(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6465077号
(P6465077)

(45)発行日　平成31年2月6日(2019.2.6)　(24)登録日　平成31年1月18日(2019.1.18)

(51)Int.Cl.		F I		
G10L 15/10	(2006.01)	G10L	15/10	500N
G10L 13/08	(2013.01)	G10L	13/08	124
G10L 15/22	(2006.01)	G10L	15/22	300Z
G10L 25/63	(2013.01)	G10L	25/63	
G10L 13/00	(2006.01)	G10L	13/00	100M

請求項の数 5　(全 13 頁)　最終頁に続く

(21)出願番号　特願2016-109314 (P2016-109314)
(22)出願日　平成28年5月31日 (2016.5.31)
(65)公開番号　特開2017-215468 (P2017-215468A)
(43)公開日　平成29年12月7日 (2017.12.7)
審査請求日　平成29年7月11日 (2017.7.11)

(73)特許権者　000003207
トヨタ自動車株式会社
愛知県豊田市トヨタ町１番地
(74)代理人　100100549
弁理士　川口　嘉之
(74)代理人　100085006
弁理士　世良　和信
(74)代理人　100113608
弁理士　平川　明
(74)代理人　100123319
弁理士　関根　武彦
(74)代理人　100123098
弁理士　今堀　克彦
(74)代理人　100143797
弁理士　宮下　文徳

最終頁に続く

(54)【発明の名称】音声対話装置および音声対話方法

(57)【特許請求の範囲】
【請求項１】
ユーザが発した音声と、当該音声を認識した結果を取得する音声処理手段と、
それぞれ異なる方法によって前記ユーザの感情を推定する複数の推定手段と、
前記推定したユーザの感情に基づいて応答文を生成し、前記ユーザに提供する応答手段と、を有し、
前記応答手段は、前記複数の推定手段がそれぞれ出力した感情推定結果に不一致が発生した場合に、過去に感情の推定を行った結果である推定履歴を取得し、前記推定履歴に基づいて前記不一致を解消する、
音声対話装置。
【請求項２】
前記推定履歴は、前記推定手段のそれぞれが過去に感情の推定を行った結果、正しい結果が得られたかに関する情報である正誤情報を含み、
前記応答手段は、前記正誤情報に基づいて前記推定手段ごとに重み付けを行う、
請求項１に記載の音声対話装置。
【請求項３】
音声対話装置が、
ユーザが発した音声と、当該音声を認識した結果を取得する音声処理ステップと、
それぞれ異なる方法によって前記ユーザの感情を推定する複数の推定ステップと、
前記推定したユーザの感情に基づいて応答文を生成し、前記ユーザに提供する応答ステ

ップと、を実行し、
前記応答ステップでは、前記複数の推定ステップでそれぞれ推定した感情推定結果に不一致が発生した場合に、過去に感情の推定を行った結果である推定履歴を取得し、前記推定履歴に基づいて前記不一致を解消する、
音声対話方法。
【請求項４】
前記推定履歴は、前記推定ステップのそれぞれにおいて過去に感情の推定を行った結果、正しい結果が得られたかに関する情報である正誤情報を含み、
前記応答ステップでは、前記正誤情報に基づいて、複数の前記推定ステップにおいて推定した結果に重み付けを行う、
請求項３に記載の音声対話方法。
【請求項５】
請求項３または４に記載の音声対話方法をコンピュータに実行させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声によってユーザと対話する装置に関する。
【背景技術】
【０００２】
近年、人と対話をすることによって様々な情報を提供するロボットが開発されている。例えば、特許文献１には、マイクによって入力された音声をネットワーク上で処理し、入力に対する応答を音声で返すコミュニケーションロボットが開示されている。
【０００３】
また、音声によって人と対話するシステムにおいて、ユーザの感情を読み取り、当該感情に基づいて応答を生成する技術が公知となっている。例えば、特許文献２には、ユーザが発した語句、ユーザの顔画像、ユーザの生理的情報などを取得し、感情を推定したうえで応答文を生成する対話処理装置が開示されている。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開２０１５－０１３３５１号公報
【特許文献２】特開２００１－２１５９９３号公報
【特許文献３】特開２０１０－２１７５０２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００５】
特許文献２に記載の装置のように、複数の要素に基づいて感情の推定を行った場合、それぞれが異なる結果を出力する場合がある。例えば、ユーザの顔画像からは「悲しみ」という感情が推定され、ユーザの発話内容からは「喜び」という感情が推定された場合などである。このような場合、適切な応答をどのように決定すればよいかという課題がある。
【０００６】
本発明は上記の課題を考慮してなされたものであり、複数の要素に基づいてユーザの感情を推定する音声対話装置において、推定精度を向上させることを目的とする。
【課題を解決するための手段】
【０００７】
本発明に係る音声対話装置は、
ユーザが発した音声と、当該音声を認識した結果を取得する音声処理手段と、それぞれ異なる方法によって前記ユーザの感情を推定する複数の推定手段と、前記推定したユーザの感情に基づいて応答文を生成し、前記ユーザに提供する応答手段と、を有し、前記応答

