

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5194314号
(P5194314)

(45) 発行日 平成25年5月8日(2013.5.8)

(24) 登録日 平成25年2月15日(2013.2.15)

(51) Int.Cl.	F I
G 1 0 L 15/22 (2006.01)	G 1 0 L 15/22 2 0 0 H
G 1 0 L 15/00 (2013.01)	G 1 0 L 15/00 2 0 0 Z
	G 1 0 L 15/22 2 0 0 V
	G 1 0 L 15/22 4 6 0 D

請求項の数 6 (全 32 頁)

(21) 出願番号	特願2008-69606 (P2008-69606)	(73) 特許権者	393031586
(22) 出願日	平成20年3月18日(2008.3.18)		株式会社国際電気通信基礎技術研究所
(65) 公開番号	特開2009-223171 (P2009-223171A)		京都府相楽郡精華町光台二丁目2番地2
(43) 公開日	平成21年10月1日(2009.10.1)	(74) 代理人	100090181
審査請求日	平成23年2月14日(2011.2.14)		弁理士 山田 義人
		(72) 発明者	篠沢 一彦
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内
		(72) 発明者	宮下 敬宏
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内
		(72) 発明者	光永 法明
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 コミュニケーションシステム

(57) 【特許請求の範囲】

【請求項 1】

音声認識によって人間が指示する物品を特定し、当該特定した物品が当該人間の指示したものであるかを音声にて確認するコミュニケーションシステムであって、

物品に関連する単語と当該単語の音素記号列とを対応つけて記憶する第1記憶手段、

物品の識別情報と当該物品に関連する単語とを対応つけて記憶する第2記憶手段、

音声認識により生成した音素記号列を前記第1記憶手段に参照して物品に関連する単語を特定する単語特定手段、

前記単語特定手段が特定した物品に関する単語を前記第2記憶手段に参照して物品の識別情報を特定する物品特定手段、

前記物品特定手段が特定した物品の識別情報と対応つけて前記第2記憶手段に記憶された物品に関連する単語から前記人間の近傍に存在する他の物品に関連付けられていない所定の単語を選出する単語選出手段、および

前記単語選出手段が選出した単語を用いて前記単語特定手段が特定した物品に関する単語が前記人間の指示した物品であるか否かを確認する音声出力する音声出力手段を備える、コミュニケーションシステム。

【請求項 2】

前記所定の単語は、前記第2記憶手段において前記物品特定手段が特定した物品の識別情報に対応つけて記憶されている物品に関連する単語から、前記第2記憶手段において前記物品特定手段が特定した物品以外であって前記人間の近傍に存在する物品の識別情報に

対応つけて記憶されている物品に関する単語を除いたものである、請求項 1 記載のコミュニケーションシステム。

【請求項 3】

前記音声出力手段は、前記所定の単語が存在しない場合に、指示代名詞および指示形容詞の一方と、前記物品特定手段が特定した物品の識別情報と対応つけて前記第 2 記憶手段に記憶された物品に関連する単語のうちのいずれかの単語とを用いて音声を出力する、請求項 2 記載のコミュニケーションシステム。

【請求項 4】

前記所定の単語は、音声認識による認識の成功率が高い単語である、請求項 2 または 3 記載のコミュニケーションシステム。

10

【請求項 5】

前記認識の成功率の高い単語は、音声認識により 6 個以上の音素に分解される単語である、請求項 4 記載のコミュニケーションシステム。

【請求項 6】

単語と当該単語の音声認識による認識の成功率とを対応つけて記憶する第 3 記憶手段を備え、

前記単語選出手段は、前記第 3 記憶手段への参照に基づいて前記認識の成功率の高い単語を決定する、請求項 4 または 5 記載のコミュニケーションシステム。

【発明の詳細な説明】

【技術分野】

20

【0001】

この発明はコミュニケーションシステムに関し、特にたとえば、音声認識によって人間が指示する物品を特定し、当該特定した物品が当該人間の指示したものであるかを音声にて確認するコミュニケーションシステムに関する。

【背景技術】

【0002】

この種のコミュニケーションシステムの例が開示された特許文献は見当たらないが、音声認識は人間と装置をつなぐインターフェースとしてさまざまな装置に採用されており、音声認識を採用した装置には、たとえば、特許文献 1 に開示された片付けロボットがある。

30

【0003】

特許文献 1 によると、この片付けロボットは、人間から受け取った物品の片付けや片付けた物品の人間への返却を行う。ロボットは、人間から片付ける物品を受け取ると、その物品の写真を撮って記憶してから物品を片付ける。そして、後に、人間が「この前預けた物をもってきて」などと指示すると、ロボットは、この指示を音声認識して理解し、物品を預かったときに撮影してデータベースに記憶されている複数の物品の写真を返却する物品の候補として表示する。そして、人間は、表示された写真に添付された数字を発声して返却してほしい物品をロボットに伝える。ロボットは、人間の発声した数字を音声認識し、認識された数字が添付された写真から返却する物品を特定する。

【特許文献 1】特開 2007-152443 号公報 [B25J 13/00]

40

【発明の開示】

【発明が解決しようとする課題】

【0004】

特許文献 1 に記載の片付けロボットは、人間による返却してほしい物品の指示を、写真の提示と写真の選択によって行っているが、インターフェースの簡便化のためこの物品の指示をも音声認識で実現することが考えられる。音声によって物品を指示する場合、1 つの物品を指示するための言葉（単語）は複数存在する。

【0005】

たとえば、一冊の本を指示する場合であれば、本のタイトル名に基づいて指示する場合、本の著者名に基づいて指示する場合、本の出版者名に基づいて指示する場合、本の色に

50

基づいて指示する場合、本が厚いか薄いかに基づいて指示する場合、ハードカバーであるかソフトカバーであるかに基づいて指示する場合などである。

【0006】

このように物品の指示の受け付けを音声認識で行う場合には、物品を指示するための言葉（単語）が複数存在するため、人間がどの言葉（単語）を使って物品を指示するかが、ロボットが正しく物品を特定できるか否かに影響を与える。ロボットが正しく物品を特定できるか否かに影響を与える要因としては、たとえば、個々の言葉（単語）の音声認識のしやすさの違いがあげられる。また、個々の言葉（単語）によって物品を1つに特定できるか否かという問題もある。人間が「赤色の本」と指示した場合にはロボットは人間が指示する本を一冊に特定できない可能性がある。

10

【0007】

それゆえに、この発明の主たる目的は、新規なコミュニケーションシステムを提供することである。

【0008】

また、この発明の他の目的は、システムが人間の指示する物品を特定しやすい言葉を用いて、人間に物品を指示させることができる、コミュニケーションシステムを提供することである。

【課題を解決するための手段】

【0009】

本発明は、上記の課題を解決するために、以下の構成を採用した。なお、括弧内の参照符号および補足説明等は、本発明の理解を助けるために後述する実施の形態との対応関係を示したものであって、本発明を何ら限定するものではない。

20

【0010】

第1の発明は、音声認識によって人間が指示する物品を特定し、当該特定した物品が当該人間の指示したものであるかを音声にて確認するコミュニケーションシステムであって、物品に関連する単語と当該単語の音素記号列とを対応つけて記憶する第1記憶手段、物品の識別情報と当該物品に関連する単語とを対応つけて記憶する第2記憶手段、音声認識により生成した音素記号列を第1記憶手段に参照して物品に関連する単語を特定する単語特定手段、単語特定手段が特定した物品に関する単語を第2記憶手段に参照して物品の識別情報を特定する物品特定手段、物品特定手段が特定した物品の識別情報と対応つけて第2記憶手段に記憶された物品に関連する単語から人間の近傍に存在する他の物品に関連付けられていない所定の単語を選出する単語選出手段、および単語選出手段が選出した単語を用いて単語特定手段が特定した物品に関する単語が人間の指示した物品であるか否かを確認する音声出力する音声出力手段を備える、コミュニケーションシステム。

30

【0011】

第1の発明では、コミュニケーションシステム(10)は、音声認識によって人間が指示する物品を特定し、当該特定した物品が当該人間の指示したものであるかを音声にて確認する。第1記憶手段(126、205B)は物品に関連する単語と当該単語の音素記号列とを対応つけて記憶する。第2記憶手段(122、205A)は物品の識別情報と当該物品に関連する単語とを対応つけて記憶する。単語特定手段(200、S111)は音声認識により生成した音素記号列を第1記憶手段に参照して物品に関連する単語を特定する。物品特定手段(200、S125)は単語特定手段が特定した物品に関する単語を第2記憶手段に参照して物品の識別情報を特定する。単語選出手段(200、300、S183、S201)は物品特定手段が特定した物品の識別情報と対応つけて第2記憶手段に記憶された物品に関連する単語から人間の近傍に存在する他の物品に関連付けられていない所定の単語を選出する。そして、音声出力手段(64、80、90)は単語選出手段が選出した単語を用いて単語特定手段が特定した物品に関する単語が人間の指示した物品であるか否かを確認する音声出力する。

40

【0012】

第1の発明によれば、人間が指示した物品を確認する際に、コミュニケーションシステ

50

ムは当該物品に関連する単語を用いた音声を発する。人間はコミュニケーションシステムの真似をする傾向があるので、次回に、人間が当該物品を指示する際にはコミュニケーションシステムが確認の際の音声に用いた単語を使用することになる。当該単語は、人間の指示した当該物品に関連する単語であるので、コミュニケーションシステムは当該単語を用いた人間の物品の指示によって容易に物品を特定することができる。したがって、コミュニケーションシステムが人間の指示する物品を特定しやすい言葉（単語）を用いて、人間に物品を指示させることができる。

【0013】

第2の発明は、第1の発明に従属する発明であって、所定の単語は、第2記憶手段において物品特定手段が特定した物品の識別情報に対応つけて記憶されている物品に関連する単語から、第2記憶手段において物品特定手段が特定した物品以外であって人間の近傍に存在する物品の識別情報に対応つけて記憶されている物品に関する単語を除いたものである。

10

【0014】

第2の発明では、所定の単語は、第2記憶手段において物品特定手段が特定した物品の識別情報に対応つけて記憶されている物品に関連する単語から、第2記憶手段において物品特定手段が特定した物品以外であって人間の近傍に存在する物品の識別情報に対応つけて記憶されている物品に関する単語を除いたものである。

【0015】

第2の発明によれば、所定の単語は、人間の指示した物品に関連する単語であって、人間の近傍に存在するその他の物品に関連する単語ではない。したがって、コミュニケーションシステムが人間の指示した物品を確認するために所定の単語を使用すれば、次回は、人間が真似を行い、所定の単語を用いて物品を指示することが期待できる。これによって、コミュニケーションシステムが容易に人間の指示する物品を特定することができる。

20

【0016】

第3の発明は、第2の発明に従属する発明であって、音声出力手段は、所定の単語が存在しない場合に、指示代名詞および指示形容詞の一方と、物品特定手段が特定した物品の識別情報と対応つけて第2記憶手段に記憶された物品に関連する単語のうちのいずれかの単語とを用いて音声を出力する。

【0017】

第3の発明では、音声出力手段（64、80、90）は、所定の単語が存在しない場合に、指示代名詞および指示形容詞の一方と、物品特定手段が特定した物品の識別情報と対応つけて第2記憶手段に記憶された物品に関連する単語のうちのいずれかの単語とを用いて音声を出力する。

30

【0018】

第3の発明では、人間の指示する物品を一意に示すことができる所定の単語が存在しない場合でも、人間が指示した物品に関連する単語と指示代名詞および指示形容詞の一方とを用いた音声によって、人間に物品が人間の指示したものであるか否かを確認する。したがって、人間は音声によってコミュニケーションシステムが特定した物品を確認しやすくなる。

40

【0019】

第4の発明は、第2の発明または第3の発明に従属する発明であって、所定の単語は、音声認識による認識の成功率が高い単語である。

【0020】

第4の発明では、所定の単語は、音声認識による認識の成功率が高い単語である。

【0021】

第4の発明によれば、人間に音声認識による認識の成功率の高い単語を用いて物品を指示させることができる。

【0022】

第5の発明は、第4の発明に従属する発明であって、認識の成功率の高い単語は、音声

50

認識により 6 個以上の音素に分解される単語である。

【0023】

第 5 の発明では、認識の成功率の高い単語は、音声認識により 6 個以上の音素の分解される単語である。

【0024】

第 5 の発明によれば、音声認識による認識率が高いと推定される単語を用いて人間に物品を指示させることができる。

【0025】

第 6 の発明は、第 4 の発明または第 5 の発明に従属する発明であって、単語と当該単語の音声認識による認識の成功率とを対応つけて記憶する第 3 記憶手段を備え、単語選出手段は、第 3 記憶手段への参照に基づいて認識の成功率の高い単語を決定する。

10

【0026】

第 6 の発明では、コミュニケーションシステムは、単語と当該単語の音声認識による認識の成功率とを対応つけて記憶する第 3 記憶手段 (205F) を備える。そして、単語選出手段は第 3 記憶手段への参照に基づいて音声認識による認識の成功率の高い単語を決定する。

【0027】

第 6 の発明によれば、音声認識による認識率が実際に高い単語を用いて人間に物品を指示させることができる。

20

【発明の効果】

【0028】

この発明によれば、人間の指示する物品の確認に、コミュニケーションシステムが特定しやすい言葉 (単語) を用いることによって、次回に人間が指示する物品の特定を容易にできる。

【0029】

この発明の上述の目的、その他の目的、特徴および利点は、図面を参照して行う以下の実施例の詳細な説明から一層明らかとなろう。

【発明を実施するための最良の形態】

【0030】

図 1 を参照して、この実施例のコミュニケーションロボットシステム (以下、単に「システム」ということがある。) 10 は、コミュニケーションロボット (以下、単に「ロボット」ということがある。) 12 を含む。このロボット 12 は、たとえば無線 LAN などのネットワーク 14 にアクセスすることができる。ロボット 12 はサーバ 20 と協働して人間 16 が音声、視線、および指差しで指示する物品を特定し、たとえばその物品を人間 16 に持っていくなどの動作を実行する。

