

(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2010-107982

(P2010-107982A)

(43) 公開日 平成22年5月13日(2010.5.13)

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

審査請求 有 請求項の数 9 O L (全 16 頁)

(21) 出願番号	特願2009-248013 (P2009-248013)	(71) 出願人	502192546 清華大学 中華人民共和国北京市海淀区清華大学 郵 編 1 0 0 0 8 4
(22) 出願日	平成21年10月28日 (2009.10.28)	(71) 出願人	395015319 株式会社ソニー・コンピュータエンタテイ ンメント 東京都港区南青山二丁目6番21号
(31) 優先権主張番号	200810225354.0	(74) 代理人	100105924 弁理士 森下 賢樹
(32) 優先日	平成20年10月31日 (2008.10.31)	(74) 代理人	100109047 弁理士 村田 雄祐
(33) 優先権主張国	中国 (CN)	(74) 代理人	100109081 弁理士 三木 友由

最終頁に続く

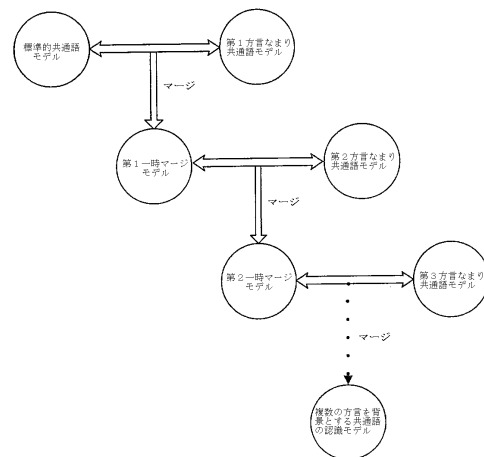
(54) 【発明の名称】 コンピュータによる複数の方言を背景とする共通語音声認識のモデリング方法及びシステム

(57) 【要約】 (修正有)

【課題】データ量が少ないまま方言なまりの共通語に対する認識率を高め、同時に標準的共通語に対する認識率が顕著に下がらないことを保証するシステムの提供。

【解決手段】まず標準的共通語のトレーニングデータに基づきトライフォンによる標準的共通語モデルを生成し、第1、第2方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第1、第2方言なまり共通語モデルをそれぞれ生成する。更に、標準的共通語モデルを用いて第1方言なまり共通語のディベロップメントデータを認識することにより得られた第1混同行列に応じ、第1方言なまり共通語モデルを標準的共通語モデルの中にマージして一時マージモデルを得る。最後は、一時マージモデルにより第2方言なまり共通語のディベロップメントデータを認識することにより得られた第2混同行列に応じて、第2方言なまり共通語モデルを一時マージモデルの中にマージして認識モデルを得る。

【選択図】 図 1



## 【特許請求の範囲】

## 【請求項 1】

複数の方言を背景とする共通語音声認識のモデリング方法であって、

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第1方言なまり共通語モデルを生成し、第2種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第2方言なまり共通語モデルを生成する工程と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて前記第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして一時マージモデルを得る工程と、

(3) 前記一時マージモデルを用いて第2種の方言なまり共通語のディベロップメントデータを認識することにより第2混同行列を生成し、当該第2混同行列に応じて前記第2方言なまり共通語モデルを前記一時マージモデルの中にマージして認識モデルを得る工程と、

を含むことを特徴とする複数の方言を背景とする共通語音声認識のモデリング方法。

## 【請求項 2】

$x$  で被認識音声の観測特徴ベクター、 $s$  で前記標準的共通語モデルにおける隠れマルコフ状態、 $d_1$  で前記第1方言なまり共通語モデルにおける隠れマルコフ状態、 $d_2$  で前記第2方言なまり共通語モデルにおける隠れマルコフ状態、を表す場合、下記の数式で与えられる前記一時マージモデルにおける確率密度関数は

$$p'(x|s) = \alpha_1 p(x|s) + (1 - \alpha_1) p(x|d_1) p(d_1|s)$$

であり、

その中で、 $\alpha_1$  は線形補間係数であり、 $0 < \alpha_1 < 1$  を満たし、

前記認識モデルの確率密度関数は

## 【数 4】

$$p''(x|s) = \sum_{k=1}^K w_k^{(sc)'} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)'} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)'} N_{pq}^{(dc2)}(\cdot)$$

であり、その中で、 $w_k^{(sc)'}$  は前記標準的共通語モデルにおける隠れマルコフ状態が占める重み、 $w_{mn}^{(dc1)'}$  と  $w_{pq}^{(dc2)'}$  はそれぞれ前記第1方言なまり共通語モデル、前記第2方言なまり共通語モデルにおける隠れマルコフ状態が占める重み、 $K$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の混合数、 $N_k^{(sc)}(\cdot)$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の要素、 $M$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間の発音バリエーションの数、 $N$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の混合数、 $N_{mn}^{(dc1)}(\cdot)$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の要素、 $P$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間の発音バリエーションの数、 $Q$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数、 $N_{pq}^{(dc2)}(\cdot)$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素、を示すことを特徴とする請求項1に記載のモデリング方法。

## 【請求項 3】

複数の方言を背景とする共通語音声認識のモデリングプログラムであって、コンピュータに

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第1方言なまり共通語モデルを生成し、第2種の方言なまり共通語のディベ

10

20

30

40

50

ロップメントデータに基づいてモノフォンによる第2方言なまり共通語モデルを生成する機能と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて前記第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして一時マージモデルを得る機能と、

(3) 前記一時マージモデルを用いて第2種の方言なまり共通語のディベロップメントデータを認識することにより第2混同行列を生成し、当該第2混同行列に応じて前記第2方言なまり共通語モデルを前記一時マージモデルの中にマージして認識モデルを得る機能と、

