

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6121187号
(P6121187)

(45) 発行日 平成29年4月26日 (2017. 4. 26)

(24) 登録日 平成29年4月7日 (2017. 4. 7)

(51) Int. Cl.

F I

G 1 O L 15/065 (2013. 01)

G 1 O L 15/065 A

G 1 O L 15/14 (2006. 01)

G 1 O L 15/14 2 O O C

G 1 O L 15/06 (2013. 01)

G 1 O L 15/06 3 O O G

G 1 O L 15/06 3 O O Y

請求項の数 5 (全 25 頁)

(21) 出願番号 特願2013-25865 (P2013-25865)
 (22) 出願日 平成25年2月13日 (2013. 2. 13)
 (65) 公開番号 特開2014-153680 (P2014-153680A)
 (43) 公開日 平成26年8月25日 (2014. 8. 25)
 審査請求日 平成27年1月14日 (2015. 1. 14)

(73) 特許権者 000004226
 日本電信電話株式会社
 東京都千代田区大手町一丁目5番1号
 (74) 代理人 100121706
 弁理士 中尾 直樹
 (74) 代理人 100128705
 弁理士 中村 幸雄
 (74) 代理人 100147773
 弁理士 義村 宗洋
 (72) 発明者 デルクロア マーク
 東京都千代田区大手町二丁目3番1号 日
 本電信電話株式会社内
 (72) 発明者 小川 厚徳
 東京都千代田区大手町二丁目3番1号 日
 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 音響モデル補正パラメータ推定装置、その方法及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

音響モデルには混合ガウス分布モデルが含まれるものとし、音響モデルパラメータには前記混合ガウス分布モデルに含まれるガウス分布の平均ベクトルが含まれるものとし、学習用音声データの特徴量及び前記学習用音声データに対する正解シンボル系列から、前記平均ベクトルを補正するための平均補正パラメータを求める音響モデル補正パラメータ推定装置であって、

予め求められた前記音響モデル及び言語モデルが記憶される記憶部と、

前記記憶部に記憶された音響モデルの平均ベクトルを、平均補正パラメータを用いて補正する音響モデル補正部と、

補正した前記平均ベクトルを含む音響モデルと前記言語モデルとに基づき、前記学習用音声データの前記特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、前記正解シンボル系列との相違度を求めるエラーカウント計算部と、

前記言語モデルによって得られる前記対立候補シンボル系列の言語確率、前記学習用音声データの前記特徴量と前記対立候補シンボル系列に基づき前記音響モデルによって得られる音響スコア及び前記相違度に基づき、前記平均補正パラメータで識別学習基準の目的関数を微分したときの微分値を求める補正パラメータ微分値計算部と、

前記微分値に応じて前記平均補正パラメータを変更することで、前記平均補正パラメータを更新する補正パラメータ更新部と、を含む、

10

20

音響モデル補正パラメータ推定装置。

【請求項 2】

請求項 1 記載の音響モデル補正パラメータ推定装置であって、

前記音響モデルパラメータには、さらに、前記混合ガウス分布モデルに含まれるガウス分布の共分散行列が含まれるものとし、

前記音響モデル補正部は、さらに、前記混合ガウス分布モデルに含まれるガウス分布の共分散行列を、分散補正パラメータを用いて補正し、

前記エラーカウント計算部は、補正した前記平均ベクトルと補正した前記共分散行列とを含む音響モデルと前記言語モデルとに基づき、前記学習用音声データの前記特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、前記正解シンボル系列との相違度を求め、

前記補正パラメータ微分値計算部は、さらに、前記言語モデルによって得られる前記対立候補シンボル系列の言語確率、前記学習用音声データの前記特徴量と前記対立候補シンボル系列に基づき前記音響モデルによって得られる音響スコア及び前記相違度に基づき、前記分散補正パラメータで識別学習基準の目的関数を微分したときの微分値を求め、

補正パラメータ更新部は、さらに、前記分散補正パラメータに対する前記微分値に応じて前記分散補正パラメータを変更することで、前記分散補正パラメータを更新する、

音響モデル補正パラメータ推定装置。

【請求項 3】

音響モデルには混合ガウス分布モデルが含まれるものとし、音響モデルパラメータには前記混合ガウス分布モデルに含まれるガウス分布の平均ベクトルが含まれるものとし、学習用音声データの特徴量及び前記学習用音声データに対する正解シンボル系列から、前記平均ベクトルを補正するための平均補正パラメータを求める音響モデル補正パラメータ推定方法であって、

記憶部には予め求められた前記音響モデル及び言語モデルが記憶され、

前記記憶部に記憶された音響モデルの平均ベクトルを、平均補正パラメータを用いて補正する音響モデル補正ステップと、

補正した前記平均ベクトルを含む音響モデルと前記言語モデルとに基づき、前記学習用音声データの前記特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、前記正解シンボル系列との相違度を求めるエラーカウント計算ステップと、

前記言語モデルによって得られる前記対立候補シンボル系列の言語確率、前記学習用音声データの前記特徴量と前記対立候補シンボル系列に基づき前記音響モデルによって得られる音響スコア及び前記相違度に基づき、前記平均補正パラメータで識別学習基準の目的関数を微分したときの微分値を求める補正パラメータ微分値計算ステップと、

前記微分値に応じて前記平均補正パラメータを変更することで、前記平均補正パラメータを更新する補正パラメータ更新ステップと、を含む、

音響モデル補正パラメータ推定方法。

【請求項 4】

請求項 3 記載の音響モデル補正パラメータ推定方法であって、

前記音響モデルパラメータには、さらに、前記混合ガウス分布モデルに含まれるガウス分布の共分散行列が含まれるものとし、

前記音響モデル補正ステップにおいて、さらに、前記混合ガウス分布モデルに含まれるガウス分布の共分散行列を、分散補正パラメータを用いて補正し、

前記エラーカウント計算ステップにおいて、補正した前記平均ベクトルと補正した前記共分散行列とを含む音響モデルと前記言語モデルとに基づき、前記学習用音声データの前記特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、前記正解シンボル系列との相違度を求め、

前記補正パラメータ微分値計算ステップにおいて、さらに、前記言語モデルによって得られる前記対立候補シンボル系列の言語確率、前記学習用音声データの前記特徴量と前記

10

20

30

40

50

対立候補シンボル系列に基づき前記音響モデルによって得られる音響スコア及び前記相違度に基づき、前記分散補正パラメータで識別学習基準の目的関数を微分したときの微分値を求め、

補正パラメータ更新ステップにおいて、さらに、前記分散補正パラメータに対する前記微分値に応じて前記分散補正パラメータを変更することで、前記分散補正パラメータを更新する、

音響モデル補正パラメータ推定方法。

【請求項 5】

請求項 1 もしくは請求項 2 記載の音響モデル補正パラメータ推定装置として、コンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力データから特徴量を抽出し、その特徴量を用いて入力データを予め定義されたクラスタに分類するパターン認識において、クラスタ分類精度を向上させるためのクラスタ分類モデルパラメータ補正技術及び特徴量補正技術に関する。例えば、入力音声から特徴量を抽出し、その特徴量を用いて入力音声を単語列に変換する音声認識における、音響モデル補正パラメータ推定装置、特徴量補正パラメータ推定装置、それらの方法及びプログラムに関する。

【背景技術】

【0002】

音声認識装置が実際に置かれる環境は様々である。このため音響モデルを学習するための音声データと実際に入力される音声の特徴は一致しないことが多い。不一致の原因は、周囲の雑音環境、話者の多様性、等であり、これらが音声認識精度を劣化させる。このため、これら周囲雑音環境や話者の多様性に対して頑健（ロバスト）な音声認識技術が求められている。ロバストな音声認識技術として、入力音声と音響モデルとが適合しやすくなるように、入力音声から抽出した特徴ベクトルや音響モデルを補正する技術が知られている。

【0003】

入力音声から抽出した特徴ベクトルを補正することでロバストな音声認識を実現する技術として、非特許文献 1 が知られている。この技術では、dMMI（differentiated Maximum Mutual Information）という基準に基づいて特徴ベクトルを補正するための補正パラメータを学習する。また、非特許文献 2 記載されているように、周囲雑音環境に対して、雑音下音声の雑音抑圧（音声強調）処理を行う場合に、dMMI 基準に基づく音響モデルの分散パラメータの補正パラメータを推定する技術がある。

【0004】

また、音響モデルを補正する技術として、音響モデルのパラメータを、線形回帰を用いて補正する線形回帰音響モデル適応技術（非特許文献 3、4）が知られている。

【先行技術文献】

【非特許文献】

【0005】

【非特許文献 1】デルクロア・マーク，小川厚徳，渡部晋治，中谷智広，中村篤，「dMMI 識別基準による特徴量変換の識別学習」，日本音響学会春季研究発表会，March 2012, pp. 121-122

【非特許文献 2】デルクロア・マーク，小川厚徳，渡部晋治，中谷智広，中村篤，「dMMI 識別基準による教師なし動的分散適応」，日本音響学会秋季研究発表会，September 2012, pp. 131-132

【非特許文献 3】Leggetter C. J. and Woodland P. C., "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models", Computer Speech & Language, 1995, vol. 9, no. 2, pp. 171-185

10

20

30

40

50

【非特許文献4】L.F. Uebel and P.C. Woodland, “Discriminative linear transforms for speaker adaptation”, in Proc. ISCA Tutorial and Research Workshop (ITRW) on Adaptation Methods for Speech Recognition, 2001, pp. 61-64.

