(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2004-536330
(P2004-536330A)
(43)公表日　平成16年12月2日(2004.12.2)

(51)Int.Cl.[7]	F I			テーマコード（参考）
G10L 15/28	G10L	3/00	571A	5D015
G10L 15/06	G10L	3/00	521T	

審査請求 未請求　予備審査請求 有　（全 64 頁）

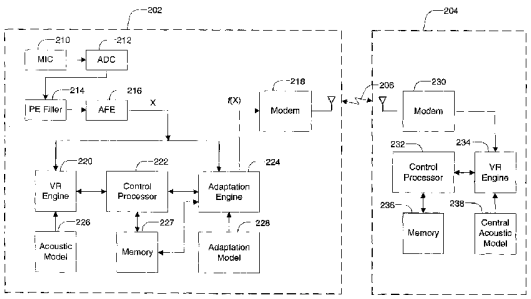
(21)出願番号	特願2002-565298(P2002-565298)
(86)(22)出願日	平成14年1月30日(2002.1.30)
(85)翻訳文提出日	平成15年7月31日(2003.7.31)
(86)国際出願番号	PCT/US2002/003014
(87)国際公開番号	WO2002/065453
(87)国際公開日	平成14年8月22日(2002.8.22)
(31)優先権主張番号	09/773,831
(32)優先日	平成13年1月31日(2001.1.31)
(33)優先権主張国	米国(US)

(71)出願人　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】音響特性ベクトル変形を使用する分散型音声認識システム

(57)【要約】
【課題】音響特性ベクトル変形を使用する分散型音声認識システム
【解決手段】音声認識システムは、話し手に依存しない音響モデル（２３８）に対する音声認識パターンマッチングに先立ち、音響特性ベクトルに話し手に依存する変形関数を適用する。順応エンジン（２２４）は、話し手に依存する特性ベクトル変形関数ｆ()を選択するために、順応モデル（２２８）で音響特性ベクトルＸの組を整合させる。ｆ()は、その後、音響特性ベクトルｆ(Ｘ)の変形された組を形成するために、Ｘに適用される。そして、音声認識は、話し手に依存しない音響モデル（２３８）で変形された音響特性ベクトルｆ(Ｘ)を相関させることにより実行される。
【選択図】図２



【特許請求の範囲】
【請求項１】
音響パターン情報を含む音響モデル、及び
選択された特性ベクトル変形関数を認識するために、音響特性ベクトルで音響パターン情報のパターンマッチングを実行するための順応エンジン、
を具備する音声認識システム。
【請求項２】
請求項１の音声認識システム、ここで、順応エンジンは、変形された音響特性ベクトルの組を生成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用するためにさらに構成される。
【請求項３】
音響モデルで変形された音響特性ベクトルの組をマッチングさせるための音声認識エンジン、をさらに具備する請求項１の音声認識システム。
【請求項４】
選択された特性ベクトル変形関数の性能を評価するため、及び評価に基づいて選択された特性ベクトル変形関数を調整するための制御プロセッサ、をさらに具備する請求項１の音声認識システム。
【請求項５】
特性ベクトル変形関数の組に対応するパラメータの少なくとも１組を記憶するためのメモリ、ここで、選択された特性ベクトル変形関数が特性変形関数の組の構成員である、をさらに具備する請求項１の音声認識システム。
【請求項６】
請求項５の音声認識システム、ここで、メモリが特性ベクトル変形関数の組に対応するパラメータの１以上の組を含む、及びここで、各パラメータの組が特定の話し手に対応する。
【請求項７】
請求項５の音声認識システム、ここで、メモリが特性ベクトル変形関数の組に対応するパラメータの１以上の組を含む、及びここで、各パラメータの組が異なる音響環境に対応する。
【請求項８】
音響パターン情報を含む順応モデル、及び
話し手に依存する特性ベクトル変形関数を認識するため、及び話し手に依存する特性ベクトル変形関数を音響特性ベクトルに適用するために、音響特性ベクトルで音響パターン情報のパターンマッチングを実行するための順応エンジン、
を具備する音声認識システム。
【請求項９】
選択された特性ベクトル変形関数を認識するために音響特性ベクトルを解析するため、及び変形された音響特性ベクトルの組を生成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用するための順応エンジン、
音響モデル、及び
音響モデルで変形された音響特性ベクトルの組をマッチングさせるための音声認識エンジン、
を具備する音声認識システム。
【請求項１０】
順応モデル、ここで、前記順応エンジンが、順応モデル中に記憶された音響パターンで音響特性ベクトルをマッチングすることにより音響特性ベクトルの解析を実行する、をさらに具備する請求項９の音声認識システム。
【請求項１１】
選択された特性ベクトル変形関数の性能を評価するため、及び評価に基づいて選択された特性ベクトル変形関数を調整するための制御プロセッサ、をさらに具備する請求項９の音

声認識システム。
【請求項１２】
特性ベクトル変形関数の組に対応するパラメータの少なくとも１組を記憶するためのメモリ、ここで、選択された特性ベクトル変形関数が特性変形関数の組の構成員である、をさらに具備する請求項９の音声認識システム。
【請求項１３】
請求項１２の音声認識システム、ここで、メモリが特性ベクトル変形関数の組に対応するパラメータの１以上の組を含む、及びここで、各パラメータの組が特定の話し手に対応する。
【請求項１４】
請求項１２の音声認識システム、ここで、メモリが特性ベクトル変形関数の組に対応するパラメータの１以上の組を含む、及びここで、各パラメータの組が異なる音響環境に対応する。
【請求項１５】
音響パターン情報を含む順応モデル、及び
選択された特性ベクトル変形関数を認識するために、音響パターン情報に対する音響特性ベクトルのパターンマッチングを実行するため、及び変形された音響特性ベクトルの組を生成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用するための順応エンジン、
を具備する遠隔局装置。
【請求項１６】
選択された特性ベクトル変形関数の性能を評価するため、及び評価に基づいて選択された特性ベクトル変形関数を調整するための制御プロセッサ、をさらに具備する請求項１５の遠隔局装置。
【請求項１７】
特性ベクトル変形関数の組に対応するパラメータの少なくとも１組を記憶するためのメモリ、ここで、選択された特性ベクトル変形関数が特性変形関数の組の構成員である、をさらに具備する請求項１５の遠隔局装置。
【請求項１８】
請求項１７の遠隔局装置、ここで、メモリが特性ベクトル変形関数の組に対応するパラメータの１以上の組を含む、及びここで、各パラメータの組が特定の話し手に対応する。
【請求項１９】
請求項１７の遠隔局装置、ここで、メモリが特性ベクトル変形関数の組に対応するパラメータの１以上の組を含む、及びここで、各パラメータの組が異なる音響環境に対応する。
【請求項２０】
変形された音響特性ベクトルを通信センタに通信するための通信インターフェース、をさらに具備する請求項１５の遠隔局装置。
【請求項２１】
音響特性ベクトルを受信するための通信インターフェース、ここで、音響特性ベクトルが特性ベクトル変形関数を使用して変形されてきている、
音響パターンを含む音響モデル、ここで、音響モデルが一人の話し手に向けられていない、
音響モデルで変形された音響特性ベクトルの組をマッチングさせるための音声認識エンジン、及び
マッチングに基づいて特性変形関数を評価するための制御プロセッサ、
を具備する音声認識通信センタ装置。
【請求項２２】
少なくとも１の遠隔局から音響特性ベクトル及び話し手認識情報を受信するための通信インターフェース、
話し手に依存する特性ベクトル変形関数パラメータを含むメモリ、及び

音響モデルで音響特性ベクトルのパターンマッチングを実行するため、パターンマッチング及び話し手認識情報に基づいて選択された話し手に依存する特性ベクトル変形関数を認識するため、及び変形された音響特性ベクトルの組を生成するために、音響特性ベクトルに選択された話し手に依存する特性ベクトル変形関数を適用するための順応エンジン、
を具備する音声認識通信センタ装置。
【請求項２３】
請求項２２の音声認識通信センタ装置、ここで、順応エンジンは、中央音響モデルで変形された音響特性ベクトルの組をマッチングさせるための音声認識エンジンをさらに具備する、ここで、中央音響モデルは一人の話し手に向けられていない。
【請求項２４】
中央音響モデル、ここで、中央音響モデルが一人の話し手に向けられていない、及びここで、順応エンジンが中央音響モデルで変形された音響特性ベクトルの組のパターンマッチングを実行するためにさらに構成される、
をさらに具備する請求項２２の音声認識通信センタ装置。
【請求項２５】
中央音響モデルで変形された音響特性ベクトルの組をマッチングさせるための音声認識エンジン、をさらに具備する請求項２２の音声認識通信センタ装置。
【請求項２６】
選択された話し手に依存する特性ベクトル変形関数の性能を評価するため、及び評価に基づいてメモリ中の選択された特性話し手に依存するベクトル変形関数のパラメータを調整するための制御プロセッサ、をさらに具備する請求項２２の音声認識通信センタ装置。
【請求項２７】
音響特性ベクトルを抽出する、
順応モデルで音響特性ベクトルの順応パターンマッチングを実行する、
順応パターンマッチングに基づいて特性ベクトル変形関数を選択する、
変形された音響特性ベクトルの組を形成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用する、及び
音響モデルで変形された音響特性ベクトルの組の音声認識パターンマッチングを実行する、
を具備する音声認識を実行する方法。
【請求項２８】
請求項２７の方法、ここで、特性ベクトル変形関数が特性ベクトル変形関数の話し手に依存する組から選択される。
【請求項２９】
順応パターンマッチングに基づいて特性ベクトル変形関数の話し手に依存する組を変形する、をさらに具備する請求項２８の方法。
【請求項３０】
音声認識パターンマッチングに基づいて話し手に依存する特性ベクトル変形関数の組を変形する、をさらに具備する請求項２８の方法。
【請求項３１】
請求項２７の方法、ここで、特性ベクトル変形関数がある音響環境に特定される。
【請求項３２】
遠隔局において、遠隔局中に記憶された順応モデルで音響特性ベクトルの順応パターンマッチングを実行する、
遠隔局において、遠隔局で記憶された特性ベクトル変形関数情報から特性ベクトル変形関数を選択する、ここで、特性ベクトル変形関数を選択することがパターンマッチングに基づく、
遠隔局において、変形された音響特性ベクトルの組を形成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用する、及び
遠隔局から通信センタへ変形された音響特性ベクトルを送る、

を具備する音声認識を実行する方法。
【請求項３３】
請求項３２の方法、ここで、特性ベクトル変形関数情報が話し手に依存する。
【請求項３４】
順応パターンマッチングに基づいて特性ベクトル変形関数情報を変形する、をさらに具備する請求項３２の方法。
【請求項３５】
通信センタから受信した情報に基づいて遠隔局において特性ベクトル変形関数情報を変形する、をさらに具備する請求項３２の方法。
【請求項３６】
遠隔局において音声認識を実行する、をさらに具備する請求項３２の方法、
ここで、音声認識を実行することが、遠隔局中に記憶された音響モデルで変形された音響特性ベクトルの音声認識パターンマッチングを実行することを具備する。
【請求項３７】
音声認識パターンマッチングに基づいて遠隔局において特性ベクトル変形関数情報を変形する、をさらに具備する請求項３６の方法。
【請求項３８】
請求項３２の方法、ここで、特性ベクトル変形関数情報が環境に依存する。
【請求項３９】
遠隔局において、音響特性ベクトルを抽出する、
遠隔局から通信センタへ音響特性ベクトルを送る、
通信センタにおいて、通信センタ中に記憶された順応モデルで音響特性ベクトルの順応パターンマッチングを実行する、
通信センタにおいて、通信センタ中に記憶されたデータベースから特性ベクトル変形関数を選択する、ここで、順応パターンマッチングに基づいて特性ベクトル変形関数を選択する、
通信センタにおいて、変形された音響特性ベクトルの組を形成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用する、及び
通信センタにおいて、通信センタ中に記憶された音響モデルで変形された音響特性ベクトルの組の音声認識パターンマッチングを実行する、
を具備する、少なくとも１の遠隔局及び１の通信センタを具備するシステムにおいて音声認識を実行する方法。
【請求項４０】
通信センタにおいて、話し手に依存する特性ベクトル変形関数の組を選択する、ここで、選択された特性ベクトル変形関数が特性ベクトル変形関数の話し手に依存する組から選択される、をさらに具備する請求項３９の方法。
【請求項４１】
順応パターンマッチングに基づいて特性ベクトル変形関数の話し手に依存する組を変形する、をさらに具備する請求項４０の方法。
【請求項４２】
音声認識パターンマッチングに基づいて特性ベクトル変形関数の話し手に依存する組を変形する、をさらに具備する請求項４０の方法。
【請求項４３】
遠隔局から通信センタへ話し手認識情報を送る、ここで、話し手認識情報に基づいて特性ベクトル変形関数の話し手に依存する組を選択する、をさらに具備する請求項４０の方法。
【請求項４４】
遠隔局において、未変形の音響特性ベクトルを抽出する、
遠隔局において、遠隔局中に記憶された順応モデルで未変形の音響特性ベクトルの順応パターンマッチングを実行する、

遠隔局において、パターンマッチングに基づいて話し手に依存する特性ベクトル変形関数を選択する、
遠隔局において、変形された音響特性ベクトルを形成するために、音響特性ベクトルに選択された話し手に依存する特性ベクトル変形関数を適用する、
遠隔局から通信センタへ変形された音響特性ベクトルを送る、
通信センタにおいて、通信センタ中に記憶された音響モデルで変形された音響特性ベクトルの組の音声認識パターンマッチングを実行する、
を具備する、少なくとも１の遠隔局及び１の通信センタを具備するシステムにおいて音声認識を実行する方法。
【請求項４５】
順応パターンマッチングに基づいて選択された話し手に依存する特性ベクトル変形関数を変形する、をさらに具備する請求項４４の方法。
【請求項４６】
遠隔局から通信センタへ未変形の音響特性ベクトルを送る、
通信センタにおいて、変形された音響特性ベクトル及び未変形の音響特性ベクトルを使用して選択された話し手に依存する特性ベクトル変形関数を解析する、及び
遠隔局において、解析に基づいて選択された話し手に依存する特性ベクトル変形関数を変形する、
をさらに具備する請求項４４の方法。
【請求項４７】
音声認識パターンマッチングに基づいて特性ベクトル変形関数の話し手に依存する組を変形する、をさらに具備する請求項４４の方法。
【請求項４８】
音響特性ベクトルを抽出する手段、
順応モデルで音響特性ベクトルの順応パターンマッチングを実行する手段、
順応パターンマッチングに基づいて特性ベクトル変形関数を選択する手段、
変形された音響特性ベクトルの組を形成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用する手段、及び
音響モデルで変形された音響特性ベクトルの組の音声認識パターンマッチングを実行する手段、
を具備する音声認識システム。
【請求項４９】
遠隔局中に記憶された順応モデルで音響特性ベクトルの順応パターンマッチングを実行する手段、
遠隔局で記憶された特性ベクトル変形関数情報から特性ベクトル変形関数を選択する手段、ここで、特性ベクトル変形関数を選択することがパターンマッチングに基づく、
変形された音響特性ベクトルの組を形成するために、音響特性ベクトルに選択された特性ベクトル変形関数を適用する手段、及び
通信センタに変形された音響特性ベクトルを送る手段、
を具備する遠隔局装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、会話信号処理に係わる。さらに詳しくは、本発明は、音響特性ベクトル変形を使用する分散型音声認識の卓越した方法及び装置に係わる。
【背景技術】
【０００２】
音声認識は、ユーザの音声命令を認識するため、及び人間と装置とのインターフェースを容易にするために、シミュレートされた情報を装置に与えるための最も重要な技術の一つである。音響会話信号から言葉のメッセージに復元する技術を採用したシステムは、音声