30

【0031】

人間 16 は、その人物が誰であることを示す無線タグ 18 を装着しているとともに、図示しないが、モーションキャプチャのためのマークが付着されている。マークは、典型的には、人間の頭頂、両肩、両肘、両手の人差し指の先端などに設定されていて、それらのマークが、人間 16 の全体とともに、サーバ 20 に制御されるカメラ 120 によって撮影される。カメラ 120 は、実施例では、3 つ設けられ、人間 16 を 3 方向から撮影し、そのカメラ映像をサーバ 20 に供給する。

40

【0032】

サーバ 20 は無線 LAN のようなネットワーク 14 に結合され、上述のようにして入力されるカメラ映像データに基づいて、マークの動きを検出するモーションキャプチャ処理を実行するとともに、たとえば肌色領域を検出することによって、人間 16 の顔の位置を特定することができる。

【0033】

このシステム 10 では、上述のように、ロボット 12 が人間 16 の指示する物品を対象物として特定するものである。対象物となり得る物品の例として、この実施例では、本 (

50

書籍) 24を用いる。本24には、その本がどのような本であることを示す無線タグ18が付着されている。本24は本棚26に収納される。

【0034】

ただし、対象物となり得る物品は実施例の書籍だけでなく、もし家庭用のシステムであれば、家庭内のあらゆる物品が考えられる。また、当然、家庭用としてだけではなく、人間と一緒に働く任意の場所(会社、事務所、工場など)での利用が考えられる。

【0035】

そして、このシステム10が対象とするすべての物品(本24)は、サーバ20に付設された物品辞書122に登録される。物品辞書122については後述する。

【0036】

また、システム10が対象とする人間16は、人間16に付着されている無線タグ18が複数存在するアンテナ124のいずれかを介して無線タグ読取装置208(図4参照)で読み取られることによって、その位置がサーバ20で把握される。つまり、人間16の位置は、人間16に付着されている無線タグ18からの電波をどのアンテナ124で受信したかによって若干大まかな位置が把握される。

【0037】

なお、図1では、簡単のため、1台のロボット12を示してあるが、2台以上であってよい。また、人間16は1人に限定される必要はなく、無線タグ18で識別できるので、複数であってよい。

【0038】

また、図1に示す実施例では、このシステム10を設置している空間のワールド座標を用いてロボット12、人間16、物品24などの位置が表現されていて、他方、ロボット12の制御はロボット座標で行なわれるので、詳細は説明しないが、ロボット12は、後述の処理における必要に応じて、ロボット座標とワールド座標との間の座標変換処理を実行するものである。

【0039】

図2を参照して、ロボット12のハードウェアの構成について説明する。また、図2はこの実施例のロボット12の外観を示す正面図である。ロボット12は台車30を含み、台車30の下面にはロボット12を自律移動させる2つの車輪32および1つの従輪34が設けられる。2つの車輪32は車輪モータ36(図3参照)によってそれぞれ独立に駆動され、台車30すなわちロボット12を前後左右の任意方向に動かすことができる。また、従輪34は車輪32を補助する補助輪である。したがって、ロボット12は、配置された空間内を自律制御によって移動可能である。

【0040】

台車30の上には、円柱形のセンサ取り付けパネル38が設けられ、このセンサ取り付けパネル38には、多数の赤外線距離センサ40が取り付けられる。これらの赤外線距離センサ40は、センサ取り付けパネル38すなわちロボット12の周囲の物体(人間や障害物など)との距離を測定するものである。

【0041】

なお、この実施例では、距離センサとして、赤外線距離センサを用いるようにしてあるが、赤外線距離センサに代えて、超音波距離センサやミリ波レーダなどを用いることもできる。

【0042】

センサ取り付けパネル38の上には、胴体42が直立するように設けられる。また、胴体42の前方中央上部(人の胸に相当する位置)には、上述した赤外線距離センサ40がさらに設けられ、ロボット12の前方の主として人間との距離を計測する。また、胴体42には、その側面側上端部のほぼ中央から伸びる支柱44が設けられ、支柱44の上には、全方位カメラ46が設けられる。全方位カメラ46は、ロボット12の周囲を撮影するものであり、後述する眼カメラ70とは区別される。この全方位カメラ46としては、たとえばCCDやCMOSのような固体撮像素子を用いるカメラを採用することができる。

10

20

30

40

50

なお、これら赤外線距離センサ４０および全方位カメラ４６の設置位置は、当該部位に限定されず適宜変更され得る。

【００４３】

胴体４２の両側面上端部（人の肩に相当する位置）には、それぞれ、肩関節４８Ｒおよび肩関節４８Ｌによって、上腕５０Ｒおよび上腕５０Ｌが設けられる。図示は省略するが、肩関節４８Ｒおよび肩関節４８Ｌは、それぞれ、直交する３軸の自由度を有する。すなわち、肩関節４８Ｒは、直交する３軸のそれぞれの軸廻りにおいて上腕５０Ｒの角度を制御できる。肩関節４８Ｒの或る軸（ヨー軸）は、上腕５０Ｒの長手方向（または軸）に平行な軸であり、他の２軸（ピッチ軸およびロール軸）は、その軸にそれぞれ異なる方向から直交する軸である。同様にして、肩関節４８Ｌは、直交する３軸のそれぞれの軸廻りにおいて上腕５０Ｌの角度を制御できる。肩関節４８Ｌの或る軸（ヨー軸）は、上腕５０Ｌの長手方向（または軸）に平行な軸であり、他の２軸（ピッチ軸およびロール軸）は、その軸にそれぞれ異なる方向から直交する軸である。

10

【００４４】

また、上腕５０Ｒおよび上腕５０Ｌのそれぞれの先端には、肘関節５２Ｒおよび肘関節５２Ｌが設けられる。図示は省略するが、肘関節５２Ｒおよび肘関節５２Ｌは、それぞれ１軸の自由度を有し、この軸（ピッチ軸）の軸回りにおいて前腕５４Ｒおよび前腕５４Ｌの角度を制御できる。

【００４５】

前腕５４Ｒおよび前腕５４Ｌのそれぞれの先端には、人の手に相当するハンド５６Ｒおよびハンド５６Ｌがそれぞれ設けられる。これらのハンド５６Ｒおよび５６Ｌは、詳細な図示は省略するが、開閉可能に構成され、それによってロボット１２は、ハンド５６Ｒおよび５６Ｌを用いて物体を把持または挟持することができる。ただし、ハンド５６Ｒ、５６Ｌの形状は実施例の形状に限らず、人間の手に酷似した形状や機能を持たせるようにしてもよい。

20

【００４６】

また、図示は省略するが、台車３０の前面、肩関節４８Ｒと肩関節４８Ｌとを含む肩に相当する部位、上腕５０Ｒ、上腕５０Ｌ、前腕５４Ｒ、前腕５４Ｌ、球体５６Ｒおよび球体５６Ｌには、それぞれ、接触センサ５８（図３で包括的に示す）が設けられる。台車３０の前面の接触センサ５８は、台車３０への人間や他の障害物の接触を検知する。したがって、ロボット１２は、その自身の移動中に障害物との接触が有ると、それを検知し、直ちに車輪３２の駆動を停止してロボット１２の移動を急停止させることができる。また、その他の接触センサ５８は、当該各部位に触れたかどうかを検知する。なお、接触センサ５８の設置位置は、当該部位に限定されず、適宜な位置（人の胸、腹、脇、背中および腰に相当する位置）に設けられてもよい。

30

【００４７】

胴体４２の中央上部（人の首に相当する位置）には首関節６０が設けられ、さらにその上には頭部６２が設けられる。図示は省略するが、首関節６０は、３軸の自由度を有し、３軸の各軸廻りに角度制御可能である。或る軸（ヨー軸）はロボット１２の真上（鉛直上向き）に向かう軸であり、他の２軸（ピッチ軸、ロール軸）は、それぞれ、それと異なる方向で直交する軸である。

40

【００４８】

頭部６２には、人の口に相当する位置に、スピーカ６４が設けられる。スピーカ６４は、ロボット１２が、その周辺の人間に対して音声ないし音によってコミュニケーションを取るために用いられる。また、人の耳に相当する位置には、マイク６６Ｒおよびマイク６６Ｌが設けられる。以下、右のマイク６６Ｒと左のマイク６６Ｌとをまとめてマイク６６とすることがある。マイク６６は、周囲の音、とりわけコミュニケーションを実行する対象である人間の音声を取り込む。さらに、人の目に相当する位置には、眼球部６８Ｒおよび眼球部６８Ｌが設けられる。眼球部６８Ｒおよび眼球部６８Ｌは、それぞれ眼カメラ７０Ｒおよび眼カメラ７０Ｌを含む。以下、右の眼球部６８Ｒと左の眼球部６８Ｌとをま

50

とめて眼球部 68 ということがある。また、右の眼カメラ 70 R と左の眼カメラ 70 L とをまとめて眼カメラ 70 ということがある。

【0049】

眼カメラ 70 は、ロボット 12 に接近した人間の顔や他の部分ないし物体などを撮影して、それに対応する映像信号を取り込む。この実施例では、ロボット 12 は、この眼カメラ 70 からの映像信号によって、人間 16 の左右両目のそれぞれの視線方向（ベクトル）を検出する。その視線検出方法は具体的には、2つのカメラを用いるものとして特開 2004 - 255074 号公報に、1つのカメラを用いるものとして特開 2006 - 172209 号公報や特開 2006 - 285531 号公報開示されるが、ここではその詳細は重要ではないので、これらの公開公報を引用するにとどめる。

10

【0050】

ただし、人間 16 の視線ベクトルの検出のためには、よく知られているアイマークレコーダなどが利用されてもよい。

【0051】

また、眼カメラ 70 は、上述した全方位カメラ 46 と同様のカメラを用いることができる。たとえば、眼カメラ 70 は、眼球部 68 内に固定され、眼球部 68 は、眼球支持部（図示せず）を介して頭部 62 内の所定位置に取り付けられる。図示は省略するが、眼球支持部は、2軸の自由度を有し、それらの各軸廻りに角度制御可能である。たとえば、この2軸の一方は、頭部 62 の上に向かう方向の軸（ヨー軸）であり、他方は、一方の軸に直交しかつ頭部 62 の正面側（顔）が向く方向に直行する方向の軸（ピッチ軸）である。眼球支持部がこの2軸の各軸廻りに回転されることによって、眼球部 68 ないし眼カメラ 70 の先端（正面）側が変位され、カメラ軸すなわち視線方向が移動される。なお、上述のスピーカ 64、マイク 66 および眼カメラ 70 の設置位置は、当該部位に限定されず、適宜な位置に設けられてよい。

20

【0052】

このように、この実施例のロボット 12 は、車輪 32 の独立 2 軸駆動、肩関節 48 の 3 自由度（左右で 6 自由度）、肘関節 52 の 1 自由度（左右で 2 自由度）、首関節 60 の 3 自由度および眼球支持部の 2 自由度（左右で 4 自由度）の合計 17 自由度を有する。

【0053】

図 3 はロボット 12 の電氣的な構成を示すブロック図である。この図 3 を参照して、ロボット 12 は、CPU 80 を含む。CPU 80 は、マイクロコンピュータ或いはプロセッサとも呼ばれ、バス 82 を介して、メモリ 84、モータ制御ボード 86、センサ入力／出力ボード 88 および音声入力／出力ボード 90 に接続される。

30

【0054】

メモリ 84 は、図示は省略をするが、ROM、HDD および RAM を含む。ROM および HDD には、ロボット 12 の動作を制御するための制御プログラムが予め記憶される。たとえば、各センサの出力（センサ情報）を検知するための検知プログラムや、外部コンピュータとの間で必要なデータやコマンドを送受信するための通信プログラムなどが記録される。また、RAM は、ワークメモリやバッファメモリとして用いられる。

【0055】

さらに、この実施例では、ロボット 12 は、人間 16 とのコミュニケーションをとるために発話したり、ジェスチャしたりできるように構成されているが、メモリ 84 に、このような発話やジェスチャのための発話／ジェスチャ辞書 85A が設定されている。

40

【0056】

モータ制御ボード 86 は、たとえば DSP で構成され、各腕や首関節および眼球部などの各軸モータの駆動を制御する。すなわち、モータ制御ボード 86 は、CPU 80 からの制御データを受け、右眼球部 68 R の 2 軸のそれぞれの角度を制御する 2 つのモータ（図 3 では、まとめて「右眼球モータ 92」と示す）の回転角度を制御する。同様にして、モータ制御ボード 86 は、CPU 80 からの制御データを受け、左眼球部 68 L の 2 軸のそれぞれの角度を制御する 2 つのモータ（図 3 では、まとめて「左眼球モータ 94」と示す

50

)の回転角度を制御する。

【0057】

また、モータ制御ボード86は、CPU80からの制御データを受け、肩関節48Rの直交する3軸のそれぞれの角度を制御する3つのモータと肘関節52Rの角度を制御する1つのモータとの計4つのモータ(図3では、まとめて「右腕モータ96」と示す)の回転角度を制御する。同様に、モータ制御ボード86は、CPU80からの制御データを受け、肩関節48Lの直交する3軸のそれぞれの角度を制御する3つのモータと肘関節52Lの角度を制御する1つのモータとの計4つのモータ(図3では、まとめて「左腕モータ98」と示す)の回転角度を制御する。

