

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3720595号
(P3720595)

(45) 発行日 平成17年11月30日(2005.11.30)

(24) 登録日 平成17年9月16日(2005.9.16)

(51) Int. Cl.⁷

F I

G 1 0 L 15/04
G 1 0 L 15/10
G 1 0 L 15/28

G 1 0 L 3/00 5 1 5 C
G 1 0 L 3/00 5 3 1 M
G 1 0 L 3/00 5 7 1 H

請求項の数 15 (全 10 頁)

(21) 出願番号	特願平10-263552	(73) 特許権者	000001007
(22) 出願日	平成10年9月17日(1998.9.17)		キヤノン株式会社
(65) 公開番号	特開2000-99070(P2000-99070A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成12年4月7日(2000.4.7)	(74) 代理人	100076428
審査請求日	平成14年9月4日(2002.9.4)		弁理士 大塚 康德
		(74) 代理人	100093908
			弁理士 松本 研一
		(74) 代理人	100101306
			弁理士 丸山 幸雄
		(72) 発明者	山田 雅章
			東京都大田区下丸子3丁目30番2号
			キヤノン株式会社内
		(72) 発明者	小森 康弘
			東京都大田区下丸子3丁目30番2号
			キヤノン株式会社内

最終頁に続く

(54) 【発明の名称】 音声認識装置及びその方法、コンピュータ可読メモリ

(57) 【特許請求の範囲】

【請求項1】

入力された音声を認識する音声認識装置であって、
前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する判定手段と

、
前記判定手段により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する報知手段と

を備えることを特徴とする音声認識装置。

【請求項2】

前記入力された音声の単音節系列に対する第1スコアを計算する計算手段と、

前記入力された音声全体を予め登録されている音声認識用パターンと比較することにより音声認識を行い、音声認識結果とそのスコアを第2スコアとして計算する認識手段とを更に備え、

前記判定手段は、前記第1スコアと前記第2スコアの大きさを比較し、前記第2スコアが前記第1スコアより大きいか否かに基づいて、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する

ことを特徴とする請求項1に記載の音声認識装置。

【請求項3】

前記判定手段は、前記第1スコアに所定数を乗じたスコア又は前記第1スコアから所定

数を減じたスコアと前記第 2 スコアの大きさを比較し、前記第 2 スコアが前記第 1 スコアより大きいか否かに基づいて、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する

ことを特徴とする請求項 2 に記載の音声認識装置。

【請求項 4】

前記計算手段は、任意の単音節と無音との組を任意数組み合わせさせた単音節連鎖を用いて、前記入力された音声の単音節系列に対する第 1 スコアを計算する

ことを特徴とする請求項 2 に記載の音声認識装置。

【請求項 5】

前記計算手段は、無音を挟んだ 2 音節連鎖を用いて、前記入力された音声の単音節系列に対する第 1 スコアを計算する

ことを特徴とする請求項 2 に記載の音声認識装置。

【請求項 6】

前記報知手段は、前記判定手段により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、表示装置を用いて、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する

ことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 7】

前記報知手段は、前記判定手段により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、音声を用いて、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する

ことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 8】

入力された音声を認識する音声認識方法であって、

前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する判定工程と

、
前記判定工程により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する報知工程と

を備えることを特徴とする音声認識方法。

【請求項 9】

前記入力された音声の単音節系列に対する第 1 スコアを計算する計算工程と、

前記入力された音声全体を予め登録されている音声認識用パターンと比較することにより音声認識を行い、音声認識結果とそのスコアを第 2 スコアとして計算する認識工程とを更に備え、

前記判定工程は、前記第 1 スコアと前記第 2 スコアの大きさを比較し、前記第 2 スコアが前記第 1 スコアより大きいか否かに基づいて、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する

ことを特徴とする請求項 8 に記載の音声認識方法。

【請求項 10】

前記判定工程は、前記第 1 スコアに所定数を乗じたスコア又は前記第 1 スコアから所定数を減じたスコアと前記第 2 のスコアの大きさを比較し、前記第 2 スコアが前記第 1 スコアより大きいか否かに基づいて、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する

