

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3948260号
(P3948260)

(45) 発行日 平成19年7月25日(2007.7.25)

(24) 登録日 平成19年4月27日(2007.4.27)

(51) Int. Cl.	F I
G 1 O L 15/22 (2006.01)	G 1 O L 15/22 4 6 O Z
G 1 O L 15/18 (2006.01)	G 1 O L 15/18 2 O O Z
G 1 O L 15/00 (2006.01)	G 1 O L 15/00 2 O O A

請求項の数 7 (全 16 頁)

(21) 出願番号	特願2001-354521 (P2001-354521)	(73) 特許権者	000005821
(22) 出願日	平成13年11月20日(2001.11.20)		松下電器産業株式会社
(65) 公開番号	特開2002-221989 (P2002-221989A)		大阪府門真市大字門真1006番地
(43) 公開日	平成14年8月9日(2002.8.9)	(74) 代理人	100097445
審査請求日	平成16年9月1日(2004.9.1)		弁理士 岩橋 文雄
(31) 優先権主張番号	特願2000-355416 (P2000-355416)	(74) 代理人	100109667
(32) 優先日	平成12年11月22日(2000.11.22)		弁理士 内藤 浩樹
(33) 優先権主張国	日本国(JP)	(74) 代理人	100109151
			弁理士 永野 大介
		(72) 発明者	遠藤 充
			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	西崎 誠
			大阪府門真市大字門真1006番地 松下
			電器産業株式会社内

最終頁に続く

(54) 【発明の名称】 テキスト入力方法及びその装置

(57) 【特許請求の範囲】

【請求項 1】

音声で文章を入力するステップと、
音声認識処理により、入力された前記音声から1乃至数単語の単位である単語列の候補を作成する候補作成ステップと、

前記作成された候補を表示する表示ステップと、

前記表示された候補をユーザが選択する選択ステップとを有し、

前記候補作成ステップと前記表示ステップと前記選択ステップとを順次繰り返して、前記単語列の単位で前記入力された音声を認識する、テキスト入力方法であって、

前記候補作成ステップにおける前記単語列の候補は、

言語モデルによって得られる単語の連鎖確率を用いて単語が連結するか否かを判断し、単語を連結させると判断した場合は単語を連結し、単語を連結させないと判断した場合は単語を連結させない、単語の伸張処理をする伸張処理ステップ

を繰り返して得られる、テキスト入力方法。

【請求項 2】

前記候補作成ステップが更に前記伸張処理の繰り返し後に音響スコアによる候補のならばの更新処理を有する請求項1に記載のテキスト入力方法。

【請求項 3】

前記伸張処理を行った単語列の候補の数が、言語スコアの最上位から所定数に達したことにより前記伸張処理の繰り返しを終了する請求項1に記載のテキスト入力方法。

10

20

【請求項 4】

音声で文章の入力を行う入力部と、
前記入力部からの音声の特徴量を抽出する音声前処理部と、
確定された単語列から言語モデルを用いて、次に続く単語列候補を作成する単語候補作成部と、

前記抽出された特徴量と前記単語候補から言語モデルを用いて 1 乃至数単語の単語列候補を作成する単語列作成部と、

前記単語列候補を表示する表示部と、

ユーザが表示された前記単語列候補を選択する操作部と、

前記操作部により選択された単語列から次の単語候補の作成を前記単語候補作成部に指示する候補作成指示部とを有するテキスト入力装置であって、 10

前記単語列作成部によって作成される前記単語列候補は、

前記言語モデルによって得られる単語の連鎖確率を用いて単語が連結するか否かを判断し、単語を連結させると判断した場合は単語を連結し、単語を連結させないと判断した場合は単語を連結させない、単語の伸張処理を繰り返して得られる、テキスト入力装置。

【請求項 5】

請求項 4 に記載のテキスト入力装置を有する携帯電話。

【請求項 6】

音声で文章を入力するステップと、

音声認識処理により、入力された前記音声から 1 乃至数単語の単位である単語列の候補 20
を作成する候補作成ステップと、

前記候補を表示する表示ステップと、

前記表示された候補をユーザが選択する選択ステップとを有し、

前記候補作成ステップと前記表示ステップと前記選択ステップとを順次繰り返して、前記単語列の単位で前記入力された音声を認識することを実行させるためのプログラムであって、

前記候補作成ステップにおける前記単語列の候補は、

言語モデルによって得られる単語の連鎖確率を用いて単語が連結するか否かを判断し、単語を連結させると判断した場合は単語を連結し、単語を連結させないと判断した場合は単語を連結させない、伸張処理をする単語の伸張処理ステップ 30

を繰り返して得られる、プログラム。

【請求項 7】

請求項 6 に記載のプログラムを記録したコンピュータに読み取り可能な記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識を利用したテキスト入力方法、特に携帯電話等の小型機器におけるテキスト入力方法及びその装置に関する。

【0002】

【従来の技術】

40

従来、音声認識を利用したテキスト入力方法としては、単語あるいは文節単位で話者が発声し、その発声毎に音声認識する方法と、文あるいは文章単位で発声し文全体を一挙に音声認識する方法とがあった。

前者の場合、特開平 2-298997 号公報に記載されているように、話者が発声した後に所定数の候補をメニュー表示しその中から話者に選択させるようにしていた。しかし、この方法では、話者がいちいち文節単位で発声を区切り、その度に正しい単語を選ばなければならないので、入力操作が大変煩わしく、また時間がかかるという課題があった。

【0003】

一方後者の場合としては、たとえば「単語を認識単位とした日本語の大語彙連続音声認識」(情報処理学会論文誌, Vol.40, No.4, pp1395-1403, Apl.1999) に記載されたものが知 50

られている。

【0004】

図12にこの従来のテキスト入力方法の動作フローを示し、その動作を説明する。

【0005】

最初に、ユーザが音声を入力する（S1201）。次に、装置は自動的に認識結果の探索を行う。認識結果の探索においては、その装置は音素等の音響単位を接続しながら発声全体の音響スコアを求める。それと同時に単語等の言語単位の系列に対して言語スコアを求める。そして、その装置はそれらを統合したスコアの高い順に認識結果の序列をつける。通常、一発声は数単語ないし数十単語からなる文章である。その装置は精度の良い認識結果を出力するために、探索途中において単語候補の組み合わせを考慮した数多くの単語列候補を残している（S1202）。 10

