

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7350376号

(P7350376)

(45)発行日 令和5年9月26日(2023.9.26)

(24)登録日 令和5年9月15日(2023.9.15)

(51)国際特許分類

F I

G 0 9 B 9/00 (2006.01)

G 0 9 B 9/00

Z

G 0 6 N 20/00 (2019.01)

G 0 6 N 20/00

G 0 6 Q 50/20 (2012.01)

G 0 6 Q 50/20

請求項の数 2 (全26頁)

(21)出願番号 特願2022-8465(P2022-8465)
(22)出願日 令和4年1月24日(2022.1.24)
(65)公開番号 特開2023-77365(P2023-77365A)
(43)公開日 令和5年6月5日(2023.6.5)
審査請求日 令和4年1月24日(2022.1.24)
(31)優先権主張番号 10-2021-0163052
(32)優先日 令和3年11月24日(2021.11.24)
(33)優先権主張国・地域又は機関
韓国(KR)
特許法第30条第2項適用 (1)掲載年月日: 2021年11月16日 (2)掲載アドレス: <https://ieeexplore.ieee.org/document/9617584>

(73)特許権者 505224569
インハ インダストリー パートナーシップ
インスティテュート
Inha-Industry Partnership Institute
大韓民国 22212 インチョン ミチ
ヨル-グ インハ-ロ 100
(74)代理人 110001519
弁理士法人太陽国際特許事務所
(72)発明者 チョ・グンシク
大韓民国 22212 インチョン ミチ
ヨル-グ インハ-ロ 100
審査官 安田 明央

最終頁に続く

(54)【発明の名称】 メタバースプラットフォームを利用した複合現実における複雑な機械類の訓練、教育、および広告システム

(57)【特許請求の範囲】

【請求項1】

航空機をはじめとする機械類の整備訓練、教育、および広報のためのメタバース複合現実(Mixed Reality)において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン(Digital Twin)を提供して3Dシミュレーションを実行するシミュレーション実行部、

2Dマニュアル、前記2Dマニュアルの作業指示書、およびSCM(Simulation Cost Model)を含む学習情報に基づいて人工知能知識(Knowledge)を提供する学習部、および

前記提供された人工知能知識、前記デジタルツインに基づいて前記3Dシミュレーションを実行するために音声要請を処理するためのニューラルネットワーク作業と言語的推論を実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知するニューラル言語音声実行部

を含み、

前記ニューラル言語音声実行部は、

スマートグラスを着用した使用者が自身のオーディオ要請を録音するようにニューラル言語音声実行部をトリガーし、ニューラル言語音声実行部がトリガー構文を感知して動的長オーディオ録音アルゴリズムを呼び出し、マイクから出力される音声信号ストリームでオーディオデータを生成するように処理する動的長オーディオ録音機、

前記オーディオデータをテキストに変換して、自動音声認識のための音声 - テキストの形態でテキスト - プログラムネットワークに伝達する音声テキストネットワーク（前記音声 - テキストネットワークは、自動音声認識ニューラルネットワーク（Automatic Speech Recognition neural network）である）、前記音声 - テキストを各ドメイン言語の実行可能なプログラムシーケンスに変換するための関数と媒介変数で構成されたテキスト - プログラムネットワーク、および
 使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する言語プログラム実行機、
 を含み、

前記動的長オーディオ録音アルゴリズムは、
 所定の第 1 時間ごとに実行されるループのすべての繰り返しに対して、前記音声信号ストリームの最初のチャンクのスペクトル平均と前記音声信号ストリームの各チャンクのスペクトル平均の差を計算し、
 前記差が閾値より小さい場合は、沈黙カウンタを増加させ、
 前記沈黙カウンタが予め決定された最大沈黙カウンタに到達すると、録音を中止して前記オーディオデータを生成するように具現され、

前記テキスト - プログラムネットワークは、
 テキストの単語を教育データセットの単語と一致させるための一般語彙（General Vocabulary）を利用して要請ベクトルに変換し、前記要請ベクトルをプログラムベクトルに変換し、前記プログラムベクトルは、少なくとも 2 つのプログラムの生成に使用されるドメイン特化言語（Domain Specific Language）の構成要素に対する参照を含み、

前記言語プログラム実行機は、実行すべき前記少なくとも 2 つのプログラムを入力として使用し、

前記少なくとも 2 つのプログラムは、第 1 プログラムおよび第 2 プログラムを含み、
 前記第 1 プログラムは、前記テキストから数字を探索して返還する FindObject 関数と要請媒介変数を含み、

前記言語プログラム実行機は、前記第 1 プログラムを実行して前記 FindObject 関数の返還値として第 1 数字を取得し、

前記第 2 プログラムは、特定の種類のマニュアルを表示し、その中に数字を強調する ShowManual 関数と、マニュアルの種類を表す媒介変数および強調される数字を表す媒介変数とを含み、

前記言語プログラム実行機は、前記第 2 プログラムを実行して前記テキストから把握される第 1 種類のマニュアルを表示し、前記第 1 数字を強調して表示するよう決定する、

メタバースプラットフォームを利用した複合現実における訓練、教育、および広告システム。

【請求項 2】

航空機をはじめとする機械類の整備訓練、教育、および広報のためのメタバース複合現実（Mixed Reality）において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン（Digital Twin）をシミュレーション実行部が提供して 3 D シミュレーションを実行する段階、

2 D マニュアル、前記 2 D マニュアルの作業指示書、および SCM（Simulation Cost Model）を含む学習情報に基づいて人工知能知識（Knowledge）を学習部が提供する段階、および

前記提供された人工知能知識、前記デジタルツインに基づいて前記 3 D シミュレーションを実行するためにニューラル言語音声実行部が音声要請を処理するためのニューラルネットワークモデルとシンボリック（Symbolic）な人工知能知識推論をとともに実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する段階、

10

20

30

40

50

を含み、
 前記提供された 2 D マニュアルおよび 3 D マニュアルに基づいてシミュレーションを実行するために、音声要請を処理するためのニューラルネットワークの作業と言語的推論をニューラル言語音声実行部で実行する段階は、
 スマートグラスを着用した使用者が自身のオーディオ要請を録音するようにニューラル言語音声実行部をトリガーし、ニューラル言語音声実行部がトリガー構文を感知して動的長オーディオ録音アルゴリズムを呼び出し、マイクから出力される音声信号ストリームでオーディオデータを生成するように処理する段階、
 音声 - テキストネットワークを介して前記オーディオデータをテキストに変換して、自動音声認識のための音声 - テキストの形態でテキスト - プログラムネットワークに伝達する段階（前記音声 - テキストネットワークは、自動音声認識ニューラルネットワーク（Automatic Speech Recognition neural network）である）、
 関数と媒介変数で構成されたテキスト - プログラムネットワークを介して前記音声 - テキストを人工知能ドメイン知識に基づいて実行可能なプログラムシーケンスに変換する段階、および
 言語プログラム実行機が使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する段階、
 を含み、
 前記動的長オーディオ録音アルゴリズムは、
 所定の第 1 時間ごとに実行されるループのすべての繰り返しに対して、前記音声信号ストリームの最初のチャンクのスペクトル平均と前記音声信号ストリームの各チャンクのスペクトル平均の差を計算し、
 前記差が閾値より小さい場合は、沈黙カウンタを増加させ、
 前記沈黙カウンタが予め決定された最大沈黙カウンタに到達すると、録音を中止して前記オーディオデータを生成するように具現され、
 前記関数と媒介変数で構成されたテキスト - プログラムネットワークを介して前記音声 - テキストを各ドメインの言語の実行可能なプログラムシーケンスに変換する段階は、
 テキストの単語を教育データセットの単語と一致させるための一般語彙（General Vocabulary）を利用して要請ベクトルに変換し、前記要請ベクトルをプログラムベクトルに変換し、前記プログラムベクトルは、少なくとも 2 つのプログラムの生成に使用されるドメイン特化言語（Domain Specific Language）の構成要素に対する参照を含み、
 前記言語プログラム実行機は、実行すべき前記少なくとも 2 つのプログラムを入力として使用し、
 前記少なくとも 2 つのプログラムは、第 1 プログラムおよび第 2 プログラムを含み、
 前記第 1 プログラムは、前記テキストから数字を探索して返還する Find Object
 関数と要請媒介変数を含み、
 前記言語プログラム実行機は、前記第 1 プログラムを実行して前記 Find Object
 関数の返還値として第 1 数字を取得し、
 前記第 2 プログラムは、特定の種類のマニュアルを表示し、その中に数字を強調する Show Manual 関数と、マニュアルの種類を表す媒介変数および強調される数字を表す媒介変数とを含み、
 前記言語プログラム実行機は、前記第 2 プログラムを実行して前記テキストから把握される第 1 種類のマニュアルを表示し、前記第 1 数字を強調して表示するよう決定する、

メタバースプラットフォームを利用した複合現実における訓練、教育、および広報方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、航空機のような複雑な機械類のレガシーマニュアル、3 D モデル、3 D シミ

ユーレータ、および（例えば、ボーイング737型機の整備訓練および教育）整備知識を提供するためのメタバースプラットフォームに関する。

【背景技術】

【0002】

エクステンデッド・リアリティ（Extended Reality：XR）とは、すべての現実世界と、バーチャル・リアリティ（Virtual Reality：VR）、拡張現実（AR）、拡張仮想（Augmented Virtuality：AV）、複合現実（Mixed Realities：MR）、音声認識などの仮想世界を融合することで現実にはないものを知覚できるようにする技術であり、人間の暮らしと産業の多様な側面において膨大な価値をもたらす。ここには、障害者支援、教育過程の改善、産業プロセスの緩和なども含まれる。

10

【0003】

従来技術では、ARと音声認識技術を利用して聴覚障害者との意思疎通の問題を解決する方法を提案した。会話の聴取中にリアルタイム拡張現実「リアルタイム字幕」を生成して展示することで聴覚障害者にナレータを提供することは、聴覚障害者には視覚的な方法で環境を感じることを可能にし、手話ができない健常者との間にはばかり意思疎通の障壁の克服を可能にした。

【0004】

教育分野においては、ARと音声認識技術を組み合わせることにより、学習から利益が得られるようにした。その例として新たな言語学習を挙げることができる。ここでの拡張現実では、非ネイティブスピーカーの言語学習過程において、生徒の経験と知識の取得に影響を及ぼす向上された環境を提供することができる。音声認識とAR学習によって興味を促進させ、基本カラーに対する単語学習、3D形態、および空間客体に関する学習のような特定の作業を、生徒が迅速かつ容易に対処できるようにサポートする。

20

【0005】

XRと音声認識により、産業過程の多様な作業を自動化したり改善したりもできる。XRは、作業プロセスをシミュレーションしたりデジタル化したりすることを可能にし、音声命令は、運営を制御しながら時間を節約することで柔軟、効率的、さらには経済的な形態の通信を可能にする。

【0006】

従来技術では、リフティング装置を制御するためのARおよび音声インタフェース実現概念を提示したが、これにより、クレーン作業のために現場に物理的に存在する必要がなくなった。他の例としては、実際の航空機の代わりに航空機のデジタルツインが使用される音声命令により、MR航空機の整備を簡単に行えるようにした。

30

【先行技術文献】

【非特許文献】

【0007】

【文献】[1] K. Yi, J. Wu, C. Gan, A. Torralba, P. Kohli, and J. B. Tenenbaum, "Neural-symbolic vqa: Disentangling reasoning from vision and language understanding," in Advances in Neural Information Processing Systems, 2018, pp. 1039-1050.

