MUSICAL-NOISE-FREE SPEECH ENHANCEMENT: THEORY AND EVALUATION

[†] *Ryoichi Miyazaki,* [†] *Hiroshi Saruwatari,* [†] *Takayuki Inoue,* [†] *Kiyohiro Shikano,* [‡] *Kazunobu Kondo*

[†] Nara Institute of Science and Technology, Nara, Japan (e-mail: ryoichi-m@is.naist.jp)
[‡] Yamaha Corporate Research & Development Center, Shizuoka, Japan

ABSTRACT

In this paper, we propose a new theory of nonlinear noise reduction with a perfectly musical-noise-free property, where no musical noise is generated even for a high signal-to-noise ratio. To achieve high-quality noise reduction with low musical noise, an iterative spectral subtraction method, i.e., recursively applied weak nonlinear signal processing, has been proposed. Although evaluation experiments indicated the existence of an appropriate parameter setting that gives a musical-noise-free state, no theoretical studies have been carried out. Therefore, in this paper, we theoretically derive pairs of internal parameters that satisfy the musical-noise-free condition by analysis based on higher-order statistics. It is clarified that finding a fixed point in the kurtosis of noise spectra enables the reproduction of the musical-noise-free state, and comparative experiments with commonly used noise reduction methods show the efficacy of the proposed method.

***Index Terms*—** Speech enhancement, musical noise free, higher-order statistics, iterative spectral subtraction

1. INTRODUCTION

In recent studies, many applications of hands-free speech communication systems have been investigated, for which noise reduction is a problem requiring urgent attention. Spectral subtraction is a commonly used noise reduction method that has high noise reduction performance [1]. However, in this method, artificial distortion, so-called *musical noise*, arises owing to nonlinear signal processing, leading to a serious deterioration of sound quality.

To achieve high-quality noise reduction with low musical noise, an *iterative spectral subtraction* method has been proposed [2]. This method is performed through signal processing in which *weak* spectral subtraction processes are recursively applied to the input signal. Also, one of the authors has reported the very interesting phenomenon that this method with appropriate parameters gives *equilibrium* behavior in the growth of higher-order statistics with increasing number of iterations [3]. This means that almost no musical noise is generated even with high noise reduction, which is one of the most desirable properties of single-channel nonlinear noise reduction methods. However, the existence of the musical-noise-free state has been discovered only experimentally, and there have been no theoretical studies on it.

Therefore, in this paper, we theoretically derive a closed-form solution of the internal parameters that satisfy the musical-noise-free condition by analysis based on higher-order statistics. It is clarified that finding a fixed point in the kurtosis of noise spectra enables the reproduction of the musical-noise-free state. In addition, comparative experiments with commonly used noise reduction methods show the efficacy of the proposed method via objective and subjective evaluations.

2. RELATED WORKS

2.1. Conventional one-shot spectral subtraction

We apply short-time Fourier analysis to the observed signal, which is a mixture of target speech and noise, to obtain the time-frequency signal. We formulate conventional *non-iterative spectral subtraction* [1] in the time-frequency domain as follows:

$$y(f,\tau) = \begin{cases} \sqrt{|x(f,\tau)|^2 - \beta \mathrm{E}[|N|^2]}\, e^{j\arg(x(f,\tau))} \\ \quad (\text{where} \quad |x(f,\tau)|^2 - \beta \cdot \mathrm{E}[|N|^2] > 0), \\ \eta x(f,\tau) \quad (\text{otherwise}), \end{cases} \tag{1}$$

where $y(f,\tau)$ is the enhanced target speech signal, $x(f,\tau)$ is the observed signal, f denotes the frequency subband, τ is the frame index, β is the oversubtraction parameter, and η is the flooring parameter. Here, $\mathrm{E}[|N|^2]$ is the expectation of the random variable $|N|^2$ corresponding to noise power spectra. In practice, we can approximate $\mathrm{E}[|N|^2]$ by averaging the observed noise power spectra $|n(f,\tau)|^2$ in the first K-sample frames, where we assume speech absence in this period; $\widehat{\mathrm{E}[|N|^2]} \approx \frac{1}{K}\sum_{\tau=1}^{K} |n(f,\tau)|^2$.

