(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2020-34835
(P2020-34835A)
(43)公開日　令和2年3月5日(2020.3.5)

(51)Int.Cl.		F I			テーマコード（参考）
G10L 15/22	(2006.01)	G10L	15/22	300U	5E555
G10L 13/00	(2006.01)	G10L	13/00	100M	
G10L 15/10	(2006.01)	G10L	15/10	500N	
G06F 3/16	(2006.01)	G10L	15/10	500Z	
G06F 3/01	(2006.01)	G06F	3/16	650	

審査請求 未請求　請求項の数 11　ＯＬ　（全 27 頁）　最終頁に続く

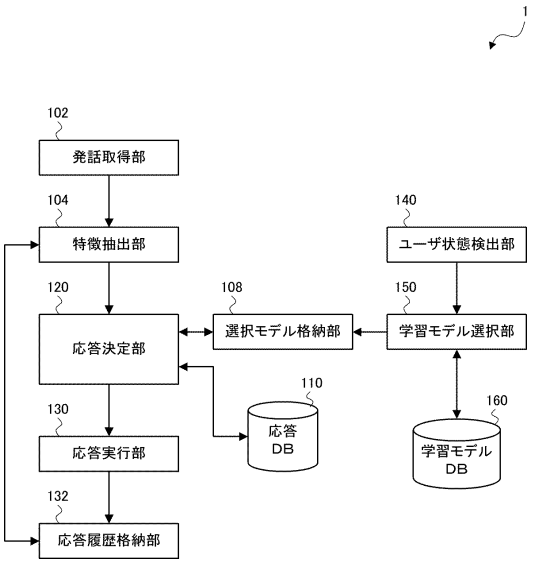
(21)出願番号　特願2018-162774 (P2018-162774)
(22)出願日　平成30年8月31日 (2018.8.31)

(71)出願人　504132272
国立大学法人京都大学
京都府京都市左京区吉田本町３６番地１
(71)出願人　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　100103894
弁理士　家入　健
(72)発明者　河原　達也
京都府京都市左京区吉田本町３６番地１
国立大学法人京都大学内
(72)発明者　堀　達朗
愛知県豊田市トヨタ町１番地　トヨタ自動車株式会社内

最終頁に続く

(54)【発明の名称】音声対話システム、音声対話方法、プログラム、学習モデル生成装置及び学習モデル生成方法

(57)【要約】
【課題】応答誤りが発生しないように適切に対処することが可能な音声対話システムを提供する。
【解決手段】発話取得部１０２は、ユーザ発話を取得する。特徴抽出部１０４は、ユーザ発話の特徴を抽出する。応答決定部１２０は、複数の学習モデルのうちのいずれかを用いて、抽出された特徴ベクトルに応じた応答を決定する。応答実行部１３０は、決定された応答を実行する。ユーザ状態検出部１４０は、ユーザ状態を検出する。学習モデル選択部１５０は、検出されたユーザ状態に応じて、複数の学習モデルから学習モデルを選択する。応答決定部１２０は、選択された学習モデルを用いて、応答を決定する。
【選択図】図２



【特許請求の範囲】
【請求項１】
ユーザと音声を用いた対話を行う音声対話システムであって、
前記ユーザの発話であるユーザ発話を取得する発話取得部と、
前記取得されたユーザ発話の特徴を少なくとも抽出する特徴抽出部と、
予め機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、前記抽出された特徴に応じた応答を決定する応答決定部と、
前記決定された応答を実行するための制御を行う応答実行部と、
前記ユーザの状態であるユーザ状態を検出するユーザ状態検出部と、
前記検出されたユーザ状態に応じて、前記複数の学習モデルから前記学習モデルを選択する学習モデル選択部と
を有し、
前記応答決定部は、前記学習モデル選択部によって選択された学習モデルを用いて、前記応答を決定する
音声対話システム。
【請求項２】
前記ユーザ状態検出部は、前記ユーザ状態として対話に対する前記ユーザの積極性の度合を検出し、
前記学習モデル選択部は、前記ユーザの積極性の度合に対応する前記学習モデルを選択する
請求項１に記載の音声対話システム。
【請求項３】
前記ユーザ状態検出部は、予め定められた期間における前記ユーザの発話量、又は、前記期間において当該音声対話システムが応答として音声を出力した時間と前記ユーザが発話した時間との合計に対する前記ユーザが発話した時間の割合を検出し、
前記学習モデル選択部は、前記ユーザの発話量又は前記ユーザが発話した時間の割合に対応する前記学習モデルを選択する
請求項２に記載の音声対話システム。
【請求項４】
前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの識別情報を検出し、
前記学習モデル選択部は、前記ユーザの識別情報に対応する前記学習モデルを選択する
請求項１に記載の音声対話システム。
【請求項５】
前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの感情を検出し、
前記学習モデル選択部は、前記ユーザの感情に対応する前記学習モデルを選択する
請求項１に記載の音声対話システム。
【請求項６】
前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの健康状態を検出し、
前記学習モデル選択部は、前記ユーザの健康状態に対応する前記学習モデルを選択する
請求項１に記載の音声対話システム。
【請求項７】
前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの覚醒状態の度合を検出し、
前記学習モデル選択部は、前記ユーザの覚醒状態の度合に対応する前記学習モデルを選択する
請求項１に記載の音声対話システム。
【請求項８】
ユーザと音声を用いた対話を行う音声対話システムを用いて行われる音声対話方法であって、
前記ユーザの発話であるユーザ発話を取得し、

前記取得されたユーザ発話の特徴を少なくとも抽出し、
予め機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、前記抽出された特徴に応じた応答を決定し、
前記決定された応答を実行するための制御を行い、
前記ユーザの状態であるユーザ状態を検出し、
前記検出されたユーザ状態に応じて、前記複数の学習モデルから前記学習モデルを選択し、
前記選択された学習モデルを用いて、前記応答を決定する
音声対話方法。
【請求項９】
ユーザと音声を用いた対話を行う音声対話システムを用いて行われる音声対話方法を実行するプログラムであって、
前記ユーザの発話であるユーザ発話を取得するステップと、
前記取得されたユーザ発話の特徴を少なくとも抽出するステップと、
予め機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、前記抽出された特徴に応じた応答を決定するステップと、
前記決定された応答を実行するための制御を行うステップと、
前記ユーザの状態であるユーザ状態を検出するステップと、
前記検出されたユーザ状態に応じて、前記複数の学習モデルから前記学習モデルを選択するステップと、
前記選択された学習モデルを用いて、前記応答を決定するステップと
をコンピュータに実行させるプログラム。
【請求項１０】
ユーザと音声を用いた対話を行う音声対話システムで用いられる学習モデルを生成する学習モデル生成装置であって、
１以上の任意ユーザと対話を行うことによって前記任意ユーザの発話であるユーザ発話を取得する発話取得部と、
前記取得されたユーザ発話の特徴を少なくとも示す特徴ベクトルを抽出する特徴抽出部と、
前記ユーザ発話に対する応答を示す正解ラベルと前記特徴ベクトルとが対応付けられたサンプルデータを生成するサンプルデータ生成部と、
前記ユーザ発話を発したときの前記任意ユーザの状態であるユーザ状態を取得して、前記取得されたユーザ状態を前記ユーザ発話に対応する前記サンプルデータに対応付けるユーザ状態取得部と、
前記ユーザ状態ごとに前記サンプルデータを分類するサンプルデータ分類部と、
前記分類された前記サンプルデータごとに、機械学習によって複数の学習モデルを生成する学習モデル生成部と
を有する学習モデル生成装置。
【請求項１１】
ユーザと音声を用いた対話を行う音声対話システムで用いられる学習モデルを生成する学習モデル生成方法であって、
１以上の任意ユーザと対話を行うことによって前記任意ユーザの発話であるユーザ発話を取得し、
前記取得されたユーザ発話の特徴を少なくとも示す特徴ベクトルを抽出し、
前記ユーザ発話に対する応答を示す正解ラベルと前記特徴ベクトルとが対応付けられたサンプルデータを生成し、
前記ユーザ発話を発したときの前記任意ユーザの状態であるユーザ状態を取得して、前記取得されたユーザ状態を前記ユーザ発話に対応する前記サンプルデータに対応付け、
前記ユーザ状態ごとに前記サンプルデータを分類し、
前記分類された前記サンプルデータごとに、機械学習によって複数の学習モデルを生成

する
学習モデル生成方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声対話システム、音声対話方法、プログラム、学習モデル生成装置及び学習モデル生成方法に関し、特に、ユーザと音声を用いた対話を行うことが可能な音声対話システム、音声対話方法、プログラム、学習モデル生成装置及び学習モデル生成方法に関する。
【背景技術】
【０００２】
ユーザが音声対話ロボット（音声対話システム）と日常会話を楽しむための技術が普及しつつある。この技術にかかる音声対話ロボットは、ユーザが発声した音声の音韻情報等を解析して、解析結果に応じた応答を実行する。ここで、音声対話ロボットが学習モデルを用いて応答を決定することが、行われている。
【０００３】
上記の技術に関連し、特許文献１は、ユーザの発する音声からそのユーザの感情状態を判断して適切な対応動作を実行する感情状態反応動作装置を開示する。特許文献１にかかる感情状態反応動作装置は、音声情報の音韻スペクトルに係る特徴量を抽出する音韻特徴量抽出手段と、音韻特徴量を入力して、予め備えた状態判断テーブルに基づき音声情報の感情状態を判断する状態判断手段と、感情状態を入力して、予め備えた対応動作選択テーブルに基づき対応動作処理を決定する対応動作選択手段とを有する。さらに、特許文献１にかかる感情状態反応動作装置は、感情状態学習用テーブルと感情状態学習手段を備え、感情状態学習手段は、感情状態学習テーブルに基づいて所定の機械学習モデルにより音韻特徴量と感情状態との関連を取得して状態判断テーブルに学習結果を保存し、状態判断手段は、状態判断テーブルに基づいて上記機械学習モデルによる感情状態の判断を行う。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開２００５－３５２１５４号公報
【発明の概要】
【発明が解決しようとする課題】
【０００５】
ユーザの状態（ユーザの違い、又は、ユーザの感情等）によっては、機械学習モデルが適切でないおそれがある。この場合、例えば、ユーザの発話と装置の発話とが衝突する発話衝突、又は、ユーザの発話と装置の発話との間の期間が長期に亘ってしまう長期沈黙といった、応答誤りが発生するおそれがある。一方、特許文献１にかかる技術では、１つの機械学習モデルを用いて対応動作処理を決定している。したがって、特許文献１にかかる技術では、学習モデルが適切でないために応答誤りが発生する場合であっても、応答誤りが発生しないように適切に対処することが困難である。
【０００６】
本発明は、応答誤りが発生しないように適切に対処することが可能な音声対話システム、音声対話方法、プログラム、学習モデル生成装置及び学習モデル生成方法を提供するものである。
【課題を解決するための手段】
【０００７】
本発明にかかる音声対話システムは、ユーザと音声を用いた対話を行う音声対話システムであって、前記ユーザの発話であるユーザ発話を取得する発話取得部と、前記取得されたユーザ発話の特徴を少なくとも抽出する特徴抽出部と、予め機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、前記抽出された特徴に応じた応答を決定

する応答決定部と、前記決定された応答を実行するための制御を行う応答実行部と、前記ユーザの状態であるユーザ状態を検出するユーザ状態検出部と、前記検出されたユーザ状態に応じて、前記複数の学習モデルから前記学習モデルを選択する学習モデル選択部とを有し、前記応答決定部は、前記学習モデル選択部によって選択された学習モデルを用いて、前記応答を決定する。

