

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-10857
(P2010-10857A)

(43) 公開日 平成22年1月14日(2010.1.14)

(51) Int.Cl.			F I			テーマコード (参考)	
HO4R	1/00	(2006.01)	HO4R	1/00	328Z	5D017	
G1OL	21/02	(2006.01)	G1OL	21/02	202A	5D018	
HO4R	3/00	(2006.01)	HO4R	3/00	320	5D020	
HO4M	1/00	(2006.01)	HO4M	1/00	H	5K027	
HO4R	1/40	(2006.01)	HO4R	1/40	320A		

審査請求 未請求 請求項の数 14 OL (全 17 頁)

(21) 出願番号 特願2008-165286 (P2008-165286)
(22) 出願日 平成20年6月25日 (2008. 6. 25)

(71) 出願人 000000295
 沖電気工業株式会社
 東京都港区西新橋三丁目16番11号
 (74) 代理人 100085198
 弁理士 小林 久夫
 (74) 代理人 100098604
 弁理士 安島 清
 (74) 代理人 100125494
 弁理士 山東 元希
 (74) 代理人 100061273
 弁理士 佐々木 宗治
 (74) 代理人 100070563
 弁理士 大村 昇
 (74) 代理人 100087620
 弁理士 高梨 範夫

最終頁に続く

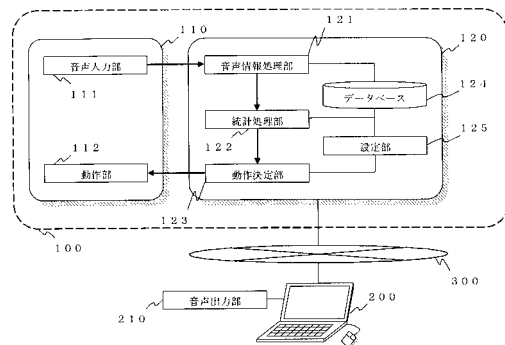
(54) 【発明の名称】 音声入力ロボット、遠隔会議支援システム、遠隔会議支援方法

(57) 【要約】

【課題】 複数人間が参加する音声コミュニケーションを効果的に支援することができる音声入力ロボットを得る。

【解決手段】 音声の入力を受け付ける音声入力部111と、音声入力部111が受け付けた音声の音源位置を推定する音源位置推定部121と、音声入力部111の位置を可変する動作部112と、を備え、音源位置推定部121は、音声入力部111が受け付けた複数の音声の音源位置を推定し、動作部112は、音源位置推定部121の推定結果に基づき、音声入力部111と複数の音声の音源位置との間の位置関係を変更する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

音声の入力を受け付ける音声入力部と、
 前記音声入力部が受け付けた音声の音源位置を推定する音源位置推定部と、
 前記音声入力部の位置を可変する動作部と、
 を備え、
 前記音源位置推定部は、
 前記音声入力部が受け付けた複数の音声の音源位置を推定し、
 前記動作部は、
 前記音源位置推定部の推定結果に基づき、
 前記音声入力部と前記複数の音声の音源位置との間の位置関係を変更する
 ことを特徴とする音声入力ロボット。

10

【請求項 2】

前記音源位置推定部の推定結果を時系列順に保持するデータベースを格納した記憶部を
 備えた
 ことを特徴とする請求項 1 記載の音声入力ロボット。

【請求項 3】

前記データベースが保持する前記推定結果を統計処理する統計処理部を備えた
 ことを特徴とする請求項 2 記載の音声入力ロボット。

【請求項 4】

前記統計処理部は、
 前記統計処理の結果を当該音声入力ロボット周辺の 2 次元平面座標上または 3 次元空間
 座標上にマッピングし、そのマッピング結果を前記記憶部に格納する
 ことを特徴とする請求項 3 記載の音声入力ロボット。

20

【請求項 5】

前記動作部は、
 前記マッピング結果に基づき前記位置関係を変更する
 ことを特徴とする請求項 4 記載の音声入力ロボット。

【請求項 6】

前記位置関係を指定する設定情報を受け付ける設定入力部を備え、
 前記動作部は、
 前記マッピング結果に基づき、
 前記位置関係が前記設定情報で指定される位置関係となるように、
 前記音声入力部の位置を可変する
 ことを特徴とする請求項 4 記載の音声入力ロボット。

30

【請求項 7】

各前記音声に対する前記音声入力部の集音状態を通知する手段を備えた
 ことを特徴とする請求項 1 ないし請求項 6 のいずれかに記載の音声入力ロボット。

【請求項 8】

各前記音声に対する前記音声入力部の集音状態を表示する表示部を備えた
 ことを特徴とする請求項 1 ないし請求項 6 のいずれかに記載の音声入力ロボット。

40

【請求項 9】

前記動作部の可変方向を可変実行前に表示する表示部を備えた
 ことを特徴とする請求項 1 ないし請求項 6 のいずれかに記載の音声入力ロボット。

【請求項 10】

前記動作部は、
 前記音源位置推定部の推定結果に基づき、
 各前記音源位置から前記音声入力部までの距離が等しくなるように、
 前記音声入力部の位置を可変する
 ことを特徴とする請求項 1 ないし請求項 9 のいずれかに記載の音声入力ロボット。

50

【請求項 1 1】

前記動作部は、
 前記音源位置推定部の推定結果に基づき、
 前記音声入力部が受け取る各前記音声の集音音量が等しくなるように、
 前記音声入力部の位置を可変する
 ことを特徴とする請求項 1 ないし請求項 9 のいずれかに記載の音声入力ロボット。

【請求項 1 2】

前記動作部は、
 前記音源位置推定部の推定結果に基づき、
 前記音源位置のうち特定のものから生じる音声の集音音量が所定条件を満たすように、
 前記音声入力部の位置を可変する
 ことを特徴とする請求項 1 ないし請求項 9 のいずれかに記載の音声入力ロボット。

10

【請求項 1 3】

請求項 1 ないし請求項 1 2 のいずれかに記載の音声入力ロボットと、
 音声を出力する音声出力部を備えた端末と、
 を有し、
 前記音声入力ロボットと前記端末はネットワークを介して接続され、
 前記音声入力ロボットは、
 前記ネットワークを介して前記音声入力部が受け付けた音声を前記端末に送信し、
 前記端末は、
 その音声を受信して前記音声出力部よりその音声を音声出力する
 ことを特徴とする遠隔会議支援システム。

