(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6673276号
(P6673276)

(45)発行日　令和2年3月25日(2020.3.25)　(24)登録日　令和2年3月9日(2020.3.9)

(51)Int.Cl.		F I		
G10L 25/84	(2013.01)	G10L	25/84	
G10L 25/51	(2013.01)	G10L	25/51	400
G10L 15/00	(2013.01)	G10L	15/00	200H

請求項の数 9　(全 26 頁)

(21)出願番号　特願2017-62756(P2017-62756)
(22)出願日　平成29年3月28日(2017.3.28)
(65)公開番号　特開2018-165759(P2018-165759A)
(43)公開日　平成30年10月25日(2018.10.25)
審査請求日　平成30年12月7日(2018.12.7)

(73)特許権者　000001443
カシオ計算機株式会社
東京都渋谷区本町１丁目６番２号
(74)代理人　100095407
弁理士　木村　満
(72)発明者　島田　敬輔
東京都羽村市栄町３－２－１　カシオ計算機株式会社　羽村技術センター内
(72)発明者　中込　浩一
東京都羽村市栄町３－２－１　カシオ計算機株式会社　羽村技術センター内
(72)発明者　山谷　崇史
東京都羽村市栄町３－２－１　カシオ計算機株式会社　羽村技術センター内

最終頁に続く

(54)【発明の名称】音声検出装置、音声検出方法、及びプログラム

(57)【特許請求の範囲】
【請求項１】
音声を検出する音声検出手段と、
前記音声検出手段により検出された音声である検出音声の音声発生源が特定の音声発生源であるか否かを判別する第１判別手段と、
前記第１判別手段の判別結果に基づいて自機を制御する制御手段と、
前記検出音声が発生した方向を判別する第２判別手段と、
前記特定の音声発生源以外の他の音声発生源の位置を示す情報を含む音声発生源位置情報を記憶した記憶部と、
前記第２判別手段による判別結果と前記記憶された音声発生源位置情報とに基づいて、前記自機に対する前記検出音声が発生した方向に前記他の音声発生源が存在するか否かを判別する第３判別手段と、を備え、
前記制御手段は、前記第３判別手段により前記検出音声が発生した方向に前記他の音声発生源が存在しないと判別されている場合に、前記自機の動作を制御する、
音声検出装置。
【請求項２】
前記制御手段は、前記第１判別手段により、前記検出された音声の音声発生源が前記特定の音声発生源であると判別された場合、前記自機の位置及び姿勢の少なくとも一方に関する制御を実行する、
請求項１に記載の音声検出装置。

【請求項３】
撮像部をさらに備え、
前記第３判別手段により前記検出音声が発生した方向に前記他の音声発生源が存在しないと判別されている場合に、前記制御手段は、前記自機の動作の制御として、前記撮像部の撮像方向を前記第２判別手段が判別した方向に向けるように、前記自機の位置及び姿勢の少なくとも一方に関する制御を実行する、
請求項１または２に記載の音声検出装置。
【請求項４】
前記音声発生源位置情報は、前記自機の周囲の複数の位置の各々に、前記特定の音声発生源以外の音声発生源が存在する確率を示す情報を含む、
請求項１から３のいずれか１項に記載の音声検出装置。
【請求項５】
前記制御手段により前記自機を移動させている間に、前記撮像部が撮像した画像から認識された音声発生源の位置を示す情報を、前記音声発生源位置情報に追加する、
請求項３または請求項３に従属する請求項４に記載の音声検出装置。
【請求項６】
前記第１判別手段は、前記音声検出手段により検出された前記音声が前記自機宛てに発せられた音声か否かを判別し、前記自機宛ての音声であると判別した場合、前記音声検出手段が検出した音声の音声発生源が前記特定の音声発生源であるか否かを判別する、
請求項１から５のいずれか１項に記載の音声検出装置。
【請求項７】
音声を検出する音声検出手段と、
前記音声検出手段により検出された音声の音声発生源が特定の音声発生源であるか否かを判別する判別手段と、
前記判別手段の判別結果に基づいて自機を制御する制御手段と、
撮像部と、
前記特定の音声発生源以外の音声発生源であって、登録された音声発生源の位置を示す情報を含む音声発生源位置情報をあらかじめ記憶した記憶部と、を備え、
前記判別手段は、前記音声検出手段により検出された前記音声の音声発生源の位置を判別し、前記判別された位置が、前記音声発生源位置情報に含まれる前記登録された音声発生源の位置であるか否かを判別し、
前記判別手段により判別された位置が、前記音声発生源位置情報に含まれる前記登録された音声発生源の位置でないと判別した場合に、前記制御手段は、前記撮像部の撮像方向を前記判別手段が判別した位置に向けるように、自機の位置、姿勢の少なくとも一方を変える、
音声検出装置。
【請求項８】
ロボットに搭載されたコンピュータが音声を検出する音声検出方法であって、
音声を検出する音声検出ステップと、
前記音声検出ステップで検出された音声である検出音声の音声発生源が特定の音声発生源であるか否かを判別する第１判別ステップと、
前記第１判別ステップの判別結果に基づいて、前記ロボットの動作を制御する制御ステップと、
前記検出音声が発生した方向を判別する第２判別ステップと、
前記第２判別ステップによる判別結果と、記憶部に記憶された、前記特定の音声発生源以外の他の音声発生源の位置を示す情報を含む音声発生源位置情報とに基づいて、前記ロボットに対する前記検出音声が発生した方向に前記他の音声発生源が存在するか否かを判別する第３判別ステップと、を備え、
前記制御ステップでは、前記第３判別ステップにより前記検出音声が発生した方向に前記他の音声発生源が存在しないと判別されている場合に、前記ロボットの動作を制御する

、
音声検出方法。
【請求項９】
ロボットに搭載されたコンピュータに、
音声を検出する音声検出機能と、
前記音声検出機能により検出された音声である検出音声の音声発生源が特定の音声発生源であるか否かを判別する第１判別機能と、
前記第１判別機能により判別された判別結果に基づいて、前記ロボットの動作を制御する制御機能と、
前記検出音声が発生した方向を判別する第２判別機能と、
前記第２判別機能による判別結果と、記憶部に記憶された、前記特定の音声発生源以外の他の音声発生源の位置を示す情報を含む音声発生源位置情報とに基づいて、前記ロボットに対する前記検出音声が発生した方向に前記他の音声発生源が存在するか否かを判別する第３判別機能と、を実現させ、
前記制御機能は、前記第３判別機能により前記検出音声が発生した方向に前記他の音声発生源が存在しないと判別されている場合に、前記ロボットの動作を制御する、
プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声検出装置、音声検出方法、及びプログラムに関する。
【背景技術】
【０００２】
人間、動物等に模した形態を有し、人間と会話等のコミュニケーションをすることができるロボットが知られている。このようなロボットには、自機に搭載されたマイクの出力に基づいてロボットの周囲に発生した音を検出し、人の声であると判別すると、人がいる方向にロボットの顔の向きあるいは体の向きを変え、その人に話しかける、手を振る等の動作をするものもある。
【０００３】
特許文献１には、ロボットが、マイクロホンに閾値以上の振幅の音が入力されることにより、音イベントが発生したことを検出し、音源方向を推定して、推定した音源方向に振り向くことが記載されている。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開２００３－２６６３５１号公報
【非特許文献】
【０００５】
【非特許文献１】Ａｎｄｒｅｗ Ｊ．Ｄａｖｉｓｏｎ，“Ｒｅａｌ－Ｔｉｍｅ Ｓｉｍｕｌｔａｎｅｏｕｓ Ｌｏｃａｌｉｚａｔｉｏｎ ａｎｄ Ｍａｐｐｉｎｇ ｗｉｔｈ ａ Ｓｉｎｇｌｅ Ｃａｍｅｒａ”，Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ｔｈｅ ９ｔｈ ＩＥＥＥ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏｎ Ｃｏｍｐｕｔｅｒ Ｖｉｓｉｏｎ Ｖｏｌｕｍｅ ２，２００３，ｐｐ．１４０３－１４１０
【非特許文献２】Ｒｉｃｈａｒｄ Ｈａｒｔｌｅｙ，Ａｎｄｒｅｗ Ｚｉｓｓｅｒｍａｎ，“Ｍｕｌｔｉｐｌｅ Ｖｉｅｗ Ｇｅｏｍｅｔｒｙ ｉｎ Ｃｏｍｐｕｔｅｒ Ｖｉｓｉｏｎ”，Ｓｅｃｏｎｄ Ｅｄｉｔｉｏｎ，Ｃａｍｂｒｉｄｇｅ．Ｕｎｉｖｅｒｓｉｔｙ Ｐｒｅｓｓ，Ｍａｒｃｈ ２００４，ｃｈａｐｔｅｒ９
【非特許文献３】Ｃｓｕｒｋａ，Ｇ．，Ｄａｎｃｅ，Ｃ．Ｒ．，Ｆａｎ，Ｌ．，Ｗｉｌｌａｍｏｗｓｋｉ，Ｊ．ａｎｄ Ｂｒａｙ，Ｃ．：Ｖｉｓｕａｌ ｃａｔｅｇｏｒｉｚａｔｉｏｎ ｗｉｔｈ ｂａｇｓ ｏｆ ｋｅｙｐｏｉｎｔｓ，ＥＣＣＶ Ｉｎｔｅｒｎａｔ

ｉｏｎａｌ Ｗｏｒｋｓｈｏｐ ｏｎ Ｓｔａｔｉｓｔｉｃａｌ Ｌｅａｒｎｉｎｇ ｉｎ Ｃｏｐｕｔｅｒ Ｖｉｓｉｎ（２００４）

【発明の概要】

【発明が解決しようとする課題】

【０００６】

しかしながら、特許文献１に記載されているロボットは、音イベントを検出すると振り向くので、実際にロボットに対して人から発せられた音だけではなく、例えば、テレビ、ラジオ等の電子機器のスピーカから出力される音声にも反応してしまうことが予想される。

【０００７】

本発明は、上記実情を鑑みてなされたものであり、ロボットに実際の人から直接発せられた音声か電子機器のスピーカから出力された音声かを判別させることで、ロボットの無駄な動作を減らすことを目的とする。

【課題を解決するための手段】

【０００８】

上記目的を達成するため、本発明に係る音声検出装置は、

音声を検出する音声検出手段と、

前記音声検出手段により検出された音声である検出音声の音声発生源が特定の音声発生源であるか否かを判別する第１判別手段と、

前記第１判別手段の判別結果に基づいて自機を制御する制御手段と、

前記検出音声が発生した方向を判別する第２判別手段と、

前記特定の音声発生源以外の他の音声発生源の位置を示す情報を含む音声発生源位置情報を記憶した記憶部と、

前記第２判別手段による判別結果と前記記憶された音声発生源位置情報とに基づいて、前記自機に対する前記検出音声が発生した方向に前記他の音声発生源が存在するか否かを判別する第３判別手段と、を備え、