【０００８】

また、本発明にかかる音声対話方法は、ユーザと音声を用いた対話を行う音声対話システムを用いて行われる音声対話方法であって、前記ユーザの発話であるユーザ発話を取得し、前記取得されたユーザ発話の特徴を少なくとも抽出し、予め機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、前記抽出された特徴に応じた応答を決定し、前記決定された応答を実行するための制御を行い、前記ユーザの状態であるユーザ状態を検出し、前記検出されたユーザ状態に応じて、前記複数の学習モデルから前記学習モデルを選択し、前記選択された学習モデルを用いて、前記応答を決定する。

【０００９】

また、本発明にかかるプログラムは、ユーザと音声を用いた対話を行う音声対話システムを用いて行われる音声対話方法を実行するプログラムであって、前記ユーザの発話であるユーザ発話を取得するステップと、前記取得されたユーザ発話の特徴を少なくとも抽出するステップと、予め機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、前記抽出された特徴に応じた応答を決定するステップと、前記決定された応答を実行するための制御を行うステップと、前記ユーザの状態であるユーザ状態を検出するステップと、前記検出されたユーザ状態に応じて、前記複数の学習モデルから前記学習モデルを選択するステップと、前記選択された学習モデルを用いて、前記応答を決定するステップとをコンピュータに実行させる。

【００１０】

応答誤りが発生する要因は、学習モデルが適切でないことが多い。本発明は、上記のように構成されているので、学習モデルが適切でない場合に、ユーザ状態に応じて適切な学習モデルに切り替えることができる。したがって、本発明は、応答誤りが発生しないように適切に対処することが可能となる。

【００１１】

また、好ましくは、前記ユーザ状態検出部は、前記ユーザ状態として対話に対する前記ユーザの積極性の度合を検出し、前記学習モデル選択部は、前記ユーザの積極性の度合に対応する前記学習モデルを選択する。

本発明は、このように構成されていることによって、ユーザの対話に対する積極性の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの積極性に合わせて応答を実行することができる。

【００１２】

また、好ましくは、前記ユーザ状態検出部は、予め定められた期間における前記ユーザの発話量、又は、前記期間において当該音声対話システムが応答として音声を出力した時間と前記ユーザが発話した時間との合計に対する前記ユーザが発話した時間の割合を検出し、前記学習モデル選択部は、前記ユーザの発話量又は前記ユーザが発話した時間の割合に対応する前記学習モデルを選択する。

本発明は、このように構成されていることによって、より正確に、ユーザの積極性の度合を判定することができる。

【００１３】

また、好ましくは、前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの識別情報を検出し、前記学習モデル選択部は、前記ユーザの識別情報に対応する前記学習モデルを選択する。

本発明は、このように構成されていることによって、ユーザに適合した学習モデルを用いて対話を行うので、対話を行うユーザに合わせて応答を実行することができる。

【００１４】

また、好ましくは、前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの感情を検出し、前記学習モデル選択部は、前記ユーザの感情に対応する前記学習モデルを選択する。

本発明は、このように構成されていることによって、ユーザの対話に対する感情の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの感情に合わせて応答を実行することができる。

【００１５】

また、好ましくは、前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの健康状態を検出し、前記学習モデル選択部は、前記ユーザの健康状態に対応する前記学習モデルを選択する。

本発明は、このように構成されていることによって、ユーザの健康状態の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの健康状態に合わせて応答を実行することができる。

【００１６】

また、好ましくは、前記ユーザ状態検出部は、前記ユーザ状態として前記ユーザの覚醒状態の度合を検出し、前記学習モデル選択部は、前記ユーザの覚醒状態の度合に対応する前記学習モデルを選択する。

本発明は、このように構成されていることによって、ユーザの覚醒状態の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの覚醒状態に合わせて応答を実行することができる。

【００１７】

また、本発明にかかる学習モデル生成装置は、ユーザと音声を用いた対話を行う音声対話システムで用いられる学習モデルを生成する学習モデル生成装置であって、１以上の任意ユーザと対話を行うことによって前記任意ユーザの発話であるユーザ発話を取得する発話取得部と、前記取得されたユーザ発話の特徴を少なくとも示す特徴ベクトルを抽出する特徴抽出部と、前記ユーザ発話に対する応答を示す正解ラベルと前記特徴ベクトルとが対応付けられたサンプルデータを生成するサンプルデータ生成部と、前記ユーザ発話を発したときの前記任意ユーザの状態であるユーザ状態を取得して、前記取得されたユーザ状態を前記ユーザ発話に対応する前記サンプルデータに対応付けるユーザ状態取得部と、前記ユーザ状態ごとに前記サンプルデータを分類するサンプルデータ分類部と、前記分類された前記サンプルデータごとに、機械学習によって複数の学習モデルを生成する学習モデル生成部とを有する。

【００１８】

また、本発明にかかる学習モデル生成方法は、ユーザと音声を用いた対話を行う音声対話システムで用いられる学習モデルを生成する学習モデル生成方法であって、１以上の任意ユーザと対話を行うことによって前記任意ユーザの発話であるユーザ発話を取得し、前記取得されたユーザ発話の特徴を少なくとも示す特徴ベクトルを抽出し、前記ユーザ発話に対する応答を示す正解ラベルと前記特徴ベクトルとが対応付けられたサンプルデータを生成し、前記ユーザ発話を発したときの前記任意ユーザの状態であるユーザ状態を取得して、前記取得されたユーザ状態を前記ユーザ発話に対応する前記サンプルデータに対応付け、前記ユーザ状態ごとに前記サンプルデータを分類し、前記分類された前記サンプルデータごとに、機械学習によって複数の学習モデルを生成する。

【００１９】

本発明は、ユーザ状態ごとにサンプルデータを分類して機械学習によって複数の学習モデルを生成することによって、ユーザ状態に対応した複数の学習モデルを生成することができる。したがって、音声対話システムは、ユーザ状態に応じて学習モデルを選択することができる。

【発明の効果】

【００２０】

本発明によれば、応答誤りが発生しないように適切に対処することが可能な音声対話シ

ステム、音声対話方法、プログラム、学習モデル生成装置及び学習モデル生成方法を提供できる。

【図面の簡単な説明】

【００２１】

【図１】実施の形態１にかかる音声対話システムのハードウェア構成を示す図である。

【図２】実施の形態１にかかる音声対話システムの構成を示すブロック図である。

【図３】実施の形態１にかかる特徴抽出部によって生成される特徴ベクトルを例示する図である。

【図４】実施の形態１にかかる学習モデルの生成方法の概略を説明するための図である。

【図５】実施の形態１にかかる学習モデルの生成方法の概略を説明するための図である。

【図６】実施の形態１にかかる学習モデルの生成方法の概略を説明するための図である。

【図７】実施の形態１にかかる音声対話システムによってなされる音声対話方法を示すフローチャートである。

【図８】実施の形態１にかかる音声対話システムによってなされる音声対話方法を示すフローチャートである。

【図９】ユーザ状態がユーザの識別情報である場合における処理を示す図である。

【図１０】ユーザ状態がユーザの対話に対する積極性の度合である場合における処理を示す図である。

【図１１】積極性の度合を判定するためのテーブルを例示する図である。

【図１２】ユーザ状態がユーザの感情である場合における処理を示す図である。

【図１３】ユーザ状態がユーザの健康状態である場合における処理を示す図である。

【図１４】ユーザ状態がユーザの覚醒状態の度合である場合における処理を示す図である。

【図１５】実施の形態２にかかる音声対話システムの構成を示すブロック図である。

【図１６】実施の形態２にかかる学習モデル生成装置の構成を示す図である。

【図１７】実施の形態２にかかる学習モデル生成装置によって実行される学習モデル生成方法を示すフローチャートである。

【発明を実施するための形態】

【００２２】

（実施の形態１）

以下、図面を参照して本発明の実施の形態について説明する。なお、各図面において、同一の要素には同一の符号が付されており、必要に応じて重複説明は省略されている。

【００２３】

図１は、実施の形態１にかかる音声対話システム１のハードウェア構成を示す図である。音声対話システム１は、ユーザと音声を用いて対話を行う。具体的には、音声対話システム１は、ユーザからの発話（ユーザ発話）に応じて、ユーザに対して音声等の応答を実行することで、ユーザと対話を行う。音声対話システム１は、例えば、生活支援ロボット及び小型ロボット等のロボット、クラウドシステム及びスマートフォン等に搭載可能である。以下の説明では、音声対話システム１がロボットに搭載された例を示している。

【００２４】

音声対話システム１は、周囲の音声を収集するマイク２と、音声を発するスピーカ４と、ユーザの状態を検出するために使用される検出装置６と、ロボットの首等を動作させるマニピュレータ８と、制御装置１０とを有する。制御装置１０は、例えばコンピュータとしての機能を有する。制御装置１０は、マイク２、スピーカ４、検出装置６及びマニピュレータ８と、有線又は無線で接続されている。検出装置６は、例えば、カメラ及び生体センサの少なくとも１つを含む。生体センサは、例えば、血圧計、体温計、脈拍計等である。

【００２５】

制御装置１０は、主要なハードウェア構成として、ＣＰＵ（Central Processing Unit）１２と、ＲＯＭ（Read Only Memory）１４と、ＲＡＭ（Random Access Memory）１６と

、インタフェース部（ＩＦ；Interface）１８とを有する。ＣＰＵ１２、ＲＯＭ１４、ＲＡＭ１６及びインタフェース部１８は、データバスなどを介して相互に接続されている。

【００２６】

ＣＰＵ１２は、制御処理及び演算処理等を行う演算装置としての機能を有する。ＲＯＭ１４は、ＣＰＵ１２によって実行される制御プログラム及び演算プログラム等を記憶するための機能を有する。ＲＡＭ１６は、処理データ等を一時的に記憶するための機能を有する。インタフェース部１８は、有線又は無線を介して外部と信号の入出力を行う。また、インタフェース部１８は、ユーザによるデータの入力の操作を受け付け、ユーザに対して情報を表示する。

【００２７】

制御装置１０は、マイク２によって集音されたユーザ発話を解析して、そのユーザ発話に応じて、ユーザに対する応答を決定して実行する。ここで、本実施の形態では、「応答」は、「沈黙」、「頷き」及び「発話」を含む。「沈黙」は、音声対話システム１が何もしないという動作である。「頷き」は、ロボットの首部を縦に振るという動作である。また、「発話」は、音声対話システム１が音声を出力するという動作である。決定された応答が「頷き」である場合、制御装置１０は、マニピュレータ８を制御して、ロボットの首部を動作させる。また、決定された応答が「発話」である場合、制御装置１０は、スピーカ４を介して、生成された応答に対応する音声（システム発話）を出力する。

【００２８】

図２は、実施の形態１にかかる音声対話システム１の構成を示すブロック図である。実施の形態１にかかる音声対話システム１は、発話取得部１０２と、特徴抽出部１０４と、選択モデル格納部１０８と、応答データベース１１０（応答ＤＢ；Database）と、応答決定部１２０と、応答実行部１３０と、応答履歴格納部１３２とを有する。さらに、実施の形態１にかかる音声対話システム１は、ユーザ状態検出部１４０と、学習モデル選択部１５０と、学習モデルデータベース１６０（学習モデルＤＢ）とを有する。

【００２９】

図２に示す各構成要素は、マイク２、スピーカ４、マニピュレータ８及び制御装置１０の少なくとも１つによって実現可能である。また、各構成要素の少なくとも１つは、例えば、ＣＰＵ１２がＲＯＭ１４に記憶されたプログラムを実行することによって実現可能である。また、必要なプログラムを任意の不揮発性記録媒体に記録しておき、必要に応じてインストールするようにしてもよい。なお、各構成要素は、上記のようにソフトウェアによって実現されることに限定されず、何らかの回路素子等のハードウェアによって実現されてもよい。さらに、図２に示す構成要素の全てが１つの装置に設けられている必要はなく、図２に示す構成要素の１つ以上は、他の構成要素とは物理的に別個の装置に設けられていてもよい。例えば、学習モデルデータベース１６０がサーバに設けられ、その他の構成要素が、サーバと通信可能な音声対話ロボット等に設けられるようにしてもよい。これらのことは、後述する他の実施の形態においても同様である。

