Musical Noise Analysis for Bayesian Minimum Mean-Square Error Speech Amplitude Estimators Based on Higher-Order Statistics

Hiroshi Saruwatari[1], *Suzumi Kanehara*[1], *Ryoichi Miyazaki*[1], *Kiyohiro Shikano*[1], *Kazunobu Kondo*[2]

[1]Nara Institute of Science and Technology, Nara, 630-0192 Japan
[2] YAMAHA Corporate Research & Development Center, Shizuoka, 438-0192 Japan

Abstract

In this study, we perform a theoretical analysis of the amount of musical noise generated in Bayesian minimum mean-square error speech amplitude estimators. In our previous study, a musical noise assessment based on kurtosis has been successfully applied to spectral subtraction. However, it is difficult to apply this approach to the methods with a decision-directed a priori SNR estimator because it corresponds to a nonlinear recursive process for noise power spectral sequences. Therefore, in this paper, we analyze musical noise generation by combining Breithaupt-Martin's approximation and our higher-order-statistics analysis. We also compare the result of theoretical analysis and that of objective experimental evaluation to indicate the validity of the proposed closed-form analysis.

Index Terms: noise reduction, musical noise, higher-order statistics, kurtosis, decision-directed a priori SNR estimator

1. Introduction

Over the past decade, the number of applications of speech communication systems, such as TV conference systems and mobile phones, has increased. These systems, however, always suffer from a problem of deterioration of speech quality under adverse noise conditions. Therefore, in speech signal processing, noise reduction is a problem requiring urgent attention. Spectral subtraction (SS) [1], Wiener filter (WF) [2], and the minimum mean-square error short-time spectral amplitude (MMSE STSA) estimator [3] are the commonly used noise reduction methods that have high noise reduction performance. However, in these methods, artificial distortion, so-called musical noise, arises owing to nonlinear signal processing, leading to a serious deterioration of sound quality [4].

Recently, an objective metric to measure how much musical noise is generated through nonlinear signal processing based on higher-order statistics has been developed by some of the authors [5, 6, 7] and others [8, 9]. Using this metric, we have successfully analyzed the amount of musical noise generated via various types of SS-based methods [10, 11, 12, 13, 14, 15]. However, it is still difficult to theoretically analyze the noise reduction methods with a decision-directed a priori SNR estimator (hereafter, this is referred to as *the DD approach*), e.g., the MMSE STSA estimator, because it corresponds to a nonlinear recursive (infinite) process for noise power spectral sequences. Several studies on the systematic analysis of the DD approach have been provided [4, 16], but they did not use the explisit metric of musical noise generation like higher-order statistics so far.

In this paper, we focus on an approximated model [16] proposed by Breithaupt et al. We propose to analyze musical noise generation in DD-approach-based WF, the MMSE STSA estimator and the minimum mean-square error log-spectral amplitude (MMSE LSA) estimator [17] by combining Breithaupt's approximation and our higher-order-statistics analysis. We also compare the result of theoretical analysis and that of objective experimental evaluation to indicate the validity of the proposed closed-form analysis.

2. Related works

2.1. Mathematical metric of musical noise generation via higher-order statistics [5, 6]

We speculate that the amount of musical noise is highly correlated with the number of isolated power spectral components and their level of isolation. In this paper, we call these isolated components *tonal components*. Since such tonal components have relatively high power, they are strongly related to the weight of the tail of their probability density function (p.d.f.). Therefore, quantifying the tail of the p.d.f. makes it possible to measure the number of tonal components. Thus, we adopt *kurtosis*, one of the most commonly used higher-order statistics, to evaluate the percentage of tonal components among the total components. A large kurtosis value indicates a signal with a heavy tail, meaning that the signal has many tonal components. Kurtosis is defined as

$$\mathrm{kurt} = \mu_4/\mu_2^2, \tag{1}$$

where kurt is the kurtosis and μ_m is the mth-order moment as

$$\mu_m = \int_0^\infty z^m p(z) dz, \tag{2}$$

where $p(z)$ is the p.d.f. of a signal z in the power spectral domain.

In this study, we apply such a kurtosis-based analysis to a *noise-only time-frequency period* of subject signals for the assessment of musical noise. Thus, this analysis should be conducted during, e.g., periods of silence during speech. This is because we aim to quantify the tonal components arising in the noise-only part, which is the main cause of musical noise perception, and not in the target-speech-dominant part.

