

(19)日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11)特許番号

第2875179号

(45)発行日 平成11年(1999) 3月24日

(24)登録日 平成11年(1999) 1月14日

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

請求項の数4 (全 9 頁)

(21)出願番号	特願平7-51045	(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)3月10日	(72)発明者	脇田 由実 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8-248976	(72)発明者	ハラルド・シンガー 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8年(1996)9月27日	(74)代理人	弁理士 青山 葆 (外2名)
審査請求日	平成7年(1995)3月10日	審査官	榎本 剛

最終頁に続く

(54)【発明の名称】 話者適応化装置及び音声認識装置

1

(57)【特許請求の範囲】

【請求項1】 話者適応前後の隠れマルコフモデルの特徴ベクトルの関係を示す移動ベクトルを用いて、話者適応学習データに基づいて初期話者モデルを話者適応して学習することにより音声認識のための隠れマルコフモデルの話者モデルを計算するための話者適応化装置において、

上記話者適応学習データが存在して話者適応学習データに基づいて話者適応された後の隠れマルコフモデルの第1の特徴ベクトルを、当該第1の特徴ベクトルと、その近傍にある話者適応された後の隠れマルコフモデルの複数の第2の特徴ベクトルとを用いて平滑化処理を実行する平滑化手段を備え、

上記平滑化手段は、上記第1の特徴ベクトルと、上記複数の第2の特徴ベクトルとの間の近傍状態と、上記第1

2

と第2の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルの特徴ベクトルの近傍状態に基づいて、上記第1の特徴ベクトルの近傍にないか、又は近傍にあっても前後の隠れマルコフモデルを構成する特徴ベクトルが近傍にない上記第2の特徴ベクトルを除外した後、上記平滑化処理を実行することを特徴とする話者適応化装置。

【請求項2】 上記平滑化手段は、上記第1の特徴ベクトルと、上記各第2の特徴ベクトルとの間の距離及び上記第1と第2の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、これらの距離が所定のしきい値未満の上記第2の特徴ベクトル以外の第2の特徴ベクトルを除外することを特徴とする請求項1記載の話者適応化装置。

10

【請求項 3】 上記平滑化手段は、上記第 1 の特徴ベクトルと、上記各第 2 の特徴ベクトルとの間の距離及び上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、より距離が短い所定数の上記第 2 の特徴ベクトル以外の第 2 の特徴ベクトルを除外することを特徴とする請求項 1 記載の話者適応化装置。

【請求項 4】 請求項 1 乃至 3 のうちの 1 つに記載の話者適応化装置と、

入力された発声音声文の音声信号に基づいて、上記話者適応化装置によって話者適応された隠れマルコフモデルの話者モデルを用いて音声認識して音声認識結果を出力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】

【産業上の利用分野】 本発明は、初期話者モデルを話者適応学習データを用いて話者適応化を行って隠れマルコフモデル（以下、HMMという。）を作成する話者適応化装置、及びそれを用いた音声認識装置に関する。

【0002】

【従来の技術】 従来、HMMを用いた音声認識装置に対して少量の学習資料によって話者適応を行なう場合、安定した適応効果を得るにはその情報不足を補うことが不可欠である。このため、不特定話者モデル等の初期話者モデルに含まれる情報を事前知識として使用する話者適応法が種々研究されている（例えば、大倉計美ほか，“混合連続分布HMMを用いた移動ベクトル場平滑化話者適応方式”，音響学会講演論文集，2-Q-17，p.p. 191-192，1992年3月（以下、文献1という。）参照。）。

【0003】 例えば、文献1に開示されている従来例の移動ベクトル場平滑化話者適応方式では、各モデル・パラメータに対して話者適応学習データを用いた推定値と初期値の差分を移動ベクトルと定義している。そして、各移動ベクトルを音響的に近傍にある移動ベクトルの情報を用いて平滑化することにより推定誤差を低減したり、対応する学習データがないことによる未学習モデル・パラメータの補間を行なって話者適応化を行っている。

【0004】

【発明が解決しようとする課題】 しかしながら、従来例における話者適応化法は、話者適応の効果は比較的高いものの、対策が困難と予想された複数の音素にわたる誤認識への対応には、いまだ不十分である。また、少数データから確実に適応させるために、学習データに存在する状態から求められた移動ベクトルを、学習データにはない近傍状態にも適応している。しかしながら、コンテキスト環境を考慮しないで近傍状態に適応すると、学習データにおけるコンテキスト特有の発声変形が、異なる

コンテキスト環境を持つ状態にも適応されるため、認識率が低下してしまう場合があるという問題点があった。

【0005】 本発明の目的は以上の問題点を解決し、適応データが少ない場合であっても誤った適応化学習を防止することができ、しかも複数の音素にわたる音声認識における誤認識を防止して、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ように話者モデルを適応化することができる話者適応化装置、及びそれを用いた音声認識装置を提供することにある。

