倍音復元技術に基づくバイアス付き事前SNR推定を導入したミュージカルノイズフリーMMSE-STSA法の音質改善に関する研究

宇根 昌和[1,a)]　宮崎 亮一[1,b)]

概要：近年，音声による情報伝達が多く利用されている．しかし，周囲の雑音によって音声の品質が劣化してしまうため，目的音声を高精度で抽出する雑音抑圧技術が必要となる．一方で，過剰な雑音抑圧は目的音声の歪みを生じさせ，かえって聞こえづらい音声となる．また，非線形雑音抑圧によって「ミュージカルノイズ」という非常に耳障りな歪みが生じる問題もある．音声成分の歪みを改善する方法として Harmonic Regeneration Noise Reduction (HRNR) が提案されている．HRNR は Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE-STSA) 法など，一般的な雑音抑圧手法に適用できるとされている．一方で，Nakai らはミュージカルノイズを発生させずに雑音抑圧を行う，MMSE-STSA 法に基づくミュージカルノイズフリー雑音抑圧手法を提案している．この手法は，MMSE-STSA 推定法にバイアス付き事前 SNR 推定を導入し，バイアス値を調整することでミュージカルノイズを発生させずに雑音抑圧を行う方法である．本研究では，バイアス付き事前 SNR 推定を導入した MMSE-STSA 法に HRNR を応用することによって，ミュージカルノイズを発生させず，かつ音声歪み量の少ない雑音抑圧手法を提案する．

Improvement of Sound Quality of Musical-Noise-Free Noise Reduction Technique for Biased MMSE Short-Time Spectral Amplitude Estimator Based on Harmonic Regeneration

Une Masakazu[1,a)]　Miyazaki Ryoichi[1,b)]

1. はじめに

音声は人間にとって最も自然で利用しやすいコミュニケーション手段の一つである．近年では，音声対話ロボットやテレビ会議システム，補聴器など，音声通信に関するシステムが増加しており，音声による情報伝達が多く利用されている．システムを用いる際には，我々の話す音声だけでなく，他人の声や車の音などの雑音も同時に入力される．周囲の雑音によって目的音声の品質が劣化するため，高精度で目的音声を抽出する雑音抑圧が必要となる．

雑音抑圧手法は主に 2 つに分けられる．1 つは信号の変形が線形な関係で表される線形処理，もう 1 つは信号の変形が非線形な関係で表される非線形処理である．線形処理に基づく雑音抑圧技術の代表例として，ビームフォーミングに基づく手法 [1] やブラインド音源分離に基づく手法 [2] がある．線形処理に基づく雑音抑圧は出力する音質が良い反面，複数のマイクロホンを必要とするためにシステムの規模やコストが大きくなる問題がある．

一方，非線形処理に基づく雑音抑圧技術は雑音抑圧性能が高く，アルゴリズムの汎用性に優れており，演算量も少ないことから盛んに研究されている技術である．中でも，Wiener Filtering (WF) [3]，Spectral Subtraction (SS) [4]，Minimum Mean-Square Error Short-Time Spectral Am-

[1] 徳山工業高等専門学校 情報電子工学専攻
Gakuendai, Shunan, Yamaguchi 745–8585, Japan
[a)] i12une@tokuyama.ac.jp
[b)] miyazaki@tokuyama.ac.jp

plitude (MMSE-STSA) 法 [5] は，古典的な雑音抑圧手法として広く研究されている．

しかし，非線形処理は過剰な雑音抑圧と雑音の推定精度の問題から，目的の音声の歪みが発生し，かえって聞こえづらい音声となることがある．また，出力信号中に「ミュージカルノイズ」と呼ばれる特有の歪みが生じる問題がある．ミュージカルノイズとは，非線形処理を適用した信号に発生する人工的な音色の歪みであり，この特有の耳障りな音色が音声の品質を著しく劣化させる．

