MUSICAL NOISE ANALYSIS BASED ON HIGHER ORDER STATISTICS FOR MICROPHONE ARRAY AND NONLINEAR SIGNAL PROCESSING

†Yu Takahashi, †Yoshihisa Uemura, †Hiroshi Saruwatari, †Kiyorhiro Shikano, and ‡Kazunobu Kondo

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

ABSTRACT

In this paper, we conduct an analysis for reduction of musical noise in integration method of microphone array signal processing and nonlinear signal processing. In these days, for better noise reduction, integration methods of microphone array signal processing and nonlinear signal processing have been researched. However, nonlinear signal processing causes musical noise. Since such musical noise make users uncomfortable, it is desired that musical noise is mitigated. Moreover, in these days, it is reported that higher-order statistics is strongly related with the amount of generated musical noise. Thus, we analyze the integrated method of microphone array signal processing and nonlinear signal processing, based on higher-order statistics. Also, we propose an architecture for reducing musical noise based on the analysis. The effectiveness of the proposed architecture and analysis correctness are shown via a computer simulation and a subjective evaluation.

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

1. INTRODUCTION

In these days, integration methods of microphone array signal processing and nonlinear signal processing have been studied for better noise reduction, e.g., [1]. It is reported that such an integration method can achieve higher noise reduction performance rather than a conventional adaptive microphone array [2], e.g., Griffith-Jim array. However, in such methods, artificial distortion (so-called musical noise) due to nonlinear signal processing arises. Since the artificial distortion makes users uncomfortable, it is desired that we take control of musical noise. However, in almost all the integration methods, to mitigate musical noise, strength of nonlinear signal processing is determined heuristically.

Recently, it is reported that the amount of generated musical noise is strongly related with the difference between higher-order statistics before/after nonlinear signal processing [3]. This fact enables us to analyze how much musical noise arises through objective nonlinear signal processing. Therefore, based on higher-order statistics, we believe that it is possible to optimize integration methods of microphone array signal processing and nonlinear signal processing from the viewpoint of not only noise reduction performance but also the sound quality. For the first step of this, we analyze the simplest case of integration of microphone array signal processing and nonlinear signal processing in this research.

Hereafter, we analyze two integration methods of microphone array signal processing and nonlinear signal processing based on higher-order statistics. Particularly, we focus on spectral subtraction (SS)[4] method, i.e., the most popular and simplest nonlinear signal processing, as a nonlinear signal processing. Figure 1 shows

This work was partly supported by MIC SCOPE project in Japan.

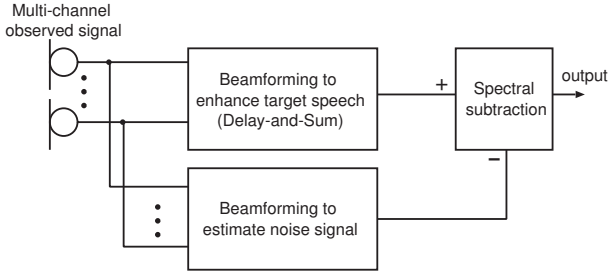

Fig. 1. Block diagram of spectral subtraction after beamforming (BF+SS).

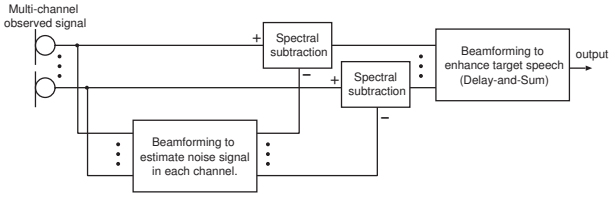

Fig. 2. Block diagram of proposed channel-wise spectral subtraction before beamforming (chSS+BF).

a typical architecture example of integration of microphone array signal processing and SS. In this architecture, SS is performed after beamforming. Thus we call this type of architecture *BF+SS*. On the other hand, we propose a new architecture illustrated in Fig. 2, which is an alternative type of integration of microphone array signal processing and SS. In this architecture, channel-wise SS is performed before beamforming. So we call this type of architecture *chSS+BF*. We analyze such two methods based on higher-order statistics, and reveal that chSS+BF can mitigate musical noise rather than BF+SS. Finally, the propriety of the analysis based on higher-order statistics is shown via a computer simulation and a subjective evaluation.

