(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6801095号
(P6801095)
(45)発行日　令和2年12月16日(2020.12.16)　(24)登録日　令和2年11月27日(2020.11.27)

(51)Int.Cl.　　F I
G10L 15/28 (2013.01)　　G10L 15/28 230K
G10L 15/14 (2006.01)　　G10L 15/14 200Z

請求項の数 7　(全 16 頁)

(21)出願番号　特願2019-517762(P2019-517762)
(86)(22)出願日　平成29年9月26日(2017.9.26)
(65)公表番号　特表2019-533193(P2019-533193A)
(43)公表日　令和1年11月14日(2019.11.14)
(86)国際出願番号　PCT/CN2017/103514
(87)国際公開番号　WO2018/059405
(87)国際公開日　平成30年4月5日(2018.4.5)
審査請求日　令和1年5月24日(2019.5.24)
(31)優先権主張番号　201610867477.9
(32)優先日　平成28年9月29日(2016.9.29)
(33)優先権主張国・地域又は機関
中国(CN)

(73)特許権者　516320344
合肥華凌股▲フン▼有限公司
HEFEI HUALING CO., LTD.
中華人民共和国　230601　安徽省合肥市合肥経済技術開発区錦繍大道176号
No. 176 JinXiu Road, Hefei Economic And Technological Development Area Hefei, Anhui, 230601, China

最終頁に続く

(54)【発明の名称】音声制御システム及びそのウェイクアップ方法、ウェイクアップ装置、並びに家電製品、コプロセッサ

(57)【特許請求の範囲】
【請求項１】
音声制御システムであって、
前記音声制御システムは、音声認識アセンブリ、及びウェイクアップ装置を含み、前記音声認識アセンブリが、前記ウェイクアップ装置のコプロセッサに接続され、
前記音声認識アセンブリは、動作アクティブ状態の際、音声認識のために使用され、音声認識後、非動作の休眠状態に入り、
前記音声認識アセンブリの前記非動作の休眠状態から前記動作アクティブ状態への切り替えは、前記コプロセッサによりウェイクアップされ、
前記音声認識アセンブリは、前記動作アクティブ状態から前記非動作の休眠状態への切り替えの前に、待ち状態に入り、
設定された期間内に、前記音声認識アセンブリがウェイクアップされていない場合、前記非動作の休眠状態に入り、前記音声認識アセンブリがウェイクアップされた場合、前記動作アクティブ状態に入り、
前記ウェイクアップ装置は、音声収集アセンブリ、及びコプロセッサを含み、
前記音声収集アセンブリは音声情報を収集するためのものであり、
前記コプロセッサは、前記音声収集アセンブリにより収集された前記音声情報を処理して前記音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離して、人間の声を含む音声情報セグメントに対してウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、音声認識アセンブリを

ウェイクアップするためのものであり、
前記コプロセッサは、収集された音声情報を処理して音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離するための、処理モジュールと、
前記処理モジュールによって分離された、人間の声を含む音声情報セグメントに対して、ウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、ウェイクアップコマンドを生成するための、認識モジュールと、
前記ウェイクアップコマンドに従って、音声認識プロセッサをウェイクアップするための、ウェイクアップモジュールと、を含む、
ことを特徴とする音声制御システム。

【請求項２】
前記処理モジュールは、分離ユニット、判定ユニットを含み、
前記分離ユニットは、非ガウス値の最大である音声信号を分離するように、デジタル信号フォーマットの前記音声情報に対してブラインド音源分離処理を行うためのものであり、
前記判定ユニットは、エネルギー閾値により前記音声信号に人間の声を含んでいるかどうかを判定し、エネルギー閾値を超える場合、人間の声を含む音声情報を分離して、人間の声を含む音声情報セグメントを得るためのものであることを特徴とする請求項１に記載の音声制御システム。

【請求項３】
前記認識モジュールは、認識ユニット、記憶ユニットを含み、
前記記憶ユニットは、ウェイクアップワードモデルを記憶するためのものであり、
前記認識ユニットは、前記判定ユニットにより分離された人間の声を含む音声情報セグメント、と前記記憶ユニットに記憶されている前記ウェイクアップワードモデルに対してウェイクアップワードの照合を行い、照合が成功した場合、ウェイクアップコマンドを生成するためのものであることを特徴とする請求項２に記載の音声制御システム。

【請求項４】
ウェイクアップワード音声モデルの構築は、
何人かの人間のウェイクアップ音声データを収集することと、
全ての前記ウェイクアップ音声データを処理して、トレーニングすることにより、ウェイクアップワードモデルを得ることと、を含むことを特徴とする請求項３に記載の音声制御システム。

【請求項５】
前記の、ウェイクアップワード音声モデルを構築することは、
オフライン状態で、様々な環境で録音された発話者からのウェイクアップワードを収集し、フレーム分割処理を行うことと、
フレーム分割の後、特徴パラメータを抽出することと、
前記特徴パラメータに対してクラスタリングを行って、隠れマルコフモデルHMM（Hidden Markov Model）の観測状態を確立することと、
Baum-Welchアルゴリズムにより隠れマルコフモデルHMMのパラメータを調整して、P（σ|λ）（ただし、λはモデルパラメータであり、σは観察状態である）を最大化し、モデルパラメータλを調整して、観察状態σの最大確率を得て、モデルトレーニングを完成させて、ウェイクアップワード音声モデルを記憶することとを含み、
前記認識モジュールは、
人間の声のデータを含む音声フレームから特徴パラメータを抽出して、新しい観察値σ'の集合を新しい観察状態として得て、P（σ'｜λ）を計算し、
P（σ'｜λ）を信頼閾値と比較して、ウェイクアップワードが認識されたかどうかを得ることを特徴とする請求項４に記載の音声制御システム。