手段は、前記複数の推定手段がそれぞれ推定したユーザの感情が一致しない場合に、前記ユーザに対して問い掛けを行い、得られた応答の内容に基づいて、いずれの推定結果を採用するかを決定することを特徴とする。

【０００８】

本発明に係る音声対話装置は、ユーザが発話した音声を認識し、当該ユーザの感情を推定したうえで応答文を生成および提供する装置である。

推定手段は、ユーザの感情を推定する手段である。ユーザの感情は、例えば、表情、声のピッチやトーン、発話内容などに基づいて推定することができるが、これらに限られない。本発明に係る音声対話装置は、それぞれ異なる方法によってユーザの感情を推定する複数の推定手段を有している。

【０００９】

また、応答手段は、複数の推定手段が推定した感情に基づいて応答文を生成する手段である。応答文は、例えば、ユーザが行った発話の内容と、推定したユーザの感情に基づいて生成することができる。

ここで問題となるのが、複数の推定手段のうち一部が、他と異なる結果を出力した場合の扱いである。例えば、推定手段のうちの一部が「喜び」という感情を推定し、一部が「悲しみ」という感情を推定した場合、どの感情に基づいて応答文を生成すればよいかが問題となる。

【００１０】

これに対し、本発明に係る音声対話装置は、応答手段がユーザに対して問い掛けを行い、得られた応答に基づいて、どの推定結果を採用するかを決定する。問い掛けとは、特定の推定結果が正しいか否かを確認するためのものである。問い掛けは、例えば、推定した感情が正しいか否かを直接確認するものであってもよいし、会話を進めることで、推定した感情が正しいか否かを間接的に確認するものであってもよい。

かかる構成によると、不確かな推定結果に基づいて応答を生成することがなくなるため、ユーザに対してより自然な応答を返すことができる。

【００１１】

また、前記複数の推定手段は、前記ユーザの発話内容に基づいて前記ユーザの感情を推定する第一の推定手段と、前記第一の推定手段と異なる方法によって前記ユーザの感情を推定する第二の推定手段と、を含み、前記応答手段は、前記第一の推定手段が推定したユーザの感情と、前記第二の推定手段が推定したユーザの感情が一致しない場合に、前記ユーザに対する問い掛けを行うことを特徴としてもよい。

【００１２】

このように、ユーザの発話内容に基づいて判定した感情と、それ以外（例えば、ユーザを観察した結果など）に基づいて判定した感情とが矛盾している場合に問い掛けを行ってもよい。かかる構成によると、発話に現れにくい感情を判定することができる。

【００１３】

また、前記問い掛けは、前記第一の推定手段が推定したユーザの感情と、前記第二の推定手段が推定したユーザの感情が異なっていることを示したうえで、実際の感情を確認するものであることを特徴としてもよい。

【００１４】

例えば、発話内容に基づいて推定した感情が、表情に基づいて推定した感情と異なっていた場合、「楽しいって言ってるけど、悲しそうに見えるよ？」といったように、何に基づいて感情を推定したかという情報をユーザに与えるようにしてもよい。かかる構成によると、ユーザからより正確な情報を引き出すことができる。

【００１５】

また、前記第一の推定手段と異なる方法とは、前記ユーザの顔を撮影した画像、あるいは、前記ユーザから取得した音声に基づいて前記ユーザの感情を推定する方法であることを特徴としてもよい。

【００１６】

このように、ユーザの表情や声をセンシングすることで、発話に現れにくい感情を判定することができる。

【００１７】

また、前記複数の推定手段は、前記ユーザの顔を撮影した画像、前記ユーザから取得した音声、前記ユーザの発話内容の少なくともいずれかに基づいて前記ユーザの感情を推定することを特徴としてもよい。

【００１８】

推定手段は、例えば、ユーザの顔を撮影した画像や取得した音声を特徴量に変換し、当該特徴量に基づいて感情の推定を行ってもよい。また、ユーザの発話を認識し、内容を解析することで感情の推定を行ってもよい。もちろん、他の方法を用いてもよい。これらの異なる手法を併用することで、感情の推定精度を向上させることができる。

【００１９】

また、前記問い掛けは、感情の推定方法についての情報を含むことを特徴としてもよい。

【００２０】

かかる構成によると、装置がどのような判断を行ったかという情報をユーザに与えることができ、ユーザからより正確な情報を引き出すことができる。

【００２１】

また、本発明の第二の形態に係る音声対話装置は、

ユーザが発した音声と、当該音声を認識した結果を取得する音声処理手段と、それぞれ異なる方法によって前記ユーザの感情を推定する複数の推定手段と、前記推定したユーザの感情に基づいて応答文を生成し、前記ユーザに提供する応答手段と、を有し、前記応答手段は、前記複数の推定手段がそれぞれ出力した感情推定結果に不一致が発生した場合に、過去に感情の推定を行った結果である推定履歴を取得し、前記推定履歴に基づいて前記不一致を解消することを特徴とする。

