(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3757638号
(P3757638)
(45)発行日　平成18年3月22日(2006.3.22)　(24)登録日　平成18年1月13日(2006.1.13)

(51)Int.Cl.　　F I
G10L 11/02 (2006.01)　　G10L 3/00 513A
G10L 15/04 (2006.01)　　G10L 3/00 513B

請求項の数 9　(全 17 頁)

(21)出願番号　特願平10-250139
(22)出願日　平成10年9月3日(1998.9.3)
(65)公開番号　特開2000-81891(P2000-81891A)
(43)公開日　平成12年3月21日(2000.3.21)
審査請求日　平成14年10月28日(2002.10.28)

前置審査

(73)特許権者　000002369
セイコーエプソン株式会社
東京都新宿区西新宿２丁目４番１号
(74)代理人　100095728
弁理士　上柳　雅誉
(74)代理人　100107076
弁理士　藤綱　英吉
(74)代理人　100107261
弁理士　須澤　修
(72)発明者　宮沢　康永
長野県諏訪市大和３丁目３番５号　セイコーエプソン株式会社内

審査官　山下　剛史

最終頁に続く

(54)【発明の名称】音声認識方法及び音声認識装置並びに音声認識処理プログラムを記録した記録媒体

(57)【特許請求の範囲】
【請求項１】
単語間に間を有して断続的に発声された複数の単語からなる組を認識する音声認識装置における音声認識方法であって、
前記複数の単語は第１から第ｎ（ｎは正の整数）の単語群に属しており、前記第１から第ｎの単語群ごとに、前記複数の単語が属する全ての単語群に対して、時間的な長さの基準を設定する工程と、
前記時間的な長さの基準を設定する工程の後に、前記単語間に間を有して断続的に発声された前記単語に対応する音声を入力してディジタル化された音声データを出力する音声入力工程と、
前記音声入力工程から出力された音声データを所定時間ごとに分析し、当該所定時間ごとの音声パワーと特徴データを算出する音声分析工程と、
前記音声分析工程によって算出された音声パワーに基づいて当該発声された単語に対応する音声の有効音声区間を検出し、当該有効音声区間の時間的な長さが前記単語群ごとに設定された基準内か否かと、当該有効音声区間内の音声パワーの大きさによって、前記発声された単語に対応する音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を出力する音声区間検出・判定工程と、
前記音声区間検出・判定工程において出力された、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を受けると、前記単語間の間に、前記発声され

た単語に対応する音声が適正な状態で入力されたことを示す情報を出力する情報出力工程と、

前記情報出力工程により情報が出力された後に、前記音声区間検出・判定工程において適正な状態で入力されたと判断された前記発声された単語に対応する音声を認識する音声認識工程と、

を有することを特徴とする音声認識方法。

【請求項２】

前記音声区間検出・判定工程が、有効音声区間の時間的な長さによって前記音声が適正な状態で入力されたか否かを判断する処理は、前記音声パワーが所定のしきい値より大きくなる時刻を始端とし、続いて、前記音声パワーが前記しきい値より小さくなる時刻を終端とし、当該終端の時刻から所定時間の間、前記音声パワーが前記しきい値より大きくならなかった場合に、前記始端から前記終端までの区間を有効音声区間として検出し、当該検出された有効音声区間が、所定の基準内である場合に、前記音声が適正な状態で入力されたと判断することを特徴とする請求項1に記載の音声認識方法。

【請求項３】

前記情報出力工程において出力される前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報は、信号音、光、音声メッセージ、表示画面上での表示の少なくとも１つであることを特徴とする請求項１又は２に記載の音声認識方法。

【請求項４】

単語間に間を有して断続的に発声された複数の単語からなる組を認識する音声認識装置であって、

前記複数の単語は第１から第ｎ（ｎは正の整数）の単語群に属しており、前記第１から第ｎの単語群ごとに、前記複数の単語が属する全ての単語群に対して、時間的な長さの基準を設定する手段と、

前記時間的な長さの基準を設定する手段によって時間的な長さの基準の設定が行われた後に、前記単語間に間を有して断続的に発声された前記単語に対応する音声を入力してディジタル化された音声データを出力する音声入力手段と、

前記音声入力手段から出力された音声データを所定時間ごとに分析し、当該所定時間ごとの音声パワーと特徴データを算出する音声分析手段と、

前記音声分析手段によって算出された音声パワーに基づいて当該発声された単語に対応する音声の有効音声区間を検出し、当該有効音声区間の時間的な長さが前記単語群ごとに設定された基準内か否かと、当該有効音声区間内の音声パワーの大きさによって、前記発声された単語に対応する音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を出力する音声区間検出・判定手段と、

前記音声区間検出・判定手段で検出された前記音声に対する有効音声区間と、前記音声分析手段で算出された特徴データと、に基づいて前記音声を認識処理する音声認識処理手段と、

前記音声区間検出・判定手段から出力された、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を受けると、前記単語間の間に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報を出力する情報出力手段と、

を有し、

前記音声認識処理手段は、前記情報出力手段により情報が出力された後に、前記音声区間検出・判定手段において適正な状態で入力されたと判断された前記発声された単語に対応する音声を認識処理することを特徴とする音声認識装置。

【請求項５】

前記音声区間検出・判定手段が、有効音声区間の時間的な長さによって前記音声が適正な状態で入力されたか否かを判断する処理は、前記音声パワーが所定のしきい値より大きくなる時刻を始端とし、続いて、前記音声パワーが前記しきい値より小さくなる時刻を終端とし、当該終端の時刻から所定時間の間、前記音声パワーが前記しきい値より大きくなら

なかった場合に、前記始端から前記終端までの区間を有効音声区間として検出し、当該検出された有効音声区間が、所定の基準内である場合に、前記音声が適正な状態で入力されたと判断することを特徴とする請求項４に記載の音声認識装置。

【請求項６】

前記情報出力手段が出力する前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報は、信号音、光、音声メッセージ、表示画面上での表示の少なくとも１つであることを特徴とする請求項４又は５に記載の音声認識装置。

【請求項７】

単語間に間を有して断続的に発声された複数の単語からなる組を認識する音声認識装置における音声認識処理プログラムを記録した記録媒体であって、

前記複数の単語は第１から第ｎ（ｎは正の整数）の単語群に属しており、前記第１から第ｎの単語群ごとに、前記複数の単語が属する全ての単語群に対して、時間的な長さの基準を設定する工程と、

前記時間的な長さの基準を設定する工程の後に、前記単語間に間を有して断続的に発声された前記単語に対応する音声を入力してディジタル化された音声データを出力する音声入力工程と、

