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

特許第4775788号
(P4775788)

(45)発行日　平成23年9月21日(2011.9.21)　(24)登録日　平成23年7月8日(2011.7.8)

(51)Int.Cl.		F I		
G09B 19/00	(2006.01)	G09B	19/00	G
G09B 19/04	(2006.01)	G09B	19/04	
G10L 15/00	(2006.01)	G10L	15/00	200E
G10L 15/04	(2006.01)	G10L	15/04	300C
G10L 15/14	(2006.01)	G10L	15/14	200Z

請求項の数 4　(全 41 頁)

(21)出願番号　特願2005-16424(P2005-16424)
(22)出願日　平成17年1月25日(2005.1.25)
(65)公開番号　特開2006-227030(P2006-227030A)
(43)公開日　平成18年8月31日(2006.8.31)
審査請求日　平成19年8月3日(2007.8.3)
(31)優先権主張番号　特願2005-13157(P2005-13157)
(32)優先日　平成17年1月20日(2005.1.20)
(33)優先権主張国　日本国(JP)

(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
(74)代理人　100115749
弁理士　谷川　英和
(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
(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
(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

最終頁に続く

(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】
比較される対象の音声に関するデータであり、1以上の音韻毎のデータであり、フレーム毎に状態識別子と状態間を遷移することで得られる遷移確率の情報を有する教師データを１以上格納している教師データ格納手段と、
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、
前記区分されたフレーム毎の音声データであるフレーム音声データを得るフレーム音声データ取得部と、
前記フレーム毎のフレーム音声データに基づいて、音素の挿入を検知する特殊音声検知部と、
前記教師データと前記フレーム毎のフレーム音声データと前記特殊音声検知部における検知結果に基づいて、前記音声受付部が受け付けた音声の評定を行う評定部と、
前記評定部の評定結果を出力する出力部を具備し、
前記特殊音声検知部は、
前記フレーム音声データ取得部が得た複数のフレーム音声データから、１以上の音素を取得し、当該１以上のいずれかの音素について、一の音素の後半部および当該音素の次の音素の前半部の評定値が所定値より低い場合、または一の音素の所定区間以上の後半部および当該音素の次の音素の所定区間以上の前半部の評定値が所定値よりすべて低い場合に、音素の挿入を検知し、

前記評定部は、
１以上のフレーム音声データに対する、前記教師データの状態間を遷移することで得られる遷移確率の累積が最も高い経路の各状態である１以上の最適状態を決定する最適状態決定部と、
前記最適状態決定部が決定した１以上の最適状態の事後確率を示す確率値を取得する最適状態確率値取得部と、
前記最適状態確率値取得部が取得した確率値をパラメータとして音声の評定値を算出する評定値算出部とを具備し、
前記特殊音声検知部が音素の挿入を検知した場合に、少なくとも音素の挿入があった旨を示す評定結果を構成する発音評定装置。
【請求項２】
比較される対象の音声に関するデータであり、1以上の音韻毎のデータであり、フレーム毎に状態識別子と状態間を遷移することで得られる遷移確率の情報を有する教師データを１以上格納している教師データ格納手段と、
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、
前記区分されたフレーム毎の音声データであるフレーム音声データを得るフレーム音声データ取得部と、
前記フレーム毎のフレーム音声データに基づいて、音素の置換を検知する特殊音声検知部と、
前記教師データと前記フレーム毎のフレーム音声データと前記特殊音声検知部における検知結果に基づいて、前記音声受付部が受け付けた音声の評定を行う評定部と、
前記評定部の評定結果を出力する出力部を具備し、
前記特殊音声検知部は、
前記フレーム音声データ取得部が得た複数のフレーム音声データから、１以上の音素を取得し、当該１以上のいずれかの音素について、一の音素の評定値が所定値より低く、当該一の音素の直前の音素の評定値または直後の音素の評定値が所定の値より高い場合に、音素の置換を検知し、
前記評定部は、
１以上のフレーム音声データに対する、前記教師データの状態間を遷移することで得られる遷移確率の累積が最も高い経路の各状態である１以上の最適状態を決定する最適状態決定部と、
前記最適状態決定部が決定した１以上の最適状態の事後確率を示す確率値を取得する最適状態確率値取得部と、
前記最適状態確率値取得部が取得した確率値をパラメータとして音声の評定値を算出する評定値算出部とを具備し、
前記特殊音声検知部が音素の置換を検知した場合に、少なくとも音素の置換があった旨を示す評定結果を構成する発音評定装置。
【請求項３】
比較される対象の音声に関するデータであり、1以上の音韻毎のデータであり、フレーム毎に状態識別子と状態間を遷移することで得られる遷移確率の情報を有する教師データを１以上格納している教師データ格納手段と、
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、
前記区分されたフレーム毎の音声データであるフレーム音声データを得るフレーム音声データ取得部と、
前記フレーム毎のフレーム音声データに基づいて、音素の欠落を検知する特殊音声検知部と、
前記教師データと前記フレーム毎のフレーム音声データと前記特殊音声検知部における検知結果に基づいて、前記音声受付部が受け付けた音声の評定を行う評定部と、

前記評定部の評定結果を出力する出力部を具備し、
前記特殊音声検知部は、
前記フレーム音声データ取得部が得た複数のフレーム音声データから、１以上の音素を取得し、当該１以上のいずれかの音素について、一の音素の評定値が所定値より低く、当該一の音素の直前の音素の評定値または直後の音素の評定値が所定の値より高く、かつ当該音素の区間長が所定の長さよりも短い場合に、音素の欠落を検知し、
前記評定部は、
１以上のフレーム音声データに対する、前記教師データの状態間を遷移することで得られる遷移確率の累積が最も高い経路の各状態である１以上の最適状態を決定する最適状態決定部と、
前記最適状態決定部が決定した１以上の最適状態の事後確率を示す確率値を取得する最適状態確率値取得部と、
前記最適状態確率値取得部が取得した確率値をパラメータとして音声の評定値を算出する評定値算出部とを具備し、
前記特殊音声検知部が音素の欠落を検知した場合に、少なくとも音素の欠落があった旨を示す評定結果を構成する発音評定装置。
【請求項４】
前記評定値算出部は、
前記最適状態確率値取得部が取得した最適状態の確率値と、当該最適状態の確率値に対応するフレームの全状態における確率値の総和とをパラメータとして音声の評定値を算出する請求項１から請求項３いずれか記載の発音評定装置。
【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、特に、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】
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】
【０００３】
しかし、従来の技術においては、無音区間があれば、類似度が低く評価されると考えられ、評価の精度が低かった。また、音素の置換や挿入や欠落など、特殊な事象が発生していることを検知できなかった。
【０００４】
一方、一般に、入力される音声には、想定されている音韻列の順序に従わない音韻の挿入や置換、および本来あるはずの音韻の欠落が起こり得る。例えば、習得段階のネイティブではない学習者の発音には、正解の音声のデータにはない虚偽の音韻の挿入、他の音韻への置換、および本来なければならない音韻の欠落が起こる．また、学習者の発音には息継ぎなどの無声区間が多数存在する。従来の技術においては、スコアが低い場合、受け付けた音声の音韻列は正しいが、単に類似度が低下したのか、音韻の挿入や置換あるいは「ｇａｒｂａｇｅ」（どのカテゴリにも含まれない雑多な音韻）により低下したのかが判断

できない。さらに、従来の技術においては、ごく短区間のみスコアが低下している場合、類似度が低いのか、正しい音韻が欠落しているのか、あるいはｇａｒｂａｇｅが混入しているのかが判断できない。

【課題を解決するための手段】

【０００５】

本第一の発明の発音評定装置は、比較される対象の音声に関するデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータである教師データを１以上格納している教師データ格納部と、音声の入力を受け付ける音声受付部と、前記音声受付部で受け付けた音声を、フレームに区分するフレーム区分部と、前記区分されたフレーム毎の入力音声データを得る入力音声データ取得部と、前記フレーム毎の入力音声データに基づいて、特殊な音声が入力されたことを検知する特殊音声検知部と、前記教師データと前記入力音声データと前記特殊音声検知部における検知結果に基づいて、前記音声受付部が受け付けた音声の評定を行う評定部と、前記評定部の評定結果を出力する出力部を具備する発音評定装置である。

かかる発音評定装置は、特殊な音声が入力されたことを検知でき、当該検知結果に基づいて音声受付部が受け付けた音声の評定を行える。

【０００６】

また、本第二の発明の発音評定装置は、第一の発明の発音評定装置において、特殊音声検知部は、無音を示すＨＭＭに基づくデータである無音データを格納している無音データ格納手段と、前記入力音声データおよび前記無音データに基づいて、無音の区間を検出する無音区間検出手段を具備する発音評定装置である。

かかる構成により、無音の区間を検出でき、精度が高い音声の評定が可能となる。

【０００７】

また、本第三の発明の発音評定装置は、第一の発明の発音評定装置において、特殊音声検知部は、一の音素の後半部および当該音素の次の音素の前半部の評定値が所定の条件を満たすことを検知し、前記評定部は、前記特殊音声検知部が前記所定の条件を満たすことを検知した場合に、少なくとも音素の挿入があった旨を示す評定結果を構成する発音評定装置である。

かかる構成により、音素の挿入を検出でき、精度が高い音声の評定が可能となる。

【０００８】

また、本第四の発明の発音評定装置は、第一の発明の発音評定装置において、特殊音声検知部は、一の音素の評定値が所定の条件を満たすことを検知し、前記評定部は、前記特殊音声検知部が前記所定の条件を満たすことを検知した場合に、少なくとも音素の置換または欠落があった旨を示す評定結果を構成する発音評定装置である。

かかる構成により、音素の置換または欠落を検出でき、精度が高い音声の評定が可能となる。

【発明の効果】

【０００９】

本発明による発音評定装置によれば、無音、挿入、置換、欠落などの特殊な場合に対応した、発音の評定ができる。

【発明を実施するための最良の形態】

【００１０】

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やステップ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）

【００１１】

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。特に、本発音評定装置は、入力音声のフレームに対する最適状態の事後確率を、動的計画法を用いて算出することから、当該事後確率

をＤＡＰ（Ｄｙｎａｍｉｃ　Ａ　Ｐｏｓｔｅｒｉｏｒｉ　Ｐｒｏｂａｂｉｌｉｔｙ）と呼び、ＤＡＰに基づく類似度計算法および発音評定装置をＤＡＰＳと呼ぶ。

【００１２】

また、本実施の形態における発音評定装置は、例えば、語学学習や物真似練習などに利用できる。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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、評定部１０６、出力部１０７を具備する。評定部１０６は、最適状態決定手段１０６１、最適状態確率値取得手段１０６２、評定値算出手段１０６３を具備する。

【００１３】

入力受付部１０１は、発音評定装置の動作開始を指示する動作開始指示や、入力した音声の評定結果の出力態様の変更を指示する出力態様変更指示や、処理を終了する終了指示などの入力を受け付ける。かかる指示等の入力手段は、テンキーやキーボードやマウスやメニュー画面によるもの等、何でも良い。入力受付部１０１は、テンキーやキーボード等の入力手段のデバイスドライバーや、メニュー画面の制御ソフトウェア等で実現され得る。

【００１４】

教師データ格納部１０２は、教師データとして比較される対象の音声に関するデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータを１以上格納している。教師データは、音韻毎の隠れマルコフモデル（ＨＭＭ）を連結したＨＭＭに基づくデータであることが好適である。また、教師データは、入力される音声を構成する音素に対応するＨＭＭを、入力順序に従って連結されているＨＭＭに基づくデータであることが好適である。ただし、教師データは、必ずしも、音韻毎のＨＭＭを連結したＨＭＭに基づくデータである必要はない。教師データは、全音素のＨＭＭの、単なる集合であっても良い。また、教師データは、必ずしもＨＭＭに基づくデータである必要はない。教師データは、単一ガウス分布モデルや、確率モデル（ＧＭＭ：ガウシャンミクスチャモデル）や、統計モデルなど、他のモデルに基づくデータでも良い。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。また、ＨＭＭに基づくデータは、例えば、複数の学習対象言語を母国語として話す外国人が発声した２以上のデータから学習した（推定した）モデルでも良い。教師データ格納部１０２は、ハードディスクやＲＯＭなどの不揮発性の記録媒体が好適であるが、ＲＡＭなどの揮発性の記録媒体でも実現可能である。

【００１５】

音声受付部１０３は、音声の入力を受け付ける。音声受付部１０３は、例えば、マイクのドライバーソフトで実現され得る。また、なお、音声受付部１０３は、マイクとそのドライバーから実現されると考えても良い。音声は、マイクから入力されても良いし、磁気テープやＣＤ－ＲＯＭなどの記録媒体から読み出すことにより入力されても良い。

【００１６】

フレーム区分部１０４は、音声受付部１０３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１０４は、通常、ＭＰＵやメモリ等から実現され得る。フレーム区分部１０４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。

【００１７】

フレーム音声データ取得部１０５は、フレーム区分部１０４が区分したフレーム毎の音声データであるフレーム音声データを１以上得る。フレーム音声データ取得部１０５は、通常、ＭＰＵやメモリ等から実現され得る。フレーム音声データ取得部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。

【００１８】

評定部１０６は、教師データ格納部１０２の教師データと、フレーム音声データ取得部

１０５が取得した１以上のフレーム音声データに基づいて、音声受付部１０３が受け付けた音声の評定を行う。評定方法の具体例は、後述する。評定部１０６は、通常、ＭＰＵやメモリ等から実現され得る。評定部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。

【００１９】

出力部１０７は、評定部１０６の評定結果を出力する。出力部１０７の出力態様は、種々考えられる。出力とは、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、記録媒体への蓄積等を含む概念である。出力部１０７は、例えば、評定部１０６の評定結果を視覚的に表示する。出力部１０７は、例えば、フレーム単位、または／および音素・単語単位、または／および発声全体の評定結果を視覚的に表示する。出力部１０７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ても含まないと考えても良い。出力部１０７は、出力デバイスのドライバーソフトまたは、出力デバイスのドライバーソフトと出力デバイス等で実現され得る。

【００２０】

評定部１０６を構成している最適状態決定手段１０６１は、１以上のフレーム音声データのうちの少なくとも一のフレーム音声データに対する最適状態を決定する。最適状態決定手段１０６１は、例えば、全音韻ＨＭＭから、比較される対象（学習対象）の単語や文章などの音声を構成する1以上の音素に対応するＨＭＭを取得し、当該取得した1以上のＨＭＭを、音素の順序で連結したデータ（比較される対象の音声に関するデータであり、音韻毎の隠れマルコフモデルを連結したＨＭＭに基づくデータ）を構成する。そして、構成した当該データ、および取得した特徴ベクトル系列を構成する各特徴ベクトルo_tに基づいて、所定のフレームの最適状態（特徴ベクトルo_tに対する最適状態）を決定する。なお、最適状態を毛低するアルゴリズムは、例えば、Ｖｉｔｅｒｂｉアルゴリズムである。また、教師データは、上述の比較される対象の音声に関するデータであり、音韻毎の隠れマルコフモデルを連結したＨＭＭに基づくデータと考えても良いし、連結される前のデータであり、全音韻ＨＭＭのデータと考えても良い。

最適状態確率値取得手段１０６２は、最適状態決定手段１０６１が決定した最適状態における確率値を取得する。

【００２１】

評定値算出手段１０６３は、最適状態確率値取得手段１０６２が取得した確率値をパラメータとして音声の評定値を算出する。評定値算出手段１０６３は、上記確率値を如何に利用して、評定値を算出するかは問わない。評定値算出手段１０６３は、例えば、最適状態確率値取得手段１０６２が取得した確率値と、当該確率値に対応するフレームの全状態における確率値の総和をパラメータとして音声の評定値を算出する。評定値算出手段１０６３は、ここでは、通常、フレームごとに評定値を算出する。

【００２２】

最適状態決定手段１０６１、最適状態確率値取得手段１０６２、評定値算出手段１０６３は、通常、ＭＰＵやメモリ等から実現され得る。最適状態決定手段１０６１等の処理手順は、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。

