(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4824286号
(P4824286)
(45)発行日　平成23年11月30日(2011.11.30)　(24)登録日　平成23年9月16日(2011.9.16)

(51)Int.Cl.　F I
G10L 15/20 (2006.01)　G10L 15/20 353
G10L 15/20 37OD

請求項の数 14 (全 17 頁)

(21)出願番号　特願2004-101400 (P2004-101400)
(22)出願日　平成16年3月30日 (2004.3.30)
(65)公開番号　特開2004-302470 (P2004-302470A)
(43)公開日　平成16年10月28日 (2004.10.28)
審査請求日　平成19年3月8日 (2007.3.8)
(31)優先権主張番号　10/403,638
(32)優先日　平成15年3月31日 (2003.3.31)
(33)優先権主張国　米国 (US)

前置審査

(73)特許権者　500046438
マイクロソフト　コーポレーション
アメリカ合衆国　ワシントン州　９８０５２－６３９９　レッドモンド　ワン　マイクロソフト　ウェイ
(74)代理人　110001243
特許業務法人　谷・阿部特許事務所
(74)復代理人　100115624
弁理士　濱中　淳宏
(72)発明者　アレサンドロ　アセロ
アメリカ合衆国　９８００６　ワシントン州　ベルビュー　１６３　プレイス　サウスイースト　６５２５

最終頁に続く

(54)【発明の名称】漸進的ベイズ学習を使用する雑音推定の方法

(57)【特許請求の範囲】
【請求項１】
雑音を含む信号内の雑音を推定する方法であって、
前記雑音を含む信号をフレームに分割すること、および、
漸進的ベイズ学習を使用して、フレームに関する平均および分散の両方を含む雑音推定を決定し、経時変化する事前雑音分布が想定され、雑音推定が、先行するフレームで計算された事後雑音に関する近似を使用して再帰的に更新されることを備え、
雑音推定を決定することは、
先行するフレームで計算された事後雑音に関する近似を使用して、前記雑音を含む信号の第１のフレームに関する雑音推定を決定すること、
前記雑音を含む信号の第２のフレームに関するデータ尤度推定を決定すること、および、
前記第２のフレームに関するデータ尤度推定および前記第１のフレームに関する前記雑音推定を使用して、前記第２のフレームに関する雑音推定を決定することを含み、
前記第２のフレームに関する前記データ尤度推定を決定することは、クリーン信号および雑音信号の非線形関数として、非線形音響ひずみモデルを使用することにより導出される非線形関数を用いることを含み、
前記第２のフレームに関する前記データ尤度推定を決定するために、前記非線形関数は、さらに近似により線形化されることを特徴とする方法。
【請求項２】

前記近似は、前記第１のフレームに関する前記雑音推定によって部分的に定義された時点での非線形関数に等しいことを特徴とする請求項１に記載の方法。
【請求項３】
前記近似は、テイラー級数展開であることを特徴とする請求項２に記載の方法。
【請求項４】
前記近似は、さらにラプラス近似を得ることを含むことを特徴とする請求項３に記載の方法。
【請求項５】
前記第２のフレームに関する前記データ尤度推定を使用することは、非線形関数のテイラー級数展開のための展開点として、前記第１のフレームに関する前記雑音推定を使用することを含むことを特徴とする請求項１に記載の方法。
【請求項６】
前記事後雑音に関する近似を使用することは、ガウス近似を使用することを含むことを特徴とする請求項１に記載の方法。
【請求項７】
前記雑音推定は、ガウス近似に基づくものであることを特徴とする請求項１に記載の方法。
【請求項８】
前記雑音推定は、各フレームに関する雑音推定を連続して決定することを含むことを特徴とする請求項７に記載の方法。
【請求項９】
雑音を含む信号内の雑音を推定するための方法であって、
前記雑音を含む信号をフレームに分割すること、および、
各フレームに関して連続的に、各フレーム内の雑音を推定し、現在のフレームに関する雑音推定は、前記現在のフレームに関するデータ尤度のガウス近似、および事前フレームのシーケンスにおける雑音のガウス近似に基づくことを備え、
前記各フレーム内の雑音を推定することは、クリーン信号および雑音信号の非線形関数として、非線形音響ひずみモデルを使用することにより導出される非線形関数を用いることを含み、
前記各フレーム内の雑音を推定するために、前記非線形関数は、さらに近似により線形化されることを特徴とする方法。
【請求項１０】
前記近似は、前のフレームに関する前記雑音推定によって部分的に定義された時点での非線形関数に等しいことを特徴とする請求項９に記載の方法。
【請求項１１】
前記近似は、テイラー級数展開であることを特徴とする請求項１０に記載の方法。
【請求項１２】
前記近似は、ラプラス近似をさらに含むことを特徴とする請求項１１に記載の方法。
【請求項１３】
前記雑音推定は、雑音平均推定および雑音分散推定を含むことを特徴とする請求項９乃至１２のいずれかに記載の方法。
【請求項１４】
請求項１乃至１３に記載の方法のいずれかを実行するように適合されることを特徴とするシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、雑音推定に関する。より詳細には、本発明は、パターン認識に使用される信号の雑音推定に関する。
【背景技術】

