

様々なミュージカルノイズフリー音声強調法における音質評価*

◎宮崎亮一, 猿渡洋, 中村哲 (奈良先端大), 近藤多伸 (ヤマハ)

1 はじめに

近年, テレビ会議システムや音声検索といった音声通信を利用したアプリケーションが増加している. しかし, 実環境においては目的音声以外の雑音が多く存在し, それがマイクロホンに混入するため, 目的音声の品質が劣化するという問題がある. そこで, 高精度に雑音抑圧し音声を強調する音声強調技術の実現が望まれている. 代表的な単一チャンネル非線形音声強調手法として, スペクトル減算法 (spectral subtraction: SS) [1], ウィーナフィルタ (Wiener filtering: WF) [2], 最小平均二乗誤差規範短時間振幅スペクトル (minimum mean-square error short-time spectral amplitude: MMSE STSA) 推定器 [3] が提案されている. しかし, これらの手法においては, ミュージカルノイズと呼ばれる非線形処理特有の歪みが生じ, 致命的な音質劣化を引き起こすことが知られている.

ミュージカルノイズの発生を抑えた高品質な音声強調手法として, 弱い非線形処理を繰り返し行う反復型 SS や反復型 WF が提案されている [4, 5, 6]. また, これらの反復型手法において, 雑音は抑圧されるがミュージカルノイズが全く発生しない適切な内部パラメータを筆者らによって数学的に明らかにされており, 「ミュージカルノイズフリー音声強調」という新たな音声強調手法として提案されている [8, 9]. しかし, MMSE STSA 推定器におけるミュージカルノイズフリー状態は確認されていない. 一般的に MMSE STSA 推定器は, 等雑音抑圧量 (noise reduction rate: NRR) 条件下で, SS や WF 等の単一チャンネル非線形音声強調手法と比べて音声歪み量の少ない高性能な音声強調手法であることが知られている. そのため, MMSE STSA 推定器にミュージカルノイズフリー理論を適用した手法が望まれる.

そこで本稿では, MMSE STSA 推定器におけるミュージカルノイズフリー状態の存在を実験的に示す. そして, 実験的に得たミュージカルノイズフリー状態を達成するパラメータを用いて, MMSE STSA 推定器にミュージカルノイズフリー理論を適用した手法を提案する. 客観評価実験より, 適切なパラメータを与えた提案手法は, 音声強調処理を行なってもミュージカルノイズがほとんど発生しないことを確認した. さ

らに, 従来の SS や WF にミュージカルノイズフリー理論を適用した手法と比べて, 音声歪みの面で有利であることを確認した.

2 先行研究

2.1 スペクトル減算法 [1]

目的音声信号と雑音信号が混合した入力観測信号に対し, 短時間フーリエ解析を行い, 時間周波数信号系列を得る. 時間周波数領域における SS は以下で定義される [1].

$$Y_{SS}(f, \tau) = \begin{cases} \sqrt{|X(f, \tau)|^2 - \beta E_{\tau}[\hat{N}(f, \tau)]^2} \exp(j \arg(X(f, \tau))) \\ \text{(if } |X(f, \tau)|^2 - \beta E_{\tau}[\hat{N}(f, \tau)]^2 > 0) \\ \eta X(f, \tau) \text{ (otherwise)} \end{cases} \quad (1)$$

ここで, $Y_{SS}(f, \tau)$ は SS 処理後の音声強調信号, $X(f, \tau)$ は入力観測信号, $\hat{N}(f, \tau)$ は推定雑音信号, f は周波数ビン, τ は時間フレームインデックス, $E_{\tau}[\cdot]$ は τ における期待値演算, β は減算係数, η はフロアリング係数を表す.