2.2. Iterative spectral subtraction

In an attempt to achieve high-quality noise reduction with low musical noise, an improved method based on iterative spectral subtraction was proposed in a previous study [2]. This method is performed through signal processing, in which the following *weak* spectral subtraction processes are recursively applied to the noise signal: (I) The average power spectrum of the input noise is estimated. (II) The estimated noise prototype is then subtracted from the input with the parameters specifically set for weak subtraction, e.g., a large flooring parameter η and a small subtraction parameter β. (III) We then return to step (I) and substitute the resultant output (partially noise reduced signal) for the input signal.

2.3. Modeling of input signal

In this paper, we assume that the input signal x in the power spectral domain is modeled using the gamma distribution as

$$P(x) = \Gamma(\alpha)^{-1}\theta^{-\alpha}x^{\alpha-1}\exp(-x/\theta), \tag{2}$$

where $x \geq 0, \alpha > 0$, and $\theta > 0$. Here, α is the shape parameter, θ is the scale parameter, and $\Gamma(\alpha)$ is the gamma function, defined as $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt$.

This work was partly supported by JST Core Research of Evolutional Science and Technology (CREST), Japan.

2.4. Mathematical metric of musical noise generation via higher-order statistics for one-shot spectral subtraction [4]

In this study, we apply the *kurtosis ratio* to a *noise-only time-frequency period* of the subject signal for the assessment of musical noise [4]. This measure is defined as

$$\text{kurtosis ratio} = \text{kurt}_{\text{proc}}/\text{kurt}_{\text{org}}, \tag{3}$$

where $\text{kurt}_{\text{proc}}$ is the kurtosis of the processed signal and kurt_{org} is the kurtosis of the observed signal. Kurtosis is defined as

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

where μ_m is the *m*th-order moment, given by

$$\mu_m = \int_0^\infty x^m P(x)dx, \tag{5}$$

and $P(x)$ is the probability density function (p.d.f.) of the random variable X. A kurtosis ratio of unity corresponds to no musical noise. This measure increases as the amount of generated musical noise increases.

The mth-order moment after spectral subtraction, μ_m^{SS}, is given by [3]

$$\mu_m^{\text{SS}} = \theta^m \mathcal{M}(\alpha, \beta, \eta, m), \tag{6}$$

where

$$\begin{aligned}\mathcal{M}(\alpha, \beta, \eta, m) = &\sum_{l=0}^{m} (-\beta\alpha_0)^l \frac{\Gamma(m+1)\Gamma(\alpha_0+m-l, \beta\alpha_0)}{\Gamma(\alpha_0)\Gamma(l+1)\Gamma(m-l+1)} \\ &+ \frac{\eta^{2m}}{\Gamma(\alpha)}\gamma(\alpha+m, \beta\alpha).\end{aligned} \tag{7}$$

$\Gamma(b, a)$ and $\gamma(b, a)$ are the upper and lower incomplete gamma functions defined as $\Gamma(b, a) = \int_b^\infty t^{a-1}\exp(-t)dt$ and $\gamma(b, a) = \int_0^b t^{a-1}\exp(-t)dt$, respectively. From (4), (6), and (7), the kurtosis after spectral subtraction can be expressed as

$$\text{kurt}(\alpha, \beta, \eta) = \mathcal{M}(\alpha, \beta, \eta, 4)/\mathcal{M}^2(\alpha, \beta, \eta, 2). \tag{8}$$

Using (3) and (8), we also express the kurtosis ratio as

$$\text{kurtosis ratio} = \frac{\mathcal{M}(\alpha, \beta, \eta, 4)/\mathcal{M}^2(\alpha, \beta, \eta, 2)}{\mathcal{M}(\alpha, 0, 0, 4)/\mathcal{M}^2(\alpha, 0, 0, 2)}. \tag{9}$$

Also, as a measure of noise reduction performance, the noise reduction rate (NRR), the output SNR minus the input SNR in dB, can be given in terms of a 1st-order moment as [3]

$$\text{NRR} = 10\log_{10}\{\alpha/\mathcal{M}(\alpha, \beta, \eta, 1)\}. \tag{10}$$

2.5. Analysis of behavior of iterative spectral subtraction

In this subsection, we formulate the amount of musical noise generated in the iterative spectral subtraction method using the analytical results obtained in Sect. 2.4. Here we conduct a *recursively applied* analysis of kurtosis in the following manner, where the subscript i represents the value in the ith iteration:

(I) First, model the input noise p.d.f. as a gamma distribution with shape parameter α_i (initially $i = 0$).
(II) Next, apply spectral subtraction to the signal using the over-subtraction parameter β and flooring parameter η. We calculate the kurtosis using (6); this is considered as the result of the ith iteration.

