(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3589044号
(P3589044)
(45)発行日　平成16年11月17日(2004.11.17)　(24)登録日　平成16年8月27日(2004.8.27)

(51)Int.Cl.[7]　F I
G10L 15/10　G10L 3/00　531K
G10L 15/06　G10L 3/00　521S

請求項の数 3　(全 28 頁)

(21)出願番号　特願平10-297924
(22)出願日　平成10年10月20日(1998.10.20)
(65)公開番号　特開2000-122689(P2000-122689A)
(43)公開日　平成12年4月28日(2000.4.28)
審査請求日　平成13年2月27日(2001.2.27)

(73)特許権者　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号
(74)代理人　100113077
弁理士　高橋　省吾
(74)代理人　100112210
弁理士　稲葉　忠彦
(74)代理人　100108431
弁理士　村上　加奈子
(74)代理人　100128060
弁理士　中鶴　一隆
(72)発明者　石井　純
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内

最終頁に続く

(54)【発明の名称】話者適応化装置

(57)【特許請求の範囲】
【請求項１】
話者の入力音声から抽出した音声特徴量と、多数の話者の音声データによりパラメータ学習を行って得た標準パタンと、を照合して認識結果を出力するとともに、前記標準パタンを、前記入力音声を発した話者に適応した話者適応パタンに更新するか否かを、前記認識結果の信頼度に応じて決定する話者適応化装置において、
前記音声特徴量と前記標準パタンから推定適応パタンを最尤推定により算出する標準パタンパラメータ最尤推定手段と、
この標準パタンパラメータ最尤推定手段により算出された推定適応パタンを構成するパラメータの値と前記標準パタンを構成するパラメータの値とを、前記信頼度に応じて線形補間することにより前記話者適応パタンを算出するパラメータ線形補間手段と、
を備えたことを特徴とする話者適応化装置。
【請求項２】
話者の入力音声から抽出した音声特徴量と、多数の話者の音声データによりパラメータ学習を行って得た標準パタンと、を照合して認識結果を出力するとともに、前記標準パタンを、前記入力音声を発した話者に適応した話者適応パタンに更新するか否かを、前記認識結果の信頼度に応じて決定する話者適応化装置において、
前記信頼度に基づいて、前記入力音声から得られる話者適応用データのパラメータ学習への重み付けを計算し、重み付けされた話者適応用データを用いて、前記標準パタンからを構成するパラメータを、前記話者適応パタンを構成するパラメータに更新する適応学習

手段を備えたことを特徴とする話者適応化装置。
【請求項３】
話者の入力音声から抽出した音声特徴量と、多数の話者の音声データによりパラメータ学習を行って得た標準パタンと、を照合して認識結果を出力するとともに、前記標準パタンを、前記入力音声を発した話者に適応した話者適応パタンに更新するか否かを、前記認識結果の信頼度に応じて決定する話者適応化装置において、
過去に出力した前記認識結果の信頼度の値に基づいて、異なる話者適用学習アルゴリズムを選択する話者適用方式選択手段を備えたことを特徴とする話者適応化装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、多数の話者の音声データによりパラメータ学習を行った標準パタンを、ある話者に適応した話者適応パタンに更新するようにした教師なし話者適応化装置、及びその話者適応パタンを用いた音声認識装置に関する。
【０００２】
【従来の技術】
音声認識のアプリケーションを想定した場合、事前の話者音声の登録を必要としない不特定話者音声認識システムの要望が高く、隠れマルコフモデル（ＨｉｄｄｅｎＭａｒｋｏｖ　Ｍｏｄｅｌ、以下HMMとする）、ニューラルネット（Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ、以下ＮＮとする）を用いた音声認識方式による実用化検討が行われている。HMM、ＮＮの詳細は、例えば「音声認識の基礎（上、下）」Ｌ．ＲＡＢＩＮＥＲ、Ｂ．Ｈ．ＪＵＡＮＧ、古井　監訳、１９９５年、１１月、ＮＴＴアドバンステクノロジ　（以下、文献１という）に記されている。これらの方法は、予め多数の話者からの単語、文などの音声データを用いた、標準パタンの学習によって不特定話者標準パタンを作成するものである。
【０００３】
しかしながら、HMMやＮＮによる不特定話者音声認識システムは、特定話者に限定した場合、その特定話者からの単語、文などの音声データによって標準パタンを学習した特定話者認識システムと比較して、単語誤り率で２～３倍程度であるのが現状である。そこで不特定話者音声認識システムの向上をはかるため、話者適応技術の研究が最近盛んに行われている。
【０００４】
話者適応化技術は、特定話者の少量の音声データ（以下適応データとする）を用いて、音声認識システムを使用する前や使用中に、不特定話者音声認識システムの標準パタンのパラメータを適応学習して認識率の向上を図るものである。話者適応化方式については、「音声認識における話者適応」松本　弘、日本音響学会平成７年春季研究発表会講演論文集、ｐｐ．２７－３０１９９５年3月（以下、文献２という）に詳しい。話者適応化法としては、適応学習データの発話の内容が既知の音声を用いるか、あるいは任意の未知の発話内容の音声を使用するかにより、「教師あり／教師なし」の２つの方法がある。教師あり話者適応方式は、適応データを用いた適応学習後の認識精度は高いが、音声認識装置の使用者が使用前に予め決められた単語や文章を発声しなければならず、使用者の負担が大きい。一方、教師なし話者適応方式は、音声認識装置の使用中に使用者が適応学習を意識することなく認識率の改善を得ようとする方法である。実際の音声認識のアプリケーションでは、教師なし話者適応の確立が望まれている。
【０００５】
従来の教師なし適応化では、入力音声に対して不特定話者用の標準パタンを用いて照合を行い、照合を行った結果として得られる認識結果を発声内容であるとして、不特定話者用標準パタンを連結し、入力音声を適応データとして標準パタンのパラメータを更新する。例えば「Ｓｐｅａｋｅｒ　Ａｄａｐｔａｔｉｏｎ　ｏｆ　ＣｏｎｔｉｎｕｏｕｓＤｅｎｓｉｔｙ　ＨＭＭｓ　Ｕｓｉｎｇ　Ｍｕｌｔｉｖａｒｉａｔｅ　Ｌｉｎｅａｒ　Ｒｅｇｒｅｓｓｉｏｎ」Ｃ．Ｌ．Ｌｅｇｇｅｔｔｅｒ　ａｎｄＰ．Ｃ．Ｗｏｏｄｌａｎ

ｄ，　Ｐｒｏｃ．　ｏｆ　ＩＣＳＬＰ９４、ｐｐ．４５１－４５４、１９９４年　（以下、文献３という）で報告されている。

【０００６】

以下に従来例として文献３に記述されている認識結果を発声内容とする教師なし話者適応化装置を図２１のブロック図を参照して説明する。図２１において、入力音声２００１は、認識装置の使用話者が発声した単語や文章の音声である。ここでの１発声はポーズからポーズの間の文節や文章として説明を行う。

【０００７】

音声特徴量抽出手段２００２は入力音声２００１の音声信号をＡ／Ｄ変換し、Ａ／Ｄ変換された信号を５ミリ秒～２０ミリ秒程度の一定時間間隔のフレームで切り出し、音響分析を行って音声特徴量を抽出する。ここで音声特徴量とは、少ない情報量で音声の特徴を表現できるものであり、例えばケプストラム、ケプストラムの動的特徴の物理量で構成する特徴量ベクトルである。

【０００８】

照合手段２００３では、認識辞書２００５でテキスト表記によって設定している認識対象の単語　［Ｗ（１），　Ｗ（２），　．．．，　Ｗ（ｗｎ）］（括弧内は単語番号、ｗｎは認識対象単語数）を認識ユニットのラベル表記へ変換し、ラベルに対応した認識ユニットの標準パタン２００４を連結することで認識対象単語の標準パタンを作成する。そして音声特徴量抽出手段２００２からの出力である発声１から発声Ｎまでの音声特徴量の時系列　Ｏ　＝　［ｏ（１），　ｏ（２）　，．．．，　ｏ（Ｔ）］（括弧内は時刻、Ｔは最大フレーム数）に対して照合を行い、話者適応学習用音声認識結果２００６を出力する。話者適応学習用音声認識結果２００６は発声に対して最も照合スコア（尤度とも言う）が高い単語番号系列　Ｒｎ■＝［ｒ■（１），　ｒ■（２）　，．．．，　ｒ■（ｍ■）］を計算し、単語番号に対応した単語のテキスト表記Ｒｗ■＝［Ｗ（ｒ■（１）），　Ｗ（ｒ■（２））　，．．．，　Ｗ（ｒ■（ｍ■））］を出力する。ここで、ｒ■（ｉ）は話者適応学習用音声認識結果２００６の単語列中のｉ番目の単語の単語番号を示す。また、ｍ■は話者適応用音声認識結果２００６の単語列数を示す。

【０００９】

標準パタン２００４は、予め用意した標準パタンであり、文献３では認識ユニットを前後音素環境　（コンテキスト）　依存の音素としたＨＭＭを用い、多数の話者の音声データでパラメータ学習を行った標準パタンを初期の標準パタンとして使用している。ＨＭＭは状態単位で以下の情報をパラメータとして有することで複数の認識ユニットの標準パタンを形成する。

（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）自己遷移確率確率及び後続状態への遷移確率

【００１０】

認識辞書２００５は、予め定めた認識対象とする単語や文章をテキストで格納し、テキスト表記から認識ユニットラベルへの変換を行って、このラベル系列にしたがって標準パタン２００４から対応する認識ユニット標準パタンを連結して照合手段２００３で用いる認識対象単語の標準パタンを生成する。例えば認識辞書２００５に「あお」が存在するならば、これは音素系列で表した場合は／ａｏ／となる。離散発声の「あお」の認識に用いる標準パタンは中心音素が／ａ／であり、先行音素が無音、後続音素が／ｏ／である認識ユニットのＨＭＭ　λ－ａｏと、中心音素が／ｏ／であり、先行音素が／ａ／、後続音素が無音の認識ユニットのＨＭＭ　λａｏ－を連結したＨＭＭによって照合を行う。最近ではこのような前後音素環境依存の音素ＨＭＭを用いて、認識対象語彙が４０，０００単語以上の音声認識システムの検討が行われている。

【００１１】

教師なし話者適応手段２００７は、照合手段２００３の出力である話者適応学習用音声認識結果２００６と標準パタン２００４を入力し、認識結果の認識ユニットラベル系列に基づき、標準パタン２００４の音素ＨＭＭを連結し、音声特徴量抽出手段２００２からの出力である音声特徴量の時系列を適応データとして標準パタンのパラメータを更新し、教師なし話者適応パタン２００８を出力する。

