### (19) 日本国特許庁(JP)

# (12) 特許公報(B2)

(11)特許番号

## 特許第4004368号 (P4004368)

(45) 発行日 平成19年11月7日(2007.11.7)

- (24) 登録日 平成19年8月31日 (2007.8.31)
- (51) Int.Cl. F I G 1 OL 15/14 (2006.01) G 1 OL 15/14 2 0 0 C

請求項の数 2 (全 17 頁)

(21) 出願番号	特願2002-277541 (P2002-277541)	(73)特許権者	<b>š</b> 393031586		
(22) 出願日	平成14年9月24日 (2002.9.24)		株式会社国際電	電気通信基础	楚技術研究所
(65) 公開番号	特開2004-117530 (P2004-117530A)		京都府相楽郡料	青華町光台二	二丁目2番地2
(43) 公開日	平成16年4月15日 (2004.4.15)	(74) 代理人	100064746		
審査請求日	平成16年6月17日 (2004.6.17)		弁理士 深見	久郎	
		(74) 代理人	100085132		
			弁理士 森田	俊雄	
		(74)代理人	100083703		
			弁理士 仲村	義平	
		(74)代理人	100096781		
			弁理士 堀井	豊	
		(74) 代理人	100098316		
			弁理士 野田	久登	
		(74)代理人	100109162		
			弁理士 酒井	將行	
					最終頁に続く

(54) 【発明の名称】 音声認識システム

(57)【特許請求の範囲】

【請求項1】

音声認識システムであって、

発話に対応する音声に基づいて、前記音声を測定するフレームごとに特徴情報を<u>観測ベ</u>クトルとして抽出するための特徴抽出手段と、

予め学習された複数の隠れマルコフモデルを格納するための記憶手段とを備え、 前記隠れマルコフモデルの各々は、

隠れ状態に依存する混合ガウス分布により<u>前記</u>観測ベクトルの出力確率を与える第1の 出力確率導出手段と、

前記隠れ状態に独立し、ガウス分布に対応する第1の潜在ベクトルを生成するための潜 10 在ベクトル生成手段と、

前記隠れ状態<u>に依存する第1の状態依存負荷マトリックス</u>と前記第1の潜在ベクトルと <u>の積と第1の混合ガウス分布で表現されるノイズとの和</u>として第2の潜在ベクトルを導出 し、前記第2の潜在ベクトルと<u>第2の状態依存負荷マトリックスの積と第2の混合ガウス</u> 分布で表現されるノイズとの和に応じて前記観測ベクトルの出力確率を与える第2の出力 確率導出手段とを含み、

前記特徴抽出部の出力を受けて、前記隠れマルコフモデルに基づく音響モデルとのマッ チングをとりつつ探索を行うことで音声認識を行う探索手段とを備える、音声認識システム。

【請求項2】

前記第1の潜在ベクトルの次元は、前記第2の潜在ベクトルの次元よりも小さい、請求 項1記載の音声認識システム。 【発明の詳細な説明】 [0001]【発明の属する技術分野】 本発明は、音声認識を行なうための音声認識システムの構成に関する。 [0002]【従来の技術】 近年の音声認識技術は、統計的手法の導入と、大規模なデータベースの構築により、不特 定話者連続音声認識においても、飛躍的に認識率が向上している。 10 [0003]このような音声認識システムへの統計的手法の導入においては、いわゆる、「隠れマルコ フモデル(以下、HMMと呼ぶ)」が採用されている。すなわち、HMMとは、一般的な マルコフモデルの確率的な自由度をより拡大したモデルといえる。 [0004]このHMMでは、状態(内部状態)と出力シンボルの2過程を考え、状態が確率的に遷移 するとともに、それに応じてシンボルを確率的に出力すると考える。そのとき、外部から は状態の遷移は直接的には観測できず、出力シンボルのみが観測可能である。この意味で 、このようなモデルが「HMM( "隠れ " マルコフモデル)」と呼ばれる。 [0005]20 図5は、このようなHMMを図解的に示す概念図である。 図5においては、状態数はn個であり、各状態間の遷移に対応して遷移確率aijが定義 され、かつ各遷移にはシンボル出力確率が対応している。図5においては、遷移確率 a 1 1の状態1から状態1への自己遷移に対応する出力確率のみを例示的に示している。  $\begin{bmatrix} 0 & 0 & 0 & 6 \end{bmatrix}$ ここでHMMは、次の諸量を規定することによって定義される。 1) 状態の集合: S = { s<sub>i</sub> } 2) 出力シンボルの集合: Y = { y ; } 3) 初期状態とその正規確率 4) 状態遷移確率の集合A = { a<sub>ii</sub> } 30 状態siから状態siに遷移する確率。マトリックス[aii]で表わす。 [0007]5) 各状態におけるシンボル出力確率の集合B = { b<sub>i</sub> ( y ) } 状態s,から状態s,に遷移して、シンボルyを出力する確率である。 [0008]6) 終端状態の集合Q = { q <sub>i</sub> } 以下、HMMによる音声認識についてさらに詳しく説明する。 [0009](1) 音声情報の離散シンボル化 マルコフモデルを利用するためには、観測シンボルに対して、音声情報を(有限個の)シ 40 ンボル集合からのシンボル系列で表現しなければならない。 [0010]しかも、そのシンボルの総数があまり多くては処理が複雑化して実用的ではなくなる。こ のような音声情報を離散化する1つの方法としては、いわゆるケプストラム分析が用いら れる。このケプストラム分析においては、まず波形を所定の時間間隔のフレームで切出し た後、各フレーム毎にフーリエ変換および所定の演算処理を行なうことで、音声情報を有 限個のケプストラム係数等の諸係数の集合として表現する。さらに、このようなケプスト ラム係数等に対して、ベクトル量子化が行なわれる。 [0011]

