(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2012-181704
(P2012-181704A)
(43)公開日　平成24年9月20日(2012.9.20)

(51)Int.Cl.		F I			テーマコード（参考）
G06T 13/00	(2011.01)	G06T	13/00	A	2C001
G06T 13/80	(2011.01)	G06F	3/048	651A	5B050
G06F 3/048	(2006.01)	A63F	13/00	C	5E501
A63F 13/00	(2006.01)	A63F	13/00	E	
A63F 13/02	(2006.01)	A63F	13/02		

審査請求 未請求　請求項の数 9　OL　（全 21 頁）

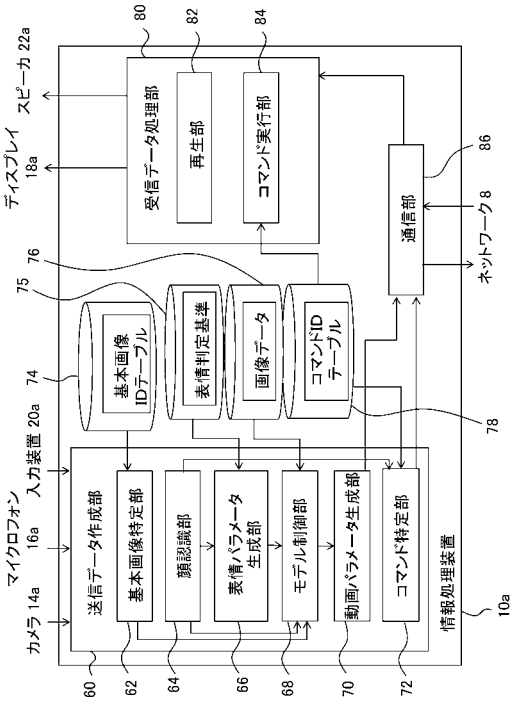
(21)出願番号　特願2011-44488(P2011-44488)
(22)出願日　平成23年3月1日(2011.3.1)

(71)出願人　310021766
株式会社ソニー・コンピュータエンタテインメント
東京都港区港南１丁目７番１号
(74)代理人　100105924
弁理士　森下　賢樹
(74)代理人　100109047
弁理士　村田　雄祐
(74)代理人　100109081
弁理士　三木　友由
(74)代理人　100134256
弁理士　青木　武司

最終頁に続く

(54)【発明の名称】情報処理装置および情報処理方法

(57)【要約】
【課題】ネットワークを介したリアルタイムな会話システムを容易な計算で実現する。
【解決手段】情報処理装置１０ａの送信データ作成部６０において、基本画像特定部６２は自ユーザを表すキャラクタの基本的な画像を特定する。顔認識部６４は自ユーザの顔認識を行う。表情パラメータ生成部６６は自ユーザの表情を数値化する。モデル制御部６８は、各時刻に対しキャラクタの出力モデルを決定する。動画パラメータ生成部７０は、キャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成する。コマンド特定部７２は、自ユーザの表情のパターンに対応するコマンドを特定する。受信データ処理部８０において、再生部８２は相手の情報処理装置から受信した動画パラメータおよび音声データに基づき音声および画像を出力する。コマンド実行部８４はコマンドの識別情報に基づきコマンドを実行する。
【選択図】図３



【特許請求の範囲】
【請求項１】
ユーザを表すキャラクタのモデルであり、異なる表情を有する複数の表情モデルのデータを格納した画像データ格納部と、
ユーザを撮影して得られる入力動画像データを逐次解析し、入力画像フレームごとに顔の部位の形状を表す数値を導出し、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、表情がなされている度合いを表情パラメータとして算出する表情パラメータ生成部と、
前記表情パラメータ生成部が算出した表情パラメータ、および、撮影と同時に取得した当該ユーザの音声データから得られる声量を用いて、前記画像データ格納部に格納された複数の表情モデルに対する重みを決定したうえで、当該複数の表情モデルを合成し、前記入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを決定するモデル制御部と、
前記モデル制御部が決定した出力モデルを含むキャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成する動画パラメータ生成部と、
前記動画パラメータ生成部が生成した動画パラメータと前記音声データとを同期させて逐次出力する出力部と、
を備えたことを特徴とする情報処理装置。
【請求項２】
前記画像データ格納部が格納する複数の表情モデルは、異なる音を発声したときの顔の状態を表す複数の発声モデルを含み、
前記モデル制御部は、前記複数の発声モデルに対する重みを異なる波形で時間変化させて合成することを特徴とする請求項１に記載の情報処理装置。
【請求項３】
前記異なる波形は、前記声量に比例した振幅を有し、周期の異なる正弦波の絶対値であることを特徴とする請求項２に記載の情報処理装置。
【請求項４】
前記発声モデルは、「あ」を発声したときの顔の状態を表す発声モデルと、「い」を発声したときの顔の状態を表す発声モデルと、「う」を発声したときの顔の状態を表す発声モデルであることを特徴とする請求項２または３に記載の情報処理装置。
【請求項５】
前記表情パラメータ生成部が生成する表情パラメータは口を開けている度合いを含み、前記モデル制御部は、前記発声モデルに対する重みの少なくともいずれかに、前記口を開けている度合いを加算することを特徴とする請求項２から４のいずれかに記載の情報処理装置。
【請求項６】
前記出力部は、ネットワークを介して他の情報処理装置へ、前記動画パラメータを逐次送信することを特徴とする請求項１から５のいずれかに記載の情報処理装置。
【請求項７】
ユーザを撮影して得られる入力動画像データを逐次解析し、入力画像フレームごとに顔の部位の形状を表す数値を導出し、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、当該表情がなされている度合いを表情パラメータとして算出するステップと、
算出した表情パラメータ、および、撮影と同時に取得した当該ユーザの音声データから得られる声量を用いて、メモリに記憶された、ユーザを表すキャラクタのモデルであり、異なる表情を有する複数の表情モデルに対する重みを決定するステップと、
メモリから前記複数の表情モデルのデータを読みだすステップと、
当該複数の表情モデルを前記重みによって重みづけして合成し、前記入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを決定するステップと、

前記出力モデルを含むキャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成するステップと、
前記動画パラメータと前記音声データとを同期させて逐次出力するステップと、
を含むことを特徴とする情報処理方法。
【請求項８】
ユーザを撮影して得られる入力動画像データを逐次解析し、入力画像フレームごとに顔の部位の形状を表す数値を導出し、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、当該表情がなされている度合いを表情パラメータとして算出する機能と、
算出した表情パラメータ、および、撮影と同時に取得した当該ユーザの音声データから得られる声量を用いて、メモリに記憶された、ユーザを表すキャラクタのモデルであり、異なる表情を有する複数の表情モデルに対する重みを決定する機能と、
メモリから前記複数の表情モデルのデータを読みだす機能と、
当該複数の表情モデルを前記重みによって重みづけして合成し、前記入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを決定する機能と、
前記出力モデルを含むキャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成する機能と、
前記動画パラメータと前記音声データとを同期させて逐次出力する機能と、
をコンピュータに実現させることを特徴とするコンピュータプログラム。
【請求項９】
ユーザを撮影して得られる入力動画像データを逐次解析し、入力画像フレームごとに顔の部位の形状を表す数値を導出し、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、当該表情がなされている度合いを表情パラメータとして算出する機能と、
算出した表情パラメータ、および、撮影と同時に取得した当該ユーザの音声データから得られる声量を用いて、メモリに記憶された、ユーザを表すキャラクタのモデルであり、異なる表情を有する複数の表情モデルに対する重みを決定する機能と、
メモリから前記複数の表情モデルのデータを読みだす機能と、
当該複数の表情モデルを前記重みによって重みづけして合成し、前記入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを決定する機能と、
前記出力モデルを含むキャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成する機能と、
前記動画パラメータと前記音声データとを同期させて逐次出力する機能と、
をコンピュータに実現させることを特徴とするコンピュータプログラムを記録した記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ユーザがネットワークを介して会話するための会話システムに用いる情報処理装置および情報処理方法に関する。
【背景技術】
【０００２】
遠隔地にいるユーザの画像および音声のデータをネットワークを介して相互に受信し、表示画像および音響として出力することにより、距離を感じずに会話することのできるテレビ会議システムが実用化されている。現在では、ウェブカメラを搭載したパーソナルコンピュータ、携帯端末、ゲーム機器の普及、ネットワークの拡充などの要因により、会議など特定の場面に限らず、より個人的に、時間、場所を問わず気軽に楽しめるビデオチャットやテレビ電話などの技術も一般的になってきた（例えば特許文献１参照）。
【先行技術文献】
【特許文献】

