

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7111682号
(P7111682)

(45)発行日 令和4年8月2日(2022.8.2)

(24)登録日 令和4年7月25日(2022.7.25)

(51)国際特許分類	F I			
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10	2 0 0 W		
G 0 6 F 3/16 (2006.01)	G 0 6 F 3/16	6 3 0		
G 1 0 L 15/28 (2013.01)	G 0 6 F 3/16	6 2 0		
G 1 0 L 15/22 (2006.01)	G 1 0 L 15/28	5 0 0		
	G 1 0 L 15/22	4 5 3		
請求項の数 20 外国語出願 (全32頁)				

(21)出願番号	特願2019-207637(P2019-207637)	(73)特許権者	595097771 アドビ インコーポレイテッド アメリカ合衆国, カリフォルニア 9 5 1 1 0 , サン ノゼ , パーク アベニュー 3 4 5
(22)出願日	令和1年11月18日(2019.11.18)	(74)代理人	100098394 弁理士 山川 茂樹
(65)公開番号	特開2020-118955(P2020-118955 A)	(74)代理人	100064621 弁理士 山川 政樹
(43)公開日	令和2年8月6日(2020.8.6)	(72)発明者	マーク・シイ・ウェブスター アメリカ合衆国・ニューヨーク州・ニュ ーヨーク
審査請求日	令和2年11月13日(2020.11.13)	(72)発明者	スコット・トーマス・ワーナー アメリカ合衆国・ニューヨーク州・ブル ックリン
(31)優先権主張番号	62/794,466		
(32)優先日	平成31年1月18日(2019.1.18)		
(33)優先権主張国・地域又は機関	米国(US)		
最終頁に続く			

(54)【発明の名称】 非表音文字体系を使用する言語のための音声支援型アプリケーションプロトタイプの実験中の音声コマンドマッチング

(57)【特許請求の範囲】

【請求項1】

音声機能を有するアプリケーションプロトタイプの実験を可能とするデジタル媒体環境において、少なくとも1つの計算デバイスによって実装される方法であって、

前記方法は：

前記少なくとも1つの計算デバイスによって、前記アプリケーションプロトタイプに関連する対話データをストレージ内に保持するステップであって、前記対話データは、保存済みの音声コマンド、各前記保存済みの音声コマンドに回答して実施されることになるアクション、及び少なくとも1つのビジュアルページを含む、ステップ；

前記アプリケーションプロトタイプのための試験段階中に、前記少なくとも1つの計算デバイスの表示デバイス上に、前記少なくとも1つのビジュアルページを表示するステップ；

少なくとも1つのマイクロフォンによって、非表音文字体系を使用する言語で発話された非表音音声コマンドをキャプチャするステップ；

前記非表音音声コマンドを発話テキスト化サービスへとリアルタイムで通信して、前記非表音音声コマンドを、前記音声コマンドの、前記非表音文字体系の非表音テキスト文字列へと変換するステップ；

前記音声コマンドの前記非表音テキスト文字列を、表音言語翻訳機へとリアルタイムで通信して、前記非表音テキスト文字列を、前記音声コマンドの、表音文字体系の表音テキスト文字列に翻訳するステップ；

10

20

前記音声コマンドの前記表音テキスト文字列を、前記アプリケーションプロトタイプに関連する前記保存済みの音声コマンドの、前記表音文字体系の表音テキスト文字列と比較して、マッチングする音声コマンドを識別するステップ；並びに

前記ストレージ中で、前記マッチングする音声コマンドに関連するアクションを実施するステップを含む、方法。

【請求項 2】

前記保存済みの音声コマンドは、前記試験段階の前に、前記非表音文字体系の非表音テキスト文字列として受信される、請求項 1 に記載の方法。

【請求項 3】

前記保存済みの音声コマンドは、前記ストレージ内に、前記保存済みの音声コマンドの前記非表音テキスト文字列として保持され、

前記表音言語翻訳機は、前記試験段階中に、前記保存済みの音声コマンドの前記非表音テキスト文字列を、前記保存済みの音声コマンドの、前記表音文字体系の前記表音テキスト文字列に翻訳する、請求項 2 に記載の方法。

【請求項 4】

前記表音言語翻訳機は、前記試験段階の前に、前記保存済みの音声コマンドの前記非表音テキスト文字列を、前記保存済みの音声コマンドの、前記表音文字体系の前記表音テキスト文字列に翻訳し、

前記保存済みの音声コマンドは、前記ストレージ内に、前記保存済みの音声コマンドの前記表音テキスト文字列として保持される、請求項 2 に記載の方法。

【請求項 5】

前記比較は、前記音声コマンドの前記表音テキスト文字列が、前記保存済みの音声コマンドの前記表音テキスト文字列と共通して有する文字の数を比較することを含む、請求項 1 に記載の方法。

【請求項 6】

前記マッチングする音声コマンドは、キャプチャされた前記音声コマンドの前記表音テキスト文字列と一致する文字を最も多く有する、前記マッチングする音声コマンドの前記表音テキスト文字列に、少なくとも部分的に基づいて識別される、請求項 5 に記載の方法。

【請求項 7】

前記音声コマンドの前記非表音テキスト文字列は、北京官話で書かれており、前記音声コマンドの、翻訳された前記表音テキスト文字列は、ピンインで書かれている、請求項 1 に記載の方法。

【請求項 8】

前記音声コマンドの前記非表音テキスト文字列は、漢字で書かれており、前記音声コマンドの、翻訳された前記表音テキスト文字列は、ローマ字で書かれている、請求項 1 に記載の方法。

【請求項 9】

前記音声コマンドの前記表音テキスト文字列と、前記保存済みの音声コマンドの前記表音テキスト文字列とは、同一の表音文字体系で書かれている、請求項 1 に記載の方法。

【請求項 10】

前記表音言語翻訳機は、非表音文字体系を使用する言語の非表音テキスト文字列を、前記非表音テキスト文字列をラテン文字へとローマ文字化することによって、翻訳する、請求項 1 に記載の方法。

【請求項 11】

アプリケーションプロトタイプの少なくとも 1 つのビジュアルページを、前記アプリケーションプロトタイプの試験段階中に、表示デバイスに表示するための、試験モジュール；前記アプリケーションプロトタイプの前記試験段階中に、少なくとも 1 つのマイクロフォンによってキャプチャされた、非表音文字体系を使用する言語で発話された非表音音声コマンドを、前記音声コマンドの、非表音文字体系の非表音テキスト文字列に変換するた

10

20

30

40

50

めの、発話テキスト化サービス；

前記音声コマンドの前記非表音テキスト文字列を、前記音声コマンドの、表音文字体系の表音テキスト文字列に翻訳するための、表音言語翻訳機；

前記音声コマンドの前記表音テキスト文字列を、前記アプリケーションプロトタイプに関連する保存済みの音声コマンドの、前記表音文字体系の表音テキスト文字列と比較して、マッチングする音声コマンドを識別するための、比較モジュール；及び

前記マッチングする音声コマンドに関連するアクションを実施するための、実施モジュール

を備える、システム。

【請求項 1 2】

前記比較モジュールは、前記マッチングする音声コマンドを：

前記音声コマンドの前記表音テキスト文字列が、前記保存済みの音声コマンドの前記表音テキスト文字列と共通して有する文字の数を比較し；

キャプチャされた前記音声コマンドの前記表音テキスト文字列と一致する文字を最も多く有する、前記マッチングする音声コマンドの前記表音テキスト文字列に、少なくとも部分的に基づいて、前記マッチングする音声コマンドを識別する

ことによって、識別するよう構成される、請求項 1 1 に記載のシステム。

【請求項 1 3】

前記音声コマンドの前記非表音テキスト文字列は、北京官話で書かれており、

前記音声コマンドの、翻訳された前記表音テキスト文字列は、ピンインで書かれている、請求項 1 1 に記載のシステム。

【請求項 1 4】

前記音声コマンドの前記非表音テキスト文字列は、漢字で書かれており、

前記音声コマンドの、翻訳された前記表音テキスト文字列は、ローマ字で書かれている、請求項 1 1 に記載のシステム。

【請求項 1 5】

前記実施モジュールは、前記アクションを：

前記マッチングする音声コマンドに関連する発話応答を識別し；

前記発話応答のテキスト文字列をテキスト発話化サービスに提供し；

前記テキスト発話化サービスから、前記発話応答のオーディオファイルを受信し；

前記オーディオファイルをプレイバックすることにより、前記発話応答を出力することによって、実施するよう構成される、請求項 1 1 に記載のシステム。

【請求項 1 6】

前記実施モジュールは、前記アクションを：

前記マッチングする音声コマンドに関連するページ遷移を識別し、ここで前記ページ遷移は、前記アプリケーションプロトタイプの追加のビジュアルページを識別し；

前記アプリケーションプロトタイプの前記追加のビジュアルページを表示することによって、実施するよう構成される、請求項 1 1 に記載のシステム。

【請求項 1 7】

前記保存済みの音声コマンドを、前記試験段階の前に、前記非表音文字体系の非表音テキスト文字列として受信する、請求項 1 1 に記載のシステム。

【請求項 1 8】

前記保存済みの音声コマンドは、ストレージ中に、前記保存済みの音声コマンドの前記非表音テキスト文字列として保持され、

前記表音言語翻訳機は更に、前記試験段階中に、前記保存済みの音声コマンドの前記非表音テキスト文字列を、前記保存済みの音声コマンドの前記表音テキスト文字列に翻訳するよう構成される、請求項 1 7 に記載のシステム。

【請求項 1 9】

前記表音言語翻訳機は更に、前記試験段階の前に、前記保存済みの音声コマンドの前記非表音テキスト文字列を、前記保存済みの音声コマンドの、前記表音文字体系の前記表音

10

20

30

40

50

テキスト文字列に翻訳するよう構成され、

前記保存済みの音声コマンドは、ストレージ中に、前記保存済みの音声コマンドの前記表音テキスト文字列として保持される、請求項 17 に記載のシステム。

【請求項 20】

前記表音言語翻訳機は、非表音文字体系を使用する言語の非表音テキスト文字列を、前記非表音テキスト文字列をラテン文字へとローマ文字化することによって、翻訳するよう構成される、請求項 11 に記載のシステム。

【発明の詳細な説明】

【背景技術】

【0001】

ユーザは、ますます多くの異なる音声支援型タスクを達成するために、音声アシスタントデバイスと対話している。一般に、このような音声アシスタントデバイスは、ユーザから音声コマンドを受信するための 1 つ以上のマイクロフォンを用いて構成される。専用の音声アシスタントデバイスとしては、Amazon (登録商標) Echo 及び Google (登録商標) Home が挙げられ、これらはそれぞれ、音声アシスタントプラットフォーム Amazon (登録商標) Alexa 及び Google (登録商標) Assistant との対話を可能とする。しかしながら、携帯電話、デスクトップコンピュータ、ラップトップコンピュータ、ゲーミングシステム等といった他のデバイスも、音声アシスタントプラットフォームの機能を活用できる音声アシスタントデバイスとして構成できる。例えば、Apple (登録商標) iPhone (登録商標) (携帯電話) は、ユーザが音声アシスタントプラットフォーム Siri (登録商標) と対話できるようにする音声アシスタントデバイスとして機能できる。

【0002】

このような音声アシスタントデバイスは、1 つ以上のスピーカーによって可聴 (音声) 応答を出力すること；視覚コンテンツを表示すること；他の視覚効果 (例えば発光ダイオード (LED) による指示) を提供すること；触覚フィードバックを提供すること等によって、ユーザの音声コマンドに対して、多様な異なる方法で応答するよう構成される。更に、音声アシスタントプラットフォームは、例えばサーモスタット、スピーカー、照明といった他のデバイスに、ユーザの音声コマンドに対して応答させることができる。家電製品に関する応答のいくつかの例としては：家屋の温度を調整するためにサーモスタットの温度を変更すること；音楽サービスからスピーカーでの音楽の再生を開始すること；家屋内の特定の照明をオン又はオフにすること等が挙げられる。

【発明の概要】

【発明が解決しようとする課題】

【0003】

音声対話がより一般的になるに従って、設計者は、音声機能をアプリケーション又はウェブページのユーザインタフェースのビジュアルページに追加すること等によって、デジタルインタフェースに音声対話機能を追加し始めている。しかしながら、従来の設計アプリケーションは、アプリケーションプロトタイプのための音声機能の設計及び試験をサポートできない。従って、アプリケーションプロトタイプの設計及び音声支援型試験は困難である。更に、場合によっては、ユーザは、中国語又は日本語といった非表音文字体系を使用する言語を用いて、アプリケーションプロトタイプと対話することを望むが、これは高い割合のマッチングエラーにつながる。

【課題を解決するための手段】

【0004】

これらの課題を克服するために、非表音文字体系を使用する言語のための音声支援型アプリケーションプロトタイプの試験中の音声コマンドマッチングについて記載する。音声支援型アプリケーションプロトタイプのための音声対話ツールは、デジタル媒体環境において活用されている。設計アプリケーションは、ユーザがあるアプリケーションのためのアプリケーションプロトタイプを設計できるようにする、設計インタフェースを提供する

10

20

30

40

50

。一般に、上記アプリケーションプロトタイプは、一連のビジュアルページ、並びに上記ビジュアルページに関連付けられた様々なトリガー及び対応するアクションを含む。上記アプリケーションプロトタイプの設計を可能とするために、上記アプリケーションプロトタイプの少なくとも1つのビジュアルページを、上記設計インタフェースに表示する。上記設計インタフェースを制御することにより、上記アプリケーションプロトタイプの上記少なくとも1つのビジュアルページに関するトリガー及び関連するアクションを受信するための、対話インタフェースを提供する。上記トリガーは、音声コマンド、ユーザジェスチャー、又は時間遅延のうちの1つに対応してよく、上記アクションは、発話応答、上記アプリケーションプロトタイプの追加のビジュアルページへのページ遷移、又はメディアファイルのプレイバックのうちの1つに対応してよい。ユーザ入力を上記設計インタフェースによって受信することで、上記トリガー及び上記アクションを提供する。上記ユーザ入力に応答して、上記アプリケーションプロトタイプの上記トリガー、上記アクション、及び上記ビジュアルページを含むように、関連対話データが生成される。上記関連対話データを保存することにより、上記アプリケーションプロトタイプの試験段階中に、上記トリガー及び上記アクションを試験することが可能となる。