【発明の概要】

【発明が解決しようとする課題】

【0006】

非特許文献1のようなdMMI基準に基づく補正パラメータの学習には、大量の学習用の音声データ（以下「学習用音声データ」ともいう）とそれに対応する正解シンボルの系列（以下「正解シンボル系列」ともいう）が必要となる。そのため、学習用音声データと正解シンボル系列からなる学習データ、特に、正解シンボル系列を準備する際に多大なコストがかかる。非特許文献2のようなdMMI基準に基づく音響モデルの分散パラメータの動的な補正パラメータの適応技術は、音声強調処理が必要なため、話者の多様性などへの適用が困難であり、汎用性に欠ける。

10

【0007】

一方、線形回帰パラメータの推定を最尤基準（MLLR (Maximum likelihood linear regression)）で行う方法（非特許文献3）または識別的基準の一種である最大相互情報量（Maximum Mutual Information : MMI）基準（MMI - LR）で行う方法（非特許文献4）に基づく補正パラメータの適応技術は、少ない音声データで実行することができるという利点がある。また、正解シンボルを必要としない教師なし適応を行うことも可能であり、この場合は正解シンボルを手で準備する必要がないという利点がある。

20

【0008】

教師あり適応の場合はMMI - LRはMLLRよりも性能が良いと報告されている（非特許文献4）。しかし、MMI - LRの識別基準による音響モデルの教師なし適応方法は、適応データを音声認識した結果を正解ラベルと見做して利用するので、正解ラベル（と見做された音声認識結果）に誤りが含まれることが多い。MMI - LRのような識別基準は、正解シンボルと他の認識仮説を考慮し、直接音響モデルパラメータを最適化することによって、認識性能を大きく向上させる技術であるため、正解シンボルに誤りがある場合はうまく音響モデルパラメータを最適化できず、性能が改善しないか悪化する可能性がある。

【0009】

本発明は、正解シンボルの誤りの悪影響を弱める仕組みを導入し、正解シンボルの誤りが多い教師なし適応の場合でも、識別基準による音響モデル適応の精度の低下を防ぎ、識別基準による教師なし音響モデル適応を可能にする音響モデル補正パラメータ推定技術及び特徴量補正パラメータ推定技術を提供することを目的とする。

30

【課題を解決するための手段】

【0010】

上記の課題を解決するために、本発明の第一の態様によれば、音響モデル補正パラメータ推定装置は、音響モデルには混合ガウス分布モデルが含まれるものとし、音響モデルパラメータには混合ガウス分布モデルに含まれるガウス分布の平均ベクトルが含まれるものとし、学習用音声データの特徴量及び学習用音声データに対する正解シンボル系列から、平均ベクトルを補正するための平均補正パラメータを求める。音響モデル補正パラメータ推定装置は、予め求められた音響モデル及び言語モデルが記憶される記憶部と、記憶部に記憶された音響モデルの平均ベクトルを、平均補正パラメータを用いて補正する音響モデル補正部と、補正した平均ベクトルを含む音響モデルと言語モデルとに基づき、学習用音声データの特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、正解シンボル系列との相違度を求めるエラーカウント計算部と、言語モデルによって得られる対立候補シンボル系列の言語確率、学習用音声データの特徴量と対立候補シンボル系列に基づき音響モデルによって得られる音響スコア及び相違度に基づき、平均補正パラメータで識別学習基準の目的関数を微分したときの微分値を求める補正パラメータ微分値計算部と、微分値に応じて平均補正パラメータを変更することで、平均補

40

50

正パラメータを更新する補正パラメータ更新部と、を含む。

【 0 0 1 1 】

上記の課題を解決するために、本発明の他の態様によれば、特徴量補正パラメータ推定装置は、学習用音声データの特徴量及び学習用音声データに対する正解シンボル系列から、認識用音声データの特徴量を補正するための特徴量補正パラメータを求める。特徴量補正パラメータ推定装置は、予め求められた、ガウス分布で表現された音響モデル及び言語モデルが記憶される記憶部と、学習用音声データの特徴量を音響モデルを表現するガウス分布のクラスタ毎の特徴量補正パラメータにより補正した補正後の特徴量を求める特徴量補正部と、予め定めた粒度で、補正後の特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、正解シンボル系列との相違度を求めるエラーカウント計算部と、言語モデルによって得られる対立候補シンボル系列の言語確率、補正後の特徴量と対立候補シンボル系列に基づき音響モデルによって得られる音響スコア及び相違度に基づき、特徴量補正パラメータで識別学習基準の目的関数を微分したときの微分値を求める補正パラメータ微分値計算部と、微分値に応じて特徴量補正パラメータを変更することで、特徴量補正パラメータを更新する補正パラメータ更新部と、を含む。

10

【 0 0 1 2 】

上記の課題を解決するために、本発明の他の態様によれば、音響モデル補正パラメータ推定方法は、音響モデルには混合ガウス分布モデルが含まれるものとし、音響モデルパラメータには混合ガウス分布モデルに含まれるガウス分布の平均ベクトルが含まれるものとし、学習用音声データの特徴量及び学習用音声データに対する正解シンボル系列から、平均ベクトルを補正するための平均補正パラメータを求める。音響モデル補正パラメータ推定方法は、記憶部には予め求められた音響モデル及び言語モデルが記憶され、記憶部に記憶された音響モデルの平均ベクトルを、平均補正パラメータを用いて補正する音響モデル補正ステップと、補正した平均ベクトルを含む音響モデルと言語モデルとに基づき、学習用音声データの特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、正解シンボル系列との相違度を求めるエラーカウント計算ステップと、言語モデルによって得られる対立候補シンボル系列の言語確率、学習用音声データの特徴量と対立候補シンボル系列に基づき音響モデルによって得られる音響スコア及び相違度に基づき、平均補正パラメータで識別学習基準の目的関数を微分したときの微分値を求める補正パラメータ微分値計算ステップと、微分値に応じて平均補正パラメータを変更することで、平均補正パラメータを更新する補正パラメータ更新ステップと、を含む。

20

30

【 0 0 1 3 】

上記の課題を解決するために、本発明の他の態様によれば、特徴量補正パラメータ推定方法は、学習用音声データの特徴量及び学習用音声データに対する正解シンボル系列から、認識用音声データの特徴量を補正するための特徴量補正パラメータを求める。特徴量補正パラメータ推定方法は、記憶部には予め求められた、ガウス分布で表現された音響モデル及び言語モデルが記憶され、学習用音声データの特徴量を音響モデルを表現するガウス分布のクラスタ毎の特徴量補正パラメータにより補正した補正後の特徴量を求める特徴量補正ステップと、予め定めた粒度で、補正後の特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、正解シンボル系列との相違度を求めるエラーカウント計算ステップと、言語モデルによって得られる対立候補シンボル系列の言語確率、補正後の特徴量と対立候補シンボル系列に基づき音響モデルによって得られる音響スコア及び相違度に基づき、特徴量補正パラメータで識別学習基準の目的関数を微分したときの微分値を求める補正パラメータ微分値計算ステップと、微分値に応じて特徴量補正パラメータを変更することで、特徴量補正パラメータを更新する補正パラメータ更新ステップと、を含む。

40

【 発明の効果 】

【 0 0 1 4 】

本発明によれば、正解シンボルの誤りの悪影響を弱めることによって、従来技術よりも適切に音響モデルパラメータまたは特徴量に対する補正パラメータを求めることができる

50

という効果を奏する。

【図面の簡単な説明】

【 0 0 1 5 】

【図 1】線形回帰音響モデル適応技術を搭載した音声認識装置の機能構成例を示す図。

【図 2】線形回帰音響モデル適応技術を搭載した音声認識装置の処理フロー例を示す図。

【図 3】音響モデル補正パラメータ学習装置の機能構成例を示す図。

【図 4】音響モデル補正パラメータ学習装置の処理フロー例を示す図。

【図 5】第一実施形態に係る音響モデル補正パラメータ学習装置の構成例を示す図。

【図 6】第一実施形態に係る音響モデル補正パラメータ学習装置の処理フロー例を示す図

。

【図 7】補正した特徴量に基づき音声認識を行う音声認識装置の機能構成例を示す図。

【図 8】補正した特徴量に基づき音声認識を行う音声認識装置の処理フロー例を示す図。

【図 9】第二実施形態に係る特徴量補正パラメータ学習装置の構成例を示す図。

【図 10】第二実施形態に係る特徴量補正パラメータ学習装置の処理フロー例を示す図。

【発明を実施するための形態】

【 0 0 1 6 】

以下、本発明の実施形態について説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。以下の説明において、テキスト中で使用する記号「^」等は、本来直前の文字の真上に記載されるべきものであるが、テキスト記法の制限により、当該文字の直後に記載する。式中においてはこれらの記号は本来の位置に記述している。また、ベクトルや行列の各要素単位で行われる処理は、特に断りが無い限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

【 0 0 1 7 】

まず、第一実施形態について説明する前に、まず、音響モデル適応技術を搭載した音声認識装置について説明する。

【 0 0 1 8 】

< 音響モデル適応技術を搭載した音声認識装置 9 0 >

図 1 に線形回帰音響モデル適応技術を搭載した音声認識装置 9 0 の機能構成例、図 2 にその処理フロー例を示す。音声認識装置 9 0 は、特徴量抽出部 9 1、単語列探索部 9 2、記憶部 9 3、音響モデル補正部 9 4 から構成される。

【 0 0 1 9 】

(記憶部 9 3)

記憶部 9 3 には、予め音響モデルと言語モデルが記憶されている。音響モデルは、音声の音響的特徴をモデル化したものである。言語モデルは音素や単語といった多数のシンボル系列から構成されている。通常、音声認識用音響モデルでは、各音素を Left to right の HMM (Hidden Markov Model: 隠れマルコフモデル) で、HMM 状態の出力確率分布を GMM (Gaussian Mixture Model: 混合ガウス分布モデル) で表現する。そのため、実際に音響モデルとして記憶部 9 3 に記憶されているのは、音素などの各シンボルにおける HMM の状態遷移確率、GMM の混合重み因子、及びガウス分布の平均ベクトル μ_m 及び共分散行列 (以下「分散パラメータ」ともいう) Σ_m 等となる。ここで、M は GMM に含まれるガウス分布の総数であり、m は、GMM のガウス分布のインデックスであり、 $m = 1, 2, \dots, M$ である。これらを音響モデルパラメータと呼び、その集合を Λ とする。言語モデルは音素や単語といった多数のシンボル系列から構成されており、図中の $P(S_j)$ は言語モデルによって得られる対立候補シンボル系列 S_j の確率 (以下「言語確率」ともいう) である。なお、対立候補シンボル系列 S_j とは音声認識結果となりうるシンボル系列であり、シンボル系列とは音素や単語等からなるシンボルの系列である。