Although kurtosis can be used to measure the number of tonal components, note that the kurtosis itself is not sufficient to measure the amount of musical noise. This is obvious since the kurtosis of some unprocessed noise signals, such as an interfering speech signal, is also high, but we do not recognize speech as musical noise. Hence, we turn our attention to the change in kurtosis between before and after signal processing to identify only the musical-noise components. Thus, we adopt

This work was partly supported by JST CREST, Japan.

the *kurtosis ratio* as a measure to assess musical noise [5]. This measure is defined as

$$\text{kurtosis ratio} = \text{kurt}_{\text{proc}}/\text{kurt}_{\text{org}}, \tag{3}$$

where $\text{kurt}_{\text{proc}}$ is the kurtosis of the processed signal and kurt_{org} is the kurtosis of the observed signal. This measure increases as the amount of generated musical noise increases. In Ref. [5], it was reported that the kurtosis ratio is strongly correlated with the human perception of musical noise.

2.2. Analysis of amount of noise reduction [12]

We analyze the amount of noise reduction via processing. Hereafter, we define the *noise reduction rate* (NRR) [18, 19] as a measure of the noise reduction performance, which is defined as the output SNR in dB minus the input SNR in dB. The NRR is

$$\text{NRR} = 10\log_{10}(\text{E}[s_{\text{out}}^2]/\text{E}[n_{\text{out}}^2])/(\text{E}[s_{\text{in}}^2]/\text{E}[n_{\text{in}}^2]), \tag{4}$$

where s_{in} and s_{out} are the input and output speech signals, and n_{in} and n_{out} are the input and output noise signals, respectively. If we assume that the amount of noise reduction is much larger than that of speech distortion in processing, i.e., $\text{E}[s_{\text{out}}^2] \simeq \text{E}[s_{\text{in}}^2]$, then

$$\text{NRR} \simeq 10\log_{10}\text{E}[n_{\text{in}}^2]/\text{E}[n_{\text{out}}^2] = 10\log_{10}\mu_1/\mu_1', \tag{5}$$

where μ_1 is the 1st-order moment of observed signal power spectra, and μ_1' is the 1st-order moment of processed signal power spectra.

2.3. DD-approach-based WF [2]

We apply short-time Fourier analysis to the observed signal, which is a mixture of target speech and noise, to obtain the time-frequency signal $X(f,\tau) = S(f,\tau) + N(f,\tau)$, where $X(f,\tau)$ is the observed signal, f denotes the frequency subband, and τ is the frame index. $S(f,\tau)$ and $N(f,\tau)$ denote the input speech and noise signals. The signal processing procedures of WF are formulated as

$$Y(f,\tau) = \xi(f,\tau)/(\xi(f,\tau)+1)\cdot X(f,\tau), \tag{6}$$

where $Y(f,\tau)$ is the enhanced target speech signal. Also, $\xi(f,\tau)$ and $\gamma(f,\tau)$ are a priori and a posteriori SNRs, which are defined as

$$\xi(f,\tau) = \text{E}[|S(f,\tau)|^2]/\text{E}[|N(f,\tau)|^2], \tag{7}$$

$$\gamma(f,\tau) = |X(f,\tau)|^2/\text{E}[|N(f,\tau)|^2]. \tag{8}$$

In (7) and (8), we can commonly estimate $\text{E}[|N(f,\tau)|^2]$ by averaging the noise power spectra in the speech absent time period, or by using other estimation methods [3, 16]. However, since we cannot estimate $\text{E}[|S(f,\tau)|^2]$ in advance, a priori SNR $\xi(f,\tau)$ is approximately calculated by the following DD approach;

$$\hat{\xi}(f,\tau) = \alpha\gamma(f,\tau-1)G^2(f,\tau-1) + (1-\alpha)F[\gamma(f,\tau)-1], \tag{9}$$

where α is a forgetting factor and $F[\cdot]$ is a flooring function.

