

1 はじめに

近年提案された雑音除去法に、複素スペクトル平面を利用することで、学習の必要なく、精度良く目的信号を復元可能な、CSCC 法 [1] がある。しかし、この手法はマイクロフォンが3個以上必要であり、装置として大掛かりになるという問題がある。これに対し我々は、マイクロフォン2個の場合でも CSCC 理論を用いて目的信号を復元する雑音除去法を提案 [2] し、更にマイクロフォン間の雑音振幅補正を行うことで実環境データに対しての性能を改善した [3]。しかしながら、理想データの場合と比較すると、十分な性能が得られていない。

本稿では、マイクロフォン間の雑音振幅比率を、事前に求めた平均値ではなく、フレームごとに推定することで、実環境での性能改善を目指した。

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と仮定した場合、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。これを離散フーリエ変換 (DFT) すると、各周波数 ω で、

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

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。

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

すると、目的信号スペクトル $S(\omega)$ は $M_1(\omega)$ および $M_2(\omega)$ から等距離であることから、 $M_1(\omega)$ 、 $M_2(\omega)$ を中心とする半径 $\|N(\omega)\|$ の円の交点に位置すると

*A noise reduction based on the amplitude modification between two microphones in a complex spectrum plane.
by Toshiya Ohkubo, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

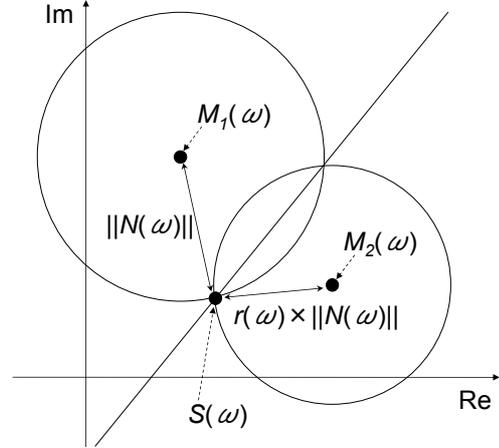


図 1: 雑音比率調整による目的信号スペクトルの推定考えられる。この交点は通常2点存在するが、推定後の目的信号のパワーは雑音除去前よりも小さくなるであろうという予想のもと、パワーの小さい方を目的信号スペクトルとして選択する。

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、各マイクロフォン間での信号振幅の差異があげられる。そこで、マイクロフォン間での振幅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 [3]。

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。 $M_1(\omega)$ 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を $\|N(\omega)\|$ とすると、式 (3) の雑音推定式は振幅比率 $r(\omega)$ を用いて次のように表わされる。

$$\|N(\omega)\| = \frac{\|M_1(\omega) - M_2(\omega)\|}{\|1 - r(\omega)e^{-j\omega\tau}\|} \quad (4)$$

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 $\|N(\omega)\|$ の円の式は、それぞれ、

$$\begin{aligned} \tilde{S}_y + \frac{M_{1x} - M_{2x}}{M_{1y} - M_{2y}} \tilde{S}_x \\ = \frac{M_{1x}^2 - M_{2x}^2 + M_{1y}^2 - M_{2y}^2 - (1 - r^2)(\|N\|)^2}{2(M_{1y} - M_{2y})} \end{aligned} \quad (5)$$

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

となる (式 (5) 中 (ω) を省略)。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を表す。式 (5) と式 (6) 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。

2.3 雑音信号の比率の逐次推定

前回の報告では、2.2節の各式における $r(\omega)$ を、ホワイトノイズを観測したデータを用いて予め計算した平均比率を用いていた。しかしながら、事前に比率学習を行う必要があり、また、周波数によっては、時間軸に対して比率が一定ではない、といった問題があった。

そこで今回、 $r(\omega)$ を各マイクロフォンの観測値の比とし、フレームごとに次式のように求めた。

$$r(\omega) = \frac{\|M_2(\omega)\|}{\|M_1(\omega)\|} \quad (7)$$

この比率を用いることで、雑音のみの区間においては正確に比率を求めることができる。

3 評価実験

3.1 実験方法

本実験での実験条件を表1に示す。実環境データは、RWCP実環境音声・音響データベース[5]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により作成した。

これらのデータに対して、各フレームごとに式(7)を用いて雑音比率 $r(\omega)$ を推定し、提案手法を適用した。評価指標には、目的信号と処理後の信号との相関係数を用いた。なお、信号到来方向は既知であるとして実験を行った。

