(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7280512号
(P7280512)

(45)発行日　令和5年5月24日(2023.5.24)

(24)登録日　令和5年5月16日(2023.5.16)

(51)国際特許分類		FI		
G06F 3/01 (2006.01)		G06F	3/01	570
G06F 3/16 (2006.01)		G06F	3/16	650
A63H 11/00 (2006.01)		A63H	11/00	Z
G06T 13/40 (2011.01)		G06T	13/40	
G10L 25/21 (2013.01)		G10L	25/21	

請求項の数　4　（全75頁）　最終頁に続く

(21)出願番号　特願2019-572290(P2019-572290)
(86)(22)出願日　平成31年2月15日(2019.2.15)
(86)国際出願番号　PCT/JP2019/005639
(87)国際公開番号　WO2019/160100
(87)国際公開日　令和1年8月22日(2019.8.22)
審査請求日　令和2年7月1日(2020.7.1)
(31)優先権主張番号　特願2018-26516(P2018-26516)
(32)優先日　平成30年2月16日(2018.2.16)
(33)優先権主張国・地域又は機関
日本国(JP)
(31)優先権主張番号　特願2018-26517(P2018-26517)
(32)優先日　平成30年2月16日(2018.2.16)
(33)優先権主張国・地域又は機関
日本国(JP)
(31)優先権主張番号　特願2018-97338(P2018-97338)
最終頁に続く

(73)特許権者　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(74)代理人　110001634
弁理士法人志賀国際特許事務所
(72)発明者　石井 亮
東京都千代田区大手町一丁目５番１号
日本電信電話株式会社内
(72)発明者　東中 竜一郎
東京都千代田区大手町一丁目５番１号
日本電信電話株式会社内
(72)発明者　片山 太一
東京都千代田区大手町一丁目５番１号
日本電信電話株式会社内
(72)発明者　富田 準二
最終頁に続く

(54)【発明の名称】　非言語情報生成装置及びプログラム

(57)【特許請求の範囲】
【請求項１】
時刻情報付きのテキスト特徴量と、学習済みの非言語情報生成モデルとに基づいて、前記時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成する非言語情報生成部を含む非言語情報生成装置であって、
前記時刻情報付きのテキスト特徴量は、テキストから抽出された特徴量と、前記テキストの所定単位毎に付与された時刻を表す時刻情報とを含んで構成され、前記非言語情報は、前記テキストに応じた行動を発現させる発現部を制御するための情報であり、
前記所定単位は、前記テキストをモーラ数又はアクセント数のいずれかで分割したものである
非言語情報生成装置。
【請求項２】
時刻情報付きの音声特徴量と、学習済みの非言語情報生成モデルとに基づいて、前記時刻情報付きの音声特徴量に対応する時刻情報付きの非言語情報を生成する非言語情報生成部を含む非言語情報生成装置であって、
前記時刻情報付きの音声特徴量は、音声情報から抽出された特徴量と、前記音声情報が発せられるときの所定単位毎の時刻を表す時刻情報とを含んで構成され、前記非言語情報は、前記音声情報に応じた行動を発現させる発現部を制御するための情報であり、
前記所定単位は、前記音声情報をモーラ数又はアクセント数のいずれかで分割したものである

非言語情報生成装置。
【請求項３】
入力されたテキストに対してテキスト解析を行い、当該テキストを音声として外部に出力すると仮定した場合の再生時刻を、所定単位毎に前記テキストに対して前記時刻情報として付与し、前記時刻情報付きのテキスト特徴量を抽出する特徴量抽出部を更に含む、
請求項１に記載の非言語情報生成装置。
【請求項４】
コンピュータを、請求項１乃至請求項３のいずれか一項に記載の非言語情報生成装置が備える各部として機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、非言語情報生成装置、非言語情報生成モデル学習装置、方法、及びプログラムに関する。
本願は、２０１８年２月１６日に日本へ出願された特願２０１８－０２６５１６号、２０１８年２月１６日に日本へ出願された特願２０１８－０２６５１７号、２０１８年５月２１日に日本へ出願された特願２０１８－０９７３３８号、２０１８年５月２１日に日本へ出願された特願２０１８－０９７３３９号、および、２０１８年１２月７日に日本へ出願された特願２０１８－２３０３１０号に基づき優先権を主張し、それらの内容をここに援用する。
【背景技術】
【０００２】
コミュニケーションにおいて、言語行動に加えて非言語行動は感情や意図を伝達する上で重要な機能を有している。そのため、コミュニケーションロボット、コミュニケーションエージェントにおいても、ユーザと円滑なコミュニケーションを行うためには非言語行動を表出することが望まれている。このような背景から、発話に合わせた非言語行動をあらかじめＤＢ（Database）に登録しておき、発話の再生に合わせて非言語行動を表出させる手法が提案されている（例えば、特許文献１を参照）。
【先行技術文献】
【特許文献】
【０００３】
【文献】日本特開２００３－１７３４５２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
しかし、上記特許文献１に記載の手法では、あらかじめ発話と非言語行動の情報（ジェスチャとセリフの組で構成されるアクションコマンド）を予め作成・登録しておく必要があるため、データ作成のコストが高かった。
【０００５】
本発明は、上記の事情に鑑みてなされたもので、音声情報又はテキスト情報の少なくとも一方と非言語情報との対応付けを自動化することができる非言語情報生成装置、非言語情報生成モデル学習装置、方法、及びプログラムを提供することを目的とする。
【課題を解決するための手段】
【０００６】
上記目的を達成するために、第１の態様に係る非言語情報生成装置は、時刻情報付きのテキスト特徴量と、学習済みの非言語情報生成モデルとに基づいて、前記時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成する非言語情報生成部を含む非言語情報生成装置であって、前記時刻情報付きのテキスト特徴量は、テキストから抽出された特徴量と、前記テキストの所定単位毎に付与された時刻を表す時刻情報とを含んで構成され、前記非言語情報は、前記テキストに応じた行動を発現させる発現部を制御する

ための情報である。
【０００７】
また、第２の態様に係る非言語情報生成装置は、時刻情報付きの音声特徴量と、学習済みの非言語情報生成モデルとに基づいて、前記時刻情報付きの音声特徴量に対応する時刻情報付きの非言語情報を生成する非言語情報生成部を含む非言語情報生成装置であって、前記時刻情報付きの音声特徴量は、音声情報から抽出された特徴量と、前記音声情報が発せられるときの所定単位毎の時刻を表す時刻情報とを含んで構成され、前記非言語情報は、前記音声情報に応じた行動を発現させる発現部を制御するための情報である。
【０００８】
本発明の非言語情報生成装置は、入力されたテキストに対してテキスト解析を行い、当該テキストを音声として外部に出力すると仮定した場合の再生時刻を、所定単位毎に前記テキストに対して前記時刻情報として付与し、前記時刻情報付きのテキスト特徴量を抽出する特徴量抽出部を更に含むようにすることができる。
【０００９】
上記のテキスト特徴量は、前記テキストから抽出される対話行為及びシソーラス情報のうち少なくとも１つを含むようにすることができる。
【００１０】
本発明の非言語情報生成装置は、前記非言語情報を発現させる発現部を更に含み、前記非言語情報生成部は、前記時刻情報付きの非言語情報が、当該非言語情報に付与された時刻情報に基づいて前記発現部から発現されるように前記発現部を制御するようにすることができる。
【００１１】
第３の態様に係る非言語情報生成モデル学習装置は、発声者の音声に対応するテキストを表すテキスト情報と、前記テキストの所定単位毎に付与された時刻を表す時刻情報とを取得する学習用情報取得部と、前記発声者が前記テキストに対応する発声を行った際の、前記発声者の行動に関する情報を表す非言語情報と、前記行動が行われた際の時刻を表す時刻情報であって、かつ前記非言語情報に対応する前記時刻情報とを取得し、時刻情報付きの非言語情報を作成する非言語情報取得部と、前記学習用情報取得部によって取得された前記テキスト情報及び該テキスト情報に対応する前記時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出する学習用特徴量抽出部と、前記学習用特徴量抽出部によって抽出された前記時刻情報付きのテキスト特徴量に基づいて、前記非言語情報取得部によって取得された前記時刻情報付きの非言語情報を生成するための、非言語情報生成モデルを学習する学習部と、を備えている。
【００１２】
また、第４の態様に係る非言語情報生成モデル学習装置は、発声者の音声に対応する音声情報と、前記音声情報が発せられるときの所定単位毎の時刻を表す時刻情報とを取得する学習用情報取得部と、前記発声者が前記音声に対応する発声を行った際の、前記発声者の行動に関する情報を表す非言語情報と、前記行動が行われた際の時刻を表す時刻情報であって、かつ前記非言語情報に対応する前記時刻情報とを取得し、時刻情報付きの非言語情報を作成する非言語情報取得部と、前記学習用情報取得部によって取得された前記音声情報及び該音声情報に対応する前記時刻情報から、該音声情報の特徴量を表す、時刻情報付きの音声特徴量を抽出する学習用特徴量抽出部と、前記学習用特徴量抽出部によって抽出された前記時刻情報付きの音声特徴量に基づいて、前記非言語情報取得部によって取得された前記時刻情報付きの非言語情報を生成するための、非言語情報生成モデルを学習する学習部と、を備えている。
【００１３】
また、第５の態様に係る非言語情報生成方法は、非言語情報生成部が、時刻情報付きのテキスト特徴量と、学習済みの非言語情報生成モデルとに基づいて、前記時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成するステップを含み、前記時刻情報付きのテキスト特徴量は、テキストから抽出された特徴量と、前記テキストの所定

単位毎に付与された時刻を表す時刻情報とを含んで構成され、前記非言語情報は、前記テキストに応じた行動を発現させる発現部を制御するための情報である。

【００１４】

また、第６の態様に係る非言語情報生成方法は、非言語情報生成部が、時刻情報付きの音声特徴量と、学習済みの非言語情報生成モデルとに基づいて、前記時刻情報付きの音声特徴量に対応する時刻情報付きの非言語情報を生成するステップを含み、前記時刻情報付きの音声特徴量は、音声情報から抽出された特徴量と、前記音声情報が発せられるときの所定単位毎の時刻を表す時刻情報とを含んで構成され、前記非言語情報は、前記音声情報に応じた行動を発現させる発現部を制御するための情報である。

【００１５】

また、本発明のプログラムは、コンピュータを、上記の非言語情報生成装置が備える各部として機能させるためのプログラムである。

【００１６】

第７の態様に係る非言語情報生成モデル学習装置は、発声者の音声に対応するテキストを表すテキスト情報と、前記テキストの所定単位毎に付与された時刻を表す時刻情報とを取得する学習用情報取得部と、前記発声者が前記テキストに対応する発声を行った際の、前記発声者の前記発声の聞き手の行動に関する情報を表す非言語情報と、前記行動が行われた際の時刻を表す時刻情報であって、かつ前記非言語情報に対応する前記時刻情報とを取得し、時刻情報付きの非言語情報を作成する非言語情報取得部と、前記学習用情報取得部によって取得された前記テキスト情報及び該テキスト情報に対応する前記時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出する学習用特徴量抽出部と、前記学習用特徴量抽出部によって抽出された前記時刻情報付きのテキスト特徴量に基づいて、前記非言語情報取得部によって取得された前記時刻情報付きの非言語情報を生成するための、非言語情報生成モデルを学習する学習部と、を含む。

【００１７】

上記のテキスト特徴量は、前記テキストから抽出される対話行為及びシソーラス情報のうち少なくとも１つを含むようにすることができる。

【００１８】

また、第８の態様に係る非言語情報生成モデル学習装置は、発声者の音声に対応する音声情報と、前記音声情報が発せられるときの所定単位毎の時刻を表す時刻情報とを取得する学習用情報取得部と、前記発声者が前記音声に対応する発声を行った際の、前記発声者の前記発声の聞き手の行動に関する情報を表す非言語情報と、前記行動が行われた際の時刻を表す時刻情報であって、かつ前記非言語情報に対応する前記時刻情報とを取得し、時刻情報付きの非言語情報を作成する非言語情報取得部と、前記学習用情報取得部によって取得された前記音声情報及び該音声情報に対応する前記時刻情報から、該音声情報の特徴量を表す、時刻情報付きの音声特徴量を抽出する学習用特徴量抽出部と、前記学習用特徴量抽出部によって抽出された前記時刻情報付きの音声特徴量に基づいて、前記非言語情報取得部によって取得された前記時刻情報付きの非言語情報を生成するための、非言語情報生成モデルを学習する学習部と、を含む。

【００１９】

また、第９の態様に係る非言語情報生成モデル学習方法は、学習用情報取得部が、発声者の音声に対応するテキストを表すテキスト情報と、前記テキストの所定単位毎に付与された時刻を表す時刻情報とを取得するステップと、非言語情報取得部が、前記発声者が前記テキストに対応する発声を行った際の、前記発声者の前記発声の聞き手の行動に関する情報を表す非言語情報と、前記行動が行われた際の時刻を表す時刻情報であって、かつ前記非言語情報に対応する前記時刻情報とを取得し、時刻情報付きの非言語情報を作成するステップと、学習用特徴量抽出部が、前記学習用情報取得部によって取得された前記テキスト情報及び該テキスト情報に対応する前記時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出するステップと、学習部が、前記学習用特徴量抽出部によって抽出された前記時刻情報付きのテキスト特徴量に基づいて、前記非言語情

報取得部によって取得された前記時刻情報付きの非言語情報を生成するための、非言語情報生成モデルを学習するステップと、を含む。

【００２０】

また、第１０の態様に係る非言語情報生成モデル学習方法は、学習用情報取得部が、発声者の音声に対応する音声情報と、前記音声情報が発せられるときの所定単位毎の時刻を表す時刻情報とを取得するステップと、非言語情報取得部が、前記発声者が前記音声に対応する発声を行った際の、前記発声者の前記発声の聞き手の行動に関する情報を表す非言語情報と、前記行動が行われた際の時刻を表す時刻情報であって、かつ前記非言語情報に対応する前記時刻情報とを取得し、時刻情報付きの非言語情報を作成するステップと、学習用特徴量抽出部が、前記学習用情報取得部によって取得された前記音声情報及び該音声情報に対応する前記時刻情報から、該音声情報の特徴量を表す、時刻情報付きの音声特徴量を抽出するステップと、学習部が、前記学習用特徴量抽出部によって抽出された前記時刻情報付きの音声特徴量に基づいて、前記非言語情報取得部によって取得された前記時刻情報付きの非言語情報を生成するための、非言語情報生成モデルを学習するステップと、を含む。

【００２１】

また、本発明のプログラムは、コンピュータを、上記の非言語情報生成モデル学習装置が備える各部として機能させるためのプログラムである。

【発明の効果】

【００２２】

以上説明したように、本発明に係る非言語情報生成装置、非言語情報生成モデル学習装置、方法、及びプログラムによれば、音声情報又はテキスト情報の少なくとも一方と非言語情報との対応付けを自動化することができる。

【図面の簡単な説明】

【００２３】

【図１】第１の実施形態に係る非言語情報生成モデル学習装置の構成の一例を示すブロック図である。

【図２】学習データの取得方法を説明するための説明図である。

【図３】本実施形態で用いる音声特徴量とテキスト特徴量とを説明するための説明図である。

【図４】第１の実施形態に係る非言語情報生成装置の構成の一例を示すブロック図である。

【図５】第１の実施形態に係る学習処理の流れの一例を示すフローチャートである。

【図６】第１の実施形態に係る非言語情報生成処理の流れの一例を示すフローチャートである。

【図７】第２の実施形態に係る非言語情報生成モデル学習装置の構成の一例を示すブロック図である。

【図８】第２の実施形態に係る学習用情報取得部の詳細な構成例を示す図である。

【図９】第２の実施形態に係る非言語情報生成装置の構成の一例を示すブロック図である。

【図１０】第２の実施形態に係る情報取得部の詳細な構成例を示す図である。

【図１１】第２の実施形態に係る学習処理の流れの一例を示すフローチャートである。

【図１２】第２の実施形態に係る非言語情報生成処理の流れの一例を示すフローチャートである。

【図１３Ａ】第３の実施形態の学習用情報取得部及び学習用特徴量抽出部を説明するための説明図である。

【図１３Ｂ】第３の実施形態の情報取得部及び特徴量抽出部を説明するための説明図である。

【図１４Ａ】第３の実施形態の変形例を説明するための説明図である。

【図１４Ｂ】第３の実施形態の変形例を説明するための説明図である。

【図１５】他の実施形態の構成の組み合わせを説明するための説明図である。

【図１６】他の実施形態の構成の組み合わせを説明するための説明図である。

【図１７】第７の実施形態に係る非言語情報生成モデル学習装置の構成の一例を示すブロック図である。
【図１８】本実施形態で用いる付加情報を説明するための説明図である。
【図１９】第７の実施形態に係る非言語情報生成装置の構成の一例を示すブロック図である。
【図２０】第７の実施形態に係る学習処理の流れの一例を示すフローチャートである。
【図２１】第７の実施形態に係る非言語情報生成処理の流れの一例を示すフローチャートである。
【図２２】第７の実施形態の他の例に係る非言語情報生成モデル学習装置の構成の一例を示すブロック図である
【図２３】第７の実施形態の他の例に係る非言語情報生成装置の構成の一例を示すブロック図である。
【図２４】付加情報により時刻情報を変更する方法を説明するための図である。
【図２５】第８の実施形態に係る非言語情報生成モデル学習装置の構成の一例を示すブロック図である
【図２６Ａ】詳細な時刻情報を付与する方法を説明するための図である。
【図２６Ｂ】第８の実施形態に係る非言語情報生成装置の構成の一例を示すブロック図である。
【図２７】第８の実施形態に係る学習処理の流れの一例を示すフローチャートである。
【図２８】第８の実施形態に係る非言語情報生成処理の流れの一例を示すフローチャートである。
【図２９】第９の実施形態に係る学習処理の流れの一例を示すフローチャートである。
【図３０】第９の実施形態に係る非言語情報生成処理の流れの一例を示すフローチャートである。
【図３１】第１１の実施形態に係る非言語情報生成装置の構成の一例を示すブロック図である。
【図３２】第１１の実施形態に係る表示画面の一例を示す図である。
【図３３】時刻情報の変更指示を説明するための図である。
【図３４】時刻情報の変更指示を説明するための図である。
【図３５】第１１の実施形態に係る非言語情報生成処理の流れの一例を示すフローチャートである。
【図３６】第１１の実施形態に係る表示制御処理の流れの一例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】
【００２４】
以下、図面を参照して、本発明を実施するための形態の一例について詳細に説明する。
【００２５】
＜本実施形態の概要＞
【００２６】
本発明の実施形態では、人間がコミュニケーションを行う際、音声情報及び音声情報に含まれる言語情報の発信と、非言語行動とが共起して起こるという特性を利用する。具体的には、本実施形態では、発話の音声情報及び発話の内容を表すテキスト情報の少なくとも一方を入力Ｘとして、話者の発話と一緒に生成される話者の非言語行動を表す非言語情報を出力Ｙとして、入力Ｘから出力Ｙを機械学習により生成する。非言語情報とは、行動に関する情報であって、言語そのもの以外の情報とする。非言語行動の例としては、例えば、頭部動作、視線方向、ハンドジェスチャ、上半身の動作、及び下半身の動作の種別（クラス）などが挙げられる。
【００２７】
本実施形態で得られる非言語情報は、人間と同じような身体性を持ち、人間とコミュニケーションを行う、コミュニケーションロボット、コミュニケーションエージェント、ゲ

ームやインタラクティブシステムで利用されるＣＧ（Computer Graphics）アニメーションにおけるジェスチャ生成等で用いられる。

【００２８】

［第１の実施形態］

【００２９】

＜非言語情報生成モデル学習装置の構成＞

【００３０】

図１は、第１の実施形態に係る非言語情報生成モデル学習装置１０の構成の一例を示すブロック図である。図１に示すように、本実施形態に係る非言語情報生成モデル学習装置１０は、ＣＰＵ（Central Processing Unit）と、ＲＡＭ（Random Access Memory）と、後述する学習処理ルーチンを実行するためのプログラムを記憶したＲＯＭ（Read Only Memory）とを備えたコンピュータで構成されている。非言語情報生成モデル学習装置１０は、機能的には、学習用入力部２０と、学習用演算部３０を備えている。

【００３１】

学習用入力部２０は、学習用の音声情報と、言語とは異なる行動に関する情報を表す学習用の非言語情報とを受け付ける。

【００３２】

本実施形態の非言語情報生成モデル学習装置１０の入力データである学習用の音声情報と学習用の非言語情報との組み合わせを表す学習データは、例えば、図２のようなシーンにおいて、発話中の話者の音声情報（Ｘ）を取得するのと同時に、所定の計測装置を用いて発話中の話者の非言語情報（Ｙ）も取得することで作成される。なお、音声情報（Ｘ）は、発話中の話者が外部へ発話する際の音声情報に相当する。

【００３３】

学習用演算部３０は、学習用入力部２０によって受け付けた学習データに基づいて、時刻情報付きの非言語情報を生成するための非言語情報生成モデルを生成する。学習用演算部３０は、図１に示されるように、学習用情報取得部３１と、学習用特徴量抽出部３２と、非言語情報取得部３３と、生成パラメータ抽出部３４と、学習部３５と、学習済みモデル記憶部３６とを備えている。

【００３４】

学習用情報取得部３１は、学習用入力部２０によって受け付けられた学習用の音声情報を取得する。また、学習用情報取得部３１は、学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【００３５】

学習用特徴量抽出部３２は、学習用情報取得部３１によって取得された学習用の音声情報及び時刻情報から、当該学習用の音声情報の特徴量を表す、時刻情報付きの学習用の音声特徴量を抽出する。

【００３６】

例えば、学習用特徴量抽出部３２は、学習用情報取得部３１によって取得された学習用の音声情報に対して所定の音声情報処理を行い、基本周波数（Ｆ０）、パワー、ＭＦＣＣ（Mel Frequency Cepstral Coefficients）等を音声特徴量として抽出する。これらの音声特徴量は、図３に示されるように、任意の時間の窓幅$T_{A,w}$と音声情報とを用いて算出することができる。図３に示されるように、これらの多次元の音声特徴量を

【００３７】

【数１】

$$X_A^{t,T_{A,w}}$$

【００３８】

とする。ここで、

【００３９】

【数２】

$$X_A^{t,T_{A,w}}$$

【００４０】

は、時刻$t_{A,s}$から、窓幅$T_{A,w}$の時間に対応する音声情報から算出された音声特徴量とする。なお、窓幅は、すべての音声特徴量で同じである必要は無く、別々に特徴量を抽出しても良い。これらの音声特徴量の抽出法は一般的であり、様々な手法がすでに提案されている（例えば、参考文献１を参照。）。このため、あらゆる手法を利用して良い。

【００４１】

［参考文献１］：中川 聖一,「音声言語処理と自然言語処理」,2013/3/1, コロナ社

【００４２】

非言語情報取得部３３は、学習用入力部２０によって受け付けられた学習用の非言語情報を取得し、当該学習用の非言語情報が表す行動が行われるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【００４３】

非言語情報取得部３３は、学習用の非言語情報として、頷き、顔向き、ハンドジェスチャ、視線、表情、及び身体の姿勢等に関する情報を取得する。頷き、顔向き、ハンドジェスチャ、視線、表情、及び身体の姿勢等に関する情報を表すパラメータの一例を以下に示す。

【００４４】

【表１】

種別	パラメータ
頷き	頷きの有無Y_N^P、回数Y_N^T、深さY_N^D
顔向き	*Yaw, Roll, Pitch* の角度（Y_{HD}^{yaw}、Y_{HD}^{roll}、Y_N^{pitch}）
ハンドジェスチャ	モーション ID（Y_{HG}^{ID}）
視線（眼球）	X,Y の位置（Y_{EB}^x、Y_{EB}^y）
表情（FACS）	47 つの AU の強度（Y_{FACS}^i (i=1,…,47)）
身体の姿勢	前後・左右の姿勢位置（Y_{BP}^{FB}、Y_{BP}^{RL}）

【００４５】

なお、FACSとは（Facial Action Coding System：顔面表情記号化システム）を表し、AUとは（Action Unit；活動単位）を表す。例えば、AU1は「眉の内側を持ち上げ（AU1）」というように、非言語情報がラベルによって表現される。上記以外の非言語情報としては、例えば、話者の視線行動、頭部動作、呼吸動作、及び口形変化等が挙げられる。

【００４６】

上記に示すように、非言語情報は、身体の関節、位置、及び動き等のイベントに関するあらゆるパラメータであって良い。計測手法は様々な手法が考えられており、どのような手法を用いても良い（例えば、参考文献２および参考文献３を参照）。

【００４７】

［参考文献２］：牧川方昭、南部雅幸、塩澤成弘、岡田志麻、吉田正樹、「ヒト心身状態の計測技術―人に優しい製品開発のための日常計測」，2010/10/1，コロナ社

［参考文献３］：Shihong Xia, Lin Gao, Yu-Kun Lai, Ming-Ze Yuan, Jinxiang Chai, "A Survey on Human Performance Capture and Animation",Journal of Computer Science and Technology, Volume 32, Issue 3, pp 536-554, (2017).