たとえば、このようなケプストラム分析によって、音声情報は、13次元のケプストラム 50

係数と13次元の ケプストラム係数と13次元の ケプストラム係数との合計39次 元の係数によって表現される。 [0012] 図5において、出力確率の横軸は、このような39次元の係数のうちの1つを表わし、縦 軸は、そのような係数の値が出力される連続的な確率を表現している。 [0013]図5に示した例では、たとえば、2つのピークを有する連続的な確率分布となっている。 [0014](2) 認識の原理 HMMを用いた音声認識では、以下のような手続が行なわれる。 10 [0015]a) 認識すべき各音声(たとえば単語)に対応して、HMMが作られているとする。 [0016]入力未知音声を分析し、音声シンボル系列を得る。 b) 観測された音声シンボル系列が、各単語モデルに従って得られる事後確率を計算す c ) る。 [0017]d) 事後確率を相互に比較して、最大の事後確率を与えるモデルに対応する単語を認識 の結果とする。 [0018]20 したがって、音声情報をどのような離散化された係数で表現するかは、HMMを用いた音 声認識の計算量に直結する。 [0019]図6は、上述したようなケプストラム係数等のうち、第1のケプストラム係数C1および 第2のケプストラム係数C2との相関関係を示す図である。 [0020]図6において、点線で示す分布が母集団分布であるものとする。このような母集団分布を 近似するためのモデル分布を実線で示す。 [0021]上述したような39次元の係数のすべての組合せについて、モデル分布を表現しようとす 30 る場合、各係数相互間でのすべての共分散のパラメータを用いることとすると、たとえば 3 9<sup>2</sup>個のパラメータが必要となる。 [0022]これだけのすべての共分散を用いてモデルを構築することとすると、上述したような、音 声認識を行なう際の計算量が膨大となってしまう。 [0023]そこで、一般には、このような39個の係数要素に対して、対角共分散のみを考慮し、複 数のガウス分布の混合分布としてシンボル出力確率を近似することが行なわれる。 [0024]図7は、このようにして、混合分布を採用した場合のモデル分布の構成を示す図である。 40 [0025]なお図7においては、4つのガウス分布によって、母集団分布が近似されている。より一 般的には、N個(N:自然数)のガウス分布によって、母集団分布が近似されているもの とする。 [0026]この場合、各ガウス分布を表現するためには、その各々の平均と分散との2つのパラメー タが必要となるため、39個の要素に対して、必要となるパラメータの個数は、たとえば 、(39×2×N)個となる。これにより、前述したような対角も非対角もすべての共分 散を用いる場合の39²個のパラメータを用いる場合に比べて、著しくパラメータ数を低

減することが可能となる。

(4) JP 4004368 B2 2007.11.7 [0027]図8は、実測されるケプストラム係数間の共分散を示す図である。 図8において、X軸およびY軸はケプストラムの次元を表わし、Z軸は、共分散の値を示 す。したがって、X軸とY軸の対角方向は、自己相関の大きさを示している。上述したよ うな各出力確率を複数のガウス分布の混合分布として近似する場合には、共分散のうち、 図8における対角方向の成分のみを考慮していることに相当する。 [0028]図8を参照すればわかるように、確かに対角方向の共分散の値が大きな値を有するものの 、非対角な共分散の成分も実際には存在しており、より正確な音声認識を行なうためには 、このような非対角の共分散成分も考慮したモデル化が必要であることがわかる。 [0029]しかしながら、単純に共分散の対角成分も考慮することとしてしまうと、上述したとおり 、計算量が増大してしまい、実用的でない。 [0030]一方で、近年、観測ベクトルに対する連続値を有する潜在的な表現が、パターン認識のた めに役立つことが見出されている。なぜならば、それらは観測ベクトル間の相関をコンパ クトに表現することができるからである。 [0031]たとえば、このような潜在的表現は、プリンシプルコンポーネント分析(PCA)(たと えば、非特許文献1を参照)および要因分析(FA)(たとえば、非特許文献2)によっ て実行されている。これらの潜在的表現は、イメージプロセッシングや、特に音声認識( たとえば、非特許文献 3 )における因子分析HMM(Facter Analysis HMM:FA-HMM)法などさまざまな応用が見出されている。 [0032]【非特許文献1】 I. T. Jolliffe, Principle Component Analysis, Springer-Verlag, 1986. [0033]【非特許文献2】 D. Rubin and D. Thayer, "EM algorithms for ML factor analysis," Psychometrika, vol. 47, no.1, pp. 69-76, 1982. [0034]【非特許文献3】 L. K. Saul and M. G. Rahim, "Maximum likelihood and minimum classification erro r factor analysis for automatic speech recognition," IEEE Trans. on SAP, vol. 8 , no.2, pp. 115-125, March 2000. [0035]【発明が解決しようとする課題】 上述したような従来の潜在表現の方法においては、導入された潜在的な空間におけるベク トルは、その空間をXで表わすとすると、ガウス分布N(・;0,I)分布するものとし ている。すなわち、平均値0であって、単位共分散を有するガウス分布であるものとされ ている。 [0036]しかしながら、このような潜在表現(潜在ベクトル)を音声認識に用いた場合に、シンボ ル出現確率を表現するための確率分布のパラメータ数を抑制しつつ、認識精度を向上させ る構成をいかにするべきかについては、必ずしもあきらかでない。 [0037]本発明は、上記のような上記のような問題点を解決するためになされたものであって、そ の目的は、モデルを表現するためのパラメータ数を抑制しつつ、認識精度を向上させるこ とが可能な音声認識システムを提供することである。

