THEORETICAL ANALYSIS OF MUSICAL NOISE IN WIENER FILTERING FAMILY VIA HIGHER-ORDER STATISTICS

1 Takayuki Inoue, 1 Hiroshi Saruwatari, 2 Kiyohiro Shikano, and 1 Kazunobu Kondo
1 Nara Institute of Science and Technology, Nara, 630-0192 Japan
2 Yamaha Corporate Research & Development Center, Shizuoka, 438-0192 Japan

ABSTRACT
Recently, one of the authors has reported that the amount of generated musical noise is strongly correlated with higher-order statistics of the power spectra. On the basis of this finding, in this paper, we provide a new theoretical analysis of the amount of musical noise generated via the Wiener filtering family. Our theoretical analysis allows the universal performance description from the viewpoint of the amount of musical noise generation and that of noise reduction, enabling reasonable sound quality comparison under the same noise reduction performance. From a mathematical analysis and evaluation experiments, we also clarify which parameter settings result in less musical noise generated in the Wiener filtering family.

Index Terms—Speech enhancement, Musical noise, Higher-order statistics, Wiener filtering, Kurtosis

1. INTRODUCTION
Spectral subtraction (SS) [1] and Wiener filtering (WF) [2] are commonly used noise reduction methods that have high noise reduction performance. However, in these methods, artificial distortion, so-called musical noise, arises owing to nonlinear signal processing, leading to a serious deterioration of sound quality. Moreover, no objective metric to measure how much musical noise is generated has been proposed in previous studies. Thus, it has been difficult to evaluate the amount of musical noise generated and to optimize the internal parameters of a system.

Recently, one of the authors has reported that the amount of generated musical noise is strongly correlated with the difference between the higher-order statistics of the power spectra before and after nonlinear signal processing [3, 4]. On the basis of this finding, an objective metric to measure how much musical noise is generated through nonlinear signal processing has been developed. Using this metric, we have analyzed the amount of musical noise generated via SS in our previous work [5, 6].

However, no theoretical analysis of the amount of musical noise generated in WF has been reported. Although various types of WF methods have often been compared in past studies [2, 7, 8], they were simply compared via experimental measurement of the resultant sound quality, and there have been few comparisons on a theoretical basis. Therefore, in this paper, we provide a new theoretical analysis of the amount of musical noise generated in several types of methods in the WF family. Our theoretical analysis allows the universal performance description from the viewpoint of the amount of musical noise generation and that of noise reduction, enabling reasonable sound quality comparison under the same noise reduction performance. From a mathematical analysis and evaluation experiments, we also clarify which parameter settings result in less musical noise generated in the WF family.

2. RELATED WORKS
2.1. Standard WF and square-root WF
In this subsection, we introduce standard WF and square-root WF, which are the most commonly used methods in the WF family [2]. 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. WF is generally formulated as follows:

\[ S(f, \tau) = G(X(f, \tau)) \exp(X(f, \tau)). \]

where \( S(f, \tau) \) is the enhanced target speech signal, \( X(f, \tau) \) is the observed signal, \( f \) denotes the frequency subband, and \( \tau \) is the time frame index. \( G \) is the gain function, defined by

\[ G = \frac{P_{ss}}{P_{ss} + P_{nn}} = \frac{P_{ss}}{P_{nn} + P_{nn}}, \]

where \( P_{ss} \), \( P_{nn} \), and \( P_{nn} \) are the power spectral densities of target speech, noise, and the observed signal, respectively. To take into account the nonstationary property of target speech, using instantaneous values of the observed and noise time-frequency signals, the gain function is reformulated in a time-varying manner as

\[ G(f, \tau) = \left\{ \begin{array}{ll} \frac{\|X(f, \tau)\|^2 - \beta \cdot \text{SNR}(f, \tau) \|X(f, \tau)\|^2}{\|X(f, \tau)\|^2 - \beta \cdot \text{SNR}(f, \tau) \|X(f, \tau)\|^2}, & \text{if } \|X(f, \tau)\|^2 - \beta \cdot \text{SNR}(f, \tau) \|X(f, \tau)\|^2 > 0, \\ 0, & \text{otherwise}, \end{array} \right. \]

where \( \text{SNR}(f, \tau) \) is the estimated noise signal, \( E(f, \tau) \) is the expected noise operator of \( f \) over \( \tau \), and \( \beta \) is the processing strength parameter.