20

【請求項 1 4】

遠隔会議を支援する方法であって、
 音声の入力を受け付ける音声入力部と、
 前記音声入力部が受け付けた音声の音源位置を推定する音源位置推定部と、
 前記音声入力部の位置を可変する動作部と、
 を備えた音声入力ロボットを会議空間に配置しておき、
 前記音声入力部が受け付けた複数の音声の音源位置を推定するステップと、
 前記音源位置推定部の推定結果に基づき、前記音声入力部と前記複数の音声の音源位置
 との間の位置関係を変更するステップと、
 を有することを特徴とする遠隔会議支援方法。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声入力部を備えた音声入力ロボット、そのロボットを有する遠隔会議支援システム、そのロボットを用いた遠隔会議支援方法に関するものである。

【背景技術】

【0002】

従来、ロボット装置に関し、『対象物に対してより自然な動作を行えて、エンターテイメント性を向上したロボット装置及びロボット装置の行動制御方法を提供する。』ことを目的とした技術として、『ロボット装置 1 は、CCDカメラ 2 2 と、マイクロホン 2 4 と、画像データから動体を検出する動体検出モジュール 3 2 及び人物の顔を検出する顔検出モジュール 3 3 と、音声データから音源方向を推定する音源方向推定モジュール 3 4 と、上記動体検出結果に基づく動体方向、上記顔検出結果に基づく顔方向、及び上記推定された音源方向の何れかに移動するよう制御する制御手段とを有し、制御手段は、動体方向又は音源推定方向に歩行中に顔検出された場合、顔方向に移動するよう制御すると共に、顔検出対象となっている対象物に所定の範囲内に近づいたとき、歩行を停止するよう制御する。』というものが提案されている（特許文献 1）。

40

【0003】

50

また、自律行動ロボットに関し、『人間に対してペットのような振る舞いで応答し、人間が親近感を感じられる自律行動ロボットのための行動制御装置を提供する。』ことを目的とした技術として、『ステレオカメラによる画像入力装置 1、画像処理によって人物が検出され、人物の顔領域を追跡する人物検出装置 2、ステレオカメラの画像から距離を算出する距離算出装置 3、人物情報記憶部 5 の情報から人物を識別する人物識別装置 4、ボディに付けられたマイクから構成される音声入力装置 6、音源方向検出装置 7、音声認識装置 8、ロボットの前後左右に設置され、障害物検出装置 10 に障害物情報を送出的る超音波センサ 9、撫でられた場合と叩かれた場合に、それぞれを識別できる信号を行動制御装置 12 に送出的るタッチセンサ 11、二つの車輪による脚部モータ 13、頭部を回転させる頭部モータ 14、およびロボットの口に付けられた音声出力装置 15 から構成される。』というものが提案されている（特許文献 2）。

10

【0004】

また、対話型ロボットに関し、『対話する人間の操作負担を増加させることなく、音声認識精度を向上させることのできる対話型ロボットを提供する。』ことを目的とした技術として、『音声認識可能な対話型ロボット 400 であって、音声認識の対象となる目的音声の音源方向を推定する音源方向推定手段と、音源方向推定手段が推定した音源方向に当該対話型ロボット自身を移動させる移動手段と、移動手段による移動後の位置において、目的音声を取得する目的音声取得手段と、目的音声取得手段が取得した目的音声に対して音声認識を行う音声認識手段とを備えた。』というものが提案されている（特許文献 3）

20

【0005】

【特許文献 1】特開 2004 - 130427 号公報（要約）

【特許文献 2】特開 2003 - 326479 号公報（要約）

【特許文献 3】特開 2006 - 181651 号公報（要約）

【発明の開示】

【発明が解決しようとする課題】

【0006】

例えば遠隔コミュニケーションを行う際のように、マイクを通して音声対話を行う環境下では、発話者とマイクの位置関係によって発話者の音声聞き取りにくい場合がある。

特に発話者が複数人存在するような状況では、発話者各人の発話音量の差、マイクとの距離・位置関係などにより、発話者毎に音声の聞き取りやすさが異なってしまう。

30

【0007】

このような状況下では、マイクで集音された音声を聴く立場の者（遠隔コミュニケーションの例では遠隔地の相手側に当たる）は、発話者に対して「声が聴こえにくい」「もう少しマイクに近づいて話をして欲しい」などの要望を出して状況改善を試みる。

しかし、このようなやり取りは発話の中断を招き、コミュニケーションの円滑な進行を妨げ、参加者に余計なストレスを与えてしまう。

【0008】

こうした課題に対して、音声を収集するマイクの性能を上げる、設置数を増やすといった改善方法も考えられるが、これらの環境を整備するためのコストを要する。

40

【0009】

一方、上記特許文献 1～3 に記載の技術では、音声を取得することで音源位置を推定してその方向にロボットが移動等することが開示されている。これは、発話者に近い位置で音声を入力することを図るものであると考えることができる。

しかし、この動作は人間とロボットが対話するためのものであり、遠隔コミュニケーションを円滑に行うためのものではない。

【0010】

例えば、上記特許文献 1～3 に記載の技術を用い、ロボットが移動等することによってロボットとその対話相手の人間との間の距離等を最適化することが考えられる。

しかし、遠隔会議のように複数の人間がコミュニケーションに参加する環境下では、口

50

ボットとその対話相手との２者間関係のみを最適化したとしても、必ずしも会議全体の進行を最適化することにはならない。

即ち、会議に複数の人間が参加している環境、換言すると、複数の音源から生じる音声を全体的に収集することが求められる環境下では、上記特許文献１～３に記載の技術は必ずしも適していない。

【００１１】

そのため、複数の人間が参加する音声コミュニケーションを効果的に支援することができる音声入力ロボットが望まれていた。

【課題を解決するための手段】

【００１２】

本発明に係る音声入力ロボットは、音声の入力を受け付ける音声入力部と、前記音声入力部が受け付けた音声の音源位置を推定する音源位置推定部と、前記音声入力部の位置を可変する動作部と、を備え、前記音源位置推定部は、前記音声入力部が受け付けた複数の音声の音源位置を推定し、前記動作部は、前記音源位置推定部の推定結果に基づき、前記音声入力部と前記複数の音声の音源位置との間の位置関係を変更するものである。

【発明の効果】

【００１３】

本発明に係る音声入力ロボットによれば、複数の音源位置から生じる音声を全体的に収集することができるので、複数の人間が参加する音声コミュニケーションを効果的に支援することができる。