[0038]

10

20

30

【課題を解決するための手段】

上記のような課題を解決するために本発明に係る音声認識システムは、発話に対応する 音声に基づいて、音声を測定するフレームごとに特徴情報を<u>観測ベクトルとして</u>抽出する ための特徴抽出手段と、予め学習された複数の隠れマルコフモデルを格納するための記憶 手段とを備え、隠れマルコフモデルの各々は、隠れ状態に依存する混合ガウス分布により 観測ベクトルの出力確率を与える第1の出力確率導出手段と、隠れ状態に独立し、ガウス <u>分布に対応する</u>第1の潜在ベクトルを生成するための潜在ベクトル生成手段と、隠れ状態 <u>に依存する第1の状態依存負荷マトリックス</u>と第1の潜在ベクトルと<u>の積と第1の混合ガ</u> <u>ウス分布で表現されるノイズとの和</u>として第2の潜在ベクトルを導出し、第2の潜在ベク トルと<u>第2の状態依存負荷マトリックスの積と第2の混合ガウス分布で表現されるノイズ</u> <u>との和</u>に応じて観測ベクトルの出力確率を与える第2の出力確率導出手段とを含み、特徴 抽出部の出力を受けて、隠れマルコフモデルに基づく音響モデルとのマッチングをとりつ つ探索を行うことで音声認識を行う探索手段とを備える。

(5)

【 0 0 3 9 】

好ましくは、第1の潜在ベクトルの次元は、第2の潜在ベクトルの次元よりも小さい。

【0040】

【発明の実施の形態】

[システムの構成]

図1は、本発明の実施の形態1の音声認識システム100の構成を示す概略ブロック図で ある。

【0041】

図1を参照して、音声認識システム100は、話者(システム利用者)2からの音声を受け取るためのマイクロフォン102と、マイクロフォン102からの音声入力を受けて音 声認識を行うための認識処理部110と、認識処理部110からの認識結果を表示するた めの表示出力部(たとえば、ディスプレイ)120とを備える。

【0042】

認識処理部110は、認識処理部110と外部とのデータの授受を制御するためのデータ 入出力部70と、音声入力を受けて、デジタル信号に変換するためのA/D変換器10と 、A/D変換器10の出力を受けて、フィルタリング等の信号処理を行なうための信号処 理部20と、信号処理部20の出力を受けて、ケプストラム係数等の特徴ベクトルを抽出 するための特徴ベクトル抽出部30と、予め学習等により生成されるHMMによる音韻モ デルデータを格納しておくための音響モデルデータベース40と、認識するべき単語の事 前確率に対応する統計データを格納している言語モデルデータベース50と、特徴ベクト ル抽出部30からの出力を受けて、音響モデルデータベース40中のデータおよび言語モ デルデータベース50中のデータとに基づいて、入力された特徴ベクトルに対応する認識 すべき各音声(たとえば単語)を探索するための探索部60とを備える。探索部60は、 データ入出力部70を介して、音声認識の結果を表示出力部120出力する。

【0043】

なお、以下に説明するような音声認識システム100の処理は、特徴ベクトル抽出部30 や探索部60が行う処理を記述するためのプログラムと音響モデルデータベース40と言 40 語モデルデータベース50とに保持されるべき初期データとを格納した記録媒体から、音 声入力機能を有するコンピュータにこれらのプログラムおよびデータをインストールする ことによっても実現可能である。このような記録媒体としては、特に限定されないが、た とえば、CD-ROMやDVD-ROM等を用いることができる。あるいは、これらのプ ログラムやデータは、インターネット等の通信回線を介して、音声入力機能を有するコン ピュータにダウンロードすることも可能である。

【0044】

「音響モデルデータベース40中の音響モデル]

以下、音響モデルデータベース40に格納される本発明の音響モデルデータについてさら に詳しく説明する。 10

20

【0045】

[一般的な因子分析 H M M ]