認識（ＶＲ）システムと呼ばれる。図１は、プリエンファシス(preemphasis)フィルタ１０２、音響特性抽出（ＡＦＥ）ユニット１０４、及びパターンマッチングエンジン(pattern matching engine)１１０を有する基本ＶＲシステムを示す。ＡＦＥユニット１０４は、デジタル音声サンプルの系列を音響特性ベクトルと呼ばれる測定値の組（例えば、抽出された周波数成分）に変換する。パターンマッチングエンジン１１０は、ＶＲ音響モデル１１２に含まれるパターンで音響特性ベクトルの系列を整合させる。ＶＲパターンマッチングエンジンは、一般に、この分野でよく知られているビタビ(Viterbi)復号技術を採用する。パターンの系列が音響モデル１１２から認識された場合、その系列は、入力発音に対応する言語学的な言葉の認識されたシーケンスのような、所望のフォーマットの出力になるように解析される。

音響モデル１１２は、種々の会話音及びそれに付随する統計的な分布情報から抽出された音響特性のデータベースとして説明される。これらの音響特性ベクトルは、音韻(phoneme)、トリホン(tri-phones)及びホールワード(whole-word)モデルのような短い会話セグメントに対応するパターンを作る。“トレーニング”は、音響モデル１１２においてパターンを生成するために、１若しくはそれ以上の話し手から特定の会話セグメント、若しくは音節の会話サンプルを集めるプロセスである。“テスチング”は、エンドユーザの会話サンプルから抽出した音響特性ベクトルの系列を音響モデル１１２の内容と相関させるプロセスである。所定のシステムの成果は、エンドユーザの会話とデータベースの内容との間の相関の程度に大きく依存する。

【０００３】

最も好ましくは、エンドユーザは、トレーニング及びテスチングの両方の期間、会話音響特性ベクトルを与え、その結果、音響モデル１１２は、エンドユーザの会話と強く整合する。しかしながら、音響モデル１１２は、一般に多数の会話節に対するパターンを表わさなければならないため、しばしば、大量のメモリを占有する。さらに、全ての可能な話し手から音響モデルを集めるために、必要な全てのデータを集めることは、実際的でない。それ故、多くの既存のＶＲシステムは、多くの代表的な話し手の会話を使用して集められた音響モデルを使用する。そのような音響モデルは、幅広い多くのユーザにわたり最善の成果を出すように設計されているが、いかなる個々のユーザに対しても最適化されない。そのような音響モデルを使用するＶＲシステムにおいて、特定のユーザの会話を認識する能力は、特定のユーザに最適化された音響モデルを使用するＶＲシステムのそれより劣るであろう。強い外国語なまりを持つユーザのようなあるユーザに関して、共有音響モデルを使用するＶＲシステムの性能は極めて悪く、ＶＲサービスを全く効果的に使用できない。

【０００４】

トレーニング及びテスト条件におけるミスマッチにより生ずる認識性能の劣化を軽減するために、順応は効果的な方法である。テスチング環境と密接に整合させるために、順応は、テスチングの期間ＶＲ音響モデルを変形する。最大ゆう度直線回帰(maximum likelihood linear regression)及びベイズ順応(Bayesian adaptation)のような、いくつかの順応スキームは、この分野ではよく知られている。

【０００５】

会話認識業務の複雑性が増加するにつれ、ワイアレス機器において認識システム全体を収容することは、ますます困難になる。それゆえ、中央通信センタに置かれた共有音響モデルは、全ユーザに対して音響モデルを提供する。中央基地局は、計算に費用がかかる音響マッチングに関しても責任がある。分散型ＶＲシステムでは、音響モデルは、多くの話し手により共有され、そのため、いかなる個々の話し手に対しても最適化されない。そこで、この分野において、計算に要求されるリソースを最小にする一方で、複数の個々のユーザに対する改善された性能を有するＶＲシステムの必要性がある。

【発明の開示】

【課題を解決するための手段】

【０００６】

ここに開示された方法及び装置は、卓越した及び改善された分散型音声認識システムを指向するものである。前記音声認識システムでは、音声認識パターンマッチングに先立ち、話し手に依存する処理が、音響特性ベクトルを変換するために使用される。話し手に依存する処理は、話し手に基づいて変化するパラメータを有する変換関数、順応モデルを使用する中間パターンマッチング処理の結果、又は両者にしたがって実行される。話し手に依存する処理は、遠隔局において、若しくは通信センタにおいて、又は二つの組み合わせで行える。音声認識パターンマッチングに先立ち、音響特性ベクトルは、環境に依存する処理を使用しても変換できる。音響特性ベクトルは、操業の音響環境（周囲ノイズ、マイクロホンの周波数応答等）における変化に順応するために変形される。環境に依存する処理は、遠隔局において、若しくは通信センタにおいて、又は二つの組み合わせで行える。

【０００７】

用語“イグゼンプラリ(exemplary)”は、ここでは“例、実例、若しくは例証として使われる”という意味で使用される。“イグゼンプラリ実施例”として説明されるいかなる実施例も、他の実施例に対して好ましい若しくは優位であると解釈される必要性はない。

【発明を実施するための最良の形態】

【０００８】

ここに開示された方法及び装置の特徴、目的、及び利点は、添付した図面とともに以下に行われる詳細な説明から、より明らかにされるであろう。図面において、参照文字は一貫して同一のものに対応する。

【０００９】

標準音声認識装置(recognizer)では、認識若しくはトレーニングのいずれかにおいて、コンピュータ処理上の複雑性の大部分は、音声認識装置のパターンマッチングサブシステムに集中している。ワイアレスシステムの背景に関連して、音声認識の適用によって使われるオーバージエア(over-the air)バンド幅を最小にするための分散型システムとして、音声認識装置は導入される。さらに、分散型ＶＲシステムは、ボコーダ(vocoder)を使用することでしばしば起きるような、音声データの無駄なソースコーデングを生じさせる性能の劣化を避ける。このような分散型構成は、米国特許番号No. 5,956,683、題名“分散型音声認識システム”に詳細に記載されている。これは、本発明の譲受人に譲渡されており、‘６８３特許としてここに引用されている。

【００１０】

デジタルワイアレス電話システムのような、イグゼンプラリワイアレス通信システムにおいて、ユーザの音声信号は、携帯電話若しくは遠隔局のマイクロホンを通して受信される。アナログ音声信号は、それから、デジタルサンプルストリーム、例えば、毎秒８０００　８－ビット会話サンプル、を生成するためにデジタルサンプルされる。ワイアレスチャネルを通して直接会話サンプルを送ることは、極めて非効率的である。それゆえ、情報は、送信する前に普通は圧縮される。ボコーデングと呼ばれる技術を介して、ボコーダは、会話サンプルのストリームをさらに小さな系列のボコーダパケットに圧縮する。そして、小さなボコーダパケットは、音声を表わす会話サンプルの代わりにワイアレスチャネルを通して送られる。そして、ボコーダパケットは、ワイアレス基地局により受信され、会話サンプルのストリームを生成するためにデボコードされる。そして、スピーカを通して聞き手に示される。

【００１１】

ボコーダの主目的は、デボコードした際に聞き手が会話を理解できるように保ちつつ、話し手の会話サンプルをできる限り圧縮することである。ボコーダアルゴリズムは、典型的にはロスのある圧縮アルゴリズムであり、その結果、デボコードされた会話サンプルは、最初にボココードされたサンプルと厳密には整合しない。その上、１若しくはそれ以上のボコーダパケットがワイアレスチャネルを介した送信において失われたとしても、理解できるデボコードされた会話を生成するために、ボコーダアルゴリズムは、しばしば最適化される。この最適化は、ボコーダに入力された会話サンプルとデボコーデングの結果との間のミスマッチをさらに引き起こす。ボコーデング及びデボコーデングに起因する会話サ

ンプルの変更は、一般に音声認識アルゴリズムの性能を劣化させる。その劣化の程度は、異なるボコーダアルゴリズムの間で大きく変化する。

【００１２】

‘６８３特許に記述されているシステムでは、遠隔局は、音響特性抽出を実行し、ボコーダパケットの代わりに音響特性ベクトルを基地局にワイアレスチャネル上で送る。音響特性ベクトルがボコーダパケットより少ないバンド幅しか占有しないため、音響特性ベクトルは、通信チャネルエラーからの保護を追加して（例えば、順方向エラー訂正（ＦＥＣ）技術を使用して）、同一のワイアレスチャネルを通して送信できる。特性ベクトルが、後で述べるように話し手に依存する特性ベクトル変形関数を使用してさらに最適化される場合、ＶＲ性能は、‘６８３特許に記載されている基本的なシステムの性能を超えていても、実現される。

【００１３】

図２は、イグゼンプラリ実施例にしたがった分散型ＶＲシステムを示す。音響特性抽出（ＡＦＥ）は、遠隔局２０２の中で生じ、音響特性ベクトルは、ワイアレスチャネル２０６を通して基地局及びＶＲ通信センタ２０４に送信される。ここに述べられている技術は、ワイアレスチャネルを含まないＶＲシステムにも同様に適用できることを、本技術分野に知識のある者は、理解するであろう。

【００１４】

示された実施例では、ユーザからの音声信号は、マイクロホン（ＭＩＣ）２１０で電気的信号に変換され、アナログ－デジタル変換機（ＡＤＣ）２１２でデジタル会話サンプルに変換される。デジタルサンプルストリームは、それからプリエンファシス（ＰＥ）フィルタ２１４、例えば、低周波数信号成分を弱めるフィニットインパルス応答(finite impulse response)（ＦＩＲ）フィルタ、を使用してフィルタされる。

【００１５】

フィルタされたサンプルは、その後、ＡＦＥユニット２１６で解析される。ＡＦＥユニット２１６は、デジタル音声サンプルを音響特性ベクトルに変換する。あるイグゼンプラリ実施例では、ＡＦＥユニット２１６は、異なる周波数ビンに対応した信号強度のベクトルを生成するために、連続したデジタルサンプルのセグメントにフーリエ変換を実施する。あるイグゼンプラリ実施例では、周波数ビンは、バークスケール(bark scale)にしたがって、変化するバンド幅を有する。バークスケールでは、各周波数ビンのバンド幅は、高周波数ビンが低周波数ビンより広い周波数幅を持つように、ビンの中心周波数にある関係を持っている。バークスケールは、ラビナーＬ．Ｒ．及びジュアングＢ．Ｈ．著、会話認識の基礎、プレンチスホール、１９９３に記載されており、本技術分野ではよく知られている。

【００１６】

あるイグゼンプラリ実施例では、各音響特性ベクトルは、一定時間間隔で集められた会話サンプルの系列から抽出される。あるイグゼンプラリ実施例では、これらの時間間隔は重なる。例えば、２つの連続する間隔のそれぞれが１０ミリ秒の区分を共有するように、音響特性は、１０ミリ秒毎に始まる２０ミリ秒間隔の会話データから得られる。ここに記載された実施例から逸脱しない範囲で、時間間隔が重ならないようにする、若しくは一定でない期間にできることを、この技術分野に知識のある者は、理解できるであろう。

【００１７】

ＡＦＥユニット２１６により生成された各音響特性ベクトル（図２で、Ｘと識別される）は、順応エンジン(adaptation engine)２２４に与えられる。順応エンジンは、順応モデル２２８の内容に基づいて音響特性ベクトルを特徴付けるためにパターンマッチングを実行する。パターンマッチングの結果に基づいて、順応エンジン２２４は、メモリ２２７から特性ベクトル変形関数ｆ()の組の一つを選択し、変形された音響特性ベクトルｆ(Ｘ)を生成するために使用する。

【００１８】

Ｘは、ここでは単一の音響特性ベクトル、若しくは連続する音響特性ベクトルの系列のど

ちらかを記述するために使用される。同様に、ｆ(Ｘ)は、単一の変形された音響特性ベクトル、若しくは連続する変形された音響特性ベクトルの系列のどちらかを記述するために使用される。

【００１９】

あるイグゼンプラリ実施例では、図２に示されるように、その後、変形されたベクトルｆ(Ｘ)は、ワイアレスモデム２１８で変調され、ワイアレスチャネル２０６を通して送信され、通信センタ２０４内のワイアレスモデム２３０で復調され、中央ＶＲエンジン２３４により中央音響モデル２３８に対して整合される。ワイアレスモデム２１８，２３０及びワイアレスチャネル２０６は、ＣＤＭＡ，ＴＤＭＡ，若しくはＦＤＭＡを含む各種のワイアレスインターフェースのいずれかを使用できる。さらに、ワイアレスモデム２１８，２３０は、他のタイプの通信インターフェースと置き換えられる。他のタイプの通信インターフェースは、説明された実施例の範囲から逸脱しないでワイアレスでないチャネルを通して通信する。例えば、遠隔局２０２は、ランドラインモデム(land-line modem)、Ｔ１／Ｅ１、ＩＳＤＮ、ＤＳＬ、イサーネット、若しくはプリント回路基板（ＰＣＢ）のトレースをも含む各種タイプの通信チャネルのいずれかを通して通信センタと通信する。

【００２０】

あるイグゼンプラリ実施例では、ベクトル変形関数ｆ()は、特定のユーザ若しくは話し手に対して最適化され、中央音響モデル２３８に対して整合された場合、会話が正確に認識される確率を最大にするように設計される。中央音響モデルは、複数のユーザ間で共有される。遠隔局２０２中の順応モデル２２８は、中央音響モデル２３８よりかなり小さく、特定ユーザに対して最適化された個別の順応モデル２２８を維持できるようにする。１若しくはそれ以上の話し手に対する特性ベクトル変形関数ｆ()のパラメータも、遠隔局２０２のメモリ２２７に記憶できるように十分に小さい。

【００２１】

他の実施例では、環境に依存するベクトル変形関数に関するパラメータの追加の組も、メモリ２２７に記憶される。環境に依存するベクトル変形関数の選択及び最適化は、本質的にはより総体的であり、一般に各通話の間に実施できる。非常に単純な環境に依存する特性ベクトル変形関数の一例は、雑音の多い環境に順応するために各音響特性ベクトルの各要素に一定の利得ｋを適用している。

【００２２】

ベクトル変形関数ｆ()は、各種形態のいずれかを持つことができる。例えば、ベクトル変形関数ｆ()は、ＡＸ＋ｂの形式の擬似変換であってよい。あるいは、ベクトル変形関数ｆ()は、初期化され、その後、連続する音響特性ベクトルの組に適用されたフィニットインパルス応答（ＦＩＲ）フィルタの組であってよい。ベクトル変形関数ｆ()の他の形式は、本技術分野に知識のある者には明らかであろうし、ここに記述した実施例の範囲内にある。

【００２３】

あるイグゼンプラリ実施例では、ベクトル変形関数ｆ()は、連続する音響特性ベクトルの組に基づいて選択される。例えば、順応エンジン２２４は、音響特性ベクトルのストリームと順応モデル２２８中の複数の会話パターンとの間の相関の程度を決めるために、ビタビデコーデング若しくはトレリス(trellis)デコーデング技術を適用できる。一旦、高い相関の程度が検出されると、ベクトル変形関数ｆ()は、検出されたパターンに基づいて選択され、音響特性ベクトルのストリームの中から対応するセグメントに適用される。このアプローチは、各音響特性ベクトルに適用されるべきｆ()を選択する前に、順応エンジン２２４が音響特性ベクトルの系列を記憶し、順応モデル２２８に対する系列のパターンマッチングを実行することが必要である。あるイグゼンプラリ実施例では、順応エンジンは、未変形の音響特性ベクトルのエラスチックバッファ(elastic buffer)を維持する。そして、送信する前に、選択されたｆ()をエラスチックバッファの内容に適用する。エラスチックバッファの内容は、順応モデル２２８中のパターンと比較される。そして、エラスチックバッファの内容と最大の相関の程度を有するパターンに関して、最大相関メトリック

(maximum correlation metric)が生成される。最大相関は、１若しくはそれ以上のしきい値に対して比較される。最大相関が検出しきい値を超えるならば、最大相関に連携したパターンに対応するｆ()が、バッファ中の音響特性ベクトルに適用され、送信される。最大相関が検出しきい値を超す前にエラスチックバッファが一杯になれば、エラスチックバッファの内容は、変形しないで送信される、若しくは、デフォールトｆ()を使用して代わりに変形される。

