

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7051799号

(P7051799)

(45)発行日 令和4年4月11日(2022.4.11)

(24)登録日 令和4年4月1日(2022.4.1)

(51)国際特許分類

F I

G 1 0 L 15/28 (2013.01)

G 1 0 L 15/28 2 3 0 J

G 1 0 L 15/10 (2006.01)

G 1 0 L 15/10 2 0 0 W

G 1 0 L 15/28 2 3 0 K

請求項の数 14 (全22頁)

(21)出願番号	特願2019-237718(P2019-237718)	(73)特許権者	512015127
(22)出願日	令和1年12月27日(2019.12.27)		バイドゥ オンライン ネットワーク テクノロジー(ペキン)カンパニー リミテッド
(65)公開番号	特開2021-56483(P2021-56483A)		中華人民共和国 ペキン ハイディエン ディストリクト シャンディー テンス
(43)公開日	令和3年4月8日(2021.4.8)		ストリート ナンバー 10 バイドゥ キャンパス 3エフ
審査請求日	令和1年12月27日(2019.12.27)		
(31)優先権主張番号	201910933124.8	(73)特許権者	521235408
(32)優先日	令和1年9月29日(2019.9.29)		シャンハイ シャオドゥ テクノロジーカンパニー リミテッド
(33)優先権主張国・地域又は機関	中国(CN)		中華人民共和国 201210 シャンハイ チャイナ (シャンハイ) パイロット
前置審査			フリー トレード ゾーン ナキシアン ロード ナンバー701 ビルディング 1
			最終頁に続く

(54)【発明の名称】 音声認識制御方法、装置、電子デバイス及び読み取り可能な記憶媒体

(57)【特許請求の範囲】

【請求項1】

ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、前記ターゲットシーンが第1動作状態に適用されるか否かを決定するステップであって、第1動作状態は、リスニング状態であり、電子デバイスが該リスニング状態にあるとき、ユーザはウェイクワードを音声で入力せずに、音声命令をリアルタイムに入力して、電子デバイスとインタラクションするステップと、

前記ターゲットシーンが前記第1動作状態に適用される場合、実行中の第2動作状態を前記第1動作状態に切り替えるステップであって、前記第2動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識し、第2動作状態は、非リスニング状態であるステップと、

前記第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識するステップと、

を含み、

前記オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識するステップの後に、

前記オーディオストリームを音声認識することにより取得された情報ストリームを取得するステップと、

前記情報ストリームから各候補意図を取得するステップと、

各候補意図から前記ターゲットシーンの制御意図とマッチングする第2制御意図をフィルタリングするステップと、
前記第2制御意図がフィルタリングして取得された場合、前記第2制御意図に対応する制御命令を実行するステップと、
を含む音声認識制御方法。

【請求項2】

前記実行中の第2動作状態を前記第1動作状態に切り替えるステップの前に、
前記第2動作状態で、前記オーディオクリップを音声認識することにより取得された第1制御意図を取得するステップと、
前記第1制御意図が前記ターゲットシーンとマッチングすると決定するステップと、
を含む請求項1に記載の音声認識制御方法。

10

【請求項3】

前記各候補意図から前記ターゲットシーンの制御意図とマッチングする第2制御意図をフィルタリングするステップの後に、
設定期間内に、前記第2制御意図が取得されなかった場合、前記第1動作状態を終了するステップを含み、
前記設定期間の範囲は、20秒～40秒である請求項1に記載の音声認識制御方法。

【請求項4】

前記情報ストリームから各候補意図を取得するステップの後に、
前記ターゲットシーンの制御意図とマッチングしない候補意図への応答を拒否するステップを含む請求項1に記載の音声認識制御方法。

20

【請求項5】

前記構成情報は、前記第1動作状態に適用されるシーンリストを含み、前記シーンリストはユーザの選択操作に応答して、音楽シーン、オーディオブックシーン及びビデオシーンから選択して生成されるものである請求項1から4のいずれかに記載の音声認識制御方法。

【請求項6】

ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、前記ターゲットシーンが第1動作状態に適用されるか否かを決定するための照会モジュールであって、第1動作状態は、リスニング状態であり、電子デバイスが該リスニング状態にあるとき、ユーザはウェイクワードを音声で入力せずに、音声命令をリアルタイムに入力して、電子デバイスとインタラクションする照会モジュールと、

30

ターゲットシーンが前記第1動作状態に適用される場合、実行中の第2動作状態を前記第1動作状態に切り替えるための切り替えモジュールであって、前記第2動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識し、第2動作状態は、非リスニング状態である切り替えモジュールと、

前記第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識する認識モジュールと、

前記オーディオストリームを音声認識することにより取得された情報ストリームを取得し、前記情報ストリームから各候補意図を取得するための第2取得モジュールと、
各候補意図から前記ターゲットシーンの制御意図とマッチングする第2制御意図をフィルタリングするためのフィルタリングモジュールと、

40

前記第2制御意図がフィルタリングして取得された場合、前記第2制御意図に対応する制御命令を実行するための実行モジュールと、
を備える音声認識制御装置。

【請求項7】

前記第2動作状態で、前記オーディオクリップを音声認識することにより取得された第1制御意図を取得するための第1取得モジュールと、

前記第1制御意図が前記ターゲットシーンとマッチングすると決定するための決定モジュールと、を含む請求項6に記載の音声認識制御装置。

50

【請求項 8】

設定期間内に、前記第 2 制御意図が取得されなかった場合、前記第 1 動作状態を終了するための終了モジュールを備え、
前記設定期間の範囲は、20 秒～40 秒である請求項 6 に記載の音声認識制御装置。

【請求項 9】

前記ターゲットシーンの制御意図とマッチングしない候補意図への応答を拒否するための拒否モジュールを含む請求項 6 に記載の音声認識制御装置。

【請求項 10】

前記構成情報は、前記第 1 動作状態に適用されるシーンリストを含み、前記シーンリストはユーザの選択操作に応答して、音楽シーン、オーディオブックシーン及びビデオシーンから選択して生成されるものである請求項 6 から 9 のいずれかに記載の音声認識制御装置。

【請求項 11】

ターゲットシーンで、第 1 動作状態の構成情報を照会し、前記ターゲットシーンが第 1 動作状態に適用されるか否かを決定するステップであって、第 1 動作状態は、リスニング状態であり、電子デバイスが該リスニング状態にあるとき、ユーザはウェイクワードを音声で入力せずに、音声命令をリアルタイムに入力して、電子デバイスとインタラクションするステップと、

前記ターゲットシーンが前記第 1 動作状態に適用される場合、実行中の第 2 動作状態を前記第 1 動作状態に切り替えるステップであって、前記第 2 動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識し、第 2 動作状態は、非リスニング状態であるステップと、

前記第 1 動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識するステップと、
を含み、

前記オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識するステップの後に、

前記オーディオストリームを音声認識することにより取得された情報ストリームを取得するステップと、

前記情報ストリームから各候補意図を取得するステップと、

各候補意図から前記ターゲットシーンの制御意図とマッチングする第 2 制御意図をフィルタリングするステップと、

前記第 2 制御意図がフィルタリングして取得された場合、前記第 2 制御意図に対応する制御命令を実行するステップと、

を含む音声認識制御方法。

【請求項 12】

ターゲットシーンで、第 1 動作状態の構成情報を照会し、前記ターゲットシーンが第 1 動作状態に適用されるか否かを決定するための照会モジュールであって、第 1 動作状態は、リスニング状態であり、電子デバイスが該リスニング状態にあるとき、ユーザはウェイクワードを音声で入力せずに、音声命令をリアルタイムに入力して、電子デバイスとインタラクションする照会モジュールと、

前記ターゲットシーンが前記第 1 動作状態に適用される場合、実行中の第 2 動作状態を前記第 1 動作状態に切り替えるための切り替えモジュールであって、前記第 2 動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識し、第 2 動作状態は、非リスニング状態である切り替えモジュールと、

