

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4299630号  
(P4299630)

(45) 発行日 平成21年7月22日 (2009.7.22)

(24) 登録日 平成21年4月24日 (2009.4.24)

(51) Int. Cl. F I  
**G 1 0 L 15/14 (2006.01)** G 1 0 L 15/14 2 0 0 Z  
**G 1 0 L 15/06 (2006.01)** G 1 0 L 15/06 3 0 0 Z

請求項の数 7 (全 22 頁)

<p>(21) 出願番号 特願2003-363900 (P2003-363900)</p> <p>(22) 出願日 平成15年10月23日 (2003.10.23)</p> <p>(65) 公開番号 特開2005-128263 (P2005-128263A)</p> <p>(43) 公開日 平成17年5月19日 (2005.5.19)</p> <p>審査請求日 平成17年5月13日 (2005.5.13)</p> <p>(出願人による申告) 国等の委託研究の成果に係る特許出願 (平成15年度通信・放送機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受けるもの)</p>	<p>(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</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 實廣 貴敏 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</p> <p>(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</p> <p>審査官 毛利 太郎</p>
--	--

最終頁に続く

(54) 【発明の名称】 モデル構造作成装置、音声認識装置、及びモデル構造作成プログラム

(57) 【特許請求の範囲】

【請求項1】

隠れマルコフモデルを記憶するための記憶手段と、

前記隠れマルコフモデルに含まれる状態を時間方向と環境方向とに分割してそれぞれ新たな隠れマルコフモデルを準備する処理を、各状態に対して行なうための分割手段と、

前記記憶手段に記憶される隠れマルコフモデルに対して、変分ベイズ法により計算される所定の評価関数の値と、前記分割手段により準備される前記新たな隠れマルコフモデルの各々に対して計算される前記評価関数の値との差分をそれぞれ算出し、前記差分が所定の選択条件を充足する分割を選択して前記記憶される隠れマルコフモデルを更新するための分割選択手段と、

所定の終了条件が成立するまで、前記記憶手段に記憶された隠れマルコフモデルに対し、前記分割手段による準備、及び前記分割選択手段による選択を繰返すように前記分割手段及び前記分割選択手段を制御するための制御手段とを含み、

前記評価関数は、変分ベイズ法により規定される自由エネルギーである、モデル構造作成装置。

【請求項2】

前記制御手段は、前記分割選択手段によって算出された差分の最大値が負となるまで、前記記憶手段に記憶された隠れマルコフモデルに対し、前記分割手段による準備及び前記分割選択手段による選択を繰返すように前記分割手段及び前記分割選択手段を制御するための手段を含む、請求項1に記載のモデル構造作成装置。

## 【請求項 3】

前記制御手段は、前記分割選択手段によって算出された差分の絶対値が所定のしきい値より小さくなるまで、前記記憶手段に記憶された隠れマルコフモデルに対し、前記分割手段による準備及び前記分割選択手段による選択を繰返すように前記分割手段及び前記分割選択手段を制御するための手段を含む、請求項 1 に記載のモデル構造作成装置。

## 【請求項 4】

前記制御手段は、前記分割選択手段によって算出された差分の最大値が負となるまで、又は前記分割選択手段によって算出された差分の絶対値が所定のしきい値より小さくなるまで、前記記憶手段に記憶された隠れマルコフモデルに対し、前記分割手段による準備及び前記分割選択手段による選択を繰返すように前記分割手段及び前記分割選択手段を制御するための手段を含む、請求項 1 に記載のモデル構造作成装置。

10

## 【請求項 5】

前記分割選択手段は、

前記記憶手段に記憶された隠れマルコフモデル、及び前記分割手段により準備された隠れマルコフモデルの各々に対し、予め定める事前ハイパーパラメータに基づいて各モデル内の各状態における事後ハイパーパラメータの推定を行なうための事後ハイパーパラメータ推定手段と、

前記事後ハイパーパラメータ推定手段によって推定された前記事後ハイパーパラメータをもとに、各モデル内の各状態における前記評価関数の値をそれぞれ計算するための評価関数値算出手段とを含む、請求項 1 ~ 請求項 4 のいずれかに記載のモデル構造作成装置。

20

## 【請求項 6】

コンピュータにより実行されると、当該コンピュータを、請求項 1 ~ 請求項 5 のいずれかに記載のモデル構造作成装置として動作させる、モデル構造作成プログラム。

## 【請求項 7】

請求項 1 ~ 請求項 5 のいずれかに記載のモデル構造作成装置と、

前記モデル構造作成装置により作成される隠れマルコフモデルを記憶するための隠れマルコフモデル記憶手段と、

音声信号から予め定める特徴情報を抽出するための特徴抽出手段と、

前記特徴抽出手段により音声信号から抽出された特徴情報を用い、前記隠れマルコフモデル記憶手段に記憶された隠れマルコフモデルを用いて音声認識を行なうための音声認識手段とを含む、音声認識装置。

30

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

この発明は音声等の情報を認識するための認識技術に関し、特に、音声の認識に用いる音響モデルの作成方法において、HMM (Hidden Markov Model : 隠れマルコフモデル) の状態共有構造を自動的に作成する方法に関する。

## 【背景技術】

## 【0002】

コンピュータ技術及び信号処理技術の発達により、話者の発話音声から、話者による内容を示す言語的情報を認識する音声認識システムは自動化の域に達しつつある。例えばコンピュータの音声入力などが典型的な例である。

40

## 【0003】

音声認識では、時系列信号である音声信号を、HMMを用いてモデル化する技術が一般的である。HMMは、複数の定常信号源(状態)の間を遷移することで、時系列信号をモデル化する確率モデルである。

## 【0004】

音声認識の音響モデルとして、HMMの状態共有構造を自動生成する代表的な方法として、音素決定木クラスタリングと、逐次状態分割法(Successive State

50

Splitting (SSS) アルゴリズム) とがある。音素決定木クラスタリングは、音声学的知識による音素カテゴリを利用して、音素環境のクラスタリングを行なう方法である。これに対して、逐次状態分割法は、音素環境方向と時間方向との2方向の状態分割をデータに依存して行なう方法である。

【0005】

音素決定木クラスタリングでは、分割する状態を選択するための規準として、尤度最大 (Maximum Likelihood: ML) 規準を用いる。しかし、一般にモデルパラメータ数が増加すると尤度は増加する。そのため、ML 規準だけを規準として状態の分割を繰返し行なってモデル構造の生成を行なう場合、状態の分割を停止することが困難であった。そのため、状態分割の停止条件として、他の何らかの規準を用いる必要があった。

10

【0006】

そこで、音素決定木クラスタリングによる方法において、分割条件及び状態分割の停止条件として情報量規準を用いる方法が後掲の非特許文献1及び2に提案されている。非特許文献1では、最小記述長 (Minimum Description Length: MDL) 規準を用いる方法が提案されている。非特許文献2では、ベイズ的情報規準を用いる方法が提案されている。これらの方法では、情報量規準の値が改善されなくなるまで状態の分割を行なう。

【0007】

また、後掲の非特許文献3に記載の技術では、決定木クラスタリングを用いた音響モデル構造の作成において、分割状態の選択条件として、変分ベイズ (Variational Bayesian) 法を用いる方法が提案されている。変分ベイズ法は、変分法を用いて、ML 規準による学習の問題である過学習を避けるためのベイズ学習を実現する方法である。

20

【0008】

逐次状態分割法では、音素決定木クラスタリング法と同様に、状態の分割条件にML 規準を用いる手法が提案された。後掲の非特許文献4では、ML - SSS (ML 規準を用いたSSSアルゴリズム) 法として不特定話者モデルの作成方法にまで拡張されている。しかし、分割条件にML 規準を用いているため、音素決定木クラスタリングによる方法において生じる問題と同様の問題が生じた。即ち、状態の分割を停止することが困難であった。

30

【0009】

そこで後掲の非特許文献5に記載では、逐次状態分割法において、情報量規準としてMDL 規準を状態分割の分割条件及び停止条件に用いている。これにより、逐次状態分割法による状態分割を自動的に停止することが可能となった。

【0010】

【非特許文献1】コウイチ シノダ、タカオ ワタナベ、「音声認識のためのMDLに基づくコンテキスト依存サブワードモデリング」日本音響学会英文誌、vol. 21、no. 2、79 - 86頁、2000。(Koichi Shinoda, Takao Watanabe, "MDL-based context-dependent subword modeling for speech recognition," The Journal of the Acoustical Society of Japan (E), vol. 21, no. 2, pp. 79 - 86, 2000)

