

(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2016-38501
(P2016-38501A)

(43) 公開日 平成28年3月22日(2016.3.22)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 13/10 (2013.01)	G 1 0 L 13/10 1 1 4	
G 1 0 L 13/00 (2006.01)	G 1 0 L 13/00 1 0 0 M	
G 1 0 L 13/033 (2013.01)	G 1 0 L 13/10 1 1 1 C	
G 1 0 L 15/02 (2006.01)	G 1 0 L 13/033 1 0 2 A	
	G 1 0 L 15/02 3 0 0 K	

審査請求 未請求 請求項の数 8 O L (全 12 頁)

(21) 出願番号 特願2014-162579 (P2014-162579)
(22) 出願日 平成26年8月8日(2014.8.8)

特許法第30条第2項適用申請有り (1) 平成26年2月26日一般社団法人 人工知能学会発行の「I S S N 0 9 1 8 - 5 6 8 2, 人工知能学会研究会資料, S I G - S L U D - B 3 0 3, 言語・音声理解と対話処理研究会(第70回)」に掲載 (2) 平成26年3月5日開催の「第70回 人工知能学会 言語・音声理解と対話処理研究会(S I G - S L U D)」にて発表

(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
京都市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 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

最終頁に続く

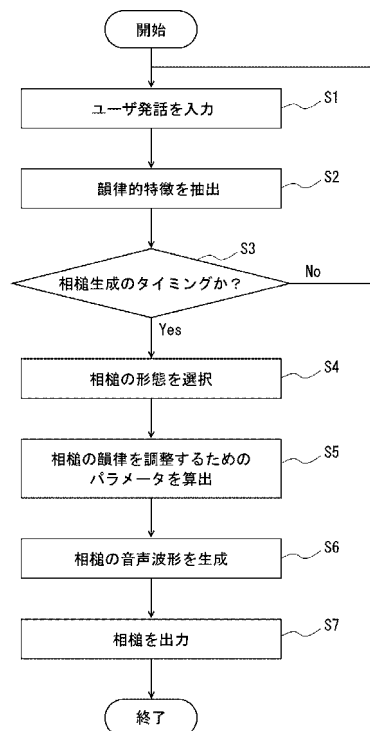
(54) 【発明の名称】 音声対話方法、及び音声対話システム

(57) 【要約】

【課題】 発話を促進させる相槌を生成することが可能な音声対話方法を提供することである。

【解決手段】 本発明にかかる音声対話方法は、ユーザ発話を入力する工程(S1)と、入力されたユーザ発話の韻律的特徴を抽出する工程(S2)と、抽出された韻律的特徴に基づき、ユーザ発話に応答する相槌を生成する工程(S4~S6)と、を備える。相槌を生成する際、相槌の韻律的特徴がユーザ発話の韻律的特徴と合うように相槌の韻律を調整することで、発話を促進させる相槌を生成することができる。

【選択図】 図2



【特許請求の範囲】

【請求項 1】

ユーザ発話を入力する工程と、
 入力された前記ユーザ発話の韻律的特徴を抽出する工程と、
 抽出された前記韻律的特徴に基づき前記ユーザ発話に応答する相槌を生成する工程と、
 を備え、
 前記相槌を生成する際、前記相槌の韻律的特徴が前記ユーザ発話の韻律的特徴と合うように前記相槌の韻律を調整する、
 音声対話方法。

【請求項 2】

前記ユーザ発話の韻律的特徴を抽出する際、前記ユーザ発話の基本周波数成分およびパワー成分を抽出し、
 前記基本周波数成分および前記パワー成分のうち、前記ユーザ発話の韻律的特徴と前記相槌の韻律的特徴との相関が高い成分を用いて前記相槌の韻律を調整する、
 請求項 1 に記載の音声対話方法。

【請求項 3】

前記ユーザ発話の韻律的特徴と前記相槌の韻律的特徴との相関を示す相関係数テーブルを予め生成し、
 前記基本周波数成分および前記パワー成分のうち、前記相槌についての相関係数が高い成分を優先的に用いて前記相槌の韻律を調整する、
 請求項 2 に記載の音声対話方法。

【請求項 4】

前記基本周波数成分は、当該基本周波数成分の最大値および平均値を含み、
 前記パワー成分は、当該パワー成分の最大値および平均値を含む、
 請求項 2 または 3 に記載の音声対話方法。

【請求項 5】

前記相槌を生成する際、下記の式を用いて、前記基本周波数成分の最大値、平均値、及び前記パワー成分の最大値、平均値の各々について韻律調整パラメータ BC_{ip} を求め、当該韻律調整パラメータ BC_{ip} を用いて前記相槌の韻律を調整する、請求項 4 に記載の音声対話方法。

【数 1】

$$BC_{ip} = \alpha \times \left\{ \frac{S_i - E(S)}{\sigma(S)} \times \sigma(BC) \right\} + E(BC)$$

