

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5120749号
(P5120749)

(45) 発行日 平成25年1月16日(2013.1.16)

(24) 登録日 平成24年11月2日(2012.11.2)

(51) Int. Cl.		F I	
G 1 O L	15/06 (2013.01)	G 1 O L	15/06 3 O O Z
G 1 O L	15/187 (2013.01)	G 1 O L	15/06 4 O O Z
G 1 O L	15/197 (2013.01)	G 1 O L	15/18 2 O O D
G 1 O L	15/183 (2013.01)	G 1 O L	15/18 2 O O F

請求項の数 3 (全 23 頁)

(21) 出願番号	特願2007-332427 (P2007-332427)	(73) 特許権者	393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2
(22) 出願日	平成19年12月25日(2007.12.25)	(74) 代理人	100099933 弁理士 清水 敏
(65) 公開番号	特開2009-156941 (P2009-156941A)	(72) 発明者	林 輝昭 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
(43) 公開日	平成21年7月16日(2009.7.16)	(72) 発明者	松田 繁樹 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
審査請求日	平成22年11月2日(2010.11.2)	(72) 発明者	清水 徹 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 木構造辞書を記録した記憶媒体、木構造辞書作成装置、及び木構造辞書作成プログラム

(57) 【特許請求の範囲】

【請求項1】

各々が子ノードを持つ複数の非終端ノードと、子ノードを持たずにそれぞれ単語に対応する複数の終端ノードとを各々含む、既知語の木構造辞書及び未知語の木構造辞書を記録したコンピュータ読取可能な記憶媒体であって、

前記既知語の木構造辞書及び未知語の木構造辞書の各々は、いずれも同じ辞書探索プログラムにより探索可能なように構成された同じ辞書構造を持ち、

前記既知語の木構造辞書及び未知語の木構造辞書の各々に含まれる前記複数の非終端ノードは、それぞれ所定の音素に対応しており、

前記既知語の木構造辞書及び未知語の木構造辞書の各々において、ルートノードから子ノードをたどっていくことにより、各非終端ノードを経由して、当該非終端ノードに対応する音素が自身の音素列の一部となっている単語に対応する終端ノードの全てに到達可能に構成されており、

前記既知語の木構造辞書内の前記複数の終端ノードの各々には、前記複数の終端ノードにそれぞれ対応する単語の生起確率を示す確率テーブルが関連付けられており、

前記未知語の木構造辞書内の前記複数の非終端ノード及び前記複数の終端ノードの各々には、前記未知語の木構造辞書内の前記複数の終端ノードに対応する単語以外の、前記既知語の木構造辞書内に存在する予め定める単語の生起確率を示す確率テーブルが関連付けられている、コンピュータ読取可能な記憶媒体。

【請求項2】

ある集合に属する単語である既知語の各々についての生起確率を示す所定の統計的言語モデルに基づいて、前記ある集合に含まれない単語である未知語の生起確率を予測するために参照される、未知語の木構造辞書を作成するための木構造辞書作成装置であって、

前記既知語は複数のクラスに分類されており、

前記未知語の木構造辞書と同じ辞書探索プログラムにより探索可能なように、前記未知語の木構造辞書と同じ辞書構造を持つ木構造辞書であって、既知語の生起確率を予測するために参照される既知語の木構造辞書を記憶するための既知語辞書記憶手段と、

前記未知語と、当該未知語の属するクラスとを対にした未知語情報の入力を受けるための未知語情報入力手段と、

前記複数のクラスのうちのあるクラスの指定を受けるためのクラス指定手段と、

前記未知語情報入力手段により入力された未知語情報から、前記クラス指定手段により指定されたクラスと対になった未知語を抽出し、前記未知語の木構造辞書を作成するための木構造辞書作成手段とを含み、

前記既知語の木構造辞書及び前記未知語の木構造辞書の各々は、各々が子ノードを持つ複数の非終端ノードと、子ノードを持たずにそれぞれ単語に対応する複数の終端ノードとから構成され、

前記複数の非終端ノードは、それぞれ所定の音素に対応しており、

前記既知語の木構造辞書と前記未知語の木構造辞書の各々において、ルートノードから子ノードをたどっていくことにより、各非終端ノードを経由して、当該非終端ノードに対応する音素が自身の音素列の一部となっている単語に対応する終端ノードの全てに到達可能に構成されており、

前記木構造辞書作成装置はさらに、

前記未知語の木構造辞書の前記複数の非終端ノード及び前記複数の終端ノードの各々に、前記所定の統計的言語モデルから得られる、前記クラス指定手段により指定されたクラスに分類されている単語であって前記既知語の木構造辞書に登録されている同一の単語の生起確率を示すファクタリングテーブルを関連付けるためのファクタリングテーブル付与手段を含む、木構造辞書作成装置。

【請求項3】

コンピュータにより実行されると、当該コンピュータを、ある集合に属する単語である既知語の各々についての生起確率を示す所定の統計的言語モデルに基づいて、前記ある集合に含まれない単語である未知語の生起確率を予測するために参照される、未知語の木構造辞書を作成するように機能させるための木構造辞書作成プログラムであって、

前記コンピュータは、前記未知語の木構造辞書と同じ辞書探索プログラムにより探索可能なように、前記未知語の木構造辞書と同じ辞書構造を持つ木構造辞書であって、前記既知語の生起確率を予測するために参照される既知語の木構造辞書を記憶するための既知語辞書記憶手段に接続されて用いられ、

前記既知語は複数のクラスに分類されており、前記プログラムは、前記コンピュータを、

前記未知語と、当該未知語の属するクラスとを対にした未知語情報の入力を受けるための未知語情報入力手段と、

前記複数のクラスのうちのあるクラスの指定を受けるためのクラス指定手段と、

前記未知語情報入力手段により入力された未知語情報から、前記クラス指定手段により指定されたクラスと対になった未知語を抽出し、前記未知語の木構造辞書を作成するための木構造辞書作成手段として機能させ、

前記既知語の木構造辞書及び前記未知語の木構造辞書の各々は、各々が子ノードを持つ複数の非終端ノードと、子ノードを持たずにそれぞれ単語に対応する複数の終端ノードとから構成され、

前記複数の非終端ノードは、それぞれ所定の音素に対応しており、

前記既知語の木構造辞書と前記未知語の木構造辞書の各々において、ルートノードから子ノードをたどっていくことにより、各非終端ノードを経由して、当該非終端ノードに対

10

20

30

40

50

応する音素が自身の音素列の一部となっている単語に対応する終端ノードの全てに到達可能に構成されており、

前記木構造辞書作成プログラムはさらに、前記コンピュータを、

前記未知語の木構造辞書の前記複数の非終端ノード及び前記複数の終端ノードの各々に、前記所定の統計的言語モデルから得られる、前記クラス指定手段により指定されたクラスに分類されている単語であって前記既知語の木構造辞書に登録されている同一の単語の生起確率を示すファクタリングテーブルを関連付けるためのファクタリングテーブル付与手段として機能させる、木構造辞書作成プログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、大語彙連続音声認識等で用いられる木構造辞書の作成方法に関し、特に、認識辞書に登録されていない未知語に関する木構造辞書を記録した記憶媒体、その作成装置、及びコンピュータプログラムに関する。

【背景技術】

【0002】

大語彙連続音声認識システムでは、音声認識の過程又は認識後の検証の過程で確率的言語モデルを用いた言語スコアを算出する。確率的言語モデルとは、ある言語において、単語列又は文字列等に対して、それらが起こる確率で自然言語をモデル化したものである。言語スコアは、音声認識の結果得られた単語列等の尤度を、言語モデルに従って算出した

20

【0003】

N個の単語又は文字からなる単語列又は文字列が生成される確率によって自然言語をモデル化したものをNグラム言語モデルと呼ぶ。特にN = 1, 2, 3の場合をそれぞれユニグラム、バイグラム、トライグラムと呼ぶ。パラメータ推定のための計算量及び精度の点から、バイグラム又はトライグラムが用いられる場合が多い。以下の説明は単語列についてのものとする。

【0004】

認識候補の言語スコアを算出するため、予め算出した言語スコアを付した先読み単語辞書を用意することが多い。また、認識候補の探索の効率を向上させるために、単語辞書を木構造ネットワークで表現することが一般的である。

30

【0005】

図1を参照して、木構造辞書の概要を説明する。今、音響モデルをモノフォンとし、語彙として「赤い」、「明るい」、「青い」、「売買」、「媒質」、及び「梅雨」を考えるものとする。これらの語の音素列20, 22, 24, 26, 28, 34を図1の上段に示す。図から明らかなように、音素列20, 22, 及び24の語頭の「a」の音30は、「赤い」、「明るい」、及び「青い」の3つの単語で共有されている。さらにその後の2つの音を含む「ak a」の音素列32は、「赤い」及び「明るい」2つの単語により共有されている。同様にして、音素列26, 28, 及び34の語頭の「b」の音38は、「売買」、「媒質」、及び「梅雨」の3つの単語で共有されている。さらにその後の2つの音を含む「ba i」の音素列36も、「売買」、「媒質」、及び「梅雨」の3つの単語により共有されている。

40

【0006】

図1の下段に、上記した単語群に対応する木構造辞書40の概念的構造を示す。木構造辞書40は、図に示すように、ルートノード80と、ルートノード80から分岐し、かつ各単語の先頭部分の音素に対応するノード50(a)、ノード82(b)、及びノード110(t)等と、さらに、これらノードから順次分岐する複数の非終端ノード50~58、62、82~90、94~98と、最終的な単語に対応する終端ノード60、64、66、92、100、102等とを含む。この例では、終端ノード60, 64, 66, 92, 100, 及び102は、それぞれ単語「明るい」、「赤い」、「青い」、「売買」、「

50

媒質」、及び「梅雨」に対応している。

【 0 0 0 7 】

例えば、前記した「赤い」、「明るい」、及び「青い」の3つの単語は、音素「a」を共有しているので、木構造辞書40のノード50をこの3つの単語で共有する。しかし、次の音素では2つのノード52(k)及び62(o)に分岐する。以下同様に木構造化されており、その結果ノード列50, 52, 54, 56, 58, 60をたどることによって単語「明るい」が、ノード列50, 52, 54, 64をたどることによって単語「赤い」が、ノード列50, 62, 66をたどることによって単語「青い」が、それぞれ探索される。