【0006】

【課題を解決するための手段】 本発明に係る請求項 1 記載の話者適応化装置は、話者適応前後の隠れマルコフモデルの特徴ベクトルの関係を示す移動ベクトルを用いて、話者適応学習データに基づいて初期話者モデルを話者適応して学習することにより音声認識のための隠れマルコフモデルの話者モデルを計算するための話者適応化装置において、上記話者適応学習データが存在して話者適応学習データに基づいて話者適応された後の隠れマルコフモデルの第 1 の特徴ベクトルを、当該第 1 の特徴ベクトルと、その近傍にある話者適応された後の隠れマルコフモデルの複数の第 2 の特徴ベクトルとを用いて平滑化処理を実行する平滑化手段を備え、上記平滑化手段は、上記第 1 の特徴ベクトルと、上記複数の第 2 の特徴ベクトルとの間の近傍状態と、上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルの特徴ベクトルの近傍状態に基づいて、上記第 1 の特徴ベクトルの近傍にないか、又は近傍にあっても前後の隠れマルコフモデルを構成する特徴ベクトルが近傍にない上記第 2 の特徴ベクトルを除外した後、上記平滑化処理を実行する。

【0007】 また、請求項 2 記載の話者適応化装置は、請求項 1 記載の話者適応化装置において、上記平滑化手段は、上記第 1 の特徴ベクトルと、上記各第 2 の特徴ベクトルとの間の距離及び上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、これらの距離が所定のしきい値未満の上記第 2 の特徴ベクトル以外の第 2 の特徴ベクトルを除外することを特徴とする。

【0008】 さらに、請求項 3 記載の話者適応化装置は、請求項 1 記載の話者適応化装置において、上記平滑化手段は、上記第 1 の特徴ベクトルと、上記各第 2 の特徴ベクトルとの間の距離及び上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、より距離が短い所定数の上記第 2 の特徴ベクトル以外の第 2 の特徴ベクトルを除外することを特徴とする。

【0009】 また、本発明に係る請求項 4 記載の音声認識装置は、請求項 1 乃至 3 のうちの 1 つに記載の話者適応化装置と、入力された発聲音声文の音声信号に基づいて、上記話者適応化装置によって話者適応された隠れマ

ルコフモデルの話者モデルを用いて音声認識して音声認識結果を出力する音声認識手段とを備えたことを特徴とする。

【0010】

【作用】請求項1記載の話者適応化装置においては、上記平滑化手段は、上記話者適応学習データが存在して話者適応学習データに基づいて話者適応された後の隠れマルコフモデルの第1の特徴ベクトルを、当該第1の特徴ベクトルと、その近傍にある話者適応された後の隠れマルコフモデルの複数の第2の特徴ベクトルとを用いて平滑化処理を実行するが、上記平滑化手段は、上記第1の特徴ベクトルと、上記複数の第2の特徴ベクトルとの間の近傍状態と、上記第1と第2の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルの特徴ベクトルの近傍状態に基づいて、上記第1の特徴ベクトルの近傍にないか、又は近傍にあっても前後の隠れマルコフモデルを構成する特徴ベクトルが近傍にない上記第2の特徴ベクトルを除外した後、上記平滑化処理を実行する。これにより、従来例に比較して適応効果を向上させることができ、特に、複数の音素にわたる誤認識を防止することができ、従って、平滑化後の移動ベクトルを用いて話者適応された話者モデルを用いて音声認識することにより、従来に比較して高い音声認識率を得ることができる。

【0011】また、請求項2記載の話者適応化装置においては、上記平滑化手段は、好ましくは、上記第1の特徴ベクトルと、上記各第2の特徴ベクトルとの間の距離及び上記第1と第2の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、これらの距離が所定のしきい値未満の上記第2の特徴ベクトル以外の第2の特徴ベクトルを除外する。これにより、より適切に近傍でない第2の特徴ベクトルを除外することができ、複数の音素にわたる誤認識を防止することができる。

【0012】さらに、請求項3記載の話者適応化装置においては、上記平滑化手段は、好ましくは、上記第1の特徴ベクトルと、上記各第2の特徴ベクトルとの間の距離及び上記第1と第2の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、より距離が短い所定数の上記第2の特徴ベクトル以外の第2の特徴ベクトルを除外する。これにより、より適切に近傍でない第2の特徴ベクトルを除外することができ、複数の音素にわたる誤認識を防止することができる。

【0013】また、請求項4記載の音声認識装置においては、請求項1乃至3のうちの1つに記載の話者適応化装置によって話者適応された隠れマルコフモデルの話者モデルが計算される。そして、上記音声認識手段は、入力された発声音声文の音声信号に基づいて、上記話者適応化装置によって話者適応された隠れマルコフモデルの

話者モデルを用いて音声認識して音声認識結果を出力する。これにより、従来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。

