

独立ベクトル分析とノイズキャンセラを用いた雑音抑圧の 柔軟索状ロボットへの適用

Applying independent vector analysis and noise cancellation
to noise reduction for a hose-shaped rescue robot

○石村大 (筑波大) 牧野昭二 (筑波大)
山田武志 (筑波大) 小野順貴 (総研大/国情研)
猿渡洋 (東大)

Masaru ISHIMURA, University of Tsukuba
Shoji MAKINO, University of Tsukuba
Takeshi YAMADA, University of Tsukuba
Nobutaka ONO, SOKENDAI/NII
Hiroshi SARUWATARI, The University of Tokyo

This paper presents a noise reduction on a hose-shaped rescue robot. The hose-shaped rescue robot is one of rescue robots developed on Tough Robotics Challenge, and it is used for searching for victims by getting one's voice with its microphone-array. However, the ego noise, caused by its vibration motors, makes it difficult to get the human voice. We propose a noise reduction method using a blind source separation technique based on Independent Vector Analysis (IVA) and noise cancellation. Our method consists of two steps: (1) estimating a speech signal and an ego-noise signal from observed multi-channel signals using the IVA-based blind source separation technique, and (2) applying the noise cancellation to the estimated speech signal using the estimated ego-noise signal as a noise reference.

Key Words: Rescue robot, Tough environment, Noise reduction, Independent vector analysis, Noise cancellation

1 はじめに

近年、地震等の自然災害発生時に、効率的に災害状況の調査を行うための遠隔操作可能なロボットの開発が、ImPACT 革新的研究開発推進プログラム「タフ・ロボティクスチャレンジ」[1]において推進されている。柔軟索状ロボット [2] とは、このプログラムの一貫で開発されているレスキューロボットの一つである。柔軟索状ロボットは、ヘビ型の細長い形状をしており、従来の遠隔操作ロボットでは探索できない狭い空間等の調査が可能である。柔軟索状ロボットの利用例の一つとして、機体の周囲に一定間隔で取り付けられたマイクロホンで人の声を捉えることによる被災者探索が挙げられる。しかし、柔軟索状ロボットは機体周囲に巻かれた繊維テープを振動モーターによって振動させることで前進するため、機体自身から発生したノイズが被災者の声とともにマイクロホンに混入してしまう。これにより、被災者の声をとらえることが困難になるという問題がある。本稿ではこのようなノイズを「エゴノイズ」と呼ぶ。

したがって本稿では、柔軟索状ロボットに取り付けられたマイクロホンを用いて災害現場で救助を求める人の声をとらえるために、柔軟索状ロボットによる観測音に含まれるエゴノイズを抑圧することを目的とする。柔軟索状ロボットには複数のマイクロホンが取り付けられているが、機体は形状を変えつつ移動するため、ビームフォーミング [3] のような固定マイクロホンアレイを前提とする雑音抑圧は困難である。そこで、音源の統計的独立性に基づく音源分離手法 [4]、特に、補助関数型ベクトル分析 (Independent Vector Analysis: IVA) に基づく音源分離手法 [5]、さらにその後段にノイズキャンセラ [6] による後処理を加えた柔軟索状ロボットにおける雑音抑圧手法を提案する。また、柔軟索状ロボットの観測音を再現したシミュレーション実験を行うことで、提案手法による雑音抑圧の効果を確認する。



Fig.1: 柔軟索状ロボットの概要

2 柔軟索状ロボット

2.1 柔軟索状ロボットの概要

図1に柔軟索状ロボットの概要を示す。柔軟索状ロボットは基本的に軸となるホース、ホースの周囲にまかれた繊維テープ、機体を振動させる振動モーターから構成される。索状ロボットは機体に巻き付けられた繊維毛と振動モーターにより前進する。また、

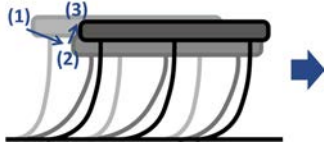


Fig.2: 柔軟索状ロボットの動作原理 [2]