3.2 実験結果

雑音除去結果を表2に、処理結果の一例を図2に示す。これらより、振幅補正を行うことで提案手法の性能が大きく改善されることがわかる。また、表2において、平均値補正に比べ、逐次推定値補正を行った場合に、若干の性能の改善が見られた。

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雑音除去法について、雑音信号の比率を逐次推定することで性能を改善する方法を提案した。実験結果より、平均比率を用いる場合に比べ、逐次比率を用いる方が、相関係数で0.03ポイントの改善が得られた。

参考文献

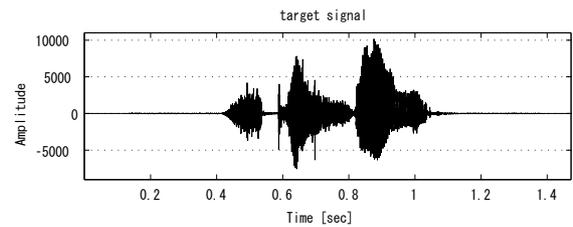
- [1] S.Sagayama, et al, "Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition," Proc. ICSLP2004, 2004.
- [2] 大久保俊也, 滝口哲也, 有木康雄, "複素スペクトル平面での2chマイクロフォンを用いた雑音除去," 日本音響学会講演論文集, 2-Q-13, pp.123-124, 2005.3.
- [3] 大久保俊也, 滝口哲也, 有木康雄, "実環境における2chマイクロフォンを用いた複素スペクトル平面での雑音除去に関する検討," 日本音響学会講演論文集, 1-P-25, pp.201-202, 2005.9.
- [4] M.Omologo and P.Svaizer, "Acoustic Event Localization in Noisy and Reverberant Environment Using CSP Analysis," Proc. ICASSP96, pp.921-924, 1996.
- [5] S. Nakamura, "Acoustic sound database collected for hands-free speech recognition and sound scene understanding," International Workshop on Hands-Free Speech Communication, pp. 43-46, 2001.

表 1: 実験条件

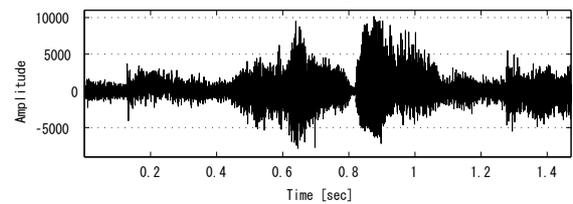
標本化	16kHz	量子化	16bit
フレーム幅	32ms	シフト幅	16ms
窓関数	Hamming	DFT	512点
マイク間隔	11.48cm	音源-マイク	2m
音源 SN 比	0dB	残響時間	0s(無響室)
目的信号	音声 (10 単語 × 3 話者), 90°方向		
雑音信号	音楽 (2 種類), 0°方向		

表 2: 実験結果 (相関係数)

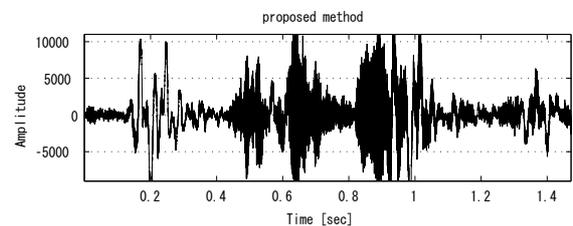
観測信号	補正無し	平均値補正	逐次推定補正
0.78	0.51	0.83	0.86



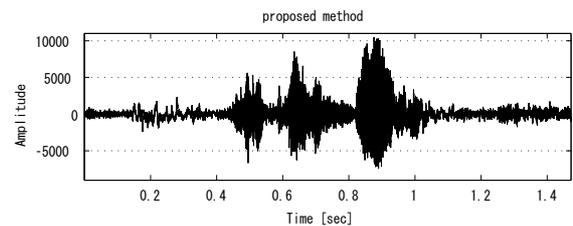
(a) 目的信号



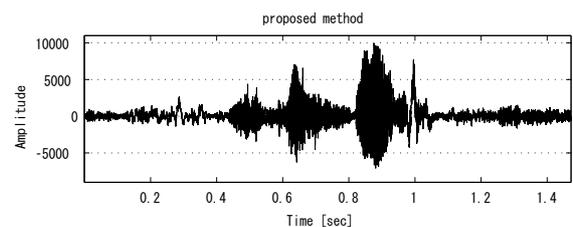
(b) 観測信号



(c) 提案手法 (補正無し)



(d) 提案手法 (平均値補正)



(e) 提案手法 (逐次推定補正)

図 2: 実験結果 (目的信号/shikisai/, 雑音信号 (音楽))