

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6697024号
(P6697024)

(45) 発行日 令和2年5月20日(2020.5.20)

(24) 登録日 令和2年4月27日(2020.4.27)

| (51) Int.Cl. | | | F I | | |
|----------------|--------------|------------------|---------|-------|---------|
| G 1 0 L | 15/22 | (2006.01) | G 1 0 L | 15/22 | 3 0 0 Z |
| G 1 0 L | 15/00 | (2013.01) | G 1 0 L | 15/00 | 2 0 0 A |
| G 1 0 L | 15/10 | (2006.01) | G 1 0 L | 15/10 | 5 0 0 T |

請求項の数 19 外国語出願 (全 42 頁)

| | | | |
|--------------------|-------------------------------------|-----------|---|
| (21) 出願番号 | 特願2018-87328 (P2018-87328) | (73) 特許権者 | 503260918 |
| (22) 出願日 | 平成30年4月27日 (2018.4.27) | | アップル インコーポレイテッド |
| (62) 分割の表示 | 特願2015-109087 (P2015-109087) の分割 | | Apple Inc. |
| 原出願日 | 平成27年5月28日 (2015.5.28) | | アメリカ合衆国 95014 カリフォル ニア州 クパチーノ アップル パーク ウェイ ワン One Apple Park Way, Cupertino, Californ ia 95014, U. S. A. |
| (65) 公開番号 | 特開2018-136568 (P2018-136568A) | (74) 代理人 | 100076428 |
| (43) 公開日 | 平成30年8月30日 (2018.8.30) | | 弁理士 大塚 康徳 |
| 審査請求日 | 平成30年5月7日 (2018.5.7) | (74) 代理人 | 100115071 |
| (31) 優先権主張番号 | 62/005,760 | | 弁理士 大塚 康弘 |
| (32) 優先日 | 平成26年5月30日 (2014.5.30) | (74) 代理人 | 100112508 |
| (33) 優先権主張国・地域又は機関 | 米国 (US) | | 弁理士 高柳 司郎 |
| (31) 優先権主張番号 | 14/502,737 | | |
| (32) 優先日 | 平成26年9月30日 (2014.9.30) | | |
| (33) 優先権主張国・地域又は機関 | 米国 (US) | | |

最終頁に続く

(54) 【発明の名称】 手動始点/終点指定及びトリガフリーズの必要性の低減

(57) 【特許請求の範囲】

【請求項1】

電子デバイスにおいて仮想アシスタントを動作させる方法であって、
前記電子デバイスでオーディオ入力を受信することと、
ユーザ要求を含む第1のユーザ発話入力を識別するために前記オーディオ入力を監視することと、

前記オーディオ入力中の前記第1のユーザ発話入力を識別することと、

前記第1のユーザ発話入力に関連する文脈情報に基づいて前記第1のユーザ発話入力に
応答するか否かを判定することであって、前記文脈情報は、前記第1のユーザ発話入力
が受信されたときのユーザの注視の方向を含み、前記判定することは、

前記第1のユーザ発話入力に関連する前記文脈情報に基づいて前記仮想アシスタント
が前記第1のユーザ発話入力に対して可聴応答を提供すべきである尤度スコアを計算する
ことであって、前記計算することは、

前記第1のユーザ発話入力が受信されたときに前記ユーザの注視の前記方向が前記
電子デバイスを指していることに応じて前記尤度スコアを増加させることと、

前記第1のユーザ発話入力が受信されたときに前記ユーザの注視の前記方向が前記
電子デバイスからそれていることに応じて前記尤度スコアを減少させることと、を含む、
ことと、

前記尤度スコアを閾値と比較することと、

前記尤度スコアが前記閾値よりも大きいとの判定に従って、前記第1のユーザ発話入

前記距離が前記距離の閾値より短いことに応じて前記尤度スコアを増加させることとを含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

【請求項 8】

前記文脈情報は、前記第 1 のユーザ発話入力自動音声認識装置により認識されたか否かを示す指標を含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記第 1 のユーザ発話入力前記自動音声認識装置により認識されたことを示す指標に応じて前記尤度スコアを増加させることと、

前記第 1 のユーザ発話入力前記自動音声認識装置により認識されなかったことを示す指標に応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

10

【請求項 9】

前記文脈情報は、前記第 1 のユーザ発話入力以前のユーザ発話入力との意味的関係を含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを判定することは、

前記意味的関係の値がユーザ発話入力の意味の閾値より大きいことに応じて前記尤度スコアを増加させることと、

前記意味的関係の値が前記ユーザ発話入力の意味の閾値より小さいことに応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

20

【請求項 10】

前記文脈情報は前記第 1 のユーザ発話入力の長さを含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記第 1 のユーザ発話入力の前記長さが長さの閾値より短いことに応じて前記尤度スコアを増加させることと、

前記第 1 のユーザ発話入力の前記長さが前記長さの閾値より長いことに応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

30

【請求項 11】

前記文脈情報は、前記第 1 のユーザ発話入力の発話者の ID を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

【請求項 12】

前記文脈情報は、前記第 1 のユーザ発話入力受信された時刻を含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記時刻が所定の 1 組の時刻の中に含まれることに応じて前記尤度スコアを増加させることと、

前記時刻が前記所定の 1 組の時刻の中に含まれないことに応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

40

【請求項 13】

前記文脈情報は、前記第 1 のユーザ発話入力、前記第 1 のユーザ発話入力受信されたときに前記電子デバイスにより表示されていたコンテンツとの間の意味的関係を含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記意味的関係の値がコンテンツ意味の閾値より大きいことに応じて前記尤度スコアを増加させることと、

前記意味的関係の前記値が前記コンテンツ意味の閾値より小さいことに応じて前記尤度スコアを減少させることと

50

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

【請求項 14】

前記文脈情報は、前記第 1 のユーザ発話入力を受信されたときに前記ユーザにより実行されていたジェスチャーを含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記ジェスチャーが所定の 1 組のジェスチャーのうち 1 つのジェスチャーであることに応じて前記尤度スコアを増加させることと、

前記ジェスチャーが前記所定の 1 組のジェスチャーのうち 1 つのジェスチャーではないことに応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

10

【請求項 15】

前記文脈情報は、前記第 1 のユーザ発話入力と前記電子デバイスの以前の出力との間の意味的關係を含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記意味的關係の値が以前の出力の意味の閾値より大きいことに応じて前記尤度スコアを増加させることと、

前記意味的關係の値が前記以前の出力の意味の閾値より小さいことに応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

20

【請求項 16】

前記文脈情報は、前記第 1 のユーザ発話入力と以前の E メールとの間の意味的關係を含み、前記仮想アシスタントが文脈情報に基づいて前記第 1 のユーザ発話入力に回答すべきである前記尤度スコアを計算することは、

前記意味的關係の値が以前の Eメールの意味の閾値より大きいことに応じて前記尤度スコアを増加させることと、

前記意味的關係の値が前記以前の Eメールの意味の閾値より小さいことに応じて前記尤度スコアを減少させることと

を含むことを特徴とする請求項 1 乃至 3 の何れか 1 項に記載の方法。

【請求項 17】

前記第 1 のユーザ発話入力に関連する文脈情報に基づいて前記第 1 のユーザ発話入力に回答するか否かを判定することは、前記第 1 のユーザ発話入力を受信されたときの前記ユーザの注視の方向を判定することを含む、請求項 1 記載の方法。

30

【請求項 18】

1 つ以上のプロセッサによる実行のためのプログラムであって、

オーディオ入力を受信することと、

ユーザ要求を含む第 1 のユーザ発話入力を識別するために前記オーディオ入力を監視することと、

前記オーディオ入力中の前記第 1 のユーザ発話入力を識別することと、

前記第 1 のユーザ発話入力に関連する文脈情報に基づいて前記第 1 のユーザ発話入力に回答するか否かを判定することであって、前記文脈情報は、前記第 1 のユーザ発話入力を受信されたときのユーザの注視の方向を含み、前記判定することは、

40

前記第 1 のユーザ発話入力に関連する前記文脈情報に基づいて仮想アシスタントが前記第 1 のユーザ発話入力に対して可聴応答を提供すべきである尤度スコアを計算することであって、前記計算することは、

前記第 1 のユーザ発話入力を受信されたときに前記ユーザの注視の前記方向が電子デバイスを指していることに応じて前記尤度スコアを増加させることと、

前記第 1 のユーザ発話入力を受信されたときに前記ユーザの注視の前記方向が前記電子デバイスからそれていることに応じて前記尤度スコアを減少させることと、を含む、

前記尤度スコアを閾値と比較することと、

50

前記尤度スコアが前記閾値よりも大きいとの判定に従って、前記第1のユーザ発話入力に
応答すると判定することと、

前記尤度スコアが前記閾値よりも小さいとの判定に従って、前記第1のユーザ発話入
力に応答しないと判定することと、を含む、ことと、

前記第1のユーザ発話入力に応答するとの判定に応じて、

前記第1のユーザ発話入力に対する応答を生成することと、

第2のユーザ発話入力を識別するために前記オーディオ入力を監視することと、

前記第1のユーザ発話入力に応答しないという判定に応じて、前記第1のユーザ発話入
力に対する応答を生成せずに、前記第2のユーザ発話入力を識別するために前記オーディ
オ入力を監視することと

10

のための命令を含むことを特徴とするプログラム。

【請求項19】

仮想アシスタントを動作させるシステムであって、

1つ以上のプロセッサと、

メモリと、

1つ以上のプログラムと

を備え、前記1つ以上のプログラムは、前記メモリに格納され、前記1つ以上のプロセッ
サによって実行されるように構成され、前記1つ以上のプログラムは、

オーディオ入力を受信することと、

ユーザ要求を含む第1のユーザ発話入力を識別するために前記オーディオ入力を監視す
ることと、

20

前記オーディオ入力中の前記第1のユーザ発話入力を識別することと、

前記第1のユーザ発話入力に関連する文脈情報に基づいて前記第1のユーザ発話入力に
応答するか否かを判定することであって、前記文脈情報は、前記第1のユーザ発話入力
が受信されたときのユーザの注視の方向を含み、前記判定することは、

前記第1のユーザ発話入力に関連する前記文脈情報に基づいて前記仮想アシスタント
が前記第1のユーザ発話入力に対して可聴応答を提供すべきである尤度スコアを計算す
ることであって、前記計算することは、

前記第1のユーザ発話入力が受信されたときに前記ユーザの注視の前記方向が電子
デバイスを指していることに応じて前記尤度スコアを増加させることと、

30

前記第1のユーザ発話入力が受信されたときに前記ユーザの注視の前記方向が前記
電子デバイスからそれていることに応じて前記尤度スコアを減少させることと、を含む、
ことと、

前記尤度スコアを閾値と比較することと、

前記尤度スコアが前記閾値よりも大きいとの判定に従って、前記第1のユーザ発話入
力に応答すると判定することと、

前記尤度スコアが前記閾値よりも小さいとの判定に従って、前記第1のユーザ発話入
力に応答しないと判定することと、を含む、ことと、

前記第1のユーザ発話入力に応答するとの判定に応じて、

前記第1のユーザ発話入力に対する応答を生成することと、

第2のユーザ発話入力を識別するために前記オーディオ入力を監視することと、

前記第1のユーザ発話入力に応答しないという判定に応じて、前記第1のユーザ発話入
力に対する応答を生成せずに、前記第2のユーザ発話入力を識別するために前記オーディ
オ入力を監視することと

40

のための命令を含む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、一般に音声処理に関し、特に文脈データを使用するユーザ発話入力の選択的
処理に関する。

50

【背景技術】

【0002】

インテリジェント自動アシスタント（又は仮想アシスタント）は、ユーザと電子デバイスとの間に直感的インタフェースを構成する。このようなアシスタントより、ユーザは、発話及び／又はテキストの形で自然言語を使用してデバイス又はシステムと対話できる。例えば、電子デバイスと関連する仮想アシスタントにユーザ発話入力を提供することにより、ユーザは電子デバイスのサービスをアクセスできる。仮想アシスタントは、ユーザ発話入力からユーザの意図を解釈し、ユーザの意図をタスクとして運用化できる。電子デバイスの1つ以上の機能を実行することにより、それらのタスクを実行でき、関連出力を自然言語の形でユーザへ返送することができる。

10

【0003】

仮想アシスタントにユーザ発話入力を適正に処理させ且つそれに適正に応答させるために、まず仮想アシスタントは、通常始点指定及び終点指定とそれぞれ呼ばれる処理を使用してオーディオ入力のストリームの中のユーザ発話入力の始点と終点を識別できる。従来の仮想アシスタントは、受信されたオーディオストリームのエネルギーレベル及び／又は音響特性に基づいて、あるいはユーザによる手動識別により、始点及び終点を識別できる。例えば、仮想アシスタントの中には、仮想アシスタントに向かって話しかける前に物理的ボタン又は仮想ボタンを押すことにより、あるいは自然言語の形で仮想アシスタントに話しかける前に特定のトリガフレーズを発話することにより、ユーザに始点識別子を入力することを要求するものがある。これらの始点識別子の1つを受信したことに応答して、仮想アシスタントは、その後受信される音声をユーザ発話入力であると解釈できる。仮想アシスタントに向けられたユーザ発話入力を明確に識別するために、これらの技術を使用できるが、この方法で仮想アシスタントと対話するのは不自然であるか、又はユーザには難しいといえる。例えば、仮想アシスタントとユーザとの間の会話のやり取りの中で、ユーザはユーザ発話入力のたびに、発話前に始点識別子を入力すること（例えば、ボタンを押すこと又は同一のトリガフレーズを繰り返すこと）を要求できる。

20

【発明の概要】

【0004】

仮想アシスタントを動作させるシステム及び方法が開示される。方法の一実施形態は、電子デバイスでオーディオ入力を受信することと、第1のユーザ発話入力を識別するためにオーディオ入力を監視することと、オーディオ入力中の第1のユーザ発話入力を識別することと、第1のユーザ発話入力に関連する文脈情報に基づいて第1のユーザ発話入力に回答すべきか否かを判定することとを含むことができる。方法は、第1のユーザ発話入力に回答すると判定に回答して、第1のユーザ発話入力に対する回答を生成することと、第2のユーザ発話入力を識別するためにオーディオ入力を監視することとを更に含むことができる。方法は、第1のユーザ発話入力に回答すべきではないという判定に基づいて、第1のユーザ発話入力に対する回答を生成せずに、第2のユーザ発話入力を識別するためにオーディオ入力を監視することを更に含むことができる。