以下に説明するように、本発明におけるアプローチの重要な点は、観測ベクトルy, R<sup>N</sup> についての音響単位(音素)従属な潜在表現ベクトル×, R<sup>L</sup>を導入することと、さらに 、潜在ベクトル×,についての音響単位独立な潜在表現ベクトルz, R<sup>K</sup>を導入すること である。潜在表現ベクトルz,は、音響単位独立であるために、音響単位従属負荷マトリ ックスによって、音響単位従属な潜在ベクトル×,を生成するためのソースと考えること ができる。以下このような潜在ベクトルを2段階に用いて、HMMを表現する本発明のモ デルを「一般的な因子分析HMM」と呼ぶ。

(6)

[0046]

図2は、一般化された因子分析HMMの図形的なモデルを示す図である。

図 2 において、丸い円と四角とは、それぞれ連続数値および離散数値のノードを示している。ハッチングされたノードは、観測シンボルy<sub>t</sub>(時刻 t) およびy<sub>t+1</sub>(時刻 t + 1) を示している。また、qtは、HMMにおける内部状態 { 1,..., S } のうち、時刻 t に おける離散的な状態を示している。Q(t) = (q1,..., qt, qt + 1,..., q<sub>T</sub>) は、状態 p から状態 q への一次の状態遷移確率 a<sub>pq</sub>による離散状態系列を示している。こ の状態系列は、音声における意味論的な系列を示すものである。

【0047】

2つの連続値変数×<sub>t</sub>およびy<sub>t</sub>は、離散的な状態系列に従属するものであるのに対し、連続変数z<sub>t</sub>は、離散的な状態系列とは独立なものである。

【0048】

状態 q t において、混合ガウス分布のうちm t で表わされるガウス分布(平均 μ<sub>qm</sub>,対角 分散 <sub>qm</sub>)で表わされる分布確率と後に説明するようにベクトル x<sub>t</sub>および q に依存し て、観測ベクトル y<sub>t</sub>が出力される。

【0049】

連続値のノードである y<sub>t</sub>、 x<sub>t</sub>および z<sub>t</sub>は、階層的な構造を有している。最も高い階層 においては、ベクトル x<sub>t</sub>は、状態 q における次元 L × K の状態依存負荷マトリックス C<sub>q</sub> によって、以下に示す式(1)~(2)にしたがって、因子分析によりベクトル z<sub>t</sub>から 生成される。

【0050】

【数1】

 $z_t \sim p(z_t) = N(z_t; 0, I)$ 

 $x_t = C_0 z_t + \zeta_{ot}$ 

【0051】

ここで、ベクトル <sub>qt</sub>は、潜在空間 X におけるノイズを表わしている。添え字 q は状態を 示し、添え字 t は時刻を示す。このノイズは、混合ガウス分布によってモデル化される。 40 この混合ガウス分布は以下の式で表わされる。 【0052】

(1)

(2)

【数 2 】

 $\{N(\zeta_{qt};\xi_{qj},V_{qj})\}_{j=1,\cdots,M_{q}^{*}}$ 

【 0 0 5 3 】

ここで、この混合ガウス分布においては、各ガウス分布は係数 c<sub>qj</sub>の重みを有している。 また、このガウス分布において、 V<sub>qj</sub> は対角共分散である。さらに、 M<sub>q</sub><sup>x</sup>は、潜在空間 X 50

10

30

において、状態 q に対する混合状態要素の数を示している。 【0054】 対角共分散 V<sub>qj</sub>における要素は、同一の値を有するようには制限されているわけではない ので、上述した関数は、状態 q における各要素 j { 1,..., M<sub>q</sub><sup>x</sup>} における潜在ベクト ル x<sub>1</sub>についての因子分析である。観測値 y<sub>1</sub>は、以下のモデルによって潜在ベクトル x<sub>1</sub> と関連している。 【0055】

【数3】

$x_t \sim Model^{FA-HMM}$	(3)
$y_t = \Lambda_q x_t + v_{qt}$	(4)

10

【0056】

この式(3)では、潜在ベクトル x<sub>t</sub>は、 FA - HMMのモデルとして表現されることを 示す。ここで、観測ノイズ v<sub>qt</sub>は、以下の混合ガウス分布に従って分布しているものとす る。

【0057】

【数4】

# $\{N(v_{qt};\mu_{qm},\Sigma_{qm})\}_{m=1,\cdots,M_o^y}$

【 0 0 5 8 】

ここで、各ガウス分布は係数 <sub>qm</sub>の重みを有しているものとする。また、 M<sub>q</sub><sup>y</sup>は、 Y 空間 における状態 q の混合要素の数を表わしている。さらに、 <sub>qm</sub>は、要素( n , n )に対す る <sup>2</sup><sub>qmn</sub>を有する対角要素である。

【0059】

<sup>2</sup>qmnの値は、n (1,...,N)に対してすべて同一の値をとるように制限されているわけではない。 。は、次元N×Lの状態依存の負荷マトリックスである。

【0060】

図 2 において見られるように、潜在ベクトル x , から観測値 y , へのリンクおよび <sub>q</sub>のリ 30 ンクが存在しない場合は、このモデルは一般的な H M M となる。

