

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5170440号
(P5170440)

(45) 発行日 平成25年3月27日(2013.3.27)

(24) 登録日 平成25年1月11日(2013.1.11)

(51) Int. Cl.		F I			
HO4R	3/00	(2006.01)	HO4R	3/00	320
HO4R	1/40	(2006.01)	HO4R	1/40	320A
B25J	13/08	(2006.01)	B25J	13/08	Z
GO1S	5/20	(2006.01)	GO1S	5/20	
GO1S	3/802	(2006.01)	GO1S	3/802	

請求項の数 5 (全 17 頁)

(21) 出願番号	特願2008-514510 (P2008-514510)	(73) 特許権者	000005326
(86) (22) 出願日	平成19年5月9日(2007.5.9)		本田技研工業株式会社
(86) 国際出願番号	PCT/JP2007/059599		東京都港区南青山二丁目1番1号
(87) 国際公開番号	W02007/129731	(74) 代理人	100077805
(87) 国際公開日	平成19年11月15日(2007.11.15)		弁理士 佐藤 辰彦
審査請求日	平成22年4月16日(2010.4.16)	(74) 代理人	100081477
(31) 優先権主張番号	60/799,106		弁理士 堀 進
(32) 優先日	平成18年5月10日(2006.5.10)	(74) 代理人	100099690
(33) 優先権主張国	米国 (US)		弁理士 鷺 健志
		(74) 代理人	100109232
			弁理士 本間 賢一
		(74) 代理人	100125210
			弁理士 加賀谷 剛

最終頁に続く

(54) 【発明の名称】 音源追跡システム、方法、およびロボット

(57) 【特許請求の範囲】

【請求項1】

音源の位置をリアルタイムで測定する音源追跡システムであって、
 移動機能を有する移動体に搭載されている第1マイクロフォンアレイからの出力と、前記第1マイクロフォンアレイの姿勢とに基づき、ビームフォーミング法にしたがって前記音源の位置または方位を測定する第1処理部と、
 固定して配置されている第2マイクロフォンアレイからの出力に基づき、前記音源の位置を測定する第2処理部と、
 前記音源の速度を測定し、当該測定結果に基づいてパーティクルフィルタの遷移モデルを切り替えながら、前記パーティクルフィルタを用いて前記第1および第2処理部のそれぞれの測定結果を統合することにより前記音源の位置を測定する第3処理部と、を備えていることを特徴とする音源追跡システム。

10

【請求項2】

請求項1記載の音源追跡システムにおいて、
 前記第2処理部が、前記第2マイクロフォンアレイを構成する複数のマイクロフォンのそれぞれと前記音源との距離を推定し、当該複数のマイクロフォンのうち当該算定距離が閾値以下の該マイクロフォンからの出力のみに基づいて前記音源の位置を測定することを特徴とする音源追跡システム。

【請求項3】

請求項1記載の音源追跡システムにおいて、

20

前記第3処理部が、前記第1および第2処理部のそれぞれによる測定結果の尤度を算定し、当該尤度の重み付き和に基づいて前記音源の位置を測定することを特徴とする音源追跡システム。

【請求項4】

音源の位置をリアルタイムで測定する音源追跡方法であって、

移動機能を有する移動体に搭載されている第1マイクロフォンアレイからの出力と、前記第1マイクロフォンアレイの姿勢とに基づき、ビームフォーミング法にしたがって前記音源の位置または方位を測定する第1処理と、

固定して配置されている第2マイクロフォンアレイからの出力に基づき、前記音源の位置を測定する第2処理と、

前記音源の速度を測定し、当該測定結果に基づいてパーティクルフィルタの遷移モデルを切り替えながら、前記パーティクルフィルタを用いて前記第1および第2処理のそれぞれの測定結果を統合することにより前記音源の位置を測定する第3処理と、を実行することを特徴とする音源追跡方法。

【請求項5】

音源の位置をリアルタイムで測定する音源追跡システムにおいて利用されるロボットであって、

前記音源追跡システムが、

第1マイクロフォンアレイからの出力と、前記第1マイクロフォンアレイの姿勢とに基づき、ビームフォーミング法にしたがって前記音源の位置または方位を測定する第1処理部と、

固定して配置されている第2マイクロフォンアレイからの出力に基づき、前記音源の位置を測定する第2処理部と、

前記音源の速度を測定し、当該測定結果に基づいてパーティクルフィルタの遷移モデルを切り替えながら、前記パーティクルフィルタを用いて前記第1および第2処理部のそれぞれの測定結果を統合することにより前記音源の位置を測定する第3処理部と、を備え、

前記第1マイクロフォンアレイが搭載され、

前記第3処理部により測定された前記音源の位置に基づいて前記ロボットの挙動を制御する制御システムを備えていることを特徴とするロボット。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音源の位置をリアルタイムで測定する音源追跡システム等に関する。

【背景技術】

【0002】

知覚のロバスト性向上の観点から、さまざまな情報を統合することは本質的である。たとえば、視聴覚の時間的な統合方法（文献1参照）、音声認識におけるMcGruck効果（文献2参照）、音源定位における視聴覚統合方法（文献3参照）等が知られている。さらに、ロボットに搭載されているマイクロフォンにより音源定位を行う手法が提案されており（文献4参照）、また、天井等に固定されたマイクロフォンアレイによる音源定位を行う手法が提案されている（文献5参照）。

文献1：「マルチドメイン音声対話システムの構築手法」長森誠ほか

情報処理学会研究報告，2000-SLP-31-7，2000

文献2：「複数の音声対話システムの統合制御手法とその評価」河口信夫ほか

情報処理学会研究報告，2001-SLP-36-10，2001

文献3：I. O'Neill et al. "Cross domain dialogue modeling." An object-based approach. In Proc. ICSLP, Vol.1, 2004

文献4：特開2004-198656号公報

文献5：特開2006-121709号公報

しかし、当該先行技術には、音源定位追跡に関するロバスト性および精度に鑑みて改善

の余地があった。

【0003】

そこで、本発明は、ロバスト性および精度の観点から音源定位追跡のパフォーマンスの向上を図りうる音源追跡システム等を提供することを解決課題とする。

【発明の開示】

【0004】

本発明の音源追跡システムは、音源の位置をリアルタイムで測定する音源追跡システムであって、移動機能を有する移動体に搭載されている第1マイクロフォンアレイからの出力と、前記第1マイクロフォンアレイの姿勢とに基づき、ビームフォーミング法にしたがって前記音源の位置または方位を測定する第1処理部と、固定して配置されている第2マイクロフォンアレイからの出力に基づき、前記音源の位置を測定する第2処理部と、前記音源の速度を測定し、当該測定結果に基づいてパーティクルフィルタの遷移モデルを切り替えながら、前記パーティクルフィルタを用いて前記第1および第2処理部のそれぞれの測定結果を統合することにより前記音源の位置を測定する第3処理部と、を備えていることを特徴とする。

10

【0005】

本発明の音源追跡システムによれば、第1マイクロフォンアレイの出力に基づく音源位置または方位測定結果と、第2マイクロフォンアレイの出力に基づく音源位置測定結果とがパーティクルフィルタを用いて統合または空間統合される。このため、異なるマイクロフォンアレイの相互の欠点や曖昧性を解消することができる。したがって、精度向上およびロバスト性に鑑みて音源定位のパフォーマンスを向上させることができる。さらに、音源速度、ひいては音源速度に鑑みて推察される当該音源の挙動態様に応じて遷移モデルが使い分けられうる。これにより、複数の音源が存在する空間において定位追跡対象である音源の異同が曖昧になる事態が回避されうる。たとえば、第1および第2音源のそれぞれの速度に鑑みて推察される各音源の挙動態様が考慮されることで、追跡対象が第1音源であるか第2音源であるかが途中で曖昧または不明瞭になる事態が回避される。そして、このような曖昧さを解消することにより、各音源の定位追跡の精度向上を図ることができる。