【請求項６】
前記音声収集アセンブリは、音声収集モジュール、及びA/D変換モジュールを含み、

前記音声収集モジュールは、アナログ信号フォーマットの音声情報を収集するためのものであり、

前記A/D変換モジュールは、前記アナログ信号フォーマットの音声情報をデジタル変換して、デジタル信号フォーマットの音声情報を得るためのものであることを特徴とする請求項１に記載の音声制御システム。

【請求項７】

請求項１～６のいずれか一項に記載の音声制御システム、及び家電製品本体を含み、前記家電製品本体が、前記音声制御システムに接続されることを特徴とするスマート家電製品。

【発明の詳細な説明】

【技術分野】

【０００１】

相互参照

本願は、2016年9月29日に提出された特許の名称が「音声制御システム及びそのウェイクアップ方法、ウェイクアップ装置、及び家電製品、コプロセッサ」である第2016108674779号の中国特許出願を引用し、その全体が参照により本願の明細書に組み込まれる。

【０００２】

本発明は、家電機器の音声制御の分野に関し、特に、音声制御システムおよびそのウェイクアップ方法、ウェイクアップ装置、並びに家電機器、コプロセッサに関する。

【背景技術】

【０００３】

人工知能技術の発展に伴い、家電業界は、新たな発展が始まり、そのうちマン-マシンの音声対話は、人間の使用習慣に合わせるものであるため、研究のホットな課題の1つとなった。図1は音声制御機能付きの家電製品の回路を示す。図1から分かるように、音声制御機能を高めるためには、従来の制御回路に音声制御回路を追加する必要がある。音声制御は、外部の音をリアルタイムに監視する必要があるので、認識を行うプロセッサは常に動作し、その結果、消費電力が増大する。

【発明の概要】

【発明が解決しようとする課題】

【０００４】

本発明は、人間の声が存在し且つ人間の声に認識すべき音声が含まれる場合にのみ、音声認識アセンブリ（音声認識プロセッサCPU）を起動するという課題を解決するために、音声制御システム及びそのウェイクアップ方法、ウェイクアップ装置、並びにスマート家電製品を提供することを目的とする。

【課題を解決するための手段】

【０００５】

本発明は、上記の課題を解決するために、音声制御システムのウェイクアップ方法を提供する。当該音声制御システムのウェイクアップ方法は、

音声情報を収集する、収集ステップと、

前記音声情報を処理して前記音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離して、認識ステップに移行する、処理ステップと、

人間の声を含む音声情報セグメントに対して、ウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、ウェイクアップステップに移行し、ウェイクアップワードが認識されない場合、前記収集ステップに戻る、認識ステップと、

音声認識プロセッサをウェイクアップする、ウェイクアップステップと、を含む。

【０００６】

いくつかの実施例では、好ましくは、前記音声情報は、異なる期間に収集された複数の音声情報セグメントから構成され、全ての前記期間が繋がって完全かつ連続的なタイムチェーンになり、及び/又は、

前記収集ステップは、
アナログ信号フォーマットの音声情報を収集することと、
前記アナログ信号フォーマットの音声情報をデジタル変換してデジタル信号フォーマットの音声情報を得ることと、を含む。
【０００７】
いくつかの実施例では、好ましくは、前記ウェイクアップステップの前に、前記ウェイクアップ方法は、ウェイクアップワード音声モデルを構築することをさらに含み、
前記認識ステップは、人間の声を含むデータを前記ウェイクアップワード音声モデルと照合し、照合が成功した場合、ウェイクアップワードが認識されたと判定し、照合が成功しなかった場合、ウェイクアップワードが認識されないと判定すること、を含む。
【０００８】
いくつかの実施例では、好ましくは、前記の、ウェイクアップワード音声モデルを構築することは、
何人かの人間のウェイクアップ音声データを収集することと、
全ての前記ウェイクアップ音声データを処理して、トレーニングするにより、ウェイクアップワードモデルを得ることと、を含む。
【０００９】
いくつかの実施例では、好ましくは、前記の、ウェイクアップワード音声モデルを構築することは、
オフライン状態で、様々な環境で録音された発話者からのウェイクアップワードを収集し、フレーム分割処理を行うことと、
フレーム分割の後、特徴パラメータを抽出することと、
前記特徴パラメータに対してクラスタリングを行って、隠れマルコフモデルHMM（Hidden Markov Model）の観測状態を確立することと、
Baum-Welchアルゴリズムにより隠れマルコフモデルHMMのパラメータを調整して、P（σ|λ）（ただし、λはモデルパラメータであり、σは観察状態である）を最大化し、モデルパラメータλを調整して、観察状態σの最大確率を得て、モデルトレーニングを完成させて、ウェイクアップワード音声モデルを記憶することと、を含む。
【００１０】
前記認識ステップは、
人間の声を含むデータの音声フレームから特徴パラメータを抽出して、新しい観察値σ'の集合を新しい観察状態として得て、P（σ'｜λ）を計算することと、
P（σ'｜λ）を信頼閾値と比較して、ウェイクアップワードが認識されたかどうかを得ることと、を含む。
【００１１】
いくつかの実施例では、好ましくは、前記処理工程は、
非ガウス値の最大である音声信号を分離するように、デジタル信号フォーマットの前記音声情報に対してブラインド音源分離処理を行う、第1の分離ステップと、
エネルギー閾値により前記音声信号に人間の声を含んでいるかどうかを判定し、エネルギー閾値を超える場合、人間の声を含んでいると判定して、第2の分離ステップに移行し、エネルギー閾値を超えない場合、人間の声を含んでいないと判定して、前記収集ステップに移行する、判定ステップと、
人間の声を含む音声情報を分離し、人間の声を含む音声情報セグメントを得る、第2の分離ステップと、を含む。
【００１２】
いくつかの実施例では、好ましくは、前記第1の分離ステップにおいて、前記ブラインド音源分離処理が採用する方法は、負のエントロピーの最大化、4次の統計量の尖度、又は時間-周波数変換に基づいた独立成分分析のICAアルゴリズムである。
【００１３】
また、本発明の他の態様は、コプロセッサを提供する。当該コプロセッサは、