【０００３】
【特許文献１】米国公開２００９／２２２５７２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
上記技術を利用してより自然な会話を楽しむには、リアルタイム性を維持することが重要である。しかしながら質の高い画像および音声のデータをレイテンシなく伝送して出力するためには、相応のデータ処理能力、通信帯域が必要となる。そのため通信帯域に制約があったり、一つの情報処理装置で並行して何らかの処理を行っていたりするような環境においても、気軽に、かつ自然な会話を楽しむことのできる技術が望まれている。
【０００５】
本発明はこのような課題に鑑みてなされたものであり、その目的は、処理のためのリソースや通信帯域を圧迫することなく、相手ユーザとのリアルタイムな会話を楽しむことができる技術を提供することにある。本発明の別の目的は、そのような会話において娯楽性を高めることのできる技術を提供することにある。
【課題を解決するための手段】
【０００６】
本発明のある態様は情報処理装置に関する。この情報処理装置は、ユーザを表すキャラクタのモデルであり、異なる表情を有する複数の表情モデルのデータを格納した画像データ格納部と、ユーザを撮影して得られる入力動画像データを逐次解析し、入力画像フレームごとに顔の部位の形状を表す数値を導出し、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、表情がなされている度合いを表情パラメータとして算出する表情パラメータ生成部と、表情パラメータ生成部が算出した表情パラメータ、および、撮影と同時に取得した当該ユーザの音声データから得られる声量を用いて、画像データ格納部に格納された複数の表情モデルに対する重みを決定したうえで、当該複数の表情モデルを合成し、入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを決定するモデル制御部と、モデル制御部が決定した出力モデルを含むキャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成する動画パラメータ生成部と、動画パラメータ生成部が生成した動画パラメータと前記音声データとを同期させて逐次出力する出力部と、を備えたことを特徴とする。
【０００７】
本発明のさらに別の態様は情報処理方法に関する。この情報処理方法は、ユーザを撮影して得られる入力動画像データを逐次解析し、入力画像フレームごとに顔の部位の形状を表す数値を導出し、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、当該表情がなされている度合いを表情パラメータとして算出するステップと、算出した表情パラメータ、および、撮影と同時に取得した当該ユーザの音声データから得られる声量を用いて、メモリに記憶された、ユーザを表すキャラクタのモデルであり、異なる表情を有する複数の表情モデルに対する重みを決定するステップと、メモリから前記複数の表情モデルのデータを読みだすステップと、当該複数の表情モデルを重みによって重みづけして合成し、入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを決定するステップと、出力モデルを含むキャラクタのアニメーション動画のフレームを各時刻において生成するための動画パラメータを生成するステップと、動画パラメータと音声データとを同期させて逐次出力するステップと、を含むことを特徴とする。
【０００８】
なお、以上の構成要素の任意の組合せ、本発明の表現を方法、装置、システム、コンピュータプログラム、コンピュータプログラムを記録した記録媒体などの間で変換したものもまた、本発明の態様として有効である。
【発明の効果】
【０００９】
本発明によると、処理のためのリソースや通信帯域を圧迫することなく、相手ユーザと

のネットワークを介した自然な会話を楽しむことが可能となる。また、会話に娯楽性を含めることができる。

【図面の簡単な説明】

【００１０】

【図１】本実施の形態に適用できる会話システムの使用環境を示す図である。

【図２】本実施の形態における情報処理装置の内部回路構成を示す図である。

【図３】本実施の形態における情報処理装置の構成を詳細に示す図である。

【図４】本実施の形態において表示される画像の例を示す図である。

【図５】本実施の形態における基本画像ＩＤテーブルのデータ構造例を示す図である。

【図６】本実施の形態における基本画像ＩＤテーブルのデータ構造の別の例を示す図である。

【図７】本実施の形態における表情判定基準のデータ構造例を示す図である。

【図８】本実施の形態においてモデル制御部がキャラクタの顔を合成する様子を模式的に示す図である。

【図９】本実施の形態におけるコマンドＩＤテーブルのデータ構造例を示す図である。

【発明を実施するための形態】

【００１１】

図１は本実施の形態を適用できる会話システムの使用環境を示す。会話システム１は、ユーザ１２ａが操作する情報処理装置１０ａ、ユーザ１２ｂが操作する情報処理装置１０ｂなど複数の情報処理装置を含む。情報処理装置１０ａ、１０ｂがそれぞれ、インターネットやＬＡＮ（Local Area Network）などのネットワーク８に接続し、画像および音声のデータを相手の装置に送ることにより、ユーザ同士の会話を実現する。なお情報処理装置１０とネットワーク８との接続は同図のとおり無線通信であってもよいし、有線ケーブルを介していてもよい。

【００１２】

本実施の形態では、ネットワーク８を介して通信を確立した情報処理装置１０ａ、１０ｂをそれぞれ操作するユーザ１２ａ、１２ｂの実物の映像に代わり、それぞれのユーザを表す仮想のキャラクタをコンピュータグラフィックス（ＣＧ）で描画して相手に見せる。すなわち各ユーザは、自分が操作する装置上に表示された、会話相手の分身ともいえるキャラクタと会話する状態となる。

【００１３】

このキャラクタが、本来の会話相手の様子をリアルタイムに反映し、あたかもしゃべっているかのように音声と合致した口の動きをすることにより、実際に相手を見ながら会話しているのと同様の印象を与えることができる。また自分の本当の顔や居場所などを見られたくない場合でも画像を用いた会話を楽しむことができる。またＣＧならではの効果を加えて娯楽性を高めることもできる。なお同時に会話するユーザは３人以上でもよく、その場合、各装置に複数のキャラクタが表示されてよい。また自分を表すキャラクタを同時に表示するようにしてもよい。

【００１４】

情報処理装置１０ａ、１０ｂは、たとえば携帯端末、ゲーム装置、パーソナルコンピュータであってよく、それぞれ会話システムを実現するためのアプリケーションプログラムをロードすることで会話機能を実現する。情報処理装置１０ａの前面には、カメラ１４ａ、マイクロフォン１６ａ、ディスプレイ１８ａを備える。情報処理装置１０ａはそのほか、アプリケーションの起動やその後の各種指示をユーザ１２ａから受け付けるための入力装置２０ａ、音声を出力するスピーカ２２ａを備える。情報処理装置１０ｂも同様である。

【００１５】

同図ではわかりやすさのためにユーザ１２ａと情報処理装置１０ａ、ユーザ１２ｂと情報処理装置１０ｂとを離して示しているが、同図に示す情報処理装置１０ａ、１０ｂはそれぞれ、ユーザ１２ａ、１２ｂが左右から把持することによって各種操作を行うことので

きる携帯可能な端末を想定している。そのため、カメラ１４ａ、マイクロフォン１６ａ、ディスプレイ１８ａ、入力装置２０ａ、スピーカ２２ａは、情報処理装置１０ａ本体と一体的に設けられている。

【００１６】

一方、カメラ１４ａ、マイクロフォン１６ａ、ディスプレイ１８ａ、入力装置２０ａ、スピーカ２２ａのいずれかあるいはすべてが、情報処理装置１０ａ本体と別の筐体を有し、有線あるいは無線で情報処理装置１０ａ本体と接続されていてもよい。たとえば入力装置２０ａはキーボード、マウス、トラックボール、リモートコントローラなどで実現してもよいし、ディスプレイ１８ａ表面や情報処理装置１０ａの外面に設けられたタッチパネルなどの形態でもよい。またディスプレイ１８ａとスピーカ２２ａとを、一般的なテレビによって実現してもよい。

【００１７】

カメラ１４ａはＣＣＤ（Charge Coupled Device）またはＣＭＯＳ（Complementary Metal Oxide Semiconductor）等の撮像素子を備えたデジタルビデオカメラであり、ユーザ１２ａの映像を撮影し、動画像データを生成する。マイクロフォン１６ａはユーザ１２ａが発する音声を取得して音声データを生成する。ディスプレイ１８ａは例えば液晶ディスプレイ、プラズマディスプレイ、有機ＥＬディスプレイ等であり、会話相手を表すキャラクタを表示する。同図の例では、ユーザ１２ａが見るディスプレイ１８ａには相手のユーザ１２ｂのキャラクタ９０が、ユーザ１２ｂが見るディスプレイ１８ｂにはユーザ１２ａのキャラクタ９２が表示されている。スピーカ２２ａは会話相手の音声を音響として出力する。

【００１８】

図２は情報処理装置１０ａの内部回路構成を示している。情報処理装置１０ａは、ＣＰＵ（Central Processing Unit）３２、ＧＰＵ（Graphics Porcessing Unit)３４、メインメモリ３６を含む。ＣＰＵ３２は、オペレーティングシステムやアプリケーションなどのプログラムに基づいて、信号処理や内部構成要素を制御する。ＧＰＵ３４はＣＰＵ３２からの指示に従い画像処理を行い、ディスプレイ１８ａなどに出力する。

【００１９】

これらの各部は、バス４０を介して相互に接続されている。バス４０にはさらに入出力インターフェース３８が接続される。入出力インターフェース３８には、ＵＳＢやＩＥＥＥ１３９４などの周辺機器インタフェースや、有線又は無線ＬＡＮ、および３ＧやＬＴＥ Long Term Evolution）などのモバイル通信とのネットワークインタフェースからなる通信部４２、ハードディスクドライブや不揮発性メモリなどの記憶部４４、音声データを再生してスピーカ２２ａへ出力する音声処理部４６、カメラ１４ａ、マイクロフォン１６ａ、入力装置２０ａからそれぞれの入力データを取得する入力部４８、磁気ディスク、光ディスクまたは半導体メモリなどのリムーバブル記録媒体を駆動する記録媒体駆動部５０が接続される。

