

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7501523号
(P7501523)

(45)発行日 令和6年6月18日(2024.6.18)

(24)登録日 令和6年6月10日(2024.6.10)

(51)国際特許分類			F I		
G 1 0 L	15/20	(2006.01)	G 1 0 L	15/20	1 5 3
G 1 0 L	15/32	(2013.01)	G 1 0 L	15/32	2 2 0 Z
G 1 0 L	15/00	(2013.01)	G 1 0 L	15/00	2 0 0 H
G 1 0 L	15/28	(2013.01)	G 1 0 L	15/28	5 0 0
B 2 5 J	13/00	(2006.01)	B 2 5 J	13/00	Z
請求項の数 9 (全38頁)					
(21)出願番号 特願2021-511308(P2021-511308)			(73)特許権者 000002185		
(86)(22)出願日 令和2年3月9日(2020.3.9)			ソニーグループ株式会社		
(86)国際出願番号 PCT/JP2020/009948			東京都港区港南1丁目7番1号		
(87)国際公開番号 WO2020/203067			(74)代理人 110002147		
(87)国際公開日 令和2年10月8日(2020.10.8)			弁理士法人酒井国際特許事務所		
審査請求日 令和5年2月22日(2023.2.22)			(72)発明者 櫻井 達馬		
(31)優先権主張番号 特願2019-65746(P2019-65746)			東京都港区港南1丁目7番1号 ソニー株式会社内		
(32)優先日 平成31年3月29日(2019.3.29)			(72)発明者 小原 一太朗		
(33)優先権主張国・地域又は機関 日本国(JP)			東京都港区港南1丁目7番1号 ソニー株式会社内		
			審査官 浜岸 広明		
最終頁に続く					

(54)【発明の名称】 情報処理装置、情報処理方法、およびプログラム

(57)【特許請求の範囲】

【請求項1】

音声認識処理の結果に基づいて動作体の動作を制御する制御部、
を備え、
前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御し、
前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果に基づいて、当該音声認識処理の結果に対応する動作を他の前記動作体に行う第1の動作体と、前記第1の辞書よりも語彙数の少ない第2の辞書を用いた第2の音声認識処理の結果に基づく動作を行う第2の動作体を含み、
前記制御部は、前記第1の動作体が収集した音声と前記第1の辞書に基づく前記第1の音声認識処理の結果に基づいて、前記第1の音声認識処理の結果に対応する動作を前記第2の動作体に行わせる、
情報処理装置。

【請求項2】

前記制御部は、前記第1の音声認識処理の結果として得られた語彙のカテゴリに対応する語彙を前記第2の辞書から抽出し、当該語彙を前記第2の動作体に伝達する、
請求項1に記載の情報処理装置。

【請求項 3】

前記制御部は、前記第 1 の音声認識処理の結果に係るログに基づいて、前記第 2 の辞書の語彙の入れ替えを制御する、

請求項 1 に記載の情報処理装置。

【請求項 4】

前記第 2 の動作体は、前記第 2 の辞書を用いた前記第 2 の音声認識処理をローカルで実行する、

請求項 1 に記載の情報処理装置。

【請求項 5】

音声認識処理の結果に基づいて動作体の動作を制御する制御部、
を備え、

10

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御し、

前記動作体は、自律移動体であり、

前記制御部は、一の前記動作体に係る音声認識環境が音声認識処理精度の低下要因となり得る環境である場合、他の前記動作体に係る音声認識環境が向上するよう他の前記動作体の動作を制御し、

前記音声認識環境は、前記動作体の動作状況を含み、

前記制御部は、一の前記動作体の動作状況に基づいて、他の前記動作体の動作を制御し、

20

前記制御部は、一の前記動作体が動作音を伴う動作を実行している場合、他の前記動作体の動作音がより低下するよう制御する、

情報処理装置。

【請求項 6】

プロセッサが、音声認識処理の結果に基づいて動作体の動作を制御すること、
を含み、

前記制御することは、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御することと、一の前記動作体が収集した音声に基づく音声認識処理の結果に基づいて、当該音声認識処理の結果に対応する動作を他の前記動作体に実行させることと、をさらに含み、

30

前記動作体は、第 1 の辞書を用いた第 1 の音声認識処理の結果に基づく動作を行う第 1 の動作体と、前記第 1 の辞書よりも語彙数の少ない第 2 の辞書を用いた第 2 の音声認識処理の結果に基づく動作を行う第 2 の動作体を含み、

前記制御することは、前記第 1 の動作体が収集した音声と前記第 1 の辞書に基づく前記第 1 の音声認識処理の結果に基づいて、前記第 1 の音声認識処理の結果に対応する動作を前記第 2 の動作体の実行させること、をさらに含む、

情報処理方法。

【請求項 7】

プロセッサが、音声認識処理の結果に基づいて動作体の動作を制御すること、
を含み、

40

前記制御することは、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御すること、をさらに含み、

前記動作体は、自律移動体であり、

前記制御することは、一の前記動作体に係る音声認識環境が音声認識処理精度の低下要因となり得る環境である場合、他の前記動作体に係る音声認識環境が向上するよう他の前記動作体の動作を制御すること、をさらに含み、

前記音声認識環境は、前記動作体の動作状況を含み、

前記制御することは、一の前記動作体の動作状況に基づいて、他の前記動作体の動作を制

50

御することと、一の前記動作体が動作音を伴う動作を実行している場合、他の前記動作体の動作音がより低下するよう制御することと、をさらに含む、

情報処理方法。

【請求項 8】

コンピュータを、

音声認識処理の結果に基づいて動作体の動作を制御する制御部、

を備え、

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御し、

10

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果に基づいて、当該音声認識処理の結果に対応する動作を他の前記動作体を実行させ、

前記動作体は、第 1 の辞書を用いた第 1 の音声認識処理の結果に基づく動作を行う第 1 の動作体と、前記第 1 の辞書よりも語彙数の少ない第 2 の辞書を用いた第 2 の音声認識処理の結果に基づく動作を行う第 2 の動作体を含み、

前記制御部は、前記第 1 の動作体が収集した音声と前記第 1 の辞書に基づく前記第 1 の音声認識処理の結果に基づいて、前記第 1 の音声認識処理の結果に対応する動作を前記第 2 の動作体を実行させる、

情報処理装置、

として機能させるためのプログラム。

20

【請求項 9】

コンピュータを、

音声認識処理の結果に基づいて動作体の動作を制御する制御部、

を備え、

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御し、

前記動作体は、自律移動体であり、

前記制御部は、一の前記動作体に係る音声認識環境が音声認識処理精度の低下要因となり得る環境である場合、他の前記動作体に係る音声認識環境が向上するよう他の前記動作体の動作を制御し、

30

前記音声認識環境は、前記動作体の動作状況を含み、

前記制御部は、一の前記動作体の動作状況に基づいて、他の前記動作体の動作を制御し、

前記制御部は、一の前記動作体が動作音を伴う動作を実行している場合、他の前記動作体の動作音がより低下するよう制御する、

情報処理装置、

として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

40

本開示は、情報処理装置、情報処理方法、およびプログラムに関する。

【背景技術】

【0002】

近年、音声認識処理に基づく動作を行う種々の装置が普及されている。また、音声認識処理の精度を向上させるための技術も多く開発されている。例えば、特許文献 1 には、雑音の音源と成り得る他の装置の音量レベルを低下させることで、音声認識精度を向上させる技術が開示されている。

【先行技術文献】

【特許文献】

【0003】

50

【文献】特開 2017-138476 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

ところで、環境中において、音声認識処理に基づく動作を行う動作体が複数存在する場合、各動作体が収集した情報に基づく認識結果を共有することで、より状況に適した動作を実現できる可能性がある。

【課題を解決するための手段】

【0005】

本開示によれば、音声認識処理の結果に基づいて動作体の動作を制御する制御部、を備え、前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御する、情報処理装置が提供される。

【0006】

また、本開示によれば、プロセッサが、音声認識処理の結果に基づいて動作体の動作を制御すること、を含み、前記制御することは、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御すること、をさらに含む、情報処理方法が提供される。

【0007】

また、本開示によれば、コンピュータを、音声認識処理の結果に基づいて動作体の動作を制御する制御部、を備え、前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御する、情報処理装置、として機能させるためのプログラムが提供される。

【図面の簡単な説明】

【0008】

【図1】本開示の一実施形態に係る情報処理システムの構成例を示す図である。

【図2】同実施形態に係る第1の動作体10の機能構成例を示すブロック図である。

【図3】同実施形態に係る第2の動作体20の機能構成例を示すブロック図である。

【図4】同実施形態に係る情報処理サーバ30の機能構成例を示すブロック図である。

【図5】同実施形態に係る音声認識結果の共有について説明するための図である。

【図6A】同実施形態に係る音声認識結果の共有における第1の動作体10および情報処理サーバ30の動作の流れを示すフローチャートである。

【図6B】同実施形態に係る音声認識結果の共有における第2の動作体20の動作の流れを示すフローチャートである。

【図7】同実施形態に係る第2の辞書242に登録される語彙セットの入れ替えについて説明するための図である。

【図8】同実施形態に係る情報処理サーバ30による認識ログの蓄積の流れを示すフローチャートである。

【図9】同実施形態に係る第2の動作体20による辞書更新の流れを示すフローチャートである。

【図10】同実施形態に係る一の動作体の動作状況に基づく他の動作体の動作制御について説明するための図である。

【図11】同実施形態に係る一の動作体の動作状況に基づく他の動作体の動作制御の流れを示すフローチャートである。

【図12A】同実施形態に係るノイズ音源の認識結果の共有について説明するための図である。

【図12B】同実施形態に係る話者認識の結果の共有について説明するための図である。

【図13A】同実施形態に係るノイズ音源や話者に係る情報共有における共有元の動作の

10

20

30

40

50

流れを示すフローチャートである。

【図 1 3 B】同実施形態に係るノイズ音源や話者に係る情報共有における共有先の動作の流れを示すフローチャートである。

【図 1 4】同実施形態に係る能動的なアクションのトリガとなる環境の共有について説明するための図である。

【図 1 5 A】同実施形態に係る能動的アクションのトリガとなる環境の共有における共有元の動作の流れを示すフローチャートである。

【図 1 5 B】同実施形態に係る能動的アクションのトリガとなる環境の共有における共有先の動作の流れを示すフローチャートである。

【図 1 6】同実施形態に係る対話エンジンの切り替えについて説明するための図である。

10

【図 1 7 A】同実施形態に係る音声認識処理結果の統合について説明するための図である。

【図 1 7 B】同実施形態に係る音声認識処理結果の統合について説明するための図である。

【図 1 8】同実施形態に係る情報処理サーバ 3 0 による音声認識結果の統合の流れを示すフローチャートである。

【図 1 9】同実施形態に係る統合認識結果に基づく動作制御の流れを示すフローチャートである。

【図 2 0】同実施形態に係る第 1 の動作体 1 0 による第 2 の動作体の状態の伝達の流れを示すフローチャートである。

【図 2 1】同実施形態に係る第 1 の動作体 1 0 の指示に基づく第 2 の動作体 2 0 によるユーザ探索の流れを示すフローチャートである。

20

【図 2 2】同実施形態に係る外部サービスから取得した情報に基づく第 2 の動作体 2 0 への指示の流れの一例を示すフローチャートである。

【図 2 3 A】同実施形態に係る第 1 の動作体 1 0 による第 2 の動作体 2 0 への移動指示の流れを示すフローチャートである。

【図 2 3 B】同実施形態に係る第 1 の動作体 1 0 からの移動指示に基づく第 2 の動作体 2 0 の動作の流れを示すフローチャートである。

【図 2 4】本開示の一実施形態に係るハードウェア構成例を示す図である。

【発明を実施するための形態】

【0009】

以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

30

【0010】

なお、説明は以下の順序で行うものとする。

1. 実施形態

1. 1. 概要

1. 2. システム構成例

1. 3. 第 1 の動作体 1 0 の機能構成例

1. 4. 第 2 の動作体 2 0 の機能構成例

1. 5. 情報処理サーバ 3 0 の機能構成例

40

1. 6. 機能の詳細

2. 情報処理サーバ 3 0 のハードウェア構成例

3. まとめ

【0011】

< 1. 実施形態 >

<< 1. 1. 概要 >>

近年、ユーザの発話などを認識し、認識結果に基づく動作を実行する種々の装置が開発されている。上記のような装置には、例えば、ユーザとの音声対話を介して種々の機能提供を行う音声エージェント装置や、認識したユーザの発話や周囲環境などに応じて振る舞いを変化させる自律動作体が挙げられる。

50

【0012】

上記のような装置には、収集した音声データをネットワークを介してサーバに送信し、サーバ側で実行された音声認識処理の結果に基づいて動作を行うものと、筐体（クライアント）に搭載した演算機で音声認識処理を実行するものとが存在する。

【0013】

サーバ側で音声認識処理を実行する場合、演算資源を豊富に用意することが比較的容易なことから、一般的に高い認識精度が期待される。一方、クライアント側で音声認識処理を実行する場合、演算資源が限定される場合が多いものの、認識可能な語彙数を抑えることなどにより、演算量の低減と認識精度の向上を図りながら、速いレスポンスを実現することが可能である。

10

【0014】

しかし、クライアント側で音声認識処理を実行する場合において、ユーザが発する語彙のセットと、予め用意された音声認識辞書の語彙のセットとに乖離がある場合、ユーザの発話を正確に認識することが困難である。

【0015】

また、例えば、音声を集める装置が比較的大きな動作音を伴う動作を行う場合、動作状況によっては、自身の動作音が音声収集の妨げとなり、音声認識の精度が低下する可能性がある。

【0016】

本開示に係る技術思想は上記のような点に着目して発想されたものであり、音声認識処理に基づく動作を行う複数の動作体が存在する環境において、各動作体がより状況に適した動作を実現することを可能とする。

