

4 チャンネル信号を四元数値信号とみなしたブラインド残響除去*

◎ 新井慶大, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

アレイ信号処理は、センサ間の位相差などの空間情報を利用する枠組みである。このようにチャンネル間の相対的な関係が重要な信号処理では、複素数を拡張した四元数が用いられることがある。例えば、3 チャンネル信号であるカラー画像に対し、RGB の各チャンネルを四元数の3つの虚部に割り当て、四元数領域で画像処理を行う手法が提案されている [1]。これに着想を得て、本稿では4チャンネルマイクロホンアレイの観測信号を四元数の各部に割り当てる音響信号処理を考え、その一例として残響除去を扱う。具体的には、複素数値信号の多チャンネル線形予測に基づく従来手法である WPE [2] を四元数領域で再定式化し、新たな残響除去アルゴリズムとして提案する。

2 線形予測に基づく多チャンネル残響除去

単一の音源が M 個のマイクロホンで観測される状況を考える。 m 番目のマイクロホンにおける観測信号の時間周波数表現から、 f 番目の周波数ビンを取り出した時系列信号を $\mathbf{x}_{m,f} \in \mathbb{C}^T$ とする。ただし、 T は時間フレーム数である。以降、各周波数ビンは独立に処理されるため、インデックス f は省略する。

多チャンネル残響除去の代表的な手法に、多チャンネル線形予測 (MCLP) に基づく WPE [2] がある。この手法では、観測信号 $\mathbf{x}_m$ を直接音と初期反射音からなる成分 $\mathbf{d}_m \in \mathbb{C}^T$ と後部残響成分 $\mathbf{r}_m \in \mathbb{C}^T$ の和として、

$$\begin{aligned} x_m[t] &= d_m[t] + r_m[t] \\ &= \sum_{l=0}^{\tau-1} h_m[l] s[t-l] + \sum_{l=\tau}^{L_h-1} h_m[l] s[t-l] \quad (1) \end{aligned}$$

と表す。ただし、 $\mathbf{s} \in \mathbb{C}^T$ は音源信号、 $\mathbf{h}_m \in \mathbb{C}^{L_h}$ は音源と m 番目のマイクロホンとの間の残響成分に対応するフィルタ係数、 L_h はフィルタ長、 τ は直接音及び初期反射音の範囲を決める予測遅延である。ここで、後部残響成分 $\mathbf{r}_m$ が過去の観測信号 $\mathbf{x}_m$ の線形結合で予測できると仮定すると、式 (1) は

$$x_m[t] = d_m[t] + \sum_{m'=1}^M \sum_{l=0}^{L_w-1} w_{m'}[l] x_{m'}[t-\tau-l] \quad (2)$$

と書き換えられる [2]。ただし、 $\mathbf{w}_{m'} \in \mathbb{C}^{L_w}$ は m' 番目のマイクロホンに対応する残響除去フィルタ、 L_w はタップ長である。このフィルタ $\mathbf{w}_{m'}$ を推定し、予

測された後部残響成分を観測信号から差し引くことで、残響除去された信号 $\mathbf{d}_m$ を得る。

3 四元数領域における4チャンネル残響除去

本稿では、4チャンネル信号を四元数値信号とみなした残響除去手法を提案する。まず四元数の基本演算を整理し、続いて提案手法の定式化について述べる。

3.1 四元数の基本演算

$\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ とする。四元数全体の集合は $\mathbb{H} = \{q_0 + q_1\iota + q_2j + q_3\kappa \mid q_0, q_1, q_2, q_3 \in \mathbb{F}\}$ と定義される。ただし、 ι, j, κ は虚数単位で、 $\iota^2 = j^2 = \kappa^2 = \iota j \kappa = -1$ を満たす。2つの四元数 $p, q \in \mathbb{H}$ の積 pq は虚数単位の乗算に従って定まり、一般に非可換 ($pq \neq qp$) である。また、 $q^\iota = -\iota q \iota$, $q^j = -j q j$, $q^\kappa = -\kappa q \kappa$ と定義する。これらは ι, j, κ に関する q の対合と呼ばれる。さらに、 $\mathbb{H}$ 上の共役を $q^* = \overline{q_0} - \overline{q_1}\iota - \overline{q_2}j - \overline{q_3}\kappa$, 絶対値を $|q| = \sqrt{|q_0|^2 + |q_1|^2 + |q_2|^2 + |q_3|^2}$, スカラー部を $\mathfrak{S}(q) = q_0$ と定義する [3]。ただし、右辺の $(\cdot)$, $|\cdot|$ はそれぞれ $\mathbb{C}$ 上の共役と絶対値である。

3.2 提案手法

以降ではマイクロホン数 $M = 4$ の場合を考える。はじめに、四元数領域で MCLP を定式化する。各チャンネルにおける観測信号 $\mathbf{x}_m \in \mathbb{C}^T$ ($m = 1, 2, 3, 4$) を用いて、四元数値信号 $\mathbf{q} \in \mathbb{H}^T$ を次のように構成する。

$$\mathbf{q} = \mathbf{x}_1 + \mathbf{x}_2\iota + \mathbf{x}_3j + \mathbf{x}_4\kappa \quad (3)$$

