

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4654452号  
(P4654452)

(45) 発行日 平成23年3月23日(2011.3.23)

(24) 登録日 平成23年1月7日(2011.1.7)

(51) Int.Cl. F I  
**G 1 0 L 15/06 (2006.01)** G 1 0 L 15/06 3 0 0 C  
**G 1 0 L 15/14 (2006.01)** G 1 0 L 15/14 2 0 0 Z

請求項の数 5 (全 21 頁)

<p>(21) 出願番号 特願2005-254424 (P2005-254424)</p> <p>(22) 出願日 平成17年9月2日(2005.9.2)</p> <p>(65) 公開番号 特開2007-65533 (P2007-65533A)</p> <p>(43) 公開日 平成19年3月15日(2007.3.15)</p> <p>審査請求日 平成20年4月2日(2008.4.2)</p> <p>(出願人による申告)平成17年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p>	<p>(73) 特許権者 393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2</p> <p>(74) 代理人 100115749 弁理士 谷川 英和</p> <p>(72) 発明者 松田 繁樹 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 ウォルフガング ヘルボート 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 中村 哲 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p style="text-align: right;">最終頁に続く</p>
---	---

(54) 【発明の名称】音響モデル生成装置、およびプログラム

(57) 【特許請求の範囲】

【請求項1】

確率モデルである音響モデルを格納している音響モデル格納部と、  
 第一の音声から構成された特徴ベクトルについての情報である第一特徴ベクトル情報を格納している第一特徴ベクトル情報格納部と、  
 モデルの構造に関する情報であるモデル構造情報を格納しているモデル構造情報格納部と、  
 第二の音声を受け付ける音声受付部と、  
 前記音響モデルに基づいて、前記第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報を2以上取得する特徴ベクトル情報取得部と、  
 前記音響モデルと、少なくとも前記第一特徴ベクトルを含む評価対象の2以上の特徴ベクトル情報に基づいて、モデルパラメータを算出するモデルパラメータ算出部と、  
 前記特徴ベクトル情報取得部が取得した2以上の第二特徴ベクトル情報と、前記モデル構造情報と前記モデルパラメータを有するモデルとの距離を、前記第二特徴ベクトル情報ごとに算出する距離算出部と、  
 前記距離算出部が算出した距離に基づいて、前記特徴ベクトル情報取得部が取得した2以上の第二特徴ベクトル情報ごとに、当該第二特徴ベクトル情報が正常データであるか異常データであるかを判断する判断部と、  
 前記判断部が最近の処理において正常データであると判断した第二特徴ベクトル情報、および前記第一特徴ベクトル情報を評価対象の特徴ベクトル情報として、前記モデルパラメ

ータ算出部にモデルパラメータの算出を指示し、かつ前記距離算出部、および前記判断部に、前記各部の前記所定の動作を行うように指示する制御部と、前記制御部の指示による繰り返し動作を終了するか否かを判断する終了判断部と、前記終了判断部が繰り返し動作を終了すると判断した場合、前記モデルパラメータ算出部が最後に算出したモデルパラメータを出力する出力部を具備する音響モデル生成装置。

【請求項 2】

前記モデル構造情報は、HMM状態を示す情報であり、前記判断部が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトルである請求項1記載の音響モデル生成装置。

【請求項 3】

前記モデル構造情報は、HMM状態を示す情報であり、前記判断部が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列である請求項1記載の音響モデル生成装置。

【請求項 4】

前記モデル構造情報は、音素HMMを示す情報であり、前記判断部が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列である請求項1記載の音響モデル生成装置。

【請求項 5】

記憶媒体に、  
確率モデルである音響モデルと、第一の音声から構成された特徴ベクトルについての情報である第一特徴ベクトル情報と、モデルの構造に関する情報であるモデル構造情報とを格納しており、

コンピュータに、  
第二の音声を受け付ける音声受付ステップと、  
前記音響モデルに基づいて、前記第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報を2以上取得する特徴ベクトル情報取得ステップと、  
前記特徴ベクトル情報取得ステップで取得した2以上の第二特徴ベクトル情報と、前記モデル構造情報と前記モデルパラメータを有するモデルとの距離を、前記第二特徴ベクトル情報ごとに算出する距離算出ステップと、  
前記距離算出ステップで算出した距離に基づいて、前記特徴ベクトル情報取得ステップで取得した2以上の第二特徴ベクトル情報ごとに、当該第二特徴ベクトル情報が正常データであるか異常データであるかを判断する判断ステップと、  
前記判断ステップで最近の処理において正常データであると判断した第二特徴ベクトル情報、および前記第一特徴ベクトル情報を評価対象の特徴ベクトル情報として、前記モデルパラメータ算出ステップ、前記距離算出ステップ、および前記判断ステップを繰り返し、前記繰り返しの動作を終了するか否かを判断する終了判断ステップと、  
前記終了判断ステップで繰り返しの動作を終了すると判断した場合、前記モデルパラメータ算出ステップで最後に算出したモデルパラメータを出力する出力ステップを実行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、音声認識等に利用する音響モデルを生成する装置等に関するものである。

【背景技術】

【0002】

近年、様々な発話スタイルの音響的な振る舞いの解析、また、これらの変動に頑健な音響モデルや言語モデルを推定するため、「日本語話し言葉コーパス」に代表される大規模な音声データベースの収録が広く行われている。ここでは、高精度な音声認識を実現するためには、クリッピングや突発性雑音を含まない音声波形と、精密な書き起こしテキストを用いて音響モデルを推定することが重要である。

## 【 0 0 0 3 】

しかしながら、データベースの規模が増大するに従い、突発性雑音や誤り記述などの異常データ (Outlier) の数も同様に増加する。実環境において音声を収録する場合、クリッピングや突発性雑音を、完全に抑えることは非常に困難である。さらに、精密な書き起こしテキストを得るための作業は、人手に頼らざるを得ず、高価かつ時間の掛る作業である。

## 【 0 0 0 4 】

書き起こしテキストを持たない大規模音声データベースを音声認識し、その認識結果から信頼度の高い単語を音響モデルの学習に用いる手法 (Unsupervised Training法) が存在する (「非特許文献 1」参照)。また、字幕などの必ずしも正確とは言えないテキストによって適応された言語モデルを用いて音声認識を行ない、信頼度の高い単語を抽出する手法が存在する (「非特許文献 2」参照)。これらの手法は、個々の単語の信頼度を計算することによって、単語毎に音響モデル推定に利用可能であるか判定している。

【非特許文献 1】 F. Wessel and H. Ney, "Unsupervised training of acoustic models for large vocabulary continuous speech recognition," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 1, pp. 23-31, 2005.

【非特許文献 2】 L. Nguyen and B. Xiang, "Light supervision in acoustic model training," in Proc. Eurospeech, vol. 3, pp. 1837-1840, 2003.

## 【 発明の開示 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 5 】

しかしながら、上記の従来技術において、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データに対して、その単語全体が学習データから削除される、さらに、異常データを含む音声データが音響モデルの学習に用いられる、という課題があった。

## 【 課題を解決するための手段 】

## 【 0 0 0 6 】

本第一の発明の音響モデル生成装置は、確率モデルである音響モデルを格納している音響モデル格納部と、第一の音声から構成された特徴ベクトルについての情報である第一特徴ベクトル情報を格納している第一特徴ベクトル情報格納部と、モデルの構造に関する情報であるモデル構造情報を格納しているモデル構造情報格納部と、第二の音声を受け付ける音声受付部と、前記音響モデルに基づいて、前記第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報を 2 以上取得する特徴ベクトル情報取得部と、前記音響モデルと、少なくとも前記第一特徴ベクトルを含む評価対象の 2 以上の特徴ベクトル情報に基づいて、モデルパラメータを算出するモデルパラメータ算出部と、前記特徴ベクトル情報取得部が取得した 2 以上の第二特徴ベクトル情報と、前記モデル構造情報と前記モデルパラメータを有するモデルとの距離を、前記第二特徴ベクトル情報ごとに算出する距離算出部と、前記距離算出部が算出した距離に基づいて、前記特徴ベクトル情報取得部が取得した 2 以上の第二特徴ベクトル情報ごとに、当該第二特徴ベクトル情報が正常データであるか異常データであるかを判断する判断部と、前記判断部が最近の処理において正常データであると判断した第二特徴ベクトル情報、および前記第一特徴ベクトル情報を評価対象の特徴ベクトル情報として、前記モデルパラメータ算出部にモデルパラメータの算出を指示し、かつ前記距離算出部、および前記判断部に、前記各部の前記所定の動作を行うように指示する制御部と、前記制御部の指示による繰り返し動作を終了するか否かを判断する終了判断部と、前記終了判断部が繰り返し動作を終了すると判断した場合、前記モデルパラメータ算出部が最後に算出したモデルパラメータを出力する出力部を具備する音響モデル生成装置である。

