

(19)日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11)特許番号

第2852210号

(45)発行日 平成11年(1999) 1月27日

(24)登録日 平成10年(1998)11月13日

(51)Int.Cl. ⁶	識別記号	F I	
G 1 0 L 3/00	5 2 1	G 1 0 L 3/00	5 2 1 N
	5 3 5		5 2 1 C
			5 3 5

請求項の数5 (全 10 頁)

(21)出願番号	特願平7-239821	(73)特許権者	593118597 株式会社エイ・ティ・アール音声翻訳通信研究所 京都府相楽郡精華町大字乾谷小字三平谷5番地
(22)出願日	平成7年(1995)9月19日	(72)発明者	外村 政啓 京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エイ・ティ・アール音声翻訳通信研究所内
(65)公開番号	特開平9-81178	(72)発明者	松永 昭一 京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エイ・ティ・アール音声翻訳通信研究所内
(43)公開日	平成9年(1997)3月28日	(74)代理人	弁理士 青山 葆 (外2名)
審査請求日	平成7年(1995)9月19日	審査官	樫本 剛

最終頁に続く

(54)【発明の名称】 不特定話者モデル作成装置及び音声認識装置

1

(57)【特許請求の範囲】

【請求項1】 入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルに基づいて、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する不特定話者モデル作成装置において、

入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルの各状態の出力ガウス分布を各状態ごとに独立にクラスタリングして合成することにより不特定話者の混合ガウス分布の隠れマルコフモデルを作成するモデル作成手段を備えたことを特徴とする不特定話者モデル作成装置。

【請求項2】 上記モデル作成手段は、
入力された複数の特定話者の発声音声データに基づいて、複数の話者に対して同一の初期話者隠れマルコフモデルを用いて所定の学習法により上記発声音声データの

2

存在する状態に対してのみ出力ガウス分布を学習することにより、複数個の特定話者用単一ガウス分布の隠れマルコフモデルを作成する学習手段と、

上記学習手段によって作成された複数個の特定話者用単一ガウス分布の隠れマルコフモデルに基づいて、各出力ガウス分布間の距離を基準にして、各クラスタにより短い距離に出力ガウス分布が含まれるように複数のクラスタにクラスタリングを行うクラスタリング手段と、

上記クラスタリング手段によって各状態毎にクラスタリングされた単一ガウス分布の隠れマルコフモデルに基づいて、各クラスタ内の複数の出力ガウス分布の隠れマルコフモデルを各状態の単一ガウス分布の隠れマルコフモデルに合成する合成手段と、

上記合成手段によって合成された各状態の単一ガウス分布の隠れマルコフモデルを混合することにより、不特定

10

話者の混合ガウス分布の隠れマルコフモデルを作成する混合手段とを備えたことを特徴とする請求項 1 記載の不特定話者モデル作成装置。

【請求項 3】 上記クラスタリング手段は、各状態毎に予め設定したしきい値以上のデータ量で学習された出力ガウス分布のみを取り出した後、クラスタリングすることを特徴とする請求項 2 記載の不特定話者モデル作成装置。

【請求項 4】 上記クラスタリング手段は、各状態においてクラスタリングされた各クラスタの中心と各出力ガウス分布間の距離の平均値が予め決めた距離以下になるまでクラスタリングを繰り返すことにより、各状態における各出力ガウス分布のバラツキが大きいほどクラスタ数が多くなるように各状態におけるクラスタ数を決定することを特徴とする請求項 2 又は 3 記載の不特定話者モデル作成装置。

【請求項 5】 入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルに基づいて、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する請求項 1 乃至 4 のうちの 1 つに記載の不特定話者モデル作成装置と、

入力された発声音声文の音声信号に基づいて、上記不特定話者モデル作成装置によって作成された不特定話者の混合分布の隠れマルコフモデルを用いて、音声認識する音声認識手段とを備えたことを特徴とする音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数の特定話者の隠れマルコフモデルに基づいて、不特定話者の隠れマルコフモデル（以下、HMMという。）を作成する不特定話者モデル作成装置、及びその不特定話者モデル作成装置を用いた音声認識装置に関する。

【0002】

【従来の技術】従来、学習用の特定話者モデルに基づいて不特定話者のHMMを作成するために、バウム・ウェルチ（Baum- Welch）の学習アルゴリズム（以下、第1の従来例という。）が広く用いられている（例えば、中川聖一著，“確率モデルによる音声認識”，p. 55 - 64，電子情報通信学会，昭和63年7月発行参照。）。この第1の従来例では、HMMにおいて時刻1から時刻tまでの間部分観測列 $\{y_1, y_2, y_3, \dots, y_t\}$ を観測した後、時刻tには状態iにいる前向き確率と、時刻tに状態iにいて時刻t+1から最後まで部分観測列 $\{y_{t+1}, y_{t+2}, y_{t+3}, \dots, y_r\}$ を観測する後向き確率とを用いて、HMMのパラメータを再推定して学習することにより、不特定話者のHMMを作成する。

