

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-164352
(P2019-164352A)

(43) 公開日 令和1年9月26日(2019.9.26)

(51) Int.Cl.	F I	テーマコード (参考)
G10L 15/22 (2006.01)	G10L 15/22 300U	3C707
G06F 3/01 (2006.01)	G06F 3/01 510	5E555
G10L 15/10 (2006.01)	G10L 15/10 500T	
G10L 13/00 (2006.01)	G10L 15/10 500N	
G10L 15/04 (2013.01)	G10L 13/00 100M	

審査請求 有 請求項の数 25 O L (全 23 頁) 最終頁に続く

(21) 出願番号 特願2019-57767 (P2019-57767)
 (22) 出願日 平成31年3月26日 (2019. 3. 26)
 (62) 分割の表示 特願2016-562886 (P2016-562886) の分割
 原出願日 平成27年4月17日 (2015. 4. 17)
 (31) 優先権主張番号 14305583.8
 (32) 優先日 平成26年4月17日 (2014. 4. 17)
 (33) 優先権主張国・地域又は機関 欧州特許庁 (EP)

(71) 出願人 311003754
 ソフトバンク・ロボティクス・ヨーロッパ
 SOFTBANK ROBOTICS EUROPE
 フランス国、75015・パリ、リュ・デュ・コロネル・ピエール・アビア・43
 (74) 代理人 110001173
 特許業務法人川口国際特許事務所
 (72) 発明者 ジェローム・モンソー
 フランス国、75013・パリ、リュ・ブッサニョー、9
 (72) 発明者 グェナエル・ガト
 フランス国、75013・パリ、リュ・デュ・ジャブロ、75

最終頁に続く

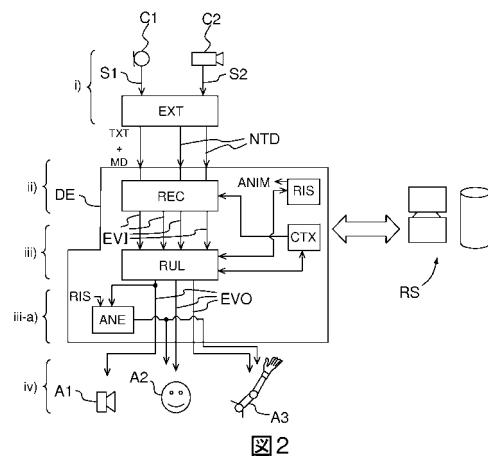
(54) 【発明の名称】 人間型ロボットとユーザーの間におけるマルチモード会話を実行する方法、前記方法を実装するコンピュータプログラム及び人間型ロボット

(57) 【要約】 (修正有)

【課題】 人間型ロボットとユーザーの間におけるマルチモード会話を実行する方法、その方法を実装するコンピュータプログラム及び人間型ロボットを提供する。

【解決手段】 マルチモード会話を実行する方法は、少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された、ユーザーによって生成された複数のイベント (EVI) を認識するべく取得された信号を解釈するステップと、少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された少なくとも1つのイベント (EVO) を有する人間型ロボットの応答を判定するステップと、を有する。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

人間型ロボット（R）と少なくとも一人のユーザー（U）の間における会話を実行する方法であって、前記人間型ロボットによって反復的に実行される、

i) 個々のセンサ（c1、c2）から複数の入力信号（s1、s2）を取得するステップであって、少なくとも1つの前記センサは、サウンドセンサであり、且つ、少なくとも1つのその他のセンサは、モーション又は画像センサである、ステップと、

ii) 少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された、前記ユーザーによって生成された複数のイベント（EVI）を認識するべく前記取得された信号を解釈するステップと、

iii) 少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された少なくとも1つのイベント（EVO）を有する前記人間型ロボットの応答を判定するステップであって、前記判定するステップは、規則の組を適用することによって実行され、それぞれの前記規則は、入力イベントの組を前記ロボットの応答に対して関連付けている、ステップと、

iv) 前記人間型ロボットにより、前記イベント又はそれぞれの前記イベントを生成するステップと、

を有し、

前記方法は、前記ステップiii)において適用された前記規則の少なくともいくつかは、応答を前記ユーザーによって一緒に生成されると共に前記ステップii)において認識される少なくとも2つのイベントの組合せに関連付けており、これらのイベントのうちの少なくとも1つは、前記ユーザーによって発声された単語又は文ではないことを特徴としている、方法。

【請求項 2】

前記ステップiii)において適用される前記規則のうちの少なくともいくつかは、前記人間型ロボットによって一緒に生成される少なくとも2つのイベントを有する応答を決定し、これらのイベントの少なくとも1つは、前記単語又は文の発声ではない請求項1に記載の方法。

【請求項 3】

前記ステップiii)において、前記人間型ロボットの応答は、会話のコンテキスト（CTX）、前記ユーザーの身元情報、前記人間型ロボットの内部状態（RIS）から選択された少なくとも1つのパラメータに基づいて判定される請求項1又は2に記載の方法。

【請求項 4】

前記ステップii)において認識された又は前記ステップiii)において判定された少なくとも1つのイベントに従って前記パラメータ又は少なくとも1つの前記パラメータの値を変更するステップを更に有する請求項3に記載の方法。

【請求項 5】

前記ステップii)は、取得された信号と前記人間型ロボットの、又はこれによってアクセス可能である、メモリ内に保存されている予想イベントのリストに属するイベントの間におけるマッチングをサーチするステップを有し、前記サーチするステップは、既定の値を上回るコンフィデンススコアを有するイベントが認識される時点まで、或いは、最高の複雑性を有する認識方法の使用が完了する時点まで、増大する複雑性を有する複数のマッチング法（MM1～MM4）を連続的に使用することによって実行される請求項1～4のいずれか1項に記載の方法。

【請求項 6】

前記使用されるマッチング方法は、会話のコンテキストに応じて選択される請求項5に記載の方法。

【請求項 7】

前記マッチング方法は、増大する複雑性の順序において、正確なマッチングについてのサーチ、概略的なマッチングについてのサーチ、音声認識のケースのみにおける音声学的

10

20

30

40

50

対応性についてのサーチ、及び意味論的な対応性についてのサーチを含む請求項 5 又は 6 に記載の方法。

【請求項 8】

音声学的対応性についてサーチする前記方法は、

- サウンドセンサによって取得されたサウンドの組の音声学的文字化のステップと、
 - 前記結果的に得られた音声学的文字化を単純化及びスムージングするステップと、
 - 前記単純化及びスムージングされた音声学的文字化と自然言語における予め定義された単語の組を単純化及びスムージングすることによって得られた複数のエントリの間における編集距離を算出するステップと、
 - 前記単純化及びスムージングされた音声学的文字化から最小の編集距離を有する前記エントリに対応した予め定義された組の自然言語単語を選択するステップと、
- を有する請求項 7 に記載の方法。

10

【請求項 9】

前記単純化及びスムージングするステップは、

- 混同されやすい音素を単一の音素によって置換するステップと、
- 単語の開始点に位置した母音及び鼻母音以外の母音を除去するステップと、
- 単語の間の休止部を除去するステップと、

を有する請求項 8 に記載の方法。

【請求項 10】

前記予想イベントのリストは、会話のコンテキストに応じて、複数の前記リストのうちから選択される請求項 5 ~ 9 のいずれか 1 項に記載の方法。

20

【請求項 11】

前記ステップ *i i i*) は、プロアクティブな規則と呼称される予め定義されたサブセット (PRO) に属する規則を適用することにより、前記ユーザーによって発声された単語又は識別されたジェスチャの欠如を含むイベントの組に対する応答を判定するステップを有する請求項 1 ~ 10 のいずれか 1 項に記載の方法。

【請求項 12】

ステップ *i i i*) において判定された前記応答が、少なくとも前記単語又は文の発声であるか、又はこれを有する場合に、発声対象の前記単語又は文の語学的分析と、前記分析の関数としての前記応答に伴う活性化の判定と、を実行するステップ *i i i - a*) の実行を更に有する請求項 1 ~ 11 のいずれか 1 項に記載の方法。

30

【請求項 13】

前記ステップ *i i i - a* は、

-) 活性化対象である前記応答の少なくとも 1 つの単語を識別するサブステップと、
-) コンセプトと、活性化対象である前記単語又はそれぞれの前記単語と関連付けられたワンオフ表現力と呼称される表現力と、を判定するサブステップと、
-) 前記人間型ロボットの、又はこれによってアクセス可能である、メモリ内に保存されている活性化のリスト (ALST) から、前記コンセプト及び前記ワンオフ表現力に基づいて、活性化を選択するサブステップと、

を有する請求項 12 に記載の方法。

40

【請求項 14】

前記サブステップ は、前記文の構造内におけるその機能に応じて、活性化対象であるそれぞれの又は前記単語を判定するべく、発声対象である文の構文分析を実行するステップを有する請求項 13 に記載の方法。