【０００２】
音声認識システムなどのパターン認識システムは、入力信号を受け取り、信号を復号して、信号によって表されたパターンを見つけようと試みる。例えば、音声認識システムでは、認識システムによって音声信号（テスト信号と呼ばれることが多い）が受け取られ、音声信号によって表された文字列を識別するために復号される。
【０００３】
入力信号は、典型的には、何らかの形の雑音によって破壊される。パターン認識システムの性能を向上させるためには、雑音を含む信号における雑音を推定することが望ましい場合が多い。
【０００４】
過去には、いくつかのフレームワークを使用して、信号内の雑音を推定してきた。第１のフレームワークでは、バッチアルゴリズムが使用され、これは、入力信号内の他のフレーム内に見られる雑音とは無関係に、入力信号の各フレーム内の雑音を推定する。個々の雑音推定値が平均されて、すべてのフレームに関するコンセンサス雑音（consensus noise）値が形成される。第２のフレームワークでは、再帰的アルゴリズムが使用され、これは、１または複数の、以前のまたは連続するフレームの雑音推定値に基づいて、現在のフレーム内の雑音を推定する。こうした再帰的技法は、経時的にゆっくりと変化する雑音を見込む。
【発明の開示】
【発明が解決しようとする課題】
【０００５】
１つの再帰的技法では、雑音を含む信号（noisy signal）は、クリーン信号（clean signal）と雑音信号（noise signal）の非線形関数となると想定される。計算を促進するために、この非線形関数は、いくつかの展開点に関して計算された、切断テイラー級数展開によって近似されることが多い。一般にテイラー級数展開は、展開点において、関数の最良推定値を表す。従って、テイラー級数近似法の良さは、展開点の選択と同程度にすぎない。ただし、従来技術の下では、テイラー級数の展開点は、フレームごとに最適化されることはなかった。従って、再帰的アルゴリズムで生成される雑音推定値は、理想的とは言えない。
【０００６】
最尤（ＭＬ：maximum-likelihood)および最大事後(ＭＡＰ：maximum a posteriori）技法が、音響の場で反復的に線形化された非線形モデルを使用して、非定常雑音の連続点推定に使用されてきた。一般に、雑音の分布に関する単純なガウスモデルを使用して、ＭＡＰ推定は、より品質の良い雑音推定を提供した。ただし、ＭＡＰ技法では、事前ガウス雑音（Gaussian noise prior）に関連付けられた平均および分散パラメータは、各々音声フリーのテスト発声（speech-free test utterance）のセグメントから決定された。非定常雑音の場合、この近似は、実際の事前雑音統計（noise prior statistics）を適正に反映するものでない可能性がある。
【０００７】
以上に鑑みて、パターン信号の雑音推定における、より効果的な雑音推定技法が求められている。
【課題を解決するための手段】
【０００８】
非定常雑音を推定するための新しい方法は、漸進的ベイズ学習（incremental bayes learning）を使用する。一態様では、この技法は、ハイパーパラメータ（平均および分散）によって定義可能な雑音推定を、経時変化する事前雑音分布の仮定として定義することができ、事前雑音分布は、先行する時間またはフレームステップで計算された事後近似（approximation posterior）を使用して、再帰的に更新される。他の態様では、この技法は、各フレームについて連続的に、各フレーム内の雑音を推定するものとして定義して、現在のフレームの雑音推定が、現在のフレームのデータ尤度のガウス近似、および以前の一

連のフレームにおける雑音のガウス近似に基づくようにすることができる。

【発明を実施するための最良の形態】

【０００９】

図１に、本発明を実施することができる好適なコンピューティングシステム環境１００の一例を示す。コンピューティングシステム環境１００は、好適なコンピューティング環境の一例を示すものに過ぎず、本発明の使用または機能の範囲に関して、どのような制限も提案することを意図するものではない。さらに、コンピューティング環境１００は、例示的オペレーティング環境１００に示された構成要素のいずれか１つまたはいずれかの組合せに関して、どのような依存性または要件をも有するものと解釈するべきではない。

【００１０】

本発明は、多数の他の汎用または特定用途向けコンピューティングシステム環境または構成でも動作可能である。本発明で使用するのに好適な、よく知られたコンピューティングシステム、環境および／または構成の例には、これに限定されないが、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースシステム、セットトップボックス、プログラム可能消費家電、ネットワークＰＣ、ミニコンピュータ、メインフレームコンピュータ、電話システム、上記いずれかのシステムを含む分散コンピューティング環境などを含む。

【００１１】

本発明は、コンピュータによって実行される、プログラムモジュールなどの、コンピュータ実行可能命令の一般的な文脈で説明することができる。一般に、プログラムモジュールには、特定のタスクを実行するか、または特定の抽象データ型を実施する、ルーチン、プログラム、オブジェクト、構成要素、データストラクチャなどが含まれる。プログラムおよびモジュールによって実行されるタスクについては、以下に図を用いて説明する。当分野の技術者であれば、本明細書の記述および／または図面を、以下で論じるいずれかの形のコンピュータ読取り可能媒体上で実施可能な、コンピュータ実行可能命令として実施することができる。

【００１２】

本発明は、通信ネットワークを介してリンクされたリモート処理デバイスによってタスクが実行される、分散コンピューティング環境でも実施可能である。分散コンピューティング環境では、プログラムモジュールは、メモリ記憶デバイスを含む、ローカルおよびリモートのどちらのコンピュータ記憶媒体にも配置することができる。

【００１３】

図１を参照すると、本発明を実施するための例示的システムには、コンピュータ１１０の形式の汎用コンピューティングデバイスが含まれる。コンピュータ１１０の構成要素は、これに限定されないが、処理ユニット１２０、システムメモリ１３０、および、システムメモリを含む様々なシステム構成要素を処理ユニット１２０に結合するシステムバス１２１を含むことができる。システムバス１２１は、メモリバスまたはメモリコントローラ、周辺バス、および、様々なバスアーキテクチャのうちのいずれかを使用するローカルバスを含む、いくつかの種類のバス構造のうちのいずれであってもよい。例を挙げると、これに限定されないが、こうしたアーキテクチャは、ＩＳＡ（Industry Standard Architecture）バス、ＭＣＡ（Micro Channel Architecture）バス、ＥＩＳＡ（Enhanced ＩＳＡ）バス、ＶＥＳＡ（Video Electronics Standards Association）ローカルバス、および、メザニンバスとも呼ばれるＰＣＩ（Peripheral Component Interconnect）バスを含むことができる。

【００１４】

コンピュータ１１０は、典型的には、様々なコンピュータ読取り可能媒体を含む。コンピュータ読取り可能媒体は、コンピュータ１１０がアクセス可能な任意の使用可能媒体であってよく、揮発性および不揮発性媒体、取外し可能および取外し不能媒体の、どちらも含む。例を挙げると、コンピュータ読取り可能媒体は、コンピュータ記憶媒体および通信媒体を含むことができるが、これらに限定されない。コンピュータ記憶媒体は、コンピュ

ータ読取り可能命令、データストラクチャ、プログラムモジュール、または他のデータなどの情報を格納するための、任意の方法または技術で実施された、揮発性および不揮発性、取外し可能および取外し不能のどちらの媒体も含む。コンピュータ記憶媒体は、これに限定されないが、ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュメモリまたは他のメモリ技術、ＣＤ－ＲＯＭ、ＤＶＤ（Digital Versatile Disk）または他の光ディスク記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶または他の磁気記憶デバイス、あるいは、所望の情報の格納に使用可能であり、コンピュータ１１０がアクセス可能な、任意の他の媒体を含む。通信媒体は、典型的には、コンピュータ読取り可能命令、データストラクチャ、プログラムモジュール、または他のデータを、搬送波または他の移送メカニズムなどの変調データ信号で具体化し、任意の情報送達媒体を含む。「変調データ信号」という用語は、その特徴のうちの１または複数を、信号内の情報を符号化するのと同じ方法で設定または変更した信号を意味する。例を挙げると、これらに限定されないが、通信媒体は、有線ネットワークまたはダイレクトワイヤード接続などの有線媒体、および音波、ＲＦ、赤外線、および他の無線媒体などの無線媒体を含む。コンピュータ読取り可能媒体の範囲内は、上記のいずれかの組合せも含む。

