明　細　書

発明の名称 ： 情報処理システム、情報処理方法及びプログラム

技術分野

[0001]　本開示は、情報処理システム、情報処理方法及びプログラムに関する。

背景技術

[0002]　人物と対話する技術が知られている。例えば、求職者による所定の質問に対する回答内容を取得し、求職者の回答内容及び評価基準情報に基づいて求職者の面接結果を数値で評価し、面接結果の情報としての採点情報を表示する技術がある。

先行技術文献

特許文献

[0003] 特許文献1：特開２０２３－３１０８９号公報

発明の概要

発明が解決しようとする課題

[0004]　本開示は、利用者と対話する装置を制御する技術を提供する。

課題を解決するための手段

[0005]　本開示の一態様による情報処理システムは、少なくとも１つのメモリと、少なくとも１つのプロセッサと、を備え、少なくとも１つのプロセッサは、利用者の発話に関する情報の取得と、利用者の発話に関する情報と機械学習モデルとに基づく対話装置の発話に関する情報の生成と、対話装置の発話に関する情報に基づく対話装置の発話の制御と、を所定の条件を満たすまで繰り返し実行し、対話装置の発話に関する情報の生成は、利用者と対話装置との関係を示す情報に基づいて実行され、当該関係は、利用者が対話装置に役務若しくは商品を提供する関係、利用者が対話装置を指導する関係、対話装置が利用者に報酬を支払う関係、又は、情報Ａを有する利用者が情報Ａを有しない対話装置に情報Ａを提供する関係の少なくとも１つを含む。

図面の簡単な説明

[0006] [図1]図１は、第１実施形態に係る対話システムの全体構成の一例を示すブロック図である。

[図2]図２は、第１実施形態に係る制御装置の機能構成の一例を示すブロック図である。

[図3]図３は、第１実施形態に係る対話方法の一例を示す処理フローである。

[図4]図４は、第２実施形態に係る対話システムの全体構成の一例を示すブロック図である。

[図5]図５は、第２実施形態に係る管理装置の機能構成の一例を示すブロック図である。

[図6]図６は、第２実施形態に係る対話方法の一例を示す処理フローである。

[図7]図７は、応用例に係る対話システムの一例を示すブロック図である。

[図8]図８は、コンピュータのハードウェア構成の一例を示すブロック図である。

発明を実施するための形態

[0007] 以下、本開示の各実施形態について添付の図面を参照しながら説明する。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した説明を省略する。

[0008] ［第１実施形態］

本開示の第１実施形態は、利用者と対話する情報処理システムの一例である。以下、本実施形態に係る情報処理システムを「対話システム」と呼ぶ。対話システムは、利用者の対話相手となる対話装置を制御することで、利用者との対話を実行してもよい。

[0009] 対話システムは、利用者と対話装置との対話を記録した対話データに基づいて、利用者に関する評価結果を出力する機能を有する。一例として、対話システムは、特定のシナリオに従って対話装置を用いて利用者と対話し、利用者と対話装置との対話を記録した対話データに基づいて、所定の評価基準に従って利用者を評価した評価結果を出力してもよい。一例として、対話の

シナリオは、利用者を小売店の店員とし、対話装置を来店客とし、利用者が来店客の要望に応対する業務を行うシナリオでもよい。この種のシナリオを用いれば、一例として、利用者の接客スキルを評価できる。

[0010] 本実施形態は、利用者と対話する装置を制御する技術を提供する。本実施形態では、利用者が発話した第１発話に関する情報を取得し、第１発話に関する情報と入力情報と機械学習モデルとに基づいて、第２発話に関する情報を生成し、第２発話に関する情報に基づいて、対話装置の発話を制御し、入力情報は、利用者と対話装置との関係を示す情報を含む。

[0011] 一の側面では、本実施形態によれば、機械学習モデルに基づいて対話装置の発話に関する情報を生成するため、利用者と対話する装置を制御できる。他の側面では、本実施形態によれば、利用者と対話装置との対話を通じて、利用者を評価するための情報を取得できる。

[0012] ＜対話システムの全体構成＞

本実施形態に係る対話システムの全体構成について、図１を参照しながら説明する。図１は、第１実施形態に係る対話システムの全体構成の一例を示すブロック図である。

[0013] 図１に示されるように、対話システム１０００は、対話装置１０、制御装置２０、生成装置３０及び端末装置４０を含む。対話装置１０、制御装置２０、生成装置３０及び端末装置４０は、通信ネットワークを介して相互にデータ通信可能に接続されていてもよい。通信ネットワークは、一例として、ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）、ＶＰＮ（Virtual Private Network）又はインターネット等のネットワークでもよい。

[0014] 対話装置１０は、対話システム１０００の利用者の一例である対話者Ｓと対話する装置の一例である。対話装置１０は、対話者Ｓによる発話を受け付けてもよい。対話装置１０は、対話装置１０による発話を対話者Ｓに提示してもよい。対話装置１０は、対話者Ｓによる発話の受け付けと、対話装置１０による発話の提示とを繰り返すことで、対話者Ｓと対話してもよい。以下、対話者Ｓによる発話を「ユーザ発話」とも呼ぶ。また、対話装置１０によ

る発話を「システム発話」とも呼ぶ。

[0015] 対話装置１０は、パーソナルコンピュータ、スマートフォン、タブレット端末等の情報処理端末でもよい。対話装置１０は、一例としてＣＧ（Computer Graphics）キャラクタを表示装置に表示し、ＣＧキャラクタを制御することで、対話者Ｓとの対話を実行してもよい。ＣＧキャラクタは、一例として、アバター等とも呼ばれるバーチャルヒューマンでもよい。ＣＧキャラクタは、人間に限定されず、実在する生物、架空の生物、ロボット、擬人化された物体等でもよい。この場合、対話者Ｓは、表示装置に表示された仮想的なＣＧキャラクタと対話することとなる。

[0016] 対話装置１０は、ロボットでもよい。対話装置１０は、ロボットを動作させることで、対話者Ｓとの対話を実行してもよい。ロボットは、一例として、人間を模したヒューマノイドでもよい。ロボットは、人間に限定されず、実在する生物、架空の生物、擬人化された物体等でもよい。この場合、対話者Ｓは、物理的実体を伴うロボットと対話することとなる。

[0017] 対話装置１０は、ユーザ発話に関する情報を生成し、制御装置２０に送信してもよい。対話装置１０は、システム発話に関する情報を制御装置２０から受信し、システム発話を対話者Ｓに提示してもよい。以下、ユーザ発話に関する情報を「ユーザ発話情報」とも呼ぶ。また、システム発話に関する情報を「システム発話情報」とも呼ぶ。

[0018] ユーザ発話情報は、ユーザ発話を収音した音声信号、ユーザ発話の発話内容を示すテキストデータ、又は対話者Ｓの顔又は身体を撮像した映像信号若しくは画像信号の少なくとも１つを含んでもよい。対話装置１０は、マイクロフォンを用いてユーザ発話を収音することで、音声信号を生成してもよい。対話装置１０は、カメラを用いて対話者Ｓの顔又は身体を撮像することで、映像信号若しくは画像信号を生成してもよい。対話装置１０は、ユーザ発話を収音した音声信号を音声認識することで、発話内容を示すテキストデータを生成してもよい。

[0019] システム発話情報は、システム発話を合成した音声信号、システム発話の

発話内容を示すテキストデータ、ＣＧキャラクタを表す映像信号若しくは画像信号、又は対話装置１０の表現を制御するための制御信号の少なくとも１つを含んでもよい。対話装置１０は、スピーカを用いてシステム発話を示す音声信号を出力することで、システム発話を対話者Ｓに提示してもよい。対話装置１０は、システム発話の発話内容を示すテキストデータを音声合成することで、システム発話を示す音声信号を生成してもよい。対話装置１０は、ＣＧキャラクタを表す映像又は画像を表示装置に表示することで、システム発話情報を対話者Ｓに提示してもよい。対話装置１０がロボットの場合、ロボットの顔又は身体を動作させることで、システム発話情報を対話者Ｓに提示してもよい。

[0020] 制御装置２０は、対話者Ｓと対話装置１０との対話を制御するパーソナルコンピュータ、ワークステーション、サーバ等の情報処理装置の一例である。制御装置２０は、対話装置１０からユーザ発話情報を取得してもよい。制御装置２０は、ユーザ発話情報に基づいて、システム発話情報を生成してもよい。制御装置２０は、システム発話情報に基づいて対話装置１０にシステム発話を発話させる制御を行ってもよい。

[0021] 制御装置２０は、対話データＤを記憶してもよい。対話データＤは、対話者Ｓと対話装置１０との対話を記録した電子データでもよい。対話データＤは、ユーザ発話情報及びシステム発話情報の少なくとも一部を含んでもよい。対話データＤは、対話者Ｓと対話装置１０との対話ごとに記憶されてもよい。対話データＤは、対話を特定可能な識別情報を含んでもよい。対話を特定可能な識別情報の一例は、対話が実行された時刻を示す時刻情報でもよい。時刻情報は、一例として、対話が開始した時刻でもよく、対話が終了した時刻でもよい。

[0022] 制御装置２０は、評価モデルＭ１を備えてもよい。評価モデルＭ１は、対話データＤに基づいて、１以上の評価値を出力するように学習された機械学習モデルの一例である。評価モデルＭ１は、対話データＤの少なくとも一部を入力とし、１以上の評価値を出力してもよい。評価モデルＭ１は、対話デ

ータＤから抽出された１以上の特徴量を入力とし、１以上の評価値を出力してもよい。

[0023] 評価モデルＭ１は、一例として、ニューラルネットワーク、決定木、ランダムフォレスト、勾配ブースティング決定木、サポートベクターマシン（SVM; Support Vector Machine）、大規模言語モデル（LLM; Large Language Model）、又は基盤モデル（Foundation Model）等でもよい。勾配ブースティング決定木は、一例として、ＬｉｇｈｔＧＢＭ（Light Gradient Boosting Machine）又はＸＧＢｏｏｓｔ（eXtreme Gradient Boosting）等でもよい。

[0024] 生成装置３０は、機械学習モデルに基づいて所定のタスクを実行するパーソナルコンピュータ、ワークステーション、サーバ等の情報処理装置の一例である。生成装置３０は、制御装置２０からの生成要求に応じて、システム発話を生成するタスクを実行してもよい。生成装置３０は、システム発話の生成結果を制御装置２０に送信してもよい。

[0025] 生成装置３０は、生成モデルＭ２を備えてもよい。生成モデルＭ２は、所定のタスクを実行するための機械学習モデルの一例である。生成モデルＭ２は、テキスト、音声、画像、動画等の各種データを生成可能な機械学習モデルであってもよい。生成モデルＭ２は、一例として、ニューラルネットワーク、大規模言語モデル、生成モデル又は基盤モデル等でもよい。生成モデルＭ２は、マルチモーダルに対応していてもよい。

[0026] 生成モデルＭ２は、１つの機械学習モデルで実現されてもよい。生成モデルＭ２は、複数の機械学習モデルが連携することで実現されてもよい。生成モデルＭ２は、実行するタスクに応じた複数の機械学習モデルで構成されてもよい。

[0027] 端末装置４０は、対話システム１０００の利用者の一例である評価者Ｅが操作するパーソナルコンピュータ、スマートフォン、タブレット端末等の情報処理端末の一例である。評価者Ｅは、対話者Ｓを評価する主体である。評価者Ｅは、一例として、求職者である対話者Ｓを採用するか否かを判断するために評価する担当者でもよい。評価者Ｅは、一例として、従業員である対

話者Ｓを人事評価する担当者でもよい。評価者Ｅは、一例として、研修の講師であってもよい。評価者Ｅは、一例として、試験の採点者であってもよい。

[0028] 端末装置４０は、制御装置２０から対話者Ｓに関する評価結果を取得してもよい。端末装置４０は、対話者Ｓに関する評価結果を評価者Ｅに提示してもよい。端末装置４０は、一例として、端末装置４０の表示装置に評価結果を表示してもよい。端末装置４０は、一例として、端末装置４０のスピーカから評価結果を合成した音声を出力してもよい。

[0029] 対話システム１０００の利用者に情報を提示することは、プロセッサが表示装置に情報を表示するために必要な処理の少なくとも一部を実行することを含んでよい。表示装置は、プロセッサと同じ装置に備えられてもよく、プロセッサと異なる装置に備えられてもよい。表示装置は、複数の表示装置であってもよい。

[0030] なお、図１に示した対話システム１０００の全体構成は一例であって、用途や目的に応じて様々なシステム構成例があり得る。対話システム１０００は、一以上の装置で構成されてもよい。対話システム１０００に含まれる各装置は、複数の装置で構成されるシステムであってもよい。対話システム１０００に含まれる各機能は、システムを構成する任意の装置で実現されてもよい。対話システム１０００に含まれる各構成要素は、システムを構成する任意の装置に含まれてもよい。

[0031] 対話装置１０、制御装置２０、生成装置３０又は端末装置４０の１つ以上が、対話システム１０００に複数台含まれていてもよい。制御装置２０又は生成装置３０は、複数台のコンピュータにより実現してもよいし、クラウドコンピューティングのサービスとして実現してもよい。制御装置２０及び生成装置３０は、一体化したスタンドアローンのコンピュータにより実現してもよい。図１に示す対話装置１０、制御装置２０、生成装置３０、端末装置４０のような装置の区分は一例である。

[0032] 対話装置１０は、対話者Ｓと対話する装置であればよく、限定されない一

例として、以下のいずれかであってもよい。対話装置１０は、１台以上の情報処理装置でもよい。対話装置１０は、複数の情報処理装置を含む情報処理システムでもよい。対話装置１０は、表示装置を有する端末装置を含んでもよい。対話装置１０は、１以上のハードウェアで制御されるバーチャルヒューマン等のＣＧキャラクタでもよい。対話装置１０は、対話システム１０００に含まれる他の装置により実現してもよい。対話装置１０は、対話システム１０００の外部の情報処理装置又は情報処理システムにより実現してもよい。なお、外部とは、対話システム１０００に含まれないことを意味する。

[0033] 一例として、対話システム１０００は、１以上のサーバ装置と、１以上のクライアント装置とで構成されてもよい。１以上のサーバ装置は、制御装置２０及び生成装置３０の各機能のうち１以上を備えてもよい。１以上のクライアント装置は、対話装置１０及び端末装置４０の各機能のうち１以上を備えてもよい。サーバ装置は、複数の情報処理装置を含むシステムで実現してもよい。サーバ装置は、クラウドコンピューティングのサービスとして実現してもよい。

[0034] 他の一例として、対話システム１０００は、１台の情報処理装置で構成されてもよい。１台の情報処理装置は、対話装置１０、制御装置２０、生成装置３０及び端末装置４０の各機能を備えてもよい。

[0035] ＜制御装置の機能構成＞

制御装置２０の機能構成について、図２を参照しながら説明する。図２は、第１実施形態に係る制御装置の機能構成の一例を示すブロック図である。

[0036] 図２に示されるように、制御装置２０は、シナリオ記憶部１０１、対話記憶部１０２、モデル記憶部１０３、評価結果記憶部１０４、発話取得部１１０、発話生成部１２０、発話提示部１３０、対話評価部１４０及び結果出力部１５０を備える。制御装置２０は、予めインストールされた制御プログラムが実行されることで、シナリオ記憶部１０１、対話記憶部１０２、モデル記憶部１０３、評価結果記憶部１０４、発話取得部１１０、発話生成部１２０、発話提示部１３０、対話評価部１４０及び結果出力部１５０として機能

する。

[0037] （シナリオ記憶部）

シナリオ記憶部１０１には、シナリオ情報が記憶される。シナリオ情報は、対話のシナリオを示す情報である。シナリオ情報は、対話の状況を示す情報を含んでもよい。対話の状況を示す情報は、一例として、対話者Ｓと対話装置１０との関係、対話における前提条件、発話に対する制約条件等を含んでもよい。対話者Ｓと対話装置１０との関係は、対話者Ｓと、対話装置１０が制御するＣＧキャラクタ又は対話装置１０としてのロボットとの関係を含んでもよい。対話者Ｓと対話装置１０との関係は、対話者Ｓの役割及び対話装置１０の役割を含んでもよい。

[0038] 対話者Ｓと対話装置１０との関係は、対話者Ｓが対話装置１０に役務若しくは商品を提供する関係、対話者Ｓが対話装置１０を指導する関係、又は対話装置１０が対話者Ｓに報酬を支払う関係、対話者Ｓが対話装置１０に情報を提供する関係の少なくとも１つ又は少なくとも２つの組み合わせを含んでもよい。具体的には、対話者Ｓと対話装置１０との関係は、小売店の店員と顧客との関係、上司と部下との関係、教師と生徒との関係、講演者と聴衆との関係、医者と患者との関係、又は被依頼者と依頼者との関係の少なくとも１つを含んでもよい。被依頼者は、一例として、弁護士又はコンサルタント等の専門家を含んでもよい。

[0039] 一例として、シナリオ情報は、小売店又は飲食店における店員と来店客との対話に関するシナリオでもよい。この場合、対話者Ｓの役割は、「顧客」、「客」、「お客さん」、「お客様」、「来店客」等のいずれかでもよい。対話装置１０の役割は、「店員」、「スタッフ」、「店の人」等のいずれかでもよい。

[0040] シナリオ情報は、１以上のシステムプロンプトと、１以上のガイダンス情報とを含んでもよい。システムプロンプトは、シナリオに沿ったシステム発話を生成モデルＭ２に生成させるために用いられる情報でもよい。ガイダンス情報は、対話を開始する際に対話者Ｓにシナリオを説明するために用いら

れる情報でもよい。ガイダンス情報は、対応するシステムプロンプトと関連付けられてもよい。ガイダンス情報は、対話者Ｓと対話装置１０との関係を示す情報を含んでもよい。

[0041] システムプロンプトは、対話の状況を示す情報を含んでもよい。システムプロンプトは、対話のシナリオに基づいて予め生成されてもよい。システムプロンプトは、対話者Ｓと対話装置１０との関係を示す情報を含んでもよい。システムプロンプトは、対話における前提条件を含んでもよい。システムプロンプトは、発話に対する制約条件を含んでもよい。