前記制御手段は、前記第３判別手段により前記検出音声が発生した方向に前記他の音声発生源が存在しないと判別されている場合に、前記自機の動作を制御する。

【発明の効果】

【０００９】

本発明によれば、ロボットに実際の人から直接発せられた音声か電子機器のスピーカから出力された音声かを判別させることで、ロボットの無駄な動作を減らすことができる。

【図面の簡単な説明】

【００１０】

【図１】本発明の実施の形態１にかかるロボットの外観図である。

【図２】ロボットの頭の自由度を説明するための図である。

【図３】ロボットの構成を示すブロック図である。

【図４】部屋内のロボットとユーザの位置の一例を示す図である。

【図５】地図作成処理のフローチャートである。

【図６】呼びかけ応答処理のフローチャートである。

【図７】音源定位の処理のフローチャートである。

【図８】仮の音源の位置を説明するための図である。

【図９】自機位置推定の処理のフローチャートである。

【図１０】実施の形態２にかかるロボットの記憶部の構成を示すブロック図である。

【図１１】呼びかけ移動処理のフローチャートである。

【図１２】顔位置推定の処理のフローチャートである。

【発明を実施するための形態】

【００１１】

（実施の形態１）

以下、図面を参照しながら本発明の実施の形態１について説明する。図１は、実施の形

態１に係るロボット１００を正面から見た場合の外観を模式的に示した図である。ロボット１００は、頭１１０と胴体１２０とを備えた人型のコミュニケーションロボットである。ロボット１００は、住宅内に設置されており、住人に呼びかけられると、呼びかけた住人と会話する。

【００１２】

図１に示すように、ロボット１００の頭１１０には、カメラ１１１と、マイク１１２と、スピーカ１１３と、が設けられている。

【００１３】

カメラ１１１（撮像手段）は、頭１１０の前面の下側、人の顔でいうところの鼻の位置に設けられている。カメラ１１１は、後述する制御部１２７の制御の下、撮像を行う。

【００１４】

マイク１１２は、１３個のマイクを含む。１３個のマイクのうちの８個のマイクが、人の顔でいうところの額の高さの位置であって、頭１１０の周周りに等間隔で配置されている。これら８個のマイクより上側に、４個のマイクが頭１１０の周回りに等間隔で配置されている。さらに、１個のマイクが頭１１０の頭頂部に配置されている。マイク１１２はロボット１００の周囲で発生した音を検出する。マイク１１２は、後述の制御部１２７と協働して、音声検出手段としての役割を果たす。

【００１５】

スピーカ１１３は、カメラ１１１より下側、人の顔でいうところの口に相当する位置に設けられている。スピーカ１１３は、後述する制御部１２７の制御の下、各種の音声を出力する。

【００１６】

首関節１２１は、頭１１０と胴体１２０とを連結する部材である。頭１１０は、破線で示される首関節１２１によって、胴体１２０に連結されている。首関節１２１は、複数のモータを含む。後述する制御部１２７がこれら複数のモータを駆動すると、ロボット１００の頭１１０が回転する。図２にロボット１００の頭１１０の回転の自由度を模式的に表した図を示す。首関節１２１により、ロボット１００の頭１１０は、胴体１２０に対して、ピッチ軸Ｘｍの軸回り、ロール軸Ｚｍの軸回り、ヨー軸Ｙｍの軸回り回転可能である。首関節１２１は、後述の足回り部１２６とともに、後述の制御部１２７と協働して、ロボット１００の各部位の動作を制御することで、自機の位置、姿勢の少なくとも一方を変える制御手段としての役割を果たす。

【００１７】

図３を参照する。上述の構成に加え、ロボット１００は、操作ボタン１２２と、センサ群１２３と、電源部１２４と、記憶部１２５と、足回り部１２６と、制御部１２７と、を備える。

【００１８】

操作ボタン１２２は、胴体１２０の背中に設けられている（図１において不図示）。操作ボタン１２２は、ロボット１００を操作するための各種のボタンであり、電源ボタン、スピーカ１１３の音量調節ボタン等を含む。

【００１９】

図１に示すように、センサ群１２３は、人の顔でいうところの目の位置と耳の位置とに設けられている。センサ群１２３は、距離センサ、加速度センサ、障害物検知センサ等を含み、ロボット１００の姿勢制御や、安全性の確保のために使用される。

【００２０】

図３を参照する。電源部１２４は、胴体１２０に内蔵された充電池であり、ロボット１００の各部に電力を供給する。

【００２１】

記憶部１２５は、ハードディスクドライブ、フラッシュメモリ等を含み、胴体１２０の内部に設けられている。記憶部１２５は、後述の制御部１２７によって実行されるプログラム、カメラ１１１が撮像した画像データ等を含む各種データを記憶する。記憶部１２５

が記憶するプログラムには、後述の呼びかけ応答処理に係る呼びかけ応答プログラム１２５１、地図作成処理に係る地図作成プログラム１２５２が含まれる。さらに、記憶部１２５には、後述のＳＬＡＭ（Ｓｉｍｕｌｔａｎｅｏｕｓ Ｌｏｃａｌｉｚａｔｉｏｎ Ａｎｄ Ｍａｐｐｉｎｇ）法で作成される部屋の地図であるＳＬＡＭ地図１２５３、撮像画像の特徴点等を格納するフレームデータベース１２５４、後述のラベリングの音声発生確率が定義された音声発生確率データベース１２５５が含まれる。

【００２２】

足回り部１２６は、胴体１２０の下側に設けられた４つの車輪（ホイール）を含む。図１に示すように、４つの車輪のうち、２つが胴体１２０の前側に、残り２つが後ろ側に（不図示）が配置されている。車輪として、例えば、オムニホイール、メカナムホイールが使用される。後述の制御部１２７が足回り部１２６の車輪を回転させると、ロボット１００は移動する。足回り部１２６は、前述の首関節１２１とともに、後述の制御部１２７と協働して、ロボット１００の各部位の動作を制御することで、自機の位置、姿勢の少なくとも一方を変える制御手段としての役割を果たす。

【００２３】

さらに、足回り部１２６の車輪にはロータリエンコーダが設けられている。ロータリエンコーダで車輪の回転数を計測し、車輪の直径や車輪間の距離等の幾何学的関係を利用することで並進移動量及び回転量を計算できる。

【００２４】

図３を参照する。制御部１２７は、ＣＰＵ（Central Processing Unit）、ＲＡＭ（Random Access Memory）等で構成される。制御部１２７は、上述のロボット１００の各部に接続されており、ＲＡＭをワークスペースとして、記憶部１２５に記憶されたプログラムを実行することにより、ロボット１００の各部を制御する。

【００２５】

本実施の形態においては、制御部１２７は、ロボット１００の各部位の動作を制御するため、前述の首関節１２１、足回り部１２６を制御することで、自機の位置、姿勢の少なくとも一方を変える制御手段の役割を果たす。

【００２６】

さらに、制御部１２７は、足回り部１２６の車輪に設けられたロータリエンコーダの回転数から、自機の位置（移動開始時の位置を基準とした自機の位置）を計測することができる。例えば、車輪の直径をＤ、回転数をＲ（足回り部１２６のロータリエンコーダにより測定）とすると、その車輪の接地部分での並進移動量はπ・Ｄ・Ｒとなる。また、車輪の直径をＤ、車輪間の距離をＩ、右車輪の回転数をＲＲ、左車輪の回転数をＲＬとすると、向き変更の回転量は（右回転を正とすると）３６０°×Ｄ×（ＲＬ－ＲＲ）／（２×Ｉ）となる。この並進移動量や回転量を逐次足し合わせていくことで、自機位置（移動開始時の位置及び向きを基準とした位置及び向き）を計測することができる。このように、制御部１２７は、オドメトリとしても機能する。

【００２７】

上述のように、ロボット１００は、住人（ユーザ）に呼びかけられると会話するので、呼びかけられたことを判別すると、呼びかけた住人（ユーザ）の顔検出を行う必要がある。以下、ロボット１００が行う顔検出の処理を説明する。ここでは、ユーザの呼びかけに応答する一連の処理（呼びかけ応答処理）の中で、ロボット１００がユーザの顔検出を行う例を説明する。図４に示すように、部屋ＲＭ内にロボット１００とユーザＰがおり、ロボット１００とユーザＰとが正対していない場合に、ユーザＰがロボット１００に呼びかける場面を想定する。

【００２８】

本実施の形態においては、部屋ＲＭ内に存在する音源の位置が登録された地図（音声発生源位置情報）があらかじめ作成されている。ロボット１００の制御部１２７は、人の声がしたことを検出したときに、まず、その音の音源の方向を判別する。そして、制御部１２７は、音源の方向と自機（ロボット１００）の位置とあらかじめ作成されている部屋Ｒ

Ｍの中の地図とに基づき、判別した音源の方向に、人以外の音源が存在するか否かを判別し、存在する否かに応じて、振り向くか振り向かないかを判別する。

【００２９】

呼びかけ応答処理に先立ってあらかじめ作成される実空間（ここでは部屋ＲＭ）内の地図の作成方法を説明する。制御部１２７の制御の下、ロボット１００は、毎日決められた時刻に、部屋の中を動き回りながら、撮像し、撮像画像に基づいて部屋の地図を作成し、作成した地図を記憶部１２５に格納する。

【００３０】

地図の作成には、ＳＬＡＭ法を採用する。ＳＬＡＭ法は、実空間の地図を作成するための手法のひとつである。この手法では、カメラの撮影する動画像の複数フレームから、同一の特徴点を追跡することで、自機の３次元位置（カメラ位置）と特徴点の３次元位置（これが集まって地図の情報を構成する）とを交互または同時に推定する処理を行う。ＳＬＡＭ法の詳細は、非特許文献１に記載されている。

【００３１】

以下、図５のフローチャートを参照しながら、制御部１２７が行うＳＬＡＭ法を採用した地図作成処理を説明する。制御部１２７は、記憶部１２５に記憶されている地図作成プログラム１２５２を実行することによって、以下の処理を実現する。

【００３２】

まず、制御部１２７は、撮像画像を取得し、撮像画像の二次元特徴点（２Ｄ特徴点）を抽出する（ステップＳ１１）。２Ｄ特徴点とは、画像中のエッジ部分など、画像内の特徴的な部分であり、ＳＩＦＴ（Scale-Invariant Feature Transform）やＳＵＲＦ（Speed-Up Robust Features）等のアルゴリズムを用いて取得することができる。

【００３３】

具体的には、ステップＳ１１において、制御部１２７は、カメラ１１１を制御して、撮像を行う。そして、撮像した画像から２Ｄ特徴点を抽出する。さらに、前述のようにオドメトリとしても機能する制御部１２７は、足回り部１２６のロータリエンコーダを使用して、自機（ロボット１００）の現在位置を計測する。制御部１２７は、２Ｄ特徴点と、自機の現在位置と、を撮像画像と対応づけて記憶部１２５に記憶する。

