

複素スペクトル平面での 2ch マイクロフォンを用いた雑音除去*

大久保俊也 滝口 哲也 有木 康雄 (神戸大・工)

1 はじめに

議事録作成といった実環境での音声認識において、目的信号に重畳する各種の雑音は認識性能を劣化させる。そのため、観測信号より雑音を除去する研究が活発に行われている。

代表的な手法である Delay-and-Sum(DS) は、学習の必要はないが性能は十分でない。また、Griffith-Jim などの適応型アレーは、ある程度性能は良いものの、環境ごとに学習の必要がある [1]。

複素スペクトル平面を利用することで、学習の必要もなく精度良く目的信号を推定可能な、CSCC 法が提案されている [2]。しかし、この手法ではマイクロフォン数が最低 3 つ必要であり、また精度を高めるためにはより多くのマイクロフォンが必要となる。

これらの従来手法では、どれもマイクロフォンアレーが必要となるが、これは大掛かりな装置となる。そこで本稿では、複素スペクトル平面を利用し、マイクロフォンの個数が 2 個の場合でも、高性能に目的信号を推定する方法について検討したので報告する。

2 複素スペクトル平面を利用した雑音除去

2.1 マイクロフォン入力の複素スペクトル平面上での配置

マイクロフォンアレーのように、マイクロフォンが一行に配置されている装置に対し、正面方向に目的音源、斜め方向に単一の雑音源が存在するとする。各音源からの音響信号が平面波で到来する場合、 i 番目のマイクロフォンで観測される信号 $m_i(t)$ は、

$$m_i(t) = s(t) + n(t - \tau_i) \quad (1)$$

と表すことができる。ここで $s(t)$ と $n(t)$ はそれぞれ目的音源と雑音源の時刻 t の信号、 τ_i は i 番目のマイクロフォンでの雑音信号の時間遅れである。これを離散時間フーリエ変換すると、周波数 ω の成分は、

$$M_i(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau_i} \quad (2)$$

となる。式 (2) を複素スペクトル平面上で幾何学的に表現すると、 $M_i(\omega)$ は $S(\omega)$ を中心、 $\|N(\omega)\|$ を半径、 $\omega\tau_i$ を偏角とする円上に配置される。この観測信号の配置より、その円の位置を推定し、目的信号スペクトルである円の中心を求めるという手法が、Complex Spectrum Circle Centroid(CSCC) 法である [2]。

このような複素スペクトル平面を利用した目的信号の推定は、フレームごとに幾何学的に目的音スペクトルを求めるため、学習の必要がなく、突発性ノイズに対しても有効であるという利点がある。

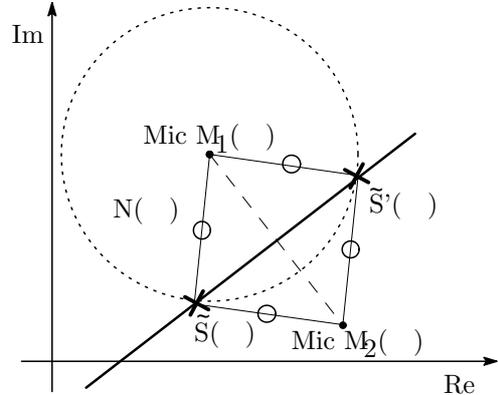


図 1: 2 つの観測信号による目的信号の幾何学的推定

2.2 2ch マイクロフォンによる目的音推定

CSCC 法は $S(\omega)$ を中心とする円の位置を推定するため、最低 3 つの観測点が必要であり、また精度を高めるためには観測点を増やさねばならず、マイクロフォンアレーのような特別な装置が必要となる。

そこで我々は 2 つの観測点のみからでも精度良く目的信号スペクトルを推定する方法を提案する。マイクロフォンが 2 個である場合、前節の議論より、 $S(\omega)$ は、 $M_1(\omega)$ と $M_2(\omega)$ からそれぞれ $\|N(\omega)\|$ の距離に位置することがわかる。 $\|N(\omega)\|$ は、 $M_1(\omega)$ と $M_2(\omega)$ を減算することで、次式のように推定することができる。

$$\|N(\omega)\| = \frac{\|M_1(\omega) - M_2(\omega)\|}{\|1 - e^{-j\omega\tau}\|} \quad (3)$$

