(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2009-3162
(P2009-3162A)
(43)公開日　平成21年1月8日(2009.1.8)

(51)Int.Cl.		F I			テーマコード（参考）
G10L 15/10	(2006.01)	G10L	15/10	500N	5D015
G10L 11/00	(2006.01)	G10L	11/00	402H	
G10L 11/04	(2006.01)	G10L	11/00	101F	
G10L 15/06	(2006.01)	G10L	11/04		
G10L 15/20	(2006.01)	G10L	15/06	400V	

審査請求 未請求　請求項の数 14　ＯＬ　（全 38 頁）　最終頁に続く

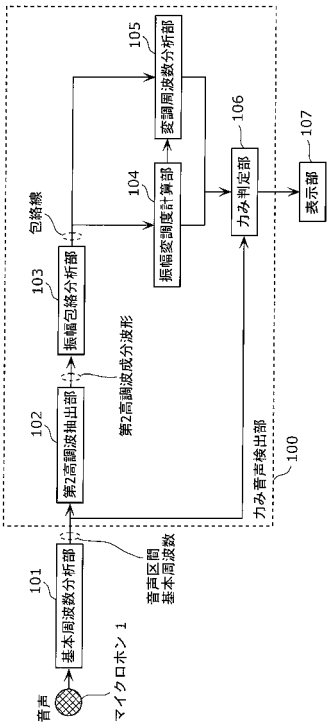
(21)出願番号　特願2007-163676 (P2007-163676)
(22)出願日　平成19年6月21日 (2007.6.21)

(71)出願人　000005821
パナソニック株式会社
大阪府門真市大字門真１００６番地
(74)代理人　100109210
弁理士　新居　広守
(72)発明者　加藤　弓子
大阪府門真市大字門真１００６番地　松下電器産業株式会社内
(72)発明者　釜井　孝浩
大阪府門真市大字門真１００６番地　松下電器産業株式会社内
(72)発明者　廣瀬　良文
大阪府門真市大字門真１００６番地　松下電器産業株式会社内

Ｆターム(参考)　5D015 AA02　HH05

(54)【発明の名称】力み音声検出装置

(57)【要約】
【課題】音韻単位という短い時間単位で感情を検出可能で、かつ個人差、言語差、地方差の影響を受けずに発話者の怒りや苛立ちを検出することのできる力み音声検出装置を提供する。
【解決手段】入力音声信号が話者が力んだ状態で発声した音声の信号であるか否かを判断する力み音声検出装置であって、入力音声信号の振幅包絡を抽出する振幅包絡分析部１０３と、前記振幅包絡分析部１０３によって抽出された振幅包絡の周期的変動を検出し、検出された周期的変動の周波数を求める変調周波数分析部１０５と、前記変調周波数分析部１０５によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に、前記入力音声信号は話者が力んだ状態で発声した音声の信号であると判定する力み判定部１０６とを備える。
【選択図】図８



【特許請求の範囲】
【請求項１】
入力音声信号が話者が力んだ状態で発声した音声の信号であるか否かを判定する力み音声検出装置であって、
入力音声信号の振幅包絡を抽出する振幅包絡抽出手段と、
前記振幅包絡抽出手段によって抽出された振幅包絡の周期的変動を検出し、検出された周期的変動の周波数を求める変調周波数分析手段と、
前記変調周波数分析手段によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に、前記入力音声信号は話者が力んだ状態で発声した音声の信号であると判定する力み判定手段と
を備える力み音声検出装置。
【請求項２】
さらに、前記入力音声信号の振幅包絡の振幅変動度合いを示す振幅変調度を計算する振幅変調度計算手段を備え、
前記力み判定手段は、前記振幅変調度計算手段によって求められた前記振幅変調度があらかじめ定められた値以上であり、かつ前記変調周波数分析手段によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に入力音声が力み音声であると判定する
請求項１に記載の力み音声検出装置。
【請求項３】
前記振幅変調度計算手段は、前記振幅包絡抽出手段によって抽出された振幅包絡に対して多項式をフィッティングさせることにより発声時に変調のない振幅包絡を推定し、前記抽出された振幅包絡の値と前記推定された振幅包絡の値との差分値と前記推定された振幅包絡の値との比を前記振幅変調度として計算する
請求項２に記載の力み音声検出装置。
【請求項４】
前記力み判定手段は、前記振幅変調度計算手段によって求められた前記振幅変調度が０．０２以上１．００以下のあらかじめ定められた値以上であり、かつ前記変調周波数分析手段によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に入力音声が力み音声であると判定する
ことを特徴とする請求項３に記載の力み音声検出装置。
【請求項５】
前記振幅包絡抽出手段は、
前記入力音声信号の基本波を抽出する基本波抽出手段と、
前記入力音声信号の振幅包絡として、前記基本波抽出手段によって抽出された基本波の振幅包絡を抽出する振幅包絡分析手段とを有する
請求項１～４のいずれか１項に記載の力み音声検出装置。
【請求項６】
前記振幅包絡抽出手段は、
前記入力音声信号の基本周波数の所定の整数倍の波である高調波を抽出する高調波抽出手段と、
前記入力音声信号の振幅包絡として、前記高調波抽出手段によって抽出された高調波の振幅包絡を抽出する振幅包絡分析手段とを有する
請求項１～４のいずれか１項に記載の力み音声検出装置。
【請求項７】
前記力み判定手段における前記あらかじめ定められた範囲の下限値は１０Ｈｚである
請求項１～６のいずれか１項に記載の力み音声検出装置。
【請求項８】
前記力み判定手段における前記あらかじめ定められた範囲は、１０Ｈｚ以上かつ１７０Ｈｚ未満である

請求項７に記載の力み音声検出装置。
【請求項９】
入力音声信号に含まれる音韻の種類を特定する音声認識装置であって、
請求項１～８のいずれか１項に記載の力み音声検出装置と、
音韻の種類ごとに、発話者が力んだ状態で発声した音声の特徴量を記憶している力み音声特徴量データベースと、
音韻の種類ごとに、発話者が通常状態で発声した音声の特徴量を記憶している標準音声特徴量データベースと、
前記力み音声検出装置において入力音声信号が力んだ状態で発声した音声の信号であると判定された場合には、前記力み音声特徴量データベースを用いて前記入力音声信号に含まれる音韻の種類を特定し、前記力み音声検出装置において前記入力音声信号が力んだ状態で発声した音声の信号であると判定されなかった場合には、前記標準音声特徴量データベースを用いて前記入力音声信号に含まれる音韻の種類を特定する音声認識手段と
を備える音声認識装置。
【請求項１０】
入力音声信号に含まれる音韻の種類を特定する音声認識装置であって、
請求項１～８のいずれか１項に記載の力み音声検出装置と、
音韻の種類ごとに音響特徴量を記憶している音響特徴量データベースと、
少なくとも読みまたは発音記号を有する単語辞書を表す言語特徴量を含む言語特徴量データベースと、
前記力み音声検出装置において入力音声信号が力んだ状態で発声した音声の信号であると判定された場合には、前記音響特徴量データベースに含まれる音響特徴量を用いた確率モデルの重みよりも前記言語特徴量データベースに含まれる言語特徴量を用いた確率モデルの重みを大きくし、重み付けされた２つの確率モデルを用いて前記入力音声信号に含まれる音韻の種類を特定する音声認識手段と
を備える音声認識装置。
【請求項１１】
入力音声信号に含まれる音韻の種類を特定するとともに話者の怒りの強度を認識する怒り認識機能付音声認識装置であって、
請求項９または１０に記載の音声認識装置と、
音韻の属性情報から発話時の力みやすさを求めるための規則を用いて、前記音声認識装置で音韻の種類が認識された音韻ごとに、発話時の力みやすさを示す力み音声発声指標を計算する力み音声発生指標計算手段と、
前記音声認識装置が備える力み検出装置により話者が力んだ状態で発声した音声の信号であると判定された入力音声信号について、前記力み音声発生指標が小さいほど怒りの強度が高くなる規則に基づいて、前記力み音声発生指標計算手段で計算された前記力み音声発生指標から怒りの強度を決定する怒り強度決定手段と
を備える怒り認識機能付音声認識装置。
【請求項１２】
入力音声信号が話者が力んだ状態で発声した音声の信号であるか否かを判定する力み音声検出方法であって、
入力音声信号の振幅包絡を抽出する振幅包絡抽出ステップと、
前記振幅包絡抽出ステップにおいて抽出された振幅包絡の周期的変動を検出し、検出された周期的変動の周波数を求める変調周波数分析ステップと、
前記変調周波数分析ステップにおいて求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に、前記入力音声信号は話者が力んだ状態で発声した音声の信号であると判定する力み判定ステップと
を含む力み音声検出方法。
【請求項１３】
入力音声信号が話者が力んだ状態で発声した音声の信号であるか否かを判定するプログ

ラムであって、
入力音声信号の振幅包絡を抽出する振幅包絡抽出ステップと、
前記振幅包絡抽出ステップにおいて抽出された振幅包絡の周期的変動を検出し、検出された周期的変動の周波数を求める変調周波数分析ステップと、
前記変調周波数分析ステップにおいて求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に、前記入力音声信号は話者が力んだ状態で発声した音声の信号であると判定する力み判定ステップと
をコンピュータに実行させるためのプログラム。
【請求項１４】
入力音声信号が話者が力んだ状態で発声した音声の信号であるか否かを判定する集積回路であって、
入力音声信号の振幅包絡を抽出する振幅包絡抽出手段と、
前記振幅包絡抽出手段によって抽出された振幅包絡の周期的変動を検出し、検出された周期的変動の周波数を求める変調周波数分析手段と、
前記変調周波数分析手段によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に、前記入力音声信号は話者が力んだ状態で発声した音声の信号であると判定する力み判定手段と
を備える集積回路。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、発話者の発話する音声中より「力み」を検出する力み音声検出装置に関する。
【背景技術】
【０００２】
自動電話応対システム、電子秘書システムおよび対話ロボット等、音声対話によるインタフェースを持つ対話システムにおいては、対話システムが、ユーザの要求により適切に対応するため、ユーザが発声した音声からユーザの感情を理解することが重要な要件となる。例えば、上記のような対話システムが、ユーザと音声による対話を行う際、対話システムの音声認識は必ずしも正確に音声を認識するとは限らない。対話システムが誤認識を起こした場合には、対話システムはユーザに対して再度音声入力を要求する。このような状況において、ユーザは少なからず怒りを覚えたり、苛立ったりする。誤認識が重なればなおさらである。怒りや苛立ちはユーザの話し方や声質を変化させ、平常時の音声とは異なるパターンとなる。このため、平常時の音声を認識用モデルとして保持する対話システムは、さらに誤認識をし易くなり、ユーザに対して同じ回答を何度も要求するなど、ユーザにとってさらに不愉快な要求をすることになる。
【０００３】
対話システムが上記のような悪循環に陥った場合、対話システムはそのインタフェースとしての用を成さなくなる。ユーザが発声する音声からユーザの怒りや苛立ちを検出することは、このような悪循環を断ち切り、対話システムとユーザとの間の音声対話を正常化するために必要不可欠である。すなわち、ユーザの怒りや苛立ちを理解することができれば、対話システムは誤認識したことに対してより丁寧な口調で聞き返したり、謝罪したりすることができる。これにより、ユーザの感情を平常に近づけ、平常の発話を導くことができ、対話システムは、音声認識率を回復することができる。ひいては、ユーザは、対話システムによる機器操作をスムーズに行うことができる。
【０００４】
従来、音声から感情を認識する方法としては、話者の発声した音声から声の高さ（基本周波数）、大きさ（パワー）、発話速度などの韻律的特徴を抽出し、入力音声全体に対して、「声が高い」「声が大きい」といった判断に基づき感情を認識する方式（例えば、特許文献１参照。）や、入力音声全体に対して、「高い周波数領域のエネルギーが大きい」

といった判断をする方式（例えば、特許文献１参照。）が知られている。また、音声のパワーと基本周波数とのシーケンスより、パワーおよび基本周波数の平均値、最大値および最小値といった統計的な代表値を求めて感情を認識する方式（例えば、特許文献２参照。）が知られている。さらに、文や単語のイントネーションやアクセントといった韻律の時間パターンを用いて感情を認識する方式（例えば、特許文献３参照。）が知られている。

【０００５】

図２１は、特許文献１に記載されている従来の音声による感情認識装置の構成を示すブロック図である。感情認識装置は、マイクロホン１と、音声コード認識手段２と、感性情報抽出手段３と、出力制御手段４と、出力装置５とを備えている。感性情報抽出手段３は、話速検出部３１と、基本周波数検出部３２と、音量検出部３３と、音声レベル判定基準格納部３４と、標準音声特徴量格納部３５と、音声レベル分析部３６と、感性レベル分析用知識ベース格納部３７と、感性レベル分析部３８と、音声スペクトル検出部３９とを備えている。出力制御手段４は、主制御部４１と、出力制御用知識ベース格納部４２とを備えている。

【０００６】

マイクロホン１は、入力音声を電気信号に変換する。音声コード認識手段２は、マイクロホン１から入力された音声の音声認識を行い、認識結果を感性情報抽出手段３および出力制御手段４へ出力する。一方、感性情報抽出手段３の話速検出部３１、基本周波数検出部３２、音量検出部３３は、マイクロホン１より入力された音声より、話速、基本周波数および音量を抽出する。音声レベル判定基準格納部３４には、入力された音声の話速、基本周波数および音量を標準の話速、基本周波数および音量とそれぞれ比較して音声レベルを決定するための基準が記憶されている。標準音声特徴量格納部３５には、音声レベルを判定する際の基準となる標準の発声速度、基本周波数および音量が記憶されている。音声レベル分析部３６は入力された音声の特徴量と標準の音声特徴量との比に基づいて、音声レベル（すなわち、話速レベルおよび基本周波数レベル）と音量レベルとを決定する。

【０００７】

また、感性レベル分析用知識ベース格納部３７は、音声レベル分析部３６で決定された各種音声レベルによって感性レベルを判定するルールを記憶している。感性レベル分析部３８は、音声レベル分析部３６からの出力と音声コード認識手段２からの出力とから、感性レベル分析用知識ベース格納部３７に記憶されているルールに基づき、感性レベルすなわち感性の種類とレベルとを判定する。出力制御手段４は、感性レベル分析部３８が出力した感性レベルに従って、出力装置５を制御して、入力された音声の感性レベルに対応する出力を生成する。ここで音声レベルの決定に用いられている情報は、１秒当たり何モーラ話しているかで表した話速や、平均基本周波数や、発話、文またはフレーズといった単位で求められる韻律情報などである。

【０００８】

また、怒りや苛立ちを認識する方式としては、特に音声の振幅に注目した技術として、２０～５０ミリ秒程度を分析フレームとして、隣接フレーム間で音量の差分として現れる子音と母音との振幅差を利用して、発話者の興奮を検出する方式（特許文献４参照）がある。

【０００９】

しかしながら、韻律情報は言語的情報を伝達するためにも使用されており、さらにその言語的情報の伝達方法が、言語の種類ごとに違うという特徴がある。たとえば日本語においては「橋」と「箸」のように基本周波数の高低によって作られるアクセントにより言葉の意味が異なる同音異義語が多くある。また、中国語においては、四声と呼ばれる基本周波数の動きにより、同じ音でもまったく異なる意味（文字）を示すことが知られている。英語ではアクセントは基本周波数よりもむしろストレスと呼ばれる音声の強度によって表現されるが、ストレスの位置は単語あるいは句の意味や、品詞を区別する手がかりとなっている。韻律による感情認識を行うためにはこのような言語による韻律パターンの差を考慮する必要があり、言語ごとに感情表現としての韻律の変化と、言語情報としての韻律の

