THEORETICAL ANALYSIS OF ITERATIVE WEAK SPECTRAL SUBTRACTION VIA HIGHER-ORDER STATISTICS

Takayuki Inoue, Hiroshi Saruwatari, Yu Takahashi, Kiyorhiro Shikano, and Kazunobu Kondo

Nara Institute of Science and Technology, Nara, 630-0192 Japan
T Corporate Research and Development Center, Yamaha Corp., Shizuoka, 438-0192 Japan

ABSTRACT

In this paper, we provide a new theoretical analysis of the amount of musical noise generated via iterative spectral subtraction based on higher-order statistics. To achieve high-quality noise reduction with low musical noise, the iterative spectral subtraction method, i.e., recursively applied weak nonlinear signal processing, has been proposed. Although the effectiveness of the method has been reported experimentally, there have been no theoretical studies. Therefore, in this paper, we formulate the generation process of musical noise by tracing the change in kurtosis, and conduct a comparison of the amount of musical noise for different parameter settings under the same noise reduction performance. It is clarified from mathematical analysis and evaluation experiments that iterative spectral subtraction with very weak processing can result in the generation of less musical noise.

1. INTRODUCTION

In recent studies, many applications of speech communication systems, such as mobile phones, have been investigated. It is, however, well known that these systems always suffer from the deterioration of speech quality under adverse noise conditions, and thus 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, 3, 4, 5]. This method is performed through signal processing in which weak spectral subtraction processes are recursively applied to the input signal. Basically, the methodology used in iterative spectral subtraction is of great interest for researchers on nonlinear signal processing and machine learning, because it addresses the inherent questions of whether or not nonrecursive weak (nonlinear) signal processing can provide better performances. Although the effectiveness of the iterative spectral subtraction method has been reported only experimentally, to the best of our knowledge there have been no studies on the theoretical advantages of iterative spectral subtraction. One reason for this is the difficulty of theoretical study due to the fact that no objective metric to measure how much musical noise is generated has been proposed.

Recently, one of the authors has reported that the amount of generated musical noise is strongly correlated with the difference between higher-order statistics of the power spectra before and after nonlinear signal processing [6, 7]. On the basis of the findings, an objective metric to measure how much musical noise is generated through nonlinear signal processing has been developed. Using this metric, in this paper we provide a new theoretical analysis of the amount of musical noise generated in iterative spectral subtraction, which is an extension of our previous theory for the recursive case. This proposed analysis enables us to formulate a closed-form estimate of the amount of musical noise, and we can thus compare the amount of musical noise generated by the original one-shot-type and iterative methods under the same noise reduction performance. We clarify from mathematical analysis and evaluation experiments that less musical noise is generated when we recursively apply weak spectral subtraction to the input signal. This is strong evidence that our theoretical justification for using the iterative spectral subtraction method is valid.

2. RELATED WORKS

2.1. Original 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 spectrum. We formulate original one-shot spectral subtraction [1] in the time-frequency domain as follows:

\[ Y(f, r) = \begin{cases} \sqrt{[X(f, r)^2 - \beta \cdot E_r[[N(f, r)]^2]]} \text{e}^{j\angle[X(f, r)]}, \\ Y(f, r) - \beta \cdot E_r[[N(f, r)]^2 > 0), \\ Y(f, r) \text{ (otherwise)}, \end{cases} \]

where \( Y(f, r) \) is the enhanced target speech signal, \( X(f, r) \) is the observed signal, and \( N(f, r) \) is the estimated noise signal. Also, \( f \) denotes the frequency subband, \( r \) is the frame index, \( E_r[\cdot] \) is the expectation operator of over \( r \), and \( \beta \) is the subtraction coefficient.

Generally speaking, conventional spectral subtraction suffers from the inherent problem of musical noise generation. For example, a large subtraction parameter allows a large noise reduction but considerable musical noise is also generated. To reduce the amount of musical noise generated, we often increase the floor parameter, but this decreases noise reduction; thus, there exists a trade-off between noise reduction and musical noise generation.