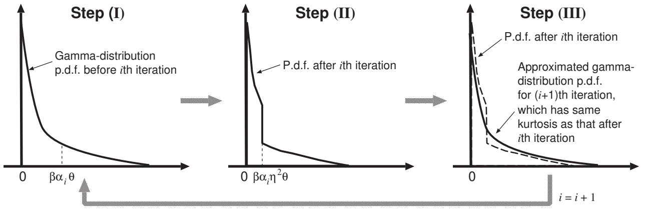
Step (I)
Gamma-distribution p.d.f. before *i*th iteration
0
βα_iθ
Step (II)
P.d.f. after *i*th iteration
0
βα_iη²θ
Step (III)
P.d.f. after *i*th iteration
Approximated gamma-distribution p.d.f. for (*i*+1)th iteration, which has same kurtosis as that after *i*th iteration
0
i = *i* + 1

Fig. 1. P.d.f. deformation and approximated gamma-distribution p.d.f. for $(i+1)$th iteration, which has same kurtosis of p.d.f. as that after ith iteration.

(III) Next, approximately remodel the resultant processed signal as a gamma distribution with the modified shape parameter α_{i+1} corresponding to the resultant kurtosis obtained in step (II) (see Fig. 1). Then return to step (I) with the updated value of α_{i+1}.

Note that this analysis includes an approximation of the p.d.f. modification in which the p.d.f. is always remodeled as a gamma distribution in each iteration. This is necessary because it is difficult to derive an exact analytical expression for the change in kurtosis of the non-gamma distribution. The proposed approximation is, however, still valid if the spectral subtraction process in each step is weak and thus does not change the p.d.f. significantly.

In the following, full details of the iterative analysis are given. The kurtosis in the ith iteration is obtained via steps (I) and (II) using (8) with $\alpha = \alpha_i$ as $\text{kurt}(\alpha_i, \beta, \eta)$. In step (III), a new α_{i+1} is calculated using the following relation between the kurtosis and the shape parameter [3]:

$$\text{kurt}(\alpha_i, \beta, \eta) = \frac{(\alpha_{i+1}+3)(\alpha_{i+1}+2)}{(\alpha_{i+1}+1)\alpha_{i+1}}. \tag{11}$$

This results in a closed-form estimate of the shape parameter from the given kurtosis as

$$\begin{aligned}&\alpha_{i+1} \\ &= \frac{\text{kurt}(\alpha_i, \beta, \eta) - 5 - \sqrt{\text{kurt}(\alpha_i, \beta, \eta)^2 + 14\,\text{kurt}(\alpha_i, \beta, \eta) + 1}}{2 - 2\,\text{kurt}(\alpha_i, \beta, \eta)} \\ &= \mathcal{A}(\text{kurt}(\alpha_i, \beta, \eta)).\end{aligned} \tag{12}$$

By applying the updated α_{i+1} to the new gamma distribution, we can obtain the following recursive equation for the kurtosis in the $(i+1)$th iteration:

$$\text{kurt}(\alpha_{i+1}, \beta, \eta) = \frac{\mathcal{M}(\mathcal{A}(\text{kurt}(\alpha_i, \beta, \eta)), \beta, \eta, 4)}{\mathcal{M}^2(\mathcal{A}(\text{kurt}(\alpha_i, \beta, \eta)), \beta, \eta, 2)}. \tag{13}$$

Thus, we can calculate the resultant kurtosis ratio as

$$\begin{aligned}\text{kurtosis ratio} =& \text{kurt}(\alpha_{i+1}, \beta, \eta)/\text{kurt}(\alpha_0, 0, 0) \\ =& \frac{\alpha_0(\alpha_0+1)}{(\alpha_0+2)(\alpha_0+3)} \frac{\mathcal{M}(\mathcal{A}(\text{kurt}(\alpha_i, \beta, \eta)), \beta, \eta, 4)}{\mathcal{M}^2(\mathcal{A}(\text{kurt}(\alpha_i, \beta, \eta)), \beta, \eta, 2)}.\end{aligned} \tag{14}$$