【0014】

【実施例】以下、図面を参照して本発明に係る実施例について説明する。図1は、本発明に係る一実施例である音声認識装置のブロック図である。本実施例の音声認識装置は、初期話者モデルメモリ30に格納された話者クラスモデルを含む初期話者モデルを、例えば文発話テキストデータであって話者適応学習データメモリ32に格納された話者適応学習データを用いて、適応化学習して音素HMMの不特定話者音素モデルに変換して音素隠れマルコフモデルメモリ(以下、音素HMMメモリという。)11に格納する話者適応化制御部31を備え、当該音素HMMメモリ11に格納された音素HMMに基づいて音声認識を行うことを特徴とする。特に、上記話者適応化制御部31は、図2に示すように、移動ベクトルの計算処理(ステップS1)と、移動ベクトルの補間処理(ステップS2)と、音素列候補の近傍状態を検出して近傍でない移動ベクトルを除外する処理(ステップS3)と、移動ベクトルの平滑化処理(ステップS4)と、処理後の移動ベクトルを用いて話者適応化する学習処理(ステップS5)とを順次実行することを特徴とする。

【0015】本実施例において、音素照合部4に接続される音素HMMメモリ11内の音素HMMは、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

(a) 状態番号

(b) 受理可能なコンテキストクラス

(c) 先行状態、及び後続状態のリスト

(d) 出力確率密度分布のパラメータ

(e) 自己遷移確率及び後続状態への遷移確率

なお、本実施例において用いる音素HMMは、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HMMを変換して作成する。ここで、出力確率密度関数は3次元の対角共分散行列をもつ混合ガウス分布(以下、ガウス分布という。)であり、各分布は、初期話者モデルメモリ30に格納された初期話者モデルを用いて話者適応化制御部31によって学習される。

【0016】一般に連続分布型HMMによるモデルに対して少量の適応データにより話者適応を行なう場合、ガウス分布の平均値の適応は他のパラメータの適応に比べて効果が大きいことが知られている(例えば、文献1参照)。本実施例においては、各ガウス分布の平均値のみの適応を行ない、分散値、状態遷移確率及び、混合ガウス分布の重み係数の適応は行なわない。

【0017】話者適応化制御部31における具体的な話者適応化処理を、図2及び図3を参照して以下に説明する。この話者適応化処理では、まず、初期話者モデルメ

メモリ 3 0 に格納された初期音素 HMM である初期話者モデルを、話者適応学習データメモリ 3 2 に格納された例えば文発話テキストデータを含む話者適応学習データ（以下、学習データという。）を用いて学習する。ここでは、文発話テキストデータに対応する音素ラベル系列に従って音素 HMM を連結して文 HMM を作成し、この文 HMM を上記話者適応学習データである文発話データを用いて学習した後、再び音素 HMM の単位に切り離すことにより、音素 HMM の学習を行う。

【0 0 1 8】すなわち、この話者適応化処理では、未知話者の音声に含まれる音素に関して、標準話者の音素 HMM の平均値を再学習する。まず、標準話者の音素 HMM を未知話者の音素 HMM の初期話者モデルとする。そして、未知話者の入力音声の音素系列に対応するように未知話者の HMM を連結し、HMM の遷移確率、出現確率の平均と分散、及び分岐確率のうち平均のみを連結学習する。具体的には、連結学習前後の HMM の平均ベクトルの差分を移動ベクトルとみなし、学習されなかった HMM の平均ベクトルの移動ベクトルを補間し平均ベクトルを移動するものである。

【0 0 1 9】まず、ステップ S 1 において、以下のよう
に移動ベクトルの計算を行う。初期話者モデル内の未知話者の全音素 HMM のガウス分布の平均ベクトルの組 $(C^I = c_1^I, \dots, c_K^I)$ 、ここで、K は全てのガウス分布の個数である。)のうち学習された k 番目の平均ベクトル c_k^I ($k \in K_1, K_1$: 学習音声中に存在した音素の HMM の平均ベクトルの番号の集合)と、話者適応学習データ内の標準話者のガウス分布の平均ベクトルの組 C^R 中で対応する c_k^R より、平均ベクトルの差分ベクトル v_k を計算し、これを話者空間の移動ベクトルとする。

【数 1】 $v_k = c_k^I - c_k^R, k \in K_1$

【0 0 2 0】ここで、 K_1 は各ガウス分布のうち学習データの存在したものの集合である。これを図示すると図 3 のようになる。図 3 に示すように、適応学習前の初期話者モデルの音響空間 A S 1 において例えば 3 個のガウス分布が存在する一方、適応学習後の話者モデルの音響空間 A S 2 において例えば 3 個のガウス分布が存在するとき、適応学習前のガウス分布の平均ベクトル c_k^R が適応学習後のガウス分布の平均ベクトル c_k^I に適応化学習されることになる。

