AUTOMATIC OPTIMIZATION SCHEME OF SPECTRAL SUBTRACTION BASED ON MUSICAL NOISE ASSESSMENT VIA HIGHER-ORDER STATISTICS

Yoshihisa Uemura, Yu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano, Kazunobu Kondo,

Nara Institute of Science and Technology, Nara, 630-0192, JAPAN
SA Group, Center for Advanced Sound Technologies, YAMAHA Corp, Shizuoka, 438-0192, JAPAN

ABSTRACT

In this paper, we propose a new optimization scheme of the strength of spectral subtraction based on musical noise assessment via higher-order statistics. Spectral subtraction often generates artificial distortion (so-called musical noise), and in this paper, we focus on such musical noise. Musical noise is related to the artificial tonal components in remnant noise. Thus, first, we propose a criterion to measure the generated tonal components based on noise's kurtosis. This criterion enables us to quantify the amount of generated musical noise. Next, we proposed a new spectral-subtraction-control scheme based on the proposed criterion. Finally, we confirm the advantages of the proposed scheme with subjective evaluation.

Index Terms— Musical noise, Spectral subtraction, Higher-order statistics, Kurtosis, Noise suppression

1. INTRODUCTION

In recent years, voice communication systems, e.g., TV conference systems or mobile phones, are used in various situations. Noise suppression techniques are indispensable for the system because noise often disturbs the smooth communication among users. Thus, a method that can reduce the noise while maintaining sound quality is required. Moreover, the method should be robust against the variation of noise environments.

Spectral subtraction is a single channel noise suppression method [1]. It is one of the most popular noise reduction techniques owing to its simple algorithm and good noise suppression performance. Although spectral subtraction is used in many noise suppression systems [2, 3], it has a critical and inherent problem. The problem is that sound quality degrades due to musical noise generated via nonlinear subtraction procedures. Musical noise consists of tonal remnant noise components [4], which are significantly disagreeable to the ear. There are many countermeasures and improved methods [5] against musical noise because it has been regarded as the biggest problem of spectral subtraction. However, we have no general measure of the amount of musical noise. Moreover, it is well known that the degree of musical noise varies by noise environment; this leads to difficulty of parameter settings in spectral subtraction. Therefore, at first, we construct a new mathematical metric of the amount of musical noise based on higher-order statistics. Secondly, we propose a new scheme to automatically determine the strength of spectral subtraction processing with the proposed measurement. Owing to these propositions, we can measure and control the quality of the processed signal, and optimize the strength of spectral subtraction processing. Finally, we have subjective evaluation to show the effectiveness of this approach.

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

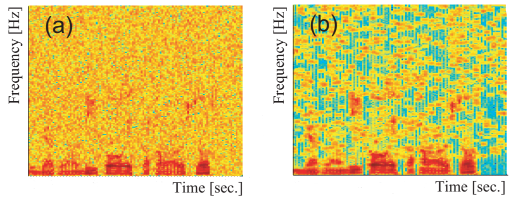

Fig. 1. (a) Observed signal spectrogram, (b) Processed signal spectrogram.

2. STATISTICAL MEASUREMENT OF MUSICAL NOISE

2.1. Spectral subtraction

At first, we analyze spectral subtraction processing to build up a measure that indicates amount of musical noise. Although various types of spectral subtraction methods are proposed, we address single-channel spectral subtraction in the power domain, which is used for any speech enhancement [6].

Let the corrupted speech signal $o(t)$ be represented as

$$o(t) = s(t) + n(t), \tag{1}$$

where $s(t)$ is a clean speech signal and $n(t)$ is a noise signal. This processing is conducted on a frame-by-frame basis. The short-time Fourier transform (STFT) is used and the previous model can be rewritten as

$$O(k, m) = S(k, m) + N(k, m), \tag{2}$$

where k denotes the frequency subband and m is the frame index. In spectral subtraction, noise reduction is achieved by subtracting the power spectrum of the estimated noise from the power spectrum of the noisy observation. This procedure is given by

$$Y(k, m) = \sqrt{|O(k, m)|^2 - \beta \cdot \mathrm{E}_m\left[|N(k, m)|^2\right]} \cdot e^{j\ \arg(O(k,m))}, \tag{3}$$

where $Y(k, m)$ is an estimated speech signal, β is a subtraction coefficient and $\mathrm{E}[\cdot]$ is an expectation operator of $\cdot$ with respect to m.