2.4. MOSIE estimator [16]

In this paper, we introduce the MOSIE estimator proposed by Breithaupt et al., which generalizes the MMSE STSA estimator and MMSE LSA estimator [16]. The processed signal $Y(f,\tau)$ via the MOSIE estimator is written as

$$Y(f,\tau) = \sqrt{\xi(f,\tau)/(\rho+\xi(f,\tau))}\sqrt{P_{\hat{N}}(f)} \left[\frac{\Gamma(\rho+\frac{\beta}{2})}{\Gamma(\rho)}\frac{\Phi(1-\rho-\frac{\beta}{2},1;-\nu(f,\tau))}{\Phi(1-\rho,1;-\nu(f,\tau))}\right]^{1/\beta} e^{j\arg(X(f,\tau))}, \tag{10}$$

where $\nu(f,\tau) = \xi(f,\tau)\gamma(f,\tau)(\rho+\xi(f,\tau))^{-1}$, $P_{\hat{N}}(f)$ is a power spectral density estimated from the speech absence period of the observed signal, $\Phi(a,b;k) = {}_1F_1(a,b;k)$ is the confluent hypergeometric function, and $\Gamma(\cdot)$ is the gamma function. Also, β is an amplitude compression parameter with the error function $e(S(f,\tau),Y(f,\tau)) = S^\beta(f,\tau) - Y^\beta(f,\tau)$, and ρ is a shape parameter of a chi-distribution that is used for modeling the p.d.f. of the speech amplitude. Equation (10) corresponds to the MMSE STSA estimator with $\rho = 1, \beta = 1$, and to the MMSE LSA estimator with $\rho = 1, \beta \rightarrow 0$.

3. Theoretical analysis of higher-order statistics in WF and MOSIE estimator

3.1. Motivation and strategy

In our previous report [20, 21], we tried to calculate the kurtosis ratio in the MMSE STSA estimator based on the multi-dimensional numerical integration. It required, however, too huge computation like Tera–Giga order multiply-accumulation, resulting in several-week or -day calculations for each parameter setting. Therefore, in this paper we propose a new computational-cost-efficient closed-form analysis.

First, we model the noise signal power spectra x using the following gamma distribution as

$$p(x) = x^{\eta-1}\exp(-x/\theta)(\theta^\eta\Gamma(\eta))^{-1}, \tag{11}$$

where η is the shape parameter corresponding to the type of noise, and θ is the scale parameter of the gamma distribution. If the input signal is Gaussian noise, the p.d.f. of its power spectra obeys the chi-square distribution with two degrees of freedom, which corresponds to the gamma distribution with $\eta = 1$. Also, if the input signal is super-Gaussian noise, the p.d.f. of its power spectra obeys the gamma distribution with $\eta < 1$. Hereafter, to analyze the NRR and kurtosis ratio in WF and the MOSIE estimator with the DD approach, we formulate the mth-order moment using several useful approximations.

3.2. Theoretical analysis of DD-approach-based WF

WF *without* the DD approach has been analized using higher-order statistics in our previous study [12], but no analysis for DD-approach-based WF has been provided so far; therefore we give it in this section. Since the DD approach (9) requires an infinite number of samples from the past, we cannot estimate how the p.d.f. of noise will change via WF. In this paper, we introduce an approximation defined below as the estimation of a priori SNR [16]

$$\hat{\xi}(f,\tau) \approx (1-\alpha)\xi^{\text{ml}}(f,\tau), \tag{12}$$

$$\xi^{\text{ml}}(f,\tau) = \max\{0, \gamma(f,\tau)-1\}, \tag{13}$$

where $\xi^{\mathrm{ml}}(f,\tau)$ corresponds to a maximum-likelihood estimate of the a priori SNR. The estimation of the processed signal via WF is rewritten as follows, where we use (12) instead of (9);

$$Y(f,\tau) \approx \frac{(1-\alpha)\xi^{\mathrm{ml}}(f,\tau)}{(1-\alpha)\xi^{\mathrm{ml}}(f,\tau)+1}X(f,\tau). \tag{14}$$