【 0 0 0 8 】

前記した「売買」、「媒質」、及び「梅雨」の3つの単語は、音素「b」を共有しているので、木構造辞書40のノード82をこの3つの単語で共有する。以下同様に木構造化されており、ノード列82, 84, 86, 88, 90, 92をたどることによって単語「売買」が、ノード列82, 84, 86, 94, 96, 98, 100をたどることによって単語「媒質」が、ノード列82, 84, 86, 102をたどることによって単語「梅雨」が、それぞれ探索される。

【 0 0 0 9 】

以上が木構造辞書の基本的構造である。

【 0 0 1 0 】

木構造辞書の終端ノード(図1の場合のノード60, 64, 66, 92, 100、及び102)には、通常、各単語のNグラム確率テーブル(図1の例の場合、確率テーブル70, 72, 74, 104, 106、及び108)、又はそうしたテーブルへのポインタが付されている。このテーブルは、木構造辞書に登録されている任意の(N-1)個の単語の組合せの各々に対し、当該単語の組合せが先行しているときに、当該終端ノードの単語が続いて生起するNグラム確率の集合である。したがって、この確率テーブルは、言語モデル中の全単語の任意の(N-1)個の組合せの数だけのエントリを持つ。つまり、確率テーブルの容量は、語彙数に依存する。

【 0 0 1 1 】

音声認識では、音声認識に同期して上記した木構造辞書の先頭ノードから終端ノードに向けて、仮説の展開を進める。終端ノードに到達した単語仮説は単語グラフと呼ばれる、仮説の集合をグラフ表現したものに登録され、再び辞書の先頭ノードから後続の単語仮説の探索を開始する。

【 0 0 1 2 】

Nグラム言語モデルを用いた大語彙連続音声認識においては、単語辞書に登録されていない単語(以下、未知語と呼ぶ)は認識できないという重要な課題がある。音声認識の対象とする文章から必要な未知語の集合を何らかの方法で事前に収集することができたときに、その単語を高精度に認識するためには、どのようにしたらよいかと考えられてきた。

【 0 0 1 3 】

従来の手法では、未知語に関するNグラム確率を計算するために、既存のNグラム言語モデルを変更していた。

【 0 0 1 4 】

例えば、未知語に対して、Nグラム言語モデルに必要なNグラム確率及びバックオフ確率を推定し、Nグラム言語モデルに追加登録する方法が考えられている(非特許文献1)。また、タスク依存の正規文法を作成して、未知語を含む文をランダムに生成し、Nグラム言語モデルを再学習する方法が考えられている(非特許文献2)。

【非特許文献1】松井淳、小林彰夫、今井亨、及び安藤彰男、「ニュース音声認識のための未知語登録法の検討」、音講論、1999年9月、2 1 7

【非特許文献2】L. ガルシュ(L. Galescu)、E. リンガー(E. Ringger)、及びJ. アレン(J. Allen)、「新しいタスク領域のための高感度言語モデルの開発」("Rapid Language Model Development

10

20

30

40

50

for New Task Domains”）、言語の資源及び評価の国際会議予稿集(In Proceedings of International Conference on Language Resources and Evaluation)、1998年、p. 807 812

【発明の開示】

【発明が解決しようとする課題】

【0015】

未知語を追加して既存のNグラム言語モデルを変更する場合、未知語に対していくらかのNグラム確率を割当てることになる。その分、既存の単語に割当てられる確率を小さくする必要がある。また、未知語を含む文を学習データとしてランダムに作成した場合、各単語のNグラム確率は、学習データ中の単語の相対頻度から求められる。したがって、未知語について作成した分の学習データが小さく、既存単語の学習データのサイズがはるかに大きい場合、既知単語のためのNグラム確率と比較すると、未知語のためのNグラム確率ははるかに小さくなってしまふ。この確率を大きくしようとしても、学習データそのものがランダムに作成されたものであるため、学習結果の信頼性は低くなる。未知語のためのNグラム確率が小さい場合には、未知語を認識することは困難であり、未知語のためのNグラム確率を大きくした場合には、認識結果の信頼性が低下するという問題がある。

10

【0016】

したがって、本発明の目的は、言語モデルを変更せずに、大語彙連続音声認識において未知語をある信頼度をもって認識できるようにするための木構造辞書を記憶した記憶媒体、その作成装置、及びその作成プログラムを提供することである。

20

【0017】

本発明の他の目的は、メモリの使用量を少なくし、かつ言語モデルを変更せずに、大語彙連続音声認識において未知語をある信頼度をもって認識できるようにするための木構造辞書の記憶媒体、その作成装置、及びその作成プログラムを提供することである。

【課題を解決するための手段】

【0018】

本発明の第1の局面に係る記憶媒体は、各々が子ノードを持つ複数の非終端ノードと、子ノードを持たずにそれぞれ単語に対応する複数の終端ノードとを含む木構造辞書を記録した記憶媒体である。木構造辞書に含まれる複数の非終端ノードは、それぞれ所定の音素に対応しており、木構造辞書において、ルートノードから子ノードをたどっていくことにより、各非終端ノードを経由して、当該非終端ノードに対応する音素が自身の音素列の一部となっている単語に対応する終端ノードの全てに到達可能に構成されている。木構造辞書内の複数の非終端ノード及び複数の終端ノードの各々には、複数の終端ノードに対応する単語以外の、予め定める単語の生起確率を示す確率テーブルが関連付けられている。

30

【0019】

当該木構造辞書の終端ノードの各々には、予め定める単語の生起確率を示す確率テーブルが関連付けられている。したがって、当該木構造辞書の各終端ノードに対応する単語が未知語であり、未知語を含む音声を音声認識するときに、当該記憶媒体を使用すれば、既存の言語モデルを変更せずに、予め定める単語の生起確率と同じ確率で未知語を認識することができる。その結果、言語モデルを変更せずに、大語彙連続音声認識において未知語をある信頼度をもって認識できるようにするための木構造辞書を記憶した記憶媒体を提供することができる。

40

【0020】

好ましくは、確率テーブルは、予め定める単語が、複数の単語の各々の後に生起する確率を記録したバイグラム確率テーブルである。

【0021】

確率テーブルは、バイグラム確率テーブルなので、当該複数の単語が既知単語である場合、バイグラム確率テーブルには、既知単語の数の分のバイグラム確率を記憶させれば良い。確率テーブルがNグラム確率テーブルである場合、予め定める単語が、複数の単語の

50

任意の(N-1)個の組合せの各々の後に生起するNグラム確率を記憶しなければならない。Nが3以上の整数であるときのNグラム確率テーブルと比較すると、バイグラム確率テーブルでは、メモリの使用量を少なくすることができる。その結果、メモリの使用量を少なくし、言語モデルを変更せずに、大語彙連続音声認識において未知語をある信頼度をもって認識できるようにするための木構造辞書を記憶した記憶媒体を提供することができる。

【0022】

本発明の第2の局面に係る木構造辞書作成装置は、ある集合に属する単語の各々についての生起確率を示す所定の統計的言語モデルに基づいて、ある集合に含まれない単語である未知語の生起確率を予測するために参照される木構造辞書を作成するための木構造辞書作成装置である。ある集合に属する単語は複数のクラスに分類されている。この木構造辞書作成装置は、未知語と、当該未知語の属するクラスとを対にした未知語情報の入力を受け取るための未知語情報入力手段と、複数のクラスのうちのいずれかの指定を受け取るためのクラス指定手段と、未知語情報入力手段から、クラス指定手段により指定されたクラスと対になった未知語を抽出し、木構造辞書を作成するための木構造辞書作成手段とを含む。木構造辞書は、各々が子ノードを持つ複数の非終端ノードと、子ノードを持たずにそれぞれ単語に対応する複数の終端ノードとから構成され、複数の非終端ノードは、それぞれ所定の音素に対応しており、木構造辞書において、ルートノードから子ノードをたどっていくことにより、各非終端ノードを経由して、当該非終端ノードに対応する音素が自身の音素列の一部となっている単語に対応する終端ノードの全てに到達可能に構成されている。木構造辞書作成装置はさらに、木構造辞書の複数の非終端ノード及び複数の終端ノードの各々に、所定の統計的言語モデルから得られる、クラス指定手段により指定されたクラスに分類されている同一の単語の生起確率を示すファクタリングテーブルを関連付けるためのファクタリングテーブル付与手段を含む。

【0023】

木構造辞書作成手段は、クラス指定手段によって指定されたクラスの未知語の木構造辞書を作成する。したがって、未知語を含む音声を、ある音声認識装置が音声認識するとき、その音声にどのクラスの未知語が含まれているかをユーザが予め知っていれば、必要なクラスの木構造辞書をユーザが選択して、選択されたクラスの木構造辞書のみを音声認識装置に探索させれば良い。全てのクラスの未知語を探索する必要がないので、処理時間を軽減することができる。また、ファクタリングテーブル付与手段は、各クラスの未知語の木構造辞書に、ファクタリングテーブルを関連付ける。したがって、あるクラスの未知語を含む音声を音声認識するとき、既存の言語モデルを変更せずに、当該クラスに分類されている同一の単語の生起確率と同じ確率で、未知語を認識することができる。その結果、言語モデルを変更せずに、大語彙連続音声認識において未知語をある信頼度をもって認識できるようにするための木構造辞書を記憶した木構造辞書作成装置を提供することができる。

【0024】

本発明の第3の局面に係る木構造辞書作成プログラムは、ある集合に属する単語の各々についての生起確率を示す所定の統計的言語モデルに基づいて、ある集合に含まれない単語である未知語の生起確率を予測するために参照される木構造辞書を作成するための木構造辞書作成プログラムである。ある集合に属する単語は複数のクラスに分類されている。この木構造辞書作成プログラムは、未知語と、当該未知語の属するクラスとを対にした未知語情報の入力を受け取るための未知語情報入力手段と、複数のクラスのうちのいずれかの指定を受け取るためのクラス指定手段と、未知語情報入力手段から、クラス指定手段により指定されたクラスと対になった未知語を抽出し、木構造辞書を作成するための木構造辞書作成手段とを含む。木構造辞書は、各々が子ノードを持つ複数の非終端ノードと、子ノードを持たずにそれぞれ単語に対応する複数の終端ノードとから構成され、複数の非終端ノードは、それぞれ所定の音素に対応しており、木構造辞書において、ルートノードから子ノードをたどっていくことにより、各非終端ノードを経由して、当該非終端ノード対