【００３４】

制御部１２７は、地図作成処理の開始後に撮像した画像が２枚以上であるか否かを判別する（ステップＳ１２）。２枚未満であると判別すると、（ステップＳ１２；Ｎｏ）、制御部１２７は、足回り部１２６を制御して、自機を所定の距離だけ移動し（ステップＳ１９）、再びステップＳ１１へ戻る。

【００３５】

一方、撮像した画像が２枚以上であると判別した場合（ステップＳ１２；Ｙｅｓ）、制御部１２７は、２つの画像の２Ｄ特徴点の対応を取得する（ステップＳ１３）。２つの画像は、例えば、今回撮像した画像と、直近に撮像した画像である。

【００３６】

ステップＳ１３で取得した２つの画像の対応する特徴点（対応特徴点）の個数が、閾値未満であるか否かを判別する（ステップＳ１４）。これは、取得した特徴点の個数が少ないと、後述のＴｗｏ－ｖｉｅｗ　Ｓｔｒｕｃｔｕｒｅ　ｆｒｏｍ　Ｍｏｔｉｏｎ法での計算ができないためである。

【００３７】

２つの画像の対応する特徴点の個数が、閾値未満であると判別した場合（ステップＳ１４；Ｎｏ）、制御部１２７は、足回り部１２６を制御して、自機を所定の距離だけ移動し（ステップＳ１９）、再びステップＳ１１へ戻る。

【００３８】

一方、２つの画像の対応する特徴点の個数が、閾値以上であると判別した場合（ステップＳ１４；Ｙｅｓ）、制御部１２７は、２つの画像間の姿勢を推定する（ステップＳ１５）。

【００３９】
具体的には、ステップＳ１５において、Ｔｗｏ－ｖｉｅｗ　Ｓｔｒｕｃｔｕｒｅ　ｆｒｏｍ　Ｍｏｔｉｏｎ法を用いて、２つの画像の間で対応する２Ｄ特徴点の２次元座標（２Ｄ座標）と、２つの画像のそれぞれの撮影位置（撮影時の自機の位置）の間の距離とから、２つの画像間の姿勢（それぞれの画像を取得した位置の差分（並進ベクトルｔ）及び向きの差分（回転行列Ｒ））を推定する。この推定は、非特許文献２に記載されているように、エピポーラ拘束式により、対応する特徴点から基礎行列Ｅを求め、基礎行列Ｅを並進ベクトルｔと回転行列Ｒとに分解することによって得られる。
【００４０】
続いて、制御部１２７は、２つの画像の間で対応する２Ｄ特徴点（２Ｄの対応特徴点）の３次元座標（３Ｄ座標）を推定する（ステップＳ１６）。具体的には、これは、ステップＳ１５で算出した２つの画像間の姿勢を表す値と、２つの画像の間で対応する２Ｄ特徴点の２Ｄ座標と、を用いて推定する。
【００４１】
制御部１２７は、ステップＳ１６で推定した推定値をデータベースに登録する（ステップＳ１７）。具体的には、制御部１２７は、ステップＳ１６で求めた「２Ｄの対応特徴点の３Ｄ座標（Ｘ、Ｙ、Ｚ）」と、「２Ｄ特徴点の特徴量」（例えばＳＩＦＴ等で得た特徴量）と、を記憶部１２５のＳＬＡＭ地図１２５３に登録する。
【００４２】
また、制御部１２７は、記憶部１２５のフレームデータベース１２５４に、画像の情報として、「ＳＬＡＭ地図内での画像の姿勢」（その画像を撮像したときの自機のＳＬＡＭ座標内での位置（並進ベクトルｔ）及び向き（回転行列Ｒ））と、「抽出した全ての２Ｄ特徴点」と、「すべての２Ｄ特徴点の中で３Ｄ位置（３Ｄ座標）が既知の点」と、「キーフレーム自体の特徴」と、を記憶部１２５のフレームデータベース１２５４に登録する。
【００４３】
ここで、キーフレームとは、処理の対象となる撮像画像のことである。キーフレーム自体の特徴とは、キーフレーム間の画像類似度を求める処理を効率化するためのデータであり、画像中の２Ｄ特徴点のヒストグラム等を用いてもよいし、画像自体を「キーフレーム自体の特徴」としてもよい。
【００４４】
制御部１２７は、処理が終了であると判別すると（ステップＳ１８；Ｙｅｓ）、地図作成処理を終了する。一方、処理が終了でないと判別すると（ステップＳ１８；Ｎｏ）、足回り部１２６を制御して、自機を所定の距離だけ移動し（ステップＳ１９）、再びステップＳ１１へ戻る。以上が地図作成処理である。
【００４５】
さらに、上述のように作成したＳＬＡＭ地図１２５３に、部屋ＲＭ内のそれぞれの位置における障害物が存在する確率を示す障害物情報として確率変数を付加してもよい。障害物情報の確率変数の値は、その値が高いほど、その位置に障害物がある可能性が高いことを表している。障害物情報の確率変数は、例えば、ＳＬＡＭ地図１２５３の作成処理におけるデータベース登録（図５のステップＳ１７）のタイミングで、ＳＬＡＭ地図１２５３に付加することができる。
【００４６】
さらに、本実施の形態においては、上述のように作成したＳＬＡＭ地図１２５３に、人以外の音声発生源情報を付加したものを使用して、ロボット１００が、検出した音が人であるか否かを判別する。
【００４７】
人以外の音声発生源情報は、例えば、ＳＬＡＭ地図１２５３の作成処理におけるデータベース登録（図５のステップＳ１７）のタイミングで、ＳＬＡＭ地図１２５３に付加することができる。
【００４８】

音声発生源の特定は、例えば以下のような方法で行う。ＳＬＡＭ地図１２５３の作成時にロボット１００が部屋ＲＭ内を動き回り撮像した画像に対して、一般画像認識（画像に含まれる物体を一般的な名称で認識する処理）を行い、音声発生源か否かのラベリングする方法を用いてもよい。画像内で音声発生源としてラベリングされた領域に存在する２Ｄ特徴点に対応する地図内の地点に対して第１の値（第２の値より大きい値）を登録する。また、それ以外の２Ｄ特徴点に対応する地点には、第２の値（第１の値より小さい値）を登録する。具体的には、ロボット１００が通過した地点には、第２の値を登録し、ロボット１００が通過した際に、接触センサ、距離センサ等により障害物に接触したと判別した地点には、第１の値を登録する。

【００４９】

上述の例では、確率変数を２値とすることを説明した。あるいは、一般画像認識結果の尤度に、ラベリングの音声発生確率を乗じた値を確率変数としてもよい。

【００５０】

ラベリングの音声発生確率、ここでは、部屋ＲＭ内のそれぞれの位置における音声発生源の確率を示す情報（確率変数）はあらかじめ記憶部１２５の音声発生確率データベース１２５５に登録されているものとする。確率変数の値は、値が高いほど、当該位置に人以外の音声発生源が存在する可能性が高いことを示す。

【００５１】

音声発生確率データベース１２５５に登録されているラベリングの音声発生確率として、例えば、換気扇：０．８、ドア：０．５、観葉植物：０といった値が登録されている。換気扇は、動作中にそれなりの音を出し、ドアは、開け閉めする人により出る音の大きさに差があり、置かれているだけの観葉植物については、音は発生しないといった観点で、このような値が規定される。

【００５２】

また、ラベリングの音声発生確率は、時刻、季節、気温等に応じて、複数の値を規定しておいてもよい。季節に応じたラベリングの場合、例えば、夏：０．８、冬：０とする。夏場であれば、窓を開けることが多いため、室内でも室外で発生した音が聞こえることがあり、冬場の窓を閉め切った状態であれば、室外の音はほぼ聴こえないからである。

【００５３】

また、一般画像認識結果の尤度を使用するのは次のような理由による。一般画像認識を使用した場合、どのような画像に対しても、認識の精度が高いというわけではない。一般画像認識結果の尤度を用いることで、一般画像認識が誤認識した場合の影響を減らすことができる。

【００５４】

また、一般画像認識ではなく他の手法を用いてもよい。非特許文献３に記載されているBag-of-featuresという手法がある。この手法は、画像中の物体がどのカテゴリに属するかを求める画像分類問題の手法である。

【００５５】

あるいは、一般画像認識ではなく、ユーザが指定した音声発生源の領域、音声発生源となる物体を示す情報を、作成したＳＬＡＭ地図１２５３に追加してもよい。この場合、例えば、ロボット１００は、タッチパネル、ディスプレイ等の表示装置と、タッチパネル、キーボード等の入力装置を備え、ユーザに対して作成したＳＬＡＭ地図１２５３を提示して、ユーザに音声発生源を入力されるようにしてもよい。

【００５６】

あるいは、ロボット１００は、ＳＬＡＭ地図１２５３後に、部屋Ｒ内を動き回り、部屋ＲＭ内にある物体を指さしして、ユーザに、当該物体が音声発生源であるか等を尋ねてもよい。ユーザの回答に基づく音声発生源の情報を、ＳＬＡＭ地図１２５３に追加することができる。

【００５７】

あるいは、ＳＬＡＭ地図１２５３後に、部屋ＲＭ内の物体を撮像し、撮像画像を表示装

置に表示し、ユーザに、当該物体が音声発生源であるか等を尋ねてもよい。この場合も、ユーザの回答に基づく音声発生源の情報を、ＳＬＡＭ地図１２５３に追加することができる。

【００５８】

次に、音を検出した場合に、地図を使用して、検出した音の音源が人であるか否かを判別し、判別結果に応じて応答する呼びかけ応答処理を説明する。呼びかけ応答処理の開始に先立って、上述の地図作成処理はすでに実行されているものとし、ＳＬＡＭ地図１２５３、フレームデータベース１２５４、音声発生確率データベース１２５５には、適宜の情報が登録済みであるとする。

【００５９】

制御部１２７は、記憶部１２５の呼びかけ応答プログラム１２５１を実行することで、以下の呼びかけ応答処理を行い、検出した音声発生源が特定の音声発生源（ここでは人間）であるか否かを判別する判別手段として機能する。

【００６０】

図６のフローチャートを参照しながら、呼びかけ応答処理を説明する。制御部１２７は、ロボット１００の周辺である程度の大きさの音を検出したか否かを判別する（ステップＳ１０１）。具体的には、制御部１２７は、１つ以上のマイク１１２に所定の閾値以上の振幅の音が入力されたか否かを判別する。なお、所定の大きさとは、マイク１１２の感度によるものとする。

【００６１】

マイク１１２により所定の大きさの音が検出できない場合（ステップＳ１０１；Ｎｏ）、制御部１２７は、音を検出するまで待ち受ける。

【００６２】

一方、ある程度の大きさの音を検出したと判別した場合（ステップＳ１０１；Ｙｅｓ）、制御部１２７は、マイク１１２により検出した音が人間の声か否かを判別する（ステップＳ１０２）。具体的には、制御部１２７は、ステップＳ１０１で検出した音が特定の周波数帯域の音であるか否かを判別する。ステップＳ１０１で検出した音が人間の声でない場合（ステップＳ１０２；Ｎｏ）、制御部１２７はステップＳ１０１へ戻り、音を検出するまで待ち受ける。