【００２０】

ＣＰＵ３２は、記憶部４４に記憶されているオペレーティングシステムを実行することにより情報処理装置１０ａの全体を制御する。ＣＰＵ３２はまた、リムーバブル記録媒体から読み出されてメインメモリ３６にロードされた、あるいは通信部４２を介してダウンロードされた各種プログラムを実行する。

【００２１】

ＧＰＵ３４は、ジオメトリトランスファエンジンの機能とレンダリングプロセッサの機能とを有し、ＣＰＵ３２からの描画命令に従って描画処理を行い、表示画像を、図示しないフレームバッファに格納する。そしてフレームバッファに格納された表示画像をビデオ信号に変換してディスプレイ１８ａに出力する。あるいは通信部４２を介して相手の情報処理装置１０ｂへ送信する。音声処理部４６は、通信部４２を介して相手の情報処理装置１０ｂから取得した音声データをＣＰＵ３２からの命令に従って再生しスピーカ２２ａにオーディオ出力する。

【００２２】
ネットワーク８を介した相手の情報処理装置との通信の確立手順、相手のユーザの音声データ取得、およびオーディオ出力手順については、電話、電話会議、テレビ電話、ボイスチャット、ビデオチャットなどの技術で一般的に用いられている手法のいずれかを適用すればよいため、詳細な説明は省略する。以後、ユーザを表すキャラクタを動作させる手法に主眼を置き説明する。
【００２３】
図３は情報処理装置１０ａの構成を詳細に示している。同図において、さまざまな処理を行う機能ブロックとして記載される各要素は、ハードウェア的には、ＣＰＵ、メモリ、その他のＬＳＩで構成することができ、ソフトウェア的には、メモリにロードされたプログラムなどによって実現される。したがって、これらの機能ブロックがハードウェアのみ、ソフトウェアのみ、またはそれらの組合せによっていろいろな形で実現できることは当業者には理解されるところであり、いずれかに限定されるものではない。
【００２４】
情報処理装置１０ａは、当該装置を操作するユーザ１２ａの映像および音声を取得し、それに基づき相手のユーザが操作する情報処理装置１０へ送信するデータを作成する送信データ作成部６０、相手のユーザが操作する情報処理装置１０から送信された情報を処理して相手のユーザを表すキャラクタがしゃべる様を表現する受信データ処理部８０、ネットワーク８を介して必要なデータを送受信する通信部８６を含む。通信部８６はハードウェアとしては図２の通信部４２で実現できる。
【００２５】
情報処理装置１０ａはさらに、基本画像ＩＤテーブルを格納した基本画像情報記憶部７４、表情判定基準を格納した表情判定基準記憶部７５、画像データを格納した画像データ記憶部７６、およびコマンドＩＤテーブルを記憶したコマンド情報記憶部７８を含む。以後の説明では、図１に示すように情報処理装置１０ａが情報処理装置１０ｂと通信を確立し、それぞれを操作しているユーザ１２ａとユーザ１２ｂが会話するものとする。
【００２６】
このとき、送信データ作成部６０はユーザ１２ａの映像および音声のデータを取得し、それに基づき、ユーザ１２ａを表すキャラクタの動画パラメータを生成して、音声データと同期させて逐次、通信部８６から相手の情報処理装置１０ｂへ送信する。ここで動画パラメータとはキャラクタの動きなどを制御するパラメータであり、アニメーションの各フレームに対応するデータ群である。最初にキャラクタのポリゴンセットなど基本のデータを送信しておくことにより、当該動画パラメータを用いて各フレームの画像をその場で作成することができる。具体的な制御パラメータの種類はアニメーションの表示手法によって様々である。なお動画パラメータとして、動画のフレームデータそのものを送信してもよい。
【００２７】
さらに、ユーザ１２ａの表情が特定のパターンとなったときに、それに対応するコマンドの識別情報を、通信部８６から相手の情報処理装置１０ｂへ送信する。受信データ処理部８０は、相手の情報処理装置１０ｂから受信した動画パラメータと音声データを用いて動画生成および表示、音声再生をするとともに、受信したコマンドの識別情報に基づきコマンドを実行する。
【００２８】
なお、自分の情報処理装置１０ａに、相手のキャラクタとともに自分のキャラクタを表示する態様においては、受信データ処理部８０は、自装置内の送信データ作成部６０が生成した動画パラメータを用いた動画表示を行う。さらにコマンドの内容によっては、自装置内でコマンドを実行する。
【００２９】
送信データ作成部６０は主にＣＰＵ３２、ＧＰＵ３４、入力部４８によって実現され、ユーザ１２ａを表すキャラクタの基本的な画像を特定する基本画像特定部６２、カメラ１

４ａが撮影しているユーザ１２ａの映像を画像フレームごとに解析して顔認識を行う顔認識部６４、顔の所定の部位の形状に基づきユーザ１２ａの表情を数値化する表情パラメータ生成部６６、各時刻に対しキャラクタの出力モデルを決定するモデル制御部６８、出力モデルを含む画像を生成するための動画パラメータを生成する動画パラメータ生成部７０、および、ユーザ１２ａの表情のパターンに対応するコマンドを特定するコマンド特定部７２を含む。

【００３０】

受信データ処理部８０は主にＣＰＵ３２、ＧＰＵ３４、音声処理部４６によって実現され、相手の情報処理装置１０ｂから受信した動画パラメータおよび音声データを用いて動画作成および音声再生をしてディスプレイ１８ａ、スピーカ２２ａにそれぞれ出力する再生部８２、およびコマンドの識別情報に基づきコマンドを実行するコマンド実行部８４を含む。

【００３１】

まず送信データ作成部６０の基本画像特定部６２は、基本画像情報記憶部７４に記憶された基本画像ＩＤテーブルを参照してユーザ１２ａの基本画像を特定する。ここで「基本画像」は、ユーザ１２ａを表すキャラクタのモデルであり、ユーザごとに設定可能とする。たとえば図示しないキャラクタ作成ツールを別に用意しておき、ユーザは当該ツールに表示された複数種類の頭部形状、顔の各部位、髪型、髪色、衣服、装身具などからそれぞれ好みのものを選択して組み合わせることにより、自分のキャラクタの３次元モデルを作成する。背景画像や枠なども選択できるようにしてもよい。このようなツールは、一般的なアプリケーションで実用化されているものを適用できる。

【００３２】

そのようにユーザ１２ａが自分のキャラクタのモデルなどを決定したら、当該モデルを含む基本画像の識別情報とユーザ１２ａの識別情報とを対応づけて基本画像ＩＤテーブルに記録する。基本画像特定部６２は、ユーザ１２ａが情報処理装置１０ａを起動させたときなどに当該ユーザの識別情報、例えばログインＩＤなどを取得し、それに基づき基本画像ＩＤテーブルを参照することにより、基本画像の識別情報、例えばモデルの識別番号、またはジオメトリデータの格納先アドレスなどを特定する。

【００３３】

一方、ユーザ１２ａの識別情報以外の情報、例えばユーザ１２ａが会話開始時点で置かれていた状況も用いて基本画像を決定するようにしてもよい。例えば情報処理装置１０ａがゲーム装置を兼ねており、ゲーム実行中、またはゲームから切り替えて、他のユーザ１２ｂと会話を開始するような場合、キャラクタの顔のモデルはユーザ１２ａの識別情報によって決定する一方で、当該キャラクタが身につける衣服などは、直前に行っていたゲームのキャラクタの衣装とすることで、会話を盛り上げたり娯楽性を高めたりすることができる。同様に、ユーザ１２ａの居場所、時刻、季節などによって衣装や背景画像を決定してもよい。

【００３４】

いずれの場合も、基本画像とするモデルや背景画像のデータを準備できるという条件のもと、基本画像ＩＤテーブルにおいて、当該データと、そのいずれかを選択するための属性の識別情報とを対応づけておけばよい。基本画像特定部６２は、会話を開始する際に当該属性の識別情報を取得することにより基本画像を特定する。具体例は後に詳述する。

【００３５】

顔認識部６４は、カメラ１４ａから取得したユーザ１２ａの動画像データを逐次解析し、入力画像フレームごとにユーザ１２ａの顔認識を行う。具体的には平均的な顔画像とのテンプレートマッチングなどによって、画像中の顔領域、目、鼻、口など顔の各部位の位置および形状を特定し、それぞれにおける特徴点の座標を抽出する。顔認識部６４が行うこれらの処理は、顔認識技術として実用化されている様々な画像解析技術のいずれを適用してもよい。

【００３６】

表情パラメータ生成部６６は、顔認識部６４が抽出した各部位の特徴点の座標から得られるパラメータを、あらかじめ設定した基準値と比較することにより、表情の類型ごとに、表情がなされている度合いを数値化する（以後、これを「表情パラメータ」と呼ぶ）。ここで用いる基準値は、表情判定基準として表情判定基準記憶部７５に格納しておく。

