

位相差拘束付複素スペクトル円心 (CSCC) 法による 雑音抑圧の検討*

松本恭輔, 小野順貴, 嵯峨山茂樹 (東大院・情報理工)

1 はじめに

我々は近年、マイクロフォンアレイによる雑音抑圧手法として、複素スペクトル円心 (Complex Spectrum Circle Centroid: CSCC) 法を提案している [1, 2]. CSCC 法は、方向既知の目的音源と未知の単一雑音源から平面波音響信号が到来するという仮定の下で、各フレーム・周波数毎の独立な非線形演算によって雑音抑圧を行なう手法であり、音源信号のスパースな性質を利用しやすい、学習・適応が不要、突発的な雑音にも対応可能、といった特長をもつが、様々なモデル化誤差に対する頑健性を向上させることが課題であった。本稿では、マイクロフォンアレイが直線等間隔アレイの場合に着目し、新たな拘束条件を導入することで、各フレーム・周波数毎の処理であるという特徴を有したまま、頑健性を向上させる手法について報告する。

2 位相差拘束付き複素スペクトル円心法

2.1 観測モデル

本手法では、一直線上・等間隔に配置され、順に 1 から K の番号が振られた K 個のマイクロフォンが、それぞれ別の方向から到来する目的音信号及び、単一の雑音信号を受信する状況を考える。また、各信号は平面波伝播し、目的信号の到来方向は既知であると仮定する。また本稿では、各マイクロフォンで目的信号に対する同相化を行って得られる信号を観測信号として扱う。このとき、 k 番目のマイクロフォンの観測信号 $m_k(t)$ は、1 番目のマイクロフォンが受信する目的信号 $s(t)$ 、雑音信号 $n(t)$ を用いて、

$$m_k(t) = s(t) + n(t - (k-1)\tau) \quad (1)$$

と表される。ここで $(k-1)\tau$ は、 k 番目のマイクロフォンでの目的信号と雑音信号の到達時間差である。式 (1) を時間周波数分解して得られる、

$$M_k(j, \omega) = S(j, \omega) + N(j, \omega)e^{-i(k-1)\omega\tau} \quad (2)$$

を観測モデルとする。ここで、 j は時間フレーム番号、 ω は各周波数であるが、本稿では一度に複数の j, ω にまたがって議論することはない。以降では表記の煩雑さを避けるため、 (j, ω) は省略し、 $q := \omega\tau$ と表記する。また、以下のベクトル表記を用いることがある。

$$\begin{aligned} M &:= [M_1 \ \dots \ M_K]^T, \\ e_s &:= [1 \ 1 \ \dots \ 1]^T, \\ e_n &:= [1 \ e^{-iq} \ \dots \ e^{-i(K-1)q}]^T. \end{aligned} \quad (3)$$

2.2 従来の CSCC 法の原理

幾何学的な観点から式 (2) を解釈すると、観測スペクトル M は、複素スペクトル平面上で中心 S 、半径

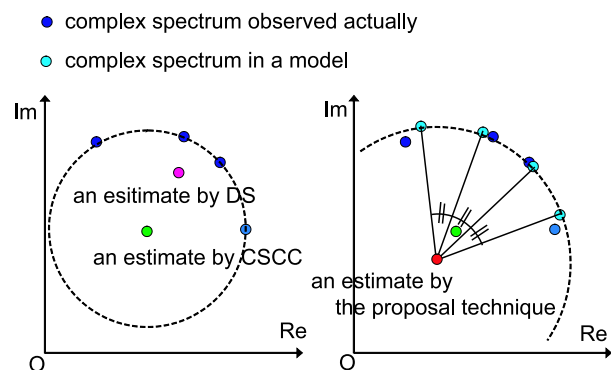


Fig. 1 CSCC 法と DS 法 (左), 提案法 (右)

$|N|$ の円上に分布することができる。よって、観測スペクトル M に円をフィッティングし円心を求めることにより、雑音抑圧が行なわれる。これが従来の CSCC 法の原理である [1].

