

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-54588

(P2010-54588A)

(43) 公開日 平成22年3月11日(2010.3.11)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/14 (2006.01)	G 1 0 L 15/14 2 0 0 A	5 D 0 1 5
G 1 0 L 15/06 (2006.01)	G 1 0 L 15/06 3 0 0 C	

審査請求 未請求 請求項の数 10 O L (全 24 頁)

(21) 出願番号	特願2008-216640 (P2008-216640)	(71) 出願人	000004226
(22) 出願日	平成20年8月26日 (2008. 8. 26)		日本電信電話株式会社
			東京都千代田区大手町二丁目3番1号
		(74) 代理人	100121706
			弁理士 中尾 直樹
		(74) 代理人	100066153
			弁理士 草野 卓
		(74) 代理人	100128705
			弁理士 中村 幸雄
		(72) 発明者	渡部 晋治
			東京都千代田区大手町二丁目3番1号 日
			本電信電話株式会社内
		(72) 発明者	中村 篤
			東京都千代田区大手町二丁目3番1号 日
			本電信電話株式会社内
		Fターム(参考)	5D015 HH00

(54) 【発明の名称】 音響モデル作成装置、その装置を用いた音声認識装置、これらの方法、これらのプログラム、およびこれらの記録媒体

(57) 【要約】

【課題】 演算量を少なく、音響モデルを作成する。

【解決手段】 今回の適応用音声データの特徴量系列を抽出し、音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ、共分散行列パラメータに対するスケーリング因子、初期の共分散行列で表されることに基づき、前回のガウス分布の平均ベクトルパラメータの事後確率分布、今回まで累積された特徴量系列の一部を用い、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ及び、スケーリング因子を計算することで、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布を求め、今回の音響モデルパラメータの事後確率分布を新たな音響モデルパラメータに変換して更新する。

【選択図】 図3

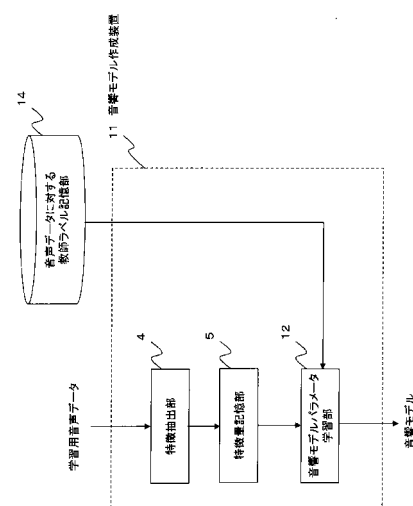


図3

【特許請求の範囲】

【請求項 1】

適応用音声データの特徴量系列を抽出する特徴抽出部と、

音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ、当該事後確率分布の共分散行列パラメータに対するスケーリング因子、初期音響モデルパラメータ中の共分散行列で表されることに基づき、前回までの累積された特徴量系列が加味された、前回求めた音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布及び今回まで累積された特徴量系列の一部を用いて、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ及び当該事後確率分布の共分散行列パラメータのスケーリング因子を計算することで、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布を求める逐次学習部と、

前記今回の音響モデルパラメータの事後確率分布を新たな音響モデルパラメータに変換して更新するモデル更新部と、を具備する音響モデル作成装置。

【請求項 2】

請求項 1 記載の音響モデル作成装置であって、

前記逐次学習部は、音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布の共分散ベクトルパラメータに対する今回のスケーリング因子 $r^{\wedge}_{t+1}$ を、前回のスケーリング因子 $r^{\wedge}_t$ 、今回の事後占有確率値の和 ζ_{t+1} とから求め、

音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布の今回の平均ベクトルパラメータ $\mu^{\wedge}_{t+1}$ を、前回の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトル v_{t+1} 、今回の事後占有確率値の和 ζ_{t+1} 、今回の各時点における ζ と特徴量との積和 M_{t+1} 、今回のスケーリング因子 $r^{\wedge}_{t+1}$ とから求めることを特徴とする音響モデル作成装置。

【請求項 3】

請求項 2 記載の音響モデル作成装置であって、

前記逐次学習部は、今回の音響モデルパラメータのスケーリング因子 $r^{\wedge}_{t+1}$ および今回の音響モデルパラメータの平均ベクトルパラメータ $\mu^{\wedge}_{t+1}$ を以下の式により求め、

$$r^{\wedge}_{t+1} = ((\mu^0)^{-1} + (r^{\wedge}_t)^{-1})^{-1} + \zeta_{t+1}$$

$$\mu^{\wedge}_{t+1} = \mu^{\wedge}_t + v_{t+1} + \frac{M_{t+1} - \zeta_{t+1}(\mu^{\wedge}_t + v_{t+1})}{r^{\wedge}_{t+1}}$$

ただし、 μ^0 は予め定められた定数であり、 $r^{\wedge}_t$ は前回のスケーリング因子であり、 ζ_{t+1} は今回の事後占有確率値の和であり、 M_{t+1} は今回の各時点における ζ と特徴量との積和であり、 v_{t+1} は今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトルであることを特徴とする音響モデル作成装置。

【請求項 4】

認識用音声データの音響的特徴を持つ適応用音声データに適応化させた音響モデルを、請求項 1～3 何れかに記載した音響モデル作成装置により作成して、音響モデルを更新する認識用モデル更新部と、

前記更新された音響モデルを用いて、前記音響的特徴を持った入力音声データに対する音声認識を行う認識部とを具備する音声認識装置。

【請求項 5】

適応用音声データの特徴量系列を抽出する特徴抽出過程と、

音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ、当該事後確率分布の共分散行列パラメータに対するスケーリング因子、初期音響モデルパラメータ中の共分散行列で表

されることに基づき、前回までの累積された特徴量系列が加味された、前回求めた音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布及び今回まで累積された特徴量系列の一部を用いて、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ及び当該事後確率分布の共分散行列パラメータのスケーリング因子を計算することで、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布を求める逐次学習過程と、

前記今回の音響モデルパラメータの事後確率分布を新たな音響モデルパラメータに変換して更新するモデル更新過程と、を有する音響モデル作成方法。

【請求項 6】

10

請求項 5 記載の音響モデル作成方法であって、

前記逐次学習過程は、音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布の共分散ベクトルパラメータに対する今回のスケーリング因子 $r^{\wedge}_{t+1}$ を、前回のスケーリング因子 $r^{\wedge}_t$ 、今回の事後占有確率値の和 ζ_{t+1} とから求め、

音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布の今回の平均ベクトルパラメータ $\mu^{\wedge}_{t+1}$ を、前回の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトル v_{t+1} 、今回の事後占有確率値の和 ζ_{t+1} 、今回の各時点における ζ と特徴量との積和 M_{t+1} 、今回のスケーリング因子 $r^{\wedge}_{t+1}$ とから求めることを特徴とする音響モデル作成方法。

【請求項 7】

20

請求項 6 記載の音響モデル作成方法であって、

前記逐次学習過程は、今回の音響モデルパラメータのスケーリング因子 $r^{\wedge}_{t+1}$ および今回の音響モデルパラメータの平均ベクトルパラメータ $\mu^{\wedge}_{t+1}$ を以下の式により求め、

【数 2】

$$r^{\wedge}_{t+1} = ((\mu^0)^{-1} + (r^{\wedge}_t)^{-1})^{-1} + \zeta_{t+1}$$

$$\mu^{\wedge}_{t+1} = \mu^{\wedge}_t + v_{t+1} + \frac{M_{t+1} - \zeta_{t+1}(\mu^{\wedge}_t + v_{t+1})}{r^{\wedge}_{t+1}}$$

ただし、 μ^0 は予め定められた定数であり、 $r^{\wedge}_t$ は前回のスケーリング因子であり、 ζ_{t+1} は今回の事後占有確率値の和であり、 M_{t+1} は今回の各時点における ζ と特徴量との積和であり、 v_{t+1} は今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトルであることを特徴とする音響モデル作成方法。

30

【請求項 8】

認識用音声データの音響的特徴を持つ適応用音声データに適応化させた音響モデルを、請求項 5～7 何れかに記載した音響モデル作成方法により作成して、音響モデルを更新する認識用モデル更新過程と、

前記更新された音響モデルを用いて、前記音響的特徴を持った入力音声データに対する音声認識を行う認識過程とを有する音声認識方法。

【請求項 9】

40

請求項 1～3 何れかに記載の音響モデル作成装置または、請求項 4 記載の音声認識装置としてコンピュータを動作させるプログラム。

【請求項 10】

請求項 9 記載のプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、音声認識時に使用する音響モデルを逐次適応法により作成する音響モデル作成装置、および、その装置で作成された音響モデルを用いて音声認識を行う音声認識装置、これらの方法、これらのプログラム、およびこれらの記録媒体に関する。

50

【背景技術】

【0002】

〔音声認識〕

従来の音声認識装置の機能構成例を図1に示し、従来の音声認識装置の処理の主な流れを図2に示す。音声認識装置2は主に、特徴抽出部4と単語列探索部6と音響モデル記憶部8と言語モデル記憶部10とで構成されている。

【0003】

まず、音響モデル記憶部8中の音響モデルの読み込みを行う（ステップS2）。なお、場合によっては、音響モデルの他に、単語モデル、コンテキスト依存音素モデル等の読み込みを行う。また、言語モデル記憶部10の読み込みを行う（ステップS4）。入力された認識用音声データは音声認識装置2に読み込まれ（ステップS6）、認識用音声データは特徴抽出部4に入力され、認識用音声データはフレーム（一定時間区間）ごとにMFCC（メルフィルタバンクケプストラム係数）ベクトルなどの音響的特徴量系列（以下、「特徴量系列」という。）に変換される（ステップS8）。変換された特徴量系列は図に示していないが、一旦、特徴量記憶部に記憶される。記憶された特徴量系列は、読み出されて、単語列探索部6に入力される。

【0004】

単語列探索部6では、音響モデル記憶部8の音響モデルを用いて認識用音声データの特徴量系列に対しスコアを算出し、これに言語モデル記憶部10の言語モデル等に対するスコアを参照して単語列探索を行う（ステップS10）。また、場合によっては、音素列探索や孤立単語探索を行う。最終的に認識結果を単語列として出力し（ステップS12）、場合によっては、音素列、孤立単語のみを出力する。

【0005】

〔音響モデル作成〕

次に、音響モデルの作成方法について説明する。音響モデルは、音声の音響的特徴をモデル化したものであり、認識用音声データと音響モデルを参照することにより、音声データを音素や単語といったシンボルに変換する。そのため、音響モデルの作成は、音声認識装置の性能を大きく左右する。通常、音声認識用音響モデルでは、各音素をLeft to rightの隠れマルコフモデル（HMM）で、HMM状態の出力確率分布を混合ガウス分布モデル（GMM）で表現する。そのため、実際に音響モデルとして記憶部に記憶されているのは、音素などの各シンボルにおける、HMMの状態遷移確率 a 、GMMの混合重み因子 w 、及び音響モデル中のガウス分布の平均ベクトルパラメータ μ 、及び音響モデル中のガウス分布の共分散行列パラメータ Σ となる。これらを音響モデルパラメータと呼びその集合を θ とする。つまり、 $\theta = \{a, w, \mu, \Sigma\}$ とする。音響モデルパラメータ θ の値を正確に求めるのが音響モデルの作成過程となり、この過程を音響モデル作成方法と呼ぶ。