収集された音声情報を処理して音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離するための、処理モジュールと、
前記処理モジュールによって分離された、人間の声を含む音声情報セグメントに対して、ウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、ウェイクアップコマンドを生成するための、認識モジュールと、
前記ウェイクアップコマンドに従って、音声認識プロセッサをウェイクアップするための、ウェイクアップモジュールと、を含む。

【００１４】

いくつかの実施例では、好ましくは、前記処理モジュールは、分離ユニット、判定ユニットを含み、
前記分離ユニットは、非ガウス値の最大である音声信号を分離するように、デジタル信号フォーマットの前記音声情報に対してブラインド音源分離処理を行うためのものであり、
前記判定ユニットは、エネルギー閾値により前記音声信号に人間の声を含んでいるかどうかを判定し、エネルギー閾値を超える場合、人間の声を含む音声情報を分離して、人間の声を含む音声情報セグメントを得るためのものである。

【００１５】

いくつかの実施例では、好ましくは、前記認識モジュールは、認識ユニット、記憶ユニットを含み、
前記記憶ユニットは、ウェイクアップワードモデルを記憶するためのものであり、
前記認識ユニットは、前記判定ユニットにより分離された人間の声を含む音声情報セグメント、と前記記憶ユニットに記憶されている前記ウェイクアップワードモデルに対してウェイクアップワードの照合を行い、照合が成功した場合、ウェイクアップコマンドを生成するためのものである。

【００１６】

いくつかの実施例では、好ましくは、前記ウェイクアップワード音声モデルの構築は、
何人かの人間のウェイクアップ音声データを収集することと、
全ての前記ウェイクアップ音声データを処理して、トレーニングすることにより、ウェイクアップワードモデルを得ることと、を含む。

【００１７】

いくつかの実施例では、好ましくは、前記の、ウェイクアップワード音声モデルを構築することは、
オフライン状態で、様々な環境で録音された発話者からのウェイクアップワードを収集し、フレーム分割処理を行うことと、
フレーム分割の後、特徴パラメータを抽出することと、
前記特徴パラメータに対してクラスタリングを行って、隠れマルコフモデルHMM（Hidden Markov Model）の観測状態を確立することと、
Baum-Welchアルゴリズムにより隠れマルコフモデルHMMのパラメータを調整して、P（σ|λ）（ただし、λはモデルパラメータであり、σは観察状態である）を最大化し、モデルパラメータλを調整して、観察状態σの最大確率を得て、モデルトレーニングを完成させて、ウェイクアップワード音声モデルを記憶することと、を含む。

【００１８】

前記認識ステップは、
人間の声のデータを含む音声フレームから特徴パラメータを抽出して、新しい観察値σ'の集合を新しい観察状態として得て、P（σ'｜λ）を計算することと、
P（σ'｜λ）を信頼閾値と比較して、ウェイクアップワードが認識されたかどうかを得ることと、を含む。

