

ロバスト主成分分析を用いた動作雑音抑圧に基づく 柔軟索状ロボットのための音声強調

○坂東 宜昭¹ 糸山 克寿¹ 昆陽雅司² 田所 諭² 中臺 一博³ 吉井 和佳¹ 奥乃 博⁴

¹ 京都大学 大学院情報学研究科 ² 東北大学 大学院情報科学研究科

³ 東京工業大学 大学院情報理工学研究科, (株) ホンダ RI-JP ⁴ 早稲田大学 理工学術院

1. はじめに

柔軟索状ロボット [1, 2] は細長い形状が特徴のレスキューロボットで、瓦礫の隙間に挿入し被災者を捜索するために開発されている。例えば、小型車輪で駆動する Active-Hose II [1] や、織毛の振動で駆動する Active Scope Camera (ASC) [2] などが報告されている。

柔軟索状ロボットでは、被災者の声を聞くために自身の動作雑音の抑圧が不可欠である。より広い範囲を限られた時間で探索するためにロボットは駆動し続ける必要があるが、従来は声を聞くために定期的にアクチュエータを静止させる必要があった。

柔軟索状ロボットの動作雑音は、周期的に変化し接地面の材質・形状に依存する。また、雑音はロボット自身から、目的の音声は遠方から到来するので低信号対雑音比 (低 SNR) での頑健な動作が不可欠である。ロボットの動作雑音を推定し抑圧する従来法 [3-6] では、事前学習が必要である、雑音の定常性を仮定するため性能が劣化するという問題があった。低 SNR での音声強調や音源分離は、マイクロホンアレイを用いた手法が広く研究されている [7, 8]。しかし、マイクロホンアレイ信号処理ではマイクロホンの配置情報が必要で柔軟に形状が変化する本ロボットには不適、計算コストが大きく実時間処理が困難といった問題がある。

Robust principal component analysis (RPCA) [9] を用いた雑音抑圧が柔軟索状ロボットでは有望である。RPCA は、入力行列を低ランク成分とスパース成分へ分解する手法である。RPCA を音響信号へ適用すると、事前情報を用いずに、動作雑音などの高頻度に現れる周波数成分と、音声などの低頻度に顕れる周波数成分へ分離することができる [10]。さらに、雑音を複数の基底で表現するので周期的に変化する雑音も対応できる。

本研究では、低 SNR 環境での頑健性を向上させるために、ロボットに装着したマイクロホンアレイ (図 1) へ RPCA を適用する手法を開発する。目的の音声は各マイクロホンに共通して観測され、動作雑音はマイクロホンごとに異なって観測される点に注目し、1) まず各マイクロホンごとに online RPCA (ORPCA) [11] を適用し、2) 各マイクロホンの抑圧結果について共通成分を抽出することで雑音を抑圧する。実際のロボットの走行雑音を用いた実験により、提案法を評価した。

2. 柔軟索状ロボットの動作雑音抑圧

提案法では、ロボット上の複数のマイクロホンへ個別に ORPCA [11] で雑音を抑圧し、各チャンネルの抑圧結果から共通成分を中央値で抽出する (図 2)。

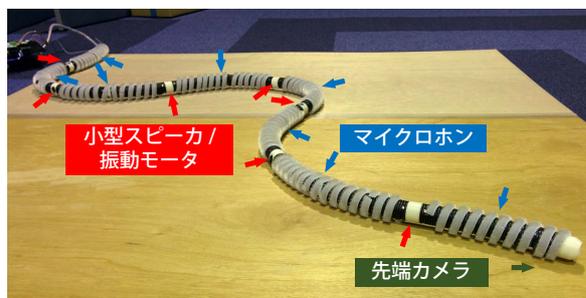


図 1: 8 チャンネルのマイクロホンアレイを搭載した全長 3m のプロトタイプ・柔軟索状ロボット。

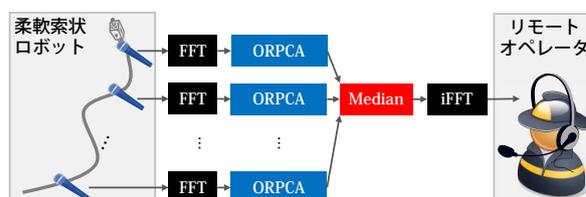


図 2: 動作雑音抑圧システムの概要

2.1 プロトタイプ：柔軟索状ロボット

図 1 に、マイクロホンアレイを搭載した柔軟索状ロボットの写真を示す。本体は、直径 38 mm のコルゲートチューブからなり、全長 3m である。8 つのマイクロホンアレイ ($M=8$) をロボット表面に 40 cm 間隔で 90 度ずつ回転して装着した。両端のマイクロホン間の距離は 2.8m である。マイクロホンは手元から順番にインデックス m で区別する ($m = 1, \dots, M$)。本ロボットは、Namari らの Tube-type Active Scope Camera [2] と同様、織毛と振動モータを用いた駆動で前進する。振動モータはロボット内に 40 cm 間隔で 7 つ直列に装着されている。