【００６３】

一方、人間の声であると判別すると（ステップＳ１０２；Ｙｅｓ）、制御部１２７は、音源の位置（ここではユーザＰの声が発せられた位置）を求めるため、音声定位を行う（ステップＳ１０３）。ここでは、音源の位置を推定するため、音源定位のひとつの手法であるＭＵＳＩＣ（ＭＵｌｔｉｐｌｅ　ＳＩｇｎａｌ　Ｃｌａｓｓｉｆｉｃａｔｉｏｎ）を採用することとする。なお、音源定位の最中に音源であるユーザＰは移動せず、静止しているものとする。

【００６４】

図７を参照して音源定位を説明する。まず、マイク１１２に入力された音声を時間周波数変換する（ステップＳ１０３０１）。ここでは、時間周波数変換として、ＳＴＦＴ（Ｓｈｏｒｔ－Ｔｉｍｅ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ）（短時間フーリエ変換）を行う。

【００６５】

音源数をＮとすると、第ｎ番目の音源の信号S_nは、下記式（１）で表せる。

$S_n(\omega, f)$（ｎ＝１，２，…，Ｎ）…（１）

ωは角周波数、ｆはフレーム番号である（以下の説明でも同様）。

【００６６】

マイク１１２で観測される信号は、マイク１１２の数をＭとすると、下記式（２）で表せる。

$X_m(\omega, f)$（ｍ＝１，２，…，Ｍ）…（２）

【００６７】

音源から出た音は、空気を伝わってマイク１１２で観測されるが、そのときの伝達関数を$H_{nm}(\omega)$とすると、音源の信号を表す数式に、伝達関数を乗じることで、マイク１１２で観測される信号を求めることができる。m番目のマイク１１２で観測される信号$X_m(\omega, f)$は下記式（３）のように表される。

【００６８】

【数１】

$$X_m(\omega,f)=\sum_{n=1}^{N}S_n(\omega,f)H_{nm}(\omega) \quad \cdots (3)$$

【００６９】

ロボット１００は、マイク１１２を複数有しているので、マイク１１２全体で観測される信号$x(\omega, f)$は下記式（４）で表すことができる。

【００７０】

【数２】

$$x(\omega,f)=\begin{bmatrix} X_1(\omega,f) \\ X_2(\omega,f) \\ \vdots \\ X_M(\omega,f) \end{bmatrix} \quad \cdots (4)$$

【００７１】

同様に、全音源の信号$s(\omega, f)$も下記式（５）で表すことができる。

【００７２】

【数３】

$$s(\omega,f)=\begin{bmatrix} S_1(\omega,f) \\ S_2(\omega,f) \\ \vdots \\ S_N(\omega,f) \end{bmatrix} \quad \cdots (5)$$

【００７３】

同様に、第n番目の音源の伝達関数$h_n(\omega)$は下記式（６）で表すことができる。

【００７４】

【数４】

$$\mathrm{h_n}(\omega)=\begin{bmatrix}\mathrm{H_{n1}}(\omega)\\ \mathrm{H_{n2}}(\omega)\\ \vdots\\ \mathrm{H_{nM}}(\omega)\end{bmatrix} \quad \cdots (6)$$

【００７５】

全ての伝達関数を下記式（７）のように表記する。

ｈ（ω）＝［ｈ$_1$（ω），ｈ$_2$（ω），…ｈ$_N$（ω）］ …（７）

【００７６】

上記の式（７）で表される伝達関数を、上述の式（３）に適用すると、下記式（８）のように表される。

ｘ（ω，ｆ）＝ｈ（ω）ｓ（ω，ｆ） …（８）

【００７７】

ｈ$_n$（ω）は音源位置毎に独立であり、ある程度のフレーム数（例えば、フレーム数をＬとする））で見ればＳ$_n$（ω，ｆ）は無相関とみなせるので、ｘ（ω，ｆ）は音源数ＮをＲＡＮＫとする超平面を構成する。このとき、距離で正規化した音量が大きな音源の伝達関数方向に分布が広がりやすい。そこで、部分空間とゼロ空間に分解することを考える。

【００７８】

再び図７を参照する。次の式（９）に示すように相関行列を計算する（ステップＳ１０３０２）。ここで、＊は複素共役転置を意味する。

【００７９】

【数５】

$$\mathrm{R}(\omega,\mathrm{f})=\sum_{\mathrm{l}=0}^{\mathrm{L}-1}\mathrm{x}(\omega,\mathrm{f}+\mathrm{l})\mathrm{x}^{*}(\omega,\mathrm{f}+\mathrm{l}) \quad \cdots (9)$$

【００８０】

続いて、固有値分解する（ステップＳ１０３０３）。ここで、固有値λｍ（ω，ｆ）と固有ベクトルｅｍ（ω，ｆ）は固有値が降順になるように並べ替えられているものとする。

【００８１】

原理的には、ｈ$_n$（ω）は部分空間の固有ベクトルｅｍ（ω，ｆ）（ｍ＝１～Ｎ）の重み付け加算から復元できるが、実際には復元が困難であるためゼロ空間を構成する固有ベクトルｅｍ（ω，ｆ）（ｍ＝Ｎ＋１～Ｍ）がｈ$_n$（ω）と直交することを使って音源定位を実現する。

【００８２】

しかし、音源であるユーザＰが部屋ＲＭ内を移動する可能性があるため、音源位置を予め知ることはできず、音源位置の伝達関数を予め取得しておくことは難しい。このため、

仮の音源位置を決め、仮の音源位置の伝達関数をあらかじめ用意しておき、音源定位を行う。

【００８３】

図８に、仮の音源位置とマイクの配置の一例を示す。図８では、太線の円がロボット１００の頭１１０を表し、太線上の黒丸がマイク１１２を表す。なお、ここでは、便宜上１３個のマイク１１２の全てを表示していない。ロボット１００の回りには４個の仮の音源位置があるものとする。

【００８４】

複数のマイク１１２は、ロボット１００の頭１１０に配置されていることから、円周に沿って配置されているとみなすことができる。Ｘ軸の正の向きと、マイク１１２が成す円の中心（ロボット１００の頭１１０の中心位置に相当）と仮の音源１～４とをそれぞれ結んだ線と、がなす角度をθ１、θ２、θ３、θ４とし、それぞれの伝達関数$h_{\theta}(\omega)$を予め計算しておく。

【００８５】

図８では、音源が４個の例を示したが、音源数がＮ個の場合、θ１、θ２、…θＮのそれぞれの伝達関数$h_{\theta}(\omega)$を予め計算しておけばよい。また、あるいは、仮の音源位置の伝達関数を用意するのではなく、幾何的な情報をもとに予め伝達関数を計算しておいてもよい。

【００８６】

再び図７を参照する。次の式（１０）を使用して、周波数帯毎のＭＵＳＩＣスペクトルを計算する（ステップＳ１０３０４）。

【００８７】

【数６】

$$M_{\theta}(\omega,f)=\frac{h^{*}_{\theta}(\omega)h_{\theta}(\omega)}{\sum_{m=N+1}^{M}\left|h^{*}_{\theta}(\omega)e_{m}(\omega.f)\right|^{2}} \quad \cdots (10)$$

【００８８】

ここで、式（１０）の分母は、ノイズや誤差、ＳＴＦＴの周波数帯間の信号漏洩の影響等からゼロにはならない。また、音源の方向とあらかじめ決めた角度θ（θ１、θ２、…θＮ）のいずれかが近い場合、つまり$h_{n}(\omega)$と$h_{\theta}(\omega)$が近い場合、式（１０）の値は極端に大きなものになる。図８に示す例では、音源である人と仮の音源２の位置が近いため、θ２の伝達関数を使用した場合、式（１０）の値が極端に大きくなることが想定される。

【００８９】

そして、統合したＭＵＳＩＣのパワーを求めるため、式（１１）に示すように周波数帯毎のＭＵＳＩＣスペクトルを重み付け加算する（ステップＳ１０３０５）。

【００９０】

【数７】

$$M(f)=\sum_{\omega}w(\omega)M\ (\omega,f) \quad \cdots (11)$$

【００９１】

重み付け係数は、固有値λｍ（ω，ｆ）が大きいほど大きくすれば、$S_{n}(\omega,f)$に

含まれるパワーに応じた計算をすることもできる。この場合は$S_n(\omega, f)$に殆どパワーがない場合の悪影響を軽減できる。

【００９２】

続いて、パワースペクトルから適切なピーク（極大値）を選択する（ステップＳ１０３０６）。具体的には、まず、複数のピークを求め、その中から適切なピークを選択し、選択したピークにおけるθを音源方向とする。ここで、ピークを求めるのは以下のような理由による。本来の音源方向のθのパワーが必ずしも一番大きいとは限らず、本来の音源方向に近いθのパワーは総じて大きくなるので、音源方向は複数のピークの何れかに正解があるからである。

【００９３】

また、テレビが点いている、ドアホンが鳴る等の部屋ＲＭ内に他の音源がある場合でも、多くの場合、人は、テレビ、ドアホン等の周囲の音より大きな声でロボット１００に呼びかけると考えられる。よって、人の声のパワーの方が、人以外のテレビ、ドアホン等の音源から発せられる音のパワーより大きくなることが想定される。よって、単純にパワーが最大となる仮の音源位置を示すθを音源方向として選択しても問題はない。ただし、周囲の環境などによっては、パワーが最大となる仮の音源位置ではなく、パワーが２番目あるいはそれ以降となる仮の音源位置を、音源方向と選択することが適切な場合もある。このようにして、制御部１２７は、音源方向、ここでは、ロボット１００の位置から見たユーザＰがいる方向、を判別することができる。

【００９４】

音源定位の処理は以上である。ここでは、平面を仮定して説明したが、３次元を仮定しても上記説明は成り立つ。

【００９５】

再び図６を参照する。ステップＳ１０３の音源定位を実行して音源方向を判別すると、制御部１２７は、音源方向を示す情報として、ロボット１００の向いている方向に対する音源の方向を示す角度θを記憶部１２５に記憶する。続いて、制御部１２７は、ステップＳ１０４へ進み、撮影画像と、地図（ＳＬＡＭ地図１２５３、フレームデータベース１２５４）を用いて自機位置推定の処理を実行する。

【００９６】

図９を参照して、自機位置の推定の処理を説明する。制御部１２７は、カメラ１１１により撮像された画像の二次元特徴点（２Ｄ特徴点）を抽出する（ステップＳ１０４０１）。具体的には、制御部１２７は、カメラ１１１を制御して撮像し、撮像した画像から２Ｄ特徴点を抽出する。

