰直線より上側の自己相関波形のピーク点を、自己相関波形の山として検出してもよい。

## 【0031】

上述した実施形態では、（ピッチ周波数，分散）を判断材料として感情推定を実施する。しかしながら、実施形態はこれに限定されるものではない。例えば、少なくともピッチ周波数を判断材料として感情推定を実施してもよい。また例えば、このような判断材料を時系列に収集した時系列データを判断材料として感情推定を実施してもよい。また例えば、過去に推定した感情を判断材料に加えることで、感情の変化傾向を加味した感情推定を実現してもよい。また例えば、音声認識した意味情報を判断材料に加えることにより、会話内容を加味した感情推定を実現してもよい。

40

## 【0032】

また、上述した実施形態では、回帰分析によりピッチ周波数を求めている。しかしながら、実施形態はこれに限定されるものではない。例えば、自己相関波形の山（または谷）の間隔を求めて、ピッチ周波数としてもよい。また例えば、山（または谷）の間隔ごとにピッチ周波数を求め、これら複数のピッチ周波数を母集団として統計処理を実施し、ピッチ周波数およびその分散度合いを決定してもよい。

## 【0033】

なお、上述した実施形態では、話し声についてピッチ周波数を求め、そのピッチ周波数の時間変化（抑揚的な変化量）に基づいて、感情推定用の対応関係を作成することが好ま

50



しい。

【 0 0 3 4 】

本発明者は、この話し声から実験的に作成された対応関係を使用して、歌声や楽器演奏などの楽曲（音声信号の一種）についても感情推定を試みた。

【 0 0 3 5 】

具体的には、音符よりも短い時間間隔でピッチ周波数の時間変化をサンプリングすることにより、単純な音程変化とは異なる抑揚的な情報を得ることが可能になる。（なお、一つのピッチ周波数を求めるための音声区間は、音符よりも短くても長くしてもよい）

また別の手法として、節単位などの複数の音符を含む長い音声区間でサンプリングしてピッチ周波数を求めることで、複数の音符を反映した抑揚的な情報を得ることが可能になる。

10

この楽曲による感情推定では、楽曲を聴いたときに人間が感じる感情（或いは楽曲作者が楽曲に込めたであろう感情）とほぼ同じ傾向の感情出力が得られることが分かった。

例えば、長調／短調といった調子の違いに応じて、喜び／悲しみという感情を検出することが可能になる。また、浮き浮きするようなテンポの良いサビ部分では、強い喜びを検出することが可能になる。また、激しいドラム音からは、怒りを検出することが可能になる。

【 0 0 3 6 】

なお、ここでは話し声から作成した対応関係をそのまま兼用しているが、楽曲専用の感情検出装置であれば、楽曲に特化した対応関係を実験的に作成することももちろん可能である。

20

このように、本実施形態の感情検出装置を用いることで、楽曲に表れる感情を推定することも可能になる。これを応用することによって、人間の音楽鑑賞状態をシミュレーションする装置や、楽曲の示す喜怒哀楽に応じて反応するロボットなどを作成することができる。

【 0 0 3 7 】

また、上述した実施形態では、ピッチ周波数を基準にして、対応する感情状態を推定する。しかしながら、本発明はこれに限定されるものではない。例えば、下記のパラメータの少なくとも1つを加味して、感情状態を推定してもよい。

- (1) 時間単位における周波数スペクトラムの変化量
- (2) ピッチ周波数の揺らぎ周期、立上がり時間、維持時間、または立下がり時間
- (3) 低域側の山（谷）から求めたピッチ周波数と平均ピッチ周波数との差
- (4) 高域側の山（谷）から求めたピッチ周波数と平均ピッチ周波数との差
- (5) 低域側の山（谷）から求めたピッチ周波数と、高域側の山（谷）から求めたピッチ周波数との差異、または増減傾向
- (6) 山（谷）の間隔の最大値、または最小値
- (7) 山（谷）の連続する数
- (8) 発話スピード
- (9) 音声信号のパワー値、またはその時間変動
- (10) 音声信号における人間の可聴域を外れた周波数域の状態