ことを特徴とする請求項 9 に記載の音声認識方法。

【請求項 11】

前記計算工程は、任意の単音節と無音との組を任意数組み合わせさせた単音節連鎖を用いて、前記入力された音声の単音節系列に対する第 1 スコアを計算する

ことを特徴する請求項 9 に記載の音声認識方法。

【請求項 12】

10

20

30

40

50

前記計算工程は、無音を挟んだ2音節連鎖を用いて、前記入力された音声の単音節系列に対する第1スコアを計算する

ことを特徴とする請求項9に記載の音声認識方法。

【請求項13】

前記報知工程は、前記判定工程により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、表示装置に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を表示する

ことを特徴とする請求項8に記載の音声認識方法。

【請求項14】

前記報知工程は、前記判定工程により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、音声で、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する

ことを特徴とする請求項8に記載の音声認識方法。

【請求項15】

入力された音声を認識する音声認識のプログラムコードが格納されるコンピュータ可読メモリであって、

前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する判定工程のプログラムコードと、

前記判定工程により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する報知工程のプログラムコードと

を備えることを特徴とするコンピュータ可読メモリ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、入力された音声を認識する音声認識装置及びその方法、コンピュータ可読メモリに関するものである。

【0002】

【従来の技術】

音声認識技術では、あらかじめ標準音声の標準パターン（単語や文）を何らかの形式で登録し、その登録された標準パターンと入力された音声を比較することで音声認識を行う。登録の形式としては、例えば、音素表現や生成文法を利用した形式などがある。そして、音声認識を行う際には、入力された音声と標準パターンとを比較することによって得られる一致度を示すスコアを定め、スコアの最も良い標準パターンを音声認識結果とする。

【0003】

音声認識対象の音声入力の方法としては、ユーザが単音節に区切った音声を発声して入力する方法がある。例えば、「カナガワ」と入力する場合は、「カ、ナ、ガ、ワ」と音節毎に区切った音声を発声して入力する。この入力を単音節発声と呼ぶことにする。

【0004】

この単音節発声で入力された音声に対する音声認識では、従来、以下の2つの方法が用いられている。

【0005】

1. 入力された音声のうち、無音と思われる部分を取り除いた音声に対して音声認識を行う。

【0006】

2. 単音節発声のパターンも音声認識対象として登録し、各パターンの音声認識を含む音声認識を行う。

【0007】

1の方法は、入力された音声のうち、無音と思われる部分を取り除き、残った音声を接続して得られる音声に対し音声認識を行う（図7）。

10

20

30

40

50

【0008】

2の方法は、入力された音声「カナガワ」である場合、「カナガワ」だけでなく、「カ（無音）ナ（無音）ガ（無音）ワ」というパターンも標準パターンとして登録する。そして、入力された音声と「カ（無音）ナ（無音）ガ（無音）ワ」と登録された標準パターンとのスコアが最も良かった場合には、「カナガワ」を音声認識結果として採用するというものである。

【0009】

【発明が解決しようとする課題】

しかしながら、上記従来音声認識では、以下の問題点があった。

【0010】

まず、1の方法では、無音かどうかの判定を誤ると認識結果に悪影響を与えるという問題点があった。一方、無音かどうかを精度良く判定するためには、音声認識と同様の処理が必要となり、この場合には、以下に述べる2の方法における問題点と同様の問題点が生じていた。

【0011】

2の方法では、登録しておく各標準パターンについて、単音節発声で音声が入力された場合とそうでない場合の2通りの標準パターンを用意しておかなければならず、処理量の増加を招く。また、一般に、語頭の環境（無音の直後）では認識率が悪い場合が多く、単音節発声では全ての音節が語頭の環境となるため、認識結果の信頼性が低下する。さらに、次のような問題がある。多くの場合、音声認識は発声の開始点と終了点を自動的に検出する音声切出処理と併用されて実行される。そして、単音節発声の場合、音節と音節との間に無音があるため、単語中に挿入された無音を発声終了と誤って認識する音声切出の誤りが生じやすい。この音声切出の誤りが生じた場合、切り出された音声に対して音声認識を行っても正確な音声認識結果を得られる確率は低いものであった。

