THEORETICAL MUSICAL-NOISE ANALYSIS AND ITS GENERALIZATION FOR METHODS OF INTEGRATING BEAMFORMING AND SPECTRAL SUBTRACTION BASED ON HIGHER-ORDER STATISTICS

†Yu Takahashi, †Hiroshi Saruwatari, †Kiyorhiro 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 conduct a theoretical analysis of the amount of musical noise generated via methods of integrating beamforming and spectral subtraction (SS) based on higher-order statistics under the same noise reduction performance condition. In our previous analysis, we did not consider the effect of flooring technique in SS and the fact that the noise reduction performances of the integration methods are not equivalent. Then, in this study, we analyze the amount of generated musical noise with consideration of such problems. As a result of the analysis, it is clarified that an appropriate structure depends on both the parameters of SS and the statistical characteristics of the input signal. Moreover, it is also revealed that a specific structure is proper to reduce the musical noise for almost all cases.

Index Terms— Musical noise, higher-order statistics, spectral subtraction, acoustic arrays, speech enhancement

1. INTRODUCTION

In these days, integration methods of microphone array signal processing and nonlinear signal processing have been studied for better noise reduction [1, 2, 3]. It is reported that such an integration method can achieve higher noise reduction performance rather than a conventional adaptive microphone array [4], e.g., Griffith-Jim array. However, in such methods, artificial distortion (so-called musical noise) due to nonlinear signal processing arises. Since the artificial distortion makes users uncomfortable, it is desired that we take control of musical noise. However, in almost all the integration methods, the strength of nonlinear signal processing is determined heuristically to mitigate musical noise. Indeed there exists an objective analysis on the target-signal distortion and remained noise components thorough the short-time spectral attenuation techniques [5], but the analysis did not refer to the amount of musical noise generated.

Recently, it is reported that the amount of generated musical noise is strongly related with the difference between higher-order statistics before/after nonlinear signal processing [6]. Based on the fact, the objective metric to measure how much musical noise arises through nonlinear signal processing has been developed [6]. This objective metric would enable us to optimize methods of integrating microphone array signal processing and nonlinear signal processing from the viewpoint of not only noise reduction performance but also the sound quality to human hearing. As a first step toward to achieve the goal, we have analyzed the amount of musical noise generated through two methods of integrating beamforming and spectral subtraction (SS) [7] based on the metric [8]. Figure 1(a) shows a typical architecture example of integration of microphone array signal processing and SS. In this architecture, SS is performed after beamforming. Thus we call this type of architecture *BF+SS*. On the other hand, there exists an alternative architecture illustrated in Figure 1(b). In this architecture, channelwise SS is performed before beamforming. So we call this type of architecture *chSS+BF*.

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

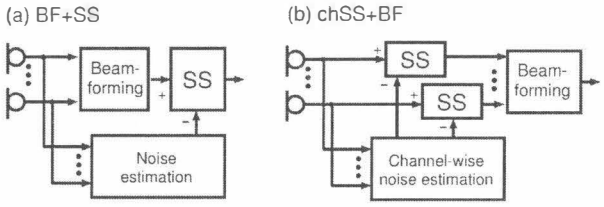

Fig. 1. (a) Block diagram of spectral subtraction after beamforming (BF+SS), and (b) channelwise SS before beamforming (chSS+BF).

As a result of our previous analysis, it has been clarified that chSS+BF can mitigate the amount of musical noise [8]. However, in that analysis, some approximation was introduced to analyze the amount of musical noise generated through SS, and we did not consider the effect of flooring technique in SS. Moreover, we have found that the noise reduction performances of chSS+BF and BF+SS are not equivalent even if the same parameters are set. Therefore we could say that our previous analysis was not a generalized analysis but a limited analysis. Therefore, in this study, the following additional analysis are given:

- The mathematical analysis of the amount of musical noise generated via SS without any approximation unlike Ref. [6],
- The analysis of the effect of flooring technique in SS, and
- The fair comparison of the amount of musical noise via chSS+BF and BF+SS under the same noise reduction performance condition.