【００２４】

ｆ()の話し手に依存した最適化は、多くの方法のいずれかにより達成できる。第１のイグゼンプラリ実施例では、制御プロセッサ２２２は、ユーザの会話と多数の言葉にわたる順応モデル２２８との間の相関の程度をモニタする。ｆ()の変化がＶＲ性能を改善するであろうと制御プロセッサ２２２が決める場合、制御プロセッサ２２２は、ｆ()のパラメータを変形し、メモリ２２７に新しいパラメータを記憶する。あるいは、制御プロセッサ２２２は、ＶＲ性能を改善するために直接順応モデル２２８を変形できる。

【００２５】

図２に示されたように、遠隔局２０２は、個別のＶＲエンジン２２０及び遠隔局音響モデル２２６を付加的に含むことができる。メモリ容量の制限から、ワイアレス電話のような遠隔局２０２における遠隔局音響モデル２２６は、一般に小さいはずであり、それゆえ少数の句若しくは音素に制限される。一方、遠隔局音響モデル２２６が、少数のユーザにより使用される遠隔局に含まれているため、遠隔局音響モデル２２６は、ＶＲ性能を改善するために１若しくはそれ以上の特定ユーザに最適化されることができる。例えば、“コール”及び各１０の数字(ten digits)のような言葉に対する会話パターンは、ワイアレス電話の所有者に整合される。そのようなローカルな遠隔局音響モデル２２６は、遠隔局２０２が言葉の小さな組に対して非常によいＶＲ性能を有することを可能にする。さらに、遠隔局音響モデル２２６は、通信センタ２０４にワイアレスリンクを確立しなくとも、遠隔局２０２がＶＲを達成することを可能にする。

【００２６】

ｆ()の最適化は、管理された若しくは管理されない学習のいずれかを通して起きる。管理された学習は、所定の単語若しくは文章をユーザが発音することで生じるトレーニングを、一般に指す。所定の単語若しくは文章は、遠隔局音響モデルを正確に最適化するために使用される。ＶＲシステムは、入力として使用された単語若しくは文章のプリオリ(priori)知識を有するため、所定の単語若しくは文章を認識するために管理された学習期間中ＶＲを実行する必要がない。管理された学習は、特定ユーザに対する音響モデルを生成する最も正確な方法であると一般に考えられている。管理された学習の一例は、遠隔局２０２の遠隔局音響モデル２２６の中に１０の数字に関する会話をユーザが最初にプログラムする場合である。遠隔局２０２が、話された数字に対応する会話パターンのプリオリ知識を有するため、遠隔局音響モデル２２６は、ＶＲ性能を劣化させる小さいリスクで個別ユーザに整合される。

【００２７】

管理された学習とは対照的に、管理されない学習は、発音された会話パターン若しくは単語のプリオリ知識を持つＶＲシステムなしで生じる。発音が誤った会話パターンとマッチングするというリスクのため、管理されない学習に基づいた遠隔局音響モデルの変形は、非常に慎重なやり方で行われなければならない。例えば、多くの過去の発音は、互いに似ており、そして他の会話パターンより音響モデルの中のある会話パターンにより近いものを、発生したであろう。これらの全ての過去の発音が、モデル中のある会話パターンと正確に合っているならば、音響モデル中のその会話パターンは、同様の発音の組にさらによく整合させるように変形されるであろう。しかし、それらの過去の発音の多くが、モデル中のある会話パターンに対応しなければ、その会話パターンを変形することは、ＶＲ性能を劣化させるであろう。好ましくは、ＶＲシステムは、過去のパターンマッチングの精度に関してユーザからのフィードバックを集められる。しかし、このようなフィードバックは、頻繁には利用できない。

【００２８】
あいにく、管理された学習は、ユーザにとって長たらしくたいくつであり、多数の会話パターンを持つ音響モデルを生成することを非現実的にしている。しかし、管理された学習は、ベクトル変形関数ｆ()の組を最適化する際に、若しくは順応モデル２２８においてさらに制限された会話パターンを最適化する際にさえ、まだ有効である。ユーザの強いなまりによって生じる会話パターンの差異は、管理された学習が必要とされる応用の一例である。音響特性ベクトルは、なまりを補正するために大きな変形を必要とするため、その変形において正確さに対する必要性が、大きい。

【００２９】
管理されない学習は、最適化がＶＲエラーの直接原因になりにくいと思われる特定ユーザに対するベクトル変形関数ｆ()を最適化するためにも使用できる。例えば、普通より長い音声トラクト長さ(vocal tract length)若しくは平均音声ピッチを有する話し手に順応するために必要なベクトル変形関数ｆ()の調整は、なまりを補正するために要求される調整より、本質的により全体的である。そのような全体的なベクトル変形における大きな不正確さは、ＶＲ有効性に強く影響を与えずにできる。

【００３０】
一般に、順応エンジン２２４は、小さな順応モデル２２８をベクトル変形関数ｆ()を選択するためにだけ使用し、全体のＶＲを実行するためには使用しない。サイズが小さいために、順応モデル２２８は、順応モデル２２８若しくはベクトル変形関数ｆ()のいずれかを最適化するためのトレーニングを実行するためには、同様に適さない。順応モデル２２８若しくはベクトル変形関数ｆ()の調整は、順応モデル２２８に対する話し手の音声データのマッチングの程度を改善するために現れる。その順応モデル２２８若しくはベクトル変形関数ｆ()の調整は、大きな中央音響モデル２３８に対するマッチングの程度を実際に劣化させる。中央音響モデル２３８は、実際にＶＲに使用されるものであるため、そのような調整は、最適化というよりむしろ誤りであろう。

【００３１】
あるイグゼンプラリ実施例では、遠隔局２０２及び通信センタ２０４は、順応モデル２２８若しくはベクトル変形関数ｆ()のいずれかを変形するために管理されない学習を使用する際に協力する。順応モデル２２８若しくはベクトル変形関数ｆ()のいずれかを変形するか否かの決定は、中央音響モデル２３８に対して改善されたマッチングに基づいて行われる。例えば、遠隔局２０２は、通信センタ２０４に、音響特性ベクトル、未変形の音響特性ベクトルＸ及び変形された音響特性ベクトルｆ(Ｘ)の複数の組を送ることができる。あるいは、遠隔局２０２は、変形された音響特性ベクトルf_1(Ｘ)及びf_2(Ｘ)を送ることができる。ここで、f_2()は、仮の改善された特性ベクトル変形関数である。他の実施例では、遠隔局２０２は、Ｘ及び特性ベクトル変形関数f_1()及びf_2()両者に関するパラメータを送る。遠隔局２０２は、通信センタ２０４に情報の第２の組を送ることが固定時間間隔基づいてできるか否かの、複数の組の決定を送ることができる。

【００３２】
変形された音響特性ベクトル若しくは特性ベクトル変形関数に関するパラメータのいずれかの、音響特性情報の複数の組を受信すると、通信センタ２０４は、自身のＶＲエンジン２３４及び中央音響モデル２３８を使用して結果としての変形された音響特性ベクトルのマッチングの程度を評価する。通信センタ２０４は、それから、変更がＶＲ性能の改善をもたらすか否かを指示する情報を、遠隔局２０２に送り返す。例えば、通信センタ２０４は、音響特性ベクトルの各組に関する会話パターン相関メトリックを遠隔局２０２に送る。音響特性ベクトルの各組に関する会話パターン相関メトリックは、音響特性ベクトルの組と中央音響モデル２３８の内容との間の相関の程度を示す。２つの組のベクトルの間の相対的な相関の程度に基づいて、遠隔局２０２は、その順応モデル２２８を調整できる、又は１若しくはそれ以上の特性ベクトル変形関数ｆ()を調整できる。遠隔局２０２は、実際の言葉の認識に対して使われるどちらのベクトルの組を使用するかを特定できる。若しくは、通信センタ２０４は、その相関メトリックスに基づいてベクトルの組を選択できる

。代わりの実施例では、遠隔局２０２は、通信センタ２０４から結果としての相関メトリックスを受信した後、ＶＲに関して使用される音響特性ベクトルの組を同定する。

【００３３】

代わりの実施例では、遠隔局２０２は、特性ベクトル変形関数ｆ()を認識するためにローカルな順応エンジン２２４及び順応モデル２２８を使用し、通信センタ２０４にｆ()とともに未変形の音響特性ベクトルＸを送る。それから通信センタ２０４は、ｆ()をＸに適用し、変形した及び未変形のベクトルの両方を使用してテスチングを実行する。その後、遠隔局２０２によって特性ベクトル変形関数のより正確な調整ができるように、通信センタ２０４は、遠隔局２０２にテスチングの結果を送り返す。

【００３４】

他の実施例では、順応エンジン２２４及び順応モデル２２８は、遠隔局２０２の代わりに通信センタ２０４に取り込まれる。通信センタ２０４内の制御プロセッサ２３２は、モデム２３０を通して未変形の音響特性ベクトルのストリームを受信し、通信センタ２０４内の順応エンジン及び順応モデルにそれらを与える。この中間パターンマッチングの結果に基づいて、制御プロセッサ２３２は、通信センタメモリ２３６に記憶されているデータベースから特性ベクトル変形関数ｆ()を選択する。あるイグゼンプラリ実施例では、通信センタメモリ２３６は、特定のユーザに対応する特性ベクトル変形関数ｆ()の組を含む。これは、前記の遠隔局２０２に記憶されている特性ベクトル変形関数情報に追加される、若しくは代わりのいずれかである。通信センタ２０４は、特性ベクトルが抽出された音声データを提供している個々の話し手を認識するために、各種のタイプの話し手認識情報のいずれもが使用できる。例えば、特性ベクトル変形関数の組を選択するために使用された話し手認識情報は、ワイアレスチャネル２０６の反対側の一端におけるワイアレス電話の移動認識数（ＭＩＮ）でありうる。あるいは、ユーザは、ＶＲサービスを向上させる目的で自身を認識するためにパスワードを登録できる。さらに、環境に依存する特性ベクトル変形関数は、会話データの観測に基づいたワイアレス電話の通話の間に順応でき、適用できる。他の多くの方法も、ここで述べられた実施例の範囲から逸脱しないで、話し手に依存するベクトル変形関数の組を選択するために使用できる。

【００３５】

本技術分野において知識のある者は、遠隔局２０２の中の複数のパターンマッチングエンジン２２０，２２４が、ここに記述した範囲から逸脱しないで統合できることも、理解するであろう。しかも、遠隔局２０２中の異なる音響モデル２２６，２２８は、同様に統合できる。さらに、１若しくはそれ以上の音響モデル２２０，２２４は、遠隔局２０２の制御プロセッサ２２２に取り込むことができる。１若しくはそれ以上の音響モデル２２６，２２８も、制御プロセッサ２２２により使用されるメモリ２２７の中に含めることができる。

【００３６】

通信センタ２０４において、中央会話パターンマッチングエンジン２３４は、ここに記述した範囲から逸脱しないで、もし存在するならば、順応エンジン（図示しない）と統合されることができる。しかも、中央音響モデル２３８は、順応モデル（図示しない）と統合されることができる。さらに、もし通信センタ２０４中に存在するならば、中央会話パターンマッチングエンジン２３４及び順応エンジン（図示しない）のいずれか、若しくは両者は、通信センタ２０４の制御プロセッサ２３２の中に含めることができる。もし通信センタ２０４中に存在するならば、中央音響モデル２３８及び順応エンジン（図示しない）のいずれか、若しくは両者は、通信センタ２０４の制御プロセッサ２３２の中に含めることもができる。

【００３７】

図３は、分散型ＶＲを実行する方法のフローチャートである。分散型ＶＲでは、Ｘ及びｆ()の変形が、遠隔順応モデルに収束することに基づいて遠隔局２０２において全て発生する。ステップ３０２において、遠隔局２０２は、デジタル音声サンプルのストリームを生成するために、マイクロホンからアナログ音声信号を採取する。ステップ３０４において

、会話サンプルは、そして、例えば前記のプリエンファシスフィルタを使用してフィルタされる。ステップ３０６において、音響特性ベクトルＸのストリームは、フィルタされた会話サンプルから抽出される。前記のように、音響特性ベクトルは、重なる若しくは重ならない間隔で会話サンプルから抽出されうる。その間隔は、固定若しくは可変の期間のいずれかである。

【００３８】

ステップ３０８において、遠隔局２０２は、音響特性ベクトルのストリームと順応モデル（図２の２２８のような）に含まれる複数のパターンとの間の相関の程度を決定するために、パターンマッチングを実施する。ステップ３１０において、遠隔局２０２は、順応モデル中のパターンを選択する。パターンは、音響特性ベクトルＸのストリームに最も密接に整合する。選択されたパターンは、ターゲットパターンと呼ばれる。前に議論したように、Ｘとターゲットパターンとの間の相関の程度は、検出しきい値に対して比較される。相関の程度が検出しきい値より大きければ、その後、遠隔局２０２は、ターゲットパターンに対応する特性ベクトル変形関数ｆ()を選択する。相関の程度が検出しきい値より小さければ、その後、遠隔局２０２は、ｆ(Ｘ)＝Ｘになるように音響特性ベクトル認識関数ｆ()を選択するか、若しくはあるデフォールトｆ()を選択する。あるイグゼンプラリ実施例では、遠隔局２０２は、自身のローカル順応モデルにある各種のパターンに対応する特性ベクトル変形関数のローカルデータベースから特性ベクトル変形関数ｆ()を選択する。遠隔局２０２は、ステップ３１２において音響特性ベクトルＸのストリームに選択された特性ベクトル変形関数ｆ()を適用する。そのようにしてｆ(Ｘ)が生成される。

【００３９】

あるイグゼンプラリ実施例では、遠隔局２０２は、Ｘとターゲットパターンとの間の相関の程度を示す相関メトリックを生成する。遠隔局２０２は、ｆ(Ｘ)とターゲットパターンとの間の相関の程度を示す相関メトリックも生成する。管理されない学習の一例では、遠隔局２０２は、ステップ３１４において、１若しくはそれ以上の特性ベクトル変形関数ｆ()を変形するか否かを決定するために、過去の相関メトリックの値とともに２つの相関メトリックを使用する。ステップ３１４においてｆ()を変形する決定がなされれば、その後、ｆ()はステップ３１６において変形される。あるイグゼンプラリ実施例では、変形されたｆ()は、新たな変形された音響特性ベクトルｆ(Ｘ)を形成するために、ステップ３１８においてＸに直ちに適用される。代わりの実施例では、ステップ３１８が省略され、新たな特性ベクトル変形関数ｆ()は、音響特性ベクトルＸの後の組まで効果を生じない。

【００４０】

ステップ３１４において、若しくはステップ３１６及び３１８の後で、ｆ()を変形しない決定がなされれば、遠隔局２０２は、ステップ３２０において通信センタ２０４にワイアレスチャネル２０６を通して現在のｆ(Ｘ)を送信する。その後、ＶＲパターンマッチングは、ステップ３２２において通信センタ２０４の中で起きる。

【００４１】

代わりの実施例では、通信センタ２０４は、ＶＲパターンマッチングステップ３２２の間に会話パターン相関メトリックスを生成し、ｆ()の最適化を助けるために、遠隔局３０２にこれらのメトリックスを送り返す。会話パターン相関メトリックスは、いくつかの方法のうちの一つでフォーマットされる。例えば、通信センタ２０４は、音響特性ベクトル変形エラー関数$f_E()$を返信することができる。$f_E()$は、中央音響モデル中に見つけられたパターンで正確な相関を作り出すために、ｆ(Ｘ)に適用できる。あるいは、通信センタ２０４は、ターゲットパターン若しくはｆ(Ｘ)と最大の相関の程度を有すると認められた中央音響モデル中のパターンに対応する音響特性ベクトルの組を単純に返信できる。あるいは、通信センタ２０４は、ターゲットパターンを選択するために使用された、ハードデシジョン(hard-decision)若しくはソフトデシジョン(soft-decision)ビタビデコーデングプロセスから導かれる枝のメトリックを返信できる。会話パターン相関メトリックスは、情報のこれらのタイプの組み合わせも含むことができる。この返信情報は、その後、ｆ()の最適化において遠隔局２０２によって使用される。あるイグゼンプラリ実施例では、ス

テップ３１８におけるｆ(Ｘ)の再生成は省略され、遠隔局２０２は、通信センタ２０４からフィードバックを受信した後、ｆ()の変形（ステップ３１４，３１６）を実行する。

