

(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2012-137776
(P2012-137776A)

(43) 公開日 平成24年7月19日 (2012.7.19)

(51) Int.Cl.
G 1 O L 15/28 (2006.01)

F I
G 1 O L 15/28 3 5 3

テーマコード (参考)
5 D O 1 5

審査請求 有 請求項の数 9 O L (全 19 頁)

(21) 出願番号	特願2012-39034 (P2012-39034)	(71) 出願人	505367431 ノヴォーリス テクノロジーズ リミテッド Novauris Technologies Ltd. イギリス国 グロスターシャー州 チュル テンハム ビショップス クリーヴ スト ーク ロード ミルバンク
(22) 出願日	平成24年2月24日 (2012.2.24)		
(62) 分割の表示	特願2006-509541 (P2006-509541) の分割	(74) 代理人	100146318 弁理士 岩瀬 吉和
原出願日	平成16年3月31日 (2004.3.31)	(74) 代理人	100114188 弁理士 小野 誠
(31) 優先権主張番号	10/401,572	(74) 代理人	100119253 弁理士 金山 賢教
(32) 優先日	平成15年3月31日 (2003.3.31)		
(33) 優先権主張国	米国 (US)		
		最終頁に続く	

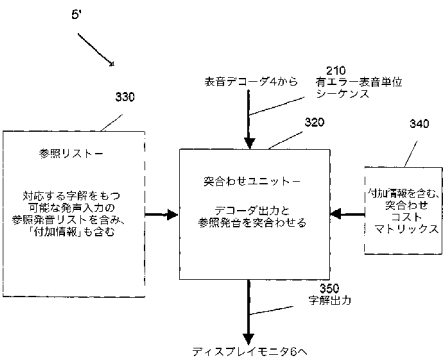
(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】図1に示される構成と同様の構成を有するシステムの確度を高めて計算負荷を軽減し、それによってレスポンスを高速化したり、処理能力がさほど強力でないハードウェアを使用できるようにすること。

【解決手段】音声認識方法は表音認識器によって出力される表音シーケンスを受け取る工程を含む。方法は、参照リストに格納された複数の参照音素シーケンスの内の、表音シーケンスに最もぴったり合った1つと表音シーケンスを突き合わせる工程も含む。参照リストに格納された複数の参照音素シーケンスの内の少なくとも1つは、表音認識器によって出力され得る表音シーケンスに関する付加情報を含む。

【選択図】図3



【特許請求の範囲】**【請求項 1】**

表音認識器から出力される、強勢レベルを伴わない表音シーケンスを受け取る入力ユニットと、

強勢レベルを伴う複数の参照音素シーケンスを格納する参照リストと、

前記表音シーケンスにおける表音単位の挿入、脱落または置換の確率に関する情報を、表音単位および強勢レベル毎に有するコストマトリックスと、

前記入力ユニットが受け取った前記表音シーケンス出力を、前記参照リストに格納された前記複数の参照音素シーケンスのうちの 1 つと突き合わせて、前記コストマトリックスからの情報を利用して前記表音シーケンスに最も適合する参照音素シーケンスを決定する突合わせユニットと

10

を備えることを特徴とする音声認識装置。

【請求項 2】

前記コストマトリックスが有する前記情報は、表音単位および強勢レベル毎に前記確率の逆対数で表されるコスト値であり、

前記突合わせユニットは、コスト値の合計が最小となる参照音素シーケンスを前記最も適合する参照音素シーケンスと決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】

前記強勢レベルは 2 レベル又は 3 レベルであることを特徴とする請求項 1 又は 2 に記載の音声認識装置。

20

【請求項 4】

前記表音認識器は、前記表音認識器への音響的シーケンス入力に出現する母音直後子音および母音直前子音を弁別できないことを特徴とする請求項 1 から 3 のいずれかに記載の音声認識装置。

【請求項 5】

前記コストマトリックスは、子音を母音直後子音と母音直前子音とに区別して、前記確率に関する情報を有していることを特徴とする請求項 1 から 4 のいずれかに記載の音声認識装置。

【請求項 6】

30

表音認識器から出力される表音シーケンスを受け取る入力ユニットであって、前記音響的シーケンス入力に出現する母音直後子音および母音直前子音を弁別できない入力ユニットと、

母音直後子音と母音直前子音との弁別を伴う複数の参照音素シーケンスを格納する参照リストと、

前記表音シーケンスにおける表音単位の挿入、脱落または置換の確率に関する情報を、表音単位、および、母音直後子音と母音直前子音との弁別毎に有するコストマトリックスと、

前記入力ユニットが受け取った前記表音シーケンス出力を、前記参照リストに格納された前記複数の参照音素シーケンスのうちの 1 つと突き合わせて、前記コストマトリックスからの情報を利用して前記表音シーケンスに最も適合する参照音素シーケンスを決定する突合わせユニットと

40

を備えることを特徴とする音声認識装置。

【請求項 7】

コンピュータを、

表音認識器から出力される、強勢レベルを伴わない表音シーケンスを受け取る入力ユニットと、

強勢レベルを伴う複数の参照音素シーケンスを格納する参照リストと、

前記表音シーケンスにおける表音単位の挿入、脱落または置換の確率に関する情報を、表音単位および強勢レベル毎に有するコストマトリックスと、

50

前記入力ユニットが受け取った前記表音シーケンス出力を、前記参照リストに格納された前記複数の参照音素シーケンスのうちの１つと突き合わせて、前記コストマトリックスからの情報を利用して前記表音シーケンスに最も適合する参照音素シーケンスを決定する突合わせユニットと
を備える音声認識装置として機能させるためのプログラム。

【請求項 8】

前記音声認識装置は請求項 2 から 5 のいずれかに記載の音声認識装置であることを特徴とするプログラム。

【請求項 9】

コンピュータを、

表音認識器から出力される表音シーケンスを受け取る入力ユニットであって、前記音響的シーケンス入力に出現する母音直後子音および母音直前子音を弁別できない入力ユニットと、

母音直後子音と母音直前子音との弁別を伴う複数の参照音素シーケンスを格納する参照リストと、

前記表音シーケンスにおける表音単位の挿入、脱落または置換の確率に関する情報を、表音単位、および、母音直後子音と母音直前子音との弁別毎に有するコストマトリックスと、

前記入力ユニットが受け取った前記表音シーケンス出力を、前記参照リストに格納された前記複数の参照音素シーケンスのうちの１つと突き合わせて、前記コストマトリックスからの情報を利用して前記表音シーケンスに最も適合する参照音素シーケンスを決定する突合わせユニットと
を備える音声認識装置として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声認識装置に関する。

【背景技術】

【0002】

自動音声認識を実行するための従来のシステム及び方法の１つが図 1 に示される。

【0003】

入力音声が入力された音声をデジタル信号に変換するために、マイクロホン（図 1 のモジュール 1）で電気信号に変換されて、アナログ - デジタル変換器（モジュール 2）に送られる。例として、アナログ - デジタル変換器モジュール 2 は 10 kHz のサンプリングレートで動作し、16 ビットの数値精度を有する。市販のほとんどのパーソナルコンピュータ（PC）はそのようなアナログ - デジタル変換器を備える。

【0004】

別の構成において、音声信号は電話回線またはその他の通信システム（例えば LAN または WAN）を通じて送信することができ、あるいは記録装置によって供給することができよう。

【0005】

デジタル化された音声信号は次いで、音響特性を生成するスペクトルアナライザ（モジュール 3）に供給される。例えば、スペクトルアナライザモジュール 3 は 10 ミリ秒または 20 ミリ秒毎に 1 つのレートで音響特性を生成する。音響特性のセットは、例えば、音声認識技術の当業者に知られているように、第 1 の 13 個のケプストラム係数並びにこれらのケプストラム係数の時間に関する一次微分係数及び二次微分係数からなる。