【請求項 15】

前記サブステップ において、前記ワンオフ表現力は、前記単語の表現力、前記単語に関係付けられた 1 つ又な複数のその他の単語の表現力、及び前記全体的な応答の全体的な表現力から選択された少なくとも 1 つのパラメータに基づいて判定される請求項 13 又は 14 に記載の方法。

【請求項 16】

50

前記リストのそれぞれの活性化は、1つ又は複数のコンセプトと関連付けられ、且つ、特定の表現力を有しており、前記サブステップは、前記リストのうちにおいて、前記サブステップにおいて判定された前記コンセプトと関連付けられると共に前記ワンオフ表現力に最も近接した特定の表現力を有する活性化を選択するステップを含む請求項13～15のいずれか1項に記載の方法。

【請求項17】

前記特定の表現力及び前記ワンオフ表現力に基づいて、最終的な表現力と呼称される表現力を判定するサブステップを更に有する請求項16に記載の方法。

【請求項18】

前記ワンオフ表現力又は前記最終的な表現力は、前記活性化の少なくとも1つのジェスチャの速度及び振幅のうちから選択された少なくとも1つのパラメータを決定する請求項13～17のいずれか1項に記載の方法。

10

【請求項19】

前記ステップi)～iv)と同時に前記ロボットによって反復的に実装される、
A)前記ロボット(R)に設定された基準フレームとの関係において前記ユーザー(U)の身体の少なくとも一部分の位置を判定するステップと、
B)予め定義された値の範囲において前記ロボット又はその1つの要素と前記ユーザーの前記少なくとも1つの又は前記身体部分の間の距離を維持するべく、前記ロボットの少なくとも1つのアクチュエータを駆動するステップと、
を更に有する請求項1～18のいずれか1項に記載の方法。

20

【請求項20】

前記ステップB)は、既定の角度範囲内において前記ユーザーとの関係における前記ロボットの向きを維持するべく、前記ロボットの少なくとも1つのアクチュエータを駆動するステップを更に有する請求項19に記載の方法。

【請求項21】

C)前記既定の値の範囲内において前記距離を維持すると共に、適宜、前記既定の角度範囲内において前記向きを維持しつつ、前記ロボットの前記疑似ランダムな変位を生成するように前記アクチュエータ又は少なくとも1つの前記アクチュエータを駆動するステップを更に有する請求項19又は20に記載の方法。

30

【請求項22】

D)前記ユーザーと前記人間型ロボットの間における継続中の会話の意味論的分析を実行し、且つ、前記分析に従って、前記既定の距離値の範囲を変更すると共に、適宜、前記既定の角度範囲を変更するステップを更に有する請求項19～21のいずれか1項に記載の方法。

【請求項23】

前記ステップA)は、前記ロボットに設定された前記基準フレームとの関係において前記ユーザーの下半身の位置を判定するステップを有する請求項19～22のいずれか1項に記載の方法。

【請求項24】

プログラムコード命令を有するコンピュータプログラムプロダクトであって、前記プログラムコード命令は、前記プログラムが、人間型ロボット(R)上において組み込まれた少なくとも1つのプロセッサによって実行された際に、請求項1～23のいずれか1項に記載の方法を実行し、前記ロボットは、個々の入力信号を取得するべく、前記プロセッサ又は少なくとも1つのプロセッサに対して動作可能に接続されると共に少なくとも1つのサウンドセンサ及び少なくとも1つの画像又は運動センサを有する複数のセンサ(c1、c2)と、単語又は文を発声するべく、前記プロセッサ又は少なくとも1つの前記プロセッサによって制御される発話合成モジュールと、前記ロボットが複数の運動又はジェスチャを実行できるようにする前記プロセッサ又は少なくとも1つの前記プロセッサによって駆動されるアクチュエータの組(A1、A2、A3)と、を有する、コンピュータプログラムプロダクト。

40

50

【請求項 25】

人間型ロボット（R）であって、

- 少なくとも1つの組込み型のプロセッサと、
 - 個々の入力信号を取得するべく、前記プロセッサ又は少なくとも1つの前記プロセッサに対して動作可能に接続されると共に少なくとも1つのサウンドセンサ及び少なくとも1つの画像又は運動センサを有するセンサ組立体（c1、c2）と、
 - 単語又は文を発声するべく、前記プロセッサ又は少なくとも1つの前記プロセッサによって駆動される発話合成モジュールと、
 - 前記ロボットが複数の運動又はジェスチャを実行できるようにする前記プロセッサ又は少なくとも1つの前記プロセッサによって駆動されるアクチュエータの組（A1、A2、A3）と、
- を有し、

前記人間型ロボットは、前記プロセッサ又は少なくとも1つの前記プロセッサが、請求項1～23のいずれか1項に記載の方法を実行するようにプログラム又は構成されていることを特徴としている、人間型ロボット。

【請求項 26】

少なくとも1つのリモートサーバーに接続するための装置を更に有し、前記プロセッサ又は少なくとも1つの前記プロセッサは、請求項1～23のいずれか1項に記載の方法を実行するべく、前記リモートサーバー又は少なくとも1つの前記リモートサーバーと協働するようにプログラム又は構成されている請求項25に記載の人間型ロボット。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、人間型ロボットとユーザー、即ち、通常は人間である、対話者の間における所謂「マルチモード」会話を実行する方法に関する。又、本発明は、このような方法を実装するためのコンピュータプログラムプロダクト及び人間型ロボットにも関する。

【背景技術】**【0002】**

「人間型ロボット」は、胴体、頭部、腕部、脚部、音声認識及び音声合成などを使用して人間と口頭で通信する能力などの人間の外観及び機能の特定の属性を有するロボットとして定義することができる。この種のロボットは、人間と機械の間における認知距離の低減を目的としている。人間型ロボットの最も重要な特性の1つは、人間対話者との間において可能な限り自然な会話をサポートするその能力にある。この能力は、日々の暮らしにおいて必要とされる際に、高齢の、病気の、或いは、単純に孤独である人々を支援するための、且つ、これらの人々に人間の個人的な支援の存在の感情的な観点においても受け入れ可能な代用物を提供するための、「コンパニオンロボット」の開発に必須である。これを実現するには、人間の振る舞いを可能な限り密接にエミュレートする方式によって人間とやり取りするこのような人間型ロボットの能力の開発が必須である。具体的には、ロボットは、人間のものに対応した豊富な表現と、通常は人間のものであるタイプの振る舞い及び感情と相乗効果を有する表現のモードと、を伴って、人間の質問又は主張を解釈できると共に会話モードにおいて複製を生成することができる必要がある。

【0003】

この方向における第一歩は、本出願人によって市販されると共に、ロボットプレーヤーに関する（特許文献1）及び自然なインターフェイスの会話を伴う人間型ロボットに関する（特許文献2）において開示されているNaο（商標）人間型ロボットをプログラムする方法に起因し、実現されている。

【0004】

但し、これらの文献によって開示されているロボットは、限られると共に既定された会話の要素しか実行することができない。

【0005】

10

20

30

40

50

(特許文献3)は、既存のロボットのものよりも豊富な会話能力を示す、会話エージェント、音声認識ツール、及び対話者の振る舞いを分析するツールを有する人間型ロボットについて記述している。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】国際公開第2012/000927号

【特許文献2】国際公開第2012/010451号

【特許文献3】国際公開第2013/150076号

【発明の概要】

【発明が解決しようとする課題】

【0007】

本発明は、このような人間型ロボットを改善し、これにより、人間対話者との間におけるやり取りを相対的に豊富に且つリアルにすることを目的としている。本発明は、具体的には、ユーザーの動作を解釈するための能力をロボットに提供することにより、人間-ロボット間のやり取りを改善することを目的とした「Juliette」と呼称されるプロジェクトを含む。

【課題を解決するための手段】

【0008】

このような目標の実現を許容する本発明の目的は、請求項1に記載の人間型ロボットと少なくとも1つのユーザーの間における会話を実行する方法であり、この方法は、前記人間型ロボットによって反復的に実行される、

i) 個々のセンサから複数の入力信号を取得するステップであって、少なくとも1つの前記センサは、サウンドセンサであり、且つ、少なくとも1つのその他のセンサは、モーション又は画像センサである、ステップと、

ii) 少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された、前記ユーザーによって生成される複数のイベントを認識するべく、取得された信号を解釈するステップと、

iii) 少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された少なくとも1つのイベントを有する前記人間型ロボットの応答を判定するステップであって、前記判定するステップは、規則の組を適用することにより、実行され、それぞれの前記規則は、入力イベントの組をロボットの応答に関連付けている、ステップと、

iv) 前記人間型ロボットにより、前記イベント又はそれぞれの前記イベントを生成するステップと、

を有し、

この方法は、前記ステップiii)において適用される前記規則の少なくともいくつかは、応答を前記ユーザーによって一緒に生成されると共に前記ステップii)において認識される少なくとも2つのイベントの組合せに関連付けており、これらのイベントの少なくとも1つは、前記ユーザーによって発声された単語又は文ではない、ことを特徴としている。