2. SPECTRAL SUBTRACTION AFTER BEAMFORMING

In BF+SS, first, the single-channel speech enhanced signal is obtained by beamforming, e.g., delay-and-sum (DS) [5]. Next, the single-channel estimated noise signal is also obtained by beamforming technique, e.g., null beamformer [6] or adaptive beamforming [5]. Finally, we obtain the speech enhanced signal based on SS. The detailed signal processing is shown below.

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

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

where $\boldsymbol{x}(f,\tau) = [x_1(f,\tau),\ldots,x_J(f,\tau)]^{\mathrm{T}}$ is the observed signal vector, $\boldsymbol{h}(f) = [h_1(f),\ldots,h_J(f)]^{\mathrm{T}}$ is the transfer function vector, $s(f,\tau)$ is the target speech, and $\boldsymbol{n}(f,\tau) = [n_1(f,\tau),\ldots,n_J(f,\tau))]^{\mathrm{T}}$ is the noise vector. For enhancing the target speech, DS is applied to the ob-

served signal. This can be represented by

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

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

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

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

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

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

3. PROPOSED METHOD AND ANALYSIS

3.1. Overview

In the proposed chSS+BF, channel-wise noise estimation is conducted firstly. Next, SS is applied to the multi-channel input signal channel-wisely. Finally, we perform DS to the SS-applied multi-channel signal to obtain the speech enhanced signal. This architecture can mitigate the musical noise (details are shown in Sect. 3.3).

3.2. Algorithm

In the proposed method, first, we perform SS in each input channel. Consequently, we obtain the multi-channel target speech enhanced signal by channel-wise SS. This can be designated as

$$|y_j^{(\mathrm{chSS})}(f,\tau)| = \begin{cases}\sqrt{|x_j(f,\tau)|^2-\beta\cdot|\tilde{n}_j(f)|^2} \\ \quad(\text{where } |x_j(f,\tau)|^2-\beta\cdot|\tilde{n}_j(f)|^2>0),\\ 0 \quad(\text{otherwise}),\end{cases} \tag{6}$$

where $y_j^{(\mathrm{chSS})}(f,\tau)$ is the target speech enhanced signal by SS at j channel, and $\tilde{n}_j(f)$ is the estimated noise signal in j channel.

Finally, we obtain the target speech enhanced signal by applying DS to $\boldsymbol{y}_{\mathrm{chSS}}(f,\tau)$. This procedure can be represented by

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

$$\boldsymbol{y}_{\mathrm{chSS}}(f,\tau) = [y_1^{(\mathrm{chSS})}(f,\tau),\ldots,y_J^{(\mathrm{chSS})}(f,\tau)]^{\mathrm{T}}, \tag{8}$$

where $y(f,\tau)$ is the final output of the proposed method.

3.3. Kurtosis based analysis

3.3.1. Analysis strategy

It has been reported by the authors that the amount of generated musical noise is strongly related with the difference between the before-and-after kurtosis of a signal in nonlinear signal processing [3]. Thus, in this section, we analyze the amount of generated musical noise through the proposed chSS+BF and BF+SS, based on kurtosis. Basically, kurtosis increases through nonlinear signal processing, and larger increment of the kurtosis by nonlinear signal processing leads to more amount of musical noise generation [3]. Thus, it can be expected that the generated musical noise becomes smaller with a lower-kurtosis-increment signal processing. In the following subsections, hence, we analyze the kurtosis of BF+SS and the proposed chSS+BF, and prove which method can reduce the resultant kurtosis. Note that our analysis has no limitation in assumption of noise model, thus any noises including Gaussian and non-Gaussian can be under consideration.

3.3.2. Kurtosis

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

$$\mathrm{kurt}_x = \frac{\mu_4}{\mu_2^2}, \tag{9}$$

where x is the probability variable, kurt_x is the kurtosis of x, and μ_n is the n-th order moment of x. Although kurt_x becomes 3 if x is Gaussian signal, note that the kurtosis of Gaussian signal in power spectral domain becomes 6. This is because Gaussian signal in time domain obeys chi-square distribution with two degrees of freedom in power spectral domain. In chi-square distribution with two degrees of freedom, $\mu_4/\mu_2^2 = 6$.