【発明を実施するための最良の形態】

【００１４】

実施の形態１．

図１は、本発明の実施の形態１に係る遠隔会議支援システムの構成図である。

本実施の形態１に係る遠隔会議支援システムは、音声入力ロボット１００、会議端末２００を有する。音声入力ロボット１００と会議端末２００は、例えばＬＡＮ（Local Area Network）やインターネットのようなネットワーク３００を介して遠隔接続されている。

【００１５】

音声入力ロボット１００は、ロボット本体部１１０、ロボット制御部１２０を備える。

ロボット本体部１１０は、音声入力ロボット１００の本体筐体と、本体筐体に取り付けられた各構成部分とを備える。具体的な構成は後述する。

ロボット制御部１２０は、音声入力ロボット１００の動作を制御する。具体的な構成は後述する。ロボット制御部１２０およびその各構成部は、その機能を実現する回路デバイスのようなハードウェアで構成することもできるし、マイコンやＣＰＵ（Central Processing Unit）のような演算装置とその動作を規定するソフトウェアで構成することもできる。また、必要な記憶装置やネットワークインターフェースを適宜備える。

【００１６】

ロボット本体部１１０とロボット制御部１２０は、同一筐体上に構成してもよいし、例えばロボット制御部１２０をロボット本体部１１０から切り離して外部に構成し、有線または無線により相互に通信するように構成してもよい。

【００１７】

ロボット本体部１１０は、音声入力部１１１、動作部１１２を備える。

【００１８】

音声入力部１１１は、例えば複数のマイクロフォンを備えたマイクロフォンアレイなどから構成され、音声入力ロボット１００が存在する周辺の音声を収集する。

音声入力ロボット１００が姿勢を変えることなく全方位からの音声を収集できるようにするためには、マイクロフォンアレイで音声入力部１１１を構成するのが好適である。例えば、単一指向性マイクを円周上に複数配置し、指向方向を円の外側に向ける、といった

10

20

30

40

50

手法が考えられる。

音声入力部 1 1 1 が収集した音声は、後述の音声情報処理部 1 2 1 に出力される。

【 0 0 1 9 】

動作部 1 1 2 は、音声入力ロボット 1 0 0 が存在する空間において、動作決定部 1 2 3 の指示に基づき音声入力部 1 1 1 の空間位置を可変する機能を備える。動作部 1 1 2 の具体的な構成例については、後述の図 2 で説明する。

【 0 0 2 0 】

ロボット制御部 1 2 0 は、音声情報処理部 1 2 1、統計処理部 1 2 2、動作決定部 1 2 3、データベース 1 2 4、設定部 1 2 5 を備える。

【 0 0 2 1 】

音声情報処理部 1 2 1 は、音声入力部 1 1 1 が収集した音声を受け取り、その音声の音源位置を推定し、その推定音源の音量を算出する。推定結果や算出結果は、データベース 1 2 4 に格納される。音源位置の推定手法は、任意の公知技術など適当なものを適宜用いる。

また、音声情報処理部 1 2 1 は、ネットワーク 3 0 0 を介して、音声入力部 1 1 1 から受け取った音声を会議端末 2 0 0 に送信する。

【 0 0 2 2 】

統計処理部 1 2 2 は、データベース 1 2 4 に蓄積されたデータと設定部 1 2 5 が受け取った設定情報から、後述の図 3 ~ 図 5 で説明する統計処理を行い、音声入力ロボット 1 0 0 が存在する空間の音声環境をマップ化 (マッピング) して音声分布マップを作成する。作成したマップはデータベース 1 2 4 に格納される。

統計処理部 1 2 2 が行う統計処理の対象となるのは、音声情報処理部 1 2 1 が処理した前述の各情報、即ち音源の推定位置、推定音源位置の音量、時間 (サンプルングタイム) などである。

【 0 0 2 3 】

動作決定部 1 2 3 は、統計処理部 1 2 2 が作成した音声分布マップと設定部 1 2 5 が受け取った設定情報から、音声入力部 1 1 1 の空間位置可変を実行するか否か、および可変先位置を決定する。決定した結果は可変指令として動作部 1 1 2 に出力される。

【 0 0 2 4 】

データベース 1 2 4 は、音声情報処理部 1 2 1 が処理した前述の各情報、即ち音源の推定位置、推定音源位置の音量などを時系列順に保持する。データベース 1 2 4 は、保持する情報を格納する HDD (Hard Disk Drive) などの記憶装置を用いて構成することができる。情報の格納形式は任意のものでよい。

【 0 0 2 5 】

設定部 1 2 5 は、音声の聴き手側が望む音声環境や音声収集状況、即ち、聴き手側がどのようにして発話者側からの音声を聴きたいのかを設定するための設定情報の入力を受け付ける。設定内容の具体例は後述する。

具体的には、例えばネットワークインターフェースや画面入力を介して、上述の設定情報の入力を受け付ける、といった構成が考えられる。

設定部 1 2 5 が受け取った設定情報と統計処理部 1 2 2 が作成した音声分布マップから、動作決定部 1 2 3 が出力する可変指令の内容が決定される。

【 0 0 2 6 】

なお、設定部 1 2 5 が受け付ける「音声の聴き手側が望む音声環境」とは、例えば以下の (1) ~ (3) のようなことを言う。

【 0 0 2 7 】

(1) 各発話者からほぼ等距離となるような位置関係で発話を聴きたい。

この場合は、複数音源からの音声を同時に取得する。音量が発話者の発話音量に大きく影響を受けるため、声量による発話者の感情を読み取りやすい。

【 0 0 2 8 】

(2) 発話者達の発話を同等の音量で聴きたい。

10

20

30

40

50

この場合も、複数音源からの音声を同時に取得する。発話・主張の強さに関して、音量差の影響を受けにくい。

【0029】

(3) 特定発話者の発話を聴きやすい状況で聴きたい。

これは、特定発話者の発言が多い状況、例えば、ある発話者が資料説明を行っているような状況に相当する。

この場合は、発話頻度の高い話者の位置に着目し、発話者の声が大きすぎるときは音量を小さくし、声が小さすぎるときは音量を大きくしたい、といった要望があるものと想定される。

【0030】

このように、設定部125を介して集音状態の設定変更を可能としたことにより、人間同士のコミュニケーションを音声入力ロボット100が効果的に支援することができるのである。

この点は、上記特許文献1～3に記載の対話型ロボットのように、プログラムされた規定の目的に従った動作しか行わない、人間とロボットの間のコミュニケーションを前提とした技術とは異なる。

