(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2022-534708
(P2022-534708A)
(43)公表日　令和4年8月3日(2022.8.3)

(51)国際特許分類		FI			テーマコード（参考）
Ｇ０６Ｔ	*13/40 (2011.01)*	Ｇ０６Ｔ	13/40		５Ｂ０５０
Ｇ１０Ｌ	*15/10 (2006.01)*	Ｇ１０Ｌ	15/10	５００Ｎ	５Ｅ５５５
Ｇ１０Ｌ	*15/22 (2006.01)*	Ｇ１０Ｌ	15/22	３００Ｚ	
Ｇ０６Ｆ	*3/01 (2006.01)*	Ｇ０６Ｆ	3/01	５１０	
Ｇ０６Ｆ	*3/0481(2022.01)*	Ｇ０６Ｆ	3/0481		

審査請求　未請求　予備審査請求　未請求　（全34頁）　最終頁に続く

(21)出願番号　特願2021-569969(P2021-569969)
(86)(22)出願日　令和2年6月4日(2020.6.4)
(85)翻訳文提出日　令和4年1月24日(2022.1.24)
(86)国際出願番号　PCT/US2020/036068
(87)国際公開番号　WO2020/247590
(87)国際公開日　令和2年12月10日(2020.12.10)
(31)優先権主張番号　62/858,234
(32)優先日　令和1年6月6日(2019.6.6)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,
最終頁に続く

(71)出願人　521511564
アーティー，インコーポレイテッド
ＡＲＴＩＥ，ＩＮＣ．
アメリカ合衆国 ９００２６ カリフォルニア州，ロサンゼルス，バークレー サークル ３０３３
３０３３ Ｂｅｒｋｅｌｅｙ Ｃｉｒｃｌｅ，Ｌｏｓ Ａｎｇｅｌｅｓ，Ｃａｌｉｆｏｒｎｉａ ９００２６（ＵＳ）
(74)代理人　110002734
特許業務法人藤本パートナーズ
(72)発明者　マッキンタイヤ－カーウィン，アルマンド
アメリカ合衆国 ９００２６ カリフォルニア州，ロサンゼルス，バークレー サ
最終頁に続く

(54)【発明の名称】　動的に反応する仮想キャラクターのためのマルチモーダルモデル

(57)【要約】
本開示の実施形態は、マルチモーダルモデルを使用して、仮想キャラクター（又は「アバター」）を制御するための方法に関する。マルチモーダルモデルは、ユーザに関連する様々な入力情報を処理して、複数の内部モデルを使用して入力情報を処理し得る。マルチモーダルモデルは、内部モデルを組み合わせて、本物らしく感情的に魅力的な応答を作成し得る。仮想キャラクターへのリンクは、ウェブブラウザに埋め込まれ得、アバターは、仮想キャラクターと対話するためのユーザによる選択に基づいて、動的に生成され得る。クライアントについて、レポートが生成され得、このレポートは、クライアントに関連付けられた仮想キャラクターと対話するユーザの特性に関するインサイトを提供する。
【選択図】図１

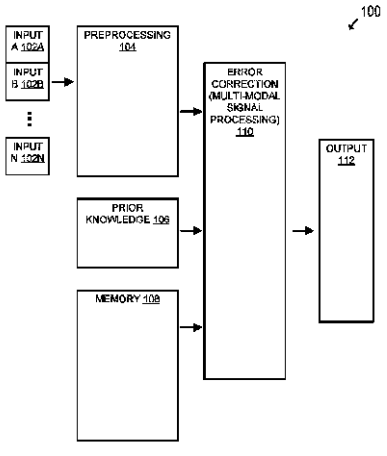


FIGURE 1

【特許請求の範囲】
【請求項１】
仮想キャラクターを制御するための方法であって、
マルチモーダル入力情報をデバイスから受信するステップであって、前記マルチモーダル入力情報が、発話情報、顔面表情情報、及び、前記デバイスを取り巻く環境を表す環境情報のいずれかを含む、ステップと、
前記デバイスに提示される表示環境内の位置に、前記仮想キャラクターを表示するステップと、
少なくとも２つの内部モデルを実装するステップであって、それによって前記マルチモーダル入力情報の特性を識別する、ステップと、
前記少なくとも２つの内部モデルの識別された前記特性を検査するステップであって、それによって、前記識別された特性のうち第１の識別された特性が、前記識別された特性のうち第２の識別された特性における類似の特徴の閾値数を含むかどうかを決定する、ステップと、
前記第１の識別された特性及び前記第２の識別された特性を、仮想キャラクター知識モデルに含まれる前記仮想キャラクターに特有の情報と比較するステップであって、それによって、前記第１の識別された特性が前記識別された特性のうち前記第２の識別された特性における類似の特徴の前記閾値数を含むと決定することに基づいて、選択された特性を選択するステップと、
前記仮想キャラクターに関連付けられた潜在的なアクションのライブラリにアクセスするステップであって、それによって、前記選択された特性に整合するアクションを決定し、前記アクションは、前記仮想キャラクターによって実行されることとなるアニメーションと、該アニメーションに関連付けられたオーディオとの両方が含まれる、ステップと、
前記デバイス上に提示された環境内における前記仮想キャラクターを修正すると共に前記関連付けられたオーディオを出力することによって、決定された前記アクションを実行するステップと、
を含む、方法。
【請求項２】
前記少なくとも２つの内部モデルは、前記発話情報から発話感情を解析可能な発話認識モデルと、前記顔面表情情報に基づいて顔面特徴感情を検出可能な顔面特徴認識モデルとを含み、前記選択された特性は、前記発話感情と前記顔面特徴感情との間で共通の感情であり、前記決定されたアクションは、前記感情に基づいて決定される、請求項１に記載の方法。
【請求項３】
前記少なくとも２つの内部モデルは、ユーザとの以前のエンゲージメントに関連する情報を含む事前知識情報を引出し可能な事前知識モデルを含んでおり、前記選択された特性は、前記事前知識モデルを使用して処理された前記事前知識情報に基づいて選択される、請求項１に記載の方法。
【請求項４】
前記内部モデルは、聴覚情報からコンテキスト及び意味を導出するように構成された自然言語理解モデル、環境情報を識別するように構成された認識モデル、及び、ユーザ及び他の仮想キャラクターに関連するデータを識別するように構成されたソーシャルシミュレーションモデルを含む、請求項１に記載の方法。
【請求項５】
前記仮想キャラクターに対して、前記デバイス上におけるユーザへの質問を表す初期アクションを実行するように指示するステップであって、前記入力情報は、前記ユーザによる前記質問に対する応答を表す、ステップをさらに含む、請求項１に記載の方法。
【請求項６】
ネットワークを介して複数のユーザと埋め込みリンクを共有するステップと、
デバイスのセットのいずれかから、前記埋め込みリンクが選択されたことを示す選択を

受信するステップと、
前記選択の受信に応答して、前記選択を送信した前記デバイスのセットのうち前記デバイスへとデータのストリームを送信するステップであって、それによって、前記仮想キャラクターを前記デバイス上に表示する、ステップと
をさらに含む、請求項１に記載の方法。
【請求項７】
第１の時間に、前記データのストリームの第１のバッチを送信するステップであって、前記第１のバッチは、前記デバイスの前記ディスプレイ上に前記仮想キャラクターを最初に生成するための情報を含む、ステップと、
前記第１の時間の後の第２の時間に、前記データのストリームの第２のバッチを送信するステップであって、前記第２のバッチは、前記仮想キャラクターによる第１のアクションを出力するための情報を含み、前記第２の時間では、前記第１のバッチが破棄される、ステップと
をさらに含む、請求項６に記載の方法。
【請求項８】
前記環境情報を検査するステップであって、それによって、前記環境の床を表す前記環境の部分を識別するステップと、
前記仮想キャラクターを、前記環境の床を表す前記環境の前記部分の上方の第１の位置に配置するステップと
をさらに含む、請求項１に記載の方法。
【請求項９】
デバイスによってキャプチャされたユーザに関連するマルチモーダル入力に対する応答を提供するように構成された該デバイスであって、
少なくとも１つのメモリであって、
マルチモーダル入力情報から特性を識別するように構成された、少なくとも２つの内部モデルと、
仮想キャラクターに特有の情報を含む仮想キャラクター知識モデルと、
前記仮想キャラクターに関連付けられた潜在的アクションのライブラリであって、各アクションは、前記仮想キャラクター及びそれに関連付けられたオーディオによって実行されるアニメーションに関連付けられている、ライブラリと
を含むメモリと、
少なくとも１つのプロセッサであって、
発話情報、顔面表情情報、及び、環境を表す環境情報のうち少なくとも１つを含むマルチモーダル入力情報を受信するステップと、
前記少なくとも２つの内部モデルによって識別された前記特性を検査するステップであって、それによって、第１の識別された特性が、第２の識別された特性との閾値類似度の範囲内にあるかどうかを決定する、ステップと、
前記第１の識別された特性及び前記第２の識別された特性を、仮想キャラクター知識モデルと比較するステップであって、それによって、選択された特性を識別する、ステップと、
前記仮想キャラクターに関連付けられた前記潜在的アクションのライブラリを検査することによって、前記選択された特性に整合するアクションを決定するステップであって、前記アクションは、前記デバイス上に出力されることになるオーディオを含む、ステップと、
前記オーディオを前記デバイス上に出力するステップと
を行うように構成された、プロセッサと、
を含む、デバイス。
【請求項１０】
前記少なくとも１つのプロセッサは、さらに、
前記環境情報から導出された前記環境内の位置内において、前記デバイスの前記ディス

プレイ上に前記仮想キャラクターを表示するステップと、
前記仮想キャラクターを前記デバイス上に表示される前記環境内において修正することにより、前記デバイスに出力されることとなる前記オーディオと、前記仮想キャラクターによって実行されることとなる選択されたアニメーションとの両方を含むアクションを実行するステップと
を行うように構成される、請求項９に記載のデバイス。
【請求項１１】
前記少なくとも２つの内部モデルは、前記ユーザとの以前のエンゲージメントに関連する情報を含む前記事前知識情報を引出し可能な事前知識モデルを含んでおり、前記選択された特性は、前記事前知識モデルを使用して処理された前記事前知識情報に基づいて選択される、請求項９に記載のデバイス。
【請求項１２】
前記少なくとも２つの内部モデルは、前記発話情報から発話感情を解析可能な発話認識モデルと、前記顔面表情情報に基づいて顔面特徴感情を検出可能な顔面特徴認識モデルとを含み、前記選択された特性は、前記発話感情と前記顔面特徴感情との間で共通の感情であり、前記決定されたアクションは、前記感情に基づいて決定される、請求項９に記載のデバイス。
【請求項１３】
ユーザデバイスのウェブブラウザ上で仮想キャラクターを動的に生成するためのコンピュータ実装される方法であって、
前記ユーザデバイスの前記ウェブブラウザにリンクを埋め込むステップであって、前記リンクは、前記ウェブブラウザを前記ユーザデバイス上で実行されているアプリケーションにリンクさせる、ステップと、
前記ユーザデバイスから、前記リンクが選択されたという通知を受信するステップと、
情報を表す前記アプリケーションからのデータのストリームを前記ウェブブラウザへと送信するステップであって、それによって、前記仮想キャラクターを生成する、ステップと、
前記仮想キャラクターを、前記ユーザデバイスの前記ウェブブラウザ上に表示するステップと、
前記ユーザデバイスからマルチモーダル入力情報を受信するステップであって、前記マルチモーダル入力情報が、発話情報、顔面表情情報、及び、環境を表す環境情報を含む、ステップと、
少なくとも２つの内部モデルを実装するステップであって、それによって前記マルチモーダル入力情報の特性を識別する、ステップと、
前記少なくとも２つの内部モデルによって識別された前記特性を検査するステップであって、それによって、第１の識別された特性が、第２の識別された特性との閾値類似度の範囲内にあるかどうかを決定する、ステップと、
前記第１の識別された特性及び前記第２の識別された特性を、仮想キャラクター知識モデルに含まれる前記仮想キャラクターに特有の情報と比較するステップであって、それによって、前記第１の識別された特性が前記識別された特性のうち前記第２の識別された特性における類似の特徴の前記閾値数を含むと決定することに基づいて、選択された特性を選択するステップと、
前記仮想キャラクターに関連付けられた潜在的なアクションのライブラリにアクセスするステップであって、それによって、前記選択された特性に整合するアクションを選択し、前記アクションは、前記仮想キャラクターによって実行されることとなるアニメーションと、該アニメーションに関連付けられたオーディオとが含まれる、ステップと、
前記環境内において前記アクションを実行している前記仮想キャラクターを表示すると共に前記関連付けられたオーディオを出力するステップと、
を含む、コンピュータ実装される方法。
【請求項１４】

前記ウェブブラウザは、前記ユーザデバイス上で実行されるモバイルアプリケーション上に表示されるページを含む、請求項１３に記載のコンピュータ実装される方法。
【請求項１５】
前記少なくとも２つの内部モデルは、発話情報から発話感情を解析可能な発話認識モデルと、前記顔面表情情報に基づいて顔面特徴感情を検出可能な顔面特徴認識モデルとを含み、前記選択された特性は、前記発話感情と前記顔面特徴感情との間で共通の感情であり、前記決定されたアクションは、前記感情に基づいて決定される、請求項１３に記載のコンピュータ実装される方法。
【請求項１６】
前記少なくとも２つの内部モデルは、ユーザとの以前のエンゲージメントに関連する情報を含む事前知識情報を引出し可能な事前知識モデルを含んでおり、前記選択された特性は、前記事前知識モデルを使用して処理された前記事前知識情報に基づいて選択される、請求項１３に記載のコンピュータ実装される方法。
【請求項１７】
ネットワークを介して複数のユーザと埋め込みリンクを共有するステップと、
デバイスのセットのいずれかから、前記リンクが選択されたことを示す選択を受信するステップと、
前記選択の受信に応答して、前記選択を送信した前記デバイスのセットのうち前記ユーザデバイスへとデータのストリームを送信するステップであって、それによって、前記仮想キャラクターを前記ユーザデバイス上に表示する、ステップと
をさらに含む、請求項１３に記載のコンピュータ実装される方法。
【請求項１８】
第１の時間に、前記データのストリームの第１のバッチを送信するステップであって、前記第１のバッチは、前記ユーザデバイスのディスプレイ上に前記仮想キャラクターを最初に生成するための情報を含む、ステップと、
前記第１の時間の後の第２の時間に、前記データのストリームの第２のバッチを送信するステップであって、前記第２のバッチは、前記仮想キャラクターによる第１のアクションを出力するための情報を含み、前記第２の時間では、前記第１のバッチが破棄される、ステップと
をさらに含む、請求項１３に記載のコンピュータ実装される方法。
【請求項１９】
前記環境情報を検査するステップであって、それによって、前記環境の床を表す前記環境の部分を識別するステップと、
前記仮想キャラクターを、前記環境の床を表す前記環境の前記部分の上方の第１の位置に配置するステップと
をさらに含む、請求項１３に記載のコンピュータ実装される方法。
【請求項２０】
前記選択された特性と前記アクションとに関連する情報を格納するステップと、
複数のユーザについての一連の選択された特性及びアクションを集約するステップと、
前記複数のユーザについての前記一連の選択された特性及びアクションを処理するステップであって、それによって、前記仮想キャラクターと前記複数のユーザとのエンゲージメントに関連する分析のセットを導出するステップと、
前記仮想キャラクターと前記複数のユーザとのエンゲージメントに関連する分析の前記セットを表示する分析ダッシュボードを提示するステップと
をさらに含む、請求項１３に記載のコンピュータ実装される方法。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願への相互参照
本実施形態は、発明の名称が“ MULTI-MODAL MODEL FOR DYNAMICALLY RES

