

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7286772号
(P7286772)

(45)発行日 令和5年6月5日(2023.6.5)

(24)登録日 令和5年5月26日(2023.5.26)

| | | | | |
|-------------------------|---------|-------|---------|--|
| (51)国際特許分類 | F I | | | |
| G 1 0 L 15/22 (2006.01) | G 1 0 L | 15/22 | 2 0 0 H | |
| G 0 6 F 3/16 (2006.01) | G 1 0 L | 15/22 | 3 0 0 Z | |
| | G 0 6 F | 3/16 | 6 2 0 | |
| | G 0 6 F | 3/16 | 6 5 0 | |

請求項の数 26 (全40頁)

| | | | |
|-------------------|-------------------------------|----------|---|
| (21)出願番号 | 特願2021-537862(P2021-537862) | (73)特許権者 | 502208397 |
| (86)(22)出願日 | 平成31年2月7日(2019.2.7) | | グーグル エルエルシー |
| (65)公表番号 | 特表2022-516101(P2022-516101 A) | | Google LLC |
| (43)公表日 | 令和4年2月24日(2022.2.24) | | アメリカ合衆国 カリフォルニア州 94043 マウンテン ビュー アンフィシアター パークウェイ 1600 |
| (86)国際出願番号 | PCT/US2019/017043 | | 1600 Amphitheatre P |
| (87)国際公開番号 | WO2020/139408 | | arkway 94043 Mounta |
| (87)国際公開日 | 令和2年7月2日(2020.7.2) | | in View, CA U.S.A. |
| 審査請求日 | 令和3年8月24日(2021.8.24) | (74)代理人 | 100108453 |
| (31)優先権主張番号 | 62/785,842 | | 弁理士 村山 靖彦 |
| (32)優先日 | 平成30年12月28日(2018.12.28) | (74)代理人 | 100110364 |
| (33)優先権主張国・地域又は機関 | 米国(US) | | 弁理士 実広 信哉 |
| | | (74)代理人 | 100133400 |
| | | | 弁理士 阿部 達彦 |

最終頁に続く

(54)【発明の名称】 選択されたサジェスションによる自動アシスタントへのボイス入力への補足

(57)【特許請求の範囲】

【請求項1】

1つまたは複数のプロセッサによって実施される方法であって、ユーザによって提供された口頭発話を特徴づけるデータに対して、音声テキスト処理を実施するステップであって、第1の候補テキストセグメントおよび第2の候補テキストセグメントを決定するステップを含み、前記第1の候補テキストセグメントおよび前記第2の候補テキストセグメントが、前記口頭発話の異なる解釈に対応し、

前記口頭発話が、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される、ステップと、

前記口頭発話を特徴づける前記データに対して、前記音声テキスト処理を実施するステップに基づいて、前記口頭発話が完全であるか否かを決定するステップであって、

少なくとも、自動アシスタントが、前記自然言語コンテンツに基づいて、1つまたは複数のアクションが達成されることを引き起こすことができるか否かを決定するステップを含む、ステップと、

前記口頭発話が不完全であると決定されるとき、

前記口頭発話が不完全であるとの決定に回答して、前記コンピューティングデバイスの前記ディスプレイパネルが、1つまたは複数のサジェスション要素を提供することを引き起こすステップであって、

前記1つまたは複数のサジェスチョン要素が、前記ユーザによって前記自動アシスタントインターフェースに対して話されるとき、前記自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、前記ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含み、前記特定のサジェスチョン要素が、前記第1の候補テキストセグメントに基づいて決定され、少なくとも1つの他のサジェスチョン要素が、前記第2の候補テキストセグメントに基づいて決定される、ステップと、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数のサジェスチョン要素を提供することを引き起こすステップの後に、前記ユーザが、前記特定のサジェスチョン要素の前記他の自然言語コンテンツに関連付けられる、別の口頭発話を提供したと決定するステップと、

10

前記ユーザが前記別の口頭発話を提供したとの決定に回答して、前記口頭発話および前記別の口頭発話の組合せが完全であるか否かを決定するステップと、

前記口頭発話および前記別の口頭発話の前記組合せが完全であると決定されるとき、
前記1つまたは複数のアクションが、前記自然言語コンテンツおよび前記別の口頭発話に基づいて、前記自動アシスタントを介して実施されることを引き起こすステップとを含む方法。

【請求項2】

1つまたは複数のプロセッサによって実施される方法であって、ユーザによって提供された口頭発話を特徴づけるデータに対して、音声テキスト処理を実施するステップであって、

20

前記データから、1つまたは複数の候補テキストセグメントを生成するステップを含み、前記口頭発話が、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される、ステップと、

前記口頭発話を特徴づける前記データに対して、前記音声テキスト処理を実施するステップに基づいて、前記口頭発話が完全であるか否かを決定するステップであって、少なくとも、自動アシスタントが、前記自然言語コンテンツに基づいて、1つまたは複数のアクションが達成されることを引き起こすことができるか否かを決定するステップを含む、ステップと、

30

前記口頭発話が不完全であると決定されるとき、前記口頭発話が不完全であるとの決定に回答して、前記コンピューティングデバイスの前記ディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすステップであって、

前記1つまたは複数のサジェスチョン要素が、前記ユーザによって前記自動アシスタントインターフェースに対して話されるとき、前記自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、前記ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含む、ステップと、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数のサジェスチョン要素を提供することを引き起こすステップの後に、前記ユーザが、前記特定のサジェスチョン要素の前記他の自然言語コンテンツに関連付けられる、別の口頭発話を提供したと決定するステップと、

40

前記ユーザが前記別の口頭発話を提供したとの決定に回答して、前記口頭発話および前記別の口頭発話の組合せが完全であるか否かを決定するステップと、

前記口頭発話および前記別の口頭発話の前記組合せが完全であると決定されるとき、前記1つまたは複数のアクションが、前記自然言語コンテンツおよび前記別の口頭発話に基づいて、前記自動アシスタントを介して実施されることを引き起こすステップと、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数の候補テキストセグメントのうちの少なくとも1つの候補テキストセグメントのグラフィカル表現を提供することを引き起こすステップと、

50

前記口頭発話が完全であると決定されるとき、

1人または複数のユーザによって前記自動アシスタントインターフェースに対して話されるとき、前記ユーザからの前記口頭発話に関連付けられた音声テキスト処理の量と比較して、低減された量の音声テキスト処理を生じ、前記自動アシスタントが前記アクションを完成させることを促進するために動作することを引き起こす、自然言語コマンドを識別するステップと、

前記自然言語コマンドを識別するステップに基づいて、前記自然言語コマンドを特徴づける、1つまたは複数の他のサジェスチョン要素を生成するステップと、

前記口頭発話が完全であるとの決定に少なくとも応答して、前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数の他のサジェスチョン要素を提供することを引き起こすステップと

10

を含む方法。

【請求項3】

前記ユーザによって提供された前記口頭発話の特徴づける前記データに対して、音声テキスト処理を実施するステップが、前記データから、1つまたは複数の候補テキストセグメントを生成するステップを含み、前記方法が、

前記口頭発話が不完全であると決定されるとき、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数の候補テキストセグメントのうちの少なくとも1つの候補テキストセグメントのグラフィカル表現を提供することを引き起こすステップ

20

をさらに含む、請求項1に記載の方法。

【請求項4】

前記口頭発話が、完全であり、前記1つまたは複数のアクションのためのすべての必須パラメータを含むと決定されるとき、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数の候補テキストセグメントのうちの前記少なくとも1つの候補テキストセグメントの前記グラフィカル表現を提供することをバイパスすること、および、前記1つまたは複数のサジェスチョン要素を提供することをバイパスすることを引き起こすステップ

をさらに含む、請求項2または3に記載の方法。

【請求項5】

30

前記口頭発話が完全であると決定されるとき、

1人または複数のユーザによって前記自動アシスタントインターフェースに対して話されるとき、前記ユーザからの前記口頭発話に関連付けられた音声テキスト処理の量と比較して、低減された量の音声テキスト処理を生じ、前記自動アシスタントが前記アクションを完成させることを促進するために動作することを引き起こす、自然言語コマンドを識別するステップと、

前記自然言語コマンドを識別するステップに基づいて、前記自然言語コマンドを特徴づける、1つまたは複数の他のサジェスチョン要素を生成するステップと、

前記口頭発話が完全であるとの決定に少なくとも応答して、前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数の他のサジェスチョン要素を提供することを引き起こすステップと

40

をさらに含む、請求項3または3に従属する4に記載の方法。

【請求項6】

前記口頭発話が不完全であると決定されるとき、

少なくとも1つのサジェスチョン要素が、前記ディスプレイパネルを介して提供されるとき、前記ユーザから期待される口頭発話のタイプに基づいて、複数の異なる音声テキスト処理モデルから、音声テキスト処理モデルを選択するステップ

をさらに含む、請求項1から5のいずれか一項に記載の方法。

【請求項7】

前記口頭発話が不完全であると決定されるとき、

50

前記1つまたは複数のサジェスチョン要素の1つまたは複数の用語の方に向けて、および/あるいは、前記1つまたは複数のサジェスチョン要素に対応するコンテンツの1つまたは複数の予想されるタイプの方に向けて、音声テキスト処理にバイアスをかけるステップをさらに含む、請求項1から6のいずれか一項に記載の方法。

【請求項8】

口頭入力無音のしきい値持続時間が、前記口頭発話に後続したか否かを決定するステップであって、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数のサジェスチョン要素を提供することを引き起こすステップが、少なくとも前記口頭入力無音のしきい値持続時間が前記口頭発話に後続したとの決定にさらに応答するものである、ステップ

10

をさらに含む、請求項1から7のいずれか一項に記載の方法。

【請求項9】

前記コンピューティングデバイスに接続された前記ディスプレイパネルが、前記1つまたは複数のサジェスチョン要素を提供することを引き起こすステップが、

前記コンピューティングデバイスに接続された前記ディスプレイパネルが、前記特定のサジェスチョン要素に隣接して、前記第1の候補テキストセグメントをグラフィカルに表現し、前記少なくとも1つの他のサジェスチョン要素に隣接して、前記第2の候補テキストセグメントをグラフィカルに表現することを引き起こすステップを含む、請求項1に記載の方法。

20

【請求項10】

前記ユーザが、前記特定のサジェスチョン要素の前記他の自然言語コンテンツに関連付けられる、前記別の口頭発話を提供したと決定するステップが、

前記別の口頭発話に基づいて、前記ユーザが前記第1の候補テキストセグメントを識別したか、前記第2の候補テキストセグメントを識別したかを決定するステップを含む、請求項9に記載の方法。

【請求項11】

前記口頭発話が不完全であると決定されるとき、

前記口頭発話が不完全であるとの決定に応答して、前のアクションが、前記自動アシスタントを介して達成されており、前記ユーザが、前記ユーザと前記自動アシスタントとの間の1つまたは複数の前の対話の間に、前記他の自然言語コンテンツの少なくとも一部分を識別した、前記1つまたは複数の前の対話の特徴づける履歴データに基づいて、前記他の自然言語コンテンツを生成するステップ

30

をさらに含む、請求項1から10のいずれか一項に記載の方法。

【請求項12】

前記口頭発話が不完全であると決定されるとき、

前記口頭発話が不完全であるとの決定に応答して、前記ユーザに関連付けられた様々なデバイス間の関係の特徴づけるデバイスポートロギーデータに基づいて、前記他の自然言語コンテンツを生成するステップであって、

前記特定のサジェスチョン要素が、前記デバイスポートロギーデータに基づくものであり、前記ユーザに関連付けられた前記様々なデバイスのうちの1つまたは複数のデバイスを識別する、ステップ

40

をさらに含む、請求項1から10のいずれか一項に記載の方法。

【請求項13】

前記1つまたは複数のアクションが、前記自然言語コンテンツおよび前記別の口頭発話に基づいて、前記自動アシスタントを介して実施されることを引き起こすステップが、デバイスを制御するステップを含む、請求項1から12のいずれか一項に記載の方法。

【請求項14】

前記特定のサジェスチョン要素が、前記アクションを示すグラフィカル要素をさらに備える、請求項1から13のいずれか一項に記載の方法。

50

【請求項 15】

前記口頭発話が不完全であると決定されるとき、

前記1つまたは複数のサジェスチョン要素のプロビジョニング後、前記別の口頭発話を待つための特定の持続時間を決定するステップであって、前記特定の持続時間が、前記口頭発話が不完全であると決定されることに基づいて決定される、ステップをさらに含む、請求項1から13のいずれか一項に記載の方法。

【請求項 16】

前記口頭発話が完全であると決定されるとき、

前記完全な口頭発話に基づく、1つまたは複数の他のサジェスチョン要素を生成するステップと、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数の他のサジェスチョン要素を提供することを引き起こすステップと、

前記1つまたは複数の他のサジェスチョン要素のプロビジョニング後、さらなる口頭発話を待つための代替の特定の持続時間を決定するステップであって、前記代替の特定の持続時間が、前記特定の持続時間よりも短く、前記代替の特定の持続時間が、前記口頭発話が完全であると決定されることに基づいて決定される、ステップとをさらに含む、請求項15に記載の方法。

【請求項 17】

1つまたは複数のプロセッサによって実施される方法であって、

自動アシスタントがアクションを実施することを引き起こすことを促進するために、ユーザによって提供された口頭発話を特徴づけるデータに対して、音声テキスト処理を実施するステップであって、

前記口頭発話が、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される、ステップと、

前記口頭発話を特徴づける前記データに対して、前記音声テキスト処理を実施するステップに基づいて、前記口頭発話が完全であるか否かを決定するステップであって、

少なくとも、前記自然言語コンテンツに、前記アクションに関連付けられた機能を制御するための1つまたは複数のパラメータ値がないか否かを決定するステップを含む、ステップと、

前記口頭発話が不完全であると決定されるとき、

前記口頭発話が不完全であるとの決定に応答して、前記コンピューティングデバイスに接続された前記ディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすステップであって、

前記1つまたは複数のサジェスチョン要素が、前記ユーザによって前記自動アシスタントインターフェースに対して話されるとき、前記自動アシスタントが前記アクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、前記ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含む、ステップと、

前記ユーザが、前記自動アシスタントインターフェースにおいて受信された別の口頭発話を介して、前記1つまたは複数のサジェスチョン要素のうちの前記特定のサジェスチョン要素を選択したと決定するステップと、

前記ユーザが前記特定のサジェスチョン要素を選択したとの決定に応答して、前記アクションが、前記特定のサジェスチョン要素の前記他の自然言語コンテンツと、前記口頭発話の前記自然言語コンテンツとに基づいて実施されることを引き起こすステップと、前記ユーザが前記特定のサジェスチョン要素を選択したとの決定に応答して、前記口頭発話が不完全であると決定される前に前記他の自然言語コンテンツに関連付けられた前の優先度から、前記他の自然言語コンテンツに関連付けられた優先度が修正されることを引き起こすステップと、

前記口頭発話が完全であると決定されるとき、

10

20

30

40

50

前記口頭発話が完全であるとの決定に回答して、前記アクションが、前記口頭発話の前記自然言語コンテンツに基づいて実施されることを引き起こすステップとを含む方法。

【請求項 18】

1つまたは複数のプロセッサによって実施される方法であって、

自動アシスタントがアクションを実施することを引き起こすことを促進するために、ユーザによって提供された口頭発話の特徴づけるデータに対して、音声テキスト処理を実施するステップであって、

第1の候補テキストセグメントおよび第2の候補テキストセグメントを決定するステップを含み、前記第1の候補テキストセグメントおよび前記第2の候補テキストセグメントが、前記口頭発話の異なる解釈に対応し、

10

前記口頭発話が、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される、ステップと、

前記口頭発話の特徴づける前記データに対して、前記音声テキスト処理を実施するステップに基づいて、前記口頭発話が完全であるか否かを決定するステップであって、

少なくとも、前記自然言語コンテンツに、前記アクションに関連付けられた機能を制御するための1つまたは複数のパラメータ値がないか否かを決定するステップを含む、ステップと、

前記口頭発話が不完全であると決定されるとき、

20

前記口頭発話が不完全であるとの決定に回答して、前記コンピューティングデバイスに接続された前記ディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすステップであって、

前記1つまたは複数のサジェスチョン要素が、前記ユーザによって前記自動アシスタントインターフェースに対して話されるとき、前記自動アシスタントが前記アクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、前記ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含み、前記1つまたは複数のサジェスチョン要素のうちの少なくとも1つのサジェスチョン要素が、前記第1の候補テキストセグメントに基づいて決定され、少なくとも1つの他のサジェスチョン要素が、前記第2の候補テキストセグメントに基づいて決定される、ステップと、

30

前記ユーザが、前記自動アシスタントインターフェースにおいて受信された別の口頭発話を介して、前記1つまたは複数のサジェスチョン要素のうちの前記特定のサジェスチョン要素を選択したと決定するステップと、

前記ユーザが前記特定のサジェスチョン要素を選択したとの決定に回答して、前記アクションが、前記特定のサジェスチョン要素の前記他の自然言語コンテンツと、前記口頭発話の前記自然言語コンテンツとに基づいて実施されることを引き起こすステップと、