【0006】

音響特性のシーケンスは次いで表音デコーダ（モジュール 4）への入力を形成し、表音デコーダモジュール 4 はこの入力と複数の表音単位を表す 1 組のモデルとを突合わせ、よって、このプロセスは最終的に認識されることになる単語とは無関係である。表音デコ

10

20

30

40

50

ーダモジュール 4 を構成する表音単位は、言語の音素、すなわち単子音及び単母音、にほぼ対応していると言って差し支えない。いくつかの、または全ての母音の異なる強勢レベルに対して別々の音標が用いられることもある。

【 0 0 0 7 】

公開ツールキット H T K を、例えば、表音デコーダを構成し、トレーニング発声のコーパスに必要なモデルを構築するために用いることができる。このツールキットはインターネットアドレス ' <http://htk.eng.cam.ac.uk> ' からインターネットを通じて入手することができ、このアドレスは非特許文献 1 へのアクセスも提供する。H T K は適したスペクトルアナライザの例も提供する。

【 0 0 0 8 】

表音デコーダモジュール 4 の出力は音素様単位のシーケンスである。いくつかのタイプの表音デコーダについては、出力は単シーケンスではなく表音単位のマトリックスに対応するが、簡単のため、以降の説明では単シーケンスを想定する。音声認識に内在する困難さのため、表音シーケンスは通常、入力 of 正確な表音のコピーではなく、そのようなコピーを大雑把に近似する、誤りを含むシーケンスである。

【 0 0 0 9 】

図 1 において、話者はマイクロホン 1 に ' James Smith ' と発声するが、表音デコーダモジュール 4 は結果的に (正しい表音参照 / j h e y m z s m i h t h / と比較して) 誤りを含むシーケンス / c h e y m s n i h / に対応する出力を行っている。

【 0 0 1 0 】

発声に対応する単語または単語シーケンスを判断するため、字解器 (モジュール 5) において、表音デコーダモジュール 4 から表音シーケンス出力と入力の可能な字解に対応する参照表音シーケンスのセットとが突き合わされる。既知の選択発音を受け入れるために、1 つより多くの参照表音シーケンスを使って、特定の単語または単語シーケンスを表すこともできる。次に図 2 に移れば、字解器モジュール 5 は表音デコーダモジュール 4 から出力される表音単位シーケンス 2 1 0 を受け取る。字解器モジュール 5 は、表音デコーダモジュール 4 の出力と参照リスト 2 3 0 から得られるような参照発音とを突き合わせる突合わせユニット 2 2 0 も備える。参照リスト 2 3 0 は対応する字解を含む可能な音声入力の参照発音のリストを収める。突合わせユニット 2 2 0 は突合わせ判断に際して突合わせしコストマトリックス 2 4 0 からの情報も利用する。突合わせユニット 2 2 0 は字解出力 2 5 0 を出力し、字解出力 2 5 0 は、ユーザに表示するためにディスプレイモニタモジュール 6 に供給される。

【 0 0 1 1 】

字解器で実施される突合わせプロセスは、表音認識器モジュール 4 の出力と、話者が紡ごうとした語彙シーケンスに最も密に対応する参照シーケンスの間の相違を許容しなければならない。これらの相違には、(図 1 の表音デコーダモジュール 4 に相当する) 表音認識器の確度において予期されていない発音の変異または脱落を含む、様々な原因があり得る。これらの相違は、対応する参照シーケンスに関する表音認識器からの出力において、表音単位の挿入、脱落または置換という形で現れ得る。

【 0 0 1 2 】

そのような相違に出会うと、有効な突合わせプロセスはそれぞれの表音単位の挿入及び脱落の確率並びに (表音認識器で生成される表音単位と対応する参照単位が同じである確率を含む) それぞれの表音単位対間の置換の確率の推定値を利用する。次いで、ダイナミックプログラミング (D P) 文字列突合わせしプロセスを用いて全ての参照シーケンスに対して一致評点を得ることができ、次いで評点が最も高い一致を選択することができる。そのようなダイナミックプログラミング文字列突合わせプロセスの 1 つが非特許文献 2 に収められた非特許文献 3 に説明されている。

【 0 0 1 3 】

上記確率の逆対数は突合わせプロセスにおけるコストとして現れ、コストマトリックスで表される。挿入、脱落及び置換の必要な確率は、上述した突合わせプロセスを用いるこ

10

20

30

40

50

とにより、対応する表音参照シーケンスが分かっている、字解的にコピーされたトレーニング資料から反復推定することができる。

【 0 0 1 4 】

例えば、認識のためのシステムで用いたいトレーニング資料と同様の発声を含むが、正確な正書法発音表記が知られている、トレーニング資料のコーパスをとることができる。この場合、それぞれの発声に対応する認識音標のシーケンスを生成するために表音デコーダが用いられ、それぞれのデコーダシーケンスが対応する参照シーケンスに対して突き合わされる。参照シーケンスに出現するそれぞれの種類の音標に対して、突合わせプロセスがデコーダシーケンスのそれぞれを参照シーケンスにリンクする回数にしたがって判断がなされる。すべてのデコーダ音標が各参照音標に割り当てられない回数及び各デコーダ音標がいずれかの参照音標に割り当てられない回数をカウントすることもできる。これらのカウント値は、確率の対数をとることによりペナルティ値に変換される、割当確率及び脱落確率を計算するために必要な情報を提供する。

10

【 0 0 1 5 】

ペナルティ値を得るために必要な確率を導くアラインメントを生成するために用いられる突合わせ器には動作のためにペナルティ値セットが必要である。この点に関し、最善推定ペナルティ値の精度を反復して高めることができる。例として、それぞれのデコーダ音標が同じ音素に対応する参照音標のセットと確率 P_1 で関係付けられ、デコーダ音標と参照音標の間の他の全ての関連の確率は P_1 より低い P_2 である、非常に単純なペナルティ値セットから手順が開始される。両セットの全ての音標において、 P_2 と同じとすることができるであろう確率 P_3 で脱落が可能であると仮定される。

20

【 0 0 1 6 】

P_1 に対する妥当な初期値を 0.6 とすることができよう。確率の総和は 1 でなければならないから、 $P_3 = P_2$ であるとすれば、 P_2 は 0.4 をデコード音標の個数 (N) で割った値となる。これは、値 0.6 がデコーダ音標の 1 つに対応し、したがって、残りの音標数 ($N - 1$) + 脱落 (1) の確率で割ることになるからである。

【 0 0 1 7 】

原則として、ダイナミックプログラミング文字列突合わせプロセスにより、音声入力参照リストのそれぞれのシーケンスに対応する尤度推定値の完全なリストを提供できる。これは、デコーダによって生成されるいずれの音標もあるペナルティコストで参照セットのいずれかの音標と関連付けることができ、参照シーケンス及びデコーダシーケンスの両者のいずれの音標もあるペナルティコストで無視（すなわち、脱落）することが可能だからである。DP 突合わせプロセスは、いかなる与えられたデコーダシーケンスについても、いかなる単一の与えられた参照シーケンスについても、どの音標を無視すべきでありどの音標を関連付けるべきであるかを選んで、2つのシーケンスをリンクさせるに必要な最小総ペナルティコストを見いだす。このプロセスにおける主要な拘束は、シーケンスのそれぞれにおける音標の順序を変更できないことだけである。

