

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7585501号
(P7585501)

(45)発行日 令和6年11月18日(2024.11.18)

(24)登録日 令和6年11月8日(2024.11.8)

(51)国際特許分類	F I			
G 0 6 F 3/16 (2006.01)	G 0 6 F 3/16	6 5 0		
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10	2 0 0 W		
G 1 0 L 15/00 (2013.01)	G 1 0 L 15/00	2 0 0 N		
	G 0 6 F 3/16	6 3 0		
	G 0 6 F 3/16	6 2 0		
請求項の数 26 (全35頁)				

(21)出願番号	特願2023-537389(P2023-537389)	(73)特許権者	502208397
(86)(22)出願日	令和3年11月29日(2021.11.29)		グーグル エルエルシー
(65)公表番号	特表2024-505794(P2024-505794 A)		Google LLC
(43)公表日	令和6年2月8日(2024.2.8)		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/US2021/061000		1600 Amphitheatre P
(87)国際公開番号	WO2023/009157		arkway 94043 Mounta
(87)国際公開日	令和5年2月2日(2023.2.2)		in View, CA U.S.A.
審査請求日	令和5年8月10日(2023.8.10)	(74)代理人	100108453
(31)優先権主張番号	63/226,431		弁理士 村山 靖彦
(32)優先日	令和3年7月28日(2021.7.28)	(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
(31)優先権主張番号	17/532,199	(74)代理人	100133400
(32)優先日	令和3年11月22日(2021.11.22)		弁理士 阿部 達彦
最終頁に続く		最終頁に続く	

(54)【発明の名称】 ユーザが音声発話、または音声発話のシーケンスを反復的に与えるときの、自動アシスタントによるグラフィカルユーザインターフェース要素の動的適合

(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装される方法であって、
 コンピューティングデバイスにおいて、音声発話内に含まれる、ユーザからの要求の第1の部分を受け取るステップであって、前記音声発話が、前記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自動アシスタントに向けられたものである、ステップと、
 前記要求の前記第1の部分処理することに基づいて、前記要求の前記第1の部分に関連付けられる1つまたは複数の候補意図を決定するステップと、
 前記1つまたは複数の候補意図に基づいて、前記コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせるステップであって、
 前記汎用コンテナグラフィカル要素が、複数の異種の調整されたコンテナグラフィカル要素のいずれか1つに動的に適合され得、
 前記複数の異種の調整されたコンテナグラフィカル要素のそれぞれが、前記1つまたは複数の候補意図のうちの対応する1つに関連付けられ、前記要求に基づいて前記自動アシスタントが制御するアプリケーションまたはデバイスの設定の状態を示すグラフィカル要素の配列を含む、ステップと、
 前記コンピューティングデバイスにおいて、前記音声発話内に含まれ、または前記音声発話に続いて受け取られる追加の音声発話内に含まれる、前記ユーザからの前記要求の第2の部分を受け取るステップと、

前記要求の前記第2の部分処理することに基づいて、前記要求が前記1つまたは複数の候補意図のうち特定の意図に対応することを決定するステップと、

前記特定の意図に基づいて、前記ディスプレイインターフェースにおいてレンダリングされる前記汎用コンテナグラフィカル要素を、前記複数の異種の調整されたコンテナグラフィカル要素のうち特定の調整されたコンテナグラフィカル要素に動的に適合させるステップであって、前記特定の調整されたコンテナグラフィカル要素のみが前記ディスプレイインターフェースにレンダリングされる、ステップと

を含む方法。

【請求項2】

前記特定の調整されたコンテナグラフィカル要素が、前記特定の意図に関連付けられる、対応するパラメータについてのスロット値を特徴付け、

10

前記要求の前記第1の部分および前記要求の前記第2の部分が、前記スロット値を特定しない、請求項1に記載の方法。

【請求項3】

前記要求の前記第1の部分を受け取ったことに応答して、前記要求の前記第1の部分の特徴付ける自然言語コンテンツを前記コンピューティングデバイスの前記ディスプレイインターフェースに視覚的にレンダリングさせるステップをさらに含み、

前記ディスプレイインターフェースが、前記汎用コンテナグラフィカル要素をレンダリングすると同時に、前記要求の前記第1の部分の前記自然言語コンテンツをレンダリングする、請求項1または2に記載の方法。

20

【請求項4】

前記要求の前記第1の部分を受け取った後に、しきい持続時間が経過したと判定するステップをさらに含み、

前記汎用コンテナグラフィカル要素を前記ディスプレイインターフェースにおいてレンダリングさせる前記ステップが、前記しきい持続時間の経過に基づいて実施される、請求項1から3のいずれか一項に記載の方法。

【請求項5】

前記コンピューティングデバイスにおいて、前記音声発話、前記追加の音声発話、または前記音声発話もしくは前記追加の音声発話の後に受け取られるさらに追加の音声発話内に含まれる、前記ユーザからの前記要求の第3の部分を受け取るステップであって、前記要求の第3の部分が、前記1つまたは複数の設定についての更新後状態を含む、ステップと

30

、前記自動アシスタントにより、前記コンピューティングデバイスあるいは1つまたは複数の追加のコンピューティングデバイスの前記1つまたは複数の設定を、現在の状態から前記更新後状態に変更させるステップとをさらに含む、請求項1に記載の方法。

【請求項6】

前記ディスプレイインターフェースにおいてレンダリングされる前記汎用コンテナグラフィカル要素を前記特定の調整されたコンテナグラフィカル要素に動的に適合させる前記ステップが、

40

第2の音声発話内で識別されるスロット値のタイプに基づいて、前記複数の異種の調整されたコンテナグラフィカル要素の中から、前記特定の調整されたコンテナグラフィカル要素を選択するステップであって、

スロット値の前記タイプが数値の範囲に限定される数値に対応する、ステップを含む、請求項1から5のいずれか一項に記載の方法。

【請求項7】

前記汎用コンテナグラフィカル要素が、前記要求の前記第2の部分を受け取る前に、前記コンピューティングデバイスの前記ディスプレイインターフェースにおいてレンダリングされる、請求項1から6のいずれか一項に記載の方法。

【請求項8】

50

前記汎用コンテナグラフィカル要素が、前記要求の前記第2の部分が受け取られている間に、前記コンピューティングデバイスの前記ディスプレイインターフェースにおいてレンダリングされる、請求項1から7のいずれか一項に記載の方法。

【請求項9】

前記要求の前記第1の部分を処理することに基づいて、前記要求の前記第1の部分に関連付けられる1つまたは複数の候補意図を決定する前記ステップが、

ストリーミング自動音声認識(ASR)モデルを使用して、前記コンピューティングデバイスの1つまたは複数のマイクロフォンによって生成されたオーディオデータのストリームを処理し、ASR出力を生成するステップであって、オーディオデータの前記ストリームが、前記要求の前記第1の部分を取り込む、ステップと、

自然言語理解(NLU)モデルを使用して、前記ASR出力を処理し、NLU出力を生成するステップと、

前記NLU出力に基づいて、前記要求の前記第1の部分に関連付けられる前記1つまたは複数の候補意図を決定するステップと

を含む、請求項1から8のいずれか一項に記載の方法。

【請求項10】

前記要求の前記第2の部分を処理することに基づいて、前記要求が前記1つまたは複数の候補意図のうちの特定の意図に対応することを決定する前記ステップが、

ストリーミングASRモデルを使用して、オーディオデータの前記ストリームを処理し、追加のASR出力を生成するステップであって、オーディオデータの前記ストリームが、前記要求の前記第2の部分も取り込む、ステップと、

前記NLUモデルを使用して、前記追加のASR出力を処理し、追加のNLU出力を生成するステップと、

前記追加のNLU出力に基づいて、前記1つまたは複数の候補意図の中から前記特定の意図を選択するステップと

を含む、請求項9に記載の方法。

【請求項11】

1つまたは複数のプロセッサによって実装される方法であって、

コンピューティングデバイスにおいて、ユーザによってサブMITされた要求の部分を受け取るステップであって、前記要求の前記部分が、前記コンピューティングデバイスを介してアクセス可能な自動アシスタントに向けられた音声発話内に含まれる、ステップと、

前記自動アシスタントにより、前記要求の前記部分が、前記コンピューティングデバイスまたは前記自動アシスタントを介して前記コンピューティングデバイスと通信している1つまたは複数の追加のコンピューティングデバイスの1つまたは複数の設定の現在の状態を修正することに関連付けられると判定するステップと、

前記1つまたは複数の設定の前記現在の状態に基づいて、前記1つまたは複数の設定の前記現在の状態を特徴付ける、調整されたコンテナグラフィカル要素データを決定するステップと、

前記調整されたコンテナグラフィカル要素データに基づいて、前記1つまたは複数の設定の前記現在の状態を示す1つまたは複数の調整されたコンテナグラフィカル要素を前記コンピューティングデバイスのディスプレイインターフェースにレンダリングさせるステップであって、前記1つまたは複数の調整されたコンテナグラフィカル要素のそれぞれが、前記1つまたは複数の設定の現在の状態を示すグラフィカル要素の配列を含む、ステップと、

前記1つまたは複数の設定の前記現在の状態を示す前記1つまたは複数の調整されたコンテナグラフィカル要素を前記コンピューティングデバイスの前記ディスプレイインターフェースにレンダリングさせたことに応答して、

前記コンピューティングデバイスにおいて、前記ユーザによってサブMITされた前記要求の追加の部分を受け取るステップであって、前記要求の前記追加の部分が、前記音声発話、または前記音声発話の後に受け取られる追加の音声発話内に含まれ、前記要求の

10

20

30

40

50

前記追加の部分が、前記1つまたは複数の設定についての更新後状態を含む、ステップと、
前記自動アシスタントにより、前記コンピューティングデバイスあるいは前記1つまたは複数の追加のコンピューティングデバイスの前記1つまたは複数の設定を、前記現在の状態から前記更新後状態に変更させるステップと
を含み、

前記1つまたは複数の設定を前記現在の状態から前記更新後状態に変更させるステップに
応答して、前記更新後状態を示す制御グラフィカル要素の配列を含む特定の調整されたコン
テナグラフィカル要素がレンダリングされる、方法。

【請求項 1 2】

前記要求の前記部分を受け取ったことに応答して、前記要求の前記部分を特徴付ける自然
言語コンテンツを前記コンピューティングデバイスの前記ディスプレイインターフェー
スに視覚的にレンダリングさせるステップをさらに含み、

10

前記1つまたは複数の調整されたグラフィカルコンテナ要素が、前記ディスプレイイン
ターフェースが前記自然言語コンテンツをレンダリングすると同時にレンダリングされ
る、請求項 11 に記載の方法。

【請求項 1 3】

前記要求の前記部分が、前記1つまたは複数の設定の前記現在の状態を含まない、請求項
11 または 12 に記載の方法。

【請求項 1 4】

前記調整されたコンテナグラフィカル要素データに基づいて、前記1つまたは複数の設
定の前記現在の状態を示す前記1つまたは複数の調整されたコンテナグラフィカル要素を
前記コンピューティングデバイスの前記ディスプレイインターフェースにレンダリングさ
せる前記ステップが、

20

前記コンピューティングデバイスとは別の、前記1つまたは複数の追加のコンピュー
ティングデバイスのうちの第1のコンピューティングデバイスの前記1つまたは複数の設定
のうちの第1の設定を示す、前記1つまたは複数の調整されたコンテナグラフィカル要素の
うちの第1の調整されたコンテナグラフィカル要素をレンダリングさせるステップと、

前記コンピューティングデバイスとは別の、前記1つまたは複数の追加のコンピュー
ティングデバイスのうちの第2のコンピューティングデバイスの前記1つまたは複数の設定
のうちの第2の設定を示す、前記1つまたは複数の調整されたコンテナグラフィカル要素の
うちの第2の調整されたコンテナグラフィカル要素をレンダリングさせるステップと
を含む、請求項 11 から 13 のいずれか一項に記載の方法。

30

【請求項 1 5】

前記第1のコンピューティングデバイスの前記第1の設定が、前記第1のコンピューティ
ングデバイスの音量設定に対応し、

前記第2のコンピューティングデバイスの前記第2の設定が、前記第2のコンピューティ
ングデバイスの音量設定に対応する、請求項 14 に記載の方法。

【請求項 1 6】

前記第1のコンピューティングデバイスの前記第1の設定が、前記第1のコンピューティ
ングデバイスの輝度設定に対応し、

40

前記第2のコンピューティングデバイスの前記第2の設定が、前記第2のコンピューティ
ングデバイスの輝度設定に対応する、請求項 14 に記載の方法。

【請求項 1 7】

前記要求の前記部分を処理することに基づいて、前記ユーザが前記要求を与える際に休
止したと判定するステップをさらに含み、

前記1つまたは複数の設定の前記現在の状態を示す前記1つまたは複数の調整されたコン
テナグラフィカル要素を前記コンピューティングデバイスの前記ディスプレイインターフ
ェースにレンダリングさせる前記ステップが、前記ユーザが前記要求を与える際に休止し
たとの判定に応答するものである、請求項 11 から 16 のいずれか一項に記載の方法。

【請求項 1 8】

50

前記要求の前記部分を処理することに基づいて、前記ユーザが前記要求を与える際に休止したと判定する前記ステップが、

前記要求の前記部分を処理することに基づいて、前記ユーザが前記要求を与える際に特定のワードまたはフレーズを与えた後に休止したと判定するステップを含む、請求項17に記載の方法。

【請求項19】

前記1つまたは複数の設定の前記現在の状態を示す前記1つまたは複数の調整されたコンテナグラフィカル要素を前記コンピューティングデバイスの前記ディスプレイインターフェースにレンダリングさせる前記ステップが、前記ユーザが前記要求を与える際にしきい持続時間にわたって休止したとの判定に応答するものである、請求項17に記載の方法。

10

【請求項20】

前記要求の前記部分を処理することに基づいて、前記ユーザが前記要求を与える際に休止したと判定する前記ステップが、

ストリーミング自動音声認識(ASR)モデルを使用して、前記コンピューティングデバイスの1つまたは複数のマイクロフォンによって生成されたオーディオデータのストリームを処理し、ASR出力を生成するステップであって、オーディオデータの前記ストリームが、前記要求の前記部分を取り込む、ステップと、

自然言語理解(NLU)モデルを使用して、前記ASR出力を処理し、NLU出力を生成するステップと、

前記NLU出力に基づいて、前記ユーザが前記要求を与える際に休止したと判定するステップと

20

を含む、請求項17に記載の方法。

【請求項21】

前記要求の前記部分を処理することに基づいて、前記ユーザが前記要求を与える際に休止したと判定する前記ステップが、

前記要求の前記部分を処理することに基づいて、前記要求の前記部分に関連付けられるオーディオベースの特性を決定するステップと、

前記要求の前記部分に関連付けられる前記オーディオベースの特性に基づいて、前記ユーザが前記要求を与える際に休止したと判定するステップと

を含む、請求項17に記載の方法。

30

【請求項22】

1つまたは複数のプロセッサによって実装される方法であって、

コンピューティングデバイスにおいて、音声発話内に含まれる、ユーザからの要求の第1の部分を受け取るステップであって、前記音声発話が、前記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自動アシスタントに向けられたものである、ステップと、

前記要求の前記第1の部分の処理することに基づいて、前記要求の前記部分が、前記コンピューティングデバイスまたは前記コンピューティングデバイスと通信している1つまたは複数の追加のコンピューティングデバイスを制御することに関連付けられる特定のワードまたはフレーズを含むと判定するステップと、

前記特定のワードまたはフレーズを含む前記要求の前記第1の部分に基づいて、前記コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせるステップであって、