かかる構成により、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データのみを除いた音響モデルが生成できる。

## 【 0 0 0 7 】

また、本第二の発明の音響モデル生成装置は、第一の発明に対して、前記モデル構造情報は、HMM状態を示す情報であり、前記判断部が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトルである音響モデル生成装置である。

【0008】

また、本第三の発明の音響モデル生成装置は、第一の発明に対して、前記モデル構造情報は、HMM状態を示す情報であり、前記判断部が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列である音響モデル生成装置である。

【0009】

また、本第四の発明の音響モデル生成装置は、第一の発明に対して、前記モデル構造情報は、音素HMMを示す情報であり、前記判断部が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列である音響モデル生成装置である。

かかる第二から第四の発明の構成により、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データのみを除いた音響モデルが生成できる。

【0010】

また、本第五の発明の音響モデル生成装置は、第一から第四いずれかの発明に対して、前記終了判断部は、予め決められた回数だけ、前記制御部の指示による繰り返し動作を行った場合に、前記繰り返し動作を終了すると判断する音響モデル生成装置である。

かかる構成により、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データのみを除いた音響モデルが生成でき、かつ少ない処理量音響モデルが生成できる。

【0011】

また、本第六の発明の音響モデル生成装置は、第一から第五いずれかの発明に対して、前記判断部は、予め決められた割合に適合するように、当該第二特徴ベクトル情報中の正常データおよび異常データを決定する音響モデル生成装置である。

かかる構成により、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データのみを除いた音響モデルが、精度高く生成できる。

【発明の効果】

【0012】

本発明による音響モデル生成装置によれば、誤り音素記述や突発性雑音など単語中の一部分だけに異常データが存在する場合において、その単語中の一部の異常データのみを学習データから削除できる。さらに、異常データを含む音声データが音響モデルの学習に用いられる可能性を回避できる。

【発明を実施するための最良の形態】

【0013】

以下、音響モデル生成装置等の実施形態について図面を参照して説明する。なお、実施の形態において同じ符号を付した構成要素は同様の動作を行うので、再度の説明を省略する場合がある。

(実施の形態1)

図1は、本実施の形態における音響モデル生成装置のブロック図である。

【0014】

本音響モデル生成装置は、音響モデル格納部101、第一特徴ベクトル情報格納部102、モデル構造情報格納部103、音声受付部104、第一特徴ベクトル情報取得部105、第一特徴ベクトル情報蓄積部106、特徴ベクトル情報取得部107、モデルパラメータ算出部108、距離算出部109、判断部110、制御部111、終了判断部112、出力部113を具備する。音声受付部104は、第一音声受付手段1041、第二音声受付手段1042を具備する。また、本音響モデル生成装置は、音声受付部104に音声を入力するマイク305や、ハードディスク3017を有しても良い。さらに、本音響モデル生成装置は、出力部113の出力対象であるディスプレイ304や、ハードディスク

10

20

30

40

50

3017を有しても良い。なお、音声を入力するハードディスク3017と、出力対象であるハードディスク3017は、同じハードディスクでも、異なるハードディスクでも良い。

【0015】

音響モデル格納部101は、隠れマルコフモデル(以下、適宜「HMM」と言う。)などの確率モデルである音響モデルを格納している。音響モデルは、混合ガウス分布を持つ状態の情報と状態遷移確率の情報から構成されているHMMの情報である。音響モデルは、例えば、一のファイルに格納されているデータである。音響モデル格納部101は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0016】

第一特徴ベクトル情報格納部102は、第一の音声から構成された特徴ベクトルについての情報である第一特徴ベクトル情報を格納している。第一の音声は、音響モデルを新たに生成するために前もって学習しておくための音声データである。第一特徴ベクトル情報は、第一の音声から構成される特徴ベクトル情報の集合であり、以下、適宜、「基本学習データ」とも言う。基本学習データは、通常、信頼できるデータである。「信頼できるデータ」とは、電気的な雑音や突発性雑音を含まない音声ファイル、および精密かつ正確に記述された(単語や音素による)書き起こしテキストを持つ音声データベースのことである。また、第一の音声は、マイクから取得されても良いし、磁気テープやCD等の光ディスクやハードディスク等の記録媒体から読み出されても良い。第一特徴ベクトル情報は、基本学習データをV i t e r b iアラインメントして得られる特徴ベクトルの集合である。第一特徴ベクトル情報は、特徴ベクトルの集合の場合もあるし、特徴ベクトル時系列の集合の場合もある。なお、第一特徴ベクトル情報が特徴ベクトルの集合である場合、後述する正常データか異常データかの判断対象の単位が一の特徴ベクトルである。また、第一特徴ベクトル情報が特徴ベクトル時系列の集合である場合、後述する正常データか異常データかの判断対象の単位が一の特徴ベクトル時系列である。音声は、V i t e r b iアラインメントにより、1以上の特徴ベクトル、および1以上の特徴ベクトル時系列が取得され得る。V i t e r b iアラインメントは、公知の技術であるので説明を省略する。また、特徴ベクトルは、例えば、三角型フィルタを用いたチャンネル数24のフィルタバンク出力を離散コサイン変換したMFCCであり、その静的パラメータ、デルタパラメータおよびデルタデルタパラメータをそれぞれ12次元、さらに正規化されたパワーとデルタパワーおよびデルタデルタパワー(39次元)を有する。また、スペクトル分析において、ケプストラム平均除去を施すことは好適である。また、特徴ベクトルは、上記に限られず、例えば、フィルタバンクのチャンネル数は20で、12MFCCsとデルタパラメータ、デルタパワーの計25次元等でも良い。第一特徴ベクトル情報格納部102は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0017】

モデル構造情報格納部103は、モデルの構造に関する情報であるモデル構造情報を格納している。モデルの構造は、ここでは、「HMM状態」、または「音素HMM」である。モデル構造情報は、モデルの構造を示す情報であり、例えば、文字列「HMM状態」、または文字列「音素HMM」である。また、モデル構造情報は、HMM状態を示す識別子(例えば、「1」)、音素HMMを示す識別子(例えば、「0」)などでも良い。モデル構造情報は、例えば、本音響モデル生成装置を使用するユーザが、本音響モデル生成装置に対して与える。なお、モデル構造情報の与えられ方は問わない。モデル構造情報格納部103は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0018】

音声受付部104は、ここでは、第一の音声、および第二の音声を受け付ける。ただし、音声受付部104は、第二の音声のみを受け付けても良い。第二の音声は、音響モデルを新たに生成する元になるデータである。本音響モデル生成装置は、第二の音声から上手く異常データを取り除いて、新たに音響モデルを生成する装置である。第二の音声等の入力手段は、マイクや記録媒体からの読み出し等、何でも良い。音声受付部104は、マイ

10

20

30

40

50

ク等の入力手段のデバイスドライバーや、記録媒体からのデータ読み出し制御のソフトウェア等で実現され得る。なお、音声受付部104を構成する第一音声受付手段1041は、第一の音声を受け付ける。第二音声受付手段1042は、第二の音声を受け付ける。

【0019】

第一特徴ベクトル情報取得部105は、音響モデルに基づいて、第一音声受付手段1041が受け付けた第一の音声をV i t e r b iアラインメントして、第一特徴ベクトル情報の集合を取得する。第一特徴ベクトル情報取得部105は、通常、MPUやメモリ等から実現され得る。第一特徴ベクトル情報取得部105の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

10

【0020】

第一特徴ベクトル情報蓄積部106は、第一特徴ベクトル情報取得部105が取得した第一特徴ベクトル情報を第一特徴ベクトル情報格納部102に、少なくとも一時的に蓄積する。第一特徴ベクトル情報蓄積部106は、通常、MPUやメモリ等から実現され得る。第一特徴ベクトル情報蓄積部106の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

