

音源振動を補助情報とする時間領域単チャネル残響除去*

☆ 森下優理, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

マイクロホンによる楽器演奏の収録において、観測信号に含まれる室内残響は収録後の編集の柔軟性を低下させる要因となるため、それを除去する技術が求められている。しかし、持続的な楽器音を単チャネルで収録する場合、直接音と残響成分が重複するため、残響成分のみを区別して除去することは困難である。一方、接触型センサの一種であるピエゾセンサは、室内残響の影響を受けないという特徴を持つ。しかし、ピエゾセンサ信号はマイクロホン信号とは音響特性が異なるほか、演奏に伴う非音響的振動を含む可能性があるため、そのまま音楽制作などに用いるには適さない場合がある。そこで本稿では、ピエゾセンサ信号を補助情報として利用する残響除去手法を提案する。マイクロホンでの観測信号に対してピエゾセンサ信号を参照することで、直接音と室内残響を区別し、残響成分のみを除去するフィルタの推定を実現する。

2 逆フィルタ推定に基づく残響除去

マイクロホンで楽器音を収録する際、直接音信号 $s \in \mathbb{R}^T$ と室内インパルス応答 $h \in \mathbb{R}^L$ の畳み込みにより、時間領域信号 $x = h * s \in \mathbb{R}^T$ を得るものとする。ただし、 T は信号長、 L はインパルス応答長、 $*$ は畳み込みを表す。本稿では、単チャネル観測信号 x から所望の直接音信号 $s = w * x$ を得る h の逆フィルタ $w \in \mathbb{R}^T$ を構成することを目的とする。

従来手法の多くは、フィルタ長削減のために、時間領域での畳み込みを時間周波数領域での畳み込みで近似しているため、この近似誤差が性能の低下を引き起こす懸念がある。また、観測信号のみから逆フィルタを推定する問題は解が不定であるため、所望の逆フィルタを得るためには、直接音に関する補助情報を導入することが必要である。一般的には直接音のスパース性が補助情報として用いられるが、持続音のようなスパース性が成立しない信号に対しては、別の補助情報を導入する必要がある。

3 提案手法

本稿では、室内残響の影響を受けずに音源振動を取得するピエゾセンサ信号を補助情報とし、時間領域で逆フィルタを推定する残響除去手法を提案する。提

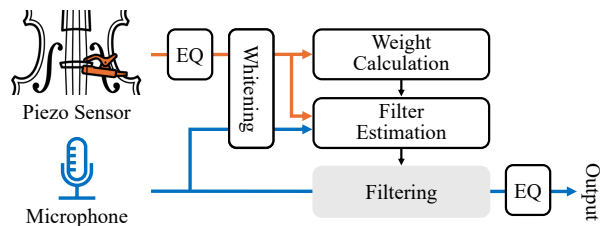


図-1 提案手法の概要。EQは周波数特性を補正する線形位相フィルタを表す。橙色の矢印はピエゾセンサ信号、青色の矢印はマイクロホン信号を表す。

案手法の概要を図-1に示す。周波数特性補正と白色化を行った後、ピエゾセンサ信号に基づいて作成した重みを用いて、最適化による逆フィルタ推定を行う。そして、推定されたフィルタを観測信号に適用し、周波数特性補正を行うことで直接音信号を推定する。

3.1 周波数特性補正と白色化

ピエゾセンサ信号はマイクロホン信号と周波数特性が異なるため、そのままフィルタ推定に用いると推定信号の周波数特性に歪みが生じる。そこで、ピエゾセンサ信号に対してマイクロホンの周波数特性に近づける線形位相フィルタを適用する。また、特定の周波数帯域にエネルギーが偏った信号で二乗誤差を取ると、エネルギーが小さい周波数帯域の誤差が小さく評価されてしまう。そこで、線形予測分析に基づく白色化フィルタを用いて、マイクロホン信号とピエゾセンサ信号のスペクトル包絡を平坦化し、全周波数帯域の誤差を均一に評価できるようにする。推定されたフィルタを適用した信号に対しても、周波数特性をマイクロホンの特性に近づける線形位相フィルタを適用することで、直接音信号を推定する。