【 0 0 2 0 】

(特徴量抽出部 9 1)

特徴量抽出部 9 1 は、認識用音声データを読み込み (s 9 3)、音声の特徴量を抽出す

10

20

30

40

50

る (s 9 5) 。特徴量としては、例えば、M F C C (Mel Frequency Cepstral Coefficient) 、 Δ M F C C 、 $\Delta \Delta$ M F C C 、対数パワー、 Δ 対数パワー等があり、これらが 1 0 ~ 1 0 0 次元程度の特徴量ベクトル o を構成する。さらに、時系列の特徴量ベクトルである特徴量ベクトル系列 O を以下のように表現できる。

【 0 0 2 1 】

【数 1】

$$O = \{o_1, o_2, \dots, o_N \mid o_n \in R^D\} \quad (1)$$

【 0 0 2 2 】

ただし、 N はフレームの数、 n は 1 から N の整数、 R は実数の集合である。つまり、 O は 1 から N フレーム目までの D 次元特徴量ベクトルで表現されるデータである。例えば、分析フレーム幅は 3 0 m s 程度、分析フレームシフト幅は 1 0 m s 程度で分析が実行される。

10

【 0 0 2 3 】

(音響モデル補正部 9 4)

音響モデル補正部 9 4 は、補正前の (記憶部 9 3 に記憶された) 音響モデルパラメータ Λ を含む音響モデルと、予め学習し記憶部 9 3 に記憶しておいた音響モデル補正パラメータ $\theta^\wedge$ を読み込み (s 9 1 、 s 9 4) 、音響モデル補正パラメータ $\theta^\wedge$ を用いて、音響モデルパラメータ Λ を含む音響モデルを補正し (s 9 6) 、補正した音響モデルパラメータ $\Lambda^\wedge$ を単語列検索部 9 2 に送る。この例では、線形回帰音響モデル適応は以下の式 (2) のように、音響モデルパラメータに含まれる平均ベクトル $\mu = \{ \mu_1, \mu_2, \dots, \mu_M \}$ を補正する。

20

【 0 0 2 4 】

【数 2】

$$\begin{aligned} \hat{\mu}_m &= A\mu_m + b \\ &= W\xi_m \end{aligned} \quad (2)$$

$$W := [A \quad b]$$

$$\xi_m = [\mu_m' \quad 1]'$$

30

【 0 0 2 5 】

ただし、 $\mu^\wedge_m$ は補正後の音響モデルパラメータにおける m 番目のガウス分布の平均ベクトル、 A は平均ベクトルに対する変換行列、 b は平均ベクトルに対するバイアスベクトルである。' はベクトルまたは行列の転置を表す。今後、 A 、 b もしくは W を平均補正パラメータともいう。また、式 (2) は平均ベクトル μ_m の補正の例を示したが、分散パラメータ Σ_m についても同様の補正を行うことができる。音響モデルパラメータを補正するためのパラメータを音響モデル補正パラメータと呼び、平均補正パラメータや分散パラメータを補正するためのパラメータ (以下「分散補正パラメータ」ともいう) は音響モデル補正パラメータに含まれるものとする。この例では、音響モデル補正パラメータ θ は平均補正パラメータ A 、 b のみからなるため、平均補正パラメータ A 、 b のことを音響モデル補正パラメータ $\theta = (A, b)$ ともいう。平均補正パラメータ A 、 b を、変換行列 A とバイアスベクトル b とからなる行列として表現した場合を $W = [A \quad b]$ とし、変換行列 A とバイアスベクトル b とからなる集合として表現した場合を $\theta = (A, b)$ とする。

40

【 0 0 2 6 】

より詳細な補正を行うため、音響モデル補正パラメータはよく音響モデルのガウス分布のクラスタ毎に推定される。その場合は、補正後の平均ベクトル $\mu^\wedge_m$ は式 (3) のようになる。クラスタの作り方は、例えば非特許文献 3 のような方法がある。

【 0 0 2 7 】

【数 3】

$$\hat{\mu}_m = A_k \mu_m + b_k = W_k \xi_m \quad (3)$$

【0028】

k はガウス分布のクラスタのインデックス、 A_k 、 b_k はクラスタ k の平均補正パラメータである。クラスタ毎に平均補正パラメータを推定した場合、 $\theta_k = (A_k, b_k)$ とし、K をクラスタの総数とし、 $k = 1, 2, \dots, K$ とし、 $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ とする。また、 $W_k = [A_k \quad b_k]$ とし、 $W = (W_1, W_2, \dots, W_K)$ とする。

【0029】

(単語列探索部 92)

単語列探索部 92 は、音響モデル補正部 94 から取得した補正後の音響モデルパラメータ $\hat{\Lambda}$ に基づき、特徴量ベクトル系列 O に対する J 個の対立候補シンボル系列 S_j を生成して、対立候補シンボル系列 S_j 毎に音響スコアを算出する。ただし、 $j = 1, 2, \dots, J$ であり、J は 1 以上の整数である。さらに、単語列探索部 92 は、予め言語モデルを記憶部 93 から読み込んでおき (s 92)、この言語モデルに基づき、対立候補シンボル系列 S_j 毎に言語スコアを算出する。さらに、音響スコアと言語スコアとを統合して、J 個の対立候補シンボル系列 S_j の中から、認識用音声データに対応する文として最も確からしい (最も音響スコアと言語スコアとを統合したスコアが高い) 対立候補シンボル系列を探索し (s 97)、その対立候補シンボル系列を認識結果 (単語列) $S^{\wedge}$ として出力する (s 98)。

【0030】

< 音響モデル補正パラメータ学習装置 80 >

上記の音声認識装置 90 では、音響モデルパラメータを、線形回帰を用いて補正する。つまり、音響モデル補正部 94 で用いる平均補正パラメータ $\hat{\theta}$ は線形回帰パラメータである。

【0031】

線形回帰パラメータを学習する方法として、線形回帰パラメータの推定を最尤基準 (Maximum likelihood linear regression) で行う方法 (非特許文献 3) と、識別的基準の一種である最大相互情報量 (Maximum Mutual Information: MMI) 基準 (MMI - LR) で行う方法 (非特許文献 4) が知られている。最尤基準 (非特許文献 3) よりも識別的基準 (非特許文献 4) により推定された補正パラメータを用いる方が、最終的な音声認識精度が向上することが多い。

【0032】

以下では、非特許文献 4 の音響モデル補正パラメータ学習装置 80 の具体的な処理を、図 3 及び図 4 を用いて説明する。図 3 に音響モデル補正パラメータ学習装置 80 の機能構成例、図 4 にその処理フロー例を示す。音響モデル補正パラメータ学習装置 80 は、特徴量抽出部 81、音響モデル補正パラメータ計算部 83 及び記憶部 93 を備える。音響モデル補正パラメータ学習装置 80 は、学習用音声データとその学習用音声データに対する正解シンボル系列 S_r とからなる学習データを入力とし、音響モデル補正パラメータ $\hat{\theta}$ を出力する。

【0033】

(特徴量抽出部 81)

特徴量抽出部 81 は、学習用音声データを読み込み (s 83)、音声の特徴量ベクトル系列 O を抽出する (s 85)。特徴量抽出の具体的な処理は、上述の音声認識装置 90 の特徴量抽出部 91 と同じである。

【0034】

(音響モデル補正パラメータ計算部 83)

音響モデル補正パラメータ計算部 83 は、記憶部 93 から音響モデルと言語モデルとを読み込み (s 81、s 82)、さらに、正解シンボル系列 S_r を読み込み (s 84)、特徴量抽出部 81 で抽出した音声の特徴量 O を用いて、音響モデル補正パラメータ $\theta = (\theta$

10

20

30

40

50

$\theta_1, \theta_2, \dots, \theta_K$) を推定し (s 8 6) 、出力する (s 8 7) 。音響モデル補正パラメータ θ は、適応データ (学習用音声データの特徴量ベクトル系列 O) とその特徴量ベクトル系列 O に対応する正解シンボル系列 S_r とを用いて、次式のように目的関数 F_θ を最大化する形で推定される。

【 0 0 3 5 】

【数 4】

$$\hat{\theta} = \arg \max_{\theta} F_{\theta}(O, S_r) \quad (4)$$

【 0 0 3 6 】

非特許文献 4 では、目的関数として M M I 基準を用いるので、式 (4) の代わりに式 (5) を使用する。

【 0 0 3 7 】

【数 5】

$$\hat{\theta} = \arg \max_{\theta} F_{\theta}^{MMI}(O, S_r) \quad (5)$$

【 0 0 3 8 】

ここで、M M I 目的関数は次式のように書ける。

【 0 0 3 9 】

【数 6】

$$F_{\theta}^{MMI}(O, S_r) = \frac{1}{\psi} \log \frac{P(S_r)^{\psi\eta} p_{\Lambda}(O|S_r; \theta)^{\psi}}{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O|S_j; \theta)^{\psi}} \quad (6)$$

【 0 0 4 0 】

ここで、 S_j は特徴量ベクトル系列 O を音声認識することによって得られる対立候補シンボル系列、 $P(S_r)$ 及び $P(S_j)$ は言語モデルによってそれぞれ得られる正解シンボル系列 S_r の言語確率及び対立候補シンボル系列 S_j の言語確率、 $p_{\Lambda}(O|S_r)$ 及び $p_{\Lambda}(O|S_j)$ はそれぞれ正解シンボル系列 S_r 及び対立候補シンボル系列 S_j において音響モデル (H M M) によって得られる音響スコア、 ψ は音響スコアに対するスケールパラメータ、 η は言語確率に対するスケールパラメータを表す。

【 0 0 4 1 】

また、音響スコア $p_{\Lambda}(O|S_j)$ は以下の式で書くことができる。

【 0 0 4 2 】

【数 7】

$$p_{\Lambda}(O|S_j) = \sum_{\{n_1:T\}} \prod_{t=1}^T p(o_t|n_t) p(n_t|n_{t-1}) \quad (7)$$