30

【 0 0 1 8 】

抽象的な例として、表音デコーダの出力がシーケンス AB であるとし、これが参照シーケンス CD と比較されたとする。1つの可能なアラインメントは、 A を C と関連付け、 B を D と関連付ける。あるいは、これらの2つの関連付けのペナルティ値は大きい、 B を C と関連付けるペナルティ値が小さければ、総ペナルティ値が最小の関連付けは、 A 及び D をある脱落コストで脱落させ、 B をある小さい関連付けコストで C と関連付けることであることがわかる。DP 突合わせで許されないのは、 B が C と関連付けられ且つ A が D と関連付けられるような並べ替えである（逆進方向に実行するため、これには時間が必要であろう）。

40

【 0 0 1 9 】

図 1 に示される発声 'James Smith' を用いて、比較プロセスがどのようににはたらき得るかを示すための一例の説明が以下に与えられる。別の実施形態に関する上述した表記において James Smith に対応する参照シーケンスは、

50

/JH EY M Z S M IH1 TH/

である。

【 0 0 2 0 】

次に、表記：

/JH EY N S M AY DH/

で与えられるであろう、‘ Jane Smythe ’ に対応する、別の、同様な参照シーケンスがあると仮定する。

【 0 0 2 1 】

表音デコーダ 4 からのもっともらしい出力は、

/CH EY N S N IH TH S/

であり得る。

【 0 0 2 2 】

以下に与えられる表 1 及び 2 は、D P 突合わせ器が、この代表例において、2 つの名前の表音デコーダ出力のペナルティコストが最小のアラインメントとして何を見いだし得るかを示す。

【表 1】

表1:“James Smith”としてのデコーダシーケンスの解釈

デコーダ	CH	EY	N	脱落	S	N	IH	TH	S
参照	JH	EY	M	Z	S	M	IH	TH	脱落
	1	2	3	4	5	6	7	8	9

【表 2】

表2:“Jane Smythe”としてのデコーダシーケンスの解釈

デコーダ	CH	EY	N		S	N	IH	TH	S
参照	JH	EY	N		S	M	AY	DH	脱落
	1	2	3	4	5	6	7	8	9

【 0 0 2 3 】

列 1 において、デコーダ音標 /CH/ は両方の突合わせにおいて参照音標 /JH/ に関連付けられる。音素 /CH/ 及び /JH/ は音響的にかなり似ており、したがってペナルティ値は大きくないであろうし、いずれにしても両アラインメントに等しく適用される。列 2 のアラインメントも、音素を同じ音素と関連付けるから、小さいペナルティ値を有するであろう。ペナルティ値は、対数確率に対応し、正しく認識されているいかなる音素の確率も 1 0 0 % ではないから、實際上ペナルティ値はゼロではないであろう。列 3 において、デコーダ音標 /N/ は James Smith の場合は /M/ に関連付けられ、Jane Smythe の場合は /N/ に関連付けられる。この場合、多くはないが /M/ と /N/ は混同されやすいので、Jane Smythe 仮説はより小さいペナルティ値を受ける。列 4 において、James Smith 仮説は、デコーダが /Z/ を検出できなかったと仮定しなければならないためにペナルティ値を受ける（これは、/S/ が後続しているこの状況では実際上非常におこり易いが、D P は状況について知ることがなく、標準の脱落ペナルティ値を適用する）。Jane Smythe 仮説の場合、列 4 は単にスペース維持欄であり、いかなるペナルティ値も受けない。列 5 及び 6 は飛ばして、列 7 においてデコーダシーケンスの母音 /IH/ は、James Smith 参照シーケンスの一致している母音である /IH/ と突き合わされ、Jane Smythe 参照シーケンスの異なる母音である /AY/ と突き合わされる。/IH/ を /AY/ と関連付けるために払われるペナルティ値は、2 つの母音が音響的に極めて異なっているから、大きくなりそうである。これは、Jane Smythe 仮説が他の全ての箇所において同じかまたはより良い評点をつけるとしても、それにもかかわらず James Smith 仮説がより高い総評点を得ることを十分に意味し得る。列 8 は列 3 と同様である。最後に、列 9 が示しているのは、発声の終わりに少量の雑音があるときにしばしば発生するも

10

20

30

40

50

ので、デコーダによって無声摩擦音と誤解釈される。いずれの仮説もこれを見無視するためには脱落ペナルティ値を払わなければならない。

【 0 0 2 4 】

通常、一致度の最も高い単一の予想単語シーケンスが字解器モジュール 5 によって選択されるであろう。いくつかの別の構成において、上位 N 個の一致する発声を選択することができ、ここで N はあらかじめ定められた整数であるかまたは一致の相対品質に関する規準に基づく。簡単のため、以下では、一致度の最も高い単一の予想単語の出力についてのみ考察する。例として、字解器モジュール 5 の出力は ASCII 文字としての単語に対応する。

【 0 0 2 5 】

用途に応じて、字解は、単に（例えば P C のモニタにより）視覚表示してもよいし、またはテキスト - 音声合成器を用いて音声に再び変換してもよい。あるいは、（例えばデータベースから得られる）関連情報を調べるために用いてもよいし、または何か別の作業を制御するために用いてもよい。図 1 に示される構成において、発声は ' James Smith ' という名前であり、出力はその名前を（ ' James Smith ' を電話番号 613-929-1234 と関係付けるデータベースから得られるような）対応電話番号とともに含む。2 つの情報はディスプレイモニタ（モジュール 6 ）に送られる。

【 0 0 2 6 】

図 1 のモジュール 3 , 4 及び 5 は汎用コンピュータのプログラムとして実現できる。

【 0 0 2 7 】

図 1 に示されるシステムはうまくはたらくことができるが、偶因エラーはあらゆる音声認識システムにおこる。エラーレートは、語彙の大きさ及び混同し易さ、音響信号の品質等の要因によって増加する。このため、難易度の高いある種の用途では音声認識システムが実用不能になり得る。さらに、音声認識システムを、より小規模で、より低コストの処理ハードウェア上でさらに迅速に応答または機能させるかあるいは応答及び機能させることが常に望ましい。

【先行技術文献】

【非特許文献】

【 0 0 2 8 】

【非特許文献 1】エス・ジェイ・ヤング（S . J. Young）等「HTKブック（The HTK Book）」

【非特許文献 2】「タイムワープ、文字列編集及びマクロモレキュール：シーケンス比較の理論及び実際（Time Warps, String Edits and Macromolecules: the Theory and Practice of Sequence Comparison）」, 1983 年, アディソン - ウェズリー（Addison Wesley）

【非特許文献 3】ジェイ・ビー・クラスカル（J. B. Kruskal）「シーケンス比較概説（An Overview of Sequence Comparison）」

【発明の概要】

【発明が解決しようとする課題】

【 0 0 2 9 】

本発明の課題は、図 1 に示される構成と同様の構成を有するシステムの確度を高めて計算負荷を軽減し、それによってレスポンスを高速化したり、処理能力がさほど強力でないハードウェアを使用できるようにすることにある。

【課題を解決するための手段】

【 0 0 3 0 】

本発明の一実施形態にしたがえば、音響入力を受け取る工程、音響入力をデジタル信号に変換する工程、音響特性を得るためにデジタル信号にスペクトル分析を施す工程、デジタル信号の音響特性から表音シーケンスを得る工程及び表音シーケンスに最もよく一致する表音参照を得るために字解を実施する工程を含み、字解を実施する工程中に得られる表音参照における表音単位のセットがデジタル信号の音響特性から表音シーケンスを得る工