2.2. Relationship between kurtosis and musical noise

Spectral subtraction often generates isolated power spectral components (see Fig. 1 (a) and (b)). In this paper, we define the musical noise as the generated audible isolated spectral components through processing. Thus we speculate that the amount of musical noise is highly related to the number of isolated components and the isolated level of them.

Hence, secondly, we adopt kurtosis to quantify the isolated spectral components, and focus on the kurtosis changes. Since isolated spectral components have relatively sufficient power, we would hear them as *tonal sound*, which results in musical noise. Therefore, it is expected that the measurement of the amount of prominence of tonal components enables us to measure or quantify the amount of musical noise. However, such a measurement is extremely complicated, so instead we introduce a simple statistical estimate, i.e., kurtosis.

This adoption allows us to obtain the characteristics about tonal components. The adopted kurtosis can evaluate the width of the probability density function (p.d.f.) and the weight of their skirts. We could say that kurtosis can evaluate the percentage of tonal components in total components. Bigger value indicates a signal with heavy skirt; it means that a signal has a lot of tonal components. Also kurtosis is calculated in concise algebraic form. Thus, kurtosis is a suitable measure for the tonal components in computers.

Kurtosis is defined as

$$kurt = \frac{\mu_4}{{\mu_2}^2}, \tag{4}$$

where *kurt* denotes kurtosis and μ_n is the n-th order moment which is given by

$$\mu_n = \int_0^\infty x^n P(x)\, dx. \tag{5}$$

Here, $P(x)$ is p.d.f. of the signal. We consider the spectral subtraction in power spectral domain, so the integral range is only positive.

Although we can measure the number of the tonal components by kurtosis, note that kurtosis itself is not enough to measure the musical noise. This is obvious in that kurtosis of some unprocessed signals, e.g., speech signals, is also high, but we don't recognize speech as musical noise. Since we want to check only the musical noise components, it should not consider genuine tonal components. In order to address the above-mentioned aim, we focus on the fact that musical noise is generated only in artificial signal processing. Hence, we turn our attention to *change* of kurtosis between before/after signal processing.

3. RESULTANT KURTOSIS IN SPECTRAL SUBTRACTION

We derive the relationship between kurtosis and the strength of spectral subtraction. Moreover, the relationship between kurtosis of processed signal and kurtosis of unprocessed signal are revealed.

3.1. Gamma distribution modeling

We utilize the gamma distribution as a model of speech or noise signal [7, 8]. The gamma distribution have a lot of useful mathematically attributes which are derived from the gamma function.

The p.d.f. of the gamma distribution is written as

$$P(x) = \frac{1}{\Gamma(\alpha)\, \theta^{\alpha}} \cdot x^{\alpha-1}\, e^{-\frac{x}{\theta}}, \tag{6}$$

where $x \geq 0$, $\alpha > 0$ and $\theta > 0$. Also α denotes the shape parameter and θ is the scale parameter. The Gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \cdot dx. \tag{7}$$

Note that the gamma function has a famous functional equation as follows;

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) = (\alpha-1)(\alpha-2)\Gamma(\alpha-2) = (\alpha-1)\cdots(\alpha-j)\Gamma(\alpha-j). \tag{8}$$

Hereafter, in this paper, let $C = 1/[\Gamma(\alpha)\, \theta^{\alpha}]$. If $\alpha = 1$, this is the exponential distribution. It is well known that the average of the gamma distribution is given by

$$\mathrm{E}\,[P(x)] = \alpha\theta, \tag{9}$$

where $\mathrm{E}[\cdot]$ is an expectation operator. The gamma distribution modeling is the estimation of the shape and the scale parameters from the raw input signal. In this paper, we use the maximum likelihood estimation method for estimating two parameters α and θ, as follows,

$$\hat{\alpha} = \frac{3 - \gamma + \sqrt{(\gamma-3)^2 + 24\gamma}}{12\gamma}, \tag{10}$$

$$\hat{\theta} = \frac{\mathrm{E}\,[\,x\,]}{\hat{\alpha}}, \tag{11}$$

where $\gamma = \log(\,\mathrm{E}\,[x]\,) - \mathrm{E}\,[\,\log x\,]$ (see Refs. [9, 10]).

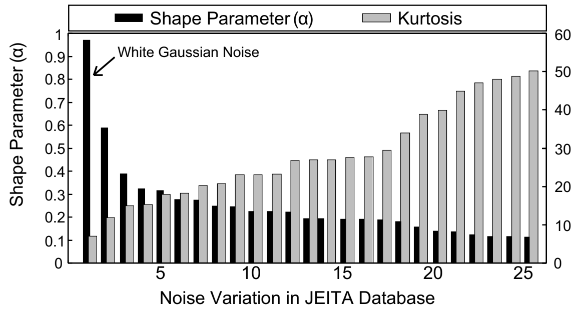

Fig. 2. Shape parameter and kurtosis of the real-world signals.