【0003】上記第1の従来例の方法を用いて、多様な話者の音声の音響的特徴量の変動に対応するために多数

話者の音声データでモデルを学習することが望ましく学習データが多量になる傾向があり、多数の話者による多量の音声データでモデルを学習することが望ましい。しかしながら、このような多量のデータを取り扱う場合、その膨大な計算量はコンピュータの処理速度が高速化しつつある現在においても問題となっている。

【0004】このような不特定話者モデルの計算量を削減するために、既に小坂らによって特定話者モデルによる話者クラスタリングとモデル合成によるCCL法（以下、第2の従来例という。）が提案されている（従来文献2「小坂ほか，“クラスタリング手法を用いた不特定話者モデル作成法”，日本音響学会論文集，1-R-12，1994年11月」参照。）。この第2の従来例の方法では、各話者の音声の音響的特徴の類似性がすべての音響空間で等しいという仮定のもとに、すべての音韻にわたるモデルセット全体を単位としてクラスタリングを行っている。具体的には、十分に学習された特定話者モデルをモデル間の距離を定義することによってクラスタリングした後、各特定話者モデルを合成することにより不特定話者モデルを作成している。

【0005】

【発明が解決しようとする課題】第2の従来例の方法では、少ない計算量で不特定話者モデルを作成することが可能であるが、特定話者モデルのすべてのパラメータが十分学習されていない場合には性能のよいモデルが得られないため各話者に対して多くの発声データが必要となる。また、HMMの全ての状態において混合出力ガウス分布の混合数が必ず同じになり、話者による特徴量のバラツキの少ない状態に対して無駄なパラメータが増えるという問題があった。

【0006】本発明の第1の目的は以上の問題点を解決し、各特定話者モデルのすべてのパラメータが学習されている必要がなく、また話者毎に学習されているパラメータが異なっている場合においても不特定話者モデルを作成でき、しかも処理装置のメモリ容量が少なくすみ、その計算時間を短縮することができる不特定話者モデル作成装置を提供することにある。また、本発明の第2の目的は、上記第1の目的に加えて、作成された不特定話者モデルを用いて音声認識することができ、従来例に比較して音声認識率を改善することができる音声認識装置を提供することにある。

【0007】

【課題を解決するための手段】本発明に係る請求項1記載の不特定話者モデル作成装置は、入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルに基づいて、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する不特定話者モデル作成装置において、入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルの各状態の出力ガウス分布を各状態ごとに独立にクラスタリングして合成することにより不特定話者の混合

ガウス分布の隠れマルコフモデルを作成するモデル作成手段を備えたことを特徴とする。

【0008】また、請求項2記載の不特定話者モデル作成装置は、請求項1記載の不特定話者モデル作成装置において、上記モデル作成手段は、入力された複数の特定話者の発声音声データに基づいて、複数の話者に対して同一の初期話者隠れマルコフモデルを用いて所定の学習法により上記発声音声データの存在する状態に対してのみ出力ガウス分布を学習することにより、複数個の特定話者用単一ガウス分布の隠れマルコフモデルを作成する学習手段と、上記学習手段によって作成された複数個の特定話者用単一ガウス分布の隠れマルコフモデルに基づいて、各出力ガウス分布間の距離を基準にして、各クラスタにより短い距離に出力ガウス分布が含まれるように複数のクラスタにクラスタリングを行うクラスタリング手段と、上記クラスタリング手段によって各状態毎にクラスタリングされた単一ガウス分布の隠れマルコフモデルに基づいて、各クラスタ内の複数の出力ガウス分布の隠れマルコフモデルを各状態の単一ガウス分布の隠れマルコフモデルに合成する合成手段と、上記合成手段によって合成された各状態の単一ガウス分布の隠れマルコフモデルを混合することにより、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する混合手段とを備えたことを特徴とする。

【0009】さらに、請求項3記載の不特定話者モデル作成装置は、請求項2記載の不特定話者モデル作成装置において、上記クラスタリング手段は、各状態毎に予め設定したしきい値以上のデータ量で学習された出力ガウス分布のみを取り出した後、クラスタリングすることを特徴とする。

【0010】またさらに、請求項4記載の不特定話者モデル作成装置は、請求項2又は3記載の不特定話者モデル作成装置において、上記クラスタリング手段は、各状態においてクラスタリングされた各クラスタの中心と各出力ガウス分布間の距離の平均値が予め決めた距離以下になるまでクラスタリングを繰り返すことにより、各状態における各出力ガウス分布のバラツキが大きいほどクラスタ数が多くなるように各状態におけるクラスタ数を決定することを特徴とする。

【0011】また、本発明に係る請求項5記載の音声認識装置は、入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルに基づいて、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する請求項1乃至4のうちの1つに記載の不特定話者モデル作成装置と、入力された発聲音声文の音声信号に基づいて、上記不特定話者モデル作成装置によって作成された不特定話者の混合分布の隠れマルコフモデルを用いて、音声認識する音声認識手段とを備えたことを特徴とする。