3.2 逆フィルタ推定のための最適化問題

本稿では、周波数特性を補正したピエゾセンサ信号を疑似的な正解信号とする。これにより、マイクロホン信号をピエゾセンサ信号に近づけるためのデータ忠実性項を考えることができる。また、ピエゾセンサ信号から各時間周波数ビンにおける直接音の有無を判定し、直接音が存在しないビンに対してスパース性を課すことができる。これを利用し、逆フィルタ推定を以下の凸最適化問題として定式化する。

$$\min_w \frac{1}{2} \|Xw - s_p\|_2^2 + \lambda \|\Omega_s \odot \mathcal{G}(Xw)\|_1 + \mu \|\omega_w \odot w\|_2 \quad (1)$$

*Single-channel time-domain dereverberation guided by sound source vibration. By Yuri MORISHITA, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

Algorithm 1 PDS による逆フィルタ推定**Input:** $w_0 \in \mathbb{R}^T, v_0 \in \mathbb{C}^{M \times N}, \tau, \sigma, \lambda, \mu, I > 0$ **Output:** w_I

```

1: for  $i = 0, 1, \dots, I - 1$  do
2:    $w_{i+1} = \text{prox}_{\tau\mu\|\omega_w \odot (\cdot)\|_2} (w_i - \tau(\nabla h(w_i) + \sigma\mathcal{L}^*v_i))$ 
3:    $t_i = v_i + \mathcal{L}(2w_{i+1} - w_i)$ 
4:    $v_{i+1} = t_i - \text{prox}_{\sigma^{-1}\lambda\|\Omega_s \odot (\cdot)\|_1} (t_i)$ 
5: end for

```

ここで、 $\mathbf{X} \in \mathbb{R}^{T \times T}$ は白色化した観測信号 $\mathbf{x}$ で構成される畳み込み行列、 $\mathbf{s}_p \in \mathbb{R}^T$ は周波数補正と白色化を行ったピエゾセンサ信号、 $\mathcal{G}: \mathbb{R}^T \rightarrow \mathbb{C}^{M \times N}$ は離散ガボール変換、 M は周波数点数、 N は時間フレーム数、 $\lambda, \mu > 0$ は正則化パラメータ、 $\odot$ は要素ごとの積である。また、 $\Omega_s \in \{0, 1\}^{M \times N}$ 、 $\omega_w \in \mathbb{R}_+^T$ は3.3節で述べる重みである。推定信号は $\hat{\mathbf{s}} = \mathbf{X}\mathbf{w}$ で得られる。式 (1) の第1項は、直接音を保持した残響除去の実現のため、推定信号を $\mathbf{s}_p$ に近づけるデータ忠実性項である。第2項は残響成分を抑圧するスパース正則化項である。第3項は、インパルス応答として自然な減衰を得るための時間領域での正則化項である。

3.3 重みの設計

重み行列 $\Omega_s \in \{0, 1\}^{M \times N}$ は、時間周波数領域において、推定信号の残響成分に対してのみスパース性を誘導することを目指して設計する。ピエゾセンサ信号のエネルギーが小さい領域は、室内残響成分が支配的であるとみなして重みを1に設定することでスパース性を課し、他の領域の重みを0に設定する。

観測信号とフィルタの線形畳み込み $\mathbf{x} * \mathbf{w}$ を高速フーリエ変換の要素積による巡回畳み込みで実装するため、 $\mathbf{w}$ のタップ長を調節する重みベクトル $\omega_w \in \mathbb{R}^T$ を導入する。 $\mathbf{x}$ の末尾にタップ長以上の長さをゼロ埋めして巡回畳み込みを行うと、巡回が起らず、線形畳み込みとなる。

