Theoretical Analysis of Musical Noise in Generalized Spectral Subtraction Based on Higher-Order Statistics

Takayuki Inoue, Hiroshi Saruwatari, *Member, IEEE,* Yu Takahashi, *Student Member, IEEE,* Kiyohiro Shikano, *Fellow, IEEE,* and Kazunobu Kondo

Abstract

In this paper, we provide a new theoretical analysis of the amount of musical noise generated via generalized spectral subtraction based on higher-order statistics. Power spectral subtraction is the most commonly used spectral subtraction method, and in our previous study a musical noise assessment theory limited to the power spectral domain was proposed. In this paper, we propose a generalization of our previous theory on spectral subtraction for arbitrary exponent parameters. We can thus compare the amount of musical noise between any exponent domains from the results of our analysis. We also clarify that less musical noise is generated when we choose a lower-exponent spectral domain; this implies that there is no theoretical justification for using power/amplitude spectral subtraction.

Index Terms

Speech enhancement, Musical noise, Higher-order statistics, Generalized spectral subtraction, Wiener filtering

 This work was partly supported by MIC Strategic Information and Communications R&D Promotion Programme (SCOPE) in Japan, and JST Core Research of Evolutional Science and Technology (CREST) in Japan. T. Inoue, H. Saruwatari, and K. Shikano are with Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, Japan (e-mail: sawatari@is.naist.jp). K. Kondo and Y. Takahashi is with MSP Group, Corporate Research and Development Center, Yamaha Corp. Shizuoka, 438-0192, Japan.

I. Introduction

Over the past decade, the number of applications of speech communication systems, such as TV conference systems, hearing aids, and mobile phones, has increased because speech is the most convenient media for communication among human beings. These systems, however, always suffer from a problem of deterioration of speech quality under adverse noise conditions in real environments such as noisy offices, crowded public spaces, and railway stations. Therefore, in speech signal processing, noise reduction is a problem requiring urgent attention. In this paper, we particularly address noise reduction technology for listening use, highly evaluating the quality of speech-enhanced signals according to human perceptual impressions as well as the amount of noise reduction.

Spectral subtraction is a commonly used noise reduction method that has high noise reduction performance [1], [2]. However, in this method, artificial distortion, so-called *musical noise*, arises owing to nonlinear signal processing, leading to a serious deterioration of sound quality. To cope with the problem, many studies on the analysis of musical noise generation in nonlinear signal processing and its mitigation have been presented (see, e.g., [3], [4]). However, no objective metric to measure how much musical noise is generated has been proposed in previous studies. Thus, it has been difficult to evaluate the amount of musical noise generated and to optimize the internal parameters of a system.

Generally speaking, conventional spectral subtraction methods have a parameter that determines the domain in which the exponent is applied in the spectral subtraction process [5], e.g., the power spectral domain [6], [7], amplitude spectral domain [2], or other domains [8], [9], [4], [10]. We investigated the domain in which the exponent has been used in conventional spectral subtraction methods via Google Scholar, and we found that spectral subtraction is most commonly performed in the power spectral domain with an exponent value of 2 (see Fig. 1). However, to the best of our knowledge, there have been no theoretical studies on the advantages of spectral subtraction in the power spectral domain and no theoretical analysis of the amount of musical noise in domains with different values of the exponent parameter.

Recently, some of the authors have reported that the amount of generated musical noise is strongly correlated with the difference between the higher-order statistics of the power spectra before and after nonlinear signal processing [11], [12], [13]. On the basis of the findings, an objective metric to measure how much musical noise is generated through nonlinear signal processing has been developed. Hence, using this metric, we were able to analyze the amount of musical noise generated via spectral subtraction only in the power spectral domain. However, it still remains as an open problem that there is no theoretical analysis of the amount of musical noise generated in a general setting, where the exponent value may differ from the value of 2 in the power spectral domain.

In this paper, we provide a new theoretical analysis of the amount of musical noise generated, which is a generalization of our previous theory on spectral subtraction, in the case of an arbitrary exponent parameter. We can thus compare the amount of musical noise between any exponent domains from the results of our analysis. We also clarify from mathematical analysis and evaluation experiments that less musical noise is generated when we choose a spectral domain with a lower exponent; this implies a lack of theoretical justification for using the conventional methods of power/amplitude spectral domain subtraction.

In this paper, we also include a theoretical analysis of the amount of musical noise generated in *Wiener filtering*. Historically, two conventional noise reduction methods, namely, spectral subtraction and Wiener filtering, were often compared in past studies (see, e.g., [14]). However, they were simply compared via an experimental measurement of the resultant sound quality, and there have been few comparisons on a theoretical basis. Our theoretical analysis allows the universal performance comparison between spectral subtraction and Wiener filtering from the viewpoint of the amount of musical noise generation and that of noise reduction, enabling the description of the advantages and disadvantages of each method. Note that the main contribution of this paper is not the development of new algorithms but the proposal of a versatile method of theoretical analysis for generalized spectral subtraction.

The rest of this paper is organized as follows. In Sect. II, we describe related works on spectral subtraction and the musical noise metric. In Sect. III, a theoretical analysis of spectral subtraction and its behavior under typical noise conditions is performed. In Sect. IV, noise reduction experiments are described. Following a discussion on the results of the experiments, we present our conclusions in Sect. V.

II. Related Works

A. Formulation of Generalized Spectral Subtraction

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. We formulate *generalized spectral subtraction* [5], [8], [9] in

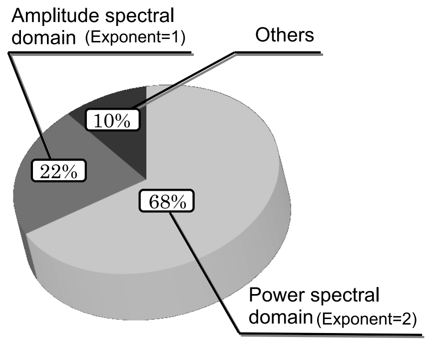

Fig. 1. Value of exponent used in conventional spectral subtraction methods. This investigation was conducted via Google Scholar by surveying 50 highly ranked articles retrieved by the keywords "spectral subtraction."

the time-frequency domain as follows:

$$\hat{S}_{\mathrm{GSS}}(f,\tau) = \begin{cases} \sqrt[2n]{|X(f,\tau)|^{2n} - \beta \cdot \mathrm{E}_{\tau}[|\hat{N}(f,\tau)|^{2n}]}\, e^{j\arg(X(f,\tau))} \\ \qquad (\text{where} \quad |X(f,\tau)|^{2n} - \beta \cdot \mathrm{E}_{\tau}[|\hat{N}(f,\tau)|^{2n}] > 0), \\ 0 \qquad (\text{otherwise}), \end{cases} \tag{1}$$

where $\hat{S}_{\mathrm{GSS}}(f,\tau)$ is the enhanced target speech signal, $X(f,\tau)$ is the observed signal, and $\hat{N}(f,\tau)$ is the estimated noise signal. Also, f denotes the frequency subband, τ is the frame index, $\mathrm{E}_{\tau}[\cdot]$ is the expectation operator of $\cdot$ over τ, β is the subtraction coefficient, and n is the exponent parameter. The case of $n = 1$ corresponds to power spectral subtraction, and the case of $n = 1/2$ corresponds to amplitude spectral subtraction. A block diagram of generalized spectral subtraction is shown in Fig. 2.