3.3.3. Resultant kurtosis in spectral subtraction [3]

In this section, we analyze the kurtosis after SS. For evaluating resultant kurtosis of SS, we utilize gamma distribution as a model of input signal in power domain [7]. The probability density function (p.d.f.) of the gamma distribution for probability variable x is defined as

$$P(x) = \Gamma^{-1}(\alpha)\,\theta^{-\alpha}\cdot x^{\,\alpha-1}\,e^{-\frac{x}{\theta}}, \tag{10}$$

where $x \geq 0$, $\alpha > 0$ and $\theta > 0$. Here, α denotes the shape parameter and θ is the scale parameter. Besides, $\Gamma(\cdot)$ is the gamma function. Gamma distribution with $\alpha = 1$ corresponds to chi-square distribution with two degrees of freedom. Moreover, it is well-known that the average of the gamma distribution is $\mathrm{E}[P(x)] = \alpha\theta$, where $\mathrm{E}[\cdot]$ is an expectation operator. Furthermore, the kurtosis of Gamma distribution, $\mathrm{kurt}_{\mathrm{GM}}$, can be designated as [3]

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

In SS, the average of observed power spectrum is utilized as an estimated noise power spectrum. So the amount of subtraction is $\beta\cdot\alpha\theta$. Subtraction of the estimated noise power spectrum in each frequency band can be regarded as deforming of the p.d.f., which is the lateral shift of the p.d.f. to zero power direction. As a result, the probability of the negative power component arises. To avoid this, such a negative component probability is replaced by zero (so-called flooring technique). The resultant p.d.f. after SS can be written as

$$P(x) = \begin{cases} C\cdot(x+\beta\cdot\alpha\theta)^{\alpha-1}\,e^{-\frac{x+\beta\cdot\alpha\theta}{\theta}} & (x>0),\\ C\int_0^{\beta\cdot\alpha\theta} x^{\alpha-1}e^{-\frac{x}{\theta}}dx & (x=0),\end{cases} \tag{12}$$

where $C = 1/[\Gamma(\alpha)\theta^{\alpha}]$. Thus, the resultant kurtosis of SS, $\mathrm{kurt}_{\mathrm{SS}}$, can be given as

$$\mathrm{kurt}_{\mathrm{SS}} \geq \frac{e^{\alpha\beta}}{\alpha(\alpha+1)}\left\{(\alpha+2)(\alpha+3)+\beta\alpha(\alpha+2)(\alpha-1)+\frac{(\beta\alpha)^2}{2}(\alpha-3)(\alpha-1)\right\}. \tag{13}$$

Although we cannot describe details of the derivation of (13) due to the limitation of the paper space, reference [3] helps you to understand the derivation of (13).

3.3.4. 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.

Now let x_j ($j = 1,\ldots,J$) be J-channel input signal, and we assume they are i.i.d. signal each other. Moreover, we assume that the p.d.f. of x_j is both side symmetry and its average is zero. These assumptions make odd order cumulants zero except the first order cumulant. For cumulants, it is well known that the following relation holds;

$$\mathrm{cum}_n(aX+bY) = a^n\,\mathrm{cum}_n(X) + b^n\,\mathrm{cum}_n(Y), \tag{14}$$

where $\mathrm{cum}_n(X)$ expresses the n-th order cumulant of probability variable X. Based on the relation (14), the resultant cumulant after DS,

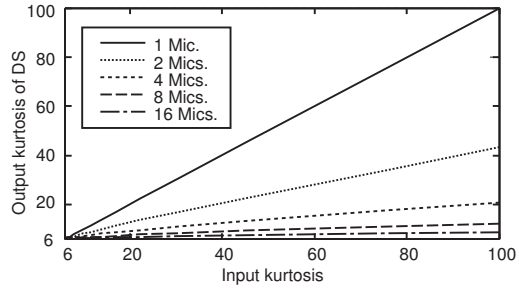

Fig. 3. Relation between input kurtosis and output kurtosis of DS.

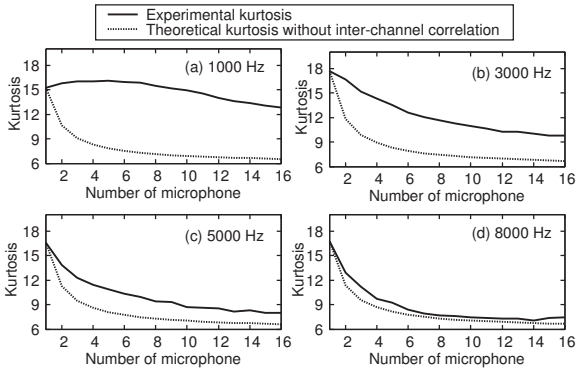

Fig. 4. Simulation result for noise with inter-channel correlation (solid line), and theoretical effect of DS considers no inter-channel correlation (dotted line) in each frequency subband.