[0042] システムプロンプトが対話者Ｓと対話装置１０との関係を示す情報を含む場合、システムプロンプトは、少なくとも対話装置１０の役割に関する情報を含んでいればよく、対話者Ｓの役割に関する情報を明示的に含んでいなくてもよい。システムプロンプトは、対話者Ｓの役割を間接的に示す情報を含んでいてもよい。対話装置１０の役割は、顧客、部下、生徒、聴衆、患者又は依頼者の少なくとも１つを含んでもよい。

[0043] システムプロンプトは、一例として、下記の文字列でもよい。

・「あなたは服飾又は雑貨などを扱う店舗に来た男性のお客さんです。店員に尋ねられながら服を購入します。買いたい服が決まっているわけではありません。好みは、デザインがシンプル、素材は青色の綿、用途は普段使い、予算は５万円です。情報は小出しにしてください。服を探していることは伝えてかまいません。お客さん役の発話のみを５０文字以内で出力してください。」

[0044] ここで、「お客さん」及び「店員」は、対話装置１０と対話者Ｓとの関係を示す情報の一例である。なお、「お客さん」は対話装置１０の役割であり、「店員」は対話者Ｓの役割である。「お客さん役の発話のみを５０文字以内で出力してください。」は、発話に対する制約条件の一例である。その他の部分は、対話における前提条件の一例である。

[0045] システムプロンプトは、他の一例として、下記の文字列でもよい。下記の文字列は、対話者Ｓの役割を間接的に示す情報を含むシステムプロンプトの

、限定されない一例である。

・「あなたは服飾又は雑貨などを扱う店舗に来た男性のお客さんです。服を購入します。買いたい服が決まっているわけではありません。好みは、デザインがシンプル、素材は青色の綿、用途は普段使い、予算は５万円です。情報は小出しにしてください。服を探していることは伝えてかまいません。お客さん役の発話のみを５０文字以内で出力してください。」

[0046] ここで、「お客さん」は、対話装置１０の役割に関する情報の一例である。また、「お客さん」は、対話者Ｓの役割を間接的に示す情報の一例である。対話装置１０の役割が「お客さん」であることが示されていることで、対話者Ｓの役割は「店員」であることが間接的に示されている。

[0047] シナリオ情報は、複数のシステムプロンプトを含んでもよい。複数のシステムプロンプトは、一例として、対話の進行度合いに応じたシステムプロンプトでもよい。対話の進行度合いは、一例として、対話のターン数又は対話が開始してからの経過時間でもよい。

[0048] ガイダンス情報は、システムプロンプトと関連する内容でもよい。ガイダンス情報は、対話者Ｓと対話装置１０との関係を示す情報を含んでもよい。対話者Ｓと対話装置１０との関係を示す情報は、システムプロンプトと同じ内容でもよい。ガイダンス情報は、システムプロンプトよりも情報量が少なくてもよい。ガイダンス情報は、システムプロンプトから対話装置１０に関する前提条件の少なくとも一部が省略されていてもよい。ガイダンス情報は、対話の開始を指示する情報が含まれてもよい。対話の開始を指示する情報は、ガイダンス情報の末尾に位置してもよい。

[0049] ガイダンス情報は、一例として、下記の文字列でもよい。

・「あなたは衣料品店の店員です。お客さんとして私が立っていますので、店員になったつもりで話しかけてください。それではどうぞ。」

[0050] ここで、「店員」及び「お客さん」は、対話者Ｓと対話装置１０との関係を示す情報の一例である。なお、「お客さん」は対話装置１０の役割であり、「店員」は対話者Ｓの役割である。「それではどうぞ」は、対話の開始を

指示する情報の一例である。

[0051] シナリオ記憶部１０１には、複数のシナリオ情報が記憶されてもよい。複数のシナリオ情報は、一例として、業種又は職種に応じたシナリオに関するシナリオ情報を含んでもよい。複数のシナリオ情報は、シナリオを示す識別情報と関連付けて、シナリオ記憶部１０１に記憶されてもよい。識別情報とシナリオ情報とを関連付ける方法は、一例として、識別情報とシナリオ情報とを１つのセットとして記憶してもよいし、一方の情報から他方の情報を取得可能な情報を記憶してもよいし、一方の情報と他方の情報の識別情報とを１つのセットとして記憶してもよい。

[0052] シナリオ情報は、評価者Ｅが作成してもよい。シナリオ情報は、予め作成されたテンプレートに基づいて作成されてもよい。一例として、シナリオ情報は、テンプレート中の固有名詞（例えば、店舗名、商品名、サービス名等）を変更することで作成されてもよい。シナリオ記憶部１０１には、予め作成された複数のシナリオ情報のうち、評価者Ｅが選択したシナリオ情報が記憶されてもよい。

[0053] （対話記憶部）

対話記憶部１０２には、対話データＤが記憶される。対話データＤは、対話者Ｓと対話装置１０との対話ごとに記憶されてもよい。対話データＤは、対話者Ｓと対話装置１０との対話が開始してから終了するまでの間に、制御装置２０が取得したユーザ発話情報及びシステム発話情報を含んでもよい。対話データＤに含まれるユーザ発話情報は、発話取得部１１０が取得してもよい。対話データＤに含まれるシステム発話情報は、発話生成部１２０が生成してもよい。

[0054] 対話データＤは、一例として、時刻情報、対話者Ｓの属性情報、対話者Ｓの音声情報、対話者Ｓの画像情報、対話装置１０の音声情報、対話装置１０の画像情報、対話装置１０の制御情報、対話全体の録画データ、対話システム１０００のログ情報の少なくとも１つを含んでもよい。

[0055] 対話者Ｓの属性情報は、対話者Ｓのプロフィールを含んでもよい。プロフ

ィールは、一例として、履歴書又は職務経歴書の記載事項を含んでもよい。対話者Ｓの属性情報は、予め設定されてもよく、対話の開始時に対話者Ｓにより入力されてもよい。

[0056] 対話者Ｓの音声情報は、対話者Ｓの発話を収音した音声信号を含んでもよい。対話者Ｓの音声情報は、音声信号を認識したテキストデータを含んでもよい。対話者Ｓの画像情報は、対話者Ｓの顔又は身体の少なくとも一部を撮像した静止画又は動画を含んでもよい。対話者Ｓの画像情報は、対話者Ｓの画像から認識した情報を含んでもよい。一例として、対話者Ｓの画像情報は、表情ラベル又は姿勢情報等を含んでもよい。表情ラベルは、人物の表情を分類したクラスを示すラベルである。姿勢情報は、人物の特徴点の位置を示す情報である。例えば、姿勢推定技術によって各関節位置を得ることで、姿勢を推定してもよい。

[0057] 対話装置１０の音声情報は、システム発話を合成した音声信号を含んでもよい。対話装置１０の音声情報は、音声信号の合成に用いたテキストデータを含んでもよい。対話装置１０の画像情報は、ＣＧキャラクタの顔又は身体の少なくとも一部を含む静止画又は動画を含んでもよい。対話装置１０の画像情報は、ＣＧキャラクタの画像の生成に用いた情報を含んでもよい。一例として、対話装置１０の画像情報は、表情ラベル又は姿勢情報等を含んでもよい。対話装置１０の制御情報は、対話装置１０がロボットである場合に、当該ロボットの表情や身振り手振りなどを含む、身体制御情報を含んでもよい。

[0058] 対話全体の録画データは、対話者Ｓの音声信号及び映像信号と、対話装置１０の音声信号及び映像信号とを合成した動画データでもよい。対話全体の録画データは、対話装置１０により合成されてもよく、制御装置２０により合成されてもよい。

[0059] 対話システム１０００のログ情報は、対話者Ｓと対話装置１０との対話において、対話システム１０００が行った処理の結果を記録した履歴情報でもよい。対話システム１０００のログ情報は、対話装置１０の処理結果、制御

装置２０の処理結果、又は生成装置３０の処理結果の少なくとも１つを含んでもよい。一例として、対話システム１０００のログ情報は、ユーザ発話の認識結果、生成モデルＭ２への入力情報、生成モデルＭ２からの出力情報、システム発話の発話内容等を含んでもよい。

[0060] 対話データＤに含まれる各情報は、時刻情報と関連付けて記憶されてもよい。各情報と時刻情報とを関連付ける方法は、一例として、各情報と時刻情報とを１つのセットとして記憶してもよいし、一方の情報から他方の情報を取得可能な情報を記憶してもよいし、一方の情報と他方の情報の識別情報とを１つのセットとして記憶してもよい。

[0061] （モデル記憶部）

モデル記憶部１０３には、評価モデルＭ１が記憶される。評価モデルＭ１は、予めモデル記憶部１０３に記憶されてもよい。評価モデルＭ１は、予め用意された教師データを学習することで、制御装置２０により生成されてもよい。評価モデルＭ１は、制御装置２０と異なる他の情報処理装置又は情報処理システムにより生成され、モデル記憶部１０３に記憶されてもよい。

[0062] モデル記憶部１０３には、複数の評価モデルＭ１が記憶されてもよい。複数の評価モデルＭ１は、評価値ごとに構築された評価モデルを含んでもよい。評価値は、対話者Ｓに関する表情、身体動作（一例として、頭、手、腕、姿勢等）、声、対話の円滑度、対話内容の少なくとも１つを評価する指標を含んでもよい。

[0063] 評価モデルＭ１は、対話者Ｓの表情を評価する評価モデルを含んでもよい。表情を評価する評価モデルは、対話者Ｓの顔が撮像された画像情報を入力とし、対話者Ｓの表情に関する評価値を出力する機械学習モデル（例えば、ニューラルネットワーク等）を含んでもよい。表情を評価する評価モデルＭ１は、顔の特徴点や顔のランドマーク、ＦＡＵ（Facial Action Units）などの画像情報から抽出された特徴量を入力とし、対話者Ｓの表情に関する評価値を出力する機械学習モデル（例えば、決定木、ランダムフォレスト、サポートベクターマシン、ニューラルネットワーク等）を含んでもよい。

[0064] 評価モデルＭ１は、対話者Ｓの身体動作を評価する評価モデルを含んでもよい。身体動作を評価する評価モデルは、対話者Ｓを撮像した映像信号を入力とし、対話者Ｓの身体動作に関する評価値を出力する機械学習モデル（例えば、基盤モデル、姿勢推定モデル、行動認識モデル、ジェスチャ認識モデルや、それらのモデルで得られる特徴量から評価値を出力するモデル）を含んでもよい。

[0065] 評価モデルＭ１は、対話者Ｓの声を評価する評価モデルを含んでもよい。声を評価する評価モデルは、対話者Ｓの発話を収音した音声信号を入力とし、対話者Ｓの音声の非言語情報に関する評価値を出力する機械学習モデル（例えば、ニューラルネットワーク等）を含んでもよい。非言語情報は、一例として、トーン、抑揚、話速、声の高さ、明瞭さ、発話時間、フィラーの回数等を含んでもよい。

[0066] 評価モデルＭ１は、対話の円滑度を評価する評価モデルを含んでもよい。対話の円滑度を評価する評価モデルは、対話全体の音声信号を入力とし、対話の円滑度に関する評価値を出力する機械学習モデル（例えば、基盤モデル、音声認識モデル、音声区間検出モデル、話す速度の推定モデル、フィラー検出モデルや、それらから得られる特徴量を使って評価値を学習したモデル）を含んでもよい。対話の円滑度に関する評価値は、一例として、対話のやりとりの長さ、対話者Ｓの発話のタイミング、対話者Ｓが話している割合、無言の時間等を含んでもよい。対話の円滑度に関する評価値の少なくとも一部は、ルールベースで解析してもよい。

[0067] 評価モデルＭ１は、対話内容を評価する評価モデルを含んでもよい。対話内容を評価する評価モデルＭ１は、発話内容を示すテキストデータの少なくとも一部を入力とし、対話内容に関する評価値を出力する機械学習モデル（例えば、対話内容を評価するよう指示するシステムプロンプトを与えた大規模言語モデル）を含んでもよい。

[0068] 評価モデルＭ１は、評価者Ｅの評価基準に基づいて構築されてもよい。評価モデルＭ１は、複数の評価者Ｅそれぞれの評価基準ごとに構築されてもよ

い。評価モデルＭ１は、評価者Ｅの評価基準に基づいて評価値の正解値が付与された対話の録画データに基づいて構築されてもよい。

[0069] （評価結果記憶部）

評価結果記憶部１０４には、対話者Ｓに関する評価結果が記憶される。対話者Ｓに関する評価結果は、対話評価部１４０が生成してもよい。評価結果記憶部１０４には、対話者Ｓに関する複数の評価結果が記憶されてもよい。複数の評価結果は、異なる複数の対話に関する対話データＤそれぞれに基づいて生成されてもよい。複数の対話は、時刻、シナリオ、評価基準、評価値等の少なくとも１つが異なってもよい。評価結果記憶部１０４には、複数の評価結果に示される評価値の推移又は統計情報が記憶されてもよい。統計情報は、一例として、最大、最小、平均等でもよい。

[0070] 評価結果記憶部１０４には、外部評価値が記憶されてもよい。外部評価値は、対話データＤ以外の情報に基づいて、対話者Ｓを評価した評価値である。外部評価値は、対話システム１０００の外部の情報処理装置又は情報処理システムで生成され、評価結果記憶部１０４に記憶されてもよい。一例として、外部評価値は、対話者Ｓの適性検査の結果、語学力を示す情報、職務経験を示す情報、特定の分野における専門知識を有することを示す情報等を含んでもよい。

[0071] （発話取得部）

発話取得部１１０は、ユーザ発話情報を取得する。発話取得部１１０は、ユーザ発話を収音した音声信号に基づいて、ユーザ発話情報を生成してもよい。発話取得部１１０は、ユーザ発話を収音した音声信号を、対話装置１０から取得してもよい。

[0072] 発話取得部１１０は、対話者Ｓを撮像した映像信号に基づいて、ユーザ発話情報を生成してもよい。発話取得部１１０は、対話者Ｓを撮像した映像信号を、対話装置１０から取得してもよい。

[0073] 発話取得部１１０は、ユーザ発話の発話内容を示すテキストデータに基づいて、ユーザ発話情報を生成してもよい。発話取得部１１０は、ユーザ発話

を収音した音声信号を音声認識することで、ユーザ発話の発話内容を示すテキストデータを生成してもよい。

[0074] 発話取得部１１０は、対話装置１０から取得した映像信号が、対話者Ｓが適切に撮像されている映像信号であるか否かを判定してもよい。発話取得部１１０は、映像信号に対話者Ｓの顔が撮像されているか否かを判定してもよい。発話取得部１１０は、映像信号に撮像された対話者Ｓの顔のサイズが所定の範囲内であるか否かを判定してもよい。発話取得部１１０は、顔の検出器を用いて、映像信号に撮像されている対話者Ｓの顔を検出してもよい。顔の検出器は、一例として、機械学習モデルを用いて構築されてもよい。

[0075] 対話者Ｓの身体動作に関する評価値を計算する場合、発話取得部１１０は、映像信号に対話者Ｓの身体が撮像されているか否かを判定してもよい。発話取得部１１０は、映像信号に撮像された対話者Ｓの身体のサイズが所定の範囲内であるか否かを判定してもよい。発話取得部１１０は、身体の検出器を用いて、映像信号に撮像されている対話者Ｓの身体を検出してもよい。身体の検出器は、一例として、機械学習モデルを用いて構築されてもよい。

[0076] 発話取得部１１０は、映像信号に対話者Ｓが適切に撮像されていない場合、その旨を対話者Ｓに通知してもよい。一例として、発話取得部１１０は、対話者Ｓが適切に撮像されていない旨を示す通知を対話装置１０に送信してもよい。対話装置１０は、対話者Ｓが適切に撮像されていない旨を示すメッセージを表示装置に表示してもよい。対話装置１０は、対話者Ｓが適切に撮像されていない旨を発話してもよい。発話取得部１１０は、映像信号に対話者Ｓが適切に撮像されていない場合、対話者Ｓとの対話を中止してもよい。

[0077] 発話取得部１１０は、取得したユーザ発話情報を対話記憶部１０２に記憶してもよい。発話取得部１１０は、リアルタイムにユーザ発話情報を記憶してもよく、所定の処理単位で一括してユーザ発話情報を記憶してもよい。所定の処理単位は、一例として、ユーザ発話情報のデータ量、対話のターン数、対話の経過時間等でもよい。

[0078] 発話取得部１１０は、対話記憶部１０２に記憶されている対話データＤに

ユーザ発話情報を含めてもよい。発話取得部１１０は、対話データＤに含まれる時刻情報と関連付けてユーザ発話情報を対話記憶部１０２に記憶してもよい。

[0079] （発話生成部）

発話生成部１２０は、システム発話情報を生成する。発話生成部１２０は、発話取得部１１０が取得したユーザ発話情報に基づいて、システム発話情報を生成してもよい。発話生成部１２０は、生成モデルＭ２に基づいて、システム発話情報を生成してもよい。発話生成部１２０は、ユーザ発話情報と入力情報と生成モデルＭ２とに基づいて、システム発話情報を生成してもよい。

[0080] なお、ある情報と機械学習モデルとに基づいて、他の情報を生成することは、以下のいずれか１つ以上の処理を実行することを含んでもよい。

・ある情報を機械学習モデルに入力して他の情報を生成すること。

・ある情報に基づいて生成した情報を機械学習モデルに入力して他の情報を生成すること。

・ある情報を用いて抽出された情報を用いて機械学習モデルを用いた生成処理を実行し、他の情報を生成すること。

[0081] 入力情報は、シナリオ情報に基づいて生成されてもよい。シナリオ情報は、予め設定されてもよい。シナリオ情報は、所定の基準に基づいて設定されてもよい。シナリオ情報は、評価者Ｅの指示に基づいて設定されてもよい。

