Theoretical analysis of musical noise in nonlinear noise reduction based on higher-order statistics

Yu Takahashi*, Ryoichi Miyazaki†, Hiroshi Saruwatari† and Kazunobu Kondo*
* Corporate Research & Development Center, Yamaha Corporation, Hamamatsu, Japan
† Nara Institute of Science and Technology, Ikoma, Japan

***Abstract*—In this paper, we review a musical-noise-generation analysis of nonlinear noise reduction techniques with using higher-order statistics (HOS). Recently, an objective metric based on HOS to analyze nonlinear artifacts, i.e., musical noise, caused by nonlinear noise reduction techniques has been proposed. Such metric enables us to perform objective comparison of any nonlinear methods from the perspective of the amount of musical noise generated. Furthermore, such metric enables us to control the musical noise generated by nonlinear noise reduction techniques. In the paper, first, the mathematical principle of the analysis for the amount of musical noise based on HOS is described, and analyses and comparison examples of typical nonlinear noise reduction techniques are demonstrated. Next, it is clarified that to find a fixed point in HOS leads to no-musical noise property in noise reduction. Finally, several expansions on the theory are discussed.**

I. Introduction

Many applications of speech communication systems, such as hearing aids, mobile phones, and teleconference systems, have been investigated in recent years. It is, however, well known that these systems are always suffer from noise condition. Since noise causes a serious problem of speech quality, thus noise reduction is an essential technique to achieve high quality speech communication systems.

Various methods have been presented for noise reduction techniques and they can be generally classified into two groups; methods based on single-channel input [1]–[6], and those based on multichannel input, e.g., microphone array signal processing [7]. Moreover, methods integrating microphone array signal processing and nonlinear signal processing have been actively researched in recently years, e.g., [8], [9]. Above all, we focus our attention on nonlinear single-channel noise reduction techniques in this paper.

Spectral subtraction (SS) [1]–[3], Wiener filtering [4], [5], and minimum mean-square error short-time spectral amplitude estimator (MMSE-STSA) [6] are commonly used nonlinear single-channel noise reduction techniques. Actually these methods are powerful noise reduction techniques, but these methods often cause nonlinear artifacts, so-called *musical noise*.

Recently, it was reported that the amount of generated musical noise is strongly related to the difference between higher-order statistics (HOS) before and after nonlinear signal processing [10]. Based on this fact, the authors have proposed a HOS-based objective metric for the amount of musical noise generated [10]. This HOS-based objective metric enables us to analyze/optimize nonlinear signal processing from the viewpoint of musical-noise generation by mathematical manner. Actually the HOS-based analysis has been applied to nonlinear signal processing. For instance, generalized spectral subtraction (GSS) has been analyzed on the basis of the measure in Ref. [11], and a parameter to reduce the amount of musical noise generated was clarified as a result of the analysis. Also the analysis of Wiener filtering family was performed in Ref. [12]. These analyses provided a new fact that commonly-used parameters are not appropriate and there exists more appropriate parameters for less amount of musical-noise generation. Interestingly, it was also revealed that output signal of SS with an optimized parameter contains less amount of musical-noise than that of Winer filtering [12]. The validity of these results were also confirmed by subjective evaluations as well as mathematical analyses. Furthermore, in Refs. [13], [14], one of the author have proposed SS with a special parameter, which does not cause any musical noise. This method was established by analyzing the change of HOS through SS.

As we described, the HOS-based analysis makes it possible to compare nonlinear noise reduction techniques from the viewpoint of the amount of musical-noise generation by *objective manner*. Moreover, the HOS-based analysis allows us to control the amount of musical noise generated by nonlinear signal processing. In this paper, we show the mathematical manner to analyze the amount of musical noise generated by typical nonlinear signal processing on the basis of HOS, and demonstrate a *musical-noise-free* noise reduction method that do not yield any musical noise, as an application of the analysis.

The rest of the paper is organized as follows. In Sect. II, the metric based on HOS used for the amount of musical noise generated is described. Following the section, we denote analysis examples based on HOS in Sect. III, and we give comparison results based on the results of the analyses in Sect. IV. In Sect. V, we demonstrate the musical-noise-free nonlinear signal processing as an application of HOS-based analysis. Finally we give our conclusion in Sect. VI.

II. Objective metric for musical noise generated

A. Overview

An objective metric is indispensable for us to perform objective comparison of noise reduction techniques. Moreover, it is desirable that the metric can be derived by mathematically-closed form. Actually, various kinds of objective metric for

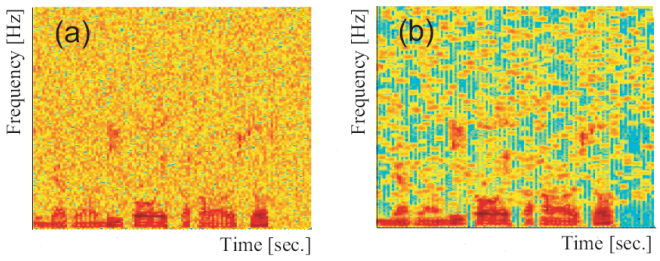

Fig. 1. (a) Observed spectrogram, and (b) processed spectrogram

noise reduction techniques have been proposed, for instance, signal-to-noise ratio (SNR) and cepstral distortion (CD) [15] are widely-used metrics. These metrics are clearly objective and mathematically-closed-form metrics. Generally, SNR only considers power of noise and source signal, and CD considers speech distortion. Then, the amount of musical noise generated cannot be measured by these typical metrics. Therefore, an objective metric designed for the amount of musical noise generated is needed. In this section, we review the HOS-based objective metric for the amount of musical noise generated based on HOS proposed by the authors.

B. Objective metric for musical noise generated based on higher-order statistics

Generally, nonlinear noise reduction techniques reduce noise drastically but often provide musical noise at the same time. This musical noise can be considered as the audible isolated spectral components generated through such nonlinear signal processing. Fig. 1(b) shows an example of a spectrogram of musical noise in which many isolated components can be observed. Then, it can be speculated that the amount of musical noise is strongly related to the number of such isolated components and their level of isolation. Hence, Uemura et al. have introduced kurtosis, i.e., 4th-order statistics, to quantify the isolated spectral components, and they focus their attention on the changes in kurtosis [10]. Since isolated spectral components are dominant, they are heard as tonal sounds, which results in our perception of musical noise. Therefore, it is expected that obtaining the number of tonal components will enable us to quantify the amount of musical noise. However, such a measurement is extremely complicated, so instead they have introduced a simple statistical estimate, i.e., kurtosis. This strategy allows us to obtain the characteristics of tonal components. The adopted kurtosis can be used to evaluate the width of the probability density function (p.d.f.) and the weight of its tails, i.e., kurtosis can be used to evaluate the percentage of tonal components among the total components. A larger value indicates a signal with a heavy tail in its p.d.f., meaning that it has a large number of tonal components. Also, kurtosis has the advantageous property that it can be easily calculated in a concise algebraic form.

C. Kurtosis

Kurtosis is one of the most commonly used HOS for the assessment of non-Gaussianity. Kurtosis is defined as

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

where x is a random variable, kurt_x is the kurtosis of x, and μ_n is the nth-order moment of x. Here μ_n is defined as

$$\mu_n = \int_{-\infty}^{+\infty} x^n P(x) \mathrm{d}x, \tag{2}$$

where $P(x)$ denotes the p.d.f. of x. Note that this μ_n is not a central moment but *a raw moment*. Thus, (1) is not kurtosis according to the mathematically strict definition, but a modified version; however, we refer to (1) as kurtosis in this study.

