解説

能動スコープカメラの極限ロボット聴覚

Extreme Audition for Active Scope Camera

坂 東 宜 昭^{*1} 安 部 祐 一^{*2} 糸 山 克 寿^{*3} 奥 乃 博^{*4}

^{*1}産業技術総合研究所 ^{*2}東北大学 ^{*3}東京工業大学／ホンダ RI-JP ^{*4}早稲田大学

Yoshiaki Bando^{*1}, Yuichi Ambe^{*2}, Katsutoshi Itoyama^{*3} and Hiroshi G. Okuno^{*4}

^{*1}National Institute of Advanced Industry and Technology ^{*2}Tohoku University ^{*3}Tokyo Institute of Technology / Honda RI-JP ^{*4}Waseda University

1. は じ め に

人や動物が侵入できない環境を探索するレスキューロボットにおいて、ロボット聴覚技術は、目視できない被災者捜索に活用されているうえ、自己位置の推定にも有用である。クローラロボットやドローンにマイクロホンアレイを取り付ければ、目視困難な被災者の音声を検出・定位できる[1]～[3]。また、スピーカをピーコンとしてマイクロホンとの位置関係を推定すれば、自己位置推定に応用できる[4]。例えば、全地球測位システム（GPS）や地磁気センサの使えない屋内環境で、音響センサによる位置推定が広く研究されている[4][5]。電波が強く減衰する海中でも、海面ブイから発する基準音を用いた自己位置推定が活用されている[6]。

本稿では、能動スコープカメラ[7]でのロボット聴覚技術について解説する。本ロボットの目的は、瓦礫の隙間のような狭隘環境の探索である。直径約 50 [mm]・全長 5 [m] 以上の柔軟な本体が特徴で、体表の繊毛を振動させ、外環境との摩擦率を用いて推進する。本ロボットには関節がないため、狭く複雑な瓦礫環境でも、柔軟に本体を曲げながら侵入できる。また、米国の Jacksonville [8] や熊本震災[9]での探索タスクに適用され、有効性が検証された。一定の探査能力が確認できた一方、繊毛の摩擦率のみでは高い段差を踏破できないうえ、先端のカメラとモノラルマイクロホンのみでの情報収集は不十分であることが判明していた。

情報収集能力の観点では、柔軟なロボットをオペレータが容易に操縦でき、被災者の手がかりを効率的に発見できる能力が不可欠である。本ロボットの操縦では、ロボットの姿勢（形状）に応じて取るべき操作が異なる。しかし、従来システムでオペレータが確認できる情報は、視界が制限された先端カメラの映像のみであり、操縦に熟練を要した。ロボットの全身に装着した慣性センサ（加速度・角速度センサ）アレイを用いる姿勢推定法も提案されたが、累積誤差が

図 1 感覚統合型能動スコープカメラ

原理的に除去できず長時間の運用には不向きであった[10]。また、被災者の手がかり発見のために、マイクロホンで周囲の音を録音しているが、自身の走行雑音が原因で音声聞き取りづらかった。従来は、音の有無を確認するため逐一ロボットを静止させており、捜索に時間を要した。

ImPACT-TRC [11] では、飛躍的な駆動性能と情報収集能力の向上を目指し、空気噴射ノズルを先端に搭載する感覚統合型能動スコープカメラ（図 1）を開発した。空気噴射により先端部を約 30 [cm] 浮上させ推進できるようになり、段差の踏破性能が向上した。また本ロボットは先端カメラに加え、全身にマイクロホン・スピーカ、慣性センサおよび振動センサアレイを搭載する。これらを用いて、被災者捜索の効率化のためのロボットビジョン [12][13]・聴覚 [14]～[16] および触覚機能 [13][17] を開発した。

能動スコープカメラのロボット聴覚機能としては、音声強調 [14][16][18] と姿勢推定 [15] を開発した。自身の走行雑音のなかから音声を抽出する音声強調は、効率的な被災者の捜索に不可欠である。また、ロボット上のマイクロホンとスピーカを用いた姿勢推定は、慣性センサの累積誤差を除去できる。これらの機能を実現するには、ロボット自身と瓦礫環境に起因する以下の三つの課題があった：

- 環境依存の走行雑音：走行雑音には地面との接地音が含まれるため、走行環境に応じて雑音の音色が変動する。
- マイクロホンの移動：ロボット上のマイクロホンの配置は、ロボットの運動に伴って時々刻々と変化する。
- マイクロホンの遮へい：瓦礫に遮へいされ、一部のマイクロホンで目的音を十分な音量で収録できないことがある。