【0006】

近年、音響モデルは確率統計的手法により大量の音声データとその教師ラベルの情報から、音響モデルパラメータ θ を学習することにより作成される。通常学習データに対しては、その何れの部分は何れの音素であるかを示す教師ラベル情報が与えられている。教師ラベル情報が与えられていない場合は、実際人が聞いて教師ラベル情報を付けたり、また音声認識装置を用いることにより教師ラベル情報を付与する。以降では学習用音声データには教師ラベル情報が前記のような方法で付与されているとして説明を進める。

【0007】

従来の音響モデル作成装置の機能構成例を図3に示し、従来の音響モデル作成装置の処理の主な流れを図4に示す。図3及び図4において、教師ラベル情報の付与については省略する。

【0008】

音響モデル作成装置11は、特徴抽出部4、特徴量記憶部5、音響モデルパラメータ学習部12、とで構成されている。学習用音声データが音響モデル作成装置11により読み込まれる（ステップS22）。読み込まれた学習用音声データは、特徴抽出部4で特徴量

系列に変換される（ステップ S 2 4）。変換された特徴量系列は特徴量記憶部 5 に記憶される。記憶された特徴量系列は読み出されて、音響モデルパラメータ学習部 1 2 に入力される。教師ラベルが存在していなければ（ステップ S 2 6）、音声認識装置、若しくは人手によって教師ラベル情報が与えられる（ステップ S 2 8）

【0009】

次に、音響モデルパラメータ学習部 1 2 による音響モデルパラメータの学習について説明する。教師ラベル情報により得られる学習データ中の各音素に対応するデータから、音響モデルパラメータ θ （HMM の状態遷移確率 a 、GMM の混合重み因子 w 、及び GMM の平均ベクトルパラメータ μ 及び共分散行列パラメータ Σ ）を推定することを音響モデルパラメータの学習と呼ぶ。パラメータを学習する手法としては最尤学習法がある。また、

10

【0010】

音響モデルパラメータ学習部 1 2 は、教師ラベル記憶部 1 4 に予め用意された音声データに対応する教師ラベル情報を用いて、音響モデルパラメータの学習を行う（ステップ S 3 0）。音響モデル作成装置 1 1 で作成された音響モデルが出力される（ステップ S 3 2）。また、ステップ S 2 6 において、教師ラベルが存在していれば、直接ステップ S 3 0 に進む。

【0011】

音響モデルパラメータは数百万の自由度を持つため、これらを学習するためには数百時間に及ぶ大量の学習用音声データが必要となる。しかし、事前に話者、雑音、発話スタイルといった全ての音響的変動要因を含む音声データを数百万のパラメータを十分に学習するほど、大量に収集するのは不可能である。そこで、少量の学習用音声データから音響モデルパラメータを推定する手法として、適応学習が非常に重要な技術となる。

20

【0012】

〔音響モデルパラメータの変換にもとづく適応学習〕

音響モデルパラメータに対しての適応学習は、パラメータあたりの学習データ量が少ない場合に初期モデルを先験知識として用い、少ないデータで学習を行う手法である。通常の学習方法との違いは学習データのみならず初期モデルを用いて音響モデルを構築する点である。このように初期モデルと学習データから新たに音響モデルを構築する学習方法を

30

【0013】

適応学習では、一般的に初期音響モデルパラメータ θ_0 と新たに作られる音響モデルパラメータ θ の変換に注目する。例えば、N 個の D 次元特徴ベクトルで表現される特徴量系列 $O = \{o_1, o_2, \dots, o_N \mid o_n \in \mathbb{R}^D\}$ が与えられたとき、音響モデルパラメータ θ の推定を考えるのではなく、その変換パラメータを考えるのが変換パラメータ推定法である。つまり初期モデルのパラメータ θ_0 と特徴量系列 O から適応後の音響モデルパラメータ θ を $\theta = f(\theta_0, O)$ として求めるときの、 $f(\cdot)$ を求め、それにより新たに音響モデルパラメータ θ を得る手法である。

【0014】

$f(\cdot)$ がパラメトリックに表現される（関係が数式で表現される。）とすれば、適応学習は $f(\cdot)$ のパラメータである変換パラメータ W （後に具体的に説明する）の推定を初期モデルパラメータ θ_0 と適応用音声データ O から求めることになる。これを音響モデルパラメータの変換にもとづく適応学習と呼ぶ。

40

【0015】

〔線形回帰法〕

適応学習の中では、音響モデル中のガウス分布の平均ベクトルパラメータ μ に対する線形回帰行列を推定する手法が広く用いられている（非特許文献 1、2 参照）。線形回帰行列を用いた場合の音響モデル作成装置の機能構成例を図 5 に示し、この場合の音響モデル作成装置の主な処理の流れを図 6 に示す。この手法を用いた音響モデル作成装置 2 1 は、

50

特徴抽出部 4、特徴量記憶部 5、パラメータ適応部 22、とで構成されており、パラメータ適応部 22 は変換パラメータ推定部 24、変換パラメータ記憶部 26、モデルパラメータ変換部 28、とで構成されている。

【0016】

まず、初期音響モデルパラメータ θ_0 が初期音響モデルパラメータ記憶部 30 に読み込まれる（ステップ S40）。そして、適応音声データ 20 が読み込まれ（ステップ S42）、特徴抽出部 4 に入力され、特徴量系列 O に変換される（ステップ S44）。変換された特徴量系列 O は一旦、特徴量記憶部 5 に記憶される。記憶された特徴量系列 O は変換パラメータ推定部 24 に入力される。以下に変換パラメータ推定部 24、モデルパラメータ変換部 28 の処理を説明する。

10

【0017】

初期音響モデルパラメータ θ_0 中のあるガウス分布の平均ベクトルパラメータ μ_0 は以下の式（1）により線形変換される。

$$\mu = A \mu_0 + v \quad (1)$$

ここで、A は $D \times D$ の行列であり、平均ベクトルパラメータ μ_0 の回転、伸縮をさせる行列である。v は D 次元ベクトルであり平均ベクトルパラメータ μ_0 の平行移動をさせるベクトルを表す。このとき、変換パラメータ $W = (v, A)$ である。変換パラメータ W は特徴量系列 O から期待値最大化（Expectation Maximization）アルゴリズム（以下 EM アルゴリズムという）やその一種である MLLR（Maximum Likelihood Linear Regression）アルゴリズムを用いて繰り返し計算により効率よく求められる（ステップ S46）。推定すべき変換パラメータ W のパラメータ数は $D^2 + D = D(D+1)$ となる。何故なら、行列 A の要素数は D^2 であり、ベクトル v の要素数は D であるからである。平均ベクトルのパラメータ数 D よりもパラメータ数が多いが、複数のガウス分布で同一の変換パラメータを共有することにより、推定すべきパラメータ数を減らすことが可能である。推定された変換パラメータ W は一旦変換パラメータ記憶部 26 に記憶される。

20

【0018】

記憶された変換パラメータ W はモデルパラメータ変換部 28 に入力される。モデルパラメータ変換部 28 で、得られた変換パラメータ W と（初期音響モデルパラメータ θ_0 中の）初期平均ベクトルパラメータ μ_0 をもとに前記式（1）から新たな平均ベクトルパラメータ μ を得る（ステップ S48）。平均ベクトルパラメータ μ が音響モデルパラメータ μ として出力される（ステップ S50）。

30

【0019】

[音響モデルパラメータ変換にもとづく逐次適応]

以上までは、一まとまりの特徴量系列 $O = \{o_1, o_2, \dots, o_n, \dots, o_N\}$ （ただし、N はフレーム数である）に対しての適応学習を考えた。しかし、音声は雑音などの外的要因や発声のなまり等の内的要因によって、時々刻々その音響的特徴を大きく変化させている。このような変化に追従していくためには、時系列的に与えられるまとまった量の音声データに対して逐次モデルを適応させる逐次適応学習が有効である。このとき、特徴量系列を 1 まとまりとして捉えず、複数のまとまりが時系列的に与えられる場合の適応を考える。つまり以下の式（2）（3）のように考える。

40

【数 3】

$$O = \underbrace{\{o_1, o_2, \dots, o_{N_1}\}}_{O_1}, \underbrace{\{o_{N_1+1}, o_{N_1+2}, \dots, o_{N_1+N_2}\}}_{O_2}, \dots, \underbrace{\{o_{N_i+1}, o_{N_i+2}, \dots, o_{N_i+N_{i+1}}\}}_{O_i}, \dots \quad (2)$$

$$O = \{O_1, O_2, \dots, O_i, \dots, O_T\} \quad (3)$$

【0020】

ただし、t は前回の時刻、t+1 は今回の時刻を示し、式（3）中の T は t の総数を示し、 θ_{t+1} および θ_t は今回および前回の音響モデルパラメータである。このとき、あるまとまり t+1 での音響モデルパラメータ θ_{t+1} は、その前のまとまり t において得

50

られた音響モデルパラメータ θ_t 及び特徴量系列のまとまり O_{t+1} から求められる。つまり、以下の式 (4) に示す漸化式で表現することにより、時々刻々音響モデルを求めることができる。これをパラメータ変換に基づく逐次適応法と呼ぶ。

$$\theta_{t+1} = f(\theta_t, O_{t+1}) \quad (4)$$

【0021】

図7に、逐次適応法を用いた場合の音響モデルパラメータが変換される手順を示す。まず、特徴量系列 O_1 と初期音響モデルパラメータ θ_0 を用いてモデルパラメータ変換部28で音響モデルパラメータ θ_1 が求められる。そして、今度は、音響モデルパラメータ θ_1 と次の特徴量系列 O_2 と用いて、音響モデルパラメータ θ_2 が求められる。このようにして、前回の音響モデルパラメータ θ_t と今回の特徴量系列 O_{t+1} とを用いて、今回の音響モデルパラメータ θ_{t+1} が求められる。

10

【0022】

[線形回帰法]

このとき、変換パラメータ推定法の逐次適応への適用を考察する(非特許文献2参照)。先ほどは、変換パラメータ W は全ての特徴量系列から推定されたとしたが、逐次適応においては各まとまりごと (t ごと) に W を推定する。それを $W_t = \{v_t, A_t\}$ とすれば、パラメータ変換に基づく逐次適応法における平均パラメータの更新式(前記式(4)に示す)は前記式(1)を基に、以下の式(5)のように漸化式で表現することができる。

$$\mu_{t+1} = A_{t+1} \mu_t + v_{t+1} \quad (5)$$

20

これによって、パラメータ変換に基づく逐次適応が実現される。以下の説明では、 A_{t+1} は「今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数行列」といい、 v_{t+1} は「今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトル」という。

【0023】

以上の逐次適応法は得られた音響モデルパラメータ $\theta_0, \dots, \theta_{t+1}$ にどの程度推定による誤差が含まれるかが考慮されていない。そのため、学習に悪影響を及ぼすような音声データが存在した場合、学習が失敗した場合等は、その影響がそのまま認識性能に出てしまい、頑健性が低いものになってしまう。

【0024】

30

[分布変換にもとづく逐次適応法]

次に、本発明の基本概念となる「分布変換にもとづく逐次適応法」について説明する。本手法では、音響モデルパラメータ θ_t そのものの推定を考えるのではなく、音響モデルパラメータの分布 $p(\theta_t)$ を考える(特許文献1、非特許文献3、4参照)。