前記第 1 動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識する認識モジュールと、

前記オーディオストリームを音声認識することにより取得された情報ストリームを取得し、前記情報ストリームから各候補意図を取得するための第 2 取得モジュールと、

各候補意図から前記ターゲットシーンの制御意図とマッチングする第 2 制御意図をフィルタリングするためのフィルタリングモジュールと、

10

20

30

40

50

前記第 2 制御意図がフィルタリングして取得された場合、前記第 2 制御意図に対応する制御命令を実行するための実行モジュールと、
を備える音声認識制御装置。

【請求項 1 3】

少なくとも 1 つのプロセッサと、
前記少なくとも 1 つのプロセッサに通信可能に接続されるメモリと、を備え、
前記メモリには前記少なくとも 1 つのプロセッサによって実行可能な命令が記憶され、前記命令が前記少なくとも 1 つのプロセッサによって実行されることにより、前記少なくとも 1 つのプロセッサが請求項 1 ~ 6 のいずれかに記載の音声認識制御方法、又は、請求項 1 1 に記載の音声認識制御方法を実行可能である電子デバイス。

10

【請求項 1 4】

コンピュータ命令が記憶されている非一時的なコンピュータ読み取り可能な記憶媒体であって、
前記コンピュータ命令は、前記コンピュータに請求項 1 から 5 のいずれかに記載の音声認識制御方法、又は、請求項 1 1 に記載の音声認識制御方法を実行させる非一時的なコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本願は、音声認識及び人工知能の技術分野に関し、特に音声認識制御方法、装置、電子デバイス及び読み取り可能な記憶媒体に関する。

20

【背景技術】

【0002】

人工知能技術及び端末技術の継続的な発展に伴って、人工知能製品、たとえばスマートスピーカ等の電子デバイスが普及しつつあり、ユーザは電子デバイスが対応する制御命令を実行するように音声で制御することができる。現在、ユーザが電子デバイスと連続的に音声インタラクションをする必要があり、すなわち、ユーザに連続的に会話を開始させるニーズがある場合、電子デバイスが設定された動作モードに入るように音声で制御し、全二重継続会話 (Full-duplex Continued Conversation) の機能を体験することができる。

30

【0003】

しかしながら、ユーザに連続的に会話を開始させるニーズがない場合、たとえば、ユーザがその他のユーザと会話する場合には、対応する制御命令を意図せずにトリガーするシーンにおいて、電子デバイスが設定された動作モードに入ると、ユーザが迷惑を被ることがあり、ユーザ体験を著しく悪化させる。

【発明の概要】

【発明が解決しようとする課題】

【0004】

本願は、関連技術の技術課題の 1 つを少なくともある程度解決することを目的とする。

【課題を解決するための手段】

40

【0005】

本願は、音声認識制御方法、装置、電子デバイス及び読み取り可能な記憶媒体を提供し、さまざまなユーザのパーソライズニーズを満たすように、ユーザが自分のニーズに応じて、第 1 動作状態をトリガーするシーンを設定することを達成し、ターゲットシーンが第 1 動作状態に適用される場合のみ、装置が第 1 動作状態に入るように制御し、ユーザに対する迷惑を減らすことで、ユーザ体験を改善することができる。

【0006】

本願の第 1 態様の実施例は、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第 1 動作状態の構成情報を照会し、前記ターゲットシーンが第 1 動作状態に適用されるか否かを決定するステップと、前記タ

50

ターゲットシーンが前記第1動作状態に適用される場合、実行中の第2動作状態を前記第1動作状態に切り替えるステップであって、前記第2動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識するステップと、前記第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、前記オーディオストリームを音声認識するステップと、を含む音声認識制御方法を提供する。

【0007】

本願の第2態様の実施例は、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、前記ターゲットシーンが第1動作状態に適用されるか否かを決定するための照会モジュールと、ターゲットシーンが前記第1動作状態に適用される場合、実行中の第2動作状態を前記第1動作状態に切り替えるための切り替えモジュールであって、前記第2動作状態

10

で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識する切り替えモジュールと、前記第1動作状態

【0008】

本願の第3態様の実施例は、ターゲットシーンで、第1動作状態の構成情報を照会し、前記ターゲットシーンが第1動作状態に適用されるか否かを決定するステップと、ターゲットシーンが前記第1動作状態に適用される場合、実行中の第2動作状態を前記第1動作状態に切り替えるステップであって、前記第2動作状態

20

で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識するステップと、前記第1動作状態

【0009】

本願の第4態様の実施例は、ターゲットシーンで、第1動作状態の構成情報を照会し、前記ターゲットシーンが第1動作状態に適用されるか否かを決定するための照会モジュールと、ターゲットシーンが前記第1動作状態に適用される場合、実行中の第2動作状態を前記第1動作状態に切り替えるための切り替えモジュールであって、前記第2動作状態

30

で、ウェイクワードに基づいてオーディオクリップを収集し、前記オーディオクリップを音声認識する切り替えモジュールと、前記第1動作状態

【0010】

本願の第5態様の実施例は、少なくとも1つのプロセッサと、前記少なくとも1つのプロセッサに通信可能に接続されるメモリと、を備え、前記メモリには、前記少なくとも1つのプロセッサによって実行可能な命令が記憶され、前記命令が前記少なくとも1つのプロセッサによって実行されることにより、前記少なくとも1つのプロセッサに本願の第1態様の実施例に係る音声認識制御方法を実行させ、又は、本願の第3態様の実施例に係る音声認識制御方法を実行させる電子デバイスを提供する。

40

【0011】

本願の第6態様の実施例は、コンピュータ命令の非一時的なコンピュータ読み取り可能な記憶媒体を提供し、前記コンピュータ命令は、前記コンピュータに本願の第1態様の実施例に係る音声認識制御方法、又は、本願の第3態様の実施例に係る音声認識制御方法を実行させる。

【発明の効果】

【0012】

上記出願の1つの実施例は、以下の利点又は有益な効果を有する。

ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作

50

状態に適用されるか否かを決定し、ターゲットシーンが第1動作状態に適用される場合、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する実行中の第2動作状態を第1動作状態に切り替え、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザが自分のニーズに応じて、第1動作状態をトリガーするシーンを設定することができ、また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御し、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

【0013】

以下、具体的な実施例を参照しながら、上記好ましい態様が有するその他の効果について説明する。

10

【図面の簡単な説明】

【0014】

図面は、本実施案を理解するためのものにすぎず、本願を限定するものではなく、本願の上記及び/又は付加的な態様及び利点は、図面を組み合わせた実施例についての以下の説明から明瞭かつ理解しやすくなる。

【図1】本願の実施例1に係る音声認識制御方法の概略フローチャートである。

【図2】本願の実施例の設定インターフェイス模式図である。

【図3】本願の実施例2に係る音声認識制御方法の概略フローチャートである。

【図4】本願の実施例3に係る音声認識制御方法の概略フローチャートである。

20

【図5】本願の実施例4に係る音声認識制御方法の概略フローチャートである。

【図6】本願の実施例5に係る音声認識制御装置の構造概略図である。

【図7】本願の実施例6に係る音声認識制御装置の構造概略図である。

【図8】本願の実施例7に係る音声認識制御装置の構造概略図である。

【図9】本願の実施例8に係る電子デバイスの構造概略図である。

【発明を実施するための形態】

【0015】

以下、図面を参照しながら本願の例示的な実施例を説明する。理解しやすくするために、本願の実施例の様々な詳細を含むが、それらは例示的なものにすぎない。従って、当業者であれば、本願の範囲及び精神から逸脱することなく、ここで説明された実施例に対して様々な変更や改定を行うことができる。同様に、明瞭化かつ簡素化するために、以下の説明では、周知機能及び構造についての説明が省略される。

30

【0016】

以下、図面を参照しながら、本願の実施例の音声認識制御方法、装置、電子デバイス及び読み取り可能な記憶媒体を説明する。

【0017】

図1は、本願の実施例1に係る音声認識制御方法の概略フローチャートである。

【0018】

本願の実施例は、該音声認識制御方法が音声認識制御装置に配置されることを例として説明し、該音声認識制御装置は、任意の電子デバイスに適用されることにより該電子デバイスは音声認識制御機能を実行できる。