【0021】

特徴ベクトル情報取得部107は、音響モデル格納部101の音響モデルに基づいて、第二音声受付手段1042が受け付けた第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報を2以上取得する。特徴ベクトル情報取得部107は、第一特徴ベクトル情報取得部105と同様、第二の音声をV i t e r b iアラインメントして、第二特徴ベクトル情報を取得する。第二特徴ベクトル情報は、特徴ベクトルの集合または、特徴ベクトル時系列の集合である。なお、第二特徴ベクトル情報が特徴ベクトルの集合である場合、後述する正常データか異常データかの判断対象の単位が一の特徴ベクトルである。また、第二特徴ベクトル情報が特徴ベクトル時系列の集合である場合、後述する正常データか異常データかの判断対象の単位が一の特徴ベクトル時系列である。第二特徴ベクトル情報は、以下、「追加学習データ」と、適宜、言う。特徴ベクトル情報取得部107は、通常、MPUやメモリ等から実現され得る。特徴ベクトル情報取得部107の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

20

30

【0022】

モデルパラメータ算出部108は、音響モデル格納部101の音響モデルと、少なくとも第一特徴ベクトルを含む評価対象の2以上の特徴ベクトル情報に基づいて、モデルパラメータを算出する。モデルパラメータ算出部108は、例えば、HMMに対してEM法を適用したアルゴリズムであるBWアルゴリズムにより、モデルパラメータを算出する。なお、ページアンネットのモデルパラメータをモデルパラメータ算出部108で算出しても良い。EM法、およびBWアルゴリズム、ページアンネットは公知技術であるので説明は省略する。また、モデルパラメータとは、学習データの統計的な性質を表すパラメータ(例えば、平均や分散、状態遷移確率など)の集合である。さらに、モデルパラメータのデータ構造は、HMMである。モデルパラメータ算出部108は、通常、MPUやメモリ等から実現され得る。モデルパラメータ算出部108の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

40

【0023】

距離算出部109は、特徴ベクトル情報取得部107が取得した2以上の第二特徴ベクトル情報と、モデルとの距離を、第二特徴ベクトル情報ごとに算出する。ここで、モデルとは、モデル構造情報格納部103のモデル構造情報モデルパラメータ算出部108が算出したモデルパラメータを有する情報のことである。また、モデル構造情報は、例えば、「HMM状態」や「音素HMM」などである。また、ここで、距離とは、例えば、モデル

50

中に第二特徴ベクトル情報が現れる確率をパラメータとする減少関数の演算結果である。距離算出部109は、例えば、フォワード法により距離を算出する。フォワード法は、公知技術であるので説明を省略する。距離算出部109は、通常、MPUやメモリ等から実現され得る。距離算出部109の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

#### 【0024】

判断部110は、距離算出部109が算出した距離に基づいて、特徴ベクトル情報取得部107が取得した2以上の第二特徴ベクトル情報ごとに、当該第二特徴ベクトル情報が正常データであるか異常データであるかを判断する。判断部110は、例えば、予め決められた割合に適合するように、当該第二特徴ベクトル情報中の正常データおよび異常データを決定する。この予め決められた割合を、以下、適宜「削減率」という。削減率は、全体の第二特徴ベクトル情報中に、異常データが占める割合である。判断部110は、距離算出部109が算出した距離が、予め決められた閾値以上である場合に、当該特徴ベクトル情報を異常データと判断し、閾値未満の場合に、当該特徴ベクトル情報を正常データと判断しても良い。なお、閾値を示すデータは、予め判断部110が保持している。判断部110は、通常、MPUやメモリ等から実現され得る。判断部110の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

#### 【0025】

制御部111は、判断部110が最近の処理において正常データであると判断した第二特徴ベクトル情報、および第一特徴ベクトル情報を評価対象の特徴ベクトル情報として、モデルパラメータ算出部108にモデルパラメータの算出を指示し、かつ距離算出部109、および判断部110に、各部の所定の動作を行うように指示する。各部の所定の動作とは、距離算出部109の上述した動作、および判断部110の上述した動作である。制御部111および後述する終了判断部112の処理により、モデルパラメータ算出部108、距離算出部109、および判断部110の動作は繰り返し行われる。かかる繰り返しの動作を「繰り返し動作」という。繰り返し動作は、終了判断部112が、繰り返し動作を終了すると判断するまで繰り返される。制御部111は、かかる繰り返し動作を司る。つまり、制御部111は、ループ処理を行う。なお、制御部111は、本音響モデル生成装置がループ処理を行えば、必ずしも、明示的にモデルパラメータ算出部108等に指示する(メッセージ送信など)を行う必要はない。また、最近の処理とは、繰り返し動作において、現時点での最終の処理のことである。また、評価対象の特徴ベクトル情報は、繰り返し動作でループする間、通常、変化する。ループする間、通常、第二特徴ベクトル情報が変化するからである。制御部111は、通常、MPUやメモリ等から実現され得る。制御部111の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

#### 【0026】

終了判断部112は、制御部111の指示による繰り返し動作を終了するか否かを判断する。終了判断部112は、例えば、予め決められた回数(例えば、10回)だけ、制御部111の指示による繰り返し動作を行った場合に、繰り返し動作を終了すると判断する。また、終了判断部112は、最近に算出されたモデルパラメータと、一つ前に算出されたモデルパラメータとの差が所定の差以内(例えば、「0」)である場合に、繰り返し動作を終了すると判断しても良い。最近に算出されたモデルパラメータと、一つ前に算出されたモデルパラメータとの差が0である場合、モデルパラメータが動いておらず、収束したことを示す。終了判断部112は、通常、MPUやメモリ等から実現され得る。終了判断部112の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

#### 【0027】

出力部113は、終了判断部112が繰り返し動作を終了すると判断した場合、モデル

パラメータ算出部 108 が最後に算出したモデルパラメータを出力する。出力とは、記録媒体への蓄積、ディスプレイへの表示、プリンタへの印字、音出力、外部の装置への送信等を含む概念である。出力部 113 は、ディスプレイや記録媒体等のデバイスを含むと考へても含まないと考へても良い。出力部 113 は、デバイスのドライバーソフトまたは、デバイスのドライバーソフトとデバイス等で実現され得る。

次に、音響モデル生成装置の動作について図 2 のフローチャートを用いて説明する。

(ステップ S201) 第一特徴ベクトル情報取得部 105 は、音響モデル格納部 101 に格納されている HMM の音響モデルを読み込む。

(ステップ S202) 第一音声受付手段 1041 は、第一の音声を受け付けたか否かを判断する。第一の音声を受け付ければステップ S203 に行き、第一の音声を受け付けなければステップ S202 に戻る。

【0028】

(ステップ S203) 第一特徴ベクトル情報取得部 105 は、ステップ S201 で読み込んだ音響モデルに基づいて、ステップ S202 で受け付けた第一の音声を V i t e r b i アラインメントして、第一特徴ベクトル情報の集合を取得する。

【0029】

(ステップ S204) 第一特徴ベクトル情報取得部 105 は、ステップ S203 で取得した第一特徴ベクトル情報の集合を第一特徴ベクトル情報格納部 102 に蓄積する。この蓄積は、一時的な蓄積でも良いことは言うまでもない。

【0030】

(ステップ S205) 第二音声受付手段 1042 は、第二の音声を受け付けたか否かを判断する。第二の音声を受け付ければステップ S206 に行き、第二の音声を受け付けなければステップ S205 に戻る。なお、例えば、第一の音声、および第二の音声は、講演で講演者から発声される音声であり、第一の音声は、講演の前半部の音声、第二の音声は、講演の後半部の音声である。

【0031】

(ステップ S206) 特徴ベクトル情報取得部 107 は、ステップ S201 で読み込んだ音響モデルに基づいて、ステップ S205 で受け付けた第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報の集合を取得する。

(ステップ S207) 特徴ベクトル情報取得部 107 は、ステップ S206 で取得した第二特徴ベクトル情報の集合を図示しない手段に蓄積する。この蓄積は、一時的な蓄積でも良いことは言うまでもない。

【0032】

(ステップ S208) モデルパラメータ算出部 108 は、評価対象の特徴ベクトル情報の集合を取得する。ここで、取得する評価対象の特徴ベクトル情報の集合は、例えば、第一特徴ベクトルの集合のみである。また、評価対象の特徴ベクトル情報の集合は、例えば、第一特徴ベクトルの集合と、第二特徴ベクトルの集合でも良い。

