一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
EA2018-82 (2018-12)

バイアス付き倍音復元技術における内部パラメータと音質の関係性の調査

宇根　昌和†　　宮崎　亮一†

† 徳山工業高等専門学校情報電子工学専攻　〒745–8585 山口県周南市学園台
E-mail: †{i12une,miyazaki}@tokuyama.ac.jp

あらまし　これまでに音声の歪みを改善する手法として Harmonic Regeneration Noise Reduction (HRNR) と呼ばれる手法が提案されている．また，我々は音声の歪みとミュージカルノイズの両問題に対し，HRNR にバイアスを導入（バイアス付き HRNR）することを提案している．本稿では，バイアス付き HRNR について，各パラメータと音質の関係について実験的に調査する．さらに，事前 SNR の推定精度と音質への影響について考察する．
キーワード　倍音復元，音声の歪み，ミュージカルノイズ，事前 SNR 推定

Relationship between internal parameters and sound quality in biased harmonic regeneration technique

Masakazu UNE† and Ryichi MIYAZAKI†

† Department of computer Science and Electronic Engineering, National Institute of Technology Tokuyama College　Gakuendai, Shunan-shi, Yamaguchi, 745–8585 Japan
E-mail: †{i12une,miyazaki}@tokuyama.ac.jp

Abstract　Harmonic Regeneration Noise Reduction (HRNR) has been proposed to improve the speech distortion. We introduced bias into HRNR and proposed new noise reduction technique (biased HRNR) which approach two main problems: speech distortion and musical noise. In this paper, we investigate the relationship between the internal parameters and the sound quality in biased HRNR. Additionally, we consider the effect on the sound quality and the estimation accuracy of a priori SNR.
Key words　Harmonic regeneration, speech distortion, musical noise, a priori SNR estimation

1. ま え が き

近年のスマートフォンには通話機能だけでなく，ボイスメモ機能を含む録音システムや音声認識システムが搭載されている．これらのシステムを快適に利用するために，周囲の雑音による音質の劣化は深刻な問題である．そこで，これまでに様々な雑音抑圧手法が提案されてきた [1]〜[8]．ビームフォーミングに基づく雑音抑圧 [1] や音源分離に基づく手法 [2] に代表されるマルチチャネル雑音抑圧は複数のマイクを必要とし，小型のデバイスに搭載するには規模やコストの面で問題がある．また，複数のマイクの入力を同時に扱うために逆行列の計算が必要となり，演算量や安定性の問題もある．一方，シングルチャネル雑音抑圧 [3]〜[8] は 1 つのマイクのみで処理するため規模やコストも小さく，演算量も少ない．このような背景から，小型のデバイスを用いて快適な通話やクリアな録音を行うためには，高品質なシングルチャネル雑音抑圧が必要不可欠である．

しかし，シングルチャネル雑音抑圧による出力音声には目的音声の歪みとミュージカルノイズの発生の 2 つの問題が発生する [9],[10]．音声の歪みの問題は雑音とともに目的の音声成分も抑圧してしまうことによって発生し，聞き取りづらい音声になる．また，この問題により音声認識システムは著しく認識精度を落とす．音声の歪みの問題に対して，Harmonic Regeneration Noise Reduction (HRNR) という手法が提案されている [11],[12]．HRNR は音声の歪みの多くが倍音成分の歪みであることに着目し，倍音復元信号と呼ばれる信号により事前 Signal to Noise Ratio (SNR) を推定することで，音声の歪みを低減する手法である．

一方，ミュージカルノイズの発生の問題は非線形処理特有の歪みで，ミュージカルノイズが発生した音声は聴覚的に非常に不快である．この問題は通話や録音などの人が聞くシステムの利用を考えた場合，大きな弊害となる．ミュージカルノイズの発生の問題に対して，ミュージカルノイズフリー雑音抑圧と呼ばれ

る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手法が提案されている [13]〜[15]．中でも，Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE-STSA) 法に基づくミュージカルノイズフリー雑音抑圧 (ミュージカルノイズフリー MMSE-STSA 法) [14] は他のミュージカルノイズフリー雑音抑圧手法に比べ，音声歪みの少ない手法として知られている．[14] において Nakai らは Decision Directed (DD) と呼ばれる事前 SNR の推定法 [5] にバイアスを導入することで，ミュージカルノイズを発生させず，かつ音声の歪みを抑える手法を提案した．これまでに，ミュージカルノイズフリー MMSE-STSA 法に基づき，HRNR の事前 SNR 推定式にバイアスを導入し，ミュージカルノイズを発生させず，より音声の歪みを抑えた新たな手法 (バイアス付き HRNR) を提案している [16].

上に述べたように，HRNR では事前 SNR の推定に倍音復元信号を用い，また，バイアス付き HRNR では事前 SNR の推定式にバイアスを導入する処理を行っている．これらを含む多くの雑音抑圧手法は事前 SNR の推定に様々な改良が加えられており，事前 SNR の推定が出力音声の品質に大きく影響する [17]．一方，HRNR やバイアス付き HRNR には多くの内部パラメータが存在し [11], [16]，この内部パラメータは事前 SNR の推定精度，ひいては音質を決定づける．[16] では，内部パラメータと音質の関係性について非常に限定的な条件でのみ述べており，様々な条件における内部パラメータと音質の関係については言及していない．内部パラメータと音質の関係を明らかにすることは，最適なパラメータを決定する緒となり，実用的なシステムへの応用も考えられる [18], [19].