40

【非特許文献2】ウ・チョウ、W・ライヒル、「罰則付ベイズ的情報規準に基づく状態決定木の結合」、ICASSP '99 予稿集、vol. I、345 - 348頁、1999。(Wu Chou, W. Reichl, "Decision Tree State Tying Based on Penalized Bayesian Information Criterion," Proc. of ICASSP '99, vol. I, pp. 345 - 348, 1999)

【非特許文献3】S・ワタナベ、Y・ミナミ、A・ナカムラ、N・ウエダ、「ベイズ法に

50

基づくHMM状態共有構造の構築」、ICSLP予稿集、vol.4、2669-2672頁、2002。(S. Watanabe, Y. Minami, A. Nakamura and N. Ueda, "Constructing shared-state HMMs based on A Bayesian Approach," Proc. of ICSLP, vol.4, pp.2669-2672, 2002)

【非特許文献4】M. オステンドルフ、H. シンガー、「尤度最大規準逐次分割法を用いたHMM構造の設計」、Computer Speech and Language、vol.11、17-41頁、1997。(M. Ostendorf, H. Singer, "HMM topology Design using maximum likelihood successive state splitting," Computer Speech and Language, vol.11, pp.17-41, 1997)

10

【非特許文献5】T. ジツヒロ、T. マツイ、S. ナカムラ、「MDL規準に基づく一様でないコンテキスト依存HMM構造の自動生成」、EUROSPEECH'03予稿集、vol.4、2721-2724頁、2003。(T. Jitsuhiro, T. Matsui, S. Nakamura, "Automatic Generation of Non-Uniform Context-Dependent HMM Topologies Based on the MDL Criterion," Proc. of EUROSPEECH'03, vol.4, pp.2721-2724, 2003)

20

【発明の開示】

【発明が解決しようとする課題】

【0011】

非特許文献1及び2に記載の方法で用いる情報量規準は、漸近理論を用いて導出されたものである。そのため、学習データが少ないときに、これらの規準で適切なモデルを選択することは困難である。

【0012】

非特許文献3に記載の方法では、漸近性を仮定しないベイズ的規準を用いるため、非特許文献1及び2の問題を解決することができる。しかし、この方法では、従来の決定木クラスタリングと同様に、状態ごとの区間を既知として定義している。即ち、音素環境方向の分割のみを考慮しているため、時間方向の詳細なモデル構造を自動的に作成できない。

30

【0013】

非特許文献4に記載のML-SSS法では、ML規準を用いて状態分割を停止することが困難であった。そのため、状態分割の停止条件として、最終的なモデル構造における状態数 $N_s$ 及び時間方向の状態数制限 $N_p$ 、尤度の上昇率に対するしきい値など、他の何らかの規準を用いる必要があった。これらの他の規準は経験的に与える必要がある。さらには、あるデータに対して最適なモデルを1つ決定するためには、これらの他の規準を実験的に求める必要がある。さらに、学習データ、評価データが変わるたびに、これらの規準を調整しなければならない。

【0014】

40

非特許文献5に記載の方法では、逐次状態分割法において、状態分割を自動的に停止できる。しかし、非特許文献5に記載の方法で用いる情報量規準は、いくつかの仮定の上に導出されており、HMMのような複雑なモデルを厳密に扱うことができない。また、非特許文献1及び2に記載の方法と同様に、学習データが少ないときに、これらの情報量規準で適切なモデルを選択することは困難である。

【0015】

それゆえに、本発明の目的は、状態分割の停止条件を人手で与えることなく、状態分割を自動的に停止することができるモデル構造の作成方法を提供することである。

【0016】

本発明の別の目的は、音素環境方向及び時間方向の両方について、厳密なモデル評価を

50

行なうことができ、少数の学習データであっても、有効なモデル構造を自動的に作成することのできるモデル構造の作成方法を提供することである。

【課題を解決するための手段】

【0017】

本発明の第1の局面に係るモデル構造作成装置は、隠れマルコフモデルを記憶するための記憶手段と、隠れマルコフモデルに含まれる状態を時間方向と環境方向とに分割してそれぞれ新たな隠れマルコフモデルを準備する処理を、各状態に対して行なうための分割手段と、記憶手段に記憶される隠れマルコフモデルに対して、変分ベイズ法により計算される所定の評価関数の値と、分割手段により準備される新たな隠れマルコフモデルの各々に対して計算される評価関数の値との差分をそれぞれ算出し、差分が所定の選択条件を充足する分割を選択して記憶手段に記憶される隠れマルコフモデルを更新するための分割選択手段と、所定の終了条件が成立するまで、記憶手段に記憶された隠れマルコフモデルに対し、分割手段による準備、及び分割選択手段による選択を繰返すように分割手段及び分割選択手段を制御するための制御手段とを含む。

10

【0018】

好ましくは、制御手段は、分割選択手段によって算出された差分の最大値が負となるまで、記憶手段に記憶された隠れマルコフモデルに対し、分割手段による準備及び分割選択手段による選択を繰返すように分割手段及び分割選択手段を制御するための手段を含む。

【0019】

又は、分割選択手段によって算出された差分の絶対値が所定のしきい値より小さくなるまで、記憶手段に記憶された隠れマルコフモデルに対し、分割手段による準備及び分割選択手段による選択を繰返すように分割手段及び分割選択手段を制御するための手段を含んでもよい。

20

【0020】

より好ましくは、制御手段は、分割選択手段によって算出された差分の最大値が負となるまで、又は分割選択手段によって算出された差分の絶対値が所定のしきい値より小さくなるまで、記憶手段に記憶された隠れマルコフモデルに対し、分割手段による準備及び分割選択手段による選択を繰返すように分割手段及び分割選択手段を制御するための手段を含む。

【0021】

分割選択手段は、記憶手段に記憶された隠れマルコフモデル、及び分割手段により準備された隠れマルコフモデルの各々に対し、予め定める事前ハイパーパラメータに基づいて各モデル内の各状態における事後ハイパーパラメータの推定を行なうための事後ハイパーパラメータ推定手段と、事後ハイパーパラメータ推定手段によって推定された事後ハイパーパラメータをもとに、各モデル内の各状態における評価関数の値をそれぞれ計算するための評価関数値算出手段とを含んでもよい。

30

【0022】

好ましくは、評価関数は、変分ベイズ法により規定される自由エネルギーである。

【0023】

本発明の第2の局面に係るモデル構造作成プログラムは、コンピュータにより実行されると、当該コンピュータを、上記したいずれかのモデル構造作成装置として動作させるものである。

40

【0024】

本発明の第3の局面に係る音声認識装置は、上記したいずれかのモデル構造作成装置と、モデル構造作成装置により作成される隠れマルコフモデルを記憶するための隠れマルコフモデル記憶手段と、音声信号から予め定める特徴情報を抽出するための特徴抽出手段と、特徴抽出手段により音声信号から抽出された特徴情報を用い、隠れマルコフモデル記憶手段に記憶された隠れマルコフモデルを用いて音声認識を行なうための音声認識手段とを含む。

【発明を実施するための最良の形態】

50

## 【 0 0 2 5 】

< 変分ベイズ法を用いた逐次状態分割法 の概念 >

( 逐次状態分割法 の概念 )

逐次状態分割法は、複数の状態の連結によって表現されるHMMの状態共有構造を自動生成する方法である。逐次状態分割法の基本原理は、状態間の確率的な遷移により、音声の特徴パラメータの時間的な推移を表現した確率モデルに対して、状態を分割する操作を繰返すことにより、逐次的にHMMの状態共有構造を精密化するというものである。

## 【 0 0 2 6 】

図1を用いて、状態分割の概念について説明する。図1を参照して、逐次状態分割法における状態分割では、状態(白抜きの丸印で示す)と、状態の始端及び終端(黒塗りの丸印で示す)と、状態と状態の始点及び終点とを結ぶパス(矢印で示す)とを含む初期モデル20を設定する。図1に示す初期モデル20は、1つの状態により形成された初期モデルである。

10

## 【 0 0 2 7 】

逐次状態分割法では、まず、モデルの各状態に対して、パスの分割を伴わない時間方向への分割(以下、「時間方向分割」と呼ぶ。)22及びパスの分割を伴う音素環境方向への分割(以下、「音素環境方向分割」)24を試みる。時間方向分割22を行なうと、初期モデルから、状態が直列に並ぶ第1の分割モデル26が作成される。これに対して、音素環境方向分割24を行なうと、時間方向に状態が並列する第2の分割モデル28が作成される。このような分割を繰返すことで、複数の状態を複数のパスで連結した構造を有するHMM状態ネットワーク30を作成する。