10

20

30

40

50

応する音素が自身の音素列の一部となっている単語に対応する終端ノードの全てに到達可能に構成されている。この木構造辞書作成プログラムはさらに、木構造辞書の複数の非終端ノード及び複数の終端ノードの各々に、所定の統計的言語モデルから得られる、クラス指定手段により指定されたクラスに分類されている同一の単語の生起確率を示すファクタリングテーブルを関連付けるためのファクタリングテーブル付与手段を含む。

【発明を実施するための最良の形態】

【0025】

以下、本発明の実施の形態について、添付図面に基づき詳細に説明する。なお、以下の実施の形態の説明では、同一の部品には同一の参照番号を付してある。それらの機能及び名称も同一である。したがって、それらについての詳細な説明は繰返さない。

10

【0026】

[第1の実施の形態]

図1において説明した木構造辞書に登録されている既知単語は、品詞属性によって予め複数のクラスに分類されている。例えば、図1において、「明るい」は明度を示す形容詞なので、「明るい」は「明度を示す形容詞」というクラスに属する。「赤い」及び「青い」は、色相を示す形容詞なので、「色相を示す形容詞」というクラスに属する。「売買」、「媒質」、及び「梅雨」は名詞なので、「名詞」というクラスに属する。本実施の形態に係る音声認識装置140を利用するユーザは、クラスを自由に決めることが可能であり、どの単語をどのクラスに属するかも自由に決めることが可能である。

20

【0027】

<機能的構成>

図2は、本実施の形態に係る音声認識装置140の機能ブロック図である。図2を参照して、音声認識装置140は、予め用意されている既知単語のための木構造辞書を記憶するための単語辞書150と、既知単語のユニグラム確率及びバイグラム確率を記憶するための言語モデル152と、所定の言語における音声から得られた音響的特徴量の系列が、ある音素から得られた尤度を計算するための統計的確率モデルである音響モデル154と、クラス毎に未知語のための木構造辞書を記憶するための未知語を含む木構造辞書156と、クラスと未知語を含む木構造辞書156に含まれる木構造辞書との間のマッピングテーブルを記憶するためのクラス 木構造辞書マッピングテーブル158とを含む。

【0028】

音声認識装置140はさらに、未知語とその属するクラスとの組からなるテキストである未知語に関する情報162と、クラスと、当該クラスに属する未知語に対して割り当てられるNグラム確率を与える既知単語(この単語を、これ以降、そのクラスの「代表単語」と呼ぶ。)との組からなるテキストである代表単語に関する情報176とを受けて、単語辞書150及び言語モデル152を用いて、未知語を含む木構造辞書156と、クラス 木構造辞書マッピングテーブル158とを作成したり、未知語を含む木構造辞書156に登録されていない未知語とそのクラスとの組である追加単語に関する情報168を入力として、未知語を含む木構造辞書156に新たな未知語を追加したりすることにより、木構造辞書156を管理するための未知語辞書管理部160を含む。

30

【0029】

音声認識装置140はさらに、複数個ある未知語の木構造辞書のうち、音声認識に使用する木構造辞書のクラス名を列挙した使用クラスの情報172を入力として、単語辞書150、言語モデル152、未知語を含む木構造辞書156、及びクラス 木構造辞書マッピングテーブル158を用いて音声データ174を音声認識し、音声認識されたテキスト178を出力するための音声認識部170を含む。

40

【0030】

(未知語辞書管理部160)

未知語辞書管理部160は、未知語に関する情報162、及び未知語の属する各クラスの代表単語に関する情報176を入力として、単語辞書150及び言語モデル152を用いて、未知語を含む木構造辞書156及びクラス 木構造辞書マッピングテーブル158

50

を作成するための未知語辞書作成部 164 と、追加単語に関する情報 168 を入力として、クラス 木構造辞書マッピングテーブル 158 を用いて未知語を含む木構造辞書 156 に新規の未知語を登録して、更新するためのメンテナンス部 166 とを含む。

【0031】

(先読み単語辞書 150 及び言語モデル 152)

図3は、図2に示す先読み単語辞書 150 及び言語モデル 152 の詳細について説明するための図である。図3を参照して、言語モデル 152 は、既知単語のユニグラム確率を記憶するためのユニグラムテーブル 224 と、既知単語のバイグラム確率を記憶し、先読み単語辞書 150 の全ての既知単語に対して関連付けられているバイグラムテーブル 226 とを含む。

10

【0032】

単語辞書 150 は、ルートノード 200 と、ルートノード 200 から順次分岐する複数の非終端ノード 202 ~ 208 等と、ルートノード 200 から非終端ノード列をたどることにより到達可能な、既知単語に対応する終端ノード 210 ~ 216 等と、既知単語のバイグラムの全ての組合せについて算出されたバイグラム確率の中の最大値を記憶するための、既知単語の各々について、当該既知単語が先行単語であるときに、ルートノード 200 との距離が k (k は 1、2、...、 $m-1$ の任意の数) である非終端ノードの各々につき、その非終端ノードをルートとするサブツリーに属する終端ノードの単語が続いて生起するバイグラム確率の最大値を記憶するための、その非終端ノードに関連付けられているファクタリングテーブル 222 等と、ルートノード 200 との距離が $(k+1)$ 以上である非終端ノードの各々について、当該非終端ノードをルートとするサブツリーに属する終端ノードの既知単語が生起するユニグラム確率の最大値を記憶する領域 228 等とを含む。なお、ここでルートノード 200 は、対応する音素がないダミーノードである。

20

【0033】

例えば、図3に示す例では、単語辞書 150 は、ノード 200 からノード列 202、208、及び 210 をたどることによって単語「亜鉛」が、ノード 202 からノード 212 までのノード列をたどることによって単語「赤い」が、ノード 204 からノード 214 までのノード列をたどることによって単語「金閣寺」が、ノード 206 からノード 216 までのノード列をたどることによって単語「二条城」が、それぞれ探索される構造となっている。

30

【0034】

(言語モデル)

以下、言語モデルの構造について説明する。以下の説明では、全ての既知単語を w_1 、 w_2 、...、 w_m で表わす。

【0035】

図4は、図3に示すユニグラムテーブル 224 の詳細な図である。図4を参照して、ユニグラムテーブル 224 には、既知単語 w_1 、 w_2 、...、 w_m の各々に対して、その単語が生起するユニグラム確率 $P(w_1)$ 、 $P(w_2)$ 、...、 $P(w_m)$ が記憶されている。

【0036】

図5は、図3に示すバイグラムテーブル 226 の詳細な図である。図5を参照して、バイグラムテーブル 226 は、全ての既知単語の各々に対して、当該既知単語が先行単語であるときに単語 w_1 が続いて生起するバイグラム確率を記憶するテーブル 250 と、同様に、 w_2 が生起するバイグラム確率を記憶するテーブル 252 と、 w_m が生起するバイグラム確率を記憶するテーブル 254 等とを含む。

40

【0037】

テーブル 250 の 1 行目には、先行単語が w_1 であるときに w_1 が続いて生起するバイグラム確率 $P(w_1 | w_1)$ が記憶される。同様に、2 行目には、先行単語が w_2 であるときに w_1 が続いて生起するバイグラム確率 $P(w_1 | w_2)$ が、 m 行目には、先行単語が w_m であるときに w_1 が続いて生起するバイグラム確率 $P(w_1 | w_m)$ が記憶される。その他の行、及びその他のテーブルについても同様である。

50

【0038】

図6は、図3に示す終端ノードに関連して記憶されるものについて説明するための図である。図6を参照して、単語辞書150の終端ノードの各々には、単語Index、クラス、単語、及び発音が記憶される。単語Indexは、当該単語に対応する確率を言語モデルから得るための、言語モデル中のエントリを一意に識別するためのものである。単語辞書150においては、単語Indexにはその単語そのものが記憶される。図3に示すノード210の単語「亜鉛」の単語Index、クラス、及び発音は、それぞれ「亜鉛」、「化学物質」、及び「アエン」である。同様に、ノード212の単語「赤い」の単語Index、クラス、及び発音は、それぞれ、「赤い」、「形容詞」、及び「アカイ」であり、ノード214の単語「金閣寺」の単語Index、クラス、及び発音は、それぞれ、「金閣寺」、「地名」、及び「キンカクジ」であり、ノード216の単語「二条城」の単語Index、クラス、及び発音は、それぞれ、「二条城」、「地名」、及び「ニジョウジョウ」である。

10

【0039】

図7は、図3に示すファクタリングテーブル222の詳細について説明するための図である。図3に示すノード204をルートとするサブツリーに属する終端ノードの全ての既知単語を、 w_{i1} 、 w_{i2} 、 \dots 、 w_{ip} とする。図7を参照して、テーブル222の1列目には、 w_{i1} 、 w_{i2} 、 \dots 、 w_{ip} に対して、単語 w_1 が先行単語であるときのバイグラム確率 $P(w_{i1} | w_1)$ 、 $P(w_{i2} | w_1)$ 、 \dots 、及び $P(w_{ip} | w_1)$ の最大値が記憶されている。同様に、任意の $p \in \{2, \dots, m\}$ において、テーブル222の p 列目には、全ての w_{i1} 、 w_{i2} 、 \dots 、 w_{ip} に対して、単語 w_p が先行単語であるときのバイグラム確率 $P(w_{i1} | w_p)$ 、 $P(w_{i2} | w_p)$ 、 \dots 、及び $P(w_{ip} | w_p)$ の最大値が記憶されている。

20

【0040】

(未知語を含む木構造辞書156)

未知語として登録するクラスの数 p とする。図8は、未知語を含む木構造辞書156の詳細な図である。図8を参照して、未知語を含む木構造辞書156は、クラス1~クラス p の木構造辞書270、272、及び274等を含む。

