(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2019年8月1日(01.08.2019)

WIPO | PCT

(10) 国際公開番号
WO 2019/146199 A1

(51) 国際特許分類:
G06F 3/16 (2006.01)　*G10L 13/08* (2013.01)
G06F 3/01 (2006.01)　*G10L 15/22* (2006.01)
G10L 13/00 (2006.01)

(21) 国際出願番号：　PCT/JP2018/040151

(22) 国際出願日：　2018年10月29日(29.10.2018)

(25) 国際出願の言語：　日本語

(26) 国際公開の言語：　日本語

(30) 優先権データ：
特願 2018-009150　2018年1月23日(23.01.2018)　JP

(71) 出願人：ソニー株式会社 (**SONY CORPORATION**) [JP/JP]; 〒1080075 東京都港区港南１丁目７番１号 Tokyo (JP).

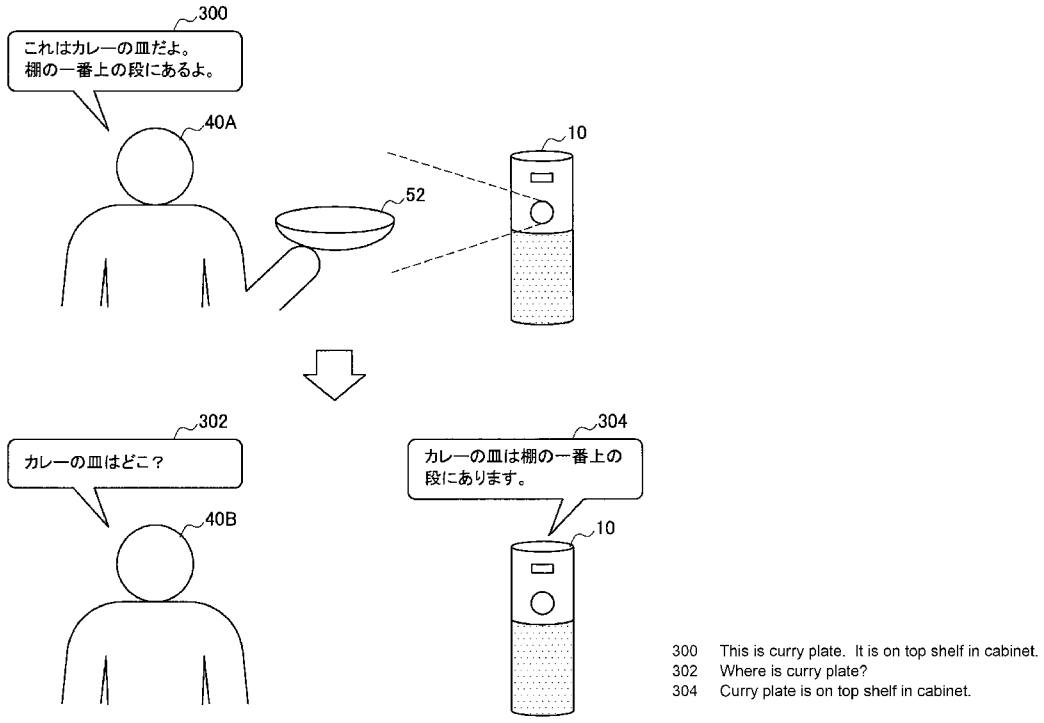
(72) 発明者：斎藤 真里 (**SAITO, Mari**); 〒1080075 東京都港区港南１丁目７番１号 ソニー株式会社内 Tokyo (JP).

(74) 代理人：特許業務法人酒井国際特許事務所 (**SAKAI INTERNATIONAL PATENT OFFICE**); 〒1000013 東京都千代田区霞が関３丁目８番１号 虎の門三井ビルディング Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能)：AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,

(54) **Title:** INFORMATION PROCESSING DEVICE AND INFORMATION PROCESSING METHOD

(54) 発明の名称： 情報処理装置、及び情報処理方法



(57) **Abstract:** [Problem] To more smoothly communicate with a user. [Solution] Provided is an information processing device comprising an output control unit that controls presentation of information to a user, wherein, when a specific expression to a group including the user is applicable to the contents of the information to be presented, the output control unit causes presentation of the information including the specific expression to be executed by using a voice and/or an image. In addition, provided is an information processing device comprising a learning unit that learns an object to be

WO 2019/146199 A1

[続葉有]

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

recognized and a language expression concerning the object to be recognized, in association with each other, wherein the language expression includes a specific expression to a group including a user, and the learning unit learns the specific expression on the basis of a collected voice and/or image.

(57) 要約: 【課題】ユーザとのコミュニケーションをより円滑に行うこと。 【解決手段】ユーザに対する情報提示を制御する出力制御部、を備え、前記出力制御部は、提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち少なくとも一方を用いて実行させる、情報処理装置が提供される。また、認識対象と前記認識対象に係る言語表現とを対応付けて学習する学習部、を備え、前記言語表現は、ユーザを含む集団内における特有表現を含み、前記学習部は、収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習する、情報処理装置が提供される。

明　細　書

発明の名称 ： 情報処理装置、及び情報処理方法

技術分野

[0001] 本開示は、情報処理装置、及び情報処理方法に関する。

背景技術

[0002] 近年、ユーザに対し音声や視覚情報を用いて情報提示を行う種々の装置が普及している。また、上記のような情報提示に係るユーザの利便性を向上させる技術も多く開発されている。

[0003] 上述の技術に関連し、例えば、下記非特許文献１には、一般物体認識を用いて画像に含まれる物体を認識し、当該物体の一般的な名称（以下では、一般表現とも称される）を取得し、物体を一般表現で表現することに関する技術が開示されている。上述の一般物体認識の機能を搭載した端末は、当該一般物体認識により物体を認識し、認識した物体を一般表現で表現することができる。

先行技術文献

非特許文献

[0004] 非特許文献1：アレックス　クリシュフスキー（Ａｌｅｘ　Ｋｒｉｚｈｅｖｓｋｙ）、外２名、“深層畳み込みニューラルネットワークを用いたＩｍａｇｅＮｅｔ分類”（“ＩｍａｇｅＮｅｔ　Ｃｌａｓｓｉｆｉｃａｔｉｏｎ　ｗｉｔｈ　Ｄｅｅｐ　Ｃｏｎｖｏｌｕｔｉｏｎａｌ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ”）、［ｏｎｌｉｎｅ］、平成２４年１２月４日、ニューラル情報処理システム（Ｎｅｕｒａｌ　Ｉｎｆｏｒｍａｔｉｏｎ　Ｐｒｏｃｅｓｓｉｎｇ　Ｓｙｓｔｅｍｓ）、［平成３０年１月１９日検索］、インターネット（ＵＲＬ：https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf）

発明の概要

発明が解決しようとする課題

[0005] しかしながら、同じ物体に対して、個人または集団（例えば家族）によっては一般表現とは異なる表現（以下では、特有表現とも称される）がされることがある。上記非特許文献１に記載されるような技術では、一般物体認識により物体の一般表現を取得、利用、及び認識することはできるが、物体の特有表現を取得、利用、及び認識することはできない。

[0006] そこで、本開示では、ユーザとのコミュニケーションをより円滑に行うことが可能な、新規かつ改良された情報処理装置、及び情報処理方法を提案する。

課題を解決するための手段

[0007] 本開示によれば、ユーザに対する情報提示を制御する出力制御部、を備え、前記出力制御部は、提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち少なくとも一方を用いて実行させる、情報処理装置が提供される。

[0008] また、本開示によれば、認識対象と前記認識対象に係る言語表現とを対応付けて学習する学習部、を備え、前記言語表現は、ユーザを含む集団内における特有表現を含み、前記学習部は、収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習する、情報処理装置が提供される。

[0009] また、本開示によれば、ユーザに対する情報提示を制御することと、提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち少なくとも一方を用いて実行させることと、を含む、プロセッサにより実行される、情報処理方法が提供される。

[0010] また、本開示によれば、認識対象と前記認識対象に係る言語表現とを対応付けて学習することと、前記言語表現は、ユーザを含む集団内における特有表現を含み、収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習することと、を含む、プロセッサにより実行される、情

報処理方法が提供される。

発明の効果

[0011] 以上説明したように本開示によれば、ユーザとのコミュニケーションをより円滑に行うことが可能となる。

[0012] なお、上記の効果は必ずしも限定的なものではなく、上記の効果とともに、または上記の効果に代えて、本明細書に示されたいずれかの効果、または本明細書から把握され得る他の効果が奏されてもよい。

図面の簡単な説明

[0013] [図1]本開示の一実施形態の概要を示す説明図である。

[図2]同実施形態に係る情報処理システムの構成例を示すブロック図である。

[図3]同実施形態に係る情報処理端末の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の構成例を示すブロック図である。

[図5]同実施形態に係る特有表現ＤＢの構成例を示す説明図である。

[図6]同実施形態に係る集団に属さないユーザへの情報提示の例を示す説明図である。

[図7]同実施形態に係る特有表現を学習する際の情報処理サーバの動作例を示すフローチャートである。

[図8]同実施形態に係る特有表現を適用する際の情報処理サーバの動作例を示すフローチャートである。

[図9]同実施形態に係るユーザが提示する音声と画像に基づく特有表現の適用例を示す説明図である。

[図10]同実施形態に係るユーザが提示する音声と画像に基づく特有表現の学習例及び適用例を示す説明図である。

[図11]同実施形態に係る画像に基づく特有表現の学習例を示す説明図である。

[図12]同実施形態に係る情報処理端末が自発的に取得した音声と画像に基づく特有表現の学習例を示す説明図である。

[図13]同実施形態に係るユーザが特有表現を理解していないと推定される場

合の情報提示の例を示す説明図である。

[図14]本開示の一実施形態に係るハードウェア構成例を示す説明図である。

発明を実施するための形態

[0014] 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。

[0015] なお、説明は以下の順序で行うものとする。

１．本開示の実施形態

１．１．概要

１．２．システム構成例

１．３．情報処理端末１０の機能構成例

１．４．情報処理サーバ２０の機能構成例

１．５．動作例

２．変形例

３．ハードウェア構成例

４．まとめ

[0016] ＜１．本開示の実施形態＞

＜＜１．１．概要＞＞

近年、機械学習などの手法を用いてユーザに情報提示を行う種々の装置が普及している。上記のような装置には、例えば、音声発話や視覚情報を用いてユーザに情報を提示するエージェント装置が挙げられる。エージェント装置は、例えば、ユーザの問い合わせに対する応答情報などを、音声発話の出力や視覚情報の表示などにより行うことができる。

[0017] この際、エージェント装置が出力する応答情報は、エージェント装置が学習した情報に基づき提示され得る。例えば、エージェント装置は、一般物体認識により物体等の認識対象を認識し、認識対象の一般的な名称（以下では、一般表現とも称される）を取得することで学習する。そして、エージェン

ト装置は、学習した一般表現を用いた音声発話や視覚情報の表示等を行うことができる。

[0018] しかしながら、同じ物体に対して、個人または集団（例えば家族）によっては一般表現とは異なる表現（以下では、特有表現とも称される）がされることがある。しかし、上述のエージェント装置は、一般表現を学習することしか想定されていない。したがって、ユーザからの入力情報に特有表現が含まれていた場合、エージェント装置は、特有表現を認識することができず、エージェント装置がユーザに対して適切な情報を提示できないことが想定される。

[0019] 本開示の一実施形態に係る情報処理装置、及び情報処理方法は、上記の点に着目して発想されたものであり、認識対象と当該認識対象に対する特有表現とを対応付けて学習することで、ユーザとのコミュニケーションをより円滑に行うことを可能とする。このために、本実施形態に係る情報処理方法を実現する情報処理装置は、入力された情報に基づき、特有表現を用いた応答情報の出力を制御することを特徴の一つとする。

[0020] ここで、特有表現とは、一般的な表現とは異なる表現のことである。例えば、食事に利用する皿の一般的な表現は「皿」である。当該皿をカレーの盛り付けに使用する場合、ある家庭では当該皿のことを「カレーの皿」と呼ぶ場合がある。この「カレーの皿」という表現が特有表現に該当する。

[0021] なお、特有表現は、上述した皿等の物体に対する呼称に限定されない。例えば、特有表現は、物体等の呼称以外に、物体等の性質、状態等を示す形容表現であってもよい。また、特有表現は、物体等の動作を示す動作表現であってもよい。また、特有表現は、方言であってもよい。このように、特有表現に含まれる表現が呼称に限定されないことで、情報処理端末１０は、ユーザ４０とのコミュニケーションの幅を広げることができる。

[0022] 図１は、本開示の一実施形態の概要を示す説明図である。図１の上部には、情報処理端末１０が特有表現を学習する際の一例が示されている。図１の上部に示す一例では、ユーザ４０Ａが情報処理端末１０に対して、皿５２を

提示している。同時に、ユーザ４０Ａは情報処理端末１０に対して、「これはカレーの皿だよ。棚の一番上の段にあるよ。」（ユーザ発話３００）という発話をしている。

[0023] 上述のユーザの動作に対して、情報処理端末１０は、ユーザ４０Ａが提示した皿５２の画像を入力情報として取得し、当該画像を一般物体認識により解析することで皿５２が一般表現で「皿」であると認識する。また、情報処理端末１０は、ユーザ発話３００の音声を入力情報として取得し、当該音声を解析することで「カレーの皿」を特有表現として取得する。そして、情報処理端末１０は、「皿」であると認識した皿５２と、取得した「カレーの皿」という特有表現とを対応付けることで学習する。なお、情報処理端末１０は、ユーザ発話３００の「棚の一番上の段にあるよ。」という皿５２の場所を示す情報を、皿５２の特有表現と対応付けて学習してもよい。

[0024] 一方、図１の下部には、情報処理端末１０が特有表現を提示情報に適用する際の一例が示されている。図１の下部に示す一例では、ユーザ４０Ｂが情報処理端末１０に対して、「カレーの皿はどこ？」（ユーザ発話３０２）という発話をしている。ユーザ４０Ｂが発話すると、情報処理端末１０は、ユーザ発話３０２の音声を入力情報として取得する。そして、情報処理端末１０は、ユーザ発話３０２の音声を解析することで、ユーザがカレーの皿の場所を質問していると認識する。同時に、情報処理端末１０は、ユーザ発話３０２に特有表現が含まれているか否かを確認する。

[0025] なお、音声に特有表現が含まれている場合、情報処理端末１０は、当該特有表現を適用した情報提示を行う。ユーザ４０が特有表現を用いた場合、当該ユーザには特有表現が通じることを示しており、また当該特有表現がユーザ４０にとってはより親しみやすい、あるいは分かりやすい表現であることが想定される。

[0026] また、ユーザ４０が特有表現を用いているのに対し、情報処理端末１０が一般表現を用いると、ユーザ４０は、どこか会話がかみ合っていないと感じたり、機械と会話をしていると強く感じたりしてしまう場合がある。一方、