【0012】

【発明の実施の形態】以下、図面を参照して本発明に係

る実施形態について説明する。図1は、本発明に係る一実施形態である音声認識装置のブロック図である。本実施形態の音声認識装置は、特に、特定話者の発聲音声データ30のメモリに格納された複数N人の特定話者の発聲音声データに基づいて公知の最尤推定法を用いてデータの存在する状態に対してのみ出力ガウス分布を学習し、上記特定話者モデルの中から学習された出力ガウス分布のパラメータのみを取り出しHMMの対応する状態毎にクラスタリングを行った後合成及び混合を行って混合ガウス分布の隠れマルコフ網(以下、HM網という。)を作成し、作成したHM網をHM網11のメモリに格納する不特定話者モデル作成部31を備え、HM網11のメモリに格納されたHM網を参照して音声認識を行うことを特徴とする。

【0013】この音声認識装置は、マイクロホン1と、特徴抽出部2と、バッファメモリ3と、音素照合部4と、文脈自由文法データベース20のメモリに格納された所定の文脈自由文法に基づいて作成された、メモリに格納されたLRテーブル13のメモリを参照して音声認識処理を実行する音素コンテキスト依存型LRパーザ(以下、LRパーザという。)5とを備える。

【0014】図2は、不特定話者モデル作成部31によって実行される不特定話者モデル作成処理を示すフローチャートである。当該作成処理においては、まず、ステップS1において、複数N人の特定話者の発聲音声データに基づいて、当該発聲音声データの特徴パラメータを抽出し、抽出した特徴パラメータに基づいて、複数N人の全ての話者に対して同一のHM網である初期話者モデル(各状態1混合)を用いて公知の最尤推定法によりデータの存在する状態に対してのみ出力ガウス分布の平均値と分散を学習することにより、N個の特定話者用単一ガウス分布のHM網を作成する。

【0015】次いで、ステップ2では、図3に示すように、作成されたN個の特定話者用単一ガウス分布のHM網に基づいて、各状態毎に予め設定したしきい値以上のデータ量で学習された出力ガウス分布のみを取り出した後、図4に示すように、出力ガウス分布間の公知のバタチャ(Bhattacharyya)距離を基準にして、各クラスタにより短い距離に出力ガウス分布が含まれるように複数のクラスタにクラスタリングを行なう。ここで、取り出す学習データ量にしきい値を設けたのは信頼性の低い出力ガウス分布がクラスタリングに悪影響を及ぼさないようにするためである。これにより、信頼性の高いHM網11を得ることができ、当該HM網11を用いて音声認識することにより、従来例に比較して高い音声認識率で音声認識することができる。また、当該クラスタリングでは、各状態においてクラスタリングされ各クラスタの中心と各出力ガウス分布間の公知のバタチャ(Bhattacharyya)距離の平均値が予め決めた距離以下になるまでクラスタリングを繰り返

7

すことにより、各状態における各メンバーの出力ガウス分布のパラツキに応じてクラスタ数 K を決定する。ここで、パラツキが大きい場合はクラスタ数 K を比較的多く設定する一方、パラツキが小さい場合はクラスタ数 K を比較的小く設定する。また、上記クラスタ数 K の決定においては、最大のクラスタ数 K_{max} 及び最小のクラスタ数 K_{min} を設定してもよい。さらに、学習データ量が小さい場合は、好ましくは、クラスタ数 K を小さく設定する。

【0016】次いで、ステップ $S3$ においては、上記ステップ $S2$ で各状態ごとにクラスタリングされた結果を用いて、図5に示すように、クラスタ内の複数の出力ガウス分布を各状態の単一ガウス分布に合成する。合成は出力ガウス分布の総数、及びクラスタリング結果が各状態ごとに異なること以外は、従来文献2の方法と同様の方法で行なった。当該ステップ $S3$ の合成方法については詳細後述する。さらに、ステップ $S4$ においては、各*

$$S h_j = \sum_i w_j^{(i)} S_j^{(i)} + \sum_i$$

【数3】

$$w_j^{(i)} = n_j^{(i)} / \left\{ \sum_i n_j^{(i)} \right\}$$

【0019】数1と数2はそれぞれ、複数のガウス分布を単一ガウス分布と見なして求めた場合の平均値、分散を表す。ここで、 $\mu_j^{(i)}$ と $S_j^{(i)}$ は自然数 i 番目のHM網のの状態 j における単一ガウス分布である出力確率密度関数の平均値と分散を表わす。また、 $n_j^{(i)}$ は i 番目のHM網の状態 j におけるサンプル数を表す。すなわち、数1から明らかなように、合成後の平均値 μh_j と分散 $S h_j$ とはそれぞれ、合成前の平均値 μ_j と分散 S_j を、各状態におけるサンプル数 $n_j^{(i)}$ に応じてサンプル数 $n_j^{(i)}$ が大きいほど大きい重み係数 $w_j^{(i)}$ で重み付けされて計算される。

【0020】本実施形態においては、音声認識のための統計的音素モデルセットとしてHM網11を使用している。当該HM網11は効率的に表現された音素環境依存モデルである。1つのHM網は多数の音素環境依存モデルを包含する。HM網11はガウス分布を含む状態の結合で構成され、個々の音素環境依存モデル間で状態が共有される。このためパラメータ推定のためのデータ数が不足する場合も、頑健なモデルを作成することができる。このHM網11は逐次状態分割法 (Successive State Splitting: 以下、SSSという。) を用いて自動作成される。上記SSSではHM網のトポロジーの決定、異音クラスタの決定、各々の状態におけるガウス分布のパラメータの推定を同時に行なう。本実施形態においては、HM網のパラメータとして、ガウス分布で表現される出力確率及び遷移確率を有する。このため認識時には