ImPACT-TRC では、これらの課題を克服する姿勢推定と音声強調を目指し研究を進めてきた。以降では、従来のロ

原稿受付 2019 年 8 月 9 日

キーワード：Disaster Robotics, Active Scope Camera, Robot Audition, Posture Estimation, Speech Enhancement

^{*1}〒135-0064 江東区青海 2-4-7

^{*2}〒980-8579 仙台市青葉区荒巻字青葉 6-6-01

^{*3}〒152-8552 目黒区大岡山 2-12-1

^{*4}〒169-0072 新宿区大久保 2-4-12

^{*1}Koto-ku, Tokyo

^{*2}Aoba-ku, Sendai-shi, Miyagi

^{*3}Meguro-ku, Tokyo

^{*4}Shinju-ku, Tokyo

ロボット聴覚・音響信号処理技術を整理しながら、これらの課題に対処した音声強調と姿勢推定について解説する。

2. 感覚統合型能動スコープカメラ

ImPACT-TRC で開発した感覚統合型能動スコープカメラは、ロボット本体の長さ 7 [m]、外径 50 [mm]、重さが 5 [kg] の長尺で柔軟、軽量なロボットである。写真を図 1 に、ロボット構成図を図 2 に示す。本プロジェクトでは、特に従来の能動スコープカメラの飛躍的な駆動性能の向上と、新たな高度情報収集能力の付与を目指して開発に取り組んできた。以下では、関連するロボットを概観したのち、開発した能動スコープカメラの概要を説明する。

2.1 従来の柔軟索状ロボット

ホース型の柔軟な索状ロボットは、瓦礫内探査や大型機械等の点検に用いられるスコープカメラを能動化し、より遠方かつ複雑な環境を探索できるように開発されている。例えば、Active Hose [19] は、コルゲートチューブ状の本体に能動小型車輪と、先端に首振り機構を持つ。本ロボットは、能動車輪の間隔より狭い障害物に乗り上げると推力を失ってしまう。本問題を解決するため、Glow Hose [20] や Hawkes ら [21] のロボットは、空気で膨らませたビニール状の本体を持ち、先端部に本体内部から新たなビニールを展開する機構を持つ。外部との摩擦で推進するのではなく、本体の延伸により先端部を推進させるため、外界との接触度合いによらず推進できる。別の解決策として、能動スコープカメラ [22] は、コルゲートチューブ全体に振動する絨毛を装着し、体表全体で推進力を得ることで小型車輪での問題を解決した。Glow Hose らと比較し、ロボット内が中空のため、長尺の本体全体にセンサを装着することができ、瓦礫環境の分散センシングが可能になる利点がある。

2.2 感覚統合型能動スコープカメラ

感覚統合型能動スコープカメラでは従来の絨毛駆動機構に加えて、空気噴射ノズルによる駆動機構、根本からロボットを送り出す機構を搭載して駆動性能の向上を目指した (図 2)。ロボット胴体全体には傾斜した絨毛を巻きつけ、絨毛駆動機構を搭載した。振動モータで胴体を振動させることで、傾斜した絨毛が固着滑りを繰り返して推進力を得る [23]。

図 2 感覚統合型能動スコープカメラの構成

全身に絨毛が巻かれているため、胴体が接地するほど推進力が発生し、瓦礫内部などの狭隘空間への挿入性が高い。先端には二自由度能動ノズル [24] を搭載した。根本からコンプレッサで圧縮空気を送り、ノズルから空気を音速で噴射することで反力を得る。ノズルの噴射方向は二軸回りに制御可能であり、噴射方向を制御することで、先端を安定に浮上させたり (高さ 300 [mm])、左右に移動させることができる。この浮上機能により、これまで難しかった段差踏破や進入経路の選択が可能となったうえ、俯瞰視点による状況の把握ができるようになった。さらに、瓦礫挿入口でロボットを送り出すための送り出し機構を開発し [25]、ロボットの挿入を機械化した。ロボット挿入長の計測や、一部操作の自動化が可能となった。従来は最低二人必要だったロボットの操縦を、一人のオペレータが統合インターフェースを用いながら実施できるようになった。

