

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4859125号
(P4859125)

(45) 発行日 平成24年1月25日(2012.1.25)

(24) 登録日 平成23年11月11日(2011.11.11)

(51) Int.Cl.		F I			
G 1 O L	15/00	(2006.01)	G 1 O L	15/00	2 O O E
G 1 O L	11/00	(2006.01)	G 1 O L	11/00	4 O 2 F
G 1 O L	15/10	(2006.01)	G 1 O L	15/10	2 O O C
G 1 O L	15/14	(2006.01)	G 1 O L	15/10	3 O O F
			G 1 O L	15/14	2 O O B

請求項の数 9 (全 32 頁)

(21) 出願番号 特願2007-27903 (P2007-27903)
 (22) 出願日 平成19年2月7日(2007.2.7)
 (65) 公開番号 特開2008-191551 (P2008-191551A)
 (43) 公開日 平成20年8月21日(2008.8.21)
 審査請求日 平成22年1月25日(2010.1.25)

(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 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
 (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

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】

1以上の音素毎の音響モデル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、
 評価対象の2以上の音素の並びであり、同一の音素が連続する1以上の最適音素系列の集合である最適音素部分系列を1組以上格納している最適音素部分系列格納部と、
 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、フレームに区分し、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部と、
 前記1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、前記最適音素部分系列格納部の各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する特徴ベクトル部分系列取得部と、
 前記教師データ格納部から教師データを読み出し、当該教師データを用いて、前記特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列が評価対象の音素である事後確率を算出し、当該事後確率から音声の評価値を算出する評価値算出部と、
 前記評価値算出部が算出した評価値を出力する出力部を具備する発音評定装置。

【請求項 2】

評価対象の2以上の音素の並びの情報である評価対象音素系列を格納している評価対象音

素系列格納部と、

前記評定対象音素系列に沿った音響モデルを前記教師データ格納部から読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、

前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の集合である最適音素部分系列を1組以上取得する最適音素部分系列取得部をさらに具備し、

前記特徴ベクトル部分系列取得部は、

前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する請求項1記載の発音評定装置。

【請求項3】

前記評定値算出部は、

全ての音響モデルである教師データが与えられたという条件のもと、音素区間音声データの時刻ごとに、当該音素区間音声データが、評定対象の正解となる音素に対応した音響モデルの持つ状態である事後確率を算出し、当該事後確率から音素の評定値を算出する請求項1または請求項2記載の発音評定装置。

【請求項4】

前記評定値算出部は、

音素区間音声データの時刻ごとに、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出する請求項1記載または請求項2記載の発音評定装置。

【請求項5】

前記評定値算出部は、

音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出する

請求項1または請求項2記載の発音評定装置。

【請求項6】

前記評定値算出部は、

フレーム毎または音素毎の前記評定値を基にして、文章や単語の評定値を算出する請求項1から請求項5いずれか記載の発音評定装置。

【請求項7】

各音素の時間に関する情報である音素時間情報を取得する音素時間情報取得部をさらに具備し、

前記評定値算出部は、

音素時間長を重みとした音素の評定値の加重平均を算出し、当該加重平均から文章または単語の評定値を算出する請求項6記載の発音評定装置。

【請求項8】

記録媒体に、

1以上の音素毎の音響モデルである1以上の教師データと、

評定対象の2以上音素の並びであり、同一の音素が連続する1以上の最適音素系列の集合である1組以上の最適音素部分系列を格納しており、

コンピュータを、

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、フレームに区分し、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部と、

前記1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、前記記録媒体に格納され

10

20

30

40

50

ている各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する特徴ベクトル部分系列取得部と、
前記記録媒体に格納されている教師データを読み出し、当該教師データを用いて、前記特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列が評価対象の音素である事後確率を算出し、当該事後確率から音声の評価値を算出する評価値算出部と、
 前記評価値算出部が算出した評価値を出力する出力部として機能させるためのプログラム。

【請求項9】

前記記憶媒体に、

評価対象の2以上の音素の並びの情報である評価対象音素系列をさらに格納しており、
 コンピュータを、

前記記憶媒体に格納されている評価対象音素系列に沿った音響モデルを読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、

前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の集合である最適音素部分系列を1組以上取得する最適音素部分系列取得部として、さらに機能させ、

前記特徴ベクトル部分系列取得部は、

前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得するように機能させるための請求項8記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された音声の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価装置等に関するものである。

【背景技術】

【0002】

従来、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以上格納しており、音声入力を受け付ける音声受付部と、前記受け付けた音声を、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1以上得るフレーム音声データ取得部と、前記教師データと前記1以上のフレーム音声データに基づいて、前記音声受付部が受け付けた音声の評価を行う評価部と、前記評価部の評価結果を出力する出力部を具備する発音評価装置があった（例えば、特許文献1参照）。上記の発音評価装置において、例えば、p-DAPと呼ばれるアルゴリズムを用いて発音評価を行っていた。p-DAP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）を表すように評価値を算出する。

【特許文献1】特開2006-227587号公報（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】

しかしながら、従来の発音評価装置においては、フレームの属する音素区間以外の音声データの情報も評価点に混入していた。そのために、音素単位の評価精度が十分である、とは言えなかった。

【課題を解決するための手段】

【0004】

本第一の発明の発音評価装置は、1以上の音素毎の音響モデル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、評価対象の1以上音素の並びのであり、同一の音素

10

20

30

40

50

が連続する1以上の最適音素系列の集合である最適音素部分系列を1組以上格納している最適音素部分系列格納部と、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、フレームに区分し、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部と、前記1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、前記最適音素部分系列格納部の各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する特徴ベクトル部分系列取得部と、前記教師データ格納部から教師データを読み出し、当該教師データを用いて、前記特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列が評定対象の音素である事後確率を算出し、当該事後確率から音声の評定値を算出する評定値算出部と、前記評定値算出部が算出した評定値を出力する出力部を具備する発音評定装置である。

10

【0005】

かかる構成により、音素単位の評定を高い精度で行うことができる。

【0006】

また、本第二の発明の発音評定装置は、第一の発明に対して、評定対象の1以上音素の並びの情報である評定対象音素系列を格納している評定対象音素系列格納部と、前記評定対象音素系列に沿った音響モデルを前記教師データ格納部から読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の組である最適音素部分系列を1組以上取得する最適音素部分系列取得部をさらに具備し、前記特徴ベクトル部分系列取得部は、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する発音評定装置である。

20

【0007】

かかる構成により、音素単位の評定を高い精度で行うことができる。また、予め1組以上の最適音素部分系列を用意しておく必要がない。

30

【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、前記評定値算出部は、全ての音響モデルである教師データが与えられたという条件のもと、音素区間音声データの時刻ごとに、当該音素区間音声データが、評定対象の正解となる音素に対応した音響モデルの持つ状態である事後確率を算出し、当該事後確率から音素の評定値を算出する発音評定装置である。

【0009】

かかる構成により、音素単位の評定を高い精度で行うことができる。

【0010】

また、本第四の発明の発音評定装置は、第一記載、第二いずれかの発明に対して、前記評定値算出部は、音素区間音声データの時刻ごとに、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出する発音評定装置である。

40

【0011】

かかる構成により、高速に音素単位の評定をより高い精度で行うことができる。

【0012】

また、本第五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、前記評定値算出部は、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出する発音評定装置である。

【0013】

50

かかる構成により、より高速に音素単位の評定を高い精度で行うことができる。

【0014】

また、本第六の発明の発音評定装置は、第五の発明に対して、前記評定値算出部は、前記同時生起確率値を、バックワードアルゴリズムを用いずにフォワードアルゴリズムにより取得する発音評定装置である。

【0015】

かかる構成により、高速に音素単位の評定をより高い精度で行うことができる。

【0016】

また、本第七の発明の発音評定装置は、第一から第六いずれかの発明に対して、前記評定値算出部は、フレーム毎または音素毎の前記評定値を基にして、文章や単語の評定値を算出する発音評定装置である。

10

【0017】

かかる構成により、文章や単語の評定値を算出できる。

【0018】

また、本第八の発明の発音評定装置は、第七の発明に対して、各音素の時間に関する情報である音素時間情報を取得する音素時間情報取得部をさらに具備し、前記評定値算出部は、音素時間長を重みとした音素の評定値の加重平均から文章または単語の評定値を算出する発音評定装置である。

【0019】

かかる構成により、文章や単語の評定値を精度高く算出できる。

20

【発明の効果】

【0020】

本発明による発音評定装置によれば、音素単位の評定を高い精度で行うことができる。

【発明を実施するための最良の形態】

【0021】

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。

(実施の形態1)

【0022】

30

本実施の形態において、最適状態系列から求めることができる音素区間ごとに評定点を算出するアルゴリズムにより発音評定を行う発音評定装置について説明する。このアルゴリズムは、音素区間ごとに動的事後確率を算出するということと、後に解説するアルゴリズムとの区別のために修飾子を付けて、PDAPS-PE (Phoneme Ergodic Phoneme unit Dynamic A posteriori Probability based pronunciation Scoring / 音韻エルゴディック音韻単位動的事後確率に基づく発音評定) と呼ぶ。

【0023】

PDAPS-PEアルゴリズムは、評定値の算出に用いる音声データを音素区間に限定し、その音素区間以外の情報を排除したアルゴリズムである。

【0024】

40

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、評定対象音素系列格納部103、音声受付部104、フレーム音声データ取得部105、特徴ベクトル系列取得部106、最適状態系列取得部107、最適音素部分系列取得部108、特徴ベクトル部分系列取得部109、評定値算出部110、出力部111を具備する。

【0025】

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、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

50

実現され得る。

【 0 0 2 6 】

教師データ格納部 1 0 2 は、1以上の音素毎の音響モデル（以下、適宜、音響モデルパラメータ、とも言う。）である教師データを1以上格納している。教師データは、隠れマルコフモデル（HMM）に基づくデータであることは好適である。さらに、本実施の形態における発音評定装置（PDAPS-PEアルゴリズムで利用する場合）では、教師データは、HMMであり、一の音素HMMの終端状態から、当該一の音素または他のすべての音素の始端状態へ連結された音響モデルであることは好適である。つまり、実施の形態2以降の発音評定装置（PDAPS、PAPPS、PAPPS-FNのアルゴリズムで利用する場合）では、教師データは、一の音素HMMの終端状態から、当該一の音素または他のすべての音素の始端状態へ連結された音響モデルではない。

10

【 0 0 2 7 】

なお、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：ガウシア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。なお、音響モデルは、例えば、音響を識別するIDと音響を特徴付ける特徴ベクトルの組である。教師データ格納部 1 0 2 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 2 8 】

評価対象音素系列格納部 1 0 3 は、評価対象の1以上音素の並びの情報である評価対象音素系列を格納している。評価対象音素系列は、ユーザが、本来、発音しなければならない単語や文章（正解の単語や文章）などの音素列を示す情報である。音素の情報は、例えば、音素のIDや音素を示す文字コードなどである。1以上音素の並びの情報とは、例えば、[a , o , i]などの音素の文字コード列や、[1 , 5 , 2]などの音素のID列などである。また、評価対象音素系列格納部 1 0 3 の評価対象音素系列は、入力受付部 1 0 1 が受け付けた情報でも良いし、予め格納されていても良い。評価対象音素系列格納部 1 0 3 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。