前記音声入力工程から出力された音声データを所定時間ごとに分析し、当該所定時間ごとの音声パワーと特徴データを算出する音声分析工程と、

前記音声分析工程によって算出された音声パワーに基づいて当該発声された単語に対応する音声の有効音声区間を検出し、当該有効音声区間の時間的な長さが前記単語群ごとに設定された基準内か否かと、当該有効音声区間内の音声パワーの大きさによって、前記発声された単語に対応する音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を出力する音声区間検出・判定工程と、

前記音声区間検出・判定工程において出力された、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を受けると、前記単語間の間に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報を出力する情報出力工程と、

前記情報出力工程により情報が出力された後に、前記音声区間検出・判定工程において適正な状態で入力されたと判断された前記発声された単語に対応する音声を認識する音声認識工程と、

を音声認識装置に実行させるための音声認識処理プログラムを記録した記録媒体。

【請求項８】

前記音声区間検出・判定工程が、有効音声区間の時間的な長さによって前記音声が適正な状態で入力されたか否かを判断する処理は、前記音声パワーが所定のしきい値より大きくなる時刻を始端とし、続いて、前記音声パワーが前記しきい値より小さくなる時刻を終端とし、当該終端の時刻から所定時間の間、前記音声パワーが前記しきい値より大きくならなかった場合に、前記始端から前記終端までの区間を有効音声区間として検出し、当該検出された有効音声区間が、所定の基準内である場合に、前記音声が適正な状態で入力されたと判断することを特徴とする請求項７に記載の音声認識処理プログラムを記録した記録媒体。

【請求項９】

前記情報出力工程において出力される前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報は、信号音、光、音声メッセージ、表示画面上での表示の少なくとも１つであることを特徴とする請求項７又は８に記載の音声認識処理プログラムを記録した記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、話者の発話する音声を認識する際、話者の音声が適正な状態で入力されたか

否かを話者に報知するようにした音声認識方法及びそれを用いた音声認識装置並びに音声認識処理プログラムを記録した記録媒体に関する。

【０００２】

【従来の技術】

近年、音声認識技術は様々な分野で幅広く利用されてきている。特に最近では、日用品とも言える家電製品や、子供用の玩具にまで用いられている。

【０００３】

このように、特定のユーザが対象ではなく、不特定多数の幅広いユーザが使用すると考えられる機器に音声認識技術を用いるとなれば、ユーザの発話した音声を高い確率で認識できるように、ユーザに対し、適正な音声の入力の仕方など使い方について分かり易くガイドするなどして、使い勝手のよいものとすることが重要な課題である。

【０００４】

たとえば、幅広いユーザを対象とした音声認識を用いた機器の１つとして最近では音声時計なるものが開発されつつある。これは、時計に設けられたボタンなどを押すと、現在の時刻を「何時何分」というように音声で教えてくれるものである。

【０００５】

この音声時計は、暗闇の中でも簡単に現在の時刻を知ることができることから、真夜中に目を覚ましたとき現在時刻を知りたいときや、夜間に外出中、暗闇の中で時刻を知るときに便利なものであり、さらには、目の不自由な人にとっても大変便利なものとなる。また、このような音声時計は子供の玩具にも適用される可能性もある。

【０００６】

また、このような音声時計は、時刻を音声によって出力するだけでなく、現在時刻合わせやアラーム時刻の設定なども音声によって行うことができる。たとえば、現在時刻が午前６時３０分であれば、音声時計を現在時刻設定モードとして、ユーザが「午前」、「６時」、「３０分」というように、必要な単語を決められた順番に発話する。そして、音声時計側では、ユーザの発話した音声を認識し、その認識結果に基づいて時刻合わせ処理を行う。アラーム時刻の設定も同様であり、アラーム時刻設定モードとして希望のアラーム時刻を発話する。

【０００７】

このような操作により時刻合わせが行われるが、このとき、ユーザ側からすると、自分の発話した音声が適正な状態（認識処理を行う上で適正な状態）で入力されたかどうかが不安となることが多い。

【０００８】

これを解消するために、ユーザの発話した１単語ごとにその単語に対する認識結果を応答しながら音声入力する方式のものもある。たとえば、前述した発話内容例においては、ユーザが「午前」と発話し、それに対する認識結果として、装置側からは、「午前」という応答が返ってきて、続いて、ユーザが「６時」と発話すると、装置側からは、「６時」という応答が返り、さらに、ユーザが「３０分」と発話すると、装置側からは、「３０分」という応答が返ってくるというような動作を行う。なお、この場合、ユーザの発話した音声が不適切であって、認識ができなかった場合には、装置側からの応答がなかったり、「もう一度発話して下さい」といった応答がなされるようにすることもできる。

【０００９】

このように、ユーザの発話した１単語ごとにその認識結果を応答したり、認識できなかった場合には、それに対する何らかの応答が返ってくるというようにすれば、ユーザ側にとっては、自分の発話した内容が適切であるか否かがわかり、しかも、それがどのように認識されたかがわかるので安心感が得られ、使い勝手のよいものとなる。

【００１０】

【発明が解決しようとする課題】

しかしながら、前述のように、１単語ごとにそれを認識して応答するのでは、たとえば、時刻合わせというような１つの設定動作を行わせるために多くの時間を要する問題がある

。また、このような音声認識技術を日用品や玩具など低価格が要求される機器に適用する場合、コストをできる限り低く抑えることが必要となってくるため、ＣＰＵの処理能力やメモリの容量には大ききな制約がある。したがって、装置側ではＣＰＵに大きな負担をかけたり、メモリを多く使う処理はできるだけ少なくすることが要求される。

【００１１】

これに対処するには、前述の時刻合わせを例に取れば、ユーザが１単語発話するとそれを認識してその認識結果を応答するというのではなく、ユーザに時刻合わせに必要な内容として、たとえば、「午前」・「６時」・「３０分」を１つの組とし、この１つの組を構成する単語を、１単語ごとに少し間をおきながら断続的に発話してもらい、その発話内容について音声認識するということが考えられる。この場合、１つの組を構成する複数の単語の１つ１つに対して装置側から、前述したような認識結果の応答はないので、確かに、時刻設定時間は短くできる。

【００１２】

しかし、このように複数の単語からなる比較的長い一連の音声を始めから最後まで発話することにより装置側に入力する方法では、前述したように、ユーザにとっては、自分の発話した１単語ごとの音声が、適正な状態で入力されたかどうかということが不安となって残る。したがって、ユーザの発話した音声が、機器に適正な状態で入力されたか否かを、面倒な処理を行うことなく、何らかの形でユーザに報知することが必要となってくる。

【００１３】

そこで本発明は、ユーザの発話した音声に対し、その音声が音声認識を行う上で適正に入力されたか否かを、 簡単な処理を行うだけで、ユーザに報知可能とし、音声入力操作時の使い勝手を向上させることを目的としている。