In a general setting of spectral subtraction, non-zero flooring is often introduced; thus the second branch in (1) is set to non-zero small value to mitigate musical noise. However, in this paper, we omit it because such a non-zero flooring simply improves the sound quality at a sacrifice of degradation of noise reduction performance. Hence there is a tradeoff between the flooring and noise reduction, and consequently the non-zero flooring is *never* an essential solution for musical noise problem.

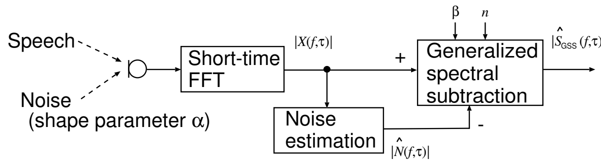

Fig. 2. Block diagram of generalized spectral subtraction.

B. Formulation of Wiener Filtering

Wiener filtering is generally formulated as follows:

$$\hat{S}_{\rm WF}(f,\tau) = G|X(f,\tau)|e^{j\arg(X(f,\tau))}, \tag{2}$$

where $\hat{S}_{\rm WF}(f,\tau)$ is the enhanced target speech signal. G is the gain function, defined by

$$G = \frac{P_{ss}}{P_{ss}+P_{nn}} = \frac{P_{ss}/P_{nn}}{P_{ss}/P_{nn}+1}, \tag{3}$$

where P_{ss} and P_{nn} are the power spectral densities of target speech and noise signal, respectively.

We cannot calculate the a priori signal-to-noise ratio (SNR) P_{ss}/P_{nn} in (3) because we have no information on P_{ss}. In addition, to take into account the nonstationary property of target speech, we use instantaneous values of the observed and noise time-frequency signals. Therefore, we replace the a priori SNR in the gain function with the a posteriori SNR $|X(f,\tau)|/{\rm E}_\tau[|\hat{N}(f,\tau)|]$, and the gain function is reformulated in a time-varying manner as

$$G(f,\tau) \approx \frac{|X(f,\tau)|/{\rm E}_\tau[|\hat{N}(f,\tau)|]}{|X(f,\tau)|/{\rm E}_\tau[|\hat{N}(f,\tau)|]+1} = \frac{|X(f,\tau)|}{|X(f,\tau)|+{\rm E}_\tau[|\hat{N}(f,\tau)|]}. \tag{4}$$

Moreover, we extend (4) to a square-root and parametric form to achieve better and flexible noise reduction; the gain function is given by [15], [16]

$$G(f,\tau) = \sqrt{\frac{|X(f,\tau)|^2}{|X(f,\tau)|^2+\xi{\rm E}_\tau[|\hat{N}(f,\tau)|^2]}}, \tag{5}$$

where ξ is the processing strength parameter.

Note that there exists a conventional approach in which the a priori SNR is replaced with a function of the instantaneous observed signal, i.e., the relation, $P_{ss} = P_{xx} - P_{nn} \approx |X(f,\tau)|^2 - P_{nn}$, is used in (3) [1]. However, in our preliminary experiments, it has been clarified that the conventional method is inferior to (5) in terms of musical noise generation and speech distortion. In addition, another approach for updating P_{ss} in a *decision-directed* fashion is often used [1]. However, this paper would not address it because its mathematical analysis of higher-order statistics is difficult, remaining as a future work. Although, strictly speaking, (5) is not a standard Wiener filtering but a modified Wiener filtering method, we still call (5) Wiener filtering in this paper and regard it as a subject to be analyzed.

C. Mathematical Metric of Musical Noise Generation via Higher-Order Statistics [11]

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 skirt of their probability density function (p.d.f.). Therefore, quantifying the skirt 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 skirt, meaning that the signal has many tonal components. Kurtosis is defined as

$$\mathrm{kurt} = \frac{\mu_4}{\mu_2^2}, \tag{6}$$

where "kurt" is the kurtosis and μ_m is the mth-order moment, given by

$$\mu_m = \int_0^\infty x^m P(x) dx, \tag{7}$$

where $P(x)$ is the p.d.f. of a power spectral component x. Note that μ_m is not a central moment but a raw moment. Thus, (6) is not kurtosis in the mathematically strict definition but a modified version; we still refer to (6) as kurtosis in this paper.

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, even though these signals contain target-speech-dominant periods. Thus, this analysis should be conducted during, for example, 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 [12], 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 the *kurtosis ratio* as a measure to assess musical noise [11]. This measure is defined as

$$\text{kurtosis ratio} = \frac{\mathrm{kurt}_{\mathrm{proc}}}{\mathrm{kurt}_{\mathrm{org}}}, \tag{8}$$

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. [11], it was reported that the kurtosis ratio is strongly correlated with the human perception of musical noise.

III. Theoretical analysis of generalized spectral subtraction and Wiener filtering

A. Analysis Strategy

In this section, we analyze the amount of noise reduction and musical noise generated through generalized spectral subtraction and Wiener filtering using kurtosis. In the analysis, we first model a noise signal by a gamma distribution and formulate the resultant p.d.f. after generalized spectral subtraction (see Sect. III-B). Then, kurtosis is obtained from the 2nd- and 4th-order moments, and the amount of noise reduction is calculated from the 1st-order moment (see Sect. III-C). Also, we analyze the amount of musical noise and noise reduction in Wiener filtering (see Sect. III-D). Finally, we compare the kurtosis values upon changing the exponent parameter (n in (1)) under the same amount of noise reduction (see Sect. III-E).

B. Process of Deforming P.d.f. of Input Noise Signal via Generalized Spectral Subtraction

1) Modeling of Input Signal: The p.d.f. is deformed via multiple processes in generalized spectral subtraction (see Fig. 3). These processes are as follows: the nth-exponentiation operation, subtraction in the spectral domain, and the extraction of the nth root. In this section, we formulate the p.d.f. in each process.

We assume that the input signal x in the power spectral domain can be modeled by the gamma distribution as [17], [18]

$$P(x) = \frac{x^{\alpha-1}\exp(-\frac{x}{\theta})}{\Gamma(\alpha)\theta^{\alpha}}, \tag{9}$$

where α is the shape parameter corresponding to the type of noise (e.g., $\alpha = 1$ is Gaussian and $\alpha < 1$ is super-Gaussian), θ is the scale parameter of the gamma distribution, and $\Gamma(\alpha)$ is the *gamma function*, defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}\exp(-t)dt. \tag{10}$$

Full details of the three processes involved in the deformation of the p.d.f. are described in the following sections.