20

【 0 0 2 9 】

音声受付部 1 0 4 は、評価対象の音声の入力を受け付ける。ここで受け付けとは、通常、マイクからの受け付けであるが、記録媒体に格納された音声を読み込む処理、外部装置からの受信処理等でも良い。音声受付部 1 0 4 は、例えば、マイクのドライバースoftwareで実現され得る。また、なお、音声受付部 1 0 4 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。

30

【 0 0 3 0 】

フレーム音声データ取得部 1 0 5 は、音声受付部 1 0 4 が受け付けた音声を、フレームに区分し、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以上得て、当該1以上のフレーム音声データをメモリ上に配置する。フレーム音声データ取得部 1 0 5 は、公知技術により実現され得る。フレーム音声データ取得部 1 0 5 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 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。

40

【 0 0 3 1 】

特徴ベクトル系列取得部 1 0 6 は、フレーム音声データ取得部 1 0 5 が取得した1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である特徴ベクトル系列「 $O = o_1, o_2, \dots, o_T$ 」を取得する。特徴ベクトル系列取得部 1 0 6 は、フレーム音声データを、スペクトル分析することにより、特徴ベクトル（ o_t ）を取得する。また、特徴ベクトルは、例えば、三角型フィルタを用いたチャンネル数24のフィルタバンク出力を離散コサイン変換したMFCCであり、その静的パラメータ、デルタパラメータお

50

よびデルタデルタパラメータをそれぞれ1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。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。なお、特徴ベクトル系列取得部106は、公知技術により実現され得る。特徴ベクトル系列取得部106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。

【0032】

最適状態系列取得部107は、評定対象音素系列格納部103から評定対象音素系列を読み出し、当該読み出した評定対象音素系列に沿った音響モデルを、教師データ格納部102から読み出し、メモリ上に配置する。そして、最適状態系列取得部107は、当該読み出した音響モデルと、特徴ベクトル系列取得部106が取得した特徴ベクトル系列($O = o_1, o_2, \dots, o_t, \dots, o_T$)であり、音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態(特徴ベクトル o_t に対する最適状態)の集合である最適状態系列($q^*_1, q^*_2, \dots, q^*_t, \dots, q^*_T$)を取得する。最適状態系列取得部107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。最適状態系列取得部107が最適状態を決定するアルゴリズムは、例えば、Viterbialゴリズムによる。かかる場合、最適状態系列取得部107は、通常、上記で評定対象の音素系列に沿って連結した音響モデル(ここでは、HMM)を用いて最適状態を決定する。最適状態系列取得部107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。最適状態系列取得部107は、通常、MPUやメモリ等から実現され得る。最適状態系列取得部107の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている。但し、ハードウェア(専用回路)で実現しても良い。

【0033】

最適音素部分系列取得部108は、最適状態系列取得部107が取得した最適状態系列($q^*_1, q^*_2, \dots, q^*_t, \dots, q^*_T$)の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の集合である最適音素部分系列を1組以上取得する。つまり、最適状態(q^*_t)の属する音響モデルを(p^*_t)とすると、最適音素部分系列取得部108は、最適状態系列($q^*_1, q^*_2, \dots, q^*_t, \dots, q^*_T$)中の各最適状態に対応する音響モデルを、教師データ格納部102から読み出し、最適音素系列($p^*_1, p^*_2, \dots, p^*_t, \dots, p^*_T$)を取得する。そして、最適音素部分系列取得部108は、最適音素系列($p^*_1, p^*_2, \dots, p^*_t, \dots, p^*_T$)の中で同じ音素が連続した部分を検出し、音素系列の部分系列($\{p^*_1, p^*_2, \dots\}^{(1)} \dots \{p^*_t, \dots, p^*_T\}^{(n)} \dots \{p^*_T\}^{(s)}\}$)を得る。なお、音素(p^*_t)は、例えば、音素を識別するIDや音素を示す文字のコードなどである。また、 $\{p^*_t, \dots, p^*_T\}^{(n)}$ が、最適音素部分系列の一つである。最適音素部分系列取得部108は、通常、MPUやメモリ等から実現され得る。最適音素部分系列取得部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。

【0034】

特徴ベクトル部分系列取得部109は、特徴ベクトル系列取得部106が取得した特徴ベクトル系列から、最適音素部分系列取得部108が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列($\{o^*_1, o^*_2, \dots\}^{(1)} \dots \{o^*_t, \dots, o^*_T\}^{(n)} \dots \{o^*_T\}^{(s)}\}$)を1組以上取得する。なお、この特徴ベクトル部分系列が評定対象の音素区間音声データであり、それに対応した最適音素部分系列の代表音素が評定対象の(正解の)音

10

20

30

40

50

素となる。特徴ベクトル部分系列取得部 109 は、通常、MPU やメモリ等から実現される。特徴ベクトル部分系列取得部 109 の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアは ROM 等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

【0035】

評定値算出部 110 は、音素の評定値を音素区間データから算出する。評定値算出部 110 は、音素の評定値を算出する場合に、入力された音声データをすべて用いるようなことはしない。つまり、評定値算出部 110 は、全ての音響モデルである教師データが与えられたという条件のもと、音素区間音声データの時刻ごとに、当該音素区間音声データが、評定対象の正解となる音素に対応した音響モデルの持つ状態である事後確率を算出し、当該事後確率から音素の評定値を算出する。なお、評定対象の状態は、正解となる音素に対応した音響モデルが有するものである。具体的には、評定値算出部 110 は、例えば、以下の数式 1、数式 2 により、音声の評定値 (PDAPS-PE(t)) を算出する。つまり、評定値算出部 110 は、最適音素部分系列取得部 108 が取得した最適音素部分系列と、その部分系列に対応する特徴ベクトル部分系列取得部 109 が取得した特徴ベクトル部分系列「 $O^{(n)}$ 」と、教師データ格納部 102 から読み出した全ての音響モデルを用いて、最適音素部分系列の時刻ごとの最適音素「 P_t^* 」ごとに、特徴ベクトル部分系列の時刻（フレーム）「 t 」における状態が「 j 」である事後確率を最適音素「 P_t^* 」に含まれる全ての状態において取得し、その事後確率の総和を取得することで、音声の評定値 (PDAPS-PE(t)) を算出する。さらに詳細には、評定値算出部 110 は、最適音素部分系列取得部 108 が取得した最適音素部分系列と、その部分系列に対応する特徴ベクトル部分系列取得部 109 が取得した特徴ベクトル部分系列「 $O^{(n)}$ 」と、教師データ格納部 102 から読み出した全ての音響モデルを用いて、最適音素部分系列の時刻ごとの最適音素「 P_t^* 」ごとに、全ての音響モデルの全状態において、全ての音響モデルパラメータ「 a^{ll} 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の時刻（フレーム）「 t 」における状態が「 j 」である確率を取得し、その取得した状態ごとの確率値から、全ての音響モデルパラメータ「 a^{ll} 」と特徴ベクトル系列「 $O^{(n)}$ 」が与えられたときに、そのベクトル系列の時刻（フレーム）「 t 」における状態が「 j 」である事後確率を最適音素「 P_t^* 」に含まれる全ての状態において取得し、その事後確率の総和を取得することで、音声の評定値 (PDAPS-PE(t)) を算出する。

【数 1】

$$\begin{aligned} \text{PDAPS-PE}(t) &= \sum_{j \in P_t^*} \Pr(q_t = j | O^{(n)}, \lambda^{\text{all}}) \\ &= \sum_{j \in P_t^*} \frac{\Pr(q_t = j, O^{(n)} | \lambda^{\text{all}})}{\Pr(O^{(n)} | \lambda^{\text{all}})} \\ &= \sum_{j \in P_t^*} \frac{\Pr(q_t = j, O^{(n)} | \lambda^{\text{all}})}{\sum_{i=1}^N \Pr(q_t = i, O^{(n)} | \lambda^{\text{all}})} \end{aligned}$$

【数 2】

$$\begin{aligned} \Pr(q_t = j, O^{(n)} | \lambda^{\text{all}}) &= \Pr(q_t = j, \{o_1, o_2, \dots, o_t\}^{(n)} | \lambda^{\text{all}}) \times \Pr(\{o_{t+1}, o_{t+2}, \dots, o_T\}^{(n)} | q_t = j, \lambda^{\text{all}}) \\ &= \alpha_t(j) \beta_t(j) \end{aligned}$$

【0036】

なお、数式 1 において、「 a^{ll} 」は全ての音響モデル（教師データ）のパラメータであり、「 N 」は教師データ全体の総状態数である。また、数式 2 は、公知のアルゴリズムである、前向き後ろ向き（Forward-Backward）アルゴリズムにより、確率値が算出される

ことを示す。また、数式 1、2 等において、「 $\Pr(q_t = j, O^{(n)} | a^{11})$ 」は、全ての音響モデルパラメータ「 a^{11} 」が与えられたという条件のもと、特徴ベクトル部分系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の時刻（フレーム）「 t 」における状態が「 j 」である確率である。

【0037】

つまり、数式 2 は、前向き後ろ向きアルゴリズムを用いて、全ての音響モデルパラメータ「 a^{11} 」が与えられたという条件のもと、特徴ベクトル部分系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の時刻（フレーム）「 t 」における状態が「 j 」である確率を算出する数式である。

【0038】

以下、数式 2 における確率値を算出する方法について説明する。従来の発音評定や音声認識などに用いる音素単位の音響モデル（HMM）は、音素ごとに独立しており、1つの音響モデルのみで実行するように定義されている。前向き後ろ向きアルゴリズムは、数式 2 のように全ての音響モデルパラメータ「 a^{11} 」を用いて実行することは出来ない。そこで独立した音素ごとの音響モデルをあたかも 1つの大きな音響モデルとして振舞うように改造する。

【0039】

具体的には、例えば、レフトtoライト型の 1つの音素HMMは始端と終端の状態を持ち、その間が自己遷移を含めた状態遷移でつながれた形で独立している。その 1つの音素HMMの終端状態から自分自身もあわせて全ての音素HMMの始端状態へと状態遷移を仮想的につなぎ合わせる。こうして終端から始端への仮想状態遷移を全ての音素HMMにおいて作成し、つなぎ合わせることにより、複数の音素HMMを 1つの大きな音響モデルとして実装する。したがって、教師データは、隠れマルコフモデル（HMM）であり、一の音素HMMの終端状態から、当該一の音素または他のすべての音素の始端状態へ連結された音響モデルであることは好適である。

【0040】

以上のようにして改造された音響モデルを用いることで数式 2 の前向き後ろ向きアルゴリズムは実行される。なお、この改造された音響モデルならびにその音響モデルを用いて実行される前向き後ろ向きアルゴリズムのことを、音素エルゴディック音響モデルならびに音素エルゴディック前向き後ろ向きアルゴリズムと呼ぶこととする。

【0041】

つまり、評定値算出部 110 は、取得した 1 以上の事後確率を基にした音素の確率値をパラメータとして音声の評定値を算出する。評定値算出部 110 は、例えば、時間ごとの音素の確率値の時間区間における平均値や中央値を算出し、当該算出した値を評定値とする。そして、時間区間を一文章や一単語や一音素などの区間にすれば、評定値算出部 110 は、文章ごとや単語ごとや音素ごとの評定値を算出できる。評定値算出部 110 は、通常、MPU やメモリ等から実現され得る。評定値算出部 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。