Equation (3) is referred to as the standard WF gain function. We cannot calculate the a priori signal-to-noise ratio (SNR) in (2) because we have no information on \( P_{nn} \). Therefore, instead of the a priori SNR in (2), we replace the a priori SNR in the gain function with the a posteriori SNR \( \|X(f, \tau)\|^2 / P_{nn} \) resulting in

\[ G(f, \tau) = \frac{P_{ss}}{P_{ss} + P_{nn}} = \frac{P_{ss}}{P_{nn} + 1}, \]

where \( \|X(f, \tau)\|^2 / P_{nn} + 1 = \|X(f, \tau)\|^2 + P_{nn}. \)

Moreover, we extend (5) to a parametric form to achieve flexible noise reduction; the gain function is given by [9]

\[ G(f, \tau) = \left( \frac{\|X(f, \tau)\|^2}{\|X(f, \tau)\|^2 + \beta \|E(f, \tau)\|^2} \right)^\xi, \]

where \( \xi \) is the signal exponent parameter and \( \eta \) is the gain exponent parameter.
level of isolation. In this paper, we call these isolated components **tonal components**. Since such tonal components have relatively high power, they are strongly related to the weight of the skirt of their probability density function (p.d.f.). Therefore, quantifying the skirt of the p.d.f. makes it possible to measure the number of tonal components. Thus, we adopt kurtosis, one of the most commonly used higher-order statistics, to evaluate the percentage of tonal components among the total components. A larger kurtosis value indicates a signal with a heavy skirt, meaning that the signal has many tonal components. Kurtosis is defined as

$$\mu_4 = \int_0^\infty x^4 \, P(x) \, dx,$$

where $\mu_4$ is the 4th central moment of the probability density function (p.d.f.). Therefore, quantifying the skirt of the p.d.f. makes it possible to measure the number of tonal components. Thus, we adopt kurtosis, one of the most commonly used higher-order statistics, to evaluate the percentage of tonal components among the total components. A larger kurtosis value indicates a signal with a heavy skirt, meaning that the signal has many tonal components. Kurtosis is defined as

$$\mu_4 = \int_0^\infty x^4 \, P(x) \, dx,$$

where $\mu_4$ is the 4th central moment of the probability density function (p.d.f.).

In this study, we apply such a kurtosis-based analysis to a noise-only time-frequency period of subject signals for the assessment of musical noise, even though these signals contain target-speech-dominant periods. Thus, this analysis should be conducted during, for example, periods of silence during speech. This is because we aim to quantify the tonal components arising in the noise-only part, which is the main cause of musical noise perception [4], and not in the target-speech-dominant part. Although kurtosis can be used to measure the number of tonal components, note that the kurtosis itself is not sufficient to measure the amount of musical noise. This is obvious since the kurtosis of a signal $x$ in the power spectral domain can be modeled by

$$\mu_m = \int_0^\infty x^m \, P(x) \, dx,$$

where $\mu_m$ is the $m$th-order moment of the processed signal and $\mu_{m\text{org}}$ is the $m$th-order moment of the original signal. The measure increases as the amount of generated musical noise increases. In Ref. [3], it was reported that the kurtosis ratio is strongly correlated with the human perception of musical noise.

### 3. THEORETICAL ANALYSIS OF WF FAMILY

#### 3.1. Analysis strategy

In this section, we analyze the amount of noise reduction and musical noise generated through the WF family using kurtosis. In the analysis, we first model a noise signal by a gamma distribution and formulate the resultant p.d.f. and moments after WF (see Sect. 3.2). Then, kurtosis is obtained from the 2nd- and 4th-order moments, and the amount of noise reduction is calculated from the 1st-order moment (see Sect. 3.3). Finally, we compare the kurtosis values among the WF family under the same amount of noise reduction (see Sect. 3.4).