PONSIVE AVATARS（動的応答性アバターのためのマルチモーダルモデル）"である、２０１９年６月６日に出願された米国仮特許出願第６２／８５８，２３４号に対する優先権を主張し、その全体が参照により本明細書に組み込まれる。

【０００２】

技術分野

本開示の教示は、概して、デバイス上の仮想キャラクターを制御することに関する。より詳細には、本開示の教示は、マルチモーダルモデルを使用して、デバイス上で動的に応答する仮想キャラクターを制御することに関する。

【背景技術】

【０００３】

仮想キャラクター（又は「アバター」）は、ユーザデバイス（例えば、スマートフォン、コンピュータ、拡張現実デバイス）上において、ユーザとの対話を促進し得る。仮想キャラクターは、ユーザデバイスのディスプレイに表示される環境に描写されるキャラクターの仮想表現を含み得る。仮想キャラクターは、視覚的形態である必要はなく、ユーザデバイスを通じて、ユーザとコミュニケーション可能でありさえすればよい。仮想キャラクターが行うためのアクションを決定するために、ユーザからの入力が識別及び検査され得る。仮想キャラクターは、決定されたアクション（例えば、アニメーション、発話の実行）を行うことができ、それによって、仮想キャラクターとユーザとの間の継続的な対話を促進し得る。

【０００４】

仮想キャラクターを生成するための重要な考慮事項は、ユーザが受信した入力に応答して仮想キャラクターが行うアクションの正確さである。全般的に、入力に対する応答が正確であるほど、仮想キャラクターとの対話におけるユーザエクスペリエンスが向上し得る。逆に、ユーザが提供した入力を仮想キャラクターが誤って解釈すると、仮想キャラクターは入力に対して不正確な応答を示すことがあり、仮想キャラクターのユーザエクスペリエンスが低下することになる。

【図面の簡単な説明】

【０００５】

【図１】図１は、様々な実施形態に係る、仮想キャラクターの出力を制御するためのマルチモーダルモデルのフロー図を示す。

【０００６】

【図２】図２は、様々な実施形態に係る、仮想キャラクターのエラーをマルチモーダルを使用して修正するためのフロー図を示す。

【０００７】

【図３】図３は、様々な実施形態に係る、仮想キャラクターを制御するためのマルチモーダルモデルのフロー図を示す。

【０００８】

【図４】図４は、様々な実施形態に係る、マルチモーダルモデル上に実装された仮想キャラクターと対話するためのシステムを示している。

【０００９】

【図５】図５は、様々な実施形態に係る、デバイス上に表示される例示的な仮想キャラクターを示す。

【００１０】

【図６】図６は、様々な実施形態に係る、ユーザと仮想キャラクターを表示するユーザデバイスとの間の対話のフロー図を示す。

【００１１】

【図７】図７は、様々な実施形態に係る、仮想キャラクター反応からインサイトを抽出するフロー図を示す。

【００１２】

【図８】図８は、様々な実施形態に係る、インサイトダッシュボードを示す。

【００１３】
【図９】図９は、様々な実施形態に係る、仮想キャラクターをマルチモーダルモデルを使用して制御する方法のブロック図を示す。
【００１４】
【図１０】図１０は、本明細書に記載の少なくともいくつかの動作を実装可能な処理システムの例を示すブロック図である。
【００１５】
これらの図面及び表は、様々な実施形態について、例示のみを目的として表すものである。当業者は、本技術の原理から逸脱することなく、代替となる実施形態を採用できることを認識するであろう。したがって、これらの図面には特定の実施形態が示されているが、本技術は、様々な修正がなされる余地がある。
【発明を実施するための形態】
【００１６】
以下に示す実施形態は、当業者が実施形態を実践するために必要な情報を示し、実施形態を実践するための最良の態様を例示するものである。添付の図に照らして以下の説明を読めば、当業者は、本開示の概念を理解し、本明細書で特に言及されていないこれらの概念の応用を認識するであろう。これらの概念及び応用は、本開示及び添付の特許請求の範囲に含まれる。
【００１７】
実施形態は、特定のコンピュータプログラム、システム構成、ネットワークなどを参照して説明される場合がある。しかしながら、当業者は、これらの特徴が、他のコンピュータプログラムのタイプ、システム構成、ネットワークのタイプ等にも等しく適用可能であることを認識するであろう。例えば、用語「Ｗｉ－Ｆｉネットワーク」は、ネットワークを説明するために使用され得るが、関連する実施形態は、別のタイプのネットワークに展開され得る。
【００１８】
さらに、開示の技術は、特殊用途のハードウェア（例えば、回路）、ソフトウェア及び／若しくはファームウェアで適切にプログラムされたプログラマブル回路、又は、特殊用途のハードウェアとプログラマブル回路との組み合わせを使用して具現化することができる。したがって、実施形態は、電子デバイスによって生成されたビデオコンテンツを検査し、ビデオコンテンツに含まれる要素を識別し、分類モデルを適用して適切なアクションを決定して、適切なアクションを実行するように、コンピューティングデバイス（例えば、基地局又はネットワーク接続されたコンピュータサーバ）をプログラムするために使用され得る命令を有する機械可読媒体を含んでいてもよい。
【００１９】
本開示において、用語「アバター」は、「仮想キャラクター」と交換可能である。仮想キャラクターには、ユーザとコミュニケーション可能なユーザデバイス上における仮想エンティティが含まれ得る。このコミュニケーションは、例えば、声（会話）、テキスト（例えば、チャット又はメッセージングアプリケーション）、画像／ビデオ、音楽、及び／又は、任意のプログラムの実行（例えば、仮想キャラクターが、調整の許可を有するＩｏＴ組み込みシステム内の照明スイッチをオンにする）等の、種々異なるタイプの対話によって生じてもよい。仮想キャラクターは、外観としてアニメーションする体を有していてもよいが、例えば、声のみ又はテキストベースのエージェントのみであってもよく、必ずしもアニメーションする体を有していなくてもよい。
【００２０】
用語
本明細書で使用される用語の目的は、実施形態を説明することのみであり、本開示の範囲を限定することを意図するものではない。文脈が許す場合、単数形又は複数形を使用する単語は、それぞれ、複数形又は単数形を含み得る。
【００２１】

本明細書で使用される場合、特に別記のない限り、用語「処理（processing）」、「計算（computing）」、「算定（calculating）」、「決定（determining）」、「表示（displaying）」、「生成（generating）」等は、コンピュータのメモリ又はレジスタ内で物理（電子）量として表されるデータを、コンピュータのメモリ、レジスタ又は他のそのような記憶媒体、伝送デバイス若しくはディスプレイデバイス内で物理量として同様に表される他のデータへ操作及び変換する、コンピュータ又は同様の電子コンピューティングデバイスのアクション及びプロセスを意味する。

【００２２】

本明細書で使用される場合、用語「接続（connected）」、「結合（coupled）」等は、２つ以上の要素間の直接的又は間接的な任意の接続又は結合を指し得る。要素間の結合又は接続は、物理的、論理的又はそれらの組み合わせであり得る。

【００２３】

「実施形態（an embodiment）」又は「一実施形態（one embodiment）」への言及は、記載される特定の特徴、機能、構造又は特性が、少なくとも１つの実施形態に含まれることを意味する。そのような句が現れても、それらは必ずしも同じ実施形態に言及しているわけではなく、必ずしも相互に排他的である代替の実施形態に言及しているわけでもない。

【００２４】

文脈が明らかに他を要求しない限り、単語「含む（comprise）」及び「含んでいる（comprising）」は、排他的又は網羅的な意味ではなく、包括的意味で（すなわち、「含むが限定されない」という意味で）解釈されるべきである。

【００２５】

用語「基づく（based on）」もまた、排他的又は網羅的な意味ではなく、包括的意味で解釈されるべきである。したがって、特に明記されない限り、用語「基づく」は、「少なくとも部分的に基づく」を意味することが意図される。

【００２６】

用語「モジュール」は、ソフトウェアコンポーネント、ハードウェアコンポーネント及び／又はファームウェアコンポーネントを広く示す。モジュールは、典型的には、指定された入力に基づいて有用なデータ又は他の出力を生成可能な機能的コンポーネントである。モジュールは通常、指定された入力に基づいて有用なデータ又は他の出力を生成できる機能コンポーネントである。モジュールは、自己完結型であってもよい。コンピュータプログラムは、１つ又は複数のモジュールを含んでいてもよい。したがって、コンピュータプログラムは、種々異なるタスクを完遂することを担当する複数のモジュールを含んでいてもよく、複数のタスクを完遂することを担当する単一のモジュールを含んでいてもよい。

【００２７】

複数の項目のリストに関して使用される場合、単語「又は」は、リスト内の項目のいずれか、リスト内の項目の全て、及び、リスト内の項目の任意の組み合わせという解釈の全てを網羅することを意図している。

【００２８】

本明細書に記載のプロセスのいずれかにおいて実行されるステップの順序は、例示的なものである。しかしながら、物理的可能性に反しない限り、これらのステップは、様々な順序及び組み合わせで実行されてもよい。例えば、本明細書に記載のプロセスにステップが追加されてもよく、本明細書に記載のプロセスからステップが削除されてもよい。同様に、ステップは、置換されてもよく、並べ替えられてもよい。したがって、任意のプロセスについての記載は、オープンエンドであることが意図される。

【００２９】

システム概要

仮想キャラクター（又は「アバター」）は、ユーザデバイスを介したユーザとのコミュニケーションを促進することができる。仮想キャラクターは、キャラクター（例えば、本

、映画、テレビのキャラクター）の仮想表現、有名人の表現、人間の表現、動物の表現、架空の生き物の表現、及び、物体（典型的には、現実世界では無生物であるが、対話の目的のため（喋る本のように）人間とコミュニケーションすることができる）の表現であってもよい。加えて、仮想キャラクターは、幽霊、霊魂、感情のように、物理的形態を有さないエンティティ又はアニマシーを表すことができる。例えば、ユーザは、ユーザに関連付けられたスマートフォンを介して、スーパーヒーローを表す仮想キャラクターと対話してもよい。仮想キャラクターを生成するシステムの観点から、該システムは、ユーザからの入力を受信し、その入力に基づいて応答を決定してもよい。

【００３０】

仮想キャラクターを生成するための重要な考慮事項は、ユーザが受信した入力に応答して仮想キャラクターが行うアクションの正確さである。全般的に、入力に対する応答が正確であるほど、仮想キャラクターとの対話におけるユーザエクスペリエンスが向上し得る。逆に、ユーザが提供した入力を仮想キャラクターが誤って解釈すると、仮想キャラクターは入力に対して不正確な応答を示すことがあり、仮想キャラクターのユーザエクスペリエンスが低下することになる。

【００３１】

多くの場合では、仮想キャラクターに提供される入力は、分離されて個別に処理される。例えば、仮想キャラクターは、ユーザの声データと、顔面特徴情報とを受信し得る。しかしながら、これらの情報は、仮想キャラクターについて出力するアニメーション／発話の識別に伴って、分離されて個別に処理されてもよいが、仮想キャラクターのアニメーション／発話の精度が低下することがある。

【００３２】

仮想キャラクター制御及びエラー訂正のためのマルチモーダルモデル

本実施形態は、キャラクターのアクションを自動的に制御すると共に、仮想キャラクターによって実行される正確なアクションを提供するためにマルチモーダルモデルを使用してエラーを修正するプログラムに、少なくとも部分的に関連し得る。マルチモーダルモデルは、ユーザに関連する情報の複数のモード（例えば、発話、顔面表情、環境情報、ユーザとの以前の対話から収集された情報、及び、システムがユーザについて知っている情報）を識別し、該情報を、ユーザ入力に応答して仮想キャラクターについての正確なアニメーション／アクションを動的に生成するための様々な内部モデルに組み合わせるモデルを含んでいてもよい。

【００３３】

マルチモーダルモデルは、マルチモーダルセンサのアレイによって生成されたデータ信号から抽出された情報について、エラー訂正を実行してもよい。換言すれば、現実世界の環境から読取値を収集するマルチモーダルデジタルセンサの任意のアレイと、データ信号の各タイプに合わせて調整された任意の前処理及び情報抽出プロセスのセットとが、各センサから抽出された情報を使用し、それによって、他のデータ信号から抽出された情報のエラーを訂正してもよい。マルチモーダルエラー訂正では、対話の時にセンサによって収集された情報に加えて、過去からユーザについて収集された情報を使用することもできる。例えば、マルチモーダルエラー訂正（ＭＭＥＣ）は、エラー訂正を促進するために、ユーザが過去にキャラクターに対して発言したことを記録したもの（transcript）を使用することができる。また、ＭＭＥＣは、エラー訂正をガイドするために、ユーザについての情報を含む知識ベース（オントロジー）内に格納されている情報を使用することもできる。現在収集中のノイズの多い信号を訂正するために、ユーザの過去についての検証済みの情報を使用することができる。知識ベース内の情報は、コンピュータによって自動的に抽出されたものでもよく、人間によって手動で入力されたより構造化された情報でもよい。