【0025】

これにより、推定による誤差を例えばその分布の分散から考慮することができる。さらに音響モデルパラメータの分布として累積された特徴量系列 $O^t = \{O_1, O_2, \dots, O_t\}$ が与えられたときの事後確率分布を考える。つまり、 $p(\theta_t)$ ではなく、 $p(\theta_t | O^t)$ を推定対象とする。ここで、 O^{t+1} および O^t は今回および前回までに累積された特徴量系列であることを示す。

40

【0026】

ここで、 $p(\alpha | \beta)$ はある事象 β が起こるという条件下で、別の事象 α が起こる確率である事後確率(条件付き確率)である。つまり、 $p(\theta_t | O^t)$ は特徴量系列 O^t が与えられた時の音響モデルパラメータが θ_t である事後確率であることを示す。これにより、累積された特徴量系列 O^t の情報を音響モデルパラメータに加味することができるため、頑健性を確保することができる。従って、以下の式(8)に示す漸化式

$$p(\theta_{t+1} | O^{t+1}) = F[p(\theta_t | O^t)] \quad (8)$$

を用いて時間発展、つまり、音声の音響的特徴の変化として対応した漸化式を記述することにより、前記式(4)で注目した音響モデルパラメータ θ ではなく、音響モデルパラメータの事後確率分布 $p(\theta | O)$ に基づく逐次適応を実現することができる。ここで、 $F[$

50

・]は $p(\theta | O)$ を引数として持つ汎関数である。また、 $F[\cdot]$ は今回まで累積された特徴量系列 O^{t+1} の一部の特徴量系列に基づいて表現されるものである。以下の説明では、 $F[\cdot]$ は、今回まで累積された特徴量系列 O^{t+1} に基づいて、表現されるものとする。このとき、 $F[\cdot]$ をパラメトリックに表現し、その変換パラメータ W を例えば特徴量 O_t から適切に推定することにより前記式(8)で表現される逐次適応を実現できる。ただし、変換パラメータの推定は、特徴量 O_t のみではなく、特徴量系列 $O_1, O_2, \dots, O_t$ のうちの一部を用いてもよく、特徴量系列 O^t を用いてもよい。

【0027】

前記式(4)と前記式(8)を見比べてわかることは、前記式(8)はパラメータを逐次変換させるのではなく、その事後確率分布を逐次変換させていることである。また、時刻 t での事後確率分布 $p(\theta_t | O^t)$ のパラメータを ω_t とすると、 $p(\theta_t | O^t)$ の逐次更新はパラメータ ω_t の逐次更新で表現できる。従って、時々刻々、事後確率分布パラメータ ω_t を求めることにより逐次適応が実現できる。従って、分布変換にもとづく逐次適応法では、事後確率分布 $p(\theta_t | O^t)$ ではなく、事後確率分布パラメータ ω_t を更新していく。

【0028】

図8に当該逐次的応法を適用した場合の機能構成例を示し、図9に事後確率分布のパラメータ ω を逐次適応させる順序を示し、図10に主な処理の流れを示す。図8に示す音響モデル作成装置48は、特徴抽出部4、特徴量記憶部5、モデル適応化部50、とで構成され、モデル適応化部50は逐次学習部52、事後確率分布記憶部54、モデル更新部56、とで構成されている。

【0029】

まず、前回の事後確率分布のパラメータ ω_t がモデル適応化部50で読み込まれる(ステップS60)。次に、適応用音声データが読み込まれ(ステップS62)、適応用音声データが特徴抽出部4に入力され、特徴量系列 O_{t+1} に変換される(ステップS64)。変換された特徴量系列 O_{t+1} は一旦、特徴量記憶部5に記憶され逐次学習部52に入力される。

【0030】

逐次学習部52では、前記式(8)のように、前回までの累積された特徴量系列が加味された前回求めた音響モデルパラメータの事後確率分布 $p(\theta_t | O^t)$ と、前記今回抽出した特徴量系列 O_{t+1} とに基づき、今回の特徴量系列に適応化させた今回の音響モデルパラメータの事後確率分布 $p(\theta_{t+1} | O^{t+1})$ を求める(ステップS68)。以下に、逐次学習部52による更に具体的な事後確率分布 $p(\theta_{t+1} | O^{t+1})$ の求め方を説明する。

【0031】

$p(\theta_t | O^t)$ から $p(\theta_{t+1} | O^{t+1})$ への時間発展を記述する前記式(8)中の関数 $F[\cdot]$ には任意の形を与えることが可能であり様々な変換を考えることができる。この実施形態では、具体的な関数系のひとつとして、確率の積の公式とベイズの定理から理論的に近似無く導出される漸化式を紹介する。はじめに $p(\theta_{t+1} | O^{t+1})$ はベイズの定理から次のように表現される。

【数4】

$$p(\theta_{t+1} | O^{t+1}) = \frac{p(O_{t+1} | \theta_{t+1}, O^t) p(\theta_{t+1} | O^t)}{p(O_{t+1} | O^t)} \quad (9)$$

ここで式(9)の右辺にある $p(\theta_{t+1} | O^t)$ は $p(\theta_t | O^t)$ を用いると次のように表現される。

$$p(\theta_{t+1} | O^t) = \int p(\theta_{t+1} | \theta_t, O^t) p(\theta_t | O^t) d\theta_t \quad (10)$$

従って式(10)を式(9)に代入することにより次式(11)のような漸化式を導出することができる。

10

20

30

40

【数 5】

$$p(\theta_{t+1} | O^{t+1}) = \frac{p(O_{t+1} | \theta_{t+1}, O^t) \int p(\theta_{t+1} | \theta_t, O^t) p(\theta_t | O^t) d\theta_t}{p(O_{t+1} | O^t)} \quad (11)$$

【0032】

式(11)の右辺には前回(時刻 t)での事後確率分布 $p(\theta_t | O^t)$ が含まれており、 $p(\theta_t | O^t)$ から現在(次の時刻 $t+1$)での事後確率 $p(\theta_{t+1} | O^{t+1})$ を求める式となっている。従って、式(11)を音響モデルパラメータの事後確率分布の漸化式と呼ぶ。この漸化式を用いることにより、前回まで累積された、特徴量系列 O^{t+1} の情報が加味された音響モデルパラメータの事後確率分布 $p(\theta_{t+1} | O^{t+1})$ の逐次推定を逐次学習部52で行うことが出来る。また式(11)の積分計算はモンテカルロ法(Monte Carlo method)などの数値計算で解くことが出来る。またここで時間発展を最初の一ステップに限る。つまり $t \rightarrow 0$, $t+1 \rightarrow 1$ とすれば以下の式(12)のようになる。

【数 6】

$$p(\theta_1 | O_1) = \frac{p(O_1 | \theta_1) \int p(\theta_1 | \theta_0) p(\theta_0) d\theta_0}{p(O_1)} \quad (12)$$

式(12)は、逐次的ではなく、与えられた適応用データ O_1 から $p(\theta_1 | O_1)$ を推定する通常の適応を示している。つまり、本発明は逐次適応のみならず通常の適応においてもその効果を与えることができる。

【0033】

前記式(11)による逐次適応を実現するためには、右辺は次の4つの確率分布である $p(O_{t+1} | O^t)$ 、 $p(\theta_t | O^t)$ 、 $p(O_{t+1} | \theta_{t+1}, O^t)$ 、 $p(\theta_{t+1} | \theta_t, O^t)$ に具体系を与える必要がある。ここで $p(O_{t+1} | O^t)$ は求めたい分布である $p(\theta_{t+1} | O^{t+1})$ の引数 θ_{t+1} に依存しないため、規格化定数として扱うことができるため、具体形を与えなくても良い。残りの3つである $p(\theta_t | O^t)$ 、 $p(O_{t+1} | \theta_{t+1}, O^t)$ 、 $p(\theta_{t+1} | \theta_t, O^t)$ について考察を行う。

【0034】

$p(\theta_t | O^t)$ は前述した音響モデルパラメータの事後確率分布であり、適切に初期分布を設定することにより逐次求めることが可能である。 $p(O_{t+1} | \theta_{t+1}, O^t)$ は O_{t+1} の出力分布であり、HMMやGMMといった音響モデルの設定によって与えられるものである。最後に $p(\theta_{t+1} | \theta_t, O^t)$ は音響モデルパラメータ θ の確率的ダイナミクスである。従って、前記式(11)の漸化式は、初期分布、出力分布及び確率的ダイナミクスによって構成されている。

【0035】

説明を図8に戻すと、逐次学習部52により求められた今回の事後確率分布 $p(\theta_{t+1} | O^{t+1})$ は一旦、事後確率分布記憶部54に記憶される。そして今回の事後確率分布 $p(\theta_{t+1} | O^{t+1})$ はモデル更新部56に入力される。

モデル更新部56で、音響モデル記憶部58内の音響モデルとしての前回の事後確率分布 $p(\theta_t | O^t)$ が、今回の音響モデルパラメータの事後確率分布 $p(\theta_{t+1} | O^{t+1})$ に新たな音響モデルとして更新する(ステップS70)。

【0036】

また、図9について説明すると、求められた前回の事後確率分布 $p(\theta_t | O^t)$ は一旦、音響モデル(分布モデル)記憶部58に一旦、記憶される。逐次学習部52で、前回の事後確率分布 $p(\theta_t | O^t)$ と、今回の特徴量系列 O^{t+1} とを用いて、前記式(11)から、今回の事後確率分布 $p(\theta_{t+1} | O^{t+1})$ を求める。このようにして、音響モデルパラメータの事後確率分布を逐次的に更新する。

【0037】

[マルコフ過程の導入]

次に前記式(11)の演算処理をマルコフ過程を仮定することで簡単にする手法を説明

する。 $p(O_{t+1} | \theta_{t+1}, O^t)$ 及び $p(\theta_{t+1} | \theta_t, O^t)$ は累積された特徴量系列に直接依存する。これらを全ての累積特徴量系列から推定しようとした場合、時が経つにつれ累積データは多くなるため、その推定は大変計算量が多くなり現実的でない。そこで、マルコフ過程を仮定すると、 $p(O_{t+1} | \theta_{t+1}, O^t)$ と $p(\theta_{t+1} | \theta_t, O^t)$ はそれぞれ式(13)のように近似される。

$$\frac{p(O_{t+1} | \theta_{t+1}, O^t)}{p(\theta_{t+1} | \theta_t, O^t)} = \frac{p(O_{t+1} | \theta_{t+1})}{p(\theta_{t+1} | \theta_t)} \quad (13)$$

【0038】

この近似により、逐次学習部52は前回の音響モデルパラメータの事後確率分布 $p(\theta_t | O^t)$ と、今回の出力分布 $p(O_{t+1} | \theta_{t+1})$ と、今回の確率的ダイナミクス $p(\theta_{t+1} | \theta_t)$ と、を用いて今回の音響モデルパラメータの事後確率分布 $p(\theta_{t+1} | O^{t+1})$ を求める。具体的には以下の式(14)のように近似される。

$$p(\theta_{t+1} | O^{t+1}) \propto p(O_{t+1} | \theta_{t+1}) \int p(\theta_{t+1} | \theta_t) p(\theta_t | O^t) d\theta_t \quad (14)$$

ここで、 $A \propto B$ はAとBは比例しているということを表す。前記式(14)によって、シンプルな出力分布及び確率的ダイナミクスを設定することができる。図8中の逐次学習部52は、この式(14)を計算することになる。

【0039】

[ガウス分布の平均ベクトルの考察]

