

チャンネル間クロススペクトルの虚部に基づく 小サイズのアレイによる拡散性雑音抑圧*

伊藤 信貴、小野 順貴、嵯峨山 茂樹 (東大院情報理工)

1 はじめに

我々は背景雑音や室内残響などの拡散性雑音を、実用性の高い小サイズのマイクフォンアレイを用いて抑圧するための研究を行っている。これまでアレイ配置の対称性を利用した手法を提案してきた [1–5] が、本発表では任意配置のアレイに適用可能な新しい抑圧法を提案する。

2 提案法

2.1 記法

複素共役を $*$ 、転置を \top 、Hermite 転置を H で表す。信号は短時間 Fourier 変換 (STFT) 領域で表現し、フレーム番号、角周波数をそれぞれ τ 、 ω で表す。スカラーの信号 $\alpha(\tau, \omega)$ 、 $\beta(\tau, \omega)$ のクロススペクトルを $\phi_{\alpha\beta}(\tau, \omega) \triangleq E[\alpha(\tau, \omega)\beta^*(\tau, \omega)]$ 、ベクトルの信号 $\gamma(\tau, \omega)$ の共分散行列を $\Phi_{\gamma\gamma}(\tau, \omega) \triangleq E[\gamma(\tau, \omega)\gamma^{\text{H}}(\tau, \omega)]$ と表記する ($E[\cdot]$: 期待値)。

2.2 観測モデル

拡散性雑音の存在下で既知の方向から到来する目的信号を M 素子からなるアレイで観測するとする。観測信号を $x(\tau, \omega) \in \mathbb{C}^M$ 、目的信号を $s(\tau, \omega)$ 、ステアリングベクトル (既知) を $d(\omega) \in \mathbb{C}^M$ 、拡散性雑音を $v(\tau, \omega) \in \mathbb{C}^M$ と書く。このとき観測モデルは

$$x(\tau, \omega) = s(\tau, \omega)d(\omega) + v(\tau, \omega) \quad (1)$$

と書ける。 $s(\tau, \omega)$ と $v(\tau, \omega)$ は零平均・無相関と仮定すると、 $x(\tau, \omega)$ は零平均で共分散行列は

$$\Phi_{xx}(\tau, \omega) = \phi_{ss}(\tau, \omega)d(\omega)d^{\text{H}}(\omega) + \Phi_{vv}(\tau, \omega) \quad (2)$$

となる。

2.3 Wiener ポストフィルタ

拡散性雑音は指向性制御だけでは十分に抑圧できないが、最小分散ビームフォーマの出力

$$y(\tau, \omega) \triangleq \frac{d^{\text{H}}(\omega)\Phi_{xx}^{-1}(\tau, \omega)x(\tau, \omega)}{d^{\text{H}}(\omega)\Phi_{xx}^{-1}(\tau, \omega)d(\omega)} \quad (3)$$

を更に Wiener ポストフィルタ (時間周波数マスク)

$$p(\tau, \omega) \triangleq \phi_{ss}(\tau, \omega)/\phi_{yy}(\tau, \omega) \quad (4)$$

により後処理する枠組みは効果的であり、平均二乗誤差規範で最適であることも示されている [6]。Wiener ポストフィルタ (4) の設計においては、分子の目的信号のパワースペクトル $\phi_{ss}(\tau, \omega)$ を雑音の重畳した観測信号から精確に推定する方法が重要である。

Zelinski [7] の推定法はチャンネル間で雑音が無相関であるという仮定に基づく。しかし、この仮定はマイクフォン間の距離が波長に対して十分大きい場合にのみ妥当な仮定であり、アレイサイズが小さい場合には Zelinski 法の性能は大きく低下する。小サイズのアレイで推定を行うために、我々は結晶型と呼ぶ対称なアレイを用いた雑音無相関化に基づく方法を提案してきた [1–5]。本稿では更にアレイ配置に対する制約を取り除くため、チャンネル間クロススペクトルの位相に着目した新しい方法を提案する。

2.4 提案するポストフィルタ設計法

拡散性雑音が等方的であり、チャンネル間クロススペクトルがマイクフォン間の距離のみで決まると仮定すると、任意のマイクフォン対に対して、

$$\phi_{v_m v_n}(\tau, \omega) = \phi_{v_n v_m}(\tau, \omega) \quad (5)$$

が成り立つ。これは拡散性雑音のクロススペクトルは常に実数でモデル化できることを意味する。一方、特定方向から到来する目的信号に対しては、クロススペクトルは到来時間差で決まる位相を持ち、実数とは限らないため、理想的には観測信号のクロススペクトルの虚部は、

$$\Im[\phi_{x_m x_n}(\tau, \omega)] = \phi_{ss}(\tau, \omega)\Im[d_m(\omega)d_n^*(\omega)] \quad (6)$$