【 0 0 4 3 】

ただし、 T は適応データ (学習用音声データの特徴量ベクトル系列 O) の特徴量ベクトル系列の長さ、 t はフレーム番号またはそのフレームに対応する時刻 (以下「フレーム時刻」ともいう) を表し、 1 から T の整数である。つまり、特徴量ベクトル系列 O は 1 から T フレーム目までの D 次元特徴量ベクトルで表現されるデータである (式 (1) 参照) 。また、 $\{n_1 : T\}$ は対立候補シンボル系列 S_j に対応する H M M 状態シーケンス (フレーム時刻 1 から T まで) であり、 $\sum_{\{n_1 : T\}}$ (ただし、下付添字 $\{n_1 : T\}$ は $\{n_1 : T\}$ を表す) は対立候補シンボル系列 S_j に対応する可能なあらゆる H M M 状態シーケンスの足し算、 $p(o_t|n_t)$ はフレーム時刻 t における H M M 状態 n_t から特徴量ベクトル o_t が出力される確率 (なお、H M M 状態は一般に G M M で表現される) 、 $p(n$

10

20

30

40

50

$t | n_{t-1}$) はフレーム時刻 $t - 1$ におけるある HMM 状態 n_{t-1} からフレーム時刻 t におけるある HMM 状態 n_t への遷移確率である。

【 0 0 4 4 】

〔 第一実施形態 〕

〔 第一実施形態のポイント 〕

しかし、前述の通り、MMI - LR の識別基準による音響モデルの教師なし適応方法は、適応データを音声認識した結果を正解ラベルと見做して利用するので、正解ラベル（と見做された音声認識結果）に誤りが含まれることが多く、うまくモデルパラメータを最適化できず、性能が改善しないか悪化する可能性がある。

【 0 0 4 5 】

10

そこで、第一実施形態では、正解シンボルの誤りを考慮し、識別学習による音響モデル補正パラメータ推定を行うために、音響モデル補正パラメータを推定する基準として参考文献 1 及び参考文献 2 に記述されている dMMI 導関数基準（differenced MMI: dMMI）を応用する。

〔 参考文献 1 〕 McDermott, E., Watanabe, S. and Nakamura, A., “Discriminative training based on an integrated view of MPE and MMI in margin and error space”, In Proc. ICASSP'10, 2010, pp. 4894 - 4897

〔 参考文献 2 〕 特願 2 0 0 9 - 1 9 8 3 6 2 号公報

【 0 0 4 6 】

すなわち、dMMI - LR を開発した。以下で数式を用いて dMMI - LR による音響モデル補正パラメータの推定方法について述べる。

20

【 0 0 4 7 】

まず、以下の Ψ 関数を以下のように定義する。

【 0 0 4 8 】

【 数 8 】

$$\Psi_{\theta, \sigma} := \sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O | S_j; \theta)^{\psi} e^{\psi\sigma\epsilon_{j,r}} \quad (8)$$

【 0 0 4 9 】

ここで σ はマージンパラメータ、 $\epsilon_{j,r}$ は正解シンボル系列 S_r に対する対立候補シンボル系列 S_j の相違度（例えば、単語エラー数、音素エラー数等）を表している。すなわち、マージンパラメータ σ は相違度 $\epsilon_{j,r}$ に応じて音響モデル補正パラメータ推定時に対立候補シンボル系列 S_j をどれだけ重視するかをコントロールするパラメータである。マージンパラメータ σ は、 $-\infty \sim +\infty$ の値を取り得る。マージンパラメータ σ がマイナスの値を取れば、小さい相違度 $\epsilon_{j,r}$ を持つ対立候補シンボル系列 S_j ほど、すなわち、エラー数が少ない対立候補シンボル系列 S_j ほど重視される。逆に、マージンパラメータ σ がプラスの値を取れば、大きい相違度 $\epsilon_{j,r}$ を持つ対立候補シンボル系列 S_j ほど、すなわち、エラー数が多い対立候補シンボル系列 S_j ほど重視されることになる。この Ψ 関数を用いると、dMMI 識別学習基準の目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ （ただし、下付添字 σ_1, σ_2 は、 σ_1, σ_2 を表す）は次式のように書くことができる。

30

40

【 0 0 5 0 】

【 数 9 】

$$\begin{aligned} F_{\theta, \sigma_1, \sigma_2}^{dMMI} &= \frac{1}{\psi(\sigma_2 - \sigma_1)} \log \frac{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O | S_j; \theta)^{\psi} e^{\psi\sigma_1\epsilon_{j,r}}}{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O | S_j; \theta)^{\psi} e^{\psi\sigma_2\epsilon_{j,r}}} \\ &= \frac{1}{\psi(\sigma_2 - \sigma_1)} \log \frac{\Psi_{\theta, \sigma_1}}{\Psi_{\theta, \sigma_2}} \quad (9) \end{aligned}$$

50

【 0 0 5 1 】

分子の第一マージンパラメータ σ_1 はマイナスの値を取る ($\sigma_1 < 0$)。すなわち、分子では相違度 $\varepsilon_{j,r}$ が小さい対立候補シンボル系列 S_j ほど重視される。一方、分母の第二マージンパラメータ σ_2 はプラスの値を取る ($\sigma_2 > 0$)。すなわち、分母では相違度 $\varepsilon_{j,r}$ が大きい対立候補シンボル系列 S_j ほど重視される。

【 0 0 5 2 】

この d M M I 識別学習基準は、第一マージンパラメータ σ_1 及び第二マージンパラメータ σ_2 を調整することによって、M P E (Minimum Phone Error) 識別学習基準 (参考文献 3) または B M M I (boosted-MMI) 識別学習基準 (参考文献 4) に近づく。

[参考文献 3] Povey, D., Woodland, P.C., "Minimum Phone Error and l-smoothing for improved discriminative training", In Proc. ICASSP, 2002, vol.1, pp.1-105-1-108

[参考文献 4] Povey, D., Kanevsky, D., Kingsbury, B., Ramabhadran, B., Saon, G. and Visweswariah, K., "Boosted MMI for model and feature-space discriminative training", In Proc. ICASSP, 2008, pp.4057-4060, ここで σ_2 は、例えば、+ 0.1 という 0 に近い小さなプラスの値に設定しておけばよい (参考文献 5)。

[参考文献 5] Saon, G. and Povey, D., "Penalty function maximization for large margin HMM training", In Proc. Interspeech, 2008, pp.920-923

例えば、 σ_1 を大きなマイナスの値 (理論的には $-\infty$ 、実装上は、例えば、- 50) に設定する。このとき、次式に示すように、d M M I 識別学習基準の目的関数 $F^{d M M I}_{\theta, \sigma_1, \sigma_2}$, σ_1, σ_2 は B M M I 識別学習基準の目的関数 $F^{B M M I}_{\theta, \sigma_2}$ に近づく。

【 0 0 5 3 】

【数 10】

$$\lim_{\sigma_1 \rightarrow -\infty} F_{\theta, \sigma_1, \sigma_2}^{dMMI} \propto F_{\theta, \sigma_2}^{BMMI}$$

$$= \frac{1}{\psi} \log \frac{P(S_r)^{\psi\eta} p_{\Lambda}(O|S_r; \theta)^{\psi}}{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O|S_j; \theta)^{\psi} e^{\psi\sigma_2 \varepsilon_{j,r}}} \quad (10)$$

【 0 0 5 4 】

式 (10) から明らかなように、B M M I 識別学習基準の目的関数 $F^{B M M I}_{\theta, \sigma_2}$ では、分母の第二マージンパラメータ σ_2 のみが残る。すなわち B M M I 識別学習基準による音響モデル補正パラメータ推定では、エラー数が多い (相違度 $\varepsilon_{j,r}$ が大きい) 対立候補シンボル系列 S_j ほど重視される (参考文献 6 参照)。

[参考文献 6] Povey, D., Kanevsky, D., Kingsbury, B., Ramabhadran, B., Saon, G. and Visweswariah, K., "Boosted MMI for model and feature-space discriminative training", In Proc. ICASSP, 2008, pp.4057-4060

【 0 0 5 5 】

B M M I の分子は直接正解シンボルの貢献が考慮されるので、正解シンボルの誤りの影響を受けやすい。一方、d M M I の場合、 σ_1 をより大きい値 (例えば - 10) に設定することによって、分子では認識の対立候補シンボル系列 S_j の貢献の足し算になる。マージン $\exp(\psi \sigma_1 \varepsilon_{j,r})$ を重みとして、正解シンボル系列 S_r に近い対立候補シンボル系列 S_j (正解シンボル S_r に対する誤りが少ない) が考慮される。そのため、分子には、正解シンボル系列 S_r だけでなく、それに近い対立候補シンボル系列 S_j も考慮されることによって、正解シンボルの誤りの悪影響を弱めることができる。その結果、正解シンボル系列 S_r に誤りがあっても、音響モデル補正パラメータを識別基準により安定して精度よく推定することができる。マージンパラメータ σ_1 の値は、分子でどのぐらいの対立候補シンボル系列 S_j を考慮するかを決める値である。 σ_1 の値はタスクの認識率などに

依存する。ただし、例えば -3 ~ -10 の間の値に設定すると良い。

【0056】

ここで、音響モデル補正パラメータ群の集合 θ は、次式のように、上記の dMMI 識別学習基準の目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を最大化するように推定される。

【0057】

【数11】

$$\hat{\theta} = \arg \max_{\theta} F_{\theta, \sigma_1, \sigma_2}^{dMMI}(O, S_r) \quad (11)$$

【0058】

ここでは、 W_k の推定方法について述べる。dMMI 識別学習基準の目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を最大化するような W_k を求めるために、まず目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を W_k で微分する。ここで、対立候補シンボル系列 S_j が単語（または音素）ラティスの形態で表現される場合、ラティス上で勾配を計算することにより、 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を W_k で微分した値は以下のように表現される。