Such mathematically strict musical-noise analysis would be an unprecedentedly analysis as far as we know. Although we have concluded that chSS+BF can always mitigate the amount of musical noise in our previous analysis [8], we reveal that BF+SS can reduces the amount of musical noise in the case where some extreme parameter settings in SS as a result of the analysis in this study. However, such cases are significantly rare case, and chSS+BF is an appropriate structure to mitigate musical noise rather than BF+SS for almost all practical cases.

2. INTEGRATING BEAMFORMING AND SS

2.1. Formulation of BF+SS

We consider the following J-channel observed signal in time-frequency domain as

$$\boldsymbol{x}(f,\tau) = \boldsymbol{h}(f)s(f,\tau) + \boldsymbol{n}(f,\tau), \tag{1}$$

where $\boldsymbol{x}(f,\tau) = [x_1(f,\tau), \ldots, x_J(f,\tau)]^{\mathrm{T}}$ is the observed signal vector, $\boldsymbol{h}(f) = [h_1(f), \ldots, h_J(f)]^{\mathrm{T}}$ is the transfer function vector, $s(f,\tau)$ is the target speech, and $\boldsymbol{n}(f,\tau) = [n_1(f,\tau), \ldots, n_J(f,\tau))]^{\mathrm{T}}$ is the noise vector. For enhancing the target speech, delay-and-sum (DS) is firstly applied to the observed signal in BF+SS. This can be represented by

$$y_{\mathrm{DS}}(f,\tau) = \boldsymbol{g}_{\mathrm{DS}}(f,\theta_{\mathrm{U}})^{\mathrm{T}}\boldsymbol{x}(f,\tau), \tag{2}$$

$$\boldsymbol{g}_{\mathrm{DS}}(f,\theta) = [g_1^{(\mathrm{DS})}(f,\theta), \ldots, g_J^{(\mathrm{DS})}(f,\theta)]^{\mathrm{T}}, \tag{3}$$

$$g_j^{(\mathrm{DS})}(f,\theta) = J^{-1} \cdot \exp\left(-i2\pi(f/M)f_s d_j \sin\theta/c\right), \tag{4}$$

where $\boldsymbol{g}_{\mathrm{DS}}(f,\theta)$ is the coefficient vector of DS array, and $\theta_{\theta U}$ is the look direction. Also, f_s is the sampling frequency and d_j ($j = 1, \cdots, J$) is the microphone position. Besides, M is the DFT size, and c is the sound velocity. Finally, we obtain the target-speech-enhanced spectral amplitude based on SS. This procedure can be given as

$$|y_{\mathrm{SS}}(f,\tau)| = \begin{cases} \sqrt{|y_{\mathrm{DS}}(f,\tau)|^2 - \beta \cdot |\hat{n}(f)|^2} \\ \quad (\text{where } |y_{\mathrm{DS}}(f,\tau)|^2 - \beta \cdot |\hat{n}(f)|^2 \geq 0), \\ \eta \cdot |y_{\mathrm{DS}}(f,\tau)| \quad (\text{otherwise}), \end{cases} \tag{5}$$

where $y_{\mathrm{SS}}(f,\tau)$ is the enhanced target speech signal, β is the subtraction coefficient, η is the flooring coefficient, and $\hat{n}(f)$ is the estimated noise signal. $\hat{n}(f,\tau)$ is ordinarily estimated by some beamforming techniques, e.g., fixed or adaptive beamforming [2].

