

ブラインド信号分離と Audio-Language モデルを融合した音源再構成に基づく音響ノイズ低減 Noise Reduction through Sound Source Reconstruction by Combining Blind Signal Separation and Audio-Language Model

東京大学 ○正田 晃己, ルイ笠原 純ユネス, 安琪, 山下 淳

Abstract 本稿では、目的音の事前収集と到来方向の設定を必要としない、新たなノイズ低減手法を提案する。音響点検は不可視の欠陥を検出することができるため、製造工場や建築物の安全性および信頼性を確保する上で極めて重要である。一方で、点検現場には多種多様な騒音が存在し、これらが点検対象の音響信号に混入することで、点検の精度は低下する。既存のノイズ低減手法は、目的音の事前収集や到来方向の仮定が必要であり、その適用範囲は限定的である。これらの制約を取り除いた手法を実現するため、ブラインド信号分離と Audio-Language モデルを融合した音源再構成に基づくノイズ低減手法を提案する。実験では打音点検を対象に提案手法を検証し、幅広い信号ノイズ比の条件において効果的なノイズ低減を達成した。本手法は、音響点検に限らず、施設保安のための音響イベント検出、自動運転車両の環境音認識など、様々な音響計測タスクにおける貢献が期待できる、汎用的なノイズ低減アプローチである。

1 序論

点検は、製造工場や建築物の安全性および信頼性を確保する上で極めて重要なプロセスである。特に、音響点検は機械や建築構造物の内部における不可視な欠陥を検出することができるため、ロボット等を使用した点検自動化が積極的に試みられている。しかしながら、実際の点検現場では多種多様なノイズが存在するため、ノイズの低減は容易ではない。

ノイズ低減には、ハードウェアを使った手法とソフトウェアによる手法がある。ハードウェアを使った手法では、ヘッドホンのように目的音以外を物理的に遮断することでノイズを低減する。一方で、それぞれの目的音源に対して接触して点検を行うことは、検査時間を増加させるだけでなくロボットや検査対象の破損リスクとなる。

目的音源との接触を必要としない、ソフトウェアによるノイズ低減手法は、大きく2つに分類することができる。1つ目は、ノイズ除去フィルタを構成する手法であり、深層学習を用いたアプローチ [1] は訓練された目的音に対して高い性能を示す。しかし、点検現場において目的音の事前収集が必要であり、特に異常検知タスクにおいては事前に異常状態の目的音を収集することは難しく、この手法の適用は困難である。

2つ目は、ブラインド信号分離 (BSS) に基づく手法であり、マイクアレイで捕捉された混合信号のみから音源を分離する。この手法では、分離した音源から何らかの手段で目的音を特

定する必要がある。Wang らは到来方向を設定することで、目的音の特定を達成している [2]。一方でこの手法を適用する場合は、目的音の位置を事前に把握しておく必要がある。

以上のように、既存のノイズ低減手法は目的音に関する事前情報が必要であり、その適用範囲は限定的である。そこで本研究では、目的音の事前収集と到来方向の設定を必要としないノイズ低減を目的とする。本稿では、Audio-Language モデルを用いることで BSS の音源特定の問題を解決し、本研究の目的達成を目指す。

2 提案手法

2.1 コンセプト

目的音の事前収集や到来方向の仮定を必要としないノイズ低減を実現するため、図1に示す手法を提案する。本手法は、BSS による音響信号の分解と、Consistency-Aware Attention (CA-Attention) による音源再構成の2つのプロセスから構成される。

まず、複数のマイク信号から空間情報 (音の到来時間差) と統計的独立性を利用して高精度な信号分解が可能な BSS を採用する。BSS により、事前の訓練や音源方向の仮定なしに、混合音から尤もらしい個別の音響信号への分解が可能となる。しかし、BSS のハイパーパラメタである信号分解数には、単一の音源が複数の信号に分解されるリスクと、複数の音源が単一の信号に統合されるリスクのトレードオフが存在する。さ