【００３０】

発話取得部１０２は、マイク２を含み得る。発話取得部１０２は、ユーザ発話（及び音声対話システム１の発話）を取得する。具体的には、発話取得部１０２は、ユーザの発話（及び音声対話システム１の発話）を集音してデジタル信号に変換する。そして、発話取得部１０２は、ユーザ発話の音声データ（ユーザ音声データ）を、特徴抽出部１０４に対して出力する。

【００３１】

特徴抽出部１０４は、少なくともユーザ発話の特徴を抽出する。具体的には、特徴抽出部１０４は、ユーザ発話について、発話の具体的な意味内容を示す言語情報とは異なる非言語情報の解析を行う。また、特徴抽出部１０４は、非言語情報の解析結果である非言語情報解析結果として、後述する特徴ベクトルを生成する。そして、特徴抽出部１０４は、非言語情報解析結果（特徴ベクトル）を、応答決定部１２０に対して出力する。なお、特徴抽出部１０４は、ユーザ発話以外のユーザの特徴を抽出して特徴ベクトルを生成しても

よい。

【００３２】

ここで、非言語情報とは、処理対象のユーザ発話の言語情報（文字列）とは異なる情報であり、ユーザ発話の韻律情報、及び、応答履歴情報の少なくとも一方を含む。韻律情報とは、ユーザ発話の音声波形の特徴を示す情報であり、例えば、基本周波数、音圧、周波数等の変化量、変動帯域、振幅の最大値及び平均値等である。また、応答履歴情報とは、応答決定部１２０によって決定（生成）され、応答実行部１３０によって実行された応答に関する過去の履歴を示す情報である。応答履歴格納部１３２は、応答実行部１３０によって応答が実行されると、この応答履歴情報を格納（更新）する。

【００３３】

具体的には、特徴抽出部１０４は、発話取得部１０２によって取得されたユーザ音声データについて音声分析等を行って、音声波形から韻律情報を解析する。そして、特徴抽出部１０４は、韻律情報を示す特徴量を示す値を算出する。なお、特徴抽出部１０４は、ユーザ音声データについて、例えば３２ｍｓｅｃごとに区切られたフレームごとに、基本周波数等を算出してもよい。また、特徴抽出部１０４は、応答履歴格納部１３２から応答履歴情報を抽出して、応答履歴の特徴を示す特徴量を算出する。

【００３４】

なお、ユーザ発話の言語情報を用いた構文解析は、パターン認識等を用いるため、多大な時間を要することが多い。一方、非言語情報の解析（韻律情報の解析及び応答履歴情報の解析）については、解析に用いられるデータ量が構文解析と比較して少なく、演算手法が、構文解析と比較して単純である。したがって、非言語情報の解析に要する時間は、構文解析と比較してかなり短くなり得る。

【００３５】

選択モデル格納部１０８は、後述する学習モデル選択部１５０によって選択された学習モデルを格納する。ここで、本実施の形態においては、学習モデル選択部１５０は、学習モデルデータベース１６０に格納された複数の学習モデルから、後述する方法によって、適切な学習モデルを選択する。なお、音声対話の開始前など、学習モデル選択部１５０によって学習モデルの選択処理がなされていない場合は、選択モデル格納部１０８は、任意の１つの学習モデルを格納してもよい。

【００３６】

応答データベース１１０は、音声対話システム１が応答を行う際に必要なデータを格納する。例えば、応答データベース１１０は、応答が「発話」である場合のシステム発話を示す複数のシステム音声データを、予め記憶している。

【００３７】

応答決定部１２０は、非言語情報解析結果（特徴ベクトル）に応じて、どの応答を実行するかを決定する。ここで、本実施の形態においては、応答決定部１２０は、予め、教師あり学習等の機械学習によって生成された複数の学習モデルのうちのいずれかを用いて、抽出された特徴（特徴ベクトル）に応じた応答を決定する。詳しくは後述する。

【００３８】

本実施の形態においては、応答決定部１２０は、「沈黙」、「頷き」及び「発話」のうちの１つを、応答として決定する。応答決定部１２０は、決定された応答を示すデータ（応答データ）を、応答実行部１３０に対して出力する。なお。応答決定部１２０は、応答として「発話」を行うと決定した場合、応答データベース１１０に記憶された複数のシステム発話から、順番に、又はランダムに、システム発話（システム音声データ）を選択してもよい。応答決定部１２０は、選択されたシステム音声データを、応答実行部１３０に対して出力する。

【００３９】

応答実行部１３０は、応答決定部１２０によって決定された応答を実行するための制御を行う。具体的には、応答決定部１２０から出力された応答データが「沈黙（沈黙応答）」を示す場合、応答実行部１３０は、スピーカ４及びマニピュレータ８を動作させないよ

うに制御する。また、応答決定部１２０から出力された応答データが「頷き（頷き応答）」を示す場合、応答実行部１３０は、マニピュレータ８を制御してロボットの首部を動作させる。また、応答決定部１２０から出力された応答データが「発話（発話応答）」を示す場合、応答実行部１３０は、スピーカ４を制御して、応答決定部１２０によって選択されたシステム音声データを示す音声を出力させる。

【００４０】

応答履歴格納部１３２は、応答実行部１３０によって実行された応答を識別するデータを、応答履歴情報として格納する。さらに、応答履歴格納部１３２は、応答履歴情報として対話に関する時間を含む場合に、その時間を計測し、計測された時間を応答履歴情報として格納してもよい。

【００４１】

図３は、実施の形態１にかかる特徴抽出部１０４によって生成される特徴ベクトルを例示する図である。なお、図３に例示する特徴ベクトルは、一例にすぎず、他の様々な特徴ベクトルが可能である。ｉ番目のユーザ発話についての特徴ベクトルをv_iとすると、特徴ベクトルのｎ個の成分は、$v_i=(v_{i1}, v_{i2}, \cdots, v_{i(m-1)}, v_{im}, v_{i(m+1)}, \cdots, v_{in})$と表される。ここで、ｉ、ｎ及びｍは整数（但しｎ＞ｍ）である。また、v_{i1}～$v_{i(m-1)}$が、ｉ番目のユーザ発話情報に関する韻律情報の解析結果に対応する。また、v_{im}～v_{in}が、応答履歴情報の解析結果に対応する。なお、v_{im}～v_{in}については、応答履歴格納部１３２に格納された情報そのものであってもよい。つまり、特徴抽出部１０４は、応答履歴情報については、応答履歴格納部１３２から応答履歴を抽出するのみでもよく、特別な解析を行わなくてもよい。

【００４２】

図３に示す例では、例えば、v_{i1}は、ｉ番目のユーザ発話の句末のＴ１ｍｓｅｃ（ユーザ発話の終了時点からＴ１ｍｓｅｃ（Ｔミリ秒）遡った時間から終了時点までの期間）における基本周波数ｆ０（$f0_{T1}$）についてのパラメータを示す。また、v_{i7}は、ｉ番目のユーザ発話の長さ（ユーザ発話長）Ｌ１［ｓｅｃ］を示す。なお、基本周波数ｆ０は、フレームごとに、ＳＰＴＫ（Speech Signal Processing Toolkit）のＳＷＩＰＥ（Saw-tooth Waveform Inspired Pitch Estimation）のロジックを用いて算出され得る。

【００４３】

また、v_{im}は、直前の応答タイプを示す。直前の応答タイプは、直前（ｉ番目のユーザ発話の直前）に、応答実行部１３０によって実行された応答のタイプ（「沈黙」、「頷き」、及び「発話」のいずれか）である。ここで、v_{im}のような、数値ではない成分の成分値（特徴量）については、各タイプに数値が割り当てられている。例えば、v_{im}において、成分値「１」は「沈黙」を示し、成分値「２」は「頷き」を示し、成分値「３」は「発話」を示す。

【００４４】

ユーザ状態検出部１４０（図２）は、ユーザ発話を発したユーザの状態（ユーザ状態）を検出する。詳しくは後述する。ユーザ状態検出部１４０は、検出されたユーザ状態を、学習モデル選択部１５０に対して出力する。ここで、ユーザ状態とは、例えば、ユーザの識別情報、対話に対するユーザの積極性、ユーザの感情、ユーザの健康状態、又は、ユーザの覚醒状態であるが、これらに限定されない。なお、ユーザの感情とは、例えば、ユーザの喜怒哀楽、又は驚き等であるが、これらに限定されない。また、ユーザの健康状態とは、例えば、ユーザの脈拍、体温又は血圧等であるが、これらに限定されない。ユーザ状態検出部１４０は、カメラによって撮影されたユーザの画像、生体センサによって検出されたユーザの脈拍、体温若しくは血圧、又はマイク２によって集音されたユーザ音声を用いて、上述したようなユーザ状態を検出する。

【００４５】

学習モデル選択部１５０は、ユーザ状態検出部１４０によって検出されたユーザ状態に応じて、学習モデルデータベース１６０に記憶された複数の学習モデルから学習モデルを選択する。詳しくは後述する。学習モデルデータベース１６０は、予め機械学習によって

生成された複数の学習モデルを記憶する。複数の学習モデルの生成方法の具体例については後述する。

【００４６】

応答誤りが発生する要因は、学習モデルが適切でないことが多い。例えば、あるユーザにとって適切な学習モデルが、別のユーザにとっては適切でないことがある。また、同じユーザであっても、そのユーザの感情等の変化によって、適切であった学習モデルが適切でなくなることがある。ここで、学習モデルが適切でないとは、ユーザ発話に対する応答の精度が悪いことである。言い換えると、学習モデルが適切でないとは、対話のテンポ及びリズムが良好でないということである。ユーザによって、適した対話のテンポ及びリズムは異なり得るので、あるユーザにとって適切な学習モデルが、別のユーザにとっては適切でないことがある。また、同じユーザであっても、そのユーザの感情等の変化によって、適した対話のテンポ及びリズムは異なり得るので、適切であった学習モデルが適切でなくなることがある。学習モデルの応答の精度が悪いと、ロボットは、あるユーザ発話に対して「沈黙応答」を実行すべきときに「発話応答」を実行してしまい、又は、あるユーザ発話に対して「発話応答」を実行すべきときに「沈黙応答」を実行してしまう。

【００４７】

これに対し、本実施の形態においては、ユーザ状態に応じて、学習モデルを適切なものに切り替えることができる。したがって、本実施の形態にかかる音声対話システム１は、応答誤りが発生しないように適切に対処することが可能となる。つまり、本実施の形態にかかる音声対話システム１は、応答の精度を良好にすることが可能となる。

【００４８】

次に、学習モデルの生成方法の概略を説明する。

図４～図６は、実施の形態１にかかる学習モデルの生成方法の概略を説明するための図である。まず、学習モデルを生成するためのサンプルデータを取得する。図４で示すように、ユーザＡの発話に対して、ロボット（音声対話システム１）が応答するといった、ユーザＡとロボットとの対話によって、特徴ベクトルと正解ラベルとが対応付けられたサンプルデータ（教師データ）を収集する。このとき、オペレータは、ユーザＡの発話に対して適切な応答をロボットが実行するように、ロボット（音声対話システム１）を操作する。

【００４９】

図４に示す例では、正解ラベル「Ａ」が、「沈黙応答」に対応する。正解ラベル「Ｂ」が、「頷き応答」に対応する。正解ラベル「Ｃ」が、「発話応答」に対応する。オペレータは、ユーザＡの発話の途中では、沈黙応答が実行されるように、ロボットを操作する。このとき、オペレータは何もしなくてもよい。また、オペレータは、ユーザＡの発話の読点レベルの切れ目で頷き応答が実行されるように、ロボットを操作する。このとき、ロボットは、オペレータの操作によって頷く動作を行う。また、オペレータは、ユーザＡの発話の句点レベルの切れ目で発話応答が実行されるように、ロボットを操作する。このとき、ロボットは、オペレータの操作によって発話を行う。