【００１４】

【課題を解決するための手段】

前述した目的を達成するために、本発明の音声認識方法は、単語間に間を有して断続的に発声された複数の単語からなる組を認識する音声認識装置における音声認識方法であって、前記複数の単語は第１から第ｎ（ｎは正の整数）の単語群に属しており、前記第１から第ｎの単語群ごとに、前記複数の単語が属する全ての単語群に対して、時間的な長さの基準を設定する工程と、前記時間的な長さの基準を設定する工程の後に、前記単語間に間を有して断続的に発声された前記単語に対応する音声を入力してディジタル化された音声データを出力する音声入力工程と、前記音声入力工程から出力された音声データを所定時間ごとに分析し、当該所定時間ごとの音声パワーと特徴データを算出する音声分析工程と、前記音声分析工程によって算出された音声パワーに基づいて当該発声された単語に対応する音声の有効音声区間を検出し、当該有効音声区間の時間的な長さが前記単語群ごとに設定された基準内か否かと、当該有効音声区間内の音声パワーの大きさによって、前記発声された単語に対応する音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を出力する音声区間検出・判定工程と、前記音声区間検出・判定工程において出力された、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を受けると、前記単語間の間に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報を出力する情報出力工程と、前記情報出力工程により情報が出力された後に、前記音声区間検出・判定工程において適正な状態で入力されたと判断された前記発声された単語に対応する音声を認識する音声認識工程と、を有することを特徴とする。

【００１５】

また、本発明の音声認識方法において、前記音声区間検出・判定工程が、有効音声区間の時間的な長さによって前記音声が適正な状態で入力されたか否かを判断する処理は、前記音声パワーが所定のしきい値より大きくなる時刻を始端とし、続いて、前記音声パワーが前記しきい値より小さくなる時刻を終端とし、当該終端の時刻から所定時間の間、前記音声パワーが前記しきい値より大きくならなかった場合に、前記始端から前記終端までの区間を有効音声区間として検出し、当該検出された有効音声区間が、所定の基準内である

場合に、前記音声が適正な状態で入力されたと判断することを特徴とする。

【００１６】

また、本発明の音声認識方法において、前記情報出力工程において出力される前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報は、信号音、光、音声メッセージ、表示画面上での表示の少なくとも１つであることを特徴とする。

【００１８】

また、本発明の音声認識装置は、単語間に間を有して断続的に発声された複数の単語からなる組を認識する音声認識装置であって、前記複数の単語は第１から第ｎ（ｎは正の整数）の単語群に属しており、前記第１から第ｎの単語群ごとに、前記複数の単語が属する全ての単語群に対して、時間的な長さの基準を設定する手段と、前記時間的な長さの基準を設定する手段によって時間的な長さの基準の設定が行われた後に、前記単語間に間を有して断続的に発声された前記単語に対応する音声を入力してディジタル化された音声データを出力する音声入力手段と、前記音声入力手段から出力された音声データを所定時間ごとに分析し、当該所定時間ごとの音声パワーと特徴データを算出する音声分析手段と、前記音声分析手段によって算出された音声パワーに基づいて当該発声された単語に対応する音声の有効音声区間を検出し、当該有効音声区間の時間的な長さが前記単語群ごとに設定された基準内か否かと、当該有効音声区間内の音声パワーの大きさによって、前記発声された単語に対応する音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を出力する音声区間検出・判定手段と、前記音声区間検出・判定手段で検出された前記音声に対する有効音声区間と、前記音声分析手段で算出された特徴データと、に基づいて前記音声を認識処理する音声認識処理手段と、前記音声区間検出・判定手段から出力された、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を受けると、前記単語間の間に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報を出力する情報出力手段と、を有し、前記音声認識処理手段は、前記情報出力手段により情報が出力された後に、前記音声区間検出・判定手段において適正な状態で入力されたと判断された前記発声された単語に対応する音声を認識処理することを特徴とする。

【００１９】

また、本発明の音声認識装置において、前記音声区間検出・判定手段が、有効音声区間の時間的な長さによって前記音声が適正な状態で入力されたか否かを判断する処理は、前記音声パワーが所定のしきい値より大きくなる時刻を始端とし、続いて、前記音声パワーが前記しきい値より小さくなる時刻を終端とし、当該終端の時刻から所定時間の間、前記音声パワーが前記しきい値より大きくならなかった場合に、前記始端から前記終端までの区間を有効音声区間として検出し、当該検出された有効音声区間が、所定の基準内である場合に、前記音声が適正な状態で入力されたと判断することを特徴とする。

【００２０】

また、本発明の音声認識装置において、前記情報出力手段が出力する前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報は、信号音、光、音声メッセージ、表示画面上での表示の少なくとも１つであることを特徴とする。

【００２２】

また、本発明の音声認識処理プログラムを記録した記録媒体は、単語間に間を有して断続的に発声された複数の単語からなる組を認識する音声認識装置における認識対象音声の入力状態報知プログラムを記録した記録媒体であって、前記複数の単語は第１から第ｎ（ｎは正の整数）の単語群に属しており、前記第１から第ｎの単語群ごとに、前記複数の単語が属する全ての単語群に対して、時間的な長さの基準を設定する工程と、前記時間的な長さの基準を設定する工程の後に、前記単語間に間を有して断続的に発声された前記単語に対応する音声を入力してディジタル化された音声データを出力する音声入力工程と、前記音声入力工程から出力された音声データを所定時間ごとに分析し、当該所定時間ごとの音声パワーと特徴データを算出する音声分析工程と、前記音声分析工程によって算出され

た音声パワーに基づいて当該発声された単語に対応する音声の有効音声区間を検出し、当該有効音声区間の時間的な長さが前記単語群ごとに設定された基準内か否かと、当該有効音声区間内の音声パワーの大きさによって、前記発声された単語に対応する音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を出力する音声区間検出・判定工程と、

前記音声区間検出・判定工程において出力された、前記発声された単語に対応する音声が適正な状態で入力されたことを示す信号を受けると、前記単語間の間に、前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報を出力する情報出力工程と、前記情報出力工程により情報が出力された後に、前記音声区間検出・判定工程において適正な状態で入力されたと判断された前記発声された単語に対応する音声を認識する音声認識工程と、を音声認識装置に実行させるための音声認識処理プログラムを記録した記録媒体である。

【００２３】

また、本発明の音声認識処理プログラムを記録した記録媒体において、前記音声区間検出・判定工程が、有効音声区間の時間的な長さによって前記音声が適正な状態で入力されたか否かを判断する処理は、前記音声パワーが所定のしきい値より大きくなる時刻を始端とし、続いて、前記音声パワーが前記しきい値より小さくなる時刻を終端とし、当該終端の時刻から所定時間の間、前記音声パワーが前記しきい値より大きくならなかった場合に、前記始端から前記終端までの区間を有効音声区間として検出し、当該検出された有効音声区間が、所定の基準内である場合に、前記音声が適正な状態で入力されたと判断することを特徴とする。