[0082] 発話生成部１２０は、ユーザ発話情報に基づいて生成モデルＭ２への入力情報を生成し、生成装置３０に送信してもよい。以下、生成モデルＭ２への入力情報を、「生成モデル入力情報」とも呼ぶ。

[0083] 生成モデル入力情報は、入力情報の一例であるシステムプロンプト及び対話履歴情報を少なくとも含んでもよい。発話生成部１２０は、シナリオ記憶部１０１に記憶されたシナリオ情報からシステムプロンプトを読み出してもよい。発話生成部１２０は、対話の進行度合いに応じて、読み出すシステムプロンプトを選択してもよい。発話生成部１２０は、業種又は職種に応じて

、読み出すシステムプロンプトを選択してもよい。発話生成部１２０は、複数のシステムプロンプトを読み出し、それらを結合してもよい。一例として、発話生成部１２０は、業種に応じたシステムプロンプトと、職種に応じたシステムプロンプトとを読み出し、それらを結合したシステムプロンプトを生成してもよい。

[0084] 発話生成部１２０は、機械学習モデルに基づいて、システムプロンプトを選択してもよい。機械学習モデルは、生成モデルＭ２でもよく、他の機械学習モデルでもよい。他の機械学習モデルは、一例として、大規模言語モデル又は基盤モデルでもよい。発話生成部１２０は、対話者Ｓと対話装置１０との対話に関する情報と、複数のシステムプロンプトとを機械学習モデルに入力し、１以上のシステムプロンプトを選択するための情報を取得してもよい。対話に関する情報は、対話履歴情報でもよい。また、発話生成部１２０は、機械学習モデルに基づいて、システムプロンプトを生成してもよい。

[0085] 対話履歴情報は、対話者Ｓの発話内容及び対話装置１０の発話内容を時系列に連結した情報でもよい。対話履歴情報は、対話が開始してから直前の発話までの発話内容が含まれてもよい。対話履歴情報は、直前の発話から時間方向に遡って所定数の発話内容を含んでもよい。各発話内容は、発話者（対話者Ｓ又は対話装置１０）を示す情報が関連付けられてもよい。発話者を示す情報は、発話内容の前又は後ろに付加されてもよい。発話者を示す情報は、所定の記号により識別可能に付加されてもよい。各発話内容の境界は、所定の記号により示されてもよい。対話履歴情報は、発話を生成する対象の発話者を示す情報を含んでもよい。発話を生成する対象の発話者を示す情報には、発話内容が関連付けられていなくてよい。

[0086] 対話履歴情報は、一例として、下記の文字列でもよい。

・「店員：いらっしゃいませ￥ｎお客さん：」

[0087] ただし、「いらっしゃいませ」は、直前の発話の発話内容の一例である。「店員」及び「お客さん」は、発話者を示す情報の一例である。「：」は発話者とその発話内容を区切る記号の一例である。「￥ｎ」は発話内容の境界

（この場合は直前の発話内容の終端）を示す記号の一例である。上記の例では、「お客さん」には発話内容が関連付けられていないため、お客さん（すなわち、対話装置１０）の発話が生成対象となる。上記の例では、対話履歴情報に直前の発話内容のみが含まれているが、それ以前に１以上のターンが行われていれば、「店員」より前方に以前のターンにおける発話内容が含まれてもよい。

[0088] 対話履歴情報はテキスト情報に限定されるものではなく、音声情報又は画像情報を含んでもよく、テキスト情報と音声情報又は画像情報の少なくとも一方とを含んでもよい。具体的には、対話履歴情報は、対話者Ｓの音声情報、対話者Ｓの画像情報、対話装置１０の音声情報、対話装置１０の画像情報を時系列に連結した情報を含んでもよい。

[0089] 発話生成部１２０は、システムプロンプト及び対話履歴情報を所定のテンプレートに埋め込むことで、生成モデル入力情報を生成してもよい。発話生成部１２０は、テンプレート、システムプロンプト及び対話履歴情報を所定のルールに基づいて加工することで、生成モデル入力情報を生成してもよい。発話生成部１２０は、テンプレート、システムプロンプト及び対話履歴情報を機械学習モデルに入力することで、生成モデル入力情報を生成してもよい。機械学習モデルは、生成モデルＭ２でもよく、他の機械学習モデルでもよい。他の機械学習モデルは、一例として、大規模言語モデル又は基盤モデルでもよい。

[0090] 生成モデル入力情報を生成するためのテンプレートは、システムプロンプト又は対話履歴情報を埋め込む１以上のプレースホルダを含んでもよい。テンプレートは、発話の生成を指示する指示情報を含んでもよい。指示情報は、出力情報の出力形式を指示する情報を含んでもよい。出力情報の出力形式は、一例として、テキストデータ、音声信号、映像信号又は制御信号の少なくとも１つを含んでもよい。テンプレートは、出力情報に関する制約情報を含んでもよい。制約情報は、発話の長さや話しても良い内容、聞かれたら話しても良い内容、聞かれても話してはいけない内容、話し方などを含んでも

よい。指示情報及び制約情報は、予め定めた固定文でもよい。

[0091] テンプレートは、任意のプロンプトチューニング手法により、良好な出力情報が得られるように最適化されていてもよい。テンプレートは、一例として、テンプレートを用いて生成したプロンプトを評価したベンチマークに基づいて、遺伝的アルゴリズム等により最適なテンプレートを探索することで最適化されていてもよい。

[0092] 生成モデル入力情報は、テキストデータ、画像データ又は音響データを含んでもよい。テキストデータは、一例として、プロンプトと呼ばれる自然言語文でもよい。画像データは、一例として、静止画でもよく、動画でもよい。テキストデータは、音響データ又は動画を音声認識したテキストデータでもよい。テキストデータは、画像データを文字認識したテキストでもよい。画像データは、一例として、利用者を撮影した画像を含んでもよい。音響データは、一例として、利用者の発話した音声が含まれてもよい。音響データは、テキストデータを音声合成した音声データでもよい。

[0093] 発話生成部１２０は、生成装置３０が生成モデル入力情報を生成モデルＭ２に入力することで生成された出力情報を、生成装置３０から受信してもよい。生成モデル入力情報は、システムプロンプトと対話履歴情報とが別々に生成モデルＭ２に入力されてもよい。生成モデル入力情報は、システムプロンプトと対話履歴情報とが結合されて生成モデルＭ２に入力されてもよい。

[0094] 発話生成部１２０は、生成モデルＭ２の出力情報に基づいて、システム発話情報を生成してもよい。生成モデルＭ２の出力情報が、システム発話を示す音声信号又はＣＧキャラクタを表す映像信号を含む場合、発話生成部１２０は、当該音声信号又は当該映像信号の少なくとも一方をシステム発話情報として用いてもよい。生成モデルＭ２の出力情報が、発話内容を示すテキストデータを含む場合、発話生成部１２０は、当該テキストデータに基づいて音声信号を合成してもよい。生成モデルＭ２の出力情報が、対話装置１０の表現を制御するための情報（一例として、表情ラベル、発話時の口の形の情報、又は姿勢情報等）を含む場合、発話生成部１２０は、当該情報に基づい

てＣＧキャラクタを表す映像信号を合成してもよい。

[0095] 発話生成部１２０は、生成モデルＭ２の出力情報の一部に基づいて、システム発話情報を生成してもよい。発話生成部１２０は、生成モデルＭ２の出力情報が所定のデータ長又はランダムに決定したデータ長以上である場合、当該データ長以下となるように生成モデルＭ２の出力情報を短縮してもよい。生成モデルＭ２がストリームで出力情報を出力する場合、データ長を満たした時点で、生成モデルＭ２に後続の出力を停止させてもよい。

[0096] 一例として、発話生成部１２０は、生成モデルＭ２の出力情報の先頭から順にＮ個の文を取得してもよい。なお、Ｎは正の整数であり、ランダムに定めてもよい。一例として、発話生成部１２０は、最大M字以下となるように、生成モデルＭ２の出力情報の先頭から順にＮ′個の文を取得してもよい。なお、Ｍ，Ｎ′は正の整数である。文の境界は、所定の記号により特定してもよい。一例として、句点（「。」又は「．」等）、記号（「？」又は「！」等）を区切り文字として用いてもよい。

[0097] 発話生成部１２０は、生成したシステム発話情報を対話記憶部１０２に記憶してもよい。発話生成部１２０は、リアルタイムにシステム発話情報を記憶してもよく、所定の処理単位で一括してシステム発話情報を記憶してもよい。所定の処理単位は、一例として、システム発話情報のデータ量、対話のターン数、対話の経過時間等でもよい。

[0098] 発話生成部１２０は、対話記憶部１０２に記憶されている対話データＤにシステム発話情報を含めてもよい。発話取得部１１０は、対話データＤに含まれる時刻情報と関連付けてシステム発話情報を対話記憶部１０２に記憶してもよい。

[0099] （発話提示部）

発話提示部１３０は、対話装置１０の発話を制御する。発話提示部１３０は、システム発話を合成した音声信号を対話装置１０に送信してもよい。対話装置１０は、制御装置２０から受信した音声信号に基づいて、システム発話を示す音声をスピーカから出力してもよい。

[0100] 発話提示部１３０は、発話内容を示すテキストデータを対話装置１０に送信してもよい。対話装置１０は、制御装置２０から受信したテキストデータを合成することで、音声信号を生成し、スピーカから出力してもよい。

[0101] 発話提示部１３０は、ＣＧキャラクタを表す映像信号を対話装置１０に送信してもよい。対話装置１０は、制御装置２０から受信した映像信号に基づいて、ＣＧキャラクタを表す映像を表示装置に表示してもよい。

[0102] 発話提示部１３０は、映像信号の合成に用いる情報（一例として、表情ラベル、発話時の口の形の情報、又は姿勢情報等）を対話装置１０に送信してもよい。対話装置１０は、制御装置２０から受信した情報に基づいて映像信号を合成し、ＣＧキャラクタを表す映像を表示装置に表示してもよい。

[0103] 発話提示部１３０は、ＣＧキャラクタの表現を制御するための制御信号を対話装置１０に送信してもよい。対話装置１０は、制御装置２０から受信した制御信号に基づいて、ＣＧキャラクタの表現を制御してもよい。

[0104] 発話提示部１３０は、対話の状況を説明するシステム発話（以下、「説明発話」とも呼ぶ）を、対話装置１０に発話させる制御を行ってもよい。発話提示部１３０は、シナリオ記憶部１０１に記憶されたガイダンス情報に基づいて、説明発話を対話装置１０に発話させてもよい。発話提示部１３０は、シナリオ記憶部１０１に記憶されたガイダンス情報を対話装置１０に送信してもよい。発話提示部１３０は、ガイダンス情報を合成した音声信号を対話装置１０に送信してもよい。

[0105] 発話提示部１３０は、ユーザ発話に応答するシステム発話を、対話装置１０に発話させる制御を行ってもよい。発話提示部１３０は、発話生成部１２０が生成したシステム発話情報に基づいて、ユーザ発話に応答するシステム発話を対話装置１０に発話させてもよい。発話提示部１３０は、システム発話情報に含まれるテキストデータ、音声信号、映像信号又は制御信号の少なくとも１つを対話装置１０に送信してもよい。

[0106] （対話評価部）

対話評価部１４０は、対話者Ｓに関する評価結果を生成する。対話評価部

１４０は、対話記憶部１０２から読み出した対話データＤに基づいて、対話者Ｓに関する評価結果を生成してもよい。対話評価部１４０は、モデル記憶部１０３から読み出した評価モデルＭ１に基づいて、対話者Ｓに関する評価結果を生成してもよい。対話評価部１４０は、対話データＤに基づいて評価モデルＭ１への入力情報を生成し、評価モデルＭ１への入力情報を評価モデルＭ１へ入力してもよい。評価モデルＭ１への入力情報は、対話データＤから抽出した特徴量を含んでもよい。以下、評価モデルＭ１への入力情報を、「評価モデル入力情報」とも呼ぶ。

[0107] 対話者Ｓに関する評価結果は、１以上の評価値を含んでもよい。対話評価部１４０は、評価者Ｅの評価基準に従って、評価結果に含める評価値を選択してもよい。対話評価部１４０は、評価者Ｅの評価基準に応じた評価モデルＭ１に基づいて、対話者Ｓに関する評価結果を生成してもよい。対話評価部１４０は、評価結果に含める評価値に応じて、評価モデル入力情報に含める特徴量を対話データＤから抽出してもよい。

[0108] 対話評価部１４０は、モデル記憶部１０３に複数の評価モデルＭ１が記憶されている場合、複数の評価モデルＭ１それぞれに基づいて、対話者Ｓに関する評価結果を生成してもよい。対話者Ｓに関する評価結果は、複数の評価モデルＭ１それぞれに基づく複数の評価結果を含んでもよい。対話者Ｓに関する評価結果は、複数の評価モデルＭ１それぞれが出力する複数の評価値を含む１つの評価結果を含んでもよい。

[0109] 対話者Ｓに関する評価結果は、評価値の確からしさを示す信頼度を含んでもよい。評価値の信頼度は、評価モデルＭ１により評価値と共に出力されてもよい。一例として、評価モデルＭ１が分類モデルであれば、評価モデルＭ１は、分類対象の各クラスに対して確率を計算し、確率が最も高いクラスを評価値とし、そのクラスに対する確率を評価値の信頼度として出力すればよい。

[0110] 対話評価部１４０は、限定されない一例として、評価モデル入力情報を生成モデルＭ２に入力することで、対話者Ｓに関する評価結果を生成してもよ

い。評価モデルＭ１は、生成モデルＭ２に基づいて生成されてもよい。評価モデルＭ１は、一例として、生成モデルＭ２を転移学習することで生成されてもよい。評価モデルＭ１は、一例として、生成モデルＭ２をファインチューニングすることで生成されてもよい。

[0111] 対話評価部１４０は、任意のタイミングで、対話者Ｓに関する評価結果を生成してもよい。対話評価部１４０は、バッチ処理で評価結果を生成してもよく、アドホック処理で評価結果を生成してもよい。一例として、対話評価部１４０は、所定の時間間隔で、蓄積された対話データＤそれぞれについて評価結果を生成してもよい。対話評価部１４０は、対話が終了したことを契機として、当該対話に関する対話データＤについて評価結果を生成してもよい。

[0112] 対話評価部１４０は、対話者Ｓに関する評価結果を評価結果記憶部１０４に記憶してもよい。対話評価部１４０は、対話を特定可能な識別情報と関連付けて、評価結果を評価結果記憶部１０４に記憶してもよい。一例として、対話評価部１４０は、時刻情報と評価結果とを関連付けて評価結果記憶部１０４に記憶してもよい。

[0113] （結果出力部）

結果出力部１５０は、対話者Ｓに関する評価結果を出力する。結果出力部１５０は、評価結果記憶部１０４に記憶されている対話者Ｓに関する評価結果を読み出してもよい。

[0114] 結果出力部１５０は、対話者Ｓに関する外部評価値を、対話者Ｓに関する評価結果に含めて出力してもよい。結果出力部１５０は、評価結果記憶部１０４に記憶されている対話者Ｓに関する外部評価値を対話者Ｓに関する評価結果と共に読み出し、対話者Ｓに関する評価結果に外部評価値を含めてもよい。

[0115] 結果出力部１５０は、対話者Ｓの評価に用いた対話データＤを、対話者Ｓに関する評価結果に含めてもよい。一例として、結果出力部１５０は、対話データＤに含まれる録画データ、又は録画データから抽出された特徴量を、

対話者Ｓに関する評価結果に含めてもよい。録画データから抽出された特徴量は、評価モデルＭ１に入力された特徴量を含んでもよい。

[0116] 結果出力部１５０は、対話者Ｓに関する評価結果を端末装置４０に送信してもよい。結果出力部１５０は、端末装置４０からの出力要求に応じて、対話者Ｓに関する評価結果を端末装置４０に送信してもよい。結果出力部１５０は、対話者Ｓに関する評価結果が複数ある場合、評価結果に示される評価値の推移又は統計情報を端末装置４０に送信してもよい。端末装置４０は、対話者Ｓに関する評価結果を評価者Ｅに提示してもよい。

[0117] 結果出力部１５０は、対話者Ｓに関する評価結果を対話装置１０に送信してもよい。対話装置１０は、対話者Ｓに関する評価結果を対話者Ｓに提示してもよい。結果出力部１５０は、対話者Ｓに関する評価結果を外部の情報処理装置又は情報処理システムに送信してもよい。外部の情報処理装置又は情報処理システムは、対話者Ｓに関する評価結果を対話者Ｓ以外の利用者に提示してもよい。対話者Ｓ以外の利用者は、評価者Ｅを含んでもよい。

[0118] 結果出力部１５０は、一例として、評価結果を表示する画面データを端末装置４０に送信してもよい。画面データは、一例として、ＨＴＭＬ（Hyper Text Markup Language）等で記述された電子データでもよい。結果出力部１５０は、評価結果を表示する画面データを対話装置１０に送信してもよい。

[0119] 結果出力部１５０は、一例として、電子メール又はメッセージングサービス等を通じて評価結果を評価者Ｅ又は対話者Ｓのアカウントに送信してもよい。結果出力部１５０は、評価結果を表示する画面へのリンク情報を電子メール又はメッセージングサービス等を通じて評価者Ｅ又は対話者Ｓのアカウントに送信してもよい。

[0120] 図２に示す制御装置２０の機能構成は一例であり、用途や目的に応じて様々な機能構成例があることは言うまでもない。図２に示すシナリオ記憶部１０１、対話記憶部１０２、モデル記憶部１０３、評価結果記憶部１０４のような記憶部の区分は一例である。図２に示す発話取得部１１０、発話生成部１２０、発話提示部１３０、対話評価部１４０、結果出力部１５０のような

処理部の区分は一例である。

[0121] 例えば、シナリオ記憶部１０１、対話記憶部１０２、モデル記憶部１０３又は評価結果記憶部１０４の少なくとも２つを１つの記憶部に統合してもよい。また、例えば、対話記憶部１０２、モデル記憶部１０３又は評価結果記憶部１０４の少なくとも１つを複数の記憶部に分割してもよい。

