

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7313455号

(P7313455)

(45)発行日 令和5年7月24日(2023.7.24)

(24)登録日 令和5年7月13日(2023.7.13)

(51)国際特許分類

F I

G 1 0 L 13/00 (2006.01)

G 1 0 L 13/00 1 0 0 M

G 0 6 N 20/00 (2019.01)

G 0 6 N 20/00 1 3 0

G 1 0 L 15/10 (2006.01)

G 1 0 L 15/10 5 0 0 N

G 1 0 L 15/00 (2013.01)

G 1 0 L 15/00 2 0 0 A

G 0 6 F 3/16 (2006.01)

G 0 6 F 3/16 6 9 0

請求項の数 12 (全29頁)

(21)出願番号 特願2021-544347(P2021-544347)

(86)(22)出願日 令和2年1月10日(2020.1.10)

(65)公表番号 特表2022-523504(P2022-523504
A)

(43)公表日 令和4年4月25日(2022.4.25)

(86)国際出願番号 PCT/US2020/013160

(87)国際公開番号 WO2020/159693

(87)国際公開日 令和2年8月6日(2020.8.6)

審査請求日 令和4年12月13日(2022.12.13)

(31)優先権主張番号 15/929,095

(32)優先日 平成31年1月29日(2019.1.29)

(33)優先権主張国・地域又は機関
米国(US)

早期審査対象出願

(73)特許権者 521335616

グリッドスペース インコーポレイテッド
アメリカ合衆国 カリフォルニア州 9 4
1 0 8, サンフランシスコ カーニー ス
トリート 2 3 7 スペース 3 8 6

(74)代理人 100107766

弁理士 伊東 忠重

(74)代理人 100070150

弁理士 伊東 忠彦

(74)代理人 100135079

弁理士 宮崎 修

(72)発明者 スコデアリー, アンソニー, スコデアリ
ーアメリカ合衆国 カリフォルニア州 9 0
0 2 9, ロサンゼルス ノース ヴァージ
最終頁に続く

(54)【発明の名称】 発話エージェント

(57)【特許請求の範囲】

【請求項1】

発話合成会話エージェントを作動させる方法であって、

コンピュータが、通話セッション中に発呼者オーディオ信号を受信するようオーディオインタフェースを作動させることと、前記コンピュータが、感情モデルによって構成される感情分析エンジンの作動を通じて前記発呼者オーディオ信号からの感情スコアを含むオーディオトランスクリプトを生成することと、前記コンピュータが、ユーザモデルから入力を受信するように構成されるユーザインタフェーススイッチに前記オーディオトランスクリプトを通信することと、前記コンピュータが、前記ユーザモデルからの履歴会話データで訓練される発話シンセサイザエンジンに前記ユーザインタフェーススイッチからの応答制御を通信することと、前記コンピュータが、前記発話シンセサイザエンジンを作動させて、

前記履歴会話データによって構成される応答論理エンジンの作動を通じて前記オーディオトランスクリプト及び前記発呼者オーディオ信号についての応答信号を生成し、

前記履歴会話データによって構成される発話合成モデルの作動を通じて前記応答信号から合成されたユーザモデル応答及び周囲信号を含む合成されたオーディオ応答を生成し、
且つ前記コンピュータが、前記通話セッション中に前記オーディオインタフェースを通じて、
前記発呼者オーディオ信号に応答して、前記合成されたオーディオ応答を通信することと、

前記コンピュータが、前記オーディオインタフェースを通じて前記発呼者オーディオ信号を受信することに対応して、前記ユーザモデルから前記ユーザインタフェーススイッチを通じてユーザモデル入力を受信することと、

前記コンピュータが、前記発呼者オーディオ信号に対応して、応答オーディオと、バックグラウンドノイズとを含む、ユーザモデルオーディオ応答を、前記オーディオインタフェースに通信することと、

前記コンピュータが、前記オーディオトランスクリプト、前記発呼者オーディオ信号、及び前記ユーザモデルオーディオ応答を、制御されたメモリデータ構造内に前記履歴会話データとして格納することと、

前記コンピュータが、前記発話合成モデルを作動させて、前記履歴会話データ内のユーザモデル応答の前記バックグラウンドノイズから前記周囲信号を生成することと、を含む、方法。

【請求項 2】

前記コンピュータが、前記発話合成モデルを作動させて、前記応答論理エンジンから文言対発話応答を受信することに対応して、前記発呼者オーディオ信号についての合成されたユーザモデル応答として合成された発話を生成することを更に含む、請求項 1 に記載の方法。

【請求項 3】

前記コンピュータが、前記発話合成モデルを作動させて、前記応答論理エンジンから非言語応答を受信することに対応して、前記発呼者オーディオ信号についての前記合成されたユーザモデル応答としてリスニング応答キューを生成することを更に含む、請求項 1 に記載の方法。

【請求項 4】

前記コンピュータが、教師付き訓練フェーズ中に前記発話シンセサイザエンジンを作動させて、

前記ユーザモデルを用いて複数の通話セッションについての通話要約トランスクリプトを受信することと、

前記ユーザモデルから応答監査を受信することに対応して、前記応答論理エンジンの作動を通じて応答状態を識別し、確実性スコアを備える応答オプションを生成することと、

前記応答状態、前記応答オプション、及び前記確実性スコアに対応して、前記ユーザモデルからフィードバック制御を受信することと、を含む、

前記通話要約トランスクリプトは、識別される実体、通話セッション意図、前記感情スコア、及び前記応答論理エンジンの訓練インタフェースを通じたユーザモデル応答を含む、請求項 1 に記載の方法。

【請求項 5】

命令を含むソフトウェアを格納する非一時的コンピュータ読取可能格納媒体であって、

前記命令は、コンピュータによって実行されるときに、前記コンピュータに、

通話セッション中にオーディオインタフェースを作動させて発呼者オーディオ信号を受信させ、

感情モデルによって構成される感情分析エンジンの作動を通じて前記発呼者オーディオ信号から感情スコアを含むオーディオトランスクリプトを生成させ、

ユーザモデルから入力を受信するように構成されるユーザインタフェーススイッチに前記オーディオトランスクリプトを通信させ、

前記ユーザモデルからの履歴会話データで訓練された発話シンセサイザエンジンに前記ユーザインタフェーススイッチからの応答制御を通信させ、

前記発話シンセサイザエンジンを作動させて、

前記履歴会話データによって構成される応答論理エンジンの作動を通じて前記オーディオトランスクリプト及び前記発呼者オーディオ信号についての応答信号を生成させ、

前記履歴会話データによって構成される発話合成モデルの作動を通じて前記応答信号からの合成されたユーザモデル応答及び周囲信号を含む合成されたオーディオ応答を生

10

20

30

40

50

成させ、

前記通話セッション中に前記オーディオインタフェースを通じて、前記発呼者オーディオ信号に回答して、前記合成されたオーディオ回答を通信させ、

前記オーディオインタフェースを通じて前記発呼者オーディオ信号を受信することに応答して、前記ユーザモデルから前記ユーザインタフェーススイッチを通じてユーザモデル入力を受信し、

前記発呼者オーディオ信号に回答して、応答オーディオと、バックグラウンドノイズとを含む、ユーザモデルオーディオ回答を、前記オーディオインタフェースに通信し、

前記オーディオトランスクリプト、前記発呼者オーディオ信号、及び前記ユーザモデルオーディオ回答を、制御されたメモリデータ構造内に前記履歴会話データとして格納し、

前記発話合成モデルを作動させて、前記履歴会話データ内のユーザモデル回答の前記バックグラウンドノイズから前記周囲信号を生成させる、

非一時的コンピュータ読取可能格納媒体。

【請求項 6】

前記命令は、

前記発話合成モデルを作動させて、前記応答論理エンジンから文言対発話回答を受信することに応答して、前記発呼者オーディオ信号についての前記合成されたユーザモデル回答として合成された発話を生成する、

ように前記コンピュータを更に構成する、

請求項 5 に記載の非一時的コンピュータ読取可能格納媒体。

【請求項 7】

前記命令は、

前記発話合成モデルを作動させて、前記応答論理エンジンから非言語回答を受信することに応答して、前記発呼者オーディオ信号についての前記合成されたユーザモデル回答としてリスニング回答キューを生成する、

ように前記コンピュータを更に構成する、

請求項 5 に記載の非一時的コンピュータ読取可能格納媒体。

【請求項 8】

前記命令は、

教師付き訓練フェーズ中に前記発話シンセサイザエンジンを作動させて、

前記ユーザモデルを用いて複数の通話セッションについての通話要約トランスクリプトを受信し、

前記ユーザモデルから応答監査を受信することに応答して、前記応答論理エンジンの作動を通じて応答状態を識別し、確実性スコアを備える応答オプションを生成し、且つ

前記応答状態、前記応答オプション、及び前記確実性スコアに応答して、前記ユーザモデルからフィードバック制御を受信する、

ように前記コンピュータを更に構成し、

前記通話要約トランスクリプトは、識別される実体と、通話セッション意図と、前記感情スコアと、前記応答論理エンジンの訓練インタフェースを通じたユーザモデル回答とを含む、

請求項 5 に記載の非一時的コンピュータ読取可能格納媒体。

【請求項 9】

プロセッサと、

命令を含むソフトウェアを格納するメモリと、を含む、

コンピューティング装置であって、

前記命令は、前記プロセッサによって実行されるときに、

通話セッション中に発呼者オーディオ信号を受信するようにオーディオインタフェースを作動させ、

感情モデルによって構成される感情分析エンジンの作動を通じて前記発呼者オーディオ信号からの感情スコアを含むオーディオトランスクリプトを生成し、

10

20

30

40

50

ユーザモデルからの入力を受信するように構成されるユーザインタフェーススイッチに前記オーディオトランスクリプトを通信し、

前記ユーザモデルからの履歴会話データで訓練された発話シンセサイザエンジンに前記ユーザインタフェーススイッチからの応答制御を通信し、

前記発話シンセサイザエンジンを作動させて、

前記履歴会話データによって構成される応答論理エンジンの作動を通じて前記オーディオトランスクリプト及び前記発呼者オーディオ信号についての応答信号を生成し、且つ

前記履歴会話データによって構成される発話合成モデルの作動を通じて前記応答信号からの合成されたユーザモデル応答及び周囲信号を含む合成されたオーディオ応答を生成し、

10

前記通話セッション中に前記オーディオインタフェースを通じて前記発呼者オーディオ信号に応答して前記合成されたオーディオ応答を通信し、

前記オーディオインタフェースを通じて前記発呼者オーディオ信号を受信することに応答して、前記ユーザモデルから前記ユーザインタフェーススイッチを通じてユーザモデル入力を受信し、

