

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4684409号
(P4684409)

(45) 発行日 平成23年5月18日 (2011.5.18)

(24) 登録日 平成23年2月18日 (2011.2.18)

(51) Int.Cl.

F I

G 1 O L 15/18 (2006.01)

G 1 O L 15/18 3 O O H

請求項の数 4 (全 15 頁)

(21) 出願番号 特願2000-379315 (P2000-379315)
 (22) 出願日 平成12年12月13日 (2000.12.13)
 (65) 公開番号 特開2001-175277 (P2001-175277A)
 (43) 公開日 平成13年6月29日 (2001.6.29)
 審査請求日 平成19年3月9日 (2007.3.9)
 (31) 優先権主張番号 99124759.4
 (32) 優先日 平成11年12月13日 (1999.12.13)
 (33) 優先権主張国 欧州特許庁 (EP)

(73) 特許権者 598094506
 ソニー インターナショナル (ヨーロッ
 パ) ゲゼルシャフト ミット ベシュレ
 ンクテル ハフツング
 ドイツ連邦共和国 10785 ベルリン
 ケンパーブラッツ 1
 (74) 代理人 100067736
 弁理士 小池 晃
 (74) 代理人 100086335
 弁理士 田村 榮一
 (74) 代理人 100096677
 弁理士 伊賀 誠司

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

(57) 【特許請求の範囲】

【請求項 1】

通常型の認識文法を複数の高混同可能性ワードクラスに分類し、上記複数の高混同可能性ワードクラスから音素又は音節の組み合わせとして構成されるサブワードにより記述された高混同可能性ワードクラス用サブワードユニット文法を生成し、上記高混同可能性ワードクラス用サブワードユニット文法と上記通常型の認識文法の分類されなかった低混同可能性部分とを併合し、低混同可能性認識文法を生成する文法生成ステップと、

上記低混同可能性認識文法を用いて、入力された音声フレーズからワード及びサブワードユニットから構成される推測グラフを生成する音声認識ステップと、

上記推測グラフに対する制約された検索処理を実行し、認識すべき音声フレーズを表す最終ワードグラフを出力する制約検索処理ステップとを有し、

上記制約検索処理ステップでは、上記高混同可能性ワードクラス用サブワードユニット文法と、低混同可能性部分と高混同可能性部分との間の意味論的及び語用論的な制約情報とを利用して、上記推測グラフからサブワードユニットを削除し、ワードから構成される最終ワードグラフを生成する音声認識方法。

【請求項 2】

上記制約検索処理ステップでは、上記制約情報に基づいて、上記推測グラフから候補を削減し、最も可能性が高いフレーズに対応する分岐のないグラフを生成する請求項 1 記載の音声認識方法。

【請求項 3】

10

20

上記制約検索処理ステップでは、上記音声フレーズにおける上記サブフレーズを除く残りの部分のうちの少なくとも1つの部分に上記低混同可能性認識文法における語彙を制約する請求項1又は2記載の音声認識方法。

【請求項4】

通常型の認識文法を複数の高混同可能性ワードクラスに分類し、上記複数の高混同可能性ワードクラスから音素又は音節の組み合わせとして構成されるサブワードにより記述された高混同可能性ワードクラス用サブワードユニット文法を生成し、上記高混同可能性ワードクラス用サブワードユニット文法と上記通常型の認識文法の分類されなかった低混同可能性部分とを併合し、低混同可能性認識文法を生成する文法生成部と、

上記低混同可能性認識文法を用いて、入力された音声フレーズからワード及びサブワードユニットから構成される推測グラフを生成する音声認識回路と、

上記推測グラフに対する制約された検索処理を実行し、認識すべき音声フレーズを表す最終ワードグラフを出力する制約検索処理回路とを有し、

上記制約検索処理回路では、上記高混同可能性ワードクラス用サブワードユニット文法と、低混同可能性部分と高混同可能性部分との間の意味論的及び語用論的な制約情報とを利用して、上記推測グラフからサブワードユニットを削除し、ワードから構成される最終ワードグラフを生成する音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識方法及び音声認識装置に関し、特に、可能性のある候補の集合内の検索処理の処理負担を軽減する音声認識方法及び音声認識装置に関する。

【0002】

【従来の技術】

近年、連続的に発話された音声を自動的に認識するための装置及び方法の重要性が高まっている。情報サービス、顧客サポートサービス等、様々なサービス分野において、ユーザからの問い合わせに自動的に応答する装置を利用すれば、このためのコストを大幅に削減することができる。

【0003】

音声認識を自動的に行う装置及び方法においては、発話の速度、イントネーション、発音、背景の雑音等、ユーザ毎に条件が異なる発話の入力に対して、発話を正しく認識及び理解することが重要となる。

【0004】

自動電話対応サービス、スケジュール情報サービス等に関連する様々な手法及び装置が提案されている。これら手法及び装置は、ユーザの発音が非常に限定された狭い範囲内にある場合にのみ、正しく動作する。すなわち、これら手法及び装置は、非常に狭い語彙範囲及び音声条件のみを対象として設計されている。

【0005】

語彙量の多い音声認識の分野においては、多くの手法及び装置は、次のように動作する。まず、音声フレーズを受信し、この音声フレーズを表す信号が生成される。次に、デジタル化、フーリエ分析、及びその他の信号評価技術等を含む所定の規則により、音声フレーズを表す信号が前処理される。この前処理により得られた結果は、次の処理のために保存される。

【0006】

次に、この前処理された信号に基づいて、少なくとも1組の一連の推測音声要素(hypothetic speech elements)が生成される。この一連の推測音声要素に基づいて、受信された音声フレーズに対応する候補となる少なくとも一組の一連のワードが判定される。この一連の判定は、推測音声要素に所定の言語モデルを適用して行われる。