【００２２】

複数の推定手段がそれぞれ出力した感情推定結果間に不一致が発生した場合、過去の推定履歴に基づいて不一致を解消してもよい。例えば、過去の傾向から大きく外れた推定結果を出力した推定手段がある場合、信頼度が低いと判定し、当該推定手段が推定した結果を修正ないし破棄してもよい。また、これ以外の方法によって、特定の推定手段の信頼度が低いことを判定し、当該推定手段が推定した結果を修正ないし破棄してもよい。

【００２３】

また、前記推定履歴は、前記推定手段のそれぞれが過去に感情の推定を行った結果、正しい結果が得られたかに関する情報である正誤情報を含み、前記応答手段は、感情を推定する際に、前記正誤情報に基づいて前記推定手段ごとに重み付けを行うことを特徴としてもよい。

【００２４】

複数の推定手段は、例えば、表情や声、発話内容など、異なる方法によって感情の推定を行うが、感情がどこに表れやすいかは対象者によって異なる場合がある。そこで、過去に感情推定を行った際の正誤に関する情報を推定履歴として残しておき、当該推定履歴に基づいて、推定手段ごとに重み付けを行う。これにより、より正確に感情が推定できる推定手段についてはより大きい重みを与え、正確性が高くない推定手段についてはより小さい重みを与えるといったことが可能になる。すなわち、対象者にあわせて最適な方法で感情の推定が行えるようになる。

【００２５】

なお、本発明は、上記手段の少なくとも一部を含む音声対話装置として特定することができる。また、前記音声対話装置が行う音声対話方法として特定することもできる。上記処理や手段は、技術的な矛盾が生じない限りにおいて、自由に組み合わせて実施することができる。

【発明の効果】

【００２６】
本発明によれば、複数の要素に基づいてユーザの感情を推定する音声対話装置において、推定精度を向上させることができる。
【図面の簡単な説明】
【００２７】
【図１】実施形態に係る音声対話システムのシステム構成図である。
【図２】実施形態に係る音声対話システムの処理フロー図である。
【図３】実施形態に係る音声対話システムの処理フロー図である。
【図４】ユーザの感情を確認するための質問の例である。
【図５】実施形態に係る音声対話システムの処理フロー図である。
【発明を実施するための形態】
【００２８】
以下、本発明の好ましい実施形態について図面を参照しながら説明する。
本実施形態に係る音声対話システムは、ユーザが発した音声を取得して音声認識を行い、認識結果に基づいて応答文を生成することでユーザとの対話を行うシステムである。
【００２９】
（第一の実施形態）
<システム構成>
図１は、本実施形態に係る音声対話システムのシステム構成図である。本実施形態に係る音声対話システムは、ロボット１０と、制御装置２０と、サーバ装置３０から構成される。
【００３０】
ロボット１０は、スピーカやマイク、カメラ等を有しており、ユーザとのインタフェースを担う手段である。ロボット１０は、人型やキャラクター型であってもよいし、他の形状であってもよい。
また、制御装置２０は、ロボット１０に対して制御命令を発行する装置である。本実施形態では、ロボット１０はユーザインタフェースとしてのみ機能し、発話内容の認識、その他の処理など、システム全体を制御する処理は制御装置２０が行う。
また、サーバ装置３０は、制御装置２０から送信された要求に応じて、ユーザに提供する応答（応答文）を生成する装置である。
【００３１】
まず、ロボット１０について説明する。
ロボット１０は、画像取得部１１、音声入力部１２、音声出力部１３、近距離通信部１４から構成される。
【００３２】
画像取得部１１は、不図示のカメラを用いて、ユーザの顔が含まれた画像（以下、顔画像）を取得する手段である。本実施形態では、ロボットの正面に取り付けられたカメラを用いて、ユーザの顔を撮像する。カメラは、ＲＧＢ画像を取得するカメラであってもよいし、グレースケール画像や、赤外線画像を取得するカメラであってもよい。画像取得部１１が取得した顔画像は、後述する近距離通信部１４を介して制御装置２０に送信される。
【００３３】
音声入力部１２は、ユーザが発した音声を取得する手段である。具体的には、不図示のマイクを用いて、音声を電気信号（以下、音声データ）に変換する。取得した音声データは、顔画像と同様に近距離通信部１４を介して制御装置２０へ送信される。
【００３４】
音声出力部１３は、ユーザに提供する音声を出力する手段である。具体的には、不図示のスピーカを用いて、制御装置２０から送信された音声データを音声に変換する。
【００３５】
近距離通信部１４は、制御装置２０と近距離無線通信を行う手段である。本実施形態では、近距離通信部１４は、Ｂｌｕｅｔｏｏｔｈ（登録商標）接続を利用した通信を行う。