【００１５】

システムメモリ１３０は、読取り専用メモリ（ＲＯＭ）１３１およびランダムアクセスメモリ（ＲＡＭ）１３２などの、揮発性および／または不揮発性メモリの形のコンピュータ記憶媒体を含む。起動時などに、コンピュータ１１０内の要素間で情報を転送するのに役立つ、基本ルーチンを含む基本入出力システム１３３（ＢＩＯＳ）は、通常、ＲＯＭ１３１内に格納される。ＲＡＭ１３２は、典型的には、処理ユニット１２０によって即時にアクセス可能であり、および／または、現在作動中である、データおよび／またはプログラムモジュールを含む。例を挙げると、これらに限定されないが、図１は、オペレーティングシステム１３４、アプリケーションプログラム１３５、他のプログラムモジュール１３６、およびプログラムデータ１３７を示す。

【００１６】

コンピュータ１１０は、他の取外し可能／取外し不能、揮発性／不揮発性の、コンピュータ記憶媒体を含むこともできる。単なる例に過ぎないが、図１には、取外し不能の不揮発性磁気媒体から読み取るかまたはこれに書き込むハードディスクドライブ１４１、取外し可能な不揮発性磁気ディスク１５２から読み取るかまたはこれに書き込む磁気ディスクドライブ１５１、および、ＣＤ　ＲＯＭまたは他の光媒体などの取外し可能な不揮発性光ディスク１５６から読み取るかまたはこれに書き込む光ディスクドライブ１５５が示されている。例示的なオペレーティング環境で使用可能な、他の取外し可能／取外し不能、揮発性／不揮発性のコンピュータ記憶媒体は、これらに限定されないが、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、ソリッドステートＲＡＭ、ソリッドステートＲＯＭなどを含む。ハードディスクドライブ１４１は、典型的には、インタフェース１４０などの取外し不能メモリインタフェースを介して、システムバス１２１に接続され、磁気ディスクドライブ１５１および光ディスクドライブ１５５は、典型的には、インタフェース１５０などの取外し可能メモリインタフェースによって、システムバス１２１に接続される。

【００１７】

上述し、図１に示した、ドライブおよびそれらの関連付けられたコンピュータ記憶媒体は、コンピュータ読取り可能命令、データストラクチャ、プログラムモジュール、およびコンピュータ１１０に関する他のデータの記憶域を提供するものである。例えば、図１では、ハードディスクドライブ１４１は、オペレーティングシステム１４４、アプリケーションプログラム１４５、他のプログラムモジュール１４６、およびプログラムデータ１４７を格納するものとして示されている。これらの構成要素は、オペレーティングシステム１３４、アプリケーションプログラム１３５、他のプログラムモジュール１３６、およびプログラムデータ１３７と同じであっても異なっていてもよいことに留意されたい。オペレーティングシステム１４４、アプリケーションプログラム１４５、他のプログラムモジ

ュール１４６、およびプログラムデータ１４７には、少なくとも異なるコピーであることを示すために、ここでは異なる番号が与えられている。

【００１８】

ユーザは、キーボード１６２、マイクロフォン１６３、および、マウス、トラックボールまたはタッチパッドなどのポインティングデバイス１６１を介して、コマンドおよび情報をコンピュータ１１０に入力することができる。他の入力デバイス（図示せず）には、ジョイスティック、ゲームパッド、衛星放送用パラボラアンテナ、スキャナなどが含まれる。これらおよび他の入力デバイスは、システムバスに結合されたユーザ入力インタフェース１６０を介して、処理ユニット１２０に接続されることが多いが、パラレルポート、ゲームポート、またはユニバーサルシリアルバス（ＵＳＢ）などの、他のインタフェースおよびバス構造によって接続することもできる。モニタ１９１または他のタイプのディスプレイデバイスも、ビデオインタフェース１９０などのインタフェースを介して、システムバス１２１に接続される。コンピュータは、モニタに加えて、出力周辺インタフェース１９５を介して接続可能な、スピーカ１９７およびプリンタ１９６などの他の周辺出力デバイスを含むこともできる。

【００１９】

コンピュータ１１０は、リモートコンピュータ１８０などの１または複数のリモートコンピュータへの論理接続を使用して、ネットワーク化環境で動作可能である。リモートコンピュータ１８０は、パーソナルコンピュータ、ハンドヘルドデバイス、サーバ、ルータ、ネットワークＰＣ、ピアデバイス、または他の共通ネットワークノードであってよく、典型的には、コンピュータ１１０に関して上記で述べた要素の多くまたはすべてを含む。図１に示された論理接続には、ローカルエリアネットワーク（ＬＡＮ）１７１およびワイドエリアネットワーク（ＷＡＮ）１７３が含まれるが、他のネットワークを含むこともできる。こうしたネットワーキング環境は、オフィス、企業規模のコンピュータネットワーク、イントラネット、およびインターネットでよく見られる。

【００２０】

ＬＡＮネットワーキング環境で使用される場合、コンピュータ１１０は、ネットワークインタフェースまたはアダプタ１７０を介してＬＡＮ１７１に接続される。ＷＡＮネットワーキング環境で使用される場合、コンピュータ１１０は、通常、インターネットなどのＷＡＮ１７３を介した通信を確立するために、モデム１７２または他の手段を含む。内蔵または外付けが可能なモデム１７２は、ユーザ入力インタフェース１６０または他の適切なメカニズムを介して、システムバス１２１に接続することができる。ネットワーク化環境では、コンピュータ１１０関して示されたプログラムモジュールまたはその一部を、リモートメモリ記憶デバイスに格納することができる。例を挙げると、図１では、リモートアプリケーションプログラム１８５がリモートコンピュータ１８０上に常駐しているように示されているが、これに限定されるものではない。示されたネットワーク接続は例示的なものであり、コンピュータ間に通信リンクを確立する他の手段が使用可能であることを理解されよう。