【００５０】

図４の例では、ユーザＡの発話「結局」と「１人で聴いたよ。」との間には切れ目がないので、オペレータは、ユーザＡの発話の途中であると判断し、沈黙応答が実行されるように、ロボットを操作する。また、ユーザＡの発話「１人で聴いたよ。」が終了すると、句点レベルの切れ目があったと判断し、発話応答が実行されるようにロボットを操作する。このとき、ロボットは、発話「本当ですか」を出力する。

【００５１】

さらに、ユーザＡのユーザ状態が検出される。ユーザ状態は、例えばオペレータによって判断されてもよいし、上述したユーザ状態検出部１４０のような機能により自動的に検出されてもよい。これにより、ユーザ状態＃１～＃Ｎのいずれかが、サンプルデータに対応付けられる。ここで、Ｎは、２以上の整数であり、ユーザ状態の個数を示す。このＮが、学習モデルの数に対応する。

【００５２】
図５は、図４の例によって取得された特徴ベクトルと正解ラベルとの組であるサンプルデータを例示する図である。ユーザ発話「結局」のユーザ発話長は０．５秒であったので、特徴ベクトルの成分（図３の$v_{i\,7}$）に「０．５」が入力される。また、ユーザ発話「結局」に対する応答は「沈黙応答」であったので、ユーザ発話「結局」の特徴ベクトルには、正解ラベル「Ａ」が対応付けられる。
【００５３】
また、ユーザ発話「１人で聴いたよ。」のユーザ発話長は１．５秒であったので、特徴ベクトルの成分（図３の$v_{i\,7}$）に「１．５」が入力される。また、ユーザ発話「１人で聴いたよ。」に対する応答は「発話応答」であったので、ユーザ発話「１人で聴いたよ。」の特徴ベクトルには、正解ラベル「Ｃ」が対応付けられる。さらに、この一連のユーザ発話「結局１人で聴いたよ。」では、ユーザ状態（例えばユーザの識別情報）が「ユーザ状態＃１」（例えば「ユーザＡ」）であったので、ユーザ発話「結局１人で聴いたよ。」に対応するサンプルデータ群に、ユーザ状態＃１が対応付けられる。
【００５４】
図６は、分類されたサンプルデータ群から学習モデルが生成される態様を例示する図である。上記のようにして収集されたサンプルデータ群が、ユーザ状態＃１～＃Ｎごとに、Ｎ個のグループに分類される。ユーザ状態＃１のサンプルデータ群（例えば「ユーザＡ」のユーザ発話に対応するサンプルデータ群）から、例えば教師あり学習等の機械学習によって、学習モデル＃１が生成される。同様にして、ユーザ状態＃Ｎのサンプルデータ群（例えば「ユーザＮ」のユーザ発話に対応するサンプルデータ群）から、機械学習によって、学習モデル＃Ｎが生成される。学習モデル＃１～＃Ｎの正解ラベル「Ａ」，「Ｂ」，「Ｃ」の境界が互いに異なっているので、学習モデル＃１～＃Ｎそれぞれに同じ特徴ベクトルを入力した場合であっても、出力される応答は異なり得る。このように生成された複数の学習モデルが、学習モデルデータベース１６０に格納される。学習モデルを生成するために使用される機械学習の方法は、例えば、ランダムフォレストであってもよいし、サポートベクターマシン（ＳＶＭ；Support Vector Machine）であってもよいし、ディープラーニングであってもよい。
【００５５】
なお、学習モデルは、ユーザ状態に応じて、正解ラベル「Ａ」，「Ｂ」，「Ｃ」の境界が定められている。例えば、ユーザ状態が「積極性」である場合、積極性の度合が大きなユーザ状態に対応する学習モデルであるほど、「発話」が選択される確率が低くなり得、「沈黙」が選択される確率が高くなり得る。これは、対話に対する積極性の度合が大きいということは、ユーザが積極的に発話する傾向にあるということであるので、発話衝突を抑制するため、音声対話システム１はあまり発話しないようにするということである。逆に、対話に対する積極性の度合が小さい場合には、ユーザがあまり積極的に発話しない傾向にあるということであるので、長期沈黙を抑制するため、音声対話システム１がより発話するようにする。
【００５６】
また、学習モデルは、上述したユーザ状態検出部１４０によって検出されるユーザ状態に対応している。例えば、ユーザ状態検出部１４０がユーザ状態として「積極性の度合」を検出する場合、学習モデルは、積極性の度合ごとに、複数設けられている。また、ユーザ状態検出部１４０がユーザ状態として「ユーザの識別情報」を検出する場合、学習モデルは、ユーザの識別情報（ユーザＡ、ユーザＢ、・・・、ユーザＮ等）ごとに、複数設けられている。
【００５７】
図７及び図８は、実施の形態１にかかる音声対話システム１によってなされる音声対話方法を示すフローチャートである。まず、発話取得部１０２は、上述したようにユーザ発話を取得する（ステップＳ１０２）。特徴抽出部１０４は、上述したように、取得されたユーザ発話について非言語情報（韻律情報及び応答履歴情報）の解析を行って、ユーザ発

話の特徴（特徴ベクトル）を抽出する（ステップＳ１０４）。
【００５８】
次に、応答決定部１２０は、現在の学習モデル（選択モデル格納部１０８に格納された学習モデル）を用いて、抽出された特徴ベクトルに応じた、ユーザ発話に対する応答を決定する（ステップＳ１１０）。応答実行部１３０は、上述したように、Ｓ１１０で決定された応答を実行する（ステップＳ１２０）。
【００５９】
図８は、Ｓ１１０の処理を示すフローチャートである。応答決定部１２０は、抽出された特徴ベクトルを、学習モデルに入力する（ステップＳ１１２）。応答決定部１２０は、学習モデルの出力を判定する（ステップＳ１１４）。
【００６０】
出力が「沈黙応答」である場合（Ｓ１１４の「沈黙」）、応答決定部１２０は、沈黙応答を実行すると決定する（ステップＳ１１６Ａ）。つまり、応答決定部１２０は、その特徴ベクトルに対応するユーザ発話に対して、何もしないと決定する。また、出力が「頷き応答」である場合（Ｓ１１４の「頷き」）、応答決定部１２０は、頷き応答を実行すると決定する（ステップＳ１１６Ｂ）。つまり、応答決定部１２０は、その特徴ベクトルに対応するユーザ発話に対して、ロボットの首部を縦に振るようにマニピュレータ８を動作させると決定する。また、出力が「発話応答」である場合（Ｓ１１４の「発話」）、応答決定部１２０は、発話応答を実行すると決定する（ステップＳ１１６Ｃ）。つまり、応答決定部１２０は、その特徴ベクトルに対応するユーザ発話に対して、システム発話を出力させるようにスピーカ４を動作させると決定する。
【００６１】
次に、ユーザ状態検出部１４０は、上述したように、ユーザ状態を検出する（ステップＳ１３０）。学習モデル選択部１５０は、Ｓ１３０の処理で検出されたユーザ状態に対応する学習モデルを選択する（ステップＳ１４０）。具体的には、現在の学習モデルが、検出されたユーザ状態に対応するものと異なる場合、学習モデル選択部１５０は、現在の学習モデルを、検出されたユーザ状態に対応する学習モデルに切り替える。一方、現在の学習モデルが、検出されたユーザ状態に対応するものである場合、学習モデル選択部１５０は、学習モデルを変更しない。このように、実施の形態１にかかる学習モデル選択部１５０は、ユーザ状態に応じた新たな学習モデルを選択するように構成されているので、応答の精度がより良くなる学習モデルを選択することが可能となる。
【００６２】
以下、ユーザ状態の具体例を説明する。第１の例は、ユーザ状態がユーザの識別情報である場合の例である。第２の例は、ユーザ状態が対話に対するユーザの積極性の度合である場合の例である。第３の例は、ユーザ状態がユーザの感情の度合である場合の例である。第４の例は、ユーザ状態がユーザの健康状態の度合である場合の例である。第５の例は、ユーザ状態がユーザの覚醒状態の度合である場合の例である。
【００６３】
（ユーザ状態の第１の例）
図９は、ユーザ状態がユーザの識別情報である場合における処理を示す図である。図９は、ユーザ状態がユーザの識別情報である場合における、Ｓ１３０，Ｓ１４０（図７）の具体的な処理を示す。ユーザ状態検出部１４０は、カメラである検出装置６から、ユーザの画像を取得する（ステップＳ１３２Ａ）。なお、「画像」とは、情報処理の対象としての、画像を示す画像データをも意味し得る（以下の説明において同様）。
【００６４】
ユーザ状態検出部１４０は、画像に対して顔認識処理を行って、ユーザの識別情報を検出する（ステップＳ１３４Ａ）。具体的には、ユーザ状態検出部１４０は、例えば、画像の中からユーザの顔領域を決定し、顔特徴点の検出を行って、目、鼻、口端などの顔の特徴点位置を判定する。そして、ユーザ状態検出部１４０は、特徴点位置を用いて顔領域の位置及び大きさを正規化した後、予め登録された人物の画像との顔照合処理を行う。これ

により、ユーザ状態検出部１４０は、照合された人物の識別情報を取得する。
【００６５】
次に、学習モデル選択部１５０は、検出された識別情報に対応する学習モデルを選択する（ステップＳ１４２Ａ）。なお、予め、ユーザの識別情報ごとに、複数の学習モデルが学習モデルデータベース１６０に格納されているとする。例えば、ユーザ状態検出部１４０によって「ユーザＡ」の識別情報が検出された場合、学習モデル選択部１５０は、「ユーザＡ」に対応する学習モデルを選択する。
【００６６】
このようにして、第１の例にかかる音声対話システム１は、ユーザに適合した学習モデルを用いて対話を行うので、対話を行うユーザに合わせて応答を実行することができる。したがって、第１の例にかかる音声対話システム１は、応答誤りが発生しないように適切に対処することが可能となる。また、応答誤りが発生するということは、現在の対話のテンポ又はリズムが、そのユーザに適していないということである。第１の例にかかる音声対話システム１は、ユーザに対応する学習モデルを選択することによって、対話のテンポ又はリズムをそのユーザに適したものすることが可能となる。
【００６７】
また、第１の例においては、学習モデルを生成する際に、ユーザ状態としてユーザの識別情報が対応付けられる。言い換えると、ユーザの識別情報ごとに、複数の学習モデルが生成される。学習モデルを生成する際には、例えばオペレータが、ユーザの識別情報を入力することで、サンプルデータとユーザの識別情報とが対応付けられる。これにより、ユーザの識別情報ごとに、サンプルデータが分類され、分類されたサンプルデータを用いて、機械学習によって複数の学習データが生成される。したがって、例えば、ユーザＡに対応する学習モデル、ユーザＢに対応する学習モデル、及び、ユーザＣに対応する学習モデルが生成されることとなる。
【００６８】
なお、上述した例では、画像を用いた顔認識処理によってユーザを識別するとしたが、ユーザを識別する方法は、この方法に限られない。ユーザ発話に対して話者認識処理を行うことによって、ユーザ発話を発したユーザを識別してもよい。さらに、ユーザの識別情報（ＩＤ）を入力することによって、ユーザを識別してもよい。
【００６９】
（ユーザ状態の第２の例）
図１０は、ユーザ状態がユーザの対話に対する積極性の度合である場合における処理を示す図である。図１０は、ユーザ状態がユーザの積極性の度合である場合における、Ｓ１３０，Ｓ１４０（図７）の具体的な処理を示す。ユーザ状態検出部１４０は、過去Ｔ分間におけるユーザ発話割合Ｒｓを取得する（ステップＳ１３２Ｂ）。ここで、Ｔは、予め定められた期間を示す。例えば、Ｔ＝１［分］であるが、これに限定されない。「過去Ｔ分間」とは、現在からＴ分間遡った時刻から現在までの期間である。ユーザ発話割合Ｒｓは、過去Ｔ分間における、音声対話システム１が応答として音声を出力した時間ｔｒ［分］とユーザ発話した時間ｔｕ［分］との合計ｔｕ＋ｔｒ［分］に対するユーザが発話した時間ｔｕの割合である。つまり、Ｒｓ［％］＝１００＊ｔｕ／（ｔｕ＋ｔｒ）である。
【００７０】
ユーザ状態検出部１４０は、ユーザ発話割合Ｒｓに対応する積極性の度合を検出する（ステップＳ１３４Ｂ）。具体的には、ユーザ状態検出部１４０は、図１１に例示するテーブルを、予め記憶している。ユーザ状態検出部１４０は、このテーブルを用いて、ユーザ発話割合Ｒｓが積極性のどの段階に対応するのかを判定する。
【００７１】
図１１は、積極性の度合を判定するためのテーブルを例示する図である。図１１に例示したテーブルでは、積極性の度合とユーザ発話割合Ｒｓとが対応付けられている。図１１の例では、積極性の度合が、＃１～＃４の４つの段階で定められている。度合＃１から度合＃４にかけて、積極性の度合が大きくなる。ユーザ状態検出部１４０は、取得されたユ