以上の議論では、HMMの状態遷移確率 a 、GMMの混合重み因子 w 、及びガウス分布の平均ベクトルパラメータ μ 及び共分散行列パラメータ Σ といった全ての音響モデルパラメータ θ の事後確率分布 $p(\theta | O)$ についての処理を行った。一般に、音響モデルにおいて最も性能を左右するパラメータはガウス分布の平均ベクトルパラメータ μ であり、またそれ以外のパラメータの事後確率分布を推定対象とした場合、分布変換関数 F の推定すべきパラメータ数が多くなるため、少量データ適応において効果が十分でなくなる。そのため、以降ではガウス分布の平均ベクトルパラメータ μ のみに焦点を当て、つまり、音響モデルパラメータ θ に代えて、ガウス分布の平均ベクトルパラメータ μ を用いて、図8の逐次学習部52では演算する。演算された事後確率分布 $p(\mu | O)$ の時間発展について考察する。つまり、前記式(14)においてガウス分布の平均ベクトルパラメータ μ のみを考えるため時間発展は次式(15)を逐次学習部52で演算する。

$$p(\mu_{t+1} | O^{t+1}) \propto p(O_{t+1} | \mu_{t+1}) \int p(\mu_{t+1} | \mu_t) p(\mu_t | O^t) d\mu_t \quad (15)$$

なお、式(15)は音響モデル中の各ガウス分布の平均ベクトルパラメータに独立に与えられる。その際の各ガウス分布のインデックスは文中では省略する。

【0040】

[線形ダイナミクス]

次に、前記式(15)の解析解を導出することを考える。これを用いて、逐次学習を行う。式(15)にはさまざまな解析解が存在するが、最も単純な解析解として確率的ダイナミクスが線形で表現される場合を考える。つまり、確率的ダイナミクスとして、以下の式(16)を仮定することが出来る。

$$\mu_{t+1} = A_{t+1} \mu_t + v_{t+1} + \varepsilon_{t+1} \quad (16)$$

【0041】

ここで ε_{t+1} は平均0、共分散行列 U のガウシアンノイズである。式(16)は、前記式(5)における線形変換が確率的に揺らいでいるといえる。このとき、確率ダイナミクスの分布具体系は、以下の式(17)として与えられる。

$$p(\mu_{t+1} | \mu_t) = N(\mu_{t+1} | A_{t+1} \mu_t + v_{t+1}, U) \quad (17)$$

【0042】

ここで式(17)の $N(x | m, S)$ は、 x を引き数とする平均パラメータ m 、共分散行列パラメータ S のガウス分布である。さらに通常のHMM、GMMで表現される音響モデルに対して一まとまりの特徴量系列 $O_t = \{o_{N_t+1}, \dots, o_{N_t+N_{t+1}}\}$ が出力

される出力分布 $p(O_t | \mu_t)$ は以下の式 (18) で表すことができる。

【数 7】

$$p(O_{t+1} | \mu_{t+1}) = \prod_{n=N_{t+1}}^{N_{t+1}+N_{t+2}} (N(o_n | \mu_{t+1}, \Sigma))^{\zeta_n} \quad (18)$$

【0043】

ここで、 ζ_n は、対象のガウス分布に割り当てられた O_n の事後占有確率値である。また、状態遷移確率 a および混合重み因子 w は $p(\mu | O)$ の推定に関係ないため無視した。また HMM や GMM の潜在変数は無視したが、これらは EM アルゴリズム（期待値最大化アルゴリズム）を用いることによって対処可能である。実際、式 (18) は EM アルゴリズムにおける補助関数の形式で表現されている。 10

【0044】

最後に音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布 $p(\mu_t | O^t)$ がガウス分布で表現されると仮定し、事後確率分布 $p(\mu_t | O^t)$ の平均ベクトルパラメータを $\mu^{\wedge}_t$ とし、事後確率分布 $p(\mu_t | O^t)$ の共分散行列パラメータが $Q^{\wedge}_t$ で表現されたとすると関数形は以下の式 (19) で表すことができる。

$$p(\mu_t | O^t) = N(\mu_t | \mu^{\wedge}_t, Q^{\wedge}_t) \quad (19)$$

【0045】

従って、式 (17)、(18)、及び (19) を式 (15) に代入することにより以下の式 (20) で示される解析解を導出することができる。 20

$$p(\mu_{t+1} | O^{t+1}) = N(\mu_{t+1} | \mu^{\wedge}_{t+1}, Q^{\wedge}_{t+1}) \quad (20)$$

ここで、

$$Q^{\wedge}_{t+1} = ((U + A_{t+1} Q^{\wedge}_t A_{t+1}')^{-1} + \zeta_{t+1} \Sigma^{-1})^{-1} \quad (21)$$

$$K^{\wedge}_{t+1} = Q^{\wedge}_{t+1} \zeta_{t+1} \Sigma^{-1} \quad (22)$$

$$\mu^{\wedge}_{t+1} = A_{t+1} \mu^{\wedge}_t + v_{t+1} + K^{\wedge}_{t+1} (M_{t+1} / \zeta_{t+1} - A_{t+1} \mu^{\wedge}_t - v_{t+1}) \quad (23)$$

【0046】

ただし、 $\mu^{\wedge}_{t+1}$ は今回の事後確率分布 $p(\mu_{t+1} | O^{t+1})$ をガウス分布で表現した際の平均ベクトルパラメータであり、 $Q^{\wedge}_{t+1}$ は今回の事後確率分布 $p(\mu_{t+1} | O^{t+1})$ をガウス分布で表現した際の共分散行列パラメータであり、 $K^{\wedge}_{t+1}$ はカルマンゲインであり、 A_{t+1} 、 v_{t+1} 、及び U はそれぞれ音響モデルパラメータ中の平均の前記確率的ダイナミクスを線形表現した時の係数行列、係数ベクトル、及びガウシアンノイズの共分散行列であり、 Σ は初期音響モデルパラメータ中の共分散行列であり、 A_{t+1}' は行列 A_{t+1} の転置を表す。 ζ_{t+1} は、今回の事後占有確率値の和であり、 M_{t+1} は今回の各時点における ζ と特徴量との積和であり、 ζ_t 、 M_t はガウス分布の平均ベクトルパラメータの十分統計量であり以下の式 (24) のように定義される。 30

【数 8】

$$\begin{cases} \zeta_t \equiv \sum_{n=N_t+1}^{N_t+N_{t+1}} \zeta_n \\ M_t \equiv \sum_{n=N_t+1}^{N_t+N_{t+1}} \zeta_n o_n \end{cases} \quad (24)$$

40

【0047】

ζ_t 、 M_t は Forward-backward アルゴリズムや Viterbi アルゴリズム、`kmeans` 法などのアライメント手法によって効率よく求めることができる。式 (21) ~ (23) の更新を音響モデル中の全てのガウス分布に対して行うことにより、全ての事後確率分布を更新することができる。

【0048】

このようにして、今回の事後確率分布 $p(\mu_{t+1} | O^{t+1})$ の分布パラメータ ω_{t+1} は $Q^{\wedge}_{t+1}$ 、 $\mu^{\wedge}_{t+1}$ となり（図 9 の音響モデルパラメータ記憶部 58 の括弧書 50

き参照)、式(21)(22)(23)から求めることができる。なお、式(22)における $K^{\wedge}_{t+1}$ は数式の表現のしやすさのために導入したものである。実際の計算では、式(22)と(23)を同時に行っても良い。その場合、 $K^{\wedge}_{t+1}$ は求める必要がない。

【0049】

つまり、音響モデルパラメータの事後確率分布 $p(\mu | O)$ の漸化式はそのパラメータ $(Q^{\wedge}, \mu^{\wedge})$ の漸化式(21)(22)(23)によって求めることができる。これは線形動的システムにおけるカルマンフィルタの解と類似している。しかし、カルマンフィルタの解は $o_n \rightarrow o_{n+1}$ のように各音声分析フレームごとの更新となっている。一方、本発明では $O_t \rightarrow O_{t+1}$ のように1まとまりのフレームごとの更新となっているのが違いとなっている。そのため、パラメータ $Q^{\wedge}, \mu^{\wedge}$ は、1フレームの特徴量 o_n ではなく、その統計量であらわされている。従って、これを巨視的な線形動的システムと呼ぶ。

【0050】

$Q^{\wedge}, \mu^{\wedge}$ を用いた場合の逐次学習部52の具体的構成例を図11に示す。逐次学習部52は、 $Q^{\wedge}$ 更新部520、 $K^{\wedge}$ 更新部522、 $\mu^{\wedge}$ 更新部524、事後確率計算部526とで構成されている。

$Q^{\wedge}$ 更新部520では前記式(21)が計算され、 $K^{\wedge}$ 更新部522では前記式(22)が計算され、 $\mu^{\wedge}$ 更新部524では前記式(23)が計算され、事後確率計算部526では前記式(20)が計算される。

【0051】

従って、 $Q^{\wedge}, \mu^{\wedge}$ を求めるためには、線形変換パラメータ $W_{t+1} = \{v_{t+1}, A_{t+1}\}$ 、システムノイズ U_{t+1} 、初期パラメータ $Q^{\wedge}_0$ 、及び $\mu^{\wedge}_0$ の4つを設定する必要がある。ここで、 $Q^{\wedge}_0$ は初期音響モデルの共分散行列パラメータから与えられるものであり、 $\mu^{\wedge}_0$ は初期音響モデルの平均ベクトルパラメータから与えられるものである。

【0052】

このうち線形変換パラメータ $W = \{v_{t+1}, A_{t+1}\}$ は、今回まで累積された特徴量系列 O^t のうち少なくとも1つの特徴量系列を用いて、推定される。よく知られた手法の一例としては上述したEMアルゴリズムやMLLRアルゴリズムを用いて繰り返し計算により効率よく求められる。また、複数のガウス分布で同一の変換パラメータを共有することにより、推定すべきパラメータ数を減らすことが可能である。

【0053】

システムノイズ U も線形変換パラメータ W と同様に学習によって求めることができる。または、行列成分すべてを特徴量系列やその他のデータから先験的に与えることもできる。最も単純な方法は、システムノイズ U を $(u^0)^{-1} \Sigma$ としておき、システムノイズの共分散行列が出力分布の共分散行列と比例関係にあるとすると、 u^0 を予め与えられるパラメータとする。つまり、1つだけパラメータが導入される。システムノイズ U と線形変換パラメータ W が、前記式(8)の分布変換関数 F における変換パラメータとなる。

【0054】

このとき更新式は、以下の式(25)(26)(27)で表され、 $Q^{\wedge}$ 更新部520では前記式(25)が計算され、 $K^{\wedge}$ 更新部522では前記式(26)が計算され、 $\mu^{\wedge}$ 更新部524では前記式(27)が計算される。

$$Q^{\wedge}_{t+1} = ((u^0)^{-1} \Sigma + A_{t+1} Q^{\wedge}_t A_{t+1}^T)^{-1} + \zeta_{t+1} \Sigma^{-1} \quad (25)$$

$$K^{\wedge}_{t+1} = Q^{\wedge}_{t+1} \zeta_{t+1} \Sigma^{-1} \quad (26)$$

$$\mu^{\wedge}_{t+1} = A_{t+1} \mu^{\wedge}_t + v_{t+1} + K^{\wedge}_{t+1} (M_{t+1} / \zeta_{t+1} - A_t \mu^{\wedge}_t - v_{t+1}) \quad (27)$$

以上によってパラメータ u^0 によって制御される分布変換にもとづく逐次適応法を実現できる。

【0055】

10

20

30

40

50

[平行移動適応]

前記線形ダイナミクスの式(16)の平均ベクトル μ_t の平行移動 v_{t+1} にだけ注目することにより、推定すべきパラメータを少なくしてより少量データでの適応を実現できる。このとき、前記式(25)(26)(27)における行列 A_{t+1} を単位行列 I とする、つまり、 $A_{t+1} = I$ とすると、 $Q^{\wedge}$ 、 $K^{\wedge}$ 、 $\mu^{\wedge}$ は以下の式(28)(29)(30)で計算される。

【0056】

$$Q^{\wedge}_{t+1} = ((u^0)^{-1} \Sigma + Q^{\wedge}_t)^{-1} + \zeta_{t+1} \Sigma^{-1} \quad (28)$$

$$K^{\wedge}_{t+1} = Q^{\wedge}_{t+1} \zeta_{t+1} \Sigma^{-1} \quad (29)$$

$$\mu^{\wedge}_{t+1} = \mu^{\wedge}_t + v_{t+1} + Q^{\wedge}_{t+1} \zeta_{t+1} \Sigma^{-1} (M_{t+1} / \zeta_{t+1} - \mu^{\wedge}_t - v_{t+1}) \quad (30)$$

10

この場合、 $Q^{\wedge}$ 更新部520では前記式(28)が計算され、 $K^{\wedge}$ 更新部522では前記式(29)が計算され、 $\mu^{\wedge}$ 更新部524では前記式(30)が計算される。

【非特許文献1】C.J.Leggetter and P.C.Woodland, Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Computer Speech and Language, Vol.9, pp.171-185, 1995.