40

前記汎用コンテナグラフィカル要素が、複数の異種の調整されたコンテナグラフィカル要素のいずれか1つに動的に適合され得、

前記複数の異種の調整されたコンテナグラフィカル要素のそれぞれが、前記音声発話の前記第1の部分の処理することに基づいて決定された、対応する意図に関連付けられ、前記要求に基づいて前記自動アシスタントが制御するアプリケーションまたはデバイスの設定の状態を示すグラフィカル要素の配列を含む、ステップと、

前記コンピューティングデバイスにおいて、前記音声発話内に含まれ、または前記音声発話に続いて受け取られる追加の音声発話内に含まれる、前記ユーザからの前記要求の第

50

2の部分を受け取るステップと、

前記要求の前記第2の部分进行处理することに基づいて、前記要求が前記1つまたは複数の候補意図のうち特定の意図に対応することを決定するステップと、

前記特定の意図に基づいて、前記ディスプレイインターフェースにおいてレンダリングされる前記汎用コンテナグラフィカル要素を、前記複数の異種の調整されたコンテナグラフィカル要素のうち特定の調整されたコンテナグラフィカル要素に動的に適合させるステップであって、前記特定の調整されたコンテナグラフィカル要素のみが前記ディスプレイインターフェースにレンダリングされる、ステップと

を含む方法。

【請求項 2 3】

前記要求の前記第1の部分进行处理することに基づいて、前記要求の前記部分が、前記コンピューティングデバイスまたは前記コンピューティングデバイスと通信している前記1つまたは複数の追加のコンピューティングデバイスを制御することに関連付けられる前記特定のワードまたはフレーズを含むと判定する前記ステップが、

ストリーミング自動音声認識(ASR)モデルを使用して、前記コンピューティングデバイスの1つまたは複数のマイクロフォンによって生成されたオーディオデータのストリームを処理し、ASR出力を生成するステップであって、オーディオデータの前記ストリームが、前記要求の前記第1の部分を取り込む、ステップと、

前記ASR出力に基づいて、前記要求の前記部分が前記コンピューティングデバイスまたは前記1つまたは複数の追加のコンピューティングデバイスを制御することに関連付けられる前記特定のワードまたはフレーズを含むと判定するステップと

を含む、請求項22に記載の方法。

【請求項 2 4】

前記特定のワードまたはフレーズを含む前記要求の前記第1の部分に基づいて、前記コンピューティングデバイスの前記ディスプレイインターフェースにおいて前記汎用コンテナグラフィカル要素をレンダリングさせる前記ステップが、

前記特定のワードまたはフレーズが前記コンピューティングデバイスのオンデバイスメモリ内で前記汎用コンテナグラフィカル要素にマッピングされると判定するステップと、

前記特定のワードまたはフレーズが前記汎用コンテナグラフィカル要素にマッピングされるとの判定に応答して、

自然言語理解(NLU)モデルを使用して前記ASR出力を処理することなく、前記コンピューティングデバイスの前記ディスプレイインターフェースにおいて前記汎用コンテナグラフィカル要素をレンダリングさせるステップと

を含む、請求項23に記載の方法。

【請求項 2 5】

少なくとも1つのプロセッサと、

実行されるとき、請求項1から24のいずれか一項に対応する動作を前記少なくとも1つのプロセッサに実行させる命令を記憶するメモリと

を備えるシステム。

【請求項 2 6】

実行されるとき、請求項1から24のいずれか一項に対応する動作を少なくとも1つのプロセッサに実行させる命令を記憶する非一時的コンピュータ可読記憶媒体。

【発明の詳細な説明】

【背景技術】

【0001】

人間は、本明細書において「自動アシスタント」と呼ばれる(「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」、「会話型エージェント」などとも呼ばれる)対話型ソフトウェアアプリケーションとの間で人間-コンピュータ対話を行い得る。たとえば、人間(自動アシスタントと対話するとき、「ユーザ」と呼ばれることが

10

20

30

40

50

ある)は、いくつかのケースではテキストに変換され、次いで処理され得る音声自然言語入力(すなわち、発話)を使用して、かつ/またはテキスト(たとえば、タイプ入力された)自然言語入力を与えることによって、自動アシスタントにコマンドおよび/または要求を与え得る。

【0002】

多くの場合、自動アシスタントは、完全なコマンドフレーズを念頭に置いていないことがあるユーザによって呼び出され得る。たとえば、自動アシスタントが対応することを求める要求の部分を含む、「アシスタント、...を設定して(Assistant, set ...)」という音声発話をユーザが与えると仮定する。この例では、要求の部分は、ユーザがスマートスピーカについての音量を設定し、スマートサーモスタットについての温度を設定し、スマート電球についての輝度レベルを設定することなどを意図することを示し得る。しかしながら、こうした例の多くでは、ユーザが一定の時間窓以内に完全なコマンドフレーズを発音しない場合、曖昧過ぎるので自動アシスタントは要求に基づいて対応しないことがあり、または自動アシスタントは、要求に基づいて対応し、ユーザが意図していなかった何らかのアクションを実施し得る。したがって、ユーザは、自動アシスタントを再呼出して、完全なコマンドフレーズを用いて音声発話を再び与える必要があり得、それによって、ユーザと自動アシスタントとの間の1つまたは複数の対話セッションが延長され、コンピューティングデバイスにおいて受け取られるユーザ入力の量が増大する。

【0003】

ある場合には、自動アシスタントは、完全なコマンドフレーズを念頭に置いていても、コマンドフレーズに関連する特定のスロット値を念頭に置いていないことがあるユーザによって呼び出され得る。たとえば、自動アシスタントが対応することを求める要求の部分を含む「アシスタント、音量を...に設定して(Assistant, set the volume to ...)」という音声発話をユーザが与えると仮定する。この例では、要求の部分は、ユーザが、スピーカについての音量を、スピーカについての音量を設定するためのコマンドに関連する特定のスロット値として設定することを意図することを示し得る。しかしながら、こうした例の多くでは、ユーザは、スピーカの音量の現在の状態を知らないことがあり、その結果、現在のステータスに対してスピーカの音量をどのように修正するかを知らないことがある。したがって、ユーザは、音声発話の最初の部分(たとえば、「アシスタント、音量を...に設定して」)を与えた後に、どのようにスピーカの音量を修正するかについて考えるために休止し得る。上記の例と同様に、ユーザが一定の時間窓以内に特定のスロット値を発音しない場合、曖昧過ぎるので自動アシスタントは要求に基づいて対応しないことがあり、または自動アシスタントは、要求に基づいて対応し、ユーザが意図していなかった何らかのアクションを実施し得る。この場合も、ユーザは、自動アシスタントを再呼出して、完全なコマンドフレーズおよび特定のスロット値を用いて音声発話を再び与える必要があり得、それによって、ユーザと自動アシスタントとの間の1つまたは複数の対話セッションが延長され、コンピューティングデバイスにおいて受け取られるユーザ入力の量が増大する。その結果、こうした例では、不完全な音声発話の処理によって計算資源が浪費され得、ユーザが自動アシスタントとの対話セッションを再び行うことが必要となる。

【発明の概要】

【課題を解決するための手段】

【0004】

本明細書において説明される実装は、ユーザのコンピューティングデバイスにおいて少なくとも部分的に実行される自動アシスタントに向けられた要求を含む音声発話、または音声発話のシーケンスをユーザが反復的に与えることに基づいて、グラフィカルユーザインターフェース(GUI)要素を動的に適合させ得る自動アシスタントに関する。GUI要素は、ユーザからの着信要求の部分の特徴付け得、かつ/またはユーザが自動アシスタントに対する要求をより効率的かつ正確に説明するのを援助し得る提案を提供し得る。いくつかの実装では、要求の部分処理することに基づいて、候補意図が決定され得、ユーザが要求を完了する前に、候補意図に関連する汎用コンテナグラフィカル要素が、コンピューティン

10

20

30

40

50

グデバイスのディスプレイインターフェースにおいてレンダリングされ得る。さらに、要求の追加の部分処理することに基づいて、候補意図から特定の意図が決定され得、汎用コンテナグラフィカル要素が、コンピューティングデバイスにおいて異なるディスプレイインターフェースをレンダリングすることなく、特定の意図に関連する特定の調整されたコンテナグラフィカル要素に動的に適合され得る。追加または代替の実装では、要求の部分に含まれる特定のワードまたはフレーズが、候補意図を決定することを必要とせずに、汎用コンテナグラフィカル要素に直接的にマッピングされ得る。追加または代替の実装では、特定の調整されたコンテナグラフィカル要素は、要求の部分がコンピューティングデバイスおよび/または追加のコンピューティングデバイスの1つまたは複数の設定の現在の状態を修正することに関連するとの判定に回答して、コンピューティングデバイスおよび/

10

【0005】

たとえば、スマートスピーカについての音量、スマートサーモスタットについての温度、スマート電球についての輝度レベルなどのデバイスの状態を自動アシスタントが調節することを求める要求の部分を含む、「アシスタント、...を設定して」という音声発話を与えることをユーザが開始すると仮定する。ユーザが要求の部分を与えるとき、自動アシスタントは、ストリーミング自動音声認識(ASR)モデルを使用して、要求の部分を取り込むオーディオデータのストリームを処理させ、ASR出力を生成し得る。さらに、自動アシスタントは、自然言語理解(NLU)モデルを使用して、ASR出力を処理させ、NLU出力を生成し得る。汎用コンテナグラフィカル要素は、(たとえば、要求の部分が「設定して(set)」または別の特定のワードまたはフレーズを含むことを示す)ASR出力および/または(たとえば、要求の部分が汎用コンテナグラフィカル要素に関連する候補意図を含むことを示す)NLU出力に基づいて、コンピューティングデバイスのディスプレイインターフェースにおいてレンダリングされ得る。

20

【0006】

さらに、自動アシスタントがデバイスの状態を調節することを求める要求の追加の部分を含む、「...スピーカについての音量...(the volume for the speakers ...)」という音声発話を(または音声発話に続く追加の音声発話として)ユーザが引き続き与えると仮定する。同様に、ユーザが要求の追加の部分を与えるとき、自動アシスタントは、ストリーミングASRモデルを使用して、要求の追加の部分をも取り込むオーディオデータのストリームを処理させ、追加のASR出力を生成し得る。さらに、自動アシスタントは、NLUモデルを使用して、追加のASR出力を処理させ、追加のNLU出力を生成し得る。追加のASR出力および/または追加のNLU出力に基づいて、自動アシスタントは、ユーザがスマートスピーカについての音量を設定したいと判定し得る。したがって、汎用コンテナグラフィカル要素は、スマートスピーカについての音量を設定することに特有の、調整されたコンテナグラフィカル要素に動的に適合され得る。たとえば、スマートスピーカについての音量を設定することに特有の、調整されたコンテナグラフィカル要素は、スマートスピーカについての音量の現在の状態、タッチ入力を使用してユーザが音量を設定することを可能にする音量制御グラフィカル要素、音量がスマートスピーカについて設定されていることを示すメディアコンテンツ、スマートスピーカに関連付けられるデバイス識別子、および/またはスマートスピーカに関連付けられる任意の他のコンテンツを含み得る。

30

40

【0007】

一方、音声発話を与えることを続行する際に、ユーザが「...温度...(the temperature ...)」を示した場合、汎用コンテナグラフィカル要素が、スマートスピーカについての音量を設定することに特有の、調整されたコンテナグラフィカル要素とは別個の、スマートサーモスタットについての温度を設定することに特有の、調整されたコンテナグラフィカル要素に動的に適合され得る。たとえば、スマートサーモスタットについての温度を設定することに特有の、調整されたコンテナグラフィカル要素は、温度の現在の状態、温度が

50

スマートサーモスタットについて設定されていることを示すメディアコンテンツ、タッチ入力を使用してユーザが温度を設定することを可能にする温度制御グラフィカル要素、スマートサーモスタットに関連付けられるデバイス識別子、および/またはスマートサーモスタットに関連付けられる任意の他のコンテンツを含み得る。それでも、こうした例のどちらでも、どんな追加のユーザインターフェースもレンダリングすることなく、同一の汎用コンテナグラフィカル要素が、こうした様々な調整されたコンテナグラフィカル要素に動的に適合され得る。

【0008】

汎用コンテナグラフィカル要素は、複数の異種の意図のうちの対応する1つにそれぞれ関連付けられ、または特定のワードまたはフレーズに直接的にマッピングされる複数の異種の調整されたコンテナグラフィカル要素のいずれか1つについてのブレースホルダとして動作し得る。したがって、ユーザが自動アシスタントに要求を引き続き与えるとき、汎用コンテナグラフィカル要素は、要求の追加の部分を処理することに基づいて決定された特定の意図に関連付けられる特定の調整されたグラフィカル要素に動的かつシームレスに適合され得る。たとえば、上記の例では、「アシスタント、...を設定して」という要求の部分に回答して、汎用コンテナグラフィカル要素が、当初はディスプレイインターフェースにおいてレンダリングされ得る。汎用コンテナグラフィカル要素は、たとえば、値の範囲を示すためのグラフィカル要素の配列(たとえば、ドット形状の配列)を含み得る。その後で、ユーザが要求の追加の部分(たとえば、「...スピーカの音量...」、「...輝度...」、「...温度...」など)を与えたとき、自動アシスタントは、要求の追加の部分に基づいてグラフィカル要素の配列を適合させ得る。たとえば、ユーザが「...スピーカの音量...」という要求の追加の部分を与えたことに基づいて、グラフィカル要素の配列が、ユーザがどのように音量を修正するかを決定するのを援助するために、スマートスピーカの音量に関連付けられる値の範囲を反映し、スマートスピーカの音量の現在の状態を含むように適合され得る。さらに、たとえば、ユーザが「...輝度...」という要求の追加の部分を与えたことに基づいて、グラフィカル要素の配列が、ユーザがどのように輝度を修正するかを決定するのを援助するために、スマート電球の輝度に関連付けられる値の範囲を反映し、スマート電球の輝度の現在の状態を含むように適合され得る。

【0009】

いくつかの実装では、自動アシスタントは、候補意図を処理して、要求を与える際にユーザが制御しようと試みている可能性がある特定のデバイスおよび/またはアプリケーションを識別し得る。自動アシスタントが特定のデバイスおよび/またはアプリケーションを識別したとき、自動アシスタントは、汎用コンテナグラフィカル要素内のグラフィカル要素の配列に、特定のアプリケーションおよび/またはデバイスの現在の状態を表現させ、調整されたコンテナグラフィカル要素がもたらされ得る。たとえば、グラフィカル要素の配列は、7つの塗りつぶされた円と、その後続く3つの空の円を含み得、それによって、特定のデバイスおよび/またはアプリケーションに関連付けられるスマート電球が、スマート電球の輝度設定の現在の状態として、現在は最大輝度レベルの70%であることを示す。代替または追加として、自動アシスタントは、ユーザが参照していると予測される特定のデバイスおよび/またはアプリケーションを表すアイコン(たとえば、台所の照明を表すアイコン)を識別し得る。自動アシスタントは、ユーザからの要求に回答して制御するように自動アシスタントが選択した特定のデバイスおよび/またはアプリケーションを識別するために、調整されたコンテナグラフィカル要素に対するアイコンを含み得る。このようにして、ユーザは、(たとえば、輝度を70%から50%に変更するように)特定のアプリケーションおよび/またはデバイスを指定するために音声発話を介して要求の別の部分を与えることを回避し、その代わりにタッチ入力を利用することを選び得、それによって、普通なら音声発話または追加の音声発話を処理するときに消費されることとなるような計算資源が保持される。