ーザ発話割合Ｒｓが、度合＃１～＃４のどの度合に対応するのかを判定する。例えば、Ｒｓ＝２０［％］である場合、ユーザ状態検出部１４０は、積極性の度合を＃１と判定する。また、Ｒｓ＝８０［％］である場合、ユーザ状態検出部１４０は、積極性の度合を＃４と判定する。

【００７２】

次に、学習モデル選択部１５０は、検出された積極性の度合に対応する学習モデルを選択する（ステップＳ１４２Ｂ）。なお、予め、ユーザの積極性の度合ごとに、複数の学習モデルが学習モデルデータベース１６０に格納されているとする。例えば、ユーザ状態検出部１４０によって「積極性の度合＃１」が検出された場合、学習モデル選択部１５０は、「積極性の度合＃１」に対応する学習モデルを選択する。また、ユーザ状態検出部１４０によって「積極性の度合＃４」が検出された場合、学習モデル選択部１５０は、「積極性の度合＃４」に対応する学習モデルを選択する。

【００７３】

なお、上述した説明では、ユーザ発話割合に応じてユーザの積極性の度合を判定するとしたが、ユーザの発話量に応じてユーザの積極性の度合を判定してもよい。具体的には、ユーザ状態検出部１４０は、過去Ｔ分間におけるユーザ発話量［分］を取得する（Ｓ１３２Ｂ）。ユーザ状態検出部１４０は、ユーザ発話量に対応する積極性の度合を検出する（Ｓ１３４Ｂ）。この場合、図１１に例示したテーブルと同様に、ユーザ状態検出部１４０は、ユーザ発話量と積極性の度合（段階）とが対応付けられたテーブルを記憶していてもよい。ユーザ状態検出部１４０は、このテーブルを用いて、ユーザ発話量が積極性のどの段階に対応するのかを判定し得る。

【００７４】

このようにして、第２の例にかかる音声対話システム１は、ユーザの対話に対する積極性の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの積極性に合わせて応答を実行することができる。したがって、第２の例にかかる音声対話システム１は、応答誤りが発生しないように適切に対処することが可能となる。また、応答誤りが発生するということは、現在の対話のテンポ又はリズムが、ユーザの積極性の度合に適していないということである。第２の例にかかる音声対話システム１は、ユーザの積極性の度合に対応する学習モデルを選択することによって、対話のテンポ又はリズムをユーザの積極性の度合に適したものすることが可能となる。また、ユーザの対話に対する積極性の度合は、対話の話題等によって変化し得る。第２の例にかかる音声対話システム１は、積極性の度合の変化に応じて学習モデルを変更することができる。

【００７５】

また、第２の例においては、学習モデルを生成する際に、ユーザ状態としてユーザの積極性の度合が対応付けられる。言い換えると、積極性の度合ごとに、複数の学習モデルが生成される。学習モデルを生成する際には、例えばオペレータが、対話中のユーザの積極性の度合を入力することで、サンプルデータとユーザの積極性の度合とが対応付けられる。また、学習モデルの生成の際でも、図１０に示したように、ユーザ発話割合又はユーザ発話量を用いて対話中のユーザの積極性の度合が判定されてもよい。この場合、オペレータが、期間Ｔを適宜設定してもよい。例えば、対話の話題が変更されたときに、ユーザの積極性の度合を計算するようにしてもよい。

【００７６】

これにより、ユーザの積極性の度合ごとに、サンプルデータが分類され、分類されたサンプルデータを用いて、機械学習によって複数の学習データが生成される。したがって、例えば、積極性の度合＃１に対応する学習モデル、積極性の度合＃２に対応する学習モデル、積極性の度合＃３に対応する学習モデル、及び、積極性の度合＃４に対応する学習モデルが生成されることとなる。

【００７７】

なお、上述したように、学習モデルは、ユーザ状態に応じて、正解ラベル「Ａ（沈黙）」，「Ｂ（頷き）」，「Ｃ（発話）」の境界が定められている。ユーザ状態が「積極性」

である場合、積極性の度合が大きなユーザ状態に対応する学習モデルであるほど、「発話」が選択される確率が低くなり得、「沈黙」が選択される確率が高くなり得る。つまり、学習モデル＃１（度合＃１）で「Ａ（沈黙）」が選択される確率よりも、学習モデル＃４（度合＃４）で「Ａ（沈黙）」が選択される確率の方が高くなるように、学習モデルが生成される。これにより、積極性の度合が大きなユーザとの対話において、発話衝突を抑制するため、音声対話システム１はあまり発話しないようにすることが可能となる。また、積極性の度合が小さいユーザとの対話において、長期沈黙を抑制するため、音声対話システム１はシステム発話を多くするようにすることが可能となる。

【００７８】

なお、上述した例では、ユーザ発話割合又はユーザ発話量を用いてユーザの対話に対する積極性の度合を検出するとしたが、ユーザの積極性の度合を検出する方法は、この方法に限られない。例えば、ユーザ状態検出部１４０は、ユーザの画像を取得することで、積極性の度合を検出してもよい。具体的には、ユーザ状態検出部１４０は、ユーザの顔画像に示されたユーザの表情及び視線を解析してユーザの積極性を判定し、積極性の度合を数値化してもよい。また、例えば、ユーザ状態検出部１４０は、ユーザ発話を取得することで、積極性の度合を検出してもよい。具体的には、ユーザ状態検出部１４０は、ユーザ発話の韻律を解析してユーザの積極性を判定し、積極性の度合を数値化してもよい。しかしながら、上述したように、ユーザ発話割合又はユーザ発話量を用いて積極性の度合を判定することにより、より正確に、ユーザの積極性の度合を判定することができる。したがって、ユーザ発話割合又はユーザ発話量を用いることにより、第２の例にかかる音声対話システム１は、応答誤りが発生しないように、より適切に対処することが可能となる。

【００７９】

（ユーザ状態の第３の例）

図１２は、ユーザ状態がユーザの感情である場合における処理を示す図である。図１２は、ユーザ状態がユーザの感情の度合である場合における、Ｓ１３０，Ｓ１４０（図７）の具体的な処理を示す。「感情の度合」とは、例えば「喜び」の度合である。しかしながら、「感情の度合」は、怒りの度合、悲しみの度合、又は驚きの度合であってもよい。

【００８０】

ユーザ状態検出部１４０は、カメラである検出装置６から、ユーザの顔画像を取得する（ステップＳ１３２Ｃ）。ユーザ状態検出部１４０は、顔画像を解析して、表情及び視線等から、ユーザの感情（喜び）の度合を検出する（ステップＳ１３４Ｃ）。例えば、ユーザ状態検出部１４０は、「Ａｆｆｄｅｘ」又は「Ｅｍｏｔｉｏｎ　ＡＰＩ」等の人工知能を用いて、ユーザの感情（喜び）を数値化してもよい。そして、ユーザ状態検出部１４０は、図１１に例示したような、感情を示す数値と感情の度合とを対応付けたテーブルを用いて、感情の度合を検出してもよい。

【００８１】

次に、学習モデル選択部１５０は、検出された感情（喜び）の度合に対応する学習モデルを選択する（ステップＳ１４２Ｃ）。なお、予め、ユーザの感情の度合ごとに、複数の学習モデルが学習モデルデータベース１６０に格納されているとする。例えば、ユーザ状態検出部１４０によって「感情（喜び）の度合＃１」が検出された場合、学習モデル選択部１５０は、「感情（喜び）の度合＃１」に対応する学習モデルを選択する。また、ユーザ状態検出部１４０によって「感情（喜び）の度合＃４」が検出された場合、学習モデル選択部１５０は、「感情（喜び）の度合＃４」に対応する学習モデルを選択する。

【００８２】

このようにして、第３の例にかかる音声対話システム１は、ユーザの対話に対する感情の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの感情に合わせて応答を実行することができる。したがって、第３の例にかかる音声対話システム１は、応答誤りが発生しないように適切に対処することが可能となる。また、応答誤りが発生するということは、現在の対話のテンポ又はリズムが、ユーザの感情の度合に適していないということである。第３の例にかかる音声対話システム１は、ユーザの感情の度合に対応

する学習モデルを選択することによって、対話のテンポ又はリズムをユーザの感情の度合に適したものすることが可能となる。また、ユーザの対話に対する感情の度合は、対話の話題等によって変化し得る。第３の例にかかる音声対話システム１は、感情の度合の変化に応じて学習モデルを変更することができる。

【００８３】

また、第３の例においては、学習モデルを生成する際に、ユーザ状態としてユーザの感情の度合が対応付けられる。言い換えると、感情の度合ごとに、複数の学習モデルが生成される。学習モデルを生成する際には、例えばオペレータが、対話中のユーザの感情の度合を入力することで、サンプルデータとユーザの感情の度合とが対応付けられる。また、学習モデルの生成の際でも、ユーザの顔画像を用いて対話中のユーザの感情の度合が判定されてもよい。

【００８４】

これにより、ユーザの感情の度合ごとに、サンプルデータが分類され、分類されたサンプルデータを用いて、機械学習によって複数の学習データが生成される。したがって、例えば、感情の度合＃１に対応する学習モデル、感情の度合＃２に対応する学習モデル、感情の度合＃３に対応する学習モデル、及び、感情の度合＃４に対応する学習モデルが生成されることとなる。

【００８５】

なお、上述した例では、ユーザの顔画像を用いてユーザの感情の度合を検出するとしたが、ユーザの感情の度合を検出する方法は、この方法に限られない。例えば、ユーザ状態検出部１４０は、ユーザ発話を取得することで、感情の度合を検出してもよい。具体的には、ユーザ状態検出部１４０は、ユーザ発話の韻律を解析してユーザの感情を判定し、感情の度合を数値化してもよい。

【００８６】

また、上述した例では、ユーザ状態検出部１４０が感情の度合を検出するとした。しかしながら、ユーザ状態検出部１４０は、感情の種類、つまり、喜び、悲しみ、怒り、驚き等を検出してもよい。具体的には、ユーザ状態検出部１４０は、喜び、悲しみ、怒り、驚きのそれぞれを示す数値を検出する。そして、ユーザ状態検出部１４０は、これらの数値のうち最も大きな値に対応する感情（例えば「怒り」）を、ユーザの感情として検出してもよい。この場合、学習モデルが、感情の種類ごとに複数設けられている。そして、学習モデル選択部１５０は、検出された感情の種類（例えば「怒り」）に応じた学習モデルを選択するようにしてもよい。

