

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4716125号  
(P4716125)

(45) 発行日 平成23年7月6日(2011.7.6)

(24) 登録日 平成23年4月8日(2011.4.8)

(51) Int.Cl.	F I				
G 1 0 L 15/00	(2006.01)	G 1 0 L 15/00	2 0 0 E		
G 1 0 L 15/14	(2006.01)	G 1 0 L 15/14	2 0 0 A		
G 1 0 L 11/00	(2006.01)	G 1 0 L 11/00	4 0 2 F		

請求項の数 6 (全 27 頁)

<p>(21) 出願番号 特願2006-212704 (P2006-212704)</p> <p>(22) 出願日 平成18年8月4日(2006.8.4)</p> <p>(65) 公開番号 特開2008-40035 (P2008-40035A)</p> <p>(43) 公開日 平成20年2月21日(2008.2.21)</p> <p>審査請求日 平成21年4月8日(2009.4.8)</p> <p>(出願人による申告) 国等の委託研究成果に係る特許出願(平成18年度独立行政法人新エネルギー・産業技術総合開発機構、開発項目「半導体アプリケーションチッププロジェクト(情報家電用半導体アプリケーションチップ技術開発)マルチメディア多機能チップの研究開発」の「発音の評定を行うソフトウェアの多機能マルチメディアチップへの適応に関する研究開発」、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 渡辺 秀行 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 田川 博章 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 足立 隆弘 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】 発音評定装置、およびプログラム

(57) 【特許請求の範囲】

【請求項1】

音韻毎の隠れマルコフモデルに基づくデータである母語音素HMMを格納している母語音素HMM格納部と、  
 発音評定の対象である正解トランスクリプションに従ってHMMに基づくデータを連結したデータである連結HMMを格納している連結HMM格納部と、  
 音声の入力を受け付ける音声受付部と、  
 前記音声受付部が受け付けた音声を、フレームに区分するフレーム区分部と、  
 前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得部と、  
 前記特徴ベクトル系列と前記母語音素HMMを取得し、前記特徴ベクトル系列と前記母語音素HMMを用いて、ピタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得部と、  
 前記特徴ベクトル系列と前記連結HMMを取得し、前記特徴ベクトル系列と前記連結HMMを用いて、ピタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得部と、  
 前記特徴ベクトル系列、前記母語音素HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコアを算出するフレーム状態最適パススコア算出部と、

前記フレーム状態最適パススコア算出部が算出したフレームごとのスコアを出力する出力部を具備する発音評定装置。

【請求項 2】

音韻毎の隠れマルコフモデルに基づくデータである母語音素 HMM を格納している母語音素 HMM 格納部と、

発音評定の対象である正解トランスクリプションに従って HMM に基づくデータを連結したデータである連結 HMM を格納している連結 HMM 格納部と、

音声の入力を受け付ける音声受付部と、

前記音声受付部が受け付けた音声を、フレームに区分するフレーム区分部と、

前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得部と、

前記特徴ベクトル系列と前記母語音素 HMM を取得し、前記特徴ベクトル系列と前記母語音素 HMM を用いて、ビタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得部と、

前記特徴ベクトル系列と前記連結 HMM を取得し、前記特徴ベクトル系列と前記連結 HMM を用いて、ビタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得部と、

前記特徴ベクトル系列、前記母語音素 HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適音素内の 1 以上の各状態に対する確率を算出し、当該 1 以上の確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコアを算出するフレーム音素最適パススコア算出部と、

前記フレーム音素最適パススコア算出部が算出したフレームごとのスコアを出力する出力部を具備する発音評定装置。

【請求項 3】

前記音声受付部が受け付けた音声の発音区間を取得する発音区間取得部と、

前記発音区間ごとに、当該発音区間に対応する 1 以上のフレームごとの 1 以上のスコアを用いて、当該発音区間の代表のスコアを算出する代表値演算部をさらに具備し、

前記出力部は、少なくとも前記代表値演算部が算出した各発音区間の代表のスコアを出力する請求項 1 または請求項 2 記載の発音評定装置。

【請求項 4】

前記母語音素 HMM 格納部の母語音素 HMM から、発音評定の対象である正解トランスクリプションに従って連結 HMM を構成する連結 HMM 構成部をさらに具備し、

前記連結 HMM 格納部の連結 HMM は、前記連結 HMM 構成部が構成した連結 HMM である請求項 1 から請求項 3 いずれか記載の発音評定装置。

【請求項 5】

コンピュータに、

音声の入力を受け付ける音声受付ステップと、

前記音声受付ステップで受け付けた音声を、フレームに区分するフレーム区分ステップと、

前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得ステップと、

前記特徴ベクトル系列と格納している母語音素 HMM を取得し、前記特徴ベクトル系列と前記母語音素 HMM を用いて、ビタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得ステップと、

前記特徴ベクトル系列と格納している連結 HMM を取得し、前記特徴ベクトル系列と前記連結 HMM を用いて、ビタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得ステップと、

前記特徴ベクトル系列、前記母語音素 HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコ

10

20

30

40

50

アを算出するフレーム状態最適パススコア算出ステップと、  
前記フレーム状態最適パススコア算出ステップで算出したフレームごとのスコア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。

【請求項6】

コンピュータに、

音声の入力を受け付ける音声受付ステップと、

前記音声受付ステップで受け付けた音声を、フレームに区分するフレーム区分ステップと、

前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得ステップと、

10

前記特徴ベクトル系列と格納している母語音素HMMを取得し、前記特徴ベクトル系列と前記母語音素HMMを用いて、ビタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得ステップと、

前記特徴ベクトル系列と格納している連結HMMを取得し、前記特徴ベクトル系列と前記連結HMMを用いて、ビタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得ステップと、

前記特徴ベクトル系列、前記母語音素HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適音素内の1以上の各状態に対する確率を算出し、当該1以上の確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコアを算出するフレーム音素最適パススコア算出ステップと、  
前記フレーム音素最適パススコア算出ステップで算出したフレームごとのスコア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。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された音声を評価する装置等に関し、例えば、語学学習等に利用できる発音評定装置等に関するものである。

【背景技術】

【0002】

30

従来の技術として、以下の語学学習装置がある（特許文献1参照）。本語学学習装置は、学習者が選択した役割の発音をレファランスデータと比較して一致度によって点数化して表示し、点数によって適当な次の画面を自動に表示することにより、学習能率を向上させる装置である。本従来の語学学習装置は、入力された音声信号は音声認識技術により分析された後、学習者発音のスペクトルと抑揚とが学習者発音表示ボックスに表れるという構成になっている。そして、従来の技術においては、標準音データと学習者の発音のスペクトル、および抑揚が比較されて点数が表示される。

【特許文献1】特開2003-228279（第1頁、第1図等）

【発明の開示】

【発明が解決しようとする課題】

40

【0003】

しかし、従来の技術においては、標準音データと学習者の発音のスペクトル、および抑揚が比較されて点数が表示される構成であるので、両者の類似度の評定の精度が低く、また、リアルタイムに高速に点数を表示するためには、処理能力が極めて高いCPU、多量のメモリが必要であった。

【課題を解決するための手段】

【0004】

本第一の発明の発音評定装置は、音韻毎の隠れマルコフモデルに基づくデータである母語音素HMMを格納している母語音素HMM格納部と、発音評定の対象である正解トランスクリプションに従ってHMMに基づくデータを連結したデータである連結HMMを格納

50

している連結HMM格納部と、音声の入力を受け付ける音声受付部と、前記音声受付部が受け付けた音声を、フレームに区分するフレーム区分部と、前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得部と、前記特徴ベクトル系列と前記母語音素HMMを取得し、前記特徴ベクトル系列と前記母語音素HMMを用いて、ビタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得部と、前記特徴ベクトル系列と前記連結HMMを取得し、前記特徴ベクトル系列と前記連結HMMを用いて、ビタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得部と、前記特徴ベクトル系列、前記母語音素HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、前記区分された  
10  
フレームごとに発音の良し悪しを示すスコアを算出するフレーム状態最適パススコア算出部と、前記フレーム状態最適パススコア算出部が算出したフレームごとのスコアを出力する出力部を具備する発音評定装置である。

【0005】

かかる構成により、音韻毎に連結した連結HMMとフレーム分割されたデータを用いて、比較対象の音声と入力音声の類似度を評定することができるので、両者の類似度の評定の精度を高く、高速に行うことができる。

【0006】

また、本第二の発明の発音評定装置は、音韻毎の隠れマルコフモデルに基づくデータである母語音素HMMを格納している母語音素HMM格納部と、発音評定の対象である正解  
20  
トランスクリプションに従ってHMMに基づくデータを連結したデータである連結HMMを格納している連結HMM格納部と、音声の入力を受け付ける音声受付部と、前記音声受付部が受け付けた音声を、フレームに区分するフレーム区分部と、前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得部と、前記特徴ベクトル系列と前記母語音素HMMを取得し、前記特徴ベクトル系列と前記母語音素HMMを用いて、ビタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得部と、前記特徴ベクトル系列と前記連結HMMを取得し、前記特徴ベクトル系列と前記連結HMMを用いて、ビタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得部と、前記特徴ベクトル系列、前記母語音素HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記  
30  
区分されたフレームごとに最適音素内の1以上の各状態に対する確率を算出し、当該1以上の確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコアを算出するフレーム音素最適パススコア算出部と、前記フレーム音素最適パススコア算出部が算出したフレームごとのスコアを出力する出力部を具備する発音評定装置である。

【0007】

かかる構成により、状態のもっともらしさではなく、母語音素のもっともらしさを評定するので、第一の発明と比較しても、精度高く発音評定ができる。

【0008】

また、本第三の発明の発音評定装置は、第一、第二いずれかの発明に対して、前記音声  
40  
受付部が受け付けた音声の発音区間を取得する発音区間取得部と、前記発音区間ごとに、当該発音区間に対応する1以上のフレームごとの1以上のスコアを用いて、当該発音区間の代表のスコアを算出する代表値演算部をさらに具備し、前記出力部は、少なくとも前記代表値演算部が算出した各発音区間の代表のスコアを出力する発音評定装置である。