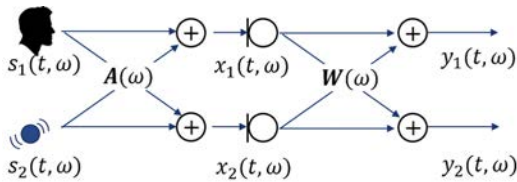


Fig.3: 音源の統計的独立性に基づく音源分離

マイクロホン・カメラ・ガスセンサ等の様々なセンサを取り付け各種センシングを行うことを可能とする。

また、柔軟索状ロボットの動作原理を図2に示す。図2は図1に示したロボットと床との接地面を模式的に表したものである。全ての繊維は機体が進む方向とは反対の方向に傾斜している。この状態で機体を振動モーターにより振動させると、(1)の状態から繊維と床との摩擦により(2)の状態へ、さらに繊維が滑ることによって(3)の状態へと遷移する。索状ロボットはこの動作を繰り返すことで前進する。

2.2 柔軟索状ロボットにおける音声収録上の問題

索状ロボットでは、2.1節で示した動作原理から、次のように人の声の検出・聞き分けを行う際に問題が生じる。

- 振動モーターの駆動音。
- 繊維と接地面との摩擦音。
- マイクロホン自体の振動による雑音。

本稿では、これらの雑音をまとめて「エゴノイズ」と呼び、IVAに基づく音源分離とノイズキャンセラによりこのエゴノイズを抑圧する手法について検討する。

3 提案法

本稿で対象とする柔軟索状ロボットは複数のマイクロホンを持つ。そこで従来から研究されているマルチチャネル音源分離手法の適用を試みる。マルチチャネル音源分離手法としては、ビームフォーマや音源の統計的独立性に基づく手法が挙げられるが、前者はマイクロホンの位置情報を必要とする。そこで、本稿では音源の統計的独立性に基づく手法、特に Independent Vector Analysis (IVA) に基づく音源分離手法を利用する。

但し、IVAに基づく音源分離は独立性という高次の統計量を規範とするため、統計量の分析にある程度の長さの信号長を必要とする。したがって、推定される分離フィルタは時不変なフィルタとなりその仮定に従わない成分の残留等、分離しきれない雑音が生じる可能性が考えられる。そこで本研究では、IVAの後段に時変のノイズキャンセラを適用し、時不変フィルタでキャンセルしきれない残留雑音を除去することを提案する。通常のノイズキャンセラにおけるノイズ観測用の参照マイクの代わりに、本研究ではIVAをノイズ推定器として用いる。

以下では、まずIVAに基づく音源分離、ノイズキャンセラについて説明し、次に提案法について述べる。

3.1 IVAに基づく音源分離

IVAに基づく音源分離手法として、今回利用する補助関数型独立ベクトル分析によるブラインド音源分離 [5] について説明する。例として、 K 個の音源が K 個のマイクロホンにより観測される状況を想定する。音源信号、観測信号の短時間フーリエ変換表現をそれぞれ $s_k(\omega)$, $x_k(\omega)$ とする。但し、 k は音源とマイクロホンのインデックス、 ω は周波数ビンのインデックスを表し、時

間フレームのインデックスは省略する。このとき混合モデルは次のように表せる。

$$\mathbf{x}(\omega) = \mathbf{A}(\omega)\mathbf{s}(\omega) \quad (1)$$

また推定音源信号 $\mathbf{y}(\omega)$ は次式で表現できる。

$$\mathbf{y}(\omega) = \mathbf{W}(\omega)\mathbf{x}(\omega) \quad (2)$$

ここで、

$$\mathbf{A}(\omega) = (\mathbf{a}_1(\omega), \mathbf{a}_2(\omega), \dots, \mathbf{a}_K(\omega))^h \quad (3)$$

は混合行列、

$$\mathbf{W}(\omega) = (\mathbf{w}_1(\omega), \mathbf{w}_2(\omega), \dots, \mathbf{w}_K(\omega))^h \quad (4)$$