ユーザ４０が特有表現を用いているのに対し、情報処理端末１０がユーザ４０に合わせて同様な特有表現を用いることで、ユーザ４０は、情報処理端末１０に対して人間らしさを感じることができる。このため、情報処理端末１０が特有表現を用いた情報提示を行うことで、ユーザ４０が提示された情報をより直感的に理解することができ、またユーザ４０が情報処理端末１０により親しみを感じる等の効果が期待される。

[0027] また、特有表現が含まれていない場合、情報処理端末１０は、特有表現を用いずに一般表現を用いた情報提示を行う。図１に示す例の場合、ユーザ発話３０２に「カレーの皿」という特有表現が含まれているため、情報処理端末１０は、当該特有表現を用いて「カレーの皿は棚の一番上の段にあります。」（端末発話３０４）と情報提示を行う。

[0028] ＜＜１．２．システム構成例＞＞

次に、図２を参照しながら、本実施形態に係る情報処理システムの構成例について説明する。図２は、本実施形態に係る情報処理システムの構成例を示すブロック図である。図２を参照すると、本実施形態に係る情報処理システムは、情報処理端末１０及び情報処理サーバ２０を備える。また、情報処理端末１０と情報処理サーバ２０は、ネットワーク３０を介して互いに通信が行えるように接続される。

[0029] （１）情報処理端末１０

本実施形態に係る情報処理端末１０は、情報処理サーバ２０による制御に基づいて、ユーザに対し音声や視覚情報を用いた応答情報の出力を行う情報処理装置である。本実施形態に係る情報処理端末１０は、情報処理サーバ２０が特有表現を適用した出力情報に基づき、応答情報を出力することを特徴の一つとする。

[0030] 本実施形態に係る情報処理端末１０は、音声や視覚情報を出力する機能を有する種々の装置として実現され得る。本実施形態に係る情報処理端末１０は、例えば、携帯電話、スマートフォン、タブレット、ウェアラブル装置、汎用コンピュータ、据え置き型または自律移動型の専用装置などであっても

よい。

[0031] また、本実施形態に係る情報処理端末１０は、ユーザや周囲環境に係る種々の情報を収集する機能を有する。情報処理端末１０は、例えば、ユーザの発話を含む音情報、ユーザが機器操作により入力した入力文章、ユーザや周囲を撮像した画像情報、その他の種々のセンサ情報を収集し、情報処理サーバ２０に送信する。

[0032] （２）情報処理サーバ２０

本実施形態に係る情報処理サーバ２０は、入力された情報に基づき特有表現を学習する情報処理装置である。例えば、本実施形態に係る情報処理サーバ２０は、入力された情報に基づき、認識対象の特有表現を学習することを特徴の一つとする。具体的には、本実施形態に係る情報処理サーバ２０は、収集された音声、または画像の少なくとも一方の入力情報に基づき、認識対象と当該認識対象に係る言語表現を対応付けて学習する。なお、言語表現は、ユーザを含む集団内における特有表現を含む。

[0033] また、本実施形態に係る情報処理サーバ２０は、入力された情報に基づき、ユーザに対する応答情報の出力を制御する情報処理装置でもある。そして、情報処理サーバ２０は、ユーザが属する集団に対して特有表現を用いた応答情報の提示が可能な場合、音声、または画像の少なくとも一方を用いて特有表現を用いた応答情報の出力を制御する。

[0034] （３）ネットワーク３０

ネットワーク３０は、情報処理端末１０と情報処理サーバ２０を接続する機能を有する。ネットワーク３０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３０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３０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。

[0035] 以上、図２を参照しながら、本実施形態に係る情報処理システムの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端末１０及び情報処理サーバ２０が有する機能は、単一の装置により実現されてもよい。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。

[0036] ＜＜１．３．情報処理端末１０の機能構成例＞＞

次に、図３を参照しながら、本実施形態に係る情報処理端末１０の機能構成例について説明する。図３は、本実施形態に係る情報処理端末１０の機能構成例を示すブロック図である。図３を参照すると、本実施形態に係る情報処理端末１０は、表示部１１０、音声出力部１２０、音声入力部１３０、撮像部１４０、センサ入力部１５０、制御部１６０、及びサーバ通信部１７０を備える。

[0037] （１）表示部１１０

本実施形態に係る表示部１１０は、画像やテキストなどの視覚情報を出力する機能を有する。本実施形態に係る表示部１１０は、例えば、情報処理サーバ２０による制御に基づいて、応答情報に対応するテキストや画像を表示する。

[0038] このために、本実施形態に係る表示部１１０は、視覚情報を提示する表示デバイスなどを備える。上記の表示デバイスには、例えば、液晶ディスプレイ（ＬＣＤ：Ｌｉｑｕｉｄ　Ｃｒｙｓｔａｌ　Ｄｉｓｐｌａｙ）装置、ＯＬＥＤ（Ｏｒｇａｎｉｃ　Ｌｉｇｈｔ　Ｅｍｉｔｔｉｎｇ　Ｄｉｏｄｅ）装置、タッチパネルなどが挙げられる。また、本実施形態に係る表示部１１０は、プロジェクション機能により視覚情報を出力してもよい。

[0039] （２）音声出力部１２０

本実施形態に係る音声出力部１２０は、音声発話を含む種々の音を出力する機能を有する。本実施形態に係る音声出力部１２０は、例えば、情報処理

サーバ２０による制御に基づいて、応答情報に対応する音声発話を出力する。このために、本実施形態に係る音声出力部１２０は、スピーカやアンプなどの音声出力装置を備える。

[0040] （３）音声入力部１３０

本実施形態に係る音声入力部１３０は、ユーザによる発話や、情報処理端末１０の周囲で発生する周囲音などの音情報を収集する機能を有する。音声入力部１３０が収集する音情報は、情報処理サーバ２０による音声認識や周囲環境の認識などに用いられる。本実施形態に係る音声入力部１３０は、音情報を収集するためのマイクロフォンを備える。

[0041] （４）撮像部１４０

本実施形態に係る撮像部１４０は、ユーザや周囲環境の画像を撮像する機能を有する。撮像部１４０が撮像した画像情報は、情報処理サーバ２０によるユーザの行動認識や状態認識、周囲環境の認識に用いられる。本実施形態に係る撮像部１４０は、画像を撮像することが可能な撮像装置を備える。なお、上記の画像には、静止画像のほか動画像が含まれる。

[0042] （５）センサ入力部１５０

本実施形態に係るセンサ入力部１５０は、周囲環境やユーザの行動、状態に関する種々のセンサ情報を収集する機能を有する。センサ入力部１５０が収集したセンサ情報は、情報処理サーバ２０による周囲環境の認識やユーザの行動認識、状態認識に用いられる。センサ入力部１５０は、例えば、赤外線センサを含む光センサ、加速度センサ、ジャイロセンサ、地磁気センサ、熱センサ、振動センサ、ＧＮＳＳ（Ｇｌｏｂａｌ　Ｎａｖｉｇａｔｉｏｎ　Ｓａｔｅｌｌｉｔｅ　Ｓｙｓｔｅｍ）信号受信装置などを備える。

[0043] また、本実施形態に係るセンサ入力部１５０は、ユーザが機器操作により入力した入力文章を検出する機能を有する。このために、本実施形態に係るセンサ入力部１５０は、例えば、キーボード、タッチパネル、マウス、各種のボタンなどを備える。

[0044] （６）制御部１６０

本実施形態に係る制御部１６０は、情報処理端末１０が備える各構成を制御する機能を有する。制御部１６０は、例えば、各構成の起動や停止を制御する。また、制御部１６０は、情報処理サーバ２０により生成される制御信号を表示部１１０や音声出力部１２０に入力する。また、本実施形態に係る制御部１６０は、後述する情報処理サーバ２０の出力制御部２５０と同等の機能を有してもよい。

[0045] （７）サーバ通信部１７０

本実施形態に係るサーバ通信部１７０は、ネットワーク３０を介して情報処理サーバ２０との情報通信を行う機能を有する。具体的には、サーバ通信部１７０は、音声入力部１３０が収集した音情報や、撮像部１４０が撮像した画像情報、センサ入力部１５０が収集したセンサ情報を情報処理サーバ２０に送信する。また、サーバ通信部１７０は、情報処理サーバ２０から応答情報の出力に係る制御信号などを受信する。

[0046] 以上、図３を参照しながら、本実施形態に係る情報処理端末１０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、本実施形態に係る情報処理端末１０の機能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端末１０は、図３に示す構成のすべてを必ずしも備えなくてもよい。例えば、情報処理端末１０は、表示部１１０やセンサ入力部１５０などを備えない構成をとることもできる。また、上述したように、本実施形態に係る制御部１６０は、情報処理サーバ２０の出力制御部２５０と同等の機能を有してもよい。本実施形態に係る情報処理端末１０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。

[0047] ＜＜１．４．情報処理サーバ２０の機能構成例＞＞

次に、図４～６を参照しながら、本実施形態に係る情報処理サーバ２０の機能構成例について詳細に説明する。図４は、本実施形態に係る情報処理サーバ２０の機能構成例を示すブロック図である。図４を参照すると、本実施形態に係る情報処理サーバ２０は、通信部２１０、入力解析部２２０、学習部２３０、記憶部２４０、出力制御部２５０を備える。

[0048] （１）通信部２１０

本実施形態に係る通信部２１０は、ネットワーク３０を介して情報処理端末１０との情報通信を行う機能を有する。具体的には、通信部２１０は、情報処理端末１０から音情報、画像情報、センサ情報などを受信する。また、通信部２１０は、応答情報の出力に係る制御信号を情報処理端末１０に送信する。

[0049] （２）入力解析部２２０

本実施形態に係る入力解析部２２０は、入力された情報を解析する機能を有する。例えば、本実施形態に係る入力解析部２２０は、情報処理端末１０が収集したユーザの発話に係る音情報を解析する機能を有する。より具体的に、入力解析部２２０は、音情報の解析結果に基づき、音情報を他の構成が利用可能な情報に変換する。例えば、入力解析部２２０は、ユーザの発話に係る音情報に基づいてのユーザの発話内容をテキストに変換する。そして、入力解析部２２０は、当該テキストに基づき、ユーザの発話の意味を解析する。

[0050] 上述のように、入力解析部２２０は、入力されたユーザの発話に係る音情報を解析することで、ユーザの発話に含まれる特有表現を取得することができる。特有表現を取得後、入力解析部２２０は、当該特有表現を他の構成に出力する。例えば、入力解析部２２０は、特有表現の学習時に特有表現を取得した場合、当該特有表現を学習部２３０へ出力する。また、入力解析部２２０は、特有表現の適用時に特有表現を取得した場合、当該特有表現を出力制御部２５０に出力する。

[0051] また、本実施形態に係る入力解析部２２０は、情報処理端末１０が収集した画像を解析する機能を有する。例えば、入力解析部２２０は、入力された画像を解析することで、認識対象を認識し、当該認識対象に関する情報（以下では、認識対象情報とも称される）を取得することができる。

[0052] また、入力解析部２２０は、当該画像を解析し、ユーザの状態や行動、物体、周囲環境に係る認識を行う。入力解析部２２０は、当該画像を解析する

ことで、ユーザの視線や表情、感情、行動などを認識することが可能である。また、入力解析部２２０は、例えば、画像情報やセンサ情報に基づいて、ユーザが位置する場所の特性を推定することも可能である。

[0053] 上述のように、入力解析部２２０は、画像を解析することで認識対象情報を取得後、入力解析部２２０は、当該認識対象情報を他の構成に出力する。例えば、入力解析部２２０は、特有表現の学習時に認識対象情報を取得した場合、当該認識対象情報を学習部２３０、及び記憶部２４０へ出力する。また、入力解析部２２０は、特有表現の適用時に認識対象情報を取得した場合、当該認識対象情報を出力制御部２５０に出力する。

[0054] （３）学習部２３０

学習部２３０は、Ｄｅｅｐ　Ｌｅａｒｎｉｎｇなどのアルゴリズムを用いて入力情報に基づく学習を行う機能を有する。例えば、本実施形態に係る学習部２３０は、入力情報に基づき認識した認識対象と、入力情報に基づき取得した特有表現を対応付けることで学習する。より具体的に、学習部２３０は、入力解析部２２０から入力された認識対象情報と特有表現を対応付けた上で記憶部２４０に出力して記憶させる。

[0055] 学習部２３０における特有表現の学習例として、例えば、学習部２３０は、画像に基づいて認識される物体と呼称とを対応付けて学習する。具体的に、学習部２３０は、入力解析部２２０が画像を解析することで認識された認識対象と、その呼称とを対応付ける。そして、学習部２３０は、対応付けた情報を記憶部２４０へ出力する。

[0056] また、学習部２３０は、画像から認識される物体と、音声から認識された物体に対してユーザが用いた特有表現とを対応付けて学習する。より具体的に、学習部２３０は、入力解析部２２０が画像を解析することで認識された認識対象と、入力解析部２２０が音情報を解析することで認識された、ユーザが認識対象に対して用いた特有表現とを対応付ける。そして、学習部２３０は、対応付けた情報を記憶部２４０へ出力する。

[0057] 上述のように、画像と音声を用いることは、情報処理サーバ２０が音声を

うまく認識できなかった時に効果がある。例えば、ユーザ４０の発話をはっきりと認識できず、複数の認識候補がある場合、ユーザ４０が提示している情報を示す画像に基づき、当該情報の内容に一番近い認識候補を選択することができる。

[0058] また、学習部２３０は、音声から認識される特有表現と、ユーザに関する実績情報に基づき認識される物体とを対応付けて学習する。より具体的に、学習部２３０は、入力解析部２２０が音情報を解析することで認識された特有表現と、スケジュール等の実績情報に基づき特定される認識対象とを対応付ける。そして、学習部２３０は、対応付けた情報を記憶部２４０へ出力する。

[0059] 例えば、ユーザ４０が情報処理端末１０に対して「この前の結婚式で使った鞄はどれ？」と発話したとする。この時、情報処理サーバ２０は、ユーザ４０のスケジュールから結婚式の日を検索し、その日の写真を解析することで、ユーザ４０が結婚式で使った鞄を特定する。そして、学習部２３０は、音声から認識された「結婚式の鞄」という特有表現と、ユーザに関する実績情報に基づき認識された物体である「鞄」とを対応付けて学習する。

[0060] また、ユーザに関する実績として、過去に他のユーザ４０が発話した際に、情報処理サーバ２０が発話内容と、当該発話内容に対応する画像を登録していたとする。情報処理サーバ２０がユーザ４０の発話をはっきりと認識できなかったとしても、当該ユーザ４０の発話が他のユーザ４０の過去の発話と類似していた場合、情報処理サーバ２０は、当該画像に基づき、誤って認識した音声を正すことができる。