2.2 問題設定

本稿で扱う動作雑音抑圧の問題設定は以下の通り。

入力	M ch マイクロホンアレイで録音した音響信号の振幅スペクトル $\mathbf{y}_{1t}, \dots, \mathbf{y}_{Mt} \in \mathbb{R}^F$
出力	動作雑音を抑圧した振幅スペクトル $\mathbf{s}_t \in \mathbb{R}^F$

t, F, f をそれぞれ、時間フレームインデックス、周波数ビンの個数、周波数ビンインデックスとする。振幅スペクトル $\mathbf{y}_{tm} = [y_{tm1}, \dots, y_{tmF}]^T$ は録音信号を短時間フーリエ変換した振幅値とする。

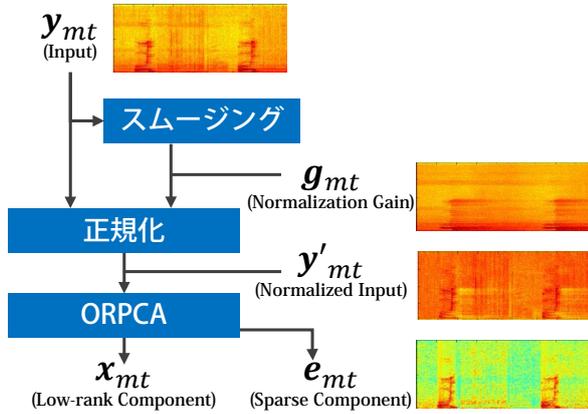


図3: 各チャンネルの入力スペクトルへのORPCA適用の流れ。

2.3 Online Robust PCA による雑音抑圧

各マイクロホンで観測された音響信号 \mathbf{y}_{mt} は、ORPCA を用いて低ランク成分 $\mathbf{x}_{mt} \in \mathbb{R}^F$ とスパース成分 $\mathbf{e}_{mt} \in \mathbb{R}^F$ に分解する (図3)。

$$\mathbf{y}_{mt} = \mathbf{x}_{mt} + \mathbf{e}_{mt}$$

周期的に変化する動作雑音は低ランク成分へ、低頻度で現れる音声や他のスパースなノイズはスパース成分へ分解される。

ORPCA では各周波数成分を同じ重みで学習するので、強いピーク周波数を持つ本ロボットの雑音に対して過学習を起こすことがある。そこで、入力スペクトルの各周波数成分 y_{mtf} を正規化係数 $g_{mtf} \in \mathbb{R}$ を用いて正規化してから ORPCA を適用する。

$$y'_{mtf} = \frac{1}{g_{mtf}} y_{mtf} \quad (1)$$

本ロボットの動作雑音は環境に依存するので、正規化係数も走行面の材質・形状に依存する。提案法では動作雑音の平均スペクトルは頻繁に大きく変動しないと仮定し、以下で与えるように正規化係数 $\mathbf{g}_{mt} = [g_{mt1}, \dots, g_{mtF}]^T$ をオンラインで更新する。

$$\mathbf{g}_{mt} = \alpha \mathbf{y}_{mt} + (1 - \alpha) \mathbf{g}_{m(t-1)} \quad (2)$$

ここで、 α は学習率パラメータを表し、小さい値 (e.g. $1e-2$) を設定する。

2.4 Online Robust PCA

提案法はRPCAのオンライン拡張の一つであるORPCA [11] を使用する。ORPCAの処理はインデックス m に対して独立なので、本節ではインデックス m を省略する。入力、低ランク成分、スパース成分の列をそれぞれ、 $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_t]$ 、 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_t] \in \mathbb{R}^{F \times t}$ 、 $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_t] \in \mathbb{R}^{F \times t}$ としたとき、オフラインRPCAでは、以下を最適化する問題として定義される。

$$\min_{\mathbf{X}, \mathbf{E}} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{X} - \mathbf{E}\|_F^2 + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{E}\|_1 \right\} \quad (3)$$