8

* 状態ごとに全てのクラスタの合成された単一ガウス分布を公知の話者混合法を用いて混合することにより混合ガウス分布のHM網を作成してHM網11のメモリに格納する。混合比率は各クラスタのメンバーの出力ガウス分布の学習データ量の総和の比に比例する値とした。すなわち、各クラスタのメンバーの学習データ量が大きいほど、混合比率を大きく設定する。

【0017】上記ステップ $S3$ において用いられる各クラスタにおける合成後の平均値 μh_j と分散 $S h_j$ は、次の数1及び数2で表される。なお、重み係数 $w_j^{(i)}$ は次の数3で表される。

【0018】

【数1】

$$\mu h_j = \sum_i w_j^{(i)} \mu_j^{(i)}$$

【数2】

$$w_j^{(i)} (\mu_j^{(i)} - \mu h_j)^2$$

一般のHMMと同様に扱うことができる。

【0021】次いで、上述の本実施形態の音声認識方法を用いた、SSS-LR (left-to-right rightmost型) 不特定話者連続音声認識装置について説明する。この装置は、メモリに格納されたHM網11と呼ばれる音素環境依存型の効率のよいHMMの表現形式を用いている。また、上記SSSにおいては、音素の特徴空間上に割り当てられた確率的定常信号源 (状態) の間の確率的な遷移により音声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、尤度最大化の基準に基づいて個々の状態をコンテキスト方向又は時間方向へ分割するという操作を繰り返すことによって、モデルの精密化を逐次的に実行する。

【0022】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、対数パワー及び16次ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して音素照合部4に入力される。

【0023】音素照合部4に接続されるメモリ内のHM網11は、各状態をノードとする複数のネットワークとして表され、各状態はそれぞれ以下の情報を有する。

(a) 状態番号

(b) 受理可能なコンテキストクラスタ

(c) 先行状態、及び後続状態のリスト

(d) 出力確率密度分布のパラメータ

(e) 自己遷移確率及び後続状態への遷移確率

【0024】音素照合部4は、音素コンテキスト依存型

L R パーザ 5 からの音素照合要求に応じて音素照合処理を実行する。そして、不特定話者モデルを用いて音素照合区間内のデータに対する尤度が計算され、この尤度の値が音素照合スコアとして L R パーザ 5 に返される。このときに用いられるモデルは、HMM と等価であるために、尤度の計算には通常の HMM で用いられている前向きパスアルゴリズムをそのまま使用する。

【0025】一方、メモリ内の所定の文脈自由文法 (CFG) データベース 20 を公知の通り自動的に変換して L R テーブルを作成して L R テーブル 13 のメモリに格納される。L R パーザ 5 は、上記 L R テーブル 13 を参照して、入力された音素予測データについて左から右方向に、後戻りなしに処理する。構文的にあいまいさがある場合は、スタックを分割してすべての候補の解析が平行して処理される。L R パーザ 5 は、L R テーブル 13 から次にくる音素を予測して音素予測データを音素照合部 4 に出力する。これに回答して、音素照合部 4 は、その音素に対応する HM 網 11 内の情報を参照して照合し、その尤度を音声認識スコアとして L R パーザ 5 に戻し、順次音素を接続していくことにより、連続音声の認識を行い、その音声認識結果データを出力する。上記連続音声の認識において、複数の音素が予測された場合は、これらすべての存在をチェックし、ビームサーチの方法により、部分的な音声認識の尤度の高い部分木を残すという枝刈りを行って高速処理を実現する。

【0026】以上の実施形態において、特定話者の発声音声データ 30 と、HM 網 11 と、L R テーブル 13 と、文脈自由文法データベース 20 とはそれぞれ、例えばハードディスクメモリに格納される。また、音素照合部 4 と L R パーザ 5 と不特定話者モデル作成部 31 は例えばデジタル電子計算機によって構成される。

【0027】以上の実施形態においては、図 2 の不特定話者モデル作成処理によって不特定話者モデルを作成しているが、当該作成処理によって作成された HM 網に対して公知のバーム・ウェルチの学習アルゴリズムを用いて再学習して、HM 網を作成してもよい。

【0028】

【実施例】本発明者は、図 1 の音声認識装置の有効性を確かめるために、以下の通り実験を行った。当該実験には、コンテキスト依存型の音素 HMM の状態を効果的に共有した HM 網 (例えば、従来文献 3 「鷹見ほか、"音素コンテキストと時間に関する逐次状態分割による隠れマルコフ網の自動生成"、電子通信情報学会技術研究報 *