変化とを分離して、感情認識用のデータを生成する必要があった。また、同一言語内においても、韻律を用いる感情認識においては、早口の人や、声の高い（低い）人、等の個人差があり、例えば、普段から大声で早口で話す声の高い人は常に怒っていると認識されてしまうことになる。そのため、個人ごとの標準データを記憶し、個人ごとに標準データと比較することで各個人に合わせた感情認識を行い、個人差による感情の認識間違いを防ぐという方法も必要であった（例えば、特許文献３参照）。

【００１０】

一方、声質による感情の認識については、「息漏れ」音を検出して発話者の聴取者に対する親近感や気を使っている態度を認識する方式（特許文献５参照）や、テンションの低さや、リラックスしている状態、あるいは苦しみを表現し、日本語ではフレーズ境界の基本周波数が低めの部分に見られる「りきみ（Ｖｏｃａｌ　Ｆｒｙ）」の検出を行う方法（非特許文献１参照）がある。また、音声認識に用いる音響特徴量を各感情の表出確率と対応付けて、分析フレームごとの感情表出確率より入力音声の一定区間に対して感情表出尤度を算出して、感情を認識する方式（特許文献６参照）が提案されている。しかし、スペクトル情報をあらわす音響特徴量については、音韻による差、言語差、地方差、個人差等の感情以外による分散が大きく、感情表出確率つきのコードブックを生成するには感情表現がラベルされた膨大なデータが必要である。このように、韻律以外の指標によって発話者の怒りや苛立ちを簡易に認識する方法については提案されていない。

【特許文献１】特開平９－２２２９６号公報（第６－９頁、表１－５、図２）
【特許文献２】特開２００３－９９０８４号公報
【特許文献３】特開２００５－２８３６４７号公報
【特許文献４】特開２００４－３１７８２２号公報
【特許文献５】特開２００６－８４６１９号公報
【特許文献６】特開２００５－３４５４９６号公報
【非特許文献１】石井カルロス寿憲、他著、「Ｖｏｃａｌ　Ｆｒｙ発声区間の自動検出法」、電子情報通信学会論文誌Ｄ、Ｊ８９－Ｄ巻１２号２６７９頁－２６８７頁、２００６

【発明の開示】

【発明が解決しようとする課題】

【００１１】

前述のように、韻律による感情認識では、韻律情報のうち言語情報を表すために使われている変動と感情表現としての変動とを分離するために、言語ごとに大量の音声データ、分析処理および統計処理が必要となる。さらに、同一言語であっても、地方差や年齢等による個人差も大きく、同一話者による音声であったとしても体調等により大きく変動する。このため、ユーザ個人ごとに標準データを持たない場合には、韻律による感情認識では、不特定多数の音声に対して常に安定した結果を生成することが困難であった。

【００１２】

さらに、不特定多数の使用を想定するコールセンターや駅などの公共の場所での案内システム等の場合には、話者ごとのデータを用意することができない。このため、個人ごとに標準データを用意する方式が採用できない。また、韻律データは１秒あたりのモーラ数、平均値もしくはダイナミックレンジのような統計的代表値、または時間パターンなど、発話、文またはフレーズといった音声としてまとまった長さで分析する必要がある。このため、音声の特徴が短時間で変化する場合には追随が困難であり、これが原因で音声による感情認識を高い精度で行うことができないという課題を有していた。

【００１３】

また、韻律以外の指標によって発話者の怒りや苛立ちを検出する方法はこれまでに提案されておらず、言語や個人による差にかかわらず安定して怒りや苛立ちを検出する方法がないという課題を有していた。

【００１４】

本発明は、前記従来の課題を解決するもので、音韻単位という短い時間単位で感情を検出可能で、かつ個人差、言語差、地方差の影響を受けずに発話者の怒りや苛立ちを検出す

ることのできる力み音声検出装置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】
【００１５】
上記目的を達成するために、本発明に係る力み音声検出装置は、入力音声信号が話者が力んだ状態で発声した音声の信号であるか否かを判定する力み音声検出装置であって、入力音声信号の振幅包絡を抽出する振幅包絡抽出手段と、前記振幅包絡抽出手段によって抽出された振幅包絡の周期的変動を検出し、検出された周期的変動の周波数を求める変調周波数分析手段と、前記変調周波数分析手段によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に、前記入力音声信号は話者が力んだ状態で発声した音声の信号であると判定する力み判定手段とを備える。
【００１６】
後述するように、力み音声の入力音声信号の振幅包絡には周期的変動が見られる。このような特徴は、発話者、発話者の喋る言語または発話者の住む地方が異なっていても共通に見られる性質である。上記した力み音声検出装置の構成によると、入力音声信号中に振幅包絡の周期的変動が見られるか否かを判断することにより、入力音声が力み音声か否かを判定している。よって、個人差、言語差、地方差の影響を受けずに発話者の怒りや苛立ちを検出することができる。また、力み音声か否かの判定を音韻単位で行なうことにより、短い時間単位で入力音声に力みが生じているか否かの判定を行なうことができる。
【００１７】
好ましくは、上述の力み音声検出装置は、さらに、前記入力音声信号の振幅包絡の振幅変動度合いを示す振幅変調度を計算する振幅変調度計算手段を備え、前記力み判定手段は、前記振幅変調度計算手段によって求められた前記振幅変調度があらかじめ定められた値以上であり、かつ前記変調周波数分析手段によって求められた前記周期的変動の周波数があらかじめ定められた範囲内にある場合に入力音声が力み音声であると判定する。
【００１８】
後述するように、力み音声の入力音声信号の振幅包絡には振幅変動が生じる。このため、振幅変動の度合いである振幅変調度が大きい部分の入力音声信号は力みの音声信号であると判定することができる。
【００１９】
本発明の他の局面に係る音声認識装置は、入力音声信号に含まれる音韻の種類を特定する音声認識装置であって、上述の力み音声検出装置と、音韻の種類ごとに、発話者が力んだ状態で発声した音声の特徴量を記憶している力み音声特徴量データベースと、音韻の種類ごとに、発話者が通常状態で発声した音声の特徴量を記憶している標準音声特徴量データベースと、前記力み音声検出装置において入力音声信号が力んだ状態で発声した音声の信号であると判定された場合には、前記力み音声特徴量データベースを用いて前記入力音声信号に含まれる音韻の種類を特定し、前記力み音声検出装置において前記入力音声信号が力んだ状態で発声した音声の信号であると判定されなかった場合には、前記標準音声特徴量データベースを用いて前記入力音声信号に含まれる音韻の種類を特定する音声認識手段とを備える。
【００２０】
この構成によると、力み音声の発生位置において、標準的な音声の特徴量が適合しないことによる音声認識精度の低下を防ぐことができる。これにより、正確に音声認識を行なうことができる。
【００２１】
本発明のさらに他の局面に係る音声認識装置は、入力音声信号に含まれる音韻の種類を特定する音声認識装置であって、上述の力み音声検出装置と、音韻の種類ごとに音響特徴量を記憶している音響特徴量データベースと、少なくとも読みまたは発音記号を有する単語辞書を表す言語特徴量を含む言語特徴量データベースと、前記力み音声検出装置において入力音声信号が力んだ状態で発声した音声の信号であると判定された場合には、前記音響特徴量データベースに含まれる音響特徴量を用いた確率モデルの重みよりも前記言語特

徴量データベースに含まれる言語特徴量を用いた確率モデルの重みを大きくし、重み付けされた２つの確率モデルを用いて前記入力音声信号に含まれる音韻の種類を特定する音声認識手段とを備える。

【００２２】

この構成によると、力み音声の発生位置において、音響特徴量を用いた確率モデルが適合しないことによる音声認識精度の低下を防ぐことができる。これにより、正確に音声認識を行なうことができる。

【００２３】

本発明のさらに他の局面に係る怒り認識機能付音声認識装置は、入力音声信号に含まれる音韻の種類を特定するとともに話者の怒りの強度を認識する怒り認識機能付音声認識装置であって、上述の音声認識装置と、音韻の属性情報から発話時の力みやすさを求めるための規則を用いて、前記音声認識装置で音韻の種類が認識された音韻ごとに、発話時の力みやすさを示す力み音声発声指標を計算する力み音声発生指標計算手段と、前記音声認識装置が備える力み検出装置により話者が力んだ状態で発声した音声の信号であると判定された入力音声信号について、前記力み音声発生指標が小さいほど怒りの強度が高くなる規則に基づいて、前記力み音声発生指標計算手段で計算された前記力み音声発生指標から怒りの強度を決定する怒り強度決定手段とを備える。

【００２４】

この構成によると、力み音声が発生しにくい部分で入力音声に力み音声が発生している場合には怒りの強度が高いと判断することができる。これにより、入力音声信号から、発話者の怒りや苛立ちの強度を個人差や地方差に影響されることなく正確に認識することができる。

【００２５】

なお、本発明は、このような特徴的な手段を備える力み音声検出装置として実現することができるだけでなく、力み音声検出装置に含まれる特徴的な手段をステップとする力み音声検出方法として実現したり、力み音声検出方法に含まれる特徴的なステップをコンピュータに実行させるプログラムとして実現したりすることもできる。そして、そのようなプログラムは、ＣＤ－ＲＯＭ（Ｃｏｍｐａｃｔ Ｄｉｓｃ－Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）等の記録媒体やインターネット等の通信ネットワークを介して流通させることができるのは言うまでもない。

【発明の効果】

【００２６】

本発明によれば、話者が怒りや苛立ちのために思わず喉頭を力んでしまい、生理的に発生する特徴的な声質である、力み音声を検出することができる。また、力み音声を検出することにより、言語の種類や話者の特性による個人差や地方差に影響されることなく、話者の怒りや苛立ちを音韻単位で認識することができる。このため、対話システム等において話者の怒りや苛立ちを緩和するような応対動作を行ったり、音声認識部の動作を変更したりする対応を取ることができる。よって、音声対話インタフェースにおいて誤認識がさらに誤認識を呼ぶという悪循環を断ち、快適で使い易い音声対話インタフェースを実現することができる。

【発明を実施するための最良の形態】

【００２７】

まず、本願が対象とする「力み」音声について、説明する。本願では、怒鳴ったり、興奮したりする際の発声中に生じる「力み」音声を対象としているが、本願が対象とする「力み」音声とは異なる定義がされた「りきみ」音声として、「きしる声」（creaky）あるいは「フライ」（vocal fry）とも呼ばれる音声の研究がなされている（石井カルロス寿憲、石黒浩および萩田紀博、「りきみの自動検出のための音響分析」電子情報通信学会技術研究報告，ＳＰ２００６－０７巻、ｐｐ．１－６，２００６）。

【００２８】

そこで、本願の「力み」音声について定義する。本願の「力み」音声とは、力を入れて

発話する際に、通常より発声器官に力が入るあるいは発声器官が強く緊張するために起こるもので、発声器官が「力み」の音声を生成しやすい状況が作られる場合に発声される音声であると定義する。具体的には、「力み」の音声は、力が入った発声であるため、音声の振幅はどちらかといえば大きく、当該モーラが両唇音や歯茎音でかつ鼻音や有声破裂音であり、文末や句末というよりアクセント句の先頭から３番目の間に位置するモーラである、といった実際の音声中の一部で起こる状況で発声され易い声質の音声である。また、「力み」の音声は感動詞や感嘆詞に限らず、自立語と付属語との違いを問わず様々な品詞中に見られる。

【００２９】

次に、本発明の基礎となる、音声中の力み音声の特徴について述べる。

感情や表情を伴った音声においては、様々な声質の音声が混在し、音声の感情や表情を特徴付け、音声の印象を形作っていることが知られている（例えば日本音響学会誌５１巻１１号（１９９５），ｐｐ８６９－８７５，粕谷英樹・楊長盛，“音源から見た声質”、特開２００４－２７９４３６号公報参照。）。本願発明に先立って、同一テキストに基づいて発話された５０文について無表情な音声と感情を伴う音声との調査を行った。

【００３０】

図１は、録音された音声における力み音声の感情種類による発生頻度を示すグラフである。図１は、４名の話者について「平静」、「怒り」、「激怒」、「朗らか」、「明るく元気」の５種類の感情表現を伴った音声中で、力み音声（ｈａｒｓｈ　ｖｏｉｃｅ）で発声されているとしてラベルされた、モーラ数を示したものである。力み音声は「激怒」、「怒り」の感情を伴った音声に多く出現し、「平静」、「朗らか」のような穏やかな音声では出現頻度が低い。「平静」、「朗らか」のような穏やかな音声は、音声認識における音響モデルを生成する際に使用される話し方であり、このような話し方の音声に対しては音声認識の認識率が高く、誤認識が起こりにくい。力み音声を検出することにより怒りや苛立ちのような声を荒げた状況、すなわち誤認識が起こりやすい状況を検知することができる。

【００３１】

「激怒」および「怒り」の感情を伴った音声における力み音声の波形の調査により、力み音声の波形の多くに振幅の周期的変動が見られることが明らかになった。図２（ａ）は、「特売（とくばい）してますよ」の「ばい」部分について、感情を伴わず「平静」に発声した音声より切り出した通常発声の音声波形とその振幅包絡の概形を示した図である。図２（ｂ）は、「激怒」の感情を伴って発声された、同じく「特売してますよ」中の「ばい」部分の波形とその振幅包絡の概形を示した図である。両波形とも、音素の境界を破線で示している。図２（ａ）の波形の／ａ／、／ｉ／を発声している部分では、振幅が滑らかに変動していく様子が見える。通常の発声においては、図２（ａ）の波形のように母音の立ち上がりで滑らかに振幅が大きくなり、音素の中央付近で最大値となり、音素境界に向けて振幅が小さくなる。母音の立下りがある場合には滑らかに無音あるいは後続子音に向けて振幅が小さくなる。図２（ａ）のように母音が続く場合は、緩やかに後続の母音に向けて振幅が小さくあるいは大きくなる。通常発声においては、１つの母音内において、図２（ｂ）のように振幅の増減を繰り返すことはほとんどなく、このような基本周波数との関係が一見してわからない振幅の変動を持つ音声についての報告はない。そこで振幅変動が力み音声の特徴であると考え、力み音声であるとラベルされた音声について、以下の処理によって振幅包絡の変動周期を求めた。

【００３２】

まず、音声波形を代表する正弦波成分を抽出するため、音声波形の振幅包絡曲線を求める。つまり、対象となる音声波形の基本周波数の第２高調波を中心周波数とするバンドパスフィルタを逐次求め、そのフィルタに音声波形を通過させる。フィルタを通過した波形に対してヒルベルト変換を施して解析信号を求め、その絶対値によってヒルベルト包絡曲線を求める。求められた振幅包絡曲線をさらにヒルベルト変換し、瞬時角速度をサンプル点ごとに計算し、サンプリング周期に基づいて角速度を周波数に変換する。サンプル点ご

とに求められた瞬時周波数について音韻ごとにヒストグラムを作成し、最頻値をその音韻の音声波形の振幅包絡の変動周波数とみなした。

【００３３】

図３は、このような方法で求められた「力み」音声の音韻ごとの振幅包絡の変動周波数を、音韻ごとの平均基本周波数に対してプロットした図である。男性話者、女性話者共に基本周波数に関わらず、振幅包絡の変動周波数は８０Ｈｚ－９０Ｈｚを中心として、５０Ｈｚ－１１０Ｈｚに分布している。力み音声の特徴の１つとして、５０Ｈｚ－１１０Ｈｚの周波数帯域に振幅の周期変動があることが発見された。このような周期変動は怒りや苛立ちによって発声器官に力が入った場合の生理的な反応であると考えられ、言語や個人による差は小さいことが期待される。そのため、音声中の５０Ｈｚ－１１０Ｈｚの振幅の周期変動を検出することで、言語差や個人差の影響を受けずに怒りや苛立ちを検出することができるはずである。