2.3 位相差拘束付 CSCC 法

従来の CSCC 法による雑音抑圧は、半径の拘束条件のみを用いて円心 S を推定する (Fig. 1 左). この手法は円心を求めるという非線形演算が 1 つの特色であり、モデルどおりの状況では高い推定精度を発揮するが、非線形フィルタの安定性が保証されておらず、観測誤差によっては推定精度が大幅に低下することもあった。一方、式 (2) を見ると、マイクロフォンアレイが直線等間隔である場合には、雑音項の位相回転も等間隔であることがわかる。本稿での提案は、この位相差の拘束条件を制約条件として加え、複素スペクトル平面上の観測信号に対し、角度等間隔に分布する同一円周上の点をフィッティングし、その円心を求めることにより雑音抑圧を試みるものであり (Fig. 1 右)、この手法を以下では、位相差拘束付 CSCC 法と呼ぶ。

観測スペクトル M に含まれる誤差としてガウス雑音を仮定すれば、

$$J(S, N, q) := \sum_{k=1}^K |M_k - S - Ne^{-i(k-1)q}|^2 \quad (4)$$

を最小にするような S, N, q が最尤解となる。しかしながら、この問題は位相差 q に関しては非線形であり、解を得るには反復解法が必要となる。ここでは実時間での雑音抑圧への応用などを想定し、1) 別の枠組みでの位相差 q の推定、2) q が既知の元での観測モデルへのフィッティング、の 2 段階で解析解を得る手法を提案する。

Step 1: 位相差 q の推定 q の推定を容易にするために、隣接する観測信号の差信号を考え、未知数を減らすことを考える。これは式 (2) の観測モデルの下

*"Complex Spectrum Circle Centroid method with the phase difference constraint for Microphone-Array-Based noise reduction" by MATSUMOTO, Kyosuke, ONO, Nobutaka, and SAGAYAMA, Shigeki (The University of Tokyo).

では,

$$\tilde{M}_k := M_k - M_{k+1} = e^{-(k-1)iq} \{1 - e^{-iq}\} \quad (5)$$

のように表され、複素スペクトル平面上で元の観測スペクトルと同様に、同一円周上に位相差 $|q|$ 毎に分布することがわかる。よって、隣り合う \tilde{M}_k を要素に持つ複素 2 次元ベクトル

$$d_k := \begin{bmatrix} \tilde{M}_k \\ \tilde{M}_{k+1} \end{bmatrix} = e^{-i(k-1)q} N \{1 - e^{-iq}\} \begin{bmatrix} 1 \\ e^{-iq} \end{bmatrix} \quad (6)$$

は、理想的には同一の方向 $d := [1, e^{-iq}]^T =$ を向くことになる。実際の観測スペクトル M から得られる d_k の方向は誤差要因によるばらつきを持つが、おおよそ d の方向を示す傾向があると考えられる。これは d_k の共分散行列

$$V := \sum_{k=1}^{K-1} d_k d_k^T \quad (7)$$

の最大固有値 λ_a に対応する固有ベクトル d_a に反映されるため、これを $d_a = [1, z]^T$ のように正規化し、以下のように位相差 q の推定値を得る。

$$\hat{q} = \operatorname{argmin} |d_a - d|^2 = -\operatorname{arg} z \quad (8)$$

Step 2: モデルへのフィッティング Step 1 で推定した位相差 \hat{q} の元での S, N の最尤推定値は、

$$\begin{aligned} \hat{S} &= \frac{K e_s^\dagger M - (e_s^\dagger e_n)(e_n^\dagger M)}{K^2 - |e_n^\dagger e_s|^2} \\ \hat{N} &= \frac{K e_n^\dagger M - (e_n^\dagger e_s)(e_s^\dagger M)}{K^2 - |e_s^\dagger e_n|^2} \end{aligned} \quad (9)$$