(ステップ S209) モデルパラメータ算出部 108 は、音響モデル格納部 101 の音響モデルと、ステップ S208 で取得した評価対象の 2 以上の特徴ベクトル情報に基づいて、モデルパラメータを算出する。

(ステップ S210) 距離算出部 109 は、カウンタ  $i$  に 1 を代入する。

【0033】

(ステップ S211) 距離算出部 109 は、 $i$  番目の第二特徴ベクトル情報が存在するか否かを判断する。 $i$  番目の第二特徴ベクトル情報が存在すればステップ S212 に行き、 $i$  番目の第二特徴ベクトル情報が存在しなければステップ S215 に行く。

(ステップ S212) 距離算出部 109 は、 $i$  番目の第二特徴ベクトル情報と、モデル構造情報格納部 103 のモデル構造情報とステップ S209 で取得したモデルパラメータを有するモデルとの距離を算出する。

(ステップ S213) 距離算出部 109 は、 $i$  番目の第二特徴ベクトル情報に対応付けて、ステップ S212 で算出した距離を一時蓄積する。

10

20

30

40

50



(ステップS214) 距離算出部109は、カウンタ*i*を1、インクリメントし、ステップS211に戻る。

(ステップS215) 判断部110は、距離をキーとして、第二特徴ベクトル情報群の中で、距離の近い方から、例えば、99.5%を選択する。

【0034】

(ステップS216) 判断部110は、例えば、ステップS215で選択した上位99.5%を正常データの第二特徴ベクトル情報とし、下位0.5%を異常データの第二特徴ベクトル情報とする。そして、判断部110は、各第二特徴ベクトル情報に対応付けて、正常データか異常データかを示すラベルを付与する。ラベルは、例えば、正常データが「1」、異常データが「0」である。また、ソートした上位99.5%を正常データの第二特徴ベクトル情報とし、下位0.5%を異常データの第二特徴ベクトル情報とする場合、削減率が「0.5%」である、とする。

【0035】

(ステップS217) 終了判断部112は、処理を終了するか否かを判断する。終了判断部112は、例えば、モデルパラメータ算出部108が、ループにより、10回、モデルパラメータを算出した場合に、処理を終了する、と判断する。また、終了判断部112は、例えば、ステップS217を5回通過した後、6回目の通過の時点で、処理を終了する、と判断しても良い。また、終了判断部112は、例えば、最後にステップS209で算出したモデルパラメータと、その前のループにより算出したモデルパラメータとを比較し、同じ場合に処理を終了する、と判断しても良い。さらに、終了判断部112は、例えば、最後にステップS209で算出したモデルパラメータと、その前のループにより算出したモデルパラメータとを比較し、所定以内の差である場合に処理を終了する、と判断しても良い。ここで、「所定以内の差」とは、一般的には、個々の学習データに対する確率の積(学習データに対するモデルの尤度)の差である。直前に推定されたモデルパラメータを $H(n-1)$ 、新たに推定されたモデルパラメータを $H(n)$ とした場合、各々の尤度は以下の様に計算される。そして、これらの尤度の差( $L(n) - L(n-1)$ )が、予め設定した値以下の場合に、推定処理を終了する。なお、「Prod」は、「product」を意味する。

【数1】

$$L(n-1) = \text{Prod } P(x | H(n-1))$$

$$L(n) = \text{Prod } P(x | H(n))$$

そして、終了判断部112が処理を終了すると判断した場合、ステップS219に行き、終了判断部112が処理を終了しないと判断した場合、ステップS218に行く。

【0036】

(ステップS218) モデルパラメータ算出部108は、評価対象の特徴ベクトル情報の集合を取得する。評価対象の特徴ベクトル情報の集合は、ここでは、正常データのラベルが付された第二特徴ベクトル情報と、第一特徴ベクトル情報である。本ステップの処理終了後、ステップS209に戻る。

【0037】

(ステップS219) 出力部113は、モデルパラメータ算出部108が最後に算出したモデルパラメータを出力する。このモデルパラメータは、ステップS209で算出した最後のモデルパラメータである。本ステップの処理終了後、処理を終了する。

【0038】

なお、図2のフローチャートにおいて、第一特徴ベクトル情報の集合は、予め第一特徴ベクトル情報格納部102に格納されていても良い。かかる場合、ステップS202、S203、S204の処理は、不要である。また、かかる場合、第一特徴ベクトル情報の集合を読み込む処理は、別途、必要である。

【0039】

以下、本実施の形態における音響モデル生成装置の具体的な動作について説明する。本

10

20

30

40

50

音響モデル生成装置は、最尤法を基礎とした異常データ検出手法によって、受け付けた音声から異常データを除去し、正常データのみを採用し、当該正常データから音響モデルを生成する装置である。本音響モデル生成装置は、任意の確率モデル $\mu$ と、長さ $T_i$ の特徴ベクトル時系列 $X_i = (x_i, 1, \dots, x_i, T_i)$ から構成された追加学習データ $W$ に対して適用することができる。また、追加学習データは特徴ベクトル時系列の集合である。

【0040】

今、音響モデル格納部101の音響モデルは、混合ガウス分布を持つ状態と状態遷移確率から構成されている。音響モデルは、隠れマルコフモデル(HMM)、混合ガウスモデル(GMM)、ページアンネットなど、任意の確率モデルである。

10

また、ここで、講演者が、講演を開始する、とする。

まず、本音響モデル生成装置の第一特徴ベクトル情報取得部105は、音響モデル格納部101に格納されているHMMの音響モデルを読み込む。

【0041】

そして、本音響モデル生成装置は、講演者の講演を、例えば、30分間、受け付け、本音響モデル生成装置の第一特徴ベクトル情報取得部105が、読み込んだ音響モデルに基づいて、当該30分間の音声(第一の音声)をViterbiアラインメントし、第一特徴ベクトル情報を得る。第一特徴ベクトル情報(音響特徴量の集合)の例を図3に示す。図3は、講演者が「はい、そうです」と発声した場合の、音響特徴量を示す。図3の音響特徴量は、1stMFCC、2ndMFCC、3rdMFCC、4thMFCCの4つのMFCCである。また、図3において、横軸は発話開始からの絶対時間(ms)、縦軸は各音響特徴量の値である。そして、第一特徴ベクトル情報取得部105は、図3に示す第一特徴ベクトル情報の集合を第一特徴ベクトル情報格納部102に蓄積する。なお、第一の音声は、基本学習データである。

20

【0042】

次に、講演開始から30分以降、60分までの間、第二音声受付手段1042は、音声を受け付ける。かかる音声は第二の音声である。なお、ここでは、第一音声受付手段1041と第二音声受付手段1042は、通常、物理的に一の手段である。

【0043】

次に、特徴ベクトル情報取得部107は、読み込んだ音響モデルに基づいて、受け付けた第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報の集合を取得する。この第二特徴ベクトル情報の集合も、図3に示すのと同様の構造を有するデータである。特徴ベクトル情報取得部107は、第二特徴ベクトル情報の集合を蓄積する。

30

【0044】

次に、モデルパラメータ算出部108は、評価対象の特徴ベクトル情報の集合(ここでは、第一特徴ベクトル情報の集合)を取得する。そして、モデルパラメータ算出部108は、BWアルゴリズムにより、第一特徴ベクトル情報の集合からモデルパラメータ( $\theta_1$ )を算出する。算出したモデルパラメータと状態の関係を表した図の例を、図4に示す。図4において、「s」は状態、「a」は状態遷移確率、「b」は状態出力確率分布である。つまり、図4において、「 $S_1$ 」「 $S_2$ 」「 $S_3$ 」は状態、「 $a_{01}$ 」「 $a_{11}$ 」「 $a_{12}$ 」「 $a_{22}$ 」「 $a_{23}$ 」「 $a_{34}$ 」は状態遷移確率、「 $b_1$ 」「 $b_2$ 」「 $b_3$ 」は状態出力確率分布である。モデルパラメータ( $\theta_1$ )は、図4において、「 $a_{01}$ 」「 $a_{11}$ 」「 $a_{12}$ 」「 $a_{22}$ 」「 $a_{23}$ 」「 $a_{33}$ 」「 $a_{34}$ 」、および「 $b_1$ 」「 $b_2$ 」「 $b_3$ 」の集合である。なお、状態出力確率は、例えば、混合ガウス分布である。また、 $\theta_1$ は、最初に算出されたモデルパラメータである。

