

ホース型ロボットのマイクロホンアレイを用いた姿勢推定

○坂東 宜昭¹ 大塚 琢馬¹ 水本 武志^{1*} 糸山 克寿¹ 中臺 一博² 奥乃 博¹

¹ 京都大学 大学院情報学研究科 ²(株) ホンダ・リサーチ・インスティテュート・ジャパン

1. はじめに

災害現場でのレスキューロボットによる、人では捜索が危険な場所や困難な場所の探索が期待されている。ホース型ロボットは、細長く、ロボット体表と環境との接地面での摩擦を利用して推進できるため、リモートオペレータによる狭い隙間の探索が可能である。リモートオペレータは、先端のカメラから情報を収集し、手元のホースの抜き差しと先端のアクチュエータを用いて目的の方向へロボットを進行させる。ホース型ロボットは盛んに開発されており、例えば、タイミングベルトと小型車輪を用いた Active-Hose [1] や、繊毛の振動を用いた Active Scope Camera (ASC) [2] などが報告されている。本稿では、レスキューロボットの種類として研究開発を行っているホース型ロボットの利便性 (usability) 向上を目的に、マイクロホンアレイを用いて本ロボットの姿勢を推定する技術を報告する。

Robin Murphy は、アメリカでの5つの災害における9つのレスキューロボットの適応事例から navigational usability と mission usability が不可欠であると指摘している [3]。前者は、ロボットの航行能力についての有用性を指し、推進機構や、操縦に必要な情報収集能力の欠如が問題点である。後者は各種センサデータとリモートオペレータへの情報提供についての有用性を指し、データの統合方法、センサシステムの欠陥、オペレータへの提示方法が問題点である。

ホース型ロボットの navigational usability と mission usability を向上させるためには以下の2つの機能 (図1) の実現が不可欠である。

1) 音による姿勢推定 ホース型ロボットが瓦礫の隙間などに進入すると、リモートオペレータはその姿勢を視認できない。狭い隙間ではホースがたわむことがあり、押すのか引くのかかわらず、ロボットの進入に支障をきたすことがある。また、従来のジャイロセンサや、GPS、オドメトリだけでは十分に姿勢推定できない。音による姿勢推定が navigational usability の向上が不可欠である。

2) 音源定位・分離 従来のホース型ロボットには被災者発見と位置推定のため、先端にビデオカメラと単一マイクロホンが搭載されている。瓦礫などの隙間は暗く遮蔽物が多いため、ビデオカメラのみで被災者を発見することは困難である。音は暗闇でも伝わり、また遮蔽物を回りこむため、被災者の音声などの音源定位・分離による被災者位置推定は mission usability の向上に寄与する。遠隔地の音源方向提示による聴覚アウェアネスの有効性は、HARK [4] を用いたテレプレゼンスロボットの開発 [5] でも指摘されている。

*現在は (株) ホンダ・リサーチ・インスティテュート・ジャパン

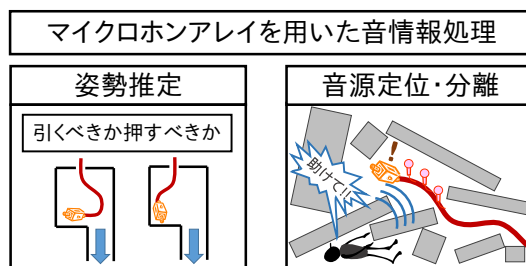


図1 ホース型ロボットの利便性 (usability) 向上には姿勢推定と音源定位・分離機能の実現が必要である。

これらの2機能を実現するために、複数の小型スピーカとマイクロホンアレイを装着したホース型ロボットを開発中である。小型スピーカから試験音を発し、マイクロホン間の到達時間差を手がかりに、ロボットの姿勢を推定する。また、小型スピーカを用いて被災者に呼びかけ、マイクロホンアレイを用いて被災者の声を収録し、音源定位・分離による被災者位置推定を行う。音源定位にはマイクロホンの位置が必要であり、ホース型ロボットではマイクロホンの位置関係が変化するため、音源定位のためにもロボットの姿勢推定が不可欠である。本稿ではホース型ロボットの姿勢推定を取り上げ、関連研究、提案手法、ホース型ロボットのモックアップを用いた評価実験について述べる

