STRUCTURE SELECTION ALGORITHM FOR LESS MUSICAL-NOISE GENERATION IN INTEGRATION SYSTEMS OF BEAMFORMING AND SPECTRAL SUBTRACTION

[†]*Yu Takahashi,* [†]*Yoshihisa Uemura,* [†]*Hiroshi Saruwatari,* [†]*Kiyohiro Shikano, and* [‡]*Kazunobu Kondo*

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

ABSTRACT

In this paper, we propose an appropriate structure selection algorithm for less musical-noise generation in integration methods of microphone array and spectral subtraction. In our previous work, we have analyzed musical-noise reduction structure in integration methods of microphone array and spectral subtraction based on higher-order statistics. However, that analysis can be applied to only Gaussian and super-Gaussian noises. In this paper, we extend the analysis into sub-Gaussian noise case. As a result of the extended analysis, we find the fact that an appropriate structure is depending on the type of input noise. Based on this fact, we propose an appropriate structure selection algorithm in integration methods of microphone array and spectral subtraction for less musical-noise generation. The effectiveness of the proposed algorithm are shown via a computer simulation.

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

1. INTRODUCTION

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

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 tc 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, strength of nonlinear signal processing is determined by hand heuristically to mitigate musical noise.

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 nonlinear signal processing. Based on the higher-order statistics, we have given a preliminary analysis of integration methods of microphone array and spectral subtraction (SS) [4] for less musical-noise generation [5]. In that analysis, we have analyzed the following two structures; spectral subtraction after beamforming (BF+SS) and channel-wise spectral subtraction *before* beamforming (chSS+BF). Figure 1 shows a block diagram of BF+SS. This is a kind of traditional integration method of microphone array and SS. In this architecture, SS is performed after beamforming. On the other hand, a block diagram of chSS+BF is depicted 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. As a result of the analysis, we have found that chSS+BF structure can reduce musical noise for various types of noise. However, the analysis was applied to only Gaussian noise or super-Gaussian noise case.

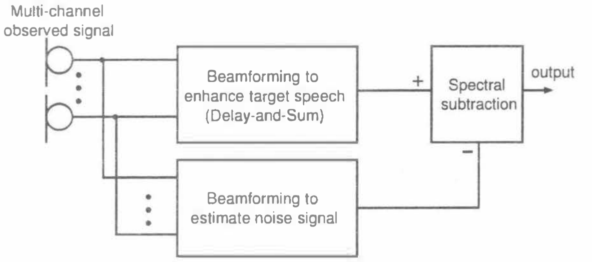

Fig. 1. Block diagram of spectral subtraction after beamforming (BF+SS).

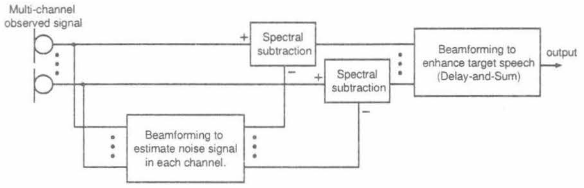

Fig. 2. Block diagram of channel-wise spectral subtraction before beamforming (chSS+BF).

In this paper, we extend the analysis of the two structures, i.e., BF+SS and chSS+BF, into sub-Gaussian noise case. As a result of the extended analysis, we clarify that the appropriate structure depends on input noise type. Based on this fact, we newly propose a structure selection algorithm for less musical-noise generation. Finally, the efficacy of the proposed selection algorithm is shown via a computer simulation.

2. INTEGRATION METHODS OF MICROPHONE ARRAY AND SPECTRAL SUBTRACTION

2.1. Spectral subtraction after beamforming

Figure 1 shows the block diagram of BF+SS. In BF+SS, first, the single-channel speech enhanced signal is obtained by beamforming, e.g., delay-and-sum (DS) [8]. Next, the single-channel estimated

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

978-1-4244-2710-9/09/$25.00 © 2009 IEEE

noise signal is also obtained by beamforming technique, e.g., null beamformer [9] or adaptive beamforming [8]. 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 $\boldsymbol{x}(f,\tau) = [x_1(f,\tau),\dots,x_J(f,\tau)]^{\mathrm{T}}$ is the observed signal vector, $\boldsymbol{h}(f) = [h_1(f),\dots,h_J(f)]^{\mathrm{T}}$ is the transfer function vector, $s(f,\tau)$ is the target speech, and $\boldsymbol{n}(f,\tau) = [n_1(f,\tau),\dots,n_J(f,\tau))]^{\mathrm{T}}$ is the noise vector. For enhancing the target speech, DS is applied to the observed signal. This can be represented by

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

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

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

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

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

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 [1, 2].