【0 0 2 1】次いで、図 2 のステップ S 2 においては、以下の通り、移動ベクトルの補間処理を実行する。すなわち、未知話者の全音素 HMM のガウス分布の平均ベクトルの組 C^I のうち、話者適応学習データが存在しなかった音素に対する未学習の HMM に属するガウス分布の平均ベクトル c_n^I (ここで、 $n \in K_2$ であり、 K_2 は各ガウス分布のうち話者適応学習データの存在しなかったものの集合である。)を、学習された k 番目 ($k \in K_1$) の移動ベクトル v_k と、平均ベクトル c_n^R と平均ベク

トル c_n^R 間のファジイ級関数 $\mu_{n,k}$ から求めた移動ベクトル v_n を用いて c_n^I に移動する。

【0 0 2 2】

【数 2】

$$v_n = \sum_{k \in K_1} \mu_{n,k} v_k$$

【数 3】

$$c_n^I = c_n^R + v_n$$

【数 4】

$$\mu_{n,k} = 1 / [\sum_{j \in K_1} (d_{n,k} / d_{n,j})^{1/(m-1)}]$$

【0 0 2 3】ここで、 $d_{n,k}$ は、平均ベクトル c_n^R と平均ベクトル c_k^R の距離を表す。上記の移動ベクトルの計算処理と補間処理を図 4 を用いて説明する。図 4 (a) 及び (b) は、すべての HMM に含まれるガウス分布の合計が 4 個である場合について示してある。連結学習により平均ベクトル c_1^R, c_2^R, c_3^R がそれぞれ、平均ベクトル c_1^I, c_2^I, c_3^I にそれぞれ移動し、平均ベクトル c_n^R は学習されなかった場合を示している。この場合の平均ベクトル c_n^I は、 c_1^R, c_2^R, c_3^R と移動ベクトル v_1, v_2, v_3 及びファジイ級関数 $\mu_{n,1}, \mu_{n,2}, \mu_{n,3}$ を用いて計算される。

【0 0 2 4】図 4 の (a) に示すように、話者適応学習データが存在しなかった未学習のガウス分布の平均ベクトル c_n^R の近傍に 3 個の平均ベクトル c_1^R, c_2^R, c_3^R が存在する。そして、図 4 の (b) に示すように、これらの移動ベクトル v_n ($n = 1, 2, 3$) に基づいて、数 3 を用いて平均ベクトル c_n^R の移動ベクトル v_n を求めて、移動ベクトルの補間処理を行い、未学習のガウス分布の平均ベクトル c_n^I を求めている。

【0 0 2 5】上述のステップで得られたモデルは、十分な適応語数が得られていない場合に推定誤差を含んでいる。このような推定誤差を含むものから求められた移動ベクトルの方向は、非連続的な動きをしていると考えられる。そこで、話者空間を移動するための移動ベクトルに連続性の拘束条件を入れ、移動ベクトルの方向性を揃える、すなわち平滑化を行うことにより推定誤差の吸収を行う。

【0 0 2 6】本実施例においては、平滑化処理の前に、図 2 のステップ S 3 において、音素列候補の近傍状態を検出して音素列候補のコンテキストを考慮して近傍でない移動ベクトルを除外する処理を実行する。ここでは、平均ベクトル c_k^R の近傍にある平均ベクトル c_m^R と c_k^R との間のファジイ級関数 $\mu_{k,m}$ ($k \in m$) を計算する。そして、コンテキスト環境を考慮して上記距離 $d_{n,k}$ が所定のしきい値未満の音素列候補以外の音素列候補を除外する。

10

20

30

40

50

【0027】このステップS3の処理の一例を表1の音素列候補を用いて説明する。表1の各音素列候補のうち真ん中の音素がいま適応すべき音素であるとし、初期話者モデルの音素列をA1とする。

【0028】

【表1】

学習データ中の音素列候補

A1 : o / a / o

A2 : i / a / i

A3 : o / e / o

A4 : i / e / i

【0029】ここで、音素aを構成する特徴ベクトルの中でn番目の特徴ベクトル $c^R(a, n)$ と、音素eを構成する特徴ベクトルの中でm番目の特徴ベクトル $c^R(e, m)$ との間の距離は比較的短く近傍に存在するが、音素oを構成するすべての特徴ベクトル $c^R(o, /)$ と、音素iを構成するすべての特徴ベクトル $c^R(i, /)$ との距離は、どの1対1の組み合わせをとっても近傍にないと仮定する。ここで、ベクトル間の距離は、例えば特徴ベクトル $c^R(a, n)$ と特徴ベクトル $c^R(e, m)$ との距離は、次の数5で表されて計算することができる。