D. Kurtosis ratio [10]

Although we can measure the number of tonal components by kurtosis, it is worth mentioning that kurtosis itself is not sufficient to measure musical noise. This is because that the kurtosis of some unprocessed signals such as speech signals is also high, but we do not perceive speech as musical noise. Since we aim to count only the musical-noise components, we should not consider genuine tonal components. To achieve this aim, we should focus on the fact that musical noise is generated only in artificial signal processing. Hence, we should consider the change in kurtosis during signal processing. Consequently, the following *kurtosis ratio* [10] has been proposed to measure the kurtosis change:

$$\text{kurtosis ratio} = \frac{\mathrm{kurt}_{\mathrm{proc}}}{\mathrm{kurt}_{\mathrm{input}}}, \tag{3}$$

where $\mathrm{kurt}_{\mathrm{proc}}$ is the kurtosis of the processed signal and $\mathrm{kurt}_{\mathrm{input}}$ is the kurtosis of the input signal. A larger kurtosis ratio ($\gg 1$) indicates a marked increase in kurtosis as a result of processing, implying that a larger amount of musical noise is generated. On the other hand, a smaller kurtosis ratio ($\simeq 1$) implies that less musical noise is generated. It has been confirmed that this kurtosis ratio closely matches the amount of musical noise in a subjective evaluation based on human hearing [10].

III. Theoretical Analysis Examples Based on Higher-Order Statistics

A. Overview

In this section, we give the way to analyze kurtosis ratio after nonlinear signal processing thorough analysis examples of typical nonlinear noise reduction techniques. Particularly, our analyses include generalized spectral subtraction (GSS) and Wiener filtering family. GSS is an expansion of SS, and parametrized by an exponent parameter [3]. GSS involves the standard power- and amplitude-domain SS. Comparison results of the analyzed methods based on kurtosis ratio will be demonstrated in Sect. IV.

B. Signal model

We introduce a gamma distribution to model time-frequency power-domain signal [16], [17]. The p.d.f. of the gamma distribution $P_{\mathrm{GM}}(x)$ for random variable x is defined by

$$P_{\mathrm{GM}}(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \cdot x^{\alpha-1} \exp\left\{-\frac{x}{\theta}\right\}, \tag{4}$$

where $x \geq 0$, $\alpha > 0$, and $\theta > 0$. Here, α is the shape parameter, θ is the scale parameter, and $\Gamma(\cdot)$ is the gamma function. The gamma distribution with $\alpha = 1$ corresponds to the chi-square distribution with 2 degrees of freedom. Moreover, it is well known that the mean of x for a gamma distribution is $E[x] = \alpha\theta$.

In the following, we mathematically analyze how the distribution of input noise is deformed via nonlinear signal processing. To describe the change of the distribution, we formulate the mth-order moment of the p.d.f. deformed after the signal processing. Based on this mth-order moment, kurtosis ratio and noise reduction rate (NRR) are derived. NRR and SNR are very similar but NRR indicates SNR improvement.

C. Analysis Example 1: Generalized spectral subtraction

First of all, short-time analysis of input signal is conducted by frame-by-frame discrete Fourier transform. As a result, we obtain time-frequency domain observation $X(f,\tau)$ where f is frequency bin, and τ is time frame index.

GSS that is a generalized form of SS can be formulated as [3]

$$\hat{S}_{\mathrm{GSS}}(f,\tau) = \begin{cases} \sqrt[2n]{|X(f,\tau)|^{2n} - \beta \cdot \mathrm{E}_{\tau}[|\hat{N}(f,\tau)|^{2n}]}\, e^{j\arg(X(f,\tau))} \\ \quad (\text{where} \quad |X(f,\tau)|^{2n} - \beta \cdot \mathrm{E}_{\tau}[|\hat{N}(f,\tau)|^{2n}] > 0), \\ \rho X(f,\tau) \quad (\text{otherwise}) \end{cases} \tag{5}$$

where $\hat{S}_{\mathrm{GSS}}(f,\tau)$ is a recovered speech signal, and $\hat{N}(x,\tau)$ is an estimated noise signal. Besides $\mathrm{E}_{\tau}[\cdot]$ expresses time-averaging operator. β, ρ, and n are parameters of GSS and they are oversubtraction parameter, flooring parameter, and exponent parameter, respectively. GSS with $n = 1$ corresponds to power-domain SS and GSS with $n = 0.5$ corresponds to amplitude-domain SS. As shown in Fig 2, the p.d.f. of the observation modeled by the gamma distribution is deformed via GSS. To calculate kurtosis ratio, the 4th- and the 2nd-order moment of the deformed p.d.f. is needed. The mth-order moment of the p.d.f. after performing GSS by (5) can be derived as [11]

$$\mu_m = \theta^m \mathcal{M}_{\mathrm{GSS}}(\alpha, \beta, m/n, \rho) \tag{6}$$

where

$$\begin{aligned} &\mathcal{M}_{\mathrm{GSS}}(\alpha, \beta, m/n, \rho) \\ &= \frac{1}{\Gamma(\alpha)} \sum_{l=0}^{m/n} \left[\left\{ -\beta \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \right\}^{l} \frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)} \right. \\ &\quad \left. \cdot \Gamma(\alpha + m - ln, (\beta\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}) \right] \\ &\quad + \frac{\rho^{2m}}{\Gamma(\alpha)} \gamma(\alpha+m, \beta\alpha). \end{aligned} \tag{7}$$

Here, $\gamma(\alpha, z)$ and $\Gamma(\alpha, z)$ are lower and upper incomplete gamma function, respectively. They are defined as

$$\gamma(\alpha, z) = \int_0^z t^{\alpha-1} \exp(-t) dt, \tag{8}$$

$$\Gamma(\alpha, z) = \int_z^{\infty} t^{\alpha-1} \exp(-t) dt. \tag{9}$$

From the derived the mth-order moment, kurtosis ratio between original signal and signal after GSS is designated as

$$\mathrm{KR}_{\mathrm{GSS}} = \frac{\mathcal{M}_{\mathrm{GSS}}(\alpha, \beta, 4/n, \rho)/\mathcal{M}^2_{\mathrm{GSS}}(\alpha, \beta, 2/n, \rho)}{\mathcal{M}_{\mathrm{GSS}}(\alpha, 0, 4/n, \rho)/\mathcal{M}^2_{\mathrm{GSS}}(\alpha, 0, 2/n, \rho)}. \tag{10}$$

Here, note that $\mathcal{M}_{\mathrm{GSS}}(\alpha, 0, 4/n, \rho)$ means the mth-order moment of the original observation because oversubtraction parameter is 0.