【００３４】

図４は、男性話者による「激怒」の感情を伴う発声を対象として、分析した力み音声の振幅包絡の変動周波数の分布をヒストグラムと累積度数とで示したものである。表１は、図４に示した力み音声の振幅包絡の変動周波数の頻度および累積度数を示す表である。

【００３５】

【表１】

データ区間	頻度	累積％
0	0	0.00%
10	1	0.18%
20	6	1.29%
30	11	3.33%
40	17	6.47%
50	27	11.46%
60	45	19.78%
70	41	27.36%
80	60	38.45%
90	73	51.94%
100	76	65.99%
110	77	80.22%
120	43	88.17%
130	31	93.90%
140	11	95.93%
150	11	97.97%
160	4	98.71%
170	2	99.08%
180	0	99.08%
190	2	99.45%
200	3	100.00%
次の級	0	100.00%

【００３６】

力み音声でない通常の音声では、その振幅包絡に周期的変動が無い。このため、力み音声を検出するためには周期的変動がない状態と変動がある状態とを区別する必要がある。

【００３７】

図４のヒストグラムにおいて、力み音声の頻度は振幅変動の周波数が１０Ｈｚから２０Ｈｚの間で立ち上がり、４０Ｈｚから５０Ｈｚの範囲で急激に増加している。周波数の下限は４０Ｈｚ付近が妥当と考えられるが、より広い範囲で網羅的に力み音声を検出する際には下限を１０Ｈｚとしても良い。累積度数より力みとラベルされた音韻のうち９０％は４７．１Ｈｚ以上の周波数で振幅が変動している。これらより、周波数の下限として４７．１Ｈｚを用いることができる。振幅変動の周波数が高くなりすぎると人間の聴覚は振幅の変動を捉えることができなくなる特性を考えると、振幅変動によって力み音声を検出するためには上限を設けるのが望ましい。聴覚の特性としては、７０Ｈｚ付近が「粗さ」を最もよく感じる周波数であり、変調を受ける元の音にもよるが、１００Ｈｚから２００Ｈ

ｚにかけて「粗さ」の感覚は小さくなる。

【００３８】

図４のヒストグラムにおいて、力み音声の頻度は１１０Ｈｚから１２０Ｈｚの範囲で急激に減少しており、さらに１３０Ｈｚから１４０Ｈｚの範囲で半減している。力み音声を特徴付ける振幅変動の周波数の上限は１３０Ｈｚ付近に設定されるべきである。さらに下限同様により広い範囲で網羅的に力み音声を検出する際には、図４において１７０Ｈｚから１８０Ｈｚの範囲で一端頻度が０まで低下することに基づいて、上限を１７０Ｈｚとしても良い。４７．１Ｈｚの下限とあわせて累積度数より力みとラベルされた音韻のうち８０％が含まれることになる１２３．２Ｈｚを上限として用いることは有効である。

【００３９】

図５は、力み音声の振幅包絡の変調度を説明するための図である。振幅変動の変調度（振幅変調度）については、変調される信号である音声波形にもともと振幅の変化があるため、振幅一定のキャリア信号の振幅を変調するいわゆる振幅変調とは異なる。したがって、振幅変動の変調度を以下のように定義した。図５（ａ）に示すように、第２高調波を中心周波数とするバンドパスフィルタを通過した波形のヒルベルト包絡曲線として求められた振幅包絡曲線を多項式近似し、多項式によるフィッティング関数を作成する。図５（ａ）は、振幅包絡曲線に５次関数をフィッティングさせることにより、フィッティング関数を求めている。フィッティング関数を変調前の波形の振幅包絡曲線とみなす。図５（ｂ）に示すように、第２高調波成分の振幅包絡曲線のピークごとに、当該振幅包絡曲線の値とフィッティング関数の値との差分を求め、それを振幅変動量とみなす。振幅変動量とフィッティング関数の値とは、共に一定の値ではないため、振幅変動量とフィッティング関数の値との両者について音韻内での中央値を求め、両中央値の比を変調度とする。

【００４０】

図６は、このようにして求めた変調度のヒストグラムと累積度数とを示す図である。表２は、図６に示した変調度の頻度および累積度数を示す表である。

【００４１】

【表２】

データ区間	頻度	累積％
0	0	0.00%
0.02	7	1.29%
0.04	52	10.91%
0.06	60	22.00%
0.08	75	35.86%
0.1	62	47.32%
0.12	42	55.08%
0.14	32	61.00%
0.16	35	67.47%
0.18	32	73.38%
0.2	38	80.41%
0.22	16	83.36%
0.24	22	87.43%
0.26	9	89.09%
0.28	6	90.20%
0.3	14	92.79%
0.32	8	94.27%
0.34	4	95.01%
0.36	2	95.38%
0.38	4	96.12%
0.4	2	96.49%
0.42	6	97.60%
0.44	2	97.97%
0.46	4	98.71%
0.48	3	99.26%
0.5	1	99.45%
0.52	1	99.63%
0.54	0	99.63%
0.56	0	99.63%
0.58	0	99.63%
0.6	1	99.82%
0.62	0	99.82%
0.64	0	99.82%
0.66	0	99.82%
0.68	0	99.82%
0.7	0	99.82%
0.72	0	99.82%
0.74	0	99.82%
0.76	0	99.82%
0.78	0	99.82%
0.8	0	99.82%
0.82	0	99.82%
0.84	0	99.82%
0.86	0	99.82%
0.88	1	100.00%
0.9	0	100.00%
0.92	0	100.00%
0.94	0	100.00%
0.96	0	100.00%
0.98	0	100.00%
1	0	100.00%
次の級	0	100.00%

【００４２】

図６に示すヒストグラムは、男性話者による「激怒」の感情を伴う発声中（力み音声の発声中）に見られた振幅変動の変調度の分布を示している。聴取者が振幅変動を知覚する

ためには変動の大きさすなわち変調度が一定以上である必要がある。図６のヒストグラムにおいて、振幅変動の変調度の頻度が０．０２から０．０４の範囲で急激に頻度が高くなっており、力み音声を特徴付ける振幅変動の変調度の下限を０．０２付近とすることが妥当である。また、累積度数を見ると、９０％の音韻は変調度が０．０３８以上であるため、変調度の下限として０．０３８を用いることもできる。

【００４３】

図５に示した振幅変調の定義から変調度は最大１であることが分かる。図７は変調度１の場合の変換元波形と変換結果とを模式的に示したものである。変換元波形として、例えば、振幅が一定の波形を仮定する。図７（ａ）は、振幅が１である極短い周期の正弦波を模式的に示している。図７より、変調度が１を超えた場合は変換元の波形を反転させることとなり、振幅を変動させる処理から逸脱する。従って、振幅変動の変調度の定義より、振幅変動の変調度は最大１である。

【００４４】

図６に示すヒストグラムでは、さらに、０．０３８の下限とあわせて、力みとラベルされた音韻のうち８０％が含まれることになる、０．２７６を振幅変動の変調度の上限として用いることも有効である。上記のことから、力み音声を検出するための１つの基準として、振幅包絡の周期変動が４０Ｈｚ－１２０Ｈｚ、変調度が０．０４以上という基準を用いることができる。

【００４５】

上記のように力み音声は言語差や個人差とかかわりのない怒りや苛立ちによる生理的反応として検出可能なものであり、力み音声の検出により話者の怒りや苛立ちを検知できる可能性を示している。

【００４６】

以下本発明の実施の形態について、図面を参照しながら説明する。

【００４７】

（実施の形態１）

図８は本発明の実施の形態１における力み音声検出装置の構成を示す機能ブロック図である。図９は実施の形態１における力み音声検出装置の動作を示したフローチャートである。

【００４８】

図８に示されるように、力み音声検出装置は、入力音声より力み音声を検出する装置であり、マイクロホン１と、基本周波数分析部１０１と、力み音声検出部１００とを備えている。マイクロホン１は、入力音声を電気信号に変換する装置である。基本周波数分析部１０１は、入力音声の周期性を分析し、入力音声中の有声区間を抽出すると共に、有声区間の基本周波数を求める処理部である。力み音声検出部１００は、入力音声中の有声区間から力み音声の区間を検出する処理部である。表示部１０７は力み判定部１０６が判定した内容を表示する表示装置であり、例えば、ＬＣＤ（Liquid Crystal Display）などである。

【００４９】

力み音声検出部１００の構成についてより詳細に説明すると、力み音声検出部１００は、第２高調波抽出部１０２と、振幅包絡分析部１０３と、振幅変調度計算部１０４と、変調周波数分析部１０５と、力み判定部１０６とを備えている。

【００５０】

第２高調波抽出部１０２は、基本周波数分析部１０１より出力された入力音声の有声区間について、同じく基本周波数分析部１０１より出力された有声区間の基本周波数に基づいて、有声区間の第２高調波成分を抽出する処理部である。振幅包絡分析部１０３は、第２高調波抽出部１０２より出力された有声区間の第２高調波成分波形を受付け、その振幅包絡曲線を求める処理部である。

【００５１】

振幅変調度計算部１０４は、振幅包絡分析部１０３より出力された第２高調波成分の振

幅包絡曲線を受付け、第２高調波成分の振幅変調度を求める処理部である。変調周波数分析部１０５は振幅包絡分析部１０３より出力された第２高調波成分の振幅包絡曲線を受付け、包絡線の変動周波数すなわち第２高調波の振幅変調周波数を求める処理部である。力み判定部１０６は、振幅変調度計算部１０４より出力される第２高調波の振幅変調度と変調周波数分析部１０５より出力される第２高調波の振幅変調周波数とより、入力音声の有声区間が力み音声の区間であるか否かを判定する処理部である。

【００５２】

このように構成された力み音声検出装置の動作を図９に従って説明する。

マイクロホン１より音声が入力される（ステップＳ１００１）。基本周波数分析部１０１は、入力音声の周期性の有無を分析すると同時に、周期性のある部分についてはその周波数を求める（ステップＳ１００２）。周期性および周波数の分析は、例えば以下のようにして行われる。つまり、入力音声の自己相関係数を求め、５０Ｈｚから５００Ｈｚに相当する周期で相関係数が一定以上になる部分を周期性のある部分すなわち有声区間とみなし、相関係数が最大となる周期に対応する周波数を基本周波数とする。さらに基本周波数分析部１０１は、ステップＳ１００２において音声中の有声区間とみなされた区間を抽出する（ステップＳ１００３）。音声中に有声区間がない場合には（ステップＳ１００４でＮＯ）、表示部１０７は、音声中に力み音声がないことを表示する（ステップＳ１０１６）。

【００５３】

音声中に有声区間がある場合には（ステップＳ１００４でＹＥＳ）、第２高調波抽出部１０２は、音声中の未処理の有声区間のうち１区間についてステップＳ１００２で求められた当該有声区間の基本周波数の２倍の周波数を中心としたバンドパスフィルタを設定し、当該有声区間の音声波形をフィルタリングして第２高調波成分を抽出する（ステップＳ１００５）。

【００５４】

振幅包絡分析部１０３は、ステップＳ１００５で抽出された第２高調波成分の振幅包絡を抽出する（ステップＳ１００６）。振幅包絡は全波整流を行ってそのピーク値をスムーシングして求める方法や、ヒルベルト変換を行ってその絶対値を求める方法等を用いて抽出される。

【００５５】

振幅変調度計算部１０４は、ステップＳ１００６で抽出した振幅包絡を多項式で近似し、振幅変調前の音声の包絡線を推定する（ステップＳ１００７）。なお、実験的には多項式として３次式または５次式が望ましい。

【００５６】

振幅変調度計算部１０４は、ステップＳ１００６で求めた振幅包絡のピークごとに振幅包絡の値とステップＳ１００７で求めた多項式による近似値との差分を求め（ステップＳ１００８）、当該分析区間中の全ピークでの差分の中央値と当該分析区間内での近似式の値の中央値との比より変調度を求める（ステップＳ１００９）。変調度は振幅包絡の凸のピーク値の平均または中央値と、凹のピーク値の平均または中央値との比など、他の定義をすることも可能であるが、その際には変調度の基準値はその定義に基づいて設定される必要がある。

【００５７】

力み判定部１０６は、ステップＳ１００９で求められた変調度があらかじめ定められた基準値、例えば０．０４を超えているか否かを判断する（ステップＳ１０１０）。図６のヒストグラムに示すとおり、力み音声の頻度は、変調度が０．０２から０．０４の間で急激に増加していることから、基準値をここでは０．０４とする。変調度が基準値を超えていない場合（ステップＳ１０１０でＮＯ）、力み判定部１０６は、当該有声区間を力み音声でない、すなわち通常音声の区間と判断し（ステップＳ１０１１）、ステップＳ１００４に戻り、次の有声区間の処理を行う。

【００５８】

変調度が基準値を超えている場合（ステップＳ１０１０でＹＥＳ）、変調周波数分析部１０５は、ステップＳ１００６で抽出された振幅包絡について分析フレームごとに瞬時周波数を求める。変調周波数分析部１０５は、さらに、当該有声区間で求められた瞬時周波数の中央値を求め、これを変調周波数とする（ステップＳ１０１２）。

【００５９】

力み判定部１０６は、ステップＳ１０１２で求められた変調周波数が、あらかじめ定められた基準範囲内（例えば４０Ｈｚ以上１２０Ｈｚ未満）にあるか否かを判断する（ステップＳ１０１３）。図４のヒストグラムに示すとおり、力み音声の頻度が３０Ｈｚから４０Ｈｚの間で急激に増加し、７０Ｈｚから８０Ｈｚをピークに１１０Ｈｚから１２０Ｈｚで急激に減少していることから、基準範囲をここでは４０Ｈｚ以上１２０Ｈｚ未満とした。変調周波数が基準範囲外である場合（ステップＳ１０１３でＮＯ）、力み判定部１０６は、当該有声区間を力み音声でない、すなわち通常音声の区間と判断し（ステップＳ１０１４）、ステップＳ１００４に戻り、次の有声区間の処理を行う。

【００６０】

変調周波数が基準範囲内である場合（ステップＳ１０１３でＹＥＳ）、力み判定部１０６は、当該有声区間を力み音声の区間であると判定する（ステップＳ１０１５）。その後ステップＳ１００４に戻り、次の有声区間の処理を行う。ステップＳ１００４からステップＳ１０１５の動作を繰り返し、入力音声中のすべての有声区間の処理を行った後、表示部１０７は力み音声区間を表示する（ステップＳ１０１６）。

【００６１】

かかる構成によれば、入力された音声の振幅包絡の周期的変動の有無を分析することにより、入力音声中の力み音声を抽出することができる。例えば、コールセンターにおいて利用者の電話音声から力み音声を抽出することで、利用者の怒りや苛立ちをモニターして適切な対応を取ることができる。

【００６２】

なお、本実施の形態では、ステップＳ１００５において第２高調波抽出部１０２は基本周波数の２倍の周波数を中心周波数とするバンドパスフィルタにより第２高調波を抽出するものとしたが、基本周波数を中心周波数とするバンドパスフィルタあるいは基本周波数を含むローパスフィルタによって基本波を抽出するものとしても良い。また、解析信号を求めてヒルベルト包絡を計算するために、音声信号から特定の高調波を抽出することにより正弦波の信号を取り出すものであれば何でも良く、第３高調波以上が安定して取り出せるのであればそれを用いてももちろん構わない。