前記口頭発話が完全であると決定されるとき、

前記口頭発話が完全であるとの決定に回答して、前記アクションが、前記口頭発話の前記自然言語コンテンツに基づいて実施されることを引き起こすステップとを含む方法。

40

【請求項 19】

前記ディスプレイパネルにおける、少なくとも1つの他のサジェスチョン要素に対する、前記特定のサジェスチョン要素の提示の順序が、前記他の自然言語コンテンツへの割り当てられた優先度に少なくとも部分的に基づく、請求項17または18に記載の方法。

【請求項 20】

口頭入力無音のしきい値持続時間が、前記口頭発話に後続したか否かを決定するステップであって、

前記コンピューティングデバイスの前記ディスプレイパネルが、前記1つまたは複数のサジェスチョン要素を提供することを引き起こすステップが、少なくとも口頭入力無音の前記しきい値持続時間が前記口頭発話に後続したとの決定にさらに応答するものである

50

、ステップ

をさらに含む、請求項17から19のいずれか一項に記載の方法。

【請求項21】

前記口頭発話が不完全であると決定される時、

前記口頭発話が不完全であるとの決定に回答して、前のアクションが、前記自動アシスタントを介して達成されており、前記ユーザが、前記ユーザと前記自動アシスタントとの間の1つまたは複数の前の対話の間に、前記他の自然言語コンテンツの少なくとも一部分を識別した、前記1つまたは複数の前の対話の特徴づける履歴データに基づいて、前記他の自然言語コンテンツを生成するステップ

をさらに含む、請求項17から20のいずれか一項に記載の方法。

10

【請求項22】

前記口頭発話が不完全であると決定される時、

前記口頭発話が不完全であったとの決定に回答して、前記ユーザに関連付けられた様々なデバイス間の関係の特徴づけるデバイスポートロジータに基づいて、前記他の自然言語コンテンツを生成するステップであって、

前記特定のサジェスション要素が、前記デバイスポートロジータに基づくものであり、前記ユーザに関連付けられた前記様々なデバイスのうちの1つまたは複数のデバイスを識別する、ステップ

をさらに含む、請求項17から20のいずれか一項に記載の方法。

【請求項23】

前記コンピューティングデバイスに接続された前記ディスプレイパネルが、1つまたは複数のサジェスション要素を提供することを引き起こすステップが、

前記コンピューティングデバイスに接続された前記ディスプレイパネルが、前記少なくとも1つのサジェスション要素に隣接して、前記第1の候補テキストセグメントをグラフィカルに表現し、前記少なくとも1つの他のサジェスション要素に隣接して、前記第2の候補テキストセグメントをグラフィカルに表現することを引き起こすステップを含む、請求項18に記載の方法。

20

【請求項24】

1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサが請求項1から23のいずれか一項に記載の方法を行うことを引き起こす命令を備える、コンピュータプログラム。

30

【請求項25】

1つまたは複数のプロセッサによって実行されると、前記1つまたは複数のプロセッサが請求項1から23のいずれか一項に記載の方法を行うことを引き起こす命令を備える、コンピュータ可読記憶媒体。

【請求項26】

請求項1から23のいずれか一項に記載の方法を行うための1つまたは複数のプロセッサを備える、システム。

【発明の詳細な説明】

【背景技術】

40

【0001】

人間は、本明細書では「自動アシスタント」と呼ぶ(「デジタルエージェント」、「チャットボット」、「対話型パーソナルアシスタント」、「インテリジェントパーソナルアシスタント」、「アシスタントアプリケーション」、「会話エージェント」などとも呼ぶ)、対話型ソフトウェアアプリケーションとのヒューマンコンピュータダイアログに携わることがある。たとえば、人間(自動アシスタントと対話するとき、「ユーザ」と呼ぶことがある)は、場合によってはテキストに変換され、次いで処理され得る、口頭自然言語入力(すなわち、発話)を使用して、および/またはテキストの(たとえば、タイピングされた)自然言語入力を提供することによって、コマンドおよび/または要求を自動アシスタントに提供することがある。自動アシスタントは、可聴および/または視覚的ユーザインターフェース

50

出力を含み得る、応答ユーザインターフェース出力を提供することによって、要求に応答する。したがって、自動アシスタントは、ボイスベースのユーザインターフェースを提供し得る。

【0002】

いくつかの事例では、ユーザは、自動アシスタントアクションの実施を引き起こすように、ユーザによって意図されるが、意図された自動アシスタントアクションの実施を生じない、口頭発話(spoken utterance)を提供することがある。たとえば、口頭発話は、自動アシスタントによって理解可能ではない構文において提供されることがあり、かつ/または、自動アシスタントアクションのための必須パラメータを欠くことがある。結果として、自動アシスタントは、口頭発話を完全に処理すること、および、それが自動アシスタントアクションのための要求であると決定することができないことがある。このことは、自動アシスタントが口頭発話に対する応答を提供しないこと、または、「すみません、その役には立てません」などの応答におけるエラー、および/もしくはエラー音を提供することにつながる。自動アシスタントが、口頭発話の意図されたアクションを実施することに失敗するにもかかわらず、様々なコンピュータおよび/またはネットワークリソースが、それにもかかわらず、口頭発話の処理、および適切なアクションを解決しようとする試みにおいて消費される。たとえば、口頭発話に対応するオーディオデータは、送信される、音声テキスト処理を受ける、かつ/または自然言語処理を受けることがあり得る。リソースのそのような消費は、意図された自動アシスタントアクションが実施されないので、無駄であり、ユーザは、意図された自動アシスタントアクションの実施を後で求めるとき、再構築された口頭発話を提供するように試みる可能性が高くなる。さらに、そのような再構築された口頭発話もまた、処理されなければならないことになる。さらに、これによって、ユーザが代わりに好適な口頭発話を自動アシスタントに最初に提供した場合と比較して、自動アシスタントアクションの実施においてレイテンシが生じる。

【発明の概要】

【課題を解決するための手段】

【0003】

本明細書で説明する実装形態は、自動アシスタントに向けられた口頭発話を完成させるためのサジェスチョンを、表示モダリティを介して提供すること、および場合によっては、口頭発話が進行中である間に、様々なファクタに基づいて、更新されたサジェスチョンを継続的に提供することに関する。たとえば、サジェスチョンは、自動アシスタントが、ユーザによって最初に意図された自動アシスタントアクションを総合的に実施することを引き起こすようになる、実行可能なダイアログフレーズを開発すると同時に、ユーザと自動アシスタントとの間の対話の合理化も行うことを促進するために提供され得る。追加または代替として、サジェスチョンは、ユーザが自動アシスタントとの現在のダイアログセッション、および/または後続のダイアログセッションに参加することになる、頻度および/または時間の長さを低減することができる、アクションを実施するために、自動アシスタントを採用するようにユーザに勧めることを促進するために提供され得る。

【0004】

いくつかの実装形態では、ユーザと自動アシスタントとの間のダイアログセッションの間に、ユーザは、自動アシスタントが1つまたは複数のアクションを実施することを引き起こすことを促進するために、口頭発話を提供することができる。口頭発話は、オーディオデータと、各々が口頭発話の対応する解釈である1つまたは複数のテキストセグメントを生成するために、オーディオデータ上で実施された音声テキスト処理とによって取り込まれ得る。1つまたは複数のテキストセグメントは、口頭発話が完全な要求および/またはアクション可能な要求に対応するか否かを決定するために、処理され得る。本明細書で使用する「完全な要求」または「アクション可能な要求」は、自動アシスタントによって完全に処理されると、1つまたは複数の対応する自動アシスタントアクションの実施を引き起こすものであり、ここで、実施された自動アシスタントアクションは、デフォルトエラータイプアクションではない(たとえば、「すみません、その役には立てません」という可

10

20

30

40

50

聴応答ではない)。たとえば、アクション可能な要求は、1つまたは複数のスマートデバイスが制御されることを引き起こすことができ、特定の可聴および/またはグラフィカルコンテンツがレンダリングされることを引き起こすことができる、などである。口頭発話が不完全な要求に対応すると決定されるとき、1つまたは複数のテキストセグメントは、要求を完成させるための1つまたは複数のサジェスチョンを決定および提供するために、さらに処理され得る。1つまたは複数のサジェスチョンのうちの各サジェスチョンは、1つまたは複数のテキストセグメントのうちの対応するものと組み合わせられると、完全な要求を提供する、追加のテキストを含み得る。言い換えれば、ユーザが、特定のサジェスチョンのテキスト(または、特定のサジェスチョンに基づくテキストに概して適合する代替テキスト)を復唱することによって、自分のダイアログセッションを継続した場合、自動アシスタントは、対応する自動アシスタントアクションを実施することになる。したがって、サジェスチョンの生成は、タスクを実施することにおいて、ユーザを支援する。ユーザが要求を入力中である間、ユーザには、アシスタントとの対話においてユーザをガイドする情報が提供される。ユーザに提示される情報は、ユーザの既存の入力に基づくものであり、したがって、ユーザに情報を提供するために、客観分析を使用する。したがって、ユーザには、タスクを行うことにおいてユーザを支援するために、客観的に関連のある情報が提供される。

【0005】

いくつかの実装形態では、コンピューティングデバイスは、口頭発話の異なる解釈を生成することができ、それらの解釈における変形形態は、口頭発話に回答して提供されるサジェスチョンへの対応する変形形態を生じることができる。たとえば、「アシスタント、変更してもらえますか...」などの口頭発話は、コンピューティングデバイスによる複数の解釈に対応し得る。上述の口頭発話は、自動アシスタントを介してIoTデバイス設定を変更するための要求として、または、自動アシスタントを介してアプリケーション設定を変更するための要求として解釈され得る。各解釈は、コンピューティングデバイスと通信しているディスプレイパネルを介して、ユーザに提示され得る、対応するサジェスチョンを生成するために処理され得る。たとえば、コンピューティングデバイスは、ディスプレイパネルが、口頭発話の両方の解釈に基づくサジェスチョンのリストとともに、口頭発話の最初のテキスト(たとえば、「アシスタント、変更してもらえますか...」)を示すことを引き起こすことができる。サジェスチョンは、ある解釈に基づく少なくとも1つのサジェスチョンと、別の解釈に基づく少なくとも1つの他のサジェスチョンとを含み得る。

【0006】

たとえば、「IoTデバイス設定を変更する」解釈に基づく少なくとも1つのサジェスチョンは、ユーザによって話されると、元の口頭発話を補足するテキストであり得る、「サーモスタットを72に」であり得る。一例として、ディスプレイパネルは、少なくとも1つのサジェスチョンのテキストを話すユーザに回答して、元の口頭発話、および少なくとも1つのサジェスチョンを一緒に提示することができる。言い換えれば、ディスプレイパネルは、以下のテキスト、すなわち、「アシスタント、サーモスタットを72に変更してもらえますか」を提示することができる。ユーザは、提示されたテキストを見て、その部分(すなわち、「サーモスタットを72に」)がサジェストされたものであり、自分の元の口頭発話(すなわち、「アシスタント、変更してもらえますか...」)内に含まれていなかったことを識別することができる。サジェストされた部分を識別すると、ユーザは、サジェストされた部分のテキストを復唱し、自動アシスタントが、ユーザのサーモスタットの設定を72に変更することなどのアクションを完成させることを促進するために作用することを引き起こすことができる。いくつかの実装形態では、ディスプレイパネルは、サジェスチョンとして、サジェストされた、まだ話されていない部分(たとえば、「アシスタント、してもらえますか」を提示することなく、「サーモスタットを72に」)のみを含む、テキストを提示することができる。これによって、サジェスチョンを容易に一瞥することができるようになり、ユーザがサジェスチョンを迅速に見ること、および、「サーモスタットを72に」または(サジェスチョンとは異なるが、サジェスチョンに基づいて確認可能である)「サーモ

10

20

30

40

50

スタットを70に」というさらなる口頭入力など、ユーザがアクション可能な要求を構築するために提供することができる、さらなる口頭入力を確認することが可能になり得る。

【0007】

様々な実装形態では、サジェスションのテキスト部分は、場合によっては、サジェスションのテキスト部分に適合しているさらなる口頭入力を提供する結果として実施されることになるアクションを示す、アイコンまたは他のグラフィカル要素と組み合わせて提示され得る。一例として、サジェスションは、サジェスションに適合しているさらなる口頭入力が、サーモスタットの設定点における変更を生じるようになることを示すために、サーモスタットのグラフィカル指示、および上下の矢印とともに、「サーモスタットを72に変更してください」というテキスト部分を含み得る。別の例として、サジェスションは、サジェスションに適合しているさらなる口頭入力が、テレビにおけるストリーミングコンテンツの再生を生じるようになることを示すために、再生アイコンを内部に有するテレビの輪郭とともに、「リビングルームのTVでTV局を再生して」というテキスト部分を含み得る。また別の例として、サジェスションは、サジェスションに適合しているさらなる口頭入力が、ライトをオフにすることを生じるようになることを示すために、その上に重ねられたXを有する電球アイコンとともに、「リビングルームのライトをオフにして」を含み得る。そのようなアイコンによっても、サジェスションをより容易に一瞥することができるようになり、ユーザが最初にアイコンを迅速に見ること、および、アイコンがユーザの所望のアクションに適合している場合、サジェスションのテキスト部分を見ることが可能になり得る。これによって、ユーザが、複数の同時に提示されたサジェスションのグラフィカル要素を(対応するテキスト部分を最初に読むことなく)迅速に一瞥して、ユーザの意図に適合するものを確認し、次いで、そのものの対応するテキスト部分のみを読み取ることが可能になり得る。これによって、ユーザが複数のサジェスションを提示される時、さらなる口頭入力を提供する際のレイテンシを低減することができ、結果的に、生じたアクションの実施のレイテンシが低減される。

【0008】

いくつかの実装形態では、口頭発話は、ユーザが特定の語を明確に述べた、かつ/または発音した方法の結果として、異なるように解釈され得る。たとえば、「アシスタント、変更してもらえますか...」という上述の口頭発話を受信するコンピューティングデバイスは、X%の確実性ととともに、口頭発話が「変更する」という語を含むこと、およびY%の確実性(ただし、Yは、X未満である)とともに、口頭発話が「整理する」という語を含み、したがって、「整理する」機能を実施するための要求を指すことを決定することができる。たとえば、口頭発話の音声テキスト処理は、異なる確実性をもつ、2つの別個の解釈(「変更する」および「整理する」)を生じることができる。結果として、口頭発話に回答するタスクを課せられる自動アシスタントは、各解釈(「変更する」および「整理する」)に基づくサジェスションを提示することができる。たとえば、自動アシスタントは、「変更する」解釈に対応する「...サーモスタットを72に」などの第1のサジェスションと、「整理する」解釈に対応する「...マイデスクトップファイルフォルダ」などの第2のサジェスションとをグラフィカルに提示することができる。さらに、ユーザが「サーモスタットを72に」(または、「サーモスタットを80に」などの同様のコンテンツ)を復唱することによって、第1のサジェスションのコンテンツを後で復唱する場合、ユーザは、「変更する」機能が自動アシスタントを介して(および、ユーザのさらなる復唱に基づいて)実施されることを引き起こすことができる。しかしながら、ユーザが「マイデスクトップファイルフォルダ」(または、「マイドキュメントフォルダ」などの同様のコンテンツ)を復唱することによって、第2のサジェスションのコンテンツを後で復唱する場合、ユーザは、「整理する」機能が自動アシスタントを介して(および、ユーザのさらなる復唱に基づいて)実施されることを引き起こすことができる。言い換えれば、自動アシスタントは、復唱されたサジェスションに適合する解釈(「変更する」または「整理する」)を利用することになる。たとえば、「整理する」が20%の確率のみで予測され、「変更する」が80%の確率で予測された場合でも、それにもかかわらず、ユーザが「マイデスクトップファイルフォルダ」とさ

10

20

30

40

50

らに話す場合、「整理する」が利用され得る。別の例として、ならびに「変更する」および「整理する」例の変形態として、「サーモスタットを75に変更してください」という第1のサジェスチョン、および「マイデスクトップファイルフォルダを整理してください」という第2のサジェスチョンが提示され得る。ユーザが、タッチ入力を通して「マイデスクトップファイルフォルダを整理してください」を選択するか、または「マイデスクトップファイルフォルダを整理してください」もしくは「第2のもの」など、さらなる口頭入力を提供する場合、「整理する」解釈が、「変更する」解釈の代わりに利用されることになる。言い換えれば、「変更する」が、より高い確率に基づいて、正確な解釈として最初に選択されるにもかかわらず、「整理する」は、それにもかかわらず、「整理する」サジェスチョンに向けられたユーザ選択またはさらなる口頭入力に基づいて、その選択に取って代わることがある。これらおよび他の方法において、音声テキスト精度を向上させることができ、より小さい確実性の解釈が、いくつかの状況において利用され得る。これによって、そうでない場合、最初の口頭発話の不正確な解釈が選択され、ユーザが、正確な解釈が選択されるように試みて、自分の最初の口頭発話を繰り返さなければならなかった場合に費やされ得る、計算リソースを保護することができる。さらに、これによって、ユーザが同じことを繰り返して言うことが減り、自動アシスタントが可聴応答をレンダリングするまで待機するようになるとすれば、ユーザが自動アシスタントに関与するのに費やす時間の量を低減することができる。