20

【0017】

このために、本開示の一実施形態に係る情報処理方法は、プロセッサが、音声認識処理の結果に基づいて動作体の動作を制御すること、を含む。また、上記制御することは、一の動作体が収集した音声に基づく音声認識処理の結果、または一の動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の動作体を制御すること、をさらに含む。

【0018】

<<1. 2. システム構成例>>

30

図1は、本開示の一実施形態に係る情報処理システムの構成例を示す図である。本実施形態に係る情報処理サーバ30は、音声認識処理に基づく動作を行う複数の動作体を備えてよい。図1に示す一例の場合、本実施形態に係る情報処理システムは、第1の動作体10、第2の動作体20、および情報処理サーバ30を備えている。また、各構成は、ネットワーク40を介して互いに通信が可能なように接続される。

【0019】

(第1の動作体10)

本実施形態に係る第1の動作体10は、音声認識処理に基づいて動作する動作体（情報処理装置）の一例である。本実施形態に係る第1の動作体10は、収集した音声の情報処理サーバ30に送信し、情報処理サーバ30による音声認識処理の結果に基づいて動作する。本実施形態に係る第1の動作体10は、上記の音声認識処理の結果に基づいて、ユーザの発話に対するレスポンス発話を行うことや、認識した環境などに基づいて能動的にユーザに語りかけることなどができてよい。

40

【0020】

また、本実施形態に係る第1の動作体10は、例えば、底部に備える車輪により走行や回転などが可能な自律移動体であってもよい。本実施形態に係る第1の動作体10は、例えば、卓上に設置が可能な程度の大きさや形状を有してもよい。本実施形態に係る第1の動作体10は、情報処理サーバ30による音声認識処理の結果や、各種の認識結果に基づいて、自律的に移動を行いながら、発話を含む各種の動作を実行する。

【0021】

50

(第2の動作体20)

本実施形態に係る第2の動作体20は、第1の動作体10と同様に、音声認識処理に基づいて動作する動作体(情報処理装置)の一例である。一方、本実施形態に係る第2の動作体20は、第1の動作体10とは異なり、自身、すなわちローカル側で音声認識処理を実行する。

【0022】

本実施形態に係る第2の動作体20は、例えば、図示するように、イヌなどの動物を模した四足歩行型の自律移動体であってもよい。本実施形態に係る第2の動作体20は、ローカル側で実行した音声認識処理の結果などに基づいて、ユーザに対する応答を動作や鳴き声などにより実行する。

10

【0023】

(情報処理サーバ30)

本実施形態に係る情報処理サーバ30は、第1の動作体10が収集した音声に基づく音声認識処理や自然言語理解処理を行う情報処理装置である。また、本実施形態に係る情報処理サーバ30は、一の動作体が収集した音声に基づく音声認識処理の結果や一の動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の動作体の動作を制御してもよい。

【0024】

(ネットワーク40)

ネットワーク40は、上記の各構成を接続する機能を有する。ネットワーク40は、インターネット、電話回線網、衛星通信網などの公衆回線網や、Ethernet(登録商標)を含む各種のLAN(Local Area Network)、WAN(Wide Area Network)などを含んでもよい。また、ネットワーク40は、IP-VPN(Internet Protocol-Virtual Private Network)などの専用回線網を含んでもよい。また、ネットワーク40は、Wi-Fi(登録商標)、Bluetooth(登録商標)など無線通信網を含んでもよい。

20

【0025】

以上、本実施形態に係る情報処理システムの構成例について述べた。なお、図1を用いて説明した上記の構成はあくまで一例であり、本実施形態に係る情報処理システムの構成は係る例に限定されない。例えば、本実施形態に係る情報処理システムは、3種以上の動作体を備えてもよい。また、動作体の一部は自律移動体に限定されず、据え置き型や屋内埋め込み型のエージェント装置であってもよい。本実施形態に係る情報処理システムの構成は、仕様や運用に応じて柔軟に変形可能である。

30

【0026】

<<1.3. 第1の動作体10の機能構成例>>

次に、本実施形態に係る第1の動作体10の機能構成例について述べる。図2は、本実施形態に係る第1の動作体10の機能構成例を示すブロック図である。図2に示すように、本実施形態に係る第1の動作体10は、音入力部110、撮影部120、センサ部130、トリガ検出部140、制御部150、駆動部160、音出力部170、表示部180、および通信部190を備える。

40

【0027】

(音入力部110)

本実施形態に係る音入力部110は、ユーザの発話音声を含む各種の音を収集する。このために、本実施形態に係る音入力部110は、1つ以上のマイクロフォンを備える。

【0028】

(撮影部120)

本実施形態に係る撮影部120は、ユーザや周囲環境の画像を撮影する。このために、本実施形態に係る撮影部120は、撮像素子を備える。

【0029】

本実施形態に係るセンサ部130は、各種のセンサデバイスによりユーザや周囲環境、

50

また第1の動作体10に係るセンサ情報を収集する。本実施形態に係るセンサ部130は、例えば、ToFセンサ、慣性センサ、赤外線センサ、照度センサ、ミリ波レーダ、タッチセンサ、GNSS (Global Navigation Satellite System) 信号受信機などを備える。

【0030】

(トリガ検出部140)

本実施形態に係るトリガ検出部140は、音入力部110、撮影部120、およびセンサ部130が収集した各種の情報に基づいて、音声認識処理の開始に係る各種のトリガを検出する。

【0031】

例えば、本実施形態に係るトリガ検出部140は、音入力部110が収集した発話音声と、ユーザにより自由に登録された特定発話表現とに基づいて、特定ワード(起動ワード)を検出してもよい。

【0032】

また、例えば、本実施形態に係るトリガ検出部140は、撮影部120が撮影した画像に基づいて、ユーザの顔や体の検出や、特定ジェスチャの検出を行ってもよい。

【0033】

また、例えば、本実施形態に係るトリガ検出部140は、センサ部130が収集した加速度情報に基づいて、ユーザによる第1の動作体10の持ち上げや静置を検出してもよい。

【0034】

(制御部150)

本実施形態に係る制御部150は、第1の動作体10が備える各構成を制御する。また、本実施形態に係る制御部150は、後述するように、情報処理サーバ30による音声認識処理の結果などを第2の動作体20に伝達し、間接的あるいは直接的に第2の動作体20の動作を制御してもよい。

【0035】

(駆動部160)

本実施形態に係る駆動部160は、制御部150による制御に基づいて、各種の動作を行う。本実施形態に係る駆動部160は、例えば、複数のアクチュエータ(モータ等)や車輪等を備えてもよい。

【0036】

(音出力部170)

本実施形態に係る音出力部170は、制御部150による制御に基づいてシステム音声等の出力を行う。このために、本実施形態に係る音出力部170は、アンプやスピーカを備える。

【0037】

(表示部180)

本実施形態に係る表示部180は、制御部150による制御に基づいて、視覚情報の提示を行う。本実施形態に係る表示部180は、例えば、眼に対応するLEDやOLED等を備える。

【0038】

(通信部190)

本実施形態に係る通信部190は、ネットワーク40を介して第2の動作体20や情報処理サーバ30との情報通信を行う。例えば、本実施形態に係る通信部190は、音入力部110が収集したユーザの発話音声を情報処理サーバ30に送信し、当該発話音声に対応する音声認識結果や応答情報を受信する。

【0039】

以上、本実施形態に係る第1の動作体10の機能構成例について説明した。なお、図2を用いて説明した上記の構成はあくまで一例であり、本実施形態に係る第1の動作体10の機能構成は係る例に限定されない。本実施形態に係る第1の動作体10の機能構成は、

10

20

30

40

50

仕様や運用に応じて柔軟に変形可能である。

【0040】

<<1. 4. 第2の動作体20の機能構成例>>

次に、本実施形態に係る第2の動作体20の機能構成例について述べる。図3は、本実施形態に係る第2の動作体20の機能構成例を示すブロック図である。図3に示すように、本実施形態に係る第2の動作体20は、音入力部210、撮影部220、センサ部230、認識部240、制御部250、駆動部260、音出力部270、表示部280、および通信部290を備える。

【0041】

(音入力部210)

本実施形態に係る音入力部210は、ユーザの発話音声を含む各種の音を収集する。このために、本実施形態に係る音入力部210は、1つ以上のマイクロフォンを備える。

【0042】

(撮影部220)

本実施形態に係る撮影部220は、ユーザや周囲環境の画像を撮影する。このために、本実施形態に係る撮影部220は、撮像素子を備える。撮影部220は、例えば、第2の動作体20の鼻先と腰部に2つの広角カメラを備えてもよい。この場合、鼻先に配置される広角カメラは、第2の動作体20の前方視野（すなわち、イヌの視野）に対応した画像を撮像し、腰部の広角カメラは、上方を中心とする周囲領域の画像を撮像する。第2の動作体20は、例えば、腰部に配置される広角カメラにより撮像された画像に基づいて、天井の特徴点などを抽出し、SLAM (Simultaneous Localization and Mapping) を実現することができる。

【0043】

本実施形態に係るセンサ部230は、各種のセンサデバイスによりユーザや周囲環境、また第2の動作体20に係るセンサ情報を収集する。本実施形態に係るセンサ部230は、例えば、測距センサ、慣性センサ、赤外線センサ、照度センサ、タッチセンサ、接地センサなどを備える。

【0044】

(認識部240)

本実施形態に係る認識部240は、音入力部210、撮影部220、およびセンサ部230が収集した情報に基づいて、各種の認識処理を実行する。例えば、本実施形態に係る認識部240は、音入力部210が収集したユーザの発話音声に基づく音声認識処理をローカル側で実行する。また、認識部240は、話者識別、表情や視線の認識、物体認識、動作認識、空間領域認識、色認識、形認識、マーカー認識、障害物認識、段差認識、明るさ認識などを行ってよい。

【0045】

(制御部250)

本実施形態に係る制御部250は、認識部240による各種の認識処理の結果に基づいて、第2の動作体20が備える各構成を制御する。また、本実施形態に係る制御部250は、認識部240による各種の認識処理の結果を他の動作体（例えば、第1の動作体または第2の動作体）に伝達し、当該他の動作体の動作を間接的あるいは直接的に制御してもよい。

【0046】

(駆動部260)

駆動部260は、制御部250による制御に基づいて、第2の動作体20が有する複数の関節部を屈伸させる機能を有する。より具体的には、駆動部260は、制御部250による制御に基づき、各関節部が備えるアクチュエータを駆動させる。

【0047】

(音出力部270)

本実施形態に係る音出力部270は、制御部250による制御に基づいてイヌの鳴き声

10

20

30

40

50

を模した音等の出力を行う。このために、本実施形態に係る音出力部 170 は、アンプやスピーカを備える。

【0048】

(表示部 280)

本実施形態に係る表示部 280 は、制御部 250 による制御に基づいて、視覚情報の提示を行う。本実施形態に係る表示部 280 は、例えば、眼に対応する LED や OLED 等を備える。

【0049】

(通信部 290)

本実施形態に係る通信部 290 は、ネットワーク 40 を介して第 1 の動作体 10 や情報処理サーバ 30 との情報通信を行う。例えば、本実施形態に係る通信部 290 は、第 1 の動作体 10 または情報処理サーバ 30 から、他の動作体が収集した音声に基づく音声認識処理の結果を受信する。

【0050】

以上、本実施形態に係る第 2 の動作体 20 の機能構成例について説明した。なお、図 3 を用いて説明した上記の構成はあくまで一例であり、本実施形態に係る第 2 の動作体 20 の機能構成は係る例に限定されない。本実施形態に係る第 2 の動作体 20 の機能構成は、仕様や運用に応じて柔軟に変形可能である。

【0051】

<< 1. 5. 情報処理サーバ 30 の機能構成例 >>

次に、本実施形態に係る情報処理サーバ 30 の機能構成例について述べる。図 4 は、本実施形態に係る情報処理サーバ 30 の機能構成例を示すブロック図である。図 4 に示すように、本実施形態に係る情報処理サーバ 30 は、音声認識部 310、自然言語処理部 320、制御部 330、応答生成部 340、通信部 350 を備える。

【0052】

(音声認識部 310)

本実施形態に係る音声認識部 310 は、第 1 の動作体 10 から受信した発話音声に戻づく音声認識処理を行い、当該発話音声を文字列に変換する。

【0053】

(自然言語処理部 320)

本実施形態に係る自然言語処理部 320 は、音声認識部 310 が生成した文字列に基づく自然言語理解処理を行い、ユーザの発話の意図を抽出する。

【0054】

(制御部 330)

本実施形態に係る制御部 330 は、一の動作体が収集した音声に基づく音声認識処理の結果や、一の動作体が収集したセンサ情報に基づいて認識された環境に基づいて、一の動作体や他の動作体の動作を制御する。本実施形態に係る制御部 330 が有する機能の詳細については、別途後述する。

【0055】

(応答生成部 340)

本実施形態に係る応答生成部 340 は、制御部 330 による制御に基づいて、自然言語処理部 320 が抽出したユーザの発話意図に対応する応答情報を生成する。応答生成部 340 は、例えば、ユーザの発話に対する応答音声等を生成する。

【0056】

(通信部 350)

本実施形態に係る通信部 350 は、ネットワーク 40 を介して第 1 の動作体 10 や第 2 の動作体 20 との情報通信を行う。例えば、通信部 350 は、第 1 の動作体 10 から発話音声を受信する。また、通信部 350 は、上記発話音声に基づく音声認識処理および自然言語理解処理の結果に係る情報を第 2 の動作体 20 に送信する。

【0057】

10

20

30

40

50

以上、本実施形態に係る情報処理サーバ30の機能構成例について説明した。なお、図4を用いて説明した上記の構成はあくまで一例であり、本実施形態に係る情報処理サーバ30の機能構成は係る例に限定されない。本実施形態に係る情報処理サーバ30の機能構成は、仕様や運用に応じて柔軟に変形可能である。