20

【0006】

前記第2処理部が前記第2マイクロフォンアレイを構成する複数のマイクロフォンのそれぞれと前記音源との距離を推定し、当該複数のマイクロフォンのうち当該算定距離が閾値以下の該マイクロフォンからの出力のみに基づいて前記音源の位置を測定することが好ましい。

30

【0007】

当該構成の音源追跡システムによれば、第2マイクロフォンアレイを構成する複数のマイクロフォンのそれぞれと音源との距離に基づいて選択された、当該第2マイクロフォンアレイの部分集合のみが利用されるサブアレイ法が採用されている。これにより、第2処理部による計算量の低減および計算時間の短縮が図られうる。

【0010】

前記第3処理部が、前記第1および第2処理部のそれぞれによる測定結果の尤度を算定し、当該尤度の重み付き和に基づいて前記音源の位置を測定することが好ましい。

40

【0011】

当該構成の音源追跡システムによれば、異なるマイクロフォンアレイによる異なる音源の定位結果、すなわち、第1処理部による音源の位置または方位測定結果および第2処理部による音源の位置測定結果が、各測定結果の尤度の重み付き和によって統合されうる。したがって、当該測定結果が直接的に統合される場合と比較して、当該測定結果が簡易に統合されうる。

【0012】

本発明の音源追跡方法は、音源の位置をリアルタイムで測定する音源追跡方法であって、移動機能を有する移動体に搭載されている第1マイクロフォンアレイからの出力と、前

50

記第 1 マイクロフォンアレイの姿勢とに基づき、ビームフォーミング法にしたがって前記音源の位置または方位を測定する第 1 処理と、固定して配置されている第 2 マイクロフォンアレイからの出力に基づき、前記音源の位置を測定する第 2 処理と、前記音源の速度を測定し、当該測定結果に基づいてパーティクルフィルタの遷移モデルを切り替えながら、前記パーティクルフィルタを用いて前記第 1 および第 2 処理のそれぞれの測定結果を統合することにより前記音源の位置を測定する第 3 処理と、を実行することを特徴とする。

【 0 0 1 3 】

本発明の音源追跡方法によれば、精度向上およびロバスト性に鑑みて音源定位のパフォーマンスを向上させることができる。

【 0 0 1 4 】

本発明のロボットは、音源の位置をリアルタイムで測定する音源追跡システムにおいて利用されるロボットであって、前記音源追跡システムが、第 1 マイクロフォンアレイからの出力と、前記第 1 マイクロフォンアレイの姿勢とに基づき、ビームフォーミング法にしたがって前記音源の位置または方位を測定する第 1 処理部と、固定して配置されている第 2 マイクロフォンアレイからの出力に基づき、前記音源の位置を測定する第 2 処理部と、前記音源の速度を測定し、当該測定結果に基づいてパーティクルフィルタの遷移モデルを切り替えながら、前記パーティクルフィルタを用いて前記第 1 および第 2 処理部のそれぞれの測定結果を統合することにより前記音源の位置を測定する第 3 処理部と、を備え、前記第 1 マイクロフォンアレイが搭載され、前記第 3 処理部により測定された前記音源の位置に基づいて前記ロボットの挙動を制御する制御システムを備えていることを特徴とする。

【 0 0 1 5 】

本発明のロボットによれば、精度向上およびロバスト性に鑑みて音源定位のパフォーマンスを向上させることができるので、当該音源定位結果に鑑みて適当に行動が制御される。

【 図面の簡単な説明 】

【 0 0 1 6 】

【 図 1 】 音源追跡システムの構成説明図

【 図 2 】 ロボットおよび R E M A の構成説明図

【 図 3 】 I R M A の構成説明図

【 図 4 】 音源追跡システムの機能説明図

【 図 5 a 】 音源追跡の曖昧性に関する説明図その 1

【 図 5 b 】 音源追跡の曖昧性に関する説明図その 2

【 図 5 c 】 音源追跡の曖昧性に関する説明図その 3

【 図 6 】 音源追跡の実験状況説明図その 1

【 図 7 a 】 音源の比較追跡結果の説明図その 1

【 図 7 b 】 本願発明の手法にしたがった音源の比較追跡結果の説明図その 1

【 図 8 】 音源追跡の実験状況説明図その 2

【 図 9 a 】 音源の比較追跡結果の説明図その 2

【 図 9 b 】 本願発明の手法にしたがった音源の比較追跡結果の説明図その 2

【 図 1 0 a 】 音源追跡結果に基づくロボット行動制御の説明図

【 図 1 0 b 】 音源追跡結果に基づくロボット行動制御の説明図

【 発明を実施するための最良の形態 】

【 0 0 1 7 】

本発明の音源追跡システム等の実施形態について図面を用いて説明する。

【 0 0 1 8 】

図 1 に示されている音源追跡システム 1 は、ロボット 1 0 に搭載されている R E M A (ロボット搭載型マイクロフォンアレイ (Robot-Embedded Microphone Array) の略。第 1 マイクロフォンアレイに該当する。) 1 1 と、ロボット 1 0 が存在する環境に固定して配置された I R M A (室内設置型マイクロフォンアレイ (In-Room Microphone Array) の略

10

20

30

40

50

。第2マイクロフォンアレイに該当する。)12とを用いて、音源(声を発する人間)20の位置を測定するシステムである。音源追跡システム1は、REMAシステム(第1処理部に該当する。)110と、IRMAシステム(第2処理部に該当する。)120と、MAI(マイクロフォンアレイインテグレータの略。第3処理部に該当する。)130とを備えている。

【0019】

図2に示されているようにロボット10は基体(胴体部)P0と、基体P0の上部に設けられた頭部P1と、基体P0の上部左右両側から延設された左右の腕部P2と、腕部P2の先端部に設けられた手部P4と、基体P0の下部から下方に延設された左右の脚部P3と、脚部P3の先端部に取り付けられている足部P5とを備えている。ロボット1は、再表03/090978号公報や、再表03/090979号公報に開示されているように、制御システム100によりアクチュエータ(またはモータ)の動作が制御され、当該アクチュエータから伝達される力によって、人間の肩関節、肘関節、手根関節、股関節、膝関節、足関節等の複数の関節に相当する複数の関節部分において腕部P2や脚部P3を屈伸運動させることができる。ロボット1は、左右の脚部P3(または足部P5)のそれぞれの離床および着床の繰り返しを伴う動きによって自律的に移動することができる。

【0020】

REMA11は頭部P1に搭載されている。REMA11は基本水平面に平行な面において環状に配置された8つの無指向性のマイクロフォン112により構成されている。ロボット1は頭部P1を基体P0に対してヨー軸回りに回転させることができ、かつ、頭部P1を基体P0に対して傾けることができる。なお、REMA11はロボット10の基体P0等、頭部P1とは異なる部位に搭載されてもよい。

【0021】

頭部P1には、左右に並んでロボット1の前方に向けられた一对の頭カメラ(CCDカメラ)C1が搭載されている。基体P0には前側下部に腰カメラ(赤外線カメラ)C2が搭載されている。ロボット1には基体P0の加速度に応じた信号を出力するジャイロセンサ、各関節の関節角度に応じた信号を出力するロータリエンコーダ、脚部P3に作用する床反力に応じた信号を出力する6軸力センサ等、種々のセンサが搭載されている。

