

分散型マイクロホンアレーを用いたブラインド音源分離と ポストフィルタによる性能向上の検討*

☆小野拓磨, 伊藤信貴, 宮部滋樹, 小野順貴, 嵯峨山茂樹 (東大院・情報理工)

1 はじめに

近年、複数の録音機器を分散配置させ、録音信号を後で組織的に統合する分散型マイクロホンアレーという新しい枠組みが試みられてきている [1, 2]。分散型マイクロホンアレーは、1) 広範囲にマイクロホンを配置するため、音源-マイクロホンの距離が小さくなりやすく、高 SN 比の入力を得ることが可能、2) 複数録音機器を用いることにより多素子数化が容易、3) 固定配置を必要とせずハードウェア構成が簡便といった魅力的な特長を有す。このような特長に着目し我々はこれまで、分散型マイクロホンアレーを用いたブラインド音源分離 (Blind Source Separation; BSS) 問題に取り組み、周波数領域での独立成分分析 (Independent Component Analysis; ICA) [3] におけるパーミュテーション問題 [4] の解決のために、近接させたマイクロホンをペアとして分散配置させ、ペア内マイクロホン間で推定される音源の到来時間差をクラスタリングする手法を検討してきた [5, 6]。

本報告では、分散型マイクロホンアレーによる BSS 手法の発展として、ポスト処理による拡散性雑音の抑制を検討した。BSS と残響抑制の手法には、ICA に基づく雑音パワー推定とスペクトルサブトラクションによる手法 [7] や EM アルゴリズムによる音源分離と雑音抑制の組み合わせ手法 [8] 等があるが、本研究では分散型配置の特長を積極的に利用することを考える。雑音が一定の到来方向を持たない拡散性雑音とみなせる場合、それらは距離の離れたマイクロホンペア間では無相関とみなせるのではないかと、との着眼に基づき、ペア毎の ICA とマルチチャネル Wiener フィルタによる後段処理に基づく分離手法を試みた。特に本研究ではその典型として、対象とする音源数は 2、全チャネルはワイアレスシステム等で正確な時間同期がされている場合を仮定し、拡散性雑音が重畳する環境でのシミュレーション実験によりこのような後段処理の有効性を検討した。

2 マイクロホンペア分散配置による BSS

本手法は、ICA の出力信号に対しペア間での拡散性雑音の無相関性に基づくマルチチャネル Wiener フィルタによるポスト処理を行う。フィルタ設計のために、複数チャネル間の相関を計算する必要があるため、以下 L 組のマイクロホンペアごとの ICA による分離行列の更新と復元信号の推定を行う。短時間フーリエ変換による時間周波数表現では l 組目のマイクロホンペアの 2 つの観測信号のベクトル列 $\mathbf{X}_l(\tau, \omega) = [X_{l1}(\tau, \omega), X_{l2}(\tau, \omega)]^T$ ($l = 1, \dots, L$) は $\mathbf{X}_l(\tau, \omega) \simeq A_l(\omega)\mathbf{S}(\tau, \omega)$ ($l = 1, \dots, L$) と表せる。ここで τ, ω はフレーム、周波数のインデックス、 $A_l(\omega)$ は l 組目のペアでの混合行列である。周波数領域 ICA により、混合過程が未知の条件で観測信号から元信号を復

元した信号 $\mathbf{Y}_l(\tau, \omega) = W_l(\omega)\mathbf{X}_l(\tau, \omega)$ ($l = 1, \dots, L$) を推定する。ここで $W_l(\omega)$ は l 組目のペアでの分離行列である。分離行列 $W_l(\omega)$ の更新には補助関数法に基づく手法 [9] を用いた。

このように、各マイクロホンペア、各周波数ごとに独立に分離行列が更新されるが、解の中にスケールとパーミュテーションの任意性が残る。対角行列 $B_l(\omega)$ と置換行列 $P_l(\omega)$ が推定できれば $\hat{\mathbf{S}}_l(\tau, \omega) = B_l(\omega)P_l(\omega)W_l(\omega)\mathbf{X}_l(\tau, \omega)$ ($l = 1, \dots, L$) により元信号の推定値を得る。ここで、両任意性の解決は我々の発表文献 [10] に準ずる。

3 ポストフィルタによる拡散性雑音抑制

各マイクロホンペア間の距離は十分に大きくとることができるので、ペア間の拡散性雑音成分は無相関であるとみなせる。この性質を用い、ペア毎での ICA により得られた複数の出力信号から、フィルタ生成のための元信号のパワースペクトルとステアリングベクトルの推定と、マルチチャネル Wiener フィルタ設計による拡散性雑音の抑制を試みる。以下では ICA の出力信号を並び替えた形で改めて定式化を行う。音源ごとに ICA の出力信号を並べたベクトル $\tilde{\mathbf{Y}}_n(\tau, \omega) = [Y_{1n}(\tau, \omega), \dots, Y_{Ln}(\tau, \omega)]^T$ は ICA により抑制不十分な拡散性雑音を含んだ形で

$$\tilde{\mathbf{Y}}_n(\tau, \omega) = S_n(\tau, \omega)\mathbf{v}_n(\omega) + \mathbf{N}_n(\tau, \omega) \quad (n = 1, 2) \quad (1)$$