【0009】

かかる構成により、発音区間ごとのスコアを得ることができる。

【0010】

また、本第四の発明の発音評定装置は、第一から第三いずれかの発明に対して、前記母語音素HMM格納部の母語音素HMMから、発音評定の対象である正解トランスクリプションに従って連結HMMを構成する連結HMM構成部をさらに具備し、前記連結HMM格納部の連結HMMは、前記連結HMM構成部が構成した連結HMMである発音評定装置で  
50

ある。

【 0 0 1 1 】

かかる構成により、予め連結HMMを取得する必要がなく、連結HMMを準備する手間を省くことができる。

【 発明の効果 】

【 0 0 1 2 】

本発明による発音評定装置によれば、精度高く発音評定ができる。

【 発明を実施するための最良の形態 】

【 0 0 1 3 】

以下、発音評定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。

10

(実施の形態1)

【 0 0 1 4 】

本実施の形態において、比較対象の音声と入力音声の類似度の評定を精度高く、かつ高速にできる発音評定装置について説明する。本実施の形態における発音評定装置は、入力された音声、比較対象の音声(例えば、ネイティブが発音した音声)と比較して、音声を構成する各フレームの状態の類似度を算定するものである。

【 0 0 1 5 】

図1は、本実施の形態における発音評定装置のブロック図である。

20

【 0 0 1 6 】

発音評定装置は、母語音素HMM格納部101、連結HMM格納部102、音声受付部103、フレーム区分部104、特徴ベクトル系列取得部105、第一最適状態番号系列取得部106、第二最適状態番号系列取得部107、フレーム状態最適パススコア算出部108、発音区間取得部109、代表値演算部110、出力部111、連結HMM構成部112を具備する。

【 0 0 1 7 】

フレーム状態最適パススコア算出部108は、演算式格納手段1081、フレーム状態最適パススコア算出手段1082を具備する。

【 0 0 1 8 】

30

母語音素HMM格納部101は、音韻毎の隠れマルコフモデル(HMM)に基づくデータである母語音素HMMを格納している。母語音素HMMとは、ネイティブ発音の音声データベースから学習したネイティブ発音の音韻HMMである。HMMに基づくデータは、例えば、フレーム毎に、状態識別子と遷移確率の情報を有する。母語音素HMM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。

【 0 0 1 9 】

連結HMM格納部102は、発音評定の対象である正解トランスクリプションに従ってHMMに基づくデータを連結したデータである連結HMMを格納している。連結HMM格納部102の連結HMMは、後述する連結HMM構成部112が構成したものであることが好適である。ただし、本発音評定装置で連結HMMを構成せず、他の装置で構成した連結HMMを、連結HMM格納部102に記憶させても良い。連結HMM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。

40

【 0 0 2 0 】

音声受付部103は、音声の入力を受け付ける。音声の入力手段は、例えば、マイクや、記録媒体や、受信手段である。つまり、音声受付部103は、マイクからユーザの音声を受け付けても良いし、ハードディスクやCD-ROMなどの記録媒体から音声データを読み出しても良いし、外部装置から音声データを受信しても良い。音声受付部103は、例えば、マイクと、そのデバイスドライバー等で実現され得る。また、音声をユーザから受け付ける場合、ユーザが発音すべき音声を、図示しない手段により、ディスプレイに文字列などで出力することは好適である。

50

## 【 0 0 2 1 】

フレーム区分部 1 0 4 は、音声受付部 1 0 3 が受け付けた音声を、フレームに区分する。音声をフレームに区分する技術は公知技術であるので、詳細な説明を省略する。フレーム区分部 1 0 4 は、通常、MPU やメモリ等から実現され得る。フレーム区分部 1 0 4 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。

## 【 0 0 2 2 】

特徴ベクトル系列取得部 1 0 5 は、区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する。特徴ベクトル系列は、各フレームの特徴ベクトルの列である。なお、各フレームの特徴ベクトルを取得する技術は公知技術であるので、詳細な説明を省略する。例えば、特徴ベクトルは、三角型フィルタを用いたチャンネル数 2 4 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 1 2 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（3 9 次元）を有する。特徴ベクトル系列取得部 1 0 5 は、通常、MPU やメモリ等から実現され得る。特徴ベクトル系列取得部 1 0 5 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。

## 【 0 0 2 3 】

第一最適状態番号系列取得部 1 0 6 は、特徴ベクトル系列取得部 1 0 5 が取得した特徴ベクトル系列と、母語音素 HMM 格納部 1 0 1 の母語音素 HMM を読み出し、当該特徴ベクトル系列と当該母語音素 HMM を用いて、ビタビアルゴリズムにより全フレームの最適状態を決定する。各フレームの特徴ベクトルと母語音素 HMM を用いて、ビタビアルゴリズムにより、当該フレームの最適状態を決定する処理は公知技術であるので詳細な説明を省略する。第一最適状態番号系列取得部 1 0 6 が取得した全フレームの最適状態を、第一の最適状態番号系列という。第一最適状態番号系列取得部 1 0 6 は、通常、MPU やメモリ等から実現され得る。第一最適状態番号系列取得部 1 0 6 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。

## 【 0 0 2 4 】

第二最適状態番号系列取得部 1 0 7 は、特徴ベクトル系列取得部 1 0 5 が取得した特徴ベクトル系列と、連結 HMM 格納部 1 0 2 の連結 HMM を読み出し、当該特徴ベクトル系列と当該連結 HMM を用いて、ビタビアルゴリズムにより全フレームの最適状態を決定する。各フレームの特徴ベクトルと連結 HMM を用いて、ビタビアルゴリズムにより、当該フレームの最適状態を決定する処理は公知技術であるので詳細な説明を省略する。第二最適状態番号系列取得部 1 0 7 が取得した全フレームの最適状態を、第二の最適状態番号系列という。第二最適状態番号系列取得部 1 0 7 は、通常、MPU やメモリ等から実現され得る。第二最適状態番号系列取得部 1 0 7 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。

## 【 0 0 2 5 】

フレーム状態最適パススコア算出部 1 0 8 は、特徴ベクトル系列、母語音素 HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、区分されたフレームごとに発音の良し悪しを示すスコアを算出する。フレーム状態最適パススコア算出部 1 0 8 は、特徴ベクトル系列、母語音素 HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに、状態らしさを算出することとなる。また、フレーム状態最適パススコア算出部 1 0 8 は、後述する演算式格納手段 1 0 8 1 に格納されている演算式を用いて、状態らしさを算出するが、その演算式は後述の数式 1 に限られない。つまり、数式 1 のパラメータを若干、変更している演算式でも良い。また、フレーム状態最適パススコア算出部 1 0 8 は、例えば、数式 1 の「 $a_{q_t - 1}^{(1)} q_t^{(2)}$ 」が

10

20

30

40

50

「0」である場合に、予め決められた0でない値に変更して、「 $OPS_1(t)$ 」を算出しても良い。同様に、フレーム状態最適パススコア算出部108は、例えば、数式1の「 $a_{q_t^{(2)} q_{t+1}^{(1)}}$ 」が「0」である場合に、予め決められた0でない値に変更して、「 $OPS_1(t)$ 」を算出しても良い。つまり、フレーム状態最適パススコア算出部108は、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、区分されたフレームごとに発音のよし悪しを示すスコアを算出すれば、その算出に用いる演算式は問わない。フレーム状態最適パススコア算出部108は、通常、MPUやメモリ等から実現され得る。フレーム状態最適パススコア算出部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。

10

## 【0026】

演算式格納手段1081は、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列をパラメータとして、フレーム状態最適パススコアを算出する演算式の情報格納している。演算式格納手段1081は、例えば、以下の数式1の演算式の情報格納している。

## 【数1】

$$OPS_1(t) = \frac{a_{q_{t-1}^{(1)} q_t^{(2)}} \cdot a_{q_t^{(2)} q_{t+1}^{(1)}} \cdot b_{q_t^{(2)}}(o_t)}{\sum_{j=1}^M (a_{q_{t-1}^{(1)} j} \cdot a_{j q_{t+1}^{(1)}} \cdot b_j(o_t))}$$

20

## 【0027】

数式1において、「 $OPS_1(t)$ 」は、 $t$ 番目のフレームのフレーム状態最適パススコアを示す関数である。「 $t$ 」はフレームを識別する番号であり、「 $t$ 」は、( $t=1, 2, \dots, T$ ,  $T$ はフレーム総数)の範囲を取り得る。また、 $t$ 番目のフレームにおける特徴ベクトルを「 $o_t$ 」、 $t$ 番目分析フレームに対応するHMMの状態(状態番号)を $q_t$ 、特徴ベクトル系列を $O = \{o_1, o_2, \dots, o_T\}$ 、 $t$ 番目フレームに対する状態 $q_t$ を除いた系列を $Q \setminus \{q_t\} = \{q_1, \dots, q_{t-1}, q_{t+1}, \dots, q_T\}$ 、すべての母語音素モデルの(連結順序を考えない)集合を $a^{11}$ 、母語音素モデルを評価対象の正解トランスクリプションに従って連結したモデル(連結モデル)を $c^o$ とする。また、 $a^{11}$ のもとで、音素系列の制約なしに求めた最適状態系列(第一の最適状態番号系列)を $Q_1 = \{q_1^{(1)}, q_2^{(1)}, \dots, q_T^{(1)}\}$ 、 $c^o$ のもとで求めた最適パス(forced alignment 最適パス[第二の最適状態番号系列])を $Q_2 = \{q_1^{(2)}, q_2^{(2)}, \dots, q_T^{(2)}\}$ とする。また、 $a_{ij}$ は、 $i$ 番目状態から $j$ 番目状態へ遷移する確率、 $b_j(o)$ は $j$ 番目状態における特徴ベクトル $o$ の出力確率密度関数である。さらに、 $q_{T+1}$ は特徴ベクトルを出力しない最終状態であり、状態系列 $Q$ には含まれない便宜的に設定した状態である。なお、数式1において、 $t=1$ の場合、 $a_{q_0^{(1)} j}$ を $a_j$ に置き換える( $j=1, 2, \dots, N$ )。なお、 $a_j$ は、 $j$ 番目状態に対する初期状態確率である。なお、数式1において、「 $q_t^{(1)}$ 」の添え字の「 $t$ 」と「 $(1)$ 」は上下に配置されているが、「 $q_t^{(1)}$ 」と同意義である、とする。その他、本明細書の数式における添え字が上下に配置されているものと、「 $q_t^{(1)}$ 」の「 $t$ 」と「 $(1)$ 」のように横に並んで配置されているものとは同意義である、とする。また、「 $N$ 」は、すべての母語音素モデルにわたる総状態数である。