【0059】

【数12】

$$\frac{\partial F_{\theta, \sigma_1, \sigma_2}^{dMMI}(O, S_r)}{\partial W_k} = - \sum_{t, q_t, n_t, m} \gamma_{q_t}^{dMMI} \gamma_{n_t, m} \Delta_{m, k} \Sigma_{n_t, m}^{-1} (o_t - \mu_{n_t, m}) \xi_{n_t, m} \quad (12)$$

$$\Delta_{m, k} = \begin{cases} 1 & m \in k \text{ 番目のガウス分布のクラス} \\ 0 & m \notin k \text{ 番目のガウス分布のクラス} \end{cases}$$

【0060】

ここで、 q_t はフレーム時刻 t におけるラティスのアークを、 n_t はフレーム時刻 t における音響モデル（例えば、HMM からなる音響モデル）の状態を、 m は状態 n_t におけるガウス分布のインデックスを表す（例えば、HMM の状態の出力確率分布は GMM で表現されるものとする）。また、 $\gamma^{dMMI}_{q_t}$ （ただし、下付添字 q_t は q_t を表す）は単語（または音素）ラティスのアーク q_t の事後確率であり、同じラティスについて、第一

マージンパラメータ σ_1 または第二マージンパラメータ σ_2 を使って、二度、Forward-Backward algorithm を実行して計算される（参考文献 7 参照）。

【参考文献 7】E. McDermott, T.J. Hazen, J.L. Roux, A. Nakamura and S. Katagiri, "Discriminative Training for Large-Vocabulary Speech Recognition Using Minimum Classification Error", IEEE Trans. ASLP, 2007, vol. 15, no. 1, pp. 203 - 223

【参考文献 8】V. Valtchev, J.J. Odell, P.C. Woodland, and S.J. Young, "Lattice-based discriminative training for large vocabulary speech recognition", In Proc. ICSLP, 1996, vol. 2, pp. 605-609

【0061】

例えば R-Prop（参考文献 9）のような勾配法を用いる場合は、微分値（ $\partial F^{dMMI}_{\theta, \sigma_1, \sigma_2} / \partial W_k$ ）が 0 に近づくように W_k を更新すればよい。または例えば、Extended-Baum Welch のようなアルゴリズム（参考文献 3）を用いて、 W_k を更新することもできる。このように、上記の式が得られれば、平均補正パラメータ W_k は容易に推定することが可能である。

【参考文献 9】Riedmiller, M. and Braun, H., "A direct adaptive method for faster backpropagation learning: The RPROP algorithm", In Proc. ICNN '93, 1993, p

10

20

30

40

50

p. 586-591

【 0 0 6 2 】

< 音響モデル補正パラメータ学習装置 1 0 0 >

以上の原理に基づく、第一実施形態の音響モデル補正パラメータ学習装置 1 0 0 の構成について説明する。装置構成図を図 5 に、処理フローを図 6 に示す。

【 0 0 6 3 】

音響モデル補正パラメータ学習装置 1 0 0 は、特徴量抽出部 1 1 0、音響モデル補正部 1 2 0、エラーカウント計算部 1 3 0、補正パラメータ微分値計算部 1 4 0、補正パラメータ更新部 1 5 0、収束判定部 1 6 0、音響モデル記憶部 1 7 0 及び言語モデル記憶部 1 8 0 を含む。

10

【 0 0 6 4 】

音響モデル補正パラメータ学習装置 1 0 0 は、学習用音声データ（以下「適応用音声データ」ともいう）とその正解シンボル系列 S_r と音響モデル補正パラメータの初期値 θ^0 とを受け取り、音響モデル補正パラメータを更新し、最適な音響モデル補正パラメータを求め、出力する。なお、本実施形態では、音響モデル補正パラメータは、平均補正パラメータのみからなり、クラスタ毎に求められるものとしているが、他の音響モデル補正パラメータを含んでもよいし、また、クラスタ毎に求めなくともよい。

【 0 0 6 5 】

（特徴量抽出部 1 1 0）

特徴量抽出部 1 1 0 は、学習用音声データを読み込み（s 1 0 3）、その特徴量ベクトル系列 O を抽出し（s 1 0 5）、音響モデル補正部 1 2 0 に出力する。特徴量抽出の具体的な処理は、既存の技術を用いることができる。例えば、上述の音声認識装置 9 0 の特徴量抽出部 9 1 と同じ方法により特徴量を抽出すればよい。

20

【 0 0 6 6 】

（音響モデル記憶部 1 7 0 及び言語モデル記憶部 1 8 0）

音響モデル記憶部 1 7 0 及び言語モデル記憶部 1 8 0 には、それぞれ予め求められた音響モデル及び言語モデルが記憶される。音響モデル及び言語モデルとしては、既存のモデルを用いればよい。例えば、記憶部 9 3 において説明した音響モデルと言語モデルを用いることができる。

30

【 0 0 6 7 】

（音響モデル補正部 1 2 0）

音響モデル補正部 1 2 0 は、補正前の音響モデル Λ を音響モデル記憶部 1 7 0 から読み込み（s 1 0 1）、音響モデル補正パラメータの初期値 θ^0 または更新された音響モデル補正パラメータ θ^{i-1} （ただし、 i は、繰り返し回数を表すインデックスを示す）を受け取り、式（3）により音響モデルの平均ベクトルを補正し（s 1 0 6）、補正後の音響モデル $\Lambda^{\wedge}$ を、エラーカウント計算部 1 3 0 に出力する。

【 0 0 6 8 】

【数 1 3】

$$\hat{\mu}_m = A_k \mu_m + b_k = W_k \xi_m \quad (3)$$

40

【 0 0 6 9 】

ただし、 $\theta^0 = \{\theta_1^0, \theta_2^0, \dots, \theta_K^0\}$ であり、 $\theta_k^0 = \{W_k^0\}$ である。同様に、 $\theta^{i-1} = \{\theta_1^{i-1}, \theta_2^{i-1}, \dots, \theta_K^{i-1}\}$ であり、 $\theta_k^{i-1} = \{W_k^{i-1}\}$ である。初期値 W_k^0 を構成する A_k^0 、 b_k^0 としては、例えば、それぞれ単位行列、ゼロベクトル（全ての要素が 0 のベクトル）等が考えられる。

【 0 0 7 0 】

（エラーカウント計算部 1 3 0）

エラーカウント計算部 1 3 0 は、言語モデル記憶部 1 8 0 から言語モデルを読み込み（s 1 0 2）、この言語モデルと、音響モデル補正部 1 2 0 から受け取った補正後の音響モデル $\Lambda^{\wedge}$ とを用いて、特徴量抽出部 1 1 0 から受け取った特徴量ベクトル系列 O を音声認

50

識することによって得られるJ個の対立候補シンボル系列 S_j を求める。さらに、エラーカウント計算部 130 は、入力された正解シンボル系列 S_r を読み込み (s 104)、予め定めた粒度で、対立候補シンボル系列 S_j 毎に、正解シンボル系列 S_r との相違度 $\varepsilon_{j,r}$ を求め (s 107)、補正パラメータ微分値計算部 140 に出力する。特に、予め定めた粒度を音素以下の粒度とすれば、相互情報量最大化の枠組みで粒度の細かい相違度を用いることが可能となる。例えば、予め定めた粒度 (音素や単語等) で読み込んだ正解シンボル系列 S_r と求めた対立候補シンボル系列 S_j との異なる部分をカウントし、カウント値を相違度 $\varepsilon_{j,r}$ として求める。

【0071】

(補正パラメータ微分値計算部 140)

10

補正パラメータ微分値計算部 140 は、言語モデル記憶部 180 から言語モデルを読み込み (s 102)、入力された正解シンボル系列 S_r を読み込み (s 104)、補正後の音響モデル Λ を受け取り、エラーカウント計算部 130 から受け取った対立候補シンボル系列 S_j と相違度 $\varepsilon_{j,r}$ とを用いて、式 (9) で表される目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を求める。

【0072】

【数 14】

$$F^{dMMI}_{\theta, \sigma_1, \sigma_2} = \frac{1}{\psi(\sigma_2 - \sigma_1)} \log \frac{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O|S_j; \theta)^{\psi} e^{\psi\sigma_1 \varepsilon_{j,r}}}{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(O|S_j; \theta)^{\psi} e^{\psi\sigma_2 \varepsilon_{j,r}}} \quad (9)$$

$$= \frac{1}{\psi(\sigma_2 - \sigma_1)} \log \frac{\Psi_{\theta, \sigma_1}}{\Psi_{\theta, \sigma_2}} \quad (9)$$

20

【0073】

ただし、第一マージンパラメータ σ_1 の調整は、学習用音声データの特徴と認識用音声データの特徴との不一致の度合いを考慮して人手により行われるものとする。第二マージンパラメータ σ_2 は、例えば、+0.1 という 0 に近い小さなプラスの値とする。

【0074】

30

さらに、補正パラメータ微分値計算部 140 は、目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を音響モデル補正パラメータ $W_k = [A_k \quad b_k]$ で微分する (式 (12)、s 108)。

【0075】

【数 15】

$$\frac{\partial F^{dMMI}_{\theta, \sigma_1, \sigma_2}(O, S_r)}{\partial W_k} = - \sum_{t, q_t, n_t, m} \gamma_{q_t}^{dMMI} \gamma_{n_t, m} \Delta_{m, k} \Sigma_{n_t, m}^{-1} (o_t - \mu_{n_t, m}) \xi_{n_t, m} \quad (12)$$

$$\Delta_{m, k} = \begin{cases} 1 & m \in k \text{ のガウスクラスタ} \\ 0 & m \notin k \text{ のガウスクラスタ} \end{cases}$$

40

【0076】

算出した微分値 ($\partial F^{dMMI}_{\theta, \sigma_1, \sigma_2} / \partial W_k$) を補正パラメータ更新部 150 に出力する。

【0077】

(補正パラメータ更新部 150)

補正パラメータ更新部 150 は、微分値 ($\partial F^{dMMI}_{\theta, \sigma_1, \sigma_2} / \partial W_k$) に応じて平均補正パラメータ W_k を変更することで、平均補正パラメータを更新する。つまり、式 (9) の目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を最大化するように、式 (11) に従い、 W_k 、すなわち A_k および b_k を同時に更新する (s 109)。