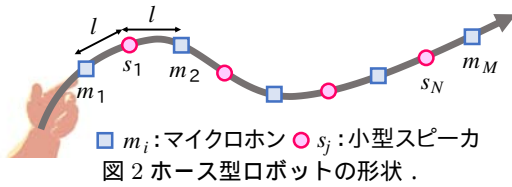
2. 関連研究

音の到達時間差を用いた姿勢推定法の開発のため、ホース型ロボットの姿勢推定と、音を用いたマイクロホン位置推定に関する従来法を概観する。

Hatazaki ら [6] は、ホース型ロボットの種類である ASC の姿勢推定に、加速度センサとジャイロセンサを使用し、これらのセンサから得られる情報に対して、Unscented Kalman Filter を適用して統計的に観測誤差を軽減した。これらのセンサ情報は直前の時刻との差分から得られるので、本手法では、姿勢の初期値が正確でないと、誤差が蓄積する。

Ono ら [7] は、音源からの直接音到達時間の差から、音源位置と各マイクの位置、録音開始時刻を推定する Blind Alignment 問題を定義し、補助関数法による解法を示した。本手法は計算コストが高いことから、オフライン処理を想定しているため、ロボットの姿勢推定のような逐次的にマイクロホン位置を推定する問題には不向きである。また、入力である到達時間差に含まれる誤差は考慮されていない。

Miura ら [8] は、Simultaneous Localization and Mapping (SLAM) に基づくオンラインマイクロホン位置推定法として、ロボットの周りを拍手をしながら周回する人 (音源) を仮定し、人の位置および、マイクロホンの位置と録音開始時刻を同時に推定した。本手



□ m_i :マイクロホン ○ s_j :小型スピーカ
図2 ホース型ロボットの形状.

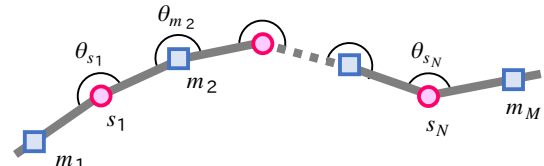


図3 ホース型ロボットの姿勢のモデル.

法では、周回する人の仮定を用いて計算コストを削減している。しかし、瓦礫内でホース型ロボットの周囲を周回する音源を設置することは難しい。

本稿では、試験音を複数回再生し、ホースの形状制約を用いた状態空間モデルを用いて、試験音の到達時間差を手掛かりとしたオンライン姿勢推定を行う。到達時間差には、現在のマイクロホン位置と音源位置に関する情報が含まれるので、誤差の蓄積を軽減できる。また、ホース形状はホース上のマイクロホンと小型スピーカの位置に制約を加える。これを用いて、マイクロホンと小型スピーカの位置を表す状態空間の低次元化を行う。

3. 音によるホース型ロボットの姿勢推定

図2にホース型ロボットの形状を示す。ロボットには、マイクロホンと小型スピーカを交互に間隔 l だけ離して配置する。各マイクロホンと小型スピーカをそれぞれ、リモートオペレータ側から順に $m_1, s_1, \dots, s_N, m_M$ とする。ただし、 M, N はそれぞれマイクロホンと小型スピーカの個数を表し、 $N = M - 1$ である。本稿では、ロボットの姿勢は二次元平面上での曲線で表現されるとし、 m_i, s_j の各座標は、 $x_{m_i}(t), x_{s_j}(t) \in \mathbb{R}^2$ とする。以下に本稿が扱う問題設定を示す。

