

## Research Article

# Musical-Noise Analysis in Methods of Integrating Microphone Array and Spectral Subtraction Based on Higher-Order Statistics

Yu Takahashi,<sup>1</sup> Hiroshi Saruwatari (EURASIP Member),<sup>1</sup>  
Kiyohiro Shikano (EURASIP Member),<sup>1</sup> and Kazunobu Kondo<sup>2</sup>

<sup>1</sup>Graduate School of Information Science, Nara Institute of Science and Technology, Nara 630-0192, Japan

<sup>2</sup>SP Group, Center for Advanced Sound Technologies, Yamaha Corporation, Shizuoka 438-0192, Japan

Correspondence should be addressed to Yu Takahashi, yuu-t@yuu-t.sakura.ne.jp

Received 5 August 2009; Revised 3 November 2009; Accepted 16 March 2010

Academic Editor: Simon Doclo

Copyright © 2010 Yu Takahashi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We conduct an objective analysis on musical noise generated by two methods of integrating microphone array signal processing and spectral subtraction. To obtain better noise reduction, methods of integrating microphone array signal processing and nonlinear signal processing have been researched. However, nonlinear signal processing often generates musical noise. Since such musical noise causes discomfort to users, it is desirable that musical noise is mitigated. Moreover, it has been recently reported that higher-order statistics are strongly related to the amount of musical noise generated. This implies that it is possible to optimize the integration method from the viewpoint of not only noise reduction performance but also the amount of musical noise generated. Thus, we analyze the simplest methods of integration, that is, the delay-and-sum beamformer and spectral subtraction, and fully clarify the features of musical noise generated by each method. As a result, it is clarified that a specific structure of integration is preferable from the viewpoint of the amount of generated musical noise. The validity of the analysis is shown via a computer simulation and a subjective evaluation.

## 1. Introduction

There have recently been various studies on microphone array signal processing [1]; in particular, the delay-and-sum (DS) [2–4] array and the adaptive beamformer [5–7] are the most conventionally used microphone arrays for speech enhancement. Moreover, many methods of integrating microphone array signal processing and nonlinear signal processing such as spectral subtraction (SS) [8] have been studied with the aim of achieving better noise reduction [9–15]. It has been well demonstrated that such integration methods can achieve higher noise reduction performance than that obtained using conventional adaptive microphone arrays [13] such as the Griffith-Jim array [6]. However, a serious problem exists in such methods: artificial distortion (so-called *musical noise* [16]) due to nonlinear signal processing. Since the artificial distortion causes discomfort to users, it is desirable that musical noise is controlled through signal processing. However, in almost all nonlinear noise reduction

methods, the strength parameter to mitigate musical noise in nonlinear signal processing is determined heuristically. Although there have been some studies on reducing musical noise [16] and on nonlinear signal processing with less musical noise [17], evaluations have mainly depended on subjective tests by humans, and no objective evaluations have been performed to the best of our knowledge.

In our recent study, it was reported that the amount of generated musical noise is strongly related to the difference between higher-order statistics (HOS) before and after nonlinear signal processing [18]. This fact makes it possible to analyze the amount of musical noise arising through nonlinear signal processing. Therefore, on the basis of HOS, we can establish a mathematical metric for the amount of musical noise generated in an objective manner. One of the authors has analyzed single-channel nonlinear signal processing based on the objective metric and clarified the features of the amount of musical noise generated [18, 19]. In addition, this objective metric suggests the possibility that

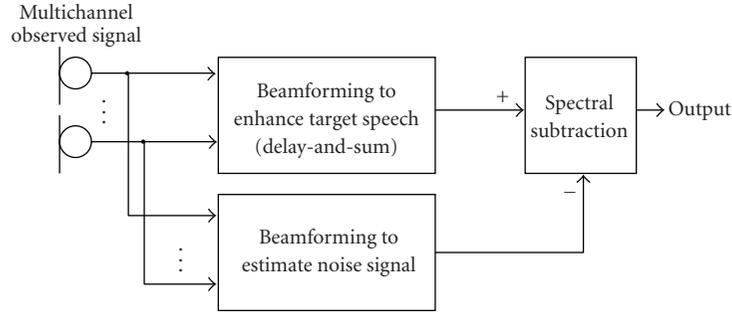


FIGURE 1: Block diagram of architecture for spectral subtraction after beamforming (BF+SS).

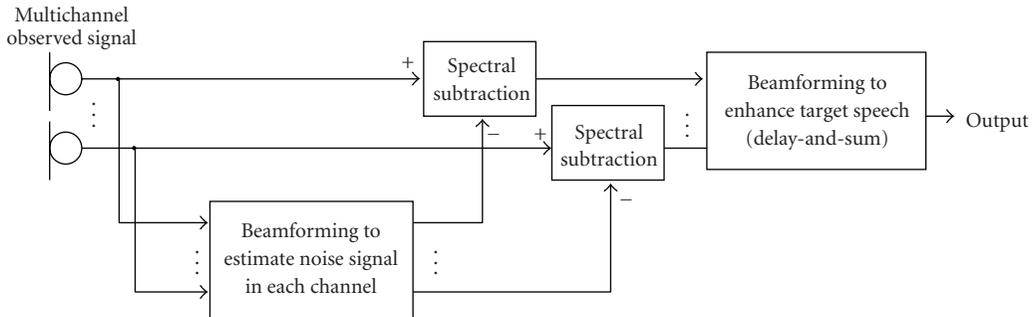


FIGURE 2: Block diagram of architecture for channelwise spectral subtraction before beamforming (chSS+BF).

methods of integrating microphone array signal processing and nonlinear signal processing can be optimized from the viewpoint of not only noise reduction performance but also the sound quality according to human hearing. As a first step toward achieving this goal, in this study we analyze the simplest case of the integration of microphone array signal processing and nonlinear signal processing by considering the integration of DS and SS. As a result of the analysis, we clarify the musical-noise generation features of two types of methods on integration of microphone array signal processing and SS.

Figure 1 shows a typical architecture used for the integration of microphone array signal processing and SS, where SS is performed after beamforming. Thus, we call this type of architecture BF+SS. Such a structure has been adopted in many integration methods [11, 15]. On the other hand, the integration architecture illustrated in Figure 2 is an alternative architecture used when SS is performed before beamforming. Such a structure is less commonly used, but some integration methods use this structure [12, 14]. In this architecture, channelwise SS is performed before beamforming, and we call this type of architecture chSS+BF.

We have already tried to analyze such methods of integrating DS and SS from the viewpoint of musical-noise generation on the basis of HOS [20]. However, in the analysis, we did not consider the effect of flooring in SS and the noise reduction performance. On the other hand, in this study we perform an exact analysis considering the effect of flooring in SS and the noise reduction performance. We analyze these two architectures on the basis of HOS and obtain the following results.

- (i) The amount of musical noise generated strongly depends on not only the oversubtraction parameter of SS but also *the statistical characteristics of the input signal*.
- (ii) Except for the specific condition that the input signal is Gaussian, the noise reduction performances of the two methods are not equivalent even if we set the same SS parameters.
- (iii) Under equivalent noise reduction performance conditions, chSS+BF generates less musical noise than BF+SS for almost all practical cases.

The most important contribution of this paper is that these findings are mathematically proved. In particular, the amount of musical noise generated and the noise reduction performance resulting from the integration of microphone array signal processing and SS are analytically formulated on the basis of HOS. Although there have been many studies on optimization methods based on HOS [21], this is the first time they have been used for musical-noise assessment. The validity of the analysis based on HOS is demonstrated via a computer simulation and a subjective evaluation by humans.

The rest of the paper is organized as follows. In Section 2, the two methods of integrating microphone array signal processing and SS are described in detail. In Section 3, the metric based on HOS used for the amount of musical noise generated is described. Next, the musical-noise analysis of SS, microphone array signal processing, and their integration methods are discussed in Section 4. In Section 5, the noise reduction performances of the two integration methods are discussed, and both methods are compared under equivalent

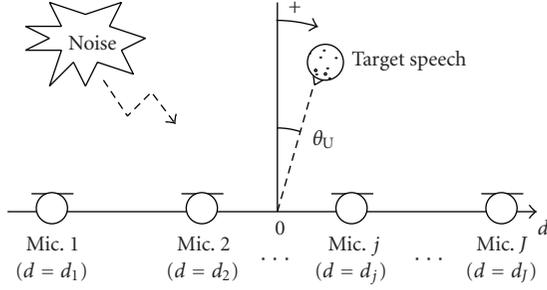


FIGURE 3: Configuration of microphone array and signals.

noise reduction performance conditions in Section 6. Moreover, the result of a computer simulation and experimental results are given in Section 7. Following a discussion of the results of the experiments, we give our conclusions in Section 8.

## 2. Methods of Integrating Microphone Array Signal Processing and SS

In this section, the formulations of the two methods of integrating microphone array signal processing and SS are described. First, BF+SS, which is a typical method of integration, is formulated. Next, an alternative method of integration, chSS+BF, is introduced.

**2.1. Sound-Mixing Model.** In this study, a uniform linear microphone array is assumed, where the coordinates of the elements are denoted by  $d_j$  ( $j = 1, \dots, J$ ) (see Figure 3) and  $J$  is the number of microphones. We consider one target speech signal and an additive interference signal. Multiple mixed signals are observed at each microphone element, and the short-time analysis of the observed signals is conducted by a frame-by-frame discrete Fourier transform (DFT). The observed signals are given by

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

where  $\mathbf{x}(f, \tau) = [x_1(f, \tau), \dots, x_J(f, \tau)]^T$  is the observed signal vector,  $\mathbf{h}(f) = [h_1(f), \dots, h_J(f)]^T$  is the transfer function vector,  $s(f, \tau)$  is the target speech signal, and  $\mathbf{n}(f, \tau) = [n_1(f, \tau), \dots, n_J(f, \tau)]^T$  is the noise signal vector.

**2.2. SS after Beamforming.** In BF+SS, the single-channel target-speech-enhanced signal is first obtained by beamforming, for example, by DS. Next, single-channel noise estimation is performed by a beamforming technique, for example, null beamformer [22] or adaptive beamforming [1]. Finally, we extract the resultant target-speech-enhanced signal via SS. The full details of signal processing are given below.

To enhance the target speech, DS is applied to the observed signal. This can be represented by

$$y_{\text{DS}}(f, \tau) = \mathbf{g}_{\text{DS}}(f, \theta_U)^T \mathbf{x}(f, \tau),$$

$$\mathbf{g}_{\text{DS}}(f, \theta_U) = [g_1^{(\text{DS})}(f, \theta_U), \dots, g_J^{(\text{DS})}(f, \theta_U)]^T, \quad (2)$$

$$g_j^{(\text{DS})}(f, \theta_U) = J^{-1} \cdot \exp\left(-\frac{i2\pi(f/M)f_s d_j \sin \theta_U}{c}\right),$$

where  $\mathbf{g}_{\text{DS}}(f, \theta_U)$  is the coefficient vector of the DS array and  $\theta_U$  is the specific fixed look direction known in advance. Also,  $f_s$  is the sampling frequency,  $M$  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 expressed as

$$|y_{\text{SS}}(f, \tau)| = \begin{cases} \sqrt{|y_{\text{DS}}(f, \tau)|^2 - \beta \cdot \text{E}_\tau[|\hat{n}(f, \tau)|^2]} & (\text{where } |y_{\text{DS}}(f, \tau)|^2 \\ & - \beta \cdot \text{E}_\tau[|\hat{n}(f, \tau)|^2] \geq 0), \\ \eta \cdot |y_{\text{DS}}(f, \tau)| & (\text{otherwise}), \end{cases} \quad (3)$$

where this procedure is a type of extended SS [23]. Here,  $y_{\text{SS}}(f, \tau)$  is the target-speech-enhanced signal,  $\beta$  is the oversubtraction parameter,  $\eta$  is the flooring parameter, and  $\hat{n}(f, \tau)$  is the estimated noise signal, which can generally be obtained by a beamforming techniques such as fixed or adaptive beamforming.  $\text{E}_\tau[\cdot]$  denotes the expectation operator with respect to the time-frame index. For example,  $\hat{n}(f, \tau)$  can be expressed as [13]

$$\hat{n}(f, \tau) = \lambda(f) \mathbf{g}_{\text{NBF}}^T(f) \mathbf{x}(f, \tau), \quad (4)$$

where  $\mathbf{g}_{\text{NBF}}(f)$  is the filter coefficient vector of the null beamformer [22] that steers the null directivity to the speech direction  $\theta_U$ , and  $\lambda(f)$  is the gain adjustment term, which is determined in a speech break period. Since the null beamformer can remove the speech signal by steering the null directivity to the speech direction, we can estimate the noise signal. Moreover, a method exists in which independent component analysis (ICA) is utilized as a noise estimator instead of the null beamformer [15].

**2.3. Channelwise SS before Beamforming.** In chSS+BF, we first perform SS independently in each input channel and then we derive a multichannel target-speech-enhanced signal

by channelwise SS. This can be expressed as

$$|y_j^{(\text{chSS})}(f, \tau)| = \begin{cases} \sqrt{|x_j(f, \tau)|^2 - \beta \cdot \text{E}_\tau \left[ |\tilde{n}_j(f, \tau)|^2 \right]} & \left( \text{where } |x_j(f, \tau)|^2 - \beta \cdot \text{E}_\tau \left[ |\tilde{n}_j(f, \tau)|^2 \right] \geq 0 \right), \\ \eta \cdot |x_j(f, \tau)| & \text{(otherwise),} \end{cases} \quad (5)$$

where  $y_j^{(\text{chSS})}(f, \tau)$  is the target-speech-enhanced signal obtained by SS at a specific channel  $j$  and  $\tilde{n}_j(f, \tau)$  is the estimated noise signal in the  $j$ th channel. For instance, the multichannel noise can be estimated by single-input multiple-output ICA (SIMO-ICA) [24] or a combination of ICA and the projection back method [25]. These techniques can provide the multichannel estimated noise signal, unlike traditional ICA. SIMO-ICA can separate mixed signals not into monaural source signals but into SIMO-model signals at the microphone. Here SIMO denotes the specific transmission system in which the input signal is a single source signal and the outputs are its transmitted signals observed at multiple microphones. Thus, the output signals of SIMO-ICA maintain the rich spatial qualities of the sound sources [24]. Also the projection back method provides SIMO-model-separated signals using the inverse of an optimized ICA filter [25].

Finally, we extract the target-speech-enhanced signal by applying DS to  $\mathbf{y}_{\text{chSS}}(f, \tau) = [y_1^{(\text{chSS})}(f, \tau), \dots, y_j^{(\text{chSS})}(f, \tau)]^T$ . This procedure can be expressed by

$$y(f, \tau) = \mathbf{g}_{\text{DS}}^T(f, \theta_U) \mathbf{y}_{\text{chSS}}(f, \tau), \quad (6)$$

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

Such a chSS+BF structure performs DS after (multichannel) SS. Since DS is basically signal processing in which the summation of the multichannel signal is taken, it can be considered that interchannel smoothing is applied to the multichannel spectral-subtracted signal. On the other hand, the resultant output signal of BF+SS remains as it is after SS. That is to say, it is expected that the output signal of chSS+BF is more natural (contains less musical noise) than that of BF+SS. In the following sections, we reveal that chSS+BF can output a signal with less musical noise than BF+SS in almost all cases on the basis of HOS.

### 3. Kurtosis-Based Musical-Noise Generation Metric

**3.1. Introduction.** It has been reported by the authors that the amount of musical noise generated is strongly related to the difference between the kurtosis of a signal before and after signal processing. Thus, in this paper, we analyze the amount of musical noise generated through BF+SS and chSS+BF on the basis of the change in the measured kurtosis. Hereinafter, we give details of the kurtosis-based musical-noise metric.

**3.2. Relation between Musical-Noise Generation and Kurtosis.** In our previous works [18–20], we defined musical noise as the audible isolated spectral components generated through signal processing. Figure 4(b) shows an example of a spectrogram of musical noise in which many isolated components can be observed. We speculate that the amount of musical noise is strongly related to the number of such isolated components and their level of isolation.

Hence, we introduce kurtosis to quantify the isolated spectral components, and we focus on the changes in kurtosis. Since isolated spectral components are dominant, they are heard as tonal sounds, which results in our perception of musical noise. Therefore, it is expected that obtaining the number of tonal components will enable us to quantify the amount of musical noise. However, such a measurement is extremely complicated; so instead we introduce a simple statistical estimate, that is, kurtosis.

This strategy allows us to obtain the characteristics of tonal components. The adopted kurtosis can be used to evaluate the width of the probability density function (p.d.f.) and the weight of its tails; that is, kurtosis can be used to evaluate the percentage of tonal components among the total components. A larger value indicates a signal with a heavy tail in its p.d.f., meaning that it has a large number of tonal components. Also, kurtosis has the advantageous property that it can be easily calculated in a concise algebraic form.