[0061] 例えば、過去にユーザ４０Ｂが「銀行に行く」と発話した際に、当該銀行の画像を登録していたとする。そして、ユーザ４０Ａが「銀行に行く」と発話したが、情報処理サーバ２０が「近郊に行く」と認識したとする。この時、情報処理サーバ２０は、「近郊」と「銀行」の発音が似ていることから、音声を誤って認識したと推定した場合、ユーザ４０Ｂが過去に登録した銀行の画像を提示してユーザ４０Ａに確認することができる。

[0062] また、学習部２３０は、発話者が発話対象者に用いた特有表現と、発話対象者とを対応付けて学習する。例えば、ある母と子供がいたとして、母が発話者であり、子供が発話対象者であるとする。子供の名前は「ゆうた」であるが、母が「ゆうちゃん」と子供に呼びかけても子供が反応したとする。この場合、入力解析部２２０は、発話者である母が発話対象者である子供に用いた特有表現を「ゆうちゃん」と認識する。そして、学習部２３０は、「ゆうちゃん」という特有表現と、発話対象者である子供とを対応付けて学習する。

[0063] また、学習部２３０は、発話対象者の特徴に応じて、発話対象者と特有表現の対応付けを再評価する。例えば、父、母、２人の子供で構成される家族があり、母が発話者であり、父、２人の子供が発話対象者であるとする。母が「男性の人」と呼びかけると、父と２人の子供が反応する。しかし、母が「男の子の人」と呼びかけると、２人の子供が反応し、父は反応しない。上述のどちらの表現も、性別が男である人に対する表現であるが、発話対象者が大人であるか子供であるかにより反応に差が出た。

[0064] よって、情報処理端末１０が上述の例の様子を見て、情報処理サーバ２０が学習していたとすると、情報処理サーバ２０は、「大人」と「子供」が発話対象者の特徴であると認識する。また、情報処理サーバ２０は、「男性」という表現が父と２人の子供に対する特有表現と認識する。また、情報処理サーバ２０は、「男の子」という表現は、２人の子供に対する特有表現と認識する。

[0065] もし、２人の子供が成長して大人になった場合、情報処理サーバ２０は、当該２人に対する特有表現を再評価し、「男の子」から「大人」に更新してもよい。

[0066] なお、学習部２３０は、特有表現の学習の他に、ユーザの問い合わせに対する回答などの学習、画像認識、音声認識、機械制御に係る学習を行ってもよい。また、本実施形態に係る学習アルゴリズムは、上記の例に限定されず、生成される応答情報の特性に応じて適宜採択され得る。また、学習部２３

０は、上述した学習例における学習方法を組み合わせて用いてもよい。

[0067] （４）記憶部２４０

記憶部２４０は、入力解析部２２０が取得した情報を記憶する機能を有する。例えば、記憶部２４０は、入力解析部２２０が取得したユーザに関する情報をユーザＤＢに記憶し、入力解析部２２０が取得した物体に関する情報を一般物体ＤＢに記憶する。

[0068] ユーザＤＢは、ユーザに係る種々の情報を記憶するデータベースである。ユーザＤＢは、ユーザの名前やＩＤのほか、例えば、年齢、性別や、ユーザ間の関係性などの属性情報を記憶する。上記のユーザの間の関係性には、例えば、妻と夫、母親と子供、などの家族内における属性や、上司と部下などの組織内における属性などが含まれてよい。

[0069] また、ユーザＤＢは、ユーザのスケジュールや、ユーザが認識対象を使用した実績等を記憶してよい。上記の情報は、学習部２３０が認識対象と特有表現の対応付けを行う際に用いられ得る。さらに、ユーザＤＢは、ユーザの画像情報や音声特徴などを記憶してよい。入力解析部２２０は、ユーザＤＢに記憶される上記の情報に基づいて、ユーザを識別することも可能である。

[0070] 一般物体ＤＢは、物体に関する種々の情報を記憶するデータベースである。一般物体ＤＢは、物体の一般的な名称のほか、例えば、色、大きさ、形状等の物体の特徴を示す情報を記憶してもよい。

[0071] また、記憶部２４０は、学習部２３０が学習した情報を記憶する機能も有する。例えば、記憶部２４０は、学習部２３０が学習した情報を特有表現ＤＢに記憶する。より具体的に、記憶部２４０は、学習部２３０が対応付けた認識対象と特有表現に関する情報を特有表現ＤＢに記憶する。

[0072] ここで、図５を参照しながら、特有表現ＤＢの構成について説明する。図５は、本実施形態に係る特有表現ＤＢの構成例を示す説明図である。図５に示すように、特有表現ＤＢでは、認識対象の一般表現と、当該認識対象に対してユーザ４０が用いる特有表現とが対応付けて記憶されている。例えば、図１で説明した例の場合、入力解析部２２０が入力された画像を解析して取

得した「皿」という一般表現と、入力解析部２２０が入力された音声を解析して取得した「カレーの皿」という特有表現とが対応付けられて記憶される。

[0073] また、特有表現ＤＢは、一般表現と特有表現の他に、対象ユーザ、回数、及び画像に関する情報も記憶してよい。対象ユーザは、登録された特有表現を用いられたユーザであり、特有表現と対応付けられて記憶される。例えば、母親が赤ちゃんに対して幼児言葉を使ったとする。この時、特有表現は幼児言葉であり、対象ユーザは赤ちゃんであるため、幼児言葉と赤ちゃんが対応付けられて記憶される。そして、情報処理端末１０は、赤ちゃんに対して幼児言葉を用いた発話を行うことができる。もし、対象ユーザが特有表現を用いたユーザである場合、幼児言葉と母親が対応付けられてしまう。そして、情報処理端末１０が母親に対して幼児言葉を用いてしまうことになるが、これは間違いである。よって、対象ユーザは、特有表現を用いられたユーザであることが重要である。

[0074] また、回数は、認識対象が認識された回数であり、特有表現と対応付けられて記憶される。また、画像は、認識対象の画像、または認識対象に関する情報を示す画像のことであり、特有表現と対応付けられて記憶される。なお、回数、及び画像は、必ずしも記憶されなくてもよい。

[0075] なお、図５に示す特有表現ＤＢでは、複数の対象者が１つのレコードに登録されているが、対象者１人に１つのレコードを作成して登録してもよい。

[0076] （５）出力制御部２５０

出力制御部２５０は、ユーザに対する応答情報の出力を制御する機能を有する。本実施形態に係る出力制御部２５０は、学習部２３０の学習により記憶部２４０に記憶された特有表現と、入力解析部２２０が取得した情報に基づいて応答情報の出力表現を制御することを特徴の一つとする。

[0077] 出力表現を制御するにあたり、出力制御部２５０は、まず、入力された音声または画像の少なくとも一方の入力情報に基づき、特有表現を提示される情報に適用可能か否かを判定する。

[0078] 例えば、出力制御部２５０は、入力された画像に基づき判定を行う場合、画像に基づき認識された認識対象に適用可能な特有表現が存在するか否かを確認する。この時、出力制御部２５０は、記憶部２４０の特有表現ＤＢに認識対象と対応付けられた特有表現が登録されているか否かを確認する。確認の結果、適用可能な特有表現が存在する場合、出力制御部２５０は、提示される情報に特有表現を適用可能と判定する。

[0079] また、出力制御部２５０は、入力された音声に基づき判定を行う場合、音声に基づき認識された言語表現に適用可能な特有表現が存在するか否かを確認する。この時、出力制御部２５０は、記憶部２４０の特有表現ＤＢに言語表現と対応する特有表現が登録されているか否かを確認する。確認の結果、適用可能な特有表現が存在する場合、出力制御部２５０は、提示される情報に特有表現を適用可能と判定する。

[0080] 上述の各々の判定処理の結果、出力制御部２５０は、提示される情報に特有表現を適用可能と判定した場合、特有表現を適用した応答情報を通信部２１０に出力する。ここで出力制御部２５０が出力する応答情報とは、例えば、ユーザの発話に対する応答に関する音情報である。出力制御部２５０は、当該音情報に特有表現を適用する。なお、情報処理端末１０は、ユーザ４０からの発話が入力されなくても、自発的に情報を出力してもよい。

[0081] なお、出力制御部２５０は、特有表現と対応する画像が記憶部２４０の特有表現ＤＢに登録されている場合、当該画像と音声の両方を用いて情報提示を実行させてもよい。出力制御部２５０は、ユーザの発話に対して、音声だけでなく画像も用いて応答することで、より詳細な情報をユーザに提示することができる。

[0082] なお、上述の各々の判定処理の結果、出力制御部２５０は、検出された複数のユーザに基づき、提示される情報に特有表現を適用しないと判定する場合、一般表現を用いた情報提示を情報処理端末１０に実行させる。例えば、出力制御部２５０は、検出された複数のユーザから集団に属さないユーザが識別された場合、提示される情報に特有表現を適用しないと判定し、一般表

現を用いて情報提示を実行させる。具体例について、図６を参照しながら説明する。図６は、本実施形態に係る集団に属さないユーザへの情報提示の例を示す説明図である。

[0083] 図６の左側は、特有表現を適用可能な集団にユーザ４０Ａが属している場合における、情報処理端末１０による情報提示の例を示している。例えば、ユーザ４０Ａが「交通費は誰に申請すればいい？」（ユーザ発話３８０）という質問を情報処理端末１０にする。当該質問に対し、情報処理端末１０は、「ＵＬに申請してください。」（端末発話３８２）と特有表現を用いた情報提示をする。

[0084] 上述の例に対し、図６の右側は、特有表現を適用可能な集団にユーザ４０Ｂが属していない場合における、情報処理端末１０による情報提示の例を示している。例えば、ユーザ４０Ｂが「交通費は誰に申請すればいい？」（ユーザ発話３８４）という質問を情報処理端末１０にする。当該質問に対し、情報処理端末１０は、「ユニットリーダーのＤＤＤさんに申請してください。」（端末発話３８６）と一般表現を用いた情報提示をする。

[0085] 以上、図４～６を参照しながら、本実施形態に係る情報処理サーバ２０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機能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２０の機能構成は係る例に限定されない。例えば、入力解析部２２０、学習部２３０、記憶部２４０等は、情報処理サーバ２０とは異なる別の装置に備えることも可能である。

[0086] また、上述したように、本実施形態に係る出力制御部２５０が有する機能は、情報処理端末１０の制御部１６０の機能として実現されてもよい。すなわち、本実施形態に係る出力制御部２５０の機能は、サーバ側及びクライアント側の両方の機能として実現され得る。例えば、当該機能を情報処理サーバ２０の機能として備える場合、ユーザは、様々な情報処理端末１０においてサービスを享受することが可能である。一方、情報処理端末１０が出力制御部２５０と同等の機能や、入力解析部２２０、学習部２３０、及び記憶部

２４０等を備える場合、オフラインにおける使用や個人情報等のよりセキュアな保管が可能となる。本実施形態に係る情報処理サーバ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。

[0087] ＜＜１．５．動作例＞＞

以下では、図７、８を参照しながら、本実施形態に係る情報処理サーバ２０の動作例について説明する。まず、図７を参照しながら、ユーザからの入力情報に基づき特有表現を学習する際の情報処理サーバ２０の動作例について説明する。図７は、本実施形態に係る特有表現を学習する際の情報処理サーバ２０の動作例を示すフローチャートである。

[0088] まず、情報処理サーバ２０の通信部２１０が、情報処理端末１０が取得したユーザの発話を、入力情報としてネットワーク３０を介して受信する（ステップＳ１０００）。入力情報を受信した通信部２１０は、当該入力情報を入力解析部２２０に出力する。

[0089] 通信部２１０から入力情報を入力された入力解析部２２０は、認識対象を含む画像を受信したか否かを確認する（ステップＳ１００２）。当該画像を受信しなかった場合（ステップＳ１００２／ＮＯ）、情報処理サーバ２０は、通常依頼として処理を行う（ステップＳ１０１４）。

[0090] 認識対象を含む画像を受信した場合（ステップＳ１００２／ＹＥＳ）、入力解析部２２０は、入力情報の解析処理を行う。なお、入力解析部２２０は、画像を解析して一般表現を取得するステップＳ１００４、１００６の処理と、発話を解析して特有表現を取得するステップＳ１００８、１０１０の処理を並列処理として実行してもよい。

[0091] １つ目の並列処理として、入力解析部２２０は、入力された画像を解析する（ステップＳ１００４）。そして、入力解析部２２０は、画像の解析結果に基づき一般表現を取得する（ステップＳ１００６）。一般表現を取得後、入力解析部２２０は、当該一般表現を学習部２３０へ出力する。

[0092] ２つ目の並列処理として、入力解析部２２０は、入力された発話を解析する（ステップＳ１００８）。そして、入力解析部２２０は、発話の解析結果

に基づき特有表現を取得する（ステップＳ１０１０）。特有表現を取得後、入力解析部２２０は、当該特有表現を学習部２３０へ出力する。

[0093] 並列処理終了後、学習部２３０は、入力解析部２２０から入力された一般表現、特有表現、及び認識対象の画像を関連付けて、記憶部２４０に登録する（ステップＳ１０１２）。

[0094] 以上、図７を参照しながら、情報処理サーバ２０が認識対象の特有表現を学習する際の動作例について説明した。

[0095] 続いて、図８を参照しながら、ユーザからの入力情報に基づき情報処理サーバ２０が特有表現を用いた情報提示の制御を行う際の動作例について説明する。図８は、本実施形態に係る特有表現を適用する際の情報処理サーバ２０の動作例を示すフローチャートである。

[0096] まず、情報処理サーバ２０の通信部２１０が、情報処理端末１０が取得したユーザの発話を、入力情報としてネットワーク３０を介して受信する（ステップＳ２０００）。入力情報を受信した通信部２１０は、当該入力情報を入力解析部２２０に出力する。

[0097] 通信部２１０から入力情報を入力された入力解析部２２０は、入力情報に含まれる発話を解析する（ステップＳ２００２）。解析結果に基づき、入力解析部２２０は、記憶部２４０に登録されている特有表現が発話に含まれているか否かを確認する（ステップＳ２００４）。

[0098] 特有表現が発話に含まれていた場合（ステップＳ２００４／ＹＥＳ）、入力解析部２２０は、当該特有表現を用いて、ユーザの発話に回答する（ステップＳ２００６）。

[0099] また、特有表現が発話に含まれていなかった場合、情報処理サーバ２０は、通常依頼として処理を行う（ステップＳ２００８）。

[0100] なお、上述した動作例では、ステップＳ１０００、及びステップＳ２０００にて、通信部２１０が入力情報としてユーザの発話を受信したことをトリガーとして処理が開始されているが、トリガーは上述の例に限定されない。例えば、通信部２１０が入力情報として画像を受信したことをトリガーとし

て処理が開始されてもよい。

[0101] 以上、図８を参照しながら、情報処理サーバ２０が特有表現を用いた情報提示の制御を行う際の動作例について説明した。

[0102] 以上、図１～８を参照しながら、本開示の実施形態に係る情報処理システムについて説明した。続いて、本開示の実施形態に係る変形例について説明する。

[0103] ＜２．変形例＞

以下では、本開示の一実施形態の変形例を説明する。なお、以下に説明する変形例は、単独で本開示の実施形態に適用されてもよいし、組み合わせで本開示の実施形態に適用されてもよい。また、変形例は、本開示の実施形態で説明した構成に代えて適用されてもよいし、本開示の実施形態で説明した構成に対して追加的に適用されてもよい。