10

を実行させることを特徴とするコンピュータプログラム。

【請求項4】

n種(nは2以上の自然数)の方言を背景とする共通語音声認識のモデリング方法であって、

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1～n種の方言なまり共通語のそれぞれに対し、そのディベロップメントデータに基づいてモノフォンによる第1～第n方言なまり共通語モデルを生成する工程と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして第1一時マージモデルを得る工程と、

20

(3) 第(i-1)一時マージモデル(iは $2 < i < n$ を満たす自然数)を用いて第i種の方言なまり共通語のディベロップメントデータを認識することにより第i混同行列を生成し、当該第i混同行列に応じて第i方言なまり共通語モデルを前記第(i-1)一時マージモデルの中にマージする動作を、 $i = 2$ から $i = n$ まで順に繰り返すことにより、認識モデルを得る工程と、

を含むことを特徴とする複数の方言を背景とする共通語音声認識のモデリング方法。

【請求項5】

n種(nは2以上の自然数)の方言を背景とする共通語音声認識のモデリングプログラムであって、コンピュータに

30

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1～n種の方言なまり共通語のそれぞれに対し、そのディベロップメントデータに基づいてモノフォンによる第1～第n方言なまり共通語モデルを生成する機能と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして第1一時マージモデルを得る機能と、

(3) 第(i-1)一時マージモデル(iは $2 < i < n$ を満たす自然数)を用いて第i種の方言なまり共通語のディベロップメントデータを認識することにより第i混同行列を生成し、当該第i混同行列に応じて第i方言なまり共通語モデルを前記第(i-1)一時マージモデルの中にマージする動作を、 $i = 2$ から $i = n$ まで順に繰り返すことにより、認識モデルを得る機能と、

40

を実行させることを特徴とするコンピュータプログラム。

【請求項6】

請求項3または5に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。

【請求項7】

複数の方言を背景とする共通語音声認識のモデリングシステムであって、モデル生成ユ

50

ニットと、当該モデル生成ユニット全体の動作を制御する制御ユニットとを備え、

前記モデル生成ユニットは、

標準的共通語のトレーニングデータが記憶されている標準的共通語トレーニングデータベースと、

第1、第2種の方言なまり共通語のディベロップメントデータがそれぞれ記憶されている第1、第2ディベロップメントベースと、

前記標準的共通語トレーニングデータベースに記憶されている標準的共通語のトレーニングデータに基づいて、トライフォンによる標準的共通語モデルを生成するための標準的共通語モデル生成部と、

前記第1、第2ディベロップメントデータベースにそれぞれ記憶されている第1、第2種の方言なまり共通語のディベロップメントデータに基づいて、モノフォンによる第1、第2方言なまり共通語モデルを生成するための第1、第2モデル生成部と、

前記標準的共通語モデル生成部により生成された標準的共通語モデルを用いて、前記第1ディベロップメントデータベースに記憶されている第1種の方言なまり共通語のディベロップメントデータを認識することにより、第1混同行列を生成するための第1混同行列生成部と、

前記第1混同行列生成部により生成された第1混同行列に応じて、前記第1モデル生成部により生成された第1方言なまり共通語モデルを、前記標準的共通語モデル生成部により生成された標準的共通語モデルの中にマージして一時マージモデルを生成するための第1モデルマージ部と、

前記第1モデルマージ部により生成された一時マージモデルを用いて、前記第2ディベロップメントデータベースに記憶されている第2種の方言なまり共通語のディベロップメントデータを認識することにより、第2混同行列を生成するための第2混同行列生成部と

、  
前記第2混同行列生成部により生成された第2混同行列に応じて、前記第2モデル生成部により生成された第2方言なまり共通語モデルを、前記第1モデルマージ部により生成された一時マージモデルの中にマージして認識モデルを生成するための第2モデルマージ部と

を備えることを特徴とする複数の方言を背景とする共通語音声認識のモデリングシステム。

【請求項8】

xで被認識音声の観測特徴ベクター、sで前記標準的共通語モデルにおける隠れマルコフ状態、d1で前記第1方言なまり共通語モデルにおける隠れマルコフ状態、d2で前記第2方言なまり共通語モデルにおける隠れマルコフ状態、を表す場合、下記の数式で与えられる前記一時マージモデルにおける確率密度関数は

$$p'(x|s) = \alpha_1 p(x|s) + (1 - \alpha_1) p(x|d_1) p(d_1|s)$$

であり、

その中で、 $\alpha_1$ は線形補間係数であり、 $0 < \alpha_1 < 1$ を満たし、

前記認識モデルの確率密度関数は

【数5】

$$p''(x|s) = \sum_{k=1}^K w_k^{(sc)'} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)'} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)'} N_{pq}^{(dc2)}(\cdot)$$

であり、その中で、 $w_k^{(sc)'}$ は前記標準的共通語モデルにおける隠れマルコフ状態が占める重み、 $w_{mn}^{(dc1)'}$ と $w_{pq}^{(dc2)'}$ はそれぞれ前記第1方言なまり共通語モデル、前記第2方言なまり共通語モデルにおける隠れマルコフ状態が占める重み、Kは標準的共通語モデルの隠れマルコフ状態sの混合正規分布の混合数、 $N_k^{(sc)}(\cdot)$ は標準的共通語モデルの隠れマルコフ状態sの混合正規分布の要素、Mは前記第1方言なまり共通語モデルの隠れマルコフ状態d1と標準的共通語モデルの隠れマルコフ状態sの間での発音バリエーションの数、Nは前記第1方言なまり共通語モデルの隠れマル