【0042】

出力部 111 は、評定値算出部 110 が算出した評定値を出力する。評定値の出力態様は問わない。評定値は数値で出力されても良いし、折れ線グラフや棒グラフ等で出力されても良い。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考えても良い。出力部 111 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。

【0043】

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。

【0044】

10

20

30

40

50

(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。

【0045】

(ステップS202) 音声受付部104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。

【0046】

(ステップS203) フレーム音声データ取得部105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。

10

【0047】

(ステップS204) フレーム音声データ取得部105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し、当該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取得し、1以上のフレーム音声データをメモリ上に配置する。

【0048】

(ステップS205) 特徴ベクトル系列取得部106は、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をそれぞれ音声分析し、1以上の特徴ベクトルを抽出し、特徴ベクトル系列($O = o_1, o_2, \dots, o_t, \dots, o_T$)を得る。本特徴ベクトルは、例えば、三角型フィルタを用いたチャンネル数24のフィルタバンク出力を離散コサイン変換したMFCCであり、その静的パラメータ、デルタパラメータおよびデルタデルタパラメータをそれぞれ12次元、さらに正規化されたパワーとデルタパワーおよびデルタデルタパワー(39次元)を有する。

20

【0049】

(ステップS206) 最適状態系列取得部107は、評定対象音素系列格納部103から評定対象音素系列を読み出す。

【0050】

(ステップS207) 最適状態系列取得部107は、ステップS206で読み出した評定対象音素系列に沿った音響モデルを、教師データ格納部102から読み出し、メモリ上に配置する。

【0051】

30

(ステップS208) 最適状態系列取得部107は、ステップS207で読み出した音響モデルと、ステップS205で取得した特徴ベクトル系列($O = o_1, o_2, \dots, o_t, \dots, o_T$)とを比較し、フレーム毎の最適状態(特徴ベクトル o_t に対する最適状態)の集合である最適状態系列($q^*_1, q^*_2, \dots, q^*_t, \dots, q^*_T$)を、Viterbialgorithmにより取得する。

【0052】

(ステップS209) 最適音素部分系列取得部108は、ステップS208で取得した最適状態系列($q^*_1, q^*_2, \dots, q^*_t, \dots, q^*_T$)の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の組である最適音素部分系列を1組以上取得し、メモリ上に1組以上の最適音素部分系列を配置する。各最適音素部分系列の区切りの情報は、何でも良い。つまり、各最適音素部分系列は、異なるバッファに格納されても良いし、区切りの情報(例えば「,」)が、最適音素部分系列の間に挿入されていても良い。

40

【0053】

(ステップS210) 特徴ベクトル部分系列取得部109は、ステップS205で取得した特徴ベクトル系列から、ステップS209で取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列($\{o^*_1, o^*_2, \dots\}^{(1)} \dots \{\dots, o^*_t, \dots, o^*_T\}^{(n)} \dots \{\dots, o^*_T\}^{(s)}$)を1組以上取得する。特徴ベクトル部分系列取得部109は、最適音素部分系列と特徴ベクトル部分系列の対応を、フレームIDや時刻でとる。

50

【 0 0 5 4 】

(ステップS 2 1 1) 評価値算出部 1 1 0 は、評価値を算出する。評価値を算出するアルゴリズム例の詳細は、図 3 のフローチャートを用いて説明する。

【 0 0 5 5 】

(ステップS 2 1 2) 出力部 1 1 1 は、評価値算出部 1 1 0 が算出した評価値を出力する。

【 0 0 5 6 】

(ステップS 2 1 3) 音声受付部 1 0 4 は、タイムアウトか否かを判断する。つまり、音声受付部 1 0 4 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S 2 0 1 に戻り、タイムアウトでなければステップS 2 0 2 に戻る。

10

【 0 0 5 7 】

(ステップS 2 1 4) 入力受付部 1 0 1 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なければ処理を終了し、終了指示を受け付けなければステップS 2 0 1 に戻る。

【 0 0 5 8 】

なお、図 2 のフローチャートにおいて、電源オフや処理終了の割り込みにより処理は終了する。

【 0 0 5 9 】

次に、ステップS 2 1 1 の評価値を算出するアルゴリズム例の詳細について、図 3 のフローチャートを用いて説明する。

20

【 0 0 6 0 】

(ステップS 3 0 1) 評価値算出部 1 1 0 は、評価値を代入するバッファを初期化する。

【 0 0 6 1 】

(ステップS 3 0 2) 評価値算出部 1 1 0 は、カウンタ i に 1 を代入する。なお、カウンタ i は、評価値を代入するバッファ (例えば、配列) のアドレス (配列の添え字) を示すカウンタである。

【 0 0 6 2 】

(ステップS 3 0 3) 評価値算出部 1 1 0 は、教師データ格納部 1 0 2 から、すべての音響モデルを読み出し、メモリ上に配置する。

30

【 0 0 6 3 】

(ステップS 3 0 4) 評価値算出部 1 1 0 は、カウンタ n に 1 を代入する。

【 0 0 6 4 】

(ステップS 3 0 5) 評価値算出部 1 1 0 は、 n 番目の最適音素部分系列が存在するかどうかを判断する。 n 番目の最適音素部分系列が存在すればステップS 3 0 6 に行き、 n 番目の最適音素部分系列が存在しなければステップS 3 1 8 に行く。

【 0 0 6 5 】

(ステップS 3 0 6) 評価値算出部 1 1 0 は、 n 番目の最適音素部分系列を取得し、メモリ上に配置する。

40

【 0 0 6 6 】

(ステップS 3 0 7) 評価値算出部 1 1 0 は、 n 番目の特徴ベクトル部分系列を取得し、メモリ上に配置する。

【 0 0 6 7 】

(ステップS 3 0 8) 評価値算出部 1 1 0 は、特徴ベクトル部分系列の時刻ごと、全ての音響モデルの状態ごとの確率を、全ての音響モデルパラメータを用いて、前向き後ろ向きアルゴリズムにより算出し、バッファに格納する。

【 0 0 6 8 】

(ステップS 3 0 9) 評価値算出部 1 1 0 は、カウンタ t に 1 を代入する。

【 0 0 6 9 】

50

(ステップS310) 評価値算出部110は、「 $t \leq T(n)$ 」を満たすか否か、つまり、評価対象の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S311に行き、t番目のフレームが存在しなければステップS317に行く。

【0070】

(ステップS311) 評価値算出部110は、t番目のフレームの状態確率の総和を算出し、メモリ上に配置する。

【0071】

(ステップS312) 評価値算出部110は、最適音素「 P_t^* 」に含まれる状態(j)を取得する。

10

【0072】

(ステップS313) 評価値算出部110は、状態(j)の事後確率を算出する。さらに具体的には、評価値算出部110は、「状態(j)の確率/状態確率の総和」により、状態(j)の事後確率を算出する。

【0073】

(ステップS314) 評価値算出部110は、ステップS313で算出した事後確率をi番目の評価値に加算し、i番目の評価値のバッファに代入する。

【0074】

(ステップS315) 評価値算出部110は、最適音素「 P_t^* 」に、次の状態が存在するか否かを判断する。次の状態が存在すればステップS312に行き、次の状態が存在しなければステップS316に行く。

20

【0075】

(ステップS316) 評価値算出部110は、カウンタt、およびカウンタiを1、インクリメントし、ステップS310に戻る。

【0076】

(ステップS317) 評価値算出部110は、カウンタnを1、インクリメントし、ステップS305に戻る。

【0077】

(ステップS318) 評価値算出部110は、フレーム毎の評価値から、音素区間、単語区間、文章区間の評価値を算出し、算出値をメモリ上に配置する。かかる、音素区間、単語区間、文章区間の評価値の算出方法は、フレーム毎の評価値の各区間における平均値や中央値を算出する方法等がある。上位処理にリターンする。

30

【0078】

以上、本実施の形態によれば、精度の高い発音評価ができる。具体的には、本実施の形態によれば、音素単位の評価を正確に行うことができる。

【0079】

なお、本実施の形態における発音評価装置の評価実験の結果について、他の発音評価装置の評価実験の結果にまとめて後述する。

【0080】

また、本実施の形態の具体例によれば、評価値算出部110は、数式1、2により、上記の事後確率値を算出した。しかし、評価値算出部110が評価値を算出するアルゴリズムは、数式1、2によるものとは限らない。例えば、数式1の事後確率の算出において、音響モデルの状態としての、起こりうる全ての事象が母音の音素を表すもののみになっても良い。かかる場合、数式2の確率を求めるために用いる音響モデルパラメータ「 a^1 」も母音の音素を表すもののみとなり、評価対象の音素は、いずれかの母音を表すもののみが許されることは明らかである。つまり、本実施の形態における発音評価装置は、全ての音響モデルパラメータが与えられたという条件のもと、音素区間音声データの時刻ごとに、音素区間音声データが評価対象の状態である事後確率から音素の評価値を算出することができる。

40

【0081】

50

また、本実施の形態において、評定対象の音声データが、予め音素区間に区切られた音声データである音素区間音声データの集合でも良い。かかる場合、発音評定装置において、最適状態系列取得部や最適音素部分系列取得部は不要である。かかる場合、発音評定装置は、同一の音素が連続する1以上の最適音素系列の集合である最適音素部分系列を1組以上格納している最適音素部分系列格納部を有する。

【0082】

また、本実施の形態において、発音評定装置が、フレーム毎の評定値を基にして、文章や単語の評定値を算出することは好適である。具体的には、発音評定装置は、文章や単語に含まれるフレーム毎の複数の評定値の平均値や中央値から評定値を算出することは好適である。

10

【0083】

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、コンピュータを、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、フレームに区分し、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部と、前記1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、格納されている各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する特徴ベクトル部分系列取得部と、格納されている教師データを読み出し、当該教師データを用いて、前記特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列が評定対象の音素である事後確率を算出し、当該事後確率から音声の評定値を算出する評定値算出部と、前記評定値算出部が算出した評定値を出力する出力部として機能させるためのプログラム、である。

20

【0084】

また、上記プログラムにおいて、コンピュータを格納されている評定対象音素系列に沿った音響モデルを読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の集合である最適音素部分系列を1組以上取得する最適音素部分系列取得部として、さらに機能させ、前記特徴ベクトル部分系列取得部は、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得するように機能させるためのプログラム、であることは好適である。

30

【0085】

また、上記プログラムにおいて、前記評定値算出部が、全ての音響モデルである教師データが与えられたという条件のもと、音素区間音声データの時刻ごとに、当該音素区間音声データが、評定対象の正解となる音素に対応した音響モデルの持つ状態である事後確率を算出し、当該事後確率から音素の評定値を算出するように機能させるためのプログラム、であることは好適である。

40

【0086】

また、上記の教師データは、隠れマルコフモデル(HMM)であり、一の音素HMMの終端状態から、当該一の音素または他のすべての音素の始端状態へ連結された音響モデルである、ことは好適である。

(実施の形態2)

