

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4714523号
(P4714523)

(45) 発行日 平成23年6月29日(2011.6.29)

(24) 登録日 平成23年4月1日(2011.4.1)

(51) Int.Cl.		F I			
G 1 0 L	17/00	(2006.01)	G 1 0 L	17/00	2 0 0 Z
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	5 0 0 P
			G 1 0 L	15/06	4 0 0 U
			G 1 0 L	15/06	3 0 0 D

請求項の数 7 (全 13 頁)

(21) 出願番号	特願2005-217478 (P2005-217478)	(73) 特許権者	310022372
(22) 出願日	平成17年7月27日(2005.7.27)		富士通東芝モバイルコミュニケーションズ株式会社
(65) 公開番号	特開2007-33898 (P2007-33898A)		神奈川県川崎市中原区上小田中四丁目1番1号
(43) 公開日	平成19年2月8日(2007.2.8)	(74) 代理人	100070150
審査請求日	平成20年7月16日(2008.7.16)		弁理士 伊東 忠彦
		(72) 発明者	杉浦 千加志
			東京都青梅市末広町2丁目9番地 株式会社東芝青梅事業所内
		(72) 発明者	井阪 岳彦
			東京都青梅市末広町2丁目9番地 株式会社東芝青梅事業所内
		審査官	菊池 智紀

最終頁に続く

(54) 【発明の名称】 話者照合装置

(57) 【特許請求の範囲】

【請求項1】

話者の入力音声¹を複数のフレームに時分割し、この分割された複数のフレームからそれぞれ特徴ベクトルを生成する特徴ベクトル生成手段と、

前記特徴ベクトル生成手段により生成された複数の特徴ベクトルを代表する特徴代表ベクトルを生成する特徴代表ベクトル生成手段と、

登録対象²の話者³について、前記特徴ベクトル生成手段により生成された第1の特徴ベクトルをベクトル量子化することで得られるコードベクトル及び、前記特徴代表ベクトル生成手段により生成された第1の特徴代表ベクトルを記憶するコードブックと、

照合対象⁴の話者⁵について、前記特徴ベクトル生成手段により生成された第2の特徴ベクトルに対し、当該第2の特徴ベクトルと前記特徴代表ベクトル生成手段により生成された第2の特徴代表ベクトルとの間の距離を縮小する変換処理を行う手段と、

前記コードブックに記憶されたコードベクトルに対し、当該コードベクトルと前記第1の特徴代表ベクトルとの間の距離を縮小する変換処理を行う手段と、

前記変換処理後のコードベクトルと前記変換処理後の第2の特徴ベクトルとの間のベクトル量子化歪みを算出する手段と、

前記算出されたベクトル量子化歪みを予め設定されたしきい値と比較し、その比較結果を話者の照合結果として出力する判定手段と

を具備し、

前記特徴代表ベクトル生成手段は、

10

20

前記特徴ベクトル生成手段により生成された複数の特徴ベクトルをそれぞれベクトル量子化する手段と、

前記ベクトル量子化された複数のコードベクトルの重心を算出し、この算出された重心を特徴代表ベクトルとする手段と

を備えることを特徴とする話者照合装置。

【請求項 2】

前記変換処理を行う手段は、特徴ベクトルの変換機能を有する変換行列に、特徴ベクトルと第 1 の重み行列を乗じた荷重特徴代表ベクトルとの差分を乗じた行列を求め、この行列に対して、第 2 の重み行列を乗じた荷重特徴代表ベクトルを加算することで変換演算を行うことを特徴とする請求項 1 記載の話者照合装置。

10

【請求項 3】

前記変換処理を行う手段は、前記変換行列が対角行列からなる場合に、当該対角行列の各対角成分を、当該対角成分の平均値、及び特徴代表ベクトルの次元ごとのばらつきから算出することを特徴とする請求項 2 記載の話者照合装置。

【請求項 4】

前記特徴ベクトル生成手段は、

入力音声の複数のフレームのうち有声音フレームを抽出する手段と、

前記抽出された有声音フレームから特徴ベクトルを生成する手段と

を備えることを特徴とする請求項 1 記載の話者照合装置。

【請求項 5】

前記特徴代表ベクトル生成手段は、

前記特徴ベクトル生成手段により生成された複数の特徴ベクトルのうち有声音フレームのみからなる特徴ベクトルを抽出する手段と、

前記抽出された有声音のみからなる特徴ベクトルから特徴代表ベクトルを算出する手段と

を備えることを特徴とする請求項 1 記載の話者照合装置。

20

【請求項 6】

前記特徴代表ベクトル生成手段は、

前記特徴ベクトル生成手段により生成された複数の特徴ベクトルのうち入力音声の母音に該当する特徴ベクトルを抽出する手段と、

前記抽出された母音に該当する特徴ベクトルから特徴代表ベクトルを算出する手段と

を備えることを特徴とする請求項 1 記載の話者照合装置。

30

【請求項 7】

前記変換処理に使用する変換係数及び前記しきい値の両方を可変設定する手段と、

前記変換係数及びしきい値の両方が可変設定されるごとに、前記コードベクトル及び第 2 の特徴ベクトルに対する変換処理、変換処理後のコードベクトル及び第 2 の特徴ベクトル間のベクトル量子化歪みの算出処理、及び算出されたベクトル量子化歪みをしきい値と比較する処理を繰り返す手段と、