【0012】

本発明は上記の問題点に鑑みてなされたものであり、音声を効率良く、かつ精度良く認識することができる音声認識装置及びその方法、コンピュータ可読メモリを提供することを目的とする。

【0013】

【課題を解決するための手段】

上記の目的を達成するための本発明による音声認識装置は以下の構成を備える。即ち、入力された音声を認識する音声認識装置であって、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する判定手段と

、
前記判定手段により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する報知手段と

を備える。

【0014】

また、好ましくは、前記入力された音声の単音節系列に対する第1スコアを計算する計算手段と、

前記入力された音声全体を予め登録されている音声認識用パターンと比較することにより音声認識を行い、音声認識結果とそのスコアを第2スコアとして計算する認識手段とを更に備え、

前記判定手段は、前記第1スコアと前記第2スコアの大きさを比較し、前記第2スコアが前記第1スコアより大きいか否かに基づいて、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する。

【0016】

また、好ましくは、前記判定手段は、前記第1スコアに所定数を乗じたスコアと前記第2スコアを比較し、その比較結果に基づいて、前記入力された音声に単音節に区切られた音

10

20

30

40

50

声が含まれるか否かを判定する。

【0017】

また、好ましくは、前記判定手段は、前記第1スコアに所定数を減じたスコアと前記第2スコアを比較し、その比較結果に基づいて、前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する。

【0018】

また、好ましくは、前記計算手段は、任意の単音節と無音との組を任意数組み合わせさせた単音節連鎖を用いて、前記入力された音声の単音節系列に対する第1スコアを計算する。

【0019】

また、好ましくは、前記計算手段は、無音を挟んだ2音節連鎖を用いて、前記入力された音声の単音節系列に対する第1スコアを計算する。 10

【0020】

また、好ましくは、前記報知手段は、表示装置を用いてユーザに音声の再入力を促す旨を報知する。

【0021】

また、好ましくは、前記報知手段は、音声を用いてユーザに音声の再入力を促す旨を報知する。

【0022】

上記の目的を達成するための本発明による音声認識方法は以下の構成を備える。即ち、
入力された音声を認識する音声認識方法であって、
前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する判定工程と 20

、
前記判定工程により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する報知工程と
を備える。

【0023】

上記の目的を達成するための本発明によるコンピュータ可読メモリは以下の構成を備える。即ち、

入力された音声を認識する音声認識のプログラムコードが格納されたコンピュータ可読メモリであって、 30

前記入力された音声に単音節に区切られた音声が含まれるか否かを判定する判定工程のプログラムコードと、

前記判定工程により、前記入力された音声に単音節に区切られた音声が含まれると判定された場合に、ユーザに、該入力された音声を連続発声で再入力させることを促す旨を報知する報知工程のプログラムコードと

を備える。

【0024】

【発明の実施の形態】

以下、図面を参照して本発明の好適な実施形態を詳細に説明する。 40

【0025】

図1は本発明の実施形態の音声認識装置のハードウェア構成を示す図である。H1は数値演算・制御等の処理を行う中央処理装置であり、本実施形態で実行される処理手順にしたがって演算を行う。H2はRAM・ROM等の記憶装置であり、本実施形態で実行される処理手順や処理に必要な一時データが格納される。H3はハードディスク装置等の外部記憶装置であり、音声認識用の標準パターン(テンプレート)や音韻モデル等が格納される。

【0026】

H4はディスプレイ等の出力装置であり、音声認識結果、あるいは音声認識結果を処理することによって得られた情報を出力する。但し、本実施形態では、他の装置・プログラム 50

