(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4951664号
(P4951664)

(45)発行日　平成24年6月13日(2012.6.13)　(24)登録日　平成24年3月16日(2012.3.16)

(51)Int.Cl.　F I
G10L 15/06 (2006.01)　G10L 15/06 300D
G10L 15/14 (2006.01)　G10L 15/14 200C
G10L 15/14 200A

請求項の数 9　(全 12 頁)

(21)出願番号　特願2009-248013 (P2009-248013)
(22)出願日　平成21年10月28日 (2009.10.28)
(65)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)
審査請求日　平成22年1月7日 (2010.1.7)
(31)優先権主張番号　200810225354.0
(32)優先日　平成20年10月31日 (2008.10.31)
(33)優先権主張国　中国 (CN)

(73)特許権者　502192546
清華大学
中華人民共和国北京市海淀区清華大学　郵編１００００８４
(73)特許権者　310021766
株式会社ソニー・コンピュータエンタテインメント
東京都港区港南１丁目７番１号
(74)代理人　100105924
弁理士　森下　賢樹
(74)代理人　100109047
弁理士　村田　雄祐
(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)【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

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

であり、

その中で、λ_1は線形補間係数であり、$0<\lambda_1<1$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、$N_{pq}^{(dc2)}(\cdot)$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、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

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

であり、

その中で、λ_1は線形補間係数であり、$0<\lambda_1<1$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、$N_{pq}^{(dc2)}(\cdot)$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。

【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。従来の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。

【発明の概要】

【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；

（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：

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

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

であり、

その中で、λ_1は線形補間係数であり、$0<\lambda_1<1$を満たし、

前記認識モデルの確率密度関数は

【数１】

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

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。$N_{mn}^{(dc1)}(\cdot)$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。$N_{pq}^{(dc2)}(\cdot)$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、

（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て、

（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】
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。
【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；

（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_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

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

である。

その中で、λ_1は線形補間係数であり、$0<\lambda_1<1$を満たす。

【００１５】

認識モデルの確率密度関数は

【数２】

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

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_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。$N_{mn}^{(dc1)}(\cdot)$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。$N_{pq}^{(dc2)}(\cdot)$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。

【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。λ_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λ_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。λ$_2$は０＜λ$_2$＜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λ$_2$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$$^{（ｓｃ）}$（・）は標準的共通語モデルの隠れマルコフ状態ｓの混合正規分布の要素である。Ｍは前記第１方言なまり共通語モデルの隠れマルコフ状態ｄ$_1$と標準的共通語モデルの隠れマルコフ状態ｓの間での発音バリエーションの数である。Ｎは前記第１方言なまり共通語モデルの隠れマルコフ状態ｄ$_1$の混合正規分布の混合数である。Ｎ$_{mn}$$^{（ｄｃ１）}$（・）は前記第１方言なまり共通語モデルの隠れマルコフ状態ｄ$_1$の混合正規分布の要素である。Ｐ（ｄ$_{1m}$｜ｓ）は発音変化モデルの確率を表す。Ｐは前記第２方言なまり共通語モデルの隠れマルコフ状態ｄ$_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$の混合正規分布の混合数である。Ｎ$^{（ｄｃ２））}$（・）は前記第２方言なまり共通語モデルの隠れマルコフ状態ｄ$_2$の混合正規分布の要素である。Ｐ（ｄ$_{2p}$｜ｓ）は発音変化モデルの確率を表す。

【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$_k$$^{（ｓｃ）'}$、ｗ$_{mn}$$^{（ｄｃ１）'}$及びｗ$_{pq}$$^{（ｄｃ２）'}$は上記数式における三つのモデルそれぞれの混合重みを表す。混同行列Ｐ（ｄ$_{1m}$｜ｓ）とＰ（ｄ$_{2p}$｜ｓ）、及び重み係数λ$_1$とλ$_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：

【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

【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、▲ミン▼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%

【００２２】

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。

【００２３】

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%

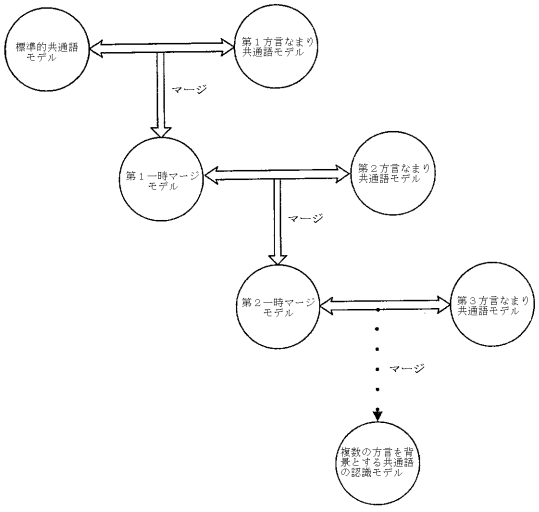
【００２４】

上記から明らかなように、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。

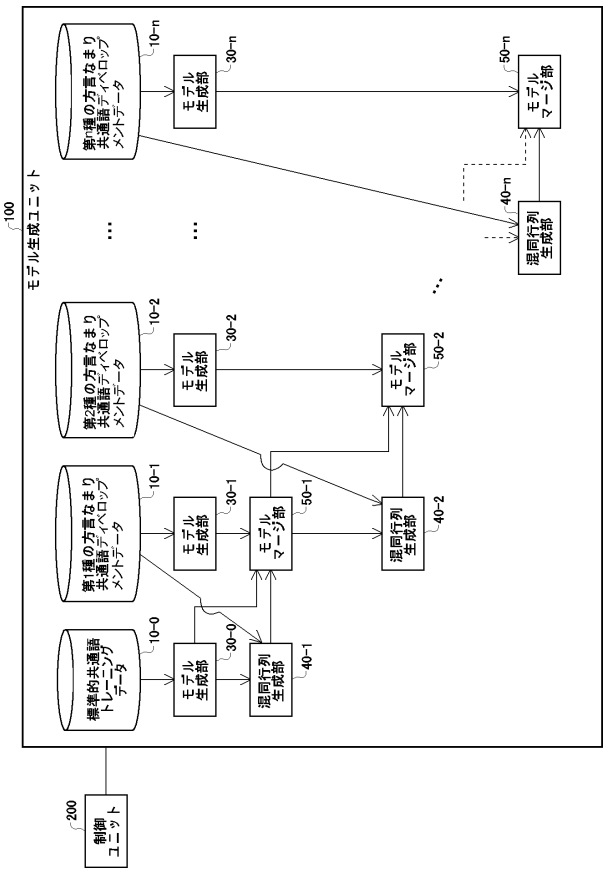
【００２５】

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。

【図１】



【図２】



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・ヒルスデイル・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

審査官 山下 剛史

(56)参考文献 特開平２－１７３６９９（ＪＰ，Ａ）
特開昭５８－７２９９６（ＪＰ，Ａ）
特開平４－３２６４００（ＪＰ，Ａ）
特表２００６－５２６１７４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １５／００－１７／００