【0061】

図3は、このような一般的なHMMの例を示す図である。

従来の H M M においては、状態 q t において、混合ガウス分布のうちm t で表わされるガ ウス分布 (平均 μ<sub>qm</sub>,分散 <sub>qm</sub>)で表わされる分布確率で、観測ベクトル y t が出力され る。

【 0 0 6 2 】

さらに、図2において、ベクトルz,からy,への確率的なリンクを付加し、かつベクトルx,から観測値y,へのリンクを消去すると、これは、上述した非特許文献3に示される因子分析HMM(FA-HMM)が得られる。

[0063]

図4は、このような従来のFA-HMMの構成を示す図である。

図 2 と図 4 とを比べると、図 4 では、潜在ベクトル x<sub>t</sub>と観測ベクトル y<sub>t</sub>とは同一の次元 である必要があるのに対し、図 2 では、潜在ベクトル z<sub>t</sub>の導入により、より少ない次元 のベクトル z<sub>t</sub>から潜在ベクトル x<sub>t</sub>が生成されるので、より少ないパラメータで、非対角 の共分散の効果を表す潜在ベクトル x<sub>t</sub>を表現することができる。

【0064】

すなわち、上述した関数(1)および(2)は潜在ベクトル×<sub>t</sub>の簡潔な表現である。ベ クトルztは、意味論的には独立である(すなわち音響的に独立である)。状態に依存す るC<sub>a</sub>は、いわゆる「ボーカルトラクトフィルタ(Vocal Tract Filter)」として動作す 20

るので、意味論的に従属したベクトル×→が生成され得る。このようなわけで、上述した ようなモデルを「一般化された因子分析HMM」と呼ぶ。 [0065] [一般化された因子分析HMM(GFA-HMM)における最尤パラメータ評価] 系列Q(T)、X(T)、Ζ(T)、M(T)およびJ(T)は、隠れ状態であるため、 モデルパラメータ の最尤評価を反復的ないわゆるEM(Expectation-Maximization)ア ルゴリズムによって実行することができる。EMアルゴリズムにおいては、以前のモデル パラメータ から計算された隠れ状態系列の事後確率について、現在のモデルパラメータ ハット (の上部に " ^ " がつけられたもの、以下他の変数でも " ^ " がつけられたも のをハットと呼ぶ)で計算された結合ログ尤度の平均として補助関数Qが定義される。 10 [0066]すなわち以下のとおりの定義が行なわれる。 [0067]【数5】  $Q(\Theta, \hat{\Theta}) = E_{\Theta}[\log \prod_{t=1}^{I} p(y_t, x_t, z_t, q_t, m_t, j_t | \hat{\Theta})]$  $= E_{\Theta}[\log \prod_{i=1}^{T} \{\hat{a}_{q_{i-1}q_{i}} \hat{\pi}_{q_{i}m_{i}}\}^{\delta_{qm}(t)}] \\ + E_{\Theta}[\log \prod_{t=1}^{T} \{p(z_{t})\hat{c}_{q_{i}j_{t}}p(x_{t}|z_{t},q_{t},j_{t};\hat{\Theta})\}^{\delta_{qj}(t)}] \\ + E_{\Theta}[\log \prod_{t=1}^{T} \{p(y_{t}|x_{t},q_{t},m_{t};\hat{\Theta})\}^{\delta_{qm}(t)}]$ 20 (5)[0068]つまり、モデルパラメータ の初期値を設定したのち、補助関数Qを最大とする ハット を最尤推定し、モデルパラメータ の値を ハットに設定して、所定の収束条件が満たさ 30 れるまで、補助関数Qによる ハットの最尤推定を繰り返すという手続きをとる。 [0069]ここで、 <sub>ami</sub>(t)は、以前のステップで得られているモデルパラメータ から計算さ れるものである。 [0070]各要素は、それぞれ、{a<sub>qp</sub>ハット, <sub>qm</sub>ハット}、{C<sub>q</sub>ハット,c<sub>qj</sub>ハット, <sub>qj</sub>ハット , V<sub>g</sub> i ハット } および { µ<sub>g</sub> m ハット, <sub>g</sub> n ハット } のそれぞれの関数であるので、パラメー タ評価は、これらに対して各々別々に実施することができる。補助関数を解くことは、そ れらの事後統計が必要となる。 [0071]40 (事後統計の計算) 時刻 t における、状態 q における尤度、空間 Y における混合要素 m、空間 X における混合 要素iの値が以下のとおり与えられる。 [0072] 【数6】

 $p(y_t|q,m,j;\Theta) = N(y_t;\mu_{qm} + \Lambda_q\xi_{qj};\Sigma_{qm} + \Lambda_q(V_{qj} + C_qC_q^T)\Lambda_q^T)$ (6)