の一部として組み込まれることも可能であり、この場合には出力は他の装置・プログラムの入力に接続されるものとなる。H5はマイクロホン等の入力装置であり、ユーザが発声した音声を入力する。但し、本実施形態では、他の装置・プログラムの一部として組み込まれることも可能であり、この場合には入力は他の装置・プログラムを通じて間接的に行われることになる。H6はバスであり、上記中央処理装置H1～入力装置H5の各種装置を相互に接続する。

【0027】

以上のハードウェア構成を踏まえて、本実施形態で実行される処理について、図2を用いて説明する。

【0028】

図2は本発明の実施形態で実行される処理を示すフローチャートである。

【0029】

まず、ステップS1において、ユーザが発声した音声を入力する。次に、ステップS2において、ステップS1で入力された音声に対してLPC分析やケプストラム分析等の信号処理を行い、音声認識の際の標準パターンと入力された音声とを比較するスコア計算に用いるパラメータを計算する。

【0030】

次に、ステップS3において、ステップS2で計算されたパラメータの単音節系列に対するスコア s_0 を計算する。単音節系列に対するスコアは、任意の単音節系列を認識候補として入力された音声と比較する音声認識アルゴリズムを適用し、その結果によって得られた第1位候補に対するスコアを単音節系列に対するスコアとする。ここで、任意の単音節系列を認識対象とした場合の認識対象の表現方法の一例について、図3を用いて説明する。

【0031】

図3は本発明の実施形態の認識対象の表現方法の一例を示す図である。

【0032】

図3中、ア～は対応する各音節の発声を受理し、「無音」は適当な時間長の無音を受理する。これらの受理を任意数実行して得られる単音節と無音の組を組み合わせた単音節連鎖が認識対象となる。は図を見やすくするために設けられたNULLノードであり、入力無しで次の状態に遷移する。

【0033】

次に、ステップS4において、ステップS2で計算されたパラメータを用いて音声認識を行う。音声認識の結果、1～複数個の認識結果文字列 $r_i (i = 1, 2, \dots)$ 及びそのスコア s_i が求められる。その例を図4に示す。

【0034】

次に、ステップS5において、ステップS3で計算されたスコア s_0 と、ステップS4で計算された第1位候補の音声認識結果のスコア s_1 とを比較する。比較の結果、 s_1 (認識スコア) $>$ s_0 (単音節スコア) (スコアが大きいほど認識結果としてもっともらしいものとする)の場合(ステップS5でYES)、ステップS6に進み、ステップS4で得られた音声認識結果を出力し処理を終了する。一方、 s_1 (認識スコア) $>$ s_0 (単音節スコア)でない場合(ステップS5でNO)、ステップS7に進み、単音節発声が発見されたことをユーザに知らせるメッセージ、あるいは図5に示すような連続発声を促す等のメッセージを出力装置H4に出力し、ステップS1に戻る。

【0035】

以上説明したように、本実施形態によれば、計算量をあまり増加させることなくユーザの単音節発声に対応することが可能となる。また、単音節発声か否かという情報のみを用いることで、単音節発声の内容の認識結果(何という音節系列が発声されたかという音声認識結果)には依存しないので、単音節に対する音声認識率に影響されない。また、単音節発声の途中で音声切出部が発声終了を誤検出し発声が途中で区切られたとしても、正常に動作することができる。また、最終的な音声認識の入力として単音節発声による入力を排

10

20

30

40

50

除することが可能となり、良好な認識率を確保することが可能となる。

【0036】

上記実施形態では、ステップS3及びステップS4を明示的に処理がわかれているものとして説明したが、ステップS3及びステップS4が1つのステップで実行することも可能である。

【0037】

即ち、図6に示すような認識文法を用いて音声認識を行い、第1位候補の音声認識結果が単音節系列であった場合に、ユーザから入力された音声が単音節発声であると見做すことにより、上記実施形態と等価な処理を行うことが可能である。この場合、ステップS3及びステップS5は、ステップS4に含まれることになる。