入力 波形が既知の試験音を録音した M チャンル同期音響信号

出力 各マイクロホンとスピーカの座標

仮定 (1) マイクロホンとスピーカはロボット上に配置される。

(2) 再生しているスピーカは既知。

(3) $x_{m_1}(t)$ と $x_{s_1}(t)$ は既知。

入力は到達時間差を得るために使用される。ここで、試験音とは到達時間差推定のためにスピーカで再生する元信号である。出力は、マイクロホンとスピーカの二次元座標であり、各マイクロホンとスピーカは仮定(1)によりホース上に配置されるため、ホースの姿勢を表す。仮定(2)により、複数のスピーカを同じスピーカと混同しない。仮定(3)により、推定姿勢の回転と平行移動が制限される。

3.1 概要

本手法では、ホース型ロボットの姿勢を音の到達時間差を手がかりとし、Particle Filter [9] を用いて推定する。Algorithm 1 に本オンライン姿勢推定法の1ステップでの動作の概略を示す。次に本手順で使用するホース型ロボットの姿勢を表す状態空間モデル、Particle Filter について述べる。

3.2 姿勢の定式化

図3にモデル化したホースの姿勢を示す。ロボットの姿勢は、マイクロホンとスピーカを関節とする区分

Algorithm 1 マイクロホン位置の逐次推定

for $j = 1 \rightarrow N$ do

試験音をスピーカ s_j から再生

M チャンルマイクロホンアレイで録音

試験音の到達時間差を推定

end for

得られた到達時間差を Particle Filter に入力
事後確率最大となるように姿勢を更新

線形曲線により近似する。状態変数 $\xi(t)$ は、ホースの姿勢を表し、各関節の角度からなる $M + N - 2$ 次元ベクトルである。

$$\xi(t) = [\theta_{m_2}(t), \dots, \theta_{m_{M-1}}(t), \theta_{s_1}(t), \dots, \theta_{s_N}(t)]^T$$

各マイクロホンとスピーカの座標はリモートオペレータ側のマイクロホンとスピーカの座標 $x_{m_1}(t), x_{s_1}(t)$ を用いて再帰的に計算される。

$$x_{m_i}(t) = x_{s_{i-1}}(t) + l[\cos \theta'_{m_i}(t), \sin \theta'_{m_i}(t)]^T$$

$$x_{s_i}(t) = x_{m_i}(t) + l[\cos \theta'_{s_i}(t), \sin \theta'_{s_i}(t)]^T$$

$$\theta'_{m_i}(t) = \pi + \theta'_{s_{i-1}}(t) - \theta_{m_i}(t)$$

$$\theta'_{s_i}(t) = \pi + \theta'_{m_i}(t) - \theta_{s_i}(t), \theta'_{s_1}(t) = \theta_{s_1}(t)$$

状態遷移モデル 状態遷移はランダムウォークで表現する。

$$p(\xi(t)|\xi(t-1)) \sim \mathcal{N}(\xi(t-1), \Sigma_r) \quad (1)$$

$$\Sigma_r = \sigma_r^2 I_{M+N-2}$$

ここで、 σ_r はランダムウォークの標準偏差で、 I は単位行列である。また、ホースに力が働いていないときは各関節角が直線となることを考慮して、初期状態は角度 π を平均とする正規分布に従うとする。

$$p(\xi(0)) \sim \mathcal{N}(\xi_\pi, \Sigma_0) \quad (2)$$

$$\xi_\pi = [\pi, \dots, \pi]^T \in \mathbb{R}^{M+N-2}$$

$$\Sigma_0 = \sigma_0^2 I_{M+N-2}$$

ここで、 σ_0 はホースの曲がりやすさを表す標準偏差である。

観測モデル 観測変数 $y(t) \in \mathbb{R}^{N \times (M-2)}$ は、全スピーカに対する各マイクロホン間の到達時間差とする。つまり、 $\tau_{m_j \rightarrow i}^{s_j}$ を要素とし、以下のように表す。

$$y(t) = \left[\begin{array}{c} \tau_{m_j \rightarrow i}^{s_j} \\ \vdots \\ \tau_{m_j \rightarrow i}^{s_j} \end{array} \middle| \begin{array}{c} j \in [1, \dots, N], \\ i \in [1, \dots, j-1, j+2, \dots, M] \end{array} \right] \quad (3)$$

