

(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2018-40897
(P2018-40897A)

(43) 公開日 平成30年3月15日(2018.3.15)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10 5 0 0 Z	
G 1 0 L 15/00 (2013.01)	G 1 0 L 15/00 2 0 0 H	
G 1 0 L 13/00 (2006.01)	G 1 0 L 13/00 1 0 0 M	
G 1 0 L 13/02 (2013.01)	G 1 0 L 13/02 1 3 0 Z	
G 1 0 L 15/22 (2006.01)	G 1 0 L 15/22 3 0 0 Z	

審査請求 有 請求項の数 7 O L (全 16 頁) 最終頁に続く

(21) 出願番号	特願2016-173862 (P2016-173862)	(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
(22) 出願日	平成28年9月6日(2016.9.6)	(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	高梨 克也 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

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装置を用いた自動対話方法

(57) 【要約】

【課題】ユーザの発話から解析される韻律からでは、次の発話を対話ロボットがすべきか、ユーザの発話を待つべきか、明りょうに判断できない場合があった。このような場合は、両者の間で沈黙が続いたり、両者の発話が被ったりして、円滑な会話を楽しむ妨げとなっていた。

【解決手段】入力部が入力したユーザ発話に基づいて、次のユーザ発話を待つべきか、自身からの発話を実行すべきかを推定する推定部と、推定部が、ユーザ発話を待つべきとも、自身からの発話を実行すべきとも推定できなかった場合に、対話内容とは関係の無い繋ぎ発声であるフィラーを出力部に出力させるフィラー出力、および疑似頭部を前後させる頷き動作の少なくともいずれかを実行して応答する応答制御部とを備える音声対話装置を提供する。

【選択図】 図3

交代確率P	0.0-0.5	0.5-0.8	0.8-1.0
応答種類	応答無し	フィラー発声	応答発話

【特許請求の範囲】**【請求項 1】**

ユーザの発話である入力発話を入力する入力部と、前記ユーザへの発話である出力発話を出力する出力部とを有する音声対話装置であって、

前記入力部が入力した前記入力発話に基づいて、前記入力部による前記入力発話の入力を待つべきか、前記出力部による前記出力発話の出力を実行すべきかを推定する推定部と

、
前記推定部が、前記入力部による前記入力発話の入力を待つべきとも、前記出力部による前記出力発話の出力を実行すべきとも推定できなかった場合に、対話内容とは関係の無い繋ぎ発声であるフィラーを前記出力部に出力させるフィラー出力、および疑似頭部を前後させる頷き動作の少なくともいずれかを実行して応答する応答制御部とを備える音声対話装置。

10

【請求項 2】

前記入力部が入力した前記入力発話の韻律を解析する解析部を備え、

前記推定部は、前記解析部が解析した前記韻律に基づいて、前記入力部による前記入力発話の入力を待つべきか、前記出力部による前記出力発話の出力を実行すべきかを推定する請求項 1 に記載の音声対話装置。

【請求項 3】

前記推定部は、前記解析部が解析した前記韻律に基づいて、前記入力部による前記入力発話の入力を待つべきか、前記出力部による前記出力発話の出力を実行すべきかを推定する推定値を演算し、

20

前記応答制御部は、前記推定部が演算した前記推定値が、前記入力部による前記入力発話の入力を待つべきとも、前記出力部による前記出力発話の出力を実行すべきとも推定されない所定範囲に含まれる場合に、前記フィラー出力および前記頷き動作の少なくともいずれかを実行して応答する請求項 2 に記載の音声対話装置。

【請求項 4】

前記応答制御部は、推定値が、前記推定値を前記出力発話の出力を実行すべき確率として表した場合の前記所定範囲に前記含まれる場合であって、前記所定範囲に設定される第 1 閾値よりも小さい場合には前記頷き動作を実行して応答し、前記所定範囲に設定され、前記第 1 閾値以上の値に設定される第 2 閾値よりも大きい場合には前記フィラー出力を実行して応答する請求項 3 に記載の音声対話装置。

30

【請求項 5】

前記推定部は、前記出力部による前記出力発話の出力を実行することなく前記応答制御部が前記フィラー出力を続けて実行した場合は、前記出力発話の出力を実行すべきと推定する度合いを高くする請求項 1 から 4 のいずれか 1 項に記載の音声対話装置。

【請求項 6】

前記推定部は、前記フィラー出力の実行中に前記入力部が前記入力発話を入力したことを検知した場合は、前記入力発話の入力を待つべきと推定する度合いを高くする請求項 1 から 5 のいずれか 1 項に記載の音声対話装置。

【請求項 7】

40

ユーザの発話である入力発話を入力する入力部と、前記ユーザへの発話である出力発話を出力する出力部とを有する音声対話装置を用いた自動対話方法であって、

前記入力部が入力した前記入力発話に基づいて、前記入力部による前記入力発話の入力を待つべきか、前記出力部による前記出力発話の出力を実行すべきかを推定する推定ステップと、

前記推定ステップで、前記入力部による前記入力発話の入力を待つべきとも、前記出力部による前記出力発話の出力を実行すべきとも推定できなかった場合に、対話内容とは関係の無い繋ぎ発声であるフィラーを前記出力部に出力させるフィラー出力、および疑似頭部を前後させる頷き動作の少なくともいずれかを実行して応答する応答ステップとを含む自動対話方法。

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声対話装置および音声対話装置を用いた自動対話方法に関する。

【背景技術】

【0002】

ユーザが音声対話ロボットと日常会話を楽しむための技術が普及しつつある。例えば、違和感のない会話を実現するための技術として、ユーザの韻律を解析することにより、対話ロボットが適当なタイミングで相槌するものが知られている（特許文献1を参照）。また、ユーザとロボットの間で話者が交代するタイミングを模索する研究もなされている（非特許文献1を参照）。

10

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2009-3040号公報

【非特許文献】

【0004】

【非特許文献1】認知科学 研究論文 Vol.7(2000)No.1 93頁~106頁

【発明の概要】

20

【発明が解決しようとする課題】

【0005】

ユーザの発話から解析される韻律からでは、次の発話に対話ロボットがすべきか、ユーザの発話を待つべきか、明りょうに判断できない場合があった。このような場合は、両者の間で沈黙が続いたり、両者の発話が被ったりして、円滑な会話を楽しむ妨げとなっていた。