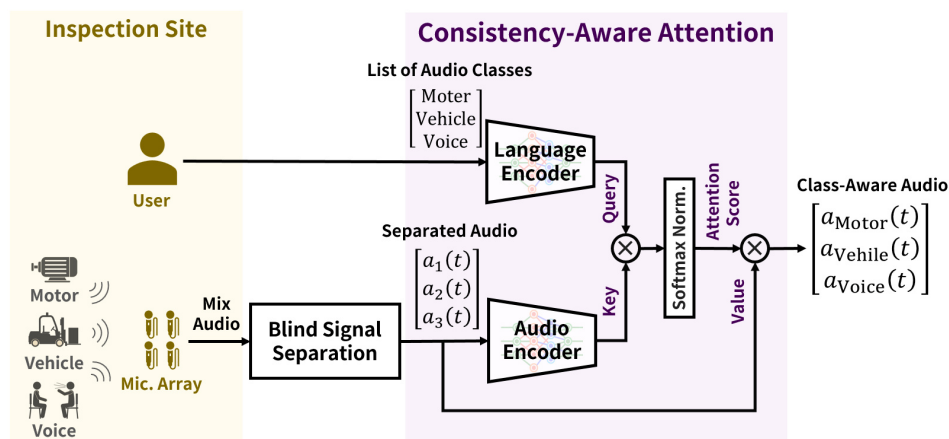


図1 提案手法の概略: 分解された音響信号とあらかじめ定義された音響クラスとの類似度を、Audio-Language モデルを使用して計算する。この類似度を正規化することで、各分解信号が各音響クラスにどれだけ関連しているかを示す Attention Score を得る。最後に、Attention Score と分解された音響信号を掛け合わせることで、音源を各クラスごとに再構成する。

表 1 スペクトル矛盾による音響クラスの認識失敗の例。表中の値はそれぞれの分散表現の類似度、太字の値はクラス別の類似度の最大値を示す。

Key \ Query	クラス A	クラス B
クラス A に対応する分解信号 (スペクトル矛盾なし)	0.8	0.4
クラス B に対応する分解信号 (スペクトル矛盾あり)	0.1	0.3

らに、BSS によって分解された各信号が、どの目的音に対応するか特定することは困難である。

これらの BSS の課題を解決するために、CA-Attention を提案する。CA-Attention ではまず、モダリティの異なる分解信号と音響クラスを Audio-Language モデルにより比較可能な分散表現に埋め込み、互いの類似度を計算する。そして、類似度の重みを付けた分解信号を足し合わせることで、音源別の再構成を目指す。さらに本研究では、莫大なデータセットを使用した事前訓練により高い汎用性をもつ Audio-Language モデル [3] を用いることで、訓練過程を省略する。

以上のように、BSS による教師なし信号分解と CA-Attention による音源再構成を組み合わせ、目的音の事前収集や到来方向の設定を必要としないノイズ低減を達成する。本手法では、ユーザは音響クラスの一覧を自然言語で設定し、目的とするクラスを選択するだけで、ノイズ低減が可能である。

2.2 Consistency-Aware Attention

BSS と Audio-Language モデルを融合させるうえで、BSS にもう 1 つ大きな問題が存在する。BSS はスペクトル矛盾性 [4] を考慮せずに混合音を分解するため、分解信号は現実には存在する音響信号から乖離することがある。さらに、Audio-Language モデルは現実の音響信号で訓練されているため、この乖離により音響クラスとの類似度が意図せず低下する。類似度の低下による音響クラスの認識失敗の例を表 1 に示す。本例において、クラス A に対応する分解信号は、スペクトル矛盾がなく全体として類似度が高いため、A と B の両方に対応する信号として誤認識される。