音声成分の歪みを改善する方法として倍音復元に基づく雑音抑圧 (Harmonic Regeneration Noise Reduction: HRNR) が提案されている [6]．発声された音声には倍音成分が多く含まれており，雑音抑圧を行うと音声中の多くの倍音成分が失われる．HRNR は雑音抑圧によって失われた倍音成分を復元し，より品質の良い音声を得る技術である．HRNR を提案した論文 [6] では，一般的な雑音抑圧手法に対して HRNR を適用できるとされているが，WF についての評価実験のみ述べられていた．そこで我々は，SS と MMSE-STSA 法に対して HRNR を適用した信号について評価を行い，音質が改善されることを明らかにした [7]．

一方で，非線形雑音抑圧においてミュージカルノイズを全く発生させずに雑音抑圧を行う「ミュージカルフリー雑音抑圧」が提案されている [8–10]．前述の通り，ミュージカルノイズは人にとって耳障りな音色であるため，ミュージカルノイズフリー雑音抑圧は，人が聞くシステムにおいて有用な技術である．中でも，Nakai らは MMSE-STSA 法に基づくミュージカルノイズフリー雑音抑圧手法を提案している [9]．この手法は MMSE-STSA 法にバイアス付き事前 SNR 推定を導入することにより，ミュージカルノイズを発生させず，かつ，音声の歪み量の少ない信号を得られる方法である．

本研究ではバイアス付き事前 SNR 推定を導入した MMSE-STSA 法に対し，HRNR を応用することでミュージカルノイズを全く発生させず，より音声の歪み量が少ない雑音抑圧手法を提案する．

2. 関連研究

2.1 信号の定義

雑音を含む観測信号 $x(t)$ は，元の音声信号 $s(t)$ と雑音信号 $n(t)$ から成り次の式で表される．

$$x(t) = s(t) + n(t) \tag{1}$$

式 (1) を短時間フーリエ変換することで，次の式に表される複素スペクトルを得る．

$$X(p,k) = S(p,k) + N(p,k) \tag{2}$$

ここで，p は短時間フレームのインデックス，k はフレーム内の周波数インデックスを表す．

雑音抑圧とは，観測信号のスペクトル $X(p,k)$ にスペクトルゲイン $G(p,k)$ を掛け合わせ，次の式に示す音声信号のスペクトルの推定値 $\hat{S}(p,k)$ を求めることである．

$$\hat{S}(p,k) = G(p,k)X(p,k) \tag{3}$$

ここで，スペクトルゲインは雑音抑圧手法によって異なるが，一般的には事前 SNR $\xi(p,k)$ と事後 SNR $\gamma(p,k)$ の関数であり，次のように表せる．

$$G(p,k) = g(\xi(p,k), \gamma(p,k)) \tag{4}$$

関数 g には様々な雑音抑圧手法のスペクトルゲイン関数を用いる（WF，MMSE-STSA 法など）[3,5]．また，$\xi(p,k)$ と $\gamma(p,k)$ は次の式で定義される．

$$\xi(p,k) = \frac{\mathrm{E}[|S(p,k)|^2]}{\mathrm{E}[|\hat{N}(p,k)|^2]} \tag{5}$$

$$\gamma(p,k) = \frac{|X(p,k)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]} \tag{6}$$

ここで，$\hat{N}(p,k)$ は推定した雑音スペクトル，$\mathrm{E}[\cdot]$ は期待値演算子を表す．

2.2 MMSE-STSA 法

MMSE-STSA 法は，元の音声信号と推定音声信号の振幅スペクトルの平均二乗誤差を最小にする手法である [5]．MMSE-STSA 法のスペクトルゲイン $G_{\mathrm{STSA}}(p,k)$ は次の式で表される．

$$G_{\mathrm{STSA}}(p,k) = \frac{\sqrt{\nu(p,k)}}{\gamma(p,k)}\Gamma\left(\frac{3}{2}\right) M\left(-\frac{1}{2}; 1; \nu(p,k)\right) \tag{7}$$