2.2. Formulation of chSS+BF

In chSS+BF, we first perform SS independently in each input channel then we derive a multichannel target-speech-enhanced signal by channelwise SS. This can be designated as

$$|y_j^{(\mathrm{chSS})}(f,\tau)| = \begin{cases} \sqrt{|x_j(f,\tau)|^2 - \beta \cdot \mathrm{E}_\tau[|\bar{n}_j(f,\tau)|^2]} \\ \quad (\text{where } |x_j(f,\tau)|^2 - \beta \cdot \mathrm{E}_\tau[|\bar{n}_j(f,\tau)|^2] \geq 0), \\ \eta \cdot |x_j(f,\tau)| \quad (\text{otherwise}), \end{cases} \tag{6}$$

where $y_j^{(\mathrm{chSS})}(f,\tau)$ is the target-speech-enhanced signal obtained by SS at a specific channel j and $\bar{n}_j(f,\tau)$ is the estimated noise signal in the jth channel. For instance, the multichannel noise can be estimated by single-input multiple-output independent component analysis (SIMO-ICA) [9]. Finally, we extract the target-speech-enhanced signal by applying DS to $\boldsymbol{y}_{\mathrm{chSS}}(f,\tau) = [y_1^{(\mathrm{chSS})}(f,\tau), \ldots, y_J^{(\mathrm{chSS})}(f,\tau)]^{\mathrm{T}}$. This procedure can be given by

$$y(f,\tau) = \boldsymbol{g}_{\mathrm{DS}}^{\mathrm{T}}(f,\theta_{\mathrm{U}})\boldsymbol{y}_{\mathrm{chSS}}(f,\tau), \tag{7}$$

where $y(f,\tau)$ is the final output of chSS+BF.

3. KURTOSIS-BASED MUSICAL-NOISE ANALYSIS

3.1. Analysis strategy

It has been reported by the authors that the amount of generated musical noise is strongly related to the difference between the before-and-after kurtosis of a signal in nonlinear signal processing [6]. Thus, in this section, we analyze the amount of musical noise generated through chSS+BF and BF+SS based on kurtosis. Basically, kurtosis increases through nonlinear signal processing, and larger increment of the kurtosis by nonlinear signal processing leads to more amount of musical noise generation [6]. Thus, the generated musical noise becomes smaller with a lower-kurtosis-increment signal processing. In the following subsections, hence, we analyze the kurtosis of BF+SS and chSS+BF, and prove which method can reduce the resultant kurtosis. Note that our analysis has no limitation in assumption of noise model, thus any noises including Gaussian and non-Gaussian can be under consideration.

In our previous study [8], we have already given a kurtosis-based musical-noise analysis of BF+SS and chSS+BF. However, the analysis did not consider the effect of flooring technique in SS, and some approximation was introduced. Unlike that analysis, in this study, we consider the effect of flooring technique, and do not use approximation in the analysis of SS.

3.2. Kurtosis

Kurtosis is one of the popular higher-order statistics for assessment of non-Gaussianity. Kurtosis is defined as

$$\mathrm{kurt}_x = \mu_4/\mu_2^2, \tag{8}$$

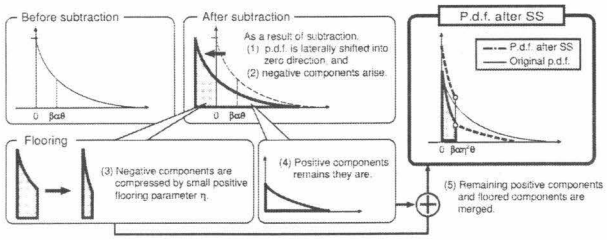

Fig. 2. Deformation of original p.d.f. of power-domain signal via SS.

where x is the probability variable, kurt_x is the kurtosis of x. Note that μ_n is not a central moment but a *raw moment*. Thus, (8) is not kurtosis in the mathematically strict definition, but a modified version, however, we refer to (8) as kurtosis in this paper. Although kurt_x becomes 3 if x is Gaussian signal, note that the kurtosis of Gaussian signal in power spectral domain becomes 6. This is because Gaussian signal in time domain obeys chi-square distribution with two degrees of freedom in power spectral domain. In chi-square distribution with two degrees of freedom, $\mu_4/\mu_2^2 = 6$.