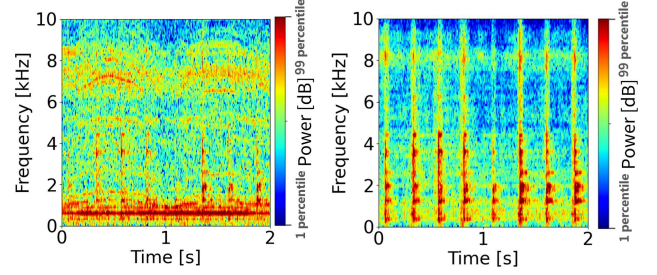
この問題を解決するため、CA-Attention を提案する。CA-Attention は式 (1) で表され、 E_L は音響クラスの分散表現を並べた行列 (Query)、 E_A は分解信号の分散表現を並べた行列 (Key)、 A は分解信号を並べた行列 (Value)、 τ は温度パラメタ、 $*$ は行列転置を表す。Transformer における Cross-Attention との極めて重要な違いは softmax^2 の箇所であり、Query に対してのみならず Key に対しても正規化を行うことを意味する。この Key に対する正規化により、類似度は分解信号ごとの相対値に変換され、音響クラスへの適切な注意機構が実現される。

$$\text{CA-Attention}(E_L, E_A, A) = \text{softmax}^2\left(\frac{E_L E_A^*}{\tau}\right) A. \quad (1)$$

3 実験

3.1 設定

提案手法の有効性を検証するため、最も重要な音響点検タスクの 1 つである打音検査を対象に実験を行った。本実験では、コンクリートの試験体を連続的に叩くことで発生する打叩音を目的音とし、ノイズ低減性能を評価する。スピーカとインパクトドライバを使い、エンジン・モータ、ファン・モータの 2 セットの音を発生させ、それらをノイズ音源として用意した。



(a) ノイズ低減前の打叩音。 (SNR: 0 dB) (b) ノイズ低減後の打叩音。 (SNR: 9.52 dB)

図 2 打叩音、ファン、モータの混合音におけるノイズ低減結果。

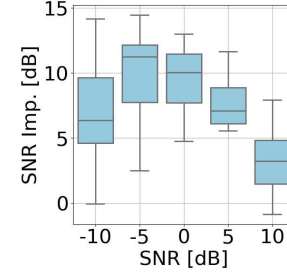


図 3 信号ノイズ比 (SNR) に対する SNR 改善量の箱ひげ図。

録音には 4 チャンネルのマイクアレイを使い、それぞれの音源別に音を記録した。また、音源とマイクの距離は約 1 メートル、各音源は 90 度間隔を開けて配置した。BSS の信号分解数は手で 4 に設定し、音響クラスは continuous tapping, engine, fan, motor の 4 つを自然言語で指定した。

3.2 結果

ノイズ低減結果の一例を図 2 に示す。6 kHz – 8 kHz における強いスペクトルはモータの音であり、回転数の変化に伴い音の高さが変化していることがわかる。また、約 1 kHz の強いスペクトルはファンの音、周期的な縦縞のスペクトルは打叩音である。図 2(b) を見ると、打叩音のスペクトルがより明瞭になっていることが確認できる。信号ノイズ比 (SNR) に対する SNR 改善量を図 3 に示しており、幅広い SNR 条件で効果的なノイズ低減を達成した。

4 結論

本稿では、目的音の事前収集と到来方向の設定を必要としないノイズ低減手法を提案し、幅広い SNR 条件で効果的なノイズ低減を達成した。本手法は、音響点検に限らず、施設保安のための音響イベント検出、自動運転車両の環境音認識など、様々な音響計測タスクにおける貢献が期待できる、汎用的なノイズ低減アプローチである。

参考文献

- [1] H. Song, M. Kim, D. Park, Y. Shin, and J.-G. Lee, “Learning from noisy labels with deep neural networks: A survey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 8135–8153, 2023.
- [2] L. Wang and A. Cavallaro, “A blind source separation framework for ego-noise reduction on multi-rotor drones,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2523–2537, 2020.
- [3] Y. Wu, K. Chen, T. Zhang, Y. Hui, T. Berg-Kirkpatrick, and S. Dubnov, “Large-scale contrastive language-audio pre-training with feature fusion and keyword-to-caption augmentation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023.
- [4] D. Kitamura and K. Yatabe, “Consistent independent low-rank matrix analysis for determined blind source separation,” *Computing Research Repository*, vol. abs/2007.00274, 2020.