

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6285914号  
(P6285914)

(45) 発行日 平成30年2月28日 (2018. 2. 28)

(24) 登録日 平成30年2月9日 (2018. 2. 9)

(51) Int. Cl.	F I
G 1 O L 15/32 (2013. 01)	G 1 O L 15/32 2 2 O Z
G 1 O L 15/10 (2006. 01)	G 1 O L 15/10 2 O O W
G 1 O L 15/28 (2013. 01)	G 1 O L 15/28 2 3 O K
G O 6 F 3/16 (2006. 01)	G O 6 F 3/16 6 3 O

請求項の数 24 (全 19 頁)

(21) 出願番号	特願2015-509065 (P2015-509065)	(73) 特許権者	309027126
(86) (22) 出願日	平成25年4月23日 (2013. 4. 23)		ニュアンス コミュニケーションズ, イン
(65) 公表番号	特表2015-520409 (P2015-520409A)		コーポレイテッド
(43) 公表日	平成27年7月16日 (2015. 7. 16)		アメリカ合衆国 マサチューセッツ州 O
(86) 国際出願番号	PCT/US2013/037679		1 8 0 3 パーリントン ワン・ウェイサ
(87) 国際公開番号	W02013/163113		イド・ロード (番地なし)
(87) 国際公開日	平成25年10月31日 (2013. 10. 31)	(74) 代理人	100107364
審査請求日	平成28年3月24日 (2016. 3. 24)		弁理士 齊藤 達也
(31) 優先権主張番号	13/456, 959	(72) 発明者	ニューマン, マイケル ジェイ.
(32) 優先日	平成24年4月26日 (2012. 4. 26)		アメリカ合衆国 O 2 1 4 3 マサチュー
(33) 優先権主張国	米国 (US)		セッツ州 サマービル, マウンテン アベ
			ニュー 1 2

最終頁に続く

(54) 【発明の名称】 ユーザ定義可能な制約条件を有する省スペースの音声認識を構築する為の埋め込みシステム

(57) 【特許請求の範囲】

【請求項 1】

コンピュータに実装される、音声認識の管理方法であって、

電子機器の音声起動式ウェイクアップ機能において構成入力を受け取るステップであって、前記構成入力はトリガフレーズを含み、前記構成入力は、前記電子機器の第1のプロセッサにおいて受け取られ、前記電子機器は、前記第1のプロセッサに加えて第2のプロセッサを有しており、前記構成入力及び少なくとも1つのデコイの言葉が前記第2のプロセッサにてまず受け取られる、前記受け取るステップと、

前記トリガフレーズに対応する、音声認識状態のネットワークを有する有限状態トランスデューサネットワークを作成するステップであって、前記音声認識状態のネットワークは、前記第1のプロセッサにおいて、前記第1のプロセッサが実行する第1の音声認識エンジンを用いて作成され、前記トリガフレーズは、語彙モデルの適用なしで認識文法を用いて評価される、前記作成するステップと、

前記音声認識状態のネットワークを前記第1のプロセッサから前記第2のプロセッサに転送するステップと、

前記第2のプロセッサにおいて、前記トリガフレーズに対応する、前記音声認識状態のネットワークを用いて、第2の音声認識エンジンを実行するステップであって、前記第2のプロセッサは、前記第1のプロセッサの前記第1の音声認識エンジンが非アクティブ状態にある間に、前記第2の音声認識エンジンを実行し、前記第2の音声認識エンジンは、前記有限状態トランスデューサネットワークを実行するように構成された有限状態トラン

10

20

スデューサデコーダを備える、前記実行するステップと、  
を含む、コンピュータに実装される方法。

【請求項 2】

前記第 2 の音声認識エンジンを実行する前記ステップは、  
前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 のプロセッサを用いて、音声発話を連続的に分析するステップと、  
特定の音声発話が前記トリガフレーズを含むことを識別したら、前記第 1 のプロセッサに対し、前記第 1 の音声認識エンジンをアクティブ状態に戻すように伝える信号を送るステップと、  
を含む、請求項 1 に記載の、コンピュータに実装される方法。 10

【請求項 3】

前記第 1 の音声認識エンジンをアクティブ状態に戻す前記ステップは、前記第 1 のプロセッサが後続の音声コマンドに応答することを含む、請求項 2 に記載の、コンピュータに実装される方法。

【請求項 4】

前記第 1 のプロセッサの前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 の音声認識エンジンを実行する前記ステップは、前記第 1 のプロセッサが非アクティブ音声コマンドモードであるステップと、  
特定の音声発話が前記トリガフレーズを含むことを識別したら、前記電子機器が非アクティブ音声コマンドモードからアクティブ音声コマンドモードに切り替わるように仕向けるステップと、を含む、  
請求項 1 に記載の、コンピュータに実装される方法。 20

【請求項 5】

前記第 2 のプロセッサは、前記電子機器がスタンバイモードにある間に、前記第 2 の音声認識エンジンを実行して前記音声起動式ウェイクアップ機能を提供する、請求項 1 に記載の、コンピュータに実装される方法。

【請求項 6】

前記音声認識状態を前記第 2 のプロセッサに転送する前記ステップは、前記音声認識状態をデジタル信号プロセッサ (DSP) に転送するステップを含む、請求項 1 に記載の、コンピュータに実装される方法。 30

【請求項 7】

前記構成入力を受け取る前記ステップは、前記電子機器のユーザインタフェースを介して、前記トリガフレーズをテキスト入力として受け取るステップを含む、請求項 1 に記載の、コンピュータに実装される方法。

【請求項 8】

前記構成入力を受け取る前記ステップは、前記テキスト入力を追認する音声発話を受け取るステップを含む、請求項 7 に記載の、コンピュータに実装される方法。

【請求項 9】

前記第 1 のプロセッサは中央演算処理装置であり、前記第 2 のプロセッサはデジタル信号プロセッサである、請求項 1 に記載の、コンピュータに実装される方法。 40

【請求項 10】

前記第 2 のプロセッサは、前記第 1 のプロセッサがアクティブ状態にあるときに使用する電力より、低い電力を使用する、請求項 9 に記載の、コンピュータに実装される方法。

【請求項 11】

前記第 1 のプロセッサ及び前記第 2 のプロセッサは、携帯電話の内部で共存する、請求項 9 に記載の、コンピュータに実装される方法。

【請求項 12】

音声認識を管理するシステムであって、  
第 1 のプロセッサと、  
第 2 のプロセッサと、

前記第 1 及び第 2 のプロセッサと結合されたメモリと、を備え、前記メモリは、前記プロセッサによって実行された場合に、

電子機器の音声起動式ウェイクアップ機能において構成入力を受け取る動作であって、前記構成入力はトリガフレーズを含み、前記構成入力は、前記電子機器の第 1 のプロセッサにおいて受け取られ、前記電子機器は、前記第 1 のプロセッサに加えて第 2 のプロセッサを有しており、前記構成入力及び少なくとも 1 つのデコイの言葉が前記第 2 のプロセッサにてまず受け取られる、前記受け取る動作と、

前記トリガフレーズに対応する、音声認識状態のネットワークを有する有限状態トランスデューサネットワークを作成する動作であって、前記音声認識状態のネットワークは、前記第 1 のプロセッサにおいて、前記第 1 のプロセッサが実行する第 1 の音声認識エンジンを用いて作成され、前記トリガフレーズは、語彙モデルの適用なしで認識文法を用いて評価される、前記作成する動作と、

10

前記音声認識状態のネットワークを前記第 1 のプロセッサから前記第 2 のプロセッサに転送する動作と、

前記第 2 のプロセッサにおいて、前記トリガフレーズに対応する、前記音声認識状態のネットワークを用いて、第 2 の音声認識エンジンを実行する動作であって、前記第 2 のプロセッサは、前記第 1 のプロセッサの前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 の音声認識エンジンを実行し、前記第 2 の音声認識エンジンは、前記有限状態トランスデューサネットワークを実行するように構成された有限状態トランスデューサデコーダを備える、前記実行する動作と、を前記システムに行わせる命令を記憶する、