$K_n^{(\mathrm{DS})}$, can be given by,

$$K_n^{(\mathrm{DS})} = K_n / J^{n-1}, \tag{15}$$

where K_n is the n-th order cumulant of x_j. Using (15) and well-known mathematical relation between cumulant and moment, the power-spectral-domain kurtosis of DS can be expressed by

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

Considering an actual acoustic signal and its cumulants, we can illustrate the relation between input and output kurtosis via DS as Fig. 3. This relation can be approximated as

$$\mathrm{kurt}_{\mathrm{DS}} \simeq J^{-1} \cdot (\mathrm{kurt}_{\mathrm{in}} - 6) + 6, \tag{17}$$

where $\mathrm{kurt}_{\mathrm{in}}$ is the input kurtosis. As we can see from Fig. 3, the output kurtosis decreases in proportion to the number of microphones.

When input signals have inter-channel correlation, the relation between input and output kurtosis via DS approaches to the case of only 1 microphone. If all input signals are the same signal, i.e., these signals are completely correlated, the output of DS is also the same signal. In such the case, the effect of DS corresponds to the case of only 1 microphone. In particular, inter-channel correlation is not completely unit within all frequency subbands. It is well known that the intensity of the inter-channel correlation is strong in lower frequency subbands, and is weak in higher frequency subbands [5]. Therefore, in lower frequency subbands, it can be expected that DS cannot reduce the kurtosis of the signal well.

Figure 4 shows the preliminary simulation result of DS. In this preliminary simulation, first, SS is applied to multi-channel Gaussian signal with actual inter-channel correlation. Next, DS is applied to such the spectral-subtraction-applied signal. From this result, we can confirm that the above mentioned fact, i.e., the effect of DS is weak in lower frequency subbands. Indeed the effect of DS becomes weak, note that the effect is not lost completely in lower frequency subbands. Also, we can see that theoretical kurtosis curve is proper to the actual result in higher frequency subbands. This is because that inter-channel correlation is weak in higher frequency subband. Consequently, DS can reduce the kurtosis of the input signal even if inter-channel correlation exists.

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

In the previous subsections, we have discussed the resultant kurtosis of SS and DS. In this subsection, we discuss the resultant kurtosis of the proposed chSS+BF and BF+SS. As described in Sect. 3.3.1, it can be expected that the smaller kurtosis increment leads to the less amount of generated musical noise.

In BF+SS, first, DS is applied to multi-channel input signal. At this point, the resultant kurtosis in power spectral domain, $\mathrm{kurt}_{\mathrm{DS}}$, is

$$\mathrm{kurt}_{\mathrm{DS}} = J^{-1} \cdot (\mathrm{kurt}_{\mathrm{in}} - 6) + 6, \tag{18}$$

where $\mathrm{kurt}_{\mathrm{in}}$ is the kurtosis of the input signal in power spectral domain. Using (11), we can derive a shape parameter of gamma distribution corresponds to $\mathrm{kurt}_{\mathrm{DS}}$ as

$$\hat{\alpha} = \frac{\sqrt{\mathrm{kurt}_{\mathrm{DS}}^2 + 14\,\mathrm{kurt}_{\mathrm{DS}} + 1} - \mathrm{kurt}_{\mathrm{DS}} + 5}{2\,\mathrm{kurt}_{\mathrm{DS}} - 2}, \tag{19}$$

where $\hat{\alpha}$ is the shape parameter of gamma distribution corresponds to $\mathrm{kurt}_{\mathrm{DS}}$. Consequently, using (13), the resultant kurtosis of BF+SS, $\mathrm{kurt}_{\mathrm{BF+SS}}$, can be written as

$$\mathrm{kurt}_{\mathrm{BF+SS}} \geq \frac{e^{\hat{\alpha}\beta}}{\hat{\alpha}(\hat{\alpha}+1)}\left\{(\hat{\alpha}+2)(\hat{\alpha}+3) + \beta\hat{\alpha}(\hat{\alpha}+2)(\hat{\alpha}-1) + \frac{(\beta\hat{\alpha})^2}{2}(\hat{\alpha}-3)(\hat{\alpha}-1)\right\}. \tag{20}$$