40

【0045】

次に、距離算出部109は、モデルパラメータ算出部108が算出したモデルパラメータ( $\theta_1$ )とモデル構造情報格納部103のモデル構造を有するモデルに対する、各第二特徴ベクトル情報の出現確率( $d(k)_1$ )を算出する。ここで、第二特徴ベクトル情報は、特徴ベクトル時系列である、とする。

50

## 【 0 0 4 6 】

$d(k)_1$  は、数式 2 により算出される。また、第二特徴ベクトル情報は、特徴ベクトル時系列（例えば、特徴ベクトル時系列を「 $X_i = (x_{i,1}, \dots, x_{i,T_i})$ 」とする）であるので、数式 3 に示すように、第二特徴ベクトル情報ごとに、最大の確率の積が得られる学習データ  $U$  を選択し、かかる学習データ  $U$  の最大の確率の積を出現確率とする。

## 【数 2】

$$d(k)_1 := P((x_{k,1}, x_{k,2}, \dots, x_{k,T_k}) | \theta_1).$$

## 【数 3】

$$h = \sum_{i \in UCW} T_i$$

$$\hat{U} = \operatorname{argmax}_U \prod_{i \in UCW} P(\theta | X_i)$$

10

## 【 0 0 4 7 】

ここで、 $P(x_{k,1}, x_{k,2}, \dots, x_{k,T_k})$  は、モデル ( $\theta_1$ ) に対する特徴ベクトル時系列の確率を意味する。また、 $h$  は、特徴ベクトルの個数である。なお、最大の確率の積が得られる学習データ  $U$  を選択する問題は、Knapsack 問題と同じであり、動的計画法により効率的に解くことができる。数式 2 の第 2 番目の式の左辺 ( $U$  ハット) が、最大の確率の積が得られる学習データである。なお、動的計画法は、公知技術であるので説明を省略する。

20

そして、距離算出部 109 は、例えば、出現確率 ( $d(k)_1$ ) を距離として、算出する。

## 【 0 0 4 8 】

また、距離算出部 109 が出現確率を算出する場合の概念図を図 5 に示す。図 5 において、特徴ベクトル時系列 (Feature Vector Sequences) が長さ (Length) 「 $L$ 」の特徴ベクトルを有する場合 (特徴ベクトル時系列により長さは異なり得る)、上述した動的計画法により、確率の積が最大となるように特徴ベクトル時系列を選択し、当該最大の積を出現確率とする。つまり、図 5 は、数式 1 の説明となる。図 5 において、各特徴ベクトル時系列の出現確率「Probability」は、Knapsack 問題の解法である動的計画法により算出することを示す。

30

そして、距離算出部 109 は、算出した距離と、各第二特徴ベクトル情報を対応づけて蓄積する。かかる距離算出等の処理を、全第二特徴ベクトル情報に対して行う。

そして、ここで、削減率「0.5%」である、とする。そして、判断部 110 は、距離をキーとして、距離が小さい 99.5% の第二特徴ベクトル情報を選択する。

## 【 0 0 4 9 】

そして、判断部 110 は、例えば、選択した上位 99.5% を正常データの第二特徴ベクトル情報とし、下位 0.5% を異常データの第二特徴ベクトル情報として、正常データには「1」を、異常データには「0」のデータ (ラベル) を付加する。

## 【 0 0 5 0 】

次に、終了判断部 112 は、処理を終了するか否かを判断する。ここで、処理終了の条件は、例えば、処理終了の判断の回数が 10 回まで、とする。つまり、ここでは、処理終了の判断は、1 回目の判断であるので、処理を終了しない。なお、終了判断部 112 は、処理終了の判断を 1 回行うごとに、カウンタを 1、インクリメントする (カウンタの初期値は、1 である)。

40

## 【 0 0 5 1 】

次に、モデルパラメータ算出部 108 は、第二特徴ベクトル情報の集合の中で、正常データには「1」のラベルが付された第二特徴ベクトル情報と、第一特徴ベクトル情報の集合を、次の評価対象の特徴ベクトル情報の集合とし、取得する。

次に、制御部 111 は、モデルパラメータ算出部 108 にモデルパラメータの算出を指

50

示する。

【0052】

次に、モデルパラメータ算出部108は、音響モデル格納部101の音響モデルと、先に取得した評価対象の特徴ベクトル情報の集合（正常データには「1」のラベルが付された第二特徴ベクトル情報と、第一特徴ベクトル情報の集合）に基づいて、上記と同様に、BWアルゴリズムにより、モデルパラメータ（ $\theta_2$ ）を算出する。

【0053】

次に、制御部111は、距離算出部109、および判断部110に、各部の所定の動作を行うように指示する。なお、かかる指示は、実際に指示しなくとも（メッセージ送信など行わなくとも）、距離算出部109、および判断部110が所定の処理を行えば、制御部111が指示をしたと考えることとする。例えば、実際に指示しない場合、制御部111は、モデルパラメータ算出部108、距離算出部109、および判断部110の一連の処理をループさせる制御等を行い、指示するのと同様の動作を結果として行うからである。

そして、10回モデルパラメータが算出され、最終的にモデルパラメータ（ $\theta_{10}$ ）が得られる。

【0054】

そして、終了判断部112は、処理を終了すると、判断する。出力部113は、モデルパラメータ算出部108が最後に算出したモデルパラメータ（ $\theta_{10}$ ）を、例えば、予め決められたファイルに出力する。モデルパラメータ（ $\theta_{10}$ ）のデータ例は、例えば、図4である。

【0055】

以下、本音響モデル生成装置が生成した音響モデルを用いて音声認識性能を確認する実験について述べる。この実験は、日本語大語彙連続音声認識実験である。本実験に使用した音声認識エンジンは、本出願人が開発した「ATRASSR Ver 3.3」という公知の音声認識エンジンである。

【0056】

まず、以下の実験は、3つの手法で行った。3つの手法とは、手法A、手法B、手法Cである。図6に、本音響モデル生成装置が行う3手法による処理であり、異常データに頑健な音響モデル推定法の処理の概念図を示す。手法A（Feature vectors on HMM state）は、モデル構造情報は、HMM状態を示す情報であり、判断部110が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトルである場合である。手法Aの場合、第一特徴ベクトル情報も、特徴ベクトルである。また、手法B（Feature vector sequences on HMM state）は、モデル構造情報（Strage構造の情報）は、HMM状態を示す情報であり、判断部110が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報（検出単位）は、特徴ベクトル時系列である場合である。手法Bの場合、第一特徴ベクトル情報も、特徴ベクトル時系列である。手法Bは、状態を学習するために用いられる特徴ベクトル時系列の集合に対して異常データに頑健な推定が行われる。

さらに、手法C（Feature vector sequences on phoneme HMM）は、音素HMM（Phoneme HMM（HMM状態系列））を学習するために用いられる特徴ベクトル時系列の集合に対して異常データに頑健な推定が行われる手法である。手法Cにおいて、モデル構造情報は、音素HMMを示す情報であり、判断部110が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列である場合である。手法Cの場合、第一特徴ベクトル情報も、特徴ベクトル時系列である。なお、特徴ベクトルの集合は、学習データをViterbiアラインメントすることによって得られることは上述した通りである。

【0057】

また、図6において、「Strage」とは、学習データの集合Wのことである。図6

10

20

30

40

50

において、「S t r a g e」の中で繋がっていないノードは特徴ベクトルであり、繋がっているノード群は、特徴ベクトル時系列である。手法Aにおいて、モデル構造(S t r a g e構造)は、HMM状態であり、判断部110が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報、および第一特徴ベクトル情報(「検出単位」とも言う。)は、特徴ベクトルである。ここでは、学習データはHMM状態毎にV i t e r b iアラインメントを行う、とする。従って、推定中に状態アラインメントが変化することは無い。

#### 【0058】

次に、本実験において、言語モデルの学習には、旅行会話基本表現集B T E C及び、自然発話音声、自然発話音声・言語データベースS D B、S L D B、L D Bに含まれる6. 1M単語を用いた。第1パスは多重クラス複合2 - g r a mを使用し、第2パスでは単語3 - g r a mを用いた。辞書サイズは34kである。音響モデル学習用音声データは、A T R旅行対話データベースである。407名が発話した対話及び、音素バランス503文(計26380発話)である。評価用音声データとして、A T R旅行会話基本表現集B T E C t e s t s e t - 01(510文、男性4名、女性6名、各々51文の発声データ)を用いた。