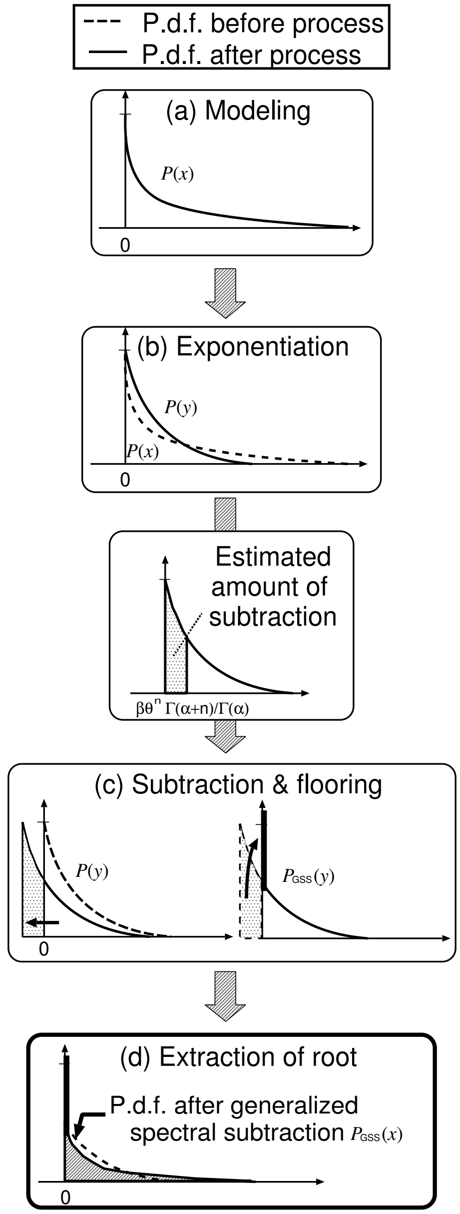

Fig. 3. Deformation of p.d.f. in generalized spectral subtraction.

2) Exponentiation Operation: The original p.d.f. $P(x)$ is first deformed by the exponentiation operation (see Fig. 3(b)). We can calculate the resultant p.d.f. $P(y)$ 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

$$P(y) = P(g^{-1}(y))|J|, \tag{11}$$

where $|J|$ is the Jacobian of the transformation, defined by

$$|J| = \left|\frac{\partial g^{-1}}{\partial y}\right|. \tag{12}$$

We apply (11) to (9). Since x is the power spectral domain signal, y is expressed as $y = x^n$, i.e., the Jacobian is

$$|J| = \left|\frac{\partial x}{\partial y}\right| = \left|\frac{1}{nx^{n-1}}\right| = \left|\frac{1}{ny^{(n-1)/n}}\right|. \tag{13}$$

Consequently,

$$P(y) = P(x)|J| = \frac{y^{\alpha/n-1}\exp(-\frac{y^{1/n}}{\theta})}{n\Gamma(\alpha)\theta^{\alpha}}. \tag{14}$$

3) Subtraction Process in Exponent Spectral Domain: Next, the amount of subtraction in the generalized spectral subtraction is estimated. This corresponds to the estimated noise spectrum multiplied by the oversubtraction parameter β, where the estimated noise spectrum is the mean of noise, $\mathrm{E}[y]$, given by

$$\mathrm{E}[y] = \int_0^\infty yP(y) = \int_0^\infty \frac{y^{\alpha/n}\exp(-\frac{y^{1/n}}{\theta})}{n\Gamma(\alpha)\theta^{\alpha}}dy. \tag{15}$$

Here, we let $t = y^{1/n}/\theta$, then $dy = n\theta(\theta t)^{n-1}dt$, and the range of the integral does not change. Consequently,

$$\mathrm{E}[y] = \frac{\theta^n}{\Gamma(\alpha)}\int_0^\infty t^{\alpha+n-1}\exp(-t)dt, \tag{16}$$

and, from $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt$, the amount of estimated noise is

$$\mathrm{E}[y] = \frac{\theta^n\Gamma(\alpha+n)}{\Gamma(\alpha)}. \tag{17}$$

In the subtraction process, the p.d.f. in the exponent spectral domain undergoes a lateral shift of $\beta\mathrm{E}[y]$ in the zero-power direction. As a result, a negative power component with a nonzero probability arises. To avoid this, the negative component is replaced with zero (see Fig. 3(c)). Thus, the resultant p.d.f. after subtraction is

$$\begin{aligned} &P_{\mathrm{GSS}}(y) \\ &= \begin{cases} \frac{1}{n\theta^{\alpha}\Gamma(\alpha)}(y+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{\alpha/n-1} \\ \exp\left(-\frac{(y+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}}{\theta}\right) & (y>0), \\ \frac{1}{n\theta^{\alpha}\Gamma(\alpha)}\int_0^{\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha)} z^{\alpha/n-1}\exp(-\frac{z^{1/n}}{\theta})dz & (y=0). \end{cases} \end{aligned} \tag{18}$$

4) Extraction of nth Root: We apply the extraction of the nth root to $P_{\mathrm{GSS}}(y)$ given by (18), and reconstruct the p.d.f. in the power spectral domain, $P_{\mathrm{GSS}}(x)$. In a similar way to in Sect. III-B2, we let $x = y^{1/n}$ and apply a change of variables, where the Jacobian is

$$|J| = |\frac{\partial y}{\partial x}| = \frac{n}{y^{(1-n)/n}} = \frac{n}{x^{1-n}}. \tag{19}$$

Consequently, the resultant p.d.f. after generalized spectral subtraction, $P_{\mathrm{GSS}}(x)$, is given by

$$\begin{aligned} P_{\mathrm{GSS}}(x) &= P_{\mathrm{GSS}}(y)|J| \\ &= \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{n-1}(x^n + \beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{\alpha/n-1} \\ \exp\left(-\frac{(x^n+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}}{\theta}\right) & (x > 0), \\ \frac{1}{\theta^\alpha\Gamma(\alpha)} \int_0^{\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha)} z^{\alpha-1}\exp(-\frac{z}{\theta})dz & (x = 0). \end{cases} \end{aligned} \tag{20}$$

C. Estimation of Amount of Musical Noise and Noise Reduction

1) The mth-order moment of $P_{\mathrm{GSS}}(x)$: The mth-order moment of $P_{\mathrm{GSS}}(x)$ is given by

$$\begin{aligned} \mu_m &= \int_0^\infty x^m P_{\mathrm{GSS}}(x)dx \\ &= \frac{1}{\theta^\alpha\Gamma(\alpha)} \int_0^\infty x^{m+n-1}(x^n + \beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{\alpha/n-1} \\ &\quad \exp\left(-\frac{(x^n + \beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}}{\theta}\right)dx. \end{aligned} \tag{21}$$

Let $t = (x^n + \beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}/\theta$, then $dy = n\theta(\theta t)^{n-1}dt$, and the range of the integral changes from $[0, \infty]$ to $[(\beta\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}, \infty]$. Thus, μ_m is given by

$$\mu_m = \frac{\theta^m}{\Gamma(\alpha)} \int_{\left\{\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{1/n}}^\infty \left\{t^n - \beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{m/n} t^{\alpha-1}\exp(-t)dt. \tag{22}$$