【0009】

このような方法の特定の実施形態は、従属請求項の主題を構成している。

【0010】

本発明の別の目的は、プログラムコード命令を有するコンピュータプログラムプロダクトであり、プログラムコード命令は、前記プログラムが、人間型ロボット上において組み込まれた少なくとも1つのプロセッサによって実行された際に、このような方法を実行し、前記ロボットは、個々の入力信号を取得するべく、前記プロセッサ又は少なくとも1つのプロセッサに対して動作可能に接続されると共に少なくとも1つのサウンドセンサ及び少なくとも1つの画像又は運動センサを有する複数のセンサと、単語又は文を発声するべ

10

20

30

40

50

く、前記プロセッサ又は少なくとも1つの前記プロセッサによって制御される発話合成モジュールと、前記ロボットが複数の運動又はジェスチャを実行できるようにする前記プロセッサ又は少なくとも1つの前記プロセッサによって駆動されるアクチュエータの組と、を有する。

【0011】

本発明の更に別の目的は、

- 少なくとも1つの組込み型のプロセッサと、
 - 個々の入力信号を取得するべく、前記プロセッサ又は少なくとも1つの前記プロセッサに対して動作可能に接続されると共に、少なくとも1つのサウンドセンサと、少なくとも1つの画像又は運動センサと、を有するセンサ組立体と、
 - 単語又は文を発声するべく、前記プロセッサ又は少なくとも1つの前記プロセッサによって駆動される発話合成モジュールと、
 - 前記ロボットが複数の運動又はジェスチャを実行できるようにする前記プロセッサ又は少なくとも1つの前記プロセッサによって駆動されるアクチュエータの組と、
- を有する人間型ロボットであり、

このロボットは、前記プロセッサ又は少なくとも1つの前記プロセッサが、本発明の一実施形態による方法を実行するようにプログラム又は構成されているを特徴としている。

【0012】

このような人間型ロボットは、少なくとも1つのリモートサーバーに接続するための装置を更に有してもよく、前記プロセッサ又は少なくとも1つの前記プロセッサは、本発明の一実施形態による方法を実行するべく、前記リモートサーバー又は少なくとも1つの前記リモートサーバーと協働するようにプログラム又は構成されている。

【0013】

本発明のその他の特徴、詳細、及び利点については、一例として付与される添付図面との関連において以下の説明を参照することにより、明らかとなる。

【図面の簡単な説明】

【0014】

【図1】本発明を実装するのに適した人間型ロボットの物理的アーキテクチャを示す。

【図2】本発明の一実施形態による方法のステップと、その実装のためのハードウェア及びソフトウェア手段の構成と、を示す図である。

【図3】本発明の一実施形態による「プロアクティブ」な会話の実装形態を示す図である。

【図4】本発明の一実施形態による人間型ロボットの応答を活性化するステップを示す図である。

【図5a】活性化対象である1つ又は複数の単語を判定するための文の構文分析の3つの例である。

【図5b】活性化対象である1つ又は複数の単語を判定するための文の構文分析の3つの例である。

【図5c】活性化対象である1つ又は複数の単語を判定するための文の構文分析の3つの例である。

【図6】本発明の一実施形態によるユーザーとの関係におけるロボットの位置のサーボ制御を示す。

【図7】本発明の一実施形態によるイベントを識別するステップを示す図である。

【図8】本発明の一実施形態による音声学的発話認識のステップを示す図である。

【発明を実施するための形態】

【0015】

図1は、本発明のいくつかの実施形態における人間型ロボットの物理的アーキテクチャを示している。

【0016】

この図の特定のロボットRは、本発明を実装しうる人間型ロボットの一例に過ぎないも

10

20

30

40

50

のと解釈されたい。この図のロボットの下肢は、歩行の機能を有していないが、自身が位置している表面上において転動するその基部 R B 上において任意の方向に運動することができる。本発明は、歩行に適合したロボットにおいて容易に実装することができる。例として、このロボットは、約 120 cm であってもよい高さ H と、約 65 cm の奥行 D と、約 40 cm の幅 W と、を有する。特定の実施形態においては、本発明のロボットは、メッセージ（オーディオ、ビデオ、ウェブページ）をその環境に伝達することができるか又はタブレットの触覚インターフェイスを通じてユーザーからエントリを受け取ることができるタブレット R T を有する。又、タブレットのプロセッサに加えて、本発明のロボットは、その独自のマザーボードのプロセッサをも使用しており、これは、例えば、Intel（商標）の ATOM（商標）Z530 であってもよい。又、本発明のロボットは、有利には、特定の実施形態において、マザーボードと、特に、磁気回転式エンコーダ（MRE：Magnetic Rotary Encoder）及び四肢内の関節のモーター及びロボットが車輪として使用しているボールを制御するセンサを有する基板の間におけるデータフローの処理に専用であるプロセッサをも含む。モーターは、確実な結合に必要とされる最大トルクの大きさに応じて、異なるタイプであってもよい。例えば、e-minebea（商標）のブラシ DC コードレスモーター（例えば、SE24P2CTCA）を使用することが可能であり、或いは、Maxon（商標）のブラシレス DC モーター（例えば、EC45_70W）を使用することもできる。MRE は、好ましくは、ホール効果を使用したタイプであり、12 又は 14 ビットの精度を有する。

10

20

30

40

【0017】

又、本発明の実施形態においては、図 1 に示されているロボットは、様々な種類のセンサを有する。これらのうちのいくつかは、ロボットの位置及び運動を制御するべく使用されている。これは、例えば、3 軸ジャイロメータと、3 軸加速度計と、を有するロボットの胴体内に配置された慣性ユニットのケースに当て嵌まる。又、ロボットは、5 フレーム / 秒における 5 メガピクセルの分解能と、水平方向の約 57° 及び垂直方向の約 44° の視野（FOV）と、を有する、Shenzhen V-Vision Technology Ltd.（商標）の製品（OV5640）などのシステムオンチップ（SOC：System On Chip）タイプの 2 つの 2D カラー RGB カメラをロボットの額（上部及び下部）において含むこともできる。又、20 フレーム / 秒において 0.3 メガピクセルの分解能を有し、且つ、2D カメラとほぼ同一の FOV を有する、ASUS XTION（商標）の SOC センサなどの 1 つの 3D センサをロボットの目の背後において含むこともできる。又、本発明のロボットは、その環境内の物体 / 生き物までのその相対的な位置を検知しうるように、例えば、頭部内において、3 つだけ、且つ、基部内において、3 つだけ、レーザーライン生成器を装備することもできる。又、本発明のロボットは、その環境内のサウンドを検知する能力を有するように、マイクロフォンを含むことができる。一実施形態においては、1 kHz における 300 mV / Pa + / - 3 dB の感度と、300 Hz ~ 12 kHz の周波数範囲（1 kHz との関係において -10 dB）と、を有する 4 つのマイクロフォンをロボットの頭部において埋植することができる。又、本発明のロボットは、その環境内の物体 / 人間までの距離を計測するべく、恐らくは、その基部の前面及び背面において配置された、2 つのソナーセンサを含むこともできる。又、ロボットは、人間とのやり取りを許容するべく、その頭部上において、且つ、その手の上部において、触覚センサを含むこともできる。又、ロボットは、自身のルート上において遭遇する障害物を検知するべく、その基部上において、バンパーを含むこともできる。

【0018】

又、その感情を変換し、且つ、その環境内の人間と通信するべく、本発明のロボットは、

- 例えば、その眼、耳の内部の、且つ、その肩の上部の LED と、
 - その耳の内部に配置された、例えば、2 つのラウンドスピーカと、
- を含むこともできる。

【0019】

50

本発明のロボットは、ベースステーション又はその他のロボットとEthernet RJ45又はWiFi802.11接続を通じて通信してもよい。

【0020】

本発明のロボットは、約400Whのエネルギーを有するリン酸鉄リチウム電池によって電力供給することができる。ロボットは、自身が含む電池のタイプに適合した充電ステーションにアクセスすることができる。

【0021】

ロボットの位置/運動は、センサの計測値に鑑み、それぞれの四肢によって定義されたチェーンと、それぞれの四肢の端部において定義されたエフェクタと、を起動するアルゴリズムを使用することにより、そのモーターによって制御されている。

10

【0022】

図2は、本発明の一実施形態による会話の方法を示している。このような方法の実装によって得られる会話は、「マルチモード」と呼称することが可能であり、その理由は、ロボットが、その応答を生成すべく、ユーザー（対話者）によって生成される、発話された単語、ジェスチャ、身体の姿勢、顔面の表現などのような質的に異なるイベントの組合せを考慮しているからである。上述の国際公開第2013/150076号パンフレットは、ロボットが、対話者の、例えば、手を振るなどの、ジェスチャに対しては反応するが、一緒に生成された口頭の及び非口頭のイベントの特定の組合せに対しては反応しない方法をも開示していることに留意されたい。