【0058】

さらに、モータ制御ボード86は、CPU80からの制御データを受け、首関節60の直交する3軸のそれぞれの角度を制御する3つのモータ(図3では、まとめて「頭部モータ100」と示す)の回転角度を制御する。そして、モータ制御ボード86は、CPU80からの制御データを受け、車輪32を駆動する2つのモータ(図3では、まとめて「車輪モータ36」と示す)の回転角度を制御する。

【0059】

モータ制御ボード86にはさらにハンドアクチュエータ108が結合され、モータ制御ボード86は、CPU80からの制御データを受け、ハンド56R、56Lの開閉を制御する。

【0060】

なお、この実施例では、車輪モータ36を除くモータは、制御を簡素化するためにステッピングモータ(すなわち、パルスモータ)を用いる。ただし、車輪モータ36と同様に直流モータを用いるようにしてもよい。また、ロボット12の身体部位を駆動するアクチュエータは、電流を動力源とするモータに限らず適宜変更された、たとえば、他の実施例では、エアアクチュエータが適用されてもよい。

【0061】

センサ入力／出力ボード88は、モータ制御ボード86と同様に、DSPで構成され、各センサからの信号を取り込んでCPU80に与える。すなわち、赤外線距離センサ40のそれぞれからの反射時間に関するデータがこのセンサ入力／出力ボード88を通じてCPU80に輸入される。また、全方位カメラ46からの映像信号が、必要に応じてセンサ入力／出力ボード88で所定の処理を施してからCPU80に輸入される。眼カメラ70からの映像信号も、同様に、CPU80に輸入される。また、上述した複数の接触センサ58(図3では、まとめて「接触センサ58」と示す)からの信号がセンサ入力／出力ボード88を介してCPU80に与えられる。音声入力／出力ボード90もまた、同様に、DSPで構成され、CPU80から与えられる音声合成データに従った音声または声がスピーカ64から出力される。また、マイク66からの音声入力、音声入力／出力ボード90を介してCPU80に与えられる。

【0062】

また、CPU80は、バス82を介して通信LANボード102に接続される。通信LANボード102は、たとえばDSPで構成され、CPU80から与えられた送信データを無線通信装置104に与え、無線通信装置104は送信データを、ネットワーク14を介してサーバ20に送信する。また、通信LANボード102は、無線通信装置104を介してデータを受信し、受信したデータをCPU80に与える。たとえば、送信データとしては、ロボット12からサーバ20への信号(コマンド)であったり、ロボット12が行ったコミュニケーションについての動作履歴情報(履歴データ)などであったりする。このように、コマンドのみならず履歴データを送信するのは、メモリ84の容量を少なくするためと、消費電力を抑えるためである。この実施例では、履歴データはコミュニケーションが実行される度に、サーバ20に送信されたが、一定時間または一定量の単位でサーバ20に送信されるようにしてもよい。

【0063】

さらに、CPU80は、バス82を介して無線タグ読取装置106が接続される。無線タグ読取装置106は、アンテナ（図示せず）を介して、無線タグ18（RFIDタグ）から送信される識別情報の重畳された電波を受信する。そして、無線タグ読取装置106は、受信した電波信号を増幅し、当該電波信号から識別信号を分離し、当該識別情報を復調（デコード）してCPU80に与える。図1によれば無線タグ18は、ロボット12が配置された会社の受付や一般家庭の居間などに居る人間16に装着され、無線タグ読取装置106は、通信可能範囲内の無線タグ18を検出する。なお、無線タグ18は、アクティブ型であってもよいし、無線タグ読取装置106から送信される電波に応じて駆動されるパッシブ型であってもよい。

【0064】

10

図4を参照して、サーバ20のハードウェアの構成について説明する。図4に示すように、サーバ20は、CPU200を含む。CPU200は、プロセッサとも呼ばれ、バス202を介して、メモリ204、カメラ制御ボード206、無線タグ読取装置208、LAN制御ボード210、入力装置制御ボード212、およびモニタ制御ボード214に接続される。

【0065】

CPU200は、サーバ20の全体の制御を司る。メモリ204は、ROM、RAM、およびHDDなどを包括的に示したものであり、サーバ20の動作のためのプログラムを記録したり、CPU200が動作する際のワークエリアとして機能したりする。カメラ制御ボード206は、当該制御ボード206に接続されるカメラ120を制御するためのもの

20

【0066】

無線タグ読取装置208は、当該制御ボード208に接続されるアンテナ124を介して人間16や物品（本）24に装着された無線タグ18から送信される識別情報の重畳された電波を受信する。そして、無線タグ読取装置208は、受信した電波信号を増幅し、当該電波信号から識別信号を分離し、当該識別情報を復調（デコード）してCPU200に与える。アンテナ124は、ロボット12が配置された会社の受付や一般家庭の各部屋などにくまなく配置され、システム10が対象とするすべての物品（本）24および人間16の無線タグ18から電波を受信できるようになっている。したがって、アンテナ124は複数存在するが、図1および図4では包括的に示している。

30

【0067】

また、LAN制御ボード210は、当該制御ボード210に接続される無線通信装置216を制御し、サーバ20が外部のネットワーク14に無線によってアクセスできるようにするものである。さらに、入力装置制御ボード212は、当該制御ボード212に接続される入力装置としてのたとえば、キーボードやマウスなどによる入力を制御するものである。そして、モニタ制御ボード214は、当該制御ボード214に接続されるモニタへの出力を制御するものである。

【0068】

また、サーバ20は、図示しないインターフェースによって、サーバ20に付設された物品辞書122および音声認識辞書126（図1参照）に接続されている。

40

【0069】

メモリ204には、物品ローカル辞書205A、音声認識ローカル辞書205B、発話パターン辞書205C、発話辞書205D、個人正誤DB205E、および音声認識率DB205Fが設定されている。

【0070】

物品ローカル辞書205Aは、後述する物品辞書122から抽出された内容が登録される辞書である。サーバ20は、ロボット12が人間16を認識した際に、当該人間16の近傍に存在する物品（本）24の情報だけを物品辞書122から抽出して物品ローカル辞書205Aに登録する。音声認識ローカル辞書205Bは、後述する音声認識辞書126から抽出された内容が登録される辞書である。サーバ20は、ロボット12が人間16を

50

認識して物品ローカル辞書 205A を作成すると、当該物品ローカル辞書 205A に登録されている単語を音声認識するために必要な情報を音声認識辞書 126 から抽出して音声認識ローカル辞書 205B に登録する。したがって、物品ローカル辞書 205A および音声認識ローカル辞書 205B は、人間 16 の位置の変化に応じて動的に書き換えられる。このように音声認識辞書 126 から音声認識ローカル辞書 205B を作成し、音声認識に使用する辞書を小さくすることによって音声認識の対象となる単語（音素記号列）の数を少なくし、音声認識の処理にかかる時間を短くするとともに正しく音声認識できる割合を高めることができる。

【0071】

発話パターン辞書 205C は、人間 16 の行った発話の内容が特定のパターンであるか否かを判断するための辞書である。発話辞書 205D は、サーバ 20 がロボット 12 に、人間 16 に対して発話させる音声の内容を決定するために必要な情報を記憶している。また、個人正誤 DB 205E は、システム 10 が、人間 16 が指示した物品（本）24 を特定することに最終的に成功したか否かを人間 16 の ID 別に記憶している。そして、音声認識率 DB 205F は、音声認識辞書 126 に登録されている単語のそれぞれについての音声認識における認識率、つまり、実際の音声認識により当該単語を正しく認識できた割合を記憶している。

【0072】

次に、図 5 を参照して物品辞書 122 を説明する。この図 5 に示す物品辞書 122 は、たとえばユーコード（Ucode）のような ID をそれぞれの物品の 1 つに割り当て、物品毎にその名称、属性、位置（座標）などの必要な情報を文字列として登録している。なお、ユーコードは、具体的には、128 ビットの数字からなり、340 兆の 1 兆倍のさらに 1 兆倍の数の物品を個別に識別できるものである。ただし、この物品辞書 122 に使う ID は必ずしもこのようなユーコードである必要はなく、適宜の数字や記号の組み合わせからなるものであってよい。

【0073】

このような物品辞書 122 は、システム 10（ロボット 12 およびサーバ 20）が識別すべき対象物となるすべての、たとえば家庭内の物品を ID と文字列とで登録するものであり、いわばグローバル辞書に相当する。

【0074】

物品辞書 122 には、1 つの物品（本）24 についての情報が 1 つのレコードとして登録されている。そして、たとえば、1 つのレコードには、本の ID 以外に「名称」、「属性」、「著者」、「出版社」などの項目が記憶されている。「属性」の項目には、本のカバーの色や体裁など当該本を補足的に説明する単語が記憶されている。各項目の情報はテキスト形式の単語で記憶されている。しかし、「属性」の項目においてダブルコーテーション（"）で囲まれている情報は、音声認識における音素記号形式（音素記号列）の単語で記憶されている。

【0075】

この音素記号列の情報は、ロボット 12 がマイク 66 で拾った人間 16 が発する音声をサーバ 20 が音声認識する処理の過程で生成される情報であり、音声認識ローカル辞書 205B（音声認識辞書 126）に登録されておらず、音声認識をすることができなかったが、人間 16 の視線や指差しにより、人間 16 がどの本を指示しているか判断できた場合に、その認識できなかった音声に相当する音素記号列をその本のレコードの「属性」として記憶したものである。

【0076】

また、音素記号列の後に括弧書きで付記されているのは、当該音素記号列に対応つけて記憶された人間 16 の ID の情報である。これは、当該音素記号列の示す音声が、当該 ID が示す人間 16 によって発せられたことを示している。

【0077】

なお、前述したように、物品辞書 122 には各物品（本）24 が存在する場所の情報も

10

20

30

40

50

記憶されており、物品辞書 1 2 2 の 1 つのレコードには、図示しないが当該物品（本） 2 4 が存在する場所を記憶する項目も含まれている。

【0 0 7 8】

次に、図 1 0 を参照して音声認識辞書 1 2 6 を説明する。一般的に音声認識辞書には単語辞書と文法辞書とが存在するが音声認識辞書 1 2 6 は単語辞書である。文法辞書についての説明は省略する。音声認識辞書 1 2 6 は、図 1 0 に示すように、テキスト形式の単語を記憶する項目、テキスト形式の単語に対応する音素記号形式（音素記号列）の単語を記憶する項目、および当該テキスト形式の単語が物品の名称、たとえば、本のタイトル（名称）であるか否かを示す情報を記憶する項目を含むレコードからなっている。テキスト形式の単語を記憶する項目に記憶されているのが物品の名称であるか否かを示す情報を記憶する項目には、テキスト形式の単語を記憶する項目に物品の名称が記憶されている場合には「1」が記憶され、そうでない場合は「0」が記憶されている。

10

【0 0 7 9】

音声認識の処理では、入力された音声を音素に分解し、分解した各音素について当該音素を表す記号を生成する。そして、入力された音声の単語に相当するこの記号の列が音声認識辞書 1 2 6 に記憶されている音素記号列である。音声認識の処理では、音声から生成した音素記号列ともっとも近い音素記号列を音声認識辞書 1 2 6（音声認識ローカル辞書 2 0 5 B）内で特定し、この特定した音素記号列に対応して記憶されている単語を音声認識結果として出力する。

【0 0 8 0】

20

前述したように、このシステム 1 0 では、人間 1 6 が音声と視線および指差しによって物品（本） 2 4 を指示すると、ロボット 1 2 とサーバ 2 0 とが協働して、人間 1 6 が指示した物品（本） 2 4 を特定し、その特定した物品（本） 2 4 をロボット 1 2 が人間 1 6 のところに運搬などする。以下において、この人間 1 6 とシステム 1 0 とのやり取りをコミュニケーションと呼ぶことがある。

【0 0 8 1】

より詳細に述べると、このシステム 1 0 では、人間 1 6 がロボット 1 2 に近づくと、ロボット 1 2 が人間 1 6 を無線タグ 1 8 によって認識する。サーバ 2 0 には、システム 1 0 が対象とする物品（本） 2 4 のすべてが登録された物品辞書 1 2 2、および音声認識によって物品（本） 2 4 を特定するための単語が登録された音声認識辞書 1 2 6 が付設されている。ロボット 1 2 が人間 1 6 を認識すると、ロボット 1 2 はサーバ 2 0 に対して物品辞書 1 2 2 および音声認識辞書 1 2 6 のローカル辞書（物品ローカル辞書 2 0 5 A、音声認識ローカル辞書 2 0 5 B）の作成を指示する。

30

【0 0 8 2】

ローカル辞書の作成の指示を受けると、サーバ 2 0 では、ロボット 1 2 が認識した人間 1 6 の位置を特定し、特定した当該人間 1 6 から所定の範囲内、たとえば、半径 5 m 以内にある物品（本） 2 4 のレコードのみを物品辞書 1 2 2 から抽出して物品ローカル辞書 2 0 5 A を作成する。そして次に、音声認識辞書 1 2 6 から、物品ローカル辞書 2 0 5 A に登録されている物品（本） 2 4 を音声認識するため必要な情報のみを抽出して音声認識ローカル辞書 2 0 5 B を作成する。

40

【0 0 8 3】

その後、ロボット 1 2 は、認識した人間 1 6 に対して、たとえば、「何か本を持ってきましょうか？」という発話を行う。この発話に対し、人間 1 6 は、持ってきてほしい物品（本） 2 4 に視線を向けつつ当該物品（本） 2 4 を指差しながら、「化学の謎持ってきて」などと答える。

【0 0 8 4】

するとロボット 1 2 は、「化学の謎持ってきて」という人間 1 6 の声を音声認識し、人間 1 6 の視線を推定し、指差した指が向かっている方向を推定することによって、人間 1 6 が指示している物品（本） 2 4 を特定する。