前記発呼者オーディオ信号に応答して、応答オーディオと、バックグラウンドノイズとを含む、ユーザモデルオーディオ応答を、前記オーディオインタフェースに通信し、

前記オーディオトランスクリプト、前記発呼者オーディオ信号、及び前記ユーザモデルオーディオ応答を、制御されたメモリデータ構造内に前記履歴会話データとして格納し、且つ

20

前記発話合成モデルを作動させて、前記履歴会話データ内のユーザモデル応答の前記バックグラウンドノイズから前記周囲信号を生成させる、

ように当該コンピューティング装置を構成する、

コンピューティング装置。

【請求項 10】

前記命令は、

前記発話合成モデルを作動させて、前記応答論理エンジンから文言対発話応答を受信することに応答して、前記発呼者オーディオ信号についての前記合成されたユーザモデル応答として合成された発話を生成する、

ように当該コンピューティング装置を更に構成する、

30

請求項 9 に記載のコンピューティング装置。

【請求項 11】

前記命令は、

前記発話合成モデルを作動させて、前記応答論理エンジンから非言語応答を受信することに応答して、前記発呼者オーディオ信号についての前記合成されたユーザモデル応答としてリスニング応答キューを生成する

ように当該コンピューティング装置を更に構成する、

請求項 9 に記載のコンピューティング装置。

【請求項 12】

前記命令は、

40

教師付き訓練フェーズ中に前記発話シンセサイザエンジンを作動させて、

前記ユーザモデルを用いて複数の通話セッションについての通話要約トランスクリプトを受信し、

前記ユーザモデルから応答監査を受信することに応答して、前記応答論理エンジンの作動を通じて応答状態を識別し、確実性スコアを備える応答オプションを生成し、

前記応答状態、前記応答オプション、及び前記確実性スコアに応答して、前記ユーザモデルからフィードバック制御を受信する、

ように当該コンピューティング装置を更に構成し、

通話要約トランスクリプトは、識別された実体と、通話セッション意図と、前記感情スコアと、前記応答論理エンジンの訓練インタフェースを通じたユーザモデル応答とを含む、

50

請求項 9 に記載のコンピューティング装置。

【発明の詳細な説明】

【背景技術】

【0001】

発呼者(caller)とサービスエージェント(service agent)／代表者(representative)との間のオーディオ通信に依存する多くのサービス対話(service interaction)は、自動化エージェント(automated agents)（すなわち、ロボット(bots)）の使用から利益を得ることがある。自動化エージェントは、発呼者からの定型データ（例えば、住所、電話番号、口座番号など）を収集することのような、単調なタスクを処理することができるが、質問に応答すること又は会話に関与することのような、より複雑なタスクを処理することができない。多くの個人は、自動化エージェントと会話することが不快(off putting)であると知り、これらのシステムと対話する(interact)ことを強制されることは、個人の情緒的状态に否定的な影響を及ぼすことがある。解決策として、ある者は、人間の発話(speech)を合成できるロボットを使用することを提案する。これは可能性であるが、現在の発話合成システムは、より複雑な会話を説明することを困難にする、大量の手作業チューニング及び設計で構築されたオントロジー(ontologies)を利用する。これらのシステムは、非常に堅牢(ロバスト)に見えるが、それらの深さは限られている。従って、自動化エージェントに対する改良の必要性が存在する。

10

【発明の概要】

【0002】

任意の特定の要素又は行為の議論を容易に識別するために、参照番号中の最上位の桁(digit)又は数桁(digits)は、その要素が最初に導入された図番号を指している。

20

【図面の簡単な説明】

【0003】

【図 1】 1つの実施形態に従ったシステム 100 を図示している。

【0004】

【図 2】 1つの実施形態に従った方法 200 を図示している。

【0005】

【図 3】 1つの実施形態に従ったシステム 300 を図示している。

【0006】

【図 4】 1つの実施形態に従った方法 400 を図示している。

【0007】

【図 5】 1つの実施形態に従ったプロセス 500 を図示している。

【0008】

【図 6】 1つの実施形態に従った教師付きトレーニングフェーズ 600 を図示している。

【0009】

【図 7】 1つの実施形態に従った教師付きトレーニングフェーズ 700 を図示している。

【0010】

【図 8】 1つの実施形態に従った教師付きトレーニングフェーズ 800 を図示している。

【0011】

【図 9】 1つの実施形態に従った畳み込みニューラルネットワーク 900 を図示している。

【0012】

【図 10】 1つの実施形態に従った畳み込みニューラルネットワーク層 1000 を図示している。

30

【0013】

【図 11】 1つの実施形態に従った VGG ネット 1100 を図示している。

【0014】

【図 12】 1つの実施形態に従った畳み込み層フィルタリング 1200 を図示している。

【0015】

【図 13】 1つの実施形態に従ったプーリング層機能 1300 を図示している。

50

40

【0016】

【図14】1つの実施形態に従った図1400を図示している。

【発明を実施するための形態】

【0017】

部分的に自動化された会員サービス代表(MSR: member services representative)ロボットは、人間MSRが利用可能であるときに、人間MSRへの引き渡し(hand off)をシームレスに処理することがある。MSRロボットは、通話(call)の自動化された部分を要約する能力と、ロボットの挙動を監視し且つ検討するツールとを含むことがある。記録されたMSR通話を利用して、自然に聞こえるように発話シンセサイザエンジン(speech synthesizer engine)を訓練し、発呼者(caller)の要求に正確に反応するように会話エージェント(conversational agent)を訓練することがある。MSRロボットは、綺麗な演じられた発話(speech)ではなく、衛生的にされていない会話的な発話を使用して、訓練されることがある。MSRロボットは、個々のMSRの語調(tone)、韻律(prosody)、及び方言(dialect)をモデル化することができることがある。MSRロボットは、人間トランスクリプト(human transcripts)ではなく、発話認識トランスクリプト(speech recognition transcripts)を使用して、訓練されることがある。

【0018】

畳み込みニューラルネットワーク(CNN: convolutional neural networks)は、二次元又は三次元においてモデル化されるデータセットにおける特徴(features)を分類するのに特に適している。これは画像分類(image classification)についてCNNを評判の良いものにする。何故ならば、三次元(幅及び高さについて二次元、色成分(color components)及び強度(intensity)のようなピクセル特徴について三次元)において画像をコンピュータメモリ内に表現することができるからである。例えば、480×480のサイズのピクセルのカラーJPEG画像は、480×480×3であるアレイを用いて、コンピュータメモリ内でモデル化されることができ、ここで、第3の次元の値の各々は、0～255に亘るピクセルについての赤色、緑色、又は青色の色成分強度である。この数字のアレイを訓練されたCNNに入力することは、画像が特定の等級(class)である確率(猫について0.80、犬について0.15、鳥について0.05など)を記述する出力を生成する。画像分類は、入力画像を撮り、画像を最も良く記述する等級(猫、犬など)又は等級の確率を出力する、タスクである。

【0019】

基本的に、CNNは、データセットを入力し、それを一連の畳み込み変換、非線形活性化関数(例えば、ReLU)、及びプーリング演算(ダウンサンプリング(downsampling)、例えば、マックスプूल(maxpool))、及び出力層(output layer)(例えば、ソフトマックス(softmax))を通過させて、分類を生成する。

【0020】

発話合成会話エージェント(speech synthesizing conversation agent)を作動させる方法は、通話セッション中に発呼者オーディオ信号を受信するためにオーディオインタフェースを作動させることを含む。方法は、感情モデル(センチメントモデル)(sentiment model)によって構成される感情分析エンジン(センチメント分析エンジン)(sentiment analysis engine)の作動を通じて発呼者オーディオ信号からの感情スコア(センチメントスコア)(sentiment score)を含むオーディオトランスクリプトを生成する。方法は、ユーザモデルから入力を受信するように構成されるユーザインタフェーススイッチにオーディオトランスクリプトを通信する。方法は、ユーザインタフェーススイッチからの応答制御を、ユーザモデルからの履歴会話データで訓練された発話シンセサイザエンジンに通信する。次に、方法は、発話シンセサイザエンジンを作動させる。発話シンセサイザエンジンは、履歴会話データによって構成される応答論理エンジンの作動を通じてオーディオトランスクリプト及び発呼者オーディオ信号についての応答信号を生成する。発話シンセサイザエンジンは、履歴会話データによって構成される発話合成モデルの作動を通じた応答信号からの合成されたユーザモデル応答と周囲信号とを含む、合成オーディオ応答を生成

する。次に、方法は、通話セッション中に、オーディオインタフェースを通じた発呼者オーディオ信号に回答して合成オーディオ応答を通信する。

【0021】

発話合成会話エージェントを作動させる方法は、応答論理エンジンからの文言対発話応答の受信に回答して、発話合成モデルを作動させて、発呼者オーディオ信号についての合成ユーザモデル応答として合成発話を生成することを含むことがある。

【0022】

発話合成会話エージェントを作動させる方法は、応答論理エンジンから非言語応答を受信することによって回答して、発話合成モデルを作動させて、発呼者オーディオ信号のための合成されたユーザモデル応答としてリスニング応答キュー(listening response cues)を生成することを含んでよい。

10

【0023】

発話合成会話エージェントを作動させる方法は、オーディオインタフェースを通じて発呼者オーディオ信号を受信することによって回答して、ユーザモデルからユーザインタフェーススイッチを通じてユーザモデル入力を受信することを含んでよい。方法は、発呼者オーディオ信号に回答して、ユーザモデルオーディオ応答をオーディオインタフェースに通信してよく、ユーザモデルオーディオ応答は、応答オーディオと、バックグラウンドノイズとを含む。次に、方法は、オーディオトランスクリプト、発呼者オーディオ信号、及びユーザモデルオーディオ応答を、制御されたメモリデータ構造内に履歴会話データとして格納してよい。幾つかの構成において、方法は、発話合成モデルを作動させて、履歴会話データ内のユーザモデル応答のバックグラウンドノイズから周囲信号を生成する。

20

【0024】

発話合成会話エージェントを作動させる方法は、ユーザモデルを用いて複数の通話セッションについての通話要約トランスクリプトを受信するために、教師付き訓練フェーズ中に発話シンセサイザエンジンを作動させることを含んでよい。通話要約トランスクリプトは、識別された実体と、通話セッション意図と、感情スコアと、応答論理エンジンの訓練インタフェースを通じたユーザモデル応答とを含んでよい。教師付き訓練フェーズ中に、方法は、ユーザモデルから応答監査を受信することによって回答して、応答状態を識別し、応答論理エンジンの作動を通じて確実性スコアを備える応答オプションを生成してよい。方法は、応答状態、応答オプション、及び確実性スコアに回答して、ユーザモデルからフィードバック制御を受信してよい。

