

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6655552号  
(P6655552)

(45) 発行日 令和2年2月26日(2020.2.26)

(24) 登録日 令和2年2月5日(2020.2.5)

(51) Int.Cl.			F I		
<b>G 1 0 L</b>	<b>15/22</b>	<b>(2006.01)</b>	G 1 0 L	15/22	3 0 0 Z
<b>G 1 0 L</b>	<b>13/10</b>	<b>(2013.01)</b>	G 1 0 L	13/10	1 1 4
<b>B 2 5 J</b>	<b>19/00</b>	<b>(2006.01)</b>	B 2 5 J	19/00	Z

請求項の数 20 (全 20 頁)

(21) 出願番号	特願2016-562858 (P2016-562858)	(73) 特許権者	311003754
(86) (22) 出願日	平成27年4月17日 (2015.4.17)		ソフトバンク・ロボティクス・ヨーロッパ
(65) 公表番号	特表2017-517028 (P2017-517028A)		SOFTBANK ROBOTICS E
(43) 公表日	平成29年6月22日 (2017.6.22)		UROPE
(86) 国際出願番号	PCT/EP2015/058356		フランス国、75015・パリ、リュ・デ
(87) 国際公開番号	W02015/158878		ュ・コロネル・ピエール・アビア・43
(87) 国際公開日	平成27年10月22日 (2015.10.22)	(74) 代理人	110001173
審査請求日	平成28年11月15日 (2016.11.15)		特許業務法人川口国際特許事務所
(31) 優先権主張番号	14305579.6	(72) 発明者	メゾニエ, ブリュノ
(32) 優先日	平成26年4月17日 (2014.4.17)		フランス国、75014・パリ、アブニュ
(33) 優先権主張国・地域又は機関	欧州特許庁 (EP)		・デュ・ジェネラル・ルクレール、19、
前置審査		(72) 発明者	ウーサン, ダビド
			フランス国、75015・パリ、リュ・エ
			ムリオール・22

最終頁に続く

(54) 【発明の名称】 ロボットとの対話を取り扱う方法とシステム

(57) 【特許請求の範囲】

【請求項1】

ヒューマノイドロボットと人間ユーザ間の音声対話を取り扱うコンピュータ実施方法であって、前記方法は、

前記音声対話中に、音声データを受信し前記音声データをテキストデータへ変換する工程と、

前記テキストデータの1つまたは複数の対話モード実行規則の検証に応じて修正対話モードを選択する工程と、を含み、

対話モードは1つまたは複数の対話コンテンツと前記1つまたは複数の対話コンテンツのための1つまたは複数の対話音声スキンを含み、対話コンテンツは所定文章の集合を含み、前記集合は質問文と回答文を含み、対話音声スキンは、周波数、トーン、速度およびピッチを含む音声レンダリングパラメータを含み、前記1つまたは複数の対話コンテンツおよび音声スキンはウェブプラットフォームを使用することによりオンラインでオーサリングされ、

1つまたは複数の対話コンテンツの選択は1つまたは複数のフィルタを適用することにより緩和され、前記フィルタは1つまたは複数の単語のブラックリスト、1つまたは複数の単語のホワイトリスト、および/または対話モード実行規則を含み、

前記ロボットにより表現される対話に対する前記1つまたは複数の対話コンテンツおよび/または音声スキンの使用の規則はセキュアブート方法の使用を含む、方法。

【請求項2】

前記1つまたは複数の対話コンテンツおよび音声スキンのオーサリングは、所定の音声スキンを混合する工程および/またはウェブプラットフォームユーザに音声スキンを微調整する生成コントロールを提供する工程および/または音声または記録された文章またはそれらの組合せをアップロードし共有する工程を含む1つまたは複数の工程を含む、請求項1に記載の方法。

【請求項3】

前記1つまたは複数の対話コンテンツおよび/または音声スキンのうち1つまたは複数の所定の対話コンテンツおよび/または音声スキンは複数のエンティティにより修正される、請求項1または2に記載の方法。

【請求項4】

前記1つまたは複数の対話コンテンツのうち1つまたは複数の所定の対話コンテンツはインターネットから抽出される、請求項1に記載の方法。

【請求項5】

1つまたは複数の対話コンテンツを修正する工程と、前記1つまたは複数の修正対話コンテンツに関して1つまたは複数のユーザフィードバックを受信する工程と、前記1つまたは複数の修正対話コンテンツを別のロボットへ伝達する工程とをさらに含む請求項1に記載の方法。

【請求項6】

前記1つまたは複数の対話コンテンツの使用を規制する工程をさらに含む請求項1に記載の方法であって、前記規制工程は、1つまたは複数の対話コンテンツおよび/または音声スキンおよび/または関連実行規則の使用を変調、フィルタ処理、減衰、増幅、増加、奨励、低減、禁止、制限、回避、または禁止する工程を含む1つまたは複数の工程を含む、方法。

【請求項7】

前記ロボットの製造者は、アナログ音声復元前の最終音声レンダリングまたは最後の点を制御する、請求項1に記載の方法。

【請求項8】

前記選択された修正対話モードを実行する工程をさらに含む請求項1に記載の方法。

【請求項9】

前記修正対話モードは、現在の対話コンテンツおよび/または現在の対話の現在の対話音声スキンを修正することにより得られる請求項1または2に記載の方法。

【請求項10】

前記現在の対話コンテンツを修正する工程は、前記対話コンテンツの単語の同義語を使用する工程と、前記対話コンテンツの1つまたは複数の単語の挿入または順列または置換を含む構文修正を適用する工程とを含む、請求項9に記載の方法。

【請求項11】

前記現在の対話音声スキンを修正する工程は前記現在の対話音声スキンの周波数および/またはトーンおよび/または速度および/またはピッチを修正する工程を含む、請求項9に記載の方法。

【請求項12】

前記修正対話モードは所定の対話モードを活性化することにより得られ、前記所定の対話モードは1つまたは複数の対話実行規則により選択され、前記所定の対話モードは所定の対話コンテンツおよび/または所定の対話音声スキンを含む、請求項1に記載の方法。

【請求項13】

対話モード実行規則は前記ロボットにより認識された環境に依存する、請求項1乃至12のいずれか一項に記載の方法。

【請求項14】

対話モード実行規則は、ユーザの年齢、ユーザの性別、ユーザの気分、ユーザの感情、ユーザの数、ユーザとの相互作用履歴、ユーザ嗜好、ロボットおよび/またはユーザの空間的配置、ロボットおよび/またはユーザのジェスチャまたはジェスチャの組み合わせ、

10

20

30

40

50

ロボットの環境内に検知された事象、ローカル天気、地理的位置、日付、時間およびそれらの組み合わせを含むリストから選択されたパラメータを含む、請求項 1 3 に記載の方法。

【請求項 1 5】

対話モード実行規則はインターネットから動的に取り出されるおよび/またはユーザ構成可能である、請求項 1 に記載の方法。

【請求項 1 6】

対話コンテンツおよび/または対話音声スキンは音声命令またはユーザ要求により選択される、請求項 1 に記載の方法。

【請求項 1 7】

欠落対話モードを識別する工程と、前記ユーザとの対話中に前記欠落対話モードを取り出し前記ロボットへインストールする工程とをさらに含む請求項 1 に記載の方法。

【請求項 1 8】

選択された対話モードを実行する前に前記ユーザの確認を受信する工程をさらに含む請求項 1 に記載の方法。

【請求項 1 9】

コンピュータプログラムであって、好適なコンピュータ装置上で請求項 1 乃至 1 5 のいずれか一項に記載の方法の工程を実行するための命令を含むコンピュータプログラム。

【請求項 2 0】

請求項 1 乃至 1 5 のいずれか一項に記載の方法の工程を実行するようにされたコンピュータを含むシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本特許は、特にロボットと人間ユーザ間の会話の特定状況における、デジタルデータ処理の領域に関し、より具体的には音声合成および双方向対話の取り扱いに関する。

