

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4517163号
(P4517163)

(45) 発行日 平成22年8月4日(2010.8.4)

(24) 登録日 平成22年5月28日(2010.5.28)

(51) Int.Cl. F I
G 1 0 L 15/02 (2006.01) G 1 0 L 15/02 2 0 0 B
G 1 0 L 15/06 (2006.01) G 1 0 L 15/06 3 1 0 Z

請求項の数 4 (全 13 頁)

<p>(21) 出願番号 特願2004-71058 (P2004-71058)</p> <p>(22) 出願日 平成16年3月12日 (2004. 3. 12)</p> <p>(65) 公開番号 特開2005-258186 (P2005-258186A)</p> <p>(43) 公開日 平成17年9月22日 (2005. 9. 22)</p> <p>審査請求日 平成19年2月26日 (2007. 2. 26)</p> <p>(出願人による申告) 平成15年9月30日付け「第年5E1 8号」平成15年度、支出負担行為担当官 総務省大臣官房会計課企画官、研究テーマ「携帯電話等を用いた多言語自動翻訳システム」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p>	<p>(73) 特許権者 393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 黒岩 真吾 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 遠藤 俊樹 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 中村 哲 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p style="text-align: right;">最終頁に続く</p>
--	--

(54) 【発明の名称】 周波数特性等化装置

(57) 【特許請求の範囲】

【請求項1】

入力音声信号から抽出された音響特徴量からなる特徴ベクトルに対し、当該入力音声信号の入力系の周波数特性の等化处理を行なう周波数特性等化装置であって、前記周波数特性等化装置は、前記特徴ベクトルに対し、予め準備されたベクトル量子化(VQ)コードブックを使用してベクトル量子化を行なうベクトル量子化装置に接続して用いられるものであり、

前記周波数特性等化装置は、

入力される特徴ベクトルから減算されるべきバイアスを記憶するためのバイアス記憶手段と、

前記バイアス記憶手段に記憶されたバイアスを、入力される前記特徴ベクトルから減算してベクトル量子化のために前記ベクトル量子化装置に与えるためのバイアス減算手段と

、
 予め定める基準により決定され、かつ前記VQコードブックを用いてVQコード化された複数のリファレンスベクトルの各々と、前記バイアス減算手段の出力ベクトルとの間で、前記ベクトル量子化装置において算出された距離を記憶するための距離記憶手段と、

前記複数のリファレンスベクトルのうち、前記バイアス減算手段の出力ベクトルと最も近いものを前記距離記憶手段に記憶された距離に基づいて選択し、前記バイアス減算手段の出力ベクトルと前記出力ベクトルに最も近いリファレンスベクトルに基づいて、予め定める算出式に従って前記バイアス記憶手段に記憶された前記バイアスを更新するためのバ

イアス更新手段とを含む、周波数特性等化装置。

【請求項 2】

前記距離記憶手段は、

VQコード化された前記複数のリファレンスベクトルを記憶するためのリファレンスベクトル記憶手段と、

前記VQコードブックに記憶されたセントロイドの各々と、前記バイアス減算手段の出力ベクトルとについて、前記ベクトル量子化装置において算出された距離を受け、当該距離のうちで、前記リファレンスベクトル記憶手段に記憶された前記複数のリファレンスベクトルに対応するものを記憶するための手段とを含む、請求項 1 に記載の周波数特性等化装置。

10

【請求項 3】

前記バイアス更新手段は、

前記複数のリファレンスベクトルの各々のうち、前記バイアス減算手段の出力ベクトルとの間の距離が最も小さなものを前記距離記憶手段に記憶された距離に基づいて選択し、前記バイアス減算手段の出力ベクトルと前記選択されたリファレンスベクトルとに基づき、前記バイアス減算手段の出力を前記選択されたリファレンスベクトルに近づけるためのバイアスを算出するためのバイアス算出手段と、

前記バイアス算出手段により算出されたバイアスを用いて、前記バイアス記憶手段に記憶された前記バイアスを更新するための更新手段とを含む、請求項 1 又は請求項 2 に記載の周波数特性等化装置。

20

【請求項 4】

前記複数のリファレンスベクトルは、所定の学習データに基づき得られた音素ごとの特徴パラメータの平均ベクトルであり、かつ前記VQコードブックを用いてVQコード化したものとして、音素ごとに予め選択される、請求項 1 ~ 請求項 3 のいずれかに記載の周波数特性等化装置。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は分散型音声認識のための端末装置に関し、特に、端末装置における入力系の周波数特性の等化技術の改善に関する。