【００４８】

上記図２に示されるように、例えば、顔の向きは、人間や対話ヒューマノイドの頭部方向を頭部計測装置（ヘッドトラッカ）などの行動キャプチャ装置で計測することが可能である。また、ＣＧなどのアニメーションデータから顔の向きを取得することも可能である。また、図２に示されるように、顔の向きは、例えば、Yaw, Roll, Pitchの3軸の角度で表現される。

【００４９】

生成パラメータ抽出部３４は、非言語情報取得部３３で取得された学習用の非言語情報が表すパラメータ（感覚尺度）を離散化して、時刻情報付きの離散化された非言語情報を抽出する。例えば、顔向きは、Yaw, Roll, Pitchの角度情報によって表されるが、任意の閾値α，β（α＜β）を予め定めて、以下に示すような名義尺度に変換しても良い。なお、以下に示す例はYawのみが提示されている。

【００５０】

－α＜Yaw＜α：正面

【００５１】

α≦Yaw＜β：少し左を向いている

【００５２】

β≦Yaw：大きく左を向いている

【００５３】

－β＜Yaw≦－α：少し右を向いている

【００５４】

－β≧Yaw：大きく右を向いている

【００５５】

このように、非言語情報取得部３３で取得された学習用の非言語情報を離散化し、時刻情報を付けたものを、多次元ベクトル

【００５６】

【数３】

$$Y^{t_{N,s},t_{N,e}}$$

【００５７】

とする。$t_{N,s}$，$t_{N,e}$は、非言語情報の得られた開始時刻、終了時刻とする。

【００５８】

学習部３５は、学習用特徴量抽出部３２によって抽出された時刻情報付きの学習用の音声特徴量と、生成パラメータ抽出部３４によって取得された、時刻情報付きの離散化された非言語情報とに基づいて、時刻情報付きの音声特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。

【００５９】

具体的には、学習部３５は、学習用特徴量抽出部３２によって抽出された時刻情報付きの学習用の音声特徴量

【００６０】

【数４】

$$X_A^{t,T_{A,w}}$$

【００６１】
を入力として、時刻情報付きの非言語情報
【００６２】
【数５】

$$Y^{t_{N,s},t_{N,e}}$$

【００６３】
を出力する非言語情報生成モデルを構築する。非言語情報生成モデルを構築するにあたり、あらゆる機械学習手法を用いて良いが、本実施形態ではＳＶＭ（Support Vector Machine）を用いる。例えば、ＳＶＭを利用して、
【００６４】
【数６】

$$Y^{t_{N,s},t_{N,e}}$$

【００６５】
における各次元のパラメータのクラス分類器を構築したり、ＳＶＭを回帰へ適用したＳＶＲ（Support Vector Machine for Regression）による回帰モデルを構築したりする。
【００６６】
　また、本実施の形態では、非言語情報が表す行動の種類毎に、当該種類の行動の有無を推定するためのＳＶＭモデルを作成する。
【００６７】
　なお、非言語情報生成モデルは、どのような時間分解能で、どの時刻のパラメータを使用して、非言語情報を推定するかは任意である。ここで、任意の時間区間Ｔ１～Ｔ２におけるジェスチャ
【００６８】
【数７】

$$Y^{T1,T2}$$

【００６９】
を推定する場合を例に使用する特徴量の一例を示す。推定対象となる時刻Ｔ１～Ｔ２の時刻で得られる、音声特徴量
【００７０】
【数８】

$$X_A^{T1,T_{A,w}} \sim X_A^{T2,T_{A,w}}$$

【００７１】
と出力となるジェスチャ
【００７２】
【数９】

$$Y^{T1,T2}$$

【００７３】
をペアとし、これらのペアの複数データである学習データを用いて学習が行われる。その学習済みの非言語情報生成モデルを、
【００７４】
【数１０】

$$M^{T1,T2}$$

【００７５】
とする。
【００７６】
学習済みモデル記憶部３６には、学習部３５によって学習された学習済みの非言語情報生成モデルが格納される。学習済みの非言語情報生成モデルは、時刻情報付きの音声特徴量から時刻情報付きの非言語情報を生成する。
【００７７】
＜非言語情報生成装置のシステム構成＞
【００７８】
図４は、第１の実施形態に係る非言語情報生成装置４０の構成の一例を示すブロック図である。図４に示すように、本実施形態に係る非言語情報生成装置４０は、ＣＰＵ（Central Processing Unit）と、ＲＡＭ（Random Access Memory）と、後述する非言語情報生成処理ルーチンを実行するためのプログラムを記憶したＲＯＭ（Read Only Memory）とを備えたコンピュータで構成されている。非言語情報生成装置４０は、機能的には、入力部５０と、演算部６０と、発現部７０とを備えている。
【００７９】
入力部５０は、音声情報と当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報とを受け付ける。
【００８０】
演算部６０は、情報取得部６１と、特徴量抽出部６２と、学習済みモデル記憶部６３と、非言語情報生成部６４とを備えている。
【００８１】
情報取得部６１は、入力部５０によって受け付けられた、音声情報と当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報とを取得する。
【００８２】
特徴量抽出部６２は、学習用特徴量抽出部３２と同様に、情報取得部６１によって取得された音声情報及び時刻情報から、該音声情報の特徴量を表す、時刻情報付きの音声特徴量を抽出する。
【００８３】
学習済みモデル記憶部６３には、学習済みモデル記憶部３６に格納されている学習済みの非言語情報生成モデルと同一の学習済みの非言語情報生成モデルが格納されている。
【００８４】
非言語情報生成部６４は、特徴量抽出部６２によって抽出された時刻情報付きの音声特徴量と、学習済みモデル記憶部６３に格納された学習済みの非言語情報生成モデルとに基づいて、特徴量抽出部６２によって抽出された時刻情報付きの音声特徴量に対応する時刻情報付きの非言語情報を生成する。
【００８５】

例えば、非言語情報生成部６４は、学習済みモデル記憶部６３に格納された学習済みの非言語情報生成モデル

【００８６】

【数１１】

$$M^{T1,T2}$$

【００８７】

を用いて、時刻情報付きの音声特徴量としての任意の特徴量

【００８８】

【数１２】

$$X_A^{T1,T_{A,w}} \sim X_A^{T2,T_{A,w}}$$

【００８９】

を入力として、時刻情報付きの非言語情報としてのジェスチャ

【００９０】

【数１３】

$$Y^{T1,T2}$$

【００９１】

を取得する。

【００９２】

そして、非言語情報生成部６４は、生成された時刻情報付きの非言語情報が、当該非言語情報に付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御する。

【００９３】

具体的には、非言語情報生成部６４は、発現部７０において、ジェスチャ

【００９４】

【数１４】

$$Y^{T1,T2}$$

【００９５】

を任意の対象（例えば、アニメーションキャラクタ、ロボット等）の動作として反映させる。

【００９６】

発現部７０は、非言語情報生成部６４による制御に応じて、入力部５０によって受け付けられた音声情報と非言語情報生成部６４によって生成された非言語情報とを発現させる。

【００９７】

発現部７０の一例としては、コミュニケーションロボットや、ディスプレイに表示されるコミュニケーションエージェント、ゲームやインタラクティブシステムで利用されるＣＧアニメーション等が挙げられる。

【００９８】

＜非言語情報生成モデル学習装置１０の作用＞
【００９９】
次に、本実施形態に係る非言語情報生成モデル学習装置１０の作用について説明する。まず、非言語情報生成モデル学習装置１０の学習用入力部２０に、複数の学習用の音声情報と複数の学習用の非言語情報との組み合わせを表す学習データが入力されると、非言語情報生成モデル学習装置１０は、図５に示す学習処理ルーチンを実行する。
【０１００】
まず、ステップＳ１００において、学習用情報取得部３１は、学習用入力部２０によって受け付けられた複数の学習データのうちの、学習用の音声情報と当該学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。
【０１０１】
ステップＳ１０２において、非言語情報取得部３３は、学習用入力部２０によって受け付けられた複数の学習データのうちの、学習用の非言語情報と当該学習用の非言語情報が表す行動が行われるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。
【０１０２】
ステップＳ１０４において、学習用特徴量抽出部３２は、上記ステップＳ１００で取得された学習用の音声情報及び時刻情報から、時刻情報付きの学習用の音声特徴量を抽出する。
【０１０３】
ステップＳ１０６において、生成パラメータ抽出部３４は、上記ステップＳ１０２で取得された学習用の非言語情報及び時刻情報から、時刻情報付きの離散化された非言語情報を抽出する。
【０１０４】
ステップＳ１０８において、学習部３５は、上記ステップＳ１０４で抽出された時刻情報付きの学習用の音声特徴量と、上記ステップＳ１０６で取得された時刻情報付きの非言語情報とに基づいて、時刻情報付きの音声特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。
【０１０５】
ステップＳ１１０において、学習部３５は、上記ステップＳ１０８で得られた学習済みの非言語情報生成モデルを、学習済みモデル記憶部３６へ格納して、学習処理ルーチンを終了する。
【０１０６】
＜非言語情報生成装置４０の作用＞
【０１０７】
次に、本実施形態に係る非言語情報生成装置４０の作用について説明する。まず、非言語情報生成モデル学習装置１０の学習済みモデル記憶部３６に格納された学習済みの非言語情報生成モデルが、非言語情報生成装置４０へ入力されると、非言語情報生成装置４０の学習済みモデル記憶部６３へ学習済みの非言語情報生成モデルが格納される。そして、入力部５０に、非言語情報生成対象の音声情報が入力されると、非言語情報生成装置４０は、図６に示す非言語情報生成処理ルーチンを実行する。
【０１０８】
ステップＳ２００において、情報取得部６１は、入力部５０によって受け付けられた、音声情報と当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報とを取得する。
【０１０９】
ステップＳ２０２において、特徴量抽出部６２は、学習用特徴量抽出部３２と同様に、上記ステップＳ２００で取得された音声情報及び時刻情報から、時刻情報付きの音声特徴量を抽出する。
【０１１０】
ステップＳ２０４において、非言語情報生成部６４は、学習済みモデル記憶部６３に格

納された学習済みの非言語情報生成モデルを読み出す。

【０１１１】

ステップＳ２０６において、非言語情報生成部６４は、上記ステップＳ２０２で抽出された時刻情報付きの音声特徴量と、上記ステップＳ２０４で読み出された学習済みの非言語情報生成モデルとに基づいて、上記ステップＳ２０２で抽出された時刻情報付きの音声特徴量に対応する時刻情報付きの非言語情報を生成する。

【０１１２】

ステップＳ２０８において、非言語情報生成部６４は、上記ステップＳ２０６で生成された時刻情報付きの非言語情報が、当該非言語情報に付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御して、非言語情報生成処理ルーチンを終了する。

【０１１３】

以上説明したように、第１の実施形態に係る非言語情報生成装置４０によれば、音声情報及び時刻情報から時刻情報付きの音声特徴量を抽出し、抽出された時刻情報付きの音声特徴量と、時刻情報付きの非言語情報を生成するための学習済みの非言語情報生成モデルとに基づいて、時刻情報付きの音声特徴量に対応する時刻情報付きの非言語情報を生成する。これにより、音声情報と非言語情報との対応付けを自動的に行い、そのコストを低減させることができる。

【０１１４】

コミュニケーションロボットや対話エージェント等の発話音声、またはそれに対応するテキストに合わせ、ジェスチャ等の動作を行わせる際には、発話に合わせてどのようなタイミングで、どのような動作を行わせるかを決める必要がある。従来では、これらがシナリオとして全て人手で作成、設定されており、作成コストが高かった。

【０１１５】

これに対し、本実施形態では、時刻情報付きの音声特徴量から時刻情報付きの非言語情報を生成するための学習済みの非言語情報生成モデルを生成することにより、音声情報を入力とし、入力に対応する動作に対応する時刻情報付きの非言語情報（出力のタイミングが付与された非言語情報）が出力される。

【０１１６】

これにより、本実施形態によれば、音声情報から非言語情報を自動的に生成することが可能となるため、従来技術のように、発話に対して非言語情報を個別に登録する必要がなく、コストが大幅に削減される。また、本実施形態を用いることで、入力された音声情報に対し、人間らしい自然なタイミングでの非言語行動を生成可能であり、エージェント、及びロボット等の人間らしさや自然さの向上や、非言語行動による意図の伝達の円滑化、会話の活性化などが効果として得られる。

【０１１７】

また、予め学習済みの非言語情報生成モデルを用いて、発話音声またはテキストを入力とし、入力に対応する動作とそのタイミングの情報を出力する。これにより、シナリオ作成コストを下げることができる。また、実際の人間の動作を元に動作が生成されるので、より自然なタイミングで動きを再現できる。

【０１１８】

また、第１の実施形態に係る非言語情報生成モデル学習装置１０によれば、時刻情報付きの学習用の音声特徴量と時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きの音声特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。これにより、音声情報と非言語情報との対応付けのコストを低減させつつ、音声特徴量から非言語情報を生成するための非言語情報生成モデルを得ることができる。

【０１１９】

また、学習済みの非言語情報生成モデルを用いることにより、自然なタイミングでの非言語行動を生成することができる。

【０１２０】

上記第１の実施形態では音声特徴量から非言語情報を生成する場合を例に説明した。上記第１の実施形態では、話されている内容には踏み込まず、音声特徴量として表れる情報（例えば、感情など）を手掛かりに、必要最小限の構成で非言語情報を生成することが可能である。

【０１２１】

なお、上記第１の実施形態では、話している内容には踏み込まない（言語情報を用いない）ため、例えば動物にセンサをつけて非言語情報と音声情報（例えば、鳴き声等）とを取得し、動物型のロボットを動かしてもよい。

【０１２２】

［第２の実施形態］

【０１２３】

＜非言語情報生成モデル学習装置の構成＞

【０１２４】

次に、本発明の第２の実施形態について説明する。なお、第１の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。

【０１２５】

第２の実施形態では、入力として音声情報の代わりにテキスト情報を用いる。そして、テキスト情報から非言語情報を生成する非言語情報生成モデルの学習を行う点が第１の実施形態と異なる。なお、第２の実施形態で用いるテキスト情報は、話者が音声によって外部へ発話する際の、発話内容を表すテキスト情報である。

【０１２６】

図７は、第２の実施形態に係る非言語情報生成モデル学習装置２１０の構成の一例を示すブロック図である。図７に示すように、第２の実施形態に係る非言語情報生成モデル学習装置２１０は、ＣＰＵと、ＲＡＭと、後述する学習処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成されている。非言語情報生成モデル学習装置２１０は、機能的には、学習用入力部２２０と、学習用演算部２３０とを備えている。

【０１２７】

学習用入力部２２０は、学習用のテキスト情報と学習用の非言語情報とを受け付ける。

【０１２８】

学習用演算部２３０は、学習用入力部２２０によって受け付けた学習データに基づいて、時刻情報付きの非言語情報を生成するための非言語情報生成モデルを生成する。学習用演算部２３０は、図７に示されるように、学習用情報取得部２３１と、学習用特徴量抽出部２３２と、非言語情報取得部３３と、生成パラメータ抽出部３４と、学習部２３５と、学習済みモデル記憶部２３６とを備えている。

【０１２９】

学習用情報取得部２３１は、学習用のテキスト情報に対応する学習用の音声情報を取得し、当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。学習用情報取得部２３１は、図８に示されるように、学習用テキスト解析部２３７と、学習用音声合成部２３８とを備えている。

【０１３０】

学習用テキスト解析部２３７は、学習用のテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。例えば、学習用テキスト解析部２３７は、学習用のテキスト情報に対して形態素解析をはじめとするテキスト解析を行って、形態素毎に、単語表記（形態素）情報、品詞、カテゴリ情報、評価表現、感情表現、感性表現、擬音語・擬態語・擬声語、固有表現、主題、文字数、位置、及びシソーラス情報等を抽出し、文毎に、発話の対話行為等を抽出する。なお、形態素毎ではなく、文節毎に、単語表記（形態素）情報、品詞、カテゴリ情報、評価表現、感情表現、固有表現、主題、文字数、位置、及びシソーラス情報等を抽出してもよい。また、形態素や文節以外の任意の単位で、単語表記（形態素）情報、品詞、カテゴリ情報、評価表現、感情表現、感性表現、擬音語・擬

態語・擬声語、固有表現、主題、文字数、位置、及びシソーラス情報等を抽出してもよい。例えば、文字単位でもよく、英語であれば、空白で区切った文字列単位でもよいし、句単位でもよい。また、主題の抽出については、文ごとや発話単位ごとに行ってもよい。ここで、対話行為とは、発話における意図を抽象化し、ラベルとして抽象化したものである。主題とは、テキストにおける話題や焦点を表す情報である。文字数とは、形態素又は文節内の文字数である。位置とは、文頭、文末からの形態素又は文節の位置である。シソーラス情報とは、日本語語彙大系に基づく形態素又は文節内の単語のシソーラス情報である。これらのテキスト特徴量の抽出法は一般的なものでよく、様々な手法がすでに提案されている（上記参考文献１及び下記参考文献４～６を参照）。本実施形態では、これらの情報のうち、単語表記（形態素）情報、品詞、対話行為、文字数、位置、及びシソーラス情報を用いる場合を例に説明を行う。

【０１３１】

［参考文献４］：R Higashinaka, K Imamura, T Meguro, C Miyazaki, N Kobayashi, H Sugiyama, T Hirano， T Makino, Y Matsuo， "Towards an open-domain conversational system fully based on natural language processing", In Proceedings of International conference on Computational linguistics, pp. 928-939, 2014

【０１３２】

［参考文献５］：日本特開2014-222399号公報

【０１３３】

［参考文献６］：日本特開2015-045915号公報

【０１３４】

学習用音声合成部２３８は、学習用テキスト解析部２３７によって取得されたテキスト解析結果に基づいて、学習用のテキスト情報に対応する学習用の音声情報を合成する。例えば、学習用音声合成部２３８は、テキスト解析結果を用いて音声合成を行い、テキスト情報に対応する発話を生成し、学習用のテキスト情報に対応する学習用の音声情報とする。

【０１３５】

また、学習用音声合成部２３８は、学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。具体的には、学習用音声合成部２３８は、音声合成により生成された発話の音声の開始時刻から終了時刻に対応する時刻情報を取得する。この時刻情報は、発話に対応するテキスト情報の各形態素に対応したものとする。なお、テキスト情報に含まれる文字毎に開始時刻と終了時刻とを取得するようにしてもよい。

【０１３６】

学習用特徴量抽出部２３２は、学習用音声合成部２３８によって取得された学習用のテキスト情報及び時刻情報から、当該学習用のテキスト情報の特徴量を表す、時刻情報付きの学習用のテキスト特徴量を抽出する。具体的には、学習用特徴量抽出部２３２は、所定の解析単位毎に当該テキスト情報に対して時刻情報を付与して、時刻情報付きのテキスト特徴量を抽出する。

【０１３７】

具体的には、学習用特徴量抽出部２３２は、学習用音声合成部２３８によって出力された学習用のテキスト情報に対して文分割を行う。次に、学習用特徴量抽出部２３２は、学習用テキスト解析部２３７で得られた対話行為に関するテキスト特徴量

【０１３８】

【数１５】

$$X_D^{t_{S,s},t_{S,e}}$$

【０１３９】

を一文ごとに抽出する。なお、
【０１４０】
【数１６】

$$t_{S,s}, t_{S,e}$$

【０１４１】
は、一文に該当する発話の開始時刻及び終了時刻とする。
【０１４２】
　また、学習用特徴量抽出部２３２は、分割により得られた各文を構成する複数の形態素の各々について、当該形態素の単語表記情報、品詞、カテゴリ情報（例えば、名詞、固有表現、及び用言等）、評価表現、感情表現、固有表現、文字数、文中の位置、シソーラス情報等のうち、少なくとも単語表記情報を抽出する。そして、学習用特徴量抽出部２３２は、これらの多次元特徴量を
【０１４３】
【数１７】

$$X_P^{t_{P,s},t_{P,e}}$$

【０１４４】
とする。なお、
【０１４５】
【数１８】

$$t_{P,s}, t_{P,e}$$

【０１４６】
は、形態素単位に該当する発話音声の開始時刻及び終了時刻とする。時刻情報付きのテキスト特徴量の一例を図３に示す。図３に示されるように、テキスト情報の各形態素の開始時刻及び終了時刻が取得されている。
【０１４７】
　なお、分割により得られた各文を構成する複数の文節の各々について、当該文節の単語表記情報、品詞、カテゴリ情報（例えば、名詞、固有表現、及び用言等）、評価表現、感情表現、固有表現、文字数、文中の位置、シソーラス情報等を抽出してもよい。そして、学習用特徴量抽出部２３２は、これらの多次元特徴量を
【０１４８】
【数１９】

$$X_C^{t_{C,s},t_{C,e}}$$

【０１４９】
とする。なお、
【０１５０】
【数２０】

$$t_{C,s}, t_{C,e}$$

【０１５１】
は、文節単位に該当する発話音声の開始時刻及び終了時刻とする。
【０１５２】
学習用情報取得部２３１において、音声認識及び音声合成を行った際に得られる情報を流用してもよい。
【０１５３】
学習部２３５は、学習用特徴量抽出部２３２によって抽出された時刻情報付きの学習用のテキスト特徴量と、生成パラメータ抽出部３４によって抽出された時刻情報付きの学習用の離散化された非言語情報とに基づいて、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。
【０１５４】
具体的には、学習部２３５は、学習用特徴量抽出部２３２によって抽出された時刻情報付きの学習用のテキスト特徴量
【０１５５】
【数２１】

$$X_D^{t_{S,s},t_{S,e}}$$

【０１５６】
【数２２】

$$X_P^{t_{P,s},t_{P,e}}$$

【０１５７】
を入力として、非言語情報
【０１５８】
【数２３】

$$Y^{t_{N,s},t_{N,e}}$$

【０１５９】
を出力する非言語情報生成モデルを構築する。非言語情報生成モデルを構築する際には、あらゆる機械学習手法を用いて良いが、本実施形態ではＳＶＭを用いる。
【０１６０】
なお、非言語情報生成モデルは、どのような時間分解能で、どの時刻のパラメータを使用して、非言語情報を推定するかは任意である。ここで、任意の時間区間Ｔ１～Ｔ２におけるジェスチャ
【０１６１】
【数２４】

$$Y^{T1,T2}$$

【０１６２】
を推定する場合を例に使用する特徴量の一例を示す。推定対象となる時刻Ｔ１～Ｔ２の時刻で得られる、言語特徴量
【０１６３】
【数２５】

$$X_D^{T1,T2} 、 X_P^{T1,T2}$$

【０１６４】
と出力となるジェスチャ
【０１６５】
【数２６】

$$Y^{T1,T2}$$

【０１６６】
をペアとし、これらのペアの複数データである学習データを用いて学習が行われる。その学習済みの非言語情報生成モデルを、
【０１６７】
【数２７】

$$M^{T1,T2}$$

【０１６８】
とする。なお、Ｔ１，Ｔ２の設定方法として、例えば、形態素単位で非言語情報の推定が行われる場合には、各形態素の開始時刻及び終了時刻をＴ１，Ｔ２とする。この場合には、形態素毎にＴ２からＴ１までのウィンドウ幅は異なる。
【０１６９】
学習済みモデル記憶部２３６には、学習部２３５によって学習された学習済みの非言語情報生成モデルが格納される。学習済みの非言語情報生成モデルは、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成する。
【０１７０】
＜非言語情報生成装置の構成＞
【０１７１】
図９は、第２の実施形態に係る非言語情報生成装置２４０の構成の一例を示すブロック図である。図９に示すように、本実施形態に係る非言語情報生成装置２４０は、ＣＰＵ（Central Processing Unit）と、ＲＡＭ（Random Access Memory）と、後述する非言語情報生成処理ルーチンを実行するためのプログラムを記憶したＲＯＭ（Read Only Memory）とを備えたコンピュータで構成されている。非言語情報生成装置２４０は、機能的には、入力部２５０と、演算部２６０と、発現部７０とを備えている。
【０１７２】
入力部２５０は、テキスト情報を受け付ける。

【０１７３】
演算部２６０は、情報取得部２６１と、特徴量抽出部２６２と、学習済みモデル記憶部２６３と、非言語情報生成部２６４とを備えている。
【０１７４】
情報取得部２６１は、入力部２５０によって受け付けられたテキスト情報を取得する。また、情報取得部２６１は、テキスト情報に対応する音声情報を取得し、当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。情報取得部２６１は、図１０に示されるように、テキスト解析部２６５と、音声合成部２６６とを備えている。
【０１７５】
テキスト解析部２６５は、学習用テキスト解析部２３７と同様に、入力部２５０によって受け付けられたテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。
【０１７６】
音声合成部２６６は、学習用音声合成部２３８と同様に、テキスト解析部２６５によって取得されたテキスト解析結果に基づいて、テキスト情報に対応する音声情報を合成する。そして、音声合成部２６６は、音声合成により生成された発話の音声の開始時刻から終了時刻に対応する時刻情報を取得する。
【０１７７】
特徴量抽出部２６２は、学習用特徴量抽出部２３２と同様に、情報取得部２６１によって取得されたテキスト情報及び時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出する。
【０１７８】
学習済みモデル記憶部２６３には、学習済みモデル記憶部２３６に格納されている学習済みの非言語情報生成モデルと同一の学習済みの非言語情報生成モデルが格納されている。
【０１７９】
非言語情報生成部２６４は、特徴量抽出部２６２によって抽出された時刻情報付きのテキスト特徴量と、学習済みモデル記憶部２６３に格納された学習済みの非言語情報生成モデルとに基づいて、特徴量抽出部２６２によって抽出された時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成する。
【０１８０】
例えば、非言語情報生成部２６４は、学習済みモデル記憶部２６３に格納された学習済みの非言語情報生成モデル
【０１８１】
【数２８】

$$M^{T1,T2}$$

【０１８２】
を用いて、時刻情報付きのテキスト特徴量としての任意の特徴量
【０１８３】
【数２９】

$$X_D^{T1,T2}、X_P^{T1,T2}$$

【０１８４】
を入力として、時刻情報付きの非言語情報に対応する生成パラメータとしてのジェスチャ
【０１８５】

【数３０】

$$Y^{T1,T2}$$

【０１８６】
を取得する。
【０１８７】
そして、非言語情報生成部２６４は、生成された時刻情報付きの生成パラメータが、発現部７０から出力されるように発現部７０を制御する。
【０１８８】
具体的には、非言語情報生成部２６４は、発現部７０において、ジェスチャ
【０１８９】
【数３１】

$$Y^{T1,T2}$$

【０１９０】
を任意の対象（例えば、アニメーションキャラクタ、ロボット等）の動作として反映させる。
【０１９１】
発現部７０は、非言語情報生成部２６４による制御に応じて、入力部２５０によって受け付けられたテキスト情報に対応する音声情報と非言語情報生成部２６４によって生成された非言語情報とを発現させる。
【０１９２】
＜非言語情報生成モデル学習装置２１０の作用＞
【０１９３】
次に、第２の実施形態に係る非言語情報生成モデル学習装置２１０の作用について説明する。まず、非言語情報生成モデル学習装置２１０の学習用入力部２２０に、複数の学習用のテキスト情報と複数の学習用の非言語情報との組み合わせを表す学習データが入力されると、非言語情報生成モデル学習装置２１０は、図１１に示す学習処理ルーチンを実行する。
【０１９４】
まず、ステップＳ３００において、学習用情報取得部２３１は、学習用入力部２２０によって受け付けられた複数の学習データ（具体的には、テキスト情報と非言語情報との対）のうちの、学習用のテキスト情報を取得する。
【０１９５】
ステップＳ３０３において、学習用テキスト解析部２３７は、上記ステップＳ３００で取得された学習用のテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、学習用音声合成部２３８は、学習用テキスト解析部２３７によって取得されたテキスト解析結果に基づいて、学習用のテキスト情報に対応する学習用の音声情報を合成する。そして、学習用音声合成部２３８は、学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。
【０１９６】
ステップＳ３０４において、学習用特徴量抽出部２３２は、上記ステップＳ３０３で取得された学習用のテキスト情報及び時刻情報から、時刻情報付きの学習用のテキスト特徴量を抽出する。
【０１９７】
ステップＳ３０８において、学習部２３５は、上記ステップＳ３０４で抽出された時刻

