THEORETICAL ANALYSIS OF MUSICAL NOISE IN GENERALIZED SPECTRAL SUBTRACTION: WHY SHOULD NOT USE POWER/AMPLITUDE SUBTRACTION?

†*Takayuki Inoue*, †*Yu Takahashi*, †*Hiroshi Saruwatari*, †*Kiyohiro Shikano, and* ‡*Kazunobu Kondo*

†Nara Institute of Science and Technology, Nara, 630-0192 Japan
‡ SP Group, Center for Advanced Sound Technologies, Yamaha Corp., Shizuoka, 438-0192 Japan

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. Therefore, 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 the lower exponent spectral domain; this implies that there is no theoretical justification for using power/amplitude spectral subtraction.

1. INTRODUCTION

Over the past decade, the number of applications of speech communication systems, such as TV conference systems and mobile phones, has increased. These systems, however, always suffer from a problem of deterioration of speech quality under adverse noise conditions. Therefore, in speech signal processing, noise reduction is a problem requiring urgent attention.

Spectral subtraction is a commonly used noise reduction method that has high noise reduction performance [1]. However, in this method, artificial distortion, so-called *musical noise*, arises owing to nonlinear signal processing, leading to a serious deterioration of sound quality. Moreover, 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 in which domain the exponent is applied in the spectral subtraction process, e.g., power spectral domain [2], amplitude spectral domain [1], or others [3, 4, 5, 6]. We investigated in which domain 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 studies on the theoretical 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, one of the authors has 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 [7, 8, 9]. 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 for 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.

Therefore, 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. 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. This is the world's first mathematical leap in the analysis as far as we know.

This work was partly supported by MIC Strategic Information and Communications R&D Promotion Programme in Japan.

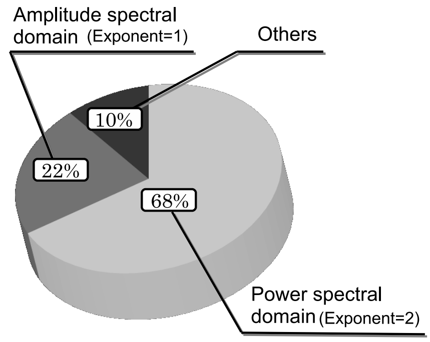

Figure 1: Value of exponent used in conventional spectral subtraction methods. This investigation was conducted via Google Scholar by surveying 50 highly ranked articles retrieved by the keywords "spectral subtraction."

2. RELATED WORKS

2.1 Formulation of Generalized Spectral Subtraction

We apply short-time Fourier analysis to the observed signal which is a mixture of target speech and noise, and then obtain the time-frequency signal. We formulate *generalized spectral subtraction* [3, 4] in the time-frequency domain as follows:

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

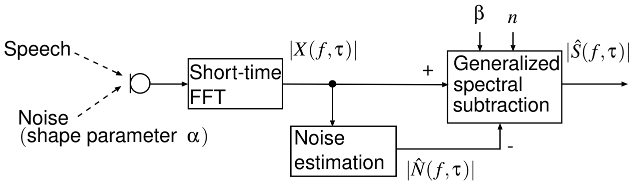

Figure 2: Block diagram of generalized spectral subtraction.

where $\hat{S}(f,\tau)$ 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.

2.2 Mathematical Metric of Musical Noise Generation via Higher-Order Statistics [7]

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

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

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

$$\mu_m = \int_0^\infty x^m P(x)dx, \tag{3}$$

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, (2) is not kurtosis in the mathematically strict definition but a modified version; we still refer to (2) 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, 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 [7]. This measure is defined as

$$\text{kurtosis ratio} = \frac{\mathrm{kurt}_{\mathrm{proc}}}{\mathrm{kurt}_{\mathrm{org}}}, \tag{4}$$