本モデルでは $\tau_{m_j \rightarrow j}^{s_j}, \tau_{m_j \rightarrow j+1}^{s_j}$ は定数なので観測変数に含めない。このときの観測モデルを以下に示す。

$$p(y(t)|\xi(t)) = \prod_j \prod_i p(\tau_{m_j \rightarrow i}^{s_j} | \xi(t))$$

$$p(\tau_{m_j \rightarrow i}^{s_j} | \xi(t)) \sim \mathcal{N}\left(\frac{D_{m_i}^{s_j}(t) - D_{m_j}^{s_j}(t)}{c}, \sigma_{m_i}^{s_j}(t)\right)$$

$$\sigma_{m_i}^{s_j}(t) = d_1 \times D_{m_i}^{s_j}(t) + d_2 \quad (4)$$

ここで、 c は音速を表し、 $D_{m_i}^{s_j}(t), D_{m_j}^{s_j}(t)$ はそれぞれ、 m_i, s_j 間、 m_j, s_j 間の距離である。また、式(4)は到達時間差推定値の誤差の標準偏差が m_i, s_j 間の距離に線形依存することを表し、 d_1, d_2 は距離と分散の関係を表すパラメータである。

3.3 Particle Filter による姿勢推定

Particle Filter を用いたオンライン姿勢推定法を述べる。推定対象は、観測 $\mathbf{y}(t)$ が与えられたときの姿勢 $\xi(t)$ の事後確率である。この分布を P 個のパーティクルを用いて近似計算する。

$$p(\xi(t)|\mathbf{y}(t)) \approx \sum w^p(t) \delta_{\xi^p(t)}(\xi(t)) \quad (5)$$

ここで、 $w^p(t)$ はパーティクル p の重み、 $\xi^p(t)$ は状態変数である。 $w^p(t), \xi^p(t)$ は次の2ステップにより計算する。1) 提案分布を用いて $\xi^p(t-1)$ を更新する。

$$\begin{aligned} \xi^p(t) &\sim p(\xi^p(t)|\mathbf{y}(t), \xi^p(t-1)) \\ &= \alpha \mathcal{N}(\xi(t-1), \Sigma_r) + \beta \mathcal{N}(\xi_\pi, \Sigma_0) \end{aligned}$$

提案分布には、ランダムウォーク(式(1))と初期状態(式(2))の分布を重み $\alpha, \beta(\alpha + \beta = 1)$ で混合した分布を用いる。これにより、ひとつの姿勢へのパーティクルの集中を軽減し、状態空間の探索を効率化する。2) 各パーティクルの重み $w^p(t)$ を算出する。

$$w^p(t) \propto \frac{p(\mathbf{y}(t)|\xi^p(t))p(\xi^p(t)|\xi^p(t-1))}{p(\xi^p(t)|\mathbf{y}(t), \xi^p(t-1))}$$

全パーティクルの重み計算の後、重み $w^p(t)$ は $\sum_p w^p(t) = 1$ となるように正規化する。推定値は最大事後確率推定により得る。

$$\hat{\xi}(t) = \xi^{\hat{p}}(t) \text{ s.t. } \hat{p} = \underset{p}{\operatorname{argmax}} w^p(t)$$

本手法では、Sampling Importance Resampling (SIR) アルゴリズムに基づき、各ステップごとにパーティクルが持つ重みに比例して、リサンプリング処理を行う。