【００３７】

各表情の度合いを数値化するためには、表情判定基準として、当該表情を全くしていないときの値と、最大限しているときの値を設定しておく。例えば口を開けている度合いを数値化する場合、上下の唇の間隔に着目し、平均的な顔において口を閉じているときの間隔を下限値、口を最大に開けているときの間隔を上限値として設定する。そして実際の間隔が下限値から上限値までの範囲のどこに位置するかを確認し、下限値を０、上限値を１として正規化することで、口を開けている度合いが０以上１以下の数値として得られる。

【００３８】

基準値との比較に際しては、顔領域の大きさなどを元に正規化したり、顔の傾きを補正したりする処理を適宜行う。口の開閉のみならず、目の開閉、笑顔、悲しい顔、怒った顔など、複数の表情に対して下限基準値、上限基準値を設定しておくことにより、各表情に対する表情パラメータが入力画像フレームごとに得られる。

【００３９】

モデル制御部６８は、基本画像特定部６２が特定した基本画像に基づき、画像データ記憶部７６から、対応するモデルや画像のデータを読みだす。さらに顔認識部６４が検出した顔領域の傾きや、表情パラメータ生成部６６が算出した表情パラメータを用いて、入力画像フレームに対応する各時刻におけるキャラクタの出力モデルを生成する。画像データ記憶部７６が記憶するデータは、ユーザが作成するなどしたキャラクタや衣装の３次元モデルのジオメトリデータ、背景画像のデータ、３次元グラフィックスの描画に用いる各種パラメータなどを含む。

【００４０】

ここでキャラクタの３次元モデルとして、基準となる表情のモデルのほかに、異なる表情を有する複数の表情モデルのデータを準備しておく。そしてモデル制御部６８が、表情パラメータ生成部６６から取得した表情パラメータに応じた重みで複数の表情モデルを合成することにより、ユーザの表情に近い表情を有する出力モデルを各時刻に対して生成する。このとき同時に、顔認識部６４が検出した顔領域の傾きを出力モデルの顔の傾きに反映させる。また体の部分は基本画像に含まれる衣装を着ているようにする。このようにして生成した出力モデルのパラメータは、背景画像のデータとともに動画パラメータ生成部７０に供給する。

【００４１】

本実施の形態ではキャラクタがユーザの実際の表情と同様の表情をするのみならず、キャラクタがあたかもしゃべっているように見せることで、より高度に臨場感を演出する。例えば、キャラクタの口の動きをユーザの発話と同期させ、口（唇）の形状をユーザの発音に対応させることができれば、ユーザの音声とキャラクタの画像が合致し、より自然な会話を楽しむことができる。

【００４２】

ただしユーザの実際の口の形状を１フレームごとに詳細に取得してトレースしたり、音声解析によって得られた音に合わせて口の形状を決定したりすると、処理コストが増大してレイテンシが生じやすくなる。その結果、処理リソースや通信帯域に制約があるほどリアルタイムな画像表現が困難となる。そこで本実施の形態では、声量、口の形状、発話期間を総合的に考慮することにより、処理リソースなどに制約がある環境においても、音声と合致してしゃべるキャラクタの画像生成を実現する。

【００４３】

具体的には、人が発話しているときに現れる代表的な口の形状を抽出、選択してその口の形を有する発声モデルを表情モデルに含め、実際に発声している期間において、時間に対して重みを変化させながら合成する。例えば母音を発音するための典型的な口の形とし

て、口を上下または全体的に広げた形（「あ」と発生したときの口）、横に広げた形（「い」と発生したときの口）、すぼめた形（「う」と発生したときの口）を表情モデルとして準備し、各重みを時間変化させながら合成する。

【００４４】

選択する口の形状はこれに限らず、言語によっても異なってよい。例えば上下に広げた口、前に突き出した口などでもよい。いずれにしろ代表的な音を発声したときの口の形を優先的に選択する。このようにすることで、選択した複数個の口形状の組み合わせが、それぞれの度合いを変えながら時間変化するため、少ないモデル数でも、単純に口がパクパクと開閉しているのでない微妙な動きを容易な演算で演出できる。またこのときの重みに、実際の声量、実際のユーザの口の開き具合を加味することにより、上記のように別の表情の３次元モデルと合成しても、全体的な口の開き具合が、それらの実際の値を反映したものとなる。具体例は後に述べる。

【００４５】

動画パラメータ生成部７０は、キャラクタをアニメーション表示するのに必要が動画パラメータを生成する。例えば時間的に連続する出力モデルをモーフィングしながら描画していくことにより、よりスムーズなアニメーションの動画を生成できる。動画パラメータ生成部７０はさらに、生成した動画パラメータを圧縮符号化しながら通信部８６に逐次出力する。当該動画パラメータは通信部８６で音声データと同期させて相手の情報処理装置１０ｂへ送信する。

【００４６】

コマンド特定部７２は、顔認識部６４が抽出した、顔の各部位の特徴点の座標の時間変化、あるいは当該特徴点から認識できる表情の時間変化を監視し、コマンド情報記憶部７８に格納されたコマンドＩＤテーブルに設定されたパターンに該当する表情の変化が発生したか否かを判定する。ここで設定するパターンは、「５秒以上笑顔とみなされる表情が続いた」など、会話の途中で発生した自然な表情でもよいし、「左目をつぶった後に口を開けて右目をつぶった」など、ユーザがコマンドを発生させるために意識的に行うものでもよい。いずれの場合も、頭部や顔の部位のいずれかの変化、またはその組み合わせと発生順、１つのパターンとして認識する制限時間を設定する。

【００４７】

そしていずれかのパターンに該当する変化があったと判定されたら、コマンドＩＤテーブルにおいてそのパターンに対応づけられたコマンドの識別情報を特定する。ここで「コマンド」とは、自分が操作する情報処理装置１０ａと相手の情報処理装置１０ｂの少なくともいずれかにおいて表示される画像に変化を与える命令である。例えば「５秒以上笑顔とみなされる表情が続いた」場合には自分のキャラクタが踊り出すコマンドを発生させることにより、より効果的にそのときの感情を表現できる。

【００４８】

「左目をつぶった後に口を開けて右目をつぶった」場合には、意図的に画像に与えたい変化、例えば画面に光のシャワーが流れるように画像を加工するコマンドを発生させる。コマンド特定部７２はコマンドＩＤテーブルから該当コマンドの識別情報を特定すると、当該識別情報を通信部８６に出力する。

【００４９】

受信データ処理部８０の再生部８２は、相手の情報処理装置１０ｂから送信された動画パラメータおよび音声のデータを通信部８６から取得し、フレーム描画、再生処理を行いディスプレイ１８ａおよびスピーカ２２ａへ出力する。コマンド実行部８４は、相手の情報処理装置１０ｂから送信されたコマンドの識別情報を通信部８６から取得しコマンドを実行する。具体的には、コマンド情報記憶部７８に格納されたコマンドＩＤテーブルを参照して、コマンドの識別情報にさらに対応づけられたコマンド実行ファイルをメインメモリ３６などから読み出して記述されたスクリプトを実行したりする。

【００５０】

このように、相手の情報処理装置１０ｂへ動画パラメータやコマンドの識別情報のみを

送信し、表示画像に係る処理の一部を送信先の情報処理装置１０ｂで行わせることにより、送信元の情報処理装置１０ａにおける処理の負荷が軽減する。さらに総合的にみて、会話に際して装置間で送受信すべきデータを減らすことができ、データ伝送に必要な帯域を節約することができる。あるいは動画データを送信元で生成してしまい送信先ではそれを再生するのみとしてもよい。これらの態様のいずれかを、使用可能な通信帯域や処理能力によって適応的に選択するようにしてもよい。

【００５１】

図４はディスプレイ１８ａに表示される画像の別の例を示している。この表示画像例９４では、相手のキャラクタ９０を表示する相手キャラクタウィンドウ９６と、自分のキャラクタ９２を表示する自キャラクタウィンドウ９８とを含む。相手キャラクタウィンドウ９６には、相手の情報処理装置１０ｂから送信された、相手のキャラクタ９０の動画パラメータによって作成した動画像を出力し、自キャラクタウィンドウ９８には、自分のキャラクタ９２の動画を出力する。この場合、自分のキャラクタ９２の動画パラメータは、受信データ処理部８０が、自装置内の送信データ作成部６０から直接受け取ることにより生成される。

【００５２】

またコマンド特定部７２は、コマンドＩＤテーブルに設定されたパターンに該当する表情の変化が発生したと判定したら、対応するコマンドの識別情報を相手の情報処理装置１０ｂに送信するとともに、自装置内のコマンド実行部８４にも当該情報を通知する。そしてコマンド実行部８４は、自キャラクタウィンドウ９８内の画像を対象にコマンドを実行する。

【００５３】

ただし、自分のキャラクタ９２が踊り出すなどキャラクタのアニメーション表示を伴うコマンドを実行する場合は、相手の情報処理装置１０ｂに当該コマンドの識別情報を送信する代わりに、動画パラメータを送信したり、自装置内で生成したアニメーションの動画データを送信したりしてもよい。当該アニメーションを定型的なものとする場合は、あらかじめ、動画パラメータや、操作するユーザのキャラクタのアニメーションを作成して記憶部などに記憶させておいてもよい。