30

40

## 【0028】

演算式格納手段1081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。

## 【0029】

フレーム状態最適パススコア算出手段1082は、演算式格納手段1081から演算式の情報を読み出し、当該演算式に従って、パラメータの値を取得し、当該パラメータの値を、読み出した演算式に代入し、スコアを算出する。さらに、具体的には、フレーム状態

50

最適パススコア算出手段 1082 は、上記の数式 1 を読み出し、1 番目のフレームから、T 番目のフレームまで、フレームごとに、フレーム状態最適パススコアを算出する。フレーム状態最適パススコアは、各フレームの状態らしさを示すスコアである。フレーム状態最適パススコア算出手段 1082 は、フレームごとに、例えば、以下のように分母の値を算出する。つまり、まず、フレーム状態最適パススコア算出手段 1082 は、「 $j = 1$ 」から、「 $a_{q_{t-1}^{(1)}}^{(1)}$ 」( $q_{t-1}^{(1)}$  番目状態から 1 番目状態へ遷移する確率) を母語音素 HMM 格納部 101 から読み出す(なお、 $t = 1$  の場合は、「 $a_{q_t^{(1)}}^{(1)}$ 」は  $1$  になる。)。また、フレーム状態最適パススコア算出手段 1082 は、「 $a_{1q_{t+1}^{(1)}}^{(1)}$ 」(1 番目状態から  $q_{t+1}^{(1)}$  番目状態へ遷移する確率) を母語音素 HMM 格納部 101 から読み出す。そして、フレーム状態最適パススコア算出手段 1082 は、 $t$  番目のフレームの特徴ベクトルを取得し、当該特徴ベクトルを出力密度関数に代入し、 $b_1(o_t)$  を得る。そして、フレーム状態最適パススコア算出手段 1082 は、得た「 $a_{q_{t-1}^{(1)}}^{(1)}$ 」、「 $a_{1q_{t+1}^{(1)}}^{(1)}$ 」、および  $b_1(o_t)$  を乗算し、その結果を一時的にメモリに記憶する。次に、フレーム状態最適パススコア算出手段 1082 は、「 $j = 2$ 」として、同様に演算し、その結果を一時的にメモリに記憶する。そして、フレーム状態最適パススコア算出手段 1082 は、以上の処理を「 $j = N$ 」まで繰り返す。そして、フレーム状態最適パススコア算出手段 1082 は、メモリ上に一時記憶した乗算結果をすべて読み出し、それらの和を算出する。そして、フレーム状態最適パススコア算出手段 1082 は、数式 1 の分母の演算結果を得る。なお、「 $1$ 」とは、1 番目の状態に対する初期状態確率であり、母語音素 HMM 格納部 101 に予め格納されている。また、フレーム状態最適パススコア算出手段 1082 は、出力密度関数の情報を予め格納しており、かかる出力密度関数の情報を読み出し、パラメータを代入し、 $b_1(o_t)$  を得る。

#### 【0030】

次に、フレーム状態最適パススコア算出手段 1082 は、フレームごとに、以下のように分子の値を算出する。つまり、フレーム状態最適パススコア算出手段 1082 は、「 $a_{q_{t-1}^{(1)}q_t^{(2)}}^{(1)}$ 」( $q_{t-1}^{(1)}$  番目状態から  $q_t^{(2)}$  番目状態へ遷移する確率) を母語音素 HMM 格納部 101 から読み出す(なお、 $t = 1$  の場合は、「 $a_{q_t^{(1)}q_t^{(2)}}^{(1)}$ 」は  $q_t^{(2)}$  になる。)。また、フレーム状態最適パススコア算出手段 1082 は、「 $a_{q_t^{(2)}q_{t+1}^{(1)}}^{(1)}$ 」( $q_t^{(2)}$  番目状態から  $q_{t+1}^{(1)}$  番目状態へ遷移する確率) を母語音素 HMM 格納部 101 から読み出す。また、フレーム状態最適パススコア算出手段 1082 は、 $t$  番目のフレームの特徴ベクトルを取得し、当該特徴ベクトルを出力密度関数に代入し、 $b_{q_t^{(2)}}(o_t)$  を得る。そして、フレーム状態最適パススコア算出手段 1082 は、得た 3 つの値を乗算し、分子の値を取得し、メモリに一時記憶する。

#### 【0031】

次に、フレーム状態最適パススコア算出手段 1082 は、フレームごとに、分母の値と分子の値からフレーム状態最適パススコアを算出し、少なくとも、メモリ上に一時記憶する。

#### 【0032】

フレーム状態最適パススコア算出手段 1082 は、通常、MPU やメモリ等から実現され得る。フレーム状態最適パススコア算出手段 1082 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。

#### 【0033】

発音区間取得部 109 は、音声受付部 103 が受け付けた音声の発音区間を取得する。発音区間は、音素単位でも良いし、単語単位でも良いし、文単位等でも良い。発音区間取得部 109 が取得する発音区間は、フレームの番号列でも良いし、音声受付部 103 が受け付けた音声データを発音区間に分割した結果でも良い。発音区間取得部 109 は、通常、MPU やメモリ等から実現され得る。発音区間取得部 109 の処理手順は、通常、ソフ

10

20

30

40

50



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。

【0034】

代表値演算部110は、発音区間ごとに、当該発音区間に対応する1以上のフレームごとの1以上のスコアを用いて、当該発音区間の代表のスコアを算出する。代表値演算部110は、1以上のフレームごとの1以上のスコアの中央値を算出することが好適であるが、平均値や最大値などの他の代表値を算出するようにしても良い。代表値演算部110は、通常、MPUやメモリ等から実現され得る。代表値演算部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。

10

【0035】

出力部111は、フレーム状態最適パススコア算出部108が算出したフレームごとのスコアを出力する。出力部111は、代表値演算部110が算出した各発音区間の代表のスコアを出力しても良い。出力部111は、フレームごとのスコアと、各発音区間の代表のスコアの両方を出力しても良いし、どちらか一方を出力しても良い。ここで、出力とは、ディスプレイへの表示、プリンタへの印字、音声出力、外部の装置への送信、記録媒体への蓄積等を含む概念である。出力部111は、ディスプレイやスピーカー等の出力デバイスを含むと考えると含まないとも考えても良い。出力部111は、出力デバイスのドライバソフトまたは、出力デバイスのドライバソフトと出力デバイス等で実現され得る。

【0036】

20

連結HMM構成部112は、母語音素HMM格納部101の母語音素HMMから、発音評定の対象である正解トランスクリプションに従って連結HMMを構成し、連結HMM格納部102に少なくとも一時記憶する。母語音素HMMから、発音評定の対象である正解トランスクリプションに従って連結HMMを構成する技術は公知技術であるので詳細な説明を省略する。なお、正解トランスクリプションの情報は、連結HMM構成部112が、予め保持していても良いし、ユーザが手動で与えても良い。連結HMM構成部112は、通常、MPUやメモリ等から実現され得る。連結HMM構成部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。

【0037】

30

次に、発音評定装置の動作について図2、図3のフローチャートを用いて説明する。

【0038】

（ステップS201）音声受付部103は、ユーザから、発音評定の対象の音声を受け付けたか否かを判断する。なお、ユーザから音声を受け付ける前に、図示しない手段により、所定の単語や文の発音をユーザに促すようにしても良い。かかる処理は、例えば、「learn と発音してください。」とディスプレイに表示したり、スピーカーで音声出力したりする処理である。音声受付部103が音声を受け付ければステップS202に行き、音声を受け付けなければステップS201に戻る。

【0039】

（ステップS202）連結HMM構成部112は、母語音素HMM格納部101から母語音素HMMを読み出し、メモリ上に配置する。

40

【0040】

（ステップS203）連結HMM構成部112は、正解トランスクリプションを取得する。正解トランスクリプションは、例えば、図示しない記憶手段に予め格納されていても良いし、ユーザが入力して与えても良い。

【0041】

（ステップS204）連結HMM構成部112は、ステップS202で読み出した母語音素HMMから、ステップS203で取得した正解トランスクリプションに従って連結HMMを構成し、メモリ上に配置する。かかる場合、連結HMM格納部102は、メモリである。

50

## 【 0 0 4 2 】

(ステップS 2 0 5) フレーム区分部 1 0 4 は、ステップS 2 0 1 で受け付けた音声をフレームに区分し、1以上の各フレームに対応する1以上のフレーム音声データを取得する。

## 【 0 0 4 3 】

(ステップS 2 0 6) 特徴ベクトル系列取得部 1 0 5 は、ステップS 2 0 5 で取得された1以上の各フレーム音声データを音声分析し、1以上の各フレームに対応する1以上の特徴ベクトルを算出し、特徴ベクトル系列(O)を取得する。

## 【 0 0 4 4 】

(ステップS 2 0 7) 第一最適状態番号系列取得部 1 0 6 は、ステップS 2 0 6 で取得した特徴ベクトル系列と、母語音素HMMを用いて、ビタビアルゴリズムにより全フレームの最適状態を決定する。ここでの最適状態の集合は、第一最適状態番号系列( $Q_1$ )である。

10

## 【 0 0 4 5 】

(ステップS 2 0 8) 第二最適状態番号系列取得部 1 0 7 は、ステップS 2 0 6 で取得した特徴ベクトル系列と、連結HMMを用いて、ビタビアルゴリズムにより全フレームの最適状態を決定する。ここでの最適状態の集合は、第二最適状態番号系列( $Q_2$ )である。

## 【 0 0 4 6 】

(ステップS 2 0 9) フレーム状態最適パススコア算出部 1 0 8 は、カウンタ t に 1 を代入する。ここで「t」は、フレーム番号である。

20

## 【 0 0 4 7 】

(ステップS 2 1 0) フレーム状態最適パススコア算出部 1 0 8 は、t 番目のフレームが存在するか否かを判断する。t 番目のフレームが存在すればステップS 2 1 1 に行き、存在しなければステップS 2 1 3 に行く。