情報付きの学習用のテキスト特徴量と、ステップＳ１０６で取得された時刻情報付きの学習用の生成パラメータとに基づいて、時刻情報付きのテキスト特徴量から時刻情報付きの生成パラメータを生成するための非言語情報生成モデルを学習する。

【０１９８】

＜非言語情報生成装置２４０の作用＞

【０１９９】

次に、第２の実施形態に係る非言語情報生成装置２４０の作用について説明する。まず、非言語情報生成モデル学習装置２１０の学習済みモデル記憶部２３６に格納された学習済みの非言語情報生成モデルが、非言語情報生成装置２４０へ入力されると、非言語情報生成装置２４０の学習済みモデル記憶部２６３へ学習済みの非言語情報生成モデルが格納される。そして、入力部２５０に、非言語情報生成対象のテキスト情報が入力されると、非言語情報生成装置２４０は、図１２に示す非言語情報生成処理ルーチンを実行する。

【０２００】

ステップＳ４００において、情報取得部２６１は、入力部２５０によって受け付けられた、テキスト情報を取得する。

【０２０１】

ステップＳ４０１において、テキスト解析部２６５は、上記ステップＳ４００で取得されたテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、音声合成部２６６は、テキスト解析部２６５によって取得されたテキスト解析結果に基づいて、テキスト情報に対応する音声情報を合成する。そして、音声合成部２６６は、音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０２０２】

ステップＳ４０２において、特徴量抽出部２６２は、上記ステップＳ４０１で取得されたテキスト情報及び時刻情報から、時刻情報付きのテキスト特徴量を抽出する。

【０２０３】

ステップＳ４０４において、非言語情報生成部２６４は、学習済みモデル記憶部２６３に格納された学習済みの非言語情報生成モデルを読み出す。

【０２０４】

ステップＳ４０６において、非言語情報生成部２６４は、上記ステップＳ４０２で抽出された時刻情報付きのテキスト特徴量と、上記ステップＳ４０４で読み出された学習済みの非言語情報生成モデルとに基づいて、上記ステップＳ４０２で抽出された時刻情報付きのテキスト特徴量に対応する時刻情報付きの生成パラメータを生成する。

【０２０５】

なお、第２の実施形態に係る非言語情報生成装置及び非言語情報生成モデル学習装置の他の構成及び作用については、第１の実施形態と同様であるため、説明を省略する。

【０２０６】

以上説明したように、第２の実施形態に係る非言語情報生成装置２４０によれば、テキスト情報に対応する音声情報を取得し、音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得し、時刻情報付きのテキスト特徴量と時刻情報付きの非言語情報を生成するための学習済みモデルとに基づいて、時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成する。これにより、テキスト情報と非言語情報との対応付けを自動的に行い、そのコストを低減させることができる。

【０２０７】

また、本実施形態では、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための学習済みの非言語情報生成モデルを生成することにより、テキスト情報を入力とし、入力に対応する動作に対応する時刻情報付きの非言語情報（出力のタイミングが付与された非言語情報）が出力される。

【０２０８】

これにより、本実施形態によれば、テキスト情報から非言語情報を自動的に生成することが可能となるため、従来技術のように、発話に対して非言語情報を個別に登録する必要

がなく、コストが大幅に削減される。また、本実施形態を用いることで、入力されたテキスト情報に対し、人間らしい自然なタイミングでの非言語行動を生成可能であり、エージェント、及びロボット等の人間らしさや自然さの向上や、非言語行動による意図の伝達の円滑化、会話の活性化などが効果として得られる。

【０２０９】

また、第２の実施形態に係る非言語情報生成モデル学習装置２１０によれば、時刻情報付きの学習用のテキスト特徴量と時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。これにより、テキスト情報と非言語情報との対応付けのコストを低減させつつ、テキスト特徴量から非言語情報を生成するための非言語情報生成モデルを得ることができる。

【０２１０】

また、上記第２の実施形態ではテキスト特徴量から非言語情報を生成する場合を例に説明した。上記第２の実施形態では、単語表記や品詞、対話行為等の情報を手掛かりに、非言語情報を生成することが可能である。この構成を用いれば、チャットにおける対話のように、入力に音声を伴わない場合に、必要最小限の構成で非言語情報を生成することが可能である。

【０２１１】

［第３の実施形態］

【０２１２】

次に、本発明の第３の実施形態について説明する。なお、第１又は第２の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。

【０２１３】

第３の実施形態では、入力として音声情報及びテキスト情報の両方を用いる。音声情報及びテキスト情報から非言語情報を生成する点が第１及び第２の実施形態と異なる。なお、第３の実施形態で用いるテキスト情報は、話者が音声によって外部へ発話する際の、発話内容を表すテキスト情報である。

【０２１４】

第３の実施形態の非言語情報生成モデル学習装置における学習用情報取得部３３１と学習用特徴量抽出部３３２との構成例を図１３Ａに示す。

【０２１５】

図１３Ａに示されるように、学習用の音声情報が入力されると、学習用音声認識部３３７は、学習用の音声情報に対して所定の音声認識処理を行い、学習用の音声情報に対応するテキスト情報（以下、学習用の認識テキストと称する。）を取得する。

【０２１６】

そして、学習用テキスト解析部３３８は、学習用の認識テキストに対して所定のテキスト解析を行い、テキスト解析結果を取得する。

【０２１７】

そして、学習用特徴量抽出部３３２は、学習用の音声情報から時刻情報付きの音声特徴量を抽出する。また、学習用特徴量抽出部３３２は、学習用の認識テキストから時刻情報付きのテキスト特徴量を抽出する。

【０２１８】

そして、第３の実施形態の学習部（図示省略）は、学習用の時刻情報付きの音声特徴量及び学習用の時刻情報付きのテキスト特徴量と、学習用の非言語情報とに基づいて、非言語情報生成モデルを学習させる。これにより、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを得ることができる。

【０２１９】

第３の実施形態の非言語情報生成装置における情報取得部３６１と特徴量抽出部３６２との構成例を図１３Ｂに示す。

【０２２０】
図１３Ｂに示されるように、音声情報が入力されると、音声認識部３６５は、音声情報に対して所定の音声認識処理を行い、音声情報に対応するテキスト情報（以下、認識テキストと称する。）を取得する。そして、テキスト解析部３６６は、認識テキストに対して所定のテキスト解析を行い、テキスト解析結果を取得する。
【０２２１】
そして、特徴量抽出部３６２は、音声情報から時刻情報付きの音声特徴量を抽出する。また、特徴量抽出部３６２は、認識テキストから時刻情報付きのテキスト特徴量を抽出する。
【０２２２】
そして、第３の実施形態の非言語情報生成部（図示省略）は、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量と学習済みの非言語情報生成モデルとに基づいて、時刻情報付きの非言語情報を生成する。これにより、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量の両方を用いて、非言語情報を適切に生成することができる。
【０２２３】
なお、音声特徴量とテキスト特徴量との両方を用いて非言語情報を生成する場合には、時刻情報の時間軸は、上記図３に示されるように、音声特徴量とテキスト特徴量との間で対応が取られ一致していることが望ましい。
【０２２４】
また、音声特徴量とテキスト特徴量との両方を用いて非言語情報を生成し、非言語情報を発現部により発現させる場合には、入力となる音声情報やテキスト情報、テキスト情報から得られた合成音声や音声情報から得られた認識テキストと合わせて提示することも可能である。
【０２２５】
以上説明したように、第３の実施形態に係る非言語情報生成装置によれば、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量と、時刻情報付きの非言語情報を生成するための学習済みの非言語情報生成モデルとに基づいて、時刻情報付きの非言語情報を生成する。これにより、音声情報及びテキスト情報と非言語情報との対応付けのコストを低減させることができる。
【０２２６】
また、第３の実施形態に係る非言語情報生成モデル学習装置によれば、学習用の時刻情報付きの音声特徴量及び学習用の時刻情報付きのテキスト特徴量に基づいて、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを得ることができる。
【０２２７】
なお、上記第３の実施形態では、入力情報が音声情報である場合を例に説明したが、これに限定されるものではない。例えば、入力情報がテキスト情報であってもよい。
【０２２８】
入力情報がテキスト情報である場合の、非言語情報生成モデル学習装置における学習用情報取得部４３１と学習用特徴量抽出部４３２との構成例を図１４Ａに示す。
【０２２９】
図１４Ａに示されるように、学習用のテキスト情報が入力されると、学習用テキスト解析部４３７は、学習用のテキスト情報に対して所定のテキスト解析処理を行い、学習用のテキスト情報に対応するテキスト解析結果を取得する。
【０２３０】
そして、学習用音声合成部４３８は、テキスト解析結果に対して所定の音声合成処理を行い、学習用の音声情報を取得する。
【０２３１】
そして、学習用特徴量抽出部４３２は、学習用の音声情報から時刻情報付きの音声特徴量を抽出する。また、学習用特徴量抽出部４３２は、学習用のテキスト情報から時刻情報

付きのテキスト特徴量を抽出する。
【０２３２】
そして、学習部（図示省略）は、学習用の時刻情報付きの音声特徴量及び学習用の時刻情報付きのテキスト特徴量と、学習用の非言語情報とに基づいて、非言語情報生成モデルを学習させる。これにより、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを得ることができる。
【０２３３】
また、入力情報がテキスト情報である場合の、非言語情報生成装置における情報取得部４６１と特徴量抽出部４６２との構成例を図１４Ｂに示す。
【０２３４】
図１４Ｂに示されるように、テキスト情報が入力されると、テキスト解析部４６５は、テキスト情報に対して所定のテキスト解析処理を行い、テキスト情報に対応するテキスト解析結果を取得する。
【０２３５】
そして、音声合成部４６６は、テキスト解析結果に対して所定の音声合成処理を行い、音声情報を取得する。
【０２３６】
そして、特徴量抽出部４６２は、音声情報から時刻情報付きの音声特徴量を抽出する。また、特徴量抽出部４６２は、テキスト情報から時刻情報付きのテキスト特徴量を抽出する。
【０２３７】
そして、非言語情報生成部（図示省略）は、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量と、予め学習された非言語情報生成モデルとに基づいて、非言語情報を表す生成パラメータを取得する。
【０２３８】
なお、本発明は、上述した実施形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。
【０２３９】
例えば、上記各実施形態における、情報取得部（又は学習用情報取得部）と特徴量抽出部（又は学習用特徴量抽出部）との組み合わせに対応する構成は、図１５に示される全４パターンである。また、学習時と非言語情報生成時との構成の組み合わせとして、可能なバリエーションは、図１６に示されるパターンとなる。なお、学習時と生成時の特徴量が一致している方が、非言語情報生成時の精度は高くなる。
【０２４０】
また、本発明は、周知のコンピュータに媒体もしくは通信回線を介して、プログラムをインストールすることによっても実現可能である。
【０２４１】
また、上述の装置は、内部にコンピュータシステムを有しているが、「コンピュータシステム」は、WWW（World Wide Web）システムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。
【０２４２】
また、本願明細書中において、プログラムが予めインストールされている実施形態について説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。
【０２４３】
なお、以下に、他の実施形態を説明する。
【０２４４】
[他の実施形態の概要]
【０２４５】

非言語情報生成モデル学習装置において用いる学習データとして、例えば、上記図２に示されるようなシーンにおいて、発話中の話者の音声情報（Ｘ）を取得するのと同時に、計測装置を用いて発話中の話者の話し相手である会話相手の非言語情報（Ｙ）を取得することにより作成する。

【０２４６】

そのようにして作成された学習データに基づいて、非言語情報生成モデルの学習を行うと、入力情報である音声情報又はテキスト情報に対し、適切なタイミングでリアクション（例えば、相槌など）の非言語行動を行うエージェント及びロボット等が実現可能となる。

【０２４７】

さらに、上記図２のようなシーンにおいて、発話中の話者の音声情報（Ｘ）を取得するのと同時に、計測装置を用いて他の参加者の非言語情報（Ｙ）を取得することにより、学習データを作成することもできる。

【０２４８】

そのようにして複数の参加者のそれぞれの非言語情報を元に、非言語情報生成モデルの学習を行えば、複数ロボットやエージェントに、入力情報である音声情報又はテキスト情報に対し、適切かつ異なるタイミングでリアクションを行わせることが可能となる。

【０２４９】

この場合には、話者から取得された音声情報と当該話者の相手（例えば、会話の聞き手又は会話の参加者）の行動に関する情報を表す非言語情報との組み合わせを学習データとして、学習済みの非言語情報生成モデルを学習させる。ここで、当該話者の相手として、会話の聞き手、会話の参加者だけでなく、会話の傍聴者等も含めてもよく、これら「話者の音声（及びその内容）を受けて何らかの反応を示す実体」について、「発声の聞き手」とも表現する。

【０２５０】

また、テキスト情報を対象とする場合には、話者から取得された音声情報に対応するテキスト情報と当該話者の相手（例えば、会話の聞き手又は会話の参加者）の行動に関する情報を表す非言語情報との組み合わせを学習データとして、学習済みの非言語情報生成モデルを学習させる。

【０２５１】

なお、以下の第４の実施形態～第６の実施形態は、第１の変形例～第３の変形例である。

【０２５２】

[第４の実施形態]

【０２５３】

第４の実施形態では、音声情報を対象とし、非言語情報生成モデル学習装置は、話者（図２の「話者」）から出力された学習用の音声情報から抽出される時刻情報付きの学習用の音声特徴量と、話者の相手（図２の「他の参加者」または「話者の会話相手」）の行動に関する情報を表す学習用の非言語情報とに基づいて、時刻情報付きの音声特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習させる。これにより、話者から取得される音声情報に応じて、話者の相手の行動を表現する学習済みの非言語情報生成モデルが得られる。なお、図２に示されるように、「話者」とは、発話等の音声を発する人物を表す。また、「話者の相手」とは、例えば、話者から発せられた発話等を聞いている人物を表し、図２に示される「他の参加者」及び「話者の会話相手」に相当する。また、話者の相手の行動とは、話者が発した音声に対する、話者の相手のリアクション等である。

【０２５４】

また、非言語情報生成対象の音声情報が入力された場合には、非言語情報生成装置は、話者から取得された音声情報から抽出される時刻情報付きの音声特徴量と、学習済みの非言語情報生成モデルとに基づいて、時刻情報付きの非言語情報を生成する。

【０２５５】

なお、第４の実施形態の非言語情報生成モデル学習装置及び非言語情報生成装置の他の

構成及び作用については、第１の実施形態と同様であるため、説明を省略する。
【０２５６】
[第５の実施形態]
【０２５７】
第５の実施形態では、テキスト情報を対象とし、非言語情報生成モデル学習装置は、話者から取得された学習用の音声情報に対応するテキストから抽出される時刻情報付きの学習用のテキスト特徴量と、話者の相手の行動に関する情報を表す学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習させる。これにより、話者から取得される音声情報に対応するテキスト情報に応じて、話者の相手の行動を表現する学習済みの非言語情報生成モデルが得られる。
【０２５８】
また、非言語情報生成対象のテキスト情報が入力された場合には、非言語情報生成装置は、話者から取得された音声情報に対応するテキスト情報から抽出される時刻情報付きのテキスト特徴量と、学習済みの非言語情報生成モデルとに基づいて、時刻情報付きの非言語情報を生成する。
【０２５９】
なお、第５の実施形態の非言語情報生成モデル学習装置及び非言語情報生成装置の他の構成及び作用については、第２の実施形態と同様であるため、説明を省略する。
【０２６０】
[第６の実施形態]
【０２６１】
第６の実施形態では、音声情報とテキスト情報との両方を対象とし、非言語情報生成モデル学習装置は、時刻情報付きの学習用の音声特徴量及び時刻情報付きの学習用のテキスト特徴量と、話者の相手の行動に関する情報を表す学習用の非言語情報とに基づいて、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習させる。これにより、音声情報及びテキスト情報に応じて、話者の相手の行動を表現する学習済みの非言語情報生成モデルが得られる。
【０２６２】
また、非言語情報生成対象の音声情報とテキスト情報との両方が入力された場合には、非言語情報生成装置は、時刻情報付きの音声特徴量及び時刻情報付きのテキスト特徴量と、学習済みの非言語情報生成モデルとに基づいて、時刻情報付きの非言語情報を生成する。
【０２６３】
なお、第６の実施形態の非言語情報生成モデル学習装置及び非言語情報生成装置の他の構成及び作用については、第３の実施形態と同様であるため、説明を省略する。
【０２６４】
[第７の実施形態]
【０２６５】
＜非言語情報生成モデル学習装置の構成＞
【０２６６】
次に、本発明の第７の実施形態について説明する。なお、第１の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。
【０２６７】
第７の実施形態では、入力された発話に対応するジェスチャを、学習済みの機械学習モデルに基づいて生成する。その際、ジェスチャとして、行動の有無だけでなく、大きさ、回数、ポーズ長の割合の情報も生成する。また、入力として用いる発話音声またはテキストに加えて、ジェスチャの生成に影響を与える変数を「付加情報」として更に用いて、非言語情報を生成する非言語情報生成モデルの学習を行う点が第１の実施形態と異なる。
【０２６８】

なお、第７の実施形態で用いるテキスト情報は、話者が音声によって外部へ発話する際の、発話内容を表すテキスト情報である。

【０２６９】

図１７は、第７の実施形態に係る非言語情報生成モデル学習装置７１０の構成の一例を示すブロック図である。図１７に示すように、第７の実施形態に係る非言語情報生成モデル学習装置７１０は、ＣＰＵと、ＲＡＭと、後述する学習処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成されている。非言語情報生成モデル学習装置７１０は、機能的には、学習用入力部７２０と、学習用演算部７３０とを備えている。

【０２７０】

学習用入力部７２０は、学習用のテキスト情報と学習用の非言語情報と学習用の付加情報との組み合わせを含む学習データを受け付ける。

【０２７１】

受け付ける非言語情報は、行動の有無だけでなく、大きさ、回数、ポーズ長の割合の情報も含む。

【０２７２】

具体的には、非言語情報には、以下の表２～表４に示す行動リスト表に含まれる何れかの行動が含まれる。

【０２７３】

【表２】

No.	大項目	中項目	内容	終了位置
1	Head_pitch	initial	平常状態(正面)	Initial
2		nod_1l	1回大きく頷く	Initial
3		nod_1m	1回中くらいに頷く	Initial
4		nod_1s	1回小さく頷く	Initial
5		nod_2l	2回大きく頷く	Initial
6		nod_2m	2回中くらいに頷く	Initial
7		nod_2s	2回小さく頷く	Initial
8		nod_3l	3回大きく頷く	Initial
9		nod_3m	3回中くらいに頷く	Initial
10		nod_3s	3回小さく頷く	Initial
11		nod_4l	3回大きく頷く	Initial
12		nod_4m	3回中くらいに頷く	Initial
13		nod_4s	3回小さく頷く	Initial
14		nod_5l	5回以上大きく頷く	Initial
15		nod_5m	5回以上中くらいに頷く	Initial
16		nod_5s	5回以上小さく頷く	Initial
17		upper_l	大きく上に頭を回転	Upper_L
18		down_l	大きく下に頭を回転	Down_L
19		upper_m	中くらいで上に頭を回転	Upper_M
20		down_m	中くらいで下に頭を回転	Down_M
21		upper_s	小さく上に頭を回転	Upper_S
22		down_s	小さく下に頭を回転	Down_S
24	Head_yaw	initial	平常状態(正面)	Initial
25		right_l	大きく右を見る	Right_L
26		left_l	大きく左を見る	Left_L
27		right_m	中くらいで右を見る	Right_M
28		left_m	中くらいで左を見る	Left_M
29		right_s	小さく右を見る	Right_S
30		left_s	小さく左を見る	Left_S
31		shake_l	大きく頭を左右に振る	Initial
32		shake_m	中くらいに頭を左右に振る	Initial
33		shake_s	小さく頭を左右に振る	Initial

【０２７４】

【表３】

35	Head_roll	initial	平常状態(正面)	Initial
36		tilt_right_l	大きく右に傾げる	Tilt_right_L
37		tilt_left_l	大きく左に傾げる	Tilt_left_L
38		tilt_right_m	中くらいに右に傾げる	Tilt_right_M
39		tilt_left_m	中くらいに大きく左に傾げる	Tilt_left_M
40		tilt_right_s	小さく右に傾げる	Tilt_right_S
41		tilt_left_s	小さく左に傾げる	Tilt_left_S
43	Hand_gesture	initial_down	平常状態(手を下に)	Initial_down
44		initial_chest	平常状態(手を曲げて広げて腕の高さに)	Initial_chest
46a		iconic_1	情景描写、動作を表現	Initial_chest
46b		iconic_2	情景描写、動作を表現	Initial_chest
46c		iconic_6	情景描写、動作を表現	Initial_chest
47a		metaphoric_1	絵画的、図形的なジェスチャ	Initial_chest
47b		metaphoric_2	絵画的、図形的なジェスチャ	Initial_chest
47c		metaphoric_7	絵画的、図形的なジェスチャ	Initial_chest
48a		beat_1	発話の調子を整えたり、発言を強調	Initial_chest
48b		beat_2	発話の調子を整えたり、発言を強調	Initial_chest
48c		beat_8	発話の調子を整えたり、発言を強調	Initial_chest
49a		deictic_1	指差し	Initial_chest
49b		deictic_2	指差し	Initial_chest
49c		deictic_9	指差し	Initial_chest
50a		feedback_1	他人の発話に同調・同意・呼応	Initial_chest
50b		feedback_2	他人の発話に同調・同意・呼応	Initial_chest
50c		feedback_10	他人の発話に同調・同意・呼応	Initial_chest
51a		compellation_1	相手に呼びかけを行う	Initial_chest
51b		compellation_2	相手に呼びかけを行う	Initial_chest
51c		compellation_11	相手に呼びかけを行う	Initial_chest
52a		hesitate_1	言いよどみ時	Initial_chest
52b		hesitate_2	言いよどみ時	Initial_chest
52c		hesitate_12	言いよどみ時	Initial_chest
53a		others_1_1	頬を触る	Others_1_1
53b		others_1_2	頬を触る	Others_1_2
53c		others_1_13	頬を触る	Others_1_13
54		others_2	教を教える	Others_2
55		others_3	腕を組む	Others_3

【０２７５】

【表４】

56	Facial_expression	initial	平常	Initial
57		smile_l	強い笑顔	Smile_L
58		smile_m	中くらいな笑顔	Smile_M
59		smile_s	少しの笑顔	Smile_S
60		anger_l	強い怒り	Anger_L
61		anger_m	中くらいな怒り	Anger_M
62		anger_s	少しの怒り	Anger_S
63		sad_l	強い悲しみ	Sad_L
64		sad_m	中くらいな悲しみ	Sad_M
65		sad_s	少しの悲しみ	Sad_S
66		surprise_l	強い驚き	Surprise_L
67		surprise_m	中くらいな驚き	Surprise_M
68		surprise_s	少しの驚き	Surprise_S
69		dislike_l	強い嫌悪	Dislike_L
70		dislike_m	中くらいな嫌悪	Dislike_M
71		dislike_s	少しの嫌悪	Dislike_S
72		fear_l	強い恐れ	Fear_L
73		fear_m	中くらいな恐れ	Fear_M
74		fear_s	少しの恐れ	Fear_S
75	Upper_body_posture	initial	直立(適当に揺れてる)	Initial
76		forward_l	大きく前傾(適当に揺れてる)	Forward_L
77		forward_m	中くらい前傾(適当に揺れてる)	Forward_M
78		forward_s	小さく前傾(適当に揺れてる)	Forward_S
79		backward_l	大きく前傾(適当に揺れてる)	Backward_L
80		backward_m	中くらい前傾(適当に揺れてる)	Backward_M
81		backward_s	小さく前傾(適当に揺れてる)	Backward_S
82		bow	お辞儀する	Bow

【０２７６】

ここで、学習用のテキスト情報は、発話音声データを基に、それぞれ1人ずつの発話を人手で書き起こしたものであり、発話区間は、200msec以上の無音区間で区切った。

【０２７７】

また、学習用の非言語情報は、発話音声データに対応する発話時の動きを考慮して付与されたものである。非言語情報は、例えば、頷き、顔向き、ハンドジェスチャ、視線、表情、身体の姿勢、身体の関節、位置、動き、瞳孔径の大きさ、及び瞬きの有無等である。

学習用の非言語情報に含まれるハンドジェスチャは、例えば、発話時の手の動きに加えて、発話内容を考慮しながらアノテーションを行ったものである。また、学習用の非言語

情報に含まれるハンドジェスチャは、自動認識により得られたものでもよい。ハンドジェスチャの自動認識については、画像処理などを利用した方法（例えば、参考文献７）が多く提案されており、どのような方法でもよい。

【０２７８】

［参考文献７］Siddharth S. Rautaray, Anupam Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey", Artificial Intelligence Review, January 2015, Volume 43, Issue 1, pp. 1-54.