【００３４】

図１は、様々な実施形態に係る、仮想キャラクターの出力を制御するためのマルチモーダルモデルのフロー図を示す。マルチモーダルモデルは、発話認識技術、自然言語理解、

コンピュータビジョン（例えば、ユーザの顔面認識、ユーザの物理的環境及び他の物理的オブジェクトについての「世界認識」データ）、知識ベース（オントロジー）、特定の仮想キャラクターのペルソナ、及び、統合されたマルチモーダルモードへのソーシャルシミュレーションを組み合わせてもよい。マルチモーダルモデルは、本物らしく感情的に魅力的な人工知能アルゴリズムを作成し、その結果として正確な仮想キャラクターを生み出すために、複数の内部モデルを共に使用することが可能であってもよい。

【００３５】

図１に示すように、システムは、センサのアレイから生の入力（入力１０２Ａ～Ｎ）を受信し得る。該システムは、シングルモード信号処理を実行する前処理１０４を含み得る。該システムは、知識ベース及びモデル内に符号化された世界についての情報を含む事前知識１０６と、ユーザとの過去の経験について符号化された情報を格納するメモリ１０８とを含み得る。該システムは、エラー訂正を実行可能なマルチモーダル信号処理１１０を含み得る。該システムは、マルチモーダル信号処理に基づいて、増強された結果１１２を出力し得る。

【００３６】

いくつかの実施形態では、コンピュータシステムは、聴覚（audio）信号を生成するマイク、ビデオストリーム信号を生成するカメラ、温度信号を生成可能な熱抵抗器、どのような臭いが存在するかについての信号を生成する嗅覚センサ、ＧＰＳセンサ、加速度計等、多くの種々異なるセンサを含んでいてもよい。システムは、これらのセンサから信号を受信して、現実世界で起こっていることを定量化することができ、これらの信号のデジタル表現を、互いのコンテンツを訂正するために使用することができる。

【００３７】

ユーザに関連して受信された入力は、マルチモーダルモデルに含まれる内部モデルと組み合わせられてもよい。上記のように、該モデルには、自然言語理解、知識モデル、ペルソナ、ユーザの生活／経歴についての構造的知識、及び、ソーシャルシミュレーションが含まれていてもよく、これらは、仮想キャラクターと対話する際の精度、及び、ユーザエクスペリエンスを向上させ得る。

【００３８】

例として、入力は「私は嬉しい」と言うユーザの声を含むことがあり、内部モデルは「私は嬉しい」という単語を識別するために自然言語理解を利用し得る。この例では、第２の入力は、ユーザの表情を識別する顔面表情情報を含み得る。内部モデルは、ユーザの顔面表情を決定するために、コンピュータビジョン等の顔面認識技術を含み得る。該システムは、ユーザが微笑んでいるかどうかの決定をチェックしてもよい。ユーザが微笑んでいることは、ユーザが実際に喜んでいることの確認を表し得る。該顔面表情が、ユーザが眉をひそめている又は悲しいことを決定した場合、該システムは、「嬉しい」の声入力は正しくなく、実際には「悲しい」ことを識別するために、全ての入力を検査し得る。この例に示されるように、マルチモーダルモデルは、仮想キャラクターのアクティビティの精度を高めるために、様々なサブシステム又は内部モデルの精度を相互参照及び確認することができる。

【００３９】

別の例として、ユーザが話しており、スーパーヒーローを表す仮想キャラクターに対して声入力を提供している場合、該システムは、「私のお気に入りの悪役はチョーカーである」という声入力をユーザから受信することがある。声入力の単語を識別した後、該システムは、知識モデルを利用して「チョーカー」という単語を検査し、内部モデルでスーパーヒーローに関連するものとして知られているキャラクターである単語「ジョーカー」に文法的に近いことを識別し得る。知識モデルは、ユーザが実際に「ジョーカー」を意味していることを示唆することができ、これにより、スーパーヒーローを表す仮想キャラクターによって提供される出力の精度が向上し得る。

【００４０】

いくつかの実施形態では、マルチモーダルモデルは、デバイスに接続されたセンサ（例

えば、カメラ、マイク）からキャプチャされた声データ、顔面表情データ及び環境情報を受信してもよい。声データは、ユーザの単語（例えば、ユーザの発言で構成されるテキストの線形ストリーム）の自動発話認識（ＡＳＲ）を使用して、検査及び処理され得る。顔面表情情報は、顔面表情認識アルゴリズム（例えば、多くの一般的な顔面表情を認識可能なアルゴリズム）を使用して、検査及び処理され得る。環境情報は、多くのオブジェクト（例えば、犬、植物、コーヒーカップ、他の人々等の一般的なオブジェクト）を認識する世界認識コンピュータビジョンアルゴリズムによって、検査及び処理され得る。世界認識アルゴリズムは、ユーザの物理的な世界についての他のデータポイント（内部又は外部、明るい又は暗い、等）を認識し得る。

【００４１】

いくつかの実施形態では、ユーザデバイスが複数のカメラを備えたデバイス（例えば、スマートフォン）上で拡張現実出力を利用している場合にのみ、１つ又は複数のコンピュータビジョンアルゴリズムがアクティブであり得る。いくつかの場合には、ユーザデバイスは、拡張現実表示を出力しておらず、１台のカメラにしかアクセスできないか、どのカメラにもアクセスできないことがある。

【００４２】

１つの内部モデルは、声入力情報に基づいて、ＡＳＲデータから、主題、視点（ＰＯＶ）、出来事、意図、名詞、動詞等を識別可能な自然言語理解（ＮＬＵ）モデルを含んでいてもよい。換言すれば、内部モデルがＡＳＲを介して認識される単語をユーザから受信した際、内部モデルは、ユーザの声データに関連するインサイト及びコンテキストの追加レイヤを追加してもよい。

【００４３】

一例として、ユーザは「僕は一日中学校にいて、ギターの弾き方を学んだ。すごく楽しかった」と発言することがある。内部モデルは、ユーザが一人称で話していること、彼らが主題であること、そして彼らの意図がギターを学ぶことであったことを抽出し得る。マルチモーダルモデルは、内部モデルを検査して、「ギターを楽しく学んでくれて嬉しいよ！」のような応答を生成し得る。これは、ただＡＳＲ技術を単独で使用するだけよりも正確であり得る。ＡＳＲ技術単独では、システムは単語「ギター」を理解するが、それが他のものとの関係でどのように使用されているか、話者がそれについてどのように感じているか、又は、彼らの目標は何であるかを判断しない。したがって、該システムは、ユーザの声入力に対して、肯定的又は否定的に応答することができないと考えられる。

【００４４】

いくつかの実施形態では、内部モデルは、仮想キャラクターの世界を表す知識モデルを含んでいてもよい。例えば、クライアントが、彼らのキャラクターの仮想キャラクターを生成する映画スタジオである場合、知識モデルには、キャラクターを最もよく描写するために、そのキャラクターに関連する情報が含まれ得る。この情報には、キャラクターの性格の特定の側面を表す「ペルソナ」が含まれていてもよい。知識モデルには、他のユーザや仮想キャラクターの原動力を解釈するソーシャルシミュレーションが含まれていてもよい。いくつかの実施形態では、知識モデルは、様々なソース（例えば、スクリプト、漫画本、本、ユーザ入力等）から、そのキャラクターに関する情報を受け取ることによって、キャラクターに関連するより多くの情報を自動的に生成し得る。

【００４５】

該システムは、生データ（例えば、ストリーミング発話データ、リアルタイムの発話－テキスト変換（speech-to-text））を受信し、顔面表情認識データをキャプチャし、そしてユーザの物理環境及び物理オブジェクトについての世界認識データをキャプチャしてもよい。該システムは、ユーザから追加のコンテキスト及び意味を導出するために、内部モデル、ＮＬＵアルゴリズム、知識モデル、ペルソナ及びソーシャルシミュレーションを実行しつつ、生データを相互に比較し得る。この情報及びモデルに基づいて、マルチモーダルモデルは、仮想キャラクターの出力をより正確に生成し得る。

【００４６】

コンピュータが世界についての情報を取得するためにセンサを使用する際、センサによって変換されたデータは、センサが描写しようとした現実に対してエラーを包含していることがある。センサは、現実世界の出来事によって生成された情報をサンプリングすることがある。この情報の損失に加えて、センサによって生成されたデータストリームを解釈するための計算方法もまた、エラーを生じやすいことがある。この典型には、エラーについて２つの主要な箇所があり得る。すなわち、センサからデータ信号への変換、及び、データストリームの解釈である。マルチモーダルエラー訂正は、センサ出力がデータ信号に変換される際に発生するエラーの一部を低減する助けとなる。加えて、データ信号の精度が高い場合には、データ信号は現実をより正確に表しているため、データ信号の解釈がより容易になり得る。このことは、マルチモーダルエラー訂正が、複数の面における精度を向上させるのに役立つことができる。

【００４７】

図２は、様々な実施形態に係る、仮想キャラクターのエラーをマルチモーダルを使用して修正するためのフロー図２００を示す。図２に示すように、複数のセンサ（例えば、画像センサ２０２Ａ、聴覚センサ２０２Ｂ、嗅覚センサ２０２Ｃ、センサＮ２０２Ｎ）は、入力層２０４にてデータを提供し得る。入力層２０４からのデータは、デジタル画像信号２０６Ａ、デジタル聴覚信号２０６Ｂ、デジタル嗅覚特徴信号２０６Ｃ、デジタル信号Ｎ２０６Ｎ等を含み得る。この入力データは、シングルモード情報抽出２０８を使用して処理されてもよく、このシングルモード情報抽出２０８は、コンピュータビジョン前処理２１０Ａ、聴覚前処理２１０Ｂ、嗅覚前処理２１０Ｃ、前処理タイプＮ２１０Ｎ等を含み得る。マルチモーダルエラー訂正２１６は、処理された情報を受け取り、情報ストリーム内のエラーを検出し得る。例えば、視覚情報ストリーム２１８Ａ、聴覚情報ストリーム２１８Ｂ、嗅覚情報ストリーム２１８Ｃ、情報ストリームＮ２１８Ｎ、事前知識２２０及びメモリは、出力層２２４内の拡張ストリーム（例えば、２２６Ａ～Ｎ）を導出するために、マルチモーダルエラー訂正２１６によって処理されてもよい。

【００４８】

各デバイスは、他のデバイスと対話する際に、固有のバージョンのマルチモーダルエラー訂正を使用してもよい。現実世界において実際に起こっていることを解読するため、情報は、共通認識及び世界知識と共に、全ての認識から併合され得る。コンピュータは、種々異なるモードの情報と、世界の事前情報（知識ベース及びモデルへと符号化されている）とをどのように合成するかについて、指示される場合がある。最終的なデータストリームを改善するために、マルチモーダルエラー訂正ステージの各接続は、種々異なる発信元からのデータ信号を使用することを表し得る。

【００４９】

種々異なるセンサからの情報を使用してデータストリームの品質を改善する方法のいくつかの例には、聴覚ベースのＡＳＲをコンピュータビジョンベースのリップリードと共に使用してＡＳＲ結果を改善すること、嗅覚センサを使用してＡＳＲ又は画像認識を訂正すること、画像／オブジェクト認識を使用してＡＳＲを訂正する、ＡＳＲを使用して画像認識を訂正すること、世界（現実世界又は架空世界）の知識を使用してＡＳＲ、ＣＶを改善すること等が含まれる。

【００５０】

いくつかの実施形態では、仮想キャラクターが具体的に言及されていない理由は、このエラー訂正プロセスが種々異なるタイプのシステムで使用され得るためであることがある。該システムは、何が言われているのか、どのような音が出ているのか、物理的世界について、及び、キャラクターに対するユーザの物理的な反応について、仮想キャラクターに認識させるために使用されてもよいが、自動運転車又はスマートホーム等、他のタイプのコンピュータシステムによっても使用され得る。

【００５１】

いくつかの実施形態では、マルチモーダルエラー訂正フェーズの目的は、フォールスネガティブを検出して、それらの値を再分類することであり得る。これにより、情報抽出パ

イプラインの再現率（recall）が改善され得る。
【００５２】
図２に示すように、世界知識アセットは、統合されていてもよい。図２では、世界知識は、パイプラインのマルチモーダルエラー訂正フェーズへの入力を含んでいることがある。世界知識は、入力層におけるセンサによって抽出された信号とは別に表されてもよい。入力層におけるセンサによって抽出された情報は、現時点の世界に起こっている出来事に関する情報をキャプチャし得る。世界情報の例には、ユーザの発言は何か、カメラフレーム内に検出可能なオブジェクトは何か、等が含まれる。
【００５３】
いくつかの実施形態では、いくつかの世界知識アセットが事前に収集されてもよい。世界知識アセットには、現実の世界についての事実情報であれ、（架空の人物の世界のような）物語の世界についての架空の情報であれ、世界についての事実が含まれ得る。世界知識アセットには、共通認識の知識が含まれ得る。例えば、オブジェクト「マグ」は「カップ」のタイプであり、飲食に意味的に関連する出来事に対して使用可能である。加えて、架空の人物の大砲についての情報もまた、世界知識アセットとして体系化可能である。例えば、バットマン（登録商標）の世界では、「ジョーカー（登録商標）」は主人公「バットマン」で心理ゲームをプレイする悪役である。バットマンの世界の他の悪役には、「ポイズン・アイビー」と「サブゼロ」が含まれ得る。この情報は、構造化された知識ベース又はオントロジーに符号化可能である。
【００５４】
世界知識／アセットは、過去に収集された情報を表すため、センサデータストリームから分離され得る。一方で、入力層におけるセンサからの信号は、コンピュータシステムとの対話がアクティブである間に継続的に取得されるデータ信号を表し得る。世界の知識ベースは静的であってもよく、対話／ユーザセッション中に成長しなくてもよい。この知識ベースは、コンピュータがこのコード化された情報を効率的に使用し、それによってより適切な決定を行うか、いくつかのコンテキストではマルチモーダルエラー訂正プロセスを支援できるように、事前にキュレーション及び設定されてもよい。
【００５５】
該情報は、マルチモーダルエラー訂正プロセスを強化することができる。例として、情報には、ユーザがコンピュータシステムに対して（ゴッサムシティにおける）バットマンの世界について話していることが含まれていてもよい。ユーザが何かを発言した後、ＡＳＲシステムは、ユーザが「チョーカーは悪い奴だ」と発言したと考える。バットマンの世界では、「ジョーカー」が悪役であり、ユーザが「ジョーカーは悪い奴だ」と発言しようとした可能性がはるかに高い。該システムは、スマートホーム、自動車、又は、ユーザと対話する他のデバイスで使用されてもよい。
【００５６】
図３は、様々な実施形態に係る、仮想キャラクターを制御するためのマルチモーダルモデルのフロー図３００を示す。該システムは、ユーザからの入力を受信して、マルチモーダルモデルのための内部モデルを実装することを含んでいてもよい。
【００５７】
仮想キャラクターの特徴を実装する内部モデルは、仮想キャラクターがリアルタイム又はほぼリアルタイムで行動及び応答することを可能にし得る。様々な入力（ユーザからの声入力、ユーザの顔面認識）を分離して個別に処理する代わりに、本実施形態は、仮想キャラクターによってアクションに表される、より正確なアニメーション／発話／出力を理解及び生成するために、複数の入力を入力及び検査し得る。受信した入力の例には、ユーザの声３０２、ユーザの顔面表情３０４、及び、（例えば、ビデオフィードの背景をスキャンして、コンピュータビジョンアルゴリズムが、ペット、テレビ、家具等の一般的なオブジェクトを認識できるかどうかを確認することによって）ユーザの環境３０６に関連する情報が含まれ得る。
【００５８】