30

40

ピッチ周波数と上記のパラメータの実験データと、被験者の申告する感情状態（怒り、喜び、緊張、または悲しみなど）とを対応付けることによって、感情推定用の対応関係を予め作成することができる。対応記憶部17は、この対応関係を記憶する。一方、感情推定部18は、音声信号から求めたピッチ周波数と上記パラメータとを、対応記憶部17の対応関係に照会することにより、感情状態を推定する。

【 0 0 3 8 】

[ ピッチ周波数の応用例 ]

(1) 音声や音響からの感情要素のピッチ周波数の抽出(本実施形態)により、周波数特性やピッチが求められる。さらに、フォルマント情報やパワー情報についても、時間軸での変化から容易に求めることができる。さらに、これら情報を可視化することも可能になる

50

。また、ピッチ周波数の抽出により、時間変化による音声や音響、音楽などの揺らぎの状態が明確になるため、スムーズな音声や音楽の感情感性リズム解析や音色分析も可能になる。

【0039】

(2) 本実施形態でのピッチ解析で得られた情報の時間変化における変化パターン情報などを感性会話以外にも、映像、アクション(表情や動作)、音楽、映像、構文などに応用することも可能である。

【0040】

(3) また、映像、アクション(表情や動作)、音楽、映像、構文などのリズムを有する情報(リズム情報という)を音声信号と見なしてピッチ解析することも可能である。さらに、リズム情報について時間軸での変化パターン分析も可能である。これらの解析結果に基づいてリズム情報を可視化したり、音声化することにより、別の表現形態の情報に変換することも可能になる。

10

【0041】

(4) また、感情や感性、リズム情報、音色分析手段などで得られた、変化パターンなどを感情感性心理特性解析などに応用することもできる。その結果を用いて、共有もしくは連動する感性の変化パターンやパラメータ、閾値などを求めることも可能になる。

【0042】

(5) 二次利用として、感情要素のばらつき度合いや多感情の同時検出状態などから、真意といった心理情報を推測して、心理や精神の状態を推測することも可能になる。その結果、顧客やユーザーや相手の心理状態による、金融やコールセンタなどでの商品顧客分析管理システム、真偽分析などへの応用が可能になる。

20

【0043】

(6) また、ピッチ周波数による感情要素の判断では、人間が持つ心理特性(感情、指向性、嗜好性、思考(心理意思))を分析して、シミュレーション構築する要素を得ることが可能になる。この人間の心理特性を、既存のシステム、商品、サービス、ビジネスモデルに応用することも可能である。

【0044】

(7) 上述したように、本発明の音声解析では、不明瞭な歌声、鼻歌、楽器音などからもピッチ周波数を安定かつ確実に検出できる。これを応用することによって、従来は評価が困難であった不明瞭な歌声などについても、歌唱の正確さを的確に評価判定するカラオケシステムを実現することができる。

30

また、ピッチ周波数やその変化を画面に表示することにより、歌声の音程や抑揚やピッチ変化を可視化することが可能になる。このように可視化された音程や抑揚やピッチ変化を参考にすることにより、正確な音程や抑揚やピッチ変化をより短時間に感覚的に習得することが可能になる。さらに、上級者の音程や抑揚やピッチ変化を可視化してお手本とすることにより、上級者の音程や抑揚やピッチ変化をより短時間に感覚的に習得することも可能になる。

【0045】

(8) また、本発明の音声解析を実施することにより、従来は困難であった不明瞭な鼻歌やアカペラからもピッチ周波数を検出できるため、安定かつ確実に譜面を自動作成することが可能になる。

40

【0046】

(9) 本発明の音声解析を、言語教育システムに応用することも可能である。すなわち、本発明の音声解析を用いることにより、不馴れな外国語や標準語や方言の発話音声からもピッチ周波数を安定かつ確実に検出することができる。このピッチ周波数に基づいて、外国語や標準語や方言の正しいリズムや発音を誘導する言語教育システムを構築することが可能になる。