40

C. Han, J. Mao, C. Gan, J. Tenenbaum, and J. Wu, "Visual concept-metaconcept learning," in Advances in Neural Information Processing Systems, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alche-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: <https://proceedings.neur>

50

ips.cc/paper/2019/file/98d8a23fd60826a2a474c5b4f5811707-Paper.pdf
J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu, "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision," in 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6 - 9, 2019. OpenReview.net, 2019. [Online]. Available: <https://openreview.net/forum?id=rJgMlhRctm>

10

【発明の概要】

【発明が解決しようとする課題】

【0008】

本発明は、レガシーマニュアル、3Dモデル、3Dシミュレータ、および航空機の整備知識を提供するボーイング737型機の整備訓練および教育のためのメタバースを提供することを技術的課題とする。また、メンテナンスマニュアルが厳格に遵守されるメタバースの運営フローを探索して制御するためのコンテキスト認識音声理解モジュールNSS E (Neuro-Symbolic Speech Executor)を提供することを他の技術的課題とする。

【課題を解決するための手段】

20

【0009】

メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広告システムを提供する。本発明に係るメタバースを利用した複合現実における複雑な機械類の訓練、教育、および広告システムは、航空機をはじめとする機械類の整備訓練、教育、および広報のためのメタバース複合現実 (Mixed Reality) において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン (Digital Twin) を提供して3Dシミュレーションを実行するシミュレーション実行部、2Dマニュアル、前記2Dマニュアルの作業指示書、およびSCM (Simulation Cost Model) を含む学習情報に基づいて人工知能知識 (Knowledge) を提供する学習部、および前記提供された人工知能知識、前記デジタルツインに基づいて前記3Dシミュレーションを実行するために音声要請を処理するためのニューラルネットワーク作業と言語的推論を実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知するニューラル言語音声実行部を含む。

30

【0010】

本発明の実施形態に係る前記ニューラル言語音声実行部は、スマートグラスを着用した使用者が自身のオーディオ要請を録音するようにニューラル言語音声実行部をトリガーし、ニューラル言語音声実行部がトリガー構文を感知して動的長オーディオ録音アルゴリズムを呼び出し、マイクから出力される音声信号ストリームでオーディオデータを生成するように処理する動的長オーディオ録音機、前記オーディオデータをテキストに変換して自動音声認識のための音声-テキストの形態でテキスト-プログラムネットワークに伝達する音声-テキストネットワーク (前記音声-テキストネットワークは、自動音声認識ニューラルネットワーク (Automatic Speech Recognition neural network) である)、前記音声-テキストを各ドメインの言語の実行可能なプログラムシーケンスに変換するための関数と媒介変数で構成されたテキスト-プログラムネットワーク、および使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する言語プログラム実行機を含む。

40

【0011】

前記テキスト-プログラムネットワークは、テキストの単語を教育データセットの単語と一致させるための一般語彙 (General Vocabulary) を利用して要請

50

ベクトルに変換して前記要請ベクトルをプログラムベクトルに変換し、前記プログラムベクトルは、プログラムの生成に使用されるドメイン特化言語 (Domain Specific Language) の構成要素に対する参照を含む。

【0012】

前記言語プログラム実行機は、実行すべきプログラムを入力として使用し、各プログラムは関数および該当の媒介変数で構成され、与えられたプログラムの各プログラムに対して繰り返しを入力すれば関数および媒介変数が抽出され、以前の繰り返しの結果を説明する変数 (Prev) が媒介変数に追加され、関数と媒介変数の準備が整えば、Execute関数は各関数を呼び出して抽出された媒介変数を伝達し、各関数は返還値を有するため各繰り返しで変数 (Prev) がアップデートされ、前記手順をすべてのプログラムに適用する。

10

【0013】

また他の側面において、本発明で提案する、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広報方法は、航空機を含む機械類の整備訓練、教育、および広報のためのメタバース複合現実 (Mixed Reality) において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン (Digital Twin) をシミュレーション実行部が提供して3Dシミュレーションを実行する段階、2Dマニュアル、前記2Dマニュアルの作業指示書、およびSCM (Simulation Cost Model) を含む学習情報に基づいて人工知能知識 (Knowledge) を学習部が提供する段階、および前記提供された人工知能知識、前記デジタルツインに基づいて前記3Dシミュレーションを実行するためにニューラル言語音声実行部が音声要請を処理するためのニューラルネットワーク作業と言語的推論を実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する段階を含む。

20

【発明の効果】

【0014】

本発明の実施形態に係るコンテキスト認識音声理解モジュールNSSE (Neuro-Symbolic Speech Executor) は、従来の音声認識方式とは異なり、ニューラルネットワークと伝統的な言語的推論を結合したニューラル言語 (Neuro-Symbolic) AIを適用することで、コンテキストと航空機関連知識を土台に使用者の要請と応答を理解することができる。また、提案するメタバースを利用した航空機の整備訓練、教育方法、およびシステムは、高価な物理的航空機を容易に修正およびアップデートすることのできる仮想航空機に代替可能であることから、航空技術のための安価かつ拡張可能なソリューションであると言える。さらに、現場の専門家の代わりとなるNSSEは、航空機の整備に対する効果的な訓練および教育を容易に行うことができるため、技術的な指導とすべての資源を提供することができる。

30

【図面の簡単な説明】

【0015】

【図1】本発明の一実施形態における、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広告システムの構成を示した図である。

40

【図2】本発明の一実施形態における、航空機整備メタバースの1人称ビュースナップショットの例示図である。

【図3】本発明の一実施形態における、航空機整備マニュアル3Dシミュレータの例示図である。

【図4】本発明の一実施形態における、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広報方法を説明するためのフローチャートである。

【図5】本発明の一実施形態における、ニューラル言語音声実行部の動作過程を示した図である。

【図6】本発明の一実施形態における、動的長オーディオ録音アルゴリズムを示した図である。

50

【図 7】本発明の一実施形態における、テキスト - プログラムネットワークの動作過程を示した図である。

【図 8】本発明の一実施形態における、テキスト - プログラムネットワークのアクテクチャを示した図である。

【図 9】本発明の一実施形態における、言語プログラム実行機の動作過程を示した図である。

【図 10】本発明の一実施形態における、言語プログラム実行機アルゴリズムを示した図である。

【図 11】本発明の一実施形態における、サンプル使用者の要請から結果までの過程を示した例示図である。

10

【図 12】本発明の一実施形態における、ニューラル言語音声実行部のコンテキスト管理を説明するための図である。

【図 13】本発明の一実施形態における、ニューラル言語音声実行部のアクテクチャを示した図である。

【発明を実施するための形態】

【0016】

X Rに音声認識を開発して内蔵するためには、使用者要請の特性、音声通信アドレス関数、およびアプリケーションが構築された環境を理解することが重要となる。例えば、「再生」、「停止」、「次のイメージ」のような事前の定義、静的および短文の命令で構成された音声要請であれば、スマートグラスホロレンズのようなモバイル機器のオフライン内臓音声制御によって簡単に処理することができる。

20

【0017】

しかし、文章が長く、意味構造が柔軟であり、同じ関数を参照する使用者の要請には、分類ニューラルネットワークが適用される。例えば、「次の客体を提示」、「次の客体を表示」、「次の項目に移動」などの命令は、次の客体を順に表示する動作にマッピングされるため、分類モデルは、音声信号関数を事前に定義された範疇のうちの1つの動作クラスにマッピングされるようになる。一般的に、C N N基盤のニューラルモデル構造は、データ関数を抽出する機能によってオーディオ分類に活用される。オーディオ関数のタイプによって1 Dまたは2 D畳み込みフィルタが使用される。原始オーディオ信号処理の場合は1 D畳み込みが適用され、M F C Cまたはログスペクトル特徴の場合は2 Dが使用される。同じように、英語と韓国語からなる音声データからM F C C関数を抽出して特化動作を実行する8個のクラスのうちの1つに変換する従来技術である、オーダーメイド型二重言語C N Nニューラルネットワークに基づいて音声命令の実現に焦点を合わせた。例えば、「チュートリアル（使用指針）ビデオ再生」を選択すれば、メディアプレイヤーが参照ビデオを開始させる。ネットワークは、オーディオM F C C関数を使用して作業クラスと識別言語の2つの結果を生成する。ここで、音声通信は、アプリケーションの運営関数を呼び出す。

30

【0018】

X Rシステムで音声表記が必要な場合があるが、信号波をシーケンスにマッピングするための音響モデルを構築した自動音声認識（Automatic Speech Recognition: A S R）技法に基づいてオーディオ信号をテキストに変換する。従来技術では、原始オーディオを入力として受けて音声表現を計算する完全畳み込みモデルを適用したし、他の従来技術では、回帰型ニューラルネットワーク（Recurrent Neural Network: R N N）を活用した。さらに他の従来技術では、アテンション基盤のデコーダとL S T M基盤の言語モデルの予測を結合することによって結果を得た。X Rシステムでは、このようなネットワークを活用するためには空間と処理能力が必要となるため、実際のモバイル機器ではなくオンラインで使用される。それにもかかわらず、分類またはA S Rネットワークのいずれもニューラルモデルの推論は各状況の情報に依存しない。しかし、本論文では、脈絡が重要な場合の事例と要求について扱った。