【0073】

ここで、上付き添え字のTは転置を意味している。マージナライズをすることによって、  
確率 p(y,|q,m)と確率 p(y,|q)とを得ることができる。  
[0074]  
(0074]  
(0074]  
(0074]  
(0075]  
(0075]  
[数7]  

$$\gamma_{qmj}(t) = \frac{\pi_{qm}c_{qj}p(y_l|q,m,j;\Theta)}{\sum_{q}\sum_{m}\sum_{j}\pi_{qm}c_{qj}p(y_l|q,m,j;\Theta)}$$
(7)  
[0076]  
上述した事後確率をマージナライズすることで、状態 q と混合要素mにおける事後確率  
(1)、状態 q における事後確率 q(t)とを得ることができる。  
[0077]  
連続値の隠れ系列×(T)の事後分布に閉しては、ペイズの規則に従えば、以下のとおり  
与えられる。  
[0078]  
上述した間数の各要素は、ガウス分布であるので、事後分布は、またガウス分布である。  
事後分布り、(x, | y, q, m, j]  
こて、以下の式が成り立つ。  
[0080]  
 $\rho(x_l|y_l,q,m,j;\Theta) = \frac{p(y_l|x_l,q,m;\Theta)p(x_l|q,j;\Theta)}{p(y_l|q,m,j;\Theta)}$ 
(8)  
[0079]  
上述した間数の各要素は、ガウス分布であるので、事後分布は、またガウス分布である。  
事後分布り(x, | y, q, m, j]  
 $\phi_{qmj}^*(t) = E_{\Theta}[x_l|y_l,q,m,j]$   
 $= \psi_{qmj}^*[(V_{sl} + C_q C_q^T)^{-1}\xi_{aj} + \Lambda_q^T \Sigma_{qm}^{-1}(y_l - \mu_{qm})]$ (9)  
 $\psi_{qmj}^x = E_{\Theta}[\delta x_l \delta x_l^T |y_l,q,m,j]$   
 $= [(V_{sj} + C_q C_q^T)^{-1} + \Lambda_q^T \Sigma_{qm}^{-1} \Lambda_q^{-1}$ (10)  
40

は 0 0 8 2 ] さらに、関数 <sup>×</sup>qmj(t)を以下のとおり定義する。 【 0 0 8 3 】 【 数 1 0 】  $\Phi_{qmj}^{x}(t) = E_{\Theta}[x_t x_t^T | y_t, q, m, j]$  $= \Psi_{qmj}^{x} + \phi_{qmj}^{x}(t)\phi_{qmj}^{x}(t)^T$ 

[0084]

式(9)と式(10)を組合せることとすると、事後平均<sup>×</sup><sub>am</sub>(t)と、事後分散<sup>×</sup><sub>am</sub> (t)とは以下のとおり表わされる。 [0085]【数11】

$$\phi_{qm}^{x}(t) = \sum_{j} \gamma_{qmj}(t) \phi_{qmj}^{x}(t)$$

$$\Psi_{qm}^{x}(t) = \sum_{j} \gamma_{qmj} \Psi_{qmj}^{x}(t)$$

[0086]

同様に、 <sup>×</sup><sub>q</sub>(t)と <sup>×</sup><sub>q</sub>(t)も表わすことができる。

ここで、p(z<sub>t</sub>)~N(z<sub>t</sub>;0,I)とp(x<sub>t</sub>|z<sub>t</sub>,q,m,j; )はガウス分布 であるので、潜在ベクトル z,の事後分布もまたガウス分布 N ( z,; <sup>z</sup> qmj(t), <sup>z</sup> qmj(t ))である。

[0087]

このようにして、第1および第2のオーダの統計のみが必要となるだけである。

[0088]

潜在ベクトルェ,に対する事後統計の計算をこのようにして簡略化することができる。な 20 ぜならば、 x t の事後平均ベクトルである <sup>x</sup> q m i (t)は、関数(1)および(2)に対 して変数 x,の観測ベクトルとして得られるからである。

[0089]

このようなわけで同様にして式(9)および(10)において、変数z,の事後統計は、 以下の式で与えられる。

[0090]

【数12】

$$\begin{split} \phi_{qmj}^{z}(t) &= E_{\Theta}[z_{t}|x_{i}, y_{t}, q, m, j] \\ &= \Psi_{qmj}^{z}(t)C_{q}^{T}V_{qj}^{-1}(\phi_{qm}^{z}(t) - \xi_{qj}) \quad (11) \\ \Psi_{qmj}^{z}(t) &= E_{\Theta}[\delta z_{t}\delta z_{t}^{T}|x_{t}, y_{t}, q, m, j] \\ &= [I + C_{q}^{T}V_{qj}^{-1}C_{q}]^{-1} \quad (12) \end{split}$$

[0091]ここでも、関数 <sup>z</sup>qmj(t)を以下のとおり定義する。 [0092]【数13】

$$\Phi_{qmj}^{z}(t) = E_{\Theta}[z_{t}z_{t}^{T}|y_{t}, q, m, j]$$

$$=\Psi^z_{qmj}(t) + \phi^z_{qmj}(t)\phi^z_{qmj}(t)^T$$

[0093]

ここで、モデルパラメータ ハットを更新するためのEMアルゴリズムは、上述した事後 統計の積算を含んでいる。このようなパラメータ評価のためのEMアルゴリズムの手続き は次のとおりである。 [0094]

(パラメータ評価のためのEMアルゴリズム)