【0006】

本発明は、このような問題を解決するためになされたものであり、対話ロボットを一例とする音声対話システムとユーザとの間で、より円滑な会話を実現する技術を提供するものである。

30

【課題を解決するための手段】

【0007】

本発明の第1の態様における音声対話装置は、ユーザの発話である入力発話を入力する入力部と、ユーザへの発話である出力発話を入力する出力部とを有する音声対話装置であって、入力部が入力した前記入力発話に基づいて、入力部による入力発話の入力を待つべきか、出力部による出力発話の出力を実行すべきかを推定する推定部と、推定部が、入力部による入力発話の入力を待つべきとも、出力部による出力発話の出力を実行すべきとも推定できなかった場合に、対話内容とは関係の無い繋ぎ発声であるフィラーを出力部に出力させるフィラー出力、および疑似頭部を前後させる頷き動作の少なくともいずれかを実行して応答する応答制御部とを備える。

40

【0008】

本発明の第2の態様における自動対話方法は、ユーザの発話である入力発話を入力する入力部と、ユーザへの発話である出力発話を入力する出力部とを有する音声対話装置を用いた自動対話方法であって、入力部が入力した前記入力発話に基づいて、入力部による入力発話の入力を待つべきか、出力部による出力発話の出力を実行すべきかを推定する推定ステップと、推定ステップで、入力部による入力発話の入力を待つべきとも、出力部による出力発話の出力を実行すべきとも推定できなかった場合に、対話内容とは関係の無い繋ぎ発声であるフィラーを出力部に出力させるフィラー出力、および疑似頭部を前後させる頷き動作の少なくともいずれかを実行して応答する応答ステップとを含む。

【0009】

50

以上の各実施態様の構成により、フィラーの発声により音声対話装置側に発話権があることを仄めかしたり、頷き動作によりユーザの発話を促したりできるので、両者の間で沈黙が続いたり、両者の発話が被ったりすることを抑制できる。

【発明の効果】

【0010】

本発明により、音声対話システムとユーザとの間で、より円滑な会話を実現することができる。

【図面の簡単な説明】

【0011】

【図1】ユーザが本実施形態にかかる対話ロボットと会話する様子を示す図である。 10

【図2】第1の実施例における、対話ロボットのシステム構成を示す図である。

【図3】第1の実施例における、応答決定テーブルを示す図である。

【図4】第1の実施例における、対話処理フローを示す図である。

【図5】第2の実施例における、応答決定テーブルを示す図である。

【図6】第2の実施例における、対話処理フローの特徴部を示す図である。

【図7】第3の実施例における、対話ロボットのシステム構成を示す図である。

【図8】第3の実施例における、応答決定テーブルを示す図である。

【図9】第3の実施例における、対話処理フローの特徴部を示す図である。

【図10】第4の実施例における、応答決定テーブルを示す図である。

【図11】第5の実施例における、対話処理フローの特徴部を示す図である。 20

【図12】第6の実施例における、対話処理フローの特徴部を示す図である。

【図13】他の実施態様にかかるスマートフォンの外観を示す図である。

【発明を実施するための形態】

【0012】

以下、発明の実施の形態を通じて本発明を説明するが、特許請求の範囲に係る発明を以下の実施形態に限定するものではない。また、実施形態で説明する構成の全てが課題を解決するための手段として必須であるとは限らない。

【0013】

図1は、ユーザが本実施形態にかかる対話ロボット100と会話する様子を示す図である。対話ロボット100は、音声対話装置の一態様であり、ユーザの発話に応じた返答を行う。対話ロボット100は、例えば人間のような生物を模した筐体を有することで、ユーザに与える無機質感を和らげている。図示するように、対話ロボット100は、手足を含む体部120の上に擬似的な頭部110が配された外観を有する。そして、頭部110には、耳が位置する側頭部辺りにマイク111が配されており、マイク111は、ユーザの発話を集音する。また、擬似的な口にはスピーカ112が配されており、スピーカ112は、対話ロボット100が生成した応答を発話、発声する。なお、以下の説明においては、対話内容として意味を有する音声を「発話」とし、対話内容とは関係の無い繋ぎの音声を「発声」とする。これらを特に区別する必要が無い場合には「発話」と表現する場合がある。 30

【0014】

図2は、本実施形態の第1の実施例における、対話ロボット100のシステム構成を示す図である。対話ロボット100は、主な機能要素として、発話入力部210、解析部220、推定部230、応答制御部240、出力生成部250、発話出力部260を備える。 40

【0015】

発話入力部210は、マイク111を含み、ユーザの発話を集音してデジタル信号に変換し、解析部220へ引き渡す。解析部220は、デジタル信号として受け取ったユーザ発話の韻律を解析する。韻律は、発話を音波として捉えた場合の、その物理的特徴を表すものである。

【0016】

韻律の解析は、ユーザが発する一塊の音波において、例えば基本周波数やその変動帯域、振幅の最大値および平均値を特徴量として演算し、数値化する処理である。解析部 220 は、受け取ったデジタル信号を、例えば 25 msec のウィンドウ幅で切り取り、それを一塊の音波としてそれぞれの特徴量を演算する。そして、例えば 10 msec の遅延時間でウィンドウを順次シフトさせ、シフトさせるごとに特徴量を演算し、演算した特徴量を推定部 230 へ逐次引き渡す。

【0017】

また、解析部 220 は、ユーザの発話の節目を韻律の変化から予測して、節目のタイミングを応答制御部 240 へ出力する。発話の節目は、一連の会話の区切りであり、解析部 220 は、振幅レベルの低下や、語尾の抑揚によって節目を予測する。