【００２１】

図２は、例示的なコンピューティング環境である、移動デバイス２００を示す構成図である。移動デバイス２００は、マイクロプロセッサ２０２、メモリ２０４、入出力（Ｉ／Ｏ）構成要素２０６、および、リモートコンピュータまたは他の移動デバイスと通信するための通信インタフェース２０８を含む。一実施形態では、前述の構成要素は、相互に通信するために、好適なバス２１０を介して結合される。

【００２２】

メモリ２０４は、移動デバイス２００への一般的な電力が遮断されたときに、メモリ２０４に格納された情報が失われないように、バッテリバックアップモジュール（図示せず）を備えた、ランダムアクセスメモリ（ＲＡＭ）などの不揮発性電子メモリとして実装される。好ましくは、メモリ２０４の一部がアドレス可能メモリとして割り振られ、好ましくは、メモリ２０４の他の部分が、ディスクドライブ上の記憶域をシミュレートするよう

な記憶域に使用される。

【００２３】

メモリ２０４は、オペレーティングシステム２１２、アプリケーションプログラム２１４、ならびにオブジェクトストア２１６を含む。オペレーション時には、オペレーティングシステム２１２は、好ましくは、メモリ２０４からプロセッサ２０２によって実行される。好ましい一実施形態では、オペレーティングシステム２１２は、Microsoft Corporationから市販されている、ＷＩＮＤＯＷＳ（登録商標）ＣＥブランドのオペレーティングシステムである。オペレーティングシステム２１２は、好ましくは移動デバイス用に設計され、アプリケーション２１４が、露出された（exposed）アプリケーションプログラミングインタフェースおよびメソッドのセットを介して使用可能なデータベース機能を実施する。オブジェクトストア２１６内のオブジェクトは、露出されたアプリケーションプログラミングインタフェースおよびメソッドへの呼び出しに応答して、少なくとも部分的に、アプリケーション２１４およびオペレーティングシステム２１２によって維持される。

【００２４】

通信インタフェース２０８は、移動デバイス２００が情報を送受信できるようにする、多数のデバイスおよび技術を表す。デバイスは、一例を挙げると、有線および無線モデム、衛星受信機、ならびに放送チューナを含む。移動デバイス２００は、データを交換するために、コンピュータに直接接続することもできる。このような場合には、通信インタフェース２０８は、赤外線受信機、ならびにシリアルまたはパラレル通信接続であってよく、それらはすべて、ストリーミング情報を伝送することができる。

【００２５】

入出力構成要素２０６は、タッチセンシティブスクリーン、ボタン、ローラ、およびマイクロフォンなどの様々な入力デバイス、ならびに、音声生成器、バイブレーティングデバイス、およびディスプレイを含む、様々な出力デバイスを含む。上述したデバイスは、例として示したものであって、移動デバイス２００上にすべてが存在する必要はない。さらに、他の入出力デバイスは、本発明の範囲内にある移動デバイス２００に取り付けられるか、またはこのデバイスが見つけることができる。

【００２６】

本発明の一態様の下で、パターン認識信号における雑音を推定するシステムおよび方法が提供される。これを実行するために、本発明は、再帰的アルゴリズムを使用して、少なくとも１つの隣接フレームに関して見つけられた雑音推定に部分的に基づき、雑音を含む信号の各フレームでの雑音を推定する。本発明の下では、漸進的ベイズ学習を使用することによって、単一のフレームに関する雑音を推定するが、ここでは、経時変化する事前雑音分布が想定され、雑音推定値は、先行フレームで計算された事後雑音に関する近似を使用して、再帰的に更新される。この再帰的プロセスを通して、雑音推定は非定常雑音を追跡することができる。

【００２７】

【数１】

$$\mathbf{y}_1^t = y_1, y_2, \cdots, y_\tau, \cdots, y_t$$

【００２８】

を、対数領域（ログスペクトラムまたはケプストラム）で表された、一連の雑音を含む音声観察データとし、普遍性を失うことなしにスカラ値付けされるものと想定する。データ

【００２９】

【数２】

$$\mathbf{y}_1^t$$

【００３０】

を使用して、同じデータ長さｔを持つ、破壊した雑音シーケンス

【００３１】

【数３】

$$\mathbf{n}_1^t = \mathbf{n}_1, \mathbf{n}_2, \ldots, \ldots, \mathbf{n}_t$$

【００３２】

を逐次推定する。ベイズの学習フレームワークでは、雑音ｎ（未知のパラメータとして扱われる）に関する知識は、所与のｐ（ｎ）の事前分布に含まれると想定される。雑音シーケンスが定常である場合、すなわち、雑音の統計プロパティが経時的に変化しない場合、雑音パラメータｎに関する従来のベイズ推論（すなわち事後を算出する）は、いかなる時点でも、以下の「バッチモード」ベイズ規則を介して実施することが可能であり、

【００３３】

【数４】

$$p\left(n\middle|y_1^t\right) = \frac{p\left(y_1^t\middle|n\right)p(n)}{\int_{(\Theta)} p\left(y_1^t\middle|n\right)p(n)dn'}$$

【００３４】

ここで、Θは、雑音パラメータ空間の許容領域である。

【００３５】

【数５】

$$\mathrm{p}\left(\mathrm{n}\middle|\mathrm{y}_1^t\right)$$

【００３６】

が与えられると、原則として、雑音ｎに関する任意の推定が可能である。例えば、雑音ｎに関する従来のＭＡＰ点推定は、事後

【００３７】

【数６】

$$\mathrm{p}\left(\mathrm{n}\middle|\mathrm{y}_1^t\right)$$

【００３８】

のグローバルまたはローカルの最大として計算される。最小平均二乗誤差（ＭＭＳＥ）推定は、事後

【００３９】

【数７】

$$\mathrm{p}\left(\mathrm{n}\middle|\mathrm{y}_1^t\right)$$

【００４０】

を介した期待値である。

【００４１】

ただし、雑音シーケンスが非定常であり、雑音を含む音声

【００４２】

【数８】

$$\mathbf{y}_1^t$$

【００４３】

のトレーニングデータが、ほとんどの実際の音声の特徴を拡張した応用の場合と同様に、逐次提示される場合、経時的に変化している雑音統計を追跡するために、新しい雑音推定技法が求められる。反復的な応用では、ベイズ規則は以下のように表すことができる。