【0041】

以後、クラス1を「地名」を表わすクラスとする。図2に示す代表単語に関する情報176には、クラス1「地名」を代表する単語として、既知単語「金閣寺」が割当てられているとする。図9は、図8に示すクラス1の木構造辞書の詳細な図である。図9を参照して、木構造辞書270は、ルートノード280と、ルートノード280から順次分岐する複数の非終端ノード282~288等と、クラス1の未知語にそれぞれ対応する終端ノード290~296等と、全ての既知単語の各々について、当該既知単語が先行単語であるときに、クラス1を代表する既知単語が続いて生起するバイグラム確率を記憶するためのテーブルであり、全ノードに関連付けられているクラス1のファクタリングテーブル300とを含む。

30

【0042】

ノード280からノード列282、288、及び290をたどることによって単語「建仁寺」が、ノード282からノード292までのノード列をたどることによって単語「高德庵」が、ノード284からノード294までのノード列をたどることによって単語「三千院」が、ノード286からノード296までのノード列をたどることによって単語「知恩院」が、それぞれ探索される構造となっている。

40

【0043】

図10は、図9に示す終端ノードに記憶されるものについて説明するための図である。図10を参照して、クラス1の木構造辞書270の終端ノードの各々には、単語Index、クラス、単語、及び発音が記憶され、単語Indexは、クラス1を代表する既知単語の単語Indexが記憶される。したがって、各終端ノードの単語Indexには既知単語である「金閣寺」が記憶される。したがって、この木構造辞書270に属する全ての

50

単語についての生起確率は、金閣寺と同じ値になる。クラス1が「地名」なので、全ての未知語に対して「地名」が入る。単語の列には、各単語の左隣に@を付したものが記憶される。@は、その単語が未知語であることを示す。

【0044】

図9に示すノード290の単語「建仁寺」の単語Index、クラス、単語、及び発音は、それぞれ「金閣寺」、「地名」、「@建仁寺」、及び「ケンニンジ」である。同様に、ノード292の単語「高德庵」の単語Index、クラス、単語、及び発音は、それぞれ「金閣寺」、「地名」、「@高德庵」、及び「コートクアン」であり、ノード294の単語「三千院」の単語Index、クラス、単語、及び発音は、それぞれ「金閣寺」、「地名」、「@三千院」、及び「サンゼンイン」であり、ノード296の単語「知恩院」の単語Index、クラス、単語、及び発音は、それぞれ「金閣寺」、「地名」、「@知恩院」、及び「チオンイン」である。

10

【0045】

図11は、図9に示すクラス1のファクタリングテーブル300の詳細な図である。図11を参照して、ファクタリングテーブル300には、任意の $i \in \{1, \dots, m\}$ において、ファクタリングテーブル300の*i*列目には、単語 w_i が先行単語であるときに、クラス1を代表する既知単語である「金閣寺」が続いて生起するバイグラム確率 $P(\text{金閣寺} | w_i)$ が記憶される。

【0046】

<ハードウェア構成>

20

図12は音声認識装置140で用いられるコンピュータシステム330の外観を示し、図13はコンピュータシステム330のブロック図を示す。なお、ここで示すコンピュータシステム330は単なる例であって、他の構成も利用可能である。

【0047】

図12を参照して、コンピュータシステム330は、コンピュータ340と、全てコンピュータ340に接続された、モニタ342と、キーボード346と、マウス348と、スピーカ372と、マイクロフォン370とを含む。さらに、コンピュータ340はDVD(Digital Versatile Disk: デジタル多用途ディスク)ドライブ350と、半導体メモリドライブ352とを含む。

【0048】

30

図13を参照して、コンピュータ340はさらに、DVDドライブ350と半導体メモリドライブ352とに接続されたバス366と、全てバス366に接続された、CPU356と、コンピュータ340のブートアッププログラムを記憶するROM358と、CPU356によって使用される作業領域を提供するとともにCPU356によって実行されるプログラムのための記憶領域となるRAM360と、プログラム等の各種ファイルを記憶するためのハードディスクドライブ354とを含む。

【0049】

上述の実施の形態に係る音声認識装置140のシステムを実現するソフトウェアは、DVD-ROM362又は半導体メモリ364等の媒体に記録されたオブジェクトコードファイル又はスクリプトファイルの形で流通し、DVDドライブ350又は半導体メモリドライブ352等の読出装置を介してコンピュータ340に提供され、ハードディスクドライブ354に記憶される。CPU356がプログラムを実行する際には、プログラムはハードディスクドライブ354から読出されてRAM360に記憶される。図示しないプログラムカウンタによって指定されたアドレスから命令がフェッチされ、その命令が実行される。CPU356は、CPU356内のレジスタ、RAM360、又はハードディスクドライブ354から処理すべきデータを読出し、処理の結果を、これもまたCPU356内のレジスタ、RAM360、又はハードディスクドライブ354に記憶する。

40

【0050】

コンピュータシステム330の一般的動作は周知であるので、詳細な説明は省略する。

【0051】

50

<ソフトウェア構成>

図14～18は、未知語辞書作成部164及びメンテナンス部166を実現するためのプログラムの制御構造を示すフローチャートである。未知語辞書作成部164及びメンテナンス部166のプログラムは、実質的にはコンピュータである音声認識装置140によって実行される。

【0052】

(未知語辞書作成部164のソフトウェア構成)

図14は、未知語辞書作成部164の機能を実現するためのプログラムの制御構造を示すフローチャートである。図14を参照して、このプログラムは、単語辞書150、言語モデル152、未知語に関する情報162、及び代表単語に関する情報176を読み込むステップ400と、ステップ400の後、未知語に関する情報162に挙げられているクラスの中で、まだ木構造辞書が作成されていないクラスが存在するか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ402を含む。

10

【0053】

このプログラムはさらに、ステップ402での判定結果がYESのときに、まだ木構造辞書が作成されていないクラスのうちの1つを選択するステップ404と、ステップ404の後、ルートノードからたどって、当該クラスに属する全ての未知語が探索可能となる木構造辞書を作成するステップ406と、ステップ406の後、当該クラスに属する全ての終端ノードの各々に対して、当該終端ノードの未知語の単語Index、クラス、単語、及び発音を、それぞれ代表単語に関する情報176の中で当該クラスに割当てられている既知単語、当該クラス、当該未知語の左隣に@を付したものと、及び当該未知語の発音を割当てるステップ408と、ステップ408の後、図11に示すように当該クラスのファクタリングテーブル(代表単語について、各単語が先行するときの生起確率をリストしたテーブル)を生成し、当該ファクタリングテーブルを、当該クラスの木構造辞書の全ノードに関連付けるステップ410と、ステップ402での判定結果がNOのときに、本プログラムで作成した木構造辞書とクラスとの組を記憶するクラス木構造辞書マッピングテーブル158を生成するステップ412を含む。ステップ412の後、このプログラムは終了する。

20

【0054】

なお、ここでの木構造辞書の作成には、木構造辞書を作成するための既存のプログラムをそのまま流用することができる。また、各クラスのファクタリングテーブルには、言語モデルから代表単語のバイグラムテーブルをコピーすればよい。又は、すべてのノードに、言語モデル中の代表単語のバイグラムテーブルへのポインタを割当てておけばよい。

30

【0055】

(メンテナンス部166のソフトウェア構成)

図15～18は、メンテナンス部166の機能を実現するためのプログラムの制御構造を示すフローチャートである。図15を参照して、このプログラムは、ユーザに対して、ユーザが追加したい未知語のクラスを入力させるための入力画面を表示するステップ460と、ステップ460の後、ユーザからの入力を待機するステップ462と、ステップ462の後、ユーザが、クラスの入力の終了を要求するか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ464を含む。

40

【0056】

ステップ464の判定結果がYESのとき、このプログラムは終了する。このプログラムはさらに、ステップ464の判定結果がNOのときに、クラス木構造辞書マッピングテーブル158を参照して、ステップ462で入力されたクラスと一致するクラスを検索するステップ466と、ステップ466の後、ステップ466において検索しているクラスが存在しているか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ468と、ステップ468の判定結果がNOのときに、ユーザが入力したクラスが存在しないことを示すエラー画面を表示し、制御をステップ460に戻すステップ474と、ステップ468の判定結果がYESのときに、未知語を含む木構造辞書156を参照して、当

50

該クラスの木構造辞書を記憶しているデータファイルをオープンするステップ470と、ステップ470の後、ステップ470においてオープンしたデータファイルの内容をメモリに読み込むステップ472とを含む。

【0057】

図16を参照して、このプログラムはさらに、図15に示すステップ472の後、当該クラスに追加する未知語とその未知語の発音とをユーザに入力させるための入力画面を表示するステップ480と、ステップ480の後、ユーザの入力があるまで待機するステップ482と、ステップ482の後、ユーザが入力の終了を要求したか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ484とを含む。

【0058】

このプログラムはさらに、ステップ484の判定結果がYESのときに、メモリの内容をステップ470においてオープンされたデータファイルに上書きするステップ488と、ステップ488の後、当該データファイルをクローズし、図15に示すステップ460に制御を戻すステップ490と、ステップ484の判定結果がNOのときに、当該クラスの木構造辞書を参照して、入力された未知語が既に存在しているか否かを判定し、判定結果に応じて制御の流れを分岐させるステップ486と、ステップ486の判定結果がYESのときに、エラー画面を表示し、ステップ480に制御を戻すステップ492とを含む。

【0059】

図17を参照して、このプログラムはさらに、図16に示すステップ486の判定結果がNOのときに、ステップ482において入力された未知語及びその発音に基づいて、当該未知語を音素列に分解するステップ500と、ステップ500の後、変数MAXに音素列の音素数を代入するステップ502と、ステップ502の後、当該クラスの木構造辞書のルートノードに処理位置を移動するステップ504と、ステップ504の後、当該音素列のどの音素を参照しているかを示す変数iに0を代入するステップ506と、ステップ506の後、変数iに1を加算するステップ508と、ステップ508の後、変数iの値がMAXの値より大きいか否かを判定し、判定結果に応じて制御の流れを分岐するステップ510と、ステップ510の判定結果がNOのときに、現在参照しているノードの子ノードに当該未知語のi番目の音素が存在するか否かを判定し、判定結果に応じて制御を分岐するステップ512と、ステップ512の判定結果がNOのときに、現在参照しているノードに、当該未知語のi番目の音素の子ノードを追加するステップ514と、ステップ514の後、及びステップ512の判定結果がYESのときのいずれかにおいて、現在参照しているノードからi番目の音素のノードに処理位置を移動し、ステップ508に制御を戻すステップ516と、ステップ510の判定結果がYESのとき、すなわち、現在参照しているノードが終端ノードであるときに、当該終端ノードの単語Index、クラス、単語、及び発音を、それぞれ当該クラスの代表単語、現在参照している木構造辞書のクラス、追加した未知語の左隣に@を付したものと、及び当該未知語の発音として記憶させ、制御を図16に示すステップ480に戻すステップ518とを含む。