新たな情報収集能力を付与するために、本ロボットには様々なセンサを搭載した (図 2)。瓦礫内部の映像に加えて、瓦礫環境の三次元形状を把握することで、被災者の位置の把握や、瓦礫内への救助計画をたてるのに役立つ。Visual simultaneous localization and mapping (visual SLAM) とよばれるカメラの移動軌跡推定と三次元環境地図構築を同時に行う技術を付与するために、ロボット先端には最大 160 [fps] 実現可能な高速カメラを搭載した。またこのほかに、ロボット胴体に多数のマイクロホン、スピーカ、慣性センサ、および振動センサを分散的に内蔵した。マイクロホンアレイ・スピーカアレイは、要救助者の存在や状況を音声情報から確認するために搭載した。また、慣性センサアレイはロボットの姿勢推定に使用する。索状体は柔軟で環境に応じて形状が大きく変化するため、索状体全体の形状や先端位置の情報は、ロボットの経路選択や探索位置の把握に不可欠である。慣性センサ単体では、累積誤差を除去できないが、音響センサと統合することで、本問題の解決を目指した。振動センサアレイは、環境との接触をモニタするために使用する。

スコープカメラは環境との接触に応じて駆動力を発生するため、胴体と環境との接触状況を把握することが操縦において不可欠である。これらの駆動能力や情報収集能力の統合には、ロボット用ミドルウェア ROS (robot operating system) を採用することで (図 2)、開発の効率化を行った。また、メンテナンス性を確保するために、ロボットを複数のモジュールの連結構造として構成し、搭載している多数のセンサを個別に着脱可能とした。

3. 関連研究

本章では、音響センシングに基づく姿勢推定と音声強調の関連研究を概観する。

3.1 姿勢推定

姿勢推定はマイクロホンアレイの配置推定と見なすことができ、非同期分散アレイのキャリブレーション問題として広く研究されている [5] [26]。Ono ら [5] は、ある音源信号を観測した各マイクロホンの到達時間差を入力として、マ

イクロホンごとの時刻ずれと位置および、音源位置を同時推定した。また、Miura ら [4] は、同様の問題を SLAM [27] の枠組みで定式化し、拡張カルマンフィルタを用いてオンライン推論した。Ono らの手法は、音源とマイクロホン双方の静止を、Miura らの手法は、マイクロホンの静止を仮定していた。一方、Parviainen ら [28] は、マイクロホンと音源双方が移動する条件で、これらの位置とマイクロホンの時刻ずれを同時推定する手法を提案した。この手法では、マイクロホンと音源が一對ずつペアで相対位置がおおむね不変であるという仮定を用いる。このように、マイクロホンとスピーカの双方が運動する場合、何らかの配置制約が必要となる。スピーカとマイクを隣接させれば、推定する変数を削減できる。しかし、スピーカとマイクロホンは接近すると音割れしてしまうため、一定以上離す必要がある。

ホース状のマイクロホンアレイの配置推定は、海中での牽引ソナーの形状推定として広く研究されている [29]~[31]。牽引しているソナーで周囲の音源（船舶や潜水艦）を観測し、各マイクロホン間での到達時間差からソナーの形状を推定する。海中の牽引ソナーと比較し、瓦礫環境では周囲の瓦礫により音が回折することがあり、一部の到達時間差からは正確に距離推定できない問題がある。

3.2 音声強調

マイクロホンアレイを用いれば、音声と雑音の空間的特徴の差からこれらを分離することができる [32]~[34]。このアプローチでは、マイクロホン間の位相差と音量差が音源位置に応じて異なることを利用する。独立ベクトル分析 (IVA) [32] や独立低ランク行列分析 (ILRMA) [35]、複素混合ガウスモデル [33] [36] などが知られている。これらの手法は、空間情報に基づき雑音と音声を分離するため、雑音の事前学習が不要で環境適応性が高い。しかし、マイクロホンの配置が既知 [37] であるか、マイクロホンと音源の静止を仮定 [32] [33] [35] するため、マイクロホンが常に移動する本ロボットには不適である。マイクロホン間の音量差は、位相差より位置変動の影響が小さいため、音量差のみを用いる非負値テンソル因子分解 (NTF) [38] [39] が提案されている。NTF では、音量差の変化が無視できる範囲であれば、音源が移動しても分離できる。ただし、各音源のマイクロホン間の音量差 [39] や、音源・雑音スペクトルのテンプレート [38] が事前に必要である。

音声と雑音のスペクトル特徴の違いにのみ着目する、モノラル音声強調も広く研究されている。ロボットの自己雑音抑圧では、対象雑音を事前に収録して雑音テンプレートを構成し、テンプレートに基づき雑音を除去する手法 [40] [41] が広く研究されている。能動スコープカメラでは、雑音は環境依存性が高いため事前収集が困難である。事前に雑音信号を収集せずに音声強調できる手法として、低ランク・スパース分解が提案されている [42]~[44]。例えばロバスト非負値行列因子分解 (RNMF) [43] [44] は、観測振幅スペクトログラムを、高頻度（低ランク）に出現する成分と、低頻度（スパース）な成分に教師なしで分解する。能動スコープカメラの走行雑音は振動モータに起因するため低ランク性が