【0009】

いくつかの実装形態では、提供されたサジェスチョン要素は、サジェスチョンの提供に 20
応答して受信されたさらなる口頭発話の音声テキスト処理にバイアスをかけるための基礎として使用され得る。たとえば、音声テキスト処理には、提供されたサジェスチョン要素の用語の方に、および/または、サジェスチョンに対応するコンテンツの予想されるカテゴリもしくは他のタイプに適合する用語の方にバイアスをかけることができる。たとえば、「[ミュージシャンの名前]」という、提供されたサジェスチョンの場合、音声テキスト処理には、ミュージシャンの名前の方に(一般に、またはユーザのライブラリにおけるミュージシャンに)バイアスをかけることができ、「2:00に」という、提供されたサジェスチョンの場合、音声テキスト処理には、「2:00」の方に、またはより一般的に時間の方にバイアスをかけることができ、「[スマートデバイス]」という、提供されたサジェスチョンの場合、音声テキスト処理には、(たとえば、ユーザの記憶されたデバイスポロジから 30
確認されるような)ユーザのスマートデバイスの名前の方にバイアスをかけることができる。音声テキスト処理には、たとえば、音声テキスト処理を実施する際に使用するための(複数の候補音声テキストモデルからの)特定の音声テキストモデルを選択すること、ある用語に対応する候補変換に基づいて、(音声テキスト処理の間に生成された)候補変換のスコアを高めること、ある用語に対応する(音声テキスト処理において使用された)状態復号グラフ(state decoding graph)のパスを高めることによって、および/または追加のもしくは代替音声テキストバイアス技法によって、ある用語の方にバイアスをかけることができる。一例として、ユーザは、「アシスタント、~のためのカレンダーイベントを設定して」などの最初の口頭発話を提供することができ、最初の口頭発話の受信に応答して、自動アシスタントは、複数のサジェスチョン要素がコンピューティングデバイスのディスプレイ 40
パネルにおいて現れることを引き起こすことができる。サジェスチョン要素は、たとえば、「8月...9月...[現在の月]...」などであることがあり、ユーザは、「8月」などの後続の口頭発話を提供することによって、特定のサジェスチョン要素を選択することができる。サジェスチョン

要素が1月~12月を含むとの決定に応答して、自動アシスタントは、後続の口頭発話の音声テキスト処理に、1月~12月の方にバイアスをかけることができる。別の例として、数値入力を含む(または、示す)サジェスチョン要素の提供に応答して、音声テキスト処理には、数値の方にバイアスをかけることができる。これらおよび他の方法において、後続の発話の音声テキスト処理には、提供されたサジェスチョン要素の方を優先してバイアスをかけることができ、それによって、音声テキスト処理から予測されたテキストが正確であ 50

る確率を高めることができる。これによって、そうでない場合、予測されたテキストが不正確であった場合に必要となる(たとえば、不正確な解釈を訂正するための、ユーザからのさらなる入力に基づく)、さらなるダイアログの機会の発生を軽減することができる。

【0010】

いくつかの実装形態では、自動アシスタントに特定の要求を完成させるためのサジェスチョン要素を提供するためのタイミングは、それにおいてユーザが特定の要求を開始したコンテキストに基づき得る。たとえば、ユーザが自分の車両を運転中に、口頭発話を介して、要求の少なくとも一部分を提供したとき、自動アシスタントは、ユーザが車両を運転中であると決定し、要求を完成させるためのサジェスチョンを表示することを遅延させることができる。このようにして、ユーザは、グラフィックスが車両内のディスプレイパネルにおいて提示されることによって、気を散らされる頻度がより低くなる。追加または代替として、サジェスチョン要素のコンテンツ(たとえば、候補テキストセグメント)もまた、それにおいてユーザが口頭発話を開始したコンテキストに基づき得る。たとえば、自動アシスタントは、ユーザからの事前許可とともに、ユーザが自分のリビングルームから「アシスタント、オンにして...」などの口頭発話を提供したと決定することができる。それに応答して、自動アシスタントは、コンピューティングデバイスが、ユーザのリビングルーム内の特定のデバイスを識別するサジェスチョン要素を生成することを引き起こすことができる。サジェスチョン要素は、「リビングルームのTV...リビングルームのライト...リビングルームのステレオ」などの自然言語コンテンツを含み得る。ユーザは、「リビングルームのライト」などの後続の口頭発話を提供することによって、サジェスチョン要素のうちの1つの選択を行うことができ、それに応答して、自動アシスタントは、リビングルームのライトをオンにさせることができる。いくつかの実装形態では、サジェスチョン要素が識別するデバイスは、ユーザに関連付けられた様々なデバイス間の関係、および/または自宅もしくはオフィスなど、ユーザのロケーション内のコンピューティングデバイスの配置を特徴づける、デバイスポートロギーデータに基づき得る。デバイスポートロギーデータは、ロケーション内の様々なエリアのための識別子、ならびに、デバイスが各エリア内にあるか否かを特徴づけるデバイス識別子を含み得る。たとえば、デバイスポートロギーデータは、「リビングルーム」などのエリアを識別し、また、「TV」、「ライト」、および「ステレオ」など、リビングルーム内のデバイスを識別することができる。したがって、ユーザによる「アシスタント、オンにして」という上述の発話の提供に応答して、そのエリア内の特定のデバイスに対応するサジェスチョンを提供するために、デバイスポートロギーがアクセスされ、ユーザがいるエリアと比較され得る(たとえば、デバイスポートロギーデータによって「リビングルーム」内にあるとして定義されたデバイスを介して、発話が提供されることに基づいて、「リビングルーム」を識別する)。

【0011】

いくつかの実装形態では、受信された要求の少なくとも一部分に対応するアクションは、サジェスチョン要素が要求を完成させるためにユーザに提示されている間に、実施され得る。たとえば、ユーザは、「アシスタント、...のためのカレンダーイベントを作成して」などの口頭発話を提供することができる。それに応答して、自動アシスタントは、デフォルトコンテンツとともに、デフォルトカレンダーが作成されることを引き起こすことができ、また、カレンダーイベントを作成するための要求を完成させるために、サジェスチョン要素が作成されることを引き起こすこともできる。たとえば、サジェスチョン要素は、「今夜...明日...土曜日...」を含むことがあり、サジェスチョン要素は、デフォルトカレンダーイベントが作成される前、その間、またはその後提示され得る。ユーザがサジェスチョン要素のうちの1つを選択するとき、前に作成されたデフォルトカレンダーイベントは、ユーザが選択したサジェスチョン要素に従って修正され得る。

【0012】

いくつかの実装形態では、自動アシスタントは、音声処理の量、および/または、典型的にはユーザからのいくつかの入力に関連付けられる、アシスタント対話時間の量を低減するためのサジェスチョンを提供することができる。たとえば、ユーザが、カレンダーイベ

10

20

30

40

50

ントを作成するための自動アシスタントとのダイアログセッションを完了するとき、自動アシスタントは、それにもかかわらず、さらなるサジェスチョンを提供することができる。たとえば、ユーザが、「アシスタント、次の月曜日に、マシューの誕生日のためのカレンダーイベントを作成して」などの口頭発話を提供すると仮定する。それに応答して、自動アシスタントは、カレンダーイベントが生成されることを引き起こし、また、サジェスチョン要素を選択するようにユーザに勧めるために、「そして、毎年繰り返す」などのサジェスチョン要素が提示されることを引き起こし、それによって、カレンダーイベントが毎年繰り返されることを引き起こすことができる。これによって、ユーザがそのカレンダーイベントを毎年繰り返し作成する必要がなくなるので、ユーザと自動アシスタントとの間の予想される対話の数を低減することができる。また、これによって、「そして、毎年繰り返す」を将来、ボイスを介して提供して、それによって、ユーザが「毎年繰り返す」を含む将来の口頭入力を提供することが可能になり得ることを、ユーザに教示して、それによって、そのような将来の口頭の予定をより効率的にすることもできる。さらに、いくつかの実装形態では、サジェスチョン要素は、ユーザによって提供された前の口頭発話の簡約バージョンとしてサジェストされ得る。このようにして、電力およびネットワーク帯域幅などの計算リソースは、ユーザが自分の自動アシスタントと対話する時間の総量を低減することによって保護され得る。

10

【0013】

いくつかの実装形態では、ユーザに提示されるサジェスチョン要素は、ユーザがすでによく知っているいくつかの口頭発話を回避するために提供され得る。たとえば、ユーザが、「アシスタント、私のライトを調光して」などの口頭発話を提供した履歴を有するとき、自動アシスタントは、後で「アシスタント、調光して...」と言うユーザに応答して、「私のライト」というサジェスチョンを提供しないことがある。むしろ、ユーザは、「私のモニタ...私のタブレット画面...私のライトの[色]の量」など、ユーザがよく知らないことがある他のサジェスチョンを提示され得る。代替または追加として、上述のサジェスチョンは、ユーザが、不完全である口頭発話(たとえば、「アシスタント、再生して...」)を提供したときの状況において、および、ユーザが、完全である口頭発話(たとえば、「アシスタント、何か音楽を再生して」)を提供したときの他の状況において、ユーザに提示され得る。

20

【0014】

様々な実装形態では、ユーザは、いつ口頭発話がユーザによって完全であると考えられるかを明示的に示さないようになる。たとえば、ユーザは、口頭発話がユーザによって完全であると見なされると、「サブミット」ボタンを押さない、または「終了」、「実行」、「完了」、もしくは他の終わりのフレーズを話さないようになる。したがって、それらの様々な実装形態では、グラフィカルサジェスチョンが、すでに行われた口頭発話に応答して提示されることがあり、自動アシスタントは、サジェスチョンのうちの1つに適合し、すでに提供された口頭発話の継続である、さらなる口頭発話を提供するため(または、タッチ入力を通して、サジェスチョンのうちの1つを選択するため)に、より多くの時間をユーザに提供するか、代わりに、すでに提供された口頭発話のみに作用するかを決定することが必要になる。

30

40

【0015】

それらの実装形態のうちのいくつかでは、自動アシスタントは、ユーザからの口頭入力の欠如の持続時間の検出に応答して、すでに提供された口頭発話のみに作用ようになる。持続時間は、様々な実装形態では、動的に決定され得る。たとえば、持続時間は、すでに提供された口頭発話が「不完全な」要求であるか、「完全な要求」であるかに依存することがあり、かつ/または、すでに提供された口頭発話の「完全な要求」の特性に依存することがある。上述のように、「完全な要求」は、自動アシスタントによって完全に処理されると、1つまたは複数の対応する自動アシスタントアクションの実施を引き起こすものであり、ここで、実施された自動アシスタントアクションは、デフォルトエラータイプアクションではない(たとえば、「すみません、その役には立てません」という可聴応答で

50

はない)。「不完全な要求」は、自動アシスタントによって完全に処理されると、デフォルトエラータイプアクション(たとえば、エラー音、および/または「すみません、その役に立てません」などのデフォルトの口頭応答)の実施を引き起こすものである。

【0016】

様々な実装形態では、持続時間は、すでに行われた口頭発話が不完全な要求であるとき、それが完全な要求であるときと比較して、より大きくなり得る。たとえば、持続時間は、すでに行われた口頭発話が不完全な要求であるとき、5秒であり得るが、すでに行われた口頭発話が完全な要求であるとき、より短くなり得る。たとえば、「[日付]のX:00のためのカレンダーエントリ」、および「[時間]または[ロケーション]において、Xをするためのリマインダ」というサジェスチョンが、「aを作成する」という口頭発話に回答して提示され得る。「aを作成する」という口頭発話は不完全であるので、自動アシスタントは、さらなる口頭発話のために5秒待機することができ、さらなる口頭入力なしに、5秒が経過するときのみ、自動アシスタントは、不完全な口頭発話を完全に処理し、デフォルトエラータイプアクションを実施するようになる。また、たとえば、「[日付]のX:00のための」というサジェスチョンが、「お母さんに電話する、というカレンダーエントリを作成して」という口頭発話に回答して提示され得る。「お母さんに電話する、というカレンダーエントリを作成して」という口頭発話は、完全である(すなわち、タイトルをもつカレンダーエントリの作成、ならびに場合によっては、日付および時間詳細のためのさらなるプロンプトにつながるようになる)ので、自動アシスタントは、さらなる口頭発話のために3秒、または他の短縮された持続時間の間に待機することができ、さらなる口頭入力なしに、3秒が経過するときのみ、自動アシスタントは、すでに行われた口頭発話を完全に処理し、カレンダーエントリを作成し、場合によっては、カレンダーエントリのための日付および/または時間を提供するよう、ユーザにさらにプロンプトするようになる。

【0017】

別の例として、持続時間は、すべての必須パラメータを含むわけではない(かつ、必須パラメータのためにさらにプロンプトすることを生じるようになる)完全な要求のための持続時間と比較して、すべての必須パラメータを含む完全な要求の場合に、より短くなり得る。たとえば、カレンダーエントリを作成するための完全な要求は、「タイトル」、「日付」、および「時間」という必須パラメータを含み得る。「今日のPM2時にお母さんに電話する、というカレンダーエントリを作成して」という口頭発話は、すべての必須パラメータを含むが、それにもかかわらず、オプションパラメータのためのサジェスチョン(たとえば、「そして、毎日繰り返す」というサジェスチョン)が提供され得る。すべての必須パラメータが口頭発話に含まれているので、持続時間は、たとえば、口頭発話が、すべての必須パラメータを含むわけではなかった(かつ、「@[日付]のX:00」などのサジェスチョンが提示される結果となった)「お母さんに電話する、というカレンダーエントリを作成して」であった場合よりも、短くなり得る。また別の例として、持続時間は、要求に基づいて一般検索を発行し、検索結果を表示する、一般検索エージェントではなく、特定の自動アシスタントエージェントが呼び出される結果となる、完全な要求の場合に、より短くなり得る。たとえば、「時間とは何か」という口頭発話は、一般検索、および「時間」の概念に関連する情報を提供する検索結果のプロビジョンを生じる、完全な要求である。「[市]において」というサジェスチョンが、口頭発話に回答して表示されることがあり、2秒または他の持続時間の間に表示され得る。一方、「キッチンのライトをオンにして」という口頭発話は、「キッチンのライト」がオンにされることを引き起こす特定の自動アシスタントエージェントを呼び出す、完全な要求である。「X%の明るさに」というサジェスチョンが、口頭発話に回答して表示され得るが、1秒のみ(または、他の短縮された持続時間)の間に表示されることがあり、その理由は、すでに提供された口頭発話が完全であり、(一般検索の実施を引き起こすのではなく)特定の自動アシスタントエージェントを呼び出すからである。

【0018】

上記の説明は、本開示のいくつかの実装形態の概観として提供される。それらの実装形

10

20

30

40

50

態、および他の実装形態のさらなる説明について、以下でより詳細に説明する。

【0019】

他の実装形態は、上記および/または本明細書の他の場所で説明する方法のうちの1つまたは複数などの方法を実施するために、1つまたは複数のプロセッサ(たとえば、中央処理ユニット(CPU)、グラフィックス処理ユニット(GPU)、および/またはテンソル処理ユニット(TPU))によって実行可能な命令を記憶する、非一時的コンピュータ可読記憶媒体を含み得る。また他の実装形態は、上記および/または本明細書の他の場所で説明する方法のうちの1つまたは複数などの方法を実施するために、記憶された命令を実行するように動作可能な、1つまたは複数のプロセッサを含む、1つまたは複数のコンピュータ、および/あるいは1つまたは複数のロボットのシステムを含み得る。

10

【0020】

本明細書でより詳細に説明する上記の概念および追加の概念のすべての組合せが、本明細書で開示する主題の一部であるとして企図されることを諒解されたい。たとえば、本開示の最後に現れる請求される主題のすべての組合せが、本明細書で開示する主題の一部であるとして企図される。

【図面の簡単な説明】

【0021】

【図1A】ユーザが口頭発話を提供し、自動アシスタントが口頭発話を完成させるためのサジェスチョンを提供する図である。

【図1B】ユーザが口頭発話を提供し、自動アシスタントが口頭発話を完成させるためのサジェスチョンを提供する図である。

20

【図2A】ユーザが、口頭発話に回答して、および/あるいは自動アシスタントの1つまたは複数の動作に基づいて、カテゴリーサジェスチョンを受信することを示す図である。

【図2B】ユーザが、口頭発話に回答して、および/あるいは自動アシスタントの1つまたは複数の動作に基づいて、カテゴリーサジェスチョンを受信することを示す図である。

【図3】可聴プロンプトが自動アシスタントによってレンダリングされる代わりに、ユーザが、口頭発話を完成させるためのサジェスチョンを受信する図である。

【図4】表示モダリティを介して、サジェスチョンを提供し、自動アシスタントを介して、口頭発話を完成させるためのシステムの図である。

【図5】ユーザによって自動アシスタント、デバイス、アプリケーション、および/または任意の他の装置もしくはモジュールに提供された口頭発話を完成させる、かつ/または補足するための、1つまたは複数のサジェスチョン要素を提供するための方法の図である。

