Comparative Study on Various Noise Reduction Methods with Decision-Directed a Priori SNR Estimator via Higher-Order Statistics

Suzumi Kanehara*, Hiroshi Saruwatari*, Ryoichi Miyazaki*, Kiyohiro Shikano* and Kazunobu Kondo†
* Nara Institute of Science and Technology, Nara, 630-0192 Japan
E-mail: {suzumi-k, sawatari, ryoichi-m, shikano}@is.naist.jp
† Yamaha Corporate Research & Development Center, Shizuoka, 438-0192 Japan
E-mail: tashin@beat.yamaha.co.jp

***Abstract*—In this paper, we propose a new theoretical analysis of amount of musical noise generated in several noise reduction methods with a decision-directed a priori SNR estimator using higher-order statistics. In our previous study, a musical noise assessment based on kurtosis has been successfully applied to spectral subtraction and Wiener filter. However, this approach cannot be applied to some high-quality noise reduction methods, namely, the minimum mean-square error short-time spectral amplitude estimator, the minimum mean square error log-spectral amplitude estimator and the maximum a posteriori estimator, because such methods include the decision-directed a priori SNR estimator, which corresponds to a nonlinear recursive (infinite) process for noise power spectral sequences. Therefore, in this paper, we introduce a computationally efficient higher-order-moment calculation method based on generalized Gauss-Laguerre quadrature. We also mathematically clarify the justification of using a typical decision-directed parameter, namely, magic number 0.98, in the three types of the decision-directed-based estimators from a viewpoint of amounts of musical noise and speech distortion. In addition, we perform comparison between these noise reduction methods based on the mathematical analysis and human perception test.**

I. 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), *Wiener filter* (WF) [1] and Ephraim-Malah's *minimum mean-square error short-time spectral amplitude* (MMSE STSA) estimator [2] are 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.

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 [3]. Using this metric, we have successfully analyzed the amount of musical noise generated via SS [4] and WF [5]. However, any theoretical analysis of the amount of musical noise generated in noise reduction methods with a decision-directed a priori SNR estimator (hereafter it is referred to as *DD approach*), e.g., the MMSE STSA estimator, has not been reported yet. It is worth mentioning that several studies on the systematic analysis of the DD approach have been provided [6], [7], but they did not use the explicit metric of musical noise generation like higher-order statistics so far.

In this paper, we provide a new theoretical analysis of the amount of musical noise generated in the MMSE STSA estimator, the *minimum mean square error log-spectral amplitude* (MMSE LSA) estimator [8] and the *maximum a posteriori* (MAP) estimator [9] with the DD approach based on higher-order statistics. Since the DD approach corresponds to a nonlinear recursive process that requires infinite power spectral sequences, we introduce a computationally efficient higher-order moment calculation method based on generalized Gauss-Laguerre quadrature [10]. We also mathematically clarify the justification of using a typical decision-directed parameter, namely, magic number 0.98, in the three types of the decision-directed-based estimators from a viewpoint of amounts of musical noise and speech distortion. In addition, we perform comparison between these noise reduction methods based on the mathematical analysis and human perception test.

II. Related works

A. Mathematical Metric of Musical Noise Generation via High-Order Statistics [3]

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 larger 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 *m*th-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, although musical noise is not conspicuous 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 the *kurtosis ratio* as a measure to assess musical noise [3]. This measure is defined as

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

where $\mathrm{kurt}_{\mathrm{proc}}$ is the kurtosis of the processed signal and $\mathrm{kurt}_{\mathrm{org}}$ is the kurtosis of the observed signal. This measure increases as the amount of generated musical noise increases. In Ref. [3], it was reported that the kurtosis ratio is strongly correlated with the human perception of musical noise.

B. Analysis of Amount of Noise Reduction [4]

We analyze the amount of noise reduction via processing. Hereafter, we define the *noise reduction rate* (NRR) 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

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

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., $\mathrm{E}[s_{\mathrm{out}}^2] \simeq \mathrm{E}[s_{\mathrm{in}}^2]$, then

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

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.