[0122] 例えば、発話取得部１１０、発話生成部１２０、発話提示部１３０、対話評価部１４０又は結果出力部１５０の少なくとも２つを１つの処理部に統合してもよい。また、例えば、発話取得部１１０、発話生成部１２０、発話提示部１３０、対話評価部１４０又は結果出力部１５０の少なくとも１つを複数の処理部に分割してもよい。

[0123] ＜対話方法の流れ＞

対話システム１０００が実行する対話方法について、図３を参照しながら説明する。図３は、第１実施形態に係る対話方法の一例を示す処理フローである。以下では、ＣＧキャラクタを用いて、対話装置１０の一例であるパーソナルコンピュータを操作する対話者Ｓと対話する例を説明する。

[0124] ステップＳ１０１において、対話者Ｓは、対話装置１０に対して、対話を開始する操作を行う。一例として、対話者Ｓは、対話装置１０の表示装置に表示された制御装置２０の一覧から、対話を行う制御装置２０を選択する操作を行ってもよい。対話者Ｓが制御装置２０を選択する操作を行うと、対話装置１０は制御装置２０に接続する処理を実行してもよい。

[0125] 対話者Ｓは、対話装置１０のマイクロフォン又はカメラのテストを実行してもよい。対話者Ｓは、対話装置１０のマイクロフォンのスイッチをオンにし、マイクロフォンの音量レベルを確認してもよい。このとき、対話装置１０と制御装置２０は協働して、対話者Ｓに特定の文章を読み上げるように指示し、対話者Ｓが指定された文章を読み上げ、読み上げられた文章を音声認識し、認識されたテキストが指示した文章と同じ又は類似していればマイクが適切に作動していると判定してもよい。対話者Ｓは、対話装置１０のカメラのスイッチをオンにし、表示装置に表示されたカメラ画像を確認してもよ

い。このとき、対話装置１０は、顔又は身体の検出器を用いて、カメラ画像に対話者Ｓが適切に撮像されているか否かを判定してもよい。対話装置１０は、カメラ画像に対話者Ｓが適切に撮像されるまで、対話を開始しないように制御してもよい。カメラ画像に対話者Ｓが適切に撮像されていることは、例えば、顔が予め定められた大きさより大きく写っていること、顔が正面を向いていること、身体が予め定められた範囲で写っていること等を含んでもよい。

[0126] ステップＳ１０２において、制御装置２０の発話提示部１３０は、シナリオ記憶部１０１に記憶されたガイダンス情報を読み出す。発話提示部１３０は、ガイダンス情報を音声合成する。これにより、説明発話を示す音声信号が生成される。発話提示部１３０は、説明発話を示す音声信号を対話装置１０に送信する。なお、発話提示部１３０は、ガイダンス情報を音声信号と共に又は単独で対話装置１０に送信してもよい。

[0127] ステップＳ１０３において、対話装置１０は、制御装置２０から説明発話を示す音声信号を受信する。対話装置１０は、説明発話を示す音声信号に基づいて、説明発話をスピーカから出力する。対話装置１０は、制御装置２０からガイダンス情報を受信した場合、ガイダンス情報を表示装置に表示してもよい。また、対話装置１０は、制御装置２０からガイダンス情報を受信した場合、ガイダンス情報を合成した音声信号をスピーカから出力してもよい。

[0128] ステップＳ１０４において、制御装置２０は、対話者Ｓと対話装置１０との対話の記録を開始する。具体的には、制御装置２０は、ユーザ発話情報を対話記憶部１０２に記憶するように、発話取得部１１０に要求する。また、制御装置２０は、システム発話情報を対話記憶部１０２に記憶するように、発話生成部１２０に要求する。発話取得部１１０は、ユーザ発話情報を取得すると、当該ユーザ発話情報を対話記憶部１０２に記憶する機能を有効にする。発話生成部１２０は、システム発話情報を生成すると、当該システム発話情報を対話記憶部１０２に記憶する機能を有効にする。

[0129] ステップＳ１０５において、対話装置１０は、ＣＧキャラクタを表示装置に表示する。対話装置１０は、マイクロフォンにより収音される音声信号及びカメラにより撮像される映像信号を制御装置２０に送信する。以降、対話装置１０は、対話が終了するまで、音声信号及び映像信号の送信を継続する。

[0130] 対話者Ｓは、ステップＳ１０３で出力された説明発話に応じて、発話を行う。対話装置１０は、対話者Ｓによるユーザ発話を受け付ける。対話装置１０が制御装置２０に送信する音声信号には、対話者Ｓがユーザ発話を発話した音声が含まれる。また、対話装置１０が制御装置２０に送信する映像信号には、対話者Ｓがユーザ発話を発話したときの表情又は動作を撮像した画像が含まれる。

[0131] ステップＳ１０６において、制御装置２０の発話取得部１１０は、対話装置１０からユーザ発話を収音した音声信号を受信する。また、発話取得部１１０は、対話装置１０から対話者Ｓを撮像した映像信号を受信する。

[0132] 発話取得部１１０は、ユーザ発話を収音した音声信号を音声認識する。これにより、ユーザ発話の発話内容を示すテキストデータが生成される。発話取得部１１０は、ユーザ発話の発話内容を示すテキストデータを発話生成部１２０に送る。

[0133] 発話取得部１１０は、ユーザ発話を収音した音声信号、対話者Ｓを撮像した映像信号、及びユーザ発話の発話内容を示すテキストデータに基づいて、ユーザ発話情報を生成する。発話取得部１１０は、ユーザ発話情報を対話記憶部１０２に記憶する。

[0134] ステップＳ１０７において、制御装置２０の発話生成部１２０は、発話取得部１１０からユーザ発話の発話内容を示すテキストデータを受け取る。発話生成部１２０は、シナリオ記憶部１０１に記憶されたシステムプロンプトを読み出す。発話生成部１２０は、対話記憶部１０２から対話データＤを読み出す。発話生成部１２０は、ユーザ発話の発話内容を示すテキストデータ及び対話記憶部１０２から読み出した対話データＤに基づいて、対話履歴情

報を生成する。

[0135] 発話生成部１２０は、システムプロンプト及び対話履歴情報に基づいて、生成モデル入力情報を生成する。発話生成部１２０は、システムプロンプト及び対話履歴情報を所定のテンプレートに埋め込むことで、生成モデル入力情報を生成してもよい。発話生成部１２０は、生成モデル入力情報を生成装置３０に送信する。

[0136] 生成装置３０は、制御装置２０から生成モデル入力情報を受信する。生成装置３０は、受信した生成モデル入力情報を生成モデルＭ２に入力する。生成モデルＭ２は、生成モデル入力情報に基づいて所定のタスクを実行し、結果として生成された出力情報を出力する。生成モデルＭ２からの出力情報は、システム発話の発話内容を示すテキストデータを少なくとも含んでもよい。生成装置３０は、生成モデルＭ２から出力された出力情報を制御装置２０に送信する。

[0137] ステップＳ１０８において、制御装置２０の発話生成部１２０は、生成装置３０から出力情報を受信する。発話生成部１２０は、受信した出力情報の少なくとも一部に基づいて、システム発話情報を生成する。発話生成部１２０は、システム発話の発話内容を示すテキストデータを音声合成してもよい。これにより、システム発話を示す音声信号が生成される。システム発話を示す音声信号は、生成モデルＭ２からの出力情報に含まれてもよい。

[0138] 発話生成部１２０は、ＣＧキャラクタを表す映像信号を生成してもよい。発話生成部１２０は、生成モデルＭ２からの出力情報に含まれる情報に基づいて、ＣＧキャラクタを表す映像信号を合成してもよい。ＣＧキャラクタを表す映像信号は、生成モデルＭ２からの出力情報に含まれてもよい。言い換えると、ＣＧキャラクタを表す映像信号は、生成モデルＭ２が直接生成してもよい。

[0139] 発話生成部１２０は、システム発話を示す音声信号、ＣＧキャラクタを表す映像信号、及びシステム発話の発話内容を示すテキストデータに基づいて、システム発話情報を生成する。発話取得部１１０は、システム発話情報を

対話記憶部１０２に記憶する。発話生成部１２０は、システム発話情報を対話装置１０に送信する。

[0140] ステップＳ１０９において、対話装置１０は、制御装置２０からシステム発話情報を受信する。対話装置１０は、システム発話情報に基づいて、システム発話を発話する。対話装置１０は、システム発話を示す音声をスピーカから出力してもよい。対話装置１０は、ＣＧキャラクタを表す映像を表示装置に表示してもよい。対話装置１０がロボットの場合は、システム発話情報に基づいてロボットを制御してもよい。なお、対話装置１０は、システム発話を発話している間、マイクロフォンのスイッチをオフにしてもよい。言い替えると、対話装置１０は、システム発話を発話するまでは、マイクロフォンで対話者Ｓによる発話を収音してもよい。

[0141] ステップＳ１１０において、制御装置２０は、対話を終了するための終了条件を満たしたか否かを判定する。終了条件は、一例として、対話のターン数又は対話を開始してからの経過時間が所定の閾値を超えたことでもよい。図３では、ステップＳ１０８の直後のみに終了判定の処理を入れているが、経過時間で判定する場合は、ステップＳ１０６の直前からステップＳ１０８の直後の任意のタイミングで終了判定処理をしてもよい。この場合、対話者Ｓの各発話の開始から終了後あらかじめ定めておいた時間が経過するまでは終了判定を行わないようにしてもよい。このようにすることで、唐突に会話が終了することを防ぎつつ、決められた時間に近い時間経過で対話を終了することができる。

[0142] ステップＳ１０６の直前からステップＳ１０８の直後の任意のタイミングで、顔や体がカメラに写っているか否かの確認を顔や身体の検出器を用いて行ってもよい。一定時間検出できないときは、対話を終了し、ステップＳ１１１に処理を進める。対話の終了前に、対話者Ｓに顔や身体が検出できない旨のメッセージを対話装置１０がテキストで表示または音声で出力してもよい。

[0143] 制御装置２０は、対話の内容に基づいて、対話を終了するか否かを判定し

てもよい。一例として、制御装置２０は、対話履歴情報を機械学習モデルに入力し、機械学習モデルからの出力情報に基づいて、対話を終了するか否かを判定してもよい。対話の終了を判定する機械学習モデルは、生成モデルＭ２でもよく、他の大規模言語モデル又は基盤モデルでもよい。

[0144] 終了条件を満たしたと判定した場合（ＹＥＳ）、制御装置２０はステップＳ１１１に処理を進める。制御装置２０は、終了条件を満たしたと判定した場合に、所定の発話を示す音声信号を対話装置１０に送信してもよい。所定の発話は、対話が終了したことを示す発話でもよい。対話装置１０は、制御装置２０から所定の発話を示す音声信号を受信し、所定の発話をスピーカから出力してもよい。

[0145] 一方、終了条件を満たしていないと判定した場合（ＮＯ）、制御装置２０はステップＳ１０６に処理を戻す。ステップＳ１０６に処理を戻すと、制御装置２０の発話取得部１１０は、次のユーザ発話に関するユーザ発話情報を取得する。その後、制御装置２０は、新たなユーザ発話情報に基づいて、ステップＳ１０６からステップＳ１１０までの処理を再度実行する。このようにして、対話システム１０００は、ステップＳ１１０で終了条件を満たしたと判定するまで、ユーザ発話の受け付けと、システム発話の提示とを繰り返し実行する。なお、発話生成部１２０は、繰り返しの都度、システム発話の生成に用いるシステムプロンプトを変更してもよい。

[0146] ステップＳ１１１において、制御装置２０の対話評価部１４０は、対話記憶部１０２から対話者Ｓとの対話に関する対話データＤを読み出す。また、対話評価部１４０は、モデル記憶部１０３から評価者Ｅに応じた評価モデルＭ１を読み出す。

[0147] 対話評価部１４０は、読み出した対話データＤに基づいて、評価モデル入力情報を生成する。対話評価部１４０は、評価モデル入力情報を評価モデルＭ１に入力する。評価モデルＭ１は、評価モデル入力情報に基づいて、１以上の評価値を計算する。これにより、対話者Ｓに関する評価結果が生成される。対話評価部１４０は、対話者Ｓに関する評価結果を評価結果記憶部１０

４に記憶する。

[0148] ステップＳ１１２において、評価者Ｅは、端末装置４０に対して、評価結果を表示する操作を行う。一例として、評価者Ｅは、端末装置４０の表示装置に表示された評価画面において、対話者Ｓの評価結果を開く操作を行ってもよい。端末装置４０は、対話者Ｓに関する評価結果の出力要求を制御装置２０に送信する。

[0149] 制御装置２０の結果出力部１５０は、評価結果の出力要求を端末装置４０から受信する。結果出力部１５０は、対話者Ｓに関する評価結果を評価結果記憶部１０４から読み出す。結果出力部１５０は、対話者Ｓに関する評価結果を表示する画面データを生成する。結果出力部１５０は、画面データを端末装置４０に送信する。

[0150] ステップＳ１１３において、端末装置４０は、制御装置２０から画面データを受信する。端末装置４０は、画面データに基づいて、対話者Ｓに関する評価結果を表示装置に表示する。評価者Ｅは、端末装置４０の表示装置に表示された対話者Ｓに関する評価結果を参照し、対話者Ｓを評価してもよい。一例として、評価者Ｅは、対話者Ｓの評価結果を参考にして、対話者Ｓを採用するか否かを判断してもよい。一例として、研修の講師である評価者Ｅは、対話者Ｓの評価結果を参考にして、研修の良し悪しを判断してもよい。一例として、試験の採点者である評価者Ｅは、対話者Ｓの評価結果を参考にして、対話者Ｓの得点を決定し、試験の合否を判定してもよい。

[0151] 端末装置４０は、対話者Ｓに関する評価結果に含まれる録画データを表示装置に表示してもよい。評価者Ｅは、評価結果に含まれる録画データと評価値とを比較し、対話者Ｓに関する評価結果が適切か否かを判断してもよい。

[0152] 端末装置４０は、対話者Ｓに関する評価結果を修正する画面を表示装置に表示してもよい。評価者Ｅは、端末装置４０において、対話者Ｓに関する評価結果を修正してもよい。端末装置４０は、評価者Ｅにより修正された評価結果を制御装置２０に送信してもよい。制御装置２０は、端末装置４０から受信した評価結果に基づいて、評価結果記憶部１０４に記憶されている対話

者Ｓに関する評価結果を更新してもよい。制御装置２０は、更新された評価結果に基づいて、評価モデルＭ１を追加学習してもよい。

[0153] ［第１実施形態の変形例１］

対話システム１０００は、評価モデルＭ１とシステムプロンプトとを関連付けて記憶してもよい。一例として、評価モデルＭ１は、システムプロンプトを識別する識別情報と関連付けて、モデル記憶部１０３に記憶されてもよい。対話評価部１４０は、システムプロンプトに関連付けられた評価モデルＭ１をモデル記憶部１０３から読み出し、当該評価モデルＭ１に基づいて、１以上の評価値を計算してもよい。

[0154] 対話システム１０００は、評価モデルＭ１とシナリオ情報とを関連付けて記憶してもよい。一例として、評価モデルＭ１を識別する識別情報を含むシナリオ情報が、シナリオ記憶部１０１に記憶されてもよい。発話生成部１２０は、シナリオに応じたシステムプロンプトをシナリオ記憶部１０１から読み出し、生成モデルＭ２への入力情報に含めてもよい。対話評価部１４０は、シナリオに応じた評価モデルＭ１をモデル記憶部１０３から読み出し、当該評価モデルＭ１に基づいて、１以上の評価値を計算してもよい。対話システム１０００は、シナリオを選択するのみの操作で、システムプロンプト及び評価モデルＭ１を選択可能となる。

[0155] ［第１実施形態の変形例２］

制御装置２０は、対話装置１０の制御に用いた情報を記憶してもよい。対話装置１０の制御に用いた情報は、一例として、システム発話を示す音声信号の合成に用いたテキストデータ、合成した音声信号の開始時刻及び終了時刻、対話装置１０の表情又は身体動作（例えば、笑う、手を振る、頷く等）の制御に用いたラベル、当該ラベルに基づく動作を示す動画の開始時刻及び終了時刻等を含んでもよい。なお、時刻は、日本標準時（JST; Japan Standard Time）又は協定世界時（UTC; Coordinated Universal Time）、若しくは対話の録画を開始した時刻からの経過時間でもよい。

[0156] 対話装置１０の制御に用いた情報は、対話装置１０の記憶装置に記憶され

てもよい。対話装置１０の制御に用いた情報は、生成装置３０の記憶装置に記憶されてもよい。対話装置１０の制御に用いた情報は、対話装置１０、制御装置２０又は生成装置３０とは異なる他の情報処理装置又は記憶装置に記憶されてもよい。

[0157] 対話装置１０の制御に用いた情報は、音声信号又は映像信号と共に評価値の計算に用いられてもよい。一例として、音声信号の合成に用いた情報を用いれば、当該音声信号の再生の開始時刻及び終了時刻に基づいて、対話全体の録画データから対話装置１０の発話区間を特定できる。また、音声信号の合成に用いた情報を用いれば、音声認識を用いることなく、対話装置１０の発話内容をテキストとして取得できる。一例として、対話装置１０の表情又は身体動作に関する情報を用いれば、当該表情又は身体動作に対して、対話者Ｓがどのような内容を応答したのか、どのような表情をしたのか、どのような声の調子（トーン）で応答したのか等を評価モデルＭ１に追加情報として入力できる。

[0158] ［第１実施形態の変形例３］

発話生成部１２０は、検索拡張生成（RAG: Retrieval Augmented Generation）と呼ばれる手法を用いて、システム発話を生成してもよい。検索拡張生成は、機械学習モデルから良好な生成結果を得るために、外部のデータソースを検索して得た参照情報をプロンプトに含める技術である。