【0031】

本実施の形態1における「音源位置推定部」は、音声情報処理部121が相当する。

また、「動作部」は、動作部112およびその動作内容を決定する動作決定部123が相当する。

【0032】

会議端末200は、遠隔の会議参加者が使用する端末であり、例えばノートパソコンのようなコンピュータを用いて構成することができる。また、例えばスピーカ等で構成された音声出力部210を備える。

会議端末200は、ネットワーク300を介して、ロボット制御部120が送信した音声を受信し、音声出力部210よりその音声を音声出力する。遠隔の会議参加者は、その音声を聴取することにより、音声入力ロボット100周辺の会議参加者の音声を聴くことができる。

【0033】

以上、本実施の形態1に係る遠隔会議支援システムの構成を説明した。

次に、動作部112の具体的な構成例を説明する。

【0034】

図2は、音声入力ロボット100の外観構成例を示す図である。図2(a)は自走式、図2(b)は固定可動式の構成例である。

【0035】

図2(a)に示す自走式構成では、動作部112は平面上を任意方向に移動可能な車両で構成され、その車両台座の上にマイクで構成された音声入力部111を複数設置する構成とした。

車両で構成された動作部112は、動作決定部123の指示に基づき車輪を駆動させ、指示された方向に音声入力ロボット100を移動させる。

【0036】

図2(b)に示す固定可動式構成では、動作部112は底部台座に固定された可動スイングアームで構成され、可動スイングアームの上に固定された台座上にマイクで構成された音声入力部111を複数設置する構成とした。

可動スイングアームで構成された動作部112は、動作決定部123の指示に基づきアームの姿勢(ヨー・ピッチ角)や長さを可変することで、音声入力部111の空間位置を移動させる。

【0037】

以上、動作部112の具体的な構成例を説明した。

次に、統計処理部122が作成する音声分布マップの例について、先に述べた設定部1

10

20

30

40

50

25にとの関連から説明する。

【0038】

図3は、音源位置のみを基準として作成した音声分布マップの例である。以下、図3を用いて音声入力部111の空間位置を変更する過程を説明する。

なお、ここでは上述の設定情報として、「(1)各発話者からほぼ等距離となるような位置関係で発話を聴きたい。」が設定部125に入力されたものとする。

【0039】

図3(a)は、会議参加者と音声入力ロボット100の初期状態を示す図である。同図において、1~3は会議参加者の位置、黒三角は音声入力ロボット100の初期位置を示している。

図3(a)の状態では、音声入力ロボット100と会議参加者2の距離が最も近く、他の会議参加者と音声入力ロボット100の距離は遠い。

【0040】

音声情報処理部121は、所定のサンプリングタイム内で、音声入力部111より会議参加者1~3の発話音声を受け取り、各会議参加者の音源位置を推定してデータベース124に格納する。

統計処理部122は、各会議参加者の音源位置の推定結果を用いて、図3(a)のような2次元平面座標上に各会議参加者の位置をマッピングした音声分布マップを作成する。

【0041】

図3(b)は、動作決定部123が音声入力ロボット100の移動先を決定する様子を示す図である。

動作決定部123は、図3(a)に示す音声分布マップおよび設定部125が受け取った設定情報に基づき、音声入力ロボット100(または音声入力部111、以下同様)と各会議参加者の距離がそれぞれ等距離となるように、音声入力ロボット100の移動先を決定する。

【0042】

図3(c)は、音声入力ロボット100が移動した後の音声分布マップである。音声入力ロボット100の空間位置が移動することにより、音声入力ロボット100と各会議参加者の距離がそれぞれ等距離となる。

【0043】

図4は、音源位置と各音源の音量を基準として作成した音声分布マップの例である。以下、図4を用いて音声入力部111の空間位置を変更する過程を説明する。

なお、ここでは上述の設定情報として、「(2)発話者達の発話を同等の音量で聴きたい。」が設定部125に入力されたものとする。

【0044】

図4(a)は、会議参加者と音声入力ロボット100の初期状態を示す図である。同図において、1~3は会議参加者の位置、円の大きさは各会議参加者の発話音量、黒三角は音声入力ロボット100の初期位置を示している。

図4(a)の状態では、音声入力ロボット100と会議参加者1の距離が最も近く、これに対応して会議参加者1から集音される音量が最も大きくなっている。

【0045】

音声情報処理部121は、所定のサンプリングタイム内で、音声入力部111より会議参加者1~3の発話音声を受け取り、各会議参加者の音源位置を推定してデータベース124に格納する。また、各会議参加者の発話音量を算出してデータベース124に格納する。

ここでいう発話音量とは、例えばサンプリングタイム内での最大/最小音量、あるいはサンプリングタイム内での音量の平均値、といった値のことである。

統計処理部122は、各会議参加者の音源位置の推定結果を用いて、図4(a)のような2次元平面座標上に各会議参加者の位置と発話音量をマッピングした音声分布マップを作成する。

10

20

30

40

50

【 0 0 4 6 】

図 4 (b) は、動作決定部 1 2 3 が音声入力ロボット 1 0 0 の移動先を決定する様子を示す図である。

動作決定部 1 2 3 は、図 4 (a) に示す音声分布マップおよび設定部 1 2 5 が受け取った設定情報に基づき、音声入力ロボット 1 0 0 が集音する各会議参加者の発話音量がそれぞれ同等になるように、音声入力ロボット 1 0 0 の移動先を決定する。

【 0 0 4 7 】

図 4 (c) は、音声入力ロボット 1 0 0 が移動した後の音声分布マップである。音声入力ロボット 1 0 0 の空間位置が移動することにより、音声入力ロボット 1 0 0 が集音する各会議参加者の発話音量 (円の大きさ) がそれぞれ同等になる。

10

【 0 0 4 8 】

図 5 は、音源位置、各音源の音量、および各音源の音声発生頻度を基準として作成した音声分布マップの例である。以下、図 5 を用いて音声入力部 1 1 1 の空間位置を変更する過程を説明する。

なお、ここでは上述の設定情報として、「(3) 特定発話者の発話を聴きやすい状況で聴きたい。」が設定部 1 2 5 に入力されたものとする。

【 0 0 4 9 】

図 5 (a) は、会議参加者と音声入力ロボット 1 0 0 の初期状態を示す図である。同図において、1 ~ 3 は会議参加者の位置、円の大きさは各会議参加者の発話音量、円の輪数は発話回数、黒三角は音声入力ロボット 1 0 0 の初期位置を示している。