## 【 0 0 4 8 】

(ステップS 2 1 1) フレーム状態最適パススコア算出部 1 0 8 は、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、t 番目のフレームの最適状態に対する確率を算出し、当該確率を用いて、t 番目のフレームについて、発音のよし悪しを示すスコアを算出する。フレーム状態最適パススコアを算出する詳細な処理例については、図3のフローチャートを用いて説明する。

30

## 【 0 0 4 9 】

(ステップS 2 1 2) フレーム状態最適パススコア算出部 1 0 8 は、カウンタ t を 1、インクリメントし、ステップS 2 1 0 に戻る。

## 【 0 0 5 0 】

(ステップS 2 1 3) 発音区間取得部 1 0 9 は、カウンタ i に 1 を代入する。

## 【 0 0 5 1 】

(ステップS 2 1 4) 発音区間取得部 1 0 9 は、i 番目の発音区間が存在するか否かを判断する。i 番目の発音区間が存在すればステップS 2 1 5 に行き、i 番目の発音区間が存在しなければ処理を終了する。

40

## 【 0 0 5 2 】

(ステップS 2 1 5) 発音区間取得部 1 0 9 は、i 番目の発音区間に対応する1以上のフレームを特定し、代表値演算部 1 1 0 は、当該発音区間に対応する1以上のフレームの1以上のスコアを読み出す。

## 【 0 0 5 3 】

(ステップS 2 1 6) 代表値演算部 1 1 0 は、代表値を得るための関数の情報を読み出し、当該関数に、ステップS 2 1 5 で読み出した1以上のスコアを代入し、関数を実行し、当該発音区間の代表のスコアを算出する。

## 【 0 0 5 4 】

(ステップS 2 1 7) 出力部 1 1 1 は、ステップS 2 1 6 で算出した代表のスコアを出

50

力する。

【0055】

(ステップS218)発音区間取得部109は、カウンタ $i$ を1、インクリメントし、ステップS215に戻る。

【0056】

次に、ステップS211のフレーム状態最適パススコア算出処理について、図3のフローチャートを用いて説明する。

【0057】

(ステップS301)フレーム状態最適パススコア算出手段1082は、演算式格納手段1081からフレーム状態最適パススコアを算出するための演算式の情報を読み出す。 10

【0058】

(ステップS302)フレーム状態最適パススコア算出手段1082は、カウンタ $j$ に1を代入する。

【0059】

(ステップS303)フレーム状態最適パススコア算出手段1082は、「 $j > N$ 」であるか否かを判断する。

【0060】

(ステップS304)フレーム状態最適パススコア算出手段1082は、母語音素HMM格納部101から「 $a_{q_{t-1}}^{(1)}(j)$ 」を読み出す。なお、「 $t = 1$ 」の場合、「 $a_{q_{t-1}}^{(1)}(j)$ 」は、 $j$ ( $j$ 番目状態に対する初期状態確率)である。 20

【0061】

(ステップS305)フレーム状態最適パススコア算出手段1082は、母語音素HMM格納部101から「 $a_{jq_{t+1}}^{(1)}$ 」を読み出す。

【0062】

(ステップS306)フレーム状態最適パススコア算出手段1082は、特徴ベクトル( $O_t$ )を読み出す。

【0063】

(ステップS307)フレーム状態最適パススコア算出手段1082は、出力密度関数 $b_j(\cdot)$ の情報を読み出し、当該関数に特徴ベクトル( $O_t$ )を代入し、出力密度関数 $b_j(O_t)$ を実行し、その結果を得る。 30

【0064】

(ステップS308)フレーム状態最適パススコア算出手段1082は、得た「 $a_{q_{t-1}}^{(1)}(j)$ 」、「 $a_{jq_{t+1}}^{(1)}$ 」、および $b_j(O_t)$ を乗算し、その結果を一時的にメモリに記憶する。

【0065】

(ステップS309)フレーム状態最適パススコア算出手段1082は、カウンタ $j$ を1、インクリメントし、ステップS303に戻る。

【0066】

(ステップS310)フレーム状態最適パススコア算出手段1082は、ステップS308で一時格納したすべての乗算結果の和を算出し、その結果を一時的にメモリに記憶する。 40

【0067】

(ステップS311)フレーム状態最適パススコア算出手段1082は、「 $a_{q_{t-1}}^{(1)}(q_t^{(2)})$ 」を母語音素HMM格納部101から読み出す。

【0068】

(ステップS312)フレーム状態最適パススコア算出手段1082は、「 $a_{q_{t+1}}^{(1)}(q_t^{(2)})$ 」を母語音素HMM格納部101から読み出す。

【0069】

(ステップS313)フレーム状態最適パススコア算出手段1082は、 $t$ 番目のフレームの特徴ベクトルを取得し、当該特徴ベクトルを出力密度関数 $b_{q_t^{(2)}}(\cdot)$ に代入 50

し、 $b_{q_t}^{(2)}(o_t)$  を演算し、その結果を一時的にメモリに記憶する。

【0070】

(ステップS314) フレーム状態最適パススコア算出手段1082は、ステップS311で得た「 $a_{q_{t-1}}^{(1)} q_t^{(2)}$ 」、ステップS312で得た「 $a_{q_t}^{(2)} q_{t+1}^{(1)}$ 」、およびステップS313で得た「 $b_{q_t}^{(2)}(o_t)$ 」を乗算し、その結果を一時的にメモリに記憶する。

【0071】

(ステップS315) フレーム状態最適パススコア算出手段1082は、ステップS314における演算結果を、ステップS310における演算結果で除算し、 $t$  番目のフレームのフレーム状態最適パススコア( $OPS_1(t)$ )を得る。

10

【0072】

(ステップS316) フレーム状態最適パススコア算出手段1082は、ステップS315で得た $t$  番目のフレームのフレーム状態最適パススコア( $OPS_1(t)$ )を、一時的にメモリに記憶し、上位関数にリターンする。

【0073】

以下、本実施の形態における発音評定装置の発音評定の方法について説明する。

【0074】

まず、上記の数式1が導かれる過程について説明する。上述したように、ユーザ(学習者話者)音声の $t$  番目分析フレーム( $t = 1, 2, \dots, T$ ,  $T$ はフレーム総数)における特徴ベクトルを「 $o_t$ 」、 $t$  番目分析フレームに対応するHMMの状態(状態番号)を「 $q_t$ 」、特徴ベクトル系列を $O = \{o_1, o_2, \dots, o_T\}$ 、状態(番号)系列を $Q = \{q_1, q_2, \dots, q_T\}$ 、 $t$  番目フレームに対する状態 $q_t$ を除いた系列を $Q \setminus \{q_t\} = \{q_1, \dots, q_{t-1}, q_{t+1}, \dots, q_T\}$ 、すべての母語音素モデルの(連結順序を考えない)集合を $a^{11}$ 、母語音素モデルを評定対象の正解トランスクリプションに従って連結したモデル(連結モデル)を $c^o$ とする。また、 $a^{11}$ のもとで、音素系列の制約なしに求めた最適状態系列(最適パス)を $Q_1 = \{q_1^{(1)}, q_2^{(1)}, \dots, q_T^{(1)}\}$ 、 $c^o$ のもとで求めた最適パス(forced alignment 最適パス)を $Q_2 = \{q_1^{(2)}, q_2^{(2)}, \dots, q_T^{(2)}\}$ とする。なお、「 $Q_1$ 」は第一の最適状態番号系列、「 $Q_2$ 」は第二の最適状態番号系列である。

20

【0075】

かかる概念を図4に示す。図4において、縦は母語音素HMMの状態系列、横は学習者話者の音声特徴ベクトル系列である。

30

【0076】

この場合、フレーム状態最適パススコアを以下の数式2で定義する。

【数2】

$$OPS_1(t) = \frac{P(O, Q_1 \setminus \{q_t^{(1)}\}, q_t = q_t^{(2)} | \Lambda^{all})}{\sum_{j=1}^N P(O, Q_1 \setminus \{q_t^{(1)}\}, q_t = j | \Lambda^{all})}$$

【0077】

なお、 $P(A | a^{11})$ は、すべての音素モデル $a^{11}$ が与えられたときの事象 $A$ が起こる確率(または確率密度)、 $N$ は $a^{11}$ に含まれる総状態数である。数式2を变形すると以下の数式3となり、フレーム状態最適パススコアは、特徴ベクトル系列 $O$ 、および $a^{11}$ の下で求めた最適パスから $t$  番目フレームを除いた最適パス「 $Q \setminus \{q_t^{(1)}\} = \{q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_{t+1}^{(1)}, \dots, q_T^{(1)}\}$ 」が与えられたという条件の下での、 $t$  番目フレームに対する状態が「 $q_t^{(2)}$ 」(forced alignmentで求めた最適状態)である事後確率を表わす。

40

【数3】

$$OPS_1(t) = P(q_t = q_t^{(2)} | O, Q_1 \setminus \{q_t^{(1)}\}, \Lambda^{all})$$

50

## 【 0 0 7 8 】

また、音素モデルが HMM の場合、状態系列 Q に沿った確率（確率密度）は、以下の数式 4 で与えられる。

## 【 数 4 】

$$P(O, Q | \Lambda^{all}) = \pi_{q_1} \prod_{t=1}^T a_{q_t q_{t+1}} b_{q_t}(o_t)$$

## 【 0 0 7 9 】

ここで  $a_{ij}$  は  $i$  番目状態に対する初期状態確率、 $a_{ij}$  は  $i$  番目状態から  $j$  番目状態へ遷移する確率、 $b_j(o)$  は  $j$  番目状態における特徴ベクトルの出力確率密度関数である。また、 $q_{T+1}$  は特徴ベクトルを出力しない最終状態であり、状態系列 Q には含まれない便宜的に設定した状態である。