【0010】

いくつかの実装では、ユーザは、1つまたは複数の追加の音声発話を介して要求を完了

10

20

30

40

50

することによって、調整されたコンテナグラフィカル要素内に含まれるグラフィカル要素の配列によって示される現在の状態を目の当たりにすることに基づいて、自動アシスタントに特定のデバイスおよび/またはアプリケーションを制御させ得る。たとえば、調整されたコンテナグラフィカル要素内に含まれるグラフィカル要素の配列を目の当たりにすることによって、ユーザは、ユーザの要求の最終部分を考慮し得る。ユーザは、「...30%に(... to 30%)」などの最終音声発話を与え、それによって、輝度レベルを70%から30%に調節するように特定のデバイスおよび/またはアプリケーションを制御するように自動アシスタントに指示し得る。代替または追加として、ユーザは、「30%ドット」に対応するグラフィカル要素の配列の部分をタップして、自動アシスタントに輝度レベルを70%から30%に同様に調節させ得る。

10

【0011】

いくつかの実装では、要求の最初の部分がユーザによって与えられたことに応答して、自動アシスタントは、複数の調整されたコンテナ要素をレンダリングさせ得る。たとえば、「アシスタント、...の[曲名1]を再生して(Assistant, play [song title 1] by ...)」などの音声発話を介してユーザが要求の最初の部分を与えたとき、自動アシスタントは、コンピューティングデバイスのディスプレイインターフェースにおいて、複数の異なる調整されたコンテナ要素をレンダリングさせ得る。調整されたコンテナグラフィカル要素のそれぞれは、要求に関連付けられ得る、異なる動作および/または解釈に対応し得る。たとえば、第1の調整されたコンテナグラフィカル要素は、コンピューティングデバイスまたは追加のコンピューティングデバイスにおいて「[アーティスト1]」による「[曲名1]」を再生するための動作に対応し得、第2の調整されたコンテナグラフィカル要素は、コンピューティングデバイスまたは追加のコンピューティングデバイスにおいて「[アーティスト2]」による「[曲名1]」を再生するための別の動作に対応し得る。いくつかの実装では、追加または代替として、調整されたコンテナグラフィカル要素のそれぞれは、コンピューティングデバイスまたは追加のコンピューティングデバイスの現在の状態(たとえば、第1のデバイスおよび/または第2のデバイスにおいて何が現在再生されているか)を含む。ユーザは(たとえば、音声発話および/またはタッチ入力を介して)要求を完了し得、自動アシスタントは、それに応じて要求を履行させ得る。

20

【0012】

いくつかの実装では、ストリーミングASRモデルを使用して生成されたASR出力は、たとえば、要求の様々な部分に対応すると予測される予測音声仮説、要求の様々な部分に対応すると予測される予測音素、予測音声仮説および/または予測音素が要求の様々な部分に対応する可能性がどれほどであるかを示す予測ASR尺度、および/または他のASR出力を含み得る。さらに、NLUモデルを使用して生成されたNLU出力は、たとえば、要求の様々な部分を与える際のユーザの実際の意図に対応すると予測される候補意図、候補意図に関連付けられる、対応するパラメータについての1つまたは複数のスロット値、および/または他のNLU出力を含み得る。さらに、NLU出力に基づいて1つまたは複数の構造化要求が生成され、様々なデバイスおよび/またはアプリケーションによって処理されて、要求についての履行データが生成され得る。履行データは、実装されるとき、ユーザによって与えられた要求を自動アシスタントに履行させ得る。

30

40

【0013】

いくつかの実装では、本明細書において説明される汎用コンテナグラフィカル要素および/または調整されたコンテナグラフィカル要素は、ユーザが要求を与えることを休止したとの判定に応答して、コンピューティングデバイスのディスプレイインターフェースにおいてレンダリングされるだけであり得る。自動アシスタントは、たとえば、コンピューティングデバイスにおいて受け取られる要求の部分に関連付けられるNLUデータおよび/またはオーディオベースの特性に基づいて、ユーザが要求を与えることを休止したと判定し得る。要求の部分に関連付けられるオーディオベースの特性は、イントネーション、音調、強勢、リズム、テンポ、ピッチ、および引き延ばされた音節のうちの1つまたは複数を含み得る。たとえば、自動アシスタントに向けられた音声発話内に含まれる、「アシスタ

50

ント、音量を...に設定して」という要求をユーザが与えると仮定する。この例では、自動アシスタントは、たとえば、前述の「に(to)」から経過するしきい持続時間と、予測されるスマートスピーカの音量を変更する意図に関連付けられる音量パラメータについてのスロット値をユーザが与えなかったことを示すNLU出力とに基づいて、ユーザが休止したと判定し得る。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のディスプレイインターフェースにおいて、スマートスピーカについての音量コンテナグラフィカル要素をレンダリングさせ得る。コンピューティングデバイスのディスプレイインターフェースにおいてレンダリングされる、スマートスピーカについての音量コンテナグラフィカル要素は、ユーザがどのように現在の音量に対して音量を修正するかを決定するのを援助するために、スマートスピーカの現在の音量を含み得る。代替または追加として、「に(to)」を与える際にユーザが引き延ばされた音節を含める(たとえば、「アシスタント、音量を...に設定して(Assistant, set the volume toooooo ...)」)とさらに仮定する。この例では、自動アシスタントは、たとえば、少なくとも要求を与える際の引き延ばされた音節に基づく、現在の音量に対して音量をどのように修正するかに関する不確実性を反映するオーディオベースの特性に基づいて、ユーザが休止したと判定し得る。したがって、音量コンテナグラフィカル要素は、ユーザがどのように現在の音量に対して音量を修正するかを決定するのを援助し得る。

10

【0014】

本明細書において説明される技法を使用することにより、様々な技術的利点が達成され得る。非限定的な一例として、本明細書において説明される技法により、ユーザが音声発話または音声発話のシーケンスを与える間に、自動アシスタントが、汎用GUI要素からの様々なGUI要素を調整されたGUI要素に動的に適合させることが可能となる。たとえば、ユーザは、要求の部分を含む音声発話を与え得、自動アシスタントは汎用GUI要素をレンダリングし得、その後で、汎用GUI要素は、音声発話の追加の部分または追加の音声発話を処理することに基づいて、調整されたGUI要素に適合される。こうした調整されたGUI要素は、ユーザが要求を完了するのを援助し得、それによって、ユーザと自動アシスタントとの間の対話セッションがより迅速かつ効率的に終了し、コンピューティングデバイスにおいて受け取られるユーザ入力の量が削減され得る。さらに、ユーザが一定の時間窓以内に要求を完了しないために自動アシスタントが失敗する事例が軽減され得る。その結果、コンピューティングデバイスでの計算資源が保持され、要求を履行する際の待ち時間が短縮され得る。

20

30

【0015】

上記の説明は、本開示のいくつかの実装の概要として与えられるものである。そうした実装のより詳しい説明、および他の実装が、以下でより詳細に説明される。

【図面の簡単な説明】

【0016】

【図1A】様々な実装による、ユーザが自動アシスタントに例示的要求を反復的に与え、自動アシスタントが例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする図である。

【図1B】様々な実装による、ユーザが自動アシスタントに例示的要求を反復的に与え、自動アシスタントが例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする図である。

40

【図1C】様々な実装による、ユーザが自動アシスタントに例示的要求を反復的に与え、自動アシスタントが例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする図である。

【図2A】様々な実装による、ユーザが自動アシスタントに追加の例示的要求を反復的に与え、自動アシスタントが追加の例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする図である。

【図2B】様々な実装による、ユーザが自動アシスタントに追加の例示的要求を反復的に与え、自動アシスタントが追加の例示的要求の部分に対応するグラフィカル要素を反復的

50

にレンダリングする図である。

【図2C】様々な実装による、ユーザが自動アシスタントに追加の例示的要求を反復的に与え、自動アシスタントが追加の例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする図である。

【図3】様々な実装による、ユーザが自動アシスタントに向けられた要求を与えるとき、グラフィカル要素を反復的に構築する自動アシスタントを提供するシステムを示す図である。

【図4A】様々な実装による、ユーザが自動アシスタントに向けられた要求を与えるとき、ディスプレイインターフェースにおいてグラフィカル要素を反復的に与えるように自動アシスタントを制御するための方法を示す図である。

10

【図4B】様々な実装による、ユーザが自動アシスタントに向けられた要求を与えるとき、ディスプレイインターフェースにおいてグラフィカル要素を反復的に与えるように自動アシスタントを制御するための方法を示す図である。

【図5】様々な実装による例示的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】

【0017】

図1A、図1B、および図1Cは、それぞれ、ユーザ102が自動アシスタントに例示的要求を反復的に与え、自動アシスタントが例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする概観100、概観120、および概観140を示す。要求は、自動アシスタントに向けられた音声発話、または自動アシスタントに向けられた音声発話のシーケンス内に含まれ得る。たとえば、図1Aの例では、ユーザの自宅内のコンピューティングデバイス104を介してアクセス可能である自動アシスタントに向けられた要求108の第1の部分を含む「アシスタント、...を設定して」という音声発話をユーザ102が与えると仮定する。ユーザ102は、自動アシスタントに特定のアシスタントアクションを実施させ、かつ/または要求108の第1の部分に関連付けられる特定の意図を実行させることを促進するために、音声発話を与え得る。自動アシスタントは、ストリーミング自動音声認識(ASR)モデルを使用して、要求108の第1の部分を取り込むオーディオデータのストリームを処理させて、ASR出力を生成し得る。さらに、自動アシスタントは、自然言語理解(NLU)モデルを使用してASR出力を処理させ、NLU出力を生成し得る。特に、ユーザ102が要求108の第1の部分を含む音声発話を与えている間に、こうした動作が実施され得る。

20

30

【0018】

いくつかの実装では、自動アシスタントは、要求108の第1の部分进行处理することに基づいて、要求108の第1の部分に関連付けられる候補意図をNLU出力に基づいて決定し得る。さらに、自動アシスタントは、候補意図に基づいて、コンピューティングデバイス104のディスプレイインターフェース106においてレンダリングされるべき汎用コンテナグラフィカル要素110を決定し得る。たとえば、汎用コンテナグラフィカル要素110は、(たとえば、図1Aおよび図1Bに示されるように)要求の追加の部分进行处理することに基づいて、候補意図の中から選択された特定の意図に関連付けられる特定の調整されたコンテナグラフィカル要素に動的に適合され得るグラフィカルユーザインターフェース(GUI)要素を表現するように選択され得る。追加または代替の実装では、自動アシスタントは、ASR出力に基づいて(かつNLU出力を考慮することなく)、要求の第1の部分が特定のワードまたはフレーズ(たとえば、「設定して」)を含むことを示すと判定し得る。たとえば、ASR出力内で特定のワードまたはフレーズが検出されたとき、汎用コンテナグラフィカル要素110がコンピューティングデバイス104のディスプレイインターフェース106においてレンダリングされ得るように、特定のワードまたはフレーズがコンピューティングデバイス104のオンデバイスメモリ内の汎用コンテナグラフィカル要素110にマッピングされ得る。汎用コンテナグラフィカル要素110は、(たとえば、図1Bおよび図1Cに示されるように)要求の追加の部分进行处理することに基づいて特定の意図に関連付けられる特定の調整されたコンテナグラフィカル要素に同様に動的に適合され得る。言い換えれば、要求108の第1の部分が、グラフィカル要素の配列(たとえば、図1Aに示されるような空の円または塗り

40

50

つぶされた円の配列)によって表現され得る(たとえば、「設定して」、「変更して(chang e)」というワード、および/または他の特定のワードもしくはフレーズを含む要求108の第1の部分に基づく)「設定(setting)」を伴うアプリケーションおよび/またはデバイスを制御するための要求に対応すると予測されるので、汎用コンテナグラフィカル要素110は、コンピューティングデバイス104のディスプレイインターフェース106においてレンダリングされ得る。

【0019】

いくつかの実装では、ユーザ102が要求108の第1の部分を含む音声発話を与えることを休止したとの判定にตอบสนองして、汎用コンテナグラフィカル要素110は、コンピューティングデバイス104のディスプレイインターフェース106においてレンダリングされ得る。自動アシスタントは、たとえば、音声発話を処理する際に生成されたNLU出力、音声発話を処理することに基づいて決定されたオーディオベースの特性、および/またはユーザ102が要求108の第1の部分を与えてから経過したしきい持続時間に基づいて、ユーザ102が要求108の第1の部分を含む音声発話を与えることを休止したと判定し得る。たとえば、自動アシスタントは、ユーザ102が要求108の第1の部分を与えてから経過するしきい持続時間に基づいて、かつ候補意図に関連付けられる、対応するパラメータについてのスロット値が未知であることを示すNLU出力に基づいて、ユーザ102が音声発話を与えることを休止した(たとえば、ユーザ102が「設定して」と言ったが、何を「設定」すべきかについてどんな指示も与え損なった)と判定し得る。そうした実装のいくつかのバージョンでは、自動アシスタントはまた、予測される休止を取り囲む1つまたは複数の用語またはフレーズを考慮し得る(たとえば、前置詞、または流暢さを欠く音声(たとえば、えー(uhmmm, uhhh)などの後に休止が生じたかどうか)。代替または追加として、オーディオベースの特性は、ユーザ102が要求108の第1の部分を与えた方式が、要求を完了するために自動アシスタントに与えられる自然言語をどのように述べるかを考慮するためにユーザ102が休止したことを示すことを示し得る。

【0020】

自動アシスタントに対する完全な要求を与えることを促進するために、ユーザ102が音声発話を引き続き与え、または図1Bの概観120で示されるように「...照明...」を与えることによって要求122の第2の部分を含む追加の音声発話を与えるとさらに仮定する。この例では、要求122の第2の部分を取り込むオーディオデータのストリームが処理され、追加のASR出力および追加のNLU出力が生成され得る。追加のNLU出力に基づいて、自動アシスタントは、図1Aに関して説明された候補意図の中から、ユーザ102がユーザ102の自宅内のスマート電球の現在の状態を修正したいことを自動アシスタントに示す特定の意図を選択し得る。したがって、要求122の第2の部分を受け取ったことにตอบสนองして、自動アシスタントは、図1Aの汎用コンテナグラフィカル要素110を、要求122の第2の部分処理することに基づいて決定された特定の意図に関連付けられるコンテンツに動的に適合させ得、その結果、シームレスな形で特定の調整されたグラフィカルコンテナ要素112が得られる。

【0021】

特定の調整されたグラフィカルコンテナ要素112は、汎用グラフィカルコンテナ要素110がそれに対して動的に適合され得る複数の異種の調整されたグラフィカルコンテナ要素のうちの一つであり得、要求122の第2の部分処理することに基づいて選択される特定の意図に特有のものであり得る。言い換えれば、図1Bに示される特定の調整されたグラフィカルコンテナ要素112は、スマート電球の現在の状態を修正することに関連付けられる特定の意図に特有のものであり得、他の意図が、図1Bに示される特定の調整されたグラフィカルコンテナ要素112とは異なる、調整されたグラフィカルコンテナ要素(たとえば、温度意図を修正すること、スピーカ音量意図を修正することなど)に関連付けられ得る。特定の調整されたグラフィカルコンテナ要素112は、たとえば、自動アシスタントが要求122の第2の部分に関連付け得るデバイスおよび/またはアプリケーションを示すメディアコンテンツ124(たとえば、図1Bに示される電球アイコン)、自動アシスタントが要求122の第2

10

20

30

40

50

の部分に関連付け得、図1Aの汎用コンテナグラフィカル要素110に関連して示されるグラフィカル要素の配列から動的に適合されるデバイスおよび/またはアプリケーションの現在の状態を示す現在の状態120(たとえば、スマート電球が70%の輝度であることを示すための、図1Bに示される7つの塗りつぶされた円)、ユーザ102のタッチ入力を介して選択されたとき、(たとえば、図1Cに関連して説明されるように)デバイスおよび/またはアプリケーションの現在の状態120を制御させる1つまたは複数の制御要素144、制御されるべき様々なスマート電球に関連付けられるデバイス識別子、および/または他のコンテンツを含み得る。特に、ユーザ102が要求を与えるとき、汎用コンテナグラフィカル要素110が特定の調整されたコンテナグラフィカル要素112に動的に適合されるので、ユーザ102が要求を完了する前、かつ/または自動アシスタントが要求の履行を完了する前に、特定の調整されたコンテナグラフィカル要素112が、コンピューティングデバイス104のディスプレイインターフェース106においてレンダリングされ得る。