Using the *binomial theorem* under the condition that m/n is a natural number, we can rewrite $\{t^n - \beta\Gamma(\alpha+n)/\Gamma(\alpha)\}^{m/n}$ in (22) as

$$\begin{aligned} &\left\{t^n - \beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{m/n} \\ &= \sum_{l=0}^{m/n} \left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l \frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)} t^{n(m/n-l)}. \end{aligned} \tag{23}$$

Consequently, the mth-order moment of $P_{\mathrm{GSS}}(x)$ is given by

$$\begin{aligned}\mu_m =&\frac{\theta^m}{\Gamma(\alpha)}\sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l\frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)}\\ &\int_{\left\{\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{1/n}}^{\infty} t^{\alpha+m-ln-1}\exp(-t)dt\\ =&\frac{\theta^m}{\Gamma(\alpha)}\sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l\frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)}\\ &\Gamma(\alpha+m-ln,(\beta\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}),\end{aligned} \tag{24}$$

where $\Gamma(\alpha,z)$ is the upper incomplete gamma function defined as

$$\Gamma(\alpha,z)=\int_z^{\infty} t^{\alpha-1}\exp(-t)dt. \tag{25}$$

2) Analysis of Amount of Musical Noise: Using (24), we can obtain the kurtosis after generalized spectral subtraction as

$$\mathrm{kurt}_{\mathrm{GSS}}=\frac{\mu_4}{\mu_2^2}=\Gamma(\alpha)\frac{\mathcal{M}_{\mathrm{GSS}}(\alpha,\beta,4/n)}{\mathcal{M}^2_{\mathrm{GSS}}(\alpha,\beta,2/n)}, \tag{26}$$

where

$$\begin{aligned}\mathcal{M}_{\mathrm{GSS}}(\alpha,\beta,m/n)=&\sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l\frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)}\\ &\Gamma(\alpha+m-ln,(\beta\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}).\end{aligned} \tag{27}$$

By substituting $\beta=0$ into (26), we can estimate the kurtosis before processing. Thus, we can calculate the resultant kurtosis ratio as

$$\text{kurtosis ratio}=\frac{\mathcal{M}_{\mathrm{GSS}}(\alpha,\beta,4/n)/\mathcal{M}^2_{\mathrm{GSS}}(\alpha,\beta,2/n)}{\mathcal{M}_{\mathrm{GSS}}(\alpha,0,4/n)/\mathcal{M}^2_{\mathrm{GSS}}(\alpha,0,2/n)}. \tag{28}$$

3) Analysis of Amount of Noise Reduction: We analyze the amount of noise reduction via generalized spectral subtraction. 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 [19]. The NRR is

$$\mathrm{NRR}=10\log_{10}\frac{\mathrm{E}[s^2_{\mathrm{out}}]/\mathrm{E}[n^2_{\mathrm{out}}]}{\mathrm{E}[s^2_{\mathrm{in}}]/\mathrm{E}[n^2_{\mathrm{in}}]}, \tag{29}$$

where s_{in} and s_{out} are the input and output speech signals, respectively, and n_{in} and n_{out} are the input and output noise signals, respectively. Here, the denominator in (29) is the 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 spectral subtraction, i.e., $\mathrm{E}[s_{\mathrm{out}}^2] \simeq \mathrm{E}[s_{\mathrm{in}}^2]$, then

$$\mathrm{NRR} = 10\log_{10}\frac{\mathrm{E}[n_{\mathrm{in}}^2]}{\mathrm{E}[n_{\mathrm{out}}^2]}. \tag{30}$$

Since, $\mathrm{E}[n_{\mathrm{in}}^2] = \mu_1$ when $\beta = 0$ in (24) and $\mathrm{E}[n_{\mathrm{out}}^2] = \mu_1$ for a specific (nonzero) β,

$$\mathrm{NRR} = 10\log_{10}\frac{\mathcal{M}_{\mathrm{GSS}}(\alpha, 0, 1/n)}{\mathcal{M}_{\mathrm{GSS}}(\alpha, \beta, 1/n)}. \tag{31}$$

In summary, we can derive theoretical estimates for the amount of musical noise and NRR using (28) and (31). This greatly simplifies the analysis because both equations are expressed analytically in a form that does not include any integrals.

D. Analysis of Wiener Filtering

In the same manner as in the previous subsections, we analyze kurtosis and NRR for Wiener filtering in this section. The original p.d.f. $P(x)$ is transformed into the resultant p.d.f. $P_{\mathrm{WF}}(y)$ via Wiener filtering. We can calculate $P_{\mathrm{WF}}(y)$ by considering a change of variables of the p.d.f. Since x is the power spectral domain signal and its mean value ($\mathrm{E}_\tau[|\hat{N}(f,\tau)|^2]$) is given by $\alpha\theta$ in the gamma distribution, y for Wiener filtering is expressed as

$$y = \frac{x^2}{x + \xi\alpha\theta}. \tag{32}$$

We can obtain the mth-order moment of $P_{\mathrm{WF}}(y)$ as

$$\mu_m = \frac{\theta^m}{\Gamma(\alpha)}\mathcal{M}_{\mathrm{WF}}(\alpha, \xi, m), \tag{33}$$

where

$$\mathcal{M}_{\mathrm{WF}}(\alpha, \xi, m) = \int_0^\infty \frac{t^{\alpha+2m-1}}{(t + \xi\alpha)^m}\exp(-t)dt. \tag{34}$$

The detailed derivation of (33) and (34) is given in Appendix A. Therefore, we can calculate the resultant kurtosis as

$$\text{kurtosis ratio} = \frac{\mathcal{M}_{\mathrm{WF}}(\alpha, \xi, 4)/\mathcal{M}_{\mathrm{WF}}^2(\alpha, \xi, 2)}{\mathcal{M}_{\mathrm{WF}}(\alpha, 0, 4)/\mathcal{M}_{\mathrm{WF}}^2(\alpha, 0, 2)}, \tag{35}$$

and the resultant NRR as

$$\mathrm{NRR} = 10\log_{10}\frac{\mathcal{M}_{\mathrm{WF}}(\alpha, 0, 1)}{\mathcal{M}_{\mathrm{WF}}(\alpha, \xi, 1)}. \tag{36}$$

In summary, even for Wiener filtering, we can derive theoretical estimates for the amount of musical noise and NRR using (35) and (36). Although the internal equation (34) still contains an integral, we can calculate it using a numerical integral method in our study.

E. Comparison of Amount of Musical Noise under Same NRR Condition

According to the above analysis, we can compare the amount of musical noise between generalized spectral subtraction with different exponent parameters and Wiener filtering under the same amount of noise reduction. Figures 4–6 show the theoretical behaviors of the kurtosis ratio and NRR for various parameter values. In these figures, the shape parameter α is set to 0.2, 0.5 or 1.0, NRR is varied from 0 to 12 dB, and the exponent domain in generalized spectral subtraction is set to 2.0 (i.e., power domain spectral subtraction), 1.0 (i.e., amplitude domain spectral subtraction), 0.5, or 0.1. The subtraction coefficient β in generalized spectral subtraction and the processing strength parameter ξ in Wiener filtering are adjusted so that the target speech NRR is achieved. Note that we plot the logarithm of the kurtosis ratio because the kurtosis exponentially increases with β in spectral subtraction [11]. We call this the *log kurtosis ratio* hereafter.

