

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4736632号
(P4736632)

(45) 発行日 平成23年7月27日(2011.7.27)

(24) 登録日 平成23年5月13日(2011.5.13)

(51) Int.Cl.		F I			
G 1 0 L	11/00	(2006.01)	G 1 0 L	11/00	1 0 1 Z
G 1 0 L	15/10	(2006.01)	G 1 0 L	11/00	4 0 2 H
			G 1 0 L	11/00	4 0 2 L
			G 1 0 L	15/10	5 0 0 N
			G 1 0 L	15/10	5 0 0 Z

請求項の数 4 (全 17 頁)

(21) 出願番号 特願2005-250454 (P2005-250454)
 (22) 出願日 平成17年8月31日(2005.8.31)
 (65) 公開番号 特開2007-65226 (P2007-65226A)
 (43) 公開日 平成19年3月15日(2007.3.15)
 審査請求日 平成20年3月27日(2008.3.27)

(出願人による申告)平成17年4月1日付け、支出負担行為担当官 総務省大臣官房会計課企画官、研究テーマ「ネットワーク・ヒューマン・インターフェースの総合的な研究開発(ネットワークロボットの技術)」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願

(73) 特許権者 393031586
 株式会社国際電気通信基礎技術研究所
 京都府相楽郡精華町光台二丁目2番地2
 (74) 代理人 100099933
 弁理士 清水 敏
 (72) 発明者 カルロス・トシノリ・イシイ
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内
 (72) 発明者 石黒 浩
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内
 (72) 発明者 萩田 紀博
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 ボーカル・フライ検出装置及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

発話信号中のボーカル・フライ区間を検出するためのボーカル・フライ検出装置であって、

発話信号を、第1のフレーム長でかつ第1のフレームシフト量の第1のフレームでフレーム化するための第1のフレーム化手段と、

前記第1のフレーム化手段の出力する一連の第1のフレームの各々のパワーのピークを検出するためのパワーピーク検出手段と、

前記発話信号を、前記第1のフレーム長よりも大きな第2のフレーム長で、かつ前記第1のフレームシフト量よりも大きな第2のフレームシフト量の第2のフレームでフレーム化するための第2のフレーム化手段と、

前記第2のフレーム化手段の出力する一連の第2のフレームの各々の内部における周期性の有無を判定するための周期性判定手段と、

前記パワーピーク検出手段により検出されたパワーピークのうちで、前記周期性判定手段により周期性がないと判定された前記第2のフレーム内のパワーピークを選択するためのパワーピーク選択手段と、

前記パワーピーク選択手段により選択されたパワーピークの各々について、当該パワーピークを含む所定区間内の他のパワーピークとの間の相互相関が所定のしきい値よりも大きなパワーピークを探索し、前記発話信号中の、当該パワーピークを含む所定の区間をボーカル・フライ区間として検出するための手段とを含む、ボーカル・フライ検出装置。

【請求項 2】

前記周期性判定手段は、前記一連の第 2 のフレームの各々において、当該フレーム内での最大パワーピークの、当該フレーム内の所定の遅延範囲内での自己相関値の関数としてフレーム内の周期性の尺度を算出し、当該自己相関値のピークが所定のしきい値関数よりも大きいか否かにしたがって、周期性があるか否かを判定するための手段と、

前記判定するための手段により周期性があると判定された前記第 2 のフレームのうち、前記周期性の尺度が予め定める定数よりも大きなフレームが所定個数連続している部分以外の前記第 2 のフレームの前記周期性の尺度の値を、周期性がないと判定される値に補正するための周期性補正手段を含む、請求項 1 に記載のボカール・フライ検出装置。

【請求項 3】

前記発話信号を前記第 1 のフレーム化手段及び前記第 2 のフレーム化手段に与えるに先立って、前記発話信号の所定の周波数帯域の成分以外の成分を除波するためのフィルタリング手段をさらに含む、請求項 1 又は請求項 2 に記載のボカール・フライ検出装置。

【請求項 4】

コンピュータにより実行されると、当該コンピュータを、請求項 1 ~ 請求項 3 のいずれかに記載のボカール・フライ検出装置として動作させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は人間の声質の分析技術に関し、特に、ボカール・フライ（以下「VF」と呼ぶ。）と呼ばれる特定の声質を持つ区間を発話信号中から検出するための VF 検出装置に関する。

【背景技術】

【0002】

人間と機械との対話において、音声に含まれるテキスト的な情報以外の情報（以下これを「パラ言語情報」と呼ぶ。）を自動的に抽出することが必要となる。従来、パラ言語情報を抽出するための音響特徴量として、ピッチ、パワー及び持続時間などの音韻的特徴量が使用されてきた。しかし、最近の研究では、咽頭の声の発生源のモードによる気息性、きしり、かすれなどの声質に関する情報もパラ言語情報の知覚に重要な役割を担っていることが報告されている。

【0003】

VF、きしり、きしみ声、声門フライ、パルス・レジスタ、及び喉頭収縮音（laryngealization）という用語が、比較的離散的な、喉頭（又は声門）の一連の励振（又は短い期間のパルス）のことを表わすものとして従来技術文献で使用されている。こうした声では、連続する声門パルスの中で、声道がほぼ完全に制動され、通常は基本周波数が非常に低く、声門周期の期間が不規則となる。VF を聞いたときの知覚は、「手すりに沿って棒を動かしたときの、速く、連続した連打音」、又は「モータボートのエンジン音の口真似」、又は「熱いフライパンで料理するときの音と似た音」、等と表現される。

【0004】

VF は、言語に依存するが、重要な言語的情報に加え、重要なパラ言語的情報を伝える。ドイツ語では、形態素の境界付近で VF がよく生ずる。日本語では、緊張の解けた低い声で VF が生ずる他に、りきみ声などのように感情に満ちた強調を伴う発話でも生ずる。りきみ声は、驚き、賞賛、及び苦しみなどについての感情又は態度に主に関連するパラ言語的情報を伝える。そのようなりきみ声における VF 発話部分（以下「VF セグメント」と呼ぶ。）では、非常に低い基本周波数が見られる。

【0005】

さらに、VF セグメントには、不規則性を持つという特徴があるため、音韻情報の抽出において重要な役割を担うピッチ決定アルゴリズムに重大な誤りを引き起こすことがある。したがって、VF がどこに生じているかを知れば、パラ言語情報の抽出に役立つだけでなく、ピッチの決定性能を改善する上でも重要である。

10

20

30

40

50

【 0 0 0 6 】

V Fの生理的、知覚的、及び音響的属性に関しては、いくつかの研究分野で報告されている。それらの多くは、様々な声質と関連した音響的特徴に関する定性的な、または説明的な事項を報告している。しかし、V Fについて、自動的な検出を目的とした評価についてはわずかしら報告されていない。

【非特許文献1】イシイ、C. T.、*「きしり声検出のための自己相関に基づくパラメータの分析」*、第2回音声韻律学国際会議予稿集、pp. 643 - 646、2004年。(Ishii, C.T., "Analysis of Autocorrelation-based parameters for Creaky Voice Detection," Proc. of The 2nd International Conference on Speech Prosody: 643-646, 2004.)