20

システム。

#### 【請求項 1 3】

前記第 2 の音声認識エンジンを実行する前記動作は、

前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 のプロセッサを用いて、音声発話を連続的に分析する動作と、

特定の音声発話が前記トリガフレーズを含むことを識別したら、前記第 1 のプロセッサに対し、前記第 1 の音声認識エンジンをアクティブ状態に戻すように伝える信号を送る動作と、を含む、

請求項 1 2 に記載のシステム。

30

#### 【請求項 1 4】

前記第 1 の音声認識エンジンをアクティブ状態に戻す前記動作は、前記第 1 のプロセッサが後続の音声コマンドに応答することを含む、請求項 1 3 に記載のシステム。

#### 【請求項 1 5】

前記第 1 のプロセッサの前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 の音声認識エンジンを実行する前記動作は、前記第 1 のプロセッサが非アクティブ音声コマンドモードである動作と、

特定の音声発話が前記トリガフレーズを含むことを識別したら、前記電子機器が非アクティブ音声コマンドモードからアクティブ音声コマンドモードに切り替わるように仕向ける動作と、を含む、

40

請求項 1 2 に記載のシステム。

#### 【請求項 1 6】

前記第 2 のプロセッサは、前記電子機器がスタンバイモードにある間に、前記第 2 の音声認識エンジンを実行して前記音声起動式ウェイクアップ機能を提供する、請求項 1 2 に記載のシステム。

#### 【請求項 1 7】

前記音声認識状態を前記第 2 のプロセッサに転送する前記動作は、前記音声認識状態をデジタル信号プロセッサ(DSP)に転送する動作を含む、請求項 1 2 に記載のシステム。

#### 【請求項 1 8】

50

前記構成入力を受け取る前記動作は、前記電子機器のユーザインタフェースを介して、前記トリガフレーズをテキスト入力として受け取る動作を含む、請求項 12 に記載のシステム。

【請求項 19】

データ情報を処理する為の命令が格納されているコンピュータ記憶媒体であって、前記命令は、処理装置によって実行された場合に、

電子機器の音声起動式ウェイクアップ機能において構成入力を受け取る動作であって、前記構成入力はトリガフレーズを含み、前記構成入力は、前記電子機器の第 1 のプロセッサにおいて受け取られ、前記電子機器は、前記第 1 のプロセッサに加えて第 2 のプロセッサを有しており、前記構成入力及び少なくとも 1 つのデコイの言葉が前記第 2 のプロセッサにてまず受け取られる、前記受け取る動作と、

10

前記トリガフレーズに対応する、音声認識状態のネットワークを有する有限状態トランスデューサネットワークを作成する動作であって、前記音声認識状態のネットワークは、前記第 1 のプロセッサにおいて、前記第 1 のプロセッサが実行する第 1 の音声認識エンジンを用いて作成され、前記トリガフレーズは、語彙モデルの適用なしで認識文法を用いて評価される、前記作成する動作と、

前記音声認識状態のネットワークを前記第 1 のプロセッサから前記第 2 のプロセッサに転送する動作と、

前記第 2 のプロセッサにおいて、前記トリガフレーズに対応する、前記音声認識状態のネットワークを用いて、第 2 の音声認識エンジンを実行する動作であって、前記第 2 のプロセッサは、前記第 1 のプロセッサの前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 の音声認識エンジンを実行し、前記第 2 の音声認識エンジンは、前記有限状態トランスデューサネットワークを実行するように構成された有限状態トランスデューサデコダを備える、前記実行する動作と、を前記処理装置に行わせる、

20

コンピュータ記憶媒体。

【請求項 20】

前記第 2 の音声認識エンジンを実行する前記動作は、

前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 のプロセッサを用いて、音声発話を連続的に分析する動作と、

特定の音声発話が前記トリガフレーズを含むことを識別したら、前記第 1 のプロセッサに対し、前記第 1 の音声認識エンジンをアクティブ状態に戻すように伝える信号を送る動作と、を含む、

30

請求項 19 に記載のコンピュータ記憶媒体。

【請求項 21】

コンピュータに実装される、音声認識の管理方法であって、

電子機器の音声起動式ウェイクアップ機能において構成入力を受け取るステップであって、前記構成入力はトリガフレーズを含み、前記電子機器は、第 1 の音声認識エンジンを実行する第 1 のプロセッサを有し、前記電子機器は、前記第 1 のプロセッサに加えて第 2 のプロセッサを有しており、前記構成入力及び少なくとも 1 つのデコイの言葉が前記第 2 のプロセッサにてまず受け取られる、前記受け取るステップと、

40

前記構成入力をリモートサーバコンピュータに送信するステップと、

前記トリガフレーズに対応する、音声認識状態のネットワークを有する有限状態トランスデューサネットワークを作成するステップであって、前記音声認識状態のネットワークは、前記リモートサーバコンピュータにおいて作成され、前記トリガフレーズは、語彙モデルの適用なしで認識文法を用いて評価される、前記作成するステップと、

前記電子機器において、前記トリガフレーズに対応する、前記音声認識状態のネットワークを受け取るステップと、

前記音声認識状態のネットワークを前記第 2 のプロセッサに転送するステップと、

前記第 2 のプロセッサにおいて、前記トリガフレーズに対応する、前記音声認識状態のネットワークを用いて、第 2 の音声認識エンジンを実行するステップであって、前記第 2

50

のプロセッサは、前記第 1 のプロセッサの前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 の音声認識エンジンを実行し、前記第 2 の音声認識エンジンは、前記有限状態トランスデューサネットワークを実行するように構成された有限状態トランスデューサデコーダを備える、前記実行するステップと、  
を含む、コンピュータに実装される方法。

【請求項 2 2】

前記第 2 の音声認識エンジンを実行する前記ステップは、

前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 のプロセッサを用いて、音声発話を連続的に分析するステップと、

特定の音声発話が前記トリガフレーズを含むことを識別したら、前記第 1 のプロセッサに対し、前記第 1 の音声認識エンジンをアクティブ状態に戻すように伝える信号を送るステップと、

を含む、請求項 2 1 に記載の、コンピュータに実装される方法。

【請求項 2 3】

前記第 1 の音声認識エンジンをアクティブ状態に戻す前記ステップは、前記第 1 のプロセッサが後続の音声コマンドに応答することを含む、請求項 2 2 に記載の、コンピュータに実装される方法。

【請求項 2 4】

前記第 1 のプロセッサの前記第 1 の音声認識エンジンが非アクティブ状態にある間に、前記第 2 の音声認識エンジンを実行する前記ステップは、前記第 1 のプロセッサが非アクティブ音声コマンドモードであるステップと、

特定の音声発話が前記トリガフレーズを含むことを識別したら、前記電子機器が非アクティブ音声コマンドモードからアクティブ音声コマンドモードに切り替わるように仕向けるステップと、を含む、

請求項 2 1 に記載の、コンピュータに実装される方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、音声認識に関し、特に、音声起動機能又は音声コマンド機能を提供する音声認識に関する。

【背景技術】

【0002】

音声認識、即ち、自動音声認識は、発話された言葉を識別するコンピュータ処理を伴う。音声認識には多くの用途があり、例えば、音声書き起こし、音声翻訳、機器及びソフトウェアアプリケーションを音声で制御する機能、呼ルーティングシステム、インターネットの音声検索などの用途がある。

【0003】

任意選択で、音声認識システムと発話言語理解システムとをペアにして、システムとの対話時に、意味及び／又は実行すべき命令を抽出することが可能である。音声認識システムは、非常に複雑であり、発話の音響シグネチャと言葉の音響シグネチャとのマッチングを行うことにより動作する。このマッチングは、任意選択で、統計的言語モデルと組み合わせられてよい。従って、音声認識処理では、音響モデリングと言語モデリングの両方が用いられる。音響モデルは、音声発話の録音と、対応する書き起こしから作成されてよい。そして、音響モデルによって、対応する言葉の個々の音の統計的表現が定義される。音声認識システムは、音響モデルを使用して音の系列を識別し、統計的言語モデルを使用して、識別された音から、ありうる言葉の系列を識別する。