【００４２】

図４は、分散型ＶＲを実行する方法を示すフローチャートである。分散型ＶＲでは、Ｘ及びｆ()の変形は、中央音響モデルとの相関に基づいて通信センタ２０４の中で全て発生する。ステップ４０２において、遠隔局２０２は、デジタル音声サンプルのストリームを生成するために、マイクロホンからアナログ音声信号を採取する。ステップ４０４において、会話サンプルは、そして、例えば前記のプリエンファシスフィルタを使用してフィルタされる。ステップ４０６において、音響特性ベクトルＸのストリームは、フィルタされた会話サンプルから抽出される。前記のように、音響特性ベクトルは、重なる若しくは重ならない間隔で会話サンプルから抽出される。その間隔は、固定若しくは可変の期間のいずれかである。

【００４３】

ステップ４０８において、遠隔局２０２は、音響特性ベクトルＸの未変形のストリームをワイアレスチャネル２０６を通して送信する。ステップ４１０において、通信センタ２０４は、順応パターンマッチングを実行する。前に議論したように、順応パターンマッチングは、個別の順応モデル使用して、若しくは大きな中央音響モデル２３８を使用してのいずれかで達成できる。ステップ４１２において、通信センタ２０４は、音響特性ベクトルＸのストリームに最も密接に整合する、順応モデル中のパターンを選択する。選択されたパターンは、ターゲットパターンと呼ばれる。前記のように、Ｘとターゲットパターンとの間の相関の程度が、しきい値より大きければ、ターゲットパターンに対応するｆ()が選択される。そうでなければ、デフォールトｆ()若しくはナルｆ()が選択される。ステップ４１４において、選択された特性ベクトル変形関数ｆ()は、音響特性ベクトルｆ(Ｘ)の変形されたストリームを生成するために、音響特性ベクトルＸのストリームに適用される。

【００４４】

あるイグゼンプラリ実施例では、特性ベクトル変形関数ｆ()は、通信センタ２０４中に存在する特性ベクトル変形関数の大きなデータベースのサブセットから選択される。選択に利用できる特性ベクトル変形関数のサブセットは、話し手に依存する。そうすることにより、中央音響モデル（図２の２３８のような）を使用したパターンマッチングが、入力としてＸよりｆ(Ｘ)を使用するほうがより正確になる。前記のように、通信センタ２０４が、どのようにして話し手に依存する特性ベクトル変形関数のサブセットを選択できるかの例は、話し手のワイアレス電話のＭＩＮ、若しくは話し手により登録されたパスワードを使用することを含む。

【００４５】

あるイグゼンプラリ実施例では、通信センタ２０４は、Ｘとターゲットパターンとの間の相関、及びｆ(Ｘ)とターゲットパターンとの間の相関に関する相関メトリックスを生成する。その後、通信センタ２０４は、ステップ４１６において、１若しくはそれ以上の特性ベクトル変形関数ｆ()を変形するか否かを決めるために、過去の相関メトリック値とともに２つの相関メトリックスを使用する。ステップ４１６においてｆ()を変形すると決定されれば、その後、ｆ()はステップ４１８において変形される。あるイグゼンプラリ実施例では、変形されたｆ()は、新たな変形された音響特性ベクトルｆ(Ｘ)を生成するために、ステップ４２０においてＸに直ちに適用される。代わりの実施例では、ステップ４２０が省略され、新たな特性ベクトル変形関数ｆ()は、後の音響特性ベクトルの組まで効果を生じない。

【００４６】

ステップ４１６において、若しくはステップ４１８及び４２０の後で、ｆ()を変形しないと決定されれば、通信センタ２０４は、ステップ４２２において中央音響モデル２３８を使用してＶＲパターンマッチングを実行する。

【００４７】

図５は、分散型ＶＲを実行する方法を示すフローチャートである。ここでは、通信センタ

２０４中の中央音響モデルが、特性ベクトル変形関数若しくは順応モデルを最適化するために使用される。あるイグゼンプラリ実施例では、遠隔局２０２及び通信センタ２０４は、必要に応じて情報を交換し、特性ベクトル変形関数の最適化の精度を最大にするために協力する。

【００４８】

ステップ５０２において、遠隔局２０２は、デジタル音声サンプルのストリームを生成するために、アナログ音声信号を採取する。それから、ステップ５０４において、会話サンプルは、例えば、前記のようにプリエンファシスフィルタを使用して、フィルタされる。ステップ５０６において、音響特性ベクトルＸのストリームは、フィルタされた会話サンプルから抽出される。前記のように、音響特性ベクトルは、重なる若しくは重ならない間隔の会話サンプルのいずれかから抽出される。会話サンプルの間隔は、固定若しくは可変の期間のいずれかである。

【００４９】

ステップ５０８において、遠隔局２０２は、音響特性ベクトルのストリームと順応モデル（図２の２２８のような）に含まれる複数のパターンとの間の相関の程度を決定するために、パターンマッチングを実行する。ステップ５１０において、遠隔局２０２は、音響特性ベクトルＸのストリームに最もよく整合する順応モデル中のパターンを選択する。選択されたパターンは、ターゲットパターンと呼ばれる。前記のように、Ｘとターゲットパターンとの間の相関が、しきい値を超えるならば、第１の特性ベクトル変形関数ｆ$_1$()は、ターゲットパターンに対応するものとして選択される。そうでなければ、デフォールトｆ()若しくはヌルｆ()が選択される。遠隔局２０２は、ローカル順応モデル中の各種パターンに対応する特性ベクトル変形関数のローカルデータベースから特性ベクトル変形関数ｆ()を選択する。遠隔局２０２は、ステップ５１２において、選択された特性ベクトル変形関数ｆ()を音響特性ベクトルＸのストリームに適用する。このようにしてf(Ｘ)が生成される。

【００５０】

図３及び図４に関連して述べられた方法とは対照的に、ステップ５１４において、遠隔局２０２は、２組の音響特性ベクトル、ｆ$_1$(Ｘ)及びｆ$_2$(Ｘ)、をチャネル２０６を通して通信センタ２０４に送る。ステップ５１６において、通信センタ２０４は、入力としてｆ$_1$(Ｘ)を使用して自身の中央音響モデルに対してパターンマッチングを実行する。このＶＲパターンマッチングの結果として、通信センタ２０４は、ｆ$_1$(Ｘ)と最大の相関の程度を有するターゲットパターン若しくはパターンの組を識別する。ステップ５１８において、通信センタ２０４は、ｆ$_1$(Ｘ)とターゲットパターンとの間の相関の程度を示す第１の会話パターン相関メトリック及びｆ$_2$(Ｘ)とターゲットパターンとの間の相関の程度を示す第２の会話パターン相関メトリックを生成する。

【００５１】

音響特性ベクトルの両方の組が、中央音響モデルに対するパターンマッチングのために使用されるが、１組だけが実際のＶＲのために使用される。それ故、遠隔局２０２は、性能の予期しない劣化のリスクなしに、提案された特性ベクトル変形関数の性能を評価できる。遠隔局２０２は、ｆ()を最適化する際に、小さな、ローカル順応モデルに全てを引き継ぐ必要もない。代わりの実施例では、遠隔局２０２は、ｆ$_2$(Ｘ)＝Ｘとなるように、ｆ$_2$()にヌル関数を使用できる。このアプローチは、音響特性ベクトルの変形なしで達成されるＶＲ性能に対するｆ()の性能を、遠隔局２０２が証明することを可能にする。

【００５２】

ステップ５２０において、通信センタ２０４は、２つの会話パターン相関メトリックスをワイアレスチャネル２０６を通して遠隔局２０２に送り返す。受信した会話パターン相関メトリックスに基づいて、ステップ５２２において、遠隔局２０２は、ステップ５２４においてｆ$_1$()を変形するか否かを決定する。ステップ５２２においてｆ$_1$()を変形するか否かの決定は、１組の会話パターン相関メトリックスに基づくことができる、若しくは、ローカル順応モデルからの同一の会話パターンに関連した会話パターン相関メトリックス

の系列に基づくことができる。前に議論したように、会話パターン相関メトリックスは、音響特性ベクトル変形エラー関数ｆ$_{E}$()、ｆ(Ｘ)と最大の相関の程度を有していると認められた中央音響モデル中のパターンに対応する音響特性ベクトルの組、若しくはビタビデコーデングブランチメトリック(Viterbi decording branch metric)のような情報を含むことができる。

【００５３】

前記の技術が、各種のワイアレスチャネル２０６のいかなるタイプに同様に適用できることは、本技術分野に知識のある者は、理解するであろう。例えば、ワイアレスチャネル２０６（及びそれに応じたモデム２１８，２３０）は、符号分割多重アクセス（ＣＤＭＡ）技術、アナログセルラ、時間分割多重アクセス（ＴＤＭＡ）、若しくは他のタイプのワイアレスチャネルで利用できる。あるいは、チャネル２０６は、ワイアレス、光に限定されずに含む、赤外、及びイサーネットチャネル以外のチャネルのタイプでありうる。さらに他の実施例では、遠隔局２０２及び通信センタ２０４は、単一のシステムに統合され、チャネル２０６を全て回避する。統合されたシステムは、中央音響モデル２３８を使用するＶＲテスチングに先立ち、音響特性ベクトルの話し手に依存する変形を実行する。

【００５４】

情報及び信号が、種々の異なる技術及び手法のいずれかを使用して表わされることを、本技術分野に知識のある者は、理解するであろう。例えば、前記の記述を通して示される、データ、指示、命令、情報、信号、ビット、シンボル、及びチップは、電圧、電流、電磁波、磁場若しくは磁力粒子、光場若しくは光粒子、若しくはこれらの任意の組み合わせによって表わされる。

【００５５】

各種の解説的な論理ブロック、モジュール、回路、及びここに開示された実施例に関連して記述されたアルゴリズムステップが、電子ハードウェア、コンピュータソフトウェア、若しくは両者の組み合わせとして実施できることは、知識のある者は、さらに価値を認めるであろう。ハードウェア及びソフトウェアのこの互換性をはっきりと説明するために、各種の解説的な構成要素、ブロック、モジュール、回路、及びステップは、一般的に機能性の面からこれまでに記述されてきた。そのような機能性が、ハードウェア若しくはソフトウェアとして実行されるか否かは、個々の応用及びシステム全体に課せられた設計の制約に依存する。熟練した職人は、述べられた機能性を各個人の応用に対して違ったやり方で実行する。しかし、そのような実行の決定は、本発明の範囲から離れては説明されない。

【００５６】

ここに開示された実施例に関連して述べられた、各種の解説的な論理ブロック、モジュール、及び回路は、汎用プロセッサ、デジタルシグナルプロセッサ（ＤＳＰ）、アプリケーションスペシフィック集積回路（ＡＳＩＣ）、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）若しく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若しくはトランジスタロジック、ディスクリートハードウェア素子、若しくはここに記述した機能を実行するために設計されたこれらのいかなる組み合わせを、実施若しくは実行できる。汎用プロセッサは、マイクロプロセッサでよく、しかし代わりとして、プロセッサは、いかなる従来のプロセッサ、コントローラ、マイクロコントローラ、若しくはステートマシン(state machine)であってもよい。プロセッサは、演算デバイスの組み合わせとして実行できる。例えば、ＤＳＰとマイクロプロセッサの組み合わせ、複数のマイクロプロセッサ、ＤＳＰコアと結合した１若しくはそれ以上のマイクロプロセッサ、若しくはそのようないかなる他の構成であってもよい。

【００５７】

ここに開示された実施例に関連して述べられた方法のステップ若しくはアルゴリズムは、ハードウェアにおいて、プロセッサにより実行されるソフトウェアモジュールにおいて、若しくは、両者の組み合わせにおいて直接実現できる。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、ＥＰＲＯＭメモリ、ＥＥＰＲＯＭメモリ、

レジスタ、ハードディスク、脱着可能なディスク、ＣＤ－ＲＯＭ、若しくは、この分野で知られている他のいかなる記憶媒体の中に存在できる。あるイグゼンプラリ記憶媒体は、プロセッサが記憶媒体から情報を読み出し、そこに情報を書き込めるようなプロセッサと結合される。その代わりのものでは、記憶媒体は、プロセッサに集積できる。プロセッサ及び記憶媒体は、ＡＳＩＣ中に存在できる。ＡＳＩＣは、遠隔局中に存在できる。この代わりのものでは、プロセッサ及び記憶媒体は、遠隔局中に単体の構成部品として存在できる。

【００５８】

開示された実施例のこれまでの説明は、本技術分野に知識のあるいかなる者でも、本発明を作成し、使用することを可能にする。これらの実施例の各種の変形は、本技術分野に知識のある者に、容易に実現されるであろう。そして、ここで定義された一般的な原理は、本発明の精神及び範囲から逸脱しないで、他の実施例にも適用できる。それゆえ、本発明は、ここに示された実施例に制限することを意図したものではなく、ここに開示した原理及び卓越した特性と整合する広い範囲に適用されるものである。

【図面の簡単な説明】

【００５９】

【図１】図１は、基本音声認識システムを示す。

【図２】図２は、イグゼンプラリ実施例にしたがった分散型ＶＲシステムを示す。

【図３】図３は、分散型ＶＲを実行するための方法を示すフローチャートであり、ここで、音響特性ベクトル変形、及び特性ベクトル変形関数の選択が遠隔局において発生する。

【図４】図４は、分散型ＶＲを実行するための方法を示すフローチャートであり、ここで、音響特性ベクトル変形、及び特性ベクトル変形関数の選択が通信センターにおいて発生する。及び、

【図５】図５は、分散型ＶＲを実行するための方法を示すフローチャートであり、ここで、中央音響モデルが特性ベクトル変形関数、若しくは順応モデルを最適化するために使用される。

【符号の説明】

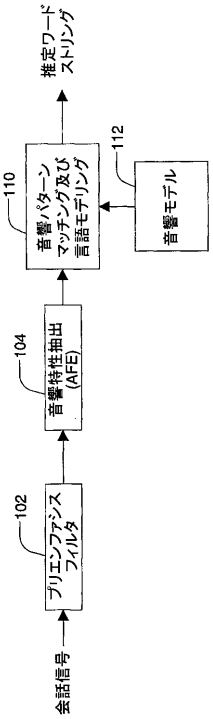
【００６０】

２０２…遠隔局，

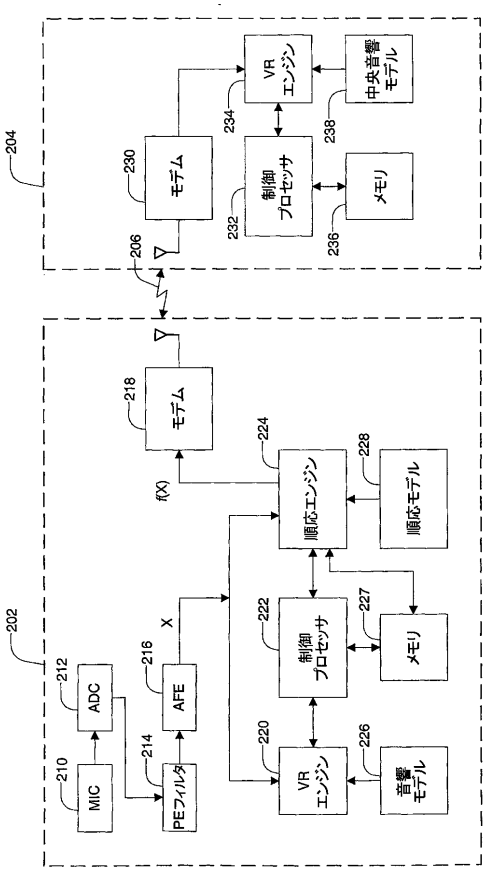
２０４…通信センタ，

２０６…ワイアレスチャネル，

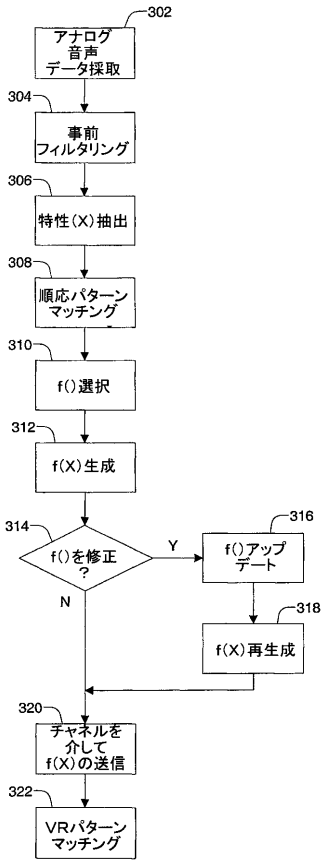
【図１】


会話信号
102
プリエンファシスフィルタ
104
音響特性抽出(AFE)
110
音響パターンマッチング及び言語モデリング
112
音響モデル
推定ワードストリング