C. Noise Reduction Methods with DD Approach

In this subsection, we briefly introduce the MMSE STSA estimator, the MMSE LSA estimator and the MAP estimator. 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 the MMSE STSA estimator, the MMSE LSA estimator and the MAP estimator are generally formulated as

$$Y(f,\tau) = G(f,\tau)|X(f,\tau)|\exp(j\arg(X(f,\tau))), \tag{6}$$

where $Y(f,\tau)$ is the enhanced target speech signal, $G(f,\tau)$ is the gain function of each method. The gain function of the MMSE STSA estimator is defined as

$$\begin{aligned} G_{\mathrm{STSA}}(f,\tau) =& \Gamma(1.5)\frac{\nu(f,\tau)}{\gamma(f,\tau)}\exp\Big(-\frac{\nu(f,\tau)}{2}\Big) \\ & \left[(1+\nu(f,\tau))I_0\left(\frac{\nu(f,\tau)}{2}\right) + \nu(f,\tau)I_1\left(\frac{\nu(f,\tau)}{2}\right)\right], \end{aligned} \tag{7}$$

that of the MMSE LSA estimator is defined as

$$G_{\mathrm{LSA}} = \frac{\xi(f,\tau)}{(1+\xi(f,\tau))}\exp\left(\frac{1}{2}\int_{\nu(f,\tau)}^{\infty}\frac{\exp(-t)}{t}dt\right), \tag{8}$$

and that of the MAP estimator is defined as

$$G_{\mathrm{MAP}}(f,\tau) = \frac{\xi(f,\tau) + \sqrt{\xi^2(f,\tau) + (1+\xi(f,\tau))\frac{\xi(f,\tau)}{\gamma(f,\tau)}}}{2(1+\xi(f,\tau))}, \tag{9}$$

where I_0 and I_1 are the modified Bessel functions of zero and first order, and $\nu(f,\tau) = \gamma(f,\tau)\xi(f,\tau)/(1+\xi(f,\tau))$. Also, $\xi(f,\tau)$ and $\gamma(f,\tau)$ are a priori and a posteriori SNRs, which are defined as

$$\xi(f,\tau) = \mathrm{E}[|S(f,\tau)|^2]/\mathrm{E}[|N(f,\tau)|^2], \tag{10}$$

$$\gamma(f,\tau) = |X(f,\tau)|^2/\mathrm{E}[|N(f,\tau)|^2]. \tag{11}$$

In (10) and (11), we can commonly estimate $\mathrm{E}[|N(f,\tau)|^2]$ by averaging the noise power spectra in the speech absent time period, or by using other estimation methods [2], [7]. However, since we cannot estimate $\mathrm{E}[|S(f,\tau)|^2]$ in advance, a priori SNR $\xi(f,\tau)$ is approximately calculated via the following DD approach;

$$\hat{\xi}(f,\tau) = \alpha\gamma(f,\tau-1)G^2(f,\tau-1) + (1-\alpha)F[\gamma(f,\tau)-1], \tag{12}$$

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

The forgetting factor α of around 0.98, namely, $0.97 \sim 0.99$, is mostly used in the past studies. These values are, however, decided experimentally, and there is no theoretical proof for the justification of using such magic number. One of the main purposes in this paper is to justify the optimal forgetting factor via theoretical analysis.

III. Theoretical Analysis of Noise Reduction Methods with DD Approach

A. Higher-Order Moment

In this section, we analyze the MMSE-STSA estimator, the MMSE LSA estimator and the MAP estimator based on the higher-order-statistics analysis. More specifically, we mathematically derive the higher-order moments of output power spectra for each method. As described in Sects. II-A and II-B, the 1st-order moment is used to calculate the NRR (amount of noise reduction), and the 2nd- and 4th-order moments are used for getting the kurtosis ratio (amount of musical noise generation).

Hereafter, random variables x and y represent the power spectra of noise $N(f,\tau)$ and its processed output $Y(f,\tau)$, respectively. Signal processing in the MMSE or the MAP estimator with the DD approach can be interpreted as *stochastic variable transform* from x to y, and more importantly, this is a nonlinear recursive (infinite) process for x because the DD approach requires infinite number of samples of the past $N(f,\tau)$ ($\tau = \tau_0, \tau_{-1}, \ldots, \tau_{-\infty}$) to output the specific $Y(f,\tau_0)$ at the time τ_0. Thus, assuming that X is independent and identically distributed (i.i.d.), we can write the mth-order moment of Y as