ここで、自己相関などの2次統計量を四元数領域で扱うために、もとの四元数に3つの対合を加えた拡張四元数を導入する [4]。時間フレーム t における $\mathbf{q}$ の拡張四元数を $\mathbf{q}_a[t] = [q[t], q^\iota[t], q^j[t], q^\kappa[t]]^\top \in \mathbb{H}^4$ と定義する。ただし、 $(\cdot)^\top$ は転置を表す。これを用いて、四元数領域での MCLP は

$$\begin{aligned} q[t] &= \tilde{d}[t] + \sum_{l=0}^{L_w-1} \mathbf{w}[l]^\top \mathbf{q}_a[t-\tau-l] \\ &= \tilde{d}[t] + \tilde{\mathbf{w}}^\top \tilde{\mathbf{q}}_\tau[t] \quad (4) \end{aligned}$$

と定式化できる。ただし、 $\tilde{\mathbf{d}} \in \mathbb{H}^T$ は所望の信号、 $\mathbf{w}[l] \in \mathbb{H}^4$ はタップ l に対応する残響除去フィルタである。また、 $\tilde{\mathbf{w}} = [\mathbf{w}[0]^\top, \dots, \mathbf{w}[L_w-1]^\top]^\top \in \mathbb{H}^{4L_w}$ はすべての l について $\mathbf{w}[l]$ を連結したベクトル、 $\tilde{\mathbf{q}}_\tau[t] =$

*Blind dereverberation for four-channel signals treated as quaternionic signals. By Keidai ARAI, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

Algorithm 1 Quaternion WPE (frequency-wise)

Input: $\mathbf{q} \in \mathbb{H}^T$, $\tilde{\mathbf{q}}_\tau[t] \in \mathbb{H}^{4L_w}$, τ , ϵ , $N_{\text{iter}} > 0$
Output: $\hat{\mathbf{s}} \in \mathbb{C}^T$

```

1: for  $n = 1, \dots, N_{\text{iter}}$  do
2:    $\mathbf{R} \leftarrow \sum_{t=1}^T \tilde{\mathbf{q}}_\tau[t] \tilde{\mathbf{q}}_\tau[t]^H / \sigma[t]$ 
3:    $\mathbf{p} \leftarrow \sum_{t=1}^T \tilde{\mathbf{q}}_\tau[t] q[t]^* / \sigma[t]$ 
4:    $\mathbf{R}_{\text{reg}} \leftarrow \epsilon \cdot \text{tr}(\mathbf{R}) \cdot \mathbf{I}$   $\triangleright$  For stabilizing inversion
5:    $\tilde{\mathbf{w}} \leftarrow ((\mathbf{R} + \mathbf{R}_{\text{reg}})^{-1} \mathbf{p})^*$   $\triangleright$  Update via (7)
6:   for  $t \leftarrow 1, \dots, T$  do
7:      $\tilde{d}[t] \leftarrow q[t] - \tilde{\mathbf{w}}^T \tilde{\mathbf{q}}_\tau[t]$ 
8:      $\sigma[t] \leftarrow \max(|\tilde{d}[t]|^2, \epsilon)$   $\triangleright$  Update via (6)
9:   end for
10: end for
11:  $\hat{\mathbf{s}} \leftarrow \mathfrak{S}(\mathbf{e})$   $\triangleright$  Reference: 1st channel

```

$[\mathbf{q}_a[t - \tau]^T, \dots, \mathbf{q}_a[t - \tau - (L_w - 1)]^T]^T \in \mathbb{H}^{4L_w}$ は拡張四元数で表した観測信号を連結したものである。

次に、四元数領域で WPE のアルゴリズムを導出する。所望信号 $\tilde{\mathbf{d}}$ に対して平均 0、時変分散 $\sigma \in \mathbb{R}_{\geq 0}^T$ の時変ガウス分布を仮定すると、全時間フレーム ($t = 1, \dots, T$) にわたる負対数尤度の最小化問題は

$$\min_{(\sigma, \tilde{\mathbf{w}})} \sum_{t=1}^T \left(\frac{|\tilde{d}[t]|^2}{\sigma[t]} + \log \sigma[t] \right) \quad (5)$$

と定式化できる。この問題は、 σ と $\tilde{\mathbf{w}}$ に関する交互最小化で解くことができ、各変数の更新則は

$$\sigma[t] = |q[t] - \tilde{\mathbf{w}}^T \tilde{\mathbf{q}}_\tau[t]|^2 \quad (t = 1, \dots, T) \quad (6)$$

$$\tilde{\mathbf{w}} = \left(\left(\sum_{t=1}^T \frac{\tilde{\mathbf{q}}_\tau[t] \tilde{\mathbf{q}}_\tau[t]^H}{\sigma[t]} \right)^{-1} \left(\sum_{t=1}^T \frac{\tilde{\mathbf{q}}_\tau[t] q[t]^*}{\sigma[t]} \right) \right)^* \quad (7)$$