【0006】

次に、その装置はその認識結果の序列における最上位の単語系列を入力した全ての音声について表示する（S1203）。次に、ユーザは表示された認識結果の内、自分の意図と異なる部分を修正する（S1204）。そして、ユーザによるすべての修正が終わったときに、その装置は一発声に対する入力操作を終了する（S1205）。

【0007】

【発明が解決しようとする課題】

しかし上記の従来技術では、文全体の認識処理を行った後に認識結果候補を修正するため、長い発声の場合等には認識処理の負担が大きく、記憶容量が多く必要となる。そのため、装置が小型化できないという課題があった。 20

【0008】

本発明は、小型化可能でかつ、1文以上の連続音声の入力が可能なテキスト入力方法を実現することを目的とする。

【0009】

【課題を解決するための手段】

上記課題を解決する本発明に係るテキスト入力方法及びその装置は、文あるいは文章単位で入力された発声を、単語あるいは文節単位でユーザが文頭から逐次、候補を選択し確定していく探索処理をするようにしたものである。

【0010】 30

これにより、多くの単語列候補の組み合わせを考慮した探索空間を保持しておく必要が無くなるため、記憶容量の大幅な削減と、音声認識処理量の削減が出来る。これにより装置の小型化が可能となる。更に、ユーザは1文以上の単位で連続音声の入力が出来るので、単語単位の入力のような煩わしさがなくなる。

【0011】

【発明の実施の形態】

本発明に係るテキスト入力方法は、音声で文章を入力するステップと、音声認識処理により、入力された前記音声から1乃至数単語の単位である単語列の候補を作成する候補作成ステップと、前記候補を表示する表示ステップと、前記表示された候補をユーザが選択する選択ステップとを有し、前記候補作成ステップと前記表示ステップと前記選択ステップとを順次繰り返して、前記単語列の単位で前記入力された音声を認識するものであって、前記候補作成ステップにおける前記単語列の候補は、言語モデルによって得られる単語の連鎖確率を用いて単語が連結するかどうかを判断し、単語を連結させると判断した場合は単語を連結し、単語を連結させないと判断した場合は単語を連結させない、単語の伸張処理をする伸張処理ステップを繰り返して得られるものである。 40

【0012】

これによって、入力音声の先頭からユーザによる候補の確定操作を行い、またその単位を数単語単位とすることで、システムが準備しなくてはならない候補が局所的な候補ですむために、少ない処理量と少ない記憶容量でテキスト入力が可能となり、装置の小型化が可能である。 50

【0013】

本発明に係るテキスト入力方法は、その候補作成ステップが更に前記伸張処理の繰返し後に音響スコアによる候補のならびの更新処理を有するものである。

【0014】

これによって、言語スコアと音響スコアを用いて候補の序列を付けることで、使用頻度が高く、かつ、発音が近い候補から順に候補を提示するので、ユーザが所望の候補を得るまでに目を通さなくてはならない候補の数が少なくてすむ。

【0015】

本発明に係るテキスト入力方法は、その伸長処理を行った単語列の候補の数が、言語スコアの最上位から所定数に達したことによりその伸長処理の繰返しを終了するものである。

10

【0016】

これによって、ユーザに提示する候補の数に制限を加えることにより、音声認識処理が不調に終わったときに、正しくない候補を延々と表示し続けることなしに、再度音声入力を促すなどの処理が可能になる。

【0017】

以下、本発明の実施の形態について、図1から図11までを用いて説明する。

【0018】

(実施の形態1)

図1は、本発明の一実施の形態となるテキスト入力装置のブロック構成図である。図1において、入力部101から取り込まれた入力音声は、音声前処理部102に入力され、A/D変換処理されたのち、特徴量抽出処理を行う。単語候補作成部103は、言語モデル104を参照して直前に確定している文節に続く単語候補を言語モデル104から所定数だけ作成する。ここで言語モデル104は単語系列における単語間の関係をモデル化したものである。最初の発声の場合には、操作部108から指示を受けた候補作成指示部109が単語候補作成部103に文頭であるという指示を伝える。単語候補作成部103はこの指示を受けると、言語モデル104を参照し、文頭に発声される確率の高い単語を候補として作成する。このように作成した単語候補を単語列作成部106に伝える。

20

【0019】

一方、作成部106は音声前処理部102から文単位で発声された音声の特徴量を受け取り、メモリ110に一旦格納する。作成部106は、単語候補作成部103からの単語候補に伸長処理及び、音響モデル105と単語辞書111を参照する音響スコア更新処理を行い、文節候補である単語列を所定の数だけ作成する。音響モデル105は音響的特徴をモデル化したものである。単語辞書111は認識されるべき単語が音響モデルが表す発音記号の例として登録されたものである。なお、伸長処理および音響スコア更新処理については、詳細に後述する。

30

【0020】

表示部107はその作成された単語列候補を表示する。ユーザは操作部108により、表示された候補の中から正しい文節を選択する。候補作成指示部109は操作部108からの指示に従い、選択された文節を単語列作成部106から受け取り、確定した文節として出力する。また一方、候補作成指示部109は単語候補作成部103へもその確定した文節を伝える。

40

【0021】

単語候補作成部103はその確定文節を受け、前述したように、言語モデル104を参照して次に続く単語候補を作成する。上記処理が入力された1文について終了するまで繰り返される。終了した後、メモリ110に格納されていた1文の特徴量データは消去される。

【0022】

図2は、本実施例の携帯電話によるマンマシンインタフェース図である。ボイスボタン201は音声認識の開始を装置に知らせるものである。候補ボタン202は文節候補の表

50

示や変更を要求するためのものである。表示画面203は確定したテキストや文節候補などを表示するものである。確定ボタン204は文節候補を確定するものである。

【0023】

図3に本発明のテキスト入力装置の動作の概要を表したフローチャートを示す。以下に、図1から図3を用いて、本発明の動作を説明する。

【0024】

初めにユーザはボイスボタン201を押し1文を発声し、音声を入力する。テキスト入力装置は、入力音声に対してA/D変換処理を行う。そして、変換された音声信号に対して例えば10msecごとのフレーム単位でLPCケプストラム係数等の特徴抽出処理を行う(S301)。