3.3. Resultant kurtosis after SS

In the analysis, we assume that the input signal x in the power-spectral domain can be modeled by gamma distribution as [10]

$$P_{\mathrm{GM}}(x) = (\Gamma(\alpha)\theta^\alpha)^{-1} \cdot x^{\alpha-1} \exp\left\{-\frac{x}{\theta}\right\}, \tag{9}$$

where α is the shape parameter and θ is the shape parameter of gamma distribution. Moreover $P_{\mathrm{GM}}(x)$ is the probability density function (p.d.f.) of gamma distribution.

In traditional SS, the long-term-averaged power spectrum of a noise signal is utilized as the estimated noise power spectrum. Then, the estimated noise spectrum multiplied by the oversubtraction parameter β is subtracted from the observation. Here, it is well known that its mean is $\mathrm{E}[x] = \alpha\theta$. Thus, the amount of subtraction is $\beta\alpha\theta$. The subtraction of the estimated noise power spectrum in each frequency band can be considered as a deformation of the p.d.f., which is a lateral shift of the p.d.f. into zero-power direction (see Fig. 2). As a result, the probability of the negative power component arises. To avoid this, such negative components are replaced by observations with multiplying a small positive value η (the so-called flooring technique). This means that the probability of the floored components, which form a section cut from the original gamma distribution, is compressed by the effect of the flooring. Finally, the floored components are stacked to laterally shifted p.d.f. (see Fig. 2). Thus, the resultant p.d.f. after SS, $P_{\mathrm{SS}}(z)$, can be written as

$$P_{\mathrm{SS}}(z) = \begin{cases} \frac{1}{\theta^\alpha\Gamma(\alpha)}(z+\beta\alpha\theta)^{\alpha-1}\exp\left\{-\frac{z+\beta\alpha\theta}{\theta}\right\} & (z \geq \beta\alpha\eta^2\theta), \\ \frac{1}{\theta^\alpha\Gamma(\alpha)}(z+\beta\alpha\theta)^{\alpha-1}\exp\left\{-\frac{z+\beta\alpha\theta}{\theta}\right\} \\ \quad + \frac{1}{(\eta^2\theta)^\alpha\Gamma(\alpha)}z^{\alpha-1}\exp\left\{-\frac{z}{\eta^2\theta}\right\} & (0 < z < \beta\alpha\eta^2\theta), \end{cases} \tag{10}$$

where z is the random variable of p.d.f. after SS.

As described in (8), kurtosis is obtained by 2nd- and 4th-order moments. Taking integration of product of (10) and z^n (n is order of moment), we can derive such moments then the kurtosis after SS can be given as

$$\mathrm{kurt}_{\mathrm{SS}} = \Gamma(\alpha)\frac{\mathcal{F}(\alpha,\beta,\eta)}{\mathcal{G}^2(\alpha,\beta,\eta)}, \tag{11}$$

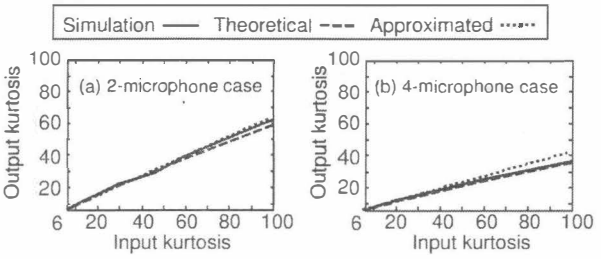

Fig. 3. Relation between input kurtosis and output kurtosis of DS.

where

$$\mathcal{F}(\alpha,\beta,\eta) = \Gamma(\beta\alpha,\alpha+4) - 4\beta\alpha\Gamma(\beta\alpha,\alpha+3) + 6\beta^2\alpha^2\Gamma(\beta\alpha,\alpha+2) - 4\beta^3\alpha^3\Gamma(\beta\alpha,\alpha+1) + \beta^4\alpha^4\Gamma(\beta\alpha,\alpha) + \eta^8\gamma(\beta\alpha,\alpha+4), \quad (12)$$

$$\mathcal{G}(\alpha,\beta,\eta) = \Gamma(\alpha)\Gamma(\beta\alpha,\alpha+2) - 2\beta\alpha\Gamma(\beta\alpha,\alpha+1) + \beta^2\alpha^2\Gamma(\beta\alpha,\alpha) + \eta^4\gamma(\beta\alpha,\alpha+2). \quad (13)$$

Here $\Gamma(x)$ is the gamma function, $\gamma(x,a)$ is the lower incomplete gamma function, and $\Gamma(x,a)$ is the upper incomplete gamma function.