入力はまた、自然言語理解データ３０８、知識モデルデータ３１０、ソーシャルシミュレーションデータ３１２等のような内部モデル入力を含んでいてもよい。マルチモーダルモデル３１４は、仮想キャラクターによって実行される（１つ以上の）アクションを表す出力３１６を生成するために、入力３０２～３１０を処理することができる。本実施形態は、仮想キャラクターについての正確な出力を生成するために、複数の入力を受信して、それらの入力を互いに比較し得る。

【００５９】

図４は、様々な実施形態に係る、マルチモーダルモデル上に実装された仮想キャラクターと対話するためのシステム４００を示す。図４に示すように、ユーザは、ユーザデバイス（例えば、スマートフォン、マイク、拡張現実グラス、コンピュータ、ゲームコンソール等）上の１つ以上のセンサ（例えば、カメラ、マイク等）から、入力情報（例えば、声、顔面認識、環境情報等）を提供してもよい。マルチモーダルモデルは、ユーザデバイスの１つで、又は、インターネット等のネットワークを介してユーザデバイスに接続された外部デバイス上で実行され得る。マルチモーダルモデルは、入力情報を検査し、内部モデルを実行し、センサによってキャプチャされたディスプレイ内の仮想キャラクターを示す表示を出力し得る。例えば、図４に示すように、仮想キャラクターは、センサによってキャプチャされた環境内の場所に表示される。

【００６０】

例として、ユーザデバイスは、オブジェクト（例えば、オブジェクト４０４）を含む環境４０２をキャプチャすることができる。表示された環境４０６及び識別されたオブジェクト４０８は、仮想キャラクター４１０を備えたインタフェース上に提供可能である。センサ４１２は、ユーザデバイスを取り巻く環境をキャプチャすることができる。例えば、センサ４１２は、ユーザ４１４に関連する情報をキャプチャすることができる。

【００６１】

マルチモーダルモデルは、アニメーションと音声の両方を含む仮想キャラクターの出力を生成してもよい。可能なアニメーションは、キャラクターについての潜在的なアニメーションのライブラリ内に格納され得る。これは、各アニメーションに対応する事前に録音された聴覚ファイルについての同様のライブラリに関連付けられていてもよい。これらのライブラリは、各キャラクターについて作成され、他のキャラクターについて個別のライブラリを作成するための基礎を提供してもよい。

【００６２】

いくつかの実施形態では、該システムは、全ての仮想キャラクターに共通するユニバーサルリギングモジュールを含んでいてもよい。リギングは、例えば、３Ｄキャラクターの骨、筋肉、皮膚を本質的に作成する、３Ｄアニメーションの手法である。ユニバーサルリギングモジュールは、ネットワーク内の１つ又は複数の仮想キャラクター全体にわたって使用することができる。

【００６３】

該システムは、ＡＩによって駆動されるプロシージャルアニメーションを生成してもよい。換言すれば、ＡＩが「私もギターが大好きです」と言いたい場合、ＡＩは仮想キャラクターに微笑むよう指示してもよく、そのような微笑みは、その特定のキャラクターが微笑む静的アニメーションファイルを全く用意することなく、プロシージャルに生じることになる。

【００６４】

いくつかの実施形態では、マルチモーダルモデルは、多くの聴覚ファイルを事前に録音する必要なく、音声を生成してもよい。マルチモーダルモデルは、比較的少ない時間量（例えば、５～１５時間）のキャラクターの声を受け取って、ＡＩが発言しようとするもの全てについての正確な音声を生成するテキスト－発話変換（text-to-speech）／発話合成システムを作成することができる。

【００６５】

本実施形態は、拡張現実ディスプレイを出力するように構成されたデバイス上で実行さ

れ得るが、本実施形態は、タッチスクリーン、モニター、スクリーン等の任意の適切なディスプレイ上で実行されてもよい。いくつかの実施形態では、本システムは、特に、システムがＦＥＲ及び世界認識内部モデルの両方を同時に利用する場合に、ユーザデバイス上で前面カメラ及び背面カメラの両方を利用してもよい。

【００６６】

いくつかの実施形態では、デバイス上に表示される仮想キャラクターを制御することは、デバイス（例えば、スマートフォン、コンピュータ、ＡＲ対応デバイス）上に配置されたセンサ（例えば、カメラ、マイク）から入力情報を受信することを含んでいてもよい。仮想キャラクターを制御することは、該入力情報の特性を識別するように構成された少なくとも２つの内部モデル（例えば、マルチモーダルモデル）が、内部モデルによって識別される特性を組み合わせて、内部モデルによって識別される特性間の類似度の識別を実施することを含んでいてもよい。特性間の類似度を見つけることによって、入力情報の最初のソースから識別された特性（例えば、「私は幸せである」と言うユーザ）が、入力情報の最初のソースから識別された特性（例えば眉をひそめている又は悲しげなユーザの顔面表情）と整合していないかどうかを識別し得る。特性の類似度が閾値より低い場合、内部モデルは、仮想キャラクターに共通の情報を含む知識モデルを検査することに基づき、より正確である可能性の高い特性を識別してもよい。

【００６７】

仮想キャラクターを制御することは、仮想キャラクターに関連付けられた潜在的なアクションのライブラリを検査し、それにより、内部モデルによって識別された特性に整合するアクションを決定することを含んでいてもよい。仮想キャラクターを制御することは、仮想キャラクターに内部モデルによって識別された特性に整合するアクションを実行させるためのアクションを出力することを含んでいてもよい。

【００６８】

いくつかの実施形態では、仮想キャラクターを制御することは、内部モデルによって識別された第１の特性の類似度が閾値類似度を下回ることを決定することを含んでいてもよい。仮想キャラクターを制御することは、仮想キャラクターに共通して関連付けられた情報を表すエントリのリストを表す第３の内部モデルを検査することを含んでいてもよい。仮想キャラクターを制御することは、第１の特性を、仮想キャラクターに関連付けられたリスト内に列挙されたエントリと照合することを含んでいてもよい。ここでは、内部モデルによって識別された特性は、リスト内に列挙されたエントリを含んでいる。

【００６９】

いくつかの実施形態では、入力情報は、ユーザの発話を表す聴覚情報、ユーザの顔面表情を表す顔面表情情報、及び、デバイスのセンサによってキャプチャされた環境を表す環境情報を含む。

【００７０】

いくつかの実施形態では、内部モデルは、聴覚情報からコンテキスト及び意味を導出するように構成された自然言語理解モデル、仮想キャラクターに特有の情報を識別するように構成された知識モデル、及び、ユーザ及び他の仮想キャラクターに関連するデータを識別するように構成されたソーシャルシミュレーションモデルを含む。

【００７１】

いくつかの実施形態では、仮想キャラクターを制御することは、デバイスのディスプレイ上に仮想キャラクターを表示し、仮想キャラクターに対して、ユーザへの質問を表す初期アクションを実行するように指示することを含んでいてもよい。ここでは、入力情報は、ユーザによる質問への応答を表す。

【００７２】

いくつかの実施形態では、センサを介して取得された情報は、知識ベース（オントロジー）内の情報を修正する必要を生じさせることがある。例えば、ユーザが「私のお気に入りのピザのトッピング（pizza topic）はアンチョビです」と発言し、知識ベースの現在のバージョンは、ユーザのお気に入りのピザのトッピングがキノコであると述べている

場合には、現時点では知識ベースが間違っており、アンチョビについての新しい好みを反映するように知識ベースを更新する必要があり得る。したがって、センサによって取得された新しい情報は、エラーを修正したり、知識ベースの情報を更新したりするために使用することができる

【００７３】

いくつかの実施形態では、仮想キャラクターは、ユーザの過去に関する情報にアクセスすることができる。この情報は、ユーザの人生における出来事のタイムラインを含むことができる。例えば、仮想キャラクターは、ユーザが５年前にパン屋を所有していたことを知っているかもしれない。この情報は、様々な方法で活用することができる。ユーザがパン屋であることをキャラクターが知っている場合、パン屋の語彙又は専門語（例えば、単語「こねる（knead）」）のセット全体があってもよい。単語「こねる（knead）」は、単語「必要（need）」と同様の響きがあり、自動発話認識ではしばしば間違って書き起こされる。多くのシステムでは、単語「こねる（knead）」を「必要（need）」と誤って書き起こす可能性がある。仮想キャラクターが、ユーザがパン屋であることを知っている場合、システムは、ユーザが単語「こねる（knead）」を発言し、それをより正確に書き起こす可能性がはるかに高いことを理解することができる。キャラクターが、ユーザがパン屋であることを知っている場合、職業に基づいて使用する単語のタイプを知ることによって、発話をより正確に書き起こすことができる。

【００７４】

ウェブブラウザで実行する完全にインタラクティブな仮想キャラクター

多くの場合、ユーザは、仮想キャラクターを生成するため、ユーザデバイス上で仮想キャラクターをレンダリングするのに必要な情報をダウンロードする。ユーザに仮想キャラクター情報をダウンロードさせるために、クライアントは、ユーザに対して、拡張現実仮想キャラクターに関連付けられたアプリケーションをダウンロードするよう納得させてもよい。特定のアプリケーションをユーザデバイス上に露骨にダウンロードする手順を行うと、ユーザとの軋轢が生じ、ユーザが仮想キャラクターを実装しなくなる可能性がある。

【００７５】

本実施形態は、本明細書に記載のマルチモーダルモデルを利用する３Ｄ仮想キャラクターをレンダリングして生成する、動的にレンダリングされた仮想キャラクター（又はインスタント仮想キャラクター）に関し得る。仮想キャラクターは、キャラクターに似た３Ｄビジュアル「ボディ」を含んでいてもよい。インスタント仮想キャラクターは、ハイパーリンクに埋め込まれていてもよい。インスタント仮想キャラクターは、ソーシャルメディアネットワーク等の様々なネットワーク上で共有されてもよい。例えば、会社は、人気のあるキャラクターや有名人の仮想キャラクターを、一連のファンと共有するエンターテインメント会社であり得る。該会社は、ハイパーリンク又は他の識別子（例えば、ＹｏｕＴｕｂｅ（登録商標）ビデオへのリンク）、テキストメッセージ等を共有してもよい。ネットワーク（例えば、ソーシャルメディア）を介して共有される際には、インスタント仮想キャラクターは、ユーザがアプリケーションを表示するためにダウンロードする必要なく、ソーシャルメディアフィード内又はウェブブラウザを介して即座に実行されてもよい。

【００７６】

インスタント仮想キャラクターは、ウェブブラウザ上で実行されてもよい。これによって、仮想キャラクターと、ユーザに関連付けられたスマートフォンとの対話が可能になり得る。マルチモーダルモデルは、ユーザをＷｅｂベースの拡張現実内に、モバイルウェブブラウザ（例えば、Ｓａｆａｒｉ（登録商標）、Ｃｈｒｏｍｅ（登録商標））を介して短いロード時間（例えば、１秒）でシームレスに移行し得る。したがって、ユーザは、「このリンクをクリックすると、有名人があなたの家に現れます」と述べるＷｅｂページ（例えば、Ｔｗｉｔｔｅｒ（登録商標））上のリンクを選択してもよい。リンクをクリックすると、仮想キャラクターがほぼリアルタイム（例えば、１秒未満）でスクリーン上に表示され得る。

【００７７】

該システムは、適切な３Ｄファイルフォーマット（例えば、ＧＬ Ｔｒａｎｓｍｉｓｓｉｏｎ Ｆｏｒｍａｔ（ＧＬＴＦ））を用いて、高度に最適化された３Ｄファイルを作成し得る。仮想キャラクターについてのアニメーション及び聴覚ファイルは、ユーザデバイスに対してバッチで（数回に分けて）ストリーミングされ、それによって、まずは仮想キャラクターの実行に必要である初期バッチがユーザデバイスに送信されて優先されてもよい。これらのバッチは、他の情報バッチのための余地を増やすために、実行後に破棄されてもよい。これらのユーザに対して送信されるバッチは、ユーザからの入力と仮想キャラクターによって生成された応答とに基づいて修正されて、動的になり得る。

【００７８】

いくつかの実施形態では、該システムは、インスタント仮想キャラクターがディスプレイ上の適切な位置に配置されるように、ユーザの環境（例えば、床）のウェブベースでの追跡を含んでいてもよい。

【００７９】

いくつかの実施形態では、該システムは、発話認識、自然言語理解、顔面表情認識及び世界認識アルゴリズムを、ユーザのモバイルウェブブラウザ内においてリアルタイム又はほぼリアルタイムで実行する。

【００８０】

いくつかの実施形態では、該システムは、前景の物体を識別するために、コンピュータビジョンアルゴリズムを使用する。例えば、該システムは、仮想キャラクターがソファの後ろに立っていると判断することがあり、ここで、アルゴリズムは、仮想キャラクターがソファによって適切に遮られ、ソファの後ろに立っているように見えるよう修正する。

【００８１】

いくつかの実施形態では、該システムはマルチユーザセッションをサポートしており、それによって、２人以上の人々が、仮想キャラクター又は仮想キャラクターを、個別のユーザデバイス（例えば、スマートフォン）を介して同じ場所で見ることができる。コンピュータビジョンアルゴリズムは、ユーザの環境を追跡し、複数のデバイスの世界観がわずかに異なる場合であっても、それらのデバイスにおいて該環境をリアルタイムで同期してもよい。