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 previous studies [2, 3, 4, 5]. 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 \( \eta \) and a small subtraction parameter \( \beta \). (III) We then return to step (I) and substitute the resultant output (partially noise-reduced signal) for the input signal.

Although the efficacy of the iterative spectral subtraction method has been reported experimentally, the theoretical or mathematical justification of its principles has not yet been presented. Intuitively, it appears that weak subtraction generates little musical noise in each iteration. However, if we require sufficient noise reduction, a large number of iterations are needed, causing the amount of musical noise to accumulate. Moreover, it is not self-evident that the accumulated musical noise is smaller than that of conventional one-shot spectral subtraction. Therefore, the lack of justification of iterative spectral subtraction reduces its applicability to general noise reduction. Also the proof of the theoretical basis of the method remains as an open problem.


We speculate that the amount of musical noise is highly correlated with the number of isolated power spectral components and their level of isolation. In this paper, we call these isolated components tonal components. Since each tonal component has 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

\[
\text{kurt} = \frac{\mu_4}{\mu_2^2}, \quad (2)
\]

where "kurt" is the kurtosis and \( \mu_m \) is the mth-order moment, given by

\[
\mu_m = \frac{1}{\mu_2} \int x^m P(x) \, dx, \quad (3)
\]

where \( P(x) \) is the p.d.f. of a signal \( x \). Note that \( \mu_2 \) is not a central moment but a raw moment. Thus, (2) is not kurtosis in the mathematically strict definition but a modified version; however, we still refer to (2) as kurtosis in this paper.

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, 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 some unprocessed noise signals, such as an interfering speech signal, is also high, but we do not recognize speech as musical noise.

Hence, we turn our attention to the change in kurtosis between before and after signal processing to identify only the musical-noise components. Thus, we adopt the kurtosis ratio as a measure to assess musical noise [6]. This measure is defined as

\[
\text{kurtosis ratio} = \frac{\text{kurt}_{\text{proc}}}{\text{kurt}_{\text{org}}}, \quad (4)
\]

where \( \text{kurt}_{\text{proc}} \) is the kurtosis of the processed signal and \( \text{kurt}_{\text{org}} \) is the kurtosis of the observed signal. This measure increases as the amount of generated musical noise increases. In Ref. [6], it was reported that the kurtosis ratio is strongly correlated with the human perception of musical noise.

3. THEORETICAL ANALYSIS OF ITERATIVE WEAK SPECTRAL SUBTRACTION

3.1. Analysis Strategy

In this section, we analyze the amounts of noise reduction and musical noise generated through iterative spectral subtraction using kurtosis. In the analysis, we first model a noise signal by a gamma distribution (see Sect. 3.2) and formulate the resultant p.d.f. after one-shot spectral subtraction (see Sect. 3.3). Then, the kurtosis is obtained from the 2nd- and 4th-order moments (see Sect. 3.4), and the amount of noise reduction is calculated from the 1st-order moment (see Sect. 3.5). Next, on the basis of the above-mentioned analysis, we formulate the behavior of recursively applied spectral subtraction and compare the kurtosis values upon changing the parameter settings under the same amount of noise reduction (see Sect. 3.6).

3.2. Modeling of Input Signal

We assume that the input signal \( x \) in the power spectral domain can be modeled by the gamma distribution as [8]

\[
P(x) = \frac{x^{\alpha-1} \exp(-\frac{x}{\theta})}{\Gamma(\alpha) \theta^\alpha}, \quad (5)
\]

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

\[
\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) \, dt. \quad (6)
\]

3.3. Process of Deforming P.d.f. of Noise via Original One-Shot Spectral Subtraction

In this section, we analyze the kurtosis after spectral subtraction. In conventional one-shot spectral subtraction, 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 Fig. 1). 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 Fig. 1). Thus, the resultant p.d.f. after spectral subtraction, \( P_{SS}(z) \), can be written as

\[
P_{SS}(z) = \frac{1}{\Theta(\alpha)} \left( \frac{z + \beta \alpha \eta}{\theta} \right)^{\alpha-1} \exp \left\{ - \frac{z + \beta \alpha \eta}{\theta} \right\}
\]

where \( z \) is the random variable of the p.d.f. after spectral subtraction.