30

【図面の簡単な説明】

【0005】

【図1】図1は、種々の実施形態に係る仮想アシスタントが動作できる例示的な環境を示す図である。

40

【図2】図2は、種々の実施形態に係る例示的なユーザデバイスを示す図である。

【図3】図3は、種々の実施形態に係る文脈情報を使用してユーザ発話入力を選択的に処理し且つそれに回答する例示的な処理を示す図である。

【図4】図4は、種々の実施形態に係る文脈情報を使用してユーザ発話入力を選択的に処理し且つそれに回答する別の例示的な処理を示す図である。

【図5】図5は、種々の実施形態に係る文脈情報を使用してユーザ発話入力を選択的に処理し且つそれに回答するように構成された電子デバイスを示す機能ブロック図である。

【発明を実施するための形態】

50

【0006】

以下の実施形態の説明中、例示を目的として実施可能な特定の実施形態を示す添付の図面を参照する。他の実施形態も使用可能であり、種々の実施形態の範囲から逸脱することなく構造の変更を実施できることを理解すべきである。

【0007】

本発明は、ユーザ発話入力を選択的に処理し且つそれに応答するシステム及び方法に関する。一実施形態の方法において、ユーザ発話入力を含むオーディオ入力ユーザデバイスで受信できる。ユーザ発話入力の始点及び終点を識別することにより、オーディオ入力からユーザ発話入力識別できる。ユーザ発話入力ユーザデバイスで実行中の仮想アシスタントに向けられたか否か及び仮想アシスタントが文脈情報に基づいてユーザ発話入力に応答すべきか否かが判定できる。この判定は、規則依存システム又は確率（例えば、機械学習）システムを使用して実行できる。ユーザ発話入力仮想アシスタントに向けられたこと及び仮想アシスタントがユーザ発話入力に応答することが判定された場合、ユーザ発話入力は処理され、適切な応答が生成される。逆に、ユーザ発話入力仮想アシスタントに向けられなかったと判定された場合、ユーザ発話入力は無視されることもあり且つ/又は応答は生成されないこともある。ユーザ発話入力仮想アシスタントに向けられた否かを判定するために文脈情報を使用することにより、ユーザは、ユーザ発話入力のたびに、発話前に始点を手動操作で識別する（例えば、ボタンを押すか又はトリガフレーズを発話することにより）必要なく仮想アシスタントと対話できるので好都合である。

【0008】

システムの概要

図1は、種々の実施形態に係る仮想アシスタントを実現する例示的なシステム100を示す。「仮想アシスタント」、「デジタルアシスタント」、「インテリジェント自動化アシスタント」又は「自動デジタルアシスタント」という用語は、ユーザの意図を推測するために発話及び/又はテキストの形の自然言語を解釈し且つ推測されたユーザの意図に基づいてアクションを実行する何らかの情報処理システムを表すことができる。例えば、推測されたユーザの意図に基づいて動作するために、システムは、推測されたユーザの意図を実現するように設計されたステップ及びパラメータを有するタスク流れを識別すること、推測されたユーザの意図からタスク流れの中に特定の要件を入力すること、プログラム、方法、サービス、APIなど呼び出すことによりタスク流れを実行すること、及び聴き取れる形（例えば、音声）及び/又は目に見える形でユーザに対して出力応答を生成することのうち1つ以上を実行できる。

【0009】

仮想アシスタントは、少なくとも一部で自然言語コマンド、要求、ステートメント、語り及び/又は問い合わせの形のユーザ要求を受け入れることが可能である。通常、ユーザ要求は、仮想アシスタントによる情報応答又はタスクの実行のどちらか一方を求める。ユーザ要求に対する満足できる応答は、要求された情報応答の提供又は要求されたタスクの実行のいずれか一方、あるいはそれら2つの組み合わせを含むことができる。例えば、ユーザは、「where am I right now?（ここはどこですか）」のような質問を仮想アシスタントにすることができる。ユーザの現在位置に基づいて、仮想アシスタントは「You are in Central Park（セントラルパークにいます）」と答えることができる。ユーザは、タスクの実行、例えば「Please remind me to call Mom at 4 PM today（今日午後4時にお母さんに電話することを気付けてください）」と要求することもできる。これに応答して、仮想アシスタントは要求を確認し、ユーザの電子スケジュールに適切な注意事項を作成することもできる。要求されたタスクの実行中、仮想アシスタントは、相当に長い時間にわたり何度も情報を交換することを含めて、連続する会話の中でユーザと対話する場合がある。情報又は種々のタスクの実行を要求するために仮想アシスタントと対話する方法は他にも数多くある。口頭による応答の提供及びプログラムに従ったアクションの実行に加えて、仮想アシスタントは、他の目に見える形又は聴き取れる形で（例えば、テキ

スト、警報、音楽、映像、アニメーションなどとして) 応答を提供でき、可能であれば複数のデバイスを使用して提供できる(例えば、電話の送受話器を介してテキストを音声として出力し且つテキストをテレビに表示する)。

【0010】

仮想アシスタントの一例は、本明細書に参考として全開示内容が取り入れられている2011年1月10日出願、名称「Intelligent Automated Assistant」の米国実用新案出願第12/987,982号公報で説明されている。

【0011】

図1に示されるように、いくつかの実施形態において、仮想アシスタントはクライアント-サーバモデルに従って実現可能である。仮想アシスタントは、ユーザデバイス102 10
で実行されるクライアント側部分とサーバシステム110で実行されるサーバ側部分とを含むことができる。ユーザデバイス102は、携帯電話、タブレットコンピュータ、ポータブルメディアプレイヤー、デスクトップコンピュータ、ラップトップコンピュータ、PDA、テレビ、テレビのセットトップボックス、ウェアラブル電子デバイスなどの何らかの電子デバイスを含むことができ、インターネット、イントラネット、あるいは他の何らかの有線又は無線のパブリックネットワーク又はプライベートネットワークを含む1つ以上のネットワーク108を介してサーバシステム110と通信可能である。ユーザデバイス102で実行されるクライアント側部分は、ユーザ向け入出力処理及びサーバシステム110との通信などのクライアント側機能性を実行可能である。サーバシステム110は、各々のユーザデバイス102に常駐する任意の数のクライアントに対してサーバ側機能性を実行可能である。 20

【0012】

サーバシステム110は1つ以上の仮想アシスタントサーバ114を含むことができ、仮想アシスタントサーバ114は、クライアント向けI/Oインタフェース122、1つ以上の処理モジュール118、データ及びモデル記憶装置120、及び外部サービスへのI/Oインタフェース116を含むことができる。クライアント向けI/Oインタフェース122は、仮想アシスタントサーバ114のクライアント向け入出力処理を容易にすることができる。1つ以上の処理モジュール118は、自然言語入力に基づいてユーザの意図を判定し且つ推測されたユーザの意図に基づいてタスクを実行するためにデータ及びモデル記憶装置120を利用可能である。いくつかの実施形態において、仮想アシスタントサーバ114は、タスク完了又は情報収集のために、電話サービス、カレンダーサービス、情報サービス、メッセージ送信サービス、ナビゲーションサービスなどの外部サービス124とネットワーク108を介して通信可能である。外部サービスへのI/Oインタフェース116は、そのような通信を容易にすることができる。 30

【0013】

サーバシステム110は、1つ以上の独立型データ処理デバイス又はコンピュータの分散ネットワークで実現可能である。いくつかの実施形態において、サーバシステム110は、サーバシステム110の下位の計算資源及び/又はインフラ構造資源を提供するために、種々の仮想デバイス及び/又は第三者サービスプロバイダ(例えば、第三者クラウドサービスプロバイダ)を使用可能である。 40

【0014】

図1では、仮想アシスタントの機能性はクライアント側部分及びサーバ側部分の双方を含むように示されているが、いくつかの実施形態において、アシスタントの機能は、ユーザデバイスにインストールされた独立型アプリケーションとして実現可能である。更に、仮想アシスタントのクライアント部分とサーバ部分への機能性の分割は、実施形態によって異なる方法により実現可能である。例えば、いくつかの実施形態において、ユーザデバイス102で実行されるクライアントは、ユーザ向け入出力処理機能のみを提供し且つ仮想アシスタントの他のすべての機能性をバックエンドサーバに委託するシンクライアントでありうる。

【0015】

ユーザデバイス

図2は、種々の実施形態に係るユーザデバイス102のブロック図である。図示されるように、ユーザデバイス102は、メモリインタフェース202、1つ以上のプロセッサ204、及び周辺装置インタフェース206を含むことができる。ユーザデバイス104の種々のコンポーネントは、1つ以上の通信バス又は信号回線により互いに結合できる。ユーザデバイス102は、周辺装置インタフェース206に結合された種々のセンサ、サブシステム及び周辺装置を更に含むことができる。センサ、サブシステム及び周辺装置は、情報を収集し且つ/又はユーザデバイス102の種々の機能を容易にする。

【0016】

例えば、ユーザデバイス102は、向き感知機能、光感知機能及び近接感知機能を容易にするために周辺装置インタフェース206に結合されたモーションセンサ210、光センサ212及び近接センサ214を更に含むことができる。測位システム(例えば、GPS受信機)、温度センサ、生体測定センサ、ジャイロ스코ープ、コンパス、加速度計などの1つ以上の他のセンサ216も、関連機能を助けるために周辺装置インタフェース206に接続される。

【0017】

いくつかの実施形態において、写真撮影及びビデオクリップの録画などのカメラ機能を助けるために、カメラサブシステム220及び光学センサ222が利用できる。1つ以上の有線通信サブシステム及び/又は無線通信サブシステム224を介して通信機能が容易にされることができ、それらのサブシステム224は、種々の通信ポート、無線周波数受信機及び送信機及び/又は光(例えば、赤外線)受信機及び送信機を含むことができる。音声認識、音声複製、デジタル記録及び電話機能などの音声有効化機能を容易にするために、オーディオサブシステム226がスピーカ228及びマイク230に結合できる。

【0018】

いくつかの実施形態において、ユーザデバイス102は、周辺装置インタフェース206に結合されたI/Oサブシステム240を更に含むことができる。I/Oサブシステム240は、タッチスクリーンコントローラ242及び/又は他の入力コントローラ244を含むことができる。タッチスクリーンコントローラ242はタッチスクリーン246に結合できる。タッチスクリーン246及びタッチスクリーンコントローラ242は、例えば、容量性感知、抵抗性感知、赤外線感知、表面弾性波技術、近接センサアレイなどの複数の接触感知技術のうちいずれかを使用して接触及び動き又はその中断を検出可能である。他の入力コントローラ244は、1つ以上のボタン、ロッカースイッチ、サムホイール、赤外線ポート、USBポート及び/又はスタイラスなどのポインタデバイスのような他の入力/制御デバイス248に結合できる。

【0019】

いくつかの実施形態において、ユーザデバイス102は、メモリ250に結合されたメモリインタフェース202を更に含むことができる。メモリ250は、何らかの電子システム、装置又はデバイス、磁気システム、装置又はデバイス、光学システム、装置又はデバイス、電磁システム、装置又はデバイス、赤外線システム、装置、又はデバイス、あるいは半導体システム、装置又はデバイス、ポータブルコンピュータディスク(磁気)、ランダムアクセスメモリ(RAM)(磁気)、読み取り専用メモリ(ROM)(磁気)、消去可能プログラム可能読み取り専用メモリ(EPROM)(磁気)、CD、CD-R、CD-RW、DVD、DVD-R又はDVD-RWなどのポータブル光ディスク、コンパクトフラッシュカード、セキュアドデジタルカード、USBメモリデバイス、メモリスティックなどのフラッシュメモリを含むことができる。いくつかの実施形態において、コンピュータベースシステム、プロセッサを含むシステム、又は命令実行システム、装置又はデバイスから命令を取り出し、それらの命令を実行することができる他のシステムのような命令実行システム、装置又はデバイスにより使用するため、あるいはそのシステム、装置又はデバイスと関連して使用するための命令(例えば、以下に説明される処理300及び/又は400を実行するための命令)を記憶するために、メモリ250の非一時的コ

10

20

30

40

50

ンピュータ可読媒体が使用できる。他の実施形態において、命令（例えば、以下に説明される処理300及び/又は400を実行するための命令）は、サーバシステム110の非一時的コンピュータ可読記憶媒体に記憶されうるが、メモリ250の非一時的コンピュータ可読記憶媒体と、サーバシステム110の非一時的コンピュータ可読記憶媒体とに分割されることも可能である。本明細書に関して、「非一時的コンピュータ可読記憶媒体」は、命令実行システム、装置又はデバイスにより使用するため、あるいはそのシステム、装置又はデバイスと関連して使用するためのプログラムを含むか又は記憶することができる何らかの媒体でありうる。

【0020】

いくつかの実施形態において、メモリ250は、オペレーティングシステム252と、通信モジュール254と、グラフィカルユーザインタフェースモジュール256と、センサ処理モジュール258と、電話モジュール260と、アプリケーションモジュール262とを記憶可能である。オペレーティングシステム252は、基本システムサービスを処理し且つハードウェア依存タスクを実行するための命令を含むことができる。通信モジュール254は、1つ以上の追加のデバイス、1つ以上のコンピュータ及び/又は1つ以上のサーバとの通信を容易にすることができる。グラフィカルユーザインタフェースモジュール256は、グラフィカルユーザインタフェース処理を容易にすることができる。センサ処理モジュール258は、センサに関連する処理及び機能を容易にすることができる。電話モジュール260は、電話に関連する処理及び機能を容易にすることができる。アプリケーションモジュール262は、電子メッセージ送信、ウェブブラウジング、メディア処理、ナビゲーション、撮像及び/又は他の処理及び機能のようなユーザアプリケーションの種々の機能性を容易にすることができる。

【0021】

本明細書において説明されるように、仮想アシスタントのクライアント側機能性を提供するために、メモリ250は、クライアント側仮想アシスタント命令を更に記憶可能であり（例えば、仮想アシスタントクライアントモジュール264に）且つ種々のユーザデータ266（例えば、特定ユーザの語彙データ、好みデータ及び/又はユーザの電子アドレス帳、トゥドゥリスト、買い物リストなどのような他のデータ）を記憶可能である。

【0022】

種々の実施形態において、仮想アシスタントクライアントモジュール264は、声入力（例えば、音声入力）、テキスト入力、タッチ入力及び/又はジェスチャー入力をユーザデバイス104の種々のユーザインタフェース（例えば、I/Oサブシステム240、オーディオサブシステム226など）を介して受信可能である。仮想アシスタントクライアントモジュール264は、更に聴き取れる形（例えば、音声出力）、目に見える形及び/又は触知できる形で出力を提供することが可能である。例えば、出力は、声、音、警報、テキストメッセージ、メニュー、グラフィックス、映像、アニメーション、振動及び/又はそれらのうち2つ以上の組み合わせとして提供できる。動作中、仮想アシスタントクライアントモジュール264は、通信サブシステム224を使用して仮想アシスタントサーバと通信可能である。更に、仮想アシスタントクライアントモジュール264は、ホームオートメーション機器などの他のデバイスと通信可能であり、それにより物理的世界に物理的影響（例えば、ドアの鍵を開ける）を与えることができるか、そのようなデバイスに埋め込むこともできる。