Figures 4–6 show that a smaller amount of musical noise is generated when a lower exponent parameter is used, regardless of the type of noise and NRR. These figures also indicate that for higher values of NRR, there is a larger difference between the kurtosis ratio for different values of the exponent parameter. This implies that humans perceive a greater variation of musical noise at a higher NRR. In addition, it is revealed that this variation is less perceptible for super-Gaussian noise.

Also, Figs. 4–6 indicate that a small amount of musical noise is generated when we use Wiener filtering in comparison with power/amplitude domain spectral subtraction, particularly at a higher NRR. In contrast, generalized spectral subtraction with a lower exponent domain (=0.5 and 0.1) generates less musical noise than Wiener filtering.

IV. Evaluation Experiments and Results

A. Experimental Conditions

We conducted objective and subjective evaluation experiments to confirm the validity of the theoretical analysis described in the previous section. Noisy observation 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

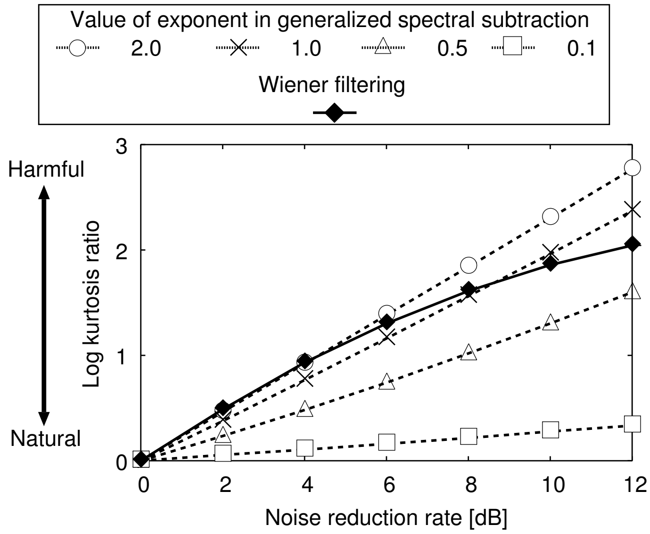

Fig. 4. Theoretical behavior of NRR and log kurtosis ratio given by (28), (31), (35), and (36) in generalized spectral subtraction and Wiener filtering for Gaussian noise (α = 1.0).

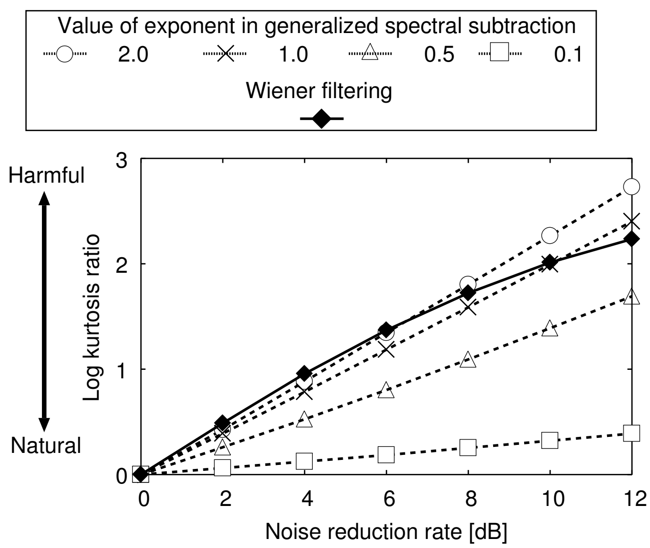

Fig. 5. Theoretical behavior of NRR and log kurtosis ratio given by (28), (31), (35), and (36) in generalized spectral subtraction and Wiener filtering for super-Gaussian noise (α = 0.5).

four speakers (4 sentences), and the noise signals were white Gaussian noise and speech noise, where the speech noise was recorded human speech emitted from 36 loudspeakers. The length of each signal was 7 s, and each signal was sampled at 16 kHz. The FFT size was 1024, and the frame shift length was 256. The shape parameter of the white Gaussian noise was 0.96 and that of the speech noise was

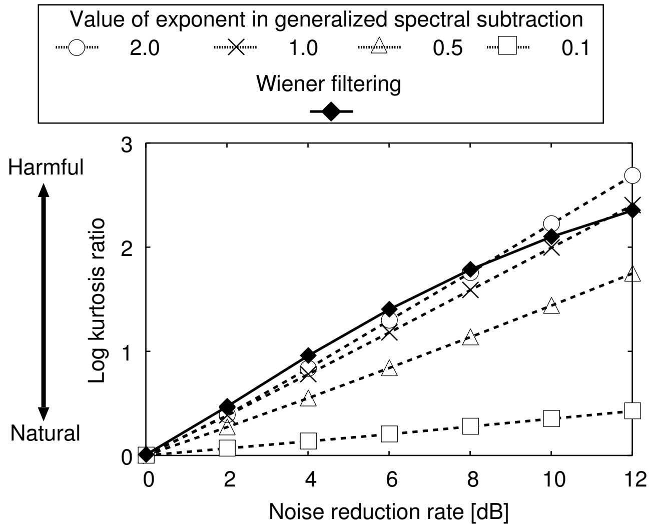

Fig. 6. Theoretical behavior of NRR and log kurtosis ratio given by (28), (31), (35), and (36) in generalized spectral subtraction and Wiener filtering for super-Gaussian noise ($\alpha = 0.2$).

TABLE I

CONDITIONS OF EVALUATION

NRR [dB]	4, 8, 12
Value of exponent	2.0, 1.0, 0.5, 0.1
Objective evaluation measure	(1) log kurtosis ratio (2) cepstral distortion
Subjective evaluation measure	preference score of 10 examinees

0.21. We conducted our experiments on Gaussian and super-Gaussian noise.

In these experiments, we assumed that the noise prototype, i.e., the average of $|\hat{N}(f,\tau)|^2$, was perfectly estimated. In addition, the log kurtosis ratio and NRR were calculated from the observed and processed signals. Other experimental conditions are listed in Table I.