Although Uemura et al. have given an approximated form (lower bound) of kurtosis after SS in Ref. [6], (11) involves no approximation through its derivation. Furthermore, (11) takes into account *the effect of the flooring technique* unlike Ref. [6].

3.4. Resultant kurtosis after DS

In this section, we analyze the kurtosis after DS. Supposed K_n is the nth-order cumulant of the *time-domain* input signal, the kurtosis after DS, $\mathrm{kurt}_{\mathrm{DS}}$, can be represented by [8]

$$\mathrm{kurt}_{\mathrm{DS}} = \frac{K_8 + 38JK_4^2 + 32JK_2K_6 + 288J^2K_2^2K_4 + 192J^3K_2^4}{2JK_4^2 + 16J^2K_2^2K_4 + 32J^3K_2^4}. \quad (14)$$

Although we cannot describe details of the derivation of (14) due to the limitation of the paper space, reference [8] helps you to understand the derivation of (14).

Regarding the power-spectral components with a gamma distribution, we can illustrate the relation between input kurtosis and output kurtosis via DS in Fig. 3. In the figure, solid lines indicate simulation results and broken lines show theoretical relations given by (14). From this figure, it is confirmed that the theoretical plots closely fit the simulation results. The relation between input/output kurtosis behaves as follows: (I) The output kurtosis is very close to a linear function of the input kurtosis, and (II) the output kurtosis is almost inversely proportional to the number of microphones. These behaviors result in the following simplified (but useful) approximation with an explicit function form:

$$\mathrm{kurt}_{\mathrm{DS}} \simeq J^{-0.7}\cdot(\mathrm{kurt}_{\mathrm{in}} - 6) + 6, \quad (15)$$

where $\mathrm{kurt}_{\mathrm{in}}$ is the input kurtosis. The approximated plots also match the simulation results in Fig. 3.

3.5. Resultant kurtosis: chSS+BF vs BF+SS

In the previous subsections, we discussed the resultant kurtosis of SS and DS. In this subsection, we analyze the resultant kurtosis for two types of composite systems, i.e., chSS+BF and BF+SS, and compete their effect on musical-noise generation.

In BF+SS, DS is first applied to a multichannel input signal. Then, the resultant kurtosis in the power spectral domain, $\mathrm{kurt}_{\mathrm{DS}}$, is

$$\mathrm{kurt}_{\mathrm{DS}} = J^{-0.7}\cdot(\mathrm{kurt}_{\mathrm{in}} - 6) + 6, \quad (16)$$

where $\mathrm{kurt}_{\mathrm{in}}$ is the kurtosis of the input signal in the power spectral domain. The shape parameter for the gamma distribution corresponding to $\mathrm{kurt}_{\mathrm{DS}}$, $\hat{\alpha}$, is given by [8]

$$\hat{\alpha} = \frac{\sqrt{\mathrm{kurt}_{\mathrm{DS}}^2 + 14\,\mathrm{kurt}_{\mathrm{DS}} + 1} - \mathrm{kurt}_{\mathrm{DS}} + 5}{2\,\mathrm{kurt}_{\mathrm{DS}} - 2}. \quad (17)$$

Consequently, using (11) and (17), the resultant kurtosis of BF+SS, $\mathrm{kurt}_{\mathrm{BF+SS}}$, can be written as

$$\mathrm{kurt}_{\mathrm{BF+SS}} = \Gamma(\hat{\alpha})\frac{\mathcal{F}(\hat{\alpha},\beta,\eta)}{\mathcal{G}^2(\hat{\alpha},\beta,\eta)}. \quad (18)$$