10

【0005】

試験段階では、上記アプリケーションプロトタイプの少なくとも1つのビジュアルページが表示され、トリガーが検出される。上記トリガーは、一部を挙げると、ユーザから受信した音声コマンド、ユーザジェスチャー、又は時間遅延の満了に対応してよい。上記トリガーを、上記アプリケーションプロトタイプに関連する保存済みのトリガーと比較することにより、マッチングするトリガーを識別する。上記トリガーが音声コマンドに対応する場合、上記音声コマンドをリアルタイムでテキスト文字列に変換する音声テキスト化サービスに、上記音声コマンドを提供する。続いて上記音声コマンドの上記テキスト文字列を、上記アプリケーションプロトタイプに関連する音声コマンドの保存済みテキスト文字列と比較することにより、マッチングする音声コマンドを識別する。

20

【0006】

上記マッチングするトリガーの識別に応答して、上記マッチングするトリガーに関連するアクションを、上記試験段階中に実施する。上記アクションは、発話応答、異なるビジュアルページへのページ遷移、又はメディアファイルのプレイバックに対応してよい。上記アクションが発話応答に対応する場合、上記発話応答をリアルタイムでオーディオファイルに変換するテキスト音声化サービスに、上記発話応答のテキスト文字列を提供してよい。

30

【0007】

1つ以上の実装形態では、上記設計アプリケーションは、表音言語翻訳機を利用して、音声コマンドのテキスト文字列を、非表音文字体系を使用する言語から、上記テキスト文字列の表音バージョンに変換する。一例として、上記表音言語翻訳機は、北京官話、並びにカタカナ、ひらがな、及び漢字の組み合わせを、それぞれピンイン及びローマ字に翻訳してよい。上記試験段階では、非表音文字体系を使用する言語で発話された非表音音声コマンドを、少なくとも1つのマイクロフォンでキャプチャする。上記非表音音声コマンドは、上記非表音音声コマンドを、上記音声コマンドの上記非表音文字体系の言語の非表音テキスト文字列へと変換するための発話テキスト化サービスへと、リアルタイムで通信される。次に、上記音声コマンドの上記非表音テキスト文字列は、上記非表音テキスト文字列を上記音声コマンドの表音文字体系の表音テキスト文字列に翻訳するための表音言語翻訳機へと、リアルタイムで通信される。上記音声コマンドの上記表音テキスト文字列を、上記アプリケーションプロトタイプに関連する上記表音文字体系の保存済み音声コマンドの表音テキスト文字列と比較して、マッチングする音声コマンドを識別する。そして、上記マッチングする音声コマンドに関連するアクションを実施する。

40

【0008】

この概要は、「発明を実施するための形態」において更に後述される複数の概念から選択されたものを、簡略化された形式で紹介している。従ってこの概要は、請求対象の主題

50

の本質的な特徴を識別することを意図したものではなく、請求対象の主題の範囲の決定を補助するものとして使用されることも意図していない。

【0009】

「発明を実施するための形態」は、添付の図面を参照して記述される。

【図面の簡単な説明】

【0010】

【図1】図1は、本明細書に記載の音声支援型アプリケーションプロトタイプのための音声対話ツールを使用するために動作可能なある例示的実装形態の環境の図である。

【図2】図2は、図1の音声対話ツールが、対話データと、アプリケーションプロトタイプのビジュアルページとの間の関連を生成し、上記対話データ及び上記ビジュアルページを保持する、ある例示的実装形態を示す。

10

【図3】図3は、図1の設計アプリケーションの試験モジュールがアプリケーションプロトタイプの試験を容易にする、ある例示的実装形態を示す。

【図4】図4は、図1の設計アプリケーションの試験モジュールが、表音言語翻訳機を利用して、アプリケーションプロトタイプの試験中にキャプチャされた音声コマンドをマッチングさせる、ある例示的実装形態を示す。

【図5A】図5Aは、音声支援型アプリケーションプロトタイプの設計に使用される設計アプリケーションの設計インタフェースの例を示す。

【図5B】図5Bは、音声支援型アプリケーションプロトタイプの設計に使用される設計アプリケーションの設計インタフェースの例を示す。

20

【図5C】図5Cは、音声支援型アプリケーションプロトタイプの設計に使用される設計アプリケーションの設計インタフェースの例を示す。

【図6A】図6Aは、音声支援型アプリケーションプロトタイプの試験段階中の、設計アプリケーションの設計インタフェースの例を示す。

【図6B】図6Bは、音声支援型アプリケーションプロトタイプの試験段階中の、設計アプリケーションの設計インタフェースの例を示す。

【図7】図7は、音声支援型アプリケーションプロトタイプを設計する例示の手順を示す。

【図8】図8は、音声支援型アプリケーションプロトタイプを試験する例示の手順を示す。

【図9】図9は、表音言語翻訳を利用して、アプリケーションプロトタイプの試験中にキャプチャされた音声コマンドをマッチングさせる、例示の手順を示す。

30

【図10】図10は、本明細書に記載の技法の実施形態を実装するために、図1～9を参照して説明及び/又は利用されているいずれのタイプの計算デバイスとして実装可能な例示的デバイスの、様々な部品を含む例示的システムを示す。

【発明を実施するための形態】

【0011】

概観

ユーザは、ますます多くの異なる音声支援型タスクを達成するために、音声アシスタントデバイスと対話している。音声支援型対話によってこれらのタスクを達成するために、これを実施するための機能を、ソフトウェアアプリケーション、ファームウェア、サービス、プラグイン、スクリプト等の設計、開発及び配備によって、実現しなければならない。しかしながら、アプリケーションプロトタイプを設計するための従来のツールは、アプリケーションプロトタイプのための音声支援型機能を設計及び試験するユーザの能力を制限するものである。

40

【0012】

これらの課題を克服するために、音声支援型アプリケーションプロトタイプのための音声対話ツールを、デジタル媒体環境で活用する。Adobe（登録商標）XD等の設計アプリケーションは、ユーザ（例えば設計者）が、あるアプリケーションのためのアプリケーションプロトタイプを設計できるようにする、設計インタフェースを提供する。一般に、上記アプリケーションプロトタイプは、上記アプリケーションのビジュアルページ、並びに上記ビジュアルページに関連する様々なトリガー及び対応するアクションを含む。ピ

50

ジュアルページの設計の一部として、上記設計アプリケーションは、上記アプリケーションプロトタイプのためのユーザインタフェースコントロール、グラフィック及びテキストの設計及びレイアウトを可能とするツール及び機能を提供する。例えば設計インタフェースは、上記アプリケーションプロトタイプの複数のビジュアルページのシーケンス（例えばホームスクリーンページ及び他の様々なビジュアルページ）を設計でき、上記アプリケーションプロトタイプとの対話中に、これらのビジュアルページへとナビゲートできる。

【0013】

機能するアプリケーションプロトタイプの設計を可能とするために、上記設計アプリケーションの音声対話ツールは、音声コマンドと発話プレイバックとの両方を、上記アプリケーションプロトタイプの設計に直接統合することを可能とする。これを実現するために、音声対話ツールは、設計インタフェースを制御して、上記アプリケーションプロトタイプのあるビジュアルページに関するトリガー及び関連するアクションを提供するためのコントロールを含む、対話インタフェースを提供する。本明細書に記載されるように、用語「トリガー (t r i g g e r) 」は、対応するアクションを誘発できるコマンドを指し、「アクション (a c t i o n) 」は、対応するトリガーの検出にตอบสนองして実施される動作である。トリガーの例としては例えば、音声コマンド、ユーザジェスチャー（例えばタップ、ドラッグ、スワイプ、キーストローク）、及び時間遅延が挙げられるが、これらに限定されない。アクションの例としては例えば、発話応答、アプリケーションプロトタイプの他のビジュアルページへのページ遷移、及びメディアファイル（例えばオーディオ又はビデオファイル）のプレイバックが挙げられるが、これらに限定されない。

【0014】

対話インタフェースにより、ユーザは、定義されたトリガー及びアクションのリストから、トリガー及び関連するアクションを選択できる。選択後、ユーザは、選択したトリガー及びアクションを記述する入力を提供する。例えば、ある音声コマンドを上記アプリケーションプロトタイプのある特定のビジュアルページに関連付けるために、音声対話ツールによって、ユーザは、（例えばタッチ機能を用いて）上記特定のビジュアルページを選択し、対話インタフェースを介して（例えばコマンド又は応答をタイピングすること、コマンド又は応答をマイクロフォン又は通信可能に連結されたデバイスに対して発話すること等によって）、上記特定の要素に対応する異なる複数の音声コマンド又は発話応答を指定できる。音声コマンドは、音声支援型アプリケーションに対するユーザの発話を示すオーディオデータ又はテキストデータとして構成してよい。発話応答は、音声コマンドの受信にตอบสนองした、音声支援型アプリケーションの（例えば音声アシスタントによる）可聴出力を表す。

【0015】

特に、対話インタフェースにより、ユーザは、様々な異なるタイプのトリガー及びアクションを、上記アプリケーションプロトタイプのビジュアルページに関連付けることができる。例えばユーザは、音声コマンドトリガーを対応する発話応答に関連付けることができる。しかしながら、設計アプリケーションにより、ユーザは、例えばあるユーザジェスチャートリガーをある発話応答に関連付けることにより、又はある音声コマンドを、上記アプリケーションプロトタイプのある異なるビジュアルページへの遷移に関連付けることにより、複数の音声対話をミックスして他のトリガー及びアクションとマッチングさせることができる。例えば、「天気はどうか (w h a t ' s t h e w e a t h e r) 」等の音声コマンドへの応答において、上記アプリケーションは、今後の天気の概要を提示するビジュアルページに遷移できる。

【0016】

続いて設計アプリケーションは、トリガー及びアクションの対話データを、対応するビジュアルページに関連付ける。これにより、上記アプリケーションプロトタイプの試験段階中に、各ビジュアルページに関連する対話データを、トリガーの識別に基づいて取得できる。場合によっては、対話データ及び関連するビジュアルページは、クライアントデバイスにローカルに保存してよい。例えば音声対話ツールは、対話データ及び関連するビジ

10

20

30

40

50

ジュアルページを含む、上記アプリケーションプロトタイプのためのファイルを生成して、上記ファイルをクライアントデバイスのストレージに保存させてよい。このようにすると、上記ファイルを他の設計者に容易に転送でき、上記他の設計者が設計を継続できる。あるいは、又は更に、対話データ及び関連するビジュアルページを、リモートストレージに保持してもよい。

【0017】

設計アプリケーションは更に、上記設計アプリケーションを用いて生成された上記音声支援型アプリケーションプロトタイプの試験を可能とする、試験モジュールと共に実装される。試験段階の開始時、試験モジュールは、上記アプリケーションプロトタイプの第1のビジュアルページ（例えばホームページ）を表示する。ユーザが、音声コマンドを発話すること又はビジュアルページをタップすること等によって上記アプリケーションプロトタイプと対話すると、試験モジュールはリアルタイムで、発話応答を出力すること、異なるビジュアルページに遷移すること、メディアコンテンツのプレイバックを開始すること、又はこれらの組み合わせ等によって、このトリガーに関連する、対応するアクションを実施させる。このようにして、本明細書に記載のシステムにより、ユーザは試験段階中に上記アプリケーションプロトタイプと対話して、上記アプリケーションプロトタイプが意図した通りに機能することを保証できる。これにより、設計者は、開発段階の前に、音声支援型対話の設計によるいずれの問題を識別して改善できる。上記アプリケーションプロトタイプの設計及び試験後、設計アプリケーションは、上記アプリケーションプロトタイプの設計に従って上記アプリケーションを開発するためにアプリケーション開発者が利用できる指示及びデータを含む、設計仕様を生成できる。あるいは、設計アプリケーションは、設計及び試験プロセスの完了時に実行可能なアプリケーションを生成するように実装してもよい。

【0018】

よって、本明細書に記載の技法は、音声コマンド及び発話プレイバックの両方をアプリケーションプロトタイプの設計に直接、簡単に統合できるようにすることにより、音声支援型アプリケーションプロトタイプの設計を簡略化し、その一方で、試験段階は、ユーザが音声コマンドを提供できるようにし、また発話プレイバックを出力して、ユーザが上記アプリケーションプロトタイプとの対話をシミュレーションできるようにする。

【0019】

特に、音声コマンドのマッチング時、設計アプリケーションの試験モジュールは、上記音声コマンドのテキスト文字列が、保存済み音声コマンドのテキスト文字列と共通して有する文字数に基づいて、マッチングする音声コマンドを選択してよい。これは、英語等の表音文字体系を使用する言語に関しては、共通の文字をマッチングさせることによって、単語の音がどの程度近いかを近似できるため、良好に機能する。例えば英語の単語「house」及び「mouse」はいずれも、5文字のうちの4文字が共通であるため、これらは音も類似している。従って、音声テキスト化サービスが音声コマンドを、意図した音声コマンドと「音が類似した」テキスト文字列（例えば「mouse」ではなく「house」、又は「ride share」ではなく「red chair」）に誤って変換するという一般的なシナリオにおいて、本明細書に記載のマッチング技法は依然として、表音文字体系において同一の音を有する単語間の文字の共通性により、マッチングする保存済み音声コマンドを高い蓋然性で識別できる。しかしながら、北京官話及び日本語の漢字等の「非表音(non phonetic)」文字体系を使用する言語に関しては、同一の音を有する単語が文字毎に容易に比較されないため、このマッチング技法は機能しない。

【0020】

この問題を解決するために、設計アプリケーションは、非表音文字体系を使用する言語の音声コマンドのテキスト文字列を、上記テキスト文字列の表音バージョンに変換するために、表音言語翻訳機を利用する。一例として、表音言語翻訳機は、北京官話、並びにひらがな、カタカナ、及び漢字の混合を、それぞれピンイン及びローマ字に翻訳してよい。試験段階では、マイクロフォンが非表音言語の音声コマンドのオーディオをキャプチャし

10

20

30

40

50

、この非表音言語の音声コマンドのオーディオを、発話テキスト化サービスに提供する。発話テキスト化サービスは、音声コマンドのオーディオを、非表音文字体系の音声コマンドのテキスト文字列に変換する。続いてこの非表音文字体系の音声コマンドのテキスト文字列を、表音言語翻訳機に提供し、上記表音言語翻訳機は、上記非表音文字体系の音声コマンドのテキスト文字列を、音声コマンドのテキスト文字列の表音バージョンに変換する。試験と同時に、又は試験前に、表音言語翻訳機は、上記アプリケーションプロトタイプに関連する保存済み音声コマンドの、非表音文字体系のテキスト文字列を、上記保存済み音声コマンドのテキスト文字列の音声バージョンに翻訳する。