【００８７】

また、ユーザ状態検出部１４０は、感情の種類ごとに度合を検出してもよい。この場合、学習モデルデータベース１６０は、例えば、怒りの度合がＸ１であり驚きの度合がＹ１である場合の学習モデル、怒りの度合がＸ１であり驚きの度合がＹ２である場合の学習モデル、怒りの度合がＸ２であり驚きの度合がＹ１である場合の学習モデル、怒りの度合がＸ２であり驚きの度合がＹ２である場合の学習モデルを格納してもよい。そして、学習モデル選択部１５０は、検出された怒りの度合及び驚きの度合に対応する学習モデルを選択してもよい。

【００８８】

（ユーザ状態の第４の例）

図１３は、ユーザ状態がユーザの健康状態である場合における処理を示す図である。図１３は、ユーザ状態がユーザの健康状態の度合である場合における、Ｓ１３０，Ｓ１４０（図７）の具体的な処理を示す。「健康状態の度合」とは、例えば、心拍数の度合である。しかしながら、「健康状態の度合」は、血圧の度合又は体温の度合であってもよい。

【００８９】

ユーザ状態検出部１４０は、生体センサである検出装置６から、ユーザの生体系パラメータを取得する（ステップＳ１３２Ｄ）。生体系パラメータは、例えば心拍数である。ユーザ状態検出部１４０は、生体系パラメータから、ユーザの健康状態の度合を検出する（

ステップＳ１３４Ｄ）。例えば、ユーザ状態検出部１４０は、図１１に例示したような、健康状態を示す数値（心拍数）と健康状態の度合とを対応付けたテーブルを用いて、健康状態の度合を検出してもよい。

【００９０】

次に、学習モデル選択部１５０は、検出された健康状態（心拍数）の度合に対応する学習モデルを選択する（ステップＳ１４２Ｄ）。なお、予め、ユーザの健康状態の度合ごとに、複数の学習モデルが学習モデルデータベース１６０に格納されているとする。例えば、ユーザ状態検出部１４０によって「健康状態（心拍数）の度合＃１」が検出された場合、学習モデル選択部１５０は、「健康状態（心拍数）の度合＃１」に対応する学習モデルを選択する。また、ユーザ状態検出部１４０によって「健康状態（心拍数）の度合＃４」が検出された場合、学習モデル選択部１５０は、「健康状態（心拍数）の度合＃４」に対応する学習モデルを選択する。

【００９１】

このようにして、第４の例にかかる音声対話システム１は、ユーザの健康状態の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの健康状態に合わせて応答を実行することができる。したがって、第４の例にかかる音声対話システム１は、応答誤りが発生しないように適切に対処することが可能となる。また、応答誤りが発生するということは、現在の対話のテンポ又はリズムが、ユーザの健康状態の度合に適していないということである。第４の例にかかる音声対話システム１は、ユーザの健康状態の度合に対応する学習モデルを選択することによって、対話のテンポ又はリズムをユーザの健康状態の度合に適したものすることが可能となる。また、ユーザの心拍数等の度合は、対話の話題等によって変化し得る。第４の例にかかる音声対話システム１は、心拍数等の健康状態の度合の変化に応じて学習モデルを変更することができる。

【００９２】

また、第４の例においては、学習モデルを生成する際に、ユーザ状態としてユーザの健康状態の度合が対応付けられる。言い換えると、健康状態の度合ごとに、複数の学習モデルが生成される。学習モデルを生成する際には、例えば、生体センサを用いて対話中のユーザの健康状態の度合を入力することで、サンプルデータとユーザの健康状態の度合とが対応付けられる。

【００９３】

これにより、ユーザの健康状態の度合ごとに、サンプルデータが分類され、分類されたサンプルデータを用いて、機械学習によって複数の学習データが生成される。したがって、例えば、健康状態の度合＃１に対応する学習モデル、健康状態の度合＃２に対応する学習モデル、健康状態の度合＃３に対応する学習モデル、及び、健康状態の度合＃４に対応する学習モデルが生成されることとなる。

【００９４】

なお、上述した例では、生体センサを用いてユーザの健康状態の度合を検出するとしたが、ユーザの健康状態の度合を検出する方法は、この方法に限られない。例えば、ユーザ状態検出部１４０は、カメラである検出装置６からユーザの顔画像を取得することで、ユーザの健康状態の度合を検出してもよい。この場合、ユーザ状態検出部１４０は、顔画像を解析してユーザの顔色（赤色、青色、白色、黄色、黒色）を検出してもよい。そして、ユーザ状態検出部１４０は、ユーザの顔色が赤色、青色、白色、黄色、黒色のどの色に近いかに応じて、健康状態を検出してもよい。この場合、ユーザの顔色ごとに、複数の学習モデルが格納されている。

【００９５】

また、ユーザ状態検出部１４０は、複数の生体系パラメータ（心拍数、血圧及び体温から、ユーザの健康状態が良好であるか劣悪であるか、又は、ユーザの疲労度を判定してもよい。また、ユーザ状態検出部１４０は、心拍数、血圧及び体温それぞれが予め定められた正常範囲にあるか否かを判定し、正常範囲を逸脱した生体系パラメータの数に応じて、健康状態の度合（健康状態が良好か劣悪かの度合）を判定してもよい。

【００９６】
（ユーザ状態の第５の例）
図１４は、ユーザ状態がユーザの覚醒状態の度合である場合における処理を示す図である。図１４は、ユーザ状態がユーザの覚醒状態の度合である場合における、Ｓ１３０，Ｓ１４０（図７）の具体的な処理を示す。
【００９７】
ユーザ状態検出部１４０は、カメラ又は生体センサである検出装置６から、ユーザの生体系パラメータを取得する（ステップＳ１３２Ｅ）。生体系パラメータは、例えば、瞬目、心拍及び脳波の少なくとも１つである。なお、瞬目は、カメラから取得されたユーザの顔画像を解析することによって取得され得る。心拍及び脳波は、それぞれ生体センサである心拍計及び脳波計を用いて取得され得る。
【００９８】
ユーザ状態検出部１４０は、生体系パラメータから、ユーザの覚醒状態の度合を検出する（ステップＳ１３４Ｅ）。例えば、ユーザ状態検出部１４０は、上述した生体系パラメータから、覚醒度を算出する。例えば、ユーザ状態検出部１４０は、瞬目の間隔、瞬目の開眼時間、目の開度等により、覚醒度を算出し得る。そして、ユーザ状態検出部１４０は、図１１に例示したような、覚醒度と覚醒状態の度合とを対応付けたテーブルを用いて、覚醒状態の度合を検出してもよい。
【００９９】
次に、学習モデル選択部１５０は、検出された覚醒状態の度合に対応する学習モデルを選択する（ステップＳ１４２Ｅ）。なお、予め、ユーザの覚醒状態の度合ごとに、複数の学習モデルが学習モデルデータベース１６０に格納されているとする。例えば、ユーザ状態検出部１４０によって「覚醒状態の度合＃１」が検出された場合、学習モデル選択部１５０は、「覚醒状態の度合＃１」に対応する学習モデルを選択する。また、ユーザ状態検出部１４０によって「覚醒状態の度合＃４」が検出された場合、学習モデル選択部１５０は、「覚醒状態の度合＃４」に対応する学習モデルを選択する。
【０１００】
このようにして、第５の例にかかる音声対話システム１は、ユーザの覚醒状態の度合に適合した学習モデルを用いて対話を行うので、対話を行うユーザの覚醒状態に合わせて応答を実行することができる。したがって、第５の例にかかる音声対話システム１は、応答誤りが発生しないように適切に対処することが可能となる。また、応答誤りが発生するということは、現在の対話のテンポ又はリズムが、ユーザの覚醒状態の度合に適していないということである。第５の例にかかる音声対話システム１は、ユーザの覚醒状態の度合に対応する学習モデルを選択することによって、対話のテンポ又はリズムをユーザの覚醒状態の度合に適したものすることが可能となる。また、ユーザの覚醒度は、対話の話題等によって変化し得る。第５の例にかかる音声対話システム１は、覚醒度の変化に応じて学習モデルを変更することができる。
【０１０１】
また、第５の例においては、学習モデルを生成する際に、ユーザ状態としてユーザの覚醒状態の度合が対応付けられる。言い換えると、覚醒状態の度合ごとに、複数の学習モデルが生成される。学習モデルを生成する際には、例えば、カメラ又は生体センサを用いて対話中のユーザの覚醒状態の度合を入力することで、サンプルデータとユーザの覚醒状態の度合とが対応付けられる。
【０１０２】
これにより、ユーザの覚醒状態の度合ごとに、サンプルデータが分類され、分類されたサンプルデータを用いて、機械学習によって複数の学習データが生成される。したがって、例えば、覚醒状態の度合＃１に対応する学習モデル、覚醒状態の度合＃２に対応する学習モデル、覚醒状態の度合＃３に対応する学習モデル、及び、覚醒状態の度合＃４に対応する学習モデルが生成されることとなる。
【０１０３】

なお、上述した例では、カメラ又は生体センサを用いてユーザの覚醒状態の度合を検出するとしたが、ユーザの覚醒状態の度合を検出する方法は、この方法に限られない。ユーザ状態検出部１４０は、ユーザ発話を取得することで、覚醒状態の度合を検出してもよい。具体的には、ユーザ状態検出部１４０は、ユーザ発話の韻律を解析してユーザの覚醒状態を判定し、覚醒状態の度合を数値化してもよい。

【０１０４】

（実施の形態２）

次に、実施の形態２について説明する。実施の形態２においては、音声対話システム１が複数の学習モデルを生成する点で、実施の形態１と異なる。なお、実施の形態２にかかる音声対話システム１のハードウェア構成については、図１に示した実施の形態１にかかる音声対話システム１のハードウェア構成と実質的に同様であるので、説明を省略する。

【０１０５】

図１５は、実施の形態２にかかる音声対話システム１の構成を示すブロック図である。実施の形態２にかかる音声対話システム１は、発話取得部１０２と、特徴抽出部１０４と、選択モデル格納部１０８と、応答データベース１１０と、応答決定部１２０と、応答実行部１３０と、応答履歴格納部１３２とを有する。また、実施の形態１にかかる音声対話システム１は、ユーザ状態検出部１４０と、学習モデル選択部１５０と、学習モデルデータベース１６０とを有する。さらに、音声対話システム１は、学習モデル生成装置２００を有する。学習モデル生成装置２００以外の構成要素については、実施の形態１にかかるものと実質的に同様の機能を有するので、説明を省略する。

【０１０６】

学習モデル生成装置２００は、後述する方法によって、複数の学習モデルを生成する。学習モデル生成装置２００によって生成された複数の学習モデルは、学習モデルデータベース１６０に格納される。学習モデルは、学習モデル生成装置２００によって自動的に格納されてもよいし、オペレータ等の作業者によって手動で格納されてもよい。

【０１０７】

なお、学習モデル生成装置２００は、その他の構成要素と物理的に一体となっている必要はない。つまり、その他の構成要素が設けられた装置（ロボット等）と、学習モデル生成装置２００が設けられた装置（コンピュータ等）とが、同一である必要はない。学習モデル生成装置２００の具体的な構成について、以下に説明する。なお、学習モデル生成装置２００の処理（後述する図１７に示す処理）は、図４～図６に対応し、ユーザとの対話（図７の処理）の前段階で行われる。

【０１０８】

図１６は、実施の形態２にかかる学習モデル生成装置２００の構成を示す図である。また、図１７は、実施の形態２にかかる学習モデル生成装置２００によって実行される学習モデル生成方法を示すフローチャートである。学習モデル生成装置２００は、発話取得部２１２、特徴抽出部２１４、サンプルデータ生成部２１６、ユーザ状態取得部２１８、サンプルデータ分類部２２０、及び、学習モデル生成部２２２を有する。なお、学習モデル生成装置２００は、図１に示した音声対話システム１のハードウェア構成と実質的に同様のハードウェア構成を、独立して有しうる。