$M_1(\omega)$ 、 $M_2(\omega)$ から等距離であることから、図 1 のように、まず $M_1(\omega)$ と $M_2(\omega)$ を結ぶ破線に対して垂直二等分線を引く。次に、 $M_1(\omega)$ を中心とする半径 $\|N(\omega)\|$ の円を描く。この垂直二等分線と円の交点に $S(\omega)$ が位置すると考えられる。垂直二等分線と円の式は、それぞれ

$$\tilde{S}_y(\omega) - C_y(\omega) = \frac{M_{1x}(\omega) - M_{2x}(\omega)}{M_{2y}(\omega) - M_{1y}(\omega)} (\tilde{S}_x(\omega) - C_x(\omega)) \quad (4)$$

$$(\tilde{S}_x(\omega) - M_{1x}(\omega))^2 + (\tilde{S}_y(\omega) - M_{1y}(\omega))^2 = \|N(\omega)\|^2 \quad (5)$$

となる。ただし、サブスクリプト x および y はそれぞれ、 $\tilde{S}(\omega)$ 、 $M_1(\omega)$ 、 $M_2(\omega)$ の実部および虚部の値であることを表す。また、 $C_x(\omega)$ および $C_y(\omega)$ は、 $M_1(\omega)$ と $M_2(\omega)$ を結ぶ破線の midpoint の実部および虚部の値である。

これらを連立させて解き、図 1 中の $\tilde{S}(\omega)$ 及び $\tilde{S}'(\omega)$ を求める。この $S(\omega)$ の候補 2 点より、本稿では、推定後の目的信号のパワーは雑音除去前よりも小さくなるであろうという予想のもと、パワーの小さい方を目的信号スペクトルとして選択する。

*Noise reduction using 2-channel microphone in complex spectrum plane.
by Toshiya Ohkubo, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

3 評価実験

3.1 実験概要

本手法の定量的評価を行うため、目的信号と雑音除去信号との相関値を求めた。更に、認識部における性能を比較するため、目的信号と雑音除去信号とのケプストラム距離 ($CepDist$) を次式により算出した。

$$CepDist = \frac{1}{T} \frac{1}{P} \sum_t \sum_p (\| S(t,p) - \tilde{S}(t,p) \|) \quad (6)$$

式 (6) 中の T は分析フレームの総数、 P はケプストラム次元数、 t, p はそれぞれ現在のフレームおよび次元を表す。相関値は 1 に近いほど目的信号成分を保持し、 $CepDist$ は値が小さいほど目的信号からのひずみが少ないことを意味する。

実験は 2.83cm 間隔で一列に並んだマイクロフォンに対し、90 度方向 (正面) から目的信号、30 度方向から単一雑音を付加した音声に対し、本手法と DS (マイクロフォン数 2 個、14 個) とで行った。なお本手法に関して、式 (3) により $\| N(\omega) \|$ を求めるために雑音源方向が必要となる。雑音源方向は CSP 法 [3] などにより推定可能であるが、本手法そのものの性能を比較するため、ここでは既知であるとした。

目的信号および雑音信号には、どちらも単語音声を用い、各信号源での SN 比が 0dB となるようにして重畳させた 10 単語を評価データとした。実験での各パラメータは、分析フレーム長を 20ms、シフト幅を 10ms、サンプリング周波数を 16kHz とし、160 点 DFT を行った。また、評価式 (6) におけるケプストラム次元数は 16 次元とした。

3.2 シミュレーションデータを用いた実験

各マイクロフォンに対して目的信号が劣化無く同時に到達し、雑音信号もマイクロフォン間での遅延のみが起る場合を考える。

結果を表 1 に示す。これより本手法が DS と比較して、性能を改善することが示された。また、本手法により雑音除去を行った信号の波形を図 2 に示す。表 1 の相関値からもわかるように、本手法によりほぼ完全に目的信号を復元できた。