##### 3.2.1. Deformation of p.d.f. via processing

In this section, we formulate the p.d.f. via WF. First, we assume that the input signal $x$ in the power spectral domain can be modeled by the gamma distribution as

$$P(x) = \frac{x^{\alpha-1} \exp(-x)}{\Gamma(\alpha) \beta^\alpha},$$

where $\alpha$ is the shape parameter corresponding to the type of noise (e.g., $\alpha = 1$ is Gaussian and $\alpha < 1$ is super-Gaussian), $\beta$ is the scale parameter of the gamma distribution, and $\Gamma(\alpha)$ is the gamma function, defined as

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} \exp(-s) \, ds.$$

The original p.d.f. $P(x)$ is transformed into the resultant p.d.f. $P(y)$ via processing. We can calculate $P(y)$ by considering a change of variables $y = g(x)$, applied to convert an integral in terms of the variable $x$ to an integral in terms of the variable $y$. The converted p.d.f. $P(y)$ can be written as

$$P(y) = P(x)^{|y|}/|y|,$$

where $|y|$ is the Jacobian of the transformation, defined by

$$\frac{dy}{dx}.$$

We apply (12) to (10) to obtain the p.d.f. after processing, $P(y)$.

##### 3.2.2. Mth-order moment of standard WF

Since $x$ is the power spectral domain signal and its mean value $E_x = E[N(f, T)]^2$ is given by $\theta_0$ in the gamma distribution, the variable $y$ used for processing is expressed as

$$y = \frac{\alpha x}{\beta},$$

This results in the following quadratic equation in $x$ to be solved:

$$x^2 - \frac{2\beta\alpha y}{\beta^2} x + \frac{\alpha^2}{\beta^2} = 0,$$

and we can derive a closed-form solution

$$x = \frac{\alpha y + \sqrt{\alpha^2 y^2 + 4\alpha^2 \beta^2}}{2\beta} = f(y).$$

Since the gain function (3) must have positive value, the range of $f(y)$ is $[\beta \theta_0, \infty]$. Since $x > 0$ and $y > 0$, the Jacobian is

$$\frac{dx}{dy} = f'(y) = |y|.$$

Consequently,

$$P(y) = P(x)/|y| = \frac{(f(y))^{|y|} \exp(-f(y))}{\Gamma(\alpha) \beta^\alpha},$$

The $m$th-order moment of $P(y)$ is given by

$$\mu_m = \int_0^\infty y^m \, P(y) \, dy = \int_0^\infty y^m \, (f(y))^{|y|} \exp(-f(y)) \, f'(y) \, dy.$$

Let $t = f(y)/\theta$, then $dy = \theta \, dt$ and the range of the integral changes from $[0, \infty]$ to $[\beta \theta_0, \infty]$. Furthermore, $f(y)$ is expressed as

$$f(y) = \theta_0 x = t \theta_0.$$

We apply (20) to (14), then $y^m$ is expressed as

$$y^m = \frac{\theta_0^m}{\theta_0} t^{m-1} \exp(-t) \, dt.$$

Then, we apply (20) and (21) to (19) to obtain

$$\mu_m = \frac{\theta_0^m}{\theta_0} M_{sw}(\alpha, \beta, m),$$

where

$$M_{sw}(\alpha, \beta, m) = \int_0^\infty (t - \beta \theta_0)^m \theta_0^{m-1} \exp(-\theta_0 t) \, dt.$$

Hereafter, we call this function the $M$ function.

#### 3.2.3. Mth-order moment of square-root WF

Square-root WF is equivalent to power domain SS. Thus, the $m$th-order moment of this method can be formulated as the following
mth-order moment of power domain SS, which we formulated previously [6]:

$$\mu_m = \frac{\gamma^m}{\Gamma(\alpha)} M_{\text{SS}(\alpha, \beta, m)},$$  \hspace{1cm} (24)

where

$$M_{\text{SS}(\alpha, \beta, m)} = \sum_{l=0}^{\infty} (-\beta \alpha)^l \frac{\Gamma(m+l)}{\Gamma(m-l+1)} \frac{\Gamma(a+m-l, \beta \alpha)}{(a+m-1, \beta \alpha)},$$  \hspace{1cm} (25)

and $\Gamma(a, z)$ is the upper incomplete gamma function defined as

$$\Gamma(a, z) = \int_{z}^{\infty} e^{-t} t^{a-1} dt,$$  \hspace{1cm} (26)