【0023】

いくつかの実施形態において、仮想アシスタントクライアントモジュール264は、ユーザ、現在のユーザとの対話及び/又は現在のユーザ入力と関連付けられた文脈を確定する目的でユーザデバイス102の周囲環境から追加情報を収集するために、種々のセンサ、サブシステム及び周辺装置を利用可能である。いくつかの実施形態において、ユーザの意図を推測するのを助けるために、仮想アシスタントクライアントモジュール264は、文脈情報又はそのサブセットをユーザ入力と共に仮想アシスタントサーバに提供可能である。仮想アシスタントは、ユーザに対する出力をどのように作成し且つ供給するかを判定

10

20

30

40

50

するために文脈情報を使用可能である。以下に更に詳細に説明されるように、文脈情報は、ユーザ発話入力が仮想アシスタントに向けられた否かを判定し且つ適切な応答を判定するためにユーザデバイス102又はサーバシステム110により更に使用できる。

【0024】

いくつかの実施形態において、ユーザ入力に付随する文脈情報は、照明、周囲の騒音、周囲の温度などのセンサ情報、周囲環境の画像又は映像、別の物体までの距離などを含むことができる。文脈情報は、ユーザデバイス102の物理的状態（例えば、デバイスの向き、デバイスの場所、デバイスの温度、電力レベル、速度、加速度、運動パターン、セルラ信号強度など）又はユーザデバイス102のソフトウェア状態（例えば、実行中の処理、インストール済プログラム、過去及び現在のネットワーク活動、バックグラウンドサービス、エラー記録、資源利用率、フロントモストアプリケーションなど）と関連付けられた情報を更に含むことができる。これらの種類の文脈情報はいずれも、ユーザ入力と関連する文脈情報として仮想アシスタントサーバに提供できる。更に、文脈情報は、脈拍、手のひらの温度、音声品質、顔の表情などの生体測定ユーザデータを更に含むことができる。

10

【0025】

いくつかの実施形態において、仮想アシスタントクライアントモジュール264は、仮想アシスタントサーバからの要求に応答して、ユーザデバイス102に記憶されている情報（例えば、ユーザデータ266）を選択的に提供可能である。仮想アシスタントクライアントモジュール264は、仮想アシスタントサーバ114による要求に応じて、自然言語会話インタフェース又は他のユーザインタフェースを介してユーザから追加入力を更に引き出せる。ユーザ要求の中で表現されるユーザの意図の推測及び/又は実現に際して仮想アシスタントサーバ114を助けるために、仮想アシスタントクライアントモジュール264は、追加入力を仮想アシスタントサーバ114へ送信可能である。

20

【0026】

種々の実施形態において、メモリ250は先に挙げた命令以外の命令又はそれより少ない数の命令を含むことができる。更に、ユーザデバイス102の種々の機能は、1つ以上の信号処理回路及び/又は特定アプリケーション向け集積回路を含めて、ハードウェアで及び/又はファームウェアで実現可能である。

30

【0027】

仮想アシスタントを動作させるための処理

図3は、種々の実施形態に係る文脈情報を使用してユーザ入力（例えば、ユーザ発話入力）を選択的に処理し且つそれに応答するために実行できる例示的な処理300を示す。いくつかの実施形態において、処理300は、図1に示されるシステム100に類似するシステム又はそれと同一のシステムを使用して実行できる。

【0028】

ブロック302において、ユーザデバイスでオーディオ入力を受信できる。オーディオ入力は、音楽、ユーザの声、背景騒音、それらの組み合わせなどの何らかの検出可能な音を含むことができる。いくつかの実施形態において、ユーザデバイス（例えば、ユーザデバイス102）は、ユーザの自然言語音声を含むオーディオ入力をマイク（例えば、マイク230）を介して受信可能である。マイクは、音声入力をアナログ表現又はデジタル表現に変換し、音声データを1つ以上のプロセッサ（例えば、プロセッサ204）に提供可能である。図には、ブロック302は処理300の他のブロックとは別であるように示されているが、いくつかの実施形態において、処理300のその他のブロックのうちいくつか又はすべてが実行中である間に、オーディオ入力はブロック302で継続的に受信されていることも可能であることを理解すべきである。

40

【0029】

ブロック304において、ユーザ発話入力を含むか又は含む可能性があるオーディオ入力のセグメントを識別するために、ブロック302で受信されたオーディオ入力が監視できる。いくつかの実施形態において、これは、オーディオ入力中のユーザ発話入力の始点

50

及び終点を識別するためにオーディオ入力の1つ以上の特性を監視することを含むことができる。始点及び終点は、ユーザ音声をオーディオ入力中の背景雑音から区別するためにオーディオ入力のエネルギー特徴（例えば、短時間エネルギー及び零交差率）に依存するアルゴリズムのような何らかの周知の始点/終点指定アルゴリズムを使用して識別できる。いくつかの実施形態において、ユーザデバイスのプロセッサは、十分に高いエネルギーを有し且つユーザ音声の零交差率特性を有するオーディオ入力セグメントを識別するために、デバイスのマイクから受信されたオーディオデータのエネルギーを解析可能である。他の実施形態において、ユーザデバイスは、ユーザ発話入力の始点及び終点を判定可能な遠隔サーバ（例えば、仮想アシスタントサーバ114）へオーディオデータを送信可能である。

10

【0030】

いくつかの実施形態において、ブロック304は、検出されたユーザ発話入力に対して音声/テキスト変換動作をデバイスでローカルに実行するか、あるいはそのような動作を実行可能な遠隔サーバへオーディオデータを送信することにより実行することを更に含むことができる。他の実施形態では、ブロック304は音声/テキスト変換動作を実行することを含まなくてもよい。その代わりに、音声/テキスト変換動作は、ブロック308で仮想アシスタントがユーザ発話入力に応答すると判定された後にブロック312で実行できる。

【0031】

ブロック306において、ブロック304でオーディオ入力を監視する間にユーザ発話入力が識別されたか否かが判定できる。ユーザ発話入力が識別されなかった場合、処理はブロック304に戻ることができる。しかし、ユーザ発話入力が識別された場合、処理はブロック308へ進むことができる。

20

【0032】

ブロック308において、ブロック304で識別されたユーザ発話入力が仮想アシスタントに向けられた否か（例えば、ユーザが仮想アシスタントに向けてユーザ発話入力を発し、そのユーザ発話入力に基づいて仮想アシスタントがタスクを実行するか又は応答を提供することを期待しているか否か）を文脈情報に基づいて判定することにより、仮想アシスタントがユーザ発話入力に回答すべきか否かが判定できる。ユーザ発話入力が仮想アシスタントに向けられた否かを判定するためにブロック308で使用できる文脈情報の種々の供給源の例は、以下に説明される。ブロック308は、ユーザデバイス、遠隔サーバ（例えば、仮想アシスタントサーバ114）又はそれらの組み合わせにおいて実行可能である。

30

【0033】

いくつかの実施形態において、ユーザがユーザ発話入力を仮想アシスタントに向ける意図した尤度又は信頼スコアを判定することにより、仮想アシスタントがユーザ発話入力に回答すべきか否かを判定するために、確率システムを使用することができる。確率システムは、ニューラルネットワークなどの機械学習システム又は分類器を含むことができる。更に、確率システムは、フィードバックループを使用して学習し、ユーザに適応できる。このような確率システムの実施形態では、尤度又は信頼スコアは、ユーザがユーザ発話入力が仮想アシスタントに向けられた計算上の確率の数値表現又は他の表現を含むことができる。その場合、仮想アシスタントがユーザ発話入力に回答すべきか否かを判定するために、計算上の尤度又は信頼スコアは閾値と比較できる。例えば、計算上の尤度又は信頼スコアが閾値より大きい場合、ユーザ発話入力は仮想アシスタントに向けられたと判定できる。しかし、計算上の尤度又は信頼スコアが閾値以下である場合、ユーザ発話入力は仮想アシスタントに向けられなかったと判定できる。

40

【0034】

尤度又は信頼性スコアはいくつかの方法で判定できる。例えば判定は、一般に何らかの数の異なる種類の文脈情報からの正の寄与、負の寄与及び/又は中立の寄与を合計することを含むことができる。例えば尤度又は信頼スコアは、一般公式 $P = C_1 + C_2 + C_3 +$

50

．．．．+CNを使用して計算でき、式中、Pは、ユーザ発話入力をユーザデバイスに向ける意図があった尤度又は信頼スコアを表し、C1．．．CNは、N個の異なる種類の文脈情報からの尤度又は信頼スコアへの正の寄与、負の寄与又は中立の寄与を表す正の値、負の値又は0の値でありうる。正の寄与は、ユーザ発話入力が入力が仮想アシスタントに向けられたことを示唆する種類の文脈情報を表し、負の寄与は、ユーザ発話入力が入力が仮想アシスタントに向けられなかったことを示唆する種類の文脈情報を表し、中立の寄与は、ユーザ発話入力が入力が仮想アシスタントに向けられた尤度に関して中立である種類の文脈情報を表すことができる。従って大きいPの値は、ユーザ発話入力が入力が仮想アシスタントに向ける意図した尤度が高くなることを示し、小さい値又は負の値のPは、ユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示せる。各文脈情報の寄与が尤度又は信頼スコアの判定に付け加える重み又は値は、均一であっても、不均一でもありうる。更に、各寄与が尤度又は信頼スコアの判定に付け加える重み又は値は、特定の文脈情報の種類の値に依存できる。例えば、寄与C1がユーザの声の音量に従属する場合、符号(例えば、+/-)及び/又はC1の大きさは、ユーザの声の音量の数値表現に依存できる。

【0035】

確率システムの例を説明したが、説明したシステム及び/又は他のスコア決定規約に対して変更を実施してもよいことを理解すべきである。例えば、先の例とは逆に、正の寄与は、ユーザ発話入力が入力が仮想アシスタントに向けられなかったことを示唆する種類の文脈情報を表し、負の寄与は、ユーザ発話入力が入力が仮想アシスタントに向けられたことを示唆する種類の文脈情報を表すことができる。他の例では、異なる種類の文脈情報からの寄与はすべて正とすることができ、大きな正の値は、ユーザ発話入力が入力が仮想アシスタントに向けられた(あるいは向けられなかった)と文脈情報が示唆することを示す。更に別の例において、異なる種類の文脈情報からの寄与はすべて負とすることができ、大きな負の値は、ユーザ発話入力が入力が仮想アシスタントに向けられた(あるいは向けられなかった)と文脈情報が示唆することを示す。

【0036】

他の実施形態において、ユーザ発話入力が入力が仮想アシスタントに向けられた否かを判定するために文脈情報に基づく何らかの数の条件付き規則を評価することにより、仮想アシスタントがユーザ発話入力に回答すべきか否かを判定するために規則依存システムが使用できる。いくつかの実施形態において、規則依存システムは決定木の使用を含むことができる。他の実施形態では、規則依存システムにより使用される規則はユーザの挙動に基づいて学習されてもよい。規則依存システムの一例を示すために、第1の規則は、ユーザがデバイスに向かい合っており且つユーザの声の音量が音量の閾値を超える場合、ユーザはユーザ発話入力は仮想アシスタントに向けられたと判定できるという条件を含むことができる。第2の規則は、ユーザのカレンダーに従って、ユーザが会議に出席中である場合、ユーザはユーザ発話入力が入力が仮想アシスタントに向けられなかったと判定できるという条件を含むことができる。ユーザ発話入力が入力が仮想アシスタントに向けられたこと又は向けられなかったことをデバイスに決定させるために、何らかの数の文脈情報によって決まる任意の数の条件を含む他の類似の規則が使用できる。いくつかの実施形態において、複数の規則が真であると評価される場合、ブロック308で実行される判定動作の結果としてランクの高い規則の結果を使用できるように、規則がランク付け可能である。更に、いくつかの実施形態において、どの規則も真であると評価されない場合、ユーザ発話入力が入力が仮想アシスタントに向けられた(又はユーザ発話入力が入力が仮想アシスタントに向けられなかった)というデフォルト判定が下される。

【0037】

ユーザ発話入力が入力が仮想アシスタントに向けられなかったので仮想アシスタントはユーザ発話入力に回答すべきでないとしてブロック308で判定された場合、ブロック310において、ユーザ発話入力を求めてオーディオ入力を監視するために、処理はブロック304に戻ることができる。いくつかの実施形態において、処理300は、ユーザ発話入力に対する回答を生成せずに、ブロック310からブロック304へ進むことができる。例えば、

処理300は、音声/テキスト変換の実行、ユーザの意図の推測、推測されたユーザの意図を実現するように設計されたステップ及びパラメータを有するタスク流れの識別、推測されたユーザの意図からタスク流れへの所定の要件の入力、プログラム、方法、サービス、APIなどを呼び出すことによるタスク流れの実行、並びに聴き取れる形及び/又は目に見える形のユーザへの出力応答の生成のうち1つ以上を実行せずに、ブロック310からブロック304へ進むことができる。逆に、ユーザ発話入力が入力が仮想アシスタントに向けられたので仮想アシスタントはユーザ発話入力に回答するとブロック308で判定された場合、処理はブロック312へ進むことができる。

【0038】

ブロック312において、ユーザデバイス及び/又は遠隔サーバにより、ユーザ発話入力に対する応答が生成される。いくつかの実施形態において、ユーザ発話入力に対する応答は、音声/テキスト変換の実行、ユーザの意図の推測、推測されたユーザの意図を実現するように設計されたステップ及びパラメータを有するタスク流れの識別、推測されたユーザの意図からタスク流れへの所定の要件の入力、プログラム、方法、サービス、APIなどを呼び出すことによるタスク流れの実行、並びに聴き取れる形（例えば、音声）及び/又は目に見える形のユーザへの出力応答の生成のうち1つ以上を含むことができる。例えば、ブロック312は、ユーザにより要求された動作（例えば、アプリケーションを開く、メッセージを送信する、連絡先に電話する、検索クエリを実行する、カレンダーで約束を作成するなど）を実行すること、ユーザにより要求された情報を提供すること（例えば、検索クエリの結果を返送する）、物理的環境を変化させるアクションを実行すること（例えば、ドアの鍵を閉めるために家庭用電気製品と通信する）などを含むことができる。それらの動作は、ユーザデバイスでローカルに実行できるが、処理のために遠隔サーバへデータを送信することにより、あるいはそれらの組み合わせにより実行できる。ブロック312で適切な応答を提供するためにユーザ発話入力を処理した後、処理はブロック304に戻ることができる。