ここで，$\Gamma(h)$，$M(a;b;z)$ はそれぞれガンマ関数，第一種合流超幾何関数を表し，$\nu(p,k)$ は次の式で表される．

$$\nu(p,k) = \frac{\xi(p,k)}{1+\xi(p,k)}\gamma(p,k) \tag{8}$$

事前 SNR $\xi(p,k)$ を求める際に音声信号の情報が必要となるが，実環境で音声信号を事前に知ることはできない．そこで，次式で表される decision-directed 法を利用し，事前 SNR $\xi(p,k)$ を推定する [5]．

$$\begin{aligned}\hat{\xi}(p,k) =& \alpha\frac{|G(p-1,k-1)X(p-1)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]} \\ &+ (1-\alpha)\mathrm{Max}[\gamma(p,k)-1, 0]\end{aligned} \tag{9}$$

ここで，α は忘却係数と呼ばれ，前フレームの情報をどの程度事前 SNR の推定に利用するかを決めるパラメータである．一般的に，$\alpha = 0.98$ と設定すると音質の面で最も良いとされる [3,5]．$\mathrm{Max}[a,b]$ は，a と b のうち大きい値を選択する関数である．

2.3 HRNR

2.3.1 HRNR の事前 SNR

一般的に，過剰に雑音抑圧を行うと音声中の倍音成分が歪み，音声の品質を劣化させる．HRNR は失われた倍音成分を復元するために提案された手法である [6]．HRNR の処理の流れを図 1 に示す．雑音抑圧によって得られた信号 $\hat{S}(p,k)$ を用いて，HRNR におけるスペクトルゲイン $G_{\mathrm{HRNR}}(p,k)$ を計算する．雑音抑圧と同様に次の式のように，観測信号に HRNR におけるスペクトルゲイン $G_{\mathrm{HRNR}}(p,k)$ を掛け合わせ，音声信号のスペクトルの推定値 $\hat{S}_{\mathrm{HRNR}}(p,k)$ を求める．

$$\hat{S}_{\mathrm{HRNR}}(p,k) = G_{\mathrm{HRNR}}(p,k)X(p,k) \tag{10}$$

HRNR におけるスペクトルゲイン $G_{\mathrm{HRNR}}(p,k)$ は以下に示すように，HRNR の事前 SNR $\hat{\xi}_{\mathrm{HRNR}}(p,k)$ と事後 SNR $\gamma(p,k)$ の関数で表される．

$$G_{\mathrm{HRNR}}(p,k) = h(\hat{\xi}_{\mathrm{HRNR}}(p,k), \gamma(p,k)) \tag{11}$$

ここで，関数 h には式 (4) の g と同様に雑音抑圧ゲインを選択できる．また，HRNR の事前 SNR $\hat{\xi}_{\mathrm{HRNR}}(p,k)$ は，次の式で表される．

$$\begin{aligned}\hat{\xi}_{\mathrm{HRNR}}(p,k) =& \rho(p,k)\frac{|\hat{S}(p,k)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]} \\ &+ (1-\rho(p,k))\frac{|S_{\mathrm{harmo}}(p,k)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]}\end{aligned} \tag{12}$$

ここで，$\rho(p,k)$ は $\hat{\xi}_{\mathrm{HRNR}}(p,k)$ の推定に雑音抑圧信号のスペクトル $\hat{S}(p,k)$ の情報をどの程度利用するかを決めるパラメータであり，式 (4) に示す各雑音抑圧手法のスペクトルゲインに設定すると良いとされる [6]．また，$S_{\mathrm{harmo}}(p,k)$ は復元信号のスペクトルと呼ばれ，次のように定義される．

$$S_{\mathrm{harmo}}(p,k) = \mathrm{FT}\left[NL\left(\mathrm{IFT}\left[\hat{S}(p,k)\right]\right)\right] \tag{13}$$

ここで，$NL(\cdot)$ は非線形関数 (絶対値，半波整流関数など)，$\mathrm{FT}[\cdot]$，$\mathrm{IFT}[\cdot]$ はそれぞれフーリエ変換，逆フーリエ変換を表す．