[0104] （１）第１の変形例

上述の実施形態は、図１を参照しながら、入力された音声に基づく判定処理を行う例について説明したが、入力された画像に基づく判定処理と、入力された音声に基づく判定処理を組み合わせた判定処理を行ってもよい。

[0105] 具体例について、図９を参照しながら説明する。図９は、本実施形態に係るユーザが提示する音声と画像に基づく特有表現の適用例を示す説明図である。図９に示すユーザ４０は、「これくらいの大きさの皿はどこ？」（ユーザ発話３２０）という発話に加え、皿の大きさを示すジェスチャーを行っている。情報処理端末１０には、ユーザ発話３２０が音声として、ユーザ４０のジェスチャーが画像として入力される。情報処理サーバ２０は、入力された音声と画像に基づき、上述の判定処理を行う。

[0106] 図９に示す例では、情報処理サーバ２０は、ユーザ４０のジェスチャーから皿の大きさを認識し、当該大きさの皿がカレーの皿であるという情報が記憶部２４０に登録されていたため、特有表現を適用可能と判定している。そして、情報処理端末１０は、情報処理サーバ２０が特有表現を適用した応答情報に基づき、「それはカレーの皿ですか？カレーの皿は棚の一番上の段に

あります。」（端末発話３２２）と発話をしている。

[0107] 上述のように、情報処理サーバ２０がユーザ４０の動作も認識できることにより、情報処理サーバ２０が認識できる特有表現の幅が広がる。そして、情報処理端末１０がユーザ４０と行うコミュニケーションの幅も広がる。よって、情報処理サーバ２０は、ユーザ４０に対する情報処理端末１０の利便性を向上することできる。

[0108] （２）第２の変形例

上述の実施形態では、情報処理サーバ２０の学習部２３０が一般表現と特有表現を対応付けて学習する例を説明したが、学習部２３０は、特有表現と特有表現に関する情報を示す画像とをさらに対応付けて学習してもよい。

[0109] 具体例について、図１０を参照しながら説明する。図１０は、本実施形態に係るユーザが提示する音声と画像に基づく特有表現の学習例及び適用例を示す説明図である。図１０の左側は、ユーザ４０が特有表現を登録済のカレーの皿５２の片付け場所を移動し、新しい片付け場所を情報処理端末１０に学習させようとしている状況を示している。具体的に、ユーザ４０Ａは、「カレーの皿を片付ける場所を変えるね。」（ユーザ発話３４０）という発話に加え、棚５４の一番上の段にあるカレーの皿５２を、棚５４の一番下の段に移動する動作を情報処理端末１０に見せて、情報処理端末１０に学習させようとしている。

[0110] この時、情報処理端末１０は、ユーザ４０Ａの動作に関する情報を画像として取得する。例えば、情報処理端末１０は、カレーの皿５２が新しい片付け場所に置いてある様子を、撮像部１４０により静止画像として取得する。また、情報処理端末１０は、ユーザ４０Ａがカレーの皿５２を移動する一連の動作を、撮像部１４０により動画像として取得してもよい。

[0111] 動作に関する情報の画像を取得した情報処理端末１０は、当該画像を情報処理サーバ２０に送信する。そして、画像を入力された情報処理サーバ２０の学習部２３０は、記憶部２４０に登録済のカレーの皿５２の特有表現と当該画像とを対応付けて学習し、記憶部２４０に記憶させる。

[0112] また、図１０の右側は、カレーの皿５２の移動後に、別のユーザ４０が情報処理端末１０にカレーの皿５２の場所を尋ねた状況を示している。具体的に、ユーザ４０Ｂは、「カレーの皿はどこ？」（ユーザ発話３４２）と情報処理端末１０に尋ねている。ユーザ４０Ｂからの発話を受けた情報処理端末１０は、当該発話を情報処理サーバ２０に解析させる。

[0113] 情報処理サーバ２０は、解析の結果からユーザ発話３４２に「カレーの皿」という特有表現が使われていることを認識し、記憶部２４０の特有表現ＤＢに「カレーの皿」という特有表現が登録されているかを確認する。この時、特有表現が登録されており、当該特有表現に対応付けられた画像が登録されている場合、情報処理サーバ２０は、情報処理端末１０に当該画像を用いた情報提示を実行させてもよい。図１０に示す例の場合、ユーザ４０Ａが「カレーの皿」に対応付けた画像が登録されている。よって、情報処理端末１０は、「カレーの皿は棚の一番下の棚に移動されました。」（端末発話３４４）に加えて、表示部１１０を用いて画像５６を表示してもよい。

[0114] 上述のように、情報処理サーバ２０が情報処理端末１０に画像も提示させることで、ユーザ４０は、情報処理端末１０が提示する情報をより詳細に理解することができる。よって、情報処理サーバ２０は、ユーザ４０に対する情報処理端末１０の利便性を向上することできる。

[0115] （３）第３の変形例

上述の実施形態では、情報処理端末１０は、ユーザ４０が提示した情報を画像として取得する例を説明したが、情報処理端末１０は、自発的に認識対象の情報を示す画像を取得し、情報処理サーバ２０が特有表現を学習するように動作してもよい。例えば、情報処理端末１０は、周囲の状況を示す画像を撮像し、当該画像を情報処理サーバ２０に送信する。情報処理サーバ２０では、入力解析部２２０が当該画像を解析する。そして、解析結果に基づき、学習部２３０が画像から認識される物体と、画像から推定される物体の用途に対応する呼称とを対応付けて学習する。

[0116] 具体例について、図１１を参照しながら説明する。図１１は、本実施形態

に係る画像に基づく特有表現の学習例を示す説明図である。図１１に示す情報処理端末１０は、自発的に周囲の状況を示す画像を取得し、当該画像を情報処理サーバ２０に送信する。情報処理サーバ２０の入力解析部２２０は、当該画像を解析することで、テーブル５８の皿５２にカレーが盛られていることを認識したとする。この時、学習部２３０は、皿５２はカレーを盛る用途に使われていると推定し、その呼称を「カレーの皿」と推定する。そして、学習部２３０は、画像から認識された物体である皿５２と、推定した用途に対応する呼称である「カレーの皿」とを対応付けて学習し、記憶部２４０に記憶させる。

[0117] また、学習部２３０は、認識対象を認識した回数に基づき、認識対象に用いる特有表現の信頼度を設定する。例えば、学習部２３０は、信頼度を示す値として当該回数を設定する。具体的に、認識対象を認識した回数が所定の回数より少ない場合、出力制御部２５０は、当該認識対象に対応する特有表現の信頼度は低いと判定する。また、認識対象を認識した回数が所定の回数より多い場合、出力制御部２５０は、当該認識対象に対応する特有表現の信頼度は高いと判定する。

[0118] そして、出力制御部２５０は、上述の判定処理の結果に基づき、特有表現を用いた応答情報の出力表現を制御する。例えば、出力制御部２５０は、皿５２に対応する特有表現である「カレーの皿」の信頼度が低い場合に、ユーザ４０にカレーの皿の場所を問う質問をされたとする。その時、出力制御部２５０は、情報処理端末１０に「カレーの皿はこの皿ですか？この皿であれば棚の一番上の段にあります。」という発話をさせると同時に、カレーの皿を示す画像を表示させてもよい。

[0119] 上述のように、情報処理端末１０が自発的に取得した情報に基づき、情報処理サーバ２０が学習することで、ユーザ４０は、情報処理サーバ２０に特有表現を学習させる手間を省くことができる。よって、情報処理サーバ２０は、ユーザ４０に対する情報処理端末１０の利便性を向上することできる。

[0120] （４）第４の変形例

上述の実施形態では、情報処理端末１０は、ユーザ４０が提示した情報を画像として、ユーザ４０が情報処理端末１０に対して行った発話を音声として取得する例を説明した。しかし、情報処理端末１０は、周囲の状況を示す画像と、周囲で発せられている音声を自発的に取得し、情報処理サーバ２０が特有表現を学習するように動作してもよい。例えば、情報処理端末１０は、周囲の状況を示す画像としてユーザが会話している画像を取得し、周囲で発せられている音声としてユーザの会話を取得し、当該画像及び音声を情報処理サーバ２０に送信する。情報処理サーバ２０では、入力解析部２２０が当該画像及び音声を解析する。そして、解析結果に基づき、学習部２３０は、画像から認識される物体と、ユーザの会話から認識された物体に対してユーザが用いた特有表現とを対応付けて学習する。

[0121] 具体例について、図１２を参照しながら説明する。図１２は、本実施形態に係る情報処理端末１０が自発的に取得した音声と画像に基づく特有表現の学習例を示す説明図である。図１２に示す情報処理端末１０は、周囲の状況を示す画像として、ユーザ４０Ａと、リモートコントローラ６０を持ったユーザ４０Ｂが会話をしている様子を画像として取得する。また、情報処理端末１０は、周囲から発せられる音声として、「チャンネル取って。」というユーザ発話３６０と「はい、チャンネル。」というユーザ発話３６２を取得する。そして、情報処理端末１０は、取得した画像と音声を情報処理サーバ２０に送信する。

[0122] 情報処理サーバ２０の入力解析部２２０は、上述の画像を解析することで、リモートコントローラ６０を認識対象として認識し、上述の音声を解析することで、ユーザ４０Ａ及びユーザ４０Ｂがリモートコントローラ６０に対して「チャンネル」という特有表現を用いていると認識する。そして、学習部２３０は、画像から認識された物体であるリモートコントローラ６０と、ユーザの会話から取得した「チャンネル」という特有表現とを対応付けて学習し、記憶部２４０に記憶させる。

[0123] 上述のように、情報処理端末１０が自発的に取得した情報に基づき、情報

処理サーバ２０が学習することで、ユーザ４０は、情報処理サーバ２０に特有表現を学習させる手間を省くことができる。よって、情報処理サーバ２０は、ユーザ４０に対する情報処理端末１０の利便性を向上することできる。

[0124] （５）第５の変形例

上述の実施形態では、特有表現を適用可能な集団にユーザ４０が属していない場合、情報処理サーバ２０の出力制御部２５０は、情報処理端末１０に一般表現を用いた情報提示を実行させる例を説明した。しかし、ユーザ４０が特有表現を理解していないことが推定された場合も、出力制御部２５０は、提示される情報に特有表現を適用しないと判定し、情報処理端末１０に一般表現を用いた情報提示を実行させる。

[0125] 具体例について、図１３を参照しながら説明する。図１３は、本実施形態に係るユーザが特有表現を理解していないと推定される場合の情報提示の例を示す説明図である。図１３に示す例では、ユーザ４０Ａが「ＵＬを呼んできて。」（ユーザ発話４００）とユーザ４０Ｂに対して発話している。ユーザ発話４００に対し、ユーザ４０Ｂは、「ＵＬ？」（ユーザ発話４０２）とユーザ発話４００の内容を理解していない旨を示す回答をしている。

[0126] この時、情報処理端末１０は、ユーザ発話４００及びユーザ発話４０２を取得し、情報処理サーバ２０に送信する。情報処理サーバ２０の入力解析部２２０は、ユーザ発話４００及びユーザ発話４０２を解析する。当該解析により、ユーザ４０Ａが用いた特有表現をユーザ４０Ｂが理解していないことが推定された場合、出力制御部２５０は、ユーザ４０Ａが用いた特有表現に対応する一般表現を用いた情報提示を、情報処理端末１０に実行させる。図１３に示す例の場合、情報処理端末１０は、「ＵＬとは、ユニットリーダーのＤＤＤさんのことです。」（端末発話４０４）と一般表現を用いた情報提示を行う。

[0127] なお、入力解析部２２０は、ユーザ４０の表情を解析することにより、ユーザ４０が特有表現を理解しているか否かを推定してもよい。

[0128] また、上述の例では、職場にて用いられる特有表現に関する例を説明した

が、本変形例は、方言が特有表現として用いられた場合にも適用可能である。例えば、ユーザ４０Ａがユーザ４０Ｂに対して方言を用いた発話を行ったとする。この時、入力解析部２２０の解析により、ユーザ４０Ａが用いた方言をユーザ４０Ｂが理解していないことが推定された場合、出力制御部２５０は、ユーザ４０が用いた方言に対応する一般表現を用いた情報提示を、情報処理端末１０に実行させる。なお、学習部２３０は、上述の実施形態、または変形例に記載の方法を用いて方言を学習することが可能である。よって、出力制御部２５０は、学習した情報を用いて方言を一般表現に変換し、一般表現を用いた情報提示を、情報処理端末１０に実行させてもよい。また、出力制御部２５０は、あらかじめ登録された方言に関する情報を用いて方言を一般表現に変換し、一般表現を用いた情報提示を、情報処理端末１０に実行させてもよい。

[0129] 上述のように、情報処理サーバ２０がユーザ４０の状態を推定し、当該状態に応じた情報を情報処理端末１０が提示することで、ユーザ４０は、分からないことがある時などに自分で調べる手間を省くことができる。よって、情報処理サーバ２０は、ユーザ４０に対する情報処理端末１０の利便性を向上することできる。

[0130] 以上、図９～１３を参照しながら、本開示の一実施形態に係る変形例について説明した。続いて、本開示の一実施形態にかかる情報処理システムの応用例について説明する。

[0131] ＜３．ハードウェア構成例＞

次に、図１４を参照しながら、本開示の一実施形態に係る情報処理端末１０及び情報処理サーバ２０に共通するハードウェア構成例について説明する。図１４は、本開示の一実施形態に係る情報処理端末１０及び情報処理サーバ２０のハードウェア構成例を示すブロック図である。図１４を参照すると、情報処理端末１０及び情報処理サーバ２０は、例えば、ＣＰＵ８７１と、ＲＯＭ８７２と、ＲＡＭ８７３と、ホストバス８７４と、ブリッジ８７５と、外部バス８７６と、インターフェース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、ハードウェア構成は、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。

[0132] （ＣＰＵ８７１）

ＣＰＵ８７１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。

[0133] （ＲＯＭ８７２、ＲＡＭ８７３）

ＲＯＭ８７２は、ＣＰＵ８７１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、ＣＰＵ８７１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。

[0134] （ホストバス８７４、ブリッジ８７５、外部バス８７６、インターフェース８７７）

ＣＰＵ８７１、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、インターフェース８７７を介して種々の構成要素と接続される。

[0135] （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。

[0136] （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。

[0137] （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。

[0138] （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。

[0139] （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。

[0140] （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。

[0141] （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。

[0142] （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、光通信用のルータ、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。