In chSS+BF, SS is first applied to each input channel. Thus, the output kurtosis of channelwise SS, $\mathrm{kurt}_{\mathrm{chSS}}$, can be given by

$$\mathrm{kurt}_{\mathrm{chSS}} = \Gamma(\alpha)\frac{\mathcal{F}(\alpha,\beta,\eta)}{\mathcal{G}^2(\alpha,\beta,\eta)}. \quad (19)$$

Finally, DS is performed and the resultant kurtosis of chSS+BF, $\mathrm{kurt}_{\mathrm{chSS+BF}}$, can be written as

$$\mathrm{kurt}_{\mathrm{chSS+BF}} = J^{-0.7}\left[\Gamma(\alpha)\frac{\mathcal{F}(\alpha,\beta,\eta)}{\mathcal{G}^2(\alpha,\beta,\eta)} - 6\right] + 6, \quad (20)$$

where we use (15). Originally, we should compare (18) and (20) here. However, one problem still remains: comparison under equivalent noise reduction performance; the noise reduction performances of BF+SS and chSS+BF are not equivalent. Then, in the next section, we conduct a noise reduction performance analysis and after that we give a kurtosis comparison of BF+SS and chSS+BF under the same noise reduction performance condition.

4. COMPARISON UNDER SAME NOISE REDUCTION PERFORMANCE CONDITION

In the previous subsection, we gave the kurtosis-based analysis of BF+SS and chSS+BF. However, the noise reduction performances (NRP) of BF+SS and chSS+BF are not equivalent even when the same parameters of SS are set. Then, in this section, we provide a flooring design strategy of BF+SS so that NRP in chSS+BF and BF+SS becomes the same. Finally, we present a kurtosis-based musical-noise comparison of BF+SS and chSS+BF under the same NRP condition.

4.1. NRP of BF+SS and chSS+BF

We utilize the following index to measure NRP

$$\mathrm{NRP} = 10\log_{10}\{E[n_{\mathrm{in}}]/E[n_{\mathrm{out}}]\}, \quad (21)$$

where n_{in} is the power-domain (noise) signal of the input and n_{out} is the power-domain (noise) signal of the output after processing.

Since we assume that the input signal can be modeled by the gamma distribution, the average power of input signal corresponds to the mean of a signal obeys gamma distribution, $\alpha\theta$. Next, the average power of the signal after SS is given by 1st-order moment of (10). This can be designated as

$$E[z] = \frac{\theta\cdot\Gamma(\beta\alpha,\alpha+1)}{\Gamma(\alpha)} - \beta\alpha\theta\cdot\frac{\Gamma(\beta\alpha,\alpha)}{\Gamma(\alpha)} + \frac{\eta\,\theta}{\Gamma(\alpha)}\gamma(\beta\alpha,\alpha+1). \quad (22)$$

Consequently, NRP of SS, $\mathrm{NRP}_{\mathrm{SS}}$, can be expressed by

$$\mathrm{NRP}_{\mathrm{SS}} = -10\log_{10}\left[\frac{\Gamma(\beta\alpha,\alpha+1)}{\Gamma(\alpha+1)} - \beta\cdot\frac{\Gamma(\beta\alpha,\alpha)}{\Gamma(\alpha)} + \eta^2\frac{\gamma(\beta\alpha,\alpha+1)}{\Gamma(\alpha+1)}\right]. \quad (23)$$

(23) involves not only β and η that are the parameter of SS but also α. This fact indicates that NRP of SS depends not only the parameter of SS but also the statistical characteristics of the input signal.