【0039】

処理300を使用して、ユーザデバイスにより実現された仮想アシスタントは、仮想アシスタントに話しかける前に物理的ボタン又は仮想ボタンを押すこと、あるいは仮想アシスタントに自然言語で話しかける前に特定のトリガフレーズ（例えば、「Hey Siri」などの所定の語又は一連の語）を発話することなどの方法によりユーザが手動操作で始点識別子を入力する必要なく、ユーザが自然言語で仮想アシスタントに話しかけることが可能になるように、ユーザ発話入力を選択的に無視するか又はユーザ発話入力に回答することができる。いくつかの実施形態において、処理300は、ユーザデバイス102により受信されたすべてのユーザ発話入力を処理するために使用できる。

【0040】

図3の動作を例示するために、ユーザと仮想アシスタントとの間の対話の一例は、仮想アシスタントに話しかける前に物理的ボタン又は仮想ボタンを押すこと、あるいは特定のトリガフレーズ（例えば、「Hey Siri」などの所定の語又は一連の語）を発話することなどによりユーザが手動操作で始点識別子を入力する必要なく、「what time is it in New York?（ニューヨークは何時ですか）」のような質問を口頭ですることにより動作を開始できる。ブロック302において、ユーザデバイスは、ユーザの質問を含むオーディオ入力を受信可能である。ブロック304において、オーディオ入力はユーザ発話入力を検出するために監視できる。オーディオ入力はユーザの質問を含んでいたため、ブロック306において、ユーザ発話入力が識別されたと判定できる。ブロック308において、識別されたユーザ発話入力に関連する文脈情報に基づいて、仮想アシスタントがユーザの質問に回答すべきか否かが判定できる。この例では、質問する間にユーザはユーザデバイスを見ており且つユーザの声の音量は音量の閾値を超えていることを文脈情報が示すので、仮想アシスタントはユーザの質問に回答すると判定できる（規則依存システム又は確率システムのいずれかを使用して）。従って、処理はブロック310及び312へ進み、ユーザの質問に対する応答が生成される。例えば、プロ

10

20

30

40

50

ック312において、ユーザの意図を判定し、実行されるタスクを識別し、ユーザの問合せに対する応答の中で「5:01 p.m. (午後5時1分)」という時間を判定し且つ表示するための機能を実行するように、ユーザの質問が処理できる。その後、処理300はブロック304に戻ることができる。

【0041】

次に、ユーザは、仮想アシスタントに話しかける前に物理的ボタン又は仮想ボタンを押すこと、あるいは特定のトリガフレーズ(例えば、「Hey Siri」のような所定の語又は一連の語)を発話することなどにより、手動操作で始点識別子を入力する必要なく、「what is the weather there? (そちらはどんな天気ですか)」のような別の質問を口頭ですることにも可能である。ユーザデバイスにより受信され、ユーザの第2の質問を含むオーディオ入力は、ブロック304で監視できる。オーディオ入力はユーザの第2の質問を含んでいたため、ブロック306において、ユーザ発話入力が識別されたと判定できる。ブロック308において、識別されたユーザ発話入力に関連する文脈情報に基づいて、仮想アシスタントがユーザの質問に回答すべきか否かが判定できる。この例では、ユーザは第1の質問に対する返答を受信してから長さの閾値の時間内に第2の質問をし、これは、第2の質問が同一の会話の一部であったことを示唆するので、仮想アシスタントはユーザの質問に回答すると(規則依存システム又は確率システムを使用して)判定できる。従って、処理はブロック310及び312へ進むことができ、ユーザの質問に対する応答が生成される。例えば、ブロック312において、ユーザの意図を判定し、実行されるタスクを識別し、ユーザの問合せに回答して「the weather is sunny (晴天です)」と述べるメッセージを判定し且つ表示するための機能を実行するように、ユーザの質問は処理できる。その後、別のユーザ発話入力を求めてオーディオ入力を監視するために、処理300はブロック304に戻ることができる。

【0042】

他の実施形態において、ユーザデバイス102は、処理300が呼び出される前にユーザが始点識別子を手動操作で入力することを要求することもできる。例えば、ユーザは、初めて仮想アシスタントに話しかける前にトリガフレーズを発話すること、あるいは物理的ボタン又は仮想ボタンを押すことを求められる。手動操作で入力された始点識別子に回答して、処理300は先に説明したように実行されることができ、その後のユーザ発話入力は、追加の始点識別子を入力することをユーザに要求することなく処理できる。図4は、この変更された対話を実行するための処理の一実施形態を示す。

【0043】

ブロック402において、始点識別子が受信できる。始点識別子は、ユーザにより話しかけられたトリガフレーズ、物理的ボタン又は仮想ボタンの選択、あるいはユーザから受信された他の手動操作入力を含むことができる。ブロック404において、先に説明したブロック302と同様に又はまったく同じようにオーディオ入力が受信できる。ブロック406において、第1のユーザ発話入力の終点を識別することにより、ブロック404で受信されたオーディオ入力から第1のユーザ発話入力が識別できる。終点は、先に説明したように、オーディオ入力のエネルギー特徴に基づいて識別できる。ブロック408において、先に説明したブロック312と同様に又はまったく同じように、第1のユーザ発話入力に対する応答が生成される。しかし、ブロック312とは異なり、ブロック402において手動操作で入力された始点識別子が受信されているので、ブロック408は、ブロック308のように仮想アシスタントが第1のユーザ発話入力に回答すべきか否かを判定することなく実行できる。ブロック408で第1のユーザ発話入力に対する応答を生成した後、処理はブロック304へ進むことができる。いくつかの実施形態において、先にブロック404でオーディオ入力が受信されているので、ブロック302は省略できる。ブロック304、306、308、310及び312は、先に図3に関して説明したように繰り返し実行できる。

【0044】

いくつかの実施形態において、処理400でブロック304が呼び出された後、ブロック304、306、308、310及び312は、後続のすべてのユーザ発話入力に対して継続して実行できる。他の実施形態において、長さの閾値の時間を超える期間、ユーザ発話入力が受信されない場合（例えば、ブロック304及び306で）、ブロック304、306、308、310及び312の実行は停止させられることができ、従って、ユーザは、次のユーザ発話入力を入力する前にブロック402で始点識別子を入力しなければならない。

【0045】

処理400を使用する場合、ユーザデバイスにより実現される仮想アシスタントは、ユーザに手動操作により始点識別子を一度入力することを求めるだけであり、後続のユーザ発話入力のたびに、それに先立ってユーザが手動操作で始点識別子を繰り返し入力する必要なく、仮想アシスタントは、後続のユーザ発話入力を選択的に無視するか又はそれに応答することができる。

【0046】

図4の動作を例示するために、ユーザと仮想アシスタントとの間の対話の一例は、仮想アシスタントに話しかける前に物理的ボタン又は仮想ボタンを押すこと、あるいは特定のトリガフレーズ（例えば、「Hey Siri」のような所定の語又は一連の語）を発話することなどにより、ユーザが手動操作で始点識別子を入力することによって開始できる。始点識別子はブロック402で受信できる。次に、ユーザは、「what time is it in New York?（ニューヨークは何時ですか）」のような質問を口頭ですることとも可能である。次に、ユーザデバイスは、ブロック404において、ユーザの質問を含むオーディオ入力を受信可能である。ブロック406において、ユーザの質問の終点を識別することにより、オーディオ入力からユーザ発話入力が識別できる。ユーザはブロック402で入力が仮想アシスタントに向けられていることを手動操作で識別しているので、ブロック408において、ブロック308に類似する判定ステップを実行することなくユーザの質問に対する応答が生成される。例えば、ブロック408において、ユーザの意図を判定し、実行されるタスクを識別し、ユーザの問合せに応答して「5:01 p.m.（午後5時1分）」という時間を判定し且つ表示するための機能を実行するように、ユーザの質問は処理できる。その後、処理400はブロック304に戻ることができる。

【0047】

次に、ユーザは、仮想アシスタントに話しかける前に物理的ボタン又は仮想ボタンを押すこと、あるいは特定のトリガフレーズ（例えば、「Hey Siri」のような所定の語又は一連の語）を発話することなどにより、手動操作で始点識別子を入力する必要なく、「what is the weather like there?（そちらはどんな天気ですか）」のような別の質問を口頭ですることとも可能である。ユーザデバイスにより受信され、ユーザの第2の質問を含むオーディオ入力は、ブロック304及び306で繰り返し監視できる。オーディオ入力はユーザの第2の質問を含んでいたため、ブロック306において、ユーザ発話入力が識別されたと判定できる。ブロック308において、識別されたユーザ発話入力に関連する文脈情報に基づいて、仮想アシスタントがユーザの質問に応答すべきか否かが判定できる。この例では、ユーザが第1の質問に対する返答を受信してから長さの閾値の時間内に第2の質問をしたことを文脈情報が示し、これは、第2の質問が同一の会話の一部であったことを示唆するので、仮想アシスタントはユーザの質問に応答すると判定できる（規則依存システム又は確率システムのいずれかを使用して）。従って、処理はブロック310及び312へ進むことができ、ユーザの質問に対する応答が生成される。例えば、ブロック312において、ユーザの意図を判定し、実行されるタスクを識別し、ユーザの問合せに応答して「the weather is sunny（晴天です）」と述べるメッセージを判定し且つ表示するための機能を実行するように、ユーザの質問は処理できる。その後、別のユーザ発話入力を求めてオーディオ入力を監視するために、処理300はブロック304に戻ることができる。

【 0 0 4 8 】

いくつかの実施形態において、処理 3 0 0 又は処理 4 0 0 のブロック 3 0 4、3 0 6、3 0 8、3 1 0 及び 3 1 2 が実行されている間、ユーザ発話入力をユーザデバイス 1 0 2 に向ける意図があることを示すためにトリガフレーズ又は他の手動入力を使用せずに、ユーザデバイス 1 0 2 が自然言語の形のユーザ発話入力を受け入れ可能であることを示すために、ユーザデバイス 1 0 2 のディスプレイに視覚標識が表示できる。

【 0 0 4 9 】

更に、処理 3 0 0 及び 4 0 0 のブロックは特定の順序で図示され且つ説明されたが、処理 3 0 0 及び 4 0 0 のブロックが他の順序で又は同時に実行できることを理解すべきである。例えば、処理 3 0 0 において、ユーザデバイス 1 0 2 は、ブロック 3 0 4、3 0 6、3 0 8、3 1 0 及び 3 1 2 のうちいくつか又はそれらすべてが実行される間、ブロック 3 0 2 でオーディオ入力を継続的に受信していることも可能である。同様に、処理 4 0 0 において、ユーザデバイス 1 0 2 は、ブロック 3 0 4、3 0 6、3 0 8、3 1 0、3 1 2、4 0 6 及び 4 0 8 のうちいくつか又はそれらすべてが実行される間、ブロック 4 0 4 でオーディオ入力を継続的に受信していることも可能である。

【 0 0 5 0 】

処理 3 0 0 及び 4 0 0 のブロックはユーザデバイス 1 0 2、サーバシステム 1 1 0、又はユーザデバイス 1 0 2 とサーバシステム 1 1 0 との組み合わせで実行できることを理解すべきである。例えば、いくつかの実施形態において、処理 3 0 0 又は 4 0 0 のすべてのブロックはユーザデバイス 1 0 2 で実行できる。他の実施形態において、処理 3 0 0 又は 4 0 0 のすべてのブロックはサーバシステム 1 1 0 で実行できる。更に別の実施形態において、処理 3 0 0 又は 4 0 0 のいくつかのブロックはユーザデバイス 1 0 2 で実行されることができ、処理 3 0 0 又は 4 0 0 の他のブロックはサーバシステム 1 1 0 で実行できる。

【 0 0 5 1 】

文脈情報

前述のように、文脈情報は、分類器又は機械学習システムにより使用される特徴を含むことができ、ユーザ発話入力が仮想アシスタントに向けられた否かを判定するために、処理 3 0 0 又は 4 0 0 のブロック 3 0 8 において、任意の数の種類の文脈情報がプロセッサ 2 0 4 及び / 又はサーバシステム 1 1 0 により使用できる。以下に、文脈情報の種類のいくつかの例と、処理 3 0 0 のブロック 3 0 8 においてユーザ発話入力が仮想アシスタントに向けられた否かを判定するために、それらの種類の文脈情報をどのように使用できるかを説明する。

【 0 0 5 2 】

いくつかの実施形態において、文脈情報は、ユーザデバイス 1 0 2 のクロック又はタイマからの時間データを含むことができる。時間データは所望の 2 つ以上のイベントの間の時間の長さを表すことができる。例えば、時間データは、ユーザ発話入力が受信された時間と、ボタンの押下、マウスのクリック、スクリーンへのタッチ、以前のユーザ発話入力などの以前のユーザ入力が受信された時間との間の時間の長さを表すことができる。一般に、いくつかの実施形態において、2 つのイベントの間の時間の長さが短いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、イベントの間の時間の長さが長いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。しかし、他の実施形態では、2 つのイベントの間の時間の長さが長いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、イベントの間の時間の長さが短いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

【 0 0 5 3 】

規則依存システムの一実施形態において、使用できる 1 つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の 1 つとして）は、連続するユーザ発話

10

20

30

40

50

入力の間の時間の長さが持続時間の閾値より短い場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、連続するユーザ発話入力の間の時間の長さが持続時間の閾値以上である場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

【0054】

確率システムの一実施形態において、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために、連続するユーザ発話入力の間の時間の長さを使用されることができ、この寄与の値は、時間の長さの値との間に線形関係又は非線形関係を有することができる。例えば、持続時間の閾値未満の時間の長さは、最終的な尤度又は信頼スコアに正の値を寄与することができ、その正の値の大きさは、時間の長さが短くなるほど大きくなりうる。同様に、持続時間の閾値以上の時間の長さは、最終的な尤度又は信頼スコアに0の値又は負の値を寄与することができ、その負の値の大きさは、時間の長さが長くなるほど大きくなりうる。いくつかの実施形態において、連続するユーザ発話入力の間の時間の長さは、確率システムの機械学習システムを訓練するために使用できる。

10

【0055】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の中に配置されるか又はユーザデバイス102から離れた場所に配置されたメモリ250又は他の記憶装置からの会話履歴データを含むことができる。会話履歴データは、ユーザから受信された何らかの数の以前のユーザ発話入力及び/又はユーザデバイスにより生成され、ユーザに提供された応答を含むことができる。いくつかの実施形態において、現在のユーザ発話入力が以前に受信されたユーザ発話入力と同一であるか否かを判定するために、以前に受信されたユーザ発話入力が現在のユーザ発話入力と比較できる。それらの実施形態において、以前のユーザ発話入力と現在のユーザ発話入力との一致（例えば、ユーザが同じことを繰り返して発言していることによって起こる）は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示し、以前のユーザ発話入力と現在のユーザ発話入力との不一致は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなるか、又はユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度に関して中立であることを示すことができる。いくつかの実施形態において、ユーザが同じことを繰り返す発言は、確率システムの機械学習システムを訓練するためにフィードバックループで使用できる。