30

【背景技術】

【0002】

近年、端末装置での音声認識として音響分析のみを端末装置で行ない、デコード等の処理をサーバで行なう分散型音声認識(DSR)が提案されている。そのフロントエンドは、欧州電気通信標準化機構(ETSI)により標準化されつつある。標準化の一環として、ETSIは2000年4月に標準DSRフロントエンド(ETSI ES201)、2002年10月に雑音に頑健なAdvanced DSRフロントエンド(ETSI ES202)(非特許文献1を参照されたい。)を、それぞれ勧告した。

【0003】

図1に、従来の分散型音声認識システム20のブロック図を示す。図1を参照して、従来の分散型音声認識システム20は、話者の発話30を受け、アナログ音声電気信号に変換するためのマイクロホン32と、マイクロホン32からのアナログ音声電気信号に対する音響分析を行ない、得られたパラメータをベクトル量子化(VQ)してVQコード列40として回線に送出するための端末装置34と、端末装置34からのVQコード列40を受け、これに対する音声認識処理をし、認識結果に対して所定の処理、例えば認識結果のテキスト列38を端末装置34に対して返送する処理を行なうためのサーバ36とを含む。

40

【0004】

端末装置34は、マイクロホン32からのアナログ音声電気信号をデジタル信号に変換するためのA/D変換機50と、A/D変換機50の出力に対して音響分析を行ない、特

50

微量として12次元MFCC (Mel Frequency Cepstrum Coefficient) 60を出力するための音響分析部52と、音響分析部52から出力されるMFCCに対し、発話環境、マイクロホン32、及びA/D変換機50の部分の周波数特性を補償し等化後のMFCC62として出力するための等化部54と、VQ処理のために予め準備されたVQコードブック58と、VQコードブック58を用いて、等化部54からの等化後のMFCC62に対するベクトル量子化を行ない、圧縮してVQコード列40として回線路上に送出するための圧縮部56とを含む。

【0005】

サーバ36は、VQコード列40を回線から受信し伸長処理を行なってMFCCを復元するための伸長部70と、伸長部70から与えられるMFCCを音響特徴量として音声認識処理を行ない、結果のテキスト列38を端末装置34に対して返送するための音声認識処理部72とを含む。もちろんサーバ36の構成はこのようなものに限定されるわけではなく、用途に応じて種々の構成があり得る。

10

【0006】

上記したES202では、音響分析部52から出力されるMFCCは、各々浮動小数点で表現される1次～12次MFCCと、パワー及びケプストラム0次とからなる特徴量ベクトルである。これらをそのままサーバ36に対して送信しようとする、データ量が多いため必要とされる通信容量が大きくなるという問題がある。そこで、ES202では、VQコードブック58を用いたベクトル量子化をすることにより、この特徴量ベクトルを44ビットで表し、送信されるデータ量を削減している。

20

【0007】

図2にベクトル量子化で使用されるVQコードブック58の概要を示す。ES202のベクトル量子化では、1次～12次MFCCとパワー及びケプストラム0次からなる14個の要素を二つずつ(1、2次、3、4次、5、6次、7、8次、9、10次、及び11、12次MFCC、並びにパワー及びケプストラム0次)の7組に分ける。そして各組を二次元ベクトルとみなして量子化する。これらは、7種類の二次元ベクトル空間を想定すると理解しやすい。ES202では、これら7種類の二次元ベクトル空間の各々に対して、64個(MFCCに対するもの)又は256個(パワー及びケプストラムに対するもの)のVQセントロイドと呼ばれるベクトルを定め、全ての二次元MFCCベクトルを、これらVQセントロイドのうちで最も近いものに割当てられた番号で表現する。セントロイドの個数は64($=2^6$)個又は256($=2^8$)個であるから、その番号は図2の最も右の欄にも示すように6ビット又は8ビットで表すことができる。VQセントロイドの概念を図3に示す。

30

【0008】

図3を参照して、例えば1、2次MFCCにより形成される二次元ベクトル空間内に、64個のセントロイド120を定める。この二次元空間における各セントロイドの座標は予め分かっている。これらセントロイドの各々には番号が割り振られている。これを以下「VQセントロイド番号」と呼ぶ。

【0009】

ベクトル量子化では、実際にサンプルから得られた1、2次のMFCCからなるベクトルを図3の1、2次空間上にプロットしたときに、最もこの点に近い位置に存在するセントロイドで、この点を表す。従ってサンプルの1、2次のMFCCは6ビットのセントロイド番号で表されることになる。

40

【0010】

同様に、3、4次の二次元ベクトルにも図3の右側に示すように64個のセントロイド122が定められる。実際にサンプルから得られた3、4次のMFCCもこの二次元空間上で最も近いセントロイド番号により表される。