$$\begin{aligned}\mu_m &= \int_0^\infty y^m p(y) dy \\ &= \int_0^\infty \int_0^\infty \cdots \int_0^\infty h^m(x_0, x_1, \ldots, x_\infty) \prod_{i=0}^{\infty} p(x_i) \\ &\quad dx_0 dx_1 \cdots dx_\infty, \end{aligned} \tag{13}$$

where the transformation function $y = h(x_0, x_1, \ldots, x_\infty)$ is defined to represent the relation among $|Y(f,\tau_0)|^2$ and $\{|N(f,\tau_0)|^2, \ldots, |N(f,\tau_{-\infty})|^2\}$ in MMSE STSA estimator as calculated via (6), (7) and (12), in the MMSE LSA estimator via (6), (8) and (12), or in the MAP estimator via (6), (9) and (12).

B. Numerical Calculation of Higher-Order Moment

In this subsection, a practical calculation method of the higher-order moments is described. Since the higher-order moment (13) cannot be expressed by an analytical form, we should introduce some approximations and a numerical integration technique for calculating the higher-order moment. First, we truncate the infinite time sequences within the past T samples, which is long enough to maintain the calculation accuracy. Next, instead of the multiple integrals, a numerical integral formula is applied in the range of x_i from 0 to Ω. For example, we can use the Newton-Cotes method that is a commonly used numerical integral formula with equally sampled function values, which is given by

$$\mu_m \simeq \sum_{k_0=1}^{c} \sum_{k_1=1}^{c} \cdots \sum_{k_T=1}^{c} h^m(x_{k_0}, x_{k_1}, \cdots, x_{k_T}) \prod_{i=0}^{T} p(x_{k_i}) w_{k_i}, \tag{14}$$

where c is the number of sample points in the integrand, $k_i = 1, 2, \ldots, c$, $x_{k_i} = (k_i - 1)\Omega/(c-1)$, and w_{k_i} is a weight for the numerical integral.

The biggest problem in this calculation is a huge computational cost. For instance, $T = 14$ and $c = 11$ were required in our preliminary experiment to maintain the calculation accuracy, resulting in $c^{(T+1)}$ multiply and accumulation (MAC) for the weight w_{k_i}. This roughly corresponds to 4 Peta MAC in floating-point arithmetic, and is obviously impossible to be processed by normal computers except supercomputer.

C. Computationally Efficient Calculation Method

To solve the problem due to huge number of MAC, we propose to introduce computationally efficient calculation of higher-order moments based on the parametric p.d.f. model and generalized Gauss-Laguerre quadrature [10]. In the following, we describe the detail of the proposed method and its advantage on the computational cost.

First, the p.d.f. of the noise power spectra is modeled by the gamma distribution as

$$p(x) = x^{\eta-1} \exp\big(-\mathrm{x}/\theta\big)/\theta^\eta/\Gamma(\eta), \tag{15}$$

where $\Gamma(\eta)$ is the gamma function, 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 $\alpha < 1$.

Then, application of the gamma distribution model to the moment calculation (13) enables us to use *generalized Gauss-Laguerre quadrature* [10], which is a computationally efficient numerical integral formula with *unequally* sampled function values. In the case of noise with the shape parameter η (the scale parameter θ can be set to 1 without loss of generality), the mth-order moments are given by

$$\begin{aligned}\mu_m^{[\eta]} &= \int_0^\infty \cdots \int_0^\infty h^m(x_0, \cdots, x_\infty) \prod_{i=0}^{\infty} \frac{x_i^{\eta-1} \exp(-x_i)}{\Gamma(\eta)} dx_0 \cdots dx_\infty \\ &\simeq \frac{1}{\Gamma(\eta)^{T+1}} \sum_{k_0=1}^{c'} \sum_{k_1=1}^{c'} \cdots \sum_{k_T=1}^{c'} h^m(x_{k_0}^{[\eta]}, x_{k_1}^{[\eta]}, \cdots, x_{k_T}^{[\eta]}) \prod_{i=0}^{T} w_{k_i}^{[\eta]}, \end{aligned} \tag{16}$$

where c' is the number of sample points in the integrand and $k_i = 1, 2, \ldots, c'$. The sampling abscissas $x_{k_i}^{[\eta]}$ are defined as the roots of the generalized Gauss-Laguerre polynomial

$$L_{c'}^{[\eta-1]}(x) = \frac{x^{-(\eta-1)} \exp(x)}{c'!} \frac{d^{c'}}{dx^{c'}} \big\{ \exp(-x) x^{c'+\eta-1} \big\}. \tag{17}$$