10

20

30

40

50

コフ状態  $d_1$  の混合正規分布の混合数、 $N_{m_n}^{(d_c 1)}(\cdot)$  は前記第 1 方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の要素、 $P$  は前記第 2 方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音バリエーションの数、 $Q$  は前記第 2 方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数、 $N_{p_q}^{(d_c 2)}(\cdot)$  は前記第 2 方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素、を示すことを特徴とする請求項 4 に記載のモデリングシステム。

【請求項 9】

請求項 7 又は 8 に記載の第 1 と第 2 モデル生成部、第 1 と第 2 混同行列生成部、第 1 と第 2 モデルマージ部のうち少なくとも一組は、単一の構成として時間分割で利用されることを特徴とする複数の方言を背景とする共通語音声認識のモデリングシステム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンピュータによる複数の方言を背景とする共通語音声認識のモデリング方法及びシステム、プログラム、並びにそのプログラムを記憶した記憶媒体に関し、コンピュータ音声認識分野に該当する。

【背景技術】

【0002】

頑健性を高めることは従来から音声認識における重要点であるとともに困難な点である。それに、言葉のなまりの問題が頑健性の低下を招く主因となっている。例えば中国語には方言が多いため、なまりの問題は顕著であり、研究の価値は非常に高い。従来音声認識システムでは、標準的共通語に対する認識率は高いが、方言によるなまりのある共通語（以下は方言なまり共通語と略称）に対する認識率は低い。この課題に対して、適応（Adaptation）などの方法を採用するのは一般的な解決策であるが、その前提としては方言なまり共通語のデータを十分に備えなければならない。また、このような方法を使用すると、標準的共通語に対する認識率は顕著に下がることがある。一方、方言の種類が多いため、それぞれの方言に対して音響モデルをトレーニングし直すと、作業の効率が低くなる。

20

【発明の概要】

30

【課題を解決するための手段】

【0003】

本発明は、データ量が少ないまま方言なまりの共通語に対する認識率を高め、同時に標準的共通語に対する認識率が顕著に下がらないことを保証するコンピュータによる複数の方言を背景とする共通語音声認識のモデリング方法及びシステムの提供を目的とする。

【0004】

本発明のコンピュータによる複数の方言を背景とする共通語音声認識のモデリング方法は、下記の工程を含む：

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第 1 種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第 1 方言なまり共通語モデルを生成し、第 2 種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第 2 方言なまり共通語モデルを生成し；

40

(2) 標準的共通語モデルを用いて第 1 種の方言なまり共通語のディベロップメントデータを認識することにより第 1 混同行列を生成し、当該第 1 混同行列に応じて第 1 方言なまり共通語モデルを標準的共通語モデルの中にマージして一時マージモデルを得て；

(3) 一時マージモデルを用いて第 2 種の方言なまり共通語のディベロップメントデータを認識することにより第 2 混同行列を生成し、当該第 2 混同行列に応じて第 2 方言なまり共通語モデルを一時マージモデルの中にマージして認識モデルを得る。

【0005】

前記方法の工程 (2) と (3) に記載のマージの方法は下記の通りである：

50

$x$  で被認識音声の観測特徴ベクター、 $s$  で前記標準的共通語モデルにおける隠れマルコフ状態、 $d_1$  で前記第1方言なまり共通語モデルにおける隠れマルコフ状態、を表す場合、下記の数式で与えられる前記一時マージモデルにおける確率密度関数は

$p'(x|s) = \alpha_1 p(x|s) + (1 - \alpha_1) p(x|d_1) p(d_1|s)$   
であり、

その中で、 $\alpha_1$  は線形補間係数であり、 $0 < \alpha_1 < 1$  を満たし、  
前記認識モデルの確率密度関数は

【数1】

$$p''(x|s) = \sum_{k=1}^K w_k^{(sc)'} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)'} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)'} N_{pq}^{(dc2)}(\cdot) \quad 10$$

であり、その中で、 $w_k^{(sc)'}$  は前記標準的共通語モデルにおける隠れマルコフ状態が占める重み、 $w_{mn}^{(dc1)'}$  と  $w_{pq}^{(dc2)'}$  はそれぞれ前記第1方言なまり共通語モデル、前記第2方言なまり共通語モデルにおける隠れマルコフ状態が占める重みを示す。 $K$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の混合数である。 $N_k^{(sc)}(\cdot)$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の要素である。 $M$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音パリエーションの数である。 $N$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の混合数である。 $N_{mn}^{(dc1)}(\cdot)$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の要素である。 $P$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音パリエーションの数である。 $Q$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数である。 $N_{pq}^{(dc2)}(\cdot)$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素である。

【0006】

本発明のコンピュータによる複数の方言を背景とする共通語音声認識のモデリング方法は下記のメリットを有する：

本発明の方法では、反復的な方法で複数の方言なまり共通語モデルを一つ一つ標準的共通語モデルの中にマージすることにより、方言ごとに音響モデルをトレーニングするような重複作業を免れ、作業の効率を高めることができる。また、本発明の方法によれば、方言なまり共通語に対する認識率を明らかに高めることができ、同時に、標準的共通語に対する認識率が下がらないばかりか、上がることもある。そのため、他の方法のように方言なまり共通語に対し適する処理をすると、標準的共通語に対する認識率は著しく下がるという課題を解決する。

【図面の簡単な説明】

【0007】

【図1】本発明のモデリング方法の原理を示す概念図である。

【図2】本発明の前記モデリング方法を実現するためのモデリングシステムの一例の機能ブロック図である。

【発明を実施するための形態】

【0008】

以下、図面を参照しながら本発明を説明する。

図1は本発明の  $n$  種（以下、 $n$  は2以上の自然数）の方言を背景とする共通語音声認識のモデリング方法の原理を示す概念図である。本モデリング方法において、