高く、この分解により雑音と音声を分離できる [18]。低ランク・スパース分解を多チャンネル法に拡張することで、さらなる性能向上が期待できる。

4. 能動スコープカメラの極限音響

本章では、能動スコープカメラのロボット聴覚を構成する、姿勢推定と音声強調について解説する。本ロボットでは、1) 環境依存の走行雑音、2) マイクロホンの移動、3) マイクロホンの遮蔽への対処が不可欠である。姿勢推定では、頑健な試験音のオンセット検出器を構成することで課題 1 と 3 に対処し、状態空間モデルに基づくオンライン姿勢推定により課題 3 に対処した。音声強調では、振幅領域での低ランク・スパース分解により課題 1 と 2 に対処し、マイクロホンごとの音声音量を同時推定することで課題 3 に対処した。

4.1 姿勢推定

音響センシングによる姿勢推定では、ロボット上のスピーカで再生した試験音の到達時間差を用いて現在の姿勢を推定する。本手法は、以下の三つのステップを繰り返すことで、逐次的に現在の姿勢を推定する：

- (1) ロボット上のスピーカを一つ選び試験音を再生する
- (2) 試験音が各マイクロホンへ到達する時間差を推定する
- (3) 到達時間差から現在の姿勢を推定する

ロボット上の s 番め ($s \in \{1, \dots, S\}$) のスピーカで再生した試験音について、 m, n 番め ($m, n \in \{1, \dots, M\}$) のマイクロホン間の到達時間差 $\tau_{ksmn} \in \mathbb{R}$ は、次式で定義される：

$$\tau_{ksmn} = t_{ksm} - t_{ksn}. \quad (1)$$

ここで k は観測番号を表し、 $t_{ksm} \in \mathbb{R}_+$ は m 番めのマイクロホンでの試験音のオンセット時刻である。試験音には、図 3 に示す時間伸長パルス (TSP) を採用した。TSP は、小型のスピーカでも大きな音量で再生でき高い時間分解能でオンセット検出できるので、走行雑音に対し頑健な到達時間差推定が可能である。図 4 にオンセット時刻推定の概要を示す。オンセット時刻の推定には、残響と反射音に対し頑健な一般化相互相関関数である GCC-PHAT [45] [46] を使用した。ただし、推定した到達時間差は、ロボット周辺の障害物の干渉により、正確に距離（差）推定できないことがある。自由空間での到達時間差はロボット上のマイクロホン・スピーカの間の距離より必ず小さい特性を用いて、はずれ値を除去する [15]。直前の推定姿勢に基づき判定する

図 3 TSP のスペクトログラム

図4 オンセット推定の概要

ことも可能である [47].

ロボットの姿勢は、推定された到達時間差 τ_{ksmn} を観測とする状態空間モデルを用いてオンライン推論する。状態空間モデルは、姿勢の状態遷移の確率モデルと、姿勢が与えられたときの観測の確率モデルで構成される。ロボットの姿勢は、ロボット上のマイクロホン・スピーカ位置を頂点とするリンクモデルとして表現し、姿勢の時間発展はランダムウォークで表現する。姿勢から計算される m 番めのマイクロホンと s 番めのスピーカの位置をそれぞれ、 $\mathbf{x}_m$ および $\mathbf{x}_s$ とすると、到達時間差 τ_{ksmn} は、マイクロホン間の距離差で表現できる:

$$\tau_{ksmn} | \mathbf{z} \sim \mathcal{N} \left(\frac{\|\mathbf{x}_m - \mathbf{x}_s\| - \|\mathbf{x}_n - \mathbf{x}_s\|}{C}, \sigma^\tau \right). \quad (2)$$

ここで、 $\sigma^\tau \in \mathbb{R}_+$ は観測誤差の標準偏差を表し、 C は音速を表す。本モデルの推論は、式 (2) に非線形変換が含まれるため、Unscented カルマンフィルタ (UKF) [48] [49] を用いる。UKF は、拡張カルマンフィルタ [27] と異なり微分計算が不要で、パーティクルフィルタ [27] より低計算量で事後分布を近似できる。これらの特性は姿勢推定システムの実装において、加速度センサやジャイロセンサといったほかの観測モデルや、姿勢モデルを個別のモジュールとして独立に定義・管理でき、開発の効率化に繋がった。