40

【0019】

ここで、電子デバイスは、パーソナルコンピュータ(Personal Computer、PCと略称する)、クラウドデバイス、モバイルデバイス、スマートスピーカなどであってもよい。モバイルデバイスは、たとえば、携帯電話、タブレットコンピュータ、携帯情報端末、ウェアラブルデバイス、カーデバイスなどの、さまざまなオペレーティングシステム、タッチスクリーンおよび/またはディスプレイを搭載したハードウェアデバイスであってもよい。

【0020】

図1に示すように、該音声認識制御方法は、ステップ101~ステップ103を含むこと

50

ができる。

【0021】

ステップ101において、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されるか否かを決定する。

【0022】

本願の実施例では、ターゲットシーンはユーザが呼び出しているアプリケーションプログラムインターフェイス(Application Programmers Interface、APIと略称される)に関連するシーンであり、たとえば、ターゲットシーンは、天気シーン、音楽シーン、ビデオシーンなどのシーンであってもよい。

10

【0023】

本願の実施例では、第1動作状態は、リスニング状態であってもよく、電子デバイスが該リスニング状態にあるとき、ユーザはウェイクワードを音声で入力せずに、音声命令をリアルタイムに入力して、電子デバイスとインタラクションすることができる。第1動作状態の構成情報はユーザが自分のニーズに応じて設定されるものであり、該構成情報は第1動作状態に適用されるシーンリストを含み、該シーンリストはユーザの選択操作にตอบสนองして、異なるシーンから選択して生成されるものであり、たとえば、該シーンリストは、ユーザの選択操作にตอบสนองして、音楽シーン、オーディオブックシーン、ビデオシーン、天気シーン、ショッピングシーンなどのシーンから選択して生成されるものであってもよい。

【0024】

本願の実施例では、あるアプリケーションプログラムインターフェイスを呼び出す場合、音声認識制御装置は、該アプリケーションプログラムインターフェイスに対応するターゲットシーンを決定できる。好ましくは、ユーザがあるアプリケーションプログラムを開き、該アプリケーションプログラムにおけるあるアプリケーションプログラムインターフェイスを呼び出す場合、音声認識制御装置は該アプリケーションプログラムインターフェイスに対応するターゲットシーンを決定でき、たとえば、ユーザが天気類アプリケーションプログラムを呼び出す場合、呼び出したアプリケーションプログラムインターフェイスに対応するターゲットシーンが天気シーンであり、又は、ユーザがインスタントメッセージング類アプリケーションプログラムにおけるあるアプリケーションプログラムインターフェイスを呼び出して、天気を照会する場合、呼び出したアプリケーションプログラムインターフェイスに対応するターゲットシーンが天気シーンである。また、たとえば、音楽類アプリケーションプログラムを呼び出す場合、呼び出したアプリケーションプログラムインターフェイスに対応するターゲットシーンが音楽シーンであり、又は、ユーザがインスタントメッセージング類アプリケーションプログラムにおけるあるアプリケーションプログラムインターフェイスを呼び出して、音楽を再生する場合、呼び出したアプリケーションプログラムインターフェイスに対応するターゲットシーンが音楽シーンであり、ここでは列挙されていない。

20

30

【0025】

本願の実施例では、音声認識制御装置はモニタリング方式によって、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、該ターゲットシーンが第1動作状態に適用されるか否かを決定することができる。具体的には、構成情報が第1動作状態のシーンリストに該ターゲットシーンが含まれることを示す場合、該ターゲットシーンが第1動作状態に適用されると決定でき、構成情報が第1動作状態のシーンリストに該ターゲットシーンが含まれないことを示す場合、該ターゲットシーンが第1動作状態に適用されないと決定する。

40

【0026】

一例としては、図2を参照すれば、図2は、本願の実施例の設定インターフェイス概略図である。ユーザは自分のニーズに応じて、音楽シーン、オーディオブックシーン及びビデオシーンから対応するシーンを選択し、第1動作状態のシーンリストを生成することができ、ユーザが音楽シーン及びビデオシーンを選択すると仮定する場合、第1動作状態のシ

50

ーンリストが音楽シーン及びビデオシーンを含み、ターゲットシーンが音楽シーンであると仮定する場合、第1動作状態の構成情報を照会することにより、ターゲットシーンが第1動作状態に適用されると決定できる。

【0027】

なお、図2は、シーンリストが音楽シーン、オーディオブックシーン及びビデオシーンから選択して生成される場合のみを例とし、つまり、図2は、シーンリストが音楽シーン、オーディオブックシーン及びビデオシーンの少なくとも1つを含む場合を例とするが、実際に適用するとき、シーンリストは、たとえば天気シーン、ショッピングシーンなど、その他のシーンをさらに含んでもよく、本願はこれについては制限しない。

【0028】

ステップ102において、ターゲットシーンが第1動作状態に適用される場合、実行中の第2動作状態を第1動作状態に切り替え、第2動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する。

【0029】

本願の実施例では、第2動作状態は、非リスニング状態であってもよく、電子デバイスが該非リスニング状態にあるとき、ユーザはウェイクワードを音声で入力し、該電子デバイスをウェイクアップして後続の音声認識を実行させることができる。ここでは、ウェイクワードは、電子デバイスの内蔵プログラムにより予め設定されるものであってもよく、又は、ユーザのパーソナライズニーズを満たすために、ウェイクワードは、ユーザが自分のニーズに応じて設定されるものであってもよく、本願はこれについて限定しない。例示すると、電子デバイスがスマートスピーカである場合、ウェイクワードは、「小度小度(ドちゃんドちゃん)」であってもよい。

【0030】

本願の実施例では、電子デバイスが第2動作状態にあるとき、音声認識制御装置は、ユーザがウェイクワードを入力したか否かを検出して入力した場合、ユーザがウェイクワードの後に入力したオーディオクリップを収集し、音声認識を行う。例示すると、電子デバイスがスマートスピーカである場合には、該スマートスピーカが該第2動作状態にあるときに、歌を再生したい場合、ユーザは、「小度小度(ドちゃんドちゃん)、歌Aを再生してください」又は「小度小度(ドちゃんドちゃん)、歌を聞きたい」と音声で入力し、電子デバイスは、ウェイクワードの後のオーディオクリップ「歌Aを再生してください」又は「歌を聞きたい」を認識し、対応する歌を再生する。

【0031】

本願の実施例では、音声認識制御装置が、ターゲットシーンが第1動作状態に適用されると決定した場合、実行中の第2動作状態を第1動作状態に切り替えることができ、音声認識制御装置が、ターゲットシーンが第1動作状態に適用されないと決定した場合、第2動作状態を実行し続けることができる。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザが自分のニーズに応じて、第1動作状態をトリガーするシーンを設定することができる。また、呼び出したターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御し、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

【0032】

なお、「第1」、「第2」の用語は、単に目的を説明するためのものであり、相対重要性を指示又は暗示するか、或いは示された技術的特徴の数を黙示的に指示すると理解してはいけない。そこで、「第1」、「第2」が限定されている特徴は、一つ又はより多くの該特徴を含むことを明示又は暗示するものである。

【0033】

ステップ103において、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。

【0034】

本願の実施例では、電子デバイスが第1動作状態にあるとき、電子デバイスは、オーディ

10

20

30

40

50

を連続的に収集してオーディオストリームを取得して、音声認識を行うことができる。これにより、該第1動作状態で、ユーザがウェイクワードを入力せずに、電子デバイスとリアルタイムにインタラクションする又は連続的にインタラクションすることができ、ユーザの操作を簡素化し、ユーザ体験を改善することができる。

【0035】

従来技術では、ユーザは連続的に会話するニーズを有していないとき、対応する制御命令を誤操作でトリガーして、電子デバイスが設定された動作モードに入るように制御する可能性があるが、このような場合には、ユーザは迷惑を被ることがあり、ユーザ体験を著しく悪化させる。