(1) 標準的共通語のトレーニングデータに基づいてトライフォン (Triphone) による標準的共通語モデルを生成し、対応しようとする第1～ $n$ 種の方言なまり共通語のそれぞれに対し、そのディベロップメントデータに基づいてモノフォン (Monophone) による第1～第 $n$ 方言なまり共通語モデルを生成し、

10

20

30

40

50

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列(Confusion Matrix)を生成し、当該第1混同行列に応じて第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして第1一時マージモデルを得る、

(3) 第 $(i - 1)$ 一時マージモデル( $i$ は $2 < i < n$ を満たす自然数)を用いて第 $i$ 種の方言なまり共通語のディベロップメントデータを認識することにより第 $i$ 混同行列を生成し、当該第 $i$ 混同行列に応じて第 $i$ 方言なまり共通語モデルを前記第 $(i - 1)$ 一時マージモデルの中にマージする動作を、 $i = 2$ から $i = n$ まで順に繰り返すことにより、最終の認識モデルを得る。

#### 【0009】

図2は上述した複数の方言を背景とする共通語音声認識のモデリングシステムの一例の機能ブロック図である。本発明のモデリングシステムはモデル生成ユニット100と制御ユニット200により構成される。図2のとおり、モデル生成ユニット100は、トレーニングデータベース(以下は「トレーニングDB」と略称)10-0と、ディベロップメントデータベース(以下は「ディベロップメントDB」と略称)10-1~10-nと、モデル生成部30-0~30-nと、混同行列生成部40-1~40-nと、モデルマージ部50-1~50-nと、を備える。

#### 【0010】

トレーニングDB10-0は、標準的共通語のトレーニングデータを記憶しているデータベースである。

ディベロップメントDB10-1~10-nは、それぞれ第1~第n種の方言なまり共通語のテストデータを記憶しているデータベースである。

モデル生成部30-0は、前記トレーニングDB10-0に記憶されている標準的共通語トレーニングデータに基づいて、トライフォンによる標準的共通語モデルを生成するためのものである。

モデル生成部30-1~30-nは、それぞれ前記ディベロップメントDB10-1~10-nに記憶されている第1~第n種の方言なまり共通語のディベロップメントデータに基づいて、モノフォンによる第1~第n方言なまり共通語モデルを生成するためのブロックである。

混同行列生成部40-1~40-nは、それぞれ対応するモデル生成部30-0~30-( $n - 1$ )により生成されたモデルを用いて、ディベロップメントDB10-1~10-nに記憶されている第1~第n種の方言なまり共通語のディベロップメントデータを認識することにより、第1~第n混同行列をそれぞれ生成するブロックである。

モデルマージ部50-1は、前記混同行列生成部40-1により生成された第1混同行列に応じて、前記モデル生成部30-1により生成された第1方言なまり共通語モデルを、前記モデル生成部30-0により生成された標準的共通語モデルの中にマージして第1一時マージモデルを生成するものである。

モデルマージ部50-2~50-( $n - 1$ )は、それぞれ対応する前記混同行列生成部40-2~40-( $n - 1$ )により生成された第2~第( $n - 1$ )混同行列に応じて、前記モデル生成部30-2~30-( $n - 1$ )により生成された第2~第( $n - 1$ )方言なまり共通語モデルを、その直前のモデルマージ部により生成された一時マージモデルの中にマージして第2~第( $n - 1$ )一時マージモデルをそれぞれ生成するものである。

モデルマージ部50-nは、前記混同行列生成部40-nにより生成された第n混同行列に応じて、前記モデル生成部30-nにより生成された第n方言なまり共通語モデルを、その直前のモデルマージ部50-( $n - 1$ )により生成された第( $n - 1$ )一時マージモデルの中にマージして最終の認識モデルを生成するものである。

#### 【0011】

制御ユニット200は、前述した本発明のモデリング方法に従って動作するよう前記モデル生成ユニット100を制御する。

#### 【0012】

10

20

30

40

50

図2において、トレーニングDB10-0、ディベロップメントDB10-1~10-nは別々のブロックとして示されているが、標準的共通語のトレーニングデータ及び第1~第n種の方言なまり共通語のディベロップメントデータを記憶する単一又は複数のデータベースとして構成されてもよい。また、図2においてモデル生成部30-0~30-nは別々のブロックとして示されているが、これらを単一又は複数のモデル生成部として、制御ユニット200からの制御に基づきこの単一又は複数のモデル生成部を時間分割で利用してもよい。また、図2において混同行列生成部40-1~40-nは別々のブロックとして示されているが、制御ユニット200からの制御に基づき、単一又は複数の混同行列生成部を時間分割で利用してもよい。また、図2においてモデルマージ部50-1~50-nは別々のブロックとして示されているが、制御ユニット200からの制御に基づき、単一又は複数のモデルマージ部を時間分割で利用してもよい。

10

## 【0013】

以下はn=2、即ち2種類の方言なまり共通語に対応できる認識モデルのモデリング方法を具体的に説明する。本モデリング方法は下記の工程を含む：

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第1方言なまり共通語モデルを生成し、第2種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第2方言なまり共通語モデルを生成する；

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を取得し、当該第1混同行列に応じて前記第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして一時マージモデルを得る；

20

(3) 前記一時マージモデルを用いて第2種の方言なまり共通語のディベロップメントデータを認識することにより第2混同行列を取得し、当該第2混同行列に応じて前記第2方言なまり共通語モデルを前記一時マージモデルの中にマージして認識モデルを得る。

## 【0014】

上記方法の工程(2)と(3)に記載のマージの方法は下記の通りである：

xで被認識音声の観測特徴ベクター、sで標準的共通語モデルにおける隠れマルコフ状態、d<sub>1</sub>で第1方言なまり共通語モデルにおける隠れマルコフ状態、を表す場合、下記の数式で与えられる一時マージモデルにおける確率密度関数は