上記の到達時間差の観測に加えて、加速度センサから得られる重力方向を統合した姿勢推定では、障害物が存在する環境下でも、リアルタイムで全長約 3 [m] のロボットの姿勢を先端位置誤差 20 [cm] 程度で推定できた [15]。ただし、ロボット上に 1 列に配置されたマイクロホン・スピーカアレイでは、原理上鏡像対称な姿勢は一意に推定できない。一方、従来の慣性センサに基づく手法 [10] は、姿勢を一意に推定できるが累積誤差の問題がある。これらの課題を相補的に解決するため、慣性センサと音響センサを統合し、慣性センサの累積誤差を音響センサで補正する姿勢推定法も開発している [50]。一意に姿勢推定できるうえ、突発的な衝撃を加えても、誤差を一定以下に抑圧できることを実験室環境にて確認している。

4.2 音声強調

本ロボットで頑健に動作する音声強調のため、低ランク・スパース分解法を多チャンネルに拡張した、ロバスト非負値テンソル因子分解法 (RNTF) [16] を開発した。本手法は、観測した混合音の振幅スペクトログラムを、各チャンネルの低ランク成分 (雑音) と、チャンネル間で共通のスパース成分 (目的音声)、マイクロホン間での音声音量に教師なしで分解する (図 5)。振幅領域で動作するためロボットの運動

図5 ロバスト非負値テンソル因子分解法

に頑健なうえ、チャンネルごとの音声音量を推定することで、アレイが一部遮蔽されても頑健に強調できる。

RNTF は、混合音に対する確率的生成モデルをベイズ推論することで音声強調する。本モデルでは、 M チャンネル振幅スペクトログラム $\mathbf{x}_{ft} \in \mathbb{R}_+^M$ を、低ランク成分 $\mathbf{n}_{ft} \in \mathbb{R}_+^M$ と共通スパース成分 $s_{ft} \in \mathbb{R}_+$ の和で表現する:

$$\mathbf{x}_{ft} \approx \mathbf{y}_{ft} = \mathbf{n}_{ft} + \mathbf{g}_t s_{ft}. \quad (3)$$

ここで、 $\mathbf{g}_t \in \mathbb{R}_+^M$ は、各チャンネルでの音声音量を表す。この近似のもっともらしさは、観測 $\mathbf{x}_{ft}$ に対する再構成音 $\mathbf{y}_{ft}$ のポワソン尤度で測る:

$$\mathbf{x}_{ft} \sim \prod_{m=1}^M \mathcal{P}(\mathbf{x}_{ft} | \mathbf{y}_{ft}). \quad (4)$$

ただし、 $\mathcal{P}(\lambda) \propto \lambda^x e^{-\lambda}$ はポワソン分布を表す。ポワソン尤度の最大化は、Kullback-Leibler 擬距離の最小化し、音源分離で広く用いられている [51]。

低ランク成分 $\mathbf{n}_{ft}$ はさらに、 K 本の基底スペクトル $\mathbf{w}_{km} \in \mathbb{R}_+^F$ と、その音量 $\mathbf{h}_{km} \in \mathbb{R}_+^T$ の積で表現する:

$$\mathbf{n}_{ft} = \sum_{k=1}^K \mathbf{w}_{km} \mathbf{h}_{km}. \quad (5)$$

この分解は、非負値行列因子分解 (NMF) とよばれ、雑音スペクトログラム (低ランク成分) をランク K 以下に制約できる。RNTF では、 $\mathbf{w}_{km}$ と $\mathbf{h}_{km}$ にガンマ事前分布を仮定し、雑音の低ランク性をさらに誘導する。一方、音声 s_{ft} は、Jeffrey's 無情報事前分布 [16] [52] を用いてスパースに誘導する。

混合音 $\mathbf{x}_{ft}$ に対する潜在変数 $\{\mathbf{w}_{km}, \mathbf{h}_{km}, s_{ft}, \mathbf{g}_t\}$ の事後分布を推定することで、音声 s_{ft} と雑音 $\mathbf{n}_{ft}$ を分離する。この事後分布は解析的に計算困難なので、変分ベイズ法 [53] を用いて近似推論する。評価実験では、振動モータによる走行雑音と音声を数値混合したデータに対し、半数のマイクロホンが遮へいされても頑健に動作した [16]。また、数フレームごとに逐次実行 [54] するミニバッチ推論法も開発した [16]。200 フレーム (2 秒) 程度のミニバッチを用いれば、組み込み GPU ボード NVIDIA Jetson TX1 (図 6) でリアルタイムに動作する。