【0036】

例えば、電子デバイスが設定された動作モードに入るように制御する条件は、ユーザが入力した音声データの意図が設定された制御意図とマッチングすることであると仮定する。設定された制御意図が「オーディオまたはビデオを再生する」場合であることを例として、ユーザAがユーザBと会話するシナリオでは、ユーザAが「このスマートスピーカをウェイクアップする方法を知っている？」と聞き、ユーザBが「もちろん。小度小度（ドちゃんドちゃん）じゃない？それで、音楽を聴いたり、映画を観たりできるんだよ。」と言った場合、スマートスピーカはユーザBの音声データを認識し、ユーザがウェイクワード「小度小度（ドちゃんドちゃん）」を入力したことを検出し、該ユーザBが入力した音声データの意図が「音楽を聴いたり、映画を観たり」することであると認識する。これは、明らかに、設定された制御意図とマッチングしているので、この場合、スマートスピーカは、音楽又は映画を再生する。もちろん、この場合、ユーザが音楽及び映画を再生したくない場合には、ユーザにとって迷惑になる。

【0037】

本願の実施例の音声認識制御方法は、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されるか否かを決定し、ターゲットシーンが第1動作状態に適用される場合、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する実行中の第2動作状態を第1動作状態に切り替え、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。これにより、ユーザは自分のニーズに応じて、さまざまなユーザのパーソナライズニーズを満たすように、第1動作状態をトリガーするシーンを設定することができる。また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御することで、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

【0038】

1つの可能な実施態様としては、ターゲットシーンが第1動作状態に適用される場合、ユーザがウェイクワードの後に入力したオーディオクリップの意図がターゲットシーンとマッチングするか否かを判断する必要があり、マッチングした場合、実行中の第2動作状態を第1動作状態に切り替え、マッチングしない場合、第2動作状態を実行し続ける。これにより、ユーザがターゲットシーンとマッチングする制御意図を入力するのみ、電子デバイスの動作状態を切り替え、ユーザに対する迷惑を回避する。以下、実施例2と組み合わせ、上記過程について詳細に説明する。

【0039】

図3は、本願の実施例2に係る音声認識制御方法の概略フローチャートである。

【0040】

図3に示すように、該音声認識制御方法は、ステップ201～ステップ207を含むことができる。

【0041】

ステップ201において、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会する。

10

20

30

40

50

【 0 0 4 2 】

ステップ 2 0 2 において、構成情報に基づいて、ターゲットシーンが第 1 動作状態に適用されるか否かを判断し、適用される場合、ステップ 2 0 3 を実行し、適用されない場合、ステップ 2 0 7 を実行する。

【 0 0 4 3 】

ステップ 2 0 1 ~ 2 0 2 の実行過程は、上記実施例におけるステップ 1 0 1 の実行過程を参照できるが、ここで詳しい説明は省略される。

【 0 0 4 4 】

ステップ 2 0 3 において、第 2 動作状態で、オーディオクリップを音声認識することにより取得された第 1 制御意図を取得する。

10

【 0 0 4 5 】

ここでは、第 2 動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する。

【 0 0 4 6 】

本願の実施例では、第 2 動作状態において、音声認識制御装置は、ユーザがウェイクワードを入力したか否かを検出し、入力した場合には、ユーザがウェイクワードの後に入力したオーディオクリップを収集して、該オーディオクリップに対して音声認識を行い、該オーディオクリップに対応する第 1 制御意図を得る。たとえば、音声認識技術に基づいて、該オーディオクリップを認識し、該オーディオクリップに対応するテキスト情報を決定した後、該テキスト情報に対してセマンティックス認識を行い、対応する第 1 制御意図を決定する。例えば、オーディオクリップが「歌を聞きたい」である場合、該オーディオクリップの意図は「歌を聞く」又は「オーディオを再生する」ことである。

20

【 0 0 4 7 】

ステップ 2 0 4 において、第 1 制御意図がターゲットシーンとマッチングするか否かを判断し、マッチングする場合には、ステップ 2 0 5 ~ 2 0 6 を実行し、マッチングしない場合には、ステップ 2 0 7 を実行する。

【 0 0 4 8 】

本願の実施例では、第 1 制御意図及びターゲットシーンの制御意図に対してセマンティックスマッチングを行い、第 1 制御意図がターゲットシーンとマッチングするか否かを決定する。

30

【 0 0 4 9 】

ステップ 2 0 5 において、実行中の第 2 動作状態を第 1 動作状態に切り替える。

【 0 0 5 0 】

本願の実施例では、ターゲットシーンが第 1 動作状態に適用され、ユーザが入力したオーディオクリップの意図がターゲットシーンとマッチングする場合のみ、実行中の第 2 動作状態を第 1 動作状態に切り替える。これにより、ユーザが入力したオーディオクリップの意図がターゲットシーンとマッチングする場合にのみ、電子デバイスの動作状態を切り替えるので、ユーザに対する迷惑を更に回避することができる。

【 0 0 5 1 】

一例としては、電子デバイスがスマートスピーカであり、ターゲットシーンがビデオシーンであることを例とする。ユーザが「小度小度（ドちゃんドちゃん）、歌を聞きたい」と音声で入力した場合、スマートスピーカは、ユーザが入力したオーディオクリップを収集して認識し、オーディオクリップ「歌を聞きたい」に対応する第 1 制御意図が「オーディオを再生する」とであると決定する。しかし、ターゲットシーンに対応する制御意図が「ビデオを再生する」とあるときには、該第 1 制御意図がターゲットシーンとマッチングしないと決定するので、スマートスピーカが第 2 動作状態を実行し続けるように制御することができる。ユーザが「小度小度（ドちゃんドちゃん）、映画 A を再生してください」と音声で入力した場合、スマートスピーカはユーザが入力したオーディオクリップを収集して認識し、オーディオクリップ「映画 A を再生してください」に対応する第 1 制御意図が「ビデオを再生する」とであると決定する。ターゲットシーンに対応する制御意図も「ビデオ

40

50

を再生する」であるときに、該第1制御意図がターゲットシーンとマッチングすると決定するので、スマートスピーカが実行中の第2動作状態を第1動作状態に切り替えるように制御することで、該第1動作状態における、ユーザとスマートスピーカとの連続的な会話、たとえば、「カンフー映画を再生してください」、「再生を終了してください」などを行うことができ、これによりマン - マシン会話が自然かつリアルになり、ユーザの没入感を向上させることができる。

【0052】

ステップ206において、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。

【0053】

ステップ207において、第2動作状態を実行し続ける。

【0054】

本願の実施例では、ターゲットシーンが第1動作状態に適用されないか、又は、ターゲットシーンが第1動作状態に適用され、且つユーザが入力したオーディオクリップの意図がターゲットシーンとマッチングしない場合、第2動作状態を実行し続ける。これにより、ユーザに対する迷惑を更に回避することができる。

【0055】

例えば、第1動作状態の構成情報が、第1動作状態のシーンリストにビデオシーンが含まれることを示す場合、ユーザが入力したオーディオクリップの意図がビデオシーンとマッチングしても、ターゲットシーンが第1動作状態に適用されない場合、電子デバイスが第1動作状態に入るように制御しない。たとえば、ユーザが天気類APPを開いたときに呼び出したアプリケーションプログラムインターフェイスに対応するターゲットシーンは天気シーンであり、明らかに、該天気シーンはビデオシーンとマッチングしていない。これは、ユーザに連続的に会話を開始させるニーズがないことを示しているので、電子デバイスが第1動作状態に入るように制御せず、ユーザに対する迷惑を減らすことができる。

【0056】

1つの可能な実施態様として、ユーザに対する迷惑を更に回避するために、第1動作状態で、ユーザが電子デバイスとリアルタイムにインタラクションする際に、ユーザが入力したオーディオデータの意図がターゲットシーンとマッチングする場合のみ、オーディオデータの意図に対応する制御命令を実行する。以下、実施例3と組み合わせて、上記過程について詳細に説明する。

【0057】

図4は、本願の実施例3に係る音声認識制御方法の概略フローチャートである。

【0058】

図4を参照して、図1に示す実施例に基づいて、ステップ103の後、該音声認識制御方法は、さらにステップ301～ステップ304を含むことができる。

【0059】

ステップ301において、オーディオストリームを音声認識することにより取得された情報ストリームを取得する。

【0060】

本願の実施例では、ユーザがオーディオデータを入力した後、音声認識制御装置は、オーディオストリームを収集して、該オーディオストリームを音声認識し、対応する情報ストリームを決定することができる。