【0 0 8 5】

50

人間 16 が指示している物品（本）24 を特定すると、サーバ 20 は、人間 16 に特定した物品（本）24 を確認するためにロボット 12 が発話する音声の内容、たとえば、「白い本ですか？」を決定し、ロボット 12 が当該物品（本）24（化学の謎の本 24）を指し示しながらこれを発話する。

【0086】

このとき、サーバ 20 は、発話の内容を、次回に人間 16 が当該物品（本）24（化学の謎の本 24）をロボット 12 に指示する際に、人間 16 に当該物品（本）24（化学の謎の本 24）を特定するための単語として使用してほしい単語を含んだものとする。

【0087】

たとえば、ロボット 12 が特定した物品（本）24（化学の謎の本 24）を確認するための発話内容が「白い本ですか？」であった場合、システム 10 は、次回に人間 16 が「化学の謎」の本を指示する際には、指差しとともに、「化学の謎持ってきて」と発話する代わりに「あの白い本持ってきて」と発話することを期待している。このことは、人間はロボットの発話内容を真似する傾向があるという知見に基づいている。この場合、人間 16 に物品（本）24 を特定するために次回に使用することを期待する単語は「白い」である。この「白い」という単語は、人間 16 の近傍に「化学の謎」の本 24 の他に「白い」物品（本）24 が存在せず、また、「白い」という単語の音声認識における認識率が高い場合にサーバ 20 によって選択される。

10

【0088】

ところで、人間 16 が指差しとともに、「化学の謎を持ってきて」などと発話する場合、「化学の謎」という単語ではなく、その人間 16 が独自に使用する、たとえば、「化学の謎」の省略形である「バケナゾ」という単語を用い「バケナゾ持ってきて」などと発話することが考えられる。この場合、「バケナゾ」という単語が物品辞書 122 および音声認識辞書 126 に登録されていなければ、システム 10 は「バケナゾ」という音声を認識し、「バケナゾ」が「化学の謎」の本を指示していることを特定することができない。

20

【0089】

このような場合においても、システム 10 は、人間 16 の視線や指差しの方向に基づいて人間 16 が指示している物品（本）24 を特定することが可能である。したがって、「バケナゾ」などのように音声認識できなかった単語がある場合には、視線や指差しの方向に基づいて特定した結果を利用して、音声認識できなかった単語を物品辞書 122 および音声認識辞書 126 に登録する。

30

【0090】

システム 10 が特定した物品（本）24 を確認するために、ロボット 12 がたとえば「白い本ですか？」と発話すると、人間 16 は、「そうです」あるいは「ちがいます」などと発話し、ロボット 12 に返答する。サーバ 20 は、この人間 16 の返答における音声を音声認識し、システム 10 が特定した物品（本）24 が、人間 16 が指示したものであるか否かを判断する。システム 10 が特定した物品（本）24 が、人間 16 が指示したものでなかった場合には、次の候補である物品（本）24 が、人間 16 が指示したものであるか否かを確認する。一方、システム 10 が特定した物品（本）24 が、人間 16 が指示したものであった場合には、ロボット 12 が当該物品（本）24 を人間 16 のところにまで運搬する。

40

【0091】

次に、図 1 に示す実施例におけるロボット 12 およびサーバ 20 の動作について、図 15 から図 22 に示すフロー図を参照して説明する。

【0092】

図 15 の最初のステップ S1 において、ロボット 12 の CPU 80（図 3）は、同じく図 3 に例示するセンサ入力／出力ボード 88 からのセンサ入力にしたがって、人間 16（図 1）を認識したかどうか判断する。具体的には、たとえば赤外線センサ 40 で人体を検知し、そのとき無線タグ読取装置 106 で人間 16 が装着している無線タグ 18 を認識したとき、人間 16 を認識したと判断する。このとき、ロボット 12 は、無線タグ 18 より

50

人間 16 のユーザ I D (たとえば、(456…0000004)) を読み取る。

【0093】

ステップ S 1 で Y E S と判断すると、次のステップ S 3 で、ロボット 12 の C P U 8 0 は、ローカル辞書、つまり、物品ローカル辞書 205 A および音声認識ローカル辞書 205 B を作成する指示を、ネットワーク 14 を介してサーバ 20 に送信する。このとき、ローカル辞書の作成指示と同時に、ステップ S 1 で認識したと判断した人間 16 のユーザ I D (456…0000004) をもサーバ 20 に送信する。

【0094】

ロボット 12 からローカル辞書の作成指示とユーザ I D とが送信されると、サーバ 20 では、図 16 のステップ S 31 で、C P U 200 は、ローカル辞書の作成指示を受信したと判断し、ステップ S 33 において、ロボット 12 が認識した人間 16 とのコミュニケーションに使用する当該人間 16 に専用の物品ローカル辞書 205 A および音声認識ローカル辞書 205 B の作成を行う。

【0095】

ステップ S 33 の「ローカル辞書の作成」の処理は、図 18 のフロー図に示す手順で実行される。まず、図 18 のステップ S 81 で、ロボット 12 から送信されたユーザ I D で特定される人間 16 の近傍にある物品 (本) 24 のレコードを物品辞書 122 から抽出して物品ローカル辞書 205 A を作成する。

【0096】

このとき、C P U 200 は、無線タグ読取装置 208 を駆動してアンテナ 124 を介して、人間 16 に付着された無線タグ 18 から電波を受信して人間 16 ごとにユーザ I D と人間の位置を特定する。そして、こうして特定した複数の人間 16 のユーザ I D と人間の位置の情報と、ロボット 12 より送信されたユーザ I D とに基づいてロボット 12 が認識した人間 16 の位置を特定する。

【0097】

ロボット 12 が認識した人間 16 の位置を特定すると、次に、物品辞書 122 に記憶されている各レコードの物品 (本) 24 の位置が記憶されている項目を参照し、図 5 に示すような物品辞書 122 から、ロボット 12 が認識した人間 16 の近傍に位置する物品 (本) 24 のレコードをすべて抽出して、図 6 に示すような物品ローカル辞書 205 A を作成する。

【0098】

こうして物品ローカル辞書 205 A を作成すると、次に、C P U 200 はステップ S 83 で、「物品ローカル辞書 205 A に登録されている単語」の一覧 300 を作成する。図 7 は、この単語の一覧 300 を示した図解図である。この単語の一覧 300 は、音声認識辞書 126 から音声認識ローカル辞書 205 B を作成する際に使用される。図 6 および図 7 からわかるように、単語の一覧 300 に登録される「物品ローカル辞書 205 A に登録されている単語」とは、物品ローカル辞書 205 A の各レコードの「名称」、「属性」、「著者」、「出版社」のそれぞれに記憶されているテキスト形式の単語である。ただし、図 7 に示されるように、この単語一覧 300 には、テキスト形式の文字列からなる単語のみではなく、先に説明した音素記号列 (図 7 の例では、“bakenazo”、“saiai”) もユーザ I D (図 7 の例では、(456…0000004)、(456…0000003)) とともに登録される。これらの音素記号列 (およびユーザ I D) は、先述したように、物品辞書 122 に単語が登録されていなかったために、登録されていない単語に対応する音素記号列が後から登録されたものである。

【0099】

ステップ S 85 では、単語の一覧 300 に登録されている音素記号列の中から、ロボット 12 が認識した人間 16 のユーザ I D 以外のユーザ I D が対応つけられている音素記号列を削除する。ロボット 12 が認識した人間 16 のユーザ I D が (456…0000004) である場合には、図 8 に示したように、他の人間 16 のユーザ I D (456…0000003) と対応つけられている “saiai” が削除される。ユーザ I D が (456…0000003) である人間 16 に特

10

20

30

40

50

有の“saiai”という本24（化学の目）の名称の呼び方は、ロボット12が認識したユーザIDが（456…0000004）である人間16とのコミュニケーションにおける音声認識には必要がないと考えられるからである。

【0100】

次に、CPU200は、ステップS87で、単語の一覧300における冗長性を排除する。つまり、図7の例において、単語の一覧300に登録されている「白色」や「ハードカバー」などの重複部分を図8に示すように削除する。

【0101】

こうして単語の一覧300が完成すると、次に、この単語の一覧300を利用して音声認識辞書126から音声認識ローカル辞書205Bを作成する。

10

【0102】

まず、ステップS89において、CPU200は、メモリ204内に設定したカウンタCt1をインクリメントする。初期状態では、カウンタCt1は「0」である。このカウンタCt1は、単語の一覧300に登録されている単語の数をカウントするものであり、単語の一覧300のポインタとして機能する。したがって、カウンタCt1のカウント値によって、単語の一覧300内において異なる単語を指定する。カウンタCt1のカウント値が、単語の一覧300に登録されている単語の数「L」に等しくなるまで、以下の動作が各単語について実行されると理解されたい。

【0103】

ステップS91では、単語の一覧300に登録されている単語が音素記号列であるか否かを判断する。たとえば、単語の一覧300に登録されている単語が図8に示すように、テキスト形式で表現された「地球温暖化」という単語であれば、ステップS91でNOと判断される。すると、ステップS93で、テキスト形式である「地球温暖化」という単語をキーとして、音声認識辞書126（図10参照）を検索し、「地球温暖化」という単語を記憶した項目を含むレコードを抽出し、図11に示すように音声認識ローカル辞書205Bに登録する。

20

【0104】

一方、単語の一覧300に登録されている単語が図8に示すように、音素記号形式で表現された“bakenazo”という単語であれば、ステップS91でYESと判断される。すると、ステップS95で、音素記号列である“bakenazo”という単語をキーとして音声認識辞書126を検索し、“bakenazo”という単語を記憶した項目を含むレコードを抽出し、図11に示すように音声認識ローカル辞書205Bに登録する。

30

【0105】

このようにして、物品辞書122から物品ローカル辞書205Aが作成され、音声認識辞書126から音声認識ローカル辞書205Bが作成される。

【0106】

一方、図15に戻って、ロボット12において、ステップS3でローカル辞書の作成指示をサーバ20に送信した後は、ステップS5で、CPU80は、メモリ84内に設定している発話／ジェスチャ辞書85Aを用いて、スピーカ64から、たとえば、「何か本を持って来ましょうか？」のような発話を行なわせる。その後、人間16がたとえば「花火の匠持ってきて」という発話をしたとすると、ロボット12が人間16の指示を認識し、ステップS7で、CPU80が、“YES”と判断する。このとき、人間16は、「花火の匠持ってきて」という発話とともに、該当本に視線を向けつつ当該本を指差すことによって、どの本持ってきてほしいかを指示する。

40

【0107】

CPU80は、ステップS7でYESと判断すると、次にステップS9において、人間16が指示した「花火の匠持ってきて」という音声の情報とともに、音声認識ならびに視線と指差しの推定を行わせる指示をサーバ20に送信する。

【0108】

サーバ20では、CPU200が、図16のステップS35において、ロボット12か

50

ら音声認識ならびに視線と指差しの推定を行わせる指示を音声の情報とともに受信したと判断すると、ステップS 3 7における「音声認識による物品の推定」の処理とステップS 3 9における「視線と指差しによる物品の推定」の処理とを並列的に実行する。

【0109】

「音声認識による物品の推定」の処理は、図19のフロー図に示す手順にしたがって実行される。まず、CPU 200は、ステップS 1 1 1で、ロボット12から送信されてきた「花火の匠持ってきて」という音声の情報を音声認識する。具体的には、「花火の匠持ってきて」という音声を音素に分割し、各音素に対応する音素記号を生成する。そして、図示しない文法辞書を参照して、生成した音素記号列“hanabinotakumi mottekite”のうち“hanabinotakumi”という音素記号列が目的語であると特定する。次に、この“hanabinotakumi”という音素記号列をキーとして音声認識ローカル辞書205B（図11参照）を検索し、もっとも近い音素記号列の単語を記憶した項目を有するレコードを特定する。そして、特定したレコードに記憶されているテキスト形式の単語である「花火の匠」を得る。また、このとき、特定したレコードの「物品名称」の項目に記憶されている情報に基づいて、音声認識されたのが物品の名称であるか、つまり、本24のタイトル（名称）であるか否かを判断することができる。この判断に基づいて後ほどフラグtwの設定を行うが、後ほど設定を行うのはフロー図の表現上の都合であり、ステップS 1 1 1においてフラグtwの設定を行えばよい。

10

【0110】

たとえば、人間16が「花火の匠持ってきて」という代わりに「黒いの持ってきて」と指示した場合には、同様にしてステップS 1 1 1における音声認識の結果として、「黒い」が得られる。なお、音声認識の方法としては特に限定されるものではなく既存の方法を採用することができる。

20

【0111】

また、人間16が「花火の匠持ってきて」という代わりに、「花火の匠」という単語の人間16独特の略語である「ハナタク」という単語を用いて、「ハナタク持ってきて」と発話した場合には、音声認識ローカル辞書205Bには「ハナタク」という単語に相当する“hanataku”という音素記号列は登録されていないので、音声認識をすることができない。

【0112】

このように、音声認識ができない単語があった場合には、CPU 200は、ステップS 1 1 3において、YESと判断し、ステップS 1 1 5において、音声認識した人間16の発話内容は特定のパターンであったか否かを判断する。ここで発話内容の特定のパターンとは、たとえば、「〇〇持ってきて」や「〇〇お願い」などであり、人間16が指示する物品を特定する単語（〇〇）が含まれるであろう発話パターンである。人間16が発した発話の内容がこれらの特定のパターンに該当するか否かは、音声認識の処理における図示しない文法辞書を用いた発話内容の文法解析とともに、メモリ204に設定されている発話パターン辞書205Cを参照することによって判定することができる。