20

30

【0056】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力が以前に受信されたユーザ発話入力と同一であるか又は一致する場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力が以前に受信されたユーザ発話入力と同一ではない場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

40

【0057】

確率システムの一実施形態において、以前に受信されたユーザ発話入力が現在のユーザ発話入力と同一であるか又は一致するという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、以前に受信されたユーザ発話入力と現在のユーザ発話入力との不一致は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

【0058】

他の実施形態において、現在のユーザ発話入力及び会話履歴データの一部又はすべてに

50

対して意味的類似性解析が実行できる。いくつかの実施形態において、この解析は、判定されたユーザの意図（例えば、タプル<コマンド、パラメータ>の形をとる自然言語解釈段階の結果）の類似性を計算することを含むことができる。他の実施形態において、意味的距離を判定するための意味的類似性解析を実行することは、類似性行列と組み合わされた編集距離を判定することを含むことができる。これらの実施形態において、ブロック308においてユーザ発話入力が仮想アシスタントに向けられた尤度又は信頼スコアを判定するために、現在のユーザ発話入力と、以前に受信されたユーザ発話入力又はユーザデバイスにより生成され、ユーザに提供された応答のうち1つ以上との間の意味的距離が判定され、使用できる。これらの実施形態において、現在のユーザ発話入力と、以前に受信されたユーザ発話入力のうち1つ以上（例えば、直前のユーザ発話入力）及び/又はユーザデバイスにより生成され、ユーザに提供された応答のうち1つ以上との間の意味的距離が短いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、現在のユーザ発話入力と、以前に受信されたユーザ発話入力のうち1つ以上（例えば、直前のユーザ発話入力）及び/又はユーザデバイスにより生成され、ユーザに提供された応答のうち1つ以上との間の意味的距離が長いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなることを示すことができる。

10

【0059】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力と、1つ以上の以前のユーザ発話入力又はユーザデバイスにより生成された応答との間の意味的距離が閾値未満の場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力と、1つ以上の以前のユーザ発話入力又はユーザデバイスにより生成された応答との間の意味的距離が閾値以上である場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

20

【0060】

確率システムの一実施形態において、現在のユーザ発話入力と、1つ以上の以前のユーザ発話入力又はユーザデバイスにより生成された応答との間の意味的距離は、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、意味的距離との間に線形関係又は非線形関係を有することができる。例えば、閾値未満の意味的距離は、最終的な尤度又は信頼スコアに正の値を寄与することができ、意味的距離が短くなるほど、正の値の大きさは増加しうる。同様に、閾値以上の意味的距離は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができ、意味的距離が長くなるほど、負の値の大きさは増加しうる。

30

【0061】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の近接センサ214などの距離センサからの距離データを含むことができる。距離データは、ユーザデバイスとユーザとの間の空間的距離（例えば、ユーザデバイスとユーザの顔との間の距離）を表すことができる。一般に、いくつかの実施形態において、ユーザデバイスとユーザとの間の距離が短いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、ユーザデバイスとユーザとの間の距離が長いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなることを示すことができる。しかし、他の実施形態では、ユーザデバイスとユーザとの間の距離が長いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、ユーザデバイスとユーザとの間の距離が短いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなることを示すことができる。

40

【0062】

50

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザデバイスとユーザとの間の距離が距離の閾値未満の場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則と組み合わせられて、又は他の規則の中の複数の条件の1つとして）は、ユーザデバイスとユーザとの間の距離が距離の閾値以上である場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

【0063】

確率システムの一実施形態において、ユーザデバイスとユーザとの間の距離は、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、ユーザデバイスとユーザとの間の距離の値との間に線形関係又は非線形関係を有することができる。例えば、距離の閾値未満の距離は、最終的な尤度又は信頼スコアに正の値を寄与することができ、距離が短くなるほど、正の値の大きさは増加することができる。同様に、距離の閾値以上の距離は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができ、距離が長くなるほど、負の値の大きさを増加することができる。

10

【0064】

いくつかの実施形態において、文脈情報は、オーディオサブシステム226からのオーディオデータを含むことができる。オーディオデータはユーザ発話入力の音量の表現を含むことができる。一般に、いくつかの実施形態において、ユーザ発話入力の音量が大きいことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、ユーザ発話入力の音量が小さいことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなることを示すことができる。しかし、他の実施形態において、ユーザ発話入力の音量が小さいことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、ユーザ発話入力の音量が大きいことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなることを示すことができる。

20

【0065】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力の音量が音量の閾値より大きかった場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力の音量が音量の閾値以下であった場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

30

【0066】

確率システムの一実施形態において、ユーザ発話入力の音量は、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、ユーザ発話入力の音量の値との間に線形関係又は非線形関係を有することができる。例えば、音量の閾値より大きい音量は、最終的な尤度又は信頼スコアに正の値を寄与することができ、音量が大きいほど、正の値の大きさは増加する。同様に、音量の閾値以下の音量は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができ、音量が小さいほど、負の値の大きさは増加する。

40

【0067】

いくつかの実施形態において、文脈情報は、オーディオサブシステム226からのオーディオデータを含むことができる。オーディオデータは、ユーザ発話入力の音量の表現を含むことができる。いくつかの実施形態において、以前のユーザ発話入力が無視されており、現在のユーザ発話入力の音量が以前のユーザ発話入力より大きい場合、これは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなったことを示すことができる。

50

【0068】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力の音量が以前のユーザ発話入力の音量より大きい場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力の音量が以前のユーザ発話入力の音量以下である場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

【0069】

確率システムの一実施形態において、ユーザ発話入力の音量は、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、ユーザ発話入力の音量の値との間に線形関係又は非線形関係を有することができる。例えば、現在のユーザ発話入力の音量が直前のユーザ発話入力の音量より大きい場合、最終的な尤度又は信頼スコアに正の値が加算される。同様に、現在のユーザ発話入力の音量が直前のユーザ発話入力の音量より小さい場合、最終的な尤度又は信頼スコアに0又は負の値が加算される。

【0070】

他の実施形態において、ユーザデバイスの可聴範囲の付近又はその範囲内の何人かの個別の発話者を判定するために、周知の発話者認識技術を使用して、オーディオデータが解析できる。これらの実施形態において、2人以上の発話者がいるという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が低くなること（及び仮想アシスタントではなく、近くの別の人物に話しかけていること）を示すことができ、発話者はただ1人であるという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができる。

【0071】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が受信されたときに2人以上の発話者がいたと判定された場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が受信されたときに発話者はただ1人であったと判定された場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。

【0072】

確率システムの一実施形態において、ユーザ発話入力が受信されたときに2人以上の発話者がいたという判定は、最終的な尤度又は信頼スコアに負の値を寄与することができ、ユーザ発話入力が受信されたときに発話者はただ1人であったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

【0073】

更に別の実施形態において、ユーザ発話入力が既知のユーザ又はユーザデバイスの正規ユーザ（例えば、デバイスの所有者）から、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されたか否かを判定するために、周知のスピーカ認識技術を使用して、オーディオデータが解析できる。これらの実施形態において、ユーザ発話入力が既知のユーザ又は正規ユーザから、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されたという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度が高くなることを示すことができ、ユーザ発話入力が既知のユーザ又は正規ユーザから、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されなかったという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける

10

20

30

40

50

意図した尤度が低くなることを示すことができる。

【0074】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力既知のユーザ又は正規ユーザから、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されたと判定された場合、ユーザは現在のユーザ発話入力は仮想アシスタントに向けられたと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力既知のユーザ又は正規ユーザから、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されなかったと判定された場合、ユーザは現在のユーザ発話入力が仮想アシスタントに向けられなかったと判定できるという規則である。

10

【0075】

確率システムの一実施形態において、ユーザ発話入力が既知のユーザ又は正規ユーザから、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されたという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、ユーザ発話入力が既知のユーザ又は正規ユーザから、あるいは以前に受信されたユーザ発話入力と同一のスピーカから受信されなかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

【0076】

いくつかの実施形態において、文脈情報はユーザデバイス102のカメラサブシステム220からの画像データを含むことができる。画像データは、カメラサブシステム220により取り込まれた画像又は映像を表すことができる。いくつかの実施形態において、画像データは、ユーザデバイスとユーザとの間の距離を推定するために使用できる。例えば、ユーザデバイスとユーザとの間の距離を推定するために、画像中のユーザの大きさが使用できる。ユーザデバイスとユーザとの間の推定距離は、先に説明した近接センサ214からの距離と同様に又はまったく同じように規則依存システム又は確率システムで使用できる。

20

【0077】

他の実施形態において、画像データは、ユーザ発話入力が受信されたときにユーザがユーザデバイスを見ていたか否か又はユーザデバイスと向かい合っていたか否かを判定するために解析できる（例えば、周知のアイトラッキング技術を使用して）。これらの実施形態において、ユーザ発話入力が受信されたときにユーザがユーザデバイスを見ていたという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザ発話入力が受信されたときにユーザがユーザデバイスを見ていなかったという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すか、あるいはユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度に関して中立である。

30

【0078】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が受信されたときにユーザはユーザデバイスを見ていたと判定された場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が受信されたときにユーザはユーザデバイスを見ていなかったと判定された場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

40

【0079】

確率システムの一実施形態において、ユーザ発話入力が受信されたときにユーザはユーザデバイスを見ていたという判定は、最終的な尤度又は信頼スコアに正の値を寄与するこ

50

とができ、ユーザ発話入力を受信されたときにユーザはユーザデバイスを見ていなかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

【0080】

更に別の実施形態において、ユーザに対するデバイスの向きを判定するために画像データが解析できる。例えば、ユーザが光学センサ222の視野の中に現れるか否かに基づいてユーザがユーザデバイスの前方に位置しているか否かを判定するために、周知の顔認識技術を使用して画像データが解析できる。同様に、ユーザが特定のアクションを実行しているか否か（例えば、ユーザデバイスを指さす、ユーザデバイスで何らかのジェスチャーをするなど）、あるいは事前に定義されたような位置にいるか否か（例えば、テレビの前に座る、リモコンを持つなど）を判定するために、周知の画像認識技術を使用して画像データが解析できる。これらの実施形態において、ユーザ発話入力を受信されたときにユーザはユーザデバイスの前方にいたか、特定のアクションを実行していたか、又は事前に定義されたような位置にいたという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザ発話入力を受信されたときにユーザはユーザデバイスの前方にいなかったか、特定のアクションを実行していなかったか、又は事前に定義されたような位置にいなかったという判定は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すか、あるいはユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度に関して中立である。

10

20

【0081】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにユーザはユーザデバイスの前方にいたか、特定のアクションを実行していたか、又は事前に定義されたような位置にいたと判定された場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにユーザはユーザデバイスの前方にいなかったか、特定のアクションを実行していなかったか、又は事前に定義されたような位置にいなかったと判定された場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

30

【0082】

確率システムの一実施形態において、ユーザ発話入力を受信されたときにユーザはユーザデバイスの前方にいたか、特定のアクションを実行していたか、又は事前に定義されたような位置にいたという判定は、最終的な尤度又は信頼スコアに正の値を寄与ことができ、ユーザ発話入力を受信されたときにユーザはユーザデバイスの前方にいなかったか、特定のアクションを実行していなかったか、又は事前に定義されたような位置にいなかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

【0083】

いくつかの実施形態において、文脈情報は、ユーザデバイス102のモーションセンサ210からの向きデータを含むことができる。モーションセンサ210は、ユーザデバイスの自由空間内の向きを表す向きデータを生成可能な傾斜計、コンパス、ジャイロスコープなどの何らかの種類の向きセンサを含むことができる。いくつかの実施形態において、ユーザデバイスの特定の向き（例えば、デバイスの前面が上向きである、デバイスが直立している、ユーザがデバイスのディスプレイを見ることができるといった向きにあるなど）は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザの他の向き（例えば、デバイスの前面が下向きである、デバイスが逆さである、ユーザがデバイスのディスプレイを見ることができないような向きにあるなど）は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤

40

50

度は低くなることを示すことができる。

【0084】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにデバイスが1組の向き（例えば、デバイスの前面が上向きである、デバイスが直立している、ユーザがデバイスのディスプレイを見ることができるといったような向きにあるなど）のうち1つの向きにあった場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにデバイスが1組の向きのうち1つの向きになかった場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

10

【0085】

確率システムの一実施形態において、ユーザ発話入力を受信されたときにユーザデバイスが1組の向き（例えば、デバイスの前面が上向きである、デバイスが直立している、ユーザがデバイスのディスプレイを見ることができるといったような向きにあるなど）のうち1つの向きにあるという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、ユーザ発話入力を受信されたときにユーザデバイスが1組の向きのうち1つの向きになかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

20

【0086】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の他のセンサ216の中のGPS受信機からの場所データを含むことができる。場所データは、ユーザデバイスの地理的場所を表すことができる。いくつかの実施形態において、ユーザデバイスが特定の場所（例えば、家庭、オフィスなど）にある間にユーザ発話入力を受信することは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザデバイスが特定の他の場所（例えば、映画館、会議室など）にある間にユーザ発話入力を受信することは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

【0087】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにデバイスが1組の場所（例えば、家庭、オフィスなど）のうち1つの場所にあった場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにデバイスが1組の場所のうち1つの場所になかった場合、ユーザはユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

30

【0088】

確率システムの一実施形態において、ユーザ発話入力を受信されたときにユーザデバイスが1組の場所（例えば、家庭、オフィスなど）のうち1つの場所にあったという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、ユーザ発話入力を受信されたときにユーザデバイスが1組の場所のうち1つの場所になかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

40

【0089】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の中に配置されるか又はユーザデバイス102から離れた場所に配置されたメモリ250又は別の記憶装置からの動作状態データを含むことができる。動作状態データは、ユーザに対してコンテンツが表示又は他の形で提示されているか否か、ユーザに対して提示されているコンテンツの

