

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3876703号
(P3876703)

(45) 発行日 平成19年2月7日(2007.2.7)

(24) 登録日 平成18年11月10日(2006.11.10)

(51) Int. Cl.

G 1 0 L 15/06 (2006.01)

F I

G 1 0 L 15/06 3 1 0 S

G 1 0 L 15/06 3 0 0 J

G 1 0 L 15/06 3 0 0 M

請求項の数 4 (全 6 頁)

(21) 出願番号 特願2001-378341 (P2001-378341)
 (22) 出願日 平成13年12月12日(2001.12.12)
 (65) 公開番号 特開2003-177779 (P2003-177779A)
 (43) 公開日 平成15年6月27日(2003.6.27)
 審査請求日 平成16年11月16日(2004.11.16)

(73) 特許権者 000005821
 松下電器産業株式会社
 大阪府門真市大字門真1006番地
 (74) 代理人 100097445
 弁理士 岩橋 文雄
 (74) 代理人 100109667
 弁理士 内藤 浩樹
 (74) 代理人 100109151
 弁理士 永野 大介
 (72) 発明者 脇田 由実
 大阪府門真市大字門真1006番地 松下
 電器産業株式会社内
 (72) 発明者 水谷 研治
 大阪府門真市大字門真1006番地 松下
 電器産業株式会社内

最終頁に続く

(54) 【発明の名称】 音声認識のための話者学習装置及び方法

(57) 【特許請求の範囲】

【請求項 1】

話者の学習用音声を用いて音響モデルパラメータを再学習し、話者に適応した音響モデルを作成する話者適応学習手段と、誤認識した単語の認識結果に相当する音素又は音節からなる音響モデル系列を正解の音素又は音節系列として発音辞書に追加する話者登録学習手段と、認識しやすさが発声内容に依存するかどうかを判断する手段と、各話者の認識しやすさと発声内容の依存の強さによって、話者適応学習手段と話者登録学習手段との選択を行い、どちらかの学習を話者に促す手段を有することを特徴とする話者学習装置。

【請求項 2】

認識しやすさが発声内容に依存するかどうかを判断した結果、依存すると判断された場合には話者登録学習手段を用い、依存しないと判断された場合には話者適応学習手段を用いることを特徴とする請求項 1 記載の話者学習装置。

10

【請求項 3】

認識のしやすさが発声内容に依存するかどうかを判断する手段は、認識スコアが所定のしきい値以下であるか、所定のしきい値以上であっても誤認識している音素又は音節の全発声に含まれる音素又は音節に対する割合により判断を行う請求項 1 記載の話者学習装置。

【請求項 4】

話者の学習用音声を用いて音響モデルパラメータを再学習し、話者に適応した音響モデルを作成する話者適応学習ステップと、誤認識した単語の認識結果に相当する音素又は音

20

節からなる音響モデル系列を正解の音素又は音節系列として発音辞書に追加する話者登録学習ステップと、認識しやすさが発声内容に依存するかどうかを判断するステップと、各話者の認識しやすさと発声内容の依存の強さによって、話者適応学習手段と話者登録学習手段との選択を行い、どちらかの学習を話者に促すステップとを有することを特徴とする話者学習方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識における話者学習装置及び方法に関するものである。

【0002】

【従来の技術】

以下、従来の話者学習法を説明する。従来の不特定話者音声認識システムでは、なるべく不特定多数の話者に対応できる標準的な音響モデルを構築して用いているが、実用上では、話者の発声特徴は多種多様であり、全ての使用話者に対して高性能を保証する音響モデルを学習することは困難である。そこで従来は、認識しない話者について、話者自身の発声を用いて音響モデルパラメータを再学習し、話者に適応した音響モデルを再構築することにより全話者に対する性能を保証する話者適応手段をとっている。この話者適応には話者の特徴を捉えるに十分な多くの学習用音声が必要であるが、発声者の負担になるので、最低限の発声回数に絞る様々な工夫がなされている（たとえば、特許第2037877）。一方、別の学習方法として、誤認識した単語の認識結果に相当する音響モデル系列を正解系列として発音辞書に追加し、誤った系列として認識したものを正しい系列として認識することを可能とする話者登録方法もある（特開平8-171396号公報）。

【0003】

【発明が解決しようとする課題】

従来の話者適応法は、学習データが十分あれば、原理的に確実に認識性能を向上できる手法であるが、ほとんど全ての実用上システムでは行われているように、話者の学習負担を考慮して発声回数が絞られた場合、学習データに存在しない一部の発声に対して、逆に認識率が低下してしまう可能性があるという問題がある。一方、従来の話者登録法は、学習された発声部分の認識率は確実に向上するが、多くの発声内容で認識しにくい話者の場合は、学習時に認識しにくい全ての発声をしなければならず学習に負担がかかる、という問題がある。