## 【 0 0 8 0 】

かかる場合、確率密度関数は、数式 5 に示すように簡略され得る。

## 【 数 5 】

$$\begin{aligned} P(O, Q_1 \setminus \{q_t^{(1)}\}, q_t = i | \Lambda^{all}) &= \pi_{q_1^{(1)}} a_{q_1^{(1)} q_2^{(1)}} b_{q_1^{(1)}}(o_1) a_{q_2^{(1)} q_3^{(1)}} b_{q_2^{(1)}}(o_2) \cdots a_{q_{t-2}^{(1)} q_{t-1}^{(1)}} b_{q_{t-2}^{(1)}}(o_{t-2}) \\ &\times a_{q_{t-1}^{(1)} q_t^{(1)}} b_{q_{t-1}^{(1)}}(o_{t-1}) a_{iq_{t+1}^{(1)}} b_i(o_t) \\ &\times a_{q_{t+1}^{(1)} q_{t+2}^{(1)}} b_{q_{t+1}^{(1)}}(o_{t+1}) \cdots a_{q_T^{(1)} q_{T+1}^{(1)}} b_{q_T^{(1)}}(o_T) \\ &= \left( a_{q_{t-1}^{(1)}} \cdot a_{iq_{t+1}^{(1)}} \cdot b_i(o_t) \right) \left( \pi_{q_1^{(1)}} \prod_{\tau=1}^{t-2} a_{q_\tau^{(1)} q_{\tau+1}^{(1)}} b_{q_\tau^{(1)}}(o_\tau) \right) \left( b_{q_{t-1}^{(1)}}(o_{t-1}) \prod_{\tau=t+1}^T a_{q_\tau^{(1)} q_{\tau+1}^{(1)}} b_{q_\tau^{(1)}}(o_\tau) \right) \end{aligned}$$

## 【 0 0 8 1 】

数式 5 を数式 2 の分子と分母に代入して、約分することにより、HMM の場合のフレーム状態最適パススコアが数式 1 のように簡略化される。

## 【 0 0 8 2 】

数式 1 において、 $t = 1$  の場合、 $a_{q_0^{(1)}} b_{q_0^{(1)}}(o_0)$  を  $b_{q_0^{(1)}}(o_0)$  に置き換える ( $j = 1, 2, \dots, N$ )。

## 【 0 0 8 3 】

また、数式 1 により算出されるフレーム状態最適パススコア ( $OPS_1(\cdot)$ ) を発音区間  $T(m)$  で評価した発音区間フレーム状態最適パススコア ( $OPS_3(m)$ ) は、数式 6 により算出される。かかる算出は、代表値演算部 110 が行う。

## 【 数 6 】

$$OPS_3(m) = Me \{ OPS_1(\tau) | \tau \in T(m) \}$$

## 【 0 0 8 4 】

数式 6 において、 $m$  は発音区間のインデクス、 $Me \{ \cdot \}$  は集合  $\{ \cdot \}$  内の要素の中央値を表わす。ただし、中央値を算出する代表値演算は、平均値や最大値などの別の代表値演算に置き換えても良い。また、発音区間は、例えば、forced alignment パス  $Q_2$  により定まる、音素区間、単語区間、フレーズ区間などである。

## 【 0 0 8 5 】

次に、数式 1 の性質、意義、つまり、上述したフレーム状態最適パススコアの性質、意義について説明する。

## 【 0 0 8 6 】

上述した数式 2 を詳細に記載すると、以下の数式 7 になる。

10

20

30

40

【数7】

$$\text{OPS}_1(t) = \frac{P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_t^{(2)}, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})}{\sum_{j=1}^N P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})}$$

【0087】

数式7において、 $Q_1 = \{q_1^{(1)}, q_2^{(1)}, \dots, q_T^{(1)}\}$ が、 $a^{11}$ のもとでの最適状態系列、つまり、「 $P(O, q_1, q_2, \dots, q_T | a^{11})$ 」を最大にする $\{q_1, q_2, \dots, q_T\}$ であるので、数式7の分母における各 $j$  ( $j = 1, 2, \dots, N$ ):  $P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | a^{11})$ において、 $j = q_t^{(1)}$ で最大となる。もし、ユーザの発話がネイティブ発音らしければ、各フレームにおいて $a^{11}$ の中のどれかの母語音素モデルと類似性が高く、 $P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | a^{11})$ は最適パス $\{q_1^{(1)}, q_2^{(1)}, \dots, q_T^{(1)}\}$ での値が優勢となるので、 $P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | a^{11})$ の各 $j$ において、「 $j = q_t^{(1)}$ 」での値が大きくなって、他の $j$ では値が大きならない(確率値が、 $j = 1, 2, \dots, N$ の中で $q_t^{(1)}$ の占める割合が高くなる)。すなわち、数式7の分母は $a^{11}$ の下での最適状態系列 $Q_1$ における確率値に近くなる。つまり、以下の数式8がいえる。かかる状況を図5に示す。図5において、縦軸はスコア、横軸は $j$ (状態番号)である。

【数8】

$$\sum_{j=1}^N P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all}) \simeq P(O, q_1^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})$$

【0088】

さらに正解トランスクリプションのネイティブ発音に近ければ、 $a^{11}$ のもとで求めた最適パス $Q_1$ と $c^o r$ のもとで求めた最適パス $Q_2$ が類似するため、多くの $t$ において「 $q_t^{(1)} = q_t^{(2)}$ 」となり、以下の数式9がいえる。

【数9】

$$\begin{aligned} \text{OPS}_1(t) &= \frac{P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_t^{(2)}, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})}{\sum_{j=1}^N P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})} \\ &\simeq \frac{P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_t^{(2)}, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})}{P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_t^{(1)}, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})} \\ &= 1 \end{aligned}$$

【0089】

つまり、「 $\text{OPS}_1(t)$ 」の値が最大値1に近くなる。

【0090】

一方、正解トランスクリプションと異なるネイティブ発音であれば、 $Q_1$ と $Q_2$ が大きく異なってくるため、「 $\text{OPS}_1(t)$ 」の値が小さくなる。また、元々発話がネイティブ発音に類似していなければ、数式7の分子の値、および分母の各 $j$ に対する値がいずれも小さくなり、 $N$ の値が大きいことから「 $\text{OPS}_1(t)$ 」の値が小さくなる。フレーム状態最適パススコア算出部108が算出するフレーム状態最適パススコアは、以上のような性質を有する。

【0091】

以上、本実施の形態によれば、音韻毎に連結した連結HMMとフレーム分割されたデータを用いて、比較対象の音声と入力音声の類似度を評定することができるので、両者の類

10

20

30

40

50

似度の評定の精度を高く、高速に行うことができる。さらに具体的には、本実施の形態によれば、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、区分されたフレームごとに発音の良し悪しを示すスコアを算出できる。このスコアは、母語音素HMMが有する各状態と比較した、ユーザが入力した音声の各状態の状態らしさを示すスコアである。

#### 【0092】

なお、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情報処理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、音声の入力を受け付ける音声受付ステップと、前記音声受付ステップで受け付けた音声を、フレームに区分するフレーム区分ステップと、前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得ステップと、前記特徴ベクトル系列と格納している母語音素HMMを取得し、前記特徴ベクトル系列と前記母語音素HMMを用いて、ビタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得ステップと、前記特徴ベクトル系列と格納している連結HMMを取得し、前記特徴ベクトル系列と前記連結HMMを用いて、ビタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得ステップと、前記特徴ベクトル系列、前記母語音素HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適状態に対する確率を算出し、当該確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコアを算出するフレーム状態最適パススコア算出ステップと、前記フレーム状態最適パススコア算出ステップで算出したフレームごとのスコア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。

#### 【0093】

また、上記プログラムに対して、コンピュータに、前記音声受付部が受け付けた音声の発音区間を取得する発音区間取得ステップと、前記発音区間ごとに、当該発音区間に対応する1以上のフレームごとの1以上のスコアを用いて、当該発音区間の代表のスコアを算出する代表値演算ステップをさらに実行させ、前記出力ステップにおいて、少なくとも前記代表値演算ステップで算出した各発音区間の代表のスコアを出力する、ことは好適である。

#### 【0094】

また、上記プログラムに対して、コンピュータに、前記母語音素HMMから、発音評定の対象である正解トランスクリプションに従って連結HMMを構成する連結HMM構成ステップをさらに実行させ、前記連結HMMは、前記連結HMM構成ステップで構成した連結HMMである、ことは好適である。

(実施の形態2)

#### 【0095】

本実施の形態において、比較対象の音声と入力音声の類似度の評定を精度高く、かつ高速にできる発音評定装置について説明する。本実施の形態における発音評定装置は、入力された音声、比較対象の音声(例えば、ネイティブが発音した音声)と比較して、音声を構成する各フレームの音素の類似度を算定するものである。

#### 【0096】

図6は、本実施の形態における発音評定装置のブロック図である。

#### 【0097】

発音評定装置は、母語音素HMM格納部101、連結HMM格納部102、音声受付部103、フレーム区分部104、特徴ベクトル系列取得部105、第一最適状態番号系列取得部106、第二最適状態番号系列取得部107、フレーム音素最適パススコア算出部

10

20

30

40

50

608、発音区間取得部109、代表値演算部110、出力部111、連結HMM構成部112を具備する。

【0098】

フレーム音素最適パススコア算出部608は、演算式格納手段6081、フレーム音素最適パススコア算出手段6082を具備する。

【0099】

フレーム音素最適パススコア算出部608は、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに最適音素内の1以上の各状態に対する確率を算出し、当該1以上の確率を用いて、区分されたフレームごとに発音の良し悪しを示すスコアを算出する。また、フレーム音素最適パススコア算出部608は、後述する演算式格納手段6081に格納されている演算式を用いて、音素らしさを算出するが、その演算式は後述の数式10に限られない。つまり、数式10のパラメータを若干、変更している演算式でも良い。また、数式10において、分子の値が「0」である場合に、予め決められた0でない値に変更して、「 $OPS_2(t)$ 」を算出しても良い。つまり、フレーム音素最適パススコア算出部608は、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに最適音素内の1以上の各状態に対する確率を算出し、当該1以上の確率を用いて、区分されたフレームごとに発音の良し悪し（音素らしさ）を示すスコアを算出すれば、その算出に用いる演算式は問わない。フレーム音素最適パススコア算出部608は、通常、MPUやメモリ等から実現され得る。フレーム音素最適パススコア算出部608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。

【0100】

演算式格納手段6081は、特徴ベクトル系列、母語音素HMM、第一の最適状態番号系列、および第二の最適状態番号系列をパラメータとして、フレーム音素最適パススコアを算出する演算式の情報に格納している。演算式格納手段6081は、例えば、以下の数式10の演算式の情報に格納している。演算式格納手段6081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。