ただし、 $\|\cdot\|_F$ 、 $\|\cdot\|_*$ 、 $\|\cdot\|_1$ および λ_1, λ_2 はそれぞれ Frobenius, 核, L1 ノルムとスケールパラメータを表す。本問題は、augmented Lagrangian multiplier (ALM) を

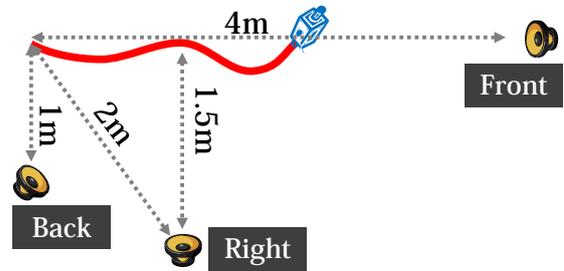


図4: 実験でのロボットと音源の配置図。3種類の音源配置 (front, right, back) で収録を行った。

用いて解かれる [9]。ALMでは、各イテレーションごとに入力の全ての要素を対象に一般化特異値分解を行うので、オンライン化が困難だった。

RPCAをオンラインで解くために、ORPCAでは式3を改変した以下の問題を最適化する。

$$\min_{\mathbf{L}, \mathbf{R}, \mathbf{E}} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{L}\mathbf{R}^T - \mathbf{E}\|_F^2 + \frac{\lambda_1}{2} (\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) + \lambda_2 \|\mathbf{E}\|_1 \right\}$$

ここで、 $\mathbf{L} \in \mathbb{R}^{F \times K}$ および $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_t] \in \mathbb{R}^{t \times K}$ はそれぞれ低ランク成分の基底とその係数を表す。 K は低ランク成分の基底数を表す。オンラインRPCAの問題のコスト関数は、式3の上限となっており、その局所解のうち1つがRPCAの大域解と等しいことが証明されている [11, 12]。さらに本式は、低ランク成分の係数 \mathbf{r}_t とスパース成分 \mathbf{e}_t について時刻に対し独立になるように式変形できる。

$$f_t(\mathbf{L}) = \frac{1}{t} \sum_{t'=1}^t l(\mathbf{y}_{t'}, \mathbf{L}) + \frac{\lambda_1}{2t} \|\mathbf{L}\|_F^2$$

$$l(\mathbf{y}_t, \mathbf{L}) = \min_{\mathbf{r}_t, \mathbf{e}_t} \left\{ \frac{1}{2} \|\mathbf{y}_t - \mathbf{L}\mathbf{r}_t - \mathbf{e}_t\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{r}_t\|_2^2 + \lambda_2 \|\mathbf{e}_t\|_1 \right\}$$

ここで、 $l(\mathbf{y}_t, \mathbf{L})$ は各時間インデックスでのコスト関数を表す。ORPCAでは、本コスト関数を各時刻ごとに基底 \mathbf{L} を固定して \mathbf{r}_t と \mathbf{e}_t を更新する。また、基底 \mathbf{L} は、各時刻ごとに以下のコスト関数 $g_t(\mathbf{L})$ を \mathbf{r}_t と \mathbf{e}_t を固定して更新する。

$$g_t(\mathbf{L}) = \frac{1}{t} \sum_{t'=1}^t l'(\mathbf{y}_{t'}, \mathbf{L}) + \frac{\lambda_1}{2t} \|\mathbf{L}\|_F^2$$

$$l'(\mathbf{y}_{t'}, \mathbf{L}) = \frac{1}{2} \|\mathbf{y}_{t'} - \mathbf{L}\mathbf{r}_{t'} - \mathbf{e}_{t'}\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{r}_{t'}\|_2^2 + \lambda_2 \|\mathbf{e}_{t'}\|_1$$

ここで、 $g_t(\mathbf{L})$ は $f_t(\mathbf{L})$ の上限となっている。これらの更新には、off-the-shelf solver および、block-coordinate descent を用いる [11]。