20

なお、聴き手側は、会議参加者 3 の発話を聴きやすい状況を希望しているものとする。

【 0 0 5 0 】

音声情報処理部 1 2 1 は、所定のサンプリングタイム内で、音声入力部 1 1 1 より会議参加者 1 ~ 3 の発話音声を受け取り、各会議参加者の音源位置を推定してデータベース 1 2 4 に格納する。また、各会議参加者の発話音量と発話回数を算出してデータベース 1 2 4 に格納する。

統計処理部 1 2 2 は、各会議参加者の音源位置の推定結果を用いて、図 5 (a) のような 2 次元平面座標上に各会議参加者の位置、発話音量、および発話回数をマッピングした音声分布マップを作成する。

【 0 0 5 1 】

30

図 5 (b) は、動作決定部 1 2 3 が音声入力ロボット 1 0 0 の移動先を決定する様子を示す図である。

動作決定部 1 2 3 は、図 5 (a) に示す音声分布マップおよび設定部 1 2 5 が受け取った設定情報に基づき、音声入力ロボット 1 0 0 が集音する会議参加者 3 の発話音量が最も大きくなるように、音声入力ロボット 1 0 0 の移動先を決定する。

【 0 0 5 2 】

図 5 (c) は、音声入力ロボット 1 0 0 が移動した後の音声分布マップである。

音声入力ロボット 1 0 0 の空間位置が移動することにより、音声入力ロボット 1 0 0 が集音する会議参加者 3 の発話音量 (円の大きさ) が最も大きくなり、他の会議参加者の発話音量は小さくなる。

40

なお、音声入力ロボット 1 0 0 が移動しても発話回数自体は変化しないため、各円の輪数は変化しない。

【 0 0 5 3 】

以上、統計処理部 1 2 2 が作成する音声分布マップの例を説明した。

【 0 0 5 4 】

なお、動作決定部 1 2 3 は、音声入力ロボット 1 0 0 自体から発生する音や、音声入力ロボット 1 0 0 が移動することで集音状態が変化することを考慮し、移動先を決定した後すぐに移動指示を出すのではなく、以下のような条件のいずれかが満たされたときに動作部 1 1 2 へ移動指示を出す。

【 0 0 5 5 】

50

(条件1) ある単位時間、各音源からの音声の発生がない状態が継続する。

(条件2) 各音源から発生する音量が一定レベル以下の状態になる。

【0056】

また、音声入力ロボット100の移動中は、上記と同様に、音声入力ロボット100自体から発生する音や、音声入力ロボット100が移動することで集音状態が変化することを考慮し、統計処理を中断する。

具体的には、動作決定部123より統計処理部122にその旨を指示するとよい。

【0057】

図6は、聴き手側の望む音声状況(音声入力ロボット100の集音状態)になるように音声入力ロボット100を動作させ、音声環境を改善する動作フローである。ここでは、遠隔会議の場面を想定する。以下、図6の各ステップについて説明する。

【0058】

(S601)

音声入力部111を通しての音声のやり取りが終了するまで、以下のステップが繰り返される。音声のやり取りが終了するとは、例えば遠隔会議が終了することを指す。

(S602)

音声入力部111は、音声入力ロボット100が存在する空間、ここでは発話側の会議室の音声を取得する。取得した音声は、ロボット制御部120へ送信される。

【0059】

(S603)

音声情報処理部121は、音声入力部111から受け取った音声に基づき、音源位置の推定、推定音源の音量、推定音源の音声出力回数、などの演算処理を実行する。また、音声入力部111から受け取った音声を会議端末200に送信する。

(S604)

音声情報処理部121は、ステップS603の結果をデータベース124に格納する。

(S605)

音声入力ロボット100が移動中である場合はステップS611へ進み、移動中でない場合はステップS606へ進む。

【0060】

(S606)

統計処理部122は、データベース124に格納されている各データ、および設定部125が受け取った設定情報(聴き手側が望む音声環境)に基づき、先に説明した統計処理を実行する。

(S607)

統計処理部122は、ステップS606の処理結果に基づき、図3~図5で説明したような音声分布マップを作成する。作成した音声分布マップは、任意のデータ形式でデータベース124に格納する。

【0061】

(S608)

動作決定部123は、ステップS607で作成された音声分布マップ、および設定部125が受け取った設定情報に基づき、音声環境を聴き手側が望むように改善するために、音声入力ロボット100の位置を変更する必要があるか否かを判定する。

位置を変更する必要がある場合はステップS609へ進み、必要がない場合はステップS602に戻って繰り返し処理を継続する。

(S609)

動作決定部123は、ステップS607で作成された音声分布マップ、および設定部125が受け取った設定情報に基づき、音声入力ロボット100の移動先位置を決定する。

【0062】

(S610)

動作決定部123は、音声入力ロボット100の移動・動作を開始・実行してよいか否

10

20

30

40

50

かを判断する。ここでの判断とは、上述の条件 1 ~ 2 が満たされているか否かを判断することを指す。

音声入力ロボット 100 の移動・動作を許可する場合はステップ S 6 1 1 へ進み、許可しない場合はステップ S 6 0 2 に戻って繰り返し処理を継続する。

(S 6 1 1)

動作決定部 1 2 3 は、動作部 1 1 2 に動作指令を出す。動作部 1 1 2 は、その動作指令に基づき音声入力ロボット 100 を駆動して音声入力部 1 1 1 の空間位置を可変する。

【0063】

以上、音声入力ロボット 100 を動作させて音声環境を改善するフローを説明した。

音声入力ロボット 100 を動作させることにより、音声入力部 1 1 1 の集音状態が聴き手側の望む状態に変化する。

【0064】

以上のように、本実施の形態 1 によれば、複数の音源位置から生じる音声を設定部 1 2 5 が受け取った設定情報に合致する条件の下で全体的に収集することができるので、遠隔会議のように複数の人間が参加する音声コミュニケーションを効果的に支援することができる。

【0065】

また、本実施の形態 1 によれば、例えば遠隔会議のように音声入力手段を通して音声のやり取りを行う環境において、聴き手側の望む音声状況（音声入力部 1 1 1 の集音状態）となるように音声入力ロボット 100 を移動させて音声環境を改善することができる。

【0066】

また、本実施の形態 1 によれば、聴き手側の望む音声環境を得るという受話者側の利点以外にも、発話者側にとっての利点もある。

【0067】

従来の遠隔会議に関する技術では、発話状況が聴き手側にどのように聴こえているかについて発話者側へのフィードバックが乏しい。

例えば、聴き手側から「声がよく聴こえない」といった会話によるフィードバックを得る以外に、フィードバックを得る手段がない。したがって、聴き手側が会話によるフィードバックをしなければ、発話者側が得られるフィードバックはない。

また、聴き手側から会話によるフィードバックを都度行っているようでは、円滑なコミュニケーションの妨げになる。

【0068】

この課題につき、本実施の形態 1 によれば、発話者側の会議空間において音声入力ロボット 100 が実際に移動することそのものが、聴き手側が集音状態の改善を望んでいるというフィードバックを発話者に与えることになる。

発話者側は、例えば音声入力ロボット 100 が自分に近づいてくるといった動作を見ることで、自分の発話が聴き手側によく聴こえていないのではないかと、といったことに気づくことができる。

【0069】

この点、音声信号の増幅演算処理などのソフトウェア処理によって集音状態を改善することも考えられる。

これに対し、本実施の形態 1 では、音声入力ロボット 100 自体が移動するという動作により、集音状態の改善と、発話者へのフィードバックとを、同時に行うことができるのである。

【0070】

実施の形態 2 .