30

【0113】

ステップS 1 1 5でYESと判断した場合は、音声認識できなかった単語であって、人間16が指示した物品を特定するであろう単語があることを示すために、ステップS 1 1 7で、メモリ204上に設定されているフラグnwに数値「1」を記憶させてフラグnwをオン状態にする。なお、フラグnwの初期状態はオフ状態（数値「0」）である。そして、ステップS 1 1 9では、ステップS 1 1 1における音声認識の処理で生成された、音声認識できなかった単語の音素記号列をメモリ204に設定されたワークエリアwaに格納する。なお、ステップS 1 1 5でNOと判断すると、ステップS 1 1 7およびステップS 1 1 9をスキップする。

40

【0114】

一方、ステップS 1 1 3においてNOと判断すると、つまり、音声認識できない単語がなかったと判断すると、ステップS 1 2 1において、CPU 200は、音声認識した単語

50

は物品の名称、つまり、本 2 4 のタイトル（名称）であったかどうかを判断する。この音声認識した単語が本 2 4 のタイトル（名称）であったかどうかは、音声認識ローカル辞書 2 0 5 B の各レコードに含まれる「物品名称」の項目に記録されている情報に基づいて判断される。

【0 1 1 5】

ステップ S 1 2 1 で Y E S と判断すると、ステップ S 1 2 3 で、本 2 4 のタイトル（名称）が音声認識されたことを示すために、メモリ 2 0 4 に設定されたフラグ t w に数値「1」を記憶させてフラグ t w をオン状態にする。なお、フラグ t w の初期状態はオフ状態（数値「0」）である。ステップ S 1 2 1 で N O と判断するとステップ S 1 2 3 をスキップする。

10

【0 1 1 6】

ステップ S 1 2 5 では、C P U 2 0 0 は、ステップ S 1 1 1 における音声認識の結果として得られたテキスト形式の単語、たとえば、「花火の匠」に基づいて物品ローカル辞書 2 0 5 A を参照し、人間 1 6 が指示した物品（本）2 4 の候補としての物品（本）2 4 を選出する。具体的には、ステップ S 1 1 1 で得られたテキスト形式の単語、たとえば、「花火の匠」をキーとして、物品ローカル辞書 2 0 5 A を検索し、「花火の匠」というテキスト形式の単語を項目に記憶しているレコードを特定する。図 6 の例では、物品 I D が「123...0000046」であるレコードが「名称」の項目に「花火の匠」という単語を記憶しているので、物品 I D が「123...0000046」である物品（本）2 4 が特定される。こうして特定された物品 I D の物品（本）2 4 が、人間 1 6 が指示した物品の候補として選出され、当該物品（本）2 4 の I D が物品リスト A（不図示）に登録される。

20

【0 1 1 7】

ここで、ステップ S 1 1 1 における音声認識の結果として得られた単語が、たとえば、「白い」であった場合は、図 6 の物品ローカル辞書 2 0 5 A の例の場合では、「属性」の項目に「白い」という単語を記憶しているレコードが複数存在するので、これらのレコードに含まれる物品 I D、つまり、物品 I D が「123...0000035」、「123...0000091」、および「123...0000102」の物品（本）2 4 が、人間 1 6 が指示した物品の候補として選出される。この場合、こうして選出された複数の物品の I D が、物品リスト A（不図示）に登録される。

【0 1 1 8】

30

ステップ S 1 2 7 では、こうして選出された物品（本）2 4 の候補が物品リスト A に存在するか否かを判断する。ステップ S 1 2 7 で Y E S と判断すると、ステップ S 1 2 9 で、C P U 2 0 0 は、音声認識の結果による物品（本）2 4 の候補が存在することを示すために、メモリ 2 0 4 に設定されているフラグ c a に数値「1」を記憶させてフラグ c a をオン状態にする。なお、フラグ c a の初期状態はオフ状態（数値「0」）である。ステップ S 1 2 7 で N O と判断すると、ステップ S 1 2 9 をスキップする。このようにして、音声認識によって、人間 1 6 が指示した物品（本）2 4 の推定が行われる。

【0 1 1 9】

図 1 6 に戻って、ステップ S 3 9 の「視線と指差しによる物品の推定」の処理は、図 2 0 のフロー図に示す手順にしたがって実行される。C P U 2 0 0 は、まず、ステップ S 1 4 1 で、メモリ 2 0 4 内に設定したカウンタ C t 2 をインクリメントする。初期状態ではカウンタ C t 2 は「0」が設定されている。このカウンタ C t 2 は、人間 1 6 の近傍に存在する物品の数をカウントするもので、物品ローカル辞書 2 0 5 A（図 6 参照）のポインタとして機能する。したがって、カウンタ C t 2 のカウント値によって、物品ローカル辞書 2 0 5 A 内において、異なる物品を指定する。カウンタ C t 2 のカウント値が、物品ローカル辞書 2 0 5 A 内にリストアップしている物品の数「m」に等しくなるまで、以下の動作が各物品について、実行されるものと理解されたい。

40

【0 1 2 0】

ステップ S 1 4 1 に続いて、C P U 2 0 0 は、人間 1 6 の視線を推定してその確信度を求める動作と、指差し方向を推定してその確信度を求める動作とを並行して実行する

50

が、ここでは便宜上、まず視線を推定し次いで指差し方向を推定する順序で説明する。

【0121】

図20のステップS141～ステップS153の処理は、図16のステップS35でロボット12から音声認識ならびに視線と指差しの推定を行わせる指示を受信した後の一定の繰り返しの時間(t_1 , t_2 , t_3 , ..., t_m)毎に実行されるが、実施例では、50Hz(1秒間に50回)で実行されるものとする。なお、ステップS141～ステップS153の処理は、たとえば、図15のステップS5でロボット12が発話を行った際に、その旨を知らせる通知をロボット12からサーバ20に送信し、当該通知を受信したサーバ20が、当該通知を受信した後の一定の繰り返しの時間(t_1 , t_2 , t_3 , ..., t_m)毎に実行するようにしてもよい。

10

【0122】

ステップS143では、CPU200は、ロボット12から、人間16の視線の方向を示す視線ベクトル情報を受信する。フロー図には明記しないが、ロボット12では、CPU80は、たとえば眼カメラ70からのカメラ映像を処理することによって、先に挙げた公開公報に記載したいずれかの方法に従って、人間16の左右のそれぞれの眼の視線ベクトルを推定する。この左右それぞれの眼の視線方向は図13において直線L1およびL2で示される。このようにして、ロボット12は、所定の時間間隔で繰り返し各視線L1およびL2を推定し、この視線の方向を示す視線ベクトル情報をサーバ20に送信する。

【0123】

サーバ20では、ステップS145において、CPU200が、カウンタCt2がそのとき物品ローカル辞書205A内でポイントしている物品と、各視線L1およびL2との距離を計算する。この計算の際には、CPU200は、ロボット12から受信した視線ベクトル情報をロボット座標系からワールド座標系に座標変換して使用する。また、この距離の計算の際には、物品ローカル辞書205Aの当該物品(本)24のレコードに記録されている当該物品(本)24の位置の情報が利用される。

20

【0124】

一方、指差し方向を推定するためには、ステップS147において、まず、CPU200は、人間16が指差し動作をした腕を特定する。具体的には、モーションキャプチャのデータを参照して、たとえば、人間16の指先と肩の高さとの差が小さい側の腕を指差し腕として推定する。なぜなら、指差し動作をする場合には、腕を持ち上げる動作をまずすると考えられるからである。このようにして、ステップS147でどちらの腕を用いて指差し動作をするかを推定した後、CPU200は、次のステップS149において、指差し方向を推定する。

30

【0125】

この実施例では、図13に示すように、指差し腕の指先と顔の中心(重心)とを通る直線L3、および指差し腕の指先とその腕の肘とを通る直線L4を想定する。そして、モーションキャプチャのデータを参照して、その直線L3およびL4を推定する。次のステップS151において、各直線L3およびL4と各物品との間の距離を計算する。

【0126】

上述のステップS143～S145およびステップS147～S151は、繰り返し時間(t_1 , t_2 , t_3 , ..., t_m)毎に行われる。そして、各繰り返しの時間(t_1 , t_2 , t_3 , ..., t_m)毎に、線L1, L2, L3, およびL4との距離が最小になる物品を求める。各線において、最小になった物品に対して高い確信度(図14でいえば「○」印)を付与する。このようにして、たとえば図14に示すような確信度表を作成する。

40

【0127】

このように直線毎に最短距離を持つ物品を算出することによって確信度表を作成するようになれば、1つの物品について2以上の直線について確信度(○)が付与されることがある。このことによって、後にステップS155で説明するような物品リストB(不図示)を作成することができるのである。

50

【0128】

この図14の確信度表において、視線L1およびL2のそれぞれについて評価される確信度は「視線確信度」ということができ、指差し方向線L3およびL4のそれぞれについて評価される確信度が「指差し方向確信度」とであるということができる。

【0129】

図14に示す例で説明すると、「123...0000001」のIDを持つ物品、実施例でいえば図1に示す「地球温暖化」という名称の本についていえば、時間t1に一方の視線L2とこの本との間の距離が最小になったものの、その他の時間区間ではどの線も当該本に最接近することはなかったと判断できる。次の、「123...0000046」のIDを持つ物品、実施例でいえば図1に示す「花火の匠」という名称の本についていえば、時間t1を除いて、各時間

10

【0130】

ステップS155においてCPU200は、図14に示す確信度表を参照して、そのとき人間16が指示したと考える物品(本)24を特定する。具体的には、確信度評価(図14で言えば丸印)が単に多い順や、繰返し時間で視線(L1またはL2)と指差し(L3またはL4)の両方に○が入っている回数が多い順などに従って、物品のIDのリストである物品リストB(不図示)を作成する。

【0131】

この確信度評価について、たとえば、図14に示す例で説明すると、「123...0000001」のIDを持つ「地球温暖化」という名称の本についていえば、確信度評価は「1」(1つの○印が付与された。)であり、「123...0000046」のIDを持つ「花火の匠」という本の確信度は「3」ということになる。したがって、この場合には、物品リストBには、ID「123...0000046」、ID「123...0000001」の順で登録される。

20

【0132】

ただし、確信度(○印)の数が同じ場合であるとか、確信度(○印)の数が所定の閾値より小さい場合など、判断に迷う場合には、たとえば、図14に示す各繰返し時間の全区間の半分以上で確信度が付与されているような物品を対象物として特定すればよい。

【0133】

このようにして、ステップS37の「音声認識による物品の推定」の処理で物品リストAが作成され、ステップS39の「視線と指差しによる物品の推定」の処理で物品リストBが作成されると、次に、CPU200は、図16のステップS41で、人間16が指示した物品の候補の一覧である候補物品一覧C(不図示)を作成する。この候補物品一覧Cを作成する処理は、図21のフロー図に示す手順で実行される。

30

【0134】

まず、CPU200は、ステップS161で、メモリ204に設定されているフラグcaがオン状態であるか、つまり、音声認識の結果により推定された物品が存在するか否か、言い換えれば、物品リストAに物品のIDが登録されているか否かを判断する。

【0135】

ステップS161でNOと判断すると、音声認識の結果によって推定された物品は存在しないので、ステップS163で、視線確信度および指差し方向確信度に基づいて候補物品一覧Cを作成する。つまり、物品リストBの内容をそのままに候補物品一覧Cを作成する。

40

【0136】

一方、ステップS161でYESと判断すると、次に、CPU200は、ステップS165で、フラグtwがオン状態であるか、つまり、音声認識により本24のタイトル(名称)(物品の名称)が認識されたか否か、言い換えれば、音声認識の結果として得られた物品リストAに登録されている物品のIDは、人間16の発話内容に含まれる本24のタイトル(名称)に基づいて決定されたか否かを判断する。

【0137】

50

ステップS165でYESと判断すると、ステップS167で、CPU200は、音声認識の結果を視線確信度および指差し方向確信度よりも優先させて候補物品一覧Cを作成する。つまり、物品リストAに登録されている物品のIDが上位となるように先に候補物品一覧Cに登録し、その後に、物品リストBに登録されている物品のIDが下位となるように登録する。これは、人間16が名指しした本24のタイトル(名称)を音声認識して物品(本)24を推定したほうが、視線や指差しに基づいて物品(本)24を推定するよりも確実であると考えられるからである。なお、物品リストAと物品リストBとに重複する物品のIDが存在する場合には、物品リストBに登録されている物品のIDを候補物品一覧Cに登録しない。

【0138】

10

一方、ステップS165でNOと判断すると、CPU200は、ステップS169で、視線確信度および指差し方向確信度を音声認識の結果よりも優先させて候補物品一覧Cを作成する。つまり、物品リストBに登録されている物品のIDが上位となるように先に候補物品一覧Cに登録し、その後に、物品リストAに登録されている物品のIDが下位となるように登録する。なお、物品リストAと物品リストBとに重複する物品のIDが存在する場合には、物品リストAに登録されている物品のIDを候補物品一覧Cに登録しない。

【0139】

このようにして候補物品一覧Cが作成されると、次に、CPU200は、図16のステップS43で、候補物品一覧Cから第1候補である物品(本)24を選出する。はじめのステップS43では、候補物品一覧Cの先頭に登録されている物品(本)24のIDを選出する。2回目以降のステップS43では、候補物品一覧Cの2番目以降に登録されている物品(本)24のIDを選出する。

20

【0140】

そして、ステップS45では、ステップS43で物品(のID)が選出されたか否か、つまり、候補となる物品(本)24が存在したか否かを判断する。ステップS45で、YESと判断すると、次に、CPU200は、ステップS47において、ステップS43で選出された物品の位置情報を取得する。つまり、CPU200は、ステップS43で選出された物品(本)24(以下、「選出物品(本)24」と呼ぶ。)のIDをキーとして物品ローカル辞書205Aを検索することによって、当該物品IDが示す物品(本)24の位置の情報を取得する。