【００２４】

また、本発明の音声認識処理プログラムを記録した記録媒体において、前記情報出力工程において出力される前記発声された単語に対応する音声が適正な状態で入力されたことを示す情報は、信号音、光、音声メッセージ、表示画面上での表示の少なくとも１つであることを特徴とする

【００２６】

本発明は、ユーザの入力した認識対象音声が適正な状態で入力されたか否かを簡単な処理を行うだけで話者に報知することを可能とし、使い勝手の向上を図るものである。これを実現するために、ユーザの発話した認識対象音声における有効音声区間の時間的な長さと、当該有効音声区間内の音声パワーの大きさによって、前記認識対象音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合には、当該認識対象音声の入力直後に適正であることを示す情報を発するようにしている。これにより、ユーザは自分の発話した音声が適正な状態で入力されたか否かを簡単に知ることができ、音声の入力操作を行う際、ユーザに対し、自分の入力した音声が本当に適正な状態で入力されたのかどうかという不安感を与えることがなくなる。

【００２７】

また、このような適正な状態で入力されたか否かの判定対象となる認識対象音声は、複数の単語を１つの組として発話された音声であって、この１つの組を構成するそれぞれの単語に対するそれぞれの音声間に、各単語の区切りを示す間を有して発話された音声を対象としている。たとえば、現在時刻などの時刻設定を音声により設定可能な時計を例に取れば、「午前」・「何時」・「何分」というように複数の単語を１つの組とし、それを構成する各単語間に区切りとしての間をおきながら断続的に発話される音声を対象としている。

【００２８】

このように、複数の単語を１つの組として、各単語間に装置側から認識結果の応答なしに、ユーザの発話を一方的に入力する状況にあっては、各単語が果たして適正な状態（認識を行う上で適正な状態）で入力されたのかどうかがユーザにとって不安なもとなる。

【００２９】

これを解消するために、複数の単語を１つの組として発話される状況の場合は、それぞれの単語間に装置側から何らかの情報を発信することで、ユーザに安心感を与えることができる。

【００３０】

その情報としては、各単語間の区切り時間内に瞬時的に発せられる信号音（たとえば「ピッ」というような信号音）、発光ダイオード（ＬＥＤ）などにより瞬時的に発光する光、音声メッセージ（たとえば「はい」というようなごく短い音声メッセージ）、液晶ディスプレイ（ＬＣＤ）などの表示部を有する装置にあっては、ＬＣＤ上での「ＯＫ」などの簡単な表示）などが考えられる。ユーザは、自分の発話した１単語ごとの音声のあとに、このような簡単な情報が装置側から瞬時的に発せられることによって、自分の発話した音声が適正な状態で入力されたことがわかるので、音声入力操作に対する安心感が得られる。

【００３１】

また、１つの組を構成する複数の単語は、第１から第ｎ（ｎは正の整数）までの単語群に属し、前述した有効音声区間の時間的な長さを判定する基準は、それぞれの単語群ごとに設定するようにしている。これは、各単語群に属する単語の長さ（発話に必要な時間的長さ）が、単語群間で大きく異なる可能性があるからである。したがって、有効音声区間の時間的な長さを判定する基準を、それぞれの単語群ごとに設定しておくことによって、各単語群に属する単語に対し適正な有効音声区間の長さの判定が可能となる。

【００３２】

また、本発明の音声認識装置は、以上説明したような認識対象音声の入力状態報知方法を採用することにより、使い勝手をよくすることができ、この種の機器の取り扱いに不慣れなユーザでも容易に取り扱うことができるようになる。

【００３３】

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照しながら説明する。なお、この実施の形態では、音声認識技術を用いた機器として前述した音声時計を例に取り、この音声時計において、午前何時何分というような時刻合わせをする例について説明する。ここでは、具体例として、「午前６時３０分」を設定することを考える。

【００３４】

図１は本発明が適用される音声時計の外観構成を示すものであって、この音声時計は、液晶表示部など時刻の表示部分を持たない音声メッセージだけの音声時計であり、筐体１には音声出力手段としてのスピーカ２とユーザからの音声コマンドが入力されるマイクロホン３が設けられる。さらに、現在時刻合わせを行ったりアラーム時刻合わせを行ったりするときにモード設定を行うためのモード設定部４、現在時刻を知りたいときに押される時刻ボタン５などを少なくとも有している。これ以外にも、機能によっては様々な構成要素が設けられるが、本発明の要旨とは直接関係しない部分の図示およびその説明は省略する。

【００３５】

このような音声時計は、現在時刻が正確に合わせられていれば、時刻ボタン５がユーザによって押されると、その時点の現在時刻として、たとえば、「午前８時３０分」などと装置側から音声メッセージによる時刻が出力される。

【００３６】

図２はこのような音声時計に用いられる音声認識装置の構成を示すブロック図であり、音声入力部１１、音声分析部１２、音声区間検出・判定部１３、音声認識処理部１４、情報出力部１５、音声認識を行うための音声認識用モデルデータ１６、装置側から発せられる様々な情報（認識結果に対応する応答内容や、ユーザに問いかけを行う際の音声メッセージ内容、さらには、話者の発話した音声が適正であると判断されたときに出力される情報）を出力するための出力用データ１７を有している。

【００３７】

音声入力部１１は、前述したマイクロホン２や図示されていないアンプ、さらには、図示

されていないＡ／Ｄ変換器などを有し、ユーザによって発話された音声をマイクロホン２を通して入力し、増幅したのちＡ／Ｄ変換を行い、たとえば、８ＫＨｚ、１０bitのディジタル化された音声データとして出力する。

【００３８】

音声分析部１２は、音声入力部１１によって出力された音声データを、たとえば、２０msec程度（シフト量は１０msec程度）の短時間ごとに音声分析を行い、その短時間（２０msec程度）ごとに音声パワーと特徴データ（たとえば、１０次元ＬＰＣケプストラム）を算出する。

【００３９】

音声区間検出・判定部１３は、音声分析部１２で算出された音声パワーを用いて、有効な音声区間（有効音声区間という）を検出し、有効音声区間の時間的な長さが予め定めた所定の時間的長さ（Ｌ１，Ｌ２で表し、Ｌ１＜Ｌ２とする）の範囲内（Ｌ１よりも長く、Ｌ２よりも短い範囲内）で、音声パワーの最大値がある２つのしきい値（ｔｈ２，ｔｈ３で表し、ｔｈ２＜ｔｈ３とする）の範囲内（ｔｈ２よりも大きく、ｔｈ３よりも小さい範囲内）に入っている場合、その有効音声区間は正常な範囲内であると判定され、その音声は音声認識を行う上で適正な状態で入力されたと判断する。そして、適正であると判断すると前記有効音声区間の終端から一定時間（Ｌ４とする）後に、当該音声が適正な状態で入力されたことを示す信号を出力する。