40

【0019】

50

本発明では、MRO (Maintenance Repair Operation) のうちでもオペレーション上のエラーによって破壊的かつ致命的な結果を招来する恐れがあるため整備士 (言い換えれば、使用者) が厳格に順守すべき法律文書である航空機整備マニュアルを考慮する。したがって、音声通信とともにオペレーションを制御するためには、各状況の情報を示すマニュアルとの強靱な関係が必要となる。マニュアルには、作業、下位作業、指針、航空機部品、2Dマニュアル、3D客体、ツール、警告、注意などに関する知識と階層構造がある。文書の全項目を連結することにより、メンテナンス過程中に参照しなければならない知識グラフを生成する。したがって、単純構造のディープラーニングネットワークによる音声通信および制御の開発は、マニュアルではすべてのリソースと関係进行处理することができず、会議中に脈絡を考慮することができない。音声相互作用システムは、脈絡情報に基づいて推論を行い、ニューラルネットワークのパターン認識能力を補う論理基盤部分を全般的に必要とする。ニューラルネットワーク分野の近年の発展 (言い換えれば、ニューラル言語AI) は、論理基盤の推論のためにニューラルネットワークと言語AIの能力を結合する。

10

【0020】

新たなAI方法論であるニューラル言語AIは、知識と推論を基盤とする言語AIまたは古典的AIの補完的能力により、マシンラーニングなどの統計的AIの強点を強化する。このとき、ニューラルという用語は、最も広範囲な意味として、人工ニューラルネットワークまたは連結主義システムの使用を意味する。言語という用語は、明示的な記号操作を基盤とするAI処理方式を示す。ニューラル言語AI接近法は、AIシステム内の情報の表現に差がある。言語システムの場合、表現が明白であり、言語手段によって操作され、人間による理解が可能である。しかし、ニューラルシステムにおける表現は、そのほとんどがニューロン同士の加重連結によってなされる。ニューラル言語AIの主な目標は、少量のデータで学習可能な能力によって複雑な問題を解決することにあり、使用者に各決定と制御可能な措置を理解させる理由を提供することにあり、これは産業にAIを統合するときに重要となる。

20

【0021】

ニューラル言語AIの増加は、このような接近の機会を新たに創出した多くの研究から始まった。従来技術では、視覚および言語理解のためのニューラル言語AIに基づいて、イメージと関連質疑応答ペアの概念共同学習を実行する技法を提案した。視覚的認識と言語理解のためのディープラーニングと推論のための言語的プログラム実行に伝統的なAIを適用することにより、接近方式では、与えられたイメージから多様な関係的および概念的質問に回答することができた。従来技術ではVQA (Visual Question Answering) システムに対するCLEVRデータセットを使用することにより、視覚データに対する質問を推論して応答した。データセットのイメージは、シリンダ、キューブ、球のような単純な3D形態で構成される。各物体は、固有のカラー (赤、緑、青など)、材料 (ゴムまたは金属)、および大きさ (小さいとか大きい) を有しており、特定の物体の前 (左側、右側、後ろ、および前) にある他の物体と特定の関係的な位置を有する。このような場面で推論を行うために、CLEVRの各質問に対する機能性プログラムを取り入れたが、ここでは、プログラムは場面グラフで実行されることができ、イメージからの質問に対する返答を提供する。提案されたプログラムには、組み合わせて特定の結果を提供するクエリ、カウント、または比較作業が含まれる。

30

40

【0022】

視覚および言語理解のためのニューラル言語AIに基づいてイメージと関連質疑応答ペアの概念共同学習を実行する技法は、視覚と言語理解を推論して分離する。まず、ニューラルネットワークによってイメージ場面を構文分析し、これを機能性プログラムに変換することによって問題が理解される。構文分析されたイメージ情報は知識で構成される。次に、この推論は、質問に対する返答を与えるために、知識を土台にしたプログラムの言語的な実行を適用する。このような技法で構造的場面表現を抽出するために、マスクR-CNNおよびCNNネットワークを適用した。質問を処理してプログラムを生成するために

50

、エンコーダとデコーダの双方向 L S T M エンコーダを使用したシーケンス・ツー・シーケンスモデルが適用される。複雑なプログラムに対する堅固性、小さな訓練データなどのような多様な長所をもつこの方法は、C L E V R データセットで優れた正確度を発揮した。

【 0 0 2 3 】

学習と推論の統合は、近年、人工知能とマシンラーニングの核心課題の 1 つとして挙げられている。さらに、ニューラル言語接近法の意味論、説明可能性、潜在的応用、最小限あるいは領域別の訓練なく作業に一般化することが可能であるなどの多くの疑点が依然として残っている。従来のディープラーニング接近法の限界が徐々に明らかになってきており、論理的推論による追加的な背景知識は、ディープラーニングシステムをさらに改善するためのものとなる。このような方法において、本発明では、トランスフォーマ (T r a n s f o r m e r) と呼ばれる構造に基づいて 1 つのニューラルネットワーク作業を統合した。

10

【 0 0 2 4 】

多様な従来技術によると、自然語処理作業では、R N N に比べてトランスフォーマの優秀性が立証された。回帰型ニューラルネットワークは、自然語を時系列で取り扱って動作するが、ここで、すべての単語は、それ以前に出たすべての単語の意味を修正する。R N N は、一度に 1 つの単語を詳察し、次の単語の表現として該当の表現をさらに脈絡化するように表現を生成する。トランスフォーマと R N N 構造を比べると、トランスフォーマは文章全体を処理する自己主義メカニズムによって情報を順に学習する反面、R N N は並列処理を許容せず単語ごとに表現を抽出するため、トランスフォーマの訓練過程を多くの G P U に分散することができて効率的である。さらに、トランスフォーマは、以前の単語との依存性を捕捉するために過去の状態に依存せず文章を全体的に処理し、マルチヘッド主義と位置埋め込みは他の単語との関係に関する情報を提供するが、R N N 構造は各状態がある過去状態に基づいて学習された情報を維持する。以前の状態だけに依存すると仮定することにより、長い従属性内で問題が発生する。したがって、トランスフォーマは、単語や単語の部分までも読み込んで周辺の単語から情報を収集し、与えられた言語のビットの意味を脈絡によって決定することができる。与えられた接近方式のすべての長所を考慮した上で、本発明では、トランスフォーマ構造に基づいて音声認識および翻訳のような言語理解モデルを構築した。

20

【 0 0 2 5 】

メタバース分野において、音声通信システムは、3 D 世界の仮想資源と相互作用するためにコンテキストを認識することが必須となる。本発明は、レガシーマニュアル、3 D モデル、3 D シミュレータ、および航空機の整備知識を提供するボーイング 7 3 7 型機の整備訓練および教育のためのメタバースを提案する。

30

【 0 0 2 6 】

また、メンテナンスマニュアルが厳格に順守されるメタバースのオペレーションフローを探索して制御するために、コンテキスト認識音声理解モジュール N S S E (N e u r o - S y m b o l i c S p e e c h E x e c u t o r) を提供する。N S S E は、従来の音声認識方式とは異なり、ニューラルネットワークと伝統的な言語的推論を結合したニューラル言語 (N e u r o - S y m b o l i c) A I を適用し、コンテキストおよび航空機関連知識に基づいて使用者の要請と応答を理解する。

40

【 0 0 2 7 】

N S S E は、訓練用合成データだけを適用することで産業的に柔軟な接近方式として開発された。これにもかかわらず、実際の使用者データに対する多様な自動音声認識マトリックで実行された評価プロセスは、平均正確度 9 4 . 7 % 、W E R (W o r d E r r o r R a t e) 7 . 5 % 、非ネイティブ発音の使用者の音声要請を処理する一般化能力として持続可能な結果を示した。

【 0 0 2 8 】

本発明が提案する、メタバースを利用した航空機の整備訓練、教育方法、およびシステムは、高価な物理的航空機を容易に修正およびアップデートすることのできる仮想航空機

50

に代替可能であることから、航空技術のための安価かつ拡張可能なソリューションであると言える。さらに、現場の専門家の代わりとなるNSS Eは、航空機整備に対する効果的な訓練および教育を容易に行うことができるため、技術的な指導とすべての資源を提供することができる。以下、本発明の実施例について、添付の図面を参照しながら詳しく説明する。

【0029】

図1は、本発明の一実施形態における、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広告システムの構成を示した図である。

【0030】

本発明が提案する、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広告システム100は、シミュレーション実行部110、人工知能知識処理部120、学習部130、およびニューラル言語音声実行部140を含む。

10

【0031】

本発明の実施形態に係るシミュレーション実行部110は、航空機をはじめとする機械類の整備訓練、教育、および広報のためのメタバース複合現実(Mixed Reality)において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン(Digital Twin)を提供して3Dシミュレーションを実行する。

【0032】

本発明の実施形態に係る学習部130は、2Dマニュアル、前記2Dマニュアルの作業指示書、およびSCM(Simulation Cost Model)を含む学習情報に基づいて、人工知能知識処理部120に人工知能知識(Knowledge)を提供する。

20

【0033】

本発明の実施形態では、航空機の整備訓練、教育、および広報を例に挙げて説明しているが、これに限定されてはならず、航空機をはじめとする多様な機械類の整備訓練、教育、および広報のためのメタバース複合現実(Mixed Reality)に適用可能である。

【0034】

図2は、本発明の一実施形態における、航空機整備メタバースの1人称ビュースナップショットの例示図である。

30

【0035】

航空機整備メタバースは、MRO(Maintenance Repair Operation)分野の使用者が一堂に会して各航空機の仮想資産をオペレーションできるようにする協業空間である。メタバースは、認識された仮想世界に連結された、持続的かつ共有された3D仮想空間で構成されたインターネットの未来帰型概念を説明するために使用される。

【0036】

これと同時に、本発明で提案するメタバースは、教育生がメンテナンス教育を容易に進めることができるように、サポート資料と関数を備えた仮想航空機でオペレーションすることが可能な学習場である。仮想マニュアルにしたがって作業をするのに必要となるすべてを備えているメタバースは、効果的な教育作業のフローを生成する。世界がいまコロナウイルスによって多くの打撃を受けており、多様な産業が伝統的な業務や正規教育からSociety 5.0を向上させるオンライン化に移行しているという事実を考慮するとき、提案されたメタバースのような仮想空間は、大流行による当面の課題を処理することが可能な潜在的ソリューションを提供する。メタバースは、すべてのものを統合する、統一されたポータルおよびハブとして機能しながらも、世界を連結するための相互運用可能なゲートウェイを生成する。同じ方式により、本発明では、現実世界を仮想航空機および整備の世界と結合させた。