#### 【0059】

また、本実験において、誤り音素記述に対する頑健性を確認するため、人工的に生成した誤り音素記述を含む音声データベースを用いて推定された音響モデルの音声認識性能を評価した。モデル推定に用いた学習データの半分に、10%の誤り音素記述が含まれている。この誤り記述は、乱数を用いて生成した。残りの半分のデータを用い、M L - S S S法により1400状態のH M n e t構造を生成した。各状態の混合数は5である。異常データの検出は、誤り記述を含む音声データベースからのみ行った。他方の誤り記述を含まない音声データベースからは異常データの検出は行わず、無条件でモデル推定に用いた。図7に学習データの削減率に対する、誤り音素記述の発見率を示す。図7に示すように、手法Bは、最も高い発見率を得た。図8に、単語誤り率を示す。手法A及び、手法Bはデータ削減率2%、手法Cはデータ削減率4%において、各々、最も低い単語誤り率が得られた。誤り記述の発見率と同様、手法Bは、最も低い単語誤り率を得た。さらに、誤り記述を取り除いた学習データを用いて推定された音響モデル(O r a c l e)に近い認識性能が得られた。これらの結果から、たとえ音声データベースが誤り記述を含んでいたとしても、本音響モデル生成装置の手法を用いることによって精密な音響モデルを推定することが可能であることが分かる。なお、「O r a c l e」とは、理論的な限界性能を意味する。つまり、人工的に生成された誤り音素記述(Outlier)を含まない学習データを用いて推定された音響モデルの認識性能である。

#### 【0060】

次に、未知の異常データに対する頑健性を確認するため、全ての音響モデル学習用データに対して本音響モデル生成装置の手法を適用し、その音声認識性能の評価を行った。本実験で用いた音声データベースは、上記の実験で用いられたような人工的な誤り音素記述は含まれていない。状態共有構造として、全ての学習データを用いて生成した1400状態のH M n e tを用いた。各状態は、前の実験と同様、5混合である。図9に単語誤り率を示す。図9のように、全ての手法において、適切なデータ削減率を設定することによって、従来のM L推定よりも低い単語誤り率が得られた。M L推定によって学習された音響モデルは7.19%の単語誤り率であるのに対して、データ削減率0.5%の手法Cは6.88%であり、4.31%の誤り削減率が得られた。

#### 【0061】

以上、本実施の形態によれば、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データに対して、その単語中の一部の異常データのみを学習データから削除できる。さらに、異常データを含む音声データが音響モデルの学習に用いられる可能性を回避できる。特に、上述の手法Bによって推定された音響モデルは、誤り音素記述を除いて学習した音響モデルに近い単語正解精度が得られた。つまり、特に手法B(モデル構造情

10

20

30

40

50

報は、HMM状態を示す情報であり、判断部110が正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列である場合は、有効である。

なお、本実施の形態によれば、モデル構造情報、および第二特徴ベクトル情報の組み合わせは、上記の手法A、B、Cにおける組み合わせのいずれでも良い。

#### 【0062】

さらに、本実施の形態における処理は、ソフトウェアで実現しても良い。そして、このソフトウェアをソフトウェアダウンロード等により配布しても良い。また、このソフトウェアをCD-ROMなどの記録媒体に記録して流布しても良い。なお、このことは、本明細書における他の実施の形態においても該当する。なお、本実施の形態における音響モデル生成装置を実現するソフトウェアは、以下のようなプログラムである。つまり、このプログラムは、コンピュータに、音響モデルを格納しており、第一の音声から構成された特徴ベクトルについての情報である第一特徴ベクトル情報を格納しており、モデルの構造に関する情報であるモデル構造情報を格納しており、第二の音声を受け付ける音声受付ステップと、格納している音響モデルに基づいて、前記第二の音声から特徴ベクトルについての情報である第二特徴ベクトル情報を2以上取得する特徴ベクトル情報取得ステップと、前記特徴ベクトル情報取得ステップで取得した2以上の第二特徴ベクトル情報と、格納しているモデル構造情報と前記モデルパラメータを有するモデルとの距離を、前記第二特徴ベクトル情報ごとに算出する距離算出ステップと、前記距離算出ステップで算出した距離に基づいて、前記特徴ベクトル情報取得ステップで取得した2以上の第二特徴ベクトル情報ごとに、当該第二特徴ベクトル情報が正常データであるか異常データであるかを判断する判断ステップと、前記判断ステップで最近の処理において正常データであると判断した第二特徴ベクトル情報、および前記第一特徴ベクトル情報を評価対象の特徴ベクトル情報として、前記モデルパラメータ算出ステップ、前記距離算出ステップ、および前記判断ステップの動作を繰り返し、前記繰り返しの動作を終了するか否かを判断する終了判断ステップと、前記終了判断ステップで繰り返しの動作を終了すると判断した場合、前記モデルパラメータ算出ステップで最後に算出したモデルパラメータを出力する出力ステップを実行させるためのプログラム、である。

#### 【0063】

上記プログラムにおいて、前記モデル構造情報は、HMM状態を示す情報であり、前記判断ステップで正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトルであっても良い。

#### 【0064】

上記プログラムにおいて、前記モデル構造情報は、HMM状態を示す情報であり、前記判断ステップで正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列であっても良い。

#### 【0065】

上記プログラムにおいて、前記モデル構造情報は、音素HMMを示す情報であり、前記判断ステップで正常データであるか異常データであるかを判断する対象の第二特徴ベクトル情報は、特徴ベクトル時系列であっても良い。

前記終了判断ステップにおいて、予め決められた回数だけ、前記繰り返し動作を行った場合に、前記繰り返し動作を終了すると判断することは好適である。

前記判断ステップにおいて、予め決められた割合に適合するように、当該第二特徴ベクトル情報中の正常データおよび異常データを決定することは好適である。

#### 【0066】

また、上記各実施の形態において、各処理（各機能）は、単一の装置（システム）によって集中処理されることによって実現されてもよく、あるいは、複数の装置によって分散処理されることによって実現されてもよい。

#### 【0067】

また、図10は、本明細書で述べたプログラムを実行して、上述した種々の実施の形態

10

20

30

40

50

の音響モデル生成装置を実現するコンピュータの外観を示す。上述の実施の形態は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムで実現され得る。図10は、このコンピュータシステム300の概観図であり、図11は、システム300のブロック図である。

【0068】

図10において、コンピュータシステム300は、FD(Flexible Disk)ドライブ、CD-ROM(Compact Disk Read Only Memory)ドライブを含むコンピュータ301と、キーボード302と、マウス303と、モニタ304と、マイク305とを含む。

【0069】

図11において、コンピュータ301は、FDドライブ3011、CD-ROMドライブ3012に加えて、CPU(Central Processing Unit)3013と、CPU3013、CD-ROMドライブ3012及びFDドライブ3011に接続されたバス3014と、ブートアッププログラム等のプログラムを記憶するためのROM(Read-Only Memory)3015と、CPU3013に接続され、アプリケーションプログラムの命令を一時的に記憶するとともに一時記憶空間を提供するためのRAM(Random Access Memory)3016と、アプリケーションプログラム、システムプログラム、及びデータを記憶するためのハードディスク3017とを含む。ここでは、図示しないが、コンピュータ301は、さらに、LANへの接続を提供するネットワークカードを含んでも良い。

【0070】

コンピュータシステム300に、上述した実施の形態の音響モデル生成装置の機能を実行させるプログラムは、CD-ROM3101、またはFD3102に記憶されて、CD-ROMドライブ3012またはFDドライブ3011に挿入され、さらにハードディスク3017に転送されても良い。これに代えて、プログラムは、図示しないネットワークを介してコンピュータ301に送信され、ハードディスク3017に記憶されても良い。プログラムは実行の際にRAM3016にロードされる。プログラムは、CD-ROM3101、FD3102またはネットワークから直接、ロードされても良い。

【0071】

プログラムは、コンピュータ301に、上述した実施の形態の音響モデル生成装置の機能を実行させるオペレーティングシステム(OS)、またはサードパーティープログラム等は、必ずしも含まなくても良い。プログラムは、制御された態様で適切な機能(モジュール)を呼び出し、所望の結果が得られるようにする命令の部分のみを含んでいれば良い。コンピュータシステム300がどのように動作するかは周知であり、詳細な説明は省略する。

