(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5998861号
(P5998861)
(45)発行日　平成28年9月28日(2016.9.28)　(24)登録日　平成28年9月9日(2016.9.9)

(51)Int.Cl.　　F I
G10L 15/22 (2006.01)　　G10L 15/22 200H
G10L 15/28 (2013.01)　　G10L 15/28 230K

請求項の数 18　(全 30 頁)

(21)出願番号　特願2012-246118(P2012-246118)
(22)出願日　平成24年11月8日(2012.11.8)
(65)公開番号　特開2014-95766(P2014-95766A)
(43)公開日　平成26年5月22日(2014.5.22)
審査請求日　平成27年1月23日(2015.1.23)

(73)特許権者　000002185
ソニー株式会社
東京都港区港南１丁目７番１号
(74)代理人　100095957
弁理士　亀谷　美明
(74)代理人　100096389
弁理士　金本　哲男
(74)代理人　100101557
弁理士　萩原　康司
(74)代理人　100128587
弁理士　松本　一騎
(72)発明者　大村　淳己
東京都港区港南１丁目７番１号　ソニー株式会社内

最終頁に続く

(54)【発明の名称】情報処理装置、情報処理方法及びプログラム

(57)【特許請求の範囲】
【請求項１】
入力画像を取得する画像取得部と、
発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させる制御部と、
前記入力画像に映るユーザの身体を認識する画像認識部と、
を備え、
前記制御部は、前記ユーザの音声について実行される音声認識を、前記画像認識部により認識される前記ユーザの身体の所定の部分と前記オブジェクトとの間の前記画面内の位置関係に基づいて、制御する、
情報処理装置。
【請求項２】
前記所定の部分は、前記ユーザの口を含み、
前記制御部は、前記ユーザの口と前記オブジェクトとの間の距離に基づいて、前記音声認識のための音声入力をアクティブ化する、
請求項１に記載の情報処理装置。
【請求項３】
前記所定の部分は、前記ユーザの手を含み、
前記制御部は、前記ユーザの手の動きに従って前記オブジェクトを前記画面内で移動させる、
請求項２に記載の情報処理装置。

【請求項４】
前記制御部は、前記入力画像に映る前記ユーザのジェスチャに応じて、前記音声認識のための音声入力を非アクティブ化する、請求項２又は３に記載の情報処理装置。
【請求項５】
前記制御部は、前記音声認識のための音声入力がアクティブ化されているか否かを、前記オブジェクトの表示属性を変化させることにより前記ユーザに通知する、請求項１～４のいずれか１項に記載の情報処理装置。
【請求項６】
前記制御部は、前記音声認識において音声が検出されているか否かを、前記オブジェクトの表示属性を変化させ又は前記オブジェクトが重畳された出力画像の状態を変化させることにより、前記ユーザに通知する、請求項１～５のいずれか１項に記載の情報処理装置。
【請求項７】
前記制御部は、前記音声認識において検出されている音声のレベルに応じて、前記オブジェクトの前記表示属性又は前記出力画像の前記状態の変化のレベルを変化させる、請求項６に記載の情報処理装置。
【請求項８】
前記音声認識は、可変的な指向性を有するマイクロフォンにより取得される音声信号を用いて実行される、請求項１～７のいずれか１項に記載の情報処理装置。
【請求項９】
前記制御部は、前記オブジェクトの位置を前記ユーザの動きに応じて変化させ、
前記マイクロフォンの指向性は、前記オブジェクトの位置に応じて設定される、
請求項８に記載の情報処理装置。
【請求項１０】
前記制御部は、前記オブジェクトの向きを前記ユーザの動きに応じて変化させ、
前記マイクロフォンの指向性は、前記オブジェクトの向きに応じて設定される、
請求項８又は９に記載の情報処理装置。
【請求項１１】
前記制御部は、前記音声認識において認識された音声の内容を表すテキストを含む第１の追加的なオブジェクトを、前記入力画像に映る前記ユーザの近傍にさらに重畳する、請求項１～１０のいずれか１項に記載の情報処理装置。
【請求項１２】
前記制御部は、前記音声認識が失敗した場合に、前記第１の追加的なオブジェクトの表示属性を変化させ又は特別な文字列を前記テキストに挿入することにより、前記音声認識の失敗を前記ユーザに通知する、請求項１１に記載の情報処理装置。
【請求項１３】
前記制御部は、前記音声認識において検出されている音声のレベルと、前記音声認識を有効に行うために求められる音声のレベルとを示す第２の追加的なオブジェクトを、前記入力画像にさらに重畳する、請求項１～１２のいずれか１項に記載の情報処理装置。
【請求項１４】
前記制御部は、１つ以上の音声コマンドの候補の各々を表すテキストオブジェクトを、前記入力画像にさらに重畳する、請求項１～１３のいずれか１項に記載の情報処理装置。
【請求項１５】
前記情報処理装置は、テレビジョン装置であり、
前記音声コマンドは、前記情報処理装置を前記ユーザが遠隔的に制御するために発せられるコマンドである、
請求項１４に記載の情報処理装置。
【請求項１６】
前記オブジェクトは、マイクロフォンを模したアイコンである、請求項１～１５のいずれか１項に記載の情報処理装置。

【請求項１７】
情報処理装置により実行される情報処理方法であって、
入力画像を取得することと、
発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させることと、
前記入力画像に映るユーザの身体を認識することと、
前記ユーザの音声について実行される音声認識を、認識された前記ユーザの身体の所定の部分と前記オブジェクトとの間の前記画面内の位置関係に基づいて、制御することと、
を含む情報処理方法。
【請求項１８】
情報処理装置を制御するコンピュータを、
入力画像を取得する画像取得部と、
発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させる制御部と、
前記入力画像に映るユーザの身体を認識する画像認識部と、
として機能させ、
前記制御部は、前記ユーザの音声について実行される音声認識を、前記画像認識部により認識される前記ユーザの身体の所定の部分と前記オブジェクトとの間の前記画面内の位置関係に基づいて、制御する、
プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示は、情報処理装置、情報処理方法及びプログラムに関する。
【背景技術】
【０００２】
従来、音声認識は、ユーザによる情報機器への入力を支援する技術として利用されている。例えば、下記特許文献１は、ユーザが発した音声の内容を音声認識によってテキストに変換し、得られたテキストを複数のユーザ間のコミュニケーションのために画面に表示する技術を開示している。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特開２０１２－５８８３８号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
しかしながら、音声認識が機能し音声入力がアクティブになっているタイミングと、ユーザが音声認識のために音声を発するタイミングとが整合しないケースが少なくない。これらタイミングが整合しなければ、認識すべき音声が認識されず、又は予期しない音声が認識されてしまうという不都合が生じ得る。
【０００５】
従って、ユーザが適切なタイミングで音声認識のために音声を発することを支援する、改善された仕組みが提供されることが望ましい。
【課題を解決するための手段】
【０００６】
本開示によれば、入力画像を取得する画像取得部と、発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させる制御部と、を備え、前記制御部は、ユーザの音声について実行される音声認識を、前記オブジェクトを用いて制御する、情報処理装置が提供される。
【０００７】
また、本開示によれば、情報処理装置により実行される情報処理方法であって、入力画

像を取得することと、発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させることと、ユーザの音声について実行される音声認識を、前記オブジェクトを用いて制御することと、を含む情報処理方法が提供される。

【０００８】

また、本開示によれば、情報処理装置を制御するコンピュータを、入力画像を取得する画像取得部と、発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させる制御部と、として機能させ、前記制御部は、ユーザの音声について実行される音声認識を、前記オブジェクトを用いて制御する、プログラムが提供される。

【発明の効果】

【０００９】

本開示に係る技術によれば、ユーザが適切なタイミングで音声認識のために音声を発することを支援することができる。

【図面の簡単な説明】

【００１０】

【図１】第１の実施形態に係る情報処理装置の概要について説明するための説明図である。

【図２】第２の実施形態に係る情報処理装置の概要について説明するための説明図である。

【図３】第１の実施形態に係る情報処理装置のハードウェア構成の一例を示すブロック図である。

【図４】第１の実施形態に係る情報処理装置の論理的機能の構成の一例を示すブロック図である。

【図５】画像認識の結果の一例について説明するための説明図である。

【図６】画像認識の結果の他の例について説明するための説明図である。

【図７】音声認識を制御するために使用される制御オブジェクトの第１の例について説明するための説明図である。

【図８】音声認識を制御するために使用される制御オブジェクトの第２の例について説明するための説明図である。

【図９】音声入力をアクティブ化するためのアクティブ化条件の第１の例について説明するための説明図である。

【図１０】音声入力をアクティブ化するためのアクティブ化条件の第２の例について説明するための説明図である。

【図１１】音声認識結果の視覚的なフィードバックの一例について説明するための説明図である。

【図１２】認識された音声の内容を表す追加的な表示オブジェクトの一例について説明するための第１の説明図である。

【図１３】認識された音声の内容を表す追加的な表示オブジェクトの一例について説明するための第２の説明図である。

【図１４】音声認識を支援する追加的な表示オブジェクトの一例について説明するための説明図である。

【図１５】マイクロフォンの指向性の制御の一例について説明するための第１の説明図である。

【図１６】マイクロフォンの指向性の制御の一例について説明するための第２の説明図である。

【図１７】マイクロフォンの指向性の制御の一例について説明するための第３の説明図である。

【図１８】出力画像のウィンドウ構成の第１の例について説明するための説明図である。

【図１９】出力画像のウィンドウ構成の第２の例について説明するための説明図である。

【図２０】第１の制御シナリオについて説明するための説明図である。

【図２１】第２の制御シナリオについて説明するための説明図である。

【図２２】第３の制御シナリオについて説明するための説明図である。
【図２３】第４の制御シナリオについて説明するための説明図である。
【図２４】第１の実施形態に係る処理の流れの一例を示すフローチャートの前半部である。
【図２５】第１の実施形態に係る処理の流れの一例を示すフローチャートの後半部である。
【図２６】第２の実施形態に係る情報処理装置のハードウェア構成の一例を示すブロック図である。
【図２７】第２の実施形態における制御シナリオの一例について説明するための説明図である。
【発明を実施するための形態】
【００１１】
以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。
【００１２】
また、以下の順序で説明を行う。

【００１３】
＜１．概要＞
本節では、図１及び図２を用いて、本開示に係る技術が適用され得る情報処理装置の概要について説明する。本開示に係る技術は、ユーザインタフェースのための手段として音声認識を活用する様々な装置及びシステムに適用可能である。一例として、本開示に係る技術は、テレビジョン装置、デジタルスチルカメラ又はデジタルビデオカメラなどのデジタル家電機器に適用されてもよい。また、本開示に係る技術は、ＰＣ（Personal Computer）、スマートフォン、ＰＤＡ（Personal Digital Assistant）又はゲーム端末などの端末装置に適用されてもよい。また、本開示に係る技術は、カラオケシステム又はアミューズメント装置のような特殊な用途を有するシステム又は装置に適用されてもよい。
【００１４】
図１は、第１の実施形態に係る情報処理装置１００の概要について説明するための説明図である。図１を参照すると、情報処理装置１００は、テレビジョン装置である。情報処理装置１００は、カメラ１０１、マイクロフォン１０２及びディスプレイ１０８を備える。カメラ１０１は、情報処理装置１００のディスプレイ１０８を見るユーザを撮像する。マイクロフォン１０２は、ユーザが発する音声を集音する。ディスプレイ１０８は、情報処理装置１００により生成される画像を表示する。ディスプレイ１０８により表示される画像は、コンテンツ画像に加えて、ユーザインタフェース（ＵＩ）画像を含み得る。図１の例では、ユーザＵａ及びＵｂがディスプレイ１０８を見ている。ディスプレイ１０８には、ＵＩ画像Ｗ０１が表示されている。ＵＩ画像Ｗ０１は、カメラ１０１により撮像される撮像画像を用いて生成され、それによりいわゆるミラー表示が実現される。情報処理装置１００は、音声認識機能を有する。ユーザＵａ及びＵｂは、マイクロフォン１０２を介して情報処理装置１００へ音声を入力することにより、情報処理装置１００を操作し又は情報処理装置１００へ情報を入力することができる。
【００１５】

図２は、第２の実施形態に係る情報処理装置２００の概要について説明するための説明図である。図２を参照すると、情報処理装置２００は、タブレットＰＣである。情報処理装置２００は、カメラ２０１、マイクロフォン２０２及びディスプレイ２０８を備える。カメラ２０１は、情報処理装置２００のディスプレイ２０８を見るユーザを撮像する。マイクロフォン２０２は、ユーザが発する音声を集音する。ディスプレイ２０８は、情報処理装置２００により生成される画像を表示する。ディスプレイ２０８により表示される画像は、コンテンツ画像に加えて、ＵＩ画像を含み得る。図２の例では、ユーザＵｃがディスプレイ２０８を見ている。ディスプレイ２０８には、ＵＩ画像Ｗ０２が表示されている。ＵＩ画像Ｗ０２は、カメラ２０１により撮像される撮像画像を用いて生成され、それによりいわゆるミラー表示が実現される。情報処理装置２００は、音声認識機能を有する。ユーザＵｃは、マイクロフォン２０２を介して情報処理装置２００へ音声を入力することにより、情報処理装置２００を操作し又は情報処理装置２００へ情報を入力することができる。