[0159] 制御装置２０は、対話装置１０に関する背景情報を含む参照情報を予め記憶しておいてもよい。発話生成部１２０は、ユーザ発話情報に基づいて、クエリを生成し、当該クエリに基づいて、参照情報を検索してもよい。発話生成部１２０は、検索された参照情報に基づいてシステムプロンプトを生成してもよい。発話生成部１２０は、生成したシステムプロンプトを生成モデルＭ２への入力情報に含め、生成モデルＭ２への入力情報を生成装置３０に送信してもよい。

[0160] 一例として、対話者Ｓが医師であり、対話装置１０が患者であるシナリオで対話を実行する例を説明する。参照情報には、症状が発生した経緯、現在

の症状、病歴、持病、投薬履歴、バイタルサイン、表情、外傷であればその様子を記述したテキスト等を含めておく。発話生成部１２０は、対話者Ｓによる質問に基づいてクエリを生成し、当該クエリに基づいて対話者Ｓによる質問に関係する参照情報を取得する。発話生成部１２０は、取得した参照情報を含むシステムプロンプトを生成し、生成モデルＭ２に基づいて対話者Ｓによる質問に応答するシステム発話を生成する。これにより、対話装置１０は、対話者Ｓによる質問に適切に応答できる。

[0161] なお、ここでは、説明を簡単にするために、参照情報はテキスト情報であることを前提に説明したが、参照情報はテキスト情報に限定されるものではなく、音声情報又は画像情報等を含んでもよい。

[0162] ［第２実施形態］

第１実施形態では、予め構築された評価モデルＭ１に基づいて、対話者Ｓに関する評価結果を生成する構成を説明した。第２実施形態では、評価者Ｅの評価基準に応じた評価モデルＭ１を構築し、評価者Ｅの評価基準に応じた評価モデルＭ１に基づいて、対話者Ｓに関する評価結果を生成する構成を説明する。

[0163] 以下、第２実施形態に係る対話システム１０００について、第１実施形態との相違点を中心に説明する。

[0164] ＜対話システムの全体構成＞

本実施形態に係る対話システムの全体構成について、図４を参照しながら説明する。図４は、第２実施形態に係る対話システムの全体構成の一例を示すブロック図である。

[0165] 図４に示されるように、対話システム１０００は、対話装置１０、制御装置２０、生成装置３０、端末装置４０及び管理装置５０を含む。すなわち、本実施形態に係る対話システム１０００は、第１実施形態に係る対話システム１０００（図１参照）と比較すると、管理装置５０をさらに含む点が異なる。

[0166] 管理装置５０は、評価モデルＭ１を管理するパーソナルコンピュータ、ワ

ークステーション、サーバ等の情報処理装置の一例である。管理装置５０は、評価モデルＭ１を生成してもよい。管理装置５０は、評価モデルＭ１の生成に用いる教師データを生成してもよい。管理装置５０は、対話者Ｓの評価に用いる評価モデルＭ１を選択してもよい。管理装置５０は、対話者Ｓの評価に用いる評価モデルＭ１を制御装置２０に送信してもよい。

[0167] 管理装置５０は、対話の録画データＶを記憶してもよい。録画データＶは、第１の対話者と第２の対話者との対話を記録した電子データでもよい。録画データＶは、第１の対話者又は第２の対話者による発話を収音した音声信号を含んでもよい。録画データＶは、第１の対話者又は第２の対話者を撮像した映像信号を含んでもよい。

[0168] 第１の対話者は、実在の人物でもよく、対話装置１０でもよい。第２の対話者は、実在の人物でもよく、対話装置１０でもよい。言い替えると、録画データＶは、実在の人物同士の対話を記録した録画データでもよく、実在の人物と対話装置１０との対話を記録した録画データでもよく、対話装置１０同士の対話を記録した録画データでもよい。

[0169] 管理装置５０は、評価モデル群Ｍ３を備えてもよい。評価モデル群Ｍ３は、複数の評価モデルＭ１を含んでもよい。評価モデル群Ｍ３は、評価基準が異なる複数の評価モデルＭ１を含んでもよい。評価モデルＭ１は、評価基準に従った正解ラベルが付与された録画データＶに基づいて、生成されてもよい。正解ラベルは、評価値の正解値でもよい。

[0170] 評価基準は、評価者Ｅごとに定められてもよい。評価基準は、評価者Ｅの属性に基づいて定められてもよい。一例として、評価基準は、評価者Ｅの業種ごとに定められてもよい。評価基準は、対話者Ｓの属性に基づいて定められてもよい。一例として、評価基準は、対話者Ｓの職種ごとに定められてもよい。一例として、評価基準は、シナリオごとに定められてもよい。にこやかに応対するほうがよいシナリオと、無表情で応対するほうがよいシナリオがあった場合、それぞれに評価基準が異なると考えられるためである。

[0171] 特定の評価基準に従った評価は、以下のように行ってもよい。まず、対話

の録画データＶに基づいて、１以上の特徴量を抽出する。１以上の特徴量は、評価基準に依存しない特徴量でよい。次に、抽出された１以上の特徴量に基づいて、評価者Ｅの評価基準に沿った評価を行う。このとき、１以上の特徴量に対する線形変換により評価値を計算してもよい。また、線形変換後の各特徴量を足し合わせた値を評価値として計算してもよい。また、１以上の機械学習モデルに基づいて評価値を取得してもよい。

[0172] なお、図４に示した対話システム１０００の全体構成は一例であって、用途や目的に応じて様々なシステム構成例があり得る。一例として、対話装置１０、制御装置２０、生成装置３０、端末装置４０又は管理装置５０の１つ以上が、対話システム１０００に含まれてもよい。管理装置５０は、複数台のコンピュータにより実現してもよいし、クラウドコンピューティングのサービスとして実現してもよい。管理装置５０は、制御装置２０又は生成装置３０と一体化したスタンドアローンのコンピュータにより実現してもよい。

[0173] 対話システム１０００は、評価者Ｅごとに設置された複数台の制御装置２０を含んでもよい。対話システム１０００は、複数の評価者Ｅにより使用される１台の制御装置２０を含んでもよい。対話システム１０００は、評価者Ｅごとに設置された複数台の管理装置５０を含んでもよい。対話システム１０００は、複数の評価者Ｅにより使用される１台の制御装置２０を含んでもよい。

[0174] ＜管理装置の機能構成＞

管理装置５０の機能構成について、図５を参照しながら説明する。図５は、第２実施形態に係る管理装置の機能構成の一例を示すブロック図である。

[0175] 図５に示されるように、管理装置５０は、データ記憶部５０１、モデル記憶部５０２、ラベル付与部５１０、特徴抽出部５２０、モデル学習部５３０、モデル評価部５４０及びモデル選択部５５０を備える。管理装置５０は、予めインストールされた管理プログラムが実行されることで、データ記憶部５０１、モデル記憶部５０２、ラベル付与部５１０、特徴抽出部５２０、モデル学習部５３０、モデル評価部５４０及びモデル選択部５５０として機能

する。

[0176] （データ記憶部）

データ記憶部５０１には、録画データＶが記憶される。録画データＶは、制御装置２０により生成されてもよい。録画データＶは、対話装置１０を用いて記録されてもよい。録画データＶは、他の情報処理装置又は情報処理システムにより生成されてもよい。

[0177] 録画データＶは、制御装置２０に記憶される対話データＤに含まれる録画データを含んでもよい。一例として、録画データＶは、予め設定したデータ収集期間中に制御装置２０に記憶された対話データＤから抽出された録画データでもよい。

[0178] 録画データＶは、不特定の利用者から収集された録画データを含んでもよい。一例として、対話の録画データは、クラウドソーシング等の手法により収集されてもよい。録画データＶは、任意の目的で記録されたものでもよい。言い替えると、録画データＶは、評価モデルＭ１の構築を目的として記録されたものでなくてもよい。

[0179] 録画データＶは、一方の対話者が、評価者Ｅにより指定された対話者Ｓである対話を記録した録画データでもよい。評価者Ｅは、評価が定まっている対話者Ｓを指定してもよい。評価者Ｅは、評価が高い対話者Ｓを指定してもよい。評価者Ｅは、評価が低い対話者Ｓを指定してもよい。

[0180] （モデル記憶部）

モデル記憶部５０２には、評価モデル群Ｍ３が記憶される。評価モデル群Ｍ３は、評価基準が異なる複数の評価モデルＭ１を含んでもよい。評価モデル群Ｍ３に含まれる評価モデルＭ１は、予めモデル記憶部５０２に記憶されていてもよい。評価モデル群Ｍ３に含まれる評価モデルＭ１は、端末装置４０から取得してもよい。端末装置４０から取得する評価モデルＭ１は、評価者Ｅにより生成されてもよい。評価モデル群Ｍ３に含まれる評価モデルＭ１は、モデル学習部５３０が生成し、モデル記憶部５０２に記憶されてもよい。

[0181] 評価モデル群Ｍ３に含まれる評価モデルＭ１は、第１実施形態と同様である。したがって、評価モデルＭ１は、対話者Ｓの表情を評価する評価モデルを含んでもよい。評価モデルＭ１は、対話者Ｓの身体動作を評価する評価モデルを含んでもよい。評価モデルＭ１は、対話者Ｓの声を評価する評価モデルを含んでもよい。評価モデルＭ１は、対話の円滑度を評価する評価モデルを含んでもよい。評価モデルＭ１は、対話内容を評価する評価モデルを含んでもよい。

[0182] （ラベル付与部）

ラベル付与部５１０は、録画データＶに正解ラベルを付与する。ラベル付与部５１０は、データ記憶部５０１に記憶されている録画データＶに正解ラベルを付与してもよい。ラベル付与部５１０は、評価基準に沿って録画データＶに正解ラベルを付与してもよい。ラベル付与部５１０は、評価者Ｅにより指定された正解ラベルを録画データＶに付与してもよい。

[0183] 正解ラベルは、評価結果に含まれる評価値の正解値を示す情報である。正解ラベルは、一例として、数値（例えば、１～１０等）、文字（例えば、アルファベット等）又は記号（例えば、〇、×、△等）を含んでもよい。

[0184] 正解ラベルは、録画データＶに含まれる音声を評価した評価値を含んでもよい。正解ラベルは、録画データＶに含まれる映像を評価した評価値を含んでもよい。正解ラベルは、録画データＶに含まれる対話の内容を評価した評価値を含んでもよい。

[0185] ラベル付与部５１０は、評価者Ｅにより指定された対話者Ｓに関する録画データに付与された正解ラベルを補正してもよい。ラベル付与部５１０は、評価者Ｅの評価が高い対話者Ｓに関する録画データに付与された正解ラベルを、評価が高い正解ラベルに補正してもよい。ラベル付与部５１０は、評価者Ｅの評価が低い対話者Ｓに関する録画データに付与された正解ラベルを、評価が低い正解ラベルに補正してもよい。

[0186] 評価者Ｅが高く評価する対話者Ｓは、評価者Ｅの評価基準に従った対話をできる対話者であると考えられる。また、評価者Ｅが低く評価する対話者Ｓ

は、評価者Ｅの評価基準に従っていない対話をする対話者であると考えられる。評価者Ｅにより指定された対話者Ｓに関する録画データに、評価者Ｅの評価に応じた正解ラベルを付与すれば、評価者Ｅの評価基準に沿った評価結果を生成できる。

[0187] ラベル付与部５１０は、機械学習モデルに基づいて、正解ラベルを付与してもよい。機器学習モデルは、学習済みの評価モデルＭ１でもよく、生成モデルＭ２でもよく、他の機械学習モデルでもよい。他の機械学習モデルは、一例として、大規模言語モデル又は基盤モデルでもよい。

[0188] ラベル付与部５１０は、機械学習モデルに基づいて付与された正解ラベルを修正してもよい。ラベル付与部５１０は、機械学習モデルに基づいて付与された正解ラベルを評価者Ｅに提示し、評価者Ｅにより指定された正解ラベルに修正してもよい。

[0189] （特徴抽出部）

特徴抽出部５２０は、録画データＶから特徴ベクトルを抽出する。特徴抽出部５２０は、評価モデル群Ｍ３に含まれる評価モデルＭ１に入力される特徴ベクトルを抽出してもよい。特徴抽出部５２０は、評価者Ｅにより指定された特徴量を含む特徴ベクトルを抽出してもよい。

[0190] 特徴ベクトルに含まれる特徴量は、第１実施形態と同様である。したがって、特徴ベクトルは、対話者Ｓの音声情報から抽出される特徴量を含んでもよい。音声情報から抽出される特徴量は、一例として、トーン、抑揚、話す速度、声の高さ、明瞭さ、話している時間、フィラーの回数等を含んでもよい。

[0191] 特徴ベクトルは、対話者Ｓの画像情報から抽出される特徴量を含んでもよい。画像情報から抽出される特徴量は、一例として、笑顔の度合い、無表情である度合い、怒り又は悲しみ等の感情を表す表情である度合い、視線の向き等を含んでもよい。

[0192] 特徴ベクトルは、対話内容から抽出される特徴量を含んでもよい。対話内容から抽出される特徴量は、一例として、肯定的な言葉の回数、否定的な言

葉の回数、丁寧な言葉遣いの割合等を含んでもよい。

[0193] （モデル学習部）

モデル学習部５３０は、評価モデル群Ｍ３に含める評価モデルＭ１を学習する。モデル学習部５３０は、録画データＶに正解ラベルが付与された教師データに基づいて、評価モデルＭ１を学習してもよい。

[0194] 具体的には、モデル学習部５３０は、特徴抽出部５２０が録画データＶから抽出した特徴ベクトルを評価モデルＭ１に入力する。評価モデルＭ１は、入力された特徴ベクトルに基づいて、１以上の評価値を予測し、評価値の予測値を出力する。モデル学習部５３０は、評価モデルＭ１から出力された予測値と、録画データＶに付与された正解ラベルとの差分情報を計算する。モデル学習部５３０は、予測値と正解ラベルとの差分情報に基づいて、評価モデルＭ１のパラメータを更新する。モデル学習部５３０は、所定の収束条件を満たすまで、評価値の予測とパラメータの更新を繰り返す。収束条件は、パラメータの更新回数が閾値以上となること、又は、パラメータの更新量が閾値未満となること、又は、差分情報から算出される数値が閾値未満となることでもよい。

[0195] モデル学習部５３０は、評価基準に応じた評価モデルＭ１を学習してもよい。モデル学習部５３０は、複数の評価基準それぞれについて、当該評価基準に応じた評価値を出力する評価モデルＭ１を学習してもよい。モデル学習部５３０は、評価基準に従った正解ラベルが付与された教師データに基づいて、当該評価基準に応じた評価モデルＭ１を学習してもよい。

[0196] モデル学習部５３０は、事前学習された評価モデルＭ１を追加学習してもよい。追加学習は、一例として、転移学習又はファインチューニングを含んでもよい。モデル学習部５３０は、録画データＶに正解ラベルが付与された教師データに基づいて、評価モデルＭ１を追加学習してもよい。一例として、モデル学習部５３０は、汎用的な評価基準で事前学習された評価モデルＭ１を、評価者Ｅの評価基準に従った正解ラベルが付与された教師データに基づいて追加学習してもよい。

[0197] モデル学習部５３０は、録画データから抽出される複数の特徴量を組み合わせることで、評価モデルＭ１を構築してもよい。一例として、モデル学習部５３０は、複数の特徴量を重み付き線形和により１つの評価値に変換する評価モデルＭ１を構築してもよい。モデル学習部５３０は、評価者Ｅによる特徴量の組み合わせ操作に基づいて、評価モデルＭ１を構築してもよい。重み付き線形和の場合、評価者Ｅは端末装置４０を使って重みを手動で操作することで評価モデルＭ１を構築してもよい。

[0198] （モデル評価部）

モデル評価部５４０は、モデル学習部５３０が学習した評価モデルＭ１を評価する。モデル評価部５４０は、予め用意された録画データを用いて、評価モデルＭ１による評価値の予測値を取得し、当該予測値を評価者Ｅに提示してもよい。予め用意された録画データは、評価モデルＭ１の学習に用いていない録画データでもよい。

[0199] 評価者Ｅは、モデル評価部５４０により提示された評価値の予測値に基づいて、評価モデルＭ１の性能を評価してもよい。一例として、評価者Ｅは、評価値の予測値が想定した結果であるか否かを判断してもよい。評価者Ｅは、評価モデルＭ１の評価結果を端末装置４０に入力してもよい。

[0200] モデル評価部５４０は、評価モデルＭ１の評価結果を取得してもよい。モデル評価部５４０は、評価者Ｅにより端末装置４０に入力された評価モデルＭ１の評価結果を取得してもよい。モデル評価部５４０は、評価モデルＭ１の評価結果を、評価モデルＭ１と関連付けて、モデル記憶部５０２に記憶してもよい。

[0201] 評価者Ｅは、評価モデルＭ１の評価結果を、対話者Ｓの評価に用いる評価モデルＭ１の選択に用いてもよい。評価モデルＭ１の評価結果は、複数の評価者Ｅの間で共有されてもよい。すなわち、第１の評価者Ｅが入力した評価モデルＭ１の評価結果は、第２の評価者Ｅにより参照できてもよい。

[0202] （モデル選択部）

モデル選択部５５０は、対話者Ｓの評価に用いる評価モデルＭ１を選択す

る。モデル選択部５５０は、評価モデル群Ｍ３に含まれる複数の評価モデルＭ１のうち、対話者Ｓの評価に用いる評価モデルＭ１を選択してもよい。モデル選択部５５０は、対話者Ｓの評価に用いる評価モデルＭ１を複数選択してもよい。

[0203] モデル選択部５５０は、評価者Ｅにより指定された１以上の評価モデルＭ１を選択してもよい。モデル選択部５５０は、評価者Ｅにより指定された複数の評価モデルＭ１を選択してもよい。評価者Ｅは、対話者Ｓの評価に用いるシナリオ情報に応じた評価モデルＭ１を選択してもよい。評価者Ｅは、モデル評価部５４０による評価結果に基づいて、１以上の評価モデルＭ１を選択してもよい。