また、上記プログラムを実行するコンピュータは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、あるいは分散処理を行ってもよい。

本発明は、以上の実施の形態に限定されることなく、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

【産業上の利用可能性】

【0072】

以上のように、本発明にかかる音響モデル生成装置等は、誤り音素記述や突発性雑音など単語中の一部分だけに存在する異常データに対して、その単語中の一部の異常データのみを学習データから削除できる、という効果を有し、音声認識装置等に利用できる音響モデルを生成する音響モデル生成装置等として有用である。

【図面の簡単な説明】

【0073】

【図1】実施の形態1における音響モデル生成装置のブロック図

【図2】同音響モデル生成装置の動作について説明するフローチャート

【図3】同第一特徴ベクトル情報の例を示す図

10

20

30

40

50

- 【図4】同モデルパラメータの例を示す図
- 【図5】同出現確率を算出する場合の概念図
- 【図6】同音響モデル生成装置が行う3手法を説明する図
- 【図7】同実験結果（誤り音素記述の発見率）を示す図
- 【図8】同実験結果（単語誤り率）を示す図
- 【図9】同実験結果（単語誤り率）を示す図
- 【図10】同音響モデル生成装置の概観図
- 【図11】同システムのブロック図
- 【符号の説明】

## 【0074】

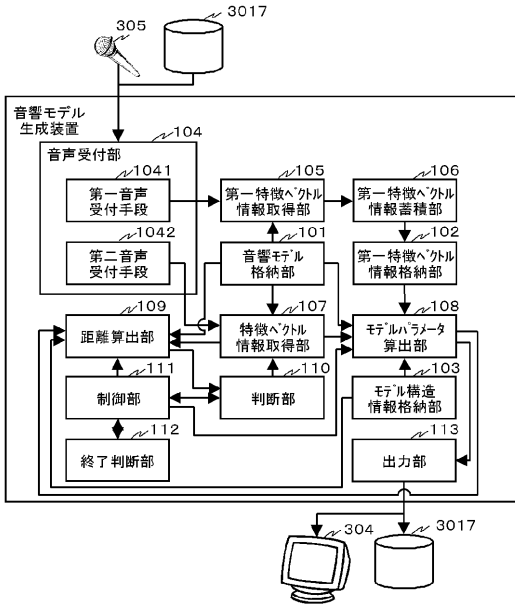
- 101 音響モデル格納部
- 102 第一特徴ベクトル情報格納部
- 103 モデル構造情報格納部
- 104 音声受付部
- 105 第一特徴ベクトル情報取得部
- 106 第一特徴ベクトル情報蓄積部
- 107 特徴ベクトル情報取得部
- 108 モデルパラメータ算出部
- 108 モデル構造情報モデルパラメータ算出部
- 109 距離算出部
- 110 判断部
- 111 制御部
- 112 終了判断部
- 113 出力部
- 1041 第一音声受付手段
- 1042 第二音声受付手段