10

20

30

40

50

程で生成され得る表音単位のセットと異なる、音声認識方法が提供される。

【 0 0 3 1 】

また、本発明の別の実施形態にしたがえば、表音認識器による表音シーケンス出力を受け取る工程を含む音声認識方法が提供される。本方法は、表音シーケンスを、参照リストに格納された複数の参照音素シーケンスの内の、表音シーケンスに最もぴったり合った１つと突き合わせる工程も含む。参照リストに格納された複数の参照音素シーケンスの内の少なくとも１つは表音認識器によって出力され得る表音シーケンスとは異なる。

【 0 0 3 2 】

本発明のまた別の実施形態にしたがえば、表音認識器によって出力される表音シーケンスを受け取るように構成された入力ユニットを備える音声認識装置が提供される。本音声認識装置は複数の参照音素シーケンスを格納するように構成された参照リストも備える。この音声認識装置はさらに、入力ユニットで受け取られる表音シーケンスを、参照リストに格納された複数の参照音素シーケンスの内の、表音シーケンスと最もぴったり合った１つと突き合わせるように構成された突合わせユニットを備える。参照リストに格納された複数の参照音素シーケンスの内の少なくとも１つは表音認識器によって出力され得る表音シーケンスと異なる。

10

【 0 0 3 3 】

本発明の上述の利点及び特徴は以下の詳細な説明及び添付図面を参照すれば明らかになるであろう。

【 図面の簡単な説明 】

20

【 0 0 3 4 】

【 図 1 】 従来の音声認識装置のブロック図である。

【 図 2 】 従来の字解器を構成する要素を示すブロック図である。

【 図 3 】 本発明の第 1 の実施形態にしたがう字解器を構成する要素を示すブロック図である。

【 発明を実施するための形態 】

【 0 0 3 5 】

図面を参照して本発明を以下に説明する。これらの図面は、本発明のシステム及び方法及びプログラムを実施する特定の実施形態のいくつかの詳細を示す。しかし、図面による本発明の説明が、図面に存在し得るいずれかの限定を本発明に課すものと解されるべきではない。本発明には、本発明の動作を達成するための方法、システム及びいずれかのコンピュータ読取可能媒体上のプログラム製品が想定されている。本発明の実施形態は、既存のコンピュータプロセッサを用いて、あるいは上記または別の目的のために導入された専用コンピュータプロセッサによって、またはハードワイヤードシステムによって実施することができる。

30

【 0 0 3 6 】

上述したように、本発明の範囲内の実施形態は、コンピュータ実行可能な命令またはデータ構造を搬送するかまたは格納するためのコンピュータ読取可能媒体を含むプログラム製品を含む。そのようなコンピュータ読取可能媒体は、汎用コンピュータまたは専用コンピュータがアクセスできる利用可能媒体であれば、どのような媒体とすることもできる。例として、そのようなコンピュータ読取可能媒体には、RAM, ROM, EPROM, EEPROM, CD-ROMまたはその他の光ディスク記憶装置、磁気ディスク記憶装置またはその他の磁気記憶装置、あるいはコンピュータ実行可能な命令またはデータ構造の形態で所望のプログラムコードを搬送または格納するために用いることができ、汎用コンピュータまたは専用コンピュータがアクセスできる、その他のいずれかの媒体があり得る。情報がネットワークまたは別の通信回線（ハードワイヤード回線、無線回線、またはハードワイヤード回線と無線回線の組合せ）を通じてコンピュータに転送されるかまたは与えられる場合、コンピュータは適宜そのような回線をコンピュータ読取可能媒体と見なす。したがって、そのような回線はいずれもコンピュータ読取可能媒体と適宜称される。上記の組合せもコンピュータ読取可能媒体な範囲内に含まれる。コンピュータ実行可能命令は

40

50

、例えば、汎用コンピュータ、専用コンピュータまたは専用処理装置にある機能または機能群を実行させる命令及びデータを含む。

【 0 0 3 7 】

本発明は、ネットワーク環境においてコンピュータによって実行される、プログラムコードのような、コンピュータ実行可能命令を含むプログラム製品により一実施形態において実施され得る方法工程の全般的状況の下に説明されるであろう。一般に、プログラムモジュールは、特定のタスクを実行するかまたは特定の抽象データタイプを実装する、ルーチン、プログラム、オブジェクト、コンポーネント、データ構造等を含む。コンピュータ実行可能命令、関連データ構造及びプログラムモジュールは本明細書に開示される方法の工程を実行するためのプログラムコードの例を表す。そのような実行可能命令または関連データ構造の特定のシーケンスは、それぞれの工程において説明される機能を実施するための対応作業例である。

10

【 0 0 3 8 】

いくつかの実施形態において、本発明はプロセッサを有する1つまたはそれより多くのリモートコンピュータとの論理接続を用いるネットワーク環境で動作することができる。論理接続には、本明細書において限定ではなく例として提示される、ローカルエリアネットワーク（LAN）及び広域ネットワーク（WAN）を含めることができる。そのようなネットワーク環境はオフィス内コンピュータネットワークまたは企業内コンピュータネットワーク、イントラネット及びインターネットにおいて珍しくはない。当業者であれば、そのようなネットワークコンピューティング環境が一般に、パーソナルコンピュータ、ハンドヘルドデバイス、マルチプロセッサシステム、マイクロプロセッサベース家電またはプログラマブル家電、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ等を含む、多くのタイプのコンピュータシステム構成を包含することを認めるであろう。本発明は、通信ネットワークを介して（ハードワイヤードリンク、無線リンクまたはハードワイヤードリンクと無線リンクの組合せにより）リンクされるローカル処理デバイスまたはリモート処理デバイスでタスクが実行される分散型コンピューティング環境において実施することもできる。分散型コンピューティング環境では、プログラムモジュールはローカル記憶装置及びリモート記憶装置のいずれにもおくことができる。

20

【 0 0 3 9 】

本発明のシステム全体またはその一部を実施するための例示的システムは、処理ユニット、システムメモリ及びシステムメモリを含む様々なシステムコンポーネントを処理ユニットに接続するシステムバスを有する、通常のコンピュータの形態の汎用コンピューティング装置を備えることができよう。システムメモリには、リードオンリメモリ（ROM）及びランダムアクセスメモリ（RAM）を含めることができる。コンピュータは、磁気ハードディスクから読み出し、磁気ハードディスクに書き込むための磁気ハードディスクドライブ、リムーバブル磁気ディスクから読み出し、リムーバブル磁気ディスクに書き込むための磁気ディスクドライブ、及びCD-ROMまたはその他の光媒体のような、リムーバブル光ディスクから読み出し、リムーバブル光ディスクに書き込むための光ディスクドライブも備えることができる。これらのドライブ及び関連するコンピュータ読取可能媒体は、コンピュータ実行可能命令、データ構造、プログラムモジュール及びコンピュータのためのその他のデータの揮発性記憶を提供する。

30

40

【 0 0 4 0 】

以下の述語は本発明の説明に用いられ、新しい述語及び特定の意味が与えられる述語を含み得る。

【 0 0 4 1 】

「言語要素」は文字言語または音声言語の単位である。

【 0 0 4 2 】

「発声要素」は関連付けられる名称をもつ発声の期間である。この名称は、発声期間中に発声される、単語、音節または音素とすることができ、あるいは発声期間中に聞き取られる音のシステムラベルを表す、自動的に生成される音標のような抽象記号とすることが

50

できる。

【0043】

本発明の目的のための「フレーム」は、与えられたシステムまたはサブシステムによって分析される最短時間単位である、固定または可変の時間単位である。フレームは、10ミリ秒毎に1回スペクトル信号処理を実施するシステムにおける10ミリ秒などの固定単位とすることができ、あるいは、推定ピッチ期間または音素認識器が特定の認識される音素または発音セグメントと関係付けられた期間などのデータ依存可変単位とすることができる。従来技術のシステムとは異なり、述語「フレーム」は、時間単位が固定期間であるかまたは与えられたシステムの全てのサブシステムにおいて同じフレームが用いられることを意味していないことに注意されたい。