10

【0022】

図1Bに示される例とは対照的に、ユーザ102が音声発話を引き続き与え、または「...スピーカの音量...」などの、要求122の第2の部分を含む追加の音声発話を与えると仮定する。この例では、要求122の第2の部分が、前述と同じまたは同様の方式で処理され得る。しかしながら、得られる特定の調整されたコンテナグラフィカル要素は、図1Bに示される特定の調整されたコンテナグラフィカル要素112とは異なるものとなる。たとえば、この対照的な例での特定の調整されたグラフィカルコンテナ要素は、たとえば、自動アシスタントが要求122の第2の部分に関連付け得るデバイスおよび/またはアプリケーションを示すメディアコンテンツ(たとえば、スピーカアイコン)、自動アシスタントが要求122の第2の部分に関連付け得、図1Aの汎用コンテナグラフィカル要素110に関連して示されるグラフィカル要素の配列から動的に適合されるデバイスおよび/またはアプリケーションの現在の状態を示す現在の状態(たとえば、塗りつぶされた円はスピーカ音量レベルに対応し得る)、ユーザ102のタッチ入力を介して選択されたとき、デバイスおよび/またはアプリケーションの現在の状態を制御させる1つまたは複数の制御要素、および/または他のコンテンツを含み得る。したがって、図1Bに示される同一の汎用コンテナグラフィカル要素110が、要求122の第2の部分に基づいて、異なって適合され得る。

20

【0023】

ユーザ102が音声発話、追加の音声発話を与えることを完了し、または自動アシスタントに対する完全な要求を与えることを促進するために、図1Cの概観140において示されるように「...30パーセントに(... to 30 percent)」を与えることによって要求142の第3の部分を含む別の追加の音声発話を与えるとさらに仮定する。この例では、要求142の第3の部分を取り込むオーディオデータのストリームが処理され、別の追加のASR出力および別の追加のNLU出力が生成され得る。別の追加のNLU出力に基づいて、自動アシスタントは、ユーザ102の音声発話を履行する際に利用される構造化要求を生成し得る。たとえば、ユーザ102から要求142の第3の部分を受け取ったことに応答して、自動アシスタントは、スマート電球の現在の状態122を70%の輝度から30%の輝度に修正させるために、コンピューティングデバイス104においてアクセス可能であるスマート電球および/またはスマート電球に関連付けられるアプリケーションに1つまたは複数の構造化要求を送信し得る。さらに、要求142の第3の部分を受け取ったことに応答して、自動アシスタントは、スマート電球が70%の輝度から30%の輝度に変化した(たとえば、スマート電球が70%の輝度であることを示すための、図1Bに示される7つの塗りつぶされた円が、スマート電球が今や30%であることを示すための、図1Cの3つの塗りつぶされた円に変化する)ことを反映するように現在の状態122を適合させることによって、スマート電球の輝度の変化を反映し得る。

30

40

【0024】

代替として、図1Cに示されるように要求142の第3の部分を含む音声発話を与える代わりに、ユーザ102は、1つまたは複数の制御グラフィカル要素144のうちの所与の1つ(たとえば、第3の円)にタッチ入力を向け得る。同様に、ユーザ102が1つまたは複数の制御

50

グラフィカル要素のうちの所与の1つにタッチ入力に向けたことに応答して、自動アシスタントは、ユーザ102の音声発話を履行する際に利用される構造化要求を生成し、スマート電球の現在の状態122を70%の輝度から30%の輝度に修正させ(ユーザ102が第3の円にタッチ入力向けると仮定する)、前述のようにスマート電球が70%の輝度から30%輝度に変化したことを反映するように現在の状態122を適合させることによってスマート電球の輝度の変化を反映し得る。

【0025】

図1A、図1B、および図1Cの例が、様々な部分(たとえば、第1の部分108、第2の部分122、および第3の部分142)でのスマート電球の輝度を変更するための要求に関連して説明されるが、それは例示のためのものであり、限定を意味するわけではないことを理解されたい。たとえば、ユーザ102が要求を与える間であっても、要求の様々な部分をユーザ102が与える間に休止しない場合であっても、自動アシスタントは、自動アシスタントがユーザ102を聴取しており、要求を履行していることをユーザ102に示すために、前述のようにGUI要素を動的に適合させ得る。その結果、本明細書において説明される動的GUI要素に基づいて、自動アシスタントが要求を履行していることをユーザ102は知っているの
10
で、ユーザ102が要求の1つまたは複数の部分を反復する事例が軽減され得る。しかしながら、ユーザ102が要求を与える際に休止した場合、本明細書において説明される動的GUI要素は、要求が与えられるときにユーザ102に提示するためにスマート電球の現在の状態122を少なくとも与えることにより、ユーザ102が要求を完了するのを援助し得る。その
20
結果、ユーザ102と自動アシスタントとの間の対話セッションが、より迅速かつ効率的に終了し得る。さらに、図1A、図1B、および図1Cの例が、スマート電球の輝度を変更するための要求に関連して説明されるが、それは例示のためのものであり、限定を意味するわけではないことを理解されたい。むしろ、本明細書において説明される技法は、自動アシスタントに向けられた任意の要求のための動的GUI要素を与えるために利用され得ることを理解されたい。

【0026】

図2A、図2B、および図2Cは、それぞれ、ユーザ202が自動アシスタントに追加の例示的要求を反復的に与え、自動アシスタントが追加の例示的要求の部分に対応するグラフィカル要素を反復的にレンダリングする概観200、概観220、および概観240を示す。要求は、自動アシスタントに向けられた音声発話、または自動アシスタントに向けられた音声発話のシーケンス内に含まれ得る。たとえば、図2Aの例では、ユーザの自宅内のコンピューティングデバイス204を介してアクセス可能である自動アシスタントに向けられた要求208の第1の部分を含む「アシスタント、...を設定して」という音声発話をユーザ202
30
が与えると仮定する。ユーザ202は、自動アシスタントに特定のアシスタントアクションを実施させ、かつ/または要求208の第1の部分に関連付けられる特定の意図を実行させることを促進するために、音声発話を与え得る。図1Aに関連して上記で与えられた説明と同様に、自動アシスタントは、ストリーミング自動音声認識(ASR)モデルを使用して、要求208の第1の部分を取り込むオーディオデータのストリームを処理させて、ASR出力を生成し得、自然言語理解(NLU)モデルを使用して、ASR出力を処理させ、NLU出力を生成し得る。要求208の第1の部分を受け取ったことに応答して、自動アシスタントは、図1Aに
40
関連して上記で説明されたのと同じまたは同様の方式で、コンピューティングデバイス204のディスプレイインターフェース206に汎用コンテナグラフィカル要素210をレンダリングさせ得る。

【0027】

この場合も、自動アシスタントに対する完全な要求を与えることを促進するために、ユーザ102が音声発話を引き続き与え、または図2Bの概観220で示されるように「...照明...」を与えることによって要求226の第2の部分を含む追加の音声発話を与えるとさらに仮定する。同様に、要求226の第2の部分を取り込むオーディオデータのストリームが処理され、追加のASR出力および追加のNLU出力が生成され得る。追加のNLU出力に基づいて、自動アシスタントは、ユーザ202がユーザ202の自宅内のスマート電球の現在の状態を
50

修正したいことを自動アシスタントに示す特定の意図を選択し得る。したがって、要求226の第2の部分を受け取ったことに応答して。しかしながら、図1Aの例とは対照的に、ユーザ202がユーザ202の自宅全体にわたって共にグループ化されたスマート電球の複数のグループ(たとえば、スマート電球の「台所」グループ、スマート電球の「地下室」グループ、およびスマート電球の「廊下」グループ)を有すると仮定する。したがって、図2Aの例では、自動アシスタントは、図2Aの汎用コンテナグラフィカル要素210の複数の例を、要求226の第2の部分の処理することに基づいて、かつスマート電球の複数のグループに基づいて決定された特定の意図に関連付けられるコンテンツに動的に適合させ得、複数の特定の調整されたグラフィカルコンテナ要素222A、222B、および222Cが得られる。

【0028】

たとえば、第1の調整されたグラフィカルコンテナ要素222Aがスマート電球の「台所」グループに関連付けられ、「台所」というデバイス識別子224と、他のコンテンツ(たとえば、5つの塗りつぶされた円によって示される、50%の輝度という「台所」照明の現在の状態、スマート電球の「台所」グループに関連付けられる制御グラフィカル要素、メディアコンテンツ、および/または他のコンテンツ)とを含み得、第2の調整されたグラフィカルコンテナ要素222Bが、スマート電球の「地下室」グループに関連付けられ、「地下室」というデバイス識別子228と、他のコンテンツ(たとえば、9つの塗りつぶされた円によって示される、90%の輝度という「地下室」照明の現在の状態、スマート電球の「地下室」グループに関連付けられる制御グラフィカル要素、メディアコンテンツ、および/または他のコンテンツ)とを含み得、第3の調整されたグラフィカルコンテナ要素222Cが、スマート電球の「廊下」グループに関連付けられ、「廊下」というデバイス識別子230と、他のコンテンツ(たとえば、7つの塗りつぶされた円によって示される、70%の輝度という「廊下」照明の現在の状態、スマート電球の「廊下」グループに関連付けられる制御グラフィカル要素、メディアコンテンツ、および/または他のコンテンツ)とを含み得る。いくつかの実装では、自動アシスタントは、スマート電球に関連付けられるアプリケーションから、スマート電球のグループのそれぞれの現在の状態を検索し得るのに対して、別の実装では、自動アシスタントは、スマート電球から直接的に、スマート電球のグループのそれぞれの現在の状態を検索し得る。したがって、ユーザ202が要求226の第2の部分を与える時間の例において、自動アシスタントは、汎用コンテナグラフィカル要素210の複数の例を動的に適合させ得、その結果、複数の別個の調整されたコンテナグラフィカル要素が得られ、スマート電球のグループのそれぞれの現ステータスに基づいて、スマート電球をどのように制御するかについての指示が動的にユーザ202に提供される。

【0029】

いくつかの実装では、ユーザ202が要求を与えるとき、ユーザ202からの要求に対応する自然言語コンテンツ232が、ディスプレイインターフェース206においてレンダリングされ得る。たとえば、自然言語コンテンツ232は、要求208の第1の部分および要求226の第2の部分を取り込むオーディオデータのストリームを処理する際にストリーミングASRモデルを使用して生成されたASR出力に基づいて決定されるストリーミングトランスクリプション(たとえば、図2Bのコンピューティングデバイス204のディスプレイインターフェース206において示されるように、「アシスタント、照明を...設定して(Assistant, set the lights ...)」)を含み得る。特に、自然言語コンテンツは、複数の調整されたコンテナグラフィカル要素222A、222B、および222Cがコンピューティングデバイス204のディスプレイインターフェース206においてレンダリングされると同時にレンダリングされ得る。こうした方式および他の方式では、自動アシスタントは、ユーザ202が要求を迅速かつ効率的に完了するのを援助しようとして、様々なグラフィカル要素(たとえば、調整されたコンテナグラフィカル要素222A、222B、および222C、自然言語コンテンツ232、および/または他のグラフィカル要素)を引き起こし得、それによって、ユーザ202と自動アシスタントとの間の対話セッションの持続時間が短縮される。

【0030】

自動アシスタントに対する完全な要求を与えることを促進するために、図2Cの概観240

10

20

30

40

50

において示されるように、「地下室内の...30パーセントに(... in the basement to 30 percent)」という要求242の第3の部分を与えることによってユーザ202が要求を完了するとさらに仮定する。この例では、要求242の第3の部分を取り込むオーディオデータのストリームが処理され、ASR出力およびNLU出力が生成され得る。NLU出力に基づいて、自動アシスタントは、ユーザ202の音声発話を履行する際に利用される構造化要求を生成し得る。たとえば、自動アシスタントは、ユーザ202から要求242の第3の部分を受け取ったことに応答して、「地下室」グループ内のスマート電球および/またはコンピューティングデバイス204においてアクセス可能であるスマート電球に関連付けられるアプリケーションに1つまたは複数の構造化要求を送信し、「地下室」グループ内のスマート電球の現在の状態を90%の輝度から30%の輝度に修正させ得る。さらに、要求242の第3の部分を受け取ったことに応答して、自動アシスタントは、スマート電球が90%の輝度から30%の輝度に変化した(たとえば、「地下室」グループ内のスマート電球が90%の輝度であることを示すための、図2Bに示される9つの塗りつぶされた円が、「地下室」グループ内のスマート電球が今や30%であることを示すための、図2Cの3つの塗りつぶされた円に変化すること)を反映するように現在の状態を適合させることによって、スマート電球の輝度の変化を反映し得る。

10

【0031】

いくつかの実装では、図2Cに示されるように、ユーザがスマート電球の特定のグループ(たとえば、図2Cの例の「地下室」グループ)を選択したことをASR出力が示すとき、自動アシスタントは、ユーザ202の特定の意図に関連付けられる、調整されたコンテナグラフィカル要素222Bが唯一の残りの調整されたコンテナグラフィカル要素となるように、コンピューティングデバイス204のディスプレイインターフェース206から、他の調整されたコンテナグラフィカル要素222Aおよび222Cを除去し得る。いくつかの実装では、自動アシスタントを介して実施されるASRおよび/またはNLUプロセスが、コンテナグラフィカル要素のコンテンツに従ってバイアスされ得る。たとえば、要求242の第3の部分を受け取ったことに応答して、自動アシスタントは、「台所」、「地下室」、および「廊下」に向けてASR処理および/またはASR出力をバイアスし得る。

20

【0032】

図3は、ユーザが自動アシスタント304に向けられた要求を与えるとき、グラフィカル要素を反復的に構築する自動アシスタント304を提供するシステム300を示す。自動アシスタント304は、コンピューティングデバイス302(たとえば、コンピューティングデバイス104の例、コンピューティングデバイス204の例など)および/またはサーバデバイスなどの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アシスタントアプリケーションの部分として動作し得る。ユーザは、アシスタントインターフェース320を介して自動アシスタント304と対話し得、アシスタントインターフェース320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のできる任意の他の装置であり得る。たとえば、ユーザは、アシスタントインターフェース320に口頭、テキスト、および/またはグラフィカルの入力を与えて、自動アシスタント304に1つまたは複数のアクションを初期化させる(たとえば、データを提供し、周辺デバイスを制御し、エージェントにアクセスし、入力および/または出力を生成するなど)ことにより、自動アシスタント304を初期化し得る。代替として、自動アシスタント304は、1つまたは複数のトレーニング済み機械学習モデルを使用するコンテキストデータ336の処理に基づいて初期化され得る。コンテキストデータ336は、自動アシスタント304がアクセス可能である環境の1つまたは複数の特徴、ならびに/あるいは自動アシスタント304と対話することを意図していると予測されるユーザの1つまたは複数の特徴を特徴付け得る。コンピューティングデバイス302はディスプレイデバイスを含み得、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コンピューティングデバイス302のアプリケーション334を制御することを可能にするためのタッチ入力および/またはジェスチャを受け取るためのタッチインターフェース(たとえば、コンピューティングデバイス104のディスプレ