上記式において、 α は相関係数、 S_i はユーザ発話の韻律的特徴、 i はサンプル数、 $E(S)$ はユーザ発話の韻律的特徴の平均値、 $E(BC)$ は相槌の韻律的特徴の平均値、 $\sigma(S)$ はユーザ発話の韻律的特徴の標準偏差、 $\sigma(BC)$ は相槌の韻律的特徴の標準偏差である。

【請求項 6】

前記ユーザ発話の韻律的特徴を用いて前記相槌を生成するタイミングを決定する工程を更に備え、
 前記ユーザ発話の韻律的特徴であるパワー成分が所定の閾値以下である場合に、前記相槌を生成する、
 請求項 1 乃至 5 のいずれか一項に記載の音声対話方法。

【請求項 7】

前記相槌には感情表出系の相槌と応答系の相槌とが含まれており、
 前記ユーザ発話が発話中である場合、前記応答系の相槌を選択し、
 前記ユーザ発話が終了している場合、前記感情表出系の相槌を選択する、
 請求項 1 乃至 6 のいずれか一項に記載の音声対話方法。

【請求項 8】

10

20

30

40

50

ユーザ発話を入力する発話入力部と、
前記発話入力部に入力された前記ユーザ発話の韻律的特徴を抽出する韻律的特徴抽出部と、

前記韻律的特徴抽出部で抽出された前記韻律的特徴に基づき前記ユーザ発話に応答する相槌を生成する相槌生成部と、を備え、

前記相槌生成部は、前記相槌の韻律的特徴が前記ユーザ発話の韻律的特徴と合うように前記相槌の韻律を調整する、

音声対話システム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は音声対話方法、及び音声対話システムに関する。

【背景技術】

【0002】

音声対話システムや人型ロボットにおいては、高齢者や認知症などの患者のケアを行うニーズが高まっており、傾聴する機能が要求されている。傾聴においては、ユーザが話しやすいように相槌を打つことが重要である。

【0003】

特許文献1には、自然で円滑な対話を実現できる音声認識装置に関する技術が開示されている。特許文献1に開示されている音声認識装置では、音声入力部に入力された音声信号を基に計算した話者の音声特徴量に基づき、話者との対話中にスピーカから相槌音を出力させる相槌タイミングを推測している。そして、相槌タイミングであるとの推測結果が得られると、相槌タイミング直前のパワーを基に相槌音を出力させるか否かを判定している。

20

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2009-3040号公報

【発明の概要】

【発明が解決しようとする課題】

30

【0005】

しかしながら、特許文献1に開示されている技術では、相槌を打つタイミングについてのみ焦点が置かれており、実際に打たれている相槌は同一の音声となっている。傾聴においては、ユーザが話しやすいように相槌を打つことが重要であるが、相槌の音声がある場合は、ユーザに機械的な印象を与えてしまい、ユーザは話を聞いてもらっているという意識を持つことができない。このため、ユーザの発話が促進されないという問題があった。

【0006】

上記課題に鑑み本発明の目的は、発話を促進させる相槌を生成することが可能な音声対話方法、及び音声対話システムを提供することである。

40

【課題を解決するための手段】

【0007】

本発明にかかる音声対話方法は、ユーザ発話を入力する工程と、入力された前記ユーザ発話の韻律的特徴を抽出する工程と、抽出された前記韻律的特徴に基づき前記ユーザ発話に応答する相槌を生成する工程と、を備え、前記相槌を生成する際、前記相槌の韻律的特徴が前記ユーザ発話の韻律的特徴と合うように前記相槌の韻律を調整する。

【0008】

本発明にかかる音声対話システムは、ユーザ発話を入力する発話入力部と、前記発話入力部に入力された前記ユーザ発話の韻律的特徴を抽出する韻律的特徴抽出部と、前記韻律的特徴抽出部で抽出された前記韻律的特徴に基づき前記ユーザ発話に応答する相槌を生成

50

する相槌生成部と、を備え、前記相槌生成部は、前記相槌の韻律的特徴が前記ユーザ発話の韻律的特徴と合うように前記相槌の韻律を調整する。

【0009】

本発明にかかる音声対話方法および音声対話システムでは、ユーザ発話の韻律的特徴を抽出し、相槌を生成する際に、相槌の韻律的特徴がユーザ発話の韻律的特徴と合うように相槌の韻律（音声波形）を調整している。このように相槌の韻律を調整することで、ユーザに機械的な印象を与えることを抑制することができ、ユーザは話を聞いてもらっているという意識を持つことができ、ユーザの発話を促すことができる。

【発明の効果】

【0010】

本発明により、発話を促進させる相槌を生成することが可能な音声対話方法、及び音声対話システムを提供することができる。

【図面の簡単な説明】

【0011】

【図1】実施の形態にかかる音声対話システムを示すブロック図である。

【図2】実施の形態にかかる音声対話方法を説明するためのフローチャートである。

【図3】ユーザと音声対話システムとが対話している状態を示す図である。

【図4】ユーザ発話の韻律的特徴と相槌の韻律的特徴との相関を示す相関係数テーブルを示す図である。

【図5】ユーザ発話の韻律的特徴と相槌の韻律的特徴との相関を示す相関係数テーブルの一例を示す図である。

【発明を実施するための形態】

【0012】

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。

図3は、ユーザと音声対話システムとが対話している状態を示す図である。図3に示すように、本実施の形態にかかる発明は、ユーザ31がロボット（音声対話システム）32と対話する際に、ロボット32が、ユーザ31の発話を促進させる相槌を発することを特徴としている。つまり、本実施の形態にかかる発明では、ユーザ31の発話の音声波形33から韻律的特徴を抽出し、相槌を生成する際に、相槌の音声波形34の韻律的特徴がユーザ31の発話の音声波形33の韻律的特徴と合うように相槌の韻律（音声波形34）を調整することを特徴としている。以下で、本実施の形態にかかる音声対話方法、及び音声対話システムについて詳細に説明する。