50

【 0 0 7 8 】

【 数 1 6 】

$$\hat{\theta} = \arg \max_{\theta} F_{\theta, \sigma_1, \sigma_2}^{dMMI}(O, S_r) \quad (11)$$

【 0 0 7 9 】

例えば R - P r o p (参考文献 9) のような勾配法を用いる場合は、微分値 ($\partial F^{dMMI}_{\theta, \sigma_1, \sigma_2} / \partial W_k$) が 0 に近づくように W_k を更新すればよい。または例えば、Extended-Baum Welch のようなアルゴリズム (参考文献 3) を用いて、 W_k を更新することもできる。更新後の音響モデル補正パラメータ $\theta^{\wedge} = (W_1, \dots, W_k, \dots, W_K)$ を収束判定部 1 6 0 に出力する。

10

【 0 0 8 0 】

(収束判定部 1 6 0)

収束判定部 1 6 0 は、音響モデル補正パラメータ $\theta^{\wedge}$ を受け取り、音響モデル補正パラメータの推定が収束したか否かを判定し (s 1 1 0)、収束していると判定した場合には、収束時の音響モデル補正パラメータ $\theta^{\wedge}$ を、音響モデル補正パラメータ推定装置の出力値として、出力する (s 1 1 1)。収束していないと判定した場合には、音響モデル補正パラメータ $\theta^{\wedge}$ を音響モデル補正部 1 2 0 に出力し、音響モデル補正部 1 2 0、エラーカウント計算部 1 3 0、補正パラメータ微分値計算部 1 4 0、補正パラメータ更新部 1 5 0、収束判定部 1 6 0 の処理を繰り返すように制御信号を出力する。収束判定部 1 6 0 は、例えば、(1) 一つ前に求めた音響モデル補正パラメータと今回求めた音響モデル補正パラメータとの差分が閾値以下になった場合や (2) 繰り返し回数が所定の回数以上になった場合に、収束していると判定する。

20

【 0 0 8 1 】

< シミュレーション結果 >

以下の表は効果の例として、大語彙連続音声認識タスクで、話者に対する教師なし音響モデル適応の実験結果を表す。このように、本発明は従来の音響モデル適応 (M L L R、M M I - L R) よりも性能を改善することがわかる。

【 0 0 8 2 】

【 表 1 】

30

	単語誤り率 (%)
ベースライン (音響モデル適応なし)	30.6
M L L R	26.9
M M I - L R	27.1
第一実施形態 (d M M I - L R)	25.8

40

【 0 0 8 3 】

< 効果 >

このような構成により、正解シンボルの誤りの悪影響を弱めることができ、従来技術 (M L L R や M M I - L R の識別基準に基づく音響モデル適応) よりも適切に音響モデルパラメータに対する補正パラメータを求めることができる。さらに、このようにして求めた音響モデル補正パラメータを用いて、補正した音響モデルを用いて音声認識を行うことで、従来技術に比べ、音声認識精度を改善できる。

【 0 0 8 4 】

< 変形例 >

第一実施形態では、音響モデル補正パラメータは平均補正パラメータのみを含むが、混

50

合ガウス分布モデルに含まれるガウス分布の分散パラメータ Σ_m を補正する分散補正パラメータも含む構成としてもよい。

【 0 0 8 5 】

この場合、以下の式 (1 3) や (1 4) により分散パラメータ Σ_m を補正することで、d M M I 基準による推定をすることができる。

【 0 0 8 6 】

【 数 1 7 】

$$\hat{\Sigma}_m = D_k \Sigma_m D_k' \quad (13)$$

【 0 0 8 7 】

10

もしくは、

【 0 0 8 8 】

【 数 1 8 】

$$\hat{\Sigma}_m = \Sigma_m + D_k \quad (14)$$

【 0 0 8 9 】

ここで、 D_k は分散補正パラメータである。

【 0 0 9 0 】

平均補正パラメータ $W_k = [A_k, b_k]$ と分散補正パラメータ D_k の同時推定は以下のように行うことができる。

20

【 0 0 9 1 】

【 数 1 9 】

$$\hat{\theta}^b = \underset{\theta^b}{\operatorname{argmax}} F_{\theta^b, \sigma_1, \sigma_2}^{dMMI}(O, S_r) \quad (15)$$

【 0 0 9 2 】

ただし、 $\theta^b = (W_1, D_1, \dots, W_k, D_k, \dots, W_K, D_K)$ である。式 (1 5) は平均と共分散パラメータの同時推定を示しているが、共分散だけ推定することも可能である。

【 0 0 9 3 】

d M M I 識別学習基準の目的関数 $F_{\theta^b, \sigma_1, \sigma_2}^{dMMI}$ (ただし、下付添字 θ^b は、 θ^b を表す。) を最大化するような D_k を求めるために、まず目的関数 $F_{\theta^b, \sigma_1, \sigma_2}^{dMMI}$ を D_k で微分する。目的関数 $F_{\theta^b, \sigma_1, \sigma_2}^{dMMI}$ を D_k で微分した値は、分散補正パラメータが式 (1 3) により補正される場合、以下のように表現される。

30

【 0 0 9 4 】

【 数 2 0 】

$$\frac{\partial F_{\theta^b, \sigma_1, \sigma_2}^{dMMI}(O, S_r)}{\partial D_k}$$

$$= - \sum_{t, q_t, n_t, m} \gamma_{q_t}^{dMMI} \gamma_{n_t, m} \Lambda_{m, k} \left[- \hat{\Sigma}_{n_t, m}^{-1} + \hat{\Sigma}_{n_t, m}^{-1} (o_t - \mu_{n_t, m}) (o_t - \mu_{n_t, m})' \left(\hat{\Sigma}_{n_t, m}^{-1} \right)' \right] \left(\Sigma_{n_t, m} D_k' + D_k \Sigma_{n_t, m} \right)$$

40

(16)

【 0 0 9 5 】

また、式 (1 5) は、分散補正パラメータが式 (1 4) により補正される場合、以下のように表現される。

【 0 0 9 6 】

【数 2 1】

$$\frac{\partial F^{dMMI}_{\theta^b, \sigma_1, \sigma_2}(O, S_r)}{\partial D_k}$$

$$= - \sum_{t, q_t, n_t, m} \gamma_{q_t}^{dMMI} \gamma_{n_t, m} \Delta_{m, k} \left[-\hat{\Sigma}_{n_t, m}^{-1} + \hat{\Sigma}_{n_t, m}^{-1} (o_t - \mu_{n_t, m})(o_t - \mu_{n_t, m})' (\hat{\Sigma}_{n_t, m}^{-1})' \right]$$

(17)

【0097】

10

また、平均ベクトルの変換行列 A_k と分散パラメータの変換行列（式（13）の D_k ）が同じになるように制約を加えてもよい。その場合は、平均ベクトルと分散パラメータの補正は式（18）のように行い、音響モデル補正パラメータの推定は式（19）で行う。

【0098】

【数 2 2】

$$\hat{\mu}_m = A_k^c \mu_m + b_k^c$$

$$\hat{\Sigma}_m = A_k^c \Sigma_m (A_k^c)' \quad (18)$$

【0099】

なお、 A_k^c は平均ベクトル及び分散パラメータに対する変換行列、 b_k^c は平均ベクトルに対するバイアスベクトルである。

20

【0100】

【数 2 3】

$$\hat{\theta}^c = \arg \max_{\theta^c} F^{dMMI}_{\theta^c, \sigma_1, \sigma_2}(O, S_r) \quad (19)$$

【0101】

ただし、 $\theta^c = (A_{k1}^c, b_{k1}^c, \dots, A_{kK}^c, b_{kK}^c, \dots, A_{Kk}^c, b_{Kk}^c)$ である。

【0102】

変形例に係る音響モデル補正パラメータ学習装置 100 の構成について、第一実施形態と異なる部分を中心に説明する。

30

【0103】

（音響モデル補正部 120）

音響モデル補正部 120 は、式（3）により、平均ベクトルを補正し、さらに、式（13）または（14）により分散パラメータ Σ_m を補正する（s106）。なお、音響モデル適応技術を搭載した音声認識装置 90 の音響モデル補正部 94 においても対応する式（式（13）または式（14））により、分散パラメータ Σ_m を補正する。

【0104】

（エラーカウント計算部 130）

エラーカウント計算部 130 は、補正した平均ベクトルと補正した共分散行列とを含む音響モデルと言語モデルとに基づき、学習用音声データの特徴量を音声認識することによって得られる対立候補シンボル系列ごとに、予め定めた粒度で、正解シンボル系列との相違度を求める（s107）。

40

【0105】

（補正パラメータ微分値計算部 140）

補正パラメータ微分値計算部 140 は、式（15）の右辺で用いられている目的関数 $F^{dMMI}_{\theta^b, \sigma_1, \sigma_2}$ を求める。さらに、補正パラメータ微分値計算部 140 は、目的関数 $F^{dMMI}_{\theta^b, \sigma_1, \sigma_2}$ を音響モデル補正パラメータ W_k 、 D_k でそれぞれ微分する（式（12）、（16）または（17）s108）。

【0106】

50

(補正パラメータ更新部 150)

補正パラメータ更新部 150 は、平均補正パラメータ W_k に対する微分値 $\partial F^{dMMI}_{\theta_b, \sigma_1, \sigma_2} / \partial W_k$ に応じて平均補正パラメータ W_k を変更することで、平均補正パラメータを更新し、さらに、分散補正パラメータ D_k に対する微分値 $\partial F^{dMMI}_{\theta_b, \sigma_1, \sigma_2} / \partial D_k$ に応じて分散補正パラメータ D_k を変更することで、分散補正パラメータ D_k を更新する (s109)。

【0107】

(収束判定部 160)

収束判定部 160 は、平均補正パラメータ及び分散補正パラメータの更新が予め定めた条件を満たすか否かを判定し (s110)、満たす場合には、更新後の平均補正パラメータ及び分散補正パラメータをそれぞれ求める平均補正パラメータ及び分散補正パラメータとして出力し (s111)、条件を満たさない場合には、音響モデル補正部 120、エラーカウンタ計算部 130、補正パラメータ微分値計算部 140 及び補正パラメータ更新部 150 の処理を繰り返す。