The p.d.f. of the observed signal, $p(x)$, is transformed into $p(y)$ by signal processing. We can calculate $p(y)$ by considering a change of variables of the p.d.f. Suppose that a change of variables, $y = g(x)$, is applied to convert an integral in terms of the variable x to an integral in terms of the variable y. The converted p.d.f. $p(y)$ can be written as $p(y) = p(g^{-1}(y))|J|$, where $|J| = |dg^{-1}/dy|$ is the Jacobian of the transformation. We apply this to (11) to obtain the p.d.f. after processing, $p(y)$. Since x is the power spectral domain signal and its mean value $\mathrm{E}[|N(f,\tau)|^2]$ is given by $\eta\theta$ in the gamma distribution, the variable y used for processing is expressed as

$$y = \begin{cases} \frac{(1-\alpha)(\frac{x^2}{\eta\theta}-x)}{(1-\alpha)(\frac{x}{\eta\theta}-1)+1}, & \text{if } x \geq \eta\theta \\ 0, & \text{if } x < \eta\theta. \end{cases} \tag{15}$$

Since $x > 0$ and $y > 0$, the Jacobian is $dx/dy = h'(y) = |J|$, where $h(\cdot)$ is the inverse function of $g(\cdot)$. Consequently,

$$p(y) = h(y)^{\eta-1}\exp(-h(y)/\theta)(\Gamma(\eta)\theta^{\eta})^{-1}h'(y). \tag{16}$$

The mth-order moment of $p(y)$ is given by

$$\mu_m = \int_{\theta\eta}^{\infty} y^m \frac{h(y)^{\eta-1}\exp(-\frac{h(y)}{\theta})}{\Gamma(\eta)\theta^{\eta}}h'(y)dy. \tag{17}$$

Let $t = h(y)/\theta$, then $dy = \theta/h'(y)dt$ and the range of the integral does not change. Furthermore, $h(y)$ is expressed as

$$h(y) = t\theta = x. \tag{18}$$

We apply (18) to (15), then y^m is expressed as

$$y^m = \left(\frac{(1-\alpha)(\frac{\theta^2 t^2}{\theta\eta} - \theta t)}{(1-\alpha)(\frac{t\theta-1}{\theta\eta})+1}\right)^m. \tag{19}$$

Then, we apply (18) and (19) to (17) to obtain

$$\mu_m = \frac{(1-\alpha)^m\theta^m}{\Gamma(\eta)}\int_{\theta\eta}^{\infty}\left(\frac{\frac{t^2}{\eta}-t}{(1-\alpha)\frac{t}{\eta}+\alpha}\right)^m t^{\eta-1}\exp(-t)dt. \tag{20}$$

Here, the integration term in (20) can be easily calculated using a simple one-dimensional numerical integration method.

3.3. Theoretical analysis of MOSIE estimator

Hereafter, we assume $\rho = 1$. First, in order to simplify the expression (10), we use the first and second terms of the Taylor series expansion $\xi(f,\tau)/(1+\xi(f,\tau)) \approx \xi(f,\tau) - \xi^2(f,\tau)$, which is valid for low SNR conditions with $\xi(f,\tau) \ll 1$. In addition, we introduce some approximations [16], then (10) can be rewritten as

$$|Y(f,\tau)| \approx \sqrt{\xi(f,\tau)-\xi^2(f,\tau)}\sqrt{P_{\hat{N}}(f)}\sqrt{R(\beta)}, \tag{21}$$

where

$$R(\beta) = \begin{cases} \Gamma(1+\beta/2)^{2/\beta}, & \text{if } \beta \neq 0 \\ \exp(-\mathrm{c}), & \text{if } \beta = 0 \end{cases} \tag{22}$$

and $\mathrm{c} = 0.5772...$ is Euler's constant.

Next, we introduce an approximation defined below as an estimate of a priori SNR [16] such as the case of WF,

$$\hat{\xi}(f,\tau) \approx h_{\xi}(f,\tau) * \xi^{\mathrm{ml}}(f,\tau), \tag{23}$$

where $*$ is an operator of convolution, and the variables in (23) are defined as

$$h_{\xi}(f,\tau) = (1-\alpha)\exp(-\lambda_{\xi}\tau), \tag{24}$$

$$\lambda_{\xi} = \ln\left(\frac{1}{\alpha R(\beta)}\right). \tag{25}$$

Equation (21) can be rewritten as follows by using (23) instead of (9),

$$\begin{aligned} |Y(f,\tau)|^2 \approx & (h_{\xi}(f,\tau) * \xi^{\mathrm{ml}}(f,\tau))R(\beta)P_{\hat{N}}(f) \\ & - (h_{\xi}(f,\tau) * \xi^{\mathrm{ml}}(f,\tau))^2 R(\beta)P_{\hat{N}}(f). \end{aligned} \tag{26}$$