【0030】

【数5】

$$d = c^R(a, n) - c^R(e, m)$$

【0031】まず、適応すべき音素について処理すると、音素aのn番目の特徴ベクトル $c^R(a, n)$ と音素eのm番目の特徴ベクトル $c^R(e, m)$ とは近傍にあるので、従来例の方法では、4つの音素列を構成する遷移状態系列が、平滑化の対象となる。本実施例においては、適応すべき音素の前後の音素がそれぞれ対応する学習データのコンテキスト、すなわち対応する前後の音素と近傍であるか否かが判断される。この例では、音素oと音素iのすべての特徴ベクトルが近傍にないので、音素列候補A2とA4とが除外される。すなわち、ステップS3の処理では、適応すべき音素が学習データの遷移状態系列の候補の対応する状態と同一又は近傍の状態であっても、適応すべき状態の前後の状態が、学習データの状態系列の候補の対応する前後の状態と近傍でない状態である場合は学習データの移動ベクトルから除外する。言い換えれば、コンテキストに基づいて近傍でない移動ベクトルを除外することになる。

【0032】本実施例においては、コンテキスト環境を考慮して上記距離 $d_{n, k}$ が所定のしきい値未満の状態系列以外の状態系列を、近傍にないとして除外しているが、本発明はこれに限らず、適応すべき状態の前後の状態が、その近傍にある所定数の近傍の状態に含まれないときに、近傍にないと判断して、学習データの移動ベク

トルから除外してもよい。また、これらの近傍状態を例えばファジイ級関数を用いて重み付けしてもよい。

【0033】さらに、ステップS4の平滑化処理においては、平均ベクトル c_k^I とその近傍にあるm番目の平均ベクトル c_m^I と c_m^R の差分ベクトル v_m を求める。次に、ファジイ級関数 $\mu_{k, m}$ を用いて、差分ベクトル v_m に平滑化処理を行い、次の数6を用いて平滑化移動ベクトル v_k^S を求める。

【0034】

10 【数6】

$$v_k^S = \left\{ \sum_{m \in N(k)} \alpha_m \mu_{k, m} v_m \right\} / \left\{ \sum_{m \in N(k)} \alpha_m \mu_{k, m} \right\}$$

【0035】ここで、 $N(k)$ は平均ベクトル c_k^R の近傍にある平均ベクトルの番号であり、 α_m は v_m の信頼度を与える定数であり、 $k = m$ の場合、 $\mu_{k, m} = 1$ とする。

【0036】最後に、ステップS5においては、処理後の移動ベクトル v_k^S と平均ベクトル c_k^R を用いて、次の数7に示すように、平均ベクトル c_k^R を初期話者モデルの未知話者へ話者適応する。すなわち計算された移動ベクトル v_k^S を用いて、メモリ30に格納された初期話者モデルを話者適応することにより学習し、これによ

20

【0037】

$$【数7】 c_s^k = c_k^R + v_k^S$$

【0038】ここで、 c_s^k は、平滑化を行って得られた話者適応後の音素HMMのガウス分布の平均ベクトルである。本実施例においては、 $\alpha_m = 1 (m = K_1)$ 、 $\alpha_m = 0 (m = K_2)$ とした。また、ファジイ級関数 $\mu_{k, m} : (k = m)$ は、 $m = K_1$ である平均ベクトル c_m^R 全てを用いて求めた。

30

【0039】上記の処理を図5を用いて説明する。図5は、全てのHMMに含まれるガウス分布の合計が4個である場合について示してある。ステップS3乃至S5による処理により、平均ベクトル $c_1^R, c_2^R, c_3^R, c_k^R$ が $c_1^I, c_2^I, c_3^I, c_k^I$ にそれぞれ移動したとする。いま、 c_k^I に対応する移動ベクトル v_k を考える。移動ベクトル v_k は、 v_1, v_2, v_3, v_k とそれぞれに対応するファジイ級関数と各移動ベクトルに対する信頼性の重み α_m により平滑化され v_k^S が計算される。

40

【0040】ただし、図5において、ステップS3の処理については図示していないが、上記ステップS3における処理を要約すると、以下の通りである。ステップS4において、話者適応用学習データが存在して話者適応用学習データに基づいて話者適応された後のHMMのガウス分布の平均ベクトル(以下、第1の平均ベクトルという。)を、当該平均ベクトルと、その近傍にある話者適応された後のHMMのガウス分布の複数の平均ベクトル(以下、第2の平均ベクトルという。)とを用いて平

50

滑化処理を実行するが、この前に、上記平均ベクトルの遷移状態系列と、上記複数の第 2 の平均ベクトルの遷移状態系列との間の近傍状態に基づいて、上記第 1 の平均ベクトルの遷移状態系列の近傍にない上記第 2 の平均ベクトルを除外する。

【0041】本実施例においては、音声認識のための統計的音素モデルセットとして音素HMMを使用している。当該音素HMMは効率的に表現された音素環境依存モデルである。この音素HMM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では音素HMMのトポロジーの決定、異音クラスの決定、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。

【0042】次いで、上述の本実施例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11に格納された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。

