(19)日本国特許庁（ＪＰ）　(12)特　許　公　報（Ｂ２）　(11)特許番号

第2886121号

(45)発行日　平成11年(1999)４月26日　(24)登録日　平成11年(1999)２月12日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５３５	Ｇ１０Ｌ　3/00	５３５
	５６１		５６１Ｇ

請求項の数４(全 14 頁)

(21)出願番号　特願平7－292685

(22)出願日　平成７年(1995)11月10日

(65)公開番号　特開平9－134192

(43)公開日　平成９年(1997)５月20日

審査請求日　平成７年(1995)11月10日

(73)特許権者　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　政瀧　浩和
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　匂坂　芳典
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(74)代理人　弁理士　青山　葆　(外２名)

審査官　樫本　剛

最終頁に続く

(54)【発明の名称】　統計的言語モデル生成装置及び音声認識装置

(57)【特許請求の範囲】

【請求項１】　所定の話者の発声音声文を書き下した学習用テキストデータに基づいて、すべての語彙を品詞毎にクラスタリングされた品詞クラスに分類し、それらの品詞クラス間のバイグラムを初期状態の統計的言語モデルとして生成する生成手段と、

上記生成手段によって生成された初期状態の統計的言語モデルに基づいて、単語の品詞クラスからの分離することができる第１の分離クラス候補と、１つの単語と１つの単語との結合、１つの単語と複数の単語の単語列との結合、複数の単語の単語列と１つの単語との結合、複数の単語の単語列と、複数の単語の単語列との結合とを含む連接単語又は連接単語列の結合によって単語の品詞クラスから分離することができる第２の分離クラス候補とを検索する検索手段と、

上記検索手段によって検索された第１と第２の分離クラス候補に対して、次単語の予測の難易度を表わす所定のエントロピーを用いて、クラスを分離することによる当該エントロピーの減少量を計算する計算手段と、

上記計算手段によって計算された上記第１と第２の分離クラス候補に対するエントロピーの減少量の中で最大のクラス分離を選択して、選択されたクラスの分離を実行することにより、品詞のバイグラムと可変長Ｎの単語のＮ－グラムとを含む統計的言語モデルを生成する分離手段と、

上記分離手段によって生成された統計的言語モデルのクラス数が所定のクラス数になるまで、上記分離手段によって生成された統計的言語モデルを処理対象モデルとして、上記検索手段の処理と、上記計算手段の処理と、上記分離手段の処理とを繰り返すことにより、所定のクラ

ス数を有する統計的言語モデルを生成する制御手段とを備えたことを特徴とする統計的言語モデル生成装置。

【請求項２】　入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、

上記音声認識手段は、品詞のバイグラムと可変長Ｎの単語のＮ－グラムとを含む統計的言語モデルを用いて音声認識することを特徴とする音声認識装置。

【請求項３】　上記統計的言語モデルは、請求項１記載の統計的言語モデル生成装置によって生成されたことを特徴とする音声認識装置。

【請求項４】　入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた連続音声認識装置において、

上記音声認識手段は、請求項１記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行うことを特徴とする連続音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、学習用テキストデータに基づいて統計的言語モデルを生成する統計的言語モデル生成装置、及び上記統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識する音声認識装置に関する。

【０００２】

【従来の技術】近年、連続音声認識装置において、その性能を高めるために言語モデルを用いる方法が研究されている。これは、言語モデルを用いて、次単語を予測し探索空間を削減することにより、認識率の向上および計算時間の削減の効果を狙ったものである。最近盛んに用いられている言語モデルとしてＮ－グラム（Ｎ－ｇｒａｍ）がある。これは、大規模なテキストデータを学習し、直前のＮ－１個の単語から次の単語への遷移確率を統計的に与えるものである。複数Ｌ個の単語列w_1^L＝w_1，w_2，…，w_Lの生成確率Ｐ（w_1^L）は次式で表される。

【０００３】

【数１】

$$P(w_1^L) = \prod_{t=1}^{L} P(w_t \mid w_{t+1-N}^{t-1})$$

【０００４】ここで、w_tは単語列w_1^Lのうちｔ番目の１つの単語を表し、w_i^jはｉ番目からｊ番目の単語列を

表わす。上記数１において、確率Ｐ（w_t｜w_{t+1-N}^{t-1}）は、Ｎ個の単語からなる単語列w_{t+1-N}^{t-1}が発声された後に単語w_tが発声される確率であり、以下同様に、確率Ｐ（Ａ｜Ｂ）は単語又は単語列Ｂが発声された後に単語Ａが発声される確率を意味する。また、数１における「Π」はｔ＝１からＬまでの確率Ｐ（w_t｜w_{t+1-N}^{t-1}）の積を意味し、以下同様である。

【０００５】Ｎ－グラムは極めて単純なものでありながら、構築の容易さ、統計的音響モデルとの相性の良さ、認識率向上や計算時間の短縮の効果が大きい等の理由で、連続音声認識には非常に有効である（例えば、従来文献１「Ｌ．Ｒ．Ｂａｈｌほか，“Ａ Ｍａｘｉｍｕｍ Ｌｉｋｅｌｉｈｏｏｄ Ａｐｐｒｏａｃｈ ｔｏ Ｃｏｎｔｉｎｕｏｕｓ Ｓｐｅｅｃｈ Ｒｅｃｏｇｎｉｔｉｏｎ”，ＩＥＥＥ Ｔｒａｎｓａｃｔｉｏｎ ｏｎ Ｐａｔｔｅｒｎ Ａｎａｌｙｓｉｓ ａｎｄ Ｍａｃｈｉｎｅ Ｉｎｔｅｌｌｉｇｅｎｃｅ，ｐｐ．１７９－１９０，１９８３年」、従来文献２「Ｐ．Ｃ．Ｗｏｏｄｌａｎｄほか，“ＴＨＥ １９９４ ＨＴＫ Ｌａｒｇｅ Ｖｏｃａｂｕｌａｒｙ Ｓｐｅｅｃｈ Ｒｅｃｏｇｎｉｔｉｏｎ Ｓｙｓｔｅｍ”，Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ＩＣＡＳＳＰ９５’，Ｖｏｌ．１，ｐｐ．７３－７６，１９９５年」、従来文献３「村上ほか，“単語のｔｒｉｇｒａｍを利用した文音声認識と自由発話認識への拡張”，電子情報通信学会技術研究報告，ＳＰ９３－１２７，ｐｐ７１－７８，平成６年」参照。）。

【０００６】一般に、Ｎ－グラムの言語モデルは、Ｎを大きくすると長い単語連鎖を取り扱うことにより次単語の精度は高くなるが、パラメータ数が多くなり、学習データ量が少ない場合は出現頻度の低い単語に信頼できる遷移確率を与えることはできない。例えば語彙数が５，０００語のとき、トライグラム（ｔｒｉｇｒａｍ）（Ｎ＝３）の全ての単語の遷移組は（５，０００）^{3}＝１，２５０億であるから、信頼できる遷移確率を求めるためには、数千億単語以上からなる膨大なテキストデータが必要となる。これだけの膨大なテキストデータを集めるのは事実上不可能である。逆に、Ｎを小さくすると、遷移確率の信頼性は高くなるが、短い単語連鎖しか取り扱うことができず、次単語の予測精度は低くなる。