【００６３】

なお、実施の形態１において、ステップＳ１０１０で変調度の基準値を０．０４としたが、０．０２以上の値であればこれ以外の値としても良い。

【００６４】

また、実施の形態１において、ステップＳ１０１３で変調周波数の基準範囲を４０Ｈｚ以上１２０Ｈｚ未満としたが、１０Ｈｚ以上１７０Ｈｚ未満の範囲であればこれ以外の範囲としても良い。

【００６５】

さらに、本実施の形態では、変調度および変調周波数を用いて力み音声の検出を行なったが、変調周波数のみを用いて力み音声の検出を行うものであってもよい。この場合には、図９のステップＳ１００９～Ｓ１０１１の処理が省略される。ただし、変調度および変動周波数を用いて力み音声の検出を行なう方が、ノイズを拾わずに、より安定的に力み音声の検出を行なうことができる。

【００６６】

（実施の形態２）

実施の形態２では、本発明に係る力み音声検出装置を音声認識装置に応用した例について説明する。

【００６７】

図１０は、本発明の実施の形態２における力み音声検出部を含む音声認識装置の構成を示す機能ブロック図である。図１１は、図１０に示した音声認識装置のうち力み音声検出部の構成を示す機能ブロック図である。図１２は、実施の形態２における力み音声検出部を含む音声認識装置の動作を示したフローチャートである。図１３は、実施の形態２における音声認識装置の動作のうち力み音声検出部の動作を示したフローチャートである。

【００６８】

図１１において、図８と同じ部分については説明を適宜省略し、図８と異なる部分を中心に説明する。図１３においても、図９と同じ部分については説明を適宜省略し、図９と異なる部分を中心に説明する。

【００６９】

図１０を参照して、音声認識装置の構成について説明する。音声認識装置は、入力音声を認識し、認識結果を出力する装置であり、マイクロホン１と、音声認識用特徴量抽出部２０１と、逆フィルタ２０２と、周期性分析部２０３と、スイッチ２０４と、特徴量データベース２０５と、音声認識部２０８と、表示部２０９とを備えている。

【００７０】

マイクロホン１は、入力音声を電気信号に変換する装置である。音声認識用特徴量抽出部２０１は、入力音声を分析し、スペクトル包絡を表すパラメータ、例えばメルケプストラム係数を抽出する処理部である。逆フィルタ２０２は、音声認識用特徴量抽出部２０１が出力するスペクトル包絡情報の逆フィルタであり、マイクロホン１より入力された音声の音源波形を出力する処理部である。

【００７１】

周期性分析部２０３は、逆フィルタより出力された音源波形の周期性を分析して有声区間を抽出する処理部である。力み音声検出部２００は、周期性分析部２０３より出力された音源波形の有声区間より力み音声の区間を検出する処理部である。

【００７２】

特徴量データベース２０５は、音韻種類ごとの音声特徴量を音声認識用に保持する記憶装置である。特徴量データベース２０５は、標準音声特徴量データベース２０６と、力み音声特徴量データベース２０７とを含む。標準音声特徴量データベース２０６は、力み音声が見られない音声データより作成された音声特徴量を保持する記憶装置であり。力み音声特徴量データベース２０７は、力み音声が見られる音声データより作成された音声特徴量を保持する音声特徴量データベース２０７である。

【００７３】

音声認識部２０８は、特徴量データベース２０５を参照し、音声認識用特徴量抽出部２０１が出力した特徴量と特徴量データベース２０５に格納された特徴量とのマッチングを行って音声認識を行う処理部である。

【００７４】

スイッチ２０４は、力み音声検出部２００で入力音声中に力み音声が検出されたか否かによって、標準音声特徴量データベース２０６および力み音声特徴量データベース２０７のうちのいずれかのデータベースを、音声認識部２０８に接続する処理部である。音声認識部２０８は、接続されたデータベースを用いて、音声認識を行なうことになる。表示部２０９は、音声認識部２０８での音声認識結果を表示する表示装置であり、例えば、ＬＣＤなどである。

【００７５】

図１１を参照して、力み音声検出部２００の構成について説明する。力み音声検出部２００は、第２高調波抽出部１０２が基本波抽出部２１２に置き換わった以外は、図８に示した実施の形態１の力み音声検出部１００と同様である。基本波抽出部２１２は、周期性分析部２０３が出力する音源波形の有声区間とその基本周波数の情報を受付け、音源波形から基本波成分を抽出する処理部である。力み音声検出部２００のそれ以外の要素は実施の形態１の力み音声検出部１００と同様であるので、説明を省略する。なお、基本波抽出部２１２の代わりに第２高調波抽出部１０２を用いてもよい。

【００７６】
このように構成された力み音声検出部を含む音声認識装置の動作について図１２に従って説明する。
【００７７】
マイクロホン１より音声が入力される（ステップＳ１００１）。音声認識用特徴量抽出部２０１は、入力音声を分析し、音声認識用の音響特徴量としてメルケプストラム係数を抽出する（ステップＳ２００２）。逆フィルタ２０２は、ステップＳ２００２で生成されたメルケプストラム係数の逆フィルタとなるようにパラメータを設定し、ステップＳ１００１でマイクロホンより入力された音声信号を通過させ、音源波形を抽出する（ステップＳ２００３）。周期性分析部２０３は、ステップＳ２００３で抽出された音源波形の中から周期性のある区間を抽出する（ステップＳ２００４）。例えば、周期性分析部２０３は、特開平１０－１９７５７５号公報に開示されている方法を用いて、周期性のある区間を抽出する。つまり、周期性分析部２０３は、音源波形を入力とし、低周波側が緩やかで高周波側が急峻な遮断特性を有するフィルタ出力の振幅変調の大きさと周波数変調の大きさとから基本波らしさを計算し、入力音声の音源波形のうち周期性のある信号の時間領域を周期性信号区間すなわち有声区間として出力する（ステップＳ２００４）。
【００７８】
力み音声検出部２００は、ステップＳ２００４で周期性分析部２０３により抽出された有声区間について、基本波成分の振幅包絡の周期的変動を検出することにより、力み音声の区間を検出する（ステップＳ２００５）。スイッチ２０４は、入力音声の有声区間において力み音声が検出されたか否かにより、特徴量データベース２０５内の標準音声特徴量データベース２０６および力み音声特徴量データベース２０７のいずれかと音声認識部２０８とを接続する（ステップＳ２００６）。つまり、スイッチ２０４は、ステップＳ２００５において力み音声が検出された場合には、力み音声特徴量データベース２０７と音声認識部２０８とを接続する。また、スイッチ２０４は、ステップＳ２００５において力み音声が検出されなかった場合には、標準音声特徴量データベース２０６と音声認識部２０８とを接続する。
【００７９】
音声認識部２０８は、特徴量データベース２０５のうちステップＳ２００６においてスイッチ２０４によって接続された特徴量データベースを参照し、ステップＳ２００２で抽出されたメルケプストラム係数を用いて音声認識を行なう。また、音声認識部２０８は、認識結果として入力音声中の時間位置情報と共に音韻列を出力する（ステップＳ２００７）。表示部２０９は、音声認識部２０８より出力された時間位置情報および音韻列を表示する（ステップＳ２００８）。
【００８０】
次に、力み音声抽出処理（ステップＳ２００５）の詳細を、図１３を参照して説明する。図１３については、図９と同じ動作については説明を適宜省略し、異なる部分を中心に説明する。
【００８１】
音声中に有声区間がない場合には（ステップＳ１００４でＮＯ）、力み判定部１０６は、力み音声は検出されなかったと判定し、スイッチ２０４が、標準音声特徴量データベース２０６と音声認識部２０８とを接続する（ステップＳ２００６）。
【００８２】
音声中に有声区間がある場合には（ステップＳ１００４でＹＥＳ）、基本波抽出部２１２は、音声中の未処理の有声区間のうちの１区間について、ステップＳ２００４で求められた当該有声区間の基本周波数の１．５倍の周波数をカットオフ周波数とするローパスフィルタを設定し、当該有声区間の音源波形をフィルタリングして基本波成分を抽出する（ステップＳ２１０５）。振幅包絡分析部１０３は、ステップＳ２１０５で抽出された基本波成分の振幅包絡を抽出する（ステップＳ２１０６）。基本波成分の振幅包絡の抽出方法は、ステップＳ１００６と同様である。

【００８３】
振幅変調度計算部１０４は、ステップＳ２１０６で抽出した振幅包絡を多項式で近似し、振幅変調前の音声の包絡線を推定する（ステップＳ２１０７）。この多項式も、実施の形態１と同様、実験的には３次式または５次式が望ましい。
【００８４】
振幅変調度計算部１０４は、ステップＳ２１０６で求めた振幅包絡のピークごとに振幅包絡の値とステップＳ２１０７で求めた多項式による近似値との差分を求め（ステップＳ１００８）、当該分析区間中の全ピークでの差分の中央値と当該分析区間内での近似式の値の中央値との比より変調度を求める（ステップＳ１００９）。
【００８５】
力み判定部１０６は、ステップＳ１００９で求められた変調度があらかじめ定められた基準値、例えば０．０４を超えているか否かを判断する（ステップＳ１０１０）。変調度の基準値については変調度の定義によって異なるが、ここではどちらも音声の低域のエネルギー変動を示すことになる基本波の振幅包絡の変動と第２高調波の振幅包絡の変動とに大きな差は無いとみなし、図６のヒストグラムより決定した０．０４以上という基準を採用する。
【００８６】
変調度が基準値を超えていない場合（ステップＳ１０１０でＮＯ）、力み判定部１０６は、当該有声区間を力み音声でない、すなわち通常音声の区間と判断し（ステップＳ１０１１）、ステップＳ１００４に戻り、次の有声区間の処理を行う。
【００８７】
変調度が基準値を超えている場合（ステップＳ１０１０でＹＥＳ）、変調周波数分析部１０５は、ステップＳ２１０６で抽出された振幅包絡について分析フレームごとに瞬時周波数を求める。変調周波数分析部１０５は、さらに、当該有声区間で求められた瞬時周波数の中央値を求め、これを変調周波数とする（ステップＳ１０１２）。
【００８８】
力み判定部１０６は、ステップＳ１０１２で求められた変調周波数が、あらかじめ定められた基準範囲内（例えば実施の形態１と同様に図４のヒストグラムより決定した４０Ｈｚ以上１２０Ｈｚ未満）であるか否かを判断する（ステップＳ１０１３）。広帯域の波形においても振幅変動が観察されることから、帯域が変わっても変調周波数は変わらないものとみなし、実施の形態１の図４に示した第２位高調波と同様の周波数範囲を採用する。
【００８９】
変調周波数が基準範囲外である場合（ステップＳ１０１３でＮＯ）、力み判定部１０６は当該有声区間を力み音声でない、すなわち通常音声の区間と判断し（ステップＳ１０１４）、ステップＳ１００４に戻り、次の有声区間の処理を行う。変調周波数が基準範囲内である場合（ステップＳ１０１３でＹＥＳ）、力み判定部１０６は当該有声区間を力み音声の区間と判定する（ステップＳ１０１５）、すなわち、入力音声中に力み音声を検出したものとし、力み音声の検出処理を終了する。続いて、スイッチ２０４が、力み音声特徴量データベース２０７と音声認識部２０８とを接続する（ステップＳ２００６）。
【００９０】
かかる構成によれば、入力された音声より力み音声を抽出し、力み音声の有無によって、力み音声を含む特徴量データベースと力み音声を含まない特徴量データベースとを切り替えて音声認識に利用することができる。このため、音声認識精度を向上させることができる。また、音声認識結果と力み音声の出現位置との対応がつくため、本実施の形態の出力を記録することで、ユーザが発話中のどの単語やフレーズに力を入れて話していたかを解析することができる。このような解析を、コールセンターの利用者の音声や、店頭での顧客の音声に適用することにより、発話中のどの単語やフレーズに力を入れていたかを知ることができ、クレーム内容をより的確に分類してマーケティングに反映させることができる。
【００９１】

また、本実施の形態に示すような力み音声検出装置を含む音声認識装置を対話制御等に用いる場合には、力み音声検出部２００の出力を利用することにより、ユーザである話者が対話動作過程のどのイベントに対して、語気を荒げたか、すなわち怒りや苛立ちを覚えたかを特定することができる。このように入力音声よりユーザの怒りや苛立ちを捉えることができ。このため、例えば、ユーザの怒りに対して、システム側の出力音声を「大変申し訳ございませんが・・・」という丁寧な謝罪や、「お手数ではございますが・・・」という丁寧な依頼の表現にしたりすることができる。これにより、ユーザの感情を平常な状態に導き、ユーザによる発話を誤認識の少ない通常音声での発話へ誘導し、対話インタフェースとしてスムーズに動作する環境を整えることができる。

【００９２】

なお、本実施の形態において音源波形はメルケプストラム係数の逆フィルタによって求めるものとしたが、声道モデルを元に声道伝達特性を求め、その逆フィルタによって音源波形を求める方法や、音源波形のモデルを元に音源波形を求める方法等、音源波形の求め方はメルケプストラム係数の逆フィルタによる方法以外の方法を用いても良い。

【００９３】

また、本実施の形態において、音声認識の音響特性モデルとしてメルケプストラム係数のパラメータを用いるものとしたが、それ以外のケプストラム係数など、音声の周波数特性を記述し、音声認識に用いられる特徴量であればどのような特徴量を用いて音声認識を行ってもよい。その際、音源波形はメルケプストラム係数の逆フィルタを用いて求めるものとしても、それ以外の方法で求めるものとしても良い。

【００９４】

さらに、本実施の形態においては、入力音声中に力み音声が１箇所検出された時点でスイッチ２０４を力み音声特徴量データベース２０７に接続するものとしたが、あらかじめ定められた数以上の箇所で力み音声が検出された場合、例えば入力音声中に２箇所（１発話２０モーラ程度として１０％）で力み音声が検出された場合に、スイッチ２０４を力み音声特徴量データベース２０７に接続するものとしてもよい。または、入力音声の一定時間あたりの力み音声の検出数があらかじめ定められた数以上となった場合、例えば１発話２０モーラ程度が３秒前後として、３秒あたりの力み音声の検出数が２つ以上になった場合に、スイッチ２０４を力み音声特徴量データベース２０７に接続するものとしてもよい。さらには、入力音声の時間長のうち、力み音声区間の占める割合があらかじめ定められた値以上であった場合に、スイッチ２０４を力み音声特徴量データベース２０７に接続するものとしても良い。

【００９５】

さらにまた、入力音声の一定時間を処理単位とし、処理単位ごとにスイッチ２０４を切り替える判断をしてもよい。また、入力音声の１フレーズごとにスイッチ２０４を切り替える判断をしてもよい。また、１発話ごとにスイッチ２０４を切り替える判断をしてもよい。また、あらかじめ定められた一定時間以上、例えば１００ｍｓ以上の無音区間によって区切られた発話単位ごとにスイッチ２０４を切り替える判断をしても良い。

【００９６】

（実施の形態３）

実施の形態３では、本発明に係る力み音声検出装置を音声認識装置に応用した例について説明する。

【００９７】

図１４は、本発明の実施の形態３における力み音声検出部を含む音声認識装置の構成を示す機能ブロック図である。図１５は、実施の形態３における力み音声検出部を含む音声認識装置の動作を示したフローチャートである。図１６は、実施の形態３における音声認識装置のうち力み音声検出部の動作の部分を示したフローチャートである。図１７は、実施の形態３の動作の具体例を示す図である。