30

40

50

インターフェース106、コンピューティングデバイス204のディスプレイインターフェース206)を含むディスプレイパネルであり得る。いくつかの実装では、コンピューティングデバイス302にはディスプレイデバイスが欠如していることがあり、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。さらに、コンピューティングデバイス302は、ユーザから音声自然言語入力を受け取るための、マイクロフォンなどのユーザインターフェースを提供し得る。いくつかの実装では、コンピューティングデバイス302はタッチインターフェースを含み得、カメラが欠けていることがあるが、任意選択で1つまたは複数の他のセンサを含み得る。

【0033】

コンピューティングデバイス302および/または他のサードパーティクライアントデバイスが、インターネットなどのネットワークを介してサーバデバイスと通信中であり得る。さらに、コンピューティングデバイス302と任意の他のコンピューティングデバイスが、Wi-Fiネットワークなどのローカルエリアネットワーク(LAN)を介して互いに通信中であり得る。コンピューティングデバイス302は、コンピューティングデバイス302での計算資源を保持するために、サーバデバイスに計算タスクをオフロードし得る。たとえば、サーバデバイスは自動アシスタント304をホストし得、かつ/またはコンピューティングデバイス302は、1つまたは複数のアシスタントインターフェース320において受け取った入力をサーバデバイスに送信し得る。しかしながら、いくつかの実装では、自動アシスタント304は、コンピューティングデバイス302においてローカルにホストされ得、自動アシスタント動作に関連する様々なプロセスが、コンピューティングデバイス302において排他的に実施され得る。

【0034】

様々な実装では、自動アシスタント304のすべての態様またはすべてよりも少ない態様が、コンピューティングデバイス302上で実装され得る。そうした実装のいくつかでは、自動アシスタント304の態様がコンピューティングデバイス302を介して実装され、サーバデバイスとインターフェースし得、サーバデバイスは、自動アシスタント304の別の態様を実装し得る。任意選択で、サーバデバイスは、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するアシスタントアプリケーションにサービスし得る。自動アシスタント304のすべての態様またはすべてよりも少ない態様がコンピューティングデバイス302を介して実装される実装では、自動アシスタント304は、コンピューティングデバイス302のオペレーティングシステムとは別々の(たとえば、オペレーティングシステム「上」にインストールされた)アプリケーションであり得、またあるいはコンピューティングデバイス302のオペレーティングシステム(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と見なされるが、オペレーティングシステムと一体化される)によって直接的に実装され得る。

【0035】

いくつかの実装では、自動アシスタント304は入力処理エンジン306を含み得、入力処理エンジン306は、コンピューティングデバイス302および/またはサーバデバイスに対する入力および/または出力を処理するための複数の異なるモジュールを利用し得る。たとえば、入力処理エンジン306は、ストリーミングASRモデルを利用する音声処理エンジン308を含み得、ストリーミングASRモデルは、アシスタントインターフェース320において受け取られたオーディオデータのストリームを処理して、オーディオデータのストリームで具体化されるテキストなどのASR出力を生成し得る。さらに、たとえば、入力処理エンジン306は、オーディオベースの機械学習モデルおよび/またはヒューリスティックベースの手法を使用して、オーディオデータのストリーム内に取り込まれる任意の音声発話/要求に関連付けられるオーディオベースの特性を決定し得る。いくつかの実装では、オーディオデータのストリームが、コンピューティングデバイス302での計算資源を保持するために、たとえばコンピューティングデバイス302からサーバデバイスに送信され得る。追加または代替として、オーディオデータのストリームは、コンピューティングデバイス302

10

20

30

40

50

において排他的に処理され得る。

【0036】

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、ニューラルネットワークを利用し得る音声認識アルゴリズム、および/またはワードまた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(たとえば、本明細書において説明されるストリーミングASRモデル)を含み得る。オーディオデータから変換されたテキストは、NLUモデルを利用するデータパーシングエンジン310によって構文解析され、ユーザによって指定されるコマンドフレーズ、意図、アクション、スロット値、および/または任意の他のコンテンツなどのNLU出力を生成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304に対して利用可能にされ得る。いくつかの実装では、データパーシングエンジン310によって提供される出力データは、パラメータエンジン312に提供され、自動アシスタント304、および/または自動アシスタント304を介してアクセスすることのできるアプリケーションもしくはエージェントによって実施することのできる特定の意図、アクション、および/またはルーチンに対応する入力をユーザが与えたどうか判定され得る。たとえば、アシスタントデータ338は、サーバデバイスおよび/またはコンピューティングデバイス302に記憶され得、自動アシスタント304によって実施することのできる1つまたは複数のアクションを定義するデータ、ならびにアクションを実施するのに必要なパラメータを含み得る。パラメータエンジン312は、意図、アクション、および/またはスロット値についての1つまたは複数のパラメータを生成し、出力生成エンジン314に1つまたは複数のパラメータを提供し得る。出力生成エンジン314は、1つまたは複数のパラメータを使用して、ユーザに出力(たとえば、視覚的出力および/または可聴出力)を提供するためのアシスタントインターフェース320と通信し、かつ/または1つまたは複数のアプリケーション334に出力を提供するための1つまたは複数のアプリケーション334と通信し得る。

10

20

【0037】

いくつかの実装では、前述のように、自動アシスタント304は、コンピューティングデバイス302のオペレーティングシステム「上」にインストールされ得るアプリケーションであり得、かつ/またはそれ自体でコンピューティングデバイス302のオペレーティングシステムの部分(または全体)を形成し得る。自動アシスタントアプリケーションは、オンデバイスASR、オンデバイスNLU、およびオンデバイス履行を含み、かつ/またはそれらに対してアクセスできる。たとえば、オンデバイスASRは、コンピューティングデバイス302にローカルに記憶されたエンドツーエンドストリーミングASRモデルを使用して(マイクروفोनによって検出された)オーディオデータのストリームを処理するストリーミングASRモデルを使用して実施され得る。オンデバイス音声認識は、オーディオデータのストリーム内に存在する(それが存在する場合)音声発話についての認識済みテキストなどのASR出力を生成する。さらに、たとえば、ストリーミングASRモデルを使用して生成されたASR出力と、任意選択でコンテキストデータとを処理するオンデバイスNLUモデルを使用して、オンデバイスNLUが実施され、NLU出力が生成され得る。NLU出力は、音声発話に対応する候補意図と、任意選択で、候補意図に関連付けられる、対応するパラメータについてのスロット値とを含み得る。

30

40

【0038】

NLU出力を利用するオンデバイス履行モデルまたは履行規則と、任意選択で他のローカルデータとを使用して、オンデバイス履行が実施され、音声発話の意図(および任意選択で、候補意図に関連付けられる、対応するパラメータについてのスロット値)を解決するために行うべきアクションを決定するための構造化要求が決定され得る。これは、音声発話および/または要求に対するローカルおよび/またはリモート応答(たとえば、回答)、音声発話および/または要求に基づいて実施するためのローカルにインストールされたアプリケーションとの対話、音声発話および/または要求に基づいて(直接的に、または対応するリモートシステムを介して)モノのインターネット(IoT)デバイスに送信するためのコマンド、ならびに/あるいは音声発話および/または要求に基づいて実施するための他の解決アクシ

50

ョンを決定することを含み得る。次いで、オンデバイス履行は、決定されたアクションのローカルおよび/またはリモート実施/実行を開始して、音声発話および/または要求を解決する。

【0039】

様々な実装では、リモート音声処理、リモートNLU、および/またはリモート履行が、コンピューティングデバイス302での計算資源を保持するために少なくとも選択的に利用され得る。たとえば、認識済みテキストが、リモート自動アシスタント構成要素リモートNLUおよび/またはリモート履行に少なくとも選択的に送信され得る。たとえば、任意選択で、認識済みテキストが、オンデバイス実施と並行して、あるいはオンデバイスNLUおよび/またはオンデバイス履行の失敗に回答して、リモート実施のために送信され得る。しかしながら、オンデバイス音声処理、オンデバイスNLU、オンデバイス履行、および/またはオンデバイス実行が、少なくとも音声発話を解決するときに実現する待ち時間短縮のために(音声発話を解決するためにクライアント-サーバラウンドトリップが不要であるために)優先され得る。さらに、オンデバイス機能は、ネットワーク接続性がなく、またはネットワーク接続性が限られた状況で利用可能な唯一の機能であり得る。

10

【0040】

いくつかの実装では、コンピューティングデバイス302は、コンピューティングデバイス302および/または自動アシスタント304を提供したエンティティとは異なるサードパーティエンティティ、あるいはコンピューティングデバイス302および/または自動アシスタント304を提供したのと同じのエンティティであるファーストパーティエンティティによって提供され得る1つまたは複数のアプリケーション334にアクセスできる。自動アシスタント304および/またはコンピューティングデバイス302は、アプリケーションデータ330にアクセスして、1つまたは複数のアプリケーション334によって実施することのできる1つまたは複数のアクション、ならびに1つまたは複数のアプリケーション334の各アプリケーションの状態および/またはコンピューティングデバイス302に関連付けられるそれぞれのデバイスの状態を決定し得る。さらに、自動アシスタント304および/またはコンピューティングデバイス302は、デバイスデータ332にアクセスして、コンピューティングデバイス302ならびに/あるいはコンピューティングデバイス302に関連付けられる1つまたは複数のデバイスによって実施することのできる1つまたは複数のアクションを決定し得る。さらに、アプリケーションデータ330および/または任意の他のデータ(たとえば、デバイスデータ332)が自動アシスタント304によってアクセスされ、コンテキストデータ336が生成され得、コンテキストデータ336は、特定のアプリケーション334および/またはデバイスが実行中のコンテキスト、ならびに/あるいは特定のユーザがコンピューティングデバイス302、アプリケーション334、および/または任意の他のデバイスもしくはモジュールにアクセスしているコンテキストを特徴付け得る。

20

30

【0041】

1つまたは複数のアプリケーション334がコンピューティングデバイス302において実行中である間、デバイスデータ332は、コンピューティングデバイス302において実行中の各アプリケーション334の現在の動作状態を特徴付け得る。さらに、アプリケーションデータ330は、1つまたは複数のアプリケーション334の指示でレンダリングされている1つまたは複数のグラフィカルユーザインターフェースのコンテンツなどの、実行中のアプリケーション334の1つまたは複数の特徴を特徴付け得る。代替または追加として、アプリケーションデータ330はアクションスキーマを特徴付け得、アクションスキーマは、それぞれのアプリケーションの現在の動作ステータスに基づいて、それぞれのアプリケーションおよび/または自動アシスタント304によって更新され得る。代替または追加として、1つまたは複数のアプリケーション334についての1つまたは複数のアクションスキーマは、静的なままであり得るが、自動アシスタント304を介して初期化するための適切なアクションを決定するために、アプリケーション状態エンジンによってアクセスされ得る。

40

【0042】

コンピューティングデバイス302は、1つまたは複数のトレーニング済み機械学習モデ

50

ルを使用して、アプリケーションデータ330、デバイスデータ332、コンテキストデータ336、および/またはコンピューティングデバイス302にとってアクセス可能である任意の他のデータを処理し得るアシスタント呼出しエンジン322をさらに含む得る。アシスタント呼出しエンジン322は、ユーザが呼出しフレーズを明示的に話して自動アシスタント304を呼び出すのを待機するか否かを判定し、またはユーザが呼出しフレーズを明示的に話すことを必要とする代わりに、データが自動アシスタントを呼び出すためのユーザによる意図を示すと見なすために、このデータを処理し得る。たとえば、1つまたは複数のトレーニング済み機械学習モデルは、複数のデバイスおよび/またはアプリケーションが様々な動作状態を示す環境内にユーザがいるシナリオに基づくトレーニングデータの例を使用してトレーニングされ得る。

10

【0043】

トレーニングデータの例は、ユーザが自動アシスタントを呼び出すコンテキスト、およびユーザが自動アシスタントを呼び出さない他のコンテキストを特徴付けるトレーニングデータを取り込むために生成され得る。1つまたは複数のトレーニング済み機械学習モデルがトレーニングデータのこうした例に従ってトレーニングされるとき、アシスタント呼出しエンジン322は、コンテキストおよび/または環境の特徴に基づいて、ユーザからの音声呼出しフレーズを自動アシスタント304に検出させ、または検出を制限させ得る。追加または代替として、アシスタント呼出しエンジン322は、コンテキストおよび/または環境の特徴に基づいて、ユーザからの1つまたは複数のアシスタントコマンドについて自動アシスタント304に検出させ、または検出を制限させ得る。いくつかの実装では、アシスタント呼出しエンジン322は、コンピューティングデバイス302が別のコンピューティングデバイスからアシスタント抑制出力を検出したことに基づいて、使用不能にされ、または制限され得る。このようにして、コンピューティングデバイス302がアシスタント抑制出力を検出しているとき、自動アシスタント304は、コンテキストデータ336に基づいて呼び出されず、さもなければ、アシスタント抑制出力が検出されていない場合に自動アシスタント304が呼び出されることになる。

20

【0044】

いくつかの実装では、システム300は、ユーザによって自動アシスタント304に与えられる要求の1つまたは複数の部分に関連付けられ得る1つまたは複数の候補意図を決定するための候補意図エンジン316を含む得る。たとえば、ユーザが「アシスタント、...を再生して(Assistant, play ...)」などの音声発話を与えたとき、候補意図エンジン316は、前述のNLU出力に基づいて音声発話に関連付けられ得る1つまたは複数の意図を識別し得る。代替または追加として、候補意図エンジン316は、ユーザによって与えられる音声発話に関連しないことがあるいくつかの候補意図を除去し得る。

30

【0045】

いくつかの実装では、システム300は、候補意図エンジン316によって識別された1つまたは複数の候補意図に基づいて1つまたは複数の汎用コンテナグラフィカル要素を取得および/または生成し得る汎用コンテナエンジン318を含む得る。汎用コンテナグラフィカル要素は、自動アシスタント304が履行することを求める要求を含む音声発話を受け取ったことに応答して、(たとえば、コンピューティングデバイス302のオンデバイスメモリから)取得または生成され得る。汎用コンテナグラフィカル要素に、音声発話を介してユーザが与えていると予測される1つまたは複数の要求に関連付けられる他の要素および/または特徴が割り当てられ得る。たとえば、「アシスタント、...を再生して」などの最初の音声発話が、特定のスロット値(たとえば、歌、アーティスト、ムービー、ストーリーミングサービスなど)を特定しないが、汎用コンテナエンジン318は、要求に関連付けられると予測されるスロット値のタイプを識別し得る。このタイプのスロット値に基づいて、汎用コンテナエンジン318は、コンピューティングデバイス302においてこれから受け取らなければならない要求の部分に基づいて動的に適合され得る汎用コンテナグラフィカル要素を取得および/または生成し得る。たとえば、「再生して」という用語を含む最初の音声発話が、メディア再生を制御するためのスロット値のタイプに関連付けられ得る。したがって、「

40

50

メディア再生」特徴に関連付けられる汎用コンテナグラフィカル要素が、汎用コンテナエンジン318によって選択され、その後で、メディア再生に関連付けられる特徴に動的に適合され得る。代替として、「ターン(turn)」という用語を含む最初の音声発話が、アプリケーションおよび/またはデバイスの出力レベルを制御するためのスロット値のタイプに関連付けられ得る。したがって、アプリケーションおよび/またはデバイスの設定を調節することに関連付けられる汎用コンテナグラフィカル要素が、汎用コンテナエンジン318によって選択され、その後で、アプリケーションおよび/またはデバイスの出力レベルを制御するための設定に関連付けられる特徴に動的に適合され得る。