3.2.4. Mth-order moment of quasi-parametric WF

Next, we consider quasi-parametric WF. In the exponent spectral domain, its mean value $E[|W(f, \eta)|^p]$ is given by $\theta^p \Gamma(\alpha + \frac{p}{2}) / \Gamma(\alpha)$ [6], and $\eta$ for quasi-parametric WF is expressed as

$$\eta = \left\{ \left( 1 + \beta \eta \right) \Gamma(\alpha + \frac{1}{2}) / \Gamma(\alpha) \right\}^{\frac{1}{2}}.$$  \hspace{1cm} (27)

Since (27) is a monotonic function with $x$, we can define the inverse function of (27) as $x = f(\eta)$. Since $x > 0$ and $y > 0$, the Jacobian is

$$\frac{dx}{d\eta} = f'(y) = |f|.$$  \hspace{1cm} (28)

Consequently,

$$P(\eta) = P(x)|f| = \frac{\beta^\eta \Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} f'(y).$$  \hspace{1cm} (29)

The mth-order moment of $P(\eta)$ is given by

$$\mu_m = \int_{0}^{\infty} \eta^m P(\eta) d\eta = \int_{0}^{\infty} \eta^m \frac{\beta^\eta \Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} f'(y) d\eta.$$  \hspace{1cm} (30)

Let $t = f(\eta)/\theta$, then $d\eta = \theta f'(\eta) dt$ and the range of the integral does not change. Furthermore, $f(\eta)$ is expressed as

$$f(\eta) = \theta = x.$$  \hspace{1cm} (31)

We apply (31) to (27) to obtain

$$\eta^m = \left( \frac{\theta t - \beta \eta \Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} \right)^{\frac{-1}{\beta \eta - \Gamma(\alpha + \frac{1}{2})}}.$$  \hspace{1cm} (32)

Then, we apply (31) and (32) to (30), giving

$$\mu_m = \frac{\beta^m}{\Gamma(\alpha)} M_{\text{QPP}(\alpha, \beta, m, \eta, \xi)}.$$  \hspace{1cm} (33)

where

$$M_{\text{QPP}(\alpha, \beta, m, \eta, \xi)} = \int_{0}^{\infty} \left( \frac{\theta t - \beta \eta \Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)} \right)^{\frac{-1}{\beta \eta - \Gamma(\alpha + \frac{1}{2})}} \exp(-t) dt.$$  \hspace{1cm} (34)

3.3. Estimation of amount of musical noise and noise reduction

3.3.1. Analysis of amount of musical noise

We can obtain the kurtosis and noise reduction performance using the $M_4$ function, which is formulated in the previous subsection, where $* = \text{SWF}$ for standard WF, $* = \text{SRWF}$ for square-root WF, and $* = \text{QPFW}$ for quasi-parametric WF. Note that when we perform the calculations for standard WF and square-root WF, we ignore the arguments of $\xi$ and $\eta$. We can obtain the kurtosis after processing as follows:

$$\text{kurt}_m = \frac{\mu_4}{\mu_2^2} = \Gamma(\alpha) \frac{M_4(\alpha, \beta, 4, \xi, \eta)}{M_2(\alpha, \beta, 2, \xi, \eta)}.$$  \hspace{1cm} (35)

By substituting $\beta = 0$ into (35), we can estimate the kurtosis before processing. Thus, we can calculate the resultant kurtosis ratio as

$$\text{kurt} \frac{\text{ratio}}{\text{m}} = \frac{M_4(\alpha, 0, 4, \xi, \eta)}{M_4(\alpha, 0, 0, 2, \xi, \eta)}.$$  \hspace{1cm} (36)

Fig. 1. Theoretical behavior of NRR and log kurtosis ratio for standard WF, square-root WF, and quasi-parametric WF.

$$NRR = \frac{10 \log_{10} \frac{E[n_{\text{in}}]}{E[n_{\text{out}}]}}{10 \log_{10} \frac{E[n_{\text{in}}]}{E[n_{\text{out}}]}}.$$  \hspace{1cm} (37)

where $x_{\text{in}}$ and $x_{\text{out}}$ are the input and output speech signals, and $n_{\text{in}}$ and $n_{\text{out}}$ are the input and output noise signals, respectively. Here, the denominator in (37) is the input SNR and the numerator is the output SNR. If we assume that the amount of noise reduction is much larger than that of speech distortion in processing, i.e., $E[n_{\text{in}}^2] = E[n_{\text{out}}^2]$, then

$$NRR = 10 \log_{10} \frac{E[n_{\text{in}}^2]}{E[n_{\text{out}}^2]}.$$  \hspace{1cm} (38)