【0007】

【発明が解決しようとする課題】

上述のような広範囲の語彙に対応する従来の音声認識方法及び音声認識装置は、構成が複雑であり、膨大な量の候補となる発話の断片（fragment）又は要素（element）を検索し、検査する必要があった。発話の主題（subject matter）すなわち語彙の範囲を限定しない場合、弁別的な検索法（distinct searching technique）を用いて、音声要素又は音声断片の可能性のある全ての候補を評価する必要があった。

【 0 0 0 8 】

そこで、本発明は、上述の課題に鑑みてなされたものであり、適応される言語モデルにおいて、検索処理の負担を軽減し、これにより短時間で信頼性の高い音声認識を行うことができる音声認識方法及び音声認識装置を提供することを目的とする。

【 0 0 0 9 】

【課題を解決するための手段】

上述の目的を達成するために、本発明に係る音声認識方法は、通常型の認識文法を複数の高混同可能性ワードクラスに分類し、上記複数の高混同可能性ワードクラスから音素又は音節の組み合わせとして構成されるサブワードにより記述された高混同可能性ワードクラス用サブワードユニット文法を生成し、上記高混同可能性ワードクラス用サブワードユニット文法と上記通常型の認識文法の分類されなかった低混同可能性部分とを併合し、低混同可能性認識文法を生成する文法生成ステップと、上記低混同可能性認識文法を用いて、入力された音声フレーズからワード及びサブワードユニットから構成される推測グラフを生成する音声認識ステップと、上記推測グラフに対する制約された検索処理を実行し、認識すべき音声フレーズを表す最終ワードグラフを出力する制約検索処理ステップとを有し、上記制約検索処理ステップでは、上記高混同可能性ワードクラス用サブワードユニット文法と、低混同可能性部分と高混同可能性部分との間の意味論的及び語用論的な制約情報とを利用して、上記推測グラフからサブワードユニットを削除し、ワードから構成される最終ワードグラフを生成する。

また、本発明に係る音声認識装置は、通常型の認識文法を複数の高混同可能性ワードクラスに分類し、上記複数の高混同可能性ワードクラスから音素又は音節の組み合わせとして構成されるサブワードにより記述された高混同可能性ワードクラス用サブワードユニット文法を生成し、上記高混同可能性ワードクラス用サブワードユニット文法と上記通常型の認識文法の分類されなかった低混同可能性部分とを併合し、低混同可能性認識文法を生成する文法生成部と、上記低混同可能性認識文法を用いて、入力された音声フレーズからワード及びサブワードユニットから構成される推測グラフを生成する音声認識回路と、上記推測グラフに対する制約された検索処理を実行し、認識すべき音声フレーズを表す最終ワードグラフを出力する制約検索処理回路とを有し、上記制約検索処理回路では、上記高混同可能性ワードクラス用サブワードユニット文法と、低混同可能性部分と高混同可能性部分との間の意味論的及び語用論的な制約情報とを利用して、上記推測グラフからサブワードユニットを削除し、ワードから構成される最終ワードグラフを生成する。

【 0 0 1 0 】

さらに、この音声認識方法において、一連のワードを判定するステップは、受信された音声フレーズ内のシードサブフレーズ（seed sub-phrase）に含まれている可能性がある少なくとも1つのサブワード、ワード、又はワードの組み合わせを判定するステップと、サブフレーズ間の追加的な対の及び／又は高次の情報を用いて、シードサブフレーズに適合するワード又はワードの組み合わせを受信された音声フレーズに含まれる少なくとも1つの第1の連続するサブフレーズとして判定し、検索処理の負担を軽減するステップとを有する。

【 0 0 1 1 】

本発明では、入力された音声フレーズにおいて、適切な高い信頼度、すなわち適切な確率で認識できる少なくとも1つのシードサブフレーズを先に認識する。このシードサブフレーズに対する1又は複数の可能な候補は、この確率に基づいて判定又は評価できる。

【 0 0 1 2 】

続いて、上述のようにして判定された最初サブフレーズ又はシードサブフレーズと、入力

10

20

30

40

50

された音声フレーズにおける他の可能なサブフレーズ間の関係に関する情報を評価し、採用されている言語モデルにおける追加的情報を用いて、不必要な検索処理を行うことなく、その他のサブフレーズを記述及び判定する。サブフレーズと追加的情報との間の関係を利用することにより、最も見込みが高いサブフレーズの候補のみを検索でき、シードサブフレーズ又はより高次のサブフレーズに対して、意味的に関連しない可能性が高いサブフレーズの候補を検索処理において無視することができる。

【0013】

すなわち、本発明に係る音声認識方法によれば、高度な信頼性をもって認識できる音声セグメントを用いて、採用されている言語モデルのみでは適切に制限できない音声信号の他の領域における検索処理を制約する。

【0014】

また、本発明の具体例においては、シードサブフレーズは、少なくとも1つの連続する推測音声要素に所定の言語モデルを適用して得られ、追加的な一対の及び／又は高次の情報は、言語モデルから得られる。すなわち、判定処理は、一連の推測音声要素又は音声断片に所定の言語モデルを適用することにより実行される。

【0015】

また、本発明の具体例においては、言語モデル内の意味論的情報及び／又は語用論的情報をサブフレーズ間の追加的情報として使用する。

【0016】

意味論的情報は、情報コンテンツ自体に関する情報を記述する。語用論的情報は、サブフレーズに相互関連する実際の状況、適用例、動作等に関する情報を表す。すなわち、語用論的情報は、サブフレーズ間の具体的な結合関係に基づく情報である。

【0017】

さらに、本発明の好ましい具体例においては、言語モデル内の追加的情報は、サブフレーズの前置詞的關係を記述するために使用される。

【0018】

特に、これら前置詞的關係は、特にサブフレーズにより記述される位置的關係、時間的關係、その他のあらゆる関係を表す。

【0019】

さらに、本発明の好ましい具体例においては、言語モデル内の追加的情報は、2個1組、3個1組、又はそれ以上の個数を1組とするサブフレーズ間の関係を記述するために使用される。