Also, the weight $w_{k_i}^{[\eta]}$ is calculated as

$$w_{k_i}^{[\eta]} = \Gamma(c'+\eta)(c'! x_{k_i}^{[\eta]})^{-1} \Big[\frac{d}{dx} L_{c'}^{[\eta-1]}(x_{k_i}^{[\eta]}) \Big]^{-2}. \tag{18}$$

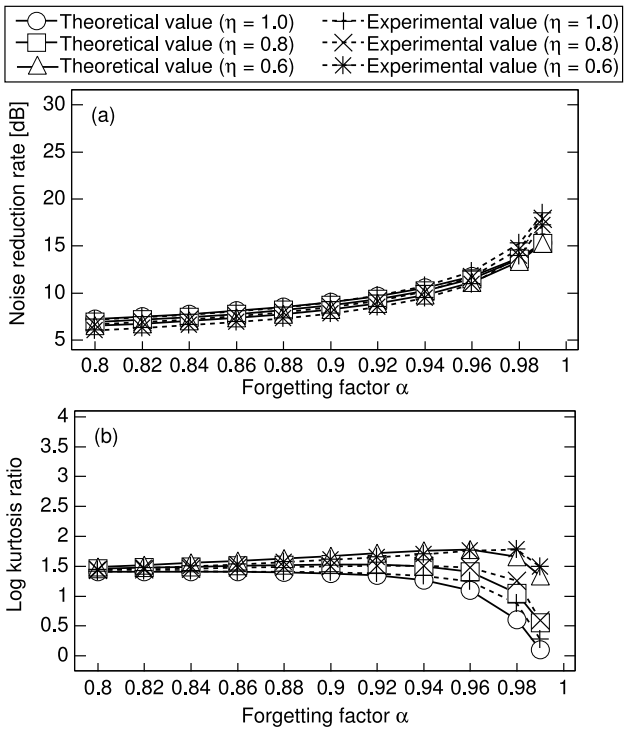

Fig. 1. Theoretical behavior and experimental results for MMSE STSA estimator: (a) noise reduction rate, and (b) log kurtosis ratio.

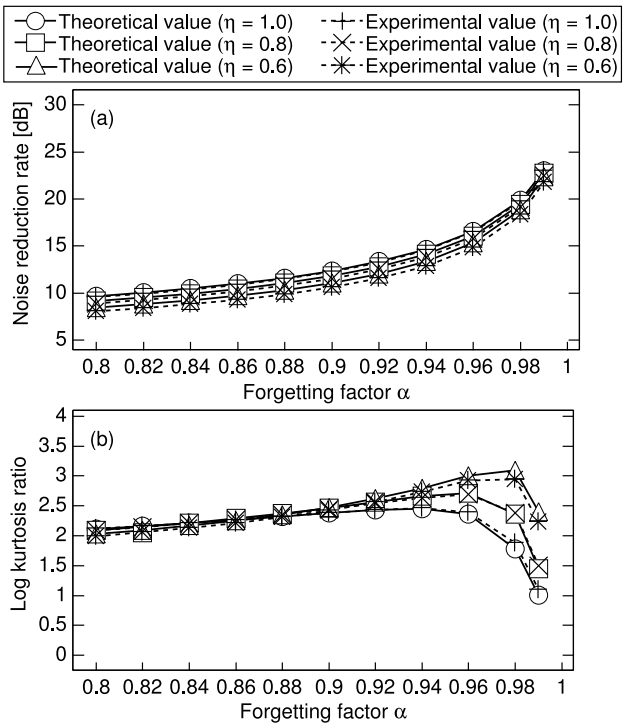

Fig. 2. Theoretical behavior and experimental results for MMSE LSA estimator: (a) noise reduction rate, and (b) log kurtosis ratio.