【０００７】

【発明が解決しようとする課題】この問題を解決するため、次のような方法が提案されている。

（１）補間による未学習遷移確率の推定方法

この方法は、例えば、Ｄｅｌｅｔｅｄ Ｉｎｔｅｒｐｏｌａｔｉｏｎ（削除補間法）（例えば、従来文献４「Ｆ．Ｊｅｌｉｎｅｋほか，“Ｉｎｔｅｒｐｏｌａｔｅｄ ｅｓｔｉｍａｔｉｏｎ ｏｆ Ｍａｒｋｏｖ Ｓｏｕｒｃｅ Ｐａｒａｍｅｔｅｒｓ ｆｒｏｍ Ｓｐａｒｓｅ Ｄａｔａ”，Ｐｒｏｃｅｅｄｉｎｇｓｏｆ Ｗｏｒｋｓｈｏｐ Ｐａｔｔｅｒｎ Ｒｅｃｏｇｎｉｔｉｏ

5

n inPractice, pp. 381−37, 1980年」参照。）や、Back−off Smoothing法（従来文献5「S. M. Katz, “Estimation of Probabilities from Sparse Data for the Language model Componentof a Speech Recognizer”, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP−35, No. 3, pp. 400−401, 1987年3月」参照。）等に代表される方法で、小さいNのN−グラム（N−gram）の値で遷移確率を補間することにより、学習用テキストデータには存在しない単語遷移に対しても、遷移確率を与えることができる。しかしながら、出現頻度の低い単語に関しては信頼できる遷移確率を与えられない恐れがある。

6

＊【０００８】（２）クラスN−グラムによるパラメータ数の削減方法

この方法は、相互情報量に基づくクラスタリング（例えば、従来文献6「P. F. Brownほか, “Class−Based n−gram modelsof natural language”, ComputationalLinguistics, Vol. 18, No. 4, pp467−479, 1992年」参照。）や、品詞（従来文献7「周ほか, “確率モデルによる日本語の大語彙連続音声認識”, 情報処理学会, 第５１回全国大会講演論文集, pp119−120, 平成7年」参照。）等によるクラス間のN−グラムを考えたもので、L個の単語の文生成確率P（w_1^L）は一般に次式で表される。

【０００９】

【数２】

$$P(w_1^L) = \prod_{t=1}^{L} P(w_t \mid c_t) \cdot P(c_t \mid c_{t-N+1}^{t-1})$$

【００１０】ここで、c_tは単語w_tの属するクラスを表し、c_i^jはｉ番目からｊ番目のクラス列を表わす。上記数２で、P（$c_t \mid c_{t-N+1}^{t+1}$）は、直前の（N−１）個の単語の属するクラスから次の単語の属するクラスへの遷移確率を表す。クラス数が５０のとき、トライグラムの全てのクラス間の遷移の組は50^3＝１２５，０００であるから、数十万単語程度と単語N−グラムに比べてかなり小規模なテキストデータで遷移確率が求められると考えられる。しかしながら、単語間の特有な連接関係を表現することができないので、次単語の予測精度は悪くなると考えられる。

【００１１】本発明の目的は以上の問題点を解決し、従来例に比較して遷移確率の予測精度及び信頼性を改善することができる統計的言語モデルを生成することができる統計的言語モデル生成装置、及び、当該統計的言語モデルを用いて従来例に比較して高い音声認識率で音声認識することができる音声認識装置を提供することにある。

【００１２】

【課題を解決するための手段】本発明に係る請求項１記載の統計的言語モデル生成装置は、所定の話者の発声音声文を書き下した学習用テキストデータに基づいて、すべての語彙を品詞毎にクラスタリングされた品詞クラスに分類し、それらの品詞クラス間のバイグラムを初期状態の統計的言語モデルとして生成する生成手段と、上記生成手段によって生成された初期状態の統計的言語モデルに基づいて、単語の品詞クラスからの分離することができる第１の分離クラス候補と、１つの単語と１つの単語との結合、１つの単語と複数の単語の単語列との結合、複数の単語の単語列と１つの単語との結合、複数の単語の単語列と、複数の単語の単語列との結合とを含む連接単語又は連接単語列の結合によって単語の品詞クラスから分離することができる第２の分離クラス候補とを検索する検索手段と、上記検索手段によって検索された第１と第２の分離クラス候補に対して、次単語の予測の難易度を表わす所定のエントロピーを用いて、クラスを分離することによる当該エントロピーの減少量を計算する計算手段と、上記計算手段によって計算された上記第１と第２の分離クラス候補に対するエントロピーの減少量の中で最大のクラス分離を選択して、選択されたクラスの分離を実行することにより、品詞のバイグラムと可変長Nの単語のN−グラムとを含む統計的言語モデルを生成する分離手段と、上記分離手段によって生成された統計的言語モデルのクラス数が所定のクラス数になるまで、上記分離手段によって生成された統計的言語モデルを処理対象モデルとして、上記検索手段の処理と、上記計算手段の処理と、上記分離手段の処理とを繰り返すことにより、所定のクラス数を有する統計的言語モデルを生成する制御手段とを備えたことを特徴とする。

【００１３】本発明に係る請求項２記載の音声認識装置は、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、品詞のバイグラムと可変長Nの単語のN−グラムとを含む統計的言語モデルを用いて音声認識することを特徴とする。

【００１４】また、請求項３記載の音声認識装置においては、上記統計的言語モデルは、請求項１記載の統計的

言語モデル生成装置によって生成されたことを特徴とする。

【００１５】本発明に係る請求項４記載の連続音声認識装置は、入力される発声音声文の音声信号に基づいて上記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた連続音声認識装置において、上記音声認識手段は、請求項１記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行うことを特徴とする。

【００１６】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。図１に本発明に係る一実施形態の連続音声認識装置のブロック図を示す。本実施形態の連続音声認識装置は、公知のワンーパス・ビタビ復号化法を用いて、入力される発声音声文の音声信号の特徴パラメータに基づいて上記発声音声文の単語仮説を検出し尤度を計算して出力する単語照合部４を備えた連続音声認識装置において、単語照合部４からバッファメモリ５を介して出力される、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、統計的言語モデル２２を参照して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行う単語仮説絞込部６を備えたことを特徴とする。

【００１７】ここで用いる統計的言語モデル２２は、学習用テキストデータに基づいて言語モデル生成部２０により生成されたものであって、統計的言語モデル２２は、品詞クラス間のバイグラム（Ｎ＝２）を基本としたものであるが、単独で信頼できる単語は品詞クラスより分離させ、単独のクラスとして取り扱い、さらに、予測精度を向上させるため、頻出単語列に関してはそれらの単語を結合して一つのクラスとして取り扱い、長い単語連鎖の表現を可能にさせ、こうして、生成されたモデルは、品詞バイグラムと可変長単語Ｎ－グラムとの特徴を併せ持つ統計的言語モデルとなり、遷移確率の精度と信頼性とのバランスをとられたものであることを特徴とする。