4. 評価実験

シミュレーテッドデータ実験と実データ実験により本手法を評価した。

4.1 シミュレーテッドデータ実験

実験設定 1) 直線, 2) 半円, 3) S字の三種の姿勢について報告する。マイクロホンの数 M は $M = 8$ で、 $l = 0.2$ [m] のホース型ロボットとした。各パーティクルは式(2)に従い初期化する。パラメータ $\sigma_0 = 0.5$ [rad] は実験的に定めた。観測データは各マイクロホンとスピーカの位置から幾何的に計算したシミュレーテッドデータを用い、全スピーカからの到達時間差の集合を1度の観測(式(3))として20回分の観測データを生成した。観測誤差の標準偏差は 30 [μ s] (0.0102 [m]) とした。

本実験では、同じ観測を用いて100回姿勢推定を試行し、推定結果の誤差の平均と標準偏差を評価した。評価尺度には、先端位置の誤差と、 m_1, s_1 以外のマイクロホンと小型スピーカの位置の平均誤差を用いた。また、本手法では m_1, s_1 を結ぶ線分からの鏡対称となる姿勢の判別はできないので、推定結果の m_M と正解データの m_M が近くなるように姿勢を反転して評価を行った。

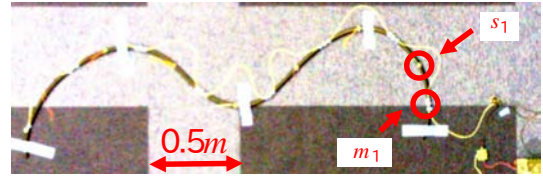


図4 実験で用いたホース型ロボットのモックアップ。

結果 図5(a)-5(c)に結果を示す。100回の試行結果について、誤差の平均を折れ線グラフで、標準偏差をエラーバーで示す。7回以上の観測でいずれの姿勢でも誤差が収束し、先端の位置誤差では 0.4 [m]、全体の誤差では 0.2 [m] 以下となっている。また、図6に各姿勢についてその推定結果の一例を示す。いずれも妥当な姿勢を推定していることが分かる。

4.2 実データ実験

モックアップ 図4に示すようにホースにマイクロホンと小型スピーカを装着したモックアップを構築した。本体はポリプロピレン製のコルゲートチューブ(内径 15 [mm])である。表面には $l = 0.2$ [m] 間隔で、小型スピーカとマイクロホンを配置した。小型スピーカには、直径 20 [mm] の磁気スピーカを用いた。マイクロホンには、デジタルMEMSマイク (17 [mm] \times 17 [mm])を使用した。マイクロホンの数 M は $M = 8$ 、両端のマイクロホン間のホースの長さは 2.8 [m] である。

実験設定 実験は残響時間 (RT60) が 1.4 [s] で、室温が 23°C である実験室で行った。試験音には TSP 信号 (1.0 [s]) を用いた。図4に示すように、モックアップを実験室の床に固定し、20回分の観測データとして、試験音を各スピーカから20回ずつ再生して録音した。Particle Filter の入力には、観測した音響信号から推定した到達時間差を用いた。到達時間差は、正確に測定した小型スピーカからの試験音と、入力との一般化相互相関 (GCC-PHAT) [10] を計算し得る。また、各パーティクルは式(2)に従い初期化し、パラメータ $\sigma_0 = 0.5$ [rad] は実験的に定めた。D/A変換器は、RASP-ZX¹を用い、録音はHARKを用いて 16 kHz、 24 bit でサンプリングを行った。評価方法は4.1と同じく、100回の試行について先端の位置誤差と全体の平均誤差を評価した。

結果 図5(d)に結果を示す。5回以上の観測で、先端位置の誤差の平均は 0.4 [m] 以下に、全体の誤差も 0.2 [m] 以下になり、収束していることが分かる。また、図7に状態推定結果の一例を示す。観測回数の増加と共に、推定値が正解に近づいている。以上より、本手法による姿勢推定が有効であることが確認された。