【0022】

IRMA12はたとえば図3に示されているように部屋全体に配置され、壁面等に固定された複数のマイクロフォン122により構成されている。IRMA用のビームフォーマを設計するため、0.25[m]のメッシュを用いて室内が離散化された。離散化領域(X-Y平面)は $X = 1.0 \sim 5.0$ [m], $Y = 0.5 \sim 3.5$ [m], $Z = 1.2$ [m]の範囲で定義される。当該離散化により221個の音源定位用の評価点が得られた。また、反響が一様でない環境や任意のマイクレイアウトに対応するため、すべての音源定位用評価点でスピーカ(音源)20をロボット10に向けた状態でインパルス応答が計測され、当該計測結果に基づいて得られる伝達関数に基づき、ビームフォーマ「M-BF」が設計された。また、M-BFに対してサブアレイ法を適用したビームフォーマ「MS-BF」が設計された。なお、M-BFやMS-BFに加えてまたは代わりに、Sim-BFやRSim-BFがIRMA用のビームフォーマとして採用されてもよい。「Sim-BF」は室内が単に自由空間であると仮定して設計されるビームフォーマである。「RSim-BF」は壁の反射音パワーによる影響が低減されるように、シミュレーション計算に基づいて設計されるビームフォーマである。

【0023】

REMAシステム110はREMA11からの出力と、REMA11の姿勢とに基づき、ビームフォーミング法にしたがって音源20の位置を測定する。IRMAシステム12はIRMA12からの出力に基づき、音源20の位置を測定する。MAI130はREMAシステム110およびIRMAシステム120のそれぞれの測定結果を、パーティクルフィルタを用いて統合することにより音源20の位置を測定し、当該測定結果をLCD132に出力させる。

10

20

30

40

50

【 0 0 2 4 】

前記構成の音源追跡システム 1 0 0 の機能について説明する。

【 0 0 2 5 】

R E M A システム 1 1 0 により第 1 処理が実行される (図 4 / S 1 1 0) 。 具体的には R E M A 1 1 からの出力に基づき、適応ビームフォーマの一種である M U S I C (Multiple Signal Classification) にしたがって実時間でロボット座標系 (または R E M A 座標系) における音源 2 0 の方位が測定される。なお、第 1 処理に際して音源 2 0 の位置が測定されてもよい。

【 0 0 2 6 】

ロボット座標系における音源 2 0 の方位を固定座標系 (図 3 に示されている X - Y 平面参照) における方位に変換するため、固定座標系における頭部 P 1 の向きが R E M A 1 1 の姿勢として測定される。固定座標系における頭部 P 1 の向きは、ロボット 1 0 の各関節角度ひいては姿勢を表すロータリエンコーダの出力に基づいて測定される。固定座標系における頭部 P 1 の姿勢測定と、R E M A 1 1 による音響信号処理とを高精度で同期させるため、エンコーダからの出力の遅れ時間が考慮されている。当該遅れ時間は、たとえば、基体 P 0 に対する頭部 P 1 のヨー軸回りの回転角度等、ロボット 1 0 における共通の角度を測定するために設けられたエンコーダおよびポテンシオメータのそれぞれからの出力の位相差に基づいて測定される。

【 0 0 2 7 】

なお、ロボット 1 0 の制御システム 1 0 0 との通信により得られる、頭カメラ C 1 または腰カメラ C 2 により撮像された部屋の壁面や床面に固定されているマークの位置に基づいて固定座標系における R E M A 1 1 の姿勢が測定されてもよい。また、部屋に設置されたカメラを通じて得られたロボット 1 0 の画像、またはモーキャプシステムによるロボット 1 0 の動作解析に基づいて固定座標系における R E M A 1 1 の姿勢が測定されてもよい。測定方法が異なっても、姿勢測定の遅れ時間等の因子が考慮されることにより、固定座標系における頭部 P 1 の姿勢測定と、R E M A 1 1 による音響信号処理とを高精度で同期させることができる。

【 0 0 2 8 】

当該測定結果に基づき、式 (1) で表現される、R E M A システム 1 1 0 による時刻 t における観測ベクトル $U_{REMA}(t)$ が得られる。

【 0 0 2 9 】

$$U_{REMA}(t) = \{ u_{a1}(t), \dots, u_{an}(t), \dots, u_{an1t}(t) \} \quad (1)$$

ここで N_{1t} は時刻 t における R E M A 1 1 の観測数である。 $u_{an}(t)$ は n 番目の R E M A マイクロフォン 1 1 2 から見た音源 2 0 の方向を表す固定座標系における水平角度 (方位) $a_{n1}(t)$ と、音源 2 0 から発せられた音の推定パワー $a_{pn}(t)$ とを成分とする、式 (2) で表現されるベクトルである。

【 0 0 3 0 】

$$u_{an}(t) = (a_{n1}(t), a_{pn}(t)) \quad (2)$$

M U S I C によれば、雑音源の方向に死角を形成した上で目的音源方向を推定することができ、かつ、環境の変化に応じて当該死角の方向を適宜調節することができるため、ロボットの柔軟性を向上させることができる (文献 6 参照) 。なお、M U S I C の実装に際して、ヒューマノイドロボットを対象に開発された実装が利用された (文献 7 参照) 。

【 0 0 3 1 】

文献 6 : F. Asano, et al. "Real-time sound source localization and separation system and its application to automatic speech recognition." In ISCA, editor, Proc. Of European Conference on Speech Processing (Eurospeech 2001), pp.1013-1016, 2001

文献 7 : I. Hara, et al. "Robust speech interface based on audio and video information fusion for humanoid HRP-2." In Proc. of IEEE/RAS International Conference on Intelligent and Systems (IROS-2004), pp.2404-2410. IEEE, 2004.

10

20

30

40

50

IRMAシステム120により第2処理が実行される(図4/S120)。具体的にはIRMA12からの出力に基づき、重み付き遅延和法(WDS-BF)にしたがって音源20の位置が測定される(文献8参照)。

文献8:K. Nakadai, et al. "Sound source tracking with directivity pattern estimation using a 64ch microphone array." In Proc. of the IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS 2005), pp.196-202, 2005.

一般に典型的なビームフォーミングでは、システム出力 $Y_p(\omega)$ は、位置 p にある音源20のスペクトル $X(\omega)$ と、音源20から n 番目のIRMAマイクロフォン122への伝達関数 $H_{p,n}(\omega)$ と、 n 番目のIRMAマイクロフォン122により集音された信号のスペクトル $X_n(\omega)$ と、 n 番目のマイクロフォン122への $G_{n,p}(\omega)$ とに基づき、式(3)により表現される。

【0032】

$$Y_p(\omega) = \sum_n G_{n,p}(\omega) X_n(\omega) \\ = \sum_n G_{n,p}(\omega) H_{p,n}(\omega) X(\omega) \quad (3)$$

WDS-BFによれば、測定結果や計算的に導出されたものなどさまざまなタイプの伝達関数 $H_{p,n}(\omega)$ を統一的に取り扱えるように一般化されている。また、伝達関数 $H_{p,n}(\omega)$ の動的変化や、入力信号 $X_n(\omega)$ のひずみなどに対するロバスト性を持たせるようにフィルタ関数 $G_{n,p}(\omega)$ のノルムが最小化されている(文献9参照)。

文献9:中島弘史「不定項を利用した平均サイドローブエネルギー最小ビームフォーミングの実現」、日本音響学会誌 Vol.62, No.10, pp.726-737, 2006.