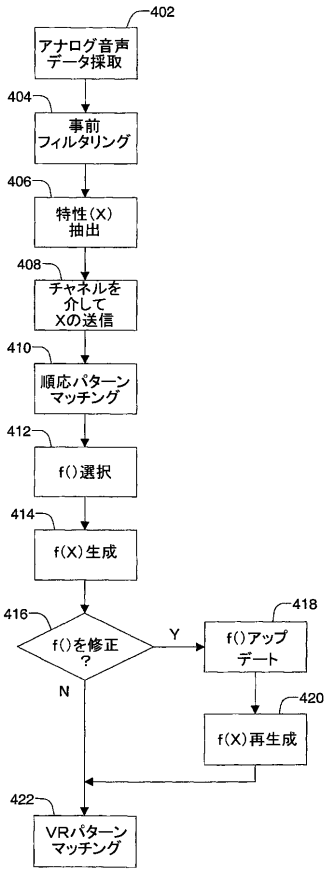
【図２】


204
206
モデム
230
制御プロセッサ
232
VRエンジン
234
メモリ
236
中央音響モデル
238
202
218
モデム
f(X)
順応エンジン
224
順応モデル
228
制御プロセッサ
222
メモリ
227
X
216
AFE
212
ADC
210
MIC
214
PEフィルタ
VRエンジン
220
音響モデル
226

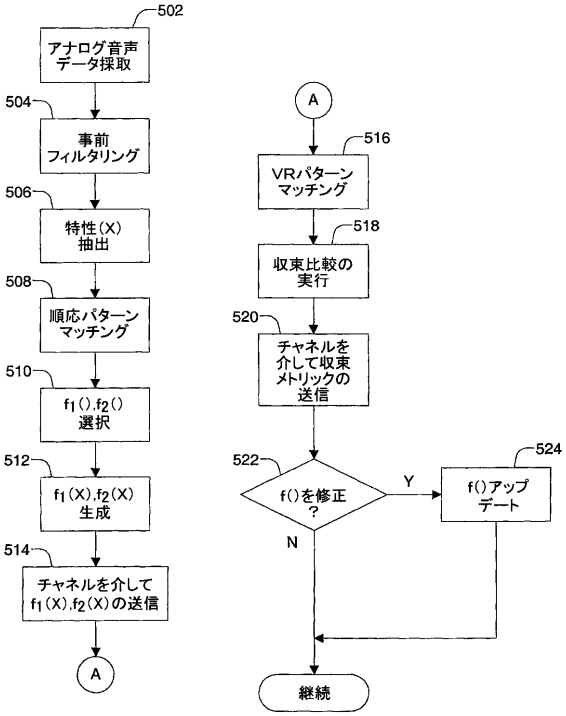
【図３】


302
アナログ音声データ採取
304
事前フィルタリング
306
特性(X)抽出
308
順応パターンマッチング
310
f()選択
312
f(X)生成
314
f()を修正?
Y
N
316
f()アップデート
318
f(X)再生成
320
チャネルを介してf(X)の送信
322
VRパターンマッチング

【図４】


402
アナログ音声データ採取
404
事前フィルタリング
406
特性(X)抽出
408
チャネルを介してXの送信
410
順応パターンマッチング
412
f()選択
414
f(X)生成
416
f()を修正?
Y
N
418
f()アップデート
420
f(X)再生成
422
VRパターンマッチング

【図５】


502
アナログ音声データ採取
504
事前フィルタリング
506
特性(X)抽出
508
順応パターンマッチング
510
f1(),f2()選択
512
f1(X),f2(X)生成
514
チャネルを介してf1(X),f2(X)の送信
A
A
516
VRパターンマッチング
518
収束比較の実行
520
チャネルを介して収束メトリックの送信
522
f()を修正?
Y
N
524
f()アップデート
継続

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/065453 A2

(51) International Patent Classification⁷: G10L 15/06

(21) International Application Number: PCT/US02/03014

(22) International Filing Date: 30 January 2002 (30.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/773,831 31 January 2001 (31.01.2001) US

(71) Applicant: QUALCOMM INCORPORATED [US/US]; 5775 Morehouse Drive, San Diego, CA 92121-1714 (US).

(72) Inventors: CHANG, Chienchung; 6076 Via Posada Del Norte, Rancho Santa Fe, CA 92067 (US). MALAYATH, Naren; 10710 Sabre Hill Drive, #229, San Diego, CA 92128 (US). YAFUSO, Byron, Yoshio; 10093 Branford Road, San Diego, CA 92129 (US).

(74) Agents: WADSWORTH, Philip, R. et al.; Qualcomm Incorporated, 5775 Morehouse Drive, San Diego, CA 92121-1714 (US).

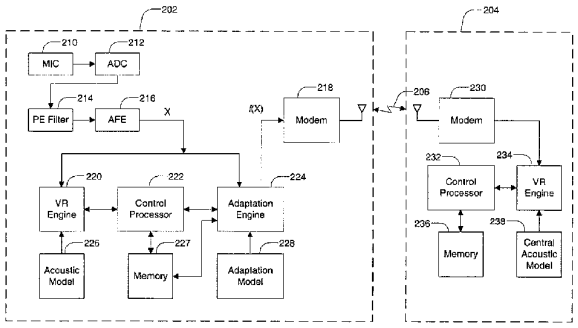
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: DISTRIBUTED VOICE RECOGNITION SYSTEM USING ACOUSTIC FEATURE VECTOR MODIFICATION



(57) Abstract: A voice recognition system applies speaker-dependent modification functions to acoustic feature vectors prior to voice recognition pattern matching against a speaker-independent acoustic model (238). An adaptation engine (224) matches a set of acoustic feature vectors X with an adaptation model (228) to select a speaker-dependent feature vector modification function f(), which is then applied to X to form a modified set of acoustic feature vectors f(X). Voice recognition is then performed by correlating the modified acoustic feature vectors f(X) with a speaker-independent acoustic model (238).

WO 02/065453 A2

WO 02/065453 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DISTRIBUTED VOICE RECOGNITION SYSTEM USING ACOUSTIC FEATURE VECTOR MODIFICATION

BACKGROUND

Field

[1001] The present invention relates to speech signal processing. More particularly, the present invention relates to a novel method and apparatus for distributed voice recognition using acoustic feature vector modification.

Background

[1002] Voice recognition represents one of the most important techniques to endow a machine with simulated intelligence to recognize user voiced commands and to facilitate human interface with the machine. Systems that employ techniques to recover a linguistic message from an acoustic speech signal are called voice recognition (VR) systems. **FIG. 1** shows a basic VR system having a preemphasis filter **102**, an acoustic feature extraction (AFE) unit **104**, and a pattern matching engine **110**. The AFE unit **104** converts a series of digital voice samples into a set of measurement values (for example, extracted frequency components) called an acoustic feature vector. The pattern matching engine **110** matches a series of acoustic feature vectors with the patterns contained in a VR acoustic model **112**. VR pattern matching engines generally employ Viterbi decoding techniques that are well known in the art. When a series of patterns are recognized from the acoustic model **112**, the series is analyzed to yield a desired format of output, such as an identified sequence of linguistic words corresponding to the input utterances.

[1003] The acoustic model **112** may be described as a database of acoustic feature vector extracted from various speech sounds and associated statistical distribution information. These acoustic feature vector patterns correspond to short speech segments such as phonemes, tri-phones and whole-word models. "Training" refers to the process of collecting speech samples of a particular speech segment or syllable from one or more speakers in order to generate

patterns in the acoustic model **112**. "Testing" refers to the process of correlating a series of acoustic feature vectors extracted from end-user speech samples to the contents of the acoustic model **112**. The performance of a given system depends largely upon the degree of correlation between the speech of the end-user and the contents of the database.

[1004] Optimally, the end-user provides speech acoustic feature vectors during both training and testing so that the acoustic model **112** will match strongly with the speech of the end-user. However, because an acoustic model **112** must generally represent patterns for a large number of speech segments, it often occupies a large amount of memory. Moreover, it is not practical to collect all the data necessary to train the acoustic models from all possible speakers. Hence, many existing VR systems use acoustic models that are trained using the speech of many representative speakers. Such acoustic models are designed to have the best performance over a broad number of users, but are not optimized to any single user. In a VR system that uses such an acoustic model, the ability to recognize the speech of a particular user will be inferior to that of a VR system using an acoustic model optimized to the particular user. For some users, such as users having a strong foreign accent, the performance of a VR system using a shared acoustic model can be so poor that they cannot effectively use VR services at all.

[1005] Adaptation is an effective method to alleviate degradations in recognition performance caused by a mismatch in training and test conditions. Adaptation modifies the VR acoustic models during testing to closely match with the testing environment. Several such adaptation schemes, such as maximum likelihood linear regression and Bayesian adaptation, are well known in the art.

[1006] As the complexity of the speech recognition task increases, it becomes increasingly difficult to accommodate the entire recognition system in a wireless device. Hence, a shared acoustic model located in a central communications center provides the acoustic models for all users. The central base station is also responsible for the computationally expensive acoustic matching. In distributed VR systems, the acoustic models are shared by many speakers and hence cannot be optimized for any individual speaker. There is

therefore a need in the art for a VR system that has improved performance for multiple individual users while minimizing the required computational resources.

SUMMARY

[1007] The methods and apparatus disclosed herein are directed to a novel and improved distributed voice recognition system in which speaker-dependent processing is used to transform acoustic feature vectors prior to voice recognition pattern matching. The speaker-dependent processing is performed according to a transform function that has parameters that vary based on the speaker, the results of an intermediate pattern matching process using an adaptation model, or both. The speaker-dependent processing may take place in a remote station, in a communications center, or a combination of the two. Acoustic feature vectors may also be transformed using environment-dependent processing prior to voice recognition pattern matching. The acoustic feature vectors may be modified to adapt to changes in the operating acoustic environment (ambiant noise, frequency response of the microphone etc.). The environment-dependent processing may also take place in a remote station, in a communications center, or a combination of the two.

[1008] The word "exemplary" is used herein to mean "serving as an example, instance, or illustration." Any embodiment described as an "exemplary embodiment" is not necessarily to be construed as being preferred or advantageous over another embodiment.

BRIEF DESCRIPTION OF THE DRAWINGS

[1009] The features, objects, and advantages of the presently disclosed method and apparatus will become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

[1010] **FIG. 1** shows a basic voice recognition system;

[1011] **FIG. 2** shows a distributed VR system according to an exemplary embodiment;

[1012] **FIG. 3** is a flowchart showing a method for performing distributed VR wherein acoustic feature vector modification and selection of feature vector modification functions occur entirely in the remote station;

[1013] **FIG. 4** is a flowchart showing a method for performing distributed VR wherein acoustic feature vector modification and selection of feature vector modification functions occur entirely in the communications center; and

[1014] **FIG. 5** is a flowchart showing a method for performing distributed VR wherein a central acoustic model is used to optimize feature vector modification functions or adaptation models.

DETAILED DESCRIPTION

[1015] In a standard voice recognizer, either in recognition or in training, most of the computational complexity is concentrated in the pattern matching subsystem of the voice recognizer. In the context of wireless systems, voice recognizers are implemented as distributed systems in order to minimize the over-the-air bandwidth consumed by the voice recognition application. Additionally, distributed VR systems avoid performance degradation that can result from lossy source coding of voice data, such as often occurs with the use of vocoders. Such a distributed architecture is described in detail in U.S. Patent No. 5,956,683, entitled "DISTRIBUTED VOICE RECOGNITION SYSTEM" and assigned to the assignee of the present invention, and referred to herein as the '683 patent.

[1016] In an exemplary wireless communication system, such as a digital wireless phone system, a user's voice signal is received through a microphone within a mobile phone or remote station. The analog voice signal is then digitally sampled to produce a digital sample stream, for example 8000 8-bit speech samples per second. Sending the speech samples directly over a wireless channel is very inefficient, so the information is generally compressed before transmission. Through a technique called vocoding, a vocoder

compresses a stream of speech samples into a series of much smaller vocoder packets. The smaller vocoder packets are then sent through the wireless channel instead of the speech samples they represent. The vocoder packets are then received by the wireless base station and de-vocoded to produce a stream of speech samples that are then presented to a listener through a speaker.

[1017] A main objective of vocoders is to compress the speaker's speech samples as much as possible, while preserving the ability for a listener to understand the speech when de-vocoded. Vocoder algorithms are typically lossy compression algorithms, such that the de-vocoded speech samples do not exactly match the samples originally vocoded. Furthermore, vocoder algorithms are often optimized to produce intelligible de-vocoded speech even if one or more vocoder packets are lost in transmission through the wireless channel. This optimization can lead to further mismatches between the speech samples input into the vocoder and those resulting from de-vocoding. The alteration of speech samples that results from vocoding and de-vocoding generally degrades the performance of voice recognition algorithms, though the degree of degradation varies greatly among different vocoder algorithms.

[1018] In a system described in the '683 patent, the remote station performs acoustic feature extraction and sends acoustic feature vectors instead of vocoder packets over the wireless channel to the base station. Because acoustic feature vectors occupy less bandwidth than vocoder packets, they can be transmitted through the same wireless channel with added protection from communication channel errors (for example, using forward error correction (FEC) techniques). VR performance even beyond that of the fundamental system described in the '683 patent can be realized when the feature vectors are further optimized using speaker-dependent feature vector modification functions as described below.

[1019] **FIG. 2** shows a distributed VR system according to an exemplary embodiment. Acoustic feature extraction (AFE) occurs within a remote station **202**, and acoustic feature vectors are transmitted through a wireless channel **206** to a base station and VR communications center **204**. One skilled in the art

will recognize that the techniques described herein may be equally applied to a VR system that does not involve a wireless channel.

[1020] In the embodiment shown, voice signals from a user are converted into electrical signals in a microphone (MIC) **210** and converted into digital speech samples in an analog-to-digital converter (ADC) **212**. The digital sample stream is then filtered using a preemphasis (PE) filter **214**, for example a finite impulse response (FIR) filter that attenuates low-frequency signal components.

[1021] The filtered samples are then analyzed in an AFE unit **216**. The AFE unit **216** converts digital voice samples into acoustic feature vectors. In an exemplary embodiment, the AFE unit **216** performs a Fourier Transform on a segment of consecutive digital samples to generate a vector of signal strengths corresponding to different frequency bins. In an exemplary embodiment, the frequency bins have varying bandwidths in accordance with a bark scale. In a bark scale, the bandwidth of each frequency bin bears a relation to the center frequency of the bin, such that higher-frequency bins have wider frequency bands than lower-frequency bins. The bark scale is described in Rabiner, L. R. and Juang, B. H., *Fundamentals of Speech Recognition*, Prentice Hall, 1993 and is well known in the art.

[1022] In an exemplary embodiment, each acoustic feature vector is extracted from a series of speech samples collected over a fixed time interval. In an exemplary embodiment, these time intervals overlap. For example, acoustic features may be obtained from 20-millisecond intervals of speech data beginning every ten milliseconds, such that each two consecutive intervals share a 10-millisecond segment. One skilled in the art would recognize that the time intervals might instead be non-overlapping or have non-fixed duration without departing from the scope of the embodiments described herein.

[1023] Each acoustic feature vector (identified as X in **FIG. 2**) generated by the AFE unit **216** is provided to an adaptation engine **224**, which performs pattern matching to characterize the acoustic feature vector based on the contents of an adaptation model **228**. Based on the results of the pattern matching, the adaptation engine **224** selects one of a set of feature vector modification functions *f*() from a memory **227** and uses it to generate a modified acoustic feature vector *f*(X).