20

## 【 0 0 2 8 】

状態分割によってHMM状態ネットワークを作成する場合、多数のモデル構造が作成され得ることになる。そのため、作成され得る多数のモデル構造の中から、音声認識にとって適切なものを選択する必要がある。そこで、モデル構造を分割する度に、分割によって作成されるモデル構造の中から最適なモデル構造を選択し、次の状態分割における分割対象とする。このように逐次的にモデル構造を選択することにより、適切なモデル構造を作成することが可能となる。ML-SSS法では、最適なモデル構造を選択するための規準に尤度最大規準を用いる。

## 【 0 0 2 9 】

< 変分ベイズ法を用いた逐次状態分割法 の原理 >

以下に、本発明の実施の形態に係る、変分ベイズ法を用いた逐次状態分割法の原理を説明する。なお、以下の説明のテキスト中で使用する記号「^」等は、本来はその直後の文字の直上に記載すべきものであるが、テキスト記法の制限により当該文字の直前に記載する。式中では、これらの記号等は本来の位置に記載してある。また、以下の説明では、変分ベイズ法を用いた逐次状態分割法を「VB-SSS法」と呼ぶ。

30

## 【 0 0 3 0 】

( 環境方向分割 )

総状態数 $N_s$ 個の状態を含むHMMがあるものとする。このHMMが含む状態は、それぞれ1個のガウス分布と $N_s$ 個の遷移確率とを持つ。このようなHMMに対し音素環境方向分割を行なった場合の、各状態における確率密度関数は下記のように表せる。

40

## 【 0 0 3 1 】

【数1】

$$p(\mathbf{O}|\Theta) = \prod_{t=1}^T p(o_t|s_t, \Theta) = \prod_{t=1}^T \mathcal{N}(o_t; \mu_{s_t}, \mathbf{S}_{s_t}^{-1})$$

ここで、 $\mathcal{N}()$  はガウス分布。

この式において、 $t$  は時刻、 $\mathbf{O} = \{o_1, \dots, o_t, \dots, o_T\}$  は学習データサンプル、 $s_t$  は

50

時刻  $t$  での状態番号、 $r_t$  は時刻  $t$  での遷移アーク番号をそれぞれ表わす。 $\mu_{s_t}$  は、時刻  $t$  での状態（状態番号  $s_t$  により表される状態）における出力確率の平均ベクトルを示す。また、 $S_{s_t}^{-1}$  は共分散行列を示す。 $S_{s_t}$  自体は精度行列と呼ばれ、共分散行列の逆行列である。

【0032】

（時間方向分割）

HMM に対して時間方向分割を行なう場合、遷移確率  $a_{s_t r_{t+1}}$  を考慮する。時間方向分割を行なった場合の確率密度関数は、下記ようになる。

【0033】

【数2】

$$\begin{aligned} p(\mathbf{O}|\Theta) &= \prod_{t=1}^T p(\mathbf{o}_t|s_t, \Theta) p(r_{t+1}|s_t, \Theta) \\ &= \prod_{t=1}^T \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_{s_t}, \mathbf{S}_{s_t}^{-1}) a_{s_t r_{t+1}} \end{aligned}$$

10

【0034】

（完全データ集合に対する確率分布）

逐次状態分割法では、音素環境方向分割に加えて時間方向分割を行なう。そのため、音素区間のみが既知で状態区間が未知という前提条件の上で状態分割を行なうことになる。このような前提条件に対して、変分ベイズ法では潜在変数が使用されるのが一般的である。潜在変数を導入した完全データに対する結合確率分布は、以下の式によって表される。

【数3】

$$p(\mathbf{O}, Z|\Theta) = \prod_{t=1}^T \prod_{i=1}^{N_s} \prod_{j=1}^{N_a} \{ \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_i, \mathbf{S}_i^{-1}) a_{ij} \}^{z_{ij}^t}$$

20

30

ただし、

$Z = \{z_{ij}^t\}_{i=1, j=1, t=1}^{N_s, N_a, T}$  は潜在変数の集合である。

【0035】

第  $i$  状態を分割するとき、分割前のパラメータ集合が  $\Theta_i$  である第  $i_1$  状態と第  $i_2$  状態とに分割した場合の分割後のパラメータ集合を  $\hat{\Theta}_i$  とすると、分割モデルを選択するための規準である分割規準となる評価関数は、ベイズ変分法における自由エネルギー  $F^{(VB)}$  を用いて下記のように表現できる。

40

【数4】

$$\Delta \mathcal{F}_{n+1}^{(VB)} = \mathcal{F}_{n+1}^{(VB)}(\hat{\Theta}_i) - \mathcal{F}_n^{(VB)}(\Theta_i)$$

ここで  $n$  は、学習回数即ち状態分割及び選択の回数を表わす。なお、 $F^{(VB)}$  の導出方法については後述する。

【0036】

（事前分布）

図2に本実施の形態に係るパラメータの関係を模式的に示す。図2を参照して、変分ベイズ法では、事前ハイパーパラメータを予め設定する必要があるが、通常、これらのハイ

50

パーパラメータに対する適切な設定値が存在する。そこで、事前ハイパーパラメータ  $\mu_0$ 、 $\nu_0$ 、 $\xi_0$ 、 $b_0$ 、及び  $s_0$  を設定する。また、現時点のモデル構造における状態数  $N_s$  は、現在のモデル構造を参照することにより得ることのできるパラメータである。各状態が持つ状態遷移確率の数  $N_a$  は、最大で  $N_s$  個あるが、VB-SS法では、1状態の遷移確率は自己ループと1つの状態への遷移とを扱う。そのため、 $N_a = 2$  となる。

【0037】

図2において、パラメータ「 $\mu$ 」54は平均ベクトルを示す。パラメータ「 $S$ 」56は精度行列を示す。パラメータ「 $a$ 」58は遷移確率を示す。パラメータ「 $z$ 」60は潜在変数である。また、パラメータ「 $o$ 」62は観測データである。

10

【0038】

本実施の形態では、確率変数間に、下記のような因子化仮定を導入する。

【0039】

【数5】

$$p(\Theta) = p(N_s, N_a) p(\mathbf{a} | N_s, N_a) p(\mathbf{S} | N_s) p(\boldsymbol{\mu} | \mathbf{S}, N_s)$$

さらに、遷移確率  $a$  (58) の事前分布にディリクレの分布を仮定する。ここで、 $D$  を特徴量次元とし、 $\mu_{ik}$  と  $s_{ik}$  とをそれぞれ第  $i$  状態での平均ベクトル  $\mu_i$  と精度行列  $S_i$  との  $k$  番目の要素とする。これら事前ハイパーパラメータを用いて、遷移確率  $a$  の事前分布に関する仮定は以下のように表される。

20

【0040】

【数6】

$$p(\mathbf{a} | N_s, N_a) = \prod_{i=1}^{N_s} \mathcal{D}(\{a_{ij}\}_{j=1}^{N_a}; \phi_0) \propto \prod_{i=1}^{N_s} \prod_{j=1}^{N_a} a_{ij}^{\phi_0 - 1}$$

$$a_{ij} \geq 0, \sum_{j=1}^{N_a} a_{ij} = 1$$

30

また、出力確率の平均ベクトル  $\mu$  と精度行列  $S$  との結合事前分布に正規ガンマ分布を仮定する。ガンマ分布は、ガンマ関数  $\Gamma(\cdot)$  を用いて、以下のように表される。

【0041】

【数7】

$$\mathcal{G}(s; \eta, \lambda) = \frac{\lambda^\eta}{\Gamma(\eta)} s^{\eta-1} \exp(-\lambda s), \quad s > 0$$

そのため、出力確率の平均ベクトル  $\mu$  と精度行列  $S$  との結合事前分布に関する仮定は、以下のように表わされる。

40

【0042】

【数8】

$$p(\boldsymbol{\mu}, \mathbf{S} | N_s) = \prod_{i=1}^{N_s} \prod_{k=1}^D \mathcal{N}(\mu_{ik}; \nu_{0k}, \xi_0^{-1} s_{ik}^{-1}) \mathcal{G}(s_{ik}; \eta_0/2, b_{0k}/2)$$

この式において、 $\nu_{0k}$  は、事前ハイパーパラメータであるベクトル  $\nu_0$  の  $k$  番目の要素を示し、 $b_{0k}$  は、同じく事前ハイパーパラメータである行列  $B_0$  の対角上の  $k$  番目の要素を示す。また、 $s_{ik}^{-1}$  は、共分散行列  $S_i^{-1}$  の対角上の  $k$  番目の要素を示す。