【０２７９】

ここで、ハンドジェスチャの種別として参考文献８に記載されたものと同様なものを用いればよい。例えば、（A) Iconic、（B) Metaphoric、（C) Beat、（D) Deictic、（E) Feedback、（F) Compellation、（G) Hesitate、（H) Othersが、ハンドジェスチャの詳細な種別である。

【０２８０】

［参考文献８］D. McNeill, "Hand and Mind: What Gestures Reveal About Thought", Chicago: University of Chicago Press, 1992

【０２８１】

（A) Iconicは、情景描写、動作を表現するために使われるジェスチャである。（B) Metaphoricは、Iconicと同様、絵画的、図形的なジェスチャであるが、指示される内容は抽象的な事柄、概念となる（時間の流れなど）。（C) Beatは、発話の調子を整えたり、発言を強調したりするジェスチャであり、手を振動させたり、発話に合わせて手を振ったりするジェスチャである。（D) Deicticは、指差しなど、直接的に方向や、場所、事物を指し示すジェスチャである。（E) Feedbackは、他人の発話に同調・同意・呼応して出るジェスチャや、他人の前の発話・ジェスチャに対して、呼応して発言したときに付随するジェスチャ、相手のジェスチャを模倣して同じ形のジェスチャを行ったものである。（F) Compellationは、相手に呼びかけを行うジェスチャである。（G) Hesitateは、言いよどみ時に出るジェスチャである。（H) Othersは、判断に迷うが、何か意味がありそうなジェスチャである。

【０２８２】

なお、ハンドジェスチャのアノテーションでは、上記のハンドジェスチャの種別だけでなく、Prep、Hold、Stroke、Returnの４つの状態を示すアノテーションも付与するようにしてもよい。

【０２８３】

Prepは、ホームポジションからジェスチャをするために手を上げる状態を示し、Holdは、空中に手をあげたままの状態(ジェスチャ開始までの待機時間)を示す。また、Strokeは、ジェスチャを実施する状態を示し、この状態の詳細情報として、上記（A）～（H）までの種別のアノテーションを与える。Returnは、手をホームポジションに戻す状態を示す。

【０２８４】

また、学習用の非言語情報に含まれる頷きジェスチャは、例えば、発話時に頭を下げて戻すまでの頷き区間に対してアノテーションを行ったものである。また、頭を前に出して戻す動作、あるいは後ろに引いて戻す動作も頷きとし、アノテーションを行った。首をかしげる動作や、左右に振る動作は、頷きとはしなかった。

【０２８５】

休止を挟まずに連続して2回以上頷いた場合は、連続した区間をひとまとまりに付け、頷いた回数を付与した。頷きの回数は、「1回、2回、3回、4回、5回以上」で分類した。

【０２８６】

受け付ける学習用の付加情報は、所定の処理単位毎（例えば、形態素毎）の付加情報である。付加情報として、個人属性、環境変数、身体的特徴、動作対象の姿勢、対話の内容、人間関係、及び感情の少なくとも１つを受け付ける。具体的には、個人属性は、性別、年齢、性格、国籍、文化圏を含み、環境変数は、対話時の人数（1対1、1対多、多対多）

、気温、屋内/屋外、陸上/空中/水中、明るい/暗い、等を含む。また、身体的特徴は、3頭身、服装：ポケットがある、スカートをはいている、帽子を被っている等、動作に影響を与えるものを含み、動作対象の姿勢は、立っている、座っている、何か手に持っている、を含む。また、対話の内容は、議論、雑談、説明などを含み、人間関係は、ジェスチャを生成させる人物と対話相手との人間関係、どちらが立場が優位か、好意をもっているかなどを含む。感情は、喜び、怒り、悲しみなどの他、緊張/リラックス等の精神状態も含む、内部状態を表すものである。

なお、付加情報が、所定の処理単位毎に変化しないもの（例えば、性別）である場合には、所定の処理単位毎に受け付けなくてもよい。この場合には、付加情報を受け付けた際に、装置側で、所定の処理単位毎の付加情報に展開すればよい。

【０２８７】

学習用演算部７３０は、学習用入力部７２０によって受け付けた学習データに基づいて、時刻情報付きの非言語情報を生成するための非言語情報生成モデルを生成する。学習用演算部７３０は、図１７に示されるように、学習用情報取得部２３１と、学習用付加情報取得部７３１と、学習用特徴量抽出部７３２と、非言語情報取得部３３と、生成パラメータ抽出部３４と、学習部７３５と、学習済みモデル記憶部７３６とを備えている。

【０２８８】

学習用情報取得部２３１は、学習用のテキスト情報に対応する学習用の音声情報を取得し、当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０２８９】

学習用付加情報取得部７３１は、学習用の付加情報を取得する。

【０２９０】

学習用特徴量抽出部７３２は、学習用情報取得部２３１によって取得された学習用のテキスト情報及び時刻情報から、当該学習用のテキスト情報の特徴量を表す、時刻情報付きの学習用のテキスト特徴量を抽出する。具体的には、学習用特徴量抽出部７３２は、所定の解析単位毎に当該テキスト情報に対して時刻情報を付与して、時刻情報付きのテキスト特徴量を抽出する。時刻情報付きのテキスト特徴量の一例を図１８に示す。図１８に示されるように、テキスト情報の各形態素の開始時刻及び終了時刻が取得されている。

【０２９１】

また、学習用特徴量抽出部７３２は、学習用付加情報取得部７３１によって取得された学習用の付加情報と、学習用情報取得部２３１によって取得された時刻情報とから、時刻情報付きの学習用の付加情報を生成する。具体的には、学習用特徴量抽出部７３２は、所定の解析単位毎に当該付加情報に対して時刻情報を付与して、時刻情報付きの付加情報を生成する。

【０２９２】

時刻情報付きの付加情報の一例を図１８に示す。図１８に示されるように、各形態素の付加情報に、当該形態素の開始時刻及び終了時刻が取得されている。また、付加情報が複数種類ある場合（例えば、対話人数と感情と気温）には、付加情報が、複数種類の付加情報を格納したベクトル配列として表されている。

【０２９３】

学習用特徴量抽出部７３２は、付加情報のベクトル配列を

【０２９４】

【数３２】

$$X_{ADD}^{t_{P,s},t_{P,e}}$$

【０２９５】
とする。なお、
【０２９６】
【数３３】

$$t_{P,s}, t_{P,e}$$

【０２９７】
は、形態素単位に該当する発話音声の開始時刻及び終了時刻とする。
【０２９８】
学習部７３５は、学習用特徴量抽出部７３２によって抽出された時刻情報付きの学習用のテキスト特徴量及び時刻情報付きの付加情報と、生成パラメータ抽出部３４によって抽出された時刻情報付きの学習用の離散化された非言語情報とに基づいて、時刻情報付きのテキスト特徴量及び付加情報から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。
【０２９９】
具体的には、学習部７３５は、学習用特徴量抽出部７３２によって抽出された時刻情報付きの学習用のテキスト特徴量
【０３００】
【数３４】

$$X_D^{t_{S,s},t_{S,e}}$$

【０３０１】
【数３５】

$$X_P^{t_{P,s},t_{P,e}}$$

【０３０２】
、及び時刻情報付きの学習用の付加情報
【０３０３】
【数３６】

$$X_{ADD}^{t_{P,s},t_{P,e}}$$

【０３０４】
を入力として、非言語情報
【０３０５】
【数３７】

$$Y^{t_{N,s},t_{N,e}}$$

【０３０６】
を出力する非言語情報生成モデルを構築する。非言語情報生成モデルを構築する際には、あらゆる機械学習手法を用いて良いが、本実施形態ではＳＶＭを用いる。例えば、上記（Ａ）～（Ｈ）までの種別毎に、ジェスチャが、当該種別に属する何れの行動であるかを推定するためのＳＶＭモデルを作成する。すわなち、上記（Ａ）～（Ｈ）までの種別毎に、ジェスチャが、上記の行動リスト表の、当該種別に属する内容として記載されている複数の行動のうち、何れの行動であるかを推定するためのＳＶＭモデルを作成する。
【０３０７】
なお、非言語情報生成モデルは、どのような時間分解能で、どの時刻のパラメータを使用して、非言語情報を推定するかは任意である。ここで、任意の時間区間Ｔ１～Ｔ２におけるジェスチャ
【０３０８】
【数３８】

$$Y^{T1,T2}$$

【０３０９】
を推定する場合を例に使用する特徴量の一例を示す。推定対象となる時刻Ｔ１～Ｔ２の時刻で得られる、言語特徴量
【０３１０】
【数３９】

$$X_D^{T1,T2} \text{、} X_P^{T1,T2}$$

【０３１１】
、及び付加情報
【０３１２】
【数４０】

$$X_{ADD}^{T1,T2}$$

【０３１３】
と、出力となるジェスチャ
【０３１４】
【数４１】

$$Y^{T1,T2}$$

【０３１５】

をペアとし、これらのペアの複数データである学習データを用いて学習が行われる。その学習済みの非言語情報生成モデルを、

【０３１６】

【数４２】

$$M^{T1,T2}$$

【０３１７】

とする。なお、Ｔ１，Ｔ２の設定方法として、例えば、形態素単位で非言語情報の推定が行われる場合には、各形態素の開始時刻及び終了時刻をＴ１，Ｔ２とする。この場合には、形態素毎にＴ２からＴ１までのウィンドウ幅は異なる。

【０３１８】

また、

【０３１９】

【数４３】

$$Y^{T1,T2}$$

【０３２０】

は、T1からT2内で得られる非言語情報の平均値や、出現した非言語情報の組み合わせや、その出現順序を考慮したパターンでも良い。例えば、ハンドジェスチャＩＤ

【０３２１】

【数４４】

$$Y_{HG}^{ID}$$

【０３２２】

が、T1からT3（T1〈T3〈T2）内で

【０３２３】

【数４５】

$$Y_{HG_1}^{ID}$$

【０３２４】

、T3からT2内で

【０３２５】

【数４６】

$$Y_{HG_2}^{ID}$$

【０３２６】

であったとき、

【０３２７】
【数４７】

$$Y^{T1,T2}$$

【０３２８】
として、より出現時間の高かったＩＤや、組み合わせ情報である
【０３２９】
【数４８】

$$\{ Y_{HG}^{ID}1,\ Y_{HG}^{ID}2\}$$

【０３３０】
、ｎgramパターンとしての
【０３３１】
【数４９】

$$Y_{HG}^{ID}1 - Y_{HG}^{ID}2$$

【０３３２】
を採用する。
【０３３３】
n-gramパターンを用いる場合、n-gramパターン内の非言語情報の時刻情報（上の例では、
【０３３４】
【数５０】

$$Y_{HG}^{ID}1 - Y_{HG}^{ID}2$$

【０３３５】
の、それぞれの開始時刻など）は、各非言語情報ごとにあらかじめ設定された所定の方法で割り振られることになる。しかし、このn-gramパターン内の非言語情報の時刻情報も推定するようにしてもよい。この場合、n-gramパターンを推定する際に用いた特徴量と、推定されたn-gramを用いて、学習データに基づき、時刻情報の推定を行うようにする。
【０３３６】
学習済みモデル記憶部７３６には、学習部７３５によって学習された学習済みの非言語情報生成モデルが格納される。学習済みの非言語情報生成モデルは、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成する。
【０３３７】
＜非言語情報生成装置の構成＞
【０３３８】
図１９は、第７の実施形態に係る非言語情報生成装置７４０の構成の一例を示すブロック図である。図１９に示すように、本実施形態に係る非言語情報生成装置７４０は、ＣＰ

Ｕ（Central Processing Unit）と、ＲＡＭ（Random Access Memory）と、後述する非言語情報生成処理ルーチンを実行するためのプログラムを記憶したＲＯＭ（Read Only Memory）とを備えたコンピュータで構成されている。非言語情報生成装置７４０は、機能的には、入力部７５０と、演算部７６０と、発現部７０とを備えている。

【０３３９】

入力部７５０は、テキスト情報及び付加情報を受け付ける。受け付ける付加情報は、所定の処理単位毎（例えば、形態素毎）の付加情報である。

【０３４０】

演算部７６０は、情報取得部２６１と、付加情報取得部７６１と、特徴量抽出部７６２と、学習済みモデル記憶部７６３と、非言語情報生成部７６４とを備えている。

【０３４１】

情報取得部２６１は、入力部７５０によって受け付けられたテキスト情報を取得する。また、情報取得部２６１は、テキスト情報に対応する音声情報を取得し、当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

付加情報取得部７６１は、入力部７５０によって受け付けられた付加情報を取得する。

【０３４２】

特徴量抽出部７６２は、学習用特徴量抽出部７３２と同様に、情報取得部２６１によって取得されたテキスト情報及び時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出する。また、特徴量抽出部７６２は、学習用特徴量抽出部７３２と同様に、付加情報取得部７６１によって取得された付加情報と、情報取得部２６１によって取得された時刻情報とから、時刻情報付きの付加情報を生成する。

【０３４３】

学習済みモデル記憶部７６３には、学習済みモデル記憶部７３６に格納されている学習済みの非言語情報生成モデルと同一の学習済みの非言語情報生成モデルが格納されている。

【０３４４】

非言語情報生成部７６４は、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び時刻情報付きの付加情報と、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルとに基づいて、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの非言語情報を生成する。

【０３４５】

例えば、非言語情報生成部７６４は、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデル

【０３４６】

【数５１】

$$M^{T1,T2}$$

【０３４７】

を用いて、時刻情報付きのテキスト特徴量としての任意の特徴量

【０３４８】

【数５２】

$$X_D^{T1,T2} \text{、} X_P^{T1,T2}$$

【０３４９】

、及び時刻情報付きの付加情報

【０３５０】
【数５３】

$$X_{ADD}^{T1,T2}$$

【０３５１】
を入力として、時刻情報付きの非言語情報に対応する生成パラメータとしてのジェスチャ
【０３５２】
【数５４】

$$Y^{T1,T2}$$

【０３５３】
を取得する。
【０３５４】
そして、非言語情報生成部７６４は、生成された時刻情報付きの生成パラメータが、当該生成パラメータに付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御する。
【０３５５】
具体的には、非言語情報生成部７６４は、発現部７０において、ジェスチャ
【０３５６】
【数５５】

$$Y^{T1,T2}$$

【０３５７】
を任意の対象（例えば、アニメーションキャラクタ、ロボット等）の動作として反映させる。
【０３５８】
発現部７０は、非言語情報生成部７６４による制御に応じて、入力部７５０によって受け付けられたテキスト情報に対応する音声情報と非言語情報生成部７６４によって生成された非言語情報とを発現させる。
【０３５９】
＜非言語情報生成モデル学習装置７１０の作用＞
【０３６０】
次に、第７の実施形態に係る非言語情報生成モデル学習装置７１０の作用について説明する。まず、非言語情報生成モデル学習装置７１０の学習用入力部７２０に、複数の学習用のテキスト情報と複数の学習用の付加情報と複数の学習用の非言語情報との組み合わせを表す学習データが入力されると、非言語情報生成モデル学習装置７１０は、図２０に示す学習処理ルーチンを実行する。
【０３６１】
まず、ステップＳ３００において、学習用情報取得部２３１は、学習用入力部７２０によって受け付けられた複数の学習データ（具体的には、テキスト情報と付加情報と非言語情報との対）のうちの、学習用のテキスト情報を取得する。
【０３６２】
ステップＳ１０２において、非言語情報取得部３３は、学習用入力部７２０によって受

け付けられた複数の学習データのうちの、学習用の非言語情報と当該学習用の非言語情報が表す行動が行われるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０３６３】

ステップＳ３０３において、学習用テキスト解析部２３７は、上記ステップＳ３００で取得された学習用のテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、学習用音声合成部２３８は、学習用テキスト解析部２３７によって取得されたテキスト解析結果に基づいて、学習用のテキスト情報に対応する学習用の音声情報を合成する。そして、学習用音声合成部２３８は、学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０３６４】

ステップＳ７００において、学習用付加情報取得部７３１は、学習用入力部７２０によって受け付けられた複数の学習データのうちの、学習用の付加情報を取得する。

【０３６５】

ステップＳ７０２において、学習用特徴量抽出部７３２は、上記ステップＳ３０３で取得された学習用のテキスト情報及び時刻情報から、時刻情報付きの学習用のテキスト特徴量を抽出する。また、学習用特徴量抽出部７３２は、上記ステップＳ７００で取得された学習用の付加情報と、上記ステップＳ３０３で取得された時刻情報とから、時刻情報付きの学習用の付加情報を生成する。

【０３６６】

ステップＳ１０６において、生成パラメータ抽出部３４は、上記ステップＳ１０２で取得された学習用の非言語情報及び時刻情報から、時刻情報付きの離散化された非言語情報を抽出する。

【０３６７】

ステップＳ７０８において、学習部７３５は、上記ステップＳ７０２で抽出された時刻情報付きの学習用のテキスト特徴量及び時刻情報付きの学習用の付加情報と、ステップＳ１０６で取得された時刻情報付きの学習用の生成パラメータとに基づいて、時刻情報付きのテキスト特徴量及び付加情報から時刻情報付きの生成パラメータを生成するための非言語情報生成モデルを学習する。

【０３６８】

ステップＳ１１０において、学習部７３５は、上記ステップＳ７０８で得られた学習済みの非言語情報生成モデルを、学習済みモデル記憶部７３６へ格納して、学習処理ルーチンを終了する。

【０３６９】

＜非言語情報生成装置７４０の作用＞

【０３７０】

次に、第７の実施形態に係る非言語情報生成装置７４０の作用について説明する。まず、非言語情報生成モデル学習装置７１０の学習済みモデル記憶部７３６に格納された学習済みの非言語情報生成モデルが、非言語情報生成装置７４０へ入力されると、非言語情報生成装置７４０の学習済みモデル記憶部７６３へ学習済みの非言語情報生成モデルが格納される。そして、入力部７５０に、非言語情報生成対象のテキスト情報及び付加情報が入力されると、非言語情報生成装置７４０は、図２１に示す非言語情報生成処理ルーチンを実行する。

【０３７１】

ステップＳ４００において、情報取得部２６１は、入力部７５０によって受け付けられた、テキスト情報を取得する。

【０３７２】

ステップＳ４０１において、テキスト解析部２６５は、上記ステップＳ４００で取得されたテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、音声合成部２６６は、テキスト解析部２６５によって取得されたテキスト解析結果に基づいて、テキスト情報に対応する音声情報を合成する。そして、音声合成部２６６は、

音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０３７３】

ステップＳ７５０において、付加情報取得部７６１は、入力部７５０によって受け付けられた、付加情報を取得する。

【０３７４】

ステップＳ７５２において、特徴量抽出部７６２は、上記ステップＳ４０１で取得されたテキスト情報及び時刻情報から、時刻情報付きのテキスト特徴量を抽出し、上記ステップＳ７５０で取得された付加情報及び上記ステップＳ４０１で取得された時刻情報から、時刻情報付きの付加情報を生成する。

【０３７５】

ステップＳ７５４において、非言語情報生成部７６４は、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルを読み出す。

【０３７６】

ステップＳ７５６において、非言語情報生成部７６４は、上記ステップＳ７５２で抽出された時刻情報付きのテキスト特徴量及び付加情報と、上記ステップＳ７５４で読み出された学習済みの非言語情報生成モデルとに基づいて、上記ステップＳ７５２で抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの生成パラメータを生成する。

【０３７７】

ステップＳ２０８において、非言語情報生成部７６４は、上記ステップＳ７５６で生成された時刻情報付きの非言語情報が、当該非言語情報に付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御して、非言語情報生成処理ルーチンを終了する。

【０３７８】

なお、第７の実施形態に係る非言語情報生成装置及び非言語情報生成モデル学習装置の他の構成及び作用については、第１の実施形態と同様であるため、説明を省略する。

【０３７９】

以上説明したように、第７の実施形態に係る非言語情報生成装置７４０によれば、付加情報を取得し、時刻情報付きのテキスト特徴量及び付加情報と、行動の回数又は大きさを含む時刻情報付きの非言語情報を生成するための学習済みモデルとに基づいて、時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成する。これにより、テキスト情報と、行動の回数又は大きさを含む非言語情報との対応付けを自動的に行い、そのコストを低減させることができる。

【０３８０】

また、非言語情報の種類を、行動の大きさ、回数を含めて細かく設定することにより、より詳細なニュアンスを表現でき、非言語情報により意図を表しやすい。また、時刻情報が対応付いていることや、付加情報と組み合わせることで、付加情報の違いによる動作の変化を、より細やかに表現でき、これにより、例えば、感情を表現しやすくなる。

【０３８１】

また、第７の実施形態に係る非言語情報生成モデル学習装置７１０によれば、時刻情報付きの学習用のテキスト特徴量及び付加情報と、行動の回数又は大きさを含む時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量及び付加情報から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。これにより、テキスト情報と、行動の回数又は大きさを含む非言語情報との対応付けのコストを低減させつつ、テキスト特徴量から非言語情報を生成するための非言語情報生成モデルを得ることができる。

【０３８２】

なお、上記の実施の形態では、付加情報が入力される場合を例に説明したが、これに限定されるものではなく、付加情報を推定するようにしてもよい。この場合には、図２２に示すように、非言語情報生成モデル学習装置７１０Ａの学習用演算部７３０Ａは、学習用

入力部７２０によって受け付けた学習データに含まれるテキスト情報から、付加情報を推定して学習用付加情報取得部７３１へ出力する学習用付加情報推定部７３１Ａを備えるようにすればよい。例えば、テキスト情報から、対話の内容や感情を推定することができる。

【０３８３】

また、図２３に示すように、非言語情報生成装置７４０Ａの演算部７６０Ａは、入力部７５０によって受け付けたテキスト情報から、付加情報を推定して付加情報取得部７６１へ出力する付加情報推定部７６１Ａを備えるようにすればよい。

【０３８４】

テキスト情報ではなく、音声情報や動画情報が入力される場合には、音声情報や動画情報から付加情報を推定するようにすればよい。例えば、動画情報から年齢や性別、環境変数、身体的特徴等を推定したり、音声情報から、対話時の人数、感情、国籍、室内/屋内等を推定したりするようにしてもよい。

【０３８５】

上記のように、付加情報を推定する場合には、付加情報にも時刻情報を自動的に付与することができる。また、付加情報を推定する場合、推定する際の単位ごとに、非言語情報の推定値を生成することができる。

【０３８６】

また、非言語情報生成装置では、推定して得られる特定の付加情報に応じて、他の付加情報を変更してもよい。具体的には、特定の付加情報に対して予め定義しておいた指定情報に切り替わるように、他の付加情報を変更してもよい。例えば、音声情報から室内/屋内が切り替わったことが推定出来た際に、身体的特徴である服装の指定内容を変更するように、他の付加情報を変更する。また、感情に応じて、音声合成の発話速度や、テキストの表示速度を変更するようにしてもよい。その他、満腹状態を検知すると体形をふくよかに設定するように付加情報を変更したり、怒りの感情を検出すると対話の内容を「議論」に変更するように付加情報を変更したりする。

【０３８７】

また、付加情報の推定の結果として、所定の処理単位の感情のラベルが時系列で「通常」「怒り」の順で並んでおり、ジェスチャと同時に音声を出力する場合、所定の処理単位毎の付加情報を音声合成部（図示省略）へのパラメータ（話速）として渡す。これにより、感情のラベル「怒り」が推定された所定の処理単位に対して、話速を所定の数足すとか、所定倍するように、テキスト情報の時刻情報を短縮するように変更する（図２４）。なお、図２４では、矢印付きの横軸が、時間軸を示している。

【０３８８】

また、「怒り」が推定された期間の時刻情報付きの音声特徴量の時刻情報を変更するようにしてもよい。

【０３８９】

また、非言語情報が、行動の大きさや回数を含む場合を例に説明したが、これに限定されるものではなく、さらに、ポーズ長の割合を含むようにしてもよい。

【０３９０】

[第８の実施形態]

【０３９１】

次に、本発明の第８の実施形態について説明する。なお、第１の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。

【０３９２】

＜第８の実施形態の概要＞

【０３９３】

入力がテキストの場合に、テキストの詳細な時刻情報が得られないことがある。例えば、（１）音声合成（または音声認識）に、各テキストに対応する必要な粒度の時刻情報を出力する機能が無い場合や、（２）ジェスチャ生成時に、テキストのみを出力する（音声合成する必要が無い）場合で、かつ音声合成を行うリソースが無い、または音声合成処理

時間が確保できない場合（入力が音声の場合は、元の音声情報を出力する必要が無い、またはリソース不足など）である。

【０３９４】

上記（１）のケースについては、ＤＮＮ（Deep Neural Network）ベース合成音声を用いる場合などに発生する。また、上記（２）のケースは、実サービスの際の制約等から発生する。

【０３９５】

そこで、本実施の形態では、上記（１）のような、必要な粒度（例えば文字や形態素）の時刻情報が得られない場合に、音声合成器を利用して、テキストに対応する音声を生成した際に取得可能な、各文節（必要となる粒度より粗いが、最も近いもの）の音声時間長を用いて、必要となる粒度の時刻情報を求める。具体的には、対象言語の発音の特性に合わせた単位数（日本語であればモーラ数、英語であればアクセント数など）により、時刻情報が獲得可能な単位を分割し、必要な単位の時刻情報として利用する。