次に、本発音評定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。

【００２３】

（ステップＳ２０１）入力受付部１０１は、発音評定装置の動作開始を指示する動作開始指示を受け付けたか否かを判断する。動作開始指示を受け付ければステップＳ２０２に行き、動作開始指示を受け付けなければステップＳ２１４に飛ぶ。

（ステップＳ２０２）音声受付部１０３は、音声の入力を受け付けたか否かを判断する。音声の入力を受け付ければステップＳ２０３に行き、音声の入力を受け付けなければステップＳ２１３に飛ぶ。

（ステップＳ２０３）フレーム区分部１０４は、ステップＳ２０２で受け付けた音声のデータを図示しないバッファに一時格納する。

【００２４】

（ステップＳ２０４）フレーム区分部１０４は、バッファに一時格納した音声データをフレームに区分する。かかる段階で、区分されたフレーム毎の音声データであるフレーム音声データが構成されている。フレーム区分部１０４が行うフレーム分割の処理は、例えば、フレーム音声データ取得部１０５がフレーム音声データを取り出す際の前処理であり、入力された音声のデータを、すべてのフレームに一度に分割するとは限らない。

（ステップＳ２０５）フレーム音声データ取得部１０５は、カウンタｉに１を代入する。

【００２５】

（ステップＳ２０６）フレーム音声データ取得部１０５は、ｉ番目のフレームが存在するか否かを判断する。ｉ番目のフレームが存在すればステップＳ２０７に行き、ｉ番目のフレームが存在しなければステップＳ２０８に行く。

【００２６】

（ステップＳ２０７）フレーム音声データ取得部１０５は、ｉ番目のフレーム音声データを取得する。フレーム音声データの取得とは、例えば、当該分割された音声データを音声分析し、特徴ベクトルデータを抽出することである。なお、フレーム音声データは、例えば、入力された音声データをフレーム分割されたデータである。また、フレーム音声データは、例えば、当該分割された音声データから音声分析され、抽出された特徴ベクトルデータを有する。本特徴ベクトルデータは、例えば、三角型フィルタを用いたチャネル数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。

（ステップＳ２０８）フレーム音声データ取得部１０５は、カウンタｉを１、インクリメントする。ステップＳ２０６に戻る。

【００２７】

（ステップＳ２０９）最適状態決定手段１０６１は、全フレームの最適状態を決定する。最適状態決定手段１０６１が最適状態を決定するアルゴリズムは、例えば、Ｖｉｔｅｒｂｉアルゴリズムによる。Ｖｉｔｅｒｂｉアルゴリズムは、公知のアルゴリズムであるので、詳細な説明は省略する。

【００２８】

（ステップＳ２１０）最適状態確率値取得手段１０６２は、全フレームの全状態の前向き尤度、および後向き尤度を算出する。最適状態確率値取得手段１０６２は、例えば、全てのＨＭＭを用いて、フォワード・バックワードアルゴリズムにより、前向き尤度、および後向き尤度を算出する。

（ステップＳ２１１）最適状態確率値取得手段１０６２は、ステップＳ２１０で取得した前向き尤度、および後向き尤度を用いて、最適状態の確率値（最適状態確率値）を、すべて算出する。

【００２９】

（ステップＳ２１２）評定値算出手段１０６３は、ステップＳ２１１で算出した1以上の最適状態確率値から、1以上のフレームの音声の評定値を算出する。評定値算出手段１０６３が評定値を算出する関数は問わない。評定値算出手段１０６３は、例えば、取得した最適状態確率値と、当該最適状態確率値に対応するフレームの全状態における確率値の総和をパラメータとして音声の評定値を算出する。詳細については、後述する。

【００３０】

（ステップＳ２１３）出力部１０７は、ステップＳ２１２における評定結果（ここでは、音声の評定値）を、設定されている出力モードに従って、出力する。ステップＳ２０２に戻る。出力モードとは、評定値を数値で画面に表示するモード、評定値の遷移をグラフで画面に表示するモード、評定値を音声で出力するモード、評定値が所定の数値より低い

場合に警告を示す情報を表示するモードなど、何でも良い。なお、ここでの出力モードは、ステップＳ２１４で設定されるモードである。

【００３１】

（ステップＳ２１４）音声受付部１０３は、タイムアウトか否かを判断する。つまり、音声受付部１０３は、所定の時間以上、音声の入力を受け付けなかったか否かを判断する。タイムアウトであればステップＳ２０１に戻り、タイムアウトでなければステップＳ２１４に戻る。

【００３２】

（ステップＳ２１５）入力受付部１０１は、出力態様変更指示を受け付けたか否かを判断する。出力態様変更指示を受け付ければステップＳ２１６に行き、出力態様変更指示を受け付なければステップＳ２１７に飛ぶ。出力態様変更指示は、上述した出力モードを有する情報である。

（ステップＳ２１６）出力部１０７は、ステップＳ２１５で受け付けた出力態様変更指示が有する出力モードを示す情報を書き込み、出力モードを設定する。ステップＳ２０１に戻る。

（ステップＳ２１７）入力受付部１０１は、終了指示を受け付けたか否かを判断する。終了指示を受け付ければ処理を終了し、終了指示を受け付なければステップＳ２０１に戻る。

なお、図２のフローチャートにおいて、本発音評定装置は、出力モードの設定機能を有しなくても良い。

以下、本実施の形態における発音評定装置の具体的な動作について説明する。本具体例において、発音評定装置が語学学習に利用される場合について説明する。

【００３３】

まず、本発音評定装置において、図示しない手段により、ネイティブ発音の音声データベースからネイティブ発音の音韻ＨＭＭを学習しておく。ここで、音韻の種類数をＬとし、ｌ番目の音韻に対するＨＭＭをλ_lとする。なお、かかる学習の処理については、公知技術であるので、詳細な説明は省略する。なお、ＨＭＭの仕様について、図３に示す。なお、ＨＭＭの仕様は、他の実施の形態における具体例の説明においても同様である。ただし、ＨＭＭの仕様が、他の仕様でも良いことは言うまでもない。

【００３４】

そして、学習したＬ種類の音韻ＨＭＭから、学習対象の単語や文章などの音声を構成する１以上の音素に対応するＨＭＭを取得し、当該取得した１以上のＨＭＭを、音素の順序で連結した教師データを構成する。そして、当該教師データを教師データ格納部１０２に保持しておく。ここでは、例えば、比較される対象の音声は、単語「ｒｉｇｈｔ」の音声である。

次に、学習者が、語学学習の開始の指示である動作開始指示を入力する。かかる指示は、例えば、マウスで所定のボタンを押下することによりなされる。

次に、学習者は、学習対象の音声「ｒｉｇｈｔ」を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。

次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。なお、フレームの間隔は、予め決められている、とする。

【００３５】

そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「$O=o_1, o_2, \cdots, o_T$」を算出する。なお、Ｔは、系列長である。ここで、特徴ベクトル系列は、各フレームの特徴ベクトルの集合である。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。なお、音声分析条

件を図４の表に示す。なお、音声分析条件は、他の実施の形態における具体例の説明においても同様である。ただし、音声分析条件が、他の条件でも良いことは言うまでもない。

【００３６】

次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルo_tに基づいて、所定のフレームの最適状態（特徴ベクトルo_tに対する最適状態）を決定する。最適状態決定手段１０６１が最適状態を決定するアルゴリズムは、例えば、Ｖｉｔｅｒｂｉアルゴリズムによる。かかる場合、最適状態決定手段１０６１は、上記で連結したＨＭＭを用いて最適状態を決定する。最適状態決定手段１０６１は、２以上のフレームの最適状態である最適状態系列を求めることとなる。

【００３７】

次に、最適状態確率値取得手段１０６２は、以下の数式１により、最適状態における最適状態確率値（$\gamma_t(q_t^*)$）を算出する。なお、$\gamma_t(q_t^*)$は、状態ｊの事後確率関数$\gamma_t(j)$のｊにq_t^*を代入した値である。そして、状態ｊの事後確率関数$\gamma_t(j)$は、数式２を用いて算出される。この確率値（$\gamma_t(j)$）は、ｔ番目の特徴ベクトルo_tが状態ｊから生成された事後確率であり、動的計画法を用いて算出される。なお、ｊは、状態を識別する状態識別子である。

【数１】

$$DAP(t) = \gamma_t(q_t^*)$$

数式１において、q_tは、o_tに対する状態識別子を表す。この確率値（$\gamma_t(j)$）は、ＨＭＭの最尤推定におけるＢａｕｍ－Ｗｅｌｃｈアルゴリズムの中で表れる占有度数に対応する。

【数２】

$$\gamma_t(j) = \frac{Pr(q_t=j,\ \mathbf{O} \mid \lambda^{all})}{Pr(\mathbf{O} \mid \lambda^{all})}$$

$$= \frac{Pr(q_t=j,\ \mathbf{O} \mid \lambda^{all})}{\sum_{k=1}^{N} Pr(q_t=k,\ \mathbf{O} \mid \lambda^{all})} \quad (1)$$

$$= \frac{\alpha_t(j)\,\beta_t(j)}{\sum_{k=1}^{N} \alpha_t(k)\,\beta_t(k)} \quad (2)$$

$$\alpha_t(j) = Pr(q_t=j,\ \{\mathbf{o}_1,\ \mathbf{o}_2,\ \cdots\cdots,\ \mathbf{o}_t\} \mid \lambda^{all})$$

$$\beta_t(j) = Pr(\{\mathbf{o}_{t+1},\ \mathbf{o}_{t+2},\ \cdots\cdots,\ \mathbf{o}_T\} \mid q_t=j,\ \lambda^{all})$$

数式２は、数式１を変形したものである。

【００３８】

数式２において、「αｔ（ｊ）」「βｔ（ｊ）」は、全部のＨＭＭを用いて、ｆｏｒｗａｒｄ－ｂａｃｋｗａｒｄアルゴリズムにより算出される。「αｔ（ｊ）」は前向き尤度、「βｔ（ｊ）」は後向き尤度である。Ｂａｕｍ－Ｗｅｌｃｈアルゴリズム、ｆｏｒｗａｒｄ－ｂａｃｋｗａｒｄアルゴリズムは、公知のアルゴリズムであるので、詳細な説明は省略する。

また、数式２において、Ｎは、全ＨＭＭに渡る状態の総数を示す。

【００３９】

なお、評定部１０６は、まず最適状態を求め、次に、最適状態の確率値（なお、確率値は、０以上、１以下である。）を求めても良いし、評定部１０６は、まず、全状態の確率値を求め、その後、特徴ベクトル系列の各特徴ベクトルに対する最適状態を求め、当該最適状態に対応する確率値を求めても良い。

【００４０】

次に、評定値算出手段１０６３は、例えば、上記の取得した最適状態確率値と、当該最適状態確率値に対応するフレームの全状態における確率値の総和をパラメータとして音声の評定値を算出する。かかる場合、もし学習者のｔフレーム目に対応する発声が、教師データが示す発音（例えば、正しいネイティブな発音）に近ければ、数式２の（２）式の分子の値が、他の全ての可能な音韻の全ての状態と比較して大きくなり、結果的に最適状態の確率値（評定値）が大きくなる。逆にその区間が、教師データが示す発音に近くなければ、評定値は小さくなる。なお、どのネイティブ発音にも近くないような場合は、評定値はほぼ１／Ｎに等しくなる。Ｎは全ての音韻ＨＭＭにおける全ての状態の数であるから、通常、大きな値となり、この評定値は十分小さくなる。また、ここでは、評定値は最適状態における確率値と全ての可能な状態における確率値との比率で定義されている。したがって、話者性や収音環境の違いにより多少のスペクトルの変動があったとしても、学習者が正しい発音をしていれば、その変動が相殺され評定値が高いスコアを維持する。よって、評定値算出手段１０６３は、最適状態確率値取得手段１０６２が取得した確率値と、当該確率値に対応するフレームの全状態における確率値の総和をパラメータとして音声の評定値を算出することは、極めて好適である。

【００４１】

かかる評定値算出手段１０６３が算出した評定値（「ＤＡＰスコア」とも言う。）を、図５、図６に示す。図５、図６において、横軸は分析フレーム番号、縦軸はスコアを％で表わしたものである。太い破線は音素境界，細い点線は状態境界（いずれもＶｉｔｅｒｂｉアルゴリズムで求まったもの）を表わしており，図の上部に音素名を表記している。図５は、アメリカ人男性による英語「ｒｉｇｈｔ」の発音のＤＡＰスコアを示す。なお、評定値を示すグラフの横軸、縦軸は、後述するグラフにおいても同様である。

【００４２】

図６は、日本人男性による英語「ｒｉｇｈｔ」の発音のＤＡＰスコアを示す。アメリカ人の発音は、日本人の発音と比較して、基本的にスコアが高い。また、図５において、状態の境界において所々スコアが落ち込んでいることがわかる。

【００４３】

そして、出力部１０７は、評定部１０６の評定結果を出力する。具体的には、例えば、出力部１０７は、図７に示すような態様で、評定結果を出力する。つまり、出力部１０７は、各フレームにおける発音の良さを表すスコア（スコアグラフ）として、各フレームの評定値を表示する。その他、出力部１０７は、学習対象の単語の表示（単語表示）、音素要素の表示（音素表示）、教師データの波形の表示（教師波形）、学習者の入力した発音の波形の表示（ユーザ波形）を表示しても良い。なお、図７において、「録音」ボタンを押下すれば、動作開始指示が入力されることとなり、「停止」ボタンを押下すれば、終了指示が入力されることとなる。なお、本発音評定装置は、学習対象の単語（図７の「ｗｏｒｄ１」など）や、音素（図７の「ｐ１」など）や、教師波形を出力されるためのデータを予め格納している、とする。

【００４４】

また、図７において、フレーム単位以外に、音素単位、単語単位、発声全体の評定結果を表示しても良い。上記の処理において、フレーム単位の評定値を算出するので、単語単位、発声全体の評定結果を得るためには、フレーム単位の１以上の評定値をパラメータとして、単語単位、発声全体の評定値を算出する必要がある。かかる算出式は問わないが、例えば、単語を構成するフレーム単位の１以上の評定値の平均値を単語単位の評定値とする、ことが考えられる。

【００４５】

なお、図７において、発音評定装置は、波形表示（教師波形またはユーザ波形）の箇所においてクリックを受け付けると、再生メニューを表示し、音素区間内ではその音素またはその区間が属する単語、波形全体を再生し、単語区間外（無音部）では波形全体のみを再生するようにしても良い。

また、出力部１０７の表示は、図８に示すような態様でも良い。図８において、音素ごとのスコア、単語のスコア、総合スコアが、数字で表示されている。

なお、出力部１０７の表示は、図５、図６のような表示でも良いことは言うまでもない。

以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。

【００４６】

また、本実施の形態によれば、連結されたＨＭＭである連結ＨＭＭを用いて最適状態を求め、評定値を算出するので、高速に評定値を求めることができる。したがって、上記の具体例で述べたように、リアルタイムに、フレームごと、音素ごと、単語ごとの評定値を出力できる。また、本実施の形態によれば、動的計画法に基づいた事後確率を確率値として算出するので、さらに高速に評定値を求めることができる。また、本実施の形態によれば、フレームごとに確率値を算出するので、上述したように、フレーム単位だけではなく、または／および音素・単語単位、または／および発声全体の評定結果を出力でき、出力態様の自由度が高い。

【００４７】

また、本実施の形態によれば、発音評定装置は、語学学習に利用することを主として説明したが、物真似練習などに利用できる。つまり、本発音評定装置は、比較される対象の音声に関するデータとの類似度を精度良く、高速に評定し、出力でき、そのアプリケーションは問わない。

【００４８】

また、本実施の形態において、音声の入力を受け付けた後または停止ボタン操作後に、スコアリング処理を実行するかどうかをユーザに問い合わせ、スコアリング処理を行うとの指示を受け付けた場合のみ、図８に示すような音素スコア、単語スコア、総合スコアを出力するようにしても良い。