【0061】

ステップ302において、情報ストリームから各候補意図を取得する。

【0062】

本願の実施例では、情報ストリームを取得した後、該情報ストリームに対してセマンティクス認識を行い、対応する各候補意図を決定することができる。

【0063】

ステップ303において、各候補意図から、ターゲットシーンの制御意図とマッチングす

10

20

30

40

50

る第2制御意図をフィルタリングする。

【0064】

ステップ304において、第2制御意図がフィルタリングして取得された場合、第2制御意図に対応する制御命令を実行する。

【0065】

本願の実施例では、各候補意図及びターゲットシーンの制御意図に対してセマンティクスマッチングを行うことができ、ある候補意図がターゲットシーンの制御意図とマッチングする場合、該候補意図を第2制御意図として、第2制御意図に対応する制御命令を実行する。

【0066】

例えば、ターゲットシーンがビデオシーンであり、ターゲットシーンに対応する制御意図が「ビデオを再生する」であり、情報ストリームが「映画Aを見たい」であると仮定すると、該情報ストリームが1つの候補意図「ビデオを再生する」のみを有すると決定できる。このとき、該候補意図に対応する制御命令を実行するので、電子デバイスが映画Aを再生するように制御することができる。

【0067】

本願の実施例の音声認識制御方法は、第1動作状態で、ユーザが電子デバイスとリアルタイムにインタラクションするとき、ユーザが入力したオーディオデータの意図がターゲットシーンとマッチングする場合のみ、オーディオデータの意図に対応する制御命令を実行し、このため、ユーザに対する迷惑を更に回避できる。

【0068】

1つの可能な実施態様としては、設定期間内に、第2制御意図が取得されなかった場合、ユーザに電子デバイスを制御する意図がないことを示す。このとき、電子デバイスのエネルギー消費量を低減させ、電子デバイスが常に第1動作状態にあることを回避するために、第1動作状態を終了することができる。

【0069】

ここでは、設定期間は予め設定されたものであり、たとえば、電子デバイスの内蔵プログラムにより予め設定されたものであってもよく、又は、ユーザにより設定されたものであってもよく、これについては限定しない。たとえば、設定期間の値の範囲は20秒～40秒であってもよい。

【0070】

1つの可能な実施態様としては、前記情報ストリームから各候補意図を取得した後、ターゲットシーンの制御意図とマッチングしない候補意図が存在する場合、該ユーザに対する迷惑を減らすために、候補意図への応答を拒否することができる。

【0071】

例えば、ビデオシーンでは、ユーザが入力したオーディオデータが「ねえ、この映画の主題歌の歌Aはかなり良いから後で歌Aを聞くよ」である場合、電子デバイスは該オーディオデータを認識し、該オーディオデータに対応する意図が「歌を再生する」ことであると決定する。これは、明らかに、ビデオシーンに対応する制御意図「ビデオを再生する」とマッチングしないので、ユーザに対する迷惑を回避するために、該オーディオデータに対応する意図への応答を拒否することができる。

【0072】

つまり、電子デバイスが第1動作状態にあるとき、ユーザが入力したオーディオデータの意図がターゲットシーンの制御意図とマッチングする場合のみ、ユーザに対する迷惑を回避するために、オーディオデータに対応する意図に応答する。

【0073】

上記実施例を達成するために、本願は、さらに、音声認識制御方法を提供する。

【0074】

図5は、本願の実施例4に係る音声認識制御方法の概略フローチャートである。

【0075】

10

20

30

40

50

図5に示すように、該音声認識制御方法は、ステップ401、ステップ402、及びステップ403を含むことができる。

ステップ401において、ターゲットシーンで、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されるか否かを決定する。

【0076】

本願の実施例では、ターゲットシーンは、電子デバイスに現在表示されているインターフェイスに対応するシーンであってもよく、たとえば、天気シーン、音楽シーン、ビデオシーンなどのシーンであってもよい。

【0077】

本願の実施例では、第1動作状態は、リスニング状態であってもよく、電子デバイスが該リスニング状態にあるとき、ユーザはウェイクワードを音声で入力せずに、音声命令をリアルタイムに入力して、電子デバイスとインタラクショナルすることができる。第1動作状態の構成情報はユーザが自分のニーズに応じて設定されるものであり、該構成情報は第1動作状態に適用されるシーンリストを含み、該シーンリストはユーザの選択操作にตอบสนองして、異なるシーンから選択して生成されるものであり、たとえば、該シーンリストは、ユーザの選択操作にตอบสนองして、音楽シーン、オーディオブックシーン、ビデオシーン、天気シーン、ショッピングシーンなどのシーンから選択して生成されるものであってもよい。

【0078】

本願の実施例では、ユーザがあるインターフェイスを開くと、音声認識制御装置は該インターフェイスに対応するターゲットシーンを決定できる。たとえば、ユーザがビデオ類のアプリケーションを開くと、ターゲットシーンがビデオシーンであると決定でき、ユーザが音楽類のアプリケーションを開くと、ターゲットシーンが音楽シーンであると決定でき、ユーザが天気問い合わせインターフェイスを開くと、ターゲットシーンが天気シーンであると決定できるが、ここでは列挙されていない。

【0079】

本願の実施例では、ターゲットシーンで、音声認識制御装置は第1動作状態の構成情報を照会し、該ターゲットシーンが第1動作状態に適用されるか否かを決定することができる。具体的には、構成情報が第1動作状態のシーンリストに該ターゲットシーンが含まれることを示す場合、該ターゲットシーンが第1動作状態に適用されると決定でき、構成情報が第1動作状態のシーンリストに該ターゲットシーンが含まれないことを示す場合、該ターゲットシーンが第1動作状態に適用されないと決定する。

【0080】

一例としては、図2を参照すれば、図2は、本願の実施例の設定インターフェイス概略図である。ユーザは自分のニーズに応じて、音楽シーン、オーディオブックシーン及びビデオシーンから対応するシーンを選択し、第1動作状態のシーンリストを生成することができる。ユーザが音楽シーン及びビデオシーンを選択すると仮定する場合、第1動作状態のシーンリストが音楽シーン及びビデオシーンを含み、ターゲットシーンが音楽シーンであると仮定する場合、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されると決定できる。

【0081】

なお、図2はシーンリストが音楽シーン、オーディオブックシーン及びビデオシーンから選択して生成される場合のみを例としており、図2は、シーンリストが音楽シーン、オーディオブックシーン及びビデオシーンの少なくとも1つを含む場合を例としている。しかし、実際に適用するときには、シーンリストは、たとえば天気シーン、ショッピングシーンなど、その他のシーンをさらにも含む場合もよく、本願はこれについては制限しない。

【0082】

ステップ402において、ターゲットシーンが第1動作状態に適用される場合、実行中の第2動作状態を第1動作状態に切り替え、第2動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する。

【0083】

10

20

30

40

50

本願の実施例では、第2動作状態は、非リスニング状態であってもよく、電子デバイスが該非リスニング状態にあるときには、ユーザはウェイクワードを音声で入力し、該電子デバイスをウェイクアップして後続の音声認識を実行させることができる。ここでは、ウェイクワードは、電子デバイスの内蔵プログラムにより予め設定されるものであってよいし、又は、ユーザのパーソナライズニーズを満たすために、ウェイクワードは、ユーザが自分のニーズに応じて設定されるものであってよく、本願はこれについては限定しない。例示すると、電子デバイスがスマートスピーカである場合、ウェイクワードは、「小度小度(ドちゃんドちゃん)」であってもよい。

【0084】

本願の実施例では、電子デバイスが第2動作状態にあるとき、音声認識制御装置は、ユーザがウェイクワードを入力したか否かを検出し、入力した場合、ユーザがウェイクワードの後に入力したオーディオクリップを収集し、音声認識を行う。例示すると、電子デバイスがスマートスピーカである場合、該スマートスピーカが該第2動作状態にあるときに、ユーザが歌を再生したい場合には、「小度小度(ドちゃんドちゃん)、歌Aを再生してください」又は「小度小度(ドちゃんドちゃん)、歌を聞きたい」と音声で入力し、電子デバイスはウェイクワード後のオーディオクリップ「歌Aを再生してください」又は「歌を聞きたい」を認識して、対応する歌を再生する。