2.3.2 HRNR の雑音抑圧量とミュージカルノイズ発生量の関係

HRNR においてミュージカルノイズが発生しない状態 (以下，ミュージカルノイズフリー状態) が存在するか実験により確認する．そこで，HRNR の内部パラメータを変化させ，雑音抑圧量とミュージカルノイズ発生量の関係を調査した．

実験にあたり，雑音抑圧量とミュージカルノイズ発生量の評価のため，それぞれ，Noise Reduction Rate (NRR) と Kurtosis Ratio (KR) を用いた [8]．また，KR が 1 以下のとき，ミュージカルノイズが発生していないことを示す．実験条件として，駅雑音を入力 SNR 0 dB で混合した信号を用いた．また，式 (11) の関数 h を式 (7) に示す MMSE-STSA 法のスペクトルゲインとした．以上の実験条件の下，式 (12) の ρ を 0 から 1.0 まで 0.05 刻みで変化させた．

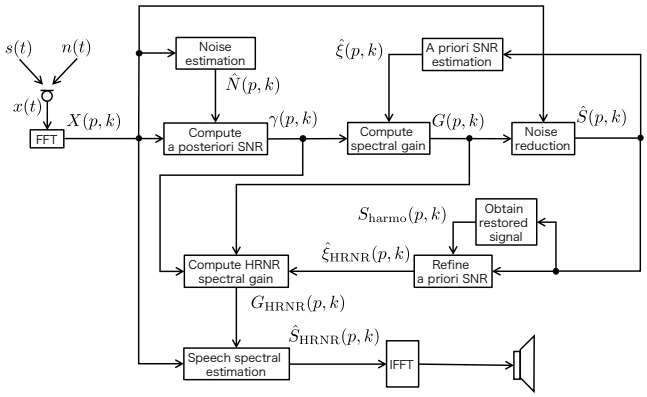

図 1 HRNR のブロック図
Fig. 1 Block diagram for HRNR.

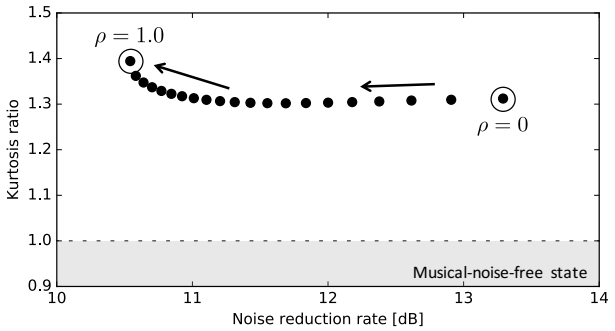

図 2 HRNR における雑音抑圧量とミュージカルノイズ発生量の関係
Fig. 2 Relation between NRR and KR for HRNR with increasing parameter ρ.

HRNR における雑音抑圧量とミュージカルノイズ発生量の関係を図 2 に示す．図 2 から，ρ を増加させると，NRR が減少することがわかる．また，KR が 1 以下となる点が存在しないことから，従来の HRNR ではミュージカルノイズフリー状態を達成しないことが確認できる．

2.4 MMSE-STSA に基づくミュージカルノイズフリー雑音抑圧

これまで, Kanehara らによって MMSE-STSA 法における雑音抑圧量とミュージカルノイズ発生量の関係が理論解析により明らかにされており，MMSE-STSA 法において式 (9) の忘却係数 α をどのような値にしてもミュージカルノイズフリー状態が存在しないことが示されている [11]．そこで，Nakai らは一般的な MMSE-STSA 法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明らかにした [9]．バイアス付き事前 SNR $\hat{\xi}_{\mathrm{bias}}$ は式 (9) の最尤推定の項にバイアス

値を設定し，次のように推定される．

$$\hat{\xi}_{\mathrm{bias}} = \alpha \frac{|G(p-1,k-1)X(p-1)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]} + (1-\alpha)\mathrm{Max}[\gamma(p,k)-1,\varepsilon] \quad (14)$$