【０３９６】

＜非言語情報生成モデル学習装置の構成＞

【０３９７】

図２５に示すように、第８の実施形態に係る非言語情報生成モデル学習装置８１０の学習用演算部８３０は、学習用情報取得部２３１と、学習用付加情報取得部７３１と、学習用特徴量抽出部７３２と、非言語情報取得部３３と、生成パラメータ抽出部３４と、学習データ作成部８３３と、学習部７３５と、学習済みモデル記憶部７３６とを備えている。

【０３９８】

学習データ作成部８３３は、学習用情報取得部２３１によって取得された学習用のテキスト情報について、詳細な時刻情報を求め、詳細な時刻情報付きのテキスト情報を生成する。

【０３９９】

ここで、詳細な時刻情報は、学習用のテキストを出力する際の時刻の範囲を、テキストを所定単位で分割したときの分割数で分割した結果に基づいて付与されたものである。

【０４００】

例えば、テキスト情報について得られている時刻情報が、発話単位の時刻情報であり、所定の単位（モーラ単位）の時刻情報を求めたい場合には、以下のように時刻情報を求める。

【０４０１】

まず、テキストの発話長を１に正規化し、時刻情報が得られる単位の発話長を所定の単位の数（この場合はモーラ数）で分割する（図２６Ａ参照）。そして、分割して得られた時刻情報を、所定の単位の時刻情報として利用する。

【０４０２】

また、学習データ作成部８３３は、学習用特徴量抽出部７３２によって抽出された時刻情報付きの学習用のテキスト特徴量及び時刻情報付きの付加情報と、生成パラメータ抽出部３４によって抽出された時刻情報付きの学習用の離散化された非言語情報とについて、上記で求めた詳細な時刻情報を用いて、詳細な時刻情報付きの学習用のテキスト特徴量、学習用の付加情報、及び学習用の非言語情報を生成する。

【０４０３】

学習データ作成部８３３は、詳細な時刻情報付きの学習用のテキスト特徴量、詳細な時刻情報付きの学習用の付加情報、及び詳細な時刻情報付きの学習用の非言語情報の組み合わせを学習データとして、学習部７３５に出力する。

【０４０４】

＜非言語情報生成装置の構成＞

【０４０５】

図２６Ｂに示すように、第８の実施形態に係る非言語情報生成装置８４０の演算部８６０は、情報取得部２６１と、付加情報取得部７６１と、特徴量抽出部７６２と、学習済み

モデル記憶部７６３と、非言語情報生成部７６４と、制御部８７０と、を備えている。
【０４０６】
発現部７０は、上記の第１の実施形態及び第２の実施形態と同様に、テキスト情報を出力すると共に、生成された時刻情報付きの非言語情報が示す行動を、時刻情報に合わせて発現する。また、テキスト情報に対応する音声も出力するようにしてもよい。
【０４０７】
特徴量抽出部７６２が、学習用特徴量抽出部７３２と同様に、情報取得部２６１によって取得されたテキスト情報及び時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出する。また、特徴量抽出部７６２は、学習用特徴量抽出部７３２と同様に、付加情報取得部７６１によって取得された付加情報と、情報取得部２６１によって取得された時刻情報とから、時刻情報付きの付加情報を生成する。
【０４０８】
また、特徴量抽出部７６２が、テキスト情報について、学習データ作成部８３３と同様に、詳細な時刻情報を求め、テキスト情報の詳細な時刻情報を求める。
【０４０９】
また、特徴量抽出部７６２が、時刻情報付きのテキスト特徴量及び付加情報について、上記で求めた詳細な時刻情報を用いて、詳細な時刻情報付きのテキスト特徴量及び付加情報を生成する。
【０４１０】
非言語情報生成部７６４は、特徴量抽出部７６２によって生成された詳細な時刻情報付きのテキスト特徴量及び詳細な時刻情報付きの付加情報と、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルとに基づいて、特徴量抽出部７６２によって抽出された詳細な時刻情報付きのテキスト特徴量及び付加情報に対応する詳細な時刻情報付きの非言語情報を生成する。
【０４１１】
そして、制御部８７０は、生成された詳細な時刻情報付きの非言語情報に対応する生成パラメータが、発現部７０から出力されるように発現部７０を制御すると共に、発現部７０により時刻情報に合わせてテキスト情報を出力するように制御する。
【０４１２】
発現部７０は、制御部８７０による制御に応じて、入力部７５０によって受け付けられたテキスト情報又はテキスト情報に対応する音声情報を詳細な時刻情報に合わせて出力すると共に、非言語情報生成部７６４によって生成された非言語情報を、詳細な時刻情報に合わせて発現させる。
【０４１３】
＜非言語情報生成モデル学習装置８１０の作用＞
【０４１４】
次に、第８の実施形態に係る非言語情報生成モデル学習装置８１０の作用について説明する。まず、非言語情報生成モデル学習装置８１０の学習用入力部７２０に、複数の学習用のテキスト情報と複数の学習用の付加情報と複数の学習用の非言語情報との組み合わせを表す学習データが入力されると、非言語情報生成モデル学習装置８１０は、図２７に示す学習処理ルーチンを実行する。
【０４１５】
まず、ステップＳ３００において、学習用情報取得部２３１は、学習用入力部７２０によって受け付けられた複数の学習データ（具体的には、テキスト情報と非言語情報と付加情報との組み合わせ）のうちの、学習用のテキスト情報を取得する。
【０４１６】
ステップＳ１０２において、非言語情報取得部３３は、学習用入力部７２０によって受け付けられた複数の学習データのうちの、学習用の非言語情報と当該学習用の非言語情報が表す行動が行われるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。
【０４１７】

ステップＳ３０３において、学習用テキスト解析部２３７は、上記ステップＳ３００で取得された学習用のテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、学習用音声合成部２３８は、学習用テキスト解析部２３７によって取得されたテキスト解析結果に基づいて、学習用のテキスト情報に対応する学習用の音声情報を合成する。そして、学習用音声合成部２３８は、学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０４１８】

ステップＳ７００において、学習用付加情報取得部７３１は、学習用入力部７２０によって受け付けられた複数の学習データのうちの、学習用の付加情報を取得する。

【０４１９】

ステップＳ７０２において、学習用特徴量抽出部７３２は、上記ステップＳ３０３で取得された学習用のテキスト情報及び時刻情報から、時刻情報付きの学習用のテキスト特徴量を抽出する。また、学習用特徴量抽出部７３２は、上記ステップＳ７００で取得された学習用の付加情報と、上記ステップＳ３０３で取得された時刻情報とから、時刻情報付きの学習用の付加情報を生成する。

【０４２０】

ステップＳ１０６において、生成パラメータ抽出部３４は、上記ステップＳ１０２で取得された学習用の非言語情報及び時刻情報から、時刻情報付きの離散化された非言語情報を抽出する。

【０４２１】

ステップＳ８００において、学習データ作成部８３３は、上記ステップＳ３０３で取得された学習用のテキスト情報及び時刻情報から、学習用のテキスト情報について、詳細な時刻情報を求める。

【０４２２】

また、学習データ作成部８３３は、学習用特徴量抽出部７３２によって抽出された時刻情報付きのテキスト特徴量及び付加情報と、生成パラメータ抽出部３４によって抽出された時刻情報付きの学習用の離散化された非言語情報とについて、上記で求めた詳細な時刻情報を用いて、詳細な時刻情報付きの学習用のテキスト特徴量、学習用の付加情報、及び学習用の非言語情報を生成する。

【０４２３】

学習データ作成部８３３は、詳細な時刻情報付きの学習用のテキスト特徴量、詳細な時刻情報付きの学習用の付加情報、及び詳細な時刻情報付きの学習用の非言語情報の組み合わせを学習データとして、学習部７３５に出力する。

【０４２４】

ステップＳ７０８において、学習部７３５は、上記ステップＳ８００で生成された詳細な時刻情報付きの学習用のテキスト特徴量及び詳細な時刻情報付きの学習用の付加情報と、ステップＳ８００で取得された詳細な時刻情報付きの学習用の生成パラメータとに基づいて、詳細な時刻情報付きのテキスト特徴量及び付加情報から詳細な時刻情報付きの生成パラメータを生成するための非言語情報生成モデルを学習する。

【０４２５】

ステップＳ１１０において、学習部７３５は、上記ステップＳ７０８で得られた学習済みの非言語情報生成モデルを、学習済みモデル記憶部７３６へ格納して、学習処理ルーチンを終了する。

【０４２６】

＜非言語情報生成装置の作用＞

【０４２７】

次に、第８の実施形態に係る非言語情報生成装置８４０の作用について説明する。まず、非言語情報生成モデル学習装置８１０の学習済みモデル記憶部７３６に格納された学習済みの非言語情報生成モデルが、非言語情報生成装置８４０へ入力されると、非言語情報生成装置８４０の学習済みモデル記憶部７６３へ学習済みの非言語情報生成モデルが格納

される。そして、入力部７５０に、非言語情報生成対象のテキスト情報及び付加情報が入力されると、非言語情報生成装置８４０は、図２８に示す非言語情報生成処理ルーチンを実行する。

【０４２８】

ステップＳ４００において、情報取得部２６１は、入力部７５０によって受け付けられた、テキスト情報を取得する。

【０４２９】

ステップＳ４０１において、テキスト解析部２６５は、上記ステップＳ４００で取得されたテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、音声合成部２６６は、テキスト解析部２６５によって取得されたテキスト解析結果に基づいて、テキスト情報に対応する音声情報を合成する。そして、音声合成部２６６は、音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０４３０】

ステップＳ７５０において、付加情報取得部７６１は、入力部７５０によって受け付けられた、付加情報を取得する。

【０４３１】

ステップＳ７５２において、特徴量抽出部７６２は、上記ステップＳ４０１で取得されたテキスト情報及び時刻情報から、時刻情報付きのテキスト特徴量を抽出し、上記ステップＳ７５０で取得された付加情報及び上記ステップＳ４０１で取得された時刻情報から、時刻情報付きの付加情報を生成する。

【０４３２】

ステップＳ８５０において、特徴量抽出部７６２が、上記ステップＳ４０１で取得されたテキスト情報及び時刻情報から、テキスト情報について詳細な時刻情報を求める。また、特徴量抽出部７６２が、時刻情報付きのテキスト特徴量について、上記で求めた詳細な時刻情報を用いて、詳細な時刻情報付きのテキスト特徴量を生成する。

【０４３３】

また、特徴量抽出部７６２が、時刻情報付きの付加情報について、上記で求めた詳細な時刻情報を用いて、詳細な時刻情報付きの付加情報を生成する。

【０４３４】

ステップＳ７５４において、非言語情報生成部７６４は、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルを読み出す。

【０４３５】

ステップＳ７５６において、非言語情報生成部７６４は、上記ステップＳ８５０で生成された詳細な時刻情報付きのテキスト特徴量及び付加情報と、上記ステップＳ７５４で読み出された学習済みの非言語情報生成モデルとに基づいて、上記ステップＳ８５０で生成された詳細な時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの生成パラメータを生成する。

【０４３６】

ステップＳ８５２において、制御部８７０は、上記ステップＳ４００で取得したテキストと、上記ステップＳ７５６で生成された時刻情報付きの非言語情報とが、時刻情報に合わせて発現部７０から出力されるように発現部７０を制御して、非言語情報生成処理ルーチンを終了する。

【０４３７】

なお、第８の実施形態に係る非言語情報生成装置及び非言語情報生成モデル学習装置の他の構成及び作用については、第１の実施形態と同様であるため、説明を省略する。

【０４３８】

以上説明したように、第８の実施形態に係る非言語情報生成装置によれば、テキスト情報の時刻情報を、所定単位毎の時刻情報に分割して付与することにより、テキスト情報について必要な粒度の時刻情報が得られない場合であっても、所定単位毎の時刻情報に従ってテキスト情報を、非言語情報が示す行動の発現と共に出力することができる。

【０４３９】
なお、上記の実施の形態では、対象言語の発音の特性に合わせた単位数として、モーラ数を用いて時刻情報を分割する場合を例に説明したが、これに限定されるものではなく、英語の場合には、アクセント数を用いて時刻情報を分割してもよい。また、モーラ数やアクセント数以外に、品詞数、シソーラスにおけるカテゴリ数などで時刻情報を分割して付与してもよい。
【０４４０】
また、所定の単位数で時刻情報を分割した後、重みづけを行うようにしてもよい。重みづけは、機械学習により決定してもよいし、重み付け用のＤＢ（対象言語の発音の特性に合わせた単位の種類ごとに、重みを設定してある）をあらかじめ用意しておいてもよい。機械学習のための学習データは、詳細な時刻情報を付与できる音声合成器を用いて作成しても、人手で作成してもよい。
【０４４１】
また、非言語情報生成装置において、テキストのみを出力すればよい場合には、発話テキストの再生速度情報を取得し、発話のテキスト表示に合わせて行動の発現を同期させればよい。また、その際には、非言語情報に付与する時刻情報について所定の単位数で分割せずに、再生される発話長（あるいは文節の時間長）に合わせて、行動を発現させればよい。
【０４４２】
［第９の実施形態］
【０４４３】
次に、本発明の第９の実施形態について説明する。なお、第１の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。
【０４４４】
＜第９の実施形態の概要＞
【０４４５】
非言語情報の生成を行う際に、時系列情報の考慮が不十分な場合や、複数の動作ごとに非言語情報生成モデルを学習し、動作ごとに個別に非言語情報の生成を行う場合には、非言語情報として自然ではないもの、不可能なものが生成される。自然ではない非言語情報とは、同時に行うと適切でない行動であり、例えば、お辞儀しながらジャンプするような行動である。また、不可能な非言語情報とは、時系列順に見ると適切でない行動であり、例えば、１形態素のみに付与された、ゆっくり５回頷く、という行動である。この行動は、１形態素のみに付与されており、時間が足りないため、不可能な行動である。
【０４４６】
そこで、本実施の形態では、生成される非言語情報に関して制約条件を設定し、生成後のデータの修正（挿入・削除・置換）を行う。
【０４４７】
例えば、自然な非言語情報を発現させるため、制約条件を用いて、学習データや生成された非言語情報から不自然な非言語情報を削除する。あるいは、制約条件を用いて非言語情報を追加する。
【０４４８】
制約条件として、非言語情報自体の制約条件、発現部の形状（ＣＧキャラクタ・ロボット）による制約条件、及び付加情報を用いた制約条件の少なくとも１つを用いる。また、制約条件は、ルールの集合として、人手で定義される。
【０４４９】
＜非言語情報生成モデル学習装置の構成＞
【０４５０】
第９の実施形態に係る非言語情報生成モデル学習装置は、第８の実施形態に係る非言語情報生成モデル学習装置８１０の構成と同様であるため、同一符号を付して説明を省略する。

【０４５１】
第９の実施形態に係る非言語情報生成モデル学習装置では、学習データ作成部８３３は、生成パラメータ抽出部３４によって抽出された時刻情報付きの学習用の離散化された非言語情報について、非言語情報の時系列に関する制約条件、又は時刻が対応する非言語情報に関する制約条件を満たすように、学習用の非言語情報又は時刻情報を変更する。
【０４５２】
例えば、行動ごとに必要最低限の時刻の情報またはテキスト単位（文節、形態素など）の数を制約条件として設定しておく。これは、発現部７０の形状や、可能な動作速度、ひとつ前の行動による現在の姿勢などから、制約条件が設定される。
【０４５３】
具体的には、文節単位で行動を生成する場合を例に説明する。ハンドジェスチャＡは、必ず３文節にまたがって行動を行う、という制約条件が設定されている場合には、1文節のみにハンドジェスチャＡが生成されると、その前、あるいは後、またはその両方の文節に、ハンドジェスチャＡのラベルを付与するように変更する。それが不可能な場合は、ハンドジェスチャＡの行動ラベルを削除する。なお、生成された非言語情報の発現が不可能であると判定された場合は、その代替行動を用意しておくとよい
【０４５４】
また、制約条件の設定について、実際の人間の行動データを元に、どの行動がどれくらいの時間（またはテキスト単位数）で生成されるべきかを事前に作成することができる。このとき、上記図２に示すように、発話中の話者の音声情報（Ｘ）を取得するのと同時に、所定の計測装置を用いて発話中の話者の非言語情報（Ｙ）も取得することで作成されるデータを用いる事が出来る。これにより、より人間として自然な動きが可能になる制約条件を設定することができる。なお、発現部７０が人間を模擬したものではない場合には、あえて不自然な動きを許容するように制約条件を設定してもよい。
【０４５５】
また、学習データ作成部８３３は、学習用特徴量抽出部７３２によって抽出された時刻情報付きの付加情報を考慮して、付加情報を用いて設定された制約条件を満たすように、学習用の非言語情報又は時刻情報を変更する。例えば、感情によって話速が上がったりすると、それまではできていた動作が出来なくなったりすることが考えられるため、学習用の非言語情報に付与された時刻情報を変更する。
【０４５６】
学習部７３５は、学習用特徴量抽出部７３２によって抽出された時刻情報付きの学習用のテキスト特徴量及び時刻情報付きの付加情報と、学習データ作成部８３３によって変更された、時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量及び付加情報から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。このとき、系列ラベリング（CRF（Conditional Random Fields））を用いて非言語情報生成モデルを学習することが好ましい。
【０４５７】
この場合には、
【０４５８】
【数５６】

$$Y^{T1,T2}$$

【０４５９】
において、ラベルの時系列関係を考慮する為に、BIO（Begin, Inside, Outside）タグなどの系列情報を付与しても良い。例えば、あるラベルが連続して出現した際に、開始ラベルにB（Begin）ラベルを付与し、以降のラベルにI（Inside）を付与する。これにより、連続したラベルを推定する際に精度が高くなる。

【０４６０】
このようにラベリングしたデータを用いて、系列ラベリングの手法によりジェスチャの生成を行う非言語情報生成モデルを学習する。これにはＳＶＭを用いる事も可能だが、ＨＭＭ（Hidden Markov Model）やＣＲＦ（Conditional Random Fields、参考文献９）を用いるとなお良い。
【０４６１】
［参考文献９］日本特許第５１５２９１８号公報
【０４６２】
＜非言語情報生成装置の構成＞
【０４６３】
第９の実施形態に係る非言語情報生成装置は、第７の実施形態に係る非言語情報生成装置７４０の構成と同様であるため、同一符号を付して説明を省略する。
【０４６４】
第９の実施形態に係る非言語情報生成装置では、非言語情報生成部７６４は、特徴量抽出部７６２によって生成された時刻情報付きのテキスト特徴量及び時刻情報付きの付加情報と、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルとに基づいて、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの非言語情報を生成する。
【０４６５】
非言語情報生成部７６４は、生成された時刻情報付きの非言語情報について、学習データ作成部８３３と同様に、非言語情報の時系列に関する制約条件、又は時刻が対応する非言語情報に関する制約条件を満たすように、非言語情報又は非言語情報に付与された時刻情報を変更する。
【０４６６】
そして、非言語情報生成部７６４は、変更された時刻情報付きの非言語情報に対応する生成パラメータが、当該生成パラメータに付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御する。
【０４６７】
発現部７０は、非言語情報生成部７６４による制御に応じて、入力部７５０によって受け付けられたテキスト情報又はテキスト情報に対応する音声情報を詳細な時刻情報に合わせて出力すると共に、非言語情報生成部７６４によって生成された非言語情報を、詳細な時刻情報に合わせて発現させる。
【０４６８】
＜非言語情報生成モデル学習装置の作用＞
【０４６９】
次に、第９の実施形態に係る非言語情報生成モデル学習装置の作用について説明する。まず、非言語情報生成モデル学習装置の学習用入力部７２０に、複数の学習用のテキスト情報と複数の学習用の付加情報と複数の学習用の非言語情報との組み合わせを表す学習データが入力されると、非言語情報生成モデル学習装置は、図２９に示す学習処理ルーチンを実行する。
【０４７０】
まず、ステップＳ３００において、学習用情報取得部２３１は、学習用入力部７２０によって受け付けられた複数の学習データ（具体的には、テキスト情報と非言語情報と付加情報の組み合わせ）のうちの、学習用のテキスト情報を取得する。
【０４７１】
ステップＳ１０２において、非言語情報取得部３３は、学習用入力部７２０によって受け付けられた複数の学習データのうちの、学習用の非言語情報と当該学習用の非言語情報が表す行動が行われるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。
【０４７２】
ステップＳ３０３において、学習用テキスト解析部２３７は、上記ステップＳ３００で

取得された学習用のテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、学習用音声合成部２３８は、学習用テキスト解析部２３７によって取得されたテキスト解析結果に基づいて、学習用のテキスト情報に対応する学習用の音声情報を合成する。そして、学習用音声合成部２３８は、学習用の音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０４７３】

ステップＳ７００において、学習用付加情報取得部７３１は、学習用入力部７２０によって受け付けられた複数の学習データのうちの、学習用の付加情報を取得する。

【０４７４】

ステップＳ７０２において、学習用特徴量抽出部７３２は、上記ステップＳ３０３で取得された学習用のテキスト情報及び時刻情報から、時刻情報付きの学習用のテキスト特徴量を抽出する。また、学習用特徴量抽出部７３２は、上記ステップＳ７００で取得された学習用の付加情報と、上記ステップＳ３０３で取得された時刻情報とから、時刻情報付きの学習用の付加情報を生成する。

【０４７５】

ステップＳ１０６において、生成パラメータ抽出部３４は、上記ステップＳ１０２で取得された学習用の非言語情報及び時刻情報から、時刻情報付きの離散化された非言語情報を抽出する。

【０４７６】

ステップＳ９００では、学習データ作成部８３３は、上記ステップＳ１０６で取得された時刻情報付きの離散化された非言語情報について、非言語情報の時系列に関する制約条件、又は時刻が対応する非言語情報に関する制約条件を満たすように、非言語情報、または非言語情報に付与された時刻情報を変更する。

【０４７７】

ステップＳ７０８において、学習部７３５は、上記ステップＳ７０２で抽出された時刻情報付きの学習用のテキスト特徴量及び時刻情報付きの学習用の付加情報と、ステップＳ９００で変更された時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量及び付加情報から時刻情報付きの生成パラメータを生成するための非言語情報生成モデルを学習する。

【０４７８】

ステップＳ１１０において、学習部７３５は、上記ステップＳ７０８で得られた学習済みの非言語情報生成モデルを、学習済みモデル記憶部７３６へ格納して、学習処理ルーチンを終了する。

【０４７９】

＜非言語情報生成装置の作用＞

【０４８０】

次に、第９の実施形態に係る非言語情報生成装置の作用について説明する。まず、非言語情報生成モデル学習装置８１０の学習済みモデル記憶部７３６に格納された学習済みの非言語情報生成モデルが、非言語情報生成装置へ入力されると、非言語情報生成装置の学習済みモデル記憶部７６３へ学習済みの非言語情報生成モデルが格納される。そして、入力部７５０に、非言語情報生成対象のテキスト情報及び付加情報が入力されると、非言語情報生成装置は、図３０に示す非言語情報生成処理ルーチンを実行する。

【０４８１】

ステップＳ４００において、情報取得部２６１は、入力部７５０によって受け付けられた、テキスト情報を取得する。

【０４８２】

ステップＳ４０１において、テキスト解析部２６５は、上記ステップＳ４００で取得されたテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、音声合成部２６６は、テキスト解析部２６５によって取得されたテキスト解析結果に基づいて、テキスト情報に対応する音声情報を合成する。そして、音声合成部２６６は、

音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０４８３】

ステップＳ７５０において、付加情報取得部７６１は、入力部７５０によって受け付けられた、付加情報を取得する。

【０４８４】

ステップＳ７５２において、特徴量抽出部７６２は、上記ステップＳ４０１で取得されたテキスト情報及び時刻情報から、時刻情報付きのテキスト特徴量を抽出し、上記ステップＳ７５０で取得された付加情報及び上記ステップＳ４０１で取得された時刻情報から、時刻情報付きの付加情報を生成する。

【０４８５】

ステップＳ７５４において、非言語情報生成部７６４は、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルを読み出す。

【０４８６】

ステップＳ９５０において、非言語情報生成部７６４は、上記ステップＳ７５２で抽出された時刻情報付きのテキスト特徴量及び付加情報と、上記ステップＳ７５４で読み出された学習済みの非言語情報生成モデルとに基づいて、上記ステップＳ７５２で抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの生成パラメータを生成する。そして、非言語情報生成部７６４は、生成された時刻情報付きの生成パラメータについて、非言語情報の時系列に関する制約条件、又は時刻が対応する非言語情報に関する制約条件を満たすように、非言語情報、または非言語情報に付与された時刻情報を変更する。

【０４８７】

ステップＳ２０８において、非言語情報生成部７６４は、上記ステップＳ９５０で変更された時刻情報付きの非言語情報が、当該非言語情報に付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御して、非言語情報生成処理ルーチンを終了する。

【０４８８】

なお、第９の実施形態に係る非言語情報生成装置及び非言語情報生成モデル学習装置の他の構成及び作用については、第１の実施形態と同様であるため、説明を省略する。

【０４８９】

以上説明したように、第９の実施形態に係る非言語情報生成装置によれば、テキスト情報に対応する音声情報を取得し、音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得し、時刻情報付きのテキスト特徴量と時刻情報付きの非言語情報を生成するための学習済みモデルとに基づいて、時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成し、制約条件を満たすように、非言語情報又は非言語情報の時刻情報を変更する。これにより、非言語情報として自然ではないものを排除して、テキスト情報と非言語情報との対応付けを自動的に行い、そのコストを低減させることができる。

【０４９０】

また、第９の実施形態に係る非言語情報生成モデル学習装置によれば、時刻情報付きの学習用の非言語情報について、制約条件を満たすように、非言語情報又は非言語情報の時刻情報を変更する。そして、時刻情報付きの学習用のテキスト特徴量と時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。これにより、非言語情報として自然ではないものを排除して、テキスト情報と非言語情報との対応付けのコストを低減させつつ、テキスト特徴量から非言語情報を生成するための非言語情報生成モデルを得ることができる。なお、制約条件を元に、書き換えに関する機械学習モデルを作成してもよい。