【００４０】

なお、ここでいう有効音声区間というのは、ユーザの音声パワーが、あるしきい値（ｔｈ１とする）より大きくなった時点を音声区間の始端として求め、その後、音声パワーがしきい値ｔｈ１より小さくなり、かつ、しきい値ｔｈ１より小さくなった時点から予め定めた所定時間（Ｌ３とする）経過しても再びしきい値ｔｈ１より大きくならない場合、しきい値ｔｈ１より小さくなった時点を音声区間の終端として求め、その音声区間の始端から終端までを有効音声区間という。

【００４１】

また、音声認識処理部１４は、検出された有効音声区間内の特徴データ（音声分析部１２で得られた前述の音声特徴データのうち有効音声区間内の特徴データ）に基づき、音声認識用モデルデータ１６を用いて音声認識処理を行う。

【００４２】

情報出力部１５は、前述したように、認識結果に対応する応答音声や、ユーザに問いかける内容の音声メッセージを出力用データ１７を用いて作成して出力するものであるが、その他に、音声区間検出・判定部１３から出力された前記適正であることを示す信号を受け取ると、音声が適正に入力されたことを示す情報（この実施の形態では、「ピッ」という信号音とする）を出力する。

【００４３】

このような構成された音声時計において、何らかの時刻設定（ここでは現在時刻とする）を行う例について説明する。ここでは、設定する時刻としては、前述したように、午前６時３０分であるとする。このとき、音声時計（装置という）を現在時刻設定モードとする。

【００４４】

図３（ａ）はユーザの発話した「午前」・「６時」・「３０分」の音声波形であり、音声入力部１１によってＡ／Ｄ変換されたあとの出力であるとする。この図３（ａ）からもわかるように、ある時刻に時刻合わせをするような場合、たとえば、「午前６時３０分」という時刻合わせ内容を発話する際、認識率を高くするために、時刻合わせ内容を構成する各単語（「午前」・「６時」・「３０分」）を１単語づつ、各単語間に少しの間（ΔＴ１，ΔＴ２時間）をおいて発話するようにユーザ側に予め知らせておくとよい結果が得られる。

【００４５】

このように、この実施の形態において用いられる音声は、複数の単語を１つの組として発

話される音声であって、この１つの組を構成するそれぞれの単語に対するそれぞれの音声間に、各単語の区切りとしての間を有して断続的に発話される音声であるとする。

【００４６】

また、ここでは、現在時刻設定モードであり、このような現在時刻設定モード（アラーム時刻設定モードの場合も同様）のときは、最初に「午前」か「午後」を発話し、２番目に「何時」、３番目に「何分」というように、発話する順番は決められているものとする。また、説明の都合上、最初に発話される部分を第１の単語群に属する単語、２番目に発話される部分を第２の単語群に属する単語、３番目に発話される部分を第３の単語群に属する単語と呼ぶことにする。

【００４７】

図３（ａ）に示すような音声データに対し、音声分析部１２によって、たとえば２０msecごとのフレームに区切って音声分析を行い、各フレームごとに音声パワーと特徴データを求める。なお、特徴データは音声認識処理のときに用いられ、ここで行われるユーザの発話した音声が適正であるか否かの判定処理では、音声パワーを用いる。

【００４８】

図３（ｂ）は各フレームごとに求められた音声パワーを曲線で結んだ音声パワー曲線を示すものである。なお、図３（ａ）では、ユーザが「午前」・「６時」・「３０分」というように、１つの組を構成する単語を順番にすべて発話したときに得られる音声データであるが、本発明が行う処理は、ユーザが「午前」と発話すると、その「午前」の発話内容について、適正であるか否かを判定する処理を行い、適正であれば装置側から「ピッ」という信号音を出し、その後に、ユーザが「６時」と発話することにより、その「６時」の発話内容について、適正であるか否かを判定する処理を行い、適正であれば装置側から「ピッ」という信号音を出し、続いて、「３０分」と発話することにより、その「３０分」の発話内容について、適正であるか否かを判定する処理を行うというように、１つの組を構成するそれぞれの単語についてその単語が適正な状態で入力されたか否かの判定処理を行う。以下、それぞれの単語ごとの処理について詳細に説明する

また、前述の有効音声区間の時間的長さを判定するための基準となる時間的長さＬ１とＬ２は、実際には、前述した第１の単語群、第２の単語群、第３の単語群のそれぞれの単語群ごとに設定されるものである。

【００４９】

この単語群とは、この場合、ユーザの発話する内容は、「午前」、「何時」、「何分」という決められたパターンであるので、「午前」の部分に発話される単語群を第１の単語群といい、この第１の単語群に属する単語は、この場合、「午前」の他には「午後」がある。また、「何時」の部分に発話される単語群を第２の単語群といい、この第２の単語群に属する単語は、この場合、「０時」、「１時」、「２時」など時間の単位を表す単語である。また、「何分」の部分に発話される単語群を第３の単語群といい、この第３の単語群に属する単語は、この場合、「０分」、「１分」、「２分」など分の単位を表す単語である。しかも、これら第１から第３の単語群は、最初に第１の単語群（たとえば「午前」）を発話し、続いて第２の単語群（たとえば「６時」）を発話し、さらに続いて第３の単語群（たとえば「３０分」）を発話するというように、発話する順番はきまっていて、装置側では、その順番に従って、入力された各単語群に対する認識処理を行うようになっている。

【００５０】

したがって、前述のＬ１とＬ２は、第１から第３の単語群ごとに設定しておく方がよい結果が得られる。以下、第１の単語群に対して設定される時間的長さをＬ１１，Ｌ２１（Ｌ１１＜Ｌ２１）とし、第２の単語群に対して設定される時間的長さをＬ１２，Ｌ２２（Ｌ１２＜Ｌ２２）とし、第３の単語群に対して設定される時間的長さをＬ１３，Ｌ２３（Ｌ１３＜Ｌ２３）とする。

【００５１】

まず、ユーザが第１の単語群に属する単語として「午前」と発話すると、音声分析部１２

によって、前述したように、たとえば２０msecごとのフレームに区切って音声分析を行い、各フレームごとに音声パワーと特徴データを求める。

【００５２】

そして、ユーザの発話した「午前」に対して得られた音声パワー曲線から、「午前」に対する音声データの有効音声区間Ｔ１を求める。

【００５３】

まず、予め設定されたしきい値ｔｈ１を基準にして、ユーザが発話して得られた音声パワーが、最初にしきい値ｔｈ１を越えた時刻を、「午前」の音声区間の始端とする。この図３（ｂ）からもわかるように、時刻ｔ１でしきい値ｔｈ１を越えているので、この時刻ｔ１を「午前」に対する音声区間の始端とする（始端ｔ１という）。