前記繰り返し処理によりえられる複数の判定結果をもとに照合結果を得る手段と

を、さらに具備することを特徴とする請求項 1 記載の話者照合装置。

40

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、例えば利用者の音声の特徴を抽出して本人認証を行う話者照合装置に関する。

【背景技術】

【0002】

近年、例えばセキュアな状態が要求される建物への入室や機器の使用に際し、利用者の音声の特徴を抽出して本人認証を行う話者照合装置が種々提案されている。

例えば、話者ごとにその入力音声をサンプリングしてその特徴量から特徴ベクトルを求

50

め、この特徴ベクトルをベクトル量子化してコードブックに登録する。そして、以後利用者の音声が入力されるごとに、この入力音声の特徴量をもとに特徴ベクトルを求め、この特徴ベクトルをベクトル量子化したのち上記コードブックに登録されたベクトルと照合することにより本人認証を行う装置が知られている（例えば、特許文献1を参照。）。

【0003】

また、別の話者照合装置として、本人話者の特徴ベクトルと他人話者の特徴ベクトルをもとに共分散行列を計算する。そして、この計算された共分散行列を用いて特徴ベクトルを変換することにより、統計的に識別性能を向上させる装置も提案されている（例えば、特許文献2を参照。）。

【特許文献1】特開平7-248791号公報

【特許文献2】特許第3080388号明細書

【発明の開示】

【発明が解決しようとする課題】

【0004】

ところが、特許文献1に記載された装置では、予めコードブックに登録された利用者本人の音声の特徴ベクトルと、入力音声の特徴ベクトルとをただ単純に照合するようにしている。このため、音声背景雑音の影響により歪んだり、また話者の発話内容が異なると、他人話者の特徴量分布がコードブックに登録された本人話者の特徴量分布に包含され、この結果他人話者を本人話者と誤認識したり、また本人話者を他人話者として誤認識する場合があります、照合精度が低かった。

【0005】

一方、特許文献2に記載された装置では、共分散行列をもとに特徴ベクトルを変換するようにしている。このため、特許文献1に記載された装置に比べ照合精度を高めることが可能である。しかしながら、共分散行列を計算するためには他人話者の音声の特徴ベクトルを取得する必要があり、その処理のために余計な手間やコストが必要となる。

【0006】

この発明は上記事情に着目してなされたもので、その目的とするところは、音声が背景雑音もしくはこの抑圧処理の影響により歪んだり、また話者の発話内容が異なる場合でも、他人話者の音声の特徴ベクトルを必要とすることなく照合精度を高めることを可能にした話者照合装置を提供することにある。

【課題を解決するための手段】

【0007】

上記目的を達成するためにこの発明は、話者の入力音声を時分割した複数のフレームからそれぞれ特徴ベクトルを生成すると共に、この生成された複数の特徴ベクトルを代表する特徴代表ベクトルを生成する手段を備える。そして、登録対象の話者については、上記生成された特徴ベクトルをベクトル量子化することで生成されるコードベクトル及び特徴代表ベクトルをコードブックに記憶する。一方照合を行う際には、上記コードブックに記憶された第1の特徴ベクトルに対し、当該第1の特徴ベクトルと上記第1の特徴代表ベクトルとの間の距離を縮小する変換処理を行い、かつ照合対象の話者の入力音声から生成される第2の特徴ベクトルに対し、当該第2の特徴ベクトルと第2の特徴代表ベクトルとの間の距離を縮小する変換処理を行う。そして、上記変換処理後のコードベクトル及び第2の特徴ベクトル間のベクトル量子化歪みを算出し、この算出されたベクトル量子化歪みを予め設定されたしきい値と比較して、その比較結果を照合結果として出力するように構成したものである。そして、特徴代表ベクトルを生成する手段は、特徴ベクトルを生成する手段により生成された複数の特徴ベクトルをそれぞれベクトル量子化する手段と、ベクトル量子化された複数のコードベクトルの重心を算出し、この算出された重心を特徴代表ベクトルとする手段とを備える。

【0008】

したがって、入力音声が背景雑音もしくはこれを抑圧処理した影響により歪んだり、また話者の発話内容が異なる場合でも、特徴ベクトルに対し上記変換処理を施すことにより

10

20

30

40

50

、コードブックに記憶された登録対象のコードベクトルの分布と、照合対象話者の特徴ベクトルの分布がそれぞれ縮小される。このため、登録対象話者のコードベクトルに対する照合対象話者の特徴ベクトル集合の包含は低減される。したがって、登録対象話者のコードベクトルと照合対象話者の特徴ベクトルとの間のベクトル量子化歪みの差は顕著になり、これにより照合精度を高めることが可能となる。