【0004】

音声認識が音声起動機能又は音声コマンド機能を提供することにより、話者が様々な命令を発話することによって機器やシステムを制御することが可能になる。例えば、話者が、コマンドを発話して特定のタスクを実行したり、クエリを発話して特定の結果を取り出

10

20

30

40

50

したりすることが可能である。発話入力は、特定のタスクを実行する厳密なフレーズセットに忠実であってよく、或いは、音声認識システムの自然言語単位によって解釈される自然言語であってよい。音声コマンド機能は、可搬機器において次第に普及しつつあり、特に、携帯電話、ラップトップ、タブレットコンピュータなどのバッテリー給電式可搬機器において普及しつつある。

【発明の概要】

【発明が解決しようとする課題】

【0005】

所与の機器の音声コマンドモードは、発話コマンドを検出するべく積極的に待機してよく、或いは、ユーザが音声コマンドモードを手動で起動するまでオフになっているとよい。例えば、ユーザの発話入力に対する応答としてタスクを実行する機能を含む携帯電話があり、このタスクは、例えば、特定の個人に電話をかけること、電子メールメッセージを取り出すこと、特定の歌を再生すること等である。典型的には、ユーザが電話機上のボタンを押す（又はアイコンを選択する）ことによって音声コマンドモードを起動し、「ジョンスミスに電話」などのコマンドを発話する。応答として、電話または機器は、対応する電話番号、例えば、連絡先リスト又は他の登録簿から取り出された電話番号に対して呼び出しを開始する。

【0006】

そのような音声コマンド機能は便利であるが、それでも、ユーザは、この機能にアクセスする為には、ボタンを押すか、他の方法で選択メニューに進まなければならない。即ち、音声コマンドのモード又は機能を開始することは、ハンズフリー操作ではない。一方、ハンズフリーであることの為に、音声コマンド音声認識エンジンを連続動作させたままにすることは、従来の音声認識エンジンの電力需要が甚大であることから、望ましくない。例えば、音声コマンドモードの一環として従来の音声認識プログラムを連続動作させている典型的な携帯電話は、（バッテリーの仕様にもよるが）バッテリーの供給電力を2時間程度で使い切ってしまう。消費電力は、有線式の機器や器具の場合も問題になる。これは、従来の音声認識システムの使用が機器の定格効率に悪影響を及ぼす可能性がある為である。そこで、携帯電話などの機器の音声認識プログラムを開始させても使用電力が格段に小さく、これによって、効率的なハンズフリー操作を実現する音声起動式「ウェイクアップ」機能が必要とされている。

【課題を解決するための手段】

【0007】

従って、本明細書に開示の手法は、機器のウェイクアップを行うか、むしろ、全てのノ残っている音声コマンドのアクティブ化を機器に行わせる音声トリガであって、音声コマンド機能の手動始動（ボタンを押すこと）と同様である音声トリガを有効化するシステム及び方法を含む。更に、そのような音声トリガは、動的にプログラム可能又はカスタマイズ可能である。例えば、話者が、特定のフレーズ（1つ以上の言葉）を音声トリガとしてプログラム又は指定することが可能であり、この音声トリガフレーズを、必要に応じて変更することが可能である。大まかには、本明細書における手法は、デジタル信号プロセッサ（DSP）又は他の、電子機器の低電力のセカンダリ処理装置において動作する音声起動式ウェイクアップシステムを実行する。従来の音声認識システムは比較的大きい為、音声認識プログラムをDSP上で動作させるのは問題が多いと考えられる。ウェイクアップ機能をDSP上で動作させることについての別の問題は、動的にプログラム可能なウェイクアップ機能を与えることである。ウェイクアップフレーズが、携帯電話の組み立て時などに、事前プログラムされれば、ウェイクアップシステムをDSP上で動作させることが容易になるが、ユーザがコマンド言葉を指定できるようにすることは難題である。

【0008】

本明細書における手法は、電子機器上で2つの音声認識システムを動作させることを含む。第1の音声システムは、比較的大きく、機器の中央演算処理装置（CPU）上で動作し、第2の音声システムは、比較的小さく、DSP上で動作する。所与の可搬電子機器の

10

20

30

40

50

CPUは、典型的には、その電子機器がアクティブ状態にある間は、ほとんどの処理をこなす。DSPは、典型的には、比較的低電力で動作し、スタンバイ処理に使用されることが多い。例えば、電子機器が携帯電話として具体化される場合、DSPは、典型的には、電話が(CPUが非アクティブ状態にある)スタンバイモードにある間に、電話の呼び出しの着信を検出するべく待機する役割を担う。第1の音声認識システムは、カスタムトリガフレーズを取得し、音声モデル及びコードを、比較的小さなシステムに変換する。この第2の音声システムがDSPに転送され、DSPは、CPUが非アクティブである間、例えば、電子機器がスタンバイモードにあるときに、第2の音声システムを実行することが可能である。DSPは、この第2の音声システムを使用して、カスタムトリガフレーズを検出するべく待機する。DSPは、カスタムトリガフレーズが発話されたことを検出したら、電子機器に対して、アクティブ状態、即ち、CPUが後続の音声コマンドに応答する状態に戻るよう伝える信号を送る。電子機器は、小さい第2の音声システムをDSP上で動作させることにより、使用電力が比較的小さいハンズフリー音声トリガウェイクアップ機能を提供することが可能である。

10

#### 【0009】

一実施形態は、電子機器上で音声認識を管理する音声認識プロセス又はシステムを実行する音声認識マネージャを含む。音声認識マネージャは、電子機器の音声起動式ウェイクアップ機能、即ち、ウェイクアップ機能を構成する為のインタフェースにおいて、構成入力を受け取る。構成入力は、トリガフレーズを含む。構成入力は、電子機器の第1のプロセッサにおいて受け取られる。この電子機器は、第1のプロセッサに加えて、第2のプロセッサを有する。音声認識マネージャは、トリガフレーズに対応する、音声認識状態のネットワークを作成する。音声認識状態のネットワークは、第1のプロセッサにおいて、第1のプロセッサが実行する第1の音声認識エンジンを用いて作成される。音声認識マネージャは、音声認識状態のネットワークを、第1のプロセッサから第2のプロセッサに、例えば、CPUからDSPに、転送する。次に、音声認識マネージャは、第2のプロセッサにおいて、トリガフレーズに対応する、音声認識状態のネットワークを用いて、第2の音声認識エンジンを実行する。第2のプロセッサは、第1のプロセッサの第1の音声認識エンジンが非アクティブ状態にある間に、第2の音声認識エンジンを実行する。この第2の音声認識エンジンを実行することは、第1の音声認識エンジン、即ち、第1のプロセッサが非アクティブ状態にある間に、音声発話を連続的に分析することを含んでよい。DSP又は音声認識マネージャは、特定の音声発話がトリガフレーズを含むことを識別したら、第1のプロセッサに対し、第1の音声認識エンジンをアクティブ状態に戻して、電子機器を制御する音声コマンドモードを有効にするよう伝える信号を送る。