Next we conduct an analysis of NRP in DS. It is well known that the NRP of DS is proportional to the number of microphones. In particular, for spatially uncorrelated multichannel signals, the noise reduction performance of DS, $\mathrm{NRP}_{\mathrm{DS}}$, can be given as [2]

$$\mathrm{NRP}_{\mathrm{DS}} = 10\log_{10} J. \quad (24)$$

Using (23) and (24), the NRP of BF+SS, $\mathrm{NRP}_{\mathrm{BF+SS}}$, can be rep-

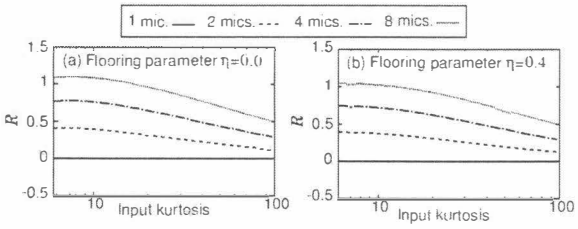

Fig. 4. Theoretical kurtosis ratio between chSS+BF and BF+SS for various values of input kurtosis.

resented by

$$\mathrm{NRP}_{\mathrm{BF+SS}} = -10\log_{10}\frac{1}{J\Gamma(\hat{\alpha})}\left[\frac{\Gamma(\beta\hat{\alpha},\hat{\alpha}+1)}{\hat{\alpha}} - \beta\cdot\Gamma(\beta\hat{\alpha},\hat{\alpha}) + \eta^2\frac{\gamma(\beta\hat{\alpha},\hat{\alpha}+1)}{\hat{\alpha}}\right], \quad (25)$$

where $\hat{\alpha}$ is the kurtosis after DS given by (17). By the same manner, the NRP of chSS+BF, $\mathrm{NRP}_{\mathrm{chSS+BF}}$, can be given by

$$\mathrm{NRP}_{\mathrm{chSS+BF}} = -10\log_{10}\frac{1}{J\cdot\Gamma(\alpha)}\left[\frac{\Gamma(\beta\alpha,\alpha+1)}{\alpha} - \beta\cdot\Gamma(\beta\alpha,\alpha) + \eta^2\frac{\gamma(\beta\alpha,\alpha+1)}{\alpha}\right]. \quad (26)$$

The above-mentioned discussion tells us that the NRPs of BF+SS and chSS+BF are not equivalent.

4.2. Flooring-parameter design in BF+SS for equivalent noise reduction performance

Using (25) and (26), the flooring parameter $\hat{\eta}$ that makes the NRP of BF+SS equal to that of chSS+BF, is

$$\hat{\eta} = \sqrt{\frac{\hat{\alpha}}{\gamma(\beta\hat{\alpha},\hat{\alpha}+1)}\cdot\left[\frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha)}\mathcal{H}(\alpha,\beta,\eta) - \mathcal{I}(\hat{\alpha},\beta)\right]}, \quad (27)$$

where

$$\mathcal{H}(\alpha,\beta,\eta) = \frac{\Gamma(\beta\alpha,\alpha+1)}{\alpha} - \beta\cdot\Gamma(\beta\alpha,\alpha) + \eta^2\frac{\gamma(\beta\alpha,\alpha+1)}{\alpha}, \quad (28)$$

$$\mathcal{I}(\hat{\alpha},\beta) = \frac{\Gamma(\beta\hat{\alpha},\hat{\alpha}+1)}{\hat{\alpha}} - \beta\cdot\Gamma(\beta\hat{\alpha},\hat{\alpha}). \quad (29)$$

Note that this $\hat{\eta}$ is valid for the case when NRP of BF+SS is superior to that of chSS+BF. Although we cannot describe the details of our analysis due to the limitation of the paper space, we have confirmed that NRP of BF+SS is always equal to or greater than that of chSS+BF for Gaussian or super-Gaussian signal.

By replacing η in (5) with this new flooring parameter $\hat{\eta}$, we can align NRPs of BF+SS and chSS+BF to ensure a fair comparison.

4.3. Kurtosis comparison under the same NRP condition

In this section, using the new flooring parameter for BF+SS, $\hat{\eta}$, we compare the output kurtosis of BF+SS and chSS+BF.