【0085】

本願の実施例では、音声認識制御装置が、ターゲットシーンが第1動作状態に適用されると決定した場合、実行中の第2動作状態を第1動作状態に切り替えることができ、音声認識制御装置が、ターゲットシーンが第1動作状態に適用されないと決定した場合、第2動作状態を実行し続けることができる。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザは自分のニーズに応じて、第1動作状態をトリガーするシーンを設定する。また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御するので、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

【0086】

ステップ403において、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。

【0087】

本願の実施例では、電子デバイスが第1動作状態にあるとき、電子デバイスはオーディオを連続的に収集してオーディオストリームを取得して、音声認識を行うことができる。これにより、該第1動作状態で、ユーザはウェイクワードを入力せずに、電子デバイスとリアルタイムにインタラクションする又は連続的にインタラクションすることができ、ユーザ操作を簡素化し、ユーザ体験を改善することができる。

【0088】

本願の実施例の音声認識制御方法は、ターゲットシーンで、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されるか否かを決定し、ターゲットシーンが第1動作状態に適用される場合、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する実行中の第2動作状態を第1動作状態に切り替え、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザは自分のニーズに応じて、第1動作状態をトリガーするシーンを設定することができる。また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御するので、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

【0089】

上記実施例を達成するために、本願は、さらに、音声認識制御装置を提供する。

【0090】

図6は、本願の実施例5に係る音声認識制御装置の構造概略図である。

10

20

30

40

50

【 0 0 9 1 】

図 6 に示すように、該音声認識制御装置 6 0 0 は、照会モジュール 6 0 1、切り替えモジュール 6 0 2 及び認識モジュール 6 0 3 を含む。

【 0 0 9 2 】

ここでは、照会モジュール 6 0 1 は、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合に、第 1 動作状態の構成情報を照会し、ターゲットシーンが第 1 動作状態に適用されるか否かを決定する。

【 0 0 9 3 】

切り替えモジュール 6 0 2 は、ターゲットシーンが第 1 動作状態に適用される場合に、実行中の第 2 動作状態を第 1 動作状態に切り替え、第 2 動作状態では、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する。

【 0 0 9 4 】

認識モジュール 6 0 3 は、第 1 動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。

【 0 0 9 5 】

更に、本願の実施例の 1 つの可能な実施態様では、図 7 を参照する。図 6 に示す実施例に基づいて、該音声認識制御装置 6 0 0 は、さらにモジュール 6 0 4 ~ 6 1 0 を含む。

【 0 0 9 6 】

第 1 取得モジュール 6 0 4 は、第 2 動作状態で、オーディオクリップを音声認識することにより取得された第 1 制御意図を取得する。

【 0 0 9 7 】

決定モジュール 6 0 5 は、第 1 制御意図がターゲットシーンとマッチングすると決定する。

【 0 0 9 8 】

第 2 取得モジュール 6 0 6 は、オーディオストリームを音声認識することにより取得された情報ストリームを取得し、情報ストリームから各候補意図を取得する。

【 0 0 9 9 】

フィルタリングモジュール 6 0 7 は、各候補意図から、ターゲットシーンの制御意図とマッチングする第 2 制御意図をフィルタリングする。

【 0 1 0 0 】

実行モジュール 6 0 8 は、第 2 制御意図がフィルタリングして取得された場合、第 2 制御意図に対応する制御命令を実行する。

【 0 1 0 1 】

終了モジュール 6 0 9 は、設定期間内に、第 2 制御意図が取得されなかった場合、第 1 動作状態を終了し、前記設定期間の範囲は 2 0 秒 ~ 4 0 秒である。

【 0 1 0 2 】

拒否モジュール 6 1 0 は、ターゲットシーンの制御意図とマッチングしない候補意図への応答を拒否する。

【 0 1 0 3 】

1 つの可能な実施態様としては、構成情報は第 1 動作状態に適用されるシーンリストを含む。シーンリストはユーザの選択操作に応答して、音楽シーン、オーディオブックシーン及びビデオシーンから選択して生成されるものである。

【 0 1 0 4 】

なお、音声認識制御方法についての上記図 1 ~ 図 4 の実施例での解釈説明も該実施例の音声認識制御装置に適用されるが、ここでは説明が省略される。

【 0 1 0 5 】

本願の実施例の音声認識制御装置は、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合には、第 1 動作状態の構成情報を照会して、ターゲットシーンが第 1 動作状態に適用されるか否かを決定し、ターゲットシーンが第 1 動作状態に適用される場合には、ウェイクワードに基づいてオーディオクリップを収集して、オーディオクリップを音声認識する実行中の第 2 動作状態を第 1 動作

10

20

30

40

50

状態に切り替え、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザは自分のニーズに応じて、第1動作状態をトリガーするシーンを設定することができる。また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御するので、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

【0106】

上記実施例を達成するために、本願は、さらに、音声認識制御装置を提供する。

【0107】

図8は、本願の実施例7に係る音声認識制御装置の概略図構造概略図である。

10

【0108】

図8に示すように、該音声認識制御装置800は、照会モジュール801、切り替えモジュール802及び認識モジュール803を備える。

【0109】

ここでは、照会モジュール801は、ターゲットシーンで、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されるか否かを決定する。

【0110】

切り替えモジュール802は、ターゲットシーンが第1動作状態に適用される場合、実行中の第2動作状態を第1動作状態に切り替え、第2動作状態で、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する。

20

【0111】

認識モジュール803は、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。

【0112】

なお、音声認識制御方法についての上記図5の実施例での解釈説明も該実施例の音声認識制御装置に適用されるが、ここでは説明が省略される。

【0113】

本願の実施例の音声認識制御装置は、ターゲットシーンで、第1動作状態の構成情報を照会して、ターゲットシーンが第1動作状態に適用されるか否かを決定し、ターゲットシーンが第1動作状態に適用される場合には、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する実行中の第2動作状態を第1動作状態に切り替え、第1動作状態では、オーディオを連続的に収集してオーディオストリームを取得し、オーディオストリームを音声認識する。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザは自分のニーズに応じて、第1動作状態をトリガーするシーンを設定することができる。また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御するので、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

30

【0114】

上記実施例を達成するために、本願は、少なくとも1つのプロセッサと、少なくとも1つのプロセッサに通信可能に接続されるメモリと、を含む電子デバイスをさらに提供する。ここでは、メモリには少なくとも1つのプロセッサによって実行している命令が記憶され、命令が少なくとも1つのプロセッサによって実行されることにより、少なくとも1つのプロセッサに本願の上記実施例に係る音声認識制御方法を実行可能である。

40

【0115】

上記実施例を達成するために、本願は、コンピュータ命令の非一時的なコンピュータ読み取り可能な記憶媒体をさらに提供し、コンピュータ命令はコンピュータに本願の上記実施例に係る音声認識制御方法を実行させる。

【0116】

本願の実施例に基づいて、本願は、さらに、電子デバイス及び読み取り可能な記憶媒体を提供する。

50

【0117】

図9に示すように、本発明の一実施例に係る音声制御方法の電子デバイスのブロック図である。電子デバイスは、ラップトップコンピュータ、デスクトップコンピュータ、ワークベンチ、携帯情報端末、サーバ、ブレードサーバ、メインフレームコンピュータ、およびその他の適切なコンピュータなど、さまざまな形式のデジタルコンピュータを表すことを目的としている。電子デバイスは、パーソナルデジタル処理、携帯電話、スマートフォン、ウェアラブルデバイス、その他の同様のコンピューティングデバイスなど、さまざまな形態のモバイルデバイスを表すこともできる。本明細書で示されるコンポーネント、それらの接続および関係、およびそれらの機能は単なる例であり、本明細書で説明および/または必要とされる本出願の実現を限定することを意図するものではない。