50

【 0 0 8 7 】

本実施の形態において、実施の形態 1 で説明した PDAPS-PE アルゴリズムをさらに音素の評定という意味で改良したアルゴリズムにより発音評定を行う発音評定装置について説明する。このアルゴリズムは、PDAPS-PE アルゴリズムの改良であるということと、その名前の修飾子であり、そのアルゴリズムの特徴であった音素エルゴディックな要素を取り除きより純粹に対象の音素区間を評定するという一方で、PDAPS (Phoneme unit Dynamic A posteriori Probability based pronunciation Scoring / 音韻単位動的事後確率に基づく発音評定) と呼ぶこととする。

【 0 0 8 8 】

実施の形態 1 で説明した PDAPS-PE アルゴリズムにおいて、評定対象の音声データに含まれる正解となる音素があきらかに限定されているにも関わらず、数式 1 の分子にある評定対象音素であるところの状態の確率を表す特徴ベクトル列「 $O^{(n)}$ 」が観測され、かつ、時刻「 t 」における状態が「 j 」であることの確率の算出に「 a^{11} 」を用いることで、評定対象音素以外の音素の確率値が混入していた。このとき同じく数式 1 の分母にある各状態の確率を表す値にも同様に、求めたい音素以外の音素の確率が混入している。ここで、かかる混入要素を取り除いたものが PDAPS アルゴリズムである。

10

【 0 0 8 9 】

図 4 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、評定対象音素系列格納部 103、音声受付部 104、フレーム音声データ取得部 105、特徴ベクトル系列取得部 106、最適状態系列取得部 107、最適音素部分系列取得部 108、特徴ベクトル部分系列取得部 109、評定値算出部 410、出力部 111 を具備する。

20

【 0 0 9 0 】

本発音評定装置は、実施の形態 1 で述べた発音評定装置と比較して、評定値算出部のみが異なる。

【 0 0 9 1 】

評定値算出部 410 は、音素の評定値を音素区間データから算出する。評定値算出部 410 は、音素区間音声データ（音素の区間に分離された音声データ）の時刻ごとに、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出する。また、さらに具体的には、評定値算出部 410 は、例えば、以下の数式 3、数式 4 により、最適音素部分系列取得部 108 が取得した最適音素部分系列と、その部分系列に対応する特徴ベクトル部分系列取得部 109 が取得した特徴ベクトル部分系列「 $O^{(n)}$ 」と、教師データ格納部 102 から全ての音響モデルを読み出し、最適音素部分系列の時刻ごとの最適音素「 P_t^* 」ごとに、全ての音響モデルの全状態において、その状態「 j 」が属する音響モデルパラメータ「 $P^{(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の時刻（フレーム）「 t 」における状態が「 j 」である確率を取得し、その取得した状態ごとの確率値から、全ての音素（音響モデル）について音響モデルパラメータ「 $P^{(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測される確率を取得し、その取得した音素ごとの確率値から、特徴ベクトル系列「 $O^{(n)}$ 」が与えられたときに、そのベクトル系列が最適音素「 P_t^* 」である事後確率を取得することで、音声の評定値（PDAPS(t)) を算出する。

30

40

【数3】

$$\begin{aligned}
 \text{PDAPS}(t) &= \Pr(p_t^* | O^{(n)}) \\
 &= \frac{\Pr(O^{(n)} | \lambda^{(p_t^*)}) \Pr(p_t^*)}{\sum_{m=1}^M \Pr(O^{(n)} | \lambda^{(m)}) \Pr(m)} \\
 &= \frac{\sum_{j \in p_t^*} \Pr(q_t = j, O^{(n)} | \lambda^{P(j)})}{\sum_{m=1}^M \sum_{i \in p_m} \Pr(q_t = i, O^{(n)} | \lambda^{P(i)})} \\
 &= \frac{\sum_{j \in p_t^*} \Pr(q_t = j, O^{(n)} | \lambda^{P(j)})}{\sum_{i=1}^N \Pr(q_t = i, O^{(n)} | \lambda^{P(i)})}
 \end{aligned}$$

10

【数4】

$$\begin{aligned}
 \Pr(q_t = j, O^{(n)} | \lambda^{P(j)}) \\
 &= \Pr(q_t = j, \{o_1, o_2, \dots, o_t\}^{(n)} | \lambda^{P(j)}) \times \Pr(\{o_{t+1}, o_{t+2}, \dots, o_{T(n)}\}^{(n)} | q_t = j, \lambda^{P(j)}) \\
 &= \alpha_t(j) \beta_t(j)
 \end{aligned}$$

20

【0092】

なお、数式3において、「 $\lambda^{(p_t^*)}$ 」は音素「 p_t^* 」の音響モデルのパラメータであり、「 $\lambda^{P(j)}$ 」は状態「 j 」を含む音響モデルのパラメータである。状態「 j 」が音素「 p_t^* 」の音響モデルの持つ状態であるならば、「 $\lambda^{P(j)}$ 」と「 $\lambda^{(p_t^*)}$ 」は同じ音響モデルのパラメータを表す。また、「 M 」は全ての音響モデル数であり、「 N 」は音響モデル全体の総状態数である。また、数式3において「 $\Pr(m)$ 」は音素「 m 」の事前確率を表し、音響モデルパラメータ「 $\lambda^{(m)}$ 」の事前確率「 $\Pr(\lambda^{(m)})$ 」を意味する（「 $\Pr(p_t^*)$ 」についても同様）。また、数式3において音素の事前確率「 $\Pr(m)$ 」はすべての音素において一定と仮定し、式の途中で省略している。数式4は、状態「 j 」が属する音響モデルパラメータ「 $\lambda^{P(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の時刻（フレーム）「 t 」における状態が「 j 」である確率を、前向き後向き（Forward-Backward）アルゴリズムにより算出することを示す。

30

【0093】

このとき、数式3の各項における分子は評価対象音素の音響モデルパラメータが与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測される確率を表しており、評価対象音素以外の確率の混入が取り除かれている。さらに、その分母においても各状態を表すところの確率は、その状態が属する音素以外の確率の混入が取り除かれている。そして、数式3により算出される評価値は、入力音素区間評価対象音声データが、評価対象の音素モデルに近い（似た）音声データであり、それ以外の音素とは遠い（似ていない）音声データであるならば、分子と分母の値が近づいてそれは大きくなり（1に近くなり）、評価対象以外の音素モデルに近ければ分子と分母の値は遠ざかって（分子<分母）それは小さくなる（0に近くなる）。また、いずれの音素モデルとも遠ければその分子と分母の各確率値は小さな値となり評価値は小さくなる。このようにしてPDAPSアルゴリズムは、評価対象音声データの音素区間を、より純粋に評価対象の音素として評価することで、高い評価精度を得ることができる。

40

【0094】

評価値算出部410は、通常、MPUやメモリ等から実現され得る。評価値算出部41

50

0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。

【0095】

次に、発音評定装置の動作について説明する。本発音評定装置の動作において、実施の形態1の発音評定装置と比較して、評定値算出処理が異なる。したがって、図5のフローチャートを用いて、本発音評定装置の評定値算出処理について説明する。図5のフローチャートにおいて、図3のフローチャートと異なるステップについてのみ説明する。

【0096】

（ステップS501）評定値算出部410は、特徴ベクトル部分系列の時刻ごと、全ての音響モデルの状態ごとの確率を、状態の属する音響モデルパラメータを用いて、前向き後向きアルゴリズムにより算出し、バッファに格納する。なお、評定値算出部410は、上記の確率の算出前に、ステップS303で取得した音響モデルの中から、状態の属する音響モデルパラメータを取得する。

10

【0097】

（ステップS502）評定値算出部410は、カウンタmに1を代入する。なお、mは、音素のカウンタである。

【0098】

（ステップS503）評定値算出部410は、m番目の音素の音響モデルが存在するかどうかを判断する。m番目の音素の音響モデルが存在すればステップS504に行き、m番目の音素の音響モデルが存在しなければステップS506に行く。

20

【0099】

（ステップS504）評定値算出部410は、m番目の音素の音響モデルに含まれる全状態の時刻tにおける確率の総和から、m番目の音素の確率を算出する。

【0100】

（ステップS505）評定値算出部410は、カウンタmを1、インクリメントし、ステップS503に戻る。

【0101】

（ステップS506）評定値算出部410は、最適音素「 P_t^* 」の事後確率を算出する。

【0102】

（ステップS507）評定値算出部410は、算出した事後確率をi番目の評定値のバッファに格納する。

30

【0103】

（ステップS508）評定値算出部410は、カウンタt、およびカウンタiを1、インクリメントし、ステップS310に戻る。

【0104】

以上、本実施の形態によれば、精度の高い発音評定ができる。具体的には、本実施の形態によれば、音素単位の評定を精度高く行うことができる。

【0105】

なお、本実施の形態における発音評定装置の評価実験の結果について、他の発音評定装置の評価実験の結果にまとめて後述する。

40

【0106】

また、本実施の形態の具体例によれば、評定値算出部410は、数式3、4により、上記の事後確率値を算出した。しかし、評定値算出部410が評定値を算出するアルゴリズムは、数式3、4によるものとは限らない。例えば、数式3の事後確率の算出において、音素としての、起こりうる全ての事象が母音の音素を表すもののみになっても良い。かかる場合、評定対象の音素は、いずれかの母音を表すもののみが許されることは明らかである。つまり、評定値算出部410は、音素区間音声データの時刻ごとに、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出すればよい。

50

【 0 1 0 7 】

また、本実施の形態において、発音評定装置が、フレーム毎の評定値を基にして、文章や単語の評定値を算出することは好適である。具体的には、発音評定装置は、文章や単語に含まれるフレーム毎の複数の評定値の平均値や中央値から評定値を算出することは好適である。

【 0 1 0 8 】

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、コンピュータを、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、フレームに区分し、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部と、前記1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、格納されている各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する特徴ベクトル部分系列取得部と、格納されている教師データを読み出し、当該教師データを用いて、前記特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列が評定対象の音素である事後確率を算出し、当該事後確率から音声の評定値を算出する評定値算出部と、前記評定値算出部が算出した評定値を出力する出力部として機能させるためのプログラム、である。

【 0 1 0 9 】

また、上記プログラムにおいて、コンピュータを格納されている評定対象音素系列に沿った音響モデルを読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の組である最適音素部分系列を1組以上取得する最適音素部分系列取得部として、さらに機能させ、前記特徴ベクトル部分系列取得部は、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得するように機能させるためのプログラム、であることは好適である。

【 0 1 1 0 】

また、上記プログラムにおいて、前記評定値算出部が、音素区間音声データの時刻ごとに、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出するように機能させるためのプログラム、であることは好適である。

(実施の形態3)

【 0 1 1 1 】

本実施の形態において、実施の形態2で説明したPDAPSアルゴリズムをさらに改良したアルゴリズムにより発音評定を行う発音評定装置について説明する。PDAPSアルゴリズムは、入力される評定対象音声データの全体と、その最適状態系列から求めた音素区間ごとに評定点を算出し、音素エルゴディックな要素を取り除いて純粋に音素区間を評定するアルゴリズムである。PDAPSアルゴリズムの評定値算出式には、評定対象の音素区間内のどの時刻(フレーム)においても評定値が同じになるという処理効率の意味から重複した計算を含んでおり、その重複を取り除いて改良したアルゴリズムが、本実施の形態において説明するアルゴリズムである。