【非特許文献2】C.J.Leggetter and P.C.Woodland, Maximum Flexible speaker using maximum likelihood linear regression In Proc ARPA Spoken Language Technology Workshop, pp.104-109, 1995.

【非特許文献3】渡部晋治、中村篤、確率分布の巨視的な時間発展システムに基づく逐次モデル適応。秋季音響学会講演論文集、2-2-10, pp. 71-72, 2006.

20

【非特許文献4】渡部晋治、中村篤、確率分布の巨視的な時間発展系に基づくモデル適応との従来型適応との関係の考察。秋季音響学会講演論文集2-3-12、2007。

【特許文献1】特開2008-64849号

【発明の開示】

【発明が解決しようとする課題】

【0057】

オンライン逐次適応タスクでは、1秒程度の非常に短い時間でのモデル更新が、実時間処理のためには必要である。しかし、分布変換に基づく適応法の更新は式(28)～(30)から分かるとおり、行列 $Q^{\wedge}_t$ や Σ の演算で表現されており、この行列の積、和および逆行列演算を全てのガウス分布について行う必要があり、計算コストが非常に高く、実時間処理が困難である。例えば、通常の音響モデルは39次元のガウス分布を数万個含むが、これに対して式(28)～(30)を実行することは 39×39 の行列 A 、 $Q^{\wedge}_t$ や Σ (ただし、 Σ は通常、対角行列を用いる)の積、和および逆行列をガウス分布数分(数万回)、行う必要があるため、非常に計算に時間がかかる。

30

【0058】

また、式(28)～(30)は逆行列の計算を含むため音声データによっては、計算が不安定になり逆行列が求まらなくなる。

また分布パラメータ $Q^{\wedge}_t$ はモデルの更新に必要なため、それらを音響モデルパラメータ記憶部58に記憶する必要がある。しかし、 $Q^{\wedge}_t$ は非対角成分が0でない全共分散行列(ただし、対称行列)であり、それが音響モデル中のガウス分布数分存在するため、大量のメモリを消費する。例えば、音響モデルは39次元のガウス分布数万個で表現される音響モデルが数メガバイト程度なのに対し、 $Q^{\wedge}_t$ だけで、音響モデルの10倍以上のメモリ(数10メガバイト)を消費する。

40

【0059】

このように、分布変換に基づく適応法は計算量が多く、計算が不安定であり、メモリを多く消費する。従って、それらを用いて、1秒程度の非常に短い逐次更新を行う、オンライン逐次適応タスクの実現をするのは困難であった。

【0060】

この発明では、従来と比べて、計算量、メモリ量を削減させ、計算の安定性を向上させ

50

た音響モデル作成装置、その装置を用いた音声認識装置、これらの方法、これらのプログラム、およびこれらの記録媒体を提供する。

【課題を解決するための手段】

【0061】

この発明の音響モデル作成装置は、特徴抽出部と、逐次学習部と、モデル更新部と、を具備する。特徴抽出部は、今回の適応用音声データの特徴量系列を抽出する。逐次学習部は、音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ、当該事後確率分布の共分散行列パラメータに対するスケーリング因子、初期音響モデルパラメータ中の共分散行列で表されることに基づき、前回までの累積された特徴量系列が加味された、前回求めた音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布及び今回まで累積された特徴量系列の一部を用いて、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、当該事後確率分布の平均ベクトルパラメータ及び当該事後確率分布の共分散行列パラメータのスケーリング因子を計算することで、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布を求める。モデル更新部は、今回の音響モデルパラメータの事後確率分布を新たな音響モデルパラメータに変換して更新する。

10

【発明の効果】

【0062】

この発明では、音響モデル中のあるガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ に対して、その共分散行列パラメータ $Q^{\wedge}_t$ を、対象とする音響モデルのガウス分布の共分散行列 Σ と、本発明で新たに導入するスカラー変数であるスケーリング因子 $r^{\wedge}_t$ の逆数 $(r^{\wedge}_t)^{-1}$ を用いて、以下の式のように掛け合わせたもので表現する。

20

$$Q^{\wedge}_t = (r^{\wedge}_t)^{-1} \Sigma \quad (31)$$

これにより、式(28)～(30)はスカラー演算に直すことができるため、計算量の削減および安定性の確保を実現することができる。また、記憶すべき更新パラメータが対称行列 $Q^{\wedge}_t$ から $r^{\wedge}_t$ となるため音響モデル記憶部中のメモリ容量を削減できる。

【発明を実施するための最良の形態】

【0063】

30

以下に、発明を実施するための最良の形態を示す。なお、同じ機能を持つ構成部や同じ処理を行う過程には同じ番号を付し、重複説明を省略する。

【実施例1】

【0064】

まず、改めて、記号について定義する。

μ_t 前回の音響モデル中のガウス分布の平均ベクトルパラメータ

Σ_t 前回の音響モデル中のガウス分布の共分散行列パラメータ

$p(\mu_t | O^t)$ 前回の音響モデル中のガウス分布の平均ベクトルパラメータ μ_t

の事後分布確率

$\mu^{\wedge}_t$ 音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 p

40

$(\mu_t | O^t)$ をガウス分布で表現した際の平均ベクトルパラメータ、もしくは、 $p(\mu_t | O^t)$ の平均ベクトルパラメータ

$Q^{\wedge}_t$ 音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ をガウス分布で表現した際の共分散行列パラメータ、もしくは、 $p(\mu_t | O^t)$ の共分散行列パラメータ

$r^{\wedge}_t$ 音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ をガウス分布で表現した際の共分散行列パラメータ $Q^{\wedge}_t$ に対するスケーリング因子、もしくは、 $p(\mu_t | O^t)$ の共分散行列パラメータ $Q^{\wedge}_t$ に対するスケーリング因子

【0065】

50

実施例 1 の音響モデル作成装置の機能構成例を図 12 に示し、処理の流れを図 10 を用いて説明し、図 13 に逐次適応法を用いた場合の音響モデルパラメータが変換される手順を示し、図 14 に逐次学習部の機能構成例を示す。図 12 に示すように、音響モデル作成装置 148 は、特徴抽出部 4、特徴量記憶部 5、モデル適応化部 150 と、で構成され、モデル適応化部 150 は逐次学習部 152、事後確率記憶部 154、モデル更新部 156 とで構成される。また、逐次学習部 152 は、 $r^{\wedge}$ 更新部 1522、 $\mu^{\wedge}$ 更新部 1524、事後確率計算部 1526、とを有する。

【0066】

まず、前回の事後確率分布の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、スケーリング因子 $r^{\wedge}_t$ ($r^{\wedge}_t$ については後述する) がモデル適応化部 150 で読み込まれる (ステップ S60)。そして、適応音声データ 20 が読み込まれ (ステップ S62)、適応音声データが特徴抽出部 4 に入力され、特徴量系列 O_{t+1} に変換される (ステップ S64)。変換された特徴量系列 O_{t+1} は一旦、特徴量記憶部 5 に記憶され、逐次学習部 152 に入力される。

【0067】

そして、逐次学習部 152 の処理としてまず、(i) 音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ が、当該事後確率分布 $p(\mu_t | O^t)$ の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、事後確率分布 $p(\mu_t | O^t)$ の共分散行列パラメータ $Q^{\wedge}_t$ に対するスケーリング因子 $r^{\wedge}_t$ 、初期音響モデルパラメータ中の共分散行列 Σ 、で表現されるガウス分布で表されることに基づく。そして、(ii) 前回までの累積された特徴量系列 O^t が加味された、前回求めた音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ 及び (iii) 今回まで累積された特徴量系列 O^t の一部を用いる。(iv) 今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現した際の、平均ベクトルパラメータ $\mu^{\wedge}_{t+1}$ 及び共分散行列のスケーリング因子 $r^{\wedge}_{t+1}$ を計算することで、今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布 $p(\mu_{t+1} | O^{t+1})$ を求める。

【0068】

以下、詳細に説明する。また、この実施例では、計算量、安定性、メモリ量、の改善に焦点を当てるため、線形回帰適応ではなく、パラメータ数の少ない平行移動適応 (前記 [平行移動適応] の段落で説明) に対して議論を進める。以下、前記 (i) ~ (iv) に分けて説明する。

前記式 (28) ~ (30) について、式 (29) に示す $K^{\wedge}_{t+1}$ を式 (30) に代入した式を以下に示す。

$$Q^{\wedge}_{t+1} = ((u^0)^{-1} \Sigma + Q^{\wedge}_t)^{-1} + \zeta_{t+1} \Sigma^{-1} \quad (32)$$

$$\mu^{\wedge}_{t+1} = \mu^{\wedge}_t + v_{t+1} + Q^{\wedge}_{t+1} \zeta_{t+1} \Sigma^{-1} (M_{t+1} / \zeta_{t+1} - \mu^{\wedge}_t - v_{t+1}) \quad (33)$$

【0069】

そして、分布パラメータ数を削減するために、前記式 (31) に示したように、音響モデルのガウス分布の平均ベクトル μ_t に対する事後分布の共分散行列 $Q^{\wedge}_t$ を音響モデルのガウス分布の共分散行列 Σ と、スケーリング因子 $r^{\wedge}_t$ または $r^{\wedge}_t$ の逆数 ($r^{\wedge}_t$)⁻¹ を掛け合わせたもので表現する。念のため、式 (31) を以下に示す。スケーリング因子 $r^{\wedge}_t$ は実数 (スカラー) で表されるパラメータである。

$$Q^{\wedge}_t = (r^{\wedge}_t)^{-1} \Sigma \quad (31)$$

この式 (31) を式 (32) に代入すると $Q^{\wedge}_{t+1}$ の更新式はそれぞれ以下のように表現できる。

$$\begin{aligned} Q^{\wedge}_{t+1} &= ((\mu^0)^{-1} \Sigma + (r^{\wedge}_t)^{-1} \Sigma)^{-1} + \zeta_{t+1} \Sigma^{-1} \\ &= ((\mu^0)^{-1} + (r^{\wedge}_t)^{-1})^{-1} + \zeta_{t+1} \Sigma^{-1} \end{aligned} \quad (34)$$