30

【図6】例示的なコンピュータシステムのブロック図である。

【発明を実施するための形態】

【0022】

図1Aおよび図1Bは、ユーザが口頭発話を提供し、自動アシスタントが口頭発話を完成させるためのサジェスチョンを提供する、図100および図120を示す。ユーザ112は、コンピューティングデバイス114の自動アシスタントインターフェースに口頭発話116を提供することができるが、口頭発話116は、少なくとも自動アシスタントに対して、不完全であり得る。たとえば、ユーザは、「アシスタント、下げてもらえますか...」などの口頭発話116を提供することができ、この口頭発話116が、コンピューティングデバイス114、および/またはサーバデバイスなどのリモートコンピューティングデバイスにおいて処理され、コンピューティングデバイス114のディスプレイパネル110においてグラフィカルにレンダリングされ得る。口頭発話116のグラフィカルレンダリングは、コンピューティングデバイス114のアシスタント対話型モジュール104において、グラフィカル要素106として提示され得る。口頭発話116の受信に回答して、コンピューティングデバイスは、口頭発話116が不完全な口頭発話であると決定することができる。不完全な口頭発話は、少なくとも1つのアクションが、自動アシスタント、および/または任意の他のアプリケーションもしくはアシスタントエージェントを介して実施される結果を生じない、口頭発話であり得る。代替または追加として、不完全な口頭発話は、機能が実行またはさもなけれ

40

50

ば制御されるために必要な1つまたは複数のパラメータ値がない、口頭発話であり得る。たとえば、不完全な口頭発話は、アシスタント機能が、自動アシスタントを介して実行されるために必要な、1つまたは複数のパラメータ値がなく、したがって、ユーザまたは他のソースからのさらなる、または補足入力を必要とする、口頭発話であり得る。

【0023】

口頭発話116が不完全な口頭発話であるとの決定にตอบสนองして、自動アシスタントは、1つまたは複数のサジェスチョン要素108が第1のユーザインターフェース102において提示されることを引き起こすことができる。しかしながら、いくつかの実装形態では、1つまたは複数のサジェスチョン要素108、および/またはそれらのコンテンツは、視覚障害があるか、またはさもなければ、第1のユーザインターフェース102を容易に見ることができないことがあるユーザに対して、可聴的にレンダリングされ得る。サジェスチョン要素108は、それからユーザ112が選択することができる候補テキストセグメントを特徴づける、自然言語コンテンツを含むことがあり、その選択は、第1のユーザインターフェース102上でタップすること、候補テキストセグメントのテキストを口頭で復唱すること、および/または、少なくとも1つのサジェスチョン要素108が、第1のユーザインターフェース102において提示されるとき、それを説明することによる。たとえば、「音」などの追加の口頭発話を提供するのではなく、ユーザ112は、「第1のもの」など、異なる追加の口頭発話を提供することができる。代替的に、ユーザが異なるサジェスチョン要素108を選択することを希望した場合、ユーザ112は、「ライト」と書いてあるサジェスチョン要素108にタッチ入力を提供し、所望のサジェスチョン要素108に対応するロケーションにおいて、第1のユーザインターフェース102をタップし、かつ/または「第2のもの」など、異なる追加の口頭発話を提供することができる。様々な実装形態では、サジェスチョン要素108のうちの1つまたは複数が、それらの選択にตอบสนองして実施されることになるアクションを示すために、アイコンおよび/または他のグラフィカル要素とともに提示され得る。たとえば、「音」が、スピーカーアイコンとともに提示されることがあり、「ライト」が、ライトアイコンとともに提示されることがあり、かつ/または「ガレージドア」が、ガレージドアアイコンとともに提示されることがある。

【0024】

各サジェスチョン要素108とともに提供されたテキストセグメントの自然言語コンテンツは、不完全な口頭発話116および候補テキストセグメントの組合せが、自動アシスタントがアクションを実施することを引き起こすようになるか否かに基づいて、生成され得る。たとえば、各候補テキストセグメントは、不完全な口頭発話116および候補テキストセグメントの組合せが、ユーザ112によって一緒に話されるとき、アクションが自動アシスタントを介して実施されることを引き起こすようになるか否かに基づいて、生成され得る。いくつかの実装形態では、ユーザ112が完全な口頭発話を提供するとき、自動アシスタント、コンピューティングデバイス114、および/または関連付けられたサーバデバイスは、サジェスチョン要素108を生成することをバイパスすることができる。代替的に、ユーザ112が完全な口頭発話を提供するとき、自動アシスタント、コンピューティングデバイス114、および/または関連付けられたサーバデバイスは、第1のユーザインターフェース102におけるプロビジョニングのために、サジェスチョン要素を生成することができる。

【0025】

いくつかの実装形態では、デバイスポロジータが、サジェスチョン要素108の候補テキストセグメントを生成するための基礎として使用され得る。たとえば、コンピューティングデバイス114は、ユーザ112に関連付けられる1つまたは複数のデバイスを識別する、デバイスポロジータへのアクセスを有することができる。さらに、デバイスポロジータは、1つまたは複数のデバイスのうちの各デバイスのロケーションを示すことができる。ユーザ112の位置が、ユーザ112からの事前許可とともに、ユーザ112のコンテキストを決定するために使用されることがあり、少なくとも、その理由は、コンテキストがユーザ112のロケーションを特徴づけるからである。デバイスポロジータは、ユーザ112の近くにあり得るデバイスを決定するために、位置データと比較され得

10

20

30

40

50

る。さらに、近くのデバイスが識別されると、不完全な口頭発話116の自然言語コンテンツを各近くのデバイスの各ステータスと比較するために、各近くのデバイスのステータスが決定され得る。たとえば、図1Aに示されているように、ユーザ112は、アシスタントが何かを「下げる」ための要求を提供したので、「下げること」のアクションを介して修正され得るステータスを有するデバイスが、サジェスチョン要素108を生成するための基礎として使用され得る。したがって、スピーカーが近くで音楽を再生中である場合、「音」というサジェスチョンが提供され得る。さらに、ライトがユーザの近くでオンである場合、「ライト」というサジェスチョンが提供され得る。追加または代替として、デバイスプロファイルデータは、ユーザ112の近くでないことがあるが、それにもかかわらず、不完全な口頭発話116に関連付けられ得る、デバイスを特徴づけることができる。たとえば、ユーザ112は、自分の寝室にいることがあるが、不完全な口頭発話116は、ユーザ112の自宅のガレージ内のガレージドアに関連付けられ得る。したがって、自動アシスタントはまた、ユーザがガレージ内にいないにもかかわらず、ユーザが「ガレージドアを下げる」ことを可能にするサジェスチョン要素108を提供することもできる。

【0026】

図1Bは、図1Aに示された第1のユーザインターフェース102において提供されたサジェスチョン要素108を選択することを促進するために、ユーザ112が追加の口頭発話128を提供する図120を示す。ユーザ112は、「ガレージドア」など、追加の口頭発話128を提供することによって、前に表示されたサジェスチョン要素108のうちの1つを選択することができる。コンピューティングデバイス114の自動アシスタントインターフェースにおける口頭発話の受信にตอบสนองして、選択されたサジェスチョン要素108に対応する候補テキストが、不完全な口頭発話116の自然言語コンテンツに隣接して提示され得る。たとえば、図1Bに示されているように、完成された候補テキスト122は、不完全な口頭発話116からの自然言語コンテンツと、選択されたサジェスチョン要素108(たとえば、追加の口頭発話128の自然言語コンテンツ)との組合せとして、第2のユーザインターフェース126において提示され得る。

【0027】

コンピューティングデバイス114、自動アシスタント、および/またはサーバデバイスは、最初の口頭発話116と組み合わせられた追加の口頭発話128が、完全な口頭発話を生じたと決定することができる。それにตอบสนองして、1つまたは複数のアクションが、自動アシスタントを介して実施され得る。たとえば、完成された口頭発話が、完成された候補テキスト122に対応するとき、自動アシスタントは、ユーザ112の自宅におけるガレージドアが電気機械的に閉じ位置に下がることを引き起こすことができる。いくつかの実装形態では、完全な口頭発話が形成されたとき、および/または不完全な口頭発話が形成されたとき、追加のサジェスチョン要素124が、第2のユーザインターフェース126において提供され得る。たとえば、完成された候補テキスト122が、自動アシスタントによってアクション可能であるにもかかわらず、コンピューティングデバイス114、自動アシスタント、および/またはサーバデバイスは、追加のサジェスチョン要素124が第2のユーザインターフェース126において提示されることを引き起こすことができる。本明細書で説明するように、様々な実装形態では、自動アシスタントは、いかなるさらなる口頭入力、または追加のサジェスチョン要素124のうちのいずれか1つのいかなる選択もなしに、時間の持続時間が経過した場合、追加のサジェスチョン要素124のうちの1つの選択もなしに、完成された候補テキスト122に基づいて、アクションを実行する(すなわち、ガレージドアが下げられることを引き起こす)ことができる。同じく本明細書で説明するように、その時間の持続時間は、その間に自動アシスタントが図1Aにおいて待機するようになる時間の持続時間よりも短くなり得る。このより短い時間の持続時間は、たとえば、完成された候補テキスト122が「完全な要求」であるのに対して、候補テキスト106(図1A)が「不完全な要求」であることに基づき得る。

【0028】

いくつかの実装形態では、実施され得る、かつ完成された候補テキスト122に関連する

10

20

30

40

50

が口頭発話218を提供したコンテキストに基づき得る。コンテキストは、コンピューティングデバイス216、コンピューティングデバイス216と通信しているリモートデバイス(サーバデバイスなど)、自動アシスタント、および/または、ユーザ214に関連付けられた任意の他のアプリケーションもしくはデバイスにおいて、利用可能なコンテキストデータによって特徴づけられ得る。たとえば、デバイスポロジータは、コンピューティングデバイス216が1つまたは複数の他のデバイスとペアリングされているか、またはさもなければ通信していることを示すことができる。このデバイスポロジータに基づいて、サジェスチョン要素210のうちの1つまたは複数は、「[デバイス名]」などの候補カテゴリーを提供することができる。代替または追加として、1つまたは複数のソースからのデータは、ユーザ214が最近アクセスしたメディアを示すアプリケーションデータを含み得る。最近のメディアが映画を含むとき、サジェスチョン要素210は、「[映画の名前]」などのコンテンツを含むように生成され得る。代替または追加として、最近のメディアが音楽を含むとき、サジェスチョン要素210は、「[曲およびアーティストの名前]」などのコンテンツを含むように生成され得る。

【0031】

図2Bは、ユーザ214が、図2Aにおいて提供されたサジェスチョン要素210から候補カテゴリーを選択するための、追加の口頭発話228を提供する図220を示す。図示のように、ユーザ214は、「キッチンのライト」という追加の口頭発話228を提供することができる。追加の口頭発話228の受信に回答して、追加のサジェスチョン要素226が、ディスプレイパネル212を介して、ユーザ214に提示され得る。追加のサジェスチョン要素226は、ユーザによる、追加の口頭発話228を提供することによる、最初の口頭発話218の完成に回答して、または追加の口頭発話228が、最初の口頭発話218を完成させないとき、ユーザ214に提示され得る。いくつかの実装形態では、追加のサジェスチョン要素226は、サジェスチョン要素210と同様のフォーマットのものであり得る、候補カテゴリーなどのコンテンツを含み得る。追加または代替として、追加のサジェスチョン要素226は、ユーザが話すとき、音声のコンテンツが、自動アシスタントを介して1つまたは複数のアクションを完成させること、修正すること、および/または初期化することを促進するために使用され得るように、コンテンツを含み得る。したがって、アシスタント対話型モジュール222は、候補テキストセグメントの様々なフォーマットを提供することができ、候補テキストセグメントは、候補カテゴリー(たとえば、サジェスチョン要素210において提供されたもの)、候補要求(たとえば、追加のサジェスチョン要素226において提供されたもの)、ならびに/あるいは、自動アシスタントおよび/または他のアプリケーションもしくはデバイスに提供されるコマンドに含めるために使用され得る任意の他のコンテンツを含み得る。

【0032】

いくつかの実装形態では、1つまたは複数のサジェスチョン要素のプロビジョニング、および/あるいは、1つまたは複数のサジェスチョン要素の提示の順序は、あるコンテンツ、アクション、口頭発話、候補テキストセグメント、および/または、それからサジェスチョンに基づき得る任意の他のデータに割り当てられた、優先度に基づき得る。たとえば、いくつかの実装形態では、サジェスチョン要素226内のあるコンテンツのプロビジョニングは、ユーザ214と自動アシスタントとの間、ユーザ214とコンピューティングデバイス216との間、ユーザ214と1つまたは複数の他のデバイスとの間、ユーザ214と1つまたは複数の他のアプリケーションとの間、ならびに/あるいは1人または複数の他のユーザと1つまたは複数のデバイスおよび/またはアプリケーションとの間の、前の対話の特徴づける履歴データに基づき得る。たとえば、「[デバイス名]」というコンテンツをもつサジェスチョン要素210の選択の受信に回答して、コンピューティングデバイス216および/またはサーバデバイスは、ユーザ214と、ユーザ214に関連付けられた1つまたは複数のデバイスとの間の前の対話の特徴づける、履歴データにアクセスすることができる。履歴データは、サジェスチョン要素226のためのコンテンツを生成するための基礎として使用され得る。たとえば、ユーザが「キッチンのライト」を選択するにもかかわらず、履歴データは、ユーザが制御することに関心があり得る他のデバイス、および/または、ユーザが明示的

10

20

30

40

50

に選択しなかったデバイスのために修正することに関心があり得る他の設定(「オン」以外)についての、サジェスチョンを生成するために使用され得る。

【0033】

一例として、およびユーザ214からの事前許可とともに、履歴データは、それにおいてユーザ214が寝る数分前に自分のキッチンのライトをオフにした、1つまたは複数の前の対話の特徴づけることができる。履歴データによって特徴づけられたこれらの対話に基づいて、自動アシスタントは、そのような対話が「アシスタント、キッチンのライトをオンにしてもらえますか?」などの完全な要求224に関連するので、そのような対話に最高優先度を割り当てることができる。さらに、自動アシスタントは、「私が寝るまで」など、サジェスチョン要素226のためのコンテンツを生成することができる。最高優先度の対話を特徴づける履歴データのコンテンツに基づいて、少なくとも、そのコンテンツが完成された要求224に関連するので、自動アシスタントは、「私が寝るまで」というサジェスチョン要素226を、第2のユーザインターフェース232において、最高優先度または最高位置のサジェスチョン要素226として提供することができる。

10

【0034】

代替または追加として、履歴データが、それにおいてユーザ214が、自動アシスタントがキッチンのライトを半分の明るさ設定、または赤色設定にすることを要求した、1つまたは複数の前の対話の特徴づけるが、それらの要求が、寝る前にライトをオフにする上述の要求よりも低い頻度で生じたとき、それらの要求のコンテンツには、上述の要求のコンテンツに割り当てられた前の優先度未満である優先度が割り当てられ得る。言い換えれば、ユーザ214が、寝るときにキッチンのライトがオフにされることをより頻繁に要求したので、「私が寝るまで」というサジェスチョン要素は、「そして、半分の明るさ設定で」および/もしくは「そして、赤色設定に」などの他のサジェスチョン要素よりも優先され、かつ/またはさもなければ、より高い位置を有するようになる。いくつかの実装形態では、サジェスチョン要素のコンテンツは、ユーザ214が前に使用していないデバイスまたはアプリケーションの機能、および/あるいは、サードパーティエンティティによってアプリケーションまたはデバイスにプッシュされた1つまたは複数の更新に基づき得る。サードパーティエンティティという用語は、コンピューティングデバイス216、自動アシスタント、および/またはコンピューティングデバイス216と通信しているサーバデバイスのメーカーとは異なるエンティティを指すことがある。したがって、キッチンのライトのメーカー(すなわち、サードパーティエンティティ)が、キッチンのライトが赤色設定に従って動作することを可能にする更新をプッシュするとき、自動アシスタントは、この更新を特徴づけるデータにアクセスし、サードパーティメーカー(すなわち、サードパーティエンティティ)によって提供された更新に基づいて、「そして、赤色設定に」というサジェスチョン要素226を生成することができる。

20

30

【0035】

代替または追加として、更新に基づく、サジェスチョン要素226のコンテンツには、ユーザ214が実施されることを頻繁に要求する1つまたは複数のアクションに基づく、別のサジェスチョン要素226に対応するコンテンツに割り当てられた別の優先度未満である、優先度が割り当てられ得る。しかしながら、更新に基づくコンテンツには、ユーザ214がディスプレイパネル212を介して、および/または自動アシスタントを介して、前に提示されたが、前に提示されたコンテンツを選択しないように選定した、コンテンツよりも高い優先度が割り当てられ得る。このようにして、ユーザに提供されるサジェスチョン要素は、ユーザに合わせてサジェスチョンを適応させるために、割り当てられた優先度に従って循環され得る。さらに、それらのサジェスチョンが、ユーザ214と自動アシスタントとの間の対話Xを低減することに関連するとき、計算リソースおよびネットワークリソースは、そのような対話を低減しないか、またはさもなければユーザに関連のないサジェスチョンを制限しながら、そのような対話を低減するより多くのサジェスチョンを提供することによって、保護され得る。

40

【0036】

50

図3は、可聴プロンプトが自動アシスタントによってレンダリングされる代わりに、ユーザ312が、口頭発話316を完成させるためのサジェスションを受信する図300を示す。ユーザ312は、「アシスタント、カレンダーエントリを作成してもらえますか?」などの口頭発話316を提供することができる。それに応答して、自動アシスタントを介して制御され得る、アシスタント対話型モジュール304は、口頭発話316のグラフィカル表現306を提供することができ、グラフィカル表現306は、コンピューティングデバイス314のディスプレイパネル320において提示され得る。