【0060】

<動作>

上記した構成を持つ音声認識装置140は以下のように動作する。

【0061】

(未知語辞書作成部164の動作)

図2を参照して、単語辞書150、言語モデル152、及び音響モデル154は予め用意されているものとする。

【0062】

ユーザは、登録したい未知語に関する情報162及び代表単語に関する情報176を予め用意する。

【0063】

未知語辞書管理部160を実現するプログラムは、単語辞書150と、言語モデル15

10

20

30

40

50

2と、未知語に関する情報162と、代表単語に関する情報176とを読み込み(図14に示すステップ400)、未知語に関する情報162に挙げられているクラスの中で、まだ木構造辞書が作成されていないものが存在するか否かを判定する(図14に示すステップ402)。その判定結果がYESの場合、プログラムは、木構造辞書が作成されていないクラスのいずれかを選択し(図14に示すステップ404)、選択したクラスの木構造辞書を作成する(図14に示すステップ406)。このとき、その木構造辞書の名称又はパスを記憶する。プログラムはさらに、ステップ406で作成した木構造辞書の全ての終端ノードに対して、図10に示すように単語Index、クラス、単語、発音を割当て(図14に示すステップ408)、当該クラスのファクタリングテーブルを生成し、当該クラスの木構造辞書の全ノードに、当該ファクタリングテーブルを関連付ける(図14に示すステップ410)。プログラムは、未知語に関する情報162に挙げられたクラスのみだけステップ402~ステップ410の処理を繰り返す。

10

【0064】

こうした処理を繰り返した結果、図14に示すステップ402において、判定結果がNOとなる。この場合、プログラムは、各木構造辞書に対して記憶しておいたその名称又はパス名と、対応するクラス名とを対応付けてクラス 木構造辞書マッピングテーブル58を生成し(図14に示すステップ412)、処理を終了する。

【0065】

(メンテナンス部166の動作)

未知語を新たに木構造辞書に登録するときには、メンテナンス部166は以下のように動作する。図2を参照して、未知語辞書作成部164により、未知語を含む木構造辞書156及びクラス 木構造辞書マッピングテーブル158が予め作成されているものとする。

20

【0066】

以下の説明では、新たに登録する未知語、そのクラス、及びその発音を予め決定しているものとする。

【0067】

メンテナンス部166を実現するプログラムは、ユーザに対してクラスを指定させるための入力画面を表示し(図15に示すステップ460)、ユーザが入力画面にクラス名を入力するまで待機する(図15に示すステップ462)。ここでは、クラス名又は処理の終了の指示のいずれかの入力が可能である。

30

【0068】

ユーザが、新規に登録したい未知語のクラスを入力したものとする。プログラムは、ユーザが、入力を終了するか否かを判定する(図15に示すステップ464)。ユーザがクラスを入力したので、判定結果はNOであり、プログラムはクラス 木構造辞書マッピングテーブル158を参照して、ユーザが入力したクラスを検索し(図15に示すステップ466)、そのクラスの木構造辞書が存在するか否かを判定する(図15に示すステップ468)。ユーザの入力したクラスが存在せず、判定結果がNOであれば、プログラムはエラー画面を表示してステップ460に処理を戻し、次の入力を待つ。ユーザが入力したクラスが存在し、ステップ468での判定結果がYESであれば、プログラムは、検索された木構造辞書156のデータファイルをオープンし(図15に示すステップ470)、オープンしたデータファイルの内容をメモリに読み込む(図15に示すステップ472)。

40

【0069】

次に、プログラムは、当該クラスに登録する未知語とその発音とをユーザに入力させるための入力画面を表示し(図16に示すステップ480)、ユーザからの入力を待つ(図16に示すステップ482)。ここでは、未知語と発音、又はこの処理を終了する指示のいずれかの入力が可能である。

【0070】

ユーザが、新たに登録したい未知語とその発音とを入力したものとする。図16に示すステップ484における判定結果はNOとなり、プログラムは、メモリに読み込んだ木構造

50

辞書を参照して、ユーザによって入力された未知語が既に存在しているか否かを判定する（図16に示すステップ486）。判定結果がYESのとき、プログラムは、エラー画面を表示して（図16に示すステップ492）、ステップ480に処理を戻す。ステップ486の判定結果がNOのとき、プログラムは以下のような処理を実行する。

【0071】

プログラムはまず、ユーザによって入力された未知語を音素列に分解する（図17に示すステップ500）。次に、未知語の単語の音素数を示す変数MAXに音素数を代入する（図17に示すステップ502）。プログラムは、当該クラスの木構造辞書のルートノードに処理位置を移動し（図17に示すステップ504）、未知語の音素を1つずつ参照するために、先ず変数iに0を代入して（図17に示すステップ506）、変数iに1を加算する（図17に示すステップ508）。このとき、プログラムは、登録したい未知語の1番目の音素を参照していることになる。プログラムは、変数iの値がMAXの値を超えているか否かを判定する（図17に示すステップ510）。現在、変数iの値は1なので、判定結果はNOである。プログラムは、ルートノードの子ノードに、登録したい未知語のi番目の音素が存在するか否かを判定し（図17に示すステップ512）、存在すればi番目の音素の子ノードに処理位置を移動し（図17に示すステップ516）、存在しなければ、現在参照しているノードにi番目の音素の子ノードを新たに追加して（図17に示すステップ514）、その子ノードに処理位置を移動する（図17に示すステップ516）。その後、プログラムは、ステップ508に処理を戻し、2番目以降の音素について同様の処理を繰り返す。

【0072】

こうした処理を繰り返した結果、ステップ510において変数iの値がMAXの値を超えたと判定されたとき、プログラムは、それまでの処理で追加された終端ノードに、登録したい未知語のための単語Index、当該未知語の属するクラス、当該未知語、及び当該未知語の発音を記憶させ（図17に示すステップ518）、図16に示すステップ480以後の処理を繰り返す。ここでは、単語Index及びクラスとしては、同じ木構造辞書中の任意の終端ノードに記憶されている単語Index及びクラスと同じものをそれぞれ記憶させればよい。

【0073】

ステップ484において、ユーザが入力の終了を要求した場合、プログラムは、ステップ470においてオープンされたデータファイルにメモリに記憶された木構造辞書の内容を上書きし（図16に示すステップ488）、データファイルをクローズし（図16に示すステップ490）、図15に示すステップ460に処理を戻す。

【0074】

一方、ステップ464において、ユーザが入力の終了を要求した場合、プログラムは、処理を終了する。

【0075】

（音声認識部170による未知語を含む木構造辞書156の探索方法）

音声認識部170は、単語辞書150、言語モデル152、音響モデル154、未知語を含む木構造辞書156、クラス木構造辞書マッピングテーブル158、使用クラスの情報172、及び音声データ174を読む。

【0076】

未知語を含む木構造辞書156にはクラス1～クラスpまでの未知語の木構造辞書が記憶されており、使用クラスの情報172には、少なくともクラス1及びクラスpのクラス名が挙げられており、少なくともクラス2のクラス名が挙げられていないものとする。

【0077】

図18は、未知語を含む木構造辞書156の探索方法について説明するための図である。図18を参照して、音声認識部170は、音声認識するために、どのクラスの未知語を探索するかを決定するためのスイッチであるスイッチ550、552、及び554等と、単語辞書150と、未知語を含む木構造辞書156とを用いる。これらスイッチは、プロ

10

20

30

40

50

グラム上では、探索先の単語辞書を使用するかしないかを指定するフラグとして保持される。

【0078】

スイッチ550、552、及び554は、それぞれクラス1、クラス2、及びクラスpに対応している。その他のクラスについても同様である。スイッチがONの場合、音声認識部170は、当該スイッチに対応するクラスの木構造辞書を探索し、OFFの場合、当該スイッチに対応するクラスの木構造辞書を探索しない。プログラムは、使用クラスの情報172にクラスの名前が挙げられている場合、そのクラスに対応するスイッチをONとし、クラスの名前が挙げられていない場合、そのクラスに対応するスイッチをOFFとする。

10

【0079】

一方、単語辞書150にはそのようなスイッチはなく、常に探索の対象となる。

【0080】

したがって、音声認識部170は、音声認識する場合、単語辞書150と、対応するスイッチ(フラグ)がオンとなっている未知語の木構造辞書を探索する。

【0081】

<本実施の形態の効果>

以上の説明から明らかなように、本実施の形態に係る音声認識装置140を使用することにより、未知語辞書作成部164が未知語を含む木構造辞書156を作成するので、音声認識部170は、音声データ174に未知語が含まれていても、言語モデルを変更せずに、代表単語の生起確率と同じ確率で未知語を認識することができる。言語モデルが変更されないので、言語モデル中の確率の信頼性が低下することもない。

20

【0082】

また、使用クラスの情報172により、音声認識時に探索させる未知語の木構造辞書をユーザが選択することができるので、未知語を含む木構造辞書156のクラスの数が増大であっても、音声認識部170は、全てのクラスの未知語の木構造辞書を探索する必要はなく、処理時間に支障はない。

【0083】

さらに、未知語を含む木構造辞書156が、本実施の形態に係る未知語辞書作成部164によっていったん作成されたとしても、ユーザは、未知語を含む木構造辞書156に、新しい未知語を自由に登録することができる。

30

【0084】

<変形例>

上述した実施の形態では、図3に示す言語モデル152には、ユニグラムテーブル224と全てのバイグラム確率が予め記憶されたバイグラムテーブル226とが含まれていた。しかし、本発明はそのような実施の形態には限定されない。Nを3以上の整数として、言語モデル152は、ユニグラムテーブル224及びバイグラムテーブル226に代えて、Nグラム確率を記憶させたNグラム木構造を含んでも良い。