【００１８】まず、本実施形態において用いる可変長Ｎ－グラムの概念について以下に説明する。Ｎ－グラムは、（Ｎ－１）重のマルコフモデルであり、これは、過去（Ｎ－１）回の状態遷移を記憶するように単純（１重）マルコフモデルの各状態が分離されたものと解釈される。例として、図３にバイグラムをマルコフモデルとして図式化した状態遷移図を示し、図４にトライグラムをマルコフモデルとして図式化した状態遷移図を示す。

【００１９】図３においては、状態s_1においてシンボルａを出力されたとき状態s_1のままであるが、状態s_1でシンボルｂを出力した状態s_2に遷移する。状態s_2でシンボルｂを出力したときは状態s_2のままであるが、状態s_2でシンボルａを出力したとき状態s_1に戻る。図４のトライグラムは、バイグラムの状態s_1を状態s_{11}と状態s_{12}とに分離しかつ、状態s_2を状態s_{21}と状態s_{22}とに分離したものと考えられる。さらに、全ての状態の分離を進めることにより、より高次のＮ－グラムとなる。

【００２０】図５に示す可変長Ｎ－グラムは、単純マルコフモデルの状態を部分的に分離させたものである。すなわち、図３のバイグラムにおいて、状態s_2から、シンボルａが出力される際に、続けてシンボルｂを出力する場合（これをａｂと表わし、シンボルａｂを出力するという。）、続けてｂ以外のシンボルを出力する場合（これをａ（／ｂ）と表し、シンボルａ（／ｂ）を出力するという。ここで、／は否定の意味を表しバー（上線）である。）とに分け、前者の場合、状態s_1から状態s_{12}に遷移させる一方、後者の場合、状態s_2から状態s_{11}に遷移させる。すなわち、前者の場合において、状態s_1から状態s_{12}へと分離させ、シンボルａを出力する残りの遷移（ａ（／ｂ））を状態s_{11}に残したものである。なお、このモデルにおいて、状態s_{11}でシンボルａｂを出力したとき状態s_{12}に遷移する一方、状態s_{11}でシンボルａ（／ｂ）を出力したとき状態s_{11}のままである。また、状態s_{12}でシンボルａｂを出力したとき状態s_{12}のままである一方、状態s_{12}でシンボルａ（／ｂ）を出力したとき状態s_{11}に遷移する。

【００２１】このモデルは、複数の連続したシンボルを新しいシンボルとみなすことで、単純マルコフモデルの構造のまま、長い連鎖を表すことができるという特徴がある。同様の状態分離を繰り返すことで、局所的にさらに長い連鎖を表すことができる。これが可変長Ｎ－グラムである。すなわち、シンボルを単語とみなした言語モデルとしての可変長単語Ｎ－グラムは、単語列（１単語も含む）間のバイグラムと表される。

【００２２】次いで、可変長Ｎ－グラムの動作について説明する。本実施形態で用いる統計的言語モデル２２は、品詞クラスと単語との可変長Ｎ－グラムであり、次の３種類のクラス間のバイグラムとして表現する。

（１）品詞クラス（以下、第１のクラスという。）、

（２）品詞クラスから分離した単語のクラス（以下、第２のクラスという。）、及び、（３）連接単語が結合してできたクラス（以下、第３のクラスという。）。

【００２３】上記第１のクラスに属する単語は、主として出現頻度の小さいもので、単語単独で取り扱うよりも遷移確率の信頼性が高められる。また、第２のクラスに属する単語は、主として出現頻度が高いもので、単独で

取り扱っても十分な信頼性があり、さらに、連接単語が結合して上記第３のクラスに分類されることにより、可変長Ｎ－グラムとして動作し、次単語の予測精度が高められる。ただし、本実施形態において、連接する品詞クラスと品詞クラス、および、品詞クラスと単語の結合は＊

＊考えない。複数Ｌ個の単語からなる文の生成確率Ｐ（w_1^L）は、次式で与えられる。

【００２４】

【数３】

$$P(w_1^L) = \prod_{t=1}^{K} P(ws_t \mid c_t) \cdot P(c_t \mid c_{t-1})$$

【００２５】ここで、ws_tは文章を上記のクラスに分類した時の、ｔ番目の単語列（単独の単語も含める）を意味する。従って、Ｐ（$ws_t \mid c_t$）は、ｔ番目のクラスがわかったときに単語列ws_tが出現する確率であり、Ｐ（$c_t \mid c_{t-1}$）は１つ前の（ｔ－１）番目のクラスから当該ｔ番目のクラスの単語が出現する確率である。また、文章のＫは単語列の個数を表し、Ｋ≦Ｌである。従って、数３のΠはｔ＝１からＫまでの積である。※

※ここで、例として、次の７単語からなる発声音声文の文章を考える。

【００２６】

【数４】「わたくし－村山－と－言－い－ま－す」

【００２７】この文章の生成確率Ｐ（w_1^L）は、数３を用いて、次の式で与えられる。

【００２８】

【数５】

Ｐ（w_1^L）＝Ｐ（わたくし｜｛わたくし｝）・Ｐ（｛わたくし｝）
・Ｐ（村山｜＜固有名詞＞）・Ｐ（＜固有名詞＞｜｛わたくし｝）
・Ｐ（と｜｛と｝）・Ｐ（｛と｝｜＜固有名詞＞）
・Ｐ（言います｜［言います］）・Ｐ（［言います］｜｛と｝）

【００２９】ただし、＜＞，｛｝，［］はそれぞれ、第１のクラス、第２のクラス、第３のクラスに属していることを表す。ただし、各単語および単語列は次のように属している。

（１）「村山」は固有名詞なので、第１のクラスに属する。

（２）「わたくし」、「と」はそれぞれ、名詞から分離した単語、助詞から分離した単語であり、第２のクラスに属する。

★（３）「言います」は動詞と、動詞の接尾辞と、助動詞と、助動詞の接尾辞との組み合わせであり、第３のクラスに属する。ここで、第２と第３のクラスにおいて、単語とクラスの出現頻度は等しいので、Ｐ（わたくし｜｛わたくし｝）＝１、Ｐ（と｜｛と｝）＝１、Ｐ（言います｜［言います］）＝１であり、従って、上記数５は次の式のようになる。

【００３０】

【数６】

Ｐ（w_1^L）＝Ｐ（わたくし）
・Ｐ（村山｜＜固有名詞＞）・Ｐ（＜固有名詞＞｜わたくし）
・Ｐ（と｜＜固有名詞＞）
・Ｐ（言います｜と）

【００３１】次いで、本実施形態で用いる可変長Ｎ－グラムである統計的言語モデル２２を生成するための言語モデル生成処理について参照して説明する。本実施形態で用いる統計的言語モデル２２は、品詞クラスのバイグラムを初期状態とし、エントロピーの最小化の基準によるクラス分離という形で生成される。エントロピーの減少は正になることが保証されており、クラス分離によって、学習用テキストデータに関してエントロピーは単調に減少する。ここで用いるエントロピーは、一般には、☆