[0204] 一例として、評価者Ｅは、対話者Ｓの接客スキルを評価する場合、接客用の評価基準で構築された評価モデルＭ１を選択してもよい。評価者Ｅは、対話者Ｓによる商品説明を評価する場合、商品説明用の評価基準で構築された評価モデルＭ１を選択してもよい。評価者Ｅは、対話者Ｓによる技術解説を評価する場合、技術解説用の評価基準で構築された評価モデルＭ１を選択してもよい。

[0205] モデル選択部５５０は、選択した評価モデルＭ１を制御装置２０に送信してもよい。モデル選択部５５０は、評価者Ｅによる指示に応じて、評価モデルＭ１を制御装置２０に送信してもよい。制御装置２０は、管理装置５０から受信した評価モデルＭ１を、モデル記憶部１０３に記憶してもよい。

[0206] モデル選択部５５０は、シナリオ情報と関連付けられた評価モデルＭ１を選択してもよい。モデル選択部５５０は、制御装置２０でシナリオ情報が選択された場合、当該シナリオ情報と関連付けられた評価モデルＭ１を選択し、制御装置２０に送信してもよい。

[0207] 図５に示す管理装置５０の機能構成は一例であり、用途や目的に応じて様々な機能構成例があることは言うまでもない。図５に示すデータ記憶部５０１、モデル記憶部５０２のような記憶部の区分は一例である。図５に示すラベル付与部５１０、特徴抽出部５２０、モデル学習部５３０、モデル評価部

５４０、モデル選択部５５０のような処理部の区分は一例である。

[0208] 例えば、データ記憶部５０１及びモデル記憶部５０２を１つの記憶部に統合してもよい。また、例えば、データ記憶部５０１又はモデル記憶部５０２を複数の記憶部に分割してもよい。

[0209] 例えば、ラベル付与部５１０、特徴抽出部５２０、モデル学習部５３０、モデル評価部５４０又はモデル選択部５５０の少なくとも２つを１つの処理部に統合してもよい。また、例えば、発話取得部１１０、発話生成部１２０、発話提示部１３０、対話評価部１４０又は結果出力部１５０の少なくとも１つを複数の処理部に分割してもよい。

[0210] ＜対話方法の流れ＞

対話システム１０００が実行する対話方法について、図６を参照しながら説明する。図６は、第２実施形態に係る対話方法の一例を示す処理フローである。

[0211] ステップＳ２０１において、端末装置４０は、管理装置５０から録画データＶを取得する。端末装置４０は、録画データＶを表示装置に表示する。評価者Ｅは、表示装置に表示された録画データＶを参照し、当該録画データＶに関する評価値を入力する。評価者Ｅは、複数の録画データＶを繰り返し評価してもよい。

[0212] 端末装置４０は、評価者Ｅにより入力された評価値を管理装置５０に送信する。端末装置４０は、録画データＶに関する評価値が入力される都度、当該評価値を管理装置５０に送信してもよい。端末装置４０は、複数の録画データＶに関する評価値を一括して管理装置５０に送信してもよい。

[0213] ステップＳ２０２において、管理装置５０のラベル付与部５１０は、端末装置４０から評価者Ｅにより入力された評価値を受信する。ラベル付与部５１０は、評価者Ｅにより入力された評価値に基づいて、録画データＶに正解ラベルを付与する。これにより、録画データＶに正解ラベルが付与された教師データが生成される。ラベル付与部５１０は、教師データをモデル学習部５３０に送る。

[0214] ステップＳ２０３において、管理装置５０の特徴抽出部５２０は、録画データＶから特徴ベクトルを抽出する。特徴抽出部５２０は、ステップＳ２０２で正解ラベルが付与された録画データＶについて、特徴ベクトルを抽出してもよい。特徴抽出部５２０は、特徴ベクトルをモデル学習部５３０に送る。

[0215] ステップＳ２０４において、管理装置５０のモデル学習部５３０は、ラベル付与部５１０から教師データを受け取る。また、モデル学習部５３０は、特徴抽出部５２０から特徴ベクトルを受け取る。

[0216] モデル学習部５３０は、教師データ及び特徴ベクトルに基づいて、評価モデルＭ１を学習する。モデル学習部５３０は、学習済みの評価モデルＭ１をモデル記憶部５０２に記憶する。これにより、学習済みの評価モデルＭ１が、評価モデル群Ｍ３に含まれる。

[0217] ステップＳ２０５において、評価者Ｅは、端末装置４０に対して、評価モデルＭ１を評価する操作を行う。端末装置４０は、評価者Ｅによる操作に応じて、管理装置５０に評価要求を送信する。

[0218] 管理装置５０のモデル評価部５４０は、端末装置４０からの評価要求に応じて、評価モデルＭ１を評価する。具体的には、モデル評価部５４０は、予め用意された録画データから特徴ベクトルを抽出し、評価モデルＭ１に入力する。評価モデルＭ１は、入力された特徴ベクトルに基づいて、１以上の評価値を予測し、評価値の予測値を出力する。モデル評価部５４０は、評価値の予測値を端末装置４０に送信する。端末装置４０は、管理装置５０から受信した評価値の予測値を表示装置に表示する。

[0219] 評価者Ｅは、表示装置に表示された評価値の予測値を参照し、評価モデルＭ１の性能を評価してもよい。評価者Ｅは、評価モデルＭ１の評価結果を端末装置４０に入力してもよい。端末装置４０は、評価者Ｅにより評価モデルＭ１の評価結果が入力されると、入力された評価モデルＭ１の評価結果を管理装置５０に送信する。

[0220] ステップＳ２０６において、管理装置５０のモデル評価部５４０は、端末

装置４０から評価モデルＭ１の評価結果を受信する。モデル評価部５４０は、評価モデルＭ１の評価結果を、評価モデルＭ１と関連付けて、モデル記憶部５０２に記憶する。

[0221] ステップＳ２０７において、評価者Ｅは、端末装置４０に対して、評価モデルＭ１を選択する操作を行う。端末装置４０は、評価者Ｅによる操作に応じて、管理装置５０に選択要求を送信する。

[0222] 管理装置５０のモデル選択部５５０は、端末装置４０からの選択要求に応じて、評価モデル群Ｍ３に含まれる評価モデルＭ１の一覧と、評価モデルＭ１それぞれの評価結果とを端末装置４０に送信する。端末装置４０は、管理装置５０から受信した評価モデルＭ１の一覧及び評価結果を表示装置に表示する。

[0223] 評価者Ｅは、表示装置に表示された評価モデルＭ１の一覧から、対話者Ｓの評価に用いる評価モデルＭ１を選択してもよい。端末装置４０は、評価者Ｅにより評価モデルＭ１が選択されると、評価モデルＭ１の選択結果を管理装置５０に送信する。選択結果には、評価者Ｅにより選択された評価モデルＭ１を示す情報が含まれる。

[0224] ステップＳ２０８において、管理装置５０のモデル選択部５５０は、端末装置４０から評価モデルＭ１の選択結果を受信する。モデル選択部５５０は、評価モデルＭ１の選択結果に基づいて、評価モデル群Ｍ３に含まれる評価モデルＭ１のうち、対話者Ｓの評価に用いる１以上の評価モデルＭ１を選択する。モデル選択部５５０は、選択した１以上の評価モデルＭ１を制御装置２０に送信する。

[0225] ステップＳ２０９において、制御装置２０は、管理装置５０から１以上の評価モデルＭ１を受信する。制御装置２０は、受信した１以上の評価モデルＭ１をモデル記憶部１０３に記憶する。

[0226] ステップＳ２１０において、対話者Ｓは、対話装置１０との対話を実行する。制御装置２０は、対話者Ｓと対話装置１０との対話を制御し、対話者Ｓと対話装置１０との対話を記録した対話データＤを、対話記憶部１０２に記

憶する。対話システム１０００は、ステップＳ２１０として、第１実施形態に係る対話方法のステップＳ１０１からステップＳ１１０までの処理（図３参照）を実行してもよい。

[0227] ステップＳ２１１において、制御装置２０の対話評価部１４０は、対話記憶部１０２から対話者Ｓとの対話に関する対話データＤを読み出す。また、対話評価部１４０は、モデル記憶部１０３から評価者Ｅに応じた評価モデルＭ１を読み出す。

[0228] 対話評価部１４０は、読み出した対話データＤに基づいて、評価モデル入力情報を生成する。対話評価部１４０は、評価モデル入力情報を評価モデルＭ１に入力する。評価モデルＭ１は、評価モデル入力情報に基づいて、１以上の評価値を計算する。これにより、対話者Ｓに関する評価結果が生成される。対話評価部１４０は、対話者Ｓに関する評価結果を評価結果記憶部１０４に記憶する。

[0229] ステップＳ２１２において、評価者Ｅは、端末装置４０に対して、評価結果を表示する操作を行う。一例として、評価者Ｅは、端末装置４０の表示装置に表示された評価画面において、対話者Ｓの評価結果を開く操作を行ってもよい。端末装置４０は、対話者Ｓに関する評価結果の出力要求を制御装置２０に送信する。

[0230] 制御装置２０の結果出力部１５０は、評価結果の出力要求を端末装置４０から受信する。結果出力部１５０は、対話者Ｓに関する評価結果を評価結果記憶部１０４から読み出す。結果出力部１５０は、対話者Ｓに関する評価結果を表示する画面データを生成する。結果出力部１５０は、画面データを端末装置４０に送信する。

[0231] ステップＳ２１３において、端末装置４０は、制御装置２０から画面データを受信する。端末装置４０は、画面データに基づいて、対話者Ｓに関する評価結果を表示装置に表示する。評価者Ｅは、端末装置４０の表示装置に表示された対話者Ｓに関する評価結果を参照し、対話者Ｓを評価してもよい。一例として、評価者Ｅは、対話者Ｓの評価結果を参考にして、対話者Ｓを採

用するか否かを判断してもよい。一例として、研修の講師である評価者Ｅは、対話者Ｓの評価結果を参考にして、研修の良し悪しを判断してもよい。一例として、試験の採点者である評価者Ｅは、対話者Ｓの評価結果を参考にして、対話者Ｓの得点を決定し、試験の合否を判定してもよい。

[0232] ［第２実施形態の変形例１］

制御装置２０のモデル記憶部１０３は、複数の評価基準に応じた複数の評価モデルＭ１を記憶してもよい。制御装置２０のモデル記憶部１０３は、管理装置５０のモデル記憶部５０２に記憶される評価モデル群Ｍ３に含まれるすべての評価モデルＭ１を記憶してもよい。複数の評価基準は、評価者Ｅに依存しない汎用的な評価基準を含んでもよい。複数の評価基準は、複数の汎用的な評価基準を含んでもよい。

[0233] 制御装置２０の対話評価部１４０は、１つの対話データＤに対して、複数の評価基準それぞれに応じた評価結果を生成してもよい。制御装置２０の対話評価部１４０は、モデル記憶部１０３に記憶されたすべての評価モデルＭ１を用いて、１つの対話データＤに対して、すべての評価基準に応じた評価結果を生成してもよい。

[0234] 制御装置２０が複数の評価基準に応じた評価結果を生成する場合、評価者Ｅは自らの評価基準による評価結果のみを参照できてもよい。制御装置２０が複数の評価基準に応じた評価結果を生成する場合、評価者Ｅは複数の評価基準に応じた評価結果を参照できてもよい。言い替えると、評価者Ｅは、他の評価者Ｅの評価基準による評価結果を参照できてもよい。評価者Ｅが他の評価者Ｅの評価基準による評価結果を参照できるか否かは、評価者Ｅの利用プランにより設定されてもよい。評価者Ｅが参照可能な評価基準は、評価者Ｅ又は他の評価者Ｅの利用条件又は入力情報等に基づいて選択してもよい。

[0235] 制御装置２０が記憶するシナリオ情報は、評価者Ｅと関連付けられてもよい。シナリオ情報は、評価者Ｅが生成してもよい。シナリオ情報は、当該シナリオ情報を生成した評価者Ｅと関連付けられてもよい。制御装置２０の対話評価部１４０は、評価者Ｅの評価基準に応じた評価結果を生成する場合、

当該評価者Ｅに関連付けられたシナリオ情報を用いて実行した対話に関する対話データＤのみを用いて、対話者Ｓに関する評価結果を生成してもよい。

[0236] ［応用例］

第２実施形態に係る対話システム１０００は、複数の評価者Ｅに対して、対話者Ｓに関する評価結果を提供するサービスを実現できる。本応用例では、サービス提供者が対話者Ｓと対話装置１０との対話に関する対話データＤを収集し、複数の評価者Ｅそれぞれに、評価者Ｅの評価基準に従った評価結果を提供する構成を説明する。

[0237] 本応用例に係る対話システムの全体構成について、図７を参照しながら説明する。図７は、応用例に係る対話システムの全体構成の一例を示すブロック図である。なお、本応用例では、２名の評価者Ｅ（Ｅ－１，Ｅ－２）に、対話者Ｓに関する評価結果を提供するサービスの一例を説明するが、評価者Ｅの数は限定されず、３名以上でもよい。

[0238] 図７に示されるように、対話システム１０００は、対話装置１０、制御装置２０、２台の評価装置２５（２５－１，２５－２）、生成装置３０、２台の端末装置４０（４０－１，４０－２）及び管理装置５０を含む。

[0239] 制御装置２０及び管理装置５０は、サービス提供者により設置される。一例として、制御装置２０及び管理装置５０は、サービス提供者が管理する施設に設置されてもよく、データセンタ等の遠隔地に設置されてもよい。生成装置３０は、サービス提供者により設置されてもよく、サービス提供者以外の主体により設置されてもよい。

[0240] 対話装置１０は、対話者Ｓごとに設置される。対話装置１０は、対話者Ｓにより設置されてもよく、サービス提供者により設置されてもよい。一例として、対話装置１０は、対話者Ｓが存在する空間に設置されてもよい。

[0241] 評価装置２５及び端末装置４０は、評価者Ｅごとに設置される。評価装置２５及び端末装置４０は、評価者Ｅにより設置されてもよく、サービス提供者により設置されてもよい。一例として、評価装置２５は、サービス提供者が管理する施設に設置されてもよく、データセンタ等の遠隔地に設置されて

もよく、評価者Ｅが存在する空間に設置されてもよい。端末装置４０は、評価者Ｅが存在する空間に設置されてもよい。

[0242] 複数の評価者Ｅは、評価基準が異なる主体でもよい。一例として、評価者Ｅ－１は、第１の企業又はその従業員であり、評価者Ｅ－２は、第２の企業又はその従業員であってもよい。一例として、評価者Ｅ－１は、第１の業種の企業又はその従業員であり、評価者Ｅ－２は、第２の業種の企業又はその従業員であってもよい。

[0243] 管理装置５０は、評価モデル群Ｍ３を記憶する。評価モデル群Ｍ３は、評価モデルＭ１－１及び評価モデルＭ１－２を含む。評価モデルＭ１－１は、評価者Ｅ－１により評価者Ｅ－１の評価基準に従った正解ラベルが付与された教師データに基づいて構築されている。評価モデルＭ１－２は、評価者Ｅ－２により評価者Ｅ－２の評価基準に従った正解ラベルが付与された教師データに基づいて構築されている。

[0244] なお、本応用例では、評価者Ｅごとの評価モデルＭ１が構築されている例を示したが、複数の評価者Ｅについて１つの評価モデルＭ１が構築されていてもよい。一例として、評価基準が同一の評価者Ｅについては、同一の評価モデルＭ１が構築されてもよい。一例として、評価モデルＭ１は、評価者Ｅを識別する識別情報を入力として、入力された識別情報で識別される評価者Ｅの評価基準に沿った評価値を出力するように構築されてもよい。

[0245] 管理装置５０は、評価モデルＭ１－１を評価装置２５－１に送信する。評価装置２５－１は、管理装置５０から受信した評価モデルＭ１－１を記憶する。管理装置５０は、評価モデルＭ１－２を評価装置２５－２に送信する。評価装置２５－２は、管理装置５０から受信した評価モデルＭ１－２を記憶する。

[0246] 対話装置１０は、対話者Ｓと対話する。制御装置２０は、対話者Ｓと対話装置１０との対話を制御し、対話者Ｓと対話装置１０との対話を記録した対話データＤを記憶する。制御装置２０は、対話データＤを評価装置２５－１，２５－２それぞれに送信する。評価装置２５－１，２５－２は、それぞれ

対話データＤを記憶する。

[0247] 評価装置２５－１は、評価モデルＭ１－１に基づいて、対話者Ｓに関する評価結果を生成する。評価装置２５－１は、対話者Ｓに関する評価結果を端末装置４０－１に送信する。端末装置４０－１は、対話者Ｓに関する評価結果を評価者Ｅ－１に提示する。評価者Ｅ－１は、評価者Ｅ－１の評価基準に従った評価結果を取得できる。

[0248] 評価装置２５－２は、評価モデルＭ１－２に基づいて、対話者Ｓに関する評価結果を生成する。評価装置２５－２は、対話者Ｓに関する評価結果を端末装置４０－２に送信する。端末装置４０－２は、対話者Ｓに関する評価結果を評価者Ｅ－２に提示する。評価者Ｅ－２は、評価者Ｅ－２の評価基準に従った評価結果を取得できる。

[0249] 評価者Ｅ－１，Ｅ－２は、予め作成された評価基準のテンプレートをカスタマイズすることで、それぞれの評価基準を作成してもよい。評価基準のテンプレートは、サービス提供者等により提供されてもよく、他の評価者Ｅにより提供されてもよい。評価者Ｅ－１，Ｅ－２は、予め作成された複数の評価基準から、対話者Ｓの評価に用いる評価基準を選択してもよい。サービス提供者は、評価者Ｅ－１，Ｅ－２の属性又は対話者Ｓの属性に基づいて、評価基準を提案してもよい。一例として、評価者Ｅ－１，Ｅ－２の属性は業種でもよい。一例として、対話者Ｓの属性は職種でもよい。