【００１６】

これら装置において、音声認識機能が動作し音声入力がアクティブになっている間、ユーザが音声認識のための音声のみを発するとは限らない。また、音声入力がアクティブになっていない時にユーザが音声認識のための音声を発する可能性もある。このようなタイミングの不整合は、認識しなくてもよい音声の認識又は音声認識の不成功などといった、ユーザにとって不都合な結果を招来し得る。そこで、情報処理装置１００及び２００は、次節より詳細に説明する仕組みに従って、ユーザが適切なタイミングで音声認識のために音声を発することを支援する。

【００１７】

＜２．第１の実施形態＞

［２－１．ハードウェア構成例］

図３は、情報処理装置１００のハードウェア構成の一例を示すブロック図である。図３を参照すると、情報処理装置１００は、カメラ１０１、マイクロフォン１０２、入力デバイス１０３、通信インタフェース（Ｉ／Ｆ）１０４、メモリ１０５、チューナ１０６、デコーダ１０７、ディスプレイ１０８、スピーカ１０９、遠隔制御Ｉ／Ｆ１１０、バス１１１及びプロセッサ１１２を備える。

【００１８】

（１）カメラ

カメラ１０１は、ＣＣＤ（Charge Coupled Device）又はＣＭＯＳ（Complementary Metal Oxide Semiconductor）などの撮像素子を有し、画像を撮像する。カメラ１０１により撮像される画像（動画を構成する各フレーム）は、情報処理装置１００による処理のための入力画像として扱われる。

【００１９】

（２）マイクロフォン

マイクロフォン１０２は、ユーザにより発せられる音声を集音し、音声信号を生成する。マイクロフォン１０２により生成される音声信号は、情報処理装置１００による音声認識のための入力音声として扱われる。マイクロフォン１０２は、無指向性マイクロフォンであってもよく、又は固定的な若しくは可変的な指向性を有していてもよい。あるシナリオにおいて、マイクロフォン１０２は可変的な指向性を有し、その指向性は動的に制御される。

【００２０】

（３）入力デバイス

入力デバイス１０３は、ユーザが情報処理装置１００を直接的に操作するために使用されるデバイスである。入力デバイス１０３は、例えば、情報処理装置１００の筐体に配設されるボタン、スイッチ及びダイヤルなどを含み得る。入力デバイス１０３は、ユーザ入力を検出すると、検出されたユーザ入力に対応する入力信号を生成する。

【００２１】

（４）通信インタフェース
通信Ｉ／Ｆ１０４は、情報処理装置１００による他の装置との間の通信を仲介する。通信Ｉ／Ｆ１０４は、任意の無線通信プロトコル又は有線通信プロトコルをサポートし、他の装置との間の通信接続を確立する。

【００２２】
（５）メモリ
メモリ１０５は、半導体メモリ又はハードディスクなどの記憶媒体により構成され、情報処理装置１００による処理のためのプログラム及びデータ、並びにコンテンツデータを記憶する。メモリ１０５により記憶されるデータは、例えば、後に説明する画像認識及び音声認識のための特徴データを含み得る。なお、本明細書で説明するプログラム及びデータの一部又は全部は、メモリ１０５により記憶されることなく、外部のデータソース（例えば、データサーバ、ネットワークストレージ又は外付けメモリなど）から取得されてもよい。

【００２３】
（６）チューナ
チューナ１０６は、アンテナ（図示せず）を介して受信される放送信号から、所望のチャンネルのコンテンツ信号を抽出し及び復調する。そして、チューナ１０６は、復調したコンテンツ信号をデコーダ１０７へ出力する。

【００２４】
（７）デコーダ
デコーダ１０７は、チューナ１０６から入力されるコンテンツ信号からコンテンツデータを復号する。デコーダ１０７は、通信Ｉ／Ｆ１０４を介して受信されるコンテンツ信号からコンテンツデータを復号してもよい。デコーダ１０７により復号されるコンテンツデータに基づいて、コンテンツ画像が生成され得る。

【００２５】
（８）ディスプレイ
ディスプレイ１０８は、ＬＣＤ（Liquid Crystal Display）、ＯＬＥＤ（Organic light-Emitting Diode）又はＣＲＴ（Cathode Ray Tube）などにより構成される画面を有し、情報処理装置１００により生成される画像を表示する。例えば、図１及び図２を用いて説明したコンテンツ画像及びＵＩ画像が、ディスプレイ１０８の画面に表示され得る。

【００２６】
（９）スピーカ
スピーカ１０９は、振動板及びアンプなどの回路素子を有し、情報処理装置１００により生成される出力音声信号に基づいて、音声を出力する。スピーカ１０９の音量は、変更可能である。

【００２７】
（１０）遠隔制御インタフェース
遠隔制御Ｉ／Ｆ１１０は、ユーザにより使用されるリモートコントローラから送信される遠隔制御信号（赤外線信号又はその他の無線信号）を受信するインタフェースである。遠隔制御Ｉ／Ｆ１１０は、遠隔制御信号を検出すると、検出された遠隔制御信号に対応する入力信号を生成する。

【００２８】
（１１）バス
バス１１１は、カメラ１０１、マイクロフォン１０２、入力デバイス１０３、通信Ｉ／Ｆ１０４、メモリ１０５、チューナ１０６、デコーダ１０７、ディスプレイ１０８、スピーカ１０９、遠隔制御Ｉ／Ｆ１１０及びプロセッサ１１２を相互に接続する。

【００２９】
（１２）プロセッサ
プロセッサ１１２は、例えば、ＣＰＵ（Central Processing Unit）又はＤＳＰ（Dig

ital Signal Processor）などであってよい。プロセッサ１１２は、メモリ１０５又は他の記憶媒体に記憶されるプログラムを実行することにより、後に説明する情報処理装置１００の様々な機能を動作させる。

【００３０】

［２－２．機能構成例］

図４は、図３に示した情報処理装置１００のメモリ１０５及びプロセッサ１１２により実現される論理的機能の構成の一例を示すブロック図である。図４を参照すると、情報処理装置１００は、画像取得部１２０、音声取得部１３０、アプリケーション部１４０、認識部１５０、特徴データベース（ＤＢ）１６０及び制御部１７０を備える。認識部１５０は、画像認識部１５２及び音声認識部１５４を含む。制御部１７０は、認識制御部１７２及び表示制御部１７４を含む。なお、図４に示した機能ブロックの一部は、情報処理装置１００の外部の（例えば、クラウドコンピューティング環境内の）装置において実現されてもよい。例えば、画像認識部１５２は、以下に説明する画像認識処理を自ら実行する代わりに、当該処理を外部の画像認識機能に実行させてもよい。同様に、音声認識部１５４は、以下に説明する音声認識処理を自ら実行する代わりに、当該処理を外部の音声認識機能に実行させてもよい。

【００３１】

（１）画像取得部

画像取得部１２０は、カメラ１０１により撮像される画像を入力画像として取得する。入力画像は、典型的には、ユーザが映る動画を構成する一連のフレームの各々である。そして、画像取得部１２０は、取得した入力画像を認識部１５０及び制御部１７０へ出力する。

【００３２】

（２）音声取得部

音声取得部１３０は、マイクロフォン１０２により生成される音声信号を入力音声として取得する。そして、音声取得部１３０は、取得した入力音声を認識部１５０へ出力する。

【００３３】

（３）アプリケーション部

アプリケーション部１４０は、情報処理装置１００が有する様々なアプリケーション機能を実行する。例えば、テレビジョン番組再生機能、電子番組表表示機能、録画設定機能、写真再生機能、動画再生機能、音楽再生機能及びインターネットブラウジング機能などが、アプリケーション部１４０により実行されてよい。アプリケーション部１４０は、アプリケーション機能を通じて生成される（コンテンツ画像を含み得る）アプリケーション画像及び音声を、制御部１７０へ出力する。

【００３４】

本実施形態において、アプリケーション部１４０により実行されるアプリケーション機能の少なくとも一部は、後述する音声認識部１５４と連携し、ユーザからの音声入力を受け付ける。例えば、テレビジョン番組再生機能は、音声認識部１５４により認識される音声コマンドに従って、再生されるチャンネル及び音量などの設定を変更し得る。電子番組表表示機能は、音声認識部１５４により認識される音声コマンドに従って、表示すべき電子番組表のチャンネル及び時間帯を変更し得る。写真再生機能は、音声認識部１５４により認識される指定日に撮像された写真を再生し得る。インターネットブラウジング機能は、音声認識部１５４により認識されるキーワードを用いたインターネット検索を実行し得る。

【００３５】

（４）画像認識部

画像認識部１５２は、画像取得部１２０から入力される入力画像に映るユーザの身体を認識する。例えば、画像認識部１５２は、入力画像から抽出される画像特徴量をユーザの

身体の所定の部分について特徴ＤＢ１６０により予め記憶される画像特徴量と照合することにより、当該所定の部分を認識する。所定の部分とは、例えば、ユーザの手、口及び顔のうちの少なくとも１つを含み得る。

【００３６】

図５は、画像認識部１５２による画像認識の結果の一例について説明するための説明図である。図５を参照すると、入力画像Ｗ０３にユーザＵａが映っている。ユーザＵａは、カメラ１０１の方向を向き、左手を挙げている。画像認識部１５２は、画像特徴量の照合又はその他の公知の手法を用いて、入力画像Ｗ０３内の手領域Ａ０１、口領域Ａ０２及び顔領域Ａ０３を認識し得る。そして、画像認識部１５２は、認識したこれら領域の画像内の位置を示す位置データを、制御部１７０へ出力する。

【００３７】

一例として、画像認識部１５２は、入力画像内で認識した顔領域の部分画像（顔画像）を特徴ＤＢ１６０により予め記憶される既知のユーザの顔画像データと照合することにより、ユーザを識別してもよい。画像認識部１５２によるユーザ識別結果は、例えば、音声認識の調整、ＵＩ画像に表示されるメニューの個人化又はアプリケーション部１４０によるコンテンツの推薦などの用途に使用され得る。なお、ユーザの識別（即ち、個人認識）は、入力画像ではなく、入力音声に基づいて行われてもよい。

【００３８】

本実施形態において、画像認識部１５２は、入力画像に映るユーザのジェスチャをも認識し得る。なお、本明細書において、ジェスチャとの用語は、ユーザの身体の動的な動きを伴わないいわゆるポーズ（形状）をも含むものとする。

【００３９】

図６は、画像認識部１５２による画像認識の結果の他の例について説明するための説明図である。図６を参照すると、入力画像Ｗ０４にユーザＵａ及びＵｂが映っている。ユーザＵａは、右手の人差し指を口に当てるジェスチャを行っている。画像認識部１５２は、入力画像Ｗ０４内の手領域Ａ０４を認識し、ユーザＵａの上記ジェスチャをさらに認識し得る。ユーザＵｂは、両手で口を塞ぐジェスチャを行っている。画像認識部１５２は、入力画像Ｗ０４内の手領域Ａ０５を認識し、ユーザＵｂの上記ジェスチャをさらに認識し得る。画像認識部１５２は、ユーザのジェスチャを認識すると、認識したジェスチャの種類を示すジェスチャデータを、制御部１７０へ出力する。

【００４０】

（５）音声認識部

音声認識部１５４は、音声取得部１３０から入力される入力音声に基づいて、ユーザの音声を認識する。本実施形態において、音声取得部１３０から音声認識部１５４への音声入力は、後述する認識制御部１７２によりアクティブ化され、又は非アクティブ化される。音声入力がアクティブである間、音声認識部１５４は、入力音声をその内容を示すテキストに変換する。実行中のアプリケーションがフリーテキストの入力を受け付ける場合には、音声認識部１５４は、認識した音声の内容を示すテキストを、アプリケーション部１４０へ出力し得る。その代わりに、実行中のアプリケーションが所定の音声コマンドセット内の音声コマンドの入力を受け付ける場合には、音声認識部１５４は、ユーザの音声から認識した音声コマンドを識別する識別子を、アプリケーション部１４０へ出力してもよい。音声入力が非アクティブである間、音声認識部１５４は、音声認識を実行しない。

【００４１】

音声認識部１５４は、音声取得部１３０から入力される入力音声のレベルを判定し、判定したレベルを制御部１７０へ通知してもよい。後述する認識制御部１７２は、音声認識部１５４から通知される入力音声のレベルに応じて、画面上でのユーザへの様々なフィードバックを行い得る。