【0021】

次に、キャプチャされた音声コマンドのテキスト文字列の、翻訳された表音バージョンを、保存済み音声コマンドのテキスト文字列の翻訳された表音バージョンと比較する。キャプチャされた音声コマンドのテキスト文字列の表音バージョンと、保存済み音声コマンドのテキスト文字列の表音バージョンとの間の、共通する文字の数の決定に基づいて、マッチングする音声コマンドを識別する。非表音テキスト文字列を最初に音声バージョンに変換することにより、共通する文字の数に基づくこのような比較は、マッチングする音声コマンドを合理的な信頼性レベル内で識別するにあたって有効となる。よって、本明細書に記載の表音言語翻訳技法は、音声支援型アプリケーションプロトタイプの試験中の音声コマンドマッチングの精度を向上させる。

【0022】

本明細書に記載されているように、用語「非表音テキスト文字列(non phonetic text string)」とは、北京官話又は漢字といった非表音文字体系で書かれたテキスト文字列を指し、非表音音声コマンドとは、ユーザが非表音文字体系を使用する言語で音声コマンドを発話したときにマイクロフォンがキャプチャしたオーディオデータを指す。対照的に、「表音テキスト文字列(phonetic text string)」「表音テキスト文字列」又は「翻訳された表音テキスト文字列(translated phonetic text string)」は、ピンイン又はローマ字といった表音的表現に翻訳された、非表音文字体系のテキスト文字列を指す。

【0023】

以下の議論では、まず、本明細書に記載の技法を使用できる例示的環境について説明する。続いて、この例示的環境及び他の環境において実施できる、例示的な実装の詳細及び手順について説明する。従って、上記例示の手順の実施は、上記例示的環境に限定されず、また上記例示的環境は、上記例示の手順の実施に限定されない。

【0024】

例示的環境

図1は、本明細書に記載の、非表音文字体系を使用する言語のための音声支援型アプリケーションプロトタイプの試験中に、音声コマンドマッチングを使用するために動作可能な、例示的な実装形態における環境100の図である。図示されている環境100は、計算デバイス102、音声対話設計システム104、音声アシスタントプラットフォーム106、及びネットワーク110を介して互いに通信可能に連結される様々な音声アシスタントデバイス108を含む。

【0025】

計算デバイス102、音声対話設計システム104、及び音声アシスタントプラットフォーム106の実装に使用できるデバイスは、多様な方法で構成できる。例えばこれらのデバイスは、デスクトップコンピュータ、ラップトップコンピュータ、(例えばタブレット又は携帯電話等のハンドヘルド構成を想定した)移動体デバイス等として構成してよい。よって、デバイスは、相当なメモリ及びプロセッサリソースを有するフルリソースデバイス(例えばパーソナルコンピュータ、ゲームコンソール)から、限定的なメモリ及び/又は処理リソースを有する低リソースデバイス(例えば移動体デバイス)にまで及んでよい。更に、「デバイス(device)」は、図8に関連して更に説明されるように、「クラウドを介した(over the cloud)」動作を実施するためにビジネス

10

20

30

40

50

で利用される複数のサーバー等の、複数の異なるデバイスを表す場合がある。

【0026】

音声アシスタントデバイス108もまた、多様な方法で構成できる。一般に、音声アシスタントデバイス108は、音声コマンド112をユーザ114から受信するための1つ以上のマイクロフォンを用いて構成される。専用の音声アシスタントデバイス108の例としては、Amazon（登録商標）Echo及びGoogle（登録商標）Homeが挙げられ、これらはそれぞれ、音声アシスタントプラットフォームAmazon（登録商標）Alexa及びGoogle（登録商標）Assistantとの対話を可能とする。携帯電話、デスクトップコンピュータ、ラップトップコンピュータ、ゲーミングシステム等といった他のデバイスも、音声アシスタントプラットフォームの機能を活用できる音声アシスタントデバイス108として構成できる。例えば、Apple（登録商標）iPhone（携帯電話）は、ユーザ114が音声アシスタントプラットフォームSiri（登録商標）と対話できるようにする音声アシスタントデバイス108として機能できる。この目的のために、音声アシスタントデバイス108は、音声コマンド112の受信に加えて、インタフェース（例えばタッチ、カメラ等）を通してユーザ入力を受信するよう構成できる。

10

【0027】

いずれの場合においても、これらの音声アシスタントデバイス108は、それぞれの音声アシスタントプラットフォーム106と対話して、音声コマンド112に回答する。音声アシスタントプラットフォーム106は、音声コマンド処理モジュール116を有するものとして図示されている。音声コマンド処理モジュール116は、音声コマンド112によって開始されるリクエストを受信し、上記リクエストにどのように回答するかを決定し、音声アシスタントデバイス108、又は接続されたホーム、（例えばeコマースプラットフォームによる商品の配送を容易にする）1つ以上のサービスそれぞれを提供するサービスプロバイダデバイス等といった、音声コマンド112が影響を及ぼすことができる他のいずれのデバイスに、回答を提供する機能を示す。従って、音声アシスタントデバイス108は、1つ以上のスピーカーを介して可聴発話応答115を出力すること、視覚コンテンツを表示すること、他の視覚効果（例えば発光ダイオード（LED）による指示）を提供すること、触覚フィードバックを提供すること等によって、無数の方法でユーザ114の音声コマンドに回答するよう構成してよい。更に、音声アシスタントプラットフォーム106は、例えばサーモスタット、スピーカー、照明といった他のデバイスに、ユーザ114の音声コマンド112に対して回答させることができる。家電製品に関する回答のいくつかの例としては：家屋の温度を調整するためにサーモスタットの温度を変更すること；音楽サーバーからスピーカーでの音楽の再生を開始すること；家屋内の特定の照明をオン又はオフにすること等が挙げられる。音声コマンド112は、音声アシスタントデバイス108を含むがこれに限定されない1つ以上のデバイスによる回答を開始するために、多様な設定（例えば住宅、職場等）で活用され得る。

20

30

【0028】

図示されている環境100の文脈において、設計アプリケーション118は、計算デバイス102のユーザ（例えば設計者）がソフトウェアアプリケーション、ファームウェア、サービス、プラグイン、スクリプト等といったアプリケーションに関するアプリケーションプロトタイプ120を設計できるようにする機能を示す。場合によっては、アプリケーションプロトタイプ120は、統合型ディスプレイスクリーンを含むAmazon（登録商標）Show等の音声アシスタントプラットフォーム106のために設計される。しかしながら、アプリケーションプロトタイプ120は、音声支援型機能を含むいずれのタイプのアプリケーションのために設計してよい。設計アプリケーション118は、Adobe（登録商標）XD又はAdobe（登録商標）Sketchといった設計アプリケーションとして実装してよい。しかしながら、設計アプリケーション118は、多様な異なるサードパーティコンテンツ生成又は編集アプリケーション、音声アプリケーション開発アプリケーション、開発環境等として実装してよい。

40

50

【 0 0 2 9 】

設計アプリケーション 1 1 8 が、ユーザがアプリケーションプロトタイプ 1 2 0 のビジュアルページ 1 2 2 のシーケンスを設計できるようにする、様々なインタフェース及びコントロールを提供する。ビジュアルページ 1 2 2 の設計の一部として、設計アプリケーション 1 1 8 は、アプリケーションプロトタイプ 1 2 0 のための様々なユーザインタフェースコントロール（例えば選択可能なボタン及びドロップダウンリスト）、グラフィック、テキスト等を設計及びレイアウトするためのツールを提供する。例えばユーザは、アプリケーションのホームページ、及びエンドユーザをナビゲートできる他のビジュアルページ 1 2 2 を設計してよい。

【 0 0 3 0 】

本明細書に記載の技法によると、設計アプリケーション 1 1 8 は音声対話ツール 1 2 6 を更に含み、これは、音声コマンド及び発話応答の形式の対話データ 1 2 4 を、アプリケーションプロトタイプ 1 2 0 の異なる複数のビジュアルページ 1 2 2 と関連付けるのを容易にする。これを実現するために、音声対話ツール 1 2 6 は、設計アプリケーション 1 1 8 の設計インタフェースを制御して、アプリケーションプロトタイプ 1 2 0 のビジュアルページ 1 2 2 のためのトリガー及び関連するアクションを受信するための対話インタフェースを提供する。本明細書に記載されているように、トリガーは、対応するアクションを誘発できる様々なユーザコマンドを含む。トリガーの例のごく一部としては、音声コマンド、ユーザジェスチャー（例えばタップ、ドラッグ、スワイプ、キーストローク）、及び時間遅延が挙げられる。アクションの例のごく一部としては、発話応答、アプリケーションプロトタイプの他のビジュアルページへのページ遷移、及びメディアファイル（例えば音声ファイル）のプレイバックが挙げられる。

【 0 0 3 1 】

場合によっては、対話インタフェースにより、ユーザは、定義されたトリガー及びアクションのリストから、トリガー及び関連するアクションを選択できる。選択後、ユーザは、選択したトリガー及びアクションを記述する入力を提供する。例えば、ある音声コマンドを上記アプリケーションプロトタイプ 1 2 0 のある特定のビジュアルページ 1 2 2 に関連付けるために、音声対話ツール 1 2 6 によって、ユーザは、（例えばタッチ機能を用いて）上記特定のビジュアルページ 1 2 2 を選択し、対話インタフェースを介して（例えばコマンド又は応答をタイピングすること、コマンド又は応答をマイクロフォン又は通信可能に連結されたデバイスに対して発話すること等によって）、上記特定のビジュアルページに対応する異なる複数の音声コマンド又は発話応答を指定できる。本明細書で使用される場合、対話データ 1 2 4 は、音声コマンド及び発話応答、並びに他のタイプのトリガー（例えばユーザジェスチャー及び時間遅延）並びにアクション（例えばページ遷移及びメディアのプレイバック）を含む。

【 0 0 3 2 】

特に、設計アプリケーション 1 1 8 により、ユーザは、様々な異なるタイプのトリガー及びアクションを、アプリケーションプロトタイプ 1 2 0 のビジュアルページ 1 2 2 に関連付けることができる。例えばユーザは、音声コマンドトリガーを対応する発話応答に関連付けることができる。しかしながら、設計アプリケーション 1 1 8 により、ユーザは、例えばあるユーザジェスチャートリガーをある発話応答に関連付けることにより、又はある音声コマンドを、アプリケーションプロトタイプ 1 2 0 のある異なるビジュアルページ 1 2 2 への遷移に関連付けることにより、複数の音声対話をミックスして他のトリガー及びアクションとマッチングさせることができる。

【 0 0 3 3 】

アプリケーションプロトタイプ 1 2 0 のための音声支援型対話の設計を容易にするの一部として、音声対話ツール 1 2 6 は、音声対話設計システム 1 0 4 の機能を活用してよい。図示されている環境 1 0 0 では、音声対話設計システム 1 0 4 が、発話テキスト化サービス 1 2 8 及びテキスト発話化サービス 1 3 0 と共に図示されている。一般に、発話テキスト化サービス 1 2 8 は、1 つ以上のマイクロフォンがキャプチャした音声データ中

10

20

30

40

50

の、発話された単語及びフレーズを識別し、上記発話された単語及びフレーズをテキスト文字列に変換する機能を表す。テキスト発話化サービス130は、テキストデータ（例えば設計アプリケーションのユーザインタフェースへのユーザ入力を介して受信したテキストデータ）を発話応答に加工する機能を表す。よって、発話テキスト化サービス128及びテキスト発話化サービス130により、ユーザは、アプリケーションプロトタイプ120の設計及び試験中に、設計アプリケーション118への可聴入力及びテキスト入力の両方を提供できる。音声対話設計システム104は、これらの特定の部品と共に図示されているものの、本発明の精神又は範囲から逸脱せずに本明細書に記載の機能を提供するために、より多数の、より少数の、及び異なる部品を含んでよく、又はこれらに対する（例えばクラウドコンピューティングサービスを介した）アクセスを有してよいことを理解されたい。更に場合によっては、発話テキスト化サービス128若しくはテキスト発話化サービス130を、計算デバイス102において実装してよく、又は別個のサードパーティ製サーバーにおいて実装してよい。発話テキスト化サービス128及びテキスト発話化サービス130をサードパーティ製サービスとして実装する場合、音声対話ツール126は、発話テキスト化サービス128及びテキスト発話化サービス130への安全な通信チャネルを確立するために、認証トークンを利用できる。

10

【0034】

設計アプリケーション118は更に、設計アプリケーション118を用いて生成された音声支援型アプリケーションプロトタイプ120の試験を可能とする、試験モジュール132と共に実装される。これを実現するために、試験モジュール132は、試験段階において、アプリケーションプロトタイプ120の第1のビジュアルページ（例えばホームページ）を表示する。ユーザが、音声コマンドを発話すること又はビジュアルページ122をタップすること等によってアプリケーションプロトタイプ120と対話すると、試験モジュール132はリアルタイムで、発話応答を出力すること、異なるビジュアルページに遷移すること、メディアのプレイバックを開始すること、又はこれらの組み合わせ等によって、このトリガーに関連する、対応するアクションを実施させる。このようにして、本明細書に記載のシステムにより、ユーザは試験段階中にアプリケーションプロトタイプ120と対話して、上記アプリケーションプロトタイプが意図した通りに機能することを保証できる。これにより、ユーザは、開発段階の前に、音声支援型対話の設計によるいずれの問題を識別して改善できる。

20

30

【0035】

音声対話設計システム104は更に、計算デバイス102において、又は計算デバイス102から離れて（例えばクラウドサービスとして）実装できる、表音言語翻訳機134と共に実装される。一般に、表音言語翻訳機134は、非表音文字体系のテキスト文字列を取得し、この非表音テキスト文字列を、異なる言語又は文字の表音テキスト文字列に変換するよう構成される。例えば表音言語翻訳機は、非表音文字体系のテキスト文字列をローマ文字又はラテン文字に変換するための、ローマ文字化技法を利用してよい。これを実施するために、表音言語翻訳機は、書かれたテキストを表すための翻字技法、及び発話された単語を表すための転写技法を利用してよい。転写方法は、発話中の音素又は意味の単位を記録する音素転写と、発話音声を正確に記録する、より厳密な音声転写とに細分化できる。