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

$$\text{kurt}_x = \frac{\mu_4}{\mu_2^2}, \quad (7)$$

where  $x$  is a random variable,  $\text{kurt}_x$  is the kurtosis of  $x$ , and  $\mu_n$  is the  $n$ th-order moment of  $x$ . Here  $\mu_n$  is defined as

$$\mu_n = \int_{-\infty}^{+\infty} x^n P(x) dx, \quad (8)$$

where  $P(x)$  denotes the p.d.f. of  $x$ . Note that this  $\mu_n$  is not a central moment *but a raw moment*. Thus, (7) is not kurtosis according to the mathematically strict definition, but a modified version; however, we refer to (7) as kurtosis in this paper.

**3.4. Kurtosis Ratio.** Although we can measure the number of tonal components by kurtosis, it is worth mentioning that kurtosis itself is not sufficient to measure musical noise. This is because that the kurtosis of some unprocessed signals such as speech signals is also high, but we do not perceive speech as musical noise. Since we aim to count only the musical-noise components, we should not consider genuine tonal components. To achieve this aim, we focus on the fact that musical noise is generated only in artificial signal processing. Hence, we should consider the change in kurtosis during signal processing. Consequently, we introduce the following *kurtosis ratio* [18] to measure the kurtosis change:

$$\text{kurtosis ratio} = \frac{\text{kurt}_{\text{proc}}}{\text{kurt}_{\text{input}}}, \quad (9)$$

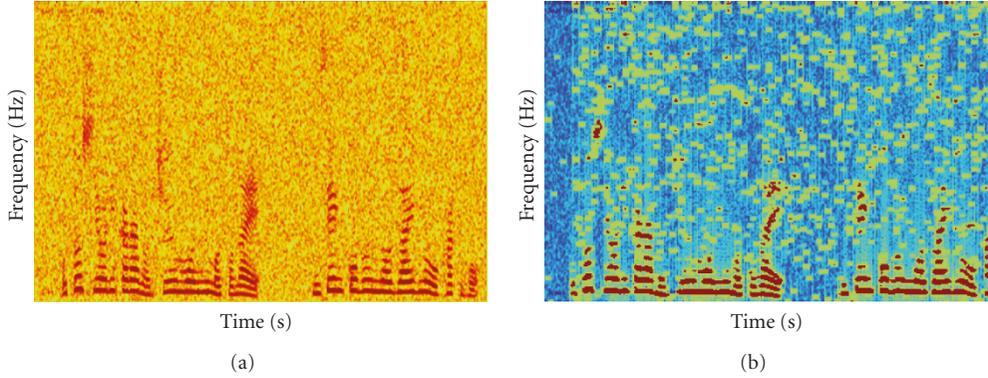


FIGURE 4: (a) Observed spectrogram and (b) processed spectrogram.

where  $\text{kurt}_{\text{proc}}$  is the kurtosis of the processed signal and  $\text{kurt}_{\text{input}}$  is the kurtosis of the input signal. A larger kurtosis ratio ( $\gg 1$ ) indicates a marked increase in kurtosis as a result of processing, implying that a larger amount of musical noise is generated. On the other hand, a smaller kurtosis ratio ( $\approx 1$ ) implies that less musical noise is generated. It has been confirmed that this kurtosis ratio closely matches the amount of musical noise in a subjective evaluation based on human hearing [18].

#### 4. Kurtosis-Based Musical-Noise Analysis for Microphone Array Signal Processing and SS

**4.1. Analysis Flow.** In the following sections, we carry out an analysis on musical-noise generation in BF+SS and chSS+BF based on kurtosis. The analysis is composed of the following three parts.

- (i) First, an analysis on musical-noise generation in BF+SS and chSS+BF based on kurtosis that does not take noise reduction performance into account is performed in this section.
- (ii) The noise reduction performance is analyzed in Section 5, and we reveal that the noise reduction performances of BF+SS and chSS+BF are not equivalent. Moreover, a flooring parameter designed to align the noise reduction performances of BF+SS and chSS+BF is also derived to ensure the fair comparison of BF+SS and chSS+BF.
- (iii) The kurtosis-based comparison between BF+SS and chSS+BF under the same noise reduction performance conditions is carried out in Section 6.

In the analysis in this section, we first clarify how kurtosis is affected by SS. Next, the same analysis is applied to DS. Finally, we analyze how kurtosis is increased by BF+SS and chSS+BF. Note that our analysis contains no limiting assumptions on the statistical characteristics of noise; thus, all noises including Gaussian and super-Gaussian noise can be considered.

**4.2. Signal Model Used for Analysis.** Musical-noise components generated from the noise-only period are dominant in spectrograms (see Figure 4); hence, we mainly focus our attention on musical-noise components originating from input noise signals.

Moreover, to evaluate the resultant kurtosis of SS, we introduce a gamma distribution to model the noise in the power domain [26–28]. The p.d.f. of the gamma distribution for random variable  $x$  is defined as

$$P_{\text{GM}}(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \cdot x^{\alpha-1} \exp\left\{-\frac{x}{\theta}\right\}, \quad (10)$$

where  $x \geq 0$ ,  $\alpha > 0$ , and  $\theta > 0$ . Here,  $\alpha$  denotes the shape parameter,  $\theta$  is the scale parameter, and  $\Gamma(\cdot)$  is the gamma function. The gamma distribution with  $\alpha = 1$  corresponds to the chi-square distribution with two degrees of freedom. Moreover, it is well known that the mean of  $x$  for a gamma distribution is  $E[x] = \alpha\theta$ , where  $E[\cdot]$  is the expectation operator. Furthermore, the kurtosis of a gamma distribution,  $\text{kurt}_{\text{GM}}$ , can be expressed as [18]

$$\text{kurt}_{\text{GM}} = \frac{(\alpha + 2)(\alpha + 3)}{\alpha(\alpha + 1)}. \quad (11)$$

Moreover, let us consider the power-domain noise signal,  $x_p$ , in the frequency domain, which is defined as

$$\begin{aligned} x_p &= |x_{\text{re}} + i \cdot x_{\text{im}}|^2 \\ &= (x_{\text{re}} + i \cdot x_{\text{im}})(x_{\text{re}} + i \cdot x_{\text{im}})^* \\ &= x_{\text{re}}^2 + x_{\text{im}}^2, \end{aligned} \quad (12)$$

where  $x_{\text{re}}$  is the real part of the complex-valued signal and  $x_{\text{im}}$  is its imaginary part, which are independent and identically distributed (i.i.d.) with each other, and the superscript  $*$  expresses complex conjugation. Thus, the power-domain signal is the sum of two squares of random variables with the same distribution.

Hereinafter, let  $x_{\text{re}}$  and  $x_{\text{im}}$  be the signals after DFT analysis of signal in a specific microphone  $j$ ,  $x_j$ , and we suppose that the statistical properties of  $x_j$  equal to  $x_{\text{re}}$  and  $x_{\text{im}}$ . Moreover, we assume the following;  $x_j$  is i.i.d. in each channel, the p.d.f. of  $x_j$  is symmetrical, and its mean is zero. These assumptions mean that the odd-order cumulants and moments are zero except for the first order.

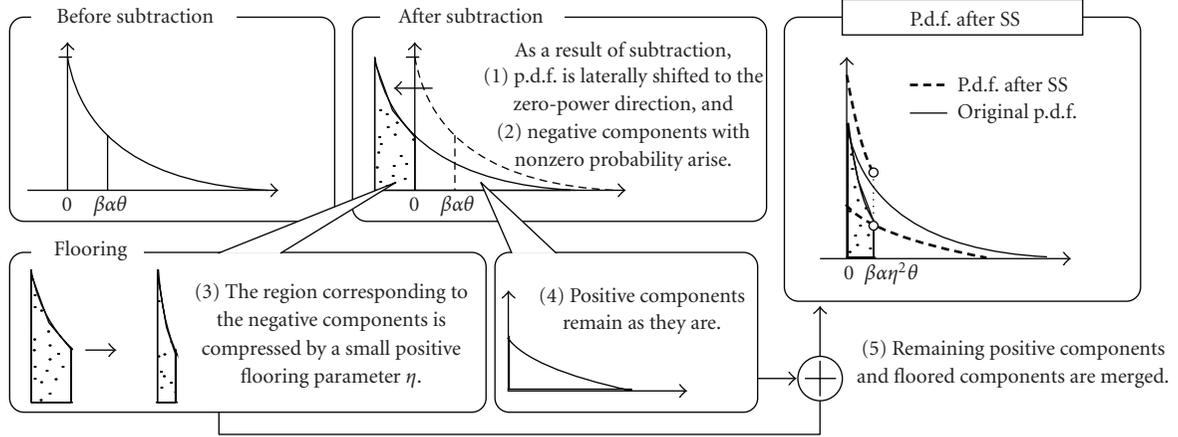


FIGURE 5: Deformation of original p.d.f. of power-domain signal via SS.

Although  $\text{kurt}_x = 3$  if  $x$  is a Gaussian signal, note that the kurtosis of a Gaussian signal in the power spectral domain is 6. This is because a Gaussian signal in the time domain obeys the chi-square distribution with two degrees of freedom in the power spectral domain; for such a chi-square distribution,  $\mu_4/\mu_2^2 = 6$ .

**4.3. Resultant Kurtosis after SS.** In this section, we analyze the kurtosis after SS. 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 power spectrum multiplied by the oversubtraction parameter  $\beta$  is subtracted from the observed power spectrum. When a gamma distribution is used to model the noise signal, its mean is  $\alpha\theta$ . 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 shift of the p.d.f. to the zero-power direction (see Figure 5). As a result, negative-power components with nonzero probability arise. To avoid this, such negative components are replaced by observations that are multiplied by a small positive value  $\eta$  (the so-called flooring technique). This means that the region corresponding to the probability of the negative components, which forms a section cut from the original gamma distribution, is compressed by the effect of the flooring. Finally, the floored components are superimposed on the laterally shifted p.d.f. (see Figure 5). Thus, the resultant p.d.f. after SS,  $P_{SS}(z)$ , can be written as

$$P_{SS}(z) = \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} (z + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{z + \beta\alpha\theta}{\theta}\right\} & (z \geq \beta\alpha\eta^2\theta), \\ \frac{1}{\theta^\alpha \Gamma(\alpha)} (z + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{z + \beta\alpha\theta}{\theta}\right\} \\ + \frac{1}{(\eta^2\theta)^\alpha \Gamma(\alpha)} z^{\alpha-1} \exp\left\{-\frac{z}{\eta^2\theta}\right\} & (0 < z < \beta\alpha\eta^2\theta), \end{cases} \quad (13)$$

where  $z$  is the random variable of the p.d.f. after SS. The derivation of  $P_{SS}(z)$  is described in Appendix A.

From (13), the kurtosis after SS can be expressed as

$$\text{kurt}_{SS} = \Gamma(\alpha) \frac{\mathcal{F}(\alpha, \beta, \eta)}{\mathcal{G}^2(\alpha, \beta, \eta)}, \quad (14)$$

where

$$\begin{aligned} \mathcal{G}(\alpha, \beta, \eta) &= \Gamma(\alpha)\Gamma(\beta\alpha, \alpha + 2) - 2\beta\alpha\Gamma(\beta\alpha, \alpha + 1) \\ &\quad + \beta^2\alpha^2\Gamma(\beta\alpha, \alpha) + \eta^4\gamma(\beta\alpha, \alpha + 2), \\ \mathcal{F}(\alpha, \beta, \eta) &= \Gamma(\beta\alpha, \alpha + 4) - 4\beta\alpha\Gamma(\beta\alpha, \alpha + 3) \\ &\quad + 6\beta^2\alpha^2\Gamma(\beta\alpha, \alpha + 2) - 4\beta^3\alpha^3\Gamma(\beta\alpha, \alpha + 1) \\ &\quad + \beta^4\alpha^4\Gamma(\beta\alpha, \alpha) + \eta^8\gamma(\beta\alpha, \alpha + 4). \end{aligned} \quad (15)$$

Here,  $\Gamma(b, a)$  is the upper incomplete gamma function defined as

$$\Gamma(b, a) = \int_b^\infty t^{a-1} \exp\{-t\} dt, \quad (16)$$

and  $\gamma(b, a)$  is the lower incomplete gamma function defined as

$$\gamma(b, a) = \int_0^b t^{a-1} \exp\{-t\} dt. \quad (17)$$

The detailed derivation of (14) is given in Appendix B. Although Uemura et al. have given an approximated form (lower bound) of the kurtosis after SS in [18], (14) involves no approximation throughout its derivation. Furthermore, (14) takes into account the effect of the flooring technique unlike [18].

Figure 6(a) depicts the theoretical kurtosis ratio after SS,  $\text{kurt}_{SS}/\text{kurt}_{GM}$ , for various values of oversubtraction parameter  $\beta$  and flooring parameter  $\eta$ . In the figure, the kurtosis of the input signal is fixed to 6.0, which corresponds

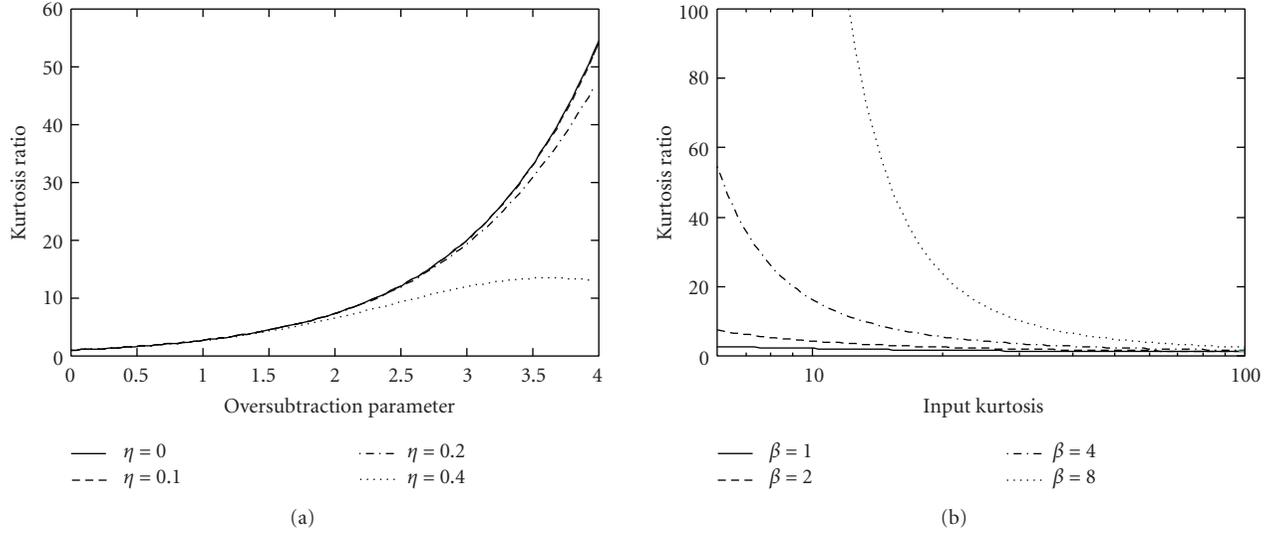


FIGURE 6: (a) Theoretical kurtosis ratio after SS for various values of oversubtraction parameter  $\beta$  and flooring parameter  $\eta$ . In this figure, kurtosis of input signal is fixed to 6.0. (b) Theoretical kurtosis ratio after SS for various values of input kurtosis. In this figure, flooring parameter  $\eta$  is fixed to 0.0.

to a Gaussian signal. From this figure, it is confirmed that the kurtosis ratio is basically proportional to the oversubtraction parameter  $\beta$ . However, kurtosis does not monotonically increase when the flooring parameter is nonzero. For instance, the kurtosis ratio is smaller than the peak value when  $\beta = 4$  and  $\eta = 0.4$ . This phenomenon can be explained as follows. For a large oversubtraction parameter, almost all the spectral components become negative due to the larger lateral shift of the p.d.f. by SS. Since flooring is applied to avoid such negative components, almost all the components are reconstructed by flooring. Therefore, the statistical characteristics of the signal never change except for its amplitude if  $\eta \neq 0$ . Generally, kurtosis does not depend on the change in amplitude; consequently, it can be considered that kurtosis does not markedly increase when a larger oversubtraction parameter and a larger flooring parameter are set.

The relation between the theoretical kurtosis ratio and the kurtosis of the original input signal is shown in Figure 6(b). In the figure,  $\eta$  is fixed to 0.0. It is revealed that the kurtosis ratio after SS rapidly decreases as the input kurtosis increases, even with the same oversubtraction parameter  $\beta$ . Therefore, the kurtosis ratio after SS, which is related to the amount of musical noise, strongly depends on the statistical characteristics of the input signal. That is to say, SS generates a larger amount of musical noise for a Gaussian input signal than for a super-Gaussian input signal. This fact has been reported in [18].

**4.4. Resultant Kurtosis after DS.** In this section, we analyze the kurtosis after DS, and we reveal that DS can reduce the kurtosis of input signals. Since we assume that the statistical properties of  $x_{re}$  or  $x_{im}$  are the same as that of  $x_j$ , the effect of DS on the change in kurtosis can be derived from the cumulants and moments of  $x_j$ .