50

10

30

---

(11)

再評価のための式は以下の式(13)~(15)として与えられる。 【0095】 【数14】

$$\sum_{t} \gamma_{qm}(t) \Sigma_{qm}^{-1} \hat{\Lambda}_{q} \Phi_{qm}^{x}(t)$$
  
= 
$$\sum_{t} \gamma_{qm}(t) \Sigma_{qm}^{-1}(y_{t} - \mu_{qm}) \phi_{qm}^{x}(t)^{T}$$
(13)

$$\hat{\mu}_{qm} = \frac{1}{\sum_t \gamma_{qm}(t)} \sum_t \gamma_{qm}(t) [y_t - \Lambda_q \phi_{qm}^x(t)]$$
(14)

$$\hat{\Sigma}_{qm} = diag \frac{1}{\sum_{t} \gamma_{qm}(t)} \sum_{t} \gamma_{qm}(t) [(y_t - \mu_{qm})(y_t - \mu_{qm})^T - (\lambda_q \phi_{qm}^x(t)(y_t - \mu_{qm})^T - (y_t - \mu_{qm})\phi_{qm}^x(t)^T \Lambda_q^T + \Lambda_g \Phi_{qm}^x(t) \Lambda_q^T]$$
(15)

【 0 0 9 6 】 負荷行列 <sub>q</sub>ハットは、行ごとに評価される。新たな負荷行列 <sub>q</sub>ハットの第 n 番目の行べ クトル <sub>qn</sub>ハットは、以下の式(1 6)で表現される。 【 0 0 9 7 】 【数 1 5 】

$$\hat{\lambda}_{qn} = \mathbf{k}_{qn}^T G_{qn}^{-1} \tag{16}$$

【0098】 ここで、L×L行列のG<sub>qn</sub>とL次元のベクトルk<sub>qn</sub>は、以下のとおり定義される。 【0099】 【数16】

$$G_{qn} = \sum_{t} \gamma_{qm}(t) \frac{1}{\sigma_{qmn}^2} \Phi_{qm}^x(t)$$
(17)

$$\mathbf{k}_{qn} = \sum_{t} \gamma_{qm}(t) \frac{1}{\sigma_{qmn}^2} (y_{tn} - \mu_{qmn}) \phi_{qm}^x(t)^T$$
(18)

【0100】 ここで、y<sub>tn</sub>とµ<sub>qmn</sub>とは、それぞれ、現在の観測ベクトルと観測ノイズの平均ベクトル のn番目の要素である。 【0101】 空間 X における「観測ベクトル」を <sup>x</sup>qmj(t)とすると{C<sub>q</sub>ハット, <sub>qj</sub>ハット, V<sub>qj</sub>ハッ ト}の再評価式も上記の式と同様にして導き出される。このような再評価式は以下のとお 40 りである。 【0102】

【数17】

30

$$\hat{C}_{q}\left[\sum_{t}\sum_{j}\gamma_{qj}(t)\Phi_{qj}^{z}(t)\right] = \sum_{t}\sum_{j}\left[\phi_{qj}^{x}(t) - \xi_{qj}\right]\phi_{qj}^{z}(t)^{T}$$
(19)

(12)

$$\hat{\xi}_{qj} = \frac{1}{\sum_{t} \gamma_{qj}(t)} \sum_{t} \gamma_{qj}(t) (\phi_{qj}^{x}(t) - C_{q} \phi_{qj}^{z}(t))$$
(20)

$$\hat{V}_{qj} = diag \frac{1}{\sum_{t} \gamma_{qj}(t)} \sum_{t} \gamma_{qj}(t) \{ [\phi_{qj}^{x}(t) - \xi_{qj} - C_{q} \phi_{qj}^{z}(t)]$$

$$[\phi_{qj}^{x}(t) - \xi_{qj} - C_{q} \phi_{qj}^{z}(t)]^{T} + C_{q} \Psi_{qj}^{z}(t) C_{q}^{T} \}$$
(21)

【0103】

式(13)におけるのと同様に、負荷行列C<sub>q</sub>ハットも、行ごとに計算される。混合分布の重みについての最大化の補助関数(5)は、最終的に以下のような更新式(22)~( 23)になる。

[0104]

【数18】

$$\hat{\pi}_{qm} = \frac{\sum_{t} \sum_{j} \gamma_{qmj}(t)}{\sum_{t} \sum_{m} \sum_{j} \gamma_{qmj}(t)}$$

$$\hat{c}_{qj} = \frac{\sum_{t} \sum_{m} \gamma_{qmj}(t)}{\sum_{t} \sum_{m} \sum_{j} \gamma_{qmj}(t)}$$
(22)
(23)

【0105】

以上のような手続きにより、モデルを表現するためのパラメータ数を抑制しつつ、認識精 度を向上させることが可能となる。 30

【 0 1 0 6 】

[実験結果]

(実験条件)

上述したような一般化された因子分析 H M M (G F A - H M M )を、通常の H M M と比較 した結果を以下説明する。

【0107】

このような比較は、オーロラ 2 データベースに対して行なわれた。このようなオーロラ 2 データベースについては、文献: D. Pearce, "Aurora project: Experimental framewor k for the performance evaluation of distributed speech recognition front-ends," in ISCA ITRW ASR2000, Sep. 2000.に開示されている。