10

【0018】

推定部 230 は、受け取った物理的特徴量としての韻律に基づいて、ユーザからの発話を待つべきか、自身が発話すべきかを推定する。本実施形態において具体的には、推定部 230 は、話者がユーザから対話ロボット 100 へ移る確率である交代確率 P を推定値として算出する。交代確率 $P = 0$ は、発話権が継続してユーザに有ることを意味し、交代確率 $P = 1$ は、発話権がユーザから対話ロボット 100 へ移ることを意味する。交代確率 P は、0 以上 1 以下の値として算出される。したがって、交代確率 P が 0 に近いほど発話権が継続してユーザに有ると強く推定され、1 に近いほど発話権がユーザから対話ロボット 100 へ移ると強く推定されると言える。

【0019】

推定部 230 は、韻律 DB 310 に接続されている。韻律 DB 310 は、例えばハードディスクドライブの記録媒体であり、韻律を表現するそれぞれの特徴量が、どのような数値である場合に交代確率に対してどれくらいの影響を与えるかを表現する、ルックアップテーブルや評価関数が格納されたデータベースである。推定部 230 は、解析部 220 から受け取った韻律のそれぞれの特徴量を、韻律 DB 310 のルックアップテーブルや評価関数に当てはめてそれぞれの評価値に換算し、これらの評価値から多変量解析などの手法を用いて交代確率 P を算出する。具体的手法については、非引用文献 1 などに詳しい。

20

【0020】

推定部 230 は、算出した交代確率 P を応答制御部 240 へ引き渡す。応答制御部 240 は、解析部 220 からユーザ発話の節目のタイミングを知らされた場合に、交代確率 P を参照して、いかなる応答を実行するかを決定する。そして、決定した応答を実行するための具体的なデータを、出力生成部 250 から取得する。

30

【0021】

本実施形態の第 1 の実施例では、応答制御部 240 は、交代確率 P の値に応じて「何も応答しない」「フィルター発声を行う」「応答発話を行う」の 3 つの中から決定する。「応答しない」は、特に何ら反応を示さないという応答であるが、形式的には無発話時間を設定するなどのために出力生成部 250 を参照する。

【0022】

応答制御部 240 は、「フィルター発声を行う」ことを決定した場合は、出力生成部 250 の一機能として実装されるフィルター生成部 251 にフィルターを生成させる。フィルターは、対話内容とは関係の無い繋ぎ発声である。例えば、「あー」「あのー」「えー」「えーと」「そうですねー」などが挙げられる。これらのフィルターは、再生可能な発声データとして、フィルター DB 320 に格納されている。

40

【0023】

フィルター DB 320 は、フィルター生成部 251 に接続されているデータベースであり、例えばハードディスクドライブの記録媒体によって構成されている。フィルター生成部 251 は、応答制御部 240 からの要求に従ってフィルター DB 320 に格納された発声データを例えばランダムに抽出することによりフィルターを生成し、応答制御部 240 へ引き渡す。

50

【 0 0 2 4 】

応答制御部 2 4 0 は、「応答発話を行う」ことを決定した場合は、出力生成部 2 5 0 の一機能として実装される応答生成部 2 5 2 に応答発話を生成させる。応答発話は、ユーザ発話の内容に対して意味を有する対話として応答する発話である。

【 0 0 2 5 】

発話 DB 3 3 0 は、応答生成部 2 5 2 に接続されているデータベースであり、例えばハードディスクドライブの記録媒体によって構成されている。発話 DB 3 3 0 は、例えばコーパスとして体系化されており、個々の用語は、再生可能な発話データを伴って格納されている。応答生成部 2 5 2 は、応答制御部 2 4 0 からの要求に従って発話 DB 3 3 0 から発話データを収集し、会話として成立するセンテンスを構築して、応答制御部 2 4 0 へ引き渡す。

10

【 0 0 2 6 】

応答生成部 2 5 2 は、出力生成部 2 5 0 に生成させたフィルア発声データ、応答発話データを発話出力部 2 6 0 へ引き渡す。発話出力部 2 6 0 は、スピーカ 1 1 2 を含み、応答生成部 2 5 2 から受け取ったフィルア発声データ、応答発話データをアナログ信号に変換して、スピーカ 1 1 2 から音として出力する。

【 0 0 2 7 】

図 3 は、第 1 の実施例における、応答決定テーブルを示す図である。図示するように、応答制御部 2 4 0 は、交代確率 P が 0 . 0 以上 0 . 5 未満である場合に、「何も応答しない」と決定する。同様に、交代確率 P が 0 . 5 以上 0 . 8 未満である場合に、「フィルア発話を行う」と決定し、交代確率 P が 0 . 8 以上 1 . 0 以下である場合に、「応答発話を行う」と決定する。このように、応答制御部 2 4 0 がいかなる応答を行うかは、交代確率 P の値に応じた分類ルールとして、予め定められている。

20

【 0 0 2 8 】

図 3 に示す分類ルールは、発話入力部 2 1 0 による入力発話の入力、すなわちユーザが発話することを待つべきとする交代確率 P の範囲を、0 . 0 以上 0 . 5 未満と定めている。同様に、発話出力部 2 6 0 による出力発話の出力、すなわち対話ロボット 1 0 0 自身が意味内容を有する発話をすべきとする交代確率 P の範囲を 0 . 8 以上 1 . 0 以下と定めている。そして、その間の 0 . 5 以上 0 . 8 未満の範囲を、発話入力部 2 1 0 による入力発話の入力を待つべきとも、発話出力部 2 6 0 による出力発話の出力を実行すべきとも推定できない範囲として定めている。

30

【 0 0 2 9 】

本実施例においては、対話ロボット 1 0 0 が入力発話の入力を待つべきとも出力発話の出力を実行すべきとも推定できない場合に、ユーザとの間で沈黙が続いたり、ユーザの発話と被ったりすることを抑制するために、フィルア発話を行う。フィルア発話には、発話した話者側に発話権があることを相手に仄めかす効果があると知られている。

【 0 0 3 0 】

図 4 は、第 1 の実施例における、対話処理フローを示す図である。図示する対話処理フローは、例えばスイッチが操作されるなどの対話開始の検知に伴って開始される。

【 0 0 3 1 】

応答制御部 2 4 0 は、ステップ S 1 0 1 で、ユーザが発話を開始したか否かを判断する。具体的には、例えば、発話入力部 2 1 0 の AD 変換出力を監視して判断する。ユーザが発話を開始したと判断したら、ステップ S 1 0 2 へ進む。