For cumulants, when  $X$  and  $Y$  are independent random variables it is well known that the following relation holds:

$$\text{cum}_n(aX + bY) = a^n \text{cum}_n(X) + b^n \text{cum}_n(Y), \quad (18)$$

where  $\text{cum}_n(\cdot)$  denotes the  $n$ th-order cumulant. The cumulants of the random variable  $X$ ,  $\text{cum}_n(X)$ , are defined by a cumulant-generating function, which is the logarithm of the moment-generating function. The cumulant-generating function  $\mathcal{C}(\zeta)$  is defined as

$$\mathcal{C}(\zeta) = \log(\mathbb{E}[\exp\{\zeta X\}]) = \sum_{n=1}^{\infty} \text{cum}_n(X) \frac{\zeta^n}{n!}, \quad (19)$$

where  $\zeta$  is an auxiliary variable and  $\mathbb{E}[\exp\{\zeta X\}]$  is the moment-generating function. Thus, the  $n$ th-order cumulant  $\text{cum}_n(X)$  is represented by

$$\text{cum}_n(X) = \mathcal{C}^{(n)}(0), \quad (20)$$

where  $\mathcal{C}^{(n)}(\zeta)$  is the  $n$ th-order derivative of  $\mathcal{C}(\zeta)$ .

Now we consider the DS beamformer, which is steered to  $\theta_U = 0$  and whose array weights are  $1/J$ . Using (18), the resultant  $n$ th-order cumulant after DS,  $\mathcal{K}_n = \text{cum}_n(y_{DS})$ , can be expressed by

$$\mathcal{K}_n = \frac{1}{J^{n-1}} K_n, \quad (21)$$

where  $K_n = \text{cum}_n(x_j)$  is the  $n$ th-order cumulant of  $x_j$ . Therefore, using (21) and the well-known mathematical relation between cumulants and moments, the power-spectral-domain kurtosis after DS,  $\text{kurt}_{DS}$  can be expressed by

$$\text{kurt}_{DS} = \frac{\mathcal{K}_8 + 38\mathcal{K}_4^2 + 32\mathcal{K}_2\mathcal{K}_6 + 288\mathcal{K}_2^2\mathcal{K}_4 + 192\mathcal{K}_2^4}{2\mathcal{K}_4^2 + 16\mathcal{K}_2^2\mathcal{K}_4 + 32\mathcal{K}_2^4}. \quad (22)$$

The detailed derivation of (22) is described in Appendix C.

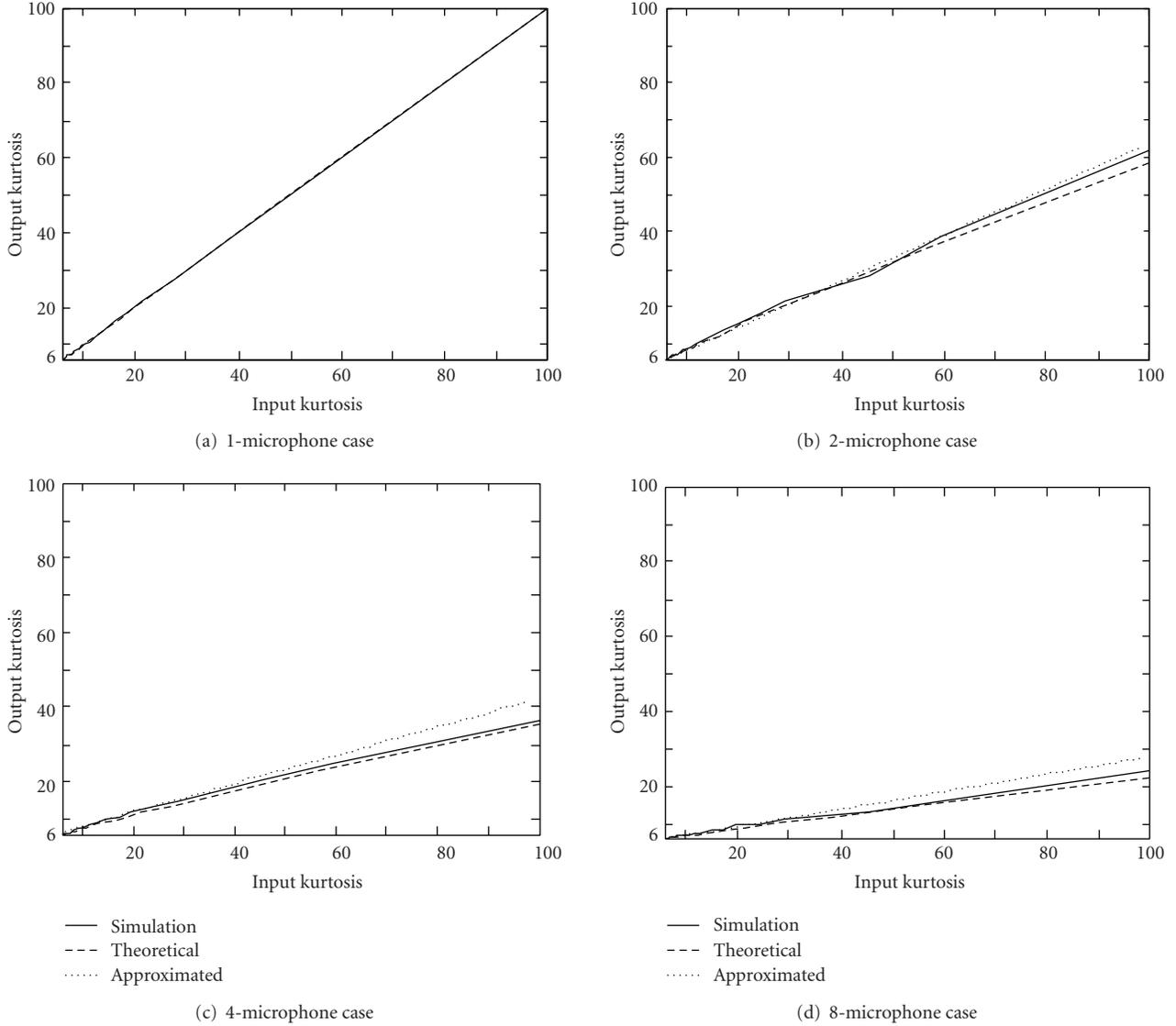


FIGURE 7: Relation between input kurtosis and output kurtosis after DS. Solid lines indicate simulation results, broken lines express theoretical plots obtained by (22), and dotted lines show approximate results obtained by (23).

Regarding the power-spectral components obtaining from a gamma distribution, we illustrate the relation between input kurtosis and output kurtosis after DS in Figure 7. In the figure, solid lines indicate simulation results and broken lines show theoretical relations given by (22). The simulation results are derived as follows. First, multichannel signals with various values of kurtosis are generated artificially from a gamma distribution. Next, DS is applied to the generated signals. Finally, kurtosis after DS is estimated from the signal resulting from DS. 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:

$$\text{kurt}_{\text{DS}} \simeq J^{-0.7} \cdot (\text{kurt}_{\text{in}} - 6) + 6, \quad (23)$$

where  $\text{kurt}_{\text{in}}$  is the input kurtosis. The approximated plots also match the simulation results in Figure 7.

When input signals involve interchannel correlation, the relation between input kurtosis and output kurtosis after DS approaches that for only one microphone. If all input signals are identical signals, that is, the signals are completely correlated, the output after DS also becomes the same as the input signal. In such a case, the effect of DS on the change in kurtosis corresponds to that for only one microphone. However, the interchannel correlation is not equal to one within all frequency subbands for a diffuse noise field that is a typically considered noise field. It is well known that the

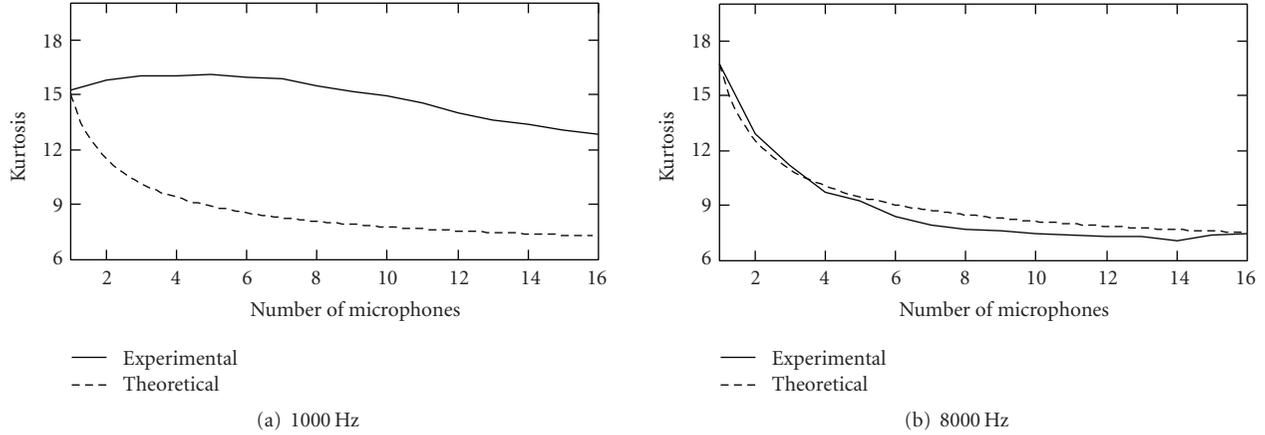


FIGURE 8: Simulation result for noise with interchannel correlation (solid line) and theoretical effect of DS assuming no interchannel correlation (broken line) in each frequency subband.

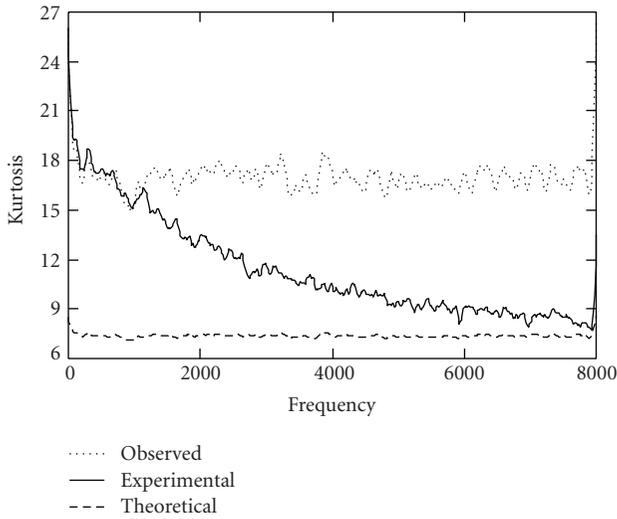


FIGURE 9: Simulation result for noise with interchannel correlation (solid line), theoretical effect of DS assuming no interchannel correlation (broken line), and kurtosis of the observed signal without any signal processing (dotted line) in eight-microphone case.

intensity of the interchannel correlation is strong in lower-frequency subbands and weak in higher-frequency subbands for the diffuse noise field [1]. Therefore, in lower-frequency subbands, it can be expected that DS does not significantly reduce the kurtosis of the signal.

As it is well known that the interchannel correlation for a diffuse noise field between two measurement locations can be expressed by the sinc function [1], we can state how array signal processing is affected by the interchannel correlation. However, we cannot know exactly how cumulants are changed by the interchannel correlation because (18) only holds when signals are mutually independent. Therefore, we cannot formulate how kurtosis is changed via DS for signals with interchannel correlation. For this reason,

we experimentally investigate the effect of interchannel correlation in the following.

Figures 8 and 9 show preliminary simulation results of DS. In this simulation, SS is first applied to a multichannel Gaussian signal with interchannel correlation in the diffuse noise field. Next, DS is applied to the signal after SS. In the preliminary simulation, the interelement distance between microphones is 2.15 cm. From the results shown in Figures 8(a) and 9, we can confirm that the effect of DS on kurtosis is weak in lower-frequency subbands, although it should be noted that the effect does not completely disappear. Also, the theoretical kurtosis curve is in good agreement with the actual results in higher-frequency subbands (see Figures 8(b) and 9). This is because the interchannel correlation is weak in higher-frequency subbands. Consequently, for a diffuse noise field, DS can reduce the kurtosis of the input signal even if interchannel correlation exists.

If input noise signals contain no interchannel correlation, the distance between microphones does not affect the results. That is to say, the kurtosis change via DS can be well fit to (23). Otherwise, in lower-frequency subbands, it is expected that the mitigation effect of kurtosis by DS degrades with decreasing distance between microphones. This is because the interchannel correlation in lower-frequency subbands increases with decreasing distance between microphones. In higher-frequency subbands, the effect of the distance between microphones is thought to be small.

**4.5. Resultant Kurtosis: BF+SS versus chSS+BF.** In the previous subsections, we discussed the resultant kurtosis after SS and DS. In this subsection, we analyze the resultant kurtosis for two types of composite systems, that is, BF+SS and chSS+BF, and compare their effect on musical-noise generation. As described in Section 3, it is expected that a smaller increase in kurtosis leads to a smaller amount of musical noise generated.

In BF+SS, DS is first applied to a multichannel input signal. At this point, the resultant kurtosis in the power spectral domain,  $kurt_{DS}$ , can be represented by (23). Using

(11), we can derive a shape parameter for the gamma distribution corresponding to  $\text{kurt}_{\text{DS}}$ ,  $\hat{\alpha}$ , as

$$\hat{\alpha} = \frac{\sqrt{\text{kurt}_{\text{DS}}^2 + 14 \text{kurt}_{\text{DS}} + 1} - \text{kurt}_{\text{DS}} + 5}{2 \text{kurt}_{\text{DS}} - 2}. \quad (24)$$

The derivation of (24) is shown in Appendix D. Consequently, using (14) and (24), the resultant kurtosis after BF+SS,  $\text{kurt}_{\text{BF+SS}}$ , can be written as

$$\text{kurt}_{\text{BF+SS}} = \Gamma(\hat{\alpha}) \frac{\mathcal{F}(\hat{\alpha}, \beta, \eta)}{\mathcal{G}^2(\hat{\alpha}, \beta, \eta)}. \quad (25)$$

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

$$\text{kurt}_{\text{chSS}} = \Gamma(\alpha) \frac{\mathcal{F}(\alpha, \beta, \eta)}{\mathcal{G}^2(\alpha, \beta, \eta)}. \quad (26)$$

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

$$\text{kurt}_{\text{chSS+BF}} = J^{-0.7} \left[ \Gamma(\alpha) \frac{\mathcal{F}(\alpha, \beta, \eta)}{\mathcal{G}^2(\alpha, \beta, \eta)} - 6 \right] + 6, \quad (27)$$

where we use (23).

We should compare  $\text{kurt}_{\text{BF+SS}}$  and  $\text{kurt}_{\text{chSS+BF}}$  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 as described in the next section. Moreover, the design of a flooring parameter so that the noise reduction performances of both methods become equivalent will be discussed in the next section. Therefore,  $\text{kurt}_{\text{BF+SS}}$  and  $\text{kurt}_{\text{chSS+BF}}$  will be compared in Section 6 under equivalent noise reduction performance conditions.

## 5. Noise Reduction Performance Analysis

In the previous section, we did not discuss the noise reduction performances of BF+SS and chSS+BF. In this section, a mathematical analysis of the noise reduction performances of BF+SS and chSS+BF is given. As a result of this analysis, it is revealed that the noise reduction performances of BF+SS and chSS+BF are not equivalent even if the same parameters are set in the SS part. We then derive a flooring-parameter design strategy for aligning the noise reduction performances of BF+SS and chSS+BF.

*5.1. Noise Reduction Performance of SS.* We utilize the following index to measure the noise reduction performance (NRP):

$$\text{NRP} = 10 \log_{10} \frac{E[n_{\text{out}}]}{E[n_{\text{in}}]}, \quad (28)$$

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

First, we derive the average power of the input signal. We assume that the input signal in the power domain can be modeled by a gamma distribution. Then, the average power of the input signal is given as

$$\begin{aligned} E[n_{\text{in}}] &= E[x] = \int_0^{\infty} x P_{\text{GM}}(x) dx \\ &= \int_0^{\infty} x \cdot \frac{1}{\theta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} \exp\left\{-\frac{x}{\theta}\right\} dx \\ &= \frac{1}{\theta^{\alpha} \Gamma(\alpha)} \int_0^{\infty} x^{\alpha} \exp\left\{-\frac{x}{\theta}\right\} dx. \end{aligned} \quad (29)$$

Here, let  $t = x/\theta$ , then  $\theta dt = dx$ . Thus,

$$\begin{aligned} E[n_{\text{in}}] &= \frac{1}{\theta^{\alpha} \Gamma(\alpha)} \int_0^{\infty} (\theta t)^{\alpha} \exp\{-t\} \theta dt \\ &= \frac{\theta^{\alpha+1}}{\theta^{\alpha} \Gamma(\alpha)} \int_0^{\infty} t^{\alpha} \exp\{-t\} dt \\ &= \frac{\theta \Gamma(\alpha+1)}{\Gamma(\alpha)} = \theta \alpha. \end{aligned} \quad (30)$$

This corresponds to the mean of a random variable with a gamma distribution.