【0108】

このような構成により、適切に、平均補正パラメータに加え分散補正パラメータを求めることができる。

【0109】

<その他の変形例>

音響補正パラメータをクラス毎に求めなくともよい。その場合、式 (2) で平均ベクトルを補正し、式 (13a) または式 (14a) で分散パラメータを補正する。

【0110】

【数24】

$$\hat{\Sigma}_m = D \Sigma_m D' \quad (13a)$$

$$\hat{\Sigma}_m = \Sigma_m + D \quad (14a)$$

【0111】

また、本実施形態では、音響モデル補正パラメータ学習装置 100 は、特徴量抽出部 110 を含むが、学習用音声データに対する特徴量を入力される場合には、特徴量抽出部 110 を含まなくともよい。

【0112】

図示しない記憶部に予め目的関数 $F^{dMMI}_{\theta_b, \sigma_1, \sigma_2}$ を音響モデル補正パラメータで微分したもの (例えば式 (12)、(16)、(17) で表される) を計算式として記憶しておいてもよい。この場合、実際の微分値は以下のようにして求める。補正パラメータ微分値計算部 140 は、計算式を記憶部から読み込み、さらに、言語モデル記憶部 180 から言語モデルを読み込み、正解シンボル系列 S_r を読み込み、補正後の音響モデル Λ と対立候補シンボル系列 S_j と相違度 $\varepsilon_{j,r}$ とを受け取り、計算式に代入し、微分値 ($\partial F^{dMMI}_{\theta_b, \sigma_1, \sigma_2} / \partial W_k$) や ($\partial F^{dMMI}_{\theta_b, \sigma_1, \sigma_2} / \partial D_k$) を算出し (s108)、補正パラメータ更新部 109 に出力する。

【0113】

〔第二実施形態〕

式 (3) の音響モデルパラメータの補正は式 (20) のように、特徴量の補正と等しくなることを示すことができる。

【0114】

【数25】

$$\hat{o}_t = A_k^f o_t + b_k^f = W_k^f [o_t' \quad 1]' \quad (20)$$

【0115】

A_k^f は特徴量に対する変換行列、 b_k^f は特徴量に対するバイアスベクトルである。W

10

20

30

40

50

$k^f := [A_k^f \ b_k^f]$ である。

目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を特徴量補正パラメータ W_k^f で微分すると、次式のようになる。

【0116】

【数26】

$$\frac{\partial F_{\theta, \sigma_1, \sigma_2}^{dMMI}(O, S_r)}{\partial W_k^f} = - \sum_{t, q_t, n_t, m} \gamma_{q_t}^{dMMI} \gamma_{n_t, m} \Delta_{m, k} \Sigma_{n_t, m}^{-1} (\mu_{n_t, m} - \hat{o}_t) [o_t' \ 1]' \quad (21)$$

【0117】

10

第一実施形態では、音響モデルを補正することで話者適応を行う構成を前提とし、音響モデル補正パラメータを推定する構成を説明したが、式(20)を用いると、本発明は特徴量の補正パラメータ(以下「特徴量補正パラメータ」ともいう)の推定にも応用することができる。

【0118】

まず、補正した特徴量に基づき音声認識を行う音声認識装置70について説明する。

【0119】

<音声認識装置70>

図7に音声認識装置70の機能構成例、図8にその処理フロー例を示す。音声認識装置70は、特徴量抽出部91、特徴量補正部71、単語列探索部72、記録部74から構成

20

【0120】

予め音響モデル及び言語モデルを記録部74に記録している。さらに、予め特徴量補正パラメータ W^f を記録部74に記録している。 $W^f = (W_1^f, W_2^f, \dots, W_K^f)$ とし、 $W_k^f = \{A_k^f \ b_k^f\}$ とする。なお、本実施形態では、 $W^f = \theta$ 、 $W_k = \theta_k^f$ とする。

【0121】

特徴量補正部71は特徴量補正パラメータ W^f を読み込む(s71)。特徴量抽出部91で抽出した特徴量ベクトル系列 O を、単語列探索部72に送る前に、特徴量補正部71において、予め求めておいた特徴量補正パラメータ W^f を用いて式(20)により、補正

30

【0122】

【数27】

$$\hat{o}_t = A_k^f o_t + b_k^f = W_k^f [o_t' \ 1]' \quad (20)$$

【0123】

単語列探索部72は、音響モデルと言語モデルを読み込む(s71、s72)。単語列探索部72は、まず、音響モデルに基づき、特徴量補正部71で補正された特徴量ベクトル系列 $O^{\wedge}$ に対する J 個の対立候補シンボル系列 S_j を生成して、対立候補シンボル系列 S_j 毎に音響スコアを算出する。次に、言語モデルに基づき、対立候補シンボル系列 S_j 毎に言語スコアを算出する。さらに、音響スコアと言語スコアとを統合して、 J 個の対立候補シンボル系列 S_j の中から、認識用音声データに対応する文として最も確からしい(最も音響スコアと言語スコアとを統合したスコアが高い)対立候補シンボル系列を探索し(s75)、その対立候補シンボル系列を認識結果(単語列) $S^{\wedge}$ として出力する(s76)。

40

【0124】

この特徴量ベクトル系列 O の補正は、最終的な音声認識精度を向上させることを目的として行われる。すなわち特徴量補正技術のポイントは、最終的な音声認識精度を向上させるための特徴量補正パラメータ W^f をいかに推定するか、という点にある。

【0125】

50

本実施形態では、特徴量補正パラメータ W^f を以下のように推定する。第一実施形態と異なる部分を中心に説明する。

【0126】

<特徴量補正パラメータ推定装置200>

特徴量補正パラメータ推定装置200の構成を図9に、処理フローを図10に示す。第一実施形態とは異なる処理についてのみ説明する。特徴量補正パラメータ推定装置200は、特徴量抽出部110、音響モデル補正部220、エラーカウント計算部230、補正パラメータ微分値計算部240、補正パラメータ更新部250、収束判定部260、音響モデル記憶部170及び言語モデル記憶部180を含む。

【0127】

10

(特徴量補正部220)

特徴量補正部220は、特徴量補正パラメータの初期値 W^{f0} または更新された特徴量補正パラメータ $W^{f(i-1)}$ と、補正前の特徴量ベクトル系列 O とを受け取り、式(20)に基づき特徴量ベクトル系列 O を補正し($s206$)、補正後の特徴量ベクトル系列 $O^{\wedge}$ をエラーカウント計算部230に出力する。

【0128】

【数28】

$$\hat{o}_t = A_k^f o_t + b_k^f = W_k^f [o_t^T \ 1]^T \quad (20)$$

【0129】

20

ただし、 $W^{f0} = \{W_1^{f0}, W_2^{f0}, \dots, W_K^{f0}\}$ であり、 $W_k^{f0} = \{A_k^{f0}, b_k^{f0}\}$ である。初期値 A_k^{f0} 、 b_k^{f0} としては、例えば、それぞれ単位行列、ゼロベクトル(全ての要素が0のベクトル)等が考えられる。同様に、 $W^{f(i-1)} = \{W_1^{f(i-1)}, W_2^{f(i-1)}, \dots, W_K^{f(i-1)}\}$ であり、 $W_k^{f(i-1)} = \{A_k^{f(i-1)}, b_k^{f(i-1)}\}$ である。また、本実施形態では、特徴量補正パラメータを θ とも記載する。

【0130】

(エラーカウント計算部230)

エラーカウント計算部230は、補正前の特徴量ベクトル系列 O の代わりに補正後の特徴量ベクトル系列 $O^{\wedge}$ を用いる点を除いては、第一実施形態と同じである($s102$ 、 $s104$ 、 $s207$)。

30

【0131】

(補正パラメータ微分値計算部240)

補正パラメータ微分値計算部240は、補正前の特徴量ベクトル系列 O の代わりに補正後の特徴量ベクトル系列 $O^{\wedge}$ を用い、補正後の音響モデル $\Lambda^{\wedge}$ の代わりに音響モデル Λ を用い、目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を微分する際に音響モデル補正パラメータ W_k の代わりに特徴量補正パラメータ W_k^f を用いる。

【0132】

よって、補正パラメータ微分値計算部240は、音響モデル記憶部170及び言語モデル記憶部180からそれぞれ音響モデル及び言語モデルを読み込み($s101$ 、 $s102$)、入力された正解シンボル系列 S_r を読み込み($s104$)、エラーカウント計算部130から受け取った対立候補シンボル系列 S_j と相違度 $\varepsilon_{j,r}$ とを用いて、次式で表される目的関数 $F^{dMMI}_{\theta, \sigma_1, \sigma_2}$ を求める。

40

【0133】

【数 2 9】

$$F_{\theta, \sigma_1, \sigma_2}^{dMMI} = \frac{1}{\psi(\sigma_2 - \sigma_1)} \log \frac{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(\hat{O} | S_j; \theta)^{\psi} e^{\psi\sigma_1 \varepsilon_{j,r}}}{\sum_j P(S_j)^{\psi\eta} p_{\Lambda}(\hat{O} | S_j; \theta)^{\psi} e^{\psi\sigma_2 \varepsilon_{j,r}}}$$

$$= \frac{1}{\psi(\sigma_2 - \sigma_1)} \log \frac{\Psi_{\theta, \sigma_1}}{\Psi_{\theta, \sigma_2}}$$

【0134】

10

ただし、第一マージンパラメータ σ_1 の調整は、学習用音声データの特徴と認識用音声データの特徴との不一致の度合いを考慮して人手により行われるものとする。第二マージンパラメータ σ_2 は、例えば、+0.1という0に近い小さなプラスの値とする。さらに、補正パラメータ微分値計算部240は、目的関数 $F_{\theta, \sigma_1, \sigma_2}^{dMMI}$ を特徴量補正パラメータ $W_k^f = [A_k^f \quad b_k^f]$ で微分する(式(21)、s208)。