Finally, we derive the NRR of GSS. As we mentioned, NRR indicates SNR improvement, which is defined by

$$\mathrm{NRR}\ [\mathrm{dB}] = 10 \log_{10} \frac{\mathrm{E}[s^2_{\mathrm{out}}]/\mathrm{E}[n^2_{\mathrm{out}}]}{\mathrm{E}[s^2_{\mathrm{in}}]/\mathrm{E}[n^2_{\mathrm{in}}]}, \tag{11}$$

where s_{in} and s_{out} are input and processed target signal components, respectively. Besides, n_{in} and n_{out} are input and processed noise signals, respectively. In (11), the denominator corresponds to input SNR and the numerator is the output SNR. If we assume that the amount of noise reduction is much larger than that of speech distortion in nonlinear noise reduction techniques, i.e., $\mathrm{E}[s^2_{\mathrm{out}}] \simeq \mathrm{E}[s^2_{\mathrm{in}}]$, then

$$\mathrm{NRR}\ [\mathrm{dB}] \simeq 10 \log_{10} \frac{\mathrm{E}[n^2_{\mathrm{in}}]}{\mathrm{E}[n^2_{\mathrm{out}}]}. \tag{12}$$

Therefore, NRR can be approximated by using 1st-order moment. This can be written as

$$\mathrm{NRR}_{\mathrm{GSS}} = 10 \log_{10} \frac{\mathcal{M}_{\mathrm{GSS}}(\alpha, 0, 1/n, \rho)}{\mathcal{M}_{\mathrm{GSS}}(\alpha, \beta, 1/n, \rho)}. \tag{13}$$

D. Analysis Example 2: Standard Wiener filtering

In the following, we denote analyses of Wiener filtering family. Generally, Wiener filtering is defined under assumption that a target signal is stationary, as

$$\hat{S}(f,\tau) = G|X(f,\tau)|e^{j\arg(X(f,\tau))}, \tag{14}$$

where $\hat{S}(f,\tau)$ is an estimated target signal, and G is a spectral gain formulated by [4], [5]

$$G = \frac{P_{ss}}{P_{ss} + P_{nn}} = \frac{P_{xx} - P_{nn}}{P_{xx}}. \tag{15}$$

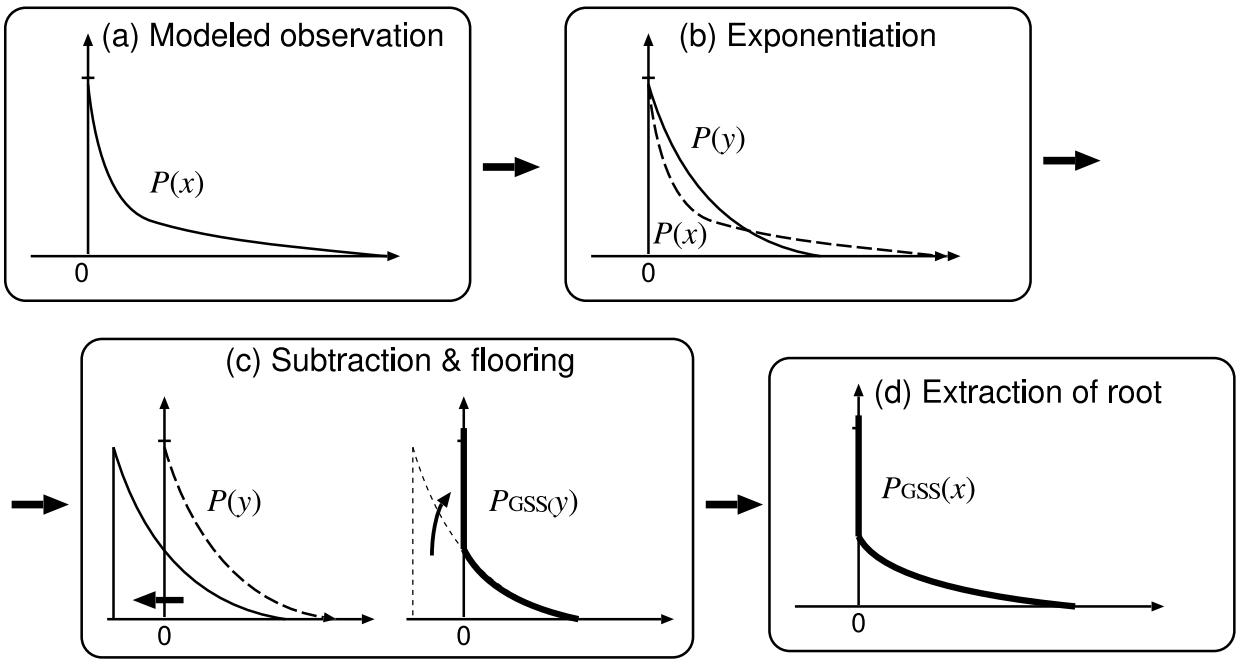

Fig. 2. Deformation of p.d.f. via GSS in the case that flooring parameter $\rho = 0$.

Here P_{ss}, P_{nn}, P_{xx} are power spectral density of target, noise, and observed signal, respectively.

As for an actual speech enhancement problem, an instantaneous observation is utilized to consider nonstationarity of a target speech signal. The spectral gain of this type of Wiener filtering is formulated as

$$G(f,\tau) = \begin{cases} \left(|X(f,\tau)|^2 - \beta \cdot \mathrm{E}_\tau[|\hat{N}(f,\tau)|^2]\right) / |X(f,\tau)|^2 \\ \qquad (|X(f,\tau)|^2 - \beta \cdot \mathrm{E}_\tau[|\hat{N}(f,\tau)|^2] > 0), \\ 0 \quad \text{(otherwise)} \end{cases}, \tag{16}$$

where β is a parameter to control the amount of noise reduction. We refer to this type of Wiener filtering as *standard WF*. Anyway, there exists an alternative approach *decision-directed a priori SNR estimator* [4], [5] to estimate a priori SNR P_{ss}/P_{nn}. In the approach, a priori SNR P_{ss}/P_{nn} is estimated by using an estimated target speech in the previous frame. Although we do not treat this decision-directed a priori SNR estimator in this paper, but Refs. [18], [19] have analyzed this type of Wiener filtering.

The mth-order moment of the p.d.f. after performing the standard WF can be represented by [12]

$$\mu_m^{(\mathrm{SWF})} = \theta^m \mathcal{M}_{\mathrm{SWF}}(\alpha, \beta, m), \tag{17}$$

where

$$\mathcal{M}_{\mathrm{SWF}}(\alpha, \beta, m) = \frac{1}{\Gamma(\alpha)} \int_{\beta\alpha}^{\infty} (t - \beta\alpha)^{2m} t^{\alpha-m-1} \exp(-t) dt. \tag{18}$$

Based on this mth-order moment, kurtosis ratio after applying the standard WF to observation can be written by

$$\mathrm{KR}_{\mathrm{SWF}} = \frac{\mathcal{M}_{\mathrm{SWF}}(\alpha, \beta, 4) / \mathcal{M}^2_{\mathrm{SWF}}(\alpha, \beta, 2)}{\mathcal{M}_{\mathrm{SWF}}(\alpha, 0, 4) / \mathcal{M}^2_{\mathrm{SWF}}(\alpha, 0, 2)}. \tag{19}$$

Also NRR of the standard WF can be formulated by using 1st-order moment, as

$$\mathrm{NRR}_{\mathrm{SWF}} = 10\log_{10} \frac{\mathcal{M}_{\mathrm{SWF}}(\alpha, 0, 1)}{\mathcal{M}_{\mathrm{SWF}}(\alpha, \beta, 1)}. \tag{20}$$

E. Analysis Example 3: Square-root Wiener filtering

Unlike the standard WF mentioned in the previous subsection, there exists an different kind of Wiener filtering defined as

$$G(f,\tau) = \begin{cases} \left(|X(f,\tau)|^2 - \beta \cdot \mathrm{E}_\tau[|\hat{N}(f,\tau)|^2]\right)^{\frac{1}{2}} / |X(f,\tau)| \\ \qquad (|X(f,\tau)|^2 - \beta \cdot \mathrm{E}_\tau[|\hat{N}(f,\tau)|^2] > 0), \\ 0. \quad \text{(otherwise)} \end{cases}. \tag{21}$$

We refer to this type of Wiener filtering as *square-root WF*. This is very similar to the standard WF, but the difference is taking square root to determine its spectral gain.

