

単チャネル Cross-Relation 法に基づく時間周波数領域ブラインド残響除去*

◎ 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

単一マイクロホンによる観測では空間情報を取得できないため、単チャネルブラインド残響除去は不良設定問題である。この不足する情報を補うため、WPE [1] や、行列復元問題として定式化するアプローチ [2] に代表されるように、従来法の多くは除去後の信号の時間周波数成分に対するスパース性を仮定してきた。しかしこれらの手法は、仮定が成立しにくい音楽信号などへの適用には限界があった。そこで本稿では、畳み込みの可換性から導かれる、マルチチャネルの観測信号間の関係を利用して残響除去を行う Cross-Relation 法の考え方にに基づき、スパース性などの音源の性質に依存しない単チャネル残響除去手法を提案する。

2 マルチチャネル Cross-Relation 法

室内での残響付加過程は、室内インパルス応答 (RIR) の畳み込みによる線形時不変システムとしてモデル化できるため、残響除去はシステム同定問題と深く関連している。例えば、マルチチャネルシステムの同定手法に Cross-Relation (CR) 法がある。文献 [3] では、時間周波数 (TF) 領域での CR (TF-CR) によって RIR を推定し、その逆フィルタ処理で残響を除去する手法を提案している。本節では TF-CR による RIR の推定法を、2 チャネル観測を例に説明する。

TF-CR では、周波数サブバンドごとに伝達特性がかかる畳み込み伝達関数 (CTF) モデルを用いて、各周波数で CTF を推定する。本モデルでは、サブバンド f での RIR としての役割を果たす CTF $h_1^{(f)}, h_2^{(f)}$ と、残響を含まない信号 $s^{(f)} \in \mathbb{C}^d$ の線形畳み込み $x_i^{(f)} = h_i^{(f)} * s^{(f)}$ ($i = 1, 2$) により、各チャネルの観測信号 $x_1^{(f)} \in \mathbb{C}^{n_1}, x_2^{(f)} \in \mathbb{C}^{n_2}$ を表す。ただし $*$ は畳み込みを表す。すべての周波数で CTF を推定した後、それらを全周波数で結合し、逆短時間 Fourier 変換によって時間領域の RIR を推定する。以降添字 f の表記を省略する。

CTF h_1 及び h_2 の推定は、畳み込みの可換性より導かれる等式 $x_1 * h_2 = s * h_1 * h_2 = x_2 * h_1$ を整理することで得られる、 s が消去された過決定方程式

$$\underbrace{\begin{bmatrix} \mathbf{M}_{n_2-d+1}(x_2) & -\mathbf{M}_{n_1-d+1}(x_1) \end{bmatrix}}_{=: \mathbf{S}_d \in \mathbb{C}^{(n_1+n_2-d) \times (n_1+n_2-2d+2)}} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \mathbf{0} \quad (1)$$

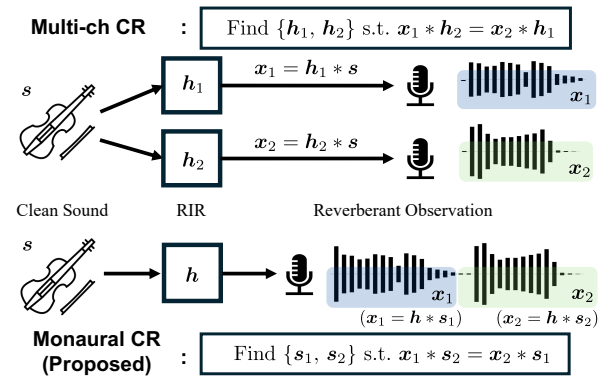


図-1 マルチチャネル CR 法及び提案法の概念図

の求解に帰着させる。ここで、 $\mathbf{M}_p(\mathbf{a}) \in \mathbb{C}^{(p+q-1) \times p}$ は長さ q のベクトル $\mathbf{a}$ の畳み込み演算を表現する列数 p の Toeplitz 行列であり、行列 $\mathbf{S}_d$ は d 次 Sylvester 行列である。CTF h_1, h_2 が互いに素である (e.g. Z 変換が共通の零点を持たない) なら、 $\mathbf{S}_d$ の右零空間は 1 次元であるため、その基底として (スケール不定性を除いて) h_1, h_2 を一意に復元可能である。