【００１９】

また、本発明の別の態様は、音声制御システムのウェイクアップ装置を提供する。当該音声制御システムのウェイクアップ装置は、音声収集アセンブリ、及び前記のコプロセッ

サを含み、そのうち、
前記音声収集アセンブリは音声情報を収集するためのものであり、
前記コプロセッサは、前記音声収集アセンブリにより収集された前記音声情報を処理して前記音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離して、人間の声を含む音声情報セグメントに対してウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、音声認識アセンブリをウェイクアップするためのものである。
【００２０】
いくつかの実施例では、好ましくは、前記音声収集アセンブリは、音声収集モジュール、及びA/D変換モジュールを含み、
前記音声収集モジュールは、アナログ信号フォーマットの音声情報を収集するためのものであり、
前記A/D変換モジュールは、前記アナログ信号フォーマットの音声情報をデジタル変換して、デジタル信号フォーマットの音声情報を得るためのものである。
【００２１】
また、本発明の別の態様は、音声制御システムを提供する。当該音声制御システムは、音声認識アセンブリ、及び前記のウェイクアップ装置を含み、前記音声認識アセンブリが、前記ウェイクアップ装置のコプロセッサに接続され、
前記音声認識アセンブリは、動作アクティブ状態の際、音声認識のために使用され、音声認識後、非動作の休眠状態に入り、
前記音声認識アセンブリの前記非動作の休眠状態から前記動作アクティブ状態への切り替えは、前記コプロセッサによりウェイクアップされる。
【００２２】
いくつかの実施例では、好ましくは、前記音声認識アセンブリは、前記動作アクティブ状態から前記非動作の休眠状態への切り替えの前に、待ち状態に入る。
【００２３】
設定された期間内に、前記音声認識アセンブリがウェイクアップされていない場合、前記非動作の休眠状態に入り、前記音声認識アセンブリがウェイクアップされた場合、前記動作アクティブ状態に入る。
【００２４】
また、本発明の他の態様は、スマート家電製品を提供する。当該スマート家電製品は、前記音声制御システム、及び家電製品本体を含み、前記家電製品本体が、前記音声制御システムに接続される。
【発明の効果】
【００２５】
本発明に提供される技術は、ウェイクアップ技術を追加し、音声のウェイクアップ装置を利用して補助処理又は前処理装置とし、常に音声情報を収集し、且つ音声情報を分析し認識して、音声にウェイクアップワードを含んでいると確認した場合、音声認識プロセッサをウェイクアップして、音声認識を行う。このような形態により、音声認識プロセッサは、音声認識が必要とされる時のみ動作し、全天候型の連続稼動を回避し、エネルギー消費は明らかに削減される。音声ウェイクアップ装置は、音声全体を認識する必要がなく、ウェイクアップワードのみを認識するため、消費電力が低い。全天候型の連続稼動をしても、エネルギー消費も非常に低いので、従来の音声認識の高消費電力の問題を解決できる。
【図面の簡単な説明】
【００２６】
【図１】従来の技術に係る音声制御機能付きの家電製品の回路の概略構成図である。
【図２】本発明の一実施例に係るコプロセッサの概略構成図である。
【図３】本発明の一実施例に係る音声制御システムのウェイクアップ装置の概略構成図である。

【図４】本発明の一実施例に係る、ウェイクアップ装置を有する音声制御システムの概略構成図である。
【図５】本発明の一実施例に係る音声制御システムのウェイクアップ方法のフローチャートである。
【図６】本発明の一実施例に係るウェイクアップワード認識において使用されるパスワード認識モデルである。
【図７】本発明の一実施例に係るウェイクアップワードモデルの構築のフローチャートである。
【図８】本発明の一実施例に係るウェイクアップワードの認識のフローチャートである。
【図９】本発明の一実施例に係る音声認識アセンブリの状態変換の概略図である。
【発明を実施するための形態】
【００２７】
以下に、図面および実施形態を参照して本発明の具体的な実施形態をさらに詳細に説明する。以下の実施例は本発明を説明するためのもので、本発明の範囲を制限するものではない。
【００２８】
本発明の説明では、明確な規定と限定がない限り、「取り付け」、「互いに接続」、「接続」の用語の意味は広く理解されるべきであり、例えば、固定接続や、取り外し可能な接続や、または一体型接続でも可能であり、機械的な接続や、電気的接続でも可能であり、直接接続することや、中間媒体を介して間接接続することでも可能であり、また2つの素子の内部が連通することでも可能である。
【００２９】
本発明は、家庭用電器の音声制御回路の消費電力を低減するために、音声制御システムのウェイクアップ方法、ウェイクアップ装置、音声制御システム、及びスマート家電製品を提供する。
【００３０】
以下は、基本設計、代替設計、および拡張設計により、本技術を詳細に説明する。
【００３１】
音声認識のエネルギー消費量を削減するコプロセッサであって、図2に示すように、主に従来の音声認識プロセッサのフロントエンドに適用され、早期の音声処理に使用されており、ウェイクアップコマンドを取得するにより、音声認識プロセッサをウェイクアップし、音声認識プロセッサの動作時間の長さを、音声認識を必要とする期間に短縮し、低電力のコプロセッサはエネルギー損失が低く、損失を大幅に低減することができる。この機能に基づいて、当該コプロセッサは、収集された音声情報を処理して、音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離する処理モジュールと、処理モジュールによって分離された、人間の声を含む音声情報セグメントに対して、ウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、ウェイクアップコマンドを生成する認識モジュールと、ウェイクアップコマンドに従って、音声認識プロセッサをウェイクアップするウェイクアップモジュールと、を主に含む。その作動過程は図5に示す。
【００３２】
収集された音声には収集環境における様々な音を含み、人間の声を効果的に分離して認識することはその後の処理の最初のステップであるので、処理モジュールにより人間の声を含む音声セグメントを分離する必要がある。しかし、人間の声を含む音声セグメントの内容には大量の情報を含み、すべての情報が音声認識を必要とするわけではないので、音声セグメントに含まれるいくつかの特別なワードを認識し、これらの特別なワードによって、当該音声セグメントが音声認識を必要とする情報であることを決定するため、従来の音声認識処理装置の負担をさらに軽減することができる。従って、本実施形態では、特別なワードをウェイクアップワードとし、ウェイクアップワードで、音声認識処理装置をウェイクアップすることを決定する。