実施の形態 1 では、音声入力ロボット 100 が移動する際に、音声入力ロボット 100 自身から発生する音の影響や、音声入力ロボット 100 が移動することによる集音状態の変化に鑑み、所定の条件を満たすまでは音声入力ロボット 100 の移動を許可しないこととした。

10

20

30

40

50

【 0 0 7 1 】

こうした動作を行う場合、音声入力ロボット 1 0 0 に対する移動指示が出てから実際に移動するまでにタイムラグが生じる。したがって、音声入力ロボット 1 0 0 の移動により発話者側へ聴き手側の要望を間接的にフィードバックするのが遅れてしまう。

音声入力ロボット 1 0 0 の移動やフィードバックが遅れば、その分だけ聴き手側の要望が反映されるのが遅れ、発話を聴き取りづらい状態が継続することを余儀なくされる。

【 0 0 7 2 】

そこで、本実施の形態 2 では、上述のようなフィードバックの遅れを解消し、発話者側の注意を喚起して発話状況の改善（発話者が位置を変える、音量を上げる、など）を促すことを図る。

【 0 0 7 3 】

図 7 は、本発明の実施の形態 2 に係る遠隔会議支援システムの構成図である。

本実施の形態 2 に係る遠隔会議支援システムは、実施の形態 1 の図 1 で説明した構成に加え、ロボット本体部 1 1 0 に表示部 1 1 3 を備える。その他の構成は図 1 と概ね同様であるため、以下では差異点を中心に説明する。

【 0 0 7 4 】

表示部 1 1 3 は、動作決定部 1 2 3 の指示に基づき、音声入力ロボット 1 0 0 の移動方向や移動位置を表示する機能部である。

動作決定部 1 2 3 は、統計処理部 1 2 2 の統計処理に基づき音声入力ロボット 1 0 0 の移動先位置や方向を決定した後、動作部 1 1 2 にその旨の指示を出す前に、表示部 1 1 3 にその位置や方向を表示させる。

【 0 0 7 5 】

このように、音声入力ロボット 1 0 0 に対する移動指示が生じた際に、実際の移動によって初めてその内容を表面化させるのではなく、事前に表示することにより、発話者は聴き手側にどのように音声が行っているのかを間接的に知ることができる。

また、表示のみを行うので、音声入力ロボット 1 0 0 の移動による音声環境の変化を生じさせることもない。

【 0 0 7 6 】

一方、移動方向や位置を表示することによって、音声入力ロボット 1 0 0 が移動しようとしていることを発話者に知らせ、以下の効果を発揮する。

即ち、音声入力ロボット 1 0 0 が移動を開始するために、発話者は発話を一時中断し、音声入力ロボット 1 0 0 の移動が完了するまで発話の間を空ける、といった行動をとることが可能になる。

【 0 0 7 7 】

図 8 は、表示部 1 1 3 の構成例を示す図である。図 8 (a) はプロジェクタを用いて表示部 1 1 3 を構成した例、図 8 (b) は LED (Light Emitting Diode) を用いて表示部 1 1 3 を構成した例を示している。

【 0 0 7 8 】

図 8 (a) の例では、プロジェクタを用いて構成された表示部 1 1 3 は、音声入力ロボット 1 0 0 が移動しようとしている方向を、矢印のような図形や文字等を用いて音声入力ロボット 1 0 0 の周辺空間に投射する。

具体的には、例えば矢印の向きで移動方向を表し、矢印の長さで移動距離を表す、といった手法が考えられる。これ以外の手法でもよいし、矢印や文字以外の表現方法を用いてもよい。

【 0 0 7 9 】

図 8 (b) の例では、音声入力ロボット 1 0 0 の周辺に円周方向に LED を複数配置して、音声入力ロボット 1 0 0 が移動しようとしている方向の LED を点灯させることにより、移動方向を表示する。

【 0 0 8 0 】

図 8 (a) (b) いずれの場合でも、音声入力ロボット 1 0 0 が移動しようとしていな

10

20

30

40

50

いときは表示をOFFしておく。

【0081】

図9は、本実施の形態2において、聴き手側の望む音声状況（音声入力ロボット100の集音状態）になるように音声入力ロボット100を動作させ、音声環境を改善する動作フローである。図6と同様に、遠隔会議の場面を想定する。以下、図9の各ステップについて説明する。

【0082】

(S901)～(S909)

図6のステップS601～S609と同様であるため、説明を省略する。

(S910)

動作決定部123は、表示部113に音声入力ロボット100の移動方向を表示するよう指示を出す。表示部113は、その指示に基づき音声入力ロボット100の移動方向を表示する。

【0083】

(S911)

動作決定部123は、音声入力ロボット100の移動・動作を開始・実行してよいか否か、即ち実施の形態1で説明した条件1～2が満たされているか否かを判断する。

音声入力ロボット100の移動・動作を許可する場合はステップS912へ進み、許可しない場合はステップS901のループを継続する。

(S912)

図6のステップS611と同様であるため、説明を省略する。

【0084】

以上、本実施の形態2において、音声入力ロボット100を動作させて音声環境を改善するフローを説明した。

【0085】

表示部113の表示内容は、必ずしも音声入力ロボット100の移動先に関する情報のみでなくともよい。即ち、聴き手側が聴取している集音状態を、直接・間接を問わず発話者に何らかの形でフィードバックすることができればよい。

例えば、発話を中断したくないが、現在の発話内容に対して質問がある、といった聴き手側の意思を表示して発話者にその旨をフィードバックしてもよい。このような表示により、コミュニケーションの円滑を図ることができる。

