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

特許第6270661号
(P6270661)

(45)発行日　平成30年1月31日(2018.1.31)　(24)登録日　平成30年1月12日(2018.1.12)

(51)Int.Cl.		F I		
G10L 13/10	(2013.01)	G10L	13/10	114
G10L 13/00	(2006.01)	G10L	13/00	100M
G10L 13/033	(2013.01)	G10L	13/10	111C
G10L 15/02	(2006.01)	G10L	13/033	102A
		G10L	15/02	300K

請求項の数 7　(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)
(65)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)
審査請求日　平成28年10月7日(2016.10.7)

特許法第３０条第２項適用　（１）平成２６年２月２６日一般社団法人　人工知能学会発行の「ＩＳＳＮ　０９１８－５６８２，人工知能学会研究会資料，ＳＩＧ－ＳＬＵＤ－Ｂ３０３，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（ＳＩＧ－ＳＬＵＤ）」にて発表

(73)特許権者　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
(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　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　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)【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み、
前記相槌を生成する際、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１に記載の音声対話方法。
【請求項３】

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む、請求項１または２に記載の音声対話方法。

【請求項４】

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３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、σ（Ｓ）はユーザ発話の韻律的特徴の標準偏差、σ（ＢＣ）は相槌の韻律的特徴の標準偏差である。

【請求項５】

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４のいずれか一項に記載の音声対話方法。

【請求項６】

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５のいずれか一項に記載の音声対話方法。

【請求項７】

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、

前記韻律的特徴抽出部は、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み、

前記相槌生成部は、前記相槌を生成する際、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】

【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。

【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】

【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。

【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。

【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Ｆ０（以下、単にＦ０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Ｆ０として、Ｆ０の対数を用いてもよい。例えば、Ｆ０の対数は、発話音声を用いて１０ｍ秒毎にＦ０を算出し、この算出されたＦ０に対して１０を底とする対数を取ることで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値を算出することで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、５００ｍ秒）における基本周波数成分Ｆ０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。
【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。

【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Ｆ０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Ｆ０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Ｆ０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。

【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αとを対応付けたテーブルである。相関係数αは、韻律的特徴の特徴量毎に求める。つまり、相関係数α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算出する。

【００２２】

例えば、相関係数α（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Ｆ０の最大値を用いて求めた相関係数である。相関係数α（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Ｆ０の平均値を用いて求めた相関係数である。相関係数α（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α（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、この収録した対話の音声を分析して、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。

【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。

【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、α（１、１）、α（１、２）、α（１、３）、α（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。

【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。

【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、・・・、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（相槌の韻律的特徴の平均値）である。σ（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）、σ（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）、σ（Ｓ）、σ（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σ（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α（１、１）、α（１、２）、α（１、３）、α（１、４）が供給される。

【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）、σ（Ｓ）を求める。なお、Ｅ（ＢＣ）、σ（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α（１、１）、基本周波数成分Ｆ０の最大値に対応したＳ$_{i}$、Ｅ（Ｓ）、σ（Ｓ）、Ｅ（ＢＣ）、σ

（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。

【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α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。

【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。

【００３６】

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Ｆ０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）、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Ｆ０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、５００ｍ秒）における基本周波数成分Ｆ０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。

【００４３】

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（相槌の形態）を選択する（ステップＳ４）。また、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。

【００４４】

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（ステップＳ５）。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。

【００４５】

具体的には、韻律調整パラメータ生成部１７は、上記式を用いて韻律調整パラメータＢ

C_{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。

【００４６】

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（ステップＳ６）。ここで、韻律調整パラメータ２７は、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）、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）、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。

【００４７】

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（ステップＳ７）。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。このとき、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。

【００４８】

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。

【００４９】

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。

【００５０】

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。

【００５１】

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３１の発話の音声波形３３から韻律的特徴Ｓｉを抽出し、この抽出した韻律的特徴Ｓｉ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３４の韻律的特徴BC_{ip}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）を調整することができる。

【００５２】

ここで、上記式におけるＥ（ＢＣ）は、相槌の韻律的特徴（Ｆ０、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Ｓｉ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。

【００５３】

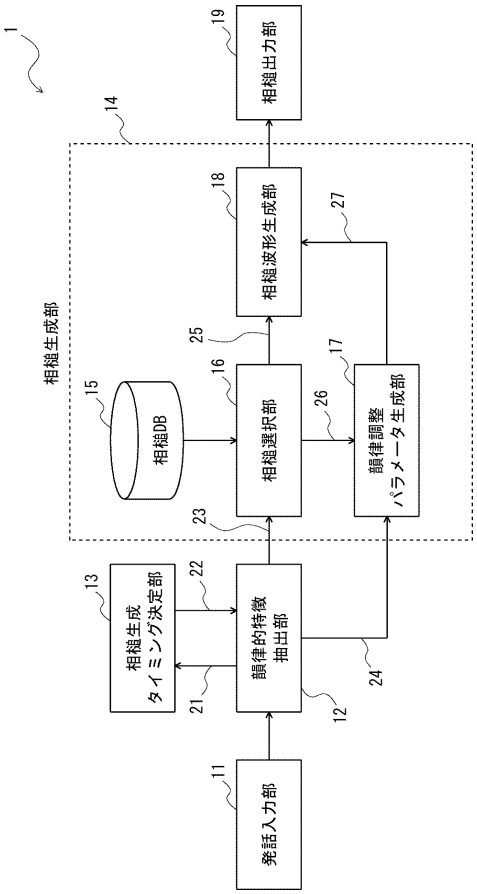
以上、本発明を上記実施形態に即して説明したが、本発明は上記実施の形態の構成にのみ限定されるものではなく、本願特許請求の範囲の請求項の発明の範囲内で当業者であればなし得る各種変形、修正、組み合わせを含むことは勿論である。