Next, the average power of the signal after SS is calculated. Here, let  $z$  obey the p.d.f. of the signal after SS,  $P_{\text{SS}}(z)$ , defined by (13); then the average power of the signal after SS can be expressed as

$$\begin{aligned} E[n_{\text{out}}] &= E[z] \\ &= \int_0^{\infty} z P_{\text{SS}}(z) dz \\ &= \int_0^{\infty} \frac{z}{\theta^{\alpha} \Gamma(\alpha)} (z + \beta \alpha \theta)^{\alpha-1} \exp\left\{-\frac{z + \beta \alpha \theta}{\theta}\right\} dz \\ &\quad + \int_0^{\beta \alpha \eta^2 \theta} \frac{z}{(\eta^2 \theta)^{\alpha} \Gamma(\alpha)} z^{\alpha-1} \exp\left\{-\frac{z}{\eta^2 \theta}\right\} dz. \end{aligned} \quad (31)$$

We now consider the first term of the right-hand side in (31). We let  $t = z + \beta \alpha \theta$ , then  $dt = dz$ . As a result,

$$\begin{aligned} &\int_0^{\infty} \frac{z}{\theta^{\alpha} \Gamma(\alpha)} (z + \beta \alpha \theta)^{\alpha-1} \exp\left\{-\frac{z + \beta \alpha \theta}{\theta}\right\} dz \\ &= \int_{\beta \alpha \theta}^{\infty} (t - \beta \alpha \theta) \cdot \frac{1}{\theta^{\alpha} \Gamma(\alpha)} \cdot t^{\alpha-1} \exp\left\{-\frac{t}{\theta}\right\} dt \\ &= \int_{\beta \alpha \theta}^{\infty} \frac{1}{\theta^{\alpha} \Gamma(\alpha)} \cdot t^{\alpha} \exp\left\{-\frac{t}{\theta}\right\} dt \\ &\quad - \int_{\beta \alpha \theta}^{\infty} \frac{\beta \alpha \theta}{\theta^{\alpha} \Gamma(\alpha)} \cdot t^{\alpha-1} \exp\left\{-\frac{t}{\theta}\right\} dt \\ &= \frac{\theta \cdot \Gamma(\beta \alpha, \alpha + 1)}{\Gamma(\alpha)} - \beta \alpha \theta \cdot \frac{\Gamma(\beta \alpha, \alpha)}{\Gamma(\alpha)}. \end{aligned} \quad (32)$$

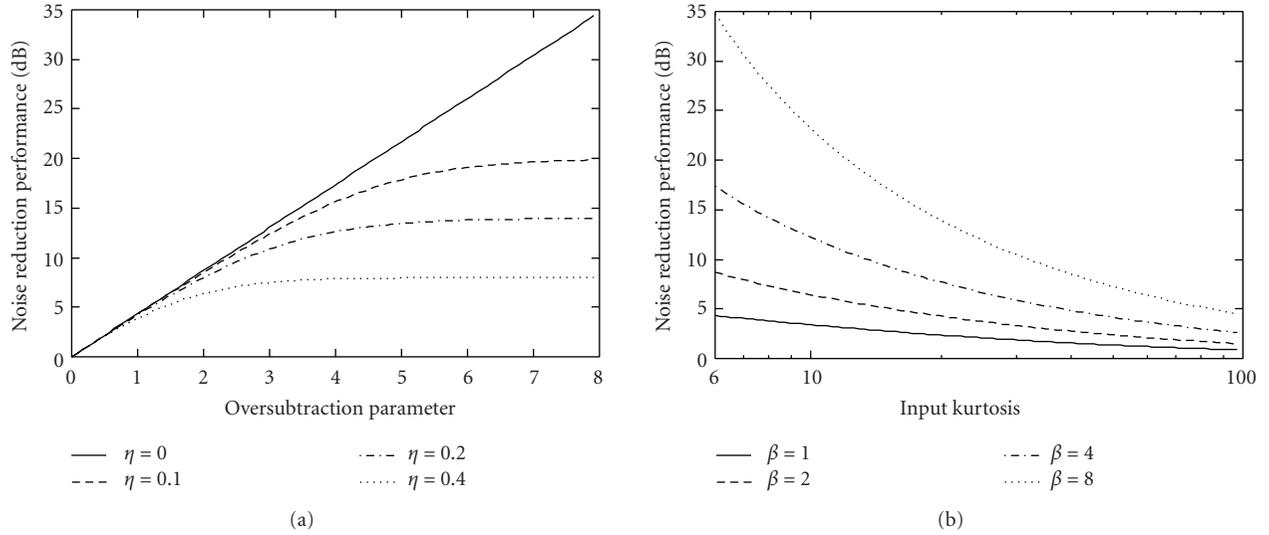


FIGURE 10: (a) Theoretical noise reduction performance of SS with various oversubtraction parameters  $\beta$  and flooring parameters  $\eta$ . In this figure, kurtosis of input signal is fixed to 6.0. (b) Theoretical noise reduction performance of SS with various values of input kurtosis. In this figure, flooring parameter  $\eta$  is fixed to 0.0.

Also, we deal with the second term of the right-hand side in (31). We let  $t = z/(\eta^2\theta)$  then  $\eta^2\theta dt = dz$ , resulting in

$$\begin{aligned} & \int_0^{\beta\alpha\eta^2\theta} \frac{z}{(\eta^2\theta)^\alpha \Gamma(\alpha)} z^{\alpha-1} \exp\left\{-\frac{z}{\eta^2\theta}\right\} dz \\ &= \frac{1}{(\eta^2\theta)^\alpha \Gamma(\alpha)} \int_0^{\beta\alpha} (\eta^2\theta t)^\alpha \cdot \exp\{-t\} \eta^2\theta dt \quad (33) \\ &= \frac{\eta^2\theta}{\Gamma(\alpha)} \gamma(\beta\alpha, \alpha + 1). \end{aligned}$$

Using (30), (32), and (33), the noise reduction performance of SS,  $\text{NRP}_{\text{SS}}$ , can be expressed by

$$\begin{aligned} \text{NRP}_{\text{SS}} &= 10 \log_{10} \left( \frac{E[z]}{E[x]} \right) \\ &= -10 \log_{10} \left[ \frac{\Gamma(\beta\alpha, \alpha + 1)}{\Gamma(\alpha + 1)} \right. \\ &\quad \left. - \beta \cdot \frac{\Gamma(\beta\alpha, \alpha)}{\Gamma(\alpha)} + \eta^2 \frac{\gamma(\beta\alpha, \alpha + 1)}{\Gamma(\alpha + 1)} \right]. \quad (34) \end{aligned}$$

Figure 10(a) shows the theoretical value of  $\text{NRP}_{\text{SS}}$  for various values of oversubtraction parameter  $\beta$  and flooring parameter  $\eta$ , where the kurtosis of the input signal is fixed to 6.0, corresponding to a Gaussian signal. From this figure, it is confirmed that  $\text{NRP}_{\text{SS}}$  is proportional to  $\beta$ . However,  $\text{NRP}_{\text{SS}}$  hits a peak when  $\eta$  is nonzero even for a large value of  $\beta$ . The relation between the theoretical value of  $\text{NRP}_{\text{SS}}$  and the kurtosis of the input signal is illustrated in Figure 10(b).

In this figure,  $\eta$  is fixed to 0.0. It is revealed that  $\text{NRP}_{\text{SS}}$  decreases as the input kurtosis increases. This is because the mean of a high-kurtosis signal tends to be small. Since the shape parameter  $\alpha$  of a high-kurtosis signal becomes small, the mean  $\alpha\theta$  corresponding to the amount of subtraction also becomes small. As a result,  $\text{NRP}_{\text{SS}}$  is decreased as the input kurtosis increases. That is to say, the  $\text{NRP}_{\text{SS}}$  strongly depends on the statistical characteristics of the input signal as well as the values of the oversubtraction and flooring parameters.

**5.2. Noise Reduction Performance of DS.** It is well known that the noise reduction performance of DS ( $\text{NRP}_{\text{DS}}$ ) is proportional to the number of microphones. In particular, for spatially uncorrelated multichannel signals,  $\text{NRP}_{\text{DS}}$  is given as [1]

$$\text{NRP}_{\text{DS}} = 10 \log_{10} J. \quad (35)$$

**5.3. Resultant Noise Reduction Performance: BF+SS versus chSS+BF.** In the previous subsections, the noise reduction performances of SS and DS were discussed. In this subsection, we derive the resultant noise reduction performances of the composite systems of SS and DS, that is, BF+SS and chSS+BF.

The noise reduction performance of BF+SS is analyzed as follows. In BF+SS, DS is first applied to a multichannel input signal. If this input signal is spatially uncorrelated, its noise reduction performance can be represented by  $10 \log_{10} J$ . After DS, SS is applied to the signal. Note that DS affects the kurtosis of the input signal. As described in Section 4.4, the resultant kurtosis after DS can be approximated as  $J^{-0.7} \cdot (\text{kurt}_{\text{in}} - 6) + 6$ . Thus, SS is applied to the kurtosis-modified signal. Consequently, using (24), (34), and (35),

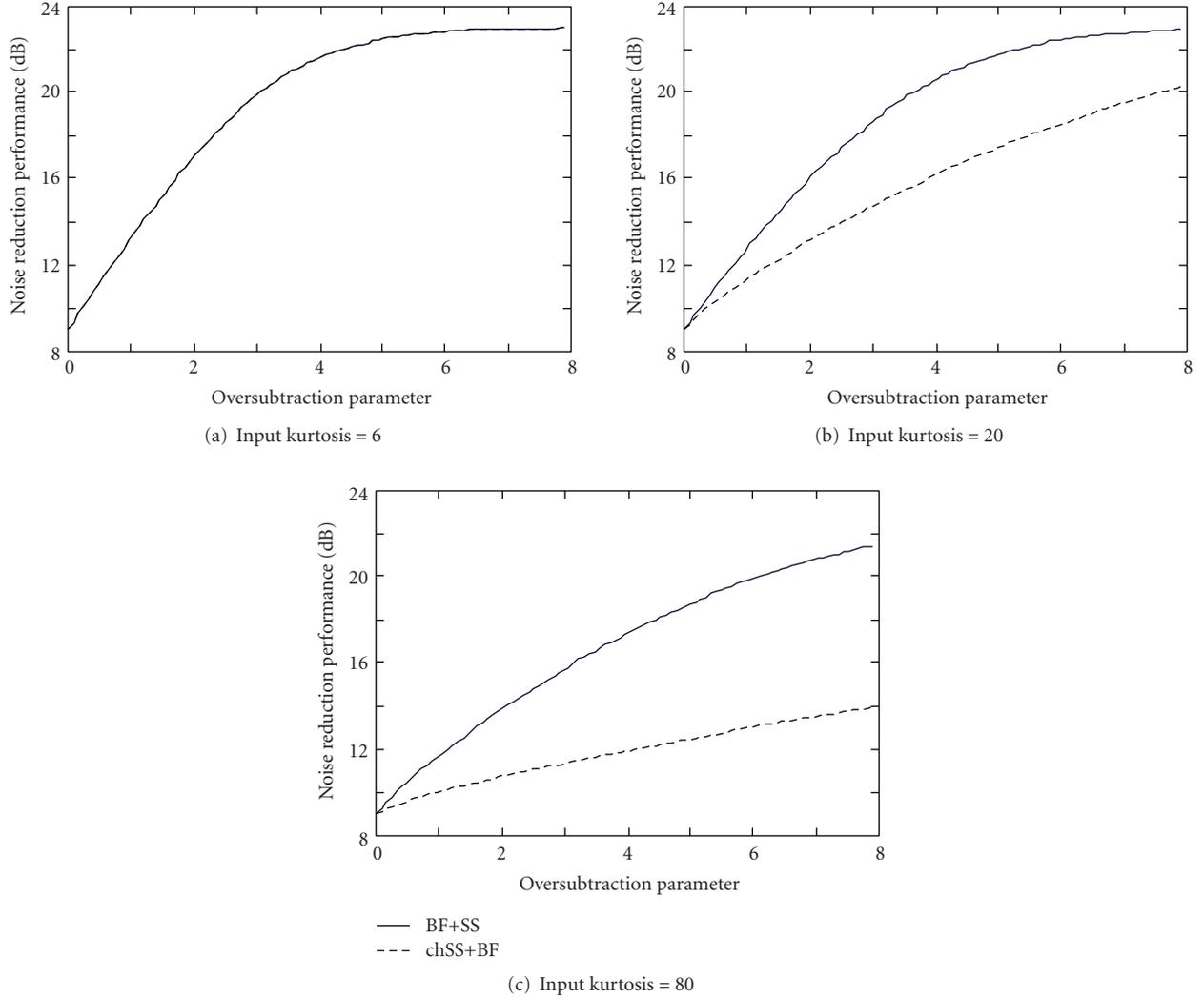


FIGURE 11: Comparison of noise reduction performances of chSS+BF with BF+SS. In this figure, flooring parameter is fixed to 0.2 and number of microphones is 8.

the noise reduction performance of BF+SS,  $\text{NRP}_{\text{BF+SS}}$ , is given as

$$\begin{aligned}
 \text{NRP}_{\text{BF+SS}} &= 10 \log_{10} J - 10 \log_{10} \\
 &\times \left[ \frac{\Gamma(\beta \hat{\alpha}, \hat{\alpha} + 1)}{\Gamma(\hat{\alpha} + 1)} - \beta \cdot \frac{\Gamma(\beta \hat{\alpha}, \hat{\alpha})}{\Gamma(\hat{\alpha})} + \eta^2 \frac{\gamma(\beta \hat{\alpha}, \hat{\alpha} + 1)}{\Gamma(\hat{\alpha} + 1)} \right] \quad (36) \\
 &= -10 \log_{10} \frac{1}{J \cdot \Gamma(\hat{\alpha})} \\
 &\times \left[ \frac{\Gamma(\beta \hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} - \beta \cdot \Gamma(\beta \hat{\alpha}, \hat{\alpha}) + \eta^2 \frac{\gamma(\beta \hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} \right],
 \end{aligned}$$

where  $\hat{\alpha}$  is defined by (24).

In chSS+BF, SS is first applied to a multichannel input signal; then DS is applied to the resulting signal. Thus, using

(34) and (35), the noise reduction performance of chSS+BF,  $\text{NRP}_{\text{chSS+BF}}$ , can be represented by

$$\begin{aligned}
 \text{NRP}_{\text{chSS+BF}} &= -10 \log_{10} \frac{1}{J \cdot \Gamma(\alpha)} \\
 &\times \left[ \frac{\Gamma(\beta \alpha, \alpha + 1)}{\alpha} - \beta \cdot \Gamma(\beta \alpha, \alpha) + \eta^2 \frac{\gamma(\beta \alpha, \alpha + 1)}{\alpha} \right]. \quad (37)
 \end{aligned}$$

Figure 11 depicts the values of  $\text{NRP}_{\text{BF+SS}}$  and  $\text{NRP}_{\text{chSS+BF}}$ . From this result, we can see that the noise reduction performances of both methods are equivalent when the input signal is Gaussian. However, if the input signal is super-Gaussian,  $\text{NRP}_{\text{BF+SS}}$  exceeds  $\text{NRP}_{\text{chSS+BF}}$ . This is due to the fact that DS is first applied to the input signal in BF+SS; thus, DS reduces the kurtosis of the signal. Since  $\text{NRP}_{\text{SS}}$  for

