

RESEARCH

Open Access

Noise reduction for periodic signals using high-resolution frequency analysis

Toshio Yoshizawa, Shigeki Hirobayashi* and Tadanobu Misawa

Abstract

The spectrum subtraction method is one of the most common methods by which to remove noise from a spectrum. Like many noise reduction methods, the spectrum subtraction method uses discrete Fourier transform (DFT) for frequency analysis. There is generally a trade-off between frequency and time resolution in DFT. If the frequency resolution is low, then the noise spectrum can overlap with the signal source spectrum, which makes it difficult to extract the latter signal. Similarly, if the time resolution is low, rapid frequency variations cannot be detected. In order to solve this problem, as a frequency analysis method, we have applied non-harmonic analysis (NHA), which has high accuracy for detached frequency components and is only slightly affected by the frame length. Therefore, we examined the effect of the frequency resolution on noise reduction using NHA rather than DFT as the preprocessing step of the noise reduction process. The accuracy in extracting single sinusoidal waves from a noisy environment was first investigated. The accuracy of NHA was found to be higher than the theoretical upper limit of DFT. The effectiveness of NHA and DFT in extracting music from a noisy environment was then investigated. In this case, NHA was found to be superior to DFT, providing an approximately 2 dB improvement in SNR.

1. Introduction

Noise reduction to recover a target signal from an input waveform is important in a number of fields. We usually use a frequency spectrum to remove noise from the input waveform. Although it is difficult to distinguish a signal from the noise in the time domain, this task tends to become easier in the frequency domain. However, it is difficult to filter out noise that is similar to a signal. For example, the consonant, which is the part of the sound that has a frequency spectrum that is similar to a noise. This study proposes a basic technology by which to remove a noise from musical sound including several periodic signals. We selected white noise and pink noise as the noise signals. These noises are common in cities as well as in nature and have a continuous spectrum. Based on this study, we can remove white noise, including wideband noise such as pulse and white noise, from an old music recording in order to apply digital remastering in multimedia industries. We will also be able to remove noise from a recording of a singing voice because this is a periodic signal. When listening to music in a high-noise environment, difficulty in hearing the music and the

presence of ambient noise can decrease the level of enjoyment. Therefore, various noise reduction methods are being investigated, and a number of noise reduction techniques have been proposed. The spectral subtraction method (SS method) is a widely used approach [1] in which the target signal is extracted from a noisy signal by measuring the noise in advance and modeling the statistical spectral envelope characteristics [2-4]. The SS method does not require multiple microphones, and highly effective results can be obtained by using a relatively simple algorithm. For this reason, many techniques for improving the SS method have been proposed. Sorensen and Andersen [5] also used the SS method in combination with speech presence detection. Soon and Koh [6] and Ding et al. [7] treated audio signals as graphics and applied 2D and 1D Wiener filters in the frequency domain for noise reduction. The advantage of this method is the possibility of frame-to-frame correlation. In addition, the amplitude in the frequency domain can be adjusted and an unmodified initial phase can be used. Finally, Virag [8] and Udrea et al. [9] suggested an SS method based on the characteristics of the human auditory system.

However, using unmodified noisy phases limits the noise reduction effect. In general, the discrete Fourier

* Correspondence: hirobays@eng.u-toyama.ac.jp
Department of Intellectual Information Systems Engineering, Faculty of Technology, University of Toyama, 3190 Gofuku, Toyama-shi, Toyama, Japan

transform (DFT) is used to obtain the spectral characteristics during preprocessing for the SS method. The frequency resolution of the DFT is restricted because it depends on the analytical frame length and the window function. If the frequency resolution is low, the noise spectrum can overlap the spectrum of the signal source, which makes it difficult to extract the original signal. Energy leaks into another band and side lobes are generated when the frequency of the analytic signal does not correspond to an integral multiple of the base frequency. In harmonic frequency analysis, there is then a high probability of overlap between the side-lobes of the source spectrum and the noise spectrum. If the side-lobes are removed, then the signal source can fully be recovered. Similarly, if the time resolution is low, then rapid frequency variations cannot be detected. In order to solve this problem, Kauppinen and Roth attempted to increase the frequency resolution by applying an extrapolation method to the signal frame in the time domain [10]. In this study, we have applied non-harmonic analysis (NHA), which has a high frequency resolution with limited influence of the frame length [11], to the problem of noise reduction. For a similar frame length, NHA is expected to achieve better frequency resolution than the length extrapolation method used in [10]. Therefore, we investigated the use of NHA as an alternative preprocessing method to DFT for noise reduction. Since the effects of frequency resolution can best be evaluated for periodic signals, sounds produced by musical instruments were used in this study, and preliminary noise reduction experiments were performed.

The remainder of this article is organized as follows. In Section 2, we provide an introduction to the NHA algorithm. In Section 3, we investigate noise reduction using single sinusoidal waves. Section 4 describes the side-lobe suppression experiments. In Section 5, noise reduction experiments are carried out using sounds produced by musical instruments, and the results are described in Section 6.