3.2. Kurtosis of modeling signal

In this way of modeling by the gamma distribution, kurtosis is determined by the shape and the scale parameters as below. At first, we represent the 4th-order moment as

$$\mu_4 = \int_0^\infty x^4 P(x)\, dx = \int_0^\infty x^4 \left[C \cdot x^{\alpha-1} e^{-\frac{x}{\theta}}\right] dx.$$

Here, let $X = x/\theta$, this moment can be rewritten as

$$\mu_4 = C \cdot \theta^{\alpha+4} \int_0^\infty X^{(\alpha+4)-1} e^{-X} dX = C \cdot \theta^{\alpha+4} \cdot \Gamma(\alpha+4). \tag{12}$$

Next, by the same manner, 2nd-order moment can be designated as

$$\mu_2 = \int_0^\infty x^2 P(x)\, dx = C \cdot \theta^{\alpha+2} \cdot \Gamma(\alpha+2). \tag{13}$$

Consequently, using (8), we can obtain kurtosis of the modeled raw signal by the gamma distribution as follows.

$$\begin{aligned} kurt_{org} &= \frac{\mu_4}{\mu_2^2} = \frac{C\theta^{\alpha+4} \cdot \Gamma(\alpha+4)}{[C\theta^{\alpha+2} \cdot \Gamma(\alpha+2)]^2} = \frac{(\alpha+2)(\alpha+3) \cdot \Gamma(\alpha+2)}{C\theta^{\alpha} \cdot \Gamma(\alpha+2)^2} \\ &= \frac{[\Gamma(\alpha)\theta^{\alpha}]\,(\alpha+2)(\alpha+3)}{\theta^{\alpha}\alpha(\alpha+1) \cdot \Gamma(\alpha)} = \frac{(\alpha+2)(\alpha+3)}{\alpha(\alpha+1)}. \end{aligned} \tag{14}$$

Figure 2 indicates the values of the shape parameter and kurtosis of the real-world noise, in the real-world noise database of Japan Electronics and Information Technology Industries Association (JEITA). This figure shows the shape parameter is about 0.1 to 0.6, and kurtosis is about 10 to 50 in the real-world noise.

3.3. Logarithmic kurtosis ratio

In this paper, we propose a new metric of logarithmic kurtosis ratio (log kurtosis ratio) as a measure of the amount of generated musical noise. We use the average of observed signal power spectrum as estimated noise power spectrum, so the amount of subtraction is $\beta \cdot \alpha\theta$, using (9). To subtract the estimated noise power spectrum in each frequency band can be regarded as deforming of the p.d.f., which is the parallel translation of the p.d.f. to zero power direction. As a result, the probability of the negative power component arises. To avoid this, such a negative component probability is replaced to zero (so-called flooring technique). The resultant p.d.f. of the processed signal is written as

$$P(x) = \begin{cases} C \cdot (x + \beta \cdot \alpha\theta)^{\alpha-1}\, e^{-\frac{x+\beta\cdot\alpha\theta}{\theta}} & (x > 0)\,, \\ C \int_0^{\beta\cdot\alpha\theta} x^{\alpha-1} e^{-\frac{x}{\theta}} dx & (x = 0)\,. \end{cases} \tag{15}$$

Here we can approximate $(x+\beta\alpha\theta)^{\alpha-1}$ in (15) by Taylor expansion as

$$(x+\beta\alpha\theta)^{\alpha-1} \approx x^{\alpha-1} + \beta\alpha\theta(\alpha-1)x^{\alpha-2} + \frac{(\beta\alpha\theta)^2}{2}(\alpha-1)(\alpha-2)x^{\alpha-3}. \tag{16}$$

Using (15) and (16), we have

$$\begin{aligned} \mu_4 \approx C e^{-\alpha\beta} \Big[&\int_0^\infty x^{(\alpha+4)-1} e^{-\frac{x}{\theta}} dx + \beta\alpha\theta(\alpha-1) \int_0^\infty x^{(\alpha+3)-1} e^{-\frac{x}{\theta}} dx \\ &+ \frac{(\beta\alpha\theta)^2}{2}(\alpha-2)(\alpha-1) \int_0^\infty x^{(\alpha+2)-1} e^{-\frac{x}{\theta}} dx \Big]. \end{aligned} \tag{17}$$