【００４４】

【数９】

$$p(n_t|y_1^t)=\frac{1}{C_t}p(y_t|y_1^{t-1},n_t)p(n_t|y_1^{t-1})$$

上式で、$C_t=p(y_1^t|y_1^{t-1})=\int_\Theta p(y_t|t_1^{t-1},n_t)p(n_t|y_1^{t-1})dn_t$

【００４５】

雑音の多い音声y_tとその過去の

【００４６】

【数１０】

$$y_1^{t-1}$$

【００４７】

所与のn_t、または

【００４８】

【数１１】

$$p(y_t|y_1^{t-1},n_t)=p(y_t|n_t)$$

【００４９】

との間に、条件付き独立性があると想定し、事後

【００５０】

【数１２】

$$p(n_t|y_1^{t-1})\approx p(n_{t-1}|y_1^{t-1})$$

【００５１】

内が平滑であると想定すると、以前の数式は、以下のように表すことができる。

【００５２】

【数１３】

$$p(n_t|y_1^t)\approx\frac{1}{C_t}p(y_t|n_t)p(n_{t-1}|y_1^{t-1}) \quad (1)$$

【００５３】

非定常雑音の漸進的な学習を、数式（１）を以下のように繰返し使用することにより、確立することができる。初期には、雑音を含む音声データｙがなく、既知の事前ｐ（n_0｜y_0）＝ｐ（n_0）から事後ＰＤＦが得られるが、ここで、ｐ（n_0）は、既知の雑音のみのフレームの分析および推定されたガウスから取得される。次にｔ＝１の場合、式（１）を使用すると、

【００５４】

【数１４】

$$p(n_1|y_1)\approx\frac{1}{C_1}p(y_1|n_1)p(n_0) \quad (2)$$

【００５５】

が生成され、ｔ＝２の場合は、すでに数式２から計算されたｐ（n_1｜y_1）を使用して、

【００５６】

【数１５】

$$p(n_2|y_1,y_2)\approx\frac{1}{C_2}p(y_2|n_2)p(n_1|y_1)$$

【００５７】
が生成される。ｔ＝３の場合、式（１）は、
【００５８】
【数１６】

$$p\left(n_3\middle|y_1^3\right)\approx\frac{1}{C_3}p\left(y_3\middle|n_3\right)p\left(n_2\middle|y_1,y_2\right)$$

【００５９】
となり、以下同様である。従って、このプロセスは、以下のように、再帰的に一連の事後（ｐ（ｙ$_t$｜ｎ$_t$）が使用可能であるとすると）を生成し、
【００６０】
【数１７】

$$p\left(n_1\middle|y_1\right).p\left(n_2\middle|y_1^2\right)\ldots.p\left(n_\tau\middle|y_1^\tau\right)\ldots.p\left(n_t\middle|y_1^t\right)\ldots \qquad (3)$$

【００６１】
これは、非定常雑音シーケンス
【００６２】
【数１８】

$$\mathbf{n}_1^t$$

【００６３】
に関する、漸進的ベイズの推定を行うための基本を提供する。次に、これまで論じてきた漸進的ベイズの推定の一般原理を、事前雑音がガウスであるという単純化された想定の下で、フレーム形式のデータＰＤＦ（ｐ（ｙ$_t$｜ｎ$_t$））を供給する特定の音響ひずみモデルに適用する。
【００６４】
雑音に適用すると、漸進的ベイズの学習は、最も近い過去まで観察されたデータが与えられた事後を使用して、雑音に関する現在の「事前」分布を更新するが、これは、この事後が、現時点以前のパラメータに関する最も完全な情報であるためである。この方法が図３に示されており、第１のステップでは、雑音を含む信号３００がフレームに分割される。ステップ３０２では、各フレームについて、漸進的ベイズ学習が適用され、各フレームの雑音推定は、経時変化する事前雑音分布を想定し、雑音推定は、先行する時間フレームで計算された事後雑音に関する近似を使用して、再帰的に更新される。従って、式（３）の事後シーケンスは、（１フレームの大きさのタイムシフトを伴う）雑音分布パラメータに関して、経時変化する事前シーケンス（すなわち、事前進化）となる。一実施形態では、ステップ３０２は、式（１）で示された事後における平滑を想定する、先行フレーム、好ましくは直前のフレームにおける雑音推定を使用しながら、現在のフレームに関するデータ尤度ｐ（ｙ$_t$｜ｎ$_t$）の計算を含むことが可能である。
【００６５】
データ尤度ｐ（ｙ$_t$｜ｎ$_t$）が非ガウスである（簡単に説明する）場合、事後は、必然的に非ガウスである。式（１）を連続して適用すると、結果として、高速に展開する以前の事後の組合せとなり、処理が困難形となる。この困難を克服するために、近似法が求められる。使用される近似法は、ｙ$_t$とｎ$_t$の間の非線形関係を線形化するために、１次テイラー級数展開を適用することである。これにより、ｐ（ｙ$_t$｜ｎ$_t$）のガウス形となる。従って、過去のデータ履歴
【００６６】
【数１９】

$$p\left(n_\tau\middle|y_1^\tau\right)$$

【００６７】

について事後から継承された、経時変化する事前雑音ＰＤＦｐ（$n_{\tau+1}$）は、ガウスによって以下のように近似することができ、

【００６８】

【数２０】

$$p\left(n_\tau \middle| y_1^\tau\right) = \frac{1}{(2\pi)^{1/2}\sigma_{n_\tau}} \exp\left[-\frac{1}{2}\left(\frac{n_\tau - \mu_{n_\tau}}{\sigma_{n_\tau}}\right)^2\right] = N\left[n_\tau; \mu_{n_\tau}, \sigma_{n_\tau}^2\right] \quad (4)$$

【００６９】

上式で、$\mu_{n\tau}$および

【００７０】

【数２１】

$$\sigma_{n\tau}^2$$

【００７１】

は、事前ＰＤＦを特徴付けるハイパーパラメータ（平均および分散）と呼ばれる。次に、再帰的ベイズ規則である式（１）から計算された式（３）の事後シーケンスは、ハイパーパラメータの一時進化を決定する、原則的な方法を提供し、これについて以下で論じる。

【００７２】

データ尤度ｐ（$y_t | n_t$）を計算するための音響ひずみおよびクリーン音声モデルを示す。第１に、クリーン音声ｘの対数スペクトラムについて、以下のような経時変化しないガウスモデル混合を想定する。

