

複素スペクトル円心 (CSCC) 法による マイクロホンアレーを用いた雑音除去*

井上和士 鎌本優 岡嶋崇 西本卓也 嵯峨山茂樹 (東大・情報理工)

1 はじめに

マイクロホンアレーによる雑音低減技術の研究が盛んである。音声認識では雑音による認識性能低下が大きな問題であり、雑音の低減により性能向上が期待される。またハンズフリー電話などでは目的音の SNR の向上にマイクロホンアレー技術が有用と考えられる。

マイクロホンアレー信号処理の手法はさまざま提案されているが、Griffith-Jim[1], AMNOR[2], ICA[3] などの適応による手法ではあらかじめ雑音方向を学習することが必要である。しかし、実際の応用では雑音方向が固定でなく移動・出没する場合も多く、適応に必要な時間遅れのため十分な性能を発揮できないことがある。また、雑音が音声などの非定常信号である場合も多い。

雑音方向の移動に伴って適応学習を行わない手法としては、Delay-and-Sum(DS) があるがその性能は十分とはいえない。我々は学習なしで非線形なフィルタを構成する「複素スペクトル円心 (Complex Spectrum Circle Centroid: CSCC) 法」[4, 5] を提案し、音声認識によって効果を評価した。本稿では推定した短時間スペクトルを逆変換し音声として再合成する手法およびシミュレーションと実環境データで行った評価について述べる。

2 CSCC 法

異なる方向の目的音源及び単一の雑音源から平面波の音響信号が n 個のマイクロホンに入力すると仮定する。あるいは平面波でなく一般化して、マイクロホンごとに両者の信号間で異なる時間差をもって加算され観測されると仮定する。各マイクロホンで観測した信号にそれぞれ適当な遅延を加え目的音源からの信号の時間差を補正すれば、 i 番目のマイクロホンの観測信号 $m_i(t)$ は以下のように表せる。

$$m_i(t) = s(t) + n(t - \tau_i), \quad i = 1, 2, \dots, n \quad (1)$$

ここで $s(t)$ と $n(t)$ はそれぞれ時刻 t における目的信号と雑音信号、 τ_i は i 番目のマイクロホンでの目的信号と雑音の間の到達時間差である。これをフーリエ変換すれば、周波数 ω の成分は

$$M_i(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau_i}, \quad i = 1, 2, \dots, n \quad (2)$$

となる。式 (2) を幾何学的に解釈すると、図 1 に示すような複素平面上で全ての $M_i(\omega)$ は $S(\omega)$ を中心とした半径 $\|N(\omega)\|$ の円上に位置する。したがってこれらの円周上の点が 3 個以上あれば、円の中心が求められ、目的信号のスペクトル $S(\omega)$ が得られる。

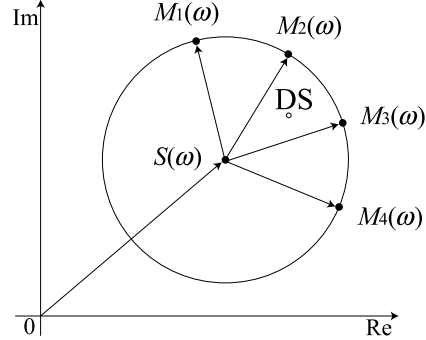


図 1: 各マイクロホンで受信した信号の周波数 ω における複素平面上での配置。図中の点 DS は Delay-and-Sum 法で得られ、 $S(\omega)$ は本手法で得られる。

この原理に基づく信号処理手法を CSCC 法と呼ぶ。一方、DS によって求められる信号の短時間スペクトルは、幾何学的には観測信号の短時間スペクトルの重心であり、一般に真のスペクトルとは異なる。

以上の原理は異なる周波数ごとに独立に成り立つ。すなわち、周波数ごとに雑音方向が異なっても良い。雑音が複数の音声であるような場合でも、それらの基本周波数やフォルマント周波数が異なれば、周波数ごとに優勢な雑音成分は一方向のみであり、効果が期待できる。