そこで本稿では，HRNR の内部パラメータとミュージカルノイズの発生量を含む音質との関係を実験的に調査する．また，バイアス付き HRNR における内部パラメータと音質の関係についても同様に調査する．最後に，内部パラメータによって決定される事前 SNR の推定値と音質の関係について考察する．

2. 雑音抑圧手法と事前 SNR の推定

本章では，Ephraim らが提案した古典的な事前 SNR 推定法とそれを利用した MMSE-STSA 法について述べる．また，音声の歪みやミュージカルノイズの発生の問題を解決するために提案された事前 SNR 推定法について述べる．

2.1 古典的な事前 SNR 推定法

目的の音声 $s(t)$ には雑音 $n(t)$ が加算され，観測信号 $x(t)$ は次のように表される．

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

式 (1) に Short-time fourier transform (STFT) を適用し，時間フレーム p $(0 \leq p \leq P)$，周波数 k $(0 \leq k \leq K)$ における観測信号のスペクトル $X(p,k)$ は目的信号のスペクトル $S(p,k)$ と雑音信号のスペクトル $N(p,k)$ を用いて次のように表せる．

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

以後，特に明示しない限り (p,k) を省略する．得られる信号は X のみであるため，次式のように X に適当なスペクトルゲイン G を乗算し，目的音声のスペクトルの推定値 $\hat{S}$ を得ることを考える．

$$\hat{S} = GX \tag{3}$$

Wiener Filtering や MMSE-STSA 法を始めとする多くの雑音抑圧手法のスペクトルゲインは，事前 SNR ξ と事後 SNR γ の関数として，次式のように表せる．

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

ここで，$g(\cdot,\cdot)$ はゲイン関数である．以降，本稿ではこのスペクトルゲイン g には後述する MMSE-STSA 法のゲイン関数の式 (9) を用いる．また，事前 SNR ξ と事後 SNR γ は次式のように定義される．

$$\xi = \frac{\mathrm{E}\left[|S|^2\right]}{\mathrm{E}\left[|N|^2\right]} \tag{5}$$

$$\gamma = \frac{|X|^2}{\mathrm{E}\left[|N|^2\right]} \tag{6}$$

ここで，$\mathrm{E}[\cdot]$ は期待値演算子を表す．本研究では $\mathrm{E}\left[|N|^2\right]$ をフレーム T までの非音声区間の期待値 $\mathrm{E}\left[\left|\hat{N}\right|^2\right]$ として次式のように近似する．

$$\begin{aligned}\mathrm{E}\left[|N(p,k)|^2\right] &\approx \mathrm{E}\left[\left|\hat{N}\right|^2\right] \\ &= \frac{1}{T}\sum_{\tau=0}^{T}|X(\tau,k)|^2\end{aligned} \tag{7}$$

また，ξ は得ることができないため，一般的に以下に示す DD を用いて推定される [5].

$$\hat{\xi}^{\mathrm{DD}}(p,k) = \alpha\frac{\left|\hat{S}(p-1,k)\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} + (1-\alpha)\,\mathrm{Max}\left[\gamma(p,k)-1, 0\right] \tag{8}$$

ここで，α は忘却係数であり，一般的に 0.98 に設定すると最も音質が良いことが実験的に明らかにされている [5]．また，$\mathrm{Max}\,[a,b]$ は a と b のうち大きい値を出力する関数である．

2.2 MMSE-STSA 法

MMSE-STSA 法は振幅領域での真の音声信号と推定した音声信号の誤差を最小化する手法である [5]．MMSE-STSA 法のスペクトルゲインは事前 SNR と事後 SNR の関数 $g(\xi,\gamma)$ として次式のように表される．

$$g(\xi,\gamma) = \frac{\sqrt{\nu}}{\gamma}\Gamma\left(\frac{3}{2}\right)M\left(-\frac{1}{2};1;-\nu\right) \tag{9}$$

$$\nu = \frac{\xi}{1+\xi}\gamma \tag{10}$$

ここで，$\Gamma(\cdot)$ と $M(a;b;z)$ は，それぞれガンマ関数，クンマー関数である．式 (8) を用いて事前 SNR を推定し，次式より最終的な目的音声の推定値 $\hat{S}_{\mathrm{STSA}}$ を得る．

$$\hat{S}_{\mathrm{STSA}} = G_{\mathrm{STSA}}X \tag{11}$$

$$= g(\hat{\xi}^{\mathrm{DD}}, \gamma) X \tag{12}$$

2.3 HRNR

人間の言語において，発音される言葉の約 80 % は有声音であり，有声音のパワースペクトルは高域になるほど小さいことが知られている．この高域成分のパワーが小さいことが原因で，雑音抑圧時にそれらの成分は雑音とみなされ，抑圧される．HRNR はこの点に着目し，主に抑圧される高域の成分 (= 倍音成分) を復元することにより，音声歪みの問題を解決する手法である [11]．HRNR のブロック図を図 1 に示す．HRNR は二度の雑音抑圧を行う．一度目は 2.2 節で述べたような一般的な雑音抑圧を行い，一時的に目的音声の推定値 $\hat{S}$ を得る．一時的な推定音声 $\hat{S}$ に対し，次式のように非線形関数を適用することで倍音復元信号 S_{harmo} を得る．本稿では非線形関数として，Max 関数を用いる．

$$S_{\mathrm{harmo}} = \mathcal{F}\left[\mathrm{Max}\left[\mathcal{F}^{-1}\left[\hat{S}\right], 0\right]\right] \tag{13}$$

ここで，$\mathcal{F}[\cdot]$ と $\mathcal{F}^{-1}[\cdot]$ はそれぞれフーリエ変換，逆フーリエ変換である．S_{harmo} は雑音抑圧後のスペクトルの倍音成分を擬似的に復元した信号となり，元のクリーンな音声には存在しない不自然なスペクトルを持つため，直接倍音復元信号を出力音声とするのは適切ではない．しかし，倍音成分に対する有用な情報を持つため，倍音復元信号を用いて新たな事前 SNR の推定に用いる．HRNR における事前 SNR $\hat{\xi}_{\mathrm{HRNR}}$ を次式に示す．

$$\hat{\xi}^{\mathrm{HRNR}} = \frac{\rho\left|\hat{S}\right|^2 + (1-\rho)\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{14}$$

ここで，ρ $(0 \le \rho \le 1)$ は $\left|\hat{S}\right|^2$ と $|S_{\mathrm{harmo}}|^2$ の重みを決めるパラメータである．また，この ρ については前段のスペクトルゲインを適用すると良いとされている [11]．即ち，前段の雑音抑圧手法が MMSE-STSA 法の場合，スペクトルゲインは式 (9) より計算され，式 (14) において $\rho = G$ とする．定数パラメータを適用した場合の事前 SNR と，スペクトルゲインを適用した場合の事前 SNR を明示的に区別するため，両者を次のように表す．

$$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}} = \frac{\rho\left|\hat{S}\right|^2 + (1-\rho)\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{15}$$

$$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{gain}} = \frac{G\left|\hat{S}\right|^2 + (1-G)\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{16}$$