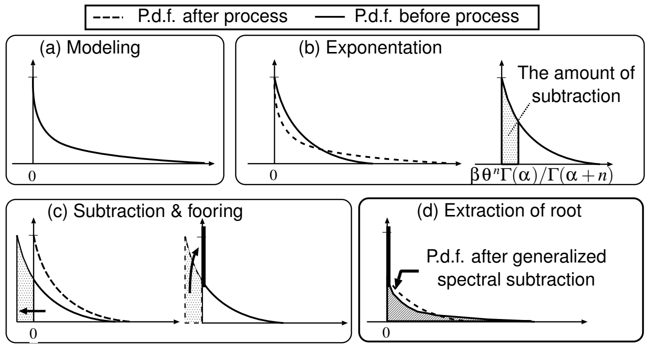

Figure 3: Process of deforming p.d.f.

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

3. THEORETICAL ANALYSIS OF GENERALIZED SPECTRAL SUBTRACTION

3.1 Analysis Strategy

In this section, we analyze the amount of noise reduction and musical noise generated through generalized spectral subtraction 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. 3.2). 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. 3.3). Finally, we compare the kurtosis values upon changing the exponent parameter (n in (1)) under the same amount of noise reduction (see Sect. 3.4).

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

3.2.1 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 [10]

$$P(x) = \frac{x^{\alpha-1}\exp(-\frac{x}{\theta})}{\Gamma(\alpha)\theta^\alpha}, \tag{5}$$

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

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt. \tag{6}$$

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

3.2.2 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

$$P(y) = P(g^{-1}(y))|J|, \tag{7}$$

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

$$|J| = \left|\frac{\partial g^{-1}}{\partial y}\right|. \tag{8}$$

We apply (7) to (5). Since x is the power spectral domain signal, y is expressed as $y = x^n$, i.e., the Jacobian is

$$|J| = \left|\frac{\partial x}{\partial y}\right| = \left|\frac{1}{nx^{n-1}}\right| = \left|\frac{1}{ny^{(n-1)/n}}\right|. \tag{9}$$

Consequently,

$$P(y) = P(x)|J| = \frac{y^{\alpha/n-1}\exp(-\frac{y^{1/n}}{\theta})}{n\Gamma(\alpha)\theta^{\alpha}}. \tag{10}$$

3.2.3 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, $\mathrm{E}[y]$. $\mathrm{E}[y]$ is given by

$$\mathrm{E}[y] = \int_0^{\infty} yP(y) = \int_0^{\infty} \frac{y^{\alpha/n}\exp(-\frac{y^{1/n}}{\theta})}{n\Gamma(\alpha)\theta^{\alpha}} dy. \tag{11}$$

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,

$$\mathrm{E}[y] = \frac{\theta^n}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha+n-1}\exp(-t)dt, \tag{12}$$

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

$$\mathrm{E}[y] = \frac{\theta^n\Gamma(\alpha+n)}{\Gamma(\alpha)}. \tag{13}$$

As a result of the subtraction process, the p.d.f. in the exponent spectral domain undergoes a lateral shift 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

$$P_{\mathrm{gss}}(y) = \begin{cases} \frac{1}{n\theta^{\alpha}\Gamma(\alpha)}(y+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{\alpha/n-1} \\ \exp\left(-\frac{(y+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}}{\theta}\right) & (y>0), \\ \frac{1}{n\theta^{\alpha}\Gamma(\alpha)}\int_0^{\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha)} y^{\alpha/n-1}\exp(-\frac{y^{1/n}}{\theta}) & (y=0). \end{cases} \tag{14}$$

3.2.4 Extraction of nth Root

We apply the extraction of the nth root to $P_{\mathrm{gss}}(y)$, given by (14), and reconstruct the p.d.f. in the power spectral domain, $P_{\mathrm{gss}}(x)$. In a similar way to in Sect. 3.2.2, we let $x = y^{1/n}$ and apply a change of variables, where the Jacobian is

$$|J| = |\frac{\partial y}{\partial x}| = \frac{n}{y^{(1-n)/n}} = \frac{n}{x^{1-n}}. \tag{15}$$