【００１２】

文献３では、数式１で示される重回帰写像モデルに基づき、ＨＭＭのパラメータの一つであるガウス分布の平均ベクトルを線形変換することで教師なし話者適応パタン２００８を計算する。数式１においてμ（ｑ）、μａ（ｑ）は更新前後のガウス分布番号ｑの平均ベクトルであり、次元数はｄであり音声特徴量ベクトルの次元数と同じである。Ａはｄ×ｄの変換行列であり、ｖはｄ次元の定数項ベクトルである。変換行列Ａとｖは数式２によってＡのｐ行目、ｖのｐ次元目を算出する。数式２において、Ψは更新を行うガウス分布番号の集合、ｒ（ｉ，ｔ）は時刻ｔにガウス分布ｉに特徴ベクトルｏ（ｔ）が存在する期待値、μ（ｉ，ｒ）はガウス分布ｉの平均ベクトルのｒ次元目の要素、σ２（ｉ，ｐ）はガウス分布ｉの共分散行列のｐ行ｐ列目の要素、ｏ（ｔ，ｐ）は特徴ベクトルｏ（ｔ）のｐ次元目の要素、Ｔは適応データの総フレーム数である。

【数１】

$$\mu_a(q) = A\mu(q) + v$$

【数２】

$$\begin{pmatrix} a(p,1) \\ a(p,2) \\ \vdots \\ a(p,d) \\ v(p) \end{pmatrix} = \begin{pmatrix} f(1,1) & f(1,2) & \cdots & f(1,d+1) \\ f(2,1) & f(2,2) & \cdots & f(2,d+1) \\ \vdots & & \cdots & \vdots \\ f(d,1) & f(d,2) & \cdots & f(d,d+1) \\ f(d+1,1) & f(d+1,2) & \cdots & f(d+1,d+1) \end{pmatrix}^{-1} \begin{pmatrix} z(1) \\ z(2) \\ \vdots \\ z(d) \\ z(d+1) \end{pmatrix}$$

$$f(r,s) = \sum_{i\in\Psi}\sum_{t=1}^{T} \gamma(i,t)\frac{\mu(i,r)\mu(i,s)}{\sigma^2(p)}, (1 \le r,s \le d)$$

$$f(r,d+1) = f(d+1,r) = \sum_{i\in\Psi}\sum_{t=1}^{T} \gamma(i,t)\frac{\mu(i,r)}{\sigma^2(i,p)}, (1 \le r \le d)$$

$$f(d+1,d+1) = \sum_{i\in\Psi}\sum_{t=1}^{T} \gamma(i,t)\frac{1}{\sigma^2(i,p)}$$

$$z(r) = \sum_{i\in\Psi}\sum_{t=1}^{T} \gamma(i,t)\frac{o(t,p)\cdot\mu(i,r)}{\sigma^2(i,p)}, (1 \le r \le d)$$

$$z(d+1) = \sum_{i\in\Psi}\sum_{t=1}^{T} \gamma(i,t)\frac{o(t,p)}{\sigma^2(i,p)}$$

【００１３】

教師なし話者適応パタン２００８は、教師なし話者適応手段２００７からの出力であり、この標準パタンを用いて音声認識装置などで音声認識が行われる。

【００１４】

【発明が解決しようとする課題】

しかし、従来の教師なし話者適応化装置では、照合を行って得られた話者適応用認識結果

を発声内容として標準パタンのパラメータの更新を行っていたため、話者適応学習用認識結果が誤った場合には、パラメータの誤った推定が行われ認識率が低下する、という問題点があった。

【００１５】

そこで、本発明は、以上の問題点を解決し、教師なし話者適応方式において話者適応学習用認識結果が誤った場合においても、標準パタンのパラメータ誤推定を防ぎ、認識率を向上させることのできる話者適応化装置、およびその話者適応化装置により更新された教師なし話者適応パタン使用して音声認識を行う音声認識装置を提供することを目的とする。

【００１６】

【課題を解決するための手段】

この発明に係る話者適応化装置では、話者の入力音声から抽出した音声特徴量と、多数の話者の音声データによりパラメータ学習を行って得た標準パタンと、を照合して認識結果を出力するとともに、前記標準パタンを、前記入力音声を発した話者に適応した話者適応パタンに更新するか否かを、前記認識結果の信頼度に応じて決定する話者適応化装置において、

前記音声特徴量と前記標準パタンから推定適応パタンを最尤推定により算出する標準パタンパラメータ最尤推定手段と、

この標準パタンパラメータ最尤推定手段により算出された推定適応パタンを構成するパラメータの値と前記標準パタンを構成するパラメータの値とを、前記信頼度に応じて線形補間することにより前記話者適応パタンを算出するパラメータ線形補間手段と、

を備えるものである。

【００１７】

また、次の発明に係る話者適応化装置では、話者の入力音声から抽出した音声特徴量と、多数の話者の音声データによりパラメータ学習を行って得た標準パタンと、を照合して認識結果を出力するとともに、前記標準パタンを、前記入力音声を発した話者に適応した話者適応パタンに更新するか否かを、前記認識結果の信頼度に応じて決定する話者適応化装置において、

前記信頼度に基づいて、前記入力音声から得られる話者適応用データのパラメータ学習への重み付けを計算し、重み付けされた話者適応用データを用いて、前記標準パタンからを構成するパラメータを、前記話者適応パタンを構成するパラメータに更新する適応学習手段を備えたことを特徴とする。

【００１８】

また、次の発明に係る話者適応化装置では、話者の入力音声から抽出した音声特徴量と、多数の話者の音声データによりパラメータ学習を行って得た標準パタンと、を照合して認識結果を出力するとともに、前記標準パタンを、前記入力音声を発した話者に適応した話者適応パタンに更新するか否かを、前記認識結果の信頼度に応じて決定する話者適応化装置において、

過去に出力した前記認識結果の信頼度の値に基づいて、異なる話者適用学習アルゴリズムを選択する話者適用方式選択手段を備えたことを特徴とする。

【００３９】

【発明の実施の形態】

実施の形態１.

図１は、請求項１記載の発明による話者適応化装置の１構成である実施の形態１を示すブロック図である。図１において従来技術の説明図である図２１と同一の機能ブロックは同一の符号を付し説明を省略する。従来技術と異る本発明の特徴的な部分は、認識結果信頼度演算手段１０１を備えたことと、教師なし話者適応手段２００７の代りに認識結果信頼度付き教師なし話者適応手段１０２を備えたことである。

【００４０】

次に図１を参照しながら動作について説明する。認識結果信頼度演算手段１０１は、照合手段２００３からの出力である話者適応学習用音声認識結果２００６と音声特徴量抽出手

段２００２からの出力である音声特徴量、及び標準パタン２００４を入力し、話者適応学習用認識結果２００６に対する信頼度を演算する。認識結果の信頼度は、例えば「種々の統計量を用いた単語リジェクト方式の検討」花沢、阿部、日本音響学会平成１０年春期研究発表会講演論文集、ｐｐ．１４１－１４２、１９９８年３月（以降、文献４という）に示されている統計量を用いる。

【００４１】

文献４では、認識結果の信頼度を得るために（１）音響尤度差、（２）音素継続時間長、（３）音素混同行列の３種類の統計量を用いている。

（１）の音響尤度差は、入力音声の話者適応学習用音声認識結果２００６であるＲｗ■のフレーム尤度と、全音素接続の音素タイプライタによる音声認識装置の認識結果Ｒｗ■の区間に対しての尤度の差を数式３により計算して信頼度とするものである。数式３においてｌｔはフレームｔにおける認識結果Ｒｗ■の対数フレーム尤度、Ｌｔは、音素タイプライタによる対数フレーム尤度である。また、ＮはＲｗ■の音素数、ｂｉとｅｉは、ｉ番目の音素の始端と終端フレームである。Ｓａは値が小さいほど信頼性が高い統計量であるので通常はマイナスを乗じた値として信頼度とする。

【数３】

$$S_a = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{e_i - b_i + 1}\sum_{t=b_i}^{e_i}(l_t - L_t)$$

【００４２】

（２）の音素継続時間長は、入力音声に対する話者適応学習用音声認識結果Ｒｗ■の各音素の隣接音素間の継続時間長の整合性に基づく信頼性の統計量であり、数式４によって信頼度を計算する。数式４においてｄｉはＲｗ■を構成するｉ番目の音素を中心として前後１音素づつの継続時間長を並べた３次元のベクトルであり、Ｄｉは他の多数話者の音声データを用いて事前に求めた前記３音素の継続時間長の平均値を並べた３次元ベクトルである。数式４によって演算するＳｄは、認識結果Ｒｗ■中の隣接する３音素間の継続時間長の比が、学習データによって求めた平均時間長の比に近いほど大きな値をとる。したがって、Ｓｄは値が大きいほど認識結果の信頼度が高い統計量である。

【数４】

$$S_d = \frac{1}{N}\sum_{i=1}^{N}\log\frac{d_i \cdot D_i}{|d_i||D_i|}$$

【００４３】

（３）の音素混同行列は、音素タイプライタによる音素認識を並行して行い、話者適応学習用音声認識結果Ｒｗ■を構成する音素系列と音素タイプライタによる認識結果である音素系列とを時間軸上で対応づけ、事前に求めた音素混同行列を用いて数式５によって信頼度を計算する。数式５において、ｈｉはＲ■ｗを構成するｉ番目の音素モデル、ｐｉｋは音素タイプライタによる音素系列中でｈｉと区間が重なる音素、Ｋｉはｈｉと区間が重なる音素数、ｍ（ｈ，ｐ）は事前に求めた音素ｈ音素ｐの混同率、ｗｉｋはｈｉとｐｉｋとの区間重なり率であり、数式６によって計算する。数式５のＳｃは、値が大きいほど認識結果の信頼度が高い統計量である。最終的なＲｗ■の認識結果信頼度は上記の３種類の統計量を用い、数式７によって計算する。数式７においてｗ２、ｗ３は重み係数であり実験的に設定する。

【数５】

$$S_c = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_i} w_{ik} * \log\{m(h_i, p_{i,k})\}$$

【数６】

$$w_{ik} = \frac{\text{音素モデル}h_i\text{と音素}p_{ik}\text{の重なりフレーム数}}{\text{音素モデル}h_i\text{のフレーム数}}$$