【０４９１】

［第１０の実施形態］

【０４９２】
次に、本発明の第１０の実施形態について説明する。なお、第１の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。
【０４９３】
＜第１０の実施形態の概要＞
【０４９４】
制約条件により不可能な非言語情報を検出した場合に、制約条件を満たすように、音声データ（合成音声を含む）にポーズを入れる、またはテキストの表示速度（合成音声の話速）を変更するために、テキスト情報に付与された時刻情報を変更する点が、第９の実施形態と異なっている。
【０４９５】
本実施形態は、非言語情報の方が、音声（テキスト）より重要な場合に有効である。特に、出力である音声を音声合成で作成する場合や、テキストを出力する場合において、高い効果が見込める。
【０４９６】
＜非言語情報生成モデル学習装置の構成＞
【０４９７】
第１０の実施形態に係る非言語情報生成モデル学習装置は、第８の実施形態に係る非言語情報生成モデル学習装置８１０の構成と同様であるため、同一符号を付して説明を省略する。
【０４９８】
第１０の実施形態に係る非言語情報生成モデル学習装置では、学習データ作成部８３３は、学習用のテキスト情報、及び学習用特徴量抽出部７３２によって取得された時刻情報付きの学習用のテキスト特徴量について、非言語情報の時系列に関する制約条件を満たすように、学習用のテキスト情報の時刻情報、及び学習用のテキスト特徴量に付与された時刻情報を変更する。
【０４９９】
例えば、制約条件を満たすように、非言語情報に合わせてポーズを入れるように、学習用のテキスト情報及び学習用のテキスト特徴量に付与された時刻情報を変更したり、非言語情報に合わせてテキストの表示速度（合成音声の話速）を変更するように、学習用のテキスト情報及びテキスト特徴量に付与された時刻情報を変更したりする。
【０５００】
また、学習データ作成部８３３は、学習用特徴量抽出部７３２によって抽出された時刻情報付きの付加情報を考慮して、付加情報を用いて設定された制約条件を満たすように、テキスト情報に付与された時刻情報を変更する。
【０５０１】
学習部７３５は、学習データ作成部８３３によって変更された時刻情報付きの学習用のテキスト特徴量と、学習用特徴量抽出部７３２によって抽出された時刻情報付きの付加情報と、生成パラメータ抽出部３４によって抽出された、時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量及び付加情報から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。
【０５０２】
＜非言語情報生成装置の構成＞
【０５０３】
第１０の実施形態に係る非言語情報生成装置は、第７の実施形態に係る非言語情報生成装置７４０の構成と同様であるため、同一符号を付して説明を省略する。
【０５０４】
第１０の実施形態に係る非言語情報生成装置では、非言語情報生成部７６４は、特徴量抽出部７６２によって生成された時刻情報付きのテキスト特徴量及び時刻情報付きの付加情報と、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルとに

基づいて、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの非言語情報を生成する。

【０５０５】

非言語情報生成部７６４は、テキスト情報及び時刻情報付きのテキスト特徴量について、学習データ作成部８３３と同様に、非言語情報の時系列に関する制約条件を満たすように、時刻情報を変更する。

【０５０６】

そして、非言語情報生成部７６４は、生成された時刻情報付きの非言語情報に対応する生成パラメータが、当該生成パラメータに付与された時刻情報に基づいて発現部７０から出力されるように発現部７０を制御する。

【０５０７】

発現部７０は、非言語情報生成部７６４による制御に応じて、入力部７５０によって受け付けられたテキスト情報又はテキスト情報に対応する音声情報を、変更された時刻情報に合わせて出力すると共に、非言語情報生成部７６４によって生成された非言語情報を、時刻情報に合わせて発現させる。

【０５０８】

なお、第１０の実施形態に係る非言語情報生成装置及び非言語情報生成モデル学習装置の他の構成及び作用については、第９の実施形態と同様であるため、説明を省略する。

【０５０９】

以上説明したように、第１０の実施形態に係る非言語情報生成装置によれば、テキスト情報に対応する音声情報を取得し、音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得し、時刻情報付きのテキスト特徴量と時刻情報付きの非言語情報を生成するための学習済みモデルとに基づいて、時刻情報付きのテキスト特徴量に対応する時刻情報付きの非言語情報を生成し、制約条件を満たすように、テキスト情報の時刻情報を変更する。これにより、自然ではないものを排除して、テキスト情報と非言語情報との対応付けを自動的に行い、そのコストを低減させることができる。

【０５１０】

また、第１０の実施形態に係る非言語情報生成モデル学習装置によれば、時刻情報付きの学習用のテキスト特徴量と学習用の非言語情報について、制約条件を満たすように、テキスト特徴量の時刻情報を変更する。そして、時刻情報付きの学習用のテキスト特徴量と時刻情報付きの学習用の非言語情報とに基づいて、時刻情報付きのテキスト特徴量から時刻情報付きの非言語情報を生成するための非言語情報生成モデルを学習する。これにより、自然ではないものを排除して、テキスト情報と非言語情報との対応付けのコストを低減させつつ、テキスト特徴量から非言語情報を生成するための非言語情報生成モデルを得ることができる。

【０５１１】

［第１１の実施形態］

【０５１２】

次に、本発明の第１１の実施形態について説明する。なお、第１の実施形態と同様の構成となる部分については、同一符号を付して説明を省略する。

【０５１３】

＜第１１の実施形態の概要＞

【０５１４】

ジェスチャシナリオや非言語情報生成モデルのための学習データを作成する際には、発話に対して適切な動作、すなわち意図した通りの動作が行われるかの確認や修正作業が必須となる。しかし、動作の確認や修正作業の際に、どのような発話に対し、どのような動作が割り当てられているか分かりづらく、作業コストが高くなりやすい。そこで、本実施の形態では、どのようなテキスト情報に対し、どのような非言語情報が割り当てられているのかを可視化することで、動作の確認や修正作業を容易にする。

【０５１５】

具体的には、所定の単位で分割された言語情報（テキスト又は音声）に対し、非言語情報を対応付けて表示し、各所定の単位ごとに実際の動作を確認可能にすることで、修正しやすいインタフェースを提供する。また、修正結果に応じて、学習データを追加／修正することができ、さらに、非言語情報生成モデルを再学習することができる。

ここで、本実施の形態で説明するインタフェースの使用シーンは、例えば、以下の５つである。

（１）ジェスチャシナリオを作成する場合に、本実施の形態に係るインタフェースが使用される。例えば、入力されたテキスト情報について、学習済みの非言語情報生成モデルに基づいて時刻情報付きの非言語情報を生成し、ユーザの操作により、生成された非言語情報を修正し、修正結果を、固定シナリオとして出力する。

（２）学習データを修正する場合に、本実施の形態に係るインタフェースが使用される。例えば、入力された学習データを読み込み、ユーザの操作により、学習データに含まれるテキスト情報又は非言語情報を修正し、修正結果を、学習データとして出力する。

（３）学習データを追加する場合に、本実施の形態に係るインタフェースが使用される。例えば、入力されたテキスト情報について、学習済みの非言語情報生成モデルに基づいて時刻情報付きの非言語情報を生成し、ユーザの操作により、生成された非言語情報を修正し、修正結果を、当該非言語情報生成モデルに対応する学習データとして出力する。

（４）学習済みの非言語情報生成モデルを再学習する場合に、本実施の形態に係るインタフェースが使用される。例えば、入力されたテキスト情報について、学習済みの非言語情報生成モデルに基づいて時刻情報付きの非言語情報を生成し、ユーザの操作により、生成された非言語情報を修正し、修正結果を、当該非言語情報生成モデルに対応する学習データとして追加して、当該非言語情報生成モデルを再学習する。

（５）上記第９の実施の形態、第１０の実施の形態で説明した制約条件を生成する場合に、本実施の形態に係るインタフェースが使用される。例えば、上記（１）～（４）の何れかと同様の方法で得られた修正結果を用いて、制約条件を規定する。

【０５１６】

＜非言語情報生成装置の構成＞

【０５１７】

図３１は、第１１の実施形態に係る非言語情報生成装置１１４０の構成の一例を示すブロック図である。図３１に示すように、第１１の実施形態に係る非言語情報生成装置１１４０は、ＣＰＵと、ＲＡＭと、後述する非言語情報生成処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成されている。非言語情報生成装置１１４０は、機能的には、入力部７５０と、演算部１１６０と、表示部１１９０と、出力部１１９２を備えている。

【０５１８】

入力部７５０は、テキスト情報及び付加情報を受け付ける。受け付ける付加情報は、所定の処理単位毎（例えば、形態素毎や文節毎）の付加情報である。なお、付加情報が、所定の処理単位毎に変化しないもの（例えば、性別）である場合には、所定の処理単位毎に受け付けなくてもよい。この場合には、付加情報を受け付けた際に、装置側で、所定の処理単位毎の付加情報に展開すればよい。

【０５１９】

演算部１１６０は、情報取得部２６１と、付加情報取得部７６１と、特徴量抽出部７６２と、学習済みモデル記憶部７６３と、非言語情報生成部７６４と、制御部１１７０と、学習データ生成部１１７２と、再学習制御部１１７４とを備えている。なお、付加情報取得部７６１を省略してもよい。上記（１）、（５）の使用シーンでは、更に、学習データ生成部１１７２及び再学習制御部１１７４を省略してもよい。また、上記（２）、（３）の使用シーンでは、更に、再学習制御部１１７４を省略してもよい。

【０５２０】

情報取得部２６１は、上記第２の実施の形態に係る非言語情報生成装置２４０の情報取得部２６１と同様に、入力部７５０によって受け付けられたテキスト情報を取得すると共

に、テキスト情報に対応する音声情報を取得し、当該音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

付加情報取得部７６１は、上記第７の実施の形態に係る非言語情報生成装置７４０の付加情報取得部７６１と同様に、入力部７５０によって受け付けられた付加情報を取得する。

特徴量抽出部７６２は、上記第７の実施の形態に係る非言語情報生成装置７４０の特徴量抽出部７６２と同様に、情報取得部２６１によって取得されたテキスト情報及び時刻情報から、該テキスト情報の特徴量を表す、時刻情報付きのテキスト特徴量を抽出する。また、特徴量抽出部７６２は、付加情報取得部７６１によって取得された付加情報と、情報取得部２６１によって取得された時刻情報とから、時刻情報付きの付加情報を生成する。

学習済みモデル記憶部７６３には、上記第７の実施の形態に係る非言語情報生成装置７４０の学習済みモデル記憶部７６３と同様に、学習済みモデル記憶部７３６に格納されている学習済みの非言語情報生成モデルと同一の学習済みの非言語情報生成モデルが格納されている。

非言語情報生成部７６４は、上記第７の実施の形態に係る非言語情報生成装置７４０の非言語情報生成部７６４と同様に、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び時刻情報付きの付加情報と、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルとに基づいて、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの非言語情報を生成する。

制御部１１７０は、非言語情報生成部７６４によって生成された時刻情報付きの非言語情報と、入力部７５０によって受け付けられたテキスト情報及び付加情報とを表示するように表示部１１９０を制御する。

【０５２１】

表示部１１９０は、表示画面１１９０Ａと、発現部１１９０Ｂとを備えている。なお、本実施の形態では、発現部１１９０Ｂが、表示部１１９０に含まれている場合を例に説明するが、これに限定されるものではなく、表示部１１９０とは別の装置（例えば、ロボット）で構成されていてもよい。

【０５２２】

発現部１１９０Ｂは、テキスト情報に対応する音声を出力するとともに、生成された時刻情報付きの非言語情報が示す行動を、時刻情報に合わせて発現する。あるいは、テキスト情報を含む吹き出しを表示するようにしてもよい。

【０５２３】

このとき、表示部１１９０により表示される表示画面１１９０Ａの一例を図３２に示す。

【０５２４】

表示画面１１９０Ａでは、テキストを所定単位毎に分割して表示すると共に、テキスト特徴量に付与された時刻情報及び非言語情報に付与された時刻情報に基づいて、テキストの所定単位毎に対応付けて、非言語情報を示すラベルを表示する。また、表示画面１１９０Ａでは、テキスト情報に対応する音声の音声波形も表示するようにしてもよい。

【０５２５】

なお、付与された時刻情報は、テキストを出力する際の時刻であり、例えば、上記第８の実施の形態と同様に、テキストを出力する際の時刻の範囲を、テキストを所定単位で分割したときの分割数で分割した結果に基づいて付与されたものであってもよい。

【０５２６】

また、表示部１１９０は、非言語情報を示す行動を発現させる発現部１１９０Ｂを含み、発現部１１９０Ｂによる行動の発現の開始、停止、所定単位ずつ（例えば、１形態素ずつや、１文節ずつ）の早送り、又は巻き戻しの指示を受付可能な状態で表示画面１１９０Ａを表示する。例えば、再生ボタン、ポーズボタン、巻き戻しボタン、及び早送りボタンを表示画面１１９０Ａに表示する。

【０５２７】

なお、発現部１１９０Ｂによる行動の発現の早送り、又は巻き戻しの指示を受付可能な

スライドバーを表示画面１１９０Ａに表示してもよい。
【０５２８】
制御部１１７０は、発現の開始、停止、早送り、又は巻き戻しの指示を受け付けると、当該指示に応じて、発現部１１９０Ｂによる行動の発現を制御する。
また、表示部１１９０は、発現部１１９０Ｂによって発現されている行動が、テキスト中のどの部分に対応するものであるか識別可能となるように表示画面１１９０Ａに表示してもよい。例えば、発現部１１９０Ｂによって発現されている行動に対応するテキスト中の対応部分に、再生バーを表示したり、発現部１１９０Ｂによって発現されている行動に対応するテキスト中の対応部分のマス目の色を変更したり、点滅させたりしてもよい。
【０５２９】
また、表示部１１９０は、付加情報の設定を受付可能な状態で表示画面１１９０Ａを表示する。制御部１１７０は、付加情報の設定を受け付けると、当該付加情報を、特徴量抽出部７６２へ出力し、当該付加情報を更に用いて非言語情報生成部７６４により生成された非言語情報を示すラベルと、テキストとを表示画面１１９０Ａに表示させるように表示部１１９０を制御する。
【０５３０】
また、表示部１１９０は、非言語情報を示すラベルの変更指示を受付可能な状態で表示画面１１９０Ａを表示する。
【０５３１】
学習データ生成部１１７２は、非言語情報を示すラベルの変更指示を受け付けると、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び付加情報と、変更指示に従って変更された非言語情報を示すラベルとの組み合わせを、非言語情報生成モデルを学習するための学習データとして生成する。
【０５３２】
また、表示部１１９０は、非言語情報生成モデルの再学習指示、及び学習データ生成部１１７２により生成された学習データに対する重みの設定を受付可能な状態で表示画面１１９０Ａを表示する。ここで、学習データに対する重みとは、再学習時に、既存の学習データと比較して、追加する学習データをどれぐらい重視するかに応じて設定されるものである。例えば、この重みを最大値に設定した場合、追加した学習データ中の対（Ｘ，Ｙ）のＸに対し、必ずＹが生成されるように、非言語情報生成モデルが再学習される。
【０５３３】
再学習制御部１１７４は、再学習指示及び重みの設定を受け付けると、学習データ生成部１１７２によって生成された学習データ及び設定された重みを用いて、非言語情報生成モデル学習装置７１０に非言語情報生成モデルを学習させる。
【０５３４】
具体的には、表示画面１１９０Ａに対して以下のＳｔｅｐ１～Ｓｔｅｐ５のようにユーザにより操作される。
【０５３５】
（Ｓｔｅｐ１）非言語情報を生成する対象となる発話文を示すテキストを入力するか、あるいは選択することでテキストを設定する。例えば、上記（１）の使用シーンのように、ジェスチャシナリオを生成する場合や、上記（３）、（４）の使用シーンのように、学習データを追加する場合には、テキストを入力する。また、上記（２）の使用シーンのように、学習データを修正する場合には、学習データの集合が提示され、修正対象となる学習データのテキストを選択する。
【０５３６】
（Ｓｔｅｐ２）発話文を示すテキストが所定単位ごとに分割され、分割された単位ごとに生成された非言語情報を示すラベルＹが表示される。
【０５３７】
（Ｓｔｅｐ３）発現の開始が指示されると、生成された非言語情報により、発現部１１９０Ｂが動く。

【０５３８】
（Ｓｔｅｐ４）ユーザが、発現部１１９０Ｂの動作を目視で確認する。
【０５３９】
（Ｓｔｅｐ５）おかしな動きをする時のマスＭ（ラベル付きのマスＭor空白のマスＭ）をクリックすると、正しい非言語情報を示すラベルに書き換えることができる。その結果、当該ラベルが、入力発話に対する、正解の非言語情報を示すラベルとして、学習データに追加できるようにしてもよい（その際、重みづけも設定できるようにしてもよい）。
【０５４０】
なお、所定単位ごとの時刻情報を表示し、時刻情報の変更指示を受け付け可能に表示画面１１９０Ａを表示してもよい（図３３参照）。例えば、時刻情報の値をクリックすることで値が編集可能となり、時刻情報の値を変更することができる。あるいは、所定単位ごとの開始時刻を示す縦棒の位置を左右に変更することにより、所定単位ごとの開始時刻を変更することができる。
【０５４１】
また、音声データ（合成音声を含む）にポーズを入れる変更指示を受け付け可能に表示画面１１９０Ａに表示してもよい。例えば、図３４に示すように、テキスト情報の所定の単位毎に、ポーズＰの開始位置（図３４の点線の縦線参照）を挿入する変更指示を受け付けると共に、ポーズＰの開始位置を調整することによりポーズ長の割合を変更する変更指示を受け付けることができる。また、上記第１０の実施の形態で説明したように制約条件を満たすように挿入されたポーズに対しても同様に、ポーズＰの開始位置を調整することによりポーズ長の割合を変更する変更指示を受け付けるようにしてもよい。
【０５４２】
また、所定単位（例えば文節）ごとのテキスト特徴量と非言語情報に対応する生成パラメータとが、変更指示を受付可能に表示されても良い。例えば、図３２に示すように、ユーザが、テキスト情報のマスＭまたは非言語情報のマスＭにマウスカーソルを合わせ、右クリックした際に、マスＭに対応するテキスト特徴量がオーバーレイ表示され、変更指示が受付可能となるようにしてもよい。また、テキスト情報のマスＭに対応するテキスト特徴量がオーバーレイ表示される場合、マスＭのテキスト情報から抽出されたテキスト特徴量全てが表示され、非言語情報のマスＭに対応するテキスト特徴量がオーバーレイ表示される際、非言語情報が生成される根拠となったテキスト特徴量がオーバーレイ表示されるようにしてもよい。
【０５４３】
第１１の実施形態に係る非言語情報生成モデル学習装置は、第７の実施形態に係る非言語情報生成モデル学習装置７１０と同様であるため、同一符号を付して説明を省略する。
【０５４４】
＜非言語情報生成装置の作用＞
【０５４５】
次に、第１１の実施形態に係る非言語情報生成装置の作用について説明する。まず、非言語情報生成モデル学習装置７１０の学習済みモデル記憶部７３６に格納された学習済みの非言語情報生成モデルが、非言語情報生成装置へ入力されると、非言語情報生成装置の学習済みモデル記憶部７６３へ学習済みの非言語情報生成モデルが格納される。そして、入力部７５０に、非言語情報生成対象のテキスト情報及び付加情報が入力されると、非言語情報生成装置は、図３５に示す非言語情報生成処理ルーチンを実行する。
【０５４６】
ステップＳ４００において、情報取得部２６１は、入力部７５０によって受け付けられた、テキスト情報を取得する。
【０５４７】
ステップＳ４０１において、テキスト解析部２６５は、上記ステップＳ４００で取得されたテキスト情報に対して所定のテキスト解析を行い、テキスト解析結果を取得する。また、音声合成部２６６は、テキスト解析部２６５によって取得されたテキスト解析結果に

基づいて、テキスト情報に対応する音声情報を合成する。そして、音声合成部２６６は、音声情報が発せられるときの開始時刻から終了時刻までの時刻を表す時刻情報を取得する。

【０５４８】

ステップＳ７５０において、付加情報取得部７６１は、入力部７５０によって受け付けられた、付加情報を取得する。

【０５４９】

ステップＳ７５２において、特徴量抽出部７６２は、上記ステップＳ４０１で取得されたテキスト情報及び時刻情報から、時刻情報付きのテキスト特徴量を抽出し、上記ステップＳ７５０で取得された付加情報及び上記ステップＳ４０１で取得された時刻情報から、時刻情報付きの付加情報を生成する。

【０５５０】

ステップＳ７５４において、非言語情報生成部７６４は、学習済みモデル記憶部７６３に格納された学習済みの非言語情報生成モデルを読み出す。

【０５５１】

ステップＳ７５６において、非言語情報生成部７６４は、上記ステップＳ７５２で抽出された時刻情報付きのテキスト特徴量及び付加情報と、上記ステップＳ７５４で読み出された学習済みの非言語情報生成モデルとに基づいて、上記ステップＳ７５２で抽出された時刻情報付きのテキスト特徴量及び付加情報に対応する時刻情報付きの生成パラメータを生成する。

【０５５２】

ステップＳ１１００において、制御部１１７０は、非言語情報生成部７６４によって生成された時刻情報付きの非言語情報と、入力部７５０によって受け付けられたテキスト情報及び付加情報とを表示するように表示部１１９０を制御して、非言語情報生成処理ルーチンを終了する。

【０５５３】

上記ステップＳ１１００の処理は、図３６に示す処理ルーチンによって実現される。

【０５５４】

まず、ステップＳ１１５０において、制御部１１７０は、テキストを所定単位毎に分割して表示すると共に、テキスト特徴量に付与された時刻情報及び非言語情報に付与された時刻情報に基づいて、テキストの所定単位毎に対応付けて、非言語情報を示すラベルを表示画面１１９０Ａに表示する。

【０５５５】

ステップＳ１１５２では、制御部１１７０は、表示画面１１９０Ａに対する操作を受け付けたか否かを判定する。表示画面１１９０Ａに対する操作を制御部１１７０が受け付けた場合には、処理がステップＳ１１５４へ進む。

【０５５６】

ステップＳ１１５４において、制御部１１７０は、上記ステップＳ１１５２で受け付けた操作の種別が、変更指示であるか、発現指示であるか、再学習指示であるかを判定する。受け付けた操作が、付加情報の設定や、非言語情報を示すラベルの変更指示である場合には、ステップＳ１１５６において、制御部１１７０は、変更指示に従った変更を反映した結果を表示画面１１９０Ａに表示する。受け付けた操作が付加情報の設定であった場合には、制御部１１７０は、更に、当該付加情報を、特徴量抽出部７６２へ出力し、当該付加情報を更に用いて非言語情報生成部７６４により生成された非言語情報を示すラベルと、テキストとを表示画面１１９０Ａに表示する。

【０５５７】

また、受け付けた操作が非言語情報を示すラベルの変更指示であった場合には、学習データ生成部１１７２は、特徴量抽出部７６２によって抽出された時刻情報付きのテキスト特徴量及び付加情報と、変更指示に従って変更された非言語情報を示すラベルとの組み合わせを、非言語情報生成モデルを学習するための学習データとして生成して、非言語情報生成モデル学習装置７１０へ出力する。そして、処理が上記ステップＳ１１５２へ戻る。

【０５５８】
　また、発現部１１９０Ｂによる行動の発現の開始、停止、１文節ずつの早送り、又は巻き戻しの指示を受け付けた場合には、ステップＳ１１５８において、制御部１１７０は、受け付けた指示に従って、発現部１１９０Ｂによる行動の発現を制御して、処理が上記ステップＳ１１５２へ戻る。
【０５５９】
　また、再学習指示及び重みの設定を受け付けた場合には、ステップＳ１１６０において、制御部１１７０は、学習データ生成部１１７２によって生成された学習データ及び設定された重みを用いて、非言語情報生成モデル学習装置７１０に非言語情報生成モデルを学習させて、処理ルーチンを終了する。
【０５６０】
　上記のように、入力されたテキスト情報及び付加情報に対して非言語情報生成モデルを用いて非言語情報が生成された結果に対して、ユーザが修正を行うことによって、非言語情報生成モデルの学習データを生成することができ、非言語情報生成モデルの学習データを追加することができる。また、非言語情報生成モデルの再学習をユーザが指示することにより、追加された学習データを用いた再学習を行って、非言語情報生成モデルを更新することができる。
【０５６１】
　以上説明したように、第１１の実施形態に係る非言語情報生成装置によれば、どのようなテキスト情報に対し、どのような非言語情報が割り当てられているのかを可視化することにより、非言語情報の修正作業を容易にする。
【０５６２】
　なお、ユーザが修正を行うことによって生成された時刻情報付きのテキスト情報及び非言語情報からなるジェスチャシナリオを、固定シナリオとして出力するようにしてもよい。
【０５６３】
　また、上記第９の実施の形態又は第１０の実施の形態で説明した方法と同様に、制約条件を元に時刻情報付きのテキスト情報及び非言語情報の組み合わせに対して書き換えを行った場合に、書き換え結果に対してユーザによる修正を行い、そのデータを学習データとして、書き換えに関する機械学習モデルを作成するようにしてもよい。
【０５６４】
　また、制約条件を元に時刻情報付きのテキスト情報及び非言語情報の組み合わせに対して書き換えを行った場合には、書き換え履歴や、どの制約条件が適用させたかを表示するようにしてもよいし、さらに、制約条件自体の修正を受け付けるようにしてもよい。
【０５６５】
　また、学習データから、当該非言語情報を示すラベルと対応付けられたテキスト特徴量のランキングを求めて、ユーザに提示してもよい。この場合には、まず、学習データからテキスト特徴量と非言語情報を示すラベルの対を取得し、非言語情報を示すラベルごとに、対となっているテキスト特徴量の種類と、その出現数をカウントする。そして、非言語情報を示すラベルごとに、カウント数が多い順にテキスト特徴量の種類を並べ替えたものを、テキスト特徴量のランキングとして提示すればよい。
　また、テキスト特徴量のランキングを提示した後、学習データの選択を受け付け、選択された学習データの編集を受け付けるようにしてもよい。例えば、図３２に示すように、ユーザが、非言語情報のマスＭにマウスカーソルを合わせ、右クリックした際に、マスＭのラベルに対する特徴量のランキングがオーバーレイ表示される。そして、特徴量のランキングの各特徴量名の選択指示（例えば、クリック）が受付可能となっており、特徴量名が選択されると、当該マスＭの非言語情報ラベルと選択されたテキスト特徴量とからなる学習データが表示される。このとき、学習データは、直接編集可能に表示され、削除や追記、編集等が行われることにより、学習データが編集される。
　また、学習データを修正した場合に、修正前の学習データを負例として追加できるようにしてもよい。