10

【0038】

また、単音節系列に対するスコアを計算する際に、任意の単音節と無音の組を組み合わせた単音節連鎖を用いるものとしたが、ユーザが発声すると予想される単音節系列のみを用いても良い。例えば、認識対象語に含まれる2音節連鎖の組のみを用いてもよい。

【0039】

また、単音節系列に対するスコアs0に対してペナルティを課した後、音声認識結果のスコアs1と比較しても良い。例えば、正数となるスコアを用いた場合、乗数(0<<1)を用いてs0とs1を比較しても良い。あるいは、(>0)を用いてs0とs1を比較してもよい。

【0040】

20

また、ステップS7では、出力装置H4にヘルプメッセージを表示することで、音声の再入力をユーザに促すようにしていたが、出力装置H4に音声出力装置を用いて、音声によるヘルプメッセージを出力することで、音声の再入力をユーザに促すようにしても良い。

【0041】

尚、本発明は、複数の機器(例えばホストコンピュータ、インタフェース機器、リーダ、プリンタなど)から構成されるシステムに適用しても、一つの機器からなる装置(例えば、複写機、ファクシミリ装置など)に適用してもよい。

【0042】

また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ(またはCPUやMPU)が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

30

【0043】

この場合、記憶媒体から読出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【0044】

プログラムコードを供給するための記憶媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

40

【0045】

また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOS(オペレーティングシステム)などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0046】

更に、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるC

50

PUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0047】

【発明の効果】

以上説明したように、本発明によれば、音声を効率良く、かつ精度良く認識することができる音声認識装置及びその方法、コンピュータ可読メモリを提供できる。

【0048】

【図面の簡単な説明】

【図1】本発明の実施形態の音声認識装置のハードウェア構成を示す図である。

【図2】本発明の実施形態で実行される処理を示すフローチャートである。

10

【図3】本発明の実施形態の認識対象の表現方法の一例を示す図である。

【図4】本発明の実施形態の音声認識結果及びそのスコアの一例を示す図である。

【図5】本発明の実施形態のヘルプメッセージの一例を示す図である。

【図6】本発明の実施形態の他の音声認識方法を示す図である。

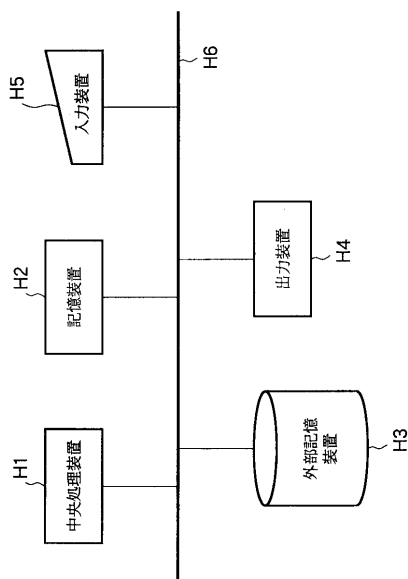
【図7】従来の音声認識方法の一例を示す図である。

【符号の説明】

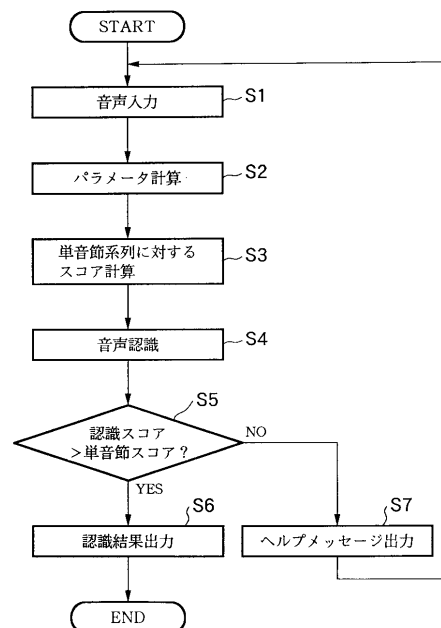
- H 1 中央処理装置
- H 2 記憶装置
- H 3 外部記憶装置
- H 4 出力装置
- H 5 入力装置
- H 6 バス