【数７】

$$S_{all} = -S_a + w_2 * S_d + w_3 * S_c$$

【００４４】
認識結果信頼度付き教師なし話者適応手段１０２は、認識結果信頼度演算手段１０１からの出力である認識結果信頼度と、照合手段２００３からの出力である話者適応学習用音声認識結果２００６と、音声特徴量抽出手段２００２からの出力である音声特徴量と、標準パタン２００４を入力して標準パタンのパラメータの更新を行い、教師なし話者適応パタン２００８を出力する。
従って、この実施の形態１の話者適応化装置によれば、上記のように認識結果に対して信頼度を付加して教師なし話者適応を行うので認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐので、認識率を向上させることができる。
【００４５】
実施の形態２．
図２は、請求項２記載の発明による話者適応化装置の１構成例である実施の形態２を示すブロック図である。図２において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明の特徴的な部分は、先行する発声によって更新した教師なし話者適応パタン２００８を標準パタン２００４へ代入し、引き続く発声に対して教師なし話者適応を行うことを特徴としたことである。
【００４６】
次に図２を参照しながら動作について説明する。認識結果信頼度付き教師なし話者適応手段１０２は、使用者の最初の発声Ｏ（１）　＝　［ｏ（ｔ１），ｏ（ｔ１＋１），　．．．，　ｏ（Ｔ１）］を用いて標準パタン２００４のパラメータを更新して教師なし話者適応パタン２００８を出力する。ここで、この最初の発声によって得られた教師なし話者適応パタンをΛ（１）とする。次にΛ（１）を標準パタン２００４とし、使用者の２番目の発声Ｏ（２）　＝　［ｏ（ｔ２），　ｏ（ｔ２＋１），　．．．，　ｏ（Ｔ２）］　を用いて教師なし話者適応処理によって更に標準パタン２００４を更新して、教師なし話者適応パタン２００８を計算する。このようにｊ番目の発声を用いた教師なし話者適応の更新前の標準パタンとして（ｊ－１）番目の発声までに逐次的に更新したΛ（ｊ－１）を用いる。
従って、この実施の形態２の話者適応化装置によれば、上記のように認識結果に対して信頼度を付加して逐次的に教師なし話者適応を行うので認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐので、認識率を向上させることができる。
【００４７】
実施の形態３．
図３は、請求項３記載の発明による話者適応化装置の認識結果信頼度演算手段の動作説明図であり、実施の形態３の特徴を示す図である。本実施の形態３の特徴的な部分は、認識

結果信頼度演算手段１０１から出力である認識結果信頼度は、ポーズで区切られた１発声毎に１つ計算することである。認識結果信頼度演算手段１０１は、図３に示すようにｋ番目の発声の始端と終端をｔｕｓ（ｋ）、ｔｕｅ（ｋ）とした場合に、ｔｕｅ（ｋ）とｔｕｅ（ｋ）との間のフレームに関して１つの認識結果信頼度Ｓｕ（ｋ）を計算して、ｔｕｅ（ｋ）とｔｕｅ（ｋ）との間の各フレームの認識結果信頼度をＳｕ（ｋ）とする。
従って、この実施の形態３の話者適応化装置によれば、上記のように１発声毎に認識結果に対して信頼度を付加して教師なし話者適応を行うので、認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐので、認識率を向上させることができる。

【００４８】

実施の形態４．

図４は、請求項４記載の発明による話者適応化装置の認識結果信頼度演算手段の動作説明図であり、実施の形態４の特徴を示す図である。本実施の形態４の特徴的な部分は、認識結果信頼度演算手段１０１からの出力である認識結果信頼度は、認識ユニットに１つ計算することである。認識ユニットとは標準パタンの基本単位であり、認識ユニットを連結することで認識対象の単語、文章を認識する標準パタンを構成する。認識結果信頼度演算手段１０１は入力音声の話者適応学習用音声認識結果２００６に基づき、認識ユニットラベル系列にしたがって標準パタンを連結し、この標準パタンによって音声特徴量の時系列を認識ユニットに分割する。分割されたｕ番目の認識ユニットの始端と終端をｔｒｓ（ｕ）、ｔｒｅ（ｕ）とした場合に、ｔｒｓ（ｕ）とｔｒｅ（ｕ）の間のフレームに関して１つの認識結果信頼度Ｓｒ（ｕ）を図４のように計算し、区間内のフレームの認識結果信頼度をＳｒ（ｕ）とする。図４は認識結果が５個の認識ユニットによって構成されている例である。
従って、この実施の形態４の話者適応化装置によれば、上記のように１認識ユニット毎に認識結果に対して信頼度を付加して教師なし話者適応を行うので認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐので、認識率を向上させることができる。

【００４９】

実施の形態５．

図５は，請求項５記載の発明による話者適応化装置の認識結果信頼度演算手段の動作説明図であり、実施の形態５の特徴を示す図である。本実施の形態５の特徴的な部分は、認識結果信頼度演算手段１０１からの出力である認識結果信頼度は、音素や音節などの音声単位に１つ計算することである。以下では音声単位が音素である場合で説明する。認識結果信頼度演算手段１０１は入力音声の話者適応学習用音声認識結果２００６の音素系列にしたがって、音声特徴量の時系列を音素単位に分割する。分割されたｐ番目の音素の始端と終端をｔｐｓ（ｐ）、ｔｐｓ（ｐ）とした場合に、ｔｐｓ（ｐ）とｔｐｅ（ｐ）との間のフレームに関しては認識結果信頼度Ｓｐ（ｐ）を図５のように計算して、ｔｐｓ（ｐ）とｔｐｅ（ｐ）との区間内の各フレームの認識結果信頼度をＳｐ（ｐ）とする。図５は入力音声の話者適応学習用認識結果が／ｏｎｓｅｉ／の５音素によって構成されている例である。
従って、この実施の形態５の話者適応化装置によれば、上記のように１音素毎に認識結果に対して信頼度を付加して教師なし話者適応を行うので認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐので、認識率を向上させることができる。

【００５０】

実施の形態６．

図６は、請求項６記載の発明による話者適応化装置の認識結果信頼度演算手段の動作説明図であり、実施の形態６の特徴を示す図である。本実施の形態６の特徴的な部分は、認識結果信頼度演算手段１０１からの出力である認識結果信頼度は、一定時間間隔のフレーム単位に計算することである。以下では図６を参照しながら動作説明を行う。認識結果信頼度演算手段１０１は、入力音声を５ミリ秒～２０ミリ秒程度の一定時間間隔のフレーム単位に認識結果信頼度を出力する。図６は、フレームｔ～ｔ＋５毎に認識結果信頼度［Ｓｆ

（ｔ），　Ｓｆ（ｔ＋１），　．．．，　，　Ｓｆ（ｔ＋５）］を出力を示したものである。
従って、この実施の形態６の話者適応化装置によれば、このように一定時間間隔のフレーム単位で認識結果信頼度を計算するので、認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐことができ、認識率を向上させることができる。
【００５１】
実施の形態７．
図７は、請求項７記載の発明による話者適応化装置の１構成例である実施の形態７を示すブロック図である。図７において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明の特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２は、音声データセグメンテーション手段７０１と認識結果信頼度付き標準パタンパラメータ更新手段７０２で構成することを特徴としたことである。
【００５２】
次に図７を参照しながら動作について説明する。音声データセグメンテーション手段７０１は、話者適応学習用音声認識結果２００６に基づいて、標準パタン２００４から対応する認識ユニット標準パタンを連結し、音声特徴量の時系列を認識ユニット毎にセグメンテーションする。セグメンテーションは、例えば標準パタンがＨＭＭである場合は文献１に記載されているビタービアルゴリズムによって行う。ビタービアルゴリズムは、音声特徴量の時系列［ｏ（１），　ｏ（２），　．．．，　ｏ（ｔ）］に対する１本の最適状態系列［ｑ１，　ｑ２，　．．．，　ｑｔ］を見つけるアルゴリズムである。例えば単語標準パタンが３つの認識ユニットからなり、１認識ユニット当り１状態のＨＭＭであるとし、状態が（ｓ１，ｓ２，ｓ３）で構成されるとする。そしてビタービアルゴリズムによって得られた最適状態系列［ｓ１，　ｓ１，　ｓ２，　ｓ２，　ｓ２，　ｓ３，　ｓ３，　ｓ３］であったならば、フレーム１～２がユニット１、フレーム３～５がユニット２、フレーム６～８がユニット３にセグメンテーションされる。
【００５３】
認識結果信頼度付き標準パタンパラメータ更新手段７０２は、認識ユニットの標準パタンパラメータを、セグメンテーションによって分割された音声特徴量と認識結果信頼度を用いて更新する。
従って、この実施の形態７の話者適応化装置によれば、上記のように音声データセグメンテーションを行って識結果信頼度付き標準パタンパラメータの学習を行うので、話者適応学習用認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐことができ、認識率を向上させることができる。
【００５４】
実施の形態８．
図８は、請求項８記載の発明による話者適応化装置の１構成例である実施の形態８を示すブロック図である。図８において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２を、標準パタンパラメータ最尤推定手段８０１と、認識結果信頼度に基づくパラメータ線形補間手段８０２とで構成することである。
【００５５】
次に図８を参照しながら動作について説明する。標準パタンパラメータ最尤推定手段８０１は、音声特徴量抽出手段２００２の出力である音声特徴量と、話者適応学習用音声認識結果２００６に基づいて標準パタン２００４の認識ユニット標準パタンを連結した標準パタンを用いて、標準パタンのパラメータの最尤推定を行い、推定後の標準パタンΛｍを得る。最尤推定は、例えば文献１に記載されているＢａｕｍ－Ｗｅｌｃｈ法によってパラメータ推定を行う。
【００５６】
認識結果信頼度に基づくパラメータ線形補間手段８０２は、標準パタンパラメータ最尤推定手段８０１からの出力である最尤推定後の標準パタンΛｍ、及び推定前の標準パタンΛ

を入力し、認識結果信頼度演算手段１０１からの出力である認識結果信頼度によってΛｍとΛのパラメータの線形補間を行い、得られた値を教師なし話者適応パタン２００８のパラメータとする。例えば標準パタンがＨＭＭであり、ガウス分布の平均ベクトルμ（ｑ）（ｑはガウス分布の番号）　を更新する場合には、数式８によって教師なし話者適応パタン２００８の平均ベクトルμａ（ｑ）を計算する。数式８においてμ（ｑ）、μｍ（ｑ）は最尤推定前後の平均ベクトルの値である。またｗｍ（ｑ）は、値が０から１．０の重み係数であり、μ（ｑ）の更新に用いた適応データの認識結果信頼度によって決定する。

【数８】

$$\mu_a(q) = w_m(q)\mu_m(q) + \{1 - w_m(q)\}\mu(q)$$

従って、この実施の形態８の話者適応化装置によれば、上記のように標準パタンパラメータ最尤推定後に認識結果信頼度に基づいてパラメータの線形補間を行うので、話者適応学習用認識結果が誤っている場合でも、標準パタンのパラメータの誤った更新を防ぐことができ、認識率を向上させることができる。

【００５７】

実施の形態９．

実施の形態９は、実施の形態８の話者適応化装置における標準パタンのパラメータの線形補間において、パラメータの最尤推定に使用した適応データの認識結果信頼度の合計値が大きければ最尤推定値の重みを大きくすることを特徴とした請求項９記載の発明による話者適応化装置である。数式９は数式８の重み係数ｗｍ（ｑ）の値を計算する請求項９記載の発明の１例である。数式９においてＳｆ（ｔ）はフレームｔにおける認識結果信頼度、Ωはパラメータμの更新に用いる適応データのフレームの時刻の集合、τは値が０以上の制御定数である。

【数９】

$$w_m(q) = \frac{\sum_{t\in\Omega} S_f(t)}{\tau + \sum_{t\in\Omega} S_f(t)}$$

従って、この実施の形態９の話者適応化装置によれば、上記のように構成することで認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００５８】