【０５６６】
　また、音声合成のパラメータ（話速など）の修正を受付可能に表示するようにしてもよい。
【０５６７】
　なお、上記第７の実施形態～第１１の実施形態では、入力情報がテキスト情報である場合を例に説明したが、これに限定されるものではない。例えば、入力情報が音声情報であってもよい。入力情報が音声情報である場合の、非言語情報生成モデル学習装置における学習用情報取得部は、上記第１の実施形態の学習用情報取得部３１と同様である。入力情報が音声情報である場合の、非言語情報生成装置における情報取得部は、上記第１の実施形態の情報取得部６１と同様である。
【０５６８】
　例えば、上記各実施形態における、情報取得部（又は学習用情報取得部）と特徴量抽出部（又は学習用特徴量抽出部）との組み合わせに対応する構成は、上記図１５に示される全４パターンである。また、学習時と非言語情報生成時との構成の組み合わせとして、可能なバリエーションは、上記図１６に示されるパターンとなる。
【０５６９】
　また、上記第４の実施形態～第６の実施形態で説明した、非言語情報生成モデル学習装置において用いる学習データとして、例えば、上記図２に示されるようなシーンにおいて、発話中の話者の音声情報（Ｘ）を取得するのと同時に、計測装置を用いて発話中の話者の話し相手である会話相手の非言語情報（Ｙ）を取得することを、上記第７の実施形態～第１１の実施形態の各々に適用してもよい。
【０５７０】
[実験例]
【０５７１】
　次に、上記第５の実施形態に関する実験例について説明する。
【０５７２】
[実験データについて]
【０５７３】
　２者対話を対象に、発話を表すテキスト情報およびそれに伴う頷きを表す非言語情報を含む、コーパスデータの構築を行った。２者対話の参加者は、20～50代の日本人男性・女性であり初対面であった。参加者は計24人（12ペア）であった。参加者は互いに向き合って着座した。対話内容は、発話に伴う頷きに関するデータを多く収集するために、アニメーションの説明課題を採用した。参加者はそれぞれ、異なる内容のアニメーションを視聴した後、対話相手にアニメーション内容を説明した。10分間の対話セッションにおいて、１人の参加者が対話相手にアニメーションの内容を詳細に説明した。対話相手は、説明者に自由に質問をし、自由に会話を行うことを許可した。発話の収録のために各被験者の胸につけられた指向性ピンマイクを用いた。対話状況の全体的な外観や参加者の様子の収録として、ビデオカメラを用いた。映像は30Hzで収録された。下記に、取得したテキスト情報及び非言語情報を示す。
【０５７４】
　発話を表すテキスト情報：人手で音声情報から発声言語の書き起こしを行った後、発話内容から文を分割した。さらに、係り受け解析エンジン（参考文献１０および参考文献１１を参照）を利用して、各文を文節に分割した。分割された文節数は11877であった。
【０５７５】
[参考文献１０] 今村賢治，「系列ラベリングによる準話し言葉の日本語係り受け解析」，言語処理学会第13回年次大会発表論文集，pp. 518-521，2007.
[参考文献１１] E. Charniak，"A Maximum-Entropy-Inspired Parser"，Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference，pp. 132-139，2000.
【０５７６】

頷きを表す非言語情報：人手で映像から頷きが発生した区間をラベリングした。連続的に発生した頷きは1つの頷きイベントとして扱った。

【０５７７】

人手によるラベリング（アノテーション）では上記のすべてのデータを30Hzの時間分解能で統合した。

【０５７８】

[非言語情報生成モデル]

【０５７９】

構築したコーパスデータを用いて、単語、その品詞およびシソーラス、単語位置、テキスト情報全体の対話行為を入力として、文節単位ごとに頷きを表す非言語情報を生成する非言語情報生成モデルを構築した。それぞれのテキスト情報が有効であるかを検証するために、各テキスト特徴量を用いた非言語情報生成モデルと、全てのテキスト特徴量を用いた非言語情報生成モデルを構築した。具体的に、文節単位ごとに、対象となる文節およびその前後の文節から得られるテキスト特徴量を入力として、頷きの有無の2値を出力する非言語情報生成モデルを、決定木アルゴリズムC4.5（参考文献１２を参照）を用いて実装した。使用したテキスト特徴量は以下の通りである。

【０５８０】

[参考文献１２] J. R. Quinlan， "Improved use of continuous attributes in c4.5"， Journal of Artificial Intelligence Research, 4:77-90, 1996.

【０５８１】

文字数：文節内の文字数

【０５８２】

位置：文頭、文末からの文節の位置

【０５８３】

単語：形態素解析ツールJtag（参考文献１３を参照）により抽出された文節内の単語情報（Bag-of-words）

【０５８４】

品詞：Jtagにより抽出された文節内の単語の品詞情報

【０５８５】

シソーラス：日本語語彙大系（参考文献１４を参照）に基づく文節内の単語のシソーラス情報

【０５８６】

対話行為：単語n-gramおよびシソーラス情報を用いた対話行為推定手法（参考文献４、１５を参照）により文ごとに抽出された対話行為（33種類）

【０５８７】

[参考文献１３] Takeshi Fuchi and Shinichiro Takagi, "Japanese morphological analyzer using word co-occurrence -Jtag-", In Proceedings of International conference on Computational linguistics, pages 409-413, 1998.

【０５８８】

[参考文献１４] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦，「日本語語彙大系」， 岩波書店，1997.

【０５８９】

[参考文献１５] Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami, and Kohji Dohsaka, "Controlling listening-oriented dialogue using partially observable markov decision processes", In Proceedings of International conference on computational linguistics, pages 761-769, 2010.

【０５９０】

[実験結果について]

【０５９１】

24人の参加者のデータのうち、23人のデータを学習に用いて、残りの1人のデータで評

価を行う24交差検定法により評価を行った。これにより、他者のデータからどれだけ頷きを生成できるかを評価した。なお、各施行で頷きの有無データは、データ量が一致するように、データ数の少ない方に合わせてデータ数を削減した。よって、ベースラインであるチャンスレベルは0.50となる。性能評価結果の平均値を表５に示す。

【０５９２】

非言語情報生成モデルの評価の結果、語彙大系、品詞、単語の順で精度が良い事が分かった。機械学習においては、抽出するテキスト特徴量の分解能が高すぎる（種類数が多い）と、各テキスト特徴量の出現頻度が相対的に低くなり、学習データに一度も現れなかったテキスト特徴量が実行時に出現することが増えるため、生成の精度が低下する傾向がある。また逆に、分解能を低く（抽象化することで、種類数を少なくする）していくほど、前記した問題は起きなくなるが、データの違いが表現できなくなり、生成の精度が低下する傾向にある。

【０５９３】

シソーラス情報は単語を意味や属性でクラス化したもので、その種類数は単語より少なく、品詞より多いため、本実験の学習データ量であっても効率的に学習ができたと考えられる。シソーラス情報は階層構造を持っており、複数段階での単語の上位概念化（抽象化）が可能であるため、学習データのサイズに合わせ、抽象化の程度をコントロールしやすい。

【０５９４】

膨大な量のコーパスデータを作成するコストは高く、困難である。学習データが十分に用意できない際には、シソーラス情報を用いる事で、比較的少ないデータ量でもより良い学習効果が期待できる。

【０５９５】

【表５】

特徴量	適合率	再現率	F 値
チャンスレベル	0.500	0.500	0.500
文字数	0.561	0.556	0.558
単語	0.357	0.529	0.431
品詞	0.522	0.528	0.525
語彙大系	0.615	0.538	0.579
全て	0.578	0.601	0.593

【産業上の利用可能性】

【０５９６】

本発明は、例えば、発話の再生に合わせて非言語行動を表出させる技術に利用可能である。本発明によれば、音声情報又はテキスト情報の少なくとも一方と非言語情報との対応付けを自動化することができる。

【符号の説明】

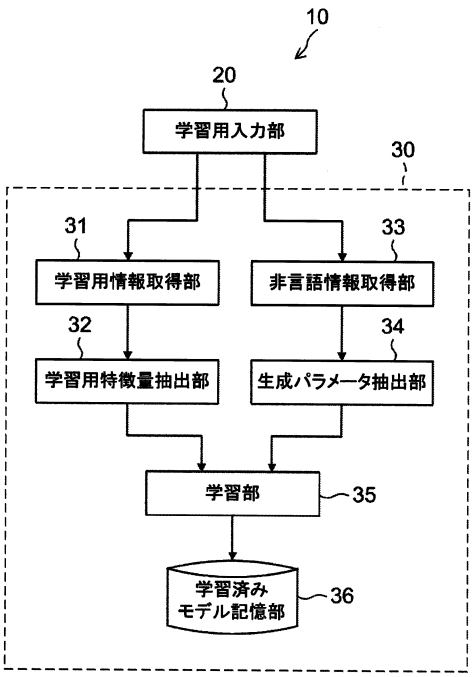
【０５９７】

１０，２１０，７１０，７１０Ａ，８１０ 非言語情報生成モデル学習装置
２０，２２０，７２０ 学習用入力部
３０，２３０，７３０，７３０Ａ，８３０ 学習用演算部
３１，２３１，３３１，４３１ 学習用情報取得部
３２，２３２，３３２，４３２，７３２ 学習用特徴量抽出部
３３ 非言語情報取得部
３４ 生成パラメータ抽出部
３５，２３５，７３５ 学習部
３６，６３，２３６，２６３，７３６，７６３ 学習済みモデル記憶部

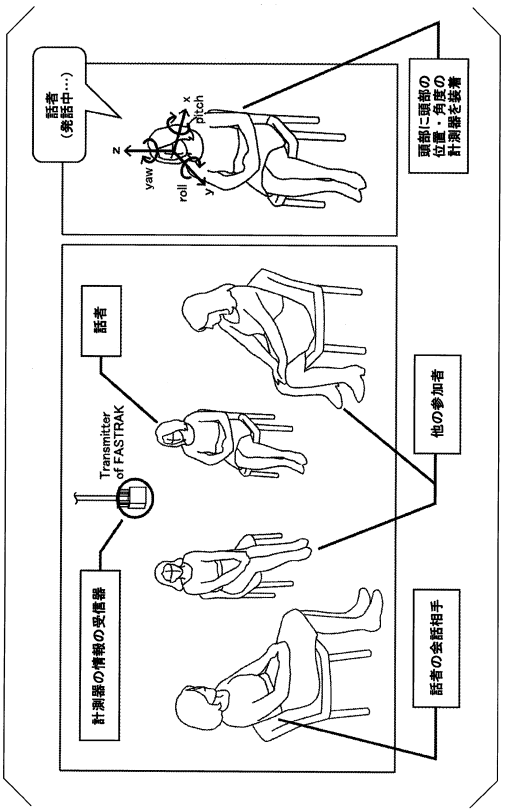
４０，２４０，７４０，７４０Ａ，８４０，１１４０ 非言語情報生成装置
５０，２５０，７５０ 入力部
６０，２６０，７６０，７６０Ａ，８６０，１１６０ 演算部
６１，２６１，３６１，４６１ 情報取得部
６２，２６２，３６２，４６２，７６２ 特徴量抽出部
６４，２６４，７６４ 非言語情報生成部
７０，１１９０Ｂ 発現部
２３７，３３８，４３７ 学習用テキスト解析部
２３８，４３８ 学習用音声合成部
２６５，３６６，４６５ テキスト解析部
２６６，４６６ 音声合成部
３３７ 学習用音声認識部
３６５ 音声認識部
７３１ 学習用付加情報取得部
７３１Ａ 学習用付加情報推定部
７６１ 付加情報取得部
７６１Ａ 付加情報推定部
８３３ 学習データ作成部
８７０，１１７０ 制御部
１１７２ 学習データ生成部
１１７４ 再学習制御部
１１９０ 表示部
１１９０Ａ 表示画面
１１９２ 出力部

【図面】

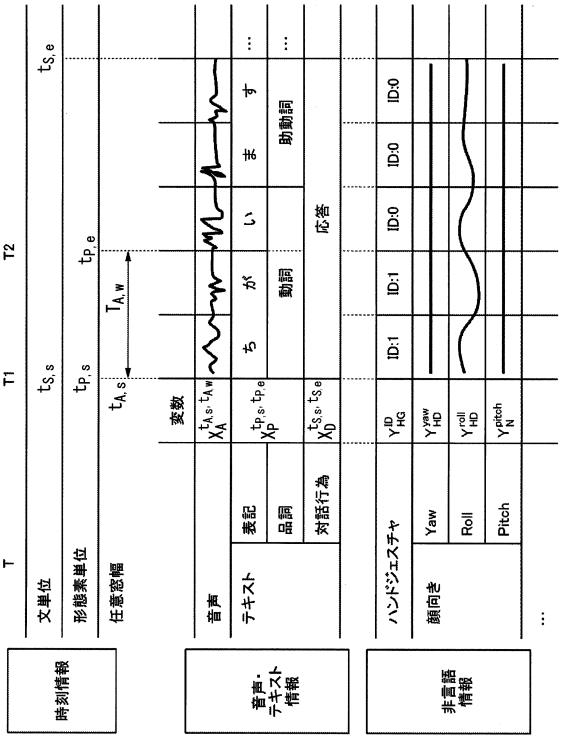
【図１】



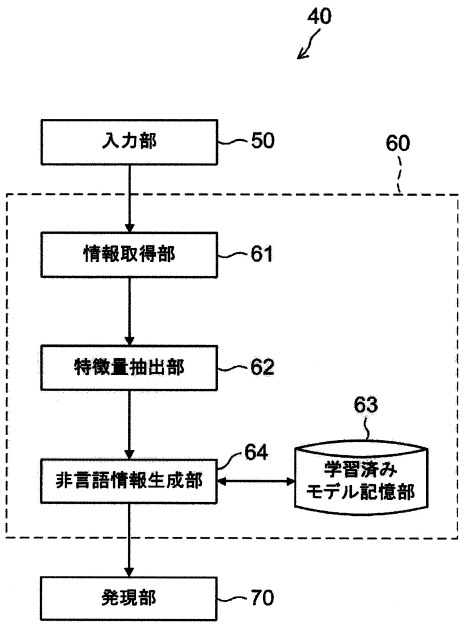
【図２】



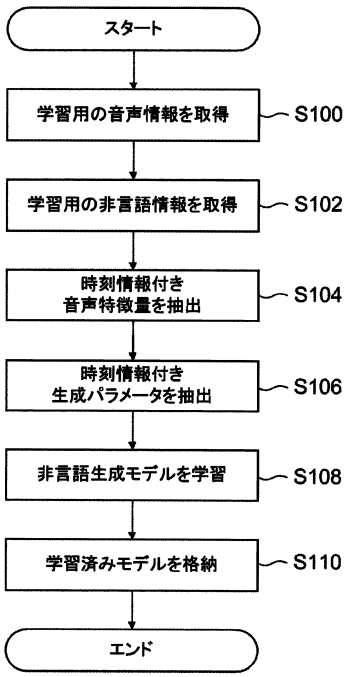
【図３】


T
T1
T2
文単位
形態素単位
任意窓幅
時刻情報
音声
テキスト
表記
品詞
対話行為
変数
ち
が
い
ま
す
動詞
助動詞
応答
音声・テキスト情報
ハンドジェスチャ
顔向き
Yaw
Roll
Pitch
ID:0
ID:1
非言語情報

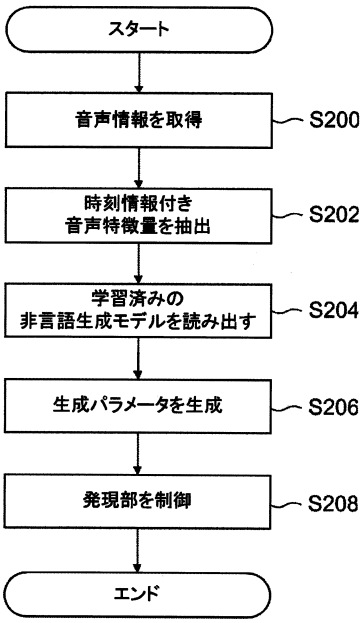
【図４】


40
入力部
50
60
情報取得部
61
特徴量抽出部
62
63
非言語情報生成部
64
学習済み
モデル記憶部
発現部
70

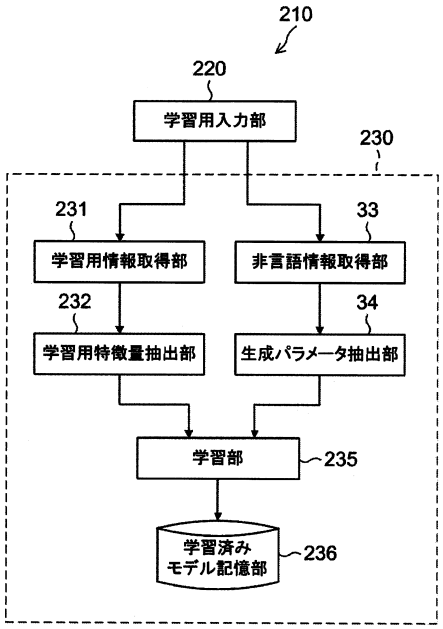
【図５】


スタート
学習用の音声情報を取得
S100
学習用の非言語情報を取得
S102
時刻情報付き
音声特徴量を抽出
S104
時刻情報付き
生成パラメータを抽出
S106
非言語生成モデルを学習
S108
学習済みモデルを格納
S110
エンド

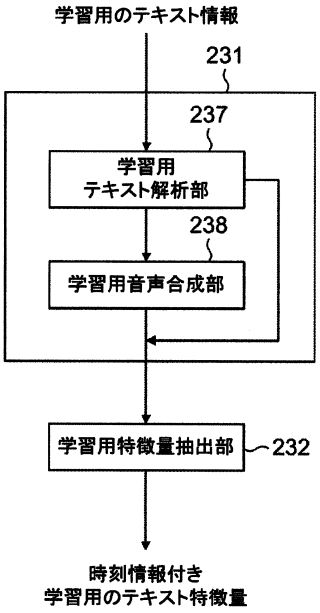
【図６】


スタート
音声情報を取得
S200
時刻情報付き
音声特徴量を抽出
S202
学習済みの
非言語生成モデルを読み出す
S204
生成パラメータを生成
S206
発現部を制御
S208
エンド

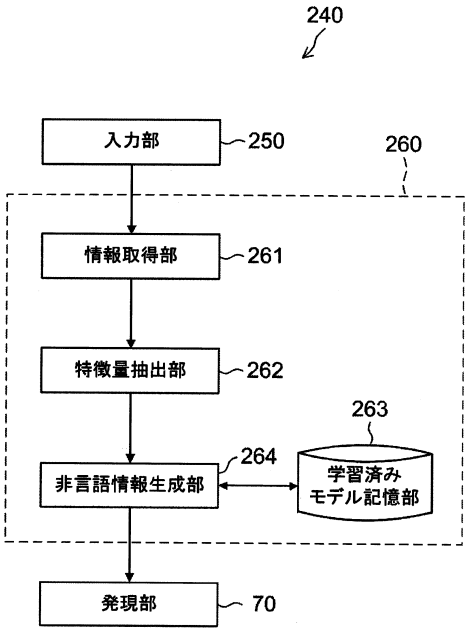
【図７】


210
220
学習用入力部
230
231
学習用情報取得部
33
非言語情報取得部
232
学習用特徴量抽出部
34
生成パラメータ抽出部
学習部
235
学習済み
モデル記憶部
236

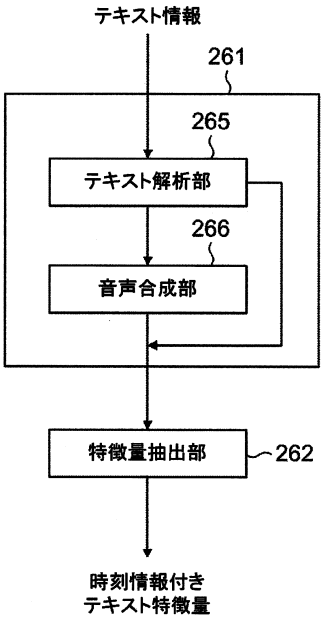
【図８】


学習用のテキスト情報
231
237
学習用
テキスト解析部
238
学習用音声合成部
学習用特徴量抽出部
232
時刻情報付き
学習用のテキスト特徴量

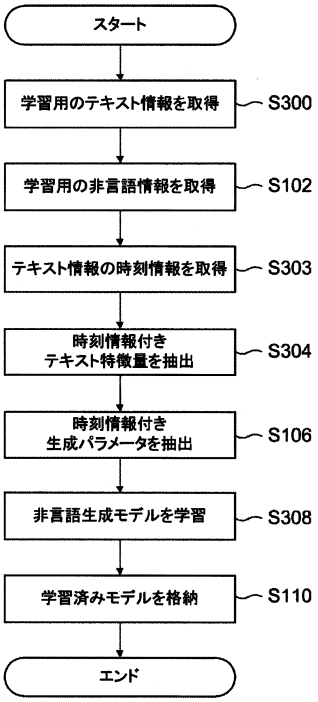
【図９】


240
入力部
250
260
情報取得部
261
特徴量抽出部
262
263
非言語情報生成部
264
学習済み
モデル記憶部
発現部
70

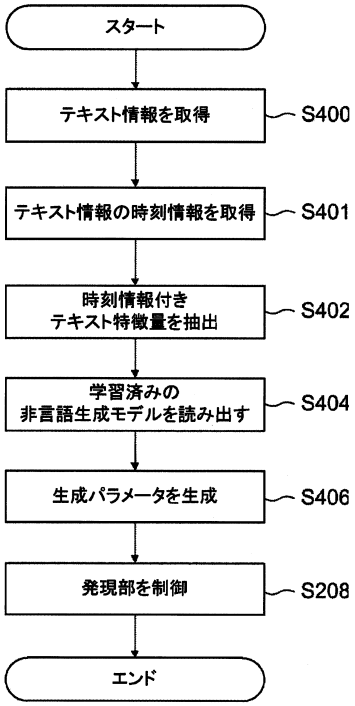
【図１０】


テキスト情報
261
265
テキスト解析部
266
音声合成部
特徴量抽出部
262
時刻情報付き
テキスト特徴量

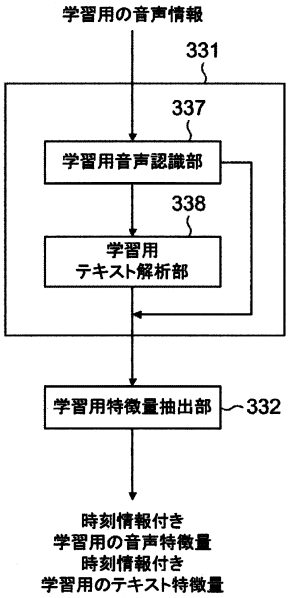
【図１１】



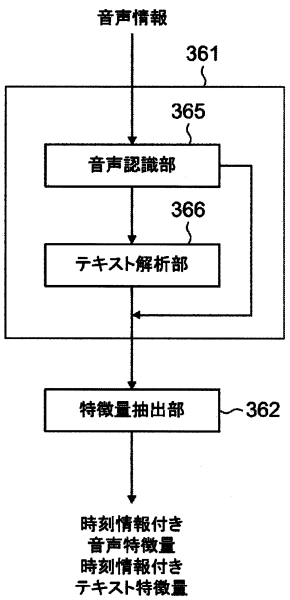
【図１２】



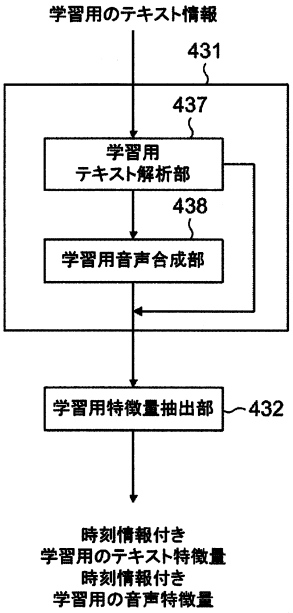
【図１３Ａ】



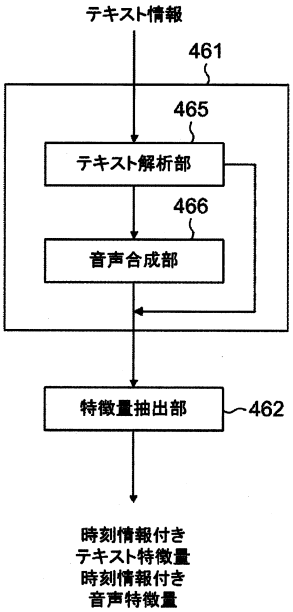
【図１３Ｂ】



【図１４Ａ】



【図１４Ｂ】



【図１５】

構成のバリエーションまとめ

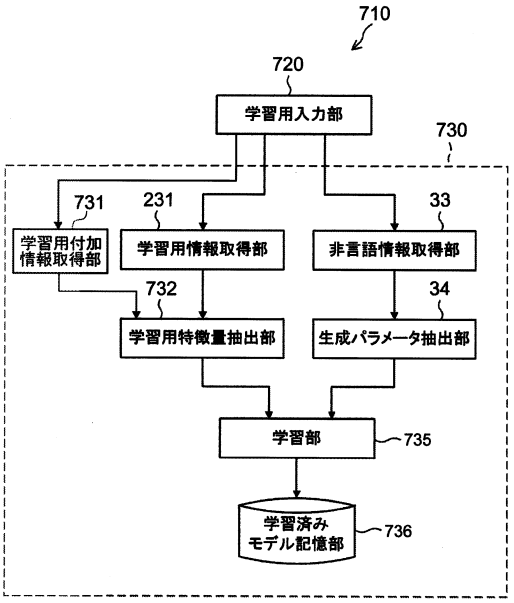
学習	学習時の入力	構成部		学習時の入力特徴量
		音声認識部	音声合成部	
パターン1	音声			音声
パターン2	テキスト		要	テキスト
パターン3	音声	要		音声、認識テキスト
パターン4	テキスト		要	合成音声、テキスト