【００８２】

いくつかの実施形態では、該システムは、インスタント仮想キャラクターの異なる場所からの同期表示をサポートする。例えば、マルチモーダルモデルが有名人についての仮想キャラクター（例えば、有名人のミュージシャン）を生成し、彼女が同時に大勢の人々のために楽曲を演奏することを望む場合、該システムは、全てが同時に又はほぼ同時に楽曲を演奏する、大勢の有名人の仮想キャラクターを生成することができる。

【００８３】

いくつかの実施形態では、該システムは、３Ｄモデルへと変換可能な自分自身又は別の対象についての画像を撮ることによって、自分自身のインスタント仮想キャラクターを作成するユーザをサポートしてもよい。

【００８４】

いくつかの実施形態では、マルチモーダルモデルを使用して仮想キャラクターを動的に生成することは、仮想キャラクターを生成するためにウェブブラウザをアプリケーションにリンクする、ウェブページへのリンクを埋め込むことを含む。仮想キャラクターを動的に生成することは、リンクが選択されたデバイスからの選択を受信することを含んでいてもよい。このことは、デバイスと対話しているユーザが、仮想キャラクターを該デバイスのディスプレイ上に表示することを希望していることを示していることがある。仮想キャラクターを動的に生成することは、仮想キャラクターを生成するための情報を表すアプリケーションからのデータのストリームを送信して、仮想キャラクターをデバイスのディスプレイ上に表示することを含んでいてもよい。

【００８５】

仮想キャラクターを動的に生成することは、デバイス上に配置されたセンサから入力情

報を受信して、マルチモーダルモデルに含まれる少なくとも２つの内部モデルを実装することを含んでいてもよい。内部モデルは、入力情報の特性を識別して、該内部モデルによって識別された特性を組み合わせ、それによって、該内部モデルによって識別された特性間の類似度を識別するように構成されていてもよい。仮想キャラクターを動的に生成することは、仮想キャラクターに関連付けられた潜在的なアクションのライブラリを検査し、それによって、内部モデルによって識別される特性に整合するアクションを決定すること、及び、仮想キャラクターにアクションを実行させるための命令を表す、内部モデルによって識別される特性に整合するアクションを出力することを含んでいてもよい。

【００８６】

いくつかの実施形態では、仮想キャラクターを動的に生成することは、仮想キャラクターについての第１のアクションを開始することを含む。ここで、第１のアクションは、仮想キャラクターによって実行される音声及びアニメーションを含むユーザへの質問を表し、入力情報は、該質問に対するユーザによる応答を表す。

【００８７】

いくつかの実施形態では、仮想キャラクターを動的に生成することは、埋め込まれたリンクを複数のユーザに対してネットワークを介して共有し、デバイスのセットからリンクが選択されたことを示す選択を受信し、仮想キャラクターを生成するための情報を表すアプリケーションからのデータのストリームを、該デバイスのセットに含まれる各デバイスへと送信することを含む。

【００８８】

いくつかの実施形態では、仮想キャラクターを動的に生成することは、第１の時点にて、仮想キャラクターを最初に生成するための情報を含む、データのストリームの第１のバッチを送信し、第１の時点の後の第２の時点にて、仮想キャラクターによる第１のアクションを出力するための情報を含む、データストリームの第２のバッチを送信することを含む。ここで、第１のバッチは、第２の時点にて破棄される。

【００８９】

いくつかの実施形態では、仮想キャラクターを動的に生成することは、環境情報を検査して環境内に含まれる床を識別し、該環境内において識別された床の真上の第１の位置に仮想キャラクターを配置することを含む。

【００９０】

仮想キャラクターとのマルチモーダル対話からのインサイトの抽出

クライアント（例えば、企業）は、本明細書に記載される実施形態の少なくともいくつかを使用して、仮想キャラクターを生成し得る。クライアントは、指定される期間の間に、多くの人（例えば、１００万人）を仮想キャラクターと対話させることがある。例として、エンターテインメント会社は、多くの人々のために、有名人の仮想キャラクターを同時に生成し得る。

【００９１】

本実施形態は、ユーザの仮想キャラクターとの対話からインサイトを抽出することに少なくとも部分的に関連し得る。インサイトは、仮想キャラクターとの対話中に任意のセンサのセットによって収集されたデータから自動的に抽出され得る。

【００９２】

図５は、様々な実施形態に係る、デバイス５００上に表示される例示的な仮想キャラクター５０２を示す。仮想キャラクターには、実在の人間、架空の人間、又は、宇宙人若しくは喋る犬のような架空のキャラクターの表現のような、生物のエンティティの仮想表現が含まれ得る。仮想キャラクターは、コンピュータのスクリーン、最も典型的には、この特許の文脈において、スマートフォン上で視認可能である。仮想キャラクターは、スクリーン上でアニメーション化されたボディの表現を有していてもよい。仮想キャラクターは、視覚的に現れている必要はなく、自身を声として表すのみであってもよい。仮想キャラクターはまた、テキスト（チャット）又は画像を介してコミュニケーションするのみであってもよい。

【００９３】
仮想キャラクターは、スマートフォンのスピーカーを通じて発話することによって、人間のユーザと話すことができる。コンピュータが発話するオーディオ（audio）は、テキスト－発話変換（text to speech）シンセサイザーによって合成され得る。仮想キャラクターは、話すことを通じて現実世界に影響を与えてもよいが、仮想キャラクターは、照明をオンオフする、音楽をオンにする、部屋全体にわたるロボットの動きを制御する等、他の方法により世界に影響を与えることもできる。
【００９４】
仮想キャラクターの重要な特徴は、それらがアニマシーを含み得ることである。アニマシーには、現実世界に影響を与えるアクションを実行するキャラクターの能力が含まれ得る。仮想キャラクターの能力は、現実世界（又は、仮想世界であっても）における何かに対してどのように反応するかを決定するものであり、仮想キャラクターに魂を宿らせる（animate）ものである。
【００９５】
図６は、様々な実施形態に係る、ユーザと仮想キャラクターを表示するユーザデバイスとの間の対話のフロー図６００を示す。仮想キャラクターとの対話は、人６０２と仮想キャラクター６０６が、デバイス６０４を介して互いにコミュニケーションする場合を含み得る。人々、及び、仮想キャラクターは、発話６０８及び／又はボディランゲージ６１０を介する等、多くの形態でコミュニケーションを行うことができる。仮想キャラクターとの対話は、人が仮想キャラクターと積極的に関わる場合を含み得る。仮想キャラクターは、人や動物が魂を宿すのと同じようには魂を宿していないかもしれないが、仮想キャラクターは魂を宿していることを模倣することがあり、仮想キャラクターは、意味的に首尾一貫しており、面白みさえあるように応答することがある。仮想キャラクターは、合成された発話６１２及び／又はアニメーション化されたボディランゲージ６１４のうち任意のものを提供することができる。
【００９６】
いくつかの対話は、会話の形であってもよい。仮想キャラクターは、人間のユーザが仮想キャラクターに話していることを聞くために、携帯電話のマイクを使用することができる。マイクは、音波を聴覚ファイルに変換し得る。この聴覚ファイルは、発話－テキスト変換モデルによってテキストへと変換することができ、該テキストは、後に自然言語理解モデルによって解釈することができる。仮想キャラクターは、人間のユーザが話したことを理解すると、発話を合成して、身体的なジェスチャー応答においてボディをアニメーションして、又は、両タイプの反応を組み合わせて、人間のユーザに対してどのように応答するかを決定し始めることができる。
【００９７】
仮想キャラクターはまた、仮想空間内にて動くことによってコミュニケーションすることもできる。仮想空間は、拡張現実若しくは仮想現実を表す携帯電話、又は、拡張現実若しくは仮想現実のいずれかを表すことができるヘッドセット等、コンピュータスクリーン上で視覚化可能な三次元空間の任意の表現を含み得る。仮想空間にて動く仮想キャラクターの例としては、話している間、唇がアニメーションされ始め、仮想唇を使用して話しているシグナルを伝えたり、ボディランゲージを変化させて種々異なる気持ち又は感情を表現したり、顔面表情を変化させて種々異なる感情のシグナルを伝えたり、ボディパーツの位置を変更（防御性のために腕を組んだり、優位を示すために腰に手を当てたりする等）して種々異なる気持ちを表したりする等を含み得る。
【００９８】
インサイトは、生データから抽出された、構造化された情報又は理解の一部を含み得る。この構造化された情報又は理解の一部は、この少しの理解が正しいことの確実性を測定する、ある種の統計値と結合され得る。例として、ジョークが面白いかどうかをテストするために、該システムは、インスタント仮想キャラクターを多数の人々（１００人又は１０００人）に送信してもよい。仮想キャラクターは、各ユーザ（又は被験者）に対して、

このジョークを伝える。次に、インスタント仮想キャラクターは、このジョークが配信された後のユーザの反応を記録する。該システムは、ユーザの反応を分析し、それによって、このジョークが面白かったかどうかを決定することができる。ユーザは、このジョークが面白かったかどうかについて、笑うことによって等の様々な方法で表現することができ、それらは、コンピュータビジョンを使用してユーザの顔が笑っていることを見つけると共に、ユーザの音声を分析して笑い声を見つけることにより、コンピュータによって自動的に分類することができる。笑いがない場合、又は、ユーザが否定的な反応を示した場合、このこともまた、コンピュータによって自動的に検出することができ、このことは、ユーザがこのジョークを面白いと思わなかったことを暗に示し得る。このジョークテストを多数の人に送信する実験の規模により、該システムは、このジョークがどれほど面白かったかについての統計値を収集することができる。この実験の結果は、このジョークが面白かったかどうか（構造化された理解の一部）を知り、このジョークがどれほど面白かったかを統計的に知る（構造化された理解の程度を測定する）インサイトである。

【００９９】

図７は、様々な実施形態に係る、仮想キャラクター反応からインサイトを抽出するフロー図７００を示す。多くの種類の情報が、人間のユーザが仮想キャラクターと対話する際に収集することができる。人間は、典型的には、仮想キャラクターに話しかけることになる。人間はまた、ボディランゲージを通じて、自分がどのように感じているかを伝えることもできる。人々が話していることのテキスト、及び、人々が使用する身体的反応又はボディランゲージは、我々が注目する２つのタイプの人間のコミュニケーションであり得る。本実施形態は、センサによって検出され、コンピュータによる解釈のためにデータ信号に変換されることが可能な、任意の他のタイプのコミュニケーションに適用可能である。

【０１００】

現実世界からの刺激をこのデータのデジタル表現へと変換して、該デジタル表現に基づいて仮想キャラクターが決定を下すための同じプロセスは、そのデータについてのインサイトを抽出するために使用されてもよい。該システムは、人間のユーザに対してどのように反応するかを決定する際に、仮想キャラクターがアクセス可能なものと同じデータからインサイトを抽出してもよい。本質的に、人間のユーザは（話すことによって、又は物理的なジェスチャーを行うことによって）反応する可能性があり、この現実世界の刺激は、入力センサアレイによって検出されて、生のデジタル信号へと変換される。信号処理は、この信号に対して（１つのモード又は複数のモードのいずれかにおいて）実行されてもよい。その結果、クリーンにされた情報信号が得られ得る。この信号は、少なくとも２つの目的に使用され得る。第１に、該情報信号は、後でこのデータからインサイトを抽出できるように、長期ストレージ内に保存され得る。第２に、該情報信号は、キャラクターが現実世界の刺激にどのように反応するかを決定できるように、仮想キャラクターの意思決定機構（decision-making facilities）へと送信され得る。

【０１０１】

この文脈において、インサイトを抽出することは、現実世界の出来事を、構造化された理解に基づいて仮想キャラクターが決定を下すための該構造化された理解へと変換することに類似していることがある。クリーンな情報信号を取得するためのデータ処理パイプラインは、複数のユースケースで同じであり得る。仮想と対話している間におけるユーザの反応についての情報が保存されると、該対話についてのインサイトを抽出可能である。ここでのインサイトは、洗練されていないデータセットから抽出された理解又はより高レベルの情報であり得る。

【０１０２】

本実施形態は、仮想キャラクターとの対話からインサイトを抽出する能力に関する。このことは、テキストストリームからインサイトを抽出することとは異なることがある。仮想キャラクターとの対話は、より構造化されており、マルチモーダルデータ信号（発話、視覚データストリーム、嗅覚データストリーム等）を生成するためである。

【０１０３】

例えば、デバイスは、ユーザ７０２によって提供される刺激を示す入力センサアレイ７０４をキャプチャすることができる。生のデジタル信号のアレイ７０６は、マルチモーダルデジタル信号処理７０８を使用して処理することができる。クリーンにされた（又は処理された）情報信号７１０は、インサイト抽出７１２のいずれかを介して処理されると共に、ユーザデータストレージ７１６内に格納され、かつ／又は、仮想キャラクター反応７１４を導出するための意思決定ロジックで処理され、それによって、ユーザデバイス７１８上に仮想キャラクター７２０による反応を出力することができる。

【０１０４】

図８は、様々な実施形態に係る、インサイトダッシュボード８００を示す。図８に示すように、クライアントは、仮想キャラクターと対話するユーザ達を表すオーディエンスに関する多くのインサイトを表示するために、ダッシュボード又はインサイトポータルを表示してもよい。ユーザ達に関連する様々な情報を受信することにより、該システムは、ユーザの様々な機能を識別することができる。例えば、該システムは、仮想キャラクターが採るアクションに対する反応を表す、ユーザの顔面表情又はユーザの語気を識別することができる。受け取られた情報には、ユーザについての周囲環境もまた含まれる。

【０１０５】

インサイトダッシュボードは、ユーザの数、新規ユーザの数、一日平均ユーザの数、及び、総対話時間等の、情報のレポートを表示してもよい。インサイトダッシュボードはまた、アクティブな注視時間、発話の冗長性、購入情報、アイテム選択情報、感情情報等の、インサイト情報のグラフィック表現を含んでいてもよい。

【０１０６】

いくつかの実施形態では、仮想キャラクターは、ユーザデバイスを介したクライアントとユーザとの間の様々な取引を促進することができる。例えば、仮想キャラクターは、ユーザがアイテムを購入したいことを識別することができ、仮想キャラクターは、ユーザを特定のウェブページに転送してもよく、取引を直接促進してもよい。

【０１０７】

インサイトは、仮想キャラクターとの対話からのユーザデータについて、感情分析及び他のアルゴリズムを実行することができる。例えば、キャラクターは、別のキャラクターからメッセージを仮想的に受信して、彼らがチームを組んで映画に出演することについてどのように感じているかをユーザに尋ねてもよい。インサイトは、総回答率を識別することができる（例えば、インサイトは、会社に対して、オーディエンスの７８％が、これは良いアイデアだと考えていることを示す）。