20

【0023】

図2に示されている方法の第1ステップi)においては、個々のセンサc1（マイクロフォン）及びc2（カメラ）からの入力信号s1、s2が、ロボットによって取得されており、且つ、抽出器モジュールEXTのバンクによって処理されている（ここでは、且つ、以下においては、「モジュール」という用語は、組み込まれているプロセッサにより、或いは、リモートセンサにより、実行されるソフトウェアモジュールを示すべく使用されており、ハードウェアの又はハードウェア-ソフトウェアのハイブリッドの実装形態が、常に可能であると共に本発明の範囲に含まれていることを理解されたい）。それぞれの抽出器モジュールは、入力信号又は所与のタイプの複数の信号を受信し、且つ、ロボットのその他のモジュールによって使用される情報を入力する。例えば、図2のケースにおいては、第1抽出器モジュールは、人間の音声と互換性を有するものとして識別されたサウンドを翻字することによって得られたテキスト出力TXTと、前記音声のイントネーション（幸せ、悲しみ、怒り、高圧的、不審など）を表すメタデータMDと、を提供すべく、マイクロフォンc1からの信号s1を処理しており、第2及び第3抽出モジュールは、前記カメラの視野内のユーザーの、それぞれ、顔面の、且つ、腕の、関心の的を表す「非テキストデータ」NTDを生成すべく、カメラc2からの信号s2を処理している。抽出器モジュールのバンクの出力は、会話エンジンモジュールDEへの入力として提供される。このモジュールによって実行される処理は、複雑なものとなる可能性があり、且つ、大きなサイズのデータベースに対するアクセスを必要とする。この理由から、この処理は、インターネット接続を通じてアクセスされる1つ又は複数のリモートサーバーRSによって部分的に実行されてもよい。

30

40

【0024】

会話エンジンモジュールは、データTXT、MD、NTDを入力として受信すると共にこれらを予め定義された「入力イベント」EVIに関連付ける認識モジュールRECを有する。例えば、モジュールRECは、テキストデータTXTを辞書の単語に関連付けてもよく、更には、ユーザーの顔面の関心の的の特定の構成を笑顔に関連付けてもよく、且つ、場合によっては、数値を前記笑顔に付与してもよく（例えば、0~5の値であり、この場合に、0は、笑顔なしを意味しており、且つ、5は、満面の笑顔を表している）、更には、ユーザーの腕の関心の的の特定の構成を、例えば、手を振るなどのジェスチャに関連付けてもよい。検討対象の特定の実施形態に応じて、認識モジュールのタスクは、抽出器モジュールによって実行することが可能であり、例えば、上述の笑顔の値を直接的に提供

50

する「笑顔抽出器」を有してもよい。

【0025】

ロボットのメモリ内に保存されている「会話のコンテキスト」又は「トピック」であるパラメータCTXは、認識モジュールの決定に対して影響を及ぼしうる。実際に、コンテキストに応じて、類似のエントリを異なるイベントとして解釈することが可能であり、例えば、異なるコンテキストにおいて、ユーザーの口が大きく開いていることは、あくびとして解釈することも可能であり、或いは、昏睡の表現として解釈することもできる。これは、本発明の方法の第2ステップ*i i*)に対応している。

【0026】

本発明の方法の第3ステップ*i i i*)は、応答を入力イベント又は入力イベントの組合せに対して関連付ける「規則適用」モジュールRULにより、実行されている。応答は、1つ又は複数の「出力イベント」EVOによって構成されており、これらの出力イベントは、ロボットによって発声されるべき単語又はフレーズ、ロボットによって放出されるべきサウンド、ロボットによって実行されるべきジェスチャ、ロボットの「顔面」の表現などであってもよい。先程引用した国際公開第2012/010451号パンフレットは、重要な変更を伴うことにはなるが、本発明において使用されうる規則適用モジュールについて記述している。実際に、本発明によれば、規則のうちの少なくともいくつかは、応答を、単一の入力イベントに対して関連付けるのではなく、少なくとも2つの一緒に生成されたイベントの組合せに対して関連付けており、これらのイベントのうちの少なくとも1つは、非口頭のものである(即ち、ユーザーによる単語又は文の発声を有していない)。

本発明の好適な実施形態によれば、規則のうちの少なくともいくつか、並びに、特に、その入力として複数のイベントを取得しているもののうちのいくつかは、その少なくとも1つが非口頭のものである出力イベントの組合せから構成された応答を判定している。

【0027】

例えば、可能な規則は、次のようなものであってもよい。

```
IF { (笑顔 > 2) AND [手を振る OR “こんにちは” OR “やあ”] }
THEN { (笑顔 = 4) AND 手を振る AND “こんにちは” }
```

【0028】

これは、ユーザーが少なくとも控えめな笑顔を伴って微笑んでおり、且つ、自身の手を振っているか、或いは、「こんにちは」又は「やあ」と言っている場合には、ロボットは、満面の笑顔、手を振ること、及び「こんにちは」という単語の発声により、回答するということを意味している。

【0029】

「一緒に生成された」イベントとは、会話を目的として、同時であると見なすべく十分に時間的に接近した2つ以上のイベントを意味している。例えば、ユーザーが自身の手を振り、且つ、次いで、一秒後に、「こんにちは」と言った場合には、これらの2つのイベントは、これらが厳密に言えば同時ではない場合にも、一緒に生成されたものと見なされる。

【0030】

その都度、適用可能な規則は、会話コンテキストCTXに依存しており、この会話コンテキストは、予め適用された規則及び/又は入力によって判定される。同一のコンテキスト又はトピックに関係する規則が、「会話」を形成し、会話は、国際公開第2011/003628号パンフレットによって開示されているように、プログラマによって編集することができる。会話トピックの例は、「フットボール」、「政治」、「料理」であってもよいが、ユーザーがロボットとの会話を開始する(或いは、後述するように、逆も又真である)際には、「出会い」であってもよく、或いは、ユーザーが去るか又は会話の終了の意思を表す際には、「さよなら」であってもよいであろう。

【0031】

更には、その都度、適用可能な規則は、ロボットの内部状態RISに依存する場合もあり、この内部状態は、予め適用された規則及び/又は入力によって判定される。内部状態

10

20

30

40

50

の例は、「幸せ」、「悲しみ」、「疲れ」であるが、「放電した電池」又は「機械的な障害」も含まれる。

【0032】

例えば、ユーザーが悲しみの表現を有しているとロボットが認識した場合には、その内部状態は、「心配」となる。次いで、ユーザーが、「今日は、身体の調子が良くない」と言った場合には、会話コンテキストは、「健康」という値（健康が会話のトピックとなることを通知する）を取得することになり、これにより、適切な規則の組を判定する。

【0033】

入力イベントの「生成」は、必ずしも、ユーザーによって実行される動作を必要とはしておらず、例えば、ユーザーがカラフルな衣服を着用しているという事実も、「イベント」となりうることを理解されたい。ユーザーによって発声された単語又は識別されたジェスチャを含んでいないイベント又はイベントの組合せに対する応答を判定するべく、「プロアクティブな規則」と呼称される特定の種類の規則が適用される。換言すれば、ロボットは、会話を開始することにより、部屋の中に存在している人々の数、沈黙しているユーザーの表現、衣服の色などのような刺激に対して反応する。本発明の特定の実施形態においては、いくつかの「雑談」トピックには、プロアクティブである、というラベルが付与されており、これは、前記トピックに関する規則のすべてがプロアクティブであることを意味している。「雑談」トピックの一例は、「笑顔」であり、これは、ユーザーが発話することなしに微笑んでいる際に適用される規則を含む。「料理」や「政治」などの相対的に具体的なトピックは、通常、プロアクティブではない。

10

20

【0034】

図3は、本発明の特定の実施形態による「プロアクティブ」な会話の実装形態を示している。抽出器バンクEXTは、シーンの異なる要素の色を識別するカラー抽出器COLと、笑顔抽出器SMIと、部屋の内部の人々の数を判定する抽出器モジュールNBPと、テキスト抽出器TXTXと、ジェスチャ抽出器GSTと、を有する。特定の状況においては、カラー抽出器が、赤色シャツを識別し、笑顔抽出器が、ユーザーの満面の笑顔（笑顔＝5）を認識し、且つ、NBPモジュールが、部屋の内部において2人の人々をカウントする一方において、モジュールTXTX及びGSTは、ユーザーが、発話もせず、且つ、明確に識別されるジェスチャをも実行していないことを通知している。次いで、会話エンジン、並びに、更に正確には、規則適用モジュールRULEは、会話データベースDDDBの、「雑談」トピックを収容したサブセットPRO内において、この状況に対して適用可能な「プロアクティブ」な規則をサーチすることになる。

30

【0035】

又、図2の方法は、ロボットの応答が、少なくとも単語又は文の発声から構成されているか又はこれを有する際に、その応答を活性化する任意選択のステップiii-a)をも有する。活性化とは、人間の「ボディトーク」をエミュレートするその発話を伴うロボットの運動の且つ/又はその他の非口頭によるイベント（例えば、表現の変化）のシーケンスである。活性化された応答は、発話及び運動を含むマルチモード応答とは弁別不能である場合もあるが、これらは、異なる方法で生成されている。マルチモード応答は、上述のように、規則適用モジュールによって直接的に判定されるのに対して、活性化は、この代わりに、図4、図5a、図5b、及び図5cを参照して後述するように、規則適用モジュールによって生成される出力固有のイベントEVO（即ち、口頭によるイベント、即ち、発声されるべき単語）をその入力として取得することにより、専用のモジュールANEにより、口頭による応答に追加される。