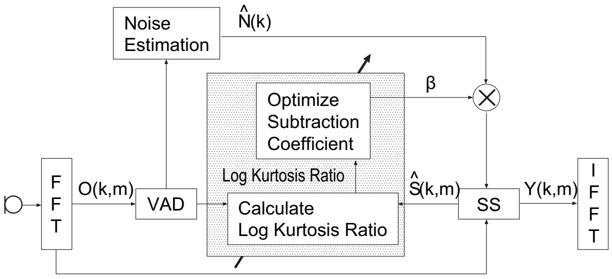

Fig. 3. Block diagram of proposed scheme.

Here we normalize the expectation of the gamma distribution with using (9), $\theta = 1/\alpha$. Also, with (12), the 4th-order moment can be represented by

$$\mu_4 \approx Ce^{-\alpha\beta}\Big[\theta^{\alpha+4}\Gamma(\alpha+4) + \beta(\alpha-1)\theta^{\alpha+3}\Gamma(\alpha+3) + \frac{\beta^2}{2}(\alpha-2)(\alpha-1)\theta^{\alpha+2}\Gamma(\alpha+2)\Big]. \quad (18)$$

Using the first term of the right-hand side in (16), maximum value of the 2nd-order moment is estimated as below,

$$\mu_2 = \int_0^{\infty} x^2 \left[C(x+\beta\cdot\alpha\theta)^{\alpha-1} e^{-\frac{x+\beta\alpha\theta}{\theta}}\right] dx \le Ce^{-\alpha\beta}\theta^{\alpha+2}\Gamma(\alpha+2). \quad (19)$$

Thus, kurtosis of the processed signal can be given as

$$kurt_{ss} \ge \frac{e^{\alpha\beta}}{\alpha(\alpha+1)}\left\{(\alpha+2)(\alpha+3)+\beta\alpha(\alpha+2)(\alpha-1)+\frac{(\beta\alpha)^2}{2}(\alpha-3)(\alpha-1)\right\}. \quad (20)$$

From (20), the kurtosis depends on only α and β, and its behavior is exponential. Therefore, it is considered reasonable and proper introducing the following log kurtosis ratio:

$$\log\left[\frac{kurt_{ss}}{kurt_{org}}\right] = \alpha\cdot\beta + \log\left\{1 + \frac{\beta\alpha(\alpha-1)}{(\alpha+3)} + \frac{(\beta\alpha)^2(\alpha-2)(\alpha-1)}{2\,(\alpha+2)(\alpha+3)}\right\}. \quad (21)$$

Note that log kurtosis ratio is a monotonically increasing function of α and β. Thus, it depends on the distribution of the original signal and the strength of spectral subtraction. Also log kurtosis ratio is confirmed by our experience, i.e., in spectral subtraction, using the smaller kurtosis signal (e.g., white Gaussian) set up more musical noise than using the bigger kurtosis signal (e.g., speech signal). These prospective log-kurtosis-ratio behaviors are confirmed by experiment in Sect. 5.

4. NEW SCHEME TO CONTROL AMOUNT OF GENERATED MUSICAL NOISE BASED ON LOG KURTOSIS RATIO

In this section, we propose a new scheme to control the amount of generated musical noise for any noise with a unified criterion that is log kurtosis ratio. This control is achieved by determining the proper subtraction coefficient automatically based on the criterion. In the conventional method, we have to experimentally select the proper subtraction coefficient about the amount of generated musical noise. This is very heuristic and not versatile, e.g., different researchers would select different subtraction coefficients because of the absence of general criteria about musical noise.

The processing flow of the proposed method can be schematized in Fig. 3, and the procedures are described as follows.

[step 1] At first, we set up the musical noise score as required level of maximum amount of generated musical noise. Also initial subtraction coefficient is set to a small value (e.g. we use 0.1).

[step 2] Next, input signal is processed as a trial.

[step 3] We calculate log kurtosis ratio using unprocessed and processed signals kurtosis.

[step 4] Finally, check the log kurtosis ratio, which is calculated back from the musical noise score and actual measurement value of the log kurtosis ratio. If actual value is less than the calculated-back one, we go back in **[step 2]** with turning the subtraction coefficient up (e.g. add 0.1).

The proposed scheme can maximize the subtraction coefficient automatically while keeping the requested noise quality, i.e., less musical noise. Also this scheme allows us to optimize β for each noise environment.