【背景技術】

【0002】

コンパニオンロボットは有利には、人間との感情的関係を確立し得る。音声スキンまたは対話スキンを介した対話の動的適応化は豊かな相互作用を可能にし得る。

【0003】

スピーチまたは音声合成のための既存システムはたいてい受動的でありかつ一様である：すなわち、男性または女性音声選択など少しの選択肢を除き、スピーチ生成エンジンのトーンはかなり中性的である。その上、提供される応答は文化的関連性(cultural references)に欠ける。工業的または大量市場音声応答システムの目的は正確には、普遍的に受け入れられる応答を提供することである、すなわち、できるだけ広く理解されることである。これは、いかなる状況のおよびさらには文化的関連性も回避することを意味する。音声命令は通常、特定状況に限定される。例えば、音声口述ソフトウェア(voice dictation software)は殆どの場合、スタンドアロンソフトウェアアプリケーション(例えばワープロソフト)の状況で使用される。最新オペレーティングシステムをますます備えるようになったいくつかのアクセス可能性特徴によると、ユーザは、いくつかの行為(例えば、アプリケーションを立ち上げる、コピーアンドペースト等)を行うために音声命令を使用し得る。これらの所定行為はかなり限定される。このような視覚的または音声相互作用モードは一般的には受動的である(例えば、ユーザが命令を能動的に与え、機械は命令を実行する)。例えば応答システムに実装されたものなど最近のコンピュータ相互作用モデルによってさえ、限定された相互作用が機械からユーザへ発生する。

【0004】

コンパニオンヒューマノイドロボットの状況では、人間ユーザとの相互作用モデルは、パーソナルコンピュータとの相互作用モデル(およびそれらの様々な形態)と比較すると

10

20

30

40

50

、著しく変わる。ロボットとの認識相互作用は、タブレットPCまたはスマートフォンのもとは基本的に異なる。特に、ロボットのスピーチ合成を変調する能力は、豊かな相互作用のカギでないにしても有益であり得、しいては関連データを収集し、ロボットまたは接続装置により与えられるサービスを改善できるようにし得る。

【0005】

特にロボットと人間ユーザ間の会話の特定状況において音声合成（形式）および関連双方向対話（実体）を取り扱う方法およびシステムの需要がある。

【発明の概要】

【課題を解決するための手段】

【0006】

ロボットと人間ユーザ間の音声対話を取り扱うコンピュータ実施方法が提供される。本方法は、前記音声対話中に、音声データを受信し前記音声データをテキストデータへ変換する工程と、前記テキストデータの1つまたは複数の対話モード実行規則の検証に応じて修正対話モードを選択する工程とを含み、対話モードは1つまたは複数の対話コンテンツと1つまたは複数の対話音声スキンを含み、対話コンテンツは所定文章の集合を含み、前記集合は質問文と回答文を含み、対話音声スキンは、周波数、トーン、速度およびピッチを含む音声レンダリングパラメータを含む。

【0007】

一つの発展形態では、本方法はさらに、前記選択された修正対話モードを実行する工程を含む。

【0008】

一つの発展形態では、修正対話モードは、現在の対話コンテンツおよび/または現在の対話の現在の対話音声スキンを修正することにより得られる。

【0009】

一つの発展形態では、現在の対話コンテンツを修正する工程は、前記対話コンテンツの単語の同義語を使用する工程と、前記対話コンテンツの1つまたは複数の単語の挿入または順列または置換を含む構文修正を適用する工程とを含む。

【0010】

一つの発展形態では、現在の対話音声スキンを修正する工程は、現在の対話音声スキンの周波数および/またはトーンおよび/または速度および/またはピッチを修正する工程を含む。

【0011】

一つの発展形態では、修正対話モードは所定対話モードを活性化することにより得られ、前記所定対話は1つまたは複数の対話実行規則により選択され、前記所定対話モードは所定対話コンテンツおよび/または所定対話音声スキンを含む。

【0012】

一つの発展形態では、対話モード実行規則はロボットにより認識される環境に依存する。

【0013】

一つの発展形態では、対話モード実行規則は、ユーザの年齢、ユーザの性別、ユーザの気分、ユーザの感情、ユーザの数、ユーザとの相互作用履歴、ユーザ嗜好、ロボットおよび/またはユーザの空間的配置、ロボットおよび/またはユーザのジェスチャまたはジェスチャの組み合わせ、ロボットの環境内で検知された事象、ローカル天候、地理的位置、日付、時間およびそれらの組合せを含むリストから選択されるパラメータを含む。

【0014】

一つの発展形態では、対話モード実行規則はインターネットから動的に取り出されるおよび/またはユーザ構成可能である。

【0015】

一つの発展形態では、1つまたは複数の対話コンテンツまたはその選択は、1つまたは複数の単語のブラックリスト、1つまたは複数の単語のホワイトリストおよび/または対

10

20

30

40

50

話モード実行規則を含む1つまたは複数のフィルタの適用により緩和される。

【0016】

一つの発展形態では、対話コンテンツおよび/または対話音声スキンは音声命令またはユーザ要求により選択される。

【0017】

一つの発展形態では、本方法はさらに、ユーザとの対話中に欠落対話モードを識別し、前記欠落対話モードを取り出しロボットヘインストールする工程を含む。一つの発展形態では、本方法はさらに、選択された対話モードを行う前にユーザの確認を受信する工程を含む。一つの発展形態では、本方法はさらに、修正対話モードの実行後にユーザのフィードバックを受信する工程を含む。一つの発展形態では、本方法の工程は繰り返され得る（例えば、対話モードはさらに修正され得る）。

10

【0018】

前記コンピュータプログラムが好適なコンピュータ装置またはロボット装置上で行われると本方法の1つまたは複数の工程を実行するための命令を含むコンピュータプログラムが開示される。本方法の1つまたは複数の工程を行うようにされた手段を含むシステムが開示される。

【0019】

コンパニオンロボットは通常、多モードである。音声相互作用は、パーソナルコンピュータとの対比でロボットを特徴付ける運動とその傾きとに沿った、ユーザとの相互作用のクリティカル部分を構成する。ユーザとロボットとの対話は、相互作用を強化または個人化し、最終的にユーザ経験を改善し得る。一実施形態では、ロボットは、その対話モードの適応化により自身を現在の認識状況に順応させる。ロボットは例えば、外国人に対し「ミスター」と言うことができる、または過去にそうすることが許されていれば人の姓を使用し、ユーザおよび/または状況に応じて多少フォーマルに話すことができる。特定の単語もまた、（例えば）ユーザ、履歴、フィードバック、気分、場所、日時に応じてフィルタ処理され得る。人が文章を理解しない場合、ロボットは、そうするように要求されればまたは主導的に、ゆっくりおよび/または同義語により繰り返し得る。ロボットはまた、ユーザの嗜好（語彙によっては多少速く話す）を学習し、ユーザの気分を改善し得る。

20

【0020】

有利には、ロボットは、新しい言語拡張を実施し、各ロボットをユニークにし、肯定的感情を起こさせ、従って人間とロボットとの関係を強化し得る。

30

【0021】

有利には、いくつかの実施形態によると、マン-マシン相互関係が能動的であり、もはや受動的ではない：すなわち、ロボットは、人間的観点から、いくつかの主導性を採り得る（例えば、ロボットは例えば曖昧性除去目的のために質問をし得る）。さらに、個人化されたやり方またはそうでなければ適切なやり方で表わされた適応化対話コンテンツまたはパターンにより、マン-マシン相互関係はさらに最適化される。

【0022】

有利には、相互作用の会話モードは、ユーザとのより「親密」な「関係」、少なくともより「自然な」相互作用を可能にする。このより良いユーザ経験は、機械による人間ユーザの「理解」の向上につながる可能性が高い。関連音声スキンおよび/または対話文章により示唆および強化された機械との付随「近親性」は、ユーザからのおよびユーザに関するデータの収集を容易にし得る。ユーザとロボットの両方は、より「表現性に富み」得る。用語「表現性」は、「マン-マシン相互関係が（より）自然であるのでユーザはより多くのデータをロボットへ伝達しロボットはユーザに関するより多くのデータを知り格納し、好循環に相互作用をさらに豊かにし得る」ということを指す。これはパーソナルコンピュータには当てはまらない。タブレットは例えばクイズまたは質問の形式でまたはスピーチ合成により「質問」を試み得る。しかし、タブレットは自力で（自律的に）動き、物体を移動し人間に追従する「コンパニオン」とはみなされないで、残留バイアスが残ることになる。捕捉され得るデータの量はコンパニオンロボットと比較して小さくなる。コン

40

50

パニオンロボットが楽しいまたはそうでなければ適切な音声スキンまたは対話パターンを使用することができるということが、データを捕捉するこの能力を強化する。

【0023】

ユーザに関し能動的または受動的に収集された情報（例えば、ユーザプロフィールまたはユーザ宣言嗜好）が、立ち上げ条件の入力として使用され得る（例えば、音声スキンまたは対話パターンは、ユーザが「*Bienvenue chez les Ch'tis*」を愛する場合だけ立ち上げられるべきである）。機械学習の機構が行われ得る、すなわち本システムにより開始または行われる音声スキンまたは対話パターンはユーザに関し学習されたものに応じて進化する。