3. THEOREM ON MUSICAL-NOISE-FREE CONDITIONS

3.1. Overview

As indicated by (13), iterative spectral subtraction theory has an interesting *domino-toppling* phenomenon as follows. Given a specific

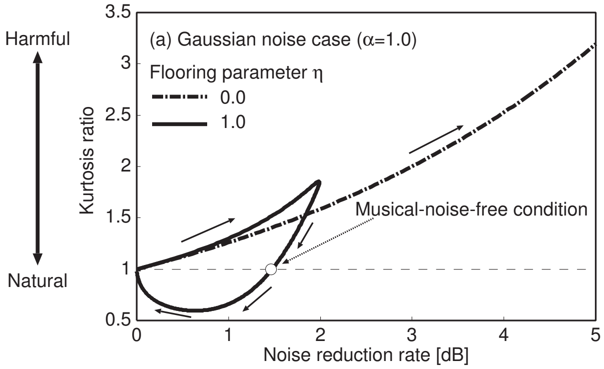

Fig. 2. NRR and kurtosis ratio obtained from theoretical analysis with increasing β. Note that *hysteresis loop* exists when $\eta = 1.0$.

parameter setting, if we are fortunate enough to obtain the same kurtosis as that of the input noise, i.e., $\mathrm{kurt}(\alpha_0, 0, 0)$, after the 1st iteration, i.e.,

$$\mathrm{kurt}(\alpha_0, \beta, \eta) = \mathrm{kurt}(\alpha_0, 0, 0) = \frac{(\alpha_0 + 3)(\alpha_0 + 2)}{(\alpha_0 + 1)\alpha_0}, \qquad (15)$$

then from (12) we have $\alpha_1 = \alpha_0$. Obviously, this leads to the relation

$$\mathrm{kurt}(\alpha_1, \beta, \eta) = \mathrm{kurt}(\alpha_0, \beta, \eta) = \mathrm{kurt}(\alpha_0, 0, 0), \qquad (16)$$

proving that the kurtosis in the 2nd iteration is also identical. The inductive result is that the kurtosis ratio never changes even at a large number of (ideally "infinite") iterations, where sufficient noise reduction is gained even if the NRR improvement in each iteration is small. This corresponds to musical-noise-free noise reduction.

In summary, we can formulate a new theorem on musical-noise-free conditions as follows.

(I) Fixed-point kurtosis condition: The kurtosis should be equal before and after spectral subtraction in each iteration. This corresponds to a fixed point for the 2nd- and 4th-order moments.

(II) NRR growth condition: The amount of noise reduction is larger than 0 dB in each iteration, relating to a change in the 1st-order moment.

This theorem should be of great interest if such conditions hold in existing signal processing. We have found a *hysteresis loop* in the relation between the NRR and kurtosis ratio in one-shot spectral subtraction (calculated by (9) and (10)) with a specific parameter setting (see Fig. 2), showing the existence of a fixed point in the kurtosis. In the following subsections, we mathematically derive more general solutions for musical-noise-free conditions.

3.2. Fixed-point kurtosis condition

Although the parameters to be optimized are η and β, we hereafter derive the optimal η given a fixed β for ease of closed-form analysis. First, we change (8) to

$$\mathrm{kurt}(\alpha_0, \beta, \eta) = \frac{\mathcal{S}(\alpha_0, \beta, 4) + \eta^8 \mathcal{F}(\alpha_0, \beta, 4)}{\left(\mathcal{S}(\alpha_0, \beta, 2) + \eta^4 \mathcal{F}(\alpha_0, \beta, 2)\right)^2}, \qquad (17)$$

$$\mathcal{S}(\alpha_0, \beta, m) = \sum_{l=0}^{m} \frac{(-\beta\alpha_0)^l \Gamma(m+1)\Gamma(\alpha_0+m-l, \beta\alpha_0)}{\Gamma(\alpha_0)\Gamma(l+1)\Gamma(m-l+1)}, \quad (18)$$

$$\mathcal{F}(\alpha_0, \beta, m) = \gamma(\alpha_0+m, \beta\alpha_0)/\Gamma(\alpha_0). \qquad (19)$$