近距離通信部１４は、ペアリング先となる制御装置２０に関する情報を記憶しており、簡便な操作で接続を行うことができる。

【００３６】

次に、制御装置２０について説明する。制御装置２０は、ロボット１０の制御を行う装置であって、典型的にはパーソナルコンピュータ、携帯電話、スマートフォンなどである。制御装置２０は、ＣＰＵ、主記憶装置、補助記憶装置を有する情報処理装置として構成することができる。補助記憶装置に記憶されたプログラムが主記憶装置にロードされ、ＣＰＵによって実行されることで、図１に図示した各手段が機能する。なお、図示した機能の全部または一部は、専用に設計された回路を用いて実行されてもよい。

【００３７】

制御装置２０は、近距離通信部２１、音声認識部２２、制御部２３、通信部２４から構成される。

【００３８】

近距離通信部２１が有する機能は、前述した近距離通信部１４と同様であるため、詳細な説明は省略する。

【００３９】

音声認識部２２は、ロボットが有する音声入力部１２が取得した音声に対して音声認識を行い、テキストに変換する手段である。音声認識は、既知の技術によって行うことができる。例えば、音声認識部２２には、音響モデルと認識辞書が記憶されており、取得した音声データと音響モデルとを比較して特徴を抽出し、抽出した特徴を認識辞書とをマッチングさせることで音声認識を行う。認識結果は、制御部２３へ送信される。

【００４０】

制御部２３は、音声認識部２２が音声認識を行った結果に基づいて、サーバ装置３０と通信を行い、応答を取得する手段である。具体的には、音声認識を行った結果得られたテキストを、通信部２４を介してサーバ装置３０（いずれも後述）に送信し、対応する応答をサーバ装置３０から受信する。また、音声合成機能によって、応答を音声データに変換し、音声出力部１３を介してユーザに提供する。これにより、ユーザは、自然言語による会話を行うことができる。

また、制御部２３は、ロボット１０から取得した情報に基づいて、ユーザの感情を推定する機能を有している。推定した感情はサーバ装置３０に送信され、応答文生成の用に供される。具体的な処理内容については後述する。

【００４１】

通信部２４は、通信回線（例えば携帯電話網）を介してネットワークにアクセスすることで、サーバ装置３０との通信を行う手段である。

【００４２】

サーバ装置３０は、ユーザに提供する応答文を生成するサーバ装置であり、通信部３１および応答生成部３２からなる。

通信部３１が有する機能は、前述した通信部２４と同様であるため、詳細な説明は省略する。

【００４３】

応答生成部３２は、制御装置２０から取得したテキストに基づいて、ユーザに提供する応答文を生成する手段である。提供する応答文は、例えば、事前に記憶された対話シナリオ（対話辞書）に基づくものであってもよいし、データベースやウェブを検索して得られた情報に基づくものであってもよい。また、応答生成部３２は、制御装置２０から取得したユーザの感情を加味して応答文を生成する。詳細な処理内容については後述する。

応答生成部３２が取得した情報は、制御装置２０へテキスト形式で送信され、その後、合成音声に変換され、ロボット１０を介してユーザに向けて出力される。

【００４４】

質問生成部３３は、制御装置２０から取得した指示に基づいて、ユーザの感情を特定するための質問を生成する手段である。例えば、制御装置２０が、ユーザの感情が一意に特

定できないと判断した場合、質問生成部３３が、当該ユーザの感情を特定するための質問を生成する。また、これに対するユーザの回答によって、制御装置２０がユーザの感情を一意に特定する。詳細な処理内容については後述する。

【００４５】

サーバ装置３０も、ＣＰＵ、主記憶装置、補助記憶装置を有する情報処理装置として構成することができる。補助記憶装置に記憶されたプログラムが主記憶装置にロードされ、ＣＰＵによって実行されることで、図１に図示した各手段が機能する。なお、図示した機能の全部または一部は、専用に設計された回路を用いて実行されてもよい。

【００４６】

<処理フローチャート>

次に、図１に示した各手段が行う処理とデータの流れについて、処理内容およびデータの流れを説明するフロー図である図２を参照しながら説明する。

まず、ステップＳ１１で、ロボット１０が有する音声入力部１２が、マイクを通してユーザが発話した音声を取得する。取得した音声は音声データに変換され、通信部を介して、制御装置２０が有する音声認識部２２へ送信される。

次に、ステップＳ１２で、ロボット１０が有する画像取得部１１が、カメラを通してユーザの顔画像を取得する。取得した顔画像は、通信部を介して制御装置２０が有する制御部２３へ送信される。

【００４７】

次に、音声認識部２２が、取得した音声データに対して音声認識を行い、テキストに変換する（ステップＳ１３）。音声認識の結果得られたテキストは、制御部２３へ送信される。また、制御部２３は、取得したテキストを一時的に記憶するとともに、サーバ装置３０が有する応答生成部３２および質問生成部３３へ送信する。また、応答生成部３２および質問生成部３３は、取得したテキストを一時的に記憶する。

