

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6696923号
(P6696923)

(45) 発行日 令和2年5月20日(2020.5.20)

(24) 登録日 令和2年4月27日(2020.4.27)

(51) Int.Cl.		F I			
G 1 0 L	13/08	(2013.01)	G 1 0 L	13/08	1 2 4
G 1 0 L	15/22	(2006.01)	G 1 0 L	15/22	3 0 0 U

請求項の数 6 (全 13 頁)

<p>(21) 出願番号 特願2017-40580 (P2017-40580)</p> <p>(22) 出願日 平成29年3月3日(2017.3.3)</p> <p>(65) 公開番号 特開2018-146715 (P2018-146715A)</p> <p>(43) 公開日 平成30年9月20日(2018.9.20)</p> <p>審査請求日 平成30年10月10日(2018.10.10)</p>	<p>(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</p> <p>(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</p> <p>(74) 代理人 100103894 弁理士 冢入 健</p> <p>(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</p> <p>(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</p>
---	---

最終頁に続く

(54) 【発明の名称】 音声対話装置、その処理方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

ユーザの音声を認識する音声認識手段と、
 前記音声認識手段により認識された音声に基づいて、前記ユーザの音声に対する応答文を生成する応答文生成手段と、
 前記ユーザとの対話間に挿入するフィラー語を生成するフィラー生成手段と、
 前記応答文生成手段により生成された応答文及び前記フィラー生成手段により生成されたフィラー語を出力する出力手段と、
 を備える音声対話装置であって、
 前記応答文生成手段により生成された応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類する分類手段を備え、
 前記出力手段が、第1の前記応答文に続く前記ユーザの音声後、前記フィラー語を出力し、第2の前記応答文を出力する場合において、
 前記分類手段は、前記第1の応答文を、前記発話パターンのうちのいずれかに分類し、
 前記フィラー生成手段は、前記分類手段により分類された発話パターンに基づいて、前記フィラー語を生成する、
 ことを特徴とする音声対話装置。

10

【請求項2】

請求項1記載の音声対話装置であって、
 前記発話パターンと、該発話パターンに対応付けられた特徴量の種類に関する情報と、

20

を含むテーブル情報を記憶する記憶手段と、

前記分類手段により分類された発話パターンに対応付けられた特徴量の種類に関する情報に基づいて、先行又は後続の発話の特徴量を算出する特徴量算出手段と、

を更に備え、

前記フィルア生成手段は、前記特徴量算出手段により算出された特徴量に基づいて、前記フィルア語を生成する、

を特徴とする音声対話装置。

【請求項 3】

請求項 2 記載の音声対話装置であって、

前記特徴量の種類に関する情報は、先行発話の韻律情報、先行発話の言語情報、後続発話の言語情報、及び、後続発話の韻律情報のうち少なくとも 1 つを含む、

ことを特徴とする音声対話装置。

10

【請求項 4】

請求項 2 又は 3 記載の音声対話装置であって、

前記記憶手段は、少なくとも 1 つの前記フィルア語を含み該フィルア語の系統を示すフィルア系に特徴量がそれぞれ対応付けられたフィルア形態情報を記憶しており、

前記フィルア生成手段は、前記分類手段により分類された発話パターンに基づいて、前記フィルア系の数を絞り込み、該絞り込んだフィルア系の中から、前記特徴量算出手段により算出された特徴量に対応付けられ 1 つのフィルア系を選択し、該選択したフィルア系に含まれる前記フィルア語を選択することで、前記フィルア語を生成する、

