

複数音源が存在する残響環境での マルチステップ線形予測の効果*

木下慶介 中谷智広 澤田宏 荒木章子 三好正人
(日本電信電話株式会社 NTT コミュニケーション科学基礎研究所)

1 はじめに

残響環境下で複数音源を多チャンネル收音する場合、観測信号は畳み込み混合としてモデル化される。この畳み込み混合から目的音源のみを抽出する問題に鑑みて、死角型ビームフォーマ [1] や、独立成分分析を用いた音源分離方法 [2] 等の種々の技術が提案されている。しかし、それらの方法では、長残響下での畳み込み混合からの精度良い目的音抽出は困難である。

本稿では、インパルス応答の後部にも比較的強いエネルギーを持つ畳み込み混合過程から得られる信号(以下、長い畳み込み混合)に対する、マルチステップ線形予測 [3] の効果を調べる。マルチステップ線形予測を用いて、長い畳み込み混合の残響成分を効果的に抑圧できることを示す。また、一応用例として、長残響下での音源分離処理の前処理にマルチステップ線形予測を適用し、その効果を調べる。

1.1 観測信号のモデル

本稿では、簡単のため、2音源の畳み込み混合を扱う。目的信号を $s(n)$ 、妨害音を $d(n)$ 、目的信号と m 番目のマイクロホン間の室内インパルス応答 (K 次) を $h_{s,m}(n)$ 、妨害音と m 番目のマイクロホン間の室内インパルス応答 (K 次) を $h_{d,m}(n)$ とすると、多チャンネル收音音声の m 番目のマイクロホン信号 $y_m(n)$ は、次式の畳み込み混合で表される。

$$y_m(n) = \sum_{t=0}^{K-1} h_{s,m}(t)s(n-t) + \sum_{t=0}^{K-1} h_{d,m}(t)d(n-t) \quad (1)$$

ここで、室内伝達関数 $h_{o,m}(z)$ ($o = s$ or d) は時不変、非最小位相であり、共通の零点を持たないものと仮定する。

また、上式を行列で表すと以下のように表せる。

$$\mathbf{y}_m = [\mathbf{s}^T(n), \mathbf{d}^T(n)]\mathbf{H}_m, \\ \triangleq \mathbf{x}^T(n)\mathbf{H}_m.$$

ここで、

$$\mathbf{s}(n) \triangleq [s(n), s(n-1), \dots, s(n-(K+N))]^T,$$

$$\mathbf{d}(n) \triangleq [d(n), d(n-1), \dots, d(n-(K+N))]^T,$$

$$\mathbf{H}_m \triangleq [\mathbf{H}_{s,m}, \mathbf{H}_{d,m}]^T,$$

$$\mathbf{H}_{o,m} \triangleq \begin{bmatrix} \mathbf{h}_{o,m} & 0 & \dots & \dots & 0 \\ 0 & \mathbf{h}_{o,m} & \ddots & & \vdots \\ \vdots & \ddots & \mathbf{h}_{o,m} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \mathbf{h}_{o,m} \end{bmatrix},$$

$$\mathbf{h}_{o,m} \triangleq [h_{o,m}(0), h_{o,m}(1), \dots, h_{o,m}(K-1)].$$

2 マルチステップ線形予測の効果

2.1 マルチステップ線形予測

M をマイクロホンの総数、 D をステップサイズ、 N を各チャンネルのフィルタ長とすると、マルチステップ線形予測は以下のように定式化される。

$$y_i(n) = \sum_{m=1}^M \sum_{p=0}^{N-1} w_{m,i}(p)y_m(n-p-D) + e_{m,i}(n). \quad (2)$$

$w_{m,i}(n)$ は、 i 番目のマイクロホン信号中の後部残響を予測するための、マイクロホン m の線形予測係数、 $e_{m,i}(n)$ は予測誤差である。