[0250] 制御装置２０は、評価者Ｅ－１，Ｅ－２の評価基準に基づいて、シナリオ情報を切り替えてもよい。制御装置２０は、評価者Ｅ－１，Ｅ－２の属性又は対話者Ｓの属性等に応じて、シナリオ情報を切り替えてもよい。

[0251] 本応用例に係る対話システム１０００を構成するために、制御装置２０は、第１実施形態に係る制御装置２０が備えるシナリオ記憶部１０１、対話記憶部１０２、発話取得部１１０、発話生成部１２０、発話提示部１３０を備えてもよい。また、評価装置２５は、第１実施形態に係る制御装置２０が備える対話記憶部１０２、モデル記憶部１０３、評価結果記憶部１０４、対話評価部１４０及び結果出力部１５０を備えてもよい。

[0252] ［その他の実施形態］

上記の各実施形態、変形例又は応用例では、対話者Ｓと対話装置１０との間で、音声情報及び画像情報を用いた対話を実行する例を説明した。上記の各実施形態、変形例又は応用例は、音声情報及び画像情報を用いた対話に限定されるものではなく、一例として、テキスト情報を用いた対話に適用してもよい。テキスト情報を用いた対話は、一例として、テキストチャット、電子メール、メッセージングサービス等を含んでもよい。

[0253] ＜まとめ＞

以上の説明から明らかなように、本開示の一実施形態に係る対話システム１０００は、利用者が発話した第１発話に関する情報を取得し、第１発話に関する情報と入力情報と機械学習モデルとに基づいて、第２発話に関する情報を生成し、第２発話に関する情報に基づいて、対話装置の発話を制御する。入力情報は、利用者と対話装置との関係を示す情報を含む。

[0254] 利用者と対話装置との関係は、利用者が対話装置に役務若しくは商品を提供する関係、利用者が対話装置を指導する関係、対話装置が利用者に報酬を支払う関係、又は、情報Ａを有する利用者が情報Ａを有しない対話装置に情報Ａを提供する関係の少なくとも１つを含んでもよい。利用者と対話装置との関係は、店員と顧客との関係、上司と部下との関係、教師と生徒との関係、講演者と聴衆との関係、医者と患者との関係、又は被依頼者と依頼者との関係の少なくとも１つを含んでもよい。

[0255] 利用者と対話装置との関係を示す情報は、少なくとも対話装置の役割に関する情報を含んでもよい。対話装置の役割は、顧客、部下、生徒、聴衆、患者又は依頼者の少なくとも１つを含んでもよい。利用者と対話装置との関係を示す情報は、利用者の役割と対話装置の役割に関する情報を含んでもよい。

[0256] 対話システム１０００は、第１発話に関する情報の少なくとも一部に基づいて、利用者に関する評価結果を生成してもよい。対話システム１０００は、第１発話に関する情報及び第２発話に関する情報を含む対話データに基づ

いて、第２の機械学習モデルへの第２の入力情報を生成し、第２の入力情報を第２の機械学習モデルへ入力することで、評価結果を生成してもよい。

[0257] 第２の入力情報は、対話データから抽出された１以上の特徴量を含んでもよい。特徴量は、利用者の表情、利用者の身体動作、利用者の声、利用者と対話装置との対話の円滑度、又は対話の内容の少なくとも１つに関する情報を含んでもよい。

[0258] 対話システム１０００は、対話データに対する評価者による評価値を取得し、対話データに評価値を付与した教師データに基づいて、第２の機械学習モデルを生成してもよい。対話システム１０００は、複数の評価基準それぞれの評価値を取得し、評価基準ごとの第２の機械学習モデルを生成してもよい。

[0259] 対話システム１０００は、評価基準ごとの第２の機械学習モデルから評価結果の生成に用いる第２の機械学習モデルを選択してもよい。対話システム１０００は、評価者に対応する評価基準の第２の機械学習モデルを選択してもよい。

[0260] 対話システム１０００は、１以上の第２の機械学習モデルに基づいて、１つの対話データに対する複数の評価結果を生成してもよい。対話システム１０００は、利用者に関する評価結果及び利用者の評価に用いた情報を出力してもよい。対話システム１０００は、第１発話に関する情報の少なくとも一部を、利用者又は他の利用者に提示してもよい。

[0261] 対話システム１０００は、予め設定されたシナリオ情報に基づいて、入力情報を生成してもよい。対話システム１０００は、所定の基準に基づいてシナリオ情報を設定してもよい。対話システム１０００は、評価者の指示に基づいてシナリオ情報を設定してもよい。

[0262] 対話システム１０００は、利用者が発話した発話に関する情報の取得と、利用者の発話に関する情報と入力情報と機械学習モデルとに基づく対話装置の発話に関する情報の生成と、対話装置の発話に関する情報に基づく対話装置の発話の制御と、を所定の条件を満たすまで繰り返し実行してもよい。対

話システム１０００は、利用者の複数の発話に関する情報と対話装置の複数の発話に関する情報とを含む対話データに基づいて、利用者の評価結果を生成してもよい。対話システム１０００は、対話データと第２の機械学習モデルとに基づいて、評価結果を生成してもよい。

[0263] これにより、本開示の一実施形態によれば、機械学習モデルに基づいて対話装置の発話に関する情報を生成するため、利用者と対話する装置を制御できる。一の側面では、一実施形態によれば、利用者と対話装置との対話を通じて、利用者を評価するための情報を取得できる。一の側面では、一実施形態によれば、機械学習モデルに基づいて利用者による発話に応答するため、利用者は対話装置と自然な流れで対話でき、評価者は利用者を適切に評価できる。他の側面では、一実施形態によれば、評価者が定めた評価基準に応じた適切な評価結果を取得できる。

[0264] ［情報処理装置のハードウェア構成］

前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）の一部又は全部は、ハードウェアで構成されていてもよいし、ＣＰＵ（Central Processing Unit）、ＧＰＵ（Graphics Processing Unit）等が実行するソフトウェア（プログラム）の情報処理で構成されてもよい。ソフトウェアの情報処理で構成される場合には、前述した実施形態における各装置の少なくとも一部の機能を実現するソフトウェアを、ＣＤ－ＲＯＭ（Compact Disc-Read Only Memory）、ＵＳＢ（Universal Serial Bus）メモリ等の非一時的な記憶媒体（非一時的なコンピュータ可読媒体）に収納し、コンピュータに読み込ませることにより、ソフトウェアの情報処理を実行してもよい。また、通信ネットワークを介して当該ソフトウェアがダウンロードされてもよい。さらに、ソフトウェアの処理の全部又は一部がＡＳＩＣ（Application Specific Integrated Circuit）、ＦＰＧＡ（Field Programmable Gate Array）等の回路に実装されることにより、当該ソフトウェアによる情報処理がハードウェアにより実行されてもよい。

[0265] ソフトウェアを収納する記憶媒体は、光ディスク等の着脱可能なものでもよいし、ハードディスク、メモリ等の固定型の記憶媒体であってもよい。また、記憶媒体は、コンピュータ内部に備えられてもよいし（主記憶装置、補助記憶装置等）、コンピュータ外部に備えられてもよい。

[0266] 図８は、前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）のハードウェア構成の一例を示すブロック図である。各装置は、一例として、プロセッサ７１と、主記憶装置７２（メモリ）と、補助記憶装置７３（メモリ）と、ネットワークインタフェース７４と、デバイスインタフェース７５と、を備え、これらがバス７６を介して接続されたコンピュータ７として実現されてもよい。

[0267] 図８のコンピュータ７は、各構成要素を一つ備えているが、同じ構成要素を複数備えていてもよい。また、図８では、１台のコンピュータ７が示されているが、ソフトウェアが複数台のコンピュータにインストールされて、当該複数台のコンピュータそれぞれがソフトウェアの同一の又は異なる一部の処理を実行してもよい。この場合、コンピュータそれぞれがネットワークインタフェース７４等を介して通信して処理を実行する分散コンピューティングの形態であってもよい。つまり、前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）は、１又は複数の記憶装置に記憶された命令を１台又は複数台のコンピュータが実行することで機能を実現するシステムとして構成されてもよい。また、端末から送信された情報をクラウド上に設けられた１台又は複数台のコンピュータで処理し、この処理結果を端末に送信するような構成であってもよい。

[0268] 前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）の各種演算は、１又は複数のプロセッサを用いて、又はネットワークを介した複数台のコンピュータを用いて、並列処理で実行されてもよい。また、各種演算が、プロセ

ッサ内に複数ある演算コアに振り分けられて、並列処理で実行されてもよい。また、本開示の処理、手段等の一部又は全部は、ネットワークを介してコンピュータ７と通信可能なクラウド上に設けられたプロセッサ及び記憶装置の少なくとも一方により実現されてもよい。このように、前述した実施形態における各装置は、１台又は複数台のコンピュータによる並列コンピューティングの形態であってもよい。

[0269] プロセッサ７１は、少なくともコンピュータの制御又は演算のいずれかを行う電子回路（処理回路、Processing circuit、Processing circuitry、ＣＰＵ、ＧＰＵ、ＦＰＧＡ、ＡＳＩＣ等）であってもよい。また、プロセッサ７１は、汎用プロセッサ、特定の演算を実行するために設計された専用の処理回路又は汎用プロセッサと専用の処理回路との両方を含む半導体装置のいずれであってもよい。また、プロセッサ７１は、光回路を含むものであってもよいし、量子コンピューティングに基づく演算機能を含むものであってもよい。

[0270] プロセッサ７１は、コンピュータ７の内部構成の各装置等から入力されたデータやソフトウェアに基づいて演算処理を行ってもよく、演算結果や制御信号を各装置等に出力してもよい。プロセッサ７１は、コンピュータ７のＯＳ（Operating System）や、アプリケーション等を実行することにより、コンピュータ７を構成する各構成要素を制御してもよい。

[0271] 前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）は、１又は複数のプロセッサ７１により実現されてもよい。ここで、プロセッサ７１は、１チップ上に配置された１又は複数の電子回路を指してもよいし、２つ以上のチップあるいは２つ以上のデバイス上に配置された１又は複数の電子回路を指してもよい。複数の電子回路を用いる場合、各電子回路は有線又は無線により通信してもよい。

[0272] 主記憶装置７２は、プロセッサ７１が実行する命令及び各種データ等を記憶してもよく、主記憶装置７２に記憶された情報がプロセッサ７１により読

み出されてもよい。補助記憶装置７３は、主記憶装置７２以外の記憶装置である。なお、これらの記憶装置は、電子情報を格納可能な任意の電子部品を意味するものとし、半導体のメモリでもよい。半導体のメモリは、揮発性メモリ又は不揮発性メモリのいずれでもよい。前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）において各種データ等を保存するための記憶装置は、主記憶装置７２又は補助記憶装置７３により実現されてもよく、プロセッサ７１に内蔵される内蔵メモリにより実現されてもよい。例えば、前述した実施形態における各記憶部は、主記憶装置７２又は補助記憶装置７３により実現されてもよい。

[0273] 前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置４０及び管理装置５０）が、少なくとも１つの記憶装置（メモリ）と、この少なくとも１つの記憶装置に接続（結合）される少なくとも１つのプロセッサで構成される場合、記憶装置１つに対して、少なくとも１つのプロセッサが接続されてもよい。また、１つのプロセッサに対して、少なくとも１つの記憶装置が接続されてもよい。また、複数のプロセッサのうち少なくとも１つのプロセッサが、複数の記憶装置のうち少なくとも１つの記憶装置に接続される構成を含んでもよい。また、複数台のコンピュータに含まれる記憶装置とプロセッサによって、この構成が実現されてもよい。さらに、記憶装置がプロセッサと一体になっている構成（例えば、Ｌ１キャッシュ、Ｌ２キャッシュを含むキャッシュメモリ）を含んでもよい。

[0274] ネットワークインタフェース７４は、無線又は有線により、通信ネットワーク８に接続するためのインタフェースである。ネットワークインタフェース７４は、既存の通信規格に適合したもの等、適切なインタフェースを用いればよい。ネットワークインタフェース７４により、通信ネットワーク８を介して接続された外部装置９Ａと情報のやり取りが行われてもよい。なお、通信ネットワーク８は、ＷＡＮ（Wide Area Network）、ＬＡＮ（Local Area

Network）、ＰＡＮ（Personal Area Network）等の何れか又はそれらの組み合わせであってよく、コンピュータ７と外部装置９Ａとの間で情報のやり取りが行われるものであればよい。ＷＡＮの一例としてインターネット等があり、ＬＡＮの一例としてＩＥＥＥ８０２．１１やイーサネット（登録商標）等があり、ＰＡＮの一例としてＢｌｕｅｔｏｏｔｈ（登録商標）やＮＦＣ（Near Field Communication）等がある。

[0275] デバイスインタフェース７５は、外部装置９Ｂと直接接続するＵＳＢ等のインタフェースである。

[0276] 外部装置９Ａはコンピュータ７とネットワークを介して接続されている装置である。外部装置９Ｂはコンピュータ７と直接接続されている装置である。

[0277] 外部装置９Ａ又は外部装置９Ｂは、一例として、入力装置であってもよい。入力装置は、例えば、カメラ、マイクロフォン、モーションキャプチャ、各種センサ、キーボード、マウス、タッチパネル等のデバイスであり、取得した情報をコンピュータ７に与える。また、パーソナルコンピュータ、タブレット端末、スマートフォン等の入力部とメモリとプロセッサを備えるデバイスであってもよい。

[0278] また、外部装置９Ａ又は外部装置９Ｂは、一例として、出力装置でもよい。出力装置は、例えば、ＬＣＤ（Liquid Crystal Display）、有機ＥＬ（Electro Luminescence）パネル等の表示装置であってもよいし、音声等を出力するスピーカ等であってもよい。また、パーソナルコンピュータ、タブレット端末又はスマートフォン等の出力部とメモリとプロセッサを備えるデバイスであってもよい。

[0279] また、外部装置９Ａまた外部装置９Ｂは、記憶装置（メモリ）であってもよい。例えば、外部装置９Ａはネットワークストレージ等であってもよく、外部装置９ＢはＨＤＤ等のストレージであってもよい。

[0280] また、外部装置９Ａ又は外部装置９Ｂは、前述した実施形態における各装置（対話装置１０、制御装置２０、評価装置２５、生成装置３０、端末装置

４０及び管理装置５０）の構成要素の一部の機能を有する装置でもよい。つまり、コンピュータ７は、外部装置９Ａ又は外部装置９Ｂに処理結果の一部又は全部を送信してもよいし、外部装置９Ａ又は外部装置９Ｂから処理結果の一部又は全部を受信してもよい。

[0281] 本明細書（請求項を含む）において、「a、b及びcの少なくとも1つ（一方）」又は「a、b又はcの少なくとも1つ（一方）」の表現（同様な表現を含む）が用いられる場合は、a、b、c、a-b、a-c、b-c又はa-b-cのいずれかを含む。また、a-a、a-b-b、a-a-b-b-c-c等のように、いずれかの要素について複数のインスタンスを含んでもよい。さらに、a-b-c-dのようにdを有する等、列挙された要素（a、b及びc）以外の他の要素を加えることも含む。

[0282] 本明細書（請求項を含む）において、「データを入力として／を用いて／データに基づいて／に従って／に応じて」等の表現（同様な表現を含む）が用いられる場合は、特に断りがない場合、データそのものを用いる場合や、データに何らかの処理を行ったもの（例えば、ノイズ加算したもの、正規化したもの、データから抽出した特徴量、データの中間表現等）を用いる場合を含む。また、「データを入力として／を用いて／データに基づいて／に従って／に応じて」何らかの結果が得られる旨が記載されている場合（同様な表現を含む）、特に断りがない場合、当該データのみに基づいて当該結果が得られる場合や、当該データ以外の他のデータ、要因、条件及び／又は状態にも影響を受けて当該結果が得られる場合を含む。また、「データを出力する」旨が記載されている場合（同様な表現を含む）、特に断りがない場合、データそのものを出力として用いる場合や、データに何らかの処理を行ったもの（例えば、ノイズ加算したもの、正規化したもの、データから抽出した特徴量、各種データの中間表現等）を出力として用いる場合を含む。

[0283] 本明細書（請求項を含む）において、「接続される（connected）」及び「結合される（coupled）」との用語が用いられる場合は、直接的な接続／結合、間接的な接続／結合、電気的（electrically）な接続／結合、通信的（communicatively）な接続／結合、機能的（operatively）な接続／結合、物理的

（physically）な接続／結合等のいずれをも含む非限定的な用語として意図される。当該用語は、当該用語が用いられた文脈に応じて適宜解釈されるべきであるが、意図的に或いは当然に排除されるのではない接続／結合形態は、当該用語に含まれるものして非限定的に解釈されるべきである。

[0284] 本明細書（請求項を含む）において、「ＡがＢするよう構成される（A configured to B）」との表現が用いられる場合は、要素Ａの物理的構造が、動作Ｂを実行可能な構成を有するとともに、要素Ａの恒常的（permanent）又は一時的（temporary）な設定（setting/configuration）が、動作Ｂを実際に実行するように設定（configured/set）されていることを含んでよい。例えば、要素Ａが汎用プロセッサである場合、当該プロセッサが動作Ｂを実行可能なハードウェア構成を有するとともに、恒常的（permanent）又は一時的（temporary）なプログラム（命令）の設定により、動作Ｂを実際に実行するように設定（configured）されていればよい。また、要素Ａが専用プロセッサ、専用演算回路等である場合、制御用命令及びデータが実際に付属しているか否かとは無関係に、当該プロセッサの回路的構造等が動作Ｂを実際に実行するように構築（implemented）されていればよい。

[0285] 本明細書（請求項を含む）において、含有又は所有を意味する用語（例えば、「含む（comprising/including）」、「有する（having）」等）が用いられる場合は、当該用語の目的語により示される対象物以外の物を含有又は所有する場合を含む、open-endedな用語として意図される。これらの含有又は所有を意味する用語の目的語が数量を指定しない又は単数を示唆する表現（a又はanを冠詞とする表現）である場合は、当該表現は特定の数に限定されないものとして解釈されるべきである。