40

【0036】

図4に示されているように、活性化モジュール又はエンジンANEは、構文分析モジュールSYNTAと、ロボット上において設置された、又はロボットによってアクセス可能である、メモリ内において保存されている活性化リストASTと、表現力値を演算する2つのモジュールIOX及びFXと、を有する。「表現力値」は、運動が、「わざとらしく」なければならない、或いは、「別個」でなければならない、程度を判定するパラメータで

50

ある。「表現力係数」は、表現力値の変更を定義している。「表現力」という用語は、表現力値と表現力係数の両方を意味している。

【0037】

構文分析によれば、図5 a、図5 b、及び図5 cを参照して後述するように、活性化対象の1つ又は複数の単語と、単独では活性化されないが、活性化された1つ又は複数の単語の表現力に影響を及ぼす関係する単語と、を判定することができる。更には、構文分析モジュールは、例えば、テキスト内の「感情的な単語」の頻度及び/又はロボットの内部状態R I Sを考慮することにより、発声対象のテキストの「全体的」な表現力を判定してもよい。活性化対象のそれぞれの単語は、それ自体で表現力を有しており、この表現力は、モジュール10Xにより、関係する単語の表現力、並びに、テキストの全体的な表現力と組み合わせられ、モジュール10Xは、「ワンオフ表現力」と呼称される表現力値を出力する。

10

【0038】

又、活性化対象であるそれぞれの単語は、「コンセプト」にも関連付けられている。コンセプト及びワンオフ表現力は、活性化リストA L S T内において活性化を選択するべく、使用される。この選択は、単語に関連付けられたコンセプトと、モジュール10Xによって演算されたワンオフ表現力と、に依存している。例えば、リストのそれぞれの活性化は、1つ又は複数のコンセプトに関連付けられていてもよく、且つ、特定の表現力値を有していてもよく、このケースにおいては、活性化対象の単語によって表現されたコンセプトに関連付けられていると共にその特定の表現力値がワンオフ表現力に最も近接している活性化が選択される。図4の例においては、選択された活性化は、anim2と呼称され、且つ、exp2の特定の表現力を有している。最後に、モジュールFXは、最終的な表現力expfを演算するべく、選択された活性化の特定の表現力及びワンオフ表現力を組み合わせている(例えば、平均化している)。活性化エンジンの出力は、<活性化,最終的な表現力>というペアである。最終的な表現力値は、例えば、活性化を構成するジェスチャの速度及び/又は振幅を決定する。

20

【0039】

図5 aは、「He loves chocolate and beer.」という活性化対象の文の構文分析を示しており、構文ツリーは、証拠として、2つの補語をリンクする接続詞「AND」を示しており、これは、列挙を示している。このケースにおいては、この接続詞が、活性化対象の単語である。これは、コンセプト「列挙」と関連付けられており、このコンセプトは、ロボットが、自身の手を閉じ、その親指を伸ばし、且つ、次いで、その人差し指を伸ばす、というジェスチャを有する「2つ」と呼称される列挙と関連付けられている。

30

【0040】

図5 bは、「I agree with you.」という活性化対象の別の文の構文分析を示している。これは、肯定形の動詞と、主語と、補語と、を有する単純な文である。「with」を除くすべての単語が活性化され、「I」は、活性化「myself」により、活性化され、この場合に、ロボットは、それ自体を指し示し、「agree」は、活性化「yeah」により、活性化され、この場合に、ロボットは、頷き、且つ、youは、ロボットによって活性化される。

40

【0041】

これらの2つの例は、非常に単純なものであり、これらのケースにおいては、表現力は、なんの役割をも果たさない。更に複雑な例は、文「I strongly disagree with you.」によって構成され、その構文ツリーは、図5 cにおいて示されている。このケースにおいては、動詞は、否定形であり(文法的にではなく、意味的にである)、このようなケースにおいては、動詞自体が、活性化されるが、主語及び補語は活性化されない。更には、非同意を強調する副詞(「強く」)が存在している。

【0042】

動詞「disagree」は、コンセプト「非同意」と関連付けられており、且つ、0

50

～10のスケールにおいて、5の表現力を有する。但し、ワンオフ表現力は、副詞「strongly」の存在に起因し、5から8に増大する。又、本発明の一実施形態においては、ロボットの内部状態RISも、ワンオフ表現力値を変更しうるであろう。

【0043】

ロボットの表現の変化のみを有する3という特定の表現力を有する「oppose1」、それぞれ、6及び9の特定の表現力を有すると共にジェスチャをも含む「oppose2」及び「oppose3」というコンセプト「非同意」と関連付けられた3つの活性化が存在している。その特定の表現力がワンオフ表現力に近接している活性化は、「oppose3」であり、これが、次いで、選択される。但し、その最終的な表現力は、特定の表現力とワンオフ表現力の平均に対応した8.5に低減される。これは、ジェスチャが、

10

【0044】

図2を再度参照すれば、出力イベント及び/又は活性化は、ロボットの異なるアクチュエータを駆動して応答を「実行」するべく、使用されることがわかる。この図の例示用の実施形態においては、アクチュエータは、ラウドスピーカA1と、顔面表現制御アクチュエータの組A2と、四肢制御アクチュエータA3と、である。これは、図2の方法のステップiv)である。

【0045】

ロボットが、ユーザーのそばに立ち、且つ、ユーザーを直接的に凝視した場合には、人間型ロボットとの間における活性化された且つ/又はマルチモードの会話も、ぎこちなく且つ不自然なものに知覚される場合がある。更には、ロボットがユーザーに過剰に近接している場合には、ロボットは、活性化された又はマルチモードの応答を生成するべく「その手を用いて発話している」間に、ユーザーを殴打する場合がある。又、機能障害のケースにおいては、ロボットがユーザー上に落下するという一般的なリスクも存在している。これを理由として、本発明の好適な実施形態によれば、ロボットは、既定の(且つ、恐らくは、コンテキストに依存した)範囲内において、ユーザーからの距離を維持するべく、サーボ制御されている。有利には、距離は、例えば、その腰などのロボットの一部分とユーザーの下半身(腰まで)の間において計測され、この結果、ユーザーは、ロボットを後退させることなしに、ロボットに向かって傾斜し、且つ、自身の手でロボットに接触することができる。又、有利には、ロボットは、既定の(且つ、恐らくは、コンテキストに依存した)角度範囲内において、ユーザーとの関係における向きを維持するべく、サーボ制御されている。好ましくは、ロボットは、不自然に静止したロボットによって誘発される不安な感覚を回避するべく、前記距離及び角度範囲内に留まりつつ、疑似ランダムな平行運動及び/又は回転運動を実行する。

20

30

【0046】

図6は、ロボットR及びユーザーUを上方から示している。ロボットを中心とした基準フレームにおいて、ユーザーは、又は更に正確には、ユーザーの下半身は、距離範囲 $[d_1, d_2]$ 及び角度範囲 $[\theta_1, \theta_2]$ によって定義された認可領域AR内において留まる必要がある。ユーザーが運動した場合には、この状態を充足状態において維持するべく、

40

【0047】

ロボットの「自然」な振る舞いを得るべく、距離及び角度範囲は、会話の際に、進行中のトピックに応じて変化してもよい。

【0048】

ロボットとの関係におけるユーザーの位置は、画像処理モジュール、レーザーライン生成器、及び/又はソナーセンサと結合されたカメラを使用することにより、判定されてもよく、図1を伴う人間型ロボットの物理的アーキテクチャに関する上述の説明を参照されたい。

50

【0049】

図2を再度参照すれば、口頭のもの又は非口頭のものという異なる種類のイベントを認識するべく入力信号を解釈するステップ(i)は、本発明による方法の非常に重要なステップであることに留意されたい。イベントを認識するステップは、人間型ロボットの、又はこれによってアクセス可能な、メモリ内に保存されている予想イベントの既定のリストの項目に入力信号をマッチングさせるステップを意味している。有利には、前記予想イベントのリストは、会話のコンテキスト又はトピックに応じて、複数の前記リストのうちから選択される。

【0050】

例えば、発話認識は、センサによって取得されたサウンド信号を辞書の1つの自然言語単語又は一連の単語とマッチングさせるステップを有しており、これは、コンテキスト固有であってもよい。通常、それぞれのマッチング結果は、コンフィデンススコアに対して関連付けられており、スコアが大きいほど、マッチングの正確性の確率も大きくなる。通常、閾値は、「成功」したマッチングと失敗したイベント識別の試みを弁別するべく、使用される。