40

【 0 0 3 2 】

ステップ S 1 0 2 で、発話入力部 2 1 0 は、ユーザの発話を入力してデジタル信号に変換し、解析部 2 2 0 へ引き渡す。ステップ S 1 0 3 で、解析部 2 2 0 は、発話入力部 2 1 0 から受け取ったデジタル信号から韻律を解析し、演算した特徴量を推定部 2 3 0 へ引き渡す。ステップ S 1 0 4 で、推定部 2 3 0 は、解析部 2 2 0 から受け取った韻律の特徴量を用いて、交代確率 P を算出し、応答制御部 2 4 0 へ引き渡す。

【 0 0 3 3 】

50

応答制御部 240 は、ステップ S105 で、ユーザ発話に節目があるか否かを判断する。具体的には、解析部 220 から節目のタイミング情報を受け取った場合に、そのタイミング情報が示す時点を節目と判断する。節目がないと判断した場合は、ステップ S102 へ戻り、ユーザ発話をさらに継続して受け入れる。節目があると判断した場合は、ステップ S106 へ進む。

【0034】

応答制御部 240 は、ステップ S106 で、対話が終了したか否かを判断する。ユーザが、例えば、「さようなら」と発話した場合や、終了ボタンを操作した場合は、対話が終了したと判断し、一連の処理を終了させる。終了と判断できない場合は、対話を継続すべく、ステップ S107 へ進む。

10

【0035】

ステップ S107 からの処理は、応答選択の処理である。応答制御部 240 は、ステップ S107 で、交代確率 P が 0.0 以上 0.5 未満であるか否かを判断する。この範囲に含まれると判断したら、ステップ S108 へ進み、ユーザの次の発話を待つ。例えば、予め定められた時間の間をおく。そして、その間にユーザが発話を開始したかを判断すべくステップ S101 へ戻り、再び一連の処理を実行する。

【0036】

応答制御部 240 は、ステップ S107 で交代確率 P が 0.0 以上 0.5 未満に含まれないと判断した場合は、ステップ S109 へ進み、交代確率 P が 0.5 以上 0.8 未満であるか否かを判断する。この範囲に含まれると判断したら、ステップ S110 へ進み、ファイラ発声を実行する。ファイラ発声を終了したらステップ S101 へ戻り、一定時間内にユーザが発話を開始したかを判断して、再び一連の処理を実行する。

20

【0037】

応答制御部 240 は、ステップ S109 で交代確率 P が 0.5 以上 0.8 未満に含まれないと判断した場合、および、ステップ S101 で一定時間内にユーザが発話を開始しなかったと判断された場合は、ステップ S111 へ進み、応答発話データを生成して、発話の出力を実行する。発話の出力が終了したらステップ S101 へ戻り、一定時間内にユーザが発話を開始したかを判断して、再び一連の処理を実行する。

【0038】

以上のように、入力発話の入力を待つべきとも出力発話の出力を実行すべきとも推定できない場合にファイラ発声を実行することにより、違和感を与える間を与えることがない円滑な対話を促進することが期待できる。

30

【0039】

次に、本実施形態にかかる第 2 の実施例を説明する。対話ロボット 100 としての装置構成は、第 1 の実施例と同様であり、フローの処理が第 1 の実施例と若干異なる。

【0040】

図 5 は、第 2 の実施例における、応答決定テーブルを示す図である。第 1 の実施例では、「ファイラ発声を行う」と判断される交代確率 P の範囲と、「応答発話を行う」と判断される交代確率 P の範囲とは、互いに重なり合うことがなかった。第 2 の実施例では、両者に重なる範囲を設ける。具体的には、交代確率 P が 0.5 以上 0.8 未満である場合に、「ファイラ発声を行う」と決定し、交代確率 P が 0.7 以上 1.0 以下である場合に、「応答発話を行う」と決定する。交代確率 P が 0.0 以上 0.5 未満である場合に、「何も応答しない」と決定することは、第 1 の実施例と同様である。

40

【0041】

すなわち、交代確率 P が 0.7 以上 0.8 未満である場合は、「ファイラ発声を行う」と「応答発話を行う」の、2 つの分類に属するものとする。これは、交代確率 P が 0.7 以上 0.8 未満である場合の対話の状況が、ファイラ発声を行うことが良いと判断される場合と、応答発話を行うことが良いと判断される場合とが混在する状況であると、実験的に明らかになってきたことに対応する処置である。

【0042】

50

このように応答決定テーブルを定めた場合の処理について説明する。図6は、第2の実施例における、対話処理フローの特徴部を示す図である。第2の実施例における対話処理フローは、図4を用いて説明した処理とは応答選択の処理の部分が異なり、他の部分では同じであるので、異なる部分を説明する。また、図4と同じ処理については同じステップ番号を付して、具体的な説明については省略する。

【0043】

交代確率Pが0.0以上0.5未満の処理であるステップS107とS108は、第1の実施例における処理と同様である。交代確率Pが0.0以上0.5未満の範囲に含まれる場合は、応答制御部240は、ユーザ発話を待って(ステップS108)、ステップS101へ戻る。

10

【0044】

ステップS201において、応答制御部240は、交代確率Pが0.5以上0.7未満であるか否かを判断する。交代確率Pがこの範囲に含まれる場合は、「フィラー発声を行う」のみに属するので、応答制御部240は、フィラー発声を実行して(ステップS110)、ステップS101へ戻る。

【0045】

ステップS201において、応答制御部240は、交代確率Pが0.7以上0.8未満であるか否かを判断する。交代確率Pがこの範囲に含まれる場合は、ステップS203へ進み、応答制御部240は、「フィラー発声のみを行う」、「応答発話のみを行う」、「フィラー発声と応答発話の両方を行う」の3つの中からランダムに1つを選択する。そして、「フィラー発声のみを行う」を選択した場合には、ステップS204(YES)を経由してステップS110へ進む。「応答発話のみを行う」を選択した場合には、ステップS204(NO)、S205(YES)を経由してステップS111へ進む。ステップS111で、応答制御部240は、応答発話データを生成して、発話の出力を実行する。発話の出力が終了したらステップS101へ戻る。