ここで，ε はバイアス値である．

3. 提案手法

3.1 バイアス値を導入した HRNR の事前 SNR

我々はミュージカルノイズを全く発生させず，音声歪み量の少ない雑音抑圧手法を目的として，HRNR に基づくミュージカルノイズフリー雑音抑圧手法を提案する．一般的に，MMSE-STSA 法にバイアス値を設けることで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 [12]．そこで，我々は HRNR の事前 SNR $\hat{\xi}_{\mathrm{HRNR}}(p,k)$ に対しバイアス値を設け，式 (12) を次のように変更する．

$$\hat{\xi}_{\mathrm{prop}}(p,k) = \rho_{\mathrm{const}} \mathrm{Max}\left[\frac{|\hat{S}(p,k)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]}, \varepsilon'\right] + (1-\rho_{\mathrm{const}})\frac{|S_{\mathrm{harmo}}(p,k)|^2}{\mathrm{E}[|\hat{N}(p,k)|^2]} \quad (15)$$

ここで，$\hat{\xi}_{\mathrm{prop}}(p,k)$ は提案手法における事前 SNR，ε' はバイアス値，ρ_{const} は $0 < \rho_{\mathrm{const}} < 1$ の定数である．$\hat{\xi}_{\mathrm{prop}}(p,k)$ を用いて，式 (11) と同様にスペクトルゲイン $G_{\mathrm{prop}}(p,k)$ を得る．

$$G_{\mathrm{prop}}(p,k) = h'(\hat{\xi}_{\mathrm{prop}}(p,k), \gamma(p,k)) \quad (16)$$

ここで，関数 h' には式 (4) の g と同様に雑音抑圧ゲインを選択できる．

3.2 提案手法の雑音抑圧量とミュージカルノイズ発生量の関係

2.3.2 節と同様に，式 (15) の内部パラメータを変更し，提案手法の音質を調査する．式 (15) には ρ_{const} と ε' の 2 つのパラメータがあるが，本研究では ρ_{const} を固定し，ε' を変化させることとした．ρ_{const} を 0.1，0.5，0.9 と固定した 3 つの値に設定し，ε' を 0 から 3.0 まで 0.05 刻みで変化させた．

提案手法における雑音抑圧量とミュージカルノイズ発生量の関係を図 3 に示す．図 3 から，ε' を増加させると NRR が減少することがわかる．さらに，$\varepsilon' = 0$ から増加し始める部分では急激に KR が減少するが，ε' が 3.0 に近づくにつれて KR はあまり変化しなくなる．また，$\rho_{\mathrm{const}} = 0.1$ の場合は，他の 2 つの場合と比べてミュージカルノイズフリー状態での NRR が大きい．このことから，雑音抑圧量とミュージカルノイズ発生量の点で，ρ_{const} は小さい値に設定することが音質の面で有効であると言える．

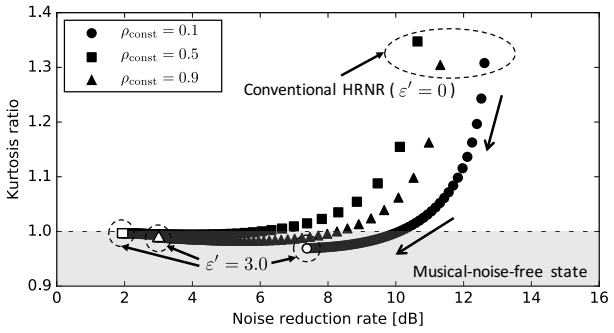

図 3 提案手法における雑音抑圧量とミュージカルノイズ発生量の関係

Fig. 3 Relation between NRR and KR for proposed method with increasing parameter ε'.