【0020】

したがって、言語モデル内に主語／述語の関係のみならず、センテンス／サブセンテンスの關係に関する記述を実現することができる。

【0021】

本発明に基づく手法は、特に、認識処理に使用される言語モデルにおいて、信号の認識を比較的確実に行える領域、すなわち、ワード混同可能性が比較的低い領域が存在する場合に、その領域を利用して、正しく認識できる可能性が比較的低い、すなわちワード混同可能性が高い領域の認識を行うものである。

【0022】

したがって、言語モデルは、好ましくは、少なくとも混同可能性が低い弁別的クラスの音声要素又は断片を表す低混同可能性部分と、混同可能性が高い弁別的クラスの音声要素又は断片を表す高混同可能性部分とから構成される認識文法を含んでいるとよい。

【0023】

ここで、混同可能性とは、通常の検索グラフ又は検索ツリーにおいて実行すべき検索処理の複雑性及び深さを表す。すなわち、混同可能性は、検索グラフ又は検索ツリーの分岐レベルを表す。

【0024】

本発明に基づく手法は、発話において音声要素が高い信頼性で認識される少なくとも1つ

10

20

30

40

50

の領域を識別することにより正しい候補の検索処理を単純化する。入力された音声フレーズにおけるこの他の部分又はサブフレーズは、混同可能性が高い部分、又は複雑性が高い部分として取り扱われる。

【0025】

すなわち、適切な候補を認識されたサブフレーズとして検索する処理においては、フレーズ全体から、高い信頼性で分析及び認識できる少なくとも1つの混同可能性が低い又は複雑性が低い部分を抽出する。入力された音声フレーズにおける残りの部分又はサブフレーズは、混同可能性が高い部分、又は複雑性が高い部分として取り扱われる。

【0026】

さらに、本発明の好ましい具体例では、認識処理において、文法又は言語モデルの所定のセグメントを「低混同可能性領域」、すなわち音声要素を比較的高い信頼度で認識できる領域としてマークする。

10

【0027】

他の具体例においては、例えば、確実性を数値的に測定する等して、認識処理自体を分析することにより高い信頼度を有するセグメントを認識する。

【0028】

また、本発明の好ましい具体例においては、音声要素又は断片のクラスとしてワードクラスを用いる。

【0029】

ここで、例えば音素や音節等、より細かい音声断片又は要素を使用してもよい。しかしながら、ワード又はサブワード等の比較的大きな要素を使用することにより抽出処理が容易になる。これは、ワード又はサブワードの組み合わせが、音素又は音節の組み合わせに比べて、認識すべき実際のフレーズにより近いためである。

20

【0030】

本発明の好ましい具体例においては、言語モデルは、通常型の認識文法から得られる低混同可能性認識文法を含む。ここでは、従来の認識文法を使用してもよく、これを修正して、通常型の文法から混同可能性が高いワードクラスを識別及び抽出することにより、単純で安価な手法により従来の認識法を向上させることができる。さらに、混同可能性が高いワードクラスにサブワードユニット文法コンパイラを適用することにより、混同可能性が高いワードクラスの音韻的、音素的、及び/又は音節的な記述を生成する。これにより、混同可能性が高い各ワードクラスに対する高混同可能性サブワードユニット文法が生成される。そして、サブワードユニット文法を通常文法における残りの低混同可能性部分に併合して低混同可能性認識文法を生成する。

30

【0031】

さらに、本発明の具体例においては、受信された音声フレーズについて、音声フレーズの候補として生成されたサブフレーズ及び/又はサブフレーズの組み合わせを含む推測グラフを生成する。サブフレーズ間の追加的情報は、推測グラフ内の最も可能性が高い候補の検索処理を限定及び制約するために使用される。

【0032】

ここで、好ましくは、検索処理において、混同可能性が高いワードクラスから候補となるサブフレーズ又はサブワードを推測グラフに挿入する。混同可能性が高いワードクラス用のサブワードユニット文法は、追加的な意味論的及び/又は語用論的情報のそれぞれとともに制約として使用できる。

40

【0033】

最終的に最も信頼性が高い候補、すなわち認識すべき入力音声フレーズに対応する候補を判定するために、基礎的推測は制約により強制された制限に基づいて拡張される。ここでは、言語モデルにおける尤度(likelihood)、音響的証拠、及び追加的な制約について得点付け(scoring)を行ってもよい。推測が入力音声信号の全体に及ぶと、その推測を示す信号が出力される。得られた推測が他の推測に比べて著しく精度が悪い場合、その推測を抑圧してもよい。このような抑圧は、推測が音声フレーズ全体に拡張される以前に行っ

50

てもよい。さらに、当分野において知られる A * 検索法を用いることにより、推測を効率的に拡張することができる。

【0034】

本発明に基づく手法では、検索処理の負担を軽減するために、制約に基づいて、推測グラフから候補を削減し、最も可能性が高いフレーズに対応する分岐のないグラフを生成することができる。

【0035】

本発明に係る音声認識方法では、混同可能性が高い、すなわち複雑性が高い領域と、高い信頼度で認識できる混同可能性が低い、すなわち複雑性が低い領域とを一对とするとともに、混同可能性が低い領域を認識することにより追加的な意味論的情報又は語用論的情報から得られる情報を用いて、混同可能性が高い領域を認識する。すなわち、容易に認識できる低混同可能性領域と追加的な情報コンテンツは、認識の信頼度が高くない音声セグメントに関する記述として利用される。

10

【0036】

このような低混同可能性部分と高混同可能性部分を一对とする手法は、話者にとって自然なものである。ユーザ及び話者は、通常、このような一对の情報又は音声フレーズにおける部分間又はサブフレーズ間の高次の構造を直感的に使用して認識を行っている。

【0037】

本発明の好ましい具体例においては、音声フレーズにおけるシードサブフレーズを除く残りの部分のうちの少なくとも1つの部分に適用できる言語モデルにおける語彙を制約し、これにより検索処理の負担を軽減する。