20

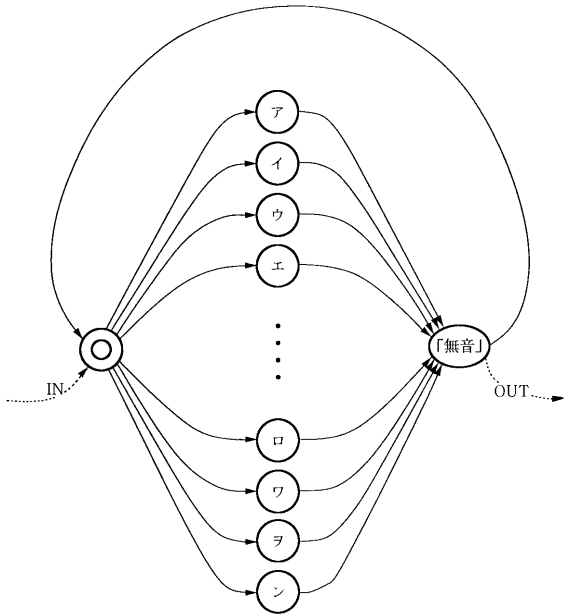
【図1】



【図2】



【図3】



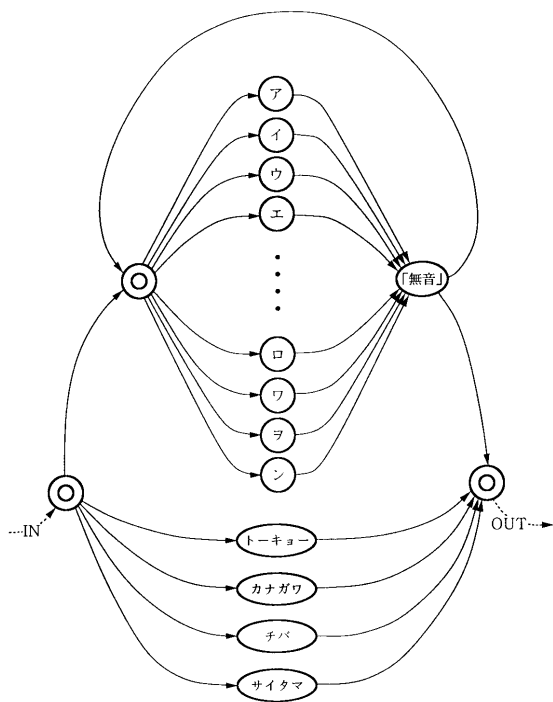
【図4】

i	認識結果 r _i	スコア S _i
1	神奈川	1032.4
2	香川	1013.8
3	佐賀	902.1

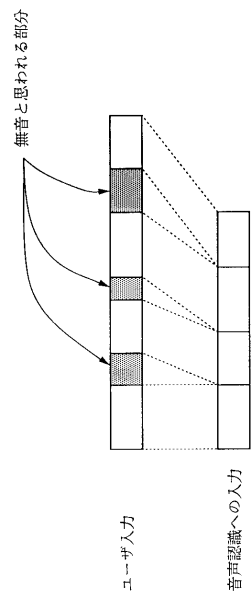
【図5】

音と音の間を切らずに一息で話してください。

【図6】



【図7】



フロントページの続き

(72)発明者 中川 賢一郎
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 荏原 雄一

(56)参考文献 特開昭56-089799(JP,A)
信学技報SP91-96
渡辺,他,音節認識を用いたゆ度補正による未知発話のリジェクション,電子情報通信学会論文誌D-II,1992年,Vol.J75-D-II,No.12,p.2002-2009

(58)調査した分野(Int.Cl.⁷,DB名)
G10L 15/00-15/28
JICSTファイル(JOIS)
IEEE Xplore