【0004】

本発明の目的は、従来の話者適応学習と話者登録学習の問題点を解決し、話者に負担にならない学習発声量で、学習後に確実に認識率を向上させる話者学習法を提供するものである。

【0005】

【課題を解決するための手段】

上述した課題を解決するために、本発明は、話者の学習用音声を用いて音響モデルパラメータを再学習し、話者に適応した音響モデルを作成する話者適応学習手段と、誤認識した単語の認識結果に相当する音素又は音節からなる音響モデル系列を正解の音素又は音節系列として発音辞書に追加する話者登録学習手段と、認識しやすさが発声内容に依存するかどうかを判断する手段と、各話者の認識しやすさと発声内容の依存の強さによって、話者適応学習手段と話者登録学習手段との選択を行うものである。

【0006】

【発明の実施の形態】

以下、図面を参照して本発明の実施形態の話者学習を説明する。

【0007】

図1は本発明の実施形態の話者学習のブロック図である。

【0008】

各話者が自分に対する認識性能を向上させる必要を感じた場合に選択するように設定され

10

20

30

40

50

た話者学習機能において、まず、システムからユーザに対し特定単語発声を促し、話者の特定単語発声が入力される。この発声内容は、各話者に対して、予め準備した標準音声がどのくらい適切かを判断するのに必要な最低限の内容であり、たとえば日本語認識の場合は、5母音を全て含む単語「マイクテスト」などの内容がふさわしい。システムが単語認識の場合には5母音が全て含まれるように対象単語から複数単語を選択しても良い。

【0009】

この発声に対して音声認識処理1で通常の認識処理が行われ、認識スコア算出処理2で認識結果と認識信頼度スコアが計算される。認識結果は、認識結果の音素または音節系列と正解の音素系列とを比較し、異なっている部分を誤りとし一致している部分を正解として、正解系列の各音素毎に正誤を記録しておく。また信頼度スコアは、たとえば正解音素または音節系列と発声された結果との各音素または音節毎の音響的距離スコアであり、距離尺度として重み付きケプストラム距離を用いた場合は、各音素の信頼度は式1で算出されるものを用いてもよい。

10

【0010】

【数1】

$$Con(k) = \sum_{m=s}^{m=e} d(m) / (e-s) \quad (式1)$$

$$\text{但し、} d(m) = \sum_{n=1}^L w(n)(c_n - c'_n)^2$$

20

$Con(k)$: 音素kの信頼度スコア

s, e : 音素kの始端及び終端フレーム

$d(m)$: 入力音声のmフレームにおける照合時に採用された距離値

$w(n)$: n次における係数間距離に対する重み係数

c_n : 入力音声のn次ケプストラム係数

c'_n : 標準音声のn次ケプストラム係数

30

【0011】

学習法決定処理3では、信頼度スコアが閾値以下であるか、閾値以上であったとしても誤認識している音素または音節（適応候補音素または音節と呼ぶ）の全発声に含まれる音素または音節に対する割合を計算する。この割合が大きい場合は、発声内容に依存せず話者の発声特徴が標準音声に適応していないことが推定され、全ての標準音声を話者に適応するように学習する必要があると考えられる。また、この割合が小さい場合には、誤認識は発声内容に依存しており、話者の発声特徴と標準音声は適応しているが、特定の発声においてのみ学習が必要であると考えられる。従って、この割合が一定値以上である場合、話者適応学習を選択し、一定値以下である場合、話者登録学習を選択する。

40

【0012】

話者適応学習を選択した場合は、話者適応処理4で、ユーザにさらに適応するに必要最低限の発声を促す。話者適応法は、たとえば、特開平5-53599に記載のVFS法を利用した場合には、標準音響モデルと学習用入力音声パラメータとをマッチングし、対応するパラメータの関係からファジー級関数を求め、求められた関数を重みとして、標準音声を学習用入力音声に近づくように標準音響モデルのパラメータを更新している。

【0013】

また、話者登録学習を選択した場合には、話者登録処理5で、学習決定処理で算出した適応候補音素または音節が含まれている単語のみの発声を促し、適応候補音素に相当する音

50

素系列を含む単語の音素系列に、発声に対する音素または音節認識結果系列を発音辞書 7 に追加する。たとえば、「メニュー」という単語が誤認識を起こす場合、この単語のみの発声を促し、その認識結果が「デニュー」であったとする。音響モデルとして音素モデルを使用している場合には、「メニュー」の正しい音素モデル系列は /m e ny u u/ であり、認識結果音素系列は /d e ny u u/ である。この話者の場合、単語の始めであり、次に /e/ が続く音素 /m/ は /d/ に誤る傾向があることがわかる。そこで、認識対象単語の中で、単語の先頭であり、次が /e/ である /m/ は /d/ と誤っても /m/ と認識するように、発音辞書に音素系列を追加する。この例の場合には、もともと辞書上で「メニュー /m e ny u u/」であったところに /d e ny u u/ を追加し、「メニュー /m e ny u u/ または /d e ny u u/」と辞書を変更する。これにより、この話者が「メニュー」を /d e ny u u/ と認識しても結果的には「メニュー」が認識できることになる。