### 3.4. Analysis of Amount of Musical Noise in Original One-Shot Spectral Subtraction

From (7), the kurtosis after spectral subtraction can be expressed as

\[
\text{Kurt}(\alpha, \beta, \eta) = \frac{\mathcal{K}(\alpha, \beta, \eta)}{\mathcal{K}(\alpha, \beta, 0)} - 3
\]

where

\[
\begin{align*}
\mathcal{K}(\alpha, \beta, \eta) &= \frac{\int_0^{\infty} z^3 P_{SS}(z) \, dz}{\left( \int_0^{\infty} z P_{SS}(z) \, dz \right)^3} \\
&= \frac{\int_0^{\infty} \left( z + \beta \alpha \eta \right)^3 \exp \left\{ - \frac{z + \beta \alpha \eta}{\theta} \right\} \, dz}{\left( \int_0^{\infty} \left( z + \beta \alpha \eta \right) \exp \left\{ - \frac{z + \beta \alpha \eta}{\theta} \right\} \, dz \right)^3}
\end{align*}
\]

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

\[
\Gamma(b, a) = \int_0^{\infty} t^{b-1} \exp(-at) \, dt,
\]

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

\[
\gamma(b, a) = \int_0^a t^{b-1} \exp(-at) \, dt.
\]

The detailed derivation of (8) is given in Appendix A. By substituting \( \beta = 0 \) and \( \eta = 0 \) into (8), we can estimate the kurtosis before processing. Thus, we can calculate the resultant kurtosis ratio as

\[
\text{kurtosis ratio} = \frac{\mathcal{K}(\alpha, \beta, \eta)}{\mathcal{K}(\alpha, 0, \eta)}
\]

### 3.5. Analysis of Amount of Noise Reduction in Original One-Shot Spectral Subtraction

We analyze the amount of noise reduction due to spectral subtraction. Hereafter we define the noise reduction rate (NRR) as a measure of the noise reduction performance, which is defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB [9]. The NRR is

\[
\text{NRR} = 10 \log \left\{ \frac{E[n_{in}]}{E[n_{out}]} \right\}
\]

where \( n_{in} \) is the power-domain (noise) signal of the input and \( n_{out} \) 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 the gamma distribution. Then, the average power of the input signal is given as \( E[n_{in}] = \alpha \beta \).

Next, the average power of the signal after spectral subtraction is calculated. Here, let \( z \) obey the p.d.f. of the signal after spectral subtraction, \( P_{SS}(z) \), defined by (7), then the average power of the signal after spectral subtraction can be expressed as

\[
E[n_{out}] = E[z] = \int_0^{\infty} z P_{SS}(z) \, dz
\]

We now consider the first term of the right-hand side in (14). We
let $t = z + \beta \theta z$, then $dt = dz$. As a result,

$$
\int_{0}^{\infty} \frac{z}{\theta^{\Gamma(\alpha)}} (z + \beta \theta z)^{-\alpha-1} \exp \left\{-\frac{z + \beta \theta z}{\theta} \right\} dz
= \int_{0}^{\infty} (t - \beta \theta z) \cdot \frac{1}{\theta^{\Gamma(\alpha)}} \cdot t^{-\alpha-1} \exp \left\{-\frac{t}{\theta} \right\} dt
= \int_{0}^{\infty} \frac{\beta \theta z}{\theta^{\Gamma(\alpha)}} \cdot t^{-\alpha} \exp \left\{-\frac{t}{\theta} \right\} dt
- \int_{0}^{\infty} \frac{\beta \theta z}{\theta^{\Gamma(\alpha)}} \cdot t^{-\alpha} \exp \left\{-\frac{t}{\theta} \right\} dt
= \frac{\beta \theta}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta \theta, \alpha + 1)}{\Gamma(\alpha)} - \frac{\beta \theta}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta \theta, \alpha)}{\Gamma(\alpha)}.
$$