Table 1. Subjective evaluation conditions

Sound	Length	10 s.
	Noise	4-type noise from JEITA
	Speech	4 sentences from Japanese News Article Sentences
Spectral subtraction set up	Noise estimation	Pause interval average power spectrum
	Subtraction coefficient	{0, 0.4, 0.8, 1.2, 1.6, 2.0}
	Flooring	0
Evaluate	Test type	Open
	Criterion	Musical noise score
	Examinees	8 males and 1 female

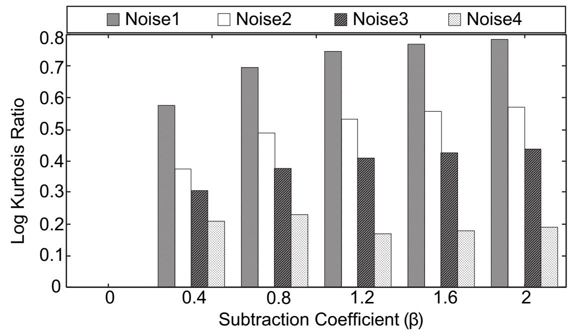

Fig. 4. Relationship between log kurtosis ratio and subtraction coefficient.

5. SUBJECTIVE EVALUATION

5.1. Experimental condition

We conduct a subjective evaluation to bring out the relationship between log kurtosis ratio and the amount of generated musical noise, and we confirm the efficacy of the proposed scheme. In this experiment, first, let examinees rate the amount of generated musical noise in five steps (0: natural, 1, $\cdots$, 4: harmful). We call this rating musical noise score. If examinees feel the stimulus sound as more musical-noise-like, a bigger value is rated to the musical noise score. Now examinees have heard known musical noise as a reference in advance. The evaluation conditions are listed in Table 1. In addition, signals used in this experiment cover possible kurtosis in the real-world, and index in ascending sequence Noise1, $\cdots$, Noise4 (c.f., Noise1: station noise, Noise2: crowd noise, Noise3: exhibition hall noise, Noise4: elevator hall noise). We depict the relationship subtraction coefficient and log kurtosis ratio in Fig. 4. From this figure, we can see that log kurtosis ratio is increasing with increase of subtraction coefficient. Note that log kurtosis ratio varies for every noise type even in the same subtraction coefficient. This means that log kurtosis ratio accurately depends on noise type, and this result is consistent with the theorem mentioned in Sect. 3.

5.2. Evaluation result

The result of the subjective evaluation is illustrated in Figs. 5, 6 and 7. Figure 5 shows the interesting various relationships. In each subtraction coefficient, the signal that has bigger kurtosis (e.g., Noise4) is rated in small musical noise score. As well, smaller kurtosis signal (e.g., Noise1) is rated in large musical noise score. It should

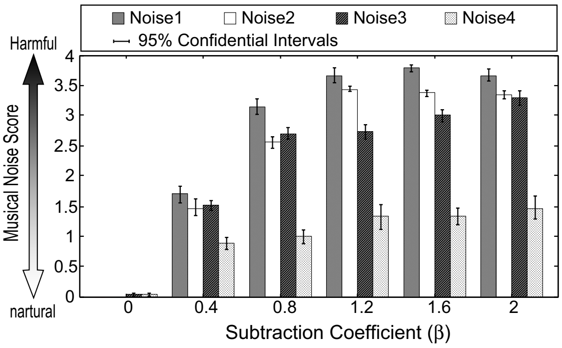

Fig. 5. Results of musical noise score.

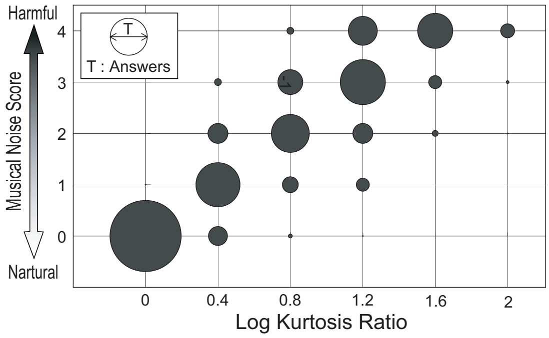

Fig. 6. Relationship between log kurtosis ratio and musical noise score.

be mentioned that musical noise score varies as noise environment (original signal's kurtosis) changes even when we fix the subtraction coefficient; this is a new finding. Compared with Figs. 4 and 5, the results of this experiment mean that log kurtosis ratio is strongly related with musical noise score.