50

種類又はID、ユーザデバイスにより実行中のアプリケーション、ユーザに対して近い時点で通知が提示されていたか否か、以前の又は直近の連絡、以前の又は直近のEメール、以前の又は直近のSMSメッセージ、以前の又は直近の電話の着呼/発呼、カレンダーへの記入、注意事項の記入、ウェブページ検索、ユーザデバイスのディスプレイのオン/オフ状態、ユーザデバイスがユーザ発話入力以外のユーザ入力を受信しているか否か、ユーザデバイスにおける設定、以前の活動などのユーザデバイスの動作状態に関連する何らかの情報を含むことができる。いくつかの実施形態において、ユーザデバイスが特定の動作状態（例えば、ユーザに対してコンテンツ又は他の情報が表示されている、コンテンツ又は他の情報が聴き取れる形でユーザに対して提示されている、ユーザと仮想アシスタントとの間の会話筆記録などの特定のコンテンツがユーザに対して提示されている、ユーザデバイスによりアプリケーションが実行されている、ユーザに対して近い時点で通知が提示されていた、ユーザデバイスのディスプレイがオンである、ユーザデバイスがマウス入力、キーボード入力、タッチ感知ディスプレイ入力などのユーザ発話入力以外のユーザ入力を受信している、近い時点で連絡先又は特定の連絡先との間でEメールの送受信があった、近い時点で連絡先又は特定の連絡先との間でSMSメッセージの送受信があった、近い時点で連絡先又は特定の連絡先との間で電話の着呼/発呼があった、ユーザデバイスで特定の設定が構成されている、以前の活動が実行されたなど）にある間にユーザ発話入力を受信することは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザデバイスが特定の他の動作状態（例えば、ユーザに対してコンテンツ又は他の情報が表示されていない、コンテンツ又は他の情報が聴き取れる形でユーザに対して提示されていない、ユーザと仮想アシスタントとの間の会話筆記録などの特定のコンテンツがユーザに対して提示されていない、ユーザデバイスによりアプリケーションが実行されていない、ユーザに対して近い時点で通知が提示されていなかった、ユーザデバイスのディスプレイがオフである、ユーザデバイスがマウス入力、キーボード入力、タッチ感知ディスプレイ入力などのユーザ発話入力以外のユーザ入力を受信していない、近い時点で連絡先又は特定の連絡先との間でEメールの送受信がなかった、近い時点で連絡先又は特定の連絡先との間でSMSメッセージの送受信がなかった、近い時点で連絡先又は特定の連絡先との間で電話の発呼/着呼がなかった、ユーザデバイスで特定の設定が構成されていない、以前の活動が実行されなかったなど）にある間にユーザ発話入力を受信することは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

【0090】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力を受信されたときにユーザデバイスのディスプレイがオンであり且つユーザデバイスがユーザに対してオーディオ情報を提示していた場合、ユーザはユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の中の1つとして）は、ユーザ発話入力を受信されたときにユーザデバイスのディスプレイがオフであり且つユーザデバイスがユーザに対してオーディオ情報を提示していなかった場合、ユーザはユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。ユーザ発話入力を仮想アシスタントに向ける意図があった又は意図がなかったことを判定させる規則を生成するために、他の種類の動作状態データが同様に使用できる。

【0091】

確率システムの一実施形態において、ユーザ発話入力を受信されたときにユーザデバイスのディスプレイがオンであり且つユーザデバイスがユーザに対してオーディオ情報を提示していたという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができる、ユーザ発話入力を受信されたときにユーザデバイスのディスプレイがオフであり且つユーザデバイスがユーザに対してオーディオ情報を提示していなかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の

10

20

30

40

50

大きさは全体的なシステム設計に基づいて調整できる。デバイスの状態が所定の1組の状態のうち1つであることを動作状態データが示すか否かに応じて、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を与えるために、他の種類の動作状態データが同様にして使用できることを理解すべきである。

【0092】

他の実施形態において、現在のユーザ発話入力及び動作状態データの一部又はすべてに対して、意味的類似性解析が実行できる。これらの実施形態において、ユーザ発話入力がユーザデバイスに向けられたか否かをブロック308で判定するために、現在のユーザ発話入力と動作状態データの1つ以上のコンポーネントとの間の意味的距離が判定され、使用できる。これらの実施形態において、現在のユーザ発話入力と動作状態データの1つ以上のコンポーネントとの間の意味的距離が短いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、現在のユーザ発話入力と動作状態データの1つ以上のコンポーネントとの間の意味的距離が長いことは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

10

【0093】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力と、動作状態データの1つ以上のコンポーネント（例えば、ユーザデバイスにより実行中であるアプリケーション、ユーザに提示された通知、連絡先リスト中の名前、以前の連絡、以前のEメール、以前のSMSメッセージ、ユーザに提示されているコンテンツ、ユーザデバイスがマップアプリケーションを実行している間の方向指示を求める要求などのユーザから受信されると予想されるコマンド、ユーザデバイスがアイズフリーモードにある間のコンテンツナビゲーション命令、先に「停止」又は「一時停止」命令を受信した後の「開始」命令など）との間の意味的距離が閾値未満の場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、現在のユーザ発話入力と動作状態データの1つ以上のコンポーネントとの間の意味的距離が閾値以上である場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

20

30

【0094】

確率システムの一実施形態において、現在のユーザ発話入力と動作状態データの1つ以上のコンポーネントとの間の意味的距離は、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、意味的距離との間に線形関係又は非線形関係を有することができる。例えば、閾値未満の意味的距離は、最終的な尤度又は信頼スコアに正の値を寄与することができ、意味的距離が短いほど、正の値の大きさは増加しうる。同様に、閾値以上である意味的距離は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができ、意味的距離が長いほど、負の値の大きさは増加しうる。

【0095】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の光センサ212からの照明データを含むことができる。照明データは、光センサ212により受け取られた周囲光の輝度を表す値を含むことができる。いくつかの実施形態において、ユーザ発話入力が受信されたときに感知された周囲光の輝度が高いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ（例えば、発話が許容される環境にユーザがいることを示す）、ユーザ発話入力が受信されたときに感知された周囲光の輝度が低いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる（例えば、映画館の中のような発話が許容されない環境にユーザがいることを示す）。しかし、他の実施形態において、ユーザ発話入力が受信されたときに感知された周囲光の輝度が低いほど、ユーザが現在の

40

50

ユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザ発話入力が受信されたときに感知された周囲光の輝度が高いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

【 0 0 9 6 】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が受信されたときに感知された周囲光の輝度が輝度の閾値より高い場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が受信されたときに感知された周囲光の輝度が輝度の閾値以下である場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

10

【 0 0 9 7 】

確率システムの一実施形態において、ユーザ発話入力が受信されたときに感知された周囲光の輝度は、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、感知された周囲光の輝度の値との間に線形関係又は非線形関係を有することができる。例えば、輝度の閾値より低い輝度は、最終的な尤度又は信頼スコアに負の値を寄与することができ、輝度の値が低いほど、負の値の大きさは増加しうる。同様に、輝度の閾値以上の輝度は、最終的な尤度又は信頼スコアに0又は正の値を寄与することができ、輝度の値が高いほど、正の値の大きさは増加しうる。

20

【 0 0 9 8 】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の中に配置されるか又はユーザデバイス102から離れた場所に配置された自動音声認識（ASR）エンジンからの（例えば、サーバシステム110からの）音声認識データを含むことができる。音声認識データは、ASRエンジンがユーザ発話入力を認識可能であったか否か及び/又はASRエンジンがユーザ発話入力に回答可能であるか否かを示す指標を含むことができる。いくつかの実施形態において、ASRエンジンがユーザ発話入力を認識可能であったこと及び/又はASRエンジンがユーザ発話入力に回答可能であることを示す指標は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ASRエンジンがユーザ発話入力を認識可能ではなかったこと及び/又はASRエンジンがユーザ発話入力に回答可能ではないことを示す指標は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

30

【 0 0 9 9 】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ASRエンジンがユーザ発話入力を認識可能であり且つ/又はユーザ発話入力に回答可能である場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ASRエンジンがユーザ発話入力を認識可能ではなく且つ/又はユーザ発話入力に回答可能ではない場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

40

【 0 1 0 0 】

確率システムの一実施形態において、ASRエンジンがユーザ発話入力を認識可能であり且つ/又はユーザ発話入力に回答可能であるという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、ASRエンジンがユーザ発話入力を認識可能ではなく且つ/又はユーザ発話入力に回答可能ではないという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシ

50

ステム設計に基づいて調整できる。

【0101】

他の実施形態において、ASRエンジンからの音声認識データは、ユーザ発話入力の長さ（例えば、語数、発話の持続時刻など）を示す指標を更に含むことができる。一般に、いくつかの実施形態において、ユーザ発話入力の長さが短いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザ発話入力の長さが長いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。しかし、いくつかの実施形態において、ユーザ発話入力の長さが長いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザ発話入力の長さが短いほど、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

10

【0102】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力の長さが長さの閾値より短い場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力の長さが閾値長さ以上である場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

20

【0103】

確率システムの一実施形態において、ユーザ発話入力の長さは、最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を計算するために使用されることができ、寄与の値は、ユーザ発話入力の長さの値との間に線形関係又は非線形関係を有することができる。例えば、長さの閾値未満の長さは、最終的な尤度又は信頼スコアに正の値を寄与することができ、長さが短くなるほど、正の値の大きさは増加しうる。同様に、長さの閾値以上の長さは、最終的な尤度又は信頼スコアに0又は負の値を寄与することができ、長さが長いほど、負の値の大きさは増加しうる。

【0104】

他の実施形態において、ASRエンジンからの音声認識データは、ユーザ発話入力の中から識別された名詞又は代名詞を更に含むことができる。例えば、音声認識データは、「ハニー」、「彼」、「彼女」、あるいは人物の苗字又は名前などの名詞又は代名詞を含むことができる。一般に、いくつかの実施形態において、それらの名詞又は代名詞のうち1つが存在することは、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができ、それらの名詞又は代名詞のうち1つが存在しないこと（又は「Siri」のような人間を指すのではない識別子が存在すること）は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができる。

30

【0105】

規則依存システムの一実施形態において、使用できる1つの規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が1組の名詞又は代名詞のうち1つを含む場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。使用できる別の規則（単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして）は、ユーザ発話入力が1組の名詞又は代名詞のうち1つを含まない（又は別の1組の名詞又は代名詞のうち1つを含む）場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。

40

【0106】

確率システムの一実施形態において、ユーザ発話入力が1組の名詞又は代名詞のうち1つを含むという判定は、最終的な尤度又は信頼スコアに負の値を寄与することができ、ユ

50

ーザ発話入力が1組の名詞又は代名詞のうち1つを含まない(又は別の1組の名詞又は代名詞のうち1つを含む)という判定は、最終的な尤度又は信頼スコアに0又は正の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

【0107】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の中に配置されるか又はユーザデバイス102から離れた場所に配置されたメモリ250又は別の記憶装置からのユーザデータを含むことができる。ユーザデータは、連絡先リスト、カレンダー、好み、個人情報、金融情報、家族情報などのユーザと関連する何らかの種類の情報を含むことができる。いくつかの実施形態において、ユーザ発話入力が仮想アシスタントに向けられたか否かの判定を助けるために、ブロック308で、ユーザデータは他の種類の文脈情報と比較できる。例えば、ユーザがユーザデバイスの仮想アシスタントと会話した尤度が高くなるようなイベント又は低くなるようなイベントに参加していたか否かを判定するために、ユーザ発話入力が受信された時刻がユーザのカレンダーと比較されることができ、ユーザの連絡先リストにある名前がユーザ発話入力の中で挙げられたか否かを判定するために、ASRエンジンからの音声認識データがユーザの連絡先リストの中の連絡先と比較されることができ、ユーザ発話入力が仮想アシスタントにより無視されるべき又は無視されるべきではない事前定義済みフレーズに対応するか否かを判定するために、ASRエンジンからの音声認識データがユーザの好みと比較できるなどである。

【0108】

規則依存システムの一実施形態において、使用できる1つの規則(単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして)は現在のユーザ発話入力が所定の1組の時刻の中の1つの時刻に受信された(例えば、ユーザが会議に出席していたか又は仮想アシスタントと会話しない活動であるとみなされるような活動に携わっていたことをユーザのカレンダーが示す時)場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。使用できる別の規則(単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして)は、現在のユーザ発話入力が所定の1組の時刻以外の時刻に受信された(例えば、ユーザが会議に出席していなかったか又は仮想アシスタントと会話しない活動であるとみなされるような活動に携わっていなかったことをユーザのカレンダーが示す時)場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。ユーザ発話入力を仮想アシスタントに向ける意図があった又は意図がなかったという判定を実行させる規則を生成するために、他の種類のユーザデータが同様に使用できる。

【0109】

確率システムの一実施形態において、ユーザが会議に出席していたか又は仮想アシスタントと会話しない活動であるとみなされるような活動に携わっていたことをユーザのカレンダーが示す時刻に現在のユーザ発話入力が受信されたという判定は、最終的な尤度又は信頼スコアに負又は0の値を寄与することができ、ユーザが会議に出席していなかったか又は仮想アシスタントと会話しない活動であるとみなされるような活動に携わっていなかったことをユーザのカレンダーが示す時刻に現在のユーザ発話入力が受信されたという判定は、最終的な尤度又は信頼スコアに正の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。最終的な尤度又は信頼スコアに対する正の寄与、負の寄与又は中立の寄与を与えるために、他の種類のユーザデータが同様にして使用できることを理解すべきである。

【0110】

いくつかの実施形態において、文脈情報は、モーションセンサ210又はユーザデバイス102の他のセンサ216の中の加速度計からの運動データを含むことができる。運動データは、ユーザデバイスの動きを表すことができ、ユーザをデバイスを振動させること、デバイスがユーザに対して接離すること(例えば、ユーザの口に対する接離)、ユーザがデバイスを着用することによる動き(例えば、時計又は他のウェアラブルデバイスとし

10

20

30

40

50

て)などによって起こるデバイスの動きを検出するために使用できる。いくつかの実施形態において、ユーザデバイスが受ける特定の動き(例えば、振動、ユーザがユーザデバイスを着用することと関連する動き、ユーザに接近する動きなど)は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は高くなることを示すことができ、ユーザデバイスが受ける他の動き(例えば、ユーザから離れる動き)は、ユーザが現在のユーザ発話入力を仮想アシスタントに向ける意図した尤度は低くなることを示すことができる。

【0111】

規則依存システムの一実施形態において、使用できる1つの規則(単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして)は、ユーザ発話入力を受信される前にユーザデバイスがユーザの口に向かって動かされたことを運動データが示す場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則(単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして)は、ユーザ発話入力を受信される前にデバイスがユーザの口に向かって動かされなかったことを運動データが示す場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。ユーザ発話入力を仮想アシスタントに向ける意図があった又は意図がなかったという判定を実行させる規則を生成するために、他の動きが同様に使用できる。

【0112】

確率システムの一実施形態において、ユーザデバイスが所定の1組の運動のうち1つで動かされた(例えば、ユーザ発話入力を受信される前にユーザの口に向かって動かされた)という判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、ユーザデバイスが所定の1組の運動のうち1つで動かされなかったという判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。最終的な尤度又は信頼スコアに対して正の寄与、負の寄与又は中立の寄与を加算するために他の動きが同様に使用できることを理解すべきである。

【0113】

いくつかの実施形態において、文脈情報は、ユーザデバイス102の他のセンサ216の中の温度センサからの温度データを含むことができる。温度データは、温度センサにより感知された温度を表すことができ、ユーザによりユーザデバイスが保持されているか否かを判定するために使用できる。例えば、高い温度又は特定の範囲内の温度は、デバイスがユーザの手で保持されていることを示唆してもよく、低い温度又は特定の範囲外の温度は、デバイスがユーザにより保持されていないことを示唆してもよい。