Similarly in the previous subsection, the mth-order moment of the p.d.f. after performing the square-root WF can be given by

$$\mu_m^{(\mathrm{SRWF})} = \theta^m \mathcal{M}_{\mathrm{SRWF}}(\alpha, \beta, m), \tag{22}$$

where

$$\mathcal{M}_{\mathrm{SRWF}}(\alpha, \beta, m) = \frac{1}{\Gamma(\alpha)} \sum_{l=0}^{m} (-\beta\alpha)^l \frac{\Gamma(m+1)\Gamma(\alpha+m-l, \beta\alpha)}{\Gamma(l+1)\Gamma(m-l+1)}. \tag{23}$$

Then, kurtosis ratio of the square-root WF is

$$\mathrm{KR}_{\mathrm{SRWF}} = \frac{\mathcal{M}_{\mathrm{SRWF}}(\alpha,\beta,4)/\mathcal{M}^2_{\mathrm{SRWF}}(\alpha,\beta,2)}{\mathcal{M}_{\mathrm{SRWF}}(\alpha,0,4)/\mathcal{M}^2_{\mathrm{SRWF}}(\alpha,0,2)}, \tag{24}$$

and we can derive NRR of square-root WF by

$$\mathrm{NRR}_{\mathrm{SRWF}} = 10\log_{10}\frac{\mathcal{M}_{\mathrm{SRWF}}(\alpha,0,1)}{\mathcal{M}_{\mathrm{SRWF}}(\alpha,\beta,1)}. \tag{25}$$

F. Analysis Example 4: Quasi-parametric Wiener filtering

There exists one more alternative type of Wiener filtering, i.e., *quasi-parametric WF*. Generally, it is difficult to know a priori SNR in (15). Therefore the quasi-parametric WF uses a posteriori SNR $|X(f,\tau)|^2/P_{nn}$ instead of a priori SNR [20]. The spectral gain of this type of Wiener filtering can be designated as

$$G(f,\tau) = \frac{|X(f,\tau)|^2}{|X(f,\tau)|^2 + P_{nn}}. \tag{26}$$

Moreover, the generalized form of this type of Wiener filtering is defined by [20]

$$G(f,\tau) = \left(\frac{|X(f,\tau)|^\xi}{|X(f,\tau)|^\xi + \beta \mathrm{E}_\tau[|\hat{N}(f,\tau)|^\xi]}\right)^\eta, \tag{27}$$

where ξ is an exponent parameter for signal, and η is an exponent parameter for gain.

The mth-order moment for the quasi-parametric WF is formulated by [12]

$$\mu_m^{(\mathrm{QPWF})} = \theta^m \mathcal{M}_{\mathrm{QPWF}}(\alpha,\beta,m,\eta), \tag{28}$$

where

$$\mathcal{M}_{\mathrm{QPWF}}(\alpha,\beta,m,\xi,\eta) = \frac{1}{\Gamma(\alpha)}\int_0^\infty \frac{t^{(\xi\eta+1)m+\alpha-1}}{\left\{t^{\frac{\xi}{2}} + \beta\frac{\Gamma(\alpha+\frac{\xi}{2})}{\Gamma(\alpha)}\right\}^{2m\eta}}\exp(-t)dt \tag{29}$$

This leads to the kurtosis ratio of the quasi-parametric WF, which can be designated as [12]

$$\mathrm{KR}_{\mathrm{QPWF}} = \frac{\mathcal{M}_{\mathrm{QPWF}}(\alpha,\beta,4,\xi,\eta)/\mathcal{M}^2_{\mathrm{QPWF}}(\alpha,\beta,2,\xi,\eta)}{\mathcal{M}_{\mathrm{QPWF}}(\alpha,0,4,\xi,\eta)/\mathcal{M}^2_{\mathrm{QPWF}}(\alpha,0,2,\xi,\eta)}. \tag{30}$$

Besides, NRR of the quasi-parametric WF can be formulated as [12]

$$\mathrm{NRR}_{\mathrm{QPWF}} = 10\log_{10}\frac{\mathcal{M}_{\mathrm{QPWF}}(\alpha,0,1,\xi,\eta)}{\mathcal{M}_{\mathrm{QPWF}}(\alpha,\beta,1,\xi,\eta)}. \tag{31}$$

Unfortunately we cannot give no detailed derivation of these kurtosis ratio and NRR of GSS and Wiener filtering family due to the limitation of the paper space, but Refs. [11], [12] would help you to understand detailed derivation.

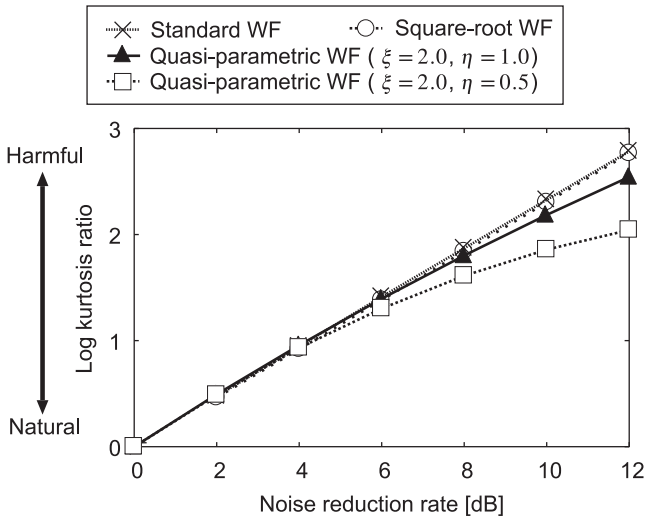

Fig. 3. Theoretical behavior of NRR and log kurtosis ratio for standard WF, square-root WF, and quasi-parametric WF.

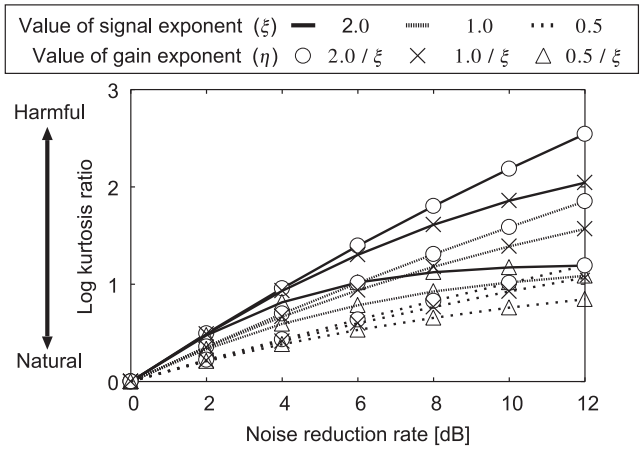

Fig. 4. Theoretical behavior of NRR and log kurtosis ratio for various exponent parameters in quasi-parametric WF.

IV. Comparison based on higher-order statistics

A. Comparison 1: Wiener filtering family

Hereinafter, we demonstrate some comparison results based on the derived kurtosis ratio and NRR in the previous section. Also we show results of objective and subjective evaluations in addition to the theoretical comparison results.

In this subsection, we compare Wiener filtering family from the perspective of the amount of musical noise generated. The analyses in the previous sections make us possible to compare the amount of musical noise generated among Wiener filtering family under the same amount of noise reduction.