【００７３】

【数２２】

$$p(x) = \sum_m p(m) N\left[x; \mu_x(m), \sigma_x^2(m)\right] \quad (5)$$

【００７４】

次に、以下のように、対数スペクトラム領域内の単純な非線形音響ひずみモデルを、使用することができ、

exp(y)=exp(x)+exp(n)　または　y=x+g(n-x)　(6)

上式で、非線形関数は、以下のようになる。

g(z)=log[1+exp(z)]

データ尤度ｐ（$y_t | n_t$）に関する有用な形を得るために、テイラー級数展開を使用して、式（６）の非線形性ｇを線形化する。これにより、線形化モデル

【００７５】

【数２３】

$$y \approx x + g(n_0 - \mu_x(m_0)) + g'(n_0 - \mu_x(m_0))(n - n_0) \quad (7)$$

【００７６】

が与えられ、ここで、n_0はテイラー級数展開点であり、１次級数展開係数は、以下のように容易に計算することができる。

【００７７】

【数２４】

$$g'(n_0 - \mu_x(m_0)) = \frac{\exp(n_0)}{\exp[\mu_x(m_0)] + \exp(n_0)}$$

【００７８】

式（７）の関数ｇおよびｇ’を評価する際に、クリーン音声値ｘは、「最適な」混合ガウス構成要素m_0の平均（$\mu_x(m_0)$）とみなされる。

【００７９】

式（７）は、（ｎを固定した後の）ランダム変数ｘからｙへの線形変換を定義する。この変換に基づいて、ラプラス近似法により、ｘに関するＰＤＦ（式５）から、以下のｙに関するＰＤＦが取得され、

【００８０】

【数２５】

$$p(y_t|n_t)=\sum_m p(m)N\left[y_t;\mu_y(m,t),\sigma_y^2(m,t)\right]$$
$$\approx N\left[y_t;\mu_y(m_0,t),\sigma_y^2(m_0,t)\right] \quad (8)$$

【００８１】

ここで、最適な混合構成要素は、以下のように決定され、

【００８２】

【数２６】

$$m_0=\arg\max_m N\left[y_t;\mu_y(m,t),\sigma_y^2(m,t)\right]$$

【００８３】

ここで、近似ガウスの平均および分散は、以下のようになる。

【００８４】

【数２７】

$$\mu_y(m_0,t)=\mu_x(m_0)+g_{m_0}+g'_{m_0}\times(n_t-n_0)$$
$$\sigma_y^2(m_0,t)=\sigma_x^2(m_0)+g'^2_{m_0}\,\sigma_{n_t}^2 \quad (9)$$

【００８５】

以下に示すように、ｐ（$y_t|n_t$）に関するガウス推定を使用して、そのアルゴリズムを展開する。上記では、ｐ（$y_t|n_t$）に関するガウス推定を提供するために、テイラー級数展開およびラプラス近似法を使用したが、本発明を逸脱することなくガウス推定を提供するために、他の技法も使用できることを理解されたい。例えば、式（８）のラプラス近似の使用に加えて、近似または（小数の構成要素を備えた）ガウス混合モデルに関する計算技法を使用することができる。

【００８６】

事前雑音において経時変化する平均および分散を推定するためのアルゴリズムを提供可能である。式（８）のようなｐ（$y_t|n_t$）に関する、および数式４のような

【００８７】

【数２８】

$$p(n_\tau|y_1^\tau)$$

【００８８】

に関する、近似ガウス形が与えられた場合、平均μ_{n_τ}および分散

【００８９】

【数２９】

$$\sigma_{n\tau}^2$$

【００９０】

の経時変化するハイパーパラメータの逐次推定として表される、事前雑音進化を決定するためのアルゴリズムを提供可能である。式（４）および（８）を式（１）に代入すると、以下が得られ、

【００９１】
【数３０】

$$N(n_t;\mu_{n_t},\sigma^2_{n_t})$$
$$\propto N[y_t;\mu_y(m_0,t),\sigma^2_y(m_0,t)]N(n_{t-1};\mu_{n_{t-1}},\sigma^2_{n_{t-1}})$$
$$\approx N[g'_{m_0}n_{t-1};\mu_1,\sigma^2_y(m_0,t)]N(n_{t-1};\mu_{n_{t-1}},\sigma^2_{n_{t-1}}) \quad (10)$$

【００９２】
ここで、μ1=yt-μx(m0)-gm0+g'm0n0であり、雑音が平滑であるという想定を使用した。左側および右側の平均および分散は、それぞれ、事前進化公式を取得するために、以下のように式（１０）で突き合わされ、
【００９３】
【数３１】

$$\mu_{n_t}=\frac{g'_{m_0}\bar{\mu}_1\sigma^2_{n_{t-1}}+\mu_{n_{t-1}}\sigma^2_y(m_0,t-1)}{g'^2_{m_0}\sigma^2_{n_{t-1}}+\sigma^2_y(m_0,t-1)} \quad (11)$$
$$\sigma^2_{n_t}=\frac{\sigma^2_y(m_0,t-1)\sigma^2_{n_{t-1}}}{g'^2_{m_0}\sigma^2_{n_{t-1}}+\sigma^2_y(m_0,t-1)}$$

【００９４】
ここで、
【００９５】
【数３２】

$$\bar{\mu}_1=y_t-\mu_x(m_0)-g_{m0}+g'_{m0}\mu_{nt-1}$$

【００９６】
である。式（１１）を確立する際には、雑音のテイラー級数展開点として、以前の時間の事前平均、すなわち
【００９７】
【数３３】

$$n_0=\mu_{n_{t-1}}$$

【００９８】
が使用される。ガウス計算で適切に確立された結果（$a_1=g'_{m0}$に設定）も、以下のように使用した。
【００９９】
【数３４】

$$N(ax;\mu_1,\sigma_1^2)N(x;\mu_2,\sigma_2^2)=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2+K\right]$$

上式で、

$$\mu=\frac{a\mu_1\sigma_2^2+\mu_2\sigma_1^2}{a^2\sigma_2^2+\sigma_1^2};\sigma^2=\frac{\sigma_1^2\sigma_2^2}{a^2\sigma_2^2+\sigma_1^2}$$

【０１００】
単純化されても依然として効果的な想定のセットに基づいて、近似再帰的ベイズ規則の２次項突合せ（quadratic term matching）が使用され、式（１１）で要約された事前雑音進化公式を首尾よく導出する。平均雑音推定は、ＲＭＳ誤差削減によって、より正確に