【００４８】

次に、ステップＳ１４で、制御部２３が、取得した顔画像に基づいてユーザの感情を推定する。本ステップでは、顔画像を特徴量に変換し、当該特徴量に基づいて感情の推定を行う。ここで用いる特徴量として、例えば、ガボールフィルタの出力結果などが挙げられるが、これ以外であってもよい。感情の推定は、例えば、学習データに基づいて構築されたモデルと特徴量とを比較することで行ってもよい。なお、本実施形態では、ユーザの感情を、「ポジティブ」および「ネガティブ」の二つの属性（以下、感情極性）に分類するものとする。

【００４９】

次に、ステップＳ１５で、制御部２３が、音声に基づく感情推定を実行する。本ステップでは、音声を特徴量に変換し、当該特徴量に基づいて感情の推定を行う。特徴量とは、例えば、音声のスペクトル、強度、ピッチ、抑揚、テンポなどが挙げられるが、これ以外であってもよい。なお、特徴量を取得する際は、個人差を吸収するために正規化を行ってもよい。また、感情の推定においては、例えば、特徴量や、特徴量の変化に基づいて、各感情極性にどの程度適合するかを算出し、決定するようにしてもよい。

【００５０】

次に、ステップＳ１６で、制御部２３が、ユーザの発話内容に基づく感情推定を実行する。例えば、ステップＳ１３において取得した発話内容に対して形態素解析を行い、結果に基づいて感情極性を推定する。感情の推定は、例えば、学習データに基づいて構築されたモデルと解析結果とを比較することで行ってもよい。

感情の推定結果は、サーバ装置３０が有する応答生成部３２および質問生成部３３へ送信され、一時的に記憶される。

【００５１】

ここで、ステップＳ１４，Ｓ１５，Ｓ１６で行った感情の推定結果が、それぞれ一致しなかった場合を考える。例えば、顔画像に基づく推定結果が「ネガティブ」、音声に基づく推定結果が「ネガティブ」であり、発話内容に基づく推定結果が「ポジティブ」であっ

た場合を考える。このように、異なる複数の基準による推定結果がそれぞれ食い違う場合、ユーザの感情に基づいた応答文が精度よく生成できなくなる。そこで、本実施形態では、三つの推定結果のうちいずれかが他と異なる場合、図３に示した処理によって、どの推定結果を採用するかを決定する。

なお、三つの推定結果がすべて同じであった場合、図３の処理は省略し、図５に示した処理に移行する（後述）。

【００５２】

図３の処理について説明する。

ステップＳ１６の終了後、サーバ装置３０に送信された推定結果のうち、他と異なるものがあった場合、応答生成部３２は、応答文の生成を一旦停止し、質問生成部３３が、ユーザの感情を確定させるための質問を生成する。

【００５３】

ステップＳ２１では、質問生成部３３が、直前に取得した三つの感情推定結果と、ユーザの発話内容に基づいて、ユーザの感情を確認するための質問文を生成する。

図４は、三つの感情推定結果の組み合わせを示した図である。例示したように、三種類の推定結果のうち、少なくともいずれかが異なるパターンは６通りある。なお、図４に示したＰはポジティブを、Ｎはネガティブを意味する。

【００５４】

質問生成部３３は、図４に示した情報を予め有しており、制御装置２０から取得した推定結果に基づいて、ユーザに問い掛けるための質問文を生成する。

例えば、顔画像に基づく推定結果が「ネガティブ」であり、音声に基づく推定結果が「ポジティブ」であり、発話内容に基づく推定結果が「ポジティブ」であった場合、「元気がないように見えるけど、本当に（ユーザの発話内容）？」といった質問を生成する。なお、（ユーザの発話内容）は、直前にユーザが発したセリフである。

【００５５】

ステップＳ２１で生成された質問文は、制御装置２０へ送信され、制御部２３によって音声合成が行われる（ステップＳ２２）。そして、音声データが、ロボット１０が有する音声出力部１３へ送信され、ユーザに提供される（ステップＳ２３）。

【００５６】

一方、質問を受け取ったユーザが音声による回答を行うと、ステップＳ２４で当該音声が取得され、ステップＳ２５でテキストへの変換が行われる。この動作は、ステップＳ１１およびＳ１３と同様であるため、説明は省略する。ステップＳ２５得られたテキストは、応答生成部３２へ送信される。

【００５７】

ステップＳ２６では、ステップＳ１４～１６で推定した感情と、ユーザから取得した回答内容に基づいて、ユーザの感情を一意に確定させる。

例えば、ユーザが「楽しかった！」とロボットに話しかけた場合であって、「顔画像：ネガティブ」「音声：ポジティブ」「発話内容：ポジティブ」という判定を行った場合を考える。システムは、「元気がないように見えるけど、本当に楽しかった？」とユーザに問いかけ、これに対してユーザが、「疲れただけ。とても楽しかったよ」と回答したとする。この場合、ユーザが「ネガティブ」という感情極性を否定する発言をしているため、システムは、ユーザの感情が「ポジティブ」であると確定させる。この結果は、制御部２３から応答生成部３２へ送信される。