Also, we deal with the second term of the right-hand side in (14). We let $t = z/(\eta^2 \theta)$, then $\eta^2 \theta t = dz$, resulting in

$$
\int_{0}^{\eta^2 \theta} \frac{z}{(\eta^2 \theta)^{\Gamma(\alpha)}} z^{-\alpha-1} \exp \left\{-\frac{z}{\eta^2 \theta} \right\} dz
= \int_{0}^{\eta^2 \theta} \frac{1}{(\eta^2 \theta)^{\Gamma(\alpha)}} \int_{0}^{\eta^2 \theta} (\eta^2 \theta)^{\alpha} \exp \{-t(\eta^2 \theta)\} dt
= \frac{\eta^2 \theta}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta \theta, \alpha + 1)}{\Gamma(\alpha)} + \frac{\eta^2 \theta}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta \theta, \alpha)}{\Gamma(\alpha)}.
$$

Using (13), (15), and (16), the noise reduction performance of spectral subtraction, NRR, can be expressed by

$$
\text{NRR}(\alpha, \beta, \eta) = -10 \log \left[ \frac{\Gamma(\beta \theta, \alpha + 1)}{\Gamma(\alpha + 1)} - \beta \cdot \frac{\Gamma(\beta \theta, \alpha)}{\Gamma(\alpha)} + \eta^2 \theta \cdot \frac{\Gamma(\beta \theta, \alpha + 1)}{\Gamma(\alpha + 1)} \right].
$$

### 3.6. Analysis of Behavior of Iterative Spectral Subtraction

#### 3.6.1. Amount of Musical Noise Generation

In this subsection, we formulate the amount of musical noise generated in the iterative spectral subtraction method using the analytical result obtained in Sect. 3.4. Here we conduct a recursively applied kurtosis analysis in the following manner, where the subscript $i$ represents the value in the $i$th iteration:

(I) First, model the input noise p.d.f. by a gamma distribution with the shape parameter $\alpha_i$ (initially $i = 0$).

(II) Next, apply spectral subtraction to the signal using the subtraction parameter $\beta$ and the flooring parameter $\eta$. We calculate the kurtosis using (8), i.e., $\text{Kurt}(\alpha_i, \beta, \eta)$. After step (II), a new $\alpha_{i+1}$ can be calculated using the following relation between the kurtosis and the shape parameter (see Appendix B),

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

This result in the following quadratic equation in $\alpha_{i+1}$ to be solved:

$$
(1 - \text{Kurt}(\alpha_i, \beta, \eta))\alpha_{i+1}^2 + (5 - \text{Kurt}(\alpha_i, \beta, \eta))\alpha_{i+1} + 6 = 0,
$$

and we can derive a closed-form estimate of the shape parameter from the given kurtosis as

$$
\alpha_{i+1} = \left[ \frac{-5 + \sqrt{\text{Kurt}(\alpha_i, \beta, \eta)^2 + 14 \text{Kurt}(\alpha_i, \beta, \eta) + 1}}{2 \text{Kurt}(\alpha_i, \beta, \eta) - 2} \right]^{-1},
$$

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

$$
\text{Kurt}(\alpha_{i+1}, \beta, \eta) = \frac{\gamma(H(\text{Kurt}(\alpha_i, \beta, \eta)))}{\gamma^2(H(\text{Kurt}(\alpha_i, \beta, \eta)), \beta, \eta)}.
$$

Thus, we can calculate the resultant kurtosis ratio as

$$
\text{kurtosis ratio} = \frac{\gamma(H(\text{Kurt}(\alpha_{i+1}, \beta, \eta)))}{\gamma^2(H(\text{Kurt}(\alpha_{i+1}, \beta, \eta)), \beta, \eta)}.
$$

#### 3.6.2. Amount of Noise Reduction

In this subsection, we analyze the amount of noise reduction by carrying out the same iterative analysis that described in Sect. 3.6.1, including the approximation of the gamma distribution modeling.

