(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4202124号
(P4202124)
(45)発行日　平成20年12月24日(2008.12.24)　(24)登録日　平成20年10月17日(2008.10.17)

(51)Int.Cl.　F I
G10L 15/06 (2006.01)　G10L 15/06 5OON
G10L 15/06 3OOC

請求項の数 26 (全 16 頁)

(21)出願番号	特願2002-512966 (P2002-512966)
(86)(22)出願日	平成13年7月11日 (2001.7.11)
(65)公表番号	特表2004-504641 (P2004-504641A)
(43)公表日	平成16年2月12日 (2004.2.12)
(86)国際出願番号	PCT/US2001/022009
(87)国際公開番号	WO2002/007145
(87)国際公開日	平成14年1月24日 (2002.1.24)
審査請求日	平成16年8月18日 (2004.8.18)
(31)優先権主張番号	09/615,572
(32)優先日	平成12年7月13日 (2000.7.13)
(33)優先権主張国	米国(US)

(73)特許権者　595020643
クゥアルコム・インコーポレイテッド
ＱＵＡＬＣＯＭＭ　ＩＮＣＯＲＰＯＲＡＴＥＤ
アメリカ合衆国、カリフォルニア州　９２１２１－１７１４、サン・ディエゴ、モアハウス・ドライブ　５７７５
(74)代理人　100058479
弁理士　鈴江　武彦
(74)代理人　100091351
弁理士　河野　哲
(74)代理人　100088683
弁理士　中村　誠
(74)代理人　100109830
弁理士　福原　淑弘

最終頁に続く

(54)【発明の名称】話者独立音声認識システムのための音声テンプレートを構成するための方法及び装置

(57)【特許請求の範囲】
【請求項１】
話者独立音声認識システムにおける使用のための音声テンプレートを生成する方法であって、前記方法は、
第１の複数の発話の各発話を細分化して、各発話に対して複数の時間クラスタセグメントを生成し、各時間クラスタセグメントはスペクトル手段によって表され、
前記第１の複数の発話のすべてに対して前記複数のスペクトル手段を量子化して複数のテンプレートベクトルを生成し、
動的時間ゆがみ演算を使用して前記複数のテンプレートベクトルの各１つと第２の複数の発話とを比較して、少なくとも１つの比較結果を生成し、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えるならば、前記第１の複数の発話と前記複数のテンプレートベクトルとのマッチングを行って、最適なマッチングパス結果を生成し、
前記最適なマッチングパス結果に従って時間上で前記第１の複数の発話を分割し、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えないようになるまで前記量子化、前記比較、前記マッチング、前記分割を反復することを具備する方法。
【請求項２】
前記比較は分散測定値を計算することを含む請求項１に記載の方法。
【請求項３】

前記比較は精度測定値を計算することを含む請求項１に記載の方法。
【請求項４】
前記比較はまず分散測定値を計算し、次に前記分散測定値が第１の所定のしきい値を越えないならば、精度測定値を計算することを含む請求項１に記載の方法。
【請求項５】
前記分散測定値が前記第１の所定のしきい値を越えるかあるいは前記精度測定値が第２の所定のしきい値を越えるならば、前記マッチングは前記第１の発話と前記複数のテンプレートベクトルとのマッチングをとることを含む請求の範囲第４項に記載の方法。
【請求項６】
前記マッチングは、動的時間ひずみの計算を実行することを含む請求項１に記載の方法。
【請求項７】
前記マッチング及び前記分割は、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えたときに、Ｋ－平均細分化の手法を用いて実行されることを含む請求項１に記載の方法。
【請求項８】
前記第１の発話の終了点を検出することをさらに含む請求項１に記載の方法。
【請求項９】
話者独立音声認識システムにおける使用のための音声テンプレートを生成するように構成された装置であって、前記装置は、
第１の複数の発話の各発話を細分化して、各発話に対する複数の時間クラスタセグメントを生成する手段であって、各時間クラスタセグメントはスペクトル手段によって表される手段と、
前記第１の複数の発話のすべてに対して前記複数のスペクトル手段を量子化して、複数のテンプレートベクトルを生成する手段と、
前記複数のテンプレートベクトルの各１つと第２の複数の発話とを比較して、少なくとも１つの比較結果を生成するために動的時間ゆがみ演算を使用する手段と、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えるならば、前記第１の複数の発話と前記複数のテンプレートベクトルとのマッチングを行って、最適なマッチングパス結果を生成する手段と、
前記最適なマッチングパス結果に従って時間上で前記第１の複数の発話を分割するための手段と、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えなくなるまで前記量子化、前記比較、前記マッチング及び前記分割を反復する手段と、
を具備する装置。
【請求項１０】
話者独立音声認識システムにおける使用のための音声テンプレートを生成するように構成された装置であって、前記装置は、
第１の複数の発話の各発話を細分化して、各発話に対する複数の時間クラスタセグメントを生成するように構成される細分化ロジックであって、各時間クラスタセグメントはスペクトル手段によって表される細分化ロジックと、
前記細分化ロジックに結合され、前記第１の複数の発話のすべてに対する前記複数のスペクトル手段を量子化して、複数のテンプレートベクトルを生成するように構成された量子化器と、
前記量子化器に結合され、動的時間ゆがみ演算を使用して前記複数のテンプレートベクトルの各１つと、第２の複数の発話とを比較して、少なくとも１つの比較結果を生成するように構成されたコンバージェンス試験器と、
前記量子化器とコンバージェンス試験器とに結合され、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えるならば、前記第１の複数の発話と前記複数のテンプレートベクトルとのマッチングをとって最適なマッチングパス結果を生成し、前記