【００５４】

コマンド識別情報の代わりに動画データを送信する態様では、相手の情報処理装置１０ｂは、送信された動画データを再生すればよいため、処理の負荷が軽減される。コマンドの識別情報、動画パラメータ、動画データのいずれを送信するかは、コマンドの内容、情報処理装置の処理能力、通信帯域などによって適応的に決定してよい。

【００５５】

別の例として、相手キャラクタウィンドウ９６および自キャラクタウィンドウ９８の双方を対象にコマンドを実行してもよい。例えば前述のとおり「５秒以上笑顔とみなされる表情が続いた」場合、自分のキャラクタ９２とともに相手のキャラクタ９０も踊り出すようにしてもよい。上記のようにコマンド識別情報に代えて動画データを送信する態様では、いったん自分のキャラクタのアニメーションの動画データを相手の情報処理装置１０ｂに送信し、情報処理装置１０ｂから、相手のキャラクタの動画データを返信してもらうことにより、双方のキャラクタが踊り出すように表示させることができる。

【００５６】

または自分と相手が所定の時間範囲で同じパターンの表情変化を行ったときのみ、コマンドを実行するようにしてもよい。例えば双方のユーザにおいて「５秒以上笑顔とみなされる表情が続いた」ら初めて、自分のキャラクタ９２と相手のキャラクタ９０が踊り出すようにしてもよい。このような場合、コマンド特定部７２がパターンの発生を認識したら、コマンド実行部８４にその旨を通知するとともに、相手の情報処理装置１０ｂにコマンドの識別情報、動画パラメータ、動画データのいずれかを送る。コマンド実行部８４は、相手の情報処理装置１０ｂから同一コマンドの識別情報または対応する動画パラメータ、動画データが送られてきたときに限り、動画の生成や再生などコマンド実行のための処理

を、双方の画像を対象に行う。

【００５７】

図５は、基本画像情報記憶部７４に格納される基本画像ＩＤテーブルのデータ構造例を示している。基本画像ＩＤテーブル１００ａは、ログインＩＤ欄１０２、キャラクタ欄１０４、および背景欄１０６を含む。ログインＩＤ欄１０２には、事前にユーザごとに付与した識別番号を記録する。キャラクタ欄１０４、背景欄１０６にはそれぞれ、上記の通りユーザが事前に作成したキャラクタのモデルを識別する情報、および、選択した背景画像を識別する情報を、ユーザごとに記録する。例えばログインＩＤ「０００３」のユーザについては、「ネコ１」なるキャラクタを、「背景２」なる画像を背景に表示したものが基本画像となる。

【００５８】

図６は基本画像ＩＤテーブルのデータ構造例の別の例を示している。基本画像ＩＤテーブル１００ｂは、上記のとおり、ユーザが直前に行っていたゲームのキャラクタの衣装を自分のキャラクタが身につける態様において、図５の基本画像ＩＤテーブル１００ａと組み合わせて参照される。基本画像ＩＤテーブル１００ｂは、ゲームを識別する情報を記録するゲーム欄１０８、および、各ゲームにおける主なキャラクタ、またはユーザが選択したキャラクタの衣装のモデルを識別する情報を記録する衣装欄１１０を含む。

【００５９】

例えば上記のユーザが直前に「ゲームＡ」なるゲームを行っていた場合、「ネコ１」なるモデルのキャラクタは、「戦闘服２」なるモデルの衣装を身につけ、これが基本画像となる。直前に行っていたゲームを識別する情報は、情報処理装置１０ａの処理履歴などに記録しておく。なお図５の基本画像ＩＤテーブル１００ａのキャラクタ欄１０４において識別されるキャラクタのモデルはデフォルトの服を着ていてよく、直前にゲームを行っていたなど特定の条件において、図６の基本画像ＩＤテーブル１００ｂに記録された情報を優先させるようにしてよい。

【００６０】

図６の基本画像ＩＤテーブル１００ｂのゲーム欄１０８は上記のとおり、ユーザの居場所、時刻、天気などユーザが置かれている状況を表す属性のいずれに置き換えてもよく、衣装欄１１０も背景画像の識別情報を記録する欄に置き換えてよい。例えばユーザの居場所を元に背景画像を決定する場合は、「学校」と対応づけて、あらかじめ準備した学校のグラフィックスを背景画像としたり、「自分の部屋」と対応づけて、ユーザが作成した仮想の部屋のグラフィックスを背景画像としたりしてもよい。また、「現在の居場所」と対応づけて、その場でユーザが撮影した風景写真などを背景画像としてもよい。

【００６１】

このような場合、ユーザが自分の居場所を、「現在の居場所」、「学校」、「自分の部屋」などから選択することにより、基本画像特定部６２がそれに対応付けられた背景画像を決定する。居場所によって衣装を変化させてもよい。なお情報処理装置１０ａの背面に、カメラ１４ａとは別のカメラ（図示せず）を設けることにより、会話中であってもユーザ１２ａが見ている風景を随時撮影することができる。そのようにして撮影した画像を背景画像とすることで、実際にユーザが居る場所に、ユーザの分身であるキャラクタが居るという、臨場感のある画像を会話相手に見せることができる。

【００６２】

カメラで撮影しつづけた風景の動画をリアルタイムで背景としてもよい。このようにすると、例えばユーザが散策しているのと同様にキャラクタが散策している様子を相手に見せることができ、会話をより魅力的に演出することができる。時刻や季節などについても、会話を開始する際の時刻や日付を、情報処理装置１０ａの内部に装備した時計から取得することにより、基本画像の決定に用いることができる。

【００６３】

図７は表情判定基準記憶部７５に格納される表情判定基準のデータ構造例を示している。表情判定基準１２０は表情類型１２２、着目量１２４、下限基準値１２６、上限基準値

１２８を対応づけた情報である。表情類型１２２は、「口開け」、「目閉じ」、「笑顔」など、表情パラメータとしてその度合いを数値化すべき表情の類型である。上述のとおり表情パラメータはキャラクタの表情を決定づける、複数の３次元モデルの合成時の重みに影響を与えるため、キャラクタの表情に含ませたい類型をあらかじめ選択する。また、各類型に対応する３次元モデルを準備する。

【００６４】

着目量１２４は、各表情の度合いを判断する際に着目するパラメータであり、特徴点の座標、分布、複数の特徴点の間隔、初期位置からの変化量など、顔認識部６４が抽出した各特徴点の座標から導出できる量である。下限基準値１２６および上限基準値１２８はそれぞれ、各表情が全くなされていないと判断するとき、最大限なされていると判断するときの、着目量の具体値である。例えば「口開け」の表情の度合いを判断するときは、「上下唇の間隔」が「Δｙ１」であれば口は閉じている、「Δｙ２」であれば口が最大限開いている、と判断する。

【００６５】

「目閉じ」の場合は、「上下まぶたの間隔」が「Δｙ３」であれば目は開いている、「Δｙ４」であれば目が閉じている、と判断する。目については右目の閉じた度合い、左目の閉じた度合いを個別に判定する。また、笑うと目頭が上がり目尻が下がる一般的な傾向を利用し、「笑顔」の表情の度合いを判断するときは、目頭と目尻の高さの差の、初期値からの変化幅を着目量とする。このとき当該パラメータが「０」であれば全く笑っていない、「Δｙ５」であれば最大限笑っている、と判断する。

【００６６】

なお同図では簡易な表記に止めているが、着目量１２４、下限基準値１２６、上限基準値１２８は実際にはさらに詳細に設定されていてよい。例えば「上下唇の間隔」は実際には、上唇および下唇の中心における特徴点の縦方向の位置を表すｙ座標の間隔、などであり、「Δｙ１」、「Δｙ２」は平均的な顔などに基づき具体的な値をあらかじめ設定する。１つの表情につき複数の条件を設定してもよい。

【００６７】

さらに同図の判断基準はあくまで例示であり、パターンマッチング、画像の周波数解析、顔面をメッシュ化したときの形状変化など、表情認識に利用できる様々な手法を用いて条件を設定してもよい。いずれの場合でも、表情判定基準としては、各表情類型が全くなされていない条件と最大限なされているときの条件とを設定しておく。そして表情パラメータ生成部６６は、下限基準値を０、上限基準値を１として実際の着目量を正規化することにより、各表情がなされている度合いを数値化する。例えば上下唇の間隔がΔｙ１とΔｙ２の中間であれば、「口開け」の表情パラメータは「０．５」である。

【００６８】

図８は、モデル制御部６８がキャラクタのモデルを合成する様子を模式的に示している。ここではキャラクタとして牛のモデルが設定されている。画像データ記憶部７６には、当該牛のキャラクタの３次元モデルとして、基本モデル１３０、右目を閉じたモデル１３２、左目を閉じたモデル１３４、「あ」と発声した状態のモデル１３６、「い」と発声した状態のモデル１３８、「う」と発声した状態のモデル（図示せず）、笑顔のモデル（図示せず）など、複数の表情モデルのデータを画像データ記憶部７６に格納しておく。