40

【0037】

数億ドル(例えば、ボーイング737型機の費用は1億ドル)以上にも達する物理的な

50

飛行機の費用を考慮するとき、提案された航空機整備メタバー스는、旧式の航空機モデルに対する訓練を必要とする、多様な航空大学や学校のための潜在的ソリューションを提案する。メタバー스에存在する航空機の仮想モデルは、簡単にアップデートしたり代替したりすることができる。また、物理的な部品（例えば、航空機着陸装置）を用いた作業には、部品の重量が大きいことから、持ち運びや設置のために特殊な装備が必要であった。これに対し、スマートグラスの多様な相互作用メカニズムは、使用者が指でタッチするという直観的な方法だけで操作を可能にする。したがって、膨大な資源節約が可能となり、産業のメタバー스가重要な役割を担うようになる。

【 0 0 3 8 】

メタバー스에接続するためには、スマートグラス（例えば、H o l l o L e n s 2 ）が使用される。スマートグラスは、複合現実を現実世界に投射することで、3 D世界に没入させる経験を提供できるようになる。図 2 の左下の写真は、1 人称の観点における、提案された航空機整備メタバー스의ナップショットを示している。

【 0 0 3 9 】

図 2 を参照すると、多様な視覚的要素が存在することが分かる。まず、主要部品（m a i n a s s e t ）は、作業しなければならない特定の航空機部品である。作業しなければならない特定の航空機部品は中央に位置し、実際のモデルのデジタルツインを示す。図 2 には、ボーイング 7 3 7 型機の主な着陸ギアが示されている。このモデルは、初心者が視覚的な手がかりを得ることができるように構成要素に注釈を付けた。モデルの右側には、ビデオを参照するためのメディアプレイヤーが配置される。チュートリアルビデオは、特定の作業に対する同僚エンジニアの作業を要約することにより、教育生が完了すべき手順を理解するのに役に立てる。次に、デジタルツインの左側では手動セクションが試演される。提案されたシステムは、従来の 2 D 航空機整備マニュアルを維持しつつも、革新的な 3 D シミュレータを取り入れている。

【 0 0 4 0 】

定義されたプロトコルを維持することが安全性と効率性のために重要となるため、システムで実現されるすべての手順は、ボーイング 7 3 7 機型の公式マニュアルと文書を土台とする。したがって、このプロジェクトで実行した最初の段階は、レガシー（l e g a c y ）文書をシステムで使用する構造化された形式に変換することである。J S O N 形式は、膨大な量のデータをカプセル化して知識に変換すると同時にメタバー스에ウェブの概念を強化するために使用されるが、ここにはすべての仮想世界とインターネットの和（s u m ）が含まれる。このようなシステムは、複合現実アニメーション、メディアコンテンツ、およびメンテナンス訓練と教育を革新する 3 D マニュアルのような新次元の情報によって手順を改善しながらも、伝統的な航空機の整備方式の信頼を維持することができる。

【 0 0 4 1 】

図 3 は、本発明の一実施形態における、航空機整備マニュアル 3 D シミュレータの例示図である。

【 0 0 4 2 】

本発明の実施形態に係る 3 D マニュアルは、従来の 2 D マニュアルを見ることのできる新たな方法を示す。一般的に、2 D レガシーマニュアルには、注釈とともに特定のプロセスを説明する図が記載されているが、これは静的であり、最終結果をスナップショットで表示したものに過ぎない。図 3 には、2 D A M M マニュアルの例が示されている。主に着陸装置の下部側面のストロット除去が示されており、2 D マニュアルは、整備プロセス中に特定の作業を実行する方法に関する参照として作動する。

【 0 0 4 3 】

3 D マニュアルは、場面と個々の構成要素を互いに異なる角度から個別に視覚化することにより、参照する情報をより適切に理解できるように手助けするモデルである。レガシーマニュアルを見るために新たな次元を紹介しながら、2 D レガシーの図を完成する 3 D マニュアルを提案した。図 3 に示した 3 D マニュアルは、2 D マニュアルに追加されたものと説明されるため、本に示された図を参照する使用者が理解することのできる視覚によ

10

20

30

40

50

って3Dマニュアルを詳察することができる。

【0044】

3Dマニュアルは、探索可能な3Dの他にも多様な機能を備えている。2Dの図は、作業、下位作業、または命令の情報をカプセル化して好ましい最終結果を表示するため、3Dで手動中間プロセスによって探求することができる。図3の2Dは、次の3つの指針をもつ下位作業の実行による最終結果を示している。

【0045】

「ボルト46からナット42、ワッシャ43を除去」

「ボルト46を除去して下部のサイドストロットアセンブリを分離」

「下部側面のストロットアセンブリからプッシュロード41を分離」

10

【0046】

これとは反対に、提案された3Dマニュアルを使用すれば、指針レベルで深層的に見ることができ、段階別に実行することもできる。すなわち、図3の「ナット42、ワッシャ43を除去」は、46から42を除去することと46から43を除去することの2つの段階に分けられているため、使用者が特定の指針を一段階で実行したり、図のように下位段階に分けたりすることができる。段階別に命令が実行されれば、プロセスをより適切に探索することができるように、「完了」アイコンで表示される視覚的な手がかりが表示される。

【0047】

各下位作業または命令には、実験するシミュレータとして考慮することのできる自体3Dマニュアルがある。この複雑なプロセスを制御するために音声命令が使用され、本発明の実施形態に係る方法によって処理される。

20

【0048】

再び図1を参照すると、本発明の実施形態に係るニューラル言語音声実行部140は、提供された人工知能知識、デジタルツインに基づいて3Dシミュレーションを実行するために音声要請を処理するためのニューラルネットワーク作業と言語的推論を実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する。

【0049】

本発明の実施形態に係るニューラル言語音声実行部140は、動的長オーディオ録音機、音声-テキストネットワーク、テキスト-プログラムネットワーク、および言語プログラム実行機を含む。

30

【0050】

本発明の実施形態に係る動的長オーディオ録音機は、スマートグラスを着用した使用者が自身のオーディオ要請を録音するようにニューラル言語音声実行部をトリガーし、ニューラル言語音声実行部がトリガー構文を感知して動的長オーディオ録音アルゴリズムを呼び出し、マイクから出力される音声信号ストリームでオーディオデータを生成するように処理する。

【0051】

本発明の実施形態に係る動的長オーディオ録音機は、システムの応答時間を改善するために録音に静的時間を設定せず、オーディオ信号を録音する動的長オーディオ録音アルゴリズムを利用する。

40

【0052】

本発明の実施形態に係る動的長オーディオ録音アルゴリズムは、入力により、原始オーディオ形式のマイクストリーム、ストリームから分析する関数の数、タイムスタンプにおけるデータ比較のための閾値、および録音が中止されるまでの最大沈黙時間を提供して、出力としてストリームで生成されたオーディオデータを取得する。

【0053】

本発明の実施形態に係る音声-テキストネットワークは、前記オーディオデータをテキストに変換し、自動音声認識のための音声-テキストの形態でテキスト-プログラムネッ

50

トワークに伝達する。

【0054】

本発明の実施形態に係るテキスト - プログラムネットワークは、前記音声 - テキストを各ドメインの言語の実行可能なプログラムシーケンスに変換するための関数と媒介変数で構成される。

【0055】

本発明の実施形態に係るテキスト - プログラムネットワークは、テキストの単語を教育データセットの単語と一致させるための一般語彙 (General Vocabulary) を利用して要請ベクトルに変換し、前記要請ベクトルをプログラムベクトルに変換し、前記プログラムベクトルはプログラム生成に使用されるドメイン特化言語 (Domain Specific Language) の構成要素に対する参照を含む。

10

【0056】

本発明の実施形態に係る言語プログラム実行機は、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する。

【0057】

本発明の実施形態に係る言語プログラム実行機は、実行しなければならないプログラムを入力として使用し、各プログラムは関数および該当の媒介変数で構成され、与えられたプログラムの各プログラムに対して繰り返しを入力すれば関数および媒介変数が抽出され、以前の繰り返しの結果を説明する変数 (Prev) が媒介変数に追加され、関数と媒介変数が準備されれば、Execute関数は各関数を呼び出して抽出された媒介変数を伝達し、各関数は返還値を有するため各繰り返しで変数 (Prev) がアップデートされ、前記手順をすべてのプログラムに適用する。図5～13を参照しながら、本発明の実施形態に係るニューラル言語音声実行部140の各構成についてより詳しく説明する。

20

【0058】

図4は、本発明の一実施形態における、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広報方法を説明するためのフローチャートである。

【0059】

提案する、メタバースを利用した複合現実における複雑な機械類の訓練、教育、および広報方法は、航空機をはじめとする機械類の整備訓練、教育、および広報のためのメタバース複合現実 (Mixed Reality) において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン (Digital Twin) をシミュレーション実行部が提供して3Dシミュレーションを実行する段階410、2Dマニュアル、前記2Dマニュアルの作業指示書、およびSCM (Simulation Cost Model) を含む学習情報に基づいて人工知能知識 (Knowledge) を学習部が提供する段階420、および提供された2Dマニュアルおよび3Dマニュアルに基づいてシミュレーションを実行するために、ニューラル言語音声実行部で音声要請を処理するためのニューラルネットワーク作業と言語的推論を実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する段階430を含む。

30

【0060】

段階410では、航空機をはじめとする機械類の整備訓練、教育、および広報のためのメタバース複合現実 (Mixed Reality) において、スマートグラスを利用して前記整備訓練、教育、および広報のための特定の視覚的要素に対してシミュレーションを実行するためのデジタルツイン (Digital Twin) をシミュレーション実行部が提供して3Dシミュレーションを実行する。

40

【0061】

段階420では、2Dマニュアル、前記2Dマニュアルの作業指示書、およびSCM (Simulation Cost Model) を含む学習情報に基づいて人工知能知識 (Knowledge) を学習部が提供する。

【0062】

50

段階 4 3 0 では、提供された人工知能知識、デジタルツインに基づいて 3 D シミュレーションを実行するために、ニューラル言語音声実行部で音声要請を処理するためのニューラルネットワーク作業と言語的推論を実行し、使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する。

【 0 0 6 3 】

段階 4 3 0 では、スマートグラスを着用した使用者が自身のオーディオ要請を録音するようにニューラル言語音声実行部をトリガーし、ニューラル言語音声実行部がトリガー構文を感知して動的長オーディオ録音アルゴリズムを呼び出し、マイクから出力される音声信号ストリームでオーディオデータを生成するように処理する段階、音声 - テキストネットワークを介して前記オーディオデータをテキストに変換して自動音声認識のための音声 - テキストの形態でテキスト - プログラムネットワークに伝達する段階、関数と媒介変数で構成されたテキスト - プログラムネットワークを介して前記音声 - テキストを各ドメインの言語の実行可能なプログラムシーケンスに変換する段階、および言語プログラム実行機から使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する段階を含む。段階 4 3 0 に含まれる具体的な各段階については、図 5 を参照しながらさらに詳しく説明する。