【0011】

このようなセントロイド番号による置き換えが、上記した7組の二次元ベクトル空間の全てにおいて行なわれる。その結果、サンプルから得られた1次～12次の12個のMF

50

CC、並びにパワー及びケプストラム0次が、7つのセントロイド番号、すなわち6ビット×6+8ビット=44ビットで表される。このようにしてベクトル量子化されたMFCCを音響特徴量としてサーバに送る。サーバでは、ベクトルコード列に対しVQコードブックを逆引きして1次~12次のMFCC並びにパワー及びケプストラム0次を求め、音声認識に用いる。

【0012】

一方、入力系の周波数特性の差異は、DSRにおける大きな問題である。この差異により、DSRにおける音声認識の認識性能が劣化することが知られている。等化部54による等化は、この入力系の周波数特性の差異を補償することを目的としている。入力系の周波数特性は未知なので、等化部54において行なわれる補償は以下に述べる方式に従い行なわれる。

10

【0013】

図4に、従来の等化部54における等化の概念を示す。図4を参照して、例えば1、2次のMFCCからなる二次元のベクトル空間内に64個のセントロイドが定められているものとする。従来の等化処理では、サンプルが現れる領域130をセントロイドが配置されている領域と重ねることを目的として、次のような処理を行っていた。

【0014】

すなわち、64個のセントロイドの中心にリファレンスベクトル132を想定し、サンプルが得られるたびに、そのサンプルのこの1、2次MFCC空間内における位置とリファレンスベクトル132との間のバイアスを計算し、次のサンプルからこのバイアスを減算する。この処理により、サンプルの現れる領域130がセントロイドの配置されている領域と重なるようになり、入力系の周波数特性が補償される。

20

【0015】

実際には、上記した処理は全次元を対象として行なわれる。全体として1次~12次MFCC並びにパワー及びケプストラム0次からなる14次元のベクトルが、同じく14次元の1つのリファレンスベクトルに近づくよう、各サンプルとリファレンスベクトルとの間のバイアスを算出し、次のサンプルからこのバイアスを減算することで周波数特性の補償を行なう。

【0016】

図5に、上記した等化を行なうための等化部54の構成を示す。図5を参照して、等化部54は、上記した考え方により算出されるバイアスを記憶するためのバイアス記憶部158と、入力されるサンプルの12次元MFCCベクトルから、バイアス記憶部158に記憶されているバイアスを減算し、圧縮部56に与えるためのバイアス減算部150と、各々浮動小数点により表現された1次~12次のMFCC並びにパワー及びケプストラム0次を要素とする14次元のリファレンスベクトルを記憶するためのリファレンスベクトル記憶部152と、バイアス減算部150の出力と、リファレンス記憶部152の出力とに基づいてバイアスを算出するためのバイアス計算部154と、バイアス計算部154により算出されたバイアスでバイアス記憶部158に記憶されたバイアスを更新するためのバイアス更新部156とを含む。

30

【0017】

この構成により、上記したように入力されたサンプルとリファレンスベクトルとの間のバイアスを算出し、次のサンプルからこのバイアスを減算することで、サンプルの現れる空間をセントロイドの位置する空間に近づけるような等化処理が実現される。

40

【0018】

なお、圧縮部56は、上記したようにVQコードブック58を用い、等化部54の出力により表される点とセントロイドとの二乗距離を算出し、二乗距離が最小となるセントロイドを選択することにより、MFCCをベクトル量子化するベクトル量子化部170を含む。圧縮部56は、ベクトル量子化部170に加えて回線に信号を送出するための種々の処理部を含むが、それらは本発明に直接は関連しないため、図示していない。

【0019】

50

【非特許文献1】ETSI ES 202 212 V.1.1.1(2003-11),「音声処理、伝送及び品質の局面(STQ);分散音声認識;拡張された先進的フロントエンド特徴抽出アルゴリズム;圧縮アルゴリズム;バックエンド音声再構築アルゴリズム」、インターネット<URL: http://webapp.etsi.org/action?PU/20031118/es_202212v010101p.pdf>、2003年(ETSI ES 202 212 v.1.1.1,“Speech Processing, Transmission and Quality Aspects(STQ);Distributed speech recognition;Extended advanced front-end feature extraction algorithm;Compression algorithms;Back-end speech reconstruction algorithm”,2003.),図4.1及びpp.26、「5.4 Blind equalization」

10

【発明の開示】

【発明が解決しようとする課題】

【0020】

しかし、上記した従来の等化処理では、バイアスの算出には各々浮動小数点で表現された多数の要素からなる二つのベクトルの間での距離計算を行わなければならない。そのため計算量が大きく、実時間で処理するためには端末装置に係る負荷が大きいという問題がある。端末装置で利用できるリソースは少ないため、計算量をできるだけ小さくすることが好ましい。また、上記した等化処理にはより効果を向上させる余地が存在するように