最適なマッチングパス結果に従って時間上で前記第１の複数の発話を分割するように構成された分割ロジックと
を具備し、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えなくなるまで前記量子化、前記比較、前記マッチング、そして前記分割を反復するように構成されている装置。
【請求項１１】
前記少なくとも１つの比較結果は分散測定値である請求項１０に記載の装置。
【請求項１２】
前記少なくとも１つの比較結果は精度測定値である請求項１０に記載の装置。
【請求項１３】
前記少なくとも１つの比較結果は分散測定値と精度測定値であり、前記コンバージェンス試験器はまず前記分散測定値を計算し、次に、この分散測定値が第１の所定のしきい値を越えないならば、前記精度測定値を計算するように構成された請求項１０に記載の装置。
【請求項１４】
前記マッチングは、前記分散測定値が前記第１の所定のしきい値を越えるかあるいは前記精度測定値が第２の所定のしきい値を越えるならば、前記第１の発話と前記複数のテンプレートベクトルとのマッチングをとることを含む請求項１３に記載の装置。
【請求項１５】
前記分割ロジックは動的時間ゆがみ計算を実行するように構成される請求項１０に記載の装置。
【請求項１６】
前記分割ロジックは、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えたときに実行されるように構成されたＫ－平均細分化ロジックを含む請求項１０に記載の装置。
【請求項１７】
前記細分化ロジックに結合され、前記第１の発話の終了点を検出するように構成された終了点検出器をさらに含む請求項１０に記載の装置。
【請求項１８】
話者独立音声認識システムにおける使用のための音声テンプレートを生成するように構成された装置であって、前記装置は、
プロセッサと、
前記プロセッサに結合され、前記プロセッサによって実行可能な一組の命令を含み、各発話に対する複数の時間クラスタセグメントを生成するために第１の複数の発話の各発話を分割し、各時間クラスタセグメントは手段によって表され、複数のテンプレートベクトルを生成するために前記第１の複数の発話のすべてに対する前記複数のスペクトル手段を量子化し、少なくとも１つの比較結果を生成するために動的時間ゆがみ演算を使用して前記複数のテンプレートベクトルの各１つと第２の複数の発話とを比較し、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えるならば、最適なマッチングパス結果を生成するために、前記第１の複数の発話と前記複数のテンプレートベクトルとのマッチングをとり、前記最適なマッチングパス結果に従って時間において前記第１の複数の発話を分割し、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えなくなるまで前記量子化、前記比較、前記マッチングそして前記分割を反復する記憶媒体と、
を具備する装置。
【請求項１９】
前記少なくとも１つの比較結果は分散測定値である請求項１８に記載の装置。
【請求項２０】
前記少なくとも１つの比較結果は精度測定値である請求項１８に記載の装置。

【請求項２１】
前記少なくとも１つの比較結果は、分散測定値と精度測定値であり、前記一組の命令は、前記分散測定値をまず計算し、次に、前記分散測定値が第１の所定のしきい値を越えないならば、前記精度測定値を計算するために前記プロセッサによって実行可能である請求項１８に記載の装置。
【請求項２２】
前記一組の命令はさらに、前記分散測定値が前記第１の所定のしきい値を越えるかあるいは前記精度測定値が第２の所定のしきい値を越えるならば、前記第１の発話と前記複数のテンプレートベクトルとのマッチングをとるために前記プロセッサによって実行可能である請求項２１に記載の装置。
【請求項２３】
前記一組の命令は、動的時間ゆがみ計算を実行することによって、前記第１の発話に一致するように構成された分割ロジックと前記複数のテンプレートベクトルとのマッチングをとるために、前記プロセッサによって実行可能である請求項１８に記載の装置。
【請求項２４】
前記一組の命令は、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えたときにＫ－平均音声細分化計算を実行することによって、前記第１の発話を分割するために、前記プロセッサによって実行可能である請求項１８に記載の装置。
【請求項２５】
前記一組の命令はさらに、前記第１の発話の終了点を検出するために前記プロセッサによって実行可能である請求項１８に記載の装置。
【請求項２６】
各発話に対する複数の時間クラスタセグメントを生成するために第１の複数の発話の各発話を分割し、各時間クラスタセグメントはスペクトル手段によって表され、
複数のテンプレートベクトルを生成するために前記第１の複数の発話のすべてに対して前記複数のスペクトル手段を量子化し、
少なくとも１つの比較結果を生成するために動的時間ゆがみ演算を使用して前記複数のテンプレートベクトルの各１つと第２の複数の発話とを比較し、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えるならば、前記第１の複数の発話と前記複数のテンプレートベクトルとのマッチングをとって最適なマッチングパス結果を生成し、
前記最適なマッチングパス結果に従って時間上で前記第１の複数の発話を分割し、
前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えなくなるまで、前記量子化、前記比較、前記マッチングそして前記分割を反復するために、
プロセッサによって実行可能な一組の命令を含むプロセッサ読取可能な媒体。
【発明の詳細な説明】
【０００１】
発明の背景
１．発明の分野
本発明は概して通信の分野に関し、より詳細には、話者独立音声認識システムのための音声テンプレートに関する。
【０００２】
２．背景
音声認識（ＶＲ）は、ユーザまたはユーザ発声コマンドを認識して機械とのヒューマンインタフェースを確立するための、シミュレートされた知能を当該機械に与える最も重要な技術の１つを代表するものである。音響音声信号から言語メッセージを再生するための技術を使用するシステムは音声認識装置と呼ばれる。“音声認識装置”の術語は、ここでは、概して話者インタフェース駆動による装置を意味するのに使用される。音声認識装置は概して、音響プロセッサとワードデコーダとを備える。音響プロセッサは、到来する生の音声のＶＲを行うのに必要な、一連の情報保持特徴またはベクトルを抽出する。ワード

デコーダは、入力音声に対応する一連の言語ワードなどの意味のある所望の出力フォーマットを得るために、前記一連の特徴あるいはベクトルを復号する。

【０００３】

音響プロセッサは、音声認識装置内のフロントエンド音声解析サブシステムを代表する。入力音声信号に応答して、音響プロセッサは、時変音声信号を特徴付けるための適切な表示を提供する。音響プロセッサは、背景ノイズ、チャネルひずみ、話者特性、そして話しかたなどの不要な情報を捨てる。効率の良い音響処理は、音声認識に強化された音響識別力を与える。この点において、解析するべき有益な特性は、短時間スペクトル包絡線である。短時間スペクトル包絡線を特徴付けるための２つの通常使用されるスペクトル解析技術は、線形予測符号化（ＬＰＣ）及びフィルタバンクに基づくスペクトルモデル化である。一般的なＬＰＣ技術は、本発明の譲受人に譲渡され、ここに参照として組み込まれている米国特許第５４１４７９６号及び、ここにその全体が参照として組み込まれているL.B. Rabiner&R.W.Schafer、音声信号のディジタル処理、396-453(1978)に開示されている。

【０００４】

ＶＲ（一般的に音声認識とも呼ばれる）の使用が安全性の観点から重要になってきている。例えば、ＶＲは、ワイヤレス電話キーパッド上の釦を押圧するというマニュアル作業に取って代わるのに使用される。このことは、自動車の運転中に電話をかけたいときに特に重要である。ＶＲなしに電話を使用する場合、運転者は一方の手をハンドルから離して電話呼をダイヤルするために釦を押している間電話キーパッドを見なければならない。これらの行為は、自動車事故の可能性を増大させる。音声駆動による電話（すなわち、音声認識用に設計された電話）は、運転者が道路を連続的に眺めながら電話をかけることを可能にする。さらに、ハンドフリーなカーキットシステムは、運転者が電話開始の間両手をハンドルに維持することを可能にする。

【０００５】

音声認識装置は、話者依存の装置と話者独立の装置とに区別される。話者依存装置はより知られているが、特定のユーザからのコマンドを認識するように学習する。これとは対照的に、話者独立装置は任意のユーザからの音声コマンドを受け入れることができる。所定のＶＲシステムのパフォーマンスを増大するために、話者依存であるか話者独立であるかにかかわらず、システムに有効なパラメータを備えさせるために学習が必要である。言い換えると、システムはそれが最適に機能できるようになる前に学習する必要がある。