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

$$P_{\mathrm{gss}}(x) = P_{\mathrm{gss}}(y)|J| = \begin{cases} \frac{1}{\theta^{\alpha}\Gamma(\alpha)}x^{n-1}(x^n+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{\alpha/n-1} \\ \exp\left(-\frac{(x^n+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}}{\theta}\right) & (x>0), \\ \frac{1}{\theta^{\alpha}\Gamma(\alpha)}\int_0^{\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha)} x^{\alpha-1}\exp(-\frac{x}{\theta}) & (x=0). \end{cases} \tag{16}$$

3.3 Estimation of Amount of Musical Noise and Noise Reduction

3.3.1 The mth-order moment of $P_{\mathrm{gss}}(x)$

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

$$\begin{aligned}\mu_m &= \int_0^{\infty} x^m P_{\mathrm{gss}}(x)dx \\ &= \frac{1}{\theta^{\alpha}\Gamma(\alpha)}\int_0^{\infty} x^{m+n-1}(x^n+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{\alpha/n-1} \\ &\quad \exp\left(-\frac{(x^n+\beta\theta^n\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}}{\theta}\right)dx.\end{aligned} \tag{17}$$

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

$$\mu_m = \frac{\theta^m}{\Gamma(\alpha)}\int_{\left\{\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{1/n}}^{\infty}\left\{t^n-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{m/n} t^{\alpha-1}\exp(-t)dt. \tag{18}$$

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 (18) as

$$\begin{aligned}&\left\{t^n-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{m/n} \\ &= \sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l \frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)} t^{n(m/n-l)}.\end{aligned} \tag{19}$$

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

$$\begin{aligned}\mu_m &= \frac{\theta^m}{\Gamma(\alpha)}\sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l\frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)} \\ &\quad \int_{\left\{\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{1/n}}^{\infty} t^{\alpha+m-ln-1}\exp(-t)dt \\ &= \frac{\theta^m}{\Gamma(\alpha)}\sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^l\frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)} \\ &\quad \Gamma(\alpha+m-ln,(\beta\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}),\end{aligned} \tag{20}$$

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

$$\Gamma(\alpha,z) = \int_z^{\infty} t^{\alpha-1}\exp(-t)dt. \tag{21}$$

3.3.2 Analysis of Amount of Musical Noise

Using (20), we can obtain the kurtosis after generalized spectral subtraction as

$$\mathrm{kurt}_{\mathrm{gss}} = \frac{\mu_4}{\mu_2^2} = \Gamma(\alpha)\frac{\mathscr{M}(\alpha,\beta,4/n)}{\mathscr{M}^2(\alpha,\beta,2/n)}, \tag{22}$$

where

$$\mathscr{M}(\alpha,\beta,m/n)=\sum_{l=0}^{m/n}\left\{-\beta\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right\}^{l}\frac{\Gamma(m/n+1)}{\Gamma(l+1)\Gamma(m/n-l+1)}$$
$$\Gamma(\alpha+m-ln,(\beta\Gamma(\alpha+n)/\Gamma(\alpha))^{1/n}). \qquad (23)$$

By substituting $\beta=0$ into (22), we can estimate the kurtosis before processing. Thus, we can calculate the resultant kurtosis ratio as

$$\text{kurtosis ratio}=\frac{\mathscr{M}(\alpha,\beta,4/n)/\mathscr{M}^2(\alpha,\beta,2/n)}{\mathscr{M}(\alpha,0,4/n)/\mathscr{M}^2(\alpha,0,2/n)}. \qquad (24)$$

3.3.3 *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 signal-to-noise ratio (SNR) in dB minus the input SNR in dB [11]. The NRR is

$$\mathrm{NRR}=10\log_{10}\frac{\mathrm{E}[s_{\mathrm{out}}^2]/\mathrm{E}[n_{\mathrm{out}}^2]}{\mathrm{E}[s_{\mathrm{in}}^2]/\mathrm{E}[n_{\mathrm{in}}^2]}, \qquad (25)$$