【0013】

図1は、本実施の形態にかかる音声対話システムを示すブロック図である。図1に示すように、本実施の形態にかかる音声対話システム1は、発話入力部11、韻律的特徴抽出部12、相槌生成タイミング決定部13、相槌データベース15、相槌選択部16、韻律調整パラメータ生成部17、相槌波形生成部18、及び相槌出力部19を備える。相槌データベース15、相槌選択部16、韻律調整パラメータ生成部17、及び相槌波形生成部18は、相槌生成部14を構成している。

【0014】

発話入力部11は、ユーザの発話を入力する。例えば、発話入力部11はマイク等を用いて構成することができる。

【0015】

韻律的特徴抽出部12は、発話入力部11に入力されたユーザ発話（先行発話）の韻律的特徴を抽出する。韻律的特徴としては、ユーザ発話の基本周波数成分 F_0 （以下、単に F_0 と記載する場合もある）やパワー成分が挙げられる。このとき、基本周波数成分 F_0 として、 F_0 の対数を用いてもよい。例えば、 F_0 の対数は、発話音声を用いて10m秒毎に F_0 を算出し、この算出された F_0 に対して10を底とする対数を取ることによって求めることができる。また、パワー成分についても、例えば10m秒毎にdB値を算出することによって求めることができる。韻律的特徴抽出部12は、抽出した韻律的特徴21を相槌生成タ

10

20

30

40

50

イミング決定部 13 に出力する。

【0016】

また、韻律的特徴抽出部 12 は、相槌生成タイミング決定部 13 から相槌生成タイミング情報 22 が供給された際、相槌選択部 16 に相槌選択信号 23 を出力する。

【0017】

また、韻律的特徴抽出部 12 は、相槌生成タイミング決定部 13 から相槌生成タイミング情報 22 が供給された際、相槌生成タイミングから所定の時間さかのぼった期間（例えば、500ms）における基本周波数成分 F_0 の最大値、平均値、最大値と最小値のレンジ等、及びパワー成分の最大値、平均値、最大値と最小値のレンジ等の特徴量を算出する。算出された特徴量 24 は、韻律調整パラメータ生成部 17 に供給される。

10

【0018】

相槌生成タイミング決定部 13 は、韻律的特徴抽出部 12 で抽出された韻律的特徴 21 を用いて、相槌を生成するタイミングを決定する。また、相槌生成タイミング決定部 13 は、相槌を生成するタイミングを決定した場合、相槌生成タイミング情報 22 を韻律的特徴抽出部 12 に出力する。

【0019】

例えば、相槌生成タイミング決定部 13 は、ユーザ発話の韻律的特徴であるパワー成分が所定の閾値以下である場合に、相槌を生成するタイミングであると決定することができる。つまり、ユーザが発話が終了したタイミングでは、ユーザ発話のパワー成分がほぼゼロになるので、このタイミングを相槌を生成するタイミングであると決定することができる。また、ユーザ発話が途中の場合であっても、ユーザ発話のパワー成分が小さい場合は、ユーザ発話の終了が近づいていると判断することができる。よって、このような場合も、相槌を生成するタイミングであると決定することができる。

20

【0020】

なお、上記では、ユーザ発話の韻律的特徴としてパワー成分を用いた場合を例として挙げたが、例えば、ユーザ発話の基本周波数成分 F_0 を用いて相槌を生成するタイミングを決定してもよい。例えば、相槌生成タイミング決定部 13 は、ユーザ発話の基本周波数成分 F_0 が所定の閾値以下である場合に、相槌を生成するタイミングであると決定してもよい。つまり、ユーザ発話の基本周波数成分 F_0 が所定の閾値以下である場合は、ユーザ発話のトーンが下がっている状態であるので、ユーザ発話の終了が近づいていると判断することができる。

30

【0021】

相槌データベース 15 は、ユーザ発話の韻律的特徴と相槌の韻律的特徴との相関を示す相関係数テーブルを格納している。この相関係数テーブルは予め生成されている。図 4 は、ユーザ発話の韻律的特徴と相槌の韻律的特徴との相関を示す相関係数テーブルを示す図である。図 4 に示すように、相関係数テーブルは、各々の相槌（相槌の形態）と相関係数とを対応付けたテーブルである。相関係数は、韻律的特徴の特徴量毎に求める。つまり、相関係数は、基本周波数成分 F_0 の最大値、平均値、及びパワー成分の最大値、平均値のそれぞれについて算出する。