B. Objective Evaluation

We first conducted an objective experiment and evaluated the sound quality of processed signals on the basis of cepstral distortion [20] and log kurtosis ratio. Here, we calculated the log kurtosis ratio from the noise-only period and the cepstral distortion from the target speech components. The cepstral distortion is

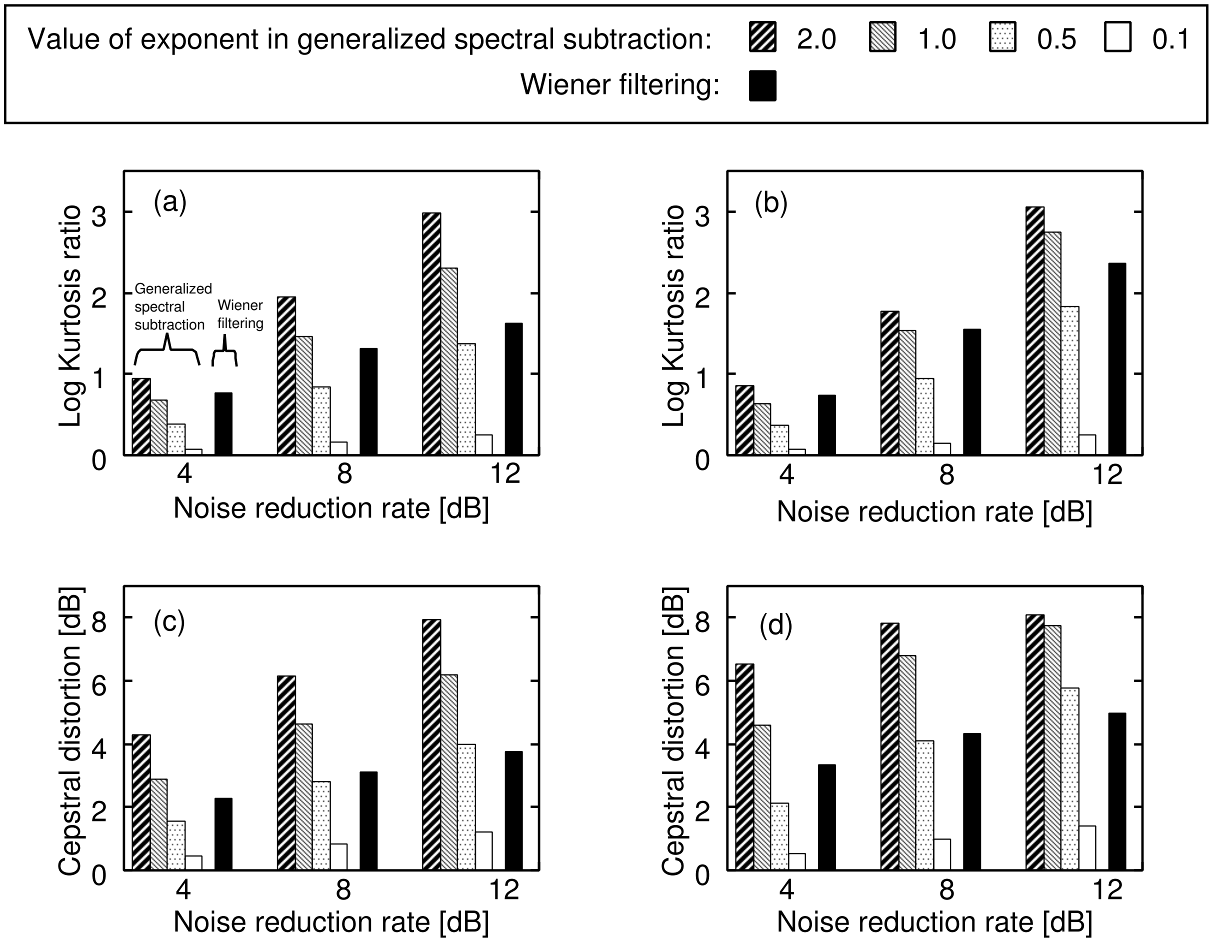

Fig. 7. Results of log kurtosis ratio and cepstral distortion for various domain values of the exponent. (a) and (c) show the results for white Gaussian noise, and (b) and (d) show the results for speech noise.

a measure of the degree of distortion via the cepstrum domain. The cepstral distortion indicates distortion among two signals, which is defined as

$$\mathrm{CD}\ [\mathrm{dB}] \equiv \frac{20}{T \log 10} \sum_{\tau=1}^{T} \sqrt{\sum_{\rho=1}^{B} 2(C_{\mathrm{out}}(\rho, \tau) - C_{\mathrm{ref}}(\rho, \tau))^2}, \tag{37}$$

where T is the frame length, $C_{\mathrm{out}}(\rho, \tau)$ is the ρth cepstral coefficient of the output signal in frame τ, and $C_{\mathrm{ref}}(\rho, \tau)$ is the ρth cepstrum coefficient of the original speech signal. B is the number of dimensions of the cepstrum used in the evaluation; we set $B = 22$. The small value of cepstral distortion indicates that the sound quality of the target speech part is high.

The result of the experiment is depicted in Fig. 7. Regarding generalized spectral subtraction, the figure shows that the log kurtosis ratio decreases as the exponent parameter becomes smaller and that

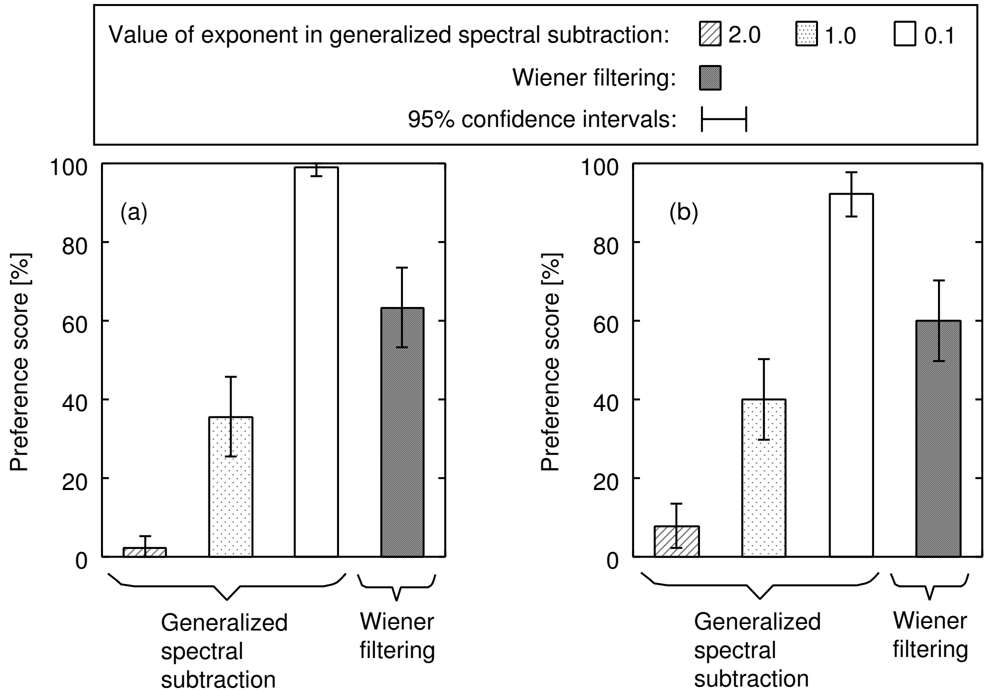

Fig. 8. Subjective evaluation results for (a) white Gaussian noise, and (b) speech noise. We presented four equi-NRR signals processed by generalized spectral subtraction and Wiener filtering in random order to 10 examinees, who selected which signal they considered to contain least musical noise.

the difference between the log kurtosis ratio of distinct exponent parameters is increased if the input noise is Gaussian. These results are consistent with the results of the theoretical analysis provided in Sect. III-E. In addition, cepstral distortion decreases when the exponent parameter is set to a small value. Consequently, in all cases, we can achieve high sound quality upon setting a lower exponent parameter in generalized spectral subtraction.