20

【0038】

【発明の実施の形態】

以下、本発明に係る音声認識方法及び音声認識装置について図面を参照して詳細に説明する。

【0039】

図1は、本発明を適用した音声認識処理装置の構成を示す図である。

【0040】

この音声認識処理装置では、入力チャンネル10を介して音声が入力され、音声認識回路(speech recognizer)11に供給される。音声認識回路11は、適応すべき言語モデルに基づく、混同可能性が低い文法である低混同可能性認識文法(low-perplexity recognition grammar)12を使用して、供給された音声の分析を行う。

30

【0041】

音声認識回路11は、供給された音声の認識の結果をワード/サブワードユニットグラフ(word/subword unit graph)14として出力する。制約検索処理回路15は、高混同可能性ワードクラス(high-perplexity word classes)用サブワードユニット文法13を用いて、ワード/サブワードユニットグラフに対する制約された検索処理を実行する。制約検索処理回路15には、意味論及び語用論的制約情報17が供給されており、制約検索処理回路15からは、認識すべき音声フレーズを表す最終ワードグラフ16が出力される。

【0042】

図1に示す具体例においては、音声認識回路11により生成されるワード/サブワードユニットグラフ14は、ワード及びサブワードユニットから構成される推測グラフとして使用される。上述のように、追加的な制約検索処理回路15は、元の高混同可能性ワードクラスから、さらなる候補となるワード又はサブワードを推測グラフであるワード/サブワードユニットグラフ14に追加する。この処理は、上述のように、高混同可能性ワードクラス用サブワードユニット文法13と、意味論及び語用論的(文脈的)制約情報17とを利用して実現される。また、制約検索処理回路15は、推測グラフであるワード/サブワードユニットグラフ14からサブワードユニットを削除し、最終ワードグラフ16は、ワードのみから構成される。したがって、最終ワードグラフ16は、この音声認識装置に入力された音声フレーズに対応する認識結果として出力される。

40

50

【 0 0 4 3 】

図 1 に示す具体例の変形例として、2 種類又は 2 つのレベルの推測を行ってもよい。2 つの推測のうちの 1 つは、図 1 に示す音声認識回路により生成される推測グラフであるワード/サブワードユニットグラフ 1 4 である。検索処理は、最も可能性が高い認識された発話の断片から開始され、制約情報を用いて、可能性が低い認識された部分に拡張する。これにより、さらなる推測が生成され、このさらなる推測は、独立した (separated) データ構造により制御及び組織化される。この独立したデータ構造においては、ワード又は文の推測が生成され、さらに、評価の結果が良好でない場合は、必要に応じてこれら推測が取り消される。

最後に、この独立した、すなわち第 2 のデータ構造には、出力される 1 又は複数の推測が含まれることとなる。この変形例では、サブワードユニットは、第 1 のデータ構造における第 1 推測グラフから取り消されない。第 1 のデータ構造における与えられた文推測内のサブワード推測は、第 1 のデータ構造内では I M I を有さないが、他の文推測にとって重要でありなんらかの価値を有する場合がある。

10

【 0 0 4 4 】

図 1 に示す具体例において用いる文法又は言語モデルとしては、図 2 に示す処理手順における元の認識文法 (original recognition grammar) 2 0 から導き出される低混同可能性認識文法 2 1 が使用される。

【 0 0 4 5 】

元の認識文法 2 0 は、1 ~ n のクラスを有する高混同可能性ワードクラス 2 2 に分類される。また、元の認識文法 2 0 において分類されなかった残りの部分は、文法における低混同可能性部分 2 6 として扱われる。

20

【 0 0 4 6 】

さらに、1 ~ n のクラスを有する高混同可能性ワードクラス 2 2 は、サブワードユニット文法コンパイラ (sub-word unit compiler) 2 3 に供給され、これにより 1 ~ n のクラスを有する高混同可能性ワードクラス用サブワードユニット文法 2 4 が生成される。

【 0 0 4 7 】

続いて、併合回路 2 5 において、元の認識文法における低混同可能性部分 2 6 と上述の 1 ~ n クラスを有する高混同可能性ワードクラス用サブワードユニット文法 2 4 とが併合 (merge) され、低混同可能性認識文法 2 1 が生成される。この低混同可能性認識文法 2 1 は、図 1 に示す具体例における制約検索処理回路 1 5 における処理に適用される。

30

【 0 0 4 8 】

通常、低混同可能性認識文法 2 1 は、認識処理に先行して生成する。例えば都市名、個人名等の 1 又は複数の高混同可能性ワードクラス 2 2 は、元の認識文法 2 0 内で識別され、抽出される。高混同可能性ワードクラス 2 2 のクラス 1 ~ n のそれぞれについて、サブワードユニット文法コンパイラ 2 3 は、音素又は音節の組み合わせとして構成されるサブワードによりこれら高混同可能性ワードクラス 2 2 を適切に記述する。このコンパイルされた文法は、残りの部分である低混同可能性部分 2 6 と併合され、最終的な低混同可能性認識文法 2 1 が生成される。この低混同可能性認識文法 2 1 は、本発明に基づく音声認識処理に使用される。

40

【 0 0 4 9 】

本発明に基づく手法では、元の認識文法 2 0 における混同可能性が高い認識結果又は混同可能性が高い部分を混同可能性が低い文法により置換することが重要である。さらに、混同可能性が低い文法は、元の混同可能性が高い文法における全てのワード又はサブワードを変換する必要がある。この処理は、音声断片 (speech fragment) 又は音声ユニットの長さをワードの長さから音節の長さに変更することにより実現できる。このため、各音声断片又は音声ユニットに関して「混同可能性」を判定でき、「混同可能性が高いワード」や「混同可能性が低い音節」等の表現を使用することができる。