【0037】

口頭発話316は、機能が実行されるための要求に対応することがあり、その機能は、ユーザ312によって指定されるか、またはさもなければ、1つもしくは複数の利用可能値によって満たされなければならない、1つまたは複数の必須パラメータを有し得る。1つまたは複数の必須パラメータを満たすための1つまたは複数の値を提供するように、ユーザにプロンプトするために、自動アシスタントは、場合によっては、特定のパラメータのための値を要求するグラフィカルプロンプト308を提供することができる。たとえば、プロンプト308は、「タイトルは何になりますか?」などのコンテンツを含み、それによって、ユーザ312がカレンダーエントリのためのタイトルを提供しなければならないことを示すことができる。しかしながら、コンピューティングデバイス314の電力および計算リソースを保護するために、自動アシスタントは、コンピューティングデバイス314が、ユーザ312のための可聴プロンプトのレンダリングをバイパスすることを引き起こすことができる。さらに、これによって、遅延時間を省くことができ、その理由は、ユーザ312が、可聴プロンプトが完全にレンダリングされるまで待機しなくてもよくなるからである。

【0038】

自動アシスタントが可聴プロンプトの提供をバイパスする代わりに、自動アシスタントは、ユーザ312によって要求された機能のための1つまたは複数の必須パラメータに基づき得る、1つまたは複数のサジェスション要素310を生成することができる。たとえば、機能が、カレンダーエントリを生成することに対応し、満たされることになるパラメータが、カレンダーエントリの「タイトル」であるので、サジェスション要素310は、カレンダーエントリのためのタイトルサジェスションを含み得る。たとえば、サジェスションのコンテンツは、「私のお母さんの誕生日のために」、「ごみを出すために」、および「明日、ドライクリーニングを引き取るために」を含み得る。

【0039】

サジェスション要素310のうちの1つを選択するために、ユーザ312は、少なくとも1つのサジェスション要素310のコンテンツを識別する、追加の口頭発話318を提供することができる。たとえば、ユーザ312は、「明日、ドライクリーニングを引き取るために」という追加の口頭発話318を提供することができる。結果として、ならびに、コンピューティングデバイス314および/または自動アシスタントの観点から、ユーザ312と自動アシスタントとの間の対話中にレンダリングされたダイアログのみが、ユーザ312によってレンダリングされているようになる。したがって、オーディオのレンダリングは、必要ではないようになるか、またはさもなければ、自動アシスタントによって採用されないようになる。さらに、ユーザによる追加の口頭発話318の提供に応答して、要求された機能のために必要であるならば、ユーザ312が別のパラメータのための値を提供することを要求するために、異なるプロンプトが、ディスプレイパネル320において提供され得る。さらに、別のパラメータが機能のために必要とされるとき、異なるサジェスション要素が、異なるコンテンツとともにディスプレイパネル320において提示され得る。その上、後続のプロンプトおよび後続のサジェスション要素が、いかなるオーディオもコンピューティングデバイス314において、またはさもなければ自動アシスタントを介してレンダリングされることなく、提示され得る。したがって、ユーザ312は、表面上、ユーザが要求した元の機能をそれからさらに完成させるためのコンテンツを識別するために、ディスプレイパネル320を同時に見ながら、進行中の口頭発話を提供しているようになる。

【0040】

図4は、自動アシスタント408を介して、口頭発話を完成させるためのサジェスチョンを、表示モダリティを介して提供するシステム400を示す。いくつかの実装形態では、サジェスチョンは、ユーザが口頭発話を完成させるか、またはユーザの前の口頭発話を少なくとも補足するために、ユーザに提供され得る。いくつかの実装形態では、サジェスチョンは、ユーザが自動アシスタントとの後続のダイアログセッションに参加することになる頻度および/または時間の長さを低減するために、提供され得る。

【0041】

自動アシスタント408は、クライアントデバイス434、および/またはサーバデバイス402など、1つまたは複数のコンピューティングデバイスにおいて提供される、自動アシスタントアプリケーションの一部として動作することができる。ユーザは、1つまたは複数のアシスタントインターフェース436を介して、自動アシスタント408と対話することができ、アシスタントインターフェース436は、マイクロフォン、カメラ、タッチスクリーンディスプレイ、ユーザインターフェース、および/または、ユーザとアプリケーションとの間のインターフェースを提供することが可能な任意の他の装置のうちの1つまたは複数を含み得る。たとえば、ユーザは、口頭の、テキストの、および/またはグラフィカル入力をアシスタントインターフェースに提供することによって、自動アシスタント408を初期化して、自動アシスタント408が機能を実施する(たとえば、データを提供する、デバイスを制御する(たとえば、IoTデバイス442を制御する)、エージェントにアクセスする、設定を修正する、アプリケーションを制御するなどの)ことを引き起こすことができる。クライアントデバイス434および/またはIoTデバイス442は、ディスプレイデバイスを含むことがあり、ディスプレイデバイスは、ユーザがタッチインターフェースを介して、クライアントデバイス434および/またはサーバデバイス402のアプリケーションを制御することを可能にするための、タッチ入力および/またはジェスチャーを受信するためのタッチインターフェースを含む、ディスプレイパネルであり得る。タッチ入力および/または他のジェスチャー(たとえば、口頭発話)はまた、ユーザがクライアントデバイス434を介して、自動アシスタント408、自動アシスタント438、および/またはIoTデバイス442と対話することを可能にすることもできる。

【0042】

いくつかの実装形態では、クライアントデバイス434および/またはIoTデバイス442は、ディスプレイデバイスを欠くことがあるが、オーディオインターフェース(たとえば、スピーカーおよび/またはマイクロフォン)を含み、それによって、グラフィカルユーザインターフェース出力を提供することなく、可聴ユーザインターフェース出力を提供すること、ならびに、ユーザからの口頭自然言語入力を受信するために、マイクロフォンなどのユーザインターフェースを提供することができる。たとえば、いくつかの実装形態では、IoTデバイス442は、1つまたは複数のボタンなど、1つまたは複数の触覚入力インターフェースを含み、グラフィックス処理ユニット(GPU)からのグラフィカルデータが提供されるようになるディスプレイパネルを省略することができる。このようにして、エネルギーおよび処理リソースが、ディスプレイパネルおよびGPUを含むコンピューティングデバイスと比較して、節約され得る。

【0043】

クライアントデバイス434は、インターネットなどのネットワーク446上で、サーバデバイス402と通信中であり得る。クライアントデバイス434は、クライアントデバイス434および/またはIoTデバイス442における計算リソースを保護するために、計算タスクをサーバデバイス402にオフロードすることができる。たとえば、サーバデバイス402は、自動アシスタント408をホストすることができ、クライアントデバイス434は、1つまたは複数のアシスタントインターフェース、および/またはIoTデバイス442のユーザインターフェース444において受信された入力を、サーバデバイス402に送信することができる。しかしながら、いくつかの実装形態では、自動アシスタント408は、クライアントデバイス434においてホストされ得る。様々な実装形態では、自動アシスタント408のすべてまたはすべてに満たない態様は、サーバデバイス402および/またはクライアントデバイス

10

20

30

40

50

434上で実装され得る。それらの実装形態のうちのいくつかでは、自動アシスタント408の態様は、クライアントデバイス434のローカル自動アシスタント438を介して実装され、自動アシスタント408の他の態様を実装することができる、サーバデバイス402とインターフェースする。サーバデバイス402は、場合によっては、複数のスレッドを介して、複数のユーザ、およびユーザの関連付けられたアシスタントアプリケーションにサービスすることができる。自動アシスタント408のすべてまたはすべてに満たない態様が、クライアントデバイス434のローカル自動アシスタント438を介して実装される実装形態では、ローカル自動アシスタント438は、(たとえば、オペレーティングシステムの「上に」インストールされた)クライアントデバイス434のオペレーティングシステムとは別個であるアプリケーションであり得るか-または代替的に、クライアントデバイス434のオペレーティングシステムによって直接実装され(たとえば、オペレーティングシステムのアプリケーションであるが、オペレーティングシステムと一体であると考えられ)得る。

10

【0044】

いくつかの実装形態では、自動アシスタント408および/または自動アシスタント438は、入力処理エンジン412を含むことがあり、入力処理エンジン412は、クライアントデバイス434、サーバデバイス402、および/あるいは1つまたは複数のIoTデバイス442のための入力および/または出力を処理するために、複数の異なるエンジンを採用することができる。たとえば、入力処理エンジン412は、オーディオデータにおいて組み入れられたテキストを識別するために、アシスタントインターフェース436において受信されたオーディオデータを処理することができる、音声処理エンジン414を含み得る。オーディオデータは、たとえば、クライアントデバイス434における計算リソースを保護するために、クライアントデバイス434からサーバデバイス402に送信され得る。

20

【0045】

オーディオデータをテキストに変換するためのプロセスは、音声認識アルゴリズムを含むことがあり、音声認識アルゴリズムは、ニューラルネットワーク、word2vecアルゴリズム、および/または、語もしくはフレーズに対応するオーディオデータのグループを識別するための統計モデルを採用することができる。オーディオデータから変換されたテキストは、データパーシングエンジン416によってパースされ、ユーザからのコマンドフレーズを生成および/または識別するために使用され得るテキストデータとして、自動アシスタント408にとって利用可能にされ得る。いくつかの実装形態では、データパーシングエンジン416によって提供された出力データは、アクションエンジン418に提供されて、ユーザが、自動アシスタント408、ならびに/または、自動アシスタント408によってアクセスされることが可能であるアプリケーション、エージェント、および/もしくはデバイスによって実施されることが可能な、特定のアクションおよび/またはルーチンに対応する入力を提供したか否かが決定され得る。たとえば、アシスタントデータ422は、サーバデバイス402および/またはクライアントデバイス434において、クライアントデータ440として記憶されることがあり、自動アシスタント408によって実施されることが可能な1つまたは複数のアクションを定義するデータ、ならびに、それらのアクションを実施することに関わるパラメータを含み得る。

30

【0046】

入力処理エンジン412が、特定のアクションまたはルーチンが実施されることを、ユーザが要求したと決定したとき、アクションエンジン418は、特定のアクションまたはルーチンのための1つまたは複数のパラメータを決定することができ、次いで、出力生成エンジン420は、特定のアクション、ルーチン、および/あるいは1つまたは複数のパラメータに基づいて、ユーザに出力を提供することができる。たとえば、いくつかの実装形態では、アシスタントインターフェース436に向けられたジェスチャーなどのユーザ入力にตอบสนองして、自動アシスタント438は、ジェスチャーを特徴づけるデータが、自動アシスタント408および/または自動アシスタント438が実施することをユーザが意図しているアクションを決定するために、サーバデバイス402に送信されることを引き起こすことができる。

40

【0047】

50

クライアントデバイス434および/またはIoTデバイス442は、ユーザからのユーザ入力に
応答して、出力を提供することが可能な、1つまたは複数のセンサーをそれぞれ含み得
る。たとえば、1つまたは複数のセンサーは、ユーザからのオーディオ入力に
応答して、出力信号を提供する、オーディオセンサー(すなわち、オーディオ
応答デバイス)を含み得る。出力信号は、Bluetooth、LTE、Wi-Fi、および/ま
たは任意の他の通信プロトコルなど、通信プロトコルを介して、クライアント
デバイス434、サーバデバイス402、IoTデバイス442に送信され得る。いくつ
かの実装形態では、クライアントデバイス434からの出力信号は、クライアント
デバイス434の1つまたは複数のプロセッサによって、ユーザ入力データに変換
され得る。ユーザ入力データは、通信プロトコルを介して、IoTデバイス442
(たとえば、Wi-Fi対応電球)に送信され、クライアントデバイス434において
処理されること、および/またはネットワーク446を介して、サーバデバイス402
に送信されることのいずれかが行われ得る。たとえば、クライアントデバイス434
および/またはサーバデバイス402は、ユーザ入力データに基づいて、ユーザ
入力データを介して制御されることになるIoTデバイス442の機能を決定する
ことができる。

10

【0048】

ユーザは、アシスタントインターフェース436において受信され、IoTデバイス442、
クライアントデバイス434、サーバデバイス402、自動アシスタント、エージェント
アシスタント、および/または任意の他のデバイスもしくはアプリケーションを
制御することを促進するために、ユーザによって提供される、口頭発話などの
入力を提供し得る。いくつかの実装形態では、システム400は、どのように
口頭発話を完成させ、口頭発話を補足し、かつ/または他の方法でユーザを
支援するかについて、ユーザにサジェスチョンを提供することができる。
たとえば、ユーザは、IoTデバイス442が特定の機能を実施することを引き
起こす意図とともに、不完全な口頭発話を提供することができる。不完全な
口頭発話は、アシスタントインターフェース436において受信され、オーディオ
データに変換されることがあり、オーディオデータは、処理のためにサーバ
デバイス402に送信され得る。サーバデバイス402は、オーディオデータ、ま
たは不完全な口頭発話に基づく他のデータを受信し、不完全な口頭発話を
完成させるかまたはさもなければ補足するための、1つまたは複数のサジェ
スチョンを決定することができる、サジェスチョンエンジン424を含み得る。

20

【0049】

いくつかの実装形態では、1つまたは複数のサジェスチョン要素、ならびに、
クライアントデバイス434および/またはIoTデバイス442のディスプレイパネルを
提供するためのタイミングは、サジェスチョンエンジン424によって決定され
得る。サジェスチョンエンジン424は、タイミングエンジン428を含むことが
あり、タイミングエンジン428は、いつユーザに追加の口頭発話のための1
つまたは複数のサジェスチョンが提示されるべきであるかを決定する
ことができる。いくつかの実装形態では、1つまたは複数のしきい値が、
サジェスチョンエンジン424にとって利用可能であるアクションデータ444
によって特徴づけられ得る。たとえば、アクションデータ444は、ユーザが
口頭発話の一部を提供した後に遅延し得る時間に対応するしきい値を特徴
づけることができる。ユーザが不完全な口頭発話を提供した後、ユーザが
遅延するならば、遅延時間の量が、444のアクションデータによって特徴
づけられたしきい値と比較され得る。遅延がしきい値を満たすとき、
タイミングエンジン428は、1つまたは複数のサジェスチョン要素が、不
完全な口頭発話を完成させるために、ユーザに提示されることを引き起こ
すことができる。

30

40

【0050】

いくつかの実装形態では、タイミングエンジン428は、ユーザによる不完全な
口頭発話の提供と、口頭発話を完成させるためのサジェスチョンの受信との
間の遅延時間のための値を適応させることができる。たとえば、サジェ
スチョンエンジン424は、アシスタントデータ422、アクションデータ444、
および/またはクライアントデータ440において組み入れられたコンテキ
ストデータに基づいて、ユーザのコンテキストを決定し、遅延時間のた
めの値を決定するための基礎として、コンテキストを使用することができる。
たとえば

50

、ユーザが、ユーザが自宅にいて、TVを見ていることによって特徴づけられた、第1のコンテキストにあるとき、タイミングエンジン428は、遅延時間を第1の値に設定することができる。しかしながら、ユーザが、ユーザが運転し、車両コンピューティングデバイスと対話することによって特徴づけられた、第2のコンテキストにあるとき、タイミングエンジン428は、遅延時間を、第1の値よりも大きい第2の値に設定することができる。このようにして、口頭発話を完成させるためのサジェスチョン要素は、ユーザが運転中に、ユーザの気を散らすことを制限するために、比較的遅延され得る。

【0051】

いくつかの実装形態では、遅延時間のための値は、ユーザが経時的にどのくらいの頻度で不完全な口頭発話を提供するかに比例するか、またはそれに従って適応させられ得る。たとえば、最初に、ユーザが自動アシスタントとの対話を開始するとき、ユーザは、期間YごとにX数の不完全な口頭発話を提供することができる。その後しばらくして、ユーザは、Z数の不完全な発話Yを提供した履歴を有することができる。ZがX未満である場合、タイミングエンジン428は、遅延時間しきい値のための値を経時的に増加させることができる。したがって、遅延時間しきい値が増加するにつれて、サーバデバイス402および/またはクライアントデバイス434における計算リソースを保護することができ、その理由は、経時的にディスプレイパネルにおいて提示されるものに対する変更が少なくなり、それによって、ディスプレイパネルのためのグラフィカルデータを処理することを担うGPUにおける負荷が低減されるからである。

【0052】

いくつかの実装形態では、サジェスチョンエンジン424は、音声テキストモデルエンジン430を含むことがあり、音声テキストモデルエンジン430は、ユーザからの期待される口頭発話に従って、複数の異なる音声テキストモデルから、音声テキストモデルを選択することができる。追加または代替として、音声テキストモデルエンジン430は、ユーザからの口頭発話の期待されるタイプに従って、複数の異なる音声テキストモデルから、音声テキストモデルを選択することができる。代替または追加として、音声テキストモデルエンジン430は、ユーザからの期待される口頭発話に基づいて、選択された音声テキストモデルにバイアスをかけることができる。いくつかの実装形態では、ユーザに提供されたサジェスチョン要素が、少なくともいくつかの数値テキストを有するコンテンツを含むとき、音声テキストモデルエンジン430は、コンテンツに基づいて、数値情報を特徴づける口頭発話を処理するために適応させられる音声テキストモデルを選択することができる。代替または追加として、ユーザに提供されたサジェスチョン要素が、市または人の名前など、少なくとも固有名詞を含むコンテンツを含むとき、音声テキストモデルエンジン430は、固有名詞をより容易に解釈するために、選択された音声テキストモデルにバイアスをかけることができる。たとえば、選択された音声テキストモデルには、そうでない場合に固有名詞を処理することから生じるようになるエラーレートが低減されるように、バイアスをかけることができる。これによって、そうでない場合に生じ得るレイテンシを除去することができ、かつ/または、サジェスチョンエンジン424からの無関係のサジェスチョンの頻度を低減することができる。