10

【0025】

次に、ユーザは候補ボタン202を押して、文節候補表示要求を行う(S302)。テキスト入力装置は、入力音声の特徴量と音響モデル及び言語モデルを用いて、文節の候補リストを作成し、表示画面203に上位の候補を1つ以上表示する(S303)。

【0026】

この文節候補リストは、音響スコアと言語スコアとの重み付けをした和である統合スコアが大きい順に単語列を並べたものである。ここで、単語列に対する音響スコアは、以下のようにして求めることができる。入力フレーム*i*、辞書フレーム*j*に対する音響スコア $s(i, j)$ は、数式(1)によって、計算できる。

【0027】

20

【数1】

$$as(i, j) = (x(i) - \mu(j))^t \sum (j)^{-1} (x(i) - \mu(j)) + \log | \sum (j) | \quad (1)$$

【0028】

ここで、 t は転置、 -1 は逆行列、 $x(i)$ は入力フレーム*i*に相当する入力ベクトル、 $((j))$ は辞書フレーム*j*に相当する特徴ベクトルの共分散行列と平均値ベクトルである。前述の音響モデルは具体的には、これら辞書フレームの共分散行列および平均値ベクトルの集合である。入力ベクトルはLPCケプストラム係数ベクトルのような入力音声の抽出された特徴ベクトルである。辞書フレームは、入力フレームに対応するとみなされた文節辞書に登録されている文節を音響モデルから取り出した同じく特徴ベクトルである。なお、特徴量データはLPCケプストラム係数ベクトルに限られず、MFCC(Mel Frequency Cepstral Coefficients)などが使用できる。

30

【0029】

単語の音響スコアは、DPマッチング等のマッチング手法により、入力フレームと辞書フレームの対応関係を求め、その対応関係を結んだ最適パス上の音響スコアを加算することにより求めることができる。さらに、単語列の音響スコアも、隣り合った単語の時間的な整合性を考慮しながら単語単位の音響スコアを加算することにより求めることができる。

40

【0030】

また、単語列に対する言語スコアは、以下のようにして求めることができる。

【0031】

前述の言語モデルは、具体的には、単語 $w(i)$ が*n*個の先行単語 $pre(i, n)$ の後に出現する連鎖確率 $P(w(i) | pre(i, n))$ の集合である。単語列に対する言語スコアは、言語モデルを参照し、各単語について先行単語を考慮して連鎖確率またはその対数値を求め、それらを加算することで得られる。

【0032】

このようにして、入力音声の特徴量と音響モデルとから音響スコアが得られ、単語列仮

50

説と言語モデルから言語スコアが得られる。それらを統合したスコアの高い単語列を文節候補としてリストに登録する。

【0033】

次に、ユーザは表示された文節の候補を確認し、所望の候補でなければ候補ボタン202を押して、次の候補を表示させる。所望の候補が表示された時点でユーザは確定ボタン204を押して文節を確定する(S304)。

【0034】

文節単位で確定操作を行い、発声の最後まで文節の確定が終わっていなければステップS302に戻り、最後の文節の確定が終わったところで終了する(S305)。

【0035】

以上のように、本発明においては、ユーザの文節候補の確定操作によって、候補を確定した後に、次に続く文節の候補作成をしていくために、それ以外の候補を保存する必要もなければ、認識処理をする必要もなくなる。これにより、装置が必要とする記憶容量が少なくて済み、装置を小型化できる。

【0036】

ここで、言語的な単位について考察する。形態素のように短い単位は、少ない種類数でカバー率を高くできるので装置の小型化に適している。しかし、ユーザが選択していく単位としては、文節等のより長い単位の方がわかりやすく好ましい。本発明は、言語的な最小単位として形態素を用いる。なお、本実施例では、人間とのインタラクションにおいてより好ましい形態素を適当に接続した文節単位に組み上げた例を提示する。この組み上げ処理を形態素の伸長処理と呼ぶ。

【0037】

以下に、単語列作成部106で行う文節候補作成処理について、図4から図6を用いて詳しく説明する。

【0038】

図4は、本発明の文節候補作成過程の処理手順を示したフローチャートである。本実施例では、最初に形態素単位の候補を伸長して文節候補リストを作った(S401~S406)。次にその結果に音響スコアを加味して最終的な文節候補リストを作成した(S407~S412)。図5は確定済みの文節「それを」500の次に接続する文節候補リストを伸長処理により作成した時の処理データの例である。図6はその伸長処理後に音響スコア更新をして文節候補リストを作成した時の処理データの例である。

【0039】

図5において、最初に確定済み文節「それを」500の次に接続する文節候補リスト510を作成する。それは、「それを」と全形態素の間の連鎖確率を、予め学習しておいた言語モデルにより求められる。この求めた形態素のリストを連鎖確率の高いものからソートしたものを文節候補リスト510とする。各文節候補には、伸長終了フラグ(図5、図8、図10、図11では「完了」と表示してある)として、今後伸長される可能性があることを表す0を初期値として与える(S401)。この伸長終了フラグは、伸長の可能性がなくなったときに「1」がセットされる。この状態では、文節候補は短すぎてわかりにくい。そこで、単語列作成部106はその文節候補とそれにつづく形態素の間に連鎖確率が比較的高いものを探して、その文節候補と接続し、より長い文節候補を作成する。

【0040】

そこで初めに、単語列作成部106は伸長する文節候補を決定する。リストの最上位から文節候補を参照し、伸長終了フラグが「0」である最初の文節候補を選ぶ(S402)。選択された候補はリスト511のようになる。

【0041】

次に、単語列作成部106は、伸長する文節候補とその候補に接続しうる各形態素との連鎖確率を求める。ここで、連鎖確率が所定の閾値未満の形態素、又は連鎖確率が句読点に対する連鎖確率よりも小さい形態素、並びに、句読点は、「その他の形態素」として一つにまとめ、それらの連鎖確率の和を求める。その連鎖確率は、「その他の形態素」の連

10

20

30

40

50

鎖確率となる（S403）。求めた連鎖確率はリスト512のようになり、「し」からの連鎖確率が比較的大きい「た」と「て」以外の確率は「（他）」としてまとめられている。図では、「その他の形態素」という概念に対応する記号として（他）を用いている。但し、伸長終了フラグ（「完了」）を表示したリスト510、520、および530については記号（他）を省略してある。（以降の図6、図8、図10、図11についても同じ。）