【００３３】
ただし、いくつかの実施例では、処理モジュールに受信された収集音声情報は、通常、期間を収集分割方式とし、音声収集アセンブリは1つの期間内に収集された音声情報セグメントを一つの送信対象として処理モジュールに送信し、そして次の期間の音声収集を続ける。当該コプロセッサは、音声収集アセンブリと音声認識プロセッサとの間に別個のハードウェアとして搭載することができる。
【００３４】
当該コプロセッサは、低消費電力のDSPを採用してもよく、従来の音声認識プロセッサの内部に搭載されたチップ、または従来の音声収集アセンブリの内部に搭載されたチップであってもよい。チップは、処理モジュール、認識モジュール、及びウェイクアップモジュールを有し、音声処理とウェイクアップ機能を実現する。
【００３５】
そのうち、処理モジュールは、主に、分離ユニットと判定ユニットとからなり、分離ユニットは、非ガウス値の最大である音声信号を分離するように、デジタル信号フォーマットの音声情報に対してブラインド音源分離処理を行う。判定ユニットは、エネルギー閾値により音声信号に人間の声を含んでいるかどうかを判定し、エネルギー閾値を超える場合、人間の声を含んでいる音声情報を分離して、人間の声を含む音声情報セグメントが得られる。
【００３６】
ブラインド音源分離の役割は、信号源が未知の場合に複数の信号源を分離することである。そのうちICAは、一般的なアルゴリズムであり、負のエントロピーの最大化、4次の統計量の尖度（kurtosis）、及び時間-周波数変換の方法に基づいて実現することができる。そして、固定小数点高速アルゴリズムは、DSP上でリアルタイムに実現しやすい。
【００３７】
音声信号は、ラプラス分布に従うため、スーパーガウス分布に属し、そして大部分ノイズの分布はガウス特性を有する。負のエントロピー、尖度（kurtosis）などは、信号の非ガウス性に対して測定できる。その値が大きいほど、非ガウス特性が大きくなるため、信号うちの該値の最大である信号を選択し分離して、処理する。
【００３８】
可能な信号を選択した後、エネルギー閾値によって発話者の音声があるかどうかを判定する。音声を含むフレームは、認識モジュールに送信されてウェイクアップワードの認識プロセスを行い、その後の処理には音声のフレームドロップを含まない。
【００３９】
認識モジュールは、認識ユニットと記憶ユニットとを含む。記憶ユニットは、ウェイクアップワードモデルを記憶する。認識ユニットは、判定ユニットにより分離された人間の声を含む音声情報セグメント、と記憶ユニットに記憶されているウェイクアップワードモデルに対してウェイクアップワードの照合を行い、照合が成功した場合、ウェイクアップコマンドを生成する。
【００４０】
ウェイクアップワードの認識は、予め設定されたウェイクアップワード（ウェイクアップワードモデルからのもの）（例えば「こんにちは、冷蔵庫」のようなもの）に基づいて、ユーザが音声制御を試してみるかどうかを確認する。基本的なプロセスは次のとおりである。
【００４１】
1.多数の発話者の音声により、ウェイクアップワードモデルを予め確立する。
【００４２】
2.トレーニング後のウェイクアップワードモデルを（ソリッドステートストレージスペース（フラッシュ；flash））に格納し、電源投入後にキャッシュ（記憶ユニット）にコピーする。
【００４３】

3.音声処理の際、先に取得した人間の声を含む音声情報セグメントをモデルと照合して、ウェイクアップワードであるか否かの判定を得る。

【００４４】

4.ウェイクアップワードであるか否かを確認する。コプロセッサがウェイクアップワードを検出した後、中断し、ウェイクアップ音声認識プロセッサが動作する。ウェイクアップワードが検出されない場合、ウェイクアップパスワードの入力を待ち続ける。

【００４５】

ウェイクアップワード音声モデルの構築は、以下のような方法を採用することができる。何人かの人間のウェイクアップ音声データを収集し、全てのウェイクアップ音声データを処理してトレーニングするにより、ウェイクアップワードモデルを得る。

【００４６】

いくつかの実施例では、ウェイクアップワードの認識は、一般的に使用されるGMM-HMM（現在、DNN-HMMモデル、ＬＳＴＭモデルも一般的に使用される）モデルを採用してYes・Noの判定を行う。そのパスワード認識モデルを図6に示す。

【００４７】

GMMモデルは音声フレームに対してクラスタリングを行うものである。

【００４８】

HMMモデルは、2つの状態集合及び3つの遷移確率によって記述できる。

【００４９】

2つの状態集合は、観察可能な状態である観測可能状態Ｏと、

状態シンボルがマルコフ特性（時刻tの状態は時刻t-1のみに相関する）に合致し、一般的に直接に観測することはできない隠れ状態Sと、を含む。

初期状態状態確率行列：初期状態の各隠れ状態の確率分布を表すものである。

状態遷移行列：時刻tからｔ＋1までの隠れ状態間の遷移確率を表すものである。

観測状態出力確率：隠れ状態がsであるという条件下で観測値がoになる確率を表すものである。

【００５０】

HMMには3つの問題がある。

【００５１】

1.評価の問題である。観測シーケンスとモデルを特定し、ある特定の出力の確率を求める。パスワード認識タスクでは、音声シーケンスとモデルに基づいて、当該シーケンスがあるセンテンスである可能性を確認することである。

【００５２】

2.デコードの問題である。観測シーケンスとモデルを特定し、観測の確率を最大にする隠れ状態のシーケンスを探す。

【００５３】

3.学習の問題である。観察シーケンスを特定し、モデルパラメータを調整することで、当該観察シーケンスを生成する確率を最大にする。パスワード認識タスクでは、多数のパスワードに基づいてモデルパラメータを調整することである。