【００４９】

また、本実施の形態において、教師データは、比較される対象の音声に関するデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータであるとして、主として説明したが、必ずしもＨＭＭに基づくデータである必要はない。教師データは、単一ガウス分布モデルや、確率モデル（ＧＭＭ：ガウシャンミクスチャモデル）や統計モデルなど、他のモデルに基づくデータでも良い。かかることは、他の実施の形態においても同様である。

【００５０】

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、前記区分されたフレーム毎の音声データであるフレーム音声データを1以上得るフレーム音声データ取得ステップと、格納されているデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータである教師データと前記1以上のフレーム音声データに基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【００５１】

また、上記プログラムにおいて、前記教師データは、音韻毎の隠れマルコフモデル（ＨＭＭ）を連結したＨＭＭに基づくデータであり、前記評定ステップは、前記1以上のフレーム音声データのうちの少なくとも一の最適状態を決定する最適状態決定サブステップと、前記最適状態決定サブステップで決定した最適状態における確率値を取得する最適状態確率値取得サブステップと、前記最適状態確率値取得サブステップで取得した確率値をパラメータとして音声の評定値を算出する評定値算出サブステップを具備するプログラム、である。

【００５２】

なお、上記プログラムにおいて、評定値算出サブステップにおいて、前記最適状態確率値取得サブステップで取得した確率値と、当該確率値に対応するフレームの全状態における確率値の総和をパラメータとして音声の評定値を算出することは好適である。

（実施の形態２）

【００５３】

本実施の形態における発音評定装置は、実施の形態１の発音評定装置と比較して、評定部における評定アルゴリズムが異なる。本実施の形態において、評定値は、各フレームにおける、すべての音韻の中で最適な音韻の事後確率（確率値）を表すように算出される。本実施の形態における発音評定装置が算出する事後確率を、実施の形態１におけるＤＡＰに対してｐ－ＤＡＰと呼ぶ。

【００５４】

図９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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、評定部９０６、出力部１０７を具備する。評定部９０６は、最適状態決定手段１０６１、音韻確率値取得手段９０６２、評定値算出手段９０６３を具備する。

音韻確率値取得手段９０６２は、最適状態決定手段１０６１が決定した最適状態を有する音韻全体の状態における１以上の確率値を取得する。ここで１以上の確率値とは、１つ以上の確率値の意味である。

【００５５】

評定値算出手段９０６３は、音韻確率値取得手段９０６２が取得した１以上の確率値をパラメータとして音声の評定値を算出する。評定値算出手段９０６３は、例えば、音韻確率値取得手段９０６２が取得した１以上の確率値の総和をパラメータとして音声の評定値を算出する。

【００５６】

音韻確率値取得手段９０６２、および評定値算出手段９０６３は、通常、ＭＰＵやメモリ等から実現され得る。音韻確率値取得手段９０６２等の処理手順は、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。

次に、本発音評定装置の動作について図１０のフローチャートを用いて説明する。図１０のフローチャートにおいて、図２と異なるステップについてのみ説明する。

【００５７】

（ステップＳ１００１）音韻確率値取得手段９０６２は、全フレームの全状態の前向き尤度と後向き尤度を算出する。そして、全フレーム、全状態の確率値を得る。具体的には、音韻確率値取得手段９０６２は、例えば、各特徴ベクトルが対象の状態から生成された事後確率を算出する。この事後確率は、ＨＭＭの最尤推定におけるＢａｕｍ－Ｗｅｌｃｈアルゴリズムの中で現れる占有度数に対応する。Ｂａｕｍ－Ｗｅｌｃｈアルゴリズムは、公知のアルゴリズムであるので、説明は省略する。

（ステップＳ１００２）音韻確率値取得手段９０６２は、全フレームの最適状態確率値を算出する。

（ステップＳ１００３）音韻確率値取得手段９０６２は、カウンタｉに１を代入する。

【００５８】

（ステップＳ１００４）音韻確率値取得手段９０６２は、ｉ番目の最適状態が存在するか否かを判断する。ｉ番目の最適状態が存在すればステップＳ１００５に行き、ｉ番目の最適状態が存在しなければステップＳ２０２に戻る。

（ステップＳ１００５）音韻確率値取得手段９０６２は、ｉ番目の最適状態を含む音韻全体の確率値をすべて取得する。

【００５９】

（ステップＳ１００６）評定値算出手段９０６３は、ステップＳ１００５で取得した１以上の確率値に基づいて、音声の評定値を算出する。評定値算出手段９０６３は、例えば、音韻確率値取得手段９０６２が取得した１以上の確率値の総和をパラメータとして音声の評定値を算出する。

（ステップＳ１００７）出力部１０７は、ステップＳ１００６で算出した評定値を出力する。

（ステップＳ１００８）音韻確率値取得手段９０６２は、カウンタｉを１、インクリメントする。ステップＳ１００４に戻る。

以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、評定値の算出アルゴリズムが実施の形態１とは異なるので、その動作を中心に説明する。

【００６０】

まず、学習者が、語学学習の開始の指示である動作開始指示を入力した後、学習対象の音声「ｒｉｇｈｔ」を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。

そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「$O=o_1, o_2, \cdots, o_T$」を算出する。

次に、音韻確率値取得手段９０６２は、各フレームの各状態の事後確率（確率値）を算出する。確率値の算出は、上述した数式１、数式２により算出できる。

【００６１】

次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルo_tに基づいて、各フレームの最適状態（特徴ベクトルo_tに対する最適状態）を決定する。つまり、最適状態決定手段１０６１は、最適状態系列を得る。

【００６２】

次に、音韻確率値取得手段９０６２は、フレーム毎に、当該フレームに対応する最適状態を含む音韻全体の確率値をすべて取得する。そして、評定値算出手段９０６３は、上記取得した１以上の確率値に基づいて、音声の評定値を算出する。具体的には、評定値算出手段９０６３は、数式３により評定値を算出する。

【数３】

$$\text{p-DAP}(t) \triangleq \sum_{j \in P(q_t^*)} \gamma_t(j)$$

なお、数式３において、Ｐ（ｉ）は、ｉ番目の状態を有しているＨＭＭの持つ全状態の集合を示す。

【００６３】

かかる評定値算出手段９０６３が算出した評定値（「ｐ－ＤＡＰスコア」とも言う。）を、図１１、図１２に示す。図１１は、アメリカ人男性による英語「ｒｉｇｈｔ」の発音のｐ－ＤＡＰスコアを示す。図１２は、日本人男性による英語「ｒｉｇｈｔ」の発音のｐ－ＤＡＰスコアを示す。アメリカ人の発音は、日本人の発音と比較して、基本的にスコアが高い。また、図１１において、音素境界でスコアの落ち込みがあるものの、ｐ－ＤＡＰは本来発音の良好なアメリカ人発音に対して，高いスコアをＤＡＰより安定して出力して

いることがわかる。かかる判断は、図５のグラフと図１１のグラフを比較して判断できる。また、図１１において、音素／ｒ／のスコアが低いが，この発音を聴いてみたところ／ｒ／の発声が若干不明瞭であった。

【００６４】

そして、出力部１０７は、算出したフレームごとの評定値を、順次出力する。かかる出力例は、図７または図８である。なお、出力部１０７は、図１１、図１２のようなグラフを出力しても良いことは言うまでもない。

以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。

【００６５】

また、本実施の形態によれば、連結されたＨＭＭである連結ＨＭＭを用いて最適状態を求め、評定値を算出するので、高速に評定値を求めることができる。したがって、上記の具体例で述べたように、リアルタイムに、フレームごと、音素ごと、単語ごとの評定値を出力できる。また、本実施の形態によれば、動的計画法に基づいた事後確率を確率値として算出するので、さらに高速に評定値を求めることができる。また、本実施の形態によれば、フレームごとに確率値を算出するので、上述したように、フレーム単位だけではなく、音素・単語単位、または／および発声全体の評定結果を出力でき、出力態様の自由度が高い。

【００６６】

また、本実施の形態によれば、評定値を、各フレームにおいて、すべての音韻の中で最適な音韻の事後確率（確率値）を表しており、実施の形態１におけるような状態単位のＤＡＰと比較して、本来、測定したい類似度を精度良く、安定して求めることができる。つまり、実施の形態１において、ＤＡＰは、全ての可能な状態に対する最適状態の事後確率を計算する。そして、総状態数Ｎは、通常、非常に大きくなり、入力音声によっては、評定値（ＤＡＰのスコア）が大きく低下する。つまり、例えば、あるフレームが存在する音韻内の２つの状態の過渡部にそのフレームが対応してしまえば、評定値が小さくなる。一方、音素に対する類似性を求める本実施の形態によれば、教師データの音韻との類似度（状態との類似度ではない）を測ることができ、好適である。

【００６７】

さらに、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、前記区分されたフレーム毎の音声データであるフレーム音声データを1以上得るフレーム音声データ取得ステップと、格納されているデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータである教師データと前記1以上のフレーム音声データに基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【００６８】

また、上記プログラムにおいて、前記教師データは、音韻毎の隠れマルコフモデル（ＨＭＭ）を連結したＨＭＭに基づくデータであり、前記評定ステップは、前記1以上のフレーム音声データのうちの少なくとも一の最適状態を決定する最適状態決定サブステップと、前記最適状態決定サブステップで決定した最適状態を有する音韻全体の状態における１以上の確率値を取得する音韻確率値取得サブステップと、前記音韻確率値取得サブステップで取得した１以上の確率値をパラメータとして音声の評定値を算出する評定値算出サブステップを具備するプログラム、である。さらに、上記プログラムにおける評定値算出サブステップにおいて、前記音韻確率値取得サブステップで取得した１以上の確率値の総和をパラメータとして音声の評定値を算出することは好適である。これは、音韻確率値取得サブステップで取得した１以上の確率値の総和を音声の評定値とすることも含む。

（実施の形態３）

【００６９】
本実施の形態における発音評定装置は、実施の形態１、２の発音評定装置と比較して、評定部における評定アルゴリズムが異なる。本実施の形態において、評定値は、発音区間ごとに算出される。本実施の形態における発音評定装置が算出する事後確率を、実施の形態１におけるＤＡＰに対してｔ－ＤＡＰと呼ぶ。
【００７０】
図１３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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、評定部１３０６、出力部１０７を具備する。評定部１３０６は、最適状態決定手段１０６１、発音区間確率値取得手段１３０６２、評定値算出手段１３０６３を具備する。
発音区間確率値取得手段１３０６２は、最適状態決定手段１０６１が決定した最適状態の確率値を、発音区間毎に取得する。ここで、発音区間とは、音韻、音節、単語など、発音の一まとまりを構成する区間である。
【００７１】
評定値算出手段１３０６３は、発音区間確率値取得手段１３０６２が取得した１以上の発音区間毎の１以上の確率値をパラメータとして音声の評定値を算出する。評定値算出手段１３０６３は、例えば、発音区間確率値取得手段１３０６２が取得した各発音区間の１以上の確率値の時間平均値を、発音区間毎に算出し、１以上の時間平均値を得て、当該１以上の時間平均値をパラメータとして音声の評定値を算出する。
【００７２】
発音区間確率値取得手段１３０６２、および評定値算出手段１３０６３は、通常、ＭＰＵやメモリ等から実現され得る。発音区間確率値取得手段１３０６２等の処理手順は、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。
次に、本発音評定装置の動作について図１４のフローチャートを用いて説明する。図１４のフローチャートにおいて、図１０と異なるステップについてのみ説明する。
（ステップＳ１４０１）発音区間確率値取得手段１３０６２は、ｊに１を代入する。
【００７３】
（ステップＳ１４０２）発音区間確率値取得手段１３０６２は、次の評定対象の発音区間である、ｊ番目の発音区間が存在するか否かを判断する。ｊ番目の発音区間が存在すればステップＳ１４０３に行き、ｊ番目の発音区間が存在しなければステップＳ２０２に行く。
（ステップＳ１４０３）発音区間確率値取得手段１３０６２は、ｊ番目の発音区間に対応する１以上の最適状態の確率値をすべて取得する。
【００７４】
（ステップＳ１４０４）評定値算出手段１３０６３は、ステップＳ１４０３で取得した１以上の発音区間毎の１以上の確率値をパラメータとして音声の評定値を算出する。例えば、ステップＳ１４０３で取得した１以上の確率値の平均値（時間平均値）を算出する。
（ステップＳ１４０５）出力部１０７は、ステップＳ１４０４で算出した確率値の平均値（評定値）を出力する。
（ステップＳ１４０６）発音区間確率値取得手段１３０６２は、カウンタｊを１、インクリメントする。ステップＳ１４０２に戻る。
以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、評定値の算出アルゴリズムが実施の形態２とは異なるので、その動作を中心に説明する。
【００７５】
まず、学習者が、語学学習の開始の指示である動作開始指示を入力した後、学習対象の音声を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレ

ームに区分する。
そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「Ｏ＝ｏ$_1$，ｏ$_2$，・・・，ｏ$_T$」を算出する。
【００７６】
次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルｏ$_t$に基づいて、各フレームの最適状態（特徴ベクトルｏ$_t$に対する最適状態）を決定する。つまり、最適状態決定手段１０６１は、最適状態系列を得る。
次に、発音区間確率値取得手段１３０６２は、各フレームの各状態の事後確率（確率値）を算出する。なお、確率値の算出は、上述した数式１、数式２により算出できる。
【００７７】
そして、発音区間確率値取得手段１３０６２は、発音区間に対応する１以上の最適状態の確率値をすべて取得する。そして、評定値算出手段１３０６３は、取得した１以上の確率値の平均値（時間平均値）を算出する。具体的には、評定値算出手段１３０６３は、数式４により評定値を算出する。
【数４】

$$\text{t-DAP}(m) \triangleq \frac{\sum_{\tau \in T(q_t^*)} \gamma_\tau(q_\tau^*)}{|T(q_t^*)|}$$

【００７８】
かかる評定値算出手段１３０６３が算出した評定値（「ｔ－ＤＡＰスコア」とも言う。）を、図１５の表に示す。図１５において、アメリカ人男性と日本人男性の評定結果を示す。ＰｈｏｎｅｍｅおよびＷｏｒｄは，ｔ－ＤＡＰにおける時間平均の範囲を示す。図１５において、アメリカ人男性の発音の評定値が日本人男性の発音の評定値より高く、良好な評定結果が得られている。
そして、出力部１０７は、算出した発音区間ごと（例えば、音素毎）の評定値を、順次出力する。かかる出力例は、図１６である。
以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。
【００７９】
また、本実施の形態によれば、連結されたＨＭＭである連結ＨＭＭを用いて最適状態を求め、評定値を算出するので、高速に評定値を求めることができる。したがって、上記の具体例で述べたように、リアルタイムに、フレームごと、音素ごと、単語ごとの評定値を出力できる。また、本実施の形態によれば、動的計画法に基づいた事後確率を確率値として算出するので、さらに高速に評定値を求めることができる。また、本実施の形態によれば、フレームごとに確率値を算出するので、上述したように、フレーム単位だけではなく、音素・単語単位、または／および発声全体の評定結果を出力でき、出力態様の自由度が高い。
また、本実施の形態によれば、評定値を、発音区間の単位で算出でき、実施の形態１におけるような状態単位のＤＡＰと比較して、本来、測定したい類似度を精度良く、安定して求めることができる。
【００８０】
さらに、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、前記区分されたフレーム毎の音声データであるフレーム音声データを1以上得るフレーム音声データ取得ステップと、格納されているデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータである教師データと前記1以上のフ

レーム音声データに基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【００８１】

また、上記プログラムにおいて、前記教師データは、音韻毎の隠れマルコフモデル（ＨＭＭ）を連結したＨＭＭに基づくデータであり、前記評定ステップは、前記1以上のフレーム音声データの最適状態を決定する最適状態決定サブステップと、前記最適状態決定サブステップで決定した最適状態の確率値を、発音区間毎に取得する発音区間確率値取得サブステップと、前記発音区間確率値取得サブステップで取得した１以上の発音区間毎の１以上の確率値をパラメータとして音声の評定値を算出する評定値算出サブステップを具備するプログラム、である。