【0046】

いくつかの実装では、システム300は、調整されたコンテナエンジン326を含み得、調整されたコンテナエンジン326は、汎用コンテナグラフィカル要素を動的に適合させることを割り当てるための制御要素を取得および/または生成し得、調整されたコンテナグラフィカル要素が得られる。たとえば、ユーザが自動アシスタント304に要求の追加の部分を与えるとき、調整されたコンテナエンジン326は、汎用コンテナグラフィカル要素における制御要素を反復的に割り当て、かつ/または除去し得る。各制御要素は、ユーザが自動アシスタント304に与えていると予測される要求に対応すると判定されるスロット値および/またはスロット値のタイプに関連付けられ得る。たとえば、自動アシスタント304が数値の範囲を有し得るデバイス設定を修正することをユーザが要求していると予測されるとき、調整されたコンテナエンジン326は、「スライディング」GUI要素(またはデバイス設定を制御するのに適した任意の他の要素)をコンテナグラフィカル要素に割り当てるように選択し得る。代替として、ユーザがメディアコンテンツ(たとえば、オーディオおよび/またはビデオ)の再生を制御することを求める要求を自動アシスタント304に与えていると予測されるとき、調整されたコンテナエンジン326は、1つまたは複数のメディア再生制御要素(たとえば、休止ボタン、再生ボタン、スキップボタンなど)をコンテナグラフィカル要素に割り当てるように選択し得る。

【0047】

いくつかの実装では、ユーザが自動アシスタント304に要求の追加の部分を与えるとき、システム300は、調整されたコンテナグラフィカル要素についてのアプリケーションおよび/またはデバイスの現在の状態を取得および/または除去し得る状態エンジン324を含み得る。さらに、状態エンジンは、要求に関連付けられる音声発話および/またはタッチ入力を処理することに基づいて、アプリケーションおよび/またはデバイスについての更新後状態を決定し得る。たとえば、自動アシスタント304が別のコンピューティングデバイスの設定を修正することをユーザが要求していると予測されるとき、状態エンジン326は、別のコンピューティングデバイスについての設定の現在の状態を決定し得る。この現在の状態に基づいて、状態エンジン326は、デバイスのステータスを特徴付け得る状態GUI要素を識別し、調整されたコンテナグラフィカル要素内に状態GUI要素を組み込ませ得る。たとえば、設定の状態が数値の範囲以内の値に対応するとき、状態エンジン326は、数値の範囲を特徴付け、設定の現在の状態を強調する状態GUI要素を生成し得る。この状態GUI要素が、コンピューティングデバイス302のインターフェースにおいて現在レンダリングされている調整されたコンテナグラフィカル要素内に組み込まれ得る。このようにして、ユーザが要求の追加の部分を引き続き与えるとき、ユーザは、設定についての現在の状態の通知を受けることができ、それによって何らかの現在の状態の記憶を新たにし、ユーザが現在の状態に対して所望の更新後状態を決定するのを援助する。

【0048】

図4Aおよび図4Bは、ユーザが自動アシスタントに向けられた要求を与えるとき、ディスプレイインターフェースにおいてグラフィカル要素を反復的に提供するように自動アシスタントを制御するための方法400および方法420を示す。方法400および方法420は、1つまたは複数のアプリケーション、コンピューティングデバイス、および/または自動アシスタントと対話することのできる任意の他のアプリケーションもしくはモジュールによって実施され得る。方法400は、ユーザが自動アシスタントに要求を与えたかどうかを判

10

20

30

40

50

定するための動作402を含み得る。要求は、たとえば、「アシスタント、...を調節して(Assistant, adjust ...)」などの音声発話内に含まれ得、音声発話は、自動アシスタントによって履行される要求の第1の部分参照し得る。

【0049】

要求の少なくとも一部が受け取られたと自動アシスタントが判定したとき、方法400は動作402から動作404に進み得る。動作404は、要求が完全な要求に対応するか、それとも不完全な要求に対応するかを判定することを含み得る。言い換えれば、自動アシスタントは、自動アシスタントが要求の履行を初期化するのに十分な情報をユーザが与えたかどうかを判定し得る。前述の例によれば、ユーザが音声発話「アシスタント、...を調節して」を与えたとき、自動アシスタントは、要求が不完全な要求に対応すると判定し得る。この判定に基づいて、方法400は、動作404から動作406に進み得る。そうではなく、要求が完全な要求に対応すると自動アシスタントが判定した場合、方法400は、図4Aおよび図4Bに示され、図4Bに関連して説明されるように、動作404から、継続要素「B」を介して、動作424に進み得る。様々な実装では、動作404において要求が完全な要求であると判定されたときであっても、方法400は動作406に進み得る。したがって、方法400および方法420は例示のためのものであり、限定を意味するわけではないことを理解されたい。

10

【0050】

動作406は、コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせることを含み得る。汎用コンテナグラフィカル要素は、汎用コンテナグラフィカル要素が動的に適合され得る他のグラフィカル要素についてのプレースホルダとして動作し得る。たとえば、汎用コンテナグラフィカル要素は、他のグラフィカル要素を割り当てるのに十分なエリアを含むボディを有する形状のグラフィカルレンダリングであり得る。他のグラフィカル要素は、限定はしないが、1つまたは複数のアプリケーションおよび/またはデバイスを制御するための制御要素、1つまたは複数のアプリケーションおよび/またはデバイスの現在の状態を示すための状態要素、1つまたは複数のアプリケーションおよび/またはデバイスのデバイス識別子、メディアコンテンツに基づくメディア要素、ならびに/あるいはディスプレイインターフェースにおいてレンダリングされ得る任意の他のタイプの要素を含み得る。方法400は、動作406から動作408に進み得、動作408は、不完全な要求の追加の部分が自動アシスタントによって受け取られたかどうかを判定することを含み得る。

20

30

【0051】

要求の追加の部分は、「...温度...」などの音声発話であり得る。不完全な要求の追加の部分が受け取られたと判定されたとき、方法400は、動作408から動作410に進み得る。そうではなく、不完全な要求の別の部分が受け取られていないと判定されたとき、方法400は、図4Aおよび図4Bに示され、図4Bに関連して説明されるように、動作408から、継続要素「A」を介して、方法420の任意選択の動作424に進み得る。図4Bに示される、任意選択の動作424は、ディスプレイインターフェースにおいて、1つまたは複数の選択可能な提案をレンダリングすることを含み得る。1つまたは複数の選択可能な提案は、自動アシスタントによってユーザから受け取られた不完全な要求の1つまたは複数の部分に基づき得る。このようにして、ユーザは完全であった要求を与えなかったが、それでも自動アシスタントは、ユーザが伝達しようとして試みている可能性のある1つまたは複数の意図に対応すると予測される1つまたは複数の選択可能な提案を提供し得る。その後で、任意選択で、方法420は、図4Aおよび図4Bに示されるように、継続要素「C」を介して動作402に戻り得る。

40

【0052】

不完全な要求の追加の部分が動作408において受け取られたと判定されたとき、方法400は、動作408から動作410に進み得る。動作410は、不完全な要求が特定の意図に対応すると判定することを含み得る。いくつかの例では、不完全な要求を完了するための1つまたは複数の追加の入力をユーザが与えた後に動作410が実施され得る。たとえば、ユーザが要求の第1の部分「アシスタント、...を調節して」と、その後続く(たとえば、同一

50

の音声発話、またはその音声発話の後に続く追加の音声発話内に含まれる)要求の第2の部分「...温度...」を与えたとき、自動アシスタントは、アプリケーションおよび/またはデバイスの温度設定を修正することをユーザが要求していると判定し得る。いくつかの実装では、特定の意図は、特定の意図に関連付けられる、対応するパラメータについてのスロット値を有し得る。いくつかの実装では、ユーザからの要求が、自動アシスタントまたは他のアプリケーションによって決定されるように、予測確率に基づいて不完全と見なされ得る。たとえば、予測確率は、特定の意図が実行されることをユーザが要求している可能性を示し得る。予測確率が確率しきい値を満たすとき、ユーザからの要求が完全であると見なされ得る。その後で、対応するパラメータについてのスロット値が、ユーザからの追加の入力および/または自動アシスタントにとって利用可能なデータに基づいて、特定の意図に割り当てられ得る。

10

【0053】

方法400は、動作410から動作412に進み得、動作412は、汎用コンテナグラフィカル要素を動的に適合させることを含み得、その結果、特定の意図に特有の、調整されたコンテナグラフィカル要素が得られ得る。調整されたコンテナグラフィカル要素は、特定のアシスタントアクションを制御するための1つまたは複数の制御要素、アプリケーションおよび/またはデバイスのステータスを示す1つまたは複数の状態要素、ならびに/あるいは本明細書において説明される他のコンテンツを含み得る。たとえば、要求の部分に基づいて(たとえば、「アシスタント、...を調節して」、)自動アシスタントは、特定のデバイス(たとえば、廊下サーモスタット)の現在の温度設定を示すための汎用コンテナグラフィカル要素に様々なグラフィカル要素を動的に適合させ得る。代替または追加として、自動アシスタントは、別のグラフィカル要素を、特定のデバイスの温度設定を調節するための汎用コンテナグラフィカル要素に割り当てさせ得る。このようにして、ユーザは、特定のデバイスの現在の状態、さらには特定のデバイスを制御するためのオプションを確認することができることになる。グラフィカル要素をコンテナグラフィカル要素に反復的に割り当てることにより、自動アシスタントおよびコンピューティングデバイスは、普通ならユーザが完成した要求を与えるのを待機して消費することになる時間およびリソースを保持し得る。

20

【0054】

方法400は、動作412から動作414に進み得、動作414は、特定のアシスタントアクションを初期化するための入力を受け取られたかどうかを判定することを含み得る。入力は、たとえば、同一の音声発話、または音声発話の後に続く追加の音声発話内に含まれる要求の別の部分(たとえば、「...72度に...」、)ならびに/あるいは特定の調整されたコンテナグラフィカル要素をレンダリングしているディスプレイインターフェースのエリアでのタッチ入力であり得る。要求の履行を初期化するための入力を受け取られたと判定される時、方法400は、図4Aおよび図4Bに示され、図4Bに関連して説明されるように、動作414から継続要素「B」を介して方法420の動作422に進む。そうでない場合、方法400は動作414から継続要素「A」を介して任意選択動作424および/または動作402に進み得る。

30

【0055】

動作422は、ユーザからの要求に基づいて履行を初期化することを含み得る。履行は、特定の意図を実行して要求を履行することに対応し得る。たとえば、ユーザは、デバイスの現在の状態が67度であることを示す特定の調整されたコンテナグラフィカル要素に基づいて、「...72度に...」などの追加の音声発話を与え得る。したがって、ユーザの記憶が、特定の調整されたコンテナグラフィカル要素において搬送された情報によって新たにされ得る。方法420は、動作422から任意選択の動作426に進み得、動作426は、履行に基づいて応答出力をレンダリングすることを含み得る。たとえば、特定の調整されたコンテナグラフィカル要素が、履行の実施に基づいて、追加のコンテンツに割り当てられ得る。前述の例によれば、コンテナグラフィカル要素に、デバイスの温度設定が、現在の状態67度から更新後状態72度に首尾よく調節または修正されたことを示すための追加のグラフィカ

40

50

ルコンテンツに割り当てられ得る。その後で、方法420は、継続要素「C」を介して動作402に戻り得る。

【0056】

図5は、例示的コンピュータシステム510のブロック図500である。コンピュータシステム510は通常、バスサブシステム512を介していくつかの周辺デバイスと通信する少なくとも1つのプロセッサ514を含む。こうした周辺デバイスは、たとえばメモリ525およびファイル記憶サブシステム526を含む記憶サブシステム524と、ユーザインターフェース出力デバイス520と、ユーザインターフェース入力デバイス522と、ネットワークインターフェースサブシステム516とを含み得る。入力および出力デバイスは、コンピュータシステム510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516は、外部ネットワークに対する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。

10

【0057】

ユーザインターフェース入力デバイス522は、キーボード、マウス、トラックボール、タッチパッド、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、ディスプレイ内に組み込まれたタッチスクリーン、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、および/または他のタイプの入力デバイスを含み得る。一般には、「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510内に、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式を含むものとする。

20

【0058】

ユーザインターフェース出力デバイス520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、または可視画像を生み出す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、非視覚ディスプレイをも提供し得る。一般には、「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510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式を含むものとする。

【0059】

記憶サブシステム524は、本明細書において説明されるモジュールの一部またはすべての機能を提供するプログラミングおよびデータ構成を記憶する。たとえば、記憶サブシステム524は、方法400の選択された態様を実施し、かつ/またはシステム300、コンピューティングデバイス104、コンピューティングデバイス204、自動アシスタント、ならびに/あるいは本明細書において論じられる任意の他のアプリケーション、デバイス、装置、および/またはモジュールのうちの1つまたは複数を実装するための論理を含み得る。

30

【0060】

こうしたソフトウェアモジュールは一般に、プロセッサ514だけで、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524内で使用されるメモリ525は、プログラム実行中の命令およびデータの記憶用のメインランダムアクセスメモリ(RAM)530と、固定命令が記憶される読取り専用メモリ(ROM)532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526は、プログラムおよびデータファイル用の永続的記憶を提供し得、ハードディスクドライブ、フロッピディスクドライブおよび関連する取外し可能媒体、CD-ROMドライブ、光学ドライブ、または取外し可能媒体カートリッジを含み得る。いくつかの実装の機能を実装するモジュールが、ファイル記憶サブシステム526によって記憶サブシステム524内に、またはプロセッサ514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

【0061】

バスサブシステム512は、コンピュータシステム510の様々な構成要素およびサブシステムが互いに所期の通りに通信することを許可する機構を提供する。バスサブシステム51

50

2が概略的に単一のバスとして示されているが、バスサブシステムの代替実装は複数のバスを使用し得る。

【0062】

コンピュータシステム510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様々なタイプであり得る。コンピュータおよびネットワークの常に変化する性質のために、図5に示されるコンピュータシステム510の説明は、いくつかの実装を示すための特定の例として意図されるに過ぎない。図5に示されるコンピュータシステムよりも多い、または少ない構成要素を有する、コンピュータシステム510の多くの他の構成が可能である。

10

【0063】

本明細書において説明されるシステムがユーザ(または、本明細書においてしばしば呼ばれるように、「参加者」)についての個人情報を収集し、または個人情報を利用し得る状況において、プログラムまたは機能がユーザ情報(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活動、職業、ユーザのプリファレンス、またはユーザの現在の地理的位置についての情報)を収集するかどうかを制御する機会、あるいはユーザにより関連し得るコンテンツをコンテンツサーバから受け取るかどうか、かつ/またはどのように受け取るかを制御する機会がユーザに提供され得る。さらに、一定のデータが、記憶または使用される前に、個人識別可能情報が除去されるように1つまたは複数の方式で処理され得る。たとえば、ユーザについての個人識別可能情報が決定できないように、または地理的位置情報が取得される場合に(都市、郵便番号、州レベルなど)、ユーザの地理的位置が一般化され得、その結果、ユーザの特定の地理的位置を決定することができないように、ユーザの識別が処理され得る。したがって、ユーザは、ユーザについて情報がどのように収集され、かつ/または使用されるかに関して制御し得る。

20

【0064】

本明細書においていくつかの実装が説明され、図示されたが、機能を実施し、結果および/または本明細書において説明される利点のうちの1つまたは複数を取得するための様々な他の手段および/または構造が利用され得、そのような変形および/または修正のそれぞれは、本明細書において説明される実装の範囲内にあると見なされる。より一般には、本明細書において説明されるすべてのパラメータ、寸法、材料、および構成は例示的なものであり、実際のパラメータ、寸法、材料、および/または構成は、特定の適用例または教示が使用される適用例に依存する。型通りの実験を用いるだけで、本明細書において説明される特定の实装に対する多くの均等物を当業者は理解し、確認することができる。したがって、前述の実装は単に例として提示され、添付の特許請求の範囲およびその均等物の範囲内で、具体的に説明され、特許請求される以外の方法で実装が実施され得ることを理解されたい。本開示の実装は、本明細書において説明されるそれぞれの個々の特徴、システム、物品、材料、キット、および/または方法に向けられた。さらに、2つ以上のそのような特徴、システム、物品、材料、キット、および/または方法の任意の組合せが、そのような特徴、システム、物品、材料、キット、および/または方法が相互に矛盾しない場合、本開示の範囲内に含まれる。