【0058】

<<1.6. 機能の詳細>>

次に、本実施形態に係る情報処理システムが有する機能について詳細に説明する。上述したように、本実施形態に係る情報処理方法は、音声認識処理に基づき動作する複数の動作体が存在する環境において、各動作体が収集した情報から認識された情報を共有することにより、各動作体がより適切な動作を行うことを実現するものである。

【0059】

このために、本実施形態に係る情報処理システムでは、機能や特性が異なる複数種類の動作体を備えてもよい。例えば、本実施形態に係る情報処理システムは、上述した第1の動作体10と第2の動作体20とを備え得る。

【0060】

本実施形態に係る第1の動作体10は、情報処理サーバ30による語彙数が豊富なクラウド辞書（以下、第1の辞書、とも称する）を用いた音声認識処理（以下、第1の音声認識処理、とも称する）、および自然言語理解処理の結果に基づいて動作することが可能であり、第2の動作体20と比較して、より広い語彙および発話意図を精度高く認識し応答動作を行うことが可能である。

【0061】

また、本実施形態に係る第1の動作体10は、底部に備える車輪により移動することから、サーボ音や接地ノイズが大きい第2の動作体20の歩行移動と比較して、移動音（動作音）が静かであることが特徴の一つである。

【0062】

また、本実施形態に係る第1の動作体10は、卓上に配置されることを想定していることから、第2の動作体20と比較して高い視点を有し、第2の動作体20よりも広い視野でユーザや物体などを認識することが可能である。

【0063】

さらには、本実施形態に係る第1の動作体10は、言語を用いてユーザとの対話が可能であり、認識したユーザの発話や環境などについて、より確実にユーザに伝達することができる。

【0064】

上記に対し、本実施形態に係る第2の動作体20は、ローカル側において、クラウド辞書よりも語彙数が少ないローカル辞書（以下、第2の辞書、とも称する）を用いて音声認識処理（以下、第2の音声認識処理、とも称する）を行い、当該音声認識処理の結果に基づいて動作する。このため、本実施形態に係る第2の動作体20は、予め登録された語彙のみを認識可能である一方、語彙数を抑えることで、演算量を効果的に低減し、速いレスポンスを行うことが可能である。

【0065】

また、本実施形態に係る第2の動作体20は、床上を歩行移動することから、第1の動作体10と比較して広い行動範囲を有し、複数の部屋間を移動することができる。さらには、本実施形態に係る第2の動作体20は、SLAMにより環境地図を生成することができ、また、ユーザからの教示などにより空間の名称（例えば、リビングルームやキッチンなど）と環境地図とを対応付けることも可能である。

【0066】

このように、本実施形態に係る第1の動作体10および第2の動作体20は、互いに対し優位な点をそれぞれに有する。以下、上記のような特徴を有する第1の動作体10と第2の動作体20との情報連携について、具体例を挙げながら説明する。

【0067】

まず、本実施形態に係る音声認識結果の共有について説明する。本実施形態に係る情報処理サーバ30の制御部330は、一の動作体が収集した音声に基づく音声認識処理の結果に基づいて、当該音声認識処理の結果に対応する動作を他の動作体に行わせてよい。例えば、制御部330は、ある第1の動作体10が収集した音声に基づく第1の音声認識処理の結果を第2の動作体20に伝達し、第2の動作体20の動作を間接的あるいは直接的に制御することが可能である。

【0068】

図5は、本実施形態に係る音声認識結果の共有について説明するための図である。図5の上段に示す一例では、ユーザUが“ That ' s my boy ”という、第2の動作体20を褒める意図（ good FB ）を有する発話U01を行っている。

10

【0069】

しかし、図5に示す一例の場合、第2の動作体20が備える第2の辞書242には、カテゴリ： good FB に対応する語彙が、“ Good boy ”のみしか登録されていないため、第2の動作体20は、ユーザUの発話U01を正確に認識することが困難である。

【0070】

一方、情報処理サーバ30が備える第1の辞書322には、“ That ' s my boy ”が登録されているため、情報処理サーバ30は、第1の音声認識処理により発話U01を精度高く認識し、また認識結果を第1の動作体10に伝達することができる。

【0071】

このように、第2の動作体20が備える第2の辞書242にユーザの発話に対応する語彙が登録されていない場合、情報処理サーバ30の制御部330は、第1の動作体10が収集した音声と第1の辞書322に基づく第1の音声認識処理の結果に基づいて、当該悔過に対応する動作を第2の動作体20に行わせてもよい。

20

【0072】

より具体的には、本実施形態に係る制御部330は、第1の音声認識処理の結果として得られた語彙のカテゴリに対応する語彙を第2の辞書242から抽出し、当該語彙を第2の動作体20に伝達してよい。

【0073】

図5の下段に示す一例の場合、制御部330は、第1の音声認識処理により得られた語彙“ That ' s my boy ”のカテゴリ： good FB に対応する語彙“ Good boy ”を第2の辞書242から抽出し、語彙“ Good boy ”を第2の動作体20に伝達している。係る制御によれば、第2の動作体20が、自力では認識することができない語彙“ That ' s my boy ”を疑似的に理解し、ユーザUの発話意図に応じた適切な動作を実行することが可能となる。なお、制御部330は、第1の音声認識処理により得られた語彙に加えて、あるいは代えて、当該語彙のカテゴリに係る情報を第2の動作体20に伝達してもよい。

30

【0074】

また、上記のような音声認識結果の共有を行う場合、本実施形態に係る制御部330は、一の動作体が収集した音声に基づく音声認識処理の結果を他の動作体に伝達したことを示す動作を当該一の動作体に行わせてもよい。

40

【0075】

例えば、本実施形態に係る制御部330は、第1の動作体10が収集した音声に基づく第1の音声認識処理の結果を第2の動作体20に伝達したことを示す動作を第1の動作体10に行わせてもよい。図5の下段に示す一例の場合、制御部330は、第1の動作体10に、“ He praised you ”というシステム発話S01を第2の動作体20に向けて出力させている。

【0076】

上記の例の他、制御部330は、例えば、「僕が教えてあげたんだ」などのシステム発話や、第1の動作体10がユーザUの発話U01を第2の動作体20に対して通訳しているようなシステム発話（例えば、犬の鳴き声を模した音を用いた発話）を第1の動作体1

50

0に実行させてもよい。

【0077】

また、制御部330は、音声認識結果の共有先である第2の動作体にも、共有が行われたことを示す動作を実行させてもよい。上記のような第1の音声認識結果の伝達を行った場合、制御部330は、例えば、通常時よりも大きな音声や動きを第2の動作体20に行わせてもよい。また、制御部330は、自力で理解できなかったことを恥ずかしそうにする振る舞いや、第1の動作体10に対して感謝を示す振る舞いなどを第2の動作体20に実行させてもよい。

【0078】

上記のような制御によれば、第1の動作体10と第2の動作体20とが実際の生物のようにコミュニケーションを行っているような表現を実現することができ、ユーザの興味をさらに引き付けることが期待される。

【0079】

以上、本実施形態に係る音声認識結果の共有について説明した。なお、上記では、共有に係る制御主体が情報処理サーバ30の制御部330である場合を例に述べたが、音声認識結果の共有に係る制御主体は、第1の動作体10の制御部150であってもよい。制御部150は、情報処理サーバ30から第1の音声認識処理の結果を受信した後、当該結果を近距離無線通信などにより第2の動作体20に伝達することが可能である。

【0080】

続いて、本実施形態に係る音声認識結果の共有の処理の流れについて説明する。図6Aは、本実施形態に係る音声認識結果の共有における第1の動作体10および情報処理サーバ30の動作の流れを示すフローチャートである。

【0081】

図6Aを参照すると、第1の動作体10は、まず、認識した環境に基づいて自律動作を行う(S1101)。

【0082】

次に、制御主体となる制御部150または制御部330は、第1の音声認識処理により認識された語彙に対応するカテゴリの語彙が第2の動作体20が備える第2の辞書242に存在するか否かを判定する(S1102)。

【0083】

ここで、対応する語彙が第2の辞書242に存在する場合(S1102: YES)、制御主体は、続いて、発話が第2の動作体20に対するものか否かを判定する(S1103)。制御部150や制御部330は、第1の動作体10のトリガ検出部140がユーザの顔が第2の動作体20に向いていることを検出したこと、第2の動作体20がユーザによる接触を検出したこと、直前の発話が第2の動作体の名前を含んでいたこと、などに基づいて、上記の判定を行うことができる。

【0084】

ここで、発話が第2の動作体20に対するものである場合(S1103: YES)、制御主体は、対応する語彙を第2の動作体20に伝達する(S1104)。

【0085】

一方、第2の辞書242に対応する語彙が存在しない場合(S1102: NO)や、発話が第2の動作体20に対するものではない場合(S1103: NO)、第1の動作体10は、ステップS1101に復帰する。

【0086】

図6Bは、本実施形態に係る音声認識結果の共有における第2の動作体20の動作の流れを示すフローチャートである。

【0087】

図6Bを参照すると、第2の動作体20は、まず、認識した環境に基づいて自律動作を行う(S1201)。

【0088】

次に、第２の動作体２０の制御部２５０は、第１の動作体１０または情報処理サーバ３０から、語彙を受信したか否かを判定する（Ｓ１２０２）。

【００８９】

ここで、語彙が受信されている場合（Ｓ１２０２：ＹＥＳ）、制御部２５０は、他装置からの伝達により語彙を理解したことを示す動作を第１の動作体１０に実行させる（Ｓ１２０３）。

【００９０】

一方、語彙が受信されていない場合（Ｓ１２０２：ＮＯ）、第２の動作体２０は、ステップＳ１２０１に復帰する。

【００９１】

以上、本実施形態に係る音声認識結果の共有の処理の流れについて説明した。続いて、本実施形態に係る第１の音声認識処理の結果のログに基づく第２の辞書の更新について述べる。

【００９２】

上述したように、本実施形態に係る第２の動作体２０は、第２の辞書２４２に登録する語彙数を抑えることで、演算量を抑え速いレスポンスを実現することができる。一方、ユーザが日常的に使用する語彙セットと第２の辞書２４２に登録されている語彙セットとの間に乖離がある場合、認識精度が低下しユーザ体験を損なう可能性が生じる。このために、本実施形態に係る情報処理システムは、第２の辞書２４２に登録される語彙セットを定期または非定期に入れ替える仕組みを有してよい。

【００９３】

図７は、本実施形態に係る第２の辞書２４２に登録される語彙セットの入れ替えについて説明するための図である。なお、図７では、情報処理システムが備える構成のうち、語彙セットの入れ替え機能に着目して作成されたブロック図であり、一部の構成が省略されている。

【００９４】

第２の辞書２４２に登録される語彙の入れ替えを実現するために、本実施形態に係る第２の動作体２０は、評価部７３５、辞書更新部７４０、および、第２の辞書２４２とは別途の候補辞書７４５を備えてよい。

【００９５】

本実施形態に係る評価部７３５は、蓄積された単独知に基づいて、ユーザの発話実績を評価する。ここで、上記の単独知とは、第２の動作体２０の独自の経験に基づく知識を指す。具体的には、本実施形態に係る単独知には、認識部２４０が認識した発話ログ７２０や発話時におけるステータスを記録した発話時ステータスログ７３０が含まれる。

【００９６】

この際、本実施形態に係る評価部７３５は、認識された回数が多い語彙ほど発話実績が高いと評価してもよい。一方、認識部２４０による誤認識や、湧き出しなどへの対応のため、評価部７３５は、発話時ステータスなどに基づいて発話実績を総合的に評価してよい。ここで、誤認識とは、ユーザの実際の発話「おはよう」に対して認識部２４０が「おはよう」以外の認識結果を出力すること、例えば、「おやすみ」と出力してしまうことを意味する。また、湧き出しとは、生活音などユーザの発話以外の音に対して認識結果を出力すること、例えば、ドアを閉めた音に対して「ばん」という認識結果を出力してしまうことを意味する。

【００９７】

また、本実施形態に係る辞書更新部７４０は、制御部２５０による制御に基づき、評価部７３５が評価した発話実績に応じた第２の辞書２４２および候補辞書７４５の更新を実行する。ここで、本実施形態に係る候補辞書７４５は、第２の辞書２４２への追加または入れ替え対象となる語彙が登録される辞書である。辞書更新部７４０は、例えば、発話実績が低い語彙を第２の辞書２４２から削除、または候補辞書７４５に移行し、候補辞書７４５に登録される優先度の高い語彙を代わりに第２の辞書２４２に登録することなどがで

10

20

30

40

50

きる。

【0098】

また、本実施形態に係る辞書更新部740は、情報処理サーバ30から未登録の語彙を取得し、当該語彙を第2の辞書242や候補辞書745に追加登録する機能を有してよい。辞書更新部740は、例えば、関連語彙931、トレンド語彙932、季節語彙933、世代別語彙934、方言935などを取得し、第2の辞書242や候補辞書745に追加登録してもよい。この際、語彙数を一定以下に保つため、辞書更新部740は、第2の辞書242や候補辞書745から、発話実績が低い語彙を削除してもよい。

【0099】

なお、上記の関連語彙931とは、ある語彙に続けて発話される別の語彙など、他の語彙や機能との関連性の高い語彙を指す。関連語彙931は、例えば、複数の第2の動作体20が記録した発話ログ720や発話時ステータスログ730を統合的に蓄積した集合知915の分析に基づき取得されてもよい。

10

【0100】

また、上記のトレンド語彙932とは世間においてトレンドとなっている語彙を、季節語彙933とは季節に応じた語彙を、世代別語彙934とは世代ごとに多用される語彙を、それぞれ指す。