【００９８】

図１４において、図８および図１０と同じ部分については説明を適宜省略し、図８およ

び図１０と異なる部分を中心に説明する。図１５においても図１２と同じ部分については説明を適宜省略し、図１２と異なる部分を中心に説明する。図１６においても図９および図１３と同じ部分については説明を適宜省略し、図９および図１３と異なる部分を中心に説明する。

【００９９】

図１４において、音声認識装置の構成は、図１０の機能ブロック図より逆フィルタ２０２、スイッチ２０４がなくなり、周期性分析部２０３が図８と同様の基本周波数分析部１０１に置き換わり、力み音声検出部２００が図８と同様の力み音声検出部１００に置き換わり、特徴量データベース２０５が音響特徴量データベース３０１と言語特徴量データベース３０２とに置き換わり、音声認識部２０８が連続単語音声認識部３０３に置き換わった以外は図１０と同様の構成である。

【０１００】

音響特徴量データベース３０１は、音韻の種類ごとに音響特徴量を記憶している記憶装置である。言語特徴量データベース３０２は、少なくとも読みまたは発音記号を有する単語辞書を表す言語特徴量を記憶している記憶装置である。連続単語音声認識部３０３は、音韻のみではなく、言語情報も含めて音声の認識を行なう処理部である。

【０１０１】

このように構成された力み音声検出装置を含む音声認識装置の動作について図１５および図１６に従って説明する。図９、図１２および図１３と同じ動作については説明を省略し、異なる部分についてのみ説明する。

【０１０２】

マイクロホン１より音声が入力される（ステップＳ１００１）。音声認識用特徴量抽出部２０１は、入力音声を分析し、メルケプストラム係数を抽出する（ステップＳ２００２）。一方、基本周波数分析部１０１は、実施の形態１と同様にして入力音声の周期性の有無を分析するとともに、周期性のある部分についてはその周波数を求める（ステップＳ１００２）。さらに、基本周波数分析部１０１は、ステップＳ１００２において音声中の有声区間とみなされた区間を抽出する（ステップＳ１００３）。

【０１０３】

力み音声検出部１００は、実施の形態１のステップＳ１００４からステップＳ１０１５で力み音声を検出した処理と同様の処理を行なうことにより、ステップＳ１００３で抽出された有声区間が力み音声の区間であるか否かを判断する（ステップＳ３００５、図１６）。

【０１０４】

連続単語音声認識部３０３は、音響特徴量データベース３０１と言語特徴量データベース３０２とを参照し、ステップＳ２００２で抽出されたメルケプストラム係数を用いて音声認識を行う（ステップＳ３００６～Ｓ３００７）。連続単語音声認識部３０３による音声認識は、例えば、音響モデルと言語モデルとからなる確率モデルを用いた音声認識方法によるものとする。音声認識は一般的に、数１に示す音響モデルと言語モデルの積が最も高くなる単語系列を選択することで行われる。

【０１０５】

【数１】

$$\hat{W} = \arg\max_{w} P(Y/W)P(W)$$

Ｗ：指定された単語系列
Ｙ：音響的な観測値系列
Ｐ（Ｙ／Ｗ）：単語列で条件付けられた音響的な観測値系列の確率（音響モデル）
Ｐ（Ｗ）：仮定された単語系列に対する確率（言語モデル）

【０１０６】

数１は対数を取ると数２のように表現できる。

【０１０７】
【数２】

$$\hat{W} = \arg\max_{w} \log P(Y/W) + \log P(W)$$

【０１０８】
音響モデルと言語モデルのバランスが等価であるとは限らないため、両モデルへの重みをつける必要がある。一般的には両重みの比として言語モデルの重みを設定することにより、数２を数３のように表現しなおす。
【０１０９】
【数３】

$$\hat{W} = \arg\max_{w} \log P(Y/W) + \alpha \log P(W)$$

α：音響モデルと言語モデルとの両モデルにおける言語モデルの重み
【０１１０】
言語モデルの重みαは、一般的な認識処理においては、時間的に一定の値を持つものとされる。しかし、連続単語音声認識部３０３は、ステップＳ３００５で検出された力み音声を含む有声区間の情報を取得し、単語ごとに言語モデル重みαを変更する。
【０１１１】
連続単語音声認識部３０３は、数４のように表現されるモデルに基づき連続音声認識を行う。
【０１１２】
【数４】

$$\hat{W} = \arg\max_{w} \log P(Y/W) + \sum_{i=1}^{n} \alpha_i \log P(w_i \mid w_1 \cdots w_{i-1})$$

w_i：ｉ番目の単語
α_i：ｉ番目の単語に適用する言語モデルの重み
【０１１３】
連続単語音声認識部３０３は、音響特徴量データベース３０１と言語特徴量データベース３０２とを参照して音声認識を行う際に、音声認識を行うフレームが力み音声を含む場合には言語モデルの重みαを大きくし、相対的に音響モデルの重みを小さくし（ステップＳ３００６）、音声認識を行う（ステップＳ３００７）。言語モデルの重みを大きくし、音響モデルの重みを小さくすることにより、力み音声により音響モデルが適合しないために認識精度が低下する影響を小さくすることができる。連続単語音声認識部３０３は、入力音声を音声認識した結果の単語列を出力し、表示部２０９は認識結果を表示する（ステップＳ２００８）。
【０１１４】
例えば、図１７（ａ）に示すように、入力音声の音韻列が「なまえおかくえんぴつがほしいんです」で、そのうち「えんぴつが」の部分が力み音声で発声されているものとする。この場合、連続単語音声認識部３０３は、ステップＳ３００５で検出された力み音声が検出された有声区間の情報を取得し、力み音声を含まない、「なまえおかく」と「ほしいんです」の部分については、力み音声ではない通常発声の学習用データより決定された言語モデルの重みα＝０．９を適用する。このとき、図１７（ｂ）に示すように従来の連続音声認識の方法すなわち言語モデルの重みαを一定として、力み音声で発声された部分についても力み音色で発声されていない場合に適用する言語モデルの重みα＝０．９を適用する。力み音声で発声された「えんぴつが」の部分が、通常発声の音響モデルにおいては「えんとつ」とのマッチングが良かったものとする。この場合、

【数５】

$$P(\text{えんとつ}|\cdots\text{書く}) < P(\text{えんぴつ}|\cdots\text{書く})$$

のように、言語モデルとしては、文頭から「書く」までの単語列に「えんとつ」が続く確率より「えんぴつ」が続く確率の方が大きいため

【数６】

$$P(W_1) < P(W_2)$$

W_1＝名前　を　書く　えんとつ　が　欲しい　ん　です

W_2＝名前　を　書く　えんぴつ　が　欲しい　ん　です

となるにもかかわらず、言語モデルの重みが小さいために相対的に音響モデルの値が大きく、数３の値は

【数７】

$$\log P(Y/W_1) + 0.9 \times \log P(W_1) > \log P(Y/W_2) + 0.9 \times \log P(W_2)$$

となる。このため、認識結果としては「名前を書く煙突が欲しいんです」が採用されることになる。

【０１１５】

しかし、本実施の形態では、連続単語音声認識部３０３は、ステップＳ３００６で、力み音声が検出された入力音声中の区間を、力み音声のない通常発声の学習データより作成された音響モデルにより音声認識する場合には認識精度が低下することに対応させて、「力み」で発声された「えんぴつが」の部分については言語モデルの重みを大きくする。すなわち図１７（ｃ）に示すように力み音声の発声を含んだデータより作成した言語モデルの重みα＝２．３を適用する。これにより、

【数８】

$$\log P(Y/W_1) + \sum_{i=1}^{n} \alpha_i \log P(w_{1,i} \mid w_{1,1} \cdots w_{1,i-1}) < \log P(Y/W_2) + \sum_{i=1}^{n} \alpha_i \log P(w_{2,i} \mid w_{2,1} \cdots w_{2,i-1})$$

となり、認識結果としては「名前を書く鉛筆が欲しいんです」が採用され、正しい認識結果を得ることができる。

【０１１６】

なお、本実施の形態において力み音声を含まない通常発声のフレームに適用する言語モデルの重みを０．９、力み音声で発声されたフレームに適用する言語モデルの重みを２．３としたが、力み音声で発声されたフレームにおいて言語モデルの重みが相対的に大きくなる限りにおいて、これ以外の値であっても良い。

【０１１７】

また、本実施の形態において、基本周波数分析部１０１がマイクロホン１から入力音声を取得して基本周波数を求めたが、実施の形態２のように逆フィルタ２０２を用いて音源波形を抽出し、音源波形から基本周波数を求め、以降の処理を行うものとしても良い。

【０１１８】

さらに、本実施の形態において音声認識の音響特性モデルはメルケプストラム係数のパラメータを用いるものとしたが、それ以外のケプストラム係数等、音声の周波数特性を記述し、音声認識に用いられる特徴量であればどのような特徴量を用いても良い。

【０１１９】

かかる構成によれば、入力された音声より怒りや苛立ちが反映された力み音声を検出し、力み音声は音響特徴量データベース内の音響モデルに合致しにくいことを考慮して言語モデルの重み係数αを大きくし、相対的に音響モデルの重みを軽くすることができる。これにより、音響モデルが合致しないことによる音韻レベルの誤認識を防ぎ、文レベルの音声認識精度を向上させることができる。さらには言語モデルの重みは、既存の言語モデルと音響モデルとのバランスを決定するものであるため、力み音声の音響モデルを生成する

必要がなく、実施の形態２のように力み音声の音響モデルを使用する場合に比べ、少量のデータで音声認識処理が可能である。

【０１２０】

本実施の形態のように音響モデルとあわせて言語モデルを使用する音声認識においては、音韻列のみでなく単語境界の判定もおこなわれる。このため、力み音声との対応によりユーザがどの単語やフレーズに力を入れて話していたかを容易に解析することができる。コールセンターの利用者の音声や、店頭での顧客の音声に本実施の形態に係る音声認識装置を適用させ、上記解析を行なうことにより、発話中のどの単語やフレーズが力んでいるかを知ることができ、クレーム内容をより的確に分類してマーケティングに反映させることができる。

【０１２１】

また音声入力による電子メール作成等に上述の音声認識装置を用いれば、単語の後ろに怒りや苛立ちを表す絵文字を自動で挿入することができる。これにより、文字だけで伝わりにくい感情を、煩雑な手間をなしに、受信者に伝えることができる。

【０１２２】

また、本実施の形態に示すような力み音声検出装置を含む音声認識装置を対話制御等に用いる場合には、力み音声検出部１００の出力を利用することにより、ユーザである話者が対話動作過程のどのイベントに対して怒りや苛立ちを覚えたかを特定することができる。このため、例えば、ユーザの怒りに対して、システム側の出力音声を「大変申し訳ございませんが・・・」というようなより丁寧な謝罪や、「お手数ではございますが・・・」というような丁寧な依頼の表現にしたりすることができる。これにより、ユーザの感情を平常な状態に導き、ユーザによる発話を誤認識の少ない通常音声での発話へ誘導し、対話インタフェースとしてスムーズに動作する環境を整えることができる。

【０１２３】

（実施の形態４）

実施の形態４では、本発明の力み音声検出装置を音声認識装置に応用した怒り認識機能付音声認識装置について説明する。

【０１２４】

図１８は、本発明の実施の形態４における怒り認識機能付音声認識装置の構成を示す機能ブロック図である。図１９は、実施の形態４における怒り認識機能付音声認識装置の動作を示したフローチャートである。また、図２０は、後述する力み音声発生指標計算規則記憶部４１４に記憶された計算規則の一例を示す図である。

【０１２５】

図１８において、図１４と同じ部分については説明を適宜省略し、図１４と異なる部分を中心に説明する。図１９においても図１５と同じ部分については説明を適宜省略し、図１５と異なる部分を中心に説明する。

【０１２６】

図１８において、怒り認識機能付音声認識装置の構成は、図１５の機能ブロック図に怒り強度判定部４１０が付け加わり、力み音声検出部１００が力み音声検出部４００に置き換わり、連続単語音声認識部３０３が連続単語音声認識部４０３に置き換わり、表示部２０９が表示部４１８に置き換わった以外は図１４と同様の構成である。

【０１２７】

力み音声検出部４００は、図８に示した実施の形態１および実施の形態３における力み音声検出部１００と同様に構成される。ただし、力み音声検出部４００は、入力音声の有声区間と基本周波数との入力を受付け、力み音声の検出結果のみではなく、実施の形態１のステップＳ１００７で求められた振幅包絡の多項式近似の結果を振幅パターンとして出力する。さらに、力み音声検出部４００は、基本周波数分析部１０１より出力された基本周波数をも出力する。

【０１２８】

連続単語音声認識部４０３は、実施の形態３における連続単語音声認識部３０３と同様

に音声認識用特徴量抽出部２０１が出力した音響特徴量と、力み音声検出部４００が出力した力み音声の検出結果とを受付け、さらに力み音声検出部４００が出力する基本周波数パターンと振幅パターン情報とを受付ける。連続単語音声認識部４０３は、これらの入力に基づき音響特徴量データベース３０１と言語特徴量データベース３０２とを参照して連続音声認識を行い、入力音声中の時間位置情報として力み音声検出部４００より出力された力み音声区間、基本周波数パターンおよび振幅パターンを、認識結果である音韻列に対してアラインメントする。連続単語音声認識部４０３は、認識結果である音韻列および単語列と共に、音韻単位で記述された力み音声発生位置、ならびに音韻列に対応付けられた基本周波数パターンおよび振幅パターンを出力する。

【０１２９】

怒り強度判定部４１０は、言語処理辞書４１１と、言語処理部４１２と、韻律情報生成部４１３と、力み音声発生指標計算部４１５と、怒り強度決定規則記憶部４１６と、怒り強度決定部４１７とを含む。

【０１３０】

言語処理辞書４１１は、単語ごとに少なくとも読み、アクセント、アクセント結合属性、品詞を記憶する記憶装置である。言語処理部４１２は、言語処理辞書４１１を参照して単語列の言語解析をし、係り受け距離に基づく単語結合度情報を出力する処理部である。韻律情報生成部４１３は、言語処理部４１２より出力された単語アクセント、アクセント結合情報および単語結合度情報と、連続単語音声認識部４０３より出力された音韻位置に対応付けられた基本周波数パターンおよび振幅パターンとを受付け、アクセント位置、アクセント句区切り、フレーズ区切りの情報を生成する処理部である。

【０１３１】

力み音声発生指標計算規則記憶部４１４は、音韻列とアクセント等の韻律情報とから音韻ごとの力み音声の発生し易さ（あるいは発生しにくさ）である力み音声発生指標を計算するための規則を記憶する記憶装置である。力み音声発生指標計算部４１５は、韻律情報生成部４１３より出力された音韻列と対応付けられたアクセント位置、アクセント句区切り、フレーズ区切りを受付け、力み音声発生指標計算規則記憶部４１４を参照して音韻ごとに力み音声発生指標を計算する処理部である。

【０１３２】

怒り強度決定規則記憶部４１６は、力み音声発生指標より怒り強度を決定するための規則を記憶する記憶装置である。怒り強度決定部４１７は、力み音声発生指標計算部４１５より出力された音韻ごとの力み音声発生指標と、音韻に対応付けられた入力音声中の力み音声発生位置とから、怒り強度決定規則記憶部４１６を参照して怒り強度を決定する処理部である。