と求まる。ただし、 $(\cdot)^{-1}$ は行列の逆、 $(\cdot)^H = ((\cdot)^T)^*$ は $\mathbb{H}$ 上の共役転置を表す。

提案手法のアルゴリズムを **Alg. 1** に示す。ただし、 $\epsilon > 0$ はゼロ除算を防ぐための微小な数、 $N_{\text{iter}} > 0$ は反復回数、 $\mathbf{R}_{\text{reg}} \in \mathbb{H}^{4L_w \times 4L_w}$ は逆行列計算の安定化のために導入した行列であり、 $\max(\cdot, \cdot)$, $\text{tr}(\cdot)$ はそれぞれ 2 入力の最大値、行列のトレースを返す関数である。Alg. 1 を周波数ビンごとに適用して時間周波数表現を更新し、時間領域へ逆変換することで、最終的な残響除去信号を得る。

4 実験

EARS データセットに含まれる計 50 発話 (男女各 25 発話) の音声信号に、Multi-Channel Impulse Response Database の室内インパルス応答¹ ($\text{RT}_{60} = 160, 360, 610$ ms) を畳み込んで得られる 4 チャンネル観測信号に対して残響除去を行った。サンプリング周波数は 16 kHz とした。短時間フーリエ変換には

¹本実験では、マイクロホン間隔 8 cm の 8 チャンネル直線アレイで収録された室内インパルス応答のうち、1, 3, 6, 8 番目の 4 チャンネルを使用した。アレイと音源間の距離は 2 m で、音源方向はマイクロホンの並ぶ方向に対して垂直とした。

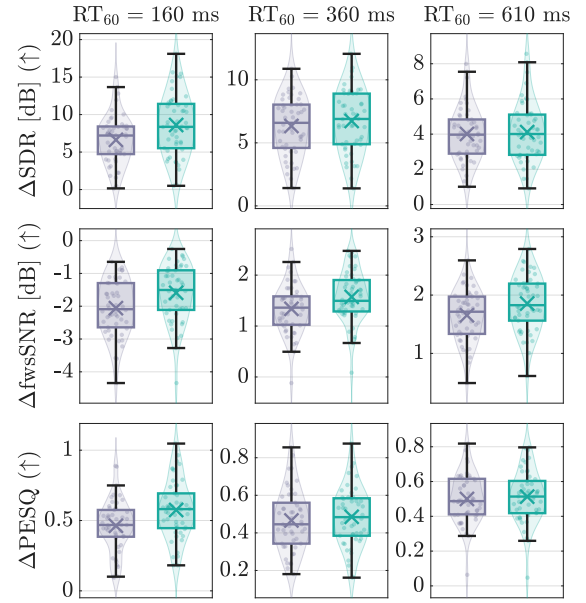


図-1 各残響時間における ΔSDR , ΔfwsSNR , ΔPESQ の比較。灰色が従来手法、緑色が提案手法に対応する。全データの平均値は×印で示し、各音源に対する評価指標の値は点で示した。

窓長が 2^{10} のハン窓を用い、合成時の窓シフトを 2^7 とした。本実験では、従来の WPE [2] と提案手法の性能を比較した。両手法について、 $L_w = 40$, $\epsilon = 10^{-6}$, $N_{\text{iter}} = 20$ とした。また、予測遅延は各残響時間に対してそれぞれ $\tau = 2, 4, 6$ とした。評価指標には ΔSDR , ΔfwsSNR , ΔPESQ を用いた。

各残響時間における残響除去性能の比較を図-1 に示す。図より、いずれの条件においても、提案手法は従来手法よりも性能が向上する傾向が見られた。特に、提案手法の平均値と中央値はすべての条件で従来手法を上回った。また、多くの条件で評価指標の最小値は従来手法と同等以上を維持しながら、最大値は向上していることが確認できた。

5 むすび

本稿では、4 チャンネル信号を単一の四元数値信号として扱い、四元数領域で残響除去を行う手法を提案した。今後は、四元数値信号の構成方法の違いによる性能の変化を調査する。

参考文献

- [1] Q. Barthélemy, A. Larue and J. I. Mars, “Color sparse representations for image processing: Review, models, and prospects,” *IEEE Trans. Image Process.*, **24**(11), 3978–3989 (2015).
- [2] 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).
- [3] A. Dagdeviren, “A generalization of complex, dual and hyperbolic quaternions: Hybrid quaternions,” *Filomat*, **37**(25), 8441–8454 (2023).
- [4] C. C. Took and D. P. Mandic, “A quaternion widely linear adaptive filter,” *IEEE Trans. Signal Process.*, **58**(8), 4427–4431 (2010).