RNTF では、音声のスパース性と雑音の低ランク性が当てはまらない音声・雑音では性能劣化する。例えば、衝撃音や急旋回時の空気噴射音のような突発性の雑音や、姿勢推定

図6 NVIDIA Jetson TX1. 単体でROS ノードとしてネットワークに接続できる

の試験音は低ランク性が低い。また、「あー」などの長音はスペクトル特徴の変化が少なく、低ランク成分に分離されやすい。これらの課題に対処するため、空気噴射音や試験音といった既知の非低ランク雑音を事前学習する半教師あり畳み込みNMFの導入[55]や、変分オートエンコーダを用いて音声特徴を事前学習する深層音声事前分布[56][57]も開発している。これらの拡張は、未だ計算量や初期値依存性の課題が残るが、今後の改良によって実時間でのより高精度な音声強調が期待できる。

5. お わ り に

本稿では、ImPACT-TRCで開発した感覚統合型能動スコープカメラのロボット聴覚機能について解説した。従来の能動スコープカメラは、踏破性能が低いという、情報収集能力に課題が残っていた。これらを解決するため、本プロジェクトでは、空気浮上機構を搭載する能動スコープカメラを開発し、スピーカ、マイクロホン、慣性センサおよび振動センサを全身に搭載して情報収集能力の向上を目指した。ロボット聴覚機能としては、操縦効率化のための姿勢推定と、被災者発見効率化のための音声強調を開発した。本ロボットのロボット聴覚では、1) 環境依存性の高い走行雑音、2) ロボットの運動に伴うマイクロホンの運動、3) 瓦礫によるマイクロホンの遮へいの三つの課題があり、既存の方法は適用困難であった。本ロボットでも頑健に動作する、音響センシングに基づく姿勢推定と、低ランク・スパース分解に基づく音声強調を開発した。今後は、より詳細な性能限界を調査するとともに、被験者実験を通して搜索タスクでの有効性を確認していく。また、開発した技術をほかのロボットへも応用し、多くのレスキューロボットに聴覚機能を付与していく。

謝 辞 本研究はImPACT「タフ・ロボティクス・チャレンジ」の支援を受けた。

参 考 文 献

[1] M. Basiri, et al.: "Audio-based relative positioning system for multiple micro air vehicle systems," *Robotics: Science and Systems RSS2013*, no.EPFL-CONF-191181, 2013.
 [2] T. Ohata, et al.: "Improvement in outdoor sound source detection using a quadrotor-embedded microphone array," *IEEE/RSJ IROS*, pp.1902–1907, 2014.
 [3] S.H. Young, et al.: "Detection and localization with an acoustic array on a small robotic platform in urban environments,"