は分離行列を表す。また h はエルミート転置、 $\mathbf{s}(\omega)$, $\mathbf{x}(\omega)$, $\mathbf{y}(\omega)$ は以下のように定義する。

$$\mathbf{s}(\omega) = (s_1(\omega), \dots, s_K(\omega))^t \quad (5)$$

$$\mathbf{x}(\omega) = (x_1(\omega), \dots, x_K(\omega))^t \quad (6)$$

$$\mathbf{y}(\omega) = (y_1(\omega), \dots, y_K(\omega))^t \quad (7)$$

ここで、 t は転置を表す。以上のモデル化により、IVAに基づく音源分離は、各音源が独立であるという仮定のもと、 $\mathbf{y}(\omega)$ の独立性が最大となるような分離行列 $\mathbf{W}(\omega)$ を見つけることによつて行われる。独立性の指標としては、非ガウス性、相互情報量、尤度などが用いられるが、いずれの場合も以下の形の目的関数の最小化問題に帰着することが知られている [4]。

$$J(\mathbf{W}) = \sum_{k=1}^K E[G(\mathbf{y}_k)] - \sum_{\omega=1}^{N_\omega} \log |\det \mathbf{W}(\omega)| \quad (8)$$

ここで、 $G(\mathbf{y}_k)$ はコントラスト関数と呼ばれ、 $p(\mathbf{y}_k)$ を \mathbf{y}_k が従う確率密度関数としたとき、 $G(\mathbf{y}_k) = -\log p(\mathbf{y}_k)$ である。今回は、音源信号の事前分布を球状ラプラス分布と仮定し $G(\mathbf{y}_k) = \|\mathbf{y}_k\|_2$ となる。導出は省略するが、(8) を最小化する \mathbf{W} を求めるアルゴリズムは補助関数法により以下のようになる [5]。

- 重み付き共分散行列の更新

$$r_k = \sqrt{\sum_{\omega=1}^{N_\omega} |\mathbf{w}_k^h(\omega)\mathbf{x}(\omega)|^2} \quad (9)$$

$$\mathbf{V}_k(\omega) = E \left[\frac{G'(r_k)}{r_k} \mathbf{x}(\omega)\mathbf{x}^h(\omega) \right] \quad (10)$$

- 分離行列の更新

$$\mathbf{w}_k(\omega) \leftarrow (\mathbf{W}(\omega)\mathbf{V}_k(\omega))^{-1} \mathbf{e}_k \quad (11)$$

$$\mathbf{w}_k(\omega) \leftarrow \mathbf{w}_k(\omega) / \sqrt{\mathbf{w}_k^h(\omega)\mathbf{V}_k(\omega)\mathbf{w}_k(\omega)} \quad (12)$$

ここで、 N_ω は周波数ビンの数を表し、 \mathbf{e}_k は k 番目の要素が1で他の要素が0の列ベクトルを表す。

3.2 ノイズキャンセラ

ノイズキャンセラでは図4に示すように、音声を収録するためのマイクロホンの他に、雑音源付近に設置することで雑音のみを収録するリファレンスマイクロホンが必要とする。これにより、音声を収録すると同時に雑音のリファレンス信号を収録可能である状況を想定する。

雑音環境下で話者が発話しているとする、時刻 t における音声収録用マイクの入力信号 $x(t)$ は、音声 $s(t)$ と雑音 $n(t)$ が混ざった信号として次式で表される。

$$x(t) = s(t) + n(t) \quad (13)$$

一方、リファレンスマイクでは同時に雑音源由来の雑音が収録される。このとき、音声収録用マイクに混入する雑音 $n(t)$ はリファレンスマイク入力信号 $r(t)$ と高い相関があると考えられる。そこで、リファレンスマイク入力信号 $r(t)$ と音声収録用マイクに

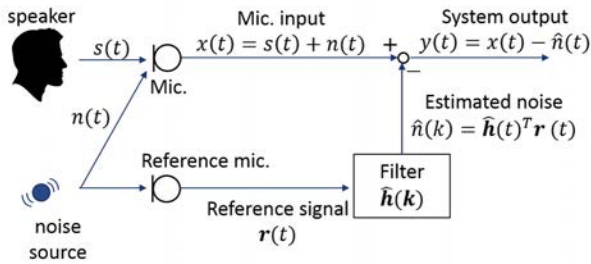


Fig.4: ノイズキャンセラの概要

混入した雑音 $n(t)$ との関係を下のようなインパルス応答の線形畳み込みで表現できると仮定する.

$$n(t) \approx \mathbf{h}(t)^T \mathbf{r}(t) \quad (14)$$

ここで、 $\mathbf{r}(t) = [r(t), r(t-1), \dots, r(t-N)]^T$ は現在の時刻 t から過去 N 時刻までのリファレンスマイク入力、 $\hat{\mathbf{h}}(t) = [\hat{h}(0), \hat{h}(1), \dots, \hat{h}(N)]^T$ はインパルス応答 (フィルタ) を表す.