10

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 7 】

V Fの基本周波数の範囲に関しては、一貫して、100 Hzより低く、平均が24 ~ 52 Hz付近にあることが報告されている。V Fにおける声門パルスは二つ、時には3つのパルスがごく短い間隔で生じ、それに続いて声門がかなり制動される。

【 0 0 0 8 】

V Fに関しては、時間領域、スペクトル領域、及びケプストラム領域での音響分析が多く報告されている。通常の方法では、固定長の短時間分析用フレームを用いて周期性（又は調波性：harmonicity）に関する属性を評価している。

20

【 0 0 0 9 】

固定長のフレームを用いると、V Fセグメントが非常に低い基本周波数を持っている（すなわち非常に長いパルス間隔を持っている）場合に問題が生ずる。標準的な（よく使用される）分析フレームのフレーム長は25ミリ秒から32ミリ秒程度であるが、そうした条件ではV Fセグメント中の分析フレーム中にたかだか一つしか声門パルスがないことが多く、時にはフレーム中に声門パルスが全く含まれない場合もある。分析フレーム中に少なくとも二つの声門パルスが存在していなければ、スペクトル中に調波構造を見出すことはできず、また声門パルス間の短期周期性を反映した相関性のピークが生ずることも難しい。

【 0 0 1 0 】

これに対する最も単純な対応策は、分析フレーム長を長くすることである。非特許文献1においては、適応的にフレーム長を変化させる技術を用いた、自己相関に基づく周期性の分析が行われている。しかし、そのような方法では問題の一部しか解決できない。なぜなら、大きな分析フレームには、異なるパルス間隔を持つ二つの声門パルスが含まれる可能性があるためである。そうした場合には、スペクトル中の調波構造が乱されるし、自己相関（又はケプストラム）のピークの大きさも下がってしまう。

30

【 0 0 1 1 】

それゆえに本発明の目的は、スペクトル中の調波構造の乱れや自己相関のピークの低下という問題を回避し、精度良くV F検出を行なうV F検出装置を提供することである。

【 0 0 1 2 】

本発明の他の目的は、スペクトル中の調波構造の乱れや自己相関のピークの低下という問題を回避し、声門パルスに同期した手法で精度良くV F検出を行なうV F検出装置を提供することである。

40

【 0 0 1 3 】

本発明のさらに他の目的は、適切な分析フレームを用いることで、スペクトル中の調波構造の乱れや自己相関のピークの低下という問題を回避し、声門パルスに同期した手法で精度良くV F検出を行なうV F検出装置を提供することである。

【課題を解決するための手段】

【 0 0 1 4 】

本発明の第1の局面に係るV F検出装置は、発話信号中のV F区間を検出するための装

50

置であって、発話信号を、第1のフレーム長でかつ第1のフレームシフト量の第1のフレームでフレーム化するための第1のフレーム化手段と、第1のフレーム化手段の出力する一連の第1のフレームの各々のパワーのピークを検出するためのパワーピーク検出手段と、発話信号を、第1のフレーム長よりも大きな第2のフレーム長で、かつ第1のフレームシフト量よりも大きな第2のフレームシフト量の第2のフレームでフレーム化するための第2のフレーム化手段と、第2のフレーム化手段の出力する一連の第2のフレームの各々の内部における周期性の有無を判定するための周期性判定手段と、パワーピーク検出手段により検出されたパワーピークのうちで、周期性判定手段により周期性がないと判定された第2のフレーム内のパワーピークを選択するためのパワーピーク選択手段と、パワーピーク選択手段により選択されたパワーピークの各々について、当該パワーピークを含む所定区間内の他のパワーピークとの間の相互相関が所定のしきい値よりも大きなパワーピークを探索し、発話信号中の、当該パワーピークを含む所定の区間をVF区間として検出するための手段とを含む。

10

【0015】

第1のフレームによりフレーム化された発話信号により、パワーピークを検出する。第2のフレームによりフレーム化された発話信号により、周期性の有無を判定する。第1のフレームは第2のフレームより短いフレーム長で、かつフレームシフト量も小さい。したがって、VFパルスのような、基本周波数の低い波形も第2のフレームを用いた場合より精度良く検出できる。一方、第2のフレームのフレーム長は第1のフレームより長いので、その中に周期性があるか否かをより精度良く判定できる。検出されたパワーピークのうちで、周期性のない部分に存在するものがVFパルスである可能性が高い。さらに、このようなVFパルス候補が、所定区間内の他の隣接するパルスとの間で高い相互相関を示せば、そのVFパルス候補がVFパルスである可能性はより高くなる。そうしたVFパルスに対応するパワーピークを含む区間をVF区間として検出することで、精度良くVF区間が検出できる。第1及び第2のフレームを処理に用いるので、信号処理に適したフレームを用いることができ、精度良くVF検出を行なうことができる。

20

【0016】

好ましくは、パワーピーク検出手段は、一連の第1のフレームのうち、当該フレームを含む所定区間内の他のフレームのいずれのパワーよりも大きく、その差が予め定められる第1のしきい値よりも大きなフレームをパワーピーク候補として検出するためのパワーピーク候補検出手段と、パワーピーク候補検出手段により検出されたパワーピーク候補のうち、当該フレームを含む、所定区間よりも広い区間内の各フレームのパワーよりも大きく、かつその差の最大値が予め定められる第2のしきい値よりも大きなフレームをパワーピークとして検出するための手段とを含む。

30

【0017】

より好ましくは、所定区間よりも広い区間は、発話信号において10ミリ秒に相当する期間である。

【0018】

さらに好ましくは、周期性判定手段は、一連の第2のフレームの各々において、当該フレーム内での最大パワーピークの、当該フレーム内の所定の遅延範囲内での自己相関値の関数としてフレーム内の周期性の尺度を算出し、当該自己相関値のピークが所定のしきい値関数よりも大きいかなにかを判定するための手段を含む。

40

【0019】

判定するための手段は、最大パワーピークに関する自己相関値に、当該フレーム内での最大パワーピークからの遅延量に関する単調減少関数となる関数を乗じて周期性の尺度を算出するようにしてもよい。

【0020】

好ましくは、所定のしきい値関数は、予め定められた0より大きく1より小さな定数に、単調減少関数を乗じて得られる。

50

【 0 0 2 1 】

より好ましくは、周期性判定手段はさらに、判定するための手段により周期性があると判定された第2のフレームのうち、周期性の尺度が予め定める定数よりも大きなフレームが所定個数連続している部分以外の第2のフレームの周期性の尺度の値を、周期性がないと判定される値に補正するための周期性補正手段を含む。

【 0 0 2 2 】

さらに好ましくは、発話信号を第1のフレーム化手段及び第2のフレーム化手段に与えるに先立って、発話信号の所定の周波数帯域の成分以外の成分を除波するためのフィルタリング手段をさらに含む。