10

【0014】

以上のように、話者の発声が発声内容に依存せずに誤るかどうかを推定し、発声内容に依存しない場合は話者適応学習、依存する場合は話者登録学習を行うことにより、従来の話者適応学習で、適応するための多くの学習発声をしたにもかかわらず認識率が低下する問題を、話者適応学習ではなく話者登録学習を行うことで解決することができる。また、従来の話者登録学習で、多くの単語を発声しなければ学習できなかった問題を、話者登録学習ではなく話者適応学習を行うことで解決することができる。

【0015】

以上詳述したように、本発明に係る実施形態の話者学習法は、各話者の認識しやすさと発声内容の依存の強さによって、話者適応学習を行うか話者登録学習を行うかの選択を行い、どちらかの学習を話者に促すことにより、従来の話者適応学習において、適応するための多くの学習発声をしたにもかかわらず認識率が低下する問題を、話者適応学習のかわりに話者登録学習を自動選択することで解決することができる。また、従来の話者登録学習において、多くの単語を発声しなければ学習できなかった問題を、話者登録学習のかわりに話者適応学習を自動選択することで解決することができる。従って、話者に負担にならない程度の学習量で、確実に認識率を向上させることが可能である話者学習法を提供するものである。

20

【0017】

以上詳述したように、本発明に係る実施形態の話者学習法は、認識しやすさが発声内容に依存するかどうかを判断した結果、依存すると判断された場合には話者登録学習を行い、依存しないと判断された場合には話者適応学習を行うことにより、従来の話者適応学習において、適応するための多くの学習発声をしたにもかかわらず認識率が低下する問題を、話者適応学習のかわりに話者登録学習を自動選択することで解決することができる。また、従来の話者登録学習において、多くの単語を発声しなければ学習できなかった問題を、話者登録学習のかわりに話者適応学習を自動選択することで解決することができる。従って、話者に負担にならない程度の学習量で、確実に認識率を向上させることが可能である話者学習法を提供するものである。

30

【0019】

以上詳述したように、本発明は、各話者の認識しやすさと発声内容の依存の強さによって、話者適応学習を行うか話者登録学習を行うかの選択を行い、どちらかの学習を話者に促すことにより、従来の話者適応学習において、適応するための多くの学習発声をしたにもかかわらず認識率が低下する問題を、話者適応学習のかわりに話者登録学習を自動選択することで解決することができる。

40

【図面の簡単な説明】

【図1】本発明の一実施例である話者学習法ブロック図

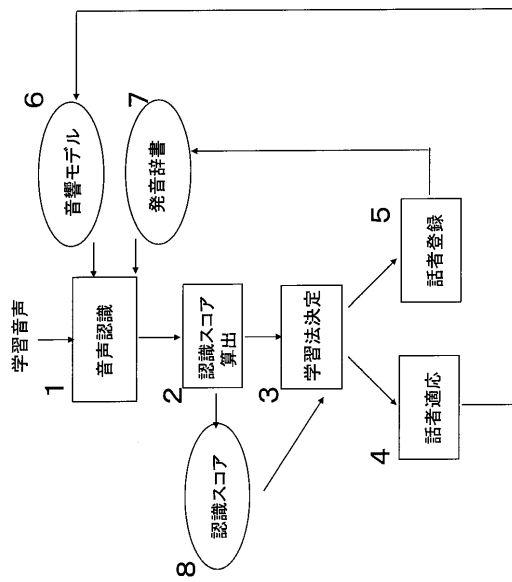
【符号の説明】

- 1 音声認識
- 2 認識スコア算出
- 3 学習法決定

50

- 4 話者適応
- 5 話者登録
- 6 音響モデル
- 7 発音辞書
- 7 発音辞書
- 8 認識スコアバッファ

【図 1】



フロントページの続き

(72)発明者 芳澤 伸一

大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式会社内

審査官 榎本 剛

(56)参考文献 特許第 3 7 9 5 4 0 9 (J P , B 2)

特許第 3 1 6 6 7 0 8 (J P , B 2)

(58)調査した分野(Int.Cl. , D B 名)

G10L 15/06

JSTPlus(JDream2)