4.3 議論と今後の課題

シミュレーテッドデータ実験と $RT60 = 1.4$ [s] の残響環境下での実データ実験により本手法の有効性を確認した。しかし本手法には次の2つの問題点が存在する。

1) 鏡対称問題。本稿で示したマイクロホンとスピーカの配置では m_1, s_1 を軸として鏡対称な姿勢はすべて同じ到達時間差を観測するため、区別ができない。また、ホースの手元から先端へ向かう途中の関節角から鏡対称となる姿勢が局所最適解となり、推定誤差を増大させることを確認した。この局所解は、観測モデル

¹株式会社 システムインフロンティア製

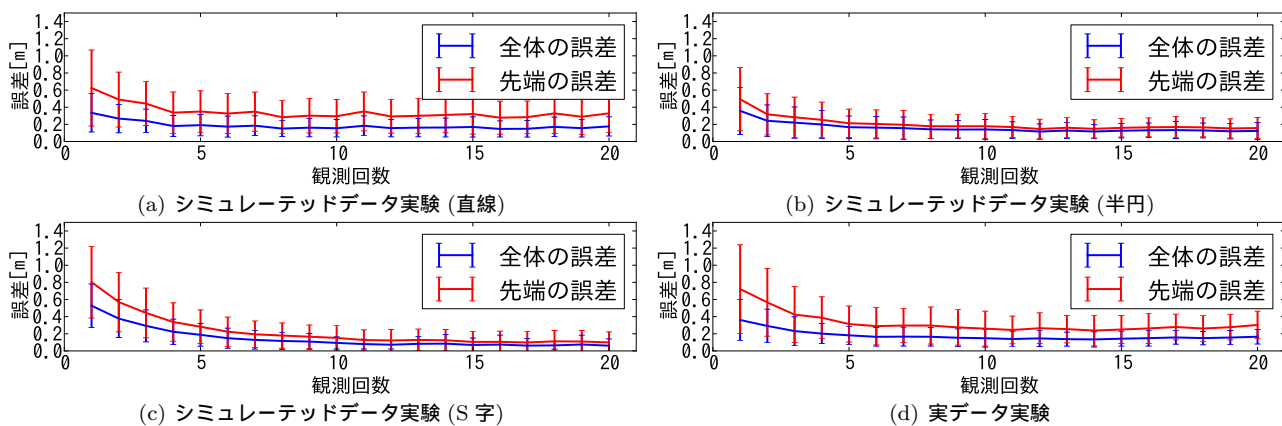


図5 実験結果の推定誤差．誤差の平均を折れ線グラフで，標準偏差をエラーバーで示す．

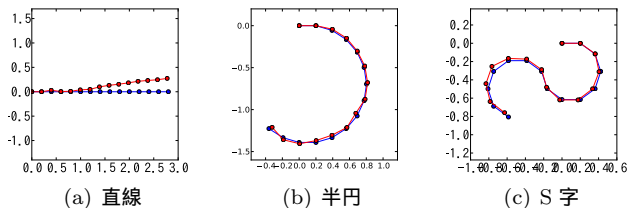


図6 シミュレーテッドデータ実験の推定結果の一例．赤線が推定値，青線が正解を示す．

の分散 (式 (4)) を大きく設定することにより回避できるが，推定精度とのトレードオフとなることが分かった．この問題は曲げセンサなどの各関節の曲り方向がわかるセンサの導入，加速度センサ・ジャイロセンサなどの従来手法との統合で対策できる．