20

30

#### 【0010】

本明細書における更に別の実施形態は、上記で概説され、後で詳細に開示されるステップ及び動作を実施するソフトウェアプログラムを含む。そのような一実施形態は、コンピュータ記憶媒体(例えば、持続的な、有形のコンピュータ可読媒体、共通点がない状態で配置された、又は共通に配置された記憶媒体、コンピュータ記憶媒体等)を有するコンピュータプログラム製品を含み、このコンピュータ記憶媒体は、プロセッサ及び対応するメモリを有するコンピュータ化された機器において実行された場合に、本明細書に開示の動作をプロセッサが実行するように(又はプロセッサに実行させるように)プログラムする、コンピュータ記憶媒体上にエンコードされたコンピュータプログラム論理を含む。そのような構成は、典型的には、コンピュータ可読記憶媒体上に配置又はエンコードされたソフトウェア、ファームウェア、マイクロコード、コードデータ(例えば、データ構造体)等として与えられ、そのようなコンピュータ可読記憶媒体として、光媒体(例えば、CD-ROM)、フロッピーディスク、ハードディスク、1つ以上のROM又はRAM又はPRROMチップ、特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)などがある。ソフトウェア又はファームウェア又は他のそのような構成は、本明細書で説明される手法を、コンピュータ化された機器に実施させる為に、そのコンピュータ化された機器にインストールされてよい。

40

50

## 【 0 0 1 1 】

従って、本開示の一特定実施形態は、命令が格納された、1つ以上の持続的なコンピュータ記憶媒体を含むコンピュータプログラム製品を対象としており、それらの命令は、例えば、電子機器の音声起動式ウェイクアップ機能において構成入力を受け取る動作であって、構成入力はトリガフレーズを含み、構成入力は、電子機器の第1のプロセッサにおいて受け取られ、電子機器は、第1のプロセッサに加えて第2のプロセッサを有する、上記受け取る動作と、トリガフレーズに対応する、音声認識状態のネットワークを作成する動作であって、音声認識状態のネットワークは、第1のプロセッサにおいて、第1のプロセッサが実行する第1の音声認識エンジンを用いて作成される、上記作成する動作と、音声認識状態のネットワークを第1のプロセッサから第2のプロセッサに転送する動作と、第2のプロセッサにおいて、トリガフレーズに対応する、音声認識状態のネットワークを用いて、第2の音声認識エンジンを実行する動作であって、第2のプロセッサは、第1のプロセッサの第1の音声認識エンジンが非アクティブ状態にある間に、第2の音声認識エンジンを実行する、上記実行する動作と、をサポートする。本明細書に記載の命令、及び方法は、各コンピュータ機器のプロセッサで実行された場合に、本明細書に開示の方法をプロセッサに実行させる。

10

## 【 0 0 1 2 】

本開示の他の実施形態は、上記で概説され、後で詳細に開示される、方法実施形態のステップ及び動作のいずれかを実施するソフトウェアプログラムを含む。

## 【 0 0 1 3 】

当然のことながら、本明細書に記載の、各ステップの説明の順序は、明確であることを意図されたものである。一般に、これらのステップは、任意の好適な順序で実施されてよい。

20

## 【 0 0 1 4 】

又、本明細書におけるシステム、方法、装置等のそれぞれは、ソフトウェアプログラムとして、又はソフトウェアとハードウェアの混成物として、又はハードウェアのみとして、例えば、プロセッサ内で、又はオペレーティングシステム内で、又はソフトウェアアプリケーション内で、又は、人が全て又は一部の操作を実施するような非ソフトウェアアプリケーションを介して、厳密に具体化されてよいことも理解されたい。

## 【 0 0 1 5 】

上述のように、本明細書における手法は、音声コマンド音声認識をサポートするソフトウェアアプリケーションでの使用に好適である。しかしながら、本明細書における実施形態は、そのような用途での使用に限定されないこと、並びに、本明細書に記載の手法は、他の用途にも好適であることに注意されたい。

30

## 【 0 0 1 6 】

更に、本明細書における様々な特徴、手法、構成等のそれぞれは、本開示の様々な場所において説明されているであろうが、それらの概念のそれぞれは、相互に独立に、又は相互に組み合わせて実行されてよいものとする。従って、本発明は、多様なかたちで実施されてよく、とらえられてよい。

## 【 0 0 1 7 】

なお、本明細書におけるこの要約セクションは、本開示の、又は特許請求される発明の実施形態及び/又は段階的に新規な態様を全て明示しているわけではない。むしろ、この要約は、様々な実施形態、及び従来手法に対して新規な対応箇所についての予備的な説明を与えているに過ぎない。本明細書及び実施形態の更なる詳細及び/又は可能な展望については、後で詳述される、本開示の「発明を実施するための形態」のセクション及び対応図面を読まれたい。

40

## 【 0 0 1 8 】

本発明の上述及び他の目的、特徴、及び利点は、添付図面において示されるように、以下の、本明細書における好ましい実施形態のより具体的な説明から明らかになるであろう。添付図面では、類似の参照符号は、様々な図面の全体を通して同じ要素を参照する。図

50



面は、必ずしも正確な縮尺では描かれておらず、むしろ、実施形態、原理、及び概念を説明することに重点が置かれている。

【図面の簡単な説明】

【0019】

【図1】本明細書における実施形態による音声トリガウェイクアップ機能をサポートする機器のシステムのブロック図である。

【図2】本明細書における実施形態による音声トリガウェイクアップをサポートするプロセスの一例を示すフローチャートである。

【図3】乃至

【図4】本明細書における実施形態による音声トリガウェイクアップをサポートするプロセスの一例を示すフローチャートである。

【図5】本明細書における実施形態による、コンピュータ/ネットワーク環境で動作する音声認識マネージャのブロック図の一例である。

【発明を実施するための形態】

【0020】

本明細書に開示の手法は、機器のウェイクアップを行うか、全ての/残っている音声コマンドのアクティブ化を機器に行わせる音声トリガであって、音声コマンド機能の手動始動（ボタンを押すこと）と同様である音声トリガを有効化するシステム及び方法を含む。更に、そのような音声トリガは、動的にプログラム可能又はカスタマイズ可能である。例えば、話者が、特定のフレーズ（1つ以上の言葉）を音声トリガとしてプログラム又は指定することが可能であり、この音声トリガフレーズは、ユーザの必要に応じた修正や変更が可能である。大まかには、本明細書における手法は、中央演算処理装置（CPU）において動作するのではなく、デジタル信号プロセッサ（DSP）又は他の、電子機器の低電力のセカンダリ処理装置において動作する音声起動式ウェイクアップシステムを実行する。

【0021】

本明細書に開示の手法は、2つのプロセッサを使用する多様な電子機器に埋め込まれてよいが、実施形態の説明の都合上、本開示は、主に、バッテリー給電式携帯電話のような携帯電話を主に参照する。従来の携帯電話は、典型的には、プロセッサが2つある。まず、メインCPU、あるいはプライマリCPUがあり、これは比較的強力である。このCPUは、電話の呼び出し、ユーティリティアプリケーション、電子メール、ゲームなど、携帯電話が積極的に使用されている場合に、ほとんどの処理をこなす。携帯電話は又、典型的には、セカンダリプロセッサを含み、これはデジタル信号プロセッサ（DSP）である。DSPは、非常に小さい電力で機能することが可能である。DSPは又、（自身に対して相対的に）大きな電力モードで動作することも可能である。携帯電話の機能のうち、非常に長い時間にわたって動作するものは、典型的には、DSP上で動作する。例えば、DSPは、携帯電話がスタンバイモードである間、或いは、積極的に使用されていない間、アクティブであって電話の呼び出しの着信を検出すべく待機する、携帯電話のコンポーネントである。

【0022】

本明細書における手法は、電子機器上で2つの音声認識システムを動作させることを含む。第1の音声システムは、比較的大きく、機器の中央演算処理装置（CPU）上で動作し、第2の音声システムは、比較的小さく、DSP上で動作する。CPU音声システムは、特定のモバイルアプリケーション向けの音声認識モデルのフルセットを含んでよく、例えば、それらは、連絡先、電子メール、歌へのアクセス、タスクの実施、アプリケーションの実行などを行う為のコマンドを認識する為のものであってよい。この、フルセットのモデルは、発話ウェイクアップトリガとして使用されるカスタムのフレーズ又は言葉をユーザが選択できることを可能にする。