実際は、短時間フレームごとの分析、マイクロホン特性のばらつき、複数方向からの雑音などの要因により、各 $M_i(\omega)$ は上記の円上からずれて位置することが多く、さらに周波数とマイクロホン間隔の組合せによっては円上で集中することもある。そこで、本方法では円の中心 $S(\omega)$ の推定値として二乗距離分散最小点:

$$\hat{S}(\omega) = \underset{Z(\omega)}{\operatorname{argmin}} \operatorname{Var} \left[\|Z(\omega) - M_i(\omega)\|^2 \right] \quad (3)$$

を用いる。ただし $\operatorname{Ver}[\cdot]$ は \cdot の分散を表す。

多角形の外接円の中心を求める演算は雑音のスペクトルに対して非線形な処理であるので、従来の線形フィルタによる処理にはない性質を持つ。また、適応学習処理を必要とせず、フレーム単位で信号のスペクトルが推定できる。

短時間フレームごとに得られた $\hat{S}(\omega)$ をフーリエ逆変換して時間領域に戻し、フレームをずらしながら加算すれば、雑音を軽減した目的信号が得られる。

3 実験評価

3.1 シミュレーションによる実験

まず提案法を計算機上のシミュレーションで実験し、CSCC 及び DS の比較を行った。マイクロホンを直線上に配置し、素子間距離を 5cm としたアレーを構成した。目的音源及び雑音源はマイクロホンと同一の平面上にあるとし、目的音源はアレー正面から見て 0° 、妨害音源は 30° 、 60° の 2 方向から SN 比

*“Complex Spectrum Circle Centroid for Microphone-Array-Based Noise Reduction” by Kazuhito INOUE, Yutaka KAMAMOTO, Takashi OKAJIMA, Takuya NISHIMOTO and Shigeki SAGAYAMA (The University of Tokyo).

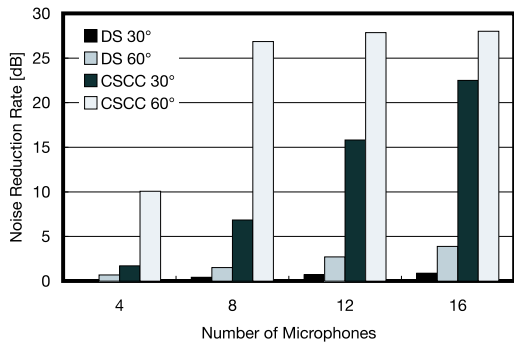


図 2: シミュレーションによる雑音低減効果: 雑音到来方向 30° 及び 60°

0dB で到来するとした．評価音声は日本音響学会の新聞記事読み上げ音声コーパスから男女 2 名を選び，また雑音として妨害音声を男女各 1 名ずつを用意し，計 8 種の組み合わせで実験を行った．短時間スペクトル分析では分析フレーム幅を 32msec，フレームシフトを 16msec，窓関数をハミング窓とした．分離性能を客観的に評価する尺度として Noise Reduction Rate(NRR = 出力 SNR - 入力 SNR) を用いた．

結果を図 2 に示す．提案法は DS に比べ，どのマイクロホンの数においても雑音の低減を行うことができた．特に雑音方向が 60° の場合に少ないマイク数でも十分な性能が得られ，最大で約 28dB に達した．

3.2 実環境データによる実験

RWCP 実環境音声・音響データベースを用い，提案法，DS に対し性能評価を行った．残響環境として E2A(残響時間 0.30sec) の特性の空間で収録された信号から，目的音源として固定音源の男女各 2 名，雑音源として移動音源の男女 1 話者を用い，計 8 通りの組み合わせを評価した．実験環境を図 4 に示す．移動音源とは図に示すように -40° から 40° まで発話中に動かしたデータである．本実験では -40° から 0° ほどの移動部分を用いた．目的音声と妨害音声を計算機上で SN 比 0dB で加算し，16kHz にダウンサンプリングしたものをマイクロホン入力信号として用いた．データベースの直線型アレー部分のみのデータを用い，目的信号に関しては全て同相であると見なして補正を行わずに実験を行った．マイクロホン特性の補正も行わなかった．