のように目的信号成分のみからなる ($\Im[\cdot]$: 虚部)。式 (6) から最小二乗法に基づき、 $\phi_{ss}(\tau, \omega)$ を

$$\hat{\phi}_{ss}(\tau, \omega) \triangleq \frac{\sum \Im[d_m(\omega)d_n^*(\omega)]\Im[\phi_{x_m x_n}(\tau, \omega)]}{\sum \Im[d_m(\omega)d_n^*(\omega)]^2} \quad (7)$$

$$= -\frac{\sum \sin[\omega(\delta_m - \delta_n)]\Im[\phi_{x_m x_n}(\tau, \omega)]}{\sum \sin^2[\omega(\delta_m - \delta_n)]} \quad (8)$$

により推定する。但し、 \sum は m, n に関する和、 δ_m は m 番目のマイクフォンに対する目的信号の時間遅れである。一方、Wiener ポストフィルタの分母 $\phi_{yy}(\tau, \omega)$ は Zelinski の推定法

$$\hat{\phi}_{yy}(\tau, \omega) \triangleq \frac{1}{M} \sum_{m=1}^M \phi_{x_m x_m}(\tau, \omega) \quad (9)$$

により推定し、 $\hat{p}(\tau, \omega) \triangleq \hat{\phi}_{ss}(\tau, \omega)/\hat{\phi}_{yy}(\tau, \omega)$ が提案するポストフィルタ設計である。なお、定義より $0 \leq p(\tau, \omega) \leq 1$ であるから、次の簡単な後処理を行う。

$$\begin{cases} \hat{p}(\tau, \omega) \leftarrow 0, & \text{if } \hat{p}(\tau, \omega) < 0 \\ \hat{p}(\tau, \omega) \leftarrow 1, & \text{if } \hat{p}(\tau, \omega) > 1 \end{cases} \quad (10)$$

3 実験

3.1 実験条件

提案法の有効性を確認するためにシミュレーションを行った。Fig. 1 に音源とアレイの配置を示す。マイクフォンは直線上に等間隔に配置した。等間隔 64 方向から異なる音声の平面波を重畳してカクテルパーティーの状況を模擬した。一方、目的信号としてアレイの軸方向から音声の平面波を加えた。音声データは ATR 音声データベース B セット [8] の連続音声を用いた。観測信号の SN 比は 1 番目のマイクフォンにおいて 0.0 dB とした。ここで、本稿での SN 比の定義は、波形をベクトルと見なしたときの目的信号方向成分とその直交成分のパワー比であり、雑音抑圧量だけでなく目的信号の歪みも考慮に入れた尺度であることに注意する。データ長は 4.096 sec、サンプリング周波数は 16 kHz とし、STFT においては

* Diffuse Noise Suppression Based on the Imaginary Parts of the Inter-Channel Cross-Spectra with a Small-Sized Microphone Array by ITO, Nobutaka, ONO, Nobutaka, and SAGAYAMA, Shigeki (The University of Tokyo)

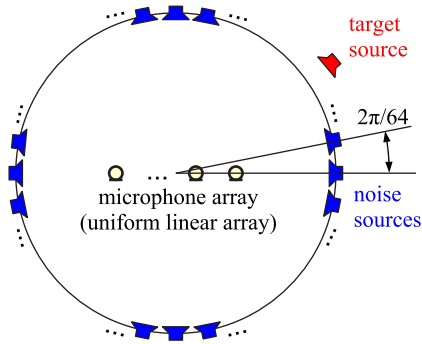


Fig. 1 Configuration of the sources and the array.

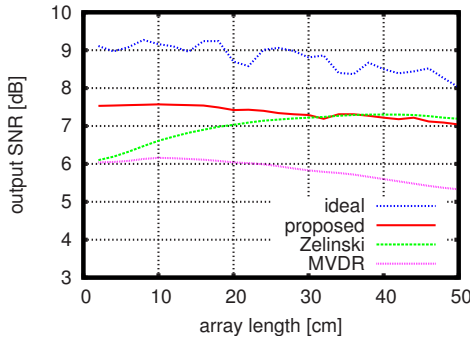


Fig. 2 Output SNR as a function of the array length.