【0024】

次に一例として本発明のいくつかの実施形態について、同様な参照子が同様な要素を表す添付図面を参照し説明する。

【図面の簡単な説明】

【0025】

【図1】本発明のグローバル技術環境を示す。

【図2】本方法の実施形態のいくつかの態様を詳述する。

【発明を実施するための形態】

【0026】

「対話」は予め作成された文章を含む。対話は質問への応答を含む所定文章の集合である。複数の質問に対する予想応答が対話を構成する。

【0027】

「対話モード」は、両方の実体（「対話パターン」または「対話コンテンツ」）に影響を与え計画文章の（「音声スキン」または「音声レンダリング」）を形成する1つまたは複数の修正操作を含む。換言すれば、「対話モード」は実体面（例えばメッセージにより伝達される事実コンテンツまたは情報）、および形式面（例えば話された言語の表現性または感情またはトーン）に関連付けられる。対話モードはダウンロード可能ソフトウェアプログラムの形式で実装され得、前記プログラムは、好適なロボット装置上で実行されると前記ロボット装置に特定の物理的行為を行わせる（プログラム対話モード（対話コンテンツおよび/または音声スキン）を実行することを含む）命令を含む。ソフトウェアプログラムは「拡張モジュール」または「プラグイン」または「アドオン」として提供され得る。追加の対話モードが、ロボットの初期設定対話コンテンツおよび音声スキンと組み合わせられ得る、またはそれに追加され得る、またはそれと置換され得る。一実施形態では、対話モードは、ロボットにインストールされた他のソフトウェアアプリケーションのサービスと呼ばれ得る。例えば、天候アプリケーションは、ある状況（例えば満月）においてダーク・ベイダーの声を使用し得る。対話モードおよび/または関連実行規則は、ネットワークを介しアクセスされてもよいしローカルにアクセスされてもよい。いくつかの実施形態では、対話モードおよび/または関連実行規則は、ネットワークおよびリモート知識ベースへのアクセスにより補完または補足される。

【0028】

「対話コンテンツ」または「対話パターン」または「対話トピック」は、例えばあるテーマまたはトピックまたは関心領域に関する質問と（予想または期待または可能）回答とに対応する所定文章の集合を指す（が、一般的範囲の文章が想定され得るので、必ずしもそうではない）。構文修正は、ロボットにインストールされた既存対話コンテンツの実体を修正し得る（例えば、「超」などの最上級の挿入、単語の置換等）。対話コンテンツまたはパターンは検閲されるべきいくつかの単語を得ることができる（例えば、所定単語の使用が、2値または当該確率または閾値であれば、禁止され得る）、またはいくつかの他の単語が許容され得る、またはいくつかの単語の使用が奨励され得る（バイアス）。対話コンテンツまたはパターンは特に、実質的コンテンツおよび他の文化的関連性を含み得る（または、修正される場合は追加し得る）。単語の選択は、状況に依存し、隠喩または文化的関連性を含み得る。従って、対話は1つまたは複数の対話コンテンツ（文章で構成さ

10

20

30

40

50

れた予め作成された対話)を含み得る。例えば、その単なる商業上の名称の代わりに、ゲームアプリケーションは、「鳥と緑色豚によるゲーム」または「ターゲットに鳥を投げなければならないゲーム」などとしてロボットにより知らされ得る。文章、潜在的質問および回答で構成されるこれらのメタ記述が対話コンテンツを構成する。このような対話コンテンツにより、ロボットがユーザと会話できるようにする。例えば、ユーザが「鳥と遊びたい」と要求すれば、ロボットはさらに、「現実鳥と遊びたいかまたは仮想鳥と遊びたいか?」と尋ね得る。ユーザが「仮想鳥と」と応答すれば、ロボットは「つまりゲームをしたいのか?!」と確認のために尋ね得る。ユーザが「はい」と応答すれば、ロボットは依然としてさらに確認を求める(例えば「緑色豚に鳥を投げなければならないゲームがある」)。

10

## 【0029】

「対話スキン」または「音声スキン」は音声レンダリング修正を指す。このような音声レンダリング修正は「形式」(例えば周波数、速度、ピッチおよびトーン)に影響を与える。換言すれば、対話スキンの適用は、根底にある予め作成された文章を修正すること無くロボットの表現性を抜本的に変更し得る。ロボットとのスピーチ相互作用の修正の影響は、様々なレベルにおいて(コンテンツ的に(実体)および/または形式(トーン等)において)評価され得る。音声スキンは、いくつかの音声を模倣することにつながるパラメータを含み得る。スピーチ合成を管理するために様々な音声パラメータを取り扱ひ得る。音声パラメータは、周波数(ロボットがより明瞭にまたはより太く話せば判断される)、速度(ロボットがどれくらい速くまたは遅く話すか)、トーン(例えば、俳優シルベスタースタローンとマスター・ヨーダ(Master Yoda)キャラクタとが同じ速度および周波数で話せば、彼らは同じトーンを有しない)を含む。一実施形態では、ユーザは、彼のコンパニオンロボットがマスター・ヨーダまたはシルベスタースタローンのように話すことを求め得る。音声パラメータを所定パラメータで適切に修正することにより、近似結果が得られる。「オンザフライ」模倣は、実現性があり(音声抜粋を記録し、パラメータを導出し、適切な修正を適用することにより)、ロボットが1人または複数のユーザを模倣できるようにすることができる。一実施形態では、複数の音声スキンが組み合わせられ得る。いくつかの音声スキンは組み合わせには不適合であり得る(互に排他的であり得る)。他のいくつかはある程度は組み合わせられ得る。他のいくつかは追加的であり得る。

20

30

## 【0030】

「対話実行規則」は、1つまたは複数の音声スキンおよび/または対話コンテンツまたはパターンの適用を規定する実行規則を指す。「実行規則」は、ロボットが言い得る語句を適合化させる(語彙、文章の前または後にいくつかの表現を追加すること等)スクリプト、プログラムコードまたはそうでなければ論理式または論理規則を含み得る。(例えばロボットは質問に答えようとするまたは状況を明確にしようとするので)ロボットが人間ユーザに何かを言おうとするときはいつも、ロボットの計画文章が1つまたはいくつかの対話実行スキン規則に一致すれば、文章はこれらの規則に従って修正され、その後、ロボットは修正された文章を言うことになる。一実施形態では、1つまたは複数の対話実行規則が1つまたは複数の文章(すなわち、ロボットにより言われるように計画された文章)に適用され得る。一実施形態では、前記規則は、ロボットにより言われる各文章に適用され得る。一実施形態では、規則は、文章のサブセット(例えば所定単語または表現を含む)に適用され得る。対話実行規則は予め定義され得る。対話実行規則はまた、インターネットから動的に取り出され得る。いくつかの規則は追加的であり得、他のいくつかは互に排他的であり得る。例えば、実行規則は年齢制限を含み得る(例えば符号化し得る)。蓄積実行規則が使用または適用され得る。例えば、特定の音声スキンは、12歳を越えるおよび/またはいくつかの状況(当日の時刻、聴衆内の測定された感情等)に従うユーザの前で承認され得る。いくつかの実行規則はユーザ(例えばペアレンタルコントロール)により構成可能であり得る。

40

## 【0031】

50

一例として、文章「私は今ダンスをすることができる」は標準所定句（ロボットメモリ内に書き込まれた）に対応する。「私は今ダンスをすることができる h e i n b i l o u t e」は「C h ' t i s」と呼ばれる対話パターンの適用後にロボットにより表現される語句に対応する。音声レンダリングまたは音声スキン「C h ' t i s」はさらに（任意選択的に）、適切な音声変調を提供し得る。形式と実体は以下のように多様に修正され得る：特定のアクセントまたはイントネーション（例えば、北フランス形式）が追加され得る、ロボットにより使用される語彙は豊富にされ得る、新しい会話トピック（例えば質問および回答のモデル）が追加され得る。

#### 【 0 0 3 2 】

対話モード（対話コンテンツおよび/または対話スキン）は、ソフトウェア編集者により定義またはプログラムされ得るソフトウェアパッケージ内に実装され得る。このようなソフトウェアは修正可能であってもなくてもよい。換言すれば、対話モード（例えば音声スキン）は完全に判断され得る（例えば、いかなるさらなるパラメータ化も公式に許容され得ない）。代替的に、対話モードは部分的にだけ判断され得る。例えば、いくつかの（例えば有限数の）ローカルパラメータがエンドユーザの管理下で残り得るが、設定の大部分は変更され得ない（例えば音声スキンの総合的健康性を維持するために）。