30

40

【0065】

いくつかの実装では、1つまたは複数のプロセッサによって実装される方法が提供され、コンピューティングデバイスにおいて、音声発話内に含まれる、ユーザからの要求の第1の部分を受け取ることを含む。音声発話は、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自動アシスタントに向けられている。方法は、要求の第1の部分処理することに基づいて、要求の第1の部分に関連付けられる1つまたは複数の候補意図を決定することと、1つまたは複数の候補意図に基づいて、コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせることをさらに含む。汎用コンテナグラフィカル要素は、複数の異種の調整されたコンテナグラフィカル要素のいずれか1つに動的に適合され得、複数の異種の調整されたコンテナ

50

グラフィカル要素のそれぞれは、1つまたは複数の候補意図のうちの対応する1つに関連付けられる。方法は、コンピューティングデバイスにおいて、音声発話内にやはり含まれ、または音声発話の後に受け取られる追加の音声発話内に含まれる、ユーザからの要求の第2の部分を受け取ることと、要求の第2の部分进行处理することに基づいて、要求が1つまたは複数の候補意図のうちの特定の意図に対応することを決定することと、特定の意図に基づいて、ディスプレイインターフェースにおいてレンダリングされる汎用コンテナグラフィカル要素を、複数の異種の調整されたコンテナグラフィカル要素のうちの特定の調整されたコンテナグラフィカル要素に動的に適合させることとをさらに含む。

【0066】

任意選択で、本明細書において説明される技術のこれらおよび他の実装は、以下の特徴のうちの1つまたは複数を含む。

10

【0067】

いくつかの実装では、特定の調整されたコンテナグラフィカル要素は、特定の意図に関連付けられる、対応するパラメータについてのスロット値を特徴付け得、要求の第1の部分および第2の部分はスロット値を特定しないことがある。

【0068】

いくつかの実装では、方法は、要求の第1の部分を受け取ったことに応答して、要求の第1の部分の特徴付ける自然言語コンテンツをコンピューティングデバイスのディスプレイインターフェースに視覚的にレンダリングさせることをさらに含み得る。ディスプレイインターフェースは、汎用コンテナグラフィカル要素をレンダリングすると同時に、要求の第1の部分の自然言語コンテンツをレンダリングし得る。

20

【0069】

いくつかの実装では、方法は、要求の第1の部分を受け取った後に、しきい持続時間が経過したと判定することをさらに含み得る。汎用コンテナグラフィカル要素をディスプレイインターフェースにおいてレンダリングさせることは、しきい持続時間の経過に基づいて実施され得る。

【0070】

いくつかの実装では、特定の調整されたコンテナグラフィカル要素は、コンピューティングデバイスあるいはコンピューティングデバイスと通信している1つまたは複数の追加のコンピューティングデバイスの1つまたは複数の設定の現在の状態に関連付けられる特定の地理的制御要素を含み得る。そうした実装のいくつかのバージョンでは、方法は、コンピューティングデバイスにおいて、音声発話、追加の音声発話、または音声発話もしくは追加の音声発話の後に受け取られるさらに追加の音声発話内にやはり含まれる、ユーザからの要求の第3の部分を受け取ることと、要求の第3の部分が、1つまたは複数の設定についての更新後状態を含むことと、自動アシスタントにより、コンピューティングデバイスの1つまたは複数の設定、あるいは追加のコンピューティングデバイスの1つまたは複数、現在の状態から更新後状態に変更させることとをさらに含み得る。

30

【0071】

いくつかの実装では、ディスプレイインターフェースにおいてレンダリングされる汎用コンテナグラフィカル要素を特定の調整されたコンテナグラフィカル要素に動的に適合させることは、第2の音声発話内で識別されるスロット値のタイプに基づいて、複数の異種の調整されたコンテナグラフィカル要素の中から、特定の調整されたコンテナグラフィカル要素を選択することを含み得る。スロット値のタイプは、数値の範囲に限定される数値に対応し得る。

40

【0072】

いくつかの実装では、汎用コンテナグラフィカル要素は、要求の第2の部分を受け取る前に、コンピューティングデバイスのディスプレイインターフェースにおいてレンダリングされ得る。

【0073】

いくつかの実装では、汎用コンテナグラフィカル要素は、要求の第2の部分が受け取ら

50

れている間に、コンピューティングデバイスのディスプレイインターフェースにおいてレンダリングされ得る。

【0074】

いくつかの実装では、要求の第1の部分処理することに基づいて、要求の第1の部分に関連付けられる1つまたは複数の候補意図を決定することは、ストリーミング自動音声認識(ASR)モデルを使用して、コンピューティングデバイスの1つまたは複数のマイクロフォンによって生成されたオーディオデータのストリームを処理し、ASR出力を生成することと、オーディオデータのストリームが要求の第1の部分を取り込むことと、自然言語理解(NLU)モデルを使用して、ASR出力を処理し、NLU出力を生成することと、NLU出力に基づいて、要求の第1の部分に関連付けられる1つまたは複数の候補意図を決定することとを含み得る。そうした実装のいくつかのバージョンでは、要求の第2の部分処理することに基づいて、要求が1つまたは複数の候補意図のうち特定の意図に対応することを決定することは、ストリーミングASRモデルを使用して、オーディオデータのストリームを処理し、追加のASR出力を生成することと、オーディオデータのストリームが要求の第2の部分も取り込むことと、NLUモデルを使用して、追加のASR出力を処理し、追加のNLU出力を生成することと、追加のNLU出力に基づいて、1つまたは複数の候補意図の中から特定の意図を選択することとを含み得る。

10

【0075】

いくつかの実装では、1つまたは複数のプロセッサによって実装される方法が提供され、方法は、コンピューティングデバイスにおいて、ユーザによってサブMITされた要求の部分を受け取ることと、要求の部分が、コンピューティングデバイスを介してアクセス可能な自動アシスタントに向けられた音声発話内に含まれることと、自動アシスタントにより、要求の部分が、コンピューティングデバイスまたは自動アシスタントを介してコンピューティングデバイスと通信している1つまたは複数の追加のコンピューティングデバイスの1つまたは複数の設定の現在の状態を修正することと関連付けられると判定することと、1つまたは複数の設定の現在の状態に基づいて、1つまたは複数の設定の現在の状態を特徴付ける、調整されたコンテナグラフィカル要素データを決定することと、調整されたコンテナグラフィカル要素データに基づいて、1つまたは複数の設定の現在の状態を示す1つまたは複数の調整されたコンテナグラフィカル要素をコンピューティングデバイスのディスプレイインターフェースにレンダリングさせることと、1つまたは複数の設定の現在の状態を示す1つまたは複数の調整されたコンテナグラフィカル要素をコンピューティングデバイスのディスプレイインターフェースにレンダリングさせたことに応答して、コンピューティングデバイスにおいて、ユーザによってサブMITされた要求の追加の部分を受け取ることと、要求の追加の部分が、音声発話、または音声発話の後に受け取られる追加の音声発話内に含まれ、要求の追加の部分が、1つまたは複数の設定についての更新後状態を含むことと、自動アシスタントにより、コンピューティングデバイスの1つまたは複数の設定、あるいは追加のコンピューティングデバイスの1つまたは複数、現在の状態から更新後状態に変更させることとを含む。

20

30

【0076】

任意選択で、本明細書において説明される技術のこれらおよび他の実装は、以下の特徴のうち1つまたは複数を含む。

40

【0077】

いくつかの実装では、1つまたは複数の調整されたコンテナグラフィカル要素のそれぞれが、1つまたは複数の設定の現在の状態を表すためのグラフィカルアイコンを含み得る。

【0078】

いくつかの実装では、方法は、要求の部分を受け取ったことに応答して、要求の部分の特徴付ける自然言語コンテンツをコンピューティングデバイスのディスプレイインターフェースに視覚的にレンダリングさせることをさらに含み得る。1つまたは複数の調整されたグラフィカルコンテナ要素は、ディスプレイインターフェースが自然言語コンテンツをレンダリングすると同時にレンダリングされ得る。

50

【 0 0 7 9 】

いくつかの実装では、要求の部分は、1つまたは複数の設定の現在の状態を含まないことがある。

【 0 0 8 0 】

いくつかの実装では、調整されたコンテナグラフィカル要素データに基づいて、1つまたは複数の設定の現在の状態を示す1つまたは複数の調整されたコンテナグラフィカル要素をコンピューティングデバイスのディスプレイインターフェースにレンダリングさせることは、コンピューティングデバイスとは別々の、1つまたは複数の追加のコンピューティングデバイスの中の第1のコンピューティングデバイスの1つまたは複数の設定の中の第1の設定を示す、1つまたは複数の調整されたコンテナグラフィカル要素の中の第1の調整されたコンテナグラフィカル要素をレンダリングさせることと、コンピューティングデバイスとは別々の、1つまたは複数の追加のコンピューティングデバイスの中の第2のコンピューティングデバイスの1つまたは複数の設定の中の第2の設定を示す、1つまたは複数の調整されたコンテナグラフィカル要素の中の第2の調整されたコンテナグラフィカル要素をレンダリングさせることとを含み得る。そうした実装のいくつかのバージョンでは、第1のコンピューティングデバイスの第1の設定は、第1のコンピューティングデバイスの音量設定に対応し得、第2のコンピューティングデバイスの第2の設定は、第2のコンピューティングデバイスの音量設定に対応し得る。そうした実装の追加または代替のバージョンでは、第1のコンピューティングデバイスの第1の設定は、第1のコンピューティングデバイスの輝度設定に対応し得、第2のコンピューティングデバイスの第2の設定は、第2のコンピューティングデバイスの輝度設定に対応し得る。

10

20

【 0 0 8 1 】

いくつかの実装では、方法は、要求の部分を処理することに基づいて、ユーザが要求を与える際に休止したと判定することをさらに含み得る。1つまたは複数の設定の現在の状態を示す1つまたは複数の調整されたコンテナグラフィカル要素をコンピューティングデバイスのディスプレイインターフェースにレンダリングさせることは、ユーザが要求を与える際に休止したとの判定に回答するものであり得る。そうした実装のいくつかのバージョンでは、要求の部分を処理することに基づいて、ユーザが要求を与える際に休止したと判定することは、要求の部分を処理することに基づいて、ユーザが要求を与える際に特定のワードまたはフレーズを与えた後に休止したと判定することを含み得る。そうした実装の追加または代替のバージョンでは、1つまたは複数の設定の現在の状態を示す1つまたは複数の調整されたコンテナグラフィカル要素をコンピューティングデバイスのディスプレイインターフェースにレンダリングさせることは、ユーザが要求を与える際にしきい持続時間にわたって休止したとの判定に回答するものであり得る。そうした実装の追加または代替のバージョンでは、要求の部分を処理することに基づいて、ユーザが要求を与える際に休止したと判定することは、ストリーミング自動音声認識(ASR)モデルを使用して、コンピューティングデバイスの1つまたは複数のマイクロフォンによって生成されたオーディオデータのストリームを処理し、ASR出力を生成することであって、オーディオデータのストリームが要求の部分を取り込むことと、自然言語理解(NLU)モデルを使用して、ASR出力を処理し、NLU出力を生成することと、NLU出力に基づいて、ユーザが要求を与える際に休止したと判定することとを含み得る。そうした実装の追加または代替のバージョンでは、要求の部分を処理することに基づいて、ユーザが要求を与える際に休止したと判定することは、要求の部分を処理することに基づいて、要求の部分に関連付けられるオーディオベースの特性を決定することと、要求の部分に関連付けられるオーディオベースの特性に基づいて、ユーザが要求を与える際に休止したと判定することとを含み得る。

30

40

【 0 0 8 2 】

いくつかの実装では、1つまたは複数のプロセッサによって実装される方法が提供され、方法は、コンピューティングデバイスにおいて、音声発話内に含まれる、ユーザからの要求の第1の部分を受け取ることを含む。音声発話は、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自動アシスタントに向けられた。方法は、要求の第1の部分処

50

理することに基づいて、要求の部分が、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1つまたは複数の追加のコンピューティングデバイスを制御することに関連付けられる特定のワードまたはフレーズを含むと判定することと、特定のワードまたはフレーズを含む要求の第1の部分に基づいて、コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせることとをさらに含む。汎用コンテナグラフィカル要素は、複数の異種の調整されたコンテナグラフィカル要素のいずれか1つに動的に適合され得、複数の異種の調整されたコンテナグラフィカル要素のそれぞれが、音声発話の第1の部分処理することに基づいて決定された、対応する意図に関連付けられる。方法は、コンピューティングデバイスにおいて、音声発話内にやはり含まれ、または音声発話に続いて受け取られる追加の音声発話内に含まれる、ユーザからの要求の第2の部分を受け取ることと、要求の第2の部分処理することに基づいて、要求が1つまたは複数の候補意図のうち特定の意図に対応することを決定することと、特定の意図に基づいて、ディスプレイインターフェースにおいてレンダリングされる汎用コンテナグラフィカル要素を、複数の異種の調整されたコンテナグラフィカル要素のうち特定の調整されたコンテナグラフィカル要素に動的に適合させることとをさらに含む。

10

【0083】

任意選択で、本明細書において説明される技術のこれらおよび他の実装は、以下の特徴のうち1つまたは複数を含む。

【0084】

いくつかの実装では、要求の第1の部分処理することに基づいて、要求の部分が、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1つまたは複数の追加のコンピューティングデバイスを制御することに関連付けられる特定のワードまたはフレーズを含むと判定することは、ストリーミング自動音声認識(ASR)モデルを使用して、コンピューティングデバイスの1つまたは複数のマイクロフォンによって生成されたオーディオデータのストリームを処理し、ASR出力を生成することであって、オーディオデータのストリームが要求の第1の部分を取り込むことと、ASR出力に基づいて、要求の部分がコンピューティングデバイスまたは1つまたは複数の追加のコンピューティングデバイスを制御することに関連付けられる特定のワードまたはフレーズを含むと判定することとを含み得る。

20

30

【0085】

そうした実装のいくつかのバージョンでは、特定のワードまたはフレーズを含む要求の第1の部分に基づいて、コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせることは、特定のワードまたはフレーズがコンピューティングデバイスのオンデバイスメモリ内で汎用コンテナグラフィカル要素にマッピングされると判定することと、特定のワードまたはフレーズが汎用コンテナグラフィカル要素にマッピングされるとの判定に回答して、自然言語理解(NLU)モデルを使用してASR出力を処理することなく、コンピューティングデバイスのディスプレイインターフェースにおいて汎用コンテナグラフィカル要素をレンダリングさせることとを含み得る。

40

【0086】

別の実装は、前述の方法および/または本明細書の別の箇所において説明される方法のうち1つまたは複数などの方法を実施するように1つまたは複数のプロセッサ(たとえば、中央演算処理装置(CPU)、グラフィックス処理装置(GPU)、および/またはテンソル処理装置(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。さらに別の実装は、記憶された命令を実行して、前述の方法および/または本明細書の別の箇所において説明される方法のうち1つまたは複数などの方法を実施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のシステムを含み得る。

【0087】

50

本明細書においてより詳細に説明される前述の概念および追加の概念のすべての組合せが、本明細書において開示される主題の部分であるものとして企図されることを理解されたい。たとえば、本開示の終わりに現れる、特許請求される主題のすべての組合せが、本明細書において開示される主題の部分であるものとして企図される。