【００９７】

続いて、制御部１２７は、記憶部１２５のフレームデータベース１２５４を参照して、フレームデータベース１２５４に登録されている以前のフレームの情報から、その画像の情報に含まれている２Ｄ特徴点のうち、３Ｄ位置が既知である２Ｄ特徴点を取得し、取得した２Ｄ特徴点から、ステップＳ１０４０１で抽出した２Ｄ特徴点と、対応が取れる特徴点を抽出する（ステップＳ１０４０２）。ここで、３Ｄ位置が既知であるとは、即ち、２Ｄ特徴点がＳＬＡＭ地図に登録されていることを意味する。

【００９８】

制御部１２７は、ステップＳ１０４０２で抽出した対応が取れる特徴点の個数が、閾値以上であるか否かを判別する（ステップＳ１０４０３）。閾値未満であると判別した場合（ステップＳ１０４０３；Ｎｏ）、制御部１２７は、足回り部１２６を制御して、自機を所定の距離だけ移動し（ステップＳ１０４０６）、再びステップＳ１０４０１へ戻る。

【００９９】

一方、ステップＳ１０４０２で抽出した対応特徴点の個数が、閾値以上であると判別した場合（ステップＳ１０４０３；Ｙｅｓ）、制御部１２７は、記憶部１２５のＳＬＡＭ地図１２５３から、ステップＳ１０４０２で抽出した対応特徴点それぞれの３Ｄ座標（Ｘｉ，Ｙｉ，Ｚｉ）を取得する（ステップＳ１０４０４）。

【０１００】
続いて、制御部１２７は、自機の姿勢を推定する（ステップＳ１０４０５）。ここでは、制御部１２７は、対応特徴点のＳＬＡＭ地図上の３Ｄ位置と、対応特徴点のフレーム座標（２Ｄ座標）の関係から自機の姿勢（並進ベクトルｔ及び回転行列Ｒで表される自機の位置及び向き）を推定する。
【０１０１】
具体的には、今撮像した画像に含まれている対応特徴点のフレーム座標を（ｕｉ，ｖｉ）とし、その対応特徴点の３Ｄ座標を（Ｘｉ，Ｙｉ，Ｚｉ）とする（ｉは１から対応特徴点の数までの値を取る）。ここで、各対応特徴点の３Ｄ位置（Ｘｉ，Ｙｉ，Ｚｉ）を下記式（１２）によってフレーム座標系に投影した値（ｕｘｉ，ｖｘｉ）とフレーム座標（ｕｉ，ｖｉ）とは理想的には一致する。
（ｕｘｉ　ｖｘｉ　１）’～Ａ（Ｒ｜ｔ）（Ｘｉ　Ｙｉ　Ｚｉ　１）’…（１２）
【０１０２】
しかし、実際には（Ｘｉ，Ｙｉ，Ｚｉ）にも（ｕｉ，ｖｉ）にも誤差が含まれているため、（ｕｘｉ，ｖｘｉ）と（ｕｉ，ｖｉ）とが一致することはめったにない。そして、未知数はＲとｔ（３次元空間ではそれぞれ３次元となり、３＋３＝６が未知数の個数である）だけなのに、数式は対応特徴点の個数の２倍存在する（対応特徴点一つに対して、フレーム座標のｕ，ｖそれぞれに対する式が存在するため）ことになるため、過剰条件の連立一次方程式になり、上述したように最小二乗法で求めることになる。
【０１０３】
具体的には、制御部１２７は、以下の式（１３）のコスト関数Ｅ１を最小化する姿勢（並進ベクトルｔ及び回転行列Ｒ）を求める。
【０１０４】
【数８】

$$E1 = \sum_{i=1}^{\text{対応特徴点数}} \left((u_i - ux_i)^2 + (v_i - vx_i)^2 \right) \quad \cdots \quad (13)$$

【０１０５】
このように求めた値が、ＳＬＡＭ法で求めたＳＬＡＭ座標での自機の姿勢（並進ベクトルｔ及び回転行列Ｒで表される自機の位置及び向き）を示す値である。このようにして算出した値により自機の姿勢が推定される。以上が自機位置推定の処理である。
【０１０６】
再び、図６を参照する。制御部１２７は、ステップＳ１０４の自機位置の推定の処理が終わると、ステップＳ１０５へ進み、ＳＬＡＭ地図１２５３と音声発生確率データベース１２５５とを参照して、ステップＳ１０４で推定した自機位置から、ステップＳ１０３で求めた音源方向に、人以外の音声発生源が存在する確率を取得する（ステップＳ１０５）。ここでは、音源方向の各点の確率の平均を求め、求めた平均を人以外の音声発生源が存在する確率としてもよい。あるいは、音源方向の各点の確率について最大値を人以外の音声発生源が存在する確率としてもよい。
【０１０７】
次に、制御部１２７は、ステップＳ１０５で求めた人以外の音声発生源が存在する確率が閾値以上であるか否かを判別する（ステップＳ１０６）。人以外の音声発生源が存在する確率が閾値以上であると判別した場合（ステップＳ１０６；Ｙｅｓ）、制御部１２７は、音源方向の音源は人以外であると判別して、首関節１２１を回転駆動させず、再びステップＳ１０１へ戻り、音の入力を待ち受ける。
【０１０８】
一方、人以外の音声発生源が存在する確率が閾値未満であると判別した場合（ステップＳ１０６；Ｎｏ）、制御部１２７は、ステップＳ１０７へ進む。

【０１０９】
続いて制御部１２７は、頭１１０の回転をさせるため、首関節１２１を回転駆動させる（ステップＳ１０７）。ここで、制御部１２７は、ロボット１００の頭１１０の正面（カメラ１１１のレンズ面）が音源（ユーザＰ）の方向に向くまで、頭１１０を回転する。具体的には、制御部１２７は、記憶部１２５に記憶されている音源定位により求められた角度θに基づいて、求めた角度だけ頭１１０を回転し、その後、回転駆動を停止する。このようにして、カメラ１１１のレンズ面を音源（ユーザＰ）がいる方向に向ける。
【０１１０】
回転駆動を停止した後、制御部１２７は、顔検出の処理を実行する（ステップＳ１０８）。まず、制御部１２７は、カメラ１１１を制御して撮像し、撮像した画像に対して以下の処理を施すことで、顔検出処理を実行する。
【０１１１】
制御部１２７は、まず、ピラミッド画像を作成する。ピラミッド画像とは、元画像を一定の比率で縮小を繰り返して作成した一連の画像群であり、ピラミッド画像の各階層に対して、固定サイズの顔検出器を適用することで様々なサイズ（つまり距離に相当）の顔を検出することができる。ここでは、回転によるカメラの見え方は対象までの距離によって変わるので、ピラミッド画像を使用して顔検出を行う。
【０１１２】
まず、顔探索対象を最初の階層に設定する。ここでは縮小前の元の画像とする。最初の検出窓を設定する。初期位置は例えば左上の隅とする。設定した検出窓に対して、固定サイズの顔検出器を適用する。この階層でのスライドによる探索が完了したかを判定する。スライドによる探索が完了でないなら、検索窓をスライドさせ、再度顔検出を行う。スライドによる探索が完了ならば、ピラミッド画像のすべての階層での処理が完了したかの判定を行う。すべての階層での処理が完了でないなら、階層を移動し、移動先の階層でもスライドによる顔検出を行う。すべての階層での処理が完了したならば、顔検出の処理を終了する。
【０１１３】
なお、ロボット１００から近い顔画像は、画角に入りきらない場合があることと、全体の計算負荷の割合が小さいことを考慮して、縮小率の大きい階層の顔探索はしないほうがより望ましい。
【０１１４】
顔検出処理により、撮像画像から顔を検出することができなかった場合（ステップＳ１０８；Ｎｏ）、制御部１２７は、再びステップＳ１０１に戻る。
【０１１５】
一方、顔検出が成功すると（ステップＳ１０８；Ｙｅｓ）、続いて、制御部１２７は、ユーザＰがロボット１００に注目しているかどうかを判別する（ステップＳ１０９）。具体的には、制御部１２７は、カメラ１１１を制御して、ユーザＰを撮像し、撮像した画像からユーザＰの顔が、ロボット１００の方を向いているか否かを判別する。ユーザＰがロボット１００に注目していないと判別すると（ステップＳ１０９；Ｎｏ）、再びステップＳ１０１へ戻り、音の入力を待ち受ける。
【０１１６】
一方、ユーザＰの顔が、ロボット１００の方を向いていると判別すると（ステップＳ１０９；Ｙｅｓ）、制御部１２７は、ユーザＰに近づくように所定の距離だけ移動し（ステップＳ１１０）、ユーザＰとの距離が決められた距離以下となったかを判別する（ステップＳ１１１）。このユーザＰとロボット１００との間の決められた距離は、ロボット１００が、ユーザＰが発声する内容を音声認識することができる程度の距離である。制御部１２７は、ロボット１００とユーザＰとの距離が決められた距離以下ではないと判別した場合に（ステップＳ１１１；Ｎｏ）。再びステップＳ１１０に戻る。
【０１１７】
一方、制御部１２７は、ユーザＰとの距離が所定の距離となったと判別した場合に（ス

テップＳ１１１；Ｙｅｓ）、ユーザＰと対話する（ステップＳ１１２）。例えば、制御部１２７は、スピーカ１１３を制御して、ユーザＰに対して、例えば、「何かご用ですか？」と話しかけ、また、マイク１１２から入力したユーザの発言を音声解析し、解析した内容に基づいて、なんらかの音声をスピーカ１１３から出力する。

【０１１８】

以上、説明したように、本実施の形態においては、ロボット１００は、あらかじめ作成したＳＬＡＭ地図１２５３に基づき、判別した音源方向に人以外の音源がある場合、人に呼ばれたのではないと判別する。よって、人以外の音源であるテレビ、ラジオ等から人の声が聞こえた場合であっても、振り向かないので、無駄な動作を減らすことができる。

【０１１９】

なお、上述の説明においては、回転駆動はｙａｗを前提で説明したが、他の方向の回転があっても成立する。

【０１２０】

実施の形態１においては、ロボット１００は、ユーザＰの方向へ近づくよう、単に移動したが、ロボット１００は、ＳＬＡＭ地図１２５３を使用して、部屋ＲＭ内を移動し、ユーザに近づいてもよい。

【０１２１】

（実施の形態２）

実施の形態２においては、ユーザＰから呼びかけられたロボット１００が、ＳＬＡＭ地図１２５３を使用して、移動経路を作成し、移動経路に沿って移動する。ロボット１００が備える構成は、実施の形態１と同様である。以下の説明においては、実施の形態２に特有の構成を中心に説明する。

【０１２２】

実施の形態１と同様に、あらかじめＳＬＡＭ地図１２５３が作成されているものとする。

【０１２３】

実施の形態２においては、図１０に示すように、記憶部１２５には後述の呼びかけ移動処理のための呼びかけ移動プログラム１２５６が記憶されているものとする。制御部１２７は、呼びかけ移動プログラム１２５６を実行することによって、以下の処理を行う。