30

$$p'(x|s) = \alpha_1 p(x|s) + (1 - \alpha_1) p(x|d_1) p(d_1|s) \quad (1)$$

である。

その中で、 $\alpha_1$ は線形補間係数であり、 $0 < \alpha_1 < 1$ を満たす。

## 【0015】

認識モデルの確率密度関数は

## 【数2】

$$p''(x|s) = \sum_{k=1}^K w_k^{(sc)} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)} N_{pq}^{(dc2)}(\cdot) \quad 40$$

であり、その中で、 $w_k^{(sc)}$ は標準的共通語モデルにおいて隠れマルコフ状態が占める重み、 $w_{mn}^{(dc1)}$ と $w_{pq}^{(dc2)}$ はそれぞれ第1方言なまり共通語モデル、第2方言なまり共通語モデルにおいて隠れマルコフ状態が占める重みを示す。Kは標準的共通語モデルの隠れマルコフ状態sの混合正規分布の混合数である。 $N_k^{(sc)}(\cdot)$ は標準的共通語モデルの隠れマルコフ状態sの混合正規分布の要素である。Mは前記第1方言なまり共通語モデルの隠れマルコフ状態d<sub>1</sub>と標準的共通語モデルの隠れマルコフ状態sの間での発音バリエーションの数である。Nは前記第1方言なまり共通語モデルの隠れマルコフ状態d<sub>1</sub>の混合正規分布の混合数である。 $N_{mn}^{(dc1)}(\cdot)$ は前記第1方言なまり共通語モデルの隠れマルコフ状態d<sub>1</sub>の混合正規分布の要素である。P

50



は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音バリエーションの数である。Qは前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数である。  $N_{p,q}^{(dc2)}(\cdot)$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素である。

【0016】

本発明の方法は、反復的な方法によって、各種の方言なまりのデータにより作られたモデルを標準的共通語モデルの中にマージするものであり、その基本的なフローは図1のとおりである。図1において二つの方言なまり共通語モデルと標準的共通語モデルとのマージを例とした場合、一時マージモデルにおける確率密度関数は

$$p'(x|s) = \lambda_1 p(x|s) + (1 - \lambda_1) p(x|d_1) p(d_1|s) \quad (1)$$

と記述できる。

【0017】

その中で、 $x$ で被認識音声の観測特徴ベクター、 $s$ で標準的共通語モデルにおける隠れマルコフ状態、 $d_1$ で第1方言なまり共通語モデルにおける隠れマルコフ状態を表す。

$\lambda_1$ は  $0 < \lambda_1 < 1$  を満たす線形補間係数であり、標準的共通語モデルが一時マージモデルにおいて占める重みを表す。実際においては最適な  $\lambda_1$  は実験を通して決められる。また、 $p(d_1|s)$  は標準的共通語モデルにおける隠れマルコフ状態  $s$  に対応する第1方言なまり共通語モデルにおける隠れマルコフ状態  $d_1$  の出力確率であり、標準的共通語に対する第1種の方言の発音の変化を示す。同じ道理で、最終マージモデルの確率密度関数は

【数3】

$$\begin{aligned} p''(x|s) &= \lambda_2 p'(x|s) + (1 - \lambda_2) p(x|d_2) p'(d_2|s) \\ &= \lambda_2 \lambda_1 p(x|s) + \lambda_2 (1 - \lambda_1) p(x|d_1) p(d_1|s) + (1 - \lambda_2) p(x|d_2) p'(d_2|s) \\ &= \lambda_2 \lambda_1 \sum_{k=1}^K w_k^{(sc)} N_k^{(sc)}(\cdot) + \lambda_2 (1 - \lambda_1) \sum_{m=1}^M P(d_{1m}|s) \cdot \sum_{n=1}^N w_{mn}^{(dc1)} N_{mn}^{(dc1)}(\cdot) + \\ &\quad (1 - \lambda_2) \sum_{p=1}^P P(d_{2p}|s) \cdot \sum_{q=1}^Q w_{pq}^{(dc2)} N_{pq}^{(dc2)}(\cdot) \\ &= \sum_{k=1}^K \lambda_2 \lambda_1 w_k^{(sc)} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N \lambda_2 (1 - \lambda_1) \cdot P(d_{1m}|s) \cdot w_{mn}^{(dc1)} N_{mn}^{(dc1)}(\cdot) + \\ &\quad \sum_{p=1}^P \sum_{q=1}^Q (1 - \lambda_2) \cdot P(d_{2p}|s) \cdot w_{pq}^{(dc2)} N_{pq}^{(dc2)}(\cdot) \\ &= \sum_{k=1}^K w_k^{(sc)'} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)'} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)'} N_{pq}^{(dc2)}(\cdot) \end{aligned}$$

と記述できる。

その中で、 $d_2$ で第2方言なまり共通語モデルにおける隠れマルコフ状態を表す。 $\lambda_2$ は  $0 < \lambda_2 < 1$  を満たす線形補間係数であり、前記一時マージモデルが最終マージモデルにおいて占める重みを表す。実際においては最適な  $\lambda_2$  は実験を通して決められる。Kは標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の混合数である。  $N_k^{(sc)}$

10

20

30

40

50

( $\cdot$ ) は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の要素である。M は前記第 1 方言なまり共通語モデルの隠れマルコフ状態  $d_1$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音バリエーションの数である。N は前記第 1 方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の混合数である。N<sub>m n</sub> ( $d_1 c_1$ ) ( $\cdot$ ) は前記第 1 方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の要素である。P ( $d_1 m | s$ ) は発音変化モデルの確率を表す。P は前記第 2 方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音バリエーションの数である。Q は前記第 2 方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数である。N ( $d_2 c_2$ ) ( $\cdot$ ) は前記第 2 方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素である。P ( $d_2 p | s$ ) は発音変化モデルの確率を表す。