4. 評価実験

4.1 実験条件

本研究では提案手法の有効性を示すため，従来手法との比較を行う．比較対象として，MMSE-STSA 法，HRNR，ミュージカルノイズフリー MMSE-STSA 法を用いた．また，評価尺度として KR，ケプストラム歪み (Cecstral Distortion: CD) [13] の 2 つの客観評価尺度を用いた．目的音声には JNAS の音声コーパス [14] より 10 文（男性 5 発話，女性 5 発話，計 10 発話）を用い，雑音は駅雑音，道雑音，白色ガウス雑音の 3 種類とし，それらをそれぞれ 0 dB，5 dB の入力 SNR で混合したものを観測信号とした．ここで，HRNR と提案手法に関して式 (11) の関数 h と式 (16) の関数 h' は式 (7) の MMSE-STSA 法のスペクトルゲインとした．雑音抑圧を行った音声の NRR が 10 dB となるように，MMSE-STSA 法においては式 (9) の α を，HRNR においては式 (12) の ρ を，ミュージカルノイズフリー MMSE-STSA 法においては式 (14) の ε を，提案手法においては式 (15) の ρ_{const} を 0.1 に固定した上で，ε' を調整した．

4.2 実験結果

客観評価実験の結果を図 4，図 5 に示す．図 4，図 5 はそれぞれ KR と CD の値を表している．また，各図において上のグラフ，下のグラフは入力 SNR を 0 dB，5 dB でそれぞれ混合した場合の結果である．

まず，図 4 より，道雑音や白色ガウス雑音においては，全ての条件で提案手法の KR が 1 以下であることから，ミュージカルノイズが発生していないことがわかる．一方，駅雑音では提案手法における KR が 1 以上である．しかし，提案手法は他の従来手法と比較して KR が最も小さく，ミュージカルノイズの発生量が少ないことが確認できる．以上をまとめると，提案手法はミュージカルノイズを発生させない，もしくは発生しても他の手法と比べて発生

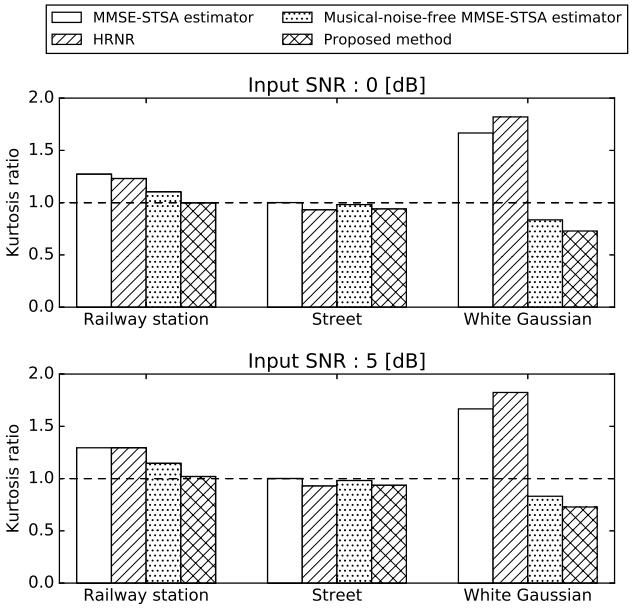

図 4 入力 SNR 0 dB，5 dB でそれぞれ混合した信号の KR
Fig. 4 KR at 0-dB and 5-dB input SNRs.