10

【0044】

「評点」は与えられた仮説が言声の何らかのセットにどれだけよく一致するかの数値評価である。特定の実施における約定に応じて、一致がよくなるほど高くなる（確率または確率の対数によるような）評点または低くなる（負のlog確率またはスペクトル距離によるような）評点で表すことができる。評点は正となる場合も、負となる場合もある。評点は、センテンス内の単語シーケンスの「先験的」確率など、与えられた仮説に関連付けられる言語要素シーケンスの相対尤度の尺度を含むこともできる。

【0045】

「ダイナミックプログラミング一致点数評価」は、ダイナミックプログラミングを用いることによる、ネットワークまたはモデルシーケンスと音響的言声シーケンスの間の一致の程度を計算するプロセスである。ダイナミックプログラミング突合わせプロセスは、2つの音響的言声シーケンスの突合わせを行うかまたは時間的アラインメントをとるため、または2つのモデルまたはネットワークの突合わせを行うために用いることもできる。ダイナミックプログラミングコンピューティングは、例えば、ネットワークを通る最善評価経路を見いだすため、またはネットワークを通る全ての経路の確率の総和を求めるために用いることができる。述語「ダイナミックプログラミング」の従来の用法が変わる。「ダイナミックプログラミング」は「最善経路突合わせ」を意味するために特に用いられることがあるが、本特許の目的のための用法は、「最善経路突合わせ」、「経路総和」突合わせ及びこれらの近似を含む、関連コンピューティング法のより広いクラスを網羅する。音響的言声シーケンスに対するモデルの時間アラインメントは一般に一致評点のダイナミックプログラミングコンピューティングの副次効果として利用できる。ダイナミックプログラミングは（モデルと言声シーケンスの間ではなく）2つのモデルまたはネットワーク間の一致の程度を計算するために用いることもできる。スペクトル距離のような、モデルのセットに基づかない距離尺度が与えられれば、ダイナミックプログラミングは発声要素の2つの事例の突合わせを行い、直接に時間アラインメントをとるために用いることもできる。

20

30

【0046】

「センテンス」は、検索または仮説評価のための完全な一単位として扱われる発声期間または発声要素シーケンスである。一般に、発声は沈黙期間などの音響的規準を用いてセンテンス長単位に分割されるであろう。しかし、センテンスは中間沈黙期間を含む場合があり、他方で、沈黙期間がない場合であっても、文法的規準によってセンテンス単位に分割することができる。述語「センテンス」は、発声が、データベース入力のように、文法的センテンス形態をとり得ないか、または、通常のセンテンスより短いフレーズなどの要素を完全単位としてシステムが分析している状況における、検索または仮説評価のための完全単位を称するためにも用いられる。

40

【0047】

「音素」は、音声言語における音の単一単位であり、文字言語の一文字に概ね対応する。

【0048】

「音標」は、特定の発声期間中に出現する音に関する音声認識システムの選択を示す、

50

音声認識システムによって生成されるラベルである。可能な音標のアルファベットは音素のアルファベットと同じであるように選ばれることが多いが、それらが同じでなければならないという要請はない。いくつかのシステムは、一方で「音素」と「音標」を弁別し、他方で「単音」と「音標」を弁別することができる。厳密に言えば、音素は言語の抽象的概念である。辞書からとられるような、単語がどのように発音されるはずであるかを表す音標は「音素」ラベルである。単語の特定の事例が特定の話者によってどのように発声されるかを表す音標は「表音」ラベルである。しかし、これらの2つの概念は混同され、いくつかのシステムはこれらを弁別しない。

【0049】

「スポッティング」は、隣接発声要素の内の1つまたはそれより多くを必ずしも初めに認識せずに、(1つまたはそれより多くの)発声要素に対する(1つまたはそれより多くの)モデルと発声期間における音響的言声の間の良好な一致の事例を直接検出することにより発声要素または発声要素シーケンスの事例を検出するプロセスである。

【0050】

「モデリング」は、一般には与えられた発声要素に対するモデルのセットが与えられた言声をどのように生成し得たかを計算することにより、与えられた発声要素シーケンスがどれだけよく与えられた言声のセットに一致するかを評価するプロセスである。確率モデリングにおいて、モデルの確率値によって特定されるランダムプロセスにおける与えられた言声セットを生成する与えられた要素シーケンスの確率を評価することにより、仮説の評点が計算され得るであろう。ニューラルネットワークのなど、その他の形態のモデルは、モデルを確率解釈と陽に關係付けることなく一致評点を直接計算することができ、あるいは関連付けられた生成確率過程を表すことなく「事後」確率分布を経験的に評価することができる。

【0051】

「トレーニング」は、要素の同定が既知であるかまたは既知であると想定される試料セットからのモデルのパラメータまたは十分な統計を評価するプロセスである。音響的モデルの管理下トレーニングにおいては、発声要素シーケンスのコピーが既知であるかまたは話者が既知の台本を声を出して拾い読みしている。非管理下トレーニングにおいては、既知の台本または校合されない認識から利用できる以外のコピーはない。半管理下トレーニングの一形態においては、ユーザがコピーを陽に校合しておくことはできないが、そうする機会が与えられた場合には、いかなるエラー訂正も行わないことで陰にそうしておくことができる。

【0052】

「音響的モデル」は、発声要素シーケンスが与えられた場合に、音響的言声シーケンスを生成するためのモデルである。音響的モデルは、例えば、隠された確率過程のモデルとすることができる。隠された確率過程は発声要素シーケンスを生成し、それぞれの発声要素に対してゼロまたはそれより多くの音響的言声のシーケンスを生成するであろう。音響的言声は、周波数及び時間の関数としての振幅のような、音響波形から導かれる(連続)物理的尺度とすることができ、あるいは音声圧縮に用いられるようなベクトル量子化器でつくられるかまたは表音認識器の出力のような、離散有限ラベルセットの言声とすることができる。連続物理的尺度は一般にガウス型分布または混合ガウス型分布のような何らかのパラメトリック確率分布の形態によってモデル化されるであろう。それぞれのガウス型分布はそれぞれの言声測定値の平均及び共変行列によって特徴付けられるであろう。共変行列が対角行列であるとすれば、多分散ガウス型分布はそれぞれの言声測定値の平均及び分散によって特徴付けられるであろう。有限ラベルセットからの言声は一般に非パラメトリック離散確率分布としてモデル化されるであろう。しかし、別の形態の音響的モデルを用いることができるであろう。例えば、「事後」確率評点を近似するためにトレーニングが行われているかまたはいない、ニューラルネットワークを用いて総一致評点を計算することができるであろう。あるいは、確率モデルを下敷きにせずスペクトル距離測定値を用いることができ、あるいは確率評価ではなくファジー論理を用いることができるである

10

20

30

40

50

う。

【 0 0 5 3 】

「言語モデル」は、文法に、または特定の発声要素に対して文脈のゼロまたはそれより多くの言語要素の値が与えられた場合の特定の言語要素の確率に対する統計モデルにかけられる言語要素シーケンスを生成するためのモデルである。

【 0 0 5 4 】

「汎言語モデル」は、純統計言語モデル、すなわち文法を明示的に含まない言語モデルであるか、または文法を明示的に含み、統計成分も有することができる、文法ベース言語モデルであり得る。