【００４２】

上述したように、あるシナリオにおいて、マイクロフォン１０２は可変的な指向性を有する。この場合、後述する認識制御部１７２により、マイクロフォン１０２の指向性が設

定される。そして、音声認識部１５４は、設定された指向性に対応する方向に位置するユーザの音声を、マイクロフォン１０２により取得される音声信号を用いて認識する。

【００４３】

（６）特徴データベース

特徴ＤＢ１６０は、画像認識部１５２により画像認識のために使用される画像特徴データ、及び音声認識部１５４により音声認識のために使用される音声特徴データを予め記憶する。画像特徴データは、例えば、ユーザの手、口又は顔などの所定の部分の既知の画像特徴量を含み得る。また、画像特徴データは、ユーザごとの顔画像データを含んでもよい。また、画像特徴データは、画像認識部１５２が認識すべきジェスチャを定義するジェスチャ定義データを含んでもよい。音声特徴データは、例えば、ユーザごとの発話の特徴を示す音声特徴量を含み得る。

【００４４】

（７）認識制御部１７２

認識制御部１７２は、入力画像に重畳されるオブジェクトであって、発話に関連する当該オブジェクトを生成する。そして、認識制御部１７２は、生成した当該オブジェクトを用いて、音声認識部１５４により実行される音声認識を制御する。以下、音声認識を制御するために使用されるこのオブジェクトを、制御オブジェクトという。制御オブジェクトは、ユーザによる操作に従って画面上で移動してもよく、又は固定的な位置に表示されてもよい。

【００４５】

図７は、制御オブジェクトの第１の例について説明するための説明図である。図７を参照すると、入力画像Ｗ０５に制御オブジェクトＩＣ１が重畳されている。制御オブジェクトＩＣ１は、手持ち型のマイクロフォンを模したアイコンである。認識制御部１７２は、例えば、ユーザからの音声入力を受け付けるアプリケーション（以下、音声対応アプリケーションという）が起動されると、画面上の規定の表示位置又は画像認識部１５２により認識されるユーザの身体の近傍に、制御オブジェクトＩＣ１を表示させる。そして、認識制御部１７２は、ユーザの動き（例えば、手領域の動き）に応じて、制御オブジェクトＩＣ１の表示位置を変化させる。認識制御部１７２は、ユーザの動き（例えば、手領域の回転）に応じて、制御オブジェクトＩＣ１の向きを変化させてもよい。音声対応アプリケーションが終了すると、制御オブジェクトＩＣ１は画面から消去され、又は非アクティブ化されて既定の表示位置若しくは画面の端部へ移動し得る。

【００４６】

図８は、制御オブジェクトの第２の例について説明するための説明図である。図８を参照すると、入力画像Ｗ０６に制御オブジェクトＩＣ２が重畳されている。制御オブジェクトＩＣ２は、スタンド型のマイクロフォンを模したアイコンである。認識制御部１７２は、例えば、音声対応アプリケーションが起動されると、画面上の既定の表示位置に制御オブジェクトＩＣ２を表示させる。制御オブジェクトＩＣ２の表示位置は移動しない。音声対応アプリケーションが終了すると、制御オブジェクトＩＣ２は画面から消去され得る。

【００４７】

なお、図７及び図８に示した制御オブジェクトＩＣ１及びＩＣ２は一例に過ぎない。例えば、口若しくは拡声器を模した他の種類のアイコン又はテキストラベルなどが制御オブジェクトとして使用されてもよい。また、制御オブジェクトの外観ではなく、制御オブジェクトの機能性が発話に関連していてもよい。

【００４８】

本実施形態において、認識制御部１７２は、画像認識部１５２により認識されるユーザの身体の所定の部分と制御オブジェクトとの間の画面内の位置関係に基づいて、音声認識部１５４により実行される音声認識を制御する。例えば、認識制御部１７２は、当該位置関係に基づくアクティブ化条件が満たされている場合に、音声認識部１５４への音声入力をアクティブ化する。認識制御部１７２は、アクティブ化条件が満たされていない場合に、音声認識部１５４への音声入力をアクティブ化しない。

【００４９】
図９は、音声入力をアクティブ化するためのアクティブ化条件の第１の例について説明するための説明図である。図９を参照すると、入力画像Ｗ０７ａ及びＷ０７ｂにユーザＵａが映っている。画像認識部１５２は、入力画像に映るユーザの口領域及び手領域を認識する。第１の例において、アクティブ化条件は、ユーザの口と制御オブジェクトとの間の距離が距離閾値Ｄ１を下回る、という条件である。図中には、口領域の中心点Ｇ１を中心とし半径が距離閾値Ｄ１に等しい円が点線で示されている。認識制御部１７２は、認識される手領域Ａ０１の動きに従って、制御オブジェクトＩＣ１を画面内で移動させる。図９の上段では、ユーザの口と制御オブジェクトＩＣ１との間の距離が距離閾値Ｄ１を上回るため、音声入力は非アクティブである。即ち、ユーザが音声を発しても（又は近傍で雑音が発生しても）、音声認識部１５４は音声を認識しない。従って、その間、ユーザが意図しない音声認識に起因して、アプリケーションが予期しない動作をすることが防止される。図９の下段において、ユーザが手を動かした結果、ユーザの口と制御オブジェクトＩＣ１との間の距離が距離閾値Ｄ１を下回っている。そこで、認識制御部１７２は、アクティブ化条件が満たされていると判定し、音声入力をアクティブ化する。すると、ユーザにより発せられる音声が音声認識部１５４により認識されるようになる。なお、ユーザの身体の口以外の部分と制御オブジェクトとの間の距離が上記距離閾値と比較されてもよい。
【００５０】
図１０は、音声入力をアクティブ化するためのアクティブ化条件の第２の例について説明するための説明図である。図１０を参照すると、入力画像Ｗ０８ａ及びＷ０８ｂにユーザＵｂが映っている。また、入力画像Ｗ０８ａ及びＷ０８ｂに制御オブジェクトＩＣ２が重畳されている。画像認識部１５２は、入力画像に映るユーザの口領域Ａ０６を認識する。第２の例において、アクティブ化条件は、ユーザの口と制御オブジェクトとの間の距離が距離閾値Ｄ２を下回る、という条件である。図中には、制御オブジェクト上の基準点Ｇ２を中心とし半径が距離閾値Ｄ２に等しい円が点線で示されている。図１０の上段では、ユーザの口と制御オブジェクトＩＣ２との間の距離が距離閾値Ｄ２を上回るため、音声入力は非アクティブである。即ち、ユーザが音声を発しても（又は近傍で雑音が発生しても）、音声認識部１５４は音声を認識しない。従って、その間、ユーザが意図しない音声認識に起因して、アプリケーションが予期しない動作をすることが防止される。図１０の下段において、ユーザが移動した結果、ユーザの口と制御オブジェクトＩＣ２との間の距離が距離閾値Ｄ２を下回っている。そこで、認識制御部１７２は、アクティブ化条件が満たされていると判定し、音声入力をアクティブ化する。すると、ユーザにより発せられる音声が音声認識部１５４により認識されるようになる。
【００５１】
なお、図９及び図１０を用いて説明したアクティブ化条件は一例に過ぎない。例えば、制御オブジェクトへのタッチ又は制御オブジェクトを高く掲げるなどといった、制御オブジェクトに関連する所定のジェスチャの検出が、アクティブ化条件として定義されてもよい。
【００５２】
音声入力が一度アクティブ化された後、認識制御部１７２は、所定の非アクティブ化条件が満たされるまで、音声入力のアクティブ状態を継続させる。非アクティブ化条件は、例えば、上記アクティブ化条件の単純な反対（例えば、ユーザの口と制御オブジェクトとの間の距離が距離閾値を上回る、など）であってもよい。その代わりに、非アクティブ化条件は、画像認識部１５２によるユーザの所定のジェスチャの認識などであってもよい。音声入力を非アクティブ化するためのジェスチャとは、例えば、人差し指を口に当てるジェスチャなどであってよい。また、非アクティブ化条件は、一単位の音声コマンドの認識の成功、又はアクティブ化からの所定の期間の経過などを含んでもよい。
【００５３】
音声入力がアクティブである間、認識制御部１７２は、音声認識部１５４による音声認識に関連するユーザへの視覚的なフィードバックをも制御する。

【００５４】
例えば、認識制御部１７２は、制御オブジェクトの表示属性を変化させることにより、音声認識部１５４への音声入力がアクティブ化されていることをユーザに通知する。認識制御部１７２により変更される制御オブジェクトの表示属性は、例えば、色、輝度、透明度、サイズ、形状及びテクスチャのうち少なくとも１つを含み得る。図９及び図１０の例では、音声入力がアクティブであるか否かが、制御オブジェクトのテクスチャの変化によって示されている。
【００５５】
また、例えば、認識制御部１７２は、音声認識部１５４から通知される入力音声のレベルをユーザへフィードバックする。入力音声のレベルのフィードバックは、制御オブジェクトの表示属性を変化させ、又は制御オブジェクトが重畳されたＵＩ画像の状態を変化させることにより行われてよい。図１１は、音声認識結果の視覚的なフィードバックの一例について説明するための説明図である。図１１を参照すると、制御オブジェクトＩＣ１が重畳されたＵＩ画像Ｗ０９に、エフェクトＦｂ１が適用されている。エフェクトＦｂ１は、制御オブジェクトＩＣ１（ユーザの口であってもよい）から波動が放出されているかのようなＵＩ画像の状態を表現する。入力音声のレベルが所定の閾値を下回る場合には、エフェクトＦｂ１は解除され得る。こうしたフィードバックによれば、ユーザは、自身が発した音声を情報処理装置１００が適切に検出しているか否かを、直感的に把握することができる。認識制御部１７２は、上記所定の閾値を上回る入力音声のレベルに応じて、制御オブジェクトの表示属性の変化のレベル又は出力画像の状態の変化のレベルを変化させてもよい。例えば、入力音声のレベルがより大きいほどより広い画像領域に、エフェクトＦｂ１が適用されてもよい。それにより、ユーザは、自身が発した音声について情報処理装置１００が検出したレベルを、直感的に把握することができる。なお、認識制御部１７２は、エフェクトＦｂ１の表示属性（例えば、色など）を、音声認識のステータス又はエラーの有無を示すように変化させてもよい。入力音声のレベルの所定の基準値との比較の結果が、ＵＩ画像Ｗ０９においてテキストで示されてもよい。
【００５６】
また、例えば、認識制御部１７２は、音声認識部１５４により認識された音声の内容を表すテキストを含む追加的な表示オブジェクトを、入力画像に映るユーザの近傍にさらに重畳してもよい。図１２及び図１３は、認識された音声の内容を表す追加的な表示オブジェクトの一例について説明するための説明図である。図１２を参照すると、制御オブジェクトＩＣ１及び追加オブジェクトＦｂ２がＵＩ画像Ｗ１０に重畳されている。追加オブジェクトＦｂ２は、ＵＩ画像Ｗ１０に映るユーザＵａが発した音声の内容を表すテキストを含む吹き出しである。こうしたフィードバックによれば、ユーザは、自身が発した音声を情報処理装置１００が正しく認識したか否かを、即座に把握することができる。図１３を参照すると、追加オブジェクトＦｂ２は、ランダム文字列Ｓｔｒ１を含む。ランダム文字列Ｓｔｒ１は、所定の閾値を上回るレベルの入力音声が検出されたものの、当該入力音声に基づく音声認識が失敗した場合に、追加オブジェクトＦｂ２に挿入され得る。こうしたフィードバックによれば、ユーザは、自身が発した音声のレベルが十分であったものの音声認識が失敗したことを、即座に把握することができる。音声認識の失敗は、追加オブジェクトＦｂ２の表示属性を変化させることによりユーザに通知されてもよい。なお、追加オブジェクトＦｂ２は、ランダム文字列の代わりに、空白を含んでもよい。ランダム文字列又は空白の長さは、音声認識が失敗した間の発話時間の長さに応じて決定されてもよい。
【００５７】
また、例えば、認識制御部１７２は、音声認識部１５４により検出されている音声のレベルと、音声認識を有効に行うために求められる音声のレベルとを示す追加的なオブジェクトを入力画像に重畳してもよい。音声認識を有効に行うために求められる音声のレベルは、メモリ１０５により予め記憶されてもよく、又は環境の雑音レベルに依存して動的に計算されてもよい。図１４は、音声認識を支援する追加的な表示オブジェクトの一例につ