と解析的に求められる。この解の形からこの手法の主な特徴として以下三点が挙げられる。

- 厳密に仮定通りの状況であれば、推定値は真のスペクトルと一致する。(導出省略)
- 分母の値は e_s / e_n のときを除き正の値をとる。そのため、2 音源の方向が異なる、という問題の仮定において、このフィルタは安定である。
- 得られた推定値は、Delay and Sum 法 (DS 法) [3] を雑音と目的信号に施した結果の重み付け和の形をした、DS 法の自然な拡張の形になっている。

3 実験と結果

3.1 実験条件

提案手法の性能を評価するために、時間周波数ごとの雑音抑圧を扱える手法として、提案手法、従来 CSCC 法、DS 法の性能を比較する。マイクロフォンと音源の配置を Fig. 2 の通り (従来 CSCC 法では 5 つ、DS 法では 7 つのマイクロフォンでも実験をおこなう) に設置し、反響音無し、球面波を仮定したシミュレーションによる複数話者の混合音声 (Target+Noise1 または Target+Noise1+Noise2) に対して雑音を抑圧する実験を行った。また、音源信号は ATR 音声データベース B セットの連続音声を、サンプリング周波数 16kHz に変換して用いた。時間周波数分解にの分析条件はどの手法も窓幅 1024 点、フレームシフト 128 点である。

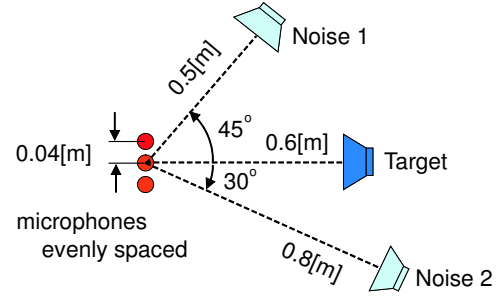


Fig. 2 音源と 3 つのマイクロフォンの配置。マイクロフォンを 5, 7 つ配置するときにはアレイの中心を一致させる。

Table 1 各手法による NRR (dB). () 内はマイクロフォンの個数を表す

	提案法 (3)	従来法 (3)	従来法 (5)	DS 法 (3)	DS 法 (7)
2 話者	+7.26	+1.83	+4.66	+0.87	+2.24
3 話者	+6.31	+0.35	+2.72	+1.73	+1.65

3.2 実験結果

各手法の性能評価指標として、100 通りの組み合わせの混合音声に対して、各手法で雑音抑圧を行ったときの NRR (Noise Reduction Ratio) の平均値を Table 1 に示す。2, 3 話者の両実験で、提案法が従来 CSCC 法よりも大きい NRR を得ているが、これは、平面波仮定に対して球面波のシミュレーションを行ったことによるモデル化誤差に対し、提案法がより頑健であることを示す。また、3 話者の混合音声の雑音抑圧においては、仮定「時間周波数毎に雑音成分は 1 つ」が崩れる可能性がある。しかし提案法は、有効に働いており、ここから、1) 3 話者の音声は時間周波数領域でほぼスパースであり、2) 提案法はその頑健性により音声のスパース性を利用した手法として有望であることが確認できる。

4 おわりに

本稿では、従来の CSCC 法に、各マイクロホン受信信号間の位相差拘束条件を導入し、時間周波数ごとに雑音を抑圧する手法を検討した。2, 3 話者混合音声に対する実験によって従来の CSCC 法に比べ推定の頑健性が向上したことを確認した。また、3 話者混合音声に対する実験結果からは、提案法は、CSCC 法の原理が時間周波数ごとの処理である特徴を活かし、音声のスパース性を利用する音源分離手法として有望であることも確認した。今後は、残響環境も含めてさらなる評価実験を行いたい。

謝辞 本研究の一部は科学研究費補助金・萌芽研究 (課題番号 17650045) の補助を受けて行なわれたので、ここに謝意を表す。

参考文献

- [1] S. Sagayama, et al., Proc. International Conference on Spoken Language Processing (ICSLP2004) (Jeju, Korea), Oct. 2004.
- [2] 井上他, 音講論 (秋), 2-3-7, pp. 619-620, Sep. 2004.
- [3] Michael Brandstein, Darren Ward, "Microphone Arrays: Signal Processing Techniques and Applications," Springer, 2001.