ここで，$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}}$ と $\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{gain}}$ は式 (14) にそれぞれ内部パラメータに定数，またはスペクトルゲインを適用した場合の事前 SNR である．以上より得られた事前 SNR を用いて，スペクトルゲインを求め，次式のように最終的な出力を得る．

$$\begin{aligned}\hat{S}_{\mathrm{HRNR}} &= G_{\mathrm{HRNR}} X \\ &= g(\hat{\xi}^{\mathrm{HRNR}}, \gamma) X\end{aligned} \tag{17}$$

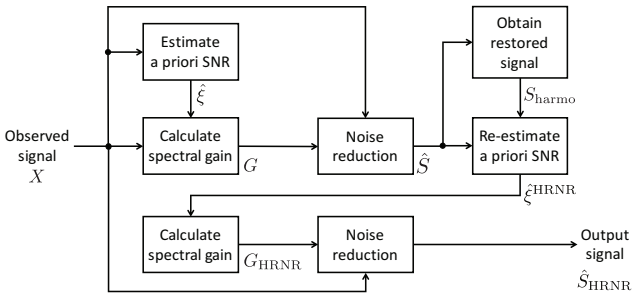

図 1　HRNR のブロック図

3. バイアス付き HRNR

一般的に，DD ではバイアス値を導入することでミュージカルノイズの発生量が少なくなることが知られている [20]．バイアス付き HRNR は従来の HRNR の事前 SNR 推定式にバイアスを導入することで，音声の歪みとミュージカルノイズの発生の両問題にアプローチした手法である [16]．HRNR における事前 SNR の推定法にバイアスを導入し，式 (14) を変更する場合，次の 3 通りが考えられる．一つ目は第一項に対してバイアスを導入した $\hat{\xi}^{\mathrm{1term}}$ (以下，1term)，二つ目は式全体にバイアスを設定した $\hat{\xi}^{\mathrm{whole}}$，三つ目は第一項を DD における最尤推定項に置き換えた $\hat{\xi}^{\mathrm{ML}}$ である．これらを式で表すと以下のようになる．

$$\hat{\xi}^{\mathrm{1term}} = \rho' \,\mathrm{Max}\left[\frac{\left|\hat{S}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]}, \varepsilon'\right] + (1-\rho')\frac{\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{18}$$

$$\hat{\xi}^{\mathrm{whole}} = \mathrm{Max}\left[\frac{\rho'\left|\hat{S}\right|^2 + (1-\rho')\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]}, \varepsilon'\right] \tag{19}$$

$$\hat{\xi}^{\mathrm{ML}} = \rho' \,\mathrm{Max}\left[\gamma - 1, \varepsilon'\right] + (1-\rho')\frac{\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{20}$$

ここで，ρ' $(0 \le \rho' \le 1)$ はバイアス付き HRNR における重みパラメータであり，ε' はバイアス値を表す．予備実験より，これらのバイアスの変化と音質に概ね違いがないことを確認している．そのため，バイアス付き HRNR の事前 SNR 推定式にはこれら 3 つを代表して 1term を採用する．HRNR と同様に，定数パラメータを適用した場合の事前 SNR と，スペクトルゲインを適用した場合の事前 SNR を明示的に区別するため，1term における両者を次のように表す．

$$\hat{\xi}^{\mathrm{1term}}_{\mathrm{const}} = \rho' \,\mathrm{Max}\left[\frac{\left|\hat{S}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]}, \varepsilon'\right] + (1-\rho')\frac{\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{21}$$

$$\hat{\xi}^{\mathrm{1term}}_{\mathrm{gain}} = G \,\mathrm{Max}\left[\frac{\left|\hat{S}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]}, \varepsilon'\right] + (1-G)\frac{\left|S_{\mathrm{harmo}}\right|^2}{\mathrm{E}\left[\left|\hat{N}\right|^2\right]} \tag{22}$$

ここで，$\hat{\xi}^{\mathrm{1term}}_{\mathrm{const}}$ と $\hat{\xi}^{\mathrm{1term}}_{\mathrm{gain}}$ は式 (18) にそれぞれ定数，またはスペクトルゲインを適用した場合の事前 SNR である．

4. 事前 SNR と音質の関係の調査

2. 章では，古典的な事前 SNR 推定法である DD をベースに，HRNR とバイアス付き HRNR における事前 SNR の推定法について述べた．両手法はそれぞれに内部パラメータによって事前 SNR が決定し，ひいては出力音声の音質を決定づける．しかし，HRNR やバイアス付き HRNR は，内部パラメータと音質の詳細な関係について明らかになっていない．本章では，それぞれの手法について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の発生量，また雑音抑圧量と音声の歪み量の関係を実験的に調査する．

4.1 DD と HRNR の内部パラメータと音質の関係

DD は忘却係数 α によって，出力音声の品質を決定づけている [21]．HRNR では前段の事前 SNR 推定の際に DD を用いるため，HRNR の出力音声も忘却係数 α によって音質が変わると考えられる．そこでまず，前段の DD の忘却係数 α と HRNR の出力の関係を明らかにするため，DD の忘却係数 α と HRNR の重みパラメータ ρ を変化させた場合の音質について網羅的に調査する．

本研究で用いる雑音には Babble noise (BB), Railway Station noise (RS) の 2 つを用い，これらを入力 SNR 10 dB それぞれで混合したものを観測信号とした．MMSE-STSA 法においては $\hat{\xi}^{\mathrm{DD}}$ の忘却係数 α を 0〜0.99 まで 0.01 刻みで変化させた．HRNR では，一段目の雑音抑圧手法に MMSE-STSA 法を採用し，$\hat{\xi}^{\mathrm{DD}}$ の忘却係数 α が 0.5, 0.7, 0.98 の 3 パターンにの場合について，$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}}$ の重みパラメータ ρ を 0.0〜1.0 まで 0.02 刻みで変化させた．さらに，重みパラメータ α を前段のスペクトルゲイン G に置き換えた場合，即ち HRNR における事前 SNR の推定に式 (16) を用いた場合についても検証した．以上の条件での出力音声に対して，雑音抑圧量とミュージカルノイズ発生量，また音声の歪み量を算出した．