【0101】

以上、第2の辞書242の更新に係る第2の動作体20の構成について述べた。上述した構成によれば、ユーザの発話実績に基づいて第2の辞書242を適宜更新することができ、ユーザが用いる語彙セットと第2の辞書242に登録される語彙セットとの乖離を効果的に抑えることができる。なお、上記で説明した辞書更新の詳細については、本開示の願人が過去に出願を行った特願2018-124856を参照されたい。

20

【0102】

一方、ユーザの発話実績を第2の動作体20が蓄積する発話ログ720や発話時ステータスログ730のみに基づいて評価する場合、分析を行うために必要な情報の蓄積には時間を要する場合も想定される。このため、本実施形態に係る制御部250は、第1の動作体10が収集した音声に基づく第1の音声認識処理の結果に係る認識ログ324をさらに用いた辞書更新を辞書更新部740に実行させてもよい。すなわち、本実施形態に係る制御部250は、認識ログ324に基づいて、第2の辞書242や候補辞書745の入れ替えを制御することができる。

30

【0103】

上記の制御によれば、第1の音声認識処理により精度高く認識された語彙とその回数に基づいて、ユーザの発話実績をさらに正確に評価することができ、辞書更新の精度を向上させるとともに、評価に必要なログの蓄積を最小化することが可能となる。

【0104】

図8は、本実施形態に係る情報処理サーバ30による認識ログの蓄積の流れを示すフローチャートである。

【0105】

図8を参照すると、まず、音声認識部310が第1の動作体10から受信した音声データに基づいて第1の音声認識処理を行う(S1301)。

40

【0106】

続いて、音声認識部310は、ステップS1301における第1の音声認識処理により複数回認識した語彙があるか否かを判定する(S1302)。

【0107】

ここで、複数回認識した語彙がある場合(S1302: YES)、音声認識部310は、認識した語彙とその回数を認識ログ324に保存する(S1303)。

【0108】

一方、複数認識した語彙がない場合(S1302: NO)、情報処理サーバ30は、ステップS1301に復帰する。

50

【0109】

図9は、本実施形態に係る第2の動作体20による辞書更新の流れを示すフローチャートである。

【0110】

図9を参照すると、まず、制御部250が、辞書の更新に係る入れ替えアルゴリズムを辞書更新部740に実行させる（S1401）。

【0111】

次に、制御部250は、情報処理サーバ30に蓄積される認識ログ324に語彙が存在するか否かを判定する（S1402）。

【0112】

ここで、認識ログ324に語彙が存在する場合（S1402：YES）、制御部250は、辞書更新部740に当該語彙と関連する他の語彙を追加で選定させる（S1403）。辞書更新部740は、認識ログ324に登録されている語彙のプロファイルに基づいて、相関語彙931、トレンド語彙932、季節語彙933、世代別語彙934、方言935などから追加で登録する語彙を選定してよい。

【0113】

次に、制御部250は、辞書更新部740を制御し、認識ログ324から取得された語彙と関連する語彙とを第2の辞書242や候補辞書745に追加させる（S1404）。

【0114】

以上、本実施形態に係る辞書更新について説明した。なお、上記では、第2の動作体20の制御部250が、辞書更新の制御主体となる場合を例に説明したが、本実施形態に係る辞書更新の制御主体は、情報処理サーバ30の制御部330であってもよい。

【0115】

次に、本実施形態に係る認識環境の共有に基づく動作制御について説明する。本実施形態に係る情報処理方法では、音声認識結果の他、各動作体が収集された情報に基づいて認識された環境を複数の動作体の間で共有し、また動作体の動作制御に利用することができてよい。

【0116】

例えば、本実施形態に係る制御部330は、一の動作体に係る音声認識環境が音声認識処理精度の低下要因となり得る環境である場合、他の動作体に係る音声認識環境が向上するよう他の動作体の動作を制御してもよい。

【0117】

ここで、上記の音声認識環境とは、音声認識精度に影響を及ぼす各種の環境を指し、例えば、動作体の動作状況を含んでよい。例えば、動作体がサーボ音や接地音などの比較的大きな動作音を伴う動作を行っている場合、当該動作音がユーザの発話音声の収集を妨害し、音声認識精度が低下する場合は想定される。

【0118】

このため、本実施形態に係る制御部330は、一の動作体の動作状況に基づいて、他の動作体の動作を制御してもよい。より詳細には、本実施形態に係る制御部330は、一の動作体が比較的大きな動作音を伴う動作を実行している場合、他の動作体の動作音がより低下するよう制御を行う。

【0119】

図10は、本実施形態に係る一の動作体の動作状況に基づく他の動作体の動作制御について説明するための図である。図10に示す一例の場合、第2の動作体20がボールで遊ぶ動作を行っている状況において、ユーザUが、発話U02を行っている。この場合、第2の動作体20は、自身の動作音に妨げられ発話U02に係る音声を精度高く収集できない可能性がある。

【0120】

この際、本実施形態に係る情報処理サーバ30の制御部330は、第2の動作体20からアップロードされる動作ステータス（2nd_MOVING：TRUE）に基づいて、

10

20

30

40

50

第2の動作体20による第2の音声認識処理の精度低下を予測することができてよい。また、制御部330は、第1の動作体10が発話UO2に係る音声を精度高く収集することができるように、第1の動作体10の動作音がより小さくなるよう制御を行う。制御部330は、例えば、第1の動作体10を静止させてもよい。

【0121】

上記の制御によれば、動作中である第2の動作体20に代わり、第1の動作体10がユーザUの発話UO2に係る音声を収集することができ、さらには、第1の音声認識処理の結果を第2の動作体20に伝達することにより、第2の動作体20が動作中であっても発話UO2に対応する動作を行うことが可能となる。

【0122】

図11は、本実施形態に係る一の動作体の動作状況に基づく他の動作体の動作制御の流れを示すフローチャートである。

【0123】

図11を参照すると、まず、第1の動作体10が自律動作を行っている状況(S1501)において、制御部330が第2の動作体20が動作中であるか否かを判定する(S1502)。上述したように、制御部330は、第2の動作体20からアップロードされる動作ステータスに基づいて上記の判定を行うことができる。

【0124】

ここで、第2の動作体20が動作中である場合(S1502: YES)、制御部330は、第1の動作体10の動作音がより小さくなるように制御を行う(S1503)。

【0125】

次に、制御部330は、第1の音声認識処理により認識された語彙に対応するカテゴリの語彙が第2の動作体20が備える第2の辞書242に存在するか否かを判定する(S1504)。

【0126】

ここで、対応する語彙が第2の辞書242に存在する場合(S1504: YES)、制御部330は、対応する語彙を第2の動作体20に伝達する(S1505)。なお、制御部330は、図6Aに示した場合と同様に、発話が第2の動作体20に対するものである場合にのみ、対応する語彙を第2の動作体20に伝達してもよい。また、第2の動作体20は、図6Bに示す流れに従って後続の処理を実行してよい。また、上記では、制御主体が情報処理サーバ30の制御部330である場合を例に述べたが、当該制御主体は、第1の動作体10の制御部150であってもよい。

【0127】

次に、本実施形態に係る他の音声認識環境の共有と動作体の動作制御について述べる。上記では、本実施形態に係る音声認識環境が動作体の動作状況を含むことを述べたが、本実施形態に係る音声認識環境には、湧き出しや話者識別に係る環境、すなわちノイズ音源や話者の認識(識別)結果が含まれてもよい。

【0128】

図12Aは、本実施形態に係るノイズ音源の認識結果の共有について説明するための図である。図12Aには、第1の動作体10aおよび10b、第2の動作体20が同じ部屋の中にいる場合において、第1の動作体10aのみが、テレビジョン装置であるノイズ源NSから音声が出力されていることを認識している状況を示している。

【0129】

この場合、第1の動作体10bおよび第2の動作体20は、ノイズ音源NSが出力する音声をユーザの発話音声と誤認し、音声認識処理を実行する可能性がある。

【0130】

上記のような事態を回避するため、第1の動作体10aは、認識したノイズ音源NSに係る情報を情報処理サーバ30にアップロードしてよい。また、情報処理サーバ30の制御部330は、第1の動作体10aによりアップロードされたノイズ音源NSに係る情報を第1の動作体10bおよび第2の動作体20に伝達し、音声認識処理が実行されないよ

10

20

30

40

50

う間接的あるいは直接的に制御してもよい。

【0131】

図12Bは、本実施形態に係る話者認識の結果の共有について説明するための図である。図12Bには、第1の動作体10aおよび10b、第2の動作体20が同じ部屋の中にいる場合において、第1の動作体10aのみが、ユーザUを識別しており、また、テレビジョン装置であるノイズ源NSから音声が出力されていないことを認識している状況を示している。

【0132】

この場合において、第1の動作体10bおよび第2の動作体20が音声を検出した場合、当該音声は、ユーザUの発話音声である可能性が高いといえる。

【0133】

このため、情報処理サーバ30の制御部330は、第1の動作体10aによりアップロードされたユーザUおよびノイズ音源NSに係る情報を第1の動作体10bおよび第2の動作体20に伝達し、音声を検出した場合、音声認識処理を実行するよう間接的あるいは直接的に制御してもよい。

【0134】

以上説明したように、本実施形態に係る制御部330は、一の動作体が収集したセンサ情報に基づいて認識されたノイズ音源や話者に係る情報を他の動作体に伝達し、また他の動作体に係る音声認識処理の実行有無を制御することができる。

【0135】

なお、ノイズ音源や話者に係る情報の共有は、必ずしも制御部330が制御主体でなくともよい。上記共有の制御主体は、共有元となる動作体（第1の動作体10または第2の動作体）であってもよく、情報処理サーバ30を介さずに直接共有先の動作体に情報を伝達してもよい。

【0136】

図13Aは、本実施形態に係るノイズ音源や話者に係る情報共有における共有元の動作の流れを示すフローチャートである。

【0137】

図13Aを参照すると、まず、共有元の動作体が自律動作を行う（S1601）。

【0138】

ここで、共有元の動作体がノイズ音源や話者を認識した場合（S1602：YES）、共有元の動作体は、認識したノイズ音源や話者に係る情報を他の動作体に伝達する（S1603）。

【0139】

一方、図13Bは、本実施形態に係るノイズ音源や話者に係る情報共有における共有先の動作の流れを示すフローチャートである。

【0140】

図13Bを参照すると、まず、共有先の動作体が自律動作を行う（S1701）。

【0141】

ここで、音声を検出した場合（S1702：YES）、共有先の動作体は、続いて、他の動作体によりノイズ音源や話者に係る情報が共有されているか否かを判定する（S1703）。

【0142】

ここで、他の動作体によりノイズ音源や話者に係る情報が共有されている場合（S1703：YES）、共有先の動作体は、共有されている情報に応じた動作を行う（S1704）。共有先の動作体は、例えば、ノイズ音源が多い環境では、「なんだかうるさいなあ、顔みせてよ」などの発話を行うことで、ユーザを識別することができる可能性を高めてもよい。

【0143】

以上説明したように、本実施形態に係る情報処理方法によれば、ある動作体が認識した

10

20

30

40

50

環境を他の動作体に伝達することで、共有先の動作体が当該環境に応じたより適切な動作を行うことが可能となる。

【0144】

なお、上記のような環境の共有は、動作体によるユーザに対する能動的なアクションに用いられてもよい。例えば、本実施形態に係る第1の動作体10は、ユーザの発話に対する応答のみではなく、ユーザに対し能動的に語り掛けることで、積極的な対話を行うことを特徴の一つとする。また、本実施形態に係る第2の動作体20も同様にユーザに対する能動的な振る舞いを行うことで積極的なインタラクションを実現することを特徴の一つとする。

【0145】

このために、本実施形態に係る情報処理方法では、認識された環境を複数の動作体で共有することで、能動的なアクションの実行頻度を効果的に高めることができよう。

【0146】

図14は、本実施形態に係る能動的なアクションのトリガとなる環境の共有について説明するための図である。図14には、キッチンに居るユーザUを同じくキッチンに居る第2の動作体20aが識別している状況が示されている。

【0147】

この際、第2の動作体20aは、ユーザUをキッチンにおいて識別したことを示す情報を情報処理サーバ30にアップロードする。また、情報処理サーバ30の制御部330は、第2の動作体20aによりアップロードされた情報をリビングルームにいる第1の動作体10および第2の動作体20bに伝達し、ユーザに対する能動的なアクションを実行するよう間接的あるいは直接的に制御してもよい。

【0148】

例えば、制御部330は、第1の動作体10に、ユーザUがキッチンにいることを把握しているようなシステム発話S03を行わせてもよい。また、例えば、制御部330は、第2の動作体20bをキッチンに移動させるなどの制御を行ってもよい。

【0149】

以上説明したように、本実施形態に係る制御部330は、一の動作体が収集したセンサ情報に基づいて認識された環境に基づいて、ユーザに対する能動的なアクションを他の動作体を実行させてよい。また、共有先の動作体が発話可能な場合、上記環境に応じた能動的な発話を共有先の動作体を実行させてよい。

【0150】

上記のような制御によれば、各動作体が自身が収集するセンサ情報に基づき認識する環境以上に多くの事象を認識することができ、ユーザに対しより積極的なアクションを行うことで、ユーザ体験を向上させることが可能となる。

【0151】

なお、ノイズ音源や話者に係る情報の共有と同様に、制御主体は、共有元となる動作体（第1の動作体10または第2の動作体）であってもよく、情報処理サーバ30を介さずに直接共有先の動作体に情報を伝達してもよい。

【0152】

図15Aは、本実施形態に係る能動的アクションのトリガとなる環境の共有における共有元の動作の流れを示すフローチャートである。

【0153】

図15Aを参照すると、まず、共有元の動作体が自律動作を行う（S1801）。

【0154】

ここで、共有元の動作体が能動的アクションのトリガとなる環境（トリガ環境、とも称する）を認識した場合（S1802：YES）、共有元の動作体は、認識した環境に係る情報を他の動作体に伝達する（S1803）。上記環境には、例えば、ユーザや他の人物の所在や行動、天気や環境音の認識などが挙げられる。