Figures 6 and 7 are the rated scatter plot in which bubble size represents the number of answers. From these results, we can see the strongly correlated relationship between log kurtosis ratio and musical noise score. Actually, the correlation coefficient is 0.84 in Fig. 6. On the other hand, Fig. 7 depicts an appearance of scattering of musical noise score from the view point of subtraction coefficient β. The results show that the fixed subtraction coefficient leads to a widely ranging musical noise score compared with the case of fixed log kurtosis ratio. This variation of musical noise score depends on the noise environment. All of this amounts to saying that the optimal subtraction coefficient from the viewpoint of quality is different for each noise environment. Besides we can determine the relationship $(\textit{musical noise score}) = 2.235 \times (\textit{log kurtosis ratio})$ from regression analysis. Finally, in Fig. 8, we can obtain smaller variance of musical noise score by using the proposed criterion. On the other hand, bigger variance of musical noise score is obtained from the conventional criterion (i.e., subtraction coefficient). The variations of musical noise score are bigger in the conventional criterion, especially on the commonly used value of the subtraction coefficient in practice. For instance, the variation of musical noise score is approximately 1.4 when the subtraction coefficient equals 2. It is to say that the conventional criterion cannot adapt to various noise environments. On the other hand, almost all the variations of musical noise score are smaller in the proposed log kurtosis ratio. Consequently we conclude that the proposed criterion can control the amount of generated musical noise and automatically adapt to the noise environment.

6. CONCLUSION

In conclusion, we proposed a log kurtosis ratio that can measure the amount of generated musical noise and a new criterion to control

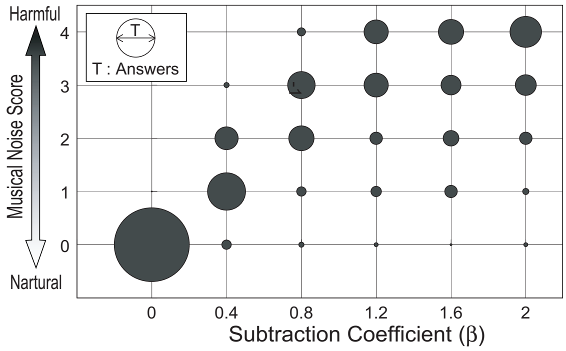

Fig. 7. Relationship between subtraction coefficient and musical noise score.

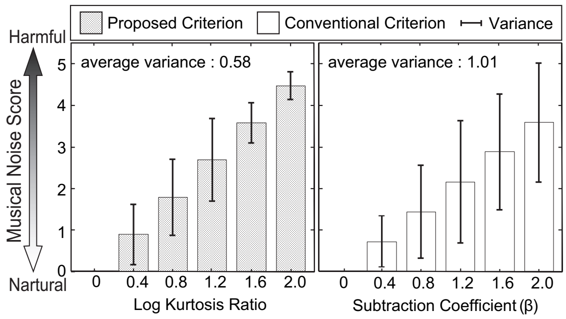

Fig. 8. Musical noise score and the variance of musical noise score of proposed criterion and conventional criterion.

processing strength of spectral subtraction based on the log kurtosis ratio. The proprieties of these propositions are confirmed by subjective evaluation. The both propositions can achieve spectral subtraction suitable for practical use.

7. 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] Y. Takahashi, et al., "Blind spatial subtraction array with independent component analysis for hands-free speech recognition," *IWAENC*, 2006.

[3] R. Mukai, et al., "Removal of residual cross-talk components in blind source separation using time-delayed spectral subtraction," *ICASSP* vol.II, pp.1789–1792, 2002.

[4] S. B. Jebara, "A perceptual approach to reduce musical noise phenomenon with wiener denoising technique," *ICASSP*, vol.III, pp.49–52, 2006.

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

[6] J. Li, et al., "noise reduction based on adaptive β-order generalized spectral subtraction for speech enhancement," *INTERSPEECH* pp,802–805, 2007.

[7] J. W. Shin, et al., "Statistical modeling of speech signal based on generalized gamma distribution," *ICASSP* vol.I, pp.781–784, 2005.

[8] T. H. Dat, et al., "Gamma modeling of speech power and its on-line estimation for statistical speech enhancement," *IEICE Trans. INF & SYST.* vol.E89-D, no.3, 2006.

[9] M. Evans, et al., *Statistical Distributions, 2nd ed.* Wiley. 1993.

[10] W. Q. Meeker and L. A. Escobar, *Statistical methods for reliability data.* Wiley. 1998.