いて説明するための説明図である。図１４を参照すると、ＵＩ画像Ｗ１２に、制御オブジェクトＩＣ１、追加オブジェクトＦｂ２及び追加オブジェクトＦｂ３が重畳されている。追加オブジェクトＦｂ２は、音声の内容を表すテキストを含む吹き出しである。ここでは、ユーザが発した音声のレベルが十分ではないことに起因して音声認識が失敗した結果、追加オブジェクトＦｂ２の背景色が暗い色に変更されている。追加オブジェクトＦｂ３は、音声のレベルを通知するインジケータである。追加オブジェクトＦｂ３の外側の点線の円周の半径は、音声認識を有効に行うために求められる音声のレベルに対応する。塗りつぶされた円の半径は、音声認識部１５４から通知される入力音声のレベルに対応する。入力音声のレベルが高くなれば、塗りつぶされた円は大きくなる。なお、追加オブジェクトＦｂ３は、図１４の例に限定されず、例えば帯状のインジケータなどであってもよい。こうしたフィードバックによれば、ユーザは、自身が発した音声のレベルが不十分であった場合に、どの程度声を大きくすれば音声認識が成功し得るかを、直感的に把握することができる。なお、認識制御部１７２は、追加オブジェクトＦｂ３の表示属性（例えば、色など）を、音声認識のステータス又はエラーの有無を示すように変化させてもよい。入力音声のレベルの所定の基準値との比較の結果が、ＵＩ画像Ｗ１２においてテキストで示されてもよい。

【００５８】

マイクロフォン１０２が可変的な指向性を有する場合には、認識制御部１７２は、制御オブジェクトを用いてマイクロフォン１０２の指向性を設定することにより、音声認識の精度を向上させてもよい。例えば、認識制御部１７２は、制御オブジェクトの画面上の位置に応じて、マイクロフォン１０２の指向性を設定してもよい。また、認識制御部１７２は、制御オブジェクトの画面上の向きに応じて、マイクロフォン１０２の指向性を設定してもよい。

【００５９】

図１５～図１７は、マイクロフォンの指向性の制御の一例について説明するための説明図である。図１５の上段において、ＵＩ画像Ｗ１３に、制御オブジェクトＩＣ１が重畳されている。制御オブジェクトＩＣ１の表示位置は、ユーザＵａの手領域の動きに応じて変化し得る。図示された時点において、制御オブジェクトＩＣ１の表示位置は、画面の中央のやや左である。図１５の下段には、ユーザＵａの頭上の視点から見た、情報処理装置１００とユーザＵａとの間の実空間における位置関係が示されている。認識制御部１７２は、例えば、カメラ１０１の画角と制御オブジェクトＩＣ１の表示位置とに基づいて、マイクロフォン１０２の指向性を角度Ｒ１に設定する。ユーザＵａは角度Ｒ１の方向に存在するため、結果として、ユーザＵａが発する音声をマイクロフォン１０２がより高い品質で集音することが可能となる。

【００６０】

図１６の上段において、ＵＩ画像Ｗ１４に、制御オブジェクトＩＣ１が重畳されている。また、ＵＩ画像Ｗ１４には、ユーザＵａ及びＵｂが映っている。図示された時点において、制御オブジェクトＩＣ１の表示位置は、ユーザＵａよりもむしろユーザＵｂの顔の近傍である。図１６の下段には、ユーザＵａ及びＵｂの頭上の視点から見た、情報処理装置１００とユーザＵａ及びＵｂとの間の実空間における位置関係が示されている。認識制御部１７２は、例えば、カメラ１０１の画角と制御オブジェクトＩＣ１の表示位置とに基づいて、マイクロフォン１０２の指向性を角度Ｒ２に設定する。角度Ｒ２の方向にはユーザＵｂが存在するため、結果として、ユーザＵｂが発する音声をマイクロフォン１０２がより高い品質で集音することが可能となる。

【００６１】

図１７の上段において、ＵＩ画像Ｗ１５に、制御オブジェクトＩＣ１が重畳されている。制御オブジェクトＩＣ１の画面上での向きは、ユーザＵａの手領域の向きに応じて変化し得る。ＵＩ画像Ｗ１５には、ユーザＵａ及びＵｂが映っている。図示された時点において、制御オブジェクトＩＣ１は、ユーザＵａより操作され、ユーザＵｂの顔領域Ａ０７の方向に向けられている。図１７の下段には、ユーザＵａ及びＵｂの頭上の視点から見た、

情報処理装置１００とユーザＵａ及びＵｂとの間の実空間における位置関係が示されている。認識制御部１７２は、例えば、制御オブジェクトＩＣ１の表示位置及び向き、並びにユーザＵｂの顔領域Ａ０７の位置に基づいて、マイクロフォン１０２の指向性を角度Ｒ３に設定する。角度Ｒ３の方向にはユーザＵｂが存在するため、結果として、ユーザＵｂが発する音声をマイクロフォン１０２がより高い品質で集音することが可能となる。

【００６２】

図１６又は図１７を用いて説明したような手法によれば、複数のユーザが存在する場合に、制御オブジェクトＩＣ１をあたかも現実のマイクロフォンであるかのように使用して、音声認識についての発話権をユーザ間で受け渡すことが可能となる。

【００６３】

ここまでに説明した例以外にも、ユーザのジェスチャに基づく様々なユーザインタフェースが実現されてよい。例えば、認識制御部１７２は、ユーザが手で口を塞ぐジェスチャの認識に応じて、音声認識部１５４によるそれまでの音声認識結果をキャンセルしてもよい。それにより、ユーザが誤った内容の音声を発し又は音声認識部１５４が音声の内容を誤って認識した場合に、ユーザが簡易に音声入力をやり直すことができる。また、認識制御部１７２は、予め定義されるジェスチャの認識に応じて、スピーカ１０９からの音声出力のボリュームを増加させ又は減少させてもよい。

【００６４】

また、認識制御部１７２は、１つ以上の音声コマンド候補の各々を表すテキストオブジェクトを、入力画像にさらに重畳してもよい。それにより、ユーザは、アプリケーション機能が受け付ける音声コマンドを事前に知っていなくても、必要とされる音声コマンドを適切に発することができる。

【００６５】

（８）表示制御部１７４

表示制御部１７４は、ディスプレイ１０８を介する画像の表示を制御する。例えば、表示制御部１７４は、アプリケーション部１４０から入力されるアプリケーション画像をディスプレイ１０８に表示させる。また、表示制御部１７４は、音声対応アプリケーションが起動された場合に、認識制御部１７２により生成されるＵＩ画像を、ディスプレイ１０８に表示させる。表示制御部１７４は、ＵＩ画像のみをディスプレイ１０８に表示させてもよく、又はアプリケーション画像及びＵＩ画像を合成することにより生成される１つの出力画像をディスプレイ１０８に表示させてもよい。

【００６６】

図１８及び図１９は、本実施形態において採用され得る出力画像のウィンドウ構成の例をそれぞれ示している。これら図において、ＵＩ用ウィンドウW_{UI}及びアプリケーション用ウィンドウW_{APP}がディスプレイ１０８により表示される。ＵＩ用ウィンドウW_{UI}は、認識制御部１７２により生成されるＵＩ画像を表示する。アプリケーション用ウィンドウW_{APP}は、アプリケーション部１４０から入力されるアプリケーション画像（例えば、コンテンツ画像）を表示する。図１８の第１の例では、アプリケーション用ウィンドウW_{APP}は、ＵＩ用ウィンドウW_{UI}の右下のコーナーに合成されている。図１９の第２の例では、ＵＩ用ウィンドウW_{UI}はアプリケーション用ウィンドウW_{APP}の一部分にブレンディングされている。こうしたウィンドウ構成によれば、ユーザは、例えばコンテンツ画像を閲覧しながら、リモートコントローラが手元になくても、制御オブジェクトを用いて情報処理装置１００を自らの音声で操作することができる。

【００６７】

［２－３．制御シナリオの例］

上述した情報処理装置１００において行われ得るいくつかの制御シナリオの例について、図２０～図２３を用いて説明する

【００６８】

（１）第１のシナリオ

図２０は、第１の制御シナリオについて説明するための説明図である。図２０を参照す

ると、５つのＵＩ画像ＳＴ１１～ＳＴ１５が時間軸に沿って示されている。
【００６９】
ＵＩ画像ＳＴ１１にはユーザＵｄが映っており、ミラー表示が実現されている。
【００７０】
次のＵＩ画像ＳＴ１２は、例えば音声対応アプリケーションが起動し、又はユーザが手を挙げるなどのジェスチャをした後に表示され得る。ＵＩ画像ＳＴ１２には、制御オブジェクトＩＣ１が重畳されている。但し、この時点では、音声認識部１５４への音声入力はアクティブ化されていない。
【００７１】
次のＵＩ画像ＳＴ１３は、例えばユーザＵｄが制御オブジェクトＩＣ１を口の近傍に移動させた後に表示され得る。認識制御部１７２は、アクティブ化条件が満たされた結果として、音声認識部１５４への音声入力をアクティブ化する。ＵＩ画像ＳＴ１３において、制御オブジェクトＩＣ１の表示属性は、アクティブ状態を示すように変化している。
【００７２】
次のＵＩ画像ＳＴ１４は、ユーザＵｄが音声を発している間に表示され得る。ＵＩ画像ＳＴ１４において、制御オブジェクトＩＣ１の表示属性は、引き続きアクティブ状態を示している。また、ＵＩ画像ＳＴ１４にはエフェクトＦｂ１が適用されると共に、認識された音声の内容を示す追加オブジェクトＦｂ２がＵＩ画像ＳＴ１４に重畳されている。
【００７３】
次のＵＩ画像ＳＴ１５は、非アクティブ化条件が満たされた場合に表示され得る。ここでは、音声入力を非アクティブ化させるジェスチャとして人差し指を口に当てるジェスチャが定義されているものとする。認識制御部１７２は、当該ジェスチャの認識に応じて、音声認識部１５４への音声入力を非アクティブ化する。制御オブジェクトＩＣ１の表示位置は例えば既定の表示位置に戻され、制御オブジェクトＩＣ１の表示属性は非アクティブ状態を示すように変更される。
【００７４】
（２）第２のシナリオ
図２１は、第２の制御シナリオについて説明するための説明図である。図２１を参照すると、５つのＵＩ画像ＳＴ２１～ＳＴ２５が時間軸に沿って示されている。
【００７５】
ＵＩ画像ＳＴ２１には、ユーザＵｄが映っている。また、ＵＩ画像ＳＴ２１に制御オブジェクトＩＣ１が重畳されている。但し、この時点では、音声認識部１５４への音声入力はアクティブ化されていない。
【００７６】
次のＵＩ画像ＳＴ２２は、例えばユーザＵｄが制御オブジェクトＩＣ１を口の近傍に移動させた後に表示され得る。認識制御部１７２は、アクティブ化条件が満たされた結果として、音声認識部１５４への音声入力をアクティブ化する。ＵＩ画像ＳＴ２２において、制御オブジェクトＩＣ１の表示属性は、アクティブ状態を示すように変化している。
【００７７】
次のＵＩ画像ＳＴ２３は、ユーザＵｄが音声を発している間に表示され得る。ＵＩ画像ＳＴ２３において、制御オブジェクトＩＣ１の表示属性は、引き続きアクティブ状態を示している。第２の制御シナリオでは、ユーザＵｄが音声を発している間、手の動きに関わらず、制御オブジェクトＩＣ１の表示位置は、ユーザＵｄの口の近傍に維持される。従って、ユーザは、例えば電子メールのメッセージのように長い文章を音声で入力するような場合に、手を挙げ続けることで疲れることなく、音声入力を継続することができる。
【００７８】
次のＵＩ画像ＳＴ２４において、ユーザＵｄは、手で口を塞ぐジェスチャをしている。認識制御部１７２は、かかるジェスチャの認識に応じて、それまでの音声認識結果をキャンセルする。第２の制御シナリオにおいて、音声認識部１５４への音声入力のアクティブ状態は、その後も維持される。