【0043】図1において、話者適応制御部31は、初期話者モデルメモリ30に格納された話者クラスモデルを含む初期話者モデルを、例えば文発話テキストデータであって話者適応学習データメモリ32に格納された話者適応学習データを用いて図2に示す話者適応化処理により移動ベクトルを計算し、計算した移動ベクトルを用いて適応化学習してHMMの不特定話者音素モデルに変換して音素HMMメモリ11に格納する。一方、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。

【0044】音素照合部4は、音素コンテキスト依存型LR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、音素HMMメモリ11に格納された音素HMMの話者モデルを用いて音素照合区間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LRパーザ5に返される。このとき、前向きパスアルゴリズムを使用する。

【0045】一方、文脈自由文法データベースメモリ20内の所定の文脈自由文法 (CFG) を公知の通り自動的に変換してLRテーブルを作成してLRテーブルメモ

リ13に格納される。LRパーザ5は、上記LRテーブルメモリ13内のLRテーブルを参照して、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。LRパーザ5は、LRテーブルメモリ13内のLRテーブル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音素HMMメモリ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LR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う。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う高速処理を実現する。

【0046】本発明者は、本実施例の音声認識装置の評価を行うために、以下のようにシミュレーションを行った。このシミュレーションには200状態のHM網を使用した。このHM網は、モデルパラメータを効果的に共有したコンテキスト依存型の音素HMMモデルのネットワークである。このシミュレーションにおいては、コンテキスト特有の適応を考慮することで、より正確な適応を行なえることを確認するため、学習データにより求められた移動ベクトルを、コンテキスト環境が同じ近傍状態に対してのみ適応した場合の適応効果を調べてみた。従来例のVFS法の利点を生かしながら、コンテキスト環境を考慮するために、一度従来例のVFS法にて話者適応を行なったHMMに対して、さらに次の2つの話者適応を行ない、適応効果を比較した。

(1) コンテキスト環境を考慮せず、再度従来例のVFS法で話者適応処理を行なった場合 (従来例)。

(2) 上述のごとく、コンテキスト環境が同じ近傍状態に対してのみ移動ベクトルを応用して、話者適応を行なった場合 (本実施例)。

このシミュレーション結果を表2に示す。ここで、テキストクローズドとは、初期話者モデルに学習データを含む場合であり、テキストオープンとは、初期話者モデルに学習データを含まない場合である。

【0047】

【表2】

	従来例	本実施例
テキストクローズド	87.1%	88.1%
テキストオープン	76.8%	78.7%

【0048】表2に示すシミュレーションの認識結果から明らかなように、クローズデータ、オープンデータともに、コンテキストを考慮したときの本実施例の認識率

10

20

30

40

50

は、コンテキスト環境を考慮しない場合の従来例に比較して向上しており、特に、オープンデータの向上率は高い。これは、オープンデータが、異なる環境のコンテキストを多く含んでいるからであると思われる。

【0049】以上説明したように、コンテキスト環境を考慮して移動ベクトルを適応することにより、コンテキスト環境を考慮しない従来例の場合に比べて適応効果は向上し、特に、複数音素にわたる誤認識に対して効果があり、従来例に比較して高い音声認識率を得ることができる。

【0050】なお、話者適応制御部 31 と、音素照合部 4 と、LR パーザ 5 は、例えばデジタル計算機により構成される。

【0051】以上の実施例においては、適応すべき音素の前後の 1 つずつの音素について図 2 のステップ S3 の処理を行っているが、本発明はこれに限らず、適応すべき音素の前後の複数の音素について処理を行ってもよい。以上の実施例においては、HMM のガウス分布の平均ベクトルの移動ベクトルを用いて話者適応した後平滑化処理しているが、ガウス分布の平均ベクトル以外の、例えば VQ コードブックなどの特徴ベクトルを用いてもよい。

【0052】

【発明の効果】以上詳述したように本発明に係る請求項 1 記載の話者適応化装置によれば、話者適応前後の隠れマルコフモデルの特徴ベクトルの関係を示す移動ベクトルを用いて、話者適応学習データに基づいて初期話者モデルを話者適応して学習することにより音声認識のための隠れマルコフモデルの話者モデルを計算するための話者適応化装置において、上記話者適応学習データが存在して話者適応学習データに基づいて話者適応された後の隠れマルコフモデルの第 1 の特徴ベクトルを、当該第 1 の特徴ベクトルと、その近傍にある話者適応された後の隠れマルコフモデルの複数の第 2 の特徴ベクトルとを用いて平滑化処理を実行する平滑化手段を備え、上記平滑化手段は、上記第 1 の特徴ベクトルと、上記複数の第 2 の特徴ベクトルとの間の近傍状態と、上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルの特徴ベクトルの近傍状態に基づいて、上記第 1 の特徴ベクトルの近傍にないか、又は近傍にあっても前後の隠れマルコフモデルを構成する特徴ベクトルが近傍にない上記第 2 の特徴ベクトルを除外した後、上記平滑化処理を実行する。これにより、従来例に比較して適応効果を向上させることができ、特に、複数の音素にわたる誤認識を防止することができ、従って、平滑化後の移動ベクトルを用いて話者適応された話者モデルを用いて音声認識することにより、従来に比較して高い音声認識率を得ることができる。