【 0 0 5 5 】

「文法」は、どの単語シーケンスまたはセンテンスが正当な（すなわち文法的な）単語シーケンスであるかの公式指定である。文法指定の実施には多くの手段がある。文法を指定するための一手段は、言語学で、及びコンピュータ言語のためのコンパイラの作成者に、良く知られた形態の書換規則のセットを用いることである。文法を指定するための別の手段は、状態空間またはネットワークとしての手段である。状態空間のそれぞれの状態またはネットワークのノードに対し、ある単語または言語要素だけがシーケンスの次の言語要素となることができる。そのような単語または言語要素のそれぞれに対し、（例えばアークの終端におけるノードに続くアークによって）次の単語の終端におけるシステムの状態がどのようになるかに関しての（例えばネットワークのラベル付アークによる）指定がある。文法表現の第3の形態は全ての正当なセンテンスのデータベースとしての表現である。

【 0 0 5 6 】

「確率文法」は、言語要素のそれぞれの正当なシーケンスの確率のモデルも含む文法である。

【 0 0 5 7 】

「純統計言語モデル」は、文法成分を有していない統計言語モデルである。純統計言語モデルにおいては一般に、全ての可能な言語要素シーケンスがゼロではない確率を有するのであろう。

【 0 0 5 8 】

「エントロピー」は、確率分布における情報量または関連付けられるランダム変数の情報理論上の尺度である。エントロピーは一般に式：

【 数 1 】

$$E = \sum_i p_i \log(p_i)$$

【 0 0 5 9 】

で与えられる。ここで対数の底は2であり、エントロピーの測定単位はビットである。

【 0 0 6 0 】

分類システムにおける「分類作業」は、目標とするクラスのセットの分割である。

【 0 0 6 1 】

（第1の実施形態）

図1に示されるシステムにおいては、表音認識器（図1のモジュール4）で生成される音標セットと（図1の字解器モジュール5によって利用されるような）参照セットとして用いられる音標セットの間に1対1の対応がある。しかし、図2に示されるような従来の字解器モジュール5の代わりに、本明細書に説明される本発明の第1の実施形態にしたがう、図3に示されるような字解器モジュール5'を用いれば、そのような1対1の対応はない。

【 0 0 6 2 】

図3に示されるように、字解器モジュール5'は表音デコーダモジュール4（図1参照）から出力される表音単位シーケンス210を受け取る。字解器モジュール5'は表音デ

10

20

30

40

50

コーダモジュール 4 の出力と参照リスト 3 3 0 から得られるような参照発音を突き合わせる突合わせユニット 3 2 0 も備える。第 1 の実施形態において、参照リスト 3 3 0 は可能な口語入力の参照発音のリストを対応する字解とともに含み、(図 2 に示される参照リスト 2 3 0 には含まれていない) 付加情報も含む。突合わせユニット 3 2 0 は突合わせ判断を行う際に突合わせコストマトリックス 3 4 0 からの情報も利用し、突合わせコストマトリックス 3 4 0 は図 2 に示されるような突合わせコストマトリックス 2 4 0 には与えられていない付加情報を含む。突合わせユニット 3 2 0 は、ユーザへの表示のためにディスプレイモニタモジュール 6 (図 1 参照) に与えられる、字解出力 2 5 0 を出力する。

【0063】

参照リスト 3 3 0 及び突合わせコストマトリックス 3 4 0 への特定の参照によれば、アメリカ英語及びその他の言語の単語の母音は強勢の置き方が異なり得る。例えば、名詞 'permit' では強勢が第 1 母音に置かれ、第 2 母音には強勢が置かれませんが、動詞 'permit' の母音には強勢が逆に置かれる。通常、単語の強勢パターンを表すには 2 つまたは 3 つの強勢レベルが用いられる。3 レベルの場合、強勢は通常、「第 1 強勢」、「第 2 強勢」及び「無強勢」とラベル付けされる。

【0064】

表音認識器を備える発声認識器は一般に、全ての母音単位の強勢レベルを弁別しようとはしない。これは、ある状況においては母音の強勢レベルを確実に同定することが困難であるからである。強勢は主として母音の相対ラウドネス、相対持続時間及び相対ピッチと関連し、これらの全ては発音同定のために用いられるスペクトルの一般的形状より決定がさらに困難であり得る。

【0065】

第 1 の実施形態においては、複数の強勢レベルが(参照リスト 3 3 0 に格納されるような)音素の参照表示に用いられる付加情報として与えられ、最善の一致を決定するために字解器モジュール 5' によって利用される突合わせコストマトリックス 3 4 0 における付加情報としても現れる。一実施形態にしたがえば、突合わせコストマトリックス 3 4 0 は発音解釈間の混同の確率をコード化し、母音の場合、混同確率は強勢レベルに依存し得る。強勢レベルが低い特定の母音は、例えば、(話者が発声しないため、または発声が短く、弱いために表音認識器モジュール 4 が検出できない結果となるため)明らかに脱落する確率が高くなり得る。表 3 に示されるように、強勢が置かれられない母音は一般に対応する強勢がおかれた母音より正しく認識される確率が低い。

【0066】

さらに詳しくは、表 3 は、本発明の第 1 の実施形態にしたがう字解器モジュール 5' によって利用される突合わせコストマトリックス 3 4 0 に与えられるような、参照発音の特定の母音(行)に対応する表音デコーダモジュール 4 からの言声母音音標出力(列)の推定確率のセットの一例を示す。表音デコーダモジュール 4 からの母音音標出力には強勢レベルがないが、(参照リスト 3 3 0 に格納された)参照発音の音標には、第 1 強勢、第 2 強勢または無強勢を意味する数字(それぞれ、「1」、「2」または「0」)が付されている。一致する母音に対応する(すなわち、母音が正しく認識される)セルは表 3 に太字で示される。これらの確率は、相応に信頼できる認識器について期待されるように、対応しない母音に対するセルの値より大きい。第 1 強勢をもつ参照母音は、対応する無強勢形態よりも高い、正認識確率を有する傾向があり、第 2 強勢をもつ母音は一般に中間の確率を有することに注意されたい。表 3 は、確率が対数確率に変換されると、第 1 の実施形態にしたがう字解器モジュール 5' によって利用されることになる突合わせコストマトリックス 3 4 0 の一部を形成するために用いることができる。表 3 の情報は、発明者等によって実施された実験的分析評価から得られた。

【表 3】

表 3

	AA	AE	AO	AW	AY
AA0	0.175	0.080	0.163	0.030	0.028
AA1	0.653	0.041	0.036	0.019	0.009
AA2	0.524	0.055	0.144	0.020	0.014
AE0	0.043	0.210	0.009	0.022	0.014
AE1	0.022	0.683	0.004	0.013	0.05
AE2	0.040	0.452	0.008	0.018	0.013
AO0	0.061	0.024	0.189	0.044	0.016
AO1	0.136	0.009	0.464	0.014	0.001
AO2	0.112	0.012	0.408	0.038	0.014
AW0	0.074	0.033	0.029	0.147	0.029
AW1	0.026	0.015	0.011	0.756	0.003
AW2	0.058	0.018	0.019	0.682	0.004
AY0	0.018	0.076	0.007	0.004	0.403
AY1	0.004	0.010	0.001	0.001	0.865
AY2	0.009	0.022	0.004	0.004	0.527