20

【0046】

応答制御部240は、「フィラー発声と応答発話の両方を行う」を選択した場合には、ステップS204(NO)、S205(NO)を経由してステップS206へ進む。応答制御部240は、ステップS206でフィラー発声を実行し、続いてステップS207で応答発話の出力を実行して、ステップS101へ戻る。

30

【0047】

以上のように本実施例によれば、対話の状況が判然としない場合に応答をランダムに変化させることで、機械的との印象を与えない、円滑な対話の実現を期待できる。また、フィラー発声に続けて応答発話の出力を実行した場合に、ユーザの発話と被ってしまうような状況においても、ユーザの発話と被るのはまずはフィラーであるので、対話に与える悪影響は相対的に小さいと言える。

【0048】

次に、本実施形態にかかる第3の実施例を説明する。第1の実施例および第2の実施例で説明したフィラー発声は、発声した話者側に発話権があることを相手に仄めかす効果を期待できるので、交代確率Pが0.5以上の場合に適用することが好ましかった。しかし、対話ロボット100が入力発話の入力を待つべきとも出力発話の出力を実行すべきとも推定できないとする交代確率Pの範囲は、0.5未満の範囲を含みうる。交代確率Pが0.5未満である場合は、ユーザに発話を促すことが好ましいとも考えられる。

40

【0049】

そこで、第3の実施例においては、ユーザに発話を促す処理を実行するために、「領き動作」と「相槌」を導入する。図7は、第3の実施例における、対話ロボット100のシステム構成を示す図である。図2で示したシステム構成と同様の構成には同じ符番を付して、その説明を省略する。

【0050】

図2のシステム構成と異なる点は、応答制御部240に接続された領き動作部270と

50

、出力生成部 250 に含まれる相槌生成部 253 と、相槌生成部 253 に接続された相槌 DB 340 を更に備える点である。頷き動作部 270 は、モータを含み、これを駆動することにより頭部 110 を前後に揺動させ、ユーザに対話ロボット 100 が頷いたように見せる動作を実行する。

【0051】

応答制御部 240 は、第 3 の実施例においては、交代確率 P の値に応じて「何も応答しない」「フィルター発声を行う」「応答発話を行う」に加えて、「相槌を発声する」も選択しうる。相槌は、ユーザが発話を継続しやすいように、次の発話を促す発声である。例えば、「うん」「そうだね」「それで？」などが挙げられる。これらの相槌は、再生可能な発声データとして、相槌 DB 340 に格納されている。

10

【0052】

相槌 DB 340 は、例えばハードディスクドライブの記録媒体によって構成されている。相槌生成部 253 は、応答制御部 240 からの要求に従って相槌 DB 340 に格納された発声データを例えばランダムに抽出することにより相槌を生成し、応答制御部 240 へ引き渡す。

【0053】

図 8 は、第 3 の実施例における、応答決定テーブルを示す図である。第 1 の実施例では、交代確率 P が 0.0 以上 0.5 未満である場合に「何も応答しない」との 1 つの分類を割り当てたが、第 3 の実施例では、これを 2 つの分類に分ける。具体的には、図示するように、交代確率 P が 0.0 以上 0.4 未満である場合に、「何も応答しない」と決定し、交代確率 P が 0.4 以上 0.5 未満である場合に、「相槌を発声する」と決定する。交代確率 P が 0.5 以上 0.8 未満である場合に、「フィルター発声を行う」と決定し、交代確率 P が 0.8 以上 1.0 以下である場合に、「応答発話を行う」と決定することは、第 1 の実施例と同様である。

20

【0054】

第 3 の実施例においては、発話入力部 210 による入力発話の入力を待つべきとも、発話出力部 260 による出力発話の出力を実行すべきとも推定できない範囲（第 1 範囲）を 0.4 以上 0.8 以下と定めている。また、ユーザが発話することを待つべきとする範囲（第 2 範囲）を 0.0 以上 0.4 未満と定め、対話ロボット 100 自身が発話をすべきとする範囲（第 3 範囲）を 0.8 以上 1.0 以下と定めている。その上で、第 1 範囲を、第 2 範囲に近い範囲（0.4 以上 0.5 未満）と第 3 範囲に近い範囲（0.5 以上 0.8 未満）とに分けている。そして、交代確率 P が、第 1 範囲に含まれるものの、第 2 範囲に近い範囲の値である場合に「相槌を発声する」を実行し、第 1 範囲に含まれるものの、第 3 範囲に近い範囲の値である場合に「フィルター発声する」を実行する。「相槌を発声する」は、「頷き動作を行う」を伴っても良いし、「頷き動作を行う」に代えても良い。

30

【0055】

このように応答決定テーブルを定めた場合の処理について説明する。図 9 は、第 3 の実施例における、対話処理フローの特徴部を示す図である。第 3 の実施例における対話処理フローは、図 4 を用いて説明した処理とは応答選択の処理の部分が異なり、他の部分では同じであるので、異なる部分を説明する。また、図 4 と同じ処理については同じステップ番号を付して、具体的な説明については省略する。

40

【0056】

応答制御部 240 は、まずステップ S 301 で、交代確率 P が 0.0 以上 0.4 未満であるか否かを判断する。交代確率 P がこの範囲に含まれる場合は、ユーザ発話を待つ（ステップ S 108）、ステップ S 101 へ戻る。

【0057】

応答制御部 240 は、ステップ S 302 で、交代確率 P が 0.4 以上 0.5 未満であるか否かを判断する。交代確率 P がこの範囲に含まれる場合は、ステップ S 303 へ進み、応答制御部 240 は、「相槌を発声する」と共に頷き動作を実行する。これらの処理が終了したらステップ S 101 へ戻る。交代確率 P が 0.5 以上 0.8 未満、および 0.8 以