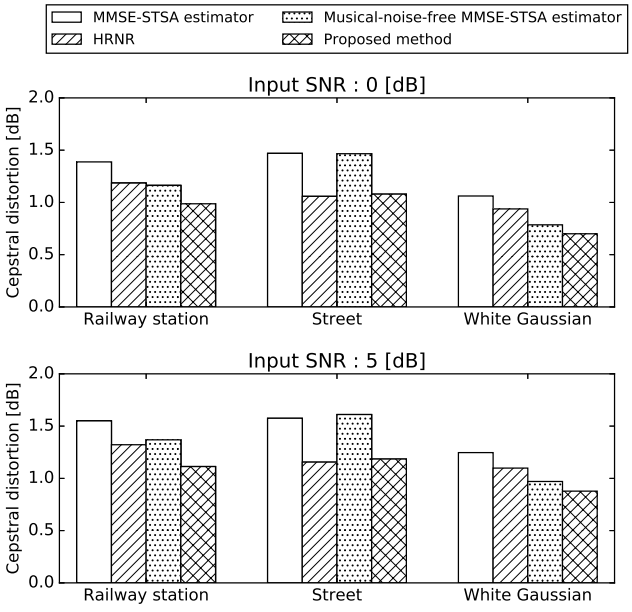

図 5 入力 SNR 0 dB，5 dB でそれぞれ混合した信号の CD
Fig. 5 CD at 0-dB and 5-dB input SNRs.

量は少ない．よって，提案手法はミュージカルノイズ発生量の点で優れていると言える．

次に，図 5 より，提案手法の CD が他の手法に比べて小さく，特に，駅雑音や白色ガウス雑音においては最も CD が小さい．このことから，提案手法は音声の歪み量の点でも優れていると言える．

以上より，提案手法は等 NRR 条件の下で従来手法より KR と CD が低く，総合的に音質が良いと言える．

5. まとめ

本稿では，HRNR に対しミュージカルノイズフリー MMSE-STSA 法を応用した手法を提案した．次に, MMSE-STSA 法, HRNR, ミュージカルノイズフリー MMSE-STSA 法との比較実験を行った．比較実験から，提案手法はミュージカルノイズ発生量，音声歪み量の点で優れていることを明らかにした．今後の課題として，提案手法におけるミュージカルノイズフリー状態の存在について理論的に明らかにする．

参考文献

[1] R. Z. J. L. Flanagan, J. D. Johnston and G. W. Elko, “Computer-streered microphon arrays for sound transduction in large rooms,” *Journal of the Acoustical Society of America*, vol.78, no.5, pp.1508–1518, 1985.

[2] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, and T. Nishikawa, “Blind source separation combining independent component analysis and beamforming,” *EURASIP Journal on Applied Signal Processing*, vol.2003, pp.1135–1146, 2003.

[3] N. Wiener, “Extrapolation, interpolation and smoothing of stationary time series with engineering applications,” *Cambridge, MA: MIT Press*, 1949.

[4] 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.

[5] 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.27, no.6, pp.1109–1121, 1984.

[6] C. Plapous, C. Marro and P. Scalart, “Improved signal-to-noise ratio estimation for speech enhancement,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.14, no.6, pp.2098–2108, 2006.

[7] M. Une, R. Miyazaki, “Evaluation of sound quality and speech recognition performance using harmonic regeneration for various noise reduction techniques,” *2017 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing*, pp.377–380, 2017.

[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] S. Nakai, H. Saruwatari, R. Miyazaki, S. Nakamura, K.Kondo, “Theoretical analysis of biased MMSE short-time spectral amplitude estimator and its extension to musical-noise free speech enhancement,” *Hands-free Speech Communication and Microphone Arrays*, pp.122–126, 2014.

[10] H. Saruwatari, “Statistical-model-based speech enhancement with musical-noise-free properties,” *Digital Signal Processing*, pp.1201–1205, 2015.

[11] S. Kanehara, H. Saruwatari, R. Miyazaki, K. Shikano, K. Kondo, “Theoretical analysis of musical noise generation in noise reduction methods with decision directed a priori SNR estimator,” *Proceedings of International*

Workshop on Acoustic Echo and Noise Control, 2012.

[12] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, vol.2, no.2, pp.345–349, 1994.

[13] L. Rabiner and B. Juang, "Fundamentals of Speech Recognition," *Upper Saddle River, NJ: Prentice-Hall*, 1993.

[14] K. Ito, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, S. Itahashi, "Jnas: Japanese speech corpus for large vocabulary continuous speech recognition research," *The Journal of Acoustical Society of Japan*, vol.20, pp.196–206, 1999.