2.2 Wiener filtering [2]

本節では, WF 族の中でも最も一般的な standard WF を導入する. 一般的に WF は次式のように定義される.

$$Y_{WF}(f, \tau) = G_{WF} |X(f, \tau)| \exp(j \arg(X(f, \tau))) \quad (2)$$

ここで, $Y_{WF}(f, \tau)$ は WF 処理後の音声強調信号, また, $G_{WF}(f, \tau)$ は次式で定義される WF のゲイン関数である.

$$G_{WF}(f, \tau) = \begin{cases} \frac{|X(f, \tau)|^2 - \beta \cdot E_{\tau}[\hat{N}(f, \tau)]^2}{|X(f, \tau)|^2} \\ \text{(if } |X(f, \tau)|^2 - \beta \cdot E_{\tau}[\hat{N}(f, \tau)]^2 > 0) \\ \eta \text{ (otherwise)} \end{cases} \quad (3)$$

2.3 MMSE STSA 推定器 [3]

MMSE STSA 推定器は一般的に次式で定義される [3].

$$Y_{MMSE}(f, \tau) = G_{MMSE} |X(f, \tau)| \exp(j \arg(X(f, \tau))) \quad (4)$$

* "Evaluation of various musical-noise-free speech enhancement methods," by Ryoichi Miyazaki, Hiroshi Saruwatari, Satoshi Nakamura (Nara Institute of Science and Technology), and Kazunobu Kondo (YAMAHA Corporate Research & Development Center).

ここで、 $Y_{\text{MMSE}}(f, \tau)$ は MMSE STSA 推定器を用いて得た音声強調信号である。また、 $G_{\text{MMSE}}(f, \tau)$ は MMSE STSA 推定器のゲイン関数であり、次式で定義される [3]。

$$G_{\text{MMSE}}(f, \tau) = \Gamma(1.5) \frac{\nu(f, \tau)}{\gamma(f, \tau)} \exp\left(-\frac{\nu(f, \tau)}{2}\right) \left[(1 + \nu(f, \tau)) I_0\left(\frac{\nu(f, \tau)}{2}\right) + \nu(f, \tau) I_1\left(\frac{\nu(f, \tau)}{2}\right) \right] \quad (5)$$

ここで、 I_0, I_1 はそれぞれ 0 次と 1 次のベッセル関数、 $\Gamma(\cdot)$ はガンマ関数、また、 $\nu(f, \tau) = \gamma(f, \tau)\xi(f, \tau)/(1 + \xi(f, \tau))$ であり、 $\xi(f, \tau), \gamma(f, \tau)$ はそれぞれ次式で定義される事前 SNR と事後 SNR である。

$$\xi(f, \tau) = E[|S(f, \tau)|^2]/E_{\tau}[|\hat{N}(f, \tau)|^2] \quad (6)$$

$$\gamma(f, \tau) = |X(f, \tau)|^2/E_{\tau}[|\hat{N}(f, \tau)|^2] \quad (7)$$

式 (6) と式 (7) において、一般的に非音声区間における雑音のパワースペクトルの期待値によって $E_{\tau}[|\hat{N}(f, \tau)|^2]$ を推定することができる。しかし、 $E[|S(f, \tau)|^2]$ を推定することはできないため、事前 SNR $\xi(f, \tau)$ は次式で定義される decision-directed 法を用いて近似的に計算される [3]。

$$\hat{\xi}(f, \tau) = \alpha\gamma(f, \tau - 1)G_{\text{MMSE}}^2(f, \tau - 1) + (1 - \alpha)\max[\gamma(f, \tau) - 1, 0] \quad (8)$$

ここで、 α は忘却係数である。

2.4 ミュージカルノイズフリー音声強調 [8, 9]

近年、Uemura らによって、音声強調によるミュージカルノイズの発生量は信号処理前後のカートシスを用いることで定量的に評価可能であるということが述べられている [7]。この尺度は値が 1 の時はミュージカルノイズが全く発生していないことを表し、値が大きくなるほど信号処理によってミュージカルノイズが多く発生したことを表す。また筆者らは、SS や WF 処理前後におけるカートシスの不動点を数学的に導出することによって、雑音は抑圧されるがミュージカルノイズが全く発生しない、ミュージカルノイズフリー条件を満たす内部パラメータを明らかにした [8, 9]。白色ガウス雑音の場合の SS と WF におけるミュージカルノイズフリー条件を満たすパラメータセットの例を Fig. 1 に示す。なお、このミュージカルノイズフリー条件の導出は既発表であるため、詳細は文献 [8, 9] を参照されたい。