【0047】

50

(10) さらに、本発明の音声解析を、台詞指導システムに応用することも可能である。すなわち、本発明の音声解析を用いることにより、不馴れな台詞のピッチ周波数を安定かつ確実に検出することができる。このピッチ周波数を、上級者のピッチ周波数と比較することにより、台詞の指導や更には演出を行う台詞指導システムを構築することが可能になる。

【0048】

(11) また、本発明の音声解析を、ボイストレーニングシステムに応用することも可能である。すなわち、音声のピッチ周波数から、音程の不安定さや、発声方法の間違いを検出してアドバイスなどを出力することにより、正しい発声方法を指導するボイストレーニングシステムを構築することが可能になる。

10

【0049】

[感情推定で得られる心的状態の応用例]

(1) 一般に、心的状態の推定結果は、心的状態に反応して処理を変化させる製品全般に使用が可能である。例えば、相手の心的状態に応じて応答(性格、会話特性、心理特性、感性、感情パターン、または会話分岐パターンなど)を変化させる仮想人格(エージェント、キャラクターなど)をコンピュータ上で構築することが可能である。また例えば、お客様の心的状態に柔軟に応じて、商品検索、商品クレーム対応、コールセンター業務、受付システム、顧客感性分析、顧客管理、ゲーム、パチンコ、パチスロ、コンテンツ配信、コンテンツ作成、ネット検索、携帯電話サービス、商品説明、プレゼンテーション、または教育支援などを実現するシステムにも応用が可能となる。

20

【0050】

(2) また、心的状態の推定結果は、心的状態をユーザーに関する校正情報とすることで処理の正確性を高める製品全般にも使用が可能である。例えば、音声認識システムにおいて、認識された語彙の候補の中から、発話者の心的状態に対して親和度の高い語彙を選択することにより、音声認識の精度を高めることが可能になる。

【0051】

(3) さらに、心的状態の推定結果は、心的状態からユーザーの不正意図を推測することにより、セキュリティを高める製品全般にも使用が可能である。例えば、ユーザー認証システムでは、不安または演技などの心的状態を示すユーザーに対して、認証拒否をしたり、追加の認証を要求することによってセキュリティを高めることが可能になる。さらには、このような高セキュリティな認証技術を基礎として、ユビキタスシステムを構築することも可能である。

30

【0052】

(4) また、心的状態の推定結果は、心的状態を操作入力として扱う製品全般にも使用が可能である。例えば、心的状態を操作入力として処理(制御、音声処理、画像処理、またはテキスト処理など)を実行するシステムを実現することができる。また例えば、心的状態を操作入力としてキャラクター動作をコントロールすることによって、ストーリーを展開させるストーリー創作支援システムを実現することが可能になる。また例えば、心的状態を操作入力として、音律、キー、または楽器構成などを変更することにより、心的状態に沿った音楽創作や編曲を行う音楽創作支援システムを実現することも可能になる。また例えば、心的状態を操作入力として、照明、BGMなどの周辺環境をコントロールする演出装置を実現することも可能である。

40

【0053】

(5) さらに、心的状態の推定結果は、精神分析、感情分析、感性分析、性格分析、または心理分析を目的とする装置全般にも使用が可能である。

【0054】

(6) また、心的状態の推定結果は、音、音声、音楽、香り、色、映像、文字、振動、または光などの表現手段を用いて、心的状態を外部出力する装置全般にも使用が可能である。このような装置を使用することで、対人間における心情のコミュニケーションを支援することが可能になる。

50

## 【 0 0 5 5 】

( 7 ) さらに、心的状態の推定結果は、心的状態を情報通信する通信システム全般にも使用が可能である。例えば、感性通信、または感性感情共鳴通信などに応用することができる。