【0053】

いくつかの実装形態では、不完全な口頭発話、または完全な口頭発話に回答して、ユーザに提供された1つまたは複数のサジェスチョン要素のうちの各サジェスチョンのコンテンツは、ユーザのコンテキストに基づき得る。たとえば、ユーザのコンテキストは、デバイスポロジエンジン426によって生成されたデータによって特徴づけられ得る。デバイスポロジエンジン426は、ユーザから不完全または完全な口頭発話を受信したデバイスを識別し、そのデバイスに関連付けられた情報を決定することができ、その情報は、受信デバイスに接続される他のデバイス、他のデバイスに対する受信デバイスのロケーション、および受信デバイスのための識別子、受信デバイスに関連付けられた1人または複数の異なるユーザ、受信デバイスの機能、受信デバイスとペアリングされるか、または別の方法で通信している、1つまたは複数の他のデバイスの機能、ならびに/あるいは、口頭

10

20

30

40

50

発話を受信したデバイスに関連付けられ得る任意の他の情報などである。たとえば、受信デバイスが、ユーザの自宅のリビングルームに位置するスタンドアロンスピーカースピーカーデバイスであるとき、デバイスプロローグエンジン426は、スタンドアロンスピーカースピーカーデバイスがリビングルーム内にあると決定し、また、リビングルーム内に位置する他のデバイスも決定することができる。

【0054】

サジェスチョンエンジン424は、ユーザに提示されることになるサジェスチョン要素を生成するために、リビングルーム内の他のデバイスの説明などの情報を使用することができる。たとえば、スタンドアロンスピーカースピーカーデバイスによるユーザからの不完全な口頭発話の受信に回答して、デバイスプロローグエンジン426は、テレビが、スタンドアロンスピーカースピーカーデバイスとともにリビングルーム内にあると決定することができる。サジェスチョンエンジン424は、テレビが同じ室内にあるとの決定に基づいて、不完全な口頭発話を完成させるためのサジェスチョン要素のためのコンテンツを生成することができる。たとえば、不完全な口頭発話が「アシスタント、変更して」であるとき、サジェスチョン要素のコンテンツは、「テレビのチャンネル」を含み得る。代替または追加として、デバイスプロローグエンジン426は、テレビがスタンドアロンスピーカースピーカーデバイスと同じ室内にあると決定し、また、スタンドアロンスピーカースピーカーデバイスがサジェスチョン要素をグラフィカルに表現することが可能ではないが、テレビがサジェスチョン要素をグラフィカルに表現することが可能であると決定することもできる。したがって、不完全な口頭発話に回答して、サジェスチョンエンジン424は、テレビディスプレイパネルが、「私のテレビのチャンネルを変更して」などのコンテンツを含む、1つまたは複数のサジェスチョン要素を提示することを引き起こすことができる。ユーザが、「私のテレビのチャンネルを変更して」というサジェスチョン要素のコンテンツを復唱するとき、自動アシスタントは、新しいテレビチャンネルがテレビにおいてレンダリングされるのと同時に提示され得る、追加のサジェスチョン要素も生成しながら、テレビチャンネルを変更させることができる。

【0055】

いくつかの実装形態では、アシスタントデータ422、アクションデータ444、および/またはクライアントデータ440などのデータは、ユーザと自動アシスタント408および/または自動アシスタント438との間の過去の対話の特徴づけることができる。サジェスチョンエンジン424は、上述のデータを使用して、アシスタントインターフェース436への不完全な口頭発話、または完全な口頭発話の提供に回答してユーザに提示される、サジェスチョン要素のためのコンテンツを生成することができる。このようにして、ユーザがいくつかの好ましいコマンドを一時的に忘れるならば、ユーザの特定の好ましいコマンドをユーザに思い出させるために、自然言語テキストが、サジェスチョン要素を介して提供され得る。そのような好ましいコマンドは、コンテンツを、前に受信された不完全、または完全な口頭発話と比較すること、およびコンテンツを、ユーザから受信された前の自然言語入力と比較することによって、識別され得る。最も頻繁に生じる自然言語入力が、好ましいコマンドに対応するとして考えられ得る。したがって、完全または不完全な口頭発話のコンテンツが、そのような自然言語入力に対応するとき、サジェスチョン要素は、それらの自然言語入力のコンテンツに基づいて、レンダリングおよび/または生成され得る。さらに、最も頻繁に処理される自然言語入力を識別すること、および/または、特定のユーザに関して好ましいコマンドを識別することによって、サジェスチョンエンジン424は、最も頻繁に、最もうまく実施されたコマンドに基づくサジェスチョンを、自動アシスタントを介して、ユーザに提供することができる。したがって、そのようなサジェスチョンを提供する結果として、最も成功しないコマンドの処理が低減されるようになるので、計算リソースが保護され得る。

【0056】

図5は、ユーザによって自動アシスタント、デバイス、アプリケーション、および/または任意の他の装置もしくはモジュールに提供された口頭発話を完成させる、かつ/または補足するための、1つまたは複数のサジェスチョン要素を提供するための方法500を示す。

10

20

30

40

50

方法500は、自動アシスタントに関連付けられ得る、1つまたは複数のコンピューティングデバイス、アプリケーション、および/または任意の他の装置もしくはモジュールによって実施され得る。方法500は、ユーザが口頭発話の少なくとも一部分を提供したと決定する動作502を含み得る。口頭発話は、ユーザによって、1つまたは複数のコンピューティングデバイスの1つまたは複数の自動アシスタントインターフェースに可聴的に提供された、口頭の自然言語入力であり得る。口頭発話は、1つまたは複数の語、1つまたは複数のフレーズ、あるいは口頭自然言語の任意の他の構成物を備え得る。たとえば、自動アシスタントインターフェースは、マイクロフォンを含むことがあり、コンピューティングデバイスは、別の自動アシスタントインターフェースとして動作することもできる、タッチディスプレイパネルを含む、スマートデバイスであり得る。いくつかの実装形態では、コンピューティングデバイス、またはコンピューティングデバイスと通信しているサーバデバイスは、ユーザが口頭発話を提供したと決定することができる。

10

【0057】

方法500は、口頭発話が完全であるか否かを決定する動作504をさらに含み得る。口頭発話が完全であるか否かを決定することは、機能の1つまたは複数のパラメータのうちの各パラメータが、割り当てられた値を有するか否かを決定することを含み得る。たとえば、口頭発話が「アシスタント、下げて」であるとき、ユーザは、デバイスの出力モダリティを下げるアクションが実行されることを要求中であり得る。アクションは、「音量を下げる」機能、または「明るさを下げる」機能などの機能に対応することがあり、それらの機能は、ユーザがデバイス名を指定することを必要とし得る。代替または追加として、口頭発話が完全であるか否かを決定することは、自動アシスタントが口頭発話の受信にตอบสนองして、アクションを実施することができるか否かを決定することを含み得る。

20

【0058】

場合によっては、動作504は、口頭発話が完全であるか否かを決定するために、アクションのいくつかのカテゴリーが、口頭発話にตอบสนองして実施され得るか否かを決定することを含み得る。たとえば、自動アシスタントが、明快のために、ユーザが言ったことを繰り返すように、ユーザに頼むなどのアクションは、自動アシスタントが口頭発話にตอบสนองするにもかかわらず、ユーザが完全な口頭発話を提供したことを示さないと考えられ得る。しかしながら、ウェブクエリを実施する、別のデバイスを制御する、アプリケーションを制御する、および/または、場合によっては意図されたアクションを実施するように自動アシスタントに要求するなどのアクションは、完全な口頭発話であると考えられ得る。たとえば、いくつかの実装形態では、口頭発話は、少なくとも1つの要求が口頭発話から識別され、要求が自動アシスタントを介して実施されることが可能であるとき、完全であると考えられ得る。

30

【0059】

口頭発話が完全(たとえば、「アシスタント、テレビを下げて」)であると決定されるとき、方法500は、実施されることになるアクションに基づいて、ユーザのためのサジェスチョンをレンダリングおよび/もしくは生成するオプション動作512、ならびに/または、口頭発話に対応するアクションを実施する動作514に進むことができる。たとえば、自動アシスタントインターフェースを介して、口頭発話を受信したコンピューティングデバイスは、口頭発話を介して指定された1つまたは複数のアクションの実施を初期化することができる。しかしながら、口頭発話が不完全であると決定されるとき、方法500は、動作506に進むことができる。

40

【0060】

動作506は、遅延のしきい値持続時間が、不完全な口頭発話の後に生じたか否かを決定することを含む、オプション動作であり得る。たとえば、口頭入力無音(spoken input silence)の持続時間がしきい値を満たすか否かは、ユーザからの許可とともに、口頭発話を提供中であるユーザ、ユーザのコンテキスト、および/または不完全な口頭発話に関連付けられた任意の他の情報にさらに基づいて決定され得る。口頭入力無音の量が、遅延のしきい値持続時間を満たすとき、方法500は、動作508に進むことができる。しかしながら、

50

口頭入力無音の量が、遅延のしきい値持続時間を満たさない場合、方法500は、動作510に進むことができる。

【0061】

方法500は、場合によっては、動作504から動作508に進むか、または場合によっては、動作504から動作506に進むことができる。動作506において、遅延のしきい値持続時間が満たされると決定されるとき、方法500は、口頭発話を完成させるための1つまたは複数のサジェスチョンをレンダリングする動作508に進むことができる。口頭発話を完成させるための1つまたは複数のサジェスチョンのコンテンツは、不完全な口頭発話を提供したユーザ、不完全な口頭発話のコンテンツ、それにおいて口頭発話が提供されたコンテキスト、口頭発話を受信したデバイスに関連付けられたデバイスポロジータを表現するデータ、ユーザと自動アシスタントと(または、1人もしくは複数のユーザと自動アシスタントと)の間の前の対話の特徴づける対話履歴データ、および/またはそれからサジェスチョンが現れ得る任意の他の情報に基づき得る。たとえば、不完全な口頭発話が「アシスタント、下げて」であるとき、ユーザに提供されたサジェスチョンのコンテンツは、「テレビ」を含むことがあり、そのことは、テレビおよび/またはユーザが、不完全な口頭発話を受信したコンピューティングデバイスと同じ室内に位置することを示す、デバイスポロジータデータに基づき得る。代替的に、不完全な口頭発話が「アシスタント、~にメッセージを送って」であるとき、ユーザに提供されたサジェスチョンのコンテンツは、それにおいてユーザがユーザの配偶者および兄弟にメッセージを送るように自動アシスタントに要求した、前のダイアログセッションの特徴づける、対話履歴データに基づき得る。したがって、サジェスチョンのコンテンツは、「私の兄弟」、および「私の妻」を含み得る。いくつかの実装形態では、1つまたは複数のサジェスチョン要素のコンテンツは、動作502の後であるが、オプション動作506の前、および動作508の前に決定および/または生成され得る。

【0062】

方法500は、動作506または動作508から、別の口頭発話がユーザによって提供されたか否かを決定する動作510に進むことができる。たとえば、ユーザは、不完全な口頭発話を提供した後、追加の口頭発話を提供することができる。そのような場合、方法500は、動作510から動作504に戻るように進むことができる。たとえば、ユーザは、サジェスチョン要素のコンテンツに対応する追加の口頭発話を提供し、それによって、ユーザがサジェスチョン要素を選択したことを自動アシスタントに示すことができる。これによって、口頭発話の編集物が生じることがあり、口頭発話の編集物は、最初の不完全な口頭発話と、ユーザが選択したサジェスチョン要素のコンテンツに対応する追加の口頭発話とから構成され得る。口頭発話のこの編集物は、口頭発話の編集物が完全な口頭発話に対応するかどうかを決定する動作504を受けることがある。実装形態では、動作510において、システムは、持続時間の間、別の口頭発話を待つことができ、他の口頭発話が持続時間において受信されない場合、動作514(または、512)に進むことができる。別の口頭発話が持続時間において受信される場合、システムは、動作504に戻るように進むことができる。それらの実装形態のうちいくつかでは、持続時間が動的に決定され得る。たとえば、持続時間は、すでに提供された口頭発話が「不完全な」要求であるか、「完全な要求」であるかに依存することがあり、かつ/または、すでに提供された口頭発話の「完全な要求」の特性に依存することがある。

【0063】

口頭発話の編集物が完全な口頭発話に対応するとき、方法500は、オプション動作512および/または動作514に進むことができる。しかしながら、口頭発話の編集物が完全な口頭発話であると決定されない場合、方法500は、場合によっては動作506に進むこと、および/または場合によっては動作508に進むことができる。たとえば、方法500が再び動作508に進むとき、口頭発話を完成させるためのサジェスチョンは、少なくとも、口頭発話の編集物のコンテンツに基づき得る。たとえば、ユーザが「私の兄弟」というサジェスチョンを選択する場合、「アシスタント、~にメッセージを送って」という不完全な口頭発

10

20

30

40

50

話と、「私の兄弟」との組合せが、提供されるさらなるサジェスションのための基礎として使用され得る。このようにして、ユーザが、ユーザのそれぞれの口頭発話を完成させるためのサジェスションを選択し続けるとき、サジェスションの循環が提供され得る。たとえば、後続のサジェスションは、ユーザが、自動アシスタントとは別個であるアプリケーションを介して、前に同じメッセージをユーザに送ったことを示す、対話履歴データに基づいて、「私は途中である」を含み得る。

【0064】

いくつかの実装形態では、方法500は、動作510から、実施されることになるアクションに基づいて、ユーザのためのサジェスションをレンダリングするオプション動作512に進むことができる。たとえば、自動アシスタントが、ユーザのためのアラームシステムをオンにするなどのアクションを実施するように要求されるとき、アクション、および/またはアクションに関連付けられた任意の情報が、ユーザがそれから選択するための追加のサジェスション要素を提供するための基礎として使用され得る。たとえば、自動アシスタントを介して実行されることになる完全な口頭発話が「アシスタント、私のアラームシステムをオンにして」である場合、1つまたは複数の他のサジェスション要素が、動作512においてレンダリングされ、「そして、～のために私のアラームを設定して」、「すべての私のライトをオフにして」、「そして、私のオーディオブックをオンにして」、および/またはユーザのための任意の他の好適なサジェスションコンテンツなどのコンテンツを含み得る。実装形態では、動作512において、システムは、さらなる口頭入力を受信に回答して、または持続時間における他の口頭発話の非受信に回答して、動作514に進む。それらの実装形態のうちいくつかでは、持続時間が動的に決定され得る。たとえば、持続時間は、すでに提供された口頭発話が「不完全な」要求を構成するか、「完全な要求」を構成するかに依存することがあり、かつ/または、すでに提供された口頭発話の「完全な要求」の特性に依存することがある。

【0065】

いくつかの実装形態では、サジェスションは、ユーザと自動アシスタントとの間の対話の量を低減し、それによって、電力、計算リソース、および/またはネットワークリソースを節約するために、レンダリングおよび/または生成され得る。たとえば、動作512においてレンダリングされた補足サジェスション要素のコンテンツは、「毎晩、PM10時の後に」を含み得る。このようにして、ユーザが、動作510において決定されるように、補足サジェスション要素を(たとえば、別の口頭発話を介して)選択したと決定される場合、方法500は、動作504に、およびその後で動作514に進み、自動アシスタントが毎晩、PM10時の後に、アラームシステムがオンにされる設定を生成することを引き起こすことができる。したがって、後続の晩の間に、ユーザは、「アシスタント、オンにして...」という同じ最初の口頭発話を提供する必要はなく、むしろ、単に選択された他のサジェスション要素を介して採用した設定に依拠することができる。

【0066】

ユーザが、最初の完全な口頭発話の後、追加の口頭発話を提供しなかったか、またはさもなければ、口頭発話の編集物に対応する1つまたは複数のコマンドが実行されることを選好することを示したとき、方法500は、動作510から、オプション動作512および/または動作514に進むことができる。たとえば、ユーザが、「アシスタント、変更して」などの口頭発話を提供し、後で「テレビのチャンネル」などのコンテンツを含むサジェスション要素を選択したとき、ユーザは、テレビを変更するための特定のチャンネルを指定しないにもかかわらず、その要求が実行されることを選定し得る。結果および動作514として、ユーザがコマンドにおける結果の命令をさらに狭くするためのいかなる補足サジェスション要素も選択しないにもかかわらず、テレビチャンネルが変更され得る(たとえば、チャンネル変更は、対話履歴データから導出された、ユーザの学習された選好に基づき得る)。

【0067】

図6は、例示的なコンピュータシステム610のブロック図である。コンピュータシステム610は、典型的には、バスサブシステム612を介していくつかの周辺デバイスと通信す

10

20

30

40

50

る、少なくとも1つのプロセッサ614を含む。これらの周辺デバイスは、たとえば、メモリ625およびファイル記憶サブシステム626を含む記憶サブシステム624と、ユーザインターフェース出力デバイス620と、ユーザインターフェース入力デバイス622と、ネットワークインターフェースサブシステム616とを含み得る。入力デバイスおよび出力デバイスは、コンピュータシステム610とのユーザ対話を可能にする。ネットワークインターフェースサブシステム616は、外部ネットワークへのインターフェースを提供し、他のコンピュータシステム内の対応するインターフェースデバイスに結合される。