実験条件

分析条件 サンプル周波数 = 12 KHz
 ハミング窓 = 20 ms
 フレーム周期 = 5 ms

使用パラメータ 16 次 LPC ケプストラム + 16 次 ケプストラム

* 告, SP91-88, 1991年12月」参照。)を使用した。HM網の構造は1人の話者の発声した2620単語の音声データを用いて決定し、総状態数200、及び600の2種類のモデルを作成した。各モデルには1状態10混合の無音モデルを付加した。特定話者モデル学習用の初期話者モデルは無音モデルを除き各状態とも単一分布としパラメータの初期値は構造決定と同じ音声データで決定した。この初期話者モデルをもとに、本特許出願人が所有する、トラベル・プランニングをタスクとした自然発話の音声認識データベース (例えば、従来文献4「T. Morimoto et al., "A Speech and Language Database for Speech Translation Research", Proc. of ICSLP '94, pp. 1791-1794, 1994年」参照)の中の男性81名の自然発話データを用いて最尤推定法により出力ガウス分布の平均値と分散を学習することにより81名分の特定話者モデルを作成した。但し、1人あたりのデータ量が20発話程度と少ないため、分散は初期パラメータより値が大きくなる場合のみ更新した。なお、今回は男性話者のみを用いて不特性話者モデルの作成、及び認識実験を行なった。認識実験は学習に用いたものと同じ自然発話データベースより選択した学習データに含まれない男性9人に対して行なった。

【0029】不特定話者モデルはHM網全体を単位としたモデルベースのクラスタリングを用いた第2の従来例のCCL法と本発明に係るHMMの状態別クラスタリングの結果を用いる方法により作成し両者の性能を音素認識実験により比較した。ただし、本発明に係る状態別クラスタリングによる方法では特定話者モデルの各状態の出力ガウス分布の内、学習時の状態占有データ量が10フレーム以上のもののみを使用した。さらに、状態別クラスタリングによって作成したモデルを初期モデルとしてバーム・ウェルチの学習アルゴリズムによって再学習したモデルの認識率との比較も行なった。またさらに、本発明に係る状態別クラスタリングによる方法でHMMを作成した後、バーム・ウェルチの学習アルゴリズムによって再学習したモデルの認識率についても実験を行った。ここで、実験条件である、分析条件、使用パラメータ、学習/認識データを表1に示す。

【0030】

【表1】

+ 対数パワー + 対数パワー

学習データ 男性 8 1 名 - - 各話者 1 会話 (合計 1 7 9 9 発声)

不特定話者モデル評価データ

男性 9 名 - - 各話者 1 会話 (1 1 ~ 2 9 発声)

【 0 0 3 1 】 表 2 及び表 3 に、第 2 の従来例の C C L 法 (以下、表においてモデルクラスタリングと略す。) 及び、本発明に係る状態別クラスタリングによる方法 (以下、表において、状態別クラスタリングと略す。) で作成した各状態、混合数の H M 網に含まれる出力ガウス分布の総数を示す。第 2 の従来例の C C L 法による場合は無音モデルを除き全ての状態に対して混合分布数が等しくなるが、本発明に係る状態別クラスタリングによる場合は各状態に対して特定話者モデルから抽出された 1 0 フレーム以上のデータで学習された出力ガウス分布数が設定した混合数より少ない場合には抽出された分布数が *

* その状態の混合分布数となるためモデルベースのクラスタリングによる場合より総分布数が少なくなっている。但し、今回は各状態における抽出した出力ガウス分布の平均値のばらつきの度合は混合数の決定において考慮していない。このように音素バランスを考慮した音声データの収集が困難な自由発話音声データベースを用いた場合には各状態ごとに混合分布数を設計することにより不要なパラメータの増加を防ぐことができる可能性があることがわかる。

【 0 0 3 2 】

【表 2】

不特定話者モデルの総分布数 - 2 0 1 状態の H M 網の場合

作成法 / 混合数	5	1 0	1 5	2 0
モデルクラスタリング	1 0 1 0	2 0 1 0	3 0 1 0	4 0 1 0
状態別クラスタリング	9 7 9	1 9 0 3	2 7 9 8	3 6 7 8

【 0 0 3 3 】

【表 3】

不特定話者モデルの総分布数 - 6 0 1 状態の H M 網の場合

作成法 / 混合数	3	5	1 0	1 5
モデルクラスタリング	1 8 1 0	3 0 1 0	6 0 1 0	9 0 1 0
状態別クラスタリング	1 6 1 7	2 5 4 0	4 6 1 4	6 4 4 7

【 0 0 3 4 】 表 4 及び表 5 は各方法により作成した不特定話者モデルを用いた音素認識実験の結果である。表中の結果は男性 9 人に対する平均値を示している。

【 0 0 3 5 】

【表 4】

モデル作成法による音素認識率 (%) の比較 - 2 0 1 状態の H M 網の場合

作成法 / 混合数	5	1 0	1 5	2 0
バーム・ウェルチ	6 5 . 9	6 6 . 8	-	-
モデルクラスタリング	6 2 . 2	6 2 . 5	6 3 . 3	6 3 . 2
状態別クラスタリング	6 3 . 6	6 4 . 1	6 4 . 0	6 4 . 5
状態別クラスタリング + バーム・ウェルチ	6 8 . 0	6 8 . 6	-	-