【数10】

$$OPS_2(t) = \frac{\sum_{i \in S(\lambda[q_t^{(2)}])} (a_{q_{t-1}^{(1)}} \cdot a_{iq_{t+1}^{(1)}} \cdot b_i(o_t))}{\sum_{j=1}^N (a_{q_{t-1}^{(1)}} \cdot a_{jq_{t+1}^{(1)}} \cdot b_j(o_t))}$$

【0101】

数式10において、「 $OPS_2(t)$ 」は、t番目のフレームのフレーム音素最適パススコアを示す関数である。数式10において、t番目フレームの最適状態「 $q_t^{(2)}$ 」を含む音素（「最適音素」という）を「 $[q_t^{(2)}]$ 」とし（図4にその概念を記載している。）、 $[q_t^{(2)}]$ に含まれる状態番号の集合を $S([q_t^{(2)}])$ とする。

【0102】

フレーム音素最適パススコア算出手段6082は、演算式格納手段6081から演算式の情報を読み出し、当該演算式に従って、パラメータの値を取得し、当該パラメータの値を読み出した演算式に代入し、スコアを算出する。さらに、具体的には、フレーム音素最適パススコア算出手段6082は、上記の数式10を読み出し、1番目のフレームから、T番目のフレームまで、フレームごとに、フレーム音素最適パススコアを算出する。つまり、フレーム音素最適パススコア算出手段6082は、フレームごとに、分母の値を算出する。フレーム音素最適パススコア算出手段6082が分母の値を算出する手順は、例えば、フレーム状態最適パススコア算出手段1082が数式1の分母の値を算出する手順

10

20

30

40

50



と同様である。

【0103】

次に、フレーム音素最適パススコア算出手段6082は、フレームごとに、数式10の分子の値を算出する。つまり、フレーム音素最適パススコア算出手段6082は、各フレームに対して、最適音素が有する1以上の状態(i)について、母語音素HMM格納部101から「 $a_{q_{t-1}}^{(1)}_i$ 」を読み出す。そして、フレーム音素最適パススコア算出手段6082は、母語音素HMM格納部101から「 $a_{i_{q_{t+1}}}^{(1)}$ 」を読み出す。そして、フレーム音素最適パススコア算出手段6082は、各フレームの特徴ベクトルを取得し、当該特徴ベクトルを出力密度関数に代入し、 $b_i(o_t)$ を得る。そして、フレーム音素最適パススコア算出手段6082は、得た「 $a_{q_{t-1}}^{(1)}_i$ 」、 $b_i(o_t)$ 、および「 $a_{i_{q_{t+1}}}^{(1)}$ 」、および $b_i(o_t)$ を乗算し、その結果を一時的にメモリに記憶する。かかる処理を最適音素が有する各状態に対して行う。そして、フレーム音素最適パススコア算出手段6082は、一時的にメモリに記憶した値の和を算出し、数式10の分子の値を得る。次に、フレーム音素最適パススコア算出手段6082は、一時記憶した分子の値から分母の値を除算し、フレームごとのフレーム状態最適パススコアを得る。

10

【0104】

次に、本発音評定装置の動作について説明する。本発音評定装置の処理は、実施の形態1の発音評定装置の図2のフローチャートにおける処理と比較して、ステップS211が「フレーム音素最適パススコアの算出」に変わるだけである。フレーム音素最適パススコアの算出処理について、図7のフローチャートを用いて説明する。図7のフローチャートにおいて、図3のフローチャートと同一のステップについて、説明を省略する。ただし、図3のフローチャートにおいてフレーム状態最適パススコア算出手段1082が行った処理は、図7のフローチャートにおいてはフレーム音素最適パススコア算出手段6082が行う。

20

【0105】

(ステップS701)フレーム音素最適パススコア算出手段6082は、カウンタiに1を代入する。

【0106】

(ステップS702)フレーム音素最適パススコア算出手段6082は、i番目の状態が存在するか否かを判断する。i番目の状態が存在すればステップS703に行き、i番目の状態が存在しなければステップS708に行く。

30

【0107】

(ステップS703)フレーム音素最適パススコア算出手段6082は、母語音素HMM格納部101から「 $a_{q_{t-1}}^{(1)}_i$ 」を読み出す。

【0108】

(ステップS704)フレーム音素最適パススコア算出手段6082は、母語音素HMM格納部101から「 $a_{i_{q_{t+1}}}^{(1)}$ 」を読み出す。

【0109】

(ステップS705)フレーム音素最適パススコア算出手段6082は、対応するフレームの特徴ベクトルを取得し、当該特徴ベクトルを出力密度関数に代入し、 $b_i(o_t)$ を得る。なお、特徴ベクトルは、特徴ベクトル系列取得部105が取得した情報である。また、出力密度関数の情報は、予め保持されており、かかる出力密度関数の情報を、フレーム音素最適パススコア算出手段6082は、読み出す。

40

【0110】

(ステップS706)フレーム音素最適パススコア算出手段6082は、ステップS703、ステップS704、およびステップS705で得た値を乗算し、その結果をメモリ上に一時格納する。

【0111】

(ステップS707)フレーム音素最適パススコア算出手段6082は、カウンタiを1、インクリメントし、ステップS702に戻る。

50

## 【 0 1 1 2 】

(ステップ S 7 0 8) フレーム音素最適パススコア算出手段 6 0 8 2 は、各状態に対応するステップ S 7 0 6 における演算結果を、メモリから読み出し、和を求め、その結果をメモリ上に一時格納する。

## 【 0 1 1 3 】

(ステップ S 7 0 9) フレーム音素最適パススコア算出手段 6 0 8 2 は、 $OPS_2(t)$  を算出する

## 【 0 1 1 4 】

(ステップ S 7 1 0) フレーム音素最適パススコア算出手段 6 0 8 2 は、ステップ S 7 0 9 で算出した  $OPS_2(t)$  をメモリ上に一時格納し、上位関数にリターンする。

10

## 【 0 1 1 5 】

以下、本実施の形態における発音評定装置の発音評定の方法について説明する。

## 【 0 1 1 6 】

まず、上記の数式 1 0 が導かれる過程について説明する。フレーム音素最適パススコアを以下の数式 1 1 で定義する。

## 【 数 1 1 】

$$OPS_2(t) = \frac{\sum_{i \in \mathcal{S}(A[q_t^{(2)}])} P(O, Q_1 \setminus \{q_t^{(1)}\}, q_t = i | \Lambda^{all})}{\sum_{j=1}^N P(O, Q_1 \setminus \{q_t^{(1)}\}, q_t = j | \Lambda^{all})}$$

20

## 【 0 1 1 7 】

そして、実施の形態 1 において数式 2 から数式 1 を算出した場合と同様に、分子と分母を約分すると、数式 1 0 が導出される。

## 【 0 1 1 8 】

また、数式 1 0 により算出されるフレーム音素最適パススコア ( $OPS_2(\tau)$ ) を発音区間  $T(m)$  で評価した発音区間フレーム音素最適パススコア ( $OPS_4(m)$ ) は、数式 1 2 により算出される。かかる算出は、代表値演算部 1 1 0 が行う。

## 【 数 1 2 】

$$OPS_4(m) = Me\{OPS_2(\tau) | \tau \in T(m)\}$$

30

## 【 0 1 1 9 】

数式 1 2 において、 $m$  は発音区間のインデクス、 $Me\{\cdot\}$  は集合  $\{\cdot\}$  内の要素の中央値を表わす。ただし中央値の代表値演算は、平均値や最大値などの別の代表値演算に置き換えても良い。また、発音区間は、例えば、forced alignment パス  $Q_2$  により定まる、音素区間、単語区間、フレーズ区間などである。

## 【 0 1 2 0 】

次に、数式 1 0 の性質、意義、つまり、上述したフレーム音素最適パススコアの性質、意義について説明する。

## 【 0 1 2 1 】

実施の形態 1 におけるフレーム状態最適パススコアの数式 7 において、最適状態 ( $q_t^{(2)}$ ) を含む音素 (最適音素) の中の各状態に対し、確率密度関数があまり変わらず、数式 7 の分母の値に対して「 $j = 1, 2, \dots, N$ 」の中で優勢となる状態が、最適状態のみならず、最適音素内の複数個の状態となる (図 8 参照)。

40

## 【 0 1 2 2 】

かかる場合、特徴ベクトル系列  $O$  が正しいトランスクリプションのネイティブ発音に類似している場合でも、状態 1 個からなる分子の確率値と比較して分母の確率値は複数状態の和となり (最適状態のみが優勢、となくなること)、分子に対する分母の割合が大きくなる。そのため、以下の数式 1 3 に示すように、スコアが高くなることが生じる。

50

## 【数 1 3】

$$\begin{aligned} OPS_1(t) &= \frac{P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_t^{(2)}, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})}{\sum_{j=1}^M P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})} \\ &\approx \frac{P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, q_t^{(2)}, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})}{\sum_{j \in S(\lambda[q_t^{(2)}])} P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})} \\ &\ll 1 \end{aligned}$$

10

## 【0 1 2 3】

そこで、最適状態 ( $q_t^{(2)}$ ) のみならず、最適音素 ( $[q_t^{(2)}]$ ) 内の複数状態で数式 1 あるいは数式 2 の分子の総和をとるフレーム音素最適パススコア (数式 1 0 あるいは数式 1 1) が、安定かつ精度の高いスコアを提供するものとして有意義である。また、実施の形態 1 における状態のもっともらしさよりも、母語音素のもっともらしさを評定した方が、発音の良し悪しをより安定して、精度高く測れるものと考えられるため、この意味でもフレーム音素最適パススコアは優れている。

## 【0 1 2 4】

以上、本実施の形態によれば、音韻毎に連結した連結 HMM とフレーム分割されたデータを用いて、比較対象の音声と入力音声の類似度を評定することができ、両者の類似度の評定の極めて精度を高く、高速に行うことができる。さらに具体的には、本実施の形態によれば、特徴ベクトル系列、母語音素 HMM、第一の最適状態番号系列、および第二の最適状態番号系列を用いて、区分されたフレームごとに最適音素内の 1 以上の各状態に対する確率を算出し、当該 1 以上の確率を用いて、区分されたフレームごとに発音の良し悪しを示すスコアを算出できる。このスコアは、母語音素 HMM が有する 1 以上の状態を有する音素と比較した、ユーザが入力した音声の各音素の音素らしさを示すスコアである。