50

上 1.0 以下の処理は、第 1 の実施例における処理と同様である。

【0058】

以上のように本実施例によれば、対話ロボット 100 が入力発話の入力を待つべきとも出力発話の出力を実行すべきとも推定できない場合であって、ユーザに発話を促すことが好ましいとも考えられるような状況においても、相槌の発声と頷き動作により、円滑な対話の実現を期待できる。

【0059】

次に第 4 の実施例について説明する。図 10 は、第 4 の実施例における、応答決定テーブルを示す図である。第 4 の実施例では、第 2 の実施例のように「フィラー発声を行う」と判断される交代確率 P の範囲と、「応答発話を行う」と判断される交代確率 P の範囲を、部分的に重複させ、更に、「何も応答しない」と判断される交代確率 P の範囲と、「相槌を発声する」と判断される交代確率 P の範囲を、部分的に重複させる。具体的には、交代確率 P が 0.0 以上 0.4 未満である場合に、「何も応答しない」と決定し、交代確率 P が 0.3 以上 0.5 以下である場合に、「相槌を発声する」と決定する。交代確率 P が 0.5 以上 0.8 未満である場合に、「フィラー発声を行う」と決定し、交代確率 P が 0.7 以上 1.0 以下である場合に、「応答発話を行う」と決定することは、第 2 の実施例と同様である。

10

【0060】

すなわち、交代確率 P が 0.3 以上 0.4 未満である場合は、「何も応答しない」と「相槌を発声する」の、2 つの分類に属するものとする。これは、交代確率 P が 0.3 以上 0.4 未満である場合の対話の状況が、何も応答しないことが良いと判断される場合と、相槌を発声することが良いと判断される場合とが混在する状況であると、実験的に明らかになってきたことに対応する処置である。交代確率 P が 0.3 以上 0.4 未満である場合は、応答制御部 240 は、「何も応答しない」と「相槌を発声する」のいずれかをランダムに選択して実行する。「相槌を発声する」を選択した場合の処理は、図 9 のステップ S303 と同様である。

20

【0061】

なお、「相槌を発声する」は、ユーザに発話を促すことを目的とし、「フィラー発声を行う」は、発声した話者側に発話権があることを相手に仄めかすことを目的とするので、「相槌を発声する」と判断される交代確率 P の範囲と、「フィラー発声を行う」と判断される交代確率 P の範囲は、重複させないことが好ましい。以上のように本実施例によれば、対話の状況が判然としない場合に、交代確率 P の値を用いて応答に様々な変化を発生させることで、機械的との印象を与えない、円滑な対話の実現を期待できる。

30

【0062】

次に第 5 の実施例について説明する。第 5 の実施例は、対話ロボット 100 としての装置構成は、第 1 の実施例と同様であり、応答決定テーブルにおける交代確率 P の範囲を動的に変更する点で第 1 の実施例と異なる。第 1 の実施例における応答決定テーブルは、図 3 に示す通りであるが、本実施例における応答決定テーブルでは、「フィラー発声を行う」と「応答発話を行う」との境界値である 0.8 を可変値 X_p とする。なお、初期値は、0.8 とする。

40

【0063】

図 11 は、第 5 の実施例における、対話処理フローの特徴部を示す図である。第 5 の実施例における対話処理フローは、図 4 を用いて説明した処理とは応答選択の処理の部分が異なり、他の部分では同じであるので、異なる部分を説明する。また、図 4 と同じ処理については同じステップ番号を付して、具体的な説明については省略する。

【0064】

交代確率 P が 0.0 以上 0.5 未満の処理であるステップ S107 と S108 は、第 1 の実施例における処理と同様である。交代確率 P が 0.0 以上 0.5 未満の範囲に含まれる場合は、応答制御部 240 は、ユーザ発話を待つ（ステップ S108）。そして、ステップ S402 へ進み、 X_p に初期値である X_{p_0} を代入する。ここでは、 X_{p_0} を上述の

50

ように 0.8 とする。その後、ステップ S 1 0 1 へ戻る。なお、ステップ S 1 0 8 とステップ S 4 0 2 は、順序が逆であっても良い。

【 0 0 6 5 】

応答制御部 2 4 0 は、ステップ S 1 0 7 で交代確率 P が 0.0 以上 0.5 未満に含まれないと判断した場合は、ステップ S 4 0 1 へ進み、交代確率 P が 0.5 以上 X p 未満であるか否かを判断する。この範囲に含まれると判断したら、ステップ S 1 1 0 へ進み、フィルア発声を実行する。フィルア発声を終了したらステップ S 4 0 3 へ進み、現在の X p の値から 0.1 減らし、新たな X p の値に更新する。そして、ステップ S 1 0 1 へ戻る。なお、ステップ S 1 1 0 とステップ S 4 0 3 は、順序が逆であっても良い。

【 0 0 6 6 】

応答制御部 2 4 0 は、ステップ S 4 0 1 で交代確率 P が 0.5 以上 X p 未満に含まれないと判断した場合は、ステップ S 1 1 1 へ進み、発話の出力を実行する。発話の出力が終了したらステップ S 4 0 4 へ進み、X p に初期値である X p₀ を代入する。その後、ステップ S 1 0 1 へ戻る。なお、ステップ S 1 1 1 とステップ S 4 0 4 は、順序が逆であっても良い。

【 0 0 6 7 】

このような処理フローによれば、処理がループされてフィルア発声を行うたびに、ステップ S 1 1 0 へ導かれる条件である交代確率 P の範囲が狭くなり、ステップ S 1 1 1 の発話出力が実行される可能性がより大きくなる。換言すれば、推定部 2 3 0 は、ユーザの発話や対話ロボット 1 0 0 の応答対話が行われることなくフィルア発声を続けた場合は、発話出力を実行すべきと推定する度合いを高くしていると言える。これにより、対話においてフィルアばかりが発声されるような、違和感のある状況を回避することが期待できる。なお、X p を更新した結果 0.5 以下になる場合は、下限値として 0.5 に固定する。この場合、フィルア発声が事実上禁止されることになる。また、減ずる値は 0.1 に限らず、他の値に調整しても良い。また、交代確率 P の範囲を調整するのではなく、フィルア発声を行った回数をカウントする変数を設定し、これが予め定められた回数に達した場合に、フィルア発声を禁止するように構成しても良い。