【０１０８】

いくつかの実施形態では、方法は、仮想キャラクターと対話するユーザ達に関連付けられたデバイスのグループのそれぞれにおいて、仮想キャラクターを生成することを含んでいてもよい。該方法は、デバイスのグループと対話するユーザ達に関連する入力情報を受信することを含んでいてもよい。この方法は、各デバイス上に、少なくとも２つの内部モデルを実装することを含んでいてもよい。内部モデルは、入力情報の特性を識別して、該内部モデルによって識別された特性を組み合わせ、それによって、該内部モデルによって識別された特性間の類似度を識別するように構成されていてもよい。デバイスのグループ内の各デバイスに関連付けられた特性について、該方法は、仮想キャラクターに関連付けられた潜在的なアクションのライブラリを検査し、それにより、内部モデルによって識別された特性に整合するアクションを決定することを含んでいてもよい。デバイスのグループ内の各デバイスについて、該方法は、仮想キャラクターにアクションを実行させるための命令を表す、内部モデルによって識別された特性に整合するアクションを出力することを含んでいてもよい。

【０１０９】

該方法は、仮想キャラクターとデバイスのグループに関連付けられたユーザ達との間の対話に関連する情報を取得することを含んでいてもよい。該方法は、対話に関して取得された情報に基づいて、仮想キャラクターとユーザ達との間の対話の特徴を表すレポートを

生成することを含んでいてもよい。
【０１１０】
いくつかの実施形態では、レポートは、仮想キャラクターと対話するユーザの数、一日平均ユーザの数、及び、ユーザによる平均対話時間を含む。
【０１１１】
いくつかの実施形態では、該方法は、仮想キャラクターと対話するユーザ達に関連付けられた、アクティブな注視時間、音声の冗長性、アイテム選択情報及び感情情報のうちの少なくとも１つのグラフィック表現を生成することを含む。ここで、該仮想表現は、レポート内に含まれている。
【０１１２】
例えば、図８に示されるように、インサイトレポートは、総ユーザインジケータ８０２、先月の新規ユーザインジケータ８０４、過去２４時間の新規ユーザインジケータ８０６、一日平均ユーザのインジケータ８０８、月間平均ユーザインジケータ８１０、総対話時間インジケータ８１２、時間対エクスペリエンスのセグメントを示す平均注視時間グラフ８１４、時間対エクスペリエンスのセグメントを示す発話冗長性グラフ８１６、仮想キャラクターとの対話に応答してユーザが採る後続のアクション（例えば、映画のチケットの購入８１８、アイテムの選択８２０）、仮想キャラクターの感情８２２等のうち任意のものを含むことができる。
【０１１３】
仮想キャラクターレイヤー
本実施形態は、仮想キャラクターが多くのアプリケーションで実装されることを可能にし得る。例えば、仮想キャラクターは、店舗内において顧客が商品を見つけるのに役立つこと、ホームセンターの顧客が新しいペンキを選ぶのに役立つこと、又は、タッチスクリーンベースの現金自動預け払い機をよりパーソナライズされたフレンドリーな顔を備えたものに置き換えることができる。いくつかの実施形態では、現実世界における仮想キャラクターの所在地は、マッピングアプリケーション内に表示され得る。
【０１１４】
いくつかの実施形態では、インサイトは、データを機械学習データセットにフィードバックして、結果を自動的に改善する改善システムを含み得る。例えば、仮想キャラクターが１００万人に対して自撮りを依頼したと仮定する。該システムは、これら全ての写真を使用して、自撮りカメラの写真からユーザの顔を検出するための新たなアルゴリズムを作成できるであろう。
【０１１５】
図９は、様々な実施形態に係る、マルチモーダルモデルを使用して仮想キャラクターを制御する方法のブロック図を示す。該方法は、発話情報、顔面表情情報、環境情報等の入力情報を受信することを含み得る（ブロック９０２）。受信される情報は、マルチモーダル（聴覚情報（audio）、視覚情報（visual）など）であり得る。
【０１１６】
該方法は、仮想キャラクターを環境内の位置に表示することを含み得る（ブロック９０４）。仮想キャラクターは、センサ／カメラによってキャプチャされた環境等、実際の環境に重ねられ又は追加され得る。いくつかの実施形態では、仮想キャラクターは、環境内に表示された際に、第１のアクションを出力してもよい。第１のアクションは、例えば、ユーザへの質問が含まれ得る。
【０１１７】
該方法は、入力情報を処理するための内部モデルを実装することを含み得る（ブロック９０６）。内部モデルは、ユーザが発話した言葉、ユーザの表情等、ユーザに関する様々な情報を決定するために、入力情報を処理し得る。
【０１１８】
該方法は、第１の識別された特性が第２の識別された特性に対する閾値類似度内にあるかどうかを決定するために、少なくとも２つの内部モデルによって識別された特性を検査

することを含み得る（ブロック９０８）。特性間の類似度を識別することは、仮想キャラクターに共通して関連付けられている特性と整合する特性を決定するために内部モデルを参照することを含み得る。

【０１１９】

一例として、第１の識別された特徴は、ユーザによって発話された言葉が「私は悲しい」であると識別することを含み得る。第２の識別された特徴は、ユーザが笑顔又は幸せであることを示す顔面表情を識別することを含み得る。したがって、これらの特性を比較すると、これらの特性が閾値類似度の範囲内にないと決定されることがある。以下に示すように、知識モデルは、どの対照的な特性がより正確である可能性が高いかを識別するために使用することができる。例えば、知識モデルは、仮想キャラクターによる以前のアクションが、ユーザからの肯定的な反応を求める可能性が最も高いことを識別する。この例では、ユーザが「悲しい」ではなく「嬉しい」と言った可能性があるため、ユーザが笑っているという特性がより正確であって、第１の特性は正しくない可能性がある。

【０１２０】

該方法は、選択された特性を識別するために、第１の識別された特性及び第２の識別された特性を、仮想キャラクター知識モデルに含まれる仮想キャラクターに特有の情報と比較することを含み得る（ブロック９１０）。知識モデルは、ユーザ達による共通の応答、仮想キャラクターに関連する用語、仮想キャラクターに関連することが多い他のキャラクター等、仮想キャラクターに特有の情報を含むことができる。

【０１２１】

該方法は、仮想キャラクターによる潜在的なアクションのライブラリにアクセスすることを含み得る（ブロック９１２）。該ライブラリは、特定の仮想キャラクターによる潜在的なアニメーション／スピーチを表す情報を含むデータベース、レジストリ、リスト等を含み得る。潜在的なアクションは、実行時に生成モデルを使用して作成することもできる。

【０１２２】

該方法は、処理された入力情報に基づいて、仮想キャラクターによって実行されるアクションを決定することを含み得る。マルチモーダルモデルは、ユーザに関連する様々な情報を処理し、様々な情報を組み合わせることによって、アクションをより正確に決定し得る。アクションを決定することは、処理された入力情報から識別された正確なアクションに最も近く整合するライブラリ内のアクションを識別することを含むことができる。

【０１２３】

該方法は、環境内でアクションを実行する仮想キャラクターを表示することを含み得る（ブロック９１４）。いくつかの実施形態では、該方法は、ユーザデバイスによる選択の識別に基づいて、仮想キャラクターを動的に生成することを含み得る（ブロック９１６）。動的に生成することは、仮想キャラクターを短い遅延（例えば、１秒未満）で制御することを含み得る。ユーザデバイスによる選択は、ユーザデバイスと対話するユーザによるハイパーリンクのクリックを含むことができる。

【０１２４】

該方法は、クライアントについてのインサイトレポートを生成することを含み得る（ブロック９１８）。インサイトレポートは、ユーザデバイスで生成された仮想キャラクターと対話するユーザ達についての様々な情報を含むことができる。

【０１２５】

第１の例示的な実施形態では、仮想キャラクターを制御するための方法が開示される。該方法は、マルチモーダル入力情報をデバイスから受信するステップであって、マルチモーダル入力情報が、発話情報、顔面表情情報、及び、デバイスを取り巻く環境を表す環境情報のいずれかを含むステップを含むことができる。該方法はまた、デバイスに提示される表示環境内の位置に、仮想キャラクターを表示するステップを含むことができる。該方法はまた、少なくとも２つの内部モデルを実装するステップであって、それによってマルチモーダル入力情報の特性を識別するステップを含むことができる。

【０１２６】
該方法はまた、少なくとも２つの内部モデルの識別された特性を検査するステップであって、それによって、識別された特性のうち第１の識別された特性が、識別された特性のうち第２の識別された特性における類似の特徴の閾値数を含むかどうかを決定するステップを含むことができる。該方法はまた、第１の識別された特性及び第２の識別された特性を、仮想キャラクター知識モデルに含まれる仮想キャラクターに特有の情報と比較するステップであって、それによって、第１の識別された特性が識別された特性のうち第２の識別された特性における類似の特徴の閾値数を含むと決定することに基づいて、選択された特性を選択するステップを含むことができる。
【０１２７】
該方法はまた、仮想キャラクターに関連付けられた潜在的なアクションのライブラリにアクセスするステップであって、それによって、選択された特性に整合するアクションを決定し、アクションは、仮想キャラクターによって実行されることとなるアニメーションと、該アニメーションに関連付けられたオーディオとの両方が含まれるステップを含むことができる。該方法はまた、デバイス上に提示された環境内におけるF仮想キャラクターを修正すると共に関連付けられたオーディオを出力することによって、決定されたアクションを実行するステップを含むことができる。
【０１２８】
いくつかの実施形態では、少なくとも２つの内部モデルは、発話情報から発話感情を解析可能な発話認識モデルと、顔面表情情報に基づいて顔面特徴感情を検出可能な顔面特徴認識モデルとを含み、選択された特性は、発話感情と顔面特徴感情との間で共通の感情であり、決定されたアクションは、感情に基づいて決定される。
【０１２９】
いくつかの実施形態では、少なくとも２つの内部モデルは、ユーザとの以前のエンゲージメントに関連する情報を含む事前知識情報を引出し可能な事前知識モデルを含んでおり、選択された特性は、事前知識モデルを使用して処理された事前知識情報に基づいて選択される。
【０１３０】
いくつかの実施形態では、内部モデルは、聴覚情報からコンテキスト及び意味を導出するように構成された自然言語理解モデル、環境情報を識別するように構成された認識モデル、及び、ユーザ及び他の仮想キャラクターに関連するデータを識別するように構成されたソーシャルシミュレーションモデルを含む。
【０１３１】
いくつかの実施形態では、該方法は、仮想キャラクターに対して、デバイス上におけるユーザへの質問を表す初期アクションを実行するように指示するステップであって、入力情報は、ユーザによる質問に対する応答を表すステップを含むことができる。
【０１３２】
いくつかの実施形態では、該方法は、ネットワークを介して複数のユーザと埋め込みリンクを共有するステップを含むことができる。該方法はまた、デバイスのセットのいずれかから、埋め込みリンクが選択されたことを示す選択を受信するステップを含むことができる。該方法はまた、選択の受信に応答して、選択を送信したデバイスのセットのうちデバイスへとデータのストリームを送信するステップであって、それによって、仮想キャラクターをデバイス上に表示するステップを含むことができる。
【０１３３】
いくつかの実施形態では、該方法は、第１の時間に、データのストリームの第１のバッチを送信するステップであって、第１のバッチは、デバイスのディスプレイ上に仮想キャラクターを最初に生成するための情報を含むステップを含むことができる。該方法はまた、第１の時間の後の第２の時間に、データのストリームの第２のバッチを送信するステップであって、第２のバッチは、仮想キャラクターによる第１のアクションを出力するための情報を含み、第２の時間では、第１のバッチが破棄されるステップを含むことができる

。

【０１３４】

いくつかの実施形態では、該方法は、環境情報を検査するステップであって、それによって、環境の床を表す環境の部分を識別するステップを含むことができる。該方法はまた、仮想キャラクターを、環境の床を表す環境の部分の上方の第１の位置に配置するステップを含むことができる。

【０１３５】

第２の例示的な実施形態では、デバイスによってキャプチャされたユーザに関連するマルチモーダル入力に対する応答を提供するように構成された該デバイスが提供される。該デバイスは、少なくとも１つのメモリを含むことができる。該メモリは、マルチモーダル入力情報から特性を識別するように構成された、少なくとも２つの内部モデルを含むことができる。該メモリはまた、仮想キャラクターに特有の情報を含む仮想キャラクター知識モデルを含むことができる。該メモリはまた、仮想キャラクターに関連付けられた潜在的アクションのライブラリであって、各アクションは、仮想キャラクター及びそれに関連付けられたオーディオによって実行されるアニメーションに関連付けられているライブラリを含むことができる。

【０１３６】

該デバイスは、少なくとも１つのプロセッサを含むことができる。該少なくとも１つのプロセッサは、発話情報、顔面表情情報、及び、環境を表す環境情報のうち少なくとも１つを含むマルチモーダル入力情報を受信するステップを行うように構成されることができる。該少なくとも１つのプロセッサはまた、少なくとも２つの内部モデルによって識別された特性を検査するステップであって、それによって、第１の識別された特性が、第２の識別された特性との閾値類似度の範囲内にあるかどうかを決定するステップを行うように構成されることができる。該少なくとも１つのプロセッサはまた、第１の識別された特性及び第２の識別された特性を、仮想キャラクター知識モデルと比較するステップであって、それによって、選択された特性を識別するステップを行うように構成されることができる。

【０１３７】

該少なくとも１つのプロセッサはまた、仮想キャラクターに関連付けられた潜在的アクションのライブラリを検査することによって、選択された特性に整合するアクションを決定するステップであって、アクションは、デバイス上に出力されることになるオーディオを含む、ステップを行うように構成されることができる。該少なくとも１つのプロセッサはまた、オーディオをデバイス上に出力するステップを行うように構成されることができる。

【０１３８】

いくつかの実施形態では、該少なくとも１つのプロセッサは、さらに、環境情報から導出された環境内の位置内において、デバイスのディスプレイ上に仮想キャラクターを表示するステップを行うように構成されることができる。該少なくとも１つのプロセッサはまた、仮想キャラクターをデバイス上に表示される環境内において修正することにより、デバイスに出力されることとなるオーディオと、仮想キャラクターによって実行されることとなる選択されたアニメーションとの両方を含むアクションを実行するステップを行うように構成されることができる。