実施の形態１０．

図９は、請求項１０記載の発明による話者適応化装置の１構成例である実施の形態１０を示すブロック図である。図９において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２は、認識結果信頼度重み付き学習データによる適応学習手段９０１で構成することである。

【００５９】

次に図９を参照しながら動作について説明する。認識結果信頼度重み付き学習データによる適応学習手段９０１は、話者適応学習用音声認識結果２００６と標準パタン２００４と認識結果信頼度演算手段１０１の出力である認識結果信頼度と音声特徴量抽出手段２００２の出力である音声特長量の時系列とを入力し、認識結果信頼度によって適応データへ重み付けしたパラメータ更新を行う。例えば、標準パタン２００４がＨＭＭである話者適応化装置では、数式１０によってガウス分布の平均ベクトル、数式１１によってガウス分布の共分散行列の更新を行う。数式１０のｏｈ（ｔ）は認識結果信頼度によって重み付けさ

れた音声特徴量であり、例えば数式１２よって計算する。
数式１２において、μ（ｑ）は更新前のガウス分布の平均ベクトル、ｏ（ｔ）は時刻ｔの音声特徴量であり、τは値が０以上の制御定数、Ｓｆ（ｔ）はフレームｔの認識結果信頼度であるので、Ｓｆ（ｔ）　が小さい場合はｏｈ（ｔ）は更新前の平均ベクトルに近い値となり、ｏ（ｔ）のパラメータ更新への寄与度が小さく、またＳｆ（ｔ）が大きい場合は、ｏｈ（ｔ）はｏ（ｔ）に近い値となりパラメータ更新への寄与度が大きくなる。数式１０においてγ（ｑ，ｔ）は、時刻ｔにガウス分布ｑに音声特徴量　ｏ（ｔ）　が存在する期待値であるが、重み付けされた音声特徴量ｏｈ（ｔ）が存在する期待値として計算してもよい。また、ここで得られたμａ（ｑ）を数式８のμｍ（ｑ）として更新前の標準パタンパラメータとの線形補間を行うことも可能である。

【数１０】

$$\mu_a(q) = \frac{\sum_{t\in\Omega}\gamma(q,t)\cdot oh(t)}{\sum_{t\in\Omega}\gamma(q,t)}$$

【数１１】

$$C_a(q) = \frac{\sum_{t\in\Omega}\gamma(q,t)\cdot(oh(t)-\mu(q))(oh(t)-\mu(q))'}{\sum_{t\in\Omega}\gamma(q,t)}$$

【数１２】

$$oh(t) = \frac{S_f(t)}{\tau + S_f(t)}o(t) + \frac{\tau}{\tau + S_f(t)}\mu(q)$$

従って、この実施の形態１０の話者適応化装置によれば、このように構成することで認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００６０】
実施の形態１１．
また、実施の形態１１は、実施の形態１０の話者適応化装置において、認識結果信頼度をフレーム毎に付与し、その値が０～１であり、信頼度が高い場合には１に近い値を出力することを特長とした請求項１１記載の話者適応化装置である。
従って、この実施の形態１１の話者適応化装置によれば、上記のように構成することで認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００６１】
実施の形態１２．
図１０は、請求項１２記載の発明による話者適応化装置の１構成例である実施の形態１２を示すブロック図である。図１０において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、照合手段２００３は複数認識結果候補出力照合手段１００１で構成し、認識結果信頼度演算手段１０１は複数認識結果候補信頼度演算手段１００２で構成し、認識結果信頼度付き教師なし話者適応手段１０２は複数認識結果候補信頼度付き教師なし話者適応手段１００３で構成することを特徴としたことである。

【００６２】

次に図１０を参照しながら動作について説明する。複数認識結果候補出力照合手段１００１は、認識辞書２００５によって定められた認識対象単語にしたがい標準パタン２００４連結して、音声特徴量抽出手段２００２の出力である音声特徴量に対して照合を行ない、予め定めた候補数の認識結果［Ｒｗ■（１），　Ｒｗ■（２），　．．．，Ｒｗ■（Ｎ）］（Ｒｗ■（ｎ）は、入力音声に対してｎ番目にスコアが高い話者適応学習用音声認識結果、Ｎは予め定めた候補数）　を照合スコアが高い認識結果候補から順に出力する。

【００６３】

複数認識結果候補信頼度演算手段１００２は、複数認識結果候補出力照合手段１００１の出力である複数認識結果候補［Ｒｗ■（１），　Ｒｗ■（２），　．．．，　Ｒｗ■（Ｎ）］と音声特徴量と標準パタン２００４とを入力して複数の認識結果候補の各々に対して認識結果信頼度［Ｓｍ（１），　Ｓｍ（２），　．．．，　Ｓｍ（Ｎ）］を計算する。ここで、Ｓｍ（ｎ）は入力音声に対するｎ番目の認識結果候補に対する認識結果信頼度の時系列である。認識結果信頼度がフレーム毎のＳｆ（ｎ，ｔ）であるならば、Ｓｍ（ｎ）＝［Ｓｆ（ｎ，１），　Ｓｆ（ｎ，２）　，．．．，Ｓｆ（ｎ，Ｔｎ）］である。複数認識結果候補信頼度付き教師なし話者適応手段１００３は、複数認識結果候補出力照合手段１００１の出力である複数認識結果候補と複数認識結果候補信頼度演算手段１００２からの出力である認識結果信頼度と標準パタン２００４を入力して標準パタンのパラメータ更新を行い、教師なし話者適応パタン２００８を出力する。

【００６４】

複数認識結果候補信頼度付き教師なし話者適応手段１００３は、例えば複数認識結果各々を用いて独立にＮ個の教師なし話者適応パタンを作成して、Ｎ個の標準パタンのパラメータを合成することで最終的な教師なし話者適応パタン２００８を得る方法がある。例えば標準パタンがＨＭＭであり更新するパラメータをガウス分布の平均ベクトル、共分散行列とした場合、数式１３によってガウス分布ｑの平均ベクトル、数式１４によって共分散行列を計算する。数式１３においてμｉ（ｎ，ｑ）は、第ｎ番目の認識結果候補を用いて更新したガウス分布ｑの平均ベクトルであり、数式１４においてＣｉ（ｎ，ｑ）はｎ番目の認識結果候補を用いて更新したガウス分布ｑの共分散行列である。数式１３、数式１４においてβ（ｎ）は第ｎ番目の認識結果候補に対する重み付けであり数式１５によって計算する。数式１５においてＳｉ（ｎ）は第ｎ番目の認識結果候補の認識結果信頼度であり、例えばフレーム毎の認識結果信頼度の合計である。

【数１３】

$$\mu_a(q) = \sum_{n=1}^{N} \beta(n)\mu_i(n,q)$$

【数１４】

$$C_a(q) = \sum_{n=1}^{N} \beta(n)\cdot C_i(n,q) + \sum_{n=1}^{N} \beta(n)\cdot \mu_i(n,q)\mu_i(n,q)' - \mu_a(q)\mu_a(q)'$$

【数１５】

$$\beta(n) = \frac{S_i(n)}{\sum_{m=1}^{N} S_i(m)}$$

従って、この実施の形態１２の話者適応化装置によれば、このように複数認識結果候補を出力し複数認識結果候補に対して認識結果信頼度を計算して、認識結果信頼度付き教師なし話者適応を行うので認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００６５】

実施の形態１３．

図１１は、請求項１３記載の発明による話者適応化装置の１構成例である実施の形態１３を示すブロック図である。図１１において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２の前段に、認識結果信頼度比較手段１１０１が付加されていることである。

【００６６】

次に図１１を参照しながら動作について説明する。認識結果信頼度比較手段１１０１は、認識結果信頼度演算手段１０１からの出力である認識結果信頼度を入力し、認識結果信頼度が予め定めた閾値より大きければ、認識結果信頼度付き教師なし話者適応手段１０２で処理を行う。一方、認識結果信頼度が予め定めた閾値より小さければ、標準パタンのパラメータの更新は行わず、標準パタン２００４の値を教師なし話者適応パタン２００８とする。

【００６７】

例えば、１発声の認識結果信頼度の合計が閾値Ｔｈ以下であるならば、この発声を用いた標準パタンのパラメータ更新は行わない話者適応化装置である。また、標準パタンのパラメータ毎にセグメンテーションによって分割された適応データの認識結果信頼度の合計を計算し、パラメータ毎の認識結果信頼度と閾値を比較し、閾値以下であるならばパラメータの更新を行わず、閾値より大きいパラメータは更新を行う話者適応化装置である。

従って、この実施の形態１３の話者適応化装置によれば、このように認識結果信頼度が予め定めた閾値以下であるならばパラメータの更新を行わないように構成することで認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００６８】

実施の形態１４．

図１２は、請求項１４記載の発明による話者適応化装置の１構成例である実施の形態１４を示すブロック図である。図１２において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２の前段に、認識結果信頼度による話者適応方式選択手段１２０１と、Ｍ個の認識結果信頼度付き教師なし話者適応手段１２０２－１～１２０２－Ｍを備えたことである。

【００６９】

次に図１２を参照しながら動作について説明する。認識結果信頼度による話者適応方式選択手段１２０１は、認識結果信頼度演算手段１０１からの出力である認識結果信頼度を入力して予め定めた方式選択閾値［Ｔｈ（１），Ｔｈ（２），．．．，Ｔｈ（Ｋ）］によって教師なし話者適応方式の選択を行う。例えば認識結果信頼度の値がＳである場合は、Ｔｈ（ｋ）≦Ｓｕ＜Ｔｈ（ｋ＋１）では認識結果信頼度付き教師なし話者適応方式１２０２－ｋを選択する。ここでＳｕは１発声の認識結果信頼度の合計値である。

【００７０】

認識結果信頼度付き教師なし話者適応手段１２０２－１～１２０２－Ｍは、例えば「Ａ Ｓｔｕｄｙ ｏｎ Ｓｐｅａｋｅｒ Ａｄａｐｔａｔｉｏｎ ｏｆ ｔｈｅ Ｐａｒａｍｅｔｅｒｓ ｏｆ Ｃｏｎｔｉｎｕｏｕｓ ＤｅｎｓｉｔｙＨｉｄｄｅｎ Ｍａｒｋｏｖ Ｍｏｄｅｌｓ」Ｃ．Ｈ．Ｌｅｅ，Ｃ．Ｈ．Ｌｉｎ，Ｂ．Ｈ．Ｊｕａｎｇ，ＩＥＥＥ ＴＲＡＮＳＡＣＴＩＯＮ ＯＮＳＩＧＮＡＬ ＰＥＯＣＥＳＳＩＮＧ，Ｖｏｌ．３９，Ｎｏ．４，１９９１年（以下、文献５という）で提案されている最大事後確率推定法が１２０２－１、「連続混合分布ＨＭＭを用いた移動ベクトル場平滑化話者適応化方式」大倉、杉山、嵯峨山、電子情報通信学会技術報告、ＳＰ９２－１６、１９９２年（以下、文献６という）で提案されている移動ベクトル場平滑化話者適応方式が１２０２－２、重回帰写像モデルに基づく話者適応方式（文献３）が１２０２－３であるとして構成できる。