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 (25) 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

$$\mathrm{NRR}=10\log_{10}\frac{\mathrm{E}[n_{\mathrm{in}}^2]}{\mathrm{E}[n_{\mathrm{out}}^2]}. \qquad (26)$$

Since, $\mathrm{E}[n_{\mathrm{in}}^2]=\mu_1$ when $\beta=0$ in (20) and $\mathrm{E}[n_{\mathrm{out}}^2]=\mu_1$ for a the specific (nonzero) β,

$$\mathrm{NRR}=10\log_{10}\frac{\mathscr{M}(\alpha,0,1/n)}{\mathscr{M}(\alpha,\beta,1/n)}. \qquad (27)$$

In summary, we can derive theoretical estimates for the amount of musical noise and NRR using (24) and (27). This greatly simplifies the analysis because both equations are expressed analytically in a form that does not include any integrals.

3.4 Comparison of Amount of Musical Noise under Same NRR Condition

According to the previous analysis, we can compare the amount of musical noise between different exponent parameters under the same amount of noise reduction. Figure 4 shows the theoretical behavior of the kurtosis ratio and NRR for various parameter values. In this figure, the shape parameter α is set to 0.1 and 1.0, NRR is varied from 0 to 12 dB, and the exponent parameter n is set to 1.0, 0.5, 0.25, and 0.125, where the oversubtraction parameter β is 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 [7]. We call this the *log kurtosis ratio* hereafter.

Figure 4 shows that a small amount of musical noise is generated when a the lower exponent parameter is used, regardless of the type of noise and NRR. This figure also indicates that for higher values of NRR, there is a larger difference of between the kurtosis ratio for different values of the exponent parameter. This implies that humans perceive a greater variation at a higher NRR. In addition, it is revealed that this variation is less perceptible for super-Gaussian noise.

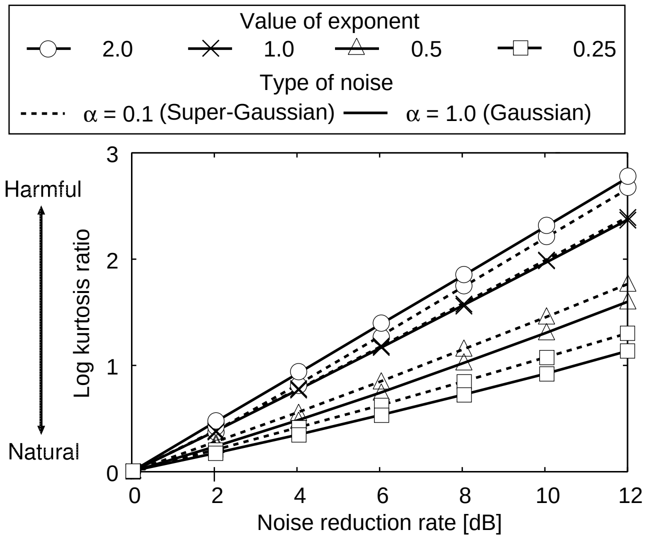

Figure 4: Relation between NRR and log kurtosis ratio.

Table 1: Conditions of evaluation

NRR [dB]	4, 8, 12
Value of exponent	2.0, 1.0, 0.5, 0.25
Objective evaluation measure	(1) log kurtosis ratio (2) cepstral distortion
Subjective evaluation measure	preference score of 7 examinees

4. EVALUATION EXPERIMENT AND RESULT

4.1 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 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 is 1024, and the frame shift length is 256. The shape parameter of the white Gaussian noise was 0.96 and that of the speech noise was 0.21. We conducted our experiments regarding 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 1.