【０１３９】

いくつかの実施形態では、少なくとも２つの内部モデルは、ユーザとの以前のエンゲージメントに関連する情報を含む事前知識情報を引出し可能な事前知識モデルを含んでおり、選択された特性は、事前知識モデルを使用して処理された事前知識情報に基づいて選択される。

【０１４０】

いくつかの実施形態では、少なくとも２つの内部モデルは、発話情報から発話感情を解析可能な発話認識モデルと、顔面表情情報に基づいて顔面特徴感情を検出可能な顔面特徴

認識モデルとを含み、選択された特性は、発話感情と顔面特徴感情との間で共通の感情であり、決定されたアクションは、感情に基づいて決定される。

【０１４１】

第３の例示的な実施形態では、ユーザデバイスのウェブブラウザ上で仮想キャラクターを動的に生成するためのコンピュータ実装される方法が提供される。コンピュータ実装される方法は、ユーザデバイスのウェブブラウザにリンクを埋め込むステップであって、リンクは、ウェブブラウザをユーザデバイス上で実行されているアプリケーションにリンクさせるステップを含むことができる。該コンピュータ実装される方法はまた、ユーザデバイスから、リンクが選択されたという通知を受信するステップを含むことができる。該コンピュータ実装される方法はまた、情報を表すアプリケーションからのデータのストリームをウェブブラウザへと送信するステップであって、それによって、仮想キャラクターを生成するステップを含むことができる。

【０１４２】

該コンピュータ実装される方法はまた、仮想キャラクターを、ユーザデバイスのウェブブラウザ上に表示するステップを含むことができる。該コンピュータ実装される方法はまた、ユーザデバイスからマルチモーダル入力情報を受信するステップであって、マルチモーダル入力情報が、発話情報、顔面表情情報、及び、環境を表す環境情報を含むステップを含むことができる。該コンピュータ実装される方法はまた、少なくとも２つの内部モデルを実装するステップであって、それによってマルチモーダル入力情報の特性を識別するステップを含むことができる。

【０１４３】

該コンピュータ実装される方法はまた、少なくとも２つの内部モデルによって識別された特性を検査するステップであって、それによって、第１の識別された特性が、第２の識別された特性との閾値類似度の範囲内にあるかどうかを決定するステップを含むことができる。該コンピュータ実装される方法はまた、第１の識別された特性及び第２の識別された特性を、仮想キャラクター知識モデルに含まれる仮想キャラクターに特有の情報と比較するステップであって、それによって、第１の識別された特性が識別された特性のうち第２の識別された特性における類似の特徴の閾値数を含むと決定することに基づいて、選択された特性を選択するステップを含むことができる。

【０１４４】

該コンピュータ実装される方法はまた、仮想キャラクターに関連付けられた潜在的なアクションのライブラリにアクセスするステップであって、それによって、選択された特性に整合するアクションを選択し、アクションは、仮想キャラクターによって実行されることとなるアニメーションと、該アニメーションに関連付けられたオーディオとが含まれるステップを含むことができる。該コンピュータ実装される方法はまた、環境内においてアクションを実行している仮想キャラクターを表示すると共に関連付けられたオーディオを出力するステップを含むことができる。

【０１４５】

いくつかの実施形態では、ウェブブラウザは、ユーザデバイス上で実行されるモバイルアプリケーション上に表示されるページを含む。

【０１４６】

いくつかの実施形態では、該コンピュータ実装される方法は、選択された特性とアクションとに関連する情報を格納するステップを含むことができる。該コンピュータ実装される方法はまた、複数のユーザについての一連の選択された特性及びアクションを集約するステップを含むことができる。該コンピュータ実装される方法はまた、複数のユーザについての一連の選択された特性及びアクションを処理するステップであって、それによって、仮想キャラクターと複数のユーザとのエンゲージメントに関連する分析のセットを導出するステップを含むことができる。該コンピュータ実装される方法はまた、仮想キャラクターと複数のユーザとのエンゲージメントに関連する分析のセットを表示する分析ダッシュボードを提示するステップを含むことができる。

【０１４７】
処理システムの例
図１０は、本明細書に記載の動作の少なくともいくつかを実装可能な処理システム１０００の例を示すブロック図である。例えば、処理システム１０００のいくつかのコンポーネントは、本実施形態に記載されるように、電子デバイス上でホストされることができる。
【０１４８】
処理システム１０００は、１つ以上の中央処理ユニット（「プロセッサ」）１００２と、メインメモリ１００６と、不揮発性メモリ１０１０と、ネットワークアダプタ１０１２（例えば、ネットワークインタフェース）と、ビデオディスプレイ１０１８と、入力／出力デバイス１０２０と、制御デバイス１０２２（例えば、キーボード及びポインティングデバイス）と、記憶媒体１０２６を含むドライブユニット１０２４と、バス１０１６に通信可能に接続された信号生成デバイス１０３０とを含むことができる。バス１０１６は、適切なブリッジ、アダプタ又はコントローラによって接続された１つ以上の物理バス及び／又はポイントツーポイント接続を表す抽象概念として図示されている。したがって、バス１０１６には、システムバス、周辺機器相互接続（Peripheral Component Interconnect、ＰＣＩ）バス又はＰＣＩ－Ｅｘｐｒｅｓｓバス、ＨｙｐｅｒＴｒａｎｓｐｏｒｔ又は業界標準アーキテクチャ（ＩＳＡ）バス、小型コンピュータシステムインタフェース（ＳＣＳＩ）バス、ユニバーサルシリアルバス（ＵＳＢ）、ＩＩＣ（Ｉ２Ｃ）バス、又は電気電子技術者協会（ＩＥＥＥ）の標準１３９４バス（すなわち、「Ｆｉｒｅｗｉｒｅ」）が含まれ得る。
【０１４９】
処理システム１０００は、デスクトップコンピュータ、タブレットコンピュータ、パーソナルデジタルアシスタント（ＰＤＡ）、スマートフォン、ゲームコンソール、音楽プレーヤ、ウェアラブル電子デバイス（例えば、時計又はフィットネストラッカー）、ネットワーク接続（「スマート」）デバイス（例えば、テレビ又はホームアシスタントデバイス）、仮想／拡張現実システム（例えば、ヘッドマウントディスプレイ）、又は、処理システム１０００によって行われることになる（１つ以上の）アクションを指定する命令のセットを（逐次的に又はそれ以外の方法で）実行可能な別の電子デバイスと同様のコンピュータプロセッサアーキテクチャを共有することができる。
【０１５０】
メインメモリ１００６、不揮発性メモリ１０１０及び記憶媒体１０２６（「機械可読媒体」とも称される）は、単一の媒体として示されているが、用語「機械可読媒体」及び「記憶媒体」は、１つ以上の命令のセット１０２８を記憶する単一の媒体又は複数の媒体（例えば、集中データベース若しくは分散データベース、並びに／又は、関連するキャッシュ及びサーバ）を含むと解されるべきである。用語「機械可読媒体」及び「記憶媒体」は、処理システム１０００により実行される命令のセットを格納、符号化又は運搬可能な任意の媒体を含むと解されるべきである。
【０１５１】
全体として、本開示の実施形態を実施するために実行されるルーチンは、オペレーティングシステム又は特定のアプリケーション、コンポーネント、プログラム、オブジェクト、モジュール若しくは命令シーケンス（あわせて「コンピュータプログラム」と称される）の一部として実装されることができる。コンピュータプログラムは、典型的には、コンピュータの様々なメモリ及び記憶デバイス内に、種々の時点にセットされた１つ以上の命令（例えば、命令１００４，１００８，１０２８）を含む。該（１つ以上の）命令は、１つ以上のプロセッサ１００２によって読み取られて実行された際に、処理システム１０００に、本開示の様々な態様を伴う要素を実行する動作を行わせる。
【０１５２】
さらに、実施形態は、完全に機能するコンピューティングデバイスの文脈にて記載されているが、当業者であれば、種々の実施形態が様々な形態のプログラム製品として配布可

能であることを理解するであろう。本開示は、実際に配布を行うために使用される機械又はコンピュータ可読媒体の特定のタイプにかかわらず適用される。

【０１５３】

機械可読記憶媒体、機械可読媒体又はコンピュータ可読媒体のさらなる例には、揮発性及び不揮発性メモリデバイス１０１０、フロッピー及び他のリムーバブルディスク、ハードディスクドライブ、光ディスク（例えば、コンパクトディスク読み取り専用メモリ（ＣＤ－ＲＯＭ）、デジタル多用途ディスク（ＤＶＤ））等のような記録可能型媒体、及び、デジタル及びアナログ通信リンクのような伝送型媒体が含まれる。

【０１５４】

ネットワークアダプタ１０１２は、処理システム１０００が、処理システム１０００及び外部エンティティによってサポートされる任意の通信プロトコルを通じて、処理システム１０００の外部にあるエンティティを用いてネットワーク１０１４内のデータを仲介することを可能にする。ネットワークアダプタ１０１２は、ネットワークアダプタカード、ワイヤレスネットワークインタフェースカード、ルータ、アクセスポイント、ワイヤレスルータ、スイッチ、マルチレイヤスイッチ、プロトコル変換器、ゲートウェイ、ブリッジ、ブリッジルータ、ハブ、デジタルメディアレシーバ及び／又はリピータを含むことができる。

【０１５５】

ネットワークアダプタ１０１２は、コンピュータネットワーク内のデータにアクセス／プロキシするためのパーミッションを統御及び／又は管理してもよく、かつ、種々異なるマシン及び／又はアプリケーションの間の様々なレベルの信頼を追跡するファイアウォールを含むことができる。ファイアウォールは、特定のマシンのセットとアプリケーションとの間、マシンとマシンとの間、及び／又は、アプリケーションとアプリケーションとの間で（例えば、これらのエンティティ間のトラフィック及びリソース共有のフローを規制するために）所定のアクセス権のセットを実施可能なハードウェア及び／又はソフトウェアコンポーネントの任意の組み合わせを有する、任意の数のモジュールであり得る。ファイアウォールはさらに、個人、マシン及び／又はアプリケーションによるオブジェクトのアクセス権及び操作権を含むパーミッション、並びに、パーミッション権が有効である状況について詳述するアクセス制御リストを管理する、及び／又は、該アクセス制御リストにアクセスすることができる。

【０１５６】

本明細書にて提示される技術は、プログラム可能な回路（例えば、１つ以上のマイクロプロセッサ）、ソフトウェア及び／若しくはファームウェア、専用のハードワイヤード（すなわち、プログラム不可能な）回路、又は、そのような形態の組み合わせによって実装され得る。専用の回路は、１つ以上の特定用途向け集積回路（ＡＳＩＣ）、プログラマブルロジックデバイス（ＰＬＤ）、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）等の形式であり得る。

【０１５７】

本明細書にて開示される技術は、製品製造プロセスの能力を補足するために設計図面から情報を抽出するためのツール又はシステムを含むことができる。本実施形態は、見積もり／概算の生成／修正、製造フィードバックの生成、製造スケジューリング／注文プロセス等を支援することができる。

【０１５８】

備考

文脈が明らかに他を要求しない限り、単語「含む（comprise）」、「含んでいる（comprising）」及びこれに類するものは、排他的又は網羅的な意味ではなく、包括的意味で（すなわち、「含むが限定されない」という意味で）解釈されるべきである。本明細書にて使用される場合、用語「接続（connect）」、「結合（couple）」又はそれらの任意の変形は、２つ以上の要素間の直接的又は間接敵な任意の接続又は結合を意味し、これらの要素間の結合又は接続は、物理的なもの、論理的なもの、又は、それらの組み合わ

せであり得る。加えて、単語「本明細書では（herein）」、「上記の（above）」、「下記の（below）」及び同様の意味の単語は、本出願において使用される場合、本出願の特定の部分ではなく、本出願全体を示す。文脈が許容する場合、上記詳細な説明における単数形又は複数形を使用する単語はまた、それぞれ複数形又は単数形を含んでいてもよい。２つ以上の項目のリストに関する単語「又は（or）」は、リスト内の項目のいずれか、リスト内の項目の全て、及び、リスト内の項目の任意の組み合わせという単語の解釈の全てを網羅する。

【０１５９】

本明細書で使用される場合、用語「実質的に」は、アクション、特性、特質、状態、構造、項目又は結果の、完全又はほぼ完全な範囲又は程度を示す。例えば、「実質的に」囲まれているオブジェクトとは、オブジェクトが完全に囲まれているか、ほぼ完全に囲まれていることを意味する。絶対的な完全性からの逸脱が正確にどの程度まで許容可能であるかは、場合により、特定の文脈に依存することがある。しかしながら、一般的に言えば、完全性の近さは、絶対的かつ完全な完全性が得られたかのように、同じ全体的な結果を有するようになるであろう。「実質的に」の使用は、アクション、特性、特質、状態、構造、項目又は結果の完全又はほぼ完全な欠如を示す否定的な意味で使用される場合にも等しく適用可能である。

【０１６０】

本発明の例の上記の詳細な説明は、網羅的であること又は本発明を上記開示の正確な形態に限定することを意図するものではない。本発明についての特定の例は、例示の目的で上述されているが、関連技術の当業者が認識するように、本発明の範囲内において様々な均等の修正が可能である。例えば、プロセス又はブロックは特定の順序で提示されているが、代替の実装では、ステップを含むルーチンを実行してもよく、ブロックを含むシステムを異なる順序で採用してもよく、いくつかのプロセス又はブロックが削除、移動、追加、細分化、結合及び／又は代替又は修正されて、代替のもの又はサブコンビネーションが提供されてもよい。これらのプロセス又はブロックのそれぞれは、種々異なる方法で実装されてもよい。また、プロセス又はブロックは、直列に実行されているように示されている場合があるが、これらのプロセス又はブロックは、代わりに並列に実行又は実装されてもよく、異なる時間に実行されてもよい。さらに、本明細書に記載される特定の数値は単なる例示であって、代替の実装では、異なる値又は範囲が採用されてもよい。

【０１６１】

本明細書で提供される本発明の教示は、他のシステムに適用可能であり、必ずしも上記のシステムに適用されなくてもよい。上記の様々な例の要素及び行為は、本発明のさらなる実装を提供するために組み合わせることができる。本発明のいくつかの代替の実装は、上記の実装に対して追加の要素を含んでいてもよいが、それに限らず、より少ない要素を含んでいてもよい。