30

【0141】

次に、CPU200は、ステップS49で、「選出物品(本)24」が、人間16が指示した物品(本)24であるか否かを人間16に確認するためにロボット12に発話させる際の発話内容を決定する。ロボット12が行うこの発話の内容には、サーバ20が、人間16が指示したと推定する物品(本)24を特定する単語(以下、「特定単語」と呼ぶ。)が含まれる。図6を参照して、サーバ20が、人間16が指示したと推定した物品(本)24が、物品IDが「123...0000046」である「花火の匠」という本であったとする。この場合、ステップS49では、たとえば、「花火の匠の本ですね」、「黒色の本ですね」、「ハードカバーの本ですね」などという発話内容を決定する。この例の場合、それぞれ、「花火の匠」、「黒色」、「ハードカバー」が先述の「特定単語」である。この「発話内容の決定」の処理は、図22のフロー図に示す手順で実行される。

40

【0142】

図22を参照して、まず、CPU200は、ステップS181で、メモリ204に設定されている個人正誤DB205Eを参照して、ロボット12が認識した人間16とコミュニケーションをとるのがはじめてであるか否か、つまり、当該人間16の指示する物品(本)24を特定するのがはじめてであるか否かを判断する。個人正誤DB205Eは、ロボット12と人間16とのコミュニケーション結果の記録である。つまり、この個人正誤DB205Eには、ロボット12とコミュニケーションを行った人間16のIDと当該人間16が指示する物品(本)24をシステム10が正しく推定することができたか否かを示す成功、不成功の別がコミュニケーションのたび毎に記録されている。個人正誤DB2

50

05Eに当該人間16のIDが記録されていなければ、当該人間16は初対面であると判断する。

【0143】

ステップS181でYESと判断すると、つまり、人間16は初対面であると判断すると、ステップS183で、CPU200は、物品(本)24の「特定単語」を本24のタイトル(名称)に決定する。この本24のタイトル(名称)の内容は、図16のステップS43で選出した物品(本)24(「選出物品(本)24」)のIDを物品ローカル辞書205Aで検索することによって得ることができる。

【0144】

そして、ステップS203では、メモリ204に設定されている発話辞書205Dを利用して、ステップS183で決定した「特定単語」である本24のタイトル(名称)、たとえば、「花火の匠」に基づいて「花火の匠の本ですね」という発話内容を決定する。なお、ステップS203では、決定した発話内容をロボット12がこれにしたがって発話するための発話内容情報の生成も行う。

【0145】

一方、ステップS181でNOと判断された場合、つまり、当該人間16のIDが個人正誤DB205Eに記録されている場合は、ステップS185で、CPU200は、同じく個人正誤DB205Eを参照して当該人間16とのコミュニケーションにおける成功率を計算し、成功率が、たとえば、70%以上であるか否かを判断する。ステップS185でNOと判断すると、先に説明したように、ステップS183で、物品(本)24の「特定単語」を本24のタイトル(名称)と決定し、ステップS203で、発話内容を決定する。

【0146】

一方、ステップS185で、YESと判断すると、ステップS187で、「選出物品(本)24」に関する単語の一覧D(不図示)を物品ローカル辞書205Aから作成する。たとえば、「選出物品(本)24」が、図6を参照して物品IDが「123...0000046」である「花火の匠」という本24であったとすると、単語の一覧Dには、物品ローカル辞書205Aの物品IDが「123...0000046」であるレコードに登録されている単語である「花火の匠」、「黒色」、「ハードカバー」、「近藤四郎」、「ATR出版株式会社」が登録される。

【0147】

単語の一覧Dが作成されると、次に、CPU200は、ステップS189で、単語の一覧Dに登録されている単語のうち、人間16の近傍に存在する物品(本)24に関連する単語と同一の単語を削除する。人間16の近傍には、物品ローカル辞書205Aに登録されている物品(本)24が存在することになる。「選出物品(本)24」が、図6を参照して物品IDが「123...0000046」である「花火の匠」という本24であったとすると、単語の一覧Dには、先述したように、「ハードカバー」という単語や「ATR出版株式会社」という単語が登録されている。

【0148】

ここで、物品(本)24の「特定単語」を「ハードカバー」という単語や「ATR出版株式会社」という単語に決定し、ロボット12が「ハードカバーの本ですね」や「ATR出版株式会社の本ですね」と発話したとする。すると、図6の物品ローカル辞書205Aからわかるように、人間16の近傍には「ハードカバーの本」や「ATR出版株式会社の本」は複数存在するので、人間16はロボット12の発話に基づいてロボット12がどの物品(本)24を示しているのかがわからない。そこで、単語の一覧Dから、物品ローカル辞書205Aに登録されている「選出物品(本)24」以外のレコードに登録されている単語と同一の単語を削除する。

【0149】

次に、ステップS191で、CPU200は、単語の一覧Dから6個未満の音素からなる単語、つまり、音声認識の処理において6個未満の音素に分解される単語を削除する。

これは、音素数の少ない単語は正しく音声認識されにくいためである。なお、分解される音素の数は、当該単語を音声認識ローカル辞書205Bに参照することによって判明する当該単語に対応する音素記号列を構成する音素記号の数によって判断することができる。ただし、図11の音声認識ローカル辞書205Bに示した音素記号列は、説明の表現の都合上、アルファベットを用いてローマ字であらわしたものであり、図11に示すアルファベットが音素記号をあらわしたものではない。そして、ステップS193では、さらに、単語の一覧Dから平仮名の「う」で始まる単語を削除する。これは、「う」から始まる単語は正しく音声認識されにくいためである。

【0150】

次に、CPU200は、ステップS195で、単語の一覧Dに単語が存在するか否かを判断する。ステップS195でNOと判断すると、「特定単語」とすべき単語が単語の一覧Dに存在しないので、ステップS197で指示形容詞である「あの」という単語を「特定単語」に決定する。そして、ステップS203では、発話辞書205Dを利用して、ステップS197で決定した「特定単語」である「あの」に基づいて「あの本ですね」などという発話内容を決定する。あるいは、指示形容詞とステップS187の段階で作成した単語の一覧Dに登録されているいずれかの単語を用いて、「あの白い本ですね」という発話内容にしてもよい。ただし、この例では、「白い本」という言葉によっては物品（本）24を一意に特定することはできない。なお、ステップS203では、指示代名詞を用いて「白いあれ（これ）ですね」という発話内容にしてもよいし、単に「あれ（これ）ですね」という発話内容にしてもよい。

【0151】

一方、ステップS195でYESと判断すると、ステップS199で、メモリ204に設定されている音声認識率DB205Fを参照し、単語の一覧Dに登録されている単語のそれぞれの音声認識率を取得して単語の一覧Dに登録されている単語を音声認識率の高い順にソートする。そして、ステップS201では、単語の一覧Dの最上位に登録されているもっとも音声認識率の高い単語を「特定単語」に決定する。ステップS203では、発話辞書205Dを利用して、ステップS201で決定した「特定単語」に基づいて発話内容を決定する。このようにして、システム10が推定した人間16が指示した物品（本）24を人間16に確認するためにロボット12が発話する内容が決定される。

【0152】

このように、人間16が指示したとシステム10が推定した物品（本）24を人間16に確認する際にロボット12が発話する内容に、音声認識率の高い「特定単語」を利用すると、人間16がこれを真似し、次回にその物品（本）24を指示する際にこの「特定単語」を利用すれば、システム10は音声認識による物品（本）24の特定を容易に行うことができるようになる。

【0153】

図16に戻って、ステップS49において発話内容が決定されると、次に、CPU200は、ステップS51で、ステップS47で取得した「選出物品（本）24」の位置情報とステップS203（図22）で生成した発話内容情報とをロボット12に送信する。

【0154】

ロボット12では、「選出物品（本）24」の位置情報と発話内容情報とを受信すると、CPU80が、図15のステップS11でYESと判断する。するとCPU80は、次のステップS13において、メモリ84に設定されている発話／ジェスチャ辞書85Aを参照して、「選出物品（本）24」の位置情報に基づいて「選出物品（本）24」を指差すとともに、発話内容情報に基づいて発話を行う。なお、このとき、CPU80は、受信した「選出物品（本）24」の位置情報をワールド座標系からロボット座標系に座標変換して利用する。

【0155】

「選出物品（本）24」が「花火の匠」というタイトル（名称）の本24である場合は、このステップS13では、図1において右から2番目に置かれている本を指差しジェス

10

20

30

40

50

チャで示すとともに、たとえば、「花火の匠の本ですね」のような発話を行う。

【0156】

このようなロボット12の指差しと発話に対し、人間16は、「そうです」とか「ちがいます」などといった応答を行う。するとロボット12では、CPU80が、マイク66を通して入力された人間16の声を取り込み、ステップS15で人間16の応答があったと判断する。そして、ステップS17では、取り込んだ人間16の音声の情報とともに、当該音声を音声認識して肯定する内容であるか否定する内容であるかを判断させる指示をサーバ20に送信する。

【0157】

サーバ20では、CPU200は、音声認識により肯定であるか否定であるかを判断させる指示と、音声認識すべき音声の情報を受信すると、図16のステップS53でYESと判断する。そして、次に、CPU200は、ステップS55で、受信した音声の情報に音声認識処理を施して、音声の内容が肯定を示すものであるか否定を示すものであるかを判断し、ステップS57でその判断結果をロボット12に送信する。

10

【0158】

ロボット12では、音声認識による肯定であるか否定であるかの判断結果を受信すると、CPU80は、ステップS19でYESと判断する。そして、さらに、ステップS21ではサーバ20における音声認識の結果が肯定を示すものであったか否かを判断する。ステップS21でNOと判断すると、CPU80は、ステップS11で、サーバ20から「選出物品(本)24」の位置情報と発話内容情報とをさらに受信したか否かを判断する。

20

【0159】

一方、ステップS21でYESと判断すると、ロボット12は、該当する物品(本)24の方向に移動し、該当する物品(本)24を把持して人間16の位置に運ぶ。つまり、物品(本)24が存在する位置の座標が既にわかっているため、ロボット12のCPU80は、車輪モータ36を制御して、ロボット12をその物品(本)24の位置に移動させ、次いでアクチュエータ108(図3)を制御することによってハンド56R(または56L)を開閉して物品(本)24をハンド56R(または56L:図2)で把持させ、その状態で再び車輪モータ36を制御してロボット12を人間16の位置にまで移動させる。このようにして、ステップS23で、サーバ20が図16のステップS43で選出した「選出物品(本)24」を人間16に運ぶことができる。

30

【0160】

一方、サーバ20では、図16のステップS57で、音声認識による人間16の音声の内容が肯定であるか否定であるかの判断結果をロボット12に送信した後、図17のステップS61で、当該判断の結果が肯定であったか否かを判断する。ステップS61でNOと判断すると、CPU200は、図16のステップS43で、音声認識の結果ならびに視線確信度と指差し方向確信度から作成した前述の候補物品一覧Cから、次の候補である物品(本)24の選出を行う。

【0161】

このとき、候補物品一覧Cに他の物品(本)24のIDが登録されていない場合には、ステップS45で物品が選出されなかったと判断される(ステップS45でNO)。すると、CPU200は、ステップS59で候補物品がない旨の通知をロボット12に送信する。

40

【0162】

ロボット12では、候補物品がない旨の通知を受信すると、CPU200が、図15のステップS25でYESと判断し、人間16が指示する物品(本)24を特定できなかったものとして処理を終了する。

【0163】

一方、図17に戻って、ステップS61でYESと判断すると、次に、CPU200は、ステップS63で、人間16の指示する物品(本)24を特定することができたことを示すために、メモリ204に設定されているフラグscに数値「1」を記憶させてフラグ

50

s cをオン状態にする。なお、フラグs cの初期状態はオフ状態（数値「0」）である。

【0164】

次に、CPU200は、ステップS65で、メモリ204に設定されているフラグnwがオン状態であるか否かを判断する。先に説明したように、このフラグnwは、人間16が発話した音声を音声認識した結果、音声認識できない単語が存在し、その単語が、人間16が指示した物品を特定するであろう単語である場合にオン状態に設定される。

【0165】

ステップS65でYESと判断すると、CPU200は、ステップS67で、音声認識できなかった単語を物品辞書122に登録して辞書を更新する。具体的には、図19のステップS119で、メモリ204のワークエリアwaに格納された音声認識できなかった単語の音素記号列を、物品辞書122に登録する。たとえば、人間16が「ハナタク持ってきて」と発話し、「ハナタク」が音声認識できなかったが、視線や指差しに基づく推定により、人間16は「花火の匠」の本を指示していたと判明したとする。この場合、ワークエリアwaには単語「ハナタク」の音素記号列“hanataku”が格納されているので、この音素記号列を、図9に示すように、物品辞書122の物品のIDが「123…0000046」である「花火の匠」という本24のレコードの「属性」の項目に、この「ハナタク」という単語を発話した人間16のIDとともに記憶する。

【0166】

次に、CPU200は、ステップS69で、音声認識できなかった単語を音声認識辞書126に新たなレコードとして登録して辞書を更新する。具体的には、ステップS67での例をそのまま用いると、図12に示すように、音声認識辞書126に、図16のステップS43で選出された物品（本）24（「選出物品（本）24」）の名称の単語である「花火の匠」を記憶した項目と、メモリ204のワークエリアwaに格納されている音素記号列である“hanataku”を記憶した項目とを含むレコードを追加する。このとき、当該レコードの「物品名称」の項目には、物品の名称（本24のタイトル（名称））であることを示す値「1」が格納される。

【0167】

このように音声認識辞書122に「花火の匠」という単語と“hanataku”という音素記号列を登録しておく、次回以降において、人間16が、たとえば、「ハナタク持ってきて」と発話すると、「ハナタク」を音声認識することができ、その音声認識の結果として物品（本）24の名称である「花火の匠」を得ることができる。