20

## 【0 1 2 5】

なお、本実施の形態における発音評定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、音声の入力を受け付ける音声受付ステップと、前記音声受付ステップで受け付けた音声を、フレームに区分するフレーム区分ステップと、前記区分された各フレームの特徴ベクトルを算出し、特徴ベクトル系列を取得する特徴ベクトル系列取得ステップと、前記特徴ベクトル系列と格納している母語音素 HMM を取得し、前記特徴ベクトル系列と前記母語音素 HMM を用いて、ピタビアルゴリズムにより第一の最適状態番号系列を取得する第一最適状態番号系列取得ステップと、前記特徴ベクトル系列と格納している連結 HMM を取得し、前記特徴ベクトル系列と前記連結 HMM を用いて、ピタビアルゴリズムにより第二の最適状態番号系列を取得する第二最適状態番号系列取得ステップと、前記特徴ベクトル系列、前記母語音素 HMM、前記第一の最適状態番号系列、および前記第二の最適状態番号系列を用いて、前記区分されたフレームごとに最適音素内の 1 以上の各状態に対する確率を算出し、当該 1 以上の確率を用いて、前記区分されたフレームごとに発音の良し悪しを示すスコアを算出するフレーム音素最適パススコア算出ステップと、前記フレーム音素最適パススコア算出ステップで算出したフレームごとのスコア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。

30

40

## 【0 1 2 6】

また、上記プログラムに対して、コンピュータに、前記音声受付部が受け付けた音声の発音区間を取得する発音区間取得ステップと、前記発音区間ごとに、当該発音区間に対応する 1 以上のフレームごとの 1 以上のスコアを用いて、当該発音区間の代表のスコアを算出する代表値演算ステップをさらに実行させ、前記出力ステップにおいて、少なくとも前

50

記代表値演算ステップで算出した各発音区間の代表のスコアを出力する、ことは好適である。

【 0 1 2 7 】

また、上記プログラムに対して、コンピュータに、前記母語音素 H M M から、発音評定の対象である正解トランスクリプションに従って連結 H M M を構成する連結 H M M 構成ステップをさらに実行させ、前記連結 H M M は、前記連結 H M M 構成ステップで構成した連結 H M M である、ことは好適である。

【 0 1 2 8 】

また、上記各実施の形態において述べた発音評定装置は、語学学習やカラオケにおける採点装置や物真似練習などに利用できる。

10

【 0 1 2 9 】

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。

【 0 1 3 0 】

また、図 9 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の発音評定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 9 は、このコンピュータシステム 3 4 0 の概観図であり、図 1 0 は、コンピュータシステム 3 4 0 のブロック図である。

20

【 0 1 3 1 】

図 9 において、コンピュータシステム 3 4 0 は、F D ( F l e x i b l e D i s k ) ドライブ、C D - R O M ( C o m p a c t D i s k R e a d O n l y M e m o r y ) ドライブを含むコンピュータ 3 4 1 と、キーボード 3 4 2 と、マウス 3 4 3 と、モニタ 3 4 4 と、マイク 3 4 5 とを含む。

【 0 1 3 2 】

図 1 0 において、コンピュータ 3 4 1 は、F D ドライブ 3 4 1 1、C D - R O M ドライブ 3 4 1 2 に加えて、C P U ( C e n t r a l P r o c e s s i n g U n i t ) 3 4 1 3 と、C P U 3 4 1 3、C D - R O M ドライブ 3 4 1 2 及び F D ドライブ 3 4 1 1 に接続されたバス 3 4 1 4 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 R O M ( R e a d - O n l y M e m o r y ) 3 4 1 5 と、C P U 3 4 1 3 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 R A M ( R a n d o m A c c e s s M e m o r y ) 3 4 1 6 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 3 4 1 7 とを含む。ここでは、図示しないが、コンピュータ 3 4 1 は、さらに、L A N 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。

30

【 0 1 3 3 】

コンピュータシステム 3 4 0 に、上述した実施の形態の発音評定装置の機能を実行させるプログラムは、C D - R O M 3 5 0 1、または F D 3 5 0 2 に記憶されて、C D - R O M ドライブ 3 4 1 2 または F D ドライブ 3 4 1 1 に挿入され、さらにハードディスク 3 4 1 7 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 3 4 1 に送信され、ハードディスク 3 4 1 7 に記憶されても良い。プログラムは実行の際に R A M 3 4 1 6 にロードされる。プログラムは、C D - R O M 3 5 0 1、F D 3 5 0 2 またはネットワークから直接、ロードされても良い。

40

【 0 1 3 4 】

プログラムは、コンピュータ 3 4 1 に、上述した実施の形態の発音評定装置の機能を実行させるオペレーティングシステム（O S）、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 3 4 0 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。

50

## 【 0 1 3 5 】

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含ま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】