【図１６】

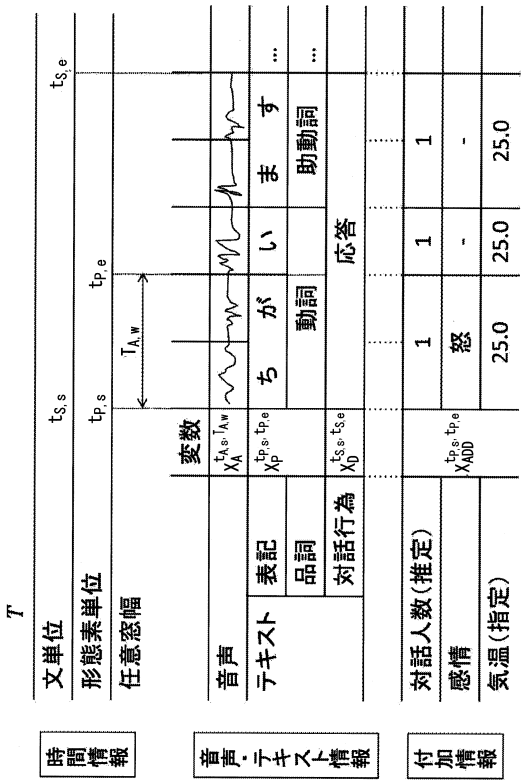
学習時と生成時の可能な構成の組み合わせバリエーション

学習時の構成パターン	生成時の構成パターン	学習時の入力特徴量	生成時の入力特徴量	学習／生成時の特徴量の一致
パターン1	パターン1	音声	音声	○
パターン1	パターン2	音声	合成音声	
パターン2	パターン1	テキスト	認識テキスト	
パターン2	パターン2	テキスト	テキスト	○
パターン3	パターン3	音声、認識テキスト	音声、認識テキスト	○
パターン3	パターン4	音声、認識テキスト	合成音声、テキスト	
パターン4	パターン3	合成音声、テキスト	音声、認識テキスト	
パターン4	パターン4	合成音声、テキスト	合成音声、テキスト	○

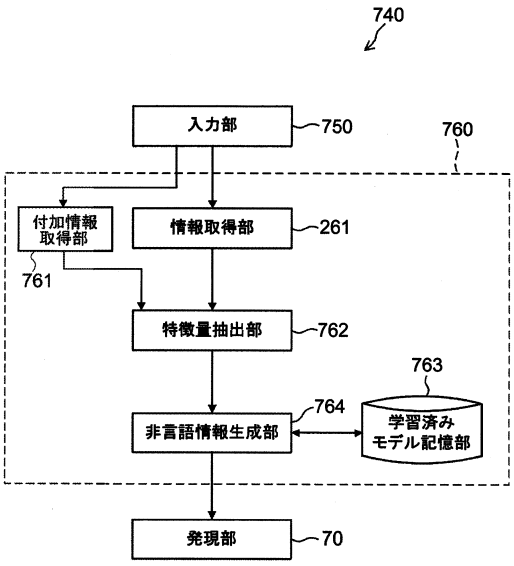
【図１７】



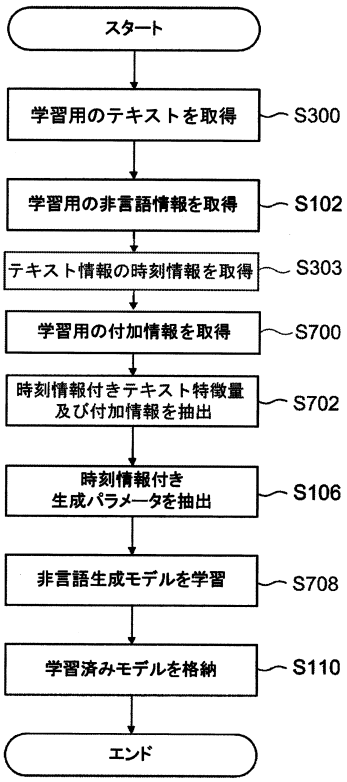
【図１８】



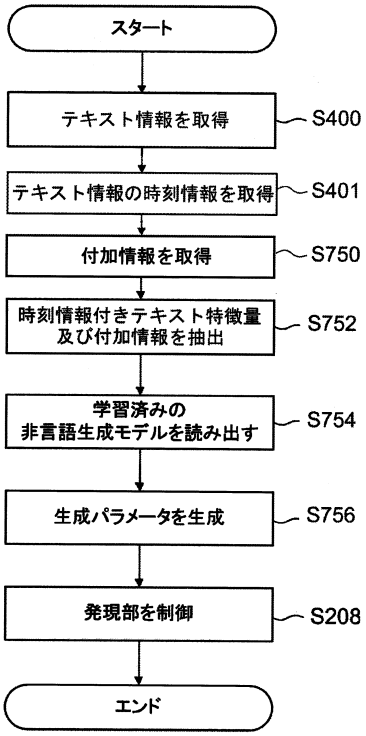
【図１９】



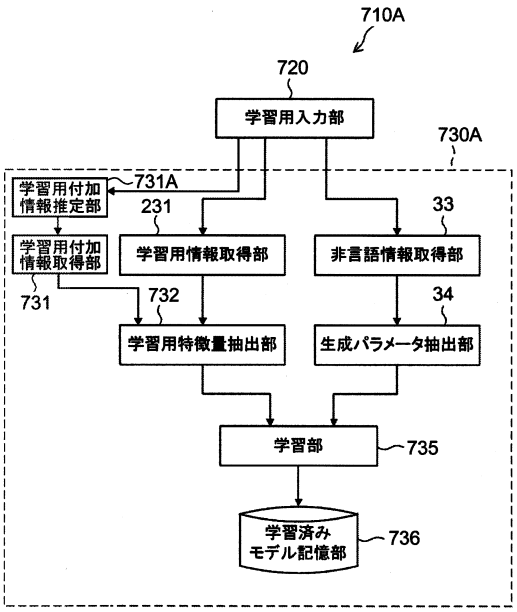
【図２０】



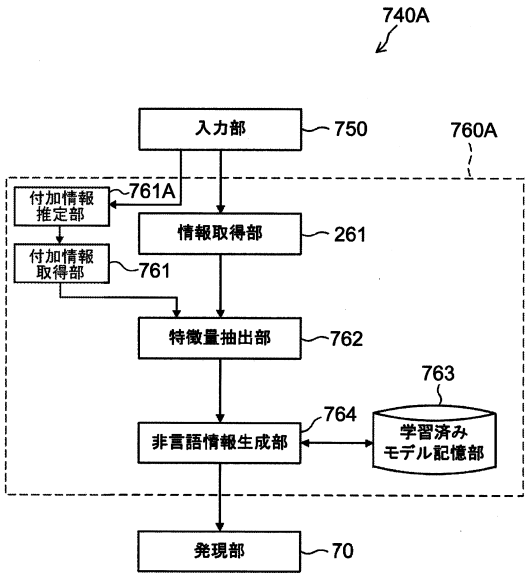
【図２１】



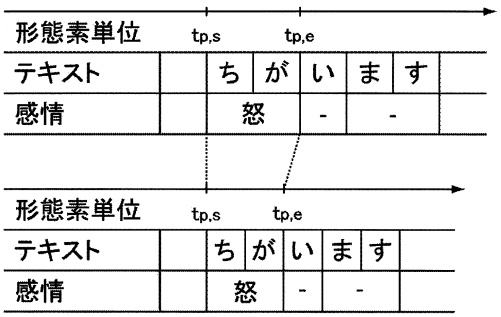
【図２２】



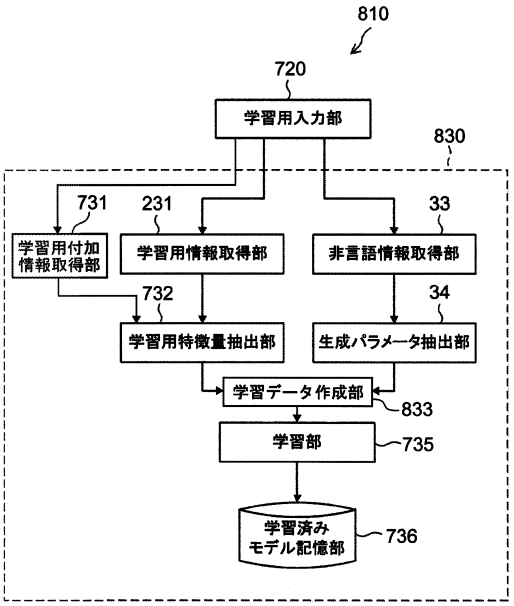
【図２３】



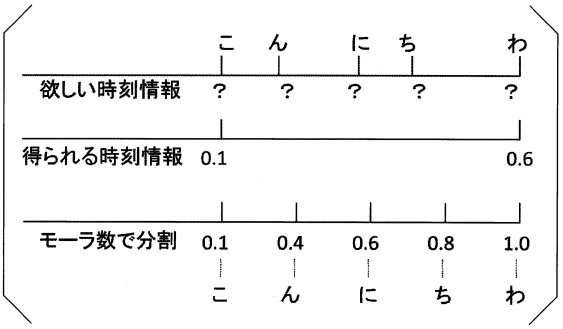
【図２４】



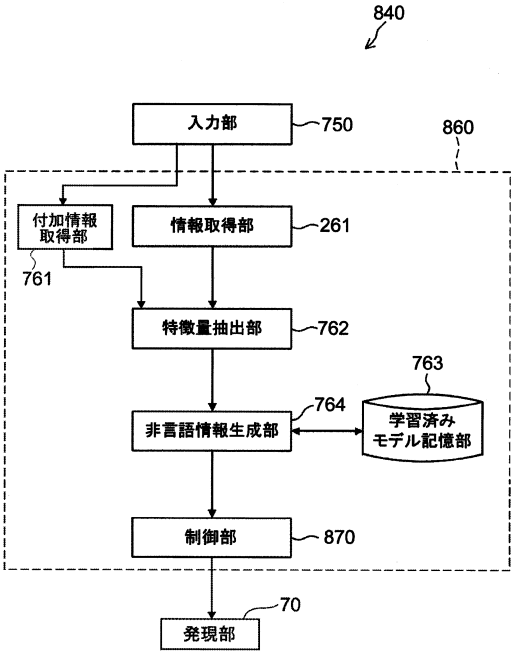
【図２５】



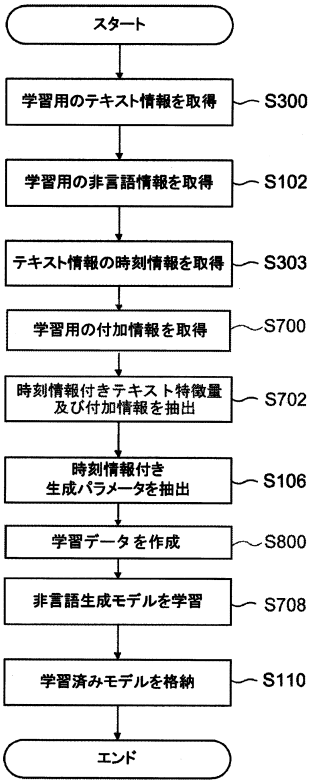
【図２６Ａ】



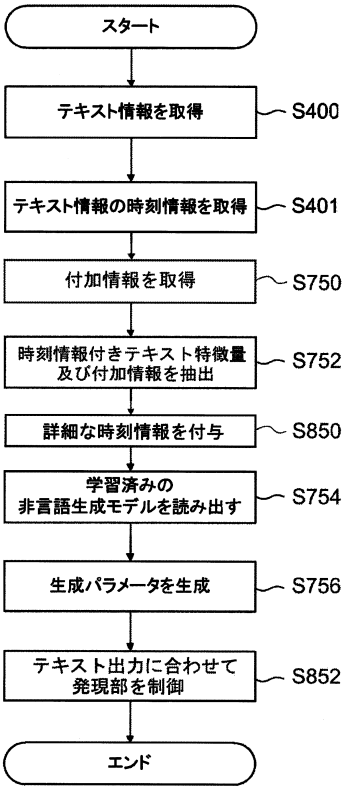
【図２６Ｂ】



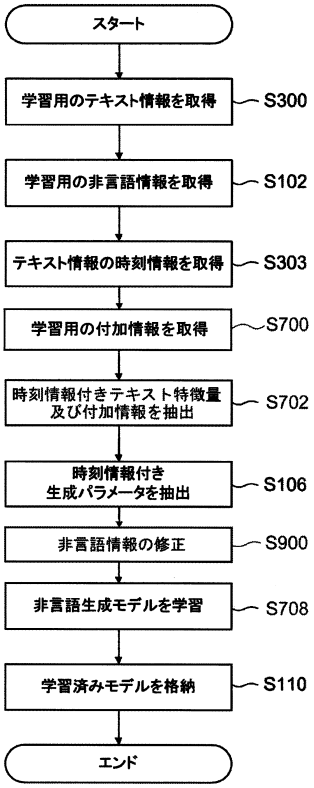
【図２７】



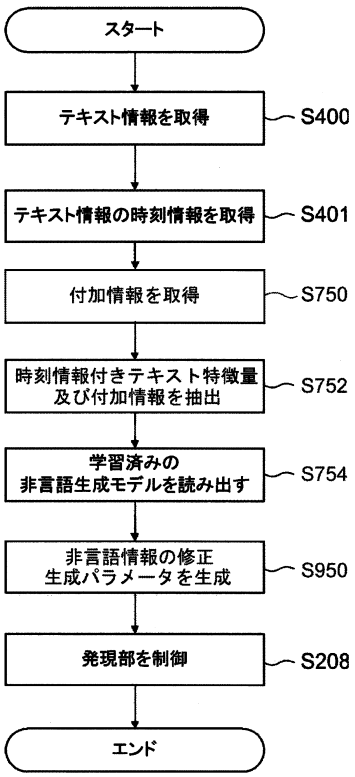
【図２８】



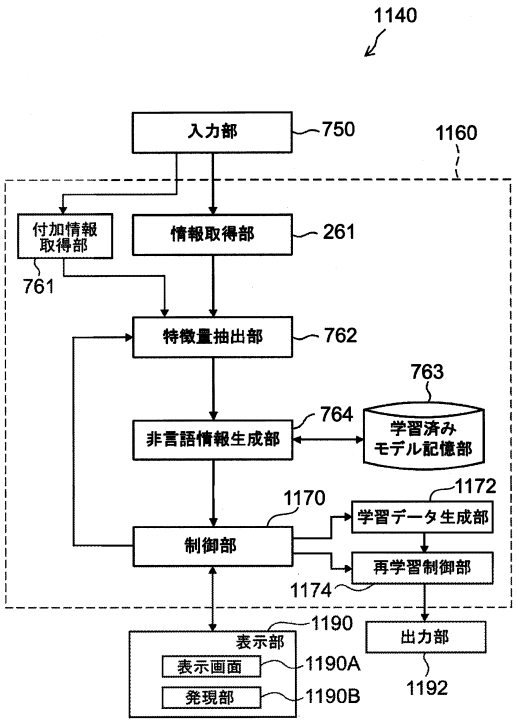
【図２９】



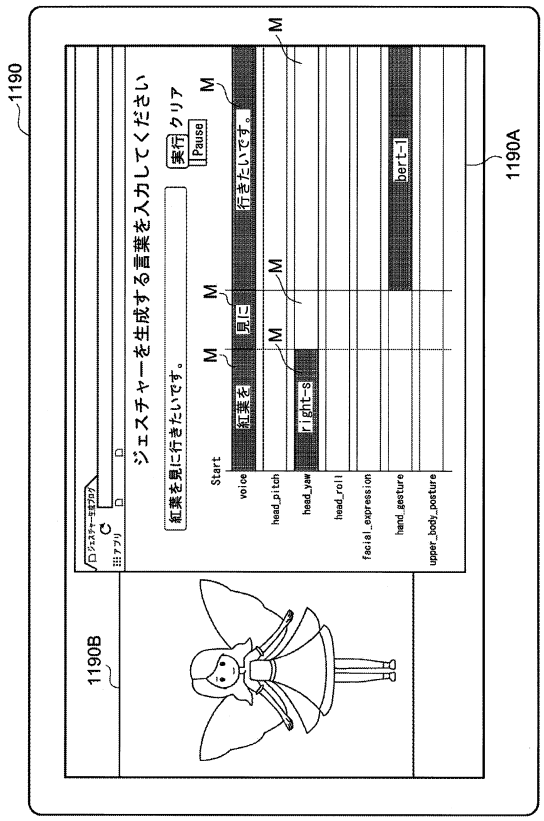
【図３０】



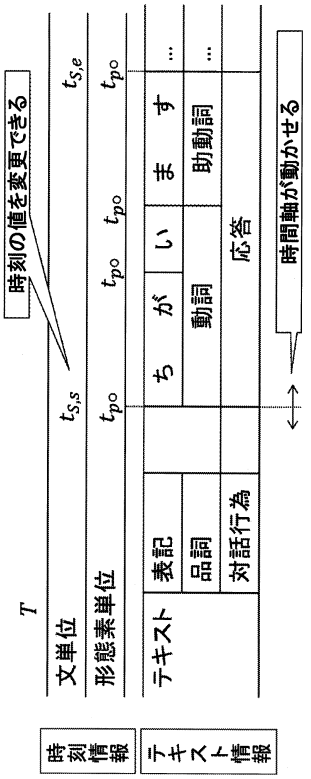
【図３１】



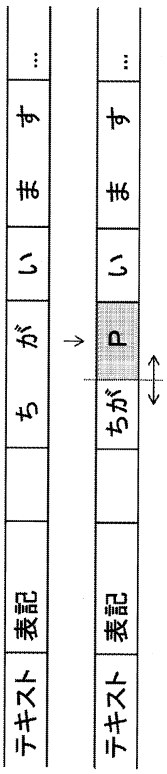
【図３２】



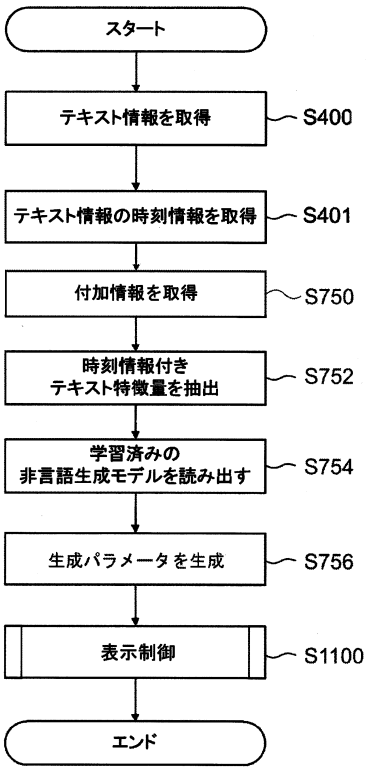
【図３３】



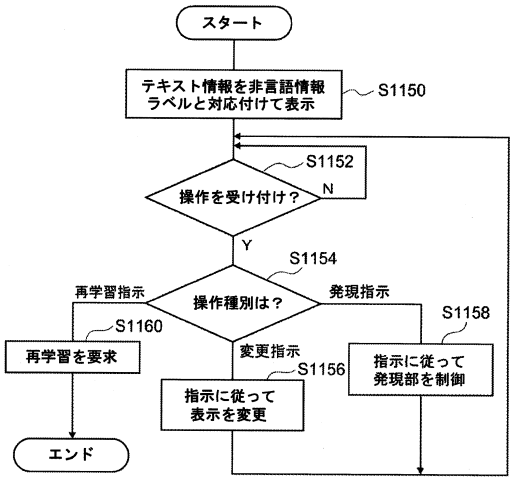
【図３４】



【図３５】



【図３６】


スタート
テキスト情報を非言語情報
ラベルと対応付けて表示
S1150
S1152
操作を受け付け？
N
Y
S1154
操作種別は？
再学習指示
発現指示
変更指示
S1160
再学習を要求
S1156
指示に従って
表示を変更
S1158
指示に従って
発現部を制御
エンド

フロントページの続き

(51)国際特許分類　　FI
G10L 25/24 (2013.01)　　G10L 25/24
G10L 25/90 (2013.01)　　G10L 25/90

(32)優先日　平成30年5月21日(2018.5.21)
(33)優先権主張国・地域又は機関
日本国(JP)
(31)優先権主張番号　特願2018-97339(P2018-97339)
(32)優先日　平成30年5月21日(2018.5.21)
(33)優先権主張国・地域又は機関
日本国(JP)
(31)優先権主張番号　特願2018-230310(P2018-230310)
(32)優先日　平成30年12月7日(2018.12.7)
(33)優先権主張国・地域又は機関
日本国(JP)
特許法第３０条第２項適用 平成３０年２月１９日～２３日開催 「ＮＴＴ Ｒ＆Ｄ フォーラム ２０１８」にて公開 平成３０年４月１９日 ｈｔｔｐ：／／ｗｗｗ．ｎｔｔ．ｃｏ．ｊｐ／ｎｅｗｓ２０１８／１８０４／１８０４１９ａ．ｈｔｍｌにて公開 平成３０年４月２８日～２９日開催 「ニコニコ超会議２０１８」にて公開 平成３０年５月７日～１２日開催 Ｅｌｅｖｅｎｔｈ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｌａｎｇｕａｇｅ Ｒｅｓｏｕｒｃｅｓ ａｎｄ Ｅｖａｌｕａｔｉｏｎ（ＬＲＥＣ ２０１８）にて公開 平成３０年５月２２日 ｈｔｔｐ：／／ｗｗｗ．ａｉ－ｇａｋｋａｉ．ｏｒ．ｊｐ／ｊｓａｉ２０１８／ ｈｔｔｐ：／／ｗｗｗ．ａｉ－ｇａｋｋａｉ．ｏｒ．ｊｐ／ｊｓａｉ２０１８／ｐｒｏｃｅｅｄｉｎｇｓにて公開 平成３０年５月３１日 ｈｔｔｐｓ：／／ｌｉｎｋ．ｓｐｒｉｎｇｅｒ．ｃｏｍ／ｃｈａｐｔｅｒ／１０．１００７／９７８－３－３１９－９１４８５－５＿２９にて公開 平成３０年６月２７日 一般社団法人情報処理学会 マルチメディア、分散、協調とモバイル（ＤＩＣＯＭＯ２０１８）シンポジウム 論文集 ｐｐ．１８６３－１８６８にて公開 平成３０年８月２７日～３１日開催 Ｔｈｅ ２７ｔｈ ＩＥＥＥ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｒｏｂｏｔ ａｎｄ Ｈｕｍａｎ Ｉｎｔｅｒａｃｔｉｖｅ Ｃｏｍｍｕｎｉｃａｔｉｏｎ（ＲＯ－ＭＡＮ２０１８）にて公開
前置審査

東京都千代田区大手町一丁目５番１号 日本電信電話株式会社内
(72)発明者　小林 *のぞみ*
東京都千代田区大手町一丁目５番１号 日本電信電話株式会社内
(72)発明者　西田 京介
東京都千代田区大手町一丁目５番１号 日本電信電話株式会社内

審査官　菅原 浩二

(56)参考文献
特開平１０－１４３３５１（ＪＰ，Ａ）
特開２０００－２００１２５（ＪＰ，Ａ）
特開２００２－３３１４８１（ＪＰ，Ａ）
特開２００３－１７３４５２（ＪＰ，Ａ）
特開２００４－０６４６７９（ＪＰ，Ａ）
特開２００５－１６５４３８（ＪＰ，Ａ）
特開２００５－３０９１７３（ＪＰ，Ａ）
特開２００７－０２７９９０（ＪＰ，Ａ）
特開２００７－０３４７８８（ＪＰ，Ａ）
国際公開第２０１７／０７２９１５（ＷＯ，Ａ１）
米国特許第０９８１２１５１（ＵＳ，Ｂ１）
CASSELL 他， BEAT: the Behavior Expression Animation Toolkit，Proceeding of the 28th annual conference on Computer graphics and interactive techniques，ACM，2001年，pp. 477-486，〈DOI: 10.1145/383259.383315〉
CHIU 他，Predicting Co-verbal Gesture: A Deep and Temporal Modeling Approach，Intelli

gent Virtual Agent. IVA2015. Lecture Notes in Computer Science，Springer，vol. 9238，pp.152-166，〈DOI: 10.1007/978-3-319-21996-7_17〉
BREITFUSS 他，Automatic Generation of Non-verbal Behavior for Agents in Virtual Worlds: A System for Supporting Multimodal Conversations of Bots and Avatars，Online Communities and Social Computing. OCSC 2009.Lecture Notes in Computer Science，Springer，2009年，vol. 5621，pp. 153-161，〈DOI: 10.1007/978-3-642-02774-1_17〉

(58)調査した分野 (Int.Cl.，ＤＢ名)

Ｇ０６Ｆ　３／０１
Ｇ０６Ｆ　３／１６
Ａ６３Ｈ　１１／００
Ｇ０６Ｔ　１３／４０
Ｇ１０Ｌ　２５／２１
Ｇ１０Ｌ　２５／２４
Ｇ１０Ｌ　２５／９０