【 0 0 2 3 】

本発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを、上記したいずれかのVF検出装置として動作させる。

【発明を実施するための最良の形態】

【 0 0 2 4 】

< 概略 >

フレーム長に関する問題を解決するために、本発明の発明者たちは、固定長の分析フレーム中において周期性が見出されない場合に声門パルスに同期した処理を行なうことにした。そのために、制動と低基本周波数というVFの属性に基づいて声門パルスの候補を検出する。これは、長いパルス間の間隔で生ずる制動には、発話信号の振幅包絡、すなわち局所的なパワーの曲線に、上下動が生ずるという現象に基づいている。

【 0 0 2 5 】

自動検出に伴うもう一つの問題は、多くの音響分析では、発話信号に関し、予めセグメント化された有音発話部分の時間的又はスペクトルの特徴を分析しているということである。子音及び非発話セグメントも含む発話全体からVFを自動的に検出するという実際の問題では、多くの挿入エラーが発生する可能性がある。なぜなら、そうしたセグメントもまた、通常は非周期性という特徴を有するためである。したがって問題は、VFにより生じた非周期性と、子音及び環境の非発話信号から生じた残響とをどのように区別するかということである。

【 0 0 2 6 】

この問題に関し、本実施の形態では、連続する（又は近接する）声門パルス間の類似性の尺度を評価することにより、問題の解決を試みる。この尺度は、二つの声門パルスの発生の間には、声門の構造は変化せず、したがって二つのタイミングでの声門の応答は類似しているだろうという仮定に基づいている。

【 0 0 2 7 】

< 構成 >

図1に、本発明の一実施の形態に係るボイカル・フライ検出装置122を採用した自動対話システム100のブロック図を示す。図1を参照して、この自動対話システム100は、入来する発話信号102に対する音声認識を行ない、音声認識結果130をテキストデータとして出力するための音声認識装置120と、発話信号102のうちのVF期間を検出し、VF区間情報132を出力するためのVF検出装置122とを含む。

【 0 0 2 8 】

自動対話システム100はさらに、音声認識装置120から音声認識結果130を、VF検出装置122からVF区間情報132を、それぞれ受け、VF区間情報132を用いたパラ言語情報処理と、音声認識結果130とを統合することにより発話者の意図を理解し、適切な応答となるテキスト情報及び声質情報を出力するための応答作成装置124と、応答作成装置124が応答を作成する際に参照する、音声のテキスト情報とパラ言語情報との組合せに対し適切な応答を作成するための知識を格納した知識ベース126と、応答作成装置124から出力された応答のテキスト情報を、応答作成装置124から指示された声質で音声合成し、音声信号104として出力するための音声合成装置128とを含む。音声信号104は図示しない回路でアナログ化され、増幅されてスピーカに供給され

10

20

30

40

50

る。

【 0 0 2 9 】

図 2 に、V F 検出装置 1 2 2 のブロック図を示す。図 2 を参照して、V F 検出装置 1 2 2 は、発話信号 1 0 2 のうち、周期性に関する大部分の情報が含まれている 1 0 0 ~ 1 5 0 0 H z の周波数成分のみを通過させるためのバンドパスフィルタ 1 6 0 を含む。1 0 0 H z 未満の周波数成分は直流成分及び徐々に上昇及び下降する成分であり、周期性分析に悪影響を与えるため、バンドパスフィルタ 1 6 0 により除波する。また 1 5 0 0 H z を超える周波数成分は、高周波数のノイズ成分を含むので、これも除波する。このバンドパスフィルタの通過帯域は、V F セグメント中の各声門パルスについて、パワーの曲線中からピークと谷とを検出できるような帯域に選ばれている。

10

【 0 0 3 0 】

V F 検出装置 1 2 2 はさらに、フレーム長が 5 ミリ秒、フレーム間隔が 2 . 5 ミリ秒のフレーム（これを本明細書では「超短期フレーム」と呼ぶ。）を用いてバンドパスフィルタ 1 6 0 の出力内の局所的なパワーのピークを V F のパルスの候補として検出し、ピーク位置情報 1 7 0 を出力するための超短期ピーク検出処理部 1 6 2 と、フレーム長 2 5 ~ 3 2 ミリ秒、フレーム長 1 0 又は 5 ミリ秒というよく用いられるフレーム（これを本明細書では「短期フレーム」と呼ぶ。）を使用し、バンドパスフィルタ 1 6 0 の出力中で V F が存在する可能性を示す、短期周期性のない部分をそれ以外の部分と区別して検出し、短期周期性情報 1 7 2 を出力するための短期周期性検出部 1 6 4 とを含む。

20

【 0 0 3 1 】

V F 検出装置 1 2 2 はさらに、超短期ピーク検出処理部 1 6 2 からピーク位置情報 1 7 0 を、短期周期性検出部 1 6 4 から短期周期性情報 1 7 2 を、それぞれ受け、ピーク位置情報 1 7 0 により示されるピークのうちから、短期周期性のない部分に存在するものを含むフレームを V F フレームの候補として選択し、V F 候補情報 1 7 6 として出力するための周期性検査部 1 6 6 と、周期性検査部 1 6 6 の出力する V F 候補情報 1 7 6 と、バンドパスフィルタ 1 6 0 の出力する 1 0 0 ~ 1 5 0 0 H z の周波数成分の発話信号 1 7 4 とを用い、前後の所定の範囲に類似したパルスを持つ V F 候補のみを V F とし、V F の存在する区間を示す V F 区間情報 1 3 2 を出力するための類似性検査部 1 6 8 とを含む。

【 0 0 3 2 】

図 3 に、超短期ピーク検出処理部 1 6 2 のブロック図を示す。図 3 を参照して、超短期ピーク検出処理部 1 6 2 は、バンドパスフィルタ 1 6 0 の出力する 1 0 0 ~ 1 5 0 0 H z の周波数成分の発話信号 1 7 4 を超短期フレームによりフレーム化するためのフレーム化処理部 1 9 0 と、フレーム化処理部 1 9 0 の出力する超短期フレームの各々に対し、パワー（これを「超短期パワー」と呼ぶ。）を算出し出力するための超短期パワー算出部 1 9 2 と、超短期パワー算出部 1 9 2 の出力する一連の超短期パワーのうち、最新の所定個数の値を格納するためのメモリ 1 9 4 と、メモリ 1 9 4 に記憶された超短期パワーのうち、前後 1 フレームの超短期パワーのいずれよりも大きく、かつその差がいずれも所定のパワーしきい値 $P_{w_{TH}}$ （例えば 6 ~ 7 d B）より大きなものを V F の声門パルスの候補と推定し、そのピーク位置をピーク位置情報 1 7 0 として出力するためのピーク比較部 1 9 6 と、ピーク比較部が使用するパワーしきい値 $P_{w_{TH}}$ を記憶するためのパワーしきい値記憶部 1 9 8 とを含む。