【0051】

当技術分野においては、認識対象の特定の種類のイベントに応じて、異なる複雑性を有するいくつかのマッチング方法が知られている。例えば、発話認識の分野においては、以下の方法(或いは、むしろ、方法の集団)が知られている。

【0052】

- 正確なマッチング:これは、入力が正確に単語又は文を含んでいるかどうかをチェックするべく有限状態機械を使用する最も単純且つ高速の方法である。コンフィデンススコアは、マッチングが確実である(スコア=1)又は識別の試みに失敗した(スコア=0)というブール値である。

【0053】

- 概略的なマッチング:こちらも、有限状態機械に基づいているが、これは、マッチングチェーン内において、ある程度の誤りを許容している。コンフィデンススコアは、誤りの数が増大するのに伴って減少する。

【0054】

- 入力と辞書の単語又は文の間における音声学的距離の判定に基づいた音声学的マッチングである(発話認識のみ)。

【0055】

- 意味論的なマッチングは、最も複雑な方法であり、入力における観察されたボキャブラリとそれぞれの会話エントリにおけるボキャブラリの間における距離の演算に基づいている。この距離は、前記入力及び前記エントリのベクトル表現の間におけるコサイン測度(cosine measure)である。ベクトルは、TF-IDF(Term Frequency - Inverse Document Frequency)重み付けを使用することにより、「bag-of-words」分散型意味表現に準拠して算出される。

【0056】

単一のマッチング法を使用するのではなく、ロボットは、最も単純な方法から開始し、コンフィデンススコアが予め設定された閾値を超過している場合には、結果を受け入れ、且つ、超過していない場合には、更に複雑な方法を試し、最も複雑なマッチング法(例えば、意味論的なもの)を使用して得られたコンフィデンススコアが、依然として閾値未満である場合には、サーチに失敗している、という階層的な方式を使用してもよい。このケースにおいては、ロボットは、入力を無視するか、或いは、解明を求める(例えば、失敗した発話認識のケースにおいては、「すみません。なんとおっしゃいましたか?」と発話することにより、実行する)。

【0057】

又、階層構造は、使用する発話認識技術などの要因に対して適合されてもよい。意味論

10

20

30

40

50

的なマッチングは、ASR (Automatic Speech Recognition) が大規模な言語モデルに基づいている場合には、好ましいものとなる一方において、音声学的なマッチングは、相対的に安定していない組込み型のASRの結果からの誤りの回復を支援することになる。

【0058】

有利には、ロボットは、異なるパラメータに応じて、且つ、具体的には、会話のコンテキスト又はトピックに応じて、マッチング方法のサブセットを選択してもよい。継続中の会話が、いくつかの異なる入力しか予想されない「閉じた」ものである場合には、正確なマッチングが成功裏に機能する可能性が高く、且つ、従って、試みる価値がある。逆に、多数の可能な入力イベントを許容する非常に広範なコンテキストのケースにおいては、正確なマッチング及び概略的なマッチングを取り止め、且つ、音声学的な方法、或いは、場合によっては、意味論的な方法により、直接的に開始することが好ましいであろう。図7の右側には、演算の複雑性が増大するマッチング方法MM1~MM4の階層的なチェーンが示されている。それぞれのマッチング方法ごとに、マッチングに成功し、入力イベントEVIが生成されるか、或いは、マッチングに失敗し、次のマッチング方法が試みられる(MM4を除く)という2つの結果が可能である。試みられるべき最初のマッチング方法は、必ずしもMM1である必要はなく、会話のコンテキストCTX、並びに、恐らくは、その他のパラメータに応じて、マッチング方式エンジンMSEによって選択される。

10

【0059】

インターネット接続が利用可能である場合には、少なくとも、1つ又は複数の最も複雑なマッチング方法は、リモートサーバーによって実行されてもよい(図2を参照されたい)。

20

【0060】

図7は、適切な抽出器によって人間の音声として認識されたサウンドを翻字することによって得られたテキストTXTを入力信号として取得する発話認識のケースを参照しているが、この方式は、相対的に一般的である。これは、「マルチモード」会話のケースに限定されるものではないことを理解されたい。

【0061】

以下、図8を参照し、音声学的マッチングに基づいた特定の発話認識方法について説明することとする。

30

【0062】

センサ(マイクロフォン)c1によって取得されたサウンドは、入力として文字化モジュールTRSCに提供され、このモジュールは、これらをテキストに変換する。次いで、このテキストは、音声学的変換モジュールPHONにより、会話の言語の特殊性(これは、例えば、当技術分野において既知のカメラ及び顔面認識モジュールの支援によって認識されたユーザーの身元情報に応じて、ロボットによって判定されたパラメータである)を考慮することにより、その音声学的均等物に変換される。又、文字化及び音声学的変換は、一緒に実行することも可能であり、これらは、一緒に「音声学的文字化」と呼称されうるものを構成する。

【0063】

次いで、音声学的文字化は、単純化モジュールSIMPにより、単純化及びスムージングされる。

40

【0064】

「単純化ステップ」は、例えば、「d」と「t」又は「k」と「g」などのように、互いに混同される可能性が高い異なる音素を単一の音素によって表すステップを有する。

【0065】

「スムージングステップ」は、もたらされる情報を保持しつつ、文字化モジュールによって提案されたステートメントセグメンテーション(しばしば、認識誤りの発生源において位置している)を無視するステップを有する。この点において、(文字化モジュールによって識別される)それぞれの単語の開始点に位置している母音と、鼻母音と、を除いて

50

、母音が無視される。INDEX内に含まれている予想単語には、同一の又は類似した処理が（有利には、オフライン状態において）適用される。距離演算モジュールDISTが、入力サウンドの単純化及びスムージングされた音声学的文字化とインデックスの単純化及びスムージングされたエントリの間における編集距離を判定する。次いで、選択モジュールSELが、最小編集距離に対応したエントリを選択する。

【0066】

例として、ユーザーが「A demain」と（即ち、「See you tomorrow」と）仏語において言った場合には、音声学的文字化は、

【数1】

「ADCEMIN」

10

となり、これは、次いで、「ATM N」として単純化される（「N」は、鼻母音を表している）。

【0067】

編集距離は、文字のストリングを別のものに変換するために必要とされる変化の最少数として定義されている。例えば、ADMNとBDLNSの間の編集距離は、3であり、その理由は、以下のように3つの変化が必要とされるからである。

- ADMN BDMN（「A」が「B」に変更される）
- BDMN BDLN（「M」が「L」に変更される）
- BDLN BDLNS（「S」が追加される）

20

【0068】

以上、マルチモード会話、活性化された発話、ロボット位置のサーボ制御、及びイベント（並びに、更に詳しくは、発話）認識の特定の方法を組み合わせた特定の実施形態を検討することにより、本発明について説明した。これらは、相乗効果をもたらす状態において最良に機能するが、本発明のこれらの異なる態様は、互いに独立した状態において実装することもできる。