【 0 0 6 4 】

図 5 は、本発明の一実施形態における、ニューラル言語音声実行部の動作過程を示した図である。

【 0 0 6 5 】

本発明の実施形態に係るニューラル言語音声実行部 (Neuro - Symbolic Speech Executor : NSSE) は、提案された、航空機整備メタバースで音声要請を処理するためのニューラルネットワークの作業と言語的推論を統合するモジュールである。パターン認識でディープラーニングの卓越な能力と推論のための伝統的な AI を結合し、ニューラル言語音声実行部は、各航空機のドメイン語彙と従来の整備マニュアルに対する多様な参照を含む多様な意味構造を備えた複雑な使用者の口語命令を理解する。例えば、「項目 8 の AM 文書表示」に対し、本発明の実施形態に係るニューラル言語音声実行部は、航空機の特定のマニュアル AM が使用者に立証されなければならない、該当の使用者を探索するために文書の 8 番の項目が強調して表示されなければならないということを認識する。

【 0 0 6 6 】

図 5 のように、ニューラル言語音声実行部には、推論を実行するための 4 つの段階がある。段階 5 1 0 で、スマートグラスを着用した使用者が自身のオーディオ要請を録音するようにニューラル言語音声実行部をトリガーする。このために「Hey、AK!」のような関連フレーズが使用されてよい。ニューラル言語音声実行部は、トリガー構文を感知して動的長オーディオ録音 (Dynamic Length Audio Recording : DLAR) アルゴリズムを呼び出し、マイクから出る音声信号ストリームでオーディオデータ生成を処理する。

【 0 0 6 7 】

段階 5 2 0 で、音声要請の長さに応じて、DLAR の出力は n 秒の持続オーディオ要請となる。次に、オーディオ要請は、音声 - テキストネットワークに伝達されるようになるが、音声 - テキストネットワークは、原始オーディオデータをテキストに変換して要請記録を抽出する自動音声認識のためのニューラルネットワークである。

【 0 0 6 8 】

段階 5 3 0 で、テキスト - プログラムネットワークは、シーケンス・ツー・シーケンスネットワークであり、英語からなる音声要請記録を招来し、関数と媒介変数で構成されて生成された各ドメインの言語の実行可能なプログラムシーケンスと一致させる。

【 0 0 6 9 】

段階 5 4 0 で、生成されたプログラム (言い換えれば、特定の関数と媒介変数の組み合わせ) は、言語プログラム実行機 (Symbolic Programs Executor

10

20

30

40

50

r)によって結果を得て使用者に視覚および音声フィードバックを送信することで、要請された作業の処理および完了を通知する。

【0070】

以下、本発明の実施形態に係る動的長オーディオ録音機について説明する。

【0071】

本発明の実施形態に係るシステムのオーディオ命令は、多様な長さからなっており、すなわち、命令である「次の命令」は1.37秒であり、「46から42、43客体の除去」は3.94秒であるため、指定された時間だけを聞くオーディオレコーダを生成することは効率的でない。可能な音声要請(45244個の要請)を分析すれば、平均2.76秒の標準偏差が0.87秒であり、最短の音声要請は0.54秒であり、最長の音声要請は7.61秒である。静的オーディオレコーダ接近法の場合、すべての要請を処理するために聴取時間を最大時間に設定しなければならない。したがって、45244個の音声命令に対する全時間は344306.84秒(45244×7.61)となり、平均浪費時間は4.85秒となる。使用者が要請を終えたとしても、静的録音は定義された時間まで続くためである。

10

【0072】

本発明では、システムの応答時間を改善するために、録音に静的時間を設定せずにオーディオ信号を録音することが可能な、動的アルゴリズムである動的長オーディオ録音(Dynamic Length Audio Recording:DLAR)を提案した。動的長オーディオ録音の論理については、アルゴリズム1で説明する。

20

【0073】

図6は、本発明の一実施形態における、動的長オーディオ録音アルゴリズムを示した図である。

【0074】

アルゴリズムに対する入力として、原始オーディオ形式のマイクストリーム、ストリームから分析する関数の数、タイムスタンプにおけるデータ比較のための閾値、および録音が中止されるまでの最大沈黙時間を提供する。出力物としてストリームで生成されたオーディオデータが得られる。アルゴリズム1によると、オーディオを0.02秒ごとに実行(6行)する間、ループのすべての繰り返しに対してマイクストリームからオーディオデータの小さなチャンク(chunk)のスペクトル平均を計算(7-8行)し、現在のスペクトルの平均差を最初のチャンクスペクトル平均と比べる(13行)。計算された差が与えられた閾値よりも小さい場合(14行)は沈黙が発生して沈黙カウンタが増加し(15行)、そうではなければカウンタは0となる(21行)。沈黙が最大音消去時間に達するたびに録音は中止(16行)され、ストリームのオーディオデータが生成される(17-18行)。この作業では、1.5秒の沈黙が発生するまでDLARが使用者の音声要請を聞き取ることができるように沈黙時間を1.5秒に設定した。

30

【0075】

提案された動的接近方式DLARと静的接近方式を比べると、可能な45244要請の場合、記録時間は192632.74秒である反面、静的接近方式の場合は344306.84秒を得た。DLARの場合、すべての要請の浪費時間は1.5秒である反面、静的接近方式は平均4.85秒であった。全体時間効率性を評価するとき、DLARは静的接近法よりも44.05%の効率性を示したが、これはシステムの推論と応答時間を相当に早める。

40

【0076】

要請音声信号がオーディオデータに変換されれば、追加処理のために音声-テキストネットワークの自動音声認識機モデルに伝達する。

【0077】

本発明の実施形態に係る音声-テキストは、音声データを読み込んで音声から音声テキストを抽出する自動音声認識ニューラルネットワークである。これは、wav2vec2.0ネットワークを基盤とする。音声-テキストはオーディオ信号のスクリプトテキスト

50

を生成し、モデルの性能はニューラル言語音声実行部の次の推論段階に直接的な影響を及ぼすため、システムで重要な役割を担う。

【 0 0 7 8 】

構造に関し、wav2vec2.0フレームワークは、音声信号の原始波形を収容して連結主の時間分類によって処理される表現を生成して信号の記録を作成する。このモデルは、マルチ畳み込みニューラルネットワークによって音声オーディオをエンコードした後、マスキング言語モデリングと類似の結果潜在音声表現の範囲をマスキングし、後にトランスフォーマを使用して脈絡と関連させる。ここで、自己アテンションメカニズム (self-attention mechanism) は、エンド・ツー・エンド方式によって潜在表現シーケンスから関係を探索する。

10

【 0 0 7 9 】

航空機整備メタバースの場合は、航空機整備の固有の単語と用語を含む専門用語が存在するため、一般的なデータセットに対して訓練されたwav2vec2.0の従来のモデルが英語で構成されていたとしても正常に作動しないことがある。しかし、ASR作業には膨大な量のデータが必要となるため、初めからwav2vec2.0を訓練するよりも、事前に訓練されたモデルを微調整する方がより効果的であると仮定した。したがって、本発明では、ニューラル言語音声実行部で音声-テキストを生成するために、LibriSpeechのような一般のデータセットに事前に制限されたwav2vec2.0を微調整することにより、膨大なデータセットの収集問題を解決する。次に、テキスト形態で作成された音声要請記録を、該当の処理のためにテキスト-プログラムネットワークに伝達する。

20

【 0 0 8 0 】

図7は、本発明の一実施形態における、テキスト-プログラムネットワークの動作過程を示した図である。

【 0 0 8 1 】

本発明の実施形態に係るニューラル言語音声実行部のテキスト-プログラムネットワークの構成要素は、音声命令のテキストを一連のプログラムに変換するディープラーニングシーケンス・ツー・シーケンスモデルである。システムにおいてプログラムは、特定のコードのピースに対する表記法であり、自体媒介変数がある関数である。したがって、テキスト-プログラムの裏面の主要直観は、要請テキストを実行する媒介変数を備えたマシン関数のシーケンスに変換することにある。

30

【 0 0 8 2 】

図7を参照すると、このようなシステムは、可能な使用者の要請から出た単語711である一般語彙 (General Vocabulary) 710と、従来の関数721、およびプログラム構成に使用することのできる媒介変数722のようなマシンとして周知の単語を示すドメイン特化言語 (Domain Specific Language) 730に対する知識を備えている。したがって、要請テキストは、テキストの単語を教育データセットの単語711と一致させる一般語彙を土台として要請ベクトルに変換する。次に、テキスト-プログラムネットワークは、要請ベクトルをプログラムベクトルに変換する。プログラムベクトルは、プログラム生成に使用されるドメイン特化言語の構成要素に対する参照を含む。したがって、例題要請テキスト「Show AMM manual of item 8」は、「Find Object (Request)」および「Show Manual (AMM, Prev)」プログラムに変換される。

40

【 0 0 8 3 】

図8は、本発明の一実施形態における、テキスト-プログラムネットワークのアーキテクチャを示した図である。

【 0 0 8 4 】

テキスト-プログラムネットワークの構造は、トランスフォーマ (Transformer) 830に基盤を置いており、トランスフォーマ830は、エンコーダ-デコーダの種類の構造であるため、翻訳作業に極めて適する。図8は、テキスト-プログラムネット

50

ワークの構造を示している。要請テキスト入力 8 1 0 とプログラム入力 8 2 0 の両方に 2 5 6 次元の単語埋め込みレイヤがある場合、埋め込みベクトルは、エンコーダとデコーダに供給される前に位置エンコードの形態で各ワードの位置情報と結合する。この作業において構造は、与えられたシーケンスに対して学習された情報の表現として、シーケンスをマッピングする 3 つの同一エンコーダと 8 つのマルチヘッドアテンション層で最適に作動する 3 つの個別のデコーダで構成される。全体的に要請語彙の大きさは 8 9 である反面、プログラム語彙の大きさは 4 9 である。トランスフォーマ 8 3 0 の出力は、出力確率を得るためのアクティブ化せず、Dropout 0.3 (8 4 0) および全結合層 (Fully Connected) 8 5 0 に伝達された。