20

思われる。

【0021】

それゆえに本発明の目的は、端末装置の計算量を小さくできる周波数特性等化装置を提供することである。

【0022】

本発明の他の目的は、端末装置の計算量を小さくし、かつ等化の効果をより高くできる周波数特性等化装置を提供することである。

【課題を解決するための手段】

【0023】

本発明に係る周波数特性等化装置は、入力音声信号から抽出された音響特徴量からなる特徴ベクトルに対し、当該入力音声信号の入力系の周波数特性の等化処理を行なう周波数特性等化装置である。この周波数特性等化装置は、特徴ベクトルに対し、予め準備されたVQコードブックを使用してベクトル量子化を行なうベクトル量子化装置に接続して用いられる。周波数特性等化装置は、入力される特徴ベクトルから減算されるべきバイアスを記憶するためのバイアス記憶手段と、バイアス記憶手段に記憶されたバイアスを、入力される特徴ベクトルから減算してベクトル量子化のためにベクトル量子化装置に与えるためのバイアス減算手段と、予め定める基準により決定され、かつVQコードブックを用いてVQコード化された複数のリファレンスベクトルの各々と、バイアス減算手段の出力ベクトルとの間で、ベクトル量子化装置において算出された距離を記憶するための距離記憶手段と、複数のリファレンスベクトルのうち、バイアス減算手段の出力ベクトルと最も近いものを距離記憶手段に記憶された距離に基づいて選択し、バイアス減算手段の出力ベクトルと、出力ベクトルに最も近いリファレンスベクトルとに基づいて、予め定める算出式に従ってバイアス記憶手段に記憶されたバイアスを更新するためのバイアス更新手段とを含む。

30

40

【0024】

好ましくは、距離記憶手段は、VQコード化された複数のリファレンスベクトルを記憶するためのリファレンスベクトル記憶手段と、VQコードブックに記憶されたセントロイドの各々と、バイアス減算手段の出力ベクトルとについてベクトル量子化装置において算出された距離を受け、当該距離のうちで、リファレンスベクトル記憶手段に記憶された複数のリファレンスベクトルに対応するものを記憶するための手段とを含む。

50

【 0 0 2 5 】

さらに好ましくは、バイアス更新手段は、複数のリファレンスベクトルのうち、バイアス減算手段の出力ベクトルとの間の距離が最も小さなものを距離記憶手段に記憶された距離に基づいて選択し、バイアス減算手段の出力ベクトルと選択されたリファレンスベクトルとに基づき、バイアス減算手段の出力を選択されたリファレンスベクトルに近づけるためのバイアスを算出するためのバイアス算出手段と、バイアス算出手段により算出されたバイアスを用いて、バイアス記憶手段に記憶されたバイアスを更新するための更新手段とを含む。

【 0 0 2 6 】

音響特徴量は、MFCC、例えば12次元MFCCを含んでもよい。

10

【 0 0 2 7 】

好ましくは、複数のリファレンスベクトルは、所定の学習データに基づき得られた音素ごとの特徴パラメータの平均ベクトルであり、かつVQコードブックを用いてVQコード化したものとして、音素ごとに予め選択される。

【 発明を実施するための最良の形態 】

【 0 0 2 8 】

図6を参照して、本発明の一実施の形態で採用される複数のリファレンスベクトルについて説明する。なお以下の説明では、説明を理解しやすくするため、特徴量として1次～12次MFCCのみを用いる場合について説明する。もちろん、従来技術でも説明した通り、これらに加えて例えばパワー及びケプストラム0次を加えて処理することも容易である。

20

【 0 0 2 9 】

図6に示されるように、例えば1次、2次MFCCからなるベクトル空間に、従来技術でも説明した通り、64個のセントロイド190が定められている。本実施の形態では、これらセントロイド190の領域内で、音声認識の対象となる言語の音素ごとの平均特徴パラメータに対応した、複数のリファレンスベクトル200、202、204、206、...、208（正しくは、複数のリファレンスベクトルの、この空間の成分）を予めVQコードブック作成データを用いて計算により求めておく。これらリファレンスベクトル200、202、204、206、...、208は、そのままではいずれもデータ量が大きい

30

【 0 0 3 0 】

このような処理をより上位の次元のMFCCベクトル空間に対しても行なうことにより、図7に示すリファレンスベクトル90のような12次元MFCCに対するリファレンスベクトルが得られる。すなわち、図7を参照して、リファレンスベクトル90は、1、2次、3、4次、...、11、12次のMFCCベクトル空間での6個のセントロイド番号100、102、...、110を含む。各セントロイド番号は6ビットで表されるので、リファレンスベクトル90は36ビットの長さを持つ。