【００８２】

さらに、評定値算出サブステップは、前記発音区間確率値取得サブステップで取得した各発音区間の１以上の確率値の時間平均値を、発音区間毎に算出し、１以上の時間平均値を得て、当該１以上の時間平均値をパラメータとして音声の評定値を算出することが好適である。

（実施の形態４）

【００８３】

本実施の形態における発音評定装置は、実施の形態１、２、３の発音評定装置と比較して、評定部における評定アルゴリズムが異なる。本実施の形態において、評定値は、最適状態を含む音韻の中の全状態の確率値を発音区間で評価して、算出される。本実施の形態における発音評定装置が算出する事後確率を、実施の形態１におけるＤＡＰに対してｔ-ｐ－ＤＡＰと呼ぶ。

【００８４】

図１７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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、評定部１７０６、出力部１０７を具備する。評定部１７０６は、最適状態決定手段１０６１、発音区間フレーム音韻確率値取得手段１７０６２、評定値算出手段１７０６３を具備する。

発音区間フレーム音韻確率値取得手段１７０６２は、最適状態決定手段１０６１が決定した各フレームの最適状態を有する音韻全体の状態における１以上の確率値を、発音区間毎に取得する。

【００８５】

評定値算出手段１７０６３は、発音区間フレーム音韻確率値取得手段１７０６２が取得した１以上の発音区間毎の１以上の確率値をパラメータとして音声の評定値を算出する。評定値算出手段１７０６３は、例えば、最適状態決定手段１０６１が決定した各フレームの最適状態を有する音韻全体の状態における１以上の確率値の総和を、フレーム毎に得て、当該フレーム毎の確率値の総和に基づいて、発音区間毎の確率値の総和の時間平均値を１以上得て、当該１以上の時間平均値をパラメータとして音声の評定値を算出する。

【００８６】

発音区間フレーム音韻確率値取得手段１７０６２、および評定値算出手段１７０６３は、通常、ＭＰＵやメモリ等から実現され得る。発音区間フレーム音韻確率値取得手段１７０６２等の処理手順は、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。

次に、本発音評定装置の動作について図１８のフローチャートを用いて説明する。図１８のフローチャートにおいて、図１４と異なるステップについてのみ説明する。

（ステップＳ１８０１）発音区間フレーム音韻確率値取得手段１７０６２は、カウンタｋに１を代入する。

【００８７】

（ステップＳ１８０２）発音区間フレーム音韻確率値取得手段１７０６２は、ｋ番目の

フレームが、ｊ番目の発音区間に存在するか否かを判断する。ｋ番目のフレームが存在すればステップＳ１８０３に行き、ｋ番目のフレームが存在しなければステップＳ１８０１０６に飛ぶ。

（ステップＳ１８０３）発音区間フレーム音韻確率値取得手段１７０６２は、ｋ番目のフレームの最適状態を含む音韻の全ての確率値を取得する。

（ステップＳ１８０４）評定値算出手段１７０６３は、ステップＳ１８０３で取得した１以上の確率値をパラメータとして、１フレームの音声の評定値を算出する。

（ステップＳ１８０５）発音区間フレーム音韻確率値取得手段１７０６２は、ｋを１、インクメントする。ステップＳ１８０２に戻る。

【００８８】

（ステップＳ１８０６）評定値算出手段１７０６３は、ｊ番目の発音区間の評定値を算出する。評定値算出手段１７０６３は、例えば、最適状態決定手段１０６１が決定した各フレームの最適状態を有する音韻全体の状態における１以上の確率値の総和を、フレーム毎に得て、当該フレーム毎の確率値の総和に基づいて、発音区間の確率値の総和の時間平均値を、当該発音区間の音声の評定値として算出する。

（ステップＳ１８０７）出力部１０７は、ステップＳ１８０６で算出した評定値を出力する。

（ステップＳ１８０８）発音区間フレーム音韻確率値取得手段１７０６２は、ｊを１、インクメントする。ステップＳ１４０２に戻る。

以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、評定値の算出アルゴリズムが実施の形態３とは異なるので、その動作を中心に説明する。

【００８９】

まず、学習者が、語学学習の開始の指示である動作開始指示を入力した後、学習対象の音声を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。

そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「$O = o_1, o_2, \cdots, o_T$」を算出する。

次に、発音区間フレーム音韻確率値取得手段１７０６２は、各フレームの各状態の事後確率（確率値）を算出する。確率値の算出は、上述した数式１、数式２により算出できる。

【００９０】

次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルo_tに基づいて、各フレームの最適状態（特徴ベクトルo_tに対する最適状態）を決定する。つまり、最適状態決定手段１０６１は、最適状態系列を得る。なお、各フレームの各状態の事後確率（確率値）を算出と、最適状態の決定の順序は問わない。

【００９１】

次に、発音区間フレーム音韻確率値取得手段１７０６２は、発音区間ごとに、当該発音区間に含まれる各フレームの最適状態を含む音韻の全ての確率値を取得する。そして、評定値算出手段１７０６３は、各フレームの最適状態を含む音韻の全ての確率値の総和を、フレーム毎に算出する。そして、評定値算出手段１７０６３は、フレーム毎に算出された確率値の総和を、発音区間毎に時間平均し、発音区間毎の評定値を算出する。具体的には、評定値算出手段１７０６３は、数式５により評定値を算出する。

【数５】

$$\text{t-p-DAP}(m) \triangleq \frac{\sum_{\tau \in T(q_t^*)} \text{p-DAP}(\tau)}{|T(q_t^*)|}$$

【００９２】

かかる評定値算出手段１７０６３が算出した評定値（「ｔ－ｐ－ＤＡＰスコア」とも言う。）を、図１９の表に示す。図１９において、アメリカ人男性と日本人男性の評定結果を示す。ＰｈｏｎｅｍｅおよびＷｏｒｄは，ｔ－ｐ－ＤＡＰにおける時間平均の範囲を示す。ここでは、ＤＡＰの代わりにｐ－ＤＡＰの時間平均を採用したものである。図１９において、アメリカ人男性の発音の評定値が日本人男性の発音の評定値より高く、良好な評定結果が得られている。

そして、出力部１０７は、算出した発音区間ごと（ここでは、音素毎）の評定値を、順次出力する。かかる出力例は、図１６である。

以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。

【００９３】

また、本実施の形態によれば、連結されたＨＭＭである連結ＨＭＭを用いて最適状態を求め、評定値を算出するので、高速に評定値を求めることができる。したがって、上記の具体例で述べたように、リアルタイムに、発音区間ごとの評定値を出力できる。また、本実施の形態によれば、動的計画法に基づいた事後確率を確率値として算出するので、さらに高速に評定値を求めることができる。

【００９４】

また、本実施の形態によれば、評定値を、発音区間の単位で算出でき、実施の形態１におけるような状態単位のＤＡＰと比較して、本来、測定したい類似度（発音区間の類似度）を精度良く、安定して求めることができる。

【００９５】

さらに、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、前記区分されたフレーム毎の音声データであるフレーム音声データを1以上得るフレーム音声データ取得ステップと、格納されているデータであり、音韻毎の隠れマルコフモデル（ＨＭＭ）に基づくデータである教師データと前記1以上のフレーム音声データに基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【００９６】

また、上記プログラムにおいて、前記教師データは、音韻毎の隠れマルコフモデル（ＨＭＭ）を連結したＨＭＭに基づくデータであり、前記評定ステップは、前記1以上のフレーム音声データの最適状態を決定する最適状態決定サブステップと、前記最適状態決定サブステップで決定した各フレームの最適状態を有する音韻全体の状態における１以上の確率値を、発音区間毎に取得する発音区間フレーム音韻確率値取得サブステップと、前記発音区間フレーム音韻確率値取得サブステップで取得した１以上の発音区間毎の１以上の確率値をパラメータとして音声の評定値を算出する評定値算出サブステップを具備するプログラム、である。

【００９７】

以上の４つの実施の形態で算出した評定値は、図２０において、それぞれ（１）～（４）の類似度である。つまり、図２０において、縦軸は、音韻毎の隠れマルコフモデル（ＨＭＭ）を、比較対象の音素の順に連結したＨＭＭである。図２０の横軸は、入力音声の特

徴ベクトル系列を示す。また、図２０の実線は、最適状態系列である。そして、黒丸の（１）はＤＡＰ、斜線部の（２）はｐ－ＤＡＰ、網掛け部の（３）はｔ－ＤＡＰを示す。ｔｐ－ＤＡＰは、発音区間において、最適状態を含む全音韻の状態の確率値に基づいて算出される。

（実施の形態５）

【００９８】

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を精度高く評定できる発音評定装置について説明する。特に、本発音評定装置は、無音区間を検知し、無音区間を考慮した類似度評定が可能な発音評定装置である。

【００９９】

また、本実施の形態における発音評定装置は、例えば、語学学習や物真似練習などに利用できる。図２１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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、特殊音声検知部２１０１、評定部２１０２、出力部１０８を具備する。評定部２１０２は、無音データ格納手段２１０２１、無音区間検出手段２１０２２、最適状態決定手段１０６１、最適状態確率値取得手段１０６２、評定値算出手段２１０２３を具備する。

【０１００】

特殊音声検知部２１０１は、フレーム毎の入力音声データに基づいて、特殊な音声が入力されたことを検知する。なお、ここで特殊な音声は、無音も含む。また、特殊音声検知部２１０１は、例えば、フレームの最適状態の確率値を、ある音素区間において取得し、ある音素区間の１以上の確率値の総和が所定の値より低い場合（想定されている音素ではない、と判断できる場合）、当該音素区間において特殊な音声が入力されたと、検知する。かかる検知の具体的なアルゴリズムの例は後述する。特殊音声検知部２１０１は、通常、ＭＰＵやメモリ等から実現され得る。特殊音声検知部２１０１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。

【０１０１】

評定部２１０２は、教師データ格納部１０２の教師データと入力音声データと特殊音声検知部２１０６における検知結果に基づいて、音声受付部１０３が受け付けた音声の評定を行う。評定方法の具体例は、後述する。評定部２１０２は、通常、ＭＰＵやメモリ等から実現され得る。評定部２１０２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。

【０１０２】

無音データ格納手段２１０２１は、無音を示すデータであり、ＨＭＭに基づくデータである無音データを格納している。無音データ格納手段２１０２１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。

【０１０３】

無音区間検出手段２１０２２は、フレーム音声データ取得部１０５が取得したフレーム音声データ、および無音データ格納手段２１０２１の無音データに基づいて、無音の区間を検出する。無音区間検出手段２１０２２は、フレーム音声データ取得部１０５が取得したフレーム音声データと無音データの類似度が所定の値以上である場合に、当該フレーム音声データは無音区間のデータであると判断しても良い。また、無音区間検出手段２１０２２は、下記で述べる最適状態確率値取得手段１０６２が取得した確率値が所定の値以下であり、かつ、フレーム音声データ取得部１０５が取得したフレーム音声データと無音データの類似度が所定の値以上である場合に、当該フレーム音声データは無音区間のデータであると判断しても良い。

【０１０４】

評定値算出手段２１０２３は、無音区間検出手段２１０２２が検出した無音区間を除い

て、かつ最適状態確率値取得手段１０６２が取得した確率値をパラメータとして音声の評定値を算出する。なお、評定値算出手段２１０２３は、上記確率値を如何に利用して、評定値を算出するかは問わない。評定値算出手段２１０２３は、例えば、最適状態確率値取得手段１０６２が取得した確率値と、当該確率値に対応するフレームの全状態における確率値の総和をパラメータとして音声の評定値を算出する。評定値算出手段２１０２３は、ここでは、通常、無音区間検出手段２１０２２が検出した無音区間を除いて、フレームごとに評定値を算出する。なお、評定値算出手段２１０２３は、かならずしも無音区間を除いて、評定値を算出する必要はない。評定値算出手段２１０２３は、無音区間の影響を少なくするように評定値を算出しても良い。

【０１０５】

無音区間検出手段２１０２２、評定値算出手段２１０２３は、通常、ＭＰＵやメモリ等から実現され得る。無音区間検出手段２１０２２等の処理手順は、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。

次に、本発音評定装置の動作について図２２のフローチャートを用いて説明する。

【０１０６】

（ステップＳ２２０１）評定値算出手段２１０２３は、ステップＳ２０７で取得したｉ番目のフレーム音声データに対応する評定値（例えば、ｐ－ＤＡＰスコア）を算出する。なお、ｐ－ＤＡＰスコアの算出方法は、実施の形態２で述べたので、ここでの説明は省略する。

【０１０７】

（ステップＳ２２０２）特殊音声検知部２１０１は、ステップＳ２２０１で算出した値が、所定の値より低いか否かを判断する。所定の値より低ければステップＳ２２０３に行き、所定の値より低くなければステップＳ２２０６に飛ぶ。

（ステップＳ２２０３）無音区間検出手段２１０２２は、無音データと全教師データの確率値を取得する。

【０１０８】

（ステップＳ２２０４）無音区間検出手段２１０２２は、ステップＳ２２０３で取得した確率値の中で、無音データの確率値が最も高いか否かを判断する。無音データの確率値が最も高ければ（かかる場合、無音の区間であると判断する）ステップＳ２２０５に行き、無音データの確率値が最も高くなければステップＳ２２０６に行く。

（ステップＳ２２０５）無音区間検出手段２１０２２は、カウンタｉを１、インクリメントする。ステップＳ２０６に戻る。

（ステップＳ２２０６）出力部１０８は、ステップＳ２２０１で算出した評定値を出力する。

【０１０９】

なお、図２２のフローチャートにおいて、出力部１０８は、無音区間と判定した区間の評定値は出力しなかった（無音区間を無視した）が、特殊音声が検知された区間が無音区間である旨を明示したり、無音区間が存在する旨を明示したりする態様で出力しても良い。また、評定値算出手段２１０２３は、発音区間や、それ以上の単位のスコアを算出する場合に、無音区間の評定値を無視して、スコアを算出することが好適であるが、無音区間の評定値の影響を、例えば、１／１０にして、発音区間や発音全体のスコアを算出するなどしても良い。評定部２１０２は、教師データと入力音声データと特殊音声検知部２１０１における検知結果に基づいて、音声受付部１０３が受け付けた音声の評定を行えばよい。

【０１１０】

また、図２２のフローチャートにおいて、特殊音声検知部２１０１は、ｉ番目のフレーム音声データのｐ－ＤＡＰスコアに基づいて特殊音声を検知したが、例えば、ＤＡＰスコアに基づいて特殊音声を検知しても良い。

【０１１１】

以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、無音区間を考慮して評定値を算出するので、評定値の算出アルゴリズムが実施の形態１等とは異なる。そこで、その異なる処理を中心に説明する。

まず、学習者が、語学学習の開始の指示である動作開始指示を入力する。

次に、学習者は、例えば、学習対象の音声を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。

次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。

そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「Ｏ＝ｏ$_1$、ｏ$_2$、・・・、ｏ$_T$」を算出する。

次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルｏ$_t$に基づいて、所定のフレームの最適状態（特徴ベクトルｏ$_t$に対する最適状態）を決定する。

次に、最適状態確率値取得手段１０６２は、上述した数式１、２により、最適状態における確率値を算出する。

【０１１２】

次に、評定値算出手段２１０２３は、例えば、最適状態決定手段１０６１が決定した最適状態を有する音韻全体の状態における１以上の確率値を取得し、当該１以上の確率値の総和をパラメータとして音声の評定値を算出する。つまり、評定値算出手段２１０２３は、例えば、ｐ－ＤＡＰスコアをフレーム毎に算出する。

【０１１３】

そして、特殊音声検知部２１０１は、算出されたフレームに対応する評定値（ｐ－ＤＡＰスコア）を用いて、特殊な音声が入力されたか否かを判断する。具体的には、特殊音声検知部２１０１は、例えば、評価対象のフレームに対して算出された評定値が、所定の数値より低ければ、特殊な音声が入力された、と判断する。なお、特殊音声検知部２１０１は、一のフレームに対応する評定値が小さいからといって、直ちに特殊な音声が入力された、と判断する必要はない。つまり、特殊音声検知部２１０１は、フレームに対応する評定値が小さいフレームが所定の数以上、連続する場合に、当該連続するフレーム群に対応する区間が特殊な音声が入力された区間と判断しても良い。