30

【0025】

命令を含む非一時的コンピュータ読取可能格納媒体であって、命令は、コンピュータによって実行されるときに、コンピュータに、通話セッション中にオーディオインタフェースを作動させて発呼者オーディオ信号を受信する。コンピュータは、感情分析エンジンの作動を通じて発呼者オーディオ信号から感情スコアを含むオーディオトランスクリプトを生成してよい。コンピュータは、ユーザモデルから入力を受信するように構成されるユーザインタフェーススイッチにオーディオトランスクリプトを通信してよい。コンピュータは、ユーザモデルからの履歴会話データで訓練された発話シンセサイザエンジンにユーザインタフェーススイッチからの応答制御を通信してよい。コンピュータは、発話シンセサイザエンジンを作動させてよい。発話シンセサイザエンジンは、履歴会話データによって構成される応答論理エンジンの作動を通じてオーディオトランスクリプト及び発呼者オーディオ信号についての応答信号を生成してよい。発話シンセサイザエンジンは、履歴会話データによって構成される発話合成モデルの作動を通じて応答信号からの合成されたユーザモデル応答及び周囲信号を含む合成されたオーディオ応答を生成してよい。次に、コンピュータは、通話セッション中にオーディオインタフェースを通じて発呼者オーディオ信号に回答して合成されたオーディオ応答を通信してよい。

40

【0026】

命令は、発話合成モデルを作動させて、応答論理エンジンから文言対発話応答を受信することによって回答して、発呼者オーディオ信号についての合成されたユーザモデル応答として

50

合成された発話を生成するように、コンピュータを更に構成する。

【0027】

命令は、発話合成モデルを作動させて、応答論理エンジンから非言語応答を受信することに応答して、発呼者オーディオ信号についての合成されたユーザモデル応答としてリスニング応答キューを生成するように、コンピュータを更に構成してよい。

【0028】

命令は、オーディオインタフェースを通じて発呼者オーディオ信号を受信することに応答して、ユーザモデルからユーザインタフェーススイッチを通じてユーザモデル入力を受信するように、コンピュータを更に構成してよい。コンピュータは、発呼者オーディオ信号に応答して、オーディオインタフェースにユーザモデルオーディオ応答を通信してよく、ユーザモデルオーディオ応答は、応答オーディオと、バックグラウンドノイズとを含む。次に、コンピュータは、オーディオトランスクリプト、発呼者オーディオ信号、及びユーザモデルオーディオ応答を、制御されたメモリデータ構造内に履歴会話データとして格納してよい。幾つかの構成において、命令は、発話合成モデルを作動させて、履歴会話データ内のユーザモデル応答のバックグラウンドノイズから周囲信号を生成するように、コンピュータを構成してよい。

【0029】

幾つかの構成において、命令は、教師付き訓練フェーズ中に発話シンセサイザエンジンを作動させるように、コンピュータを構成してよい。教師付き訓練フェーズ中に、命令は、ユーザモデルを用いて複数の通話セッションについての通話要約トランスクリプトを受信するように、コンピュータを構成してよく、通話要約トランスクリプトは、識別される実体と、通話セッション意図と、感情スコアと、応答論理エンジンの訓練インタフェースを通じたユーザモデル応答とを含む。コンピュータは、ユーザモデルから応答監査を受信することに応答して、応答論理エンジンの作動を通じて応答状態を識別し、確実性スコアを備える応答オプションを生成してよい。コンピュータは、応答状態、応答オプション、及び確実性スコアに応答して、ユーザモデルからフィードバック制御を受信してよい。

【0030】

コンピューティング装置が、プロセッサと、命令を格納するメモリとを含んでよく、命令は、プロセッサによって実行されるときに、オーディオインタフェースを作動させて、通話セッション中に発呼者オーディオ信号を受信するように、コンピューティング装置を構成する。コンピューティング装置は、感情分析エンジンの作動を通じて発呼者オーディオ信号からの感情スコアを含むオーディオトランスクリプトを生成してよい。コンピューティング装置は、ユーザモデルから入力を受信するように構成されるユーザインタフェーススイッチにオーディオトランスクリプトを通信してよい。コンピューティング装置は、ユーザモデルからの履歴会話データで訓練された発話シンセサイザエンジンにユーザインタフェーススイッチからの応答制御を通信してよい。コンピューティング装置は、発話シンセサイザエンジンを作動させてよい。発話シンセサイザエンジンは、履歴会話データによって構成される応答論理エンジンの作動を通じてオーディオトランスクリプト及び発呼者オーディオ信号についての応答信号を生成してよい。発話シンセサイザエンジンは、履歴会話データによって構成される発話合成モデルの作動を通じて応答信号から合成されたユーザモデル応答及び周囲信号を含む合成されたオーディオ応答を生成してよい。コンピューティング装置は、通話セッション中にオーディオインタフェースを通じて発呼者オーディオ信号に応答して合成されたオーディオ応答を通信してよい。

【0031】

命令は、発話合成モデルを作動させて、応答論理エンジンから文言対発話応答を受信することに応答して、発呼者オーディオ信号についての合成されたユーザモデル応答として合成された発話を生成するように、コンピューティング装置を更に構成してよい。

【0032】

命令は、発話合成モデルを作動させて、応答論理エンジンから非言語応答を受信することに応答して、発呼者オーディオ信号についての合成されたユーザモデル応答としてリス

10

20

30

40

50

ニング応答キューを生成するように、コンピューティング装置を更に構成してよい。

【0033】

命令は、オーディオインタフェースを通じて発呼者オーディオ信号を受信することに対応して、ユーザモデルからユーザインタフェーススイッチを通じてユーザモデル入力を受信するように、装置を更に構成してよい。装置は、発呼者オーディオ信号に対応して、ユーザモデルオーディオ応答をオーディオインタフェースに通信してよく、ユーザモデルオーディオ応答は、応答オーディオと、バックグラウンドノイズとを含む。装置は、オーディオトランスクリプト、発呼者オーディオ信号、及びユーザモデルオーディオ応答を、制御されたメモリデータ構造内に履歴会話データとして格納してよい。

【0034】

幾つかの構成において、命令は、発話合成モデルを作動させて、履歴会話データ内のユーザモデル応答のバックグラウンドノイズから周囲信号を生成するように、装置を更に構成してよい。

【0035】

命令は、教師付き訓練フェーズ中に発話シンセサイザエンジンを作動させるように、装置を更に構成してよい。教師付き訓練フェーズ中に、装置は、ユーザモデルを用いて複数の通話セッションについての通話要約トランスクリプトを受信してよく、通話要約トランスクリプトは、識別された実体と、通話セッション意図と、感情スコアと、応答論理エンジンの訓練インタフェースを通じたユーザモデル応答とを含む。装置は、ユーザモデルから応答監査を受信することに対応して、応答論理エンジンの作動を通じて応答状態を識別し、確実性スコアを備える応答オプションを生成してよい。装置は、応答状態、応答オプション、及び確実性スコアに対応して、ユーザモデルからフィードバック制御を受信してよい。

【0036】

ある実施形態において、感情分析エンジンによって生成されるオーディオトランスクリプトの例は、以下のように見える。

10

20

30

40

50

【表 1】

{		
"transcript": [		
{		
"pretty_transcript": "Hi I just heard a baby.",		
"index": 1,		
"sentiment": {		
"mm_v0": 0.9,	10	
}		
"word_offsets_ms": [		
870,		
1170,		
1260,		
1590,		
1920,	20	
1980,		
],		
"word_durations_ms": [		
300,		
90,		
300,		
330,		
60,	30	
750,		
240,		
690		
],		
"connection_id": "69ffe872b5dbf929",		
"start_timestamp_ms": 1531345706000,		
"channel_index": 1,	40	

Docket No: FSP1479

```
"start_ms": 0,  
"duration_ms": 4750,  
"transcript": "hi i just heard a baby",  
"offset_ms": -84  
},  
],  
"type": "offline"  
}
```

10

【0037】

オーディオトランスクリプトは、各々の語られた単語の継続時間及び単語間のオフセットについてのタイミングを示す。更に、オーディオトランスクリプトは、0.9の感情インデックススコアを例示する。

20

【0038】

幾つかの構成において、MSRロボットは、単一モデルにおいて複数のMSR間のスタイル転送(style transfer)を実行することができてよい。MSRロボットは、ラインノイズ、非発話発声(non speech vocalization)、及び現実的なコールセンター及び電話方式音響環境(telephony acoustic environment)のような音を暗黙のうちにモデル化することができてよい。通話セッション中の発呼者の情動(emotions)及び感情(sentiment)は、文言対発話(TTS:text to speech)モデルに条件付けして現実的なダイナミクスをTTS韻律(prosody)に加えるために利用されてよい(例えば、合成された音声は、通話セッション中に検出される怒り又は懸念に応答して、そのトーンを変調させてよい。会話エージェント(又はMSRロボット)は、通話セッション中に数回の対話(通話(talking)ー聴取(listening))を信じられるように処理することができてよい。

30

【0039】

会話エージェントは、通話の部分のみを取り扱うように設計されてよく、通話の自動化された部分において達成される解決策及び発呼者の意図を要約してよい。会話エージェントは、受け渡し(ハンドオフ)(handoff)において人間MSRを支援する短い要約アーチファクトを生成してよい。

【0040】

要約システムは、部分的に又は完全に、先行するロボット会話のトランスクリプトを提供するダッシュボード(dashboard)を含んでよい。このダッシュボードは、会話エージェントを訓練し且つ解釈するために利用されるツール(tooling)の一部であってよい。ロボット管理ダッシュボードの設計は、なぜロボットが特定の応答を選択したのかを説明し且つ挙動を修正するメカニズムを提供する機会を可能にすることがある。固有名詞のような特殊なトークンをラベル付けすることも、構造化された学習を可能にすることがある。加えて、このダッシュボードは、通話セッション中の受け渡しポイントまでの会話の合理的な要約を提供することがある。合成要約(synthetic summarization)は、通話に関するより小さな結論から複合要約(compound summary)の上に構築されてよい。この複合要約は、通話要約の基礎として機能することがある。

40

【0041】

会話エージェントのTTS部分は、MSR当たり20~200時間のオーダで、かなり高いデータ要件を有することがある。これは(削減、転送、クロストークなどの故に)使

50

用できないオーディオをフィルタで除去した後である。どれぐらい多くの音声モデル化されることが意図されているか及び最も多くの成功を有するスタイル転送方法に依存して、使用されるデータセットは、数十万以上の通話及びトランスクリプトを含むことができる。