従って、この実施の形態１４の話者適応化装置によれば、このように構成することで話者適応学習用音声認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００７１】

実施の形態１５．

図１３は、請求項１５記載の発明による話者適応化装置の１構成例である実施の形態１５を示すブロック図である。図１３において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段を、標準パタンパラメータクラスタリング手段１３０１と、認識結果信頼度付きパラメータグループ教師なし話者適応手段１３０２とで構成することである。

【００７２】

次に図１３を参照しながら動作について説明する。標準パタンパラメータグループ化手段１３０１は、標準パタン２００４に格納されている標準パタンパラメータをクラスタリングによってグループ化する。標準パタンがＨＭＭの場合はガウス分布［ｇ（１），ｇ（２），．．，ｇ（Ｍｇ）］（Ｍｇは全ガウス分布数）を例えば数式１６のバタチャリヤの距離によってガウス分布ｇ（ｉ）とｇ（ｊ）間の距離ｄｖ（ｇ（ｉ），ｇ（ｊ））を定義してクラスタリングを行い、グループＧ（ｘ）＝［ｇ（ｘ（１）），ｇ（ｘ（２）），．．．，ｇ（ｘ（ｎ））］（ｘ（．）は分布番号）を決定する。クラスタリング法は例えば文献１に記載されて「るＫ－平均法を用いて行う。

認識結果信頼度付きパラメータグループ教師なし話者適応手段１３０２は、標準パタンパラメータグループ化手段１３０１からの出力である標準パタンパラメータグループと認識結果信頼度演算手段１０１からの出力である認識結果信頼度を入力し、グループ毎に標準パタンパラメータの変動量の計算を行う。例えば標準パタンがＨＭＭである場合の平均ベクトルのｐ次元目の移動量は数式１７によって計算する。数式１７においてα（ｘ）は数式１８に示す信頼度によって決定される重み係数である。また、Ψｘはパラメータグループｘのガウス分布番号の集合、Ωｉはガウス分布番号ｉの適応データの時刻の集合、σ２（ｉ，ｐ）はガウス分布番号ｉの共分散行列のｐ行ｐ列目である。数式１８において、Ｓｆ（ｔ）はフレームｔの認識結果信頼度であり、τは値が０以上の制御定数である。

また、数式１９によってグループｘの平均ベクトルの共通な移動量ｖ（ｘ，ｐ）を求めることも可能である。数式１９においてｏｈ（ｔ）は数式１２に示した認識結果信頼度によって重み付けされた適応データであり、γ（ｉ，ｔ）は、時刻ｔにガウス分布ｉに音声特徴量ｏ（ｔ）が存在する期待値であるが、重み付けされた音声特徴量ｏｈ（ｔ）が存在する期待値として計算してもよい。

【数１６】

$$dv(g(i),g(j)) = \frac{1}{8}(\mu(i)-\mu(j))'\left(\frac{C(i)+C(j)}{2}\right)^{-1}(\mu(i)-\mu(j)) + \frac{1}{2}\ln\frac{|(C(i)+C(j))/2|}{|C(i)|^{1/2}|C(j)|^{1/2}}$$

【数１７】

$$v(x,p) = \frac{\sum_{i\in\Psi_x}\sum_{t\in\Omega_i}\frac{\gamma(i,t)(o(t,p)-\mu(i,p))}{\sigma^2(i,p)}}{\sum_{i\in\Psi_x}\sum_{t\in\Omega_i}\frac{\gamma(i,t)}{\sigma^2(i,p)}} \times \alpha(x)$$

【数１８】

$$\alpha(x) = \frac{\sum_{i\in\Psi_x}\sum_{t\in\Omega_i}S_f(t)}{\tau+\sum_{i\in\Psi_x}\sum_{t\in\Omega_i}S_f(t)}$$

【数１９】

$$v(x,p) = \frac{\sum_{i\in\Psi_x}\sum_{t\in\Omega_i}\frac{\gamma(i,t)(oh(t,p)-\mu(i,p))}{\sigma^2(i,p)}}{\sum_{i\in\Psi_x}\sum_{t\in\Omega_i}\frac{\gamma(i,t)}{\sigma^2(i,p)}}$$

従って、この実施の形態１５の話者適応化装置によれば、このように構成することで認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００７３】

実施の形態１６．

図１４は、請求項１６記載の発明による話者適応化装置の１構成例である実施の形態１６を示すブロック図である。図１４において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２を、標準パタンパラメータ木構造クラスタリング手段１４０１と、木構造化パラメータに基づく標準パタンパラメータグループ化手段１４０２と、認識結果信頼度付きパラメータグループ教師なし話者適応手段１３０２とで構成することである。

【００７４】

次に図１４を参照しながら動作について説明する。標準パタンパラメータ木構造クラスタリング手段１４０１は、標準パタンパラメータを例えば数式１６に示すバタチャリヤの距離によって木構造にクラスタリングする。木構造化は、まず木構造の１階層目のグループ化として全パラメータをＮ個のパラメータグループ［Ｇ（１，１，１），　Ｇ（１，１，２），．．．．，　Ｇ（１，１，Ｎ）］　（Ｇ（ｉ，ｊ，ｋ））：ｉは階層、ｊは親グループ番号、ｋはグループ番号）にクラスタリングする。

次に２階層目のクラスタリングとして、Ｇ（１，ｍ１，ｎ１）を［Ｇ（２，ｎ１，１），　Ｇ（２，ｎ１，１），．．．，　Ｇ（２，ｎ１，Ｎｎ１）］のグループにクラスタリングする。

さらに3階層目としてG（2，m2，n2）を［G（3，n2，1），G（3，n2，1），．．．，G（3，n2，Nn2）］にクラスタリングにグループ化する。このように予め定めた階層までクラスタリングを行う。木構造パラメータに基づく標準パタンパラメータグループ化手段１４０２は、認識結果信頼度演算手段１０１の出力の認識結果信頼度によって標準パラメータ木構造クラスタリングの出力である木構造化されたパラメータに基づいてパラメータをグループ化する。

【００７５】

図１５は、認識結果信頼度による木構造化パラメータのグループ化の説明図である。ノード以下に属するパラメータの適応データの認識結果信頼度の合計をノードの情報として計算する。子ノードの認識結果信頼度が予め定めた閾値ｔｈより小さく、親ノードの認識結果信頼度がｔｈ以上である場合に、親ノード以下のパラメータグループを子ノード以下に属するパラメータの推定に用いる。図１５において括弧内の数字がノード以下のパラメータの適応データの認識結果信頼度である。例えばｔｈを４０とすれば、Ｎｏｄｅ（３，１）では信頼度２０、その親ノードのＮｏｄｅ（２，１）では１００であるのでパラメータの更新には、Ｎｏｄｅ（２，１）以下のパラメータの適応データと認識結果信頼度および標準パタンパラメータを用いて、パラメータに共通の変動量を求めてＮｏｄｅ（３，２）以下のパラメータ更新を行う。パラメータグループのパラメータ変動量を演算する認識結果信頼度付きパラメータグループ教師なし話者適応手段１３０２は、実施の形態１５で記述したようにパラメータグループにおいて変動量を求め更新を行う。

従って、この実施の形態１７の話者適応化装置によれば、上記のように構成することで認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００７６】

実施の形態１７．

また、実施の形態１７の話者適応化装置は、標準パタンとして、連続混合分布型隠れマルコフモデルを用いることを特徴とした請求項１７記載の発明による話者適応化装置である。連続混合分布型隠れマルコフモデルについては文献１に詳細が記載されているので説明は省略する。

【００７７】

実施の形態１８．

また、実施の形態１８の話者適応化装置は、連続混合分布型隠れマルコフモデルのシンボル出力確率密度関数を構成する要素分布関数はガウス分布であることを特徴とする請求項１８記載の発明による話者適応化装置である。ガウス分布関数は数式２０で与えられる。数式２０において、μ（ｉ）、Ｃ（ｉ）はガウス分布ｉの平均ベクトルと共分散行列である。また、ｄは平均ベクトルの次元数であり、ｏは特徴量ベクトルである。

【数２０】

$$N(o|\mu(i),C(i)) = \frac{1}{(2\pi)^{d/2}|C(i)|^{1/2}} \times \exp\left[-\frac{1}{2}(o-\mu(i))'C^{-1}(i)(o-\mu(i))\right]$$

【００７８】

実施の形態１９．

また、実施の形態１７の話者適応化装置は、適応するパラメータはガウス分布の平均ベクトルであることを特徴とする請求項１９記載の発明による話者適応化装置である。

【００７９】

実施の形態２０．

図１６は、請求項２０記載の発明による話者適応化装置の１構成例である実施の形態２０を示すブロック図である。図１６において、実施の形態１と実施の形態７と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２を、認識結果信頼度付き標準パタンパラメータ更新

手段７０２と標準パタンパラメータ補間手段１６０１とで構成することである。

【００８０】

次に図１６を参照して動作について説明する。認識結果信頼度付きパラメータ更新手段７０２は実施の形態８や実施の形態１０に記述したパラメータ更新によってガウス分布の平均値の更新を行う。パラメータ補間手段１６０１は、適応学習データが存在しなかったガウス分布の平均ベクトルを認識結果信頼度付きパラメータ更新手段７０２によって学習されたガウス分布の平均ベクトルの更新前後の差ベクトルを用いて数式２１によって補間する。

【数２１】

$$\mu_a(q) = \mu(q) + \frac{\sum_{p \in P} \alpha_{p,q} \cdot TV(p)}{\sum_{p \in P} \alpha_{p,q}}$$

$$TV(p) = \mu_a(p) - \mu(p)$$

$$\alpha_{p,q} = \exp\left(-\frac{d_{p,q}}{f}\right)$$

$$d_{p,q} = (\mu(q) - \mu(p))' C^{-1}(q)(\mu(q) - \mu(p))$$

【００８１】

図１７はガウス分布平均値の補間の概念図である。図１７においてμ（１）、μ（２）、μ（３）は適応データが存在するガウス分布の平均ベクトルであり、μａ（１）、μａ（２）、μａ（３）は教師なし話者適応によって更新した後の平均ベクトルである。また、μ（４）は適応データが存在しない平均ベクトルである。この適応データが存在しないμ（４）は、数式２１によって、近傍の平均ベクトルの更新前後の差ベクトルによって補間を行う。数式２１において、μ（ｑ）、μａ（ｑ）はｑ番目の更新前後の平均ベクトル、αｐ，ｑは重み係数、ＴＶ（ｐ）は更新前後の平均ベクトルの差ベクトル　（移動ベクトル）、Ｐは補間に用いる近傍の平均ベクトルの集合である。またｆは制御定数であり、ｄｐ，ｑはマハラノビス距離であり、Ｃ（ｑ）はガウス分布ｑの共分散行列であり、上付き－１は逆行列を表す。