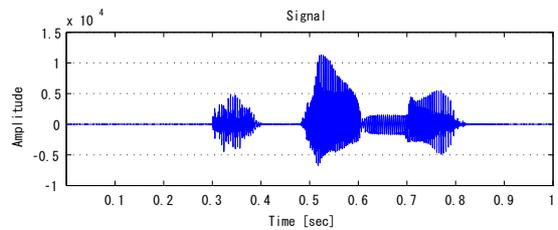
表 1: シミュレーションデータによる比較

	観測信号	DS(2)	DS(14)	本手法
相関値	0.69	0.71	0.81	0.97
$CepDist$	9.32	9.11	6.92	4.82

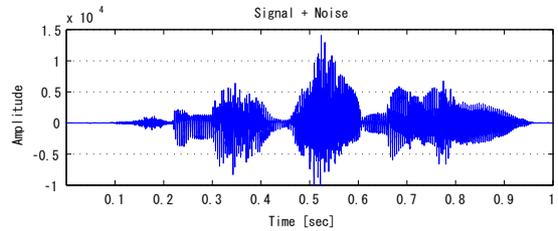
3.3 残響下実験

残響下実験では、RWCP 実環境音声・音響データベース内の実環境インパルス応答を 16kHz に変換後、目的信号および雑音信号のそれぞれに畳み込み、各マイクロフォンごとに加算することで観測信号データを作成した。今回使用したインパルス応答は残響時間 300ms、音源-マイクロフォン間 2m である。

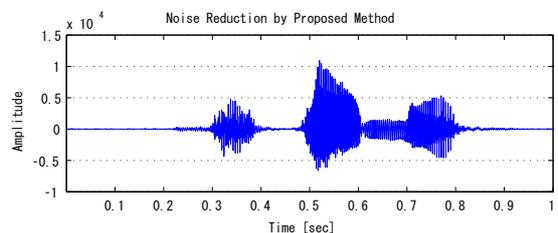
なお、この場合の相関値および $CepDist$ の計算では、目的信号として、目的信号にインパルス応答を畳み込んだ信号を用いた。これは、雑音除去の性能について比較するため、空間伝播による信号の劣化が性能比較の障害となることを避けるためである。



(a) 目的信号



(b) 雑音重畳信号



(c) 雑音除去後の信号

図 2: 本手法による雑音除去結果

表 2: 残響環境下データによる比較

	観測信号	DS(2)	DS(14)	本手法
相関値	0.66	0.67	0.73	0.49
$CepDist$	10.05	9.96	8.60	9.98

この結果を表 2 に示す。本手法における相関値および $CepDist$ は、DS と同程度にまで劣化した。これは残響の影響により、2 つのマイクロフォンで観測される信号の大きさが各周波数スペクトルにおいて異なり、 $S(\omega)$ の推定誤りが起こったためと考えられる。

4 おわりに

本研究では、複素スペクトル平面上の 2ch マイクロフォンによる観測信号スペクトルより、目的信号スペクトルを推定する手法を提案した。シミュレーションによる実験の結果、DS に比べ高性能に目的信号を復元できることを確認した。一方で、残響下においては性能が劣化することも確認した。今後は、残響下における影響について詳しく実験を行い、実環境における本手法の性能の改善を目指す。また、雑音源方向の推定などの処理も加える予定である。

参考文献

- [1] 大賀寿郎, 山崎芳男, 金田豊, “音響システムとデジタル処理,” 電子情報通信学会, 1995.
- [2] S.Sagayama, et al, “Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition,” Proc. ICSP2004, 2004.
- [3] M.Omologo and P.Svaizer, “Acoustic Event Localization in Noisy and Reverberant Environment Using CSP Analysis,” Proc. ICASSP96, pp.921-924, 1996.