50



(事後分布)

各パラメータの最適変分事後分布は、変分ベイズ法における最適変分事後分布の一般式から導出できる。以下に、遷移確率  $a$  の事後確率を示す。

【 0 0 4 3 】

【数 9】

$$q(\mathbf{a}|\mathbf{O}, N_s, N_a) = \prod_{j=1}^{N_s} \mathcal{D}(\{a_{ij}\}_{j=1}^{N_a}; \{\phi_{ij}\}_{j=1}^{N_a}) \propto \prod_{j=1}^{N_s} \prod_{i=1}^{N_a} a_{ij}^{\phi_{ij}-1}$$

$$\text{ただし、 } \phi_{ij} = \phi_0 + \bar{N}_{ij}, \bar{N}_{ij} = \sum_{t=1}^T \bar{z}_{ij}^t, \bar{z}_{ij}^t = \langle z_{ij}^t \rangle_{q(Z)} \quad 10$$

また、平均ベクトルと精度行列との結合分布は以下のように表される。

【 0 0 4 4 】

【数 1 0】

$$q(\boldsymbol{\mu}, \mathbf{S}|\mathbf{O}, N_s)$$

$$= \prod_{i=1}^{N_s} \prod_{k=1}^D \mathcal{N}(\mu_{ik}; \nu_{ik}, \xi_i^{-1} s_{ik}^{-1}) \mathcal{G}(s_{ik}; \eta_i/2, b_{ik}/2) \quad 20$$

ただし、

$$\bar{N}_i = \sum_{t=1}^T \bar{z}_i^t, \bar{z}_i^t = \langle z_i^t \rangle_{q(Z)},$$

$$\nu_{ik} = \frac{\bar{N}_i \bar{o}_{ik} + \xi_0 \nu_{0k}}{\bar{N}_i + \xi_0}, \xi_i = \xi_0 + \bar{N}_i, \eta_i = \eta_0 + \bar{N}_i, \quad 30$$

$$b_{ik} = b_{0k} + \bar{c}_{ik} + \frac{\bar{N}_i \xi_0}{\bar{N}_i + \xi_0} (\bar{o}_{ik} - \nu_{0k})^2,$$

$$\bar{o}_{ik} = \frac{1}{\bar{N}_i} \sum_{t=1}^T \bar{z}_i^t o_{tk},$$

$$\bar{c}_{ik} = \sum_{t=1}^T \bar{z}_i^t (o_{tk} - \bar{o}_{ik})^2 \quad 40$$

$$= \sum_{t=1}^T \bar{z}_i^t o_{tk}^2 - \frac{1}{\bar{N}_i} \left( \sum_{t=1}^T \bar{z}_i^t \bar{o}_{ik} \right)^2$$

潜在変数  $z$  の変分事後確率分布は、以下の式のようになる。

【 0 0 4 5 】

【数 1 1】

$$\bar{z}_{ij}^t = \exp(\gamma_{ij}^t) / \sum_{k=1}^{N_s} \sum_{l=1}^{N_a} \exp(\gamma_{kl}^t),$$

$$\gamma_{ij}^t \propto \Psi(\phi_{ij}) - \Psi\left(\sum_{j=1}^{N_a} \phi_{ij}\right)$$

$$+ \frac{D}{2} \Psi\left(\frac{\eta_i}{2}\right) - \frac{1}{2} \sum_{k=1}^D \ln \frac{b_{ik}}{2}$$

$$- \frac{\eta_i D}{2\xi_i(\eta_i - D - 1)} - \frac{1}{2} \sum_{k=1}^D \frac{\eta_i}{b_{ik}} (o_{tk} - \nu_{ik})^2,$$

10

ただし、上記式において、

20

$\Psi(x) = \partial \ln \Gamma(x) / \partial x$  はディガンマ関数である。

【0046】

(目的関数)

以下に、前述した分割モデルの選択規準となる自由エネルギーの導出について説明する。なお、以下の説明において、状態数  $N_s$  及び遷移確率数  $N_a$  における確率  $p(\cdot | N_s, N_a)$  を単に  $p(\cdot)$  と表記する。

【0047】

自由エネルギー  $F^{(VB)}$  は、以下の式となる。

30

【数 1 2】

$$\mathcal{F}^{(VB)} = \int q(Z)q(\Theta) \ln \frac{p(\mathbf{O}, Z|\Theta)p(\Theta)}{q(Z)q(\Theta)} dZ d\Theta$$

$$= \int q(Z)q(\Theta) \ln \prod_{t=1}^T p(\mathbf{o}_t, z_t|\Theta) dZ d\Theta$$

$$- \int q(Z) \ln q(Z) dZ - \int q(\Theta) \ln q(\Theta) d\Theta$$

40

$$+ \int q(\Theta) \ln p(\Theta) d\Theta.$$

上記数式の第 1 項と第 2 項とを合わせたものを  $F_{12}$  とし、第 3 項を  $F_3$ 、第 4 項を  $F_4$  とする。

【0048】

 $F_{12}$  は、以下のように導出できる。

【0049】

50

【数 1 3】

$$\begin{aligned}
\mathcal{F}_{12} &= \int q(Z)q(\Theta) \ln \prod_{t=1}^T p(\mathbf{o}, z_t | \Theta) d\Theta - \int q(Z) \ln q(Z) dZ \\
&= \sum_{t=1}^T \int q(Z) dZ \{ \langle \ln p(\mathbf{o}, z_t | \Theta) \rangle_{q(\Theta)} - \ln q(z_t) \} \\
&= \sum_{t=1}^T \ln \sum_{i=1}^{N_s} \sum_{j=1}^{N_a} \exp(\gamma_{ij}^t).
\end{aligned}$$

10

第 3 項  $F_3$  と第 4 項  $F_4$  とを变形するために、以下に示す因子化仮定を用いる。

【0 0 5 0】

【数 1 4】

$$p(\theta) = \prod_{i=1}^{N_s} p(\{a_{ij}\}_{j=1}^{N_a}) p(\boldsymbol{\mu}_i | \mathbf{S}_i) p(\mathbf{S}_i)$$

20

この因子化仮定を用いて、第 3 項  $F_3$  は、以下のように变形できる。

【0 0 5 1】

【数 1 5】

$$\begin{aligned}
\mathcal{F}_3 &= \int q(\Theta) \ln q(\Theta) d\Theta \\
&= \sum_{i=1}^{N_s} \left\{ \int q(\{a_{ij}\}_{j=1}^{N_a}) \ln q(\{a_{ij}\}_{j=1}^{N_a}) d\{a_{ij}\}_{j=1}^{N_a} \right. \\
&\quad + \int q(\boldsymbol{\mu}_i | \mathbf{S}_i) q(\mathbf{S}_i) \ln q(\boldsymbol{\mu}_i | \mathbf{S}_i) d\boldsymbol{\mu}_i d\mathbf{S}_i \\
&\quad \left. + \int q(\mathbf{S}_i) \ln q(\mathbf{S}_i) d\mathbf{S}_i \right\},
\end{aligned}$$

30

また、第 4 項  $F_4$  は、以下のように变形できる。

【0 0 5 2】

【数 1 6】

$$\begin{aligned}
\mathcal{F}_4 &= \int q(\Theta) \ln p(\Theta) d\Theta \\
&= \sum_{i=1}^{N_s} \left\{ \int q(\{a_{ij}\}_{j=1}^{N_a}) \ln p(\{a_{ij}\}_{j=1}^{N_a}) d\{a_{ij}\}_{j=1}^{N_a} \right. \\
&\quad + \int q(\boldsymbol{\mu}_i | \mathbf{S}_i) q(\mathbf{S}_i) \ln p(\boldsymbol{\mu}_i | \mathbf{S}_i) d\boldsymbol{\mu}_i d\mathbf{S}_i \\
&\quad \left. + \int q(\mathbf{S}_i) \ln p(\mathbf{S}_i) d\mathbf{S}_i \right\}.
\end{aligned} \tag{10}$$

上記のように、変形した第 3 項  $F_3$  の第 1 項、及び第 4 項  $F_4$  の第 1 項はそれぞれ第  $i$  状態の遷移確率に関する成分となる。第 3 項  $F_3$  の第 1 項を  $F_{31}$  とし、第 4 項  $F_4$  の第 1 項を  $F_{41}$  とする。 $F_{31}$  は、以下のようになる。