ARMY RESEARCH LAB ADELPHI MD, Tech. Rep. ARL-TR-2575, 2003.
 [4] H. Miura, et al.: "SLAM-based Online Calibration for Asynchronous Microphone Array," *Advanced Robotics*, vol.26, no.17, pp.1941–1965, 2012.
 [5] N. Ono, et al.: "Blind alignment of asynchronously recorded signals for distributed microphone array," *IEEE WASPAA*, pp.161–164, 2009.
 [6] J. Melo, et al.: "Guidance and control of an ASV in AUV tracking operations," *OCEANS 2008*, pp. 1–7, 2008.
 [7] J. Fukuda, et al.: "Remote vertical exploration by Active Scope Camera into collapsed buildings," *IEEE/RSJ IROS*, pp.1882–1888, 2014.
 [8] S. Tadokoro, et al.: "Application of Active Scope Camera to forensic investigation of construction accident," *IEEE ARSO*, pp.47–50, 2009.
 [9] Y. Ambe, et al.: "Use of Active Scope Camera in the Kumamoto earthquake to investigate collapsed houses," *IEEE SSR*, pp.21–27, 2016.
 [10] M. Ishikura, et al.: "Shape estimation of flexible cable," *IEEE/RSJ IROS*, pp.2539–2546, 2012.
 [11] S. Tadokoro: "Overview of the ImPACT Tough Robotics Challenge and strategy for disruptive innovation in safety and security," *Disaster Robotics*, pp.3–22, Springer, 2019.
 [12] S. Arnold, et al.: "Real-time scene parsing by means of a convolutional neural network for mobile robots in disaster scenarios," *IEEE ICIA*, pp.201–207, 2017.
 [13] M. Konyo, et al.: "ImPACT-TRC thin serpentine robot platform for urban search and rescue," *Disaster Robotics*, pp.25–76, Springer, 2019.
 [14] Y. Bando, et al.: "Low latency and high quality two-stage human-voice-enhancement system for a hose-shaped rescue robot," *Journal of Robotics and Mechatronics*, vol.29, no.1, pp.198–212, 2017.
 [15] Y. Bando, et al.: "Microphone-accelerometer based 3D posture estimation for a hose-shaped rescue robot," *IEEE/RSJ IROS*, pp.5580–5586, 2015.
 [16] Y. Bando, et al.: "Speech enhancement based on Bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms," *IEEE/ACM TASLP*, vol.26, no.2, pp.215–230, 2017.
 [17] T. Funamizu, et al.: "Visuo-haptic transmission of contact information improve operation of Active Scope Camera," *IEEE/RSJ IROS*, pp.1126–1132, 2016.
 [18] Y. Bando, et al.: "Human-voice enhancement based on online RPCA for a hose-shaped rescue robot with a microphone array," *IEEE SSR*, pp.1–6, 2015.
 [19] 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, pp.474–481, 2003.
 [20] H. Tsukagoshi, et al.: "Smooth creeping actuator by tip growth movement aiming for search and rescue operation," *IEEE ICRA*, pp.1720–1725, 2011.
 [21] E.W. Hawkes, et al.: "A soft robot that navigates its environment through growth," *Science Robotics*, vol.2, no.8, p.eaan3028, 2017.
 [22] H. Namari, et al.: "Tube-type Active Scope Camera with high mobility and practical functionality," *IEEE/RSJ IROS*, pp.3679–3686, 2012.
 [23] K. Hatazaki, et al.: "Active Scope Camera for urban search and rescue," *IEEE/RSJ IROS*, pp.2596–2602, 2007.
 [24] A. Ishii, et al.: "Design and development of biaxial active nozzle with flexible flow channel for air floating Active Scope Camera," *IEEE/RSJ IROS*, pp.442–449, 2018.
 [25] Y. Yamauchi, et al.: "A robotic thruster that can handle hairy flexible cable of serpentine robots for disaster inspection," *IEEE/ASME AIM*, pp.107–113, 2018.
 [26] T.-K. Le, et al.: "Closed-form and near closed-form solutions for tdoa-based joint source and sensor localization," *IEEE TSP*, vol.65, no.5, pp.1207–1221, 2017.
 [27] S. Thrun: *Probabilistic robotics*. MIT Press, 2005.
 [28] M. Parviainen, et al.: "Self-localization of dynamic user-worn