Since the convolution calculation exists in (26), it is difficult to directly calculate the mth-order moment of the processed signal via the MOSIE estimator. To solve this problem, first, we calculate the cumulant of the processed signal, and then we calculate the moment of the processed signal using the transformation formula of cumulants and moments. From (26), we can obtain the mth-order cumulant of the processed signal as

$$\begin{aligned} & K_m(|Y(f,\tau)|^2) = \\ & ((1-\alpha)R(\beta)\eta\theta)^m K_m(\xi^{\mathrm{ml}}(f,\tau))\sum_{\tau=1}^{\infty}\exp(-m\lambda_{\xi}\tau) \\ & - ((1-\alpha)^2 R(\beta)\eta\theta)^m K_m((\xi^{\mathrm{ml}}(f,\tau))^2) \\ & \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=1}^{\infty}\exp(-m\lambda_{\xi}\tau_1)\exp(-m\lambda_{\xi}\tau_2), \end{aligned} \tag{27}$$

where $K_m(\cdot)$ is the mth-order cumulant of $\cdot$. Using the sum of a geometric series, we can rewrite (27) as

$$\begin{aligned} & K_m(|Y(f,\tau)|^2) = \\ & \sum_{i=1}^{2}(-1)^{i-1}(1-\alpha)^{im}(R(\beta)\eta\theta)^m Z_i(m)K_m((\xi^{\mathrm{ml}}(f,\tau))^i), \end{aligned} \tag{28}$$

where $Z_i(m) = \exp(-im\lambda_{\xi})/(1-\exp(-m\lambda_{\xi}))^i$. Here, we introduce an analogy that the sequence $\xi^{\mathrm{ml}}(f,\tau) = \max\{0, \gamma(f,\tau)-1\}$ is generated from the process of normalized SS with the oversubtraction parameter of 1 and the flooring parameter of 0. Thanks to this analogy, we can obtain the mth-order moment of $\xi^{\mathrm{ml}}(f,\tau)$ in a closed form [12] as

$$\mu_{m[SS]} = \sum_{l=0}^{m}(-\eta)^l\frac{\Gamma(m+1)\Gamma(\eta+m-l,\eta)}{\Gamma(\eta)\Gamma(l+1)\Gamma(m-l+1)}. \tag{29}$$

Using the transformation formula from moments to cumulants, $K_m((\xi^{\mathrm{ml}}(f,\tau))^i)$ on the right-hand side of (28) can be expressed as

$$\begin{cases} K_1((\xi^{\mathrm{ml}}(f,\tau))^i) = & \mu_{i[SS]} \\ K_2((\xi^{\mathrm{ml}}(f,\tau))^i) = & \mu_{2i[SS]} - \mu_{i[SS]}^2 \\ K_3((\xi^{\mathrm{ml}}(f,\tau))^i) = & \mu_{3i[SS]} - 3\mu_{2i[SS]}\mu_{i[SS]} + 2\mu_{i[SS]}^3 \\ K_4((\xi^{\mathrm{ml}}(f,\tau))^i) = & \mu_{4i[SS]} - 4\mu_{3i[SS]}\mu_{i[SS]} - 3\mu_{2i[SS]}^2 \\ & +12\mu_{2i[SS]}\mu_{i[SS]}^2 - 6\mu_{i[SS]}^4 \end{cases}, \tag{30}$$

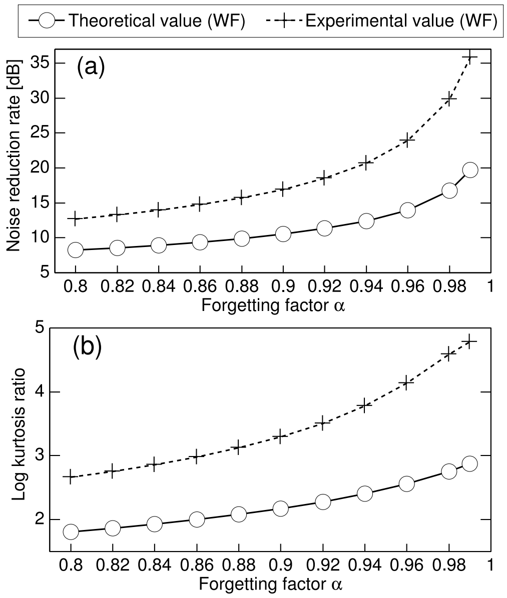

Figure 1: Theoretical behavior and experimental results for WF: (a) noise reduction rate, and (b) log kurtosis ratio.