また、 n 番目のマイクロフォン122から音源20までの距離 r_n に応じて伝達関数 $H_{p,n}(\omega)$ の値が設定される。具体的には、式(4)に示されているように伝達関数 $H_{p,n}(\omega)$ は当該距離 r_n が閾値 r_{th} 以下である場合にはそのままに維持される(n 番目のマイクロフォン122の出力が採用される)一方、当該距離 r_n が閾値 r_{th} を超えている場合には0に設定される。

【0033】

$$H_{p,n}(\omega) = H_{p,n}(\omega) \quad (r_n \leq r_{th} \text{の場合}), \text{ または} \\ = 0 \quad (r_n > r_{th} \text{の場合}) \quad (4)$$

これにより、式(5)で表現される、IRMAシステム120による時刻 t における観測ベクトル $U_{IRMA}(t)$ が得られる。

【0034】

$$U_{IRMA}(t) = \{ u_{b1}(t), \dots, u_{bm}(t), \dots, u_{bN2t}(t) \} \quad (5)$$

ここで $N2t$ は時刻 t におけるIRMA12の観測数である。 $u_{bm}(t)$ は固定座標系における音源20の位置($b_{xm}(t)$, $b_{ym}(t)$)と、固定座標系における音源20の進行方向 $b_{om}(t)$ と、音源20から発せられた音の推定パワー $b_{pm}(t)$ とを成分とする、式(6)で表現されるベクトルである。

【0035】

$$u_{bm}(t) = (b_{xm}(t), b_{ym}(t), b_{om}(t), b_{pm}(t)) \quad (6)$$

また、WDS-BFは位置 p が極座標(r, θ)により表現されることにより、指向特性推定に適用されうる。指向特性推定が応用されることにより音源20が向いている方向が推定され、あるいは実際には人間が話しているのか、それともスピーカから出力された音声なのかが判別されうる。

【0036】

そして、MAI130により第3処理が実行される(図4/S130)。具体的には、第1処理における測定結果と、第2処理における測定結果とが、パーティクルフィルタを用いて統合されることにより音源20の位置が測定される。

【0037】

パーティクルフィルタ(文献10参照)はSLAM(Simultaneous Localization And Mapping)や物体追跡などを効率的に解くために用いられる手法である(文献11参照)。基本的にはパーティクルを用いて状態がサンプリングされ、得られたパーティクルが遷

10

20

30

40

50

移モデルおよび観測モデルを用いて更新されていくことにより、観測結果から内部状態が推定される。パーティクルフィルタには、線形遷移のみを扱うKalmanフィルタなどと異なり、非線形の動きを追跡する枠組みを有していること、ガウシアン以外の分布を容易に扱えること、パーティクル数に基づき動作速度が制御可能であること等の特長がある。また、パーティクルフィルタは遷移モデルおよび観測モデルが用意されれば、データの種類にかかわらず利用される。

文献10：M.S.Arulampalam, et al. "A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking." IEEE Transactions on Signal Processing, Vol.50, No.2, pp.174-188, 2002.

文献11：S. Thrun, et al. "Probabilistic Robotics." The MIT Press, 2005. 10

しかしながら、異なる2つのマイクロフォンアレイ、すなわちREMA11およびIRMA12のそれぞれの測定結果を統合するためにパーティクルフィルタが用いられる場合には問題が生じる。すなわち、ロボット座標系および絶対座標系の併存、ならびに、極座標系およびカルテシアン座標系の併存に由来する2つの問題である。

【0038】

第1の問題は音響処理とロボット10の動作との高度な時間同期が必要となることである。これは、REMAシステム110による測定結果およびIRMAシステム120による測定結果が統合されるためには、REMAシステム110による測定結果が、ロボット座標系から固定座標系に変換される必要があるためである。

【0039】

第2の問題は測定結果の次元が異なることである。すなわち、REMAシステム110による測定結果が1次元（方位または水平角度）であるのに対して、IRMAシステム120による測定結果が2次元である。 20

【0040】

そこで、第1の問題を解決するため、動作と音響信号取得とを高精度で同期させることができるハードウェアが採用されている。また、第2の問題を解決するため、各座標系について尤度関数、尤度によって測定結果を統合する手法が採用されている。より具体的な手法について以下説明する。

【0041】

パーティクルフィルタにおいては、遷移モデル $p(s(t) | s(t-1))$ と観測モデル $p(u(t) | s(t))$ とが確率的な表現として定義される。遷移モデル $p(s(t) | s(t-1))$ は、時刻 $t-1$ における状態 $s(t-1)$ から時刻 t における状態 $s(t)$ を推定するためのモデルである。観測モデル $p(u(t) | s(t))$ は状態 $s(t)$ に鑑みて観測ベクトル $u(t)$ の確からしさを評価するためのモデルである。 30

【0042】

また、一般的なパーティクルフィルタに次のように改良されたパーティクルフィルタが用いられている。すなわち、複数の音源に対応できるように複数のパーティクルグループの併存が許容され、観測状況に応じて当該グループの数を動的に変化させる機構が実装されている（改良点その1）。また、移動音源はある程度の速度であれば急激に進行方向を変えないという仮定のもと、ランダムウォークと運動方程式とをオン減速度に応じて使い分ける非線形名遷移モデルが採用されている（改良点その2）。さらに、REMA11およびIRMA12のそれぞれから得られる次元数の異なる測定位置情報を透過的に統合するため、尤度レベルで当該測定位置情報を統合する機構が実装されている（改良点その3）。 40

【0043】

改良パーティクルフィルタを用いた処理について次に説明する。

【0044】

まず、パーティクルが初期化される（図4/S131）。 i 番目のパーティクルの内部状態 $s_i(t)$ は音源位置 $(x_i(t), y_i(t))$ 、音源速度 $v_i(t)$ および音源進行方向 $\theta_i(t)$ により式(7)で表現される。 50

【 0 0 4 5 】

$$s_i(t) = (x_i(t), y_i(t), v_i(t), o_i(t)) \quad (7)$$

初期化に際して、すべてのパーティクルが一様かつランダムに分布される。また、複数の音源を扱うため、総数 N のパーティクルが N_k 個のパーティクルを包含する $S (> 1)$ 個のパーティクルグループ P_k に分割される。さらに、パーティクルグループ P_k ごとに、音源追跡への寄与に鑑みた i 番目のパーティクルの重要度 w_i が式 (8) で表される規格化条件を満たすように導入されている。

【 0 0 4 6 】

$$\sum_{i \in P_k} w_i = 1 \quad (8)$$

また、複数の音源を扱うため、音源生成・消滅チェックが実行される (図 4 / S 1 3 2)。 10

【 0 0 4 7 】

時刻 t におけるパーティクルグループ P_k の内部状態 $S_k(t)$ が式 (9) により定義される。

【 0 0 4 8 】

$$S_k(t) = \sum_{i \in P_k} s_i(t) \cdot w_i(t) \quad (9)$$

さらに、REMA システム 1 1 0 による j 番目の観測結果 $u_{aj}(t)$ が不等式 (10 a) で表されているようにパーティクルグループ P_k の内部状態 $S_k(t)$ とのユークリッド距離が閾値 D_{th} 未満であるという条件を満たすことを要件として、当該観測結果 $u_{aj}(t)$ がパーティクルグループ P_k にアソシエートされる。同様に IRMA システム 1 2 0 による j 番目の観測結果 $u_{bj}(t)$ が不等式 (10 b) で表されているようにパーティクルグループ P_k の内部状態 $S_k(t)$ とのユークリッド距離が閾値 D_{th} 未満であるという条件を満たすことを要件として、当該観測結果 $u_{bj}(t)$ がパーティクルグループ P_k にアソシエートされる。 20