【符号の説明】

【 0 0 8 8 】

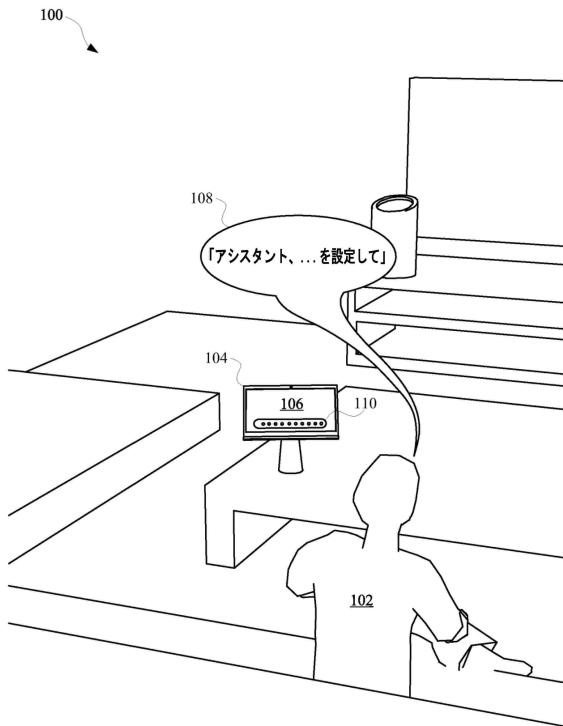
102	ユーザ	
104	コンピューティングデバイス	
106	ディスプレイインターフェース	
108	要求	10
110	汎用コンテナグラフィカル要素	
112	グラフィカルコンテナ要素	
120	現在の状態	
122	要求	
124	メディアコンテンツ	
142	要求	
144	制御要素	
202	ユーザ	
204	コンピューティングデバイス	
206	ディスプレイインターフェース	20
208	要求	
210	汎用コンテナグラフィカル要素	
222A	グラフィカルコンテナ要素	
222B	グラフィカルコンテナ要素	
222C	グラフィカルコンテナ要素	
224	デバイス識別子	
226	要求	
228	デバイス識別子	
230	デバイス識別子	
232	自然言語コンテンツ	30
242	要求	
300	システム	
302	コンピューティングデバイス	
304	自動アシスタント	
306	入力処理エンジン	
308	音声処理エンジン	
310	データパーシングエンジン	
312	パラメータエンジン	
314	出力生成エンジン	
316	候補意図エンジン	40
318	汎用コンテナエンジン	
320	アシスタントインターフェース	
322	アシスタント呼出しエンジン	
324	状態エンジン	
326	状態エンジン	
330	アプリケーションデータ	
332	デバイスデータ	
334	特定のアプリケーション	
336	コンテキストデータ	
510	コンピュータシステム	50

- 512 バスサブシステム
- 514 プロセッサ
- 516 ネットワークインターフェースサブシステム
- 520 ユーザインターフェース出力デバイス
- 522 ユーザインターフェース入力デバイス
- 524 記憶サブシステム
- 526 ファイル記憶サブシステム
- 525 メモリ
- 526 ファイル記憶サブシステム
- 530 メインランダムアクセスメモリ(RAM)
- 532 読取り専用メモリ(ROM)

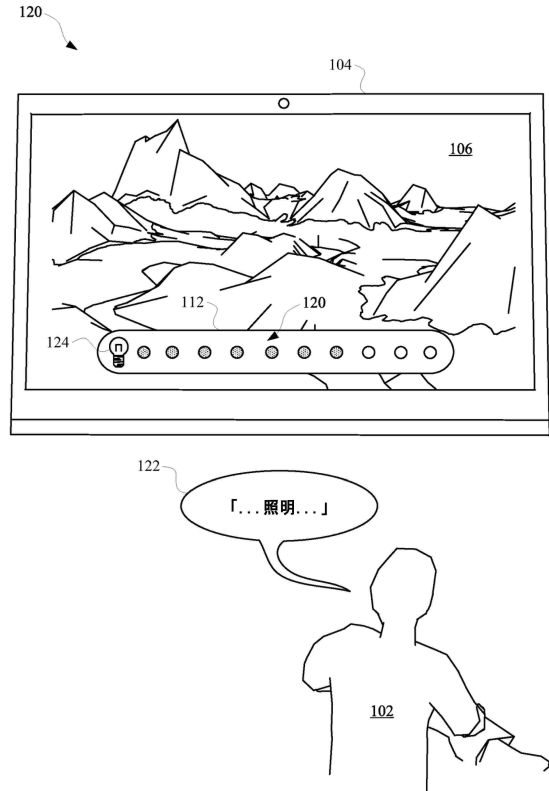
10

【図面】

【図 1 A】



【図 1 B】



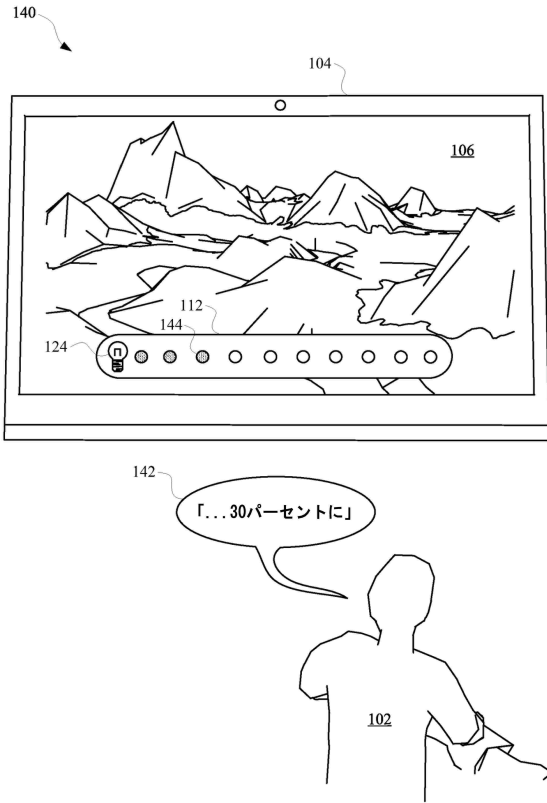
20

30

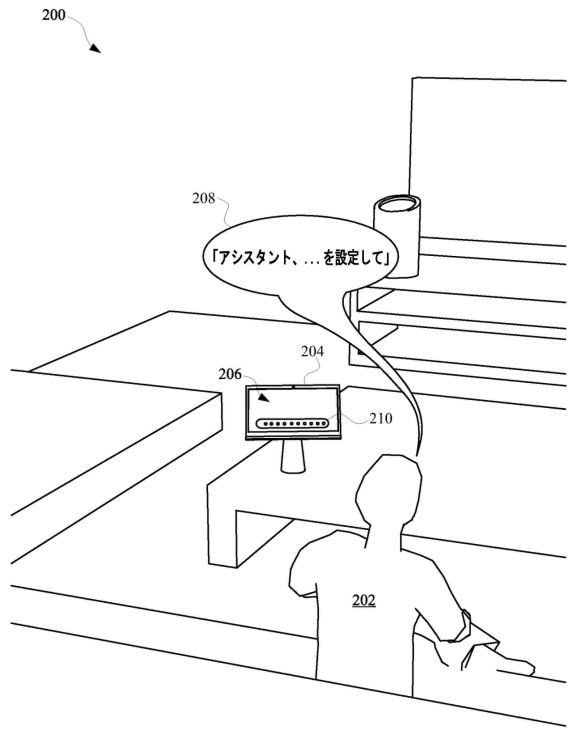
40

50

【図 1 C】



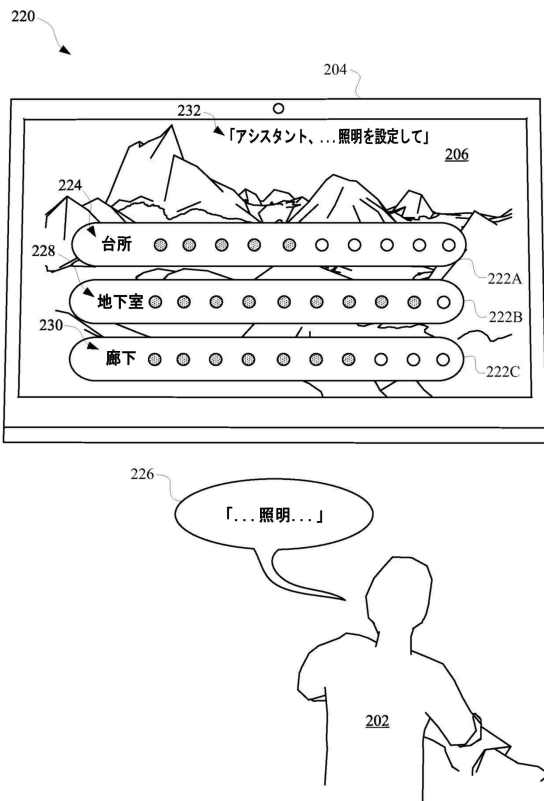
【図 2 A】



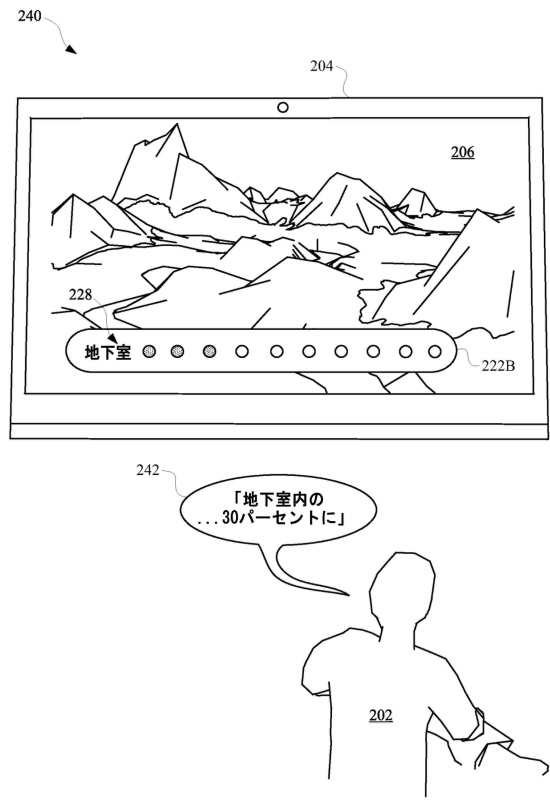
10

20

【図 2 B】



【図 2 C】

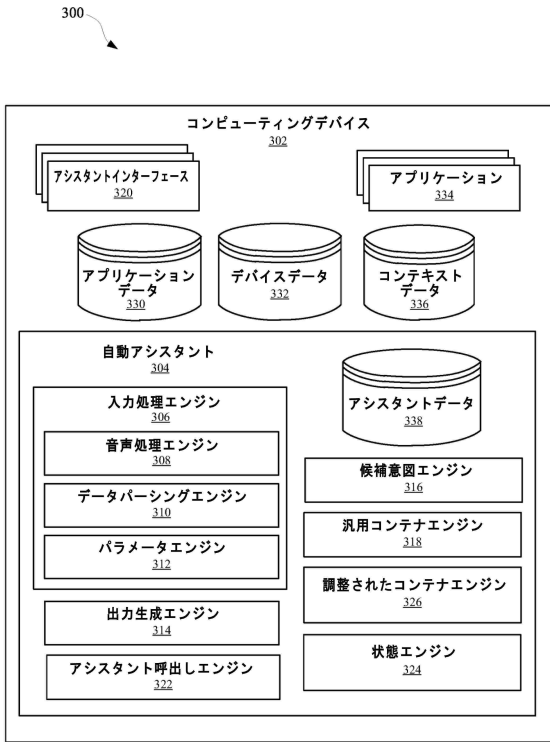


30

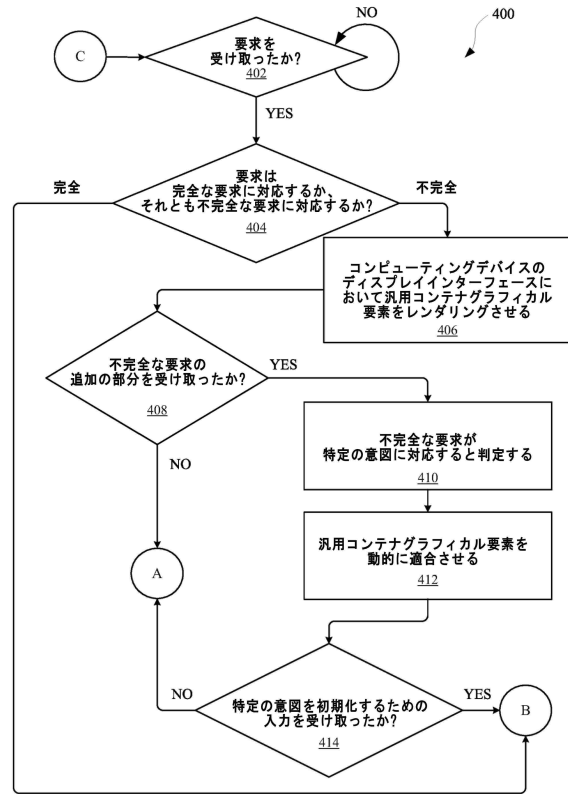
40

50

【図3】



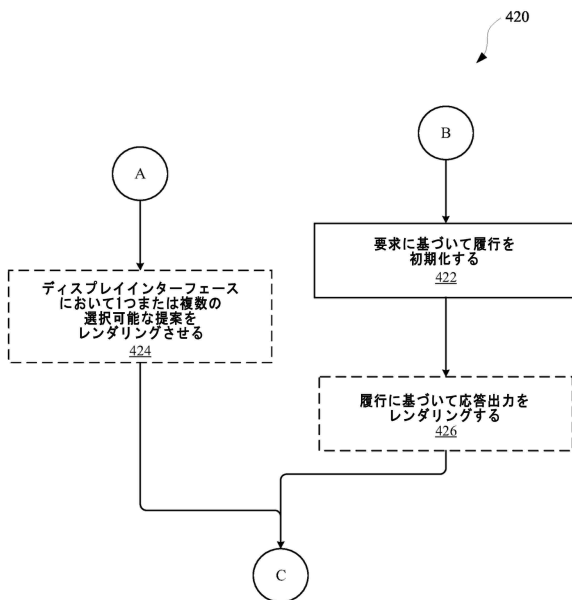
【図4A】



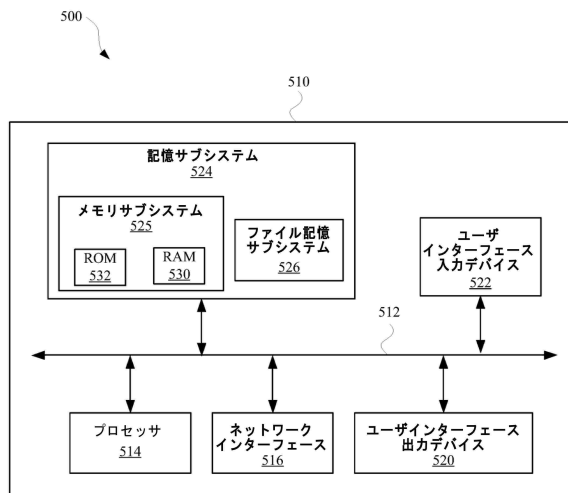
10

20

【図4B】



【図5】



30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(72)発明者 ブレット・バロス

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 ジョアン・ジェイ・ジャン

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 アンドリュー・ショーネウァイス

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 三田村 陽平

(56)参考文献 国際公開第2020/137607(WO, A1)

国際公開第2019/216874(WO, A1)

欧州特許出願公開第03783867(E P, A1)

米国特許出願公開第2018/0285070(US, A1)

特表2019-522250(JP, A)

(58)調査した分野 (Int.Cl., DB名)

G06F 3/16

G10L 15/00 - 17/26