2.5 マイクロホンの統合

ORPCAで抽出された各マイクロホンのスパース成分は、共通成分を抽出することで統合される。各スパース成分には、マイクロホンが外界と接触したときの比較的大きな突発的雑音や、ORPCAで除けなかった動作雑音などの比較的小さな雑音成分が含まれている。これらのノイズに頑健に各マイクロホンを統合するために、提案法では中央値を各周波数ビンごと取る。出力信号 $\mathbf{s}_t = [s_{t1}, \dots, s_{tF}]^T$ は、以下で与えられる。

$$\mathbf{s}_{tf} = \text{Median}(e_{1tf}, \dots, e_{Mtf}) \quad (4)$$

ここで、Median(\dots) は引数の中央値を返す手続きである。提案する雑音抑圧アルゴリズム全体の手続きを Algorithm 1 に示す。

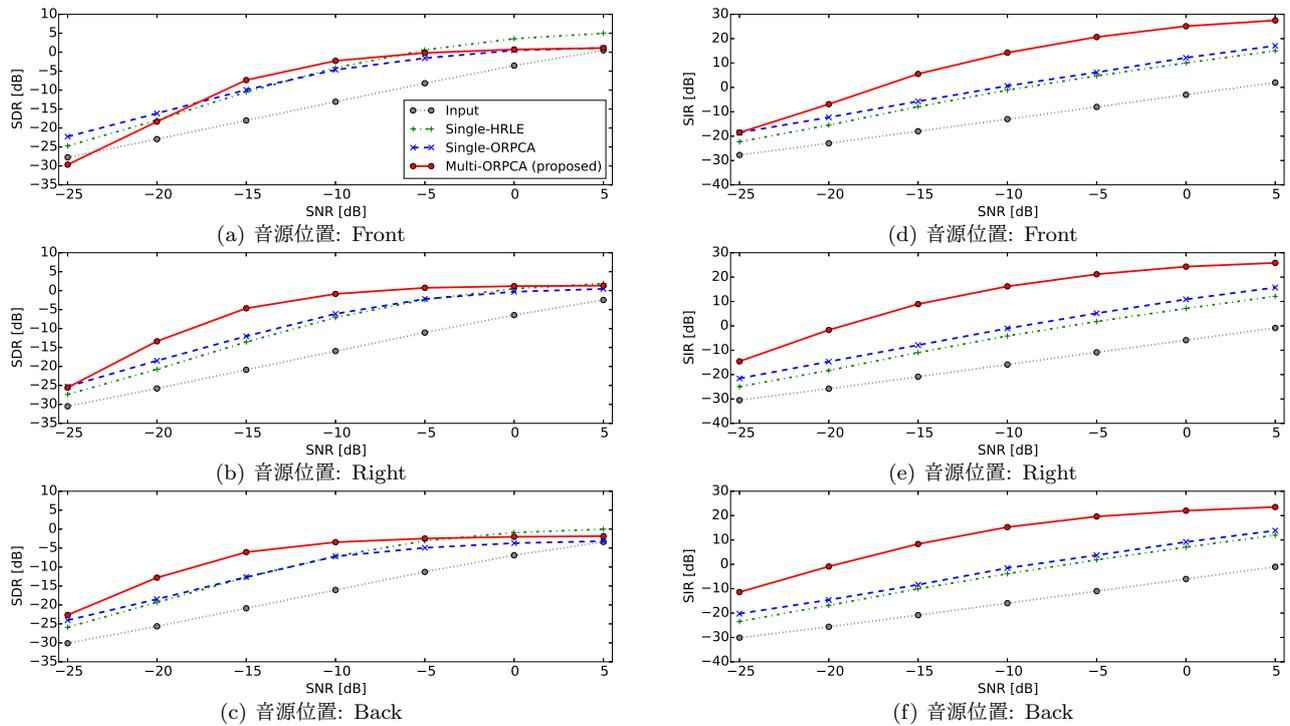


図 5: 先端マイクロホンの入力音響信号 (Input) および、雑音抑圧結果の SDR と SIR

Algorithm 1 動作雑音抑圧法の手続き

```

for  $t = 1, 2, 3 \dots$  do
  入力スペクトル  $\mathbf{y}_{mt}$  ( $m = 1, \dots, M$ ) を観測する
  for  $m = 1, \dots, M$  do
    a) 正規化係数  $\mathbf{g}_{mt}$  を更新する (式 (2))
    b) 入力  $\mathbf{y}_{mt}$  を正規化し  $\mathbf{y}'_{mt}$  を得る (式 (1))
    c)  $\mathbf{y}'_{mt}$  に ORPCA を適用しスパース成分  $\mathbf{e}'_{mt}$  を得る
    d) スパース成分  $\mathbf{e}'_{mt}$  に  $\mathbf{g}_{mt}$  を乗じ  $\mathbf{e}_{mt}$  を得る
  end for
  for  $f = 1, \dots, F$  do
     $[e_{1tf}, \dots, e_{Mtf}]$  について中央値を取り  $s_{tf}$  を得る
  end for
end for