【0168】

ステップS71では、メモリ204に設定されている個人正誤DB205Eを更新する。つまり、メモリ204に設定されているフラグs cがオン状態であるか否かを判断し、オン状態である場合には、人間16が指示する物品（本）24を特定することができたということを示している、個人正誤DB205Eに当該人間16とのコミュニケーションが成功したことを示す情報を記憶する。一方、フラグs cがオフ状態である場合には、当該人間16とのコミュニケーションが不成功であったことを示す情報を記憶する。

【0169】

なお、ステップS65でNOと判断した場合は、ステップS67およびステップS69をスキップする。

【0170】

このように、システム10では、人間16の指示する物品（本）24を特定すると、当該物品（本）24に関連する単語、つまり、物品ローカル辞書205Aに当該物品（本）24のIDに関連付けて記憶されている単語（「特定単語」）を用いた音声で、当該物品（本）24が人間16の指示した物品（本）24であるかをロボット12が確認する。

【0171】

上述したように、この「特定単語」は、物品ローカル辞書205Aにおいて人間16の指示する物品（本）24以外の他の物品（本）24のIDに関連付けられていない単語である。つまり、「特定単語」は人間16の近傍に存在する他の物品（本）24に関連付け

10

20

30

40

50

られていない単語である。また、この「特定単語」は、音声認識による認識率の高い単語である。そのため、次回に、人間 16 が同一物品(本) 24 を指示する場合に、この「特定単語」を利用すれば、システム 10 は容易に人間 16 の指示する物品(本) 24 を特定することができる。人間 16 はロボット 12 が用いる「特定単語」を真似して用いることが期待できるので、このシステム 10 によれば、人間 16 にシステム 10 が物品(本) 24 を特定しやすい単語を用いさせることができる。

【図面の簡単な説明】

【0172】

【図 1】図 1 はこの発明の一実施例を示すコミュニケーションロボットシステムの概要を示す図解図である。

10

【図 2】図 2 は図 1 に示すロボットの外観を正面から見た図解図である。

【図 3】図 3 は図 1 に示すロボットの電氣的な構成を示す図解図である。

【図 4】図 4 は図 1 に示すサーバの電氣的な構成を示す図解図である。

【図 5】図 5 は図 1 の実施例で用いられる物品辞書の一例を示す図解図である。

【図 6】図 6 は図 1 の実施例で用いられる物品ローカル辞書の一例を示す図解図である。

【図 7】図 7 は図 1 の実施例で用いられる単語の一覧の一例を示す図解図である。

【図 8】図 8 は図 1 の実施例で用いられる単語の一覧の一例を示す図解図である。

【図 9】図 9 は図 1 の実施例で用いられる物品辞書の一例を示す図解図である。

【図 10】図 10 は図 1 の実施例で用いられる音声認識辞書の一例を示す図解図である。

【図 11】図 11 は図 1 の実施例で用いられる音声認識ローカル辞書の一例を示す図解図である。

20

【図 12】図 12 は図 1 の実施例で用いられる音声認識辞書の一例を示す図解図である。

【図 13】図 13 は人間の視線および指差し方向を示す図解図である。

【図 14】図 14 は図 1 の実施例で用いられる確信度表の一例を示す図解図である。

【図 15】図 15 は図 1 の実施例におけるロボットの動作を示すフロー図である。

【図 16】図 16 は図 1 の実施例におけるサーバの動作を示すフロー図である。

【図 17】図 17 は図 1 の実施例におけるサーバの動作を示すフロー図である。

【図 18】図 18 は図 1 の実施例におけるサーバの動作を示すフロー図である。

【図 19】図 19 は図 1 の実施例におけるサーバの動作を示すフロー図である。

【図 20】図 20 は図 1 の実施例におけるサーバの動作を示すフロー図である。

30

【図 21】図 21 は図 1 の実施例におけるサーバの動作を示すフロー図である。

【図 22】図 22 は図 1 の実施例におけるサーバの動作を示すフロー図である。

【符号の説明】

【0173】

10 …コミュニケーションロボットシステム

12 …コミュニケーションロボット

14 …ネットワーク

18 …無線タグ

20 …サーバ

24 …物品(本)

40

80 …CPU

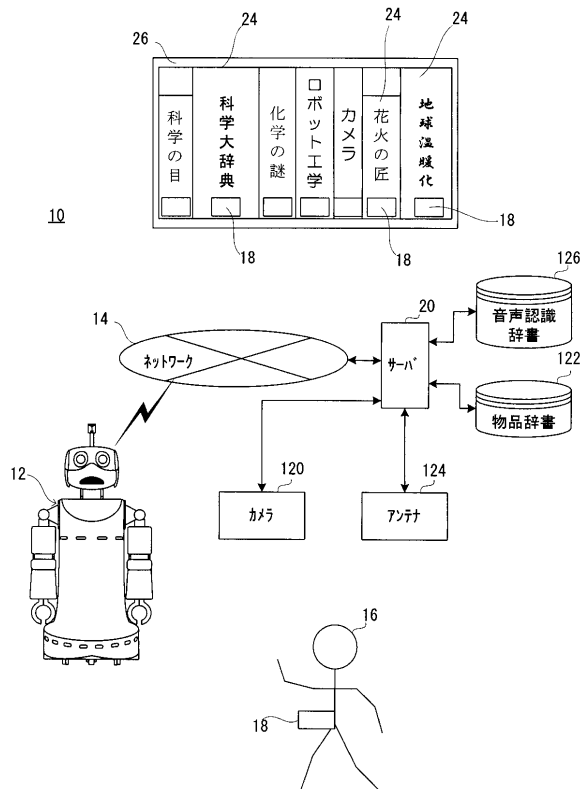
120 …カメラ

124 …アンテナ

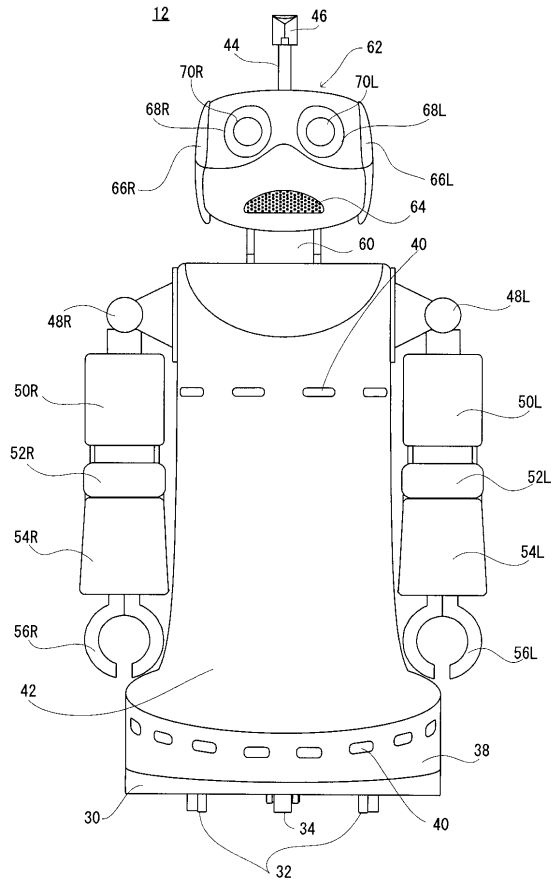
200 …CPU

208 …無線タグ読取装置

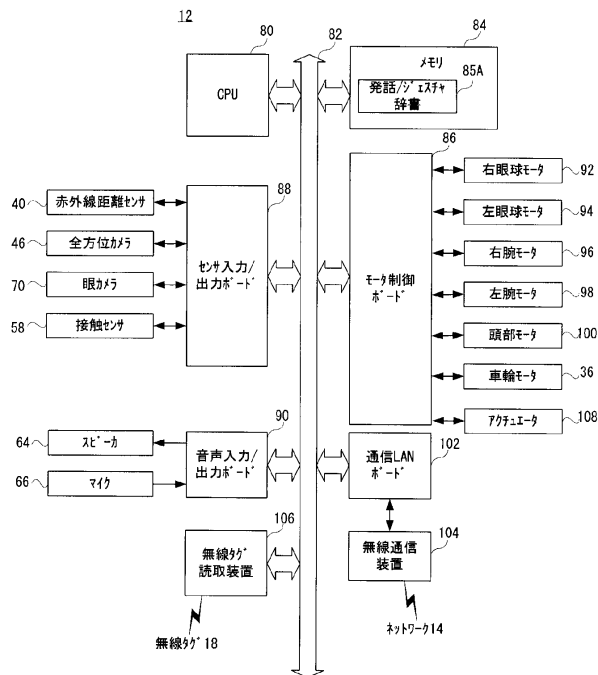
【図 1】



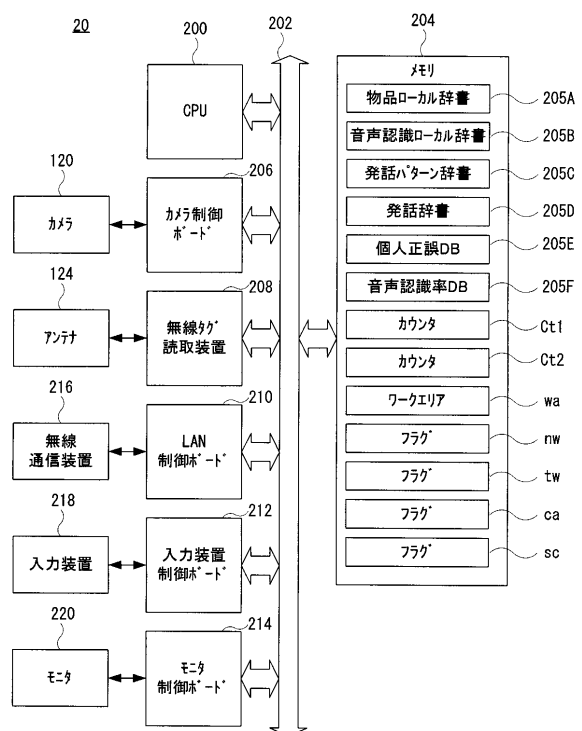
【図 2】



【図 3】



【図 4】



【図 5】

物品辞書 1 2 2

ID	文字列
123...0000001	名称: 地球温暖化 属性: 赤色、ハートカハ 著者: 山田太郎、鈴木一郎 出版社: ATR出版株式会社
123...0000002	名称: 大人の趣味 属性: 白色、雑誌、5月号 出版社: ATR出版株式会社
...	...
123...0000035	名称: カハラ 属性: 白色、雑誌、11月号 出版社: ATR出版株式会社
...	...
123...0000046	名称: 花火の匠 属性: 黒色、ハートカハ 著作: 近藤四郎 出版社: ATR出版株式会社
...	...
123...0000073	名称: ホット工学 属性: 青色、ハートカハ 著者: 田中二郎 出版社: ATR出版株式会社
...	...
123...0000091	名称: 化学の謎 属性: 白色、ハートカハ、" bakenazo" (456...0000004) 著者: 高橋三彦 出版社: ATR出版株式会社
...	...
123...0000102	名称: 科学の目 属性: 白色、ハートカハ、" saiai" (456...0000003) 著者: 斉藤五郎 出版社: ATR出版株式会社
...	...
123...0000132	名称: 科学大辞典 属性: 茶色、ハートカハ 出版社: ATR出版株式会社
...	...

【図 6】

物品辞書 2 0 5 A

ID	文字列
123...0000001	名称: 地球温暖化 属性: 赤色、ハートカハ 著者: 山田太郎、鈴木一郎 出版社: ATR出版株式会社
123...0000035	名称: カハラ 属性: 白色、雑誌、11月号 出版社: ATR出版株式会社
123...0000046	名称: 花火の匠 属性: 黒色、ハートカハ 著作: 近藤四郎 出版社: ATR出版株式会社
123...0000073	名称: ホット工学 属性: 青色、ハートカハ 著者: 田中二郎 出版社: ATR出版株式会社
123...0000091	名称: 化学の謎 属性: 白色、ハートカハ、" bakenazo" (456...0000004) 著者: 高橋三彦 出版社: ATR出版株式会社
123...0000102	名称: 科学の目 属性: 白色、ハートカハ、" saiai" (456...0000003) 著者: 斉藤五郎 出版社: ATR出版株式会社
123...0000132	名称: 科学大辞典 属性: 茶色、ハートカハ 出版社: ATR出版株式会社

【図 7】

物品辞書に登録されている単語の一覧 3 0 0

単語(音素記号列)
地球温暖化
赤色
ハートカハ
山田太郎
...
化学の謎
白色
ハートカハ
"bakenazo"(456...0000004)
...
科学の目
白色
ハートカハ
"saiai"(456...0000003)
斉藤五郎
...
花火の匠
...

【図 8】

物品辞書に登録されている単語の一覧 3 0 0

単語(音素記号列)
地球温暖化
赤色
ハートカハ
山田太郎
...
化学の謎
白色
"bakenazo"(456...0000004)
...
科学の目
斉藤五郎
...
花火の匠
...

【図 9】

物品辞書 122 (更新後)

ID	文字列
123...0000001	名称: 地球温暖化 属性: 赤色、ハードカバー 著者: 山田太郎、鈴木一郎 出版社: ATR出版株式会社
123...0000002	名称: 大人の趣味 属性: 白色、雑誌、5月号 出版社: ATR出版株式会社
...	...
123...0000035	名称: カラ 属性: 白色、雑誌、11月号 出版社: ATR出版株式会社
...	...
123...0000046	名称: 花火の匠 属性: 黒色、ハードカバー、"hanataku" (456...0000004) 著作: 近藤四郎 出版社: ATR出版株式会社
...	...
123...0000073	名称: ロボット工学 属性: 青色、ハードカバー 著者: 田中二郎 出版社: ATR出版株式会社
...	...
123...0000091	名称: 化学の謎 属性: 白色、ハードカバー、"bakenazo" (456...0000004) 著者: 高橋三郎 出版社: ATR出版株式会社
...	...
123...0000102	名称: 科学の目 属性: 白色、ハードカバー、"saiai" (456...0000003) 著者: 斉藤五郎 出版社: ATR出版株式会社
...	...
123...0000132	名称: 科学大辞典 属性: 黒色、ハードカバー 出版社: ATR出版株式会社
...	...