【００６９】

そして、表情パラメータ生成部６６が導出した表情パラメータに基づき各表情モデルに対して決定した重みｗ０、ｗ１、ｗ２、ｗ３、・・・で重み付けしたうえですべての表情モデルを合成することにより、最終的な出力モデル１４０を生成する。例えば各モデルのジオメトリデータを用い、ＧＰＵ３４に含まれるバーテクスシェーダによって、次のように頂点ブレンディングを行うことにより合成する。

【００７０】

【数１】

$$o' = o + (p0 - o) \times w0 + (p1 - o) \times w1 + (p2 - o) \times w2 + (p3 - o) \times w3 + (p4 - o) \times w4 + (p5 - o) \times w5$$

【００７１】

ここでo'は合成後のモデルの頂点座標、oは基本モデル１３０の頂点座標、ｐ０、ｐ１、ｐ２、ｐ３、ｐ４、ｐ５はそれぞれ、右目を閉じたモデル１３２の頂点座標、左目を閉じたモデル１３４の頂点座標、「あ」と発声した状態のモデル１３６の頂点座標、「い」と発声した状態のモデル１３８の頂点座標、「う」と発声した状態のモデルの頂点座標、笑顔のモデルの頂点座標である。

【００７２】

またｗ０、ｗ１、ｗ２、ｗ３、ｗ４、ｗ５はそれぞれ、右目を閉じたモデル１３２に対する重みパラメータ、左目を閉じたモデル１３４に対する重み、「あ」と発声した状態のモデル１３６に対する重み、「い」と発声した状態のモデル１３８に対する重み、「う」と発声した状態のモデルに対する重み、笑顔のモデルに対する重みであり、０以上１以下の値である。

【００７３】

ここで、右目を閉じたモデル１３２に対する重みパラメータｗ０は、表情パラメータのうち、「目閉じ」に対して得られた右目を閉じている度合いをそのまま用いる。同様に、左目を閉じたモデル１３４に対する重みパラメータｗ１は、「目閉じ」に対して得られた左目を閉じている度合いを、笑顔のモデルに対する重みｗ５は、「笑顔」に対して得られた笑顔の度合いを、そのまま用いる。これにより、例えばユーザが右目を閉じているときは出力モデルも右目を閉じ、ユーザが微笑しているときは出力モデルも微笑した状態を生成できる。

【００７４】

一方、「あ」と発声した状態のモデル１３６に対する重みｗ２、「い」と発声した状態のモデル１３８に対する重みｗ３、「う」と発声した状態のモデルに対する重みｗ４については上記のとおり、発声している期間にその値が時間変化するように決定する。例えば次に示すように、ｗ２、ｗ３、ｗ４を、異なる周期を有する正弦波の絶対値とする。

【００７５】

【数２】

$$w2 = |V\sin(0.7500f)| + m$$
$$w3 = |2V\sin(0.3750f)|$$
$$w4 = |2V\sin(0.1875f)|$$

【００７６】

ただしｗ２が１．０を超えたときはｗ２＝１．０とする。ここでＶはユーザ１２ａの声量を所定の音量に対して正規化した値であり、マイクロフォン１６ａから取得した音声データから算出する。ｆは現在認識対象となっている入力画像フレームの番号であり、時間軸上の画像フレーム列に０から昇順で与えられる値である。すなわちｆが０，１，２，・・・と進むに従い時間が経過する。ｍは、表情パラメータ生成部６６が算出した表情パラメータのうち、「口開け」に対して得られた、口を開いている度合いである。

【００７７】

上記式によれば、声量Ｖが０のとき、ｗ２＝ｍ、ｗ３＝ｗ４＝０、となる。すなわち声量Ｖが０のときはユーザ１２ａは発話していないため、しゃべっているように見せる必要はないが、口が少しでも開いていればそれをキャラクタの表情に反映させるため、その度合いｍを、「あ」と発声した状態のモデル１３６に対する重みｗ２とする。

【００７８】

声量Ｖが０より大きいときは、認識対象の入力画像フレームが先に進むのに伴う時間経過によってｗ２、ｗ３、ｗ４が異なる周期の正弦波の絶対値で変化する。正弦波の振幅は、声量Ｖによって変化する。これにより、声量Ｖが大きいほど全体として大きく開いた口となる。また上式ではｗ２、ｗ３、ｗ４の正弦波の周期を、１倍、２倍、４倍と異ならせることにより、多様かつ微妙な口の動きを表現している。ｗ３およびｗ４には正弦波にさらに「２」なる係数がかかっているが、このような係数および周期の異ならせ方は、合成後のモデルの口の動きを確認しながら、より自然に見えるように調整して決定してよい。

【００７９】

「あ」と発声した状態のモデル１３６に対する重みであるｗ２には、声量Ｖが０でないときも常に、口開けの度合いｍを加算しておくことにより、声が大きいほど、口の開きが大きいほど、「あ」の表情が平均として強く反映されることになり、正弦波の絶対値が０近傍にあるときでも口を開けた状態を維持することができる。このように声量と口の開き具合を同時に考慮しつつ、「あ」、「い」、「う」、など代表的な音の発声時の口の形状を有する複数の顔モデルに対する重みを異なる波形で時間変化させることにより、口唇の複雑な動きを表現できるとともに、大まかな動きが音声に合致しているように見せることができる。

【００８０】

ここで、「え」と発声したときの顔は「い」と発声したときの顔に類似しており、「お」と発声したときの顔は「う」と発声したときの顔と類似しているため、「あ」、「い」、「う」の３つのモデルのみを用いても、それを上記のとおり合成することで、人がしゃべるときの様々な口の形を網羅できる。結果として、少ないデータサイズおよび容易な計算で、現実感のある顔の動きを表現できる。

【００８１】

図９は、コマンド情報記憶部７８に格納されたコマンドＩＤテーブルのデータ構造例を示している。コマンドＩＤテーブル１５０は、パターン欄１５２、コマンドＩＤ欄１５４、コマンド欄１５６を含む。上述のとおりコマンド特定部７２は、顔認識部６４が会話の間中、ユーザ１２ａの顔認識処理を行っていることを利用して、顔の各部位の特徴点の時間変化を監視し、パターン欄１５２に設定されたパターンに該当する変化があったか否かを判定する。

【００８２】

上述のとおりパターン欄１５２に設定するパターンは、ユーザの自然な表情でもよいし意識的に行わないかぎり発生しないようなものでもよい。後者のパターンについては、ユーザがそれと認識して行う必要があるため、設定されているパターンとそれによって生じる変化をあらかじめユーザに提示しておく。あるいはユーザ自身が設定可能としてもよい。

【００８３】

コマンドＩＤ欄１５４およびコマンド欄１５６には、パターン欄１５２に設定したパターンに対応付けて、実行するコマンドのコマンドＩＤおよび、当該コマンドを実行するための実体であるコマンド名、実行ファイル名、または関数名などをそれぞれ記録する。パターン欄１５２に設定したパターンは、実際にはさらに、具体的な特徴点の変化を記述したデータに対応づけられていてよく、コマンド特定部７２は当該記述を参照してもよい。

【００８４】

また特徴点以外の、一般的な表情認識処理に用いられるパラメータを判定に用いてもよい。コマンド欄１５６に記録するデータは、受信データ処理部８０のコマンド実行部８４が、相手の情報処理装置１０ｂから送信されたコマンドＩＤに基づきコマンドを実行できるようにしてあればその形態は限定されない。同図の例では、すでに言及したように「５秒以上笑い続けた」、「左目をつぶった後に口を開けて右目をつぶった」といったパターンのほか、「首を縦振り」、「首を横振り」、「ウィンク５回」、「首をかしげる」、といったパターンが設定されている。

【００８５】

「首を縦振り」した場合、それを「賛成」の意思表示として、当該意思を表現する加工を表示画像に対して行う。たとえばそのようなパターンを発生させたユーザのキャラクタが親指を立てたり拍手をしたりする様を表現する。そのほか背景画像やキャラクタの顔の色を変化させたり、ハートの図形を画面全体に散らせたりするなど、様々な加工が考えられる。いずれの場合も、情報処理装置１０ａは、それを操作するユーザ１２ａが「首を縦振り」したことを検出した場合、コマンドＩＤ欄１５４から「０２２」なるコマンドＩＤを特定し、相手の情報処理装置１０ｂに送信する。あるいは対応する動画パラメータやアニメーション動画のデータを送信する。

【００８６】

当該コマンドＩＤを受信した情報処理装置１０ｂは、それに基づきコマンド欄１５６から「賛成」なるコマンド実行ファイルを特定し、それを実行することにより、送信元のユーザのキャラクタが表示されている画像に加工を施す。同様に、「首を横振り」した場合、それを「拒否」の意思表示として、「賛成」のときとは対照的な表現を行う。このようにすることで、会話中、外出の誘いなど約束をとりつける場面が発生した場合に、「賛成」や「反対」の意思表示を気軽に行える。このとき、画像上、多少大げさな演出でそれを表現することにより、実際の音声や映像のみでは実現しにくい娯楽性を提供することができる。