[1024] X is used herein to describe either a single acoustic feature vector or a series of consecutive acoustic feature vectors. Similarly, *f*(X) is used to describe a single modified acoustic feature vector or a series of consecutive modified acoustic feature vectors.

[1025] In an exemplary embodiment, and as shown in **FIG. 2**, the modified vector *f*(X) is then modulated in a wireless modem **218**, transmitted through a wireless channel **206**, demodulated in a wireless modem **230** within a communications center **204**, and matched against a central acoustic model **238** by a central VR engine **234**. The wireless modems **218**, **230** and wireless channel **206** may use any of a variety of wireless interfaces including CDMA, TDMA, or FDMA. In addition, the wireless modems **218**, **230** may be replaced with other types of communications interfaces that communicate over a non-wireless channel without departing from the scope of the described embodiments. For example, the remote station **202** may communicate with the communications center **204** through any of a variety of types of communications channel including land-line modems, T1/E1, ISDN, DSL, ethernet, or even traces on a printed circuit board (PCB).

[1026] In an exemplary embodiment, the vector modification function *f*() is optimized for a specific user or speaker, and is designed to maximize the probability that speech will be correctly recognized when matched against the central acoustic model **238**, which is shared between multiple users. The adaptation model **228** in the remote station **202** is much smaller than the central acoustic model **238**, making it possible to maintain a separate adaptation model **228** that is optimized for a specific user. Also, the parameters of the feature vector modification functions *f*() for one or more speakers are small enough to store in the memory **227** of the remote station **202**.

[1027] In an alternate embodiment, an additional set of parameters for environment-dependent feature vector modification functions are also stored in the memory **227**. The selection and optimization of environment-dependent feature vector modification functions are more global in nature, and so may generally be performed during each call. An example of a very simple environment-dependent feature vector modification function is applying a

constant gain *k* to each element of each acoustic feature vector to adapt to a noisy environment.

[1028] A vector modification function *f*() may have any of several forms. For example, a vector modification function *f*() may be an affine transform of the form AX + b. Alternatively, a vector modification function *f*() may be a set of finite impulse response (FIR) filters initialized and then applied to a set of consecutive acoustic feature vectors. Other forms of vector modification function *f*() will be obvious to one of skill in the art and are therefore within the scope of the embodiments described herein.

[1029] In an exemplary embodiment, a vector modification function *f*() is selected based on a set of consecutive acoustic feature vectors. For example, the adaptation engine **224** may apply Viterbi decoding or trellis decoding techniques in order to determine the degree of correlation between a stream of acoustic feature vectors and the multiple speech patterns in the adaptation model **228**. Once a high degree of correlation is detected, a vector modification function *f*() is selected based on the detected pattern and applied to the corresponding segment from the stream of acoustic feature vectors. This approach requires that the adaptation engine **224** store a series of acoustic feature vectors and perform pattern matching of the series against the adaptation model **228** before selecting the *f*() to be applied to each acoustic feature vector. In an exemplary embodiment, the adaptation engine maintains an elastic buffer of unmodified acoustic feature vectors, and then applies the selected *f*() to the contents of the elastic buffer before transmission. The contents of the elastic buffer are compared to the patterns in the adaptation model **228**, and a maximum correlation metric is generated for the pattern having the highest degree of correlation with the contents of the elastic buffer. This maximum correlation is compared against one or more thresholds. If the maximum correlation exceeds a detection threshold, then the *f*() corresponding to the pattern associated with the maximum correlation is applied to the acoustic feature vectors in the buffer and transmitted. If the elastic buffer becomes full before the maximum correlation exceeds the detection threshold, then the contents of the elastic buffer are transmitted without modification or alternatively modified using a default *f*().

[1030] The speaker-dependent optimization of *f*() may be accomplished in any of a number of ways. In a first exemplary embodiment, a control processor **222** monitors the degree of correlation between user speech and the adaptation model **228** over multiple utterances. When the control processor **222** determines that a change in *f*() would improve VR performance, it modifies the parameters of *f*() and stores the new parameters in the memory **227**. Alternatively, the control processor **222** may modify the adaptation model **228** directly in order to improve VR performance.

[1031] As shown in **FIG. 2**, the remote station **202** may additionally include a separate VR engine **220** and a remote station acoustic model **226**. Because of limited memory capacity, the remote station acoustic model **226** in a remote station **202** such as a wireless phone must generally be small and therefore limited to a small number of phrases or phonemes. On the other hand, because it is contained within a remote station used by a small number of users, the remote station acoustic model **226** can be optimized to one or more specific users for improved VR performance. For example, speech patterns for words like "call" and each of the ten digits may be tailored to the owner of the wireless phone. Such a local remote station acoustic model **226** enables a remote station **202** to have very good VR performance for a small set of words. Furthermore, a remote station acoustic model **226** enables the remote station **202** to accomplish VR without establishing a wireless link to the communications center **204**.

[1032] The optimization of *f*() may occur through either supervised or unsupervised learning. Supervised learning generally refers to training that occurs with a user uttering a predetermined word or sentence that is used to accurately optimize a remote station acoustic model. Because the VR system has a priori knowledge of the word or sentence used as input, there is no need to perform VR during supervised learning to identify the predetermined word or sentence. Supervised learning is generally considered the most accurate way to generate an acoustic model for a specific user. An example of supervised learning is when a user first programs the speech for the ten digits into a remote station acoustic model **226** of a remote station **202**. Because the remote station **202** has a priori knowledge of the speech pattern corresponding to the spoken

digits, the remote station acoustic model **226** can be tailored to the particular user with less risk of degrading VR performance.

[1033] In contrast to supervised learning, unsupervised learning occurs without the VR system having a priori knowledge of the speech pattern or word being uttered. Because of the risk of matching an utterance to an incorrect speech pattern, modification of a remote station acoustic model based on unsupervised learning must be done in a much more conservative fashion. For example, many past utterances may have occurred that were similar to each other and closer to one speech pattern in the acoustic model than any other speech patterns. If all of those past utterances would be correctly matched to the one speech pattern in the model, that one speech pattern in the acoustic model could be modified to more closely match the set of similar utterances. However, if many of those past utterances do not correspond to the one speech pattern in the model, then modifying that one speech pattern would degrade VR performance. Optimally, the VR system can collect feedback from the user on the accuracy of past pattern matching, but such feedback is often not available.

[1034] Unfortunately, supervised learning is tedious for the user, making it impractical for generating an acoustic model having a large number of speech patterns. However, supervised learning may still be useful in optimizing a set of vector modification functions *f*(), or even in optimizing the more limited speech patterns in an adaptation model **228**. The differences in speech patterns caused by a user's strong accent is an example of an application in which supervised learning may be required. Because acoustic feature vectors may require significant modification to compensate for an accent, the need for accuracy in those modifications is great.

[1035] Unsupervised learning may also be used to optimize vector modification functions *f*() for a specific user where optimizations are less likely to be a direct cause of VR errors. For example, the adjustment in a vector modification function *f*() needed to adapt to a speaker having a longer vocal-tract length or average vocal pitch is more global in nature than the adjustments required to compensate for an accent. More inaccuracy in such global vector modifications may be made without drastically impacting VR effectiveness.

[1036] Generally, the adaptation engine **224** uses the small adaptation model **228** only to select a vector modification function *f*(), and not to perform complete VR. Because of its small size, the adaptation model **228** is similarly unsuitable for performing training to optimize either the adaptation model **228** or the vector modification function *f*(). An adjustment in the adaptation model **228** or vector modification function *f*() that appears to improve the degree of matching of a speaker's voice data against the adaptation model **228** may actually degrade the degree of matching against the larger central acoustic model **238**. Because the central acoustic model **238** is the one actually used for VR, such an adjustment would be a mistake rather than an optimization.

[1037] In an exemplary embodiment, the remote station **202** and the communications center **204** collaborate when using unsupervised learning to modify either the adaptation model **228** or the vector modification function *f*(). A decision of whether to modify either the adaptation model **228** or the vector modification model *f*() is made based on improved matching against the central acoustic model **238**. For example, the remote station **202** may send multiple sets of acoustic feature vectors, the unmodified acoustic feature vectors X and the modified acoustic feature vectors *f*(X), to the communications center **204**. Alternatively, the remote station **202** may send modified acoustic feature vectors $f_1(X)$ and $f_2(X)$, where $f_2()$ is a tentative, improved feature vector modification function. In another embodiment, the remote station **202** sends X, and parameters for both feature vector modification functions $f_1()$ and $f_2()$. The remote station **202** may send the multiple sets decision of whether to send the second set of information to the communications center **204** may be based on a fixed time interval,

[1038] Upon receiving multiple sets of acoustic feature information, whether modified acoustic feature vectors or parameters for feature vector modification functions, the communications center **204** evaluates the degree of matching of the resultant modified acoustic feature vectors using its own VR engine **234** and the central acoustic model **238**. The communications center **204** then sends information back to the remote station **202** indicating whether a change would result in improved VR performance. For example, the communications center **204** sends a speech pattern correlation metric for each set of acoustic

feature vectors to the remote station **202**. The speech pattern correlation metric for a set of acoustic feature vectors indicates the degree of correlation between a set of acoustic feature vectors and the contents of the central acoustic model **238**. Based on the comparative degree of correlation between the two sets of vectors, the remote station **202** may adjust its adaptation model **228** or may adjust one or more feature vector modification functions *f*(). The remote station **202** may specify the use of either set of vectors to be used for actual recognition of words, or the communications center **204** may select the set of vectors based on their correlation metrics. In an alternate embodiment, the remote station **202** identifies the set of acoustic feature vectors to be used for VR after receiving the resulting correlation metrics from the communications center **204**.

[1039] In an alternate embodiment, the remote station **202** uses its local adaptation engine **224** and adaptation model **228** to identify a feature vector modification function *f*(), and sends the unmodified acoustic feature vectors X along with *f*() to the communications center **204**. The communications center **204** then applies *f*() to X and performs testing using both modified and unmodified vectors. The communications center **204** then sends the results of the testing back to the remote station **202** to enable more accurate adjustments of the feature vector modification functions by the remote station **202**.

[1040] In another embodiment, the adaptation engine **224** and the adaptation model **228** are incorporated into the communications center **204** instead of the remote station **202**. A control processor **232** within the communications center **204** receives a stream of unmodified acoustic feature vectors through the modem **230** and presents them to an adaptation engine and adaptation model within the communications center **204**. Based on the results of this intermediate pattern matching, the control processor **232** selects a feature vector modification function *f*() from a database stored in a communications center memory **236**. In an exemplary embodiment, the communications center memory **236** includes sets of feature vector modification functions *f*() corresponding to specific users. This may be either in addition to or in lieu of feature vector modification function information stored in the remote station **202** as described above. The communications center **204** can use any of a variety of types of speaker identification information to identify the particular

speaker providing the voice data from which the feature vectors are extracted. For example, the speaker identification information used to select a set of feature vector modification functions may be the mobile identification number (MIN) of the wireless phone on the opposite end of the wireless channel **206**. Alternatively, the user may enter a password to identify himself for the purposes of enhanced VR services. Additionally, environment-dependent feature vector modification functions may be adapted and applied during a wireless phone call based on measurements of the speech data. Many other methods may also be used to select a set of speaker-dependent vector modification functions without departing from the scope of the embodiments described herein.

[1041] One skilled in the art would also recognize that the multiple pattern matching engines **220**, **224** within the remote station **202** may be combined without departing from the scope of the embodiments described herein. In addition, the different acoustic models **226**, **228** in the remote station **202** may be similarly combined. Furthermore, one or more of the pattern matching engines **220**, **224** may be incorporated into the control processor **222** of the remote station **202**. Also, one or more of the acoustic models **226**, **228** may be incorporated into the memory **227** used by the control processor **222**.

[1042] In the communications center **204**, the central speech pattern matching engine **234** may be combined with an adaptation engine (not shown), if present, without departing from the scope of the embodiments described herein. In addition, the central acoustic models **238** may be combined with an adaptation model (not shown). Furthermore, either or both of the central speech pattern matching engine **234** and the adaptation engine (not shown), if present in the communications center **204**, may be incorporated into the control processor **232** of the communications center **204**. Also, either or both of the central acoustic model **238** and the adaptation model (not shown), if present in the communications center **204**, may be incorporated into the control processor **232** of the communications center **204**.

[1043] **FIG. 3** is a flowchart of a method for performing distributed VR where modifications of X and f() occur entirely in the remote station **202** based on convergence with a remote adaptation model. At step **302**, the remote station **202** samples the analog voice signals from a microphone to produce a stream

of digital voice samples. At step **304**, the speech samples are then filtered, for example using a preemphasis filter as described above. At step **306**, a stream of acoustic feature vectors X is extracted from the filtered speech samples. As described above, the acoustic feature vectors may be extracted from either overlapping or non-overlapping intervals of speech samples that are either fixed or variable in duration.

[1044] At step **308**, the remote station **202** performs pattern matching to determine the degree of correlation between the stream of acoustic feature vectors and multiple patterns contained in an adaptation model (such as **228** in **FIG. 2**). At step **310**, the remote station **202** selects the pattern in the adaptation model that most closely matches the stream of acoustic feature vectors X. The selected pattern is called the target pattern. As discussed above, the degree of correlation between X and the target pattern may be compared against a detection threshold. If the degree of correlation is greater than the detection threshold, then the remote station **202** selects a feature vector modification function *f*() that corresponds to the target pattern. If the degree of correlation is less than the detection threshold, then the remote station **202** selects either an acoustic feature vector identity function *f*() such that *f*(X)=X, or selects some default *f*(). In an exemplary embodiment, remote station **202** selects a feature vector modification function *f*() from a local database of feature vector modification functions corresponding to various patterns in its local adaptation model. The remote station **202** applies the selected feature vector modification function *f*() to the stream of acoustic feature vectors X at step **312**, thus producing *f*(X).

[1045] In an exemplary embodiment, the remote station **202** generates a correlation metric that indicates the degree of correlation between X and the target pattern. The remote station **202** also generates a correlation metric that indicates the degree of correlation between *f*(X) and the target pattern. In an example of unsupervised learning, the remote station **202** uses the two correlation metrics along with past correlation metric values to determine, at step **314**, whether to modify one or more feature vector modification functions *f*(). If a determination is made at step **314** to modify *f*(), then *f*() is modified at step **316**. In an exemplary embodiment, the modified *f*() is immediately applied

to X at step **318** to form a new modified acoustic feature vector *f*(X). In an alternate embodiment, step **318** is omitted, and a new feature vector modification function *f*() does not take effect until a later set of acoustic feature vectors X.

[1046] If a determination is made at step **314** not to modify *f*(), or after steps **316** and **318**, the remote station **202** transmits the current *f*(X) through the wireless channel **206** to the communications center **204** at step **320**. VR pattern matching then takes place within the communications center **204** at step **322**.

[1047] In an alternate embodiment, the communications center **204** generates speech pattern correlation metrics during the VR pattern matching step **322** and sends these metrics back to the remote station **302** to aid in optimizations of *f*(). The speech pattern correlation metrics may be formatted in any of several ways. For example, the communications center **204** may return an acoustic feature vector modification error function $f_E()$ that can be applied to *f*(X) to create an exact correlation with a pattern found in a central acoustic model. Alternatively, the communications center **204** could simply return a set of acoustic feature vectors corresponding to a target pattern or patterns in the central acoustic model found to have the highest degree of correlation with *f*(X). Or, the communications center **204** could return the branch metric derived from the hard-decision or soft-decision Viterbi decoding process used to select the target pattern. The speech pattern correlation metrics could also include a combination of these types of information. This returned information is then used by the remote station **202** in optimizing *f*(). In an exemplary embodiment, re-generation of *f*(X) at step **318** is omitted, and the remote station **202** performs modifications of *f*() (steps **314** and **316**) after receiving feedback from the communications center **204**.

[1048] **FIG. 4** is a flowchart showing a method for performing distributed VR where modifications of X and *f*() occur entirely in the communications center **204** based on correlation with a central acoustic model. At step **402**, the remote station **202** samples the analog voice signals from a microphone to produce a stream of digital voice samples. At step **404**, the speech samples are then filtered, for example using a preemphasis filter as described above. At

step **406**, a stream of acoustic feature vectors X is extracted from the filtered speech samples. As described above, the acoustic feature vectors may be extracted from either overlapping or non-overlapping intervals of speech samples that are either fixed or variable in duration.