【００５４】

続いて、「午前」に対する音声パワーが、しきい値ｔｈ１より小さくなる時刻を調べ、その時刻がｔ２であるとする。そして、この時刻ｔ２から予め定められたある一定時間Ｌ３が経過したのちにも音声パワーがしきい値ｔｈ１を再び越えなければ、「午前」に対する音声区間ｔ１の終端は、時刻ｔ２であるとする。この時刻ｔ２を音声区間ｔ１の終端とする（終端ｔ２という）。

【００５５】

そして、このように求められた始端ｔ１と終端ｔ２の間の区間を有効音声区間Ｔ１とし、ｔ１を有効音声区間Ｔ１の始端とし、ｔ２を有効音声区間Ｔ１の終端とする。なお、前記一定時間Ｌ３はごく短い時間が設定され、具体的には、隣接する単語間に存在する単語間の区切りとしての間の時間ΔＴ１，ΔＴ２よりもきわめて短い時間である。

【００５６】

このようにして、第１の単語群に属する「午前」の音声に対する有効音声区間Ｔ１が求められる。次に、この有効音声区間Ｔ１の時間的長さと音声パワーが、予め設定された範囲内にあるか否かを判断する。

【００５７】

つまり、前述したように、有効音声区間Ｔ１の時間的長さ（時間的長さもＴ１で表す）が予め設定された時間的長さＬ１１，Ｌ２１に対し、Ｌ１１＜Ｔ１＜Ｌ２１であって、その有効音声区間Ｔ１内の音声パワーの最大値ｍ１が予め設定されたしきい値ｔｈ２，ｔｈ３に対し、ｔｈ２＜ｍ１＜ｔｈ３を満たす場合、抽出された有効音声区間Ｔ１は正常な範囲内にあると判断され、ユーザの発話した「午前」は音声認識を行う上で適正な状態で入力されたと判断する。

【００５８】

このようにして、第１の単語群に属する「午前」に対する有効音声区間Ｔ１が所定の範囲内（ Ｌ１１＜Ｔ１＜Ｌ２１で、かつ、 ｔｈ２＜ｍ１＜ｔｈ３）であると判断されると、音声区間検出・判定部１３は、図３（ｃ）に示すように、その有効音声区間Ｔ１の終端ｔ２からＬ４時間経過後に、情報出力部１５に対し、当該音声（この場合「午前」）が適正な状態で入力された音声であることを示す信号ｓ１を出力する。

【００５９】

情報出力部１５はこの信号ｓ１を受けると、図３（ｄ）に示すように、予め決められた瞬時的な情報として信号音を出力する。この信号音は、ユーザの発話した音声が認識を行うに適正な状態で入力された音声であることをユーザに対して報知するもので、種々の報知手段が考えられるが、この実施の形態では、「ピッ」という瞬時的な信号音を出力する。

【００６０】

すなわち、ユーザが「午前」と発話してそれが適正であると判断されると、ユーザの発話した「午前」の後につづいて「ピッ」という信号音が装置側から発せられる。これにより、ユーザは自分の発話した「午前」という音声が適正な状態で装置側に入力されたということがわかる。

【００６１】

つづいて、ユーザが第２の単語群に属する「６時」と発話すると、その音声データに対す

る音声パワー曲線から、「６時」に対する音声データの有効音声区間Ｔ２を求める。
【００６２】
まず、予め設定されたしきい値ｔｈ１を基準にして、ユーザが発話して得られた音声パワーが、最初に、このしきい値ｔｈ１を越えた時刻を、「６時」の音声区間の始点とする。図３（ｂ）からもわかるように、時刻ｔ３でしきい値ｔｈ１を越えているので、この時刻ｔ３を「６前」に対する音声区間の始端とする（始端ｔ３という）。
【００６３】
続いて、「６時」という単語に対する音声パワーが、しきい値ｔｈ１より小さくなる時刻を調べ、その時刻がｔ４であるとする。そして、この時刻ｔ４から予め定められたある一定時間Ｌ３が経過したのちにも音声パワーがしきい値ｔｈ１を再び越えなければ、「６時」に対する音声区間Ｔ２の終端は、時刻ｔ４であるとする。この時刻ｔ４を音声区間の終端とする（終端ｔ４という）。
【００６４】
このように求められた始端ｔ３と終端ｔ４の間の区間を有効音声区間Ｔ２とし、ｔ３を有効音声区間Ｔ２の始端とし、ｔ４を有効音声区間Ｔ２の終端とする。次に、この有効音声区間Ｔ２の時間的長さが、予め設定された範囲内にあるか否かを判断する。この場合、Ｌ１２＜Ｔ２＜Ｌ２２か否かを判断する。また、その音声区間Ｔ２内の音声パワーの最大値ｍ２がしきい値ｔｈ２，ｔｈ３の範囲内に入っているか否かを判断する。そして、これらの条件が成立すると、抽出された有効音声区間Ｔ２は正常な範囲内にあると判断され、その音声（この場合、「６時」）は音声認識処理を行う上で適正に入力されたと判断する。
【００６５】
このようにして、第２の単語群に属する「６時」に対する有効音声区間Ｔ２が正常な範囲内であると判断されると、音声区間検出・判定部１３は、図３（ｃ）に示すように、その有効音声区間Ｔ２の終端ｔ４からＬ４時間経過後に、情報出力部１５に対し、当該有効音声区間Ｔ２が正常な範囲内であったことを示す信号ｓ２を出力する。
【００６６】
情報出力部１５はこの信号ｓ２を受けると、図３（ｄ）に示すように、適正であることを示す情報として、前述したように、「ピッ」という瞬時的な信号音を出力する。
【００６７】
すなわち、ユーザが「午前」に続いて「６時」と発話し、それが適正であると判断されると、「６時」の後につづいて「ピッ」という信号音が装置側から発せられる。これにより、ユーザは自分の発話した「６時」という音声が装置側に適正な状態で入力されたということがわかる。
【００６８】
続いて、ユーザが第３の単語群に属する「３０分」と発話すると、その音声データに対する音声パワー曲線から、「３０分」に対する音声データの有効音声区間Ｔ３を求める。
【００６９】
まず、図３（ｂ）からもわかるように、時刻ｔ５でしきい値ｔｈ１を越えているので、この時刻ｔ５を「３０分前」に対する音声区間Ｔ３の始端とする（始端ｔ５という）。そして、「３０分」という単語に対する音声パワーが、しきい値ｔｈ１より小さくなる時刻を調べ、その時刻がｔ６であるとする。この時刻ｔ６から予め定められたある一定時間Ｌ３が経過したのちにも音声パワーがしきい値ｔｈ１を再び越えなければ、「３０分」に対する音声区間の終端は、時刻ｔ６であるとする。この時刻ｔ６を音声区間の終端とする（終端ｔ６という）。
【００７０】
このように求められた始端ｔ５と終端ｔ６の間の区間を有効音声区間Ｔ３とし、ｔ５を有効音声区間Ｔ３の始端とし、ｔ６を有効音声区間Ｔ３の終端とする。次に、この有効音声区間Ｔ３の時間的長さが、予め設定された範囲内にあるか否かを判断する。この場合、Ｌ１３＜Ｔ３＜Ｌ２３か否かを判断する。また、その有効音声区間Ｔ３内の音声パワーの最大値ｍ３がしきい値ｔｈ２，ｔｈ３の範囲内に入っているか否かを判断する。そして、こ