#### 【 0 0 3 3 】

換言すれば、文字通りの意味を越えるソフトウェアアプリケーション（好適なコンピュータ装置上で実行されると1つまたは複数の工程を行い得るコンピュータプログラムコード）は、対話コンテンツ（例えば予想される質問に対する応答を含む所定の文章の集合）、および/または対話スキン（例えば、対話コンテンツの上のプログラミング、すなわち環境に応じた適応化などの実行規則、頭の運動との同期、もしあれば照明の活性化等）およびそれらの組合せ（例えばダンスをしながらの対話）であり得る（または関連付けられ得る）。ソフトウェアアプリケーションは相互依存し得る。多モード出力の結果として、ソフトウェアアプリケーションはさらに組み合わせられ得る（出力レベルでまたは下位レベルで、例えば変数またはパラメータまたはスクリプトはソフトウェアアプリケーション間で共有または修正され得る）。例えば、ロボットは、外の寒さを象徴するジェスチャの組み合わせにより、話された結果「外は - 10 である」を伴い得る。

#### 【 0 0 3 4 】

ソフトウェアアプリケーションは有利には、対話インタフェースを介し（すなわちユーザとの（「自然な」）対話の行為の過程に）ユーザへ呈示され得る。換言すれば、対話システムは、ユーザが1つまたは複数のアプリケーションを立ち上げるまたは行うことができるための「ボトルネック」として働き得る。

#### 【 0 0 3 5 】

図1は、本発明のグローバルおよび技術的環境を示す。ロボット130はセンサとアクチュエータを含む。論理または「マインド」100はロボット内に実装されるまたはそれに関連付けられ（例えば遠隔的に）、ソフトウェア110とハードウェア部品120の集合を含む。ロボット130は1人または複数のユーザ150と相互作用する（1つまたは複数の対話セッションを含む相互または双方向通信140により）。前記1人または複数のユーザは、接続装置（サーバのクラウドとおよび/または他のロボットまたは接続対象の一群などと通信する）であり得る他の計算装置160（例えばウェアラブルコンピュータまたはスマートフォンなどのパーソナルコンピュータまたはタブレット）にアクセスし得る。特に、接続装置はウェアラブルコンピュータ（例えば、時計、眼鏡、没入型ヘルメット等）であり得る。

#### 【 0 0 3 6 】

同図の特定ロボット130は、本発明が実施され得るヒューマノイドロボットの一例と考えられる。同図のロボットの下肢は、歩行には機能的でないが、置かれた表面上で転回するそのベース上で任意の方向に移動し得る。本発明は、歩行に適したロボットにおいて容易に実施され得る。

#### 【 0 0 3 7 】

10

20

30

40

50

本発明のいくつかの実施形態では、ロボットは様々な種類のセンサを含み得る。それらのいくつかは、ロボットの位置および運動を制御するために使用される。これは例えば、ロボットの胴内に配置された慣性ユニットのものであり、3軸ジャイロメータおよび3軸加速度計を含む。ロボットはまた、ロボットの額上（頂部と底部）に2つの2DカラーRGBカメラを含み得る。3Dセンサはまた、ロボットの眼の背後に含まれ得る。ロボットはまた、その環境内の物体/生物に対するその相対位置を感知することができるようにレーザー光線発生器を任意選択的に（例えば頭内とベース内に）含み得る。ロボットはまた、その環境内の音を感知することができるマイクロホンを含み得る。本発明のロボットはまた、その環境内の物体/人間までの距離を測定するために、恐らくそのベースの前部および後部に位置するソナーセンサを含み得る。ロボットはまた、人間との相互作用を可能にするためにその頭およびその手上に触覚センサを含み得る。ロボットはまた、その経路上で遭遇する障害を感知するためにそのベース上にバンパを含み得る。その感情を翻訳しその環境内の人間とコミュニケーションするために、本発明のロボットはまた、例えばその眼、耳内におよびその肩の上にLEDをそして拡声器（例えばその耳内に配置される）を含み得る。ロボットは、様々なネットワーク（3G、4G/LTE、Wifi、BLE、メッシュ等）を介し基地局と、他の接続装置と、または他のロボットと通信し得る。ロボットは電池またはエネルギー源を含む。ロボットは、ロボットが含む電池のタイプに適した充電ステーションにアクセスし得る。ロボットの位置/運動は、センサの測定という観点では、各肢により定義されたチェーンと各肢の端において定義されたエフェクタとを活性化するアルゴリズムを使用することによりそのモータにより制御される。

10

20

#### 【0038】

特定実施形態では、ロボットは、その環境へメッセージ（音声、映像、ウェブページ）を伝達し得るタブレットを埋め込み得る、またはタブレットの触覚インタフェースを介しユーザからエントリを受信し得る。別の実施形態では、ロボットはスクリーンを埋め込まないまたは提供しないが、データまたは情報がロボットの近傍の表面上に投射され得る映像プロジェクタを有する。前記表面は平ら（例えば床）であってもなくてもよい（例えば、投射面の変形はほぼ平らな投射を得るために補償され得る）。両方の実施形態（スクリーンを有するおよび/またはプロジェクタを有する）では、本発明の実施形態は有効なままである：すなわち、請求相互作用モデルは視覚的相互作用手段により補足または補完されるだけである。いずれにせよ、グラフィック手段が故障しているまたは意図的に非活性化されていたとしても、相互作用の会話モードは存続する。

30

#### 【0039】

一実施形態では、ロボットはこのようなグラフィックユーザインタフェース手段を含まない。既存ヒューマノイドロボットは通常、高度スピーチ能力を備えるがGUIを備えない。増加するユーザのコミュニティは恐らく、選択および/または必要性（実際的狀況のために若者、障害者等）により、ロボットと通信するために、グラフィック手段（例えばタブレット、スマートフォン）を補足物としてすら使用しないことになる。

#### 【0040】

ソフトウェアの集合110は（非網羅的に）、「抽出器」111、「活動示唆」112、「マインド優先順位付け」113、「パッケージ管理者」114、「ユーザ履歴データ」115、「集中自律的活動」116、「集中対話トピック」117、および「健康監視サービス」118を含む互いに相互作用するソフトウェアモジュールまたはオブジェクトまたはソフトウェアコード部品を含む。

40

#### 【0041】

「エクストラサービス」111は通常、ロボットの内部または外部にある何かを感知または認識し、短期データをロボットのメモリ中に提供する。抽出器サービスはロボットセンサから入力測定結果を受信し、これらのセンサ測定結果は、ロボットの位置、その環境内の物体/人間の識別、前記物体/人間の距離、人間により発声された単語またはその感情に関連する関連データを抽出するように前処理される。抽出器サービスは特に、顔認識、人認知、係合ゾーン、ウエイピング検出、微笑検出、凝視検出、感情検出、音声分析

50

、スピーチ認識、音声定位、運動検出、パノラマコンパス、ロボット姿勢、ロボット健康診断、電池、QRコード（登録商標）取り扱い、ホームオートメーション、種族、時間およびスケジュールを含む。

【0042】

「アクチュエータサービス」はロボット130に行為を物理的に行わせるまたは行為を行わせる。運動追跡器、LED、行動管理者は「アクチュエータサービス」である。

【0043】

「データサービス」は長期的に格納されたデータを提供する。データサービスの例は、ユーザデータとロボットで行ったもののその履歴とを格納するユーザセッションサービス115と、ロボットにより行われる手順のスケラブルストレージにそれらの高レベル定義、立ち上げ条件およびタグを与えるパッケージ管理者サービス114である。「パッケージ管理者」は特に、活動と対話のスケラブルストレージとマニフェストとを提供する。「マニフェスト」は、立ち上げ条件、タグおよび高レベル記述などのメタデータを含む。

10

【0044】

「マインドサービス」（例えばサービスマインド優先順位付け113）は、行為を開始しているときにロボットの中央「マインド」により制御されるものである。「マインドサービス」は、「アクチュエータサービス」130、「抽出器サービス」111および「データサービス」115を繋ぎ合わせる。ベーシックアウェアネス（Basic Awareness）は「マインドサービス」である。ベーシックアウェアネスは、運動サービスに移動するように告げるための人認識、運動検出および音声定位などの「エクストラサービス」を支持する。「マインド」113は、状況に基づくベーシックアウェアネスの行動を構成する。またある時には、ベーシックアウェアネスは、ひとりで行動するまたはランニング活動（Running Activity）により構成されるかのいずれである。

20

【0045】

「自律的生命（autonomous life）」はマインドサービスである。「自律的生命」は行動活動を行う。状況のコンテキストに基づき、マインドは、何の活動に焦点を合わすべきかを自律的生命に伝え得る（「集中自律的活動116」）。マニフェスト内のメタデータはこの情報をマインド内に結びつける。いかなる活動もオペレーティングシステムAPIのうちの1つまたは複数へアクセスし得る。活動はまた、何の活動に焦点を合わすべきかを自律的生命に直接伝えてもよいし、何のトピックに焦点を合わすべきかを対話サービスに伝えてもよい。

