

実環境における 2ch マイクロフォンを用いた 複素スペクトル平面での雑音除去に関する検討*

大久保俊也 滝口 哲也 有木 康雄（神戸大・工）

1 はじめに

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信号は認識性能を大きく劣化させる．そのため、観測信号より雑音信号を除去することが重要である．

雑音除去法として、複素スペクトル平面を利用することで、学習の必要がなく精度良く目的信号を復元可能な、CSCC 法 [1] がある．しかし、この手法ではマイクロフォン数が 3 つ以上必要であり、これは装置として大掛かりとなるため、適用する状況が制限される恐れがある．これに対し我々は、マイクロフォン 2 個の場合でも、複素スペクトル平面を利用して目的信号を復元する雑音除去法を提案した [2]．しかしながら提案手法は、理想データにおいては精度良く雑音除去が可能であるが、実環境データを用いた実験において、十分な効果を示さなかった．

そこで本稿では、実環境での 2 つのマイクロフォン間で観測される雑音信号の大きさの違いに着目し、雑音信号の比率を調整することで性能改善を行った結果について報告する．

2 2ch マイクロフォンによる複素スペクトル平面上での雑音除去

2.1 理想データでの目的信号スペクトルの推定方法

2 つのマイクロフォンに対し、正面方向に目的音源、正面以外の方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、これら 2 つのマイクロフォンで観測される信号 $m_1(t)$ および $m_2(t)$ は、

$$\begin{cases} m_1(t) = s(t) + n(t) \\ m_2(t) = s(t) + n(t - \tau) \end{cases} \quad (1)$$

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 はマイクロフォン間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、それぞれ、

$$\begin{cases} M_1(\omega) = S(\omega) + N(\omega) \\ M_2(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau} \end{cases} \quad (2)$$

となる．これら観測信号スペクトルは、複素スペクトル平面上で幾何学的に表現すると、 $S(\omega)$ を中心とし、 $\|N(\omega)\|$ を半径とする円上に配置される [1]．

ここで、 $\|N(\omega)\|$ 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、 $M_1(\omega)$ と $M_2(\omega)$ の垂直二等分線と、 $M_1(\omega)$ を中心とする半径 $\|N(\omega)\|$ の円との交点に $S(\omega)$ が位置すると考えられる．よって、それらを連立させて解き、目的信号スペクトル $S(\omega)$ の候補 2 点を求め、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．

2.2 マイクロフォン間での雑音信号の比率調整

前述した提案手法では、実環境にて録音したデータを用いた場合、性能が大きく劣化した [2]．この原因として、実環境で録音する場合、雑音信号が 2 つのマイクロフォン間を伝播する間に減衰し、各マイクロフォン間で雑音信号の大きさに差異が生じることがあげられる．そこで、マイクロフォン間での雑音比率を考慮して推定を行うことで、提案手法の実環境での性能が改善されると期待できる．

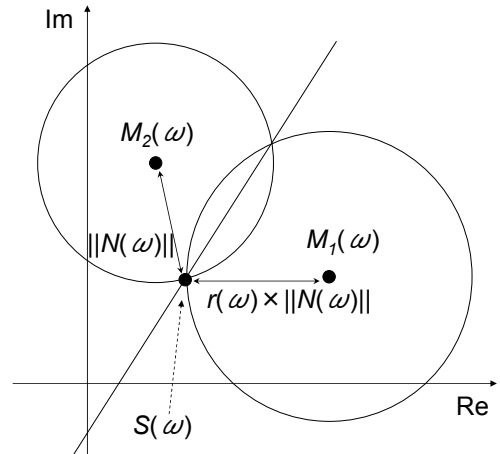


図 1: 雑音比率調整による目的信号スペクトルの推定

図 1 に、雑音比率を考慮したときの複素スペクトル平面上での各スペクトルの配置関係を示す．雑音源が $M_1(\omega)$ 側のマイクロフォンに近い場合に、 $M_2(\omega)$ 側のマイクロフォンで観測される雑音信号スペクトルの大きさ $\|N_2(\omega)\|$ を $\|N(\omega)\|$ として雑音比率を考えると、式 (3) の雑音推定式は次のようになる．

$$\|N(\omega)\| = \frac{\|M_1(\omega) - M_2(\omega)\|}{\|r(\omega) - e^{-j\omega\tau}\|} \quad (4)$$

ここで $r(\omega)$ は、 $M_1(\omega)$ と $M_2(\omega)$ で観測される雑音信号スペクトル $N_1(\omega)$ と $N_2(\omega)$ の周波数 ω 毎の比率を表わし、次のようにして求められる．

$$r(\omega) = \frac{\|N_1(\omega)\|}{\|N_2(\omega)\|} \quad (5)$$

*A study about noise reduction in real environment using a 2-channel microphone in a complex spectrum plane.
by Toshiya Ohkubo, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

また、2つの円の交点を通る直線および $M_1(\omega)$ を中心とする半径 $r(\omega) \parallel N(\omega)$ の円の式は、それぞれ、

$$\begin{aligned} \tilde{S}_y(\omega) - \frac{M_{1x}^2 - M_{2x}^2 + M_{1y}^2 - M_{2y}^2 - (r(\omega) \parallel N(\omega))^2}{2(M_{1y}(\omega) - M_{2y}(\omega))} \\ = \frac{M_{1x}(\omega) - M_{2x}(\omega)}{M_{2y}(\omega) - M_{1y}(\omega)} \tilde{S}_x(\omega) \quad (6) \end{aligned}$$

$$\begin{aligned} (\tilde{S}_x(\omega) - M_{1x}(\omega))^2 + (\tilde{S}_y(\omega) - M_{1y}(\omega))^2 \\ = (r(\omega) \parallel N(\omega))^2 \quad (7) \end{aligned}$$