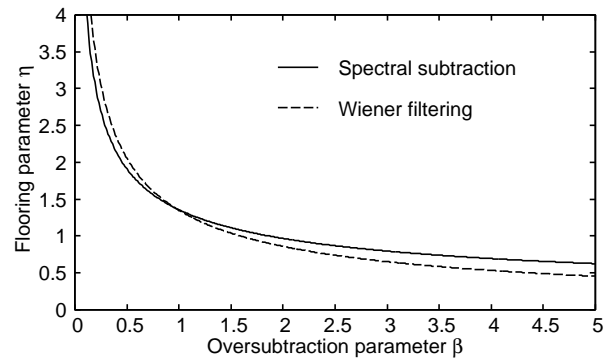


Fig. 1 Example of oversubtraction parameter β and flooring parameter η for spectral subtraction and Wiener filtering, which satisfy musical-noise-free condition.

3 ミュージカルノイズフリー MMSE STSA 推定器

3.1 MMSE STSA 推定器の雑音抑圧量とミュージカルノイズ発生量の関係 [10]

近年、Kanehara らによって MMSE STSA 推定器における雑音抑圧性能とミュージカルノイズ発生量の関係が、高次統計量に基づく理論解析によって明らかにされた [10]。Figure 2 に高次統計量に基づく理論解析より得られた MMSE STSA 推定器における忘却係数と雑音抑圧性能・ミュージカルノイズ発生量の関係を示す。本稿では、雑音抑圧性能として信号処理前後の SNR の差で定義される雑音抑圧量 (noise reduction rate: NRR) を用いる。Figure 2(a) より、忘却係数 α の増加とともに NRR が増加することが確認できる。一方、Fig. 2(b) より、忘却係数 α が 0.98 以降で急激にカートシス比が減少することが確認できる。しかし、忘却係数 α を限りなく 1.0 に近づけてもカートシス比が 1.0、すなわちミュージカルノイズフリー状態を達成することはない。

3.2 バイアス付き事前 SNR 推定 [2]

MMSE STSA 推定器におけるミュージカルノイズ発生量を減らすために、式 (8) の事前 SNR 推定にバイアスを加えた、バイアス付き事前 SNR 推定が提案されている [2]。バイアス付き事前 SNR 推定は次式で定義される。

$$\hat{\xi}(f, \tau) = \alpha\gamma(f, \tau - 1)G_{\text{MMSE}}^2(f, \tau - 1) + (1 - \alpha)\max[\gamma(f, \tau) - 1, \xi_{\min}] \quad (9)$$

ここで、 ξ_{\min} は最小許容値であり、このパラメータを増加することによってミュージカルノイズ発生量が減少することが知られている [2]。以降、このバイアス付き事前 SNR を