【0068】

ユーザインターフェース入力デバイス622は、キーボード、およびマウス、トラックボール、タッチパッド、もしくはグラフィックタブレットなどのポインティングデバイス、スキャナ、ディスプレイに組み込まれたタッチスクリーン、および音声認識システム、マイクロフォンなどのオーディオ入力デバイス、ならびに/または他のタイプの入力デバイスを含み得る。一般に、「入力デバイス」という用語の使用は、コンピュータシステム610に、または通信ネットワーク上に情報を入力するための、すべての可能なタイプのデバイスおよび方法を含むものとする。

10

【0069】

ユーザインターフェース出力デバイス620は、ディスプレイサブシステム、プリンタ、ファックスマシン、またはオーディオ出力デバイスなどの非視覚的ディスプレイを含み得る。ディスプレイサブシステムは、陰極線管(CRT)、液晶ディスプレイ(LCD)などのフラットパネルデバイス、投影デバイス、または可視の画像を作成するための何らかの他の機構を含み得る。ディスプレイサブシステムはまた、オーディオ出力デバイスを介してなど、非視覚的ディスプレイを提供し得る。一般に、「出力デバイス」という用語の使用は、コンピュータシステム610からユーザに、または別のマシンもしくはコンピュータシステムに情報を出力するための、すべての可能なタイプのデバイスおよび方法を含むものとする。

20

【0070】

記憶サブシステム624は、本明細書で説明するモジュールの一部または全部の機能を提供する、プログラミングおよびデータ構造を記憶する。たとえば、記憶サブシステム624は、方法500、および/もしくは、コンピューティングデバイス114、サーバデバイス、コンピューティングデバイス216、アシスタント対話モジュール、クライアントデバイス434、コンピューティングデバイス314、サーバデバイス402、IoTデバイス442、自動アシスタント408、自動アシスタント438、自動アシスタント、ならびに/または、本明細書で説明する任意の他のデバイス、装置、アプリケーション、および/もしくはモジュールの、選択された態様を実施するための論理を含み得る。

30

【0071】

これらのソフトウェアモジュールは、一般に、プロセッサ614によって単独で、または他のプロセッサと組み合わせて実行される。記憶サブシステム624内で使用されるメモリ625は、プログラム実行中に命令およびデータを記憶するためのメインランダムアクセスメモリ(RAM)630と、固定された命令が記憶される読取り専用メモリ(ROM)632とを含む、いくつかのメモリを含み得る。ファイル記憶サブシステム626は、プログラムおよびデータファイルのための永続的記憶を提供することができ、ハードディスクドライブ、関連付けられたリムーバブル媒体を伴うフロッピーディスクドライブ、CD-ROMドライブ、光学ドライブ、またはリムーバブル媒体カートリッジを含み得る。いくつかの実装形態の機能を実装するモジュールは、記憶サブシステム624内のファイル記憶サブシステム626によって、またはプロセッサ614によってアクセス可能な他のマシン内に記憶され得る。

40

【0072】

バスサブシステム612は、コンピュータシステム610の様々な構成要素およびサブシステムを、意図されるように互いに通信させるための機構を提供する。バスサブシステム612は、単一のバスとして概略的に示されているが、バスサブシステムの代替実装形態は、複数のバスを使用し得る。

50

【 0 0 7 3 】

コンピュータシステム610は、ワークステーション、サーバ、コンピューティングクラスタ、ブレードサーバ、サーバファーム、または任意の他のデータ処理システムもしくはコンピューティングデバイスを含む、異なるタイプのものであり得る。コンピュータおよびネットワークの常に変化し続ける性質のために、図6に示されたコンピュータシステム610についての説明は、いくつかの実装形態を例示するための具体例にすぎないものである。図6に示されたコンピュータシステムよりも多いかまたは少ない構成要素を有する、コンピュータシステム610の多数の他の構成が可能である。

【 0 0 7 4 】

本明細書で説明するシステムが、ユーザ(または、本明細書では「参加者」と呼ぶことが多い)についての個人情報を収集するか、または個人情報を利用することがある状況では、ユーザには、プログラムまたは特徴がユーザ情報(たとえば、ユーザのソーシャルネットワーク、ソーシャルアクションもしくはアクティビティ、職業、ユーザの選好、またはユーザの現在の地理的ロケーションについての情報)を収集するか否かを制御するため、あるいは、ユーザにより関連があり得る、コンテンツサーバからのコンテンツを受信するか否か、および/またはどのように受信するかを制御するための機会が提供され得る。また、あるデータは、個人を特定可能な情報が除去されるように、記憶または使用される前に1つまたは複数の方法で扱われ得る。たとえば、ユーザのアイデンティティは、ユーザについて個人を特定可能な情報を決定することができないように扱われることがあり、または、ユーザの地理的ロケーションが一般化されることがあり、その場合、地理的ロケーション情報は、ユーザの特定の地理的ロケーションを決定することができないように(市、郵便番号、または州レベルなどまで)取得される。したがって、ユーザは、情報がユーザについてどのように収集および/または使用されるかに対する制御を有し得る。

【 0 0 7 5 】

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、ユーザによって提供された口頭発話の特徴づけるデータに対して、音声テキスト処理を実施することを含む。口頭発話は、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される。方法は、口頭発話の特徴づけるデータに対して、音声テキスト処理を実施することに基づいて、口頭発話が完全であるか否かを決定することであって、少なくとも、自動アシスタントが、自然言語コンテンツに基づいて、1つまたは複数のアクションが達成されることを引き起こすことができるか否かを決定することを含む、ことをさらに含む。方法は、口頭発話が不完全であると決定されるとき、口頭発話が不完全であるとの決定に回答して、コンピューティングデバイスのディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことをさらに含む。1つまたは複数のサジェスチョン要素は、ユーザによって自動アシスタントインターフェースに対して話される時、自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含む。方法は、コンピューティングデバイスのディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことの後に、ユーザが、特定のサジェスチョン要素の他の自然言語コンテンツに関連付けられる、別の口頭発話を提供したと決定することをさらに含む。方法は、ユーザが別の口頭発話を提供したとの決定に回答して、口頭発話および別の口頭発話の組合せが完全であるか否かを決定することをさらに含む。方法は、口頭発話および別の口頭発話の組合せが完全であると決定されるとき、1つまたは複数のアクションが、自然言語コンテンツおよび別の口頭発話に基づいて、自動アシスタントを介して実施されることを引き起こすことをさらに含む。

【 0 0 7 6 】

本明細書で開示する技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含み得る。

【 0 0 7 7 】

10

20

30

40

50

いくつかの実装形態では、ユーザによって提供された口頭発話の特徴づけるデータに対して、音声テキスト処理を実施することは、データから、1つまたは複数の候補テキストセグメントを生成することを含み、方法は、口頭発話が不完全であると決定されるとき、コンピューティングデバイスのディスプレイパネルが、1つまたは複数の候補テキストセグメントのうち少なくとも1つの候補テキストセグメントのグラフィカル表現を提供することを引き起こすことをさらに含む。それらの実装形態のいくつかのバージョンでは、方法は、口頭発話が、完全であり、1つまたは複数のアクションのためのすべての必須パラメータを含むと決定されるとき、コンピューティングデバイスのディスプレイパネルが、1つまたは複数の候補テキストセグメントのうち少なくとも1つの候補テキストセグメントのグラフィカル表現を提供することをバイパスすること、および、1つまたは複数のサジェスチョン要素を提供することをバイパスすることを引き起こすことをさらに含む。それらの実装形態のいくつかの他のバージョンでは、方法は、口頭発話が完全であると決定されるとき、1人または複数のユーザによって自動アシスタントインターフェースに対して話されるとき、ユーザからの口頭発話に関連付けられた音声テキスト処理の量と比較して、低減された量の音声テキスト処理を生じ、自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、自然言語コマンドを識別すること、自然言語コマンドを識別することに基づいて、自然言語コマンドの特徴づける、1つまたは複数の他のサジェスチョン要素を生成すること、および、口頭発話が完全であるとの決定に少なくとも応答して、コンピューティングデバイスのディスプレイパネルが、1つまたは複数の他のサジェスチョン要素を提供することを引き起こすことをさらに含む。

10

20

【0078】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、少なくとも1つのサジェスチョン要素が、ディスプレイパネルを介して提供されるとき、ユーザから期待される口頭発話のタイプに基づいて、複数の異なる音声テキスト処理モデルから、音声テキスト処理モデルを選択することをさらに含む。

【0079】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、1つまたは複数のサジェスチョン要素の1つまたは複数の用語の方に、および/あるいは、1つまたは複数のサジェスチョン要素に対応するコンテンツの1つまたは複数の予想されるタイプの方に、音声テキスト処理にバイアスをかけることをさらに含む。

30

【0080】

いくつかの実装形態では、方法は、口頭入力無音のしきい値持続時間が、口頭発話に後続したか否かを決定することをさらに含む。それらの実装形態のうちいくつかでは、コンピューティングデバイスのディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことは、少なくとも口頭入力無音のしきい値持続時間が口頭発話に後続したとの決定にさらに応答するものである。

【0081】

いくつかの実装形態では、音声テキスト処理を実施することは、第1の候補テキストセグメントおよび第2の候補テキストセグメントを決定することを含み、ここで、第1の候補テキストセグメントおよび第2の候補テキストセグメントが、口頭発話の異なる解釈に対応する。それらの実装形態のうちいくつかでは、特定のサジェスチョン要素が、第1の候補テキストセグメントに基づいて決定され、少なくとも1つの他のサジェスチョン要素が、第2の候補テキストセグメントに基づいて決定される。それらの実装形態のいくつかのバージョンでは、コンピューティングデバイスに接続されたディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことは、コンピューティングデバイスに接続されたディスプレイパネルが、特定のサジェスチョン要素に隣接して、第1の候補テキストセグメントをグラフィカルに表現し、少なくとも1つの他のサジェスチョン要素に隣接して、第2の候補テキストセグメントをグラフィカルに表現することを引き起こすことを含む。それらのバージョンのうちいくつかでは、ユーザが、特定のサジェスチョン要素の他の自然言語コンテンツに関連付けられる、別の口頭発話を提供した

40

50

と決定することは、別の口頭発話に基づいて、ユーザが第1の候補テキストセグメントを識別したか、第2の候補テキストセグメントを識別したかを決定することを含む。

【0082】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、口頭発話が不完全であるとの決定に回答して、それにおいて前のアクションが、自動アシスタントを介して達成されており、ユーザが、ユーザと自動アシスタントとの間の1つまたは複数の前の対話の間に、他の自然言語コンテンツの少なくとも一部分を識別した、1つまたは複数の前の対話の特徴づける履歴データに基づいて、他の自然言語コンテンツを生成することをさらに含む。

【0083】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、口頭発話が不完全であるとの決定に回答して、ユーザに関連付けられた様々なデバイス間の関係の特徴づけるデバイスポロジータータに基づいて、他の自然言語コンテンツを生成することであって、ここで、特定のサジェスチョン要素が、デバイスポロジータータに基づくものであり、ユーザに関連付けられた様々なデバイスのうちの1つまたは複数のデバイスを識別する、ことをさらに含む。

【0084】

いくつかの実装形態では、1つまたは複数のアクションが、自然言語コンテンツおよび別の口頭発話に基づいて、自動アシスタントを介して実施されることを引き起こすことは、デバイスを制御することを含む。

【0085】

いくつかの実装形態では、特定のサジェスチョン要素は、アクションを示すグラフィカル要素をさらに含む。

【0086】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、1つまたは複数のサジェスチョン要素のプロビジョニング後、別の口頭発話を待つための特定の持続時間を決定することであって、ここで、特定の持続時間が、口頭発話が不完全であると決定されることに基づいて決定される、ことをさらに含む。それらの実装形態のいくつかのバージョンでは、方法は、口頭発話が完全であると決定されるとき、完全な口頭発話に基づく、1つまたは複数の他のサジェスチョン要素を生成すること、コンピューティングデバイスのディスプレイパネルが、1つまたは複数の他のサジェスチョン要素を提供することを引き起こすこと、および、1つまたは複数の他のサジェスチョン要素のプロビジョニング後、さらなる口頭発話を待つための代替の特定の持続時間を決定することをさらに含む。代替の特定の持続時間は、特定の持続時間よりも短く、代替の特定の持続時間は、口頭発話が完全であると決定されることに基づいて決定される。

【0087】

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、自動アシスタントがアクションを実施することを引き起こすことを促進するために、ユーザによって提供された口頭発話の特徴づけるデータに対して、音声テキスト処理を実施することを含む。口頭発話は、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される。方法は、口頭発話の特徴づけるデータに対して、音声テキスト処理を実施することに基づいて、口頭発話が完全であるか否かを決定することであって、少なくとも、自然言語コンテンツに、アクションに関連付けられた機能を制御するための1つまたは複数のパラメータ値がないか否かを決定することを含む、ことをさらに含む。方法は、口頭発話が不完全であると決定されるとき、口頭発話が不完全であるとの決定に回答して、コンピューティングデバイスに接続されたディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことをさらに含む。1つまたは複数のサジェスチョン要素は、ユーザによって自動アシスタントインターフェースに対して話されるとき、自動アシスタントがアクションを完成させることを促進するために動作することを引き起こ

10

20

30

40

50

す、他の自然言語コンテンツを、ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含む。方法は、ユーザが、自動アシスタントインターフェースにおいて受信された別の口頭発話を介して、1つまたは複数のサジェスチョン要素のうちの特定のサジェスチョン要素を選択したと決定することをさらに含む。方法は、ユーザが特定のサジェスチョン要素を選択したとの決定に応答して、アクションが、特定のサジェスチョン要素の他の自然言語コンテンツと、口頭発話の自然言語コンテンツとに基づいて実施されることを引き起こすことをさらに含む。方法は、口頭発話が完全であると決定されるとき、口頭発話が完全であるとの決定に応答して、アクションが、口頭発話の自然言語コンテンツに基づいて実施されることを引き起こすことをさらに含む。

【0088】

本明細書で開示する技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含み得る。

【0089】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、ユーザが特定のサジェスチョン要素を選択したとの決定に応答して、他の自然言語コンテンツに関連付けられた優先度が、口頭発話が不完全であると決定される前に他の自然言語コンテンツに関連付けられた前の優先度から、修正されることを引き起こすことをさらに含む。それらの実装形態のうちのいくつかでは、ディスプレイパネルにおける、少なくとも1つの他のサジェスチョン要素に対する、特定のサジェスチョン要素の提示の順序は、他の自然言語コンテンツへの割り当てられた優先度に少なくとも部分的に基づく。

【0090】

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、自動アシスタントがアクションを実施することを引き起こすことを促進するために、ユーザによって提供された口頭発話を特徴づけるデータに対して、音声テキスト処理を実施することを含む。口頭発話は、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される。方法は、口頭発話を特徴づけるデータに対して、音声テキスト処理を実施することに基づいて、口頭発話が完全であるか否かを決定することであって、少なくとも、自然言語コンテンツに、アクションに関連付けられた機能を制御するための1つまたは複数のパラメータ値がないか否かを決定することを含む、ことをさらに含む。方法は、口頭発話が不完全であると決定されるとき、コンピューティングデバイスを介してアクセス可能であるコンテキストデータに基づいて、それにおいてユーザが自動アシスタントインターフェースに口頭発話を提供したコンテキストを特徴づける、コンテキストデータを決定すること、コンテキストデータに基づいて、ディスプレイパネルを介して、1つまたは複数のサジェスチョンを提示するための時間を決定することであって、ここで、コンテキストデータが、ユーザがコンテキストにおいて、自動アシスタントに別個の口頭発話を前に提供したことを示す、こと、および、ディスプレイパネルを介して、1つまたは複数のサジェスチョンを提示するための時間を決定することに基づいて、コンピューティングデバイスのディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことをさらに含む。1つまたは複数のサジェスチョン要素は、ユーザによって自動アシスタントインターフェースに対して話されるとき、自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含む。

【0091】

本明細書で開示する技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含み得る。

【0092】

いくつかの実装形態では、1つまたは複数のサジェスチョンを提示するための時間を決定することは、コンテキストデータを、ユーザ、あるいは1人または複数の他のユーザが、コンテキストにある間に1つまたは複数の他の口頭発話を提供した、前の事例を特徴づ

10

20

30

40

50

ける他のコンテキストデータと比較することを含む。

【0093】

いくつかの実装形態では、方法は、口頭発話が不完全であると決定されるとき、1人または複数のユーザによって自動アシスタントインターフェースに対して話されるとき、ユーザからの口頭発話に関連付けられた音声テキスト処理の量と比較して、低減された量の音声テキスト処理を生じ、自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、自然言語コマンドを識別することをさらに含む。自然言語コマンドは、1つまたは複数のサジェスチョン要素のうちの別のサジェスチョン要素のあるコンテンツによって組み入れられる。