Next, the fixed-point kurtosis condition corresponds to the kurtosis being equal before and after spectral subtraction, thus

$$\frac{\mathcal{S}(\alpha_0, \beta, 4) + \eta^8 \mathcal{F}(\alpha_0, \beta, 4)}{\left(\mathcal{S}(\alpha_0, \beta, 2) + \eta^4 \mathcal{F}(\alpha_0, \beta, 2)\right)^2} = \frac{(\alpha_0 + 3)(\alpha_0 + 2)}{(\alpha_0 + 1)\alpha_0}. \qquad (20)$$

Let $\mathcal{H} = \eta^4$, then (20) yields the following quadratic equation in $\mathcal{H}$.

$$\begin{aligned}&\left(\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\right)\mathcal{H}^2\\ &-2\mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\,\mathcal{H}\\ &+\mathcal{S}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{S}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2) = 0. \quad (21)\end{aligned}$$

Thus, we can derive a closed-form estimate of $\mathcal{H}$ from the given oversubtraction parameter as

$$\begin{aligned}\mathcal{H} =& \{\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\}^{-1}\\ &\Biggl[\mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\\ &\pm \Bigl[\{\mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\}^2\\ &- \left\{\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\right\}\\ &\left\{\mathcal{S}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{S}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\right\}\Bigr]^{\frac{1}{2}}\Biggr].\end{aligned} \qquad (22)$$

Finally, $\eta = \mathcal{H}^{1/4}$ is the resultant flooring parameter that satisfies the fixed-point kurtosis condition.

3.3. NRR growth condition

In this subsection, we reveal the range of the flooring parameter η that increases the NRR. From (10), the NRR growth condition is expressed as

$$\mathrm{NRR} = 10\log_{10}\{\alpha_0/(\mathcal{S}(\alpha_0, \beta, 1) + \eta^2 \mathcal{F}(\alpha_0, \beta, 1))\} > 0. \quad (23)$$

Here, since $\eta > 0$, we can solve the inequality as

$$0 < \eta < \sqrt{\left(\alpha_0 - \mathcal{S}(\alpha_0, \beta, 1)\right)/\mathcal{F}(\alpha_0, \beta, 1)}. \qquad (24)$$

In summary, we can choose the parameters simultaneously satisfying the fixed-point kurtosis condition and NRR growth condition using (22) and (24).

3.4. Example of parameters satisfying musical-noise-free condition

According to the previous analysis, we can calculate combinations of the oversubtraction parameter β and the flooring parameter η that satisfy the musical-noise-free condition under the three types of shape parameter α. Figure 3 shows examples of traces, where the shape parameter α is set to 0.2, 0.6, and 1.0. It is worth mentioning that the specific setting $(\alpha, \beta, \eta) = (1.0, 2.0, 0.97)$ appears in Fig. 3, which was heuristically discovered in [3], but our theory can provide more wide-ranging solutions.

4. EVALUATION EXPERIMENTS AND RESULTS

4.1. Experimental conditions

We conducted objective and subjective evaluation experiments to confirm the validity of the theoretical analysis described in the previous section. The input SNR of the test data was set to 0 dB. The

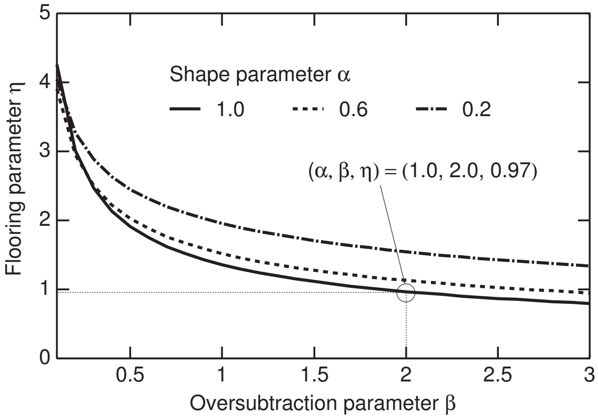

Fig. 3. Example of oversubtraction parameter β and flooring parameter η that satisfy musical-noise-free condition.