30

【0046】

「対話」サービスはマインドサービスとして構成され得る。「対話」サービスは、スピーチ認識抽出器を支持し、「アニメ化スピーチアクチュエータサービス」を使用して話すことができる。状況のコンテキストに基づき、マインドは、何のトピック（「対話トピック」）に焦点を合わすべきかを対話サービスに伝え得る。「対話」サービスはまた、会話を管理するためのアルゴリズムを有し、通常は、ひとりで行動する。対話サービスの1つの部品は「集中対話トピック」サービス117であり得る。対話トピックは、いつでも、異なる活動または対話トピックへ焦点を切り替える（またはそれを実行するまたは立ち上げる）ようにマインドにプログラマ的に告げ得る。対話トピックを判断する可能な方法の一例は以下のものである：対話トピックまたは活動の立ち上げ条件が真または偽になった瞬間に、全ての可能な活動または対話トピックのリストがさしあたりマインドへ送信される、リストは活動優先順位付けに従ってフィルタ処理される、リスト順番がランダム化される、「ユニーク」でありかつそれほど頻りに開始されなかった活動または対話トピックへ優先順位を与えるためにリストがソート（または採点）される、このリスト内の最上位対話トピックまたは活動が、実行された以前の活動と同じ活動ではないということを確認するための特別の照査。リストは再び、ユーザの嗜好に従ってソートされフィルタ処理され得る。

40

50

## 【 0 0 4 7 】

ロボットは「健康監視」サービス 1 1 8 を実施し得る。このようなサービスは、ロボットの様々な優先度を精査または制御または調節するデーモンまたは「ウォッチドッグ」として働き得る。このようなサービスは、ロボットの内部部品の状態を（連続的、間欠的または定期的に）監視し得るとともにハードウェア障害を測定または予期または予測または補正し得る。一つの発展形態では、ロボットの一団が監視される（例えばインストールベースで）。埋め込みサービスは故障状況を連続的に検知し、それを「クラウド」サービスと同期させ得る（例えば毎分毎に）。

## 【 0 0 4 8 】

ハードウェア部品 1 2 0 は、処理手段 1 2 1、メモリ手段 1 2 2、入出力 I / O 手段 1 2 3、マストレージ手段 1 2 4 およびネットワークアクセス手段 1 2 5 を含み、前記手段は互いに相互作用する（キャッシング、スワッピング、分散計算、負荷平衡等）。処理手段 1 2 1 は CPU（マルチコアまたはメニーコア（many core））または FPGA であり得る。メモリ手段 1 2 2 は、フラッシュメモリまたはランダムアクセスメモリの 1 つまたは複数を含む。I / O 手段 1 2 3 は、スクリーン（例えばタッチスクリーン）、ライトまたは LED、触覚フィードバック、バーチャルキーボード、マウス、トラックボール、ジョイスティックまたはプロジェクタ（レーザプロジェクタを含む）のうちの 1 つまたは複数を含み得る。ストレージ手段 1 2 4 は、ハードディスクドライブまたは SSD の 1 つまたは複数を含み得る。ネットワークアクセス手段は、3 G、4 G / LTE、Wi fi、BLE またはメッシュネットワークなどの 1 つまたは複数のネットワークへのアクセスを提供し得る。ネットワークトラフィックは暗号化され得る（例えばトンネル、SSL 等）。

## 【 0 0 4 9 】

一実施形態では、計算資源（計算機、メモリ、I / O 手段、ストレージおよび接続性）は例えばローカル資源（ロボット自身に利用可能な）に対する補足として遠隔的にアクセスされ得る。例えば、別の CPU ユニットが音声認識計算タスクのクラウドを介しアクセスされ得る。計算資源はまた、共有され得る。特に、複数のロボットが資源を共有し得る。ロボット近傍の接続装置はまた、例えばセキュアプロトコルを介し資源をある程度共有し得る。表示手段も共有され得る。例えば、テレビはロボットにより別のディスプレイとして使用され得る。

## 【 0 0 5 0 】

図 2 は、本方法の実施形態のいくつかの態様を詳述する。ロボット 1 3 0 は人間ユーザ 1 5 0 と相互作用する（例えば対話、ジェスチャ、命令をやり取りする）。相互作用 1 4 0 の一部は、文章（回答、質問、命令、確証、コメント等）を含む対話である。ロボットは通常、そのデフォルト標準音声スキン（形式）を使用し、標準および所定対話コンテンツ（実体）を出力する。例えば、ロボットは対話文章 1 4 1 を述べる。いくつかのパラメータ（ユーザ要求または環境パラメータ）に依存して、ロボットは、別の音声スキンおよび / または別の対話コンテンツ（例えば 1 4 2）へ切り替わり得る。ロボットはまた、初期またはデフォルト音声へ切り替え復帰し得る。詳細には、デフォルト音声スキンおよび対話コンテンツ 2 0 0（または初期 / 修正音声スキンおよび / または修正対話コンテンツ）で始まり、対話実行規則 2 2 0 は、対話が修正されなければならないかどうかまたは対話がどの程度修正されなければならないかを判断する。

## 【 0 0 5 1 】

対話実行規則 2 2 0 は、例えば、ユーザ要求 2 2 1 によりおよび / または認識された環境 2 2 2 により影響または判断される（例えば、センサまたはロボットにより判断され、抽出器によりまたはロボットのマインド内に実装されたロジックに関する上記実施形態に従ってフィルタ処理される）。例えば、環境パラメータは、近傍の 1 人または複数のユーザ（子供、大人）の年齢グループ、近傍の 1 人または複数のユーザの性別、近傍のユーザの総数、現在位置、現在の日時、1 人または複数のユーザの現在の気分（例えば、微笑む、笑う、叫ぶ等）を含む。ユーザ要求 2 2 1 は、新しい対話モードの活性化の「オンデマ

10

20

30

40

50

ンド」モードに対応する（例えば、ユーザは「今ダーク・ベイダーを真似よ」と言うことができる）。認識された環境を介した判断は、新しい対話モードの活性化の「自動トリガ」モードを明確に示す。ロボットは、対話モードの1つまたは複数のパラメータを積極的に活性化または非活性化し得る（音声スキンを弱めるまたは誇張する、対話コンテンツを適合化する等）。活性化（または非活性化）の2つのモードは組み合わせられ得る、すなわち、新しい対話モードのトリガは、ユーザ要求によりおよび環境により部分的に判断され得る。例えば、ユーザ要求に応じて、環境パラメータは対話モードの変化を確認または禁止し得る。代替的に、自動スイッチは、ユーザによる確認または承認が活性化されることを必要とし得る。対話モードは通常、いつでも活性化または非活性化され得る。一実施形態では、ロボットが複数の対話モードを行きつ戻りつ使用し得ると楽しくなると期待できる。任意選択的に、変更回数の制限が実施され得る（ユーザの没頭を回避するために、またはロボットが正気でないという印象を与えるために）。

10

#### 【0052】

依然として独立に取り扱われる対話実行規則220が実体および/または形式に作用する。規則は、音声レンダリング変更230を制御するだけでなく、新しい対話コンテンツ210のローディング（例えば、周知文章を表現することにより劇場映画への参照）を判断し得る。いかなる対話コンテンツも選択されなくても（211）、新しい音声スキンが適用され得る（230）。新または修正対話コンテンツがまた、新しい音声スキン231無しに適用され得る。判断されたパラメータが対話モードに適用される（または、新しい対話モードは、ロボットにより使用される現在の対話モードにロードされるまたはそれと置換される）。例えば、「Ch'tis」と呼ばれる対話モードが適用され、文章142が朗読される。全体としてまたは部分的に対話コンテンツおよび/または音声スキンパラメータはインターネットまたはクラウド223から取り出され得るということが観測される。

20

#### 【0053】

対話モードの例（対話コンテンツ、特に音声スキンの）について次に説明する。

#### 【0054】

一実施形態では、「ユーティリティ」対話モード（例えば対話コンテンツの修正）は、ロボットが所与の文章を様々なやり方と言うことができるように、1つまたは複数の辞書（またはソーラス）を使用できるようにする。例えば、同義語が使用され得る。有利には、このような実施形態は、ロボットが単語を繰り返すのを回避する。言語学者（予め作成された対話文章を編集する）は、ロボットが多くの事項について話すことができるようにするために、多くの対話コンテンツまたはトピックを書くことができる。同義語の使用は、人間が様々な単語を使用する際に纏めてそうするように、ロボットの表現の多様性を増加する。対話モードは対話文章を異なるようにし得る。例えば、「今日は天気が良い」と繰り返す代わりに、ロボットは「今日は天気が素晴らしい」と言うことができる。