【０１２４】

図１１にユーザＰに呼びかけられたときに、ロボット１００がユーザＰのいる場所まで移動する処理（呼びかけ移動処理）のフローチャートを示す。なお、上述の呼びかけ応答処理と同様であるので、ここでは、所定の大きさの音を検出し、検出した音が人間の声であると判別したと仮定して、説明を行う。

【０１２５】

制御部１２７は、撮像した画像とＳＬＡＭ地図１２５３を用いて自機位置推定の処理を実行する（ステップＳ２０１）。自機位置推定の処理については図９を参照して説明したため、ここでは、説明を省略する。

【０１２６】

続いて、制御部１２７は、ＳＬＡＭ法によるユーザＰの顔の位置を推定する処理（顔位置推定の処理）を実行する（ステップＳ２０２）。図１２を参照して、顔位置推定の処理を説明する。制御部１２７は、カメラ１１１を制御して撮像し、撮像した画像から二次元特徴点（２Ｄ特徴点）を抽出する（ステップＳ２０２０１）。特徴抽出にはＳＩＦＴやＳＵＲＦ等のアルゴリズムを用いる。

【０１２７】

制御部１２７は、ステップＳ２０２０１で抽出した２Ｄ特徴点のうち、撮像した画像の顔の領域内の特徴点（２Ｄ顔特徴点）を抽出する（ステップＳ２０２０２）。顔領域内に特徴点がない場合は、顔パーツ検出の結果を特徴点として用いる。

【０１２８】

制御部１２７は、顔位置推定の処理開始後に撮像した画像が２枚以上であるか否かを判

別する（ステップＳ２０２０３）。２枚未満であると判別すると、（ステップＳ２０２０３；Ｎｏ）、制御部１２７は、足回り部１２６を制御して、自機を所定の距離だけ移動し（ステップＳ２０２０８）、再びステップＳ２０２０１へ戻る。

【０１２９】

一方、撮像した画像が２枚以上であると判別した場合（ステップＳ２０２０３；Ｙｅｓ）、制御部１２７は、２つの画像の２Ｄ顔特徴点の対応を取得する（ステップＳ２０２０４）。

【０１３０】

制御部１２７は、ステップ２０２０２で抽出した対応する２Ｄ顔特徴点の個数が、閾値以上であるか否かを判別する（ステップＳ２０２０５）。閾値未満であると判別した場合（ステップＳ２０２０５；Ｎｏ）、制御部１２７は、足回り部１２６を制御して、自機を所定の距離だけ移動し（ステップＳ２０２０８）、再びステップＳ２０２０１へ戻る。

【０１３１】

一方、２つの画像の対応する２Ｄ顔特徴点の個数が、閾値以上であると判別した場合（ステップＳ２０２０５；Ｙｅｓ）、制御部１２７は、２つの画像間の姿勢を推定する（ステップＳ２０２０６）。

【０１３２】

具体的には、２つの画像の間で対応する２Ｄ顔特徴点の２次元座標（２Ｄ座標）と、２つの画像のそれぞれの撮影位置（撮影時の自機の位置）の間の距離と、に対して、Ｔｗｏ－ｖｉｅｗ　Ｓｔｒｕｃｔｕｒｅ　ｆｒｏｍ　Ｍｏｔｉｏｎ法を用いて、２つの画像間の姿勢（それぞれの画像を取得した位置の差分（並進ベクトルｔ）及び向きの差分（回転行列Ｒ））を推定する。

【０１３３】

続いて、制御部１２７は、２つの画像の間で対応する２Ｄ顔特徴点の３次元座標（３Ｄ座標）を推定する（ステップＳ２０２０７）。具体的には、これは、ステップＳ２０２０６で算出した２つの画像間の姿勢を表す値と、２つの画像の間で対応する２Ｄ顔特徴点の２Ｄ座標と、を用いて推定する。以上が、顔位置推定の処理である。

【０１３４】

図１１を再び参照する。ステップＳ２０３に進み、制御部１２７は、自機位置からユーザＰの顔位置までの経路の作成を行う（ステップＳ２０３）。

【０１３５】

実施の形態２においては、ＳＬＡＭ地図１２５３に、部屋ＲＭ内のそれぞれの位置における障害物が存在する確率を示す障害物情報の確率変数を付加したものを使用する。障害物情報の確率変数の値は、その値が高いほど、その位置に障害物がある可能性が高いことを表している。

【０１３６】

経路の作成は、まず、記憶部１２５からＳＬＡＭ地図１２５３を読み出し、ＳＬＡＭ地図１２５３上にランダムにノードを配置する（ノード情報を追加する）。このとき、自機（ロボット１００）と同じ高さにノードを配置する。また、障害物情報の確率変数が閾値以上である位置（点）を中心とした一定の範囲内には、ノードを配置しない。

【０１３７】

なお、高さは、重力方向のオフセット値を用いて推定する。具体的には、ロボット１００が過去に移動した位置から面推定を行い、法線ベクトル（重力方向）を求めて、自機位置と法線ベクトルの内積を求め自機位置の高さとする。経路中のノードも同様にして値を求める。自機位置の高さを表す値と、ノードの高さを表す値の差が決められた閾値以内であれば、自機位置の高さとノードの高さが同じであるとみなす。

【０１３８】

配置したノードについて、当該ノードを中心とした一定の範囲内に存在する他のノードとをつなぐ。これを、ランダムに配置した全てのノードについて行う。このようにして、グラフ構造を作る。

【０１３９】
ステップＳ２０１の自機位置推定で推定した自機の位置の一番近くに存在するノード、ステップＳ２０２顔位置推定で推定したユーザＰの顔の一番近くに存在するノード、をそれぞれ選択する。そして、ダイクストラ法により、選択した２つのノード間の最短経路を求める。
【０１４０】
その後、求めた最短経路に従って、移動する（ステップＳ２０４）。以上が、実施の形態２にかかる呼びかけ移動処理である。
【０１４１】
（変形例）
実施の形態２の呼びかけ移動処理では、２つの画像間の姿勢の推定（図１２のステップＳ２０２０６）に、Ｔｗｏ－ｖｉｅｗ　Ｓｔｒｕｃｔｕｒｅ　ｆｒｏｍ　Ｍｏｔｉｏｎ法を用いた。姿勢の推定はこれに限られない。
【０１４２】
上述の自機位置推定の処理における姿勢の推定（図９のステップＳ１０４０５）で行ったように姿勢を推定してもよい。この方法の方が、精度が高く、計算に要する時間も少ない。また、あるいは、被写体の顔のサイズが標準的な顔サイズであると仮定し、顔検出結果（顔のサイズ、位置）とカメラパラメータ（画角、焦点距離）を用いて、自機と顔間の相対的な姿勢を推定し、ＳＬＡＭ法により求めた自機の姿勢のＳＬＡＭ地図上における推定結果を用いて、ＳＬＡＭ地図上における顔の姿勢を算出してもよい。また、あるいは、ロボット１００に距離センサを設け、距離センサを使用して、ロボット１００とユーザＰの顔の間の距離を測定してもよい。
【０１４３】
上記の実施の形態２のＳＬＡＭ地図１２５３を使用した移動処理は、実施の形態１における呼びかけ応答処理時の、図６のステップＳ１１０の移動の際にも応用可能である。
【０１４４】
ＳＬＡＭ地図１２５３の精度を上げる方法として、次のようなものがある。フレームデータベースにある程度の撮像した画像のデータが蓄積されたところで、３Ｄ位置が既知で無い特徴点の対応を再探索し、３Ｄ位置を計算してもよい。
【０１４５】
また、バンドルアジャストメント処理を行い、キーフレーム姿勢とＭａｐ点の３Ｄ位置の精度を向上させることができる。バンドルアジャストメント処理とは、カメラ姿勢（キーフレーム姿勢）とＭａｐ点の３Ｄ位置とを同時に推定する非線形最適化法である。この方法を使用することで、ＳＬＡＭ地図上の点を、画像上に投影させたときに発生する誤差が最小になるような最適化を行うことができる。
【０１４６】
また、ループクロージング処理を行ってもよい。ループクロージング処理とは、以前に来たことのある同じ場所に戻ってきたことを認識した場合に、以前その場所にいた時の姿勢の値と現在の姿勢の値とのずれを用いて、以前に来た時から今までの軌跡中の画像や、関連するＭａｐ点の３Ｄ位置を修正することである。
【０１４７】
制御部１２７は、音源から発せられる音が、ロボット１００に向けられたものか否かを判別し、ロボット１００に向けられたものであると判別した場合だけ、音検出を行い、その音が人間の声であるか否かを判別し、人間の声であると判別した場合に、上述の処理により振り返り判定を行ってもよい。この場合、例えば、マイク１１２に含まれる１３個のマイクとして単一指向性マイクを使用することで、音源から発せられる音の方向を精度良く判別することができる。
【０１４８】
本発明は、上記実施形態に限定されず、本発明の要旨を逸脱しない部分での種々の修正は勿論可能である。

【０１４９】
上述の実施の形態では、ロボット１００、ユーザＰともに屋内（部屋ＲＭ内）にいる例を説明したが、屋外であっても同様に、本発明を採用して、ロボットは振り向き判定を行うことができる。
【０１５０】
上述の顔検出では、ピラミッド画像の階層を順次移動して、顔検出を行ったが、制御部１２７は、マイク１１２の入力音声の大きさ（振幅の大きさ）に基づいて、ロボット１００から音源までの距離を推定し、推定した距離に基づいて、ピラミッド画像の全ての階層について顔検出を行わないようにしてもよい。例えば、ユーザＰが近くにいると判別した場合、ある程度小さく縮小したピラミッド画像を使用する必要はない。
【０１５１】
また、制御部１２７は、マイク１１２の入力音声を、そのときの、人か否かの判別結果とともに、記憶部１２５に記憶しておいてもよい。再度、同じ音を検出したとき、人か否かの判別が容易となるからである。
【０１５２】
また、ユーザが、あらかじめ、ロボット１００のマイク１１２の入力とないうる人以外の音のデータを記憶させてもよい。例えば、インターホンの音、電話の呼び出し音である。よって、ロボット１００は、当該音声が聞こえた場合には、人以外であると判別することができる。
【０１５３】
また、ロボット１００が屋外にいる場合、あらかじめ、周囲を撮像して、撮像した画像を画像認識しておくことが好ましい。屋外の場合、音源になりうるものの数が室内に比べ多くなることが想定されるからである。例えば、公園内であれば、大型スピーカが設置されていることがあり、あらかじめ、撮像画像から大型スピーカを画像認識しておき、音源として記憶することで、ロボット１００が、振り向き判定をしやすくなる。
【０１５４】
上述の実施の形態では、音声発生源が人間であるか否かを判別する構成を説明した。しかし、判別する特定の音声発生源は、人間だけに限られない。音声発生源の判別の対象に、人間のように自らの意思で話す人工知能を搭載したロボットを含めることができる。本発明を採用することで、人間に加え、人間のように自らの意思で話す人工知能を搭載したロボットの音声についても、同様に判別することができる。
【０１５５】
また、本発明に係る顔認識装置は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。例えば、ネットワークに接続されているコンピュータに、上記動作を実行するためのプログラムを、コンピュータシステムが読み取り可能な記録媒体（ＣＤ－ＲＯＭ（Ｃｏｍｐａｃｔ Ｄｉｓｃ Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）、ＭＯ（Ｍａｇｎｅｔｏ－Ｏｐｔｉｃａｌ）等）に格納して配布し、当該プログラムをコンピュータシステムにインストールすることにより、上述の処理を実行する顔認識装置を構成してもよい。
【０１５６】
また、コンピュータにプログラムを提供する方法は任意である。例えば、プログラムは、通信回線の掲示板（ＢＢＳ（Ｂｕｌｌｅｔｉｎ Ｂｏａｒｄ Ｓｙｓｔｅｍ））にアップロードされ、通信回線を介してコンピュータに配信されてもよい。また、プログラムは、プログラムを表す信号により搬送波を変調した変調波により伝送され、この変調波を受信した装置が変調波を復調してプログラムを復元するようにしてもよい。そして、コンピュータは、このプログラムを起動して、ＯＳ（Ｏｐｅｒａｔｉｎｇ Ｓｙｓｔｅｍ）の制御のもと、他のアプリケーションと同様に実行する。これにより、コンピュータは、上述の処理を実行する顔認識装置として機能する。
【０１５７】
この発明は、この発明の広義の精神と範囲を逸脱することなく、様々な実施の形態及び