10

【0118】

図9に示すように、当該電子デバイスは、1つまたは複数のプロセッサ901、メモリ902、および高速インターフェイスおよび低速インターフェイスを含む様々なコンポーネントを接続するためのインターフェイスを含む。さまざまなコンポーネントは、異なるバスを使用して相互接続され、共通のマザーボードに取り付けられるか、または必要に応じて取り付けられてもよい。プロセッサは、外部入力/出力デバイス(たとえば、インターフェイスに結合されたディスプレイデバイス)にGUIのグラフィカル情報を表示するためにメモリに記憶されている命令を含む、電子デバイス内で実行される命令を処理することができる。他の実施例では、必要に応じて、複数のプロセッサおよび/または複数のバスを複数のメモリとともに使用することができる。同様に、複数の電子デバイスを接続し、それぞれの電子デバイスが必要な操作(たとえば、サーバアレイ、1グループのブレードサーバ、マルチプロセッサシステム)の一部を提供することができる。図9では、1つのプロセッサ901を例に挙げているインターフェイスを示している。

20

【0119】

メモリ902は、本出願によって提供される非一時的なコンピュータ読み取り可能な記憶媒体である。前記メモリは、少なくとも1つのプロセッサによって実行可能な命令を記憶し、その結果、前記少なくとも1つのプロセッサは、本出願に係る音声制御方法を実行する。本出願の非一時的なコンピュータ読み取り可能な記憶媒体は、本出願によって提供される音声制御方法をコンピュータに実行させるためのコンピュータ命令を記憶する。

【0120】

メモリ902は、非一時的なコンピュータ読み取り可能な記憶媒体として、非一時的なソフトウェアプログラム、非一時的なコンピュータ実行可能プログラム、および本出願の実施例の音声制御方法に対応するプログラム命令/モジュール(たとえば、図6に示される照会モジュール601、切り替えモジュール602、および認識モジュール603)を記憶することができる。プロセッサ901は、メモリ902に記憶されている非一時的なソフトウェアプログラム、命令、およびモジュールを実行することにより、サーバの様々な機能的アプリケーションおよびデータ処理を実行し、すなわち、前述の方法の実施例における音声制御方法を実現する。

30

【0121】

メモリ902は、ストレージプログラム領域およびストレージデータ領域を含むことができ、ストレージプログラム領域は、少なくとも1つの機能に必要なオペレーティングシステムおよびアプリケーションプログラムを記憶することができる。ストレージデータ領域は、電子デバイスに基づいて作成されたデータを記憶することができる。また、メモリ902は、高速ランダムアクセスメモリを含むことができ、少なくとも1つの磁気ディスクメモリ、フラッシュメモリ装置、または他の非一時的な固体メモリなどの非一時的なメモリも含むことができる。いくつかの実施例では、メモリ902は、プロセッサ901に対して遠隔に配置されたメモリを選択可能に含むことができ、これらの遠隔メモリは、ネットワークを介して電子デバイスに接続され得る。上記のネットワークの例としては、インターネット、イントラネット、ローカルエリアネットワーク、モバイル通信ネットワーク、およびそれらの組み合わせを含むが、これらに限定されない。

40

50

【 0 1 2 2 】

電子デバイスは、入力装置 9 0 3 および出力装置 9 0 4 をさらに含むことができる。プロセッサ 9 0 1、メモリ 9 0 2、入力装置 9 0 3、および出力装置 9 0 4 は、バスまたは他の方式を介して接続することができ、図 9 では、バスを介した接続を例に挙げている。

【 0 1 2 3 】

入力装置 9 0 3 は、入力された数字または文字情報を受信し、電子デバイスのユーザ設定および機能制御に関するキー信号入力を生成することができ、たとえば、タッチスクリーン、キーパッド、マウス、トラックパネル、タッチパネル、ポインティングスティック、1 つまたは複数のマウスボタン、トラックボール、ジョイスティックなどの入力デバイス電子デバイス。出力装置 9 0 4 は、表示装置、補助照明装置（たとえば、LED）、触覚フィードバック装置（たとえば、振動モータ）などを含んでもよい。当該表示装置は、液晶ディスプレイ（LCD）、発光ダイオード（LED）ディスプレイ、およびプラズマディスプレイを含んでもよいが、これらに限定されない。いくつかの実施例では、表示装置はタッチスクリーンであってもよい。

10

【 0 1 2 4 】

本明細書で説明されるシステムおよび技術の様々な実施例は、デジタル電子回路、集積回路システム、特定用途向けASIC（特定用途向け集積回路）、コンピュータハードウェア、ファームウェア、ソフトウェア、および/またはそれらの組み合わせで実現され得る。これらの様々な実施例は、少なくとも1つのプログラマブルプロセッサを含むプログラマブルシステム上で実行可能および/または解釈可能な1つまたは複数のコンピュータプログラムで実装されてもよく、当該プログラマブルプロセッサは、ストレージシステム、少なくとも1つの入力装置、および少なくとも1つの出力装置からデータおよび命令を受信し、データおよび命令を当該ストレージシステム、当該少なくとも1つの入力装置、および当該少なくとも1つの出力装置に送信することができる専用または汎用のプログラマブルプロセッサであってもよい。

20

【 0 1 2 5 】

これらのコンピューティングプログラム（プログラム、ソフトウェア、ソフトウェアアプリケーション、またはコードともいう）は、プログラマブルプロセッサの機械命令を含み、さらに高レベルの手続き型および/またはオブジェクト指向プログラミング言語、および/またはアセンブリ/機械言語を用いてこれらのコンピューティングプログラムを実施することができる。本明細書で使用される「機械読み取り可能な媒体」および「コンピュータ読み取り可能な媒体」という用語は、機械命令および/またはデータをプログラマブルプロセッサに提供するための任意のコンピュータプログラム製品、デバイス、および/または装置（たとえば、磁気ディスク、光ディスク、メモリ、プログラマブルロジックデバイス（PLD）を指し、機械読み取り可能な信号としての機械命令を受信する機械読み取り可能な媒体を含む、「機械読み取り可能な信号」という用語は、機械命令および/またはデータをプログラマブルプロセッサに提供するための信号を指す。

30

【 0 1 2 6 】

ユーザとのインタラクションを提供するために、本明細書に記載のシステムおよび技術は、ユーザに情報を表示するための表示装置（たとえば、CRT（陰極線管）またはLCD（液晶ディスプレイ）モニター）と、キーボードおよびポインティング装置（たとえば、マウスまたはトラックボール）とを有するコンピュータで実施されてもよく、ユーザは当該キーボードおよび当該ポインティング装置を介してコンピュータに入力を提供できる。他の種類の装置を使用してユーザとのインタラクションを提供することもできる。たとえば、ユーザに提供されるフィードバックは、任意の形式のセンシングフィードバック（たとえば、視覚フィードバック、聴覚フィードバック、または触覚フィードバック）であってもよく、任意の形式（たとえば、音響入力、音声入力、または触覚入力）を用いてユーザからの入力を受信することができる。

40

【 0 1 2 7 】

本明細書に記載のシステムおよび技術の各実施例は、バックエンドコンポーネントを含む

50

コンピューティングシステム（たとえば、データサーバとして）、またはミドルコンポーネントを含むコンピューティングシステム（たとえば、アプリケーションサーバ）、またはフロントエンドコンポーネントを含むコンピューティングシステム（たとえば、グラフィカルユーザインターフェイスまたはWebブラウザを備えたユーザコンピュータであってもよい。ユーザは、当該グラフィカルユーザインターフェイスまたは当該Webブラウザを介して、ここで説明されるシステムおよび技術の実施例とインタラクションできる）、またはこのようなバックエンドコンポーネント、ミドルコンポーネント、またはフロントエンドコンポーネントの任意の組み合わせを含むコンピューティングシステムで実施することができる。デジタルデータ通信の任意の形式または媒体（たとえば、通信ネットワーク）によってシステムのコンポーネントを相互接続してもよい。通信ネットワークの例は、ローカルエリアネットワーク（LAN）、ワイドエリアネットワーク（WAN）、およびインターネットを含む。

10

【0128】

コンピュータシステムは、クライアントとサーバとを含むことができる。通常、クライアントとサーバとは互いにリモートであり、通常は通信ネットワークを介してインタラクションする。それぞれのコンピュータで実行され、互いにクライアントとサーバとの関係を持つコンピュータプログラムによってクライアントとサーバとの関係を生成する。