【 0 0 6 8 】

次に第 6 の実施例について説明する。第 6 の実施例は、対話ロボット 1 0 0 としての装置構成は、第 1 の実施例と同様であり、応答決定テーブルにおける交代確率 P の範囲を動的に変更する点で第 1 の実施例と異なる。第 1 の実施例における応答決定テーブルは、図 3 に示す通りであるが、本実施例における応答決定テーブルでは、「何も応答しない」と「フィルア発声を行う」との境界値である 0.5 を可変値 X q とする。なお、初期値は、0.5 とする。

【 0 0 6 9 】

図 1 2 は、第 6 の実施例における、対話処理フローの特徴部を示す図である。第 6 の実施例における対話処理フローは、図 4 を用いて説明した処理とは応答選択の処理の部分と異なり、他の部分では同じであるので、異なる部分を説明する。また、図 4 と同じ処理については同じステップ番号を付して、具体的な説明については省略する。

【 0 0 7 0 】

応答制御部 2 4 0 は、ステップ S 5 0 1 で交代確率 P が 0.0 以上 X q 未満であるか否かを判断する。交代確率 P が 0.0 以上 X q 未満の範囲に含まれる場合は、応答制御部 2 4 0 は、ユーザ発話を待つ（ステップ S 1 0 8 ）。そして、ステップ S 5 0 5 へ進み、X q に初期値である X q₀ を代入する。ここでは、X q₀ を上述のように 0.5 とする。その後、ステップ S 1 0 1 へ戻る。なお、ステップ S 1 0 8 とステップ S 5 0 5 は、順序が逆であっても良い。

【 0 0 7 1 】

応答制御部 2 4 0 は、ステップ S 5 0 1 で交代確率 P が 0.0 以上 X q 未満に含まれないと判断した場合は、ステップ S 5 0 2 へ進み、交代確率 P が X q 以上 0.8 未満であるか否かを判断する。この範囲に含まれると判断したら、ステップ S 1 1 0 へ進み、フィル

10

20

30

40

50

ー発声を実行する。応答制御部240は、フィルラ発声を行っている間にユーザ発話と被ったか否かを判断する(ステップS503)。具体的には、例えば、発話入力部210のAD変換出力を監視して判断する。被ったと判断したらステップS504へ進み、ステップS504で、現在のXqの値に0.1を加え、新たなXqの値に更新する。そして、ステップS101へ戻る。被っていないと判断したらステップS505へ進み、Xqに初期値であるXq₀を代入する。

【0072】

応答制御部240は、ステップS502で交代確率PがXq以上0.8未満に含まれないと判断した場合は、ステップS111へ進み、発話の出力を実行する。発話の出力が終了したらステップS505へ進み、Xqに初期値であるXq₀を代入する。その後、ステップS101へ戻る。なお、ステップS111とステップS505は、順序が逆であっても良い。

10

【0073】

このような処理フローによれば、処理がループされてフィルラ発声とユーザ発話が被るたびに、ステップS110へ導かれる条件である交代確率Pの範囲が狭くなり、ステップS108の処理である「何も応答しない」が実行される可能性がより大きくなる。換言すれば、フィルラ発声中にユーザの発話を検知した場合は、ユーザの発話を待つべきと推定する度合いを高くしている。すなわち、対話においてフィルラ発声とユーザ発話が被ることが続く状況は、互いに発話権を主張し合うような状況であるので、これを回避すべく、対話ロボット100は、ユーザがより発話しやすいように「何も応答しない」の割合を大きくする。このような処理により、対話が詰まるような状況を回避することが期待できる。なお、Xqを更新した結果0.8以上になる場合は、上限値として0.8に固定する。この場合、フィルラ発声が事実上禁止されることになる。また、加える値は0.1に限らず、他の値に調整しても良い。また、交代確率Pの範囲を調整するのではなく、フィルラ発声とユーザ発話が被った回数をカウントする変数を設定し、これが予め定められた回数に達した場合に、フィルラ発声を禁止するように構成しても良い。

20

【0074】

以上に説明した第5の実施例と第6の実施例は、それぞれ第1の実施例をベースとして説明したが、適宜数値を調整して第2から第4の実施例と組み合わせても良い。また、第5の実施例と第6の実施例を組み合わせても良い。

30

【0075】

また、第3の実施例と第4の実施例においては、相槌発声と頷き動作を共に実行するものとして説明したが、いずれか一方の動作を実行させるように構成しても良い。頷き動作の処理を省くのであれば、頷き動作部270を除いても良い。また、第3の実施例と第4の実施例においては、応答決定テーブルに「相槌を発声する」交代確率Pの範囲と「フィルラ発声を行う」交代確率Pの範囲をそれぞれ設けたが、「フィルラ発声を行う」交代確率Pの範囲を省いても良い。この場合は、「応答発話を行う」交代確率Pの範囲を、0.5以上1.0以下とすれば良い。

【0076】

以上に説明した実施態様においては、音声対話装置として対話ロボット100を説明した。しかし、音声対話装置は、人型を模したような立体構造を有していなくても良く、様々な実施態様を採用し得る。図13は、他の実施態様にかかるスマートフォン700の外観を示す図である。

40

【0077】

スマートフォン700は、スマートフォンとしての通常の機能に加え、音声対話装置としての機能も発揮する。スマートフォン700は、表示部710、マイク711、およびスピーカ712を備える。マイク111は、対話ロボット100のマイク111と同様に機能し、スピーカ712は、対話ロボット100のスピーカ112と同様に機能する。また、スマートフォン700は、図2に示すシステム構成を内蔵する。

【0078】

50

応答制御部 240 は、表示部 710 に、例えばロボットを表現したキャラクタ 800 を CG として表示しても良い。キャラクタ 800 は、対話ロボット 100 のように、頭部 810 と体部 820 を有し、応答制御部 240 は、頷き動作を実行させる場合に、頭部 810 を前後に揺動するアニメーションを表示する。