【0042】

幾つかの構成において、会話エージェントは、情動及び感情をTTSへの条件付け信号としてモデル化することを例外として、音声に依存しないことがある。しかしながら、アーキテクチャに含まれるあらゆる生殖モデル(generative model)は、通話トランスクリプトの最大の可能なコーパス(corpus)上で訓練されることがある。

【0043】

会話エージェントの合成要約部分は、小さな検証データセットに依存することがあるが、分類器が意図を検出するために通話について訓練されているならば、データセットは、分類器当たり数千の通話トランスクリプトを含むことがある。

【0044】

幾つかの構成において、以前の会話からの訓練データは、以下のように、発呼者からの検索用語及び検索語句修飾子を用いて検索されてよい。

【0045】

(直定数(リテラル)(Literals))

訓練中、単語や語句を検索するために、それを引用符内に置く。

“crash”(「衝突」)

“lost credit card”(「紛失したクレジットカード」)

引用される語句の前に1～3の上付き波形符号(～)を加えると、類似のマッチ(一致)を緩める(意味論的には、音ではなく意味が類似することを意味する)。

単一の上付き波形符号を利用して、複数(plurals)又は同根(conjugates)のような類似の形態をマッチさせてよい。例えば、以下の通りである。

～“crash”は、“crashes”又は“crashing”とマッチする。

2つの上付き波形符号を利用して、同義語をマッチさせてよい。例えば、以下の通りである。

～～“crash”は、“accident”(「事故」)又は“collision”(「衝突」)とマッチする。

3つの上付き波形符号を利用して、関連する語句をマッチさせてよい。例えば、以下の通りである。

～～～“have a nice day”(「良い1日を」)は、“I hope your day is great”(「良い1日でありますように」)とマッチする。

【0046】

(語句演算子(Phrase Operators))

1つのものについて1つの発話セグメント内を検索するために、ユーザは、検索用語を演算子‘near’、‘or’、又は‘then’と組み合わせてよい。例えば、以下の通りである。

～～“crash”near“honda”

～～“crash”及び“honda”の両方を探す。

別の例は、以下の通りである。

～～“crash”or“ticket”(「交通違反切符」)

～～“crash”又は“ticket”の一方又は両方を探す。

更に別の例は、以下の通りである。

～～“crash”then“report”(「報告書」)

～～“crash”及び“report”の両方を順に探す。

例えば、“I had an accident and then they wrote a police report”(「私は事故を起こし、次に、彼らは事故報告書を書いた」)はマッチするが、“I found the police report after the crash”(私は衝突の後に事故報告書を発見した)はマッチしない。

【0047】

10

20

30

40

50

(会話演算子(Conversation Operators))

2つのものについて会話全体に亘って検索するために、ユーザは検索用語を演算子and、or又はlaterと組み合わせてよい。

and演算子は、両方の直定数(リテラル)を含む会話を探す。

～～“lost card”(「紛失したカード」) and “two weeks”(「2週間」)は、以下のような会話とマッチする。

Hello thanks for calling. (お電話ありがとうございます。)

...

I want to report a missing card. (私はカードの紛失を報告したいです。)

...

The new card should arrive in one to two weeks. (新しいカードは1～2週間で届くはずです。)

しかしながら、対照的に、near演算子はマッチしない。何故ならば、それらは異なる発話セグメントに亘るからである。

or演算子は、直定数(リテラル)の一方又は両方を含む会話を探す。その使用は、語句スキナに関する文脈(コンテキスト)によって決定される。

Caller ～～“lost card” or caller “two weeks”は、

以下のような会話とマッチする。

Hello thanks for calling. (お電話ありがとうございます。)

...

I want to report a missing card. (私は紛失したカードを報告したいです。)

...

The new card should arrive in five days. (新しいカードは5日以内に届くはずです。)

...

Later演算子は、直定数(リテラル)の両方を順に含む会話を探す。

～～～“reset my password”(「パスワードを再設定する」) later ～“thanks”(「ありがとう」)は、

以下のような会話とマッチする。

Hello thanks for calling. (電話ありがとう。)

...

I need my password reset. (私はパスワードを再設定する必要があります。)

...

Thanks! (ありがとう!)

しかしながら、最後の「ありがとう」が省略されたならば、会話の前の方でありがとう(thanks)が述べられていたとしても、会話はマッチしない。

【0048】

(セグメント修飾子(Segment Modifiers))

追加的な修飾子をセグメントの左側に置いて、それを特定の特性(property)に限定したり、それを何らかの他の方法で修飾したりしてよい。

エージェント	エージェントが次の語句を言う場合にのみ当て嵌まる	Agent ～～“great to hear”
発呼者	発呼者が次の語句を言う場合に当て嵌まる	Caller ～～“very helpfull”
否定(or)	次の語句が起こらない場合にのみ当て嵌まる	Not ～～“claim”

【0049】

(複合クエリ(Compound Queries))

10

20

30

40

50

更に複雑なクエリが括弧を用いて構築されてよい。内部スキャナが評価され、次に、外部スキャナと組み合わせられる。

(~~“crash” near ~~“police report”) or ~~~“file a claim”(「申立を提出する」)

この語句は、crash及びpolice reportの両方が述べられている場合又はclaimが提出されている場合(又は両方)にのみ一致する。しかしながら、“police report”のみでは一致しない。これは制限なく行なわれてよい。

((~~“crash” near ~~“police report”) or ~~~“file a claim”) later agent ~~“sorry”) and caller not ~~“thank you”) or “thank you for your help with the claim”

10

修飾子がスタックされてもよい(しかしながら、順序は意味に影響を及ぼし得る)ことに留意のこと。

not agent ~~“sorry”は、

エージェントが誤らない会話とマッチする。

【0050】

(抽出器(Extractors))

抽出器は、概念を表す中括弧「{}」中の特殊な語句である。デフォルトで、全ての抽出器は、それらが恰も2つの上付き波形符号を有するように扱われ、これは省略されることがある。

~~“hello my name is {name}”(「こんにちは、私の名前は{名前}です」)は、

20

「hi my name is George」(「やあ!私の名前はジョージです」)とマッチする。

現在スキャナ内にある幾つかの抽出器は、以下を含む。

{firstName}	Anthony, Steve
{surname}	Richardson, Hernandez
{fullName}	Anthony Richardson, Steve Hernandez
{date}	March Fifth, Christmas
{time}	Five thirty a.m., Noon
{greeting}	Hi there, good morning
{polite}	Thanks, please
{positive}	Great, wonderful, amazing
{negative}	Terrible, awful, sad
{company}	Microsoft, Gridspace
{zipCode}	Nine oh two one oh
{title}	Mister, Miss, Doctor
{phoneNumber}	Eight six even five three oh nine

30

40

50

【 0 0 5 1 】

(時間演算子(Time Operators))

ユーザは、スキャナに対して時間的制約をつけてよい。

【 0 0 5 2 】

(最大持続時間(Maximum Duration))

角括弧、未満演算子、数字、及び単位で、ある量未満の時間が経過したことを指定する。

[<30s]	Less than 30 seconds (30秒未満)
[<5s]	Less than five seconds (5秒未満)
[<5m]	Less than five minutes (5分未満)

10

例えば、以下の通りである。

“interest rate” [<30s] “a.p.r.” は、
「interest rate」の後に30以内の「a.p.r」を探す。

【 0 0 5 3 】

(最小持続時間)

[>20s]	More than 20 seconds (20秒以上)
[>100s]	More than one hundred seconds (100秒以上)
[>15m]	More than fifteen minutes (15分以上)

20

【 0 0 5 4 】

(開始及び終了トークン(Start and End Tokens))

通話の開始及び終了は、通話の開始及び終了に対して時間的制約を課するための特殊な抽出器を用いて指定されてよい。

例えば、以下の通りである。

{start} [<30s] “thanks for calling”

“thanks for calling” (「電話ありがとう」) という要求は、最初の30秒内に言われる。

同様に、{end} は、通話の終了を示すことができる。

“anything else today” [>1m] {end}

これは“anything else today” (「今日は他にありますか」) が通話終了の1分以上前に言われたことを強制する。

【 0 0 5 5 】

(メタデータ(Metadata))

検索ビルダ(search builder)において、ユーザは、日付、開始時刻、持続時間、ユーザ提供メタデータのような通話セッションメタデータに対する制約をつけてもよい。

メタデータクエリが先ず実行され、次に、スキャナが結果として得られるサブセットに対して実行される。

【 0 0 5 6 】

図1は、通話セッション中にメンバーサポート代表者(MSR: member support representative)に倣ってモデル化された合成されたオーディオ応答(synthesized audio responses)を生成するための発話合成会話エージェント(speech synthesizing conversation agent)を作動させるシステム100を図示している。システム100は、オ

30

40

50

オーディオインタフェース110と、感情分析エンジン106と、ユーザインタフェーススイッチ116と、制御されたメモリデータ構造118(controlled memory data structure)と、発話シンセサイザエンジン114とを含む。システム100の作動中、個人はコールセンターを呼び出してMSRと話す。発呼者のオーディオ信号は、発呼者オーディオ信号120としてオーディオインタフェース110を通じて受信される。システム100において、発呼者オーディオ信号120は、発呼者が彼らのアカウントに問題があることを述べる、発呼者のオーディオを含む。発呼者オーディオ信号120がシステム100を更に下方に進む前に、発呼者オーディオ信号120は、感情分析エンジン106に通信される。感情分析エンジン106は、転写サービス102(transcription service)と、感情モデル104とを含む。転写サービス102は、発呼者オーディオ信号120をオーディオトランスクリプト108に転写(トランスクリライブ)し、各々の話された単語のタイミングと、発呼者オーディオ信号120内の各々の話された単語の間の一時停止とを記す。感情モデル104は、オーディオトランスクリプト108のタイミングから感情スコア112を決定するために感情分析エンジン106によって利用される訓練された言語処理モデルである。

【0057】

感情スコア112を備えるオーディオトランスクリプト108は、ユーザインタフェーススイッチ116に通信される。ユーザインタフェーススイッチ116は、MSR(ユーザモデル136)が、感情スコア112でオーディオトランスクリプト108を検討し、それらが発呼者に個人的に回答したいかどうかを決定すること、又は発話シンセサイザエンジン114が合成オーディオ応答130を生成することを可能にすることを可能にする。ユーザモデル136と発呼者との間の初期的な通信は、完全に文言(テキスト)ベースであるので、感情スコア112は、発呼者オーディオ信号120の調律的特徴についての洞察をユーザに提供する。