のように定式化できる。ここで、 n は音源のインデックス、 $\mathbf{v}_n(\omega)$ はステアリングベクトル、 $\mathbf{N}_n(\tau, \omega)$ は n 番目の音源以外の拡散性雑音成分である。元信号と雑音は無相関が仮定できるので、短時間でのエルゴード性を仮定すれば式 (1) は以下のように表せる。

$$\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega) = \phi_{SS}^n(\tau, \omega)\mathbf{v}_n(\omega)\mathbf{v}_n^H(\omega) + \Phi_{\mathbf{N}\mathbf{N}}^n(\tau, \omega) \quad (2)$$

ただし $\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega) = E[\tilde{\mathbf{Y}}_n(\tau, \omega)\tilde{\mathbf{Y}}_n^H(\tau, \omega)]$, $\phi_{SS}^n(\tau, \omega) = E[|S_n(\tau, \omega)|^2]$, $\Phi_{\mathbf{N}\mathbf{N}}^n(\tau, \omega) = E[\mathbf{N}_n(\tau, \omega)\mathbf{N}_n^H(\tau, \omega)]$ である。

3.1 元信号のパワースペクトルとステアリングベクトルの同時推定

元信号のパワースペクトルとステアリングベクトル推定のために、式 (2) の非対角成分を考えると、ペア間での拡散性雑音の無相関性から以下が成り立つ。

$$\phi_{\tilde{\mathbf{Y}}_i\tilde{\mathbf{Y}}_j}^n(\tau, \omega) \simeq \phi_{SS}^n(\tau, \omega)v_{ni}(\omega)v_{nj}^*(\omega) \quad (i \neq j) \quad (3)$$

ただし $\phi_{\tilde{\mathbf{Y}}_i\tilde{\mathbf{Y}}_j}^n(\tau, \omega)$ は $\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)$ の $\{ij\}$ 成分、 $v_{ni}(\omega)$ は $\mathbf{v}_n(\omega)$ の i 行目の成分である。これを用いて、式 (3) の両辺の差の二乗和を目的関数とし、

$$J_n(\omega) = \sum_{i,j,i \neq j} E \left[|\phi_{\tilde{\mathbf{Y}}_i\tilde{\mathbf{Y}}_j}^n(\tau, \omega) - \phi_{SS}^n(\tau, \omega)v_{ni}(\omega)v_{nj}^*(\omega)|^2 \right] \quad (4)$$

*Blind source separation with distributed microphone arrays and investigation of performance gain using post-filter. by ONO Takuma, ITO Nobutaka, MIYABE Shigeki, ONO Nobutaka, SAGAYAMA Shigeki (The University of Tokyo)

Table 1 シミュレーション実験の条件

Sources	ATR database
Room size	6.3 m × 7.7 m × 2.7 m
Reverberation time	300 ms
Sampling rate	16 kHz
Data length	4.0 s
Frame length (ICA)	4096 points(256 ms)
Frame shift (ICA)	1024 points
Frame length (Post filter)	1024 points(64 ms)
Frame shift (Post filter)	64 points

を最小にすることで最適な元信号のパワースペクトルの推定値 $\hat{\phi}_{SS}^n(\tau, \omega)$ とステアリングベクトルの推定値 $\hat{\mathbf{v}}_n(\omega)$ が同時に求まる [11]。推定のための更新式は余白の都合上省略する。

3.2 マルチチャネル Wiener フィルタの設計

二乗誤差を最小にする意味で最適なフィルタ処理として、前節で推定したステアリングベクトル $\hat{\mathbf{v}}_n(\omega)$ を用いた最小分散ビームフォーマ (MVDR: Minimum Variance Distortionless Response) の出力

$$Z_n(\tau, \omega) = \frac{\hat{\mathbf{v}}_n^H(\omega) \Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)^{-1} \tilde{\mathbf{Y}}_n(\tau, \omega)}{\hat{\mathbf{v}}_n^H(\omega) \Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)^{-1} \hat{\mathbf{v}}_n(\omega)} \quad (5)$$

に対して、更に後段に Wiener フィルタを用いる。

前節で推定した元信号のパワースペクトルの推定値 $\hat{\phi}_{SS}^n(\tau, \omega)$ を使い、Wiener フィルタ $H_n(\tau, \omega) = \hat{\phi}_{SS}^n(\tau, \omega) / \hat{\phi}_{ZZ}^n(\tau, \omega)$ を設計した。ここで、 $\hat{\phi}_{ZZ}^n(\tau, \omega)$ は MVDR の出力パワースペクトル $E[|Z_n(\tau, \omega)|^2]$ である。このようにポスト処理を行った最終的な出力を

$$\tilde{Z}_n(\tau, \omega) = \frac{\hat{\phi}_{SS}^n(\tau, \omega) \hat{\mathbf{v}}_n^H(\omega) \Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)^{-1} \tilde{\mathbf{Y}}_n(\tau, \omega)}{\hat{\phi}_{ZZ}^n(\tau, \omega) \hat{\mathbf{v}}_n^H(\omega) \Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)^{-1} \hat{\mathbf{v}}_n(\omega)} \quad (6)$$