2.2. Channel-wise spectral subtraction before beamforming

Figure 2 illustrates the block diagram of chSS+BF. In chSS+BF, 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

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

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. For instance, such the multi-channel noise can be estimated by single-input multiple-output independent component analysis (SIMO-ICA) [6] or combination of ICA and projection back method [7].

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

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

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

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

3. KURTOSIS-BASED MUSICAL NOISE ANALYSIS

3.1. Analysis strategy

It has been reported by the authors that the amount of generated musical noise is strongly related with the difference between before/after kurtosis of a signal in nonlinear signal processing [3]. Thus, in this section, we analyze the amount of generated musical noise based on kurtosis. In the following subsections, we will mention that kurtosis-based analysis on DS and SS which are parts of BF+SS and chSS+BF, respectively. Also, not only Gaussian and super-Gaussian noise but also sub-Gaussian noise are analysis target unlike our previous analysis in Ref. [5]. The analysis on BF+SS and chSS+BF will be denoted in Sect. 4.1.

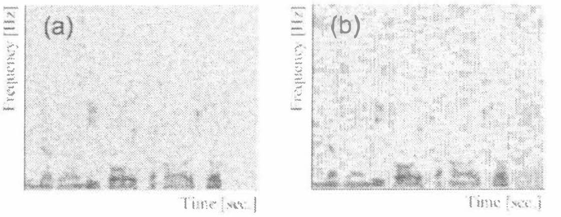

Fig. 3. (a) Observer spectrogram, and (b) processed spectrogram.

3.2. Kurtosis-based musical-noise generation metric [3]

In our related works, we define the musical noise as the generated audible isolated spectral components thorough processing. Figure 3(b) illustrates an example spectrogram of musical noise. In the figure, we can see isolated components. We speculate that the amount of musical noise is highly related to the number of such isolated components and the isolated level of them.

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

This adoption allows us to obtain the characteristics about tonal components. The adopted kurtosis can evaluate the width of the probability density function (p.d.f.) and the weight of their skirts. We could say that kurtosis can evaluate the percentage of tonal components in total components. Bigger value indicates a signal with heavy skirt; it means that a signal has a lot of tonal components. Also kurtosis is calculated in concise algebraic form. Thus, kurtosis is a suitable measure for the tonal components in computers. Kurtosis is 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, $\mu_4/\mu_2^2 = 6$.

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

sis ratio for measuring the kurtosis change;

$$\text{kurtosis ratio} = \frac{\text{kurt}_{\text{proc}}}{\text{kurt}_{\text{input}}} \tag{10}$$

where $\text{kurt}_{\text{proc}}$ is the kurtosis of processed signal, and $\text{kurt}_{\text{input}}$ is the kurtosis of input signal. Larger kurtosis ratio ($\gg 1$) indicates the larger kurtosis increment through a processing. This is equal to huge amount of musical-noise generation. On the other hand, smaller kurtosis ratio ($\simeq 1$) implies that less musical-noise generation.

3.3. Resultant kurtosis in SS [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 [10]. The probability density function (p.d.f.) of the gamma distribution for probability variable x is defined as

$$P(x) = \Gamma^{-1}(\alpha)\, \theta^{-\alpha} \cdot x^{\alpha-1}\, e^{-\frac{x}{\theta}}, \tag{11}$$

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 $\text{E}\left[P(x)\right] = \alpha\theta$, where $\text{E}[\cdot]$ is an expectation operator. Furthermore, the kurtosis of Gamma distribution, kurt_{GM}, can be designated as [3]

$$\text{kurt}_{\text{GM}} = \frac{(\alpha+2)(\alpha+3)}{\alpha(\alpha+1)}. \tag{12}$$

Using such a gamma distribution model, the resultant kurtosis of SS, kurt_{SS}, can be given as

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

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).