【0022】

40

例えば、相関係数（1、1）は、ユーザ発話（先行発話）と相槌「あー」との相関を示す相関係数のうち、基本周波数成分 F_0 の最大値を用いて求めた相関係数である。相関係数（1、2）は、ユーザ発話（先行発話）と相槌「あー」との相関を示す相関係数のうち、基本周波数成分 F_0 の平均値を用いて求めた相関係数である。相関係数（1、3）は、ユーザ発話（先行発話）と相槌「あー」との相関を示す相関係数のうち、パワー成分の最大値を用いて求めた相関係数である。相関係数（1、4）は、ユーザ発話（先行発話）と相槌「あー」との相関を示す相関係数のうち、パワー成分の平均値を用いて求めた相関係数である。

【0023】

相関係数は、話し役（複数のサンプル）と聞き役（カウンセラ）の対話を収録し、この

50

収録した対話の音声进行分析して、ユーザ発話と相槌との相関を相槌の形態別に調べることで推定することができる。ここで、話し役は主にユーザ発話を発し、聞き役は主に相槌を発する。相関係数を求める場合、相槌の開始から終了までの韻律的特徴と、相槌の直前のユーザ発話の有声区間（例えば、500m秒）の韻律的特徴を使用する。使用する韻律的特徴の種類は、該当区間の対数F0の最大値、平均値、及びパワー成分の最大値、平均値とすることができる。

【0024】

なお、図4に示すように、相槌の種類には感情表出系の相槌と応答系の相槌とがある。感情表出系の相槌は、「あー」、「はー」等の興味、理解、共感等の感情を示す相槌である。応答系の相槌は、「ふーん」、「はい」等の相手の発話に対する応答を示す相槌である。

10

【0025】

図1に示す相槌選択部16は、韻律的特徴抽出部12から相槌選択信号23が供給されると、相槌データベース15に格納されている相槌の形態の中から、所定の相槌を選択する。このとき選択される相槌は任意に決定することができる。一例を挙げると、相槌生成タイミング決定部13で決定されたタイミングがユーザ発話の途中のタイミングである場合、応答系の相槌（つまり、相手の発話に対する応答を示す相槌）の中から相槌を選択してもよい。一方、相槌生成タイミング決定部13で決定されたタイミングがユーザ発話が終了したタイミングである場合、感情表出系の相槌（つまり、興味、理解、共感等の感情を示す相槌）の中から相槌を選択してもよい。

20

【0026】

相槌選択部16は、選択した相槌に関する相槌情報25（例えば、テキストデータ）を相槌波形生成部18に出力する。また、相槌選択部16は、選択した相槌の相関係数に関する情報26を、韻律調整パラメータ生成部17に出力する。相槌選択部16は、相関係数に関する情報を相槌データベース15から取得することができる。相槌選択部16は、例えば、相槌として図4に示す「あー」を選択した場合、相関係数に関する情報26として、（1、1）、（1、2）、（1、3）、（1、4）の値を韻律調整パラメータ生成部17に出力する。

【0027】

韻律調整パラメータ生成部17は、相槌選択部16で選択された相槌の韻律的特徴が、ユーザ発話の韻律的特徴と合うように相槌の韻律を調整するパラメータを生成する。このとき、韻律調整パラメータ生成部17は、韻律的特徴抽出部12から供給された特徴量24と、相槌選択部16から供給された相関係数に関する情報26とを用いて、韻律調整パラメータを生成する。生成された韻律調整パラメータ27は、相槌波形生成部18に供給される。

30

【0028】

具体的には、韻律調整パラメータ生成部17は、下記の式を用いて韻律調整パラメータ BC_{ip} を求める。このとき、韻律調整パラメータ生成部17は、基本周波数成分F0の最大値、平均値、及びパワー成分の最大値、平均値の各々について韻律調整パラメータ BC_{ip} を求める。

40

【0029】

【数1】

$$BC_{ip} = \alpha \times \left\{ \frac{S_i - E(S)}{\sigma(S)} \times \sigma(BC) \right\} + E(BC)$$

【0030】

上記式において、 BC_{ip} は韻律調整パラメータ（相槌の韻律的特徴の目標値）、 α は相関係数、 S_i はユーザ発話の韻律的特徴を示す。 i はサンプル数であり、 $i = 1, 2, \dots, N$ である。 $E(S)$ はユーザ発話の直前Nターンの発話（ $N - 1$ ）における平均値（ユーザ発話の韻律的特徴の平均値）、 $E(BC)$ は相槌データベースにおける平均値

50

(相槌の韻律的特徴の平均値)である。(S)はユーザ発話の直前Nターンの発話(N-1)における標準偏差(ユーザ発話の韻律的特徴の標準偏差)、(BC)は相槌データベースにおける標準偏差(相槌の韻律的特徴の標準偏差)である。本実施の形態では、 S_i 、 $E(S)$ 、 $E(BC)$ 、(S)、(BC)は、基本周波数成分F0の最大値、平均値、及びパワー成分の最大値、平均値のそれぞれについて求める。 $E(BC)$ および(BC)は、相槌データベース15に予め格納されている。なお、ユーザ発話は、初対面なら直前のターンだけで推測、リピーター(かつ、ユーザ判別可能)なら過去の対話履歴全体から推測してもよい。