【０００６】

話者依存ＶＲ装置は概して、トレーニング段階と認識段階の２つの段階において動作する。トレーニング段階において、ＶＲシステムはシステムのボキャブラリ内のワードのそれぞれを一度あるいは二度（通常は二度）話すことをユーザに促し、これにより、システムは特定のワードまたはフレーズに対するユーザ音声の特徴を学ぶことができる。ハンドフリーカーキットに対する例示的なボキャブラリは、キーパッド上の数字、“発呼”、“送信”、“ダイヤル”、“キャンセル”、“クリア”、“追加”、“削除”、“ヒストリ”、“プログラム”、“イエス”、“ノー”のキーワード、そして、通常、仕事仲間、友達あるいは家族のメンバ、と呼ばれる特定の数の人の名前を含む。学習がいったん完了したならば、ユーザは学習されたキーワードを話すことによって認識段階において発呼を開始することができ、ＶＲ装置は話された音声を（テンプレートとして記憶されている）予め学習された音声と比較し、最善のマッチングを行うことによって認識する。例えば、“ジョン”という名前が学習された名前の１つであるならば、ユーザは“ジョンを呼べ”というフレーズを発音することによってジョンに対する発呼を開始する。ＶＲシステムは“呼べ”及び“ジョン”というワードを認識し、ユーザが予めジョンの電話番号として入力した番号をダイヤルする。

【０００７】

話者独立ＶＲ装置は、所定のサイズの予め記録されたボキャブラリ（例えば、ある種の制御ワード、０から９までの数、イエス及びノー）を含む学習テンプレートを使用する。多数の話者（例えば１００）がボキャブラリ内の各ワードを発音して記録しなければならな

い。

【０００８】

従来、話者独立ＶＲテンプレートは、第１の組の話者（概して１００の話者）により話されたワードを含む試験データベースと、第2の組の話者（第１の組と同じ程度の話者）により話された同じワードを含む学習データベースとを比較することによって構成される。一人のユーザにより話された１つのワードは概して発話(utterance)と呼ばれる。トレーニングデータベースの各発話はまず正規化され、試験データベースの発話によりコンバージェンス（convergence）が試験される前に、量子化（概して既知の技術に従ったベクトル量子化）される。しかしながら、時間正規化技術は、以前のフレームとは最大の相異をもつ個々のフレーム（発話の周期的セグメント）からのみ獲得される情報に依存する。所定の発話における情報の多くを使用する話者独立ＶＲテンプレートを構築するための方法を提供することが好ましい。発話のタイプに基づいて話者独立ＶＲテンプレートを構築するための従来の技術の精度またはコンバージェンスをさらに増大させることが望ましい。すなわち、増大した精度を提供するとともに発話における大量の情報を使用する話者独立音声認識テンプレートを構築する方法に対する要求が存在する。

【０００９】

発明の要約

本発明は、増大した精度を提供するとともに発話における大量の情報を使用する話者独立音声認識テンプレートを構築する方法に関している。すなわち、本発明の一側面において、話者独立音声認識システムにおける使用のための音声テンプレートを生成する方法が提供される。この方法は好ましくは、第１の複数の発話の各発話を分割して、各発話に対する複数の時間クラスタセグメントを生成し、各時間クラスタセグメントはスペクトル手段により表され、前記第１の複数の発話のすべてに対する前記複数のスペクトル手段を量子化して複数のテンプレートベクトルを生成し、前記複数のテンプレートベクトルの各々と第２の複数の発話とを比較して少なくとも１つの比較結果を生成し、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値を越えるならば、前記第１の複数の発話と前記複数のテンプレートベクトルとのマッチングを行って、最適なマッチングパス結果を生成し、前記最適なマッチングパス結果に従って、時間上で前記第１の複数の発話を分割し、前記少なくとも１つの比較結果が少なくとも１つの所定のしきい値のどれをも越えなくなるまで前記量子化、前記比較、前記マッチング、前記分割を反復する。

【００１０】

好ましい実施形態の詳細な説明

一実施形態によれば、図１に示すように、話者独立音声認識のための音声テンプレートを構築して実行するためのシステム１０は、話者独立テンプレート構築サブシステム１２及び音声認識サブシステム１４を含む。話者独立テンプレート構築サブシステム１２は、音声認識サブシステム１４に結合される。

【００１１】

話者独立音声テンプレートは、図４～図６を参照して後述するように、話者独立テンプレート構築サブシステム１２により構築される。テンプレートは、図２－３を参照して後述するように、ユーザからの入力音声を認識するにおいて使用される音声認識サブシステム１４に供給される。

【００１２】

一実施形態によれば、図２に示すように、音声認識サブシステム１００は、アナログ／ディジタル変換器（Ａ／Ｄ）１０２、フロントエンド音響プロセッサ１０４、特徴抽出器１０６、音声テンプレートデータベース１０８、パターン比較ロジック１１０、決定ロジック１１２を含む。特定の実施形態において、音響プロセッサ１０４及び特徴抽出器１０６は１つのデバイス例えばパラメータ抽出器として実現される。一実施形態において音響プロセッサ１０４は周波数解析モジュール１１４を含む。一実施形態において特徴抽出器１０６は、終了点検出器１１６、時間クラスタリング音声細分化モジュール１１８、音声レベル正規化器１２０を含む。