[1049] At step **408**, the remote station **202** transmits the unmodified stream of acoustic feature vectors X through the wireless channel **206**. At step **410**, the communications center **204** performs adaptation pattern matching. As discussed above, adaptation pattern matching may be accomplished using either a separate adaptation model or using a large central acoustic model **238**. At step **412**, the communications center **204** selects the pattern in the adaptation model that most closely matches the stream of acoustic feature vectors X. The selected pattern is called the target pattern. As described above, if the correlation between X and the target pattern exceeds a threshold, an *f*() is selected that corresponds to the target pattern. Otherwise, a default *f*() or a null *f*() is selected. At step **414**, the selected feature vector modification function *f*() is applied to the stream of acoustic feature vectors X to form a modified stream of acoustic feature vectors *f*(X).

[1050] In an exemplary embodiment, a feature vector modification function *f*() is selected from a subset of a large database of feature vector modification functions residing within the communications center **204**. The subset of feature vector modification functions available for selection are speaker-dependent, such that pattern matching using a central acoustic model (such as **238** in **FIG. 2**) will be more accurate using *f*(X) as input than X. As described above, examples of how the communications center **204** may select a speaker-dependent subset of feature vector modification functions include use of a MIN of the speaker's wireless phone or a password entered by a speaker.

[1051] In an exemplary embodiment, the communications center **204** generates correlation metrics for the correlation between X and the target pattern and between *f*(X) and the target pattern. The communications center **204** then uses the two correlation metrics along with past correlation metric values to determine, at step **416**, whether to modify one or more feature vector modification functions *f*(). If a determination is made at step **416** to modify *f*(), then *f*() is modified at step **418**. In an exemplary embodiment, the modified *f*() is

immediately applied to X at step **420** to form a new modified acoustic feature vector *f*(X). In an alternate embodiment, step **420** is omitted, and a new feature vector modification function *f*() does not take effect until a later set of acoustic feature vectors X.

[1052] If a determination is made at step **416** not to modify *f*(), or after steps **418** and **420**, the communications center **204** performs VR pattern matching at step **422** using a central acoustic model **238**.

[1053] **FIG. 5** is a flowchart showing a method for performing distributed VR wherein a central acoustic model within the communications center **204** is used to optimize feature vector modification functions or adaptation models. In an exemplary embodiment, the remote station **202** and the communications center **204** exchange information as necessary and collaborate to maximize the accuracy of optimizations of feature vector modification functions.

[1054] At step **502**, the remote station **202** samples the analog voice signals from a microphone to produce a stream of digital voice samples. At step **504**, the speech samples are then filtered, for example using a preemphasis filter as described above. At step **506**, a stream of acoustic feature vectors X is extracted from the filtered speech samples. As described above, the acoustic feature vectors may be extracted from either overlapping or non-overlapping intervals of speech samples that are either fixed or variable in duration.

[1055] At step **508**, the remote station **202** performs pattern matching to determine the degree of correlation between the stream of acoustic feature vectors and multiple patterns contained in an adaptation model (such as **228** in **FIG. 2**). At step **510**, the remote station **202** selects the pattern in the adaptation model that most closely matches the stream of acoustic feature vectors X. The selected pattern is called the target pattern. As described above, if the correlation between X and the target pattern exceeds a threshold, a first feature vector modification function $f_1()$ is selected that corresponds to the target pattern. Otherwise, a default *f*() or a null *f*() is selected. The remote station **202** selects the feature vector modification function *f*() from a local database of feature vector modification functions corresponding to various patterns in its local adaptation model. The remote station **202** applies the

selected feature vector modification function $f()$ to the stream of acoustic feature vectors X at step **512**, thus producing $f(X)$.

[1056] In contrast to the methods described in association with **FIG. 3** and **FIG. 4**, at step **514**, the remote station **202** sends two sets of acoustic feature vectors, $f_1(X)$ and $f_2(X)$, through the channel **206** to the communications center **204**. At step **516**, the communications center **204** performs pattern matching against its central acoustic model using $f_1(X)$ as input. As a result of this VR pattern matching, the communications center **204** identifies a target pattern or set of patterns having the greatest degree of correlation with $f_1(X)$. At step **518**, the communications center **204** generates a first speech pattern correlation metric indicating the degree of correlation between $f_1(X)$ and the target pattern and a second speech pattern correlation metric indicating the degree of correlation between $f_2(X)$ and the target pattern.

[1057] Though both sets of acoustic feature vectors are used for pattern matching against the central acoustic model, only one set is used for actual VR. Thus, the remote station **202** can evaluate the performance of a proposed feature vector modification function without risking an unexpected degradation in performance. Also, the remote station **202** need not rely entirely on its smaller, local adaptation model when optimizing $f()$. In an alternate embodiment, the remote station **202** may use a null function for $f_2()$, such that $f_2(X)=X$. This approach allows the remote station **202** to verify the performance of $f()$ against VR performance achieved without acoustic feature vector modification.

[1058] At step **520**, the communications center **204** sends the two speech pattern correlation metrics back to the remote station **202** through the wireless channel **206**. Based on the received speech pattern correlation metrics, the remote station **202** determines, at step **522**, whether to modify $f_1()$ at step **524**. The determination of whether to modify $f_1(X)$ at step **522** may be based on one set of speech pattern correlation metrics, or may be based on a series of speech pattern correlation metrics associated with the same speech patterns from the local adaptation model. As discussed above, the speech pattern correlation metrics may include such information as an acoustic feature vector modification error function $f_E()$, a set of acoustic feature vectors corresponding

to patterns in the central acoustic model found to have had the highest degree of correlation with *f*(X), or a Viterbi decoding branch metric.

[1059] One skilled in the art will recognize that the techniques described above may be applied equally to any of a variety of types of wireless channel **206**. For example, the wireless channel **206** (and accordingly the modems **218**, **230**) may utilize code division multiple access (CDMA) technology, analog cellular, time division multiple access (TDMA), or other types of wireless channel. Alternatively, the channel **206** may be a type of channel other than wireless, including but not limited to optical, infrared, and ethernet channels. In yet another embodiment, the remote station **202** and communications center **204** are combined into a single system that performs speaker-dependent modification of acoustic feature vectors prior to VR testing using a central acoustic model **238**, obviating the channel **206** entirely.

[1060] Those of skill in the art would understand that information and signals may be represented using any of a variety of different technologies and techniques. For example, data, instructions, commands, information, signals, bits, symbols, and chips that may be referenced throughout the above description may be represented by voltages, currents, electromagnetic waves, magnetic fields or particles, optical fields or particles, or any combination thereof.

[1061] Those of skill would further appreciate that the various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the embodiments disclosed herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, circuits, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present invention.

[1062] The various illustrative logical blocks, modules, and circuits described in connection with the embodiments disclosed herein may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general purpose processor may be a microprocessor, but in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[1063] The steps of a method or algorithm described in connection with the embodiments disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may reside in RAM memory, flash memory, ROM memory, EPROM memory, EEPROM memory, registers, hard disk, a removable disk, a CD-ROM, or any other form of storage medium known in the art. An exemplary storage medium is coupled to the processor such the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. The processor and the storage medium may reside in an ASIC. The ASIC may reside in a remote station. In the alternative, the processor and the storage medium may reside as discrete components in a remote station.

[1064] The previous description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the present invention. Various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without departing from the spirit or scope of the invention.

Thus, the present invention is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

[1065] WHAT IS CLAIMED IS:

CLAIMS

1. A voice recognition system comprising:
an adaptation model containing acoustic pattern information; and
an adaptation engine for performing pattern matching of acoustic feature vectors with the acoustic pattern information to identify a selected feature vector modification function.

2. The voice recognition system of claim 1 wherein the adaptation engine is further configured to apply the selected feature vector modification function to the acoustic feature vectors to produce a set of modified acoustic feature vectors.

3. The voice recognition system of claim 1 further comprising a voice recognition engine for matching the set of modified acoustic feature vectors with an acoustic model.

4. The voice recognition system of claim 1 further comprising a control processor for evaluating the performance of the selected feature vector modification function and adjusting the selected feature vector modification function based on the evaluating.

5. The voice recognition system of claim 1 further comprising a memory for storing at least one set of parameters corresponding to a set of feature vector modification functions, wherein the selected feature vector modification function is a member of the set of feature modification functions.

6. The voice recognition system of claim 5 wherein the memory contains more than one set of parameters corresponding to a set of feature vector modification functions, and wherein each set of parameters corresponds to a specific speaker.

7. The voice recognition system of claim 5 wherein the memory contains more than one set of parameters corresponding to a set of feature vector modification functions, and wherein each set of parameters corresponds to a different acoustic environment.

8. A voice recognition system comprising:

an adaptation model containing acoustic pattern information; and

an adaptation engine for performing pattern matching of acoustic feature vectors with the acoustic pattern information to identify a speaker-dependent feature vector modification function and apply the speaker-dependent feature vector modification function to the acoustic feature vectors.

9. A voice recognition system comprising:

an adaptation engine for analyzing acoustic feature vectors to identify a selected feature vector modification function and applying the selected feature vector modification function to the acoustic feature vectors to produce a set of modified acoustic feature vectors;

an acoustic model; and

a voice recognition engine for matching the set of modified acoustic feature vectors with the acoustic model.

10. The voice recognition system of claim 9 further comprising an adaptation model, wherein said adaptation engine performs the analyzing acoustic feature vectors by matching the acoustic feature vectors with acoustic patterns stored in the adaptation model.

11. The voice recognition system of claim 9 further comprising a control processor for evaluating the performance of the selected feature vector modification function and adjusting the selected feature vector modification function based on the evaluating.

12. The voice recognition system of claim 9 further comprising a memory for storing at least one set of parameters corresponding to a set of feature vector

modification functions, wherein the selected feature vector modification function is a member of the set of feature modification functions.

13. The voice recognition system of claim 12 wherein the memory contains more than one set of parameters corresponding to a set of feature vector modification functions, and wherein each set of parameters corresponds to a specific speaker.

14. The voice recognition system of claim 12 wherein the memory contains more than one set of parameters corresponding to a set of feature vector modification functions, and wherein each set of parameters corresponds to a different acoustic environment.

15. A remote station apparatus comprising:

an adaptation model containing acoustic pattern information; and

an adaptation engine for performing pattern matching of acoustic feature vectors against the acoustic pattern information to identify a selected feature vector modification function, and applying the selected feature vector modification function to the acoustic feature vectors to produce a set of modified acoustic feature vectors.

16. The remote station apparatus of claim 15 further comprising a control processor for evaluating the performance of the selected feature vector modification function and adjusting the selected feature vector modification function based on the evaluating.

17. The remote station apparatus of claim 15 further comprising a memory for storing at least one set of parameters corresponding to a set of feature vector modification functions, wherein the selected feature vector modification function is a member of the set of feature modification functions.

18. The remote station apparatus of claim 17 wherein the memory contains more than one set of parameters corresponding to a set of feature vector

modification functions, and wherein each set of parameters corresponds to a specific speaker.

19. The remote station apparatus of claim 17 wherein the memory contains more than one set of parameters corresponding to a set of feature vector modification functions, and wherein each set of parameters corresponds to a different acoustic environment.

20. The remote station apparatus of claim 15 further comprising a communications interface for communicating the modified acoustic feature vectors to a communications center.

21. A voice recognition communication center apparatus comprising:

a communications interface for receiving acoustic feature vectors, wherein the acoustic feature vectors have been modified using a feature vector modification function;

an acoustic model containing acoustic patterns, wherein the acoustic model is not trained to a single speaker;

a voice recognition engine for matching the set of modified acoustic feature vectors with the acoustic model; and

a control processor for evaluating the feature modification function based on the matching.

22. A voice recognition communication center apparatus comprising:

a communications interface for receiving acoustic feature vectors and speaker identification information from at least one remote station;

a memory containing speaker-dependent feature vector modification function parameters; and

an adaptation engine for performing pattern matching of the acoustic feature vectors with an acoustic model, identifying a selected speaker-dependent feature vector modification function based on the pattern matching and the speaker identification information, and applying the selected speaker-

dependent feature vector modification function to the acoustic feature vectors to produce a set of modified acoustic feature vectors.

23. The voice recognition communication center apparatus of claim 22 wherein the adaptation engine further comprises a voice recognition engine for matching the set of modified acoustic feature vectors with a central acoustic model, wherein the central acoustic model is not trained to a single speaker.

24. The voice recognition communication center apparatus of claim 22 further comprising a central acoustic model, wherein the central acoustic model is not trained to a single speaker, and wherein the adaptation engine is further configured to perform pattern matching of the set of modified acoustic feature vectors with the central acoustic model.

25. The voice recognition communication center apparatus of claim 22 further comprising a voice recognition engine for matching the set of modified acoustic feature vectors with a central acoustic model.

26. The voice recognition communication center apparatus of claim 22 further comprising a control processor for evaluating the performance of the selected speaker-dependent feature vector modification function and adjusting the parameters of the selected feature speaker-dependent vector modification function in the memory based on the evaluating.

27. A method of performing voice recognition comprising:

extracting acoustic feature vectors;

performing adaptation pattern matching of the acoustic feature vectors with an adaptation model;

selecting a feature vector modification function based on the adaptation pattern matching;

applying the selected feature vector modification function to the acoustic feature vectors to form a set of modified acoustic feature vectors; and

performing voice recognition pattern matching of the set of modified acoustic feature vectors with an acoustic model.

28. The method of claim 27 wherein the feature vector modification function is selected from a speaker-dependent set of feature vector modification functions.

29. The method of claim 28 further comprising modifying the speaker-dependent set of feature vector modification functions based on the adaptation pattern matching.

30. The method of claim 28 further comprising modifying the speaker-dependent set of feature vector modification functions based on the voice recognition pattern matching.

31. The method of claim 27 wherein the feature vector modification function is specific to an acoustic environment.

32. A method of performing voice recognition comprising:

at a remote station, performing adaptation pattern matching of acoustic feature vectors with an adaptation model stored in the remote station;

at the remote station, selecting a feature vector modification function from feature vector modification function information stored at the remote station, wherein the selecting a feature vector modification function is based on the pattern matching;

at the remote station, applying the selected feature vector modification function to the acoustic feature vectors to form a set of modified acoustic feature vectors; and

sending the modified acoustic feature vectors from the remote station to a communications center.

33. The method of claim 32 wherein the feature vector modification function information is speaker-dependent.

34. The method of claim 32 further comprising modifying the feature vector modification function information based on the adaptation pattern matching.

35. The method of claim 32 further comprising modifying the feature vector modification function information at the remote station based on information received from the communications center.

36. The method of claim 32 further comprising performing voice recognition at the remote station, wherein the performing voice recognition comprises performing voice recognition pattern matching of the modified acoustic feature vectors with an acoustic model stored in the remote station.

37. The method of claim 36 further comprising modifying the feature vector modification function information at the remote station based on the voice recognition pattern matching.

38. The method of claim 32 wherein the feature vector modification function information is environment-dependent.

39. A method of performing voice recognition in a system comprising at least one remote station and a communications center, the method comprising:

at the remote station, extracting acoustic feature vectors;

sending the acoustic feature vectors from the remote station to the communications center;

at the communications center, performing adaptation pattern matching of the acoustic feature vectors with an adaptation model stored in the communications center;

at the communications center, selecting a feature vector modification function from a database stored in the communications center, wherein the selecting a feature vector modification function is based on the adaptation pattern matching;

at the communications center, applying the selected feature vector modification function to the acoustic feature vectors to form a set of modified acoustic feature vectors; and

at the communications center, performing voice recognition pattern matching of the set of modified acoustic feature vectors with an acoustic model stored in the communications center.

40. The method of claim 39 further comprising, at the communications center, selecting a speaker-dependent set of feature vector modification functions, wherein the selected feature vector modification function is selected from the speaker-dependent set of feature vector modification functions.

41. The method of claim 40 further comprising modifying the speaker-dependent set of feature vector modification functions based on the adaptation pattern matching.