[0143] 以上、図１４を参照しながら、本開示の一実施形態に係る情報処理端末１０及び情報処理サーバ２０のハードウェア構成例について説明した。

[0144] ＜４．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２０は、ユーザ４０を含む集団内における特有表現を適用可能な場合、音声または画像のうち少なくとも一方を用いて、ユーザ４０に対する当該特有表現を含む情報提示を、情報処理端末１０に実行させる。

[0145] その結果、情報処理端末１０は、ユーザ４０が用いた特有表現を用いた情報提示を行うことができる。また、ユーザ４０は、ユーザ４０が用いた特有表現を情報処理端末１０が理解した上で、ユーザ４０の発話に対する情報提示を行っているように感じることができる。よって、ユーザとのコミュニケーションをより円滑に行うことが可能な、新規かつ改良された情報処理装置、及び情報処理方法を提供することが可能である。

[0146] 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。

[0147] また、本明細書の情報処理サーバ２０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、情報処理サーバ２０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。

[0148] 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）

ユーザに対する情報提示を制御する出力制御部、

を備え、

前記出力制御部は、提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち少なくとも一方を用いて実行させる、

情報処理装置。

（２）

前記出力制御部は、入力された音声または画像の少なくとも一方の入力情報に基づき、前記特有表現を前記提示される情報に適用可能か否かを判定する、前記（１）に記載の情報処理装置。

（３）

前記出力制御部は、前記画像に基づき認識された認識対象に適用可能な前記特有表現が存在する場合、前記提示される情報に前記特有表現を適用可能と判定する、前記（２）に記載の情報処理装置。

（４）

前記出力制御部は、前記音声に基づき認識された言語表現に適用可能な前記特有表現が存在する場合、前記提示される情報に前記特有表現を適用可能と判定する、前記（２）または（３）に記載の情報処理装置。

（５）

前記出力制御部は、前記特有表現と対応する前記画像が登録されている場合、前記音声及び前記画像を用いて前記情報提示を実行させる、前記（１）～（４）のいずれか一項に記載の情報処理装置。

（６）

前記出力制御部は、検出された複数のユーザに基づき、前記提示される情報に前記特有表現を適用しないと判定する場合、一般表現を用いて前記情報提示を実行させる、前記（２）に記載の情報処理装置。

（７）

前記出力制御部は、前記複数のユーザから前記集団に属さないユーザが識別された場合、前記提示される情報に前記特有表現を適用しないと判定し、一般表現を用いて前記情報提示を実行させる、前記（６）に記載の情報処理装置。

（８）

前記出力制御部は、前記ユーザが前記特有表現を理解していないことが推定された場合、前記提示される情報に前記特有表現を適用しないと判定し、一般表現を用いた前記情報提示を実行させる、前記（６）または（７）に記載の情報処理装置。

（９）

認識対象と前記認識対象に係る言語表現とを対応付けて学習する学習部、

を備え、

前記言語表現は、ユーザを含む集団内における特有表現を含み、

前記学習部は、収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習する、

情報処理装置。

（１０）

前記特有表現は、物体に対する呼称を含み、

前記学習部は、前記画像に基づいて認識される前記物体と前記呼称とを対応付けて学習する、前記（９）に記載の情報処理装置。

（１１）

前記学習部は、前記画像から認識される物体と、前記音声から認識された前記物体に対して前記ユーザが用いた特有表現とを対応付けて学習する、前記（９）に記載の情報処理装置。

（１２）

前記学習部は、前記画像から認識される物体と、前記画像から推定される前記物体の用途に対応する前記呼称とを対応付けて学習する、前記（１０）に記載の情報処理装置。

（１３）

前記学習部は、前記画像から認識される物体と、ユーザの会話から認識された前記物体に対して前記ユーザが用いた特有表現とを対応付けて学習する、前記（９）に記載の情報処理装置。

（１４）

前記学習部は、前記音声から認識される前記特有表現と、前記ユーザに関する実績情報に基づき認識される物体とを対応付けて学習する、前記（９）に記載の情報処理装置。

（１５）

前記学習部は、発話者が発話対象者に用いた特有表現と、前記発話対象者とを対応付けて学習する、前記（９）に記載の情報処理装置。

（１６）

前記学習部は、前記発話対象者の特徴に応じて、前記発話対象者と前記特有表現の対応付けを再評価する、前記（１５）に記載の情報処理装置。

（１７）

前記学習部は、前記認識対象を認識した回数に基づき、前記認識対象に用いる前記特有表現の信頼度を設定する、前記（９）に記載の情報処理装置。

（１８）

前記特有表現は、呼称、形容表現、または動作表現の少なくともいずれか１つを含む、前記（９）～（１７）のいずれか一項に記載の情報処理装置。

（１９）

ユーザに対する情報提示を制御することと、

提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち

少なくとも一方を用いて実行させることと、

を含む、プロセッサにより実行される、情報処理方法。

（２０）

認識対象と前記認識対象に係る言語表現とを対応付けて学習することと、

前記言語表現は、ユーザを含む集団内における特有表現を含み、

収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習することと、

を含む、プロセッサにより実行される、情報処理方法。

符号の説明

[0149] １０　情報処理端末

２０　情報処理サーバ

３０　ネットワーク

１１０　表示部

１２０　音声出力部

１３０　音声入力部

１４０　撮像部

１５０　センサ入力部

１６０　制御部

１７０　サーバ通信部

２１０　通信部

２２０　入力解析部

２３０　学習部

２４０　記憶部

２５０　出力制御部

請求の範囲

[請求項1] ユーザに対する情報提示を制御する出力制御部、

を備え、

前記出力制御部は、提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち少なくとも一方を用いて実行させる、

情報処理装置。

[請求項2] 前記出力制御部は、入力された音声または画像の少なくとも一方の入力情報に基づき、前記特有表現を前記提示される情報に適用可能か否かを判定する、請求項１に記載の情報処理装置。

[請求項3] 前記出力制御部は、前記画像に基づき認識された認識対象に適用可能な前記特有表現が存在する場合、前記提示される情報に前記特有表現を適用可能と判定する、請求項２に記載の情報処理装置。

[請求項4] 前記出力制御部は、前記音声に基づき認識された言語表現に適用可能な前記特有表現が存在する場合、前記提示される情報に前記特有表現を適用可能と判定する、請求項２に記載の情報処理装置。

[請求項5] 前記出力制御部は、前記特有表現と対応する前記画像が登録されている場合、前記音声及び前記画像を用いて前記情報提示を実行させる、請求項１に記載の情報処理装置。

[請求項6] 前記出力制御部は、検出された複数のユーザに基づき、前記提示される情報に前記特有表現を適用しないと判定する場合、一般表現を用いて前記情報提示を実行させる、請求項２に記載の情報処理装置。

[請求項7] 前記出力制御部は、前記複数のユーザから前記集団に属さないユーザが識別された場合、前記提示される情報に前記特有表現を適用しないと判定し、一般表現を用いて前記情報提示を実行させる、請求項６に記載の情報処理装置。

[請求項8] 前記出力制御部は、前記ユーザが前記特有表現を理解していないこ

とが推定された場合、前記提示される情報に前記特有表現を適用しないと判定し、一般表現を用いた前記情報提示を実行させる、請求項６に記載の情報処理装置。

[請求項9] 認識対象と前記認識対象に係る言語表現とを対応付けて学習する学習部、

を備え、

前記言語表現は、ユーザを含む集団内における特有表現を含み、

前記学習部は、収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習する、

情報処理装置。

[請求項10] 前記特有表現は、物体に対する呼称を含み、

前記学習部は、前記画像に基づいて認識される前記物体と前記呼称とを対応付けて学習する、請求項９に記載の情報処理装置。

[請求項11] 前記学習部は、前記画像から認識される物体と、前記音声から認識された前記物体に対して前記ユーザが用いた特有表現とを対応付けて学習する、請求項９に記載の情報処理装置。

[請求項12] 前記学習部は、前記画像から認識される物体と、前記画像から推定される前記物体の用途に対応する前記呼称とを対応付けて学習する、請求項１０に記載の情報処理装置。

[請求項13] 前記学習部は、前記画像から認識される物体と、ユーザの会話から認識された前記物体に対して前記ユーザが用いた特有表現とを対応付けて学習する、請求項９に記載の情報処理装置。

[請求項14] 前記学習部は、前記音声から認識される前記特有表現と、前記ユーザに関する実績情報に基づき認識される物体とを対応付けて学習する、請求項９に記載の情報処理装置。

[請求項15] 前記学習部は、発話者が発話対象者に用いた特有表現と、前記発話対象者とを対応付けて学習する、請求項９に記載の情報処理装置。

[請求項16] 前記学習部は、前記発話対象者の特徴に応じて、前記発話対象者と

前記特有表現の対応付けを再評価する、請求項１５に記載の情報処理装置。

[請求項17] 前記学習部は、前記認識対象を認識した回数に基づき、前記認識対象に用いる前記特有表現の信頼度を設定する、請求項９に記載の情報処理装置。

[請求項18] 前記特有表現は、呼称、形容表現、または動作表現の少なくともいずれか１つを含む、請求項９に記載の情報処理装置。

[請求項19] ユーザに対する情報提示を制御することと、

提示される情報の内容に関し、前記ユーザを含む集団内における特有表現を適用可能な場合、前記特有表現を含む情報提示を、音声または画像のうち少なくとも一方を用いて実行させることと、

を含む、プロセッサにより実行される、情報処理方法。

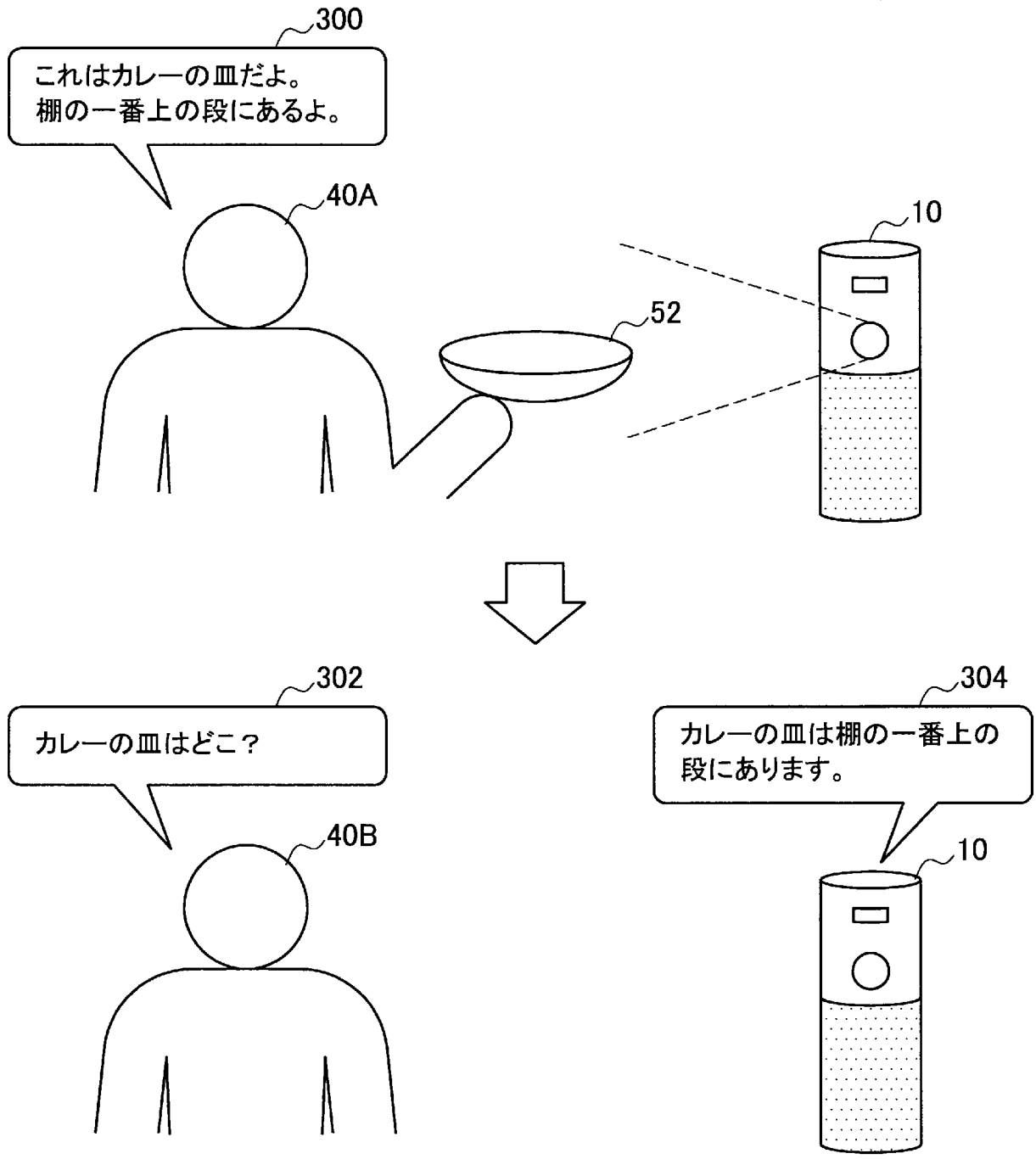
[請求項20] 認識対象と前記認識対象に係る言語表現とを対応付けて学習することと、

前記言語表現は、ユーザを含む集団内における特有表現を含み、

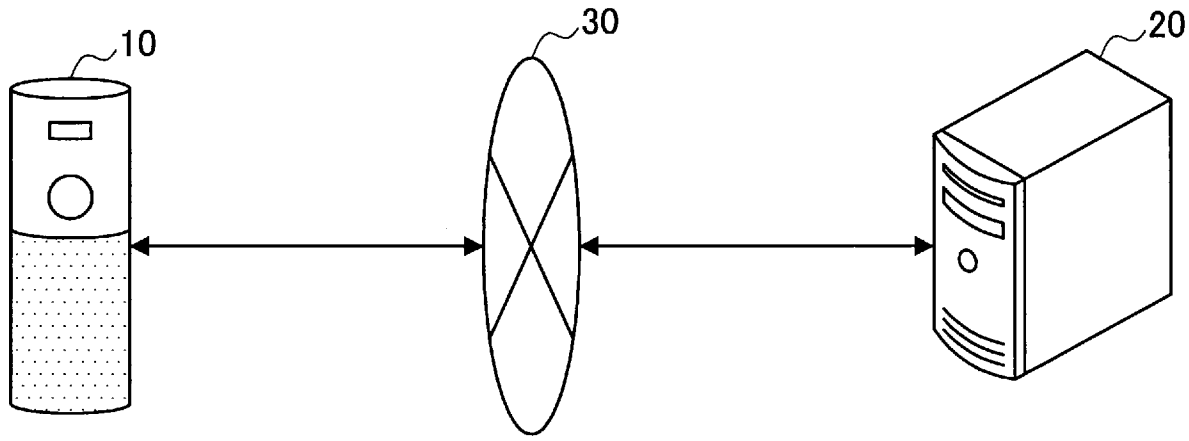
収集された音声または画像のうち少なくとも一方に基づいて、前記特有表現を学習することと、