Figures 3 and 4 depict the theoretical behavior of the kurtosis ratio and NRR of Wiener filtering family with various parameter values [12]. In these figures, the shape parameter α corresponding to noise type was set to 1.0, and the processing strength parameter β was adjusted so that the target speech NRR is achieved. The target NRR was configured from 0.0 dB to 12.0 dB. Note that we utilized logarithmic kurtosis ratio

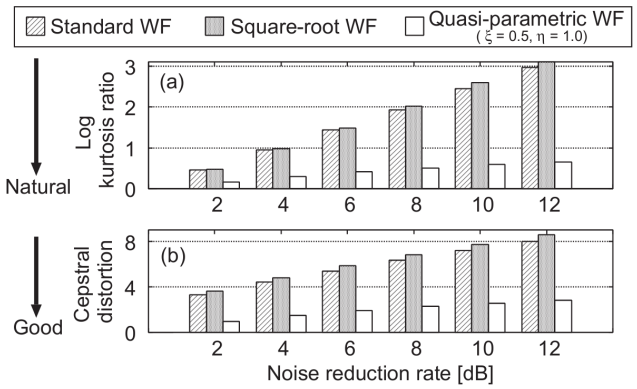

Fig. 5. Objective evaluation results: (a) Log kurtosis ratio, and (b) cepstral distortion of processed signals.

in the figures because the kurtosis exponentially increases with β [10]. We call this *log kurtosis ratio* hereinafter. As for the quasi-parametric WF, the signal exponent parameter ξ was set to 2.0, 1.0, and 0.5, and the gain exponent parameter η was set to $2.0/\xi$, $1.0/\xi$, and $0.5/\xi$.

From Fig. 3, we can see that large amount of musical noise is generated when we use standard WF and square-root WF. However, it also shows that a smaller amount of musical noise is generated when we use quasi-parametric WF with a lower gain exponent parameter. Figure 4 shows that a small amount of musical noise is generated when either of the exponent parameters, ξ or η, is set to a lower value. Consequently, we can achieve high sound quality upon setting lower exponent parameters in the quasi-parametric WF.

We also conducted objective and subjective evaluations to confirm the validity of the theoretical comparison. In the evaluation experiments, observed signals were generated by adding noise signal to target clean speech signals with a SNR of 0 dB. The target speech signals were utterances of 4 speakers (4 sentences), and the noise signal was white Gaussian noise. The length of the each signal was 7 s, and each signal was sampled at 16 kHz. FFT size was 1024, and the frame shift length was 256. In the experiment, we assumed that noise prototype, i.e., $|\hat{N}(f,\tau)|$, was perfectly estimated.

Figure 5 illustrates the log kurtosis ratio and cepstral distortion [12]. These values were calculated from the observed and processed signal by standard WF, square-root WF, and quasi-parametric WF with $(\xi,\eta) = (0.5, 1.0)$. It can be confirmed that the result of the log kurtosis ratio is almost consistent with the theoretical behavior, and cepstral distortion is reduced when quasi-parametric WF is used.

In the subjective evaluation, we presented 3 equi-SNR signals processed by the standard WF, the square-root WF, and the quasi-parametric WF in random order to 10 subjects, who selected which signal they considered to contain least musical noise. The result is shown in Fig. 6 [12]. It can be found that musical noise is less perceptible when the quasi-parametric WF with lower exponent parameter is utilized. This result is also consistent with the theoretical comparison result.

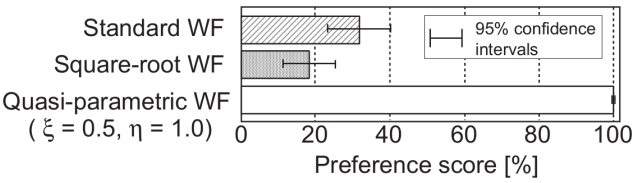

Fig. 6. Subjective evaluation result of various types of Wiener filtering family

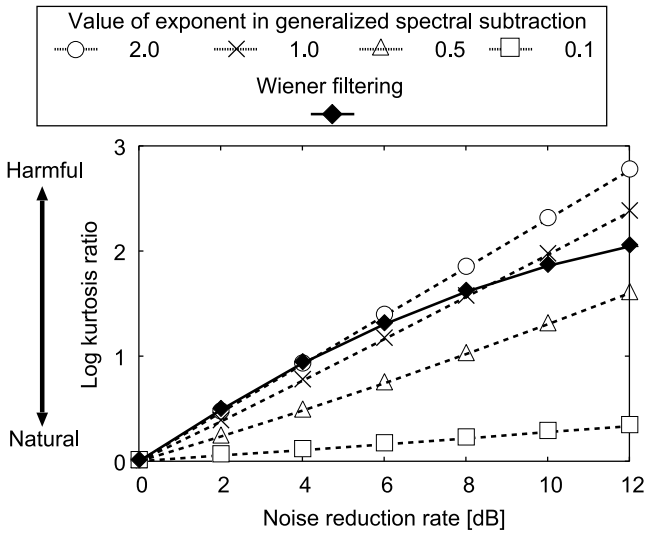

Fig. 7. Theoretical behavior of NRR and log kurtosis ratio in various exponent parameters in GSS and quasi-parametric WF for Gaussian noise ($\alpha = 1.0$)

B. Comparison 2: GSS vs. quasi-parametric WF

It is revealed that quasi-parametric WF is preferable from the viewpoint of sound quality from experiments in the previous subsection. In this subsection, we show the comparison result of GSS analyzed in Sect. III-C and quasi-parametric WF analyzed in Sect. III-F.

Figure 7 shows the theoretical behavior of GSS and quasi-parametric WF under same noise reduction performance [11]. Here parameters for quasi-parametric WF were $(\xi,\eta) = (2, 0.5)$, and exponent domain for GSS was selected from $2n = 2.0, 1.0, 0.5$, or 0.1. The oversubtraction parameter β for GSS was adjusted so that the target NRR is achieved as same as the simulation in the previous subsection.

From the result, the power- or amplitude-domain SS causes a larger amount of musical noise than that by quasi-parametric WF. On the other hand, GSS in lower exponent domain generates less amount of musical noise than that by quasi-parameter WF. This implies that GSS with an appropriate configuration achieves preferable noise reduction rather than Wiener filtering from the viewpoint of the amount of musical noise generated. The validity of the result is also confirmed by a subjective evaluation. The result of the subjective evaluation is shown in Fig. 8 [11]. In the subjective evaluation, we presented 4 equi-NRR signals processed by generalized spectral subtraction and Wiener filtering in random order to 10 examinees, who selected which signal they considered to contain least musical

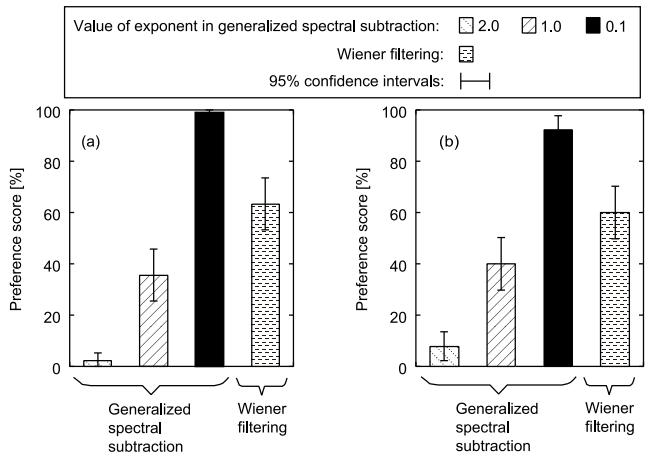

Fig. 8. Subjective evaluation results for (a) white Gaussian noise, and (b) speech noise. We presented four equi-NRR signals processed by generalized spectral subtraction and Wiener filtering in random order to 10 examinees, who selected which signal they considered to contain least musical noise.

noise. It can be confirmed that musical noise is less perceptible in the signal originating from GSS with exponent parameter $2n = 0.1$. This result also supports the result of the theoretical analysis.