【図 10】

音声認識辞書 126

単語(テキスト)	単語(音素記号)	物品名称
赤色	"akairo"	0
...	...	...
大人の趣味	"otonanosyumi"	1
...	...	...
黄色	"kiro"	0
...	...	...
化学の謎	"kagakunonazo"	1
化学の謎	"bakenazo"	1
...	...	...
化学の目	"kagakunome"	1
化学の目	"saiai"	1
...	...	...
白色	"shiroiro"	0
...	...	...
地球温暖化	"chikyuundanka"	1
...	...	...
ハードカバー	"haadokabaa"	0
花火の匠	"hanabinotakumi"	1
...	...	...

【図 11】

音声認識ローカル辞書 205B

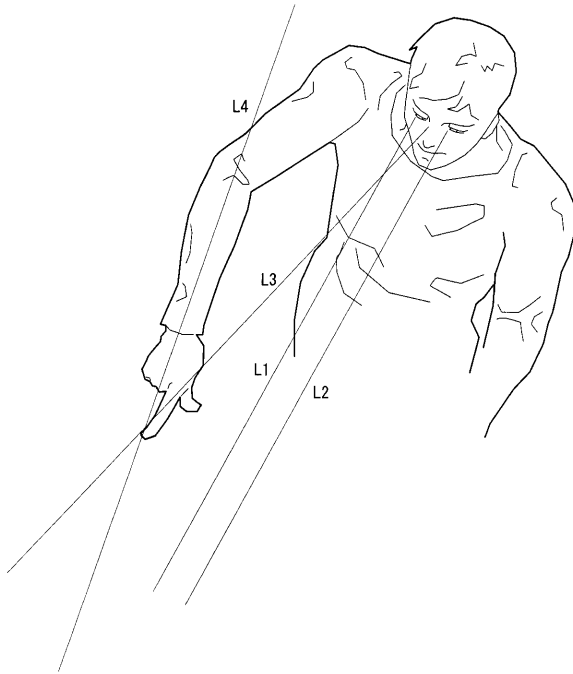
単語(テキスト)	単語(音素記号)	物品名称
地球温暖化	"chikyuundanka"	1
赤色	"akairo"	0
ハードカバー	"haadokabaa"	0
...	...	...
化学の謎	"kagakunonazo"	1
白色	"shiroiro"	0
化学の謎	"bakenazo"	1
...	...	...
化学の目	"kagakunome"	1
...	...	...
花火の匠	"hanabinotakumi"	1
...	...	...

【図 12】

音声認識辞書 126 (更新後)

単語(テキスト)	単語(音素記号)	物品名称
赤色	"akairo"	0
...	...	...
大人の趣味	"otonanosyumi"	1
...	...	...
黄色	"kiro"	0
...	...	...
化学の謎	"kagakunonazo"	1
化学の謎	"bakenazo"	1
...	...	...
化学の目	"kagakunome"	1
化学の目	"saiai"	1
...	...	...
白色	"shiroiro"	0
...	...	...
ハードカバー	"haadokabaa"	0
花火の匠	"hanabinotakumi"	1
花火の匠	"hanataku"	1
...	...	...

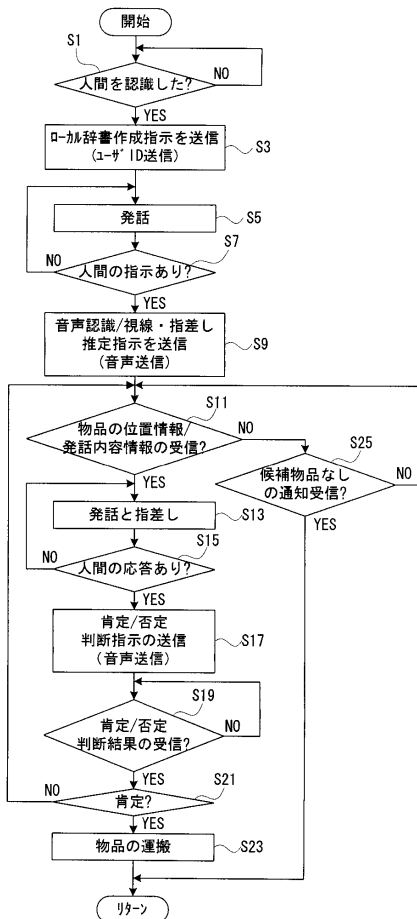
【図 1 3】



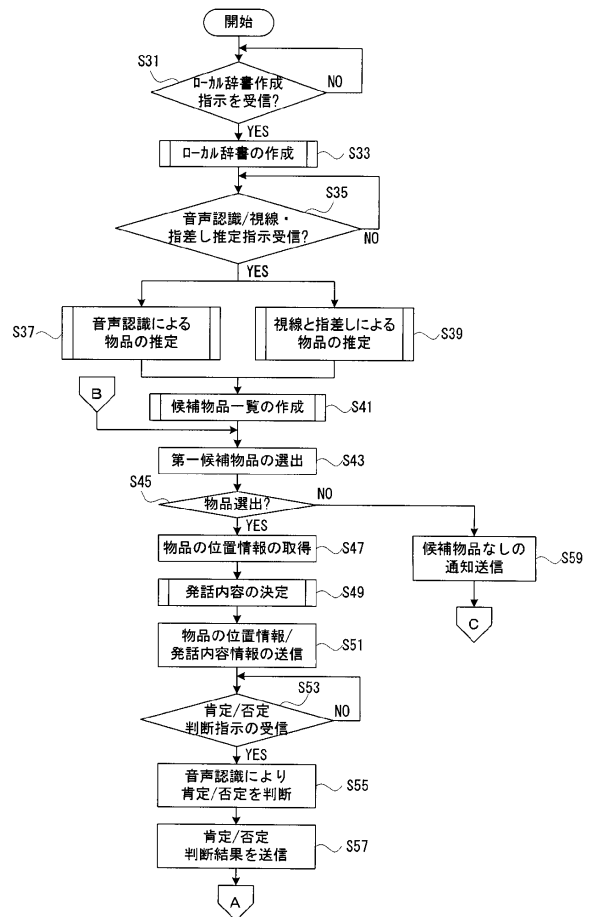
【図 14】

123...0000001	時間	L1	L2	L3	L4
	t1	-	○	-	-
	t2	-	-	-	-
	t3	-	-	-	-
	...	...	...	...	...
	tm	-	-	-	-
123...0000046	時間	L1	L2	L3	L4
	t1	-	-	-	-
	t2	-	-	○	-
	t3	-	-	○	-
	...	...	...	...	...
	tm	○	-	-	-
...	...	...	...	...	...

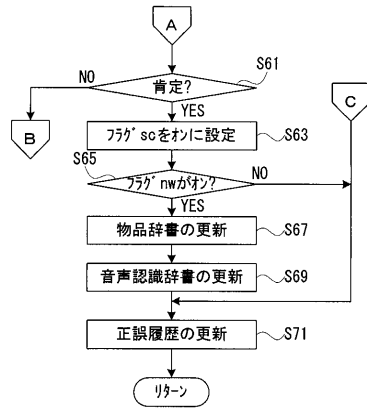
【図 1 5】



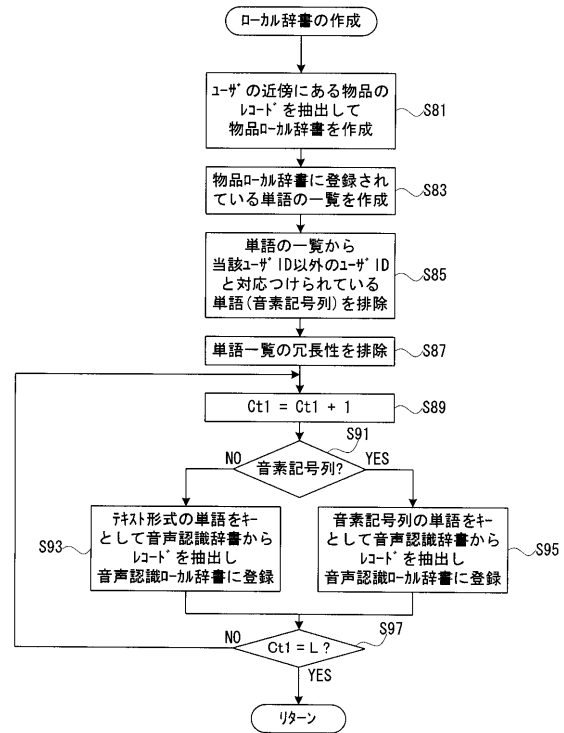
【图 16】



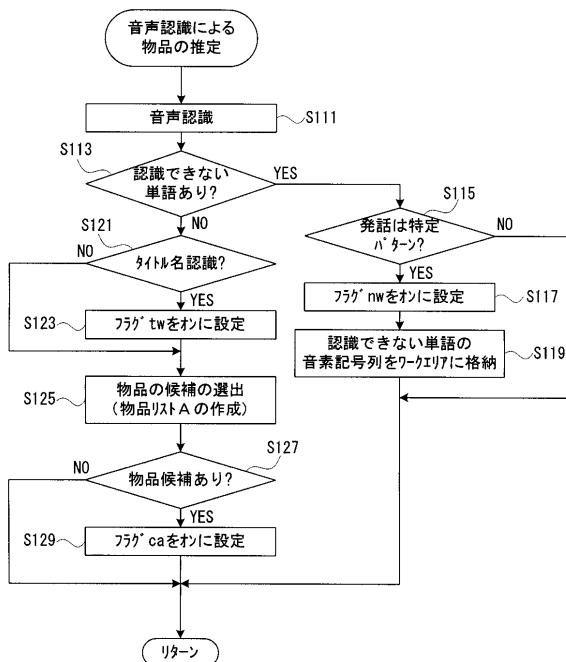
【図 17】



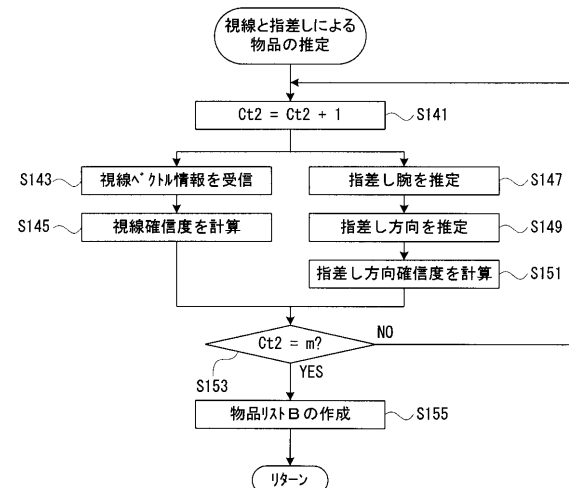
【図 18】



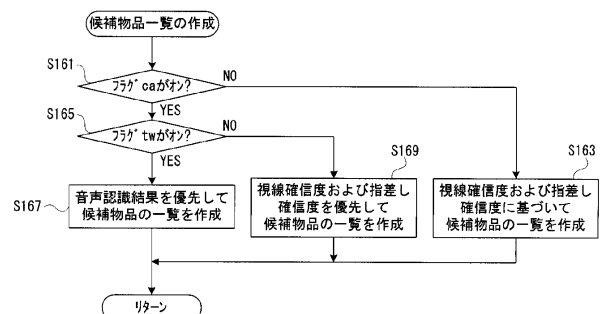
【図 19】



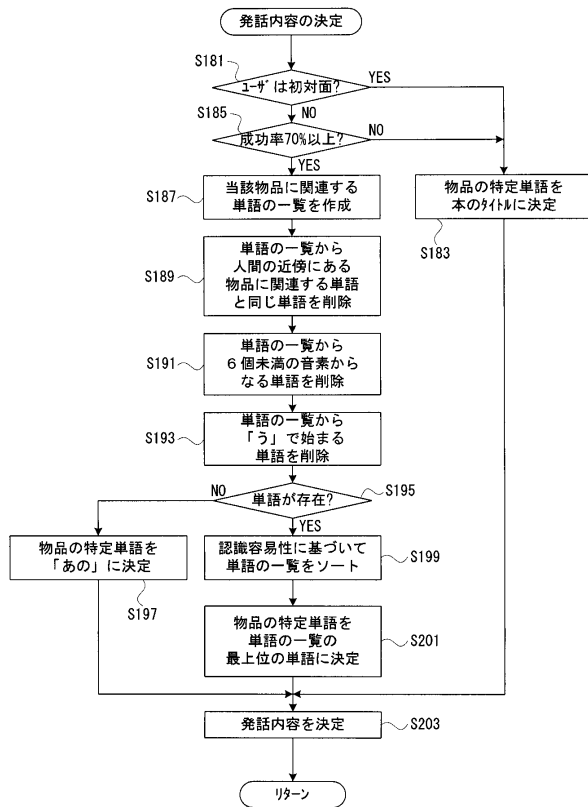
【図 20】



【図 21】



【図 22】



フロントページの続き

- (72)発明者 塩見 昌裕
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 秋本 高明
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 萩田 紀博
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官 毛利 太郎

- (56)参考文献 特開2004-240319 (JP, A)
特開2003-169147 (JP, A)
特開平08-263657 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G10L 15/00-15/28