【0 0 5 3】

【数 1 7】

$$\begin{aligned}
\mathcal{F}_{31}(\{a_{ij}\}_{j=1}^{N_a}) &= \int \mathcal{D}(\{a_{ij}\}_{j=1}^{N_a}; \{\phi_{ij}\}_{j=1}^{N_a}) \\
&\quad \times \ln \mathcal{D}(\{a_{ij}\}_{j=1}^{N_a}; \{\phi_{ij}\}_{j=1}^{N_a}) da_{ij} \\
&= \ln \Gamma \left( \sum_{j=1}^{N_a} \phi_{ij} \right) - \sum_{j=1}^{N_a} \ln \Gamma(\phi_{ij}) \\
&\quad + \sum_{j=1}^{N_a} (\phi_{ij} - 1) \left\{ \Psi(\phi_{ij}) - \Psi \left( \sum_{j=1}^{N_a} \phi_{ij} \right) \right\}
\end{aligned} \tag{30}$$

また、 $F_{41}$  は、下記のようになる。

【0 0 5 4】

【数 1 8】

$$\begin{aligned}
\mathcal{F}_{41}(\{a_{ij}\}_{j=1}^{N_a}) &= \int \mathcal{D}(\{a_{ij}\}_{j=1}^{N_a}; \{\phi_{ij}\}_{j=1}^{N_a}) \\
&\quad \times \ln \mathcal{D}(\{a_{ij}\}_{j=1}^{N_a}; \phi_0) da_{ij} \\
&= \ln \Gamma(N_a \phi_0) - N_a \ln \Gamma(\phi_0) \\
&\quad + (\phi_0 - 1) \sum_{j=1}^{N_a} \left\{ \Psi(\phi_{ij}) - \Psi \left( \sum_{j=1}^{N_a} \phi_{ij} \right) \right\}
\end{aligned} \tag{40}$$

したがって、遷移確率に対する自由エネルギーを  $F_{341}$  とすると、 $F_{341}$  は、以下のように導

出できる。

【 0 0 5 5 】

【 数 1 9 】

$$\begin{aligned}
 \mathcal{F}_{341}(\{a_{ij}\}_{j=1}^{N_a}) &= -\mathcal{F}_{31}(\{a_{ij}\}_{j=1}^{N_a}) + \mathcal{F}_{41}(\{a_{ij}\}_{j=1}^{N_a}) \\
 &= \ln \Gamma(N_a \phi_0) - \ln \Gamma\left(\sum_{j=1}^{N_a} \phi_{ij}\right) \\
 &\quad + \sum_{j=1}^{N_a} \ln \Gamma(\phi_{ij}) - N_a \ln \Gamma(\phi_0) \\
 &\quad + \sum_{j=1}^{N_a} (\phi_0 - \phi_{ij}) \left\{ \Psi(\phi_{ij}) - \Psi\left(\sum_{j=1}^{N_a} \phi_{ij}\right) \right\}
 \end{aligned}$$

10

さらに、 $F_3$ の第2項及び $F_4$ の第2項は、平均ベクトルと精度行列との結合分布に関する  
 $F_3$ の第2項を $F_{32}$ とすると、 $F_{32}$ は、以下のようになる。

20

【 0 0 5 6 】

【 数 2 0 】

$$\begin{aligned}
 \mathcal{F}_{32}(\boldsymbol{\mu}_i, \mathbf{S}_i) &= \int q(\boldsymbol{\mu}_i | \mathbf{S}_i) q(\mathbf{S}_i) \ln q(\boldsymbol{\mu}_i | \mathbf{S}_i) d\boldsymbol{\mu}_i d\mathbf{S}_i \\
 &= -\frac{D}{2} \ln 2\pi + \frac{1}{2} \ln \xi_i + \frac{1}{2} \langle \ln |\mathbf{S}_i| \rangle_{q(\mathbf{S}_i)} \\
 &\quad - \frac{1}{2} \text{Tr} \{ \xi_i \langle \mathbf{S}_i \rangle_{q(\mathbf{S}_i)} \langle (\boldsymbol{\mu}_i - \boldsymbol{\nu}_i)(\boldsymbol{\mu}_i - \boldsymbol{\nu}_i) \rangle'_{q(\boldsymbol{\mu}_i)} \} \\
 &= -\frac{D}{2} \ln 2\pi + \frac{1}{2} \ln \xi_i + \frac{D}{2} \Psi\left(\frac{\eta_i}{2}\right) \\
 &\quad - \frac{1}{2} \sum_{k=1}^D \ln \left(\frac{b_{ik}}{2}\right) - \frac{\eta_i}{2(\eta_i - D - 1)},
 \end{aligned}$$

30

40

また、 $F_4$ の第2項を $F_{42}$ とすると、 $F_{42}$ は、以下のようになる。

【 0 0 5 7 】

【数 2 1】

$$\begin{aligned}
\mathcal{F}_{42}(\boldsymbol{\mu}_i, \mathbf{S}_i) &= \int q(\boldsymbol{\mu}_i | \mathbf{S}_i) q(\mathbf{S}_i) \ln p(\boldsymbol{\mu}_i | \mathbf{S}_i) d\boldsymbol{\mu}_i d\mathbf{S}_i \\
&= -\frac{D}{2} \ln 2\pi + \frac{1}{2} \ln \xi_0 + \frac{D}{2} \Psi\left(\frac{\eta_0}{2}\right) \\
&\quad - \frac{1}{2} \sum_{k=1}^D \ln\left(\frac{b_{0k}}{2}\right) - \frac{\eta_0}{2(\eta_0 - D - 1)}.
\end{aligned}
\tag{10}$$

したがって、平均ベクトルと精度行列との結合分布に関する自由エネルギーを  $F_{342}$  とすると、 $F_{342}$  は、以下のように導出される。

【0058】

【数 2 2】

$$\begin{aligned}
\mathcal{F}_{342}(\boldsymbol{\mu}_i, \mathbf{S}_i) &= \frac{1}{2} \ln \frac{\xi_0}{\xi_i} + \frac{D}{2} \left\{ \Psi\left(\frac{\eta_0}{2}\right) - \Psi\left(\frac{\eta_i}{2}\right) \right\} - \frac{1}{2} \sum_{k=1}^D \ln\left(\frac{b_{0k}}{b_{ik}}\right) \\
&\quad - \frac{\eta_0}{2(\eta_0 - D - 1)} + \frac{\eta_i}{2(\eta_i - D - 1)}.
\end{aligned}
\tag{20}$$

【0059】

さらに、 $F_3$  の第 3 項及び  $F_4$  の第 3 項は、精度行列のみに関する。 $F_3$  の第 3 項を  $F_{33}$  とすると、 $F_{33}$  は、

【数 2 3】

$$\begin{aligned}
\mathcal{F}_{33}(\mathbf{S}_i) &= \int q(\mathbf{S}_i) \ln q(\mathbf{S}_i) d\mathbf{S}_i \\
&= \sum_{k=1}^D \int \mathcal{G}(s_{ik}; \frac{\eta_i}{2}, \frac{b_{ik}}{2}) \ln \mathcal{G}(s_{ik}; \frac{\eta_i}{2}, \frac{b_{ik}}{2}) ds_{ik} \\
&= D \left\{ \left(\frac{\eta_i}{2} - 1\right) \Psi\left(\frac{\eta_i}{2}\right) - \ln \Gamma\left(\frac{\eta_i}{2}\right) - \frac{\eta_i}{2} \right\} \\
&\quad + \sum_{k=1}^D \ln\left(\frac{b_{ik}}{2}\right),
\end{aligned}
\tag{30}$$

のようになる。

【0060】

また、 $F_4$  の第 3 項を  $F_{43}$  とすると、 $F_{43}$  は、以下のようになる。

【0061】

【数 2 4】

$$\begin{aligned}
\mathcal{F}_{43}(\mathbf{S}_i) &= \int q(\mathbf{S}_i) \ln p(\mathbf{S}_i) d\mathbf{S}_i \\
&= \sum_{k=1}^D \int \mathcal{G}(s_{ik}; \frac{\eta_i}{2}, \frac{b_{ik}}{2}) \ln \mathcal{G}(s_{ik}; \frac{\eta_0}{2}, \frac{b_{0k}}{2}) ds_{ik} \\
&= D \left\{ \left( \frac{\eta_0}{2} - 1 \right) \Psi \left( \frac{\eta_i}{2} \right) - \ln \Gamma \left( \frac{\eta_0}{2} \right) \right\} \\
&\quad + \sum_{k=1}^D \left\{ \frac{\eta_0}{2} \ln \left( \frac{b_{0k}}{b_{ik}} \right) + \ln \left( \frac{b_{ik}}{2} \right) - \frac{\eta_i b_{0k}}{2b_{ik}} \right\}.
\end{aligned}
\tag{10}$$