【００７９】
次のＵＩ画像ＳＴ２５において、ユーザＵｄは再び音声を発している。その結果、当初ユーザＵｄが発した音声の内容とは異なる内容の音声が、音声認識部１５４により適切に認識されている。
【００８０】
（３）第３のシナリオ
図２２は、第３の制御シナリオについて説明するための説明図である。図２２を参照すると、３つのＵＩ画像ＳＴ３１～ＳＴ３３が時間軸に沿って示されている。
【００８１】
ＵＩ画像ＳＴ３１にはユーザＵｄが映っており、ミラー表示が実現されている。
【００８２】
次のＵＩ画像ＳＴ３２は、例えばユーザが手を挙げるなどのジェスチャをした後に表示され得る。ＵＩ画像ＳＴ３２には、制御オブジェクトＩＣ２が重畳されている。また、ＵＩ画像ＳＴ３２には、音声対応アプリケーションが受け付ける音声コマンド候補（コマンドＡ～コマンドＤ）の各々を表す４つのテキストオブジェクトが重畳されている。
【００８３】
次のＵＩ画像ＳＴ３３において、例えばユーザＵｄが制御オブジェクトＩＣ２の近傍に近付いた結果として音声入力がアクティブ化されている。そして、ユーザＵｄがコマンドＢを読み上げる音声を発し、発せられたコマンドＢを音声認識部１５４が適切に認識している。音声コマンド候補は、例えば、情報処理装置１００をユーザが遠隔的に制御するために予め用意される１つ以上のコマンドであってよい。
【００８４】
このように、本実施形態では、ユーザの手元にリモートコントローラがなくても、ユーザが情報処理装置１００を遠隔的に制御することが可能である。例えば、リモートコントローラが紛失した状況、又は他のユーザによりリモートコントローラが保持されている状況でも、ユーザは、ストレスを感じることなく、所望のタイミングで情報処理装置１００を制御することができる。なお、ＵＩ画像ＳＴ３２が表示された後、所定の音声コマンド又はジェスチャの認識に応じて、音声コマンドＡ～Ｄを表すテキストオブジェクトが他の音声コマンド候補を表すテキストオブジェクトに置き換えられてもよい。
【００８５】
（４）第４のシナリオ
第４のシナリオは、制御オブジェクトが介在しない補足的なシナリオである。図２３は、第４の制御シナリオについて説明するための説明図である。図２３を参照すると、３つのＵＩ画像ＳＴ４１～ＳＴ４３が時間軸に沿って示されている。
【００８６】
ＵＩ画像ＳＴ４１にはユーザＵｄが映っており、ミラー表示が実現されている。
【００８７】
次のＵＩ画像ＳＴ４２において、ユーザＵｄは、耳元で手を丸めるジェスチャをしている。認識制御部１７２は、かかるジェスチャの認識に応じて、スピーカ１０９からの音声出力のボリュームを増加させる。ボリュームの増加量は、ジェスチャが認識されている時間の長さに依存して変化してもよい。
【００８８】
次のＵＩ画像ＳＴ４３において、ユーザＵｄは、人差し指を口に当てるジェスチャをしている。認識制御部１７２は、かかるジェスチャの認識に応じて、スピーカ１０９からの音声出力のボリュームを減少させる。ボリュームの減少量は、ジェスチャが認識されている時間の長さに依存して変化してもよい。
【００８９】
このように、本実施形態では、ユーザのジェスチャに基づく様々なユーザインタフェースが実現され得る。音声入力がアクティブか否か、又は音声対応アプリケーションが実行中であるか否かに依存して、同じ種類のジェスチャが互いに異なる意味に解釈されてもよ

い。なお、ユーザ独自のジェスチャをユーザに登録させるためのユーザインタフェースが提供されてもよい。例えば、“手で（制御オブジェクトを）払いのける”というジェスチャが登録され、当該ジェスチャが音声入力のアクティブ化／非アクティブ化のためのジェスチャとして定義されてもよい。個々のジェスチャのための動き、及びジェスチャと対応する処理との間のマッピングをユーザにカスタマイズさせるためのユーザインタフェースがさらに提供されてもよい。

【００９０】

［２－４．処理の流れの例］

図２４及び図２５のフローチャートは、本実施形態に係る情報処理装置１００により実行され得る処理の流れの一例を示している。ここで説明する処理は、カメラ１０１により撮像される動画を構成する一連のフレームの各々について繰り返される。

【００９１】

図２４を参照すると、まず、画像取得部１２０は、カメラ１０１により撮像される画像を入力画像として取得する（ステップＳ１００）。そして、画像取得部１２０は、取得した入力画像を認識部１５０及び制御部１７０へ出力する。

【００９２】

次に、画像認識部１５２は、画像取得部１２０から入力される入力画像に映るユーザの身体を認識する（ステップＳ１０５）。例えば、画像認識部１５２は、入力画像内のユーザの手領域及び口領域を認識し、認識したこれら領域の位置を示す位置データを制御部１７０へ出力する。また、画像認識部１５２は、予め定義されるいくつかのユーザのジェスチャを追加的に認識してもよい。

【００９３】

次に、認識制御部１７２は、音声対応アプリケーションが起動しているかを判定する（ステップＳ１１０）。音声対応アプリケーションが起動していない場合には、その後のステップＳ１１５～ステップＳ１６０の処理はスキップされる。音声対応アプリケーションが起動している場合（又はステップＳ１０５で認識されるジェスチャによって、音声対応アプリケーションが起動された場合）には、処理はステップＳ１１５へ進む。

【００９４】

ステップＳ１１５において、認識制御部１７２は、発話に関連する制御オブジェクトの表示位置及び向きを決定する（ステップＳ１１５）。制御オブジェクトの表示位置は、既定の位置であってもよく、又は画像認識部１５２により認識されるユーザの手の動きに追随して移動してもよい。同様に、制御オブジェクトの向きは、既定の向きであってもよく、又はユーザの手の動きに追随して回転してもよい。

【００９５】

次に、マイクロフォン１０２が可変的な指向性を有する場合には、認識制御部１７２は、ステップＳ１１５において決定した制御オブジェクトの表示位置及び向きに応じて、マイクロフォン１０２の指向性を設定する（ステップＳ１２０）。

【００９６】

次に、認識制御部１７２は、入力画像をミラー表示するＵＩ画像に、ステップＳ１１５において決定した表示位置及び向きを有する制御オブジェクトを重畳する（ステップＳ１２５）。ここでの制御オブジェクトの表示属性は、音声入力がアクティブ化されていないことを示す値に設定され得る。

【００９７】

図２５に移り、次に、認識制御部１７２は、上述したアクティブ化条件及び非アクティブ化条件に従って、音声入力がアクティブであるかを判定する（ステップＳ１３０）。例えば、ユーザの口領域と制御オブジェクトとの間の距離が距離閾値を下回る場合には、アクティブ化条件は満たされていると判定され得る。音声入力がアクティブであると判定されない場合には、その後のステップＳ１３５～ステップＳ１６０の処理はスキップされる。音声入力がアクティブであると判定された場合には、処理はステップＳ１３５へ進む。

【００９８】

ステップＳ１３５において、認識制御部１７２は、音声認識部１５４への音声入力を必要に応じてアクティブ化し、制御オブジェクトの表示属性を、音声入力がアクティブ化されていることを示す値に設定する（ステップＳ１３５）。

【００９９】

次に、音声取得部１３０は、マイクロフォン１０２から取得される入力音声を、音声認識部１５４へ出力する（ステップＳ１４０）。

【０１００】

次に、音声認識部１５４は、音声取得部１３０から入力される入力音声に基づいて、ユーザの音声を認識する（ステップＳ１４５）。そして、音声認識部１５４は、音声認識の結果を、アプリケーション部１４０及び認識制御部１７２へ出力する。

【０１０１】

次に、認識制御部１７２は、音声認識部１５４から入力される音声認識結果についてのフィードバックを、ＵＩ画像に適用する（ステップＳ１５０）。例えば、認識制御部１７２は、図１１に例示したエフェクトＦｂ１をＵＩ画像に適用してもよい。また、認識制御部１７２は、図１２～図１４に例示した追加オブジェクトＦｂ２又はＦｂ３をＵＩ画像に重畳してもよい。

【０１０２】

次に、認識制御部１７２は、音声認識が成功したか否かを判定する（ステップＳ１５５）。音声認識が成功していなければ、その後のステップＳ１６０の処理はスキップされる。音声認識が成功していれば、処理はステップＳ１６０へ進む。

【０１０３】

ステップＳ１６０において、アプリケーション部１４０は、音声認識結果に基づくアプリケーション処理を実行する（ステップＳ１６０）。例えば、アプリケーション部１４０は、認識された音声コマンドに対応する処理を実行してもよい。また、アプリケーション部１４０は、認識された音声の内容を示すテキストを入力情報として受け付けてもよい。

【０１０４】

次に、表示制御部１７４は、ＵＩ画像を含む出力画像をディスプレイ１０８に表示させる（ステップＳ１６５）。ここで表示される出力画像は、ＵＩ画像のみを含んでもよく、又はＵＩ画像及びアプリケーション画像の双方を含んでもよい。その後、処理は図２４のステップＳ１００へ戻る。

【０１０５】

なお、ここまで、主にＵＩ画像に１つの制御オブジェクトのみが重畳される例を説明した。しかしながら、かかる例に限定されず、ＵＩ画像に複数の制御オブジェクトが重畳されてもよい。例えば、入力画像に複数のユーザが映っている場合において、それぞれのユーザについて別個の制御オブジェクトを重畳すれば、制御オブジェクトをユーザ間で受け渡す作業を要することなく、各ユーザが所望のタイミングで音声コマンドを入力することが可能となる。

【０１０６】

＜３．第２の実施形態＞

上述したように、本開示に係る技術は、テレビジョン装置に限定されず、様々な種類の装置に適用可能である。そこで、第２の実施形態として、本開示に係る技術がメッセージ交換用アプリケーションを有する情報処理装置２００に適用される例について説明する。図２を用いて説明したように、情報処理装置２００は、タブレットＰＣである。

【０１０７】

（１）ハードウェア構成例

図２６は、情報処理装置２００のハードウェア構成の一例を示すブロック図である。図２６を参照すると、情報処理装置２００は、カメラ２０１、マイクロフォン２０２、入力デバイス２０３、通信Ｉ／Ｆ２０４、メモリ２０５、ディスプレイ２０８、スピーカ２０９、バス２１１及びプロセッサ２１２を備える。

【０１０８】

カメラ２０１は、ＣＣＤ又はＣＭＯＳなどの撮像素子を有し、画像を撮像する。カメラ２０１により撮像される画像（動画を構成する各フレーム）は、情報処理装置２００による処理のための入力画像として扱われる。

【０１０９】

マイクロフォン２０２は、ユーザにより発せられる音声を集音し、音声信号を生成する。マイクロフォン２０２により生成される音声信号は、情報処理装置２００による音声認識のための入力音声として扱われる。

【０１１０】

入力デバイス２０３は、ユーザが情報処理装置２００を操作し又は情報処理装置２００へ情報を入力するために使用されるデバイスである。入力デバイス２０３は、例えば、タッチパネル、ボタン及びスイッチなどを含み得る。入力デバイス２０３は、ユーザ入力を検出すると、検出されたユーザ入力に対応する入力信号を生成する。

【０１１１】

通信Ｉ／Ｆ２０４は、情報処理装置２００による他の装置との間の通信を仲介する。通信Ｉ／Ｆ２０４は、任意の無線通信プロトコル又は有線通信プロトコルをサポートし、他の装置との間の通信接続を確立する。

【０１１２】

メモリ２０５は、半導体メモリ又はハードディスクなどの記憶媒体により構成され、情報処理装置２００による処理のためのプログラム及びデータ、並びにコンテンツデータを記憶する。なお、プログラム及びデータの一部又は全部は、メモリ２０５により記憶されることなく、外部のデータソース（例えば、データサーバ、ネットワークストレージ又は外付けメモリなど）から取得されてもよい。

【０１１３】

ディスプレイ２０８は、ＬＣＤ又はＯＬＥＤなどにより構成される画面を有し、情報処理装置２００により生成される画像を表示する。例えば、第１の実施形態において説明したものと同様のＵＩ画像が、ディスプレイ２０８の画面に表示され得る。

【０１１４】

スピーカ２０９は、振動板及びアンプなどの回路素子を有し、情報処理装置２００により生成される出力音声信号に基づいて、音声を出力する。スピーカ２０９の音量は、変更可能である。

【０１１５】

バス２１１は、カメラ２０１、マイクロフォン２０２、入力デバイス２０３、通信Ｉ／Ｆ２０４、メモリ２０５、ディスプレイ２０８、スピーカ２０９及びプロセッサ２１２を相互に接続する。

【０１１６】

プロセッサ２１２は、例えば、ＣＰＵ又はＤＳＰなどであってよい。プロセッサ２１２は、メモリ２０５又は他の記憶媒体に記憶されるプログラムを実行することにより、第１の実施形態に係る情報処理装置１００のプロセッサ１１２と同様に、情報処理装置２００の様々な機能を動作させる。情報処理装置２００のメモリ２０５及びプロセッサ２１２により実現される論理的機能の構成は、アプリケーション機能が異なることを除き、図４に例示した情報処理装置１００の構成と同様であってよい。

【０１１７】

（２）制御シナリオの例

図２７は、第２の実施形態における制御シナリオの一例について説明するための説明図である。図２７を参照すると、４つの出力画像ＳＴ５１～ＳＴ５４が時間軸に沿って示されている。本シナリオにおいて、各出力画像は、上部のメッセージ交換用アプリケーションのアプリケーション画像と、下部のＵＩ画像とにより構成される。

【０１１８】

出力画像ＳＴ５１において、アプリケーション画像は、メッセージ入力ボックスを含む。メッセージ入力ボックスには、メッセージは入力されていない。ＵＩ画像にはユーザＵ

ｄが映っており、ミラー表示が実現されている。

【０１１９】

次の出力画像ＳＴ５２は、例えばユーザが手を挙げるなどのジェスチャをした後に表示され得る。出力画像ＳＴ５２において、ＵＩ画像に制御オブジェクトＩＣ１が重畳されている。但し、この時点では、音声入力はアクティブ化されていない。