フレーム長 512、フレームシフト 16 とし、窓関数は Hamming 窓を用いた。最小分散ビームフォーマ (3) のための Φ_{xx} は、 $x(\tau, \omega)x^H(\tau, \omega)$ の全フレームに互る時間平均により求めた。一方、ポストフィルタのための $\phi_{x_m x_n}(\tau, \omega)$ は、 $x_m(\tau, \omega)x_n^*(\tau, \omega)$ を 32 フレーム毎に時間平均して求めた。

3.2 結果

Fig. 2 はアレイ長に対して出力 SN 比をプロットしたものである。マイクロフォン数は 4 とした。提案法、Zelinski 法、最小分散ビームフォーマ単独 (ポストフィルタなし) の 3 手法を比較した。また、各マイクロフォンにおける目的信号及び雑音がそれぞれ与えられたとして計算した理想的な Wiener ポストフィルタの結果も参考のため示す。図から分かるように、アレイ長が大きい場合 (約 30 cm 以上) には提案法と Zelinski 法は同等の性能を示すが、アレイ長が小さくなると、Zelinski 法は性能が低下して最小分散ビームフォーマ単独の性能に近付いて行くのに対し、提案法の性能はほぼ不変である。これは、Zelinski 法は雑音の無相関性を仮定しているため、アレイサイズが小さいと高い雑音相関のために上手く働かないと説明できる。また、アレイサイズと並んでマイクロフォンの個数も実用上重要であるが、Fig. 3 に示すように、提案手法はマイクロフォン数が小さい場合にも良好な性能を示す。(アレイ長は 6 cm とした。) Fig. 4 はスペクトログラムの例である。Zelinski 法は雑音相関が高い低域では雑音抑圧が不十分であるが、提案法は全帯域で雑音を抑圧している。

4 まとめ

拡散性雑音抑圧のための Wiener ポストフィルタの新しい設計法として、チャンネル間クロススペクトルの虚部を用いた目的信号のパワースペクトルの推定に

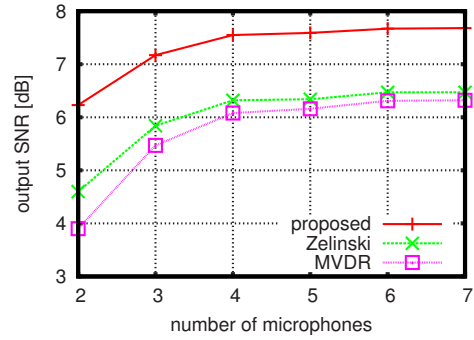


Fig. 3 Output SNR as a function of the number of microphones.

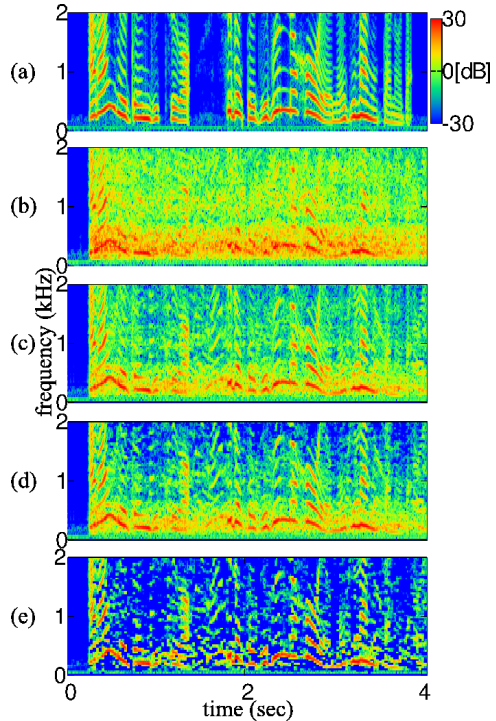


Fig. 4 Performance comparison using spectrograms. (a) Target speech; (b) observed signal (SNR: 0.0 dB); (c) minimum variance beamformer (SNR: 6.4 dB); (d) Zelinski's method (SNR: 6.6 dB); (e) proposed method (SNR: 7.8 dB).

基づく方法を提案した。本手法は任意配置かつ小サイズのアレイに対して適用できる。実験により提案法が Zelinski 法よりも高い SN 比を与えること、アレイサイズによらず有効であることが示された。

参考文献

- [1] H. Shimizu *et al.*, *Proc. WASPAA*, pp. 54–57, 2007.
- [2] N. Ito *et al.*, *Proc. ICASSP*, pp. 317–320, 2008.
- [3] N. Ono *et al.*, *Proc. SAM*, pp. 151–154, 2008.
- [4] 伊藤他, 音講論 (春), pp. 695–696, 2008.
- [5] 伊藤他, 音講論 (秋), pp. 677–678, 2008.
- [6] K. U. Simmer *et al.*, in *Microphone Arrays*, M. Brandstein *et al.*, Eds. Springer-Verlag, 2001, ch. 3, pp. 39–60.
- [7] R. Zelinski, *Proc. ICASSP*, pp. 2578–2581, 1988.
- [8] A. Kurematsu *et al.*, *Speech Commun.*, vol. 9, no. 4, pp. 357–363, 1990.