【0053】また、請求項 2 記載の話者適応化装置においては、上記平滑化手段は、好ましくは、上記第 1 の特

徴ベクトルと、上記各第 2 の特徴ベクトルとの間の距離及び上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、これらの距離が所定のしきい値未満の上記第 2 の特徴ベクトル以外の第 2 の特徴ベクトルを除外する。これにより、より適切に近傍でない第 2 の特徴ベクトルを除外することができ、複数の音素にわたる誤認識を防止することができる。

【0054】さらに、請求項 3 記載の話者適応化装置においては、上記平滑化手段は、好ましくは、上記第 1 の特徴ベクトルと、上記各第 2 の特徴ベクトルとの間の距離及び上記第 1 と第 2 の特徴ベクトルを含む隠れマルコフモデルの前後の隠れマルコフモデルを構成する特徴ベクトル間の距離を計算し、より距離が短い所定数の上記第 2 の特徴ベクトル以外の第 2 の特徴ベクトルを除外する。これにより、より適切に近傍でない第 2 の特徴ベクトルを除外することができ、複数の音素にわたる誤認識を防止することができる。

【0055】また、請求項 4 記載の音声認識装置においては、請求項 1 乃至 3 のうちの 1 つに記載の話者適応化装置によって話者適応された隠れマルコフモデルの話者モデルが計算される。そして、上記音声認識手段は、入力された発声音声文の音声信号に基づいて、上記話者適応化装置によって話者適応された隠れマルコフモデルの話者モデルを用いて音声認識して音声認識結果を出力する。これにより、従来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。

【図面の簡単な説明】

【図 1】 本発明に係る一実施例である音声認識装置のブロック図である。

【図 2】 図 1 の話者適応制御部 31 によって実行される話者適応化処理を示すフローチャートである。

【図 3】 移動ベクトルを用いて図 2 の話者適応化処理を実行する場合における、適応学習前の初期話者モデルの音響空間 AS1 から適応学習後の話者モデルの音響空間 AS2 への変換を示す概念図である。

【図 4】 (a) は、図 2 のステップ S2 で実行される移動ベクトルの計算処理を示す概念図であり、(b) は、図 2 のステップ S3 で実行される移動ベクトルの補間処理を示す概念図である。

【図 5】 図 3 のステップ S4 で実行される移動ベクトルの平滑化処理を示す概念図である。

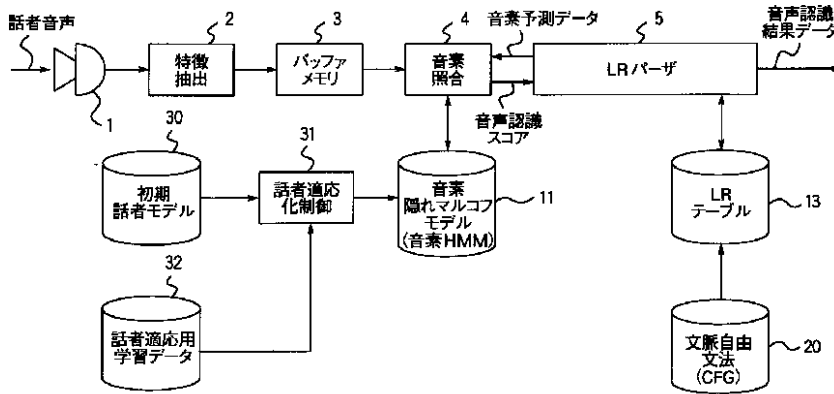
【符号の説明】

- 1 ... マイクロホン、
- 2 ... 特徴抽出部、
- 3 ... パツファメモリ、
- 4 ... 音素照合部、
- 5 ... LR パーザ、
- 11 ... 音素 HMM メモリ、
- 13 ... LR テーブルメモリ、

- 2 0 ... 文脈自由文法データベースメモリ、
- 3 0 ... 初期話者モデルメモリ、
- 3 1 ... 話者適応化制御部、
- 3 2 ... 話者適応学習データメモリ、
- S 1 ... 移動ベクトルの計算処理、
- S 2 ... 移動ベクトルの補間処理、