したがって、精度行列のみに関する自由エネルギーを  $F_{343}$  とすると、 $F_{343}$  は、以下のように導出できる。

【0062】

【数 2 5】

$$\begin{aligned}
\mathcal{F}_{343}(\mathbf{S}_i) &= D \left\{ \frac{\eta_0 - \eta_i}{2} \Psi \left( \frac{\eta_i}{2} \right) - \ln \Gamma \left( \frac{\eta_0}{2} \right) + \ln \Gamma \left( \frac{\eta_i}{2} \right) \right. \\
&\quad \left. + \frac{\eta_i}{2} \right\} + \frac{1}{2} \sum_{k=1}^D \left\{ \eta_0 \ln \left( \frac{b_{0k}}{b_{ik}} \right) - \frac{\eta_i b_{0k}}{b_{ik}} \right\}
\end{aligned}
\tag{20}$$

【0063】

最終的には、以上のようにして導出できる  $F_{12}$ 、 $F_{341}$ 、 $F_{342}$ 、及び  $F_{343}$  を用いて、自由エネルギー  $F^{(VB)}$  は、

【数 2 6】

$$\begin{aligned}
\mathcal{F}^{(VB)} &= \mathcal{F}_{12} + \sum_{i=1}^{N_s} \left\{ \mathcal{F}_{341}(\{a_{ij}\}_{j=1}^{N_a}) \right. \\
&\quad \left. + \mathcal{F}_{342}(\boldsymbol{\mu}_i, \mathbf{S}_i) + \mathcal{F}_{343}(\mathbf{S}_i) \right\},
\end{aligned}
\tag{30}$$

と表される。この式を用いて前述の自由エネルギー及びその差分を導出できる。

【0064】

&lt;構成&gt;

図3に、本実施の形態に係る音声認識システムの構成を示す。この音声認識システムは、所定の記憶装置に記憶された、状態分割を行なう前の初期モデル70と、音響モデルのモデル構造を作成するために用いる学習データ72とに基づいて、上記した原理にしたがって自動的に状態分割を行なってHMM84を生成するモデル構造作成部82と、入力音声76の音声信号について、HMM84を用いて音声認識を行ない、認識出力78を出力する音声認識装置80とを含む。

【0065】

音声認識装置80は、入力音声76の音響的特徴を抽出する特徴抽出部86と、HMM84を用いて、特徴抽出部86が抽出した音響的特徴に基づく音声の認識を行ない、認識出力78を出力する認識部88とを含む。

## 【 0 0 6 6 】

<モデル構造作成処理>

図4は、本実施の形態に係るモデル構造作成部82が行なうモデル構造作成処理のアルゴリズムを示すフローチャートである。以下の説明からも明らかとなり、この処理を含めモデル構造作成部82と音声認識装置80とは、いずれもコンピュータハードウェアと、その上で実行されるプログラムとにより実現可能である。

## 【 0 0 6 7 】

図4を参照して、モデル構造作成部82が処理を開始すると、ステップ(以下、単に「S」と表記する。)102に進む。S102では、初期モデル70のモデル構造を設定する。S104では、初期モデル70におけるパラメータを推定する。続くS106では、事前ハイパーパラメータの設定を行ない、S108に進む。

## 【 0 0 6 8 】

S108では、状態分割の対象となるモデル(最初の処理では初期モデル70、2回目の繰返し以降は分割後のモデル)構造の各状態において、変分法を用いて事後ハイパーパラメータを推定する。S110では、状態分割の対象となるモデルの自由エネルギーを計算する。続くS112では、分割前のモデル構造の自由エネルギーと分割後のモデル構造の自由エネルギーとを比較し、自由エネルギーが収束しているか否かを判定する。より具体的には、分割前の自由エネルギーと分割後の自由エネルギーとの差分の絶対値が所定のしきい値よりも小さい場合に、自由エネルギーが収束していると判定すればよい。判定結果が「YES」であった場合、状態分割を停止して、処理を終了する。判定結果が「NO」の場合、モデル中のす

べての状態に対して、S114からS118までのステップで構成される時間方向分割処理と、S120からS124までのステップで構成される音素環境方向分割処理とを行なう。なお、S112の判定処理は、繰返し処理の最初には行なわず、2回目以降の繰返し時のみ実行する。

## 【 0 0 6 9 】

S114では、モデル中の1つの状態を時間方向に分割する。S116では、S114で時間方向分割した分割モデルの事後ハイパーパラメータを推定する。続くS118では、時間方向分割した分割モデルについて、自由エネルギーを計算し、S110で求めた分割前のモデルとの間のエネルギー差分を算出する。

## 【 0 0 7 0 】

S120では、モデル中の1つの状態を音素環境方向に分割する。S122では、時間方向分割処理におけるS116と同様の処理を行ない、S120で音素環境方向分割した分割モデルの事後ハイパーパラメータを推定する。続くS124では、時間方向分割処理におけるS118と同様の処理を行ない、音素環境方向分割した分割モデルについて、自由エネルギーを計算し、S110で求めた分割前のモデルとの間のエネルギー差分を算出する。

## 【 0 0 7 1 】

分割対象となったモデル中のすべての状態に対して、S114～S118の処理及びS120～S124の処理が終了すると、S126で分割モデルの選択を行なう。即ち、S126では、時間方向分割処理又は音素環境方向分割処理により生成された分割モデルのうち、S118及びS124のいずれかで算出したエネルギー差分が最大となるのはどれかを判定する。そして、エネルギー差分が最大となる分割モデルを選択し、記憶装置に記憶された隠れマルコフモデルを更新する。

## 【 0 0 7 2 】

続くS128では、選択したモデルのエネルギー差分を参照し、状態分割によって自由エネルギーが減少したか否かを判定する。自由エネルギーが減少していた場合、状態分割を停止し、処理を終了する。それ以外の場合、S130で、選択したモデルを、次回の状態分割の対象となるモデル構造とし、モデル構造のパラメータの再推定を行ない、S108に戻る。



## 【 0 0 7 3 】

## &lt; 動作 &gt;

図 5 は、本実施の形態に係るモデル構造作成部 8 2 が行なう V B - S S S 法による状態分割を模式的に示す図である。以下、図 5 を参照しつつ、本実施の形態に係るモデル構造作成部 8 2 が行なう V B - S S S 法による状態分割の動作を説明する。なお、図 5 では図 1 と同様にモデルの状態及びパスを示す。

## 【 0 0 7 4 】

初期モデル 1 5 0 は、図 1 に示す初期モデル 2 0 と同様のものである。本実施の形態に係る状態分割では、図 1 に示す従来の逐次状態分割法における状態分割と同様の方法で、初期モデル 1 5 0 に対して時間方向分割及び音素環境方向分割を行なう。時間方向分割を行なうと第 1 の分割モデル 1 5 2 が作成される。また、音素環境方向に分割すると、第 2 の分割モデル 1 5 4 が作成される。

10

## 【 0 0 7 5 】

本実施の形態に係るモデル構造作成部 8 2 は、まず、これら作成した各分割モデルの自由エネルギーを計算し、自由エネルギーが最大の分割モデルを選択する。ここでは、第 2 のモデル 1 5 4 が選択されたものとする。

## 【 0 0 7 6 】

この後、さらに本実施の形態に係るモデル構造作成部 8 2 は、選択されたモデル 1 5 6 と、状態分割前のモデル（第 1 回目のモデル分割では初期モデル 1 5 0 がこれに相当する。）との自由エネルギーを比較する。比較の結果、自由エネルギーが減少していると判定された場合、又は収束していると判定された場合、モデル構造作成部 8 2 はこの結果に基づき状態分割を停止し、選択されたモデル 1 5 6 を出力モデル 1 5 8 として出力する。出力されたモデルが図 3 に示す H M M 8 4 となる。

20

## 【 0 0 7 7 】

それ以外の場合、選択されたモデル 1 5 6 のすべての状態に対して、時間方向及び状態方向の分割をそれぞれ行ない、新たな分割モデル 1 6 0 ~ 1 6 3 を作成する。以後、新たな分割モデル 1 6 0 ~ 1 6 3 に対し、上記したモデル 1 5 2 と 1 5 4 とに対し行なったものと同様の処理を実行する。このような処理を繰返すことにより、モデル構造作成部 8 2 は、自由エネルギーが減少するか、収束するまで、状態分割と、自由エネルギーを選択規準とする分割モデルの選択と、分割前のモデル及び分割後のモデルの比較とを繰返す。その結果、自由エネルギーが最大のモデル 1 6 4 が作成される。