【０１３３】

表示部４１８は、連続単語音声認識部４０３より出力された音声認識結果と、怒り強度決定部４１７より出力された音韻ごとの怒り強度とを対応付けて表示する表示装置である。

【０１３４】

このように構成された怒り認識機能付音声認識装置の動作について図１９に従って説明する。図１５と同じ動作については説明を省略し、異なる部分についてのみ説明する。

【０１３５】

マイクロホン１より音声が入力される（ステップＳ１００１）。音声認識用特徴量抽出部２０１は、入力音声を分析し、メルケプストラム係数を抽出する（ステップＳ２００２）。一方、基本周波数分析部１０１は、入力音声の周期性の有無を分析するとともに、周期性のある部分についてはその周波数を求める（ステップＳ１００２）。さらに、基本周波数分析部１０１は、ステップＳ１００２において音声中の有声区間とみなされた区間を抽出する（ステップＳ１００３）。

【０１３６】

力み音声検出部４００は実施の形態３と同様にしてステップＳ１００３で抽出された有

声区間において推定振幅パターンを生成して力み音声を検出する（ステップＳ４００５）。つまり、力み音声検出部４００は、音声中の各有声区間について、実施の形態３に示した図１６のステップＳ１００４からステップＳ１０１５までの処理を繰り返して、力み音声の区間を検出する。処理の概要は以下のとおりである。つまり、有声区間に対し、ステップＳ１００５で第２高調波成分を抽出し、ステップＳ１００６で第２高調波成分の振幅包絡を抽出する。ステップＳ１００７において振幅包絡を多項式で近似し、振幅変調前の音声の包絡線を推定する。ステップＳ１００８で多項式による近似値と包絡線との差分を求め、ステップＳ１００９で変調度を求める。変調度が基準値を超える場合（ステップＳ１０１０でＹＥＳ）、ステップＳ１０１２で振幅包絡の瞬時周波数の中央値を求め、これを変調周波数とする。変調周波数が基準範囲内である場合（ステップＳ１０１３でＹＥＳ）、ステップＳ１０１５において当該有声区間を力み音声の区間と判定する。力み音声検出部４００は、ステップＳ１００７において多項式近似によって推定された有声区間の変調前振幅包絡すなわち振幅パターンを、すべての有声区間について力み音声区間の検出結果と共に出力する。

【０１３７】

連続単語音声認識部４０３は、音響特徴量データベース３０１と言語特徴量データベース３０２とを参照し、ステップＳ２００２で抽出されたメルケプストラム係数を用いて音声認識を行う（ステップＳ３００６、Ｓ３００７）。つまり、連続単語音声認識部４０３は、音響特徴量データベース３０１と言語特徴量データベース３０２とを参照して音声認識を行う際に、音声認識を行うフレームが力み音声を含む場合には言語モデルの重みの値を大きくし、相対的に音響モデルの重みを小さくすることにより（ステップＳ３００６）、音声認識を行う（ステップＳ３００７）。

【０１３８】

さらに、連続単語音声認識部４０３は、認識結果である入力音声の時間位置に対する音韻ラベルを元に、力み音声検出部４００より出力された有声区間の基本周波数パターン、振幅パターンおよび力み音声区間と音韻列中の各音韻との対応付けを行う（ステップＳ４００８）。

【０１３９】

言語処理部４１２は、連続単語音声認識部４０３より出力された認識結果である音韻列および単語列に基づき、言語処理辞書４１１を参照して言語解析を行い、単語間の係り受け情報を生成する（ステップＳ４００９）。係り受け情報の解析方法としては、例えば、「情報処理学会研究報告、２０００－ＮＬ－１３８、ｐｐ７９－８６、２０００年７月」に示されるような統計学習による解析方法を用いる。言語処理部４１２は、単語の係り受け解析結果を元に隣り合う単語の結合度情報を生成し、音韻列とあわせて単語ごとの単語結合度、単語のアクセント、単語のアクセント結合情報、さらに連続単語音声認識部４０３より受付けた各音韻に対応付けられた基本周波数パターン、振幅パターンおよび力み音声検出結果を出力する。

【０１４０】

韻律情報生成部４１３は、言語処理部４１２の出力を受付け、アクセント句区切りとフレーズ区切りとを決定し、アクセント位置を決定する（ステップＳ４０１０）。つまり、韻律情報生成部４１３は、単語結合度が大きいものから順に１アクセント句が９モーラを超えない範囲でアクセント句を結合し、結合度の値が無いあるいは非常に低い「節」の切れ目をフレーズの区切りとする。韻律情報生成部４１３は、また、生成したアクセント句内に含まれる単語のアクセント結合情報に基づき、アクセント句内で１つのアクセント位置の設定を行う。このようにして単語列の情報より作成されたアクセント位置、アクセント句区切り、フレーズ区切りの情報について、音韻に対応付けられた基本周波数パターンおよび振幅パターンの立ち上がり部分または立下り部分がアクセント句区切りと一致しない場合には、韻律情報生成部４１３は、アクセント句区切りが基本周波数パターンおよび振幅パターンの立ち上がり部分または立下り部分と一致するように修正し、それに伴ってアクセント位置を修正する。フレーズ区切りが基本周波数パターンおよび振幅パターンの

立下りでない部分に設定されている場合には、韻律情報生成部４１３は、フレーズ区切りをアクセント句区切りに修正する。

【０１４１】

力み音声発生指標計算部４１５は、力み音声発生指標計算規則記憶部４１４に記憶された、子音、母音、アクセント句中の位置、アクセント核からの相対位置等の音韻属性から力み音声の発生しやすさを求める規則を用いて力み音声発生指標を音韻ごとに計算する（ステップＳ４０１１）。力み音声発生指標の計算規則は、音韻属性から力み音声の発生しやすさを数値で表現できるモデルにより表される。このようなモデルは、例えば、力み音声を含む音声データより、音韻属性を説明変数とし、力み音声が発生したか否かの２値を従属変数とし、質的データを取り扱う統計的学習手法の１つである数量化ＩＩ類を用いて統計的学習を行うことにより得られる。力み音声発生指標計算規則記憶部４１４は、例えば図２０のように統計的学習によって得られた音韻属性に対応するモデルパラメータを記憶しているものとする。力み音声発生指標計算部４１５は、各音韻の属性に従って、力み音声発生指標計算規則記憶部４１４に記憶された統計モデルを適用し、力み音声発生指標を計算する。このような発生指標の計算方法は国際公開第２００６／１２３５３９号パンフレットに詳述されている。

【０１４２】

怒り強度決定規則記憶部４１６は、力み音声発生指標計算規則を統計的に学習した際に、力み音声を含む音韻で力み音声発生指標が低い傾向が見られたか、高い傾向が見られたかにより決定された怒り強度計算規則を記憶するものである。学習データにおいて、力み音声を含む音韻で力み音声発生指標が低い傾向が見られた場合には、力み音声発生指標は「力み難さ」の指標と考えられる。怒り強度決定規則は、力み音声発生指標が高い音韻すなわち力み難さが高い音韻が力んで発声されている場合には怒りの強度が大きいと判断され、力み音声発生指標が低い音韻すなわち力み難さが低い音韻が力んで発声されている場合には怒りの強度が小さいと判断されるように設定された規則である。

【０１４３】

怒り強度決定部４１７は、ステップＳ４０１１において力み音声発生指標計算部４１５で計算された力み音声発生指標に基づき、怒り強度決定規則記憶部４１６に記憶された規則を参照して怒りの強度を決定する（ステップＳ４０１２）。表示部４１８は、ステップＳ３００７で求められた音声認識結果と共に、怒りの強度を表示する（ステップＳ４０１３）。

【０１４４】

かかる構成によれば、入力された音声より力み音声を抽出し、一方では力み音声が抽出された場合には音響特徴量データベース内の音響モデルに合致しにくいことを考慮して言語モデルの重み係数を大きくし、相対的に音響モデルの重みを軽くすることができる。これにより、音響モデルが合致しないことによる音韻レベルの誤認識を防ぎ、文レベルの音声認識精度を向上させることができる。他方では音声認識結果を利用して力み音声の発生しやすさあるいは発生しにくさを計算して、力み音声が発生しやすい部分で実際に力み音声が発生している場合には怒りの強度が低いと判断し、力み音声が発生しにくい部分で入力音声に力み音声が発生している場合には怒りの強度が高いと判断することができる。これにより、入力音声から、発話者の怒りや苛立ちの強度を個人差や地方差に影響されることなく正確に認識することができる。

【０１４５】

さらに、言語モデルの重みは、既存の言語モデルと音響モデルとのバランスを決定するものであるため、力み音声の音響モデルを生成する必要がなく、実施の形態２のように力み音声の音響モデルを使用する場合に比べ、少量のデータで音声認識処理が可能である。

【０１４６】

また、力み音声に対して無表情な音声データから作られた音響特徴量データベースを用いて音声認識を行なう場合には精度が低い。しかし、力み音声が発生している部分については音響モデルが適切でない可能性があるとして、音響モデルの重みを軽くし言語モデル

の重みを大きくすることにより不適切な音響モデルを適用することの影響を小さくすることで、音声認識精度も向上する。音声認識精度の向上により、音韻列を用いて計算する力み音声の発生しやすさの計算精度も向上するため、怒り強度の計算も精度が向上する。

【０１４７】

さらに、力み音声を音韻単位で検出し、怒り強度の判断を音韻単位で行うことで、入力音声中の感情の変化に音韻単位で追随することができる。従って本実施の形態の怒り認識機能付音声認識装置を対話制御等に用いる場合には、ユーザである話者が対話動作過程のどのイベントに対して、どのような反応をしたかを特定する場合に苛立ち出したタイミングを詳細に捉えることができる。また、怒りや苛立ちの強度も合わせて分かるため、非常に効果的である。入力音声より、ユーザの感情の変化のタイミングおよび感情の強度をともに詳細に捉えることができるため、例えば、ユーザの怒り強度に合わせて、システム側の出力音声を切り替えることができる。例えば、「大変申し訳ございませんが・・・」という丁寧な謝罪や、「お手数ではございますが・・・」という丁寧な依頼の表現のなかでも申し訳なさの程度を複数用意し、ユーザの怒りの強度に合わせて音声を出力することで、ユーザを必要以上に恐縮させてしまったり、あるいは丁寧すぎて逆効果になってしまうことがない。このため、ユーザの感情を平常な状態に導き、対話インタフェースとしてよりスムーズに動作することができる。

【０１４８】

また、認識結果と共に怒りの強度を記録することで、コールセンター等では利用者の音声から、対話内容、発話内容および怒りの強度変化の対応関係を分析することができる。このような分析結果は、クレーム分類や、対応の良し悪しを分類する際に有効となる。コールセンター等から担当者へ電話を回す際、怒り強度データもあわせて送信し、担当者側で怒り強度を表示するようにすれば、担当者が電話を受けるときには利用者の怒りの状況が分かり、正しい対応がし易くなる。

【０１４９】

なお、本実施の形態の音声認識処理では、実施の形態３と同様に、連続単語音声認識部４０３が、力み音声の有無によって重みを変えながら音響特徴量データベース３０１および言語特徴量データベース３０２を参照するものとしたが、実施の形態２のように、力み音声の有無によって標準音声特徴量データベース２０６と力み音声特徴量データベース２０７とを切り替えながら、音声認識を行うものとしてもよい。

【０１５０】

また、本実施の形態において、基本周波数分析部１０１がマイクロホン１から入力音声を取得して基本周波数を求めたが、実施の形態２のように逆フィルタ２０２を用いて音源波形を抽出し、音源波形から基本周波数および振幅パターンを求めるものとしても良い。

【０１５１】

さらに、本実施の形態において、韻律情報生成部４１３はアクセント、アクセント句区切り、フレーズ区切りを決定する際に、言語処理部４１２で求められた単語アクセント、アクセント結合情報および単語結合度情報と、連続単語音声認識部４０３で音韻と対応付けられた基本周波数パターンおよび振幅パターンとの両方の情報を用いるものとしたが、いずれか一方の情報を用いるものであってもよい。つまり、韻律情報生成部４１３は、言語処理部４１２で求められた単語アクセント、アクセント結合情報および単語結合度情報のみからアクセント、アクセント句区切り、フレーズ区切りを決定するものとしても良い。また、韻律情報生成部４１３は、連続単語音声認識部４０３で音韻と対応付けられた基本周波数パターンおよび振幅パターンのみからアクセント、アクセント句区切り、フレーズ区切りを決定しても良い。ただし、両方の情報を用いた方が、精度が向上する。なお、基本周波数パターンおよび振幅パターンのみからアクセント、アクセント句区切り、フレーズ区切りを決定する場合は、言語処理部４１２、言語処理辞書４１１は不要となり、ステップＳ４００９は省略しても良い。

【０１５２】

また、本実施の形態において、言語特徴量データベース３０２と言語処理部４１２と言

語処理辞書４１１とは独立の構成としたが、言語特徴量データベース３０２が言語処理辞書４１１の内容を含み、連続単語音声認識部４０３が認識結果である音韻列および単語列と共に、単語アクセント、アクセント結合情報、単語結合度情報をも生成するものとしても良い。その際、言語処理部４１２は、連続単語音声認識部４０３に包含され、連続単語音声認識部４０３は、韻律情報生成部４１３へ音韻列、単語列、単語アクセント、アクセント結合情報、単語結合度情報、力み音声発生位置、基本周波数パターン、振幅パターンを出力するものとする。

【０１５３】

なお、実施の形態４において力み音声発声指標の計算規則のモデルの学習には、統計的学習手法である数量化ＩＩ類を用い、説明変数には子音、母音、アクセント句中の位置、アクセント核からの相対位置を用いたが、統計的学習手法はこれ以外の方法でも良い。また、説明変数としては、上記属性のみではなく、基本周波数やパワーとそのパターン音韻の時間長等の連続量を用いてもよい。

【０１５４】

なお、実施の形態４においては、実施の形態３と同様に音声認識部は音響特徴量データベースと言語特徴量データベースを用いて、力み音声の検出により両データベースの重みを変更するものとしたが、実施の形態２のように、力み音声の検出により標準音声特徴量データベース２０６と力み音声特徴量データベース２０７とをスイッチで切り替えながら音声認識部が音声認識を行うものであってもよい。実施の形態２のようなデータベースを切り替える方法を採用する場合、力み音声検出部は入力音声中の当該処理フレームが力み音声であるか否かをスイッチに出力するのみでなく、音声認識部にも出力する。音声認識部は認識結果の音韻と合わせて、各音韻が力み音声であったか否かの情報を怒り強度判定部に出力する。怒り強度判定部の言語処理部は音韻列より言語処理辞書を参照し、言語モデルに従って単語境界、アクセント等の情報を生成して力み音声発生指標計算部に出力する。力み音声発生指標計算部は力み音声発生指標を求め、怒り強度決定部は指標に基づき怒り強度を決定する。

【０１５５】

なお、本発明の実施の形態すべてにおいて、入力音声はマイクロホン１より入力されるものとしたが、あらかじめ録音、記録された音声あるいは装置外部より入力される音声信号であっても良い。

【０１５６】

なお、本発明の実施の形態すべてにおいて、力み音声検出結果、音声認識結果、あるいは怒り強度を表示部で表示するものとしたが、記憶装置へ記録する、あるいは装置外部へ出力するものとしても良い。

【０１５７】

また、本発明の実施の形態すべてにおいて、上述した各装置を構成する構成要素の一部または全部は、１個のシステムＬＳＩ（Large Scale Integration）から構成されているとしてもよい。システムＬＳＩは、複数の構成部を１個のチップ上に集積して製造された超多機能ＬＳＩであり、具体的には、マイクロプロセッサ、ＲＯＭ、ＲＡＭなどを含んで構成されるコンピュータシステムである。前記ＲＡＭには、コンピュータプログラムが記憶されている。前記マイクロプロセッサが、前記コンピュータプログラムにしたがって動作することにより、システムＬＳＩは、その機能を達成する。