3.4. Resultant kurtosis after DS [5]

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;

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

where $\text{cum}_n(X)$ expresses the n-th order cumulant of probability variable X. Based on the relation (14), the resultant cumulant after DS, $K_n^{(\text{DS})}$, can be given by,

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

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

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

Considering an actual acoustic signal and its cumulants, we can illustrate the relation between input and output kurtosis via DS in Fig. 4. As we can see from Fig. 4, the output kurtosis decreases in proportion to the number of microphones if the input signal's kurtosis is greater than 6. If the kurtosis of input signal less than 6, the output kurtosis get closer to 6. Thus the output kurtosis is increased thorough DS when the sub-Gaussian signal case.

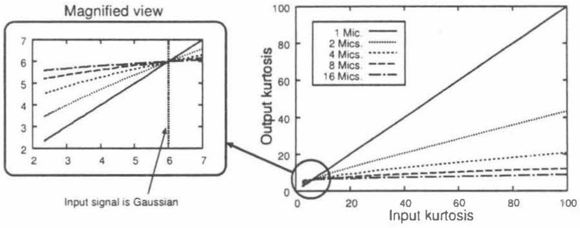

Fig. 4. Relation between input kurtosis and output kurtosis of DS.

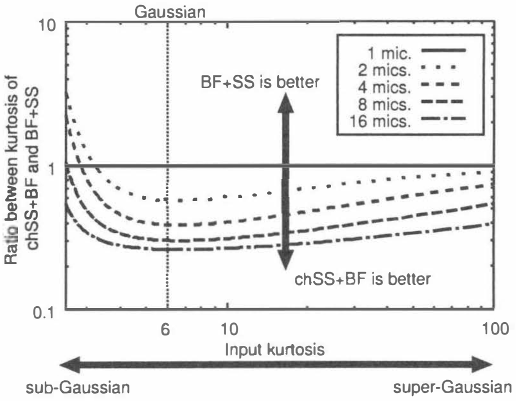

Fig. 5. Resultant kurtosis ratio between BF+SS and chSS+BF.

4. PROPOSED STRUCTURE SELECTION ALGORITHM

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

In the previous section, we have discussed the resultant kurtosis of SS and DS, respectively. In this subsection, we discuss the resultant kurtosis of two kinds of composite systems; chSS+BF and BF+SS.

To compare the output kurtosis of BF+SS and chSS+BF, we introduce the following index R defined as

$$R = \frac{\text{kurt}_{\text{chSS+BF}}}{\text{kurt}_{\text{BF+SS}}}, \tag{17}$$

where R is a ratio between resultant kurtosis of BF+SS and chSS+BF, and $\text{kurt}_{\text{BF+SS}}$ and $\text{kurt}_{\text{chSS+BF}}$ express resultant kurtosis of BF+SS and chSS+BF. As described in Sect. 3.2, the smaller kurtosis increment leads to less amount of generated musical noise. Thus, $R < 1$ implies that chSS+BF reduces musical noise rather than BF+SS.

$\text{kurt}_{\text{BF+SS}}$ is derived as following steps. Firstly, we calculate the kurtosis change via DS by applying relation in Fig. 4. Next, we estimate α in (12) corresponding to the changed kurtosis through DS. Finally, the resultant kurtosis after SS, $\text{kurt}_{\text{BF+SS}}$, is determined by inputting the above α into (13). Also, $\text{kurt}_{\text{chSS+BF}}$ is derived as follows. Firstly, we estimate α in (12) corresponding to the kurtosis of input signal. Next, we determine kurtosis change via SS by inputting the above α into (13). Finally, we calculate the resultant kurtosis after DS, $\text{kurt}_{\text{chSS+BF}}$, by using relation in Fig. 4.

Figure 5 shows R for various types of noises. From this relation, we can see that the structure of chSS+BF can reduce musical noise for signals with kurtosis of more than 4. However, for lower kurtosis signal, e.g., input kurtosis is less than 4, BF+SS reduces musical noise rather than chSS+BF. These facts indicate that the appropri-

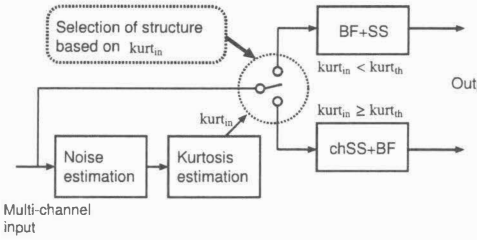
Selection of structure based on kurt$_{in}$
BF+SS
kurt$_{in}$ < kurt$_{th}$
Output
kurt$_{in}$
kurt$_{in}$ ≥ kurt$_{th}$
chSS+BF
Noise estimation
Kurtosis estimation
Multi-channel input

Fig. 6. Block diagram of proposed structure selection algorithm.

ate structure for less musical-noise generation is depending on input noise type.