40

【0036】

表音言語翻訳機134を利用して、多様な異なるタイプの非表音言語のテキスト文字列を、上記テキスト文字列の表音バージョンに変換してよい。一例として、表音言語翻訳機は、中国文字で書かれることが多い北京官話を、「ピンイン」として知られる中国語のローマ文字化バージョンに翻訳するよう構成できる。別の例では、表音言語翻訳機は、中国語から借用した表音文字（漢字）と、これもまた究極的には中国文字に由来するものである音節を表す文字（かな）との組み合わせで通常書かれる日本語を、ローマ字に翻訳するよう構成できる。日本語に関しては、（最も広く使用されている）ヘボン式ローマ字変換、訓令式ローマ字変換、及び日本式ローマ字変換を含む、複数の異なるタイプのローマ文

50

字化体系が存在する。特に、音声対話設計システム 104 は、多様な異なるタイプのローマ文字化技法を利用して、多様な異なる言語から受け取ったテキストを、上記テキストの標準的な表音バージョンに変換してよい。

【0037】

アプリケーションプロトタイプ 120 の設計及び試験後、設計アプリケーション 118 は、アプリケーションを開発するためにアプリケーション開発者に提供してよい、設計仕様を生成できる。あるいは、設計アプリケーション 118 は、設計及び試験プロセスの完了時に実行可能なアプリケーションを生成するように実装してもよい。アプリケーションプロトタイプ作成のための音声対話を可能とすることによって、本明細書に記載の技法は、設計者が、ビジュアルインタフェースを含む音声支援型アプリケーションを設計できるようにする。

10

【0038】

ある例示的環境について考察してきたが、これより、1つ以上の実装形態による、デジタル媒体環境における音声支援型アプリケーションプロトタイプのための音声対話ツールに関する技法のいくつかの例示的詳細に関する議論を考察する。

【0039】

図 2 は、図 1 の音声対話ツールが、対話データとアプリケーションプロトタイプのビジュアルページとの間の関連を生成し、上記対話データ及び上記ビジュアルページを保持する、例示的実装形態 200 を示す。図示されている例 200 は、図 1 からの設計アプリケーション 118 の音声対話ツール 126 を含む。

20

【0040】

例 200 では、ユーザ入力 202 を取得している、音声対話ツール 126 が図示されている。ユーザ入力 202 は、図 4、5 に関連して説明したものと同様のユーザインタフェース等の、設計アプリケーション 118 のユーザインタフェースを介して、音声対話ツール 126 によって受信できる。場合によっては、ユーザ入力 202 は、1つ以上のマイクロフォンによってキャプチャされた音声入力に対応してよい。大まかに言えば、ユーザ入力は、アプリケーションプロトタイプ 120 のあるビジュアルページ 122 に関連するトリガー及び対応するアクションを識別する。本明細書全体にわたって記載されているように、トリガーは、対応するアクションを誘発できる様々なユーザコマンドを含む。

【0041】

ユーザ入力 202 に基づき、音声対話ツール 126 は関連対話データ 204 を生成し、これは、対話データ 124 (例えばトリガー及びアクション)と各ビジュアルページ 122 との間の関連を記述する。この関連対話データ 204 を、設計アプリケーション 118 によって保持でき、これにより、各ビジュアルページ 122 に関連する対話データ 124 を、トリガーの検出に基づいてストレージ 206 から取得できる。

30

【0042】

ストレージ 206 は、計算デバイス 102 にローカルに実装してよい。例えば音声対話ツールは、対話データ 124 及び関連するビジュアルページ 122 を含む、アプリケーションプロトタイプ 120 のためのファイルを生成してよく、上記ファイルを、計算デバイス 102 のストレージ 206 に保存させてよい。これにより、アプリケーションプロトタイプ 120 の試験段階中に、各ビジュアルページ 122 に関連する対話データ 124 を、トリガーの識別に基づいて上記ストレージから取得できる。あるいは、又は更に、対話データ 124 及び関連するビジュアルページ 122 は、音声対話設計システム 104 のリモートストレージに保持してよい。1つ以上の実装形態では、クライアントデバイスにローカルに保存されたファイルを、音声対話設計システムのストレージにリモートで保存された関連対話データと同期させてよい。

40

【0043】

対話データ 124 は、トリガー 208 及び対応するアクション 210 を含む。この例では、トリガー 208 は、音声コマンド 212、ユーザジェスチャー 214 (例えばタップ、ドラッグ、スワイプ又はキーストローク)、及び時間遅延 216 を含み、アクションは

50

、発話応答 2 1 8、ページ遷移 2 2 0、及びメディア出力 2 2 2 を含む。場合によっては、設計アプリケーション 1 1 8 は対話インタフェースを提供し、これにより、ユーザは、定義されたトリガー及びアクションのリストから、トリガー及び関連するアクションを選択できる。選択後、ユーザは、選択したトリガー及びアクションを記述する入力を提供する。例えば、ある音声コマンドを上記アプリケーションプロトタイプ 1 2 0 のある特定のビジュアルページ 1 2 2 に関連付けるために、音声対話ツール 1 2 6 によって、ユーザは、（例えばタッチ機能を用いて）上記特定のビジュアルページ 1 2 2 を選択し、対話インタフェースを介して（例えばコマンド又は応答をタイピングすること、コマンド又は応答をマイクロフォン又は通信可能に連結されたデバイスに対して発話すること等によって）、上記特定の要素に対応する異なる複数の音声コマンド 2 1 2 又は発話応答 2 1 8 を指定
10
できる。上記音声コマンドは、音声アシスタントデバイスに対するユーザの発話を示すオーディオデータ又はテキストデータとして構成してよい。発話応答は、音声コマンド、例えば可聴発話の受信に応答した可聴出力を表す。音声コマンド 2 1 2 を、マイクロフォンを介したユーザからの音声入力として受信した場合、発話テキスト化サービス 1 2 8 を利用して、キャプチャしたオーディオをテキスト文字列へと変換できる。1 つ以上の実装形態では、音声対話ツール 1 2 6 は、音声コマンド 2 1 2 の複数のバリエーションを生成できる。例えば音声コマンド 2 1 2 を、「明日の天気はどうか（What is the weather tomorrow）」として受信した場合、音声対話ツール 1 2 6 は、「明日の天気はどうか（what is the weather going to be tomorrow）」、「明日は雨が降りそうか？（is it going to rain tomorrow?）」、「明日は晴れそうか（is it going to be sunny tomorrow）」、「明日はどの程度暑い（how hot will it be tomorrow）」等といった、上記音声コマンドの複数のバリエーションを生成できる。
20

【 0 0 4 4 】

1 つ以上の実装形態では、音声対話ツール 1 2 6 は音声コマンド 2 1 2 を、北京官話又は漢字等の非表音文字体系でのテキスト入力として受信してよい。このシナリオでは、音声対話ツール 1 2 6 は、アプリケーションプロトタイプ試験段階まで、音声コマンド 2 1 2 を非表音テキスト文字列としてストレージ 2 0 6 内に保存してよい。あるいは場合によっては、音声コマンド 2 0 6 の非表音テキスト文字列は、上記テキスト文字列をストレージ 2 0 6 に保存する前に、表音言語翻訳機 1 3 4 によって、上記テキスト文字列の表音バージョンへと変換できる。このシナリオでは、テキスト文字列の表音バージョンをストレージ 2 0 6 に保存する。
30

【 0 0 4 5 】

本明細書に記載の技法により、ユーザは、試験段階中に、アプリケーションプロトタイプ 1 2 0 と対話できるようにもなる。この文脈において、図 1 の設計アプリケーションの試験モジュールがアプリケーションプロトタイプの試験を容易にする例示的実装形態 3 0 0 を示す、図 3 について考察する。図示されている例 3 0 0 は、図 1 からの試験モジュール 1 3 2、発話テキスト化サービス 1 2 8、及びテキスト発話化サービス 1 3 0 を含む。
40

【 0 0 4 6 】

試験段階において、試験モジュール 1 3 2 は、アプリケーションプロトタイプ 1 2 0 の第 1 のビジュアルページ 1 2 2（例えばホームページ）の表示を開始する。第 1 のビジュアルページ 1 2 2 の表示中、ユーザは、音声コマンドを発話すること又はビジュアルページ 1 2 2 をタップすること等によってアプリケーションプロトタイプ 1 2 0 と対話し、試験モジュール 1 3 2 は、ユーザの対話に基づいて、トリガー 3 0 2 をリアルタイムで取得する。試験モジュール 1 3 2 は処理モジュール 3 0 4 を含むものとして図示されており、この処理モジュール 3 0 4 は、ユーザの対話を処理することによって、音声コマンド 3 0 6、ユーザジェスチャー 3 0 8 又は時間遅延 3 1 0 といったトリガー 3 0 2 を識別するための機能を表す。
50

【 0 0 4 7 】

処理モジュール304は、識別されたトリガー302を関連対話データ204と比較することにより、マッチングするトリガーを識別する。例えば、リクエスト処理モジュール304は、音声コマンド306又はユーザジェスチャー308を、それぞれ、関連対話データ204の、保存済みの音声コマンド212及びユーザジェスチャー214と比較する。マッチングするトリガーを識別するために、処理モジュール304は、一意識別子、ハッシュ、テキストベース検索等の多様な技法を活用してよい。更に、処理モジュール304は、新しいビジュアルページが表示される際に、時間遅延トリガーを継続的に監視してよい。

【0048】

トリガー302を音声コマンド306として受信した場合、処理モジュール304は、音声コマンド306のオーディオデータを発話テキスト化サービス128に提供する。発話テキスト化サービス128は上記オーディオデータをリアルタイムで処理して、音声コマンド306の上記オーディオデータをテキスト文字列に変換する。処理モジュール304は、音声コマンド306のテキスト文字列を発話テキスト化サービス128から取得し、この音声コマンド306のテキスト文字列を、音声コマンド212の保存済みのテキスト文字列と比較することによって、マッチングする音声コマンドを識別する。

10

【0049】

処理モジュール304は、受信した音声コマンド306のテキスト文字列の文字を、保存済みの音声コマンド212のテキスト文字列の文字と比較することによって、マッチングする音声コマンドを識別してよい。一致する文字の数が多いほど、マッチングの蓋然性が高いことを示す。例えばコマンドが単語「mouse」を含む場合、処理モジュール304は、単語「house」を、「mouse」とマッチングするものとして認識し得る。というのは、これらの単語は4文字が一致しているためである。場合によっては、発話テキスト化サービス128は、テキスト文字列が正しいことの信頼性のレベルを示す信頼性スコアを、テキスト文字列と共に含む。この信頼性スコアは、処理モジュール304がマッチングする音声コマンドを識別するのを支援する。1つ以上の実装形態では、処理モジュール304は、ユーザがアプリケーションプロトタイプと更に簡単に対話できるようにするために、マッチングするテキスト文字列を、低い許容誤差を用いて選択するよう構成される。

20

【0050】

トリガー302が、関連対話データ204の保存済みのトリガー208とマッチングすることが検出されたことに応答して、処理モジュール304は、関連対話データ204中の、上記トリガーに関連する、対応するアクション210を識別する。次に、発話応答314、ページ遷移316、又はメディア出力318に対応するアクション312を、実施モジュール320へと移動させる。実施モジュール320は、発話応答314を出力すること、異なるビジュアルページへのページ遷移316を実施すること、オーディオ若しくはビデオコンテンツを出力すること、又はこれらの組み合わせ等によって、識別されたトリガーに関連する、対応するアクション312を実施させる。

30

【0051】

アクション312が発話応答314に対応する場合、実施モジュールは、発話応答314のテキスト文字列を、テキスト発話化サービス130にリアルタイムで提供できる。テキスト発話化サービス130は、発話応答314のテキスト文字列を、発話応答314に対応するオーディオデータに変換する。実施モジュール320は次に、発話応答314に対応する上記オーディオデータを出力させる。

40

【0052】

このようにして、本明細書に記載のシステムにより、ユーザは試験段階中にアプリケーションプロトタイプ120と対話して、アプリケーションプロトタイプ120が意図した通りに機能することを保証できる。これにより、ユーザは、開発段階の前に、音声支援型アプリケーションプロトタイプ120の設計によるいずれの問題を識別して改善できる。

【0053】

50

特に、音声コマンドのマッチング時、試験モジュール132は、上記音声コマンドのテキスト文字列が、保存済み音声コマンドのテキスト文字列と共通して有する文字数に基づいて、マッチングする音声コマンドを選択してよい。これは、英語等の表音文字体系を使用する言語に関しては、共通の文字をマッチングさせることによって、単語の音がどの程度近いかを近似できるため、良好に機能する。しかしながら、北京官話及び日本語の漢字等の「非表音」文字体系を使用する言語に関しては、同一の音を有する単語が文字毎に容易に比較されないため、このマッチング技法は機能しない。

【0054】

この文脈において、図1の設計アプリケーションの試験モジュールが表音言語翻訳機を利用して、アプリケーションプロトタイプ120の試験中にキャプチャされた音声コマンドをマッチングさせる、例示的実装形態400を示す、図4について考察する。図示されている例400は、図1からの試験モジュール132、発話テキスト化サービス128、及び表音言語翻訳機134を含む。

10

【0055】

試験段階において、試験モジュール132は、アプリケーションプロトタイプ120の第1のビジュアルページ122（例えばホームページ）の表示を開始する。第1のビジュアルページ122の表示中、ユーザは、非表音音声コマンド402を発話することによってアプリケーションプロトタイプ120と対話する。本明細書全体を通して記載されているように、非表音音声コマンド402は、北京官話又は日本語における漢字といった非表音文字体系を使用する言語で発話された音声コマンドに相当する。非表音音声コマンド402は、少なくとも1つのマイクロフォンによって、オーディオデータとしてキャプチャされる。

20

【0056】

試験モジュール132は、キャプチャした非表音音声コマンド402のオーディオデータを発話テキスト化サービス128に通信する。発話テキスト化サービス128は、非表音音声コマンド402の上記オーディオデータをリアルタイムで処理して、非表音音声コマンド402の上記オーディオデータを、キャプチャした音声コマンドの、非表音文字体系の言語の非表音テキスト文字列404に変換する。

【0057】

試験モジュール132は、非表音テキスト文字列404を表音言語翻訳機134に通信する。表音言語翻訳機134は、キャプチャした音声コマンドの非表音テキスト文字列404を、上記キャプチャした音声コマンドの、ピンイン又はローマ字といった表音文字体系で書かれた表音テキスト文字列406に翻訳する。更に、試験段階中、表音言語翻訳機134は、保存済みの音声コマンド408の、非表音文字体系での非表音テキスト文字列410を取得する。例えば、保存済みの音声コマンド408は、ストレージ206に保持された音声コマンド212に対応してよい。表音言語翻訳機134は、保存済みの音声コマンド408の非表音テキスト文字列410を、保存済みの音声コマンド408の、表音文字体系での表音テキスト文字列412に翻訳する。あるいは、表音言語翻訳機134は、試験段階の前に、保存済みの音声コマンド408の表音テキスト文字列410を、保存済みの音声コマンド408の表音テキスト文字列412に翻訳してよい。この場合、表音テキスト文字列412は、ストレージに保持されていてよい。