【０１５８】

今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

【産業上の利用可能性】

【０１５９】

本発明にかかる音声による力み音声検出装置は、怒りや苛立ちに伴う発声器官の緊張に

よって発声する振幅の周期的変動を特徴とする、力み音声を検出するものであり、入力音声から力み音声を検出することで入力音声の話者の怒りや苛立ちを認識するという応用が可能である。従って、ロボット等の音声・対話インタフェース等として有用である。またコールセンターにおけるシステムや、電話交換の自動電話応対システム等の用途にも応用できる。

【図面の簡単な説明】

【０１６０】

【図１】録音された音声における力み音声の感情種類による発生頻度を示すグラフである。

【図２】録音された音声において観察された、通常音声と力み音声の波形と振幅包絡の一例を示す図である。

【図３】録音された音声において観察された力み音声で発声されたモーラの平均基本周波数と振幅包絡の変動周波数との関係を示すグラフである。

【図４】録音された音声において観察された力み音声で発声されたモーラの振幅包絡の変動周波数分布を示すヒストグラムと累積度数グラフである。

【図５】録音された音声において観察された、力み音声の第２高調波、振幅包絡曲線、多項式によるフィッティングの一例、および振幅変動量の計算例を示す図である。

【図６】録音された音声において観察された力み音声で発生されたモーラの振幅包絡の変調度の分布を示すヒストグラムと累積度数グラフである。

【図７】変調度１の場合の変換元波形と変換結果とを模式的に示した図である。

【図８】本発明の実施の形態１における力み音声検出装置の構成を示すブロック図である。

【図９】本発明の実施の形態１における力み音声検出装置の動作を示すフローチャートである。

【図１０】本発明の実施の形態２における力み音声検出装置を含む音声認識装置の構成を示すブロック図である。

【図１１】図１０に示した音声認識装置のうち力み音声検出部の構成を示す機能ブロック図である。

【図１２】本発明の実施の形態２における力み音声検出装置を含む音声認識装置の動作を示すフローチャートである。

【図１３】本発明の実施の形態２における力み音声検出装置を含む音声認識装置の動作の一部を示すフローチャートである。

【図１４】本発明の実施の形態３における力み音声検出装置を含む音声認識装置の構成を示すブロック図である。

【図１５】本発明の実施の形態３における力み音声検出装置を含む音声認識装置の動作を示すフローチャートである。

【図１６】本発明の実施の形態３における力み音声検出装置を含む音声認識装置の動作の一部を示すフローチャートである。

【図１７】本発明の実施の形態３における力み音声検出装置を含む音声認識装置の音声認識処理の具体例を示す図である。

【図１８】本発明の実施の形態４における怒り認識機能付音声認識装置の構成を示すブロック図である。

【図１９】本発明の実施の形態４における怒り認識機能付音声認識装置の動作を示すフローチャートである。

【図２０】本発明の実施の形態４における力み音声発生指標計算規則の一例を示す図である。

【図２１】従来の音声による感情認識装置の構成を示すブロック図である。

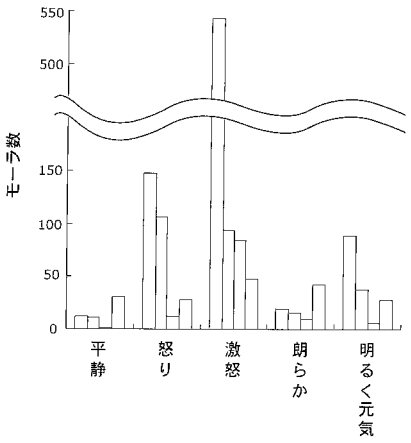
【符号の説明】

【０１６１】

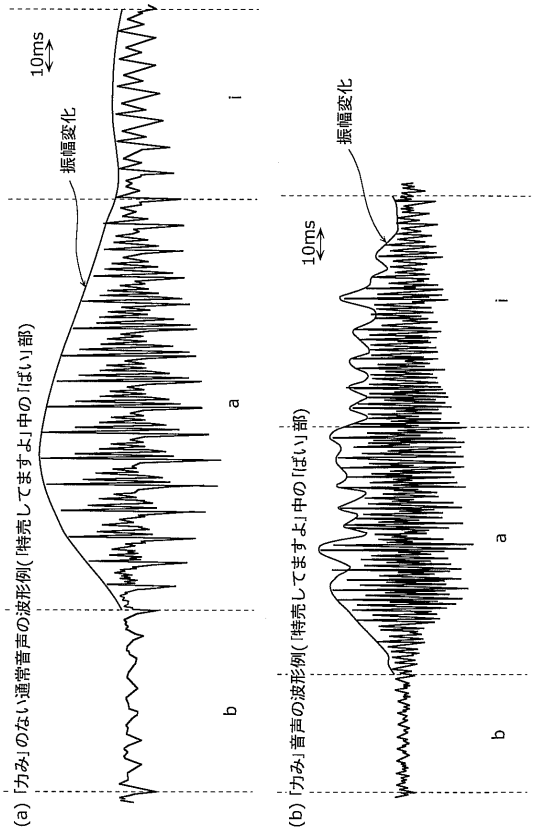
１　マイクロホン

２　音声コード認識手段
３　感性情報抽出手段
４　出力制御手段
５　出力装置
３１　話速検出部
３２　基本周波数検出部
３３　音量検出部
３４　音声レベル判定基準格納部
３５　標準音声特徴量格納部
３６　音声レベル分析部
３７　感性レベル分析用知識ベース格納部
３８　感性レベル分析部
３９　音声スペクトル検出部
４１　主制御部
４２　出力制御用知識ベース格納部
１００、２００、４００　力み音声検出部
１０１　基本周波数分析部
１０２　第２高調波抽出部
１０３　振幅包絡分析部
１０４　振幅変調度計算部
１０５　変調周波数分析部
１０６　力み判定部
１０７、２０９、４１８　表示部
２０１　音声認識用特徴量抽出部
２０２　逆フィルタ
２０３　周期性分析部
２０４　スイッチ
２０５　特徴量データベース
２０６　標準音声特徴量データベース
２０７　力み音声特徴量データベース
２０８　音声認識部
２１２　基本波抽出部
３０１　音響特徴量データベース
３０２　言語特徴量データベース
３０３、４０３　連続単語音声認識部
４１１　言語処理辞書
４１２　言語処理部
４１３　韻律情報生成部
４１４　力み音声発生指標計算規則記憶部
４１５　力み音声発生指標計算部
４１６　怒り強度決定規則記憶部
４１７　怒り強度決定部

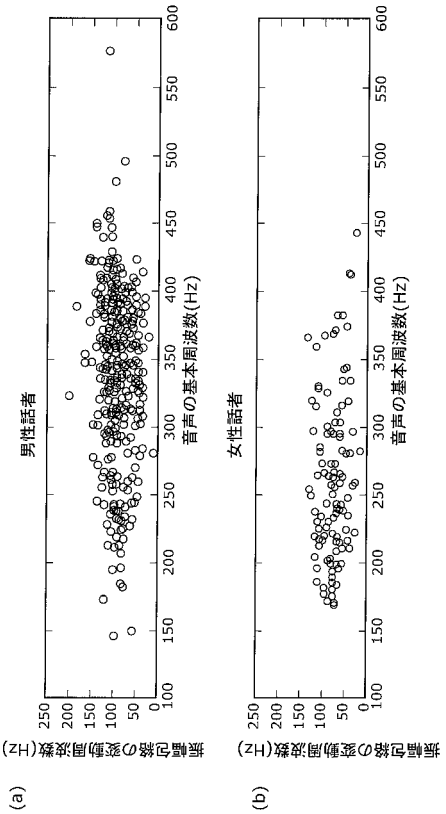
【図１】


モーラ数
550
500
150
100
50
0
平静
怒り
嫌悪
切ない
明るい

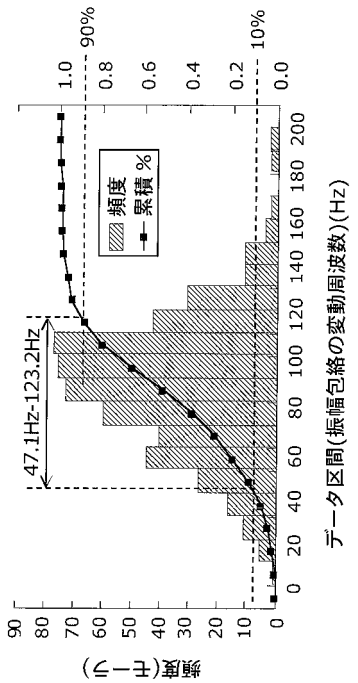
【図２】


10ms
振幅変化
i
a
b
(a)「力み」のない通常音声の波形例(「特売してますよ」中の「ばい」部)
10ms
振幅変化
i
a
b
(b)「力み」音声の波形例(「特売してますよ」中の「ばい」部)

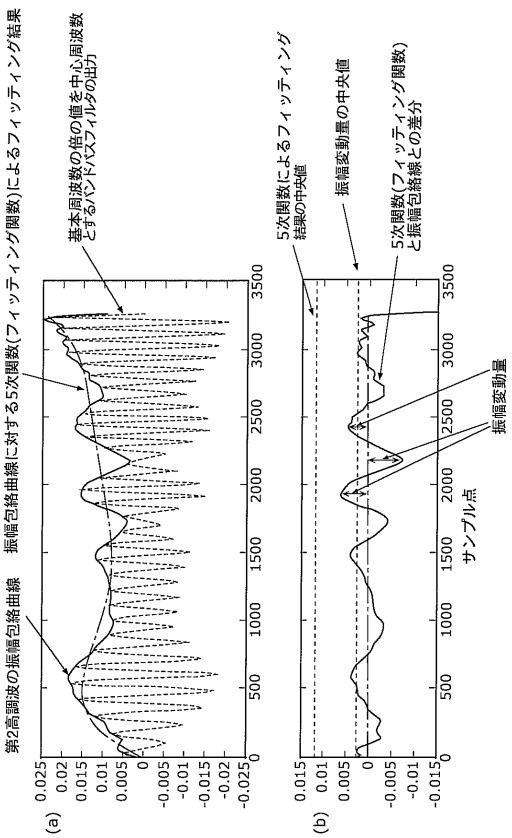
【図３】


男性話者
女性話者
音声の基本周波数(Hz)
音声の基本周波数(Hz)
振幅包絡の変動周波数(Hz)
振幅包絡の変動周波数(Hz)
(a)
(b)

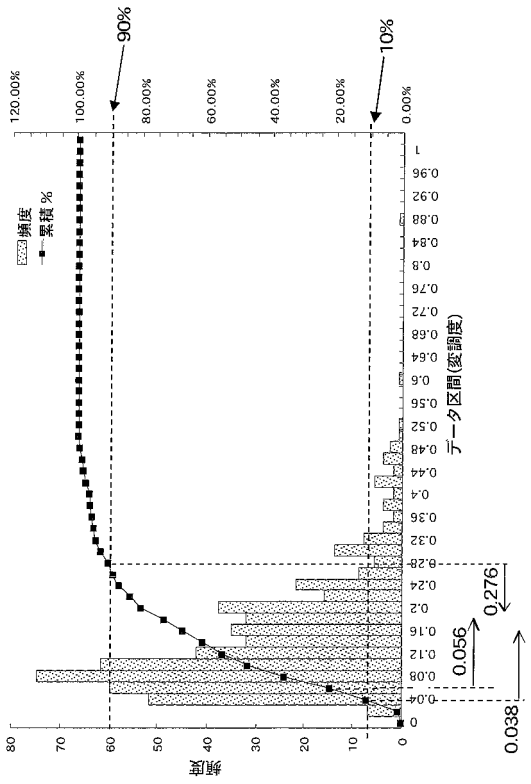
【図４】


90%
10%
47.1Hz-123.2Hz
頻度
累積 %
頻度(モーラ)
データ区間(振幅包絡の変動周波数)(Hz)

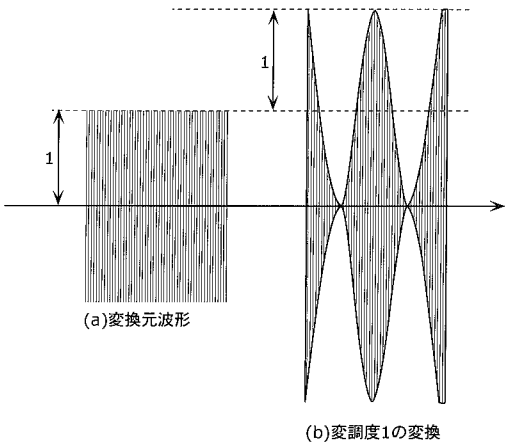
【図５】


第2高調波の振幅包絡曲線
振幅包絡曲線に対する5次関数(フィッティング関数)によるフィッティング結果
基本周波数の倍の値を中心周波数とするバンドパスフィルタの出力
5次関数によるフィッティング結果の中央値
振幅変動量の中央値
5次関数(フィッティング関数)と振幅包絡線との差分
振幅変動量
サンプル点
(a)
(b)

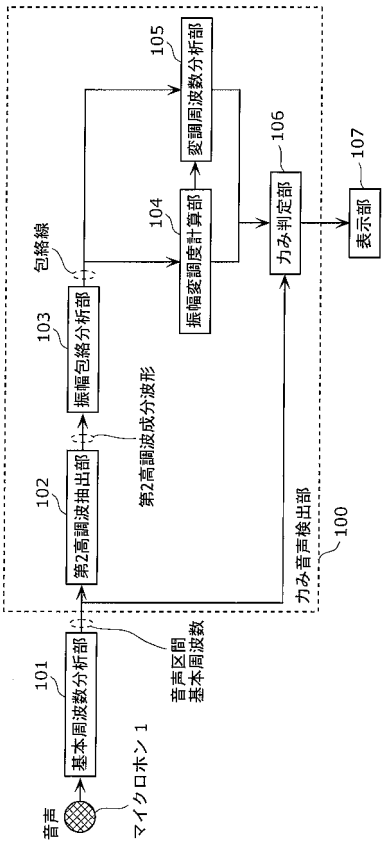
【図６】


90%
10%
頻度
累積 %
データ区間(変調度)
0.276
0.056
0.038

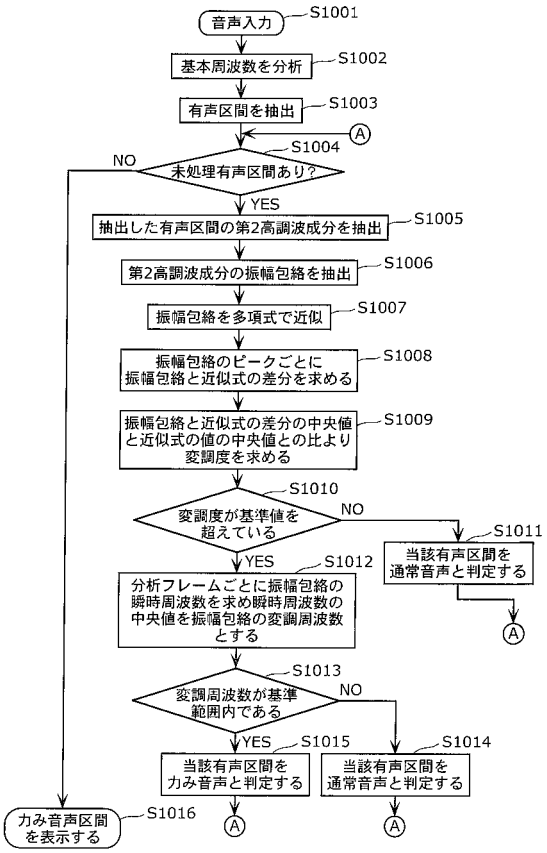
【図７】


1
1
(a)変換元波形
(b)変調度1の変換

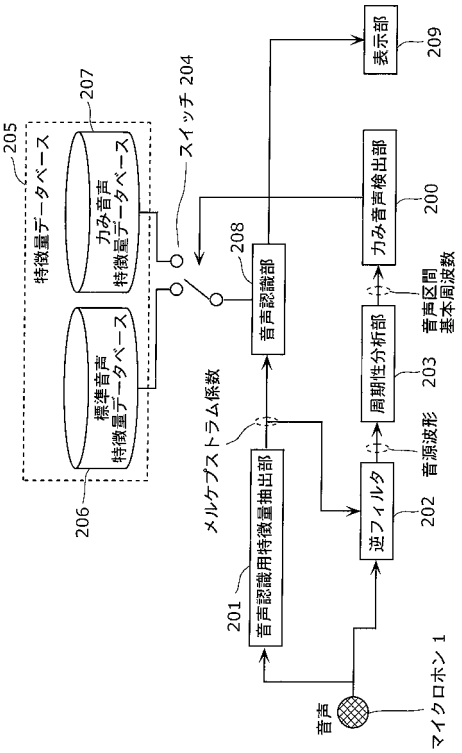
【図８】


音声
マイクロホン 1
101
基本周波数分析部
音声区間
基本周波数
102
第2高調波抽出部
第2高調波成分波形
103
振幅包絡分析部
包絡線
104
振幅変調度計算部
105
変調周波数分析部
106
力み判定部
107
表示部
力み音声検出部
100