【 0 1 1 2 】

本アルゴリズムは、PDAPSアルゴリズムから重複処理を取り除いて改良したものであり、改良の過程で評定値の算出に動的要素を直接利用しなくなったことからPAPPS(Phonemic A Posteriori Probability based pronunciation Scoring/音韻の事後確率に基づく発音評定)と呼ぶこととする。

【 0 1 1 3 】

以下、さらに詳細に、実施の形態 2 で説明した PDAPS アルゴリズムと本実施の形態における PAPPS アルゴリズムとの違いについて述べる。PDAPS アルゴリズムは、その評定値算出過程の基礎となる数式 4 で示した前向き後ろ向きアルゴリズムが、前向きアルゴリズムの算出値としての前向き確率と、後ろ向きアルゴリズムの算出値としての後ろ向き確率を、掛け合わせた値としての特徴ベクトル系列と時刻「 t 」における状態が「 j 」であることの同時生起確率を、与えられたモデルパラメータの持つ全ての状態で足し合わせることで、特徴ベクトル系列の生起確率となつて、時刻と状態の変数が無くなり、どの時刻でその生起確率を算出しても同じ値になるという特徴をもつために、数式 3 の分子と分母は評定対象音素区間の特徴ベクトル系列「 $O^{(n)}$ 」内のどの時刻においても同じ値となり、評定値も時刻の違いによる差がなくなる。さらに、前向き後ろ向きアルゴリズムを用いて特徴ベクトル系列の最終時刻「 $T^{(n)}$ 」での確率を求める場合、後ろ向きアルゴリズムは必要とせず、前向きアルゴリズムのみで求めることができる。これらの特徴を利用することで PDAPS アルゴリズムを簡略化して、高速に全く同じ評定値を算出するアルゴリズムが PAPPS である。ただし、PDASP アルゴリズムは評定対象の音素区間内では同じ値であるが時刻（フレーム）ごとに評定値を算出するが、PAPPS アルゴリズムでは評定対象の音素区間ごとに評定値を得る。

10

【 0 1 1 4 】

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。本発音評定装置は、入力受付部 101、教師データ格納部 102、評定対象音素系列格納部 103、音声受付部 104、フレーム音声データ取得部 105、特徴ベクトル系列取得部 106、最適状態系列取得部 107、最適音素部分系列取得部 108、特徴ベクトル部分系列取得部 109、評定値算出部 610、出力部 111、最適音素代表系列取得部 601 を具備する。

20

【 0 1 1 5 】

最適音素代表系列取得部 601 は、最適音素部分系列取得部 108 が取得した 1 組以上の各最適音素部分系列に対して、音素を代表する一の代表音素を取得し、当該取得した 1 以上の代表音素を有する最適音素代表系列（ $P = \{ P_1, P_2, \dots, P_n, \dots, P_s \}$ ）を取得し、当該最適音素代表系列をメモリ上に配置する。最適音素代表系列取得部 601 は、例えば、3 つの最適音素部分系列 $\{ \{ a, a, \dots, a \} \{ o, o, \dots, o \} \{ i, i, \dots, i \} \}$ の各最適音素部分系列に対して、 $\{ a, o, i \}$ という 3 つの代表音素を有する最適音素代表系列を取得する。かかる処理は、音素の ID 列や音素を示す文字コード列が存在する場合の、公知の文字列処理である。最適音素代表系列取得部 601 は、通常、MPU やメモリ等から実現され得る。最適音素代表系列取得部 601 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。

30

【 0 1 1 6 】

評定値算出部 610 は、音素の評定値を音素区間データから算出する。評定値算出部 610 は、音素区間音声データが評定対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評定値を算出する。評定値算出部 610 は、例えば、以下の数式 5、数式 6 により、最適音素代表系列取得部が取得した代表音素「 P_n 」ごとに、その音素に対応する特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列「 $O^{(n)}$ 」と、教師データ格納部から全ての音響モデルを読み出し、全ての音響モデルの全状態において、その状態「 j 」が属する音響モデルパラメータ「 $P^{(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の最終時刻（フレーム）「 $T^{(n)}$ 」における状態が「 j 」である確率を取得し、その取得した状態ごとの確率値から、全ての音素について音響モデルパラメータ「 $P^{(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測される確率を取得し、その取得した音素ごとの確率値から、特徴ベクトル系列「 $O^{(n)}$ 」が与えられたときに、そのベクトル系列が代表音素「 P_n 」である事後確率を取得することで、音声の評定値（PAPPS（ P_n ））を算出する。評定値算出部 610 は、状態ごとの確率値を、バックワード

40

50

アルゴリズムを用いずにフォワードアルゴリズムにより取得することは好適である。

【数 5】

$$\begin{aligned}
 \text{PAPPS}(p_n) &= \Pr(p_n | O^{(n)}) \\
 &= \frac{\Pr(O^{(n)} | \lambda^{(p_n)}) \Pr(p_n)}{\sum_{m=1}^M \Pr(O^{(n)} | \lambda^{(m)}) \Pr(m)} \\
 &= \frac{\sum_{j \in p_n} \Pr(q_{T^{(n)}} = j, O^{(n)} | \lambda^{P(j)})}{\sum_{m=1}^M \sum_{i \in p_m} \Pr(q_{T^{(m)}} = i, O^{(n)} | \lambda^{P(i)})} \\
 &= \frac{\sum_{j \in p_n} \Pr(q_{T^{(n)}} = j, O^{(n)} | \lambda^{P(j)})}{\sum_{i=1}^N \Pr(q_{T^{(n)}} = i, O^{(n)} | \lambda^{P(i)})}
 \end{aligned}
 \tag{10}$$

【数 6】

$$\Pr(q_{T^{(n)}} = j, O^{(n)} | \lambda^{P(j)}) = \Pr(q_{T^{(n)}} = j, \{o_1, o_2, \dots, o_{T^{(n)}}\}^{(n)} | \lambda^{P(j)}) = \alpha_{T^{(n)}}(j)
 \tag{20}$$

【0117】

数式 5 において、「 P_n 」は最適音素代表系列の「 n 」番目の音素である。また、数式 5 において、「 $\lambda^{(P_n)}$ 」は音素「 P_n 」の音響モデルパラメータであり、「 $\lambda^{P(j)}$ 」は状態「 j 」を含む音響モデルのパラメータである。状態「 j 」が音素「 P_n 」の音響モデルの持つ状態であるならば、「 $\lambda^{P(j)}$ 」と「 $\lambda^{(P_n)}$ 」は同じ音響モデルのパラメータを表す。また、「 M 」は全ての音響モデル数であり、「 N 」は音響モデル全体の総状態数である。また、数式 5 において、「 $\Pr(m)$ 」は音素「 m 」の事前確率を表し、音響モデルパラメータ「 $\lambda^{(m)}$ 」の事前確率「 $\Pr(\lambda^{(m)})$ 」を意味する（「 $\Pr(P_n)$ 」についても同様）。また、数式 5 において、音素の事前確率「 $\Pr(m)$ 」は、すべての音素において一定と仮定し、式の途中で省略している。

30

【0118】

また、数式 6 は、状態「 j 」が属する音響モデルパラメータ「 $\lambda^{(P_n)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の最終時刻（フレーム）「 $T^{(n)}$ 」における状態が「 j 」である確率を、前向き（Forward）アルゴリズムにより算出することを示す。そして、数式 5 で示す PAPPS 評定値算出式は、特徴ベクトル系列の最終時刻「 $T^{(n)}$ 」における数式 3 で示した PDAPS 評定値算出式と一致する。

【0119】

さらに、評定値算出部 610 は、数式 5 により、音素区間ごとに算出される PAPPS 評定値をメモリ上に一時記憶し、当該音素区間ごとの評定値をパラメータにして、文章や単語ごとの評定値を算出しても良い。文章や単語ごとの評定値を算出するために、評定値算出部 610 は、評定対象の文章や単語を構成する音素の PAPPS 評定値から平均値や中央値などを用いた算出方法で求めることは好適である。かかる場合、評定値算出部 610 は、最適音素部分系列取得部 108 が取得した各音素の最適音素部分系列の長さ（フレーム数）をパラメータとして、各音素の値を重み付けして、文章や単語の評定値を算出しても良いし、単純に各音素の評定値の平均値や中央値などの値を文章や単語全体の評定値としても良い。

40

【0120】

評定値算出部 610 は、通常、MPU やメモリ等から実現され得る。評定値算出部 610 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

50

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。

【0121】

次に、発音評定装置の動作について説明する。本発音評定装置の動作において、実施の形態2の発音評定装置と比較して、評定値算出処理およびその前処理が異なる。したがって、図7のフローチャートを用いて、本発音評定装置の評定値算出処理について説明する。図7のフローチャートにおいて、図3、図5のフローチャートと異なるステップについてのみ説明する。

【0122】

（ステップS701）評定値算出部610は、n番目の代表音素、および最適音素部分系列が存在するか否かを判断する。n番目の代表音素等が存在すればステップS702に行き、n番目の代表音素等が存在しなければステップS709に行く。

10

【0123】

（ステップS702）評定値算出部610は、最適音素代表系列のn番目の代表音素を取得し、メモリ上に配置する。

【0124】

（ステップS703）評定値算出部610は、n番目の特徴ベクトル部分系列を取得し、メモリ上に配置する。

【0125】

（ステップS704）評定値算出部610は、特徴ベクトル部分系列の最終時刻「 $T^{(n)}$ 」での、全ての音響モデルの状態ごとの確率を、状態の属する音響モデルパラメータを用いて、前向きアルゴリズムにより算出し、バッファに格納する。なお、評定値算出部610は、上記の確率の算出前に、ステップS303で取得した音響モデルの中から、状態の属する音響モデルパラメータを取得する。

20

【0126】

（ステップS705）評定値算出部610は、m番目の音素の音響モデルに含まれる全状態の時刻「 $T^{(n)}$ 」における確率の総和から、m番目の音素の確率を算出する。

【0127】

（ステップS706）評定値算出部610は、代表音素「 P_n 」の事後確率を算出する。

【0128】

（ステップS707）評定値算出部610は、算出した事後確率をn番目の評定値のバッファに格納する。

30

【0129】

（ステップS708）評定値算出部610は、カウンタnを1、インクリメントし、ステップS701に戻る。

【0130】

（ステップS709）評定値算出部610は、音素毎の評定値から、単語区間、文章区間の評定値を算出し、算出値をメモリ上に配置する。かかる、単語区間、文章区間の評定値の算出方法は、音素毎の評定値の各区間における平均値や中央値を算出する方法等がある。上位処理にリターンする。

40

【0131】

以上、本実施の形態によれば、精度の高い発音評定ができる。具体的には、本実施の形態によれば、音素単位の評定を高い精度で行うことができる。また、本実施の形態によれば、高速に音素単位の評定を高い精度で行うことができる。

【0132】

なお、本実施の形態によれば、評定値算出部610は、数式5、6により、評定値を算出した。しかし、評定値算出部610が評定値を算出するアルゴリズムは、数式5、6によるものとは限らない。例えば、数式6は前向き（Forward）アルゴリズムにより算出された。評定値算出部610は、数式6に代えて、後向き（Backward）アルゴリズムにより算出しても良い。また、例えば、数式5の事後確率の算出において、音素としての、起