30

40

【0058】

試験モジュール132の表音比較モジュール414は、キャプチャした音声コマンド402の表音テキスト文字列406と、保存済みの音声コマンド408の表音テキスト文字列412とを取得する。続いて表音比較モジュール414は、表音テキスト文字列406を表音テキスト文字列412と比較することにより、保存済みの音声コマンドに対応する表音テキスト文字列412の、マッチングする音声コマンド416を識別する。表音比較モジュール414は、マッチングする音声コマンド416を、表音テキスト文字列406の文字を保存済みの音声コマンド408の表音テキスト文字列412と比較することによって、識別してよい。表音テキスト文字列406、412は表音文字に翻訳されているた

50

め、表音比較モジュール414は、マッチングする音声コマンド416が、キャプチャした音声コマンドと共通して有する文字数に基づいて、マッチングする音声コマンド416を選択できる。いずれのテキスト文字列も表音文字で書かれているため、マッチングする文字数が多くなるほど、マッチングの蓋然性が高くなる。

【0059】

マッチングする音声コマンド416が表音比較モジュール414によって識別されると、試験モジュールは、本明細書全体を通して記載されているように、関連するアクションの実施を開始する。

【0060】

図5A～5Cは、音声支援型アプリケーションプロトタイプ118の設計に使用される設計アプリケーションの設計ユーザインタフェースの例500を示す。

10

【0061】

図5Aでは、設計アプリケーション118の設計インタフェース502は、表示デバイス504によって表示されている。設計インタフェース502は、第1のビジュアルページ506及び第2のビジュアルページ508を含むものとして図示されている。第1のビジュアルページ506及び第2のビジュアルページ508は、設計アプリケーション118を利用するユーザによって設計されていてもよく、又は異なる設計アプリケーションからインポートされていてもよい。特に、設計インタフェース502をユーザが利用して、いずれの個数の異なるビジュアルページを生成してよい。

【0062】

20

図示されている例500はまた、カーソル510も含み、これは、設計アプリケーション118によって提供された設計インタフェース502の様々なコントロール及び機能と対話するための入力をユーザが提供できるようにするための機能を表す。カーソル510が図示されているものの、1つ以上の実装形態では、表示されたカーソルが存在しない場合がある。場合によっては、設計インタフェース502の様々なコントロール及び機能について、他の方法で、例えばタッチ入力（又は他のジェスチャー入力）、キーボード入力、スタイラス入力、音声入力等によって、選択又は対話を行ってよい。

【0063】

ビジュアルページ506、508は、アプリケーションプロトタイプ120のビジュアルページ又はスクリーンを表し、これらは設計者によって「アートボード (art board)」と呼ばれることが多い。この例では、第1のビジュアルページ506は、アプリケーションプロトタイプ120の「ホームページ (home page)」に相当し、現在時刻、現在の温度、及び現在の天気を「晴れ (sunny)」として描画したグラフィックを含む。

30

【0064】

この例では、第2のビジュアルページ508は、アプリケーションプロトタイプ120のカレンダーページに相当し、ユーザが3つの近日中のイベントを有すること、及びユーザが「明日午前10時にカフェ・プレッセでジュディとブランチ (Brunch with Judy at Cafe' Presse, at 10 AM tomorrow)」に対応するアポイントメントを有することを示すテキストを含む。特に、第1のビジュアルページ506は、ユーザが音声コマンド「私の予定は (What's on my calendar)」を発話できることを示すテキストを含む。このテキストは、ユーザが音声コマンドの発話によってカレンダーページに遷移できることの視覚的な手がかりをユーザに提供する。

40

【0065】

この音声コマンドに基づくカレンダーページへの遷移を可能とするために、設計インタフェースにより、ユーザは、第1のビジュアルページ506を第2のビジュアルページ508にリンクさせることができる。この例ではユーザは、カーソル510を用いて第1のビジュアルページ506上でハンドル制御を選択し、カーソルを第2のビジュアルページ508へとドラッグすることによって、ビジュアルページ506、508をリンクさせる

50

。これを実施することで、設計インタフェースに、第1のビジュアルページ506と第2のビジュアルページ508との間のビジュアルリンク要素512を生成させる。

【0066】

第1のビジュアルページ506を第2のビジュアルページ508にリンクさせることに応答して、設計インタフェースは、ユーザが、図5Bに示されているこの遷移のためのトリガーを識別する入力を提供できるようにする。図5Bでは、設計アプリケーションは、アプリケーションプロトタイプのあるビジュアルページのためのトリガー及び関連するアクションを提供するためのコントロールを含む、対話インタフェース514を提示する。この例では、対話インタフェース514は、トリガーのタイプを定義する入力を受信するための、トリガーコントロール516を含む。場合によっては、対話インタフェース514のトリガーコントロール516によって、ユーザは、トリガーコントロール516付近の下向きの矢印を選択して、トリガーのリストを表示させること等によって、定義されたトリガーのリストから、トリガーのタイプを選択できる。本明細書全体を通して記載されているように、対話インタフェース514に表示された、事前に定義されたトリガーのリストは、一部の例として、音声コマンド、ユーザジェスチャー、又は時間遅延を含んでよい。この例では、ユーザは、トリガーとして音声コマンドを選択するための入力を提供した。

10

【0067】

音声コマンドをトリガーとして選択した後、対話インタフェース514を修正して、トリガーコンテンツコントロール518を提供できる。例えば図5Bでは、対話インタフェース514は、音声コマンドの単語を定義する入力を受信するよう構成されたコマンドプロンプトとして、トリガーコンテンツコントロール518を表示する。あるいは、トリガーを、ユーザジェスチャー又は時間遅延として選択した場合、それぞれユーザジェスチャーのタイプ又は時間遅延の期間を記述する入力を受信するよう、トリガーコンテンツコントロール518を修正してよい。

20

【0068】

この例では、ユーザは、音声コマンドを定義する入力を「私の予定は」として提供している。このユーザ入力は、トリガーコンテンツコントロール518へのテキスト入力として受信できる。換言すれば、ユーザは、音声コマンドの単語を、対話インタフェース514のトリガーコンテンツコントロール518に直接タイピングできる。あるいは、ユーザは音声コマンドの単語を、マイクロフォン又は通信可能に連結されたデバイスに対して発話してよく、上記マイクロフォン又はデバイスは、本明細書全体を通して記載されているように、発話された単語をテキスト入力に変換する。ユーザが音声コマンドを発話する場合、対話インタフェース514は、変換されたテキストを対話インタフェース514に表示してよく、これによりユーザは、テキストが正しく認識されていることを確認でき、また必要に応じて変更できる。

30

【0069】

対話インタフェース514は更に、アクションコントロール520及び関連するアクションコンテンツコントロール522を含むものとして図示されている。アクションコントロール520は、対応するトリガーの検出に応答して実施されることになるアクションのタイプを定義する入力を受信するよう構成される。場合によっては、対話インタフェース514のアクションコントロール520により、ユーザは、アクションコントロール520付近の下向きの矢印をクリックして、アクションのリストを提示させること等によって、定義されたアクションのリストから、アクションのタイプを選択できる。本明細書全体を通して記載されているように、アクションは、一部の例として、発話応答、ページ遷移、又はコンテンツの出力を含んでよい。この例では、ユーザは、対応する音声コマンドトリガーに関するアクションとして、ページ遷移を選択する入力を提供している。あるいは、第1のビジュアルページ506を第2のビジュアルページ508と関連付けるための入力に基づいて、ユーザに関してページ遷移アクションを自動的に選択してよい。

40

【0070】

50

ページ遷移をアクションとして選択した後、対話インタフェース 5 1 4 を修正して、アクションコンテンツコントロール 5 2 2 を提供してよい。例えば図 5 B では、対話インタフェース 5 1 4 は、ページ遷移の遷移先を定義する入力を受信するよう構成されたコマンドプロンプトとして、アクションコンテンツコントロール 5 2 2 を表示する。あるいは、アクションを発話応答又はコンテンツの出力として選択した場合、それぞれ発話応答の単語を記述する入力、又は出力されることになるコンテンツを定義する入力（例えばビデオファイルのファイル名）を受信するよう、アクションコンテンツコントロール 5 2 2 を修正してよい。

【 0 0 7 1 】

この例では、ユーザは、アプリケーションプロトタイプのカレンダーページに遷移するために、ページ遷移を定義する入力を提供している。ここでも場合によっては、第 1 のビジュアルページを第 2 のビジュアルページとリンクさせるための入力に基づいて、ページ遷移の場所をユーザに関して自動的に選択してよい。この例では、設計インタフェースにより、ユーザは、イージング、遷移エフェクト、又はページ遷移の期間といったページ遷移の特性を更に定義できる。

10

【 0 0 7 2 】

特に図 5 B では、対話インタフェース 5 1 4 への入力によって定義されたトリガー及び関連するアクションを、第 1 のビジュアルページ 5 0 6 及び第 2 のビジュアルページと関連付ける。設計インタフェースにより、ユーザは、トリガーに回答して、異なるビジュアルページに遷移することなくアクションが実施されるように、ある単一のビジュアルページに関連するトリガー及び対応するアクションを定義することもできる。

20

【 0 0 7 3 】

この文脈において、図 5 C について考察する。図 5 C では、設計アプリケーションは、アプリケーションプロトタイプのあるビジュアルページのためのトリガー及び関連するアクションを提供するためのコントロールを含む、対話インタフェース 5 1 4 を提示している。この例では、対話インタフェース 5 1 4 は、カレンダーページに相当する第 2 のビジュアルページ 5 0 8 へのユーザ入力に回答して、動的に提示される。この例では、ユーザは、音声コマンドを選択するのではなく、トリガーコントロール 5 1 6 を介して時間遅延トリガーを選択し、続いてトリガーコンテンツコントロール 5 1 8 を介して入力を提供することにより、時間遅延の期間を 0 秒となるように定義する。ユーザはまた、アクションコントロール 5 2 0 を介して、発話応答を選択するための入力も提供し、続いてアクションコンテンツコントロール 5 2 2 を介して、発話応答アクションの単語を定義する入力を提供する。発話応答を定義する上記入力は、アクションコンテンツコントロール 5 2 2 へのテキスト入力として受信できる。換言すれば、ユーザは、発話応答の単語を、対話インタフェース 5 1 4 のアクションコンテンツコントロール 5 2 2 に直接タイピングできる。あるいは、ユーザは発話応答の単語を、マイクロフォン又は通信可能に連結されたデバイスに対して発話してよく、上記マイクロフォン又はデバイスは、本明細書全体を通して記載されているように、発話された単語をテキスト入力に変換する。ユーザが発話応答を発話する場合、対話インタフェース 5 1 4 は、変換されたテキストを対話インタフェース 5 1 4 のアクションコンテンツコントロール 5 2 2 に表示してよく、これによりユーザは、テキストが正しく認識されていることを確認でき、また必要に応じて変更できる。

30

40

【 0 0 7 4 】

特に、対話インタフェース 5 1 4 によって、ユーザは、多様な異なるタイプのトリガー及びアクションを、アプリケーションプロトタイプのビジュアルページに関連付けることができる。例えば設計者は、音声コマンドトリガーを、対応する発話応答に関連付けることができる。しかしながら設計者は、例えばあるユーザジェスチャトリガーをある発話応答に関連付けることにより、又はある音声コマンドを、アプリケーションプロトタイプのある異なるビジュアルページへの遷移に関連付けることにより、複数の音声対話をミックスして他のトリガー及びアクションとマッチングさせることもできる。

【 0 0 7 5 】

50

図 6 A、6 B は、音声支援型アプリケーションプロトタイプの実験段階中の設計アプリケーションの設計ユーザインタフェースの例 6 0 0 を示す。ユーザが、図 5 A ~ 4 C に示すように、アプリケーションプロトタイプのビジュアルページ及びユーザ対話を設計すると、設計インタフェースの実験コントロールを選択することによって、実験段階を開始できる。例 6 0 0 では、設計アプリケーション 1 1 8 の試験モジュール 1 3 2 は、表示デバイス 5 0 4 上に、アプリケーションプロトタイプの第 1 のビジュアルページ 6 0 2 を表示する。特に、第 1 のビジュアルページ 6 0 2 は、図 5 A ~ 4 C に示されている第 1 のビジュアルページ 5 0 6 に相当する。実験段階において、試験モジュール 1 3 2 は、ビジュアルページ並びに関連するトリガー及びアクションのレイアウトを表示するのではなく、設計されたアプリケーションプロトタイプ 1 2 0 を、実行可能なアプリケーションのように機能するように制御する。

10

【 0 0 7 6 】

音声コマンドを発話すること又はビジュアルページをタップすること等によって、ユーザが実験段階のアプリケーションプロトタイプと対話すると、試験モジュール 1 3 2 は、可聴コンテンツを出力すること、視覚コンテンツを出力すること、異なるビジュアルページに遷移すること、又はこれらの組み合わせ等によって、上記トリガーに関連する、対応するアクションを、リアルタイムで実施させる。例えば図 6 A では、ユーザは、複数の単語「私の予定は？」を含む音声コマンド 6 0 4 を発話する。マイクロフォンは、発話された音声コマンドを検出し、発話テキスト化サービス 1 2 8 はこのオーディオデータを、複数の単語「私の予定は」を含むテキスト文字列に変換する。次にこのテキスト文字列を、アプリケーションプロトタイプの設計中にユーザによって提供された関連対話データと比較する。

20

【 0 0 7 7 】

この例では、試験モジュール 1 3 2 は、音声コマンド 5 0 4 のテキスト文字列を、図 5 B の対話インタフェース 5 1 4 のトリガーコンテンツコントロール 5 1 8 に提供された音声コマンドとマッチングするものとして認識する。マッチングの検出に基づいて、試験モジュール 1 3 2 は、ユーザが過去に図 5 B の対話インタフェース 5 1 4 のコントロール 5 2 0、5 2 2 に入力した対応するアクションを実施させる。この場合、上記対応するアクションは、カレンダーページ（例えば図 5 のビジュアルページ 5 0 8）へのページ遷移として検出される。