従って、この実施の形態２０の話者適応化装置によれば、このように適応データが存在しないガウス分布の平均ベクトルは、適応データが存在するガウス分布の平均ベクトルの差ベクトルによって補間を行って適応するので認識結果が誤っている場合でも、標準パタンのパラメータの誤った学習を防ぐことができ、認識率を向上させることができる。

【００８２】

実施の形態２１．

図１８は、請求項２１記載の発明による話者適応化装置の１構成例である実施の形態２１を示すブロック図である。図１８において、実施の形態１と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２は、認識結果信頼度付き重回帰写像モデルに基づく話者適応手段１８０１であることである。

【００８３】

次に図１８を参照しながら動作について説明する。認識結果信頼度付き重回帰写像モデルに基づく話者適応手段１８０１は、認識結果信頼度演算手段１０１の出力である認識結果信頼度と話者適応学習用音声認識結果２００６と標準パタン２００４とを入力し、数式１の重回帰写像モデルに基づく線形変換によってガウス分布の平均ベクトルを更新する。数

式１のＡとｖは数式１２に示されている認識結果信頼度によって重み付けした適応データｏｈ（ｔ）を用いて、数式２２によってＡのｐ行目、ｖのｐ次元目の要素を求める。数式２２においてｏｈ（ｔ，ｐ）は認識結果信頼度によって重み付けした適応データｏｈ（ｔ）のｐ次元目の要素であり、その他の変数に関しては数式２と同一である。また、γ（ｉ，ｔ）は、時刻ｔにガウス分布ｉに音声特徴量ｏ（ｔ）が存在する期待値であるが、重み付けされた音声特徴量ｏｈ（ｔ）が存在する期待値として計算してもよい。

【数２２】

$$\begin{pmatrix} a(p,1) \\ a(p,2) \\ \vdots \\ a(p,d) \\ v(p) \end{pmatrix} = \begin{pmatrix} f(1,1) & f(1,2) & \cdots & f(1,d+1) \\ f(2,1) & f(2,2) & \cdots & f(2,d+1) \\ \vdots & & \cdots & \vdots \\ f(d,1) & f(d,2) & \cdots & f(d,d+1) \\ f(d+1,1) & f(d+1,2) & \cdots & f(d+1,d+1) \end{pmatrix} \begin{pmatrix} z(1) \\ z(2) \\ \vdots \\ z(d) \\ z(d+1) \end{pmatrix}$$

$$f(r,s) = \sum_{i \in \Psi} \sum_{t=1}^{T} \gamma(i,t) \frac{\mu(i,r)\mu(i,s)}{\sigma^2(p)}, (1 \leq r,s \leq d)$$

$$f(r,d+1) = f(d+1,r) = \sum_{i \in \Psi} \sum_{t=1}^{T} \gamma(i,t) \frac{\mu(i,r)}{\sigma^2(i,p)}, (1 \leq r \leq d)$$

$$f(d+1,d+1) = \sum_{i \in \Psi} \sum_{t=1}^{T} \gamma(i,t) \frac{1}{\sigma^2(i,p)}$$

$$z(r) = \sum_{i \in \Psi} \sum_{t=1}^{T} \gamma(i,t) \frac{oh(t,p) \cdot \mu(i,r)}{\sigma^2(i,p)}, (1 \leq r \leq d)$$

$$z(d+1) = \sum_{i \in \Psi} \sum_{t=1}^{T} \gamma(i,t) \frac{oh(t,p)}{\sigma^2(i,p)}$$

【００８４】

また、認識結果信頼度付き重回帰写像モデルに基づく話者適応手段１８０１は、従来の重回帰写像モデルによる話者適応と同様に数式１、数式２によって平均ベクトルを更新してμａ■（ｑ）を求め、このμａ■（ｑ）を数式８のμｍ（ｑ）として認識結果信頼度によって線形補間する構成としてもよい。

従って、この実施の形態２１の話者適応化装置によれば、このように認識結果信頼度付きの重回帰写像モデルに基づく教師なし話者適応を行うので、話者適応学習用認識結果が誤った場合のパラメータの誤った更新を防ぐことができ、認識率が向上する。

【００８５】

実施の形態２２．

図１９は、請求項２２記載の発明による話者適応化装置の１構成例である実施の形態２２を示すブロック図である。図１９において、実施の形態１、実施の形態１５、及び実施の形態１８と同一の機能ブロックは同一の番号を付し説明を省略する。本発明において特徴的な部分は、認識結果信頼度付き教師なし話者適応手段１０２を、ガウス分布グループ化手段１９０１と、認識結果信頼度付き重回帰写像モデルに基づく話者適応手段１８０１とで構成することである。

【００８６】

次に図１９を参照しながら動作について説明する。ガウス分布グループ化手段１９０１は、標準パタン２００４のガウス分布をクラスタリングによってグループ化し、グループ内

のガウス分布の適応データの認識結果信頼度に基づいてグループ毎に実施の形態２１で記述した認識結果信頼度付き重回帰写像モデルに基づく話者適応を行う。
従って、この実施の形態２２の話者適応化装置によれば、このように標準パタンをグループ化して認識結果信頼度付きの重回帰写像モデルに基づいて教師なし話者適応を行うので、話者適応学習用認識結果が誤った場合のパラメータの誤った更新を防ぐことができ、認識率が向上する。
【００８７】
実施の形態２３．
図２０は、請求項２３記載の発明による音声認識装置、すなわち上記実施の形態１～２２の教師なし話者適応化装置により更新された教師なし話者適応パタン２００８を使用した音声認識装置である実施の形態２３の構成を示すブロック図である。尚、図２０において、図１等に示す話者適応化装置と同じ構成には、同一の番号を付して説明を省略する。
【００８８】
認識辞書２００５によって設定した認識対象の単語［Ｗ（１），　Ｗ（２），　．．．，　Ｗ（ｗｎ）］のテキスト表記から認識ユニットラベルへ変換し、このラベルにしたがって教師なし話者適応パタン２００８を連結し、認識対象単語の標準パタンを作成する。この認識対象単語の標準パタンを用いて、音声特徴量抽出手段２００２の出力である音声特徴量に対して照合を行い、音声認識結果２１０１を出力する。このとき、入力音声２００１は教師なし適応用に用いた単語と同一でも、それ以外の単語でも良い。音声認識結果２１０１は、入力音声２００１に対して認識対象語彙の標準パタン中で最も照合スコア（尤度）が高い単語系列のテキスト表記Ｒｗ＝［Ｗ（ｒ（１）），　Ｗ（ｒ（２）），　．．．，　Ｗ（ｒ（ｍ））］としてを出力される。ここで、ｒ（ｉ）は音声認識結果の単語時系列のｉ番目の単語の認識辞書単語番号を示す。また、ｍは認識単語系列の単語数を示す。
従って、この実施の形態２３の音声認識装置によれば、このように認識結果信頼度付きの教師なし話者適応行って得られた教師なし話者適応パタン２００８を用いて音声認識を行うので、話者適応学習用認識結果が誤った場合のパラメータの誤った更新を防ぐことができ、認識率が向上する。
【００９６】
また次の発明によれば、認識結果信頼度付き教師なし話者適応手段は、前記認識ユニットの標準パタンパラメータ更新用の分割された適応データを用い、最尤推定によって標準パタンのパラメータを推定し、認識ユニットの標準パタンのパラメータ更新に用いた適応データの認識結果信頼度の合計値に基づき、最尤推定前後のパラメータの値の線形補間によって前記標準パタンパラメータから前記話者適応パタンのパラメータへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。
【００９８】
また次の発明によれば、認識結果信頼度付き教師なし話者適応手段は、前記認識ユニットの標準パタンのパラメータ更新用の分割された適応データを用い、認識結果信頼度によって適応データのパラメータ学習への重みを計算して、重み付けされた適応データによって前記標準パタンパラメータから前記話者適応パタンのパラメータへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。
【０１０２】
また次の発明によれば、認識結果信頼度付き教師なし話者適応手段は、前記第１の発声の認識結果信頼度の値によって更新方法を切り替えるので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。
【０１０３】
また次の発明によれば、標準パタンのパラメータは、クラスタリングによってグループ化し、グループ内のパラメータの更新用の分割された適応データと認識結果信頼度を用いて

グループに共通なパラメータの変動量を演算し、前記パラメータ変動量によって前記標準パタンのグループのパラメータを前記話者適応パタンのグループのパラメータへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１０４】

また次の発明によれば、クラスタリングは、木構造クラスタリングを行って木構造状に標準パタンのパラメータをクラスタリングし、木構造のノード以下に属する標準パタンのパラメータ更新用の分割された適応データの認識結果信頼度が閾値以上であるノード以下の標準パタンのパラメータをグループとして、グループ内のパラメータの更新用の分割された適応データと認識結果信頼度を用いてグループに共通なパラメータの変動量を演算し、前記変動量によって前記標準パタンのグループのパラメータを前記話者適応パタンのグループのパラメータへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１０５】

また次の発明によれば、標準パタン、及び前記話者適応パタンとして、連続混合分布型隠れマルコフモデルを用いるので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１０６】

また次の発明によれば、連続混合分布型隠れマルコフモデルのシンボル出力確率密度関数を構成する要素分布関数は、ガウス分布であるので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１０７】

また次の発明によれば、認識結果信頼度付き教師なし話者適応手段において更新するパラメータは前記ガウス分布の平均ベクトルであるので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１０８】

また次の発明によれば、ガウス分布の平均ベクトルの更新は、適応データが存在するガウス分布の平均ベクトルは認識結果信頼度付き更新を行い、適応データが存在しないガウス分布の平均ベクトルは適応データが存在するガウス分布の更新前後の平均ベクトルの値の差分ベクトルを用いた補間によって前記標準パタンのパラメータを前記話者適応パタンのパラメータへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１０９】

また次の発明によれば、認識結果信頼度付き教師なし話者適応手段は、認識結果信頼度を用いた重回帰写像モデルに基づく話者適応によって、前記標準パタンのパラメータであるガウス分布の平均ベクトルを前記話者適応パタンのガウス分布の平均ベクトルへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１１０】

また次の発明によれば、重回帰写像モデルに基づく話者適応は、標準パタンのガウス分布をクラスタリングしてグループ化し、グループ内のガウス分布更新用の適応データと認識結果信頼度に基づいてガウス分布のグループに１つの回帰係数を演算し、標準パタンの平均ベクトルを回帰係数を用いて話者適応パタンの平均ベクトルへ更新するので、適応学習用音声認識結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【０１１１】

また次の発明によれば、請求項１～２２のうちいずれかに記載の話者適応化装置によって更新された教師なし話者適応パタンと、話者の入力音声から音声特徴量を抽出する音声特徴量抽出手段と、前記音声特徴量抽出手段が抽出した音声特徴量と前記教師なし話者適応パタンとを照合して認識結果を出力する照合手段と、を備えたので、適応学習用音声認識