4.2 Objective Evaluation

We first conducted an objective experiment and evaluated the sound quality of processed signals on the basis of cepstral distortion 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 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. 5. The figure shows that the log kurtosis ratio decreases as the exponent parameter becomes smaller and that 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 theoretical analysis provided in Sect. 3.4. In addition, cepstral distortion

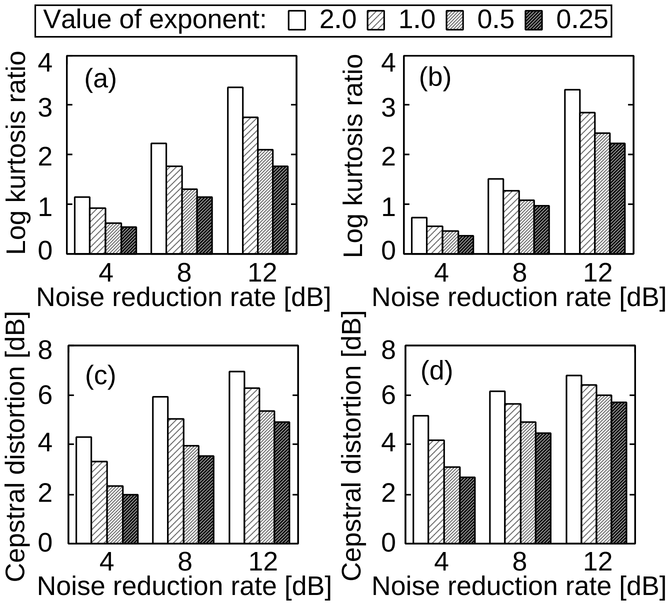

Figure 5: Results of log kurtosis ratio and cepstral distortion for various domain values of the exponent. (a) and (c) are for white Gaussian noise, and (b) and (d) are for speech noise.

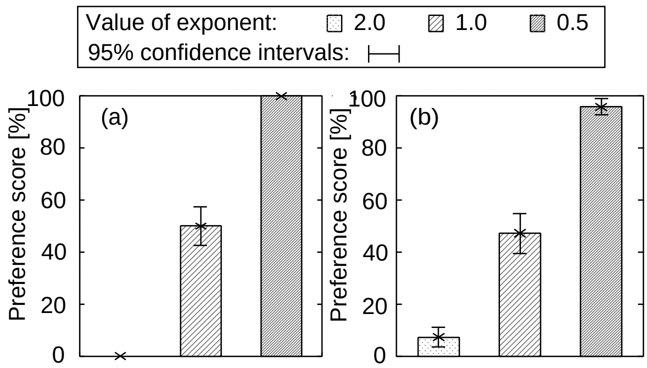

Figure 6: Subjective evaluation results for (a) white Gaussian noise, and (b) speech noise.

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.

4.3 Subjective Evaluation

We next conducted a subjective evaluation. In the evaluation, we presented three equi-NRR signals processed by the power-, amplitude-, and root-domain spectral subtraction in random order to 7 examinees, who selected which signal they considered to contain least musical noise.

The result of the experiment is shown in Fig. 6. It was found that musical noise is less perceptible when a lower exponent parameter is used. This result is also consistent with our theoretical analysis, thus confirming the validity of the proposed method of theoretical analysis.

4.4 Remark

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. 5. 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/post-processing to mitigate musical noise.

5. CONCLUSION

In this study, we performed a theoretical analysis of the amount of musical noise generated via generalized spectral subtraction based on higher-order statistics. Also, we conducted objective and subjective comparisons of the amount of musical noise for distinct exponent spectral domains under the same noise reduction performance. It was clarified from mathematical analysis and evaluation experiments that in a spectral domain with a lower exponent, less musical noise is generated.

REFERENCES

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

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

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

[4] 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.

[5] 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.

[6] 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.

[7] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, 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.

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

[9] Y. Takahashi, Y. Uemura, H. Saruwatari, K. Shikano, K. Kondo, “Musical noise analysis based on higher order statistics for microphone array and nonlinear signal processing,” *Proc. ICASSP2009*, pp.229–232, 2009.

[10] 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.

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