【発明の効果】

【0009】

要するにこの発明では、コードブックに記憶された第1の特徴ベクトル、及び照合対象話者の入力音声から生成される第2の特徴ベクトルに対し、それぞれ当該第1の特徴ベクトルとその特徴代表ベクトルとの間、及び第2の特徴ベクトルとその特徴代表ベクトルとの間の距離を縮小する変換処理を行う。そして、上記変換処理後のコードベクトル及び第2の特徴ベクトル間のベクトル量子化歪みを算出し、この算出されたベクトル量子化歪みを予め設定されたしきい値と比較して、その比較結果を話者の照合結果として出力するようにしている。

10

【0010】

したがってこの発明によれば、音声背景雑音もしくはこの抑圧処理の影響により歪んだり、また話者の発話内容が異なる場合でも、他人話者の音声の特徴ベクトルを必要とすることなく照合精度を高めることを可能にした照合装置を提供することができる。

【発明を実施するための最良の形態】

【0011】

以下、図面を参照してこの発明の実施形態を説明する。

まず、この発明の原理について説明する。

ベクトル量子化 (Vector Quantization; VQ) モデルを用いた話者照合において発生する他人受入エラーは、例えば図5に示すように他人話者の特徴ベクトルが本人コードブックに包含されて、VQ歪みが小さくなるために起こる。この問題を回避するには、本人コードブック及び他人話者の特徴ベクトルの分布をそれぞれ縮小する話者内分布補正が有効である。例えば図6に示すように、比較対象の本人コードブック及び他人話者の特徴ベクトルのそれぞれにおいて、コードベクトル及び特徴ベクトルの分布を各特徴代表ベクトル方向に近づければよい。

20

【0012】

この話者内分布補正処理は、コードベクトル及び特徴ベクトルに対しそれぞれ、当該ベクトルと特徴代表ベクトルとの間の距離を縮小する変換処理を行うことにより実現される。例えば、変換処理前の特徴ベクトルを v_i 、変換処理後の特徴ベクトルを v'_i 、特徴代表ベクトルを v_g 、変換行列を A、第1の重み行列を B、第2の重み行列を C、ベクトル数を N とすると、変換式は次の(1)式で表される。

30

【数1】

$$v'_i = A \cdot (v_i - B \cdot v_g) + C \cdot v_g, \quad (i=1,2,\dots,N) \quad \dots (1)$$

【0013】

この(1)式において、変換行列 A は例えば対角成分 $\text{diag}\{A\}$ が 0.0 以上である $D \times D$ の対角行列で表される。また、第1ならびに第2の重み行列 B、C は例えば $D \times D$ の単位行列である。なお、D はベクトル次元数を示している。

40

【0014】

特徴代表ベクトルは、例えばコードベクトル及び特徴ベクトル集合の重心として表すことができる。このようにコードベクトル及び特徴ベクトル集合の重心をそれぞれ特徴代表ベクトルとすることにより、背景雑音もしくはこの抑圧処理による影響や、発話内容の違いによるコードベクトルならびに特徴ベクトルの変動を吸収することができ、話者のみに依存させることが可能となる。

【0015】

上述した対角行列 A の各対角成分を a_d とすると、特徴ベクトルの分布を小さくするた

50

めには a_d の値の範囲は $0.0 \sim 1.0$ であり、ここで、 a_d を縮小率と呼ぶことにする。縮小率 a_d は、例えば、すべての次元に対し $a_d = 1.0$ を設定すると、無変換の状態となる。これに対しすべての次元に対し $a_d = 0.0$ を設定すると、特徴代表ベクトルのみを話者の特徴ベクトルとする状態となる。

図7及び図8は特徴代表ベクトルの話者依存性の一例を示すもので、図7は話者Xの特徴代表ベクトルを、また図8は話者Yの特徴代表ベクトルをそれぞれ示している。図9は上記話者Xと話者Yの特徴代表ベクトルを重ねて表したものである。

【0016】

同図から明らかなように、話者間における特徴代表ベクトルのばらつきが大きい次元は、より話者間の特徴の差を顕著に表す次元である。これに対し特徴代表ベクトルのばらつきが小さい次元は、話者間の特徴の差が少ない次元である。そこで、ばらつきが大きい次元に対しては縮小率 a_d を 0.0 に近い値に設定し、一方ばらつきが小さい次元については縮小率 a_d を 1.0 に近い値に設定する。この縮小率 a_d の算出式は例えば次の(2)式のように表される。

【数2】

$$a_d = p \cdot \frac{1}{\sigma_d} + q \quad \text{if } a_d < 0, \text{ then } a_d = 0 \quad \dots (2)$$

【0017】

ただし、 σ_d は話者間の特徴代表ベクトルの次元ごとの標準偏差 ($d=1, 2, \dots, D$) であり、特徴代表ベクトルにおいてどの次元に話者性が現れやすいかを示す。値が大きいほど話者性が現れやすい次元である。 p は縮小率 a_d の次元平均を B にするための調整用パラメータであり、直接値を指定することはなく、次式によって表される。