C. Summary

In this section, we gave theoretical comparison results based on the HOS-based analytic results described in the previous section. In addition to the theoretical comparison, we performed objective and subjective evaluations.

According to our results,

1) there is no theoretical justification for using power- or amplitude-domain SS, nevertheless about 90% researchers utilize power- or amplitude-domain SS according to Ref. [11]. Instead, generalized spectral subtraction with a lower exponent parameter is advantageous for achieving high-quality noise reduction.
2) With an appropriate parameter, GSS can achieve higher quality noise reduction than that by any Wiener filtering.

The validity of the result is also confirmed by subjective evaluations.

As we described in this section and Sect. III the HOS-based analysis enables us to analyze the amount of musical noise generated by mathematical manner. Actually, the analysis based on HOS reveals new facts mentioned in the section, and their results are also supported by subjective evaluations. Thus, it can be regarded that the HOS-based analysis would become an useful tool to analyze noise reduction techniques based on the perspective of the amount of musical noise generated.

V. Musical-noise-free noise reduction based on higher-order statistics

A. Overview

In this section, we provide a new nonlinear noise reduction method that do not cause any musical noise. This method is based on the analysis by HOS discussed in the previous sections. Hereinafter, we call noise reduction method without musical-noise generation *musical-noise-free* noise reduction method.

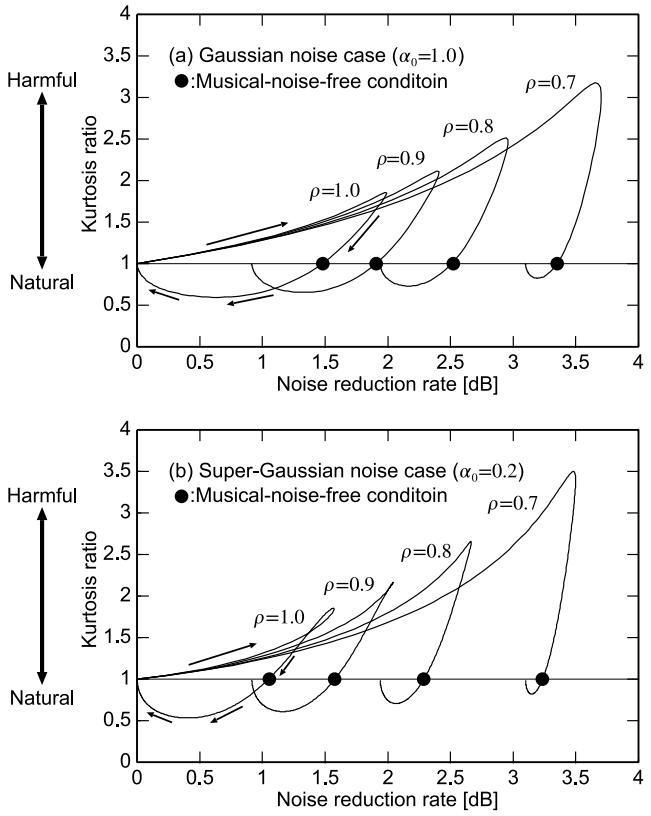

Fig. 9. Relation between NRR and kurtosis ratio from theoretical analysis with increasing β for (a) Gaussian noise case ($\alpha_0 = 1$) and (b) super-Gaussian noise case ($\alpha_0 = 0.2$).

The method is based on iterative SS [21] that iteratively performs weak SS. We found 'musical-noise-free' condition while analyzing this iterative SS. Although the amount of noise reduction generally becomes smaller with a larger flooring parameter ρ in SS, but the remained noise components approach the noise in the original observation. This phenomenon is demonstrated in Fig. 9. In the figure, we performed single shot SS with various oversubtraction parameters, then interestingly we can see a musical-noise-free condition that is a point $\mathrm{NRR} > 0$ but $\mathrm{KR} = 1$. This means a special condition of SS can achieve noise reduction without musical-noise generation. Then, we can achieve musical-noise-free noise reduction by iteratively performinig SS with parameters satisfy the musical noise free condtion.

B. Derivation of musical-noise-free condition [14]

To derive the musical-noise-free condition is equal to finding a fixed-point condition of kurtosis via SS. Although the parameters to be optimized are a flooring parameter ρ and an oversubtraction parameter β, we hereafter show the optimal η given a fixed β for ease of closed-form analysis.

First, we rewrite the kurtosis after performing SS using (6)

and (7) as,

$$\mathrm{kurt}(\alpha_0,\beta,\rho)=\frac{\mathcal{S}(\alpha_0,\beta,4)+\rho^8\mathcal{F}(\alpha_0,\beta,4)}{\left(\mathcal{S}(\alpha_0,\beta,2)+\rho^4\mathcal{F}(\alpha_0,\beta,2)\right)^2}, \tag{32}$$

where

$$\mathcal{S}(\alpha_0,\beta,m)=\sum_{l=0}^{m}(-\beta\alpha_0)^l\frac{\Gamma(m+1)\Gamma(\alpha_0+m-l,\beta\alpha_0)}{\Gamma(\alpha_0)\Gamma(l+1)\Gamma(m-l+1)} \tag{33}$$

$$\mathcal{F}(\alpha_0,\beta,m)=\frac{\gamma(\alpha_0+m,\beta\alpha_0)}{\Gamma(\alpha_0)}. \tag{34}$$

Here supposed $n=1$, which is the exponent parameter of GSS, and α_0 is a shape parameter of a noise signal in an observation. The fixed-point kurtosis condition corresponds to the kurtosis being equal to before and after SS, thus,

$$\frac{\mathcal{S}(\alpha_0,\beta,4)+\rho^8\mathcal{F}(\alpha_0,\beta,4)}{\left(\mathcal{S}(\alpha_0,\beta,2)+\rho^4\mathcal{F}(\alpha_0,\beta,2)\right)^2}=\frac{(\alpha_0+3)(\alpha_0+2)}{(\alpha_0+1)\alpha_0}. \tag{35}$$

Let $\mathcal{H}=\rho^4$ and then (35) yields the following quadratic equation in $\mathcal{H}$.

$$\begin{aligned}&\left(\mathcal{F}(\alpha_0,\beta,4)(\alpha_0+1)\alpha_0-\mathcal{F}^2(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\right)\rho^2\\&-2\mathcal{S}(\alpha_0,\beta,2)\mathcal{F}(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\rho\\&+\mathcal{S}(\alpha_0,\beta,4)(\alpha_0+1)\alpha_0-\mathcal{S}^2(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)=0.\end{aligned} \tag{36}$$

Therefore, we can derive a closed-form estimate of $\mathcal{H}$ from the given oversubtraction parameter as

$$\begin{aligned}\mathcal{H}=&\{\mathcal{F}(\alpha_0,\beta,4)(\alpha_0+1)\alpha_0-\mathcal{F}^2(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\}^{-1}\\&\Bigg[\mathcal{S}(\alpha_0,\beta,2)\mathcal{F}(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\\&\pm\Big[\{\mathcal{S}(\alpha_0,\beta,2)\mathcal{F}(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\}^2\\&-\left\{\mathcal{F}(\alpha_0,\beta,4)(\alpha_0+1)\alpha_0-\mathcal{F}^2(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\right\}\\&\left\{\mathcal{S}(\alpha_0,\beta,4)(\alpha_0+1)\alpha_0-\mathcal{S}^2(\alpha_0,\beta,2)(\alpha_0+3)(\alpha_0+2)\right\}\Big]^{\frac{1}{2}}\Bigg]\end{aligned} \tag{37}$$

Finally, $\rho=\mathcal{H}^{1/4}$ is the resultant flooring parameter that satisfies the fixed-point kurtosis condition. It is worth mentioning that this is just the fixed-point kurtosis condition but is not the musical-noise-free condition. This is because (37) do not consider NRR.