【００５４】

これらの実施形態では、具体的にウェイクアップワード音声モデルの構築を、図7に示すように、以下の方式で実施することができる。

【００５５】

オフライン状態では、様々な環境で録音された発話者からのウェイクアップワードを収集し、フレーム分割処理を行う。

【００５６】

フレーム分割の後、特徴パラメータ（MFCCなど）を抽出する。

【００５７】

GMMにより特徴パラメータに対してクラスタリングを行って、隠れマルコフモデルHMMの観測状態を確立する。

【００５８】
Baum-Welchアルゴリズムにより隠れマルコフモデルHMMのパラメータを調整して、P（σ|λ）（ただし、λはモデルパラメータであり、σは観察状態である）を最大化し、モデルパラメータλを調整して、観察状態σの最大確率を得て、モデルトレーニングを完成させて、ウェイクアップワード音声モデルを記憶する。
【００５９】
図8に示すように、ウェイクアップワードを構築するステップに基づいて、認識ステップは以下の通りである。
【００６０】
人間の声のデータを含む音声フレームから特徴パラメータを抽出して、新しい観察値σ'の集合を新しい観察状態として得て、P（σ'｜λ）を計算する
【００６１】
P（σ'｜λ）を信頼閾値と比較して、ウェイクアップワードが認識されたかどうかを得る。
【００６２】
場合によっては、閾値は実験により得られた経験値であり、異なるウェイクアップワードに必要とする閾値は実験によって調整可能である。
【００６３】
また、技術をより全面的に保護するために、音声制御システムのウェイクアップ装置も保護する。図3に示すように、当該装置は、主に音声収集アセンブリ、及び上述のコプロセッサから構成される。音声収集アセンブリは音声情報を収集するためのものである。コプロセッサは、音声収集アセンブリによって収集された音声情報を処理して音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離して、人間の声を含む音声情報セグメントに対してウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、音声認識アセンブリをウェイクアップするためのものである。
【００６４】
いくつかの実施例では、特に新製品を開発する際、音声収集アセンブリ及びコプロセッサを一体部材に統合することができる。音声認識を起動するために、両者は、収集し分析した後、音声認識プロセッサをウェイクアップするかどうかを決定するので、両者は音声認識プロセッサの作動時間を大幅に短縮し、作業損失を減少することができる。
【００６５】
そのうち、音声収集機能を持つ全ての部材は音声収集アセンブリに適用できる。音声収集アセンブリは、主に音声収集モジュール、及びA/D変換モジュールから構成される。音声収集モジュールは、アナログ信号フォーマットの音声情報を収集するためのものである。A/D変換モジュールは、アナログ信号フォーマットの音声情報をデジタル変換して、デジタル信号フォーマットの音声情報を得るためのものである。
【００６６】
いくつかの実施例では、音声収集モジュールとA/D変換モジュールは、別体のハードウェアデバイスであってもよく、音声収集アセンブリに統合される一体構造であってもよい。
【００６７】
一方、技術をより十分に保護するために、音声制御システムも提供される。図4に示すように、当該音声制御システムは、音声収集、音声処理及び音声認識に用いられ、認識結果により、音声内の制御コマンドを得る。当該音声制御システムは、主に音声認識アセンブリ（音声認識プロセッサ）とウェイクアップ装置とから構成され、音声認識アセンブリはウェイクアップ装置のコプロセッサと接続され、コプロセッサは、ウェイクアップワードを検出した後に、音声認識アセンブリをウェイクアップして音声認識動作を行わせる。音声認識アセンブリは、動作アクティブ状態の際、音声認識のために使用され、音声認識後、非動作の休眠状態に入る。音声認識アセンブリの非動作の休眠状態から動作アクティ

ブ状態への切り替えは、コプロセッサによりウェイクアップされる。
【００６８】
場合によって音声収集及び音声処理に一定の時間がかかることを考慮すると、複数の連続したウェイクアップ操作が発生することがある。このため、音声認識プロセッサが一つの人間の声を含む音声セグメントを認識した後、一定の時間の待ち状態に入る。図9に示すように、待ち状態において、認識すべき音声セグメントの情報が入ると認識し続け、認識すべき音声セグメントがなければ非動作の休眠状態に入る。即ち、音声認識アセンブリは、動作アクティブ状態から非動作の休眠状態への切り替えの前に、待ち状態に入る。設定された期間内に、音声認識アセンブリがウェイクアップされていない場合、非動作の休眠状態に入り、音声認識アセンブリがウェイクアップされた場合、動作アクティブ状態に入る。
【００６９】
上記の音声制御システムを、スマート家電製品に応用する。当該スマート家電製品は、主に音声制御システムと家電本体とから構成され、家電製品本体が、音声制御システムに接続される。
【００７０】
スマート家電製品は、制御コマンドを必要とする家庭内でのあらゆる家電機器であってもよい。
【００７１】
同時に、本発明は、スマート家電製品を動作中の電気機器、即ち他のシナリオで制御する必要がある電気機器に拡張することもできる。
【００７２】
上記の様々な保護とする装置に基づき、主に使用される音声制御システムのウェイクアップ方法は以下の通りである。
【００７３】
ウェイクアップワードの認識は、予め設定されたウェイクアップワード（ウェイクアップワードモデルからもの）（例えば「こんにちは、冷蔵庫」のようなもの）に基づいて、ユーザが音声制御を試してみるかどうかを確認する。基本的なプロセスは次のとおりである。
【００７４】
1.多数の発話者の音声により、ウェイクアップワードモデルを予め確立する。
【００７５】
2.トレーニング後のウェイクアップワードモデルを（ソリッドステートストレージスペース（フラッシュ；flash））に格納し、電源投入後にキャッシュ（記憶ユニット）にコピーする。
【００７６】
3.音声処理の際、先に取得した人間の声を含む音声情報セグメントをモデルと照合して、ウェイクアップワードであるか否かの判定を得る。
【００７７】
4.ウェイクアップワードであるか否かを確認する。コプロセッサがウェイクアップワードを検出した後、中断し、ウェイクアップ音声認識プロセッサが動作する。ウェイクアップワードが検出されない場合、ウェイクアップパスワードの入力を待ち続ける。
【００７８】
図5に示すように、以下のステップに詳細化する。
【００７９】
ステップ100：ウェイクアップワード音声モデルを構築する。
【００８０】
このステップは、初期段階の準備の時のステップであり、ウェイクアップワード音声モデルを構築した後は、その後のウェイクアップワード認識動作が容易になる。モデルの構築の場合、何人かの人間のウェイクアップ音声データを収集し、全てのウェイクアップ音