以上のモデル化によれば、音声収録用マイクに混入した雑音 $n(t)$ を予測するフィルタ $\mathbf{h}(t)$ を推定できれば、推定した雑音 $\mathbf{h}(t)^T \mathbf{r}(t)$ を音声収録用マイク入力信号から引くことで、雑音の抑圧が可能となる. このように除去したい雑音と高い相関をもつリファレンス信号から雑音を推定し、抑圧を行う手法をノイズキャンセラと呼ぶ. フィルタ $\mathbf{h}(t)$ は平均二乗誤差最小化の枠組みによって推定することが可能である. フィルタを推定するアルゴリズムは様々であるが、今回は Normalized Least Mean Square (NLMS) アルゴリズム [6] を利用する. NLMS アルゴリズムによりフィルタ $\mathbf{h}(t)$ の更新は以下ようになる.

$$\mathbf{h}(t+1) = \mathbf{h}(t) + \mu \frac{\mathbf{y}(t)}{\|\mathbf{r}(t)\|^2} \mathbf{r}(t) \quad (15)$$

3.3 提案法の処理の流れ

まず、マルチチャネル音源分離手法である IVA に基づく周波数領域ブラインド音源分離手法により音声成分とエゴノイズ成分とを可能な限り分離する. 但し、IVA は音源の独立性という高次の統計量を規範とするため、フィルタ推定に数秒の分析長を必要とし、その分析時間においてフィルタは時不変であると仮定している. これらの仮定に従わず残留してしまう雑音を除去するために、IVA により分離された音声成分に対し、雑音成分をリファレンス信号とした時間領域ノイズキャンセラを適用する. ノイズキャンセラでは、さらに時間サンプルごとのフィルタ更新が可能であるため、IVA の時不変な仮定に従わず残留した雑音の除去が期待できる.

提案法の処理の流れを図.5 に示す. 図.5 において、 s, n, n_r, r はそれぞれ音声, エゴノイズ, IVA に基づく音源分離で残留したエゴノイズ, エゴノイズのリファレンスを表す. まず IVA により多チャネル入力信号をマイク数と同じ数の独立信号に分離する. 得られた分離信号に対し、音声が最も大きく含まれる信号を人が聞いて 1 つ選び、それ以外の分離信号を全て加算した信号を雑音のリファレンス信号としてノイズキャンセラを適用する.

4 評価実験

4.1 実験条件

入力 SN 比を調節するため、実験には雑音と音声を人工的に混合したものを用いた. 雑音としては、8 チャンルのマイクロホンと 7 つの振動モータを持つ全長 3 m の柔軟索状ロボットを実際に動作させて収録した動作音、音声としては、話者位置から 8 つのマイクロホンまでのインパルス応答をドライソースの音声と畳み込んだものを用いた.

評価尺度は、Source-to-distortion ratio (SDR)[7] を用いた. これは正答となる元の音源信号と雑音抑圧後の音源信号によって算出され、雑音抑圧後の音源信号の歪みを表し、値が大きいほど雑音抑圧性能が良いことを示す. その他の実験条件を表 1 にまとめる.