30

## 【 0 0 7 8 】

図 3 を参照して、モデル構造作成部 8 2 は、作成した自由エネルギーが最大のモデル 1 6 4 を出力する。H M M 8 4 には、モデル構造作成部 8 2 が作成したモデルが保持される。

## 【 0 0 7 9 】

音声認識装置 8 0 は、このようにして作成された H M M 8 4 を用いて入力音声 7 6 に対する音声認識を行なう。入力音声 7 6 が音声認識装置 8 0 に入力されると、特徴抽出部 8 6 が、入力音声 7 6 の音響的特徴を抽出し、そのデータを認識部 8 8 に与える。認識部 8 8 は、H M M 8 4 を用いて、与えられたデータに適合するような音素など言語情報の組合せ及びその順序を推定する。認識部 8 8 は、推定した言語情報を、認識出力 7 8 として出力する。

40

## 【 0 0 8 0 】

## [ 実験 ]

本実施の形態に係るモデル構造の作成方法を用いて H M M を作成し、作成した H M M を用いて、切り出し音素認識実験と連続音声認識とを行なった。切り出し音素認識実験は音素区間ごとに切り出された区間に対して行なう音素識別実験である。この実験では、各音素区間に対し、最も良いスコアの音素モデルを認識結果とする。これにより、得られた各音素モデルの性能評価を行なうことができる。

## 【 0 0 8 1 】

本実施の形態に係るモデル構造の作成方法と比較を行なうために、尤度最大規準による

50

ML - SSS法と、MDL規準を分割及び停止規準とするMDL - SSS法とを用いて、切出し音素認識実験と連続音声認識とを行なった。さらに、ML - SSS法では最大状態長を3としたモデルと、最大状態長を4としたモデルとの2種類を作成し、実験を行なった。

【0082】

MDL - SSS法では、下記の規準を用いた。

【0083】

【数27】

$$G_c^{(MDL)}(S_i) = -G_c^{(ML)}(S_i) + C_c \frac{\alpha'_c - \alpha_c}{2} \log \Gamma(S), \quad 10$$

$$G_t^{(MDL)}(S_i) = -G_t^{(ML)}(S_i) + C_t \left\{ \frac{\alpha'_t}{2} \log \Gamma'(S) - \frac{\alpha_t}{2} \log \Gamma(S) \right\},$$

ここでは、環境分割の対数尤度ゲインを  $G_c^{(ML)}$  とし、時間方向分割の対数尤度ゲインを  $G_t^{(ML)}$  とし、また、 $p_t(S_i)$  を時刻  $t$  に状態  $S_i$  に留まる確率とし、さらに、 $\Gamma(S) = \sum_i p_t(S_i)$  をすべての状態に対するサンプル数の期待値とする。 $\Gamma'(S)$  を時間分割後の  $\Gamma(S)$  の値とし、 $\alpha_c$ 、 $\alpha'_c$  をそれぞれ分割前後のパラメータ数とする。なお、重み係数  $C_c$  と  $C_t$  とは、それぞれ  $C_c = 2$  と  $C_t = 20$  とした。

【0084】

音響モデル学習データとして、出願人において作成した旅行会話タスクデータベースより男性166人の日本語対話音声、約2.1時間分を用いた。また、評価には、同じデータベースから学習データに含まれない17人の男性話者による213発話を用いた。なお、本実験では、変分ベイズ法を分割及び停止時の規準としてのみ用いた。認識実験では、従来通りのHMMによる計算を行なった。さらに、単語連続音声認識実験を行なった。この実験には、言語モデルとして、多重クラス複合bigramモデルを用い、語彙サイズは5千単語とした。音声データのサンプリング周波数は16(kHz)、フレーム長は20ms、フレーム周期は10msとした。特徴量として12次MFCC(Mel-Frequency Cepstrum Coefficient)、MFCC、及び対数パワーを用いた。ケプストラム平均正規化は発話単位で行なった。音素体系として26種類の音素とひとつの無音を用いた。音素ごとに音響モデル構造を作成し、初期モデルはそれぞれ3状態モデルとした。構造作成中は1状態につき1ガウス分布を用いた。無音モデルは3状態で音素モデルとは別に作成した。

【0085】

上記の通り、変分ベイズ法では一般に事前ハイパーパラメータの設定が必要になる。いくつかの予備実験から、本実験では、事前ハイパーパラメータを  $\mu_0 = 0.1$ 、 $\sigma_0 = 1.0$ 、 $\nu_0 = 2.0$  とした。 $\mu_{0k}$  と  $\nu_{0k}$  については、実際の分布の平均ベクトル及び精度行列を基に設定を行なった。

【0086】

それぞれのモデルを用いた音素認識実験について、図6及び図7に、モデルの状態数と母音に対する音素認識率との関係、及び同じく状態数に対する子音に対する音素認識率をそれぞれ示す。さらに、図8に、モデルの状態数と、子音の音素認識率及び母音の音素認識率の平均値との関係を示す。

【0087】

これらの図には、最大状態長3及び4のML - SSS法によるもの、MDL - SSS法 50

によるもの、並びに本実施の形態に係る変分ベイズ法を用いた逐次分割法（VB - SSS法）によるものを示す。これらの図においては、少ない状態数であり、かつ高い音素認識率となる方法が、より高性能のモデル構造作成方法となる。

【0088】

図6、図7、及び図8より、本実施の形態に係るVB - SSS法では、特に母音に対する音素認識率は、従来の方法より高い値となることが示されている。また、MDL - SSS法では状態数が小さいモデルが推定され、他のモデル構造作成方法によるモデルに比べて、音素認識率が低い値となっている。これは、MDL規準が、少量の学習データでは十分なモデル構造を作成する性能を発揮しないことを示していると考えられる。

【0089】

さらに、切り出し音素認識実験で用いたものと同じの評価データを用いて、語彙5千単語での単語連続認識実験を行なった。なお、評価には、音素認識実験と同様に単一分布モデルを用いた。図9に3種類の方法による単語認識率を示す。

【0090】

図9を参照して、本実施の形態に係るVB - SSS法によるモデルでは、状態数1330のモデルで71パーセント以上の単語認識率となった。ML - SSS法で同等の単語認識率を得るには、状態数2200程度のモデルを作成する必要があることが示された。よって、本実施の形態に係る方法で作成したモデルは、ベースラインのML - SSS法で得られたモデルの約60%のモデルサイズで、同等の性能が得られることになる。

【0091】

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。

【図面の簡単な説明】

【0092】

【図1】音響モデル構造の作成における、状態分割の概念を示す模式図である。

【図2】変分ベイズ法を用いた逐次状態分割法における、確率変数間の関係を模式的に示す図である。

【図3】本発明の実施の形態に係る音声認識システムの機能的構成を示す図である。

【図4】モデル構造作成処理のアルゴリズムを示すフローチャートである。

【図5】モデル構造作成部82が行なうVB - SSS法による状態分割を模式的に示す図である。

【図6】モデルの状態数と母音に対する音素認識率との関係を示すグラフである。

【図7】モデルの状態数と子音に対する音素認識率との関係を示すグラフである。

【図8】モデルの状態数と音素認識率の平均値との関係を示すグラフである。

【図9】モデルの状態数と単語認識率との関係を示すグラフである。

【符号の説明】

【0093】

40、42、44、46、48 事前ハイパーパラメータ、50 状態数、52 遷移確率の数、54 平均ベクトル、56 精度行列、58 遷移確率、60 潜在変数、62 観察データ、70 初期モデル、72 学習データ、76 入力音声、78 認識出力、80 音声認識装置、82 モデル構造作成部、84 HMM、86 特徴抽出部、88 認識部