【0023】

第1の/プライマリシステムは、カスタムトリガフレーズを受け取り、このトリガフレ

10

20

30

40

50

ーズを使用して、音声モデル及びコードを、比較的小さなシステムに変換する。従って、CPUは、より小さな音声システムを事前計算する。このシステムは、2つの部分を有してよく、それらは、(1)認識状態のネットワーク、及び(2)音響モデル情報である。認識状態のネットワークは、状態の系列を通る最良のパスを検索するデコーダによって使用されてよい。音響モデル情報は、音声の記録からの測定値を含んでよい。実施形態は、DSPウェイクアップ機能に使用される、音響モデル情報の非常に小さなサブセットとともに非常に良好に機能することが可能である。これら2つの部分は、小さなシステム、即ち、CPU音声認識システムと比べて小さなシステムにまとめられる。非限定的な例として、認識状態のネットワーク(CPUによってコンパイルされたデータ)は、サイズが約5キロバイトから10キロバイトであってよく、比較として、プライマリシステムからのネットワークは、約300キロバイトから約2メガバイトであってよい。従って、この手法は、大幅なサイズ削減を実現する。

#### 【0024】

カスタム入力(カスタムトリガフレーズ)は、テキスト入力として指定されてよい。代替実施形態では、入力は、発話されてよく、且つ/又は、テキストとして入力されてよい。可聴入力は、テキスト入力と照合されることにより、認識精度の向上に役立ちうる。しかしながら、トリガフレーズに対する可聴入力は不要であることに注意されたい。システムは、最初にテキストとしてのみ入力されたトリガフレーズに基づいて、ウェイクアップ機能を実行してよい。その後、事前計算されたシステムが、DSPに転送されてよい。手法は、比較的小さな音声デコーダをDSP上で動作させることを含んでよい。このDSPデコーダは、例えば、コードサイズが30キロバイト前後であってよい。これに対し、CPU上で動作する音声デコーダは、サイズが約1~2メガバイトであってよい。例えば、DSP認識器コードは、わずか数百行のコードがDSP上で動作してよい。従来の認識器では、典型的には、たくさんのモジュールがあり、そのそれぞれが比較的大きい。これらの大きなモジュールは全て、DSP上で動作する単一モジュールに置き換えられてよい。この単一モジュールは、他のどのモジュールよりも小さくてよい。言い換えると、汎用認識エンジンを動作させる代わりに、非常に特殊な認識エンジンを、DSP上で使用する為に作成してよい。

#### 【0025】

そして、DSP上の認識器は、ウェイクアップトリガとして機能する。例えば、認識器は、本質的には、トリガフレーズが発話されたかどうかを判定する。DSPが、受け取られた音声発話からウェイクアップフレーズを認識していない場合は、何も起こらない。ウェイクアップフレーズが発話されたら、DSP認識器が判定すると、CPU認識器はアクティブになり、発話されたコマンド又はクエリに対する応答を開始又は続行することが可能である。

#### 【0026】

発話されたフレーズ又は言葉によりウェイクアップすることが可能な機器が幾つか存在するが、これらのトリガ言葉は固定である。即ち、当該電子機器では所定のトリガコマンドが固定されており、変更できない。この固定されたトリガ言葉は、典型的には、製造工程において設定される。言い換えると、トリガ言葉は、その機器の寿命にわたって不変である。これに対して、本明細書に開示の手法は、完全なカスタマイズが可能であるか、ユーザ固有であってよい、構成可能/変更可能なトリガ言葉を提供する。本明細書において提供されるそのようなカスタマイズは、オンラインの追加コードをダウンロードすること、又は、トリガコマンドを再構成する為に機器を工場に戻すことを必要としない。

#### 【0027】

実施形態では、トレーニングモード又はカスタマイズインタフェースは、音声発話を必要とせず、テキスト入力を用いて、DSP上で動作するウェイクアップシステムを作成することが可能である。実施形態によっては、テキスト入力の追認の為に音声入力を受け取ることにも可能であるが、これは必須ではない。ウェイクアップシステムをテキスト入力のみから作成することは有用であり、これは、定義済み音響モデルをかなり小さいモデルま

10

20

30

40

50

で縮小しても、所定のトリガ言葉に対して同等の認識性能／精度が得られる為である。従って、実施形態によっては、DSPシステムに送信される音響モデルデータは、既存の話者依存音響モデルから抽出することが可能である。

#### 【0028】

特徴の1つとして、DSP音響モデルは、特定のユーザの声に適合されなくてもよい。むしろ、システムは、より大きなモデルから、特定のトリガフレーズに必要な状態を選択することが可能である。ある特定のトリガフレーズに対して、対応する数の状態IDが必要である。非限定的な例として、特定のトリガフレーズが状態ID2、7、48、138、455、...を必要とする場合、音声認識マネージャ又はウェイクアップマネージャは、これらの状態IDを引き出し、より小さなモデルでの使用の為にそれらの番号を付け替えることになる。この結果として、より小さく、より効率的なモデルが得られる。従って、特定の言葉の系列に対し、プライマリシステムは、この言葉の系列を、音素及び他の従来の音声表現の系列に変換する。特定のフレーズに対して、保持されるべきモデル状態の数はごくわずかである。従って、ほとんどのモデルデータは、特定のウェイクアップトリガ言葉／フレーズに使用される必要がない。このことは又、音声認識に使用されるコードのほとんどが、DSPなどのセカンダリ処理装置上で動作する、より小さなシステムでは使用される必要がないことも意味している。

#### 【0029】

ここで図1を参照すると、ユーザ105が電子機器137を操作している。電子機器137は、カスタマイズ可能なウェイクアップ機能を含む音声コマンド機能を含む。ユーザ105は、特定のトリガフレーズをテキスト入力することが可能である。例えば、ユーザ105は、「ジーニアスポタン」、「ジョンの電話をアクティブにする(activate John's phone)」、「引きこもる(sequester)」、「パープルモンキー食器洗い機(purple monkey dishwasher)」など、任意のフレーズをタイプ入力することが可能である。なお、本明細書においては、「フレーズ」又は「トリガフレーズ」という用語は、1つの言葉又は複数の言葉を指すものであってよい。ユーザ105が(初期設定用の)トリガフレーズを入力すると、中央演算処理装置110が、デジタル信号プロセッサ120によって使用される音声認識システム又は音声認識状態を作成する。CPU110は、ウェイクアップ機能の為にデータを生成しながら、音響モデル112及び音声認識状態114にアクセスすることが可能である。CPU110は、音声認識状態114を生成することに加えて、音響モデル122として使用される音響モデルデータを音響モデル112(不特定話者音響モデル)から抽出又は収集することが可能である。そして、DSP120は、このデータを受け取り、動的音声認識状態125として記憶することが可能である。そして、DSP120は、音声認識中に、音声認識状態125及び音響モデル122の両方にアクセスすることが可能である。最初の、又は新しいトリガフレーズ及び随伴データが作成された後、DSP120は、発話を監視して、トリガ言葉が発話されたかどうかを識別する認識処理を実行することが可能である。このウェイクアップ機能が動作可能であるのは、CPU110が非アクティブである間、又はスタンバイ中、又は少なくとも、CPU110上で実行される音声コマンド機能が非アクティブである間である。CPU110は、トリガフレーズが検出されるまで非アクティブのままであってよい。トリガフレーズが発話されたことをDSP120が検出したら、DSPは、CPU120に対し、アクティブ状態、又は、音声入力に対する応答として音声エンジンを実行しタスクを実行することが可能な状態に戻るように伝える信号を送る。この特定のトリガ言葉及び対応するデータネットワークは、ユーザが新しいトリガフレーズの作成を必要とするまで、DSP上に残る。システムは、新しいトリガフレーズを受け取ったら、DSP120向けにネットワーク状態を作成する処理を繰り返すことが可能である。