予測誤差 $e_{m,i}(n)$ の平均二乗エネルギーを最小化するような予測フィルタ $w_{m,i}(n)$ は、行列の形 \mathbf{w}_i で以下のように表される [4]。

$$\mathbf{w}_i = \mathbf{H}^+ \mathbf{C} \mathbf{H}_i \mathbf{1}. \quad (3)$$

ここで、

$$\mathbf{w}_i \triangleq [\mathbf{w}_{1,i}^T, \mathbf{w}_{2,i}^T, \dots, \mathbf{w}_{M,i}^T]^T,$$

$$\mathbf{w}_{m,i} \triangleq [w_{m,i}(0), w_{m,i}(1), \dots, w_{m,i}(N-1)]^T,$$

$$\mathbf{H} \triangleq [\mathbf{H}_1, \dots, \mathbf{H}_M]^T,$$

$$\mathbf{1} \triangleq [1, 0, \dots, 0]^T.$$

また、 \mathbf{C} は以下のコンパニオン行列 \mathbf{C}_s 、 \mathbf{C}_d を要素とするブロック対角行列である。

$$\mathbf{C}_s \triangleq (\mathbf{E}[s(n-D)s^T(n-D)]) + \mathbf{E}[s(n-D)s^T(n)].$$

$$\mathbf{C}_d \triangleq (\mathbf{E}[d(n-D)d^T(n-D)]) + \mathbf{E}[d(n-D)d^T(n)].$$

2.2 畳み込み混合に対する残響除去効果

\mathbf{H} を full row rank であるとするれば、予測フィルタ \mathbf{w}_i を用いて得られる予測誤差は次のように表される。

$$e_i(n) = \mathbf{x}^T(n)\mathbf{H}_i \mathbf{1} - \mathbf{x}^T(n-D)\mathbf{H}\mathbf{w}_i, \\ = \mathbf{x}^T(n)\mathbf{H}_i \mathbf{1} - \mathbf{x}^T(n-D)\mathbf{H}\{\mathbf{H}^+ \mathbf{C} \mathbf{H}_i \mathbf{1}\}, \\ = (\mathbf{x}^T(n) - \mathbf{x}^T(n-D)\mathbf{C})\mathbf{H}_i \mathbf{1}, \\ = (\mathbf{s}^T(n) - \mathbf{s}^T(n-D)\mathbf{C}_s)\mathbf{H}_{s,i} \mathbf{1} \\ + (\mathbf{d}^T(n) - \mathbf{d}^T(n-D)\mathbf{C}_d)\mathbf{H}_{d,i} \mathbf{1}. \quad (4)$$

式 (4) より、マルチステップ線形予測は、各音源とマイクロホン i 間のインパルス応答 $\mathbf{h}_{o,i}$ の $h_{o,i}(0) \sim h_{o,i}(D-1)$ は等化せず、 $h_{o,i}(D)$ 以降の後部残響成分のみを抑圧できることがわかる。尚、精度の良い後部残響抑圧のためには、 \mathbf{C} による歪みを回避するための白色化処理等が有効である [3]。

*The effect of multi-step linear prediction based dereverberation on the highly reverberant convolutive mixture of acoustic signals, by K. Kinoshita, T. Nakatani, H. Sawada, S. Araki, and M. Miyoshi (NTT Communication Science Labs., NTT Corp.)

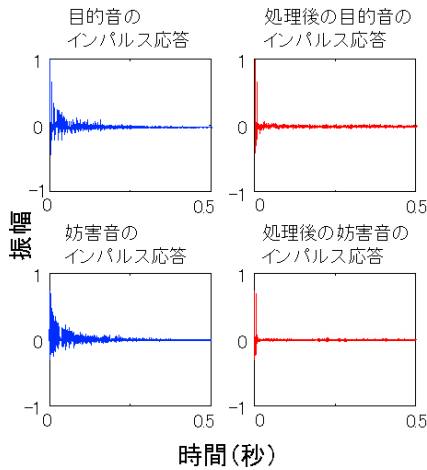


Fig. 1 残響除去処理前後の各音源とマイクロホン1間のインパルス応答：目的音-処理前(左上)、目的音-処理後(右上)、妨害音-処理前(左下)、妨害音-処理後(右下)