雑音抑圧量の評価には Noise Reduction Rate (NRR) を用いた [13]．NRR は入力 SNR と出力 SNR の差として次式より求められる．

$$\mathrm{NRR} = 10\log_{10}\frac{\mathrm{E}\left[|s_{\mathrm{out}}|^2\right]/\mathrm{E}\left[|n_{\mathrm{out}}|^2\right]}{\mathrm{E}\left[|s_{\mathrm{in}}|^2\right]/\mathrm{E}\left[|n_{\mathrm{in}}|^2\right]} \tag{23}$$

ここで，s_{out}，n_{out} はそれぞれ出力信号の音声信号と雑音信号，s_{in}，n_{in} はそれぞれ入力信号の音声信号と雑音信号である．NRR が大きいほど，より雑音を抑圧できていることを示す．

次にミュージカルノイズの発生量を測る尺度として Kurtosis Ratio (KR) を用いた [22]．KR は雑音抑圧処理前のカートシス $\mathrm{Kurt}_{\mathrm{org}}$ と処理後のカートシス $\mathrm{Kurt}_{\mathrm{proc}}$ を用いて次のように計算する．

$$\mathrm{KR} = \frac{\mathrm{Kurt}_{\mathrm{proc}}}{\mathrm{Kurt}_{\mathrm{org}}} \tag{24}$$

$\mathrm{KR} \leq 1.0$ のとき，その音声はミュージカルノイズが発生していない状態 (以下，ミュージカルノイズフリー状態) であることを示す．

最後に，音声の歪み量を測る尺度として Cepstral Distortion (CD) を用いた [23]．CD は目的信号のケプストラム係数 C_{ref} と処理後のケプストラム係数 C_{out} を用いて次のように計算する．

$$\mathrm{CD} = \frac{20}{P\log 10}\sum^{P}\sqrt{\sum^{B}2(C_{\mathrm{out}} - C_{\mathrm{ref}})} \tag{25}$$

ここで，B はケプストラムの次元数を表す．その他の条件として，用いた信号のサンプリング周波数は 16 kHz であり，STFT は長さ 512 のハミング窓を 25% オーバーラップで使用した．また，KR の計算のため，観測信号に 3 秒間の非音声区間を設けた．CD の次元数 B は 22 とした．

KR の結果を図 2 (a)，図 2 (b) に示す．図中のサイズの大きいシンボルは各パラメータが 0.0 の点を表す．図 2 (a), 図 2 (b) より，DD について，忘却係数を大きくすると NRR が増加する．一方，$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}}$ の音質について，ρ の変化の軌跡は前段の α の値に依存する．即ち，α が大きい場合，NRR も大きくなり，後段の HRNR での ρ の軌跡も NRR の大きい位置に描かれる．HRNR を適用することで NRR が増加し，概ね KR が減少するが，ρ を増加させると，NRR は減少し，KR も上昇する．$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{gain}}$ の音質については，DD のどのパラメータの場合よりも，KR は小さい値である．しかし，$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}}$ での音質と比較すると，$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{gain}}$ での音質は $\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}}$ に小さい値の定数パラメータを与えた場合，NRR と KR の両点で $\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{const}}$ に劣る．つまり，ミュージカルノイズ発生量の点では，HRNR を適用することは有用であり，値の小さな定数パラメータを用いた方が良いと言える．

次に，CD の結果を図 2 (c)，図 2 (d) に示す．図 2 (c), 図 2 (d) より，DD について，忘却係数を大きくすると CD は上昇する．HRNR のパラメータについては KR の結果と同様に，CD においても ρ の変化の軌跡が前段の NRR の変化に依存する．ρ を大きくしていくと，はじめは緩やかに CD が減少し，ある点を境に上昇する．CD における ρ の軌跡は他の条件にも見られ，CD が最も小さくなる最適なパラメータがあると考えられる．しかし，$\hat{\xi}^{\mathrm{HRNR}}_{\mathrm{gain}}$ ではどの定数パラメータの場合よりも CD が低く，音声の歪みの点では式 (16) を用いた方が良いと言える．

4.2 バイアス付き HRNR の内部パラメータと音質の関係

4.1 節では，HRNR を用いることはミュージカルノイズと音声の歪みの観点で有効であると述べた．そこで，HRNR とバイアス付き HRNR において，それぞれの内部パラメータと音質の関係について調査する．

バイアス付き HRNR には ρ' と ε' の 2 つの内部パラメータが存在する．そこで，ρ' を 0.1, 0.9 の 2 つに固定し，それぞれに対して ε' を 0.0〜1.0 まで 0.1 刻みで変化させた．バイアス付き HRNR における一段目の雑音抑圧手法には，4.1 節と同様に MMSE-STSA 法を採用し，$\hat{\xi}^{\mathrm{DD}}$ の忘却係数 α が 0.5, 0.98 の 2 パターンにの場合について，それぞれ調査した．その他の条件は 4.1 節と同様とした．