【００１３】
Ａ／Ｄ１０２は、音響プロセッサ１０４に結合されている。音響プロセッサ１０４は、特徴抽出器１０６に結合される。一実施形態において、特徴抽出器１０６内で、終了点検出器１１６は、振幅量子化器１２０に結合された時間クラスタリング音声細分化モジュール１１８に結合されている。特徴抽出器１０６はパターン比較ロジック１１０に結合されている。パターン比較ロジック１１０はテンプレートデータベース１０８及び決定ロジック１１２に結合されている。
【００１４】
音声認識サブシステム１００は例えばワイヤレス電話またはハンドフリーカーキット内に配置されている。（図示せぬ）ユーザがワードまたはフレーズを発音すると音声信号を生成する。音声信号は、従来の変換器（図示せぬ）によって電気音声信号ｓ（ｔ）に変換される。音声信号ｓ（ｔ）はＡ／Ｄ１０２に供給されて当該音声信号が例えばパルス符号変調（ＰＣＭ）、Ａ－law、あるいはμ－lawなどの既知のサンプリング方法に従ってディジタル音声サンプルｓ（ｎ）に変換される。
【００１５】
音声サンプルｓ（ｎ）はパラメータ決定のために音響プロセッサ１０４に供給される。音響プロセッサ１０４は入力音声信号ｓ（ｔ）の特性をまねた一連のパラメータを生成する。パラメータは、例えば音声符号器符号化、離散フーリエ変換（ＤＦＴ）に基づくケプストラム係数（例えば高速フーリエ変換（ＦＦＴ）に基づくケプストラム係数）、線形予測係数（ＬＰＣ）あるいはバークスケール解析、を含む既知の音声パラメータ決定技術に従って決定される。このような方法は前記した米国特許第５４１４７９６及びLawrence Rabiner & Biing-Hwang Juang、音声認識の基本（１９９３）に開示されている。パラメータの組は好ましくはフレーム（周期的フレームに分割される）に基づく。音響プロセッサ１０４はディジタルシグナルプロセッサ（ＤＳＰ）として実現される。ＤＳＰは音声符号器を含む。一方、音響プロセッサ１０４は音声符号器として実現される。
【００１６】
パラメータの各フレームは特徴抽出器１０６に供給される。特徴抽出器１０６において終了点検出器１１６は、発話（すなわちワード）の終了点を検出するために抽出されたパラメータを使用する。一実施形態において終了点検出は、米国特許出願第０９／２４６４１４号（１９９９年2月8日出願、名称：ノイズの存在下での音声の正確な終了点のための方法及び装置、この米国出願は、本発明の譲受人に譲渡され、参考としてここにその全体が組み込まれている。この出願は米国特許第６３２４５０９号として権利化された）に開示されている。この技術に従って、発話は第１の開始点及び発話の第１の終了点を決定するために、例えば信号対ノイズ比（ＳＮＲ）しきい値などの第１のしきい値と比較される。次に、前記第１の開始点に先立つ発話の部分が第２のＳＮＲしきい値と比較されて発話の第２の開始点が決定される。第１の終了点に続く発話の部分は次に第２のＳＮＲしきい値と比較されて前記発話の第２の終了点が決定される。第１及び第２のＳＮＲしきい値は好ましくは周期的に再計算され、第１のＳＮＲしきい値は好ましくは第２のＳＮＲしきい値を越える。
【００１７】
検出された発話に対する周波数領域パラメータのフレームは、時間クラスタリング音声細分化モジュール１１８に供給されて、一実施形態に従って、米国特許出願第０９／２２５８９１号（１９９９年1月4日出願、名称：音声信号の細分化及び認識のためのシステム及び方法、この米国出願は、本発明の譲受人に譲渡され、参考としてここにその全体が組み込まれている。この出願は米国特許第６３２４５０９号として権利化された）に記載された圧縮技術を実行する。この技術に従って、周波数領域パラメータにおける各音声フレームは、音声フレームに関連した少なくとも１つのスペクトル値によって表現される。次に、隣接フレームの各対に対してスペクトル相異値が決定される。隣接フレームの各対間に初期的クラスタ境界が設定され、当該パラメータにクラスタが生成され、分散値が各クラスタに割り当てられる。分散値は好ましくは、所定のスペクトル相異値の１つに等し

い。次に、複数のクラスタマージパラメータが計算される。クラスタマージパラメータの各々は一対の隣接クラスタに関連する。最小クラスタマージパラメータは複数のクラスタマージパラメータから選択される。マージされたパラメータは次に、最小クラスタマージパラメータに関連したクラスタ間のクラスタ境界を相殺して、マージされた分散値をマージされたクラスタに割り当てることによって形成される。マージされた分散値は最小クラスタマージパラメータに関連したクラスタに割り当てられた分散値を表す。このプロセスは好ましくは複数のマージされたクラスタを形成するために反復され、好ましくは複数のマージされたクラスタに従って分割された音声信号が形成される。

【００１８】

時間クラスタリング音声細分化モジュール１１８は、例えば時間正規化モジュールなどの他の装置により置き換えられる。しかしながら、時間クラスタリング音声細分化モジュール１１８は、以前のフレームと比較したときに最小の相異をもつフレームをクラスタにマージし、個々のフレームの代わりに算術平均（mean average）を使用するので、時間クラスタリング音声細分化モジュール１１８は、処理された発話におけるより多くの情報を使用する。時間クラスタリング音声細分化モジュール１１８は好ましくは、当業界で知られており以下に説明する動的時間ゆがみ（ＤＴＷ）モデルを使用するパターン比較ロジック１１０に関連して使用される。

【００１９】

クラスタ手段は音声レベル正規化器１２０に供給される。一実施形態において、音声レベル正規化器１２０は、各クラスタにチャネルあたり平均２ビット（すなわち、周波数ごとに２つのビット）を割り当てることによって音声振幅を量子化する。ケプストラム係数が抽出される他の実施形態において、音声レベル正規化器１２０は、当業者により理解されるように、クラスタ手段を量子化するのに使用されない。音声レベル正規化器１２０により生成された出力は特徴抽出器１０６によってパターン比較ロジック１１０に供給される。

【００２０】

音声認識サブシステム１００のボキャブラリワードのすべてに対するテンプレートの組は、テンプレートデータベース１０８内に永久的に記憶される。テンプレートの組は好ましくは、以下に説明するような話者独立テンプレート構築サブシステムにより構築される一組の話者独立テンプレートである。テンプレートデータベース１０８は好ましくは、例えばフラッシュメモリなどの任意の従来の形態の不揮発性記憶媒体として実現される。このことは、音声認識サブシステム１００に対する電源がＯＦＦされたときにテンプレートがテンプレートデータベース１０８内に残ることを可能にする。

【００２１】

パターン比較ロジック１１０は、特徴抽出器１０６からのベクトルをテンプレートデータベース１０８内に記憶されたすべてのテンプレートと比較する。比較結果すなわちベクトルとテンプレートデータベース１０８内に記憶されたすべてのテンプレート間の距離は、決定ロジック１１２に供給される。決定ロジック１１２は、所定のマッチングしきい値内でＮ個の最も近いマッチ（合致）を選択する従来の“Ｎ－ベスト”選択アルゴリズムを使用する。ユーザは次にどちらの選択が意図されたかについて質問される。決定ロジック１１２の出力はボキャブラリ内のどのワードが話されたかに関する決定である。

【００２２】

一実施形態において、パターン比較ロジック１１０及び決定ロジック１１２はコンバージェンスを試験するためにＤＴＷ技術を使用する。ＤＴＷ技術は当業界で知られており、その全体がここに参照として組み込まれた、Lawrence Rabiner & Biing-Hwang Juang 音声認識の基本200-238(1993)に記載されている。ＤＴＷ技術に従って、試験するべき発話の時間シーケンスをテンプレートデータベース１０８内に記憶された各発話に対する時間シーケンスに対してプロットすることによって、トレリスが形成される。次に、試験されている発話は、ポイント（例えば各１０ｍｓ）ごとに一度に１発話だけテンプレートデータベース１０８内の各発話と比較される。テンプレートデータベース１０８における各発