30

40

【 0 0 3 3 】

図 4 及び図 5 に、ピーク比較部 1 9 6 におけるピーク検出の原理を示す。図 4 を参照して、フレーム長 5 ミリ秒、フレーム間隔 2 . 5 ミリ秒の超短期フレームの各々について超短期パワー算出部 1 9 2 によりパワーを算出することにより、2 . 5 ミリ秒間隔でパワー値が得られる。これらパワー値のうち、矢印 2 1 0 , 2 1 2 , 2 1 4 , 2 1 6 , 2 1 8 等のように、前後のパワー値よりも大きなものがピーク候補となり得る。本実施の形態ではさらに、これらピーク候補の中で、次に示すような条件を充足するものをピーク候補とする。

【 0 0 3 4 】

50

図5を参照して、パワー値232の値が、前後2フレームのパワー値230及び234と比較してパワーしきい値 $P_{W_{TH}}$ より大きいものとする。本実施の形態では、そのような場合にこのパワー値を示すフレームをピーク候補とする。パワー値238のように、前後2フレームのパワー値236及び240との差のいずれかがパワーしきい値 $P_{W_{TH}}$ に満たないものはピーク候補から除外する。

【0035】

図6(A)及び(B)にそれぞれ、VFセグメントと非VFセグメント(以下「NFセグメント」と呼ぶ。)におけるピークのパワー上昇とパワー下降との分布について、実験で得られたものを示す。ここでのピーク上昇量及び下降量は、あるピークのパワー値と、そのピークより4フレーム前のフレームのパワー(すなわち、ピークの10ミリ秒前のパワー)との間の差のことをいう。図6(A)によれば、VFでは制動が起こるという特性を反映して、パワー値の上昇量と下降量との双方において、かなり大きな値が発生していることがわかる。それに対し、図6(B)によれば、NFセグメントでは、パワー値の上昇量と下降量との双方において、1~6dBの範囲が大部分であることがわかる。

10

【0036】

この図からはどの程度の値をVFとNFとを区別するためのしきい値(パワーしきい値)として選択すべきかは必ずしも明確ではない。このしきい値は後に述べるような実験の結果に基づき選択するが、例えば7dBという値を用いる。

【0037】

図2に示す短期周期性検出部164は、このようにして定められたピーク候補の各々に対して、超短期ピーク検出処理部162により抽出されたピーク候補のうちでVFセグメント中と思われるものをさらに選択する機能を持つ。

20

【0038】

図7を参照して、短期周期性検出部164は、バンドパスフィルタ160の出力を、フレーム長32ミリ秒、フレーム間隔10ミリ秒でフレーム化するためのフレーム化処理部250と、フレーム化処理部250の出力するフレーム化された発話信号を記憶するためのメモリ252と、メモリ252に記憶されたフレームごとの発話信号に基づく自己相関分析により、フレーム内周期性(Intra-frame periodicity: IFP)をフレームごとに算出するためのIFP算出部254と、IFP算出部254により各フレームについて算出されたIFP値を所定の周期性のしきい値関数 IFP_{TH} と比較し、IFP値のピークのいずれかがしきい値関数を下回っていれば周期性がないと判定して当該フレームのIFP値をヌルに設定するための周期性判定部258と、周期性判定部258により設定されたIFP値に基づき、IFP値がヌルでないフレームが3フレーム以上連続した場合のみ、短期周期性を持つセグメントと判定し、短期周期性を持つフレームか否かを示す短期周期性情報172を出力するための連続性検査部260と、周期性判定部258が使用する周期性のしきい値関数 IFP_{TH} を記憶するための周期性のしきい値関数記憶部262とを含む。

30

【0039】

IFP算出部254による自己相関分析でのIFP値は、最大ピークの相関値を「フレーム長/(フレーム長-遅延)」で正規化した値で定義される。この正規化は、遅延量が大きくなるにしたがって自己相関は小さくなるという、自己相関関数の単調減少関数としての特性に対する補償を行なうためである。

40

【0040】

IFP算出部254では、15ミリ秒より小さな遅延量(約66.7Hzより大きな基本周波数に対応)の自己相関ピークのみを周期性の分析対象とする。すなわち、分析フレーム内には少なくとも二つの声門周期が含まれることになる。

【0041】

周期性判定部258は、200Hzよりも大きな基本周波数に対応する自己相関ピークに対し、次のような処理を行なう。すなわち、66.7Hzより上の低調波の全てに関する周期性を検査する。この処理により、声門周期の繰返しによる周期性ではなく第1フォ

50

ルマント周辺の強い調波による周期性を誤って検出してしまうことを防止する。自己相関関数における低調波属性について、図8及び図9に示す。図8には1フレーム内に声門パルスを一つだけ含むVFに関する波形及び自己相関を、図9には高い基本周波数を持つ地声に関する波形及び自己相関を示す。これらは、女性話者の音声から抽出した母音/e/に関するセグメントでのものである。図8(B)及び図9(B)において、実線276及び296はしきい値関数を示す。しきい値関数は「所定の定数×(フレーム長-遅延量)/(フレーム長)」で定義される。所定の定数として、本実施の形態では0.5という値を用いる。しきい値関数もまた、自己相関関数が遅延に対する単調減少関数であるという属性を考慮したものとなっている。

【0042】

図9(B)を参照して、地声のセグメントでは、波形290(図9(A))に含まれる強い調波については、その低調波成分の自己相関294のピークも通常は大きい。66.7Hzより上の低調波(遅延が15ミリ秒以下、すなわち点線298より左側)の自己相関ピーク300は、しきい値関数296よりも高い。

【0043】

これに対し図8(B)を参照して、VFセグメントの波形270(図8(A))については、自己相関関数は強いピークを持つが、15ミリ秒以内の遅延(点線278より左側)では、低調波成分の多くは自己相関関数274の値としてしきい値関数276よりも小さな値280を持つ。本実施の形態では、IFP算出部254は、このように各低調波成分の自己相関関数を算出する機能を持つ。周期性判定部258は、IFP算出部254により各フレームに対し算出されたIFP値を検査し、そのピークのいずれかがしきい値関数の値より小さければそのフレームのIFPの値をヌルに設定する機能を持つ。連続性検査部260は、周期性判定部258が出力する各フレームに対するIFP値を検査し、IFP値がヌルとなっていないフレームが少なくとも3個連続した場合のみ、それらフレームに短期周期性があるものと判定し、それ以外の場合には短期周期性がないものと判定する。

【0044】

図10(A)及び(B)にそれぞれ、VFセグメントとNFセグメントとに対し実験で得られたIFP値の分布を白い棒グラフで示す。図10(A)及び(B)を参照して、VFセグメントではIFPの値がヌルであるフレームが圧倒的に多数であることがわかる。図10において、「null_1」は低調波成分に関する制約によりIFP値がヌルとなったフレーム(すなわち、強い自己相関ピークが存在するが、低調波には弱い自己相関ピークしか存在しないフレーム)の数を示し、「null_2」は非周期性という制約によりIFP値がヌルとなったフレーム(すなわち強い自己相関ピークがないフレーム)の数を示す。