30

#### 【0055】

一実施形態では、「ローカル」適応化は、対話コンテンツおよび/または音声スキンを、地理的位置を含む複数のパラメータに応じてカスタム化または個人化できるようにする。例えば、いくつかの地理的領域（例えばロボットが商業化される）のいくつかの言語学的特徴は適切な音声スキンのアプリケーションにより取り扱われ得る。このような実施形態では、いくつかの単語の朗読は、例えば地理的位置に基づき適応化され得る。音声スキンのトリガ条件は例えば地理的位置パラメータを含み得る。実際、フランスの南部において商業化されるロボットは南部フランスアクセントを自動的にロードし得、フランスの北部の北部アクセントについても同様である。慣用表現もまた適用され得る。

40

#### 【0056】

一実施形態では、「教育的」対話モードが実施または実行され得る。一般的に言えば、実質的コンテンツの追加は対話モードのアプリケーションの一部であり得るので、いくつかの教育的範囲が想定され得る。このようなモードでは、（例えば、前提条件、目的の定義、教育的コンテンツ従って同化の検証工程を含む）教育的対話モードが実施され得る。

50

ユーザは、コンパニオンロボットにより外国語を学習することができる。ユーザはまた、母国語のいくつかの様相を発見し得る。例えば、様々な言語スタイルを若いユーザに教えることができる（仏語「soutenu」、仏語「Verlan」、仏語「argot」等）。対話モードはまた特定の専門的隠語（医療、弁護士などの）を実装することができる。

【0057】

一実施形態では、「楽しみ（fun）」対話モードが実装され得る。ロボットは例えば、最近公開された劇場映画を真似るまたは参照することができる。例えば、音声スキンは、「ダーク・ベイダー」または「マスター・ヨーダ」スキンであり得る。（ヨーダキャラクターは動詞と主語を頻繁に並べ替える、ダーク・ベイダーは文章の終わりに「私はあなたのお父さんである」と追加し音を立てて呼吸し得る）。

10

【0058】

一つの発展形態では、映画館の公開は、ロボットにロードされると前記ロボットに特定対話モードを実施させるダウンロード可能プログラムに関連付けられる（対話コンテンツおよび/または音声スキン表現性という意味合いで）。例えば、「Bienvenue chez les Ch'tis」のような映画館の公開後、対応対話モードは「楽しいもの（goodies）」または派生商品として利用可能かもしれない。一実施形態では、DVDジャケットの裏または映画チケットの上の利用可能なQRコード（登録商標）が画像取得手段により読み取られ得、対応対話モードを実施するそれらの対応ソフトウェアプログラムがダウンロードされ、さらにインストールされ得る。任意選択的に、このようなソフトウェアプログラムの効果の期間は限定され得る（例えば2週間）。別の実施形態では、有名スターまたは有名人またはヒーローがデジタル対話モードカウンターパートを有し得る。例えば、クロードフランソワ（フランスで有名な歌手）のファンは、ロボット内に対応音声スキン（例えば声帯模写、イントネーション）および/または対話コンテンツ（例えば複製、引用、インタビューの正確な複製または復元物等）を実装し得る。多モードオブジェクトとして、上記対話モードに加えて（および任意選択的に）、ロボットは例えば、クロードフランソワのように振り付けまたはダンスをすることができる、またはもっぱら同歌手の過去の記録インタビューの実際の抜粋により応答することができる（適切な場合著作権面に関して）。

20

【0059】

一実施形態では、ロボットは、ユーザとの相互作用から長期的同化に達し得る。例えば、個人化および持続的対話モデルは、所与のロボットに次第に関連付けられ得る。長期的習性の選択および持続は、ロボットに関連付けられた「キャラクター」の一意性を規定するのを助け得る。例えば、「vin de diouss」のようないくつかの表現は、スキンの特定態様が恒久的に実施されるように何人かのユーザに沿った十分に高い評価を提示し得る。従って、スキン毎に、ロボットの「個性」は、様々な対話モード（特に音声スキン）の様々な特殊性を混合することにより強化され得る。

30

【0060】

言語相互作用は、ロボットとユーザの両方の学習曲線を意味し得る。このような相互作用は、参加者のうちの1人がロボットでも「社会的相互作用」と名付けられ得る。激励（確認）により中断される繰り返しおよび反復言語相互作用は例えば、学習活動を容易にし得る。対話相互作用は一般的には、「利用」段階と「学習」段階との分離を低減する。コンパニオンロボットは、例えば人間の言語表現を再使用することによりおよび/または同様な音声速度により「教育者」の真似をしようとし得る。ロボットとのより多くの相互作用は通常、知識ベースが豊富にされ相互査証されるのでより多くの適切な相互作用を意味する。

40

【0061】

対話モードのトリガの例（例えば実行規則による、例えば対話コンテンツおよび/または対話スキンの活性化または非活性化）について次に説明する。

【0062】

50

対話モード（例えば音声スキン）は1つまたは複数の実行規則に関連付けられ得る。対話モードはこのような実行規則に従ってトリガ（活性化または非活性化）され得る。様々な実施形態について以下に説明する。人間ユーザとロボット間の文章を含む対話中に対話コンテンツと対話音声スキンを含む対話モードの立ち上げまたは実行をトリガするいくつかのやり方がある。1つまたは複数の対話モードの立ち上げまたは実行をトリガするこれらの様々なやり方（特に以下説明される）は独立し得、さらに互いに組み合わせられ得る。

**【0063】**

一実施形態では、1つまたは複数のソフトウェアアプリケーションの立ち上げまたは実行はユーザとの対話（ユーザとの相互作用）中にトリガされる。音声信号が捕捉され、任意選択的にフィルタ処理され、強化され、スピーチツーテキスト（speech-to-text）操作が行われ（ロボット上でローカルにおよび/またはクラウド上で遠隔的に）、得られたテキストが解析され、1つまたは複数の比較が前記パターンにより行われる。1つまたは複数が一致すると、任意選択的に閾値により、1つまたは複数の対話モードが、ロボット上にインストールされたものの中から選択される。その結果、1つまたは複数の対話モードが実行される。

10

**【0064】**

一実施形態では、対話モードの立ち上げは完全に自動である、すなわちユーザの同意または確認無しに行われる。一つの発展形態では、ユーザまたはスーパーユーザ（例えば親）が、対話モードの実行を中断または中止または終結または終了し得る。別の実施形態では、対話モードの立ち上げは、ユーザによる明示的確認を必要とする。ロボットは、いくつかの対話モードを立ち上げる意図を宣言し得るが、継続する前に確認を待つことになる。例えば、ロボットは「私はダーク・ベイダーのように話すことを提案する」と宣言することができ、ユーザは依然として「今はダメ」と応答し得る。

20

**【0065】**

一実施形態では、対話モードは立ち上げ条件またはパラメータに依存して実行され得る。これらの条件またはパラメータは事実または規則またはその両方（事実に基づく規則）であり得る。これらの事実は例えば、ユーザのタイプまたはカテゴリーと、1つまたは複数の環境値（例えば現在のローカル天気、日時、検出された感情、ユーザの数等）により特徴付けられた現在のコンテキストまたは状況または環境とを含む。実行規則は簡単な規則から複雑な規則まで及ぶ。実行規則は条件付きであり得る。例えば、一実施形態では、対話モードの実行を認証または許容するために複数の規則が同時に満足されなければならない。別の実施形態では、複数の規則が連続的に満足されなければならない（例えば一定の順番でおよび/または時限または閾値により）。いくつかの実行規則は予め定義され得る。いくつかの他の実行規則は動的に定義され得る（例えば、いくつかの規則はインターネットから取り出され得る）。

30

**【0066】**

一実施形態では、実行規則は簡単な規則であり得る。例えば、実行規則は年齢制限を含み得る（例えば符号化し得る）。別の実施形態では、複数の実行規則が累積的に使用または適用され得る。例えば、特定の音声スキンは、12歳を越える年齢のユーザの前でおよび/またはいくつかの状況（当日の時刻、聴衆内で測定された感情等）に従って認証され得る。一実施形態では、対話音声スキンまたは対話コンテンツのアプリケーションは予め定義された事象が検出されるとロボットによりトリガされ、前記事象は、時間判定基準（カレンダー、当日の時刻等）と空間的判定基準（近傍において検知されたユーザの数、前記ユーザのそれぞれの年齢、前記ユーザから認識される感情的態度、例えば微笑むまたは微笑まない）との特定組み合わせを判断する。

