

独立低ランク行列分析と統計的音声強調を用いた 柔軟索状ロボットにおけるブラインド音源分離システムの開発*

☆三井祥幹, 溝口聡, 猿渡洋(東大), 越智景子(NII), 北村大地(総研大), 小野順貴(NII/総研大), 石村大, 前成美, 高草木萌, 松井裕太郎, 山岡洸瑛, 牧野昭二(筑波大)

1 はじめに

日本は世界でも有数の地震大国であり, 特に近年では, 瓦礫などの内部に進入して探索をすることが可能な柔軟索状ロボットの開発が進められている. 本ロボットには一定の間隔ごとにマイクロホンが複数個取り付けられており, 被災者の声などを検出する目的での利用が想定されるが, バイブレータの振動や地面との摩擦音などが観測音にノイズとして混入するため, 目的音成分が埋もれてしまう問題がある. このため, 録音された信号から雑音を抑圧し, 目的音を強調する技術が必要とされている.

本稿では, 多チャンネル音源分離である独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [1] およびノイズキャンセラ・統計的ポストフィルタ等を組み合わせることによって, 観測信号から目的音をオフライン処理により強調するシステムを提案する. また, 提案システムの有効性を客観評価指標や主観評価実験結果により示す.

2 柔軟索状ロボットと音響的問題

柔軟索状ロボットは災害救助を支援する目的で開発されており, 外観は Fig. 1 に示す通りである. ロボットには複数のバイブレータが搭載されており, これらを振動させると, 表面に巻きつけられた特殊な繊毛の効果によってロボットは前進する.

瓦礫内部などの状態を把握するため, ロボット本体には複数のマイクロホンなど各種センサが取り付けられている. しかし, マイクロホンではバイブレータの振動や地面との摩擦音などがエゴノイズとして観測に混入するため, 信号処理により雑音を抑圧する必要がある. 加えて, ロボットに取り付けられているマイクロホンは, ロボット本体の形状変化により位置関係が時間的に変化するため, 従来のアレイ信号処理による音源強調手法を利用することは困難である.

3 提案システムの概要

3.1 索状ロボットにおけるマイクロホン配置

今回利用した索状ロボットは, Fig. 2 に示すように 8 個のマイクロホンが 2 個ずつロボットの節部に配置されている. マイクロホン同士の間隔は 600 mm である. 各節において, 2 個のマイクロホンはそれぞれ反対方向を向けて配置されており, 索状ロボットのねじれに関わらずいずれか一方のマイクロホンが上向きとなるため, 目的音成分の SN 比を向上できる.

3.2 信号処理の方針

本稿では, オペレータが強調したい目的音を切り出すことが可能である状況を想定し, 5 秒程度のごく短い信号をバッチ処理によって強調するためのシステムを提案する. この想定より, 以下を仮定できる. (1) マイクロホンでの録音中に, 空間伝達系の大きな



Fig. 1 Overview of hose-shaped rescue robot.

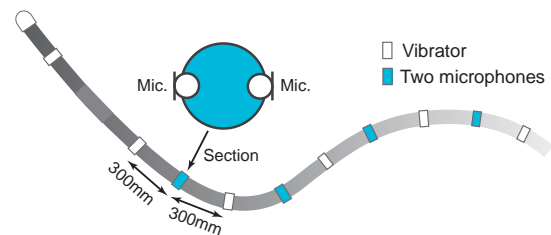


Fig. 2 Position of microphones and vibrators.

変動がない. (2) エゴノイズのパワーは概ね定常である. (3) 目的音が存在せず雑音のみの区間が存在する.

仮定 (1) より, ブラインド音源分離手法である ILRMA を利用した目的音の抽出が可能となる. また, ILRMA で捉えきれなかった細かな伝達系の変動を, ノイズキャンセラにより補償する. これらの処理後, 仮定 (2)(3) より, 統計的ポストフィルタを利用して更に雑音を抑圧できる. ここで, ポストフィルタの非線形処理に起因してミュージカルノイズが発生するため, ポストフィルタのゲイン関数に対して平滑化処理を施すことにより, ミュージカルノイズの抑圧を図る. 全体的な信号処理の流れを Fig. 3 に示す.