【 0 0 5 0 】

図 3 ~ 図 5 は、音声認識装置に供給された異なる音声フレーズ内の音声断片における高混

50

同可能性部分と低混同可能性部分のそれぞれ異なる関係を示す図である。図 3 に示す具体例では、与えられた音声フレーズ P H 内において、高混同可能性部分 (high-perplexity part) H P に続いて低混同可能性部分 (low-perplexity part) L P が出現している。図 4 に示す具体例では、低混同可能性部分 L P が高混同可能性部分 H P より前に出現している。

【 0 0 5 1 】

図 3 に示す統語図 (syntax diagram) は、話者が自分の姓の綴りを紹介している状況を表している。

【 0 0 5 2 】

姓を表す音声要素又は断片は、フレーズ P H における高混同可能性部分 H P を定義し、これに説明的な低混同可能性部分 L P が続いている。低混同可能性部分 L P は、綴りを説明することを宣言する、認識の信頼度が高い導入部 L P 1 と、低混同可能性部分 L P 2 1 ~ L P 2 n からなる実際の綴りを発話する部分とに再分割される。

10

【 0 0 5 3 】

この図 3 に示す具体例では、高混同可能性部分 H P の説明が低混同可能性部分 L P の一部、すなわち低混同可能性部分 L P 2 1 ~ L P 2 n から構成される綴りのシーケンスに含まれている。これは、低混同可能性部分 L P 自身が高混同可能性部分 H P に関する語用論的情報を含み、高混同可能性部分 H P を説明している例である。

【 0 0 5 4 】

図 4 は、低混同可能性部分 L P が高混同可能性部分 H P に関する語用論的情報を含む他の例の統語図を示す。

20

【 0 0 5 5 】

図 4 に示すフレーズ P H では、低混同可能性部分 L P は、高混同可能性部分 H P に先行している。この統語図は、都市名を郵便番号で記述している具体例を示すものである。例えばドイツでは、郵便番号を 5 桁の数字で表している。

【 0 0 5 6 】

言語モデル又は低混同可能性認識文法は、ドイツにおける都市は都市名によって記述できるとの意味論的情報を含んでいる。低混同可能性部分 L P である 5 桁の郵便番号に対し、都市名は高混同可能性部分 H P を構成する。さらに、低混同可能性部分 L P は、5 桁の郵便番号である語用論的情報をそれ自体に含んでいる。整数は、高度な信頼性をもって認識されるため、各数字 L P 1 ~ L P 5 は、低混同可能性部分の要素を構成している。

30

【 0 0 5 7 】

このように、図 3 及び図 4 に示す具体例においては、フレーズ P H における低混同可能性部分 L P と高混同可能性部分 H P 間の意味論的情報及び語用論的情報が推測ワードグラフに挿入できる特定の候補を示し、これにより認識すべき受信音声フレーズを表す可能性が最も高い候補を検索するための検索処理の負担を軽減できる。

【 0 0 5 8 】

図 5 に示す具体例では、認識すべきフレーズ P H は、高混同可能性部分 H P と、これに続く低混同可能性部分 L P により構成されている。

40

【 0 0 5 9 】

この具体例では、2 つの地理的エントリ、すなわち 2 つの都市を相互の位置関係が近接していることを宣言することにより記述している。

【 0 0 6 0 】

低混同可能性部分 L P は、第 1 の低混同可能性部分 L P 1 と、それに続く大都市の名称を記述する第 2 の低混同可能性部分 L P 2 とに分割される。第 1 の低混同可能性部分 L P 1 は、フレーズ P H における高混同可能性部分 H P により記述される小都市と低混同可能性部分 L P 2 により記述される大都市との地理的な近接関係を宣言する部分である。

【 0 0 6 1 】

この図 5 に示す具体例では、言語モデル内の意味論的情報は、小都市を近隣の大都市に対する地理的關係により特徴付けることができるという知識を含んでいる。したがって、全

50

ての小都市に対する検索処理は、一義的に又は多義的に認識された大都市に近接又は隣接している小都市からなるサブ集合に対する検索処理に制限することができる。

【0062】

図6は、図5に示す具体例における、小都市の名称のための音節モデル(syllabic model)を用いた再分析を説明する図である。このように、音声認識装置に入力されたフレーズPHにおける高混同可能性部分HPにより記述される小都市の正しい名称を検出するための検索処理の負担をさらに低減させるために、意味論的情報及び語用論的情報に加えて、音節モデル情報を導入してもよい。

【0063】

図7は、図5及び図6に示す具体例のための推測ワードグラフを図式的に示す図である。

10

【0064】

入力されたフレーズPHに対する推測ワードグラフは、例えば、文法における低混同可能性領域に対応するサブワードユニット1, 3, 5, 7と、文法における高混同可能性領域に対応するサブワードユニット2, 4, 6とから構成される。通常、文法とは異なり、ワードグラフ自体は、高混同可能性領域と低混同可能性領域とに分割することはできない。

【0065】

この具体例におけるサブワードユニット1, 3は、都市間の近接関係を記述する宣言に対応し、サブワードユニット5, 7は都市の候補を表す。

【0066】

文法における高混同可能性領域内の適切な候補を検出するためには、文法における低混同可能性部分から選択される都市の候補毎に、一連のサブワードユニットを分析する必要がある。

20

【0067】

ここで、文法における低混同可能性部分により提供される追加的な意味論的及び/又は語用論的情報が発話又はフレーズにおける混同可能性が高いサブワードユニットを判定するために不十分である場合もある。しかしながら、追加的な意味論的及び/又は語用論的情報は、与えられたフレーズの複雑性及び混同可能性を低減させるという点で有効である。