40

【 0 0 3 1 】

等化処理では、サンプルの特徴ベクトルを、最も近いリファレンスベクトルに近づけるようにバイアスが計算される。図8には、これを2次元空間で表している。図8に示す例では、サンプル240、242、244、246、...、248等を、セントロイドにより表された、それぞれの点に最も近いリファレンスベクトル220、222、224、226、...、228に近づけるためのバイアスを算出する。実際にはこの処理は12次元のベクトル空間内で行なわれ、その結果バイアスを表す一つの12次元ベクトルが得られる。そうして得られたバイアスを次のサンプルの特徴量ベクトルから減算することで等化処理

50

を行なう。

【 0 0 3 2 】

図 9 に、本実施の形態に係る端末装置 2 6 0 のブロック図を示す。サーバ側の構成は図 1 に示すものと同様である。図 9 を参照して、端末装置 2 6 0 は、図 1 に示したものと同様の A / D 変換機 5 0 及び音響分析部 5 2 と、上記した原理に従い、V Q コードブック 5 8 により表される複数のリファレンスベクトルを用いた等化处理を行なう等化部 2 7 0 と、図 1 に示す圧縮部 5 6 の機能に加えて、ベクトル量子化の際にサンプル点と各セントロイドとの間で算出された二乗距離を等化部 2 7 0 に与える機能を持つ圧縮部 2 7 2 とを含む。

【 0 0 3 3 】

図 1 0 に、等化部 2 7 0 と圧縮部 2 7 2 とのより詳細なブロック図を示す。図 1 0 を参照して、圧縮部 2 7 2 は、図 5 に示すベクトル量子化部 1 7 0 に代えて、ベクトル量子化の際に等化部 5 4 の出力により表される点と各セントロイドとの間で算出された二乗距離を等化部 2 7 0 に対して出力する機能を持つベクトル量子化部 2 9 0 を含む。ベクトル量子化部 2 9 0 の機能はこの点を除き等化部 2 7 0 と全く同様である。

【 0 0 3 4 】

等化部 2 7 0 は、図 5 に示すバイアス減算部 1 5 0 及びバイアス記憶部 1 5 8 に加えて、図 7 に示すような 3 6 ビットで表されるリファレンスベクトル $\{Re_i\}$ ($i = 1 \sim N$) を記憶するためのリファレンスベクトル記憶部 2 8 4 と、ベクトル量子化部 2 9 0 から与えられる二乗距離のうち、リファレンスベクトル記憶部 2 8 4 に記憶された V Q セントロイド番号の組合せに対応するものとの間の二乗距離を一時的に保存するための二乗距離保存部 2 8 2 とを含む。

【 0 0 3 5 】

リファレンスベクトル記憶部 2 8 4 に記憶される V Q セントロイド番号の組合せは次のように音素ごとの平均特徴パラメータを計算することにより予め決定される。すなわち、V Q コードブック 5 8 の作成データを用い、音素ごとに、特徴パラメータの平均ベクトル a_p を以下のようにして計算する。

【 0 0 3 6 】

【数 1】

$$\mathbf{a}_p = \frac{\sum_{s=1}^S \sum_{x_{st} \in p} x_{st}}{\sum_{s=1}^S N_{sp}} \quad (1)$$

ただし \mathbf{a}_p は音素 p の平均特徴パラメータ、

x_{st} は発声 s の t フレーム目の特徴パラメータ、

S はコードブック作成に用いた発声数、

N_{sp} は発声 s に含まれる音素 p の総分析フレーム数を示す。

次に、上記した式で求めた平均ベクトル \mathbf{a}_p に対し、V Q コードブック 5 8 を用いたベクトル量子化を行なう。すなわち、

【 0 0 3 7 】

【数 2】

$$Re_i = VQ(\mathbf{a}_p) \quad (2)$$

ただし、 Re_i は i 番目のリファレンスベクトル、

$VQ()$ はベクトル量子化を表す。

ベクトル量子化を行なった音素ごとの平均特徴パラメータをリファレンスベクトルとすることにより、V Q セントロイド番号の組合せとして、3 6 ビット × 音素数でリファレンスベクトルを表現できる。従って、本実施の形態によればフロントエンド側のメモリを削減することができる。

10

20

30

40

50

【 0 0 3 8 】

等化部 2 7 0 はさらに、リファレンスベクトル記憶部 2 8 4 に記憶されたリファレンスベクトルのうち、バイアス減算部 1 5 0 の出力により表される点との間の距離が最も近いものを二乗距離保存部 2 8 2 に保存された二乗距離に基づいて選択し、以下の式に従いバイアスを計算するためのリファレンス選択・バイアス計算部 2 8 0 と、リファレンス選択・バイアス計算部 2 8 0 により計算されたバイアスを用い、バイアス記憶部 1 5 8 の内容を以下に述べるように更新するためのバイアス更新部 2 8 6 とを含む。