To derive the musical-noise-free condition, we take NRR growth condition into account. From (7) and (13), the NRR growth condition can be expressed by

$$10\log_{10}\frac{\alpha_0}{\mathcal{S}(\alpha_0,\beta,1)+\rho^2\mathcal{F}(\alpha_0,\beta,1)}>0. \tag{38}$$

Since $\rho>0$ we can solve this inequality as

$$0<\rho<\sqrt{\frac{\alpha_0-\mathcal{S}(\alpha_0,\beta,1)}{\mathcal{F}(\alpha_0,\beta,1)}} \tag{39}$$

Overall, we can choose the appropriate parameters satisfying the fixed-kurtosis point condition and NRR growth condition using (37) and (39). Figure 10 illustrates examples of parameters to satisfy the musical-noise-free condition.

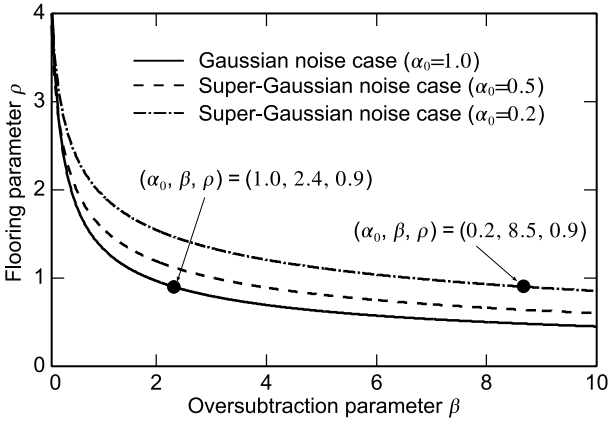

Fig. 10. Example of oversubtraction parameter β and flooring parameter ρ to satisfy musical-noise-free condition.

C. Evaluations

We conducted objective and subjective evaluations to show the efficacy of the iterative SS based on musical-noise-free theory.

First, we compared the proposed musical-noise-free iterative SS and a traditional non-iterative SS on the basis of NRR and kurtosis ratio. In the experiment, observed signals were generated by adding noise signal to target clean speech signals with a SNR of 0 dB. The target speech signals were utterances of four speakers (4 sentences). The noise were white Gaussian and babble noise. The result are illustrated in Fig. 11 [14]. All the scores are the averages in terms of four target speakers. From this result, we can see that the proposed musical-noise-free iterative SS can keep kurtosis ratio mostly closed to 1.0 by NRR = 10 dB. This fact means that the proposed musical-noise-free iterative SS causes extremely less amount of musical noise.

Next, we made a comparison of the proposed musical-noise-free iterative SS and commonly used noise reduction methods on the basis of subjective evaluation. In the evaluation, noisy signals were generated by adding noise signal to target clean speech signals with a SNR of -5, 0, 5, and 10 dB. The noise here we used are white Gaussian, babble noise, real-recorded railway-station noise, real-recorded museum noise, and real-recorded factory noise. Also we chose 4 speakers (4 sentences) for target speakers as same as the previous evaluation. We presented a pair of 10-dB-NRR signals processed by the proposed method and commonly used noise reduction methods, i.e., non-iterative SS, Wiener filtering, and MMSE-STSA estimator, in random order to 10 examinees, who selected which signal they preferred from the viewpoint of total sound quality, e.g., less musical noise, less distortion, etc. In all methods, noise estimation was done by minimum statistics [22].

The result is depicted in Fig. 12 [14]. From this result, it is revealed that resultant signal of the proposed iterative SS is

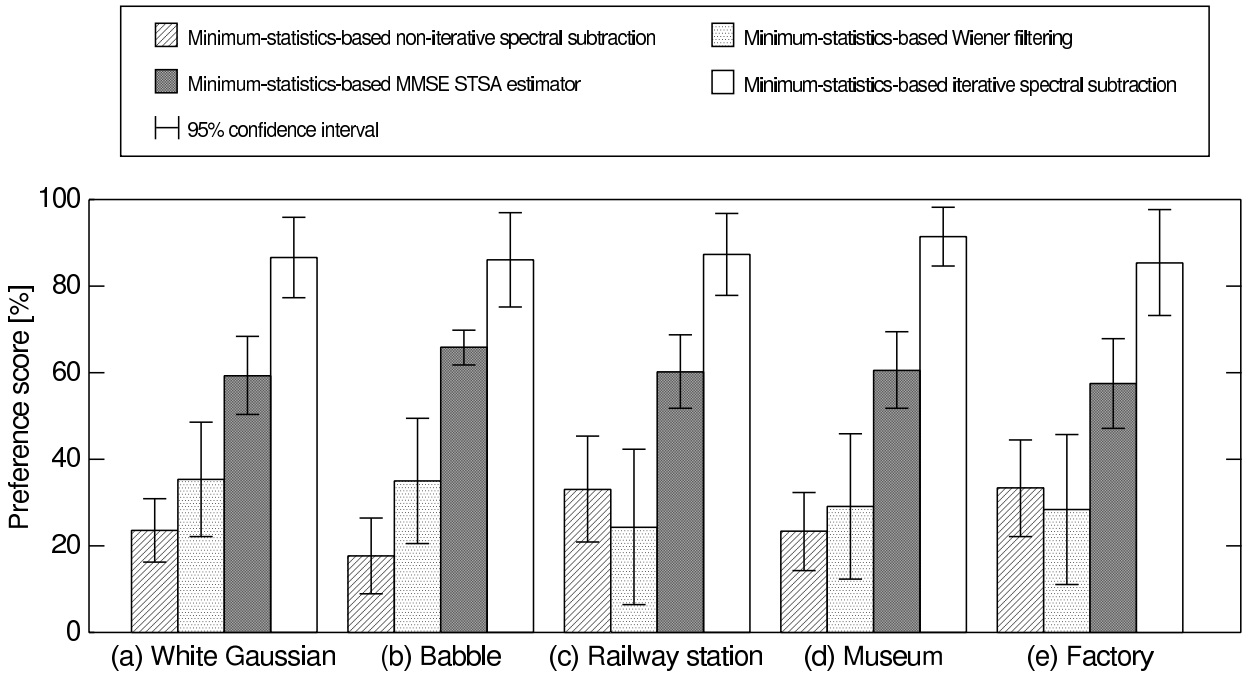

Fig. 12. Subjective evaluation results for (a) white Gaussian noise, (b) babble noise, (c) railway station noise, (d) museum noise and (e) factory noise.