【０１２０】

次の出力画像ＳＴ５３は、例えばユーザＵｄが制御オブジェクトＩＣ１を口の近傍に移動させた後に表示され得る。音声入力はアクティブ化され、制御オブジェクトＩＣ１の表示属性は、アクティブ状態を示すように変化している。メッセージ入力ボックスには、ユーザにより発せられた音声の内容が入力されている。

【０１２１】

次の出力画像ＳＴ５４は、例えばユーザＵｄが制御オブジェクトＩＣ１を口の近傍から離した後に表示され得る。音声入力は非アクティブ化され、制御オブジェクトＩＣ１の表示属性は、非アクティブ状態を示すように変化している。この状態でユーザが音声を発しても、メッセージ入力ボックスには音声の内容は入力されない。従って、ユーザは、手を動かす簡単な動作だけで、音声入力の状態を切り替えて、入力することを望む音声の内容だけをメッセージに含めることができる。

【０１２２】

＜４．まとめ＞

ここまで、図１～図２７を用いて、本開示に係る技術の実施形態について詳細に説明した。上述した実施形態によれば、入力画像に重畳して表示される制御オブジェクトを用いて、情報機器により実行される音声認識が制御される。従って、ユーザは、画面上の制御オブジェクトの状態を手掛かりとして、音声認識のための適切なタイミングを判断することができる。

【０１２３】

また、上述した実施形態によれば、入力画像内で認識されるユーザの身体の所定の部分と制御オブジェクトとの間の位置関係に基づいて、音声認識が制御される。従って、ユーザは、画面に表示される自らの身体を動かすことにより、音声認識に関連する様々な機能性を扱うことができる。

【０１２４】

また、上述した実施形態によれば、ユーザの口と制御オブジェクトとの間の距離に基づいて、音声認識のための音声入力がアクティブ化され得る。また、制御オブジェクトは、ユーザの手の動きに従って画面内で移動し得る。従って、ユーザは、制御オブジェクトを移動させ又は自ら制御オブジェクトの方へ移動することにより、意図したタイミングで所望の音声のみを容易に認識させることができる。その際にユーザに求められる動きは、現実のマイクロフォンを扱う動きに類似しているため、こうした仕組みによって、ユーザにとって直感的なユーザインタフェースを実現することができる。

【０１２５】

また、上述した実施形態によれば、音声入力がアクティブ化されているか否かが、制御オブジェクトの表示属性の変化を通じてユーザに通知される。従って、ユーザは、画面上の制御オブジェクトのみに注意を払うだけで、適切なタイミングで発話することができる。

【０１２６】

なお、本明細書において説明した各装置による一連の処理は、典型的には、ソフトウェアを用いて実現される。一連の処理を実現するソフトウェアを構成するプログラムは、例えば、各装置の内部又は外部に設けられる記憶媒体（非一時的な媒体：non-transitory media）に予め格納される。そして、各プログラムは、例えば、実行時にＲＡＭ（Random Access Memory）に読み込まれ、ＣＰＵなどのプロセッサにより実行される。

【０１２７】

以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本

開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、特許請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

【０１２８】

なお、以下のような構成も本開示の技術的範囲に属する。

（１）

入力画像を取得する画像取得部と、

発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させる制御部と、

を備え、

前記制御部は、ユーザの音声について実行される音声認識を、前記オブジェクトを用いて制御する、

情報処理装置。

（２）

前記情報処理装置は、前記入力画像に映るユーザの身体を認識する画像認識部、をさらに備え、

前記制御部は、前記画像認識部により認識されるユーザの身体の所定の部分と前記オブジェクトとの間の前記画面内の位置関係に基づいて、前記音声認識を制御する、

前記（１）に記載の情報処理装置。

（３）

前記所定の部分は、ユーザの口を含み、

前記制御部は、ユーザの口と前記オブジェクトとの間の距離に基づいて、前記音声認識のための音声入力をアクティブ化する、

前記（２）に記載の情報処理装置。

（４）

前記所定の部分は、ユーザの手を含み、

前記制御部は、ユーザの手の動きに従って前記オブジェクトを前記画面内で移動させる、

前記（３）に記載の情報処理装置。

（５）

前記制御部は、前記入力画像に映るユーザのジェスチャに応じて、前記音声認識のための音声入力を非アクティブ化する、前記（３）又は前記（４）に記載の情報処理装置。

（６）

前記制御部は、前記音声認識のための音声入力がアクティブ化されているか否かを、前記オブジェクトの表示属性を変化させることによりユーザに通知する、前記（１）～（５）のいずれか１項に記載の情報処理装置。

（７）

前記制御部は、前記音声認識において音声が検出されているか否かを、前記オブジェクトの表示属性を変化させ又は前記オブジェクトが重畳された出力画像の状態を変化させることにより、ユーザに通知する、前記（１）～（６）のいずれか１項に記載の情報処理装置。

（８）

前記制御部は、前記音声認識において検出されている音声のレベルに応じて、前記オブジェクトの前記表示属性又は前記出力画像の前記状態の変化のレベルを変化させる、前記（７）に記載の情報処理装置。

（９）

前記音声認識は、可変的な指向性を有するマイクロフォンにより取得される音声信号を用いて実行される、前記（１）～（８）のいずれか１項に記載の情報処理装置。

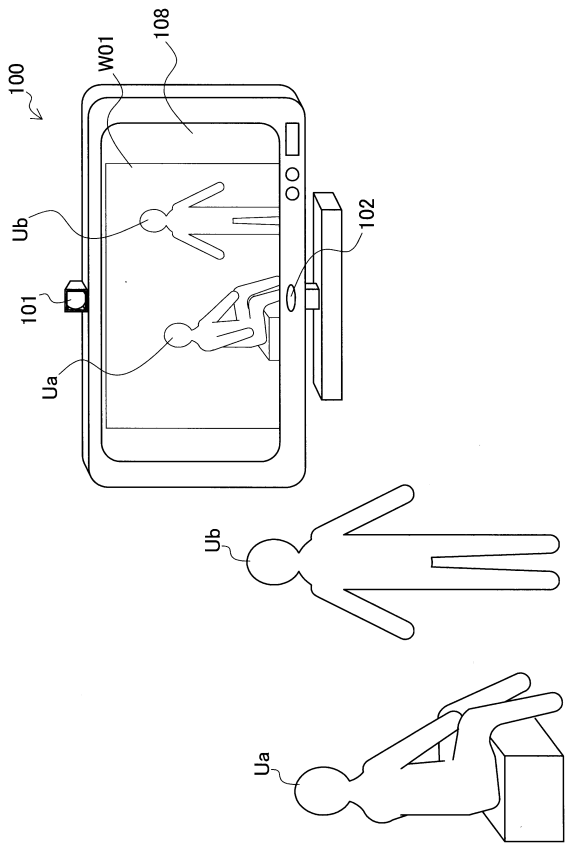
（１０）

前記制御部は、前記オブジェクトの位置をユーザの動きに応じて変化させ、

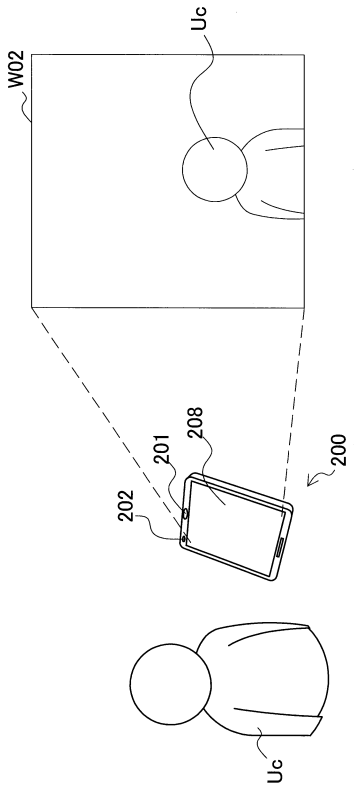
前記マイクロフォンの指向性は、前記オブジェクトの位置に応じて設定される、
前記（９）に記載の情報処理装置。
（１１）
前記制御部は、前記オブジェクトの向きをユーザの動きに応じて変化させ、
前記マイクロフォンの指向性は、前記オブジェクトの向きに応じて設定される、
前記（９）又は前記（１０）に記載の情報処理装置。
（１２）
前記制御部は、前記音声認識において認識された音声の内容を表すテキストを含む第１の追加的なオブジェクトを、前記入力画像に映るユーザの近傍にさらに重畳する、前記（１）～（１１）のいずれか１項に記載の情報処理装置。
（１３）
前記制御部は、前記音声認識が失敗した場合に、前記第１の追加的なオブジェクトの表示属性を変化させ又は特別な文字列を前記テキストに挿入することにより、前記音声認識の失敗をユーザに通知する、前記（１２）に記載の情報処理装置。
（１４）
前記制御部は、前記音声認識において検出されている音声のレベルと、前記音声認識を有効に行うために求められる音声のレベルとを示す第２の追加的なオブジェクトを、前記入力画像にさらに重畳する、前記（１）～（１３）のいずれか１項に記載の情報処理装置。
（１５）
前記制御部は、１つ以上の音声コマンドの候補の各々を表すテキストオブジェクトを、前記入力画像にさらに重畳する、前記（１）～（１４）のいずれか１項に記載の情報処理装置。
（１６）
前記情報処理装置は、テレビジョン装置であり、
前記音声コマンドは、前記情報処理装置をユーザが遠隔的に制御するために発せられるコマンドである、
前記（１５）に記載の情報処理装置。
（１７）
前記オブジェクトは、マイクロフォンを模したアイコンである、前記（１）～（１６）のいずれか１項に記載の情報処理装置。
（１８）
情報処理装置により実行される情報処理方法であって、
入力画像を取得することと、
発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させることと、
ユーザの音声について実行される音声認識を、前記オブジェクトを用いて制御することと、
を含む情報処理方法。
（１９）
情報処理装置を制御するコンピュータを、
入力画像を取得する画像取得部と、
発話に関連するオブジェクトを前記入力画像に重畳して画面に表示させる制御部と、
として機能させ、
前記制御部は、ユーザの音声について実行される音声認識を、前記オブジェクトを用いて制御する、
プログラム。
【符号の説明】
【０１２９】
１００，２００　情報処理装置
１２０　画像取得部

１５２　　画像認識部
１５４　　音声認識部
１７２　　認識制御部
１７４　　表示制御部
ＩＣ１，ＩＣ２　制御オブジェクト

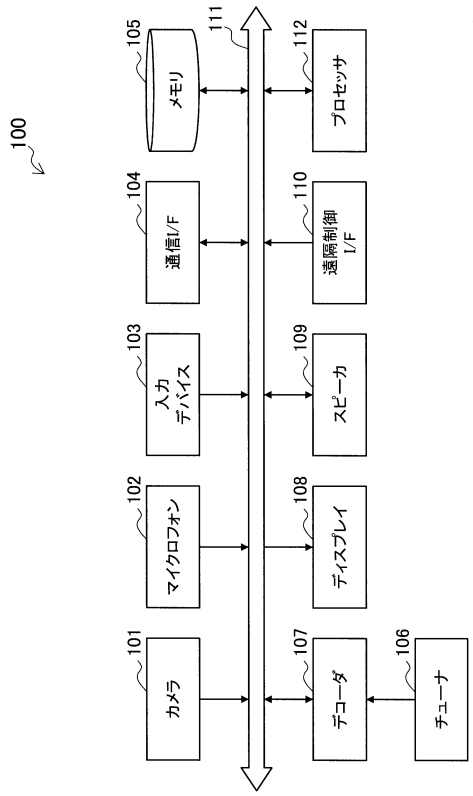
【図１】



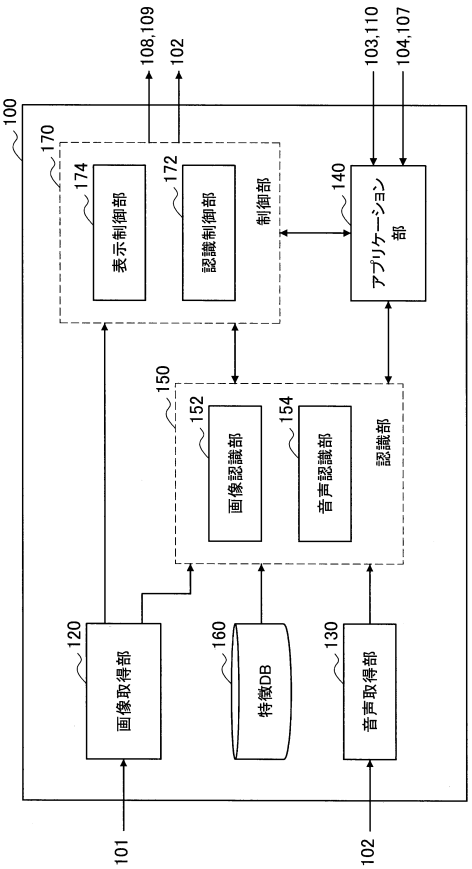
【図２】



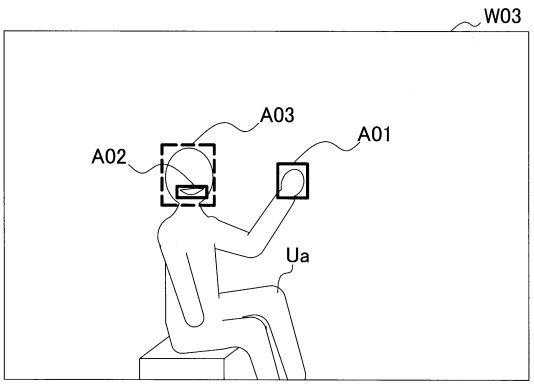
【図３】


100
105
メモリ
111
112
プロセッサ
104
通信I/F
110
遠隔制御I/F
103
入力デバイス
109
スピーカ
102
マイクロフォン
108
ディスプレイ
101
カメラ
107
デコーダ
106
チューナ