Here, we adopt the following index to compare the resultant kurtosis of chSS+BF and BF+SS:

$$R = \ln(\mathrm{kurt}_{\mathrm{BF+SS}} / \mathrm{kurt}_{\mathrm{chSS+BF}}). \quad (30)$$

where R expresses the resultant kurtosis ratio between chSS+BF and BF+SS. Note that a positive R indicates that chSS+BF reduces the kurtosis more than BF+SS, implying that less musical noise is generated in chSS+BF. The behavior of R is depicted in Figs. 4 and 5.

Figure 4 illustrates theoretical R for various values of input kurtosis. In this figure, β is fixed to 2.0, and the flooring parameter in chSS+BF is set to η = 0.0 and 0.4. The flooring parameter for BF+SS is automatically determined by (27). From this figure, we can confirm that chSS+BF reduces the kurtosis more than BF+SS for almost all input signals with various values of kurtosis. Theoretical values of R for various oversubtraction parameters are depicted in Fig. 5. Figure 5(a) shows that the output kurtosis of chSS+BF is always less than that of BF+SS for a Gaussian signal, even if η is nonzero. On the other hand, Fig. 5(b) implies that the output kurtosis of BF+SS becomes less than that of chSS+BF for some parameter settings when the kurtosis of input signal is high. However, such phenomena occur only for a large oversubtraction parameter, e.g., $\beta \geq 7$, which is not often applied in practical use. Therefore, we can conclude that chSS+BF reduces the kurtosis and musical noise more than BF+SS for almost all cases.

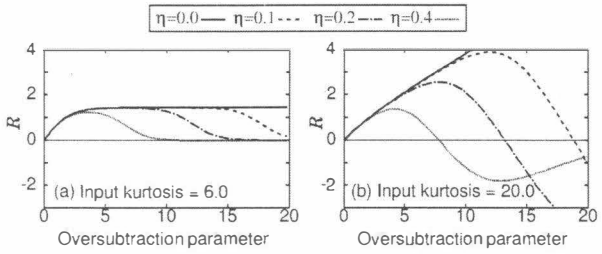

Fig. 5. Theoretical kurtosis ratio between chSS+BF and BF+SS for various oversubtraction parameters.

5. CONCLUSION

In this study, we conduct a theoretical analysis of the amount of musical noise generated via methods of integrating microphone array and SS under the same NRP condition. Moreover the analysis consider the effect of flooring technique in SS unlike our previous analysis. As a result of the analysis in this study, it is clarified that chSS+BF is proper to mitigate the musical noise for almost all practical cases. In the future, the influence by inter-channel correlation that cannot be neglect in real environment should be analyzed.

6. REFERENCES

[1] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," *Proc. of ICASSP1988*, vol.5, pp.2578–2581, 1988.

[2] M. Brandstein and D. Ward, "Microphone Arrays: Signal Processing Techniques and Applications," Springer-Verlag, 2001

[3] Y. Takahashi, et al., "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Trans. Audio, Speech and Language Proc.*, vol.17, no.4, pp.650–664, 2009.

[4] Y. Ohashi, et al., "Noise robust speech recognition based on spatial subtraction array," *Proc. of NSIP*, pp.324–327, 2005.

[5] O. Cappe and J. Laroche, "Evaluation of Short-Time Spectral Attenuation Techniques for the Restoration of Musical Recordings," *IEEE Trans. Speech and Audio Proc.*, vol.3, no.1, pp.84–93, 1995.

[6] Y. Uemura, et al., "Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics," *Proc. of IWAENC 2008*, 2008.

[7] 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.

[8] Yu Takahashi, et al., "Musical noise analysis based on higher order statistics for microphone array and nonlinear signal processing," *Proc. of ICASSP2009*, pp.229–232, 2009.

[9] T. Takatani, et al., "High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis," *IEICE Trans., Fundamentals*, vol.E87-A, no.8, pp.2063–2072, 2004.

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