【 図 1 】

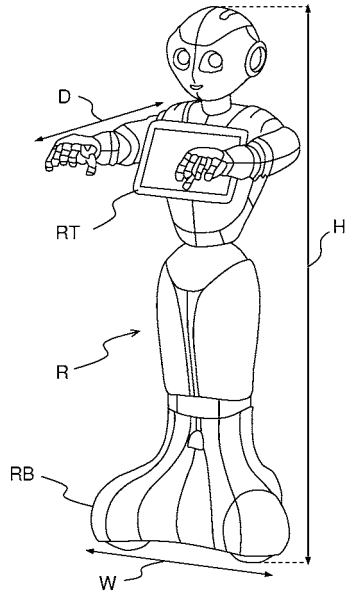


図 1

【 図 2 】

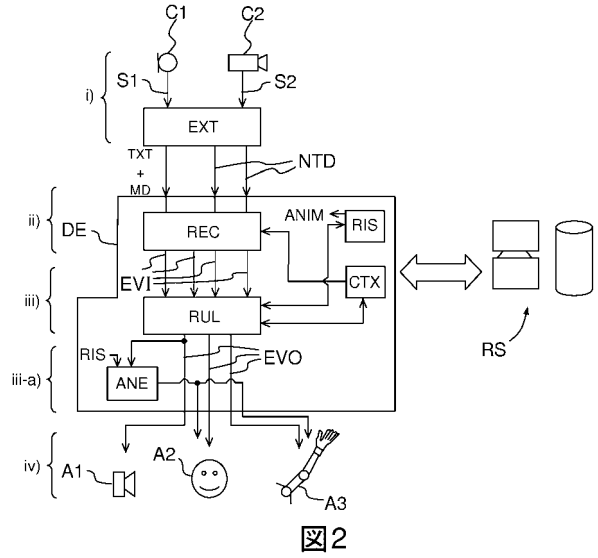


図 2

【 図 3 】

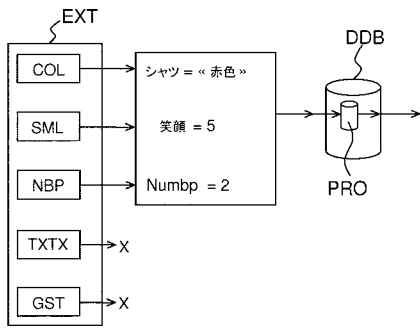


図 3

【 図 4 】

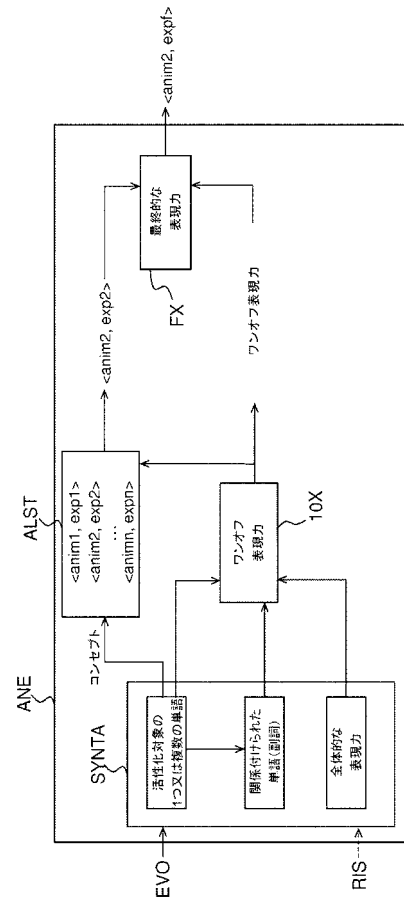


図 4

【 図 5 a 】

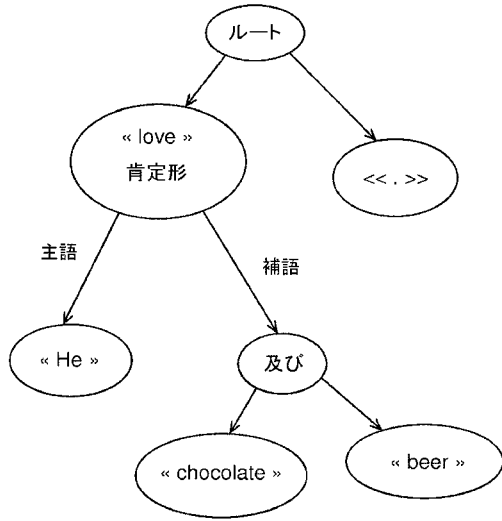


図 5a

【 図 5 b 】

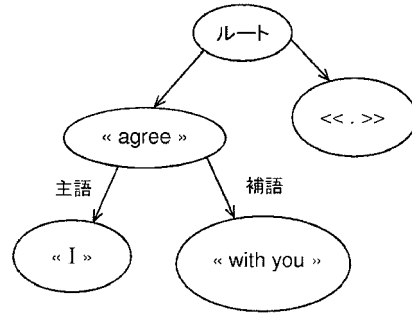


図 5b

【 図 5 c 】

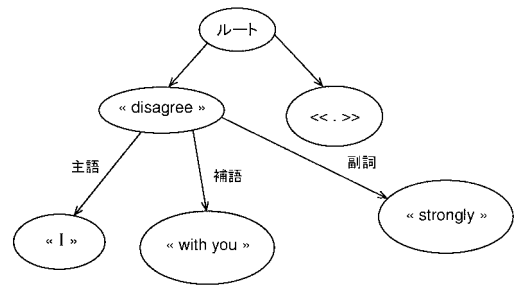


図 5c

【 図 6 】

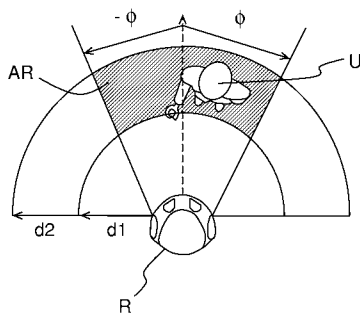


図 6

【 図 7 】

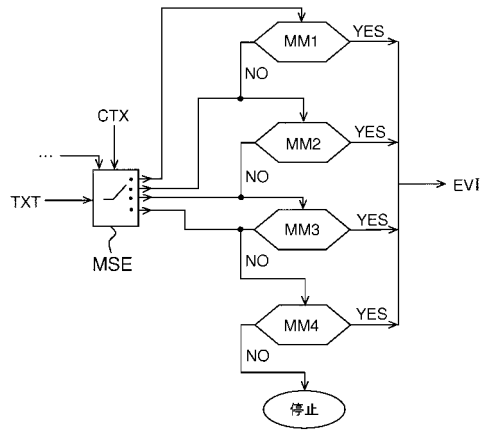


図 7

【 図 8 】

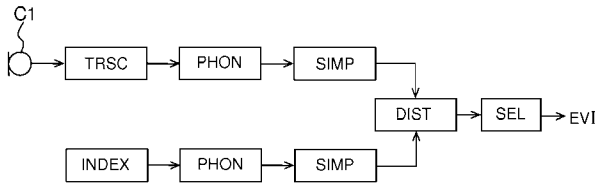


図8

【 手続補正書 】

【 提出日 】平成31年4月24日(2019.4.24)

【 手続補正 1 】

【 補正対象書類名 】特許請求の範囲

【 補正対象項目名 】全文

【 補正方法 】変更

【 補正の内容 】

【 特許請求の範囲 】

【 請求項 1 】

人間型ロボット（R）と少なくとも一人のユーザー（U）の間における会話を実行する方法であって、前記人間型ロボットによって反復的に実行される、

i) 個々のセンサ（c1、c2）から複数の入力信号（s1、s2）を取得するステップであって、少なくとも1つの前記センサは、サウンドセンサであり、且つ、少なくとも1つのその他のセンサは、モーション又は画像センサである、ステップと、

ii) 少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された、前記ユーザーによって生成された複数のイベント（EVI）を認識するべく前記取得された信号を解釈するステップと、

iii) 少なくとも単語又は文の発声、音声のイントネーション、ジェスチャ、身体の姿勢、顔面の表現を有する群から選択された少なくとも1つのイベント（EVO）を有する前記人間型ロボットの応答を決定するステップであって、前記決定するステップは、規則の組を適用することによって実行され、それぞれの前記規則は、入力イベントの組を前記ロボットの応答に対して関連付けている、ステップと、

iv) 前記人間型ロボットにより、前記イベント又はそれぞれの前記イベントを生成するステップと、

を有し、

前記ステップ i i i) において適用された前記規則の少なくともいくつかは、応答を前記ユーザーによって一緒に生成されると共に前記ステップ i i i) において認識される少なくとも2つのイベントの組合せに関連付けており、これらのイベントのうちの少なくとも1つは、前記ユーザーによって発声された単語又は文ではなく、

前記ステップ i i i) は、取得された信号と前記人間型ロボットの、又はこれによってアクセス可能である、メモリ内に保存されている予想イベントのリストに属するイベントの間におけるマッチングをサーチするステップを有し、前記サーチするステップは、既定の値を上回るコンフィデンススコアを有するイベントが認識されるまで、増大する複雑性を有する複数のマッチング法 (M M 1 ~ M M 4) を連続的に使用することによって、或いは、最高の複雑性を有する認識方法が使用された後に、実行される、方法。

【請求項 2】

前記使用されるマッチング方法は、会話のコンテキストに応じて選択される請求項 1 に記載の方法。

【請求項 3】

前記マッチング方法は、増大する複雑性の順序において、正確なマッチングについてのサーチ、概略的なマッチングについてのサーチ、音声認識のケースのみにおける音声学の対応性についてのサーチ、及び意味論的な対応性についてのサーチを含む請求項 1 又は 2 に記載の方法。

【請求項 4】

音声学の対応性についてサーチする前記方法は、

- サウンドセンサによって取得されたサウンドの組の音声学的文字化のステップと、
 - 前記結果的に得られた音声学的文字化を単純化及びスムージングするステップと、
 - 前記単純化及びスムージングされた音声学的文字化と自然言語における予め定義された単語の組を単純化及びスムージングすることによって得られた複数のエントリの間における編集距離を算出するステップと、
 - 前記単純化及びスムージングされた音声学的文字化から最小の編集距離を有する前記エントリに対応した予め定義された組の自然言語単語を選択するステップと、
- を有する請求項 3 に記載の方法。

【請求項 5】

前記単純化及びスムージングするステップは、

- 混同されやすい音素を単一の音素によって置換するステップと、
- 単語の開始点に位置した母音及び鼻母音以外の母音を除去するステップと、
- 単語の間の休止部を除去するステップと、

を有する請求項 4 に記載の方法。

【請求項 6】

前記予想イベントのリストは、会話のコンテキストに応じて、複数の前記リストのうちから選択される請求項 1 ~ 5 のいずれか 1 項に記載の方法。

【請求項 7】

前記ステップ i i i) において適用される前記規則のうちの少なくともいくつかは、前記人間型ロボットによって一緒に生成される少なくとも2つのイベントを有する応答を決定し、これらのイベントの少なくとも1つは、前記単語又は文の発声ではない請求項 1 ~ 6 のいずれか 1 項に記載の方法。