2. The NHA method

2.1 Background

The DFT is generally used for frequency analysis. A discrete spectrum X of the discrete time signal $x(n)$ of length N can be expressed as

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \quad (k = 0, 1, 2, \dots, N-1). \quad (1)$$

When the sampling frequency is Δt and the original signal $x(n)$ has a period of $N\Delta t/k$, $X(k)$ can accurately reflect the spectral structure. However, if a period other than $N\Delta t/k$ appears in $x(n)$, $X(k)$ is expressed by the combination of $N\Delta t/k$ in terms of several frequency

components, and $X(k)$ is not accurately reflected in the spectral structure.

In order to increase the frequency resolution, the value of N is generally increased. If the frequency is accompanied by a temporal fluctuation, however, then the average period is extracted and the analytical accuracy deteriorates as N is increased. Some techniques use an analysis window function for $x(n)$ in preprocessing. However, this does not improve the apparent frequency resolution.

Figure 1 shows some of the problems associated with frequency analysis. Even when analyzing the simplest frequency signal shown at the top of Figure 1, one portion of the section is removed when determining the periodicity of the analyzed signal. The center left section of Figure 1 shows the analytical accuracy. The period can accurately be identified only if the frame length is a multiple of the period of the analyzed signal. In other words, a group of different spectra appear near the true frequency because the analyzed signal is expressed as a multiple number of periods $N\Delta t/k$. In order to prevent this, an analysis window function may be used, as shown in the center right section of Figure 1. However, this will merely concentrate around the true value, making it difficult to determine the true value. We, therefore, noted that the Fourier coefficient could be estimated by solving a nonlinear equation based on the assumption of a stationary signal (see the bottom of Figure 1). Thus, the NHA developed in this study achieves a high analytical accuracy because this NHA reduces the influence of the analysis window.

2.2 Algorithm of NHA

Figure 2 shows the algorithm used by NHA. First, a frequency analysis of the input signal is carried out by fast Fourier transform (FFT) for obtaining the initial value. Next, the frequency and initial phase of the spectral component that has the largest amplitude are converged using a cost function with the steepest descent method. At this time, a weighting coefficient based on the retardation method is applied to convert the cost functions calculated by the recurrence formulas into a monotonically decreasing sequence. The amplitude is then converged using Newton's method. Following this, Newton's method is applied again to converge both the frequency and the initial phase to a high degree of accuracy. Following a final convergence of the amplitude using Newton's method, we obtain the fully converged spectrum.

Finally, we describe the motivation for the structure shown in Figure 2. For the cost function equation, given by Equation 2, although the convergence speed is slow, the steepest descent method can find the stationary point within a wide range. In contrast, the Newton method can quickly find a nearby stationary point.

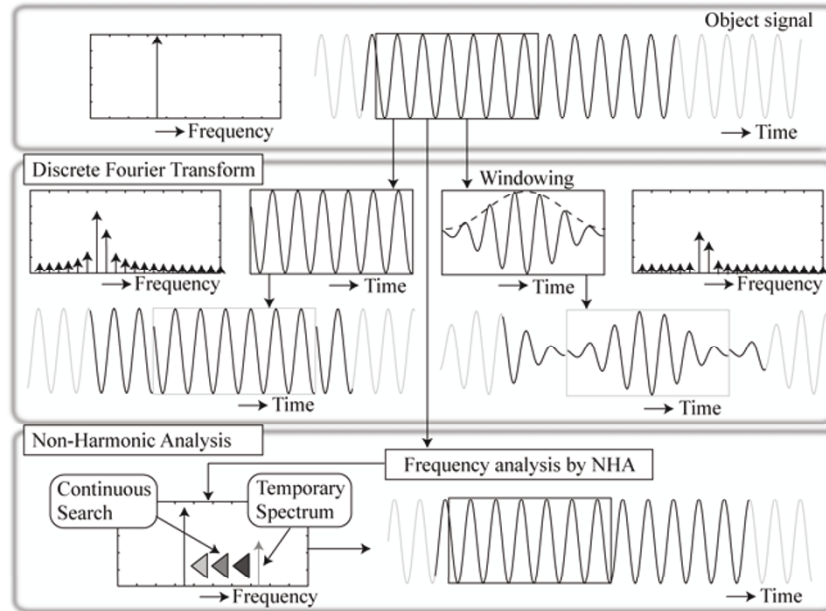


Figure 1 Fourier transform and NHA technique.

Therefore, we first use the steepest descent method to find the stationary point within a wide range. Then, we use the Newton method to quickly find a stationary point. Either way, we distinguish the convergence calculation of amplitude A from the other parameters, so that the local stationary point will not be calculated incorrectly.