【0114】

規則依存システムの一実施形態において、使用できる1つの規則(単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして)は、温度がユーザデバイスがユーザにより保持されていることに対応する特定の温度範囲にある場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図があったと判定できるという規則である。使用できる別の規則(単独で、他の規則との組み合わせで、又は他の規則の中の複数の条件の1つとして)は、温度がユーザデバイスがユーザにより保持されていることに対応する特定の温度範囲内でない場合、ユーザは現在のユーザ発話入力を仮想アシスタントに向ける意図がなかったと判定できるという規則である。

【0115】

確率システムの一実施形態において、ユーザ発話入力を受信されたときにユーザデバイスがユーザにより保持されていたという温度データに基づく判定は、最終的な尤度又は信頼スコアに正の値を寄与することができ、ユーザ発話入力を受信されたときにユーザデバイスがユーザにより保持されていなかったという温度データに基づく判定は、最終的な尤度又は信頼スコアに0又は負の値を寄与することができる。正の寄与又は負の寄与の大きさは全体的なシステム設計に基づいて調整できる。

10

20

30

40

50

【0116】

電子デバイス

いくつかの実施形態に従って、図5は、説明される種々の実施形態の原理に従って構成された電子デバイス500の機能ブロック図を示す。デバイスの機能ブロックは、説明される種々の実施形態の原理を実行するためにハードウェア、ソフトウェア又はハードウェアとソフトウェアとの組み合わせにより実現できる。説明される種々の実施形態の原理を実現するために、図5に示される機能ブロックが組み合わせられるか又は複数のサブブロックに分割できることは当業者には理解される。従って、本明細書の説明は、説明される機能ブロックの可能なすべての組み合わせ又は分割又は更なる定義を任意に支援する。

【0117】

図5に示されるように、電子デバイス500は、ユーザインタフェースを表示し且タッチ入力を受信するように構成されたタッチスクリーンディスプレイユニット502と、音声入力を受信するように構成された音声受信ユニット504とを含むことができる。いくつかの実施形態において、電子デバイス500は、音声を発生するように構成されたスピーカユニット506を含むことができる。電子デバイス500は、タッチスクリーンディスプレイユニット502及び音声受信ユニット504に結合された（任意にスピーカユニット506にも結合された）処理ユニット508を更に含むことができる。いくつかの実施形態において、処理ユニット508は、第1の監視ユニット510、音声識別ユニット512、応答判定ユニット514、応答生成ユニット516、第2の監視ユニット518及び第3の監視ユニット520を含むことができる。

【0118】

処理ユニット508は、オーディオ入力（例えば、オーディオ受信ユニット504からの）を受信するように構成可能である。処理ユニット508は、オーディオ入力中の第1のユーザ発話入力を識別するためにオーディオ入力を監視する（例えば、第1の監視ユニット510を使用して）ように構成可能である。処理ユニット508は、オーディオ入力中の第1のユーザ発話入力を識別した場合、第1のユーザ発話入力に関連する文脈情報に基づいて、第1のユーザ発話入力に対する応答を生成すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。処理ユニット508は、応答を生成すると判定に回答して、第1のユーザ発話入力に対する応答を生成し（例えば、応答生成ユニット516を使用して）且つ第2のユーザ発話要求を求めて受信されたオーディオ入力を再び監視する（例えば、第2の監視ユニット518を使用して）ように構成可能である。処理ユニット508は、応答を生成すべきでないという判定に回答して、第1のユーザ発話入力に対する応答を生成せずに、第2のユーザ発話要求を求めて受信されたオーディオ入力を再び監視する（例えば、第2の監視ユニット518を使用して）ように構成可能である。

【0119】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力の開始時に1つ以上の所定の語を識別することなく、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。他の実施形態において、処理ユニット508は、第1のユーザ発話入力の受信の前に物理的ボタン入力又は仮想ボタン入力を識別することなく、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。

【0120】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力に対して音声/テキスト変換を実行し、第1のユーザ発話入力に基づいてユーザの意図を判定し、第1のユーザ発話入力に基づいて実行されるタスクを判定し、第1のユーザ発話入力に基づいて実行されるタスクのパラメータを判定し、実行されるタスクを実行し、第1のユーザ発話入力に回答してテキストを表示するか又は第1のユーザ発話入力に対するオーディオ応答を出力することにより、第1のユーザ発話入力に対する応答を生成する（例えば、応答生成ユニット516を使用して）ように構成可能である。

【 0 1 2 1 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力に関連する文脈情報に依存する1つ以上の条件付き規則を評価することにより、第1のユーザ発話入力に応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。

【 0 1 2 2 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力に関連する文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答する尤度スコアを計算し且つ尤度スコアを閾値と比較することにより、第1のユーザ発話入力に
10 応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。

【 0 1 2 3 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力の受信と以前のユーザ入力との間の経過時間、以前のユーザ発話入力、第1のユーザ発話入力を受信されたときのユーザと電子デバイスとの間の距離、第1のユーザ発話入力を受信されたときの電子デバイスの向き、第1のユーザ発話入力を受信されたときのユーザと電子デバイスとの間の向き、第1のユーザ発話入力を受信されたときのユーザの視線の方向、第1のユーザ発話入力
20 が入力が自動音声認識装置により認識されたか否かを示す指標、第1のユーザ発話入力と以前のユーザ発話入力との意味的關係、第1のユーザ発話入力の長さ、第1のユーザ発話入力の発話者のID、第1のユーザ発話入力を受信された時刻、第1のユーザ発話入力を受信されたときに電子デバイスがユーザに対して情報を出力していたか否かを示す指標、ユーザからの入力を受信する期待値、第1のユーザ発話入力を受信されたときに電子デバイスが保持されていたか否かを示す指標、第1のユーザ発話入力を受信されたときの電子デバイスの動作状態、電子デバイスにより実行された以前のアクション、第1のユーザ発話入力を受信されたときに電子デバイスによりコンテンツが表示されていたか否かを示す指標、第1のユーザ発話入力
30 と第1のユーザ発話入力を受信されたときに電子デバイスにより表示されていたコンテンツとの意味的關係、第1のユーザ発話入力を受信されたときのユーザの位置、第1のユーザ発話入力を受信されたときにユーザにより実行されていたジェスチャー、電子デバイスの以前の出力、第1のユーザ発話入力を受信されたときの電子デバイスの場所、第1のユーザ発話入力を受信されたときに電子デバイスにより実行されていたアプリケーション、以前の連絡、以前のEメール、以前のSMSメッセージ、第1のユーザ発話入力を受信されたときの電子デバイスの動き、電子デバイスのユーザ設定、第1のユーザ発話入力を受信されたときに電子デバイスにより感知された光の量及びカレンダーデータを含む、第1のユーザ発話入力に関連する文脈情報に基づいて、第1のユーザ発話入力に
40 応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。

【 0 1 2 4 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力の受信と以前のユーザ入力との間の経過時間を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に
40 応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に
50 応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、経過時間の値が持続時間の閾値より大きいことに応答して尤度スコアを減少させることと、経過時間の値が持続時間の閾値より小さいことに応答して尤度スコアを増加させることとを含むことができる。いくつかの実施形態において、以前のユーザ入力は、以前に接触感知ディスプレイで受信されたタッチ入力、マウスのクリック、ボタンの押下又はユーザ発話入力を含むことができる。

【 0 1 2 5 】

いくつかの実施形態において、処理ユニット508は、以前のユーザ発話入力を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に
50 応答すべき尤度スコアを

計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、以前のユーザ発話入力と第1のユーザ発話入力との一致を検出したことに回答して尤度スコアを増加させることを含むことができる。

【0126】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときのユーザと電子デバイスとの間の距離を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、距離が距離の閾値より長いことに回答して尤度スコアを減少させることと、距離が距離の閾値より短いことに回答して尤度スコアを増加させることを含むことができる。いくつかの実施形態において、距離は、少なくとも一部で第1のユーザ発話入力の音量、近接センサにより測定された距離、イメージセンサにより生成された画像、又は加速度計からの加速度計データに基づいて判定できる。

10

【0127】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときの電子デバイスの向きを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、デバイスの向きが下向き又は逆さであることに回答して尤度スコアを減少させることと、デバイスの向きが上向き又は直立であることに回答して尤度スコアを増加させることを含むことができる。

20

【0128】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときのユーザと電子デバイスとの間の向きを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、電子デバイスのディスプレイがユーザに向いているような向きであることに回答して尤度スコアを増加させることと、電子デバイスのディスプレイがユーザから外れた向きであることに回答して尤度スコアを減少させることを含むことができる。

30

【0129】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときのユーザの視線の方向を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、ユーザの視線の方向が電子デバイスを指していることに回答して尤度スコアを増加させることと、ユーザの視線の方向が電子デバイスから外れていることに回答して尤度スコアを減少させることを含むことができる。

40

【0130】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力自動音声認識装置により認識されたか否かを示す指標を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力自動音声認識装置により認識されたことを示す指標に回答して尤度スコアを増加させることと、第1のユーザ発話入力自動音声認識装置により認識されな

50

かったことを示す指標に回答して尤度スコアを減少させることとを含むことができる。

【0131】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と以前のユーザ発話入力との間の意味的關係を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、意味的關係の値がユーザ発話入力の意味の閾値より大きいことに回答して尤度スコアを増加させることと、意味的關係の値がユーザ発話入力の意味の閾値より小さいことに回答して尤度スコアを減少させることとを含むことができる。

10

【0132】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力の長さを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力の長さが長さの閾値より短いことに回答して尤度スコアを増加させることと、第1のユーザ発話入力の長さが長さの閾値より長いことに回答して尤度スコアを減少させることとを含むことができる。

【0133】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力の発話者のIDを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。いくつかの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力の発話者のIDが既知の発話者又は正規発話者のリストの中の1つであることに回答して尤度スコアを増加させることと、第1のユーザ発話入力の発話者のIDが既知の発話者又は正規発話者のリストの中の1つではないことに回答して尤度スコアを減少させることとを含むことができる。他の実施形態において、尤度スコアを計算することは、第1のユーザ発話入力の発話者のIDが以前のユーザ発話入力の発話者のIDと同一であることに回答して尤度スコアを増加させることと、第1のユーザ発話入力の発話者のIDが以前のユーザ発話入力の発話者のIDとは異なることに回答して尤度スコアを減少させることとを含むことができる。

20

30

【0134】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信された時刻を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度を計算することは、時刻が所定の1組の時刻の中にあることに回答して尤度スコアを増加させることと、時刻が所定の1組の時刻の中になくことに回答して尤度スコアを減少させることとを含むことができる。

40

【0135】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときに電子デバイスがユーザに対して情報を出力していたか否かを示す指標を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力を受信されたときに電子デバイスがユーザに対して情報を出力していたことを示す指標に回答して尤度スコアを増加させることと、第1のユーザ発話入力を受信されたときに電子デバイスがユーザに対して情報を出力していなかったことを示す指標に回答して尤度スコアを減少させることとを含むことができる。

50

できる。

【0136】

いくつかの実施形態において、処理ユニット508は、ユーザからの入力を受信する期待値を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、ユーザからの入力を受信する期待値がユーザからの入力の受信が期待されていたことを示すことに応答して尤度スコアを増加させることと、ユーザからの入力を受信する期待値がユーザからの入力の受信が期待されていなかったことを示すことに応答して尤度スコアを減少させることとを含むことができる。

10

【0137】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときに電子デバイスが保持されていたか否かを示す指標を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力を受信されたときに電子デバイスが保持されていたことを指標が示すことに応答して尤度スコアを増加させることと、第1のユーザ発話入力を受信されたときに電子デバイスが保持されていなかったことを指標が示すことに応答して尤度スコアを減少させることとを含むことができる。

20

【0138】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときの電子デバイスの動作状態を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、電子デバイスの動作状態が1組の所定の動作状態のうち1つであったことに応答して尤度スコアを増加させることと、電子デバイスの動作状態が1組の所定の動作状態のうち1つではなかったことに応答して尤度スコアを減少させることとを含むことができる。

【0139】

いくつかの実施形態において、処理ユニット508は、電子デバイスにより実行された以前のアクションを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、電子デバイスにより実行された以前のアクションが1組の所定のアクションのうち1つであることに応答して尤度スコアを増加させることと、電子デバイスにより実行された以前のアクションが1組の所定のアクションのうち1つではないことに応答して尤度スコアを減少させることとを含むことができる。

30

【0140】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信されたときに電子デバイスによりコンテンツが表示されていたか否かを示す指標を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力を受信されたときに電子デバイスによりコンテンツが表示されていたことを示す指標に回答して尤度スコアを増加させることとを含むことができる。

40

【0141】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と、第1

50

のユーザ発話入力を受信されたときに電子デバイスにより表示されていたコンテンツとの間の意味的関係を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に
10 応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に
応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これ
らの実施形態において、尤度スコアを計算することは、意味的関係の値がコンテンツの意味の
閾値より大きいことに応答して尤度スコアを増加させることと、意味的関係の値がコン
텐츠の意味の閾値より小さいことに応答して尤度スコアを減少させることとを含むこと
ができる。

【0142】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信され
10 ったときのユーザの位置を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発
話入力に
応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に
応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能であ
る。これらの実施形態において、尤度スコアを計算することは、ユーザの位置が所定の1
組の位置のうち1つであることに応答して尤度スコアを増加させることと、ユーザの位置
が所定の1組の位置の1つではないことに応答して尤度スコアを減少させることとを含むこと
ができる。

【0143】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信され
20 ったときにユーザにより実行されていたジェスチャーを含む文脈情報に基づいて、仮想ア
シスタントが第1のユーザ発話入力に
応答すべき尤度スコアを計算することにより、第1
のユーザ発話入力に
応答すべきか否かを判定する（例えば、応答判定ユニット514を使
用して）ように構成可能である。これらの実施形態において、尤度スコアを計算すること
は、ジェスチャーが所定の1組のジェスチャーのうち1つであることに応答して尤度スコ
アを増加させることと、ジェスチャーが所定の1組のジェスチャーのうち1つではないこ
とに応答して尤度スコアを減少させることとを含むことができる。

【0144】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と電子デ
30 バイスの以前の出力との間の意味的関係を含む文脈情報に基づいて、仮想アシスタントが
第1のユーザ発話入力に
応答すべき尤度スコアを計算することにより、第1のユーザ発話
入力に
応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）よう
に構成可能である。これらの実施形態において、尤度スコアを計算することは、意味的関
係の値が以前の出力の意味の閾値より大きいことに応答して尤度スコアを増加させること
と、意味的関係の値が以前の出力の意味の閾値より小さいことに応答して尤度スコアを減
少させることとを含むことができる。