【 0 0 4 9 】

$$|S_k(t) - u_{aj}(t)| < D_{th} \quad (10 a)$$

$$|S_k(t) - u_{bj}(t)| < D_{th} \quad (10 b)$$

一方、観測結果に対応するパーティクルグループ P_k がない場合、新たなパーティクルグループが生成される。また、パーティクルグループ P_k にアソシエートする観測結果が所定時間にわたり得られなかった場合、当該パーティクルグループ P_k は消滅する。いずれの場合も式 (8) で示されている規格化条件を満たされるように重要度 w_i が設定される。 30

【 0 0 5 0 】

続いて、時刻 $t - 1$ における i 番目のパーティクルの状態 $s_i(t - 1)$ に基づき、遷移モデル ($s(t) | s(t - 1)$) にしたがって、時刻 t における当該 i 番目のパーティクルの状態 $s_i(t)$ が推定される (図 4 / S 1 3 3)。

【 0 0 5 1 】

遷移モデルは、前記のようにランダムウォークと運動方程式とを音源 2 0 の速度 v に応じて使い分ける非線形モデルである。具体的には、音源速度 v が閾値 v_{th} (たとえば $2 [m/s]$) 以下の場合、ランダムウォークが採用され、遷移モデル (第 1 遷移モデル) は式 (11) により表現される。 40

【 0 0 5 2 】

$$s_i(t) = s_i(t - 1) + r,$$

$$r = (r_x, r_y, r_v, r_o) \quad (11)$$

ここで r はガウス雑音を表している。

【 0 0 5 3 】

一方、音源の速度 v が閾値 v_{th} を超えている場合、遷移モデル (第 2 遷移モデル) は運動方程式 (12 a) ~ (12 d) により表現される。

【 0 0 5 4 】

$$x_i(t) = x_i(t - 1)$$

$$+ v_i(t-1) \cdot \cos(o_i(t-1)) + r_x \quad (12a)$$

$$y_i(t) = y_i(t-1)$$

$$+ v_i(t-1) \cdot \sin(o_i(t-1)) + r_y \quad (12b)$$

$$v_i(t) = \alpha \cdot v_i(t-1)$$

$$+ (1 - \alpha) \cdot \{ (x_i(t) - x_i(t-1))^2 + (y_i(t) - y_i(t-1))^2 \}^{1/2} + r_v,$$

$$x_i(t) = x_i(t) - x_i(t-1),$$

$$y_i(t) = y_i(t) - y_i(t-1) \quad (12c)$$

$$o_i(t) = \alpha \cdot o_i(t-1)$$

$$+ (1 - \alpha) \cdot \tan^{-1} \{ x_i(t) / y_i(t) \} + r_o. \quad (12d)$$

i 番目のパーティクルの重要度 $w_i(t)$ は、式(8)に示されている規格化条件下で、統合尤度 $L_i(t)$ に基づき、式(13)にしたがって更新される。

10

【0055】

$$w_i(t) = L_i(t) \cdot w_i(t-1) \quad (13)$$

統合尤度 $L_i(t)$ は REMA 尤度 L_{REMA} と、IRMA 尤度 L_{IRMA} と、適宜設定される尤度統合用重み L とに基づき、式(14)により表現される。

【0056】

$$L_i(t) = L \cdot L_{REMA}(t) + (1 - L) \cdot L_{IRMA}(t) \quad (14)$$

REMA 尤度 L_{REMA} は i 番目のパーティクルフィルタの状態(ベクトル) $s_i(t)$ が x 軸となす角度 $\theta_x(s_i(t))$ と、ロボットの位置 $P_{REMA}(t)$ と、REMA 11 による音源位置測定結果の分散 R_{REMA} とに基づき、式(15)により定義される。

20

【0057】

$$L_{REMA}(t) = \exp \left[- \frac{(\theta_x(s_i(t)) - P_{REMA}(t) - \theta_L)^2}{2 R_{REMA}} \right] \quad (15)$$

IRMA 尤度 L_{IRMA} は i 番目のパーティクルの状態 $s_i(t)$ と IRMA 12 により得られた観測ベクトル $u_{bm}(t)$ とのユークリッド距離 $|s_i(t) - u_{bm}(t)|$ と、IRMA 12 による音源位置測定結果の分散 R_{IRMA} とに基づき、式(16)により定義される。

【0058】

$$L_{IRMA}(t) = \exp \left[- |s_i(t) - u_{bm}(t)|^2 / 2 R_{IRMA} \right] \quad (16)$$

30

さらに、重要度 w_i に基づきパーティクルが更新される(図4/S134)。

【0059】

具体的には、パーティクルグループ P_k に属する i 番目のパーティクルの数が式(17)にしたがって更新される。

【0060】

$$N_{ki} = \text{round}(N_k \cdot w_i) \quad (17)$$

また、残りの未更新の $R_k (= N_k - \sum_{i \in P_k} N_{ki})$ 個のパーティクルも、式(18)で表現される残差重みパラメータ R_{wi} にしたがって分配される。この際、一般的な SIR (Sampling Importance Resampling) アルゴリズム(文献2参照)が用いられる。

【0061】

$$R_{wi} = w_i - N_{ki} / \sum_{j \in P_k} N_{kj} \quad (18)$$

40

さらに、更新後のパーティクル密度から事後確率 $p(s(t) | u_j(t))$ が推定される(図4/S135)。音源 k に対する k 番目のパーティクルグループ P_k の内部状態 $S_k(t)$ は式(9)にしたがって推定される。

【0062】

そして、推定された事後確率 $p(s(t) | u_j(t))$ に基づき、固定座標系における音源 20 の位置が推定される(図4/S136)。音源 20 の位置推定結果または追跡結果が Java (登録商標) 3D により実装されているサウンドビューワに送られ、LCD 132 において当該追跡結果が逐次 3 次元的に表示される。

【0063】

50

音源追跡が完了したか否かが判定される（図4 / S 1 4 0）。当該判定結果が否定的である場合（図4 / S 1 4 0 NO）、前記の処理が繰り返される。一方、当該判定結果が肯定的である場合（図4 / S 1 4 0 YES）、一連の処理が終了する。

【0064】

前記機能を発揮する音源追跡システム1によれば、REMA（第1マイクロフォンアレイ）11の出力に基づく音源方位測定結果と、IRMA（第2マイクロフォンアレイ）12の出力に基づく音源位置測定結果とがパーティクルフィルタを用いて統合または空間統合される（図4 / S 1 3 0 参照）。このため、異なるマイクロフォンアレイ、すなわち、REMA 11およびIRMA 12の相互の欠点や曖昧性を解消することができる。したがって、精度向上およびロバスト性に鑑みて音源定位のパフォーマンスを向上させることができる。

10

【0065】

また、IRMA（第2マイクロフォンアレイ）12を構成する複数のマイクロフォン122のそれぞれと音源20との距離 r_n に基づいて選択された、当該IRMA 12の部分集合のみが利用されるサブアレイ法が採用されている（図4 / S 1 2 0 参照）。具体的には、音源20との距離 r_n が閾値 r_{th} を超えているIRMAマイクロフォン122の伝達関数 $H_{p,n}(\)$ が0に設定されることにより、IRMA 12の部分集合が選択される（式（4）参照）。これにより、IRMAシステム（第2処理部）120による計算量の低減および計算時間の短縮が図られうる。