【００５８】

次に、図５を参照して説明する。図５は、ユーザの感情が一意に確定したあとのフロー図である。応答生成部３２は、確定したユーザの感情と、ユーザから得られた発話の内容に基づいて応答を生成する（ステップＳ３１）。なお、ユーザから得られた発話とは、ステップＳ１３で取得した内容であってもよいし、図３の処理を行っている場合、ステップＳ２５で取得した内容であってもよい。

なお、図２の処理が終わった時点で、ユーザの感情が一意に確定している場合、図３の

処理はスキップし、図５の処理が開始される。

【００５９】

前述したように、応答文は、自装置が有する対話辞書（対話シナリオ）を用いて生成してもよいし、外部にある情報ソース（データベースサーバやウェブサーバ）を用いて生成してもよい。また、当該対話辞書（対話シナリオ）は、予め感情別に分類されたものであってもよい。

生成された応答文は、制御装置２０へ送信され、音声データに変換（ステップＳ３２）されたのち、ロボット１０を介してユーザに提供される（ステップＳ３３）。この処理は、ステップＳ２２およびＳ２３と同様であるため、説明は省略する。

例えば、前述した例のように、ユーザが「疲れただけ。とても楽しかったよ」と回答した場合、「それは良かったね！」といったようなポジティブな回答がロボットから発せられる。

一方で、ユーザが「そう見える？本当は疲れてるんだよね」とネガティブな回答をした場合、システムは、ユーザの感情が「ネガティブ」であると判断する。この結果、例えば、「そうなんだ。今日はお疲れさま」といったように、ネガティブないしはユーザを労う回答がロボットから発せられる。

【００６０】

以上説明したように、本実施形態に係る音声対話システムは、複数の異なる方法によってユーザの感情を推定し、不一致が発生した場合に、ユーザに問い合わせることで当該不一致を解消する。このようにして取得したユーザの感情に基づいて応答文を生成することで、感情を誤って認識したまま応答を生成することがなくなり、対話の精度を向上させることができる。

【００６１】

（第二の実施形態）

第二の実施形態は、ステップＳ１３～Ｓ１６の処理にて、過去に行った感情推定の結果を加味して感情を推定する実施形態である。

【００６２】

第二の実施形態では、ステップＳ１３～Ｓ１６の処理において、推定した感情を時系列データとして記録する。また、複数の方法によって推定した感情に不一致が発生した場合に、当該時系列データ（すなわち過去の感情推定結果）に基づいて、推定した感情に対する信頼度を算出する。

信頼度の算出は、例えば、感情の変化量に基づいて行ってもよい。例えば、急激な感情の変化が発生したと判定した場合、信頼度を低くしてもよい。

そして、当該信頼度に基づいて推定結果を確定させる。例えば、信頼度が所定の値以下である場合、推定結果を破棄し、直前における推定結果を採用するようにしてもよい。

【００６３】

このような処理は、感情推定方法ごとに実行される。例えば、「顔画像：ネガティブ」「音声：ポジティブ」「発話内容：ポジティブ」という推定結果が得られたとする。ここで、顔画像についての過去の推定結果を参照した結果、低い信頼度が算出された場合、顔画像についての推定結果を破棄し、音声と発話内容のみに基づいて感情の推定を行ってもよい。

【００６４】

以上説明したように、第二の実施形態によると、異なる方法によって感情の推定を行った結果の間で不一致が発生した場合に、過去の感情推定結果に基づいて、推定結果を修正ないし破棄することで、当該不一致を解消する。これにより、対話の途中で一時的に推定精度の低下が発生した場合であっても、これに対応することができる。

なお、第二の実施形態では、図３に示した処理は必須ではない。例えば、ユーザへの問い掛けを行わず、前述した処理を行うことでユーザの感情を確定させてもよい。

【００６５】

（第三の実施形態）

第一の実施形態では、推定した感情に不一致が発生した場合、ユーザに問い掛けることで当該不一致を解消した。第三の実施形態は、これらの処理結果に基づいて、感情推定方法ごとの重みを算出し、当該重みを用いて感情の推定を行う実施形態である。

【００６６】

第三の実施形態では、ステップＳ２６でユーザの感情を確定する際に、「どの推定方法による感情推定が正しかったか」を判定する。例えば、「顔画像：ネガティブ」「音声：ポジティブ」「発話内容：ポジティブ」という結果が得られ、問い掛けを行った結果、「顔画像：ネガティブ」という推定が誤っていたことがわかったとする。この場合、顔画像に基づく推定が結果的に誤りであり、音声と発話内容に基づく推定が結果的に正しかったことがわかる。よって、制御部２３は、「顔画像」に対する重み係数を小さくする。または、「音声」と「発話内容」に対する重み係数を大きくする。推定方法ごとの重み係数は蓄積され、以降の感情推定において利用される。