【数3】

$$p = \frac{(B - q) \cdot D}{\sum_{i=1}^D 1/\sigma_i}$$

【0018】

同式において、 B は縮小率 a_d の次元平均値であり、話者特徴ベクトルの包含を低減するためのパラメータである。 B は値が大きいほど包含の低減効果が少なくなる。これに対し、値が小さいほど包含の低減効果が大きい、値を小さくし過ぎると話者の特徴代表ベクトルの変動によるVQ歪みの増大が顕著になるため、本人許否エラーが発生しやすくなる。このパラメータ B は、例えば 0.5 付近の値を設定するのが好ましく、値を小さくし過ぎることによる過剰な補正は避ける方が望ましい。

【0019】

q は標準偏差 σ_d の次元ごとの差をどの程度強調させるかを制御するパラメータであり、 $q = 0.0$ のときに強調の程度が最大となる。これに対し、 $q = B$ のときに次元ごとの差は無くなり、縮小率は次元一定の値 $= B$ となる (このとき $p = 0.0$)。

上記標準偏差 σ_d は、予め複数話者の複数の環境下における特徴ベクトルを分析し算出しておくことで、適当な値を設定することが可能である。ここで、算出値は十分に多くの話者数、雑音環境下での音声进行分析して得られるものであれば、音声信号から抽出される特徴ベクトルの普遍的な性質であると見なすことができる。つまり、この値は事前に算出すればよいものであり、使用環境に応じて値を事後的に算出しなくてもよい。また、特徴ベクトルにはLPCケプストラムなどのケプストラム係数を用いているので、分析用の環境と実環境でのマイクロホンなどを含む音声入力系の違いは特徴ベクトルの線形差分として現れる。このため上記音声入力系の違いは、話者間の特徴代表ベクトルの標準偏差 σ_d を計算する際に吸収される。

【0020】

次に、この発明に係わる話者照合装置の一実施形態を説明する。

10

20

30

40

50

図 1 はその機能構成を示すブロック図である。この話者照合装置は、音声特徴生成処理部 1 を備えている。音声特徴生成処理部 1 は、入力された話者の音声から特徴ベクトルを生成すると共に、この生成された特徴ベクトル集合の重心を求めてこの重心を特徴代表ベクトルとする。

【 0 0 2 1 】

図 2 はこの音声特徴生成処理部 1 の機能構成を示すブロック図である。音声特徴生成処理部 1 は、前処理部 1 1 と、L P C 係数算出部 1 2 と、L P C C 生成部 1 3 と、話者特徴代表ベクトル生成部 1 4 とを有している。

前処理部 1 1 は、入力された音声信号に対しアナログ - デジタル (A / D) 変換、ならびに雑音抑圧処理を行ったのち音声分析区間を設定し、この分析区間内の音声波形を一定の時間及び一定のシフト周期で分析窓により切り出し、音声フレームを生成して保持する。

10

【 0 0 2 2 】

L P C 係数算出部 1 2 は、上記前処理部 1 1 により形成された各音声フレームから、線形予測符号化 (Linear Prediction Coding ; L P C) により音声信号に含まれている個人性情報に関する特徴量を抽出する。L P C C 生成部 1 3 は、上記 L P C 係数算出部 1 2 により抽出された特徴量をもとに特徴ベクトル (L P C ケプストラム、パワー項 0 次を含まない 1 次以上のパラメータ) を生成する。話者特徴代表ベクトル生成部 1 4 は、上記 L P C C 生成部 1 3 により生成された特徴ベクトル集合の重心を算出する。そして、この算出された重心を特徴代表ベクトルとする。

20

【 0 0 2 3 】

またこの話者照合装置は、話者登録処理のための機能として、ベクトル量子化部 2 と、話者別コードブックデータベース 3 とを備える。ベクトル量子化部 2 は、話者登録モードが選択されている状態で、上記音声特徴生成処理部 1 により生成された特徴ベクトル集合とその特徴代表ベクトルを取り込む。そして、特徴ベクトルに対しベクトル量子化を行い、コードベクトルを出力する。話者別コードブックデータベース 3 は、例えば図 3 に示すように、上記ベクトル量子化部 2 により生成されたコードベクトルと、上記音声特徴生成処理部 1 により生成された特徴代表ベクトルを、図示しない話者名入力手段により入力された話者別名称 I D に対応付けて記憶する。

【 0 0 2 4 】

30

さらにこの話者照合装置は、話者照合処理のために必要な機能として、変換係数・しきい値制御部 4 と、変換係数・しきい値データベース 5 と、第 1 の話者特徴変換処理部 6 と、第 2 の話者特徴変換処理部 7 と、V Q 歪み算出部 8 と、しきい値判定処理部 9 と、終了判定部 1 0 とを備えている。

【 0 0 2 5 】

変換係数・しきい値制御部 4 は、他人受入エラー率を低減するための話者特徴変換係数と、この値に対応した V Q 歪み判定用のしきい値を変換係数・しきい値データベース 5 から読み込む。話者特徴変換係数は、縮小率 a のことを指している。縮小率 a は、先に示した (2) 式に従い、話者間における特徴代表ベクトルの次元ごとのばらつきを考慮して、事前に用意された値である。しきい値は、本人受入率と他人拒否率が等しくなるように設定されている。変換係数・しきい値データベース 5 は、例えば図 4 に示すように、上記算出及び設定された話者特徴変換係数及び判定しきい値を、記憶する。