【0066】

20

さらに、音源速度 v 、ひいては音源速度 v に鑑みて推察される当該音源20の挙動態様に応じて遷移モデルが使い分けられうる（図4 / S 1 3 3 参照）。たとえば、図5（a）に示されているように2つの音源20が矢印で示すように徐々に近接し、その後、徐々に離反するように挙動した状況を考える。この場合、図5（b）に示されているように第1音源21および第2音源22のそれぞれが徐々に近づくように移動し、相互に接近した後で進行方向を変更したのか、図5（c）に示されているように第1音源21および第2音源22のそれぞれが徐々に近づくように移動し、相互に接近した後も進行方向を変えずに移動したのか曖昧になる可能性がある。そこで、それぞれの速度 v に鑑みて推察される第1音源21および第2音源22のそれぞれの挙動態様が考慮されて遷移モデルが使い分けられる。

30

【0067】

すなわち、音源速度 v が閾値 v_{th} 以下の場合、音源20の挙動態様としてランダムウォークが推察され、第1遷移モデルが採用される（式（11）参照）。第1遷移モデルは、第1音源21および第2音源22のそれぞれが相互に接近した後で進行方向を変更する場合、速度 v が低下して閾値 v_{th} 以下になるという仮定に基づいて採用される遷移モデルである。これにより、第1音源21および第2音源22が相互に接近したときの各音源21, 22の速度 v が閾値 v_{th} 以下の場合、各音源21, 22が図5（b）に示されているように挙動したことが推察された上で定位追跡される。

【0068】

一方、音源速度 v が閾値 v_{th} を超えている場合、第2遷移モデルが採用される（運動方程式（12a）～（12d）参照）。第2遷移モデルは、第1音源21および第2音源22のそれぞれが接近する前後で進行方向を維持する場合、速度 v がさほど低下せず、閾値 v_{th} を超えるという仮定に基づいて採用される遷移モデルである。これにより、第1音源21および第2音源22のそれぞれが接近したときの各音源21, 22の速度 v が閾値 v_{th} を超えている場合、各音源21, 22が図5（c）に示されているように挙動したことが推察された上で定位追跡される。

40

【0069】

これにより、追跡対象が第1音源21であるか第2音源22であるかが途中で曖昧または不明瞭になる事態が回避され、各音源21, 22の定位追跡の精度向上を図ることができる。

50

【 0 0 7 0 】

また、異なるマイクロフォンアレイによる異なる音源の定位結果、すなわち、REMAシステム110による音源20の方位測定結果およびIRMAシステム120による音源20の位置測定結果が、各測定結果の尤度 L_{REMA} および L_{IRMA} の重み付き和によって統合されうる(図4/S134, 式(14)参照)。したがって、当該測定結果、すなわちREMAシステム110により得られた観測ベクトル $U_{REMA}(t)$ およびIRMAシステム120により得られた観測ベクトル $U_{IRMA}(t)$ が直接的に統合される場合と比較して、当該測定結果が簡易に統合されうる。

【 0 0 7 1 】

続いて、REMA11およびIRMA12が併用された手法による、音源の位置測定および追跡のパフォーマンスの評価結果について説明する。

10

【 0 0 7 2 】

図6に示されているようにロボット座標系 (x, y) が固定座標系 (X, Y) に対して 180° 傾いている状態で、人間(音源)20が常にロボット10を向いて発話しながら、矢印で示されているようにロボット10を中心として反時計回りに半円を描くように移動する状況が設定された。この状況で超音波3Dタグシステム(U3D-TS)(文献12参照)を用いた音源20の基準追跡結果が図7(a)および図7(b)のそれぞれに破線で示されている。

文献12: Y. Nishida, et al. "3D ultrasonic tagging system for observing human activity." In IEEE, editor, Proceedings of the 2003 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS 2003), pp.785-791, 2003.

20

また、同じ状況でIRMAシステム120から得られた測定結果のみを用い、パーティクルフィルタを適用したときの音源20の比較追跡結果が図7(a)に実線で示されている。さらに、同じ状況で本願発明の手法にしたがってREMAシステム110およびIRMAシステム120の両方の測定結果を用い、パーティクルフィルタを適用したときの音源20の追跡結果が図7(b)に実線で示されている。

【 0 0 7 3 】

図7(a)に示されている、基準追跡結果(破線)に対する比較追跡結果(実線)の誤差平均は $0.12[m]$ であり、誤差標準偏差は $0.062[m]$ であった。これに対して、図7(b)に示されている、基準追跡結果(破線)に対する本願発明の手法にしたがった追跡結果(実線)の誤差平均は $0.10[m]$ であり、誤差標準偏差は $0.040[m]$ であった。すなわち、本願発明の手法にしたがった追跡結果(図7(b)実線)のほうが、比較追跡結果(図7(a)実線)よりも、基準追跡結果(図7(a)(b)破線)に整合しており、追跡または測定精度が高いことが実験によって確認された。なお、頭部P1が人間20に向けられるように、REMA11の座標系またはロボット座標系を固定座標系に対して回転させた状況でも同様に、本願発明の手法による測定精度の高さが確認された。REMA11の座標系は、ロボット10の向きをその場で変えさせ、これに加えてまたは代えて頭部P1を基体P0に対してヨー軸回りに回転させることにより、固定座標系に対して回転させられる。

30

【 0 0 7 4 】

また、図8に示されているようにロボット座標系 (x, y) が固定座標系 (X, Y) に対して 180° 傾いている状態で、初期状態でロボット10を挟んで対向している第1の人間(音源)20Aおよび第2の人間(音源)20Bのそれぞれが常にロボット10を向いて発話しながら、矢印で示すようにロボット10を中心として反時計回りに半円を描くように移動する状況が設定された。この状況で超音波3Dタグシステム(U3D-TS)(文献15参照)を用いた音源20Aおよび20Bのそれぞれの基準追跡結果が図9(a)および図9(b)のそれぞれに破線で示されている。

40

【 0 0 7 5 】

また、同じ状況でIRMAシステム120から得られた測定結果のみを用い、パーティクルフィルタを適用したときの音源20Aおよび20Bのそれぞれの比較追跡結果が図9

50

(a) に一点鎖線および二点鎖線のそれぞれによって示されている。さらに、同じ状況で本願発明の手法にしたがって R E M A システム 1 1 0 および I R M A システム 1 2 0 の両方の測定結果を用い、パーティクルフィルタを適用したときの音源 2 0 A および 2 0 B のそれぞれの追跡結果が図 9 (b) に一点鎖線および二点鎖線のそれぞれによって示されている。

【 0 0 7 6 】

図 9 (a) に示されている、基準追跡結果 (破線) に対する比較追跡結果 (一点差線および二点鎖線) の誤差平均は 0 . 1 6 [m] であり、誤差標準偏差は 0 . 0 8 4 [m] であった。これに対して、図 9 (b) に示されている、基準追跡結果 (破線) に対する本願発明の手法にしたがった追跡結果 (実線) の誤差平均は 0 . 1 6 [m] であり、誤差標準偏差は 0 . 0 8 3 [m] であった。すなわち、本願発明の手法にしたがった追跡結果 (図 9 (b) 実線) のほうが、比較追跡結果 (図 9 (a) 実線) よりも、基準追跡結果 (図 9 (a) (b) 破線) に整合しており、追跡または測定精度が (若干ながらも) 高いことが実験によって確認された。なお、頭部 P 1 が第 1 の人間 2 0 A に向けられるように、 R E M A 1 1 の座標系またはロボット座標系を固定座標系に対して回転させた状況でも同様に、本願発明の手法による測定精度の高さが確認された。その他、さまざまな状況下においても音源追跡精度の向上が認められ、ロバスト性の向上が図られた。