【 0 0 3 9 】

すなわち、リファレンス選択・バイアス計算部 2 8 0 は、

【 0 0 4 0 】

【数 3】

$$\mathbf{xc}(t) - \text{Re}_j \quad (3)$$

ただし $\mathbf{xc}(t)$ は入力音声の t フレーム目の、バイアス除去後の特徴パラメータ、

$$j = \arg \min_j (\mathbf{xc}(t) - \text{Re}_j)$$

をバイアス更新部 2 8 6 に与える。バイアス更新部 2 8 6 は、この値を受けて、バイアス記憶部 1 5 8 に記憶されているバイアスベクトル $h(t)$ の値を以下の式に従い更新する

【 0 0 4 1 】

【数 4】

$$\mathbf{h}(t) = \mathbf{h}(t-1) + \alpha_t (\mathbf{xc}(t) - \text{Re}_j) \quad (4)$$

ただし α_t は各リファレンスベクトル Re_j の更新係数

この端末装置 2 6 0 及び等化部 2 7 0 は次のように動作する。

【 0 0 4 2 】

まず、図 1 0 に示されるリファレンスベクトル記憶部 2 8 4 に記憶するリファレンスベクトルを決定しておく。この算出方法については式 (1) (2) を参照して前述した。

【 0 0 4 3 】

続いて音声信号をサーバに送信する際の動作を説明する。図 9 を参照して、音響分析部 5 2 による音響処理までは従来のもと同様である。等化部 2 7 0 における処理は以下のように、従来のもとは異なる。

【 0 0 4 4 】

図 1 0 を参照して、バイアス減算部 1 5 0 の出力はベクトル量子化部 2 9 0 とリファレンス選択・バイアス計算部 2 8 0 とに与えられる。ベクトル量子化部 2 9 0 は、バイアス減算部 1 5 0 の出力により表される点と、VQコードブック 5 8 の各セントロイドベクトルにより表される点との間の二乗距離を算出する。バイアス減算部 1 5 0 の出力に最も近いセントロイドに対応するコードブックの番号でバイアス減算部 1 5 0 の出力をベクトル量子化し、サーバに対し VQ コード列として出力する。このとき算出される二乗距離のうち、リファレンスベクトル記憶部 2 8 4 に記憶されたリファレンスベクトルに対応するものが二乗距離保存部 2 8 2 により一時保存される。

【 0 0 4 5 】

リファレンス選択・バイアス計算部 2 8 0 は、二乗距離保存部 2 8 2 に一時保存された二乗距離の組合せから各リファレンスと入力ベクトルの距離を計算し、うち最小のリファレンスベクトルをリファレンスベクトル記憶部 2 8 4 から選択する。リファレンス選択・バイアス計算部 2 8 0 はさらに、上記した式 (3) に従ってバイアスを算出し、バイアス更新部 2 8 6 に与える。

10

20

30

40

50

【 0 0 4 6 】

バイアス更新部 2 8 6 は、リファレンス選択・バイアス計算部 2 8 0 から与えられたバイアスを使用して、式 (4) に従ってバイアス記憶部 1 5 8 に記憶されたバイアスを更新する。バイアス減算部 1 5 0 は、次の MFCC サンプルからバイアス記憶部 1 5 8 に記憶されたバイアスを減算してベクトル量子化部 2 9 0 及びリファレンス選択・バイアス計算部 2 8 0 に与える。

【 0 0 4 7 】

以下同様にして、入力される MFCC サンプルの周波数特性の等化処理を行ないながら、MFCC のベクトル量子化が行なわれ、VQ コード列がサーバに送信され音声認識処理に供される。

10

【 0 0 4 8 】

なお、上記した実施の形態に係る装置では、MFCC 1 2 次元を特徴量として使用しているが、これ以外の特徴量、たとえば LPC (線型予測符号化) ケプストラム係数等上記した等化手法を用いてもよい。

【 0 0 4 9 】

[音声認識実験]

上記した実施の形態と同様の手法の有効性を検証するため、日本音響学会新聞記事読上げ音声コーパス (J N A S) を用いて音声認識実験を行なった。

【 0 0 5 0 】

- 実験条件 -

音響モデルの学習には IPA (独立行政法人 情報処理推進機構) の学習セットから男性話者が発声した音素バランス文 (話者: 1 0 3 名、発声数: 5 , 1 6 8) を使用した。テストセットは、学習データと同様に IPA で使用されているテストセットから男性話者が発声した 1 0 0 発声を用いた。音響モデル学習には、クリーンな発声に対し、E T S I A d v a n c e d D S R フロントエンドの音響分析条件で分析した量子化前の MFCC 1 2 次元、その一次回帰係数 1 2 次元、対数パワーの一次回帰係数、合計 2 5 次元に比較対象の各手法を適用した特徴ベクトルを用いた。認識時には、量子化した MFCC 1 2 次元、その一次回帰係数 1 2 次元、対数パワーの一次回帰係数、合計 2 5 次元に比較対象の各手法を適用した特徴ベクトルを用いた。