50

りうる全ての事象が母音の音素を表すもののみになっても良い。かかる場合、評価対象の音素は、いずれかの母音を表すもののみが許されることは明らかである。評価値算出部 610 は、音素区間音声データが評価対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評価値を算出すれば良い。

【0133】

また、本実施の形態において、発音評価装置が、音素毎の評価値を基にして、文章や単語の評価値を算出することは好適である。具体的には、発音評価装置は、文章や単語に含まれる音素毎の複数の評価値の平均値や中央値から評価値を算出することは好適である。

【0134】

なお、本実施の形態における発音評価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、コンピュータを、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、フレームに区分し、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部と、前記1以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、格納されている各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得する特徴ベクトル部分系列取得部と、格納されている教師データを読み出し、当該教師データを用いて、前記特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列が評価対象の音素である事後確率を算出し、当該事後確率から音声の評価値を算出する評価値算出部と、前記評価値算出部が算出した評価値を出力する出力部として機能させるためのプログラム、である。

【0135】

また、上記プログラムにおいて、コンピュータを格納されている評価対象音素系列に沿った音響モデルを読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の組である最適音素部分系列を1組以上取得する最適音素部分系列取得部として、さらに機能させ、前記特徴ベクトル部分系列取得部は、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得するように機能させるためのプログラム、であることは好適である。

【0136】

また、上記プログラムにおいて、前記評価値算出部が、音素区間音声データが評価対象の正解となる音素である事後確率を算出し、当該事後確率から音素の評価値を算出するように機能させるためのプログラム、であることは好適である。

【0137】

また、上記プログラムにおいて、評価値算出部は、前記事後確率を、バックワードアルゴリズムを用いずにフォワードアルゴリズムにより取得するように機能させるためのプログラム、であることは好適である。

(実施の形態4)

【0138】

本実施の形態において、音素ごとに求めたPAPPS評価値をその評価対象音素の時間長(フレーム数)で正規化するアルゴリズムにより、文章や単語の発音評価を行う発音評価装置について説明する。本アルゴリズムを、実施の形態3におけるPAPPSに修飾子をつけてPAPPS-FN(Frame Normalized Phonemic A Posteriori Probability based pronunciation Scoring/フレーム正規化した音韻の事後確率に基づく発音評価)と呼ぶこととする。

【0139】

PAPPS-FNアルゴリズムにおいて音素の評価値を得ることは、実施の形態3のPAPPSアル

10

20

30

40

50

ゴリズムと全く同様である。

【0140】

図8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、評定対象音素系列格納部103、音声受付部104、フレーム音声データ取得部105、特徴ベクトル系列取得部106、最適状態系列取得部107、最適音素部分系列取得部108、特徴ベクトル部分系列取得部109、評定値算出部810、出力部111、最適音素代表系列取得部601、音素時間情報取得部801を具備する。

【0141】

音素時間情報取得部801は、各最適音素部分系列の時間に関する情報である音素時間情報を取得する。音素時間情報は、例えば、単語や文章などを構成する音素のフレーム数や、時間（通常、フレーム数から算出される）である。音素時間情報は、具体的には、最適音素部分系列または特徴ベクトル部分系列の各部分系列を構成する要素数である。音素時間情報取得部801は、通常、MPUやメモリ等から実現され得る。音素時間情報取得部801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。

【0142】

評定値算出部810は、例えば、先に示した数式5、数式6により、最適音素代表系列取得部が取得した代表音素「 P_n 」ごとに、その音素に対応する特徴ベクトル部分系列取得部が取得した特徴ベクトル部分系列「 $O^{(n)}$ 」と、教師データ格納部から全ての音響モデルを読み出し、全ての音響モデルの全状態において、その状態「 j 」が属する音響モデルパラメータ「 $P^{(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測され、かつそのベクトル系列の最終時刻（フレーム）「 $T^{(n)}$ 」における状態が「 j 」である確率を取得し、その取得した状態ごとの確率値から、全ての音素について音響モデルパラメータ「 $P^{(j)}$ 」が与えられたという条件のもと、特徴ベクトル系列「 $O^{(n)}$ 」が観測される確率を取得し、その取得した音素ごとの確率値から、特徴ベクトル系列「 $O^{(n)}$ 」が与えられたときに、そのベクトル系列が代表音素「 P_n 」である事後確率を取得し、当該1以上取得した事後確率と、各音素に対応する音素時間情報をパラメータとして、例えば、以下の数式7により、音声の評定値（PAPPS-FN(P)）を算出する。

【0143】

評定値算出部810は、上述した数式5、6により、音素の評定値を得る。次に、評定値算出部810は、文章や単語ごとの評定値を、以下の数式7により算出する。数式7は、「音素時間長を重みとした音素評定値の加重平均」を意味する。なお、数式7において、「 p 」は文章や単語を構成する音素系列である。また、「 P_n 」は音素である。数式7では、評定対象となる全音声データを1つの文章や、1つの単語であると仮定して、評定対象音素系列の全音素を用いて、評定値を算出する。音声データから複数の文章や複数の単語の評定値を算出する場合は、その文章や単語を構成する音素系列ごとに算出式を適用することにより、複数の文章や複数の単語の評定値を得ることができる。

【数7】

$$PAPPS-FN(p) = \frac{\sum_{n=1}^S PAPPS(p_n) \times T^{(n)}}{\sum_{n=1}^S T^{(n)}}$$

【0144】

数式7において、「 s 」は、最適音素部分系列の個数（音素の個数）である。また、「 $T^{(n)}$ 」は、「 n 」番目部分系列の系列長である。

【0145】

10

20

30

40

50

評定値算出部 810 は、通常、MPU やメモリ等から実現され得る。評定値算出部 810 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。

【0146】

次に、発音評定装置の動作について図 9 のフローチャートを用いて説明する。図 9 のフローチャートの先頭のステップ（ステップ S901）は、図 7 のフローチャートのステップ S709 に後続する。つまり、図 9 のフローチャートは、図 7 の全ステップを含む。

【0147】

（ステップ S901）評定値算出部 810 は、カウンタ n に 1 を代入する。

【0148】

（ステップ S902）評定値算出部 810 は、n 番目の代表音素、および最適音素部分系列が存在するか否かを判断する。n 番目の代表音素等が存在すればステップ S903 に行き、n 番目の代表音素等が存在しなければステップ S908 に行く。

【0149】

（ステップ S903）評定値算出部 810 は、n 番目の最適音素部分系列長を取得し、メモリ上に配置する。なお、最適音素部分系列長に代えて、特徴ベクトルの部分系列長でも良い。

【0150】

（ステップ S904）評定値算出部 810 は、n 番目の音素評定値を取得する。

【0151】

（ステップ S905）評定値算出部 810 は、ステップ S904 で取得した n 番目の音素評定値に、ステップ S903 で取得した n 番目の最適音素部分系列長を乗算し、その値を分子値に加算する。

【0152】

（ステップ S906）評定値算出部 810 は、n 番目の最適音素部分系列長を分母値に加算する。

【0153】

（ステップ S907）評定値算出部 810 は、カウンタ n を 1、インクリメントし、ステップ S902 に戻る。

【0154】

（ステップ S908）評定値算出部 810 は、「分子値 / 分母値」により、評定値を算出する。上位処理にリターンする。

【0155】

以上、本実施の形態によれば、単語や文章などの、ユーザが評定して欲しい評定対象の単位に好適な評価結果を精度高く得ることができる。

【0156】

なお、本実施の形態によれば、評定値算出部 810 は、数式 7 により、評定値を算出した。しかし、評定値算出部 810 が評定値を算出するアルゴリズムは、数式 7 によるものとは限らない。評定値算出部 810 は、取得した 1 以上の各音素の評定値と、各音素に対応する音素時間情報をパラメータとして音声の評定値を算出すればよい。

【0157】

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、コンピュータを、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、フレームに区分し、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部と、前記 1 以上のフレーム音声データから、フレーム毎の特徴ベクトルの集合である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取得部が取得した特徴ベクトル系列から、格納されている各最適音素部分系列に対応する 1 以上の特徴ベクトルの組である特徴ベクトル部分系列を 1 組以上取得する特徴ベクトル部分系列取得部と、格納されている教師データを読み出し、当該教師データを用いて、前記特徴ベクトル

10

20

30

40

50

ル部分系列取得部が取得した特徴ベクトル部分系列が評定対象の音素である事後確率を算出し、当該事後確率から音声の評定値を算出する評定値算出部と、前記評定値算出部が算出した評定値を出力する出力部として機能させるためのプログラム、である。

【0158】

また、上記プログラムにおいて、コンピュータを、格納されている評定対象音素系列に沿った音響モデルを読み出し、当該音響モデルと、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列であり、前記音響モデルに対応する特徴ベクトル系列とを比較し、フレーム毎の最適状態の集合である最適状態系列を取得する最適状態系列取得部と、前記最適状態系列取得部が取得した最適状態系列の中で、同一の音素が連続する1以上の最適音素系列を識別し、当該1以上の最適音素系列の組である最適音素部分系列を1組以上取得する最適音素部分系列取得部として、さらに機能させ、前記特徴ベクトル部分系列取得部は、前記特徴ベクトル系列取得部が取得した特徴ベクトル系列から、前記最適音素部分系列取得部が取得した各最適音素部分系列に対応する1以上の特徴ベクトルの組である特徴ベクトル部分系列を1組以上取得するように機能させるためのプログラム、であることは好適である。

10

【0159】

また、上記プログラムにおいて、コンピュータを、前記評定値算出部を、フレーム毎または音素毎の前記評定値を基にして、文章や単語の評定値を算出するように機能させるためのプログラム、であることは好適である。

【0160】

また、上記プログラムにおいて、コンピュータを、各音素の時間に関する情報である音素時間情報を取得する音素時間情報取得部としてさらに機能させ、前記評定値算出部が、音素時間長を重みとした音素の評定値の加重平均を算出し、当該加重平均から文章または単語の評定値を算出するように機能させるためのプログラム、であることは好適である。

20

【0161】

以下に、上記の実施の形態にかかる発音評定装置について、実験を行った結果について述べる。なお、下記の実験を行ったのは、PDAPS-PE、PDAPS、PAPPSの3つのアルゴリズムを実装する発音評定装置である。

【0162】

(実験1)

まず、日本人女性6人、日本人男性5人、英語ネイティブ男性2人に英語の文章および単語を発音してもらい、その音声を音声受付部104が受け付け、各アルゴリズムを実装する発音評定装置が発音評定を行った。そのアルゴリズムは、t-p-DAP(全時間区間の平均値を用いたもの)、t-p-DAP-PM(音素時間区間での中央値を音素数で平均したもの)、PDAPS-PE(全時間区間の平均値を用いたもの)、PDAPS(全時間区間の平均値を用いたもの)、PAPPS(音素評定値の平均値)のアルゴリズムである。そして、発音評定装置は、各アルゴリズムで求めた文章の評定値および単語の評定値の平均値を算出し、出力する。図10は、各アルゴリズムの比較結果を示す表の集合である。なお、t-p-DAPは、特許文献1で説明されている公知のアルゴリズムである。t-p-DAP-PMは、t-p-DAPのアルゴリズムでフレームごとに出力される評定値の音素時間区間での中央値を音素数で平均したものである。なお、「PM」は「Phoneme Median」を意味する。なお、PAPPS-FNアルゴリズムに関しては出力する評定値がPDAPSアルゴリズムを用いた文章・単語の評定値と全く一致するので図10等の比較対照に含めなかった。