ことを特徴とする音声対話装置。

20

【請求項 5】

ユーザの音声を認識する音声認識手段と、

前記音声認識手段により認識された音声に基づいて、前記ユーザの音声に対する応答文を生成する応答文生成手段と、

前記ユーザとの対話間に挿入するフィルア語を生成するフィルア生成手段と、

前記応答文生成手段により生成された応答文及び前記フィルア生成手段により生成されたフィルア語を出力する出力手段と、を備える音声対話装置の処理方法であって、

前記出力手段が、第 1 の前記応答文に続く前記ユーザの音声後、前記フィルア語を出力し、第 2 の前記応答文を出力する場合において、

前記第 1 の応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類するステップと、

前記分類された発話パターンに基づいて、前記フィルア語を生成するステップと、

を含む

ことを特徴とする音声対話装置の処理方法。

30

【請求項 6】

ユーザの音声を認識する音声認識手段と、

前記音声認識手段により認識された音声に基づいて、前記ユーザの音声に対する応答文を生成する応答文生成手段と、

前記ユーザとの対話間に挿入するフィルア語を生成するフィルア生成手段と、

前記応答文生成手段により生成された応答文及び前記フィルア生成手段により生成されたフィルア語を出力する出力手段と、を備える音声対話装置のプログラムであって、

前記出力手段が、第 1 の前記応答文に続く前記ユーザの音声後、前記フィルア語を出力し、第 2 の前記応答文を出力する場合において、

前記第 1 の応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類する処理と、

前記分類された発話パターンに基づいて、前記フィルア語を生成する処理と、

をコンピュータに実行させることを特徴とする音声対話装置のプログラム。

40

【発明の詳細な説明】

【技術分野】

50

【 0 0 0 1 】

本発明は、ユーザと音声対話を行う音声対話装置、その処理方法及びプログラムに関する。

【背景技術】

【 0 0 0 2 】

対話間にフィラー語（場つなぎ語）を挿入して不自然に間延びしないようにする音声対話装置が知られている（特許文献1参照）。

【先行技術文献】

【特許文献】

【 0 0 0 3 】

【特許文献1】特開2014-191030号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 4 】

しかしながら、対話間の待ち時間が生じたときの場つなぎ語として、形式的なフィラー語を出力している。このため、そのフィラー語が対話内容に合わず、対話の自然性が損なわれる虞がある。

【 0 0 0 5 】

本発明は、このような問題点を解決するためになされたものであり、より自然な対話を行うことができる音声対話装置、その処理方法及びプログラムを提供することを主たる目的とする。

【課題を解決するための手段】

【 0 0 0 6 】

上記目的を達成するための本発明の一態様は、ユーザの音声認識する音声認識手段と、前記音声認識手段により認識された音声に基づいて、前記ユーザの音声に対する応答文を生成する応答文生成手段と、前記ユーザとの対話間に挿入するフィラー語を生成するフィラー生成手段と、前記応答文生成手段により生成された応答文及び前記フィラー生成手段により生成されたフィラー語を出力する出力手段と、を備える音声対話装置であって、前記応答文生成手段により生成された応答文を、予め設定され発話の種類を示す所定の発話パターンの中のいずれかに分類する分類手段を備え、前記出力手段が、第1の前記応答文に続く前記ユーザの音声後、前記フィラー語を出力し、第2の前記応答文を出力する場合において、前記分類手段は、前記第1の応答文を、前記発話パターンの中のいずれかに分類し、前記フィラー生成手段は、前記分類手段により分類された発話パターンに基づいて、前記フィラー語を生成する、ことを特徴とする音声対話装置である。

本態様によれば、第1の応答文の発話パターンに応じてフィラー語を生成することで、その後のユーザの音声内容に対応した、最適なフィラー語を生成することができる。この最適なフィラー語を出力することで、より自然な対話を行うことができる。

この一態様において、前記発話パターンと、該発話パターンに対応付けられた特徴量の種類に関する情報と、を含むテーブル情報を記憶する記憶手段と、前記分類手段により分類された発話パターンに対応付けられた特徴量の種類に関する情報に基づいて、先行又は後続の発話の特徴量を算出する特徴量算出手段と、を更に備え、前記フィラー生成手段は、前記特徴量算出手段により算出された特徴量に基づいて、前記フィラー語を生成してもよい。

これにより、第1の応答文を、所定の発話パターンの中のいずれかに分類し、分類された発話パターンに対応した最適な先行又は後続の発話の特徴量を算出し、算出した特徴量を用いてその対話に最適なフィラー語を生成できる。

この一態様において、前記特徴量の種類に関する情報は、先行発話の韻律情報、先行発話の言語情報、後続発話の言語情報、及び、後続発話の韻律情報のうち少なくとも1つを含んでいてもよい。

これにより、処理負荷の軽減と適切なフィラー語の生成との両立を図り、対話のリズム

10

20

30

40

50

及び自然性を向上させることができる。

この一態様において、前記記憶手段は、少なくとも1つの前記フィルター語を含み該フィルター語の系統を示すフィルター系に特徴量がそれぞれ対応付けられたフィルター形態情報を記憶しており、前記フィルター生成手段は、前記分類手段により分類された発話パターンに基づいて、前記フィルター系の数を絞り込み、該絞り込んだフィルター系の中から、前記特徴量算出手段により算出された特徴量に対応付けられ1つのフィルター系を選択し、該選択したフィルター系に含まれる前記フィルター語を選択することで、前記フィルター語を生成してもよい。