【0085】

図19は、Nグラム木構造を示す図である。通常は、ユニグラム確率、バイグラム確率、及びトライグラム確率等は、予め、新聞のようなサンプル文書(以下、学習コーパス)から、1~3個の任意の単語列を計数することによって算出される。学習コーパス中の全ての単語のユニグラム確率を算出することは可能である。しかし、全てのバイグラム確率及びトライグラム確率が算出可能であるとは限らない。

40

【0086】

そこで、学習コーパスから算出されないバイグラム確率及びトライグラム確率を学習コーパスから推定するための、バックオフスムージングと称される方法がある。上述の実施の形態において、学習コーパスから全てのバイグラム確率を算出することができない場合、バックオフスムージングによってバイグラム確率を推定し、バイグラムテーブル226に記憶させれば良い。

50

【 0 0 8 7 】

以下、バックオフスムージングについて簡単に説明する。バックオフスムージングについて説明するために、準備として、学習コーパスから算出されるユニグラム確率、バイグラム確率、及びトライグラム確率等を木構造で記憶するためのNグラム木構造について説明する。

【 0 0 8 8 】

以下の説明では、全ての既知単語を w_1 、 w_2 、...、 w_m で表わす。図 19 を参照して、Nグラム木構造 600 は、ルートノード 666 と、ルートノード 666 から分岐し、且つ、全ての既知単語のそれぞれに対応するノード 602 (w_1)、ノード 604 (w_2)、ノード 606 (w_3)、ノード 608 (w_m) 等と、ルートノード 666 との距離が 1 であるノード 602 ~ 608 等の各々について、当該ノードから分岐し、且つ、学習コーパスにおいて当該ノードに対応する単語の 1 つ前に出現する全ての単語に対応するノード 610 (w_2, w_1)、ノード 612 (w_4, w_1)、ノード 614 (w_7, w_1)、ノード 616 (w_3, w_2)、ノード 618 (w_9, w_2)、及びノード 620 (w_1, w_3) 等と、ルートノード 666 との距離が 2 あるノード 610 ~ 620 等から分岐し、且つ、学習コーパスにおいて、当該ノードに対応する単語及び当該ノードの親ノードに対応する単語の順番により定まる単語列の 1 つ前に出現する全ての単語に対応するノード 622 (w_3, w_2, w_1)、ノード 624 (w_5, w_2, w_1)、ノード 626 (w_8, w_2, w_1)、ノード 628 (w_2, w_4, w_1)、ノード 630 (w_9, w_4, w_1)、及びノード 632 (w_2, w_7, w_1) 等を含む。Nグラム木構造 600 はさらに、ルートノード 666 との距離が 1 であるノードの各々について設けられ、当該ノードに対応する単語のユニグラム確率及びユニグラムバックオフ係数を記憶する領域 634 (w_1)、636 (w_2)、638 (w_3)、及び 640 (w_m) 等と、ルートノード 666 との距離が 2 であるノードの各々について設けられ、当該ノードに対応する単語の後に、当該ノードの親ノードに対応する単語が学習コーパスにおいて生起するバイグラム確率及びバイグラムバックオフ係数を記憶する領域 642 (w_2, w_1)、644 (w_4, w_1)、646 (w_7, w_1)、648 (w_3, w_2)、650 (w_9, w_2)、652 (w_1, w_3) 等と、ルートノード 666 との距離が 3 であるノードの各々について、当該ノードに対応する単語及び当該ノードの親ノードの単語の後に、当該ノードの親ノードの親ノードに対応する単語が学習コーパスにおいて生起するトライグラム確率を記憶する領域 654 (w_3, w_2, w_1)、656 (w_5, w_2, w_1)、658 (w_8, w_2, w_1)、660 (w_2, w_4, w_1)、662 (w_9, w_4, w_1)、664 (w_2, w_7, w_1) 等とを含む。なお、ここでルートノード 666 は、対応する単語がないダミーノードである。ユニグラムバックオフ係数及びバイグラムバックオフ係数は、学習コーパスから算出することができないバイグラム確率及びトライグラム確率を推定するために利用されるものである。ユニグラムバックオフ係数及びバイグラムバックオフ係数は、バックオフスムージングにおいてよく知られたものなので、ここでは、その詳細について説明しない。

【 0 0 8 9 】

Nグラム木構造 600 を利用することによって、ルートノード 600 から順次ノードをたどることによって、学習コーパスから算出されるユニグラム確率、バイグラム確率、トライグラム確率、ユニグラムバックオフ係数、及びバイグラムバックオフ係数を得ることができる。

【 0 0 9 0 】

以下、学習コーパスから算出されないバイグラム確率がある場合に、Nグラム木構造 600 を利用してバイグラム確率を推定する方法について説明する。ここで、単語 w_p の後に単語 w_q が生起するバイグラム確率 $P(w_q | w_p)$ が学習コーパスから算出されないものとする。 $p, q \in \{1, 2, \dots, m\}$ である。この場合、単語 w_q が生起するユニグラム確率を $P(w_q)$ とし、単語 w_p のユニグラムバックオフ係数を $B(w_p)$ とすると、 $P(w_q | w_p) = B(w_p) \times P(w_q)$ とすればよい。 $B(w_p)$ 及び $P(w_q)$ は、Nグラム木構造 600 においてルートノード 666 との距離が 1 であり、且つ、

10

20

30

40

50

それぞれ単語 w_p 並びに w_q に対応するノードからユニグラム確率及びユニグラムバックオフ係数から取得することができる。なお、Nグラム木構造 600 を利用してトライグラム確率を推定する方法については、バックオフスムージングにおいてよく知られた方法なので、ここではその詳細について説明しない。

【0091】

言語モデル 152 が、ユニグラムテーブル 224 及びバイグラムテーブル 226 に代えて、上記のNグラム木構造 600 を含む場合、単語辞書 150 の全ての終端ノードの各々に、Nグラム木構造 600 を関連付ける。単語辞書 150 の終端ノードのNグラム確率は、単語辞書 150 の終端ノードに到達したときに、Nグラム木構造 600 から動的に求める。ルートノードとの距離が k (k は 1、2、...、 $m-1$ の任意の数) である非終端ノードのファクタリングテーブルには、近似値として当該非終端ノードから子ノードをたどっていくことにより到達できる終端ノードの単語が生起するバイグラム確率の最大値が記憶される。

10

【0092】

図 9 に示すクラス 1 のファクタリングテーブル 300 には、バイグラム確率ではなく、任意の $(N-1)$ 個の既知単語の組合せの各々に対して、当該組合せが先行するときに、そのクラスの代表単語が生起するNグラム確率が記憶される。

【0093】

また、上述の実施の形態では、図 3 に示すノード 200 との距離が k (k は 1、2、...、 $m-1$ の任意の数) である非終端ノードに対して、ファクタリングテーブル 222 と同様なテーブルが関連づけられていた。しかし、本発明はそのような実施の形態には限定されない。 p を 2 以上の整数として、ノード 200 との距離が $(k+1)$ 以上であり、かつ p より小さい非終端ノードに、ファクタリングテーブル 222 と同様なテーブルを関連付けても良い。

20

【0094】

さらに、上述の実施の形態では、未知語を含む木構造辞書 156 のクラスを「地名」のみに限定していた。しかし、本発明はそのような実施の形態には限定されない。各国の音声データを認識したい場合、各国毎に、「ランドマーク」、「観光地」、及び「人名」等のクラスの未知語を収集して、未知語辞書作成部 164 によって、それらのクラスのための未知語を含む木構造辞書を作成させる。音声認識部 170 によって当該音声データを認識させたい場合、使用クラスの情報 172 に、対象国の言語に該当するクラスを指定して、音声認識させれば良い。

30

【0095】

また、各地方の方言を含んだ音声データを音声認識したい場合、各地方の方言を、標準語のクラスに割当てたものを未知語に関する情報 162 として、未知語辞書作成部 164 によって、未知語を含む木構造辞書 156 を作成させる。ある地方の方言を含んだ音声データを音声認識するとき、使用クラスの情報 172 において、当該地方の方言のためのクラスを指定して、音声認識部 170 によって、音声認識させれば良い。