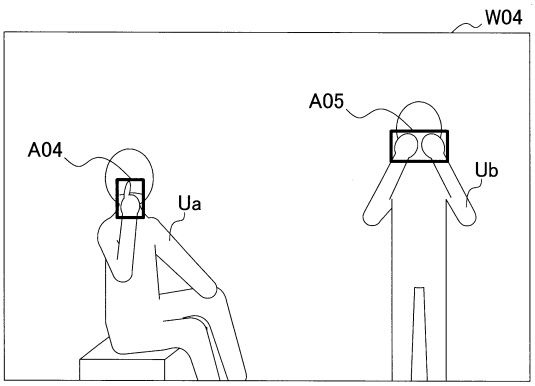
【図４】


100
108,109
102
103,110
104,107
170
174
表示制御部
172
認識制御部
制御部
140
アプリケーション部
150
152
画像認識部
154
音声認識部
認識部
120
画像取得部
160
特徴DB
130
音声取得部
101
102

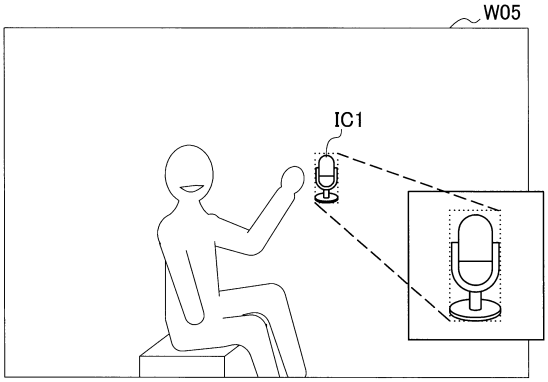
【図５】


W03
A03
A01
A02
Ua

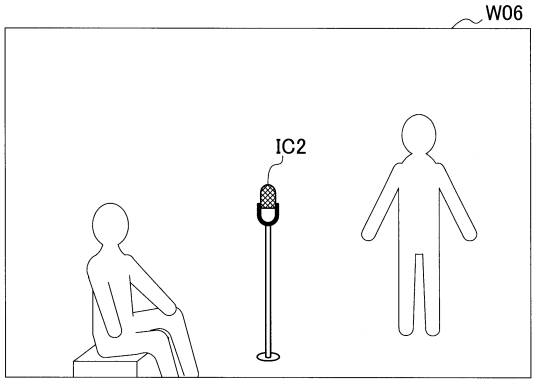
【図６】


W04
A05
A04
Ub
Ua

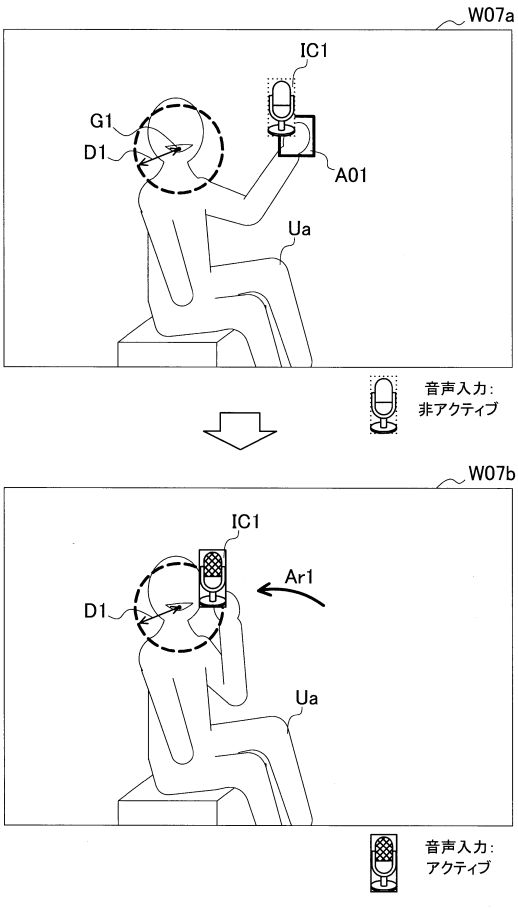
【図７】


W05
IC1

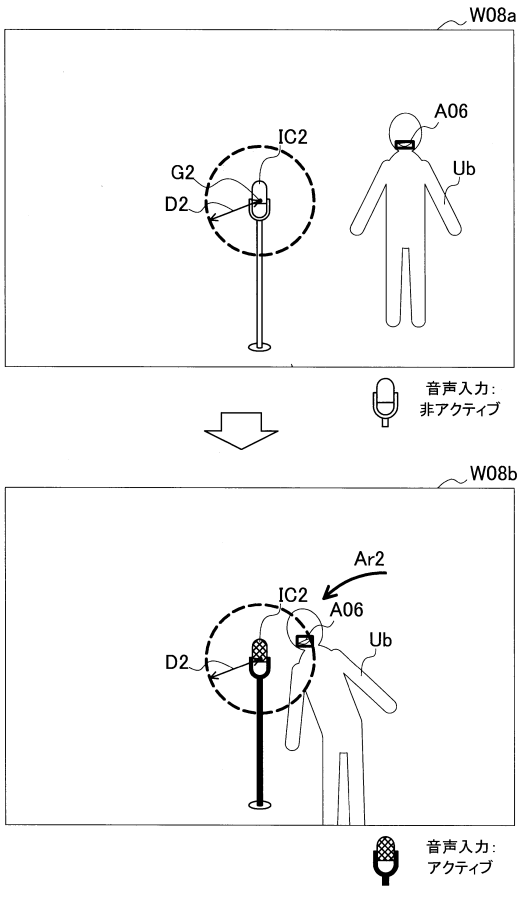
【図８】


W06
IC2

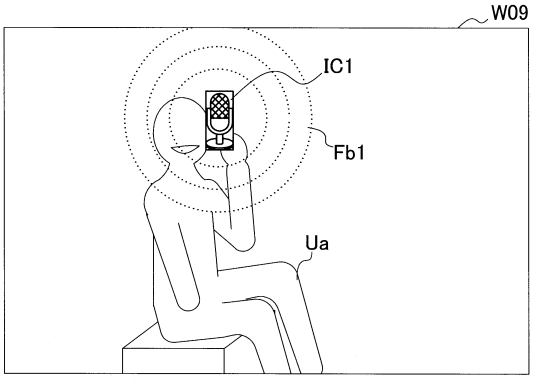
【図９】


W07a
IC1
G1
D1
A01
Ua
音声入力:
非アクティブ
W07b
IC1
Ar1
D1
Ua
音声入力:
アクティブ

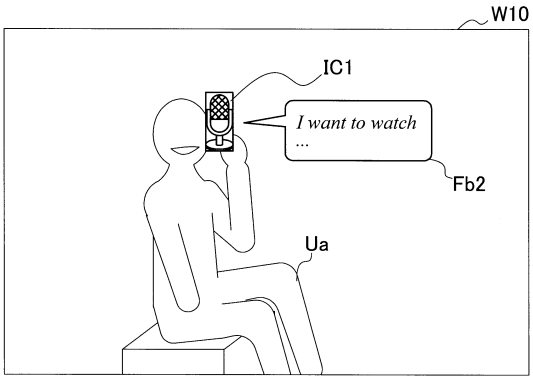
【図１０】


W08a
A06
IC2
G2
D2
Ub
音声入力:
非アクティブ
W08b
Ar2
IC2
A06
Ub
D2
音声入力:
アクティブ

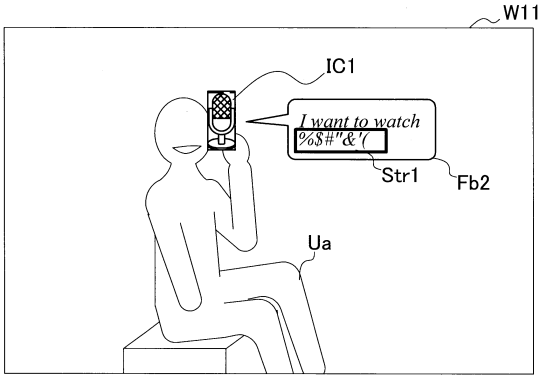
【図１１】



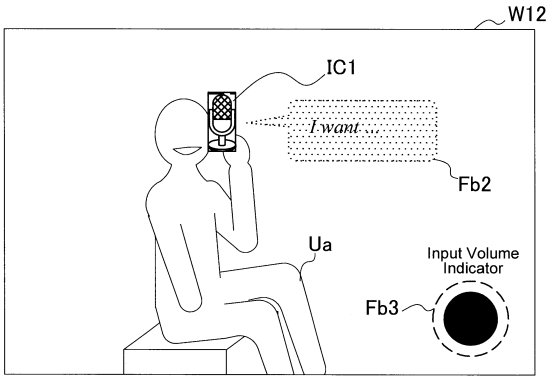
【図１２】



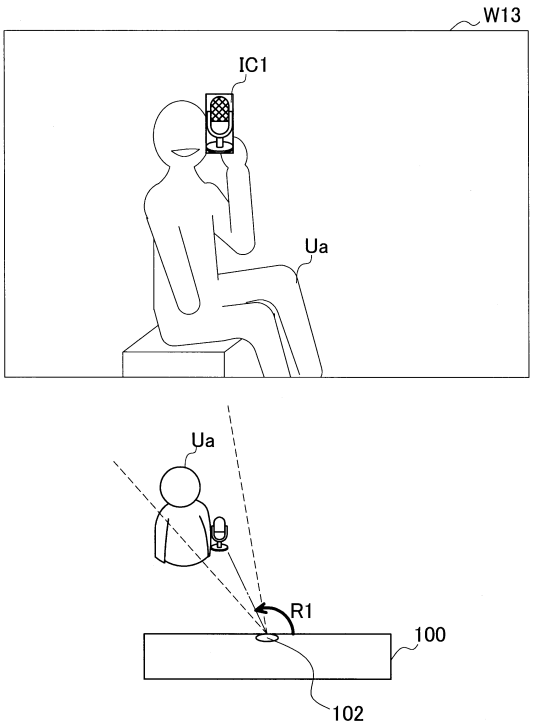
【図１３】



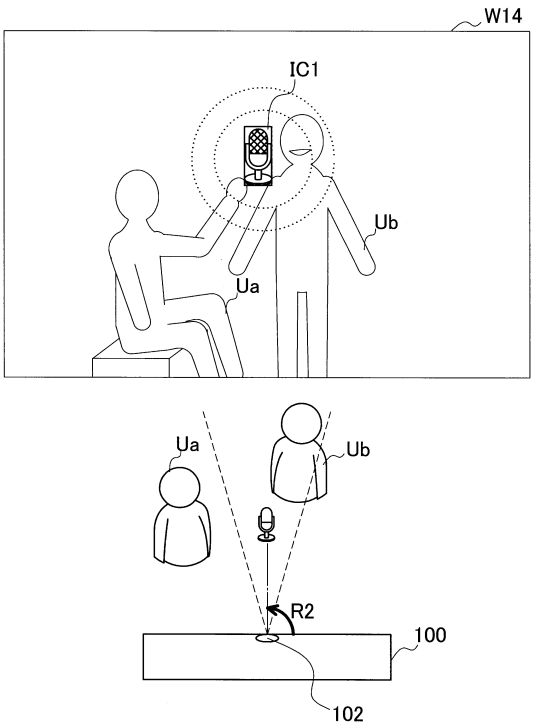
【図１４】



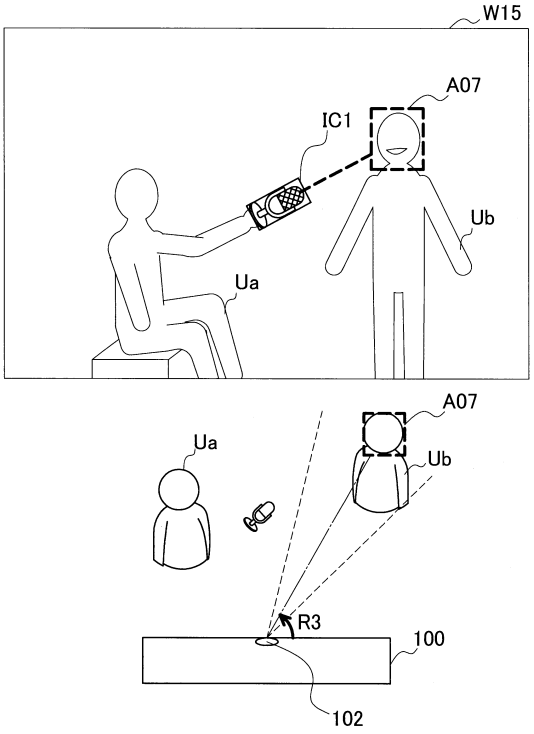
【図１５】



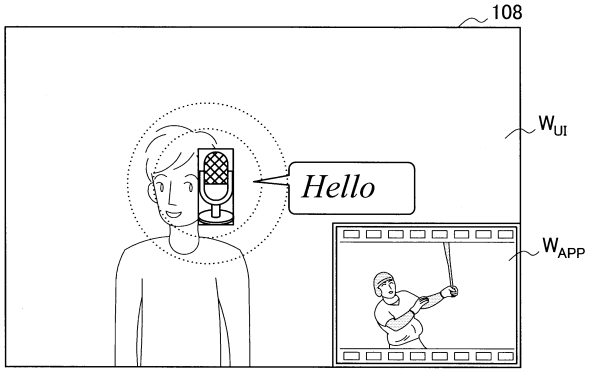
【図１６】