となる。ただし、サブスクリプト x および y はそれぞれ、目的信号スペクトルの候補 $\tilde{S}(\omega)$, $M_1(\omega)$, $M_2(\omega)$ の実部および虚部の値であること表す。式 (6) と式 (7) を連立させて $\tilde{S}(\omega)$ を求め、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。

これにより、2ch マイクロフォン間隔が一定であれば、雑音信号の到来方向毎のマイクロフォン間での雑音比率 $r(\omega)$ のみを保持し、CSP 法 [3] などにより雑音到来方向を推定し、対応する $r(\omega)$ を選択することで、雑音除去が可能であると考えられる。

3 評価実験

3.1 実験概要

本実験におけるデータの収録条件を表 1 に示す。まず雑音信号のみをスピーカから流し、各マイクロフォンで観測した信号を用いて雑音比率 $r(\omega)$ を計算する。ここでは、雑音信号がマイクロフォン列に対して 0 度方向 (真横) から到来するとした。また、この雑音信号にはホワイトノイズを用いた。比率は各フレームごとに求めた値を全フレーム数で平均した。

次に、目的信号として音声を 90 度方向 (正面) から、雑音信号として母音音声を 0 度方向 (真横) から同時にスピーカで流し、マイクロフォンで収録した。これらのデータに対して、先に求めた雑音比率を用いて提案手法を適用し、雑音除去を行った。このときの分析パラメータを表 2 に示す。

3.2 実験結果

雑音除去結果を図 2 に示す。提案手法による結果 (図 2(c)) との比較のため、マイクロフォン数が提案手法と同じ 2 個の場合の DS による結果 (図 2(b)) も示す。これにより、提案手法は DS と比較して雑音除去の性能が高いということがわかる。

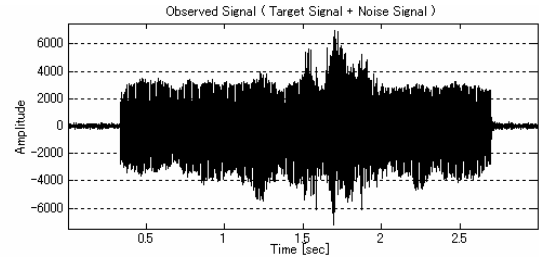
しかしながら、提案手法においても、雑音信号が完全に除去できていない。これは、ホワイトノイズにより計算した雑音比率 $r(\omega)$ が、実際の雑音信号の比率とは異なり、目的信号スペクトルの推定に誤差が生じているためであると考えられる。

表 1: 収録条件

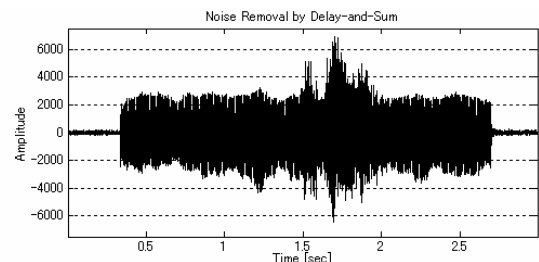
収録環境	研究室内 (定常雑音有)
スピーカ	Logitech SP-610L
マイクロフォン	SONY ECM-66B
サンプリング	32kHz 16bit
マイクロフォン間隔	2cm
音源-マイク間距離	15cm

表 2: 分析パラメータ

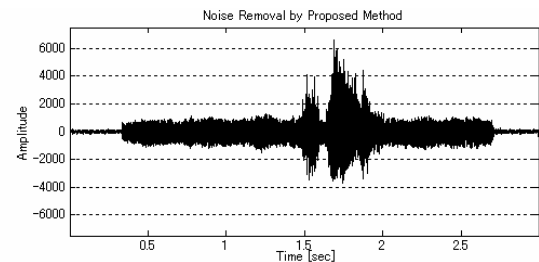
サンプリング周波数	16kHz
フレーム長	20ms
シフト幅	10ms
窓関数	ハミング窓
周波数変換	320 点 DFT



(a) 観測信号



(b) DS による雑音除去結果



(c) 提案手法による雑音除去結果

図 2: 実験結果 (目的信号/tadayou/, 雑音信号/i/)

4 おわりに

本研究では、2ch マイクロフォンによる複素スペクトル平面上での目的信号スペクトルの推定手法について、雑音信号の比率を調整することでその性能を改善する方法を提案した。実験結果より、実環境においても提案手法を用いることで雑音を抑圧できることを確認したが、雑音信号は完全には除去しきれなかった。これは、雑音信号の比率が実際とは異なることが原因であると考えられる。

今後、この雑音信号の比率についてより詳細に検討を行い、提案手法の実環境での性能改善を目指す。

参考文献

- [1] S.Sagayama, et al, "Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition," Proc. ICSLP2004, 2004.
- [2] 大久保俊也, 滝口哲也, 有木康雄, "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," 日本音響学会講演論文集, 2-Q-13, pp.123-124, 2005.3.
- [3] M.Omologo and P.Svaizer, "Acoustic Event Localization in Noisy and Reverberant Environment Using CSP Analysis," Proc. ICASSP96, pp.921-924, 1996.