なお、重み係数は、ユーザと関連付けて記憶されることが好ましい。例えば、取得した顔画像や音声に基づいてユーザを識別し、関連付けを行ってもよい。

【００６７】

第三の実施形態によると、例えば、感情が表情に出にくいユーザについては、顔画像に基づく推定結果に対して小さい重みを与え、感情が声に表れやすいユーザについては、音声に基づく推定結果に対して大きい重みを与えるといったことが可能になる。すなわち、ユーザの傾向に合った感情の推定を行うことができるようになり、感情の推定精度が向上する。

【００６８】

なお、第三の実施形態では、ユーザに問い掛けた結果に基づいて、「どの推定方法による感情推定が正しかったか」という情報を生成および蓄積したが、当該情報は、ユーザへの問い掛け以外によって生成してもよい。

【００６９】

（変形例）

上記の実施形態はあくまでも一例であって、本発明はその要旨を逸脱しない範囲内で適宜変更して実施しうる。

例えば、実施形態の説明では、音声認識部２２が音声認識を行ったが、音声認識をサーバ装置３０で行うようにしてもよい。この場合、制御装置２０が、音声データをサーバ装置に転送するようにしてもよい。

【００７０】

また、実施形態の説明では、三種類の感情推定方法を用いたが、二種類、あるいは四種類以上の感情推定方法を併用してもよい。

【００７１】

また、実施形態の説明では、ユーザの感情を「ポジティブ」と「ネガティブ」の二種類であるものとしたが、感情の種別は三種類以上であってもよい。この場合、異なる方法によって感情の推定を行うと、三種類以上の推定結果が同時に得られる場合がある。この場合、任意の方法によって絞り込みを行うようにしてもよい。また、一回の質問で絞り込むことができない場合、複数回の質問を行うことで、ユーザの感情を一意に確定させてもよい。また、ユーザの感情が一意に確定できない場合であっても、ユーザがある感情を持っている確率が高い場合、当該感情を持っているものとして処理を進めてもよい。

【００７２】

また、実施形態の説明では、「本当に楽しい？」など、ユーザに対して感情を直接確認するための質問を提示したが、ユーザの感情は間接的に確認してもよい。例えば、さらなる対話を行い、追加で得られた情報に基づいて、正解である感情を推定してもよい。

【符号の説明】

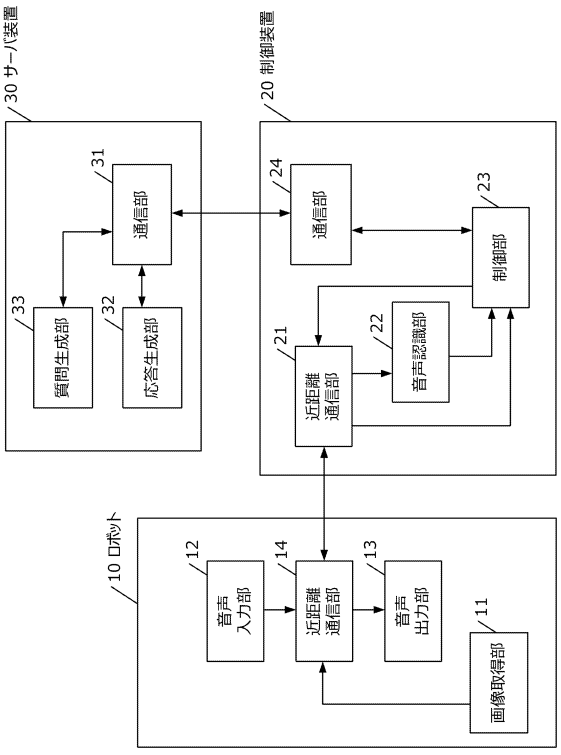
【００７３】

１０・・・ロボット

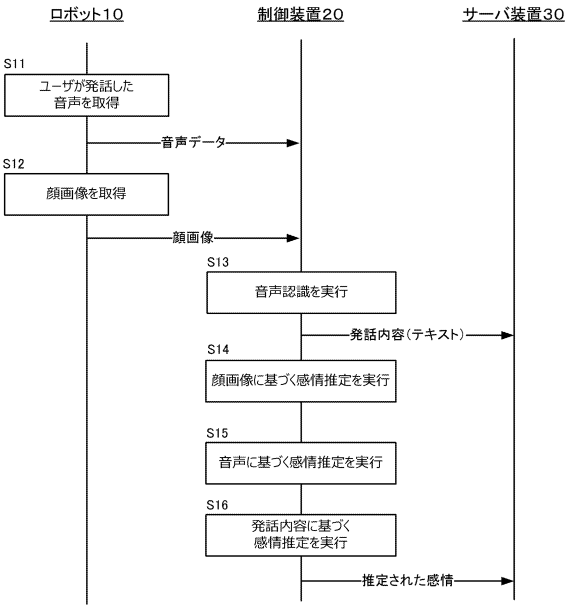
１１・・・画像取得部

１２・・・音声入力部
１３・・・音声出力部
１４，２１・・・近距離通信部
２０・・・制御装置
２２・・・音声認識部
２３・・・制御部
２４，３１・・・通信部
３０・・・サーバ装置
３２・・・応答生成部
３３・・・質問生成部