【００８７】

また「ウィンク５回」、すなわち所定の時間範囲において連続して５回、片目をつぶる動作がなされた場合は、相手を何かに誘いたい意思表示として、あらかじめ用意しておいた「お誘いアニメーション」を再生して表示する。たとえば図４に示したように、自分のキャラクタと相手のキャラクタを同じ画面上に表示するような態様において、誘う側のキャラクタから誘われる側のキャラクタへ招待状が送られるようなアニメーションを表示する。

【００８８】

また「首かしげ」、すなわち首を所定の角度以上傾ける動作がなされた場合は、疑問がある意思表示として、情報処理装置１０ａの使用方法などについて記載されたテキストファイルやホームページを別のウィンドウで表示する。例えばコマンドＩＤテーブル１５０のような表を表示することにより、ユーザがどの動作を行えばよいかわかるようにする。このウィンドウは、そのような動作を行ったユーザが操作する情報処理装置１０ａのディスプレイにのみ表示すればよいため、コマンドＩＤを相手の情報処理装置１０ｂに送信しなくてもよい。

【００８９】

以上述べた本実施の形態によれば、ネットワークを介して複数のユーザが会話する会話システムにおいて、相手のユーザの実像に変えてユーザごとに決定したキャラクタを表示させ、実際の音声に応じてリアルタイムで動作させる。このとき、一つのキャラクタにつき異なる表情の複数の３次元モデルを準備しておき、ユーザの実際の表情を反映させた重みづけで合成することによりキャラクタの表情を実際の表情に近づける。具体的には、顔認識技術を用いて各表情がなされている度合いを各時刻で数値化し、それに基づき重みを変化させる。

【００９０】

このとき、「あ」、「い」、「う」など代表的な音を発声している顔の３次元モデルも準備しておき、声量に応じた振幅で独立に時間変化する重みでそれらの３次元モデルを合成することにより、口唇の微妙かつ複雑な動きを簡易な計算で表現する。人がしゃべるときの様々な口の動きを、「あ」、「い」、「う」、のモデルのみを用いて擬似的に作り出すため、必要なモデルのデータが少なくてすみ、また計算も容易である。

【００９１】

さらに「あ」と発声している顔に対する重みには、実際の口の開き度合いをも考慮することで、実際の音声に合致した自然な口の動きを、細かい動き、大まかな動きの双方でリアルに表現することが容易にできる。複数のモデルを合成することでキャラクタの表情が

多様化するうえ、ユーザの音声、表情に合わせて細かく時間変化させることができるため、少ない処理コストで本当にしゃべっているように見せることができる。

【００９２】

またキャラクタの衣装や背景を、それまでユーザが実行していたゲームや居場所など、ユーザが置かれている状況に応じて決定することにより娯楽性を高めることができる。さらに、ユーザの表情の変化を監視しつづけることにより、有限時間においてなされた表情のパターンからコマンドを発生させる。このようにすることで、コントローラを操作するなど手を用いた操作入力を行わずにすみ操作が容易になる。このように表情を用いた操作によって会話中に画像を加工することにより、意思表示などを効果的に演出でき、会話を盛り上げることができる。

【００９３】

発生させるコマンドについては、その識別情報のみを相手の情報処理装置に送信し、実際のコマンド実行は相手の情報処理装置が担う。またキャラクタをアニメーション表示させるための動画パラメータのみを相手の情報処理装置に送信し、動画の生成は相手の情報処理装置が担う。これにより処理コストを分散させるとともに通信帯域の圧迫を軽減させる。このようにすることで、処理リソースや通信帯域に制約があっても、リアルタイムで複雑な動きをするキャラクタの画像を表示でき、音声とのずれが生じにくい自然な会話システムを実現できる。

【００９４】

以上、本発明を実施の形態をもとに説明した。上記実施の形態は例示であり、それらの各構成要素や各処理プロセスの組合せにいろいろな変形例が可能なこと、またそうした変形例も本発明の範囲にあることは当業者に理解されるところである。

【００９５】

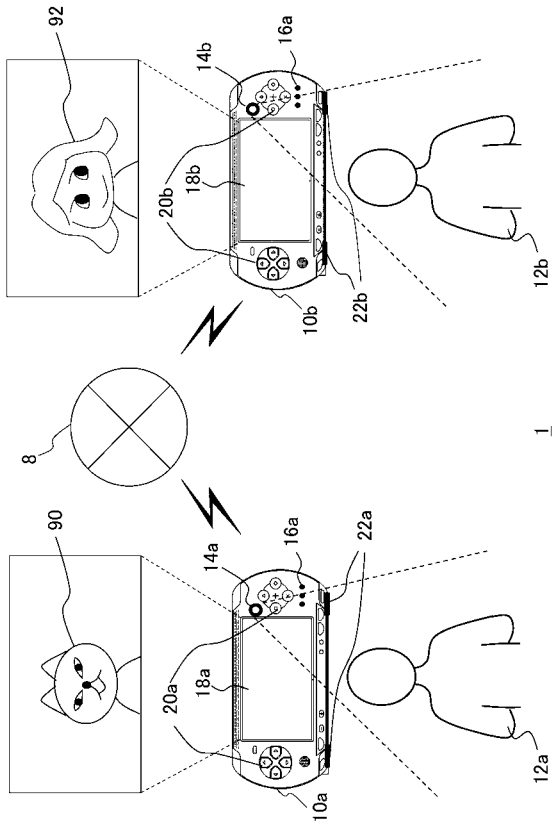
例えば本実施の形態では、直前にユーザが行っていたゲームの衣装をキャラクタに着せる態様を示したが、顔を含めたキャラクタのモデルすべてをゲームのキャラクタに入れ替えるようにしてもよい。この場合も図６に示した基本画像ＩＤテーブル１００ｂの衣装欄１１０をキャラクタ欄に代え、キャラクタのモデルを識別する情報を設定しておけば、処理手順は本実施の形態と同様に実現することができる。

【符号の説明】

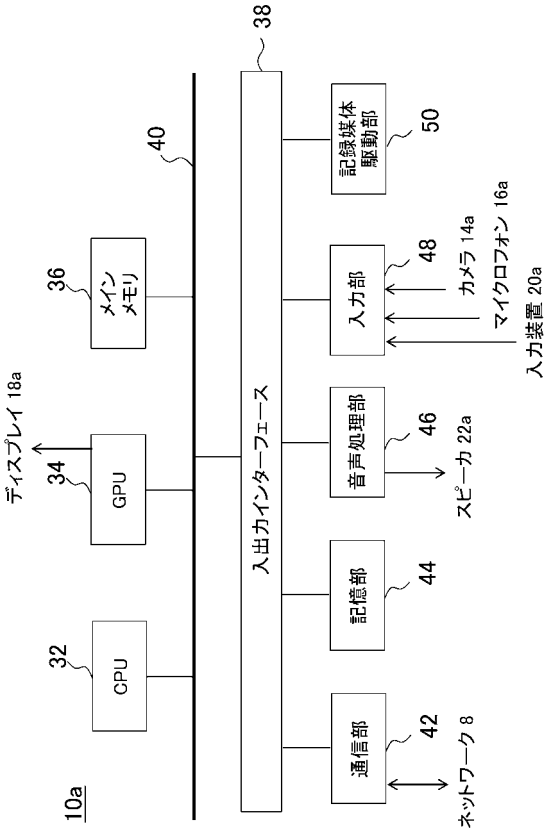
【００９６】

１　会話システム、　８　ネットワーク、　１０ａ　情報処理装置、　１４ａ　カメラ、１６ａ　マイクロフォン、　１８ａ　ディスプレイ、　２０ａ　入力装置、　２２ａ　スピーカ、　３２　ＣＰＵ、　３４　ＧＰＵ、　３６　メインメモリ、　３８　入出力インターフェース、　４０　バス、　４２　通信部、　４４　記憶部、　４６　音声処理部、　４８　入力部、　５０　記録媒体駆動部、　６０　送信データ作成部、　６２　基本画像特定部、　６４　顔認識部、　６６　表情パラメータ生成部、　６８　モデル制御部、　７０　動画パラメータ生成部、　７２　コマンド特定部、　７４　基本画像情報記憶部、　７５　表情判定基準記憶部、　７６　画像データ記憶部、　７８　コマンド情報記憶部、　８０　受信データ処理部、　８２　再生部、　８４　コマンド実行部、　８６　通信部。