## 【 0 1 3 6 】

以上のように、本発明にかかる発音評定装置は、精度高く発音評定ができるという効果を有し、語学学習装置等として有用である。

## 【図面の簡単な説明】

## 【 0 1 3 7 】

【図 1】実施の形態 1 における発音評定装置のブロック図 10

【図 2】同発音評定装置の動作について説明するフローチャート

【図 3】同フレーム状態最適パススコア算出処理について説明するフローチャート

【図 4】同発音評定装置の処理の概念を示す図

【図 5】同フレーム状態最適パススコアの概念を説明する図

【図 6】実施の形態 2 における発音評定装置のブロック図

【図 7】同フレーム音素最適パススコア算出処理について説明するフローチャート

【図 8】同フレーム音素最適パススコアの概念を説明する図

【図 9】同発音評定装置を実現するコンピュータの外観図

【図 10】同発音評定システムのブロック図

【符号の説明】 20

## 【 0 1 3 8 】

1 0 1 母語音素 H M M 格納部

1 0 2 連結 H M M 格納部

1 0 3 音声受付部

1 0 4 フレーム区分部

1 0 5 特徴ベクトル系列取得部

1 0 6 第一最適状態番号系列取得部

1 0 7 第二最適状態番号系列取得部

1 0 8 フレーム状態最適パススコア算出部

1 0 9 発音区間取得部 30

1 1 0 代表値演算部

1 1 1 出力部

1 1 2 連結 H M M 構成部

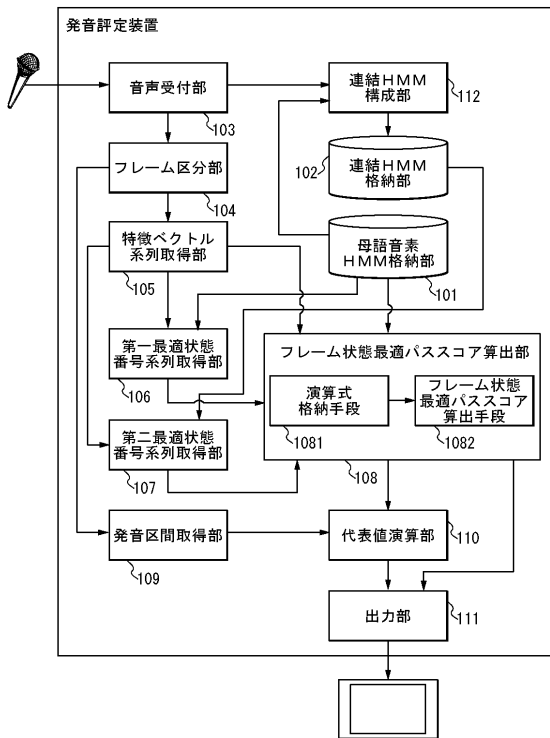
6 0 8 フレーム音素最適パススコア算出部

1 0 8 1、6 0 8 1 演算式格納手段

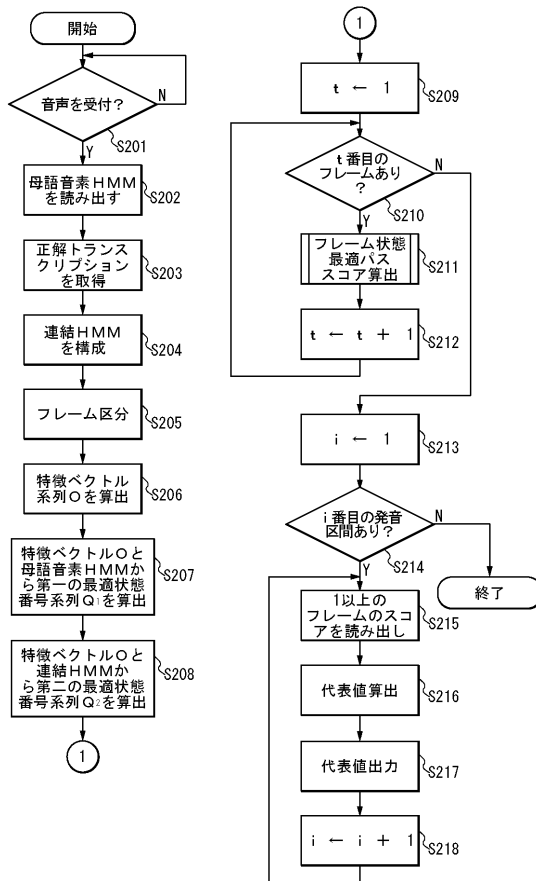
1 0 8 2 フレーム状態最適パススコア算出手段

6 0 8 2 フレーム音素最適パススコア算出手段

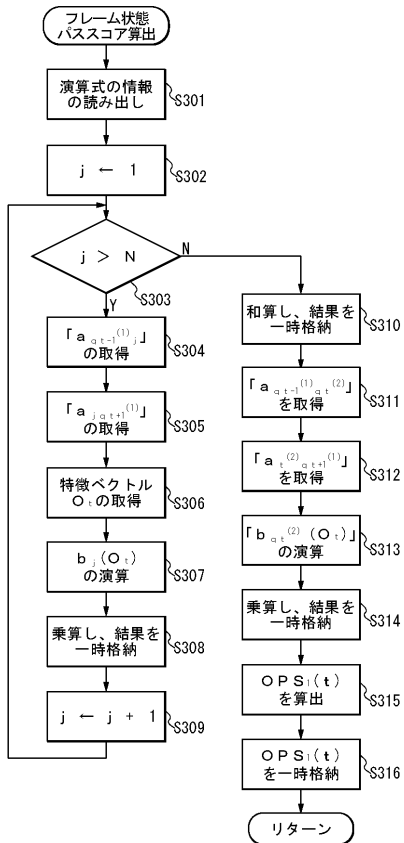
【図1】



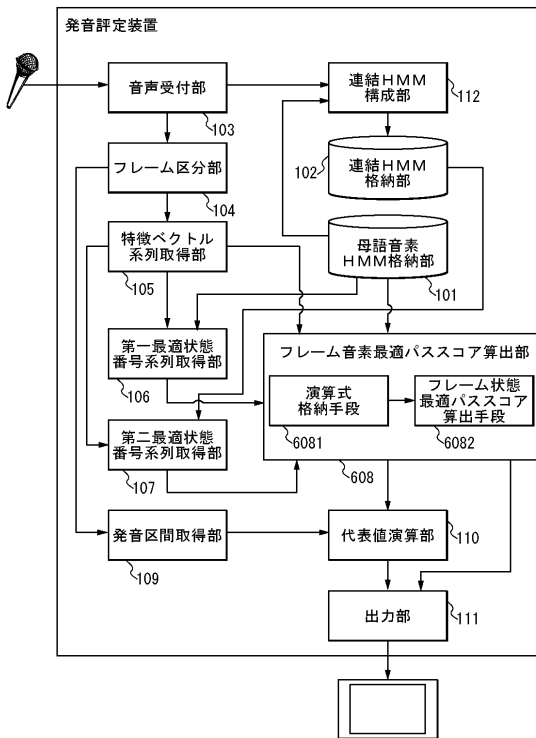
【図2】



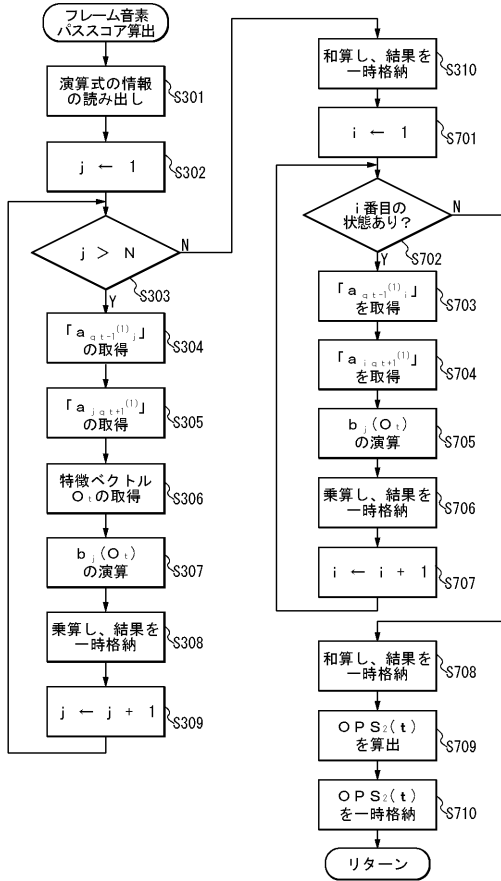
【図3】



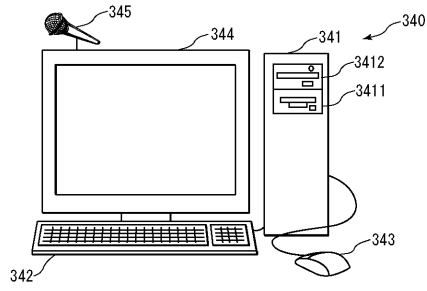
【図6】



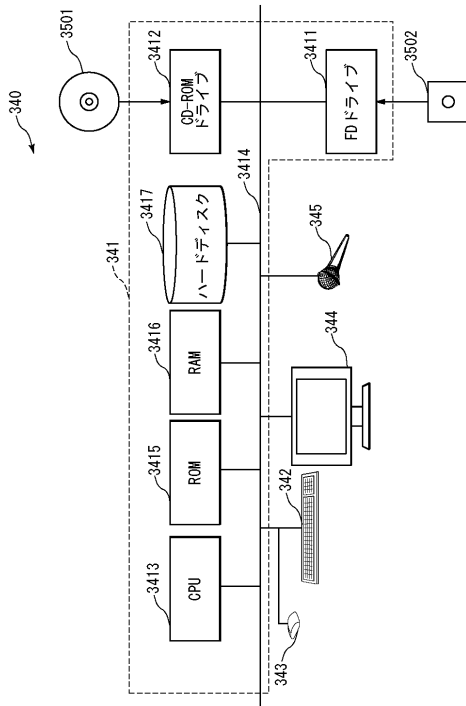
【図7】



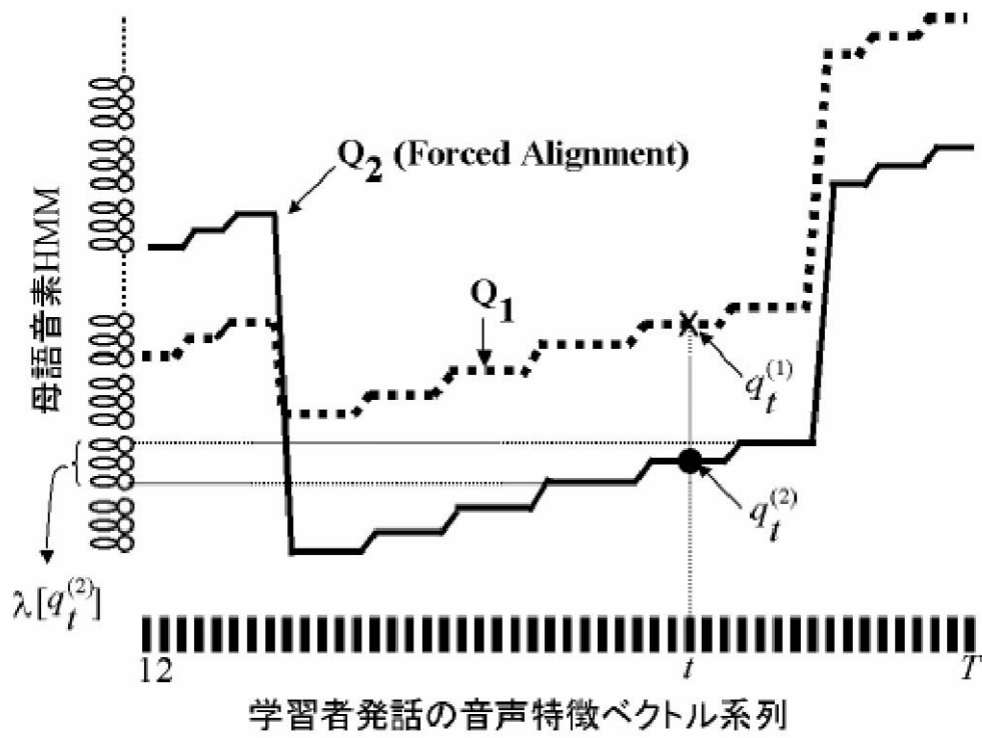
【図9】



【図10】



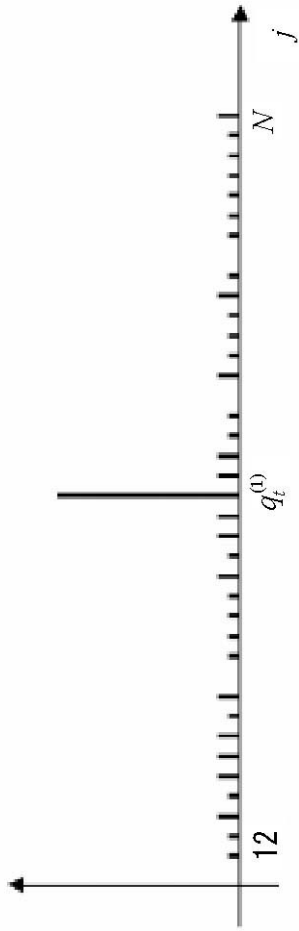
【 図 4 】





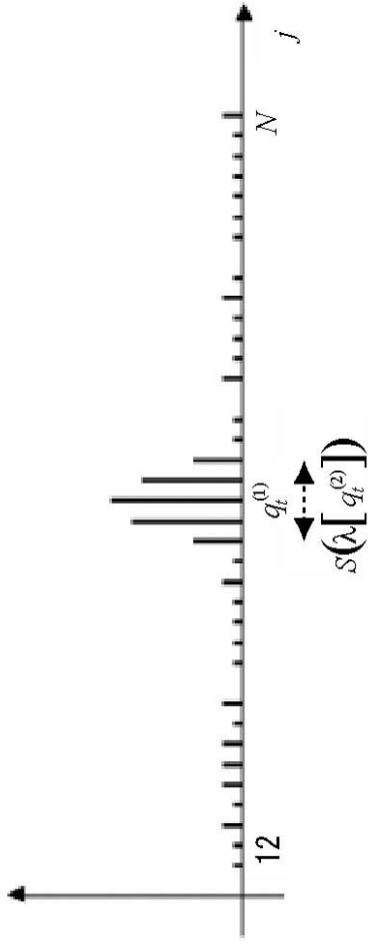
【 図 5 】

$$P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})$$



【 図 8 】

$$P(O, q_1^{(1)}, \dots, q_{t-1}^{(1)}, j, q_{t+1}^{(1)}, \dots, q_T^{(1)} | \Lambda^{all})$$



---

フロントページの続き

(72)発明者 山田 玲子

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

審査官 山下 剛史

(56)参考文献 特開2001-265211(JP,A)

特開2004-53652(JP,A)

特開平10-222190(JP,A)

特開2006-227587(JP,A)

前田直子他, "日本語・英語音素を用いた英単語発音評定方法の検討", 電子情報通信学会技術  
研究報告, Vol.101, No.604, SP2001-135(2002-01), pp.79-86

(58)調査した分野(Int.Cl., DB名)

G10L 11/00, 15/00 - 15/28

G09B 5/06, 19/00 - 19/06