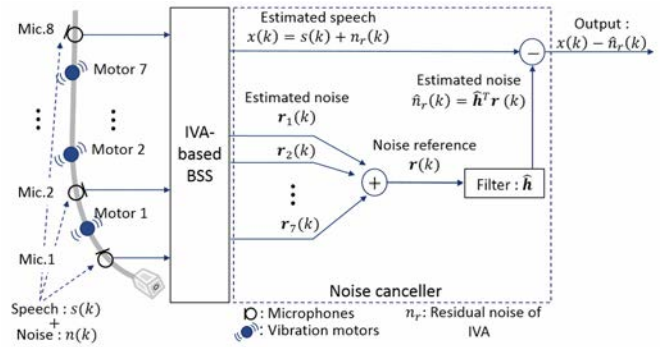


Fig.5: 提案法の処理の流れ

Table 1: 実験条件

サンプリング周波数	16 kHz
IVA の反復回数	100
分析フレーム長	1024 samples
分析フレームシフト幅	256 samples
ノイズキャンセラのフィルタ長	1600 taps
NLMS のステップサイズ	0.1
入力 SN 比	-10, -5, 0 dB

4.2 結果考察

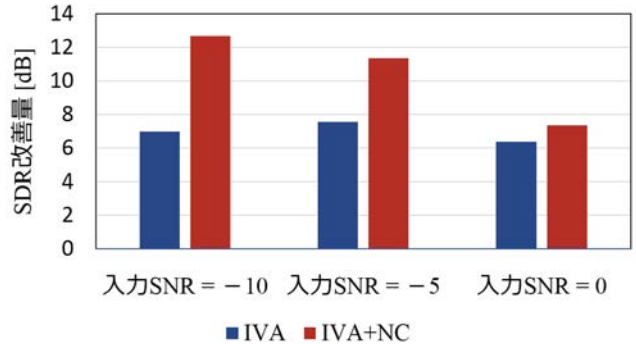


Fig.6: SDR 改善量

図 6 に IVA のみ、さらにノイズキャンセラによる後処理を適用した場合の SDR 改善量をそれぞれ示す. 図 6 から、IVA では入力 SN 比にかかわらず SDR で約 7 dB の改善が見られた. IVA の後処理としてノイズキャンセラを適用した提案法では、さらに SDR が改善することが確認できる.

但し、ノイズキャンセラの効果について、入力 SN 比が -10 dB, -5 dB の場合には約 4 dB の改善が得られる一方で、入力 SN 比が 0 dB の場合には 1 dB 程度の改善しか得られない. これは IVA に基づく音源分離による効果で音声の SN 比が改善することで、その音声にノイズキャンセラに対する外乱として作用していることが原因として考えられる.

5 おわりに

本稿では柔軟索状ロボットにおける音声強調を目的として、独立ベクトル分析に基づく音源分離とノイズキャンセラを用いた柔軟索状ロボットにおける雑音抑圧法を提案し、シミュレーション実験によりその効果を確認した. その結果 SDR において、独立ベクトル分析に基づく音源分離により約 7 dB の改善、さらにノイズキャンセラを用いた後処理により 1-4 dB 程度の改善が得られた.

謝辞

本研究は、総合科学技術・イノベーション会議により制度設計された革新的研究開発推進プログラム (ImPACT) により、科学技術振興機構を通して委託されたものである。実験データを提供して頂いた早稲田大学奥乃博教授と京都大学坂東宜昭氏に感謝の意を表す。

参考文献

- [1] ImPACT タフ・ロボティクス・チャレンジ,
<http://www.jst.go.jp/impact/program07.html>.
- [2] H. Namari, K. Wakana, M. Ishikura, M. Konyo, S. Tadokoro, "Tube-type active scope camera with high mobility and practical functionality," Proc. IEEE/RSJ International Conference on, Intelligent Robots and Systems (IROS), pp. 3679–3686, 2012.
- [3] HL. Van Trees, Optimum Array Processing (John Wiley & Sons, New York, 2002)
- [4] A. Hyvärinen, J. Karhunen and E. Oja, Independent Component Analysis (John Wiley & Sons, New York, 2001)
- [5] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," Proc. WASPAA, pp. 189–192, 2011.
- [6] E. Hansler, G. Schmidt, Acoustic echo and noise control: a practical approach (John Wiley & Sons, New York, 2004)
- [7] E. Vincent, R. Gribonval, C. Févotte, "Performance measurement in blind audio source separation," IEEE Trans. on Audio, Speech & Language Processing, vol.14, no. 4, pp. 1462–1469, 2006.