【0085】

構造側面において、本発明は、従来技術の L S T M とは対照的にトランスフォーマを適用した。従来技術の L S T M は速度が効率的でなかった。これは、シーケンスで特定の項目に対する埋め込みを生成するためには、以前のすべての単語表現を計算しなければならず、計算プロセスを G P U で実行するため並列化することができないためである。これとは反対に、トランスフォーマモデルは、並列化パイプラインを使用するため、多数の G P U にわたって教育と実行が可能となる。また、従来技術の L S T M は、トークンの前に来るトークンによってトークンの意味を理解するが、後ろに来るトークンによるトークンの意味を理解するためには脈絡化が足りなかった。しかし、トランスフォーマでは、シーケンスのすべてのトークンが該当のシーケンスの他のトークンと同時に併合されるため、脈絡を堅固にさせる。最後に、生成されたプログラム 8 6 0 は、実行のためにニューラル言語音声実行部処理の最後の部分を通過する。

【0086】

図 9 は、本発明の一実施形態における、言語プログラム実行機の動作過程を示した図である。

【0087】

本発明の実施形態に係る言語プログラム実行機 9 1 0 は、テキスト - プログラムネットワークで生成されたプログラムを実行して使用者に視覚およびオーディオフィードバックを提供する、ニューラル言語音声実行部の構成要素である。

【0088】

図 1 0 は、本発明の一実施形態における、言語プログラム実行機のアルゴリズムを示した図である。

【0089】

アルゴリズム 2 は、言語プログラム実行プロセスについて説明する。入力として、アルゴリズムは、実行すべきプログラムを使用する。各プログラムは、関数および該当の媒介変数で構成される。与えられたプログラムの各プログラムに対して繰り返しを入力すれば (2 行)、関数および媒介変数が抽出される (3 - 4 行)。この後、以前の繰り返しの結果を説明する変数 Prev が媒介変数に追加される。関数と媒介変数が準備されれば、Execute 関数は各関数を呼び出し、抽出された媒介変数を伝達する (5 行)。各関数は返還値を有するため、各繰り返しで変数 Prev がアップデートされる (5 行)。このような手順はすべてのプログラムに適用されるが、ここで、Prev の最後の値は、実行の全体結果を説明する (7 行)。返還関数の類型は互いに異なり、必要によって生成される。

【0090】

図 9 の「Show AMM manual of item 8」の例題を考慮すると、該当のプログラムは、「FindObject (Request)」と「Show Manual (AMM, Prev)」である。この場合、言語プログラム実行機が順に実行しなければならない 2 つのプログラムがある。システムには言語プログラム空間が存在し、言語プログラム実行機は、マシン空間のインスタンスと生成されたプログラムを一致させて実行を呼び出す。図 9 では、まず、FindObject 関数が呼び出され、命令のスク립ト (言い換えれば、音声 - テキストネットワークの結果) を示す要請媒介変数が使用

10

20

30

40

50

される。FindObjectは、与えられたテキストから数字を探索して返還する関数であるため、このプログラムを実行した後にPrev変数は8となる。8は、例題で言及した数字であるためである。次に、FindObjectの返還値を使用して、媒介変数AMMおよびPrevがあるShowManual関数を呼び出す。ShowManualは、特定の種類のマニュアルを表示し、その中に数字を強調する関数である。この場合、マニュアルのタイプはAMMであり、強調表示すべき番号はPrevであり、現在値8を維持している。各ドメインの言語のすべての関数には固有の義務があり、一部は処理された計算演算を返還し、一部は有効性検査などを実行する。

【0091】

図11は、本発明の一実施形態における、サンプル使用者の要請から結果までの過程を示した例示図である。

10

【0092】

すべてのプログラムの処理が完了すれば、言語プログラム実行機は、使用者にフィードバックを処理し、使用者にマニュアルを視覚的（例えば、テキストとアイコンの形態で）かつ音声的に再生して表示する進行中であるという手順を通知することにより、使用者経験を向上させる。

【0093】

本発明の実施形態に係るニューラル言語音声実行部の構成要素をすべて結合すれば、図11のようなサンプル要請が提供される。多様な持続時間を含んだ音声要請は、動的長オーディオ録音アルゴリズム1110によって処理され、音声-テキストネットワーク1120によって記録される。次に、テキスト-プログラムネットワーク1130は、多様な複雑性と大きさを備えたプログラムを生成する。テキスト-プログラムネットワーク1130が相互連結された4つのプログラムを生成したことを確認することができる。ここで、GetItems関数は、JSON知識ファイルから情報を読み込み、次には作業が含まれるため、ニューラル言語プログラム実行機1140はすべての下位作業を抽出する。下位作業は命令で構成されるため、下位作業のすべての命令を探索する。すべての項目が準備されれば、すなわち、すべての命令語が含まれたノードである数学的プログラムカウントは、以前の演算動作の項目をカウントすることで正確な命令数の要請に対する適切な返答を提供する。

20

【0094】

ニューラル言語音声実行部の作業は、ニューラル処理の長所と多様な各状況の音声要請を処理するための言語的推論を結合したニューラル言語AIを基盤とする。

30

【0095】

図12は、本発明の一実施形態における、ニューラル言語音声実行部のコンテキスト管理を説明するための図である。

【0096】

使用者の音声要請を処理して特定の状況に応じて応答するためには、ニューラル言語推論に基づいてシステムを構築することが効果的である。ニューラル言語音声実行部のニューラル構成要素が音声で複雑なパターン認識を実行するとき、言語部分は適切な応答を提供し、使用者の要請を検証するために脈絡と知識を管理する。

40

【0097】

図12は、ニューラル言語音声実行部の状況管理を説明している。まず、AMM、IPCのようなすべての従来のマニュアル1231はJSON形式で構成されるため、接近および相互参照が可能である。この航空機整備知識1232には、すべての構成要素、構築関係、および従属性が要約されている。図12の例において、多様な作業ノードは、AMMの多数の下位作業を含むと同時に、指針のある下位作業ノードは、航空機の特定の部品番号、例えば、マニュアルから出た項目51、8、42などを参照し、AMで説明するシミュレーションの手順を備えた自体3Dモデル1234を含んでいなければならない。3D仮想資産と航空機整備知識をアクティブ状態1233で参照する他に、JSONから情報を収集する。ここには、現在作業、下位作業、および命令情報、AMMマニュアルの使

50

用可能な注釈、現在場面に使用された3D資産、シミュレーションなどのような多様な環境変数とリンクが含まれる。このすべての内容は、ニューラル言語音声実行部が音声命令を処理するときに準じなければならない、考慮しなければならない脈絡を生成する。

【0098】

図12の「Show AMM manual of item 8」と「Show AMM manual of item 9」要請に対するニューラル言語音声実行部のテキスト・プログラムネットワーク1210の例をみると、言語プログラム実行機1220は、生成された同一のプログラムを生成するが、脈絡、使用可能な3D資料、および全般的な知識にしたがって要請を検証して最終的な返答を提供する。図12のAMM項目を現在状況として考慮するとき、項目8(item 8)の要請は有効であるが、項目9(item 9)はAMM注釈にないため、使用者に対するフィードバックで対応する。

10

【0099】

意味論ニューラル部分では、テキスト・プログラムネットワーク1210が要請テキストをマシン理解可能なプログラムに変換するときに、各状況の情報は考慮されない。テキスト・プログラムネットワーク1210は、言語プログラム実行機1220に結果を得るためにどのような段階を実行すべきであることを知らせるが、言語推論は脈絡基盤の有効性検査の手順を含むプログラムを実行する間に発生する。したがって、ニューラル部分と言語部分がともに作用することが必須となる。

【0100】

図13は、本発明の一実施形態における、ニューラル言語音声実行部のアキテクチャを示した図である。

20

【0101】

ニューラル言語音声実行部は、クライアント・サーバ構造による2つの装置にわたって位置する4つの主要構成要素で構成されたシステムである。図13は、ニューラル言語音声実行部のシステム構造を説明しており、使用者のスマートグラスであるクライアントマシンとすべての処理を処理するディープラーニングマシンがある。本発明の実施形態に係るシステムのクライアントは、スマートグラスHoloLens 2で実行され、音声要請作成と生成されたプログラム処理を担当する。この反面、サーバは、ニューラルネットワークとともに作動し、音声・テキストネットワークを使用してオーディオデータをテキストに変換し、テキスト・プログラムネットワークを使用してテキストをプログラムに変換する。2台のコンピュータがインターネットを介して通信しながらデータを交換する。クライアントはオーディオデータを送信し、サーバは生成されたプログラムが含まれたスクリプトをクライアントに伝達する。次の段階では、推論手順について詳しく説明する。

30

【0102】

クライアント側にある動的長オーディオの録音は、スマートグラスマイクを使用してオーディオ要請を生成する(1310)。

【0103】

オーディオデータがウェブを介してサーバマシンに送信される(1311)。

【0104】

受信されたオーディオデータは、音声・テキストネットワークで処理されて要請記録を抽出する(1320)。

40

【0105】

音声・テキストネットワークは、記録の一連のプログラムに変換される(1330)。

【0106】

要請テキストおよび生成されたプログラムがクライアントに送信される(1331)。

【0107】

言語プログラム実行機がプログラムを処理する(1340)。

【0108】

聴覚および視覚フィードバックとして生成された結果を使用者に試演する(1341)。

【0109】

50

この構造は、他の 3 D 資産を考慮せず、システムに 2 つのニューラルネットワークがあるため、クライアント装置であるスマートグラスにコンピュータ処理の過負荷がかからないように保障する。したがって、GPU を備えた強力なマシンが設置され、装置の音声要請を迅速かつ効率的に処理することができる。さらに、音声処理モジュールを機器から分離することで簡単に維持またはアップデートすることができ、スマートフォン、PC、タブレットなどのような多様なプラットフォームに構築されたアプリケーションサービスを提供することができる。

【 0 1 1 0 】

従来技術で利用した作業例では、イメージから特定の物体がどのような形態、カラー、関係を有しているかを確認するような問題に対するニューラル言語 AI を考慮する反面、本発明では、ニューラル言語 AI の概念が実際の問題を解決するために業界に適用可能であるということを示した。言い換えれば、各状況の知識と環境を参照する複雑な意味構造をもつ音声要請を処理することができる。従来技術が提案する視覚的理解と質問返答において接近方式を比較したが、本発明では、質問が何であるかを理解するために、オーディオ信号を処理し、与えられた質問に基づいて機能プログラムが生成されて航空機関連知識として実行される。さらに、従来技術では、機能的プログラムの実行のために、視覚的データの構造的場面表現を抽出するためにイメージを構文分析するが、整備マニュアルで知識を要約した JSON ファイル形式でこのような表現を作成した。これにもかかわらず、2 つの技法すべてで推論プロセスの透明性を保障するが、これは、多様な問題を追跡してこれに関する説明可能な理由を探索する機会を提供し、これは産業に使用されるシステムに極めて重要なものとなる。