40

【 0 0 2 6 】

第 1 の話者特徴変換処理部 6 は、話者照合モードが選択されているときに、上記話者別コードブックデータベース 3 から話者ごとにコードベクトル及び特徴代表ベクトルを読み出す。そして、先に (1) 式により示した変換式と、上記変換係数・しきい値制御部 4 から与えられる変換係数とを用いて、上記コードベクトルを特徴代表ベクトルに近づける変換処理を行う。

【 0 0 2 7 】

第 2 の話者特徴変換処理部 7 は、話者照合モードが選択されているときに、上記音声特

50

徴生成処理部 1 により生成された話者の特徴ベクトル集合とその特徴代表ベクトルを取り込む。そして、先に (1) 式により示した変換式と、上記変換係数・しきい値制御部 4 から与えられる変換係数とを用いて、上記話者の特徴ベクトル集合を特徴代表ベクトルに近づける変換処理を行う。

【 0 0 2 8 】

V Q 歪み算出部 8 は、上記第 1 及び第 2 の話者特徴変換処理部 6 , 7 からそれぞれ変換されたコードベクトル及び変換された話者特徴ベクトルを取り込む。そして、これらの変換後のコードベクトルと話者特徴ベクトルとの間 V Q 歪みを算出する。

しきい値判定処理部 9 は、上記 V Q 歪み算出部 8 により算出された V Q 歪みを、上記変換係数・しきい値制御部 4 から与えられるしきい値と比較し、その比較結果を表すフラグ信号を出力する。

10

【 0 0 2 9 】

終了判定部 1 0 は、上記しきい値判定処理部 9 から出力されるフラグ信号により照合対象の話者が本人であるか否かを判定する。そして、本人と判定された場合に、変換係数・しきい値制御部 4 に対し変換係数変更制御信号を与えることにより変換係数及びしきい値を更新させ、これにより上記第 1 及び第 2 の話者特徴変換処理部 6 , 7 による変換処理、及びしきい値判定処理部 9 による V Q 歪みの判定処理を繰り返し実行させる。そして、上記繰り返し処理が予め設定された回数実行されると、その総合判定結果を照合判定情報として出力する。

【 0 0 3 0 】

20

次に、以上のように構成された装置の動作を説明する。

まず、照合に先立ち照合対象となる話者、つまり本人話者の音声の特徴の登録が行われる。すなわち、本人話者が自身の音声をマイクロホンから入力すると、この入力音声は音声特徴生成処理部 1 において音声フレームに変換されたのち、このフレームごとに L P C 分析され、これにより上記入力音声の特徴量が抽出される。そして、この特徴量をもとに特徴ベクトルの集合が生成される。また、この生成された特徴ベクトル集合はの重心ベクトルがこの特徴ベクトル集合の特徴代表ベクトルとなる。

【 0 0 3 1 】

上記生成された特徴ベクトル集合はベクトル量子化部 2 によりベクトル量子化されたのち、話者別名称 I D と対応付けられて上記特徴代表ベクトルと共に話者別コードブックデータベース 3 に記憶される。他の照合対象話者についても、同様に音声の特徴ベクトル集合及びその特徴代表ベクトルが生成され、当該特徴ベクトル集合のコードベクトル及び特徴代表ベクトルが話者別コードブックデータベース 3 に記憶される。

30

また、変換係数・しきい値データベース 5 には、予め前実験を行うことによって得られる変換係数としきい値のセットが記憶される。

【 0 0 3 2 】

さて、以上のように各データベース 3 , 5 への登録処理が終了すると、続いて話者音声の照合処理が以下のように実行される。すなわち、照合対象話者の音声が入力されると、音声特徴生成処理部 1 により上記入力音声の特徴ベクトルの集合及びその特徴代表ベクトルが生成され、この特徴ベクトル集合及び特徴代表ベクトルは第 2 の話者特徴変換処理部 7 に入力される。第 2 の話者特徴変換処理部 7 では、先に (1) 式により示した変換式と、上記変換係数・しきい値制御部 4 から与えられる変換係数とを用いて、上記話者の特徴ベクトル集合をその特徴代表ベクトルに近づける変換処理が行われる。

40

【 0 0 3 3 】

またそれと並行して第 1 の話者特徴変換処理部 6 では、話者別コードブックデータベース 3 から話者ごとにコードベクトルとその特徴代表ベクトルが読み出される。そして、先に (1) 式により示した変換式と、上記変換係数・しきい値制御部 4 から与えられる変換係数とを用いて、上記コードベクトルをその特徴代表ベクトルに近づける変換処理が行われる。

【 0 0 3 4 】

50

VQ歪み算出部8では、上記変換された話者の特徴ベクトル集合と、話者別コードブックデータベース3に記憶されたコードベクトルとの間のVQ歪みが算出され、この算出されたVQ歪みはしきい値判定処理部9によりしきい値と比較される。そして、その判定結果がフラグ信号として終了判定部10に出力される。