3.3 ILRMA

ILRMA [1] は近年提案された優決定条件における多チャンネルブラインド音源分離手法であり, 線形時不変な空間フィルタを推定することによって音源分離を実現する. ILRMA では, ランク 1 空間モデル [2]・音源間の統計的独立性・各音源スペクトログラムの低ランク性 (NMF モデル) を仮定し, 空間フィルタと音源 NMF モデルを同時に推定する.

3.4 ノイズキャンセラ

ノイズキャンセラは, 目的音を収録するためのマイクロホンの他に, 雑音のみを収録するためのリファレンスマイクロホンを用意し, 平均二乗誤差最小化アルゴリズムを利用して雑音抑圧を行う手法である. ILRMA では線形時不変な空間伝達系を仮定したが, 索状ロボットは常に振動・変形するため, 空間伝達系は時間的に細かく変動しており, ノイズキャンセラによってこれを補償できる. 本システムにおいては, ILRMA の出力を「目的音成分が含まれるチャンネル」と「それ以外のチャンネルの総和 (雑音成分)」に分けることによって, 雑音成分をリファレンス信号として

*Development of blind source separation system for flexible hose-shaped robot using independent low-rank matrix analysis and statistical speech enhancement, by Yoshiaki MITSUI, Satoshi MIZOGUCHI, Hiroshi SARUWATARI (The University of Tokyo), Keiko OCHI (NII), Daichi KITAMURA (SOKENDAI), Nobutaka ONO (NII/SOKENDAI), Masaru ISHIMURA, Narumi MAE, Moe TAKAKUSAKI, Yutaro MATSUI, Kouei YAMAOKA, Shoji MAKINO (University of Tsukuba).

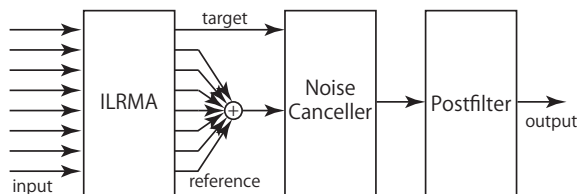


Fig. 3 Signal flow of our proposed method.

ノイズキャンセラを適用する。詳細なアルゴリズムは文献 [3]などを参照されたい。

3.5 ポストフィルタ

ロボットにおける録音信号は入力 SN 比が非常に低く、上述の 2 手法のみでは雑音を十分に取り除くことができない。ここで、エゴノイズが定常かつパワーが推定可能である場合は、統計的ポストフィルタを利用した更なる雑音抑圧が可能である。

MOSIE [4] はベイズ型ポストフィルタの一種であり、シングルチャンネルの信号に対し目的音の強調を行う。MOSIE では、時間周波数領域における振幅の事前分布として周波数ビンごとに個別のカイ分布を仮定する。カイ分布はスパース性を制御する形状母数 ρ_i をパラメータとして保有しており、ノイズパワーが推定可能な場合は、最適な ρ_i をノイズキャンセラの出力から計算できる [5]。ノイズキャンセラから出力された振幅スペクトログラムを X_{ij} とおくと、MOSIE を利用して最適なゲイン関数 G_{ij} を求めることにより、目的音成分を $G_{ij}X_{ij}$ として推定できる。この際、目的音成分が存在しない低域（概ね 300 Hz 以下）と高域（概ね 4000 Hz 以上）の帯域をカットしたほか、聴感品質を高めるために、median フィルタ及びゲイン平滑化 [6] を更に適用している。

3.6 システムの動作時間

本システムは全て MATLAB で実装されており、Intel Xeon E3-1505M (2.8 GHz) を搭載したノート PC を利用して処理時間を実測したところ、5 秒間の信号に対し約 55 秒の処理時間を必要とした。この処理時間のうち、約 52.5 秒が ILRMA による音源分離に要した時間であり、ノイズキャンセラおよびポストフィルタは合計して約 3 秒間程度の処理となっている。