The NRR in the $i$th iteration is obtained through steps (I) and (II) using (8), i.e., $\text{Kurt}(\alpha_i, \beta, \eta)$. After step (II), a new $\alpha_{i+1}$ can be calculated using the following relation between the kurtosis and the shape parameter (see Appendix B),

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

This result in the following quadratic equation in $\alpha_{i+1}$ to be solved:

$$
(1 - \text{Kurt}(\alpha_i, \beta, \eta))\alpha_{i+1}^2 + (5 - \text{Kurt}(\alpha_i, \beta, \eta))\alpha_{i+1} + 6 = 0,
$$

and we can derive a closed-form estimate of the shape parameter from the given kurtosis as

$$
\alpha_{i+1} = \left[ \frac{-5 + \sqrt{\text{Kurt}(\alpha_i, \beta, \eta)^2 + 14 \text{Kurt}(\alpha_i, \beta, \eta) + 1}}{2 \text{Kurt}(\alpha_i, \beta, \eta) - 2} \right]^{-1},
$$

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

$$
\text{Kurt}(\alpha_{i+1}, \beta, \eta) = \frac{\gamma(H(\text{Kurt}(\alpha_i, \beta, \eta)))}{\gamma^2(H(\text{Kurt}(\alpha_i, \beta, \eta)), \beta, \eta)}.
$$

Thus, we can calculate the resultant kurtosis ratio as

$$
\text{kurtosis ratio} = \frac{\gamma(H(\text{Kurt}(\alpha_{i+1}, \beta, \eta)))}{\gamma^2(H(\text{Kurt}(\alpha_{i+1}, \beta, \eta)), \beta, \eta)}.
$$

In summary, we can derive theoretical estimates for the amount of musical noise and NRR using (22) and (23). This greatly simplifies the analysis because both equations are expressed analytically in a form that does not include any integrals.
is fixed to 2, and the following parameter 0.97, corresponding to 3.6.3.

Figure 2. Musical noise generated in iterative spectral subtraction. In addition, this can be compared with that generated in the conventional one-shot spectral subtraction method under the same amount of noise reduction.

Figure 2 shows the theoretical behavior of the kurtosis ratio and NRR for several parameter settings, where the shape parameter \( \alpha \) is set to 1.0, i.e., the noise signal is assumed to be Gaussian. In the iterative spectral subtraction, the oversubtraction parameter \( \beta \) is fixed to 2, and the flooring parameter \( \eta \) is set to 0.5, 0.9, and 0.97, corresponding to normal, weak, and very weak processing in each iteration, respectively. In the conventional one-shot spectral subtraction, the oversubtraction parameter and flooring parameter are optimally adjusted so that the NRR is varied as 0, 0.5, 1.0,..., 12 dB.

From Fig. 2, we can confirm the following interesting results:

- The recursive use of very weak spectral subtraction (e.g., \( \eta = 0.97 \)) can simultaneously achieve a large NRR and a small kurtosis ratio after a large number of iterations, meaning that we can realize high-quality speech enhancement with a small amount of musical noise generated. This is strong theoretical evidence of the advantage of iterative spectral subtraction.

- Moreover, there exists an appropriate parameter setting (\( \eta = 0.97 \)) that gives equilibrium behavior in the growth of the kurtosis ratio, i.e., almost no musical noise is generated.

- In contrast, if we use strong subtraction with a small flooring parameter (e.g., \( \eta = 0.5 \)), the above-mentioned equilibrium is violated, resulting in large kurtosis ratio compared with that of conventional one-shot spectral subtraction under the same NRR. This suggests that the iterative method is not always advantageous, and that the values of the parameters should be carefully set.