【0096】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内のすべての変更を含む。

40

【図面の簡単な説明】

【0097】

【図 1】木構造辞書の概念を説明するための模式図である。

【図 2】本発明の実施の形態に係る音声認識装置 140 の概略の機能的構成を示すブロック図である。

【図 3】図 2 に示す単語辞書 150 及び言語モデル 152 の詳細な図である。

【図 4】図 3 に示すユニグラムテーブル 224 の詳細な図である。

50

【図 5】図 3 に示すバイグラムテーブル 2 2 6 の詳細な図である。

【図 6】図 3 に示す単語辞書 1 5 0 の終端ノードに記憶されているものについて説明するための図である。

【図 7】図 3 に示すファクタリングテーブル 2 2 2 に記憶されているものについて説明するための図である。

【図 8】図 2 に示す未知語を含む木構造辞書 1 5 6 の詳細について説明するための図である。

【図 9】図 8 に示すクラス 1 の木構造辞書 2 7 0 の詳細について説明するための図である。

【図 10】図 9 に示すクラス 1 の木構造辞書 2 7 0 の終端ノードに記憶されているものについて説明するための図である。 10

【図 11】図 9 に示すクラス 1 のファクタリングテーブル 3 0 0 に記憶されているものについて説明するための図である。

【図 12】コンピュータシステム 3 3 0 のハードウェア構成を示す図である。

【図 13】コンピュータシステム 3 3 0 のブロック図である。

【図 14】図 2 に示す未知語辞書作成部 1 6 4 を実現するコンピュータプログラムの制御構造を示すフローチャートである。

【図 15】図 2 に示すメンテナンス部 1 6 6 を実現するコンピュータプログラムの制御構造を示すフローチャートである。

【図 16】図 2 に示すメンテナンス部 1 6 6 を実現するコンピュータプログラムの制御構造を示すフローチャートである。 20

【図 17】図 2 に示すメンテナンス部 1 6 6 を実現するコンピュータプログラムの制御構造を示すフローチャートである。

【図 18】図 2 に示す未知語を含む木構造辞書 1 5 6 の使用方法について説明するための図である。

【図 19】N グラム木構造について説明するための図である。

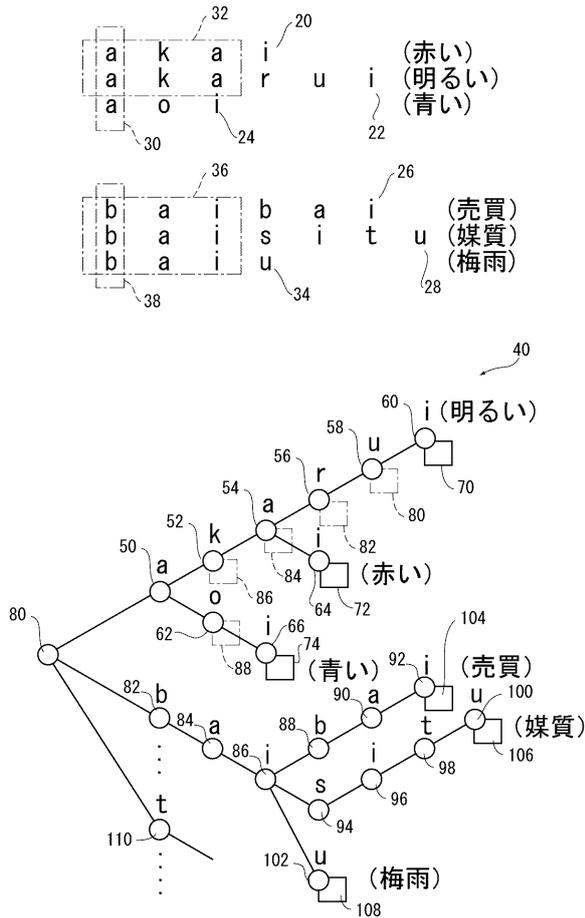
【符号の説明】

【 0 0 9 8 】

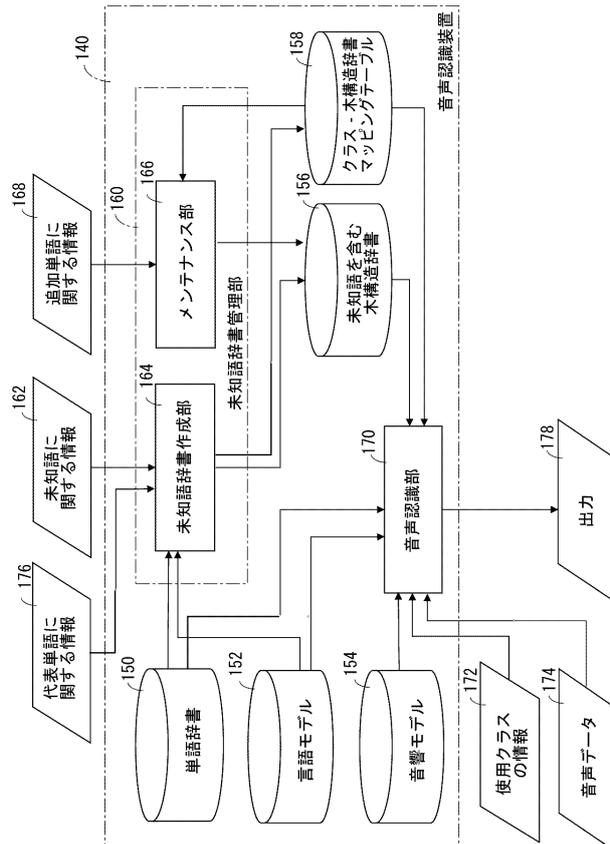
1 4 0	音声認識装置	
1 5 0	単語辞書	30
1 5 2	言語モデル	
1 5 4	音響モデル	
1 5 6	未知語を含む木構造辞書	
1 5 8	クラス 木構造辞書マッピングテーブル	
1 6 0	未知語辞書管理部	
1 6 2	未知語に関する情報	
1 6 4	未知語辞書作成部	
1 6 6	メンテナンス部	
1 6 8	追加単語に関する情報	
1 7 0	音声認識部	40
1 7 2	使用クラスの情報	
1 7 4	音声データ	
1 7 6	代表単語に関する情報	
1 7 8	音声認識されたテキスト	
2 2 2	ファクタリングテーブル	
2 2 4	ユニグラムテーブル	
2 2 6	バイグラムテーブル	
2 7 0	クラス 1 の木構造辞書	
2 7 2	クラス 2 の木構造辞書	
2 7 4	クラス P の木構造辞書	50

300 クラス1のファクタリングテーブル

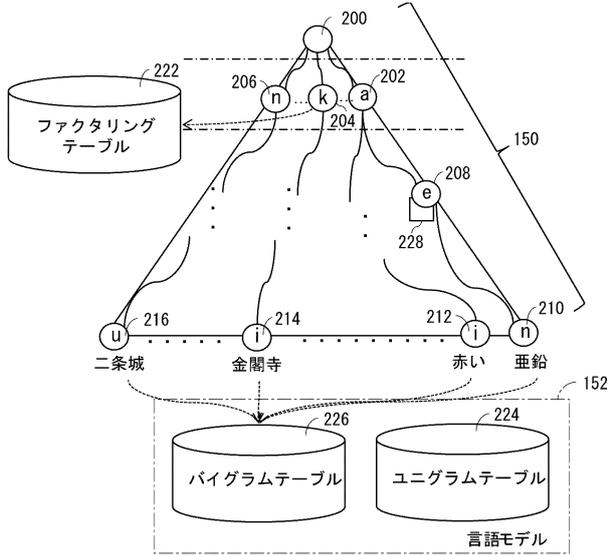
【図1】



【図2】



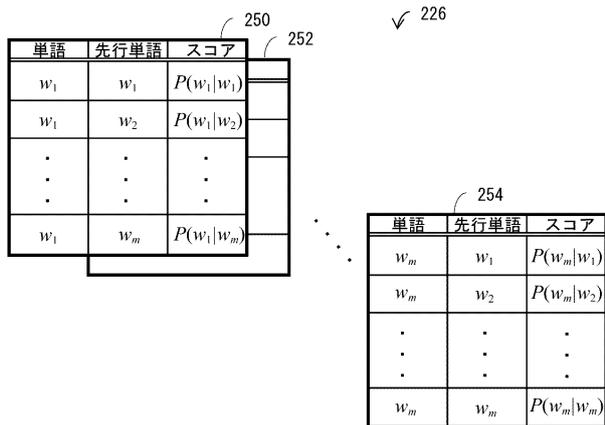
【図3】



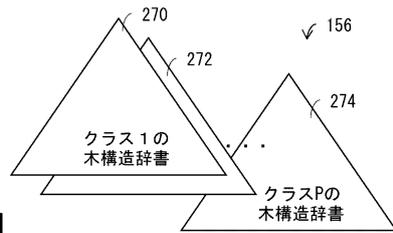
【図4】

単語	スコア
w_1	$P(w_1)$
w_2	$P(w_2)$
⋮	⋮
w_m	$P(w_m)$

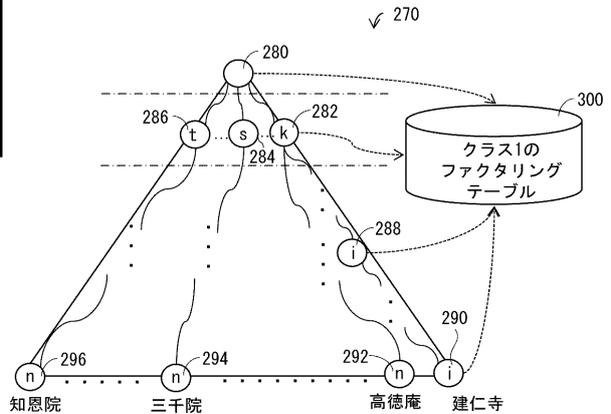
【図5】



【図8】



【図9】



【図6】

単語INDEX	クラス	単語	発音
⋮	⋮	⋮	⋮
亜鉛	化学物質	[亜鉛]	アエン
⋮	⋮	⋮	⋮
赤い	形容詞	[赤い]	アカイ
⋮	⋮	⋮	⋮
金閣寺	地名	[金閣寺]	キンカクジ
⋮	⋮	⋮	⋮
二条城	地名	[二条城]	ニジョウジョウ
⋮	⋮	⋮	⋮

【図7】

w_1	w_2	⋯	w_m
$\max_{j \in \{1,2,\dots,p\}} P(w_{i_j} w_1)$	$\max_{j \in \{1,2,\dots,p\}} P(w_{i_j} w_2)$	⋯	$\max_{j \in \{1,2,\dots,p\}} P(w_{i_j} w_m)$

$\{i_1, i_2, \dots, i_p\} \subset \{1, 2, \dots, m\}$

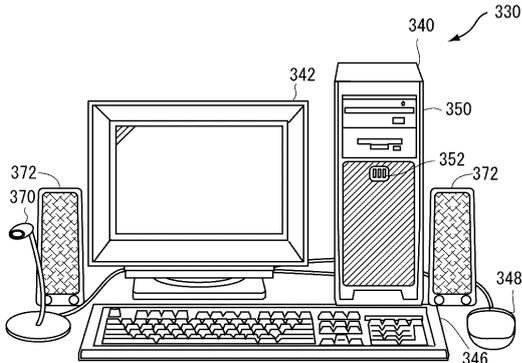
【図10】

単語INDEX	クラス	単語	発音
...
金閣寺	地名	[@建仁寺]	ケンニンジ
...
金閣寺	地名	[@高德庵]	コトクアン
...
金閣寺	地名	[@三千院]	サンゼンイン
...
金閣寺	地名	[@知恩院]	チオンイン
...