【0058】

1つのシナリオにおいて、ユーザモデル136は、それらが更なる対話(インタラクション)を処理したいことを決定し、ユーザモデル入力をユーザインタフェーススイッチ116に通信する。次に、ユーザモデル136は、ユーザモデルオーディオ応答128としてオーディオインタフェース110に通信される「それは残念です。やれるだけやってみましょう。」のような応答を発声して(vocalize)よい。ユーザモデルオーディオ応答128は、応答オーディオ140としてのユーザモデル136の発声応答(vocalized response)を含み、コールセンター環境からのバックグラウンドノイズ142を含む。ユーザモデル136と発呼者との間のこの対話に続いて、ユーザモデルオーディオ応答128、発呼者オーディオ信号120、及び感情スコア112を含むオーディオトランスクリプト108は、制御されたメモリデータ構造118に通信され、履歴会話データ132として格納される。

【0059】

別のシナリオにおいて、ユーザモデル136は、発話シンセサイザエンジン114が発呼者オーディオ信号120に対する応答を処理すべきであると決定する。ユーザモデル136は、ユーザインタフェーススイッチ116を介して発話シンセサイザエンジン114に応答制御144を通信する。発話シンセサイザエンジン114は、発話合成モデル124(speech synthesis model)と、合成されたオーディオ応答130を生成するための応答論理エンジン122とを含む。応答論理エンジン122は、履歴会話データ132を用いて訓練された会話応答モデルを利用して、発呼者オーディオ信号120に対する適切な応答を決定する。応答論理エンジン122は、発呼者オーディオ信号120と、感情スコア112を含むオーディオトランスクリプト108とを利用して、応答信号を生成し、発話合成モデル124は、合成オーディオ応答130の基礎として、応答信号を使用する。

【0060】

発話合成モデル124は、ユーザモデル136の履歴会話データ132を使用してモデル化される。発話合成モデル124は、発話合成エンジン114が、適切な調律的特徴を

10

20

30

40

50

備える応答信号から合成されたユーザモデル応答 1 3 8 を生成して、発呼者オーディオ信号 1 2 0 内の検出された感情を補完することを可能にする。合成されたユーザモデル応答 1 3 8 に加えて、合成されたオーディオ応答 1 3 0 は、ユーザモデルオーディオ応答 1 2 8 内で検出されるバックグラウンドノイズ 1 4 2 に倣ってモデル化される周囲信号 1 3 4 を含み、合成されたオーディオ応答 1 3 0 に真正性(authenticity)の層(レイヤー)を追加する。

【0061】

応答論理エンジン 1 2 2 は、通話セッション中に異なる会話／応答状態を決定することがある。これらの会話／応答状態は、能動応答状態と、受動応答状態とを含む。能動応答状態は、合成された発話応答が発呼者オーディオ信号 1 2 0 に適していることを応答論理エンジン 1 2 2 が決定する、通話セッション中の状態であってよい。合成された発話応答は、感情スコア 1 1 2 に対する応答として適切なことがある。能動応答状態において、応答論理エンジン 1 2 2 は、文言対発話応答 1 2 6 (TTS 応答)を、発話シンセサイザエンジン 1 1 4 によって発話に変換されるべき文言を含む発話合成モデル 1 2 4 に通信する。幾つかの構成において、合成された発話応答は、発呼者オーディオ信号 1 2 0 において識別された質問又は特定の陳述のような、会話における識別された意図に適することがある。

10

【0062】

受動応答状態は、非言語応答 1 4 6 が発呼者オーディオ信号 1 2 0 に適切であると発話シンセサイザエンジン 1 1 4 が決定することがある、会話状態である。受動応答状態は、通話のMSR側が依然として聞いていることを示しながら、発呼者に耳を傾け続けることが適切な応答であることを、応答論理エンジン 1 2 2 が決定するときの、会話におけるインスタンスを含んでもよい。このインスタンスにおいて、発話シンセサイザエンジン 1 1 4 は、非言語応答 1 4 6 と発話合成モデル 1 2 4 とを利用して、周囲信号 1 3 4 と、合成されたユーザモデル応答 1 3 8 としてのリスニング応答キューとを含む、合成されたオーディオ応答 1 3 0 を生成して、通話セッションのMSR側が聞いていることを示す。

20

【0063】

通話セッションの進行中に、ユーザモデル 1 3 6 は、発呼者が個人化された応答を必要とする場合があると決定し、それらが発呼者との更なる対話を処理したいことをユーザインタフェーススイッチ 1 1 6 に通信することがある。周囲信号のバックグラウンドノイズ及びユーザモデル 1 3 6 に倣ってモデル化された合成された発話の故に、合成されたオーディオ応答 1 3 0 からユーザモデルオーディオ応答 1 2 8 への移行は、発呼者にシームレスに思われることがある。

30

【0064】

発話シンセサイザエンジン 1 1 4 は、発呼者感情データと応答論理エンジン 1 2 2 によって決定されるような応答信号とを受信し、ユーザモデルの音声パターンを使用して発話を合成して、合成されたオーディオ応答 1 3 0 を感情スコア 1 1 2 と調和される(matched)べき適切な情動変調(emotional modulation)と通信するように訓練された、人工知能(AI)モデルである。

【0065】

システム 1 0 0 は、図 2、図 4 及び図 5 に記載される方法に従って作動されてよい。

40

【0066】

図 2 を参照すると、発話合成会話エージェントを作動させる方法 2 0 0 は、通話セッション中にオーディオインタフェースを作動させて発呼者オーディオ信号を受信する(ブロック 2 0 2)。ブロック 2 0 4 において、方法 2 0 0 は、感情モデルによって構成される感情分析エンジンの作動を通じて発呼者オーディオ信号からの感情スコアを含むオーディオトランスクリプトを生成する。ブロック 2 0 6 において、方法 2 0 0 は、ユーザモデルからの入力を受信するように構成されるユーザインタフェーススイッチにオーディオトランスクリプトを通信する。ブロック 2 0 8 において、方法 2 0 0 は、ユーザインタフェーススイッチからの応答制御を、ユーザモデルからの履歴会話データで訓練された音声シン

50

セサイザエンジンに通信する。ブロック 210 において、方法 200 は、発話シンセサイザエンジンを作動させる。サブルーチンブロック 212 において、発話シンセサイザエンジンは、履歴会話データによって構成される応答論理エンジンの作動を通じて、オーディオトランスクリプト及び発呼者オーディオ信号についての応答信号を生成する。サブルーチンブロック 214 において、発話シンセサイザエンジンは、履歴会話データによって構成される発話合成モデルの作動を通じて、応答信号からの合成されたユーザモデル応答及び周囲信号を含む、合成されたオーディオ応答を生成する。ブロック 216 において、方法 200 は、通話セッション中にオーディオインタフェースを通じて、発呼者オーディオ信号に応答する合成されたオーディオ応答を通信する。

【0067】

図 3 は、発話シンセサイザエンジン 326 によって受信され且つ生成される情報を示すシステム 300 を図示している。システム 300 は、制御されたメモリデータ構造 324 と、感情分析エンジン 302 と、発話シンセサイザエンジン 326 とを含む。発話シンセサイザエンジン 326 は、応答論理エンジン 334 と、発話合成モデル 336 とを含む。制御されたメモリデータ構造 324 は、感情スコアを備えるオーディオトランスクリプトのようなスコア付き会話トランスクリプト 322 と、複数の通話セッションからのユーザモデルオーディオ応答に対応する発呼者オーディオ信号のような会話オーディオ 340 とを含む、履歴会話データ 304 のための格納場所として機能する。履歴会話データ 304 、特にユーザモデルオーディオ応答を含む会話オーディオ 340 は、発話合成モデル 336 を生成／改良するために発話シンセサイザエンジン 326 に提供されてよい。

【0068】

発呼者オーディオ信号 308 は、感情分析エンジン 302 、応答論理エンジン 334 、及び制御されたメモリデータ構造 324 に、wav サンプル 328 として提供されてよい。wav サンプル 328 は、感情分析エンジン 302 及び発話シンセサイザエンジン 326 による利用を容易にする非圧縮波形オーディオである波形オーディオファイルフォーマット (.wav) を参照する。感情分析エンジン 302 は、転写サービス 338 と、感情モデル 320 とを含む。感情分析エンジン 302 は、応答論理エンジン 334 に通信される発呼者オーディオ信号 308 から、感情スコア 312 を有するオーディオトランスクリプト 306 を生成する。

【0069】

応答論理エンジン 334 は、感情スコア 312 及び発呼者オーディオ信号 308 を有するオーディオトランスクリプト 306 を受信し、応答して、応答信号 344 を生成する。発話合成モデル 336 は、合成されたオーディオ応答 318 を生成するために応答信号 344 を備えて構成される。合成されたオーディオ応答 318 は、合成されたユーザモデル応答 342 と、周囲信号 316 とを含む。会話状態に依存して、応答論理エンジン 334 は、非言語応答 332 又は文言対発話応答 330 (text to speech response) を、発話合成モデル 336 に通信されるべき応答信号 344 として生成することがある。能動会話状態が識別されるならば、応答論理エンジン 334 は、発話合成モデル 336 によって利用されるべき文言対発話応答 330 を生成して、合成された発話 310 を生成する。受動会話状態が識別されるならば、応答論理エンジン 334 は、発話合成モデル 336 によって利用されるべき非言語応答 332 を生成して、リスニング応答キュー 314 (listening response cues) を生成する。

【0070】

システム 300 は、図 2、図 4 及び図 5 に記載される処理に従って作動させられてよい。

【0071】

図 4 を参照すると、発話合成会話エージェントを作動させる方法 400 は、オーディオインタフェースを通じて発呼者オーディオ信号を受信する (ブロック 402)。ブロック 404 において、方法 400 は、オーディオインタフェースを通じて発呼者オーディオ信号を受信することに応答して、ユーザモデルから、ユーザインタフェーススイッチを通じてユーザモデル入力を受信する。ブロック 406 において、方法 400 は、発呼者オーデ

ィオ信号に応答して、ユーザモデルオーディオ応答をオーディオインタフェースに通信し、ユーザモデルオーディオ応答は、応答オーディオと、バックグラウンドノイズとを含む。ブロック408において、方法400は、オーディオトランスクリプト、発呼者オーディオ信号、及びユーザモデルオーディオ応答を、制御されたメモリデータ構造内に履歴会話データとして格納する。発話シンセサイザエンジンの作動中、音声合成モデルは、発話合成モデルを作動させて、履歴会話データにおけるユーザモデル応答のバックグラウンドノイズから周囲信号を生成する。