4. EVALUATION EXPERIMENT AND RESULT

We conducted an evaluation experiment using real noisy speech data to confirm the validity of the theoretical analysis described in the previous section. Noisy observation signals were generated by adding a noise signal to target speech signals with an SNR of 0 dB. The target speech signals were the utterances of one speaker

(1 sentence), and the noise signal was white Gaussian noise. The length of each signal was 7 s, and each signal was sampled at 16 kHz. In this experiment, we assumed that the noise prototype \( \hat{N}(f, \tau) \) was perfectly estimated. The parameter settings of \( \beta \) and \( \eta \) are the same as those in Sect. 3.6.3.

The result of the experiment is depicted in Fig. 3. The kurtosis ratio and NRR were calculated from the observed and processed signals. The figure shows that the kurtosis ratio decreases as the floor parameter increases, and we can confirm the efficacy of iterative spectral subtraction if we use weak processing in each iteration. The resultant tendency is almost consistent with the result of the theoretical analysis in Sect. 3.6.3, except for the absolute value of the kurtosis ratio. The discrepancy between the kurtosis ratio obtained from the real processed data and the theoretical estimate is mainly due to the gamma-distribution approximation introduced in our analysis. Although this result is valid only for Gaussian noise, it still remains an open problem that the objective and subjective evaluations using actual noises are conducted.

5. CONCLUSION

In this study, we performed a theoretical analysis of the amount of musical noise generated via iterative spectral subtraction based on higher-order statistics. Also, we conducted a comparison of the amount of musical noise generated for different parameter settings under the same noise reduction performance. It was clarified from mathematical analysis and evaluation experiments that iterative spectral subtraction with very weak processing can result in less musical noise being generated.

A. APPENDIX: DERIVATION OF (8)

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

\[
\mu_2 = \int_0^\infty z^2 \cdot P_{SS}(z) dz
\]

\[
= \int_0^\infty z^2 \frac{1}{\theta^\beta \Gamma(\alpha)} (z + \beta \theta)^{-\beta-1} \exp \left( -\frac{z + \beta \theta}{\theta} \right) dz
\]

\[
+ \int_0^\infty z^2 \frac{1}{(\theta^\beta)^{\Gamma(\alpha)} \exp(\frac{z}{\theta})} \exp \left( -\frac{z}{\theta} \right) dz.
\]
We now expand the first term of the right-hand side of (24). Here, let $t = (z + \beta x)/\theta$, then $\theta dt = dz$ and $z = \theta \tau - \beta x$. Consequently,
\[
\int_0^{\infty} \frac{1}{\theta^2 r(\alpha)} (z + \beta x \theta)^{r-1} \exp \left( -\frac{z + \beta x \theta}{\theta} \right) \, dz
\]
\[
= \int_0^{\infty} \theta^2 \left( \frac{1}{\theta^2 r(\alpha)} \right) (\theta t)^{r-1} \exp[-t] \, dt
\]
\[
= \frac{\theta^2}{r(\alpha)} \int_0^{\infty} (t^2 - 2\beta k\theta + \beta^2 \theta^2 \sigma^{-1}) \exp[-t] \, dt
\]
\[
= \frac{\theta^2}{r(\alpha)} \left[ \Gamma(\beta k\theta + 2) - 2\beta k\Gamma(\beta k\theta, \alpha + 1) + \beta^2 \beta^2 \sigma^{-1} \Gamma(\beta k\theta, \alpha) \right].
\] (25)

Next we consider the second term of the right-hand side of (24). Here, let $t = x/\theta(t')$, then $\theta dt = dx$. Thus,
\[
\int_0^{\infty} \frac{z^2}{\theta^2 r(\alpha)} (z + \beta x \theta)^{r-1} \exp \left( -\frac{z + \beta x \theta}{\theta} \right) \, dz
\]
\[
= \int_0^{\infty} \frac{(\theta t)^2}{\theta^2 r(\alpha)} (\theta t)^{r-1} \exp[-t] \, dt
\]
\[
= \frac{\theta^2}{r(\alpha)} \int_0^{\infty} t^2 \exp[-t] \, dt
\]
\[
= \theta^2 \frac{\gamma(\alpha, \alpha + 2)}{r(\alpha)}.
\] (26)