【0035】

終了判定部10では、上記フラグ信号をもとに照合対象の話者が本人であるか否かが判定される。そして、本人と判定されると、終了判定部10から変換係数・しきい値制御部4に対し変換係数変更制御信号が与えられる。この結果、変換係数・しきい値制御部4では変換係数及びしきい値が更新される。この更新は、他人受入エラー率を低減する方向に変換係数及びしきい値を一定量シフトすることにより行われる。第1及び第2の話者特徴変換処理部6,7ではそれぞれ、上記更新された変換係数を用いて上記コードベクトル及び話者の特徴ベクトル集合の変換処理が行われる。そして、この変換されたコードベクトル及び話者の特徴ベクトル集合間のVQ歪みがVQ歪み算出部8により算出され、この算出されたVQ歪みがしきい値判定処理部9においてしきい値と比較される。

10

【0036】

終了判定部10では、この比較結果を表すフラグ信号をもとに照合対象の話者が本人であるか否かが判定され、本人であれば再び変換係数・しきい値制御部4に対し変換係数変更制御信号が与えられる。そして、第1及び第2の話者特徴変換処理部6,7では、さらに更新された変換係数をもとにコードベクトル及び話者の特徴ベクトル集合の変換処理が行われ、この変換処理後のベクトル間のVQ歪みがしきい値判定処理部9においてしきい値と比較される。

20

【0037】

以後同様に、照合対象の話者が本人と判定されるごとに、変換係数及びしきい値が他人受入エラー率を低減する方向に順次更新され、この更新された変換係数及びしきい値をもとにコードベクトル及び話者の特徴ベクトル集合の変換処理から、VQ歪みの比較処理までの一連の照合処理が繰り返し行われる。そして、上記繰り返し回数が予め設定された1回以上の回数に達し、その時点での最終的な比較結果が本人であれば、話者は本人である旨の判定結果が出力される。これに対し、上記繰り返しの途中で他人と判定されると、話者は他人である旨の判定結果が出力される。またこの場合、照合不可と判断して、照合処理を始めからやり直すように利用者に促してもよい。

30

【0038】

以上述べたようにこの実施形態では、音声特徴生成処理部1において入力音声から特徴量を抽出して特徴ベクトルを生成するとともに、この生成された特徴ベクトル集合の重心を求めることにより特徴代表ベクトルを生成するようにしている。そして、照合モードにおいて、第1及び第2の話者特徴変換処理部6,7により、変換係数・しきい値制御部4から与えられる変換係数をもとにコードベクトル及び話者の特徴ベクトル集合をそれぞれの特徴代表ベクトルに近づける変換処理を行い、この変換処理後のベクトル間のVQ歪みをしきい値判定処理部9によりしきい値と比較し、これにより話者の本人判定を行っている。

【0039】

したがって、入力音声为背景雑音もしくはこの抑圧処理の影響により歪んだり、また話者の発話内容が異なる場合でも、話者別コードブックデータベース3に記憶されたコードベクトルの分散と、照合対象話者の特徴ベクトルの分布はそれぞれ縮小される。このため、登録話者のコードベクトルに対する照合対象話者の特徴ベクトル集合の包含は低減され、これにより登録話者のコードベクトルと照合対象話者の特徴ベクトルとの間のベクトル量子化歪みの差は顕著になり、これにより照合精度を高めることが可能となる。

40

【0040】

すなわち、この実施形態の装置は雑音環境下において高い頑健性を備える。図11に示すように、雑音環境下における特徴ベクトルは、クリーン音声の特徴ベクトルと比較すると、雑音のみの特徴ベクトルの分布に近づく傾向にある。これは、雑音が大きくなるほど

50

雑音に埋もれて雑音しか聞こえなくなることからも想像できる。この結果、雑音環境下では話者照合性能が低下するという問題を生じる。この問題を解決するために、一般には雑音付加音声に対して雑音抑圧処理を施す。しかし雑音抑圧処理を行うと、図11に例示するように過剰抑圧となって特徴ベクトルの分布が広がり、図5に示すような分布の包含を招くと云う副作用を生じる。

【0041】

これに対しこの発明の実施形態では、特徴ベクトルに対し特徴ベクトルの分布を縮小する変換処理を行うので、雑音抑圧処理を施した場合でも上記したような副作用による特徴ベクトルの分布の包含関係の強調を軽減することができ、結果として背景雑音による性能劣化を低減できる。

10

【0042】

さらにこの実施形態では、音声特徴生成処理部1において、入力音声の特徴ベクトル集合及び特徴代表ベクトルを生成する際に、有声音のみを対象としている。このため、背景雑音の影響を受けやすい無声音を予め特徴の抽出対象から排除することができ、これにより発話依存性を低減することができる。

またさらに、音声特徴生成処理部1において、特徴ベクトル集合の重心を算出し、この重心を特徴代表ベクトルとしている。このようにすると、発話偏りによる各特徴ベクトルのずれが平均をとることでキャンセルされ、結果として特徴代表ベクトルの変動を低減することができる。