In addition, the figure shows that the log kurtosis ratio in Wiener filtering is comparable to or smaller than that in power/amplitude domain spectral subtraction. However, for generalized spectral subtraction with a small exponent domain, the result is reversed, i.e., generalized spectral subtraction has less musical noise than Wiener filtering. This tendency is in good agreement with the results of the theoretical analysis in Sect. III-E. In addition, cepstral distortion has an similar tendency. Consequently, we can achieve higher-quality noise reduction in Wiener filtering than in the commonly used power/amplitude domain spectral subtraction; moreover, we can obtain a further improvement if we use generalized spectral subtraction with a lower exponent domain.

C. Subjective Evaluation

We next conducted a subjective evaluation. In the evaluation, we presented four equi-NRR signals processed by power-, amplitude-, and 0.1-exponent-domain spectral subtraction and Wiener filtering in random order to 10 examinees, who selected which signal they considered to contain least musical noise.

The result of the experiment is shown in Fig. 8. It was found that musical noise is less perceptible when generalized spectral subtraction with a lower exponent domain is used. This result is also consistent with our theoretical analysis, thus confirming the validity of the proposed method of theoretical analysis.

D. Remarks

Although the most commonly used method of noise reduction is power/amplitude spectral domain subtraction, our results clarify that there is no theoretical justification for using the corresponding exponent values (= 2 or 1); instead, we recommend that the exponent parameter should be as small as possible to minimize the amount of musical noise generated. Note that there are no side effects in the utilization of a small exponent parameter because we confirmed the decrease in both kurtosis ratio and cepstral distortion in Fig. 7. This finding is expected to be of interest to all researchers using the spectral subtraction technique. A very slight modification of the current software code will enable us to realize better-quality noise reduction without performing any additional pre/postprocessing [21], [4], [22] to mitigate musical noise.

It is worth discussing that we can obtain more better results by setting very low exponent parameter such as $n = 0$. We have already carried out an experiment under the condition that $n = 0$ or very close to zero. First, the case of $n = 0$ corresponds to a singular case, and we have no valid results from signal processing. Next, except for $n = 0$, we can obtain better noise reduction performance as n becomes smaller; e.g., for white Gaussian noise, log kurtosis ratio of 0.0028 and cepstral distortion of 0.0231 (under 12 dB NRR) are measured in the case of $n = 0.0005$. Note that, generally speaking, signal processing of generalized spectral subtraction with too small n ($\ll 0.0005$) cannot work well due to the limitation of the computational precision, causing an over/underflow in calculations.

Regarding the relationship between generalized spectral subtraction and Wiener filtering, we can rewrite (32) in Wiener filtering as

$$y = \exp\left(\log x - \log(1 + \frac{\xi\alpha\theta}{x})\right), \tag{38}$$

which approximately means that the subtraction is conducted in the logarithm domain and then the exponential transformation is applied, resulting in the power spectral domain. This process is similar to

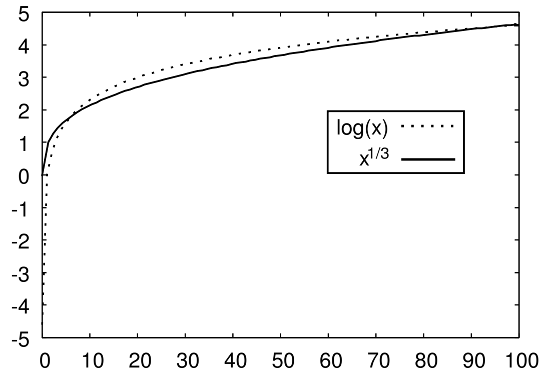

Fig. 9. Curves of logarithm function $\log(x)$ and low exponent function $x^{1/3}$.

that of generalized spectral subtraction with an exponent parameter of less than one. For example, the logarithm function $\log(x)$ is plotted with a curve of the low exponent function $x^{1/3}$ in Fig. 9, showing a good agreement in both functions' shapes. Thus, (38) roughly corresponds to (1) with $n = 1/3$. From the findings, in nonlinear speech enhancement, we consider that the above-mentioned *compress-subtract-expand* process plays an important role in achieving less musical noise generation. It can be also expected that the less-musical noise property of the logarithm-exponent process is related to a superiority in the log-spectral amplitude estimator [23]; however, the detailed theoretical analysis remains as an open problem.

V. Conclusion

In this study, we first performed a theoretical analysis of the amount of musical noise generated via generalized spectral subtraction and Wiener filtering based on higher-order statistics. Our theoretical analysis indicates that the 1st-, 2nd-, and 4th-order moments of the power spectral grids can be used to estimate the amount of noise reduction and musical noise generation. Next, we conducted experimental objective and subjective comparisons of the amount of musical noise for generalized spectral subtraction with distinct exponent parameters and Wiener filtering under the same noise reduction performance.

It was clarified from the mathematical analysis and evaluation experiments that less musical noise is generated in a spectral domain with a lower exponent. Futhermore, it was also revealed that less musical noise is generated in Wiener filtering than in power/amplitude domain spectral subtraction. However, when we use a lower exponent domain in spectral subtraction, we can obtain an enhanced speech signal with less musical noise.

In summary, our theory mathematically proves that there is no theoretical justification for using power/amplitude spectral subtraction. Instead, generalized spectral subtraction with a lower exponent parameter is advantageous for achieving high-quality noise reduction.

The method of theoretical analysis proposed in this paper was mainly aimed to address the performance assessment for generalized spectral subtraction. Needless to say, there exist a plenty of modified versions in spectral subtraction and various noise reduction methods. Some of them have been successfully analyzed by the proposed approach, e.g., integration method of spectral subtraction and beamforming [13] and iterative spectral subtraction method [24], but most of them are ongoing works and not easily analyzed. However, it is still expected that our basic idea utilizing higher-order statistics for musical noise generation analysis has a great possibility to provide a new basis of versatile sound-quality assesment as an open problem in future.