```

3. 実験

駆動機構とマイクロホンアレイを有する柔軟索状ロボットの動作雑音を用いて提案法の雑音抑圧性能を評価した。

3.1 実験設定

柔軟索状ロボットを用いて音声と動作雑音を個別に録音し、SNR を -25 dB から 5 dB まで $+5$ dB ずつ変化させて混合し、音声の分離精度を評価した。図 4 に示すように、ロボットと音声を再生するスピーカを配置した。ロボットを駆動させ、手を用いて左右にロボット振りながら、60 秒の動作雑音を録音した。音源の配置は、1) front, 2) right, 3) back の 3 種類を評価した。目的音である音声は、ノイズを軽減するために、ロボットが静止時のインパルス応答に 60 秒の男性の音声を畳み込んで作成した。録音はロボット聴覚ソフトウェア HARK (Honda Research Institute Japan Audition

for Robots with Kyoto Univ.) [7] を用いて、8 ch 同期、16 bit 量子化、16 kHz サンプリングで行った。

本実験では次の 3 手法の比較を行った：

1. Multi-ORPCA(提案法): 全てのマイクロホンに ORPCA を適用し Median で統合した場合
2. Single-ORPCA: 先端のマイクロホンのみに ORPCA を適用した場合
3. Single-HRLE: 先端のマイクロホンのみに HRLE [6] を適用した場合

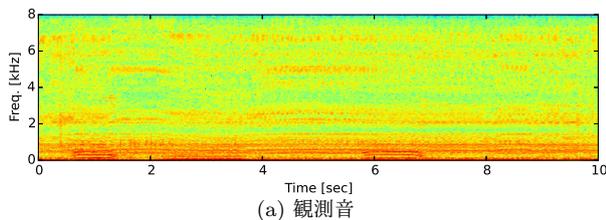
HRLE は、Single-ORPCA と同じ雑音抑圧問題を解く従来法の 1 つである。提案法のパラメータは実験的に決定した。HRLE は HARK のノードとして公開されており、パラメータは HARK のデフォルト値を使用した。

評価尺度には、信号対歪比 (SDR) と信号対妨害音比 (SIR) を用いた。SDR は総合的な分離精度を表し、SIR は抑圧音に含まれる動作雑音の度合いを表す。SDR と SIR の計算には、Python ツールキットの MIREVAL [13] を用いた。

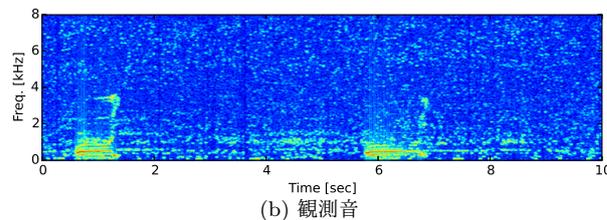
3.2 実験結果

図 5 に示すように、音源位置が front のとき、SNR が -15 dB と -10 dB において提案法が最も性能が高い。さらに、音源位置が right, back のとき、 -20 dB から -5 dB の間で最も性能が高い。また、音源位置が front かつ SNR が -25 dB の場合を除いて、SIR は最低でも 5.4 dB 以上向上している。

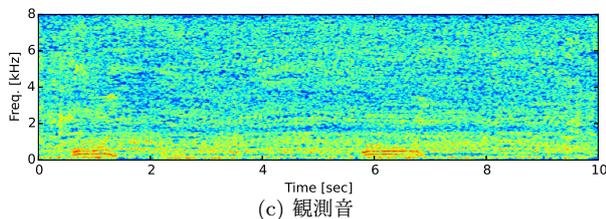
SIR は抑圧音中の雑音の少なさを表しており、図 6 のスペクトログラムが示すように、提案法が最も動作雑音を強く抑圧できている。また、図 6-(a) が示すように、本ロボットの動作雑音は約 10 Hz の周期で変動している。Single-HRLE による抑圧音では、この変動により雑音が抑圧音に縦縞として、強く残っている。一