Hereafter, we discuss the computational cost in the proposed calculation. Generally speaking, a c'-point generalized Gauss-Laguerre quadrature can yield an exact result for polynomials of up to $2c' - 1$ degrees, although a c-point Newton-Cotes method does for polynomials of up to $c-1$ degrees. Thus, we can set $c' = c/2$. This greatly reduces the number of MAC by $(1/2^{T+1})$-fold in (16), compared with (14). For example, a use of $c' = 5$ in (16) results in simple 30 Giga ($\simeq 5^{15}$) MAC, which can be processed by personal computers.

IV. Experiment

A. Experimental Conditions

We calculated the NRR and kurtosis ratio using (16). The shape parameter η of the noise p.d.f. is set to 1.0, 0.8 and 0.6. The number of sampling points in the numerical integral, c' in (16), is set to 5, and the number of frames corresponding to $T+1$ is set to 14 in the MMSE STSA estimator, and set to 15 in the MMSE LSA estimator and the MAP estimator. Since the kurtosis of 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 means 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 and subjective evaluation experiment in order to confirm the validity of our proposed theoretical analysis. The NRR and log kurtosis ratio are calculated from actual noise reduction results obtained by 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 six speakers (6 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.

B. Objective Evaluation

Figs. 1, 2 and 3 are the results of the theoretical calculations using (16) and objective evaluations for each of noise reduction methods. We can confirm that the theoretical values derived by our analysis is well consistent with those of experimental results obtained by actual noise reduction.

Regarding the detailed behavior of the NRR, the MMSE STSA estimator, the MMSE LSA estimator and the MAP estimator show the same tendency in that the NRR becomes higher as the larger forgetting factor α is used (see Figs. 1(a), 2(a) and 3(a)). In contrast, regarding the log kurtosis ratio, it is of great interest that the kurtosis ratio suddenly drops after $\alpha = 0.97$ in the MMSE STSA estimator and the MMSE LSA estimator, and after $\alpha = 0.98$ in the MAP estimator (see Figs. 1(b), 2(b) and 3(b)). From the results, we can speculate that the kurtosis-ratio drop is the key factor of less musical noise property in these three types of estimators, and we will

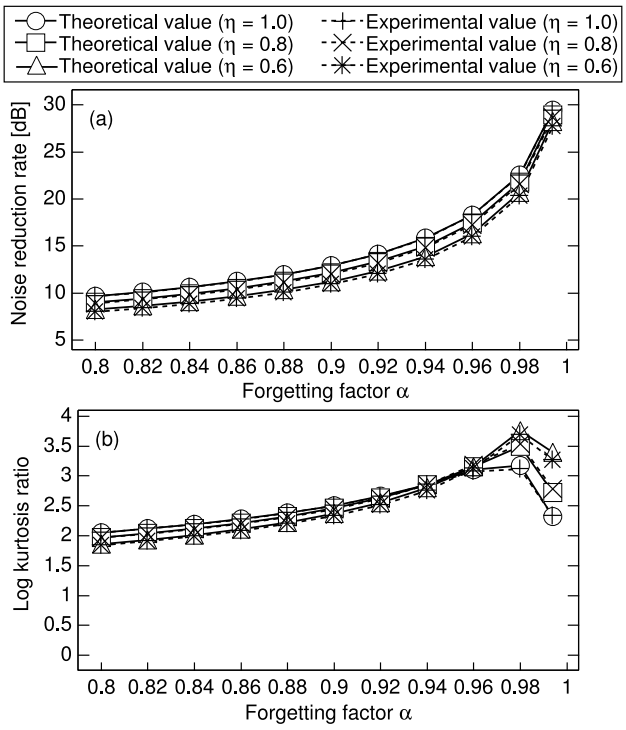

Fig. 3. Theoretical behavior and experimental results for MAP estimator: (a) noise reduction rate, and (b) log kurtosis ratio.