【0086】

このように、少なくとも集音状態を示唆する情報を発話者に通知する手段を備えることにより、本実施の形態2と同様の効果を発揮することができるのである。

【0087】

以上のように、本実施の形態2によれば、表示部113は、音声入力ロボット100が移動しようとしていることを事前に表示するので、聴き手側で音声かどのように聴こえているかを発話者に間接的にフィードバックすることが可能となる。これにより、発話者は聴き手側を意識した発話を行うことができる。

フィードバックを得た発話者は、音声入力ロボット100が移動しようとしている方向から、自己の発話状態を変更する、音声入力ロボット100の移動開始条件を満たすように発話の間を取る、といった対応を取ることができる。

【0088】

実施の形態3

以上の実施の形態1～2では、音声情報処理部121は音声入力部111から受け取った音声をそのまま会議端末200に送信することとした。

音声情報処理部121は、必要に応じて、音声入力部111から受け取った音声に対して、発話者側のノイズ除去、その他のノイズキャンセリング処理などを施した上で、会議端末200に送信するようにしてもよい。

【0089】

10

20

30

40

50

ここでいう発話者側のノイズとは、例えばPCのファン動作音などが挙げられる。

なお、ノイズキャンセリング処理を施す際には、データベース124に蓄積された音声データを用いて必要な統計処理や学習処理を行うとよい。

【0090】

実施の形態4 .

特許文献1～3に記載されているような従来技術では、対話ロボットが以後の動作を実行する方向を絞り込むために音源位置を推定し、動作方向の候補から外れた音源に関しては、以後の処理対象から除外している。

一方、以上の実施の形態1～3では、推定した音源位置や音量を除外するといった、音源の取捨選択は行わない。

これは、従来技術のように対話ロボットと発話者が1対1で対話することを意識した技術と異なり、本発明では複数の発話者の音声を集音することを目的としたものであることによる。

即ち本発明では、音源位置を処理対象から除外する必要はないため、音源位置の取捨選択は行わないのである。

【0091】

ただし、聴き手側が発した音声を発話者側で音声出力するスピーカ等の音声出力手段に関しては、発話者側の音源位置推定には不要であるため、例外的に処理対象から除外してもよい。これは、上述の各実施の形態で共通である。

【0092】

実施の形態5 .

図3～図5において、音声分布マップを2次元平面座標上で表した例を説明したが、3次元空間座標上に音声分布をマッピングしてもよい。例えば、音声の大きさや発話回数を高さで表現する、といった手法が考えられる。後者の場合は、円の輪数が等高線のように用いられて高さが表現されるイメージとなる。

【0093】

さらには、音声入力部111の配置や音声入力ロボット100の移動範囲を3次元に拡張してもよい。必要な移動手段は適宜設ける。

例えば遠隔会議では、発話者がノートパソコンを自己の目の前に広げて会議を行うことがあり、ノートパソコンが壁になって音声収集に影響を与える。そこで、上記のように高さ方向にも音声入力部111の配置や音声入力ロボット100の移動範囲を拡張し、より柔軟な音声収集を行うことができるようにするとよい。

【0094】

実施の形態6 .

以上の実施の形態1～5において、ロボット制御部120に、音声入力ロボット100の自己位置を推定する機能部を設けてもよい。

例えば、図2(a)で説明した自走式構成の場合は、車輪の回転方向、回転数、車輪直径などの値を用いて自己位置を推定する。

図2(b)で説明した固定可動式構成の場合は、アームの長さ、アームの姿勢(ヨー・ピッチ角)などの値を用いて自己位置を推定する。

自己位置推定を用いることにより、図3～図5で説明した音声分布マップは、音声入力ロボット100の位置を中心とした相対座標系ではなく、絶対座標系のマップとなる。絶対座標軸上の音源における最大/最小音量、発話発生頻度などに基づき、絶対座標上における音声入力ロボット100の理想位置が求められる。

これにより、音声入力ロボット100の理想位置を素早く判断することができる。

【0095】

実施の形態7 .

以上の実施の形態1～6では、説明の便宜上、発話者側に音声入力ロボット100を設置し、聴き手側に会議端末200を設置した例を説明した。

しかし、遠隔会議のような双方向のコミュニケーションでは、双方が発話を行うので、

10

20

30

40

50

双方の拠点に音声入力ロボット100と会議端末200を設置して同等の環境となるように構成してもよい。

【図面の簡単な説明】

【0096】

【図1】実施の形態1に係る遠隔会議支援システムの構成図である。

【図2】音声入力ロボット100の外観構成例を示す図である。

【図3】音源位置のみを基準として作成した音声分布マップの例である。

【図4】音源位置と各音源の音量を基準として作成した音声分布マップの例である。

【図5】音源位置、各音源の音量、および各音源の音声発生頻度を基準として作成した音声分布マップの例である。

【図6】聴き手側の望む音声状況になるように音声入力ロボット100を動作させ、音声環境を改善する動作フローである。

【図7】実施の形態2に係る遠隔会議支援システムの構成図である。

【図8】表示部113の構成例を示す図である。

【図9】実施の形態2において、聴き手側の望む音声状況になるように音声入力ロボット100を動作させ、音声環境を改善する動作フローである。

【符号の説明】

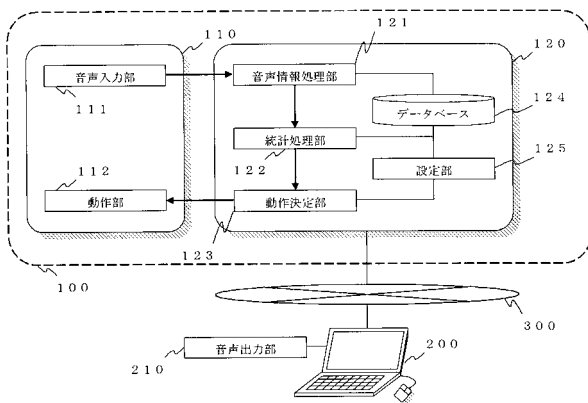
【0097】

100 音声入力ロボット、110 ロボット本体部、111 音声入力部、112 動作部、113 表示部、120 ロボット制御部、121 音声情報処理部、122 統計処理部、123 動作決定部、124 データベース、125 設定部。