【 0 0 7 7 】

なお、音源追跡結果が音源追跡システム 1 からロボット 1 0 の制御システム 1 0 0 に送信され、制御システム 1 0 0 が当該追跡結果に基づいてロボット 1 0 の行動を制御してもよい。たとえば図 1 0 (a) に示されているようにロボット 1 0 のそばで人間 (音源) 2 0 が発話している状況を考える。前記のようにロボット 1 0 には R E M A 1 1 が搭載され、その周囲には I R M A 1 2 が配置されている。ロボット 1 0 および人間 2 0 のそれぞれは固定座標系において移動していても静止していてもよい。この状況で音源追跡結果に基づき、たとえば図 1 0 (b) に示されているように人間 2 0 に頭部 P 1 の正面を向け、かつ、片方のハンド P 3 をその人間 2 0 に向けて上げるようにロボット 1 0 の行動が制御されてもよい。このように、音源 (人間) 2 0 のロバスト性のある精度の高い追跡結果に基づき、周囲の人間にロボット (ヒューマノイドロボット) 1 0 への親しみを覚えさせる等の観点からロボット 1 0 の行動が適当に制御されうる。また、複数の音源 (人間) 2 0 が存在する場合でもその挙動の曖昧さを解消して、ロバスト性を向上させることができる。

【 0 0 7 8 】

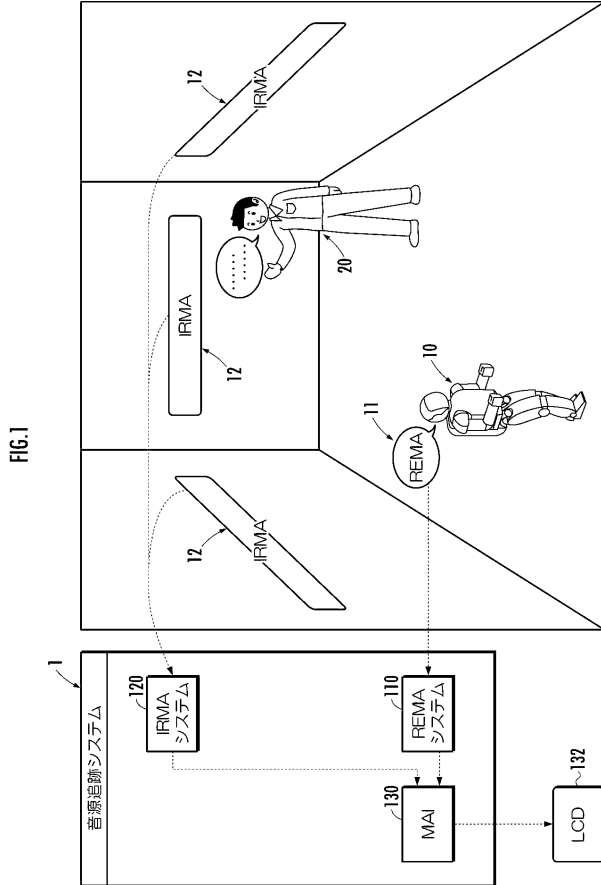
また、複数のロボット 1 0 のそれぞれに搭載された複数の R E M A 1 1 を用いて音源 2 0 が定位追跡されてもよく、配置やマイクロフォン特性などが異なる複数の I R M A 1 2 を用いて音源 2 0 が追跡されてもよく、複数の R E M A 1 1 および複数の I R M A 1 2 の組み合わせを用いて音源 2 0 が追跡されてもよい。