【0094】

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、ユーザによって提供された口頭発話の特徴づけるオーディオデータに対して、音声テキスト処理を実施することを含む。口頭発話は、自然言語コンテンツを含み、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される。方法は、口頭発話の特徴づけるオーディオデータに対して、音声テキスト処理を実施することに基づいて、第1のアクションが、自然言語コンテンツに基づいて実施されることを引き起こすことをさらに含む。方法は、ユーザが、自然言語コンテンツを含む口頭発話を提供したとの決定に応答して、コンピューティングデバイスのディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすことをさらに含む。1つまたは複数のサジェスチョン要素は、ユーザによって自動アシスタントインターフェースに対して話されるとき、自動アシスタントがアクションを完成させることを促進するために動作することを引き起こす、他の自然言語コンテンツを、ディスプレイパネルを介して提供する、特定のサジェスチョン要素を含む。方法は、ユーザが、自動アシスタントインターフェースにおいて受信された後続の口頭発話を介して、1つまたは複数のサジェスチョン要素のうちの特定のサジェスチョン要素を選択したと決定すること、および、ユーザが特定のサジェスチョン要素を選択したとの決定に応答して、第2のアクションが、特定のサジェスチョン要素によって識別された他の自然言語コンテンツに基づいて実施されることを引き起こすことをさらに含む。

【0095】

本明細書で開示する技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含み得る。

【0096】

いくつかの実装形態では、第2のアクションの実施は、第1のアクションの実施から生じた他のアクションデータの変更であり、かつ/または第1のアクションの実施から生じた他のアクションデータを補足する、アクションデータを生じる。それらの実装形態のうちのいくつかでは、方法は、ユーザが口頭発話を提供したとの決定に応答して、他の自然言語コンテンツを生成することであって、ここで、他の自然言語コンテンツが、第2のアクションの実施中に使用されたパラメータのための少なくとも1つのサジェストされた値を識別する、ことをさらに含む。

【0097】

いくつかの実装形態では、第1のアクションが実施されることを引き起こすことは、ユーザに関連付けられたあらかじめ決定されたデフォルトコンテンツに基づくデータが、コンピューティングデバイスにおいて提供されることを引き起こすことを含む。それらの実装形態のうちのいくつかでは、方法は、ユーザが特定のサジェスチョン要素を選択することに基づいて、ユーザからの後続の不完全な要求が受信されるとき、修正されたデフォルトコンテンツが利用可能であることを促進するために、デフォルトコンテンツを修正することをさらに含む。

【0098】

いくつかの実装形態では、1つまたは複数のプロセッサによって実施される方法が提供され、ユーザによって提供された口頭発話を取り込むオーディオデータに対して、音声テ

10

20

30

40

50

キスト処理を実施することに基づいて、テキストを生成することを含む。口頭発話は、ディスプレイパネルに接続されるコンピューティングデバイスの自動アシスタントインターフェースを介して受信される。方法は、テキストに基づいて、口頭発話が完全であるか、不完全であるかを決定することをさらに含む。方法は、口頭発話が完全であると決定されるか、不完全であると決定されるかに基づいて、特定の持続時間を決定することをさらに含む。特定の持続時間は、口頭発話が不完全であると決定されるときであるよりも、口頭発話が完全であると決定されるときに短い。方法は、テキストに基づいて、1つまたは複数のサジェスチョン要素を生成することであって、ここで、1つまたは複数のサジェスチョン要素が、テキストと組み合わせられるとき、自動アシスタントが対応するアクションを促進するために動作することを引き起こす、対応する追加のテキストを各々示す、ことをさらに含む。方法は、コンピューティングデバイスのディスプレイパネルが、1つまたは複数のサジェスチョン要素を提供することを引き起こすこと、および、1つまたは複数のサジェスチョン要素のプロビジョニング後、決定された持続時間の間、さらなる口頭入力について監視することをさらに含む。方法は、さらなる口頭入力、持続時間内に受信されるとき、候補テキストと、さらなる口頭入力を取り込むさらなるオーディオデータに対して、音声テキスト処理を実施することから生成された、追加のテキストとに基づいて、自動アシスタントコマンドを生成すること、および、自動アシスタントが、自動アシスタントコマンドを実行することを引き起こすことをさらに含む。方法は、さらなる口頭入力、持続時間内に受信されないとき、自動アシスタントが、候補テキストのみに基づいて、代替コマンドを実行することを引き起こすことをさらに含む。

10

20

【0099】

本明細書で開示する技術のこれらおよび他の実装形態は、以下の特徴のうちの1つまたは複数を含み得る。

【0100】

いくつかの実装形態では、口頭発話が完全であると決定され、口頭発話が完全であると決定されるか、不完全であると決定されるかに基づいて、特定の持続時間を決定することは、口頭発話が完全であり、一般検索エージェントではない特定の自動アシスタントエージェントに向けられると決定されるとき、特定の持続時間が第1の特定の持続時間であると決定すること、および、口頭発話が完全であり、一般検索エージェントに向けられると決定されるとき、特定の持続時間が第2の特定の持続時間であると決定することをさらに含む。第2の特定の持続時間は、第1の特定の持続時間よりも長い。

30

【0101】

いくつかの実装形態では、口頭発話が完全であると決定され、口頭発話が完全であると決定されるか、不完全であると決定されるかに基づいて、特定の持続時間を決定することは、口頭発話が完全であり、すべての必須パラメータを含むと決定されるとき、特定の持続時間が第1の特定の持続時間であると決定すること、および、口頭発話が完全であるが、すべての必須パラメータを欠いていると決定されるとき、特定の持続時間が第2の特定の持続時間であると決定することをさらに含む。第2の特定の持続時間は、第1の特定の持続時間よりも長い。

【0102】

各実装形態では、口頭発話が不完全であると決定されることがあり、関連付けられた処理が実施されることがある。他の実装形態では、口頭発話が完全であると決定されることがあり、関連付けられた処理が実施されることがある。

40

【符号の説明】

【0103】

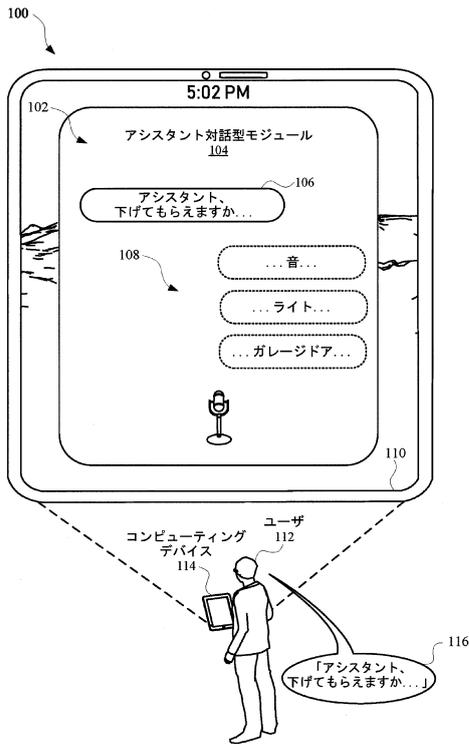
- 102、204 第1のユーザインターフェース
- 104、206、222、304 アシスタント対話型モジュール
- 106 グラフィカル要素、候補テキスト
- 108、210、310 サジェスチョン要素
- 110、212、320 ディスプレイパネル

50

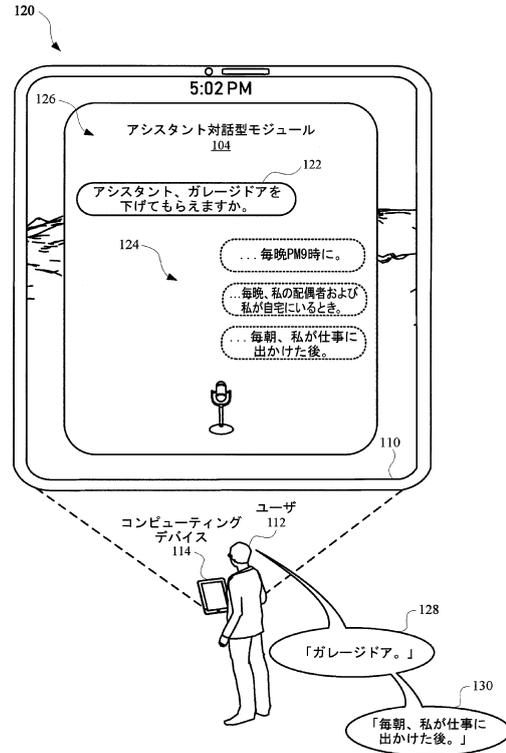
| | | |
|-------------|------------------------------|----|
| 112、214、312 | ユーザ | |
| 114、216、314 | コンピューティングデバイス | |
| 116 | 口頭発話、不完全な口頭発話、最初の口頭発話 | |
| 122 | 完成された候補テキスト、候補テキスト | |
| 124 | 追加のサジェスチョン要素、第1の追加のサジェスチョン要素 | |
| 126、232 | 第2のユーザインターフェース | |
| 128 | 追加の口頭発話、口頭発話 | |
| 130 | フレーズ | |
| 208 | 不完全なテキストセグメント | |
| 218 | 口頭発話、最初の口頭発話 | 10 |
| 224 | 完全な要求、完成された要求 | |
| 226 | 追加のサジェスチョン要素、サジェスチョン要素 | |
| 228、318 | 追加の口頭発話 | |
| 230 | 「私が寝るまで。」 | |
| 306 | グラフィカル表現 | |
| 308 | グラフィカルプロンプト、プロンプト | |
| 316 | 口頭発話 | |
| 400 | システム | |
| 402 | サーバデバイス | |
| 408 | 自動アシスタント | 20 |
| 412 | 入力処理エンジン | |
| 414 | 音声処理エンジン | |
| 416 | データパーシングエンジン | |
| 418 | アクションエンジン | |
| 420 | 出力生成エンジン | |
| 422 | アシスタントデータ | |
| 424 | サジェスチョンエンジン | |
| 426 | デバイスボロジエエンジン | |
| 428 | タイミングエンジン | |
| 430 | 音声テキストモデルエンジン | 30 |
| 432 | アクションエンジン | |
| 434 | クライアントデバイス | |
| 436 | アシスタントインターフェース | |
| 438 | 自動アシスタント、ローカル自動アシスタント | |
| 440 | クライアントデータ | |
| 442 | IoTデバイス | |
| 444 | ユーザインターフェース、アクションデータ | |
| 446 | ネットワーク | |
| 610 | コンピュータシステム | |
| 612 | バスサブシステム | 40 |
| 614 | プロセッサ | |
| 616 | ネットワークインターフェースサブシステム | |
| 620 | ユーザインターフェース出力デバイス | |
| 622 | ユーザインターフェース入力デバイス | |
| 624 | 記憶サブシステム | |
| 625 | メモリ | |
| 626 | ファイル記憶サブシステム | |
| 630 | メインランダムアクセスメモリ(RAM) | |
| 632 | 読取り専用メモリ(ROM) | |

【図面】

【図 1 A】



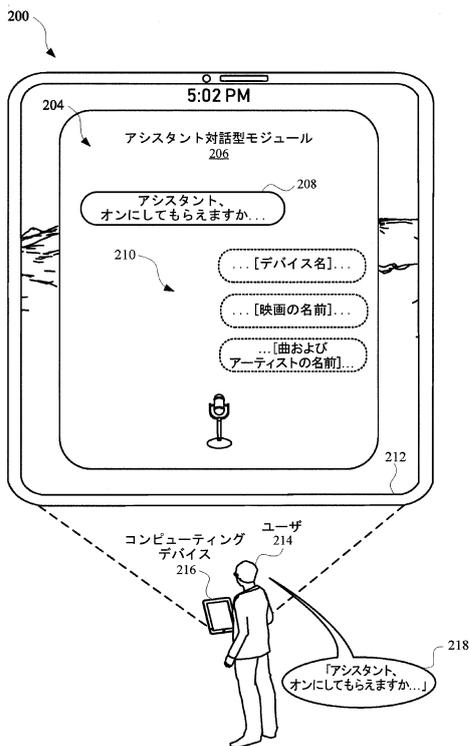
【図 1 B】



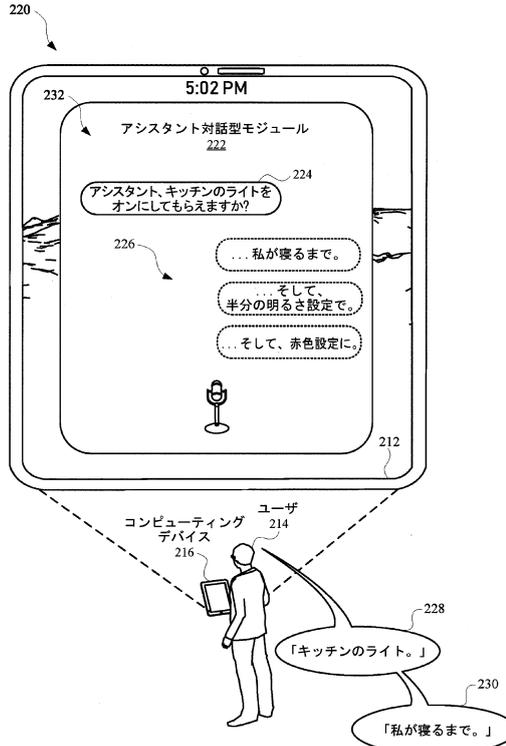
10

20

【図 2 A】



【図 2 B】

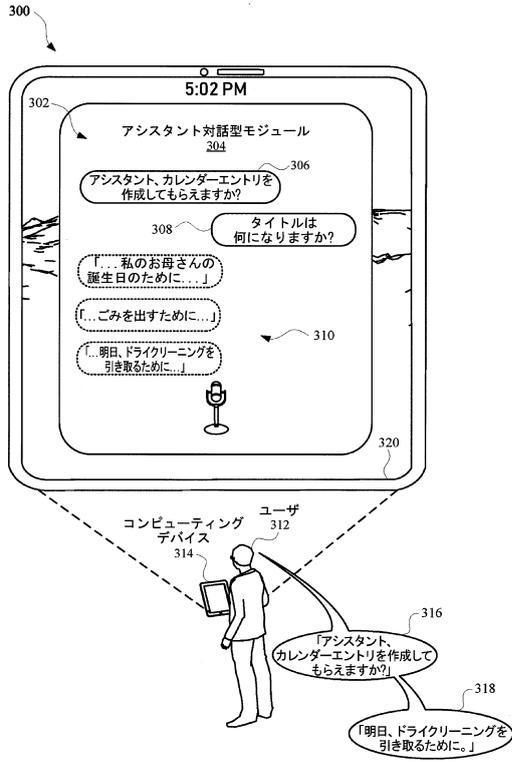


30

40

50

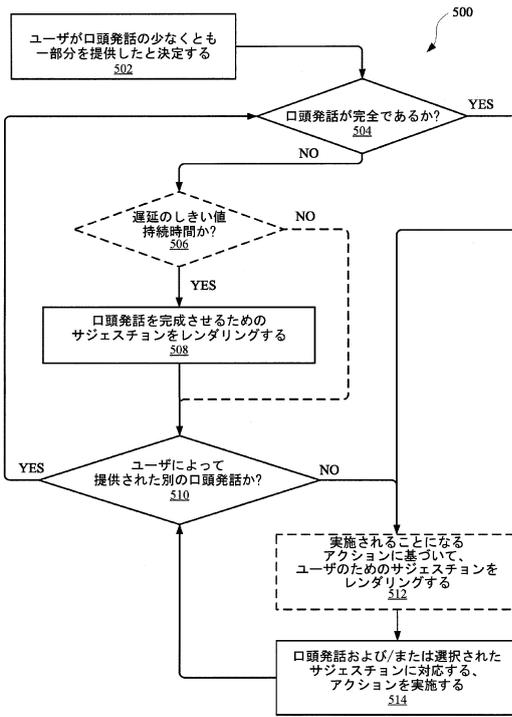
【図 3】



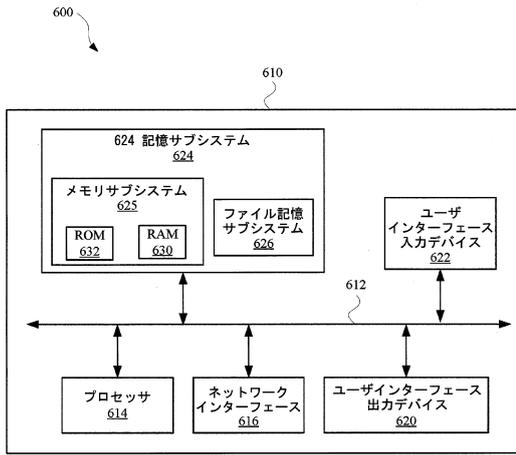
【図 4】



【図 5】



【図 6】



10

20

30

40

50

フロントページの続き

- (72)発明者 グレブ・スコベルツィン
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 オルガ・カブラロヴァ
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 コンスタンティン・シャギン
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 ウラジミール・ヴシュコヴィッチ
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 ユフェイ・ジャオ
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 ブラッドリー・ネルソン
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 アレッシオ・マクリ
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- (72)発明者 アブラハム・リー
アメリカ合衆国・カリフォルニア・94043・マウンテン・ビュー・アンフィシアター・パーク
ウェイ・1600
- 審査官 富澤 直樹
- (56)参考文献 特表2015-501022(JP,A)
特開2001-034292(JP,A)
特開2013-080476(JP,A)
特開2014-202857(JP,A)
特表2018-531404(JP,A)
特開2013-143151(JP,A)
- (58)調査した分野 (Int.Cl., DB名)
G10L 15/00-15/34
G06F 3/16