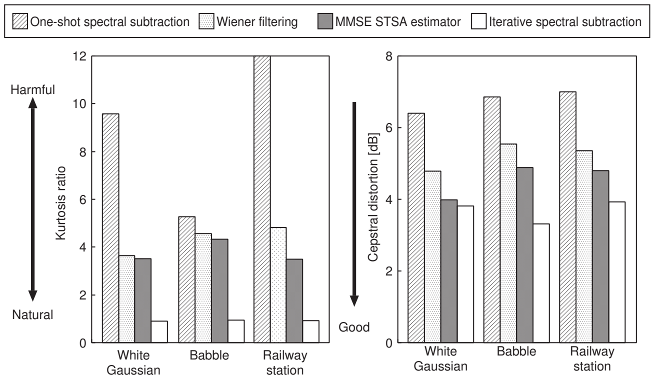

Fig. 4. Kurtosis ratio and cepstral distortion obtained from experiment with real noisy speech data under 10-dB-NRR condition.

target speech signals were the utterances of two male and two female speakers in Japanese (four sentences). The noise signal was white Gaussian noise ($\alpha_0 = 0.97$), babble noise ($\alpha_0 = 0.21$) consisting of 36 voices that simulate a crowded place or railway station noise ($\alpha_0 = 0.33$). Each signal was sampled at 16 kHz. In these experiments, the number of iterations is five in iterative spectral subtraction.

4.2. Objective evaluation

In this subsection, we compare iterative spectral subtraction with other commonly used noise reduction methods under the same NRR condition. Figure 4 shows the kurtosis ratio and cepstral distortion obtained from the experiment with real noisy speech data for white Gaussian noise and babble noise, where we evaluate 10-dB-NRR (i.e., output SNR = 10 dB) signals processed by four methods, namely, conventional one-shot spectral subtraction, Wiener filtering [5], the minimum mean-square error (MMSE) short-time spectral amplitude (STSA) estimator [6], and iterative spectral subtraction with the optimal parameter settings. In this experiment, we calculate the kurtosis ratio using (3) in the first 1 s frames, where we assume speech absence in all noise reduction methods.

From Fig. 4, we can confirm that iterative spectral subtraction outperforms the other conventional methods in terms of both the kurtosis ratio and cepstral distortion. In particular, the kurtosis ratio of the proposed method is closest to 1.0. Since Wiener filtering and the MMSE STSA estimator are often referred to as methods producing less musical noise, this result greatly emphasizes the advantageousness of iterative spectral subtraction, i.e., its *no-musical-noise property*, as theoretically predicted in Sect. 3.1.

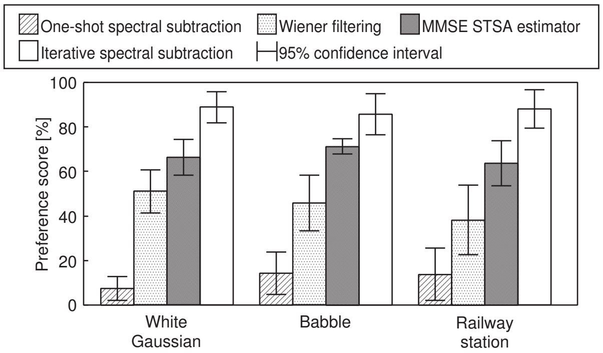

Fig. 5. Subjective evaluation results.

4.3. Subjective evaluation

We next conducted a subjective evaluation. In the evaluation, we presented a pair of 10-dB-NRR signals processed by each method in random order to 10 examinees, who selected which signal they considered to contain least musical noise. The results of the experiment are shown in Fig. 5. It was found that musical noise is less perceptible when iterative spectral subtraction with the optimal parameters is used than when conventional methods are used. This result is also consistent with our theoretical analysis, thus confirming the validity of the proposed method.

5. CONCLUSION

In this paper, we presented a new theory for musical-noise-free noise reduction based on iterative spectral subtraction. We derived the optimal parameters satisfying the musical-noise-free condition to find the fixed point in the kurtosis. Comparative experiments with commonly used methods showed the efficacy of the proposed method.

6. REFERENCES

[1] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. ASSP*, vol.27, no.2, pp.113–120, 1979.

[2] M. R. Khan, et al., “Iterative noise power subtraction technique for improved speech quality,” *Proc. ICECE2008*, pp.391–394, 2008.

[3] T. Inoue, et al., “Theoretical analysis of iterative weak spectral subtraction via higher-order statistics,” *Proc. MLSP2010*, pp.220–225, 2010.

[4] Y. Uemura, et al., “Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics,” *Proc. IWAENC2008*, 2008.

[5] P. C. Loizou, *Speech Enhancement Theory and Practice*, CRC Press, Taylor & Francis Group, FL, 2007.

[6] Y. Ephraim, et al., “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. ASSP*, vol.32, no.6, pp.1109–1121, 1984.