10

20

【0067】

表3がどのように役立ち得るかを見るために、それぞれの発音を/AE1 K L AA0 F/及び/AA1 K L AE0 F/と表すことができる固有名詞‘Ackloff’及び‘Ocklaff’を収める参照リストを考える。(強勢レベルを弁別しようとししない)表音認識器からの出力は/AA k L AA F/であるとする。強勢を無視する従来の突合わせプロセスならば、いずれの突合わせもただ1つの置換、すなわち/AA/から/AE/への置換を要求するであろうから、出力がいずれの参照発音にも等しい確率でよく一致することを見いだしたであろう。しかし、強勢を考慮に入れれば、‘Ocklaff’では/AA/の/AE0/への置換(確率=0.0043)が必要であり、‘Ackloff’解釈では/AA/の/AE1/への置換(確率=0.0022)が必要であることから、‘Ocklaff’がより尤度の高い解釈であることが明らかになるであろう。

30

【0068】

(第2の実施形態)

次に、本発明にしたがう字解器モジュール5'の第2の実施形態を以下で詳細に説明する。発声された単語は音節に分けることができ、音節は単一の母音からなり、ゼロであるか、1つまたはそれより多くの子音とその母音の前に出現しているか(「母音直前」子音)、あるいはゼロであるか、1つまたはそれより多くの子音とその母音の後に出現している(「母音直後」子音)。刊行されている多くの辞書は、発音の表示の一部として音節分界を示している。例えば、1990年にロングマングループ英国社(Longman Group UK Ltd.)によって出版された、「ロングマン発音辞典(Longman Pronunciation Dictionary)」, p.xiv-xviの、「英語音素体系及びその表記:分綴法(The English phonemic system and its notation: Syllabification)」でジェイ・シー・ウエルズ(J. C. Wells)によって述べられているように、音節分界の正確な配置には複数の規準があり、どこに音節分界が生じるかに関する全ての場合に専門家が必ずしも合意していないことも事実である。音節分界は表音突合わせプロセスの補助に用いることができ、音節分界の使用は第2の実施形態にしたがう字解器モジュール5'によって実施されるような字解プロセスの補助のために利用される。

40

【0069】

いくつかの子音が母音直前と母音直後では異なる形態で具現されることは知られた様相である。例えば、1998年にケクレイド(Kekrade)によって出版された、「自動音声認識のための発音変動モデル化に関するESCAワークショッププロシーディング(Proceedi

50

ngs of ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition)」、p.47～56の、エス・グリーンバーグ(S. Greenburg)、「早口発声 - 発音変動を理解するための音節中心の見方 (Speaking in Shorthand ■ A Syllable-Centric Perspective for Understanding Pronunciation Variation)」を参照のこと。この点に関し、破裂子音/t/は母音直後位置(例えば、'post'、'lot')においては、母音直前位置(例えば、'stem'、'top')にあるときの通常の発音よりかなり弱く発音され得る。母音直後の/t/の明確な音響的証拠がないこともある。これらの差異の結果、表音デコーダモジュール4によって生成される表音シーケンスに母音直後の/t/が見られない確率は母音直前の/t/が見られない確率よりかなり高い。そのような情報は、第2の実施形態にしたがう字解器モジュール5'によって利用される参照リスト330の参照発音リストに格納される。

10

【0070】

第2の実施形態において、子音は(例えば1つまたは複数のASCII符号により)母音直前または母音直後にあるとして(参照リスト330に格納される)参照表音表示にマークされ、この区分は次いで、表音認識器モジュール4が同じ区分化を行おうとすることなしに、(字解器モジュール5'によって用いられるような)突合わせコストマトリックス340に現れる。

【0071】

表4は、アメリカ英語において発声されない破裂子音に対する上述した様相を示す。さらに詳しくは、表4は参照発音の特定の母音(行)に対応する表音デコーダからの言声母音音標(列)の推定確率のセットの一例を与える。表音デコーダ4からの子音音標は音節内の位置に対してマークされず、参照発音における子音音標はそれぞれの子音が母音直後または母音直前にあることを示す添字(それぞれ、'<'または'>')を有する。一致する子音に対応する(すなわち、子音が正しく認識される)セルが表4に太字で示される。表4の最右列は特定の母音直前または母音直後の子音が表音デコーダからの出力に全く現れないであろう場合の推定確率を示す。母音直後の発声されない破裂子音に対するそのような脱落の推定確率が対応する母音直前破裂子音に対する推定確率よりかなり高いことに注意されたい。確率を対数確率に変換すれば、表4は、第2の実施形態にしたがう字解器モジュール5'によって利用される突合わせコストマトリックス340の一部を形成するために用いることができる。表4の情報は、発明者等によって実施された実験的分析評価から得られた。

20

30

【表4】

表4

K<	K 0.890	P 0.028	T 0.034	(脱落) 0.045
K>	0.799	0.009	0.019	0.014
P<	0.017	0.609	0.035	0.163
P>	0.008	0.855	0.003	0.017
T<	0.033	0.017	0.495	0.336
T>	0.009	0.009	0.783	0.067

40

【0072】

表4がどのように役立ち得るかを見るために、それぞれの発音を/T AA1 P/及び/P AA1 T/と表すことができる項目'top'及び'pot'を収める参照リストを考える。(強勢レベルまたは母音直前子音と母音直後子音を弁別しようとしなない)表音認識器からの出力は/P AA P/であるとする。母音直前子音と母音直後子音の間の差異を無視する従来の突合わせプロセスならば、いずれの突合わせもただ1つの置換、すなわち/P/から/T/への置換を要求するであろうから、出力がいずれの参照発音にも等しい確率でよく一致することを見いだしたであろう。しかし、母音直前子音と母音直後子音の差異を考慮すれば、'pot'で

50

は/P/の/T</への置換（確率 = 0.017）が必要であり、‘top’解釈では/P/の/T>/への置換（確率 = 0.009）が必要であろうから、‘pot’がより尤度の高い解釈であることが明らかになるであろう。

【0073】

（第3の実施形態）

次に、結合された表音単位及び自由に脱落し得る表音単位を利用する、本発明にしたがう字解器モジュール5’の第3の実施形態を以下で詳細に説明する。

【0074】

いくつかの単語が2つまたはそれより多くの許容される発音を有し得ることは周知である。音声認識においてこれを扱う一手段においては、許容される全ての発音が参照リストの個別エントリとして入力される。図3に示される字解器モジュール5’は表音認識器モジュール4の出力/ch ey s m ih/を（参照リスト330から得られる）選択発音の全てに対して突合わせし、最善の一致がこの辞書項目に対する評点としてとられる。

【0075】

第3の実施形態はある単語群が許容される発音の間で同じ種類の変化を含むという特徴を利用し、変化は1つまたはそれより多くの独立音素に局限される。例えば、‘stein’で終わる固有名詞に出現する正書シーケンス‘ei’は、（/AY/と表されることが多い）単語‘sight’に出現する母音音のように、あるいは（/IY/と表されることが多い）単語‘seed’に出現する母音音のように、発音され得る。第3の実施形態は、そのような単語の全てに対して少なくとも2つの発音をリストに載せるのではなく、いずれの母音も許容されることを意味するとして解釈される新しい音標を（図3に示されるような参照リスト330に格納される）参照発音の参照表記に導入し、この新しい音標は参照リスト330に付加情報として含められる。この音標は、例えば‘AY-IY’とすることができ、あるいは何か別の専用音標とすることができる。‘either’及び‘neither’のような、別の単語クラスに対して同じ多重音素音標を用いることもできる。

【0076】

参照リストをさらにコンパクトにして、必要な格納空間をさらに小さくすることに加えて、ある多種発音のこの種の表現は突合わせプロセスに必要な計算量を低減するために用いることができる。選択発音の全ての組合せに対応する完全な単語またはフレーズを数回突合わせしなければならないのではなく、多重音素参照音標に対して突き合わせるときに、多重音素セットにおけるそれぞれの可能な音素に対して一致評点が最高の発音が用いられる、ただ一回の突合わせを実行すればよい。