【0031】

例えば、相槌選択部16において相槌として「あー」が選択された場合、韻律調整パラメータ生成部17には、相関係数に関する情報26として(1,1)、(1,2)、(1,3)、(1,4)が供給される。

10

【0032】

韻律調整パラメータ生成部17は、韻律的特徴抽出部12から供給されたユーザ発話の基本周波数成分F0の最大値を用いて、 S_i 、 $E(S)$ 、(S)を求める。なお、 $E(BC)$ 、(BC)については、相槌データベースの値を用いて求める。その後、韻律調整パラメータ生成部17は、基本周波数成分F0の最大値に対応した相関係数(1,1)、基本周波数成分F0の最大値に対応した S_i 、 $E(S)$ 、(S)、 $E(BC)$ 、(BC)を上記式に代入して、基本周波数成分F0の最大値に対応した韻律調整パラメータ $BC_{ip}(F0_{max})$ を算出する。

20

【0033】

同様に、韻律調整パラメータ生成部17は、基本周波数成分F0の平均値に対応した韻律調整パラメータ $BC_{ip}(F0_{ave})$ 、パワーの最大値に対応した韻律調整パラメータ $BC_{ip}(P_{max})$ 、パワーの平均値に対応した韻律調整パラメータ $BC_{ip}(P_{ave})$ のそれぞれを算出する。算出されたこれらの韻律調整パラメータ27は、相槌波形生成部18に供給される。

【0034】

なお、上記では4つの韻律調整パラメータ BC_{ip} を求める場合について説明したが、求める韻律調整パラメータ BC_{ip} の数はこれ以外であってもよい。例えば、韻律調整パラメータ生成部17は、基本周波数成分F0およびパワー成分のうち、ユーザ発話の韻律的特徴と相槌の韻律的特徴との相関が高い成分(つまり、相関係数が高い成分:図5を参照)について、韻律調整パラメータ BC_{ip} を求めるようにしてもよい。換言すると、韻律調整パラメータ生成部17は、基本周波数成分F0およびパワー成分のうち、相槌についての相関係数が高い成分を優先的に用いて、韻律調整パラメータ BC_{ip} を求めるようにしてもよい。

30

【0035】

図5は、ユーザ発話の韻律的特徴と相槌の韻律的特徴との相関を示す相関係数テーブルの一例を示す図である。図5に示すように、各成分における相関係数は、相槌の形態に応じて異なってくる。例えば、相槌の形態が「はー」である場合は、相関係数の値が大きい「パワー成分の最大値(相関係数0.47)」および「パワー成分の平均値(相関係数0.29)」のそれぞれに対応した韻律調整パラメータ $BC_{ip}(P_{max})$ 、 $BC_{ip}(P_{ave})$ を求めてもよい。また、例えば、相槌の形態が「ふん」、「うん」である場合は、相関係数の値が大きい「基本周波数成分F0の最大値(相関係数0.22)」および「パワー成分の最大値(相関係数0.23)」のそれぞれに対応した韻律調整パラメータ $BC_{ip}(F0_{max})$ 、 $BC_{ip}(P_{max})$ を求めてもよい。このように、基本周波数成分F0の最大値および平均値、並びにパワー成分の最大値および平均値のうち、相関係数が高い成分を優先的に用いて韻律調整パラメータ BC_{ip} を求めることで、韻律調整パラメータの精度を向上させることができる。また、韻律調整パラメータを求める際の演算量を低減させることができる。

40

【0036】

50

図 1 に示す相槌波形生成部 18 は、相槌選択部 16 で選択された相槌に関する相槌情報 25 (例えば、テキストデータ) と、韻律調整パラメータ生成部 17 で生成された韻律調整パラメータ 27 とを用いて、相槌の音声波形を生成する。ここで、韻律調整パラメータ 27 は、基本周波数成分 F_0 の最大値に対応した韻律調整パラメータ $BC_{ip}(F0_{max})$ 、基本周波数成分 F_0 の平均値に対応した韻律調整パラメータ $BC_{ip}(F0_{ave})$ 、パワーの最大値に対応した韻律調整パラメータ $BC_{ip}(P_{max})$ 、及びパワーの平均値に対応した韻律調整パラメータ $BC_{ip}(P_{ave})$ の少なくとも 1 つである。例えば、相槌波形生成部 18 は、TTS (text to speech) 技術を用いて相槌の音声波形を生成することができる。

【0037】

このように、相槌データベース 15、相槌選択部 16、韻律調整パラメータ生成部 17、及び相槌波形生成部 18 で構成される相槌生成部 14 は、韻律的特徴抽出部 12 で抽出された韻律的特徴に基づいて、ユーザ発話に応答する相槌の音声波形を生成することができる。