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

$$\mathrm{kurt}_{\mathrm{chSS}} \geq \frac{e^{\tilde{\alpha}\beta}}{\tilde{\alpha}(\tilde{\alpha}+1)}\left\{(\tilde{\alpha}+2)(\tilde{\alpha}+3) + \beta\tilde{\alpha}(\tilde{\alpha}+2)(\tilde{\alpha}-1) + \frac{(\beta\tilde{\alpha})^2}{2}(\tilde{\alpha}-3)(\tilde{\alpha}-1)\right\}, \tag{21}$$

where $\tilde{\alpha}$ is a shape parameter of gamma distribution for the original input signal. Here, $\tilde{\alpha}$ and $\mathrm{kurt}_{\mathrm{in}}$ satisfy (11). Finally, DS is performed and its resultant kurtosis can be written as

$$\mathrm{kurt}_{\mathrm{chSS+BF}} = J^{-1} \cdot (\mathrm{kurt}_{\mathrm{chSS}} - 6) + 6, \tag{22}$$

where $\mathrm{kurt}_{\mathrm{chSS+BF}}$ is the resultant kurtosis of the proposed chSS+BF.

Here, we consider the following equation to compare the resultant kurtosis of chSS+BF and BF+SS.

$$D = \mathrm{kurt}_{\mathrm{BF+SS}} - \mathrm{kurt}_{\mathrm{chSS+BF}}, \tag{23}$$

where D expresses the difference of the output kurtosis between chSS+BF and BF+SS. Note that positive D indicates that the proposed chSS+BF reduced the resultant kurtosis compared with BF+SS. The relation about D is depicted in Fig. 5. In the figure, oversubtraction parameter β is fixed to 2. From this figure, we can confirm that the proposed chSS+BF can reduce the resultant kurtosis rather than BF+SS for almost all the input signals with various kurtosis. When input kurtosis is smaller than 4, the proposed chSS+BF cannot reduce the resultant kurtosis rather than BF+SS. However, such an input kurtosis corresponds to sub-Gaussian signal. In a common acoustical environment, such a sub-Gaussian signal cannot be expected to exist. Therefore, the proposed chSS+BF can be considered to reduce the resultant kurtosis rather than BF+SS in acoustic signals.

4. EXPERIMENT AND RESULT

4.1. Computer simulation

First, we compared BF+SS and the proposed chSS+BF in kurtosis difference and noise reduction performance. We used the following

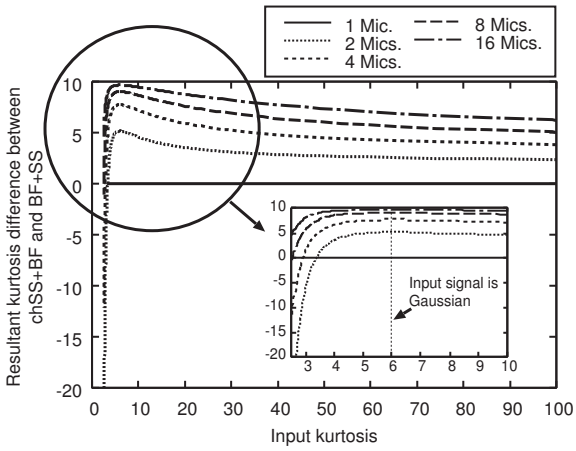

Fig. 5. Resultant kurtosis difference between chSS+BF and BF+SS.

16 kHz sampled signals as test data; the target speech is the original speech convoluted with the impulse responses which were recorded in a room with 200 ms reverberation, and to which an artificially generated spatially uncorrelated white Gaussian was added. Besides, we use 6 speakers (6 sentences) as sources of the original source. The number of microphone elements in the simulation is changed from 2 to 16. The subtraction coefficient β is set to 2.0, and the flooring parameter for BF+SS, γ, is set to 0.0, 0.1, 0.2, 0.4 and 0.8. Note that flooring is not performed in chSS+BF. In the simulation, we assume that the noise estimation is performed perfectly.