【0045】

図2に示す周期性検査部166は、超短期ピーク検出処理部162からVFセグメント候補のピーク位置情報170を、短期周期性検出部164からは短期周期性情報172を、それぞれ受け、IFP値がヌルとなっているフレームのピーク候補のみを選択し、VF候補情報176として類似性検査部168に与える機能を持つ。

【0046】

図11に、図2に示す類似性検査部168のブロック図を示す。図11を参照して、類似性検査部168は、100~1500Hzの周波数成分の発話信号174と、周期性検査部166からのVF候補情報176とに基づき、以上述べた制約をクリアしたVFセグメントのパワーピーク候補に対し、各パワーピーク付近の波形とその前のパワーピーク付近の波形との間の相互相関関数として計算されるパルス間類似性(inter-pulse similarity: IPS)値を算出するためのIPS算出部310と、後述するような実験により定められたしきい値 IPS_{TH} を記憶するためのパルス間類似性のしきい値記憶部314と、IPS算出部310から出力されるパワーピークごとのIPS値と、しきい値記憶部314に記憶されたしきい値 IPS_{TH} とを比較し、しきい値IPS

10

20

30

40

50

TH を上回るパワーピークのみを選択し、ピーク位置情報を出力するためのIPS比較部312と、IPS比較部312から出力されたピーク位置情報に基づき、隣接する(又は所定のサーチ範囲内で近接する)パルスの中でIPS値の高いものの間に存在するフレームをVFセグメントとしてマージし、VF区間情報132を出力するためのVFセグメント決定部316とを含む。

【0047】

IPS算出部310で算出されるIPS値は、前述したとおり処理対象のパワーピーク付近の波形と、その前のパワーピーク付近の波形との間の相互相関関数により算出される。相互相関計算のためのフレーム長は15ミリ秒に限定する。これは、不規則な間隔を持つ声門パルスによる、類似度計算における干渉を避けるためである。

10

【0048】

相互相関は、パワーピーク位置を中心とする、幅5ミリ秒の範囲に対し推定され、その最大値をIPS値とする。IPS値が高ければ、そのパワーピークがVFパルスを表わすものである確率が高いと考えられる。IPS値の算出においては、対象のパワーピークの前100ミリ秒の範囲に限定して他のパワーピークを探索し、そのパワーピークとの間で相互相関を算出する。100ミリ秒という値は、二つの声門の励振パルスの間の間隔として可能な最大時間間隔に対応する。励振パルスの最大値とは、基本周波数にして10Hzという非常に低い値に対応する値である。

【0049】

図10(A)及び(B)にそれぞれ、VFセグメントとNFセグメントとについて実験で算出されたIPS値の分布をハッチングした棒グラフで示す。図10(A)によれば、VFセグメントではIPS値は大きいものが圧倒的に多く、0.8~0.95の範囲を中心として集まっている。これに対しNFセグメントでは、null_2に大きな値がある。「null_2」は、探索範囲が100ミリ秒に限定されているためにヌル値に設定されたもの、つまりパワーピークの直前100ミリ秒の範囲に、他のパワーピークが存在しないためにIPS値がヌルに設定されたものを示す。一方、図10(A)ではIPS値のヌル値はほとんどない。

20

【0050】

また、図10(B)を参照して、NFセグメントではIPS値を二つのグループに分けることができる。一方はIPS値の低い範囲のグループであり、他方はIPS値の高い範囲のグループである。これらIPS値の高いものは、おそらく地声における周期性による結果と思われる。したがってこの場合にはIFP値もまた高いはずである。これに対応して、図10(B)の白い棒グラフにより、NFセグメントにおいてIFP値の高いものが多く存在していることが示されている。

30

【0051】

<動作>

以上述べた構成を有する自動対話システム100、特にVF検出装置122は以下のように動作する。図1を参照して、マイクロフォンなどから入力された発話信号102はデジタル化されて音声認識装置120及びVF検出装置122に与えられる。音声認識装置120は、この音声信号に対して音声認識処理を行ない、可能性の高い複数個の音声認識結果のテキスト情報からなる音声認識結果130を応答作成装置124に与える。一方、VF検出装置122は、以下に説明するような動作をして音声信号中でVFセグメントと思われるフレームを特定し、VF区間情報132を応答作成装置124に与える。

40

【0052】

応答作成装置124は、音声認識装置120から与えられた音声認識結果130に含まれる複数個の候補と、VF検出装置122から与えられるVF区間情報132とを用いて知識ベース126にアクセスすることにより、音声認識結果の候補とVFセグメントとの組合せから応答として最も適切と思われる応答を作成する。この応答は、応答のテキスト情報と、応答音声の声質を指定する情報とからなり、音声合成装置128に与えられる。音声合成装置128は、指定されたテキスト情報を指定された声質で再生するための音声

50

信号 104 を合成し、スピーカに与える。

【0053】

以下、VF 検出装置 122 の動作について説明する。図 2 を参照して、VF 検出装置 122 に与えられた発話信号 102 は、バンドパスフィルタ 160 に与えられる。バンドパスフィルタ 160 は、発話信号 102 のうち 100 Hz ~ 1500 Hz の周波数成分のみを発話信号 174 として通過させる。発話信号 174 は超短期ピーク検出処理部 162、短期周期性検出部 164、及び類似性検査部 168 に与えられる。

【0054】

超短期ピーク検出処理部 162 は、以下のような処理により超短期フレームでのパワーのピークを検出し、ピーク位置情報 170 として周期性検査部 166 に与える。すなわち、図 3 を参照して、フレーム化処理部 190 が 100 ~ 1500 Hz の周波数成分の発話信号 174 を超短期フレームによりフレーム化する。この超短期フレームは、フレーム長が 5 ミリ秒、フレーム間隔が 2.5 ミリ秒である。超短期フレームによりフレーム化された音声信号は超短期パワー算出部 192 に与えられる。

【0055】

超短期パワー算出部 192 は、各フレームに対し超短期パワーを算出し、結果をメモリ 194 に与え、記憶させる。メモリ 194 は最新の所定個数のフレームについて、その超短期パワーの値を記憶する。

【0056】

ピーク比較部 196 は、各フレームについて、その前後 2 フレームと比較してパワーがパワーしきい値 P_{WTH} より大きいフレームをパワーピーク候補とし、そのフレーム位置を示すピーク位置情報 170 を出力し、周期性検査部 166 に与える。

【0057】

一方、図 2 に示す短期周期性検出部 164 は以下のようにして各フレームにおける周期性を検出し、短期周期性情報 172 として周期性検査部 166 に与える。すなわち、図 7 を参照して、フレーム化処理部 250 は発話信号をフレーム長 32 ミリ秒、フレーム間隔 10 ミリ秒でフレーム化し、メモリ 252 に記憶させる。