Since, $E[n_{\text{in}}^2] = \mu_2$, when $\beta = 0$ in the $M_2$ function and $E[n_{\text{in}}^2] = \mu_2$ for a specific (nonzero) $\beta$.

$$NRR = 10 \log_{10} \frac{\mu_2}{\mu_2^{(\alpha, \beta, 0, 0, 2, \xi, \eta)}}.$$  \hspace{1cm} (39)

In summary, we can derive theoretical estimates for the amount of musical noise and NRR using (36) and (39), respectively. Although the $M_4$ function includes an integral, we can calculate it using a numerical integral method.

3.4. Comparison of amount of musical noise under same NRR condition

According to the previous analysis, we can compare the amount of musical noise among the WF family under the same amount of noise
reduction. Figures 1 and 2 show the theoretical behavior of the kurtosis ratio and NRR for various parameter values. In these figures, the shape parameter $\alpha$ is set to 1.0, NRR is varied from 0 to 12 dB, and the processing strength parameter $\beta$ is adjusted so that the target speech NRR is achieved. Note that we plot the logarithm of the kurtosis ratio because the kurtosis exponentially increases with $\beta$ [3]. We call this the log kurtosis ratio hereafter. In quasi-parametric WF, the signal exponent parameter $\xi$ is set to 2.0, 1.0, and 0.5, and the gain exponent parameter $\eta$ is set to 2.0/0.5, 1.0, and 0.5/0.5.

First, we compare the kurtosis among the typical methods in the WF family. Figure 1 shows that a large amount of musical noise is generated when we use standard WF and square-root WF. However, it also shows that a smaller amount of musical noise is generated when we use quasi-parametric WF with a lower gain exponent parameter. Next, we compare the kurtosis among various exponent parameters in quasi-parametric WF. Figure 2 shows that a small amount of musical noise is generated when either of the exponent parameters, $\xi$ or $\eta$, is set to a lower value. Consequently, we can achieve high sound quality upon setting lower exponent parameters in quasi-parametric WF.

4. OBJECTIVE AND SUBJECTIVE EVALUATIONS

We also conducted objective and subjective evaluations to confirm the validity of the theoretical analysis described in the previous section. In the evaluation experiments, noisy observation signals were generated by adding noise signals to target speech signals with a SNR of 0 dB. The target speech signals were the utterances of four speakers (4 sentences), and the noise signal was white Gaussian noise. The length of each signal was 7 s, and each signal was sampled at 16 kHz. The FFT size is 1024, and the frame shift length is 256. In these experiments, we assumed that the noise prototype, i.e., the average of $|W(i, \tau)|^2$, was perfectly estimated.

Figure 3 shows the log kurtosis ratio and cepstral distortion. These values are calculated from the observed and processed signals by standard WF, square-root WF, and quasi-parametric WF with $(\xi, \eta) = (0.5, 1.0)$. We can confirm that the result of the log kurtosis ratio is almost consistent with the theoretical behavior, and cepstral distortion is reduced when we use quasi-parametric WF.

In the subjective evaluation, we presented three equi-NRR signals processed by standard WF, square-root WF, and quasi-parametric WF in random order to 10 examinees, who selected which signal they considered to contain least musical noise. The result of the experiment is shown in Fig. 4. It is found that musical noise is less perceptible when quasi-parametric WF with lower exponent parameters is used. This result is also consistent with our theoretical analysis, thus confirming the validity of the proposed method of theoretical analysis.

5. CONCLUSION

In this study, we performed a theoretical analysis of the amount of musical noise generated among the WF family based on higher-order statistics. It was clarified from a mathematical analysis and evaluation experiments that less musical noise is generated in quasi-parametric WF than in standard WF and square-root WF. In addition, when we use lower exponent parameters in quasi-parametric WF, we can obtain an enhanced speech signal with less musical noise.

6. REFERENCES