れらの条件が成立すると、抽出された有効音声区間Ｔ３は正常な範囲内にあると判断され、その音声（この場合、「３０分」）が適正な状態で入力されたと判断する。

【００７１】

このようにして、第３の単語群に属する「３０分」に対する有効音声区間Ｔ３が正常な範囲内であると判断されると、音声区間検出・判定部１３は、図３（ｃ）に示すように、その有効音声区間Ｔ３の終端ｔ６からＬ４時間経過後に、情報出力部１５に対し、当該有効音声区間Ｔ３が正常な範囲内であったことを示す信号ｓ３を出力する。

【００７２】

情報出力部１５はこの信号ｓ３を受けると、適正であることを示す情報として、前述したように、「ピッ」という瞬時的な信号音を出力する。

【００７３】

すなわち、ユーザが「午前」、「６時」に続いて「３０分」と発話し、それが適正な状態で入力されたと判断されると、「３０分」の後につづいて「ピッ」という信号音が装置側から発せられる。これにより、ユーザは自分の発話した「３０分」が装置側に適正な状態で入力されたと判断できる。

【００７４】

このように、ユーザが、「午前」、「６時」、「３０分」と発話した場合、それぞれの単語に対する音声が適正な状態で入力されたと判断されると、「午前」、「ピッ」、「６時」、「ピッ」、「３０分」、「ピッ」というように、ユーザの発話した音声のあとに装置側から「ピッ」が応答されるので、ユーザはそれを聞くことにより自分の発話した音声が適正な状態で入力されたことがわかり、安心感が得られる。

【００７５】

なお、前述のＬ３は第１から第３の単語群において共通の時間として説明したが、これもＬ１，Ｌ２と同様に、それぞれの単語群ごとに適当な時間を設定するようにしてもよい。

【００７６】

また、ある有効音声区間（有効音声区間Ｔ１とする）が前述した条件を満たさない例として、図４（ａ）（ｂ），（ｃ）（ｄ）がある。図４（ａ）（ｂ）は有効音声区間Ｔ１の２つの最大値ｍ１がしきい値ｔｈ３を越え、しかも、有効音声区間Ｔ１の時間的長さが、Ｌ１１よりも短い場合である。また、図４（ｃ）（ｄ）は有効音声区間Ｔ１の最大値ｍ１がしきい値ｔｈ２より小さく、かつ、有効音声区間Ｔ１の時間的長さが、Ｌ２１よりも長い場合である。

【００７７】

図４（ａ）（ｂ）の例は、ユーザの発話した音声が強すぎ、しかも、きわめて早口で発話したような例であり、このような状態で発話された音声に対しては適正な音声認識が行えない可能性が高いことから、その入力音声は適正でないとする。

【００７８】

また、図４（ｃ）（ｄ）の例は、ユーザの発話した音声が小さすぎ、しかも、きわめて間延びした状態で発話したような例であり、このような状態で発話された音声に対しては、適正な音声認識が行えない可能性が高いことから、その入力音声は適正でないとする。

【００７９】

なお、図４（ａ）（ｂ），（ｃ）（ｄ）の例は、共に、音声パワーと有効音声区間の時間的長さの両方が条件を満たさない例であるが、音声パワーと有効音声区間の時間的長さのいずれか一方が条件を満たさない場合でも、その入力音声は適正でないと判断される。

【００８０】

このように、ユーザの発話した音声が適正な状態で入力されなかったと判断した場合には、「ピッ」という信号音は発しないようにする。これにより、ユーザは自分の発話した音声が適正ではないということを知ることができる。この場合、装置側からの反応が無音であることから、ユーザは、再度、入力し直すというようなことを行う。あるいは、適正な状態で入力がなされなかった場合には、再度、入力し直すことを促す音声メッセージやその他の信号などでユーザに報知することもできる。

【００８１】
以上説明したように、この実施の形態は、装置側に対し、「午前」、「６時」、「３０分」というような複数の単語から構成される内容を、１単語発話するごとに少し間をおいて次の単語を入力するという断続的な発話によって音声の入力を行うことで、この入力音声を認識させて、それに対応する動作（現在時刻設定など）を行わせるような場合を例にしている。
【００８２】
この場合、ユーザが、まず、「午前」と発話すると、その音声データから得られた有効音声区間の時間的長さと音声パワーとに基づいて、その音声が適正な状態で入力されたか否かを判定し、適正な音声であると判定した場合には、適正であることを示す「ピッ」というような瞬時的な信号音を発するようにしている。これにより、ユーザは、自分が発した音声が正常な状態で入力されたか否かを装置側からの信号音で知ることができ、音声入力操作を不安感を抱くことなく行うことができる。
【００８３】
このように、本発明は、装置との間で対話形式で音声を入力する方式でなく、１つの組に存在する複数の単語を断続的に続けて入力するような場合に特に有効なものとなる。
【００８４】
なお、本発明は以上説明した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。
【００８５】
たとえば、前述の実施の形態では、有効音声区間が適正か否かの判断条件の一つとして、音声パワーの最大値を用いているが、有効音声区間内の音声パワーの平均値を用いて有効音声区間が適正か否かの判断を行っても良い。
【００８６】
また、前述の実施の形態では、ユーザの発話した音声が適正であることを示す情報として「ピッ」というような信号音を用いたが、このような信号音に限られるものではなく、たとえば、発光ダイオードなどを点灯させるようにしてもよく、また、「はい」というような短い音声による応答でであってもよい。さらには、液晶ディスプレイのような表示部のあるものであれば、話者の音声入力直後にその音声が適正である場合、その表示部にたとえば「ＯＫ」というような表示を行うようにしてもよい。
【００８７】
また、前述の実施の形態では、音声認識技術を用いた時計における時刻合わせを例にとって説明したが、本発明は、時計以外の機器にも適用できることは勿論である。
【００８８】
また、前述の実施の形態では、３つの単語を１組として入力した例を示したが、１組を構成する単語数は３つに限られるものでないことは勿論である。
【００８９】
また、以上説明した本発明の認識対象音声の入力状態報知処理を行う処理プログラムは、フロッピィディスク、光ディスク、ハードディスクなどの記録媒体に記録させておくことができ、本発明はその記録媒体をも含むものである。また、ネットワークから処理プログラムを得るようにしてもよい。
【００９０】
以上説明したように本発明によれば、話者の発話した認識対象音声における有効音声区間の時間的な長さと、当該有効音声区間内の音声パワーの大きさによって、前記認識対象音声が適正な状態で入力されたか否かを判断し、適正であると判断した場合には、当該認識対象音声の入力直後に適正であることを示す情報を発するようにしている。これにより、ユーザが装置に対して音声の入力操作を行う際、自分の入力した音声が本当に適正な状態で入力されたのかどうかという不安感を抱くことがなくなり、音声入力を行う際の操作性の向上を図ることができる。
【００９１】