10

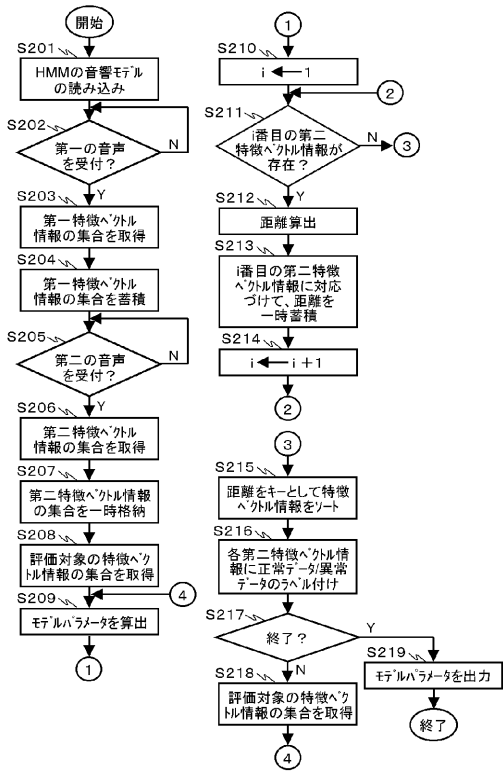
20



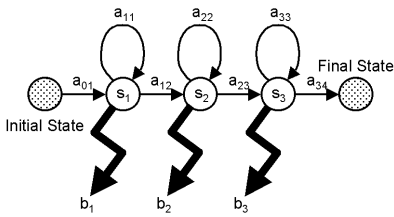
【図1】



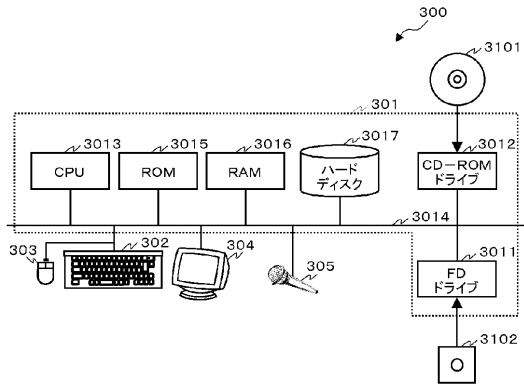
【図2】



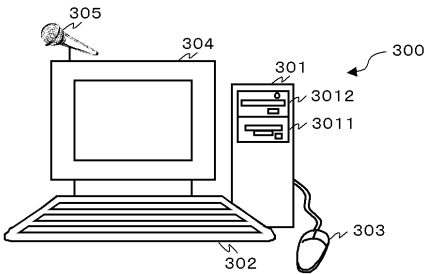
【図4】



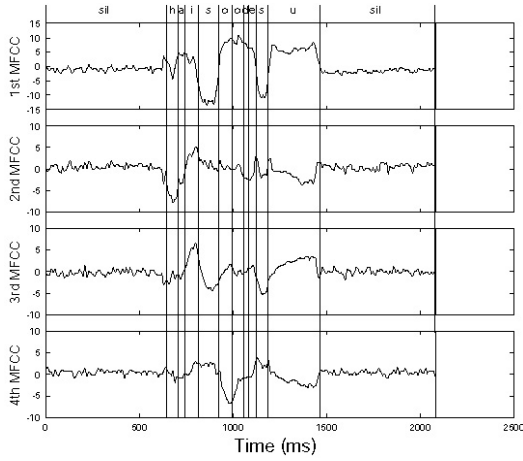
【図11】



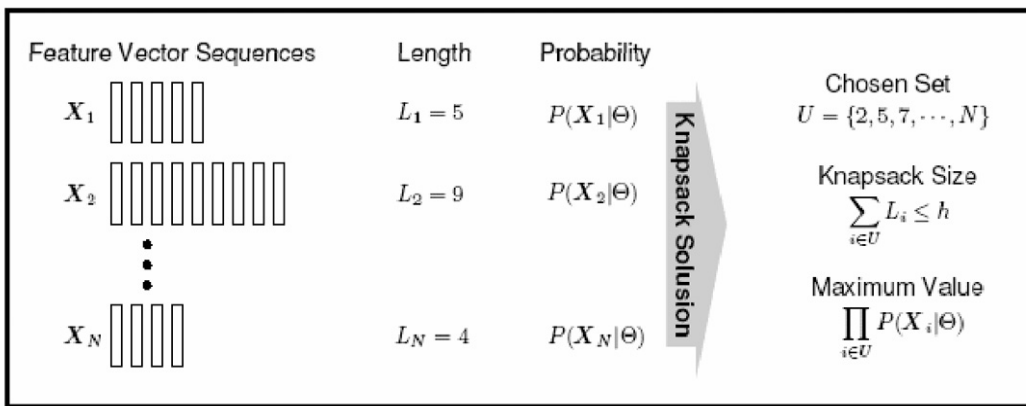
【図10】



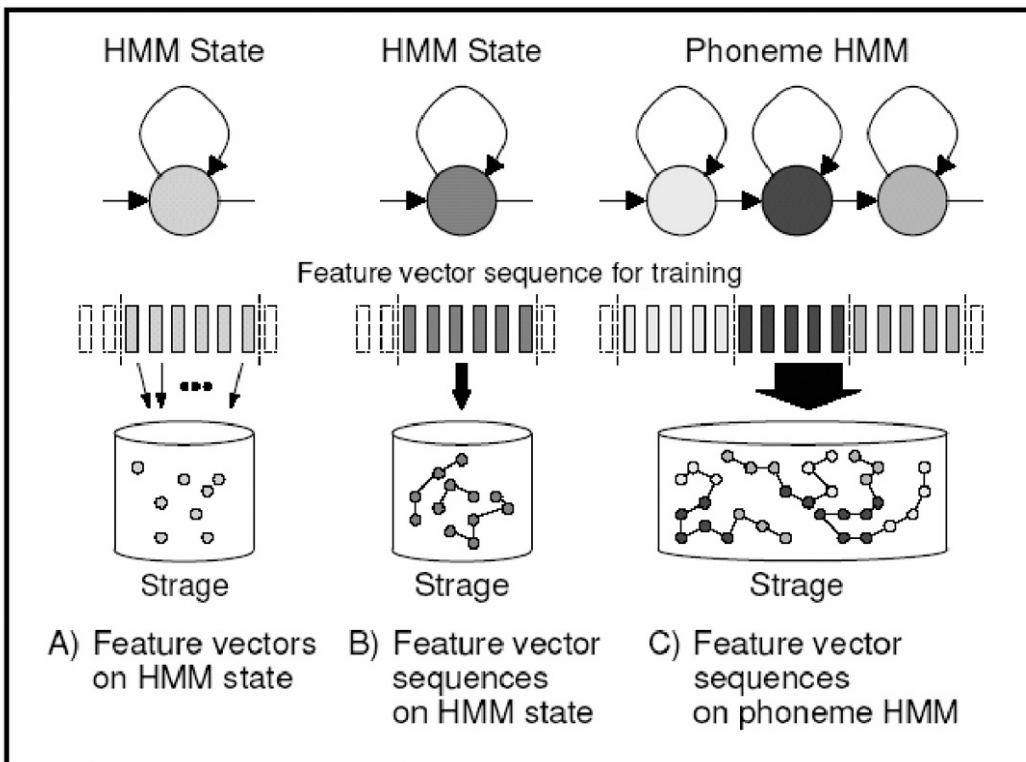
【 3 】



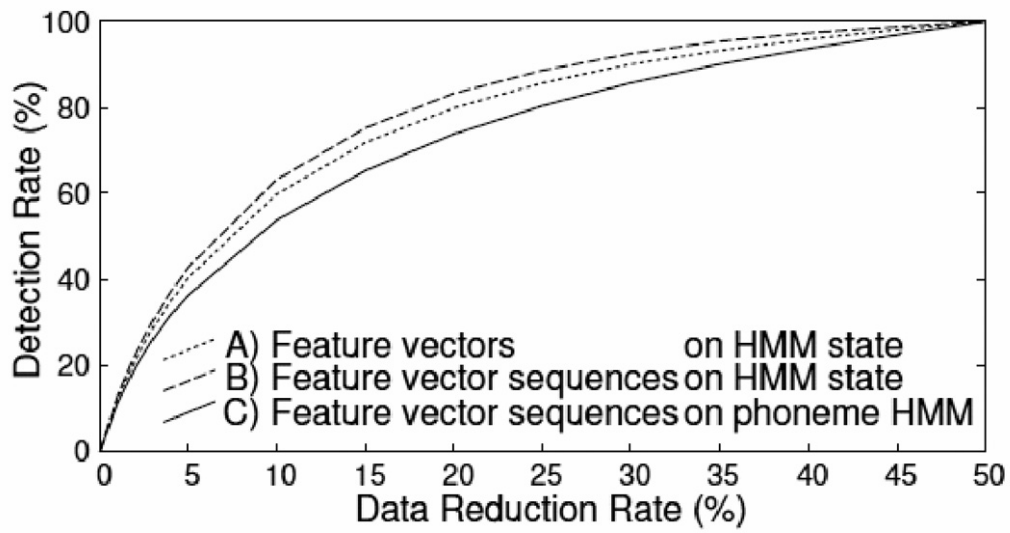
【 5 】



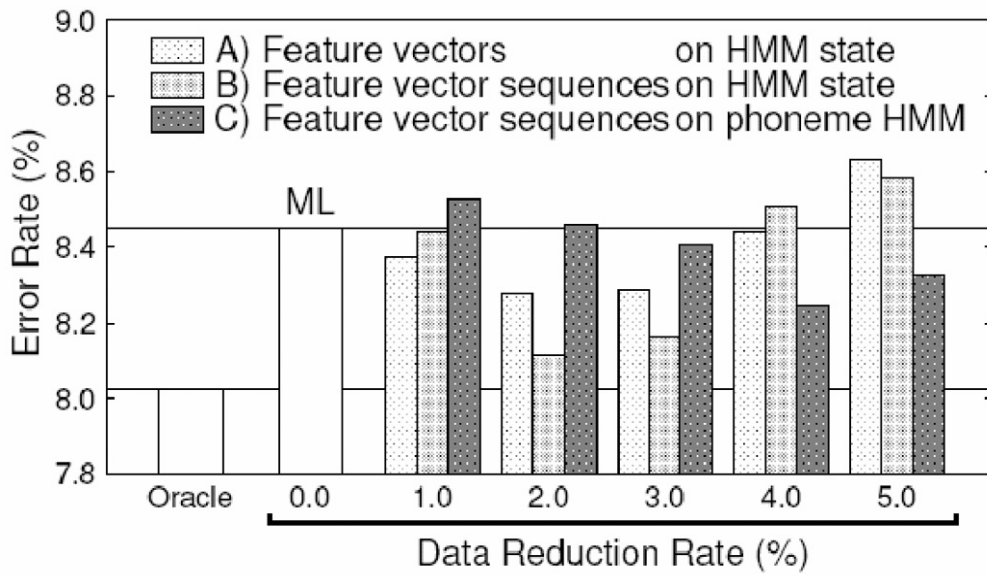
【 6 】



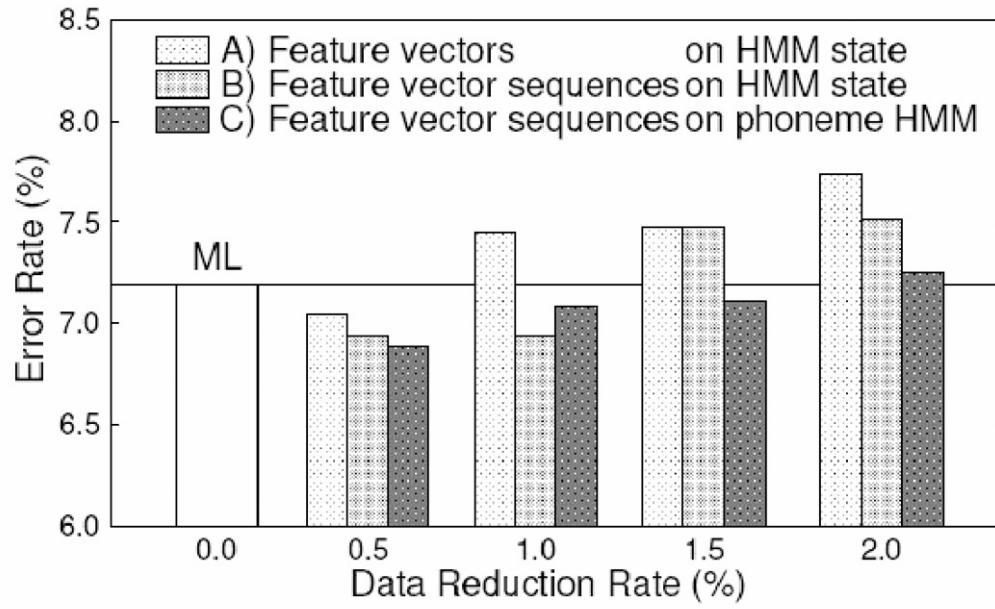
【 図 7 】



【 図 8 】



【 図 9 】



---

フロントページの続き

審査官 山下 剛史

- (56)参考文献 特開平06 - 012090 (JP, A)  
特開平09 - 062647 (JP, A)  
特開2004 - 334024 (JP, A)  
特開2001 - 083986 (JP, A)  
特開2000 - 352993 (JP, A)

- (58)調査した分野(Int.Cl., DB名)  
G10L 15/00 - 15/28