## 【 0 0 5 6 】

( 8 ) また、心的状態の推定結果は、映像や音楽などのコンテンツが人間に与える心理的な効果を判定(評価)する装置全般にも使用が可能である。さらに、この心理効果を項目としてコンテンツを分類することで、心理効果の面からコンテンツ検索が可能になるデータベースシステムを構築することも可能になる。

なお、映像や音楽などのコンテンツそのものを、音声信号と同様に分析することにより、コンテンツ出演者や楽器演奏者の音声興奮度や感情傾向などを検出することも可能である。また、コンテンツの音声を音声認識または音素片認識することでコンテンツの特徴を検出することも可能である。このような検出結果に従ってコンテンツを分類することで、コンテンツの特徴を切り口にしたコンテンツ検索が可能になる。

10

## 【 0 0 5 7 】

( 9 ) さらに、心的状態の推定結果は、商品使用時におけるユーザー満足度などを心的状態によって客観的に判定する装置全般にも使用が可能である。このような装置を使用することにより、ユーザーにとって親しみやすい製品開発や仕様作成が容易になる。

## 【 0 0 5 8 】

( 1 0 ) さらに、心的状態の推定結果は、下記の分野などにも応用が可能である。

20

介護支援システム、カウンセリングシステム、カーナビゲーション、自動車制御、運転者の状態監視、ユーザーインターフェース、オペレーションシステム、ロボット、アバター、ネットショッピングモール、通信教育システム、Eラーニング、学習システム、マナー研修、ノウハウ学習システム、能力判定、意味情報判断、人工知能分野、ニューラルネットワーク(ニューロンも含む)への応用、確率モデルが必要なシミュレーションやシステムなどの判断基準や分岐基準、経済・金融などの市場シミュレーションへの心理要素入力、アンケート収集、芸術家の感情や感性の解析、金融信用調査、与信管理システム、占いなどのコンテンツ、ウェアラブルコンピュータ、ユビキタスネットワーク商品、人間の知覚判断の支援、広告業務、ビルやホールなどの管理、フィルタリング、ユーザーの判断支援、キッチンやバスやトイレなどの制御、ヒューマンデバイス、柔らかさ、通気性が変化する繊維との連動による被服、癒しやコミュニケーションを目的とした仮想ペットやロボット、プランニングシステム、コーディネーターシステム、交通支援制御システム、料理支援システム、演奏支援、DJ映像効果、カラオケ装置、映像制御システム、個人認証、デザイン、設計シミュレーター、購買意欲を刺激するシステム、人事管理システム、オーディション、仮想の顧客集団市場調査、陪審員・裁判員シミュレーションシステム、スポーツや芸術や営業や戦略などのイメージトレーニング、故人や先祖のメモリアルコンテンツ作成支援、生前の感情や感性のパターンを保存するシステムやサービス、ナビゲーション・コンシェルジェサービス、ブログ作成支援、メッセージャーサービス、目覚まし時計、健康器具、マッサージ器具、歯ブラシ、医療器具、生体デバイス、スイッチング技術、制御技術、ハブ、分岐システム、コンデンサシステム、分子コンピュータ、量子コンピュータ、ノイマン型コンピュータ、生体素子コンピュータ、ボルツマンシステム、AI制御、ファジー制御。

30

40

## 【 0 0 5 9 】

[備考：騒音環境下での音声信号の取得について]

本発明者は、騒音環境下においても、音声のピッチ周波数を良好に検出するため、次のような防音マスクを用いた計測環境を構築した。

## 【 0 0 6 0 】

まず、防音マスクの基材として防毒マスク( TOYO製 SAFETY No1880-1 )を調達する。この防毒マスクは、口に接して覆う部分がゴム製である。このゴムは周辺騒音によって振動するため、周辺騒音がマスク内に侵入する。そこで、このゴム部分にシリコン(日新レジ