40

**【0067】**

一実施形態では、いくつかの実行規則はユーザにより構成可能（例えばペアレンタルコントロール）であり得る。いくつかの実行規則は予め定義され得、他の実行規則は例えばインターネットからおよび/または他のロボットから動的に取り出され得る。対話モード

50

は同じ対話中に活性化または非活性化され得る：すなわち、これらの活性化または非活性化は、動的であり得、例えば認識された環境に様々なやり方で依存し得る。

【0068】

一実施形態では、対話モード（すなわち、対話コンテンツおよび/または対話スキン独立に）は、「タグ」の使用および/または「条件」の使用および/または「前提条件」の使用を含むいくつかのパラメータに依存して活性化または終了され得る。

【0069】

タグは、ロボットが言い得る1つまたはいくつかの表現に関連付けられ得るマーカである。これらのタグは、動的に活性化または非活性化され得、関連表現が保存され得るかどうかを判断し得る。例えば、ロボットが「料理について話したいですか？」と尋ねた後、ユーザが「いいえ、私は、料理に興味がない」（等）と応答すれば、タグ「料理」が非活性化される。ロボットは、識別されたユーザに関連付けられた対話トピックのリストを更新する。ロボットは前記対話トピックを将来回避することになる。

10

【0070】

「条件」および「クロス条件」は、ロボットが所定変数（例えばユーザ嗜好）に応じて述べようとするものを修正できるようにする。例えば、ロボットにより尋ねられた質問「あなたは何歳ですか」に、ユーザは「私は12歳である」と答え得る。この場合、ロボットは、識別されているユーザの年齢値として値12を格納する。後で、ロボットは「今夜あなたは何ををするの」と尋ね得る。ユーザが「何もしない」と応答すれば、ロボットは「夕方活動」に関連付けられた変数としてその応答を格納する。年齢グループと夕方の占有事項が無いことから、ロボットの推論は、夕方遅く「あなたは私と遊びたいですか？」と推測または提案する。

20

【0071】

一実施形態では、トリガ（すなわち対話モードまたは音声スキンまたは対話コンテンツの活性化または非活性化）はコンテキスト（例えば環境、データ、時間、場所等）により駆動され得る。一実施形態では、ロボットは1つまたは複数のユーザ表現を監視し記録し得る。「環境」のような単語を検出すると、ロボットは「環境, 環境, e s t c e j ' a i u n e g u e u l e d ' a t m o s p h e r e ? 」と述べ得る。これは文化的関連性の例である。さらに、対話はまた、さらに詳述された事実および規則により（例えば、所謂「事象」検出により）トリガされ得る。例えば、いくつかの対話モードまたはスキンはいくつかの年齢グループの存在下で許容されないかもしれない。実際、ロボットは例えば、「少なくとも1人のユーザが12歳未満である」と評価し、会話の中で単語「フレンチフライ」を検出し、その後特定の所定スキン（「C h ' t i s 」）をロードし得る。事象の別の例は、ユーザがロボットの頭に触れると発生する。このような場合、特定のスキンがロードされ、活性化され、実行され得る。事象は、空間的配置、ジェスチャまたはジェスチャの組み合わせ、対話のコンテンツ（キーワードまたはキー表現）、年齢グループおよび/または性別の評価、ユーザ嗜好などのパラメータを含み得る。

30

【0072】

一実施形態では、1つまたは複数のアプリケーションが対話中に実行され、1つまたは複数の実行規則の達成（または検証または満足）によりトリガされる。人と機械間の対話が監視され、「パターン」が（例えば）対話フローから連続的に抽出される（「協働対話（c o l l a b o r a t i v e d i a l o g ）」モード）。一実施形態では、スピーチ・フローが受信され連続的に解析される。抽出は、マーカの有無にかかわらずスピーチ・フロー内の音声命令（例えばキー表現）の単なる抽出を越える（「O K G l a s s 、写真を撮って」）。特に、ユーザの言葉または表現が、抽出され、所定条件、タグ、マーカまたはクロス条件と比較または照合される。

40

【0073】

別の実施形態では、1つまたは複数の実行規則が予め定義される。ソフトウェアアプリケーションは、前記対話モードの実行を許容または認証できる実行規則のリストを含むファイルが編集者または発行者により提供される。実行規則は試験される、すなわち、実行

50

規則が満足または許容または検証されれば、1つまたは複数の対話モードが選択され得る。いくつかの規則は満たすべき最小判定基準であり得る。いくつかの他の規則時間は好適な立ち上げまたは実行条件を定義し得る。例えば、最小実行規則は「ユーザが12歳未満であり、午後22時前であれば、ダーク・ベイダー音声スキンが認証される」であり得、好適な規則は「3人のユーザが5m内におり、少なくとも2人のユーザが12歳未満であり、少なくとも1人が微笑んでおり、他の誰も指示に反対しなければ、ダーク・ベイダーの声で冗談を提案する」であり得る。

【0074】

様々な実施形態について説明する。

【0075】

逆行ループと関連性フィードバックについて次に説明する。一実施形態では、人間ユーザの観点からの、所与の対話モード（例えば音声スキン）の全体的成功または失敗が受信および/または定量化され得る。一つの発展形態では、きめの細かい粒度により、ロボットの各言語イニシアチブがユーザにより確認または拒絶され得る（例えばジェスチャと音声命令を組み合わせる複雑な人間行動から生じる明示的承認またはさらには暗黙的承認を検知することにより）。さらに、協働モデルが実施され得る：すなわち、所与の表現の検証または無効化がユーザ/ロボットのコミュニティの中で統計的に行われ得る。例えば、表現「*chavabilloute*」がインストールベースで75%を上回る肯定的フィードバックを受ければ、前記表現はグローバルスケールで検証され得る。反対に、表現「*al'arvoiture*」が少なすぎる肯定的フィードバックを受ければ（または否定的フィードバックを受ければ）、前記表現は、特定の対話モードまたはモデルから恒久的に除去され得る。

【0076】

一つの発展形態では、「スキン化された(*skinned*)」対話に対し人間から与えられる応答が記録され、さらに活用され得る。一実施形態では、上記応答はロボット介入の品質を計量するために使用される。別の実施形態では、これらの応答の実体は、会話モデルをさらに豊かにするために使用され得る。例えば、応答が回帰的に観測されれば、応答は対話モード（すなわち対話コンテンツ）内でさらに再使用される。

【0077】

技術事項に関わるビジネス態様に関し、一実施形態では、対話モード（例えば音声スキン）は電子マーケットを介し配布される。いくつかのスキンはダウンロードされ得る、すなわち、他のいくつかは支払いを必要とし得る。いくつかのスキンは、時間、領域または他のハードウェア要件が制限され得る。

【0078】

コンパニオンロボットが、人間が述べるあらゆる単語を理論的に記録できたとしても、プライバシー面がこのような記録を防止する。機械学習手法を使用することにより、ハイレベルおよび無侵襲特徴がロボットにより依然として取得され得る。例えば、機械学習手法により、再現パターン（語彙のタイプ、好適な表現等）が抽出され得る。同様に、劇場映画からの抽出の制限が副題の解析から発生し得る（マスター・ヨーダの例では、対話コンテンツはこのような解析から判断され得る）。トーンと周波数の面に関し、指導付き学習(*supervised learning*)は、ロボットがいくつかの指定人間を真似できるようにする。例えば、ロボットは話すことを開始することができ、さらに、いくつかのパラメータを修正する（「少しよりはっきりと話す」）ように要求され得る。一実施形態では、実装された音声スキンと実際の音声抜粋との自動比較が行われ、これにより改善フィードバックループの機会を提供し得る。

【0079】

対話コンテンツのオーサリングについて次に説明する。様々なエンティティが対話コンテンツを編集する可能性がある。一実施形態では、オペレータまたはロボットプラットフォームが対話文章をオーサリングすることができる（例えば、言語学者が対話文章を書く）。一実施形態では、対話コンテンツは第三者会社（例えばソフトウェア開発者）により書