【0058】

IFP 算出部 254 は、メモリ 252 に記憶された各フレームについて、IFP 値を算出し、周期性判定部 258 に与える。周期性判定部 258 は、IFP 算出部 254 から与えられた各フレームの IFP 値を、しきい値関数と比較することにより補正する。すなわち周期性判定部 258 は、各フレームについて、その低調波の IFP 値のいずれかがしきい値より小さければ、そのフレームの IFP 値をヌルに設定する。周期性判定部 258 は、この IFP 値をフレームごとに連続性検査部 260 に与える。

【0059】

連続性検査部 260 は、周期性判定部 258 から与えられたフレームごとの IFP 値について、その値がヌルでないフレームが少なくとも 3 フレームだけ連続していなければ、それらフレームの IFP 値をヌルに補正する。連続性検査部 260 により連続性が検査された後の各フレームの IFP 値は短期周期性情報 172 として図 2 に示す周期性検査部 166 に与えられる。

【0060】

周期性検査部 166 は、超短期ピーク検出処理部 162 から与えられたピーク位置情報 170 のうち、短期周期性検出部 164 から与えられた短期周期性情報 172 により、フレームの IFP 値がヌルとなっている部分のみを VF セグメントの候補とし、VF 候補情報 176 として類似性検査部 168 に与える。

【0061】

図 11 を参照して、類似性検査部 168 の IPS 算出部 310 は、VF 候補情報 176 により特定されるパワーピーク候補に対し、各パワーピーク付近の波形とその前のパワーピーク付近の波形との間の IPS 値を算出し、IPS 比較部 312 に与える。IPS 比較部 312 は、IPS 算出部 310 により算出された各パワーピークに対する IPS 値と、

10

20

30

40

50

しきい値記憶部 314 に記憶されたしきい値 $IPSTH$ とを比較し、しきい値 $IPSTH$ を上回るパワーピークのみを選択し、ピーク位置情報を出力する。このピーク位置情報は VF セグメント決定部 316 に与えられる。VF セグメント決定部 316 は、IPS 比較部 312 から出力されたピーク位置情報に基づき、隣接する（又は所定のサーチ範囲内で近接する）パルスの中で IPS 値の高いものの間のフレームを VF セグメントとしてマージし、VF 区間情報 132 を出力する。この VF 区間情報 132 が図 1 に示す応答作成装置 124 に与えられる。

【0062】

< 自動検出の評価 >

上記した実施の形態による VF 検出装置 122 の VF に関する自動検出を、自動検出された VF セグメントの持続期間 (VF_{dur}) 及び人手により VF として判定されラベリングされた期間 (VF_{dur_human}) を比較することにより評価した。以下、 VF_{dur} と VF_{dur_human} との比を VF 率と呼ぶ。VF とラベリングされたセグメントについては、VF 率が $2/3$ より大きい場合のみ正確に検出されたものと判定した。VF とラベリングされなかったセグメントについて自動検出により VF と判定されたもの数 (VF_{dur_ins}) を数えることにより、挿入エラーを検査した。検出結果及び挿入エラー結果を、検出性能又は挿入エラーの重大性によって二つのグループ、「検出」と「検出？」というグループに分けた。「検出？」グループは、VF 率が $1/3 \sim 2/3$ の範囲で「VF」として検出されたセグメントと、「 VF_{dur_ins} 」の値が 30 ミリ秒を下回るものを含んでいる。

【0063】

上記実施の形態に含まれる種々のパラメータに関し、いくつかの値の組合せをテストし、検出性能を低下させずに挿入エラーを減少させるようにした。最初に、IPS 値を 0.0、IFP 値を 1.0 に設定することにより、パワーピークのしきい値をリセットした。この条件は、パワーに関する情報のみを用いることに相当する。図 12 は、パワーのしきい値を様々に変えたときの検出結果を示す。図 12 を参照して、パワーのしきい値を高くすると、挿入エラーは減少する（「NF」グループの黒及び網掛けの部分）が、検出率も低下する（「VF」グループの黒及び網掛けの部分）ことが判る。

【0064】

次に、パワーのしきい値を 7 dB に固定し、IPS のしきい値を 0.0 に設定した。図 13 はこの条件での様々な IFP のしきい値についての検出結果を示す。図 13 を参照して、検出率はあまり変化しなかった（「VF」グループ）が、IFP のしきい値を 0.6 とすると挿入エラーをより削減できた（「NF」グループ）。

【0065】

最後に、パワーのしきい値を 7 dB に、IFP のしきい値を 0.6 にそれぞれ設定して、いくつかの IPS 値のしきい値について実験を行なった。図 14 を参照して、IPS 値のしきい値を 0.6 に設定すると、重大な挿入エラーをさらに削減することができた（「NF」グループの黒い部分）上に、検出率は好ましい値に維持することができた。

【0066】

「R」グループ（VF の特徴が人間には知覚されなかったセグメント）について、それらサンプルの大部分は自動検出でも VF としては検出されなかった。しかし、「VF？」グループでは、一部が「VF」として検出された。これらの結果によれば、本実施の形態に係る VF 自動検出装置によって、人間による知覚実験の結果とほぼ整合する結果が得られたといえる。

【0067】

全体的な検出率について、 VF_{dur} の合計を VF_{dur_human} の合計で割ることにより算出した。全体的な挿入誤り率については、 VF_{dur_ins} の合計を VF_{dur_human} の合計で割ることにより算出した。「パワー = 7 dB、IFP = 0.6、IPS = 0.6」というパラメータの組合せに対して、全体的な検出率として 73.3%、全体的な挿入エラー率として 3.9% という値が得られた。73.3% という検出率

10

20

30

40

50

については、検出結果を後処理することにより、さらに改善の余地がある。たとえば、近接したVFセグメントをマージする、などの方法により改善が可能と思われる。挿入エラー率がもう少し高くても問題が生じないアプリケーションにおいては、パラメータをさらに調整して検出率を高めることもできる。

【0068】

以上のように本実施の形態によれば、「パワー、IFP及びIPS」というパラメータの組合せを用いてボカール・フライを自動的に検出できる。

【0069】

<コンピュータによる実現および動作>

この実施の形態に係るVF検出装置122及び自動対話システム100は、コンピュータハードウェアと、そのコンピュータハードウェアにより実行されるプログラムと、コンピュータハードウェアに格納されるデータとにより実現できる。図15はこのコンピュータシステム330の外観を示し、図16はコンピュータシステム330の内部構成を示す。

10

【0070】

図15を参照して、このコンピュータシステム330は、FD(フレキシブルディスク)ドライブ352およびCD-ROM(コンパクトディスク読出専用メモリ)ドライブ350を有するコンピュータ340と、キーボード346と、マウス348と、モニタ342と、マイクロフォン370と、スピーカ372を含む。

【0071】

図16を参照して、コンピュータ340は、FDドライブ352およびCD-ROMドライブ350に加えて、CPU(中央処理装置)356と、CPU356、FDドライブ352およびCD-ROMドライブ350に接続されたバス366と、ブートアッププログラム等を記憶する読出専用メモリ(ROM)358と、バス366に接続され、プログラム命令、システムプログラム、および作業データ等を記憶するランダムアクセスメモリ(RAM)360と、マイク370から入力される発話信号をデジタル化したり、CPU356により処理されたデジタルの音声信号をアナログ化し、スピーカ372に与えたりするためのサウンドボード368とを含む。コンピュータシステム330はさらに、図示しないプリンタを含んでもよい。