10

【0018】

上記数式の最後の一行からわかるように、最終マージモデルは実際には標準的共通語モデル、第 1 方言なまり共通語モデル及び第 2 方言なまり共通語モデルの加重和により構成されるものである。w<sub>k</sub> ( $s c$ )、w<sub>m n</sub> ( $d_1 c_1$ ) 及び w<sub>p q</sub> ( $d_2 c_2$ ) は上記数式における三つのモデルそれぞれの混合重みを表す。混同行列 P ( $d_1 m | s$ ) と P ( $d_2 p | s$ )、及び重み係数  $\alpha_1$  と  $\alpha_2$  は既に知られているため、この三つのモデルそれぞれの混合正規分布の重みは簡単に確定することができる。

【0019】

以下は本発明の実施例を説明する：

20

【表 1】

実験データの説明

データセット	データベース	説明
標準的共通語トレーニングセット	標準的共通語トレーニングデータ	120人、200の長い文/人
標準的共通語テストセット	標準的共通語テストデータ	12人、100のコマンド/人
四川共通語ディベロップメントセット	四川なまり共通語ディベロップメントデータ	20人、50の長い文/人
四川共通語テストセット	四川なまり共通語テストデータ	15人、75のコマンド/人
閩南共通語ディベロップメントセット	閩南なまり共通語ディベロップメントデータ	20人、50の長い文/人
閩南共通語テストセット	閩南なまり共通語テストデータ	15人、75のコマンド/人

30

40

【0020】

表から明らかなように、データは、標準的共通語、四川なまり共通語、閩南なまり共通語に分けられ、更にトレーニング用又はディベロップメント用と、テスト用の二部分に分けられている。

【0021】

ベースライン：

【表 2】

## テスト基準システムの説明

テストセット 認識モデル	ワード誤り率 (WER)		
	標準的共通語テスト セット	閩南なまり共通語テ ストセット	四川なまり共通語テ ストセット
混合トレーニング認 識モデル	8.5%	21.7%	21.1%

【 0 0 2 2 】

ベースラインにおいては混合トレーニング認識モデルが用いられ、これは全部の三種類のデータを合わせてトレーニングすることにより得たものである。

【 0 0 2 3 】

実験の結果：

【表 3】

## 実験の結果

テストセット 認識モデル	ワード誤り率 (WER)		
	標準的共通語テス トセット	閩南なまり共通語テ ストセット	四川なまり共通語テ ストセット
本発明による認識モデ ル	6.3%	11.2%	15.0%

【 0 0 2 4 】

上記から明らかのように、本計算方法でトレーニングしたモデルを利用すると、二つの方言に対する認識率も明らかに上がった。同時に、標準的共通語に対する認識率も相当に改善された。このことから、本方法は実行可能且つ有効な方法であることがわかる。

【 0 0 2 5 】

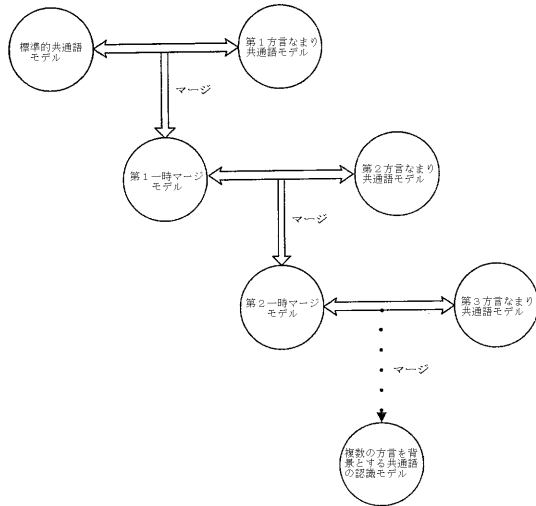
また、本発明の方法によれば、方言がいくらあっても、反復的な方法で方言なまり共通語モデルを一つ一つ標準的共通語モデルの中にマージすることによって、最終の認識モデルを得ることができる。

10

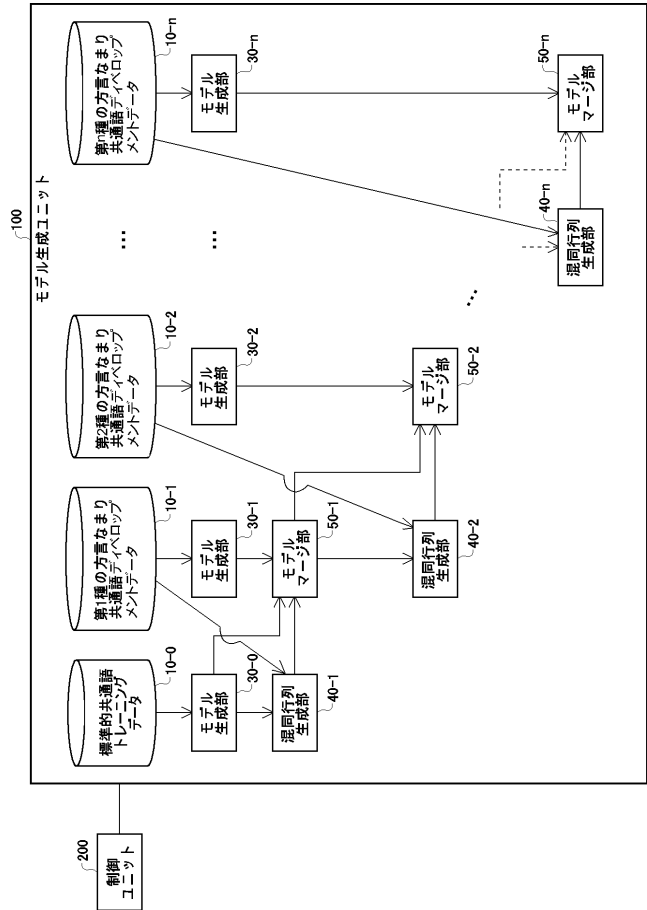
20

30

【図1】



【図2】



【手続補正書】

【提出日】平成22年1月7日(2010.1.7)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

複数の方言を背景とする共通語音声認識のモデリング方法であって、

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第1方言なまり共通語モデルを生成し、第2種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第2方言なまり共通語モデルを生成する工程と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて前記第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして一時マージモデルを得る工程と、