【0072】

図5は、発話合成会話エージェントを作動させるプロセス500を図示している。ブロック502において、プロセス500は、電話スイッチボードのようなオーディオインタフェースを通じて発呼者オーディオを受信する。ブロック504において、プロセス500は、発呼者の意図及び感情を決定する。発呼者の意図及び感情の決定は、感情分析エンジンによって達成されてよい。意図の決定は、発呼者のオーディオの転写及び転写された単語の文脈(コンテキスト)の結果として達成されてよい。発呼者の感情の決定は、感情モデルの使用を通じて達成されてよい。ある場合には、発呼者の意図及び感情の決定は、感情スコアを有するオーディオトランスクリプトとして、ユーザインタフェーススイッチに通信されてよい。発呼者の意図／感情は、ユーザインタフェーススイッチに通信されてよく、ユーザインタフェーススイッチは、どのように処理するかに関する命令をユーザ(MSR)に促す(ブロック506)。ユーザからの命令は、ユーザモデル入力の形態においてユーザインタフェーススイッチに戻されてよい。発呼者の通信された意図及び感情に基づいて、ユーザは、命令を示してよい(決定ブロック508)。ユーザがアクション(行動)を示さないならば、発話合成会話エージェントは引き継ぎ、デフォルトのアクションセットを予め形成してよい(ブロック510)。ある場合には、発話合成会話エージェントによって実行されるデフォルトのアクションセットは、合成されたオーディオ応答で発呼者オーディオに応答することであってよい。命令は、ユーザによる承認を示して、発話合成会話エージェント(ロボット(bot))が発呼者オーディオに対する応答を処理することを可能にしてよい(決定ブロック512)。ロボットが発呼者オーディオに対する応答を処理することをユーザが許容したくないことをユーザが示すならば、ユーザは、ロボットがアイドルのままである間に、発呼者オーディオに応答してよい(ブロック514)。ロボットが発呼者オーディオへの応答を処理することをユーザが許容したいことをユーザが示すならば、ロボットは、合成されたオーディオ応答を生成するためのその通話管理シーケンスを開始してよい(ブロック516)。

【0073】

図6は、応答論理エンジン604のための教師付き訓練フェーズ600を図示している。教師付き訓練フェーズ600は、発呼者トランスクリプトに対する提案される応答608を応答論理エンジン604に促してよい。次に、ユーザモデル602は、提案される応答608を評価し、応答スコアリング606を応答論理エンジン604の訓練インタフェース610に通信してよい。

【0074】

図7は、応答論理エンジン706のための教師付き訓練フェーズ700を図示している。教師付き訓練フェーズ700の間に、ユーザモデル702は、訓練インタフェース704を通じて応答論理エンジン706を訓練するエキスパートユーザとして行動してよい。ユーザモデル702は、訓練インタフェース704を通じて複数の通話要約トランスクリプト724を応答論理エンジン706に通信してよい。訓練インタフェース704は、ユーザモデル702を含む通話セッションから、通話セッション意図及び感情スコア710並びに通話要約トランスクリプト724内の識別された実体712(entities)を受信してよい。通話セッションの意図及び感情スコア710及び識別された実体712は、ユーザモデル702によって手作業で入力される注釈として、通話要約トランスクリプト724内で識別されてよい。ユーザモデル702によって利用される注釈スタイルは、応答論理エンジン706による理解のために特に構成されてよい。オーディオトランスクリプトは

10

20

30

40

50

、ユーザモデル702からの転写されたユーザモデル応答718を含んでよい。応答論理エンジン706は、通話セッション意図及び感情スコア710と、識別された実体712と、ユーザモデル応答718とを利用して、発呼者オーディオ応答を生成するための会話モデルを構築してよい。ユーザモデル702は、通話トランスクリプトの一部分についての応答監査714(response audit)を応答論理エンジン706に通信してよい。応答論理エンジン706は、現在の会話の識別された応答状態716、応答オプション722、及び応答オプションについての確実性スコア708で応答してよい。次に、ユーザモデル702は、応答オプションを検証又は調整するためのフィードバック制御720を提供してよい。

【0075】

図8は、発話合成会話エージェントのための教師付き訓練フェーズ800を図示している。教師付き訓練フェーズ800の間に、オーディオインタフェース808は、感情モデル804と転写サービス802とを含む感情分析エンジン810に通信される発呼者オーディオ信号818を受信する。感情分析エンジン810の転写サービス802は、感情スコア816を含むオーディオトランスクリプト814を生成する。オーディオトランスクリプト814は、ユーザモデル702に通信され、ユーザモデル702は、それを応答論理エンジン706の訓練インタフェース704に通信する。応答論理エンジン706は、会話のための識別された応答状態716、ならびにユーザモデル702に通信される応答オプション722及び確実性スコア708を識別する。教師付き訓練フェーズ800の間に、発話シンセサイザエンジン812は、発話合成モデル806を備える発話シンセサイザエンジン812からの生成された応答を利用して、発呼者オーディオ信号818についての合成されたオーディオ応答を生成してよい。ユーザモデル702は、フィードバック制御720を提供して、オーディオインタフェース808に通信される応答を改良してよい。

【0076】

図9は、例示的な畳み込みニューラルネットワーク900(convolutional neural network)を図示している。畳み込みニューラルネットワーク900は、畳み込み層904(convolutional layer)内に視覚化されるように、そのニューロン(neurons)を三次元(幅、高さ、深さ)において配置する。畳み込みニューラルネットワーク900の各層は、3Dボリュームの入力を3D出力ボリュームのニューロン活性化(neuron activations)に変換する。この例において、入力層902は、画像を符号化するので、その幅及び高さは、画像の寸法(dimensions)であり、深さは、3(赤色、緑色、青色)チャンネル)である。畳み込み層904は、入力層902の出力を更に変換し、出力層906は、畳み込み層904の出力を画像コンテンツの1つ以上の分類に変換する。

【0077】

図10は、例示的な畳み込みニューラルネットワーク層1000をより詳細に図示している。画像の入力層領域1002領域の入力層領域1004のサブ領域の例が、畳み込み層1006内の畳み込み層のサブ領域1008のセットによって分析される。入力層領域1002は、 32×32 の長さ及び幅のニューロン(例えば、 32×32 ピクセル)であり、3ニューロンの深さ(例えば、ピクセル毎に3つのカラーチャンネル)である。畳み込み層1006内の各ニューロンは、空間的に(高さ及び幅において)入力層領域1002内のローカル領域(局所領域)にのみ接続されるが、全ての深さ(すなわち、入力画像であるならば、全てのカラーチャンネル)に接続される。入力層領域1002の入力層領域1004のサブ領域を分析する畳み込み層サブ領域1008の深さに沿って複数のニューロン(この例では5つのニューロン)があり、畳み込み層サブ領域1008の各ニューロンは、入力層領域1004のサブ領域のあらゆるニューロンから入力を受け取る場合があることに留意のこと。

【0078】

図11は、VGGネット1100として知られる畳み込みニューラルネットワーク(CNN)の民間に普及している形態を図示している。初期的な畳み込み層1102は、未加

10

20

30

40

50

工の画像ピクセルを格納し、最終的なプーリング層 1120 は、分類スコア(class score s)を決定する。処理経路に沿った中間畳み込み層（畳み込み層 1106、畳み込み層 1112、及び畳み込み層 1116）、及び整流器活性化（RELU層 1104、RELU層 1108、RELU層 1114、及びRELU層 1118）、及び中間プーリング層（プーリング層 1110、プーリング層 1120）の各々が、列(column)として示されている。

【0079】

VGG ネット 1100 は、基本 CNN の大きな単一層フィルタを、直列の複数の 3×3 サイズのフィルタに置き換える。所与の受容フィールド(receptive field)（出力が依存する入力画像の有効領域サイズ）を用いて、複数のスタックされたより小さいサイズのフィルタは、より大きなフィルタサイズを有する単一層よりも画像特徴分類(image feature classification)において良好に機能することがある。何故ならば、複数の非線形層はネットワークの深さを増加させ、それはネットワークがより複雑な特徴を学習することを可能にするからである。VGG ネット 1100 において、各プーリング層は、 2×2 だけであってよい。

【0080】

図 12 は、畳み込み層 1202 内のニューロンのグループからの出力を次の層 1206 内のニューロンに接続する畳み込み層フィルタリング 1200 を図示している。この例では 5×5 のニューロンのセットである畳み込み層 1202 のために、受容フィールドが画定される。受容フィールドの各ニューロンの集合的な出力は、重み付けされ、次の層 1206 内の単一のニューロンにマッピングされる。この重み付けマッピングは、畳み込み層 1202 のためのフィルタ 1204 と呼ばれる（或いは畳み込み層 1202 のカーネルと呼ばれることがある）。フィルタ 1204 の深さは、この例には図示されていない（すなわち、フィルタ 1204 は、実際には、図示のような正方形ではなく、畳み込み層 1202 内のニューロンの体積(cubic volume)である）。よって、図示されているものは、全フィルタ 1204 の「スライス」である。フィルタ 1204 は、入力画像の周りでスライドされるか或いは畳み込まれ(convolved)、各々は、次の層 1206 内の異なるニューロンに時間マッピングする(time mapping)。例えば、図 12 は、フィルタ 1204 がどのように 1 ユニット(unit)（「ストライド(stride)」）だけ右に段付けられ(steped)、上からわずかにオフセットされた受容フィールドを生成し、その出力を次の層 1206 内の次のニューロンにマッピングするかを示している。ストライドは、1 以外の他の数であることができ、しばしば、1 以外の数であることができ、より大きなストライドは、受容フィールドにおけるオーバーラップ(重なり合い)を減少させ、故に、次の層 1206 のサイズを更に減少させる。この段階的な方法で画定されることができる畳み込み層 1202 内のあらゆる特異な受容フィールドは、次の層 1206 内の異なるニューロンにマッピングされる。よって、畳み込み層 1202 がスライス毎に $32 \times 32 \times 3$ ニューロンであるならば、次の層 1206 は、畳み込み層 1202 の全ての受容フィールドをカバーするために $28 \times 28 \times 1$ ニューロンのみである必要がある。これは活性化マップ(activation map)又は特徴マップ(feature map)と呼ばれる。よって、フィルタリングによる層複雑性の低減がある。 5×5 フィルタが 32×32 畳み込み層 1202 上に特異に嵌合することができる 784 の異なる方法があるので、次の層 1206 は、 28×28 のみを必要とする。畳み込み層 1202 の深さも、次の層 1206 において 3 から 1 に減少される。

【0081】

CNN において使用される層の総数、畳み込み層の数、フィルタのサイズ、及び各層でのストライドについての値は、CNN の「ハイパーパラメータ(hyperparameters)」の例である。