20

【0072】

ここでは示さないが、コンピュータ340はさらにローカルエリアネットワーク(LAN)への接続を提供するネットワークアダプタボードを含んでもよい。

30

【0073】

コンピュータシステム330に本実施の形態に係る自動対話システム100及びVF検出装置122としての動作を行なわせるためのコンピュータプログラムは、CD-ROMドライブ350またはFDドライブ352に挿入されるCD-ROM362またはFD364に記憶され、さらにハードディスク354に転送される。または、プログラムは図示しないネットワークを通じてコンピュータ340に送信されハードディスク354に記憶されてもよい。プログラムは実行の際にRAM360にロードされる。CD-ROM362から、FD364から、またはネットワークを介して、直接にRAM360にプログラムをロードしてもよい。

40

【0074】

このプログラムは、コンピュータ340にこの実施の形態に係る自動対話システム100及びVF検出装置122としての動作を行なわせる複数の命令を含む。これら命令による処理を行なうのに必要な基本的機能のいくつかはコンピュータ340上で動作するオペレーティングシステム(OS)またはサードパーティのプログラム、もしくはコンピュータ340にインストールされる各種ツールキットのモジュールにより提供される。したがって、このプログラムはこの実施の形態の自動対話システム100及びVF検出装置122としての動作を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能または「

50

ツール」を呼出すことにより、上記した自動対話システム 100 及び V F 検出装置 122 としての動作を実行する命令のみを含んでいればよい。コンピュータシステム 330 の動作は周知であるので、ここでは繰り返さない。

【0075】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内のすべての変更を含む。

【図面の簡単な説明】

【0076】

【図1】本発明の一実施の形態に係る V F 検出装置 122 を採用した自動対話システム 100 のブロック図である。

【図2】本発明の一実施の形態に係る V F 検出装置 122 のブロック図である。

【図3】超短期ピーク検出処理部 162 のブロック図である。

【図4】超短期ピーク検出処理部 162 におけるピーク検出の原理を示す図である。

【図5】超短期ピーク検出処理部 162 におけるピーク検出の原理を示す図である。

【図6】V F セグメントと N F セグメントとにおけるピークのパワー上昇とパワー下降との分布について、実験で得られた結果を示すグラフである。

【図7】短期周期性検出部 164 のブロック図である。

【図8】1 フレーム内に一つの V F パルスが存在する場合の低調波の自己相関関数の属性を示す図である。

【図9】地声に関する低調波の自己相関関数の属性を示す図である。

【図10】V F 及び N F セグメントにおける I F P 及び I P S の分布を示すグラフである。

【図11】類似性検査部 168 のブロック図である。

【図12】I F P しきい値 = 1、I P S しきい値 = 0 に固定した場合で、いくつかのパワーのしきい値について行なった実験結果を示すグラフである。

【図13】パワーのしきい値 = 7 dB、I P S しきい値 = 0 に固定した場合で、いくつかの I F P のしきい値について行なった実験結果を示すグラフである。

【図14】パワーのしきい値 = 7 dB、I F P しきい値 = 0.6 に固定した場合で、いくつかの I P S しきい値について行なった実験結果を示すグラフである。

【図15】本発明の一実施の形態に係る自動対話システム 100 及び V F 検出装置 122 を実現するコンピュータの外観を示す図である。

【図16】図15に示すコンピュータの内部構成図である。

【符号の説明】

【0077】

100 自動対話システム

102, 174 発話信号

104 音声信号

120 音声認識装置

122 V F 検出装置

124 応答作成装置

126 知識ベース

128 音声合成装置

130 音声認識結果

132 V F 区間情報

160 バンドパスフィルタ

162 超短期ピーク検出処理部

164 短期周期性検出部

166 周期性検査部

10

20

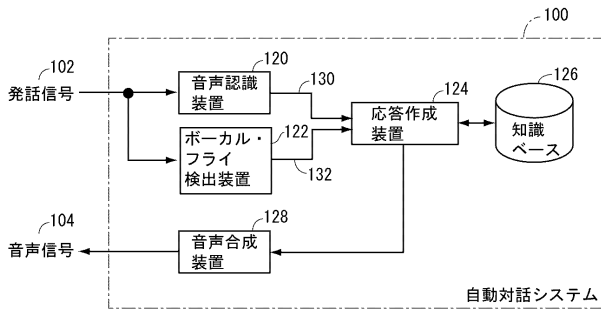
30

40

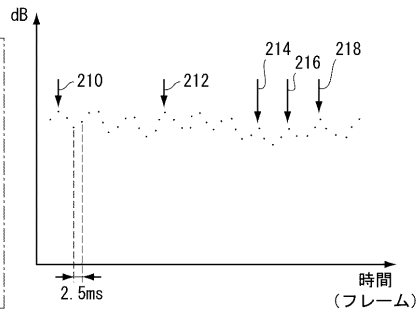
50

- 168 類似性検査部
- 170 ピーク位置情報
- 172 短期周期性情報
- 176 VF候補情報
- 190, 250 フレーム化処理部
- 192 超短期パワー算出部
- 194, 252 メモリ
- 196 ピーク比較部
- 254 I F P 算出部
- 258 周期性判定部
- 260 連続性検査部
- 310 I P S 算出部
- 312 I P S 比較部
- 314 しきい値記憶部
- 316 VFセグメント決定部