【 0 0 3 6 】

* * 【表 5】

モデル作成法による音素認識率 (%) の比較 - 6 0 1 状態の H M 網の場合

作成法 / 混合数	3	5	1 0	1 5
バーム・ウェルチ	6 7 . 6	6 7 . 8	-	-
モデルクラスタリング	6 5 . 1	6 5 . 5	6 6 . 2	6 6 . 2
状態別クラスタリング	6 7 . 8	6 7 . 9	6 7 . 8	6 7 . 8
状態別クラスタリング + バーム・ウェルチ	6 9 . 2	6 9 . 2	-	-

【 0 0 3 7 】表 4 及び表 5 の結果を表 2 及び表 3 の結果とあわせて見ると、本発明に係る状態別クラスタリングによる方法は全ての条件のもとで第 2 の従来例の C C L 法による場合より少ないパラメータ数で高い認識性能を示しており、認識率の差は H M 網の状態数が 2 0 1 状態の場合より 6 0 1 状態の場合の方が大きくなっている。実際の認識処理のスピードや話者適応を行なう場合の効率を考えた場合できるだけ少ないパラメータ数で高い認識性能が得られる方が不特定話者モデルとしての性能は良いと考えられ、このことは、本発明に係る状態別クラスタリングによる方法が性能の良いモデルを得るのに有効な方法であることを示している。

【 0 0 3 8 】また、H M 網の状態数と認識性能の関係を見た場合、6 0 1 状態の H M 網は 2 0 1 状態の H M 網より高い認識性能を示しており、これは、第 2 の従来例の C C L 法及び、本発明に係る状態別クラスタリング法のどちらの場合にも同様のことが言える。これは、2 0 1 状態ではまだ音韻環境が十分に細分化されてモデル化されていないことが原因であると考えられる。音韻環境が十分に細分化されるように状態分割されていなければ、各状態の出力ガウス分布は音韻環境及び話者環境の両方の要因による音響的特徴量の変動を同時に表現しなければならず、音韻性と話者性の区別が難しくなり、認識誤りの可能性が高くなると考えられる。

【 0 0 3 9 】さらに、表 4 及び表 5 から明らかのように、本発明に係る状態別クラスタリング法でクラスタリングした後バーム・ウェルチの学習アルゴリズムを用いて再学習した場合、他の方法に比較してより高い音素認識率が得られている。

【 0 0 4 0 】最後に、不特定話者モデルの作成時間について述べる。従来文献 2 において開示された第 2 の従来例の C C L 法では、バーム・ウェルチの学習アルゴリズムの数パーセント程度の計算時間しか要しないと報告されている。本発明に係る状態別クラスタリングを用いる場合にはクラスタリングを行なう回数が増える分、第 2

の従来例の C C L 法に比較して計算時間が増加するが、この時間はモデル作成に要する時間の大部分を占める特定話者モデルの学習時間に比較すると非常に小さいため、全体の時間で見ただけの場合には、第 2 の従来例の C C L 法と同様にバーム・ウェルチの学習アルゴリズムの数パーセント程度の計算時間で不特定話者モデルを作成可能である。

【 0 0 4 1 】以上説明したように、本発明に係る実施形態によれば、入力された複数の特定話者の単一ガウス分布の H M M の各状態の出力ガウス分布を各状態ごとに独立にクラスタリングして合成することにより不特定話者の混合ガウス分布の H M M を作成するので、各特定話者モデルの全てのパラメータが学習されている必要はなく、また話者ごとに学習されているパラメータが異なっている場合にも対応することができる。従って、発話数が少ない話者の音声データや自由発話音声のような話者ごとに発話内容が異なるデータに対しても使用することができる。さらに、H M M の状態ごとに各特定話者モデルから取り出された出力ガウス分布の平均値のばらつきやその学習データ量の情報を利用することによって状態ごとに分割するクラスタ数を決めることができるため、学習データ量や話者間の音響的特徴の変動の度合を考慮した混合分布数を H M M の各状態ごとに決定することができる。当該不特定話者モデルの H M M を用いて音声認識することにより、従来例に比較して高い音声認識率で音声認識することができる。

【 0 0 4 2 】

【発明の効果】以上詳述したように本発明に係る請求項 1 記載の不特定話者モデル作成装置によれば、入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルに基づいて、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する不特定話者モデル作成装置において、入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルの各状態の出力ガウス分布を各状態ごとに独立にクラスタリングして合成することにより不特定

話者の混合ガウス分布の隠れマルコフモデルを作成するモデル作成手段を備える。具体的には、上記モデル作成手段は、入力された複数の特定話者の発声音声データに基づいて、複数の話者に対して同一の初期話者隠れマルコフモデルを用いて所定の学習法により上記発声音声データの存在する状態に対してのみ出力ガウス分布を学習することにより、複数個の特定話者用単一ガウス分布の隠れマルコフモデルを作成する学習手段と、上記学習手段によって作成された複数個の特定話者用単一ガウス分布の隠れマルコフモデルに基づいて、各出力ガウス分布間の距離を基準にして、各クラスタにより短い距離に出力ガウス分布が含まれるように複数のクラスタにクラスタリングを行うクラスタリング手段と、上記クラスタリング手段によって各状態毎にクラスタリングされた単一ガウス分布の隠れマルコフモデルに基づいて、各クラスタ内の複数の出力ガウス分布の隠れマルコフモデルを各状態の単一ガウス分布の隠れマルコフモデルに合成する合成手段と、上記合成手段によって合成された各状態の単一ガウス分布の隠れマルコフモデルを混合することにより、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する混合手段とを備える。