話に対して、試験されている発話は時間上で調整すなわち“ゆがみが加えられ”、テンプレートデータベース１０８における発話に最も近い可能なマッチが達成されるまで、特定の点で圧縮あるいは伸張される。時間の各点で２つの発話が比較されて、マッチがその点（零コスト）で宣言されるかあるいは、ミスマッチが宣言される。特定の点でのミスマッチがあった場合には、試験されるべき発話が圧縮、伸張あるいは必要に応じてミスマッチされる。当該プロセスは２つの発話が互いに完全に比較されるまで反復される。多数（概して数千）の別々に調整された発話が可能である。最も低い（すなわち、最も少ない数の圧縮及び／又は拡大及び／又はミスマッチを要求する）コスト関数が選択される。ビタビ復号アルゴリズムの場合と同様に、最も低い全コストをもつ経路を決定するために、好ましくはテンプレートデータベース１０８における発話の各ポイントから後方を見ることによって選択が実行される。このことは、個別に調整された発話のすべての可能性を生成する“強引な”方法に頼ることなしに最も低いコスト（すなわち最も近いマッチ）の調整された発話を決定することを可能にする。次に、テンプレートデータベース１０８における発話のすべてに対する最も低いコストの調整された発話は次に比較されて、最も低いコストをもつ１つが、試験された発話に最も近くマッチした記憶された発話として選択される。

【００２３】

パターン比較ロジック１１０及び決定ロジック１１２は好ましくはマイクロプロセッサとして実現される。音声認識サブシステム１００は例えばＡＳＩＣである。音声認識サブシステム１００の認識精度は、音声認識サブシステム１００がどれぐらい正確にボキャブラリ内の話されたワードまたはフレーズを認識するかについての測定基準となる。例えば、９５％の認識精度は、音声認識サブシステム１００がボキャブラリ内のワードを１００のうち９５回正確に認識することを示す。

【００２４】

一実施形態によれば、音声認識サブシステム（図示せず）は、音声認識サブシステムに対する音声入力を認識するために図３のフローチャートに示されたアルゴリズムステップを実行する。ステップ２００において、入力音声は音声認識サブシステムに供給される。次に制御フローはステップ２０２に進む。ステップ２０２において発話の終了点が検出される。特定の実施形態において、発話の終了点は図２を参照して上記したように上記の米国特許出願第０９／２４６４１４号に記載された技術に従って検出される。制御フローは次にステップ２０４に進む。

【００２５】

ステップ２０４において時間クラスタリング音声細分化が抽出された発話に関して実行される。特定の実施形態において、使用される時間クラスタリング音声細分化技術は、図２を参照して上記したように、上記した米国特許出願第０９／２２５８９１号（米国特許第６２７８９７２号として権利化された）に記載された技術である。制御フローは次にステップ２０８に進む。ステップ２０６においてステップ２０４において生成された音声クラスタ手段とのマッチのために話者独立テンプレートが提供される。話者独立テンプレートは好ましくは図４－６を参照して以下に述べる技術に従って構成される。制御フローは次にステップ２０８に進む。ステップ２０８において特定の発話に対するクラスタとすべての話者独立テンプレート間でＤＴＷマッチが実行され、最も近いマッチングテンプレートが認識された発話として選択される。特定の実施形態においてＤＴＷマッチがLawrence Rabiner & Biing-Hwang Juang 音声認識の基本200-238(1993)に記載され、図２を参照して上記した技術に従って実行される。時間クラスタリング音声細分化以外の方法がステップ２０４において実行される。そのような方法は、例えば時間正規化を含む。

【００２６】

一実施形態によれば、図４に示すように、話者独立テンプレート構築サブシステム３００は、プロセッサ３０２と記憶媒体３０４を含む。プロセッサ１００は好ましくはマイクロプロセッサであるが、他の従来の形態のプロセッサ、専用プロセッサ、ディジタルシグナルプロセッサ（ＤＳＰ）、コントローラ、あるいはステートマシンを含む。プロセッサ３

０２は、フラッシュメモリ、ＥＥＰＲＯＭメモリ、ＲＡＭメモリ、ファームウェア命令を保持するように構成されたＲＯＭメモリ、プロセッサ３０２上で動作するソフトウェアモジュール、あるいは任意の他の従来の形態のメモリとして実現される記憶媒体３０４に結合される。話者独立テンプレート構築サブシステム３００は好ましくは、ユニックスオペレーティングシステム上で動作するコンピュータとして実現される。他の実施形態において、記憶媒体３０４はオンボードＲＡＭメモリあるいはプロセッサ３０２であり、記憶媒体３０４はＡＳＩＣ内に存在する。一実施形態において、プロセッサ３０２は、図６を参照して以下に述べるステップなどのアルゴリズムステップを実行するために記憶媒体３０４により記憶される一連の命令を実行するように構成される。

【００２７】

他の実施形態によれば、図５に示すように、話者独立テンプレート構築サブシステム４００は、終了点検出器４０２、時間クラスタリング音声細分化ロジック４０４、ベクトル量子化器４０６、コンバージェンス試験器４０８、そしてＫ－平均音声細分化ロジック４１０を含む。制御プロセッサ（図示せず）は好ましくは、話者独立テンプレート構築サブシステム４００が実行する反復の回数を制御するために使用される。

【００２８】

終了点検出器４０２は時間クラスタリング音声細分化ロジック４０４に接続される。時間クラスタリング音声細分化ロジック４０４は、ベクトル量子化器４０６に結合される。ベクトル量子化器４０６はコンバージェンス試験器４０８及びＫ－平均音声細分化ロジック４１０に結合される。制御プロセッサは好ましくは、制御バス（図示せぬ）を介して終了点検出器４０２、時間クラスタリング音声細分化ロジック４０４、ベクトル量子化器４０６、コンバージェンス試験器４０８、そしてＫ－平均音声細分化ロジック４１０に結合される。

【００２９】

学習するべき発話の学習サンプルS_x（ｎ）はフレームで終了点検出器４０２に供給される。学習サンプルは好ましくは、学習するべき発話が記憶されるトレーニングデータベース（図示せぬ）から供給される。一実施形態において、トレーニングデータベースは１００のワードを含み、各ワードは、全部で１００００の記憶された発話に対して１００の異なる話者により話される。終了点検出器４０２は発話の開始点と終了点とを検出する。一実施形態において、終了点検出器４０２は、上記した米国特許出願第０９／２４６４１４号及び図２を参照して上記した技術に従って動作する。

【００３０】

終了点検出器４０２は、検出された発話を時間クラスタリング音声細分化ロジック４０４に供給する。時間クラスタリング音声細分化ロジック４０４は検出された発話に関して圧縮アルゴリズムを実行する。一実施形態において、時間クラスタリング音声細分化ロジック４０４は、上記した米国特許出願第０９／２２５８９１号及び図２を参照して上記した技術に従って動作する。一実施形態において、時間クラスタリング音声細分化ロジック４０４は検出された発話を２０のセグメントに分割するが、各セグメントはクラスタ手段を含む。