【0155】

一方、図15Bは、本実施形態に係る能動的アクションのトリガとなる環境の共有における共有先の動作の流れを示すフローチャートである。

【0156】

図15Bを参照すると、まず、共有先の動作体が自律動作を行う（S1901）。

【0157】

ここで、他の動作体により上記トリガとなる環境が共有されている場合（S1902：YES）、共有先の動作体は、共有されている環境に応じた能動的なアクションを行う（S1903）。

【0158】

以上説明したように、本実施形態に係る情報処理方法によれば、ある動作体が認識した環境を他の動作体に伝達することで、共有先の動作体が当該環境に応じた能動的なアクションを実行することが可能となる。

【0159】

なお、本実施形態に係る第1の動作体10は、認識した環境あるいは共有された環境に応じて複数の対話エンジンを切り替えてユーザとの対話を行うことが可能である。

【0160】

図16は、本実施形態に係る対話エンジンの切り替えについて説明するための図である。図16には、環境ごとに使用されるエンジンの種別がそれぞれ示されている。本実施形態に係る第1の動作体10は、例えば、シナリオ対話、状況対話、知識対話、雑談対話に係る4つの異なるエンジンを環境に応じて切り替えることで、より豊かな対話を実現することが可能である。

【0161】

なお、上記のシナリオ対話とは、予め設定された条件と定型発話分の組から成るシナリオに基づき、条件と一致した環境に対応する発話生成を行う対話であってよい。

【0162】

また、状況対話とは、知識データベースを用いて、認識した状況（環境）を説明するような発話の生成を行う対話であってよい。

【0163】

また、上記の知識対話とは、ユーザの発話に含まれる語彙や環境中から推定される語彙を基に知識データベースから必要な他の語彙を抽出して発話生成を行う対話であってよい。

【0164】

また、上記の雑談対話とは、ドメインの定まらない自由な発話に対して、機械学習手法などを用いて発話生成を行ったり、あるいは発話文データベースから適切な定型文を抽出して発話生成を行う対話であってよい。

【0165】

なお、図中において複数のエンジンが対応している場合、優先度の高いエンジンが使用されてもよい。一方、優先度の高いエンジンによる発話生成を試みて、適切な発話が生成できなかった場合には、次に優先度の高いエンジンによる発話生成を試みてよい。

【0166】

例えば、人を見つけて話しかける場合、第1の動作体10は、シナリオ対話または状況対話に係るエンジンを使用することで、「ねえねえ」や「ひさしぶり」などの発話を行うことが可能である。

【0167】

また、人がいなくなった場合、第1の動作体10は、状況対話に係るエンジンを使用することで、「あれ？タロウどこ？」などの発話を行うことが可能である。

【0168】

また、特定の物体を見つけた場合、第1の動作体10は、状況対話に係るエンジンを使用することで、「あ、ボールだ」などの発話を行うことが可能である。

【0169】

また、その場にいる人に質問する場合、第1の動作体10は、シナリオ対話に係るエン

10

20

30

40

50

ジンを使用することで、「そういえば、好きなものは何？」などの発話を行うことが可能である。

【0170】

また、知識のある環境を認識した場合、第1の動作体10は、状況対話または知識対話に係るエンジンを使用することで、「雨だなあ、傘持っていないくちや」や「会議中かあ、静かにしよう」などの発話を行うことが可能である。

【0171】

また、時事ニュースなど外部から新たな情報を入手した場合、第1の動作体10は、知識対話に係るエンジンを使用することで、「ABC駅で遅延発生だって」などの発話を行うことが可能である。

【0172】

また、人の会話を漏れ聞いた場合、第1の動作体10は、知識対話または雑談対話に係るエンジンを使用することで、「ケーキって聞こえたよ。僕はチョコレートケーキが好きだなあ」などの発話を行うことが可能である。

【0173】

以上、本実施形態に係る対話エンジンの切り替えについて説明した。上記のような対話エンジンの切り替えによれば、環境に応じた適切かつより豊かな対話を実現することが可能となる。

【0174】

次に、本実施形態に係る音声認識結果の統合について説明する。上記では、ある動作体により収集された音声に基づく音声認識処理の結果を他の動作体に共有することについて述べたが、本実施形態に係る情報処理方法では、複数の動作体が収集した音声のそれぞれに基づく複数の音声認識処理の結果を統合的に判断し、単一の統合認識結果を決定してもよい。

【0175】

図17Aおよび図17Bは、本実施形態に係る音声認識処理結果の統合について説明するための図である。図17Aには、第1の動作体10a～10c、および第2の動作体20が共に居る環境において、ユーザUが発話U04を行った状況が示されている。

【0176】

図17Aに示す一例では、第1の動作体10aおよび10bは、発話U04に係る音声を精度高く収集し、第1の音声認識処理により正確な語彙を得ている一方、第1の動作体10aはユーザUとの距離が離れているため、発話U04に係る音声を精度高く収集することができず、第1の音声認識処理により誤った語彙を得ている。また、第2の動作体20は、ボールで遊ぶ動作を行っており、第2の音声認識処理に十分な音声を収集できていない。

【0177】

この際、情報処理サーバ30の制御部330は、各動作体が収集した音声に基づく音声認識処理の結果を統合（集計）する。なお、制御部330は、同一環境中に存在することを動作体から音声データまたは音声認識結果がアップロードされない場合には、当該動作体が音声の収集に失敗したと見做してもよい。

【0178】

また、本実施形態に係る制御部330は、統合の結果得られた統合認識結果を各動作体に伝達し、各動作体の動作を間接的あるいは直接的に制御してよい。

【0179】

例えば、図17Bに示す一例の場合、制御部330は、得られた複数の音声認識処理の結果から最も数が多い語彙“Good morning”を統合認識結果として決定し、第1の動作体10a～10cに伝達している。

【0180】

また、この際、本実施形態に係る制御部330は、統合認識結果と同一の音声認識処理の結果が得られた音声を収集した第1の動作体10bや第1の動作体10cに、上記同一

10

20

30

40

50

の音声認識処理の結果が得られなかった第1の動作体10aや第2の動作体20に関する言動を実行させてもよい。

【0181】

図17Bに示す一例の場合、制御部330は、第1の動作体10bに、第1の動作体10aに対し、正しいユーザの意図を伝えるシステム発話S03を行わせている。また、制御部330は、第1の動作体10cに、第2の動作体20が発話U04を理解できなかった旨のシステム発話S05を行わせている。

【0182】

一方、本実施形態に係る制御部は、統合認識結果と同一の音声認識処理の結果が得られなかった音声を収集した第1の動作体10aには、上記同一の音声認識処理の結果が得られなかったことを示す言動を実行させてもよい。

【0183】

図17Bに示す一例の場合、制御部330は、第1の動作体10aに、誤った語彙を認識した旨のシステム発話S04を行わせている。

【0184】

上記のような制御によれば、複数の動作体がそれぞれ収集した音声に基づく音声認識処理の結果に基づいて、精度高くユーザの発話を認識することができると共に、動作体同士が認識結果について協議を行っているような様子を表現することが可能となり、ユーザの興味をさらに引き付ける効果が期待される。

【0185】

続いて、本実施形態に係る音声認識処理結果の統合と動作制御の流れについて説明する。図18は、本実施形態に係る情報処理サーバ30による音声認識結果の統合の流れを示すフローチャートである。

【0186】

図18を参照すると、まず、通信部350が複数の動作体から音声データまたは第2の音声認識処理の結果を受信する(S2001)。

【0187】

次に、制御部330は、ステップS2001において受信した音声データに基づく第1の音声認識処理の結果やステップS2001で受信した第2の音声認識処理の結果を統合する(S2002)。

【0188】

次に、制御部330は、認識結果数や音声収集時の動作体の状態に基づいて、統合認識結果を決定する(S2003)。制御部330は、例えば、認識結果の数のみではなく、音声収集時における動作体とユーザの距離や、動作体の動作状況に基づいて重み付けを行い、統合認識結果を決定してもよい。

【0189】

次に、制御部330は、通信部350を介して、ステップS2003において決定した統合認識結果を各動作体に伝達する(S2004)。

【0190】

また、図19は、本実施形態に係る統合認識結果に基づく動作制御の流れを示すフローチャートである。

【0191】

図19を参照すると、まず、動作体(第1の動作体10または第2の動作体20)の制御部は、自身が収集した音声に対応する認識結果が棄却されたか否か、すなわち、自身が収集した音声に対応する認識結果が統合認識結果と異なっているか否か、を判定する(S2101)。

【0192】

ここで、自身が収集した音声に対応する認識結果が棄却されている場合(S2101: YES)、動作体は、上記棄却に対応する言動、すなわち統合認識結果と同一の音声認識処理の結果が得られなかったことを示す言動を実行する(S2102)。

10

20

30

40

50

【0193】

一方、自身が収集した音声に対応する認識結果が棄却されていない場合、すなわち自身が収集した音声に対応する認識結果が統合認識結果と同一であった場合（S2101：NO）、動作体の制御部は、続いて、認識結果が却下された他の動作体が所定の距離（例えば、視認が可能な距離）内に存在するか否かを判定する（S2103）。

【0194】

ここで、認識結果が却下された他の動作体が所定の距離内に存在する場合（S2103：YES）、動作体は、当該他の動作体に関する言動を実行する（S2104）。

【0195】

一方、認識結果が却下された他の動作体が所定の距離内に存在しない場合（S2103：NO）、動作体は、統合認識結果に対応する言動を実行する（S2105）。

10

【0196】

以上、本実施形態に係る情報処理システムが有する機能の一例について説明した。なお、本実施形態に係る情報処理システムは、上記で示した例以外にも複数の動作体の連携に係る機能を有してよい。

【0197】

例えば、本実施形態に係る第1の動作体10は、言語を用いたコミュニケーションを行う機能を有しない第2の動作体20に代わり、第2の動作体20の状態をユーザに伝達することも可能である。

【0198】

20

図20は、本実施形態に係る第1の動作体10による第2の動作体の状態の伝達の流れを示すフローチャートである。

【0199】

図20を参照すると、第1の動作体10は、まず、自律動作を行う（S2201）。

【0200】

ここで、ユーザの発話やユーザが操作するアプリケーションなどからの、第2の動作体の状態伝達に係る要求を検出すると（S2202）、第1の動作体10は、第2の動作体20の状態に係る問い合わせを実行する（S2203）。この際、第1の動作体10は、第2の動作体20に直接状態を問い合わせてもよいし、第2の動作体20が状態を情報処理サーバ30にアップロードしている場合には、情報処理サーバ30に問い合わせを行ってもよい。

30

【0201】

次に、第1の動作体10は、ステップS2203における問い合わせにより取得した第2の動作体20の状態をユーザに伝達する（S2204）。なお、第2の動作体20の状態には、例えば、感情、現在作成されているSLAM地図、認識した物体、バッテリー残量などの動的な状態や、個体識別番号などの静的な情報が含まれてもよい。

【0202】

また、本実施形態に係る第1の動作体10は、第2の動作体20に対する各種の動作指示を行ってもよい。

【0203】

40

上述したように、本実施形態に係る第2の動作体20は、第1の動作体10と比較して広い移動範囲を有する。このため、例えば、卓上から動けない第1の動作体10がユーザを認識できない場合、第1の動作体10は、他の部屋などにユーザを探索しに行くよう指示を行ってもよい。

【0204】

図21は、本実施形態に係る第1の動作体10の指示に基づく第2の動作体20によるユーザ探索の流れを示すフローチャートである。

【0205】

図21を参照すると、第2の動作体20は、まず、自律動作を行う（S2301）。

【0206】

50

ここで、第1の動作体10からのユーザ探索要求を受信した場合（S2302）、第2の動作体20は、家の中を歩き回るなどしてユーザの探索を行う（S2303）。

【0207】

ここで、所定時間内にユーザを認識できた場合（S2304：YES）、第2の動作体20は、ユーザを認識した場所や状況を第1の動作体10に伝達する（S2305）。この場合、第1の動作体10は、伝達された情報に基づく言動を行ってもよい。例えば、ユーザがキッチンにいることを伝達された場合、第1の動作体10は、「Mikeに教えてもらったよ、ご飯作ってるの？」などの発話を行ってもよい。

【0208】

一方、所定時間内にユーザを認識できなかった場合（S2304：NO）、第2の動作体20は、ユーザを見つけられなかったことを第1の動作体10に伝達する（S2306）。

【0209】

なお、ステップS2305やS2306における情報の伝達時に、周囲に探索対象であるユーザ以外の人物がいる場合には、第2の動作体20は、鳴き声や振る舞いなどにより探索の結果を表現してもよい。例えば、探索対象のユーザをキッチンで認識した場合には、第2の動作体20は、しっぽを振りながら物を食べるような振る舞いを行ってもよいし、ユーザを見つけることができなかった場合には、首を横に振りながら悲しそうに鳴くなどの動作を行ってもよい。

【0210】

また、例えば、本実施形態に係る第1の動作体10は、SNSやメッセージアプリケーションなどの外部サービスから取得した情報に基づいて、第2の動作体20に指示を行うことも可能である。

【0211】

図22は、本実施形態に係る外部サービスから取得した情報に基づく第2の動作体20への指示の流れの一例を示すフローチャートである。

【0212】

図22を参照すると、第1の動作体10は、まず、自律動作を行う（S2401）。

【0213】

ここで、メッセージアプリケーションやSNSからユーザの帰宅を示す情報を抽出すると（S2402）、第1の動作体10は、第2の動作体20に対し、玄関で待機するように指示を行ってよい（S2403）。この場合、第2の動作体20は、第1の動作体10からの指示に基づいて、玄関で待機する。また、第1の動作体10は、帰宅したユーザを認識した際、「Mike、玄関でちゃんと待ってた？」などの発話を行ってもよい。