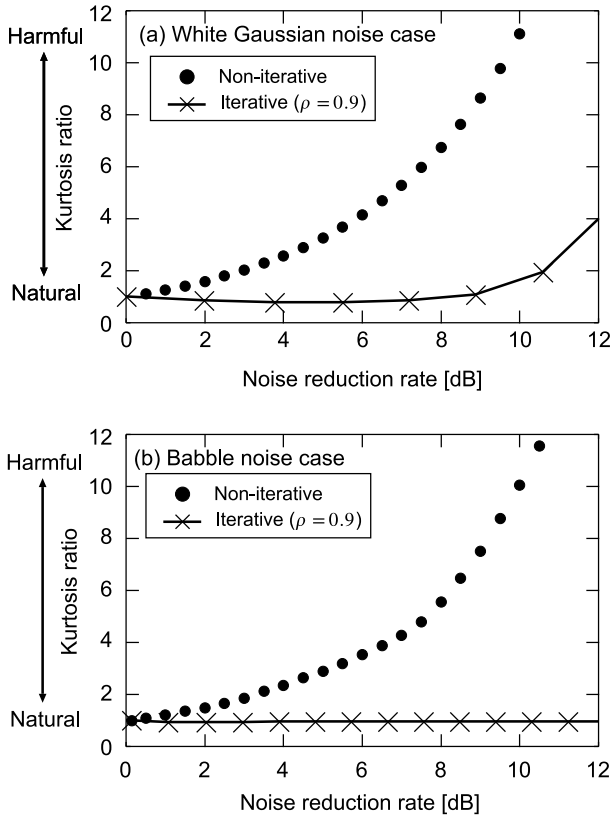

Fig. 11. Relation between NRR and kurtosis ratio obtained from experiment with noisy speech data for (a) white Gaussian noise case ($\alpha_0 = 0.97$), and (b) babble noise case ($\alpha_0 = 0.21$).

preferred to those of commonly used noise reduction methods.

Although this musical-noise-free iterative SS is one example of an optimization of nonlinear noise reduction techniques based on higher-order statistics, it shows a great possibility of using HOS to analyze or optimize noise reduction techniques from the perspective of sound quality as well as the amount of noise reduction.

VI. CONCLUSION

In the paper, we first introduced a HOS-based objective measure for the amount of musical noise generated on the basis of HOS. This objective metric enables us to measure how amount of musical noise generated. Next, we described the theoretical analysis of typical nonlinear noise reduction techniques, i.e., generalized SS and Winer filtering family, by using the HOS-based objective measure. As a result of the analyses, we revealed which method or, which parameter is appropriate for less amount of musical-noise generation. Finally, we demonstrated the musical-noise-free iterative SS that theoretically causes no musical noise. As a result of a subjective evaluation, the output signal of the proposed musical-noise-free iterative SS is surely preferred to those of commonly used noise reduction techniques.

There exists researches using HOS to optimize sound quality, as well as we described in the paper. In Refs. [23]–[25], analyses for the method integrating microphone array and nonlinear noise reduction technique were conducted. Also, an analysis for Wiener filtering with decision-directed a priori SNR estimator has been performed in Ref. [18], [19]. Furthermore, the HOS-based analysis has been applied to prediction of speech recognition performance in Ref. [26].

As described in the paper, the analysis based on higher-order statistics can be applied to various applications to ana-

lyze/optimize the method from the viewpoint of sound quality. Therefore it can be regarded that the HOS-based analysis would become the useful tool to analyze noise reduction techniques in addition to typical objective metrics, e.g., SNR, cepstral distortion.

ACKNOWLEDGMENT

This work was partly supported by the MIC Strategic Information and Communications R&D Promotion Programme (SCOPE), Japan, and JST Core Research of Evolutional Science and Technology (CREST), Japan.

REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol.ASSP-27, no.2, pp.113–120, 1979.

[2] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. ICASSP79*, pp.208–211, 1979.

[3] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Transactions on Speech and Audio Processing*, vol.6, no.4, pp.328–337, 1998.

[4] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*, MIT Press, Cambridge, MA, USA, 1949.

[5] P. C. Loizou, *Speech Enhancement Theory and Practice* CRC Press, Taylor & Francis Group FL, 2007.

[6] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-32, no.6, pp.1109–1121, 1984.

[7] G. W. Elko, "Microphone array systems for hands-free telecommunication," Speech Commun., vol.20, pp.229–240, 1996.

[8] Y. Takahashi, T. Takatani, H. Saruwatari and K. Shikano, "Blind spatial subtraction array with independent component analysys for hands-free speech recognition," *Proc. IWAENC2006*, Sept. 2006.

[9] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Transactions on Audio, Speech and Language Processing*, vol.17, no.4, pp.650-664, May. 2009.

[10] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Automatic optimization scheme of spectral subtraction based on musical noise assessment via HOS," *Proc. IWAENC2008*, 2008.

[11] 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 Transactions on Audio, Speech and Language Processing*, vol.19, no.6, pp.1770-1779, 2011.

[12] T. Inoue, H. Saruwatari, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in Wiener filtering family via higherorder statistics," *Proc. ICASSP2011*, Pp.5076–5079, 2011.

[13] R. Miyazaki, H. Saruwatari, T. Inoue, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement: Theory and evaluation," *Proc. ICASPP2012*, pp.4565–4568, 2012.

[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 Transactions on Audio, Speech and Language Processing*, vol.20, No.7, pp.2080-2094, September 2012.

[15] L. Rabiner, and B. Juang, *Fundamentals of speech recognition* Upper Saddle River, NJ: Prentice Hall PTR, 1993.

[16] E. W. Stacy, "A generalization of the gamma distribution," *Ann. Math. Statist.*, pp./1187–1192, 1962.

[17] J. W. Shin, J. Chang, and N. Kim, "Statistical modeling of speech signals based on generalized gamma distribution," *IEEE Signal Processing Letters*, vol.12, no.3, pp.258–261, 2006.

[18] S. Kanehara, R. Miyazaki, H. Saruwatari, K. Shikanno, and K. Kondo, "Mathematical metric of musical noise for various nonlinear speech enhancement algorithms," *IEICE Technical Report EA2012-44*, pp.67–72, 2012.

[19] S. Kanehara, H. Saruwatari, R. Miyazaki, K. Shikanno, and K. Kondo, "Theoretical analysis of musical noise generation in noise reduction method with decision-directed a priori SNR estimator," *Proc. IWAENC2012*.

[20] J. Even, H. Saruwatari, K. Shikano, and T. Takatani, "Speech enhancement in presence of diffuse background noise: why using blind signal extraction?," *Proc. ICASSP2010*, pp.4770–4773, 2010.

[21] K. Yamashita, S. Ogata, and T. Shimamura, "Spectral subtraction iterated with weighting factors," *Proc. IEEE Speech Coding Workshop*, pp.138–140, 2002,

[22] R. Martin, "Spectral subtraction based on minimum statistics," *Proc. EUSIPCO94*, pp.1182–1185, 1994.

[23] R. Miyazaki, H. Saruwatari, K. Shikano, and K. Kondo, "Musical-noise-free blind speech extraction using ICA-based noise estimation and iterative spectral subtraction," *Proc. ISSPA2012*, pp.322-327, 2012.

[24] 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 (doi:10.1155/2010/431347).

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

[26] R. Miyazaki, H. Saruwatari, R. Wakisaka, K. Shikano, and T. Takatani, "Theoretical analysis of parametric blind spatial subtraction array and its application to speech recognition performance prediction," *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays 2011 (HSCMA2011)*, pp.19–24, 2011.