このように、事前にフィルター系の数を絞り込むことで、さらに処理負荷を軽減することができる。

上記目的を達成するための本発明の一態様は、ユーザの音声認識する音声認識手段と、前記音声認識手段により認識された音声に基づいて、前記ユーザの音声に対する応答文を生成する応答文生成手段と、前記ユーザとの対話間に挿入するフィルター語を生成するフィルター生成手段と、前記応答文生成手段により生成された応答文及び前記フィルター生成手段により生成されたフィルター語を出力する出力手段と、を備える音声対話装置の処理方法であって、前記出力手段が、第1の前記応答文に続く前記ユーザの音声後、前記フィルター語を出力し、第2の前記応答文を出力する場合において、前記第1の応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類するステップと、前記分類された発話パターンに基づいて、前記フィルター語を生成するステップと、を含むことを特徴とする音声対話装置の処理方法であってもよい。

上記目的を達成するための本発明の一態様は、ユーザの音声認識する音声認識手段と、前記音声認識手段により認識された音声に基づいて、前記ユーザの音声に対する応答文を生成する応答文生成手段と、前記ユーザとの対話間に挿入するフィルター語を生成するフィルター生成手段と、前記応答文生成手段により生成された応答文及び前記フィルター生成手段により生成されたフィルター語を出力する出力手段と、を備える音声対話装置のプログラムであって、前記出力手段が、第1の前記応答文に続く前記ユーザの音声後、前記フィルター語を出力し、第2の前記応答文を出力する場合において、前記第1の応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類する処理と、前記分類された発話パターンに基づいて、前記フィルター語を生成する処理と、をコンピュータに実行させることを特徴とする音声対話装置のプログラムであってもよい。

【発明の効果】

【0007】

本発明によれば、より自然な対話を行うことができる音声対話装置、音声対話方法及びプログラムを提供することができる。

【図面の簡単な説明】

【0008】

【図1】本発明の一実施形態に係る音声対話装置の概略的なシステム構成を示すブロック図である。

【図2】発話パターンの一例を示す図である。

【図3】装置発話及びユーザ発話の一例を示す図である。

【図4】先行の装置発話と後続のユーザ発話との対応の一例を示す図である。

【図5】発話パターンと特徴量の種類とを対応づけたテーブル情報の一例である。

【図6】フィルター系の一例を示す図である。

【図7】本発明の一実施形態に係る音声対話装置の処理方法のフロー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】

【0009】

以下、図面を参照して本発明の実施形態について説明する。

図1は、本発明の一実施形態に係る音声対話装置の概略的なシステム構成を示すブロック図である。本実施形態に係る音声対話装置1は、例えば、ロボット、PC(PersonalCo

10

20

30

40

50

mputer)、携帯端末(スマートフォン、タブレットなど)等に搭載され、ユーザと対話を行う。

【0010】

音声対話装置1は、例えば、ユーザの音声を認識する音声認識部2と、音声の構文を解析する構文解析部3と、ユーザの音声に対する応答文を生成する応答文生成部4と、フィルター語を生成するフィルター生成部5と、音声を出力する音声出力部6と、発話の分類を行う分類部7と、特徴量を算出する特徴量算出部8と、データを記憶する記憶部9と、を備えている。

【0011】

なお、音声対話装置1は、例えば、演算処理等を行うCPU(Central Processing Unit)、CPUによって実行される演算プログラム等が記憶されたROM(Read Only Memory)やRAM(Random Access Memory)からなるメモリ、外部と信号の入出力を行うインターフェイス部(I/F)、などからなるマイクロコンピュータを中心にして、ハードウェア構成されている。CPU、メモリ、及びインターフェイス部は、データバスなどを介して相互に接続されている。

10

【0012】

音声認識部2は、音声認識手段の一具体例である。音声認識部2は、例えば、マイクを介して入力されたユーザの音声情報に基づいて音声認識処理を行い、ユーザの音声情報をテキスト化し文字列情報として認識する。

【0013】

例えば、音声認識部2は、マイクを介して入力されたユーザの音声情報をデジタル化する。音声認識部2は、そのデジタル化した情報から発話区間を検出し、検出した発話区間の音声情報に対して、統計言語モデルなどを参照してパターンマッチングを行うことで音声認識を行う。

20

【0014】

ここで、統計言語モデルは、例えば、単語の出現分布やある単語の次に出現する単語の分布等、言語表現の出現確率を計算するための確率モデルであり、形態素単位で連結確率を学習したものである。統計言語モデルは、記憶部9などに予め記憶されている。音声認識部2は、ユーザの音声情報の各形態素に対してその品詞種類(名詞、形容詞、動詞、副詞など)を付加した品詞情報付き形態素情報を生成する。音声認識部2は、認識したユーザの音声情報を構文解析部3に出力する。

30

【0015】

記憶部9は記憶手段の一具体例である。記憶部9は、例えば、メモリなどで構成されている。

【0016】

構文解析部3は、音声認識部2により認識された音声情報の構文を解析する。構文解析部3は、例えば、一般的な形態素解析器を用いて音声認識されたユーザの音声情報を示す文字列情報に対して形態素解析などを行い、文字列情報の意味解釈を行う。構文解析部3は、文字列情報の解析結果(形態素情報、係り受け情報、認識結果のテキストなどを含む)を応答文生成部4に出力する。