次に、伸長候補を作成する。「それを」→「し」の連鎖確率に、「し」→「た」の連鎖確率を掛けたものを、「それを」→「した」の連鎖確率とする。文節候補「し」は「した」に伸長されたことになる。同様にして、単語列作成部106は、伸長候補「して」を作成する。「その他の形態素」としてまとめたものは、後続の形態素への分岐が多い。つまり文節の境界として相応しいと考えられる。そのため、単語列作成部106は、「その他の形態素」については伸長が終了したとみなす。したがって、「し」のままとし、「それを」→「し」と、「し」→「（他）」の確率を掛け、それを連鎖確率とする。また、伸長終了フラグを「1」にする（S404）。その結果、伸長した候補のリスト513ができる。以上で、1回の伸長処理が終了する。

【0042】

次に、単語列作成部106は文節候補リストを更新する。即ち、単語列作成部106は文節候補リスト510から、伸長前候補511を除く。次に、単語列作成部106は、伸長後候補513を追加する。そして、それを連鎖確率の高い順位にしたがって並べ替える（S405）。その結果、更新された文節候補リスト520が得られた。

【0043】

次に、単語列作成部106は終了判定を行う。本実施例では、あらかじめ設定しておいた回数である100回の伸長処理を行ったところで終了とした（S406）。その伸長処理が100回未満のときには終了でないとして、S402へ戻る。このようにして伸長処理をつづけていくことにより、文節候補リスト530のように、「した」、「受けた」、「決めた」など、文節として適当な単位の候補が得られた。

【0044】

なお、終了判定は、伸長終了フラグが「1」にセットされている文節候補の数が連鎖確率の最上位から所定の数量に達したときに終了とすることが出来る。その他、「その他の形態素」の連鎖確率よりも大きな連鎖確率を有する伸長終了フラグが「0」の文節候補がなくなった時点で終了とすることも可能である。

【0045】

次に、音響スコアを加味して順位付けを行った文節候補リストの作成方法を説明する。

【0046】

図6において、確定済み文節「それを（終端時刻314）」600は、ステップS301において、ボイスボタン201を押した時刻を始端とし、「それを」を発声し終えた時刻（終端時刻）が314msであったことを示す。

【0047】

初めに、100回の伸長処理により作成された文節候補リスト530を元にして、連鎖確率を対数処理して言語スコアを求める。本実施例では、式（2）により言語スコアを連鎖確率から求めた。

【0048】

$$L = 20 \log_{10} l \quad (2)$$

但し、L：言語スコア、l：連鎖確率

音響スコアの初期値としては、適当に高い値（ここでは1.00）を設定した。また、言語スコアと音響スコアの和を統合スコアとした。そして、単語列作成部106は、統合スコアの高い順で文節候補リストをソートし、リスト610を求めた。また、音響マッチングにより得られる発声の終端時刻として、確定済み文節の終端時刻314を初期値として各候補にセットした（S407）。

【0049】

10

20

30

40

50

次に、単語列作成部106は音響スコアの値を更新する候補を決定した。即ち、リストの最上位から文節候補を参照し、未だ、音響スコアの更新が行われていない最初の未更新候補を選んだ(S408)。なお、音響スコアの更新が行われたか否かの判定は、確定済みの文節の終端時刻と文節候補の終端時刻とが同じであるか否かにより行われる。リスト610においては、「した」が選ばれた。

【0050】

次に、時刻314ms付近を始端として、「した」に対する音響スコアを計算する(S409)。音響マッチングの結果として、始端時刻314ms、終端時刻643msという音声区間で、比較的音響スコアの高い、0.89が(1)式により得られた(リスト612)。

10

【0051】

その音響マッチングの代表的な方法は、音声信号のAD変換、特徴パラメータへの変換、音響モデルとの局所距離の計算、DPマッチングによる局所距離の累積計算、という処理過程からなる。これらの処理は、ステップS301の音声入力において一括して行う処理と、ステップS409の音響スコアの計算において逐次的に行う処理に分散させることができる。一括して行う処理は重複計算を防ぐので処理量の点で有利であり、逐次的に行う処理は途中の結果を保存しておく必要がないので記憶容量の点で有利である。どのように分散させるかは、実際のハードウェア構成に応じて決められるべきものである。本実施例においては、音響モデルとの局所距離の計算及び、DPマッチングによる局所距離の累積計算処理をステップS409にて行った。

20

【0052】

次に、単語列作成部106は、文節候補の値を更新した。即ち、音響スコアを0.89に更新し、言語スコアと音響スコアの和を求めて統合スコアを更新した。文節候補の終端時刻はマッチング区間を参照して更新した(S410)。その結果、新しい候補はリスト613になった。

【0053】

次に、文節候補リストを更新する。即ち、単語列作成部106は文節候補リスト610から、音響スコア更新前候補611を削除する。次に、単語列作成部106は、更新後候補613を文節候補リスト610に追加する。そして、統合スコアの高い順に並べ替える(S411)。その結果、文節候補リスト620が得られた。以上の処理を、音響スコア更新処理と呼ぶ。

30

【0054】

次に、単語列作成部106は終了判定を行う。本実施例では、あらかじめ設定しておいた回数である100回数の音響スコア更新処理を行ったときに終了とした(S412)。100回未満である場合は終了でないとして、ステップS408へ戻る。このようにして音響スコア更新処理をつづけていくことにより、使用頻度が高く、かつ、発声との音響マッチングのスコアも高い文節候補リストが作成された。このリストはスコアの高い順に文節候補が並べられた形をしている。

【0055】

なお、終了判定は、終端時刻が確定済み時刻と異なっている文節候補の数が統合スコアの最上位から所定の数量に達したときに終了とすることが出来る。

40

【0056】

テキスト入力装置は、以上のようにして得られた文節候補リストの最上位の文節候補から表示する。これにより、テキスト入力装置は、その時に対象にしている文節に特定した音声認識処理で済むことにより少ない処理量と少ない記憶容量でテキスト入力が可能となる。また、統合スコアの高い順に上位の候補を1つ以上表示することができることにより、ユーザが所望の候補を得るまでに提示される候補の数が少なくて済むようになった。そして更に、文節単位で候補を表示することにより、ユーザにとって分かりやすい選択提示となった。

【0057】

50

(実施の形態 2)