☆「あいまいさ」の尺度を表わすものであり、言語モデルにおいて、エントロピーが小さいことは、言語としてあいまいさが小さく、次の単語の予測が容易であることを意味する。すなわち、エントロピーは次単語の予測の難易度を表わす。ｙという条件のもとでのｘの確率である条件付き確率Ｐ（ｘ｜ｙ）のエントロピーＨ（Ｘ｜Ｙ）は次式で表される。

【００３２】

【数７】

$$H(X \mid Y) = -\Sigma P(y) \Sigma P(x \mid y) \log_2 P(x \mid y)$$

【００３３】従って、上記数７に基づいて、本実施形態で用いるエントロピーは次式で計算される。

【００３４】

【数８】

11　　12

$$H(\{c_i\}) = -\sum_i P(c_i) \cdot \sum_k P(w_k \mid c_j) \cdot P(c_j \mid c_i) \log_2 \{P(w_k \mid c_j) \cdot P(c_j \mid c_i)\}$$

ここで、$w_k \in c_j$

【００３５】図６は、言語モデル生成部２０によって実行される言語モデル生成処理の詳細を示すフローチャートであり、以下、図６を参照して当該処理について説明する。まず、ステップＳ１では、所定の話者の発声音声文を書き下した学習用テキストデータに含まれる全語彙を品詞クラス（ここで、品詞クラスとは、品詞毎にクラスタリングされたクラスをいう。）に分類し、それらのクラス間のバイグラムを初期状態の統計的言語モデルとする。次いで、次のステップＳ２乃至Ｓ４でクラスの分離を行う。すなわち、ステップＳ２で、クラス分離することが可能な分離クラス候補を検索することによりリストアップを行う。ここでは、次の２種類のクラス分離を考える。

（１）単語の品詞クラスからの分離（以下、第１のクラス分離という。）、（２）連接単語又は連接単語列の結合によるクラス分離（以下、第２のクラス分離という。）。ここで、連接単語又は連接単語列の結合とは、連接する（時間的に隣接して入力される）１つの単語と１つの単語との結合、１つの単語と複数の単語の単語列との結合、複数の単語の単語列と１つの単語との結合、複数の単語の単語列と、複数の単語の単語列との結合とを含む。

【００３６】前者の単語の品詞クラスからの分離においては、当初品詞クラスに属している単語が、そのクラスから分離し、分離した単語は、その単語で単独のクラスを形成する。

【００３７】

【数９】$c_\xi \rightarrow \{w_x\} + c_\xi \setminus \{w_x\}$

ここで、$w_x \in c_\xi$

【００３８】ここで、$c_\xi \setminus \{w_x\}$ はクラスc_ξ から単語w_xのクラスを除いたクラスであることを意味し、単語w_xはクラスc_ξ に属している。従って、数９の意＊味するところは、例えば、名詞のクラスc_ξ は、「机」という単語w_xのクラス $\{w_x\}$ と、「机」という単語w_xのクラス $\{w_x\}$ をクラスc_ξ から除いたクラスとに分離することを意味する。

【００３９】後者の連接単語又は連接単語列の結合によるクラス分離においては、既に初期クラスより分離されている単語クラス及び単語列クラスについて、連接した２クラス間の結合を考える。結合した単語列は、その単語列で単独のクラスを形成する。

【００４０】

【数１０】$\{w_x\} \rightarrow \{w_x, w_y\} + \{w_x, /w_y\}$

【００４１】ここで、$\{w_x, w_y\}$ は連接単語列w_x, w_yのクラスを表し、$\{w_x, /w_y\}$ は単語w_xの次に単語w_yが後続しない単語w_xのクラスを表わす。すなわち、$/w_y$は単語w_y以外の単語を表わす。数１０の意味するところは、例えば、「机」という単語のクラス $\{w_x\}$ は、「机の」という単語列のクラス $\{w_x, w_y\}$ と、「机の」以外の例えば「机は」、「机が」などの単語列のクラス $\{w_x, /w_y\}$ とに分離することを意味する。上記数１０は、単語の結合に関する式であるが、単語列と単語の結合、および、単語列と単語列との結合も同様に表される。従って、第２のクラス分離では、これらのクラス分離を含む。

【００４２】次いで、ステップＳ３で、ステップＳ２でリストアップされた上記第１と第２の分離クラス候補に対して次の数１１及び数１２を用いてエントロピー減少量を計算する。ここで、上記第１のクラス分離である初期クラスの分離に対して数１１を用いる一方、上記第２のクラス分離である連接単語又は連接単語列の結合によるクラス分離に対して数１２を用いる。

【００４３】

【数１１】

$$\Delta H = H(\{c_i\}) - H(\{c_i \setminus c_\xi\} + \{w_x\} + \{c_\xi \setminus w_x\})$$

【数１２】

$$\Delta H = H(\{c_i\}) - H(\{c_i \setminus w_x\} + \{w_x, w_y\} + \{w_x, /w_y\})$$

【００４４】ここで、数１１及び数１２において、$H(\{c_i\})$ は元のすべての品詞クラスc_iについてのエントロピーであり、数１１において$H(\{c_i \setminus c_\xi\} + \{w_x\} + \{c_\xi \setminus w_x\})$ は元のすべての品詞クラスc_iから単語w_xのクラスを分離したときのエントロピーであり、数１１のΔＨは単語w_xのクラスを分離したときのエントロピーの減少量である。また、数１２において$H(\{c_i \setminus w_x\} + \{w_x, w_y\} + \{w_x, /w_y\})$ は、元のすべての品詞クラスc_iから単語列 $\{w_x, w_y\}$ のクラスを分離したときのエントロピーであり、数１２のΔＨは単語列 $\{w_x, w_y\}$ のクラスを分離したときのエントロピーの減少量である。

【００４５】次いで、ステップＳ４においては、ステップＳ２でリストアップされたすべての分離クラス候補の中で、ステップＳ３で計算したエントロピー減少量ΔＨを最大にするクラスのみを実際にクラス分離する。そして、ステップＳ５で分離クラス数が所定のしきい値の所望分離クラス数（例えば、５００、１０００など）以上になったか否かを判断し、なっていないときは、ステップＳ２に戻って上記の処理を繰り返す。一方、ステップＳ５で所望分離クラス数以上になっているときは、ステップＳ６で、得られた統計的言語モデル２２をメモリに格納した後、当該言語モデル生成処理を終了する。この言語モデル生成処理のアルゴリズムは、品詞間、および、品詞と単語間の結合は行なわないため、生成完了時点では、品詞のバイグラムと可変長単語のＮ－グラムの特徴を併せた統計的言語モデル２２となる。

【００４６】図１において、単語照合部４に接続され、例えばハードディスクメモリに格納される音素HMM１１は、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

（ａ）状態番号
（ｂ）受理可能なコンテキストクラス
（ｃ）先行状態、及び後続状態のリスト
（ｄ）出力確率密度分布のパラメータ
（ｅ）自己遷移確率及び後続状態への遷移確率

なお、本実施形態において用いる音素HMM１１は、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HMMを変換して生成する。ここで、出力確率密度関数は３４次元の対角共分散行列をもつ混合ガウス分布である。

【００４７】また、単語照合部４に接続され、例えばハードディスクに格納される単語辞書１２は、音素HMM１１の各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【００４８】図１において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して単語照合部４に入力される。