KR の結果を図 3 (a)，図 3 (b) に示す．図 3 より，バイア

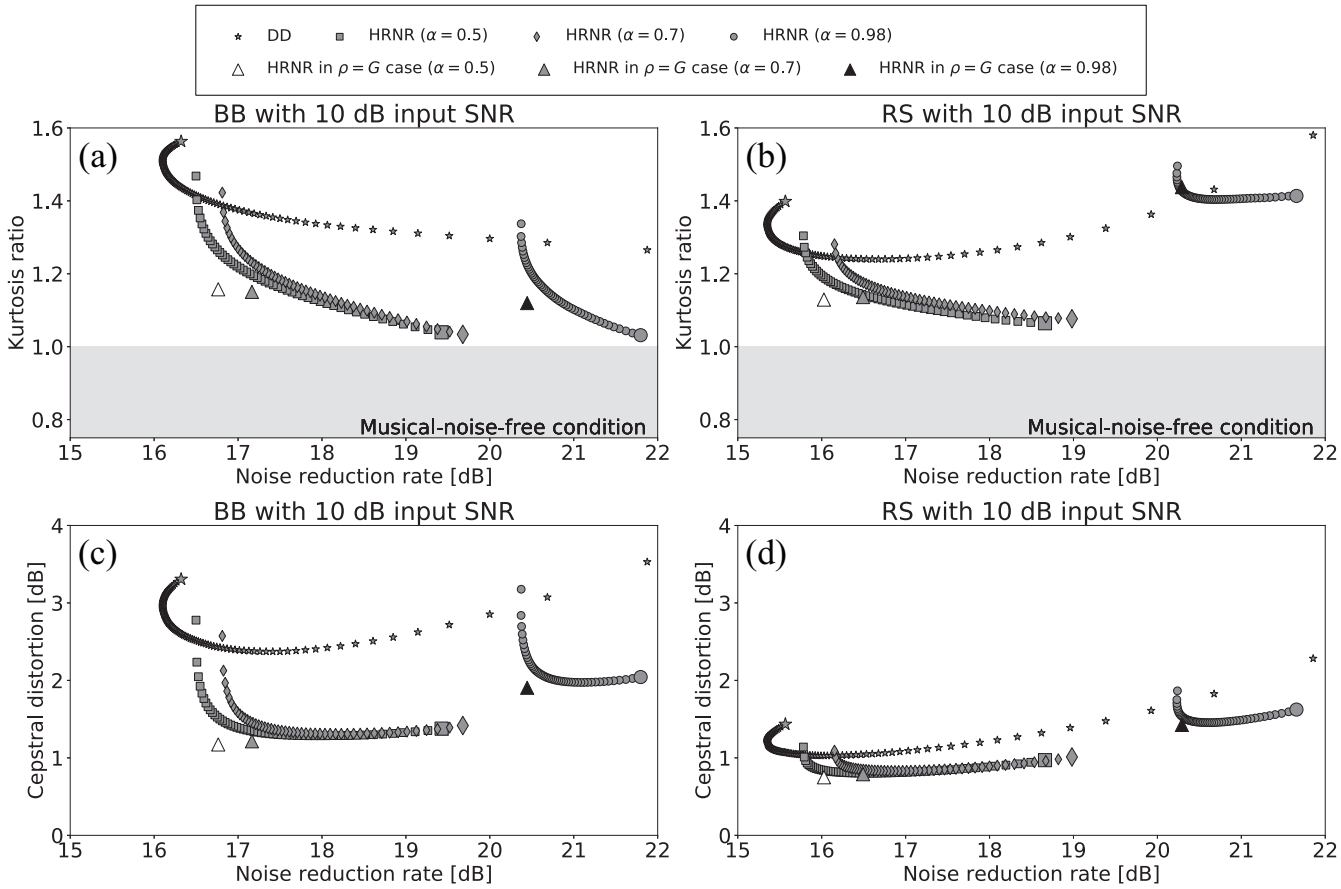

図 2　DD の忘却係数と HRNR の重みパラメータの変化における NRR と KR の関係．(a), (b): バブル雑音，駅雑音をそれぞれ入力 SNR10 dB で混合した場合の KR の結果．(c), (d): バブル雑音，駅雑音をそれぞれ入力 SNR10 dB で混合した場合の CD の結果．

ス値を大きくすることで NRR と KR が共に減少することが確認できる．また，NRR が等しい点で HRNR とバイアス付き HRNR を比較すると，バイアス付き HRNR の KR は HRNR の KR より値が小さく，ミュージカルノイズの発生量を低く抑えられていることが分かる．$\hat{\xi}^{\mathrm{1term}}_{\mathrm{const}}$ での音質について，$\rho' \neq 0.0$ では，ρ' の値に関わらず，バイアスを導入することで KR が 1.0 を下回る．即ち，定数パラメータを用いたバイアス付き HRNR は重みパラメータ ρ' の値に関わらず，バイアスによってミュージカルノイズフリー状態を達成することができる．また，NRR の値が大きい方が良いため，$\hat{\xi}^{\mathrm{1term}}_{\mathrm{const}}$ において定数パラメータは小さい値に設定すると良い．一方，$\hat{\xi}^{\mathrm{1term}}_{\mathrm{gain}}$ のバイアスの軌跡は $\hat{\xi}^{\mathrm{1term}}_{\mathrm{const}}$ の場合と同様である．しかし，前段の雑音抑圧量が小さい場合 (即ち，α が小さい場合)，バイアスを増加させても，KR が 1.0 を下回らない．従って，$\hat{\xi}^{\mathrm{1term}}_{\mathrm{gain}}$ による出力音声がミュージカルノイズフリー状態を達成するためには，前段の雑音抑圧量を大きくする必要がある．

次に，CD の結果を図 3 (c)，図 3 (d) に示す．図 3 (c)，図 3 (d) より，バイアス値を増加させると NRR と共に CD も減少する．さらに，NRR が等しい点において CD を比較すると，HRNR の CD よりバイアス付き HRNR の CD の方が小さいことから，音声の歪みも抑えられていることがわかる．各パラメータについて，バイアス値を増加させた場合の軌跡は $\varepsilon' = 0.0$ での点の NRR の大小関係を維持する．$\hat{\xi}^{\mathrm{1term}}_{\mathrm{gain}}$ における $\varepsilon' = 0.0$ での NRR より，$\hat{\xi}^{\mathrm{1term}}_{\mathrm{const}}$ における $(\rho', \varepsilon') = (0, 0)$ での NRR の方が大きいため，バイアス付き HRNR において，音声の歪みの点では小さい定数パラメータを設定する方が良い．以上をまとめると，KR と CD の観点でバイアス付き HRNR は HRNR より品質の良い雑音抑圧法であると言える．また，重みパラメータ ρ' にはスペクトルゲインより定数パラメータを用いる方が良く，その定数は小さい方が良い．