結果が誤った場合でも、標準パタンのパラメータの誤った更新を防ぐことができ認識率が向上する。

【図面の簡単な説明】

【図１】この発明による話者適応化装置の実施の形態１の構成を示すブロック図である。

【図２】この発明による話者適応化装置の実施の形態２の構成を示すブロック図である。

【図３】この発明による話者適応化装置の実施の形態３の動作説明図である。

【図４】この発明による話者適応化装置の実施の形態４の動作説明図である。

【図５】この発明による話者適応化装置の実施の形態５の動作説明図である。

【図６】この発明による話者適応化装置の実施の形態６の動作説明図である。

【図７】この発明による話者適応化装置の実施の形態７の構成を示すブロック図である。

【図８】この発明による話者適応化装置の実施の形態８の構成を示すブロック図である。

【図９】この発明による話者適応化装置の実施の形態１０の構成を示すブロック図である。

【図１０】この発明による話者適応化装置の実施の形態１２の構成を示すブロック図である。

【図１１】この発明による話者適応化装置の実施の形態１３の構成を示すブロック図である。

【図１２】この発明による話者適応化装置の実施の形態１４の構成を示すブロック図である。

【図１３】この発明による話者適応化装置の実施の形態１５の構成を示すブロック図である。

【図１４】この発明による話者適応化装置の実施の形態１６の構成を示すブロック図である。

【図１５】この発明による話者適応化装置の実施の形態１６の動作説明図である。

【図１６】この発明による話者適応化装置の実施の形態２０の構成を示すブロック図である。

【図１７】この発明による話者適応化装置の実施の形態２０の動作説明図である。

【図１８】この発明による話者適応化装置の実施の形態２１の構成を示すブロック図である。

【図１９】この発明による話者適応化装置の実施の形態２２の構成を示すブロック図である。

【図２０】この発明による音声認識装置の実施の形態２３の構成を示すブロック図である。

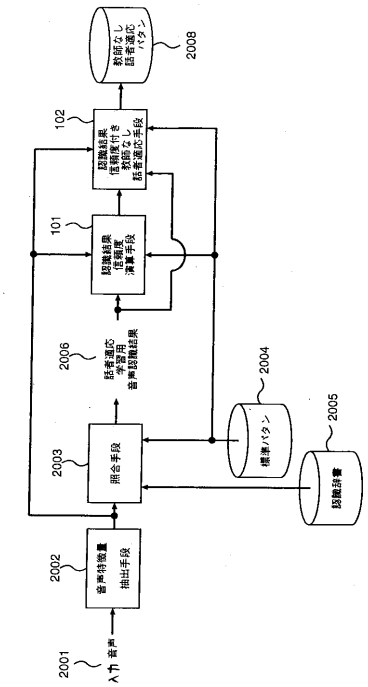
【図２１】従来の話者適応化装置の構成を示すブロック図である。

【符号の説明】

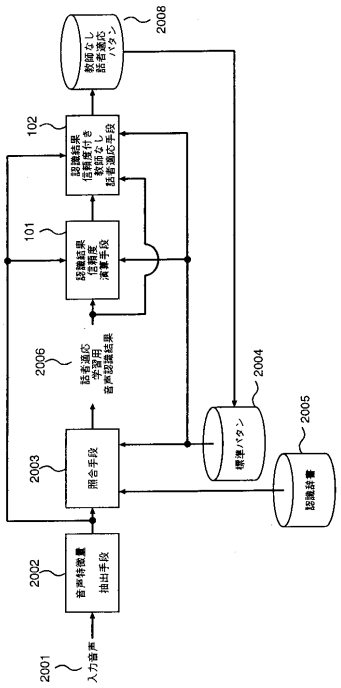
１０１　認識結果信頼度演算手段
１０２　認識結果信頼度付き教師なし話者適応手段
７０１　音声データセグメンテーション手段
７０２　認識結果信頼度付き標準パタンパラメータ更新手段
８０１　標準パタンパラメータ最尤推定手段
８０２　認識結果信頼度に基づくパラメータ線形補間手段
９０１　認識結果信頼度重み付き学習データによる適応学習手段
１００１　複数認識結果候補出力照合手段
１００２　複数認識結果候補信頼度演算手段
１００３　複数認識結果候補信頼度付き教師なし話者適応手段
１１０１　認識結果信頼度比較手段
１２０１　認識結果信頼度による話者適応方式選択手段
１２０２－１～Ｍ　認識結果信頼度付き教師なし話者適応手段　１～Ｍ
１３０１　標準パタンパラメータクラスタリング手段
１３０２　認識結果信頼度付きパラメータグループ教師なし話者適応手段

１４０１　標準パタンパラメータ木構造クラスタリング手段
１４０２　木構造化パラメータに基づく標準パタンパラメータグループ化手段
１６０１　標準パタンパラメータ補間手段
１８０１　認識結果信頼度付き重回帰写像モデルに基づく話者適応手段
１９０１　ガウス分布グループ化手段
２００１　入力音声
２００２　音声特徴量抽出手段
２００３　照合手段
２００４　標準パタン
２００５　認識辞書
２００６　話者適応学習用音声認識結果
２００７　教師なし話者適応手段
２００８　教師なし話者適応パタン
２１０１　音声認識結果

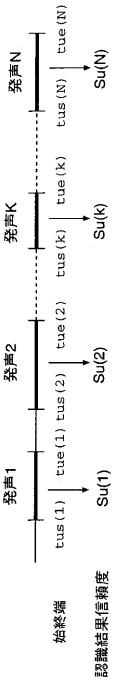
【図１】



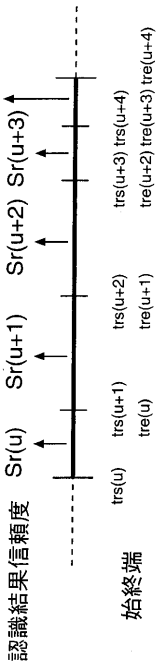
【図２】



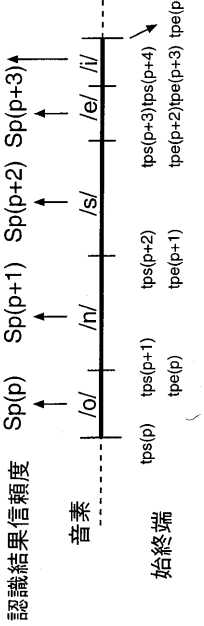
【図３】


発声1
発声2
発声K
発声N
tus(1) tue(1) tus(2) tue(2) tus(k) tue(k) tus(N) tue(N)
始終端
認識結果信頼度
Su(1) Su(2) Su(k) Su(N)

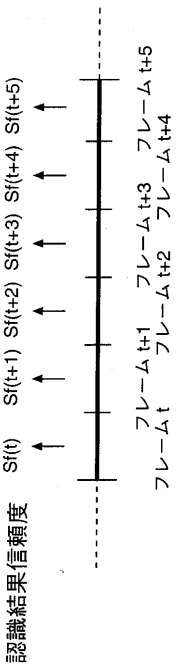
【図４】


Sr(u+4)
認識結果信頼度 Sr(u) Sr(u+1) Sr(u+2) Sr(u+3)
始終端
trs(u) trs(u+1) trs(u+2) trs(u+3) trs(u+4)
tre(u) tre(u+1) tre(u+2) tre(u+3) tre(u+4)

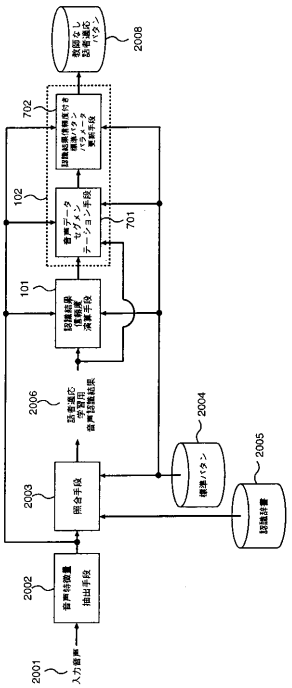
【図５】


Sp(p+4)
認識結果信頼度 Sp(p) Sp(p+1) Sp(p+2) Sp(p+3)
音素 /o/ /n/ /s/ /e/ /i/
始終端
tps(p) tps(p+1) tps(p+2) tps(p+3)tps(p+4)
tpe(p) tpe(p+1) tpe(p+2)tpe(p+3) tpe(p+4)

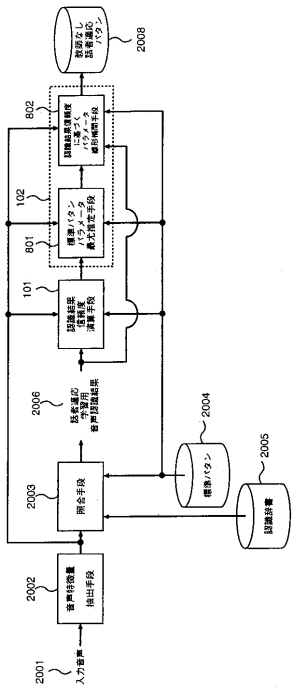
【図６】


認識結果信頼度 Sf(t) Sf(t+1) Sf(t+2) Sf(t+3) Sf(t+4) Sf(t+5)
フレーム t フレーム t+1 フレーム t+2 フレーム t+3 フレーム t+4 フレーム t+5

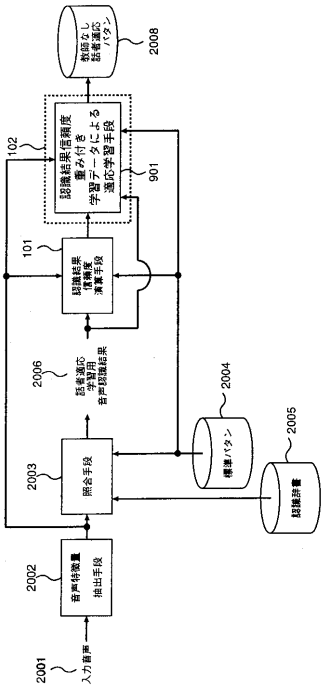
【図７】


2001
入力音声
2002
音声特徴量抽出手段
2003
照合手段
2004
標準パタン
2005
認識辞書
2006
話者適応学習用音声認識結果
101
認識結果信頼度演算手段
102
音声データモデレーション手段
701
702
認識結果信頼度付きパラメータ更新手段
2008
教師なし話者適応パタン

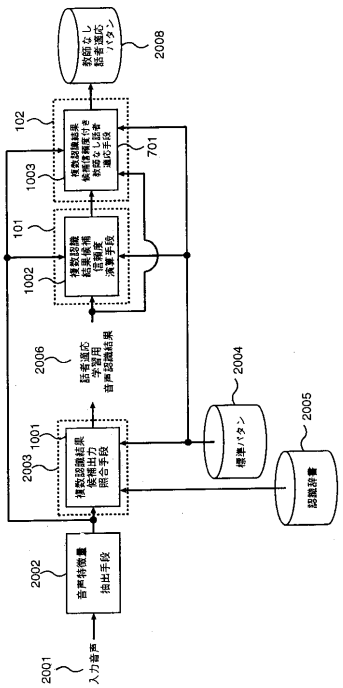
【図８】


2001
入力音声
2002
音声特徴量抽出手段
2003
照合手段
2004
標準パタン
2005
認識辞書
2006
話者適応学習用音声認識結果
101
認識結果信頼度演算手段
102
801
標準パタンパラメータ最尤推定手段
802
認識結果信頼度付きパラメータ逐次補間手段
2008
教師なし話者適応パタン