【0043】すなわち、多数の特定話者モデルから学習されている出力ガウス分布のみを取り出してHMMの各状態で独立にクラスタリングを行なうことにより、各状態における特徴量の変動の大きさや学習データ量を考慮してクラスタ数を決定することが可能となり各状態ごとに最適な出力ガウス分布数を決定することができる。また、各特定話者モデルの学習されている出力ガウス分布のみを選択的に使用することができるため各特定話者モデルの全ての出力ガウス分布が学習されている必要はなく、一人あたりの発話量の少ないデータベースに対しても有効に使用することができる。また、各話者ごとに別々にパラメータ推定を行なうため、全てのデータを一度に使う学習する第1の従来例のパーム・ウェルチの学習アルゴリズムによる方法に対して計算量を飛躍的に減らすことが可能となる。従って、不特定話者モデルの作成時間を大幅に短縮することができる。

【0044】また、請求項3記載の不特定話者モデル作成装置によれば、上記クラスタリング手段は、各状態毎に予め設定したしきい値以上のデータ量で学習された出力ガウス分布のみを取り出した後、クラスタリングする。これにより、信頼性のより高い最適な不特定話者モデルを作成することができる。従って、当該不特定話者モデルを用いて音声認識を行うことにより、従来例に比較してより高い音声認識率で音声認識することができる。

【0045】さらに、請求項4記載の不特定話者モデル作成装置によれば、上記クラスタリング手段は、各状態においてクラスタリングされた各クラスタの中心と各出力ガウス分布間の距離の平均値が予め決めた距離以下に

なるまでクラスタリングを繰り返すことにより、各状態における各出力ガウス分布のバラツキが大きいほどクラスタ数が多くなるように各状態におけるクラスタ数を決定する。従って、各状態における各出力ガウス分布のバラツキを考慮してクラスタ数を決定することが可能となり各状態ごとに最適な出力ガウス分布数を決定することができる。これにより、信頼性のより高い最適な不特定話者モデルを作成することができる。それ故、当該不特定話者モデルを用いて音声認識を行うことにより、従来例に比較してより高い音声認識率で音声認識することができる。

【0046】また、本発明に係る請求項5記載の音声認識装置によれば、入力された複数の特定話者の単一ガウス分布の隠れマルコフモデルに基づいて、不特定話者の混合ガウス分布の隠れマルコフモデルを作成する請求項1乃至4のうちの1つに記載の不特定話者モデル作成装置と、入力された発聲音声文の音声信号に基づいて、上記不特定話者モデル作成装置によって作成された不特定話者の混合分布の隠れマルコフモデルを用いて、音声認識する音声認識手段とを備える。従って、当該不特定話者モデルを用いて音声認識を行うことにより、従来例に比較してより高い音声認識率で音声認識することができる。

【図面の簡単な説明】

【図1】 本発明に係る一実施形態である音声認識装置のブロック図である。

【図2】 図1の不特定話者モデル作成部によって実行される不特定話者モデル作成処理を示すフローチャートである。

【図3】 図1の不特定話者モデル作成部によって実行される不特定話者モデル作成処理のうち特定話者モデルの学習と出力ガウス分布の抽出の処理を示す図である。

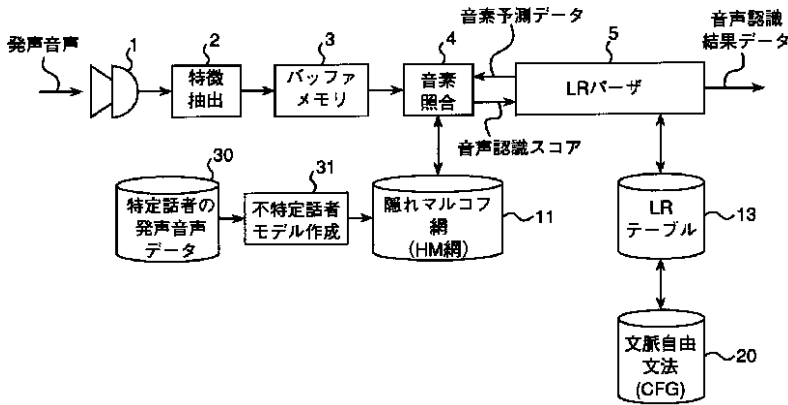
【図4】 図1の不特定話者モデル作成部によって実行される不特定話者モデル作成処理のうち各状態毎の出力ガウス分布のクラスタリングの処理を示す図である。