【0038】

相槌波形生成部 18 で生成された相槌の音声波形は、相槌出力部 19 に供給される。相槌出力部 19 は、供給された音声波形に対応した相槌を出力する。例えば、相槌出力部 19 はスピーカ等を用いて構成することができる。これにより、ロボット (音声対話システム) 32 は、相槌の韻律的特徴がユーザ発話の韻律的特徴と合うように韻律が調整された相槌を出力することができる。このように相槌の韻律を調整することで、ユーザの発話を促すことができる。

【0039】

なお、本実施の形態にかかる音声対話システムでは、相槌出力部 19 から出力される相槌に応じてロボットが首を振るように構成してもよい。このように、相槌に合わせてロボットが首を振るようにすることで、ユーザの発話を更に促すことができる。

【0040】

次に、本実施の形態にかかる音声対話システムの動作 (音声対話方法) について説明する。図 2 は、本実施の形態にかかる音声対話方法を説明するためのフローチャートである。なお、この場合も、相槌データベース 15 には、予めユーザ発話の韻律的特徴と相槌の韻律的特徴との相関を示す相関係数テーブルが格納されているものとする。

【0041】

図 1、図 2 に示すように、まず、音声対話システム 1 の発話入力部 11 は、ユーザの発話を入力する (ステップ S1)。次に、韻律的特徴抽出部 12 は、発話入力部 11 に入力されたユーザ発話 (先行発話) の韻律的特徴を抽出する (ステップ S2)。韻律的特徴としては、ユーザ発話の基本周波数成分 F_0 やパワー成分が挙げられる。次に、相槌生成タイミング決定部 13 は、韻律的特徴抽出部 12 で抽出された韻律的特徴 21 を用いて、相槌を生成するタイミングを決定する。相槌生成タイミング決定部 13 が相槌生成タイミングではないと判断した場合 (ステップ S3: No)、再度、ステップ S1 ~ S3 の動作を繰り返す。一方、相槌生成タイミング決定部 13 が相槌生成タイミングであると判断した場合 (ステップ S3: Yes)、相槌生成タイミング情報 22 を韻律的特徴抽出部 12 に出力する。例えば、相槌生成タイミング決定部 13 は、ユーザ発話の韻律的特徴であるパワー成分が所定の閾値以下である場合に、相槌を生成するタイミングであると決定することができる。

【0042】

韻律的特徴抽出部 12 は、相槌生成タイミング決定部 13 から相槌生成タイミング情報 22 が供給された場合、相槌選択部 16 に相槌選択信号 23 を出力する。また、韻律的特徴抽出部 12 は、相槌生成タイミング決定部 13 から相槌生成タイミング情報 22 が供給された場合、相槌生成タイミングから所定の時間さかのぼった期間 (例えば、500ms) における基本周波数成分 F_0 の最大値、平均値、最大値と最小値のレンジ等、及びパワー成分の最大値、平均値、最大値と最小値のレンジ等の特徴量を算出する。算出された特

10

20

30

40

50

微量 2 4 は、韻律調整パラメータ生成部 1 7 に供給される。

【 0 0 4 3 】

相槌選択部 1 6 は、韻律的特徴抽出部 1 2 から相槌選択信号 2 3 が供給されると、相槌データベース 1 5 に格納されている相槌の形態の中から、所定の相槌（相槌の形態）を選択する（ステップ S 4）。また、相槌選択部 1 6 は、選択した相槌に関する相槌情報 2 5（例えば、テキストデータ）を相槌波形生成部 1 8 に出力する。また、相槌選択部 1 6 は、選択した相槌の相関係数に関する情報 2 6 を、韻律調整パラメータ生成部 1 7 に出力する。相槌選択部 1 6 は、相関係数に関する情報を相槌データベース 1 5 から取得することができる。

【 0 0 4 4 】

韻律調整パラメータ生成部 1 7 は、相槌選択部 1 6 で選択された相槌の韻律的特徴が、ユーザ発話の韻律的特徴と合うように相槌の韻律を調整するパラメータを生成する（ステップ S 5）。このとき、韻律調整パラメータ生成部 1 7 は、韻律的特徴抽出部 1 2 から供給された特徴量 2 4 と、相槌選択部 1 6 から供給された相関係数に関する情報 2 6 とを用いて、韻律調整パラメータを生成する。生成された韻律調整パラメータ 2 7 は、相槌波形生成部 1 8 に供給される。

【 0 0 4 5 】

具体的には、韻律調整パラメータ生成部 1 7 は、上記式を用いて韻律調整パラメータ BC_{ip} を求める。このとき、韻律調整パラメータ生成部 1 7 は、基本周波数成分 F_0 の最大値、平均値、及びパワー成分の最大値、平均値の各々について韻律調整パラメータ BC_{ip} を求める。