【００３１】

時間クラスタリング音声細分化ロジック４０４は、所定のワードに対するトレーニング用発話のすべてに対するクラスタ手段をベクトル量子化器４０６に供給する。ベクトル量子化器４０６は、発話に対する（すなわち同じワードのすべての話者に対する）クラスタ手段をベクトル量子化し、最終的なベクトルを当該発話に対する潜在的な話者独立（ＳＩ）テンプレートとしてコンバージェンス試験器４０８に供給する。ベクトル量子化器４０６は好ましくは、種々の既知のベクトル量子化（ＶＱ）技術に従って動作する。種々のＶＱ技術は、例えば、Gersho & R.M. Gray,ベクトル量子化及び信号圧縮（１９９２）に記載されている。特定の実施形態において、ベクトル量子化器４０６は、４つのクラスタベクトルを生成する。すなわち、例えば、各セグメントは、各セグメントを４つのクラスタとして表すベクトル量子化器４０６に直列に供給される。各クラスタは特定のワードに対す

る各話者を表わし、ワードあたり複数のクラスタが存在する。一実施形態によれば、テンプレートあたり８０（４つのクラスタ×２０セグメント）のベクトルが存在する。

【００３２】

コンバージェンス試験器４０８は、潜在的なＳＩテンプレートと試験するべき発話の試験サンプルS_y（ｎ）とを比較する。試験サンプルはフレームの形態でコンバージェンス試験器４０８に供給される。試験用サンプルは好ましくは、試験するべき発話が記憶される試験データベース（図示せぬ）から供給される。一実施形態において、試験用データベースは１００のワードを含み、各ワードは全部で１００００の記憶された発話に対して１００の異なる話者により話される。ワードは好ましくは、トレーニングデータベースに含まれている同じワードであるが、１００の異なる話者によって話される。コンバージェンス試験器４０８は、訓練すべき発話に対する潜在的なＳＩテンプレートと試験するべき発話に対するサンプルとを比較する。一実施形態において、コンバージェンス試験器４０８は、コンバージェンスを試験するためのＤＴＷアルゴリズムを使用するように構成される。使用されるＤＴＷアルゴリズムは好ましくは、Lawrence Rabiner & Biing-Hwang Juang 音声認識の基本200-238(1993)に記載され、図２を参照して上記した技術である。

【００３３】

一実施形態において、コンバージェンス試験器４０８は、前記データベース内のすべてのワードに対する結果の精度と、前記データベースの分散とを潜在的なＳＩテンプレートで解析するように構成される。分散がまずチェックされ、この分散が所定のしきい値以下であるならば、精度がチェックされる。分散は好ましくはセグメントごとに計算され、次に加算されて全体の分散値が取得される。特定の実施形態において、分散は４つのクラスタのベストマッチ（最善の合致）に対する平均二乗誤差を計算することによって取得される。平均二乗誤差の技術は、当業界で良く知られている。試験データベースからの発話がトレーニングデータベースにより生成された潜在的なＳＩテンプレートと一致する（すなわち認識がデータベース内のすべてのワードに対して正しい）ならば、正確であるとみなされる。

【００３４】

潜在的なＳＩテンプレートは、ベクトル量子化器４０６からＫ－平均音声細分化ロジック４１０に供給される。Ｋ－平均音声細分化ロジック４１０は好ましくはフレームに分割されたトレーニングサンプルを受信する。コンバージェンス試験器４０８がコンバージェンスに対する第１の試験を実行した後に、分散または精度に対する結果は、分散及び精度に対する所定のしきい値以下になる。一実施形態において、分散または精度に対する結果が所定のしきい値以下であるならば、他の反復が実行される。したがって、制御プロセッサは、Ｋ－平均音声細分化ロジック４１０に対してトレーニングサンプルに関してＫ－平均細分化を実行するように命令し、以下に述べるような分割された音声フレームを生成する。Ｋ－平均音声細分化に従って、トレーニングサンプルは好ましくはＤＴＷ技術を用いて潜在的なＳＩテンプレートとマッチングがとられて、図２を参照して上記したような最適なパスを生成する。次にトレーニングサンプルは最適なパスに従って分割される。例えば、トレーニングサンプルの第１の５つのフレームは、潜在的なＳＩテンプレートの第１のフレームとマッチし、トレーニングサンプルの次の３つのフレームは潜在的なＳＩテンプレートの第２のフレームにマッチし、トレーニングサンプルの次の１０のフレームは、潜在的なＳＩテンプレートの第３のフレームにマッチする。この場合、トレーニングサンプルの第１の１つのフレームは１つのフレームに分割され、次の３つのフレームは第２のフレームに分割され、次の１０のフレームは第３のフレームに分割される。一実施形態において、Ｋ－平均音声細分化ロジック４１０は、Lawrence Rabiner & Biing-Hwang Juang 音声認識の基本382-384(1993)（この文献の全体が参照としてここに組み込まれている）に記載された例示的なＫ－平均細分化技術に従ってＫ－平均細分化を実行する。次にＫ－平均音声細分化ロジック４１０はクラスタ手段の更新されたフレームをベクトル量子化器４０６に供給する。ベクトル量子化器４０６は、クラスタ手段をベクトル量子化して（新たな潜在的なＳＩテンプレートを具備する）最終的なベクトルをコンバージェンス試験器

４０８に供給して他のコンバージェンス試験を実行する。当業者ならば、この反復工程が所定のしきい値以上の分散及び精度の結果を得るのに必要なだけ継続されることを認識する。

【００３５】

コンバージェンス試験がパスすると、好ましくは潜在的な（ここでは最終的な）ＳＩテンプレートが図２の音声認識サブシステムなどの音声認識サブシステムにおいて使用される。最終的なＳＩテンプレートは、図２のテンプレートデータベース１０８に記憶されるかあるいは図３のフローチャートのステップ２０６において使用される。

【００３６】

一実施形態において、話者独立テンプレート構築サブシステム（図示せず）は、発話に対する話者独立テンプレートを構築するために図６のフローチャートにおいて例示された方法ステップを実行する。ステップ５００において発話のトレーニングサンプルがトレーニングデータベース（図示せぬ）から入手される。トレーニングデータベースは好ましくは、多数のワード（例えば１００ワード）を含む。各ワードは多数の話者（ワードあたり例えば１００の話者）によって話される。制御フローは次にステップ５０２に進む。

【００３７】

ステップ５０２において、発話を検出するために終了点検出がトレーニングサンプルに関して実行される。一実施形態において終了点検出は、上記した米国特許出願第０９／２４６４１４号及び図２を参照して上記した技術に従って動作する。制御フローは次にステップ５０４に進む。

【００３８】

ステップ５０４において、検出された発話に関して時間クラスタリング音声細分化が実行され、当該発話を複数のセグメントに圧縮する。各セグメントは手段によって表示される。特定の実施形態において、発話は２０のセグメントに分割される。各セグメントはクラスタ手段を含む。一実施形態において、発話は２０のセグメントに圧縮される。各セグメントはクラスタ手段を含む。一実施形態において時間クラスタリング音声細分化は、上記した米国特許出願第０９／２２５８９１号（この出願は米国特許第６２７８９７２号として権利化された）及び図２を参照して上記した技術に従って動作される。制御フローは次にステップ５０６に進む。