【図5】 図1の不特定話者モデル作成部によって実行される不特定話者モデル作成処理のうち各クラスタ毎に複数の確率密度関数を混合する処理を示す図である。

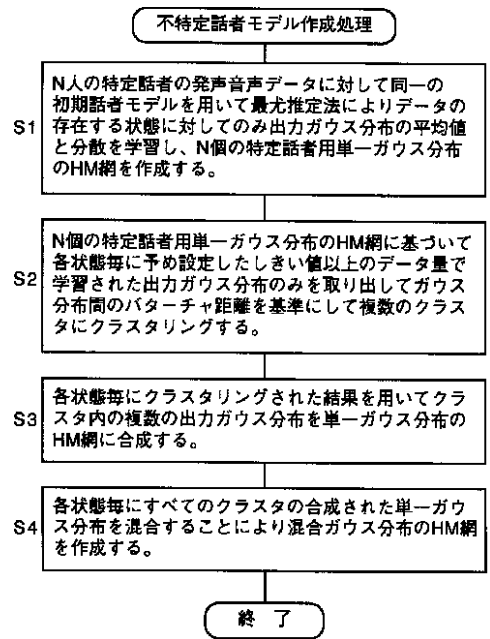
【符号の説明】

- 1...マイクロホン、
- 2...特徴抽出部、
- 3...バッファメモリ、
- 4...音素照合部、
- 5...LRパーザ、
- 11...隠れマルコフ網(HM網)、
- 13...LRテーブル、
- 20...文脈自由文法データベース、
- 30...特定話者の発聲音声データ、
- 31...不特定話者モデル作成部。

【図 1】

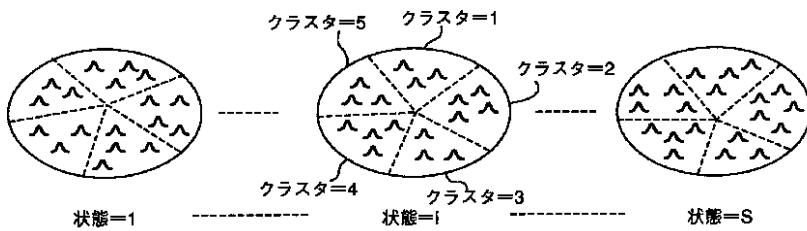


【図 2】



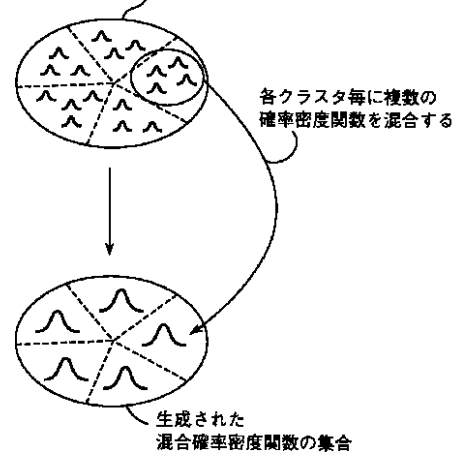
【図 4】

各状態毎の出力ガウス分布のクラスタリング

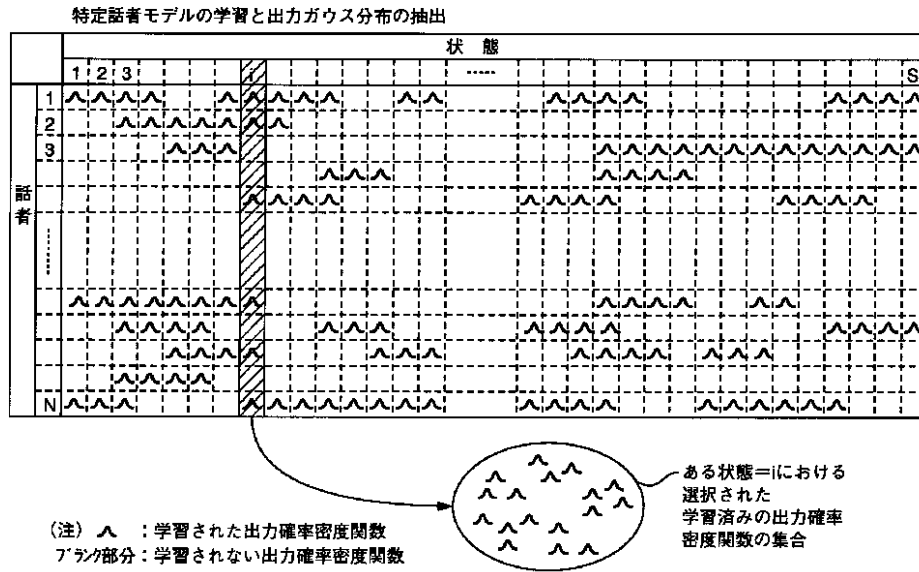


【図 5】

ある状態=iにおけるクラスタリング後の出力ガウス分布の集合



【図 3】



フロントページの続き

(56) 参考文献 特開 平 4 - 125599 (J P , A)
 特開 昭 63 - 257798 (J P , A)
 日本音響学会講演論文集 (平成 7 年 9
 月) 3 - 2 - 9 , p . 123 - 124

(58) 調査した分野(Int.Cl.⁶, DB名)
 G10L 3/00 521
 G10L 3/00 531
 G10L 3/00 535
 J I C S T ファイル (J O I S)