resulting in the closed-form calculation of $K_m(|Y(f,\tau)|^2)$ in (28). Using the transformation formula from cumulants to moments, we transform the cumulant (28) into the corresponding moment. Finally, each order moment of the processed signal is given by

$$\begin{cases}\mu_1 = & K_1(|Y(f,\tau)|^2)\\ \mu_2 = & K_2(|Y(f,\tau)|^2)+K_1^2(|Y(f,\tau)|^2)\\ \mu_4 = & K_4(|Y(f,\tau)|^2)+4K_3(|Y(f,\tau)|^2)K_1(|Y(f,\tau)|^2)\\ & +3K_2^2(|Y(f,\tau)|^2)+6K_2(|Y(f,\tau)|^2)K_1^2(|Y(f,\tau)|^2)\\ & +K_1^4(|Y(f,\tau)|^2)\end{cases} . \tag{31}$$

4. Experiment

4.1. Experimental conditions

We calculated the NRR and the kurtosis ratio using (20) and (31), respectively. The shape parameter η of the noise p.d.f. is set to 1.0, and the forgetting factor α is varied from 0.8 to 0.99. Since the kurtosis of the processed signal changes exponentially, we depict the logarithm of the kurtosis ratio, which is referred to as the *log kurtosis ratio*. If the log kurtosis ratio is large, it denotes that much musical noise generated. If the log kurtosis ratio equals zero, it means that there is no musical noise generated.

In addition, we conducted a real noise reduction experiment in order to confirm the validity of our proposed theoretical analysis. The NRR and the log kurtosis ratio are calculated from actual noise reduction results obtained from the observed signals and processed signals. In the evaluation experiment, the noisy observed signals were generated by adding noise signals to target speech signals with an SNR of 0 dB. The target speech signals were the utterances of four speakers (four sentences). The length of each signal was 15 s, and each signal was sampled at 16 kHz. The FFT size is 1024, and the frame shift length is 256. In these experiments, we calculated the noise prototype, i.e., the average of $|\hat{N}(f,\tau)|^2$, in the first 10 s frames, where the speech signal is absent.

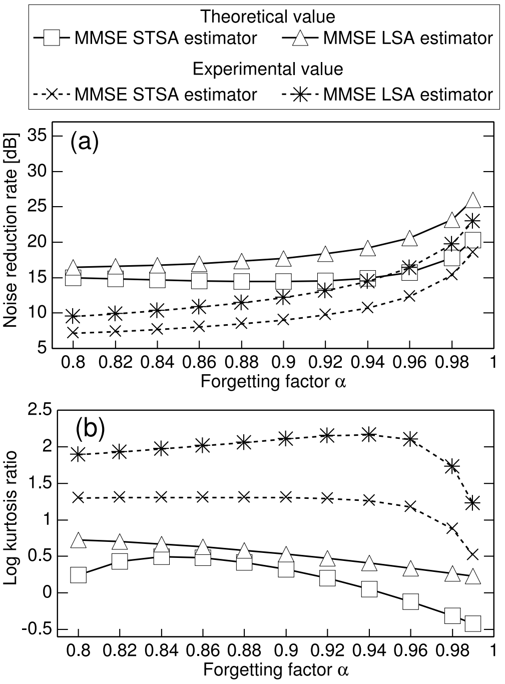

Figure 2: Theoretical behavior and experimental results for MMSE STSA estimator and MMSE LSA estimator: (a) noise reduction rate, and (b) log kurtosis ratio.

4.2. Results

Figures 1 and 2 show the theoretical behaviors calculated using (20) and (31), and the objective evaluations for each of the noise reduction methods. Regarding the detailed behavior of the NRR, all the estimators show the same tendency in that the NRR becomes higher as the larger forgetting factor α is used (see Figs. 1(a) and 2(a)).