【００３９】

ステップ５０６において、同じワードのすべての話者に対するトレーニングサンプルに対するクラスタ手段は、ベクトル量子化される。特定の実施形態において、クラスタ手段は、A.Gersho & R.M. Gray,ベクトル量子化及び信号圧縮(1992)に記載された種々の知られたＶＱ技術の１つに従ってベクトル量子化される。特定の実施形態において、４つのクラスタベクトルが生成される。すなわち、例えば、各セグメントは４つのクラスタとして表される。各クラスタは特定のワードに対する各話者を表し、ワードごとに複数のクラスタが存在する。一実施形態によれば、テンプレートあたり８０のベクトル（４つのクラスタ×２０セグメント）が生成される。制御フローは次にステップ５１０に進む。

【００４０】

ステップ５０８において試験データベース（図示せぬ）から入手した試験サンプルは、コンバージェンスが試験される。試験データベースは好ましくはトレーニングデータベースに含まれる同じワードを含む。各ワードは多数の話者（発話あたり例えば１００の話者）によって話される。

【００４１】

ステップ５１０において、量子化されたベクトルは、コンバージェンスを試験するために潜在的なＳＩテンプレートして試験サンプルと比較される。一実施形態において、コンバージェンス試験はＤＴＷアルゴリズムである。使用されるＤＴＷアルゴリズムは、Lawrence Rabiner & Biing-Hwang Juang、音声認識の基本,200-238（１９９３）及び図２を参照して上記した技術である。

【００４２】

一実施形態において、ステップ５１０のコンバージェンス試験では、データベース内のすべてのワードに対する結果の精度と、潜在的なＳＩサンプルを有するデータベースの分散である。分散がまずチェックされ、この分散が所定のしきい値以下ならば、次に精度がチェックされる。分散は好ましくは、セグメントごとに計算され、次に加算されて全体の分散値が取得される。特定の実施形態において、分散は、４つのクラスタの最善の合致に対する平均二乗誤差を計算することによって取得される。平均二乗誤差の技術は当業界でよく知られている。試験データベースによって生成された潜在的なＳＩテンプレートがトレーニングデータベースからの発話に合致するならば（すなわち、データベース内のすべてのワードに対して認識が正しいならば）、コンバージェンス試験は正確であるとみなされる。制御フローは次にステップ５１２に進む。

【００４３】

ステップ５１２において、分散あるいは精度に対するステップ５１０のコンバージェンス試験の結果が、分散及び精度に対する所定のしきい値以下ならば、他の反復が実行される。したがって、トレーニングサンプルに関してＫ－平均音声細分化が実行される。Ｋ－平均音声細分化はトレーニングサンプルを好ましくはＤＴＷ技術を用いて潜在的なＳＩテンプレートとマッチさせ、図２を参照して上記したような最適なパスを生成する。トレーニングサンプルは次に最適なパスに従って分割される。一実施形態において、Ｋ－平均音声細分化は、Lawrence Rabiner & Biing-Hwang Juang、音声認識の基本,382-384（１９９３）に記載された技術に従って実行される。制御フローは次にステップ５０６に戻って、クラスタ手段の更新されたフレームがベクトル量子化され、ステップ５１０において、試験データベースからのサンプルによってコンバージェンスが（新たな潜在的なＳＩテンプレートとして）試験される。当業者ならば、この反復工程は予め決められたしきい値以上の分散及び精度結果を達成するのに必要なだけ継続されることを認識する。

【００４４】

いったんコンバージェンス試験がパスすると（すなわち、いったんしきい値が達成されると）、潜在的な（ここでは最終的な）ＳＩテンプレートは好ましくは、図２の音声認識サブシステムなどの音声認識サブシステムにおいて使用される。最終的なＳＩテンプレートは、図２のテンプレートデータベース１０８に記憶されるかあるいは図３のフローチャートのステップ２０６において使用される。

【００４５】

すなわち、話者独立音声認識システムに対する音声テンプレートを構築するための新規でかつ改善された方法及び装置が説明された。当業者ならば、データ、命令、コマンド、情報、信号、ビット、シンボルそして上記記述において参照されるチップは好ましくは、電圧、電流、電磁波、磁界あるいは粒子、オプチカルフィールド、またはそれらの任意の組み合わせであることを認識するであろう。当業者ならば、種々の例示的な論理ブロック、モジュール、回路そしてここで開示された実施形態に関連したアルゴリズムステップは、電子ハードウェア、コンピュータソフトウェアあるいは両方の組み合わせとして実現される。種々の例示的な要素、ブロック、モジュール、回路、そしてステップは概してそれらの機能の観点から記載された。機能がハードウェアとして実現されるかソフトウェアとして実現されるかは特定のアプリケーションと全体のシステムに課される設計上の拘束に依存する。熟練した技術者ならば、このような環境の下でハードウェアとソフトウェアとを交換したり、記述された機能を各特定のアプリケーションに対して最適に実行する方法を認識するであろう。例として、種々の例示的なロジカルブロック、モジュール、回路そしてここに開示された実施形態に関連して記述されたアルゴリズムステップは、ディジタルシグナルプロセッサ（ＤＳＰ）、特定用途向け集積回路（ＡＳＩＣ）、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）あるいは他のプログラマブルロジックデバイス、ディスクリートゲートまたはトランスファーゲート、例えばレジスタやＦＩＦＯなどのディスクリートハードウェア要素、一連のファームウェア命令を実行するプロセッサ、任意の従来のプログラマブルソフトウェアモジュール及びプロセッサ、あるいはここに記載された機能を実行するために設計されたそれらの任意の組み合わせ、である。プロセッサは好ま

しくはマイクロプロセッサであるが、変形例として、プロセッサは任意の従来のプロセッサ、コントローラ、マイクロコントローラ、あるいは他のマシンである。ソフトウェアモジュールは、ＲＡＭメモリ、フラッシュメモリ、ＲＯＭメモリ、ＥＥＰＲＯＭメモリ、レジスタ、ハードウェアディスク、リムーバルディスク、ＣＤ－ＲＯＭ、あるいは任意の他の形態の公知の記憶媒体である。例示的なプロセッサは好ましくは記憶媒体に結合され、当該記憶媒体から情報を読み出したり、当該記録媒体に情報を書き込む。あるいは、記憶媒体はプロセッサと一体である。プロセッサ及び記憶媒体はＡＳＩＣ内に配置される。ＡＳＩＣは電話内に配置される。あるいは、プロセッサ及び記憶媒体が電話内に配置される。プロセッサはＤＳＰとマイクロプロセッサとの組み合わせとしてあるいはＤＳＰコアなどに関連する２つのマイクロプロセッサとして実現される。