30

【 0 0 7 8 】

図 6 B は、試験モジュール 1 3 2 が、検出されたトリガーに関連するアクションに基づいて、第 1 のビジュアルページ 6 0 2 から第 2 のビジュアルページ 6 0 6 に遷移した後の、カレンダーページに相当する第 2 のビジュアルページ 6 0 6 の表示を示す。更に、カレンダーページは、対応する発話応答を伴う 0 秒の時間遅延トリガーを含むため、試験モジュール 1 3 2 アプリケーションは、スピーカー又は関連するデバイスに、この例では複数の単語「明日午前 1 0 時にカフェ・プレッセでジュディとブランチ (Brunch with Judy at Cafe' Presse, at 10 AM tomorrow)」に対応する発話応答 6 0 8 を、可聴形式で出力させる。これを実現するために、発話応答の複数の単語を含むテキスト文字列を、テキスト発話化サービス 1 3 0 に移動させる。次に、テキスト発話化サービス 1 3 0 は、発話応答のテキスト文字列をオーディオファイルに変換し、そしてこのオーディオファイルは計算デバイスにおいてリアルタイムで出力される。このようにして、本明細書に記載のシステムにより、ユーザは、実験段階中にアプリケーションプロトタイプと対話して、上記アプリケーションプロトタイプが意図した通りに機能することを保証できる。

40

【 0 0 7 9 】

音声対話ツールに関する技法の例示的詳細について議論してきたが、これより、上記技法の更なる態様を説明するいくつかの例示的手順を考察する。

【 0 0 8 0 】

例示的手順

50

この節は、1つ以上の実装形態における、音声支援型アプリケーションプロトタイプ的设计及び試験のための例示の手順を説明する。上記手順の様子は、ハードウェア、ファームウェア、若しくはソフトウェア、又はこれらの組み合わせで実装できる。上記手順は、1つ以上のデバイスによって実行される動作を明記した複数のブロックのセットとして図示され、上記動作の実行に関して各ブロックによって示されている順序に必ずしも限定されない。

【0081】

図7は、音声支援型アプリケーションプロトタイプを設計する例示の手順700を示す。

【0082】

設計インタフェースにおいて、アプリケーションプロトタイプの少なくとも1つのビジュアルページを、計算デバイスのディスプレイ上に表示する(ブロック702)。例えば設計アプリケーション118は、アプリケーションプロトタイプ120のビジュアルページ122を、計算デバイス102のディスプレイ上の設計インタフェース502に表示する。

10

【0083】

設計インタフェースを制御して、アプリケーションプロトタイプの少なくとも1つのビジュアルページのためのトリガー及び関連するアクションを受信するための、対話インタフェースを提供する(ブロック704)。本明細書に記載の原理によると、トリガーは、音声コマンド、ユーザジェスチャー、又は時間遅延のうちの1つに対応し、アクションは、発話応答、アプリケーションプロトタイプの追加のビジュアルページへの遷移、又はメディアファイルのプレイバックに対応する。例えば音声対話ツール126は、設計インタフェース502を制御して、アプリケーションプロトタイプ120の少なくとも1つのビジュアルページ122のためのトリガー208及び関連するアクション210を受信するための、対話インタフェース514を提供する。

20

【0084】

対話インタフェースによってユーザ入力を受信して、トリガー及びアクションを提供する(ブロック706)。例えば音声対話ツール126は、設計インタフェース502を介してユーザ入力202を受信して、トリガー208及びアクション210を提供する。

【0085】

トリガー、アクション、及びアプリケーションプロトタイプのビジュアルページを含む、関連対話データを生成し、試験段階中にアプリケーションプロトタイプの試験を実施できるように、上記関連対話データを保存する(ブロック708)。例えば音声対話ツール126は、トリガー208及びアクション210を、ビジュアルページ122の指示と共に、関連対話データ204としてストレージ206に保存し、これにより、試験段階中のトリガーの検出によってアクションを実施させる。

30

【0086】

図8は、音声支援型アプリケーションプロトタイプを試験する例示の手順800を示す。

【0087】

アプリケーションプロトタイプに関して、試験段階を開始し(ブロック802)、上記アプリケーションプロトタイプの少なくとも1つのビジュアルページを試験段階において表示する(ブロック804)。例えば試験モジュール132は、アプリケーションプロトタイプ120のための試験段階を開始させ、試験段階中に、アプリケーションプロトタイプ120のための少なくとも1つのビジュアルページ122を表示する。

40

【0088】

少なくとも1つのビジュアルページが表示されている間に受信したユーザ入力に基づいて、トリガーを検出する(ブロック806)。トリガーは、一部の例として、音声コマンド、ユーザジェスチャー、時間遅延に対応してよい。例えば、第1のビジュアルページ122の表示中、ユーザは、音声コマンドを発話すること、又はビジュアルページ122をタップすること等によって、アプリケーションプロトタイプ120とリアルタイムで対話し、試験モジュール132は、ユーザの対話に基づいてトリガー302を取得する。試験

50

モジュール 132 の処理モジュール 304 は、ユーザ入力を処理して、音声コマンド 306、ユーザジェスチャー 308、又は時間遅延 310 等のトリガー 302 を識別する。

【0089】

上記トリガーを、アプリケーションプロトタイプに関連する保存済みのトリガーと比較して、マッチングするトリガーを識別する（ブロック 808）。例えば処理モジュール 304 は、識別されたトリガー 302 を関連対話データ 204 と比較することにより、マッチングするトリガーを識別する。例えば、リクエスト処理モジュール 304 は、音声コマンド 306 又はユーザジェスチャー 308 を、それぞれ、関連対話データ 204 の、保存済みの音声コマンド 212 及びユーザジェスチャー 214 と比較する。マッチングするトリガーを識別するために、処理モジュール 304 は、一意識別子、ハッシュ、テキストベース検索等の多様な技法を活用してよい。更に、処理モジュール 304 は、新しいビジュアルページが表示される際に、時間遅延トリガーを継続的に監視してよい。

10

【0090】

トリガー 302 を音声コマンド 306 として受信した場合、処理モジュール 304 は、音声コマンド 306 のオーディオデータを発話テキスト化サービス 128 に提供する。発話テキスト化サービス 128 は上記オーディオデータをリアルタイムで処理して、音声コマンド 306 の上記オーディオデータをテキスト文字列に変換する。処理モジュール 304 は、音声コマンド 306 のテキスト文字列を発話テキスト化サービス 128 から取得し、この音声コマンド 306 のテキスト文字列を、音声コマンド 212 の保存済みのテキスト文字列と比較することによって、マッチングする音声コマンドを識別する。

20

【0091】

処理モジュール 304 は、受信した音声コマンド 306 のテキスト文字列の文字を、保存済みの音声コマンド 212 のテキスト文字列の文字と比較することによって、マッチングする音声コマンドを識別してよい。一致する文字の数が多いほど、マッチングの蓋然性が高いことを示す。例えばコマンドが単語「mouse」を含む場合、処理モジュール 304 は、単語「house」を、「mouse」とマッチングするものとして認識し得る。というのは、これらの単語は 4 文字が一致しているためである。場合によっては、発話テキスト化サービス 128 は、テキスト文字列が正しいことの信頼性のレベルを示す信頼性スコアを、テキスト文字列と共に含む。この信頼性スコアは、処理モジュール 304 がマッチングする音声コマンドを識別するのを支援する。1 つ以上の実装形態では、処理モジュール 304 は、ユーザがアプリケーションプロトタイプと更に簡単に対話できるようにするために、マッチングするテキスト文字列を、低い許容誤差を用いて選択するよう構成される。

30

【0092】

試験段階中に、マッチングするトリガーに関連するアクションを実施する（ブロック 810）。本明細書に記載の原理によると、アクションは、発話応答、異なるビジュアルページへのページ遷移、又はメディアファイルのプレイバックに対応してよい。例えばトリガー 302 が、関連対話データ 204 の保存済みのトリガー 208 とマッチングすることが検出されたことに応答して、処理モジュール 304 は、関連対話データ 204 中の、上記トリガーに関連する、対応するアクション 210 を識別する。次に、発話応答 314、ページ遷移 316、又はメディア出力 318 に対応するアクション 312 を、実施モジュール 320 へと移動させる。実施モジュール 320 は、発話応答 314 を出力すること、異なるビジュアルページへのページ遷移 316 を実施すること、オーディオ若しくはビデオコンテンツを出力すること、又はこれらの組み合わせ等によって、識別されたトリガーに関連する、対応するアクション 312 を実施させる。

40

【0093】

アクション 312 が発話応答 314 に対応する場合、実施モジュールは、発話応答 314 のテキスト文字列を、テキスト発話化サービス 130 にリアルタイムで提供できる。テキスト発話化サービス 130 は、発話応答 314 のテキスト文字列を、発話応答 314 に対応するオーディオデータに変換する。実施モジュール 320 は次に、発話応答 314 に

50

対応する上記オーディオデータを出力させる。

【0094】

図9は、アプリケーションプロトタイプの実験中にキャプチャされた音声コマンドをマッチングさせるために表音言語翻訳機を利用する、例示の手順900を示す。

【0095】

アプリケーションプロトタイプに関連する対話データをストレージに保持する(ブロック902)。本明細書に記載の原理によると、対話データは、保存済みの音声コマンド、保存済みの音声コマンドそれぞれの検出に回答して実施されることになるアクション、及び少なくとも1つのビジュアルページを含む。上記少なくとも1つのビジュアルページを、アプリケーションプロトタイプのための試験段階の表示デバイス上に表示し(ブロック904)、非表音文字体系を使用する言語で発話された非表音音声コマンドをキャプチャする(ブロック906)。例えば試験段階において、試験モジュール132は、アプリケーションプロトタイプ120の第1のビジュアルページ122(例えばホームページ)の表示を開始する。第1のビジュアルページ122の表示中、ユーザは、非表音音声コマンド402を発話することによってアプリケーションプロトタイプ120と対話する。非表音文字体系を使用する言語で発話された非表音音声コマンド402は、少なくとも1つのマイクロフォンによって、オーディオデータとしてキャプチャされる。

10

【0096】

上記非表音音声コマンドを発話テキスト化サービスにリアルタイムで通信して、上記非表音音声コマンドを、上記音声コマンドの、非表音文字体系の言語の非表音テキスト文字列に変換する(ブロック908)。例えば試験モジュール132は、キャプチャした非表音音声コマンド402のオーディオデータを発話テキスト化サービス128に通信する。発話テキスト化サービス128は、非表音音声コマンド402の上記オーディオデータをリアルタイムで処理して、非表音音声コマンド402の上記オーディオデータを、キャプチャした音声コマンドの非表音テキスト文字列404に変換する。

20

【0097】

上記音声コマンドの上記非表音テキスト文字列を表音言語翻訳機にリアルタイムで通信して、上記非表音テキスト文字列を、上記音声コマンドの、表音文字体系を使用する言語の表音テキスト文字列に翻訳する(ブロック910)。例えば試験モジュール132は、非表音テキスト文字列404を表音言語翻訳機134に通信する。表音言語翻訳機134は、キャプチャした音声コマンドの、非表音文字体系を使用する言語の非表音テキスト文字列404を、上記キャプチャした音声コマンドの、表音文字体系を使用する言語の表音テキスト文字列406に翻訳する。1つ以上の実装形態では、試験段階中、表音言語翻訳機134は、保存済みの音声コマンド408の非表音テキスト文字列410を取得する。例えば、保存済みの音声コマンド408は、ストレージ206に保持された音声コマンド412に対応してよい。表音言語翻訳機134は、保存済みの音声コマンド408の非表音テキスト文字列410を、保存済みの音声コマンド408の表音テキスト文字列412に翻訳する。あるいは、表音言語翻訳機134は、試験段階の前に、保存済みの音声コマンド408の表音テキスト文字列410を、保存済みの音声コマンド408の表音テキスト文字列412に翻訳してよい。この場合、表音テキスト文字列412は、ストレージに保持されていてよい。

30

40

【0098】

上記音声コマンドの上記表音テキスト文字列を、アプリケーションプロトタイプに関連する保存済みの音声コマンドの表音テキスト文字列と比較して、マッチングする音声コマンドを識別する(ブロック912)。例えば試験モジュール132の表音比較モジュール414は、キャプチャした音声コマンド402の表音テキスト文字列406と、保存済みの音声コマンド408の表音テキスト文字列412とを取得する。続いて表音比較モジュール414は、表音テキスト文字列406を表音テキスト文字列412と比較することにより、保存済みの音声コマンドに対応する表音テキスト文字列412の、マッチングする音声コマンド416を識別する。表音比較モジュール414は、マッチングする音声コマ

50

ンド416を、表音テキスト文字列406の文字を保存済みの音声コマンド408の表音テキスト文字列412と比較することによって、識別してよい。表音テキスト文字列406、412は表音文字に翻訳されているため、表音比較モジュール414は、マッチングする音声コマンド416が、キャプチャした音声コマンドと共通して有する文字数に基づいて、マッチングする音声コマンド416を選択できる。いずれのテキスト文字列も表音文字で書かれているため、マッチングする文字数が多くなるほど、マッチングの蓋然性が高くなる。

【0099】

上記マッチングする音声コマンドに関連するアクションを実施する(ブロック914)。例えば、マッチングする音声コマンド416が表音言語比較モジュール414によって識別されると、試験モジュールは、本明細書全体を通して記載されているように、関連するアクションの実施を開始する。

10

【0100】

1つ以上の実装形態による例示の手順を説明したが、これより、本明細書に記載の様々な技法を実装するために利用できる例示的なシステム及びデバイスを考察する。

【0101】

例示的なシステム及びデバイス

図10は、本明細書に記載の様々な技法を実装できる1つ以上の計算システム及び/又はデバイスの代表である例示的な計算デバイス1002を含む、例示的なシステムを、全体として900で示す。これは、設計アプリケーション118及び音声対話設計システム104を含めることによって示されている。計算デバイス1002は例えば、サービスプロバイダのサーバー、クライアントに関連するデバイス(例えばクライアントデバイス)、オンチップシステム、及び/又は他のいずれの好適な計算デバイス若しくは計算システムであってよい。

20

【0102】

図示されているような例示的な計算デバイス1002は、処理システム1004、1つ以上のコンピュータ可読媒体1006、及び互いに通信可能に連結された1つ以上のI/Oインタフェース1008を含む。図示されていないが、計算デバイス1002は更に、上記様々な部品を互いに連結する、システムバス又は他のデータ及びコマンド転送システムを含んでよい。システムバスは、メモリバス若しくはメモリコントローラ、周辺バス、ユニバーサル・シリアル・バス、及び/又は多様なバスアーキテクチャのうちのいずれを利用したプロセッサ若しくはローカルバスといった、異なるバス構造のうちのいずれの1つ又は組み合わせを含むことができる。コントロール及びデータラインといった他の様々な例も考えられる。