【０１１４】

特殊音声検知部２１０１が、特殊音声を検知する場合について説明する図を図２３に示す。図２３（ａ）の縦軸は、ｐ－ＤＡＰスコアであり、横軸はフレームを示す。図２３（ａ）において、（Ｖ）は、Ｖｉｔｅｒｂｉアライメントを示す。図２３（ａ）において、網掛けのフレーム群のおけるｐ－ＤＡＰスコアは、所定の値より低く、特殊音声の区間である、と判断される。

【０１１５】

次に、特殊な音声が入力された、と判断した場合、無音区間検出手段２１０２２は、無音データ格納手段２１０２１から無音データを取得し、当該フレーム群の各フレームのＨＭＭと無音データとの類似度を算定し、類似度が所定値以上であれば当該フレーム群に対応する音声データが、無音データであると判断する。図２３（ｂ）は、無音データとの比較の結果、当該無音データとの類似度を示す事後確率の値（「ＡＰスコア」とも言う。）が高いことを示す。その結果、無音区間検出手段２１０２２は、当該特殊音声の区間は、無音区間である、と判断する。なお、図２３（ａ）において、網掛けのフレーム群のおけるｐ－ＤＡＰスコアは、所定の値より低く、特殊音声の区間である、と判断され、かつ、無音データとの比較の結果、ＡＰスコアが低い場合には、無音区間ではない、と判断される。そして、かかる区間において、例えば、単に、発音が上手くなく、低い評定値が出力される。なお、図２３（ａ）に示しているように、通常、無音区間は、第一のワード（「ｗｏｒｄ１」）の最終音素の後半部、および第一のワードに続く第二のワード（「ｗｏｒｄ２」）の第一音素の前半部のスコアが低い。

そして、出力部１０８は、出力する評定値から、無音データの区間の評定値を考慮しないように、無視する。

そして、出力部１０８は、各フレームに対応する評定値を出力する。この場合、例えば、無音データの区間の評定値は、出力されない。

かかる評定値の出力態様例は、例えば、図７、図８である。

なお、出力部１０８が行う出力は、無音区間の存在を示すだけの出力でも良い。

【０１１６】

以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。その場合、無音区間を考慮して類似度を評定するので、極めて正確な評定結果が得られる。

【０１１７】

なお、無音区間のデータは、無視して評定結果を算出することは好適である。ただし、本実施の形態において、例えば、無音区間の評価の影響を他の区間と比較して少なくするなど、無視する以外の方法で、無音区間のデータを考慮して、評定値を出力しても良い。

【０１１８】

また、本実施の形態の具体例によれば、ｐ－ＤＡＰスコアを用いて、評定値を算出したが、無音の区間を考慮して評定値を算出すれば良く、上述した他のアルゴリズム（ＤＡＰ、ｔ－ＤＡＰ、ｔ－ｐ－ＤＡＰ）、または、本明細書では述べていない他のアルゴリズムにより評定値を算出しても良い。つまり、本実施の形態によれば、教師データと入力音声データと特殊音声検知部における検知結果に基づいて、音声受付部が受け付けた音声の評定を行い、特に、無音データを考慮して、評定値を算出すれば良い。

また、本実施の形態によれば、まず、ＤＡＰスコアが低い区間を検出してから、無音区間の検出をした。しかし、ＤＡＰスコアが低い区間を検出せずに、無音データとの比較により、無音区間を検出しても良い。

【０１１９】

さらに、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、前記区分されたフレーム毎の入力音声データを得る入力音声データ取得ステップと、前記フレーム毎の入力音声データに基づいて、特殊な音声が入力されたことを検知する特殊音声検知ステップと、前記教師データと前記入力音声データと前記特殊音声検知ステップにおける検知結果に基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【０１２０】

また、上記プログラムは、前記特殊音声検知ステップにおいて、無音を示すＨＭＭに基づくデータである無音データを格納している無音データ格納サブステップと、前記入力音声データおよび前記無音データに基づいて、無音の区間を検出する無音区間検出サブステップを具備することは好適である。

また、上記プログラムにおいて、前記評定ステップは、前記無音の区間を除いた区間の前記入力音声データと、前記教師データに基づいて、前記音声受付部が受け付けた音声の評定を行うことは好適である。

（実施の形態６）

【０１２１】

本実施の形態において、入力音声において、特殊音声を検知し、比較対象の音声と入力音声の類似度を精度高く評定できる発音評定装置について説明する。特に、本発音評定装置は、音韻の挿入を検知できる発音評定装置である。

【０１２２】

また、本実施の形態における発音評定装置は、例えば、語学学習や物真似練習などに利用できる。図２４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。本発音評

定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、特殊音声検知部２４０１、評定部２４０２、出力部２４０３を具備する。

【０１２３】

特殊音声検知部２４０１は、一の音素の後半部および当該音素の次の音素の前半部の評定値が所定の条件を満たすことを検知する。後半部、および前半部の長さは問わない。特殊音声検知部２４０１は、通常、ＭＰＵやメモリ等から実現され得る。特殊音声検知部２４０１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。

【０１２４】

評定部２４０２は、特殊音声検知部２４０１が所定の条件を満たすことを検知した場合に、少なくとも音素の挿入があった旨を示す評定結果を構成する。なお、評定部２４０２は、実施の形態５で述べたアルゴリズムにより、特殊音声検知部２４０１が所定の条件を満たすことを検知した区間に無音が挿入されたか否かを判断し、無音が挿入されていない場合に、他の音素が挿入されたと検知しても良い。また、評定部２４０２は、無音が挿入されていない場合に、他の音韻ＨＭＭに対する確率値を算出し、所定の値より高い確率値を得た音韻が挿入された、との評定結果を得ても良い。なお、実施の形態５で述べた無音区間の検知は、無音音素の挿入の検知である、とも言える。評定部２４０２は、通常、ＭＰＵやメモリ等から実現され得る。評定部２４０２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。

【０１２５】

出力部２４０３は、評定部２４０２における評定結果を出力する。ここでの評定結果は、音素の挿入があった旨を示す評定結果を含む。評定結果は、音素の挿入があった旨、および評定値（スコア）の両方であっても良い。なお、教師データにおいて想定されていない音素の挿入を検知した場合、通常、評定値は低くなる。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２４０３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ても含まないと考えても良い。出力部２４０３は、出力デバイスのドライバーソフトまたは、出力デバイスのドライバーソフトと出力デバイス等で実現され得る。

次に、本発音評定装置の動作について図２５のフローチャートを用いて説明する。

【０１２６】

（ステップＳ２５０１）特殊音声検知部２４０１は、フレームに対応するデータを一時的に蓄積するバッファにデータが格納されているか否かを判断する。なお、格納されているデータは、ステップＳ２２０２で、所定の値より低い評定値と評価されたフレーム音声データ、または当該フレーム音声データから取得できるデータである。データが格納されていればステップＳ２５０７に行き、データが格納されていなければステップＳ２０２に戻る。

【０１２７】

（ステップＳ２５０２）特殊音声検知部２４０１は、バッファにデータが格納されているか否かを判断する。データが格納されていればステップＳ２５０７に行き、データが格納されていなければステップステップＳ２５０３に行く。

（ステップＳ２５０３）出力部２４０３は、ステップＳ２２０１で算出した評定値を出力する。

（ステップＳ２５０４）特殊音声検知部２４０１は、カウンタｉを１、インクリメントする。ステップＳ２０６に戻る。

（ステップＳ２５０５）特殊音声検知部２４０１は、バッファに、所定の値より低い評定値と評価されたフレーム音声データ、または当該フレーム音声データから取得できるデータを一時蓄積する。

（ステップＳ２５０６）特殊音声検知部２４０１は、カウンタｉを１、インクリメント

する。ステップＳ２０６に戻る。
（ステップＳ２５０７）特殊音声検知部２４０１は、カウンタｊに１を代入する。
【０１２８】
（ステップＳ２５０８）特殊音声検知部２４０１は、ｊ番目のデータが、バッファに存在するか否かを判断する。ｊ番目のデータが存在すればステップＳ２５０９に行き、ｊ番目のデータが存在しなければステップＳ２５１５に飛ぶ。
（ステップＳ２５０９）特殊音声検知部２４０１は、ｊ番目のデータに対応する最適状態の音素を取得する。
（ステップＳ２５１０）特殊音声検知部２４０１は、ｊ番目のデータに対する全教師データの確率値を算出し、最大の確率値を持つ音素を取得する。
【０１２９】
（ステップＳ２５１１）特殊音声検知部２４０１は、ステップＳ２５０９で取得した音素とステップＳ２５１０で取得した音素が異なる音素であるか否かを判断する。異なる音素であればステップＳ２５１２に行き、異なる音素でなければステップＳ２５１４に飛ぶ。
（ステップＳ２５１２）評定部２４０２は、音素の挿入があった旨を示す評定結果を構成する。
（ステップＳ２５１３）特殊音声検知部２４０１は、カウンタｊを１、インクリメントする。ステップＳ２５０８に戻る。
（ステップＳ２５１４）出力部２４０３は、バッファ中の全データに対応する全評定値を出力する。ここで、全評定値とは、例えば、フレーム毎のｐ－ＤＡＰスコアである。ステップＳ２５１３に行く。
【０１３０】
（ステップＳ２５１５）出力部２４０３は、評定結果に「挿入の旨」の情報が入っているか否かを判断する。「挿入の旨」の情報が入っていればステップＳ２５１６に行き、「挿入の旨」の情報が入っていなければステップＳ２５１７に行く。
（ステップＳ２５１６）出力部２４０３は、評定結果を出力する。
（ステップＳ２５１７）出力部２４０３は、バッファをクリアする。ステップＳ２０６に戻る。
【０１３１】
なお、図２５のフローチャートにおいて、評定値の低いフレームが２つの音素に渡って存在すれば、直ちに音素の挿入があったと判断した。つまり、一の音素の後半部（少なくとも最終フレーム）および当該音素の次の音素の第一フレームの評定値が所定値より低い場合に、音素の挿入があったと判断した。しかし、図２５のフローチャートにおいて、一の音素の所定区間以上の後半部、および当該音素の次の音素の所定区間以上の前半部の評定値が所定値よりすべて低い場合に、音素の挿入があったと判断するようにしても良い。
以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、音素の挿入の検知を行う処理が実施の形態５等とは異なる。そこで、その異なる処理を中心に説明する。
まず、学習者が、語学学習の開始の指示である動作開始指示を入力する。
次に、学習者は、例えば、学習対象の音声を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。
次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。
そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「$O = o_1, o_2, \cdots, o_T$」を算出する。
次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルo_tに基づいて、所定のフレームの最適状態（特徴ベクトルo_tに対する最適状態）を決定する。

次に、最適状態確率値取得手段１０６２は、数式１、２により、最適状態における確率値を算出する。

【０１３２】

次に、評定値算出手段２１０２３は、例えば、最適状態決定手段１０６１が決定した最適状態を有する音韻全体の状態における１以上の確率値を取得し、当該１以上の確率値の総和をパラメータとして音声の評定値を算出する。つまり、評定値算出手段２１０２３は、例えば、ｐ－ＤＡＰスコアをフレーム毎に算出する。ここで、算出するスコアは、上述したＤＡＰスコア等でも良い。

【０１３３】

そして、特殊音声検知部２１０１は、算出されたフレームに対応する評定値を用いて、特殊な音声が入力されたか否かを判断する。つまり、評定値(例えば、ｐ－ＤＡＰスコア)が、所定の値より低い区間が存在するか否かを判断する。

【０１３４】

次に、特殊音声検知部２１０１は、図２６に示すように、評定値(例えば、ｐ－ＤＡＰスコア)が、所定の値より低い区間が、２つの音素に跨っているか否かを判断し、２つの音素に跨がっていれば、当該区間に音素が挿入された、と判断する。なお、かかる場合の詳細なアルゴリズムの例は、図２５で説明した。また、図２６において、斜線部が、予期しない音素が挿入された区間である。

【０１３５】

次に、評定部２４０２は、音素の挿入があった旨を示す評定結果（例えば、「予期しない音素が挿入されました。」）を構成する。そして、出力部２４０３は、構成した評定結果を出力する。図２７は、評定結果の出力例である。なお、出力部２４０３は、通常の入力音声に対しては、上述したように評定値を出力することが好適である。

【０１３６】

以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。その場合、特殊音声、特に、予期せぬ音素の挿入を検知できるので、極めて精度の高い評定結果が得られる。

【０１３７】

なお、本実施の形態において、音素の挿入を検知できれば良く、評定値の算出アルゴリズムは問わない。評定値の算出アルゴリズムは、上述したアルゴリズム（ＤＡＰ、ｐ－ＤＡＰ、ｔ－ＤＡＰ、ｔ－ｐ－ＤＡＰ）でも良く、または、本明細書では述べていない他のアルゴリズムでも良い。

【０１３８】

さらに、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、前記区分されたフレーム毎の入力音声データを得る入力音声データ取得ステップと、前記フレーム毎の入力音声データに基づいて、特殊な音声が入力されたことを検知する特殊音声検知ステップと、前記教師データと前記入力音声データと前記特殊音声検知ステップにおける検知結果に基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【０１３９】

また、上記プログラムは、前記特殊音声検知ステップにおいて、一の音素の後半部および当該音素の次の音素の前半部の評定値が所定の条件を満たすことを検知し、前記評定ステップにおいては、前記特殊音声検知ステップにおいて前記所定の条件を満たすことを検知した場合に、少なくとも音素の挿入があった旨を示す評定結果を構成するプログラム、である。

（実施の形態７）

【０１４０】

本実施の形態において、入力音声において、特殊音声を検知し、比較対象の音声と入力音声の類似度を精度高く評定できる発音評定装置について説明する。特に、本発音評定装置は、音韻の置換を検知できる発音評定装置である。

【０１４１】

また、本実施の形態における発音評定装置は、例えば、語学学習や物真似練習などに利用できる。図２８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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、特殊音声検知部２８０１、評定部２８０２、出力部２４０３を具備する。

【０１４２】

特殊音声検知部２８０１は、一の音素の評定値が所定の値より低いことを検知する。また、特殊音声検知部２８０１は、一の音素の評定値が所定の値より低く、かつ当該音素の直前の音素および当該音素の直後の音素の評定値が所定の値より高いことをも検知しても良い。また、特殊音声検知部２８０１は、一の音素の評定値が所定の値より低く、かつ、想定していない音素のＨＭＭに基づいて算出された評定値が所定の値より高いことを検知しても良い。つまり、特殊音声検知部２８０１は、所定のアルゴリズムで、音韻の置換を検知できれば良い。そのアルゴリズムは種々考えられる。特殊音声検知部２８０１は、通常、ＭＰＵやメモリ等から実現され得る。特殊音声検知部２８０１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。

【０１４３】

評定部２８０２は、特殊音声検知部２８０１が所定の条件を満たすことを検知した場合に、少なくとも音素の置換があった旨を示す評定結果を構成する。評定部２８０２は、通常、ＭＰＵやメモリ等から実現され得る。評定部２８０２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。

次に、本発音評定装置の動作について図２９のフローチャートを用いて説明する。

【０１４４】

（ステップＳ２９０１）特殊音声検知部２８０１は、バッファに蓄積されているデータに対応するフレーム音声データ群が一の音素に対応するか否かを判断する。一の音素であればステップＳ２９０２に行き、一の音素でなければステップＳ２９１０に行く。

【０１４５】

（ステップＳ２９０２）特殊音声検知部２８０１は、バッファに蓄積されているデータに対応するフレーム音声データ群の音素の直前の音素の評定値を算出する。かかる評定値は、例えば、上述したｔ－ＤＡＰスコアである。なお、直前の音素とは、現在評定中の音素に対して直前の音素である。音素の区切りは、Ｖｉｔｅｒｂｉアルゴリズムにより算出できる。