測定されることがわかっているが、分散情報を使用して、信頼性の基準を提供することができる。

【０１０１】

上述した雑音推定技法は、雑音正規化技法または雑音除去で使用可能である（例えば、２００２年４月５日出願、米国特許出願第１０／１１７１４２号「Method of Noise Reduction Using Correction Vectors Based on Dynamic Aspects of Speech and Noise Normalization」明細書）。本発明は、クリーン信号を生成するために、各フレームに関して識別された推定雑音が雑音を含む信号から除去される、雑音削減システムの一部として、より直接的に使用することも可能である（例えば、２００２年９月６日出願、米国特許出願第１０／２３７１６３号「Non-linear Observation Model for Removing Noise from Corrupted Signals」明細書）。

【０１０２】

図４は、雑音削減を実行することができる本発明の雑音推定技法を使用可能な環境を示すブロック図である。具体的に言えば、図４に、本発明の雑音推定技法を使用することができる音声認識システムを示す。このシステムは、音響モデルをトレーニングするために使用されるトレーニング信号における雑音を削減し、および／または、テスト信号の言語内容を識別するために音響モデルに対して再度適用されるテスト信号における雑音を削減する。

【０１０３】

図４では、トレーナまたはユーザいずれかの話し手４００が、マイクロフォン４０４に向かって話す。マイクロフォン４０４は、１または複数の雑音源からの付加雑音４０２も受け取る。マイクロフォン４０４によって検出された音波信号は、電気信号に変換され、アナログ－デジタル変換器４０６に送られる。

【０１０４】

図４の実施形態では、付加雑音４０２はマイクロフォン４０４を介して入ってくるように示されているが、他の実施形態では、付加雑音４０２が、Ａ／Ｄ変換器４０６の後のデジタル信号として、入力音声信号に追加されることもある。

【０１０５】

Ａ／Ｄ変換器４０６は、マイクロフォン４０４からのアナログ信号を、一連のデジタル値に変換する。いくつかの実施形態では、Ａ／Ｄ変換器４０６は、１６ｋＨｚおよびサンプルあたり１６ビットで、アナログ信号をサンプリングし、それにより、毎秒３２キロバイトの音声データを作成する。これらのデジタル値は、フレームコンストラクタ４０７に送られ、一実施形態では、１０ミリ秒ごとに開始される２５ミリ秒フレームに、この値をグループ化する。

【０１０６】

フレームコンストラクタ４０７によって作成されたデータのフレームは、特徴抽出器４０８に送られ、各フレームから特徴を抽出する。特徴抽出モジュールの例は、線形予測符号化（ＬＰＣ）、ＬＰＣ導出ケプストラム、知覚線形予測（ＰＬＰ）、聴覚モデル特徴抽出、およびメル周波数ケプストラム係数（ＭＦＣＣ）特徴抽出を実行するためのモジュールを含む。本発明は、これらの特徴抽出モジュールに限定されるものではなく、本発明の文脈において、他のモジュールも使用可能であることに留意されたい。

【０１０７】

特徴抽出モジュールは、それぞれが音声信号のフレームに関連付けられた、特徴ベクトルのストリームを生成する。この特徴ベクトルのストリームが雑音削減モジュール４１０に送られ、本発明の雑音推定技法を使用して、各フレーム内の雑音を推定する。

【０１０８】

雑音削減モジュール４１０の出力は、一連の「クリーン」特徴ベクトルである。入力信号がトレーニング信号の場合、この一連の「クリーン」特徴ベクトルはトレーナ４２４に送られ、「クリーン」特徴ベクトルおよびトレーニングテキスト４２６を使用して、音響モデル４１８をトレーニングする。こうしたモデルをトレーニングする技法は、当分野で

知られており、本発明を理解するために、それらについて説明する必要はない。

【０１０９】

入力信号がテスト信号の場合、「クリーン」特徴ベクトルは復号器４１２に送られ、これが、特徴ベクトルのストリームに基づいて、最も可能性の高い用語のシーケンス、辞書４１４、言語モデル４１６、および音響モデル４１８を識別する。復号に使用される特定の方法は、本発明にとって重要ではなく、復号用のいくつかの知られた方法のうちのいずれも使用可能である。

【０１１０】

最も可能性の高い仮説用語のシーケンスが、信頼測度モジュール４２０に送られる。信頼測度モジュール４２０は、部分的に２次音響モデル（図示せず）に基づいて、どの用語が音声認識装置によって最も不適切に識別された可能性が高いかを識別する。次に、信頼測度モジュール４２０は、仮説用語のシーケンスを、どの用語が不適切に識別された可能性があるかを示す識別子と共に、出力モジュール４２２に送る。当分野の技術者であれば、本発明を実施するために、信頼測度モジュール４２０が必ずしも必要でないことを理解されよう。

【０１１１】

図４は音声認識システムを示しているが、本発明は、どんなパターン認識システムでも使用可能であり、音声に限定されるものではない。

【０１１２】

以上、本発明について、特定の実施形態を参照しながら説明してきたが、当分野の技術者であれば、本発明の精神および範囲を逸脱することなく、形式および詳細が変更可能であることを理解されよう。

【図面の簡単な説明】

【０１１３】

【図１】本発明を実施することができるコンピューティング環境を示すブロック図である。

【図２】本発明を実施することができる代替コンピューティング環境を示すブロック図である。

【図３】本発明の一実施形態の下で雑音を推定する方法を示す流れ図である。

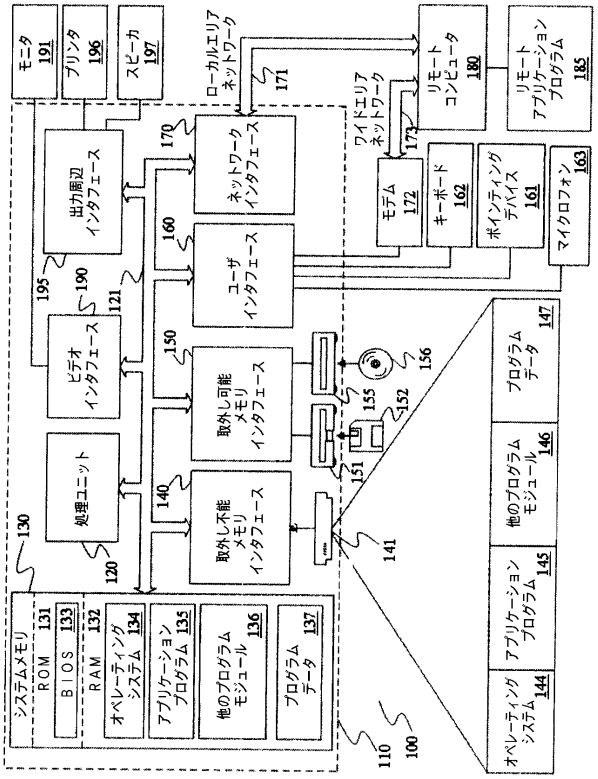
【図４】本発明を使用することができるパターン認識システムを示すブロック図である。