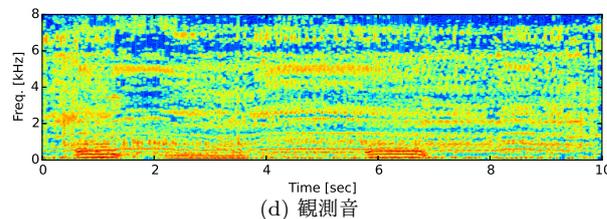
(a) 観測音



(b) 観測音



(c) 観測音



(d) 観測音

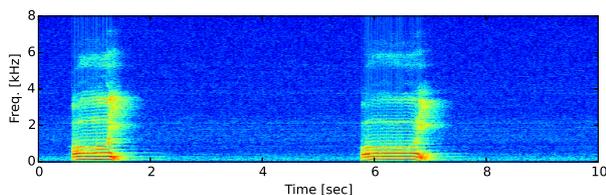
図 6: 先端のマイクロホン ($m = 8$) での観測信号および、各種法による抑圧音のスペクトログラム。

図 7: 目的音のスペクトログラムの一例

方で、Single-ORPCA と Multi-ORPCA による結果では、より抑圧されている。

音源の位置が front のとき、他の配置に比べて SDR と SIR が低い。これは、提案法の性能が音源とロボットの位置関係に依存することを表している。各マイクロホンで観測される目的音の大きさはその位置に依存して変動するが、提案法は全てのマイクロホンを同じ重みで統合しているためだと考えられる。本問題は、各マイクロホンの抑圧音について SNR を推定し、その重みでマイクロホンを統合することで改善が期待できる。

4. おわりに

本稿では、柔軟索状ロボットで収録した音響信号の音声強調のために、ORPCA をマイクロホンアレイに適用する雑音抑圧法を開発した。提案法は、まず各チャンネルにそれぞれ ORPCA を適用し、各チャンネルの共通成分を抽出するために中央値を取って統合する。駆動機構を有しマイクロホンアレイを搭載した 3m のプロトタイプ・柔軟索状ロボットを用いてその有効性を評価した。実際に録音した柔軟索状ロボットの動作雑音を用いたシミュレーション混合による実験で、マイクを一つのみ使用する従来法と比較して、提案法は SDR が最大 7.4dB、SIR が最大 17.2dB 向上することを確認した。また、性能が音源とロボットの位置関係に依存することが分かった。

今後は、音源位置や各チャンネルの抑圧音の SNR 推定の利用による性能向上を目指す。また、提案法は雑音成分の繰り返しに注目して抑圧を行うので、報道ヘリコプターや土木作業などの環境音の抑圧への応用を検討する。さらに、有効性実証のため、模擬瓦礫環境での評価、実時間処理システムの構築・評価などを行う。

謝辞 本研究は、科研費基盤 (S) No.24220006、特別研究員奨励費 No. 15J08765、および 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] A. Kitagawa *et al.* Development of small diameter Active Hose-II for search and life-prolongation of victims under debris. *JRM*, 15(5):474–481, 2003.
- [2] H. Namari *et al.* Tube-type active scope camera with high mobility and practical functionality. In *IROS-2012*, 3679–3686.
- [3] G. Ince *et al.* Incremental learning for ego noise estimation of a robot. In *IROS-2011*, 131–136.
- [4] T. Tezuka *et al.* Ego-motion noise suppression for robots based on semi-blind infinite non-negative matrix factorization. In *ICRA-2014*, 6293–6298.
- [5] B. Cauchi *et al.* Reduction of non-stationary noise for a robotic living assistant using sparse non-negative matrix factorization. In *SMIAE*, 28–33, 2012.
- [6] H. Nakajima *et al.* An easily-configurable robot audition system using histogram-based recursive level estimation. In *IROS-2010*, 958–963.
- [7] K. Nakadai *et al.* Design and implementation of robot audition system HARK – open source software for listening to three simultaneous speakers. *Advanced Robotics*, 24(5-6):739–761, 2010.
- [8] J. Taghia *et al.* A variational Bayes approach to the underdetermined blind source separation with automatic determination of the number of sources. In *ICASSP*, 253–256, 2012.
- [9] E. J. Candès *et al.* Robust principal component analysis? *JACM*, 58(3):11, 2011.
- [10] C. Sun *et al.* Noise reduction based on robust principal component analysis. *JCIS*, 10(10):4403–4410, 2014.
- [11] J. Feng *et al.* Online robust PCA via stochastic optimization. In *NIPS*, 404–412, 2013.
- [12] B. Recht *et al.* Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.
- [13] C. Raffel *et al.* mir.eval: A transparent implementation of common MIR metrics. In *ISMIR-2014*, 367–372.