10

20

30

40

50

かれ商業化される。一実施形態では、対話モードはロボットのユーザまたは所有者により書かれる。例えば、ソフトウェアツール（「DJスキン」）またはウェブプラットフォームは音声スキンの生成または修正を容易にし得る。ユーザは、新しい音声スキンを提出し、それらをオンラインで編集し、人気のあるものに投票するまたはそれを評価し得る。編集は、所定の音声スキンを混合する工程、および/または音声スキンを微調整するためのいくつかの生成コントロールをユーザへ提供する工程および/または音声または記録された文章またはそれらの組合せをアップロードし共有する工程を含み得る。利用可能スキンは無料または有償ライセンスを伴ってもよいし伴わなくても良い。音声スキンは、例えば様々な音声スキンの様々なアプリケーションを聴き1つまたは複数の好適なアプリケーションを選択し得るロボットのユーザにより選択され得る。別の実施形態では、対話は多様なエンティティにより共同執筆される。一実施形態では、対話は、インストールベースのロボットから、および様々なロボットの所有者であるユーザのまさにその応答から強化される。前記強化は初期対話の補足であり得る。別の実施形態では、また補足または代替案として、対話コンテンツはインターネットコンテンツ上で行われる抽出により書かれる（例えば、指導付きまたは指導無し方法が質問と回答を識別、抽出、使用できるようにする）。有利には、このような実施形態は、対話コンテンツの迅速な改善と分散フィードバックの受信とを可能にする。前記改善は、インストールベース全体に急速に伝達され得る。例えば、閉ループ機械学習を使用することにより、人気のある音声スキンは世界中に伝達され得る。

10

**【0080】**

20

複数のエンティティが対話モードの定義に寄与することができれば（すなわち、対話コンテンツおよび/または音声スキンおよび/または関連実行規則を提供することにより）、最終実施形態はロボットの供給者により制御され得る。規則層の別のコントロールは、対話モードの使用を、変調、フィルタ処理、減衰、増幅、増加、促進、低減、禁止、制限、回避、または禁止し得る。特に、説明したように、対話モードの使用は実行規則により調整され得る：すなわち、ロボットの製造者または供給者はこれらの規則を習得することができる、または部分的に習得することができる。

**【0081】**

一実施形態では、ロボットの供給者は最終音声レンダリング装置（すなわち類推的復元前の最新の点）を制御することができる。換言すれば、選択された対話モードまたはコンテンツまたはスキンの対話の予め作成された文章への適用から生じるロボットの計画音声表現は実効的音声復元の前にフィルタで除去され得る。

30

**【0082】**

ロボットが良くない言葉を朗読しないということを確実にするまたは合理的に確実にするために、認証された単語のホワイトリスト、禁語のブラックリスト、およびグレーリスト（認証され得る、または実際のコンテキストに依存しない単語または表現）が実装され得る。このような場合、対話コンテンツの予め作成された文章への音声スキンの適用の結果がこのようなリストと比較され得る。リストの使用を越える複雑な論理規則がまた使用され得る。成功次第、認証または許容されれば、上記文章は朗読される（および/または、それに応じて修正され得る）。

40

**【0083】**

ロボットにより表現される最終対話への1つまたは複数の対話コンテンツおよび/または音声スキンの適用の別の規制方法はセキュアブート（*secured boot*）方法の使用を含み得る。例えば、特定対話モードを符号化する各ソフトウェアパッケージがハッシュ値（例えば、2進形式のプログラムの）に関連付けられる。ロボットにインストールされたソフトウェアパッケージのハッシュ値の検証により証明された前記プログラムの適正な存在は、検証成功後のロボットの起動（または機能のいくつかの制約）を条件付きで認証し得る。

**【0084】**

対話モード（一般的には、ロボットにインストールされたソフトウェアアプリケーション

50

ン)の実行の別の規則として、ロボットの健康監視サービスが実行優先度を調整し得る。特に、ソフトウェアアプリケーションの実行はこのような「健康監視」サービスを考慮し得る。換言すれば、上位の優先度方式が対話モードを含むソフトウェアアプリケーションの実行をさらに規制し得る。一実施形態では、ロボットはユーザと相互作用しない(すなわち、誰とも相互作用しない)。このような場合、ロボットは自律的タスクを行う、または行い得る。別の実施形態では、ロボットは危険な状態(例えば「セーフガードモード」、低またはクリティカルバッテリーレベル、転倒などの障害またはリスクの存在)にある。このような場合、ロボットの優先度は、それ自身の課題を処理し解決する(例えばそれ自身のタスクを行う)ことである。例えば、バッテリーレベルがクリティカルであれば、ロボットはユーザとの対話を遮断し、エネルギー源ベースに到達しようとし得る。対話モードは、ユーザが近傍で検知されればおよび/またはロボットがクリティカル状況(ロボットはその基本機能を実行し得ない)にいなければ、活性化され得る。反対に、対話モードモジュールは、いかなるユーザも近傍で検知されなければおよび/またはロボットがクリティカル状況にいれば、非活性化され得る。

10

**【0085】**

開示された方法は、完全ハードウェア実施形態(例えばFPGA)、完全ソフトウェア実施形態またはハードウェアとソフトウェア要素の両方を含む実施形態の形式を採り得る。ソフトウェア実施形態は、限定しないがファームウェア、常駐ソフトウェア、マイクロコードなどを含む。本発明は、コンピュータまたは任意の命令実行システムにより使用されるまたはそれに関連して使用されるプログラムコードを提供するコンピュータ使用可能またはコンピュータ読み取り可能記憶媒体からアクセス可能なコンピュータプログラム製品の形式を採り得る。コンピュータ使用可能またはコンピュータ読み取り可能記憶媒体は、命令実行システム、装置、またはデバイスにより使用されるまたはそれに関連して使用されるプログラムを含み、格納し、伝達し、広め、または搬送し得る任意の装置であり得る。コンピュータ読み取り可能記憶媒体は、電子的、磁氣的、光学的、電磁氣的、赤外線、半導体システム(または装置またはデバイス)、または伝播媒体であり得る。

20

【図1】

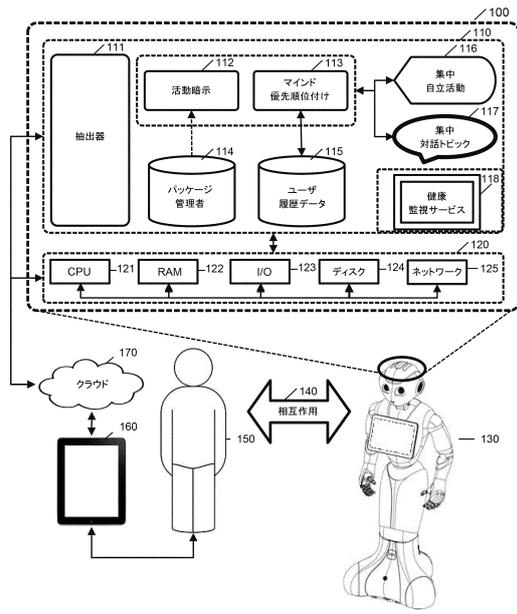


図1

【図2】

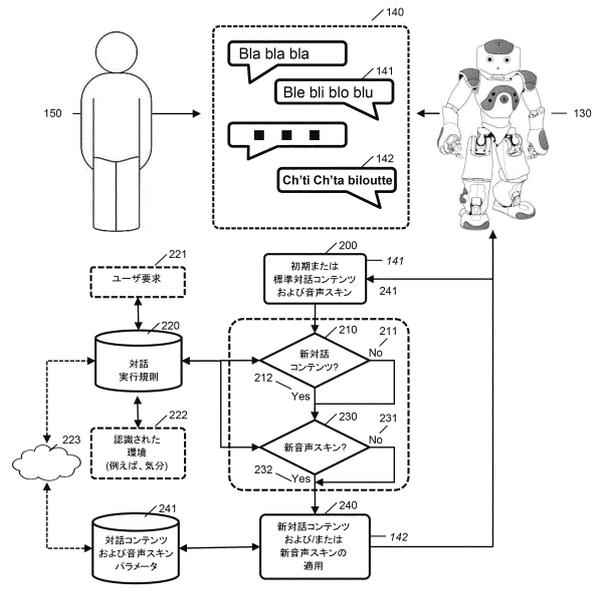


図2

---

フロントページの続き

- (72)発明者 パタイヨ, レミ  
フランス国、9 2 1 3 0・イッシー-レ-ムリノー、プロムナード・デュ・ベルジエ・1 5 2
- (72)発明者 バルビエリ, ガブリエル  
フランス国、7 5 0 1 7・パリ、リュ・ゴテエ・4 6

審査官 大野 弘

- (56)参考文献 特開2010-170033(JP, A)  
特開2004-109323(JP, A)  
特開2007-219149(JP, A)  
特開平05-165486(JP, A)  
特開2000-172289(JP, A)  
特開平10-133852(JP, A)  
再公表特許第2011/138852(JP, A1)  
再公表特許第2009/044533(JP, A1)

- (58)調査した分野(Int.Cl., DB名)
- |         |           |
|---------|-----------|
| G 1 0 L | 1 5 / 2 2 |
| B 2 5 J | 1 9 / 0 0 |
| G 1 0 L | 1 3 / 1 0 |