【0070】

そして、式(31)を変形した式 $Q^{\wedge}_{t+1} = (r^{\wedge}_{t+1})^{-1} \Sigma$ の右辺と式(34)の右辺とは等しくなるので、以下の式(35)が成り立つ。

$$r^{\wedge}_{t+1} = ((\mu^0)^{-1} + (r^{\wedge}_t)^{-1})^{-1} + \zeta_{t+1} \quad (35)$$

つまり、式(32)に示す $Q^{\wedge}_{t+1}$ の更新式を式(35)に示す $r^{\wedge}_{t+1}$ に書き直すことができる。

【0071】

一方、 $\mu^{\wedge}_{t+1}$ についての更新式について検討すると、前記式(31)を式(33)に代入すると以下の式のようにになる。

【数9】

$$\begin{aligned} \mu^{\wedge}_{t+1} &= \mu^{\wedge}_t + v_{t+1} + (r^{\wedge}_{t+1})^{-1} \Sigma \zeta_{t+1} \Sigma^{-1} \left(\frac{M_{t+1}}{\zeta_{t+1}} - \mu^{\wedge}_t - v_{t+1} \right) \\ &= \mu^{\wedge}_t + v_{t+1} + (r^{\wedge}_{t+1})^{-1} \zeta_{t+1} \left(\frac{M_{t+1}}{\zeta_{t+1}} - \mu^{\wedge}_t - v_{t+1} \right) \\ &= \mu^{\wedge}_t + v_{t+1} + \frac{M_{t+1} - \zeta_{t+1}(\mu^{\wedge}_t + v_{t+1})}{r^{\wedge}_{t+1}} \quad (36) \end{aligned}$$

となる。つまり、 $r^{\wedge}_{t+1}$ 、 $Q^{\wedge}_{t+1}$ の更新式をまとめると以下のようにになる。

$$r^{\wedge}_{t+1} = ((\mu^0)^{-1} + (r^{\wedge}_t)^{-1})^{-1} + \zeta_{t+1} \quad (35)$$

$$\mu^{\wedge}_{t+1} = \mu^{\wedge}_t + v_{t+1} + \frac{M_{t+1} - \zeta_{t+1}(\mu^{\wedge}_t + v_{t+1})}{r^{\wedge}_{t+1}} \quad (36)$$

【0072】

ここで、上述のように、 μ^0 は予め定められた定数であり、 $r^{\wedge}_t$ は前回のスケーリング因子であり、 ζ_{t+1} は今回の事後占有確率値の和であり、 M_{t+1} は今回の各時点における ζ と特徴量との積和であり、 v_{t+1} は今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトルである。また、スケーリング因子 $r^{\wedge}_t$ の逆数 $(r^{\wedge}_t)^{-1}$ を用いている理由は、 $r^{\wedge}_t$ をそのまま用いると、式(35)の左辺が $(r^{\wedge}_{t+1})^{-1}$ になってしまうという表現上の問題である。実装上はどちらを用いてもかまわない。また、初期値 ζ_0 、 M_0 、 r_0 については、任意の実数値が与えられる。また、式(20)の $Q^{\wedge}_t$ に式(31)を代入することで、以下の式(37)が求められる。

$$p(\mu_{t+1} | O^{t+1}) = N(\mu_{t+1} | \mu^{\wedge}_{t+1}, (r^{\wedge}_{t+1})^{-1} \Sigma) \quad (37)$$

【0073】

つまり、前記(i)で述べたように、前記式(37)から、音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ が、当該事後確率分布 $p(\mu_t | O^t)$ の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、当該事後確率分布 $p(\mu_t | O^t)$ の共分散行列パラメータ $Q^{\wedge}_t$ に対するスケーリング因子 $r^{\wedge}_t$ 、初期音響モデルパラメータ中の共分散行列 Σ で表現されるガウス分布で表されることが理解されよう。

【0074】

また、前記(ii)前回までの累積された特徴量系列 O^t が加味された、前回求めた音響モデル中のガウス分布の平均ベクトルパラメータ μ_t の事後確率分布 $p(\mu_t | O^t)$ を用いることについて説明する。前記式(37)を演算するために前記式(35)(36)を用いるのであるが、式(35)(36)から明らかなように、前回の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、前回のスケーリング因子 $r^{\wedge}_t$ を用いなければならない。また、式(37)より前回求めた音響モデルパラメータの事後分布確率 $p(\mu_t | O^t)$ は、

$$p(\mu_t | O^t) = N(\mu_t | \mu^{\wedge}_t, (r^{\wedge}_t)^{-1} \Sigma) \quad (37')$$

により表される。従って、前回の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、前回のスケーリング因子 $r^{\wedge}_t$ を用いるということは、事後分布確率 $p(\mu_t | O^t)$ を用いているということになる。

【0075】

10

20

30

40

50

また、(i i i) 今回まで累積された特徴量系列 O^t の一部を用いることについて説明する。今回の係数ベクトル v_{t+1} の推定は、上述のように、EMアルゴリズムやMLLRアルゴリズムを用いて、行われる。当該推定は、 $O^t = O_1, O_2, \dots, O_t$ のうちの一部を用いる。そして、式(37)により今回の音響モデルパラメータの事後確率分布 $p(\mu_{t+1} | O^{t+1})$ を求める。

【0076】

また、(i v) 今回の音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布をガウス分布で表現することについて説明すると、音響モデル中のガウス分布の平均ベクトルパラメータの事後確率分布 $p(\mu_t | O^t)$ がガウス分布で表現されると仮定することで、前記式(19)が表されるからである。

【0077】

図14について説明すると、 $r^{\wedge}$ 更新部1522が前記式(35)を演算することで $r^{\wedge}$ を更新し、 $\mu^{\wedge}$ 更新部1524が前記式(36)を演算することで $\mu^{\wedge}$ を更新する。事後確率計算部1526が前記式(37)を演算することで今回の音響モデルパラメータの事後確率分布 $p(\mu_{t+1} | O^{t+1})$ を求める。また、モデル更新部156の記載については、実施例2で説明する。また、変形例を説明すると、前記式(35)は好適な例であり、前記式(35)に近い式であれば、今回のスケーリング因子 $r^{\wedge}_{t+1}$ は、前回のスケーリング因子 $r^{\wedge}_t$ 、今回の事後占有確率値の和 ζ_{t+1} とから求めることができる。同様に、前記式(36)は好適な例であり、前記式(36)に近い式であれば、今回の平均ベクトルパラメータ $\mu^{\wedge}_{t+1}$ は、前回の平均ベクトルパラメータ $\mu^{\wedge}_t$ 、今回の音響モデルパラメータ中の平均の確率的ダイナミクスを線形表現した時の係数ベクトル v_{t+1} 、今回の事後占有確率値の和 ζ_{t+1} 、今回の各時点における ζ と特徴量との積和 M_{t+1} 、今回のスケーリング因子 $r^{\wedge}_{t+1}$ とから求めることができる。

【0078】

そして、モデル更新部56は、今回の音響モデルパラメータ中の事後確率分布 $p(\mu_{t+1} | O^{t+1})$ を新たな音響モデルパラメータに変換して更新する(ステップS70)。

【0079】

次に、この実施例の発明の効果を説明する。式(35)、式(36)からも分かるように、分布パラメータの共分散行列 $Q^{\wedge}_t$ は、前記式(31)を用いることにより、 $r^{\wedge}_t$ 、 Σ に置き換えられ、また Σ は打ち消され、式中の行列表現が取り除かれていることが分かる。また、本実施例1の更新式(35)(36)と従来の更新式(32)(33)とを見比べると、本実施例1の更新式(35)(36)は行列計算(積、和、逆行列)を必要としないため、計算が高速となり、安定性を確保できる。

また、式(35)(36)を用いることにより、逐次適応において、図9中の音響モデル記憶部58記載のように、分布パラメータの共分散行列 $Q^{\wedge}_t$ と平均ベクトルパラメータ $\mu^{\wedge}_t$ を記録するのではなく、図12、図13中の音響モデル記憶部158記載のようにスケーリング因子 $r^{\wedge}_t$ と平均ベクトル $\mu^{\wedge}_t$ を記録することにより、大幅にメモリ量を削減できる。

【実施例2】

【0080】

以上の方法により音響モデルパラメータの事後確率分布 $p(\mu | O)$ つまり、音響モデルが求まった。この実施例2では、求められた音響モデルを用いて音声認識をする、つまり音響スコアの算出の処理を説明する。図15に、この実施例の音声認識装置の機能構成例を示し、図16に、音声認識の主な処理の流れを示す。

【0081】

認識用音声データの音響的特徴量と同様な音響的特徴を持つ適応用音声データが実施例1で説明した音響モデル作成装置148に入力される。そして、音響モデル記憶部158内の音響モデルが上述したように、更新される(ステップS80)。認識用音声データがフレームに分割されて認識用音声データ x として、特徴抽出部4に入力され、特徴量系列

10

20

30

40

50

○に変換される。この特徴量系列○は、単語列探索部 6 に入力される（ステップ S 8 2）。

【0082】

単語列探索部 6 で特徴量系列○に対して、音響モデル記憶部 8 の音響モデルを用いて必要に応じて各ガウス分布の音響スコアを算出する。この音響スコア算出には例えば、以下の式（40）の計算を行う。

$$\int p(x_\tau | \mu_t) p(\mu_t | O^t) d\mu_t \quad (40)$$

ここで $p(x_\tau | \mu_t)$ は音響モデルの出力分布である。 μ_t 以外のパラメータはここでは省略する。従って、 $p(\mu_t | O^t)$ について検討すれば良い。単語列探索部 6 による複数フレームの音響スコア算出に関しては前記式（40）をもとに動的計画法（DP：Dynamic Programming マッチング）を行えばよい。音響スコアを最大とする単語列を認識単語列として出力する（ステップ S 8 4）。なお、この場合はステップ S 8 0 におけるモデル更新は、音響モデルとして事後確率分布 $p(\mu_t | O^t)$ の更新を行う（ステップ S 8 0 a）。前記式（40）の積分は数値的に解くことも可能であるが、次のような 2 種類の解析解が存在する。

【0083】

[Plug-in法]

Plug-in法では、積分をまともに扱うのではなく、 $p(\mu_t | O^t)$ の事後確率最大化（MAP）値（以下の式（41）の右辺）は、前記式（36）の μ^t である事を利用する。つまり、以下の式（41）になる。

$$\hat{\mu}_t = \arg \max_{\mu_t} p(\mu_t | O^t) \quad (41)$$

従って、音響モデル作成装置 60 による前記ステップ S 8 0 におけるモデル更新として、前記式（36）で求まる μ^t を出力分布 $p(x_\tau | \mu_t)$ の平均ベクトルパラメータ μ_t にそのまま代入（Plug-in）して音響モデルパラメータを更新する（ステップ S 8 0 b）。このようにすればスコア計算を、以下の式（42）で行うことが出来る。

【数11】

$$\int p(x_\tau | \mu_t) p(\mu_t | O^t) d\mu_t \doteq N(x_\tau | \hat{\mu}_t, \Sigma) \quad (42)$$