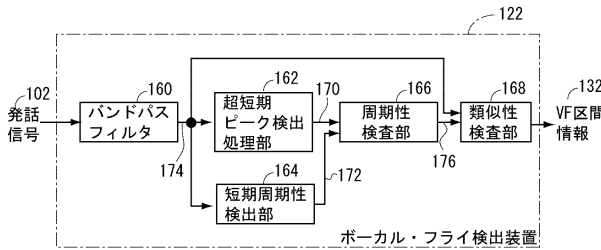
【図1】



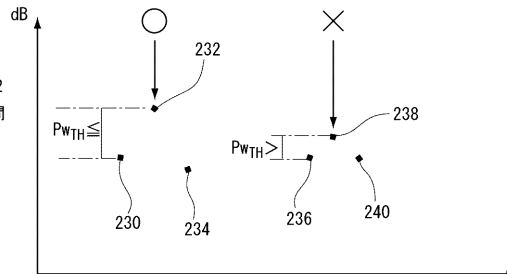
【図4】



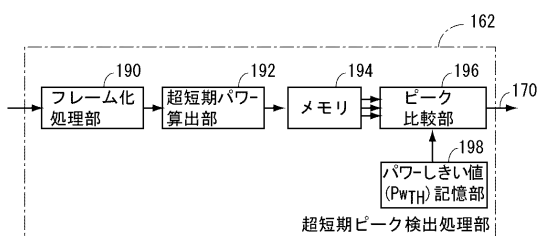
【図2】



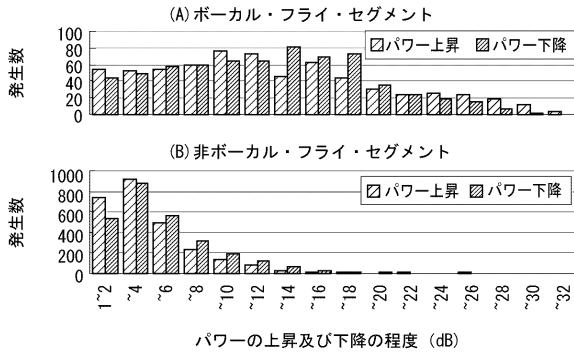
【図5】



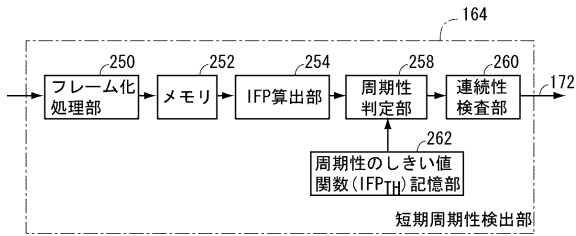
【図3】



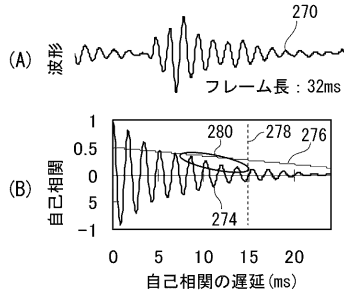
【図6】



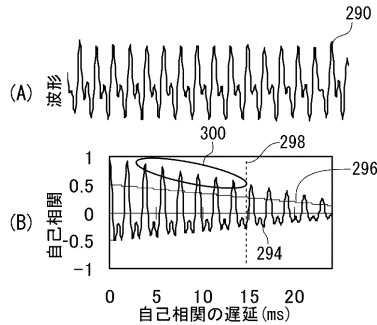
【図7】



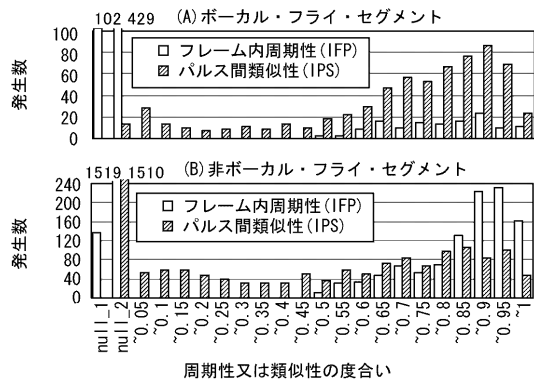
【図8】



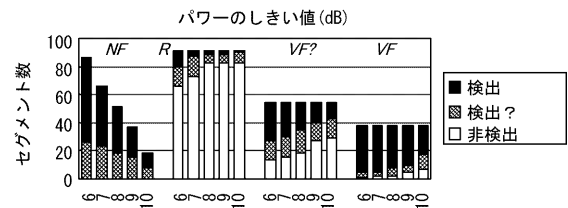
【図9】



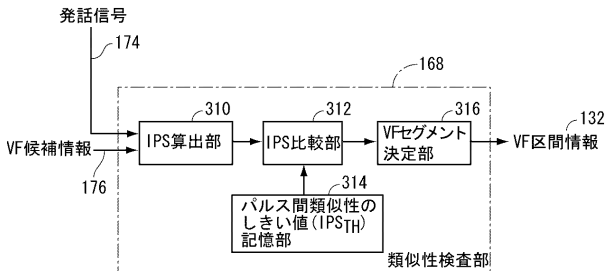
【図10】



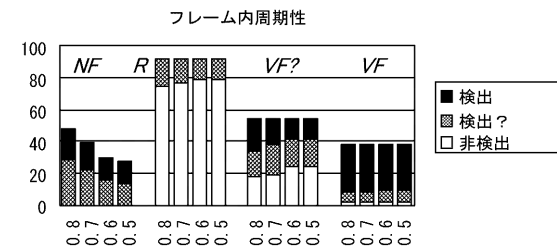
【図12】



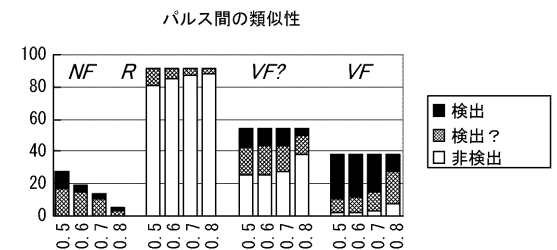
【図11】



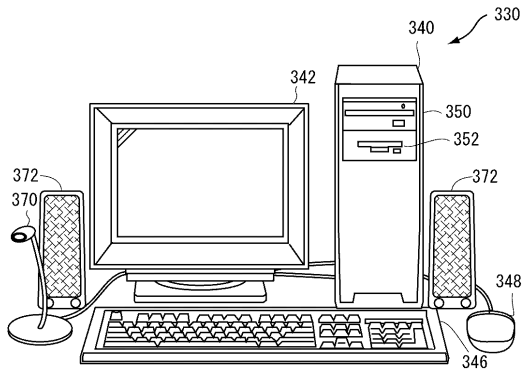
【図13】



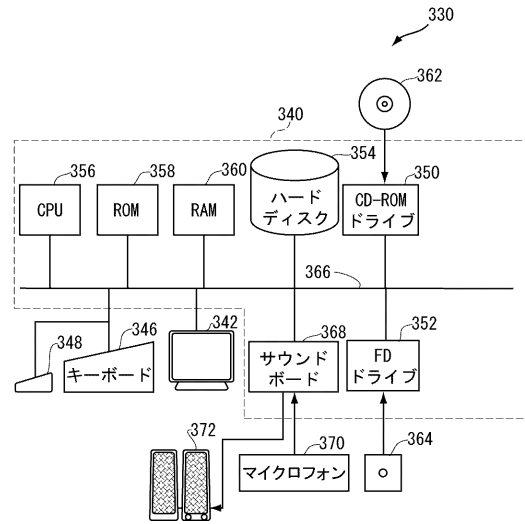
【図14】



【図15】



【図16】



フロントページの続き

審査官 山下 剛史

- (56)参考文献 ドナ・エリクソン, " 表現豊かな音声 - その生成・知覚と音声合成への応用 - ", 日本音響学会誌, Vol.61, No.6 (2005-06), pp.346-351
前川喜久雄他, " パラ言語情報の生成と知覚多次元尺度法による布置と音響特徴の関係 ", 電子情報通信学会技術研究報告, Vol.99, No.74, SP99-10 (1999-05), pp.9-16

(58)調査した分野(Int.Cl., DB名)

G 1 0 L 1 1 / 0 0 - 1 5 / 2 8

C i N i i