20

【 0 0 5 1 】

音響モデルは、木構造クラスタリングにより状態共有した 3 状態 1 6 混合の音素環境依存 HMM (4 3 音素) を用いた。総状態数は約 1 , 0 0 0 状態である。デコードには Julius - 3 . 1 p 2 を用い、単語正解精度 (word accuracy) で評価を行なった。各実験の単語正解精度は、手法ごとにデコード時に最適なパスの広さの設定を行なった結果より計算した。

30

【 0 0 5 2 】

周波数特性の際による認識性能の変化を調べるため、A U R O R A 2 データベースで使用されている G 7 1 2 , M I R S を用いた。

【 0 0 5 3 】

- 実験結果 -

表 1 に音声認識実験結果を示す。表中の「E S 2 0 2」は従来手法 (A d v a n c e d D S R フロントエンド)、「E S 2 0 2 w / o B E Q」は従来手法の等化を適用しない A d v a n c e d D S R フロントエンドの認識結果を示す。

40

【 0 0 5 4 】

【表 1】

表1

	Filter		
	clean	G712	MIRS
ES202w/oBEQ	84.78	81.10	71.27
ES202	86.24	83.64	83.33
本手法	86.69	86.24	85.36

表より、周波数特性の正規化（等化）を行わない場合（ES202w/oBEQ）、乗算性雑音の影響で単語正解精度が劣化していることが分かる。周波数特性等化を行なうと、周波数特性の変動による単語正解精度の劣化を抑制可能であることが分かる。

10

【0055】

またこの表から、本発明に係る等化手法を用いると、実験における全環境下で他の手法より高い精度を示すことが分かる。これにより、複数のリファレンスペクトルを使用する本手法では、周波数変動による特徴パラメータの移動を抑制する効果が高いことが分かる。

【0056】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内でのすべての変更を含む。

20

【図面の簡単な説明】

【0057】

【図1】従来の分散型音声認識システム20のブロック図である。

【図2】VQコードブック58の構成を説明するための図である。

【図3】セントロイドの概念を説明するための図である。

【図4】従来技術による等化手法を説明するための図である。

【図5】従来技術で使用された等化部54のブロック図である。

【図6】本発明の一実施の形態に係る複数リファレンスペクトルの概念を説明するための図である。

30

【図7】リファレンスペクトルの構成を模式的に示す図である。

【図8】複数のリファレンスを使用した等化処理を説明するための図である。

【図9】本発明の一実施の形態に係る等化処理を実施する端末装置260のブロック図である。

【図10】図9に示す端末装置260において使用される等化部270のブロック図である。

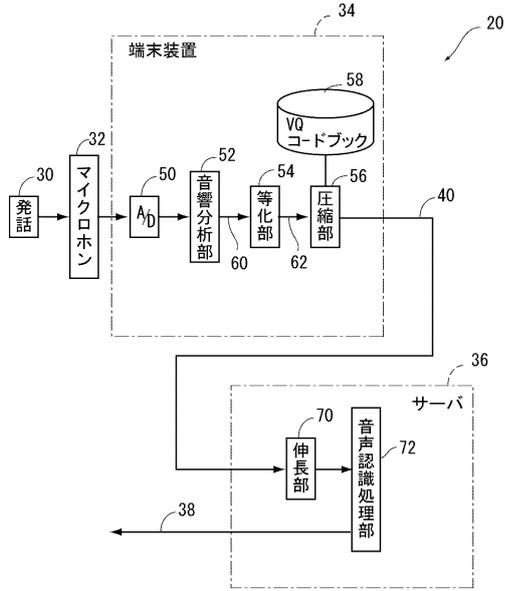
【符号の説明】

【0058】

20 従来の分散型音声認識システム、30 発話、32 マイクロホン、34、260 端末装置、36 サーバ、50 アナログ/デジタル変換機、52 音響分析部、54、270 等化部、56、272 圧縮部、58 ベクトル量子化(VQ)コードブック、70 伸長部、72 音声認識処理部、150 バイアス減算部、152、284 リファレンスペクトル記憶部、154 バイアス計算部、156、286 バイアス更新部、158 バイアス記憶部、170、290 ベクトル量子化部、282 二乗距離保存部、280 リファレンス選択・バイアス計算部