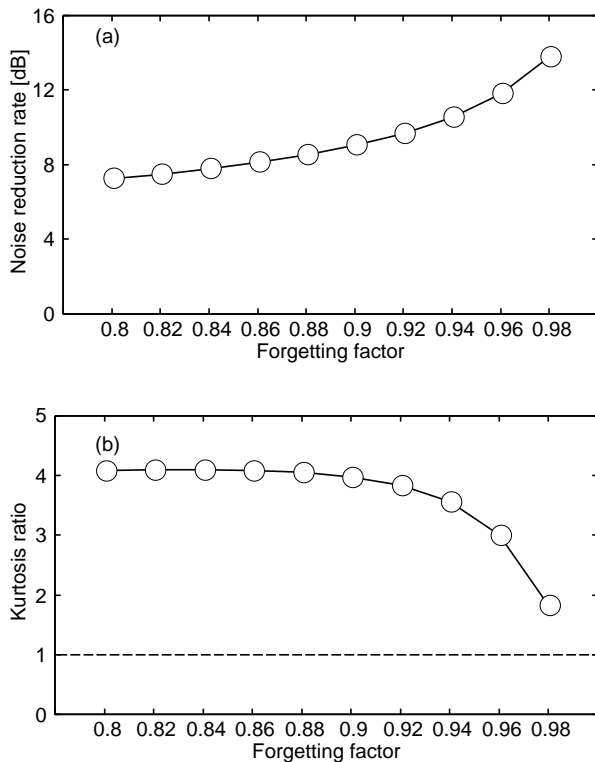


Fig. 2 Theoretical behavior for MMSE STSA estimator: (a) noise reduction rate and (b) kurtosis ratio.

導入した MMSE STSA 推定器をバイアス付き MMSE STSA 推定器と呼ぶ。

3.3 バイアス付き MMSE STSA 推定器の雑音抑圧量とミュージカルノイズ発生量の関係

本節では、前節で述べたバイアス付き MMSE STSA 推定器の雑音抑圧性能とミュージカルノイズ発生量の関係について述べる。白色ガウス雑音を用いて、最小許容値 ξ_{\min} を 0 から 1 まで変化させたときの NRR とカートシス比の関係を Fig. 3 に示す。Figure 3 より、最小許容値 ξ_{\min} を増加させると、それに伴って NRR とカートシス比が減少することが確認できる。さらに、NRR が 0 dB 以上かつカートシス比が 1、すなわちミュージカルノイズフリー状態を達成するパラメータが存在することが確認できる。

4 客観評価実験

4.1 客観評価実験条件

提案手法の妥当性を確認するために、客観評価実験を行った。テストセットの入力 SNR は 0, 5, 10 dB に設定し、目的音声信号は新聞記事読み上げ音声コーパスの 10 文（男性 5 名、女性 5 名）を用いた。雑音信号は白色ガウス雑音とバブル雑音の 2 種類を用いた。なお、バブル雑音は 36 個

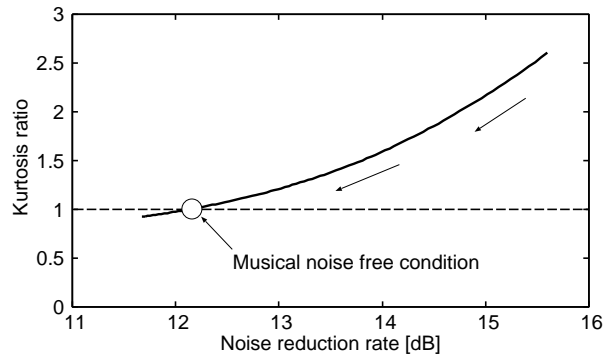


Fig. 3 Relation between NRR and kurtosis ratio obtained from objective evaluation experiment with increasing ξ_{\min} .