【0068】

本発明は、混同可能性が高い音声要素又は文法断片と、より混同可能性が低い文法部分に対応する要素との関係を解明する。すなわち、後者の混同可能性が低い文法部分に対応する要素は、前者の混同可能性が高い音声要素又は文法断片の記述又は説明として機能する。混同可能性が高い断片又は要素は、例えば、通りの名前、姓、都市名等や、例えば名称等の多数のワード、音声的に組み合わせられたワード、及び連続するワード等、認識すべき入力音声フレーズを表す一連のワードに対する候補又は認識の可能性が多数あるワードの部類(class)が文法又は言語モデル内において使用されている場合に見出される。

30

【0069】

これに対応する混同可能性が低い断片又は要素は、より高い可能性及び信頼性をもってより容易に認識されるワードでもよく、ワードの集合でもよく、連続するワードでもよく、あるいはワードの部類であってもよい。

【0070】

40

本発明の好ましい具体例においては、言語モデル又は低混同可能性認識文法は、混同可能性が高い文法断片と、この文法断片に対応する混同可能性が低い文法断片とを組として格納する追加的なデータベースを備える。このようなデータベースは、文法構造の一部となり、この言語モデルは認識処理において使用され、このデータベースに基づく文法を具現化する。

【0071】

このような文法においては、例えば図3～図4に示すように、低混同可能性断片と高混同可能性断片との相対的な位置関係は明示される。これらの位置関係は変更してもよい。

【0072】

さらに、混同可能性が高い各断片又は要素について、言語モデルにおける数が限定された

50

より小さな単位を用いた文法的又は形式的記述を与えてもよい。これらより小さな単位としては、音素、発音上の要素、音節等を用いることができる。したがって、高混同可能性部分の記述は、音節文法又は音素文法による表現を用いても実現できる。

【 0 0 7 3 】

このような文法の部分は、周知の様々な形式に基づいて表現できる。入力フレーズ内の高混同可能性部分における音素的又は音韻的關係の表現としては、例えば有限状態 (finite-state) フォーマットやコンテキストフリー (context-free) フォーマット等がある。

【 0 0 7 4 】

図 5 に対する図 6 は、有限状態音節文法 (finite-state syllable grammar) に基づく音節モデルを含む文法を示すものである。

10

【 0 0 7 5 】

図 6 に示すような文法断片は、より豊富な文法に組み込むことができる。また、同一文法内に複数の高混同可能性部分及び低混同可能性部分を設けることもできる。このように、高混同可能性部分がサブユニットモデル (sub-unit model) により表現される文法を認識文法と呼ぶ。

【 0 0 7 6 】

もちろん、当分野で知られている認識装置及び認識方法に上述の入力音声に対する認識文法を適用して複数の発話推測 (utterance hypotheses) を生成するようにしてもよい。従来から共通の慣例に基づき、複数の推測を推測グラフの形式で表現することができる。認識装置により整合された各グラフ、可能性のあるワード、サブワード、サブフレーズは、グラフにおけるエントリを構成する。通常、各エントリは、発話に対応する時間軸に沿って配列される。さらに、与えられたワードは、グラフ内において 2 回以上発生してもよく、この場合、与えられたワードは時間軸上の異なる位置に配列される。各ワードに対して、特定の時間間隔を表すワードの尤度 (likelihood) 又は確率 (probability) を表す得点 (score) を割り当てることもできる。この得点は、最も可能性が高い、したがって最良のワード連鎖又はシーケンスを判定するために使用できる。

20

【 0 0 7 7 】

低混同可能性文法の一部に対応するワード、サブワード、サブフレーズは、通常、高混同可能性文法の一部に対応するワード、サブワード、サブフレーズより高い精度及び信頼度をもって認識される。

30

【 0 0 7 8 】

図 6 に示すグラフにおいて、都市名を表す音節モデルは、異なる複数の音節エントリを含み、それぞれ異なる音節シーケンスに対応する多数の異なるグラフの経路 (path) 又は分岐 (branch) が存在し、したがって、異なる都市名がいずれも可能であると判定されてしまう。一方、フレーズ P H の後半における低混同可能性部分では、経路又は分岐の数が非常に少ない。

【 0 0 7 9 】

本発明の好ましい具体例においては、音声認識処理における検索処理は、図 7 に示す推測グラフのようなワードグラフを生成した後に開始される。検索処理は、推測グラフ内に存在し、入力フレーズを認識するために採用された文法又は言語モデルにおける低混同可能性断片又は部分に整合及び対応するワード、ワードシーケンス等から開始される。

40

【 0 0 8 0 】

図 7 に示す具体例では 4 つの大都市の名称で例示されるこれらワードシーケンスは、基礎的推測 (base hypotheses) を構成する。各基礎的推測は、それに先行又は後続するワード、サブワード、サブフレーズに拡張される。弁別方向は、高混同可能性部分に整合するサブワードユニットが基礎的推測に対応する低混同可能性部分に整合するサブワードユニットに先行するか、後続するかにより決定される。すなわち、基礎的推測は、推測ワードグラフ内の高混同可能性部分のサブワードユニットに拡張される。

【 0 0 8 1 】

通常、混同可能性が高いサブワードユニットから構成される可能なシーケンスは多数存在

50

する。さらに、通常、混同可能性が高いサブワードユニットは、基礎的推測に基づく推測ワードグラフ内に亘って分散されているため、文法に基づく場合と異なり、ワードグラフを低混同可能性部分 L P と高混同可能性部分 H P とから厳密に再構築することは困難である。

【 0 0 8 2 】

しかしながら、上述のように、基礎的推測は、可能なサブワードユニットのシーケンスに関する追加的な情報を提供する。この情報を用いて、この基礎的推測に適合しないサブワードシーケンスを無視しすることにより、検索空間を限定又は制限することができる。これにより、認識すべき入力音声フレーズに対応する一連のワードに対する可能な候補として、数が限定された一致認識結果 (consistent recognition result) を生成することができる。

10

【 0 0 8 3 】

さらに、A * 検索 (A*-search) として知られる検索法を適用することにより、処理負担の大きな検索を行うことなく、複数の基礎的推測を同時に分析して最高の尤度又は確率を有する最適な推測を得ることができる。