5. バイアス付き HRNR における事前 SNR の推定値と音質の考察

4.1 節で HRNR を適用することは有効であること，4.2 節で HRNR よりバイアス付き HRNR が有効であることを述べた．これらの優劣は事前 SNR の推定方法の違いによって決定づけられる．バイアス付き HRNR は DD や HRNR と比較して総合的に品質が良いことが示されたため，事前 SNR の推定精度が高いと考えられる．そこで，それぞれの手法における事前 SNR の推定値と真の事前 SNR を比較し，上記の仮説について実験的に調査する．

実験条件として，目的音声に BB を 10 dB の入力 SNR で加えたものを観測信号とし，各手法を用いて事前 SNR の推定値を求める．真の事前 SNR は雑音と目的信号のスペクトルは既知として，式 (5) より求める．各パラメータは NRR が 20 dB となるように，DD における忘却係数を $\alpha = 0.97$，HRNR に

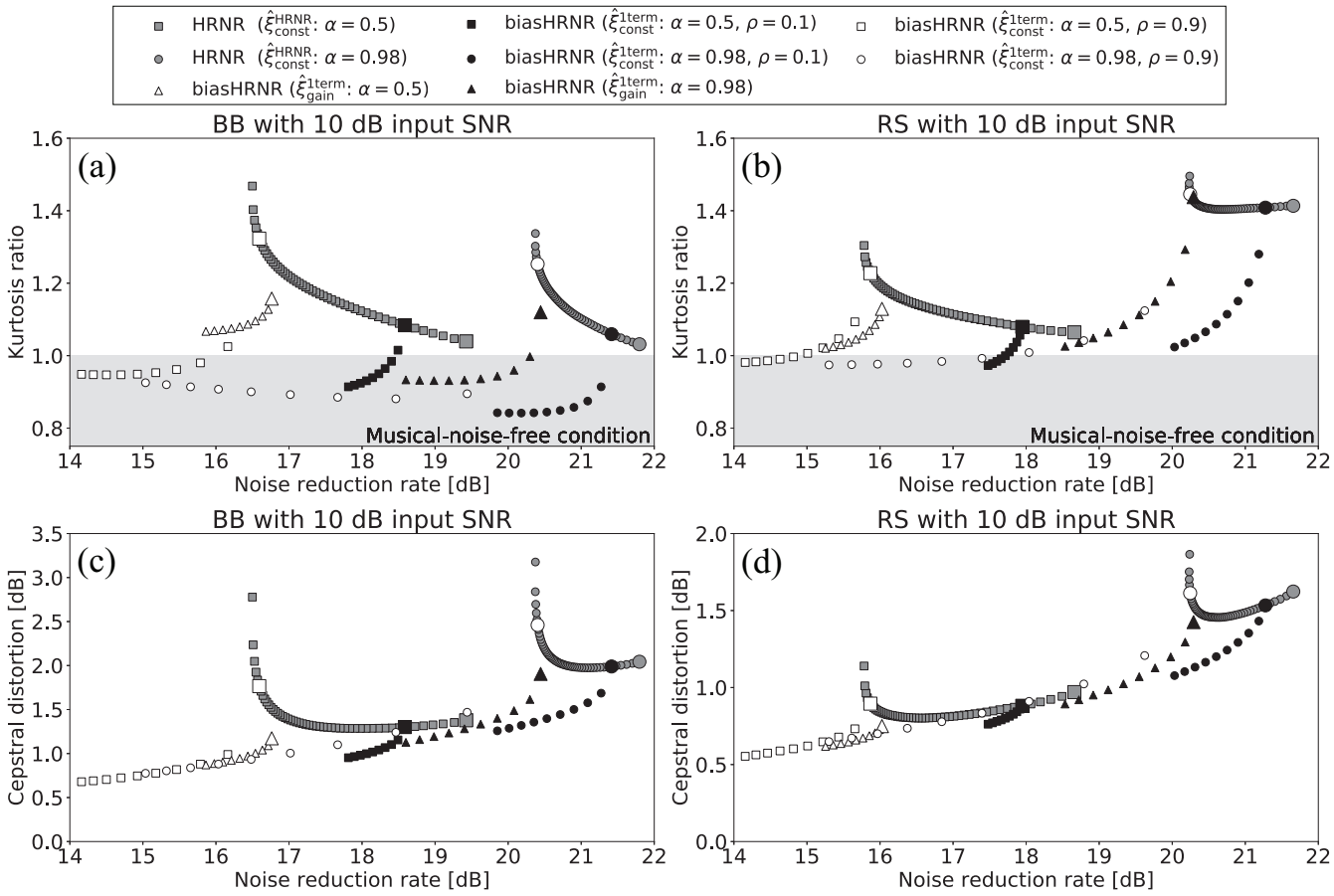

図 3　HRNR の重みパラメータとバイアス付き HRNR のバイアス値における NRR と KR の関係．(a)，(b): バブル雑音，駅雑音をそれぞれ入力 SNR10 dB で混合した場合の KR の結果．(c)，(d): バブル雑音，駅雑音をそれぞれ入力 SNR10 dB で混合した場合の CD の結果．

おける重みパラメータを $\rho = 0.04$，バイアス付き HRNR における重みパラメータを $\rho' = 0.1$，バイアスを $\varepsilon' = 0.8$ とした．その他の条件は 4.1 節と同様とした．

結果を図 4 に示す．図 4 の右上は真の事前 SNR を二次元のグラフ上に表したものであり，白線の部分で時間，または周波数方向に切り出した事前 SNR をそれぞれ左上と右下に示す．まず，真の SNR と DD による事前 SNR の推定値を比較する．時間・周波数のそれぞれの場合で，DD による事前 SNR が真の事前 SNR を下回っている．特に，図 4 左上では，高域になるにつれて事前 SNR の過小推定が著しい．事前 SNR の過小推定は出力音声に対して音声の歪みをもたらし，高域の過小推定は倍音成分の歪みの原因となる．また，図 4 右下より，DD の事前 SNR の変化は真の事前 SNR の変化に対して緩やかである．これは，DD は前のフレームを用いて平滑化されるためであり，非音声区間から音声が発される区間で遅延が発生し，事前 SNR を過小に推定している．以上から，DD による事前 SNR 推定では時間・周波数の両面において，その過小推定が原因で音声の歪みを引き起こしていることが確認できる．