Appendix A
Derivation of mth-order moment in Wiener filtering

In this section, we formulate the p.d.f. in Wiener filtering. Since x is the power spectral domain signal and its mean value ($\mathrm{E}_\tau[|\hat{N}(f,\tau)|^2]$) is given by $\alpha\theta$ in the gamma distribution, y for Wiener filtering is expressed as

$$y = \frac{x^2}{x + \xi\alpha\theta}. \tag{39}$$

This results in the following quadratic equation in x to be solved:

$$x^2 - yx - y\xi\alpha\theta = 0, \tag{40}$$

from which we can obtain the closed-form solution

$$x = \frac{y + \sqrt{y^2 + 4y\xi\alpha\theta}}{2} = f(y). \tag{41}$$

Since $x > 0$ and $y > 0$, the Jacobian is

$$\frac{dx}{dy} = f'(y) = |J|. \tag{42}$$

Consequently, the resultant p.d.f. after Wiener filtering, $P_{\mathrm{WF}}(y)$, is given by

$$P_{\mathrm{WF}}(y) = \frac{(f(y))^{\alpha-1}\exp(\frac{-f(y)}{\theta})}{\Gamma(\alpha)\theta^\alpha} f'(y). \tag{43}$$

The mth-order moment of $P_{\mathrm{WF}}(y)$ is given by

$$\begin{aligned}\mu_m &= \int_0^\infty y^m P_{\mathrm{WF}}(y)dy \\ &= \int_0^\infty y^m \frac{(f(y))^{\alpha-1}\exp(\frac{-f(y)}{\theta})}{\Gamma(\alpha)\theta^\alpha} f'(y)dy. \end{aligned} \tag{44}$$

Let $t = f(y)/\theta$, then $dy = \theta/f'(y)\ dt$ and the range of the integral does not change. Futhermore, from (41), $f(y)$ is expressed as

$$f(y) = \theta t = x. \tag{45}$$

We apply (45) to (39), then y^m is expressed as

$$y^m = \left\{\frac{(\theta t)^2}{\theta t + \xi\alpha\theta}\right\}^m = \frac{\theta^m t^{2m}}{(t+\xi\alpha)^m}. \tag{46}$$

Thus, we apply (45) and (46) to (44), then μ_m is given by

$$\mu_m = \frac{\theta^m}{\Gamma(\alpha)}\mathcal{M}_{\mathrm{WF}}(\alpha,\xi,m), \tag{47}$$

where

$$\mathcal{M}_{\mathrm{WF}}(\alpha,\xi,m) = \int_0^\infty \frac{t^{\alpha+2m-1}}{(t+\xi\alpha)^m}\exp(-t)dt. \tag{48}$$

References

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

[2] 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.

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

[4] Z. Goh, K.-C. Tan, and B. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," *IEEE Transactions on Speech and Audio Processing*, vol.6, no.3, pp.287–292, 1998.

[5] J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Prentice-Hall, Inc. NJ, p.509, 1987.

[6] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol.28, no.2, pp.137–145, 1980.

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

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

[9] 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.

[10] J. Li, H. Jiang, and M. Akagi, "Psychoacoustically-motivated adaptive β-order generalized spectral subtraction based on data-driven optimization," *Proc. Interspeech2008* pp.171–173, 2008.

[11] 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. International Workshop on Acoustic Echo and Noise Control (IWANENC2008)*, 2008.

[12] 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.

[13] 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.

[14] J. H. L. Hansen, and M. A. Clements, "Iterative speech enhancement with spectral constraints" *Proc. ICASSP87*, vol.1, pp.189–192, 1987.

[15] Y. Takahashi, K. Osako, H. Saruwatari, and K. Shikano, "Blind source extraction for hands-free speech recognition based on Wiener filtering and ICA-based noise estimation," *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, pp.164–167, 2008.

[16] 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.

[17] E. W. Stacy, "A generalization of the gamma distribution," *Ann. Math. Stat.*, vol.33, no.3, pp.1187–1192, 1962.

[18] J. W. Shin, J.-H. Chang, and N. S. Kim, "Statistical modeling of speech signal based on generalized gamma distribution," *IEEE Signal Processing Letters*, vol.12, no.3, pp.258–261, 2005.

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

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

[21] G. Whipple, "Low residual noise speech enhancement utilizing time-frequency filtering," *Proc. ICASSP94*, pp.I-5/I-8.

[22] M. R. Khan and T. Hasan, "Iterative noise power subtraction technique for improved speech quality," *Proc. International Conference on Electrical and Computer Engineering (ICECE2008)*, pp.391–394, 2008.

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

[24] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical analysis of iterative weak spectral subtraction via higher-order statistics," *Proc. MLSP2010*, pp.220–225, 2010.

PLACE PHOTO HERE

Takayuki Inoue was born in Shimane, Japan, on November 17, 1985. He received the B.E. degree in information engineering from Osaka University, Osaka, Japan, in 2009. He is currently a M.E. candidate of Nara Institute of Science and Technology. His research interests include noise reduction and nonlinear signal processing. He is a member of the Acoustical Society of Japan.

PLACE PHOTO HERE

Hiroshi Saruwatari (M'00) was born in Nagoya, Japan, on July 27, 1967. He received the B.E., M.E., and Ph.D., degrees in 1991, 1993, and 2000, respectively. He joined Intelligent System Laboratory, SECOM Co., Ltd., Tokyo, Japan, in 1993, where he engaged in the research on the ultrasonic array system for the acoustic imaging. He is currently an Associate Professor of Graduate School of Information Science, Nara Institute of Science and Technology. His research interests include noise reduction, array signal processing, blind source separation, and sound field reproduction. He received paper awards from IEICE in 2001 and 2006, from Telecommunications Advancement Foundation in 2004 and 2009, and from IEEE-IROS2005 in 2006. He won the first prize in IEEE MLSP2007 Data Analysis Competition for BSS. Prof. Saruwatari is a member of the IEICE, Japan VR Society, and the Acoustical Society of Japan.

PLACE PHOTO HERE

Yu Takahashi (S'07) was born in Kagoshima, Japan, on August 31, 1982. He received the B.E. degree in information engineering from Himeji Institute of Technology in 2005. He received the M.E. and Ph.D degrees in information science from Nara Institute of Science and Technology in 2007 and 2010, respectively. His research interests include array signal processing and blind source separation. Dr. Takahashi is a member of the Acoustical Society Japan, and a member of the Japanese Society for Artificial Intelligence.

PLACE PHOTO HERE

Kiyohiro Shikano (M'84– F'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980 respectively. He is currently a Professor of Nara Institute of Science and Technology (NAIST), where he is directing speech and acoustics laboratory. His major research areas are speech recognition, multimodal dialog system, speech enhancement, adaptive microphone array, and acoustic field reproduction. Since 1972, he had been working at NTT Laboratries, where he had been engaged in speech recognition research. During 1990–1993, he was the Exective Research Scientist at NTT Human Interface Laboratories, where he supervised the search of speech recognition and speech coding. During 1986–1990, he was the Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories, where he was directing speech recognition and speech synthesis research. Prof. Shikano received the IEEE Signal Processing Society 1990 Senior Award in 1991. He is a member of the IEICE, the IPSJ, the Acoustical Society of Japan, and the Japan VR Society.

PLACE
PHOTO
HERE

Kazunobu Kondo was born in Aichi, Japan, on Jan. 21, 1969. He received the B.E. and M.E., degrees in 1991, and 1993, respectively. He joined Electronics Development Center, Yamaha Co., Ltd., Shizuoka, Japan, in 1993, where he conducted a research and development on coding system for the musical sound sources. He is currently a Program Manager of Corporate Research and Development Center, Yamaha Corporation. His research interests include array signal processing, blind source separation, and noise reduction. Mr. Kondo is a member of the IEICE, and the Acoustical Society of Japan.