Regarding the log kurtosis ratio, WF shows a monotonic increase of the kurtosis ratio as α increases (see Fig. 1(b)), indicating that the NRR and musical noise generation share a trade-off relationship. Consequently, there is no justification of using large α in WF. In contrast, we can see that the log kurtosis ratio drops when large α is used in the MMSE STSA estimator and MMSE LSA estimator (see Fig. 2(b)), whereas WF does not show such a kurtosis-ratio drop. From the results, we can speculate that the kurtosis-ratio drop is the key factor of less musical noise property in the MOSIE estimator when we set a large α in the DD approach, unlike WF. In addition, our approximated theoretical analysis can successfully explain the tendency in good agreement with the experimantal results.

5. Conclusion

In this study, we performed a theoretical analysis of the amount of musical noise generated via WF and the MOISE estimator with the DD approach on the basis of higher-order statistics. In particular, we derived higher-order statistics in the closed form using the approximated model proposed by Breithaupt et al. From this result, we can well explain the tendency of the behaviors in the NRR and kurtosis ratio for each method. Also, we can speculate that the kurtosis-ratio drop is the key factor of less musical noise property in the MOISE estimator when we set a large forgetting factor in the DD approach, unlike WF (even when WF is based on the DD approach).

6. References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, Signal Processing,* vol.ASSP-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 Trans. Acoustics, Speech and Signal Processing*, vol.32, no.6, pp.1109–1121, 1984.

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

[5] 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," *Proc. IWAENC2008*, 2008.

[6] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Musical noise generation analysis for noise reduction methods based on spectral subtraction and MMSE STSA estimation," *Proc. ICASSP2009*, pp.4433–4436, 2009.

[7] Y. Takahashi, R. Miyazaki, H. Saruwatari, K. Kondo, "Theoretical analysis of musical noise in nonlinear noise reduction based on higher-order statistics," *Proc. 2012 APSIPA Annual Summit and Conference (APSIPA2012)*, 2012.

[8] H. Yu and T. Fingscheidt, "A figure of merit for instrumental optimization of noise reduction algorithms," *Proc. DSP in Vehicles*, 2011.

[9] H. Yu and T. Fingscheidt, "Black box measurement of musical tones produced by noise reduction systems," *Proc. ICASSP2012*, pp.4573–4576, 2012.

[10] Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Musical noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics," *EURASIP Journal on Advances in Signal Processing*, vol.2010, Article ID 431347, 25 pages, 2010.

[11] H. Saruwatari, Y. Ishikawa, Y. Takahashi, T. Inoue, K. Shikano, amd K. Kondo, "Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher-order statistics," *IEEE Trans. Audio, Speech and Language Processing*, vol.19, no.6, pp.1457–1466, 2011.

[12] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in generalized spectral subtraction based on higher-order statistics," *IEEE Trans. Audio, Speech and Language Processing*, vol.19, no.6, pp.1770–1779, 2011.

[13] R. Miyazaki, H. Saruwatari, and K. Shikano, "Theoretical analysis of amount of musical noise and speech distortion in structure-generalized parametric blind spatial subtraction array," *IEICE Transactions Fundamentals*, vol.95-A, no.2, pp.586–590, 2011.

[14] 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 Trans. Audio, Speech and Language Processing*, vol.20, no.7, pp.2080–2094, 2012.

[15] R. Miyazaki, H. Saruwatari, K. Shikano, K. Kondo, "Musical-noise-free blind speech extraction using ICA-based noise estimation and iterative spectral subtraction," *Proc. 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA2012)*, pp.322–327, 2012.

[16] C. Breithaupt and R. Martin, "Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions," *IEEE Trans. Audio, Speech, and Language Processing*, vol.19, no.2, pp.277–289, 2011.

[17] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol.33, no.2, pp.443–445, 1985.

[18] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, and T. Nishikawa, "Blind source separation combining independent component analysis and beamforming," *EURASIP J. Appl. Signal Process.*, vol.2003, pp.1135–1146, 2003

[19] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Transactions on Speech and Audio Processing*, vol.14, no.2, pp.666–678, 2006.

[20] 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," *Proc. IWAENC2012*, 2012.

[21] S. Kanehara, H. Saruwatari, R. Miyazaki, K. Shikano, and K. Kondo, "Comparative study on various noise reduction methods with decision-directed a priori SNR estimator via higher-order statistics," *Proc. 2012 APSIPA Annual Summit and Conference (APSIPA2012)*, 2012.