次に，真の事前 SNR と HRNR による事前 SNR の推定値を比較する．図 4 左上より，HRNR は真の SNR より過大に推定している．これは，S_{harmo} による擬似的に復元したスペクトルによるものと考えられる．しかし，S_{harmo} は雑音抑圧後のスペクトル $\hat{S}$ に基づき倍音成分を復元しているため，真の事前 SNR のピークに合わせた事前 SNR を推定できており，結果的に，音声の歪みを抑えることができていると考えられる．

次に，バイアス付き HRNR によるバイアスの効果について考察する．本実験では $\varepsilon' = 0.8$ としたため，$10 \times \log(0.8) \approx -0.97$ dB より，図において -0.97 dB 以下の成分にバイアスがかけられていることが確認できる．そのため，バイアス付き HRNR は真の事前 SNR が小さい部分では著しく過大推定していることがわかる．一方で，バイアス付き HRNR は他の手法と比べ出力音声の品質が良いため，バイアスによる事前 SNR の過大推定は音声の品質の劣化には大きく寄与しないと言える．

上記で述べた，各事前 SNR の推定手法の過大・過小推定を裏付けるため，これらを客観的に評価する．雑音には BB, RS に加え，Street noise (ST)，White Gaussian noise (WG) の計 4 つを用い，これらを入力 SNR 0 dB，10 dB それぞれの場合で混合したものを観測信号とした．評価尺度に Log-error (LogErr) を用いた [24]．LogErr は真の値に対する過大推定値 $\mathrm{LogErr}_{\mathrm{ov}}$ と過小推定値 $\mathrm{LogErr}_{\mathrm{un}}$ の和として次のように表される．

$$\mathrm{LogErr} = \mathrm{LogErr}_{\mathrm{un}} + \mathrm{LogErr}_{\mathrm{ov}} \tag{26}$$

$$\mathrm{LogErr}_{\mathrm{ov}} = \frac{10}{PK} \sum_{p,k} \left| \mathrm{Min} \left[\log_{10} \frac{\xi(p,k)}{\hat{\xi}(p,k)}, 0 \right] \right| \tag{27}$$

$$\mathrm{LogErr}_{\mathrm{un}} = \frac{10}{PK} \sum_{p,k} \mathrm{Max} \left[\log_{10} \frac{\xi(p,k)}{\hat{\xi}(p,k)}, 0 \right] \tag{28}$$

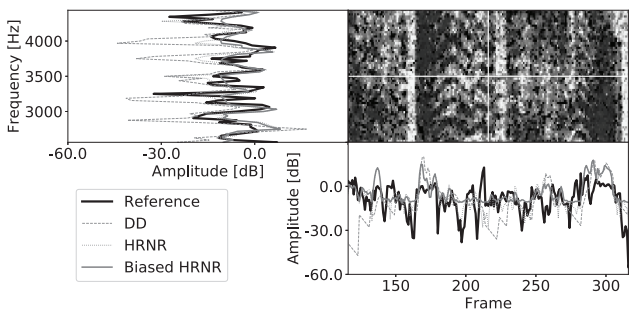

図 4　真の事前 SNR と各手法における事前 SNR の推定値との比較．周波数方向に切り出した事前 SNR（左上）．真の事前 SNR を二次元グラフ上に表現した図（右上）．時間方向に切り出した事前 SNR（右下）．

ここで，$\mathrm{Min}\,[a, b]$ は a と b のうち小さい方を出力する関数である．式 (26)〜(28) を用いて，DD，HRNR，バイアス付き HRNR における事前 SNR の推定値を算出し，それぞれの LogErr を算出した．また，各雑音ごとに NRR が概ね等しくなるよう，手法ごとにパラメータを変更した．

結果を図 5 に示す．まず，DD において，他の手法に比べて過小推定が多く発生している．この結果は図 4 の結果と一致する．次に，HRNR は DD に比べて事前 SNR の過小推定量が減少している．対して，事前 SNR の過大推定量が増加しており，HRNR においても図 4 から述べた考察と一致することがわかった．最後に，バイアス付き HRNR について，多くの場合においてバイアス付き HRNR の LogErr が大きいことから，真の事前 SNR との誤差の絶対量は他の手法に比べて大きいと言える．しかし，DD に比べ，HRNR やバイアス付き HRNR は過小推定を抑えられており，音声の歪みを改善できている．また，HRNR とバイアス付き HRNR を比べると，後者の方がさらに過小推定を抑えられている．これは，バイアスによるものと考えられる．一方，特にバイアス付き HRNR は事前 SNR を過大推定しており，図 4 から述べた考察に沿った結果であると言える．ただし，4.2 節や図 3 の結果から，バイアス付き HRNR の音質は DD や HRNR と比較して総合的に良いことが示されたため，バイアスによる事前 SNR の過大推定は音声の品質の劣化させない．

6. ま と め

本研究では，古典的な事前 SNR の推定手法である DD と対比しながら，HRNR とバイアス付き HRNR の内部パラメータと音質の関係について調査した．また，HRNR とバイアス付き HRNR において，重みパラメータにスペクトルゲインを用いた場合についても調査した．さらに，それらの手法の事前 SNR の推定値と真の事前 SNR を比較し，音質の関係について考察した．

DD と HRNR の内部パラメータと音質を比較し．HRNR を適用することは有用であることを示した．結果として，ミュージカルノイズの発生の点では重みパラメータに小さい値の定数を用いる方が有効であり，音声の歪みの点では重みパラメータ