【0082】

図 13 は、 2×2 の受容フィールドと、2 つの段(steps)とを備える、プーリング層関数 1300(pooling layer function)を図示している。プーリング層関数 1300 は、マックスプール(maxpool)プーリング技術の一例である。入力層 1302 の特定の受容フィ

10

20

30

40

50

ールド内の全てのニューロンの出力は、プーリング層 1304 内のそれらの出力のうちの最大値を有する出力によって置き換えられる。プーリングについての他のオプションは、平均プーリング及び L2 標準プーリング(L2-norm pooling)である。プーリング層を使用する理由は、特定の特徴が元の入力ボリューム（高い活性化値がある）においてひとたび認識されると、その正確な場所は他の特徴に対するその相対的な場所ほどに重要ではないからである。プーリング層は、ニューラルネットワーク内のその地点から前方で入力層 1302 の空間的寸法を劇的に減少させることができる（深さではなく、長さ及び幅が変化する）。これは 2 つの主な目的に資する。第 1 は、パラメータ又は重みの量が大幅に減少され、よって、演算コストが減少することである。第 2 は、それがオーバーフィッティング(overfitting)を制御することである。オーバーフィッティングとは、モデルが訓練例に合わせて調整されるので、それがライブデータセットに適用されるときに、それが十分に一般化することができないことを指す。

【0083】

図 14 は、WaveNet スタック及びその受容フィールドの視覚化を例示する図 1400 を図示している。発話シンセサイザエンジンは、WaveNet 自動回帰モデルから構成されてよい。図 1400 は、1 つの拡張(dilation)を伴って隠れ層 1404 内へ供給される底にある入力層 1402 を示している。隠れ層 1404 は、2 の拡張を伴って隠れ層 1406 に供給される。隠れ層 1406 は、4 の拡張を伴って隠れ層 1408 に供給される。隠れ層 1408 は、8 の拡張を伴って出力層 1410 に供給される。

【0084】

未加工のオーディオデータは、典型的には、非常に高次元（例えば、16 kHz のオーディオについて、毎秒 16,000 サンプル）であり、発話中単語又は音楽のメロディのような、数千の時間段階に亘る複雑な階層構造を含む。標準的な因果畳み込み層(causal convolution layers)でそのような長期依存性をモデリングすることは、十分に広い受容フィールドを確保するために非常に深いネットワークを必要とする。WaveNet は、受容フィールドが深さ共に指数関数的に成長することを可能にする拡張した因果畳み込みを用いることによって、この制約を回避する。

【0085】

WaveNet は、深い畳み込みニューラルネットワーク(CNN)として知られるフィールドフォワードニューラルネットワークの一種である。これらは、脳のニューロンに幾分似た相互に連結したノード(結節)の層からなる。WaveNet において、CNN は、未加工の信号を入力として取り、一度に 1 つのサンプル毎に出力を合成する。それは、 μ 則複合変換を用いて符号化され、256 の可能な値に量子化される、信号値のソフトマックス（すなわち、カテゴリー的な）分布からのサンプリングによって行われる。

【0086】

「CTC 損失関数(CTC loss function)」とは、タイミングが可変であるシーケンス問題に取り組むために、LSTM ネットワークのようなりカレントニューラルネットワーク(RNN)を訓練するための、ニューラルネットワーク出力及び関連スコアリング関数の一種である、コネクショニスト時間分類(connectionist temporal classification)を指す。CTC ネットワークは、ラベルの確率をモデル化する訓練を通じて適合される連続出力（例えば、ソフトマックス）を有する。CTC は、境界及びタイミングを学習しようとはしない。すなわち、ラベルシーケンス(label sequences)が、空白を無視して整列(アラインメント)においてのみ異なるならば、ラベルシーケンスは、等価とみなされる。等価なラベルシーケンスは、多くの方法で生じることができ、それはスコアリングを簡単でない作業とする。幸いなことに、そのための効率的なフォワード・バックワードアルゴリズムがある。次に、CTC スコアは、逆伝搬アルゴリズムと共に使用して、ニューラルネットワーク重みを更新することができる。CTC 適合ニューラルネットワークに対する代替的なアプローチは、隠れマルコフモデル(HMM: hidden Markov model)を含む。

【0087】

「ゲート付き回帰型ユニット (GRU: Gated Recurrent Unit)」とは、リカレントニューラルネットワークにおけるゲート機構を意味する。GRUは、LSTMよりも小さいデータセットでより良い性能を示すことがある。それらはLSTMよりも少ないパラメータを有する。何故ならば、それらには出力ゲートがないからである。https://en.wikipedia.org/wiki/Gated_recurrent_unitを参照

【0088】

「ビーム検索 (beam search)」とは、限定されたセット内の最も有望なノードを拡張することによってグラフを探索する発見的検索アルゴリズム (heuristic search algorithm) を指す。ビーム検索は、メモリ要求を減らす最良優先検索 (best-first search) の最適化である。最良優先検索は、ある発見的方法に従って全ての部分解 (状態) を順序付けるグラフ検索である。しかしながら、ビーム検索では、所定の数の最良部分解のみが候補として保持される。よって、それは欲張りなアルゴリズムである。ビーム検索は、その検索ツリーを構築するために幅優先検索を使用する。ツリーの各レベルで、それは現在のレベルで状態のすべての後継を生成し、発見的コストの増加する順序でそれらをソートする。しかしながら、それは (ビーム幅と呼ばれる) 各レベルで最良の状態の所定の数 β のみを格納する。それらの状態のみが次に拡張される。ビーム幅が大きいほど、より少ない状態がプルーニングされる (pruned)。ビーム幅が無限大であるならば、状態はプルーニングされず、ビーム検索は幅優先検索と同一である。ビーム幅は、検索を実行するために必要とされるメモリを制限する。目標状態 (goal state) は、潜在的にプルーニングされ得るので、ビーム検索は、完全性 (解が存在するならば、アルゴリズムが解で終了するという保証) を犠牲にする。ビーム検索は、最適ではない (すなわち、それが最良の解を見出す保証はない)。一般的に、ビーム検索は、見出された最初の解を返す。機械翻訳のためのビーム検索は、異なる場合である。設定された最大検索深さ (すなわち、翻訳長) にひとたび達すると、アルゴリズムは、様々な深さで検索に見出される解を評価し、最良の解 (最も高い確率を有する解) を返す。ビーム幅は、固定されることができ、或いは可変であることができる。可変ビーム幅を使用する1つのアプローチは、最小にある幅で開始する。解が見出されないならば、ビームは広げられ、手順を繰り返される。

【0089】

「アダム最適化 (Adam optimizer)」とは、訓練データに基づいてネットワーク重みを反復的に更新するために、古典的な確率的勾配降下手順 (stochastic gradient descent procedure) の代わりに使用することができる最適化アルゴリズムを指す。確率的勾配降下は、全ての重み更新について (アルファと呼ばれる) 単一の学習速度を維持し、学習速度は訓練中に変化しない。学習速度は、各ネットワーク重み (パラメータ) について維持され、学習が展開するにつれて別々に適応される。アダムは、確率的勾配降下の2つの他の拡張の利点を組み合わせる。具体的には、疎らな勾配の問題 (例えば、自然言語及びコンピュータビジョン問題) に対するパフォーマンスを改善するパラメータ毎の学習速度を維持する適合勾配アルゴリズム (AdaGrad: Adaptive Gradient Algorithm)、及び重みについての勾配の最近の大きさの平均 (例えば、それがどれくらい素早く変化しているか) に基づいて適合されるパラメータ毎の学習速度も維持する二乗平均平方根伝搬 (RMSProp: Root Mean Square Propagation) である。これはアルゴリズムがオンライン及び非定常問題 (例えば、ノイズ) に対してうまく機能することを意味する。アダムは、AdaGrad及びRMSPropの両方の利益を実現する。RMSPropにおけるように平均一次モーメント (平均) に基づいてパラメータ学習速度を適応させる代わりに、アダムは、勾配の二次モーメントの平均 (中心化されない分散 (uncentered variance)) も利用する。具体的には、アルゴリズムは、勾配及び二乗勾配の指数移動平均を計算し、パラメータベータ1及びベータ2は、これらの移動平均の減衰率を制御する。移動平均の初期値並びに (推奨される) 1.0に近いベータ1及びベータ2の値は、ゼロに向かうモーメント推定のバイアスを生む。このバイアスは、バイアスのかかった推定値を先ず計算し、バイアス補正された推定値を次に計算することによって、克服される。

【0090】

本明細書において、「1つの実施形態」又は「ある実施形態」への言及は、同じ実施形態を必ずしも言及しないが、それらは同じ実施形態を指すことがある。文脈が明らかに他のことを要求しない限り、明細書及び特許請求の範囲を通じて、「含む(comprise)」、「含む(comprising)」、及び同等の用語は、排他的又は網羅的な意味ではなく、包括的な意味で、すなわち、「含むが、それに限定されない」という意味で解釈されるべきである。単数又は複数の数を使用する語は、単一のもの又は複数のものに明らかに限定されない限り、それぞれ、複数の又は単数の数を含む。加えて、この出願において使用されるとき、「本明細書において(herein)」、「上記(above)」、「下記(below)」、及び類似の意味を有する語は、この出願全体を指し、この出願の如何なる特定の部分も指さない。請求項が2以上の項目のリストに関して「又は(or)」という語を使用するとき、その語は、明示的に一方に限定されない限り、その語の以下の解釈、すなわち、リスト中の項目のうちのいずれか、リスト中の項目の全て、及びリスト中の項目の任意の組み合わせの全てをカバーする。本明細書中で明確に定義されない用語は、当業者によって一般的に理解されるような、それらの従来的な意味を有する。

10

【0091】

本明細書に記載する様々な論理関数演算は、前記演算又は関数を反映する名詞又は名詞句を使用して参照される論理で実装されてよい。例えば、関連演算(association operation)は、「連想装置(associator)」又は「相関器(correlator)」によって実行されてよい。同様に、切り替え(スイッチング)は、「スイッチ」、「セレクト」による選択などによって実行されてよい。

20

【0092】

開示の主題事項は、その適用において、構造の詳細及び以下の記述に記載する或いは図面に図示する構成要素の配置に限定されないことが理解されるべきである。開示の主題事項は、他の実施形態が可能であり、様々な方法で実施され且つ実行されることができる。また、本明細書で使用する言い回し及び用語は、記載の目的のためであり、限定として解釈されてならないことが理解されるべきである。