【０１４６】

（ステップＳ２９０３）特殊音声検知部２８０１は、ステップＳ２９０２で算出した評定値が所定の値以上であるか否かを判断する。所定の値以上であればステップＳ２９０４に行き、所定の値より小さければステップＳ２９１０に行く。

（ステップＳ２９０４）特殊音声検知部２８０１は、直後の音素の評定値を算出する。かかる評定値は、例えば、上述したｔ－ＤＡＰスコアである。直後の音素とは、現在評定中の音素に対して直後の音素である。

【０１４７】

（ステップＳ２９０５）特殊音声検知部２８０１は、ステップＳ２９０４で算出した評定値が所定の値以上であるか否かを判断する。所定の値以上であればステップＳ２９０６に行き、所定の値より小さければステップＳ２９１０に行く。

【０１４８】

（ステップＳ２９０６）特殊音声検知部２８０１は、予め格納されている音韻ＨＭＭ（

予期する音韻のＨＭＭは除く）の中で、所定の値以上の評定値が得られる音韻ＨＭＭが一つ存在するか否かを判断する。所定の値以上の評定値が得られる音韻ＨＭＭが存在すればステップＳ２９０７に行き、所定の値以上の評定値が得られる音韻ＨＭＭが存在しなければステップＳ２９１０に行く。なお、予め格納されている音韻ＨＭＭは、通常、すべての音韻に対する多数の音韻ＨＭＭである。なお、本ステップにおいて、予め格納されている音韻ＨＭＭの確率値を算出し、最大の確率値を持つ音素を取得し、当該音素と最適状態の音素が異なるか否かを判断し、異なる場合に音素の置換があったと判断しても良い。

（ステップＳ２９０７）評定部２８０２は、音素の置換があった旨を示す評定結果を構成する。

（ステップＳ２９０８）出力部２４０３は、ステップＳ２９０７で構成した評定結果を出力する。

（ステップＳ２９０９）出力部２４０３は、バッファをクリアする。ステップＳ２０６に戻る。

（ステップＳ２９１０）出力部２４０３は、バッファ中の全データに対応する全評定値を出力する。

以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、音素の置換の検知を行う処理が実施の形態６等とは異なる。そこで、その異なる処理を中心に説明する。

まず、学習者が、語学学習の開始の指示である動作開始指示を入力する。

次に、学習者は、例えば、学習対象の音声を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。

次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。

そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「$O = o_1, o_2, \cdots, o_T$」を算出する。

次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルo_tに基づいて、所定のフレームの最適状態（特徴ベクトルo_tに対する最適状態）を決定する。

次に、最適状態確率値取得手段１０６２は、数式１、２により、最適状態における確率値を算出する。

【０１４９】

次に、評定値算出手段２１０２３は、例えば、最適状態決定手段１０６１が決定した最適状態を有する音韻全体の状態における１以上の確率値を取得し、当該１以上の確率値の総和をパラメータとして音声の評定値を算出する。つまり、評定値算出手段２１０２３は、例えば、ｐ－ＤＡＰスコアをフレーム毎に算出する。ここで、算出するスコアは、上述したＤＡＰスコア等でも良い。

【０１５０】

そして、特殊音声検知部２１０１は、算出されたフレームに対応する評定値を用いて、特殊な音声が入力されたか否かを判断する。つまり、評定値(例えば、ｐ－ＤＡＰスコア)が、所定の値より低い区間が存在するか否かを判断する。

【０１５１】

次に、特殊音声検知部２１０１は、図３０に示すように、評定値(例えば、ｔ－ＤＡＰスコア)が、所定の値より低い区間が、一つの音素内（ここでは音素２）であるか否かを判断する。そして、一つの音素内で評定値が低ければ、次に、特殊音声検知部２１０１は、直前の音素（音素１）または直後の音素（音素２）に対する評定値（例えば、ｔ－ＤＡＰスコア)を算出し、当該評定値が所定の値より高ければ、音素の置換が発生している可能性があると判断する。次に、特殊音声検知部２１０１は、予め格納されている音韻ＨＭＭ（予期する音韻のＨＭＭは除く）の中で、所定の値以上の評定値が得られる音韻ＨＭＭが一つ存在すれば、音素の置換が発生していると判断する。なお、図３０において、音素

２において、音素の置換が発生した区間である。なお、図３０において縦軸は評定値であり、当該評定値は、ＤＡＰ、ｐ－ＤＡＰ、ｔ－ＤＡＰ等、問わない。

【０１５２】

次に、評定部２４０２は、音素の置換があった旨を示す評定結果（例えば、「音素の置換が発生しました。」）を構成する。そして、出力部２４０３は、構成した評定結果を出力する。なお、出力部２４０３は、通常の入力音声に対しては、上述したように評定値を出力することが好適である。

【０１５３】

以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。その場合、特殊音声、特に、音素の置換を検知できるので、極めて精度の高い評定結果が得られる。

【０１５４】

なお、本実施の形態において、音素の置換を検知できれば良く、評定値の算出アルゴリズムは問わない。評定値の算出アルゴリズムは、上述したアルゴリズム（ＤＡＰ、ｐ－ＤＡＰ、ｔ－ＤＡＰ、ｔ－ｐ－ＤＡＰ）でも良く、または、本明細書では述べていない他のアルゴリズムでも良い。

【０１５５】

また、本実施の形態において、音素の置換の検知アルゴリズムは、他のアルゴリズムでも良い。例えば、音素の置換の検知において、所定以上の長さの区間を有することを置換区間の検知で必須としても良い。その他、置換の検知アルゴリズムの詳細は種々考えられる。

【０１５６】

さらに、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、前記区分されたフレーム毎の入力音声データを得る入力音声データ取得ステップと、前記フレーム毎の入力音声データに基づいて、特殊な音声が入力されたことを検知する特殊音声検知ステップと、前記教師データと前記入力音声データと前記特殊音声検知ステップにおける検知結果に基づいて、前記音声受付ステップで受け付けた音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【０１５７】

また、上記プログラムは、前記特殊音声検知ステップにおいて、一の音素の評定値が所定の条件を満たすことを検知し、前記評定ステップにおいて、前記特殊音声検知ステップで前記所定の条件を満たすことを検知した場合に、少なくとも音素の置換または欠落があった旨を示す評定結果を構成するプログラム、である。

【０１５８】

また、上記プログラムでは、前記特殊音声検知ステップにおいて、一の音素の評定値が所定の値より低く、かつ当該音素の直前の音素および当該音素の直後の音素の評定値が所定の値より高く、かつ予め格納されている音韻ＨＭＭの中で、所定の値以上の評定値が得られる音韻ＨＭＭが一つ存在することを検知し、前記評定ステップにおいて、前記特殊音声検知ステップで前記所定の条件を満たすことを検知した場合に、少なくとも音素の置換があった旨を示す評定結果を構成することは好適である。

（実施の形態８）

【０１５９】

本実施の形態において、入力音声において、特殊音声を検知し、比較対象の音声と入力音声の類似度を精度高く評定できる発音評定装置について説明する。特に、本発音評定装置は、音韻の欠落を検知できる発音評定装置である。

【０１６０】

また、本実施の形態における発音評定装置は、例えば、語学学習や物真似練習などに利

用できる。図３１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。本発音評定装置は、入力受付部１０１、教師データ格納部１０２、音声受付部１０３、フレーム区分部１０４、フレーム音声データ取得部１０５、特殊音声検知部３１０１、評定部３１０２、出力部２４０３を具備する。

特殊音声検知部３１０１は、一の音素の評定値が所定の値より低く、かつ当該音素の直前の音素または当該音素の直後の音素の評定値が所定の値より高いことを検知する。また、

特殊音声検知部３１０１は、一の音素の評定値が所定の値より低く、かつ当該音素の直前の音素または当該音素の直後の音素の評定値が所定の値より高く、かつ当該音素の区間長が所定の長さよりも短いことを検知しても良い。また、特殊音声検知部３１０１は、直前の音素に対応する確率値、または直後の音素に対応する確率値が、当該一の音素の確率値より高いことを検知しても良い。かかる場合に、特殊音声検知部３１０１は、音韻の欠落を検知することは好適である。さらに、音素の区間長が所定の長さよりも短いことを欠落の条件に含めることにより、音韻の欠落の検知の精度は向上する。特殊音声検知部３１０１は、通常、ＭＰＵやメモリ等から実現され得る。特殊音声検知部３１０１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。

【０１６１】

評定部３１０２は、特殊音声検知部３１０１が所定の条件を満たすことを検知した場合に、少なくとも音素の欠落があった旨を示す評定結果を構成する。評定部３１０２は、通常、ＭＰＵやメモリ等から実現され得る。評定部３１０２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。

次に、本発音評定装置の動作について図３２のフローチャートを用いて説明する。

【０１６２】

（ステップＳ３２０１）特殊音声検知部３１０１は、バッファに蓄積されているデータに対して、直前の音素に対応する教師データの確率値または、直後の音素に対応する教師データの確率値が、予定されている音素に対応する教師データの確率値より高いか否かを判断する。高ければステップＳ３２０２に行き、高くなければステップＳ２９１１に行く。なお、ステップＳ３２０２に行くための条件として、バッファに蓄積されているデータに対応するフレーム音声データ群の区間長が所定の長さ以下であることを付加しても良い。

（ステップＳ３２０２）評定部３１０２は、音素の欠落があった旨を示す評定結果を構成する。

なお、図３２のフローチャートにおいて、評定対象の音素（欠落したであろう音素）の区間長が、所定の長さ（例えば、３フレーム）よりも短いことを条件としたが、かかる条件は必須ではない。

以下、本実施の形態における発音評定装置の具体的な動作について説明する。本実施の形態において、音素の欠落の検知を行う処理が実施の形態７等とは異なる。そこで、その異なる処理を中心に説明する。

まず、学習者が、語学学習の開始の指示である動作開始指示を入力する。

次に、学習者は、例えば、学習対象の音声を発音する。そして、音声受付部１０３は、学習者が発音した音声の入力を受け付ける。

次に、フレーム区分部１０４は、音声受付部１０３が受け付けた音声を、短時間フレームに区分する。

そして、フレーム音声データ取得部１０５は、フレーム区分部１０４が区分した音声データを、スペクトル分析し、特徴ベクトル系列「Ｏ＝ｏ$_1$、ｏ$_2$、・・・、ｏ$_T$」を算出する。

次に、最適状態決定手段１０６１は、取得した特徴ベクトル系列を構成する各特徴ベクトルｏ$_t$に基づいて、所定のフレームの最適状態（特徴ベクトルｏ$_t$に対する最適状態）

を決定する。
次に、最適状態確率値取得手段１０６２は、数式１、２により、最適状態における確率値を算出する。
【０１６３】
次に、評定値算出手段２１０２３は、例えば、最適状態決定手段１０６１が決定した最適状態を有する音韻全体の状態における１以上の確率値を取得し、当該１以上の確率値の総和をパラメータとして音声の評定値を算出する。つまり、評定値算出手段２１０２３は、例えば、ｐ－ＤＡＰスコアをフレーム毎に算出する。ここで、算出するスコアは、上述したＤＡＰスコア等でも良い。
【０１６４】
そして、特殊音声検知部２１０１は、算出されたフレームに対応する評定値を用いて、特殊な音声が入力されたか否かを判断する。つまり、評定値(例えば、ｔ－ＤＡＰスコア)が、所定の値より低い区間が存在するか否かを判断する。
【０１６５】
次に、特殊音声検知部２１０１は、図３３に示すように、評定値(例えば、ｔ－ＤＡＰスコア)が、所定の値より低い区間が、一つの音素内（ここでは音素２）であるか否かを判断する。そして、一つの音素内で評定値が低ければ、特殊音声検知部２１０１は、直前の音素（音素１）または直後の音素（音素２）に対する評定値（例えば、ｔ－ＤＡＰスコア)を算出し、当該評定値が所定の値より高ければ、音素の欠落が発生している可能性があると判断する。そして、当該区間長が、例えば、３フレーム以下の長さであれば、かかる音素は欠落したと判断する。なお、図３３において、音素２の欠落が発生したことを示す。なお、図３３において縦軸は評定値であり、当該評定値は、ＤＡＰ、ｐ－ＤＡＰ、ｔ－ＤＡＰ等、問わない。また、上記区間長の所定値は、「３フレーム以下」ではなく、「５フレーム以下」でも、「６フレーム以下」でも良い。
【０１６６】
次に、評定部２４０２は、音素の欠落があった旨を示す評定結果（例えば、「音素の欠落が発生しました。」）を構成する。そして、出力部２４０３は、構成した評定結果を出力する。なお、出力部２４０３は、通常の入力音声に対しては、上述したように評定値を出力することが好適である。
【０１６７】
以上、本実施の形態によれば、ユーザが入力した発音を、教師データに対して、如何に似ているかを示す類似度（評定値）を算出し、出力できる。その場合、特殊音声、特に、音素の欠落を検知できるので、極めて精度の高い評定結果が得られる。
【０１６８】
なお、本実施の形態において、音素の欠落を検知できれば良く、評定値の算出アルゴリズムは問わない。評定値の算出アルゴリズムは、上述したアルゴリズム（ＤＡＰ、ｐ－ＤＡＰ、ｔ－ＤＡＰ、ｔ－ｐ－ＤＡＰ）でも良く、または、本明細書では述べていない他のアルゴリズムでも良い。
【０１６９】
また、本実施の形態において、音素の欠落の検知アルゴリズムは、他のアルゴリズムでも良い。例えば、音素の欠落の検知において、所定長さ未満の区間であることを欠落区間の検知で必須としても良いし、区間長を考慮しなくても良い。
【０１７０】
また、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、前記区分されたフレーム毎の入力音声データを得る入力音声データ取得ステップと、前記フレーム毎の入力音声データに基づいて、特殊な音声が入力されたことを検知する特殊音声検知ステップと、前記教師データと前記入力音声データと前記特殊音声検知ステップにおける検知結果に基づいて、前記音声受付ステップで受け付けた

音声の評定を行う評定ステップと、前記評定ステップにおける評定結果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。

【０１７１】

また、上記プログラムは、前記特殊音声検知ステップにおいて、一の音素の評定値が所定の条件を満たすことを検知し、前記評定ステップにおいて、前記特殊音声検知ステップで前記所定の条件を満たすことを検知した場合に、少なくとも音素の置換または欠落があった旨を示す評定結果を構成するプログラム、である。

【０１７２】

また、上記プログラムでは、前記特殊音声検知ステップにおいて、一の音素の評定値が所定の値より低く、かつ当該音素の直前の音素または当該音素の直後の音素の評定値が所定の値より高いことを検知し、前記評定ステップにおいて、前記特殊音声検知ステップで前記所定の条件を満たすことを検知した場合に、少なくとも音素の欠落があった旨を示す評定結果を構成することが好適である。

【０１７３】

さらに、上記プログラムでは、前記特殊音声検知ステップにおいて、一の音素の評定値が所定の値より低く、かつ当該音素の直前の音素または当該音素の直後の音素の評定値が所定の値より高く、かつ当該音素の区間長が所定の長さよりも短いことを検知することが好適である。

【０１７４】

また、実施の形態５から実施の形態８において検出した特殊音声は、無音、挿入、置換、欠落であった。発音評定装置は、かかるすべての特殊音声について検知しても良いことはいうまでもない。また、発音評定装置は、主として、実施の形態１から実施の形態４において述べた評定値の算出アルゴリズムを利用して、特殊音声の検出を行ったが、他の評定値の算出アルゴリズムを利用しても良い。

【０１７５】

また、特殊音声は、無音、挿入、置換、欠落に限られない。例えば、特殊音声は、ｇａｒｂａｇｅ（雑音などの雑多な音素等）であっても良い。受け付けた音声にｇａｒｂａｇｅが混入している場合、その区間は類似度の計算対象から除外するのがしばしば望ましい。例えば、発音評定においては、学習者の発声には通常、息継ぎや無声区間などが数多く表れ、それらに対応する発声区間を評定対象から取り除くことが好適である。なお、無音は、一般に、ｇａｒｂａｇｅの一種である、と考える。