【0079】

また、応答制御部 240 は、表示部 710 に吹出し 870 を表示し、フィルター発声と応答発話をテキストとしてユーザに知らせても良い。この場合、スピーカ 712 から音として出力する機能を停止させても良い。

【0080】

以上に説明した各実施態様においては、推定値として交代確率 P を算出した。しかし、受け取った物理的特徴量としての韻律に基づいて、ユーザからの発話を待つべきか、自身が発話すべきかを推定する推定手法は、推定値として数値を利用する手法に限らない。例えば、受け取った物理的特徴量としての韻律に基づいて、「何も応答しない」「フィルター発声を行う」「応答発話を行う」「相槌を発声する」の 4 つのグループのひとつを直接的に出力するものであっても良い。それぞれのグループをクラスと捉え、入力する韻律がいずれのクラスに属するかを深層学習などの手法を用いて演算することもできる。

10

【0081】

以上に説明した各実施態様においては、対話ロボット 100 あるいはスマートフォン 700 が、図 2 または図 7 に示すシステム構成を備えるものとして説明したが、一部の構成が外部機器に存在し、通信によって外部から情報を取得する構成であっても良い。例えば、韻律 DB 310、フィルター DB 320、発話 DB 330、相槌 DB 340 の各データベースは、外部機器として構成された供用のデータベースであっても良い。対話ロボット 100 あるいはスマートフォン 700 は、必要に応じて、通信機能を利用して当該データベースに必要な情報を要求する。このように、大容量の記録媒体を必要とするデータベースを外部に備えれば、音声対話装置としてのシステム構成をより簡易に最適化することができる。

20

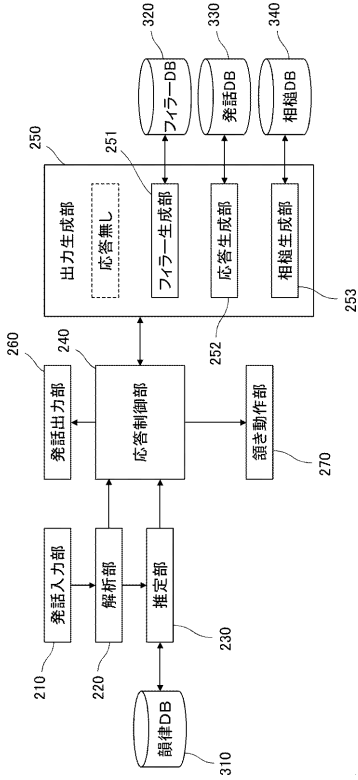
【符号の説明】

【0082】

100 対話ロボット、110 頭部、111 マイク、112 スピーカ、120 体部、210 発話入力部、220 解析部、230 推定部、240 応答制御部、250 出力生成部、251 フィルター生成部、252 応答生成部、253 相槌生成部、260 発話出力部、270 頷き動作部、310 韻律 DB、320 フィルター DB、330 発話 DB、340 相槌 DB、700 スマートフォン、710 表示部、711 マイク、712 スピーカ、800 キャラクタ、810 頭部、820 体部、870 吹出し

30

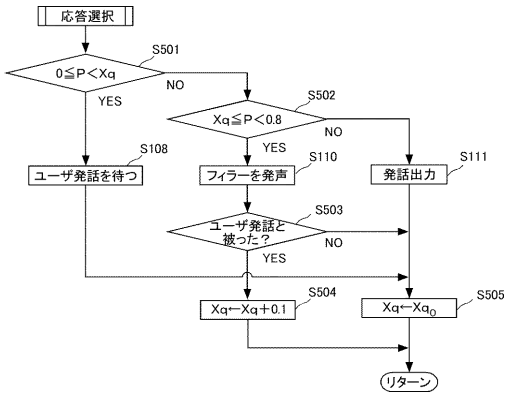
【図7】



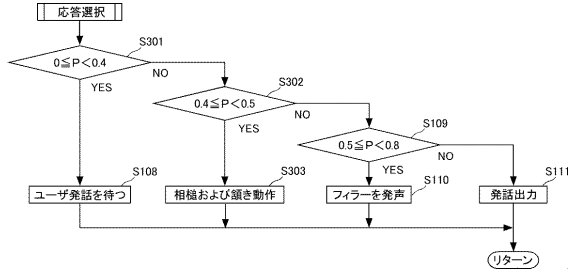
【図8】

交代確率P	0.0-0.4	0.4-0.5	0.5-0.8	0.8-1.0
応答種類	応答無し	相槌	フィルター発声	応答発話

【図12】



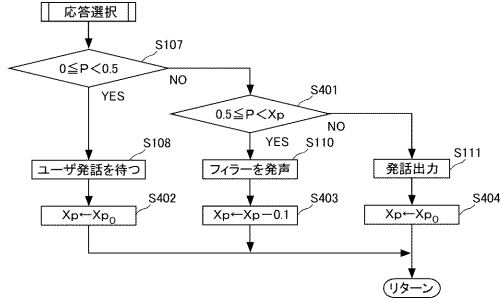
【図9】



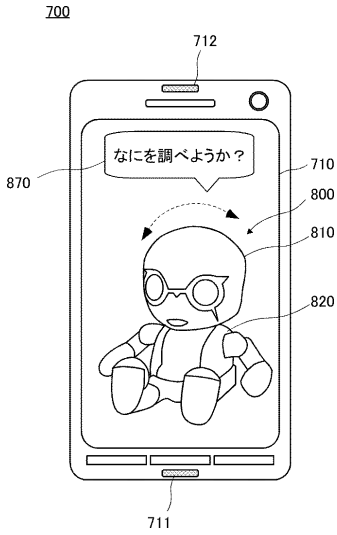
【図10】

交代確率P	0.0-0.4	0.3-0.5	0.5-0.8	0.7-1.0
応答種類	応答無し	相槌	フィルター発声	応答発話

【図11】



【図13】



フロントページの続き

(51) Int.Cl. F I テーマコード(参考)
G 0 6 F 3/16 (2006.01) G 0 6 F 3/16 6 5 0

(72) 発明者 中西 亮輔
京都府京都市左京区吉田本町 3 6 番地 1 国立大学法人京都大学内

(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