- microphones from observed speech,” *Applied Acoustics*, vol. 117, pp.76–85, 2017.
- [29] J.L. Odom, et al.: “Passive towed array shape estimation using heading and acoustic data,” *IEEE Journal of Oceanic Engineering*, vol.40, no.2, pp.465–474, 2015.
- [30] J.J. Smith, et al.: “The cramer-rao lower bound for towed array shape estimation with a single source,” *IEEE TSP*, vol.44, no.4, pp.1033–1036, 1996.
- [31] Y. Rockah, et al.: “Array shape calibration using sources in unknown locations–Part I: Far-field sources,” *IEEE TASLP*, vol.35, no.3, pp.286–299, 1987.
- [32] N. Ono: “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *IEEE WAS-PAA*, pp.189–192, 2011.
- [33] S. Araki, et al.: “Spatial correlation model based observation vector clustering and MVDR beamforming for meeting recognition,” *IEEE ICASSP*, pp.385–389, 2016.
- [34] A. Deleforge, et al.: “Phase-optimized K-SVD for signal extraction from underdetermined multichannel sparse mixtures,” *IEEE ICASSP*, pp.355–359, 2015.
- [35] D. Kitamura, et al.: “Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model,” *IEEE ICASSP*, pp.276–280, 2015.
- [36] T. Otsuka, et al.: “Bayesian nonparametrics for microphone array processing,” *IEEE/ACM TASLP*, vol.22, no.2, pp.493–504, 2014.
- [37] K. Nakadai, 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, pp.739–761, 2011.
- [38] N. Murata, et al.: “Reverberation-robust underdetermined source separation with non-negative tensor double deconvolution,” *EUSIPCO*, pp.1648–1652, 2016.
- [39] H. Chiba, et al.: “Amplitude-based speech enhancement with nonnegative matrix factorization for asynchronous distributed recording,” *IWAENC*, pp.203–207, 2014.
- [40] G. Ince, et al.: “Assessment of general applicability of ego noise estimation,” *IEEE ICRA*, pp.3517–3522, 2011.
- [41] T. Tezuka, et al.: “Ego-motion noise suppression for robots based on semi-blind infinite non-negative matrix factorization,” *IEEE ICRA*, pp.6293–6298, 2014.
- [42] P. Huang, et al.: “Singing-voice separation from monaural recordings using robust principal component analysis,” *IEEE ICASSP*, pp.57–60, 2012.
- [43] Y. Li, et al.: “Speech enhancement based on robust NMF solved by alternating direction method of multipliers,” *IEEE MMSP*, pp.1–5, 2015.
- [44] L. Zhang, et al.: “Robust non-negative matrix factorization,” *Frontiers of Electrical and Electronic Engineering in China*, vol.6, no.2, pp.192–200, 2011.
- [45] C. Knapp, et al.: “The generalized correlation method for estimation of time delay,” *IEEE TASSP*, vol.24, no.4, pp.320–327, 1976.
- [46] C. Zhang, et al.: “Why does PHAT work well in lownoise, reverberative environments?” *IEEE ICASSP*, pp.2565–2568, 2008.
- [47] 坂東宜昭, ほか: “柔軟索状レスキューロボットのためのマイクロホン・加速度センサアレイを用いた3次元姿勢推定”, ロボティクス・メカトロニクス講演会, 1A2–10a6, 2016.
- [48] E.A. Wan, et al.: “The unscented Kalman filter for nonlinear estimation,” *IEEE AS-SPCC*, pp.153–158, 2000.
- [49] S.J. Julier: “The scaled unscented transformation,” *American Control Conference*, vol.6, IEEE, pp.4555–4559, 2002.
- [50] 坂東宜昭, ほか: “慣性・音響センサアレイを用いた柔軟索状レスキューロボットのマルチモーダル姿勢推定”, ロボティクス・メカトロニクス講演会, 2A1–M01, 2018.
- [51] M.D. Hoffman: “Poisson-uniform nonnegative matrix factorization,” *IEEE ICASSP*, pp.5361–5364, 2012.
- [52] S.D. Babacan, et al.: “Sparse Bayesian methods for low-rank matrix estimation,” *IEEE TSP*, vol.60, no.8, pp.3964–3977, 2012.
- [53] C.M. Bishop: *Pattern recognition and machine learning*. Springer, 2006.
- [54] D. Gamerman, et al.: “A non-gaussian family of state-space models with exact marginal likelihood,” *Journal of Time Series Analysis*, vol.34, no.6, pp.625–645, 2013.
- [55] 坂東宜昭, ほか: “半教師ありロバスト非負値テンソル分解による柔軟索状レスキューロボットの音声強調”, 第36回日本ロボット学会学術講演会予稿集 DVD, 2J2–06, 2018.
- [56] Y. Bando, et al.: “Statistical speech enhancement based on probabilistic integration of variational autoencoder and non-negative matrix factorization,” *IEEE ICASSP*, pp.716–720, 2018.
- [57] 坂東宜昭, ほか: “柔軟索状レスキューロボットのための空気噴射音下での単チャンネル音声強調”, ロボティクス・メカトロニクス講演会, 2A2–D07, 2019.

坂東宜昭 (Yoshiaki Bando)

2018年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。同年産業技術総合研究所 研究員となり、現在に至る。2015年日本学術振興会特別研究員(DC1)。極限音響信号処理、深層ベイズ信号処理の研究に従事。情報処理学会、日本機械学会、IEEE 各会員。(日本ロボット学会正会員)

安部祐一 (Yuichi Ambe)

2016年京都大学大学院工学研究科博士後期課程研究指導認定退学。博士(工学)。同年東北大学大学院情報科学研究科研究員、2017年同特任助教(研究)、2018年同工学研究科助教となり、現在に至る。歩行ロボット、索状ロボットの運動制御の研究に従事。計測自動制御学会、日本機械学会、IEEE 各会員。

(日本ロボット学会正会員)

糸山克寿 (Katsutoshi Itoyama)

2011年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。京都大学助教を経て、2018年より東京工業大学工学院システム制御系特任講師となり現在に至る。統計的音響信号処理や機械学習に基づく音楽情報処理・ロボット聴覚・音環境理解の研究に従事。情報処理学会、日本音響学会、IEEE 各会員。

奥乃 博 (Hiroshi G. Okuno)

1972年東京大学教養学部基礎科学科卒業。博士(工学)。NTT研究所、ERATO北野プロジェクト、東京理科大学、京都大学を経て、2014年より、早稲田大学理工学術院教授(任期付)。京都大学名誉教授、インド・アミティ大学名誉教授。人工知能、音環境理解、ロボット聴覚等の研究に従事。IEEE、日本人工知能学会、情報処理学会各フェロー。(日本ロボット学会正会員・フェロー)