【０１７６】

そこで，どの音素にも属さない雑多な音素（ｇａｒｂａｇｅ音素）を設定し、ｇａｒｂａｇｅのＨＭＭをあらかじめ格納しておく。スコア低下区間において、ｇａｒｂａｇｅのＨＭＭに対する評定値（$\gamma_t(j)$）が所定の値より大きい場合，その区間はｇａｒｂａｇｅ区間と判定することは好適である。特に、発音評定において，ｇａｒｂａｇｅ区間が２つの単語にまたがっている場合、息継ぎなどが起こったものとして、評定値の計算対象から除外することは極めて好適である。

【０１７７】

また、図３４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３４は、このコンピュータシステム３４０の概観図であり、図３５は、コンピュータシステム３４０のブロック図である。

【０１７８】

図３４において、コンピュータシステム３４０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３４１と、キーボード３４２と、マウス３４３と、モニタ３４４と、マイク３４５とを含む。

【０１７９】

図１７において、コンピュータ３４１は、ＦＤドライブ３４１１、ＣＤ－ＲＯＭドライブ３４１２に加えて、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）３４１３と、ＣＰＵ３４１３、ＣＤ－ＲＯＭドライブ３４１２及びＦＤドライブ３４１１に接続されたバス３４１４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）３４１５と、ＣＰＵ３４１３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）３４１６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３４１７とを含む。ここでは、図示しないが、コンピュータ３４１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。

【０１８０】

コンピュータシステム３４０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、ＣＤ－ＲＯＭ３５０１、またはＦＤ３５０２に記憶されて、ＣＤ－ＲＯＭドライブ３４１２またはＦＤドライブ３４１１に挿入され、さらにハードディスク３４１７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３４１に送信され、ハードディスク３４１７に記憶されても良い。プログラムは実行の際にＲＡＭ３４１６にロードされる。プログラムは、ＣＤ－ＲＯＭ３５０１、ＦＤ３５０２またはネットワークから直接、ロードされても良い。

【０１８１】

プログラムは、コンピュータ３４１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（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３４０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。

【０１８２】

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。

なお、上記プログラムにおいて、ハードウェアによって行われる処理、例えば、出力ステップにおけるディスプレイなどで行われる処理（ハードウェアでしか行われない処理）は含まれない。

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】

【０１８３】

以上のように、本発明にかかる発音評定装置は、比較対象の音声と入力音声の類似度を精度高く評定できるという効果を有し、語学学習装置や物真似練習装置等として有用である。

【図面の簡単な説明】

【０１８４】

【図１】実施の形態１における発音評定装置のブロック図

【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ＨＭＭの仕様を説明する図

【図４】同音声分析条件を説明する図

【図５】同評定値算出手段が算出した評定値を示すグラフを示す図

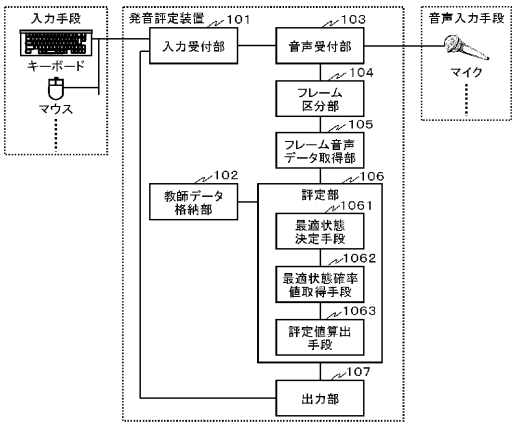
【図６】同評定値算出手段が算出した評定値を示すグラフを示す図

【図７】同出力部が出力する出力態様を示す図

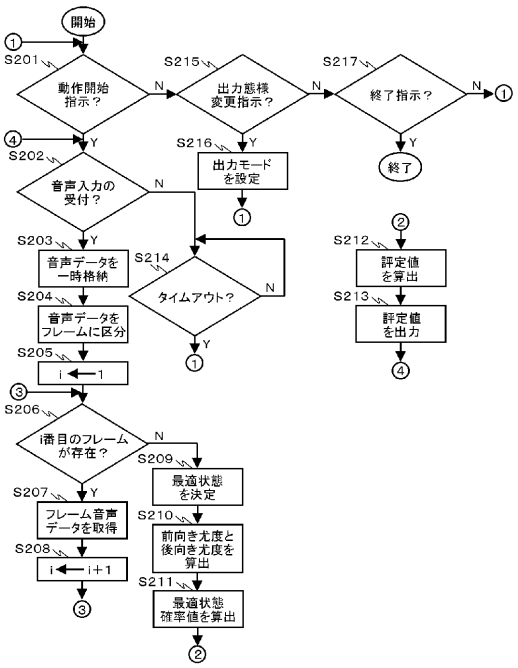
【図８】同出力部が出力する出力態様を示す図

【図９】実施の形態２における発音評定装置のブロック図
【図１０】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１１】同評定値算出手段が算出した評定値を示すグラフを示す図
【図１２】同評定値算出手段が算出した評定値を示すグラフを示す図
【図１３】実施の形態３における発音評定装置のブロック図
【図１４】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１５】同評定値算出手段が算出した評定値を示す図
【図１６】同出力部が出力する出力態様を示す図
【図１７】実施の形態４における発音評定装置のブロック図
【図１８】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１９】同評定値算出手段が算出した評定値を示す図
【図２０】４つの実施の形態で算出した評定値について説明する図
【図２１】実施の形態５における発音評定装置のブロック図
【図２２】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２３】同特殊音声の検知について説明する図
【図２４】実施の形態６における発音評定装置のブロック図
【図２５】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２６】同特殊音声の検知について説明する図
【図２７】同評定結果の出力例を示す図
【図２８】実施の形態７における発音評定装置のブロック図
【図２９】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３０】同特殊音声の検知について説明する図
【図３１】実施の形態８における発音評定装置のブロック図
【図３２】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３３】同特殊音声の検知について説明する図
【図３４】同発音評定装置を構成するコンピュータシステムの概観図
【図３５】同発音評定装置を構成するコンピュータのブロック図
【符号の説明】
【０１８５】
１０１　入力受付部
１０２　教師データ格納部
１０３　音声受付部
１０４　フレーム区分部
１０５　フレーム音声データ取得部
１０６、９０６、１３０６、１７０６，２１０２、２４０２，２８０２，３１０２　評定部
１０７、２４０３　出力部
１０６１　最適状態決定手段
１０６２　最適状態確率値取得手段
１０６３、９０６３、１３０６３、１７０６３，２１０２３　評定値算出手段
９０６２　音韻確率値取得手段
１３０６２　発音区間確率値取得手段
１７０６２　発音区間フレーム音韻確率値取得手段
２１０１、２４０１、２８０１，３１０１　特殊音声検知部
２１０２１　無音データ格納手段
２１０２２　無音区間検出手段

【図１】



【図２】



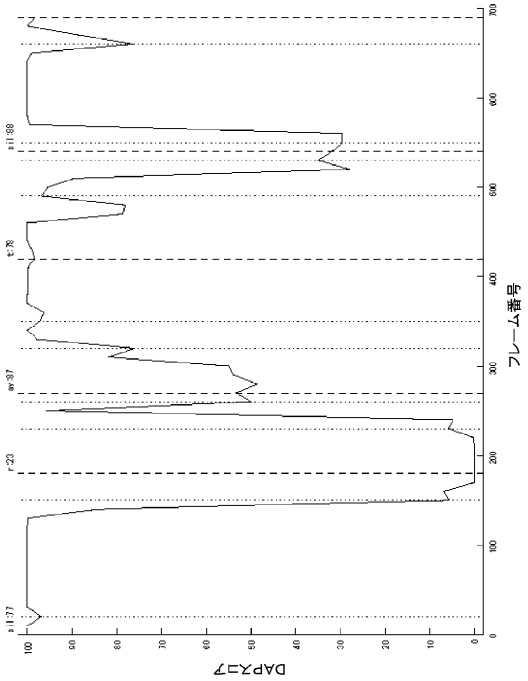
【図３】

HMMタイプ	混合ガウス型，mono phone（対角共分散行列）
状態数（モデル毎）	3
混合数（状態毎）	8
HMM総数	44
総状態数	132

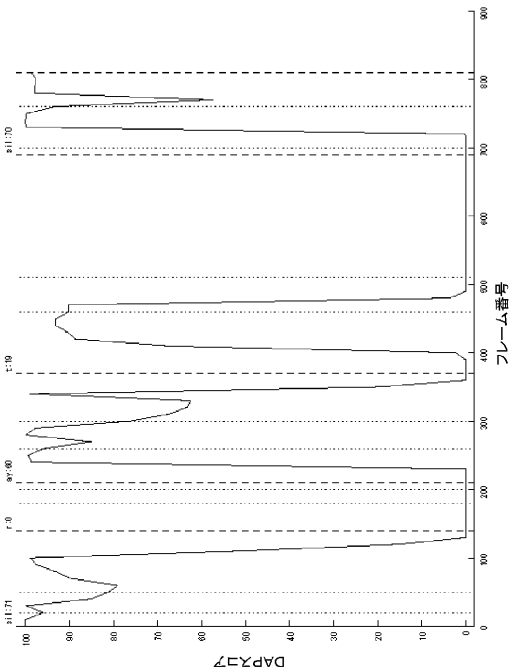
【図４】

サンプリング周波数	22.05 KH_Z
プリエンファシス	$1-0.97_{Z^{-1}}$
窓関数	Hamming窓
分析フレーム長	25 msec
フレーム周期	10 msec
特徴パラメータ	MFCC（39次元）