【 0 1 1 1 】

本発明の実施形態に係るメタバースを利用した航空機の整備訓練、教育方法、およびシステムは、航空大学の訓練と教育過程に革命を起こし得る航空機整備メタバースと呼ばれる次世代協力仮想空間を構築する。提案されたメタバースは、レガシーマニュアル、3 D モデル、およびシミュレーション、航空機知識、および確立されたメンテナンスフローのように航空機の MRO に必要なすべての資源を含んでおり、訓練のための物理的航空機を仮想航空機に代替することにより膨大な量の資源節約に繋がる。さらに、大学では資源不足によって古い航空機モデルを教育に使用してきたが、メタバースとともに最新知識を簡単に維持することができる。

【 0 1 1 2 】

本発明の実施形態によると、従来の知識のための新次元の 3 D シミュレータを構築して従来の航空機整備マニュアルを改善することを提案する。3 D マニュアルは 2、D AM マニュアルをコピーしてアニメーションと各段階の制御実行機能を追加する。一般的に、2 D マニュアルに提示された図は、情報を 1 つの観点、すなわち、静的な観点だけで表示するが、提案された 3 D マニュアルは全面観察と相互作用を可能にする。

【 0 1 1 3 】

本発明の実施形態によると、メタバースの運営フローを探索して制御するために、3 D マニュアルと相互作用するためのニューラル言語音声実行部という音声通信を提案する。このように、本発明は、航空機整備知識を土台に推論することのできる状況認識音声理解を構築することで、ニューラル言語 AI の概念を発展させることができる。

【 0 1 1 4 】

上述した装置は、ハードウェア構成要素、ソフトウェア構成要素、および / またはハードウェア構成要素とソフトウェア構成要素との組み合わせによって実現されてよい。例えば、実施形態で説明された装置および構成要素は、例えば、プロセッサ、コントローラ、ALU (arithmetic logic unit)、デジタル信号プロセッサ、マイクロコンピュータ、FPA (field programmable array)、PLU (programmable logic unit)、マイクロプロセッサ、または命令を実行して応答することができる様々な装置のように、1 つ以上の汎用コンピュータまたは特殊目的コンピュータを利用して実現されてよい。処理装置は、オペレーティングシ

システム（OS）およびOS上で実行される１つ以上のソフトウェアアプリケーションを実行してよい。また、処理装置は、ソフトウェアの実行にตอบสนองし、データにアクセスし、データを記録、操作、処理、および生成してもよい。理解の便宜のために、１つの処理装置が使用されるとして説明される場合もあるが、当業者は、処理装置が複数の処理要素および／または複数種類の処理要素を含んでもよいことが理解できるであろう。例えば、処理装置は、複数のプロセッサまたは１つのプロセッサおよび１つのコントローラを含んでよい。また、並列プロセッサのような、他の処理構成も可能である。

【０１１５】

ソフトウェアは、コンピュータプログラム、コード、命令、またはこれらのうちの１つ以上の組み合わせを含んでもよく、思うままに動作するように処理装置を構成したり、独立したまたは集成的に処理装置に命令したりしてよい。ソフトウェアおよび／またはデータは、処理装置に基づいて解釈されたり、処理装置に命令またはデータを提供したりするために、いかなる種類の機械、コンポーネント、物理装置、仮想装置、コンピュータ記録媒体または装置に具現化されてよい。ソフトウェアは、ネットワークによって接続されたコンピュータシステム上に分散され、分散された状態で記録されても実行されてもよい。ソフトウェアおよびデータは、１つ以上のコンピュータ読み取り可能な記録媒体に記録されてよい。

10

【０１１６】

実施形態に係る方法は、多様なコンピュータ手段によって実行可能なプログラム命令の形態で実現されてコンピュータ読み取り可能な媒体に記録されてよい。前記コンピュータ読み取り可能な媒体は、プログラム命令、データファイル、データ構造などを単独または組み合わせて含んでよい。前記媒体に記録されるプログラム命令は、実施形態のために特別に設定されたものであってもよいし、コンピュータソフトウェアの当業者に公知されて使用可能なものであってもよい。コンピュータ読み取り可能な記録媒体の例としては、ハードディスク、フロッピー（登録商標）ディスク、および磁気テープのような磁気媒体、CD-ROMおよびDVDのような光媒体、フロプティカルディスク（*f l o p t i c a l d i s k*）のような光磁気媒体、およびROM、RAM、フラッシュメモリなどのようなプログラム命令を記録して実行するように構成されたハードウェア装置が含まれる。プログラム命令の例としては、コンパイラによって生成されるもののような機械語コードだけではなく、インタプリタなどを使用してコンピュータによって実行される高級言語コードを含む。

20

30

【０１１７】

以上のように、実施形態を、限定された実施形態および図面に基づいて説明したが、当業者であれば、上述した記載から多様な修正および変形が可能であろう。例えば、説明された技術が、説明された方法とは異なる順序で実行されたり、かつ／あるいは、説明されたシステム、構造、装置、回路などの構成要素が、説明された方法とは異なる形態で結合されたりまたは組み合わされたり、他の構成要素または均等物によって対置されたり置換されたりしても、適切な結果を達成することができる。