【００４９】単語照合部４は、ワンーパス・ビタビ復号化法を用いて、バッファメモリ３を介して入力される特徴パラメータのデータに基づいて、音素HMM１１と単語辞書１２とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部４は、各時刻の各HMMの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素HMM１１及び単語辞書１２とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部４は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ５を介して単語仮説絞込部６に出力する。

【００５０】単語仮説絞込部６は、単語照合部４からバッファメモリ５を介して出力される単語仮説に基づいて、統計的言語モデル２２を参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果として出力する。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びをいう。

【００５１】例えば、図２に示すように、（ｉ－１）番目の単語W_{i-1}の次に、音素列a_1，a_2，…，a_nからなるｉ番目の単語W_iがくるときに、単語W_{i-1}の単語仮説として６つの仮説Ｗａ，Ｗｂ，Ｗｃ，Ｗｄ，Ｗｅ，Ｗｆが存在している。ここで、前者３つの単語仮説Ｗａ，Ｗｂ，Ｗｃの最終音素は／ｘ／であるとし、後者３つの単語仮説Ｗｄ，Ｗｅ，Ｗｆの最終音素は／ｙ／であるとする。終了時刻t_eと先頭音素環境が等しい仮説（図２では先頭音素環境が“ｘ／a_1／a_2”である上から３つの単語仮説）のうち総尤度が最も高い仮説（例えば、図２において１番上の仮説）以外を削除する。なお、上から４番めの仮説は先頭音素環境が違うため、すなわち、先行する単語仮説の最終音素がｘではなくｙであるので、上から４番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に１つのみ仮説を残す。図２の例では、最終音素／ｘ／に対して１つの仮説を残し、最終音素／ｙ／に対して１つの仮説を残す。

【００５２】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の２つの音素とを含む３つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素と連続する先行する単語仮説の少なくとも１つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

【００５３】以上の実施形態において、特徴抽出部２と、単語照合部４と、単語仮説絞込部６と、言語モデル生成部２０とは、例えば、デジタル電子計算機で構成され、バッファメモリ３，５は例えばハードデイスクメモリで構成され、音素HMM１１と単語辞書１２と学習用

テキストデータ２１と統計的言語モデル２２とは、例えばハードデイスクメモリなどの記憶装置に記憶される。

【００５４】以上実施形態においては、単語照合部４と単語仮説絞込部６とを用いて音声認識を行っているが、本発明はこれに限らず、例えば、音素HMM１１を参照する音素照合部と、例えばOne Pass DPアルゴリズムを用いて統計的言語モデル２２を参照して単語の音声認識を行う音声認識部とで構成してもよい。

【００５５】

【実施例】本発明者は、本実施形態で用いる統計的言語モデル２２の性能を確認するため、パープレキシティおよびパラメータ数について従来の単語N－グラムとの比較を行った。実験に用いたデータは本出願人が所有する自然発話旅行会話データベース（従来文献８「Morimotoほか，“A Speech and Language Database for Speech Translation Research”，ICSLP，pp1791－1794，1994年」参照。）であって、８４６対話、３５４，７００語から構成される。このうち、統計的言語モデル２２を生成するための学習用テキストデータ（トレーニングセットともいう。）として、８２８対話、３４７，２９９語を使用し、残りのデータをテスト用テキストデータ（テストセットともいう。）とした。本実施形態に係る統計的言語モデル２２は、初期クラスを活用形も含めた８０品詞とし、１０００個まで分離を行い、１００個おきにデータを採取した。また、本実施形態に係る統計的言語モデル２２と、単語N－グラムとともに、未知単語遷移に対する対策として、クラスおよび単語の遷移確率を削除補間法（従来文献４参照。）によって補間し、テストセットにおいて、未知語が出現したときは、所定の固定値（例えば、7.0×10^{-6}）を与えた。ここで、本発明に係る統計的言語モデル２２を評価するために、パープレキ＊

＊シティを用いる。例えば、複数n個の単語からなる長い単語列$w_1^n=w_1w_2\cdots w_n$があるときのエントロピーH（n）は次式で表される。

【００５６】

【数１３】

$$H(n)=-(1/n)\cdot\log_2 P(w_1^n)$$

【００５７】ここで、P（w_1^n）は単語列w_1^nの生成確率であり、パープレキシティPP（n）は次式で表される。

【００５８】

【数１４】$PP(n)=2^{H(n)}$

【００５９】ここで、単語列としてテスト用テキストデータを用いたときのパープレキシティをテストセットパープレキシティといい、単語列として学習用テキストデータを用いたときのパープレキシティをトレーニングセットパープレキシティという。

【００６０】当該実験結果におけるテストセットパープレキシティの値の変化の様子を図７に示す。図７から明らかなように、分離クラス数が増加するに従って、テストセットパープレキシティは減少し、分離クラス数が２００で単語バイグラムと、分離クラス数が６００で単語トライグラムと同程度の値となることが分かる。分離クラス数が６００以上になると、パープレキシティの減少の割合が極端に小さくなるため、分離クラス６００程度で、本実施形態の統計的言語モデル２２が最も有効に働いていると考えられる。従って、本実施形態の統計的言語モデル２２は単語バイグラム以上、単語トライグラムと同程度の予測精度の言語モデルと考えられる。

【００６１】表１にまた、分離クラス数が０，５００，１０００の時のパープレキシティの値、およびパラメータ数を示す。

【００６２】

【表１】

各言語モデルの性能比較

	バイグラム	トライグラム	本実施形態（分離クラス数）		
			0	500	1000
テストセットパープレキシティ	20.31	16.96	41.68	17.61	16.75
トレーニングセットパープレキシティ	13.50	5.99	48.77	18.77	15.05
パラメータ数(1)	4.10×10^7	2.62×10^{11}	1.28×10^4	3.43×10^5	1.17×10^6
パラメータ数(2)	７，９９１	５２，２４４	２７，８３０	１６５，１３９	４３，０７５

【００６３】ここで、パラメータ数（１）は全クラス（単語）の遷移の組み合わせ数を意味し、パラメータ数

17

（２）は、トレーニングセットにおいて実際に存在するクラス（単語）遷移の組み合わせ数を意味する。表１より、本実施形態の統計的言語モデル２２は、テストセットとトレーニングセットとのパープレキシティの差が、単語バイグラム及び単語トライグラムと比較して非常に小さいことが分かる。また、パラメータ数は、１０００クラス分離した時でも、単語バイグラムよりも少なく、単語トライグラムよりもはるかに少ない。したがって、本実施形態の統計的言語モデル２２は、与えられたパラメータで言語特徴を効率的に表現できる優れた言語モデルであると言える。従って、当該統計的言語モデル２２は従来の単語バイグラム、単語トライグラムよりも信頼性が高い言語モデルであると考えられる。