(3) 前記一時マージモデルを用いて第2種の方言なまり共通語のディベロップメントデータを認識することにより第2混同行列を生成し、当該第2混同行列に応じて前記第2方言なまり共通語モデルを前記一時マージモデルの中にマージして認識モデルを得る工程と、

を含むことを特徴とする複数の方言を背景とする共通語音声認識のモデリング方法。

【請求項2】

$x$  で被認識音声の観測特徴ベクター、 $s$  で前記標準的共通語モデルにおける隠れマルコフ状態、 $d_1$  で前記第1方言なまり共通語モデルにおける隠れマルコフ状態、 $d_2$  で前記第2方言なまり共通語モデルにおける隠れマルコフ状態、を表す場合、下記の数式で与えられる前記一時マージモデルにおける確率密度関数は

$$p'(x|s) = \alpha_1 p(x|s) + (1 - \alpha_1) p(x|d_1) p(d_1|s)$$

であり、

その中で、 $\alpha_1$  は線形補間係数であり、 $0 < \alpha_1 < 1$  を満たし、

前記認識モデルの確率密度関数は

【数4】

$$p''(x|s) = \sum_{k=1}^K w_k^{(sc)} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)} N_{pq}^{(dc2)}(\cdot)$$

であり、その中で、 $w_k^{(sc)}$  は前記標準的共通語モデルにおける隠れマルコフ状態が占める重み、 $w_{mn}^{(dc1)}$  と  $w_{pq}^{(dc2)}$  はそれぞれ前記第1方言なまり共通語モデル、前記第2方言なまり共通語モデルにおける隠れマルコフ状態が占める重み、 $K$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の混合数、 $N_k^{(sc)}(\cdot)$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の要素、 $M$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音パリエーションの数、 $N$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の混合数、 $N_{mn}^{(dc1)}(\cdot)$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の要素、 $P$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音パリエーションの数、 $Q$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数、 $N_{pq}^{(dc2)}(\cdot)$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素、を示すことを特徴とする請求項1に記載のモデリング方法。

【請求項3】

複数の方言を背景とする共通語音声認識のモデリングプログラムであって、コンピュータに

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第1方言なまり共通語モデルを生成し、第2種の方言なまり共通語のディベロップメントデータに基づいてモノフォンによる第2方言なまり共通語モデルを生成する機能と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて前記第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして一時マージモデルを得る機能と、

(3) 前記一時マージモデルを用いて第2種の方言なまり共通語のディベロップメントデータを認識することにより第2混同行列を生成し、当該第2混同行列に応じて前記第2方言なまり共通語モデルを前記一時マージモデルの中にマージして認識モデルを得る機能と、

を実行させることを特徴とするコンピュータプログラム。

【請求項4】

$n$  種 ( $n$  は2以上の自然数) の方言を背景とする共通語音声認識のモデリング方法であって、

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1～ $n$ 種の方言なまり共通語のそれぞれに対し、そのディベロップメントデータに基づいてモノフォンによる第1～第 $n$ 方言なまり共通語モデルを生成する工程と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして第1一時マージモデルを得る工程と、

(3) 第 $(i - 1)$ 一時マージモデル( $i$ は $2 < i < n$ を満たす自然数)を用いて第 $i$ 種の方言なまり共通語のディベロップメントデータを認識することにより第 $i$ 混同行列を生成し、当該第 $i$ 混同行列に応じて第 $i$ 方言なまり共通語モデルを前記第 $(i - 1)$ 一時マージモデルの中にマージする動作を、 $i = 2$ から $i = n$ まで順に繰り返すことにより、認識モデルを得る工程と、

を含むことを特徴とする複数の方言を背景とする共通語音声認識のモデリング方法。

【請求項5】

$n$ 種( $n$ は2以上の自然数)の方言を背景とする共通語音声認識のモデリングプログラムであって、コンピュータに

(1) 標準的共通語のトレーニングデータに基づいてトライフォンによる標準的共通語モデルを生成し、第1~ $n$ 種の方言なまり共通語のそれぞれに対し、そのディベロップメントデータに基づいてモノフォンによる第1~第 $n$ 方言なまり共通語モデルを生成する機能と、

(2) 前記標準的共通語モデルを用いて第1種の方言なまり共通語のディベロップメントデータを認識することにより第1混同行列を生成し、当該第1混同行列に応じて第1方言なまり共通語モデルを前記標準的共通語モデルの中にマージして第1一時マージモデルを得る機能と、

(3) 第 $(i - 1)$ 一時マージモデル( $i$ は $2 < i < n$ を満たす自然数)を用いて第 $i$ 種の方言なまり共通語のディベロップメントデータを認識することにより第 $i$ 混同行列を生成し、当該第 $i$ 混同行列に応じて第 $i$ 方言なまり共通語モデルを前記第 $(i - 1)$ 一時マージモデルの中にマージする動作を、 $i = 2$ から $i = n$ まで順に繰り返すことにより、認識モデルを得る機能と、