【0214】

このために、本実施形態に係る第1の動作体10は、メッセージアプリケーションやSNSを介してユーザとのコミュニケーションを図る機能を有してもよい。係る機能によれば、外出中であっても、ユーザが第1の動作体10との対話を楽しむことができ、また第2の動作体20や家の状況を把握することができる。

【0215】

また、例えば、本実施形態に係る第1の動作体10は、高い視点を活かして第2の動作体20に障害物回避などに係る移動指示を行うことも可能である。上述したように、本実施形態に係る第1の動作体10は、卓上に配置されることを想定していることから、床上を歩行移動する第2の動作体20と比較して高い視点を有する。

【0216】

図23Aは、本実施形態に係る第1の動作体10による第2の動作体20への移動指示の流れを示すフローチャートである。

【0217】

図23Aを参照すると、まず、第1の動作体10は、自律動作を行う（S2401）。

【0218】

10

20

30

40

50

次に、第１の動作体１０は、第２の動作体２０の周辺の映像を第２の動作体に送信する（Ｓ２４０２）。第１の動作体１０は、第２の動作体２０からの要求に基づいて上記映像の送信を行ってもよいし、第２の動作体２０が障害物に衝突しそうな状況を認識したことなどに基づいて上記映像の送信を行ってもよい。

【０２１９】

また、第１の動作体１０は、第２の動作体２０に対する移動指示に係る表出を行ってよい（Ｓ２４０３）。第１の動作体１０は、例えば、「危ないよ！みぎ、みぎ！」などの発話を行うことができる。

【０２２０】

また、図２３Ｂは、本実施形態に係る第１の動作体１０からの移動指示に基づく第２の動作体２０の動作の流れを示すフローチャートである。

【０２２１】

図２３Ｂを参照すると、まず、第２の動作体２０は、自律動作を行う（Ｓ２５０１）。

【０２２２】

ここで、第１の動作体１０から映像を受信すると、第２の動作体２０は、第１の動作体１０の視点で得られた映像を自身の周囲にマッピングする（Ｓ２５０２）。

【０２２３】

続いて、第２の動作体２０は、ステップＳ２５０２における映像のマッピングに基づいて、障害物を回避して移動する（Ｓ２５０３）。

【０２２４】

<２．情報処理サーバ３０のハードウェア構成例>

次に、本開示の一実施形態に係る情報処理サーバ３０のハードウェア構成例について説明する。図２４は、本開示の一実施形態に係る情報処理サーバ３０のハードウェア構成例を示すブロック図である。図２４に示すように、情報処理サーバ３０は、例えば、プロセッサ８７１と、ＲＯＭ８７２と、ＲＡＭ８７３と、ホストバス８７４と、ブリッジ８７５と、外部バス８７６と、インタフェース８７７と、入力装置８７８と、出力装置８７９と、ストレージ８８０と、ドライブ８８１と、接続ポート８８２と、通信装置８８３と、を有する。なお、ここで示すハードウェア構成は一例であり、構成要素の一部が省略されてもよい。また、ここで示される構成要素以外の構成要素をさらに含んでもよい。

【０２２５】

（プロセッサ８７１）

プロセッサ８７１は、例えば、演算処理装置又は制御装置として機能し、ＲＯＭ８７２、ＲＡＭ８７３、ストレージ８８０、又はリムーバブル記録媒体９０１に記録された各種プログラムに基づいて各構成要素の動作全般又はその一部を制御する。

【０２２６】

（ＲＯＭ８７２、ＲＡＭ８７３）

ＲＯＭ８７２は、プロセッサ８７１に読み込まれるプログラムや演算に用いるデータ等を格納する手段である。ＲＡＭ８７３には、例えば、プロセッサ８７１に読み込まれるプログラムや、そのプログラムを実行する際に適宜変化する各種パラメータ等が一時的又は永続的に格納される。

【０２２７】

（ホストバス８７４、ブリッジ８７５、外部バス８７６、インタフェース８７７）

プロセッサ８７１、ＲＯＭ８７２、ＲＡＭ８７３は、例えば、高速なデータ伝送が可能なホストバス８７４を介して相互に接続される。一方、ホストバス８７４は、例えば、ブリッジ８７５を介して比較的データ伝送速度が低速な外部バス８７６に接続される。また、外部バス８７６は、インタフェース８７７を介して種々の構成要素と接続される。

【０２２８】

（入力装置８７８）

入力装置８７８には、例えば、マウス、キーボード、タッチパネル、ボタン、スイッチ、及びレバー等が用いられる。さらに、入力装置８７８としては、赤外線やその他の電波

10

20

30

40

50

を利用して制御信号を送信することが可能なリモートコントローラ（以下、リモコン）が用いられることもある。また、入力装置 878 には、マイクロフォンなどの音声入力装置が含まれる。

【0229】

（出力装置 879）

出力装置 879 は、例えば、CRT（Cathode Ray Tube）、LCD、又は有機 EL 等のディスプレイ装置、スピーカ、ヘッドホン等のオーディオ出力装置、プリンタ、携帯電話、又はファクシミリ等、取得した情報を利用者に対して視覚的又は聴覚的に通知することが可能な装置である。また、本開示に係る出力装置 879 は、触覚刺激を出力することが可能な種々の振動デバイスを含む。

10

【0230】

（ストレージ 880）

ストレージ 880 は、各種のデータを格納するための装置である。ストレージ 880 としては、例えば、ハードディスクドライブ（HDD）等の磁気記憶デバイス、半導体記憶デバイス、光記憶デバイス、又は光磁気記憶デバイス等が用いられる。

【0231】

（ドライブ 881）

ドライブ 881 は、例えば、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリ等のリムーバブル記録媒体 901 に記録された情報を読み出し、又はリムーバブル記録媒体 901 に情報を書き込む装置である。

20

【0232】

（リムーバブル記録媒体 901）

リムーバブル記録媒体 901 は、例えば、DVD メディア、Blu-ray（登録商標）メディア、HD DVD メディア、各種の半導体記憶メディア等である。もちろん、リムーバブル記録媒体 901 は、例えば、非接触型 IC チップを搭載した IC カード、又は電子機器等であってもよい。

【0233】

（接続ポート 882）

接続ポート 882 は、例えば、USB（Universal Serial Bus）ポート、IEEE1394 ポート、SCSI（Small Computer System Interface）、RS-232C ポート、又は光オーディオ端子等のような外部接続機器 902 を接続するためのポートである。

30

【0234】

（外部接続機器 902）

外部接続機器 902 は、例えば、プリンタ、携帯音楽プレーヤ、デジタルカメラ、デジタルビデオカメラ、又は IC レコーダ等である。

【0235】

（通信装置 883）

通信装置 883 は、ネットワークに接続するための通信デバイスであり、例えば、有線又は無線 LAN、Bluetooth（登録商標）、又は WUSB（Wireless USB）用の通信カード、光通信のルータ、ADSL（Asymmetric Digital Subscriber Line）用のルータ、又は各種通信用のモデム等である。

40

【0236】

<3. まとめ>

以上説明したように、本開示の一実施形態に係る情報処理サーバ 30 は、音声認識処理の結果に基づいて動作体の動作を制御する制御部 330 を備える。また、本開示の一実施形態に係る制御部 330 は、一の動作体が収集した音声に基づく音声認識処理の結果、または一の動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の動作体の動作を制御すること、を特徴の一つとする。係る構成によれば、音声認識処理

50

に基づく動作を行う複数の動作体が存在する環境において、各動作体がより状況に適した動作を実現することを可能とする。

【0237】

以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

【0238】

また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

【0239】

また、コンピュータに内蔵されるCPU、ROMおよびRAMなどのハードウェアに、第1の動作体10、第2の動作体20、または情報処理サーバ30が有する構成と同等の機能を発揮させるためのプログラムも作成可能であり、当該プログラムを記録した、コンピュータに読み取り可能な非一過性の記録媒体も提供され得る。

【0240】

また、本明細書の情報処理システムの処理に係る各ステップは、必ずしもフローチャートに記載された順序に沿って時系列に処理される必要はない。例えば、情報処理システムの処理に係る各ステップは、フローチャートに記載された順序と異なる順序で処理されても、並列的に処理されてもよい。

【0241】

なお、以下のような構成も本開示の技術的範囲に属する。

(1)

音声認識処理の結果に基づいて動作体の動作を制御する制御部、
を備え、

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御する、
情報処理装置。

(2)

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果に基づいて、当該音声認識処理の結果に対応する動作を他の前記動作体を実行させる、
前記(1)に記載の情報処理装置。

(3)

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果を他の前記動作体に伝達したことを示す動作を当該一の前記動作体を実行させる、
前記(2)に記載の情報処理装置。

(4)

前記動作体は、第1の辞書を用いた第1の音声認識処理の結果に基づく動作を行う第1の動作体と、前記第1の辞書よりも語彙数の少ない第2の辞書を用いた第2の音声認識処理の結果に基づく動作を行う第2の動作体を含み、

前記制御部は、前記第1の動作体が収集した音声と前記第1の辞書に基づく前記第1の音声認識処理の結果に基づいて、前記第1の音声認識処理の結果に対応する動作を前記第2の動作体を実行させる、
前記(2)または(3)に記載の情報処理装置。

(5)

前記制御部は、前記第1の音声認識処理の結果として得られた語彙のカテゴリに対応する語彙を前記第2の辞書から抽出し、当該語彙を前記第2の動作体伝達する、

10

20

30

40

50

前記（４）に記載の情報処理装置。

（６）

前記制御部は、前記第１の音声認識処理の結果に係るログに基づいて、前記第２の辞書の語彙の入れ替えを制御する、

前記（４）または（５）に記載の情報処理装置。

（７）

前記第２の動作体は、前記第２の辞書を用いた前記第２の音声認識処理をローカルで実行する、

前記（４）～（６）のいずれかに記載の情報処理装置。

（８）

前記動作体は、自律移動体である、

前記（１）～（７）のいずれかに記載の情報処理装置。

（９）

前記制御部は、一の前記動作体に係る音声認識環境が音声認識処理精度の低下要因となり得る環境である場合、他の前記動作体に係る音声認識環境が向上するよう他の前記動作体の動作を制御する、

前記（８）に記載の情報処理装置。

（１０）

前記音声認識環境は、前記動作体の動作状況を含み、

前記制御部は、一の前記動作体の動作状況に基づいて、他の前記動作体の動作を制御する、

前記（９）に記載の情報処理装置。

（１１）

前記制御部は、一の前記動作体が動作音を伴う動作を実行している場合、他の前記動作体の動作音がより低下するよう制御する、

前記（１０）に記載の情報処理装置。

（１２）

前記音声認識環境は、ノイズ音源または話者の認識結果を含み、

前記制御部は、一の前記動作体が収集したセンサ情報に基づいて認識されたノイズ音源または話者に係る情報を他の前記動作体に伝達する、

前記（１）～（１１）のいずれかに記載の情報処理装置。

（１３）

前記制御部は、一の前記動作体が収集したセンサ情報に基づいて認識されたノイズ音源または話者に係る情報に基づいて、他の前記動作体に係る音声認識処理の実行有無を制御する、

前記（１２）に記載の情報処理装置。

（１４）

前記制御部は、一の前記動作体が収集したセンサ情報に基づいて認識された環境に基づいて、ユーザに対する能動的なアクションを他の前記動作体に行わせる、

前記（１）～（１３）のいずれかに記載の情報処理装置。

（１５）

前記制御部は、一の前記動作体が収集したセンサ情報に基づいて認識された環境に基づいて、当該環境に応じた能動的な発話を他の前記動作体に行わせる、

前記（１４）に記載の情報処理装置。

（１６）

前記制御部は、複数の前記動作体が収集した音声のそれぞれに基づく複数の音声認識処理の結果から決定される統合認識結果に基づいて、前記動作体の動作を制御する、

前記（１）～（１５）のいずれかに記載の情報処理装置。

（１７）

前記制御部は、前記統合認識結果と同一の音声認識処理の結果が得られた音声を収集し

10

20

30

40

50

た前記動作体に、前記統合認識結果と同一の音声認識処理の結果が得られなかった他の前記動作体に関する言動を実行させる、
前記（１６）に記載の情報処理装置。

（１８）

前記制御部は、前記統合認識結果と同一の音声認識処理の結果が得られなかった音声を収集した前記動作体に、前記統合認識結果と同一の音声認識処理の結果が得られなかったことを示す言動を実行させる、

前記（１６）または（１７）に記載の情報処理装置。

（１９）

プロセッサが、音声認識処理の結果に基づいて動作体の動作を制御すること、
を含み、

前記制御することは、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御すること、

をさらに含む、

情報処理方法。

（２０）

コンピュータを、

音声認識処理の結果に基づいて動作体の動作を制御する制御部、

を備え、

前記制御部は、一の前記動作体が収集した音声に基づく音声認識処理の結果、または一の前記動作体が収集したセンサ情報に基づいて認識された音声認識環境に基づいて、他の前記動作体の動作を制御する、

情報処理装置、

として機能させるためのプログラム。

【符号の説明】

【０２４２】

１０ 第１の動作体
１４０ トリガ検出部
１５０ 制御部
２０ 第２の動作体
２４０ 認識部
２４２ 第２の辞書
２５０ 制御部
３０ 情報処理サーバ
３１０ 音声認識部
３２０ 自然言語処理部
３２２ 第１の辞書
３３０ 制御部
４０ ネットワーク