つまり、平均 μ^t 、共分散行列 Σ のガウス分布で表現する。これをPlug-in法と呼ぶ。また、その他のパラメータ状態遷移確率 a 、混合重み因子 w 、共分散行列 Σ 、はそのまま適用する。ステップ S 8 0 b の後は、破線矢印で示すように、ステップ S 8 2 に移る。

【0084】

[周辺化法]

周辺化法は、Plug-in法と違い積分を解析的に解く方法である。この積分をとく方法が、平均ベクトルパラメータ μ_t についての周辺化にあたる。周辺化法は、Plug-in法と比較して、平均ベクトルパラメータの事後確率分布 $p(\mu_t | O^t)$ の分散を考慮することになる。このようにすれば、積分計算によるスコア計算は以下の式（43）で表せることになる。

$$\int p(x_\tau | \mu_t) p(\mu_t | O^t) d\mu_t = N(x_\tau | \hat{\mu}_t, (\Sigma + (r^t)^{-1} \Sigma)) \quad (43)$$

つまり、周辺化法を利用する場合はステップ S 8 0 のモデル更新において、平均ベクトルパラメータ μ を μ^t と置き換える（ステップ S 8 0 b）と共に、共分散行列パラメータ $\Sigma \rightarrow \Sigma + (r^t)^{-1} \Sigma$ と置き換えて（ステップ S 8 0 c）、音響モデルパラメータを更新する。また、その他のパラメータ、つまり、状態遷移確率 a 、混合重み因子 w 、はそのまま適用する。ステップ S 8 0 c のあとは、ステップ S 8 2 に移る。

【0085】

また、図14に示すように、モデル更新部156は、破線で示すパラメータ変換部として作用する。Plug-in法を用いる場合は、 $\mu \rightarrow \mu \wedge_t$ と置き換えて、平均ベクトルパラメータを更新し、周辺化法を用いる場合は、 $\mu \rightarrow \mu \wedge_t$ 、 $\Sigma \rightarrow \Sigma + (r \wedge_t)^{-1} \Sigma$ と置き換えて、平均ベクトルパラメータ、共分散行列パラメータを更新する。このようにすることで、分布変換に基づく逐次適応法による音響スコアを算出できる。

【0086】

〔実験結果〕

ASJ（日本音響学会）読み上げ音声データベース100時間分を用いてトライフォンHMMの総状態数2000、HMM状態あたりの混合数16の不特定話者音響モデルを構築し、日本語模擬ニュース音声に対し、逐次適応実験を行った。特徴量は12次元MFCC（メルフレクエンシイペプストラム係数）と、そのフレームのエネルギーと、MFCCのフレーム間差分 Δ と、その差分MFCCのフレーム間差分デルタ $\Delta\Delta$ として、語彙サイズ70万語のトライアングルを用いて大語彙連続音声認識実験を行った。逐次適応を行わない通常の音声認識の場合の音声認識率は81.3%であった。

【0087】

ここで、分布パラメータに共分散行列 $Q \wedge_t$ を用いた従来逐次適応では、認識率は88.5%と大きく改善した。しかし、1発話（1秒程度）で逐次更新を行う。オンライン逐次適応タスクにおいて、従来法は実時間処理できず（リアルタイムの2倍程度）、また、 $Q \wedge_t$ に用いたメモリ消費量は27メガバイトであった。

【0088】

一方、分布パラメータにスケーリング因子 $r \wedge_t$ を用いた本発明では認識率は88.5%を、従来法と同程度の性能を維持しつつ、実時間処理（リアルタイムの1倍程度）を実現した。また、 $r \wedge_t$ に用いたメモリ消費量は1.3メガバイトであり、 $Q \wedge_t$ を用いた場合と比較して、メモリを20分の1程度削減できた。

【0089】

以上の各実施形態の他、本発明である音響モデル作成装置は上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。また、音響モデル作成装置において説明した処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されるところとしてもよい。

また、この発明の音響モデル作成装置における処理をコンピュータによって実現する場合、音響モデル作成装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、音響モデル作成装置における処理機能がコンピュータ上で実現される。

【0090】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD（Digital Versatile Disc）、DVD-RAM（Random Access Memory）、CD-ROM（Compact Disc Read Only Memory）、CD-R（Recordable）／RW（ReWritable）等を、光磁気記録媒体として、MO（Magnetooptical disc）等を、半導体メモリとしてEEP-ROM（Electrically Erasable and Programmable Read Only Memory）等を用いることができる。

【0091】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバ

コンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【0092】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP（Application Service Provider）型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

10

【0093】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、音響モデル作成装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

20

【図面の簡単な説明】

【0094】

【図1】従来の音声認識装置の機能構成例を示した図。

【図2】従来の音声認識装置の処理フローを示した図。

【図3】従来の音響モデル作成装置の機能構成例を示した図。

【図4】従来の音響モデル作成装置の処理フローを示した図。

【図5】線形回帰行列を用いた場合の音響モデル作成装置の機能構成例を示した図。

【図6】図5に示す音響モデル作成装置の主な処理の流れを示したブロック図。

【図7】逐次適応法を用いた場合の音響モデルパラメータが変換される手順を示した図。

30

【図8】逐次的応法を適用した場合の音響モデル作成装置の機能構成例を示した図。

【図9】事後確率分布のパラメータを逐次適応させる順序を示した図。

【図10】図8に示す音響モデル作成装置の主な処理の流れを示したブロック図。

【図11】従来の逐次学習部52の機能構成例を示した図。

【図12】本実施例の音響モデル作成装置の機能構成例を示した図。

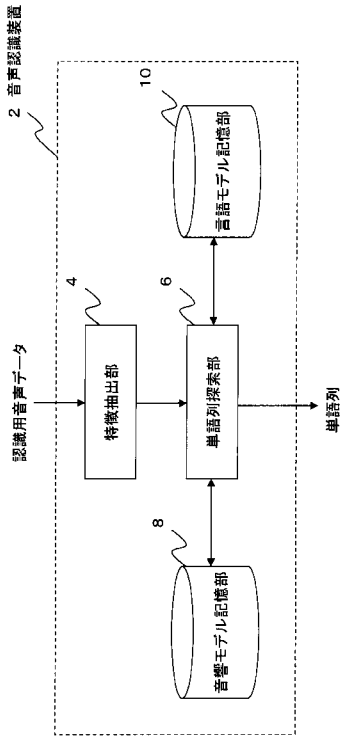
【図13】本実施例の事後確率分布のパラメータを逐次適応させる順序を示した図。

【図14】本実施例の逐次学習部などの機能構成例を示した図。

【図15】本実施例の音声認識装置の機能構成例を示した図。

【図16】本実施例の音声認識装置の処理フローを示した図。

【図 1】



【図 2】

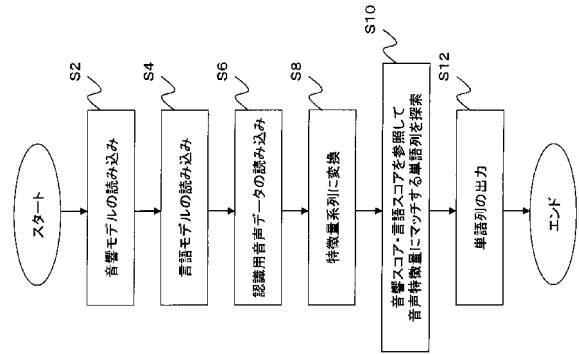
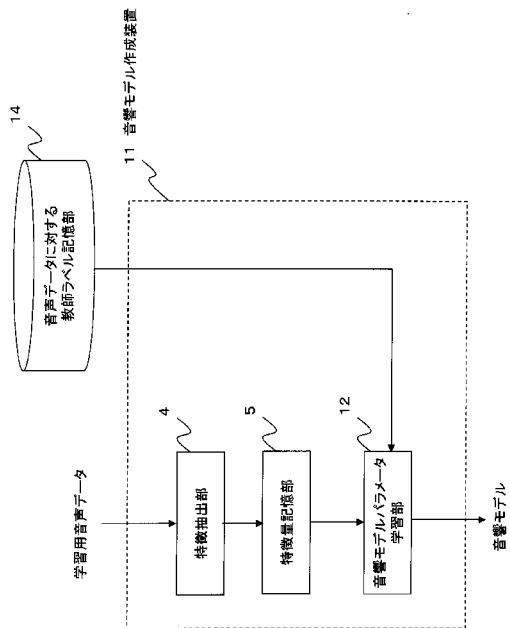