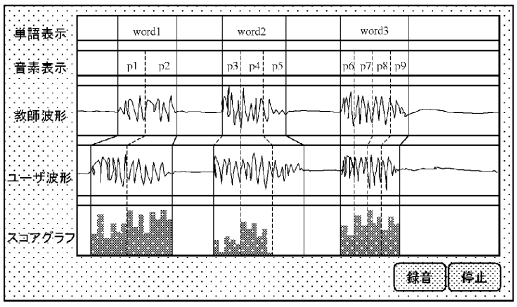
【図５】



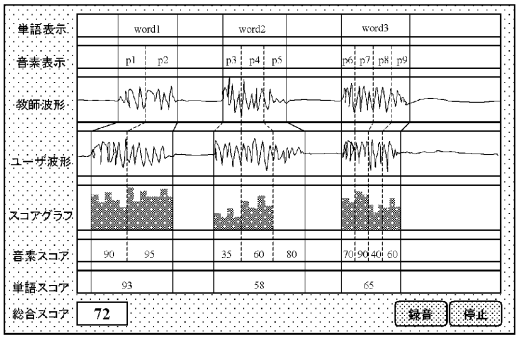
【図６】



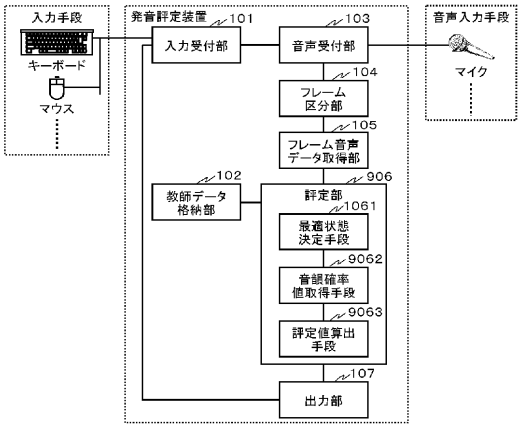
【図７】



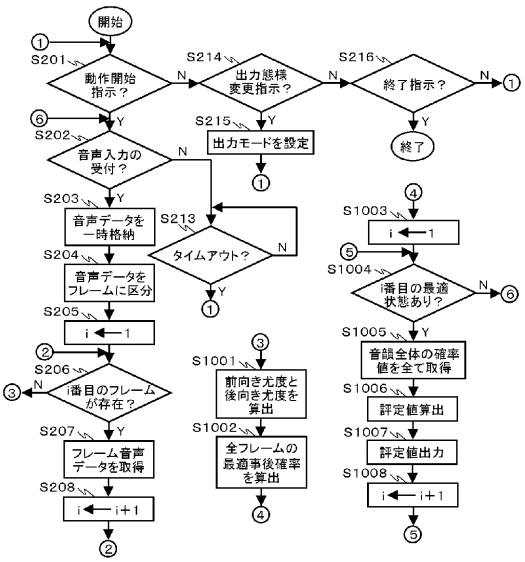
【図８】



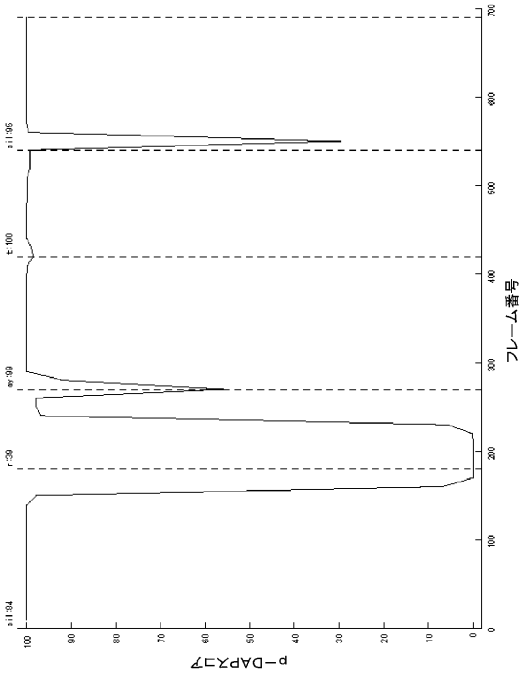
【図９】



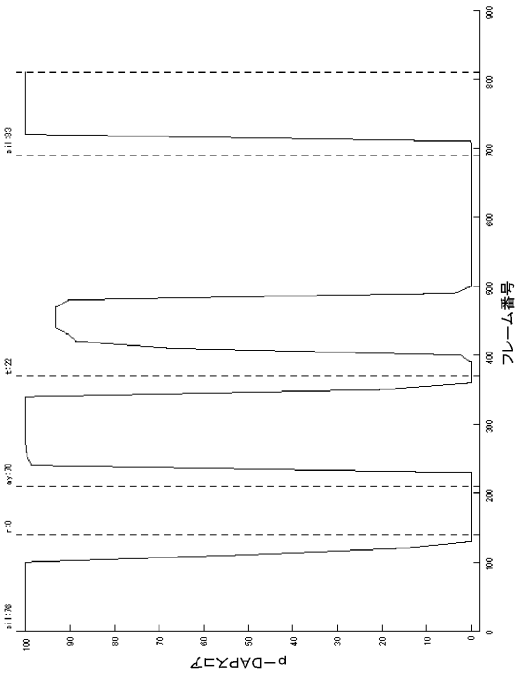
【図１０】



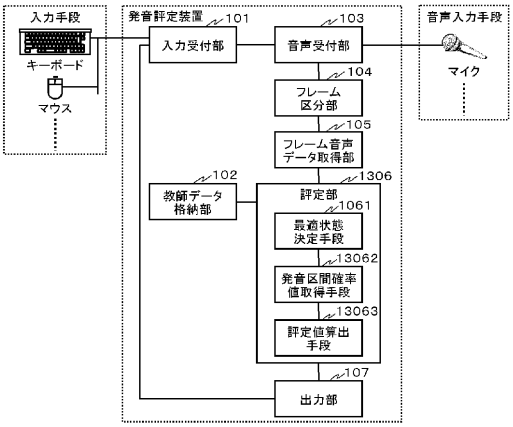
【図１１】



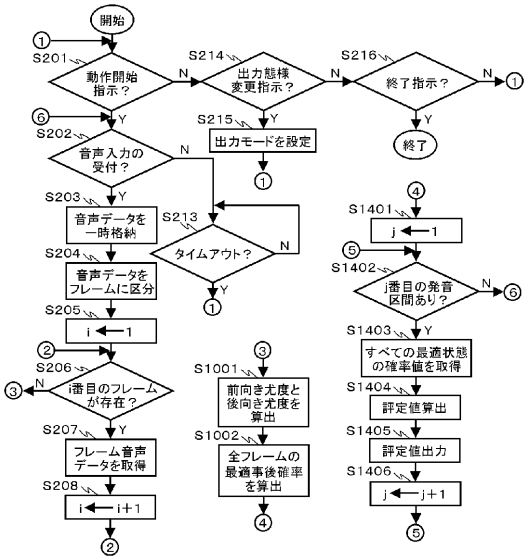
【図１２】



【図１３】



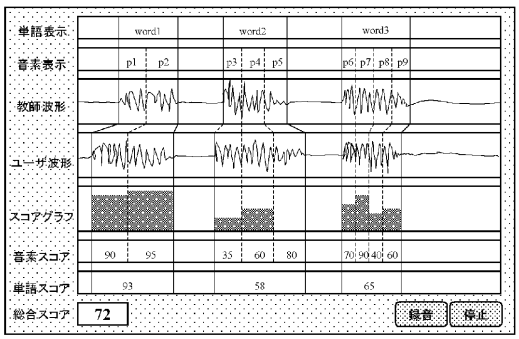
【図１４】



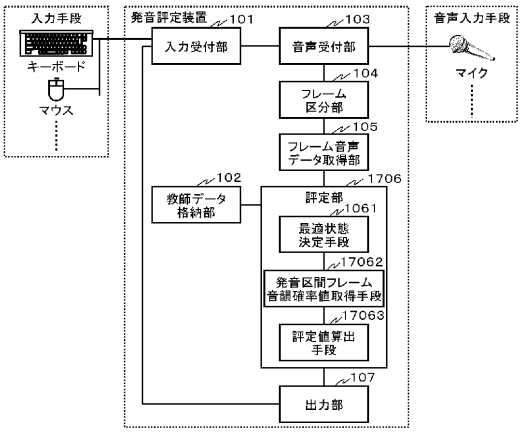
【図１５】

	Phoneme			Word
	/r/	/ay/	/t/	/right/
アメリカ人男性	23	87	78	68
日本人男性	0	60	19	29

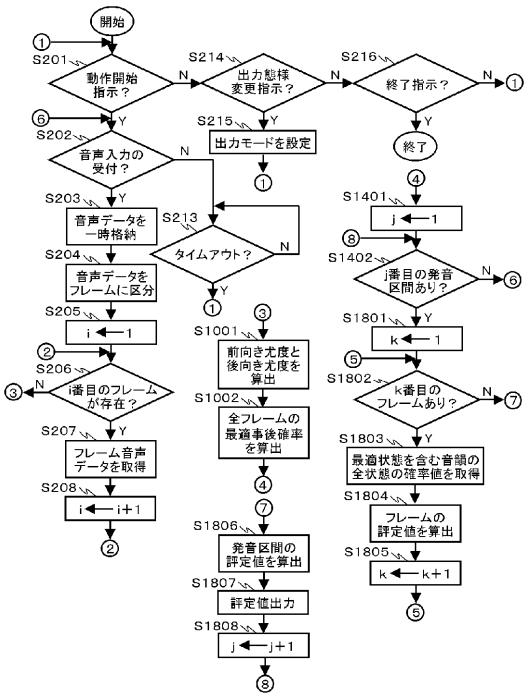
【図１６】



【図１７】



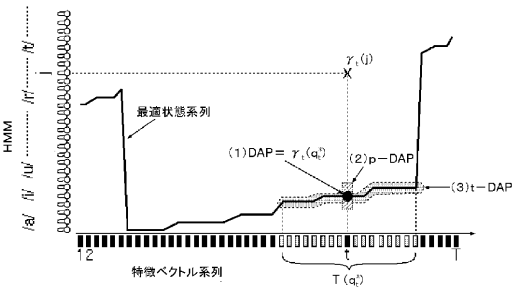
【図１８】



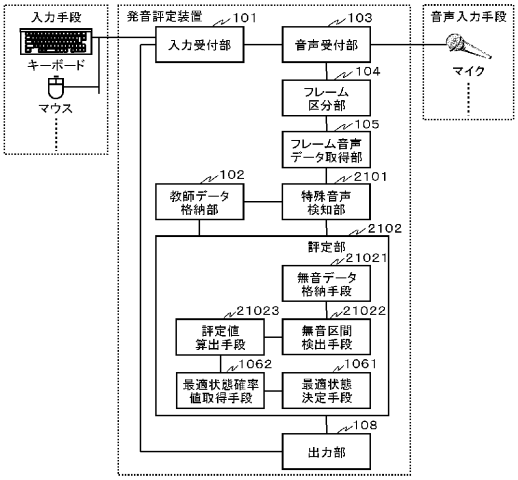
【図１９】

	Phoneme /r/	/ay/	/t/	Word /right/
アメリカ人男性	39	99	100	84
日本人男性	0	70	22	33

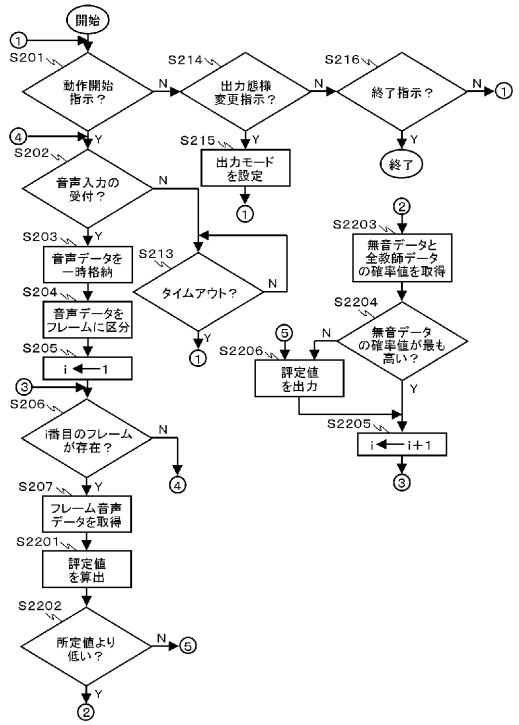
【図２０】



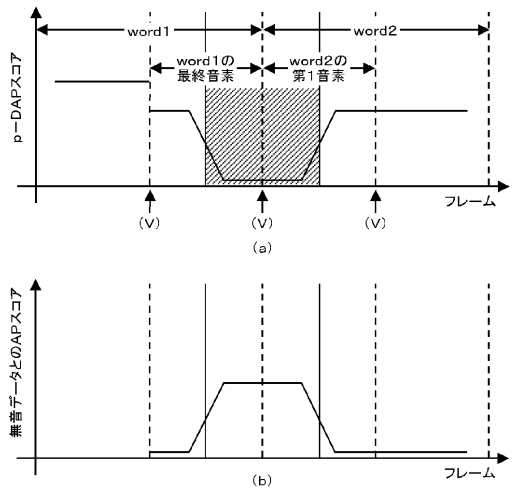
【図２１】



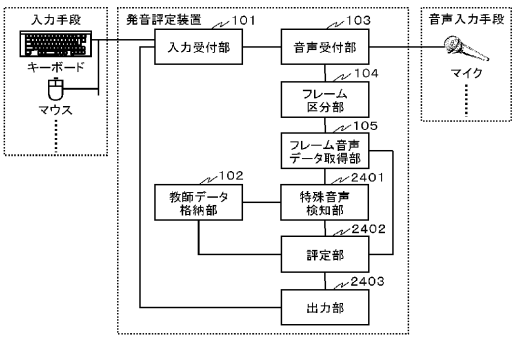
【図２２】



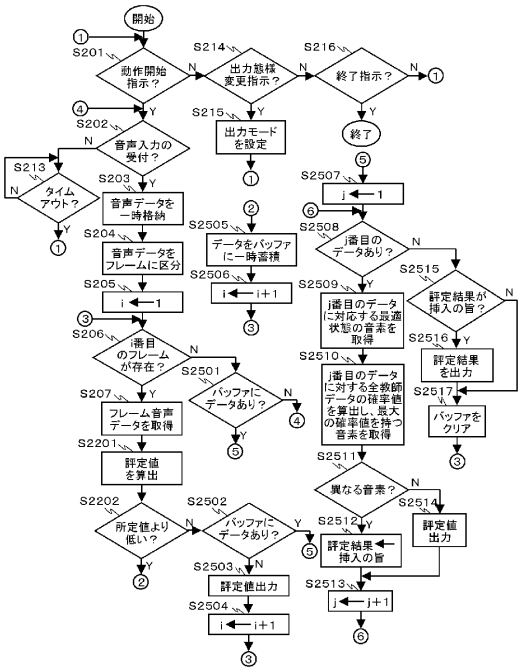
【図２３】



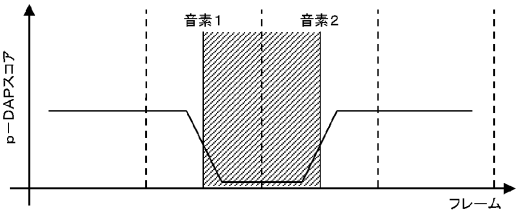
【図２４】



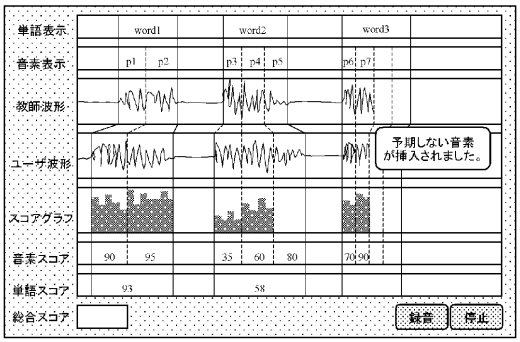
【図２５】



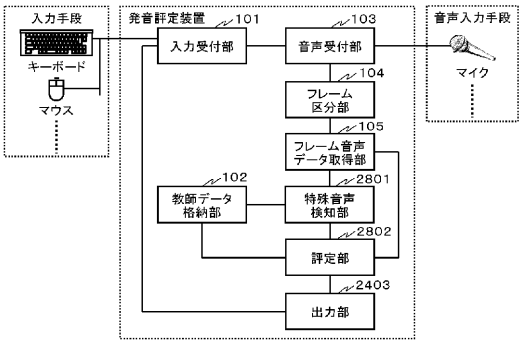
【図２６】



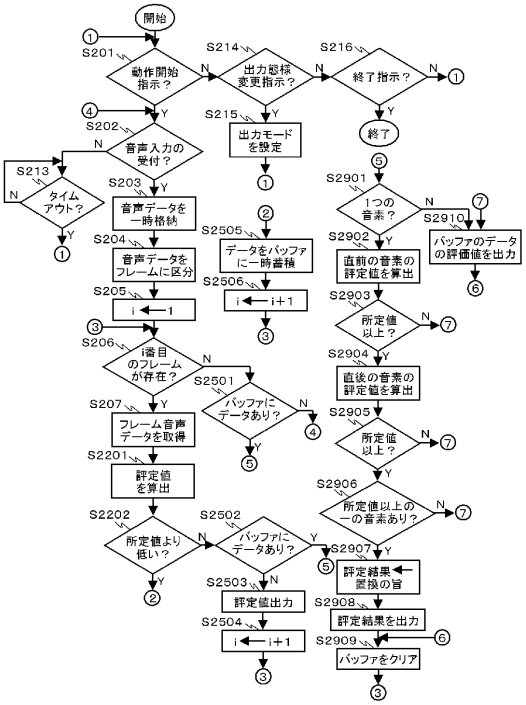
【図２７】



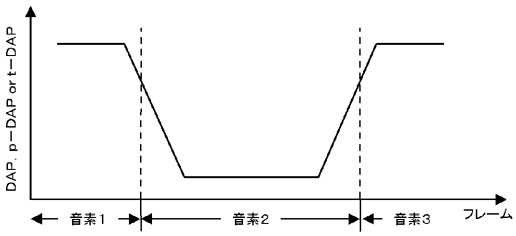
【図２８】



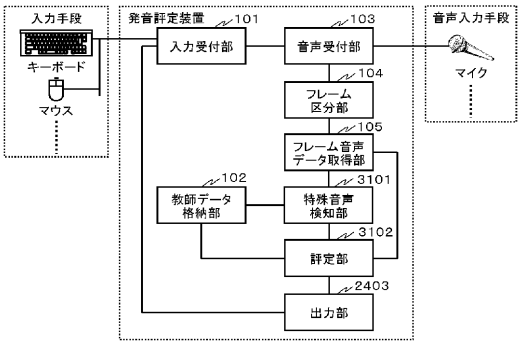
【図２９】



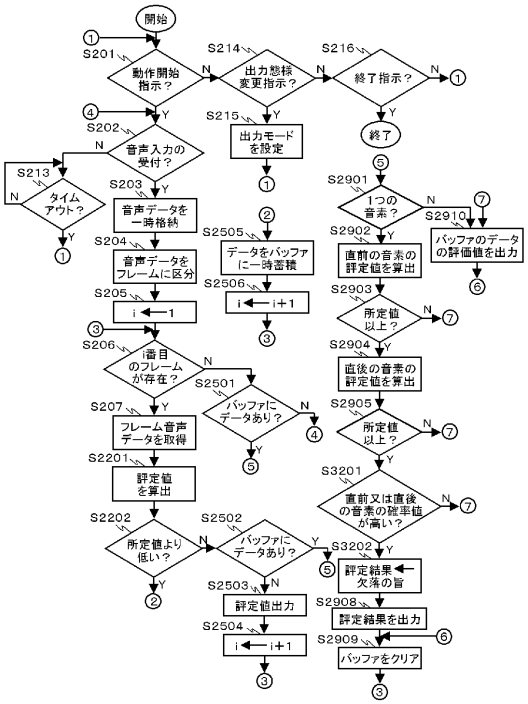
【図３０】



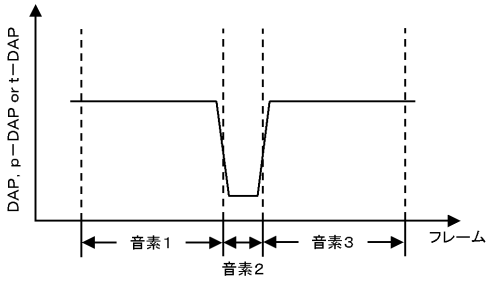
【図３１】



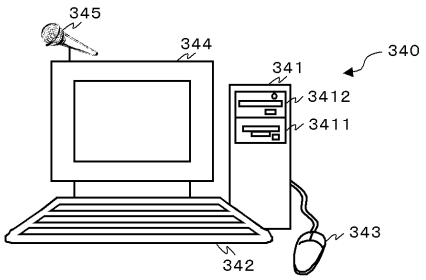
【図３２】



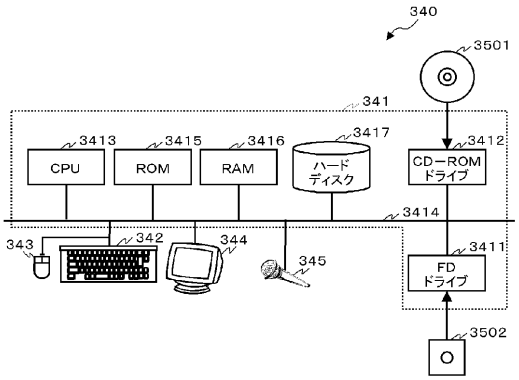
【図３３】



【図３４】



【図３５】



フロントページの続き

審査官　植田　泰輝

(56)参考文献　特開２００１－２６５２１１（ＪＰ，Ａ）
特開平９－２５８７６５（ＪＰ，Ａ）
特開平６－１１０４９４（ＪＰ，Ａ）
特開２００３－３４５３８０（ＪＰ，Ａ）
特表平７－５０３５５９（ＪＰ，Ａ）
特開２００４－１１７５３０（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０９Ｂ　５／００－７／１２，１９／００－１９／２６
Ｇ１０Ｌ　１５／００－１５／２８