50

ン株式会社製、クイックシリコーン、ライトグレー液状、比重1.3)を注入して重くする。さらに、防毒マスクの通気フィルタには、キッチンペーパー5枚以上とスポンジを多層に重ねて密閉性を高める。この状態のマスク室の中央部分に小型マイクをフィットさせて設ける。このように準備された防音マスクは、シリコンの自重と異質物の積層構造によって周辺騒音の振動を効果的に減衰させることができる。その結果、被験者の口周辺にマスク形態の小型防音室を設けることに成功し、周辺騒音の影響を抑えつつ、被験者の音声を良好に集音できるようになる。

【0061】

さらに、同様の防音対策を施したヘッドホンを被験者の耳に装着することにより、周辺騒音の影響をさほど受けずに、被験者と会話を行うことが可能になる。

10

なお、ピッチ周波数の検出には、上記の防音マスクが有効である。ただし、防音マスクの密閉空間が狭いために、音声がこもりやすい傾向となる。そのため、ピッチ周波数以外の周波数解析や音色の分析には適さない。そのような用途には、マスク同様の防音処理を施したパイプラインを防音マスクに通し、防音環境の外界(空気室)と通気させることが好ましい。この場合、呼吸に支障がないため、口だけでなく鼻も含めてマスクすることができる。この通気設備の追加によって、防音マスクにおける音声のこもりを低減することができる。さらに、被験者にとって息苦しさなどの不快感が少ないため、より自然な状態の音声を集音できるようになる。

【0062】

なお、本発明は、その精神または主要な特徴から逸脱することなく、他のいろいろな形で実施することができる。そのため、前述の実施例はあらゆる点で単なる例示に過ぎず、限定的に解釈してはならない。本発明の範囲は、特許請求の範囲によって示すものであって、明細書本文には、なんら拘束されない。さらに、特許請求の範囲の均等範囲に属する変形や変更は、すべて本発明の範囲内のものである。