30

【0103】

処理システム1004は、ハードウェアを用いて1つ以上の動作を実行する機能を表す。従って処理システム1004、プロセッサ、機能ブロック等として構成してよいハードウェア要素1010を含むものとして図示されている。これは、1つ以上の半導体を用いて形成された特定用途向け集積回路又は他の論理デバイスとしての、ハードウェアでの実装を含んでよい。ハードウェア要素1010は、これを形成する材料、又は採用される処理機構によって限定されない。例えばプロセッサは、1つ以上の半導体及び/又はトランジスタ(例えば電子集積回路(IC))からなるとよい。この文脈において、プロセッサ実行可能命令は、電子的に実行可能な命令であってよい。

40

【0104】

コンピュータ可読記憶媒体1006は、メモリ/ストレージ1012を含むものとして図示されている。メモリ/ストレージ1012は、1つ以上のコンピュータ可読媒体に関連するメモリ/ストレージ用量を表す。メモリ/ストレージ部品1012は、揮発性媒体(ランダムアクセスメモリ(RAM)等)及び/又は不揮発性媒体(読み出し専用メモリ(ROM)、フラッシュメモリ、光学ディスク、磁気ディスク等)を含んでよい。メモリ/ストレージ部品1012は、固定媒体(例えばRAM、ROM、固定ハードドライブ等

50

)、及び着脱式媒体(例えばフラッシュメモリ、着脱式ハードドライブ、光学ディスク等)を含んでよい。コンピュータ可読媒体1006は、以下で更に説明されるような、他の多様な方法で構成してもよい。

【0105】

1つ以上の入力/出力インタフェース1008は、様々な入力/出力デバイスを用いて、ユーザがコマンド及び情報を計算デバイス1002に入力できるようにし、またユーザ及び/又は他の部品若しくはデバイスに情報を提示できるようにする、機能を表す。入力デバイスの例としては、キーボード、カーソル制御デバイス(例えばマウス)、マイクロフォン、スキャナ、タッチ機能(例えば物理的な接触を検出するよう構成された容量性センサ又は他のセンサ)、(例えば接触を伴わないジェスチャーとして運動を認識するために、赤外線周波数等の可視又は不可視波長を採用してよい)カメラ等が挙げられる。出力デバイスの例としては、表示デバイス(例えばモニター又はプロジェクター)、スピーカー、プリンター、ネットワークカード、触覚応答デバイス等が挙げられる。よって計算デバイス1002は、ユーザの対話をサポートするために、以下で更に説明されるような多様な方法で構成してよい。

10

【0106】

本明細書では、ソフトウェア、ハードウェア要素、又はプログラムモジュールの一般的な文脈で、様々な技法を説明できる。一般に、このようなモジュールとしては、特定のタスクを実行する又は特定の抽象データ型を実装する、ルーチン、プログラム、オブジェクト、要素、部品、データ構造等が挙げられる。本明細書中で使用される用語「モジュール(module)」、「機能(functionality)」、及び「部品(component)」は、一般に、ソフトウェア、ファームウェア、ハードウェア、又はこれらの組み合わせを表す。本明細書に記載の技法の特徴はプラットフォームに依存せず、即ち上記技法は、多様なプロセッサを有する多様な市販の計算プラットフォーム上で実装できる。

20

【0107】

本明細書に記載のモジュール及び技法の実装は、何らかの形態のコンピュータ可読媒体上に保存でき、又は何らかの形態のコンピュータ可読媒体を介して伝送できる。コンピュータ可読媒体としては、計算デバイス1002がアクセスできる多様な媒体が挙げられる。例えば限定するものではないが、コンピュータ可読媒体としては、「コンピュータ可読記憶媒体(computer readable storage media)」及び「コンピュータ可読信号媒体(computer readable signal media)」が挙げられる。

30

【0108】

「コンピュータ可読記憶媒体」は、単なる信号伝送、搬送波又は信号自体とは対照的に、情報の持続的及び/又は非一時的な保存が可能な、媒体及び/又はデバイスを指すことができる。よって、コンピュータ可読記憶媒体は、非信号担持媒体である。コンピュータ可読記憶媒体としては、コンピュータ可読命令、データ構造、プログラムモジュール、論理要素/回路又は他のデータといった情報の保存に好適な方法又は技術で実装される、揮発性及び不揮発性の、着脱式及び非着脱式の媒体及び/又はストレージデバイスといった、ハードウェアが挙げられる。コンピュータ可読記憶媒体の例としては、限定するものではないが：RAM、ROM、EEPROM、フラッシュメモリ又は他のメモリ技術；CDROM、デジタル多用途ディスク(digital versatile disk：DVD)又は他の光学ストレージ；ハードディスク、磁気カセット、磁気テープ、磁気ディスクストレージ又は他の磁気ストレージデバイス；あるいは所望の情報の保存に好適であり、かつコンピュータがアクセス可能な、他のストレージデバイス、有形媒体、又は製造物が挙げられる。

40

【0109】

「コンピュータ可読信号媒体」は、ネットワーク等を介して計算デバイス1002のハードウェアに命令を伝送するよう構成された、信号担持媒体を指すことができる。信号媒

50

体は典型的には、コンピュータ可読命令、データ構造、プログラムモジュール、又は他のデータを、搬送波、データ信号又は他の輸送機構といった変調データ信号として具体化できる。信号媒体はまた、いずれの情報送達媒体を含んでよい。用語「変調データ信号 (modulated data signal)」は、その特徴のうちの1つ以上が、信号中の情報をコード化するような様式で設定又は変更された、信号を意味する。例えば限定するものではないが、通信媒体としては、有線ネットワーク又は直接有線接続といった有線媒体、並びに音響、RF、赤外線及び他の無線媒体といった無線媒体が挙げられる。

【0110】

上述したように、ハードウェア要素1010及びコンピュータ可読媒体1006は、いくつかの実施形態において本明細書に記載の技法のうちの少なくともいくつかの態様を実装するため（例えば1つ以上の命令を実行するため）に採用できるハードウェア形態で実装される、モジュール、プログラマブルデバイス論理、及び/又は固定デバイス論理を表す。ハードウェアとしては、集積回路又はオンチップシステムの部品、特定用途向け集積回路 (ASIC)、フィールドプログラマブルゲートアレイ (FPGA)、複合プログラマブル論理デバイス (CPLD)、及びシリコン又は他のハードウェアでの他の実装を含んでよい。この文脈において、ハードウェアは、命令によって定義されたプログラムタスク及び/又はハードウェアによって具体化された論理を実行する処理デバイスとして、並びに命令を実行のために保存するために利用されるハードウェア、例えば上述のコンピュータ可読媒体として、動作できる。

【0111】

以上の組み合わせを採用して、本明細書に記載の様々な技法を実装することもできる。従って、ソフトウェア、ハードウェア、又は実行可能なモジュールを、何らかの形態のコンピュータ可読媒体上で及び/又は1つ以上のハードウェア要素1010によって具体化される、1つ以上の命令及び/又は論理として、実装してよい。計算デバイス1002は、ソフトウェア及び/又はハードウェアモジュールに対応する特定の命令及び/又は機能を実装するよう構成してよい。従って、計算デバイス1002によって実行可能なモジュールの、ソフトウェアとしての実装は、例えば処理システム1004のコンピュータ可読記憶媒体及び/又はハードウェア要素1010の使用によって、少なくとも一部をハードウェアで達成できる。命令及び/又は機能は、本明細書に記載の技法、モジュール及び実施例を実装するために、1つ以上の製造物（例えば1つ以上の計算デバイス1002及び/又は処理システム1004）によって実行可能/実施可能とすることができる。

【0112】

本明細書に記載の技法は、多様な構成の計算デバイス1002でサポートされていてよく、また本明細書に記載の技法の具体例に限定されない。この機能はまた、その全体又は一部を、分散型システムの使用によって、例えば以下に記載されるようにプラットフォーム1016を介して「クラウド」1014上で、実装することもできる。

【0113】

クラウド1014は、リソース1018のためのプラットフォーム1016を含み、及び/又は上記プラットフォーム1016を表す。プラットフォーム1016は、クラウド1014のハードウェア（例えばサーバー）及びソフトウェアリソースの基礎となる機能を抽象化する。リソース1018としては、コンピュータ処理が計算デバイス1002から離れたサーバー上で実行されている間に利用できる、アプリケーション及び/又はデータが挙げられる。またリソース1018としては、インターネットを介して、及び/又はセルラー若しくはWi-Fiネットワークといった加入者ネットワークを通して提供される、サービスも挙げることができる。

【0114】

プラットフォーム1016は、計算デバイス1002を他の計算デバイスと接続するためのリソース及び機能を抽象化してよい。プラットフォーム1016は、プラットフォーム1016によって実装されるリソース1018に関して発生する需要に対応するレベルのスケールを提供するために、リソースのスケールを抽象化する役割も果たすことが

できる。従って、相互接続されたデバイスの実施形態では、本明細書に記載の機能の実装は、システム 1000 全体にわたって分散されていてよい。例えば上記機能は、一部が計算デバイス 1002 上で実装され、またクラウド 1014 の機能を抽象化するプラットフォーム 1016 によって実装されてもよい。

【0115】

結論

本発明を、構造的特徴及び/又は方法に関する行為に関して具体的に説明したが、添付の請求項において定義される本発明は、必ずしも上述の具体的な特徴又は行為に限定されないことを理解されたい。むしろ、上述の具体的な特徴又は行為は、請求対象である発明の例示的な実装の形態として開示されている。

10

【符号の説明】

【0116】

- 102 計算デバイス
- 120 アプリケーションプロトタイプ
- 122 ビジュアルページ
- 124 対話データ
- 128 発話テキスト化サービス
- 134 表音言語翻訳機
- 206 ストレージ
- 210 アクション
- 402 非表音音声コマンド
- 404 音声コマンドの非表音テキスト文字列
- 406 音声コマンドの表音テキスト文字列
- 408 保存済みの音声コマンド
- 412 保存済みの音声コマンドの表音テキスト文字列
- 416 マッチングする音声コマンド
- 504 表示デバイス