10

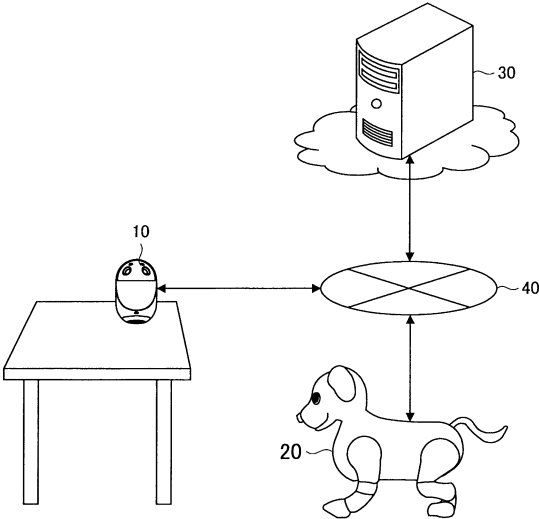
20

30

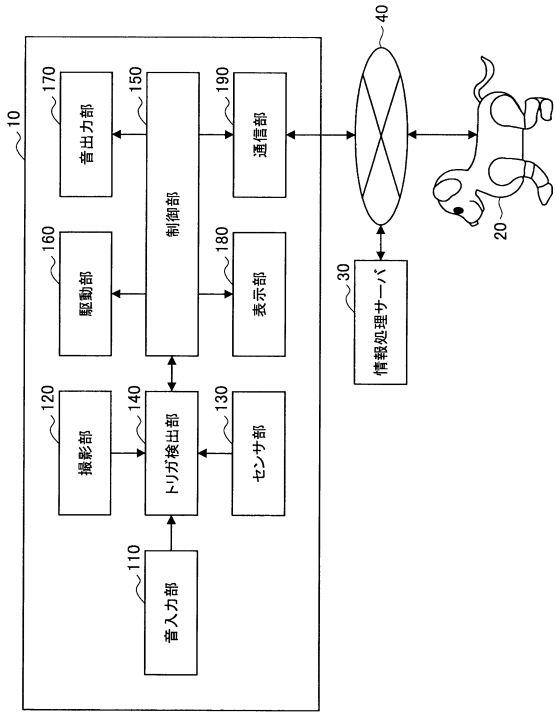
40

50

【図面】
【図 1】



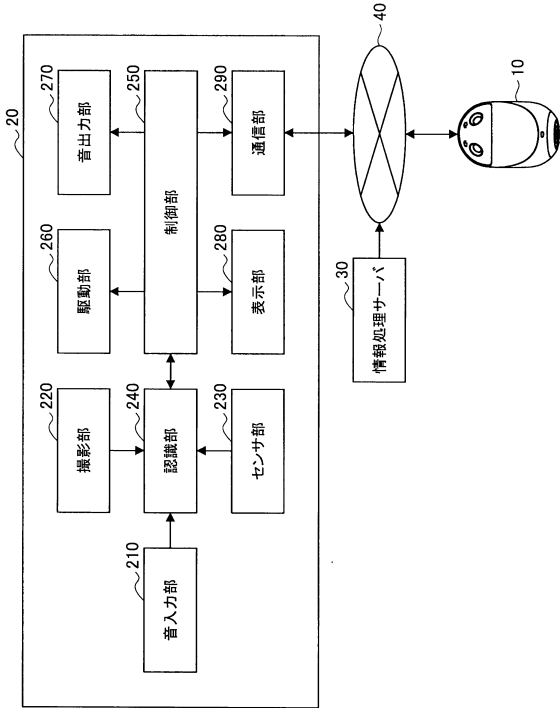
【図 2】



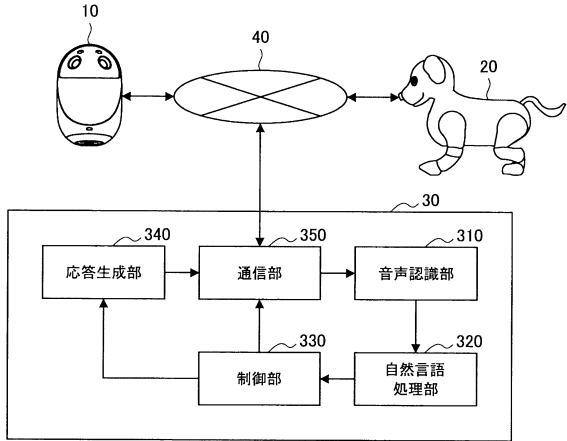
10

20

【図 3】



【図 4】

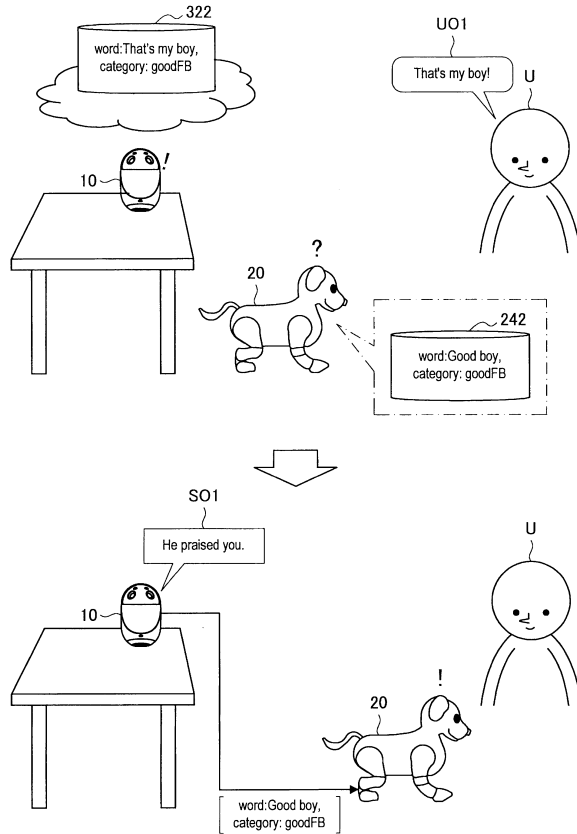


30

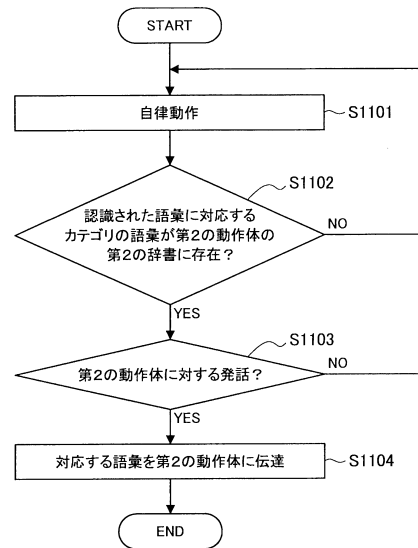
40

50

【図 5】



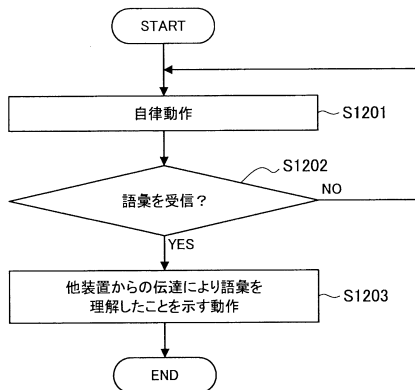
【図 6 A】



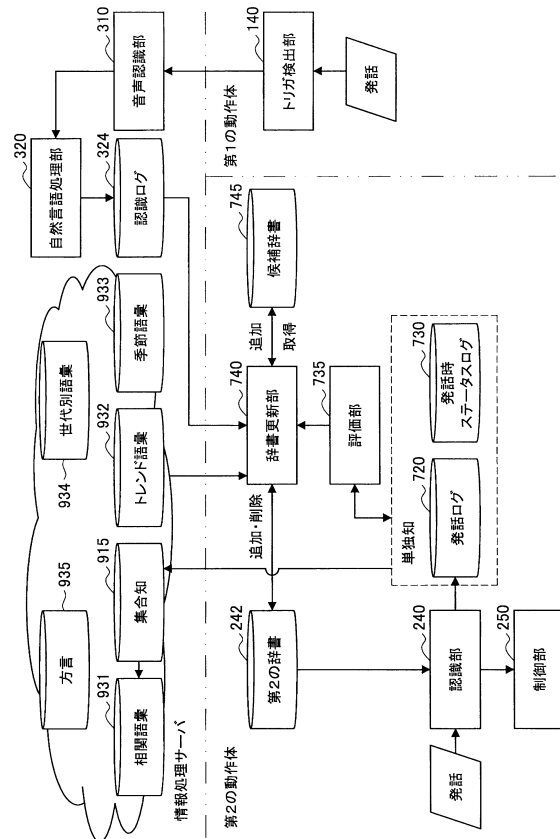
10

20

【図 6 B】



【図 7】

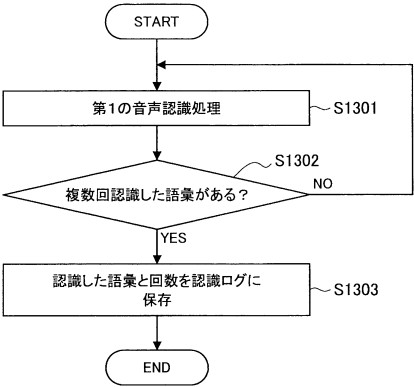


30

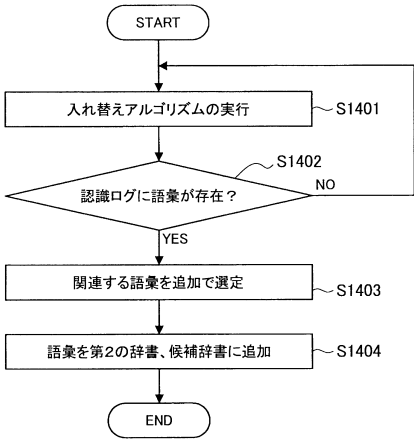
40

50

【図 8】



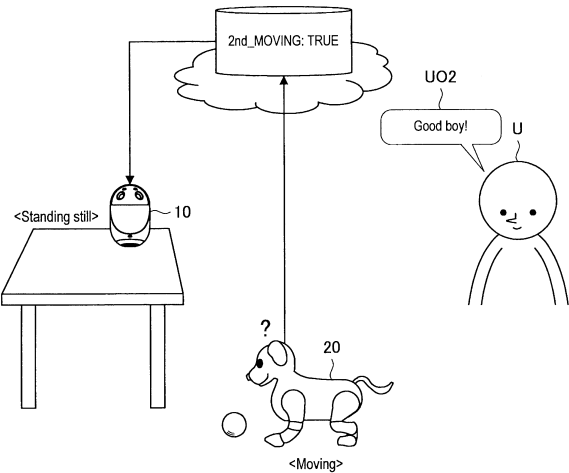
【図 9】



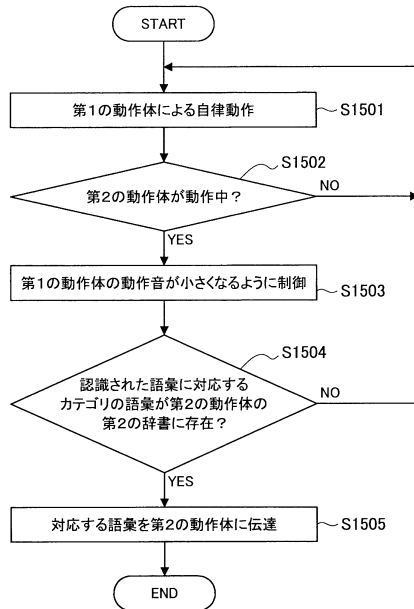
10

20

【図 10】



【図 11】

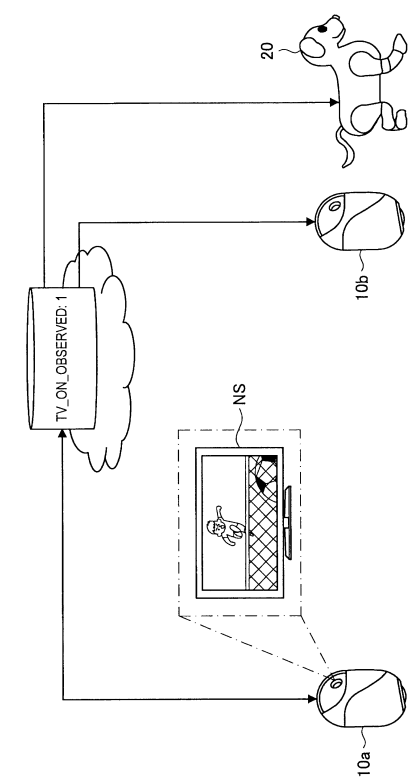


30

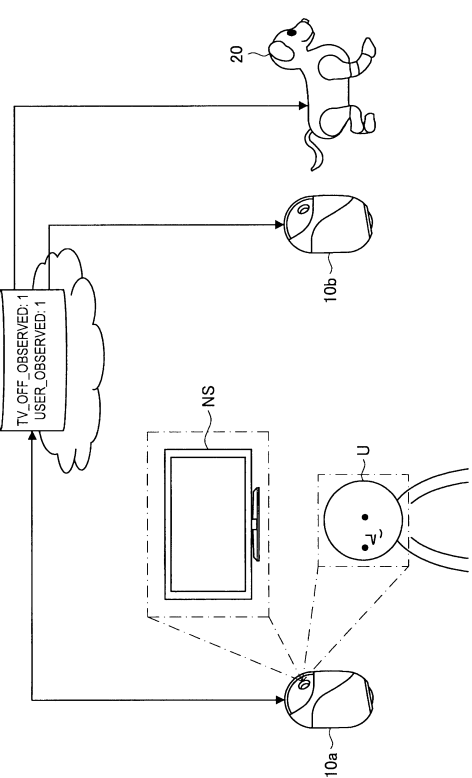
40

50

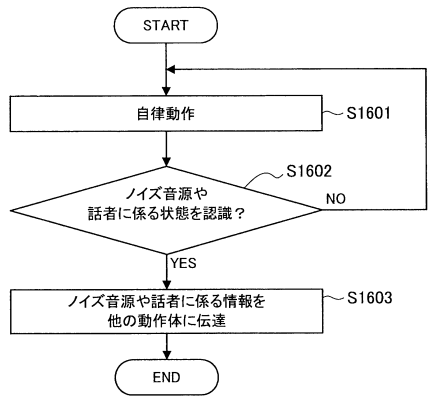
【図 1 2 A】



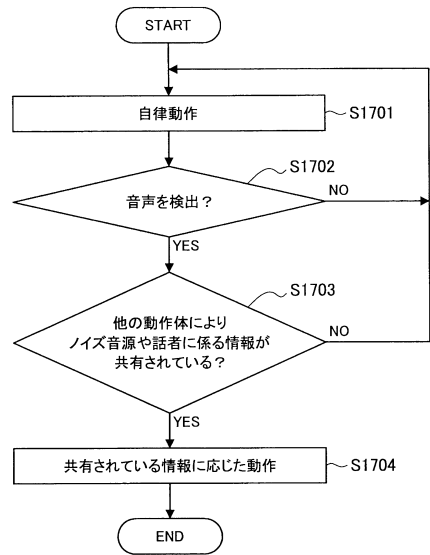
【図 1 2 B】



【図 1 3 A】



【図 1 3 B】



10

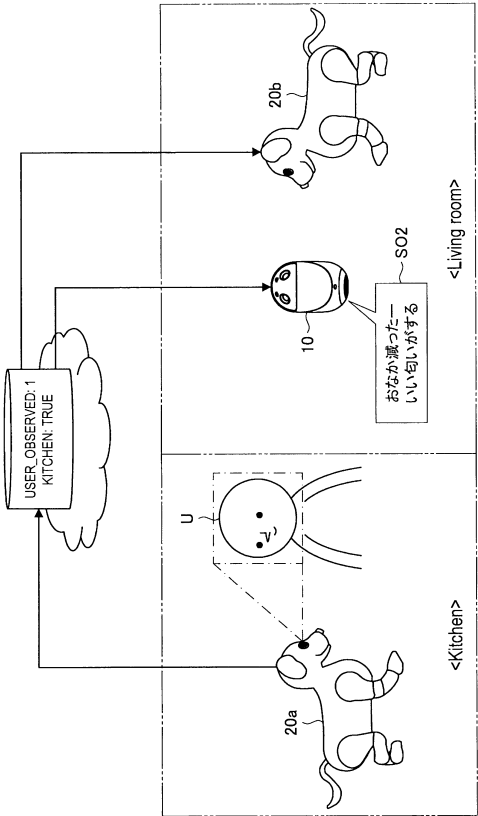
20

30

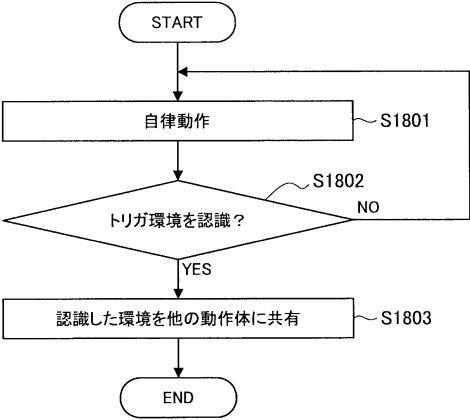
40