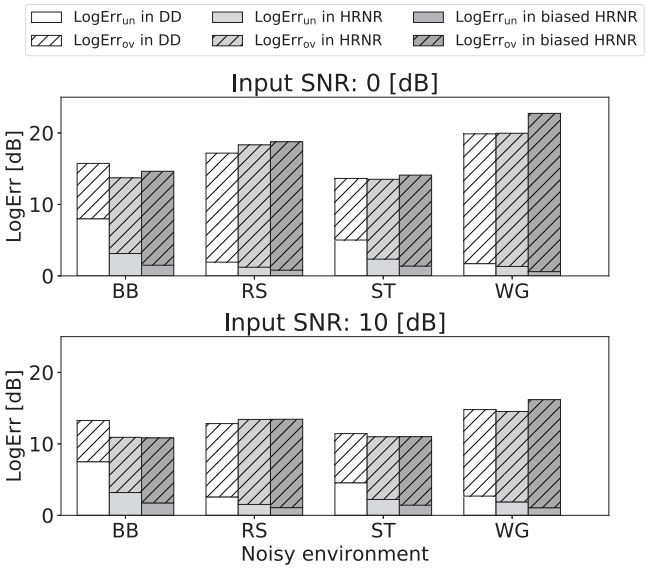

図 5　各事前 SNR 推定手法の LogErr の結果．入力 SNR が 0 dB の場合の LogErr の結果 (上) と 10 dB の場合の LogErr の結果 (下)．

にスペクトルゲインを用いることが有効であることがわかった．次に，HRNR とバイアス付き HRNR における内部パラメータと音質を比較し，バイアスを導入することはミュージカルノイズの発生，音声の歪みの両問題に対して有効であることを示した．その際の重みパラメータは定数パラメータを用い，その値は小さく設定する方が良いことがわかった．

最後に，事前 SNR の各推定手法について比較し，HRNR は DD による事前 SNR の過小推定を防ぎ，音声の歪みを改善していることを示した．バイアス付き HRNR についても，バイアスの導入することで過小推定を防ぐ一方，事前 SNR を過大推定していることもわかった．しかし，バイアス付き HRNR による音声の品質が他の手法に比べ優れているという知見から，バイアスによる事前 SNR の過大推定は音質を大きく劣化させないことがわかった．

以上をまとめると，バイアス付き HRNR は DD や HRNR と比べ，効果的な雑音抑圧手法であり，また，事前 SNR の推定においてバイアスを導入することは有効である．

文　献

[1] R. Z. J. L. Flanagan, J. D. Johnston and G. W. Elko: “Computer-streered microphone arrays for sound transduction in large rooms”, Journal of the Acoustical Society of America, **78**, 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, 11, pp. 1135–1146 (2003).

[3] S. F. Boll: “Suppression of acoustic noise in speech using spectral subtraction”, IEEE Transactions on Acoustics, Speech and Signal Processing, **27**, 2, pp. 113–120 (1979).

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

[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, **27**, 6, pp. 1109–1121 (1984).

[6] K. Yamashita, S. Ogata and T. Shimamura: "Spectral subtraction iterated with weighting factors", Proceedings of IEEE Speech Coding Workshop, pp. 138–140 (2002).

[7] C. H. You, S. N. Koh and S. Rahardja: "β-order mmse spectral amplitude estimation for speech enhancement", IEEE Transactions on Acoustics, Speech and Signal Processing, **13**, 5, pp. 475–486 (2005).

[8] J. Benesty and Y. Huang: "A single-channel noise reduction MVDR filter", Proceedings of International Conference on Acoustics, Speech and Signal Processing, pp. 273–276 (2011).

[9] P. C. Loizou: "Speech enhancement theory and practice", CRC Press, Taylor & Francis Group, FL (2007).

[10] Z. Goh, K. C. Tan and B. Tan: "Postprocessing method for suppressing musical noise generated by spectral subtraction", IEEE Transactions on Speech and Audio Processing, **6**, 3, pp. 287–292 (1998).

[11] C. Plapous, C. Marro and P. Scalart: "Improved signal-to-noise ratio estimation for speech enhancement", IEEE Transactions on Audio, Speech, and Language Processing, **14**, 6, pp. 2098–2108 (2006).

[12] M. Une and 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).

[13] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano and K. Kondo: "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction", IEEE Transactions on Audio, Speech and Language Processing, **20**, 7, pp. 2080–2094 (2012).

[14] S. Nakai, H. Saruwatari, R. Miyazaki, S. Nakamura and K. Kondo: "Theoretical analysis of biased mmse short-time spectral amplitude estimator and its extension to musical-noise free speech enhancement", Joint Workshop on Hands-free Speech Communication and Microphone Arrays, pp. 122–126 (2014).

[15] H. Saruwatari: "Statistical-model-based speech enhancement with musical-noise-free properties", Proceedings of International Conferenece on Digital Signal Processing, pp. 1201–1205 (2015).

[16] M. Une and R. Miyazaki: "Musical-noise-free speech enhancement with low speech distortion by biased harmonic regeneration technique", Proceedings of International Workshop on Acoustic Signal Enhancement, pp. 31–35 (2018).

[17] C. Breithaupt, T. Gerkmann and R. Martin: "A novel a priori snr estimation approach based on selective cepstro-temporal smoothing", Proceedings of International Conference on Acoustics, Speech and Signal Processing, pp. 4897–4900 (2008).

[18] S. Kubo and R. Miyazaki: "Estimation of spectral subtraction parameter-set for maximizing speech recognition performance", 5th IEEE Global Conference on Consumer Electronics, pp. 567–568 (2016).

[19] S. Kubo and R. Miyazaki: "Estimation of beta-order mmse-stsa parameter set for maximizing speech recognition performance with multiple regression analysis", 2018 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, pp. 180–183 (2018).

[20] O. Cappe: "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor", IEEE Transactions on Speech and Audio Processing, **2**, 2, pp. 345–349 (1994).

[21] S. Kanehara, H. Saruwatari, R. Miyazaki, K. Shikano and 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 Signal Enhancement (2012).

[22] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano and K. Kondo: "Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics", Proceedings of International Workshop on Acoustic Proceedings of International Workshop on Acoustic Echo and Noise Control (2008).

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

[24] T. Gerkmann and R. C. Hendriks: "Unbiased mmse-based noise power estimation with low complexity and low tracking delay", IEEE/ACM Transactions on Audio, Speech, and Language Processing, **20**, 4, pp. 1383–1393 (2012).