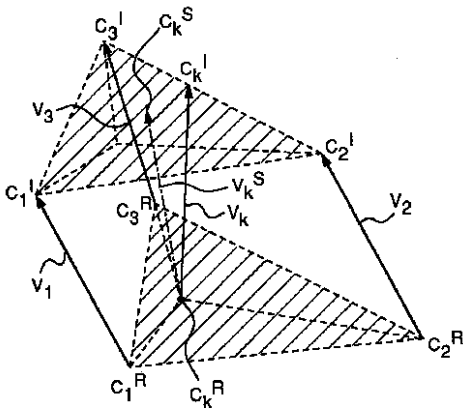
- * S 3 ... 音素列候補の近傍状態を検出して音素列候補のコンテキストを考慮して近傍でない移動ベクトルを除外する処理、
- S 4 ... 移動ベクトルの平滑化処理、
- S 5 ... 処理後の移動ベクトルを用いて話者適応化する処理。

【図 1】

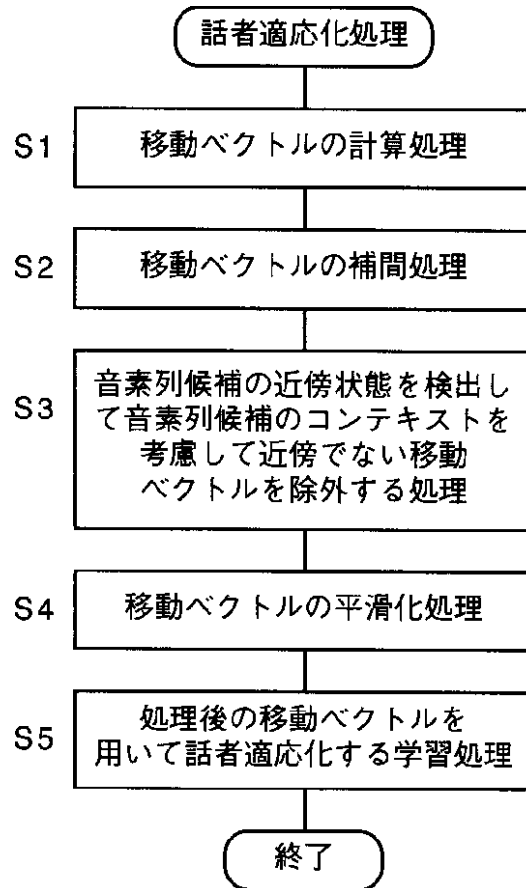


【図 5】

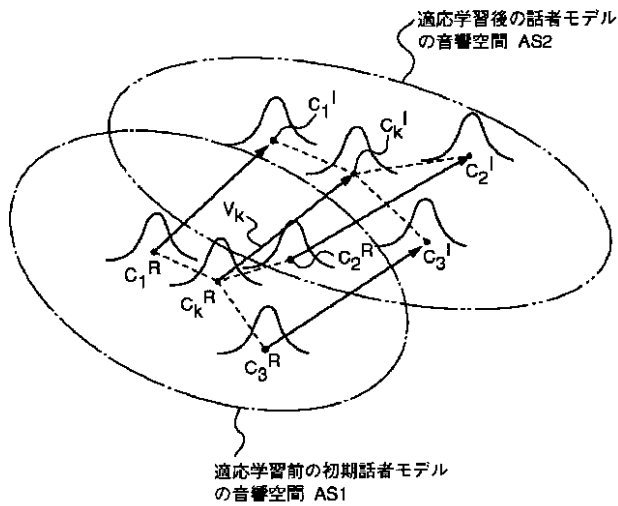
移動ベクトルの平滑化 $V_k \rightarrow V_k^S$



【図 2】

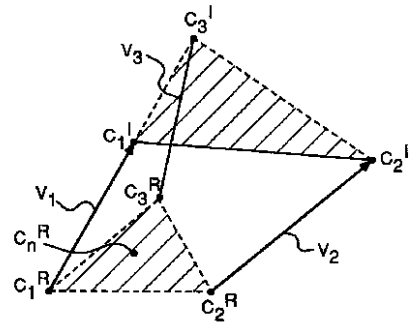


【図 3】

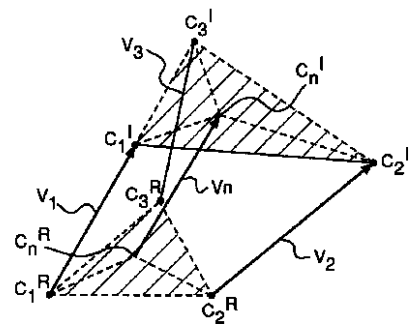


【図 4】

(a) 移動ベクトルの計算



(b) 移動ベクトルの補間



フロントページの続き

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

(56)参考文献 特開 平 6 - 60048 (J P , A)
 日本音響学会講演論文集 (平成 7 年 3
 月) 2 - 5 - 7 , p . 43 ~ 44
 電子情報通信学会技術研究報告「音
 声」 S P 95 - 30 , p . 41 ~ 47

(58)調査した分野(Int.Cl.⁶ , D B 名)

G10L 3/00 531
 G10L 3/00 535
 G10L 3/00 521

J I C S T ファイル (J O I S)