この実施例は、単語列作成部 106 で行う伸長処理と音響スコア更新処理とが同時進行的に文節候補リストを更新していく点が実施例 1 と異なる。その他のテキスト入力装置のブロック構成図、マンマシンインターフェースなどは全て同じものである。

【0058】

図 7 は、本発明の実施の形態 2 となるテキスト入力装置の文節候補作成過程の処理手順を示したフローチャートである。

【0059】

図 8 は、確定済みの文節「それを」500 の次に接続する文節候補リストを、伸長処理と音響スコア処理とを交互に繰り返すことにより作成したときの処理データの流れを示したものである。

【0060】

以下に図 7 と図 8 を用いて詳しく説明する。

【0061】

最初に、確定済みの文節「それを」500 から次に接続する文節候補リスト 801 を作成するステップ S701 は実施例 1 のステップ S401 と同じである。次にこの候補リスト 801 の連鎖確率に対数処理した言語スコアと音響スコアを加えた統合スコアを求めて、音響スコア付き候補リスト 802 を作成した (S702)。次に、その音響スコア付き候補リストの最上位から伸長処理が未処理の候補を検索し、最初の候補を伸長処理候補とする (S703)。リスト 802 においては、その候補は「し」となる。この候補に対し、言語モデルを使用して S407 と同様に「し」から連鎖確率が比較的大きい「た」と、「て」及び「(他)」が求まる (S704)。これらの文節候補を実施例 1 と同様に候補リスト 802 に追加し、統合スコアの大きい順に並べ替えて新たな文節候補 803 を得る (S705)。

【0062】

次に、この候補リストの最上位から終端時刻が確定済み文節候補の終端時刻と同じ候補を検索し、最初の候補の音響スコアを求める (S706)。リスト 803 においては、「受け」がそれに当たる。この候補に対して、S409 と同様にして音響スコアを求めると、始端時刻 314 ms、終端時刻 640 ms という音声区間で、音響スコア 0.02 が得られた。それを文節候補リスト 803 に反映させ (S707)、統合スコアの大きい順に並べ替えて新たな文節候補 804 を得た (S708)。以上のステップ S703 からステップ 708 までの処理をあらかじめ設定しておいた回数繰り返し、文節候補リスト 806 を得た。本実施例では 100 回繰り返した。本実施例の結果は、実施例 1 の結果と同じものになった。

【0063】

なお、本実施例では終了判定は、所定回数の伸長処理と音響スコア更新処理を繰り返した時点で終了と判定したが、伸長終了フラグが「1」にセットされている文節候補の数が最上位から所定の数量に達したときに終了とすることが出来る。

【0064】

また、終了判定は、終端時刻が確定済み時刻と異なっている文節候補の数が統合スコアの最上位から所定の数量に達したときに終了とすることも可能である。

【0065】

あるいは、終了判定は、上記伸長終了フラグによる方法、あるいは終端時刻による方法のいずれか早い方で終了とすることも出来る。

【0066】

(実施の形態 3)

この実施例は、単語列作成部で行う伸長処理と音響スコア更新処理とが実施例 1 と逆の順で行われる点が実施例 1 と異なる。その他のテキスト入力装置のブロック構成図、マンマシンインターフェースなどは全て同じものである。

【0067】

10

20

30

40

50

図9は、本発明の実施の形態3となるテキスト入力装置の文節候補作成過程の処理手順を示したフローチャートである。

【0068】

図10は、確定済みの文節「それを」500の次に接続する文節候補リストを、音響スコア処理を済ませた後に、伸長処理を行うことにより作成したときの処理データの流れを示したものである。

【0069】

以下に図9と図10を用いて詳しく説明する。

【0070】

最初に、確定済みの文節「それを」500から次に接続する文節候補リスト1001を作成するステップS901は実施例1のステップS401と同じである。次にこの候補リスト1001の連鎖確率に対数処理した言語スコアと音響スコアを加えた統合スコアを求めて、仮の音響スコア付き候補リスト1002を作成した(S902)。次に、候補リスト1002の最上位から確定済み文節候補の終端時刻314と異なる終端時刻の候補を検索し、最初の候補を音響スコア計算候補として決定する(S903)。この候補に対する音響スコアをS409と同様に計算する(S904)。リスト1002においては、「し」が選択され音響スコアを計算すると、始端時刻314ms、終端時刻510msという音声区間で、比較的音響スコアの高い、0.90が得られた(S904)。それを文節候補リスト1002に反映させ(S905)、統合スコアの大きい順に並べ替えて新たな文節候補1003を得た(S906)。上記のステップS903からステップS906までをあらかじめ設定してある回数繰り返した(S907)。本実施例では、100回繰り返し、候補リスト1004を得た。

【0071】

次に、この候補リスト1004に言語モデルを使用して、伸長処理を行った。初めに、候補リスト1004の最上位から伸長終了フラグが「1」にセットされていない最初の候補を選択する(S908)。

【0072】

次に、言語モデルの連鎖確率を参照し(S909)、ステップS403と同様に「し」から連鎖確率が比較的大きい「た」と、「て」及び「(他)」が求まる(S910)。

【0073】

これらの文節候補を実施例1と同様に候補リスト1004に追加し、統合スコアの大きい順に並べ替えて新たな文節候補1005を得る(S911)。

【0074】

これらのステップS908からS911をあらかじめ設定してある回数である100回繰り返し(S912)、文節候補1006を得た。この結果は、音響スコアとして一番目の形態素のみの値を用いるので、実施例1及び実施例2とは別の結果になるが、上位に表示される文節は同様のものが得られた。

【0075】

なお、本実施例では、音響スコア更新処理の終了判定は、更新処理を所定回数繰り返すことで終了としたが、終端時刻が確定済み時刻と異なっている文節候補の数が統合スコアの最上位から所定の数量に達したときに終了とすることも可能である。

【0076】

また、伸長処理の終了判定は、伸長終了フラグが「1」にセットされている文節候補の数が最上位から所定の数量に達したときに終了とすることが出来る。

【0077】

図11は図10と同じく、音響スコア更新処理を済ました後に、伸長処理を行ったものである。しかし、この処理データ例は、この伸長処理において、ステップ910の伸長済み候補の作成を行った後に、連結された形態素の分の音響スコアを計算し、伸長前の音響スコアに加えた点が図10と異なる。