とすることができる。

4 シミュレーション実験

4.1 実験目的および条件

提案手法の有用性を示す目的で、分散配置でのシミュレーション実験を行った。クリーン音声を予め録音したインパルス応答に畳み込み同期加算することで観測信号とし、平面波伝搬を仮定しランダムな方向からくる 64 個の音声を拡散性雑音とし、観測信号に付加した。Table 1 のような実験条件に対し、音源 2 個を同一平面上直交座標系 (1.0, 0), (0, 1.0) (単位 [m]) の位置に、マイクロホン 8 個 (ペア 4 個) を (0.485, 0), (0.515, 0), (0, 0.485), (0, 0.515), (-0.485, 0), (-0.515, 0), (0, -0.485), (0, -0.515) に配置し、近接したマイクロホン 2 個ずつをペアとした。

比較として、8 チャネル全てのマイクロホンを用いて PCA により音源数と同次元に圧縮後の ICA と各ペアごとの 2 チャネル ICA を実装した。補助関数を用いた ICA の反復回数は全て 30 回とし、提案法の最小分散ビームフォーマのための $\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)$ には $\tilde{\mathbf{Y}}(\tau, \omega) \tilde{\mathbf{Y}}^H(\tau, \omega)$ の全時間フレームの平均を用い、Wiener フィルタのための $\hat{\phi}_{\tilde{\mathbf{Y}}_i \tilde{\mathbf{Y}}_j}^n(\tau, \omega)$ には $\tilde{Y}_i(\tau, \omega) \tilde{Y}_j^*(\tau, \omega)$ の前後 5 フレームの平均を用いた。

Table 2 実験結果: 入力 SN 比 5 dB での分離結果

BSS method	PCA+ ICA	2ch ICA				Proposed	
		Pair1	Pair2	Pair3	Pair4		
SIR [dB]	Src1	12.3	14.2	10.5	8.4	7.2	16.3
	Src2	11.5	5.7	14.3	6.0	4.2	15.6
NRR [dB]	Src1	1.0	3.3	5.1	8.9	5.2	5.8
	Src2	0.8	2.7	4.4	4.8	5.4	6.0

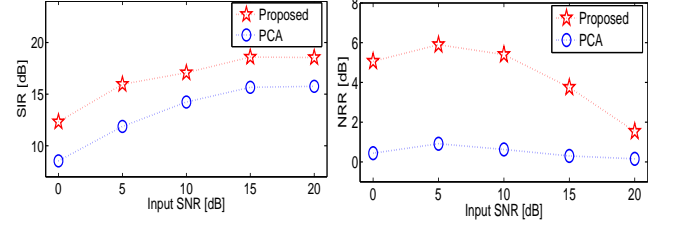


Fig. 1 入力 SN 比を変化させた場合の性能の比較 (左: SIR, 右: NRR)

4.2 実験結果

Table 2 に入力の目的音と干渉音の和と拡散性雑音の比を 5 dB とした場合の SIR (Single-to-Interference Ratio) と NRR (Noise Reduction Ratio) による評価を示す。SIR はそれぞれ目的音と干渉音のみにフィルタを通した出力信号同士の比とし、NRR は目的音と拡散性雑音にフィルタを通す前後の比の増加率として計算した。また、Fig. 1 に入力 SN 比を変化させた場合の提案法と PCA プリフィルタによる従来法の 2 音源の平均 SIR, NRR を示す。このように提案法が SIR, NRR 共に従来法を上回ることが確認できた。

5 まとめと今後の展望

本研究ではマイクロホンペアを分散させる枠組みの中で BSS を行う手法に加え、ポストフィルタによる拡散性雑音に頑健な音源分離を検討し、シミュレーション実験により提案法が拡散性雑音抑制に有効なことを示した。

今後は、室内・車内の実環境での雑音の採録とそれを用いた実験を検討している。

参考文献

- [1] Dmochowski *et al.*, *Proc. ICASSP*, pp. 89–92, 2009.
- [2] Hasegawa *et al.*, *Proc. LVA/ICA*, pp. 57–64, 2010.
- [3] Hyvärinen *et al.*, John Wiley & Sons, 2001.
- [4] Sawada *et al.*, *Proc. ICASSP*, pp. 381–384, 2003.
- [5] Ono *et al.*, *Proc. IWAENC*, 2010.
- [6] 小野他, 音講論 (春), pp. 739–740, 2010.
- [7] Takahashi *et al.*, *IEEE Trans. ASLP*, pp. 650–664, 2009.
- [8] Tran-Vu *et al.*, *Proc. IWAENC*, 2010.
- [9] Ono *et al.*, *Proc. LVA/ICA*, pp. 165–172, 2010.
- [10] 小野他, 信学技報, vol. 110, no. 331, pp. 25–30, 2010.
- [11] Ito *et al.*, *IEEE Trans. ASLP* (to appear).