【 0 0 4 6 】

相槌波形生成部 1 8 は、相槌選択部 1 6 で選択された相槌に関する相槌情報 2 5 と、韻律調整パラメータ生成部 1 7 で生成された韻律調整パラメータ 2 7 とを用いて、相槌の音声波形を生成する（ステップ S 6）。ここで、韻律調整パラメータ 2 7 は、基本周波数成分 F_0 の最大値に対応した韻律調整パラメータ $BC_{ip}(F_0_{max})$ 、基本周波数成分 F_0 の平均値に対応した韻律調整パラメータ $BC_{ip}(F_0_{ave})$ 、パワーの最大値に対応した韻律調整パラメータ $BC_{ip}(P_{max})$ 、及びパワーの平均値に対応した韻律調整パラメータ $BC_{ip}(P_{ave})$ の少なくとも 1 つである。例えば、相槌波形生成部 1 8 は、TTS (text to speech) 技術を用いて相槌の音声波形を生成することができる。

【 0 0 4 7 】

相槌波形生成部 1 8 で生成された相槌の音声波形は、相槌出力部 1 9 に供給される。相槌出力部 1 9 は、供給された音声波形に対応した相槌を出力する（ステップ S 7）。これにより、ロボット（音声対話システム）3 2 は、相槌の韻律的特徴がユーザ発話の韻律的特徴と合うように韻律が調整された相槌を出力することができる。このとき、相槌出力部 1 9 から出力される相槌に応じてロボットが首を振るように構成してもよい。

【 0 0 4 8 】

背景技術で説明したように、特許文献 1 に開示されている音声認識装置では、音声入力部に入力された音声信号を基に計算した話者の音声特徴量に基づき、話者との対話中にスピーカから相槌音を出力させる相槌タイミングを推測している。そして、相槌タイミングであるとの推測結果が得られると、相槌タイミング直前のパワーを基に相槌音を出力させるか否かを判定している。

【 0 0 4 9 】

しかしながら、特許文献 1 に開示されている技術では、相槌を打つタイミングについてのみ焦点が置かれており、実際に打たれている相槌は同一の音声となっている。傾聴においては、ユーザが話しやすいように相槌を打つことが重要であるが、相槌の音声が同一である場合は、ユーザに機械的な印象を与えてしまい、ユーザは話を聞いてもらっているという意識を持つことができない。このため、ユーザの発話が促進されないという問題があった。

10

20

30

40

50

【 0 0 5 0 】

そこで本実施の形態にかかる音声対話方法および音声対話システムでは、ユーザ発話の音声波形から韻律的特徴を抽出し、相槌を生成する際に、相槌の音声波形の韻律的特徴がユーザ発話の音声波形の韻律的特徴と合うように相槌の韻律（音声波形）を調整している。このように相槌の韻律を調整することで、ユーザに機械的な印象を与えることを抑制することができ、ユーザは話を聞いてもらっているという意識を持つことができ、ユーザの発話を促すことができる。よって、本実施の形態にかかる発明により、発話を促進させる相槌を生成することが可能な音声対話方法、及び音声対話システムを提供することができる。

【 0 0 5 1 】

つまり、本実施の形態にかかる発明では、図 3 に示すように、ユーザ 3 1 の発話の音声波形 3 3 から韻律的特徴 S_i を抽出し、この抽出した韻律的特徴 S_i を上記で示した式に代入して、相槌の韻律的特徴を予測している（つまり、 BC_{ip} を求めている）。よって、相槌を生成する際に、相槌の音声波形 3 4 の韻律的特徴 BC_{ip} がユーザ 3 1 の発話の音声波形 3 3 の韻律的特徴と合うように相槌の韻律（音声波形 3 4）を調整することができる。

【 0 0 5 2 】

ここで、上記式における $E(BC)$ は、相槌の韻律的特徴（ F_0 、パワー）の平均値であり、上記式では、この $E(BC)$ の値をベースラインとし、この $E(BC)$ に、ユーザ発話の韻律的特徴 S_i に応じた値を加算することで、相槌の韻律的特徴（韻律調整パラメータ） BC_{ip} を求めている。

【 0 0 5 3 】

以上、本発明を上記実施形態に即して説明したが、本発明は上記実施の形態の構成にのみ限定されるものではなく、本願特許請求の範囲の請求項の発明の範囲内で当業者であればなし得る各種変形、修正、組み合わせを含むことは勿論である。

【 符号の説明 】

【 0 0 5 4 】

- 1 音声対話システム
- 1 1 発話入力部
- 1 2 韻律的特徴抽出部
- 1 3 相槌生成タイミング決定部
- 1 4 相槌生成部
- 1 5 相槌データベース
- 1 6 相槌選択部
- 1 7 韻律調整パラメータ生成部
- 1 8 相槌波形生成部
- 1 9 相槌出力部
- 2 1 抽出した韻律的特徴
- 2 2 相槌生成タイミング情報
- 2 3 相槌選択信号
- 2 4 特徴量
- 2 5 相槌情報
- 2 6 相関係数に関する情報
- 2 7 韻律調整パラメータ
- 3 1 ユーザ
- 3 2 ロボット
- 3 3 ユーザ発話の音声波形
- 3 4 相槌の音声波形