As a result, the 2nd-order moment after spectral subtraction, $\mu_{2SS}$, is a composite of (25) and (26), and is given as
\[
\mu_{2SS} = \frac{\theta^2}{r(\alpha)} \left[ \Gamma(\beta k\theta + 2) - 2\beta k\Gamma(\beta k\theta, \alpha + 1) + \beta^2 \beta^2 \sigma^{-1} \Gamma(\beta k\theta, \alpha) \right] + \frac{\theta^2}{r(\alpha)} \Gamma(\beta k\theta, \alpha + 4).
\] (27)

In the same manner, the 4th-order moment after spectral subtraction, $\mu_4^{SS}$, can be represented by
\[
\mu_4^{SS} = \frac{\theta^2}{r(\alpha)} \left[ \Gamma(\beta k\theta, \alpha + 4) - 4\beta k\Gamma(\beta k\theta, \alpha + 3) + 6\beta^2 \beta^2 \Gamma(\beta k\theta, \alpha + 2) - 4\beta^3 \beta^3 \Gamma(\beta k\theta, \alpha + 1) + \beta^4 \beta^4 \Gamma(\beta k\theta, \alpha + 4) \right].
\] (28)

Consequently, using (27) and (28), the kurtosis after spectral subtraction is given as
\[
kurtSS = \frac{\mathcal{F}(\alpha, \beta, \eta)}{\mathcal{G}(\alpha, \beta, \eta)}
\] (29)
where
\[
\mathcal{G}(\alpha, \beta, \eta) = \Gamma(\beta k\theta, \alpha + 2) - 2\beta k\Gamma(\beta k\theta, \alpha + 1) + \beta^2 \beta^2 \sigma^{-1} \Gamma(\beta k\theta, \alpha + 2),
\]
\[
\mathcal{F}(\alpha, \beta, \eta) = \Gamma(\beta k\theta, \alpha + 4) - 4\beta k\Gamma(\beta k\theta, \alpha + 3) + 6\beta^2 \beta^2 \Gamma(\beta k\theta, \alpha + 2) - 4\beta^3 \beta^3 \Gamma(\beta k\theta, \alpha + 1) + \beta^4 \beta^4 \Gamma(\beta k\theta, \alpha + 4).
\] (30) (31)

**B. APPENDIX: DERIVATION OF (18)**

First, we represent the 2nd-order moment as
\[
\mu_2 = \int_0^{\infty} x^2 p(x) \, dx = \int_0^{\infty} x^2 \frac{1}{\Gamma(\alpha) \theta^\alpha} \cdot x^{\alpha-1} e^{-x/\theta} \, dx.
\] (32)

Here, let $X = x/\theta$; then this moment can be rewritten as
\[
\mu_2 = \frac{\theta^2}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+2-1} e^{-x} \, dx
\]
\[
= \frac{\theta^2}{\Gamma(\alpha)} \Gamma(\alpha + 2) = \theta^2 (\alpha + 1) \theta.
\] (33)

where we use the following well-known functional equation of the gamma function.
\[
\Gamma(\alpha) = (\alpha - 1) \cdots (\alpha - j) \Gamma(\alpha - j).
\] (34)

Next, in the same manner, the 4th-order moment can be expressed as
\[
\mu_4 = \theta^4 (\alpha + 3)(\alpha + 2)(\alpha + 1) \theta.
\] (35)

Using (33) and (35), we have
\[
kurtSS = \theta^2 + 2\theta^2 (\alpha + 3)(\alpha + 2)(\alpha + 1) \theta.
\] (36)

**C. REFERENCES**