を含む、プロセッサにより実行される、情報処理方法。

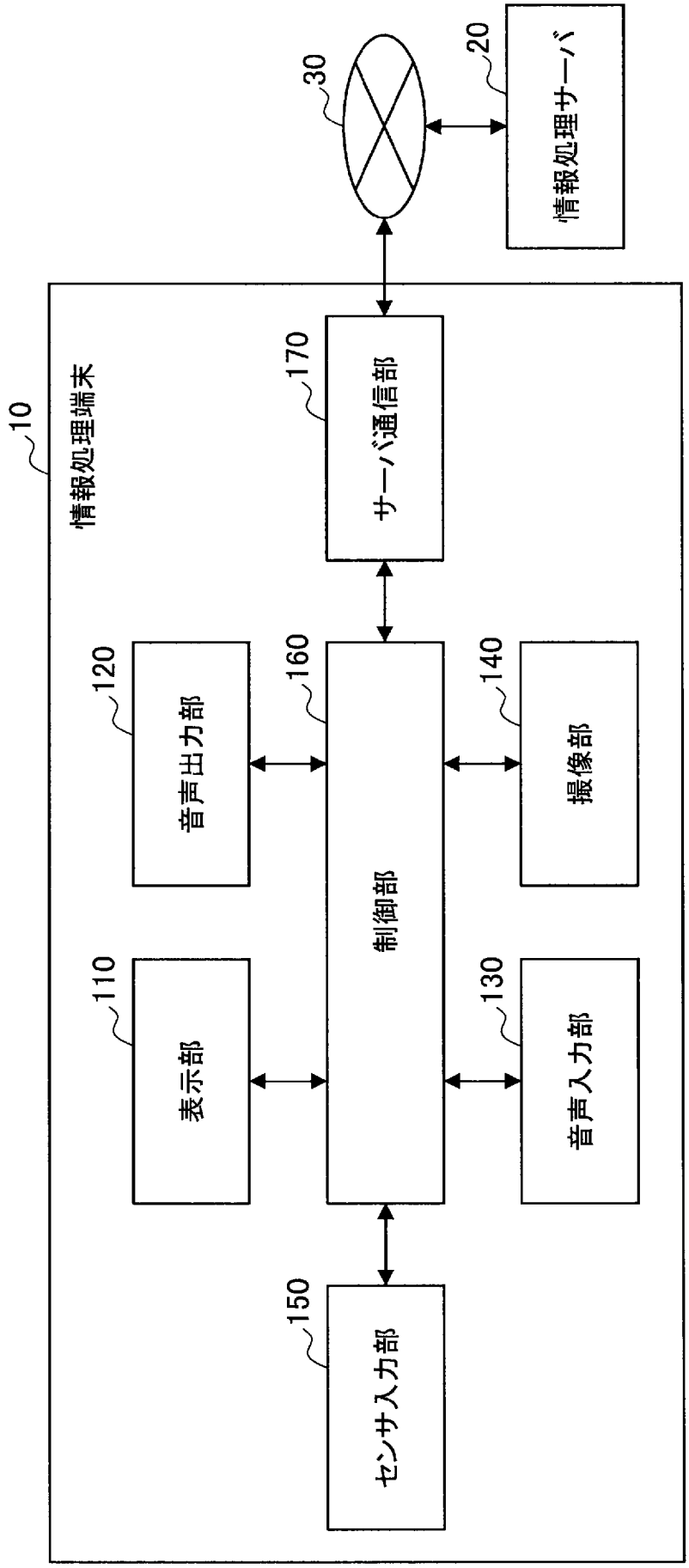
[図1]


300
これはカレーの皿だよ。
棚の一番上の段にあるよ。
40A
52
10
302
カレーの皿はどこ？
40B
304
カレーの皿は棚の一番上の段にあります。
10

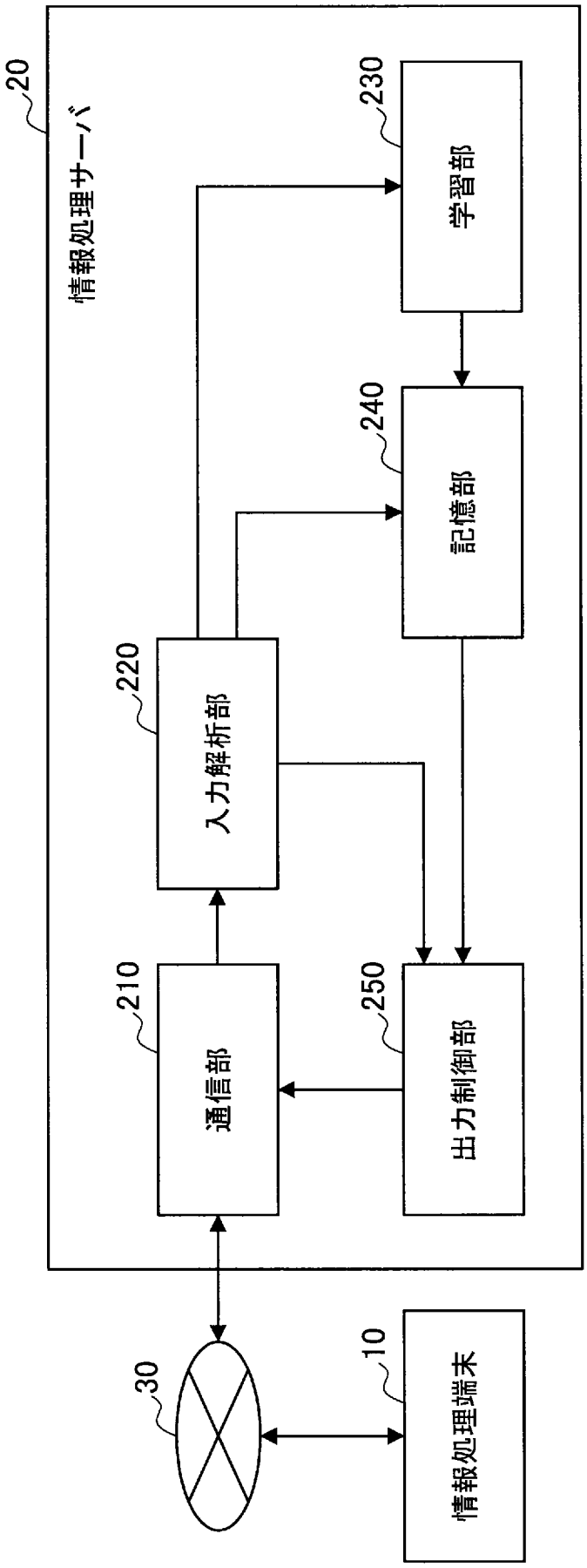
[図2]



[図3]


10
情報処理端末
170
サーバ通信部
120
音声出力部
160
制御部
140
撮像部
110
表示部
130
音声入力部
150
センサ入力部
30
20
情報処理サーバ

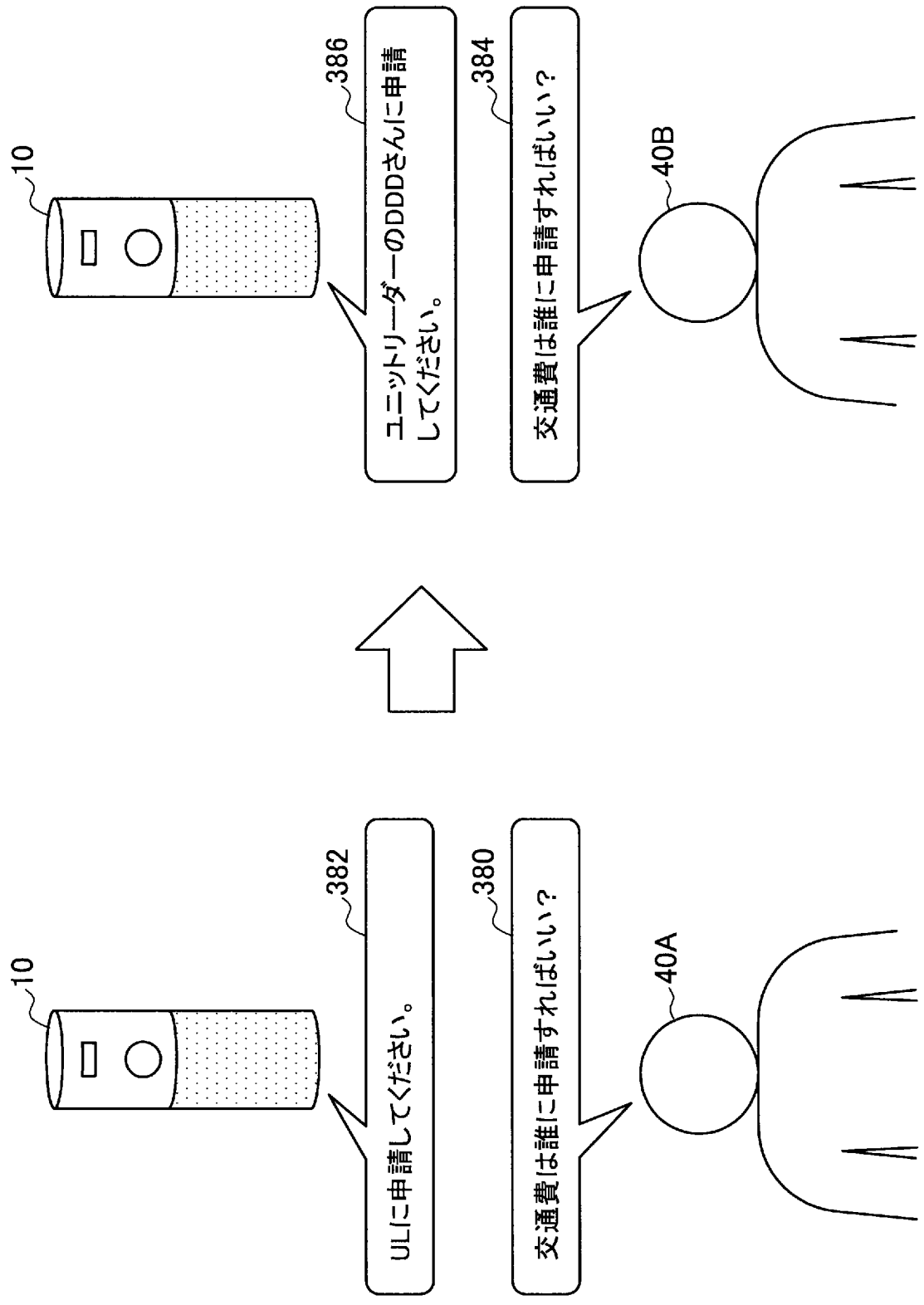
[図4]


20
情報処理サーバ
230
学習部
240
記憶部
220
入力解析部
210
通信部
250
出力制御部
30
10
情報処理端末

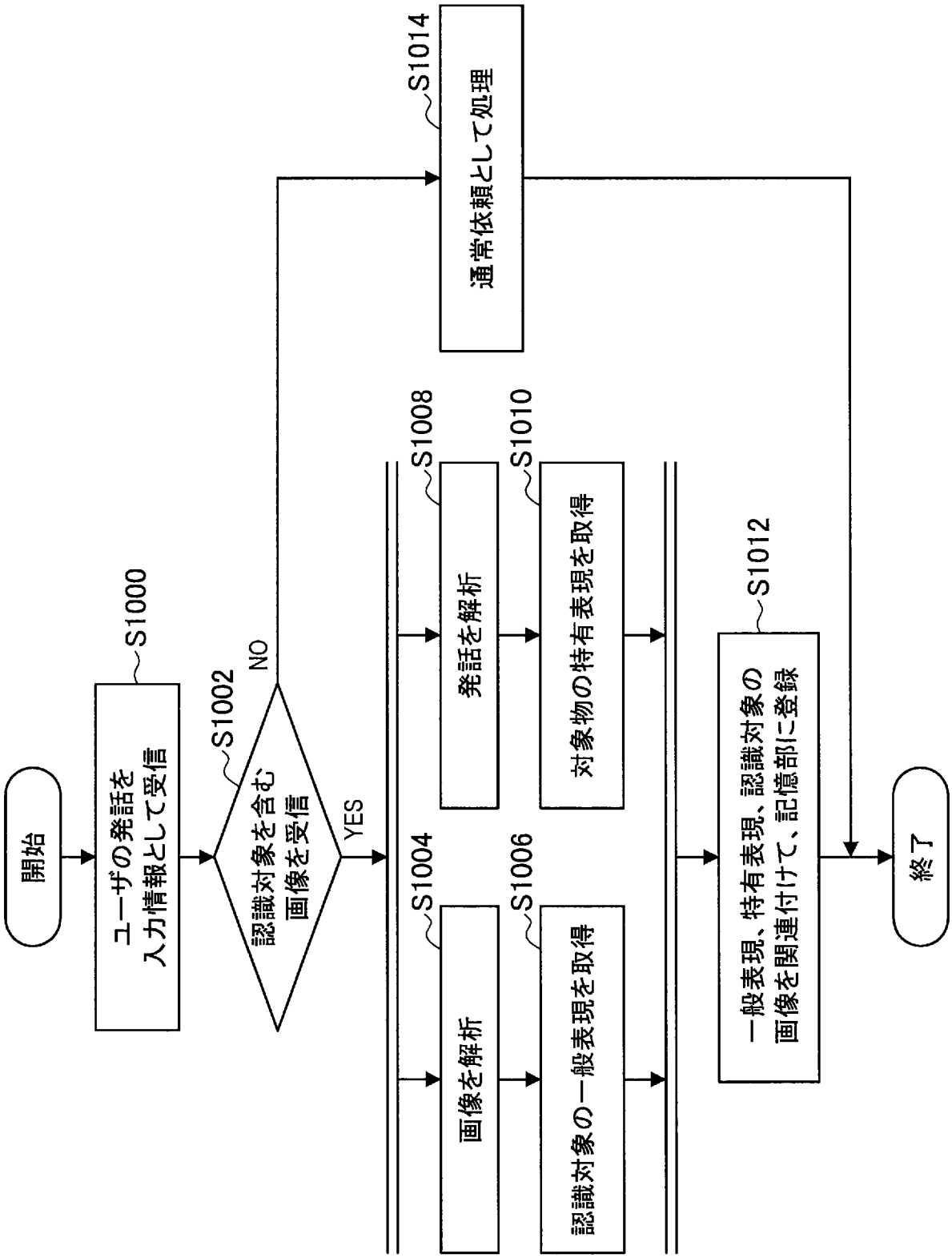
[図5]

一般表現	特有表現	対象ユーザ	回数	画像
皿	カレーの皿	AAA, BBB, CCC	5	image1
ゆうた	ゆうちゃん	AAA, BBB	—	image2
リモートコントローラ	チャンネル	AAA, BBB	—	image3
リモートコントローラ	ピッピ	CCC	—	image3