声データを処理して、トレーニングするにより、ウェイクアップワードモデルを得る。
【００８１】
図7に示すように、以下のステップに更に詳細化する。
【００８２】
オフライン状態で、様々な環境で録音された発話者からのウェイクアップワードを収集し、フレーム分割処理を行う。
【００８３】
フレーム分割の後、特徴パラメータを抽出する。
【００８４】
特徴パラメータに対してクラスタリングを行って、隠れマルコフモデルHMMの観測状態を確立する。
【００８５】
Baum-Welchアルゴリズムにより隠れマルコフモデルHMMのパラメータを調整して、P（σ|λ）（ただし、λはモデルパラメータであり、σは観察状態である）を最大化し、モデルパラメータλを調整して、観察状態σの最大確率を得て、モデルトレーニングを完成させて、ウェイクアップワード音声モデルを記憶する。
【００８６】
ステップ110：音声情報を収集する。
【００８７】
音声情報は、異なる期間に収集された複数の音声情報セグメントから構成され、全ての期間が繋がって完全かつ連続的なタイムチェーンになる。一定の期間の音声情報セグメントを単位として後の処理に送信する。一部の音声がアナログ信号として収集されることを考慮すると、後の処理には不都合であるため、アナログ-デジタル変換ステップを追加することも必要である。したがって、いくつかの実施例では、このステップを以下のように詳細化することができる。
【００８８】
ステップ1110：アナログ信号フォーマットの音声情報を収集する。
【００８９】
ステップ1120：アナログ信号フォーマットの音声情報をデジタル変換してデジタル信号フォーマットの音声情報を得る。
【００９０】
ステップ120：音声情報を処理して前記音声情報に人間の声を含むかどうかを判定し、含んでいる場合、人間の声を含む音声情報セグメントを分離して、ステップ130に移行する。
【００９１】
当該ステップは、具体的には次のとおりである。
【００９２】
ステップ1210：非ガウス値の最大である音声信号を分離するように、デジタル信号フォーマットの音声情報に対してブラインド音源分離処理を行う。
【００９３】
第1の分離ステップにおいて、ブラインド音源分離処理が採用する方法は、負のエントロピーの最大化、4次の統計量の尖度、又は時間-周波数変換に基づいた独立成分分析のICAアルゴリズムである。
【００９４】
ブラインド音源分離の役割は、信号源が未知の場合に複数の信号源を分離することである。そのうちICAは、一般的なアルゴリズムであり、負のエントロピーの最大化、4次の統計量の尖度（kurtosis）、及び時間-周波数変換方法に基づいて実現することができる。そして、固定小数点高速アルゴリズムは、DSP上でリアルタイムに実現しやすい。
【００９５】
音声信号は、ラプラス分布に従うため、スーパーガウス分布に属し、そして大部分ノイ

ズの分布はガウス特性を有する。負のエントロピー、尖度（kurtosis）などは、信号の非ガウス性に対して測定できる。その値が大きいほど、非ガウス特性が大きくなるため、信号のうちの該値の最大である信号を選択し分離して、処理する。

【００９６】

ステップ1220：エネルギー閾値により音声信号に人間の声を含んでいるかどうかを判定し、エネルギー閾値を超える場合、人間の声を含んでいると判定して、ステップ1230に移行し、エネルギー閾値を超えない場合、人間の声を含んでいないと判定して、ステップ110に移行する。

【００９７】

可能な信号を選択した後、エネルギー閾値によって発話者の声があるかどうかを判定する。音声を含むフレームは、認識モジュールに送信されてウェイクアップワードの認識プロセスを行い、その後の処理には音声のフレームドロップを含まない。

【００９８】

ステップ1230：人間の声を含む音声情報を分離し、人間の声を含む音声情報セグメントを得る。

【００９９】

ステップ130：人間の声を含む音声情報セグメントに対して、ウェイクアップワードの認識を行い、ウェイクアップワードが認識された場合、ステップ140に移行し、ウェイクアップワードが認識されない場合、ステップ110に戻る。

【０１００】

人間の声を含むデータをウェイクアップワード音声モデルと照合し、照合が成功した場合、ウェイクアップワードが認識されたと判定し、照合が成功しなかった場合、ウェイクアップワードが認識されないと判定する。

【０１０１】

具体的には、図8に示すように、人間の声を含むデータの音声フレームから特徴パラメータを抽出して、新しい観察値σ'の集合を新しい観察状態として得て、P（σ'｜λ）を計算する。