40

【0017】

応答文生成部4は、応答文生成手段の一具体例である。応答文生成部4は、構文解析部3により解析された音声情報の構文に基づいて、ユーザの音声情報に対する応答文を生成する。

【0018】

応答文生成部4は、例えば、構文解析部3から出力される文字列情報の解析結果に基づいて、ユーザの音声情報に対する応答文を生成する。より、具体的には、応答文生成部4は、文字列情報「囲碁を打ちますよ。」から一つの述語項「囲碁を打つ」を抜き出す。構文解析部3は、抜き出した形態素列「囲碁 を 打つ」を特徴ベクトルとし、SVM(support vector machine)モデルを用いて、文字列情報を解析して予め定めた二つの分類の

50

いずれかに属するかを判別する。

【0019】

応答文生成部4は、文字列情報を解析した結果、予め定めた分類のうち一方の分類に属すると判別された場合、応答文「いいですね。」を生成する。一方、応答文生成部4は、文字列情報を解析した結果、予め定めた分類のうち他方の分類に属すると判別された場合、応答文「大変ですね。」を生成する。なお、上述した応答文の生成方法は一例であり、これに限定されず、予め定めた分類の数を増やすなど、任意の生成方法を用いることができる。また、応答文生成部4は、予め定めた分類を「ポジティブ」と「ネガティブ」のように感情として感情判別を行っても良い。応答文生成部4は、生成した応答文を音声出力部6に出力する。

10

【0020】

フィルア生成部5は、フィルア生成手段の一具体例である。フィルア生成部5は、ユーザと音声対話装置1との対話間に挿入するフィルア語を生成する。フィルア語は、例えば「あー」、「えーと」、「うーん」、等の、場つなぎ語である。フィルア生成部5は、生成したフィルア語を音声出力部6に出力する。

【0021】

音声出力部6は、出力手段の一具体例である。音声出力部6は、応答文生成部4により生成された応答文及びフィルア生成部5により生成されたフィルア語を、音声で出力する。音声出力部6は、例えば、スピーカなどを用いてユーザに対して応答文及びフィルア語の音声を出力する。

20

【0022】

音声出力部6は、例えば、フィルア生成部5により生成されたフィルア語と、応答文生成部4により生成された応答文と、に基づいて、フィルア語及び応答文の音声を合成し、合成した音声を出力する。音声出力部6は、予め記憶部9などに設定された音声ファイルを選択することで、フィルア語及び応答文の音声を出力してもよい。

【0023】

ところで、従来の音声対話装置においては、例えば、対話間の待ち時間が生じたときの場つなぎ語として、形式的なフィルア語を出力している。このため、そのフィルア語が対話内容に合わず、対話の自然性が損なわれる虞がある。

【0024】

これに対し、本実施形態に係る音声対話装置1は、応答文生成部4により生成された応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類する分類部7を備えている。音声出力部6が、第1の応答文に続くユーザの音声後、フィルア語を出力し、第2の応答文を出力する。この場合において、分類部7は、第1の応答文を、発話パターンのうちのいずれかに分類する。フィルア生成部5は、分類部7により分類された発話パターンに基づいて、フィルア語を生成する。

30

【0025】

第1の応答文の内容に応じて、後続のユーザの音声及び第2の応答文の内容は変化する。したがって、第1の応答文の発話パターンに基づいてフィルア語を生成することで、ユーザの音声及び第2の応答文の内容に対応した、最適なフィルア語を生成することができる。この最適なフィルア語を出力することで、より自然な対話を行うことができる。例えば、雑談時などにおいては、最適なフィルア語を出力することで、形式的なフィルア語とは異なる、人間らしい自然な場つなぎを行うことができる。

40

【0026】

分類部7は、分類手段の一具体例である。分類部7は、音声出力部6により出力される応答文(以下、装置発話)及びユーザの音声(以下、ユーザ発話)を、複数の発話パターンのうちのいずれかに分類する。発話パターンは、予め設定され発話の種類を示すものである。

【0027】

発話パターンは、例えば、ユーザ又は音声対話装置1が行う対話行為に基づいて、複数

50

のパターンに分類されており、記憶部 9 などに設定されている。対話行為は、ユーザ発話と装置発話の関係性を示したものである。

【 0 0 2 8 】

発話パターンは、例えば、図 2 に示す如く、(1) 真偽を問う質問パターン「Yes / No 質問パターン」、(2) 集合の中から選ぶ質問パターン「Wh (When、Where、What、Why) 質問パターン」、(3) 回答、同意、受諾などを含むパターン「回答パターン」、(4) 自己開示、依頼などを含むパターン「開示パターン」、(5) 挨拶、導入などを含むパターン「それ以外パターン」、などに分類される。なお、この発話パターンは一例であり、これに限定されない。