【 0 0 8 4 】

【 発明の効果 】

以上のように、本発明に係る音声認識方法は、通常型の認識文法を複数の高混同可能性ワードクラスに分類し、上記複数の高混同可能性ワードクラスから音素又は音節の組み合わせとして構成されるサブワードにより記述された高混同可能性ワードクラス用サブワードユニット文法を生成し、上記高混同可能性ワードクラス用サブワードユニット文法と上記通常型の認識文法の分類されなかった低混同可能性部分とを併合し、低混同可能性認識文法を生成する文法生成ステップと、上記低混同可能性認識文法を用いて、入力された音声フレーズからワード及びサブワードユニットから構成される推測グラフを生成する音声認識ステップと、上記推測グラフに対する制約された検索処理を実行し、認識すべき音声フレーズを表す最終ワードグラフを出力する制約検索処理ステップとを有し、上記制約検索処理ステップでは、上記高混同可能性ワードクラス用サブワードユニット文法と、低混同可能性部分と高混同可能性部分との間の意味論的及び語用論的な制約情報とを利用して、上記推測グラフからサブワードユニットを削除し、ワードから構成される最終ワードグラフを生成する。

20

30

また、本発明に係る音声認識装置は、通常型の認識文法を複数の高混同可能性ワードクラスに分類し、上記複数の高混同可能性ワードクラスから音素又は音節の組み合わせとして構成されるサブワードにより記述された高混同可能性ワードクラス用サブワードユニット文法を生成し、上記高混同可能性ワードクラス用サブワードユニット文法と上記通常型の認識文法の分類されなかった低混同可能性部分とを併合し、低混同可能性認識文法を生成する文法生成部と、上記低混同可能性認識文法を用いて、入力された音声フレーズからワード及びサブワードユニットから構成される推測グラフを生成する音声認識回路と、上記推測グラフに対する制約された検索処理を実行し、認識すべき音声フレーズを表す最終ワードグラフを出力する制約検索処理回路とを有し、上記制約検索処理回路では、上記高混同可能性ワードクラス用サブワードユニット文法と、低混同可能性部分と高混同可能性部分との間の意味論的及び語用論的な制約情報とを利用して、上記推測グラフからサブワードユニットを削除し、ワードから構成される最終ワードグラフを生成する。これにより、正しい候補を検索するための検索処理の負担を軽減でき、短時間で信頼性の高い音声認識を行うことができる。

40

【 図面の簡単な説明 】

【 図 1 】 本発明を適用した認識処理装置の機能的ブロック図である。

【 図 2 】 低混同可能性認識文法の生成の手順を説明する図であ。

【 図 3 】 低混同可能性サブフレーズと高混同可能性サブフレーズの時間的關係の一例を示す図である。

【 図 4 】 低混同可能性サブフレーズと高混同可能性サブフレーズの時間的關係の一例を示

50

す図である。

【図5】低混同可能性サブフレーズと高混同可能性サブフレーズの時間的關係の一例を示す図である。

【図6】図5に示すフレーズを別の構造により表現した図である。

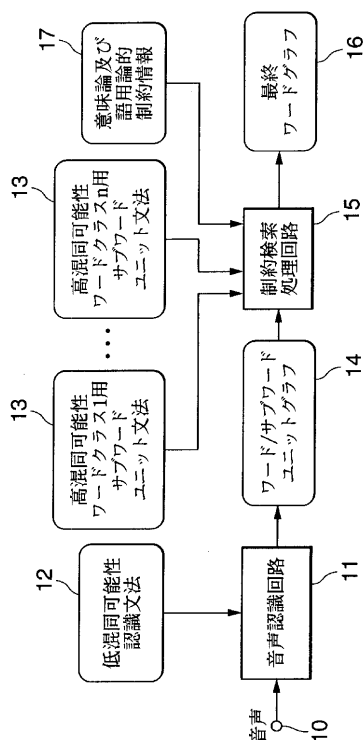
【図7】本発明により分析された低混同可能性領域と高混同可能性領域の推測グラフを示す図である。

【符号の説明】

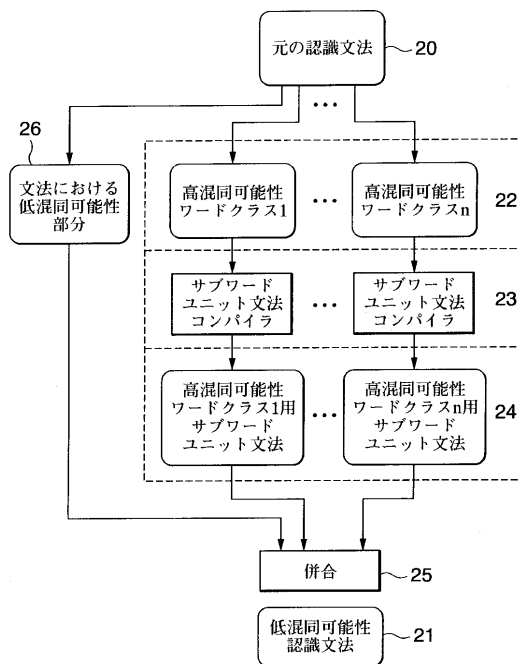
10 入力チャンネル、11 音声認識回路、12 低混同可能性文法、13 高混同可能性ワードクラス用サブワードユニット文法、14 ワード/サブワードユニットグラフ、15 制約検索処理回路、16 最終ワードグラフ、17 意味論及び語用論的制約情報