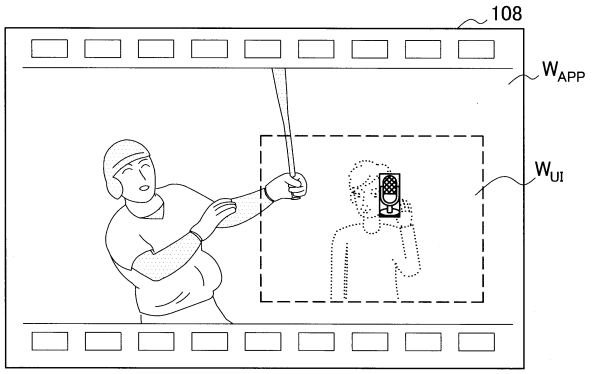
【図１７】



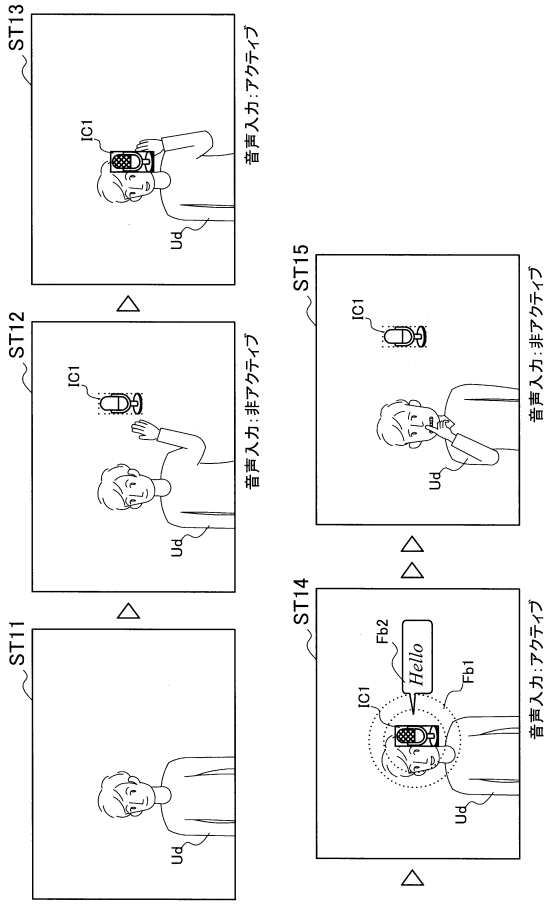
【図１８】



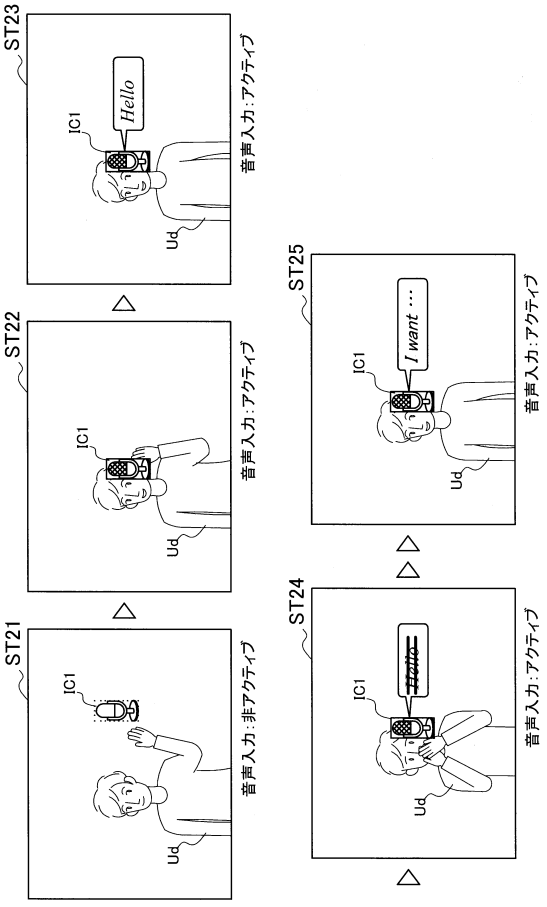
【図１９】



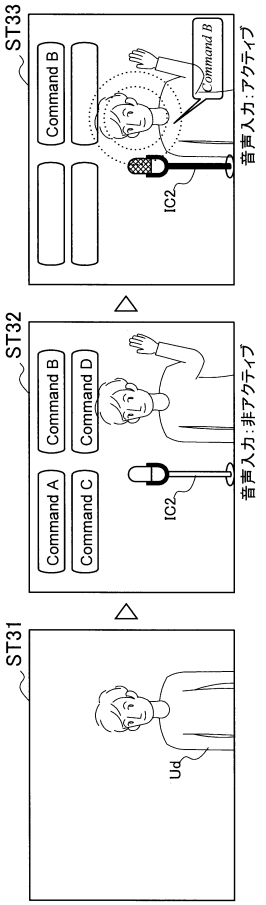
【図２０】



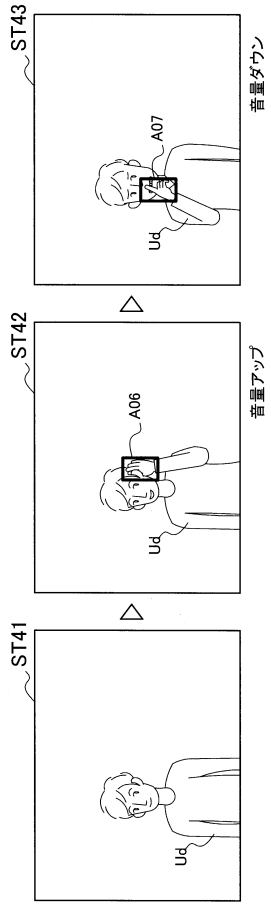
【図２１】



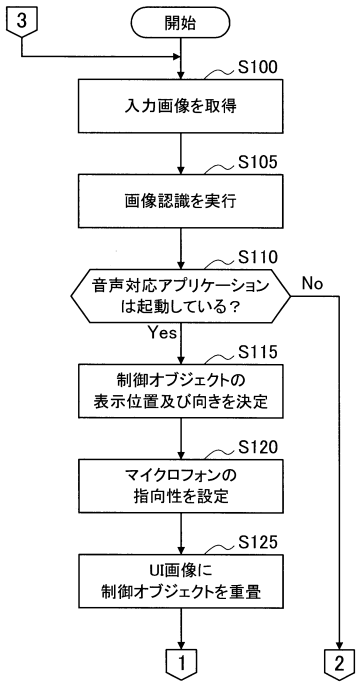
【図２２】



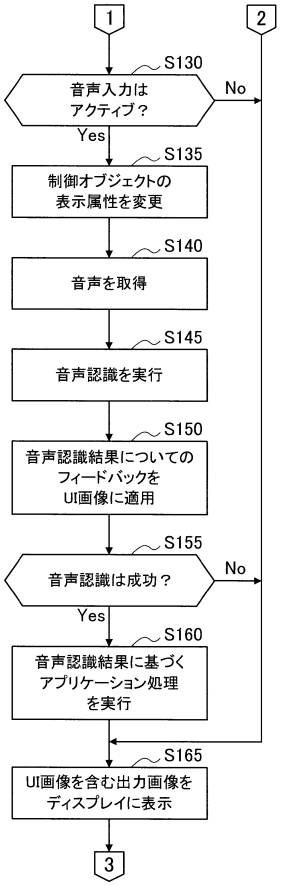
【図２３】



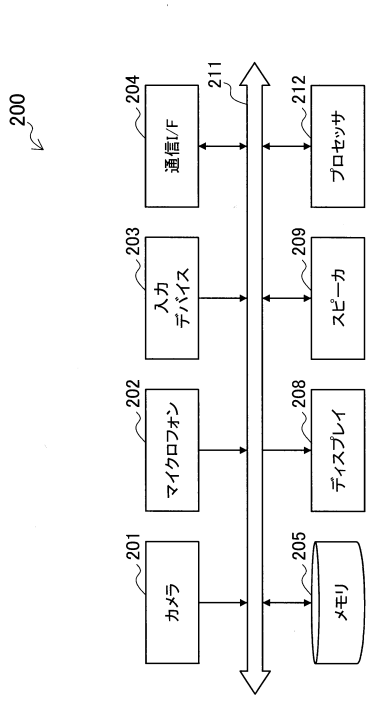
【図２４】



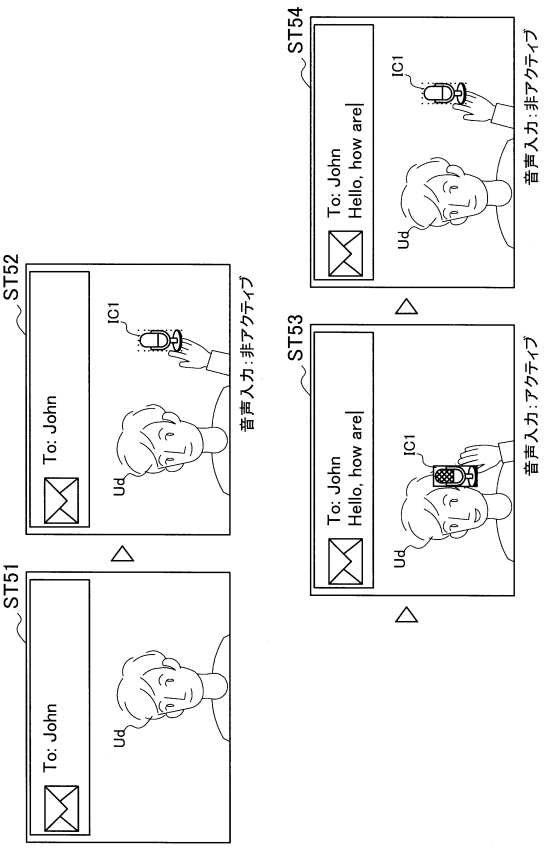
【図２５】



【図２６】


200
201
カメラ
202
マイクロフォン
203
入力デバイス
204
通信I/F
205
メモリ
208
ディスプレイ
209
スピーカ
211
212
プロセッサ

【図２７】


ST51
To: John
Ud
ST52
To: John
IC1
Ud
音声入力:非アクティブ
ST53
To: John
Hello, how are|
IC1
Ud
音声入力:アクティブ
ST54
To: John
Hello, how are|
IC1
Ud
音声入力:非アクティブ

フロントページの続き

(72)発明者　河野　道成
東京都港区港南１丁目７番１号　ソニー株式会社内
(72)発明者　池田　卓郎
東京都港区港南１丁目７番１号　ソニー株式会社内
(72)発明者　岡田　憲一
東京都港区港南１丁目７番１号　ソニー株式会社内

審査官　大野　弘

(56)参考文献　特表２００３－５２６１２０（ＪＰ，Ａ）
特開２００４－２３９９６３（ＪＰ，Ａ）
特開２００２－１０８３９０（ＪＰ，Ａ）
特開２００６－２９４０６６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／２２
Ｇ１０Ｌ　１５／２８