【０１０２】

P（σ'｜λ）を信頼閾値と比較して、ウェイクアップワードが認識されたかどうかを得る。

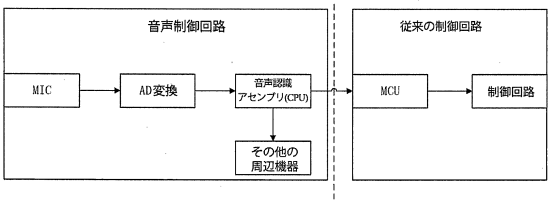
【０１０３】

ステップ140：音声認識プロセッサをウェイクアップする。

【０１０４】

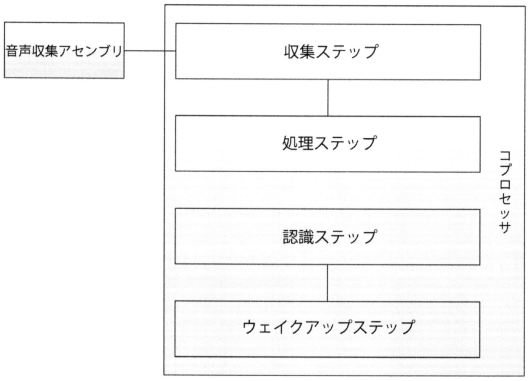
以上の説明は、本発明の好ましい実施例に過ぎず、本発明を制限するものではない。本発明の趣旨及び原則を逸脱しない範囲での修正、均等な置換、改良などは、いずれも本発明の保護の範囲に含まれるものである。

【図１】

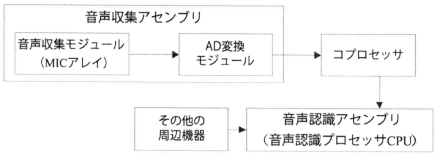


【図２】

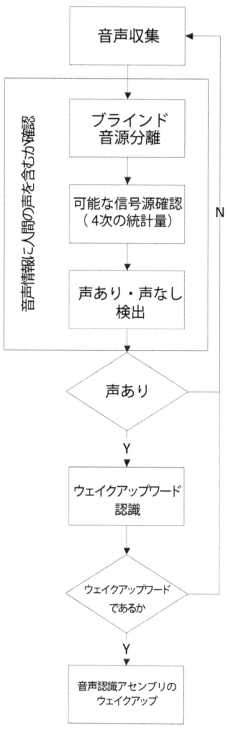
【図３】



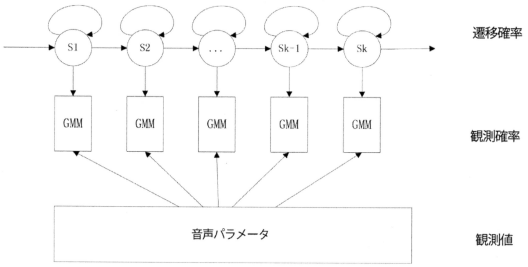
【図４】



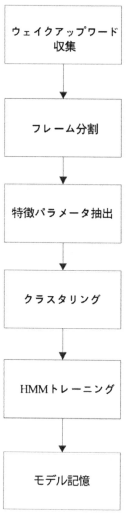
【図５】



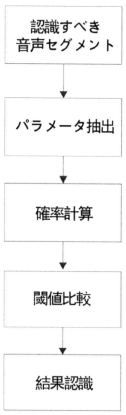
【図６】



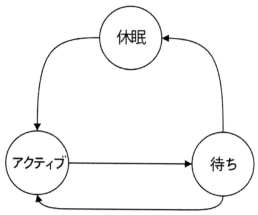
【図７】



【図８】



【図９】



フロントページの続き

(73)特許権者 517215032
合肥美的電冰箱有限公司
ＨＥＦＥＩ ＭＩＤＥＡ ＲＥＦＲＩＧＥＲＡＴＯＲ ＣＯ．，ＬＴＤ．
中国安徽省合肥市長江西路６６９号
Ｎｏ．６６９，Ｗｅｓｔ Ｃｈａｎｇｊｉａｎｇ Ｒｏａｄ，Ｈｅｆｅｉ，Ａｎｈｕｉ ２３０６０１，ＣＨＩＮＡ
(73)特許権者 512237419
美的集団股▲フン▼有限公司
ＭＩＤＥＡ ＧＲＯＵＰ ＣＯ．， ＬＴＤ．
中華人民共和国 ５２８３１１ 広東省佛山市順徳区北▲ジャオ▼鎮美的大道６号美的総部大楼Ｂ一区２６－２８楼
Ｂ２６－２８Ｆ， Ｍｉｄｅａ Ｈｅａｄｑｕａｒｔｅｒ Ｂｕｉｌｄｉｎｇ， Ｎｏ．６ Ｍｉｄｅａ Ａｖｅｎｕｅ， Ｂｅｉｊｉａｏ， Ｓｈｕｎｄｅ， Ｆｏｓｈａｎ， Ｇｕａｎｇｄｏｎｇ ５２８３１１ Ｃｈｉｎａ
(74)代理人 100146835
弁理士 佐伯 義文
(74)代理人 100129115
弁理士 三木 雅夫
(74)代理人 100203297
弁理士 橋口 明子
(72)発明者 王 岩
中華人民共和国２３０６０１安徽省合肥市合肥▲経▼▲済▼▲開▼▲発▼区▲錦▼▲綉▼大道１７６号
(72)発明者 ▲陳▼ ▲海▼雷
中華人民共和国２３０６０１安徽省合肥市合肥▲経▼▲済▼▲開▼▲発▼区▲錦▼▲綉▼大道１７６号

審査官 渡部 幸和

(56)参考文献 特表２０１６－５２６１７８（ＪＰ，Ａ）
特開平０５－２０４３９４（ＪＰ，Ａ）
特開２００５－０８４２４４（ＪＰ，Ａ）
特開平０９－０６２２９３（ＪＰ，Ａ）
特開２００７－０３６８９０（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １５／００－１５／３４