【0077】

（第4の実施形態）

それ自体、または第3の実施形態に関して説明された特徴とともに、利用され得る、本発明にしたがう字解器モジュール5’の第4の実施形態においては、それぞれの多重音素単位に対応し、個々の音素に対応するコストの最小値に対応する、特別のコストセットを突合わせコストマトリックス340に付加することによって、突合わせプロセスの効率をさらに改善することができ、この特別のコストセットは突合わせコストマトリックス340に付加情報として含められる。すなわち、上述した例において、‘stein’多重音素音標と、例えば、‘bit’に通常出現する母音の間の置換コストは、この母音と‘sight’の母音と‘seed’の母音の間の置換コストの最小値となるであろう。同様に、多重音素単位の脱落コストは‘sight’の母音及び‘seed’の母音に対する個別の脱落コストの最小値となるであろう。

【0078】

（第5の実施形態）

第5の実施形態にしたがう字解器モジュール5’によって利用されるような別の種類の発音変化においては、音素が存在するかまたは存在しない別形があり得る。一例は‘p’が発音されるか発音されない‘exempt’に出現し、別の例は単語の終わりにある第2母音が発音されるか発音されない姓‘Dicke’に出現する。やはり、これを扱う従来手段では

10

20

30

40

50

2つの発音がリストに載せられることになる。第5の実施形態にしたがう字解器モジュール5'によって利用される方法は、('exempt' の参照発音が音標 'P-' を含むであろう場合における ' - ' のような専用音標を含むように) 自由に脱落可能であるとして参照リスト330に格納された参照表記において音素をマークすることである。突合わせコストマトリックス340において、この自由に脱落可能な単位の脱落コストはゼロに、または自由に脱落可能であるとマークされない対応する参照単位 (例えば 'P') に対する脱落コストより少なくともある程度低い値に設定される。

【0079】

(第6の実施形態)

本発明の第6の実施形態においては、母音強勢情報及び母音直前 / 母音直後情報のいずれもが字解器モジュール5'によって用いられる。

10

【0080】

本明細書に与えられるフローチャートは方法工程の特定の順序を示すが、これらの工程の順序が示された順序とは当然異なり得ることに注意すべきである。また、2つまたはそれより多くの工程が同時にまたは部分的に同時に実施され得る。そのような変化は選ばれるソフトウェア及びハードウェアシステムに依存し、設計者の選択に依存するであろう。そのような変化の全てが本発明の範囲内にあることは当然である。同様に、本発明のソフトウェア及びウェブ実施は、様々なデータベース検索工程、関連工程、比較工程及び決定工程を達成するためのルールベース論理及びその他の論理による標準のプログラミング技法によって達成され得るであろう。本明細書及び特許請求の範囲に用いられる述語「モジュール」または「コンポーネント」または「ユニット」が、一行またはそれより多くの行のソフトウェアコードを用いる実施、及び / またはハードウェア実施、及び / またはマニュアル入力を受け取るための装置を包含するとされることも当然である。

20

【0081】

本発明の上述の実施形態の説明は例示及び説明の目的のために提示された。上述の説明が網羅的であるとか、または本発明を開示された正確な形態に限定するとの意図はなく、改変及び変形が上述の教示に照らして見れば可能であるか、または本発明の実施から得ることができる。実施形態は、当業者が様々な実施形態で、及び考えられる特定の用途に適するように様々な実施形態において様々な変形によって本発明を利用できるように、本発明の原理及びその実用的適用を説明するために選ばれ、説明された。例えば、字解器モジュールは、本発明の1つまたはそれより多くの実施形態に関して上述した特徴を利用する音節レベルにおける字解を実施することができ、この場合は、(表音デコードの代りに) 音節デコードが字解器モジュールに音節シーケンス (または音節マトリックス) を与えることになる。

30

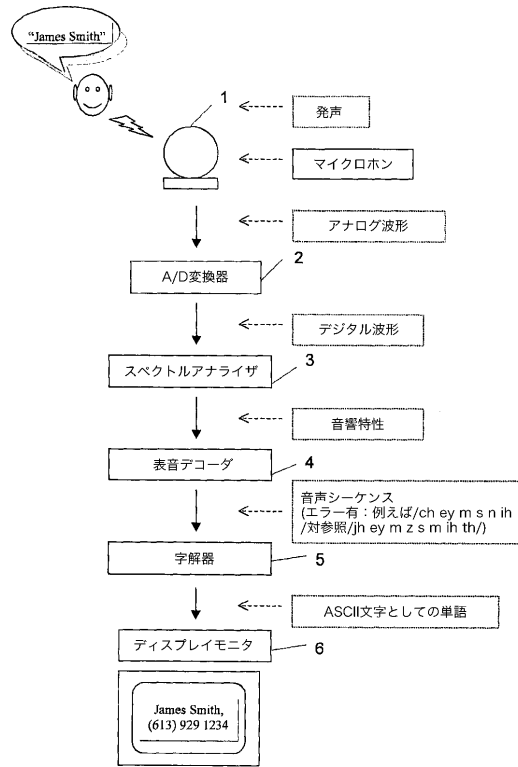
【符号の説明】

【0082】

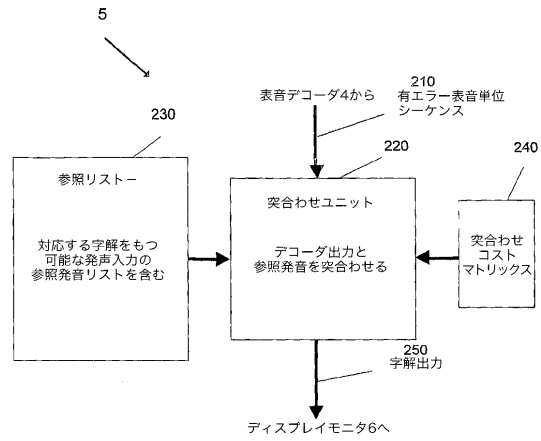
- 1 マイクロホン
- 2 アナログ - デジタル変換器
- 3 スペクトルアナライザ
- 4 表音デコード
- 5 字解器
- 6 ディスプレイモニタ
- 210 表音単位シーケンス
- 220, 320 突合わせユニット
- 230, 330 参照リスト
- 240, 340 突合わせコストマトリックス
- 250, 350 字解出力

40

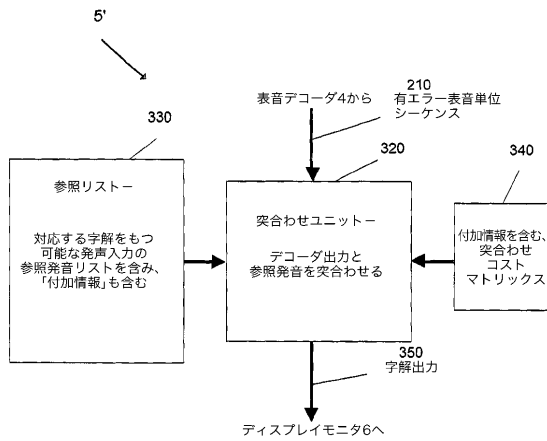
【図 1】



【図 2】



【図 3】



フロントページの続き

(74)代理人 100124855

弁理士 坪倉 道明

(74)代理人 100129713

弁理士 重森 一輝

(74)代理人 100174078

弁理士 大谷 寛

(74)代理人 230105223

弁護士 城山 康文

(72)発明者 ハント,メルヴィン ジェイ

イギリス国 ジーエル5 2 9 ユーエフ グロスターシャー州 チュルテンハム ウッドマンコー
ト デューイ クローズ 8

F ターム(参考) 5D015 BB02