実際には、スペクトルの漏れ等の影響により CTF モデルは厳密に成立しないため、式 (1) も厳密な解を持たない。そのため、 $\mathbf{S}_d$ の最小特異値に対応する右特異ベクトルで当該右零空間の基底を代替する。

3 提案法：単チャネル Cross-Relation 法

提案法では、マルチチャネル CR 法における信号と残響の関係性を反転し、単チャネルで異なる時間区間の信号を観測する設定に適用する。概念図を図-1 に示す。提案法は、RIR の推定を介さずに、残響を含まない目的信号を直接推定可能である。

提案法では、単一マイクロホンで得られた異なる時間区間の観測のサブバンド成分 $x_1 \in \mathbb{C}^{N_1}, x_2 \in \mathbb{C}^{N_2}$ を、共通の CTF $h \in \mathbb{C}^D$ と 2 つの未知目的信号 $s_1 \in \mathbb{C}^{N_1-D+1}, s_2 \in \mathbb{C}^{N_2-D+1}$ の線形畳み込み

$$x_i = s_i * h, \quad \text{for } i = 1, 2 \quad (2)$$

によりモデル化する。式 (2) に基づき、第 2 節と同様に畳み込みの可換性を用いると、未知の共通因子である h が消去され、 s_1 及び s_2 に関する以下の方程式が導かれる。

$$\underbrace{\begin{bmatrix} \mathbf{M}_{N_1-D+1}(x_2) & -\mathbf{M}_{N_2-D+1}(x_1) \end{bmatrix}}_{=: \mathbf{S}_D \in \mathbb{C}^{(N_1+N_2-D) \times (N_1+N_2-2D+2)}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \mathbf{0} \quad (3)$$

* Time-frequency domain blind dereverberation based on single-channel cross-relation method. By Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

Algorithm 1 Sub-band Monaural CR Method

Input: Subband signals $\mathbf{x}_1 \in \mathbb{C}^{N_1}, \mathbf{x}_2 \in \mathbb{C}^{N_2}$ at a certain frequency bin, prescribed CTF length D
Output: $\hat{\mathbf{s}}_1 \in \mathbb{C}^{N_1-D+1}, \hat{\mathbf{s}}_2 \in \mathbb{C}^{N_2-D+1}$
1: $\mathbf{S}_D \leftarrow [\mathbf{M}_{N_1-D+1}(\mathbf{x}_2) \quad -\mathbf{M}_{N_2-D+1}(\mathbf{x}_1)]$
2: $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \mathbf{S}_D$ {Economical SVD}
3: $\mathbf{v} \leftarrow \mathbf{V}[:, \text{end}]$
4: $\mathbf{v} \leftarrow \frac{\sqrt{\|\mathbf{x}_1\|_2^2 + \|\mathbf{x}_2\|_2^2}}{\|\mathbf{v}\|_2} \cdot \mathbf{v}$ {Scale recovery}
5: $[\hat{\mathbf{s}}_1^H \hat{\mathbf{s}}_2^H] \leftarrow \mathbf{v}^H$ {Split into two signals}

$\mathbf{s}_1$ と $\mathbf{s}_2$ が互いに素であると仮定すると、それらは前節での CTF と同様に、 $\mathbf{S}_D$ の最小特異値に対応する右特異ベクトルより復元可能である。各周波数ビンに対する提案法の手続きを Alg. 1 に示す。ただし、 $(\cdot)^H$ は複素共役転置、 $\mathbf{V}[:, \text{end}]$ は $\mathbf{V}$ の最終列を表す。