42. The method of claim 40 further comprising modifying the speaker-dependent set of feature vector modification functions based on the voice recognition pattern matching.

43. The method of claim 40 further comprising sending, from the remote station to the communications center, speaker identification information, wherein the selecting a speaker-dependent set of feature vector modification functions is based on the speaker identification information.

44. A method of performing voice recognition in a system comprising at least one remote station and a communications center, the method comprising:

at the remote station, extracting unmodified acoustic feature vectors;

at the remote station, performing adaptation pattern matching of the unmodified acoustic feature vectors with an adaptation model stored in the remote station;

at the remote station, selecting a speaker-dependent feature vector modification function based on the adaptation pattern matching;

at the remote station, applying the selected speaker-dependent feature vector modification function to the acoustic feature vectors to form a set of modified acoustic feature vectors;

sending the modified acoustic feature vectors from the remote station to the communications center;

at the communications center, performing voice recognition pattern matching of the set of modified acoustic feature vectors with an acoustic model stored in the communications center.

45. The method of claim 44 further comprising modifying the selected speaker-dependent feature vector modification function based on the adaptation pattern matching.

46. The method of claim 44 further comprising:

sending the unmodified acoustic feature vectors from the remote station to the communications center;

at the communications center, analyzing the selected speaker-dependent feature vector modification function using the modified acoustic feature vectors and the unmodified acoustic feature vectors; and

at the remote station, modifying the selected speaker-dependent feature vector modification function based on the analyzing.

47. The method of claim 44 further comprising modifying the speaker-dependent set of feature vector modification functions based on the voice recognition pattern matching.

48. A voice recognition system comprising:

means for extracting acoustic feature vectors;

means for performing adaptation pattern matching of the acoustic feature vectors with an adaptation model;

means for selecting a feature vector modification function based on the adaptation pattern matching;

means for applying the selected feature vector modification function to the acoustic feature vectors to form a set of modified acoustic feature vectors; and

means for performing voice recognition pattern matching of the set of modified acoustic feature vectors with an acoustic model.

49. A remote station apparatus comprising:

means for performing adaptation pattern matching of acoustic feature vectors with an adaptation model stored in the remote station;

means for selecting a feature vector modification function from feature vector modification function information stored at the remote station, wherein the selecting a feature vector modification function is based on the pattern matching;

means for applying the selected feature vector modification function to the acoustic feature vectors to form a set of modified acoustic feature vectors; and

means for sending the modified acoustic feature vectors to a communications center.

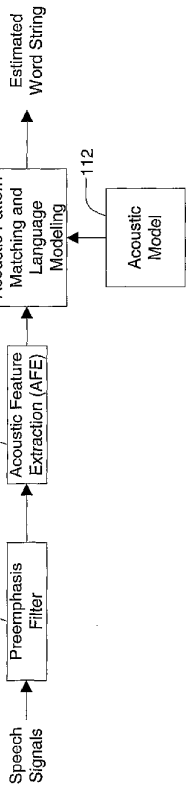

Estimated Word String
112
110
Acoustic Pattern Matching and Language Modeling
Acoustic Model
104
Acoustic Feature Extraction (AFE)
102
Preemphasis Filter
Speech Signals

FIG. 1

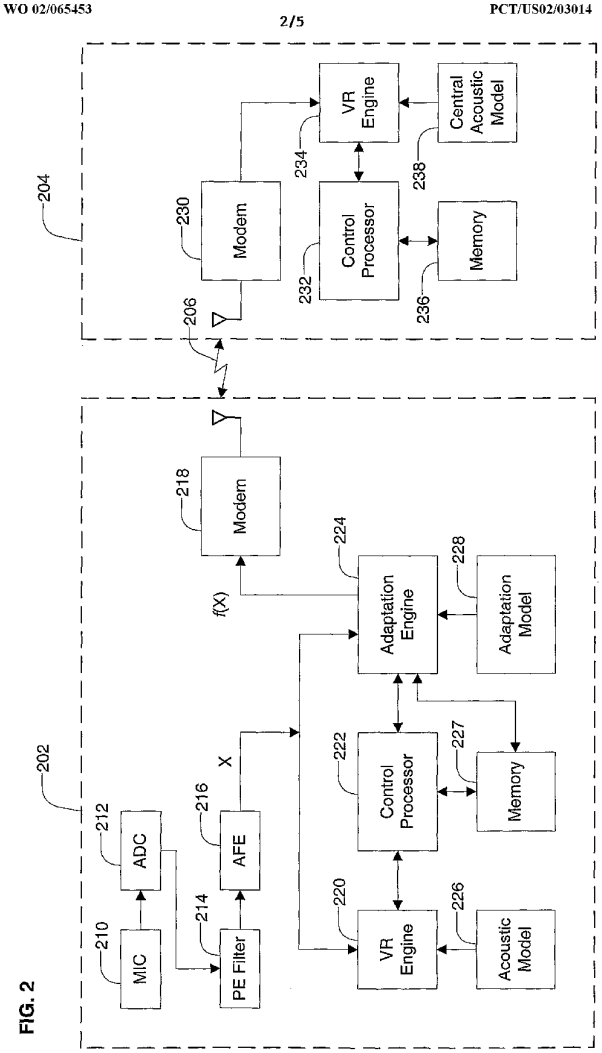

204
VR Engine
Central Acoustic Model
234
238
230
Modem
Control Processor
Memory
232
236
206
218
Modem
f(X)
224
Adaptation Engine
228
Adaptation Model
202
X
222
Control Processor
227
Memory
212
ADC
216
AFE
220
VR Engine
226
Acoustic Model
210
MIC
214
PE Filter

FIG. 2

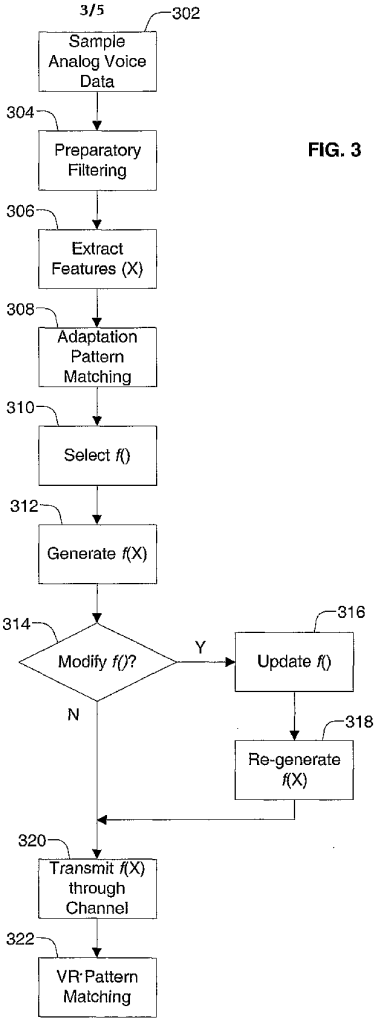

302
Sample Analog Voice Data
304
Preparatory Filtering
306
Extract Features (X)
308
Adaptation Pattern Matching
310
Select f()
312
Generate f(X)
314
Modify f()?
Y
316
Update f()
N
318
Re-generate f(X)
320
Transmit f(X) through Channel
322
VR Pattern Matching

FIG. 3

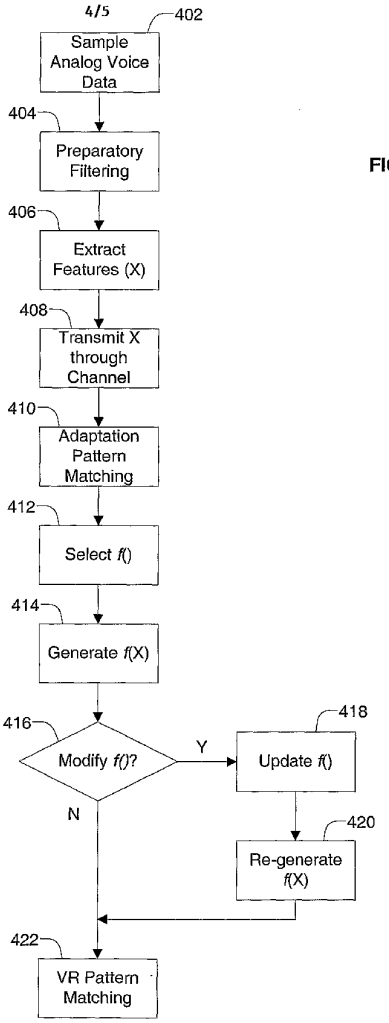

402
Sample Analog Voice Data
404
Preparatory Filtering
406
Extract Features (X)
408
Transmit X through Channel
410
Adaptation Pattern Matching
412
Select f()
414
Generate f(X)
416
Modify f()?
Y
N
418
Update f()
420
Re-generate f(X)
422
VR Pattern Matching

FIG. 4

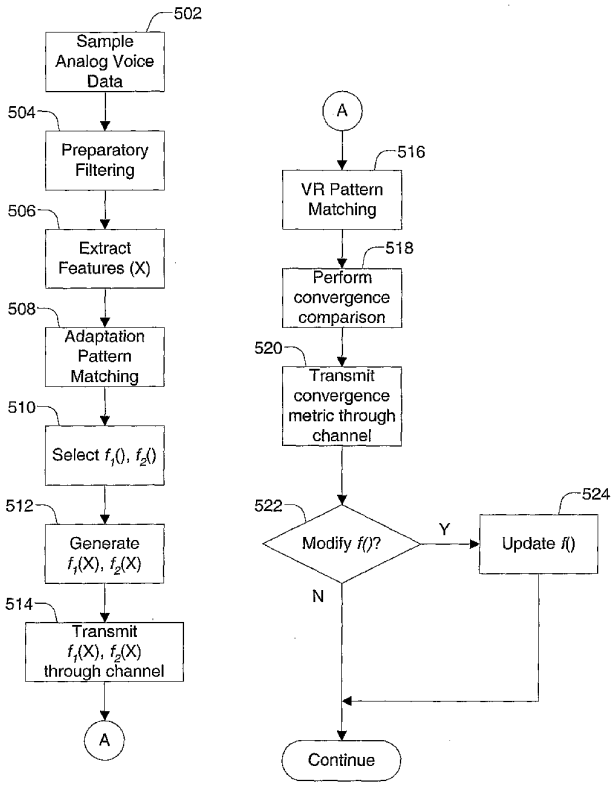

502
Sample Analog Voice Data
504
Preparatory Filtering
506
Extract Features (X)
508
Adaptation Pattern Matching
510
Select $f_1()$, $f_2()$
512
Generate $f_1(X)$, $f_2(X)$
514
Transmit $f_1(X)$, $f_2(X)$ through channel
A
A
516
VR Pattern Matching
518
Perform convergence comparison
520
Transmit convergence metric through channel
522
Modify f()?
Y
N
524
Update f()
Continue

FIG. 5

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/065453 A3

(51) International Patent Classification[7]: G10L 15/06, 15/28, 15/20

(21) International Application Number: PCT/US02/03014

(22) International Filing Date: 30 January 2002 (30.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/773,831 31 January 2001 (31.01.2001) US

(71) Applicant: QUALCOMM INCORPORATED [US/US]; 5775 Morehouse Drive, San Diego, CA 92121-1714 (US).

(72) Inventors: CHANG, Chienchung; 6076 Via Posada Del Norte, Rancho Santa Fe, CA 92067 (US). MALAYATH, Naren; 10710 Sabre Hill Drive, #229, San Diego, CA 92128 (US). YAFUSO, Byron, Yoshio; 10093 Branford Road, San Diego, CA 92129 (US).

(74) Agents: WADSWORTH, Philip, R. et al.; Qualcomm Incorporated, 5775 Morehouse Drive, San Diego, CA 92121-1714 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

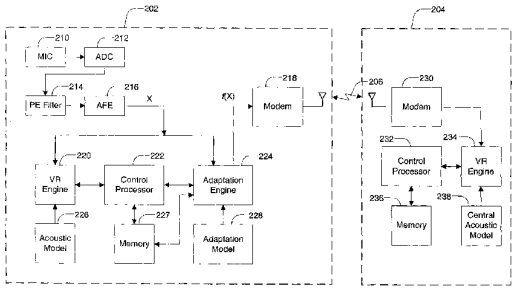
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report
before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
24 October 2002

[Continued on next page]

(54) Title: DISTRIBUTED VOICE RECOGNITION SYSTEM USING ACOUSTIC FEATURE VECTOR MODIFICATION



(57) Abstract: A voice recognition system applies speaker-dependent modification functions to acoustic feature vectors prior to voice recognition pattern matching against a speaker-independent acoustic model (238). An adaptation engine (224) matches a set of acoustic feature vectors X with an adaptation model (228) to select a speaker-dependent feature vector modification function f(), which is then applied to X to form a modified set of acoustic feature vectors f(X). Voice recognition is then performed by correlating the modified acoustic feature vectors f(X) with a speaker-independent acoustic model (238).

WO 02/065453 A3

WO 02/065453 A3

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/03014

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L15/06 G10L15/28 G10L15/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 070 139 A (MIYAZAWA ET AL) 30 May 2000 (2000-05-30)	1-3,5,6, 8,15,17, 18,20, 27,28, 32,33, 35,36, 39,40, 44,49
Y	abstract; figures 2,3,7-9	4,7, 9-14,16, 19, 21-26, 29-31, 34,37, 38, 41-43, 45-48
	--- -/--	

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 August 2002	04/09/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Quélavoine, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/03014

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 779 609 A (NIPPON ELECTRIC CO) 18 June 1997 (1997-06-18) abstract; figure 3	4,7, 9-14,16, 19,26, 29-31, 34,37, 38,41, 42,45-48
Y	EP 0 661 690 A (IBM) 5 July 1995 (1995-07-05) abstract; figures 2,3	21-26,43
A	LOGAN B: "MAXIMUM LIKELIHOOD SEQUENTIAL ADAPTATION" 6TH EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY. EUROSPEECH '99, vol. 1 OF 6, 5 - 9 September 1999, pages 17-20, XP001076054 BUDAPEST, HONGARY, BONN: ESCA, DE abstract; figure 2	4,7,11, 14,16, 19,26, 31,38
A	GALES M J F: "TRANSFORMATION SMOOTHING FOR SPEAKER AND ENVIRONMENTAL ADAPTATION" 5TH EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY. EUROSPEECH '97., vol. 4 OF 5, 22 - 25 September 1997, pages 2067-2070, XP001049266 RHODES, GREECE, GRENOBLE: ESCA, FR abstract	3-7, 11-14, 16-19, 28,31, 33,38,40
A	US 6 363 348 B1 (BESLING ET AL) 26 March 2002 (2002-03-26) abstract & EP 0 980 574 A 23 February 2000 (2000-02-23)	1,8,9, 15,21, 22,27, 32,39, 44,48,49

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 02/03014

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 6070139	A	30-05-2000	JP	9062289 A	07-03-1997
EP 0779609	A	18-06-1997	JP	3001037 B2	17-01-2000
			JP	9160584 A	20-06-1997
			DE	69614233 D1	06-09-2001
			DE	69614233 T2	08-05-2002
			EP	0779609 A2	18-06-1997
			US	5890113 A	30-03-1999
EP 0661690	A	05-07-1995	EP	0661690 A1	05-07-1995
			JP	7210190 A	11-08-1995
			US	5802251 A	01-09-1998
US 6363348	B1	26-03-2002	EP	0980574 A2	23-02-2000
			WO	9921172 A2	29-04-1999
			JP	2001506382 T	15-05-2001

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZM,ZW

(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 チャン、チエンチャン
アメリカ合衆国、カリフォルニア州 ９２０６７、ランチョ・サンタ・フェ、ビア・ポサダ・デル・ノルテ ６０７６
(72)発明者 マラヤス、ナレン
アメリカ合衆国、カリフォルニア州 ９２１２８、サン・ディエゴ、ナンバー ２２９、サブレ・ヒル・ドライブ １０７１０
(72)発明者 ヤフソ、バイロン・ヨシオ
アメリカ合衆国、カリフォルニア州 ９２１２９、 サン・ディエゴ、ブランフォード・ロード １００９３

Ｆターム(参考) 5D015 FF04 LL12