30

【0163】

図10(a)は、全話者での平均値を示す。図10(b)は、Native話者での平均値を示す。図10(c)は、日本人話者での平均値を示す。図11(a)は、タスク1での平均値を示す。図11(b)は、タスク2での平均値を示す。図10から図13において、「%」で示される値は、比較されるアルゴリズムで算出した評定値の値の割合を示す。

40

【0164】

図10、図11から、全てにおいて、t-p-DAP-PM < t-p-DAP < PDAPS-PE < P

50

APPS < PDAPSとなる結果を得た。なお、タスクとは話者が発音している文章や単語の種類を表す。つまり、タスク1とタスク2は、異なる文章や単語である。

【0165】

そして、図10、図11の結果の中では特にNative話者での平均値に関する大小関係が重要となる。本来それを教師としているので、Native話者の発音に対して、発音評定装置は満点を出力することが望ましい。しかし現実的には発音評定装置が教師としているモデル話者と、評定対象の話者が異なる場合は、話者性の違いなどで常に満点を出力することは難しい。さらにそのモデル話者の発音を録音した環境(場所やマイク)と評定対象の発音の録音環境との違いによっても常に満点を出力することは難しくなる。ただし同じモデル話者を用いて同じ評定対象音声を評定する場合、その違いはアルゴリズムによるもののみとなり、この場合Native話者の発音に対してはより高得点を出力できるアルゴリズムの方が、より望ましいといえる。このような理由から、特許文献1で説明しているDAP、p-DAPアルゴリズムよりも、今回発明したPDAPS-PE、PDAPS、PAPPS、PAPPS-FNアルゴリズムの方がより良い発音評定アルゴリズムであるといえる。

10

【0166】

(実験2)

実験1においては、Native話者に対する評定値の比較結果から、PDAPS-PE、PDAPS、PAPPS、PAPPS-FNアルゴリズムが良好な値を出力することが明らかとなった。しかしその実験では、Native話者の数が2人であることから、十分なデータ量であったとはいえない。そこで続く実験として、英語の文章および単語を発音した英語ネイティブ多数話者の音声データベースを用いて、t-p-DAP、PDAPS-PE、PDAPS、PAPPSアルゴリズムで求めた文章および単語評定値の平均値の比較を行った。なお、この音声データベースは、男女合わせて221人が総発音データ数193427の音声を入力し、音声受付部104がかかる音声データを受け付け、音声データベース(記憶媒体)に蓄積したものである。そして、発音評定装置は、音声データベースの音声データを読み出し、発音評定を行う。かかる実験結果を、図12に示す。

20

【0167】

図12によれば、全ての発音の平均評定値で、t-p-DAP < PDAPS-PE < PAPPS < PDAPSとなる結果を得ることができ、PDAPS-PE、PDAPS、PAPPS、PAPPS-FNアルゴリズムが、従来のp-DAPアルゴリズムよりもNative話者に対する評定値では、良好な結果を出力することが確認できた。

30

【0168】

また、実験2では、全ての発音データを評定するのに要する処理時間を計測し、t-p-DAP > PDAPS-PE > PDAPS > PAPPSとなる結果を得ることができた。

【0169】

(実験3)

実験1ならびに実験2で確認してきたことは、正しい発音に対して、より高得点を出力できるかということであった。しかし、それだけでは良好な発音評定アルゴリズムとしての性質を完全に満たすことはできていない。間違った発音に対しても、例えば、全く言語として理解不能な内容の発音であろうとも、高得点を出力してしまう可能性があるからである。

40

【0170】

そこで、実験3において、日本人多数話者が、完全に正しいとはいきれない英語の文章および単語を発音し、発音評定装置が当該音声を受け付け、発音評定を行った。そして、t-p-DAP、PDAPS-PE、PDAPS、PAPPSアルゴリズムで求めた文章および単語評定値の平均値の比較を行ったものが図13(a)である。

【0171】

また、図13(b)に、ヒューマンレーティングスコアと発音評定アルゴリズムにより算出するスコアとの相関係数を示す。相関係数を調べることで、人間の感覚との近さを計測し、相関が高いものほどより良好な評定値を出力するアルゴリズムであるということ

50

確認できる。なお、ヒューマンレーティングスコアとは、人手により採点した発音評定値である。

【0172】

実験3の結果、平均値の関係からは、実験1、2と同じく新たなアルゴリズム(PDAPS-PE、PDAPS、PAPPS)の方が高得点となった。ヒューマンレーティングスコアとの相関係数(図13(b))を見ると、いままで利用してきたp-DAPアルゴリズムよりも新たなアルゴリズムの方が高い相関を示す結果となり、より良い発音評定アルゴリズムであることが確認できた。さらに、相関係数の新たなアルゴリズム間での関係において、平均値の関係とは違う関係を見ることができた。平均値ではPDAPSアルゴリズムが最も高い値を出力していたが、相関係数ではPAPPSアルゴリズムが最も高い値を得ている。このことは、PAPPSアルゴリズムが、間違った(完全に正しくない)発音に対しても、より正確に評定値を算出していることを示し、人間の感覚にも近い評定値を算出できるアルゴリズムであるといえる。

10

【0173】

以上の実験結果から、本明細書で説明した発音評定アルゴリズムPDAPS-PE、PDAPS、PAPPS、PAPPS-FNが、公知のp-DAPアルゴリズムよりも優れていることを示した。

【0174】

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。

20

【0175】

また、図14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14は、このコンピュータシステム340の概観図であり、図15は、コンピュータシステム340のブロック図である。

【0176】

図14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。

30

【0177】

図15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、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。

40

【0178】

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。

50

【 0 1 7 9 】

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（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 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。