図 1

図 2

【図 3】



【図 4】

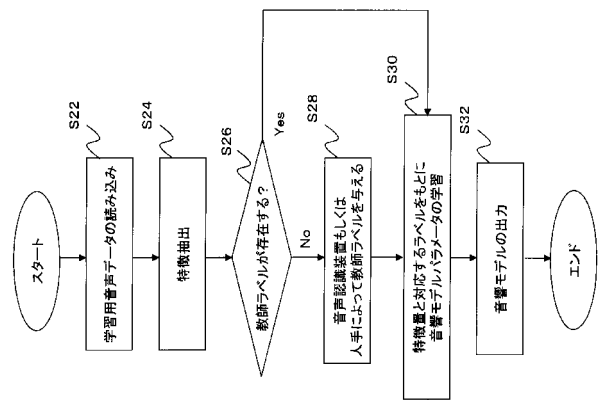


図 3

図 4

【図 5】

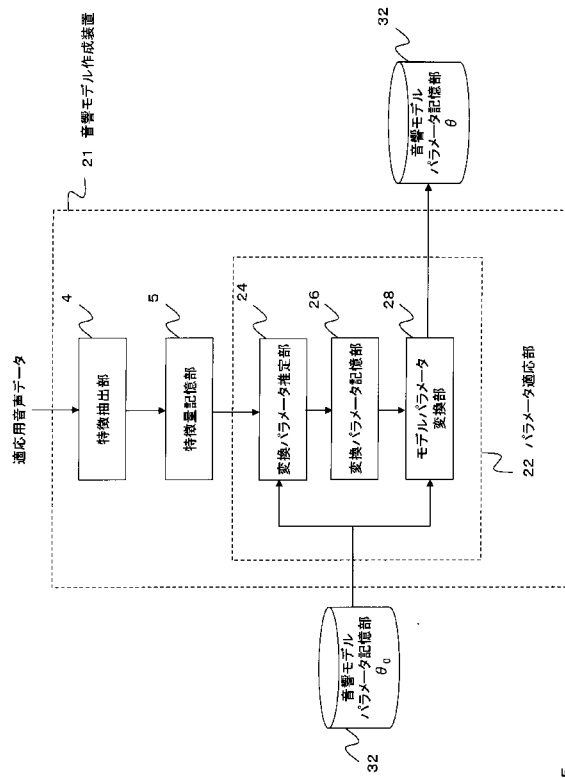


図5

【図 6】

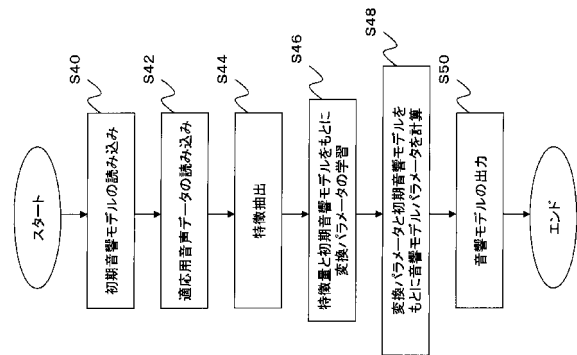


図6

【図 7】

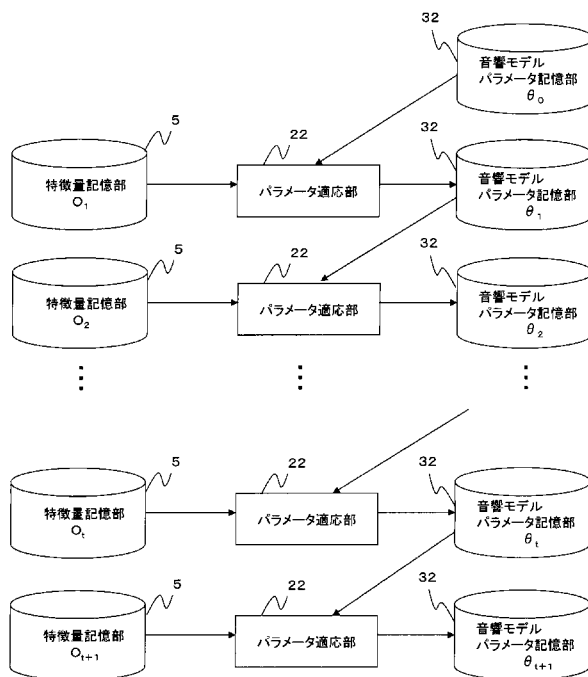


図7

【図 8】

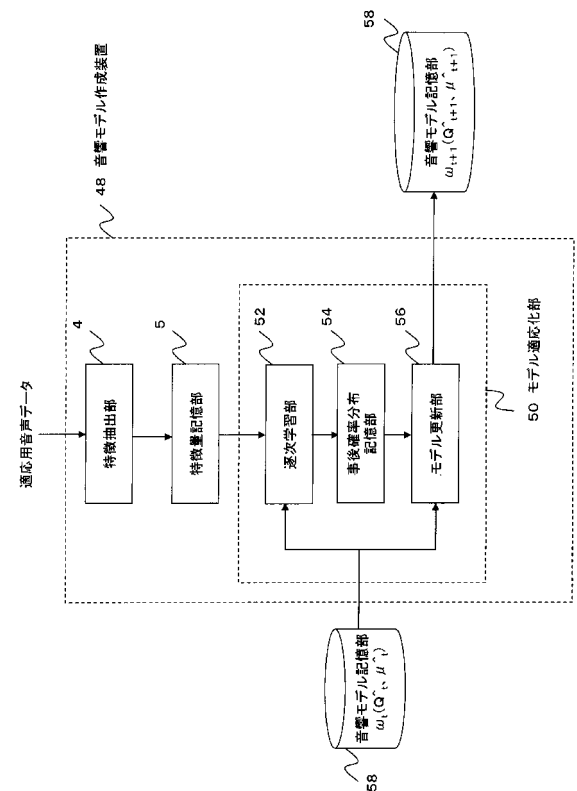


図8

【図 9】

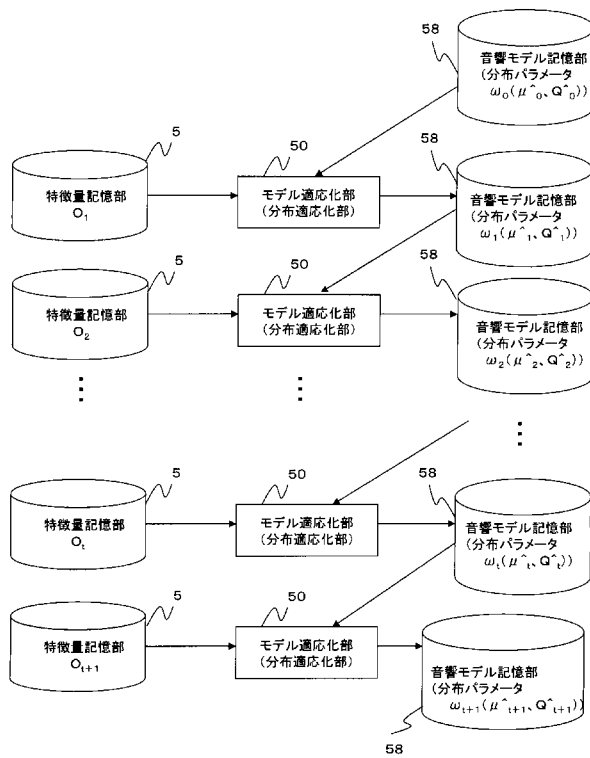


図9

【図 10】

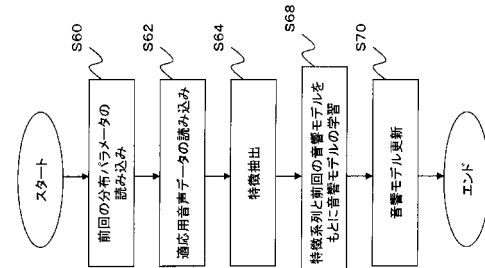


図10

【図 11】

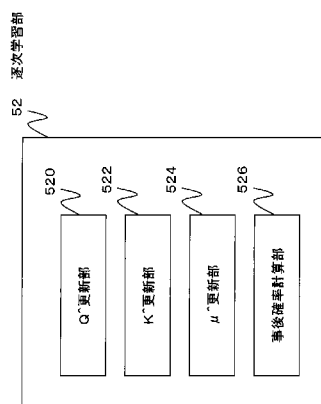


図11

【図 12】

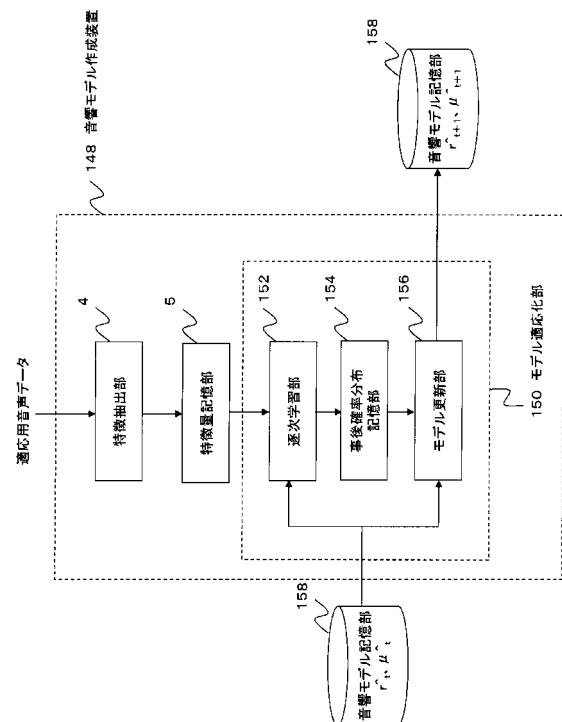


図12

【図13】

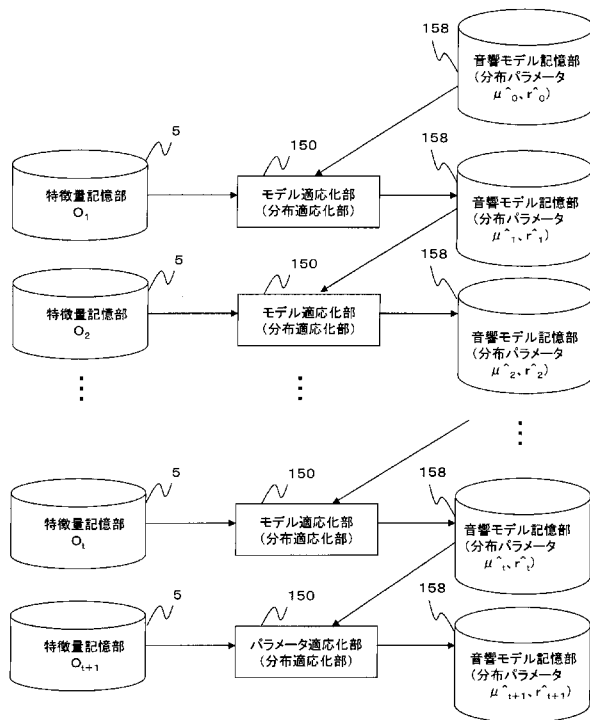


図13

【図14】

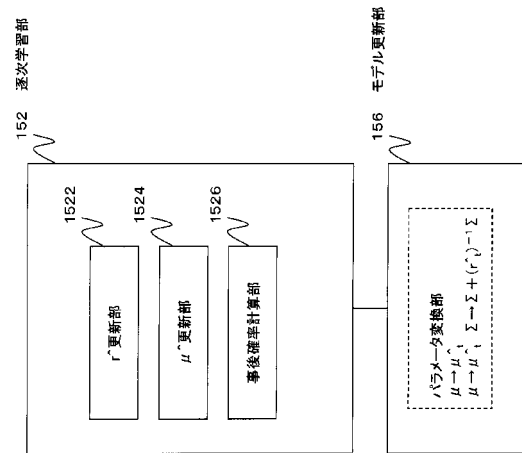


図14

【図15】

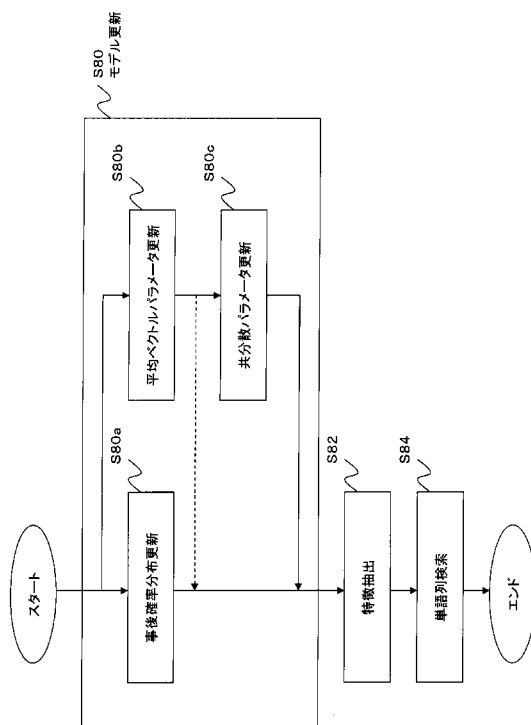


図15

【図16】

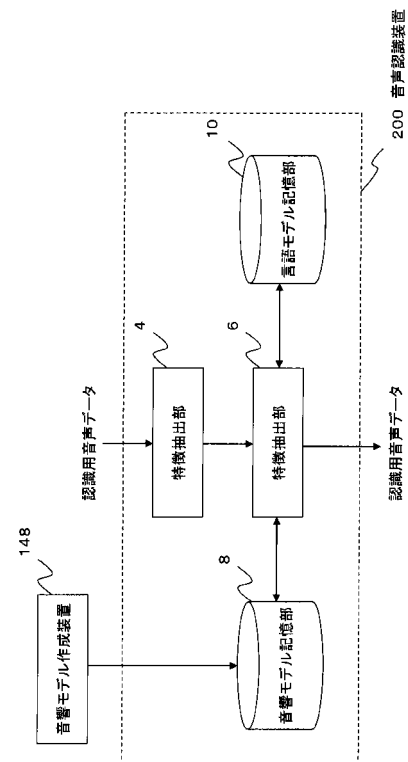


図16