【0078】

10

20

30

40

50

図 11 の候補リスト 1105 の「した」及び「して」の終端時刻がそれぞれ“643”及び“640”に更新されている。このように、音響スコアを伸長処理に併せて更新することは、文節候補の正確な音響スコアが求まりより好ましい。

【0079】

なお、実施例 1 から実施例 3 は、文節候補リストを作成し、確定ボタンの入力により候補を確定した後に次の文節候補を作成するという例で説明を行った。しかし、ユーザによる候補の確定した時点から、次の文節候補を表示するまでの時間を短縮するために、候補を表示している時点でその表示候補を用いて次の文節候補の作成処理を行うことも可能である。あるいは、候補リストの表示に所望のものが無い場合、再度ボイスボタンを押し、その認識させたい文節のみを発声することにより、候補の再作成を装置に行わせることも出来る。

10

【0080】

【発明の効果】

以上のように本発明によれば、文あるいは文章単位で入力された発声を、単語あるいは文節単位でユーザが文頭から逐次、候補を選択し確定していく探索処理をすることにより、装置の小型化と、音声入力の煩わしさの軽減を両立したテキスト入力の実現できるという有利な効果が得られる。

【図面の簡単な説明】

【図 1】 本発明の実施の形態 1 によるテキスト入力装置のブロック構成図

【図 2】 本発明の実施の形態 1 によるテキスト入力装置のマンマシンインタフェース図

20

【図 3】 本発明の実施の形態 1 によるテキスト入力装置の動作を示すフローチャート

【図 4】 本発明の実施の形態 1 によるテキスト入力装置の文節候補作成過程の処理手順のフローチャート

【図 5】 本発明の実施の形態 1 によるテキスト入力装置の伸長処理過程の処理データの例を示す図