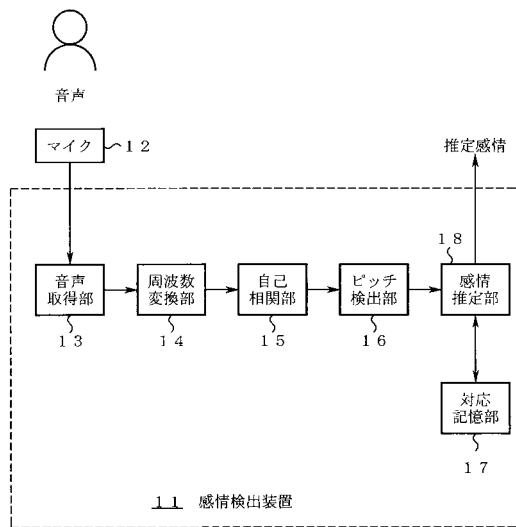
20

【産業上の利用可能性】

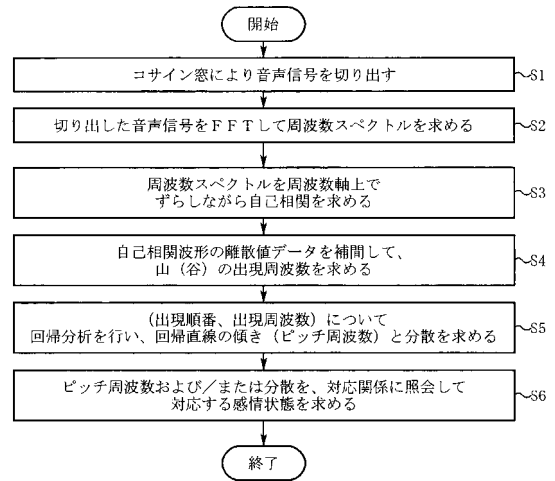
【0063】

以上説明したように、本発明は、音声解析装置などに利用可能な技術である。

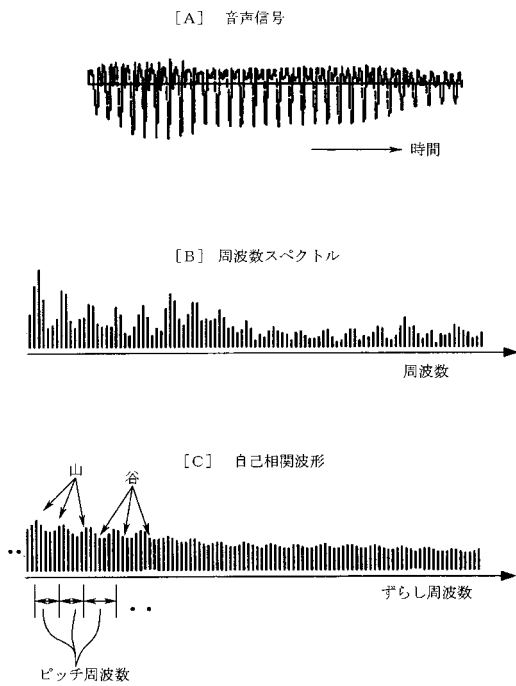
【図1】



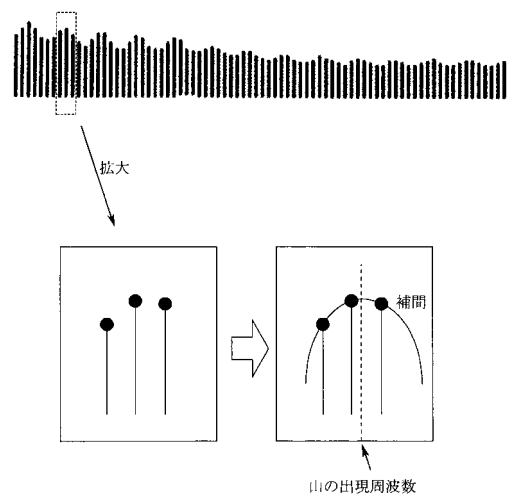
【図2】



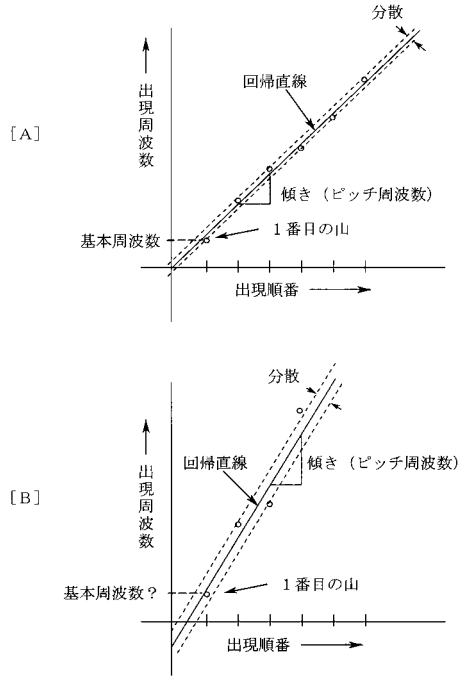
【図3】



【図4】



【図5】



---

フロントページの続き

- (72)発明者 尾形 薫  
東京都品川区北品川5丁目13-15-101
- (72)発明者 門間 史晃  
東京都品川区北品川5丁目13-15-101

審査官 菊池 智紀

- (56)参考文献 特開2004-240214(JP,A)  
国際公開第2006/112009(WO,A1)  
吉尾重治 他, "平方根及び4乗根パワースペクトルの自己相関に基づくピッチ抽出", 電子情報通信学会論文誌, 2001年 3月 1日, Vol.J84-A, No.3, p.436-440  
國枝伸行 他, "対数スペクトルの自己相関関数を利用したピッチ抽出法", 電子情報通信学会論文誌A, 1997年 3月, Vol.J80-A, No.3, p.435-443
- (58)調査した分野(Int.Cl., DB名)  
G10L 11/00-11/04  
JSTPlus(JDreamII)