4.2. Structure selection algorithm

As described in Sect. 4.1, an appropriate integration structure depends on input noise type from the viewpoint of musical-noise generation. Therefore, we propose a new appropriate structure selection algorithm in integration methods of microphone array and spectral subtraction for less musical-noise generation.

Figure 6 depicts the block diagram of the proposed algorithm. Firstly, we measure the kurtosis of the estimated noise signal, kurt$_{in}$. Next, based on the derived kurt$_{in}$, we switch the integration structures as follows;

$$\begin{cases}\text{If } \mathrm{kurt}_{in} < \mathrm{kurt}_{th} & \text{then using BF+SS structure,}\\ \text{If } \mathrm{kurt}_{in} \geq \mathrm{kurt}_{th} & \text{then using chSS+BF structure,}\end{cases} \quad (18)$$

where kurt$_{th}$ indicates the threshold for switching two structures, i.e., chSS+BF and BF+SS.

5. SIMULATION

To confirm the effectiveness of the proposed appropriate structure selection algorithm, we conducted a computer simulation. In the simulation, we compared chSS+BF, BF+SS, and proposed algorithm on the basis of kurtosis ratio described in Sect. 3.2. We used the following 16 kHz-sampled signals as test data; the target speech is the original speech convoluted with the impulse response which were real-recorded in a room with 200 ms reverberation, and to which artificially generated noises which involves various types of noises, i.e., sub-Gaussian, Gaussian and super-Gaussian noises. We used 2- or 4-element array for the simulation, and the oversubtraction parameter of SS was set to 1.5. Besides, flooring parameter of SS in BF+SS was set to 0.0. Also, threshold for proposed algorithm, kurt$_{th}$, was set to 4.0. In the simulation, we assumed that the long-term averaged noise spectrum can be estimated performed perfectly.

Figures 7 and 8 illustrate results of the simulation for 2- and 4-element array, respectively. From these results, we can confirm that the proposed algorithm can select the appropriate structure depending on input noise type. Thus, the proposed algorithm can select the appropriate structure for less musical-noise generation regardless of input noise type.

6. CONCLUSION

In this paper, we proposed an appropriate structure selection algorithm in integration methods of microphone array and spectral subtraction for less musical-noise generation. This proposed algorithm is based on the fact that the appropriate structure in integration methods of microphone array and spectral subtraction is depending on the input noise type. As a result of computer simulation, we confirmed that the proposed algorithm can select the appropriate structure from the viewpoint of musical-noise generation. In the future, the effectiveness of the proposed algorithm would be shown via subjective evaluation.

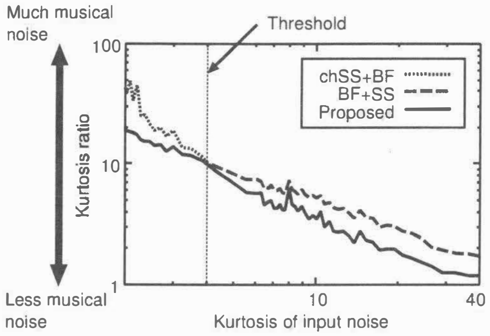

Fig. 7. Kurtosis-ratio-based comparison result for 2-element array.

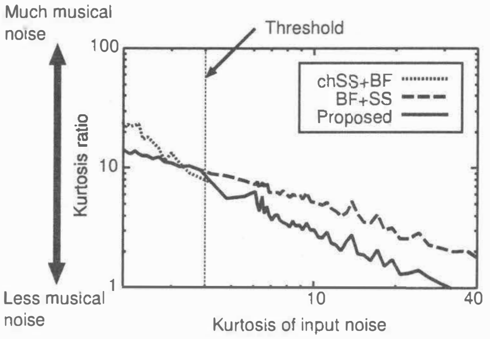

Fig. 8. Kurtosis-ratio-based comparison result for 4-element array.

7. 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 spectra subtraction," *IEEE Trans. Acoustics, Speech, Signal Proc.*, vol.ASSP-27, no.2, pp.113–120, 1979.

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

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

[7] S. Ikeda and N. Murata, "A method of ICA in the frequency domain," *Proc. Intern. Workshop on ICA and BSS*, pp.365–371, 1999.

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

[9] 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.

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