10

20

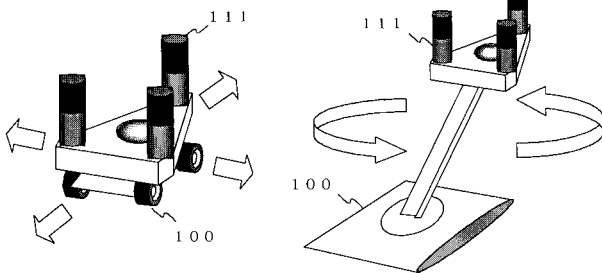
【図1】



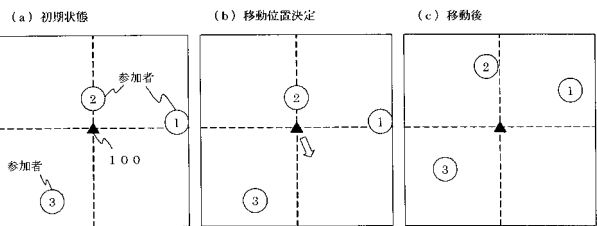
【図2】

(a) 自走式

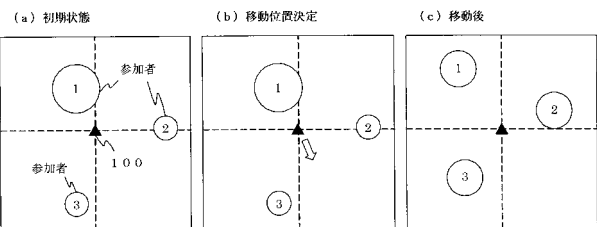
(b) 固定可動式



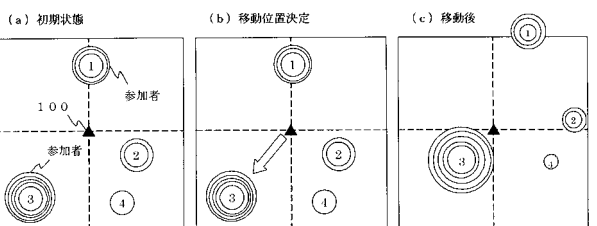
【図3】



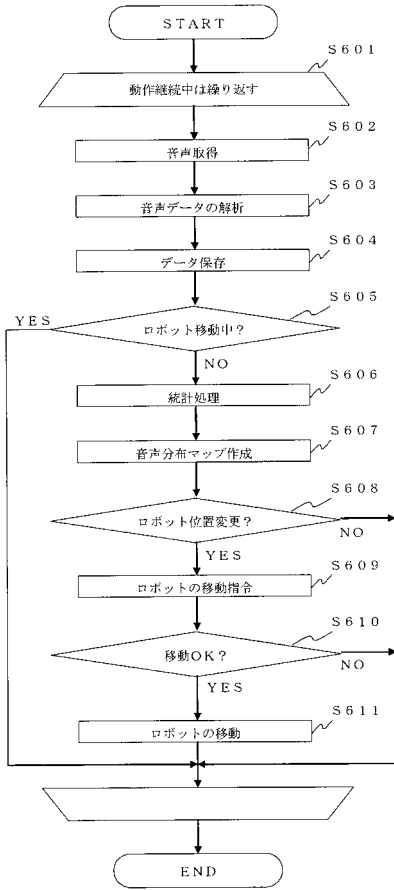
【図4】



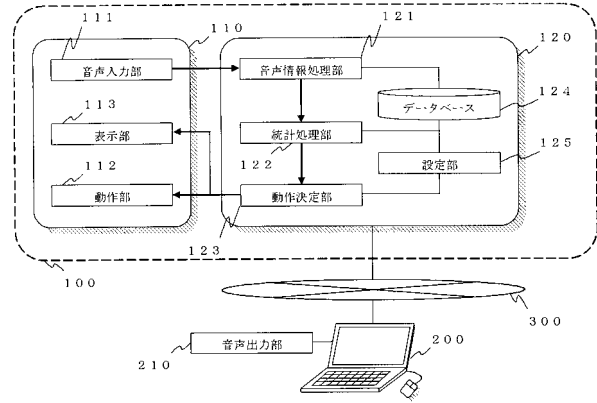
【図5】



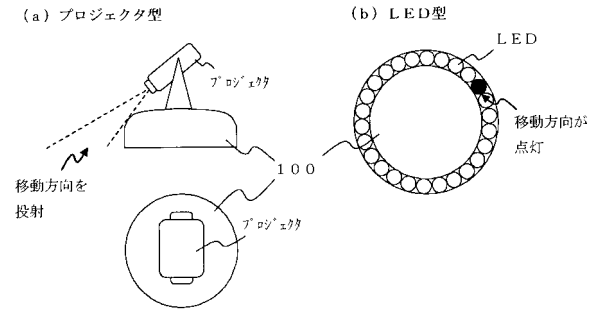
【図6】



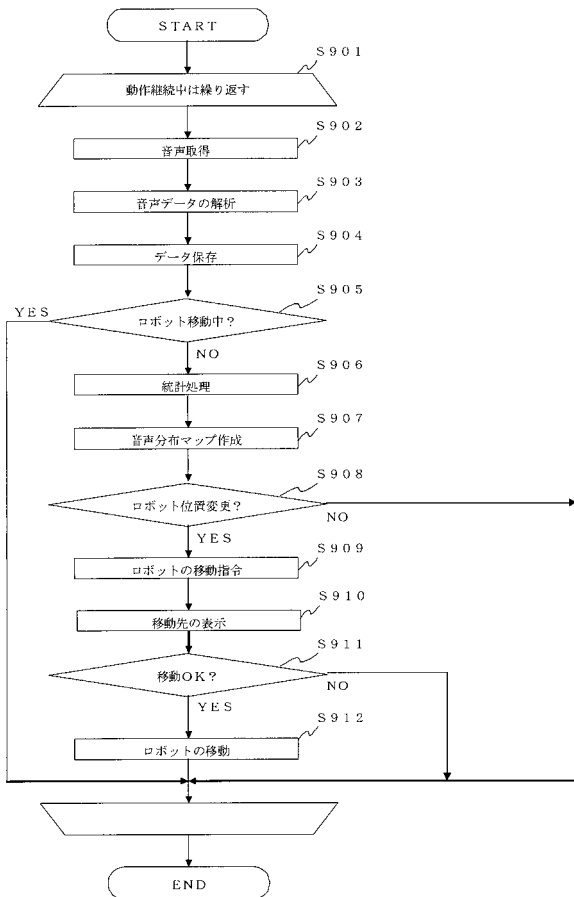
【図7】



【図8】



【図9】



フロントページの続き

(72)発明者 福島 寛之

東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社内

Fターム(参考) 5D017 BB01

5D018 BB22

5D020 BB04

5K027 AA07 BB03 DD14 FF22 HH20