【００６４】また、本実施形態の統計的言語モデル２２の信頼性を確認するため、学習単語数を変化させてテストセットパープレキシティの値の変化を調べた結果を図８に示す。この図８から明らかなように、全ての学習セット（約３５万語）を用いたときは、単語バイグラム　＊と、本実施形態の統計的言語モデル２２（２００クラス）（カッコ内の数字は分離クラス数を表す、以下同様である。）とは、ほぼ同じパープレキシティ値であるが、学習単語数を減少させても当該統計的言語モデル２２のパープレキシティの増加は比較的小さく、単語バイグラムよりも値が低くなることが分かる。同様に、単語トライグラムと、当該統計的言語モデル２２（６００クラス）とを比較しても、学習単語数が減少すると、当該統計的言語モデル２２の方が低いパープレキシティを呈する。

18

【００６５】次いで、本発明者は、本実施形態の統計的言語モデル２２を図１の連続音声認識装置に適用し、統計的言語モデル２２の効果を確認した。音素認識の実験条件を表２に示す。また、音響をパラメータもあわせて表２に示す。

【００６６】

【表２】

実験条件

分析条件	サンプリング周波数：１２ＫＨｚ， ハミング窓：２０ｍｓ， フレーム周期：１０ｍｓ
使用パラメータ	１６次ＬＰＣケプストラム＋１６次Δケプストラム ＋ｌｏｇパワー＋Δｌｏｇパワー
音響モデル	HM網の男女別不特定話者モデル ４００状態，５混合

【００６７】表２において、HM網の男女別不特定話者モデルについては、従来文献９「小坂ほか，“話者混合ＳＳＳによる不特定話者音声認識”，日本音響学会講演論文集，２－５－９，ｐｐ１３５－１３６，平成４年」に開示されている。この実験では、単語グラフを用いた連続音声認識法を用いて音響モデルおよび言語モデルを連続音声認識装置に適用した。また、認識の対象は、統※計的言語モデル２２のトレーニングセット中の１６対話であり、学習に用いられていないテストセットは１８対話である。各言語モデルで尤度１位の文認識候補の正解単語含有率を表３に示す。

【００６８】

【表３】

正解単語含有率

		バイグラム	本実施形態（分離クラス数） ０	 ５００
辞書サブセット	テストセット	７１．４	６７．３	７２．２
	トレーニング゛セット	６９．４	６３．４	６９．７
辞書フルセット	テストセット	－－	５７．１	５８．４
	ﾄﾚｰﾆﾝｸﾞ゛ｾｯﾄ	－－	５４．６	５６．０

【００６９】表３において、辞書サブセットは認識対象　50　に含まれる単語のみを辞書に登録したもの（７５０

語）、辞書フルセットは、統計的言語モデルの生成のための学習に用いた全単語よりなる辞書（６，４００語）を表す。ただし、従来の単語バイグラムは、メモリ容量と計算時間の都合上で、辞書フルセットの辞書の認識は、今回の実験では計算を行なっていない。この場合は、言い換えれば、大容量のメモリと莫大な処理時間が必要である。

【００７０】テストセットに関しては、パープレキシティの低い順、すなわち本実施形態の統計的言語モデル２２（０クラス）→単語バイグラム→本実施形態の統計的言語モデル２２（５００クラス）の順で正解単語含有率が良くなっており、本実施形態の統計的言語モデル２２（５００クラス）は、単語のバイグラムよりも若干ではあるが正解単語含有率が向上している。トレーニングセットに関しては、本実施形態の統計的言語モデル２２（５００クラス）は単語バイグラムよりも高いパープレキシティであるが、正解単語含有率は高くなっている。また、本実施形態の統計的言語モデル２２はパラメータ数が少ないので、大語彙の認識への拡張が容易ある。したがって、本実施形態の統計的言語モデル２２は連続音声認識に対しても単語バイグラムより有効な言語モデルであると考えられる。

【００７１】以上説明したように、Ｎ－グラムの精度・信頼性の向上を目的とした可変長Ｎ－グラムの統計的言語モデル２２の生成装置及びこれを用いた連続音声認識装置を実現することができる。当該統計的言語モデル２２は、品詞バイグラムを初期状態とし、品詞クラスからの単語分離、および、連接単語の結合という、２種類の状態分離を行なうことにより生成されるもので、品詞バイグラムと可変長単語Ｎ－グラムの特徴を併せ持つモデルである。当該統計的言語モデル２２の評価実験の結果、当該統計的言語モデル２２は、単語バイグラム以上、単語トライグラムと同等のパープレキシティを、はるかに少ないパラメータで実現できることが分かり、目的とした性能が実現されていることが確認できた。また、連続音声認識に適用した結果、単語バイグラムと同じ程度の正解単語含有率を得ることができた。当該統計的言語モデル２２は少ないパラメータで実現できるため、大語彙の音声認識にも容易に拡張可能である。

【００７２】従って、遷移確率の予測精度及び信頼性を改善することができる統計的言語モデル２２を生成することができる統計的言語モデル生成装置を提供することができるとともに、当該統計的言語モデル２２を用いてより高い音声認識率で連続的に音声認識することができる連続音声認識装置を提供することができる。

【００７３】

【発明の効果】以上詳述したように本発明に係る請求項１記載の統計的言語モデル生成装置によれば、所定の話者の発声音声文を書き下した学習用テキストデータに基づいて、すべての語彙を品詞毎にクラスタリングされた品詞クラスに分類し、それらの品詞クラス間のバイグラムを初期状態の統計的言語モデルとして生成する生成手段と、上記生成手段によって生成された初期状態の統計的言語モデルに基づいて、単語の品詞クラスからの分離することができる第１の分離クラス候補と、１つの単語と１つの単語との結合、１つの単語と複数の単語の単語列との結合、複数の単語の単語列と１つの単語との結合、複数の単語の単語列と、複数の単語の単語列との結合とを含む連接単語又は連接単語列の結合によって単語の品詞クラスから分離することができる第２の分離クラス候補とを検索する検索手段と、上記検索手段によって検索された第１と第２の分離クラス候補に対して、次単語の予測の難易度を表わす所定のエントロピーを用いて、クラスを分離することによる当該エントロピーの減少量を計算する計算手段と、上記計算手段によって計算された上記第１と第２の分離クラス候補に対するエントロピーの減少量の中で最大のクラス分離を選択して、選択されたクラスの分離を実行することにより、品詞のバイグラムと可変長Ｎの単語のＮ－グラムとを含む統計的言語モデルを生成する分離手段と、上記分離手段によって生成された統計的言語モデルのクラス数が所定のクラス数になるまで、上記分離手段によって生成された統計的言語モデルを処理対象モデルとして、上記検索手段の処理と、上記計算手段の処理と、上記分離手段の処理とを繰り返すことにより、所定のクラス数を有する統計的言語モデルを生成する制御手段とを備える。従って、遷移確率の予測精度及び信頼性を改善することができる統計的言語モデルを生成することができる。また、当該統計的言語モデルは少ないパラメータで実現できるため、大語彙の音声認識にも容易に拡張可能であるという特有の利点を有する。

【００７４】本発明に係る請求項２記載の音声認識装置においては、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、品詞のバイグラムと可変長Ｎの単語のＮ－グラムとを含む統計的言語モデルを用いて音声認識する。従って、遷移確率の予測精度及び信頼性を改善することができる統計的言語モデルを用いて音声認識するので、より高い音声認識率で音声認識することができる音声認識装置を提供することができる。