【0043】

20

それに加え前記実施形態では、VQ歪みとしきい値との比較の結果、話者が本人と判定された場合に、変換係数及びしきい値を他人受入エラー率が低減する方向に順次更新しながら、コードベクトル及び話者の特徴ベクトル集合の変換処理からVQ歪みの比較処理までの一連の照合処理を、複数回繰り返し実行するようにしている。このため、本人拒否率及び他人受入エラー率を共に低く保持することができる。

【0044】

図10にこの実施形態に係わる装置による効果の一例を示す。この図10は、特徴量変換無しの場合、縮小率aをベクトルの次元ごとに一定にした場合、縮小率aを次元毎に異なる値とした場合の、平均誤認識率を折れ線グラフで表したものである。

この図10からも明らかなように、この実施形態による変換処理を使用することにより、本人コードブックに対する他人話者の特徴ベクトルの分布の包含関係が低減され、変換無しの場合に比べ平均誤認識率を改善することができる。

30

【0045】

なお、この発明は上記実施形態に限定されるものではない。例えば、前記実施形態では入力音声の有声音のみを対象として特徴ベクトル及び代表ベクトルを生成するようにした。しかしこれに限らず、さらに母音のみを対象として特徴ベクトル及び代表ベクトルを生成するようにしてもよい。すなわち、通常の音声信号中には、母音が必ず含まれ、この母音は個人性を顕著に表す要素である。よって、母音の時間定常性を利用して特徴ベクトル中で母音に該当するものを抽出し、この母音ごとの特徴ベクトルを量子化したコードベクトルの平均を特徴代表ベクトルとする。このようにすると、発話偏りによる特徴代表ベクトルのずれを低減することができる。

40

【0046】

また、特徴ベクトルはLPCケプストラムの他にMFCCやスペクトル分析から生成される特徴ベクトル、ならびに照合モデルもVQのみならずGMMなどの他の照合モデルにも適用して実施できる。

その他、コードベクトル及び話者の特徴ベクトル集合の変換処理からVQ歪みの比較処理までの一連の照合処理の繰り返し回数や、変換式の構成、パラメータの値等についても、この発明の要旨を逸脱しない範囲で種々変形して実施できる。

【0047】

要するにこの発明は、上記実施形態そのままに限定されるものではなく、実施段階では

50

その要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態に亘る構成要素を適宜組み合わせてもよい。

【図面の簡単な説明】

【0048】

【図1】この発明に係わる話者照合装置の一実施形態を示す機能ブロック図。

【図2】図1に示した話者照合装置の音声特徴生成処理部の構成を示す機能ブロック図。

【図3】図1に示した話者照合装置の話者コードブックデータベースに記憶される情報要素の一例を示す図。

10

【図4】図1に示した話者照合装置の変換係数・しきい値データベースに記憶される情報要素の一例を示す図。

【図5】変換補正前の特徴ベクトル空間の一例を示す図。

【図6】変換補正後の特徴ベクトル空間の一例を示す図。

【図7】話者Xのベクトル次元に対する特徴代表ベクトルの変化を示す図。

【図8】話者Yのベクトル次元に対する特徴代表ベクトルの変化を示す図。

【図9】ベクトル次元に対する特徴代表ベクトルの話者間のばらつきを示す図。

【図10】変換補正無し、ベクトル次元一定、ベクトル次元を3パターンに可変設定した場合の誤認識率の変化を示す図。

【図11】特徴ベクトル分布とノイズ分布との関係の一例を示す図。

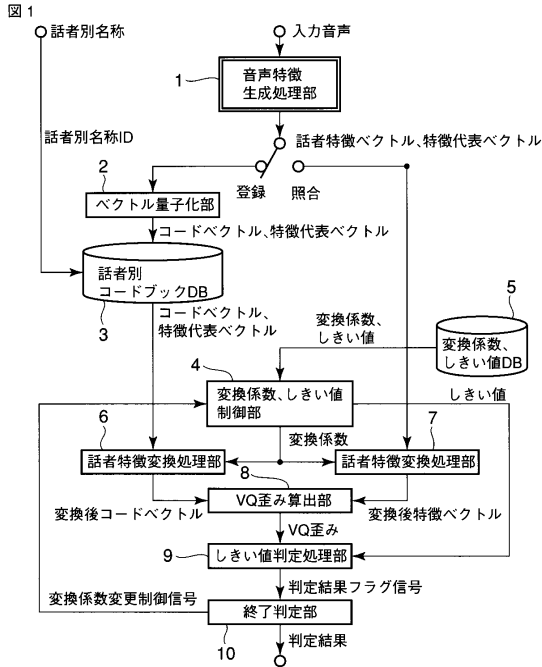
20

【符号の説明】

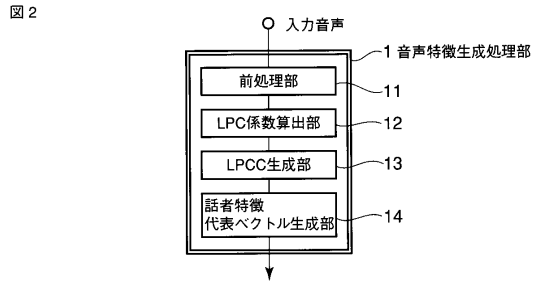
【0049】

1...音声特徴生成処理部、2...ベクトル量子化部、3...話者別コードブックデータベース、4...変換係数・しきい値制御部、5...変換係数・しきい値データベース、6,7...話者特徴変換処理部、8...VQ歪み算出部、9...しきい値判定処理部、10...終了判定部、11...前処理部、12...LPC係数算出部、13...LPC生成部、14...話者特徴代表ベクトル生成部。