【0135】

【数30】

$$\frac{\partial F_{\theta, \sigma_1, \sigma_2}^{dMMI}(O, S_r)}{\partial W_k^f} = - \sum_{t, q_t, n_t, m} \gamma_{q_t}^{dMMI} \gamma_{n_t, m} \Delta_{m, k} \Sigma_{n_t, m}^{-1} (\mu_{n_t, m} - \hat{o}_t) [o'_t \quad 1]' \quad (21)$$

20

【0136】

算出した微分値($\partial F_{\theta, \sigma_1, \sigma_2}^{dMMI} / \partial W_k^f$)を補正パラメータ更新部250に出力する。

【0137】

(補正パラメータ更新部250)

補正パラメータ更新部250は、微分値($\partial F_{\theta, \sigma_1, \sigma_2}^{dMMI} / \partial W_k^f$)を受け取り、微分値($\partial F_{\theta, \sigma_1, \sigma_2}^{dMMI} / \partial W_k^f$)に応じて W_k^f 、すなわち A_k^f および b_k^f を同時に更新する(s209)。更新後の特徴量補正パラメータ W_k^f を収束判定部160に出力する。

【0138】

30

(収束判定部260)

収束判定部260は、音響モデル補正パラメータの代わりに特徴量補正パラメータ W_k^f を用いる点を除いては、第一実施形態と同じである(s210、s211)。

【0139】

<効果>

このような構成により、正解シンボルの誤りの悪影響を弱めることができ、従来技術よりも適切に特徴量に対する補正パラメータを求めることができる。さらに、このようにして求めた特徴量補正パラメータを用いて、認識用音声データの特徴量を補正し、補正した特徴量に基づき音声認識を行うことで、従来技術に比べ、音声認識精度を改善できる。また、特徴量補正の場合は音響モデルパラメータを更新する必要がないという利点もある。

40

【0140】

<その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。

【0141】

<プログラム及び記録媒体>

また、上記の実施形態及び変形例で説明した各装置における各種の処理機能をコンピュータによって実現してもよい。その場合、各装置が有すべき機能の処理内容はプログラム

50

によって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

【0142】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等どのようなものでもよい。

【0143】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させてもよい。

【0144】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶部に格納する。そして、処理の実行時、このコンピュータは、自己の記憶部に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実施形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよい。さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP(Application Service Provider)型のサービスによって、上述の処理を実行する構成としてもよい。なお、プログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

【0145】

また、コンピュータ上で所定のプログラムを実行させることにより、各装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

10

20

30

【図 1】

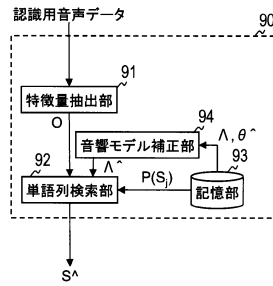


図1

【図 3】

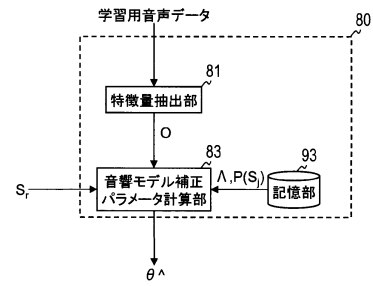


図3

【図 2】

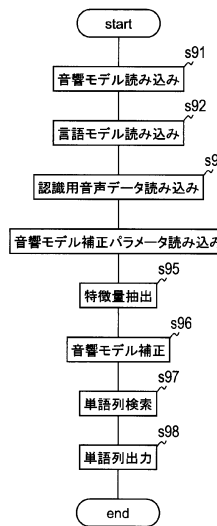


図2

【図 4】

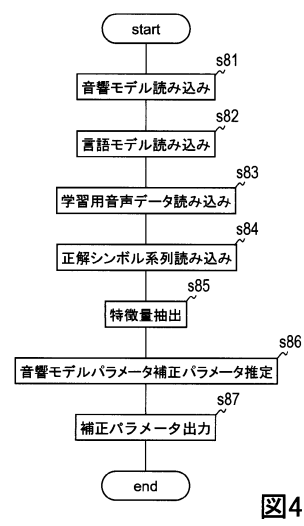


図4

【図 5】

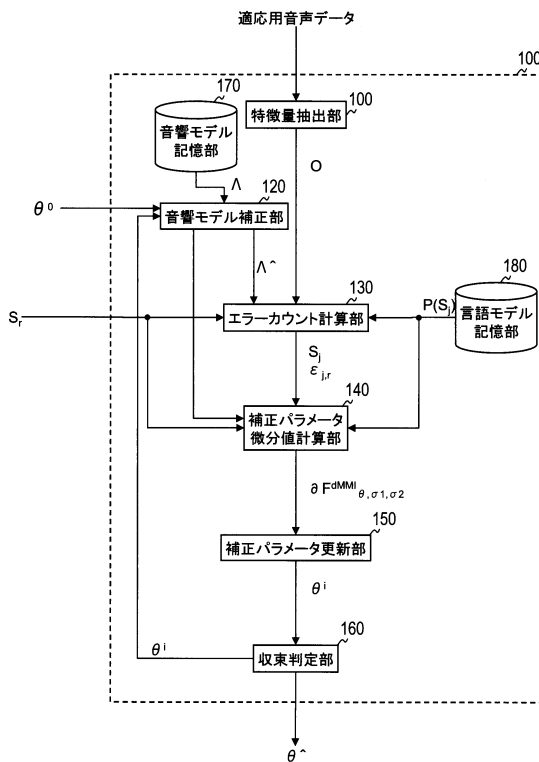


図5

【図 6】

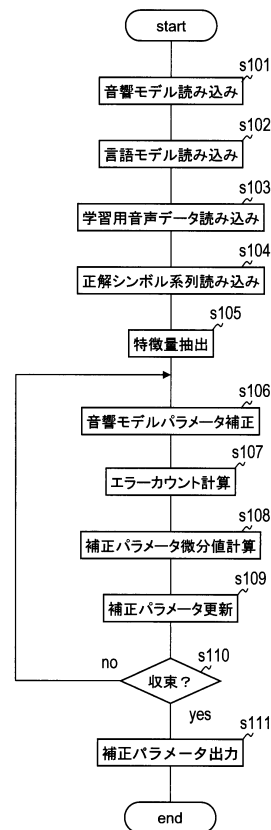


図6

【図 7】

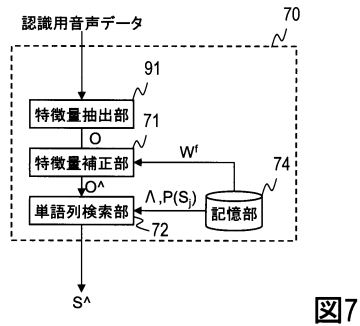


図7

【図 8】

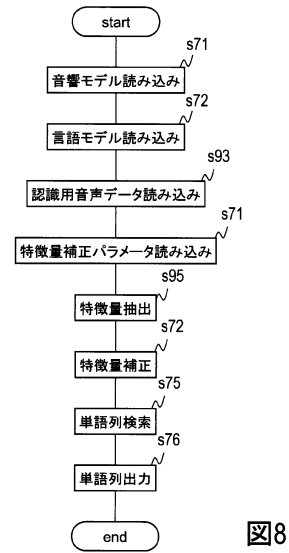


図8

【図 9】

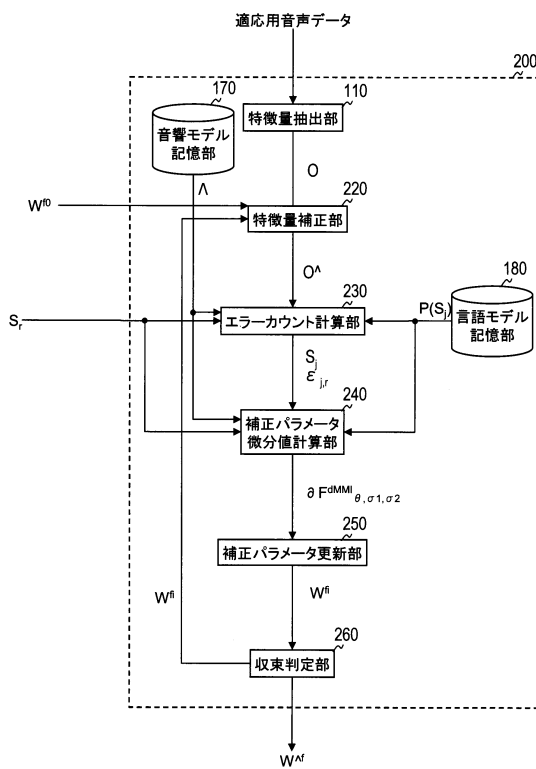


図9

【図 10】

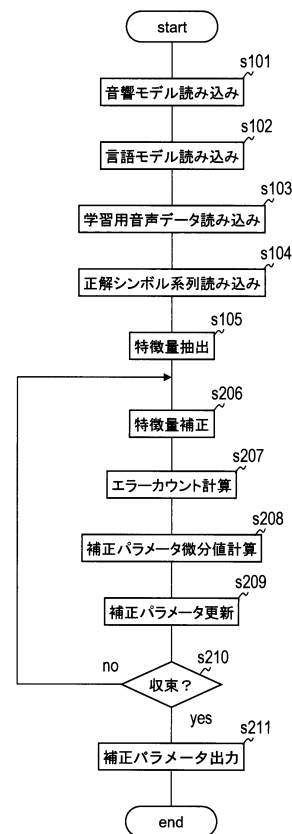


図10

フロントページの続き

(72)発明者 ハム ソンジュン

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

(72)発明者 中谷 智広

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

(72)発明者 中村 篤

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 富澤 直樹

(56)参考文献 特開2002-182682(JP,A)

特開平11-212588(JP,A)

特許第5738216(JP,B2)

特開2011-048262(JP,A)

Marc Delcroix, et al., Discriminative feature transforms using differenced maximum mutual information, ICASSP 2012, 日本, IEEE, 2012年 3月25日, pp. 4753-4756

デルクロア マーク、外4名, dMMI基準による特徴量変換の識別学習, 日本音響学会 2012年 春季研究発表会, 日本, 社団法人日本音響学会, 2012年 3月15日, pp. 121-122

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/34