4 数値実験

スパース性を満たさない信号に対する、提案法の残響除去性能及び計算時間について、従来法と比較検証するため、数値実験を行った。目的信号には MIDI 音源から合成した 10 種類の異なる楽器音ペア（信号長 2s, サンプリング周波数 11 025 Hz）を使用した。各信号に対して、様々な室内音響環境において実測された RIR を 11 025 Hz にダウンサンプリングして畳み込んだ後、信号対雑音比が 20 dB となるように加法的白色ガウス雑音を重畳して観測信号を生成した。RIR には、BFTL IR Library¹ より残響時間 (RT₆₀) がそれぞれ 0.2s, 0.7s, 1.0s, 1.3s, 1.6s, 1.8s のものを 6 種類選択し用いた。図-2 左に、実験に用いた楽器音の一例と、それに対して RT₆₀ = 1.0s の RIR を畳み込んだ信号のスペクトログラムを示す。短時間 Fourier 変換のパラメータは、窓長を 512 サンプル、シフト長を 128 サンプルとした。比較対象は WPE[1] 及び Lifting 法[2] の 2 手法とした。提案法及び Lifting 法では、RIR の RT₆₀ に相当するフレーム数を CTF 長として設定し、WPE では同値をフィルタタップ長として設定した。残響除去性能の評価指標には、SI-SDR の改善量である $\Delta\text{SI-SDR}$ を用いた。また、それぞれの手法の実行時間を計測した。

図-2 右に、残響時間ごとの各手法における $\Delta\text{SI-SDR}$ の箱ひげ図を示す。実験の結果、提案法は、RT₆₀ = 1.3s の条件では比較手法との明確な差は見られなかったものの、ほとんどの残響時間条件において比較手法を上回る残響除去性能を示した。一方、従来法は十分な除去性能を示さなかった。その原因として、対象とした楽器音信号が比較手法の前提とする時間周波数領域上のスパース性を満たさない点、及び観測信号の初期到達遅延を補正できない点が挙

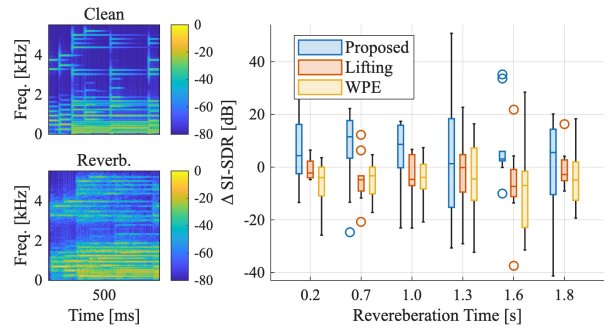


図-2 (左) 実験に用いた楽器音信号の一例と、それに対して RT₆₀ = 1.0s の RIR を畳み込んだ信号のスペクトログラム。(右) 各残響時間での残響除去性能の比較。

表-1 異なる残響時間での平均計算時間 [s]

RT ₆₀ [s]	0.2	0.7	1.0	1.3	1.6	1.8
Proposed	6.9	7.7	7.7	8.5	8.9	9.6
Lifting	44.7	129.7	194.9	288.3	353.7	470.4
WPE	0.4	0.8	1.0	3.3	3.8	4.4

げられる。

表-1 に 1 ペアあたりの平均計算時間を示した。WPE 及び Lifting 法は、CTF 長の増大に伴って計算時間が増加した。一方、提案法は WPE と比較すると計算時間を要するものの、その増加は緩やかであり、残響時間の増加に対して計算コストの上昇が抑えられたことを確認した。

5 おわりに

CR 法に基づくブラインド残響除去を単チャンネルに適用した手法を提案した。評価実験により、従来法では困難であったスパース性を満たさない楽器音信号に対しても、提案法が高い残響除去性能を達成することを確認した。今後は、Sylvester 行列の右零空間の基底ベクトルの偏角の不定性に起因する、周波数ビン間での位相の不整合問題の解決に取り組む。

謝辞 本研究は、科学技術振興機構 (JST) 創発的研究支援事業 (JPMJFR2330) の支援を受けて実施されたものである。

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 TASLP*, **18**(7), 1717–1731 (2010).
- [2] F. Yohena and K. Yatabe, "Single-channel blind dereverberation based on sparse matrix recovery with reweighting and accelerated alternating direction method of multipliers," *Acoust. Sci. & Tech.*, **46**(5), 544–552 (2025).
- [3] X. Li, S. Gannot, L. Girin and R. Horaud, "Multichannel identification and nonnegative equalization for dereverberation and noise reduction based on convolutive transfer function," *IEEE/ACM TASLP*, **26**(10), 1755–1768 (2018).

¹<https://www.cianriordan.com/barefoot>