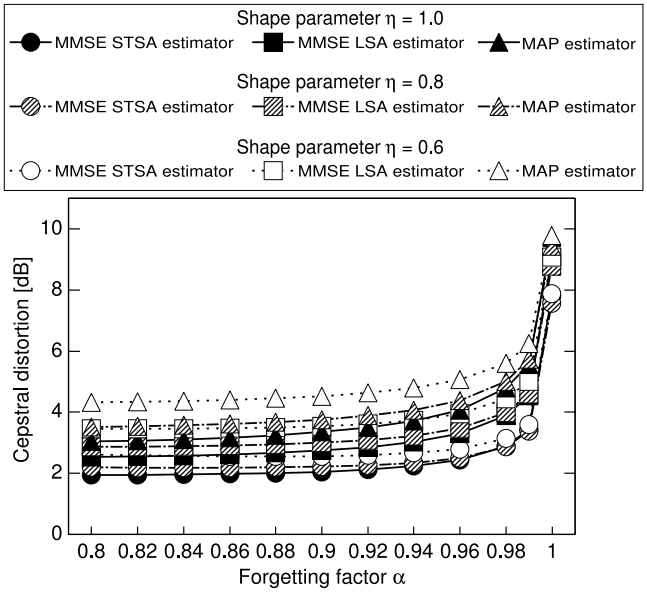

Fig. 4. Cepstral distortion for MMSE STSA estimator, MMSE LSA estimator and MAP estimator.

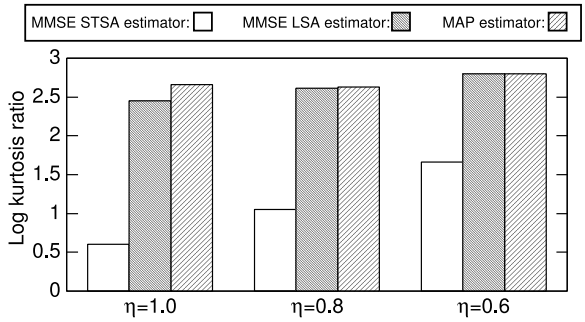

Fig. 5. Relation between a noise reduction methods and theoretical value of log kurtosis ratio under equi-NRR conditions.

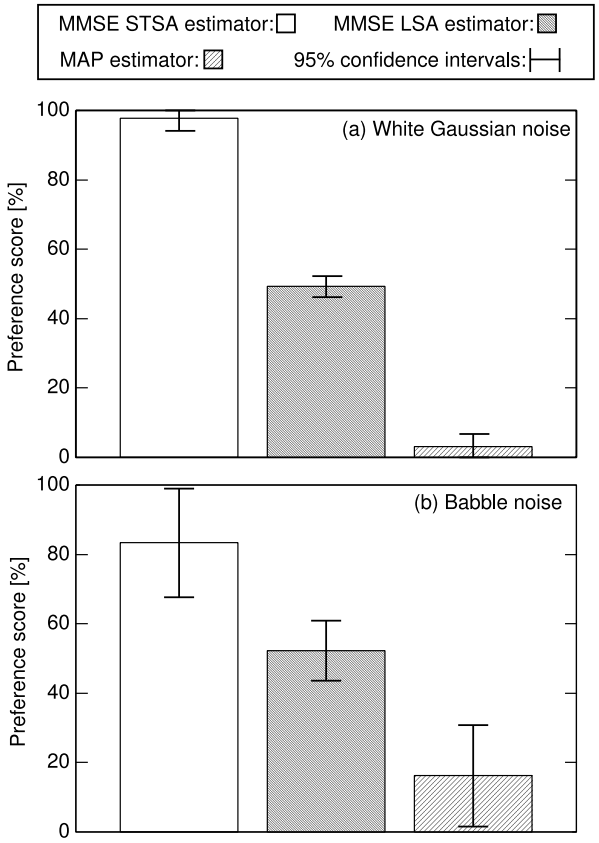

Fig. 6. Subjective evaluation results for (a) white Gaussian noise, and (b) babble noise.

discuss this issue in the next subsection.

C. Discussion on Optimal Forgetting Factor

In this subsection, we discuss the optimality on the forgetting factor for each noise reduction method. As shown in Figs. 1(b), 2(b) and 3(b), the kurtosis ratio falls after $\alpha = 0.97$ or 0.98 while higher NRRs are kept, suggesting that α should be more closer to 1.0 from the viewpoint of less musical noise. This suggestion is, however, misleading because we did not take speech distortion into account. Fig. 4 shows the cepstral distortion for each method, where we can see a monotonic increase in speech distortion along with larger α. In conclusion, our theoretical analysis proves that the forgetting factor α in these noise reduction methods with DD approach have its *sweet spot* placed from 0.97 to less than 1, namely, $0.97 \sim 0.99$, for achieving less musical noise (kurtosis-ratio drop) as well as low speech distortion.