【 0 0 2 9 】

図 3 は、装置発話及びユーザ発話の一例を示す図である。図 3 において、装置発話及びユーザ発話が時系列に並べられている。時刻 $T - 3 \sim T - 1$ は現在時刻 T 前の過去の装置発話及びユーザ発話（先行発話）であり、時刻 $T + 1 \sim T + 3$ は現在時刻 T 後の装置発話及びユーザ発話（後続発話）である。

【 0 0 3 0 】

例えば、図 3 に示す如く、分類部 7 は、時刻 $T - 3 \sim T + 3$ までのユーザ発話及び装置発話に基づいて、ルールベースや分類モデルを用いて、ユーザ発話及び装置発話を分類できる。より具体的には、分類部 7 は、時刻 T の装置発話「何か趣味はありますか？」を、ユーザへの「Wh 質問パターン」に分類し、それに続く時刻 $T + 1$ のユーザ発話「囲碁を打ちますよ。」を、「回答パターン」に分類する。分類部 7 は、その後、状態系列推定モデル（条件付き確率場）などを用いて、さらに続く時刻 $T + 3$ の装置発話「いいですね」を「回答パターン」に分類する。

【 0 0 3 1 】

分類部 7 は、先行するユーザ発話と、それに続く装置発話とを組み合わせる分類を行ってもよい。例えば、発話パターンが「yes/no 質問パターン」である場合、それに続く発話パターンは「回答パターン」に必然的に決まる。したがって、分類部 7 は、先行するユーザ発話が「yes/no 質問パターン」である場合、それに続く装置発話を「回答パターン」に分類する。同様に、分類部 7 は、先行するユーザ発話が「回答パターン」である場合、それに続く装置発話を「回答パターン」に分類する。

【 0 0 3 2 】

分類部 7 は、ルールベースを用いて、先行の装置発話に応じて、その後続のユーザ発話を分類してもよい。例えば、図 4 に示す如く、分類部 7 は、ルールベースを用いて、時刻 T の装置発話が「yes/no 質問パターン」あるいは「wh 質問パターン」の場合、時刻 $T + 1$ のユーザ発話を「回答パターン」に分類する。同様に、分類部 7 は、ルールベースを用いて、時刻 T の装置発話が「回答パターン」の場合、時刻 $T + 1$ のユーザ発話を「開示パターン」に分類する。

【 0 0 3 3 】

分類部 7 は、事前に品詞とその系列、述語項とその系列、及び、正解となる分類を付与したデータに基づいて分類モデルを学習し、その学習結果を用いて時刻 $T + 1$ のユーザ発話を分類してもよい。分類部 7 は、 $T - 3 \sim T + 1$ 発話までの系列に基づいて、時刻 $T + 2$ の装置発話の分類を行うことができる。

【 0 0 3 4 】

特徴量算出部 8 は、特徴量算出手段の一具体例である。特徴量算出部 8 は、分類部 7 により分類された第 1 の応答文（装置発話）の発話パターンに対応付けられた特徴量の種類に関する情報に基づいて、先行又は後続の発話の特徴量を算出する。

【 0 0 3 5 】

ここで、対話シーンによって、その対話間に挿入する最適なフィラー形態は異なる。例えば、時刻 T の第 1 の応答文に続く時刻 $T + 1$ のユーザの音声後、時刻 $T + 2$ のフィラー語を出力し、時刻 $T + 3$ の第 2 の応答文を出力する対話シーンでは、第 1 の応答文の内容によって、その後に挿入する最適なフィラー形態は異なる。さらに、その最適なフィラー

10

20

30

40

50

形態を判別するのに最適な特徴の捉え方も異なる。

【0036】

例えば、真偽を問う「yes / No質問パターン」の後に挿入するフィラー形態は、先行発話の韻律情報（発話長など）が深くかかわる。このため、先行発話の韻律情報の特徴を捉え、その特徴に合ったフィラー形態を選択し、フィラー語を生成することが、より自然な対話を行う上で特に有効となる。

【0037】

したがって、特徴量算出部8は、分類部7により分類された第1の応答文の発話パターンに対応付けられた特徴量の種類に関する情報に基づいて、先行又は後続の発話の特徴量を算出する。フィラー生成部5は、特徴量算出部8により算出された特徴量に応じて、最適なフィラー語を生成する。