【請求項 8】

前記ステップ i i i) において、前記人間型ロボットの応答は、会話のコンテキスト (C T X)、前記ユーザーの身元情報、前記人間型ロボットの内部状態 (R I S) から選択された少なくとも1つのパラメータに基づいて決定される請求項 1 ~ 7 のいずれか 1 項に記載の方法。

【請求項 9】

前記ステップ i i i) において認識された又は前記ステップ i i i) において決定された少なくとも1つのイベントに従って前記パラメータ又は少なくとも1つの前記パラメータ

の値を変更するステップを更に有する請求項 8 に記載の方法。

【請求項 10】

前記ステップ i i i) は、プロアクティブな規則と呼称される予め定義されたサブセット (P R O) に属する規則を適用することにより、前記ユーザーによって発声された単語又は識別されたジェスチャの欠如を含むイベントの組に対する応答を決定するステップを有する請求項 1 ~ 9 のいずれか 1 項に記載の方法。

【請求項 11】

ステップ i i i) において決定された前記応答が、少なくとも前記単語又は文の発声であるか、又はこれを有する場合に、発声対象の前記単語又は文の語学的分析と、前記分析に応じた、前記応答を伴う活性化の決定と、を実行するステップ i i i - a) の実行を更に有する請求項 1 ~ 10 のいずれか 1 項に記載の方法。

【請求項 12】

前記ステップ i i i - a は、
) 活性化対象である前記応答の少なくとも 1 つの単語を識別するサブステップと、
) コンセプトと、活性化対象である前記単語又はそれぞれの前記単語と関連付けられたワンオフ表現力と呼称される表現力と、を決定するサブステップと、
) 前記人間型ロボットの、又はこれによってアクセス可能である、メモリ内に保存されている活性化のリスト (A L S T) から、前記コンセプト及び前記ワンオフ表現力に基づいて、活性化を選択するサブステップと、
 を有する請求項 11 に記載の方法。

【請求項 13】

前記サブステップ は、前記文の構造内におけるその機能に応じて、活性化対象であるそれぞれの又は前記単語を決定するべく、発声対象である文の構文分析を実行するステップを有する請求項 12 に記載の方法。

【請求項 14】

前記サブステップ において、前記ワンオフ表現力は、前記単語の表現力、前記単語に関係付けられた 1 つ又な複数のその他の単語の表現力、及び全体的な応答の全体的な表現力から選択された少なくとも 1 つのパラメータに基づいて決定される請求項 12 又は 13 に記載の方法。

【請求項 15】

前記リストのそれぞれの活性化は、1 つ又は複数のコンセプトと関連付けられ、且つ、特定の表現力を有しており、前記サブステップ は、前記リストのうちにおいて、前記サブステップ において決定された前記コンセプトと関連付けられると共に前記ワンオフ表現力に最も近接した特定の表現力を有する活性化を選択するステップを含む請求項 12 ~ 14 のいずれか 1 項に記載の方法。

【請求項 16】

) 前記特定の表現力及び前記ワンオフ表現力に基づいて、最終的な表現力と呼称される表現力を決定するサブステップを更に有する請求項 15 に記載の方法。

【請求項 17】

前記ワンオフ表現力又は前記最終的な表現力は、前記活性化の少なくとも 1 つのジェスチャの速度及び振幅のうちから選択された少なくとも 1 つのパラメータを決定する請求項 12 ~ 16 のいずれか 1 項に記載の方法。

【請求項 18】

前記ステップ i) ~ i v) と同時に前記ロボットによって反復的に実装される、
 A) 前記ロボット (R) に設定された基準フレームとの関係において前記ユーザー (U) の身体の少なくとも一部分の位置を決定するステップと、
 B) 予め定義された値の範囲において前記ロボット又はその 1 つの要素と前記ユーザーの前記少なくとも 1 つの又は前記身体部分の間の距離を維持するべく、前記ロボットの少なくとも 1 つのアクチュエータを駆動するステップと、
 を更に有する請求項 1 ~ 17 のいずれか 1 項に記載の方法。

【請求項 19】

前記ステップ B) は、既定の角度範囲内において前記ユーザーとの関係における前記ロボットの向きを維持するべく、前記ロボットの少なくとも 1 つのアクチュエータを駆動するステップを更に有する請求項 18 に記載の方法。

【請求項 20】

C) 前記既定の値の範囲内において前記距離を維持すると共に、適宜、前記既定の角度範囲内において前記向きを維持しつつ、前記ロボットの前記疑似ランダムな変位を生成するように前記アクチュエータ又は少なくとも 1 つの前記アクチュエータを駆動するステップを更に有する請求項 18 又は 19 に記載の方法。

【請求項 21】

D) 前記ユーザーと前記人間型ロボットの間における継続中の会話の意味論的分析を実行し、且つ、前記分析に従って、前記既定の距離値の範囲を変更すると共に、適宜、前記既定の角度範囲を変更するステップを更に有する請求項 18 ~ 20 のいずれか 1 項に記載の方法。

【請求項 22】

前記ステップ A) は、前記ロボットに設定された前記基準フレームとの関係において前記ユーザーの下半身の位置を決定するステップを有する請求項 18 ~ 21 のいずれか 1 項に記載の方法。

【請求項 23】

プログラムコード命令を有するコンピュータプログラムであって、前記プログラムコード命令は、前記プログラムが、人間型ロボット (R) 上において組み込まれた少なくとも 1 つのプロセッサによって実行された際に、請求項 1 ~ 22 のいずれか 1 項に記載の方法を実行し、前記ロボットは、個々の入力信号を取得するべく、前記プロセッサ又は少なくとも 1 つのプロセッサに対して動作可能に接続されると共に少なくとも 1 つのサウンドセンサ及び少なくとも 1 つの画像又は運動センサを有する複数のセンサ (c1、c2) と、単語又は文を発声するべく、前記プロセッサ又は少なくとも 1 つの前記プロセッサによって制御される発話合成モジュールと、前記ロボットが複数の運動又はジェスチャを実行できるようにする前記プロセッサ又は少なくとも 1 つの前記プロセッサによって駆動されるアクチュエータの組 (A1、A2、A3) と、を有する、コンピュータプログラム。

【請求項 24】

人間型ロボット (R) であって、

- 少なくとも 1 つの組み込み型のプロセッサと、
 - 個々の入力信号を取得するべく、前記プロセッサ又は少なくとも 1 つの前記プロセッサに対して動作可能に接続されると共に少なくとも 1 つのサウンドセンサ及び少なくとも 1 つの画像又は運動センサを有するセンサ組立体 (c1、c2) と、
 - 単語又は文を発声するべく、前記プロセッサ又は少なくとも 1 つの前記プロセッサによって駆動される発話合成モジュールと、
 - 前記ロボットが複数の運動又はジェスチャを実行できるようにする前記プロセッサ又は少なくとも 1 つの前記プロセッサによって駆動されるアクチュエータの組 (A1、A2、A3) と、
- を有し、

前記人間型ロボットは、前記プロセッサ又は少なくとも 1 つの前記プロセッサが、請求項 1 ~ 22 のいずれか 1 項に記載の方法を実行するようにプログラム又は構成されていることを特徴としている、人間型ロボット。

【請求項 25】

少なくとも 1 つのリモートサーバーに接続するための装置を更に有し、前記プロセッサ又は少なくとも 1 つの前記プロセッサは、請求項 1 ~ 22 のいずれか 1 項に記載の方法を実行するべく、前記リモートサーバー又は少なくとも 1 つの前記リモートサーバーと協働するようにプログラム又は構成されている請求項 24 に記載の人間型ロボット。

フロントページの続き

(51) Int.Cl.	F I	テーマコード (参考)
B 2 5 J 13/00 (2006.01)	G 1 0 L 15/10 4 0 0 R	
	G 1 0 L 15/04 2 0 0	
	B 2 5 J 13/00 Z	

(72)発明者 **ダビド・ウーサン**
フランス国、7 5 0 1 5 ・パリ、リュ・エムリオール・2 2

(72)発明者 **ガブリエル・バルビエリ**
フランス国、7 5 0 1 7 ・パリ、リュ・ゴテエ・4 6

(72)発明者 **ジョスラン・マルタン**
フランス国、9 2 1 2 0 ・モンルーージュ、アブニュ・マルクス・ドルモワ・1 6 2

(72)発明者 **ジャン・テスタール**
フランス国、7 5 0 0 8 ・パリ、アブニュ・フランクリン・ルーズベルト・3 9

(72)発明者 **イルモ・グルダン**
フランス国、7 8 2 1 0 ・サン - シル・レコール、プロムナード・デ・サンジュ・7

Fターム(参考) 3C707 AS34 BS27 CS08 HS27 KS11 KS12 KT01 WA16 WL05 WL07
WL14 WL16 WM07
5E555 AA76 BA01 BB01 BC21 CA42 CA44 CA47 CB21 CB64 CB66
CC01 DA23 EA22 EA23 EA25 FA30