変形が可能とされるものである。また、上述した実施の形態は、この発明を説明するためのものであり、この発明の範囲を限定するものではない。すなわち、この発明の範囲は、実施の形態ではなく、請求の範囲によって示される。そして、請求の範囲内及びそれと同等の発明の意義の範囲内で施される様々な変形が、この発明の範囲内とみなされる。この発明の範囲内とみなされる。以下に、本願出願の当初の特許請求の範囲に記載された発明を付記する。

【０１５８】

（付記）

（付記１）

音声を検出する音声検出手段と、

前記音声検出手段が検出した音声の音声発生源が特定の音声発生源であるか否かを判別する判別手段と、

前記判別手段の判別結果に基づいて自機を制御する制御手段と、

を備える音声検出装置。

【０１５９】

（付記２）

前記制御手段は、前記判別手段が、前記音声検出手段が検出した音声の音声発生源が前記特定の音声発生源であると判別した場合、自機の位置、姿勢の少なくとも一方を変えるよう自機を制御する、

付記１に記載の音声検出装置。

【０１６０】

（付記３）

撮像部と、

前記特定の音声発生源以外の音声発生源であって、登録された音声発生源の位置を示す情報を含む音声発生源位置情報があらかじめ記憶された記憶部と、

をさらに備え、

前記判別手段は、前記音声検出手段が検出した音声の音声発生源の位置を判別し、判別した位置が、前記音声発生源位置情報に含まれる前記登録された音声発生源の位置であるか否かを判別し、

前記判別手段が判別した位置が、前記音声発生源位置情報に含まれる前記登録された音声発生源の位置でないと判別した場合に、前記制御手段は、前記撮像部の撮像方向を前記判別手段が判別した位置に向けるように、自機の位置、姿勢の少なくとも一方を変える、

付記１または２に記載の音声検出装置。

【０１６１】

（付記４）

前記音声発生源位置情報は、さらに、前記登録された音声発生源の位置に、前記特定の音声発生源以外の音声発生源が存在する確率を示す情報を含む、

付記３に記載の音声検出装置。

【０１６２】

（付記５）

前記制御手段により自機を移動させている間に、前記撮像部が撮像した画像から認識された音声発生源の位置を示す情報を、前記音声発生源位置情報に追加する、

付記３または４に記載の音声検出装置。

【０１６３】

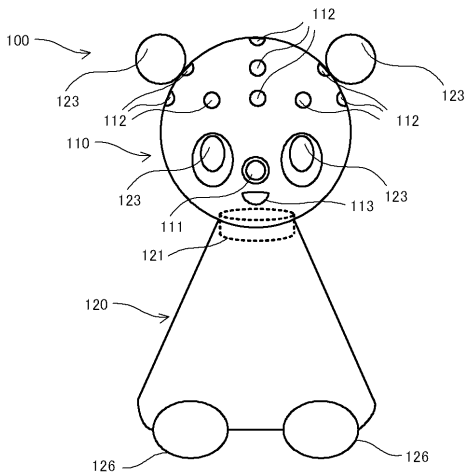
（付記６）

前記判別手段は、前記音声検出手段が検出した音声が自機宛てに発せられた音声か否かを判別し、自機宛ての音声であると判別した場合、前記音声検出手段が検出した音声の音声発生源が前記特定の音声発生源であるか否かを判別する、

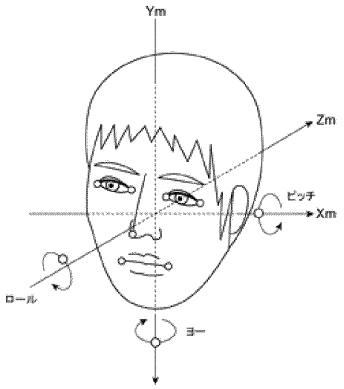
付記１から５のいずれか１つに記載の音声検出装置。

【０１６４】

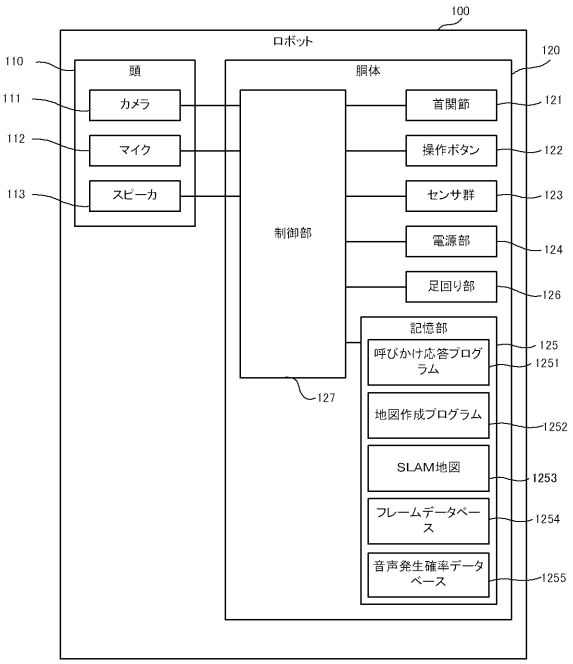
（付記７）
ロボットに搭載されたコンピュータが音声を検出する音声検出方法であって、
音声を検出する音声検出ステップと、
前記音声検出ステップで検出された音声の音声発生源が特定の音声発生源であるか否かを判別する判別ステップと、
前記判別ステップの判別結果に基づいて、前記ロボットの動作を制御する制御ステップと、
を備える音声検出方法。
【０１６５】
（付記８）
ロボットに搭載されたコンピュータに、
音声を検出する音声検出機能と、
前記音声検出機能により検出された音声の音声発生源が特定の音声発生源であるか否かを判別する判別機能と、
前記判別機能により判別された判別結果に基づいて、前記ロボットの動作を制御する制御機能と、
を実現させるプログラム。
【符号の説明】
【０１６６】
１００…ロボット、１１０…頭、１１１…カメラ、１１２…マイク、１１３…スピーカ、１２０…胴体、１２１…首関節、１２２…操作ボタン、１２３…センサ群、１２４…電源部、１２５…記憶部、１２６…足回り部、１２７…制御部、１２５１…呼びかけ応答プログラム、１２５２…地図作成プログラム、１２５３…ＳＬＡＭ地図、１２５４…フレームデータベース、１２５５…音声発生確率データベース、１２５６…呼びかけ移動プログラム、ＲＭ…部屋
【図１】



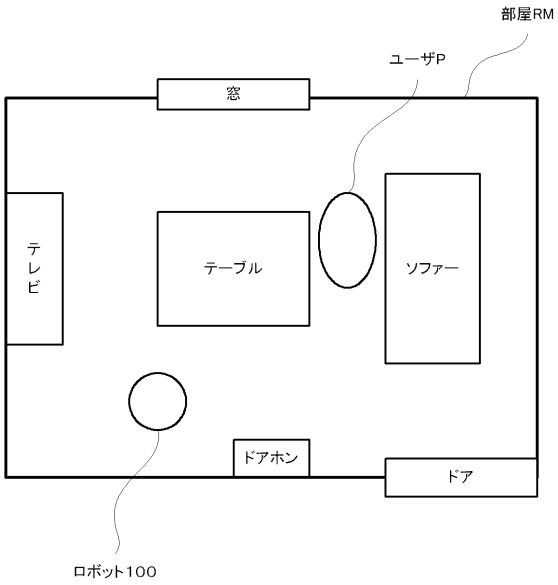
【図２】



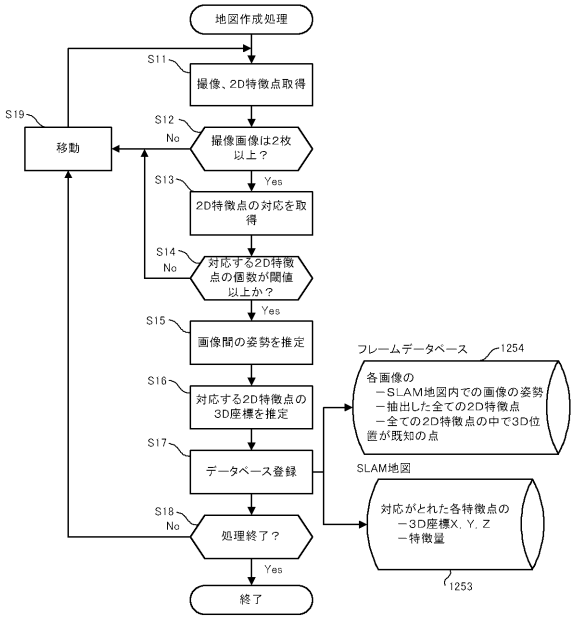
【図３】


100
ロボット
110
頭
120
胴体
111
カメラ
112
マイク
113
スピーカ
制御部
首関節
121
操作ボタン
122
センサ群
123
電源部
124
足回り部
126
127
記憶部
125
呼びかけ応答プログラム
1251
地図作成プログラム
1252
SLAM地図
1253
フレームデータベース
1254
音声発生確率データベース
1255