20

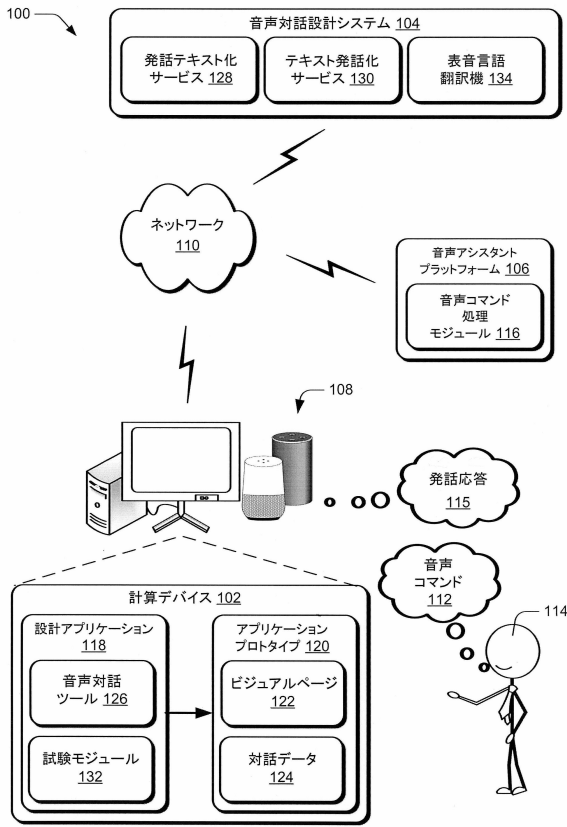
30

40

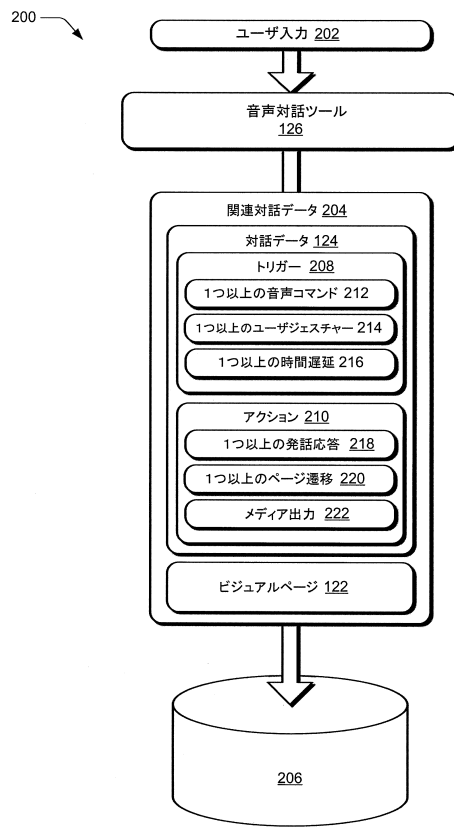
50

【図面】

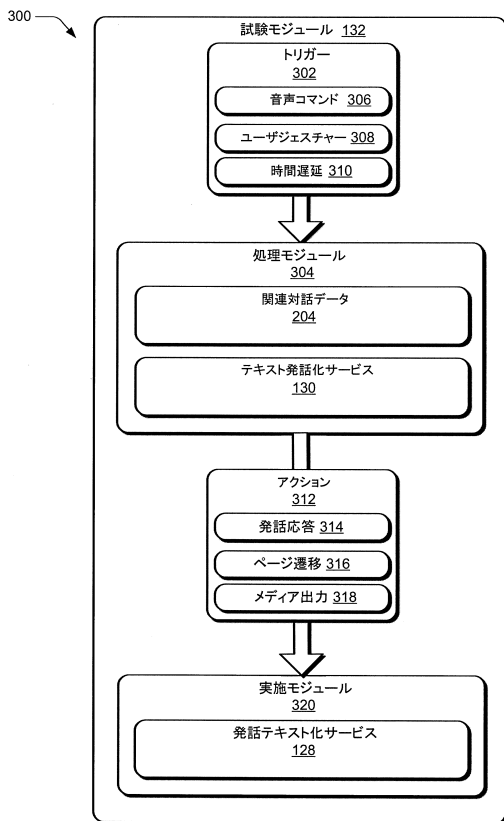
【図 1】



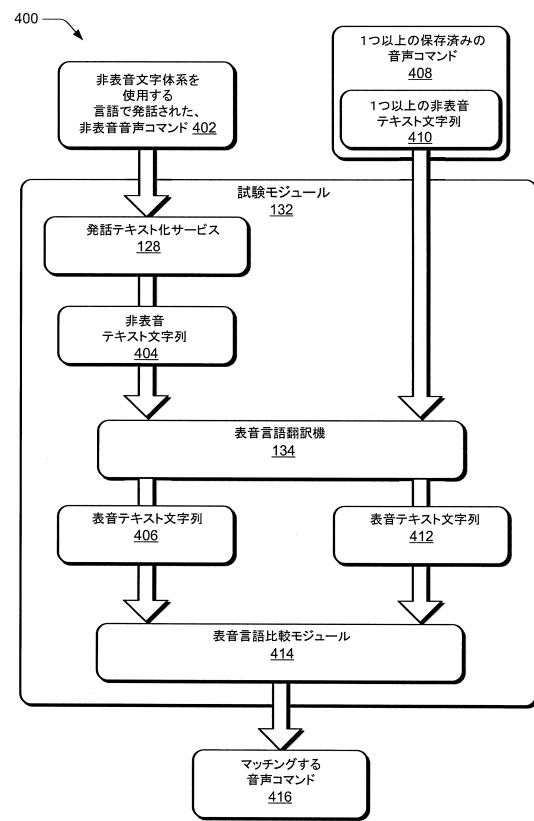
【図 2】



【図 3】



【図 4】



10

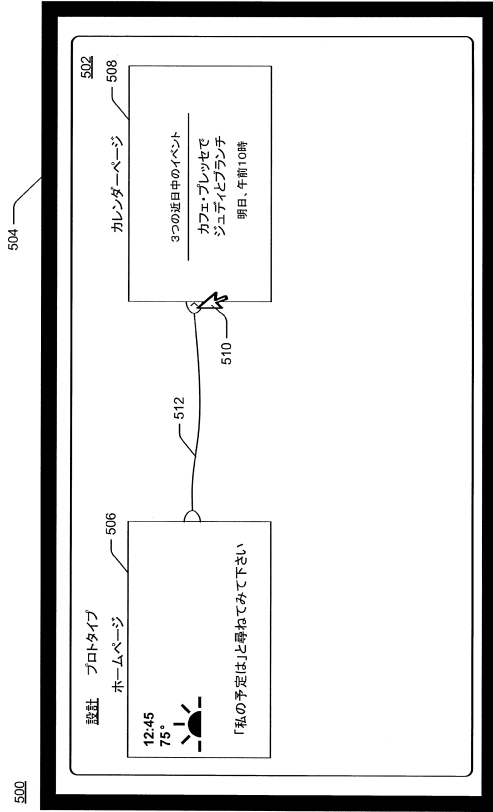
20

30

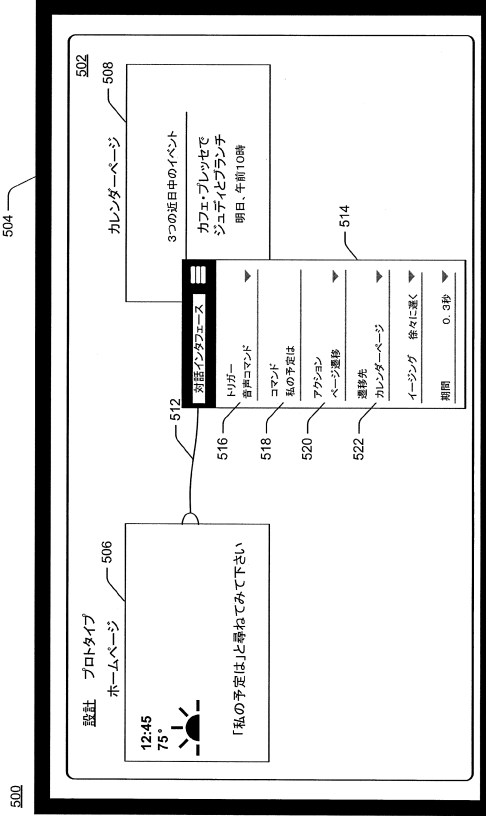
40

50

【図 5 A】



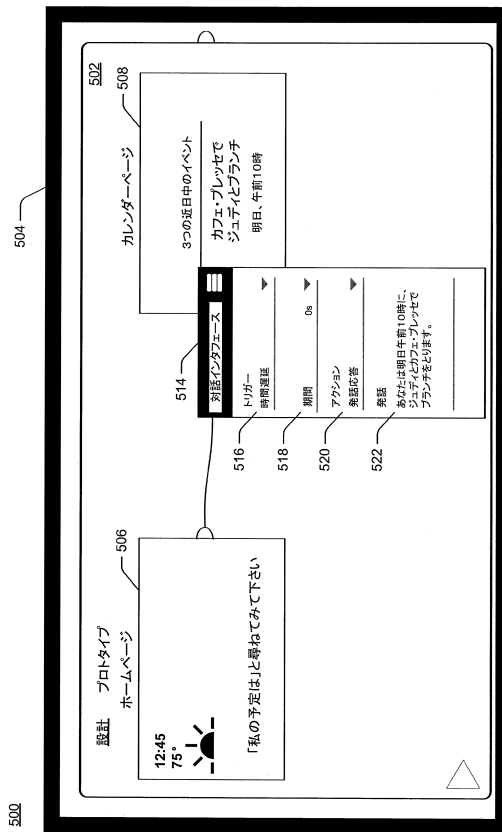
【図 5 B】



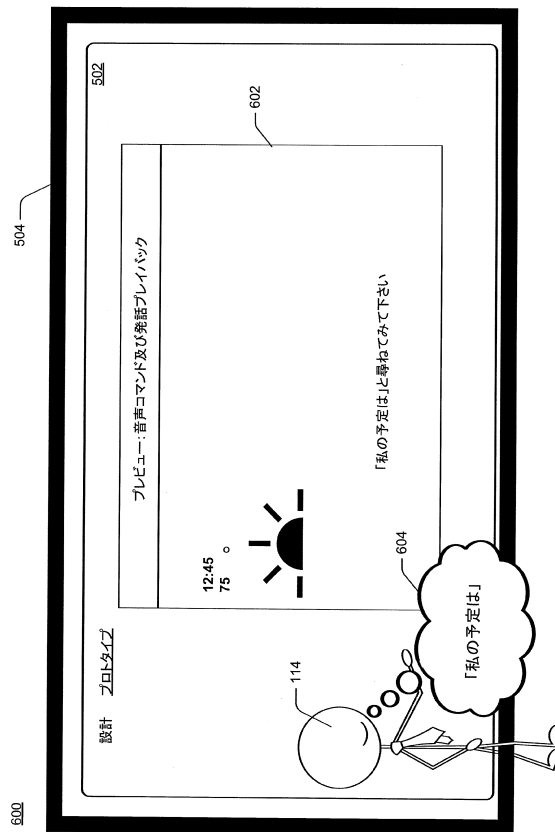
10

20

【図 5 C】



【図 6 A】

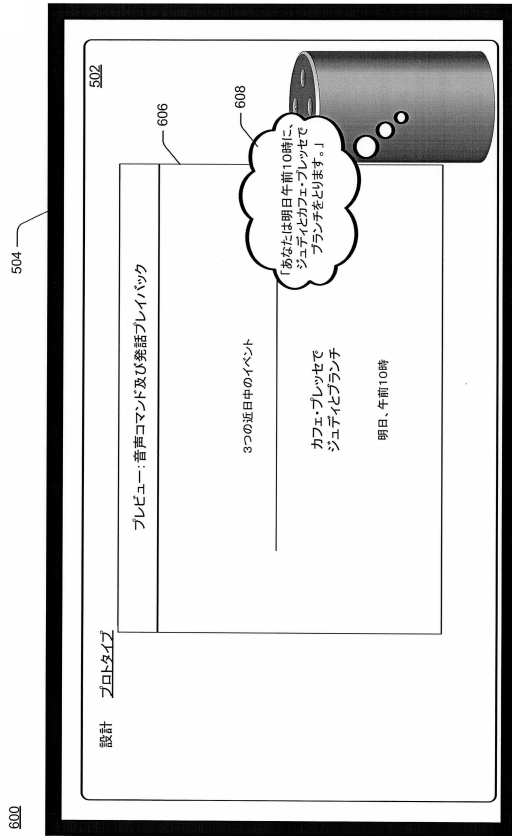


30

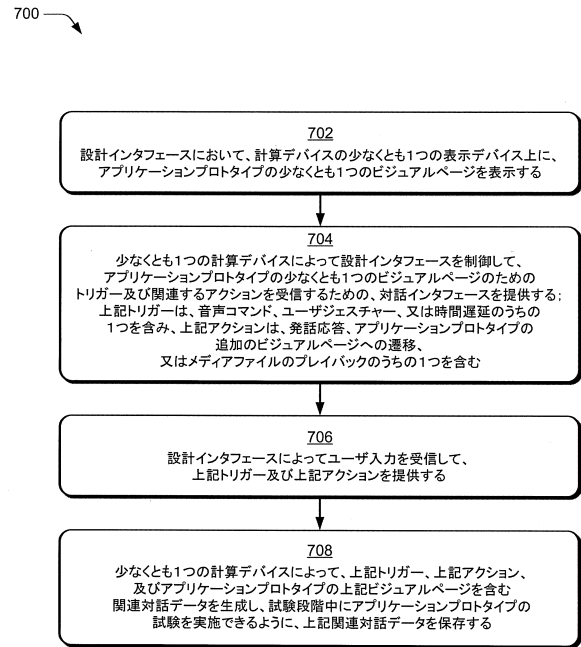
40

50

【図 6 B】



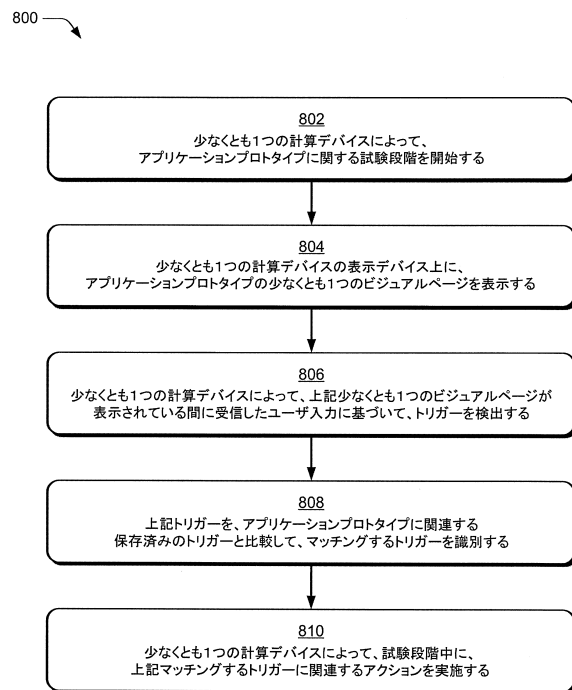
【図 7】



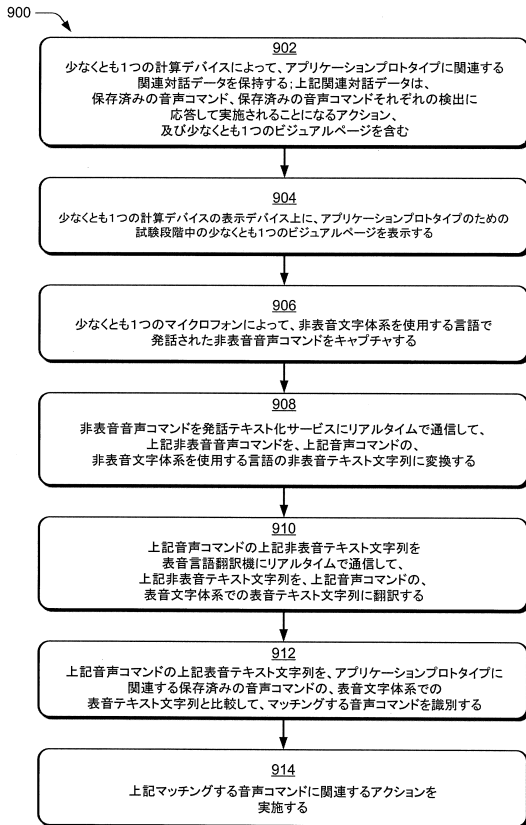
10

20

【図 8】



【図 9】



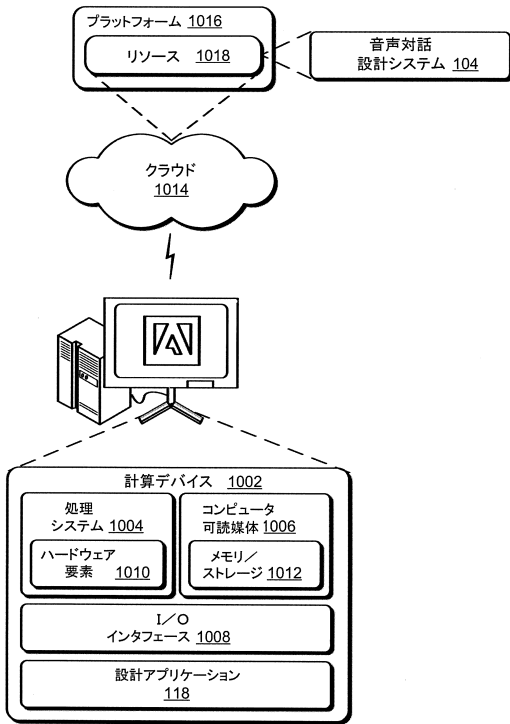
30

40

50

【図 10】

1000



10

20

30

40

50

フロントページの続き

- (72)発明者 スッセ・シェンダービー・ジェンセン
アメリカ合衆国・ニューヨーク州・ブルックリン
- (72)発明者 ダニエル・キャメロン・カンディフ
アメリカ合衆国・ニューヨーク州・ブルックリン
- (72)発明者 ブレイク・アレン・クレイトン・ソーヤー
アメリカ合衆国・ニューヨーク州・アストリア
- 審査官 中村 天真
- (56)参考文献 特開2014-081441(JP,A)
特開2011-008784(JP,A)
特開2008-293109(JP,A)
特表2007-510196(JP,A)
- (58)調査した分野 (Int.Cl., DB名)
G10L 15/00 - 15/34
G06F 3/16
IEEE Xplore