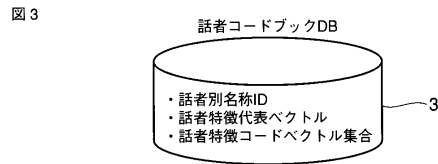
【図1】



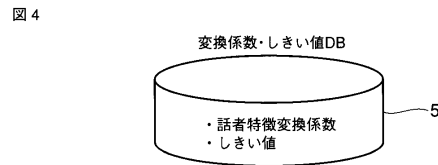
【図2】



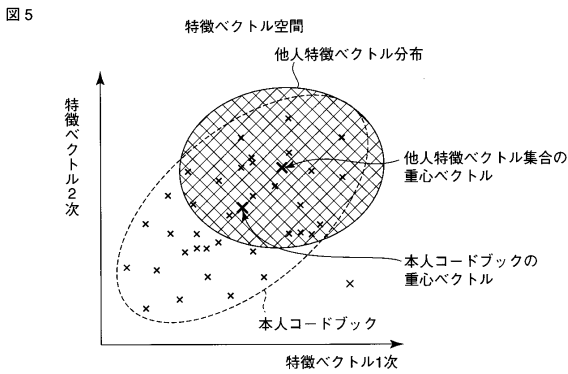
【図3】



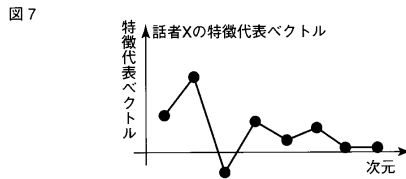
【図4】



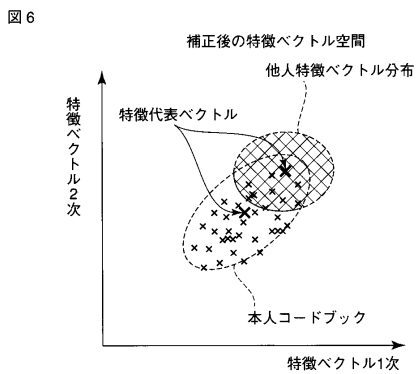
【図5】



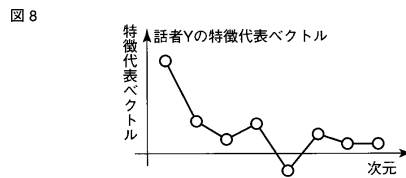
【図7】



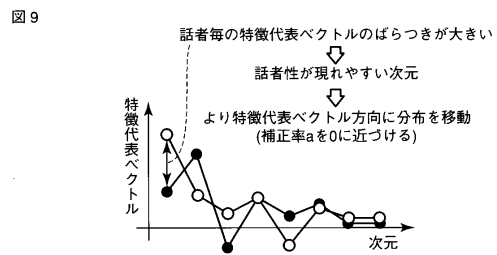
【図6】



【図8】



【図9】

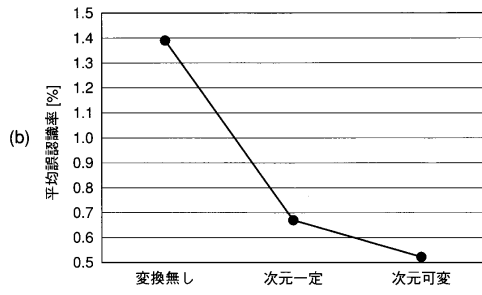


【 図 1 0 】

図 10

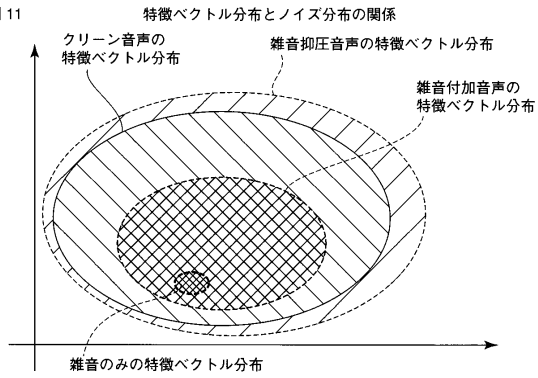
(a)

	変換無し	次元一定	次元可変
平均誤認識率	1.38889	0.67094	0.52564
閾値	0.1386	0.09092	0.08755



【 図 1 1 】

図 11



フロントページの続き

(56)参考文献 特開平03 - 274597 (JP, A)
特開平03 - 054600 (JP, A)
特開平06 - 175683 (JP, A)
特開昭63 - 213899 (JP, A)

(58)調査した分野(Int.Cl., DB名)
G10L 15/00 - 17/00
JSTPlus (JDreamII)