10

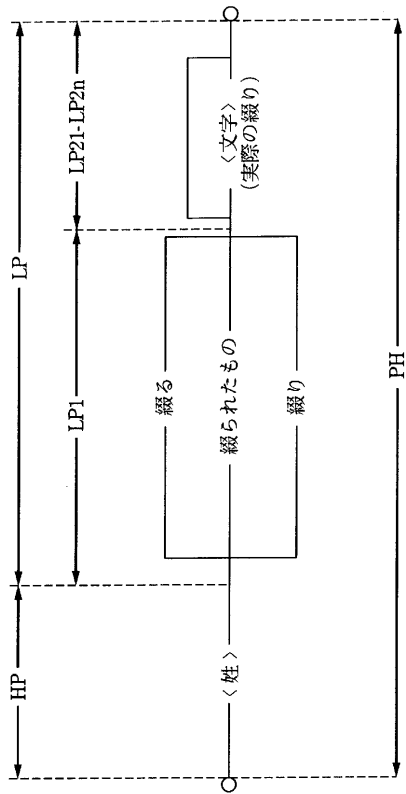
【図1】



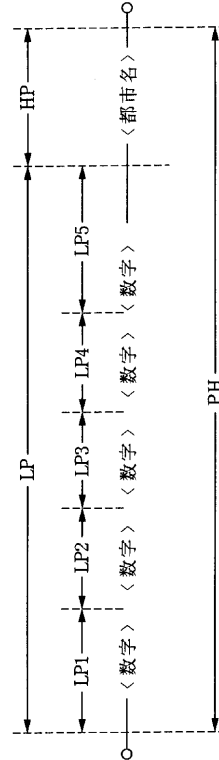
【図2】



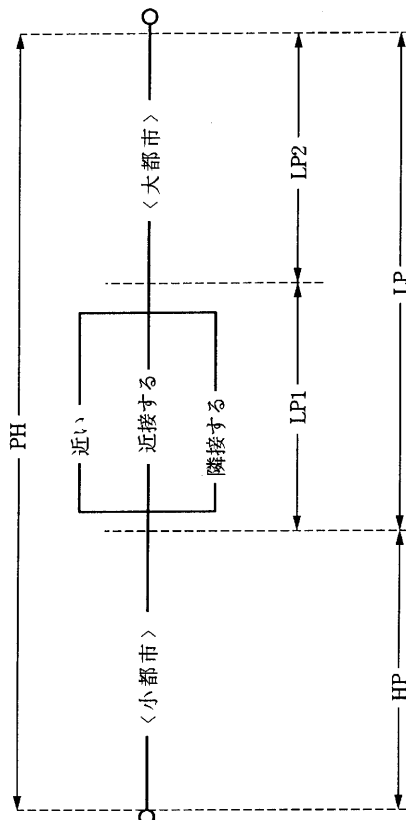
【図 3】



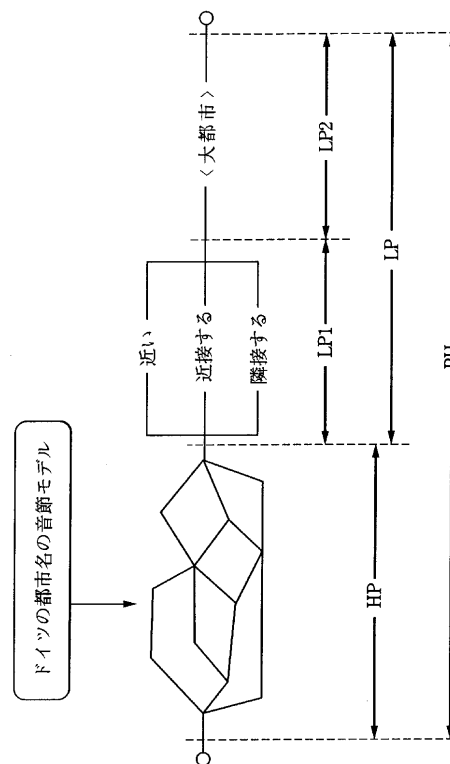
【図 4】



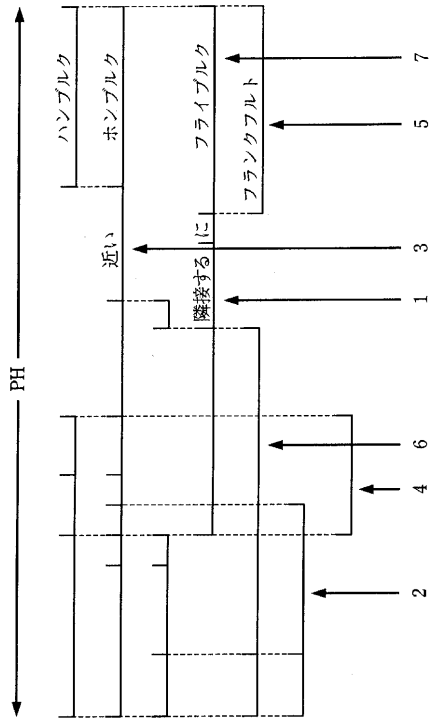
【図 5】



【図 6】



【図 7】



フロントページの続き

(72)発明者 ルッケ、ヘルムート

ドイツ連邦共和国、デー - 7 0 7 3 6 フェルバッハ、シュトゥットガルトー シュトラッセ
1 0 6、シュトゥットガルト テクノロジー センター ソニー インターナショナル (ヨー
ロッパ) ゲゼルシャフト ミット ベシュレンクテル ハフツング内

審査官 井上 健一

(56)参考文献 特開平 1 1 - 0 4 5 0 9 7 (J P , A)

李 晃伸 Akinobu LEE, 単語トレリスインデックスを用いた段階的探索による大語彙連続音声
認識 Large Vocabulary Continuous Speech Recognition Based on Multi-Pass Search Using
Word Trellis Index, 電子情報通信学会論文誌 THE TRANSACTIONS OF THE INSTITUTE OF ELECT
RONICS, INFORMATION AND COMMUNICATION ENGINEERS D-II, 日本, 社団法人電子情報通信学会 T
HE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS, 1 9 9 9 年 2 月
1 9 日, 第J82-D-II巻 第1号, p. 1 - 9

(58)調査した分野(Int.Cl., D B 名)

G10L 15/00-15/28