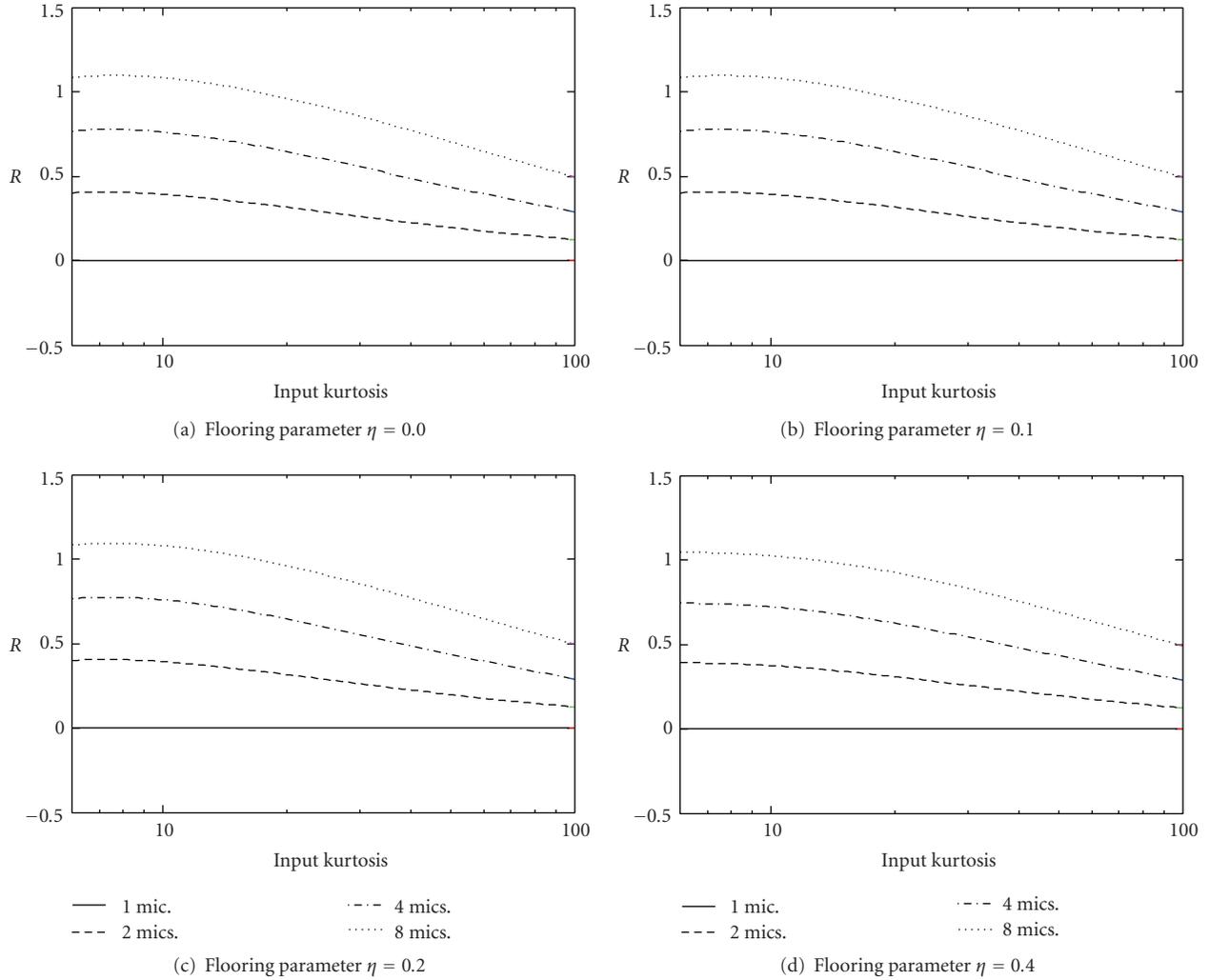


FIGURE 12: Theoretical kurtosis ratio between BF+SS and chSS+BF for various values of input kurtosis. In this figure, oversubtraction parameter is  $\beta = 2.0$  and flooring parameter in chSS+BF is (a)  $\eta = 0.0$ , (b)  $\eta = 0.1$ , (c)  $\eta = 0.2$ , and (d)  $\eta = 0.4$ .

a low-kurtosis signal is greater than that for a high-kurtosis signal (see Figure 10(b)), the noise reduction performance of BF+SS is superior to that of chSS+BF.

This discussion implies that  $\text{NRP}_{\text{BF+SS}}$  and  $\text{NRP}_{\text{chSS+BF}}$  are not equivalent under some conditions. Thus the kurtosis-based analysis described in Section 4 is biased and requires some adjustment. In the following subsection, we will discuss how to align the noise reduction performances of BF+SS and chSS+BF.

**5.4. Flooring-Parameter Design in BF+SS for Equivalent Noise Reduction Performance.** In this section, we describe the flooring-parameter design in BF+SS so that  $\text{NRP}_{\text{BF+SS}}$  and  $\text{NRP}_{\text{chSS+BF}}$  become equivalent.

Using (36) and (37), the flooring parameter  $\hat{\eta}$  that makes  $\text{NRP}_{\text{BF+SS}}$  equal to  $\text{NRP}_{\text{chSS+BF}}$ , is

$$\hat{\eta} = \sqrt{\frac{\hat{\alpha}}{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)} \cdot \left[ \frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha)} \mathcal{H}(\alpha, \beta, \eta) - \mathcal{I}(\hat{\alpha}, \beta) \right]}, \quad (38)$$

where

$$\mathcal{H}(\alpha, \beta, \eta) = \frac{\Gamma(\beta\alpha, \alpha + 1)}{\alpha} - \beta \cdot \Gamma(\beta\alpha, \alpha) + \eta^2 \frac{\gamma(\beta\alpha, \alpha + 1)}{\alpha}, \quad (39)$$

$$\mathcal{I}(\hat{\alpha}, \beta) = \frac{\Gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} - \beta \cdot \Gamma(\beta\hat{\alpha}, \hat{\alpha}). \quad (40)$$

The detailed derivation of (38) is given in Appendix E. By replacing  $\eta$  in (3) with this new flooring parameter  $\hat{\eta}$ , we can align  $\text{NRP}_{\text{BF+SS}}$  and  $\text{NRP}_{\text{chSS+BF}}$  to ensure a fair comparison.

## 6. Output Kurtosis Comparison under Equivalent 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.

Setting  $\hat{\eta}$  to (25), the output kurtosis of BF+SS is modified to

$$\text{kurt}_{\text{BF+SS}} = \Gamma(\hat{\alpha}) \frac{\mathcal{F}(\hat{\alpha}, \beta, \hat{\eta})}{\mathcal{G}^2(\hat{\alpha}, \beta, \hat{\eta})}. \quad (41)$$

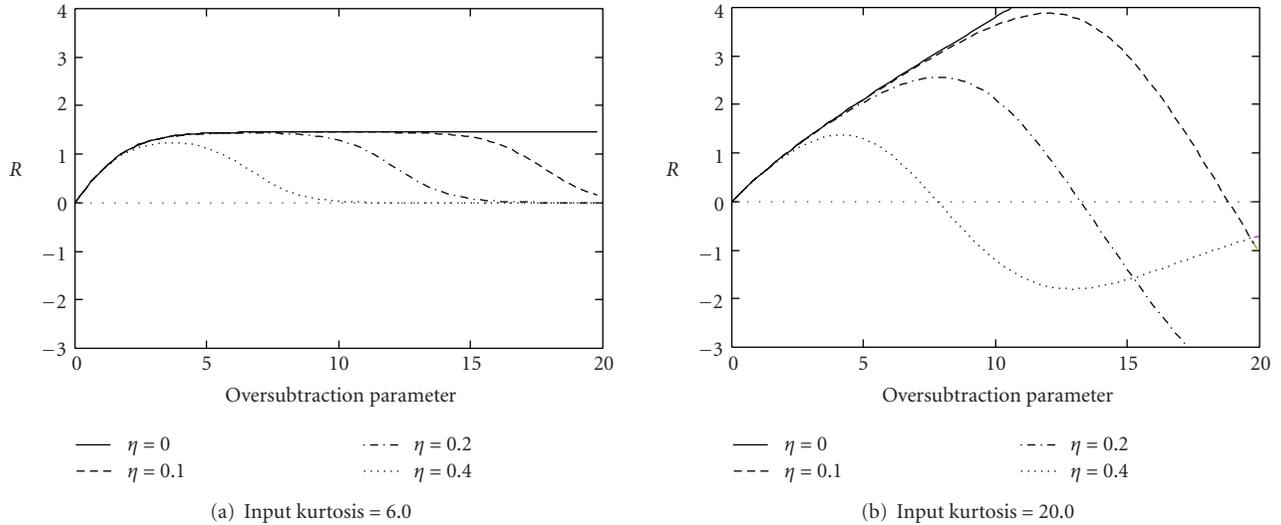


FIGURE 13: Theoretical kurtosis ratio between BF+SS and chSS+BF for various oversubtraction parameters. In this figure, number of microphones is fixed to 8, and input kurtosis is (a) 6.0 (Gaussian) and (b) 20.0 (super-Gaussian).

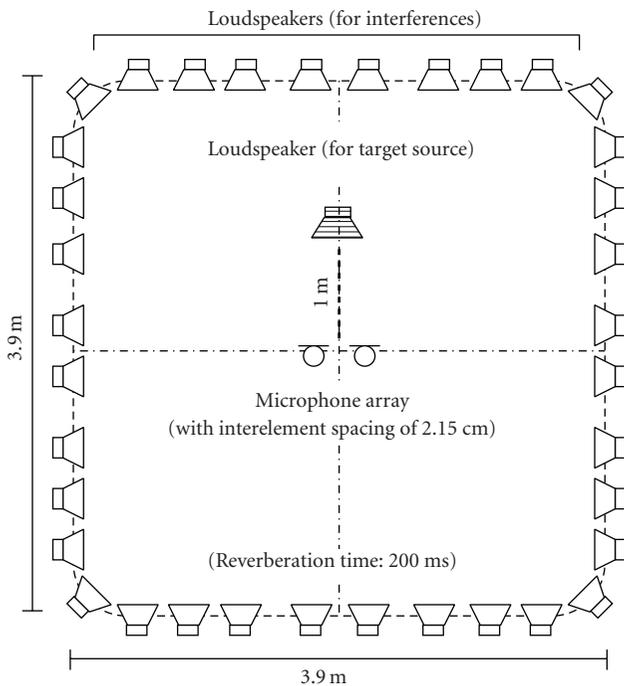


FIGURE 14: Reverberant room used in our simulations.

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

$$R = \ln \frac{\text{kurt}_{\text{BF+SS}}}{\text{kurt}_{\text{chSS+BF}}}, \quad (42)$$

where  $R$  expresses the resultant kurtosis ratio between BF+SS and chSS+BF. 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 Figures 12 and 13. Figure 12 illustrates theoretical values of  $R$  for various values of input kurtosis.

In this figure,  $\beta$  is fixed to 2.0 and the flooring parameter in chSS+BF is set to  $\eta = 0.0, 0.1, 0.2$ , and 0.4. The flooring parameter for BF+SS is automatically determined by (38). 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 input kurtosis. Theoretical values of  $R$  for various oversubtraction parameters are depicted in Figure 13. Figure 13(a) shows that the output kurtosis after chSS+BF is always less than that after BF+SS for a Gaussian signal, even if  $\eta$  is nonzero. On the other hand, Figure 13(b) implies that the output kurtosis after BF+SS becomes less than that after chSS+BF for some parameter settings. However, this only occurs for a large oversubtraction parameter, for example,  $\beta \geq 7$ , which is not often applied in practical use. Therefore, it can be considered that chSS+BF reduces the kurtosis and musical noise more than BF+SS in almost all cases.

## 7. Experiments and Results

**7.1. Computer Simulations.** First, we compare BF+SS and chSS+BF in terms of kurtosis ratio and noise reduction performance. We use 16-kHz-sampled signals as test data, in which the target speech is the original speech convoluted with impulse responses recorded in a room with 200 millisecond reverberation (see Figure 14), and to which an artificially generated spatially uncorrelated white Gaussian or super-Gaussian signal is added. We use six speakers (six sentences) as sources of the original clean speech. The number of microphone elements in the simulation is varied from 2 to 16, and their interelement distance is 2.15 cm each. The oversubtraction parameter  $\beta$  is set to 2.0 and the flooring parameter for BF+SS,  $\eta$ , is set to 0.0, 0.2, 0.4, or 0.8. Note that the flooring parameter in chSS+BF is set to 0.0. In the simulation, we assume that the long-term-averaged power spectrum of noise is estimated perfectly in advance.

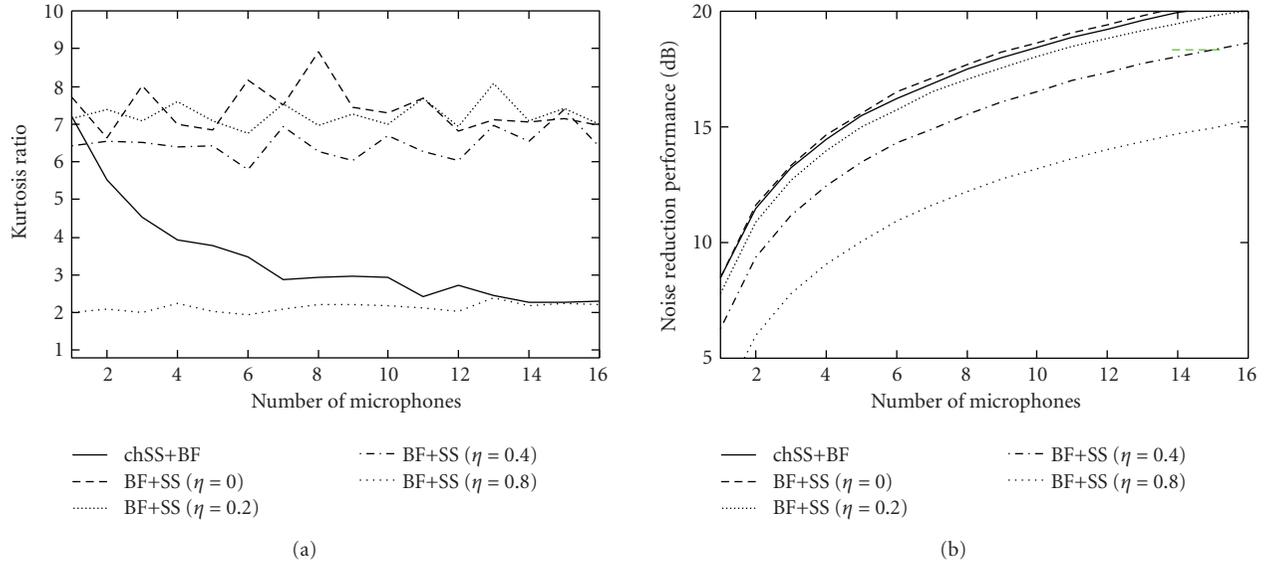


FIGURE 15: Results for Gaussian input signal. (a) Kurtosis ratio and (b) noise reduction performance for BF+SS with various flooring parameters.

Here, we utilize the kurtosis ratio defined in Section 3.4 to measure the difference in kurtosis, which is related to the amount of musical noise generated. The kurtosis ratio is given by

$$\text{Kurtosis ratio} = \frac{\text{kurt}(n_{\text{proc}}(f, \tau))}{\text{kurt}(n_{\text{org}}(f, \tau))}, \quad (43)$$

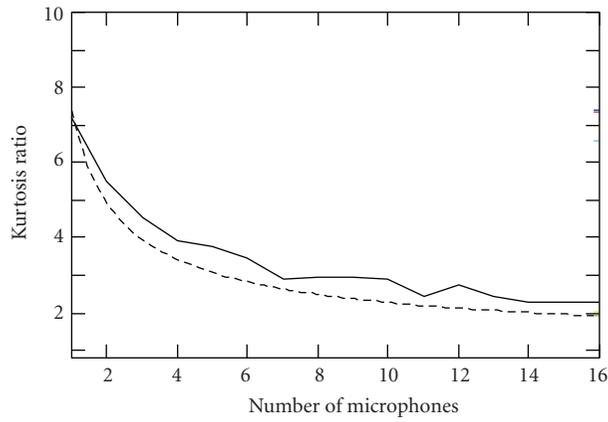
where  $n_{\text{proc}}(f, \tau)$  is the power spectra of the residual noise signal after processing, and  $n_{\text{org}}(f, \tau)$  is the power spectra of the original noise signal before processing. This kurtosis ratio indicates the extent to which kurtosis is increased with processing. Thus, a smaller kurtosis ratio is desirable. Moreover, the noise reduction performance is measured using (28).

Figures 15–17 show the simulation results for a Gaussian input signal. From Figure 15(a), we can see that the kurtosis ratio of chSS+BF decreases almost monotonically with increasing number of microphones. On the other hand, the kurtosis ratio of BF+SS does not exhibit such a tendency regardless of the flooring parameter. Also, the kurtosis ratio of chSS+BF is lower than that of BF+SS for all cases except for  $\eta = 0.8$ . Moreover, we can confirm from Figure 15(b) that the values of noise reduction performance for BF+SS with flooring parameter  $\eta = 0.0$  and chSS+BF are almost the same. When the flooring parameter for BF+SS is nonzero, the kurtosis ratio of BF+SS becomes smaller but the noise reduction performance degrades. On the other hand, for Gaussian signals, chSS+BF can reduce the kurtosis ratio, that is, reduce the amount of musical noise generated, without degrading the noise reduction performance. Indeed BF+SS with  $\eta = 0.8$  reduces the kurtosis ratio more than chSS+BF, but the noise reduction performance of BF+SS is extremely degraded. Furthermore, we can confirm from Figures 16 and 17 that the theoretical kurtosis ratio and noise

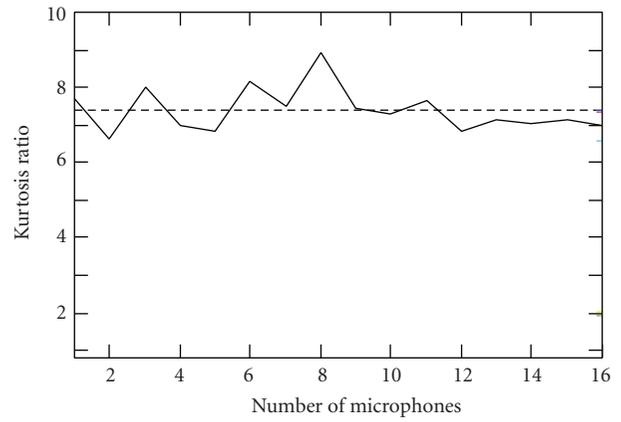
reduction performance closely fit the experimental results. These findings also support the validity of the analysis in Sections 4, 5, and 6.