結果を図 3 に示す．マイクロホン数が 10 本以上において，提案法は DS に比べ若干の雑音抑圧効果が得られ，NRR で最大約 2.3dB の向上が見られたが，シミュレーションに比べると効果は少なかった．この理由の一つとして，残響環境では目的信号と雑音とともに様々な方向から到来するため，複素スペクトル平面で目的信号を中心とする単一円上に雑音位置するという原理からのずれが大きかったことが考えられる．また，今回の実験ではマイクロホンの較正を行っていない．各マイクロホンで観測される信号のゲインが異なると，マイクロホン入力スペクトルは単一の円上に乗らず，原理からのずれが生じ，性能の低下をもたらしたと考えられる．

また，マイクロホン数 8 本において，DS よりも性能が下回っている原因として，いわゆるミュージカルノイズによる問題が挙げられる．マイクロホン数が少ない場合，円心の推定誤りが増加し，推定スペ

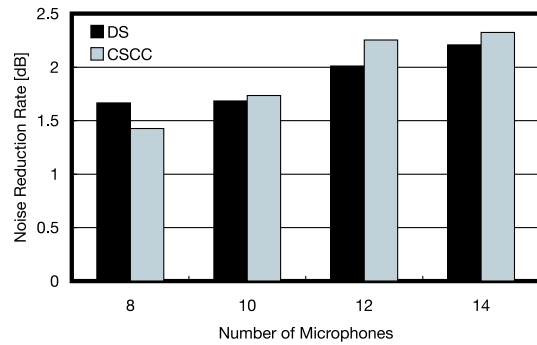


図 3: 実環境 (残響あり) における雑音低減効果

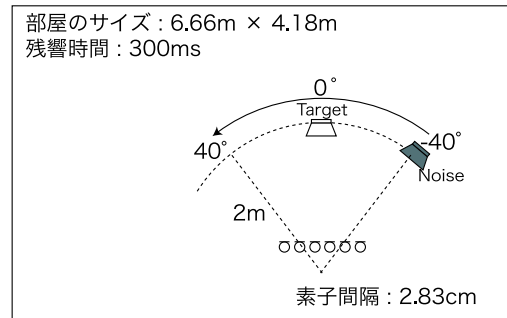


図 4: 実験環境の概念図

クトルがフレーム間で不連続になってしまうことがあった．この問題に対してはフレーム間の拘束を与えることによって，少ないマイクロホン数でも性能改善ができると期待される．

4 結論

我々が提案した学習なしで非線形なフィルタを構成する CSCC 法を用いて，目的信号の分離を行った．シミュレーション，実環境データで本手法を評価し，DS に比べ，シミュレーションでは約 28dB，実環境では約 2.3dB の NRR の向上を得た．本手法は動的に適応を行う必要は無いが，マイクロホンの周波数特性，指向特性，設置方向や位置など，本原理に関係するパラメータの事前較正は必須と考えられ，今後はこれらへの自動補償を含む手法を検討する予定である．また，実環境に対しより頑健な円心推定のアルゴリズムの検討を行う．さらに実際にマイクロホンアレーのシステムを構成し，評価実験を行う予定である．

参考文献

- [1] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag., vol. AP-30, no. 1, pp. 27-34, 1982.
- [2] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-34, no. 6, pp. 1391-1400, 1986.
- [3] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- [4] 岡嶋 鎌本 西本 嵯峨山, "マイクロホンアレー入力の周波数領域での幾何学的処理による雑音環境下での音声認識," 音講論集, vol.1, 3-10-1, pp. 583-584, 2004-Sep.
- [5] S. Sagayama, T. Okajima, Y. Kamamoto and T. Nishimoto, "Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition," Proc. ICSLP2004, to appear.