【０１６２】

これら及び他の変更は、上記の詳細な説明に照らして、本発明に対して行うことができる。上記の説明は、本発明の特定の例を説明し、企図される最良の形態について記載しているが、上記が文章においてどれほど詳細に示されていても、本発明は多くの方法で実施することができる。システムの詳細は、その特定の実装において相当に変化することがあるが、依然として本明細書に開示の本発明に包含される。上記のように、本発明の特定の機能又は態様を説明する際に使用される特定の用語は、本明細書において、関連する本発明の特定の特徴、機能、又は態様に限定されるようにその用語が再定義されることを暗に意味するものとして理解されるべきではない。全体として、以下の特許請求の範囲にて使用される用語は、上記の詳細な説明のセクションでそのような用語が明示的に定義されていない限り、本発明を本明細書に開示される特定の例にまで限定するように解釈されるべきではない。したがって、本発明の実際の範囲は、開示の例だけでなく、特許請求の範囲に基づいて本発明を実施又は実装する全ての均等な方法を包含する。

【図面】

【図１】

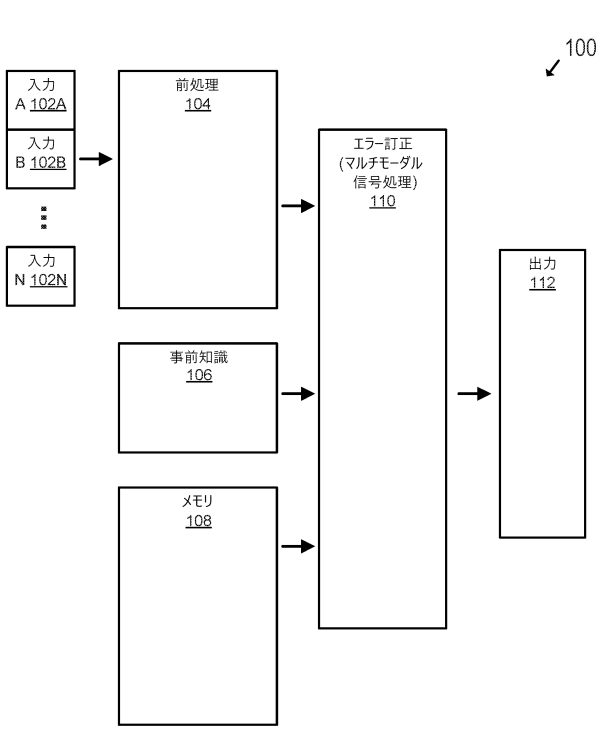


FIGURE 1

【図２】

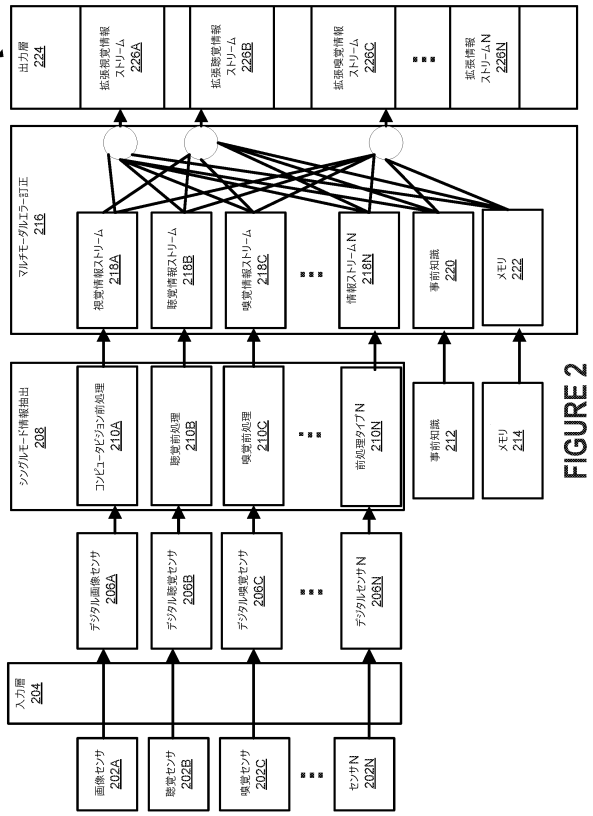


FIGURE 2

【図３】

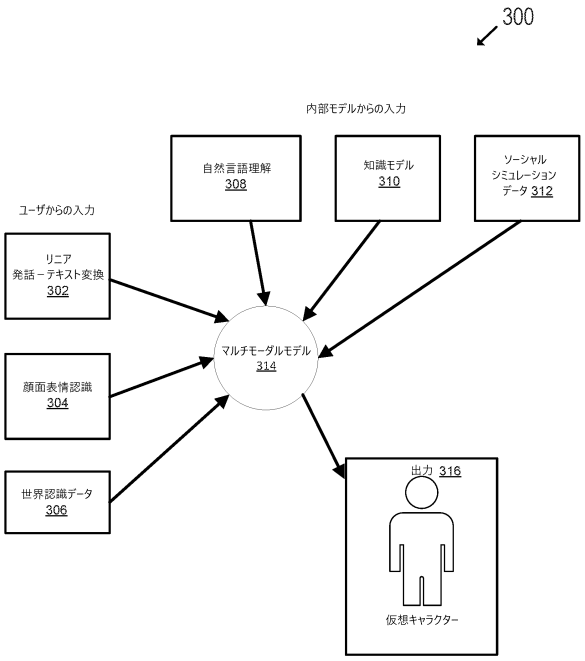


FIGURE 3

【図４】

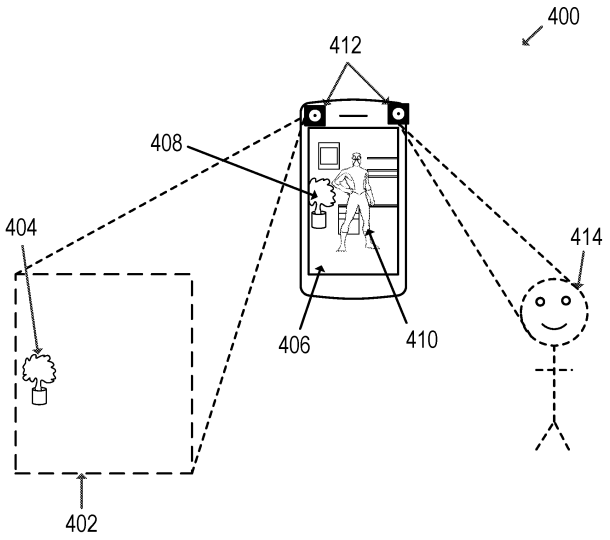


FIGURE 4

【図５】

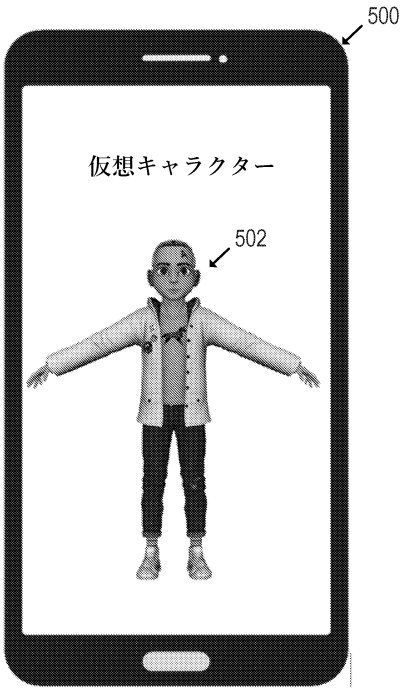


FIGURE 5

【図６】

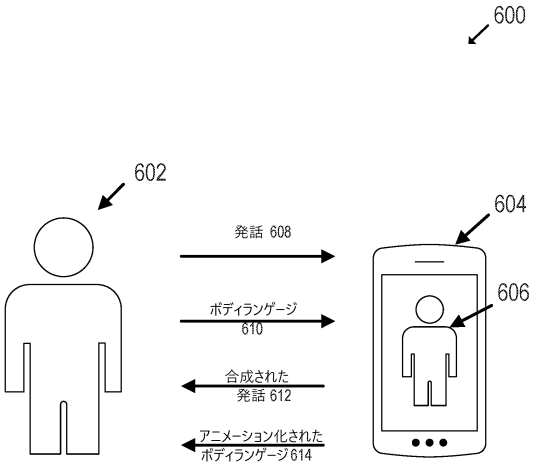


FIGURE 6

【図７】

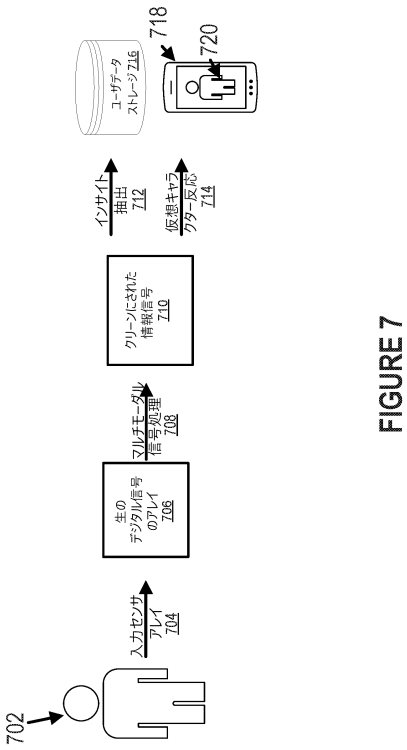


FIGURE 7

【図８】

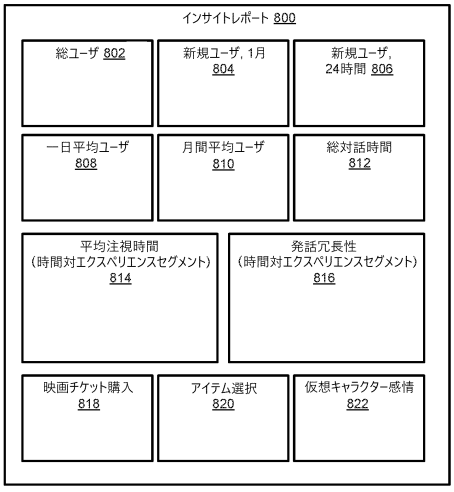


FIGURE 8

【図９】

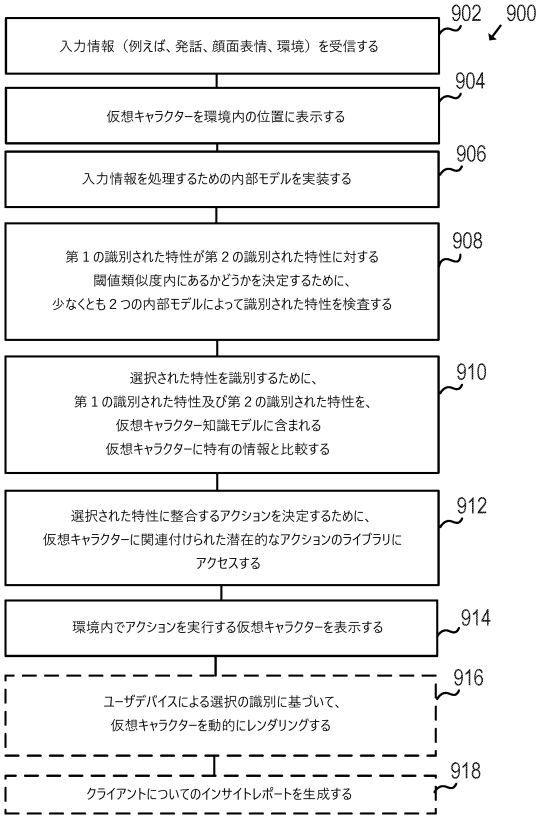


FIGURE 9

【図１０】

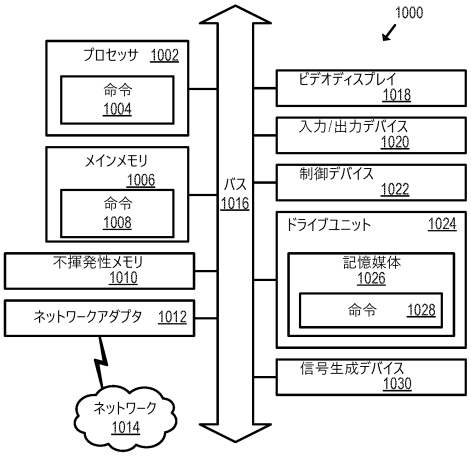


FIGURE 10

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/036068

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 3/00; G06T 13/00; G10L 15/00; G10L 15/02 (2020.01)
CPC - G06T 13/205; G06F 3/012; G06T 13/20; G10L 15/02; G10L 15/1822 (2020.08)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2017/0039750 A1 (INTEL CORPORATION) 09 February 2017 (09.02.2017) entire document	1-20
A	US 2010/0082345 A1 (WANG et al) 01 April 2010 (01.04.2010) entire document	1-20
A	WO 2017/137947 A1 (VATS) 17 August 2017 (17.08.2017) entire document	1-20
A	US 2019/0156549 A1 (FACESHIFT AG) 23 May 2019 (23.05.2019) entire document	1-20
A	US 2014/0212854 A1 (SRI INTERNATIONAL) 31 July 2014 (31.07.2014) entire document	1-20
A	ZHANG et al., "Real-Time Adaptive Behaviors in Multimodal HumanAvatar Interactions", ICMI-MLMI'10, November 8–12, 2010. Retrieved on 07.08.2020. Retrieved from < URL: https://dl.acm.org/doi/pdf/10.1145/1891903.1891909> entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 August 2020	24 AUG 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

フロントページの続き

(51)国際特許分類	FI	テーマコード（参考）
G06T 13/20 (2011.01)	G06T 13/20 500	

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

（特許庁注：以下のものは登録商標）

１．ＦＩＲＥＷＩＲＥ

ークル ３０３３，ケア オブ アーティー，インコーポレイテッド

(72)発明者 ホーリガン，ライアン

アメリカ合衆国 ９００２６ カリフォルニア州，ロサンゼルス，バークレー サークル ３０３３，ケア オブ アーティー，インコーポレイテッド

(72)発明者 アイゼンバーグ，ジョシュ

アメリカ合衆国 ９００２６ カリフォルニア州，ロサンゼルス，バークレー サークル ３０３３，ケア オブ アーティー，インコーポレイテッド

Ｆターム（参考） 5B050 AA03 BA06 BA08 BA09 BA11 BA13 CA07 CA08 DA01 EA04 EA19 EA24 EA26 FA02 FA10 GA08

5E555 AA46 BA01 BB01 BC01 CB64 CB67 DA01 DA21 DB32 DB57 DC85 EA07 FA00