Figures 18–20 illustrate the simulation results for a super-Gaussian input signal. It is confirmed from Figure 18(a) that the kurtosis ratio of chSS+BF also decreases monotonically with increasing number of microphones. Unlike the case of the Gaussian input signal, the kurtosis ratio of BF+SS with  $\eta = 0.8$  also decreases with increasing number of microphones. However, for a lower value of the flooring parameter, the kurtosis ratio of BF+SS is not degraded. Moreover, the kurtosis ratio of chSS+BF is lower than that of BF+SS for almost all cases. For the super-Gaussian input signal, in contrast to the case of the Gaussian input signal, the noise reduction performance of BF+SS with  $\eta = 0.0$  is greater than that of chSS+BF (see Figure 18(b)). That is to say, the noise reduction performance of BF+SS is superior to that of chSS+BF for the same flooring parameter. This result is consistent with the analysis in Section 5. The noise reduction performance of BF+SS with  $\eta = 0.4$  is comparable to that of chSS+BF. However, the kurtosis ratio of chSS+BF is still lower than that of BF+SS with  $\eta = 0.4$ . This result also coincides with the analysis in Section 6. On the other hand, the kurtosis ratio of BF+SS with  $\eta = 0.8$  is almost the same as that of chSS+BF. However, the noise reduction performance of BF+SS with  $\eta = 0.8$  is lower than that of chSS+BF. Thus, it can be confirmed that chSS+BF reduces the kurtosis ratio more than BF+SS for a super-Gaussian signal under the same noise reduction performance. Furthermore, the theoretical kurtosis ratio and noise reduction performance closely fit the experimental results in Figures 19 and 20.

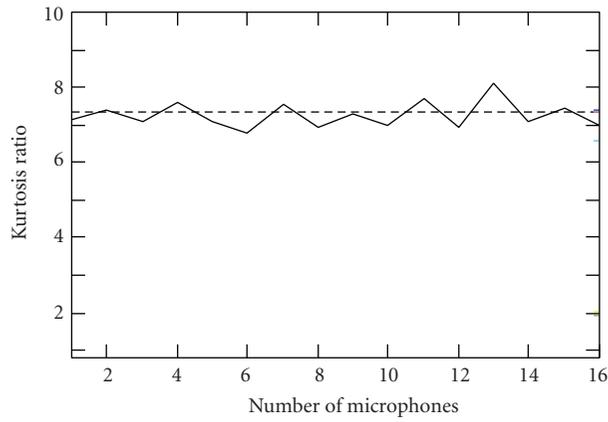
We also compare speech distortion originating from chSS+BF and BF+SS on the basis of cepstral distortion (CD) [29] for the four-microphone case. The comparison



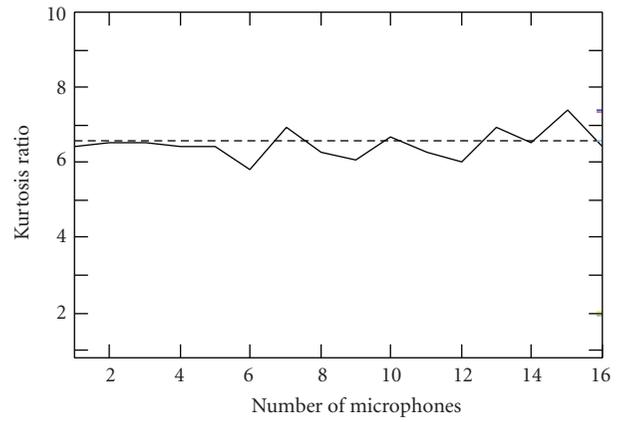
(a) chSS+BF



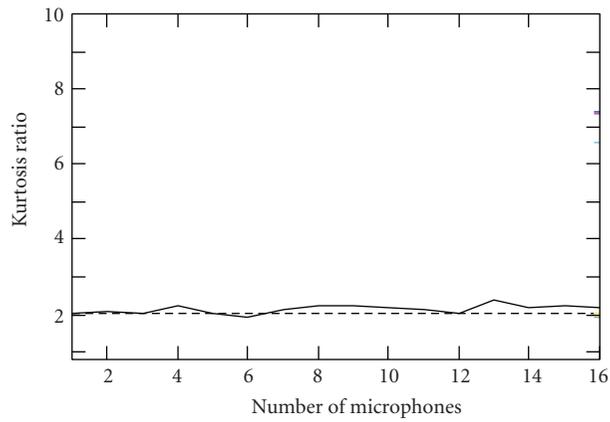
(b) BF+SS ( $\eta = 0.0$ )



(c) BF+SS ( $\eta = 0.2$ )



(d) BF+SS ( $\eta = 0.4$ )



(e) BF+SS ( $\eta = 0.8$ )

— Experimental  
 - - - Theoretical

FIGURE 16: Comparison between experimental and theoretical kurtosis ratios for Gaussian input signal.

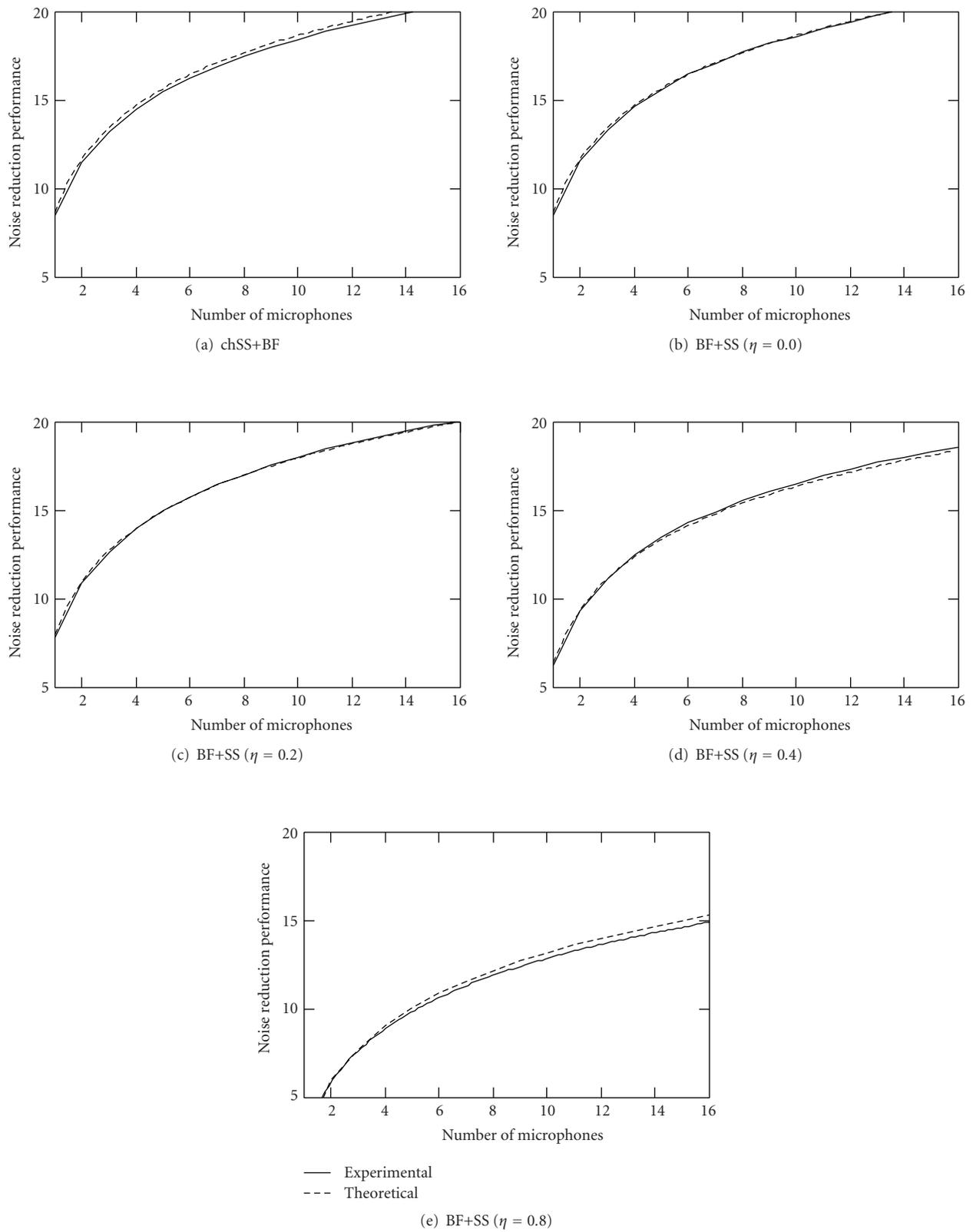


FIGURE 17: Comparison between experimental and theoretical noise reduction performances for Gaussian input signal.

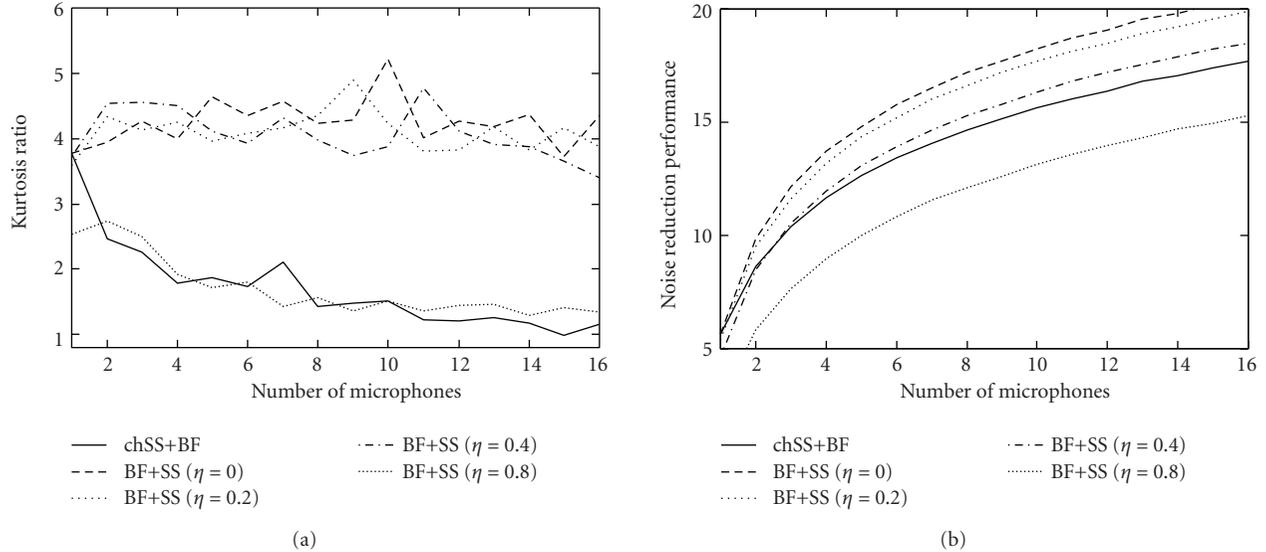


FIGURE 18: Results for super-Gaussian input signal. (a) Kurtosis ratio and (b) noise reduction performance for BF+SS with various flooring parameters.

TABLE 1: Speech distortion comparison of chSS+BF and BF+SS on the basis of CD for four-microphone case.

Input noise type	chSS+BF	BF+SS
Gaussian	6.15 dB	6.45 dB
Super-Gaussian	6.17 dB	5.12 dB

is made under the condition that the noise reduction performances of both methods are almost the same. For the Gaussian input signal, the same parameters  $\beta = 2.0$  and  $\eta = 0.0$  are utilized for BF+SS and chSS+BF. On the other hand,  $\beta = 2.0$  and  $\eta = 0.4$  are utilized for BF+SS and  $\beta = 2.0$  and  $\eta = 0.0$  are utilized for chSS+BF for the super-Gaussian input signal. Table 1 shows the result of the comparison, from which we can see that the amount of speech distortion originating from BF+SS and chSS+BF is almost the same for the Gaussian input signal. For the super-Gaussian input signal, the speech distortion originating from BF+SS is less than that from chSS+BF. This is owing to the difference in the flooring parameter for each method.

In conclusion, all of these results are strong evidence for the validity of the analysis in Sections 4, 5, and 6. These results suggest the following.

- (i) Although BF+SS can reduce the amount of musical noise by employing a larger flooring parameter, it leads to a deterioration of the noise reduction performance.
- (ii) In contrast, chSS+BF can reduce the kurtosis ratio, which corresponds to the amount of musical noise generated, without degradation of the noise reduction performance.

- (iii) Under the same level of noise reduction performance, the amount of musical noise generated via chSS+BF is less than that generated via BF+SS.
- (iv) Thus, the chSS+BF structure is preferable from the viewpoint of musical-noise generation.
- (v) However, the noise reduction performance of BF+SS is superior to that of chSS+BF for a super-Gaussian signal when the same parameters are set in the SS part for both methods.
- (vi) These results imply a trade-off between the amount of musical noise generated and the noise reduction performance. Thus, we should use an appropriate structure depending on the application.

These results should be applicable under different SNR conditions because our analysis is independent of the noise level. In the case of more reverberation, the observed signal tends to become Gaussian because many reverberant components are mixed. Therefore, the behavior of both methods under more reverberant conditions should be similar to that in the case of a Gaussian signal.

**7.2. Subjective Evaluation.** Next, we conduct a subjective evaluation to confirm that chSS+BF can mitigate musical noise. In the evaluation, we presented two signals processed by BF+SS and by chSS+BF to seven male examinees in random order, who were asked to select which signal they considered to contain less musical noise (the so-called AB method). Moreover, we instructed examinees to evaluate only the musical noise and not to consider the amplitude of the remaining noise. Here, the flooring parameter in BF+SS was automatically determined so that the output SNR of

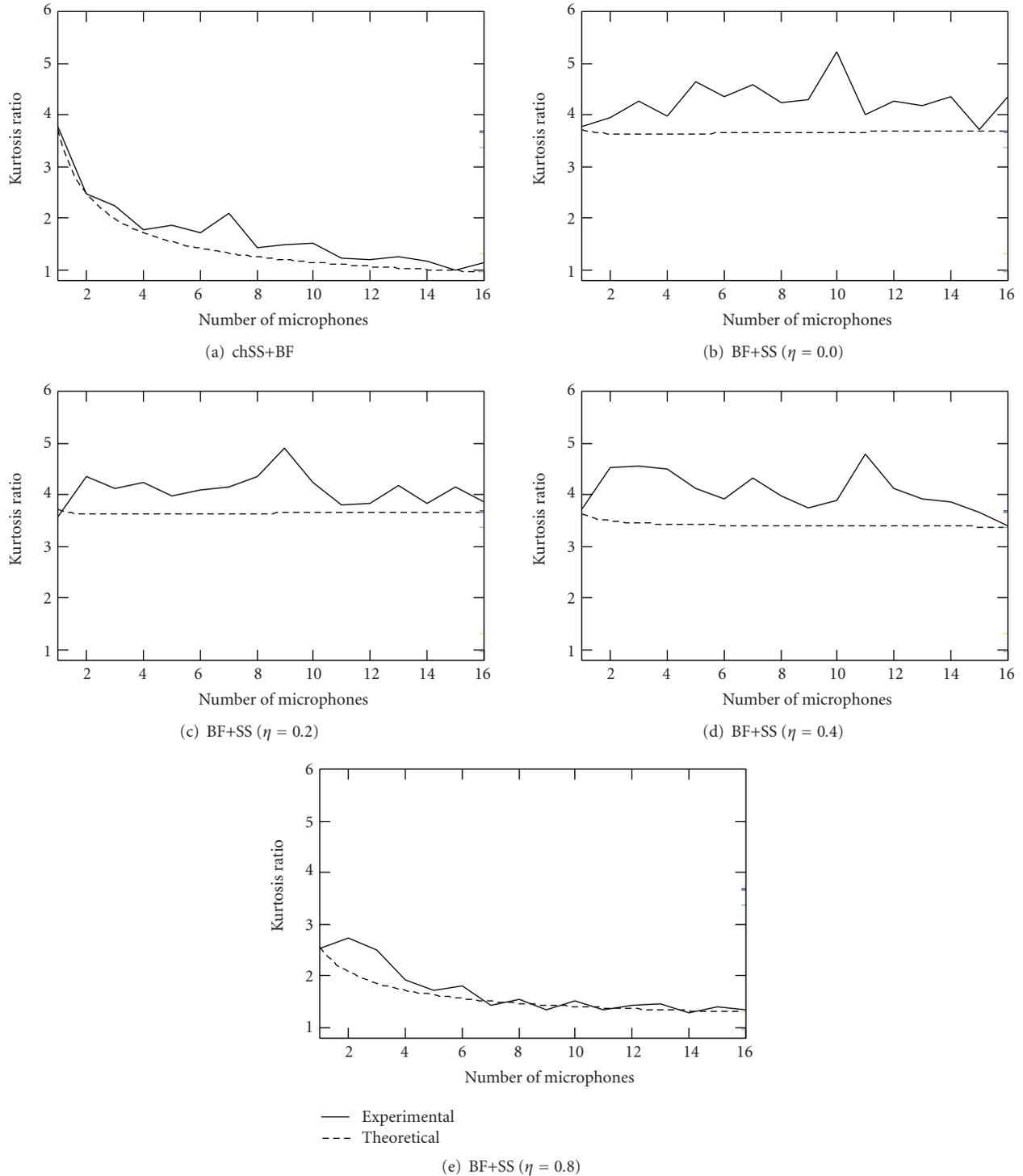


FIGURE 19: Comparison between experimental and theoretical kurtosis ratios for super-Gaussian input signal.