50

【図 1 4】



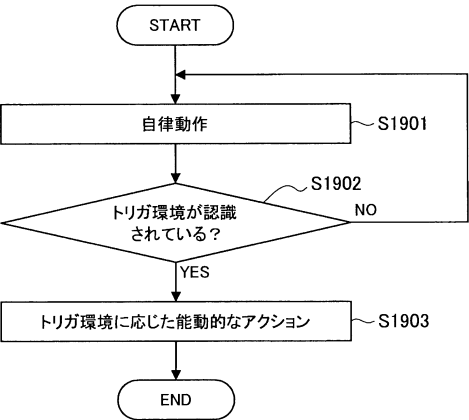
【図 1 5 A】



10

20

【図 1 5 B】



【図 1 6】

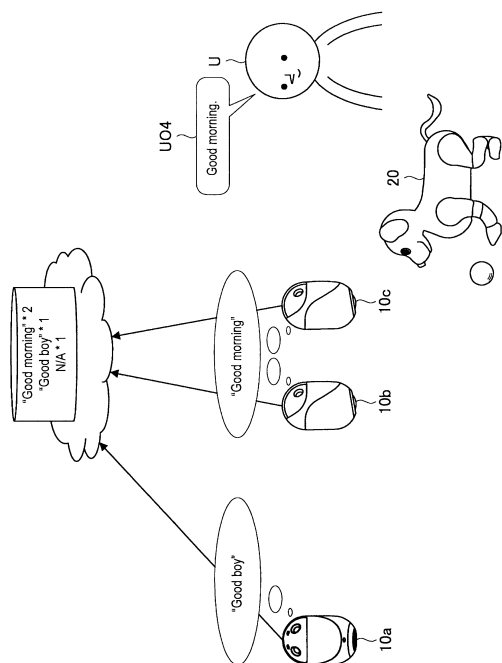
	シナリオ対話	状況対話	知識対話	雑談対話
人の会話を漏れ聞く			○	○
新たな情報を入手する			○	
知能のある環境を認識する		○	○	
その場にいる人に質問する	○			
特定の物体を見つける		○		
人がいなくなる		○		
人を見つけて話しかける	○	○		

30

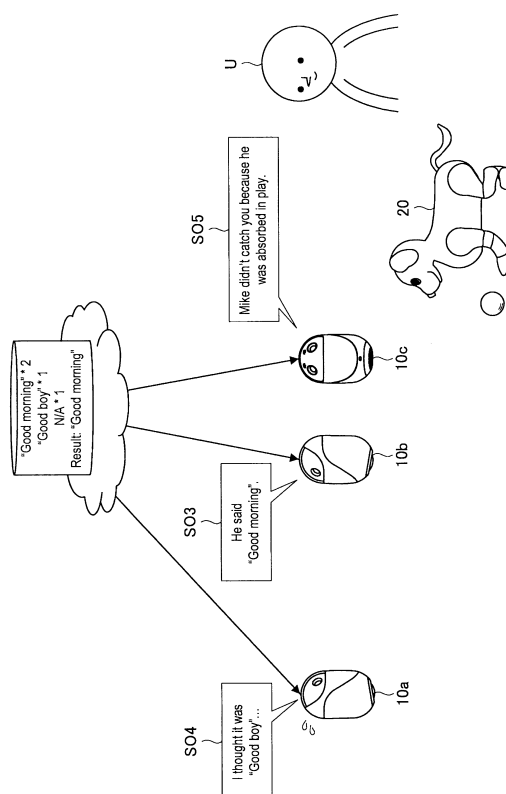
40

50

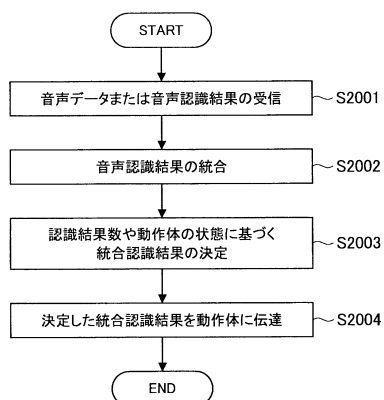
【図 17 A】



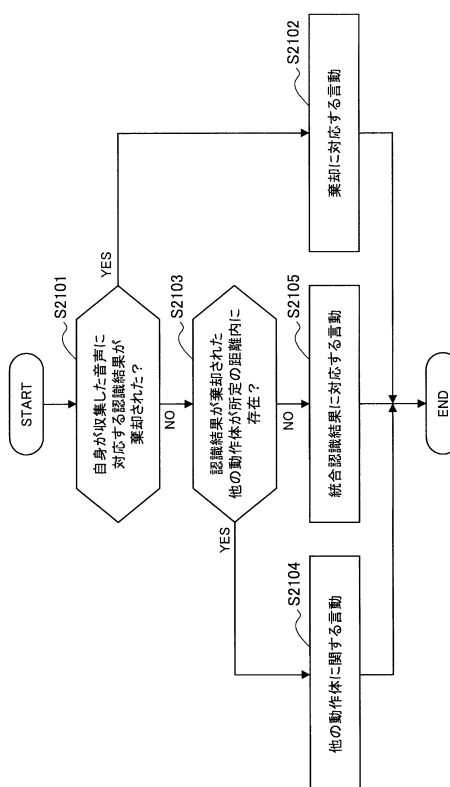
【図 17 B】



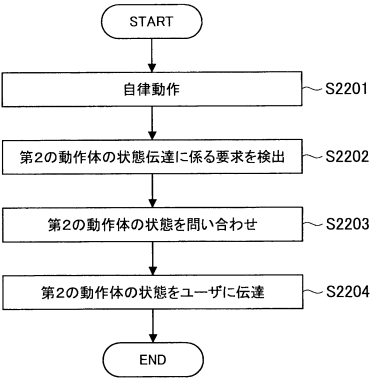
【図 18】



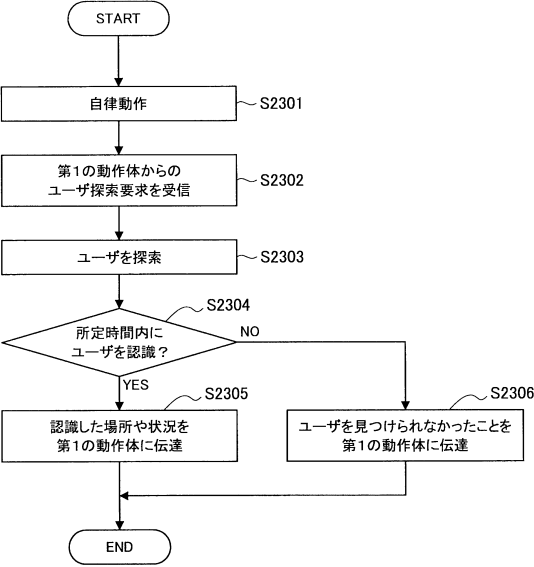
【図 19】



【図 2 0】



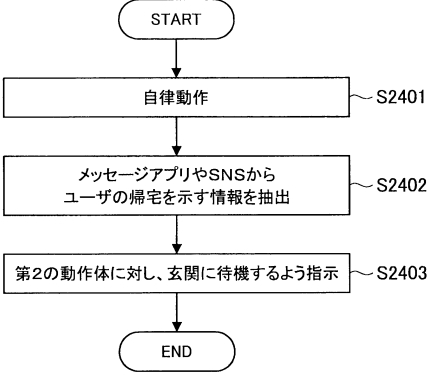
【図 2 1】



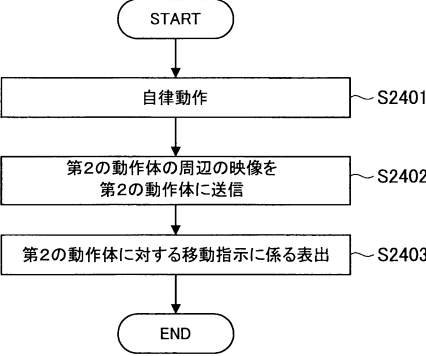
10

20

【図 2 2】



【図 2 3 A】

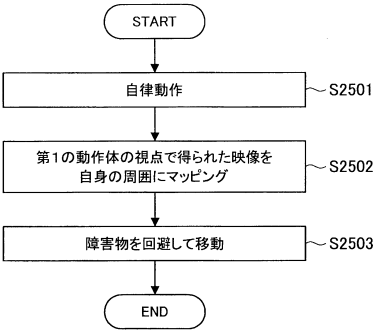


30

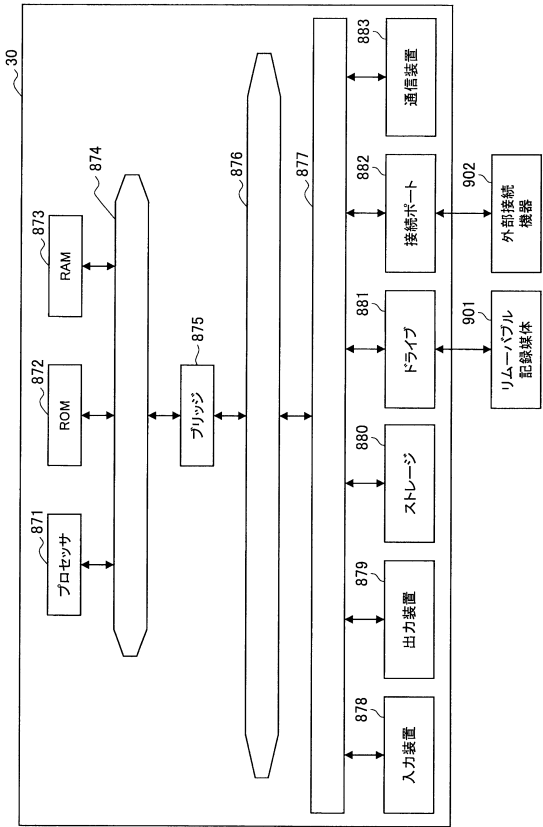
40

50

【図 2 3 B】



【図 2 4】



10

20

30

40

50

フロントページの続き

(56)参考文献 特開２０１８－０８１２３３（ＪＰ，Ａ）
 中国特許第１０７０７７８４４（ＣＮ，Ｂ）
 特開２００２－３３７０７９（ＪＰ，Ａ）
 特開２０１６－１３０８００（ＪＰ，Ａ）
 特開２０１７－１３８４７６（ＪＰ，Ａ）
(58)調査した分野 (Int.Cl., DB名)
 G 1 0 L 1 5／0 0－1 7／2 6
 B 2 5 J 1 3／0 0