【符号の説明】

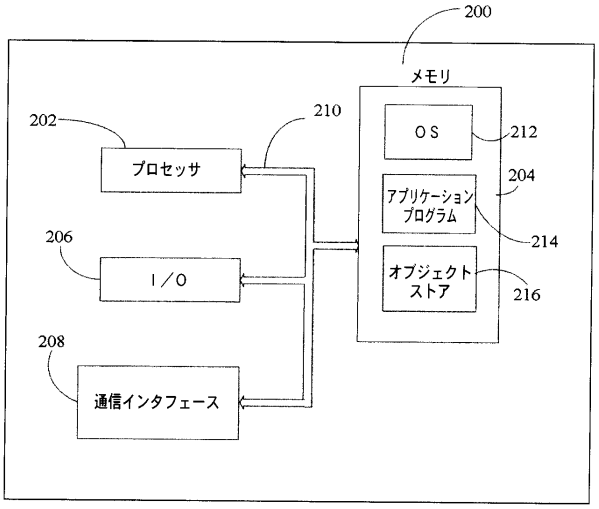
【０１１４】

４００　話し手
４０２　付加雑音
４０４　マイクロフォン
４０６　アナログ－デジタル変換器
４０７　フレームコンストラクタ
４０８　特徴抽出器
４１０　雑音削減モジュール
４１２　復号器
４１４　辞書
４１６　言語モデル
４１８　音響モデル
４２０　信頼測度モジュール
４２２　出力モジュール
４２４　トレーナ
４２６　トレーニングテキスト

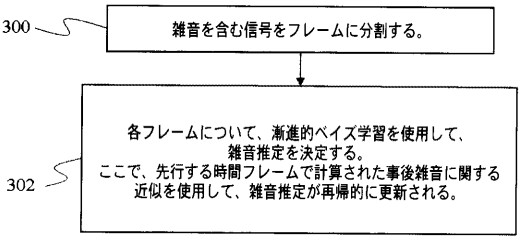
【図１】


モニタ 191
プリンタ 196
スピーカ 197
ローカルエリアネットワーク 171
ワイドエリアネットワーク 173
リモートコンピュータ 180
リモートアプリケーションプログラム 185
出力周辺インタフェース 195
ネットワークインタフェース 170
ユーザインタフェース 160
モデム 172
キーボード 162
ポインティングデバイス 161
マイクロフォン 163
ビデオインタフェース 190
取外し可能メモリインタフェース 150
取外し不能メモリインタフェース 140
処理ユニット 120
プログラムデータ 147
他のプログラムモジュール 146
アプリケーションプログラム 145
オペレーティングシステム 144
システムメモリ 130
ROM 131
BIOS 133
RAM 132
オペレーティングシステム 134
アプリケーションプログラム 135
他のプログラムモジュール 136
プログラムデータ 137
110
100

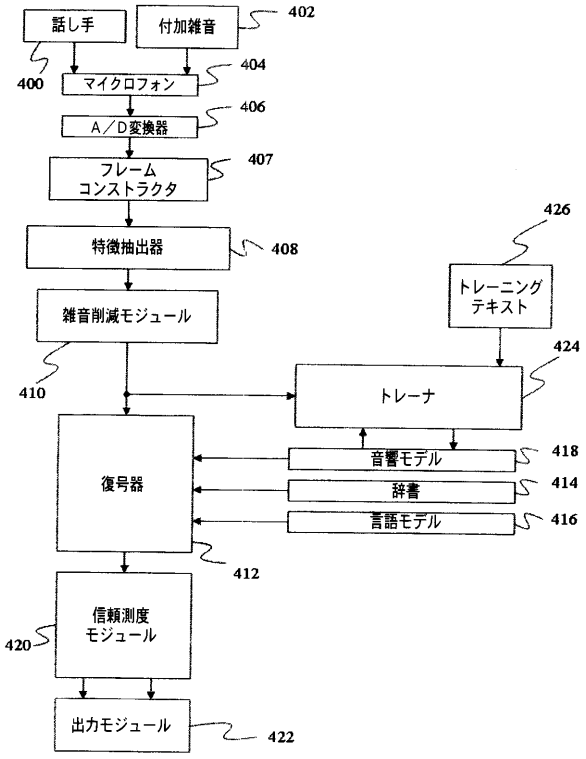
【図２】


200
メモリ
202
210
プロセッサ
OS 212
アプリケーションプログラム 204
214
206
I/O
オブジェクトストア 216
208
通信インタフェース

【図３】


300
雑音を含む信号をフレームに分割する。
各フレームについて、漸進的ベイズ学習を使用して、雑音推定を決定する。
ここで、先行する時間フレームで計算された事後雑音に関する近似を使用して、雑音推定が再帰的に更新される。
302

【図４】


話し手 400
付加雑音 402
マイクロフォン 404
A/D変換器 406
フレームコンストラクタ 407
特徴抽出器 408
雑音削減モジュール 410
トレーニングテキスト 426
トレーナ 424
復号器 412
音響モデル 418
辞書 414
言語モデル 416
信頼測度モジュール 420
出力モジュール 422

フロントページの続き

(72)発明者 デン リ
アメリカ合衆国 ９８０５３ ワシントン州 レッドモンド ノースイースト ３０ ストリート ２２３１０
(72)発明者 ジェームズ ジー. ドロッポ
アメリカ合衆国 ９８０１９ ワシントン州 ドボル ２８２ コート ノースイースト １３８２０

審査官 山下 剛史

(56)参考文献 特開平１１－２９６５１５（ＪＰ，Ａ）
特開２００２－１２３２８５（ＪＰ，Ａ）
特表平１０－５１２６８６（ＪＰ，Ａ）
国際公開第０２／２３８４２（WO，Ａ１）
特開平１１－１５４９１（ＪＰ，Ａ）
Wu Chou, et al., "Decision tree state tying based on penalized Bayesian information criterion", Proceedings of the 1999 IEEE International Conference, Acoustics, Speech, and Signal Processing(ICASSP'99), １９９９年 3月, pp.345-348

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