BF+SS and chSS+BF was equivalent. We used the preference score as the index of the evaluation, which is the frequency of the selected signal.

In the experiment, three types of noise, (a) artificial spatially uncorrelated white Gaussian noise, (b) recorded railway-station noise emitted from 36 loudspeakers, and (c) recorded human speech emitted from 36 loudspeakers, were

used. Note that noises (b) and (c) were recorded in the actual room shown in Figure 14 and therefore include interchannel correlation because they were recordings of actual noise signals.

Each test sample is a 16-kHz-sampled signal, and the target speech is the original speech convoluted with impulse responses recorded in a room with 200 millisecond

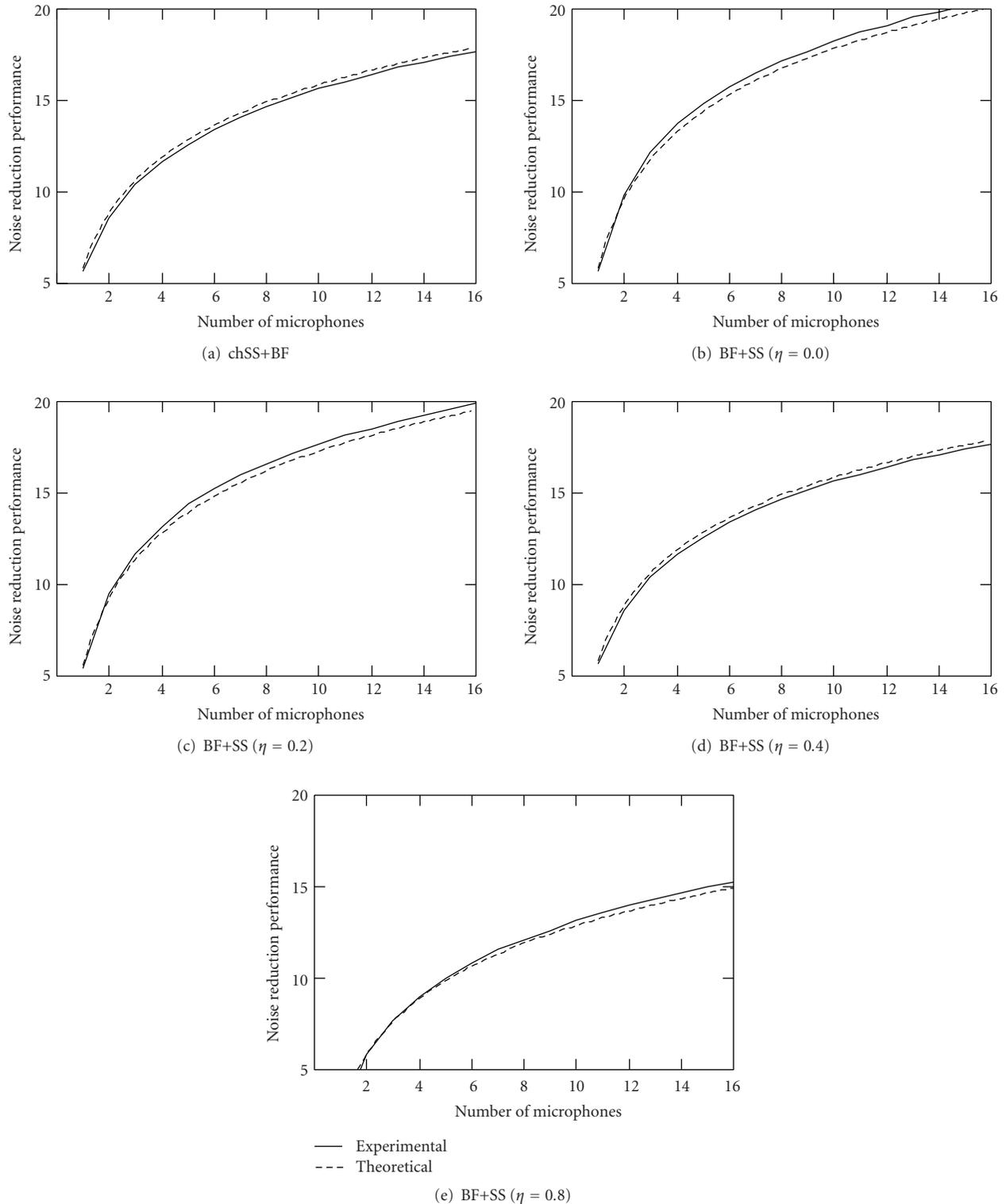


FIGURE 20: Comparison between experimental and theoretical noise reduction performances for super-Gaussian input signal.

reverberation (see Figure 14) and to which the above-mentioned recorded noise signal is added. Ten pairs of signals per type of noise, that is, a total of 30 pairs of processed signals, were presented to each examinee.

Figure 21 shows the subjective evaluation results, which confirm that the output of chSS+BF is preferred to that of BF+SS, even for actual acoustic noises including non-Gaussianity and interchannel correlation properties.

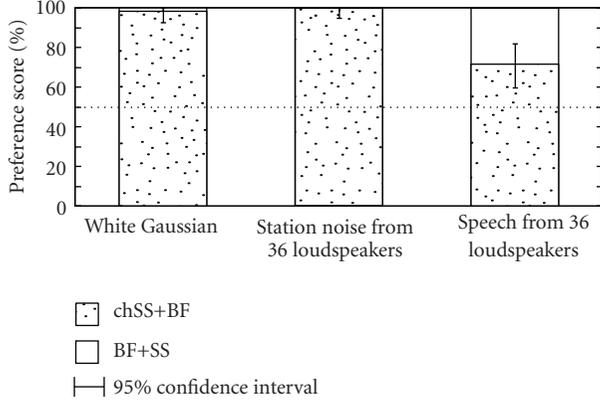


FIGURE 21: Subjective evaluation results.

## 8. Conclusion

In this paper, we analyze two methods of integrating microphone array signal processing and SS, that is, BF+SS and chSS+BF, on the basis of HOS. As a result of the analysis, it is revealed that the amount of musical noise generated via SS strongly depends on the statistical characteristics of the input signal. Moreover, it is also clarified that the noise reduction performances of BF+SS and chSS+BF are different except in the case of a Gaussian input signal. As a result of our analysis under equivalent noise reduction performance conditions, it is shown that chSS+BF reduces musical noise more than BF+SS in almost all practical cases. The results of a computer simulation also support the validity of our analysis. Moreover, by carrying out a subjective evaluation, it is confirmed that the output of chSS+BF is considered to contain less musical noise than that of BF+SS. These analytic and experimental results imply the considerable potential of optimization based on HOS to reduce musical noise.

As a future work, it remains necessary to carry out signal analysis based on more general distributions. For instance, analysis using a generalized gamma distribution [26, 27] can lead to more general results. Moreover, an exact formulation of how kurtosis is changed through DS under a coherent condition is still an open problem. Furthermore, the robustness of BF+SS and chSS+BF against low-SNR or more reverberant conditions is not discussed in this paper. In the future, the discussion should involve not only noise reduction performance and musical-noise generation but also such robustness.

## Appendices

### A. Derivation of (13)

When we assume that the input signal of the power domain can be modeled by a gamma distribution, the amount of subtraction is  $\beta\alpha\theta$ . The subtraction of the estimated noise power spectrum in each frequency subband can be considered as a lateral shift of the p.d.f. to the zero-power direction (see Figure 5). As a result of this subtraction, the

random variable  $x$  is replaced with  $x + \beta\alpha\theta$  and the gamma distribution becomes

$$\hat{P}_{\text{GM}}(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \cdot (x + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{x + \beta\alpha\theta}{\theta}\right\} \quad (\text{A.1})$$

$$(x \geq -\beta\alpha\theta).$$

Since the domain of the original gamma distribution is  $x \geq 0$ , the domain of the resultant p.d.f. is  $x \geq -\beta\alpha\theta$ . Thus, negative-power components with nonzero probability arise, which can be represented by

$$\hat{P}_{\text{negative}}(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \cdot (x + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{x + \beta\alpha\theta}{\theta}\right\}$$

$$(-\beta\alpha\theta \leq x \leq 0), \quad (\text{A.2})$$

where  $\hat{P}_{\text{negative}}(x)$  is part of  $\hat{P}_{\text{GM}}(x)$ . To remove the negative-power components, the signals corresponding to  $\hat{P}_{\text{negative}}(x)$  are replaced by observations multiplied by a small positive value  $\eta$ . The observations corresponding to (A.2),  $\hat{P}_{\text{obs}}(x)$ , are given by

$$\hat{P}_{\text{obs}}(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \cdot (x)^{\alpha-1} \exp\left\{-\frac{x}{\theta}\right\} \quad (0 \leq x \leq \beta\alpha\theta). \quad (\text{A.3})$$

Since a small positive flooring parameter  $\eta$  is applied to (A.3), the scale parameter  $\theta$  becomes  $\eta^2\theta$  and the range is changed from  $0 \leq x \leq \beta\alpha\theta$  to  $0 \leq x \leq \beta\alpha\eta^2\theta$ . Then, (A.3) is modified to

$$\hat{P}_{\text{floor}}(x) = \frac{1}{\Gamma(\alpha)(\eta^2\theta)^\alpha} \cdot (x)^{\alpha-1} \exp\left\{-\frac{x}{\eta^2\theta}\right\} \quad (\text{A.4})$$

$$(0 \leq x \leq \beta\alpha\eta^2\theta),$$

where  $\hat{P}_{\text{floor}}(x)$  is the probability of the floored components. This  $\hat{P}_{\text{floor}}(x)$  is superimposed on the p.d.f. given by (A.1) within the range  $0 \leq x \leq \beta\alpha\eta^2\theta$ . By considering the positive range of (A.1) and  $\hat{P}_{\text{floor}}(x)$ , the resultant p.d.f. of SS can be formulated as

$$P_{\text{SS}}(z) = \begin{cases} \frac{1}{\theta^\alpha\Gamma(\alpha)} (z + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{z + \beta\alpha\theta}{\theta}\right\} & (z \geq \beta\alpha\eta^2\theta), \\ \frac{1}{\theta^\alpha\Gamma(\alpha)} (z + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{z + \beta\alpha\theta}{\theta}\right\} \\ + \frac{1}{(\eta^2\theta)^\alpha\Gamma(\alpha)} z^{\alpha-1} \exp\left\{-\frac{z}{\eta^2\theta}\right\} & (0 < z < \beta\alpha\eta^2\theta), \end{cases} \quad (\text{A.5})$$

where the variable  $x$  is replaced with  $z$  for convenience.

## B. Derivation of (14)

To derive the kurtosis after SS, the 2nd- and 4th-order moments of  $z$  are required. For  $P_{SS}(z)$ , the 2nd-order moment is given by

$$\begin{aligned}\mu_2 &= \int_0^\infty z^2 \cdot P_{SS}(z) dz \\ &= \int_0^\infty z^2 \frac{1}{\theta^\alpha \Gamma(\alpha)} (z + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{z + \beta\alpha\theta}{\theta}\right\} dz \quad (\text{B.1}) \\ &\quad + \int_0^{\beta\alpha\eta^2\theta} z^2 \frac{1}{(\eta^2\theta)^\alpha \Gamma(\alpha)} z^{\alpha-1} \exp\left\{-\frac{z}{\eta^2\theta}\right\} dz.\end{aligned}$$

We now expand the first term of the right-hand side of (B.1). Here, let  $t = (z + \beta\alpha\theta)/\theta$ ; then  $\theta dt = dz$  and  $z = \theta(t - \beta\alpha)$ . Consequently,

$$\begin{aligned}\int_0^\infty z^2 \frac{1}{\theta^\alpha \Gamma(\alpha)} (z + \beta\alpha\theta)^{\alpha-1} \exp\left\{-\frac{z + \beta\alpha\theta}{\theta}\right\} dz \\ &= \int_{\beta\alpha}^\infty \theta^2 (t - \beta\alpha)^2 \frac{1}{\theta^\alpha \Gamma(\alpha)} (\theta t)^{\alpha-1} \exp\{-t\} \theta dt \\ &= \frac{\theta^2}{\Gamma(\alpha)} \int_{\beta\alpha}^\infty (t^2 - 2\beta\alpha t + \beta^2\alpha^2) t^{\alpha-1} \exp\{-t\} dt \\ &= \frac{\theta^2}{\Gamma(\alpha)} [\Gamma(\beta\alpha, \alpha + 2) - 2\beta\alpha\Gamma(\beta\alpha, \alpha + 1) + \beta^2\alpha^2\Gamma(\beta\alpha, \alpha)].\end{aligned} \quad (\text{B.2})$$

Next we consider the second term of the right-hand side of (B.1). Here, let  $t = z/(\eta^2\theta)$ ; then  $\eta^2\theta dt = dz$ . Thus,

$$\begin{aligned}\int_0^{\beta\alpha\eta^2\theta} z^2 \frac{1}{(\eta^2\theta)^\alpha \Gamma(\alpha)} z^{\alpha-1} \exp\left\{-\frac{z}{\eta^2\theta}\right\} dz \\ &= \int_0^{\beta\alpha} (\eta^2\theta t)^2 \frac{1}{(\eta^2\theta)^\alpha \Gamma(\alpha)} (\eta^2\theta t)^{\alpha-1} \exp\{-t\} \eta^2\theta dt \\ &= \frac{\eta^4\theta^2}{\Gamma(\alpha)} \int_0^{\beta\alpha} t^{\alpha+1} \exp\{-t\} dt = \eta^4\theta^2 \frac{\gamma(\beta\alpha, \alpha + 2)}{\Gamma(\alpha)}.\end{aligned} \quad (\text{B.3})$$

As a result, the 2nd-order moment after SS,  $\mu_2^{(SS)}$ , is a composite of (B.2) and (B.3) and is given as

$$\begin{aligned}\mu_2^{(SS)} &= \frac{\theta^2}{\Gamma(\alpha)} [\Gamma(\beta\alpha, \alpha + 2) - 2\beta\alpha\Gamma(\beta\alpha, \alpha + 1) \\ &\quad + \beta^2\alpha^2\Gamma(\beta\alpha, \alpha) + \eta^4\gamma(\beta\alpha, \alpha + 2)].\end{aligned} \quad (\text{B.4})$$

In the same manner, the 4th-order moment after SS,  $\mu_4^{(SS)}$ , can be represented by

$$\begin{aligned}\mu_4^{(SS)} &= \frac{\theta^4}{\Gamma(\alpha)} [\Gamma(\beta\alpha, \alpha + 4) - 4\beta\alpha\Gamma(\beta\alpha, \alpha + 3) \\ &\quad + 6\beta^2\alpha^2\Gamma(\beta\alpha, \alpha + 2) - 4\beta^3\alpha^3\Gamma(\beta\alpha, \alpha + 1) \\ &\quad + \beta^4\alpha^4\Gamma(\beta\alpha, \alpha) + \eta^8\gamma(\beta\alpha, \alpha + 4)].\end{aligned} \quad (\text{B.5})$$

Consequently, using (B.4) and (B.5), the kurtosis after SS is given as

$$\text{kurt}_{SS} = \Gamma(\alpha) \frac{\mathcal{F}(\alpha, \beta, \eta)}{\mathcal{G}^2(\alpha, \beta, \eta)}, \quad (\text{B.6})$$

where

$$\begin{aligned}\mathcal{G}(\alpha, \beta, \eta) &= \Gamma(\alpha)\Gamma(\beta\alpha, \alpha + 2) - 2\beta\alpha\Gamma(\beta\alpha, \alpha + 1) \\ &\quad + \beta^2\alpha^2\Gamma(\beta\alpha, \alpha) + \eta^4\gamma(\beta\alpha, \alpha + 2), \\ \mathcal{F}(\alpha, \beta, \eta) &= \Gamma(\beta\alpha, \alpha + 4) - 4\beta\alpha\Gamma(\beta\alpha, \alpha + 3) \\ &\quad + 6\beta^2\alpha^2\Gamma(\beta\alpha, \alpha + 2) - 4\beta^3\alpha^3\Gamma(\beta\alpha, \alpha + 1) \\ &\quad + \beta^4\alpha^4\Gamma(\beta\alpha, \alpha) + \eta^8\gamma(\beta\alpha, \alpha + 4).\end{aligned} \quad (\text{B.7})$$

## C. Derivation of (22)

As described in (12), the power-domain signal is the sum of two squares of random variables with the same distribution. Using (18), the power-domain cumulants  $K_n^{(p)}$  can be written as

$$\text{power-domain cumulants} \begin{cases} K_1^{(p)} = 2K_1^{(2)}, \\ K_2^{(p)} = 2K_2^{(2)}, \\ K_3^{(p)} = 2K_3^{(2)}, \\ K_4^{(p)} = 2K_4^{(2)}, \end{cases} \quad (\text{C.1})$$

where  $K_n^{(2)}$  is the  $n$ th square-domain moment. Here, the p.d.f. of such a square-domain signal is not symmetrical and its mean is not zero. Thus, we utilize the following relations between the moments and cumulants around the origin:

$$\text{moments} \begin{cases} \mu_1 = \kappa_1, \\ \mu_2 = \kappa_2 + \kappa_1^2, \\ \mu_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4, \end{cases} \quad (\text{C.2})$$

where  $\mu_n$  is the  $n$ th-order raw moment and  $\kappa_n$  is the  $n$ th-order cumulant. Moreover, the square-domain moments  $\mu_n^{(2)}$  can be expressed by

$$\text{squared-domain moments} \begin{cases} \mu_1^{(2)} = \mu_2, \\ \mu_2^{(2)} = \mu_4, \\ \mu_4^{(2)} = \mu_8. \end{cases} \quad (\text{C.3})$$

Using (C.1)–(C.3), the power-domain moments can be expressed in terms of the 4th- and 8th-order moments in the time domain. Therefore, to obtain the kurtosis after DS in the power domain, the moments and cumulants after DS up to the 8th order are needed.