10

【0038】

これにより、第1の応答文を、所定の発話パターンのうちのいずれかに分類し、分類された発話パターンに対応した最適な先行又は後続の発話の特徴量を算出し、算出した特徴量を用いてその対話に最適なフィラー語を生成できる。

【0039】

発話パターンには、例えば、図5に示す如く、特徴量の種類に関する情報が対応付けられている。図5は、発話パターンと特徴量の種類と、を対応づけたテーブル情報の一例である。テーブル情報は、記憶部9などに予め設定されている。なお、図5に示すテーブル情報は一例であり、これに限定されない。

20

【0040】

図5において、例えば、「Yes / No質問パターン」は、「先行発話の韻律情報」に対応付けられ、「Wh質問パターン」は、「先行発話の韻律情報」及び「先行発話の言語情報」に対応付けられ、「回答パターン」は、「後続発話の言語情報」に対応付けられている。

【0041】

ここで、「先行発話の韻律情報」は、先行発話における、例えば、音声の発話末の所定時間（100 msec程度）におけるF0（基本周波数）、パワーの最大値、最小値、平均、一次回帰係数（傾き）、発話長、ポーズ長、話速などを含む。

【0042】

「先行発話の言語情報」は、先行発話における、例えば、発話末の品詞、発話末の節境界ラベル、単語数、文節数などを含む。節境界ラベルは、節境界直後の切れ目の大きさという観点から、「思います」などの絶対境界、「けれども」などの強境界、「なら」などの弱境界、という3つのレベルに区分されている。「後続発話の言語情報」では、後続発話における、冒頭発話の品詞、単語数、文節数などを含む。

30

【0043】

特徴量算出部8は、分類部7により分類された発話パターンと、記憶部9のテーブル情報と、に基づいて、分類された発話パターンに対応する特徴量（特徴ベクトル）を算出する。特徴量算出部8は、音声認識部2からのユーザの音声情報、構文解析部3からの文字列情報、応答文生成部4からの応答文（テキストデータ）などに基づいて、先行発話の韻律情報、先行発話の言語情報、及び後続発話の言語情報の特徴ベクトルを算出する。

40

【0044】

例えば、特徴量算出部8は、記憶部9のテーブル情報に基づいて、分類部7により分類された発話パターン「回答パターン」に対応する「後続発話の言語情報」（冒頭発話の品詞、単語数、文節数など）の特徴ベクトルを算出する。

【0045】

図5に示す如く、1つの発話パターンに複数の特徴量の種類に関する情報が対応付けられていてもよい。例えば、発話パターン「Wh質問パターン」には、先行発話の韻律情報及び先行発話の言語情報が対応付けられている。この場合、特徴量算出部8は、記憶部9のテーブル情報に基づいて、分類部7により分類された発話パターン「Wh質問パターン」に対応する、「先行発話の韻律情報」の特徴ベクトルと、「先行発話の言語情報」の特徴ベ

50

クトルと、を夫々算出し、これら特徴ベクトルを統合した特徴ベクトルを算出する。

【0046】

ここで、後続発話の言語情報に基づいた処理は、後続であり構文解析を行うため、処理負荷が高くリードタイムが大きくなる（5秒程度）。同様に、先行発話の言語情報に基づいた処理も、構文解析を行うため、処理負荷が高くリードタイムが大きくなる（4秒程度）。これらの高処理負荷によって対話のリズムや自然性が低下する虞がある。このため可能であれば先行発話の韻律情報のみに基づいた処理が好ましい。しかし、適切なフィラー語を生成するためには言語情報を利用した方が良い場合もある。

【0047】

したがって、本実施形態においては、処理負荷の軽減と適切なフィラー語の生成との両立を図るように、テーブル情報において、「韻律情報」及び「言語情報」の対応付けを行っている。これにより、処理負荷の軽減と適切なフィラー語の生成との両立を図り、対話のリズム及び自然性を向上させることができる。

10

【0048】

上述の如く、発話の特徴量は、言語情報における言語的特徴（品詞列等）と、韻律情報における音響的特徴（ピッチ等）と、に大別することができる。言語的特徴は、上述の如く、処理負荷が高く出力に時間がかかる。このため、本実施形態においては、主として音響的特徴を用いつつ、適宜、言語的特徴を用いている。これにより発話の特徴を効果的に捉え最適なフィラー語を生成しつつ、同時に処理負荷の軽減を図ることができる。

【0049】

20

フィラー生成部5は、特徴量算出部8により算出された特徴量に基づいて、フィラー語を生成する。フィラー生成部5は、例えば、予め設定されたフィラー形態情報を用いて、フィラー語を生成する。

【0050】

フィラー形態情報は、例えば、記憶部9などに予め設定されている。フィラー形態情報において、例えば、図6に示す、フィラー語の各系統「以下、フィラー系」に特徴量がそれぞれ対応付けられている。フィラー系は、固有系、応答詞系、指示詞系、副詞系、気づき系、その他、なし、などを含む。