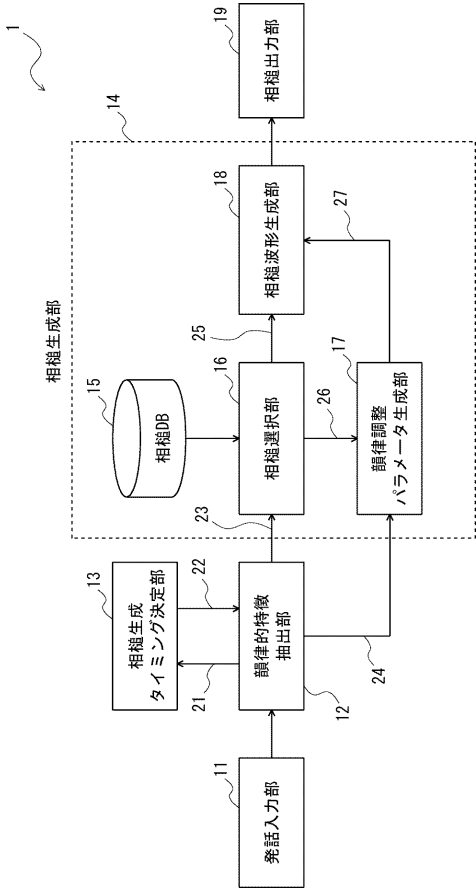
10

20

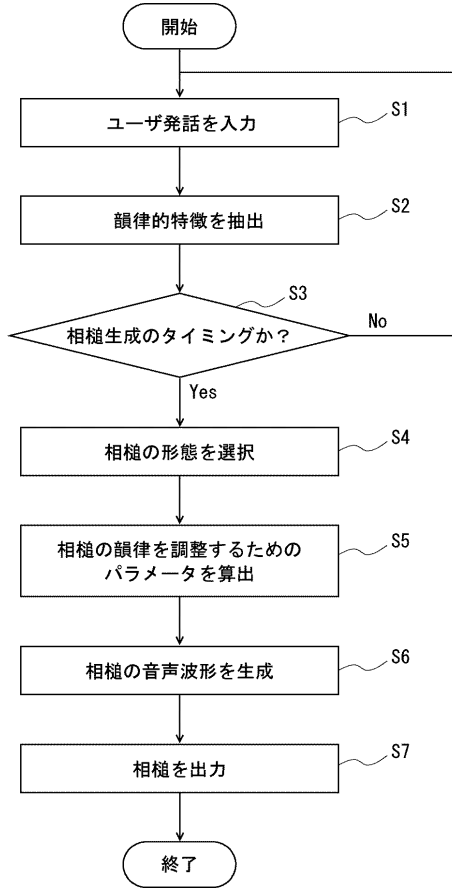
30

40

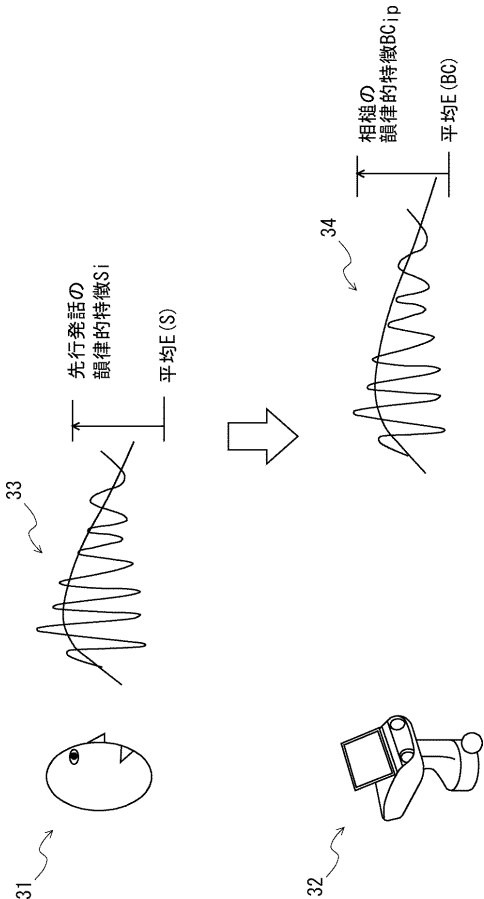
【図1】



【図2】



【図3】



【図4】

相槌の種類	相槌の形態	F0の最大値	F0の平均値	パワーの最大値	パワーの平均値
感情表出系	あー	$\alpha(1,1)$	$\alpha(1,2)$	$\alpha(1,3)$	$\alpha(1,4)$
	はー	$\alpha(2,1)$	$\alpha(2,2)$	$\alpha(2,3)$	$\alpha(2,4)$

応答系	ふーん	$\alpha(k,1)$	$\alpha(k,2)$	$\alpha(k,3)$	$\alpha(k,4)$
	はい	$\alpha(k+1,1)$	$\alpha(k+1,2)$	$\alpha(k+1,3)$	$\alpha(k+1,4)$

【図5】

相槌の種類	相槌の形態	F0の最大値	F0の平均値	パワーの最大値	パワーの平均値
感情表出系	あー	-	0.22	-	0.25
	はー	-	0.23	0.47	0.29
応答系	うーん ふーん	0.14	-	0.14	0.18
	うん ふん	0.22	0.12	0.23	-
	「うん」の繰り返し	-	-	0.34	0.35

フロントページの続き

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