【図 6】 本発明の実施の形態 1 によるテキスト入力方法の音響スコア更新過程の処理データの例を示す図

【図 7】 本発明の実施の形態 2 によるテキスト入力装置の文節候補作成過程の処理手順のフローチャート

【図 8】 本発明の実施の形態 2 によるテキスト入力装置の文節候補作成過程の処理データの例を示す図

30

【図 9】 本発明の実施の形態 3 によるテキスト入力装置の文節候補作成過程の処理手順のフローチャート

【図 10】 本発明の実施の形態 3 によるテキスト入力装置の文節候補作成過程の処理データの例を示す図

【図 11】 本発明の実施の形態 3 によるテキスト入力装置のより好ましい文節候補作成過程の処理データの例を示す図

【図 12】 従来のテキスト入力方法のフローチャート

【符号の説明】

101 入力部

40

102 音声前処理部

103 単語候補部

104 言語モデル

105 音響モデル

106 単語列作成部

107 表示部

108 操作部

109 候補作成指示部

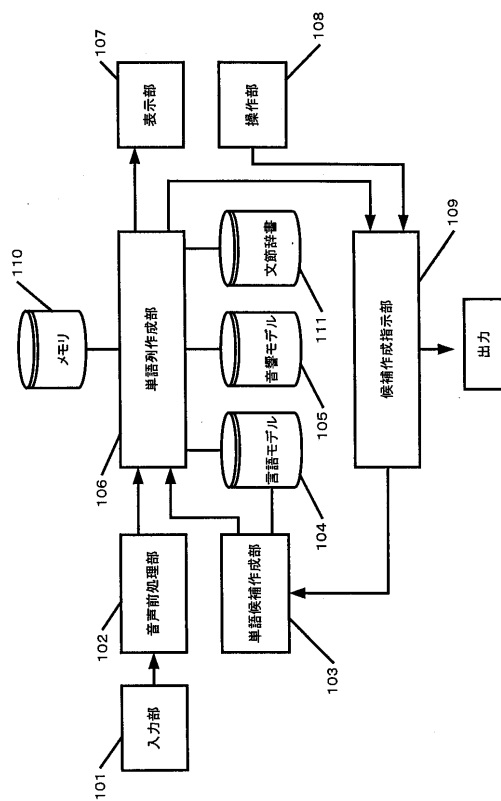
110 メモリ

201 ボイスボタン

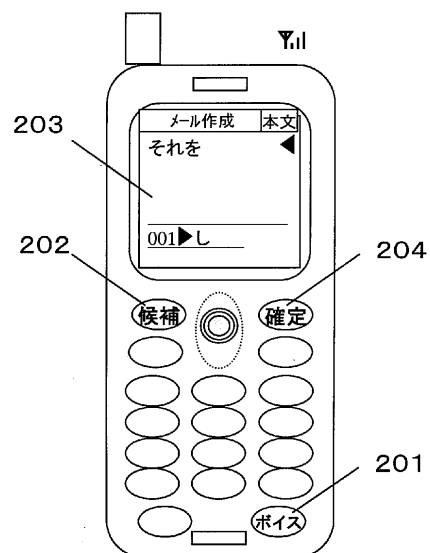
50

- 202 候補ボタン
- 203 表示画面
- 204 確定ボタン

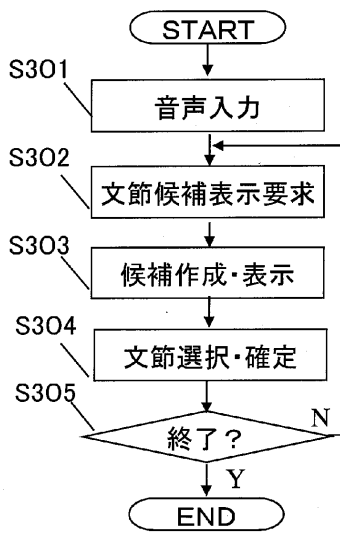
【図 1】



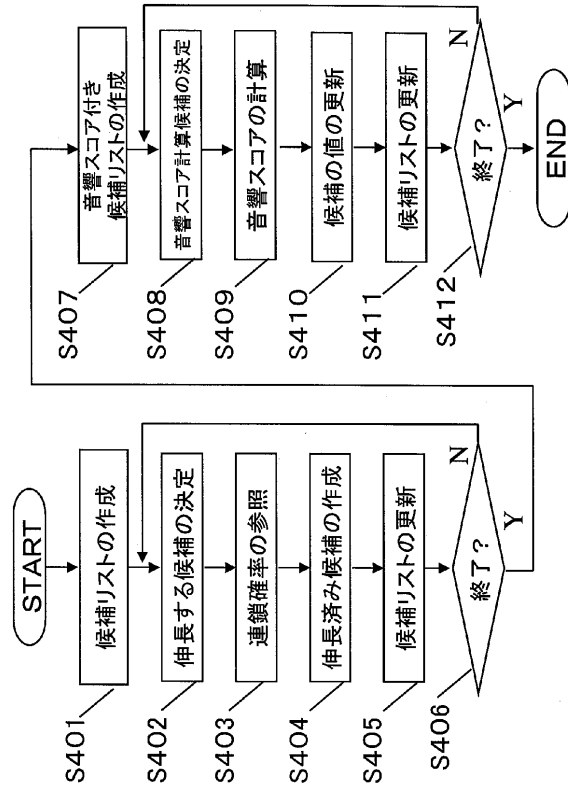
【図 2】



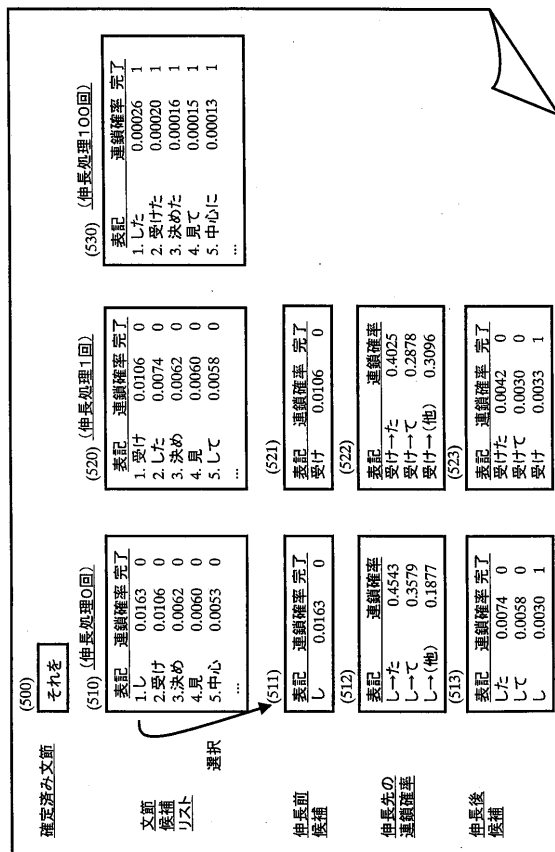
【図 3】



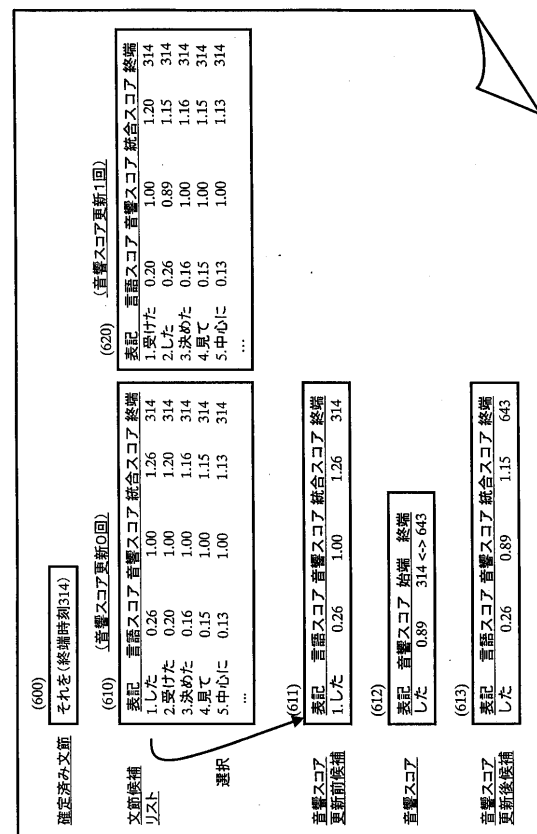
【図 4】