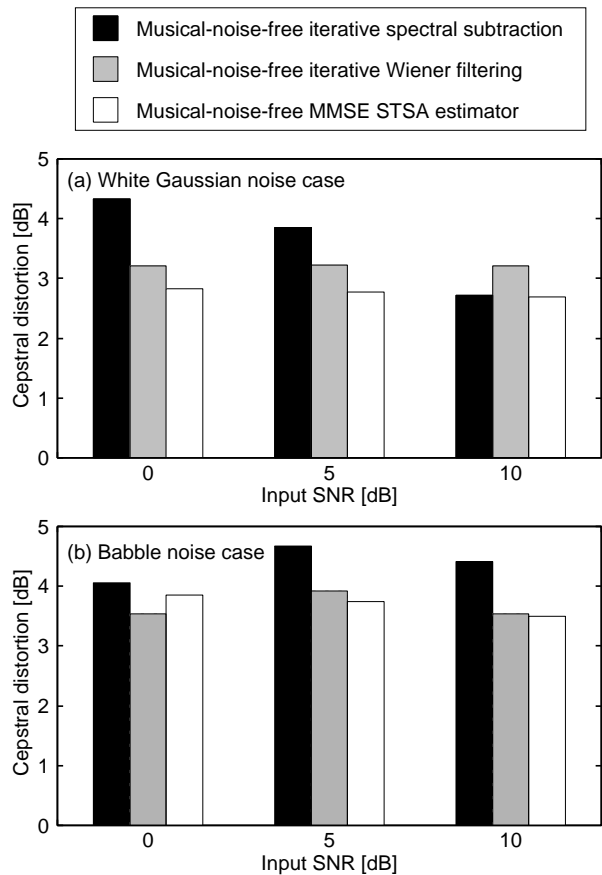


Fig. 4 Cepstral distortion obtained from experiment with (a) white Gaussian noise and (b) babble noise under 12-dB NRR condition, where we compare musical-noise-free iterative SS, musical-noise-free iterative WF, and musical-noise-free MMSE STSA estimator.

のラウドスピーカから 36 種の音声を出力したものを無指向性の単一マイクロホンで収録したものである。また、提案法の比較手法としてミュージカルノイズフリー反復型 SS とミュージカルノイズフリー反復型 WF を用いた。客観評価値として、音声強調処理によって発生した音声歪み量を測るケプストラム歪みを用いた。

4.2 客観評価実験結果

本節では、等 NRR 条件下におけるミュージカルノイズフリー音声強調手法の比較を行う。Figure 4 に、それぞれの音声強調手法における音声強調処理後のケプストラム歪みを表す。なお、ケプストラム歪みは、数値が小さいほど処理信号における音声部分の劣化程度が少ないことを表す。それぞれの手法において NRR は 12 dB であり、忘却係数 α は 0.98 に設定した。Figure 4 より、提案手法は従来のミュージカルノイズフリー音声強調手法と比べてケプストラム歪みが小さく、音声強調による音声歪み量が少ないことがわかる。

5 まとめ

本稿では、MMSE STSA 推定器にバイアス付き事前 SNR 推定を導入することで、ミュージカルノイズフリー状態が存在することを実験的に示した。また、実験的に得られたミュージカルノイズフリー状態を達成するパラメータを用いて、従来手法との比較実験を行った。客観評価実験より、ミュージカルノイズフリーを適用した MMSE STSA 推定器は、従来のミュージカルノイズフリー音声強調手法と比べて等 NRR 条件下で音声歪み量が少なく、提案法の優位性を示した。

謝辞 本研究の一部は科研費による研究助成・戦略的基礎研究推進事業 (CREST) の支援を受けた。

参考文献

- [1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.27, no.2, pp.113–120, 1979.
- [2] P. C. Loizou, *Speech Enhancement Theory and Practice*, CRC Press, Taylor & Francis Group, FL, 2007.
- [3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.32, no.6, pp.1109–1121, 1984.
- [4] M. R. Khan, T. Hansan, "Iterative noise power subtraction technique for improved

speech quality," *Proceedings of International Conference on Electrical and Computer Electrical Engineering (ICECE2008)*, pp.391–394, 2008.

- [5] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.26, no.3, pp.197–210, 1978.
- [6] J. H. Hansen and M. A. Clements, "Constrained iterative speech enhancement with application to speech recognition," *IEEE Transactions on Signal Processing*, vol.39, no.4, pp.795–805, 1991.
- [7] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, K. Kondo, "Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics," *Proceedings of International Workshop on Acoustic Signal Enhancement (IWAENC2008)*, 2008.
- [8] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano, K. Kondo, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.20, no.7, pp.2080–2094, 2012.
- [9] R. Miyazaki, H. Saruwatari, K. Shikano, K. Kondo, "Musical-noise-free speech enhancement based on iterative Wiener filtering," *Proceedings of IEEE International Symposium on Signal Processing and Information Technology (ISSPIT2012)*, 2012.
- [10] S. Kanehara, H. Saruwatari, R. Miyazaki, K. Shikano, K. Kondo, "Theoretical analysis of musical noise generation in noise reduction methods with decision-derived a priori SNR estimator," *Proceedings of International Workshop on Acoustic Signal Enhancement (IWAENC2012)*, 2012.