【０１０９】

発話取得部２１２は、１以上の任意ユーザと対話を行うことによって、図７のＳ１０２の処理と同様にして、任意ユーザの発話であるユーザ発話を取得する（ステップＳ２０２）。ここで、「任意ユーザ」とは、音声対話システム１が対話を行う相手のユーザに限られない、任意のユーザである。特徴抽出部２１４は、図７のＳ１０４の処理と同様にして、取得されたユーザ発話の特徴を少なくとも示す特徴ベクトルを抽出する（ステップＳ２０４）。

【０１１０】

次に、サンプルデータ生成部２１６は、ユーザ発話に対する応答を示す正解ラベルと特徴ベクトルとが対応付けられたサンプルデータを生成する（ステップＳ２０６）。具体的

には、サンプルデータ生成部２１６は、図４を用いて上述したように、オペレータによって判定された応答（正解ラベル）を、対応するユーザ発話の特徴ベクトルに対応付ける。これにより、サンプルデータ生成部２１６は、サンプルデータを生成する。なお、正解ラベルを自動的に判定することができれば、サンプルデータ生成部２１６は、ユーザ発話から正解ラベル（応答）を自動的に判定して、判定された正解ラベルをユーザ発話の特徴ベクトルに対応付けてもよい。次に、学習モデル生成装置２００（又は図２に示した応答実行部１３０）は、図７のＳ１２０の処理と同様にして、応答を実行する（ステップＳ２０８）。

【０１１１】

ユーザ状態取得部２１８は、ユーザ発話を発したときの任意ユーザの状態であるユーザ状態を取得して、取得されたユーザ状態をユーザ発話に対応するサンプルデータに対応付ける（ステップＳ２１０）。具体的には、ユーザ状態取得部２１８は、図９～図１４を用いて説明したように、ユーザの画像、ユーザ発話、又は生体系パラメータ等を用いて、任意ユーザのユーザ状態を取得し得る。ユーザ状態の取得方法は、ユーザ状態の種類（第１の例～第５の例）に応じて異なり得る。あるいは、ユーザ状態取得部２１８は、例えばオペレータによって判断された、任意ユーザのユーザ状態を取得してもよい。そして、ユーザ状態取得部２１８は、取得されたユーザ状態を、任意ユーザのユーザ発話に対応するサンプルデータに対応付ける。

【０１１２】

学習モデル生成装置２００は、ユーザ発話の取得を終了するか否かを判定する（ステップＳ２１２）。ユーザ発話の取得を継続する場合（Ｓ２１２のＮＯ）、学習モデル生成装置２００は、Ｓ２０２～Ｓ２１０の処理を繰り返す。一方、サンプルデータを十分に取得できたためにユーザ発話の取得を終了する場合（Ｓ２１２のＹＥＳ）、サンプルデータ分類部２２０は、図６を用いて説明したように、ユーザ状態ごとにサンプルデータを分類する（ステップＳ２２０）。そして、学習モデル生成部２２２は、図６を用いて説明したように、分類されたサンプルデータごとに、例えばランダムフォレスト又はサポートベクターマシン等の機械学習によって、複数の学習モデルを生成する（ステップＳ２２２）。

【０１１３】

このように、実施の形態２にかかる学習モデル生成装置２００は、ユーザ状態ごとにサンプルデータを分類して機械学習によって複数の学習モデルを生成することによって、ユーザ状態に対応した複数の学習モデルを生成することができる。したがって、音声対話システム１は、上述したように、ユーザ状態に応じて学習モデルを選択することができる。

【０１１４】

（変形例）

なお、本発明は上記実施の形態に限られたものではなく、趣旨を逸脱しない範囲で適宜変更することが可能である。例えば、上述したフローチャートにおいて、複数の処理の順序は、適宜、変更可能である。また、上述したフローチャートにおいて、複数の処理のうちの１つは、省略されてもよい。例えば、図７のＳ１３０の処理は、Ｓ１０２～Ｓ１２０の間に行われてもよい。

【０１１５】

また、図９～図１４を用いて説明したユーザ状態の第１の例～第５の例は、相互に適用可能である。つまり、ユーザ状態検出部１４０は、複数の種類のユーザ状態を検出してもよい。そして、学習モデル選択部１５０は、検出された複数の種類のユーザ状態に対応する学習モデルを選択してもよい。例えば、ユーザ状態検出部１４０は、ユーザの識別情報及びユーザの積極性の度合を検出してもよい。この場合、学習モデルデータベース１６０は、例えば、ユーザＡの積極性の度合ごと、ユーザＢの積極性の度合ごとに、複数の学習モデルを格納し得る。そして、ユーザ状態検出部１４０が、「ユーザＡ」の「積極性の度合＃１」を検出した場合に、学習モデル選択部１５０は、「ユーザＡ」の「積極性の度合＃１」に対応する学習モデルを選択し得る。

【０１１６】

また、上述した実施の形態では、特徴ベクトル（図３）は、ユーザ発話の韻律情報等から生成されるとしたが、このような構成に限られない。特徴ベクトルの成分は、韻律にかかるものだけでなく、カメラである検出装置６から取得されたユーザの特徴も含み得る。例えば、特徴ベクトルの成分は、ユーザの視線及び対話ロボットに対するユーザの距離を含んでもよい。

【０１１７】

また、上述した実施の形態においては、音声対話システム１がロボットに搭載された例を示しているが、このような構成に限られない。音声対話システム１は、スマートフォン又はタブレット端末等の情報端末にも搭載可能である。この場合、「頷き応答」を行うときは、マニピュレータ８を動作させる代わりに、情報端末の表示画面に、人物、動物、又はロボット等が頷くような動画を表示させてもよい。

【０１１８】

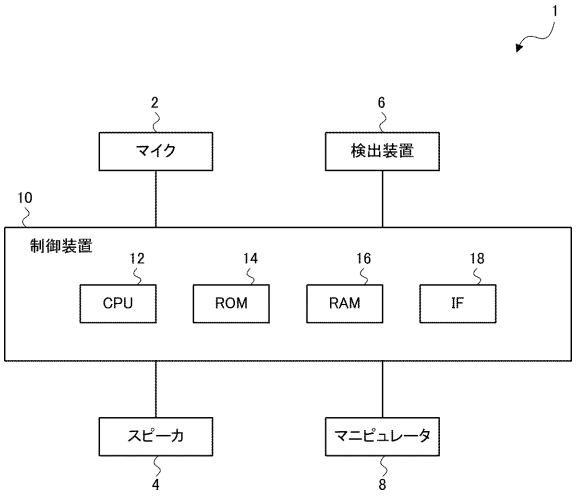
また、上述の例において、プログラムは、様々なタイプの非一時的なコンピュータ可読媒体（non-transitory computer readable medium）を用いて格納され、コンピュータに供給することができる。非一時的なコンピュータ可読媒体は、様々なタイプの実体のある記録媒体（tangible storage medium）を含む。非一時的なコンピュータ可読媒体の例は、磁気記録媒体（例えばフレキシブルディスク、磁気テープ、ハードディスクドライブ）、光磁気記録媒体（例えば光磁気ディスク）、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－Ｒ／Ｗ、半導体メモリ（例えば、マスクＲＯＭ、ＰＲＯＭ（Programmable ROM）、ＥＰＲＯＭ（Erasable PROM）、フラッシュＲＯＭ、ＲＡＭ）を含む。また、プログラムは、様々なタイプの一時的なコンピュータ可読媒体（transitory computer readable medium）によってコンピュータに供給されてもよい。一時的なコンピュータ可読媒体の例は、電気信号、光信号、及び電磁波を含む。一時的なコンピュータ可読媒体は、電線及び光ファイバ等の有線通信路、又は無線通信路を介して、プログラムをコンピュータに供給できる。

【符号の説明】

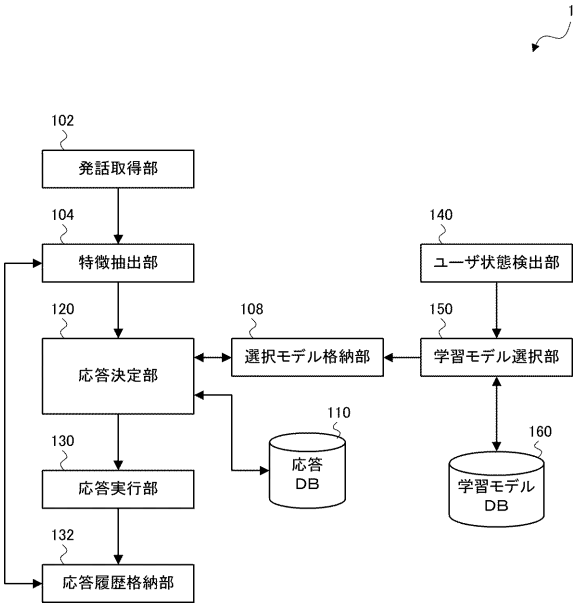
【０１１９】

１・・・音声対話システム、２・・・マイク、４・・・スピーカ、６・・・検出装置、８・・・マニピュレータ、１０・・・制御装置、１０２・・・発話取得部、１０４・・・特徴抽出部、１０８・・・選択モデル格納部、１１０・・・応答データベース、１２０・・・応答決定部、１３０・・・応答実行部、１３２・・・応答履歴格納部、１４０・・・ユーザ状態検出部、１５０・・・学習モデル選択部、１６０・・・学習モデルデータベース、２００・・・学習モデル生成装置、２１２・・・発話取得部、２１４・・・特徴抽出部、２１６・・・サンプルデータ生成部、２１８・・・ユーザ状態取得部、２２０・・・サンプルデータ分類部、２２２・・・学習モデル生成部

【図１】



【図２】



【図３】

特徴ベクトル

v_i=(v_{i1}, v_{i2}, v_{i3}, v_{i4}, v_{i5}, v_{i6}, v_{i7}, ･･･, v_{im}, ･･･)

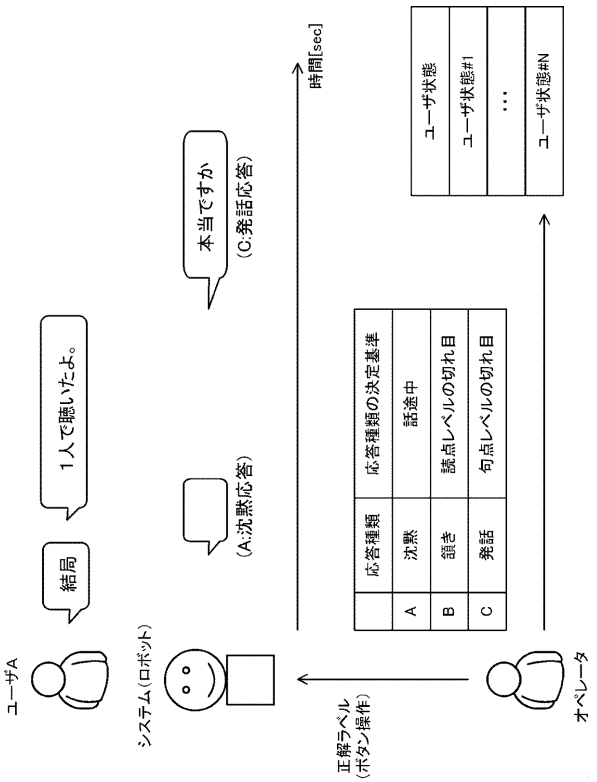
=($f0_{T1}$, V_{T1}, $f0_{T2}$, V_{T2}, f0, V, L1, ･･･,1, ･･･)