【0093】

よって、当業者は、この開示の基礎となる概念が、開示の主題事項の幾つかの目的を実行するための他の構造、システム、方法及び媒体を設計するための基礎として容易に利用される場合があることを理解するであろう。従って、特許請求の範囲は、それらが開示の主題事項の精神及び範囲から逸脱しない限り、そのような均等な構成を含むものとみなされるべきことが重要である。

30

【0094】

開示の主題事項は、前述の例示的な実施形態において記載され且つ図示されるが、本開示は一例として行われているに過ぎないこと、並びに開示の主題事項の実装の詳細における多数の変更が後続の特許請求の範囲によってのみ限定される開示の主題事項の精神及び範囲から逸脱することなく行われる場合があることが理解されるであろう。

【0095】

感情分析エンジンの1つのタイプは、2016年12月12日に出願された特許文献1「Method and system for achieving emotional text to speech」に完全に記載されている。記載のシンセサイザは、任意の数の知られている感情分析エンジン／アルゴリズムから導出される感情スコアを利用することがある。

40

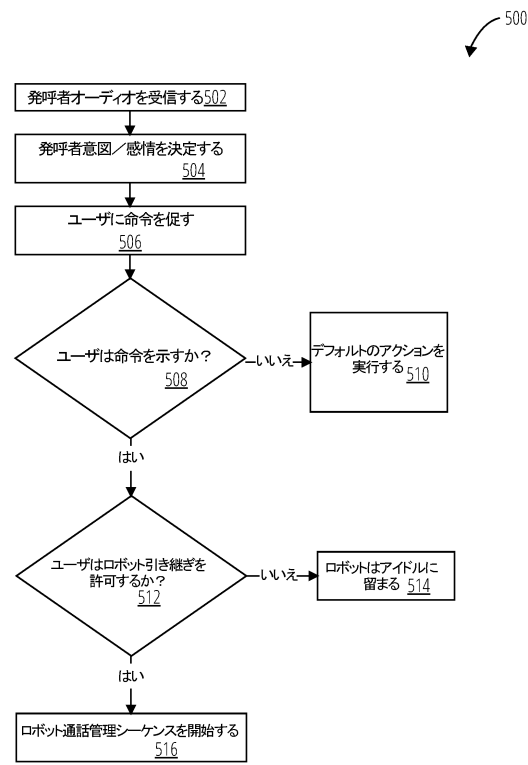
【先行技術文献】

【特許文献】

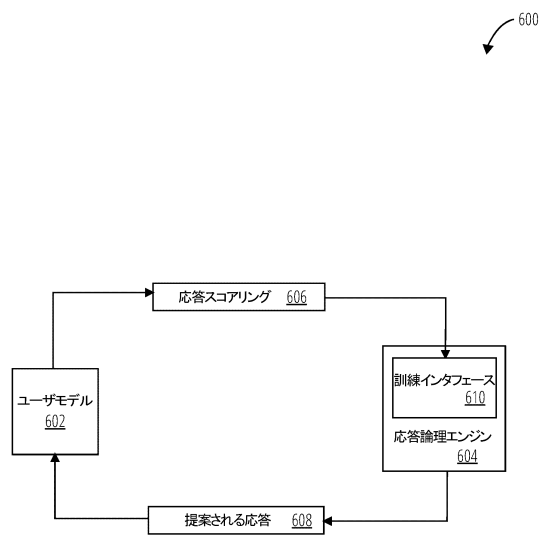
【0096】

【文献】米国特許出願第15／375,634号明細書

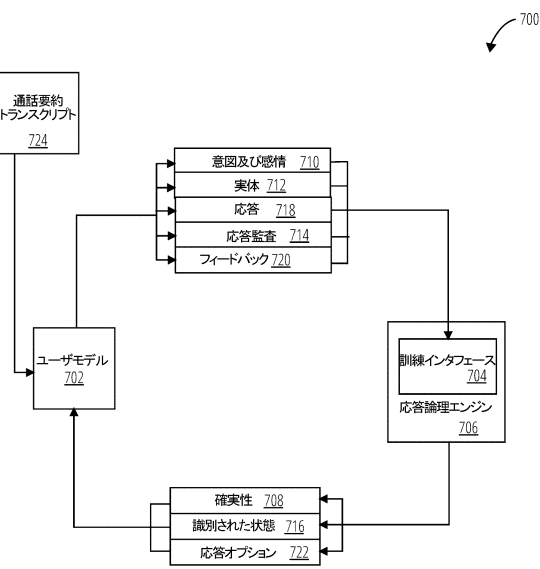
【図 5】



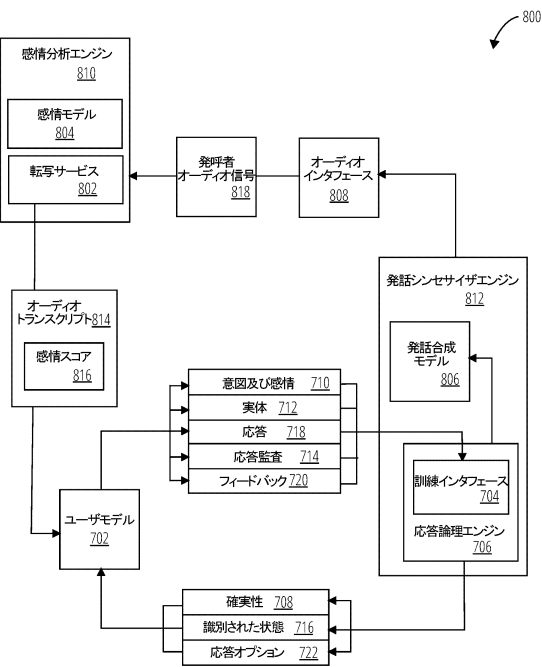
【図 6】



【図 7】



【図 8】



10

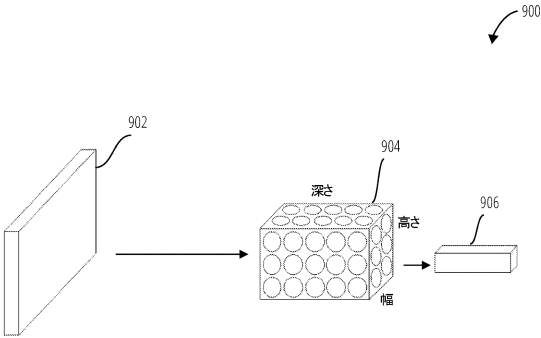
20

30

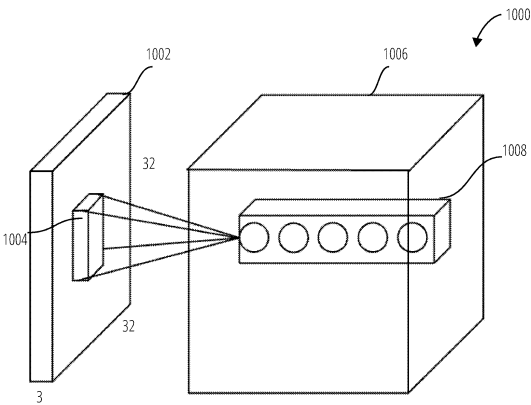
40

50

【図 9】

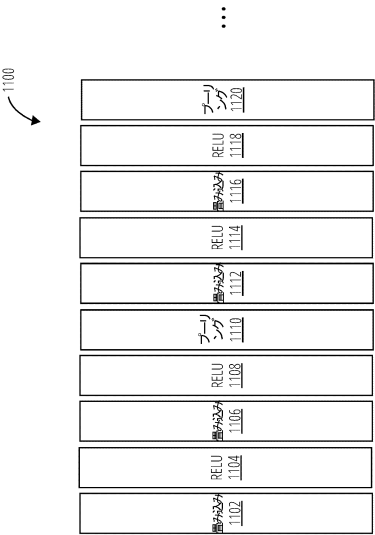


【図 10】

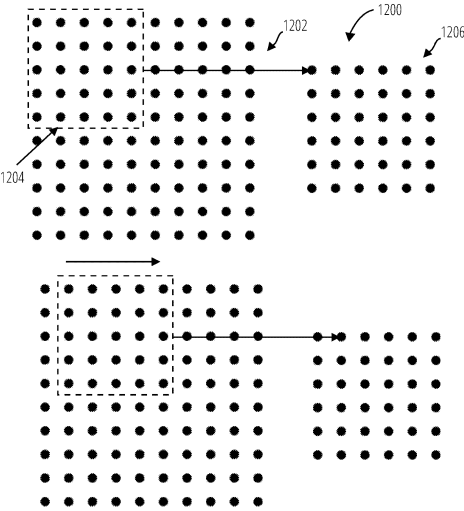


10

【図 11】



【図 12】



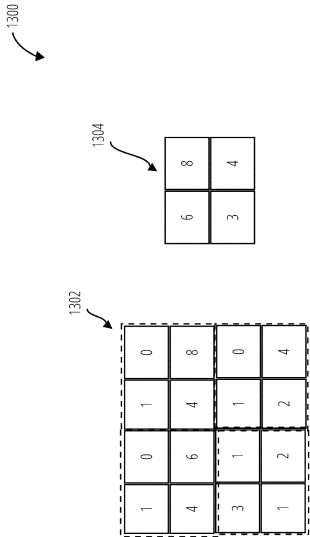
20

30

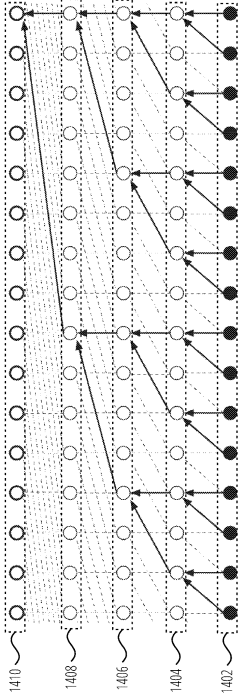
40

50

【図 1 3】



【図 1 4】



10

20

30

40

50

フロントページの続き

- ル アヴェニュー 1142
- (72)発明者 バロン, アレックス
アメリカ合衆国 カリフォルニア州 90212, ビバリーヒルズ サウス ペック ドライヴ 133
アパートメント 301
- (72)発明者 コーエン, デヴィッド
アメリカ合衆国 ワシントン州 98052, レッドモンド ノースイースト 152 アヴェニュー
2690 ユニット 623
- (72)発明者 マック ミラン, エヴァン
アメリカ合衆国 カリフォルニア州 94133, サンフランシスコ モンゴメリー ストリート 1
460 アpartment 2
- 審査官 大野 弘
- (56)参考文献 特開2005-062240 (JP, A)
- (58)調査した分野 (Int.Cl., DB名)
- G10L 13/00
G06N 20/00
G10L 15/10
G10L 15/00
G06F 3/16