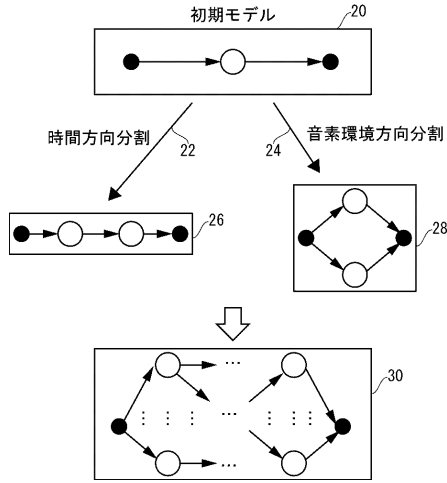
10

20

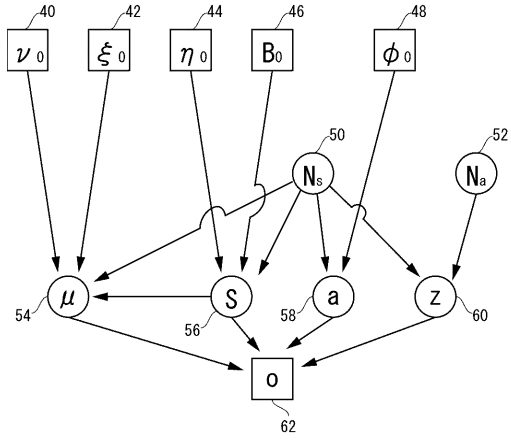
30

40

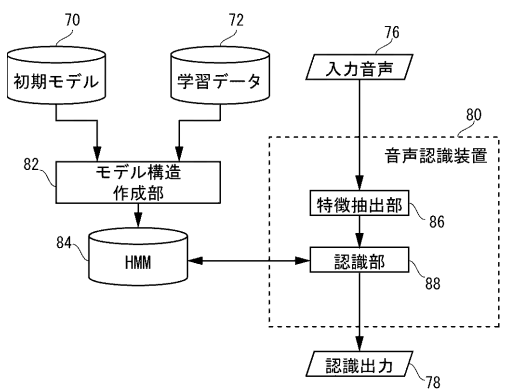
【図1】



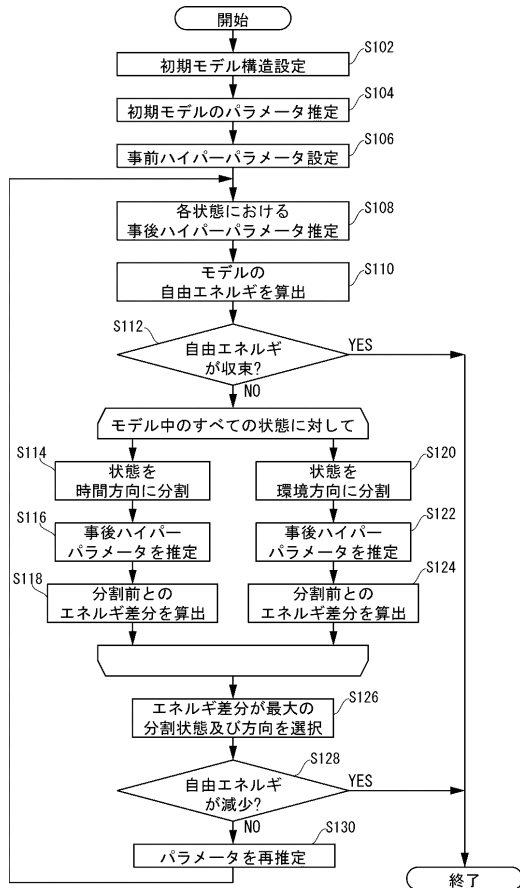
【図2】



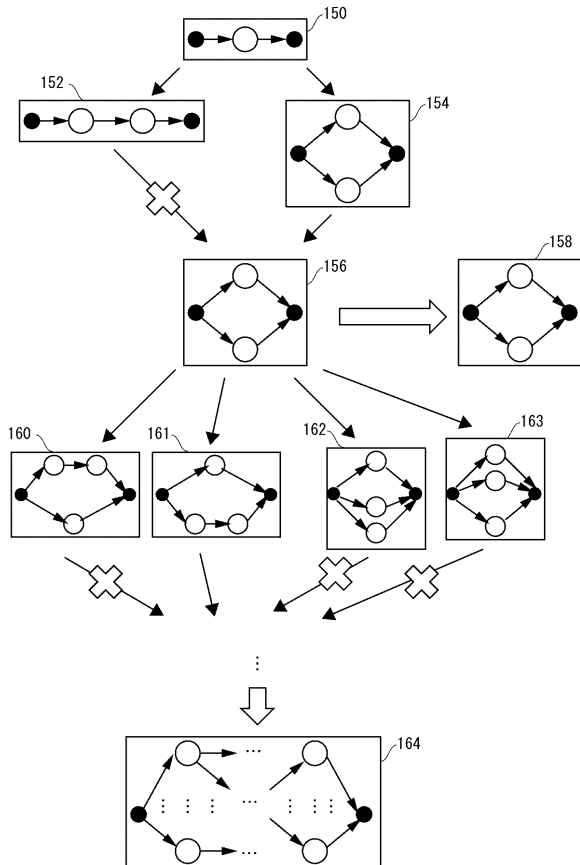
【図3】



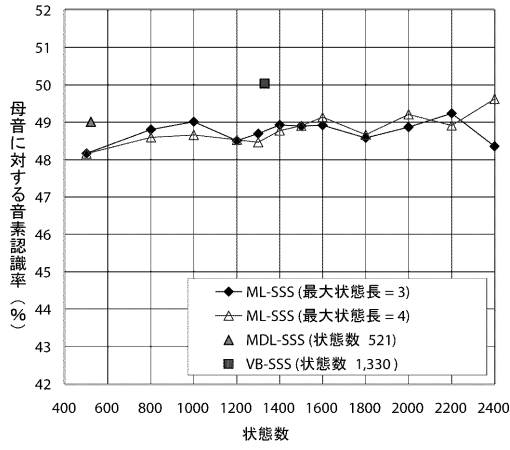
【図4】



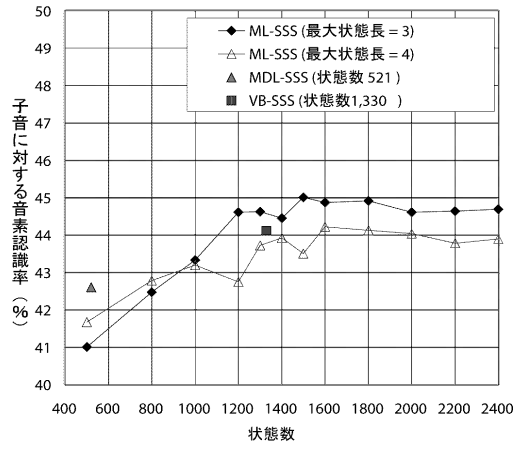
【図5】



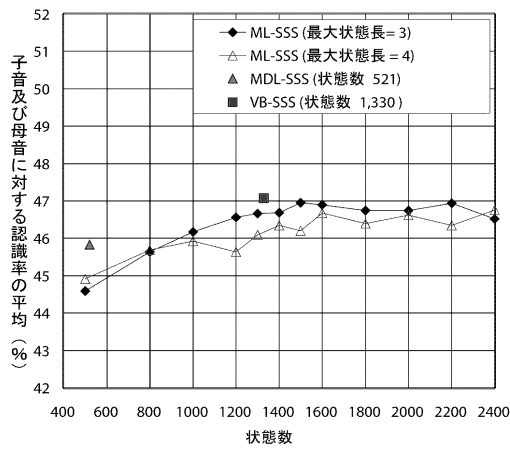
【図6】



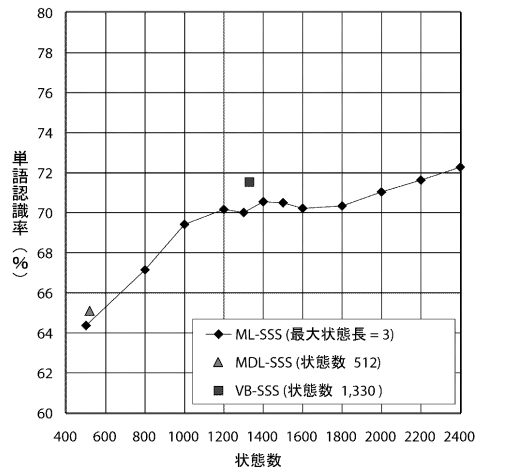
【図7】



【図8】



【図9】



---

フロントページの続き

- (56)参考文献 實廣 貴敏、松井 知子、中村 哲、MDL基準を用いた逐次状態分割法による音響モデル自動構造決定、情報処理学会研究報告、日本、社団法人情報処理学会、2002年12月20日、Vol.2002 No.121, pp.37-42  
渡部 晋治、南 泰浩、中村 篤、上田 修功、ベイズ的基準を用いた状態共有型HMM構造の選択、電子情報通信学会論文誌、日本、社団法人電子情報通信学会、2003年6月1日、Vol.J86-D-II No.6, pp.776-786

- (58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28