#### 【0030】

ここで、より具体的には、CPU110において新しいトリガフレーズが受け取られると、この文字列は、音素系列に変換されてよい。この系列は、既に存在していてよく、或

10

20

30

40

50

いは、作成される必要があつてよい。マネージャは、各音素に対して隣接する音素（各側に1つ）を識別してトライフォンを作成する。そして、各トライフォンを状態の系列に変換する。各トライフォンは、音響状態の系列のモデルを有する。一般に、ある特定のトライフォンモデルは、2個又は2～3個の状態、即ち、開始状態と終了状態、或いは、開始状態と中間状態と終了状態を有する。結果は、音響モデル状態のセットであり、これは、スコアリング用音響モデルにおいてルックアップされるものである。従つて、トライフォンが音響モデル又は音響モデル状態にマッピングされて、系列が作成される。

#### 【0031】

音声認識エンジンによる音声発話のスコアリングは、典型的には、相対スコアリング処理である。音声認識マネージャは、認識文法を使用してよい。この認識文法は、トリガフレーズを評価するとともに、一連のデコイ（おとり）の言葉又はフレーズを通る他のパスを評価して、音声認識マネージャがあまり認識しない（誤認識しない）ようにすることが可能である。認識文法は、全てのデコイの言葉及び音響モデル状態を通るパスを包含する。この構成により、マネージャは、残りの音響モデルも、いかなる語彙モデルも必要としない。この認識文法部分は、DSPによって使用される比較的コンパクトなデータにコンパイルされるものである。

#### 【0032】

音声認識状態は、確率分布、ガウス系列としてモデル化することが可能である。音声発話が検出されると、この音声発話はフレームに変換され、このフレームが確率分布と比較されて、スコアが取得される。デコイは、ランダムな言葉のセットとして選択されてよく、トリガフレーズと似ているか、全く異なつてよい。そして、音声モデルは、音声発話を評価し、デコイも1つ以上評価して、基準比較スコアを設定する。音声発話のスコアがランダム/デコイ言葉のスコアより（所定量だけ）高い場合、マネージャは、音声発話が認識されていると判定する。絶対スコアを探索するモデルの使用が任意選択で使用されてよいが、この手法は、典型的には、精度がより低い。相対スコアを使用すれば、バックグラウンドのノイズ及び音声を考慮しながら、発話された言葉を高精度で認識することが可能になる。

#### 【0033】

本明細書におけるDSP認識エンジンの利点は、（トリガフレーズ認識中の）DSP認識エンジンが、仮説を立てること、言葉を音素系列に変換すること、言葉をトライフォンに変換すること、並びに、そのトライフォンを状態の系列に変換することを行わなくてよいことである。これらの処理ステップが不要なのは、DSP認識器が、CPUによって作成された状態の系列に対して機能することが可能である為である。そのようなデコーダは、有限状態トランスデューサ（FST）と呼ばれることがある。従つて、FSTネットワークは、プライマリCPU上でコンパイルされ、DSP上で動作すべくFSTデコーダに下ろされるものであり、これは、カスタマイズ可能なトリガフレーズ入力に対する応答として動的に実行され、これによって、特定のフレーズに合わせてカスタマイズされたシステムが与えられ、このシステムは、CPU上で動作する初期システムよりかなり小さい。DSPデコーダについては、実施形態によっては、コードは同じであつてよく、CPUがコンパイルするネットワークは異なる。最初に、音響モデル122を構築する為のデータが収集されてよく、1つの音響モデルが構築されてから、音声認識マネージャが、必要なだけの異なる小さなモデル（ネットワーク状態125）を作成してよい。

#### 【0034】

音声分析中に、DSPは、音声発話を受け取り、これをネットワークに通して処理することにより、スコアを取得する。DSPデコーダは又、「犬（dog）」や「カタパルト（catapult）」などのランダム/デコイ言葉も処理する。DSPデコーダが、ネットワークを通る、音声発話のパスを識別できない場合、その仮説は棄却される。音声発話とデコイ言葉の両方がネットワークを通った場合、（トリガ言葉である）音声発話のスコアはデコイ言葉のスコアよりずっと高いはずであり、このスコア差によって、トリガ言葉が発話されたことがシステムに対して示されることが可能であり、CPUがウェイクア

10

20

30

40

50

ップするか再度アクティブになる。デコイ言葉は、いついかなる言葉が発話されても動作可能である。このウェイクアップモードでは、DSPは、聞こえる全ての言葉を分析していることが可能である。デコイ言葉の数が少なければ（例えば、デコイ言葉の数が100前後程度であれば）、処理動作の高速化に役立つ。代替として、汎用音声モデルを使用する代わりにデコイフレーズを廃棄してもよく、これは、トリガ言葉の検出において、ほとんどの精度で機能するであろう。これらのフレーズを除去することにより、メモリ消費を抑えることができるが、精度も低下する可能性がある。なお、DSP（又はセカンダリ処理装置）のプログラミングは、所与の電子機器の具体的なハードウェア及び構成の態様に依存する可能性がある。例えば、携帯電話において動作する音声認識マネージャは、タブレットコンピュータ、デスクトップコンピュータ、リモコン、家電器具、自動車等において動作する実施例とは構成が異なる場合がある。

10

#### 【0035】

図5は、本明細書における実施形態によるコンピュータ/ネットワーク環境で動作する音声認識マネージャ140の一例のブロック図を示す。図5のコンピュータシステムハードウェア態様については、フローチャートの説明に従って詳述する。

#### 【0036】

図2乃至図4のフローチャート及び図を用いて、音声認識マネージャ140に関連付けられた機能を説明する。以下の説明の都合上、音声認識マネージャ140又は他のしかるべきエンティティは、それらのフローチャートにおけるステップを実施するものとする。

#### 【0037】

20

以下、実施形態をより具体的に説明する。図2は、本明細書に開示の実施形態を示すフローチャートである。ステップ210で、音声認識マネージャは、電子機器の音声起動式ウェイクアップ機能において、構成入力を受け取る。構成入力は、トリガフレーズを含む。言い換えると、ユーザがウェイクアップカスタマイズメニューにアクセスして、特定のフレーズ（言葉又は言葉の群）を設定し、そのフレーズをタイプ入力するか、別の方法でカスタムフレーズを選択する。構成入力は、電子機器の第1のプロセッサにおいて受け取られる（送られる）。電子機器は、第1のプロセッサに加えて、第2のプロセッサを有する。

#### 【0038】

ステップ220で、音声認識マネージャは、トリガフレーズに対応する、即ち、トリガフレーズに基づく、音声認識状態のネットワークを作成する。音声認識状態のネットワークは、第1のプロセッサにおいて、第1のプロセッサが実行する第1の音声認識エンジンを用いて作成される。

30

#### 【0039】

ステップ230で、音声認識マネージャは、音声認識状態のネットワークを第1のプロセッサから第2のプロセッサに転送する。即ち、音声認識状態のネットワークは、第2のプロセッサにある記憶装置に転送されるか、第2のプロセッサからアクセス可能になる。

#### 【0040】

ステップ240で、音声認識マネージャは、第2のプロセッサにおいて、トリガフレーズに対応する、音声認識状態のネットワークを用いて、第2の音声認識エンジンを実行する。第2のプロセッサは、第1のプロセッサの第1の音声認識エンジンが非アクティブ状態にある間に、第2の音声認識エンジンを実行する。なお、第1のプロセッサ又はCPUは、アクティブ状態であってよいが、それでも、第1の音声認識エンジンは、音声コマンドタスクに対しては、比較的非アクティブであるか、応答性が鈍い。従って、電子機器がスタンバイ状態であろうが、（例えば、コンテンツの閲覧、メッセージのチェック等に）アクティブに使用されていようが、電子機器の音声コマンドモードは、実行すべきタスクに備えて積極的には待機していない。他の実施形態では、CPUは、完全に非アクティブでなくてよく、ユーザが電子機器と対話している場合のように積極的に使用されているときの消費電力に比べれば低い電力モードで動作してよい。