情報の種類	要素		成分値	成分
i番目のユーザ発話情報	句末T1msec	f0	$f0_{T1}$	v_{i1}
		ボリューム	V_{T1}	v_{i2}
	句末T2msec	f0	$f0_{T2}$	v_{i3}
		ボリューム	V_{T2}	v_{i4}
	発話区間全体	f0	f0	v_{i5}
		ボリューム	V	v_{i6}
	ユーザ発話長		L1	v_{i7}
	････		･･･	･･･
システム応答の履歴	直前の応答タイプ		1	v_{im}
	････		･･･	･･･

直前の応答タイプ：

1：沈黙、2：頷き、3：発話

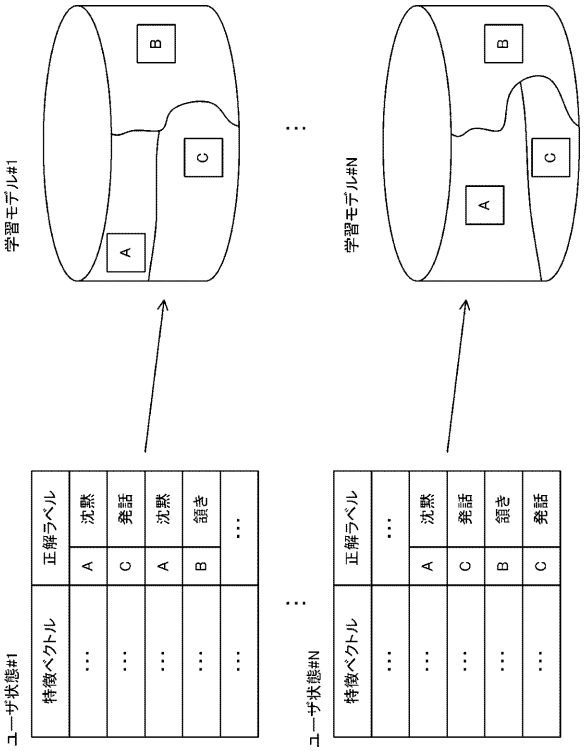
【図４】



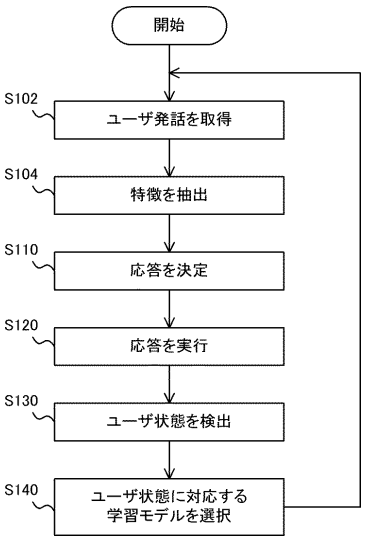
【図５】

ユーザ状態	ユーザ発話	特徴ベクトル								正解ラベル
		句末T1msec		句末T2msec			ユーザ発話長（sec）	…		
		f0	ボリューム	f0	ボリューム					
ユーザ状態#1	「結局」	…	…	…	…	…	0.5	…	A	沈黙
	「1人で聴いたよ。」	…	…	…	…	…	1.5	…	C	発話

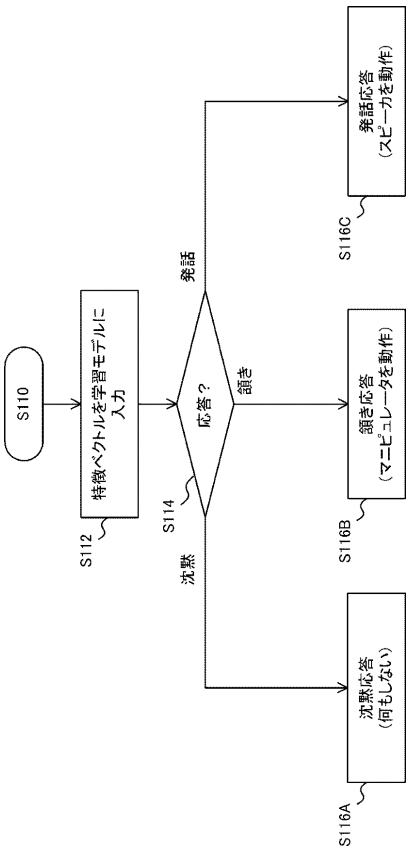
【図６】



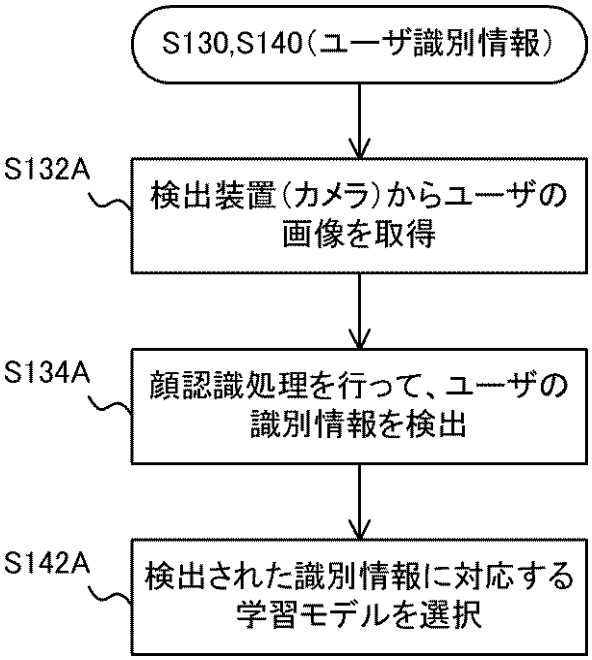
【図７】



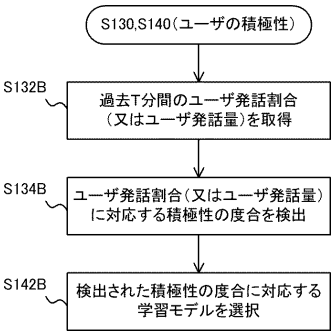
【図８】



【図９】



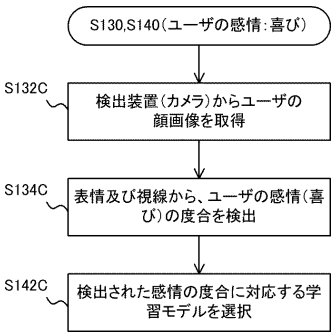
【図１０】



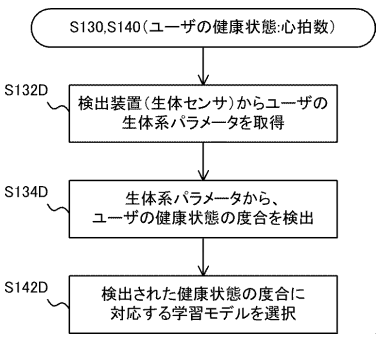
【図１１】

	積極性の度合	ユーザ発話割合
小	#1	0≦Rs＜25
	#2	25≦Rs＜50
	#3	50≦Rs＜75
大	#4	75≦Rs≦100

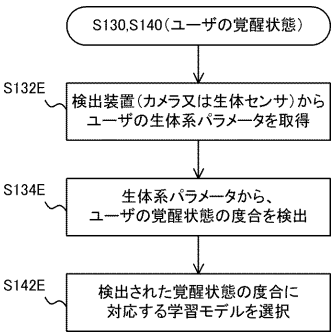
【図１２】



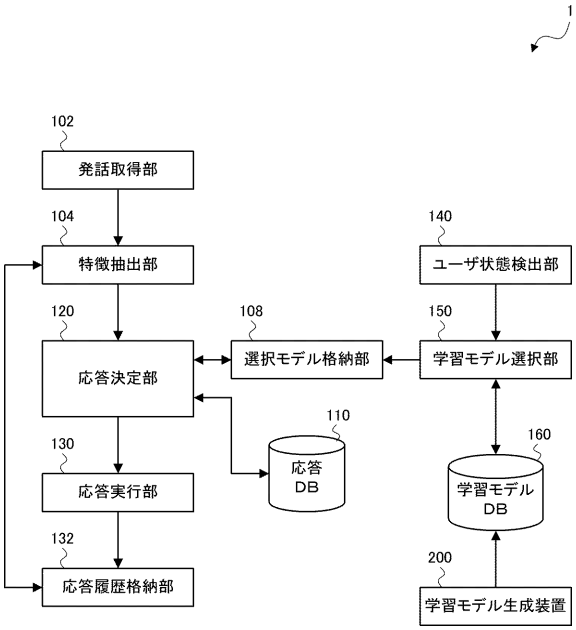
【図１３】



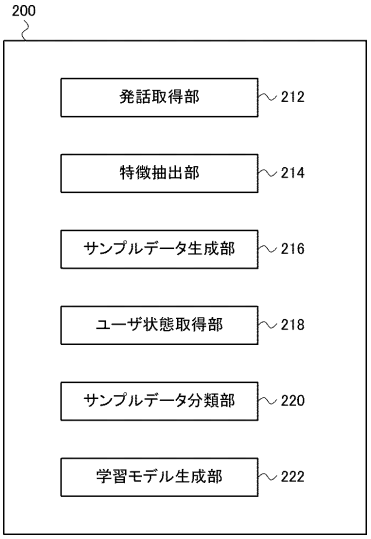
【図１４】



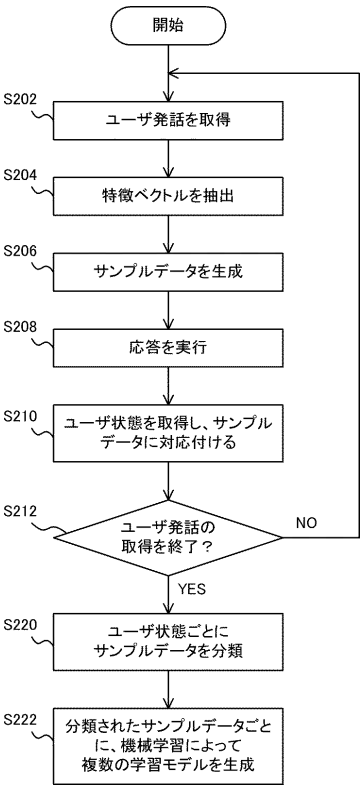
【図１５】


1
102
発話取得部
104
特徴抽出部
120
応答決定部
130
応答実行部
132
応答履歴格納部
108
選択モデル格納部
110
応答
DB
140
ユーザ状態検出部
150
学習モデル選択部
160
学習モデル
DB
200
学習モデル生成装置

【図１６】


200
発話取得部
212
特徴抽出部
214
サンプルデータ生成部
216
ユーザ状態取得部
218
サンプルデータ分類部
220
学習モデル生成部
222

【図１７】


開始
S202
ユーザ発話を取得
S204
特徴ベクトルを抽出
S206
サンプルデータを生成
S208
応答を実行
S210
ユーザ状態を取得し、サンプル
データに対応付ける
S212
ユーザ発話の
取得を終了？
NO
YES
S220
ユーザ状態ごとに
サンプルデータを分類
S222
分類されたサンプルデータごと
に、機械学習によって
複数の学習モデルを生成

フロントページの続き

(51)Int.Cl.	FI		テーマコード（参考）
	Ｇ０６Ｆ 3/16	６９０	
	Ｇ０６Ｆ 3/01	５１０	

(72)発明者　渡部　生聖
愛知県豊田市トヨタ町１番地　トヨタ自動車株式会社内

Ｆターム(参考) 5E555 AA11 AA48 AA76 BA01 BA06 BA88 BB01 BB06 BC01 BC08
BC17 CA41 CA42 CA47 CB64 CB66 CB69 CB76 DA23 EA02
EA05 EA19 EA20 EA22 EA23 EA28 FA00