【0129】

本願の実施例による技術案では、ターゲットシーンに関連するアプリケーションプログラムインターフェイスを呼び出したことが検出された場合、第1動作状態の構成情報を照会し、ターゲットシーンが第1動作状態に適用されるか否かを決定し、ターゲットシーンが第1動作状態に適用される場合、ウェイクワードに基づいてオーディオクリップを収集し、オーディオクリップを音声認識する実行中の第2動作状態を第1動作状態に切り替え、第1動作状態で、オーディオを連続的に収集してオーディオストリームを取得して、オーディオストリームを音声認識する。これにより、さまざまなユーザのパーソナライズニーズを満たすように、ユーザは自分のニーズに応じて、第1動作状態をトリガーするシーンを設定することができ、また、ターゲットシーンが第1動作状態に適用される場合のみ、装置が第1動作状態に入るように制御することができるので、ユーザに対する迷惑を減らし、ユーザ体験を改善することができる。

20

【0130】

なお、上記のさまざまな形態のプロセスを使用して、ステップを並べ替え、追加、または削除できることを理解されたい。たとえば、本出願で説明されている各ステップは、並列、順次、または異なる順序で実行できる。本出願で開示された技術的解決策の望ましい結果が達成できれば、本明細書は、ここで限定しない。

30

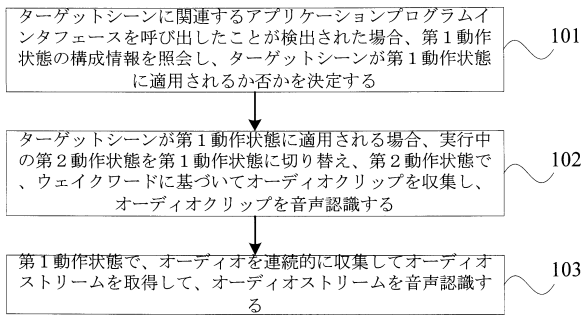
【0131】

前述した具体的な実施例は、本出願の保護範囲を制限しない。設計要件および他の要因に応じて、さまざまな修正、組み合わせ、サブ組み合わせ、および置き換えを行うことができることを当業者は理解するべきである。本出願の思想および原則の範囲内で行われた修正、同等の置き換え、および改善は、いずれも本出願の保護範囲に含まれるものとする。

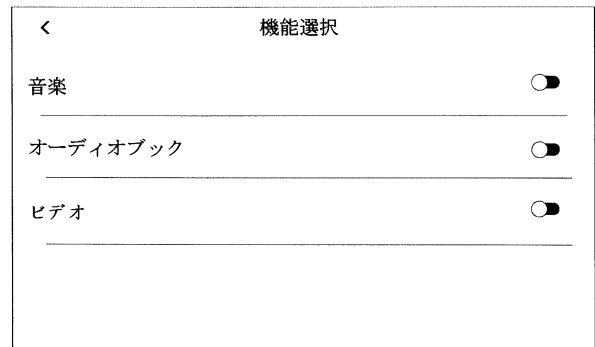
40

【図面】

【図 1】

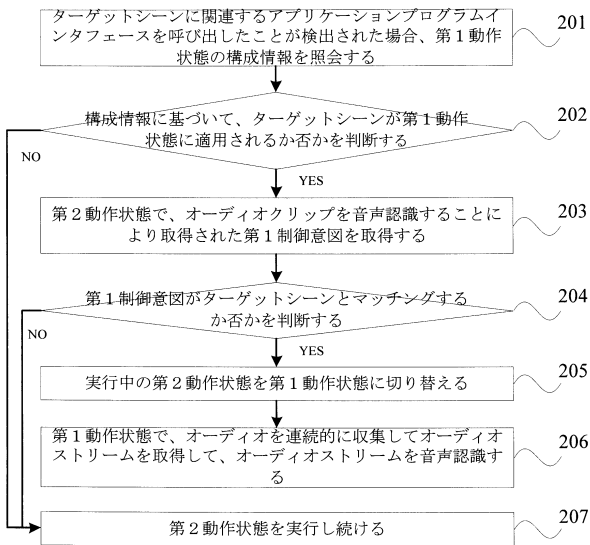


【図 2】

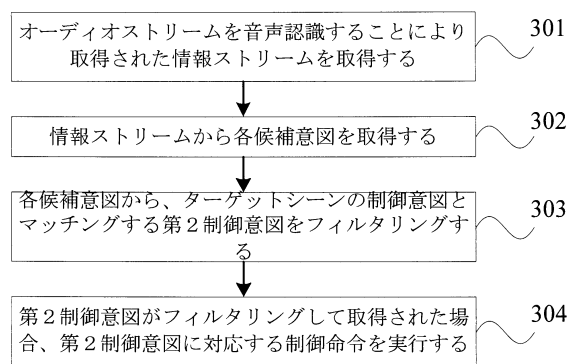


10

【図 3】



【図 4】



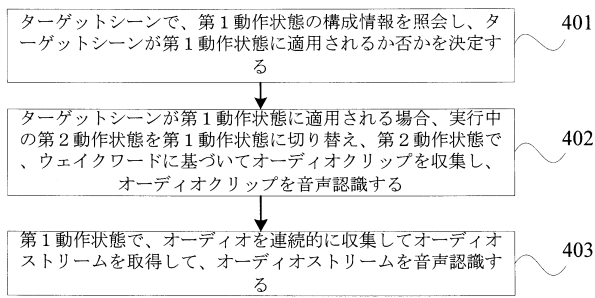
20

30

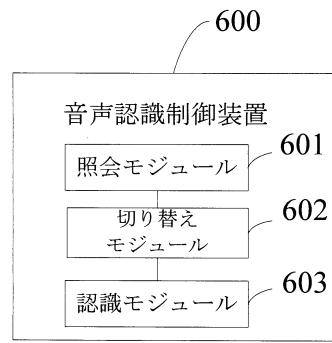
40

50

【図 5】

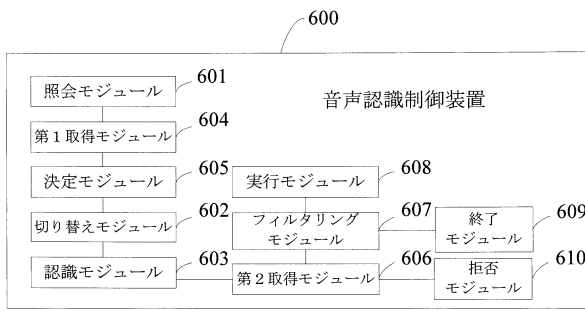


【図 6】

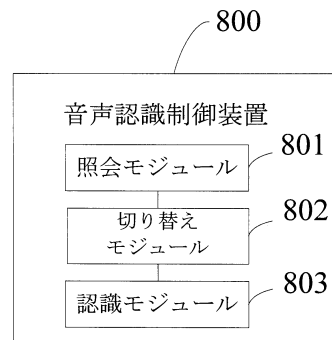


10

【図 7】

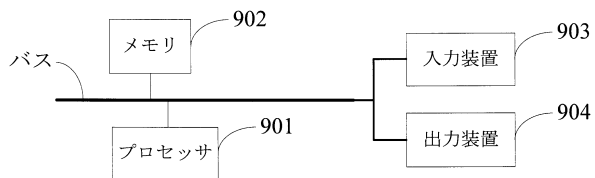


【図 8】



20

【図 9】



30

40

50

フロントページの続き

- フォース フロア
- (74)代理人 100118913
弁理士 上田 邦生
- (72)発明者 ルォ, ヨンシー
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10ティー
エイチ ストリート バイドウ キャンパス ナンバー 10 3 / フロア
- (72)発明者 ワン, シャシャ
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10ティー
エイチ ストリート バイドウ キャンパス ナンバー 10 3 / フロア
- 審査官 菊池 智紀
- (56)参考文献 米国特許出願公開第2017/0169817 (US, A1)
特開2019-139146 (JP, A)
国際公開第2019/022797 (WO, A1)
国際公開第2018/135753 (WO, A1)
米国特許出願公開第2019/0139547 (US, A1)
米国特許第09542941 (US, B1)
- (58)調査した分野 (Int.Cl., DB名)
G10L 15/00 - 15/34
G06F 3/16