【0108】

認識のための特徴は、39次元の、ケプストラム係数およびその第1および第2次の係数 であった。データベースにおける雑音のないトレーニングセット中の1000個の発声が トレーニングのための音響モデルとして用いられた。

【0109】

すべてのシステムにおける音響モデルは6回の反復を有するEMアルゴリズムによって訓 練された。すべての音響モデルにおいて、状態の数は数字に対しては10個であり、無音 状態に対しては3個であった。

【0110】

従来のHMMは、単に、混合要素の数M<sup>y</sup>qの数を適合させることができるのみである。し 50

20

たがって、モデルに対するパラメータの自由度(NoFP)は、S×(2N)×M<sup>ッ</sup> qである。

【0111】

G F A - H M M の構成はよりフレキシブルである。空間 X における混合要素の数、 M<sup>×</sup><sub>q</sub>お よび空間 X の次元 L とを変化させた。空間 Z の次元は一次元としている。また、混合要素 の数 M<sup>y</sup><sub>q</sub>も 1 とされている。潜在的なパラメータ { <sub>q</sub>, <sub>qj</sub>, V<sub>qj</sub>, C<sub>q</sub> } は、各音響モ デルに対する状態間で共有されている。G F A - H M M の単語モデルに対しては、パラメ ータの自由度 N o F P は、S x (2 N) + (n + 1) x L + (2 x L) x M<sup>×</sup><sub>q</sub>である。 【 0 1 1 2 】

(実験結果)

10

20

30

表1は、混合モデルに対するパラメータの自由度と単語の認識精度を従来のHMMのテストによって得られたものである。表1においては、これらの値が空間Yにおける混合要素の数M<sup>У</sup>gの関数として示されている。

【0113】

【表1】

$M_q^y$	1	2	3	4
NoFP	624	1248	1872	2496
W.A.	88.05	85.04	87.96	88.93

【0114】

表 1 では、混合要素の数 M<sup>y</sup> <sub>q</sub>を 4 と設定することにより、最高の認識精度 8 8 . 9 3 %が 得られている。そのような場合においてパラメータの自由度 N o F P は、単語モデルに対 して 2 4 9 6 となっている。

【0115】

一方、表2は、GFA-HMMにおける結果である。

【0116】

【表2】

次元 L	$M_q^x$	1	2	3	4
1	NoFP	666	668	670	672
	W.A.	88.80	89.73	90.30	90.93
2	NoFP	708	712	716	720
	W.A.	86.44	89.09	89.73	89.66

【0117】

GFA - HMMにおいては、同数のトレーニングデータに対して、従来のHMMを越える より高い認識精度が達成されている。たとえば、L = 1を維持したまま、空間Xにおける 混合要素を増やしていくことによって、認識精度は向上していく。最高の単語認識精度は 40 、L = 1およびM<sup>×</sup><sub>q</sub> = 4と設定した場合において90.93%である。さらに、従来のH MMによるものに比べてずっと少ないパラメータの自由度でよい。たとえば、この場合に おいては、パラメータの自由度NoFPは672である。

**[**0 1 1 8 **]** 

今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられ るべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、 特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される

【 0 1 1 9 】

【発明の効果】

(14)

以上説明したとおり、本発明に係る音声認識システムにおいては、モデルを表現するため のパラメータ数を抑制しつつ、認識精度を向上させることが可能である。 【図面の簡単な説明】

【図1】 本発明の実施の形態1の音声認識システム100の構成を示す概略ブロック図 である。

【図2】 一般化された因子分析HMMの図形的なモデルを示す図である。

【図3】 一般的なHMMの例を示す図である。

【図4】 従来のFA-HMMの構成を示す図である。

【図5】 HMMを図解的に示す概念図である。

【図6】 第1のケプストラム係数C1および第2のケプストラム係数C2との相関関係 を示す図である。

混合分布を採用した場合のモデル分布の構成を示す図である。 【図7】

【図8】 実測されるケプストラム係数間の共分散を示す図である。

【符号の説明】

2 話者、10 A/D変換器、20 信号処理部、30 特徴ベクトル抽出部、40 音響モデルデータベース、50 言語モデルデータベース、60 探索部、70 データ 入出力部、100 音声認識システム、102 マイクロフォン、110 認識処理部、 120 表示出力部。

【図1】





10













#### フロントページの続き

- (72)発明者 ヤオ・カイシェン 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 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 党 建武
  - 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

審查官 山下 剛史

(56)参考文献 特開平08-248985(JP,A)
特開平05-080792(JP,A)
特開平10-026996(JP,A)
特開平10-282986(JP,A)
特開平11-085186(JP,A)
船田哲男他,"HMM出力確率の共分散行列に対する制約的条件の実験的考察",電子情報通信
学会論文誌, Vol.J84-D-II, N0.2, 2001年2月, p.429-431

(58)調査した分野(Int.Cl., DB名)

G10L 15/00-15/28