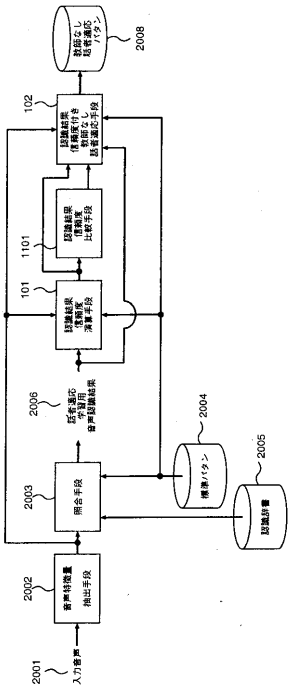
【図９】


2001
入力音声
2002
音声特徴量抽出手段
2003
照合手段
2004
標準パタン
2005
認識辞書
2006
話者適応学習用音声認識結果
101
認識結果信頼度演算手段
102
認識結果信頼度重み付き学習データによる適応学習手段
901
2008
教師なし話者適応パタン

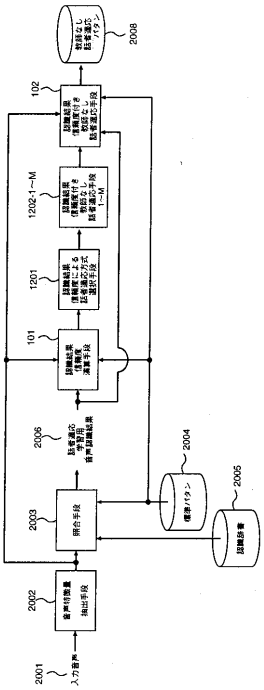
【図１０】


2001
入力音声
2002
音声特徴量抽出手段
2003
1001
複数認識結果候補出力照合手段
2004
標準パタン
2005
認識辞書
2006
話者適応学習用音声認識結果
101
1002
複数認識結果候補信頼度演算手段
102
1003
複数認識結果候補付き信頼度付き学習データによる適応学習手段
701
2008
教師なし話者適応パタン

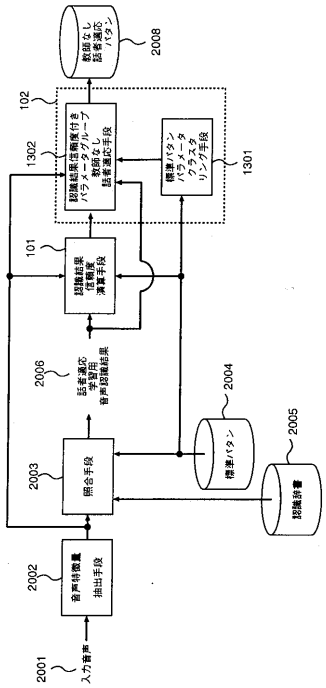
【図１１】



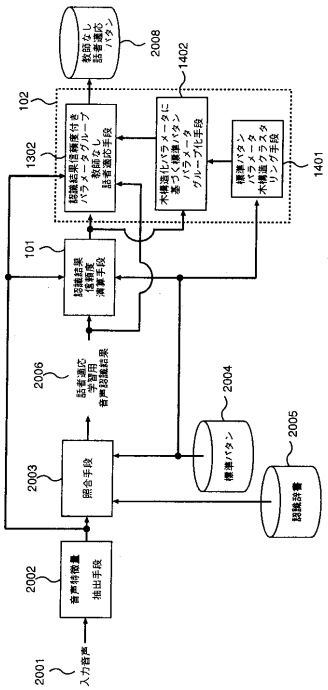
【図１２】



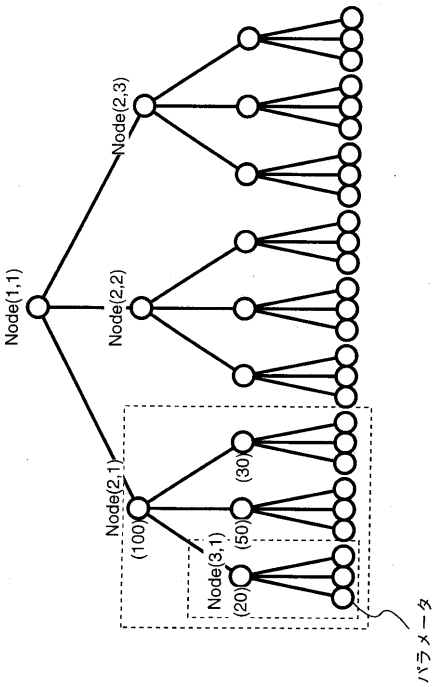
【図１３】



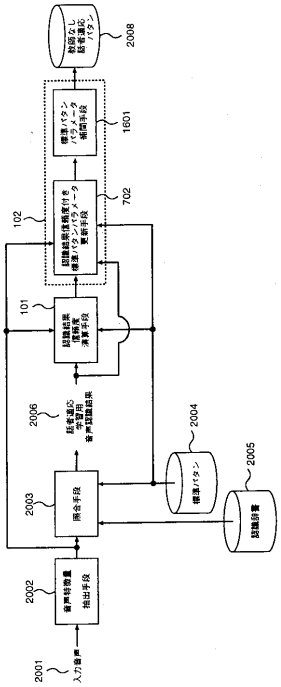
【図１４】



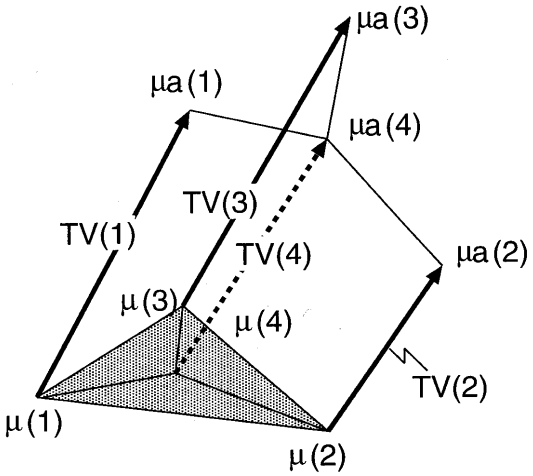
【図１５】



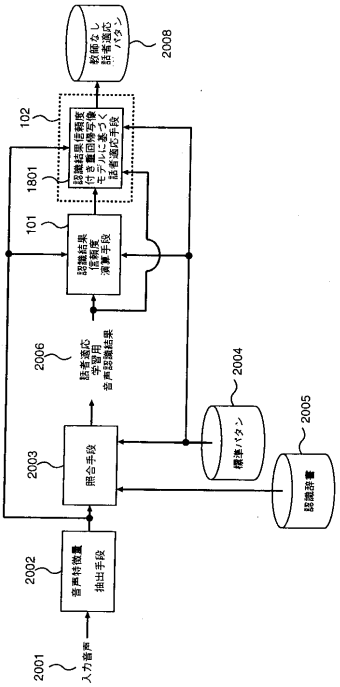
【図１６】



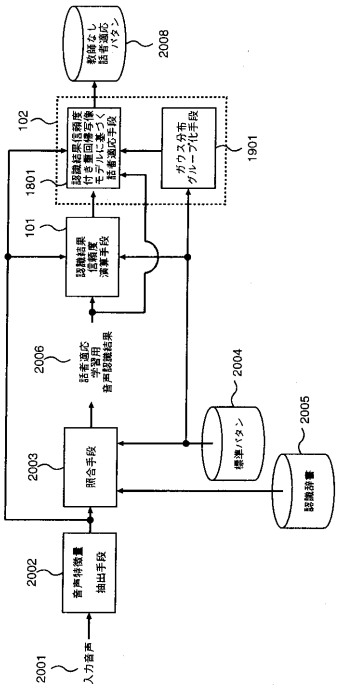
【図１７】



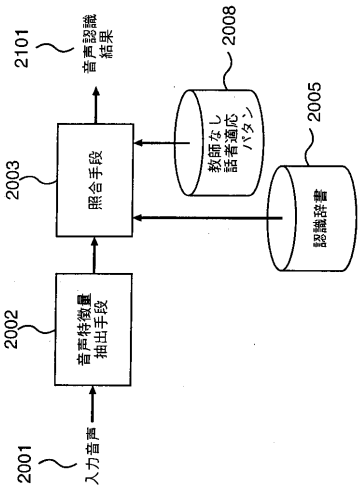
【図１８】



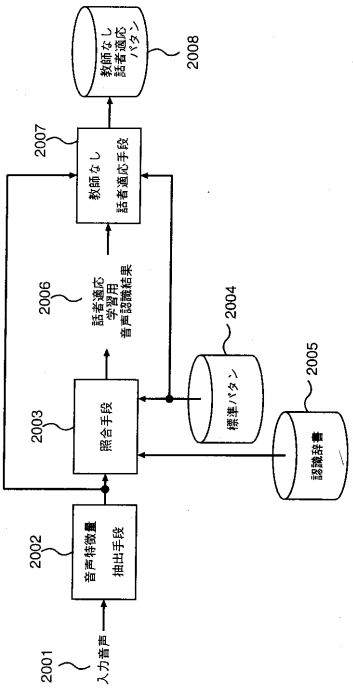
【図１９】


教師なし話者適応パターン
2008
102
認識結果信頼度尺度モデルに基づく話者適応手段
ガウス分布グループ化手段
1901
1801
101
認識結果信頼度演算手段
2006
話者適応学習用音声認識結果
2004
標準パターン
2005
認識辞書
2003
照合手段
2002
音声特徴量抽出手段
2001
入力音声

【図２０】


2101
音声認識結果
2008
教師なし話者適応パターン
2005
認識辞書
2003
照合手段
2002
音声特徴量抽出手段
2001
入力音声

【図２１】


教師なし話者適応パターン
2008
2007
教師なし話者適応手段
2006
話者適応学習用音声認識結果
2004
標準パターン
2005
認識辞書
2003
照合手段
2002
音声特徴量抽出手段
2001
入力音声

フロントページの続き

審査官　渡邊　聡

(56)参考文献　特開平０５－０８８６９３（ＪＰ，Ａ）
特開平０９－１１４４８２（ＪＰ，Ａ）
特開平０９－２５８７７０（ＪＰ，Ａ）
鈴木忠　他，母音認識のための教師なし話者適応化法，電子情報通信学会論文誌A，電子情報通信学会，１９８６年，Ｖｏｌ．６９、NO．１１，１４４６－１４５２

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G10L 15/10
G10L 15/06