4 評価実験

4.1 評価手法

本システムによる雑音抑圧の品質を評価するため、擬似的な観測信号を作成して実験を行った。目的音信号は、瓦礫中を模した環境で事前に測定した各マイクロホンのインパルス応答をドライソースに畳み込むことで作成し、雑音信号はロボットのバイブレータを実際に動作させて録音した。目的音と雑音の入力 SN 比は -25 dB から 0 dB まで 5 dB 刻みで変化させ、これら 2 種類の信号を時間領域で加算し、擬似的な観測信号とした。評価指標としては、音源分離の総合的性能を測定できる signal-to-distortion ratio (SDR) および、雑音抑圧量を測定できる signal-to-interference ratio (SIR) を利用した。評価は疑似観測信号に対して ILRMA を適用した信号 (ILRMA)・更にノイズキャンセラを適用した信号 (NC)・更にポストフィルタを適用した信号 (PF) のほか、処理を行わない疑似観測信号 (Raw) および補助関数法による独立ベクトル分析 [7] を適用した信号 (AuxIVA) の 5 種類について行った。

上述の客観評価指標の測定に加えて、入力 SN 比 -25 dB の疑似観測信号を分離して得られる信号を利

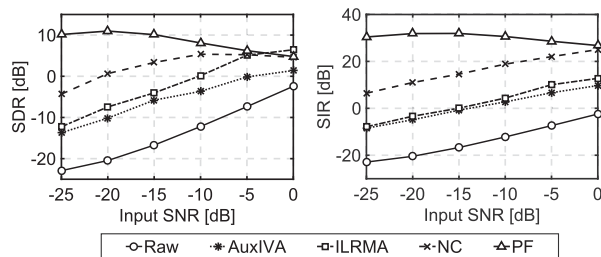


Fig. 4 SDR and SIR scores.

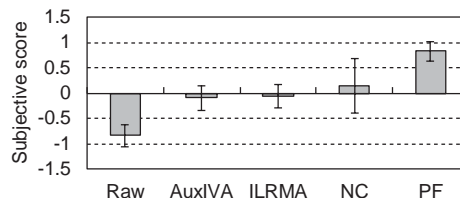


Fig. 5 Results of subjective evaluation. Error bar denotes 95% confidence interval.

用し、7 名による主観評価実験を行った。ここでは、被験者に対し前述の 5 種類の信号を 2 種類ずつ聴取させ、より音声が明瞭に聞こえるものを選択させた。

4.2 結果

Figure 4 は、あるロボットの姿勢における SDR および SIR を示す。入力 SN 比が十分に低い状況においては、処理を経るにつれて SDR・SIR が改善する様子を読み取ることができる。SN 比が -5 dB より高い状況において、処理により SDR が悪化する場合もみられるが、これはノイズキャンセラおよびポストフィルタの非線形処理から生じる歪みに起因していると考えられる。また、Fig. 5 は、主観評価により得られたスコアを示す。これらの結果より、提案システムを利用した雑音抑圧効果が確認された。

5 おわりに

本稿では、柔軟索状ロボットにおける雑音抑圧システムについて述べた。開発したシステムにおける信号処理は、ILRMA、ノイズキャンセラ、ポストフィルタの 3 段階の処理を経ており、客観評価指標および主観評価実験を通じ、処理を経るにつれて雑音抑圧の効果がみられることを示した。

謝辞 本研究は、総合科学技術・イノベーション会議による革新的研究開発推進プログラム (ImPACT) の助成を受けた。

参考文献

- [1] D. Kitamura et al., IEEE/ACM Trans. Audio, Speech, Lang. Process., 24 (9), 1626–1641, 2016.
- [2] N.Q.K. Duong et al., IEEE Trans. Audio, Speech, Lang. Process., 18 (7), 1830–1840, 2010.
- [3] M. Ishimura et al., Proc. IWAENC, 2016.
- [4] C. Breithaupt et al., IEEE Trans. Audio, Speech, Lang. Process., 19 (2), 277–289, 2011.
- [5] Y. Murota et al., Proc. ICASSP, 7540–7544, 2014.
- [6] T. Esch et al., Proc. ICASSP, 4409–4412, 2009.
- [7] N. Ono, Proc. WASPAA, 189–192, 2011.