【００４６】

本発明の好ましい実施形態が示されかつ記述された。しかしながら、本発明の精神あるいは範囲から逸脱することなしにここに開示された実施形態に対して種々の変形を行うことができる。したがって、本発明は次の特許請求の範囲に従う以外には限定されることはない。

【図面の簡単な説明】

【図１】　図１は、話者独立音声認識に対する音声テンプレートを構築して実行するためのシステムのブロック図である。

【図２】　図２は、図１のシステムにおいて使用可能な音声認識サブシステムのブロック図である。

【図３】　図３は、入力音声サンプルを認識するために、図２のサブシステムなどの、音声認識サブシステムによって実行される方法ステップを示すフローチャートである。

【図４】　図４は、図１のシステムにおいて使用可能なテンプレート構築サブシステムのブロック図である。

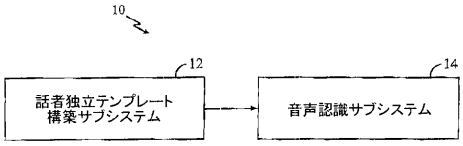
【図５】　図５は、図１のシステムにおいて使用可能なテンプレート構築サブシステムのブロック図である。

【図６】　図６は、音声テンプレートを構築するために、図４のサブシステムや図５のサブシステムなどの、テンプレート構築サブシステムによって実行される方法ステップを示すフローチャートである。

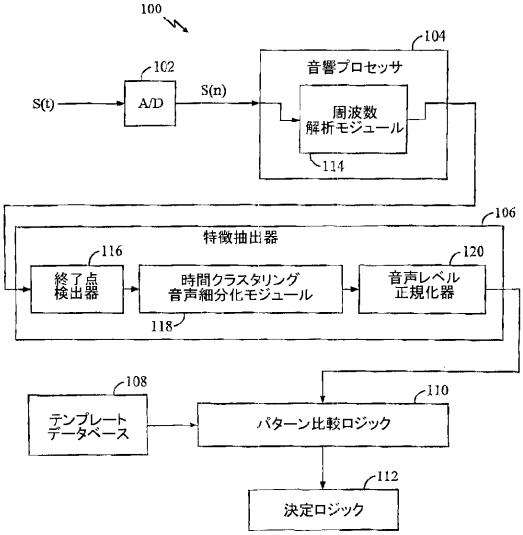
【符号の説明】

１０　システム
１２　話者独立テンプレート構築サブシステム
１４　音声認識サブシステム
１００　音声認識サブシステム
１０２　アナログ／ディジタルコンバータ（Ａ／Ｄ）
１０４　フロントエンドプロセッサ
１０６　特徴抽出器
１０８　音声テンプレートデータベース
１１０　パターン比較ロジック
１１２　決定ロジック
１１４　周波数解析モジュール
１１６　エンドポイントモジュール
１１８　時間クラスタリング音声細分化モジュール
１２０　音声レベル正規化器

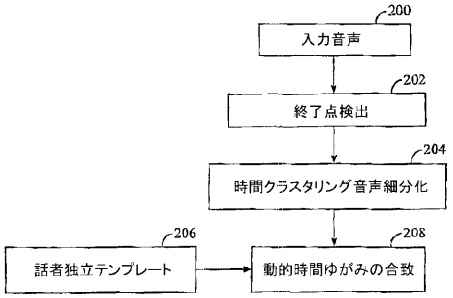
【図１】



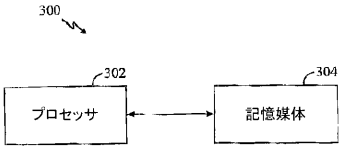
【図２】



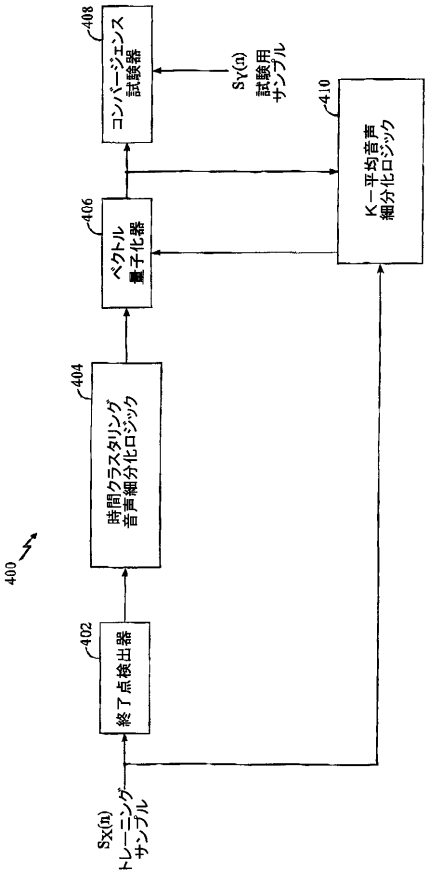
【図３】



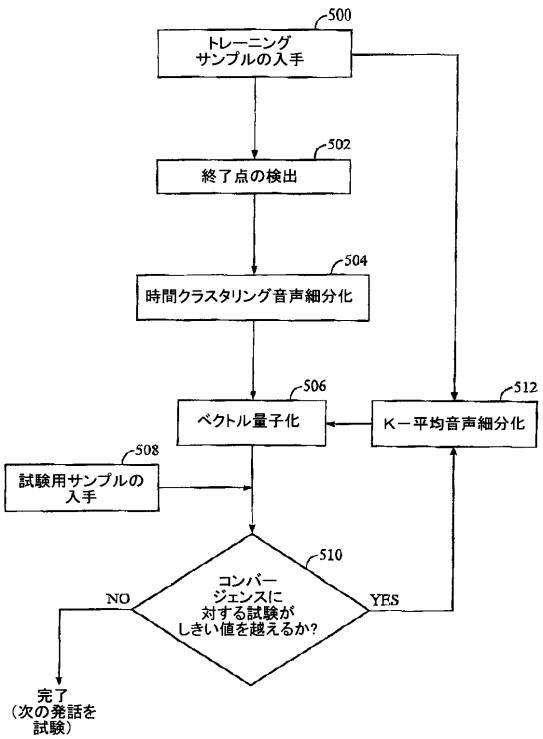
【図４】



【図５】



【図６】



フロントページの続き

(74)代理人 100084618
弁理士 村松 貞男
(74)代理人 100092196
弁理士 橋本 良郎
(72)発明者 ビー、ニン
アメリカ合衆国、カリフォルニア州 ９２１２８ サン・ディエゴ、ブリーズウェイ・ブレイス １４２０９

審査官 渡邊 聡

(56)参考文献 特開昭６２－２１７２９２（ＪＰ，Ａ）
特開昭６２－０７２０００（ＪＰ，Ａ）
特許第２７４９８１１（ＪＰ，Ｂ２）
特開平０７－５１２９６６（ＪＰ，Ａ）
特開平０５－２４１５９１（ＪＰ，Ａ）
特開平１１－２１２５９２（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G10L 15/06