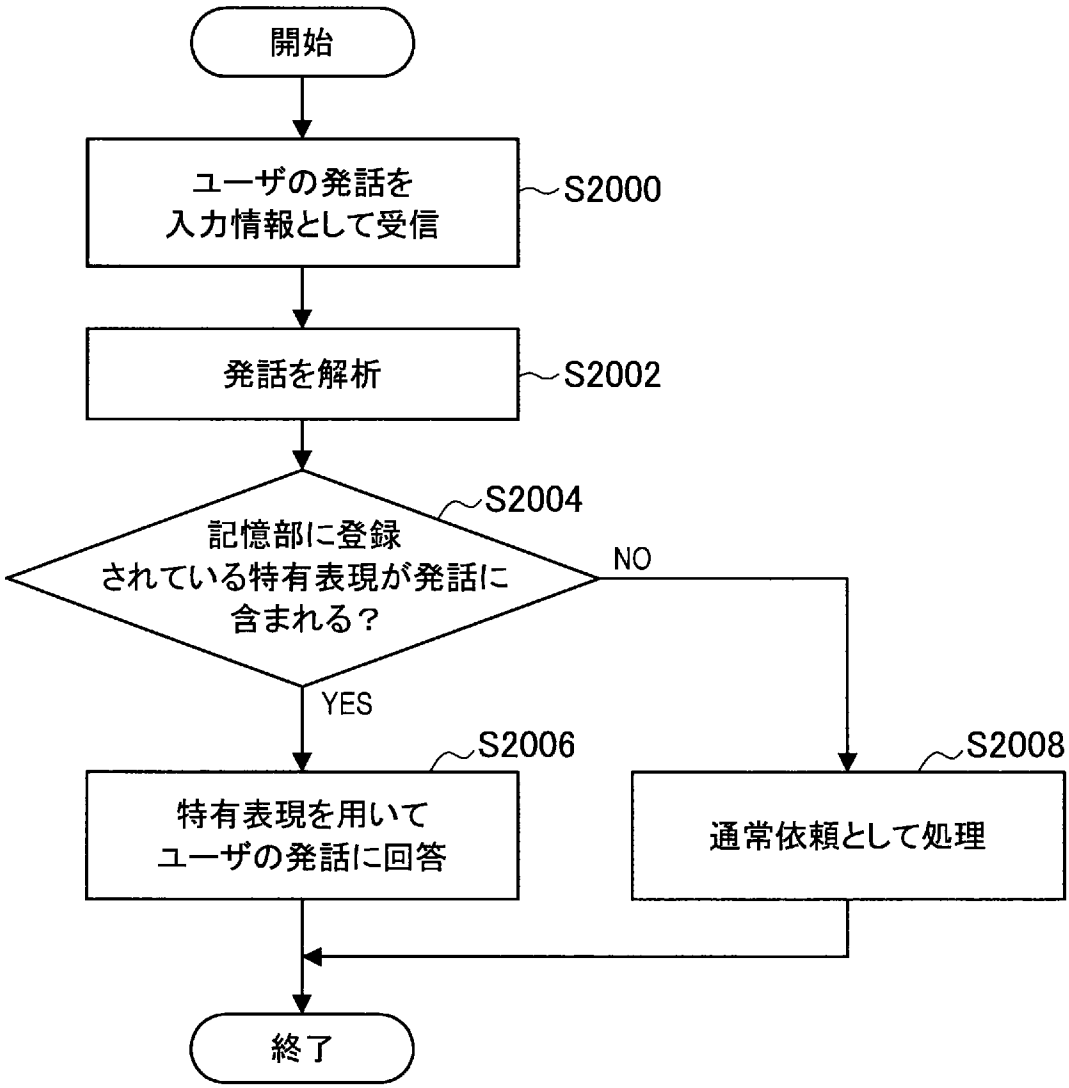
[図6]


10
382
ULに申請してください。
380
交通費は誰に申請すればいい？
40A
10
386
ユニットリーダーのDDDさんに申請してください。
384
交通費は誰に申請すればいい？
40B

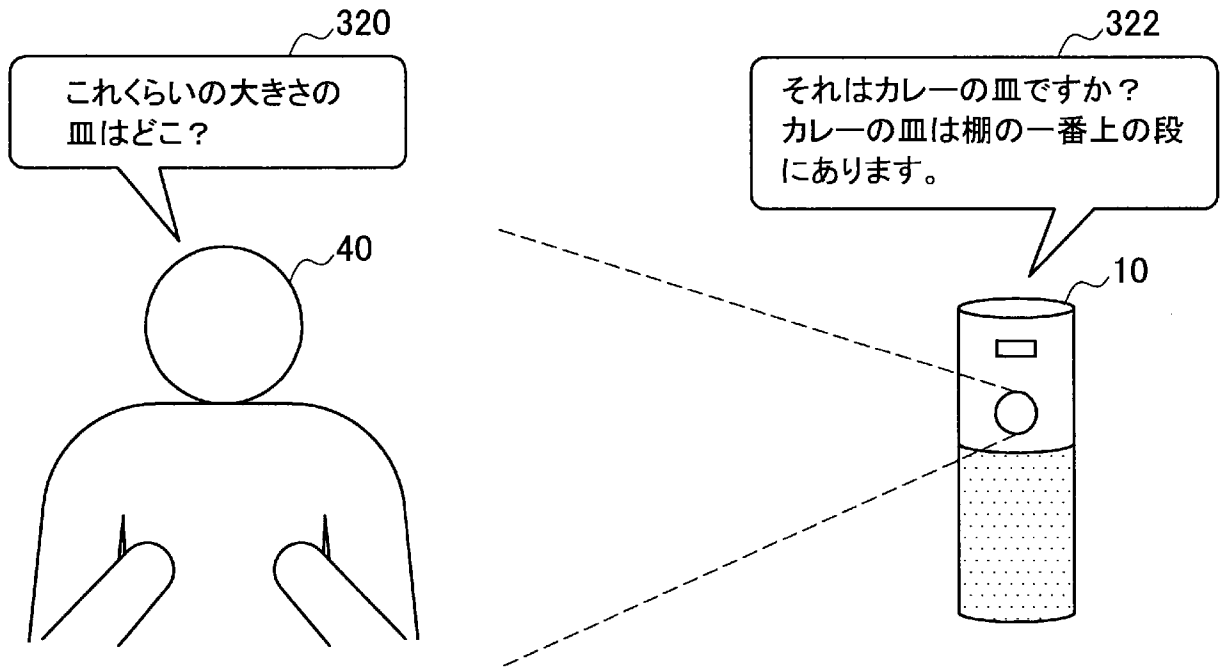
[図7]


開始
ユーザの発話を
入力情報として受信
S1000
S1002
認識対象を含む
画像を受信
NO
YES
S1014
通常依頼として処理
S1004
画像を解析
S1006
認識対象の一般表現を取得
S1008
発話を解析
S1010
対象物の特有表現を取得
S1012
一般表現、特有表現、認識対象の
画像を関連付けて、記憶部に登録
終了

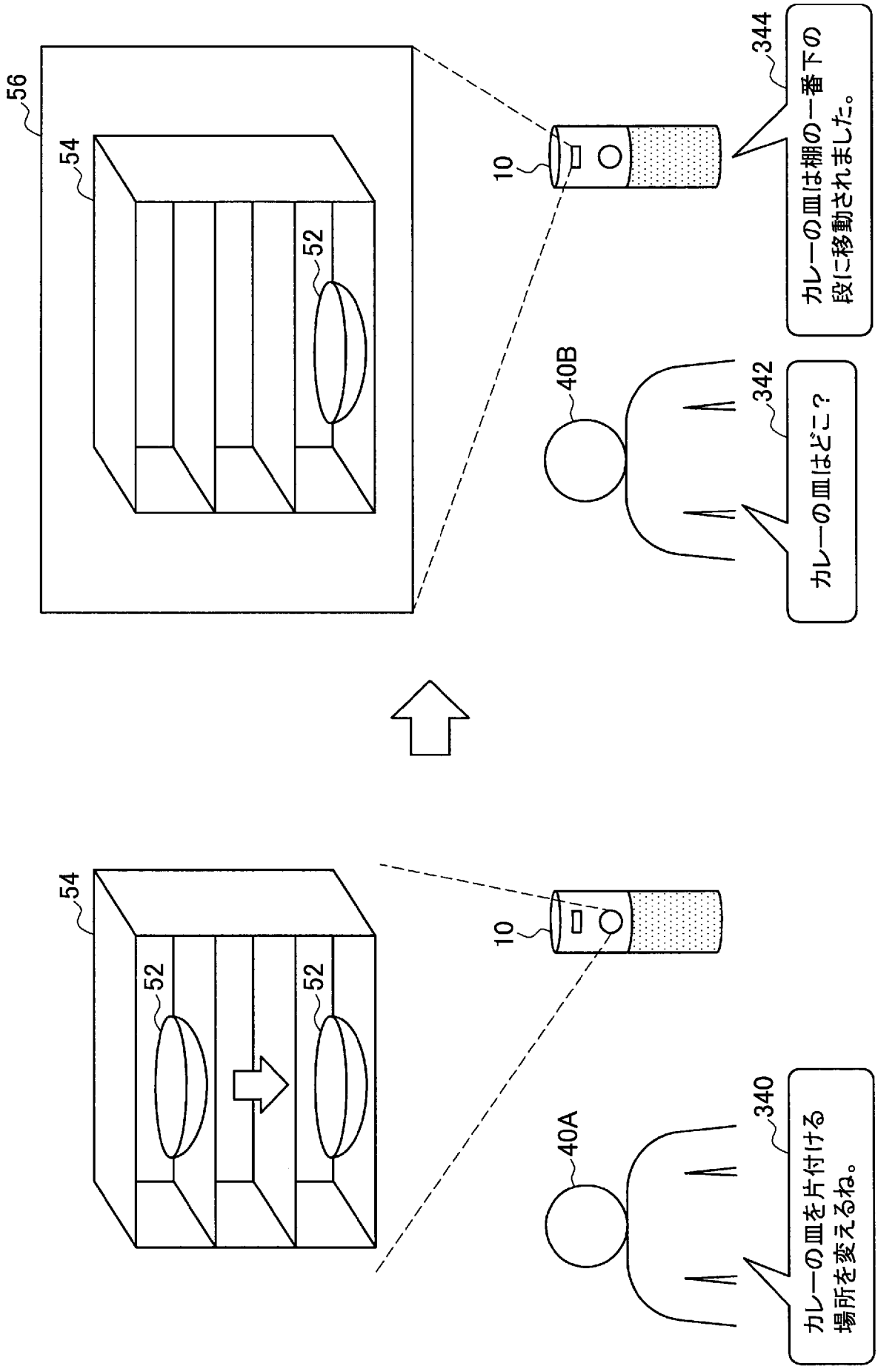
[図8]


開始
ユーザの発話を
入力情報として受信
S2000
発話を解析
S2002
S2004
記憶部に登録
されている特有表現が発話に
含まれる？
NO
YES
S2006
特有表現を用いて
ユーザの発話に回答
S2008
通常依頼として処理
終了

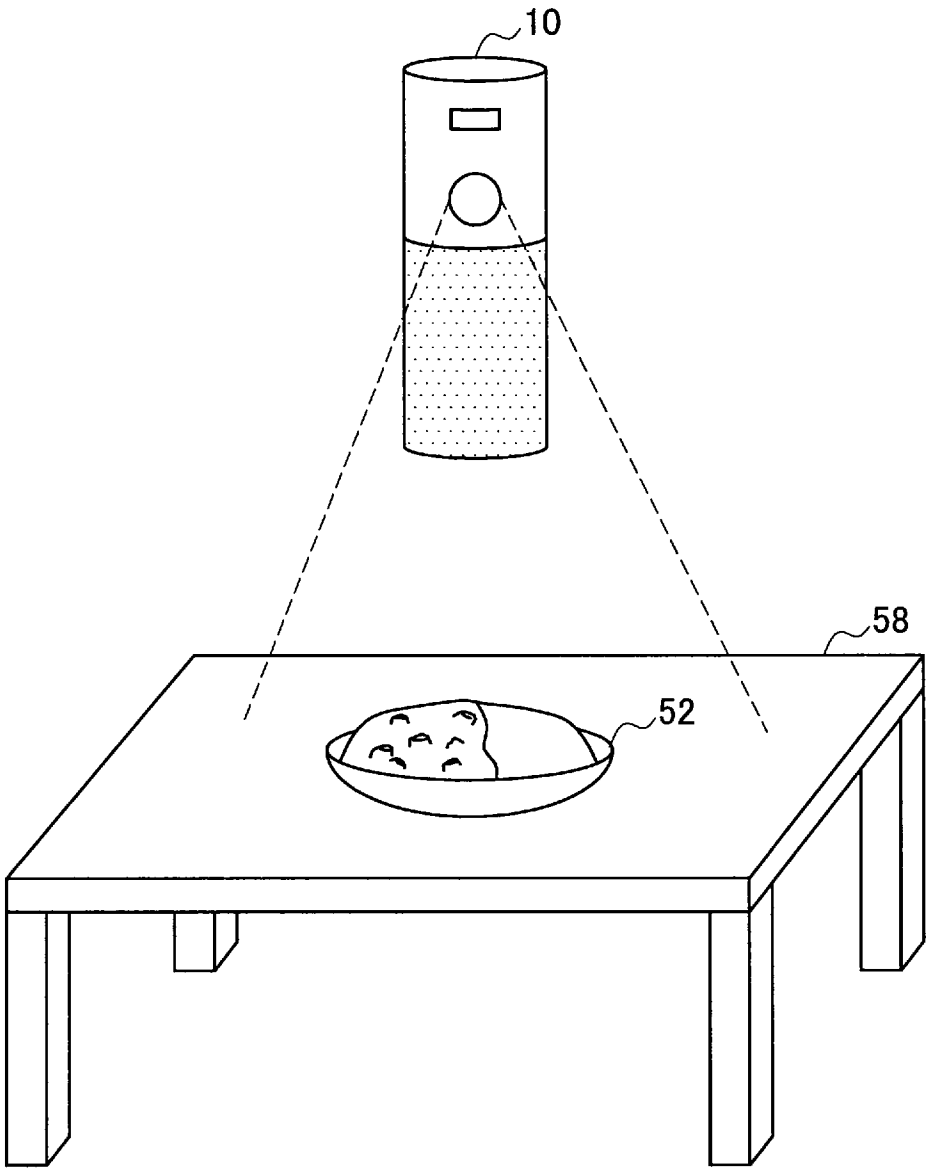
[図9]