【0145】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力を受信され
40 ったときの電子デバイスの場所を含む文脈情報に基づいて、仮想アシスタントが第1のユ
ーザ発話入力に
応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に
応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可
能である。これらの実施形態において、尤度スコアを計算することは、場所が所定の1組
の場所のうち1つであることに応答して尤度スコアを減少させることと、場所が所定の1
組の場所のうち1つではないことに応答して尤度スコアを増加させることとを含むことが
できる。

【0146】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と、第1
50 のユーザ発話入力を受信されたときに電子デバイスにより実行されていたアプリケーション
との間の意味的関係を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話
入力に
応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に
応答すべき
か否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である

。これらの実施形態において、尤度スコアを計算することは、意味的關係の値がアプリケーションの意味の閾値より大きいことに応答して尤度スコアを増加させることと、意味的關係の値がアプリケーションの意味の閾値より小さいことに応答して尤度スコアを減少させることとを含むことができる。

【 0 1 4 7 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と以前の連絡との間の意味的關係を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、意味的關係の値が以前の連絡の意味の閾値より大きいことに応答して尤度スコアを増加させることと、意味的關係の値が以前の連絡の意味の閾値より小さいことに応答して尤度スコアを減少させることとを含むことができる。

10

【 0 1 4 8 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と以前のEメールとの間の意味的關係を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、意味的關係の値が以前のEメールの意味の閾値より大きいことに応答して尤度スコアを増加させることと、意味的關係の値が以前のEメールの意味の閾値より小さいことに応答して尤度スコアを減少させることとを含むことができる。

20

【 0 1 4 9 】

いくつかの実施形態において、処理ユニット508は、第1のユーザ発話入力と以前のSMSメッセージとの間の意味的關係を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、意味的關係の値が以前のSMSメッセージの意味の閾値より大きいことに応答して尤度スコアを増加させることと、意味的關係の値が以前のSMSメッセージの意味の閾値より小さいことに応答して尤度スコアを減少させることとを含むことができる。

30

【 0 1 5 0 】

いくつかの実施形態において、処理ユニット508は、電子デバイスの動きを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、動きが所定の1組の動きのうち1つであることに応答して尤度スコアを増加させることと、動きが所定の1組の動きのうち1つではないことに応答して尤度スコアを減少させることとを含むことができる。

【 0 1 5 1 】

いくつかの実施形態において、処理ユニット508は、ユーザ設定を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答すべき尤度スコアを計算することにより、第1のユーザ発話入力に応答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、ユーザ設定が所定の1組のユーザ設定のうち1つであることに応答して尤度スコアを増加させることと、ユーザ設定が所定の1組のユーザ設定のうち1つではないことに応答して尤度スコアを減少させることとを含むことができる。

40

【 0 1 5 2 】

いくつかの実施形態において、処理ユニット508は、電子デバイスにより感知された光の量を含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に応答すべ

50

き尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、光の量が光の閾値量より多いことに応答して尤度スコアを増加させることと、光の量が光の閾値量より少ないことに応答して尤度スコアを減少させることとを含むことができる。

【0153】

いくつかの実施形態において、処理ユニット508は、カレンダーデータを含む文脈情報に基づいて、仮想アシスタントが第1のユーザ発話入力に回答すべき尤度スコアを計算することにより、第1のユーザ発話入力に回答すべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。これらの実施形態において、尤度スコアを計算することは、第1のユーザ発話入力を受信された時刻にユーザが仕事であったことを示すカレンダーデータに回答して尤度スコアを減少させることとを含むことができる。

10

【0154】

処理ユニット508は、オーディオ入力中の第2のユーザ発話入力を識別するためにオーディオ入力を監視する（例えば、第2の監視ユニット518を使用して）ように更に構成可能である。処理ユニット508は、オーディオ入力中の第2のユーザ発話入力を識別した場合、第2のユーザ発話入力に関連する文脈情報に基づいて、第2のユーザ発話入力に対する応答が生成されるべきか否かを判定する（例えば、応答判定ユニット514を使用して）ように構成可能である。処理ユニット508は、応答が生成されるとの判定に

20

【0155】

以上説明したように、本発明の1つの態様は、ユーザの関心を引くと思われる魅力あるコンテンツ又は他の何らかのコンテンツのユーザへの配信を改善するために種々のデータ源から利用可能なデータを収集すること及び利用することである。本発明は、収集されるデータは、特定の個人を一意に識別するか、あるいは特定の個人と連絡するか又は居場所を特定するために使用可能である個人情報データを含む場合もあると考える。そのような個人情報データは、デモグラフィックデータ、場所依存データ、電話番号、Eメールアドレス、自宅の住所又は他の何らかの識別情報を含む可能性がある。

30

【0156】

本発明は、本発明の技術によれば、そのような個人情報をユーザの利益になるように使用できると認識する。例えば、ユーザがより興味を引かれるように絞り込んだコンテンツを配信するために個人情報データを利用できる。従って、個人情報データの利用によって、計画的に配信コンテンツを制御することが可能になる。更に、ユーザの利益になる他の個人情報データの利用方法も本発明により意図されている。

40

【0157】

本発明は、そのような個人情報データの収集、解析、開示、転送、記憶又は他の方法による利用の責任を負うエンティティが確立したプライバシーポリシー及び/又はプライバシー慣行に従うことを更に意図する。詳細には、そのようなエンティティは、個人情報データのプライバシー及び秘密を保護するための産業要件又は政府要件に適合するか又はそれを超えると一般に認識されているプライバシーポリシー及びプライバシー慣行を実現し且つ矛盾なく使用すべきである。例えば、ユーザからの個人情報は、エンティティが合法的且つ正当に利用するために収集されるべきであり、合法的利用の範囲外で共有又は販売されるべきではない。更に、データの収集は、ユーザのインフォームドコンセントを受け

50

取った後に初めて実行されるべきである。また、エンティティは、そのような個人情報データの安全を保護し且つ個人情報データに対するアクセスを安全に保護すると共に、個人情報データにアクセスする他者がエンティティのプライバシーポリシー及びプライバシー手順を確実に厳守するように保証するために必要とされるあらゆる措置を講じるだろう。更に、広く受け入れられているプライバシーポリシー及びプライバシー慣行をエンティティが厳守していることを証明するために、エンティティは第三者による評価を受けることができる。

【0158】

以上の説明に関わらず、本発明は、ユーザが個人情報データの利用又はアクセスを選択的に阻止する実施例を更に意図する。すなわち、本発明は、個人情報データへのアクセスを防止又は阻止するためのハードウェア要素及び/又はソフトウェア要素を提供できると考える。例えば、広告配信サービスの場合、本発明は、サービスへの登録中に個人情報データの収集への参加にユーザが「オプトイン」すること又は「オプトアウト」することを選択できるように構成可能である。別の例では、ターゲットを絞り込んだコンテンツ配信サービスの場合、ユーザは場所情報を提供しないことを選択できる。更に別の例では、ユーザは厳密な場所情報を提供しないが、場所ゾーン情報の転送は許可することを選択できる。

10

【0159】

従って、本発明は、1つ以上の種々の開示された実施形態を実現するために広い範囲で個人情報データの利用に対応できるが、そのような個人情報データをアクセスする必要なく種々の実施形態が実現できると考える。すなわち、本発明の種々の実施形態は、個人情報データのすべて又は一部が欠落していても動作不能にはならない。例えば、ユーザに関連するデバイスにより要求されているコンテンツ、コンテンツ配信サービスで利用可能な他の非個人情報データ又は公開で利用可能な情報など、非個人情報データに基づいて又はごく最小限の量の個人情報に基づいて好みを推測することにより、コンテンツが選択され、ユーザに配信可能である。

20

【0160】

添付の図面を参照して実施形態を十分に説明したが、種々の変形及び変更は当業者には明らかだろうということに注意すべきである。そのような変形及び変更は、添付の特許請求の範囲により定義される種々の実施形態の範囲内に含まれると理解されるべきである。

30

【図 1】

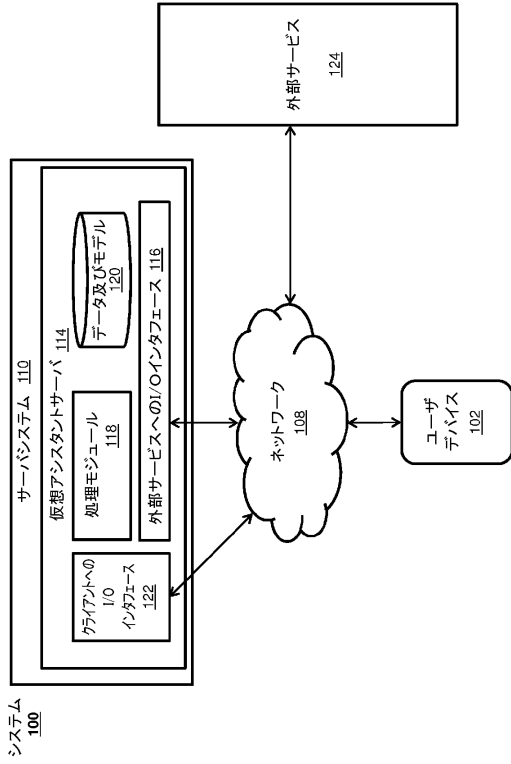


FIG. 1

【図 2】

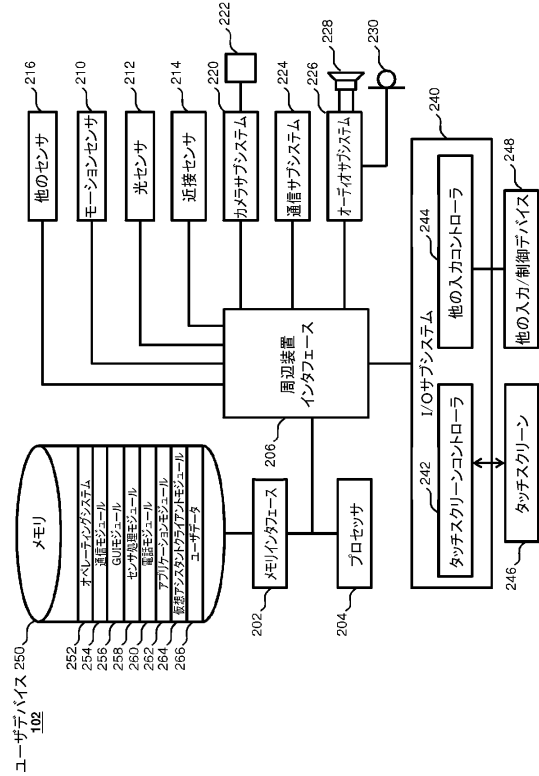


FIG. 2

【図 3】

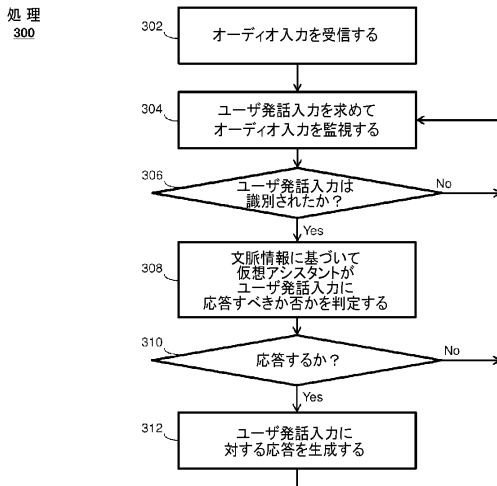


FIG. 3

【図 4】

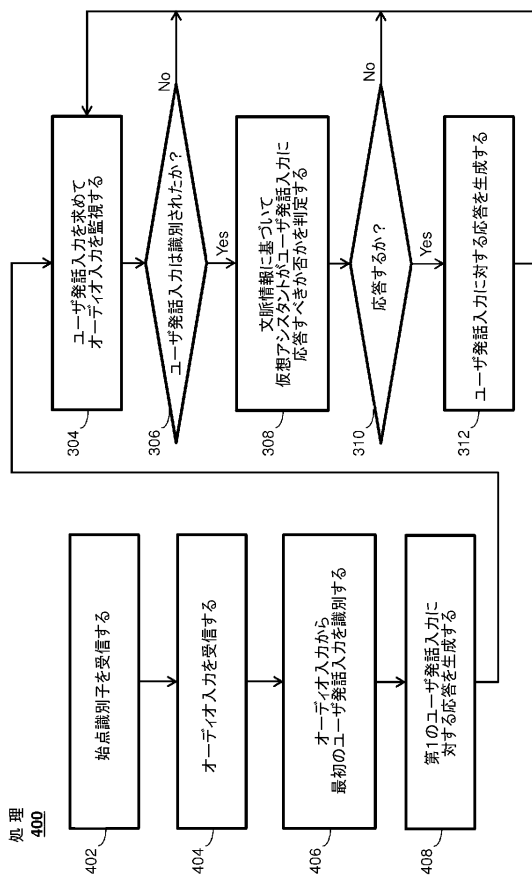
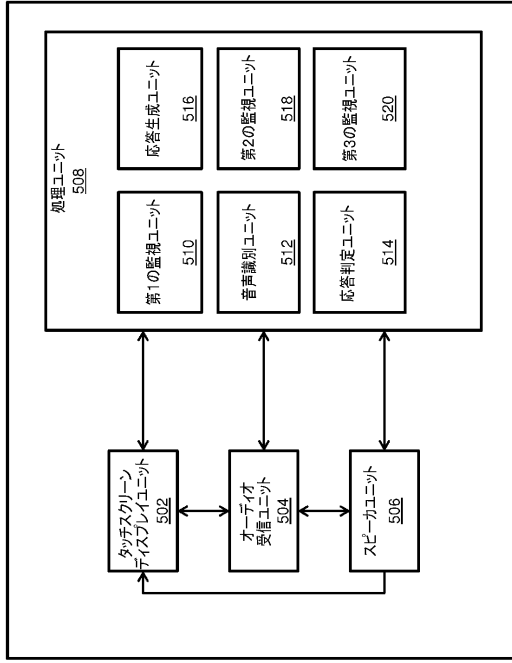


FIG. 4

【 図 5 】



電子デバイス
500

FIG. 5

フロントページの続き

- (74)代理人 100116894
弁理士 木村 秀二
- (74)代理人 100130409
弁理士 下山 治
- (74)代理人 100134175
弁理士 永川 行光
- (72)発明者 ピエノット, フィリップ ピー.
アメリカ合衆国 カリフォルニア州 95014, クパチーノ, インフィニット ループ 1
- (72)発明者 ビンダー, ジャスティン ジー.
アメリカ合衆国 カリフォルニア州 95014, クパチーノ, インフィニット ループ 1

審査官 上田 雄

- (56)参考文献 特開2014-077969(JP,A)
特開2008-033198(JP,A)
特開2003-308079(JP,A)
特開2003-202897(JP,A)
米国特許出願公開第2013/0144616(US,A1)
特開2002-123295(JP,A)
米国特許出願公開第2012/0034904(US,A1)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/34