40

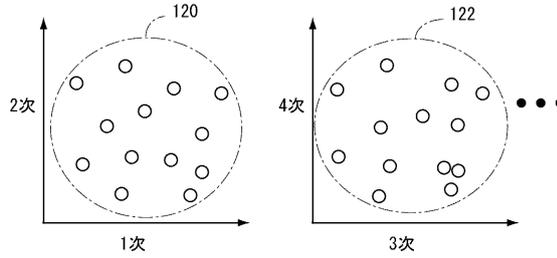
【図1】



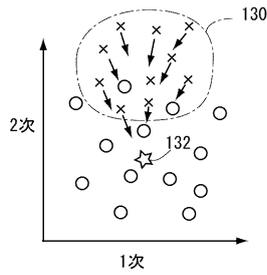
【図2】

MFCC	個数	ビット数
1, 2次	64	6
3, 4次	64	6
⋮	⋮	⋮
11, 12次	64	6
パワー及び ケプストラム0次	256	8

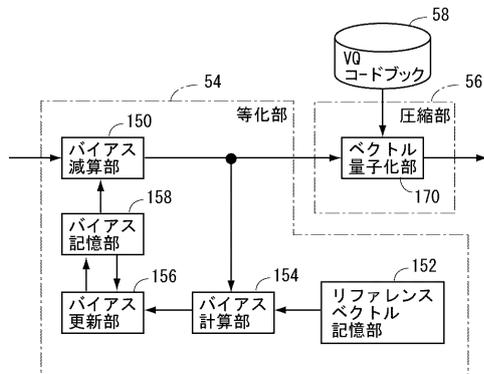
【図3】



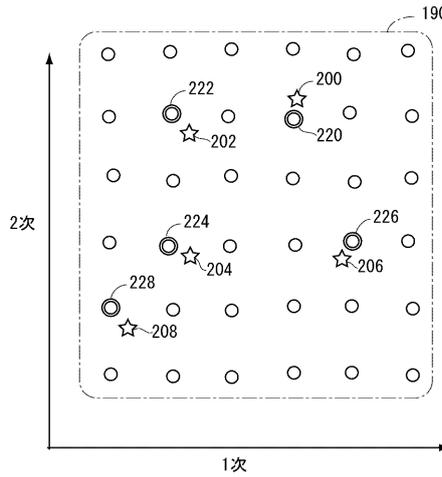
【図4】



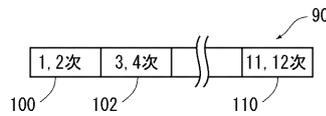
【図5】



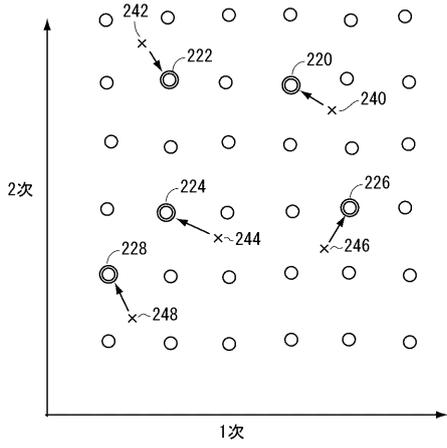
【図6】



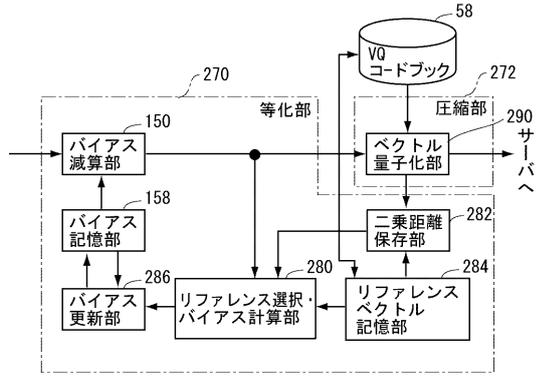
【図7】



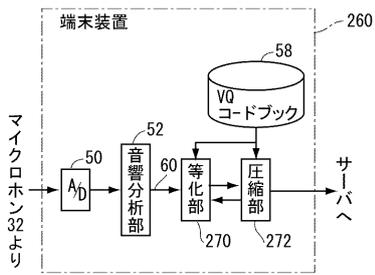
【図8】



【図10】



【図9】



フロントページの続き

審査官 涌井 智則

- (56)参考文献 特開平08-234788(JP,A)
特開2000-267692(JP,A)
特開平07-271394(JP,A)
特開平08-016311(JP,A)
特開平07-219599(JP,A)
特開平02-037400(JP,A)
柘植ら,「周波数特性の変動に頑健な実時間分散音声認識手法」,情報処理学会研究報告SLP,
2003年12月19日,Vol.2003,No.124(2003-SLP-49),pp.13-18

(58)調査した分野(Int.Cl.,DB名)

G10L 11/00-21/06
H03M 7/30
Cinii