40

#### 【0041】

50

図 3 及び図 4 は、本明細書に開示の音声認識マネージャ 140 の追加及び / 又は代替の実施形態、並びに任意選択の機能を示すフローチャートを含んでいる。ステップ 210 で、音声認識マネージャは、電子機器の音声起動式ウェイクアップ機能において、構成入力を受け取る。構成入力は、トリガフレーズを含む。構成入力は、電子機器の第 1 のプロセッサにおいて受け取られる（第 1 のプロセッサに送られる）。電子機器は、第 1 のプロセッサに加えて、第 2 のプロセッサを有する。

【0042】

ステップ 212 では、音声認識マネージャは、電子機器のユーザインタフェースを介して、トリガフレーズをテキスト入力として受け取る。例えば、ユーザが、電子機器のウェイクアップの為にユーザが発話する必要があるフレーズをタイプする。

10

【0043】

ステップ 213 で、音声認識マネージャは、テキスト入力を追認する音声発話を受け取る。テキスト入力で十分であるが、音声認識マネージャは、正確な認識を徹底する為に、テキスト入力の音声発話を処理することも可能である。

【0044】

ステップ 215 で、第 1 のプロセッサは中央演算処理装置であり、第 2 のプロセッサはデジタル信号プロセッサである。ステップ 216 で、第 2 のプロセッサは、第 1 のプロセッサがアクティブ状態にあるときに使用する電力より、低い電力を使用する。電子機器は、ウェイクアップ機能を、電力が低い方のプロセッサで動作させることにより、バッテリーの消耗が早すぎることなく、トリガ言葉を検出すべく待機することが可能である。ステップ 217 で、第 1 のプロセッサ及び第 2 のプロセッサは、携帯電話の内部で共存する。

20

【0045】

ステップ 220 で、音声認識マネージャは、トリガフレーズに対応する、音声認識状態のネットワークを作成する。音声認識状態のネットワークは、第 1 のプロセッサにおいて、第 1 のプロセッサが実行する第 1 の音声認識エンジンを用いて作成される。

【0046】

ステップ 230 で、音声認識マネージャは、音声認識状態のネットワークを第 1 のプロセッサから第 2 のプロセッサに転送する。ステップ 232 で、音声認識マネージャは、音声認識状態をデジタル信号プロセッサに転送する。

【0047】

30

ステップ 240 で、音声認識マネージャは、第 2 のプロセッサにおいて、トリガフレーズに対応する、音声認識状態のネットワークを用いて、第 2 の音声認識エンジンを実行する。第 2 のプロセッサは、第 1 のプロセッサの第 1 の音声認識エンジンが非アクティブ状態にある間に、第 2 の音声認識エンジンを実行する。

【0048】

ステップ 242 で、第 2 の音声認識エンジンは、第 1 の音声認識エンジンが非アクティブ状態にある間に、第 2 のプロセッサを用いて音声発話を連続的に分析する。その後、音声認識マネージャは、特定の音声発話トリガフレーズを含むことを識別したら、第 1 のプロセッサに対し、第 1 の音声認識エンジンをアクティブ状態に戻すように伝える信号を送る。

40

【0049】

ステップ 243 で、第 1 のプロセッサは、後続の音声コマンドに応答する。

【0050】

ステップ 246 で、第 1 のプロセッサは、非アクティブ音声コマンドモードにあり、特定の音声発話トリガフレーズを含むことを識別したら、電子機器が非アクティブ音声コマンドモードからアクティブ音声コマンドモードに切り替わるように仕向ける。

【0051】

ステップ 248 で、電子機器がスタンバイモードにある間に、第 2 のプロセッサは、第 2 の音声認識エンジンを実行して音声起動式ウェイクアップ機能を提供する。

【0052】

50

他の実施形態では、新しい状態系列及びより小さな音響モデルを、電子機器において作成する代わりに、リモートサーバにおいて作成してよい。そのような実施形態では、電子機器は、新しいトリガフレーズをサーバ又はクラウドに送信してよい。新しいトリガフレーズは、電子機器を介して、テキスト入力として入力されてよい。そして、リモートサーバは、トリガフレーズに対応する、音声認識状態のネットワークを作成し、次に、作成された状態系列及び音響モデルを電子機器に送信する。これらは、セカンダリプロセッサ又はDSPによって使用可能である。

#### 【0053】

次に図6を参照しながら、上述の音声認識マネージャ140に関連付けられた機能が実施される様子を示す基本的な実施形態を説明する。しかしながら、音声認識マネージャ140を実施する為の実際の構成は、個々の用途に応じて様々であってよいことに注意されたい。例えば、コンピュータシステム149は、本明細書に記載の処理を実施する1つ以上のコンピュータを含んでよい。

#### 【0054】

様々な実施形態では、コンピュータシステム149は、様々なタイプの機器のいずれであってもよく、そのような機器には、携帯電話、パーソナルコンピュータシステム、デスクトップコンピュータ、ラップトップコンピュータ、ノートブックコンピュータ、ネットブックコンピュータ、メインフレームコンピュータシステム、ハンドヘルドコンピュータ、ワークステーション、ネットワークコンピュータ、ルータ、ネットワーク交換機、ブリッジ、アプリケーションサーバ、記憶装置や、カメラ、カムコーダ、セットトップボックス、モバイル機器、ビデオゲームコンソール、ハンドヘルドビデオゲーム機器などの家庭用電子機器、或いは、一般的に任意のタイプのコンピューティング機器又は電子機器が含まれ、これらに限定されない。

#### 【0055】

図示されたコンピュータシステム149は、ユーザ136が入力装置135を使用して操作するグラフィカルユーザインタフェース133を表示する為のディスプレイモニタ130に接続されている。処理前と処理後の両方のデータファイル及びコンテンツを格納する為に、任意選択で、リポジトリ138が使用されてよい。入力装置135は、キーボード、コンピュータマウス、マイクなどの1つ以上の装置を含んでよい。

#### 【0056】

図示されるように、本実施例のコンピュータシステム149は、メモリシステム141、プロセッサ142、I/Oインタフェース144、及び通信インタフェース145を結合する相互接続143を含む。

#### 【0057】

I/Oインタフェース144は、コンピュータマウス、キーボード、カーソルを動かす選択ツール、ディスプレイ画面等を含む入力装置135などの周辺機器との接続性を提供する。

#### 【0058】

通信インタフェース145は、コンピュータシステム149の音声認識マネージャ140が、本明細書における実施形態に従って、ネットワーク経由の通信を行い、必要に応じて、ビューの作成、コンテンツの処理、ユーザとの通信等に必要データを取出すことを可能にする。

#### 【0059】

図示されるように、メモリシステム141は、上述され、後で更に詳しく説明される機能をサポートする音声認識マネージャ140-1によってエンコードされる。音声認識マネージャ140-1（及び/又は本明細書に記載の他のリソース）は、本明細書に記載の様々な実施形態による処理機能をサポートするデータ及び/又は論理命令などのソフトウェアコードとして埋め込まれてよい。

#### 【0060】

一実施形態の動作中に、プロセッサ142は、始動、起動、実行、解釈、その他の方法

10

20

30

40

50

で音声認識マネージャ１４０－１の論理命令を実行する為に、相互接続１４３を使用してメモリシステム１４１にアクセスする。音声認識マネージャ１４０－１の実行によって、音声認識マネージャプロセス１４０－２に処理機能が生成される。言い換えると、音声認識マネージャプロセス１４０－２は、音声認識マネージャ１４０のうちの、コンピュータシステム１４９のプロセッサ１４２内で、又はプロセッサ１４２上で実施される１つ以上の部分を表す。

【００６１】

なお、本明細書に記載のように方法動作を実施する音声認識マネージャプロセス１４０－２に加えて、本明細書における他の実施形態は、音声認識マネージャ１４０－１そのものの（即ち、実行されない論理命令及び／又はデータ）を含む。音声認識マネージャ１４０－１は、フロッピーディスク、ハードディスク、光媒体などのコンピュータ可読記憶媒体を含む、持続的な、有形のコンピュータ可読記憶媒体に格納されてよい。他の実施形態によれば、音声認識マネージャ１４０－１は、ファームウェアのようなメモリ型のシステムとして、又は読み出し専用メモリ（ＲＯＭ）として、又は、この実施例のように、実行可能コードとして、メモリシステム１４１に格納されてもよい。