【符号の説明】

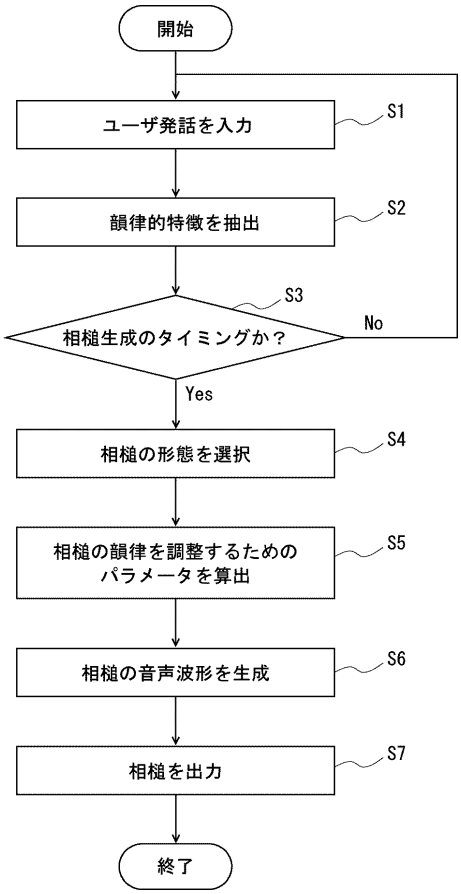
【００５４】

１　音声対話システム
１１　発話入力部
１２　韻律的特徴抽出部
１３　相槌生成タイミング決定部
１４　相槌生成部
１５　相槌データベース
１６　相槌選択部
１７　韻律調整パラメータ生成部
１８　相槌波形生成部
１９　相槌出力部
２１　抽出した韻律的特徴
２２　相槌生成タイミング情報
２３　相槌選択信号
２４　特徴量
２５　相槌情報
２６　相関係数に関する情報
２７　韻律調整パラメータ
３１　ユーザ
３２　ロボット
３３　ユーザ発話の音声波形
３４　相槌の音声波形

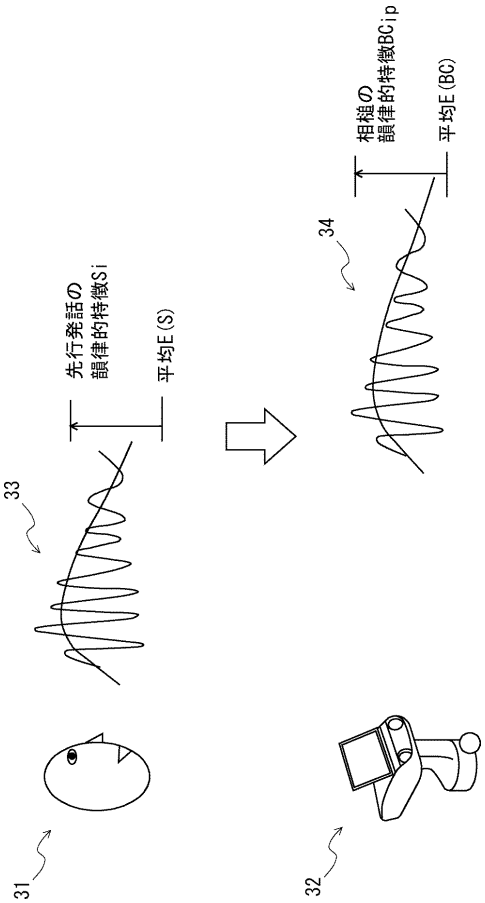
【図１】



【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	α(1,1)	α(1,2)	α(1,3)	α(1,4)
	はー	α(2,1)	α(2,2)	α(2,3)	α(2,4)
	…	…	…	…	…
応答系	ふーん	α(k,1)	α(k,2)	α(k,3)	α(k,4)
	はい	α(k+1,1)	α(k+1,2)	α(k+1,3)	α(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

審査官　鈴木　圭一郎

(56)参考文献　特開２００４－０８６００１（ＪＰ，Ａ）
特開２０１１－２１７０１８（ＪＰ，Ａ）
特開２００２－０４１０８４（ＪＰ，Ａ）
特開平１１－１７５０８２（ＪＰ，Ａ）
特開２００３－２２８４４９（ＪＰ，Ａ）
東海林圭輔，対話に関するリズムや同調作用を考慮した音声対話システム，情報処理学会研究報告，日本，一般社団法人情報処理学会，２００６年　5月１１日，Ｖｏｌ．２００６　Ｎｏ．４０，p43-48
西村良太，応答タイミングを考慮した雑談音声対話システム　A spoken dialog system for chat-like conversations considering response timing，第４６回　言語・音声理解と対話処理研究会資料，日本，一般社団法人人工知能学会，２００６年　3月　3日，ＳＩＧ－ＳＬＵＤ－Ａ５０３，p21-26

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－１３／１０
Ｇ１０Ｌ　１５／００－１５／３４
Ｇ１０Ｌ　１７／００－１７／２６