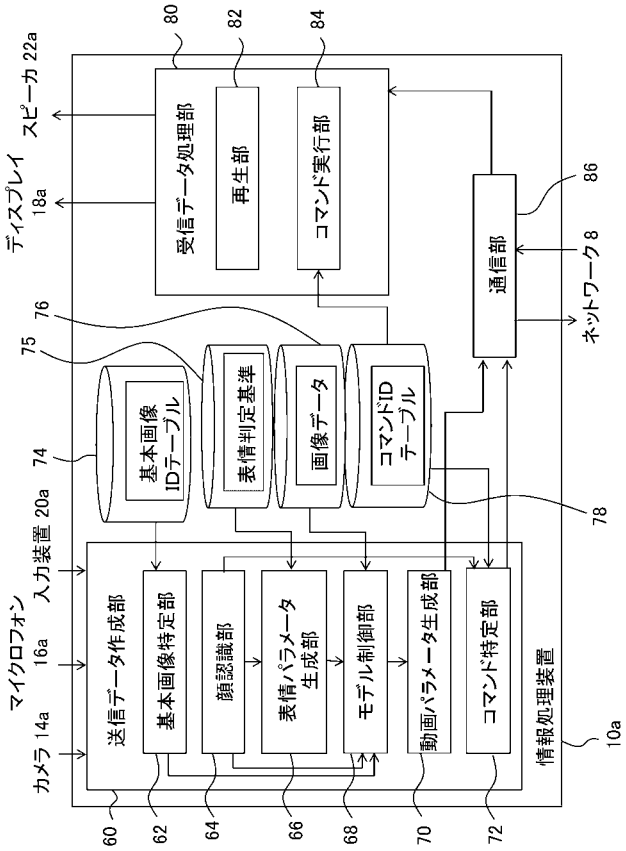
【図１】


92
14b
16a
20b
18b
10b
22b
12b
8
1
90
14a
16a
22a
20a
18a
10a
12a

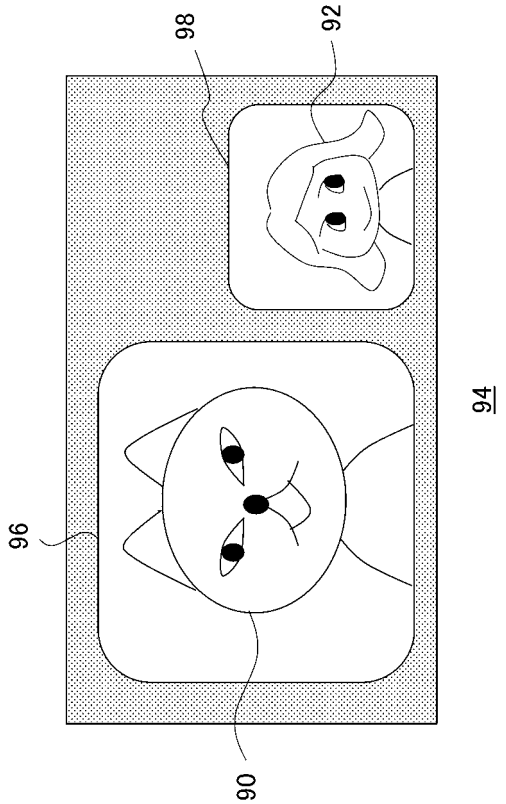
【図２】


38
40
36
メインメモリ
50
記録媒体駆動部
入力部
48
カメラ 14a
マイクロフォン 16a
入力装置 20a
ディスプレイ 18a
34
GPU
入出力インターフェース
音声処理部
46
スピーカ 22a
32
CPU
記憶部
44
10a
通信部
42
ネットワーク 8

【図３】


80
82
84
スピーカ 22a
ディスプレイ 18a
受信データ処理部
再生部
コマンド実行部
86
通信部
ネットワーク 8
76
75
74
78
基本画像IDテーブル
表情判定基準
画像データ
コマンドIDテーブル
マイクロフォン 16a
カメラ 14a
入力装置 20a
送信データ作成部
基本画像特定部
顔認識部
表情パラメータ生成部
モデル制御部
動画パラメータ生成部
コマンド特定部
情報処理装置
10a
60
62
64
66
68
70
72

【図４】


98
92
96
90
94

【図５】

102	104	106
ログインID	キャラクタ	背景
0003	ネコ1	背景2
0012	ウシ4	背景5
・・・	・・・	・・・

100a

【図６】

108	110
ゲーム	衣装
ゲームA	戦闘服2
ゲームB	制服5
・・・	・・・

100b

【図７】

122	124	126	128
表情類型	着目量	下限基準値	上限基準値
口開け	上下唇の間隔	Δy1	Δy2
目閉じ	上下まぶたの間隔	Δy3	Δy4
笑顔	目尻／目頭の変化	0	Δy5
・・・	・・・	・・・	・・・

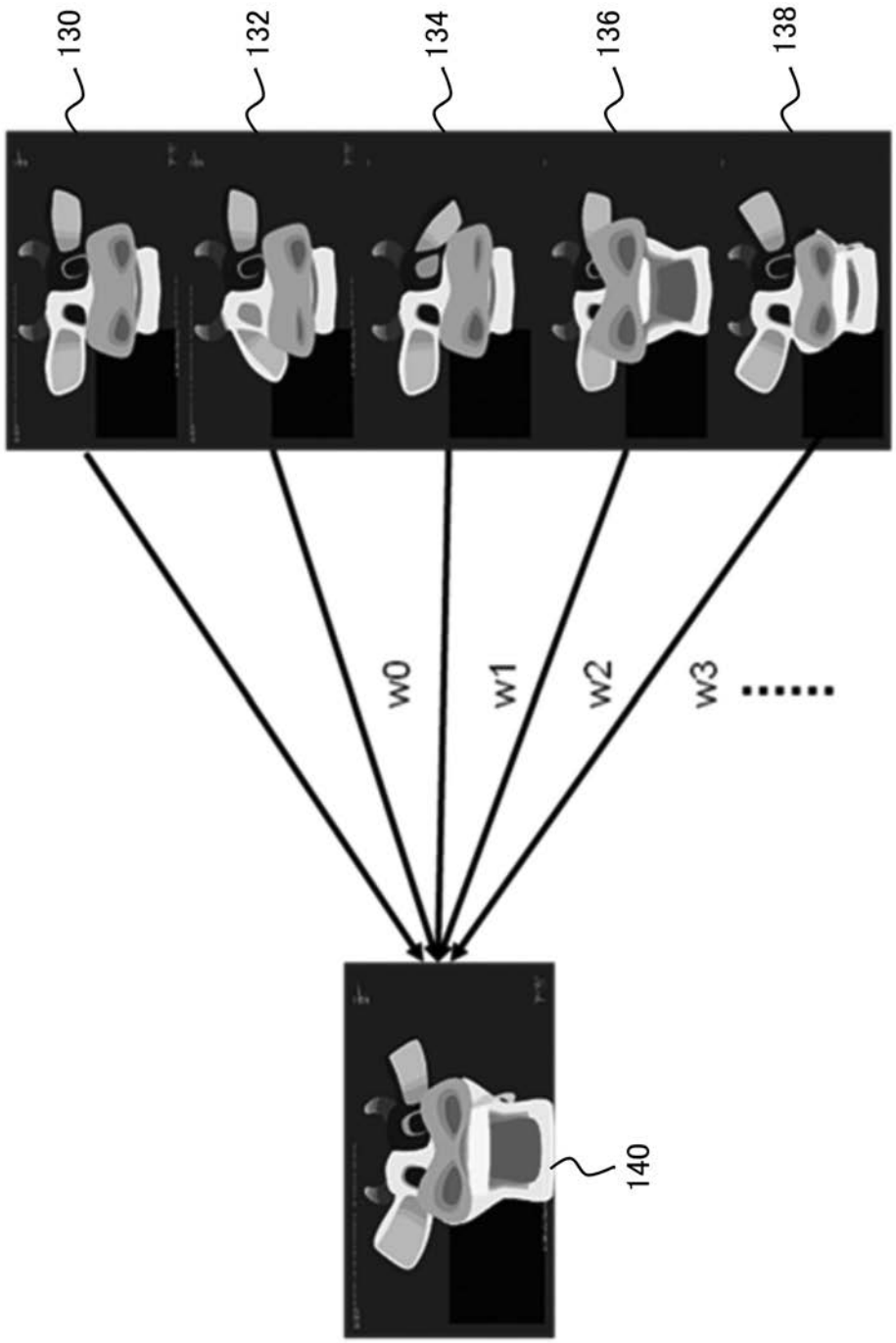
120

【図９】

152	154	156
パターン	コマンドID	コマンド
5秒間笑い続けた	007	踊り出す
左目閉→口開and右目閉	082	光のシャワー
首を縦振り	022	賛成
首を横振り	023	拒否
ウィンク5回	009	お誘いアニメーション
首かしげ	034	ヘルプ呼び出し
・・・	・・・	・・・

150

【図８】



フロントページの続き

(72)発明者　金丸　義勝
東京都港区港南１丁目７番１号　株式会社ソニー・コンピュータエンタテインメント内
(72)発明者　鈴木　章
東京都港区港南１丁目７番１号　株式会社ソニー・コンピュータエンタテインメント内
(72)発明者　清水　正朗
東京都港区港南１丁目７番１号　株式会社ソニー・コンピュータエンタテインメント内
(72)発明者　小口　貴弘
東京都港区港南１丁目７番１号　株式会社ソニー・コンピュータエンタテインメント内
(72)発明者　熊谷　史一
東京都港区港南１丁目７番１号　株式会社ソニー・コンピュータエンタテインメント内

Ｆターム(参考) 2C001 BA03 BA06 BA07 BB10 BC00 BC05 BC09 CA01 CA06 CA07
CB01 CB02 CB03 CB08 CC03 CC08
5B050 BA08 CA07 CA08 DA01 EA05 EA07 EA10 EA12 EA19 EA24
FA02 FA08 FA10 FA12 FA13
5E501 AB02 BA15 BA17 EA00 EA21 FA15 FA21 FA32