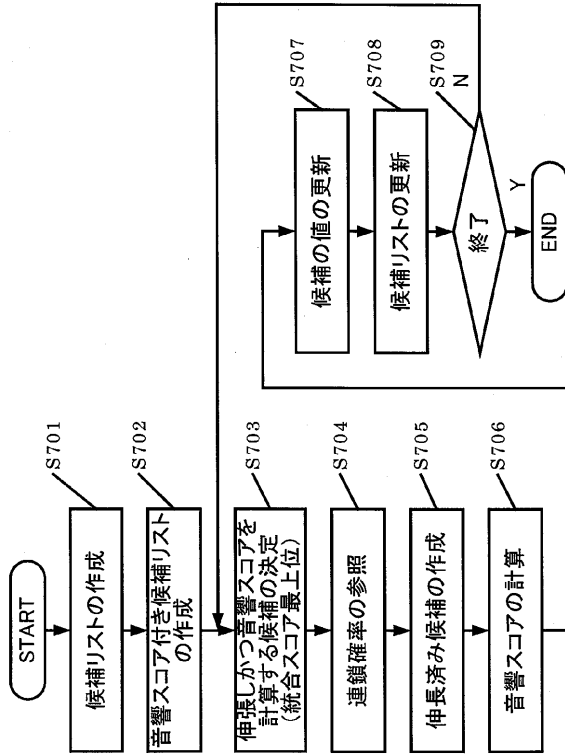
【図 5】



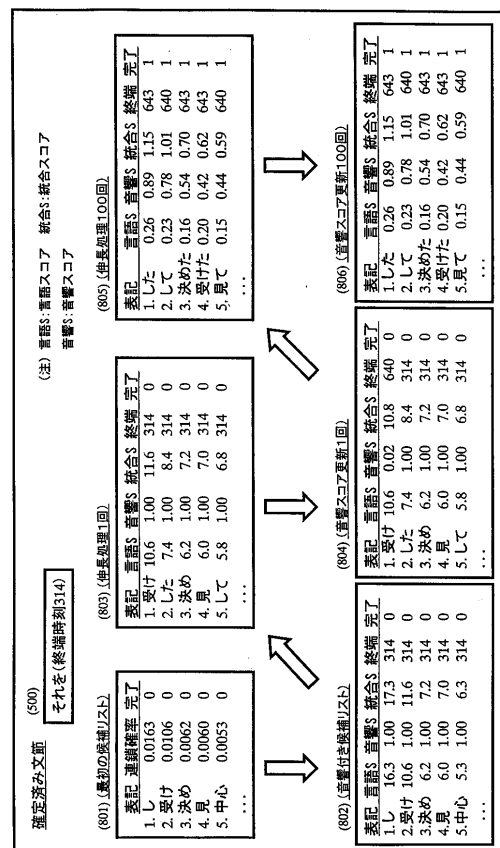
【図 6】



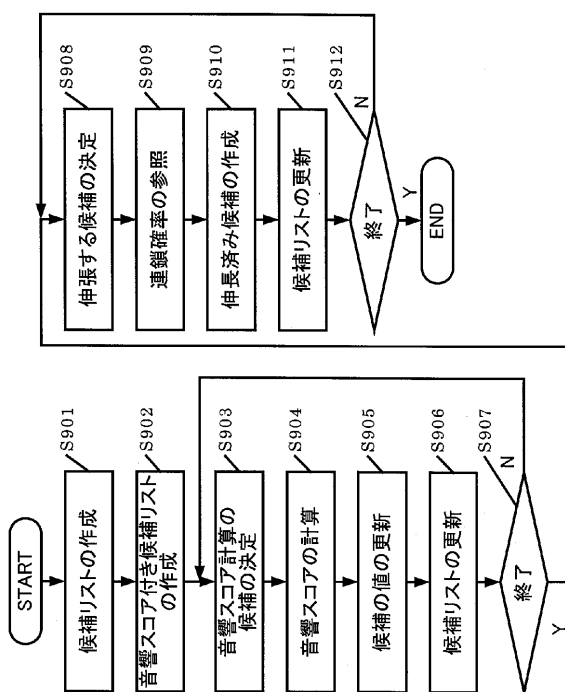
【図 7】



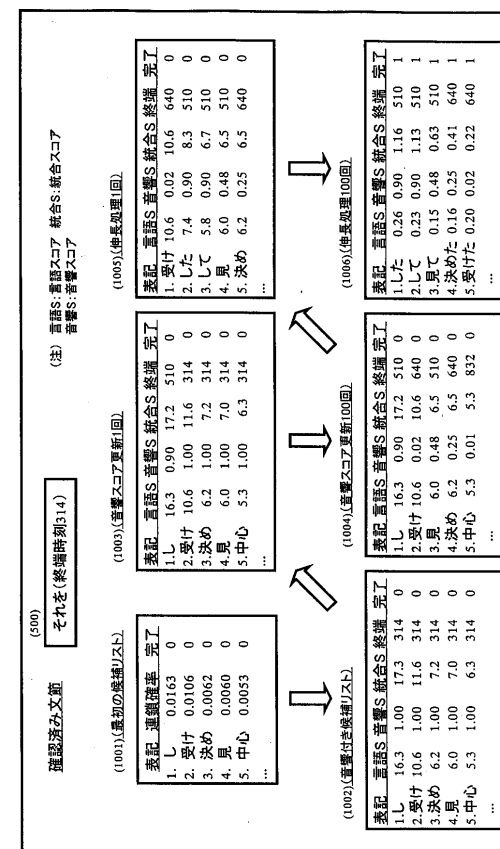
【図 8】



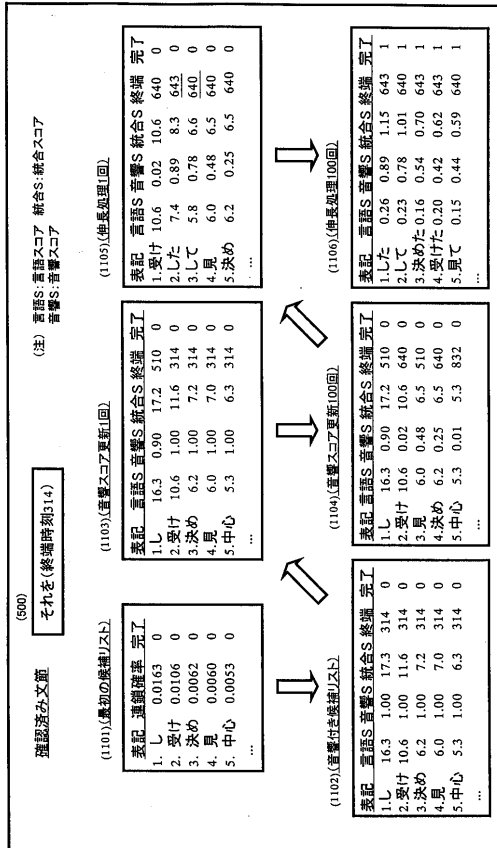
【図 9】



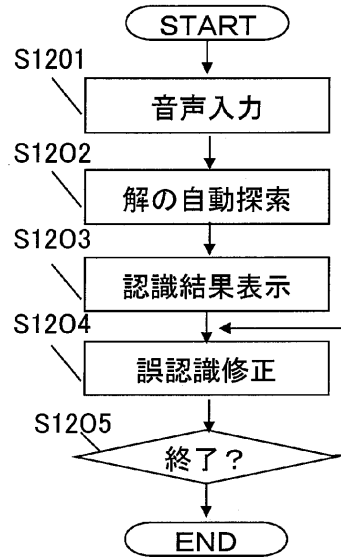
【図 10】



【図 1 1】



【図 1 2】



フロントページの続き

(72)発明者 齋藤 夏樹

大阪府門真市大字門真1006番地 松下電器産業株式会社内

審査官 榎本 剛

(56)参考文献 特開平06-282293 (JP, A)

特許第3440840 (JP, B2)

特開2000-242638 (JP, A)

特開2000-056796 (JP, A)

特開平11-168552 (JP, A)

特開平03-160868 (JP, A)

特開平08-160988 (JP, A)

特許第3758241 (JP, B2)

特開2001-249687 (JP, A)

森他, 構文構造を反映した確率的言語モデル, 情報処理学会研究報告 [音声言語情報処理], 日本, 2000年 7月14日, Vol.2000, No.64, 2000-SLP-32, 32-3, p.13-18

中川, 確率モデルによる音声認識, 電子情報通信学会, 日本, 1998年 7月 1日, p.109~128

(58)調査した分野(Int.Cl., DB名)

G10L 15/00-15/28

JSTPlus(JDream2)

IEEE