[0286] 本明細書（請求項を含む）において、ある箇所において「１つ又は複数（one or more）」、「少なくとも１つ（at least one）」等の表現が用いられ、他の箇所において数量を指定しない又は単数を示唆する表現（a又はanを冠詞とする表現）が用いられているとしても、後者の表現が「１つ」を意味することを意図しない。一般に、数量を指定しない又は単数を示唆する表現（a

又はanを冠詞とする表現）は、必ずしも特定の数に限定されないものとして解釈されるべきである。

[0287] 本明細書において、ある実施形態の有する特定の構成について特定の効果（advantage/result）が得られる旨が記載されている場合、別段の理由がない限り、当該構成を有する他の１つ又は複数の実施形態についても当該効果が得られると理解されるべきである。但し、当該効果の有無は、一般に種々の要因、条件及び／又は状態に依存し、当該構成により必ず当該効果が得られるものではないと理解されるべきである。当該効果は、種々の要因、条件及び／又は状態が満たされたときに実施形態に記載の当該構成により得られるものに過ぎず、当該構成又は類似の構成を規定したクレームに係る発明において、当該効果が必ずしも得られるものではない。

[0288] 本明細書（請求項を含む）において、複数のハードウェアが所定の処理を行う場合、各ハードウェアが協働して所定の処理を行ってもよいし、一部のハードウェアが所定の処理の全てを行ってもよい。また、一部のハードウェアが所定の処理の一部を行い、別のハードウェアが所定の処理の残りを行ってもよい。本明細書（請求項を含む）において、「１又は複数のハードウェアが第１の処理を行い、前記１又は複数のハードウェアが第２の処理を行う」等の表現（同様な表現を含む）が用いられている場合、第１の処理を行うハードウェアと第２の処理を行うハードウェアは同じものであってもよいし、異なるものであってもよい。つまり、第１の処理を行うハードウェア及び第２の処理を行うハードウェアが、前記１又は複数のハードウェアに含まれていればよい。なお、ハードウェアは、電子回路、電子回路を含む装置等を含んでよい。

[0289] 本明細書（請求項を含む）において、複数の記憶装置（メモリ）がデータの記憶を行う場合、複数の記憶装置のうち個々の記憶装置は、データの一部のみを記憶してもよいし、データの全体を記憶してもよい。また、複数の記憶装置のうち一部の記憶装置がデータを記憶する構成を含んでもよい。

[0290] 本明細書（請求項を含む）において、「第１の」、「第２の」等の用語は

、単に２つ以上の要素間を区別する方法として使用されており、その対象に対して時間的態様、空間的態様、順序、量等の技術的意味を課すことを必ずしも意図するものではない。したがって、例えば、第１の要素及び第２の要素への参照は、２つの要素のみがそこで採用され得ること、第１の要素が第２の要素に先行しなければならないこと、第２の要素が存在するために第１の要素が存在しなければならないこと等を必ずしも意味するものではない。

[0291] 以上、本開示の実施形態について詳述したが、本開示は上記した個々の実施形態に限定されるものではない。特許請求の範囲に規定された内容及びその均等物から導き出される本発明の概念的な思想と趣旨を逸脱しない範囲において、種々の追加、変更、置き換え、部分的削除等が可能である。例えば、前述した実施形態において、数値又は数式を説明に用いている場合、これらは例示的な目的で示されたものであり、本開示の範囲を限定するものではない。また、実施形態で示した各動作の順序も例示的なものであり、本開示の範囲を限定するものではない。

[0292] なお、開示の技術では、以下に記載する付記のような形態が考えられる。

[0293] （付記１）

少なくとも１つのメモリと、

少なくとも１つのプロセッサと、を備え、

前記少なくとも１つのプロセッサは、

利用者が発話した第１発話に関する情報を取得し、

前記第１発話に関する情報と入力情報と機械学習モデルとに基づいて、第２発話に関する情報を生成し、

前記第２発話に関する情報に基づいて、対話装置の発話を制御し、

前記入力情報は、前記利用者と前記対話装置との関係を示す情報を含む、

対話システム。

[0294] （付記２）

前記関係は、前記利用者が前記対話装置に役務若しくは商品を提供する関係、前記利用者が前記対話装置を指導する関係、前記対話装置が前記利用者

に報酬を支払う関係、又は、情報Ａを有する前記利用者が前記情報Ａを有しない前記対話装置に情報Ａを提供する関係の少なくとも１つを含む、

付記１に記載の対話システム。

[0295] （付記３）

前記関係は、店員と顧客との関係、上司と部下との関係、教師と生徒との関係、講演者と聴衆との関係、医者と患者との関係、又は被依頼者と依頼者との関係の少なくとも１つを含む、

付記２に記載の対話システム。

[0296] （付記４）

前記利用者と前記対話装置との関係を示す情報は、少なくとも前記対話装置の役割に関する情報を含む、

付記１乃至３のいずれかに記載の対話システム。

[0297] （付記５）

前記対話装置の役割は、顧客、部下、生徒、聴衆、患者又は依頼者の少なくとも１つを含む、

付記４に記載の対話システム。

[0298] （付記６）

前記少なくとも１つのプロセッサは、

前記第１発話に関する情報の少なくとも一部に基づいて、前記利用者に関する評価結果を生成する、

付記１乃至５のいずれかに記載の対話システム。

[0299] （付記７）

前記少なくとも１つのプロセッサは、

前記第１発話に関する情報及び前記第２発話に関する情報を含む対話データに基づいて、第２の機械学習モデルへの第２の入力情報を生成し、

前記第２の入力情報を前記第２の機械学習モデルへ入力することで、前記評価結果を生成する、

付記６に記載の対話システム。

[0300] （付記８）

前記第２の入力情報は、前記対話データから抽出された１以上の特徴量を含む、

付記７に記載の対話システム。

[0301] （付記９）

前記特徴量は、前記利用者の表情、前記利用者の身体動作、前記利用者の声、前記利用者と前記対話装置との対話の円滑度、又は前記対話の内容の少なくとも１つに関する情報を含む、

付記８に記載の対話システム。

[0302] （付記１０）

前記少なくとも１つのプロセッサは、

前記対話データに対する評価者による評価値を取得し、

前記対話データに前記評価値を付与した教師データに基づいて、前記第２の機械学習モデルを生成する、

付記６乃至９のいずれかに記載の対話システム。

[0303] （付記１１）

前記少なくとも１つのプロセッサは、

複数の評価基準それぞれの前記評価値を取得し、

前記評価基準ごとの前記第２の機械学習モデルを生成する、

付記１０に記載の対話システム。

[0304] （付記１２）

前記少なくとも１つのプロセッサは、

評価基準ごとの前記第２の機械学習モデルから前記評価結果の生成に用いる前記第２の機械学習モデルを選択する、

付記７乃至１１のいずれかに記載の対話システム。

[0305] （付記１３）

前記少なくとも１つのプロセッサは、

評価者に対応する前記評価基準の前記第２の機械学習モデルを選択する

、

付記１２に記載の対話システム。

[0306] （付記１４）

前記少なくとも１つのプロセッサは、

１以上の前記第２の機械学習モデルに基づいて、１つの前記対話データに対する複数の前記評価結果を生成する、

付記７乃至１３のいずれかに記載の対話システム。

[0307] （付記１５）

前記少なくとも１つのプロセッサは、

前記評価結果及び前記利用者の評価に用いた情報を出力する、

付記６乃至１４のいずれかに記載の対話システム。

[0308] （付記１６）

前記少なくとも１つのプロセッサは、

前記第１発話に関する情報の少なくとも一部を、前記利用者又は他の利用者に提示する、

付記１乃至１５のいずれかに記載の対話システム。

[0309] （付記１７）

前記少なくとも１つのプロセッサは、

予め設定されたシナリオ情報に基づいて、前記入力情報を生成する、

付記１乃至１６のいずれかに記載の対話システム。

[0310] （付記１８）

前記少なくとも１つのプロセッサは、

所定の基準に基づいて前記シナリオ情報を設定する、

付記１７に記載の対話システム。

[0311] （付記１９）

前記少なくとも１つのプロセッサは、

評価者の指示に基づいて前記シナリオ情報を設定する、

付記１８に記載の対話システム。

[0312] （付記２０）

前記少なくとも１つのプロセッサは、

前記利用者が発話した発話に関する情報の取得と、

前記利用者の前記発話に関する情報と前記入力情報と前記機械学習モデルとに基づく前記対話装置の発話に関する情報の生成と、

前記対話装置の前記発話に関する情報に基づく前記対話装置の発話の制御と、

を所定の条件を満たすまで繰り返し実行する、

付記１乃至１９のいずれかに記載の対話システム。

[0313] （付記２１）

前記少なくとも１つのプロセッサは、

前記利用者の複数の前記発話に関する情報と前記対話装置の複数の前記発話に関する情報とを含む対話データに基づいて、前記利用者の評価結果を生成する、

付記２０に記載の対話システム。

[0314] （付記２２）

前記少なくとも１つのプロセッサは、

前記対話データと第２の機械学習モデルとに基づいて、前記評価結果を生成する、

付記２１に記載の対話システム。

[0315] （付記２３）

少なくとも１つのプロセッサが、

利用者が発話した第１発話に関する情報を取得し、

前記第１発話に関する情報と入力情報と機械学習モデルとに基づいて、第２発話に関する情報を生成し、

前記第２発話に関する情報に基づいて、対話装置の発話を制御し、

前記入力情報は、前記利用者と前記対話装置との関係を示す情報を含む、

対話方法。

[0316] （付記２４）

少なくとも１つのプロセッサに、

利用者が発話した第１発話に関する情報を取得し、

前記第１発話に関する情報と入力情報と機械学習モデルとに基づいて、第２発話に関する情報を生成し、

前記第２発話に関する情報に基づいて、対話装置の発話を制御し、

前記入力情報は、前記利用者と前記対話装置との関係を示す情報を含む、

処理を実行させるためのプログラム。

[0317] 本願は、米国特許商標庁に２０２４年３月１３日に出願された仮出願６３／５６４，５３７号、米国特許商標庁に２０２４年５月１６日に出願された仮出願６３／６４８，３４１号、及び、日本国特許庁に２０２４年７月２９日に出願された日本国特許出願２０２４－１２２８９８号の優先権を主張するものであり、その全内容を参照することにより本願に援用する。

符号の説明

[0318] １０：対話装置

２０：制御装置

２５：評価装置

３０：生成装置

４０：端末装置

５０：管理装置

１０１：シナリオ記憶部

１０２：対話記憶部

１０３：モデル記憶部

１０４：評価結果記憶部

１１０：発話取得部

１２０：発話生成部

１３０：発話提示部

１４０：対話評価部

１５０：結果出力部

５０１：データ記憶部

５０２：モデル記憶部

５１０：ラベル付与部

５２０：特徴抽出部

５３０：モデル学習部

５４０：モデル評価部

５５０：モデル選択部

１０００：対話システム

Ｄ：対話データ

Ｍ１：評価モデル

Ｍ２：生成モデル

請求の範囲

[請求項1] 少なくとも１つのメモリと、

少なくとも１つのプロセッサと、を備え、

前記少なくとも１つのプロセッサは、

利用者の発話に関する情報の取得と、前記利用者の発話に関する情報と機械学習モデルとに基づく対話装置の発話に関する情報の生成と、前記対話装置の発話に関する情報に基づく前記対話装置の発話の制御と、を所定の条件を満たすまで繰り返し実行し、

前記対話装置の発話に関する情報の生成は、前記利用者と前記対話装置との関係を示す情報に基づいて実行され、

前記関係は、前記利用者が前記対話装置に役務若しくは商品を提供する関係、前記利用者が前記対話装置を指導する関係、前記対話装置が前記利用者に報酬を支払う関係、又は、情報Ａを有する前記利用者が前記情報Ａを有しない前記対話装置に前記情報Ａを提供する関係の少なくとも１つを含む、

情報処理システム。

[請求項2] 前記関係は、前記利用者が店員で前記対話装置が顧客の関係、前記利用者が上司で前記対話装置が部下の関係、前記利用者が教師で前記対話装置が生徒の関係、前記利用者が講演者で前記対話装置が聴衆の関係、前記利用者が医者で前記対話装置が患者の関係、又は前記利用者が被依頼者で前記対話装置が依頼者の関係の少なくとも１つを含む、

請求項１に記載の情報処理システム。

[請求項3] 前記利用者と前記対話装置との関係を示す前記情報は、少なくとも前記対話装置の役割に関する情報を含む、

請求項１に記載の情報処理システム。

[請求項4] 前記対話装置の役割は、顧客、部下、生徒、聴衆、患者又は依頼者の少なくとも１つを含む、

請求項３に記載の情報処理システム。

[請求項5] 前記機械学習モデルへの入力情報は、システムプロンプトを含み、

前記システムプロンプトは、

前記利用者と前記対話装置との関係を示す前記情報と、

少なくとも、対話の状況を示す情報、対話の前提条件、又は、発話に対する制約条件のいずれか１つと、を含む、

請求項１乃至請求項４のいずれか一項に記載の情報処理システム。

[請求項6] 前記システムプロンプトは、少なくとも前記発話に対する制約条件を含み、

前記発話に対する制約条件は、前記対話装置の発話の文字数制限を含む、

請求項５に記載の情報処理システム。

[請求項7] 前記少なくとも１つのプロセッサは、対話の進行度合いに基づいて、前記対話装置の発話に関する情報の生成に使用する前記システムプロンプトを変更する、

請求項５又は請求項６に記載の情報処理システム。

[請求項8] 前記対話の進行度合いは、少なくとも、対話のターン数、又は、対話が開始してからの経過時間のいずれか１つを含む、

請求項７に記載の情報処理システム。

[請求項9] 前記繰り返し実行の終了後、前記少なくとも１つのメモリに記憶された、少なくとも前記利用者の発話に関する情報を含む対話データを用いて、前記利用者に関する１以上の評価結果を生成する、

請求項１乃至請求項８のいずれか一項に記載の情報処理システム。

[請求項10] 前記少なくとも１つのプロセッサは、

１以上の第２の機械学習モデルに基づいて、前記１以上の評価結果を生成する、

請求項９に記載の情報処理システム。

[請求項11] 前記少なくとも１つのプロセッサは、

前記対話データの少なくとも一部に基づいて、前記１以上の第２の機械学習モデルそれぞれへの入力情報を生成し、

前記入力情報を前記１以上の第２の機械学習モデルそれぞれへ入力することで、前記１以上の評価結果を生成する、

請求項１０に記載の情報処理システム。

[請求項12] 前記入力情報は、前記対話データの少なくとも一部から抽出された１以上の特徴量を含み、

前記１以上の特徴量は、少なくとも、前記利用者の表情、前記利用者の身体動作、前記利用者の声、前記利用者と前記対話装置との対話の円滑度、又は、前記対話の内容のいずれか１つに関する情報を含む、

請求項１１に記載の情報処理システム。

[請求項13] 前記入力情報は、前記対話データの少なくとも一部から抽出された１以上の特徴量を含み、

前記１以上の特徴量は、少なくとも、前記対話データに含まる前記利用者の音声情報から抽出された特徴量、前記対話データに含まる前記利用者の画像情報から抽出された特徴量、又は、前記対話データに含まる対話内容から抽出された特徴量のいずれか１つを含む、

請求項１１に記載の情報処理システム。

[請求項14] 前記少なくとも１つのプロセッサは、

他の対話データと前記他の対話データに対する評価者による評価値とを含む教師データを用いて、前記第２の機械学習モデルを生成し、

生成した前記第２の機械学習モデルを用いて、前記評価結果の生成する、

請求項１０乃至請求項１３のいずれか一項に記載の情報処理システム。

[請求項15] 前記第２の機械学習モデルは、前記機械学習モデルと同じモデル、

又は、前記機械学習モデルに基づいて生成されたモデルである、

請求項１０乃至請求項１４のいずれか一項に記載の情報処理システム。

[請求項16] 前記少なくとも１つのプロセッサは、

所定の基準に基づいて設定されたシナリオ情報に基づいて、前記利用者と前記対話装置との関係を示す前記情報を取得する、

請求項１乃至請求項１５のいずれか一項に記載の情報処理システム。

[請求項17] 前記少なくとも１つのプロセッサは、

少なくとも業種又は職種のいずれか１つに基づいて、前記シナリオ情報を設定する、

請求項１６に記載の情報処理システム。

[請求項18] 前記少なくとも１つのプロセッサは、

前記利用者の属性に基づいて前記シナリオ情報を設定する、

請求項１７に記載の情報処理システム。

[請求項19] 前記利用者の属性は、少なくとも前記利用者の履歴書又は職務経歴書のいずれか１つを含む、

請求項１８に記載の情報処理システム。

[請求項20] 前記シナリオ情報は、ガイダンス情報を含み、

前記少なくとも１つのプロセッサは、前記繰り返し処理の開始前に、前記ガイダンス情報を前記対話装置に出力させる、

請求項１６乃至請求項１９のいずれか一項に記載の情報処理システム。

[請求項21] 前記少なくとも１つのプロセッサは、

設定された前記シナリオ情報に基づいて、前記１以上の評価結果の生成に用いる１以上の第２の機械学習モデルを選択し、

選択された前記１以上の第２の機械学習モデルを用いて、前記１以上の評価結果を生成する、

請求項１６乃至請求項１９のうち請求項１０を引用するいずれか一項に記載の情報処理システム。

[請求項22] 請求項１乃至請求項２１のいずれか一項に記載の情報処理システムを用いて、

前記利用者の発話に関する情報の取得と、前記利用者の発話に関する情報と前記機械学習モデルとに基づく前記対話装置の発話に関する情報の生成と、前記対話装置の発話に関する情報に基づく前記対話装置の発話の制御と、を所定の条件を満たすまで繰り返し実行する、

情報処理方法。

[請求項23] 少なくとも１つのプロセッサに、請求項２２に記載の情報処理方法を実行させる、

プログラム。