【0051】

各フィラー系には、単数あるいは複数のフィラー語が対応付けられている。例えば、固有系には、フィラー語「えっと」、「えー」などが対応付けられている。各フィラー系と特徴量とは、決定木の組み合わせのような構造をもつRandom forestなどの学習器を用いて、対応付けられるのが好ましい。なお、ロジスティック回帰などの学習器を用いてもよく、任意の学習器を用いることができる。

30

【0052】

フィラー生成部5は、特徴量算出部8により算出された特徴量と、記憶部9のフィラー形態情報と、に基づいて、算出された特徴量に対応付けられたフィラー系を選択し、さらに、該フィラー系の中から1つのフィラー語を選択することで、フィラー語を生成する。これにより、対話シーンの特徴を示す特徴量を用いて、その対話シーンに合わせたフィラー形態を選択でき、選択したフィラー形態に基づいて最適なフィラー語を生成できる。

40

【0053】

フィラー生成部5は、例えば、選択したフィラー系の中から1つのフィラー語をランダムに選択する。フィラー生成部5は、選択したフィラー系の中から、所定の優先順位に従って、1つのフィラー語を選択してもよく、選択方法は任意でよい。

【0054】

フィラー生成部5は、特徴量算出部8により算出された特徴量と、記憶部9のフィラー形態情報と、に基づいて、7種類のフィラー系の中から、特徴量に対応付けられ1つのフィラー系を選択しているが、これに限定されない。フィラー生成部5は、分類部7により分類された発話パターンに基づいて、フィラー系の数を絞り込み、絞り込んだ各フィラー系の中から、特徴量に対応付けられ1つのフィラー系を選択し、フィラー語を生成しても

50

よい。このように、事前にフィラー系の数を絞り込むことで、さらに処理負荷を軽減することができる。

【0055】

例えば、発話パターンが「回答」である場合、必然的に、それに対する最適なフィラー系は、予め4種類のフィラー系（固有系、副詞系、その他、なし）に絞り込むことができる。したがって、フィラー生成部5は、分類部7により分類された発話パターン「回答」に基づいて、7種類のフィラー系を4種類のフィラー系（固有系、副詞系、その他、なし）に絞り込み、絞り込んだフィラー系の中から、特徴量に対応付けられた1つのフィラー系を選択し、フィラー語を生成する。例えば、発話パターンに、絞り込んだフィラー系を対応づけた情報が記憶部9などに設定されている。フィラー生成部5は、記憶部9のこの情報に基づいて、上記フィラー系の絞り込みを行ってもよい。

10

【0056】

フィラー生成部5は、特徴量算出部8により算出された特徴量と、記憶部9のフィラー形態情報と、に基づいて、1つのフィラー系「なし」を選択した場合、フィラー語を生成しない。この場合、フィラー語は出力されない。これは、フィラー語を出力しないことがその対話のシーンに合っているという意味である。

【0057】

フィラー生成部5は、上述のように生成したフィラー語を音声出力部6に出力する。音声出力部6は、例えば、フィラー生成部5により生成された時刻T+2のフィラー語を出力し、応答文生成部4により生成された時刻T+3の応答文を出力する。

20

【0058】

図7は、本実施形態に係る音声対話装置の処理方法のフローを示すフローチャートである。

分類部7は、記憶部9のテーブル情報に基づいて、応答文生成部により生成された第1の応答文を、テーブル情報に設定された発話パターンのうちのいずれかに分類する（ステップS101）。

【0059】

特徴量算出部8は、分類部7により分類された発話パターンと、記憶部9のテーブル情報と、に基づいて、分類された発話パターンに対応する特徴量の種類の特徴ベクトルを算出する（ステップS102）。

30

【0060】

フィラー生成部5は、特徴量算出部8により算出された特徴ベクトルと、記憶部9のフィラー形態情報と、に基づいて、フィラー形態情報に設定された複数のフィラー系の中から、1つのフィラー系を選択し、さらに、該フィラー系の中からフィラー語を選択することで、フィラー語を生成する（ステップS103）。

【0061】

音声出力部6は、フィラー生成部5により生成されたフィラー語を出力する（ステップS104）。

【0062】

以上、本実施形態に係る音声対話装置1は、応答文生成部4により生成された応答文を、予め設定され発話の種類を示す所定の発話パターンのうちのいずれかに分類する分類部7を備えている。音声出力部6が、第1の応答文に続くユーザの音声後、フィラー語を出力し、第2の応答文を出力する。この場合において、分類部7は、第1の応答文を、発話パターンのうちのいずれかに分類する。フィラー生成部5は、分類部7により分類された発話パターンに基づいて、フィラー語を生成する。