【０１１８】

したがって、異なる実施形態であっても、特許請求の範囲と均等なものであれば、添付される特許請求の範囲に属する。

40

【符号の説明】

【０１１９】

１１０：シミュレーション実行部

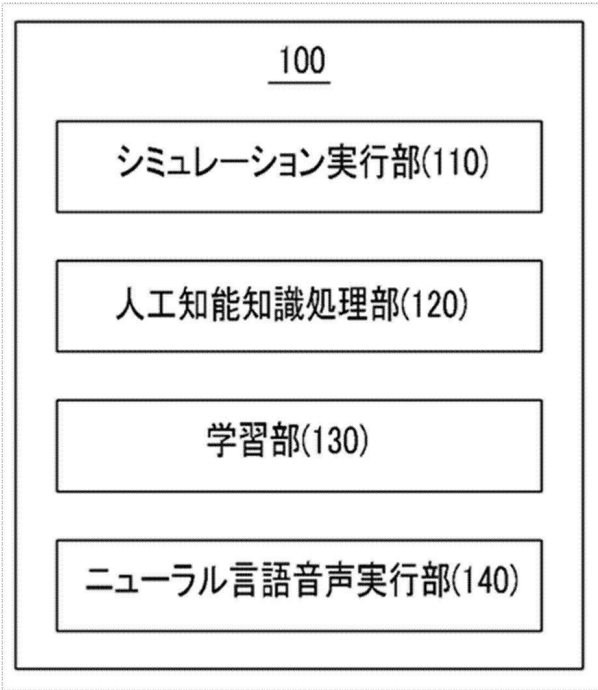
１２０：人工知能知識処理部

１３０：学習部

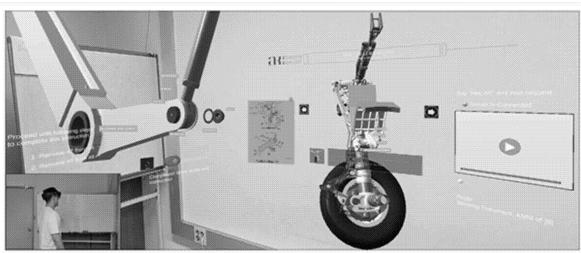
１４０：ニューラル言語音声実行部

【図面】

【図 1】



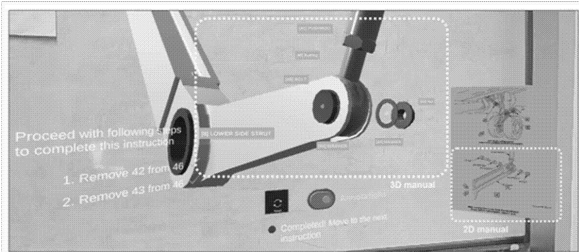
【図 2】



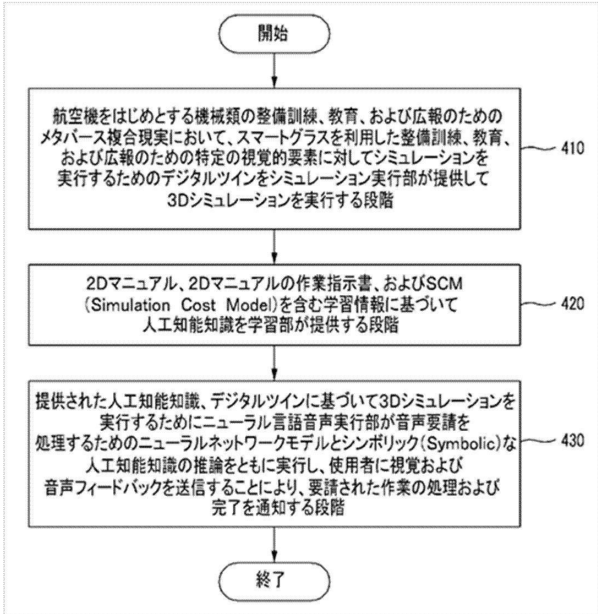
10

20

【図 3】



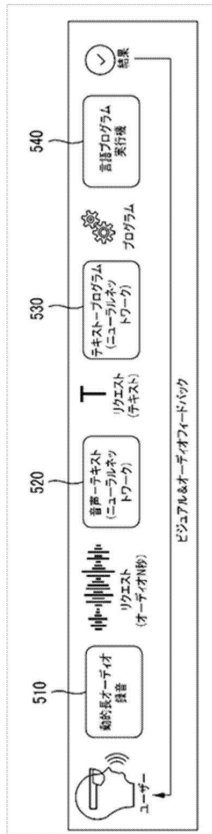
【図 4】



30

40

【図 5】



【図 6】

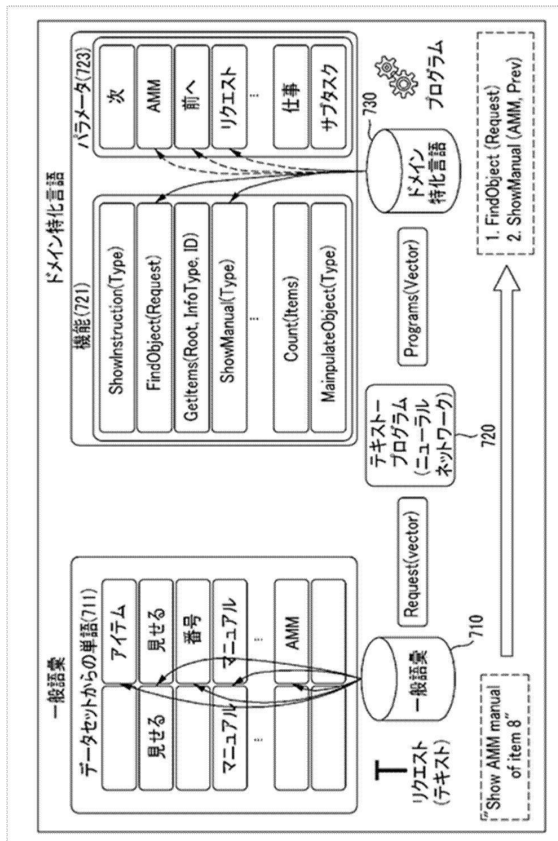
Algorithm 1 Dynamic Length Audio Recording**Input:** *MicStream*, *NofFeats*, *Thresh*, *MaxSilence***Output:** *AudioData*

```

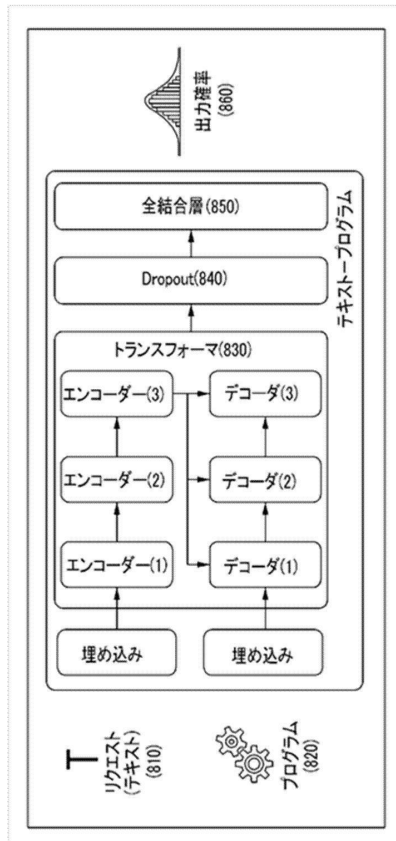
1: Recording  $\leftarrow$  True
2: FirstLog  $\leftarrow$  True
3: FirstSA  $\leftarrow$  Null
4: SilenceCounter  $\leftarrow$  0
5: AudioData  $\leftarrow$  Null
6: while Recording do
7:   S  $\leftarrow$  GetSpectrum(MicStream, NofFeats)
8:   SA  $\leftarrow$  Average(S)
9:   if FirstLog then
10:    FirstSA  $\leftarrow$  SA
11:    FirstLog  $\leftarrow$  False
12:   else
13:    Diff  $\leftarrow$  GetDiff(FirstSA, SA)
14:    if Diff < Thresh then
15:      SilenceCounter  $\leftarrow$  SilenceCounter + 1
16:      if SilenceCounter = MaxSilence then
17:        AudioData  $\leftarrow$  Stop(MicStream)
18:        Recording  $\leftarrow$  False
19:      end if
20:    else
21:      SilenceCounter  $\leftarrow$  0
22:    end if
23:  end if
24: end while
25: return AudioData

```

【図 7】



【図 8】



10

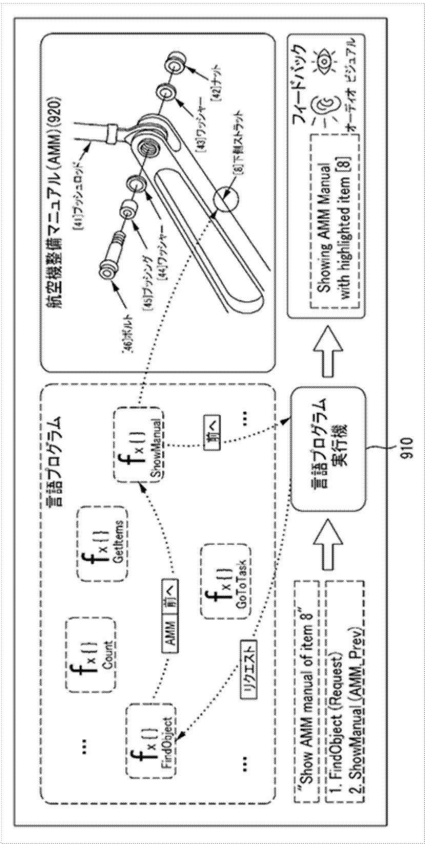
20

30

40

50

【図 9】



【図 10】

Algorithm 2 Symbolic Programs Executor

Input: *GivenPrograms*

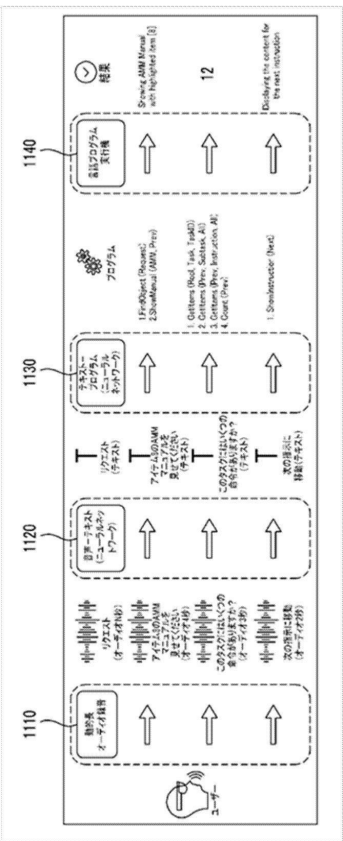
Output: *Result*

```
1: Prev ← Null
2: for Program : GivenPrograms do
3:   Function ← Program.Function
4:   Parameters ← Program.Parameters + Prev
5:   Prev ← Execute(Function, Parameters)
6: end for
7: Result ← Prev
8: return Result
```

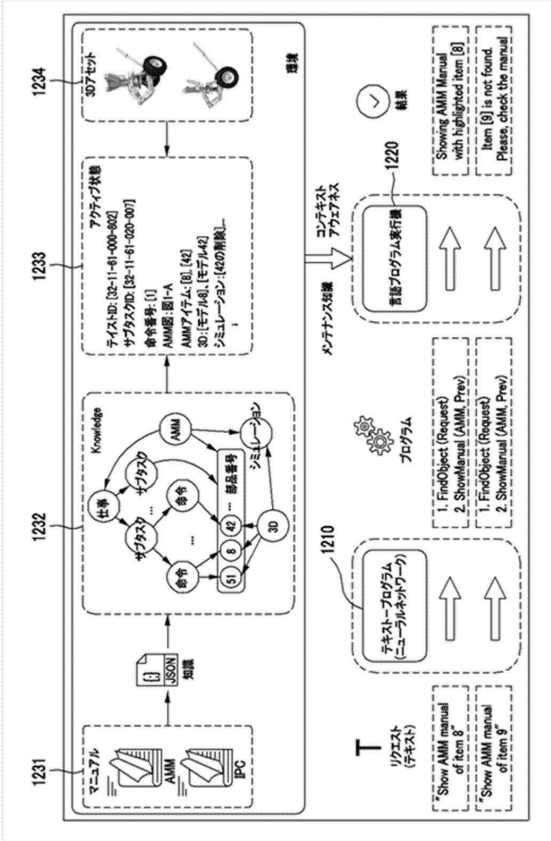
10

20

【図 11】



【図 12】

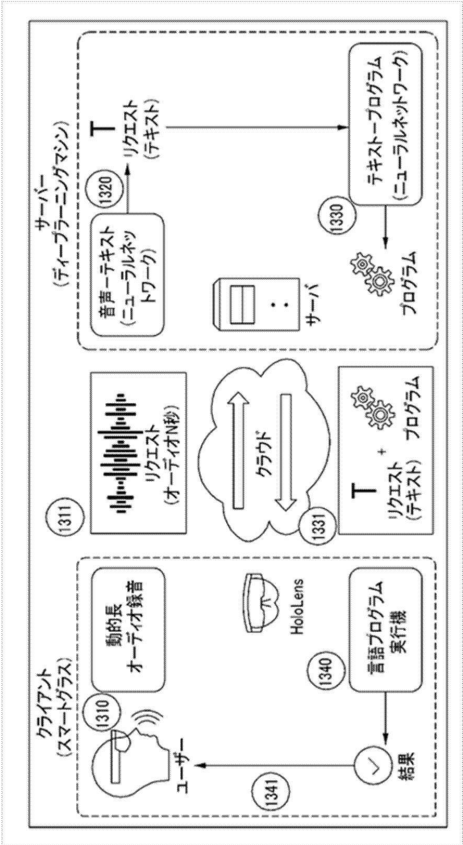


30

40

50

【図 13】



10

20

30

40

50

フロントページの続き

- (56)参考文献 国際公開第2020/032218(WO, A1)
特開2020-086437(JP, A)
特表2008-507007(JP, A)
米国特許出願公開第2020/0027459(US, A1)
韓国公開特許第10-2020-0083004(KR, A)
韓国公開特許第10-2021-0055252(KR, A)
韓国公開特許第10-2021-0096821(KR, A)
Beamo -What is a DIGITAL TWIN and why do you need it?, [online], 2021年06月07日, <https://www.youtube.com/watch?v=0i4fM48VWsM>, [検索日2023.02.20]
梓設計, [NEWS RELEASE]スマートグラスを活用したデジタルツインコンセプトプロダクトを開発, [online], 2021年05月24日, <https://www.azusasekkei.co.jp/information/news/view/270>, [検索日2023.02.16]
山田航也, 化学プラントの運転変更操作を「論理思考AI」で40%効率化 NECらが三井化学の訓練用実プラントで原料やエネルギー削減へ, [online], 2020年11月16日, <https://robotstart.info/2020/11/16/chemical-plant-ai-drivingsupport.html>, [検索日2023.02.20]
株式会社STANDARD, [プラントエンジニアリング業界でのAI活用事例]最先端技術でモノづくりに新たな息吹, [online], 2021年02月20日, https://standard-dx.com/post_blog/plant-engineering-ai, [検索日2023.02.20]
- (58)調査した分野 (Int.Cl., D B 名)
G 0 9 B 9 / 0 0
G 0 9 B 9 / 5 6
G 0 9 B 1 9 / 0 0 - 1 9 / 2 6
G 0 6 N 2 0 / 0 0
G 0 6 Q 5 0 / 2 0