【 0 1 8 0 】

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。

【 0 1 8 1 】

また、上記各実施の形態において、一の装置に存在する 2 以上の通信手段（情報送信部など）は、物理的に一の媒体で実現されても良いことは言うまでもない。

【 0 1 8 2 】

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】

【 0 1 8 3 】

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

【図面の簡単な説明】

【 0 1 8 4 】

【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】同評定値算出処理について説明するフローチャート

【図 4】実施の形態 2 における発音評定装置のブロック図

【図 5】同評定値算出処理の動作について説明するフローチャート

【図 6】実施の形態 3 における発音評定装置のブロック図

【図 7】同評定値算出処理の動作について説明するフローチャート

【図 8】実施の形態 4 における発音評定装置のブロック図

【図 9】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 10】同実験結果の表を示す図

【図 11】同実験結果の表を示す図

【図 12】同実験結果の表を示す図

【図 13】同実験結果の表を示す図

【図 14】同発音評定装置を実現するコンピュータの外観図

【図 15】同発音評定装置を実現するコンピュータシステムのブロック図

【符号の説明】

【 0 1 8 5 】

1 0 1 入力受付部

1 0 2 教師データ格納部

1 0 3 評定対象音素系列格納部

1 0 4 音声受付部

1 0 5 フレーム音声データ取得部

1 0 6 特徴ベクトル系列取得部

1 0 7 最適状態系列取得部

1 0 8 最適音素部分系列取得部

1 0 9 特徴ベクトル部分系列取得部

1 1 0、4 1 0、6 1 0、8 1 0 評定値算出部

1 1 1 出力部

6 0 1 音素代表系列取得部

10

20

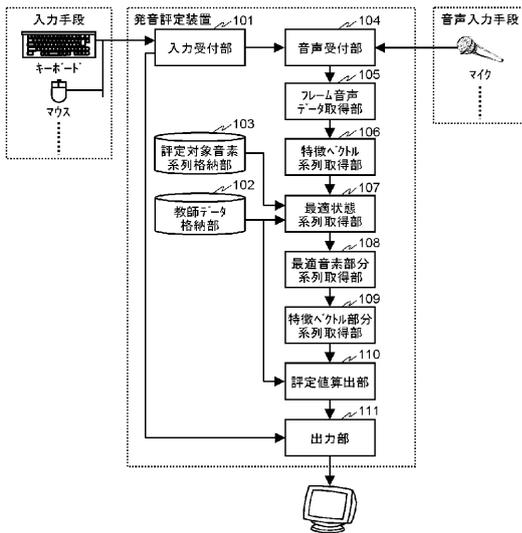
30

40

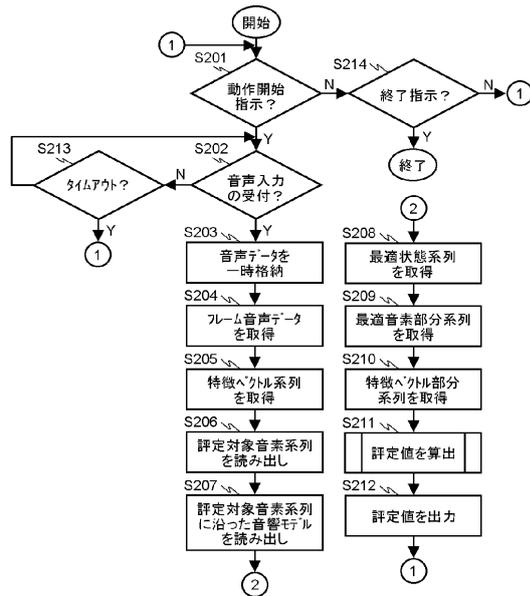
50

8 0 1 音素時間情報取得部

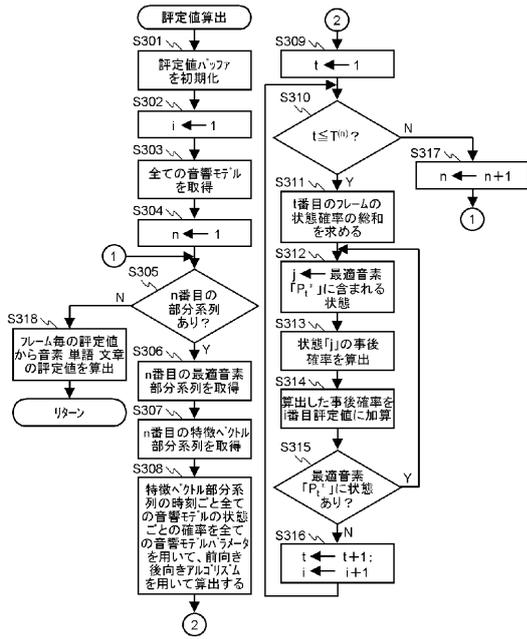
【図 1】



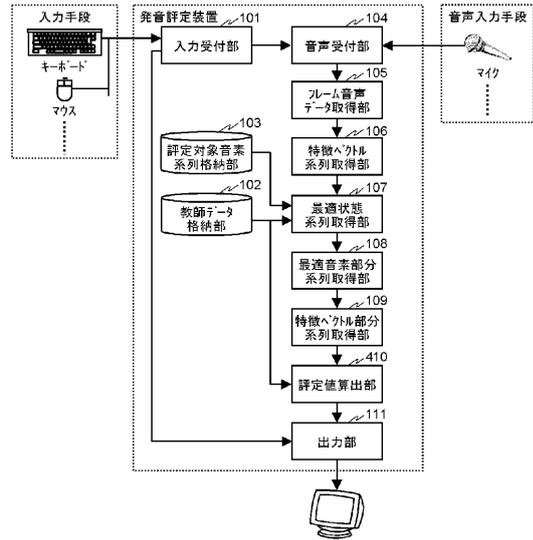
【図 2】



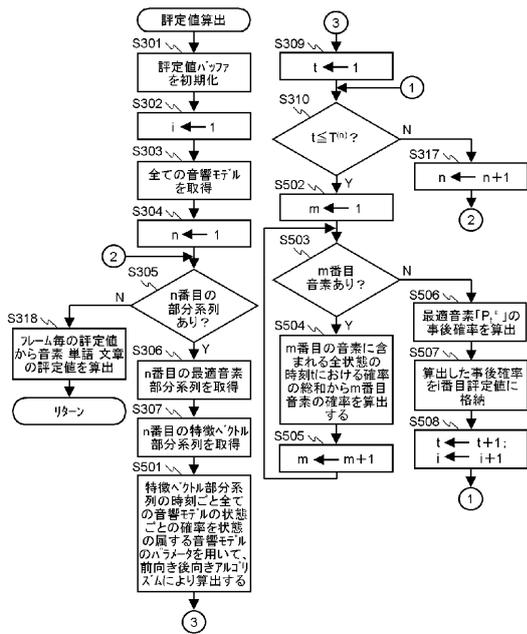
【図3】



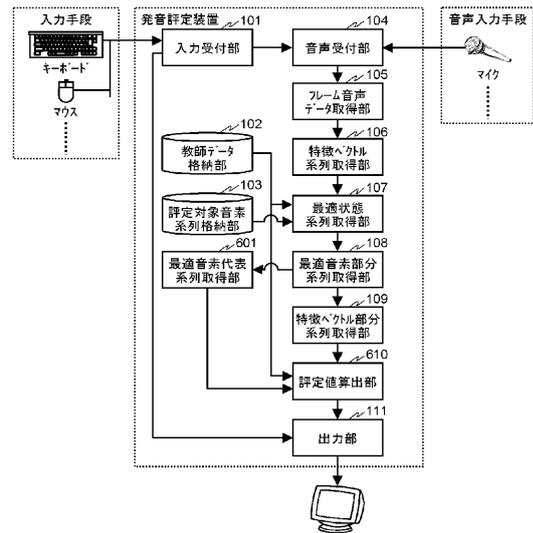
【図4】



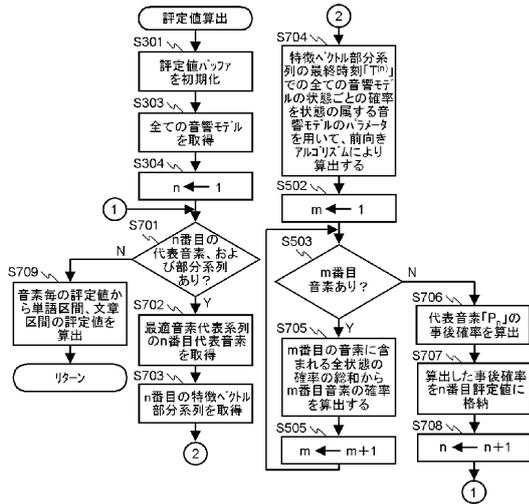
【図5】



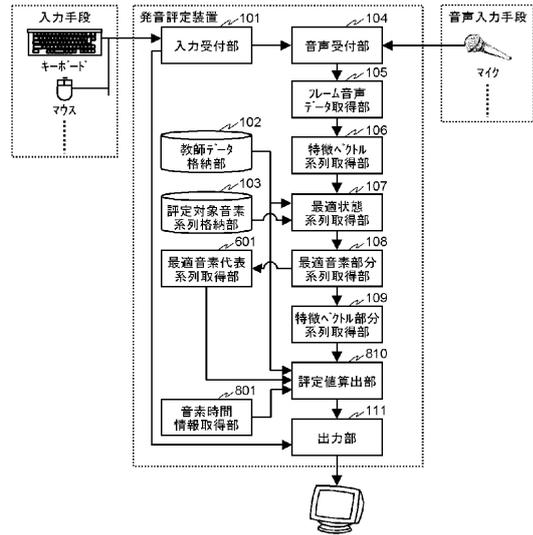
【図6】



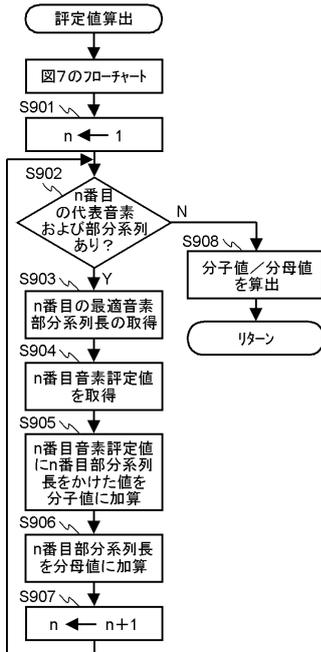
【図7】



【図8】



【図9】



【図10】

	/ t-p-DAP	/ t-p-DAP-PM	/ PDAPS-PE	/ PDAPS	/ PAPPS
t-p-DAP	100.00%	109.51%	93.03%	72.25%	86.57%
t-p-DAP-PM	91.32%	100.00%	84.95%	65.98%	79.05%
PDAPS-PE	107.49%	117.71%	100.00%	77.68%	93.05%
PDAPS	138.41%	151.57%	128.78%	100.00%	119.81%
PAPPS	115.52%	126.51%	107.47%	83.46%	100.00%

(a)

	/ t-p-DAP	/ t-p-DAP-PM	/ PDAPS-PE	/ PDAPS	/ PAPPS
t-p-DAP	100.00%	103.23%	93.23%	75.63%	85.81%
t-p-DAP-PM	96.87%	100.00%	90.31%	73.26%	83.13%
PDAPS-PE	107.26%	110.73%	100.00%	81.12%	92.04%
PDAPS	132.22%	136.49%	123.27%	100.00%	113.46%
PAPPS	116.54%	120.30%	108.64%	88.14%	100.00%

(b)

	/ t-p-DAP	/ t-p-DAP-PM	/ PDAPS-PE	/ PDAPS	/ PAPPS
t-p-DAP	100.00%	111.30%	92.98%	71.41%	86.77%
t-p-DAP-PM	89.85%	100.00%	83.54%	64.16%	77.96%
PDAPS-PE	107.55%	119.70%	100.00%	76.80%	93.32%
PDAPS	140.04%	155.86%	130.20%	100.00%	121.51%
PAPPS	115.25%	128.27%	107.16%	82.30%	100.00%

(c)

【図11】

	/ t-p-DAP	/ t-p-DAP-PM	/ PDAPS-PE	/ PDAPS	/ PAPPS
t-p-DAP	100.00%	111.51%	92.70%	71.92%	88.27%
t-p-DAP-PM	85.68%	100.00%	83.13%	64.50%	79.16%
PDAPS-PE	107.88%	120.30%	100.00%	77.59%	95.23%
PDAPS	138.04%	155.04%	128.88%	100.00%	122.73%
PAPPS	113.28%	126.32%	105.01%	81.48%	100.00%

(a)

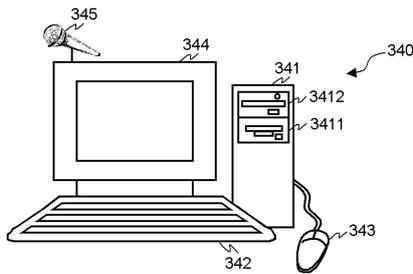
	/ t-p-DAP	/ t-p-DAP-PM	/ PDAPS-PE	/ PDAPS	/ PAPPS
t-p-DAP	100.00%	104.32%	93.97%	73.18%	82.16%
t-p-DAP-PM	95.86%	100.00%	90.08%	70.15%	78.76%
PDAPS-PE	106.42%	111.01%	100.00%	77.88%	87.43%
PDAPS	136.65%	142.55%	128.41%	100.00%	112.27%
PAPPS	121.72%	126.98%	114.38%	89.07%	100.00%

(b)

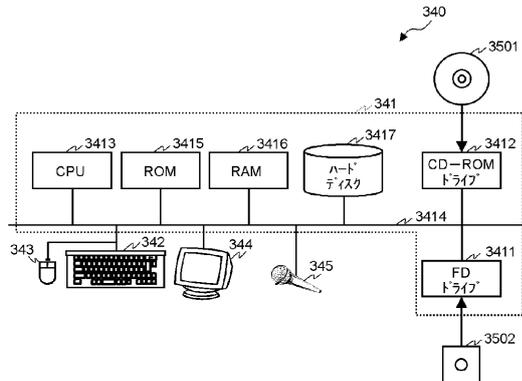
【図12】

	実行時間	平均点
t-p-DAP	13:53:26	0.4637
PDAPS-PE	12:08:19	0.4908
PDAPS-PE/t-p-DAPの割合	87.39%	105.85%
PDAPS	5:34:08	0.6416
PDAPS/t-p-DAPの割合	40.09%	138.35%
PAPPS	4:49:52	0.5849
PAPPS/t-p-DAPの割合	34.78%	126.14%

【図14】



【図15】



【図13】

	全音声データの平均点	話者ごとの平均点の平均	発話内容の種類ごとの平均点の平均
t-p-DAP	0.3305	0.3287	0.3325
PDAPS-PE	0.3581	0.3541	0.3504
PDAPS-PE/t-p-DAPの割合	107.73%	107.72%	107.76%
PDAPS	0.5283	0.4705	0.5284
PDAPS/t-p-DAPの割合	159.23%	143.05%	158.91%
PAPPS	0.4563	0.4533	0.4571
PAPPS/t-p-DAPの割合	137.78%	137.90%	137.45%

(a)

	音声データ単位	話者ごとの平均点単位	発話内容の種類ごとの平均点単位
t-p-DAP	0.2632	0.5161	0.3834
PDAPS-PE	0.2841	0.5175	0.3719
PDAPS-PE/t-p-DAPの割合	109.35%	100.27%	94.53%
PDAPS	0.2374	0.5246	0.3638
PDAPS/t-p-DAPの割合	90.22%	101.65%	92.42%
PAPPS	0.2996	0.5454	0.4150
PAPPS/t-p-DAPの割合	113.83%	105.65%	105.49%

(b)

フロントページの続き

- (72)発明者 久保 理恵子
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
- (72)発明者 駒木 亮
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 生馬 裕子
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
- (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2006-227587(JP,A)
特開2006-227030(JP,A)
特開2006-208644(JP,A)
特開2001-265211(JP,A)
渡辺秀行他, "動的事後確率を用いた発音評定システム", 日本音響学会2005年秋季研究発表会講演論文集CD-ROM, 3-8-15(2005-09), pp.871-872
- (58)調査した分野(Int.Cl., DB名)
G10L 11/00, 15/00 - 15/28