【００７５】また、請求項３記載の音声認識装置においては、上記統計的言語モデルは、請求項１記載の統計的言語モデル生成装置によって生成された。従って、遷移確率の予測精度及び信頼性を改善することができる統計的言語モデルを用いて音声認識するので、より高い音声認識率で音声認識することができる音声認識装置を提供することができる。

【００７６】本発明に係る請求項４記載の連続音声認識装置は、入力される発声音声文の音声信号に基づいて上

記発声音声文の単語仮説を検出し尤度を計算することにより、連続的に音声認識する音声認識手段を備えた連続音声認識装置において、上記音声認識手段は、請求項１記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する１つの単語仮説で代表させるように単語仮説の絞り込みを行う。すなわち、先行単語毎に１つの単語仮説で代表させる従来技術の単語ペア近似法に比較して、単語の先頭音素の先行音素（つまり、先行単語の最終音素）が等しいものをひとまとめに扱うために、単語仮説数を削減することができ、近似効果は大きい。特に、語彙数が増加した場合において削減効果が大きい。従って、当該連続音声認識装置を、間投詞の挿入や、言い淀み、言い直しが頻繁に生じる自然発話の認識に用いた場合であっても、単語仮説の併合又は分割に要する計算コストは従来例に比較して小さくなる。すなわち、音声認識のために必要な処理量が小さくなり、それ故、音声認識のための記憶装置において必要な記憶容量は小さくなる一方、処理量が小さくなるので音声認識のための処理時間を短縮することができる。さらに、遷移確率の予測精度及び信頼性を改善することができる統計的言語モデルを用いて音声認識するので、より高い音声認識率で連続的に音声認識することができる連続音声認識装置を提供することができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施形態である連続音声認識＊装置のブロック図である。

【図２】　図１の連続音声認識装置における単語仮説絞込部６の処理を示すタイミングチャートである。

【図３】　バイグラムの統計的言語モデルを示す状態遷移図である。

【図４】　トライグラムの統計的言語モデルを示す状態遷移図である。

【図５】　図１の連続音声認識装置において用いる可変長Ｎ－グラムの下のモデルを示す状態遷移図である。

【図６】　図１の言語モデル生成部２０によって実行される言語モデル生成処理を示すフローチャートである。

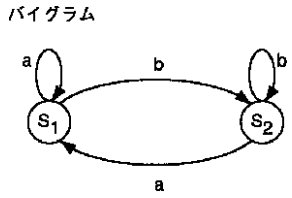
【図７】　図１の言語モデル生成部２０によって生成される統計的言語モデルにおける分離クラス数に対するテストセットパープレキシティを示すグラフである。

【図８】　図１の言語モデル生成部２０によって生成される統計的言語モデルにおける学習データの単語数に対するテストセットパープレキシティを示すグラフである。

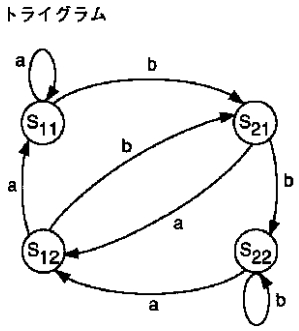
【符号の説明】

１…マイクロホン、
２…特徴抽出部、
３，５…バッファメモリ、
４…単語照合部、
６…単語仮説絞込部、
１１…音素HMM、
１２…単語辞書、
２０…言語モデル生成部、
２１…学習用テキストデータ、
２２…統計的言語モデル。