40

【0063】

第1の応答文の発話パターンに応じてフィラー語を生成することで、その後のユーザの音声内容に対応した、最適なフィラー語を生成することができる。この最適なフィラー語を出力することで、より自然な対話を行うことができる。

【0064】

50

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。

【0065】

上記実施形態において、出力手段として、音声を出力する音声出力部6が適用されているが、これに限定されない。出力手段として、例えば、文字を出力する表示部が適用されてもよい。

【0066】

本発明は、例えば、図7に示す処理を、CPUにコンピュータプログラムを実行させることにより実現することも可能である。

【0067】

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(Read Only Memory)、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(random access memory))を含む。

【0068】

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】

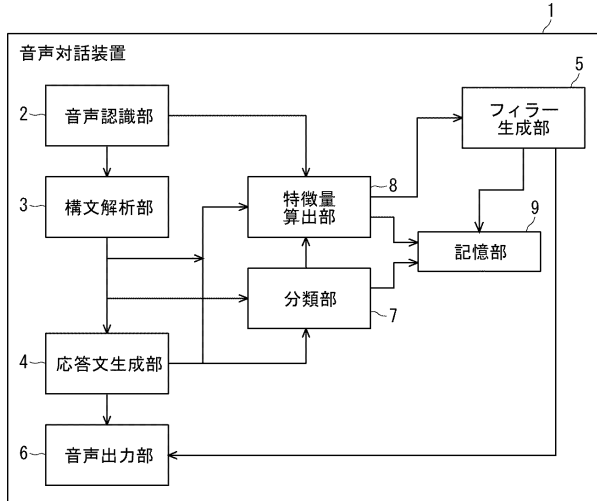
【0069】

1 音声対話装置、2 音声認識部、3 構文解析部、4 応答文生成部、5 フィラ
ー生成部、6 音声出力部、7 分類部、8 特徴量算出部、9 記憶部

10

20

【図1】



【図2】

発話パターン	内容
(1) Yes/no質問	真偽を問う質問
(2) Wh質問	集合の中から選ぶ質問
(3) 回答	回答、同意、受諾など
(4) 開示	自己開示、依頼など
(5) それ以外	挨拶、導入など

【図5】

発話パターン	特徴量の種類
(1) Yes/no質問	先行発話の韻律情報
(2) Wh質問	先行発話の韻律情報 先行発話の言語情報
(3) 回答	後続発話の言語情報
(4) 開示	先行発話の韻律情報
(5) それ以外	先行発話の韻律情報 先行発話の言語情報

【図6】

フィラー系	フィラー語
固有系	えっと、えー
応答詞系	うーん
指示詞系	あのー、そのー
副詞系	まー、もー
気づき系	あ、え
その他	なんていうか
なし	

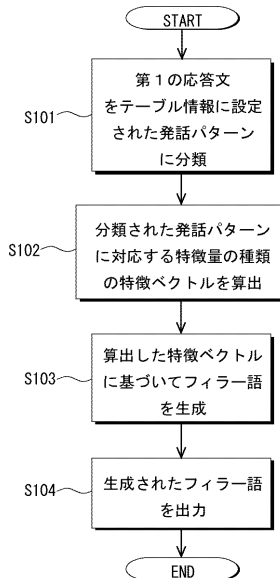
【図3】

時刻	装置発話	ユーザ発話	発話パターン
T-3	こんにちは。		(5) それ以外
T-2		こんにちは。	(5) それ以外
T-1	私は料理が好きなんですが、		(4) 開示
T		何か趣味はありますか？	(2) Wh質問
T+1		囲碁を打ちますよ。	(3) 回答
T+2	あ、		(フィラー)
T+3	いいですね。		(3) 回答

【図4】

装置発話	後続ユーザ発話
(1) Yes/no質問	(3) 回答
(2) Wh質問	(3) 回答
(3) 回答	(4) 開示
(4) 開示	(3) 回答
(5) それ以外	(1) Yes/no質問 / (2) Wh質問

【図7】



フロントページの続き

- (72)発明者 中西 亮輔
京都府京都市左京区吉田本町3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

審査官 山下 剛史

- (56)参考文献 特開2010-262147(JP,A)
米国特許出願公開第2008/0167874(US,A1)
特開2014-48443(JP,A)
特開2004-151527(JP,A)
太田健吾他, ユーザ発話の音響情報に基づく雑談対話システムの応答種別選択, 日本音響学会2017年春季研究発表会講演論文集[CD-ROM], 2017年 3月 1日, p.71-74
- (58)調査した分野(Int.Cl., DB名)
G10L 13/00 - 15/34