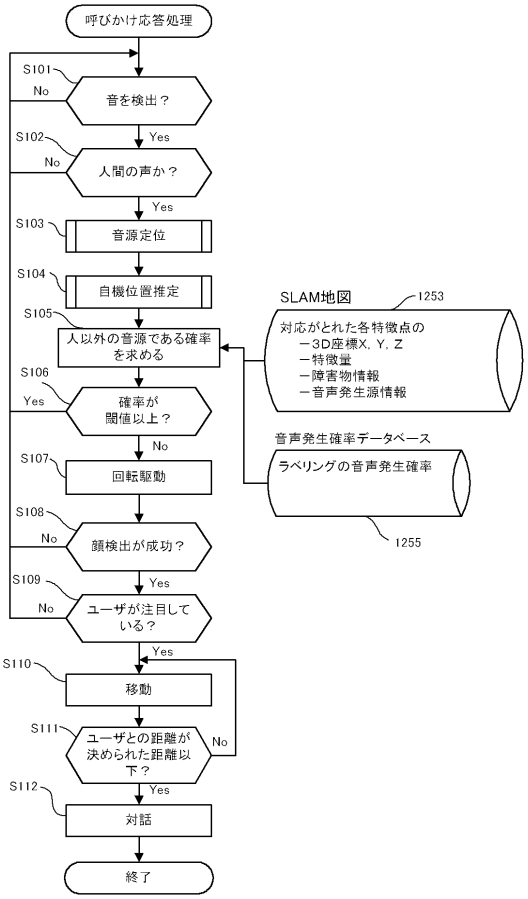
【図４】


部屋RM
ユーザP
窓
テレビ
テーブル
ソファー
ドアホン
ドア
ロボット100

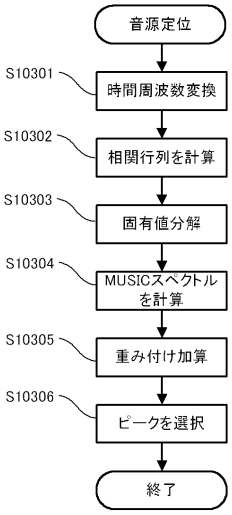
【図５】


地図作成処理
S11
撮像、2D特徴点取得
S19
移動
S12
No
撮像画像は2枚以上？
Yes
S13
2D特徴点の対応を取得
S14
No
対応する2D特徴点の個数が閾値以上か？
Yes
S15
画像間の姿勢を推定
S16
対応する2D特徴点の3D座標を推定
S17
データベース登録
S18
No
処理終了？
Yes
終了
フレームデータベース
1254
各画像の
－SLAM地図内での画像の姿勢
－抽出した全ての2D特徴点
－全ての2D特徴点の中で3D位置が既知の点
SLAM地図
対応がとれた各特徴点の
－3D座標X，Y，Z
－特徴量
1253

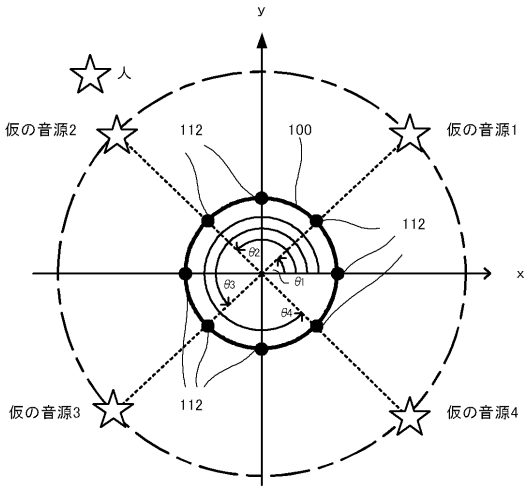
【図６】


呼びかけ応答処理
S101
No
音を検出？
Yes
S102
No
人間の声か？
Yes
S103
音源定位
S104
自機位置推定
S105
人以外の音源である確率を求める
S106
Yes
確率が閾値以上？
No
S107
回転駆動
S108
No
顔検出が成功？
Yes
S109
No
ユーザが注目している？
Yes
S110
移動
S111
ユーザとの距離が決められた距離以下？
No
Yes
S112
対話
終了
SLAM地図
1253
対応がとれた各特徴点の
－3D座標X，Y，Z
－特徴量
－障害物情報
－音声発生源情報
音声発生確率データベース
ラベリングの音声発生確率
1255

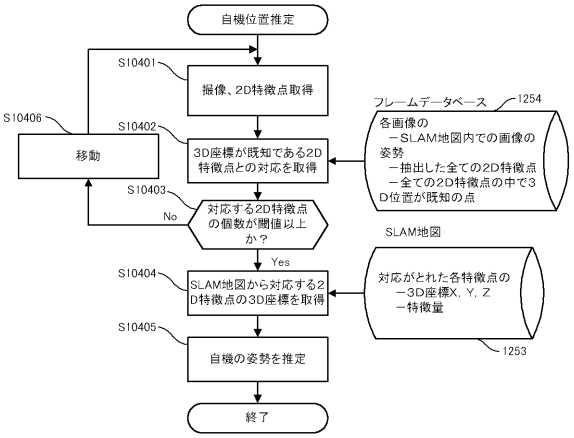
【図７】


音源定位
S10301
時間周波数変換
S10302
相関行列を計算
S10303
固有値分解
S10304
MUSICスペクトルを計算
S10305
重み付け加算
S10306
ピークを選択
終了

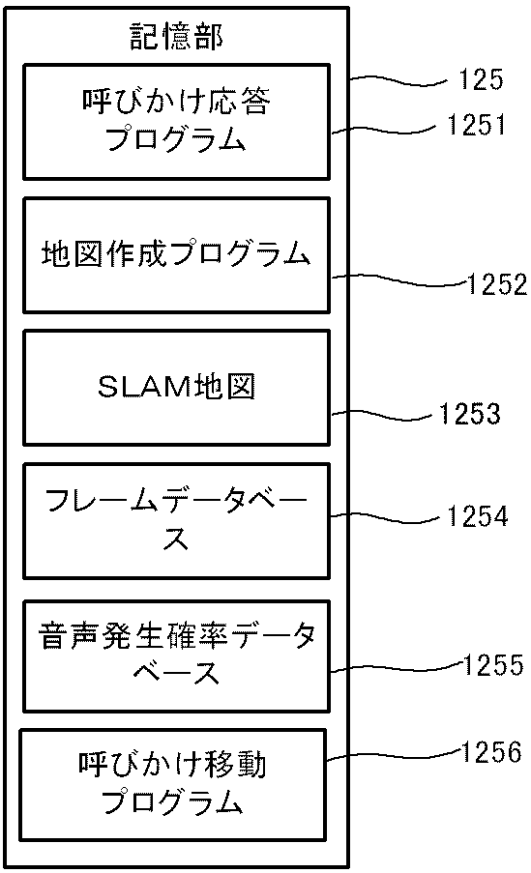
【図８】


y
x
人
仮の音源1
仮の音源2
仮の音源3
仮の音源4
112
100
112
112
θ1
θ2
θ3
θ4

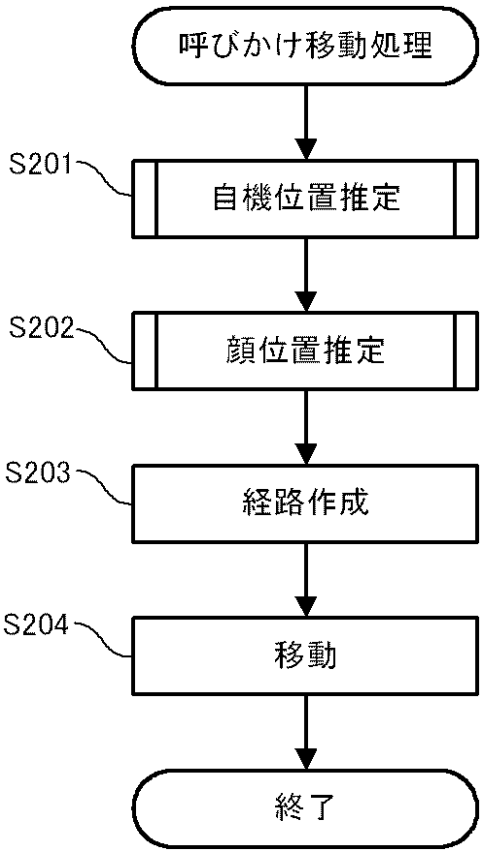
【図９】


自機位置推定
S10401
撮像、2D特徴点取得
S10406
移動
S10402
3D座標が既知である2D特徴点との対応を取得
フレームデータベース
1254
各画像の
－SLAM地図内での画像の姿勢
－抽出した全ての2D特徴点
－全ての2D特徴点の中で3D位置が既知の点
S10403
No
対応する2D特徴点の個数が閾値以上か？
Yes
SLAM地図
S10404
SLAM地図から対応する2D特徴点の3D座標を取得
対応がとれた各特徴点の
－3D座標X，Y，Z
－特徴量
1253
S10405
自機の姿勢を推定
終了

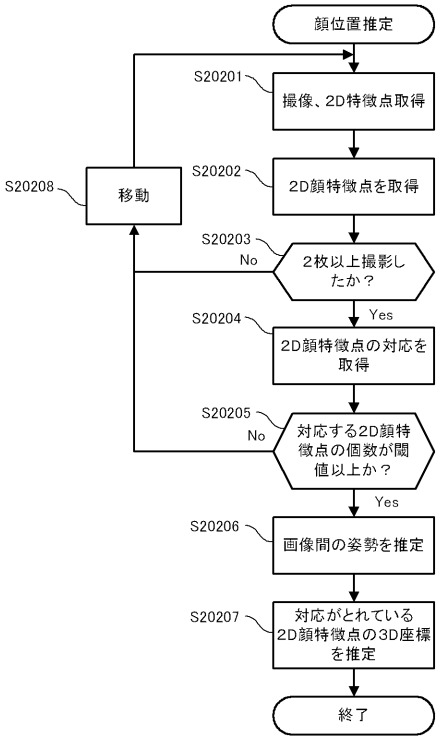
【図１０】


記憶部
呼びかけ応答プログラム
125
1251
地図作成プログラム
1252
SLAM地図
1253
フレームデータベース
1254
音声発生確率データベース
1255
呼びかけ移動プログラム
1256

【図１１】


呼びかけ移動処理
S201
自機位置推定
S202
顔位置推定
S203
経路作成
S204
移動
終了

【図１２】


顔位置推定
S20201
撮像、2D特徴点取得
S20202
2D顔特徴点を取得
S20203
No
2枚以上撮影したか？
Yes
S20204
2D顔特徴点の対応を取得
S20205
No
対応する2D顔特徴点の個数が閾値以上か？
Yes
S20206
画像間の姿勢を推定
S20207
対応がとれている2D顔特徴点の3D座標を推定
終了
S20208
移動

フロントページの続き

審査官　山下　剛史

(56)参考文献　特開２００３－２６６３５１（ＪＰ，Ａ）
特開２０１４－１３７２２６（ＪＰ，Ａ）
特開２００７－２２１３００（ＪＰ，Ａ）
特開２００３－６２７７７（ＪＰ，Ａ）
国際公開第２０１４／１６７７００（WO，Ａ１）
特開２００５－２５０２３３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－９９／００
Ａ６３Ｈ　　１／００－３７／００