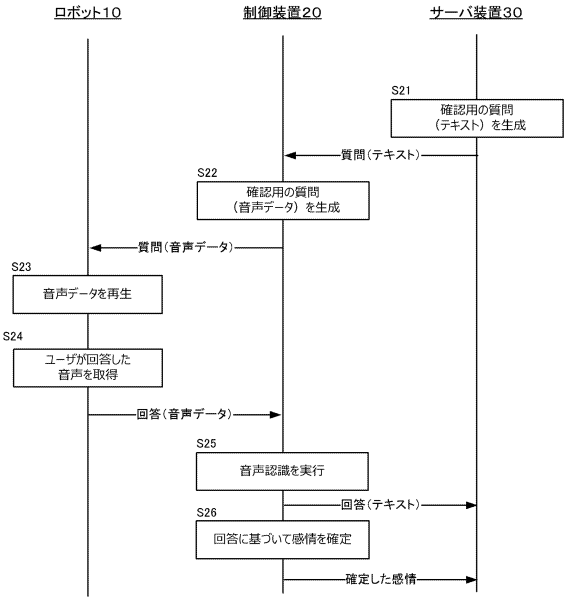
【図１】



【図２】



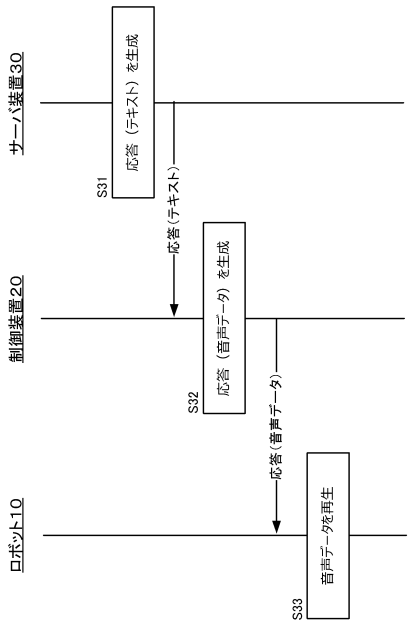
【図３】



【図４】

顔画像	音声	発話内容	質問
P	P	N	楽しそうに見えるけど、本当に（ユーザの発話内容）？
P	N	P	声が元気ないみたいだけど、本当に（ユーザの発話内容）？
N	P	P	元気がないように見えるけど、本当に（ユーザの発話内容）？
P	N	N	楽しそうに見えるけど、本当に（ユーザの発話内容）？
N	P	N	元気そうに聞こえるけど、本当に（ユーザの発話内容）？
N	N	P	元気ないように見えるけど、本当に（ユーザの発話内容）？

【図５】



フロントページの続き

(51)Int.Cl.			FI		
G06F	***3/01***	***(2006.01)***	G06F	3/01	510
G06F	***3/16***	***(2006.01)***	G06F	3/16	620
G06T	***7/20***	***(2017.01)***	G06F	3/16	610
			G06F	3/16	650
			G06T	7/20	300B

(74)代理人 100138357
弁理士 矢澤 広伸
(74)代理人 100176201
弁理士 小久保 篤史
(72)発明者 池野 篤司
東京都港区赤坂６丁目６番２０号 株式会社トヨタＩＴ開発センター内
(72)発明者 島田 宗明
東京都港区赤坂６丁目６番２０号 株式会社トヨタＩＴ開発センター内
(72)発明者 畠中 浩太
東京都港区赤坂６丁目６番２０号 株式会社トヨタＩＴ開発センター内
(72)発明者 西島 敏文
愛知県豊田市トヨタ町１番地 トヨタ自動車株式会社内
(72)発明者 片岡 史憲
愛知県豊田市トヨタ町１番地 トヨタ自動車株式会社内
(72)発明者 刀根川 浩巳
愛知県豊田市トヨタ町１番地 トヨタ自動車株式会社内
(72)発明者 梅山 倫秀
愛知県豊田市トヨタ町１番地 トヨタ自動車株式会社内

審査官 大野 弘

(56)参考文献 特開２００６－１７８０６３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ １５／１０
Ｇ０６Ｆ ３／０１
Ｇ０６Ｆ ３／１６
Ｇ０６Ｔ ７／２０
Ｇ１０Ｌ １３／００
Ｇ１０Ｌ １３／０８
Ｇ１０Ｌ １５／２２
Ｇ１０Ｌ ２５／６３