2) 姿勢推定精度．先端位置の誤差は手元に近い関節角の推定誤差が大きく影響する．本手法では，パーティクルが到達時間差に近い姿勢を表す状態に集中するように設計した．そのため，全体の姿勢は正解に近いが，手元の関節角である $\theta_{s_1}(t)$ の誤差が大きい姿勢を推定結果とする場合がある．これは，4.1 節のシミュレーテッドデータ実験の結果のうち姿勢が直線の推定誤差が他と比べて大きいことから確認できる．直線型は，他の姿勢に比べて $x_{m_1}(t)$ と $x_{m_M}(t)$ が最も離れており，その影響が出やすい．この問題は，提案分布を Kalman Filter から得て，分散と平均が動的に変更される Unscented Particle Filter (UPF) を用いれば改善が期待できる．

5. まとめ

本稿では，ホース型ロボットの navigational usability と mission usability の向上のために，Particle Filter を用いた音によるホース型ロボットの姿勢推定法を述べた．本手法は，1) ホース上に複数の小型スピーカとマイクロホンアレイを取り付け，2) 各小型スピーカから再生される試験音のマイクロホンへの到達時間差を推定し，3) ロボットの形状制約を含んだ状態空間モデルを Particle Filter による事後確率最大化により姿勢推定を行う．評価実験では，シミュレーテッドデータ実験と $RT60 = 1.4[s]$ の残響環境下での実データ実験により本手法を評価し，有効性を確認した．

今後の課題としては，navigational usability においては，鏡対称問題の解決，UPF の導入，および，温度変化や，残響・反射の強い環境，直接音が取得できない場合への対処など頑健性の向上が，mission usability においては，音源定位・分離技術を利用したりリモート

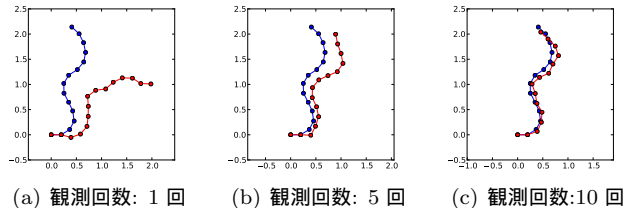


図7 実データ実験の推定結果の一例．赤線が推定値，青線が正解を示す．

オペレータへのマルチモーダル情報提示システムの開発などが挙げられる．

謝辞 本研究の一部は科研費基盤 (S) No.24220006 の支援を受けた．ホース型ロボットのサンプルをご提供いただいた東北大学大学院 田所諭教授，昆陽雅司准教授らに感謝する．

参考文献

- [1] A. Kitagawa, et al. Development of small diameter active hose-ii for search and life-prolongation of victims under debris. *Journal of Robotics and Mechatronics*, Vol. 15, No. 5 (2003), pp. 474–481.
- [2] H. Namari, et al. Tube-type active scope camera with high mobility and practical functionality. *IROS-2012*, pp. 3679–3686.
- [3] R. R. Murphy. Navigational and mission usability in rescue robots. *Journal of the Robotics Society of Japan*, Vol. 28, No. 2 (2010), pp. 142–146.
- [4] N. Kazuhiro, et al. Design and implementation of robot audition system 'HARK' open source software for listening to three simultaneous speakers. *Advanced Robotics*, Vol. 24, No. 5-6 (2010), pp. 739–761.
- [5] T. Mizumoto, et al. Design and implementation of selectable sound separation on the texai telepresence system using hark. *ICRA-2011*, pp. 2130–2137.
- [6] M. Ishikura, et al. Shape estimation of flexible cable. *IROS-2012*, pp. 2539–2546.
- [7] N. Ono, et al. Blind alignment of asynchronously recorded signals for distributed microphone array. *WASPAA-2009*, pp. 161–164.
- [8] H. Miura, et al. Slam-based online calibration of asynchronous microphone array for robot audition. In *IROS-2011*, pp. 524–529.
- [9] M.S. Arulampalam, et al. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE TSP*, Vol. 50, No. 2 (2002), pp. 174–188.
- [10] C. Knapp, et al. The generalized correlation method for estimation of time delay. *IEEE TASSP*, Vol. 24, No. 4 (1976), pp. 320–327.