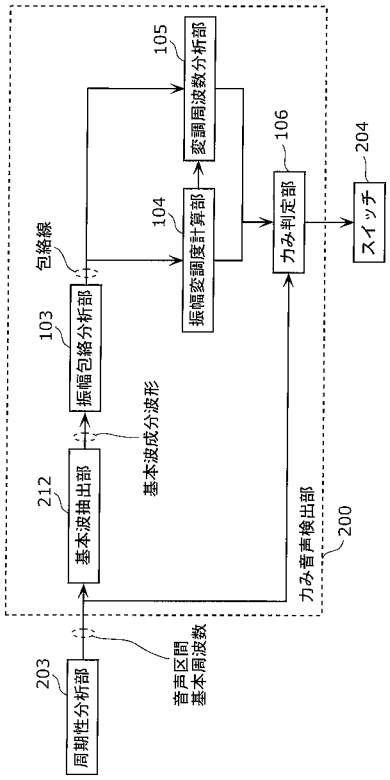
【図９】



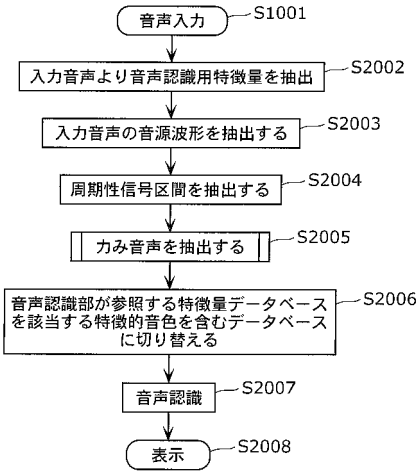
【図１０】



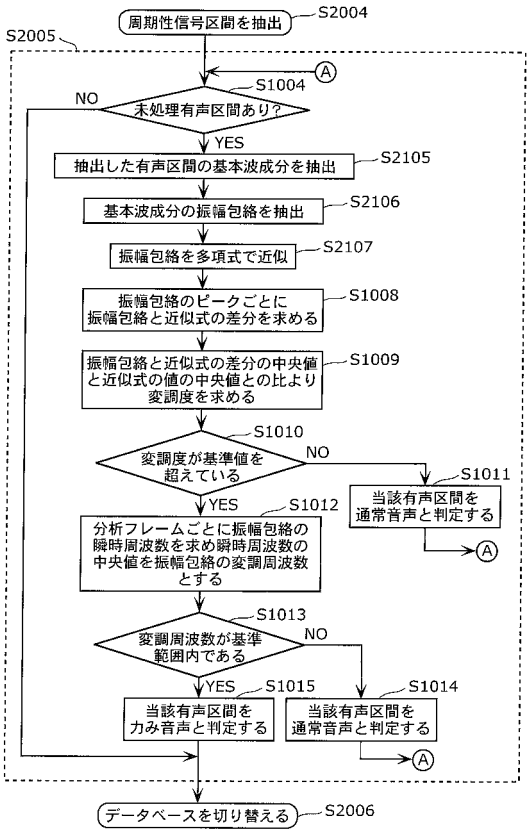
【図１１】



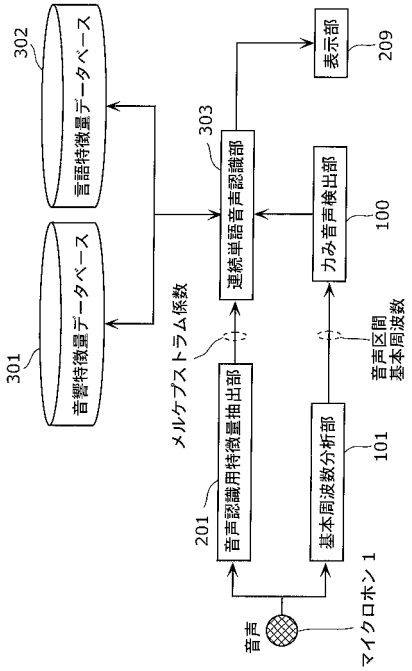
【図１２】



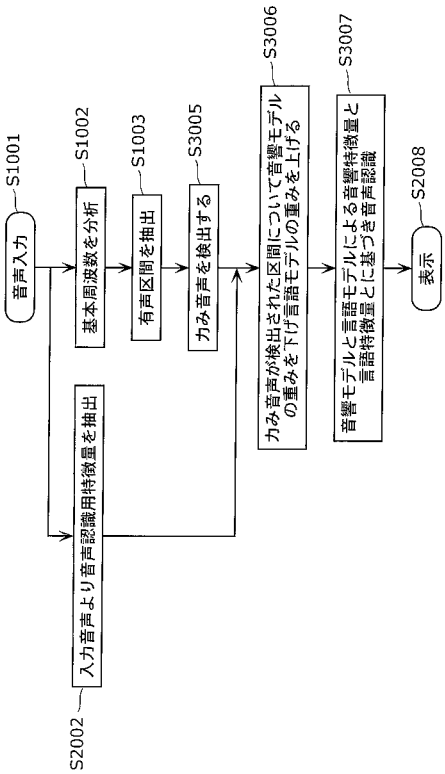
【図１３】



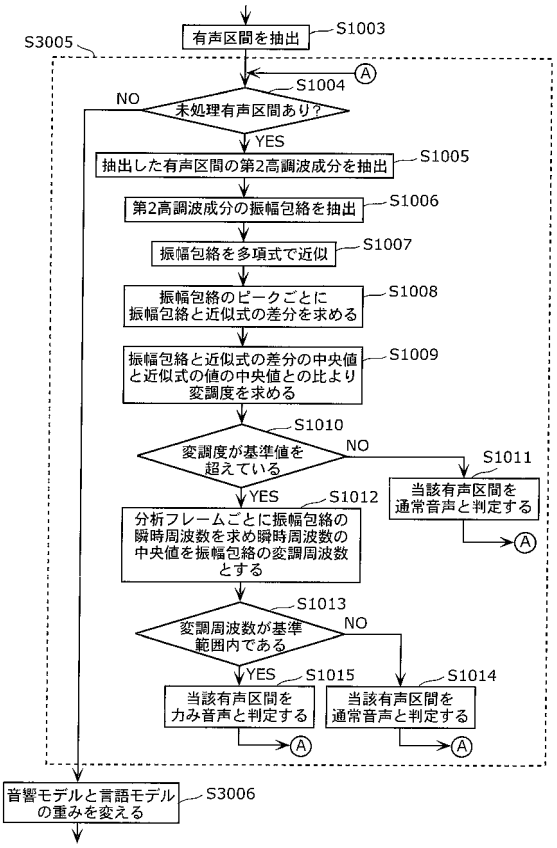
【図１４】



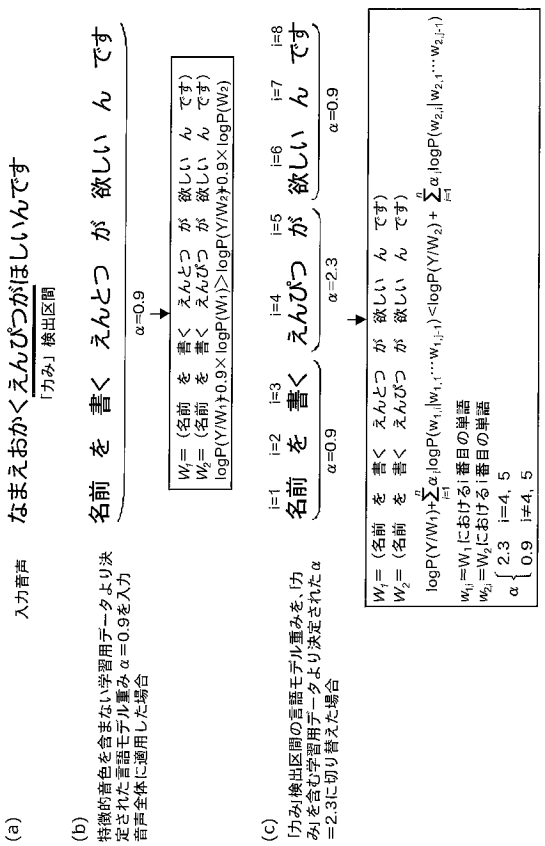
【図１５】



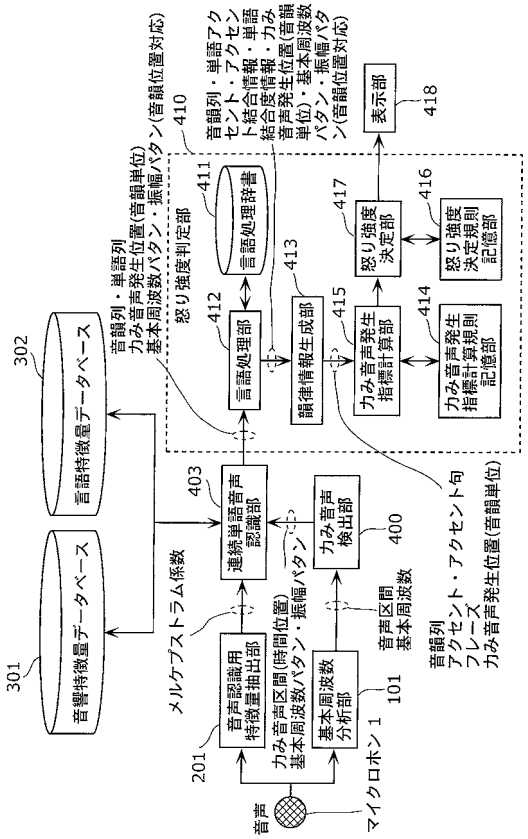
【図１６】



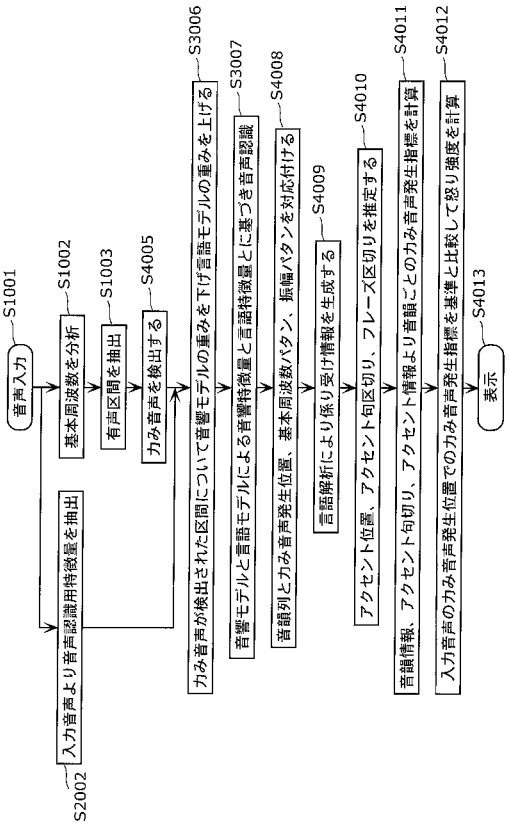
【図１７】



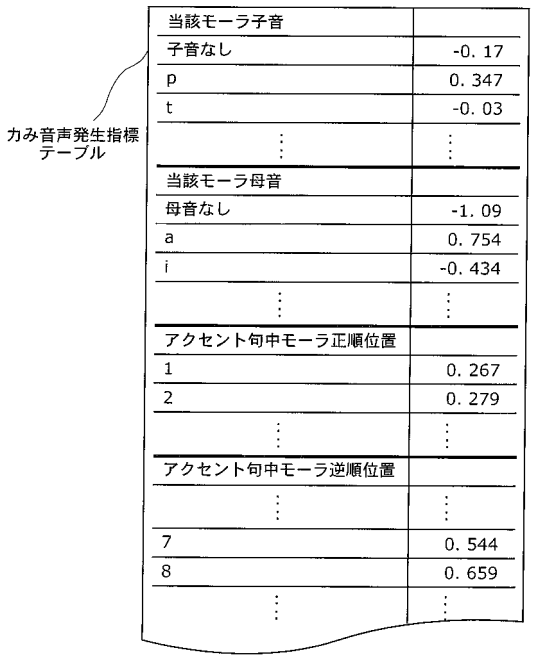
【図１８】



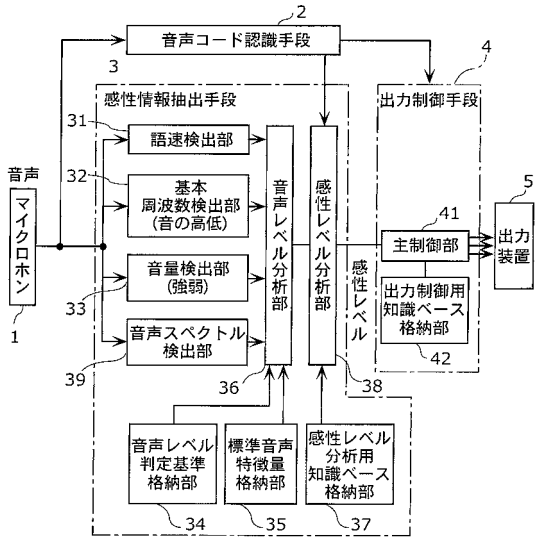
【図１９】



【図２０】



【図２１】


2
音声コード認識手段
3
4
感性情報抽出手段
出力制御手段
31
語速検出部
音声
32
基本
周波数検出部
(音の高低)
5
41
マイクロホン
主制御部
出力
装置
音声レベル分析部
感性レベル分析部
感性レベル
33
音量検出部
(強弱)
出力制御用
知識ベース
格納部
1
音声スペクトル
検出部
42
39
36
38
音声レベル
判定基準
格納部
標準音声
特徴量
格納部
感性レベル
分析用
知識ベース
格納部
34
35
37

フロントページの続き

(51)Int.Cl.	FI		テーマコード（参考）
	Ｇ１０Ｌ 15/20	２００Ｚ	
	Ｇ１０Ｌ 15/10	３００Ｇ	