320
これくらいの大きさの
皿はどこ？
40
322
それはカレーの皿ですか？
カレーの皿は棚の一番上の段
にあります。
10

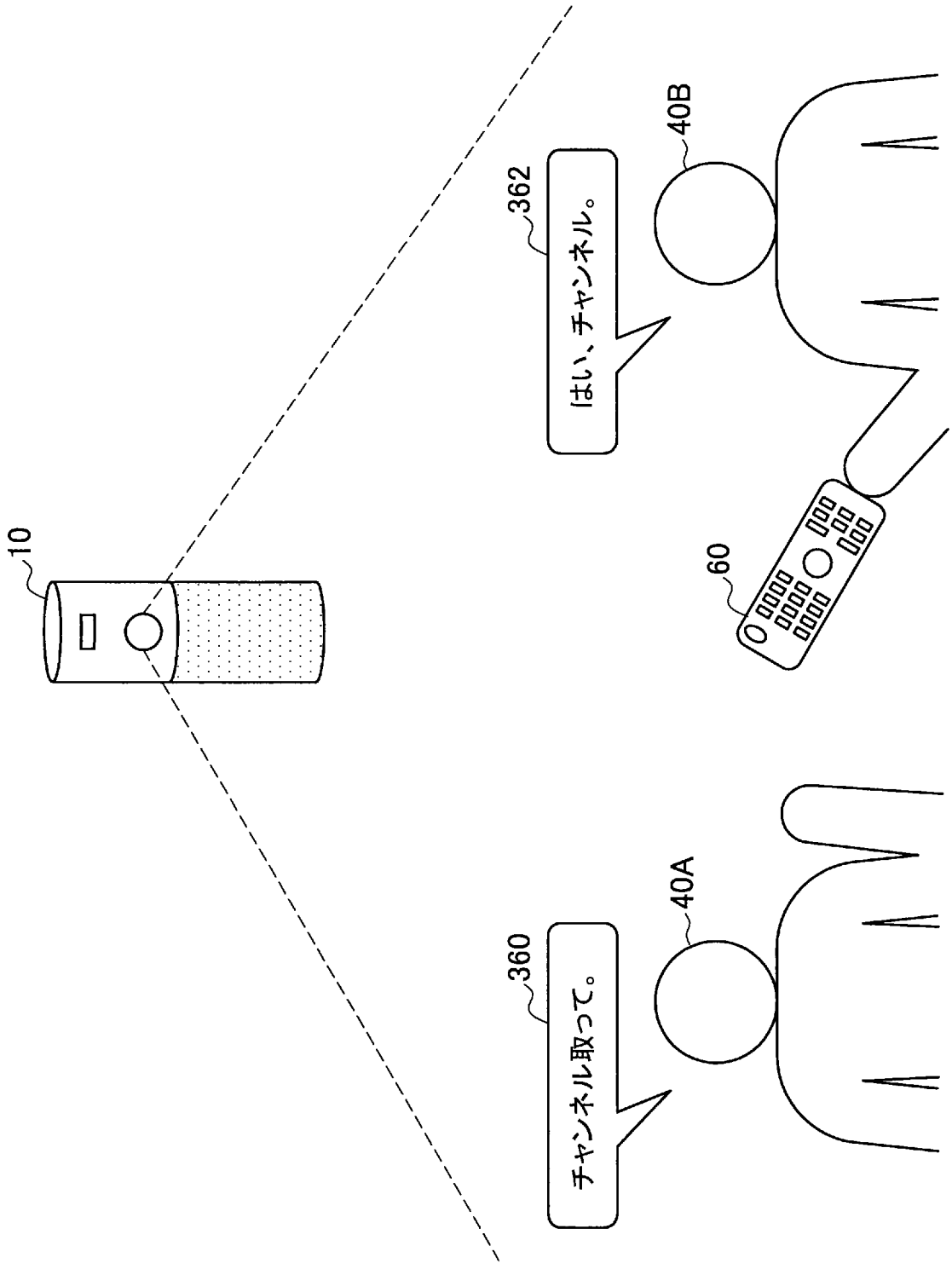
[図10]


56
54
52
10
344
カレーの皿は棚の一番下の段に移動されました。
40B
342
カレーの皿はどこ？
54
52
52
10
40A
340
カレーの皿を片付ける場所を変えるね。

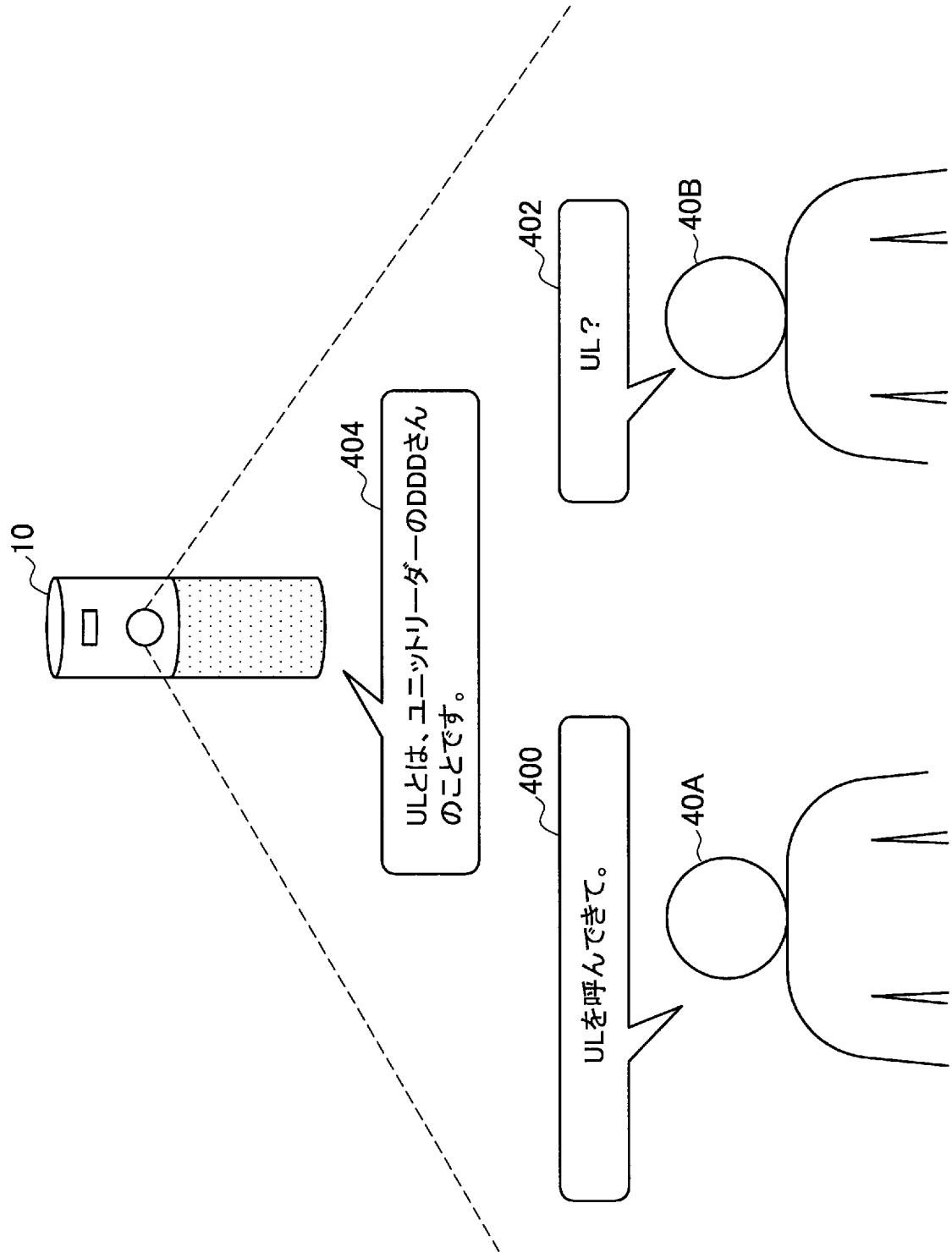
[図11]



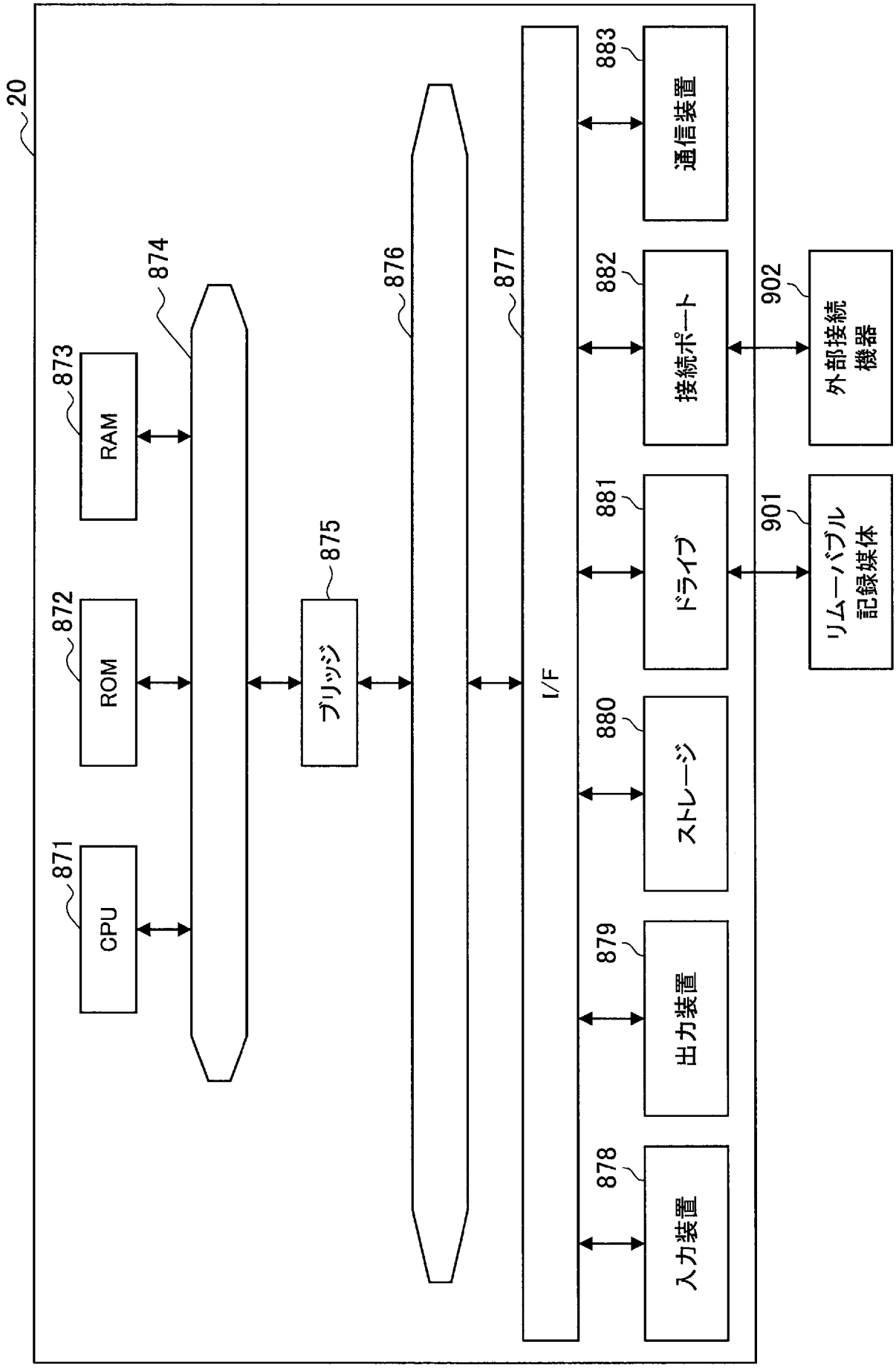
[図12]



[図13]


10
404
ULとは、ユニットリーダーのDDDさんのことです。
402
UL？
40B
400
ULを呼んできて。
40A

[図14]


20
CPU
871
ROM
872
RAM
873
874
ブリッジ
875
876
877
I/F
ストレージ
880
ドライブ
881
接続ポート
882
通信装置
883
出力装置
879
入力装置
878
リムーバブル記録媒体
901
外部接続機器
902

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/040151

A. CLASSIFICATION OF SUBJECT MATTER
Int. Cl. G06F3/16(2006.01)i, G06F3/01(2006.01)i, G10L13/00(2006.01)i, G10L13/08(2013.01)i, G10L15/22(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int. Cl. G06F3/16, G06F3/01, G10L13/00, G10L13/08, G10L15/22

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2018
Registered utility model specifications of Japan 1996-2018
Published registered utility model applications of Japan 1994-2018

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP 2010-282199 A (HONDA MOTOR CO., LTD.) 16 December 2010, paragraphs [0002]-[0007], [0056]-[0069] & US 2010/0332231 A1, paragraphs [0003]-[0022], [0115]-[0151]	1-5, 9-15, 18-20 6-8, 16-17
Y	WO 2016/136062 A1 (SONY CORP.) 01 September 2016, paragraphs [0191]-[0201] & US 2017/0337921 A1, paragraphs [0216]-[0226] & EP 3264258 A1	6-8
Y	JP 2004-252111 A (SONY CORP.) 09 September 2004, paragraphs [0036], [0054]-[0058], [0063], [0064] & US 2004/0215463 A1, paragraphs [0054], [0072]-[0076], [0081], [0082]	16

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26.12.2018	15.01.2019
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2018/040151

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2003-189148 A (KONICA CORP.) 04 July 2003, paragraphs [0001], [0047] (Family: none)	17

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１８／０４０１５１

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　G06F3/16(2006.01)i, G06F3/01(2006.01)i, G10L13/00(2006.01)i, G10L13/08(2013.01)i, G10L15/22(2006.01)i

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　G06F3/16, G06F3/01, G10L13/00, G10L13/08, G10L15/22

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１８年
日本国実用新案登録公報	１９９６－２０１８年
日本国登録実用新案公報	１９９４－２０１８年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

Ｃ．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	JP 2010-282199 A（本田技研工業株式会社）2010.12.16, 段落［０００２］－［０００７］，［００５６］－［００６９］ & US 2010/0332231 A1,	1-5, 9-15, 18-20
Y	段落［０００３］－［００２２］，［０１１５］－［０１５１］	6-8, 16-17

☑　Ｃ欄の続きにも文献が列挙されている。

☐　パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
２６．１２．２０１８	１５．０１．２０１９

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
東京都千代田区霞が関三丁目４番３号

特許庁審査官（権限のある職員）　５Ｅ　３９７９
円子　英紀
電話番号　０３－３５８１－１１０１　内線　３５２１

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
Y	WO 2016/136062 A1（ソニー株式会社）2016.09.01, 段落［０１９１］－［０２０１］ & US 2017/0337921 A1, 段落［０２１６］－［０２２６］ & EP 3264258 A1	6-8
Y	JP 2004-252111 A（ソニー株式会社）2004.09.09, 段落［００３６］，［００５４］－［００５８］， ［００６３］－［００６４］ & US 2004/0215463 A1, 段落［００５４］，［００７２］－［００７６］， ［００８１］－［００８２］	16
Y	JP 2003-189148 A（コニカ株式会社）2003.07.04, 段落［０００１］，［００４７］（ファミリーなし）	17