10

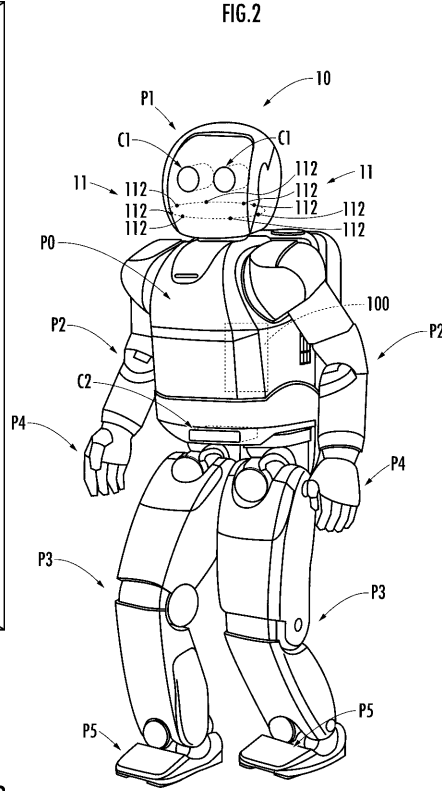
20

30

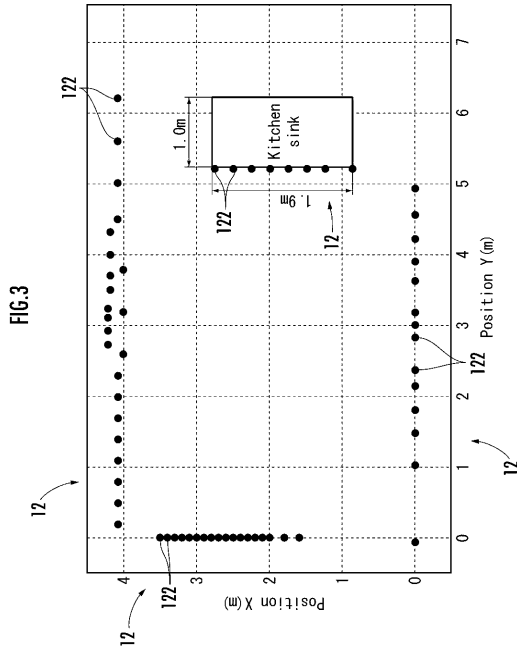
【 図 1 】



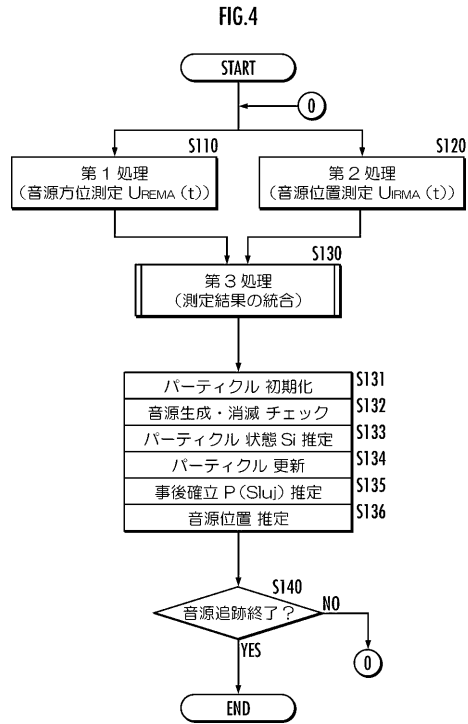
【 図 2 】

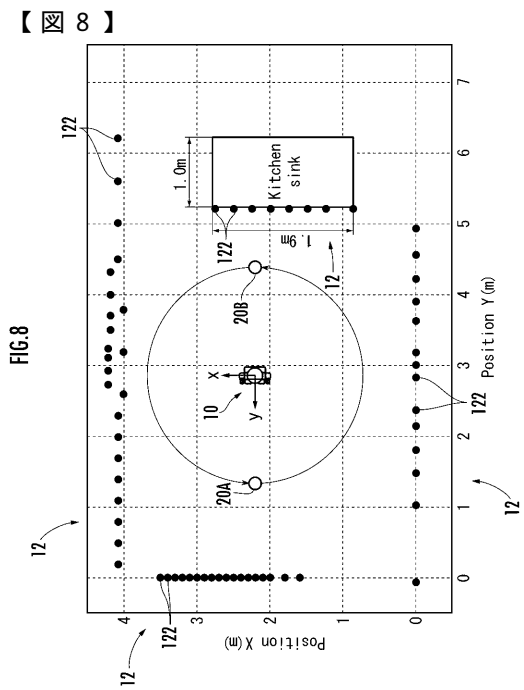
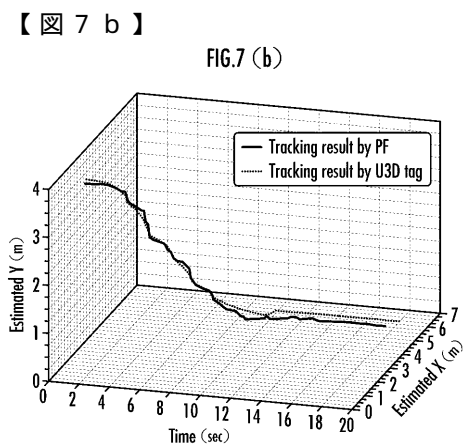
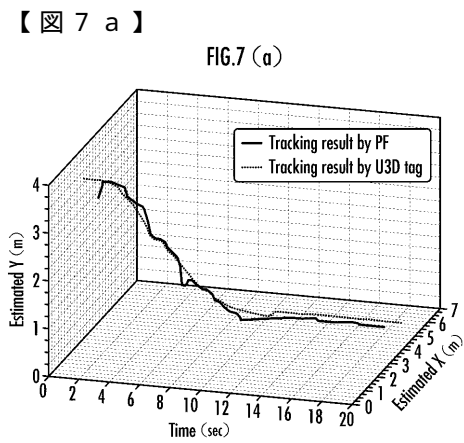
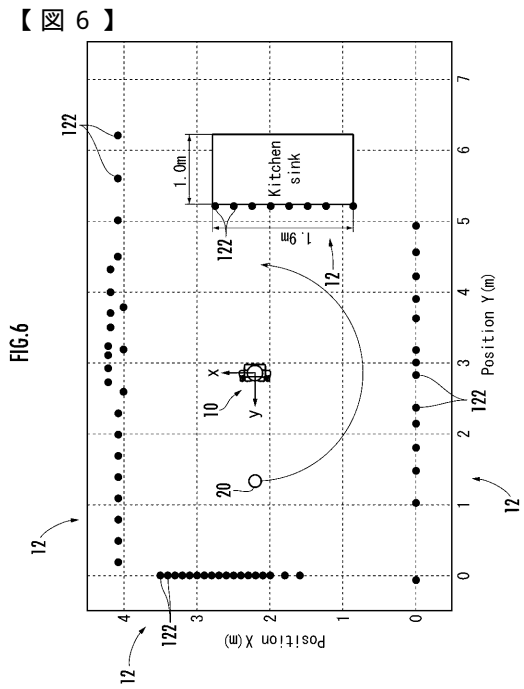
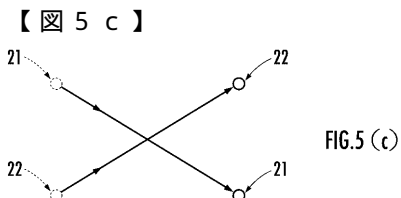
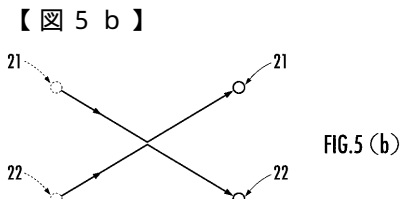
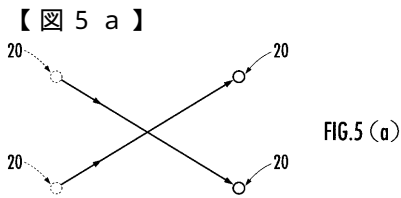


【 図 3 】

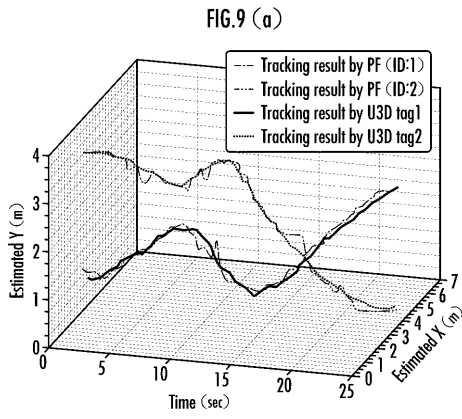


【 図 4 】

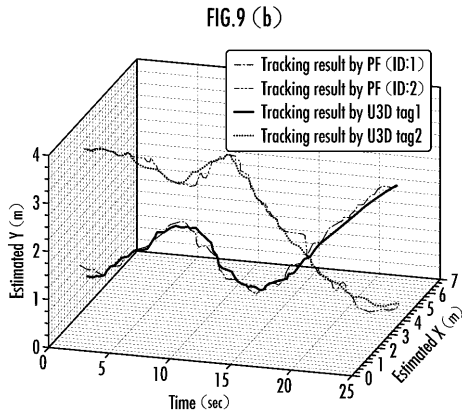




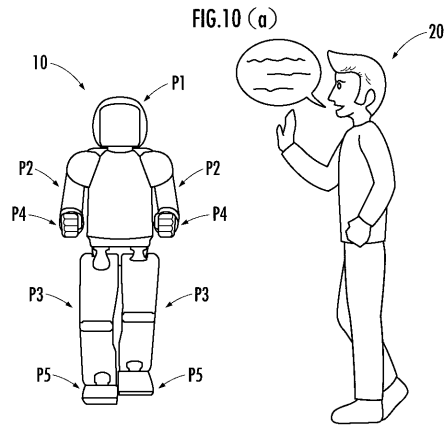
【 9 a 】



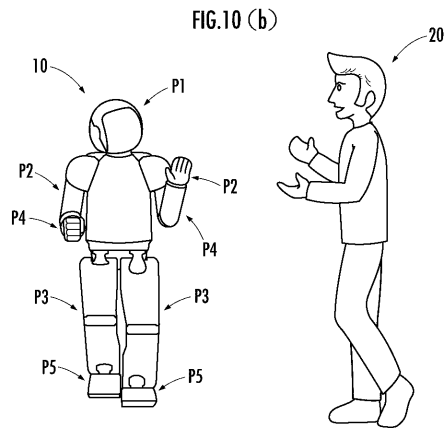
【 9 b 】



【 10 a 】



【 10 b 】



フロントページの続き

- (72)発明者 中臺 一博
埼玉県和光市本町8-1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 辻野 広司
埼玉県和光市本町8-1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 長谷川 雄二
埼玉県和光市本町8-1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 奥乃 博
京都府京都市中京区東洞院通三条下る三文字町205番地の3 フォルム東洞院三条1102号

審査官 菊地 陽一

- (56)参考文献 特開2002-366191(JP,A)
特開2004-198656(JP,A)
特開2003-251583(JP,A)
中臺 一博 Kazuhiro Nakadai, “大規模マイクロホンアレイによる室内移動音源の追跡と方向推定 Sound Source Tracking with Orientation Estimation by Using A Large Scale Microphone Array”, AIチャレンジ研究会(第22回) SIG-Challenge-0522
Proceedings of the 22nd Meeting of Special Interest Group on AI Challenges, 日本, 社団法人人工知能学会AIチャレンジ研究会 Special Interest Group on AI Challenges Japanese Society for Artificial Intelligence, 2005年12月6日, 77-82

(58)調査した分野(Int.Cl., DB名)

B25J 13/08
G01S 3/802
G01S 5/20
H04R 1/40
H04R 3/00