3 畳み込み混合の残響除去実験

ここでは、長い畳み込み混合に対する、マルチステップ線形予測の実際の動作を検証する。

3.1 音響条件

大きさ $3.5\text{m} \times 4.5\text{m} \times 2.5\text{m}$ 、残響時間 (T_{60}) 0.6 秒の残響室内のインパルス応答を、鏡像法を用いて模擬した。收音は直線上に配置されたマイクロホン 4 つを用いて行った ($M = 4$)。マイクロホン素子間距離は 20cm とした。目的音 (女性音声) はマイクロホンアレー中心から 1.5m、アレーに対して 45 度の位置に配置し、妨害音 (ピンクノイズ) はマイクロホンアレー中心から 1.5m、アレーに対して 135 度 (目的音位置から左回りに 90 度) に配置した。目的音と妨害音の混合比は、SNR5dB とした。SNR は、残響を含む目的音と残響を含む妨害音の比として定義した。

3.2 残響除去パラメータ

標準化周波数は 8kHz とし、式 (2) のフィルタ長 N とステップサイズ D は、それぞれ 4000 と 50 とした。白色化処理のためのフィルタ [3] の次数 q は 12 とした。残響除去のために 10 秒の観測音声を用いた。

3.3 畳み込み混合に対する残響除去効果

図 1 に、残響除去処理前後の、各音源とマイクロホン 1 間のインパルス応答を示した。目的音、妨害音、いずれにおいても、後部残響が処理により抑圧されている。インパルス応答中の直接音の割合¹は、処理前は平均 44% (目的音 48%、妨害音 40%) であるのに対し、処理後は平均 80% 程度 (目的音 82%、妨害音 77%) まで回復した。マルチステップ線形予測を用いて、長い畳み込み混合中の残響成分を効果的に抑圧できることを確認した。

4 音源分離の前処理としての効果

次に、マルチステップ線形予測を音源分離処理の前処理として用いた実験を行う。

$$1 \frac{\int_0^D h_{o,i}^2(t) dt}{\int_0^\infty h_{o,i}^2(t) dt} \times 100$$

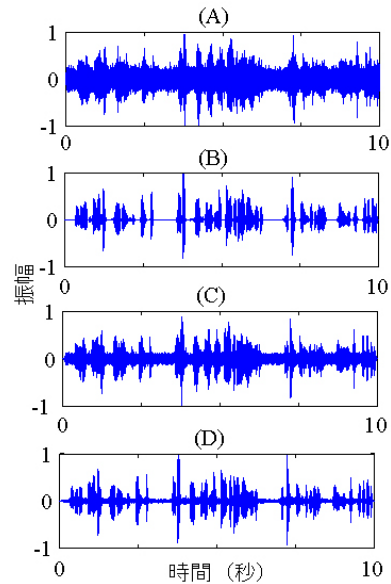


Fig. 2 音声波形：(A) 観測信号、(B) 残響除去処理により完全に後部残響を抑圧できた場合の音源分離信号、(C) 観測音をそのまま音源分離した信号、(D) マルチステップ線形予測により前処理を行った場合の音源分離信号

4.1 音源分離パラメータ

音源分離の手法としては、周波数領域独立成分分析 [2] を用いた。音源分離のためのフレーム長は、128ms とした。

4.2 残響除去による音源分離性能の向上

図 2 に、(A) 観測音、(B) 式 (1) のインパルス応答 $h_{o,m}$ を D までで打ち切った混合過程から得られる信号の音源分離信号 (残響除去処理により完全に後部残響を抑圧できた場合に相当)、マルチステップ線形予測による前処理なしの場合 (C) とありの場合 (D) の音源分離信号、それぞれの波形を示す。(C) には妨害音成分が残留しているのに対し、(D) では効果的に妨害音が抑圧されている。(B)、(C)、(D) における妨害音除去量² はそれぞれ 15.1dB、6.0dB、12.5dB である。

5 おわりに

本稿では、長残響環境下で観測される畳み込み混合に対する、マルチステップ線形予測による残響除去効果を調べた。実験の結果、マルチステップ線形予測は各音源とマイクロホン間のインパルス応答後部の残響成分を効果的に抑圧できることを示した。

参考文献

- [1] 大賀 他, “音響システムとデジタル処理,” 電子情報通信学会, 1995.
- [2] H. Sawada 他, Proc. ISCAS, pp. 3247-3250, 2007.
- [3] 木下 他, “実環境收音,” 音講論, 9 月, 2006.
- [4] T. Kailath 他, “Linear estimation,” NJ: Prentice Hall, 2000

² $20 \log_{10} \left(\frac{|\text{妨害音のインパルス応答}|}{|\text{処理後の妨害音のインパルス応答}|} \right)$