D. Comparison Among Three Methods

In this subsection, we compare the MMSE STSA estimator, the MMSE LSA estimator and the MAP estimator. From Figs. 1(b), 2(b) and 3(b), we can theoretically confirm that musical noise is the least perceptible when a signal is processed by the MMSE STSA estimator, and musical noise is the most perceptible when processed by the MAP estimator. Thus, this result suggests that the MMSE STSA estimator is the best noise reduction method from the viewpoint of less musical noise property. On the other hand, it is suggested that the MMSE STSA estimator achieves the least amount of noise suppression and the MAP estimator achieves the most amount of noise suppression from Figs. 1(a), 2(a), 3(a). Hence, there is a trade-off between amounts of noise suppression and musical noise on which method we used.

Next, we compare the log kurtosis ratio of each noise reduction method under equi-NRR conditions. We use theoretical values of the log kurtosis ratio with the same NRR; this value was taken when we use the forgetting factor $\alpha = 0.98$ in the MMSE STSA estimator in Figs. 1, 2 and 3. In other words, three methods take different values of the forgetting factor α, but NRR is the same. The result is shown in Fig. 5. From Fig. 5, we can confirm that musical noise is the least perceptible when processed by the MMSE STSA estimator, and more perceptible when the MAP estimator is used for white Gaussian noise. Further more, it is found that when the noise shape parameter η is smaller, the values of the log kurtosis ratio in all methods become close. This predicts that we hardly discriminate the difference of the amount of musical noise among three methods when noise is super-Gaussian.

E. Subjective Evaluation

We next conducted a subjective evaluation. In this subjective evaluation, we presented equi-NRR (15 dB for white Gaussian noise, 13 dB for babble noise) signals processed by the MMSE STSA estimator, the MMSE LSA estimator and the MAP estimator in random order to 11 examinees, who selected which signal they considered to contain the least musical noise. The result of the experiment is shown in Fig. 6. It is found that musical noise is the least perceptible when processed by the MMSE STSA estimator, and musical noise is the most perceptible when the MAP estimator is used. In particular, this tendency is remarkable for white Gaussian noise but slightly ambiguous for babble noise. Actually, the shape parameter of babble noise is smaller than that of white Gaussian noise. Thus, this tendency is well consistent with our prediction based on the proposed theoretical analysis as described in Sect. IV-D, confirming the validity of the proposed method for theoretical analysis.

V. Conclusion

In this study, we performed a theoretical analysis of the amount of musical noise generated in the MMSE STSA estimator, the MMSE LSA estimator and the MAP estimator based on higher-order statistics. Particularly, we calculated higher-order statistics with computationally low cost and high accuracy by using generalized Gauss-Laguerre quadrature. From mathematical analysis and evaluation experiments, we clarified the justification of using the magic number 0.98 in the MMSE STSA estimator from a viewpoint of amounts of musical noise and speech distortion. In addition, we compared three methods with DD approach and we clarified that there is the trade-off of amounts of noise suppression and musical noise on which method we used. Finally, we conducted the subjective evaluation and confirmed that the MMSE STSA estimator is the most preferred method from view point of less musical noise property.

Acknowledgment

This work was partly supported by JST Core Research of Evolutional Science and Technology (CREST), Japan.

References

[1] P. C. Loizou, *Speech enhancement theory and practice*, CRC Press, Taylor & Francis Group FL, 2007.

[2] 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.

[3] 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.

[4] 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.

[5] T. Inoue, H. Saruwatari, K. Shikano, K. Kondo, "Theoretical analysis of musical noise in wiener filter via higher-order statistics," *Proc. of APSIPA2010*, 2010.

[6] O. Cappé, "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.

[7] 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.

[8] 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.

[9] T. Lotter, P. Vary, "Speech enhancement by maximum a posteriori spectral amplitude estimation using a supergaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol.2005, no.7, pp.1110–1126, 2005.

[10] C. R. Cassity, and H. R. Hopper, "Abscissas and coefficients for the generalized Gauss-Laguerre quadrature formula using the zero ordinate," *NASA Technical Memorandum*, no.TMX-53099, 1964.