Here, we utilize the kurtosis difference as the measure for the amount of generated musical noise. This is given by

$$\text{Kurtosis difference} = \text{kurt}(n_{\text{proc}}(f,\tau)) - \text{kurt}(n_{\text{org}}(f,\tau)), \quad (24)$$

where $n_{\text{proc}}(f,\tau)$ is the power spectrum of the residual noise signal after processing, and $n_{\text{org}}(f,\tau)$ is the power spectrum of the noise signal before processing. This kurtosis difference indicates how kurtosis is increased with processing. Thus, it is desired that the kurtosis difference becomes smaller. Moreover noise reduction performance is measured based on the power of the residual noise. This is described as

$$\text{Power of residual noise [dB]} = 10\log_{10}\left\{\frac{\sum_{f,\tau}|n_{\text{proc}}(f,\tau)|^2}{\sum_{f,\tau}|n_{\text{org}}(f,\tau)|^2}\right\}. \quad (25)$$

Figure 6 shows the simulation results. From Fig. 6(a), we can see that the kurtosis difference of chSS+BF is monotonically decreasing with increasing the number of microphones. On the other hand, the kurtosis difference of BF+SS is constant regardless of the number of microphones. Indeed BF+SS with the specific flooring parameter can achieve the same kurtosis difference as chSS+BF, e.g., the case of flooring parameter 0.4 in 10 microphones. However, BF+SS with the large flooring parameter degrades the noise reduction performance itself (see Fig. 6(b)). On the other hand, the proposed chSS+BF can reduce the kurtosis difference, i.e., musical noise generation, without degradation of noise reduction performance.

4.2. Subjective evaluation

Next, we conducted a subjective evaluation to confirm that the proposed chSS+BF can mitigate the musical noise. In the evaluation, we gave two processed signals by the proposed chSS+BF and BF+SS respectively to examinees with random order, and let 7 examinees (7 males) forcedly select which signal is less amount of musical noise (so-called AB method). In the experiment, 3 types of noises, i.e., (a) artificial spatially uncorrelated white Gaussian, (b) real-recorded railway-station noise emitted from 36 loudspeakers,

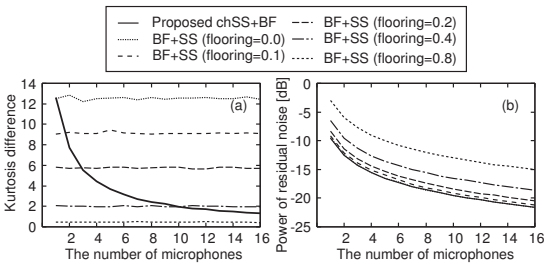

Fig. 6. Results of (a) kurtosis difference, and (b) power of residual noise, with various flooring parameters in BF+SS.

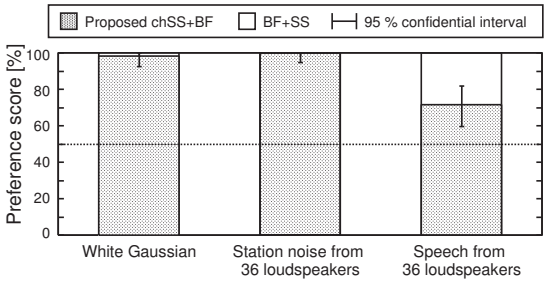

Fig. 7. Subjective evaluation results.

and (c) real-recorded human speech emitted from 36 loudspeakers, were used. Note that the noises (b) and (c) include inter-channel correlation because they were real-recorded noise signals. 10 pairs of signal per one kind of noise, totally 30 pairs of processed signal were displayed to each examinee. Figure 7 shows the subjective evaluation results, and we can confirm that the output of the proposed chSS+BF is preferred compared with that of BF+SS even for the real acoustic noises including non-Gaussianity and inter-channel correlation properties.

5. CONCLUSION

In this paper, we analyze two integrated methods of microphone array signal processing and SS, i.e., chSS+BF and BF+SS. We reveal that the proposed chSS+BF can reduce the kurtosis compared with BF+SS. Moreover, as a result of subjective evaluation, it is confirmed that the output of the proposed chSS+BF is considered as less musical noise signal compared with that of BF+SS. These analytic and experimental results imply great potential of higher-order-statistics based optimization for musical noise.

6. REFERENCES

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

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

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

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

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

[6] H. Saruwatari, et al., "Blind source separation combining independent component analysis and beamforming," *EURASIP J. Applied Signal Proc.*, vol.2003, no.11, pp.1135–1146, 2003.

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