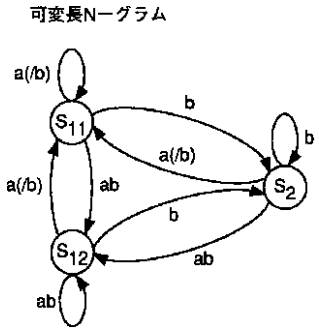
【図３】



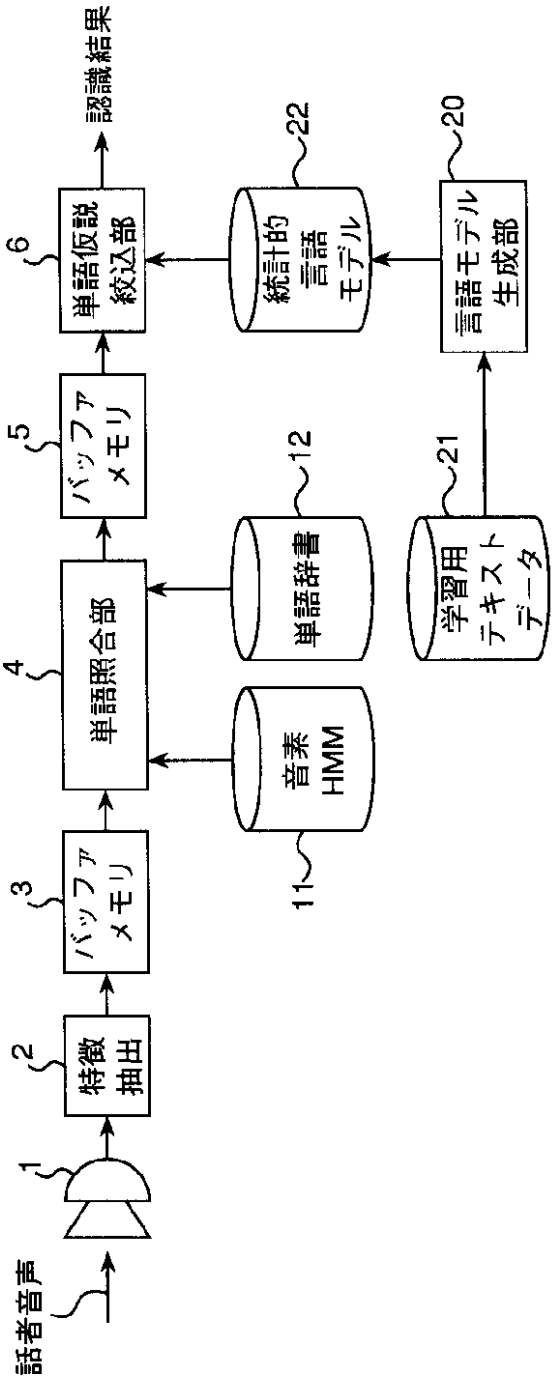
【図４】



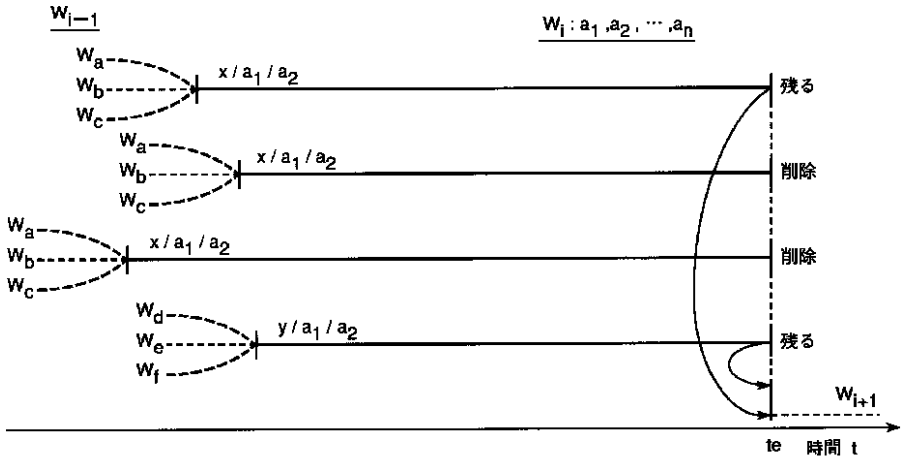
【図５】



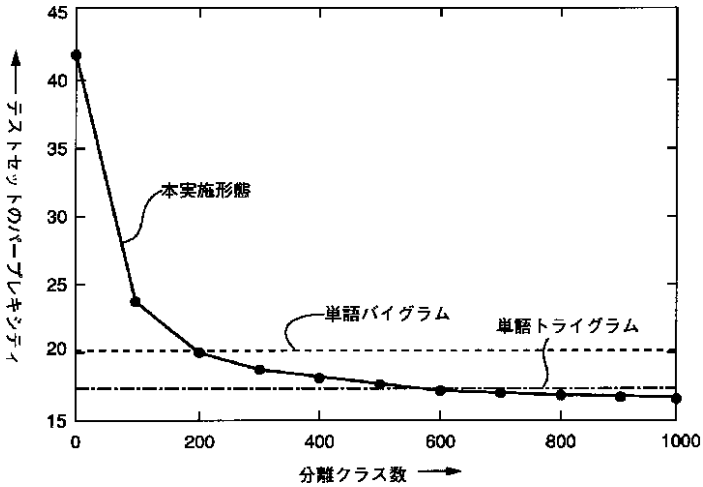
【図１】


話者音声
1
2
特徴抽出
3
バッファメモリ
4
単語照合部
5
バッファメモリ
6
単語仮説絞込部
認識結果
11
音素HMM
12
単語辞書
20
言語モデル生成部
21
学習用テキストデータ
22
統計的言語モデル

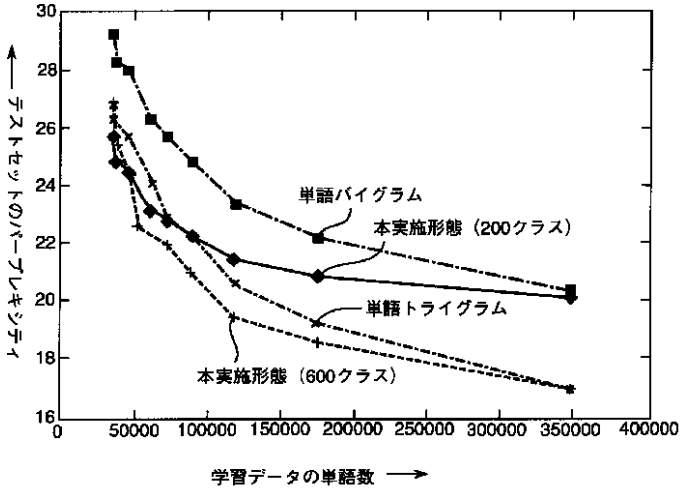
【図２】


W_{i-1}
$W_i : a_1, a_2, \cdots, a_n$
W_a
W_b
W_c
$x / a_1 / a_2$
残る
W_a
W_b
W_c
$x / a_1 / a_2$
削除
W_a
W_b
W_c
$x / a_1 / a_2$
削除
W_d
W_e
W_f
$y / a_1 / a_2$
残る
W_{i+1}
te
時間 t

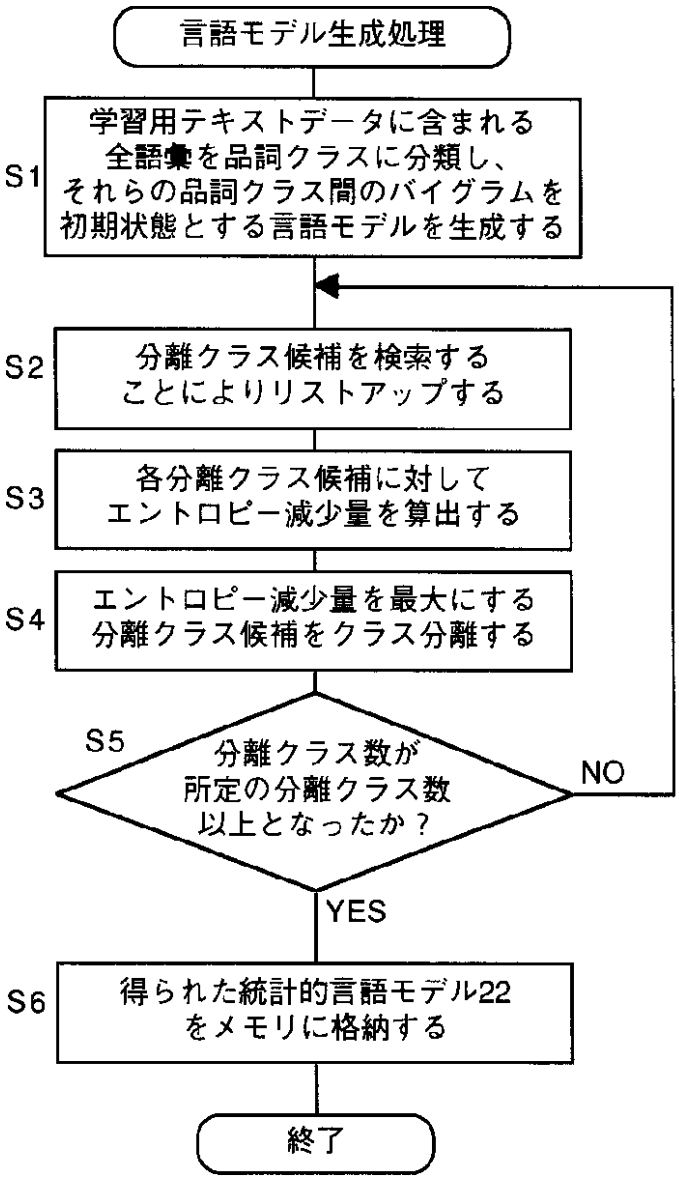
【図７】


テストセットのパープレキシティ
45
40
35
30
25
20
15
本実施形態
単語バイグラム
単語トライグラム
0
200
400
600
800
1000
分離クラス数

【図８】


テストセットのパープレキシティ
30
28
26
24
22
20
18
16
単語バイグラム
本実施形態（200クラス）
単語トライグラム
本実施形態（600クラス）
0
50000
100000
150000
200000
250000
300000
350000
400000
学習データの単語数

【図６】



フロントページの続き

(72)発明者　　松永　昭一
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(56)参考文献　　特開　平５－108704（ＪＰ，Ａ）
特開　平５－250405（ＪＰ，Ａ）
日本音響学会講演論文集（平成８年３月）１－Ｐ－17，ｐ．195～196

(58)調査した分野(Int.Cl.[6]，ＤＢ名)
G10L　3/00　535
G10L　3/00　561
ＪＩＣＳＴファイル（ＪＯＩＳ）