10

【００６２】

又、これらの実施形態に加えて、本明細書における他の実施形態は、プロセッサ１４２内の音声認識マネージャ１４０－１を音声認識マネージャプロセス１４０－２として実行することを含むことにも注意されたい。従って、当業者であれば理解されるように、コンピュータシステム１４９は、他のプロセス、及び／又はソフトウェアコンポーネント及びハードウェアコンポーネントを含んでよく、例えば、ハードウェアリソース、即ち、複数のプロセッサの割当及び使用を制御するオペレーティングシステムを含んでよい。

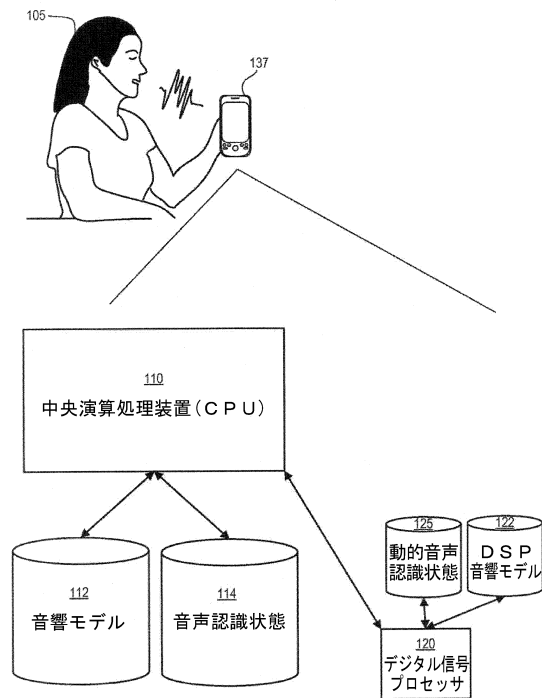
20

【００６３】

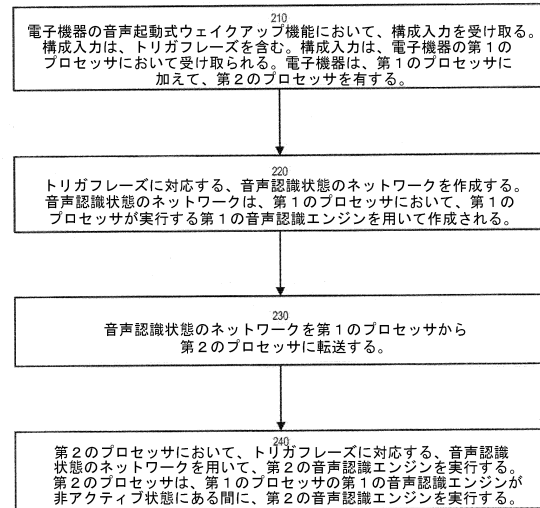
又、当業者であれば理解されるように、本発明の同じ目的を達成する限りにおいて、上述の手法の動作に対して多様な変形が行われてよい。そのような変形は、本発明の範囲によって包含されるものとする。従って、本発明の実施形態の上述の説明は、限定的ではないものとする。むしろ、本発明の実施形態に対するあらゆる限定は、後述の特許請求の範囲において示されている。



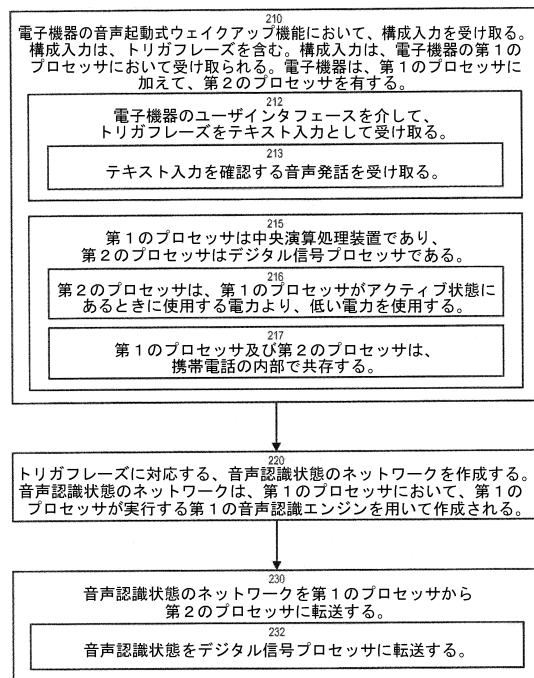
【図 1】



【図 2】



【図 3】



【図 4】

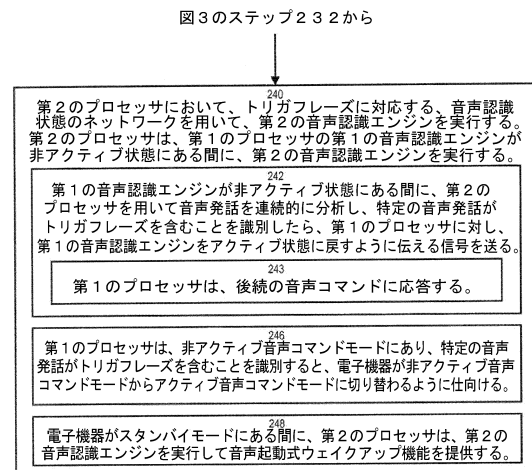
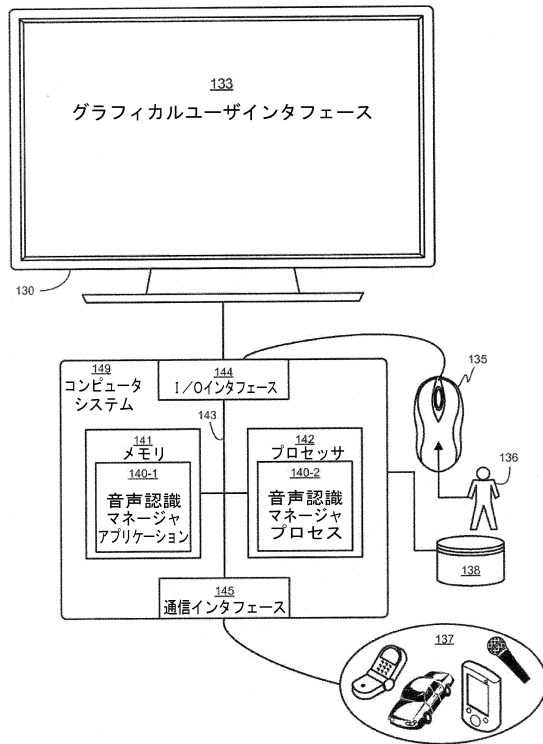


図 4 のステップ 240 へ

【図 5】



---

フロントページの続き

- (72)発明者 ロス, ロバート  
アメリカ合衆国 02460 マサチューセッツ州 ニュートン, エルム ロード 73, アパ  
ートメント 9
- (72)発明者 アレクサンダー, ウィリアム ディー.  
アメリカ合衆国 02421 マサチューセッツ州 レキシントン, レキシントン リッジ ドラ  
イブ 2121
- (72)発明者 マルブレット, ポール ヴァン  
アメリカ合衆国 01778 マサチューセッツ州 ウェイランド, ボウ ロード 30

審査官 上田 雄

- (56)参考文献 米国特許第06070140(US, A)  
特開2002-132285(JP, A)  
特開2004-294755(JP, A)  
特開2003-177790(JP, A)  
特開平09-062293(JP, A)  
特表2002-534716(JP, A)  
米国特許出願公開第2007/0028761(US, A1)  
米国特許出願公開第2010/0082343(US, A1)

- (58)調査した分野(Int.Cl., DB名)  
G10L 15/00 - 15/34  
G06F 3/16