特に、認識対象音声が、複数の単語を１つの組として発話される音声であって、このような音声を入力する際に効果が得られる。たとえば、現在時刻などの時刻設定を音声により設定可能な時計を例に取れば、「午前」・「何時」・「何分」というように複数の単語を１つの組とし、それを構成する各単語間に区切りとしての間をおきながら断続的に発話されるような場合である。このように、複数の単語を１つの組としてそれぞれの単語が断続的に発話される場合、各単語を発話したあとに、装置側から瞬時的に発せられる信号音などが返ってくるようにすることで、ユーザは、自分の発話した音声が適正な状態で入力されたことが即座にわかり、音声入力操作に対する安心感が得られる。

【００９２】

また、ユーザの発話した音声が適正な状態で入力されことを示す情報として、瞬時的な情報を発するだけであるので、たとえば、それぞれの単語を認識してその認識結果をそのまま応答するのに比べると、処理を軽いものとすることができ、処理時間も大幅に短縮することができる。

【００９３】

そして、このような認識対象音声の入力状態報知方法を採用した音声認識装置は、使い勝手のよいものとなり、この種の機器の取り扱いに不慣れなユーザでも容易に取り扱うことができるようになり、また、全体的な処理を軽いものとすることができるので、ＣＰＵやメモリに低コストなものが使用でき、装置そのものの価格も低コスト化が図れる。

【図面の簡単な説明】

【図１】本発明の実施に形態に用いられる音声時計の外観を概略的に示す図。

【図２】図１で示した音声時計に用いられる音声認識装置部分の概略構成を説明するブロック図である。

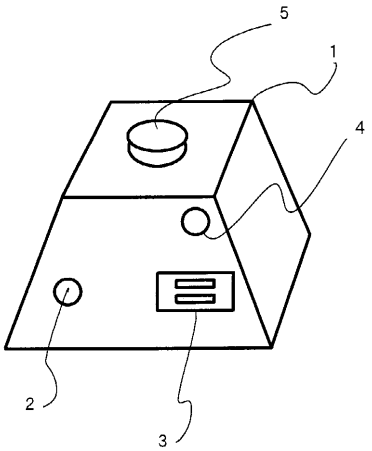
【図３】図１で示した音声時計にて時刻設定を行う際の認識対象音声の入力状態報知処理を説明するタイムチャート。

【図４】本発明の実施の形態における入力音声が適正でない例を説明する図。

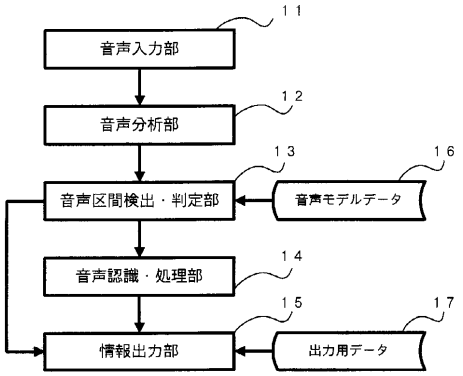
【符号の説明】

１　音声時計の筐体
２　スピーカ
３　マイクロホン
４　モード設定部
５　時刻ボタン
１１　音声入力部
１２　音声分析部
１３　音声区間検出・判定部
１４　音声認識処理部
１５　情報出力部
１６　音声認識用モデルデータ
１７　出力用モデルデータ

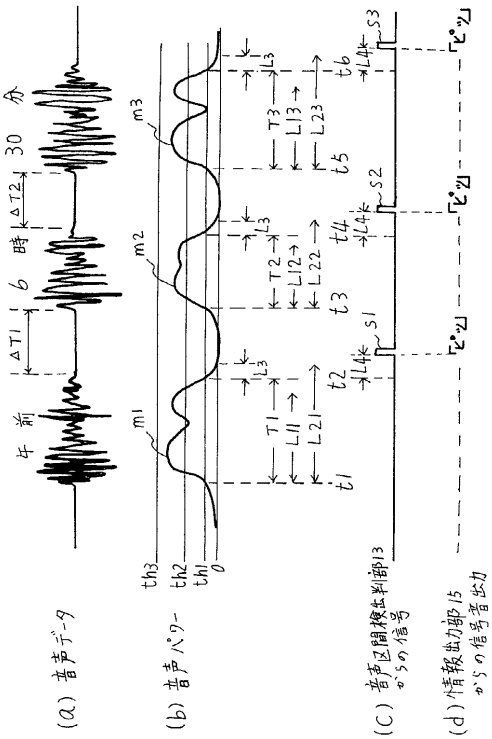
【図１】


5
1
4
2
3

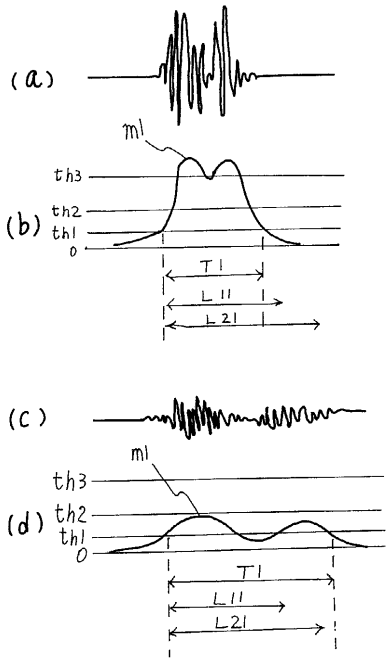
【図２】


11
音声入力部
12
音声分析部
13
音声区間検出・判定部
16
音声モデルデータ
14
音声認識・処理部
15
情報出力部
17
出力用データ

【図３】


午前
6時
30分
ΔT1
ΔT2
m1
m2
m3
th3
th2
th1
0
T1
T2
T3
L11
L12
L13
L21
L22
L23
L3
L4
t1
t2
t3
t4
t5
t6
S1
S2
S3
(a) 音声データ
(b) 音声パワー
(c) 音声区間検出判断13からの信号
(d) 情報出力部15からの信号出力

【図４】


(a)
m1
th3
th2
th1
0
(b)
T1
L11
L21
(c)
m1
th3
th2
th1
0
(d)
T1
L11
L21

フロントページの続き

(56)参考文献 特開昭５９－０８８８００（ＪＰ，Ａ）
特開昭６３－０１４２００（ＪＰ，Ａ）
特開平０５－０８０９７１（ＪＰ，Ａ）
特開平１０－０６９２９３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L 11/02
G10L 15/04