3.4 近接分離アルゴリズムによる最適化

本稿では式 (1) の最適化問題に主双対近接分離法 (PDS) を適用し、Alg. 1 を導出した。ここで、 $\tau, \sigma > 0$ はステップサイズ¹、 I は反復回数、 $\nabla h(\mathbf{w})$ は第1項の勾配、 $\mathcal{L}(\cdot) = \mathcal{G}(\mathbf{X}(\cdot))$ は第2項の線形作用素、 $\mathcal{L}^*$ はその随伴作用素である。また、 $\text{prox}_{\sigma^{-1}\lambda\|\Omega_s \odot (\cdot)\|_1}$ は第2項の近接作用素、 $\text{prox}_{\tau\mu\|\omega_w \odot (\cdot)\|_2}$ は第3項の近接作用素²であり、いずれも解析的に計算できる [2]。

¹ステップサイズ τ, σ は、 ∇h の Lipschitz 定数 $\beta = \|\mathbf{X}\|_S^2$ を用いて、アルゴリズムの収束条件 $\tau((\beta/2) + \sigma\|\mathcal{L}\|_S^2) \leq 1$ を満たすように設定する。ただし、 $\|\cdot\|_S$ はスペクトルノルムである。

²関数 f の近接作用素は以下のように定義される [2]。

$$\text{prox}_{\gamma f}(\mathbf{v}) = \arg \min_{\mathbf{x}} (f(\mathbf{x}) + (1/2\gamma)\|\mathbf{x} - \mathbf{v}\|_2^2)$$

表-1 客観評価指標の比較。太字は各指標の最良値を示す。

手法	SDR [dB]	STOI	PESQ
観測信号	4.68	0.68	2.65
ピエゾセンサ信号	5.62	0.84	2.07
Piezo WPE	5.06	0.71	2.60
提案手法	6.71	0.89	3.22

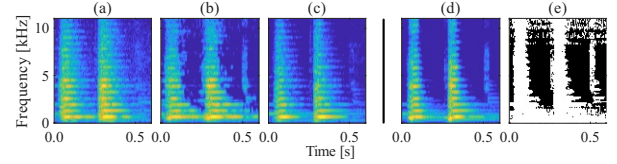


図-2 実演奏による実験結果のスペクトログラムと重み行列。(a) オフマイク信号、(b) 提案手法、(c) オンマイク信号、(d) ピエゾセンサ信号、(e) 重み行列 Ω_s 。カラーレンジは 80 dB に設定した。

4 評価実験

提案手法の性能をスピーカ再生音を用いて定量評価し、さらにヴァイオリンの実演奏音源に適用した。

スピーカを用いた実験では、スピーカにピエゾセンサを装着して無響音源を再生し、残響を含む信号をマイクロホンで収録した。サンプリング周波数は 16 kHz とした。無響信号を参照信号とする SDR, STOI, PESQ を算出した。また、ピエゾセンサ信号を用いて初期化した WPE[2] (Piezo WPE) とも比較した。結果を表-1 に示す。提案手法の指標は、観測信号、ピエゾセンサ信号および Piezo WPE を上回った。

実演奏による実験では、ヴァイオリンの演奏を、マイクロホンと胸に装着したピエゾセンサで収録した。サンプリング周波数は 22.05 kHz とした。このとき、ヴァイオリン近傍で収録したオンマイク信号と、離れた位置で収録したオフマイク信号を同時に取得した。オンマイク信号は残響の少ない比較用の信号として用い、オフマイク信号に対して残響除去を適用した。結果を図-2 に示す。観測信号 (a) には残響成分が多く含まれる一方、提案手法によって得られた信号 (b) のスペクトログラムはオンマイク信号のものに近く、残響成分が効果的に抑圧されていることが確認できた。

5 むすび

本稿では、ピエゾセンサで取得した音源振動を補助情報として、時間領域の逆フィルタ推定を行う単チャネル残響除去手法を提案した。今後は、重みの設計方法の改善による性能向上を目指す。

参考文献

- [1] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi and B. H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Audio Speech Lang. Process.*, **18**(7), 1717–1731 (2010).
- [2] N. Parikh and S. Boyd, "Proximal Algorithms," *Found. Trends Optim.*, **1**(3), 127–239 (2014).