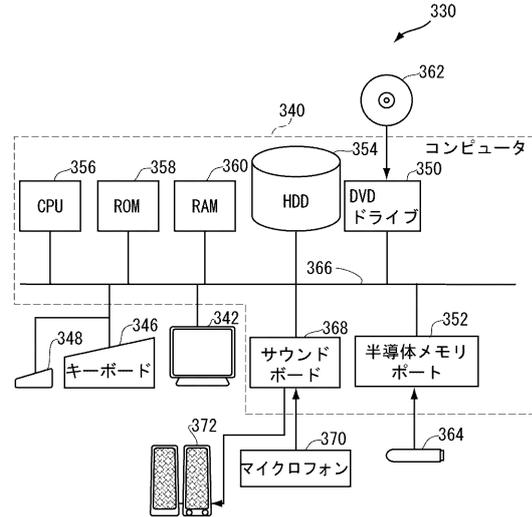
【図11】

w_1	w_2	...	w_m
$P(\text{金閣寺} w_1)$	$P(\text{金閣寺} w_2)$...	$P(\text{金閣寺} w_m)$

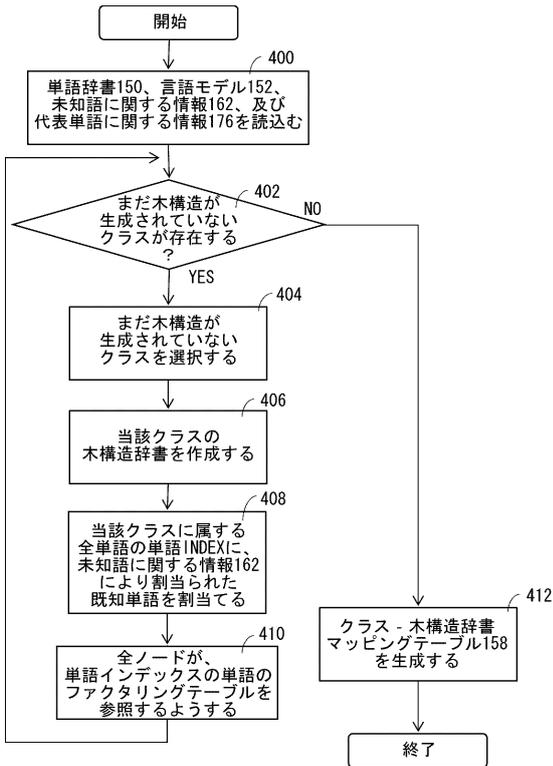
【図12】



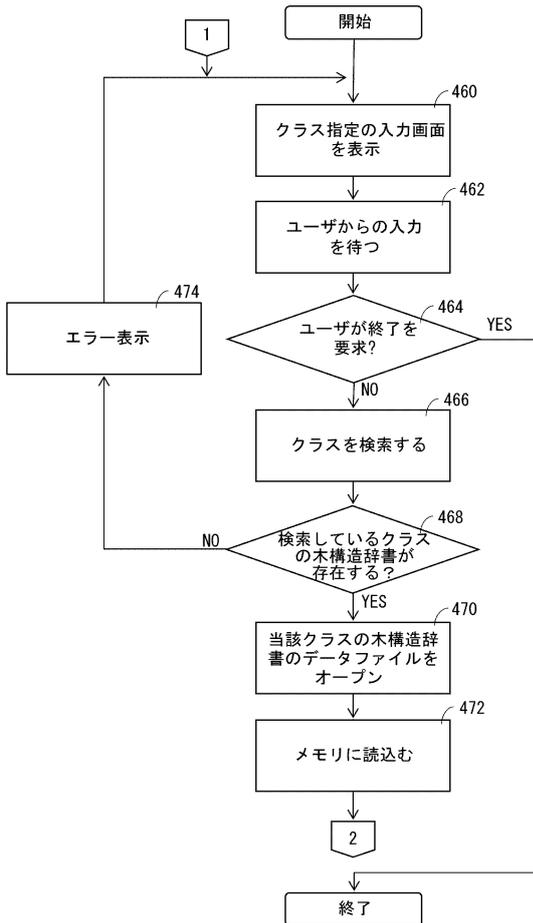
【図13】



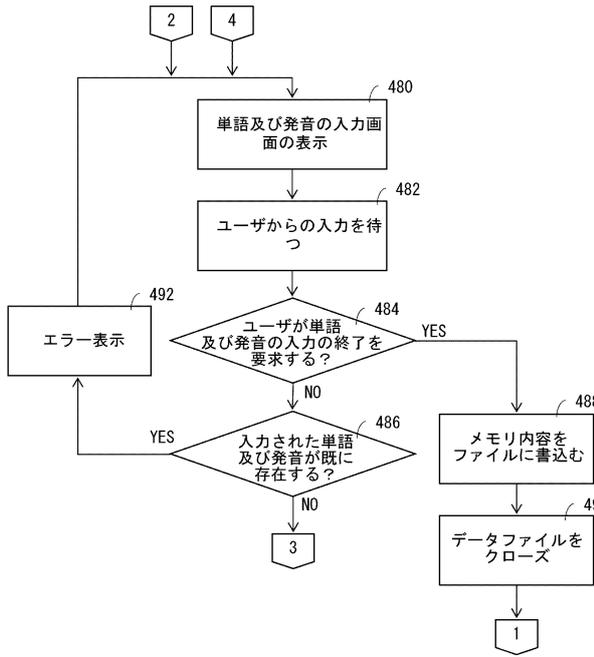
【図14】



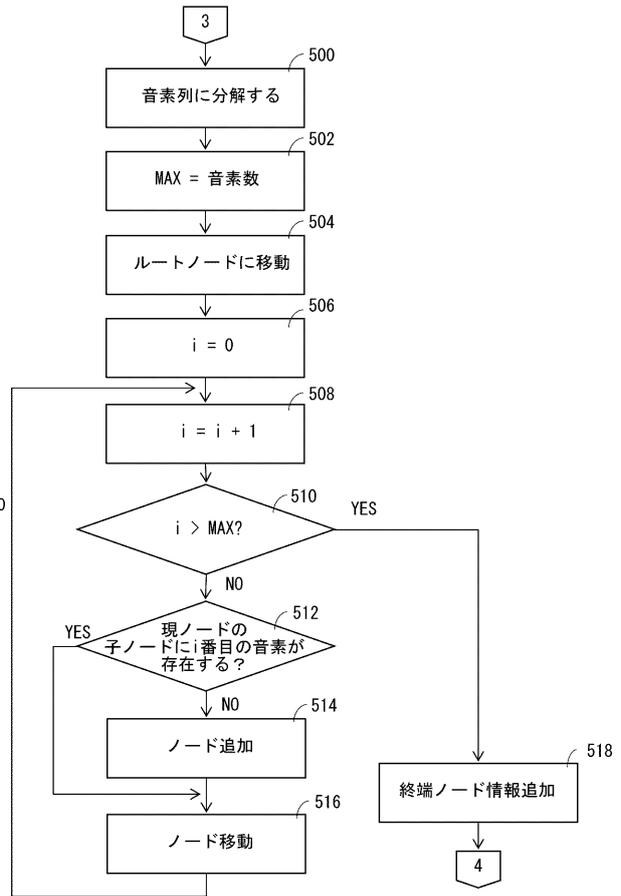
【図15】



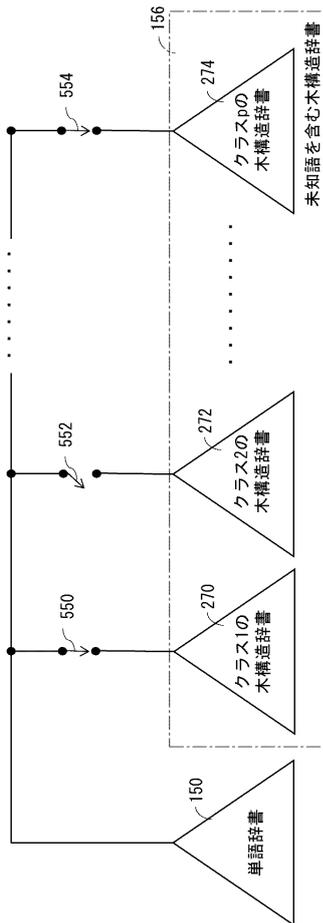
【図16】



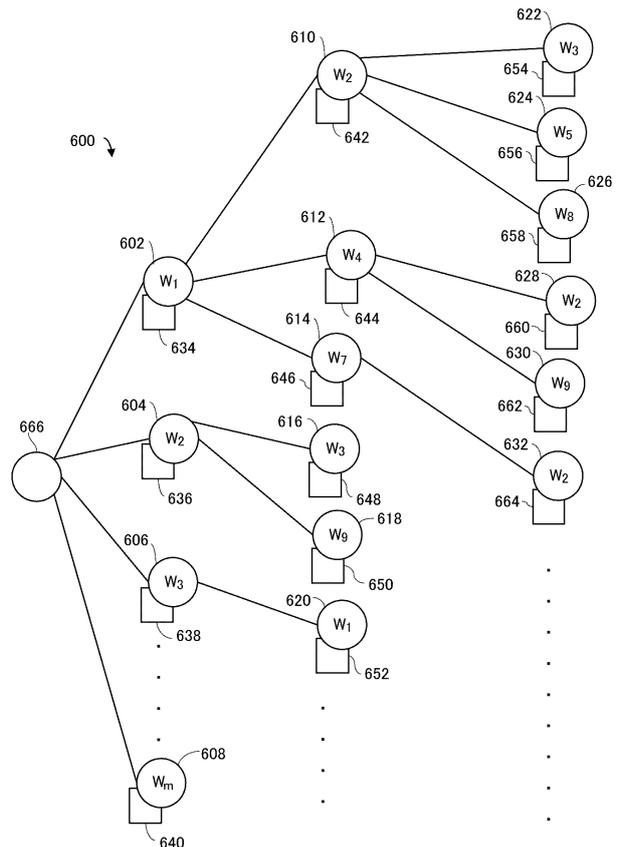
【図17】



【図18】



【図19】



フロントページの続き

- (72)発明者 中村 哲
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 葦苒 豊
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 竹澤 寿幸
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官 山下 剛史

- (56)参考文献 特開2001-255889(JP, A)
国際公開第2007/044568(WO, A2)
大西茂彦他, "大語彙連続音声認識における未知語のsub-wordモデリング手法", 電子情報通信学会技術研究報告, Vol.101, No.32(2001-04), pp.33-39
谷垣宏一他, "クラスに依存した語彙の確率的記述に基づく階層型言語モデル", 電子情報通信学会論文誌, Vol.J84-D-II, No.11(2001-11), pp.2371-2378

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28
G06F 17/22 - 17/28