を実行させることを特徴とするコンピュータプログラム。

【請求項6】

請求項3または5に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。

【請求項7】

複数の方言を背景とする共通語音声認識のモデリングシステムであって、モデル生成ユニットと、当該モデル生成ユニット全体の動作を制御する制御ユニットとを備え、

前記モデル生成ユニットは、

標準的共通語のトレーニングデータが記憶されている標準的共通語トレーニングデータベースと、

第1、第2種の方言なまり共通語のディベロップメントデータがそれぞれ記憶されている第1、第2ディベロップメントデータベースと、

前記標準的共通語トレーニングデータベースに記憶されている標準的共通語のトレーニングデータに基づいて、トライフォンによる標準的共通語モデルを生成するための標準的共通語モデル生成部と、

前記第1、第2ディベロップメントデータベースにそれぞれ記憶されている第1、第2種の方言なまり共通語のディベロップメントデータに基づいて、モノフォンによる第1、第2方言なまり共通語モデルを生成するための第1、第2モデル生成部と、

前記標準的共通語モデル生成部により生成された標準的共通語モデルを用いて、前記第1ディベロップメントデータベースに記憶されている第1種の方言なまり共通語のディベロップメントデータを認識することにより、第1混同行列を生成するための第1混同行列生成部と、

前記第1混同行列生成部により生成された第1混同行列に応じて、前記第1モデル生成部により生成された第1方言なまり共通語モデルを、前記標準的共通語モデル生成部によ

り生成された標準的共通語モデルの中にマージして一時マージモデルを生成するための第1モデルマージ部と、

前記第1モデルマージ部により生成された一時マージモデルを用いて、前記第2ディベロップメントデータベースに記憶されている第2種の方言なまり共通語のディベロップメントデータを認識することにより、第2混同行列を生成するための第2混同行列生成部と、

前記第2混同行列生成部により生成された第2混同行列に応じて、前記第2モデル生成部により生成された第2方言なまり共通語モデルを、前記第1モデルマージ部により生成された一時マージモデルの中にマージして認識モデルを生成するための第2モデルマージ部と

を備えることを特徴とする複数の方言を背景とする共通語音声認識のモデリングシステム。

【請求項8】

$x$  で被認識音声の観測特徴ベクター、 $s$  で前記標準的共通語モデルにおける隠れマルコフ状態、 $d_1$  で前記第1方言なまり共通語モデルにおける隠れマルコフ状態、 $d_2$  で前記第2方言なまり共通語モデルにおける隠れマルコフ状態、を表す場合、下記の数式で与えられる前記一時マージモデルにおける確率密度関数は

$$p'(x|s) = \alpha_1 p(x|s) + (1 - \alpha_1) p(x|d_1) p(d_1|s)$$

であり、

その中で、 $\alpha_1$  は線形補間係数であり、 $0 < \alpha_1 < 1$  を満たし、

前記認識モデルの確率密度関数は

【数5】

$$p''(x|s) = \sum_{k=1}^K w_k^{(sc)'} N_k^{(sc)}(\cdot) + \sum_{m=1}^M \sum_{n=1}^N w_{mn}^{(dc1)'} N_{mn}^{(dc1)}(\cdot) + \sum_{p=1}^P \sum_{q=1}^Q w_{pq}^{(dc2)'} N_{pq}^{(dc2)}(\cdot)$$

であり、その中で、 $w_k^{(sc)'}$  は前記標準的共通語モデルにおける隠れマルコフ状態が占める重み、 $w_{mn}^{(dc1)'}$  と  $w_{pq}^{(dc2)'}$  はそれぞれ前記第1方言なまり共通語モデル、前記第2方言なまり共通語モデルにおける隠れマルコフ状態が占める重み、 $K$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の混合数、 $N_k^{(sc)}(\cdot)$  は標準的共通語モデルの隠れマルコフ状態  $s$  の混合正規分布の要素、 $M$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音パリエーションの数、 $N$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の混合数、 $N_{mn}^{(dc1)}(\cdot)$  は前記第1方言なまり共通語モデルの隠れマルコフ状態  $d_1$  の混合正規分布の要素、 $P$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  と標準的共通語モデルの隠れマルコフ状態  $s$  の間での発音パリエーションの数、 $Q$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の混合数、 $N_{pq}^{(dc2)}(\cdot)$  は前記第2方言なまり共通語モデルの隠れマルコフ状態  $d_2$  の混合正規分布の要素、を示すことを特徴とする請求項4に記載のモデリングシステム。

【請求項9】

請求項7又は8に記載の第1と第2モデル生成部、第1と第2混同行列生成部、第1と第2モデルマージ部のうち少なくとも一組は、単一の構成として時間分割で利用されることを特徴とする複数の方言を背景とする共通語音声認識のモデリングシステム。

---

フロントページの続き

- (74)代理人 100134256  
弁理士 青木 武司
- (72)発明者 鄭 方  
中国北京市海淀区清華園清華大学内
- (72)発明者 肖 熙  
中国北京市海淀区清華園清華大学内
- (72)発明者 劉 林泉  
中国北京市海淀区清華園清華大学内
- (72)発明者 遊 展  
中国北京市海淀区清華園清華大学内
- (72)発明者 曹 文暁  
中国北京市海淀区清華園清華大学内
- (72)発明者 赤羽 誠  
東京都港区南青山2丁目6番21号 株式会社ソニー・コンピュータエンタテインメント内
- (72)発明者 陳 如新  
アメリカ合衆国カリフォルニア州フォスター・シティー、セカンド・フロアー、イースト・ヒルズ  
デイル・ブルバード 919 ソニー・コンピュータエンタテインメントアメリカ内
- (72)発明者 高橋 良和  
東京都港区南青山2丁目6番21号 株式会社ソニー・コンピュータエンタテインメント内
- Fターム(参考) 5D015 GG01