The 3rd-, 5th-, and 7th-order cumulants are zero because we assume that the p.d.f. of  $x_j$  is symmetrical and that its mean is zero. If these conditions are satisfied, the following relations between moments and cumulants hold:

$$\text{moments} \begin{cases} \mu_1 = 0, \\ \mu_2 = \kappa_2, \\ \mu_4 = \kappa_4 + 3\kappa_2^2, \\ \mu_6 = \kappa_6 + 15\kappa_4\kappa_2 + 15\kappa_2^3, \\ \mu_8 = \kappa_8 + 35\kappa_4^2 + 28\kappa_2\kappa_6 + 210\kappa_2^2\kappa_4 + 105\kappa_2^4. \end{cases} \quad (\text{C.4})$$

Using (21) and (C.4), the time-domain moments after DS are expressed as

$$\text{moments after DS} \begin{cases} \mu_2^{(\text{DS})} = \mathcal{K}_2, \\ \mu_4^{(\text{DS})} = \mathcal{K}_4 + 3\mathcal{K}_2^2, \\ \mu_6^{(\text{DS})} = \mathcal{K}_6 + 15\mathcal{K}_2\mathcal{K}_4 + 15\mathcal{K}_2^3, \\ \mu_8^{(\text{DS})} = \mathcal{K}_8 + 35\mathcal{K}_4^2 + 28\mathcal{K}_2\mathcal{K}_6 \\ \quad + 210\mathcal{K}_2^2\mathcal{K}_4 + 105\mathcal{K}_2^4, \end{cases} \quad (\text{C.5})$$

where  $\mu_n^{(\text{DS})}$  is the  $n$ th-order raw moment after DS in the time domain.

Using (C.2), (C.3), and (C.5), the square-domain cumulants can be written as

$$\text{square-domain cumulants} \begin{cases} \mathcal{K}_1^{(2)} = \mathcal{K}_2, \\ \mathcal{K}_2^{(2)} = \mathcal{K}_4 + 2\mathcal{K}_2^2, \\ \mathcal{K}_3^{(2)} = \mathcal{K}_6 + 12\mathcal{K}_2\mathcal{K}_4 + 8\mathcal{K}_2^3, \\ \mathcal{K}_4^{(2)} = \mathcal{K}_8 + 32\mathcal{K}_4^2 + 24\mathcal{K}_2\mathcal{K}_6 \\ \quad + 144\mathcal{K}_2^2\mathcal{K}_4 + 48\mathcal{K}_2^4, \end{cases} \quad (\text{C.6})$$

where  $\mathcal{K}_n^{(2)}$  is the  $n$ th-order cumulant in the square domain.

Moreover, using (C.1), (C.2), and (C.6), the 2nd- and 4th-order power-domain moments can be written as

$$\begin{aligned} \mu_2^{(\text{p})} &= 2(\mathcal{K}_4 + 4\mathcal{K}_2^2), \\ \mu_4^{(\text{p})} &= 2(\mathcal{K}_8 + 38\mathcal{K}_4^2 + 32\mathcal{K}_2\mathcal{K}_6 + 288\mathcal{K}_2\mathcal{K}_4^2 + 192\mathcal{K}_2^4). \end{aligned} \quad (\text{C.7})$$

As a result, the power-domain kurtosis after DS,  $\text{kurt}_{\text{DS}}$ , is given as

$$\text{kurt}_{\text{DS}} = \frac{\mathcal{K}_8 + 38\mathcal{K}_4^2 + 32\mathcal{K}_2\mathcal{K}_6 + 288\mathcal{K}_2^2\mathcal{K}_4 + 192\mathcal{K}_2^4}{2\mathcal{K}_4^2 + 16\mathcal{K}_2^2\mathcal{K}_4 + 32\mathcal{K}_2^4}. \quad (\text{C.8})$$

## D. Derivation of (24)

According to (11), the shape parameter  $\hat{\alpha}$  corresponding to the kurtosis after DS,  $\text{kurt}_{\text{DS}}$ , is given by the solution of the quadratic equation:

$$\text{kurt}_{\text{DS}} = \frac{(\hat{\alpha} + 2)(\hat{\alpha} + 3)}{\hat{\alpha}(\hat{\alpha} + 1)}. \quad (\text{D.1})$$

This can be expanded as

$$\hat{\alpha}^2(\text{kurt}_{\text{DS}} - 1) + \hat{\alpha}(\text{kurt}_{\text{DS}} - 5) - 6 = 0. \quad (\text{D.2})$$

Using the quadratic formula,

$$\hat{\alpha} = \frac{-\text{kurt}_{\text{DS}} + 5 \pm \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1}}{2\text{kurt}_{\text{DS}} - 2}, \quad (\text{D.3})$$

whose denominator is larger than zero because  $\text{kurt}_{\text{DS}} > 1$ . Here, since  $\hat{\alpha} > 0$ , we must select the appropriate numerator of (D.3). First, suppose that

$$-\text{kurt}_{\text{DS}} + 5 + \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1} > 0. \quad (\text{D.4})$$

This inequality clearly holds when  $1 < \text{kurt}_{\text{DS}} < 5$  because  $-\text{kurt}_{\text{DS}} + 5 > 0$  and  $\sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1} > 0$ . Thus,

$$-\text{kurt}_{\text{DS}} + 5 > -\sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1}. \quad (\text{D.5})$$

When  $\text{kurt}_{\text{DS}} \geq 5$ , the following relation also holds:

$$\begin{aligned} (-\text{kurt}_{\text{DS}} + 5)^2 &< \text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1, \\ \Leftrightarrow 24\text{kurt}_{\text{DS}} &> 24. \end{aligned} \quad (\text{D.6})$$

Since (D.6) is true when  $\text{kurt}_{\text{DS}} \geq 5$ , (D.4) holds. In summary, (D.4) always holds for  $1 < \text{kurt}_{\text{DS}} < 5$  and  $5 \leq \text{kurt}_{\text{DS}}$ . Thus,

$$-\text{kurt}_{\text{DS}} + 5 + \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1} > 0 \quad \text{for } \text{kurt}_{\text{DS}} > 1. \quad (\text{D.7})$$

Overall,

$$\frac{-\text{kurt}_{\text{DS}} + 5 + \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1}}{2\text{kurt}_{\text{DS}} - 2} > 0. \quad (\text{D.8})$$

On the other hand, let

$$-\text{kurt}_{\text{DS}} + 5 - \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1} > 0. \quad (\text{D.9})$$

This inequality is *not* satisfied when  $\text{kurt}_{\text{DS}} > 5$  because  $-\text{kurt}_{\text{DS}} + 5 < 0$  and  $\sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1} > 0$ . Now (D.9) can be modified as

$$-\text{kurt}_{\text{DS}} + 5 > \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1}, \quad (\text{D.10})$$

then the following relation also holds for  $1 < \text{kurt}_{\text{DS}} \leq 5$ :

$$\begin{aligned} (-\text{kurt}_{\text{DS}} + 5)^2 &> \text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1, \\ \Leftrightarrow 24\text{kurt}_{\text{DS}} &< 24. \end{aligned} \quad (\text{D.11})$$

This is *not* true for  $1 < \text{kurt}_{\text{DS}} \leq 5$ . Thus, (D.9) is not appropriate for  $\text{kurt}_{\text{DS}} > 1$ . Therefore,  $\hat{\alpha}$  corresponding to  $\text{kurt}_{\text{DS}}$  is given by

$$\hat{\alpha} = \frac{-\text{kurt}_{\text{DS}} + 5 + \sqrt{\text{kurt}_{\text{DS}}^2 + 14\text{kurt}_{\text{DS}} + 1}}{2\text{kurt}_{\text{DS}} - 2}. \quad (\text{D.12})$$

## E. Derivation of (38)

For  $0 < \alpha \leq 1$ , which corresponds to a Gaussian or super-Gaussian input signal, it is revealed that the noise reduction performance of BF+SS is superior to that of chSS+BF from the numerical simulation in Section 5.3. Thus, the following relation holds:

$$\begin{aligned} & -10 \log_{10} \frac{1}{J \cdot \Gamma(\hat{\alpha})} \\ & \times \left[ \frac{\Gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} - \beta \cdot \Gamma(\beta\hat{\alpha}, \hat{\alpha}) + \eta^2 \frac{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} \right] \\ & \geq -10 \log_{10} \frac{1}{J \cdot \Gamma(\alpha)} \\ & \times \left[ \frac{\Gamma(\beta\alpha, \alpha + 1)}{\alpha} - \beta \cdot \Gamma(\beta\alpha, \alpha) + \eta^2 \frac{\gamma(\beta\alpha, \alpha + 1)}{\alpha} \right]. \end{aligned} \quad (\text{E.1})$$

This inequality corresponds to

$$\begin{aligned} & \frac{1}{\Gamma(\hat{\alpha})} \left[ \frac{\Gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} - \beta \cdot \Gamma(\beta\hat{\alpha}, \hat{\alpha}) + \eta^2 \frac{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} \right] \\ & \leq \frac{1}{\Gamma(\alpha)} \left[ \frac{\Gamma(\beta\alpha, \alpha + 1)}{\alpha} - \beta \cdot \Gamma(\beta\alpha, \alpha) + \eta^2 \frac{\gamma(\beta\alpha, \alpha + 1)}{\alpha} \right]. \end{aligned} \quad (\text{E.2})$$

Then, the new flooring parameter  $\hat{\eta}$  in BF+SS, which makes the noise reduction performance of BF+SS equal to that of chSS+BF, satisfies  $\hat{\eta} \geq \eta$  ( $\geq 0$ ) because

$$\frac{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} \geq 0. \quad (\text{E.3})$$

Moreover, the following relation for  $\hat{\eta}$  also holds:

$$\begin{aligned} & \frac{1}{\Gamma(\hat{\alpha})} \left[ \frac{\Gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} - \beta \cdot \Gamma(\beta\hat{\alpha}, \hat{\alpha}) + \hat{\eta}^2 \frac{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} \right] \\ & = \frac{1}{\Gamma(\alpha)} \left[ \frac{\Gamma(\beta\alpha, \alpha + 1)}{\alpha} - \beta \cdot \Gamma(\beta\alpha, \alpha) + \eta^2 \frac{\gamma(\beta\alpha, \alpha + 1)}{\alpha} \right]. \end{aligned} \quad (\text{E.4})$$

This can be rewritten as

$$\begin{aligned} & \hat{\eta}^2 \frac{\Gamma(\alpha)}{\Gamma(\hat{\alpha})} \frac{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} \\ & = \left[ \frac{\Gamma(\beta\alpha, \alpha + 1)}{\alpha} - \beta \cdot \Gamma(\beta\alpha, \alpha) + \eta^2 \frac{\gamma(\beta\alpha, \alpha + 1)}{\alpha} \right] \\ & \quad - \frac{\Gamma(\alpha)}{\Gamma(\hat{\alpha})} \left[ \frac{\Gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)}{\hat{\alpha}} - \beta \cdot \Gamma(\beta\hat{\alpha}, \hat{\alpha}) \right], \end{aligned} \quad (\text{E.5})$$

and consequently

$$\hat{\eta}^2 = \frac{\hat{\alpha}}{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)} \left[ \frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha)} \mathcal{H}(\alpha, \beta, \eta) - \mathcal{L}(\hat{\alpha}, \beta) \right], \quad (\text{E.6})$$

where  $\mathcal{H}(\alpha, \beta, \eta)$  is defined by (39) and  $\mathcal{L}(\alpha, \beta)$  is given by (40). Using (E.3) and (E.4), the right-hand side of (E.5) is clearly greater than or equal to zero. Moreover, since  $\Gamma(\alpha) > 0$ ,  $\Gamma(\hat{\alpha}) > 0$ ,  $\hat{\alpha} > 0$ , and  $\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1) > 0$ , the right-hand side of (E.6) is also greater than or equal to zero. Therefore,

$$\hat{\eta} = \sqrt{\frac{\hat{\alpha}}{\gamma(\beta\hat{\alpha}, \hat{\alpha} + 1)} \cdot \left[ \frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha)} \mathcal{H}(\alpha, \beta, \eta) - \mathcal{L}(\hat{\alpha}, \beta) \right]}. \quad (\text{E.7})$$

## Acknowledgment

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

## References

- [1] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin, Germany, 2001.
- [2] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1508–1518, 1985.
- [3] M. Omologo, M. Matassoni, P. Svaizer, and D. Giuliani, "Microphone array based speech recognition with different talker-array positions," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, pp. 227–230, Munich, Germany, September 1997.
- [4] H. F. Silverman and W. R. Patterson, "Visualizing the performance of large-aperture microphone arrays," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, pp. 962–972, 1999.
- [5] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, pp. 926–935, 1972.
- [6] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.
- [7] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 6, pp. 1391–1400, 1986.
- [8] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

- [9] J. Meyer and K. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, pp. 1167–1170, 1997.
- [10] S. Fischer and K. D. Kammeyer, "Broadband beamforming with adaptive post filtering for speech acquisition in noisy environment," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, pp. 359–362, 1997.
- [11] R. Mukai, S. Araki, H. Sawada, and S. Makino, "Removal of residual cross-talk components in blind source separation using time-delayed spectral subtraction," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, pp. 1789–1792, Orlando, Fla, USA, May 2002.
- [12] J. Cho and A. Krishnamurthy, "Speech enhancement using microphone array in moving vehicle environment," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 366–371, Graz, Austria, April 2003.
- [13] Y. Ohashi, T. Nishikawa, H. Saruwatari, A. Lee, and K. Shikano, "Noise robust speech recognition based on spatial subtraction array," in *Proceedings of the International Workshop on Nonlinear Signal and Image Processing*, pp. 324–327, 2005.
- [14] J. Even, H. Saruwatari, and K. Shikano, "New architecture combining blind signal extraction and modified spectral subtraction for suppression of background noise," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '08)*, Seattle, Wash, USA, 2008.
- [15] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 4, pp. 650–664, 2009.
- [16] S. B. Jebara, "A perceptual approach to reduce musical noise phenomenon with Wiener denoising technique," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, vol. 3, pp. 49–52, 2006.
- [17] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.
- [18] 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," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '08)*, Seattle, Wash, USA, 2008.
- [19] 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 estimatio," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 4433–4436, 2009.
- [20] Y. Takahashi, Y. Uemura, H. Saruwatari, K. Shikano, and K. Kondo, "Musical noise analysis based on higher order statistics for microphone array and nonlinear signal processing," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 229–232, 2009.
- [21] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [22] 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, no. 11, pp. 1135–1146, 2003.
- [23] M. Mizumachi and M. Akagi, "Noise reduction by paired-microphone using spectral subtraction," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 2, pp. 1001–1004, 1998.
- [24] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E87-A, no. 8, pp. 2063–2072, 2004.
- [25] S. Ikeda and N. Murata, "A method of ICA in the frequency domain," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 365–371, 1999.
- [26] E. W. Stacy, "A generalization of the gamma distribution," *The Annals of Mathematical Statistics*, pp. 1187–1192, 1962.
- [27] K. Kokkinakis and A. K. Nandi, "Generalized gamma density-based score functions for fast and flexible ICA," *Signal Processing*, vol. 87, no. 5, pp. 1156–1162, 2007.
- [28] J. W. Shin, J.-H. Chang, and N. S. Kim, "Statistical modeling of speech signals based on generalized gamma distribution," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 258–261, 2005.
- [29] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall PTR, 1993.