2.3 Details of NHA

In this section, we present a more detailed description of the NHA method. Since the Fourier coefficient is estimated by solving a nonlinear equation, NHA enables the frequency and its associated parameters to be accurately estimated without being significantly affected by the frame length. In order to minimize the sum of

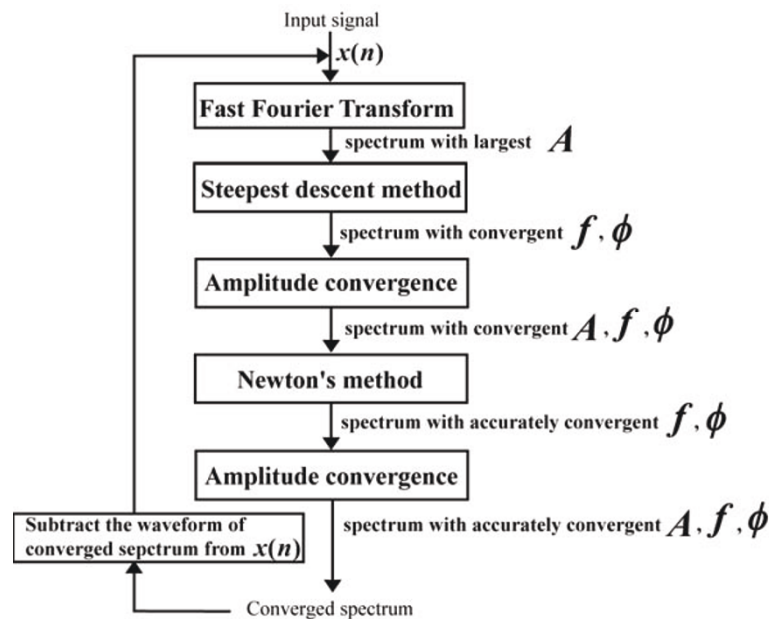


Figure 2 NHA algorithm.

squares of the difference between the object signal and the sinusoidal model signal, the frequency \hat{f} , amplitude \hat{A} , and initial phase $\hat{\phi}$ are calculated using the cost function, as follows:

$$F(\hat{A}, \hat{f}, \hat{\phi}) = \frac{1}{N} \sum_{n=0}^{N-1} \left\{ x(n) - \hat{A} \cos \left(2\pi \frac{\hat{f}}{f_s} n + \hat{\phi} \right) \right\}^2, \quad (2)$$

where N is the frame length and f_s is the sampling frequency ($f_s = 1/\Delta t$).

2.3.1. Steepest descent method

George and Smith [12,13] attempted to introduce the signal parameter A and the initial phase ϕ by applying the least mean squares method to the difference signal between the analyzed signal and the modulated harmonic sinusoidal wave.

However, this method is strongly dependent on the frame length and is difficult to apply to the analysis of signals that do not have a simple frequency harmonic structure because frequencies that are dependent on the frame length are used for the group of harmonic frequencies, as in DFT. In other words, small frequency changes cannot be detected.

By focusing on the problem of solving a nonlinear equation, we apply the nonlinear equation process to Equation 2 for optimum calculation of the frequency f , as well as the parameter amplitude A and initial phase ϕ . Figure 3 shows an example of the characteristics of \hat{f} and $\hat{\phi}$ in the evaluation function of Equation 2, enlarged

around the true value, where N is 512, f_s is 512, and the true values of A , f , and ϕ are 1, 100 Hz, and 0.5π rad, respectively. Since small values are given in black, troughs appear as black and peaks as white. In other words, Equation 2 is a multimodal nonlinear evaluation function. Around the true value ($\hat{f} = 100$, $\hat{\phi}/(2\pi) = 0.5$), minimum and maximum values are aligned vertically. This is because the true value is a minimum but becomes a maximum for the antiphase case ($\phi(2\pi) = 0, 1$). Since the trough at the minimum value is 2 Hz wide, the minimum of the evaluation function can be estimated only if the initial value lies in the trough when solving the nonlinear equation. Since the DFT frequency resolution is 1 Hz, one or two points can be contained in a trough that is 2 Hz wide. At the point on the frequency axis where the DFT amplitude becomes maximum (i.e., the integral frequency when the frame length is 1 s), the evaluation function of Equation 2 is minimized at the initial phase determined by DFT.

If the maximum amplitude A determined by DFT and the frequency f and initial phase ϕ are used as initial values ($A_{0,0}$, $f_{0,0}$, $\phi_{0,0}$), then the initial values can be given inside the trough containing the minimum of cost function in Figure 3.

Therefore, in order to obtain an accurate spectrum, we use the initial value ($A_{0,0}$, $f_{0,0}$, $\phi_{0,0}$), which is converged using the nonlinear equation process. Considering Equation 2 as the cost function, this nonlinear problem is converted into a minimization problem, and $\hat{f}_{m,p}$ and $\hat{\phi}_{m,p}$ are determined using the steepest descent

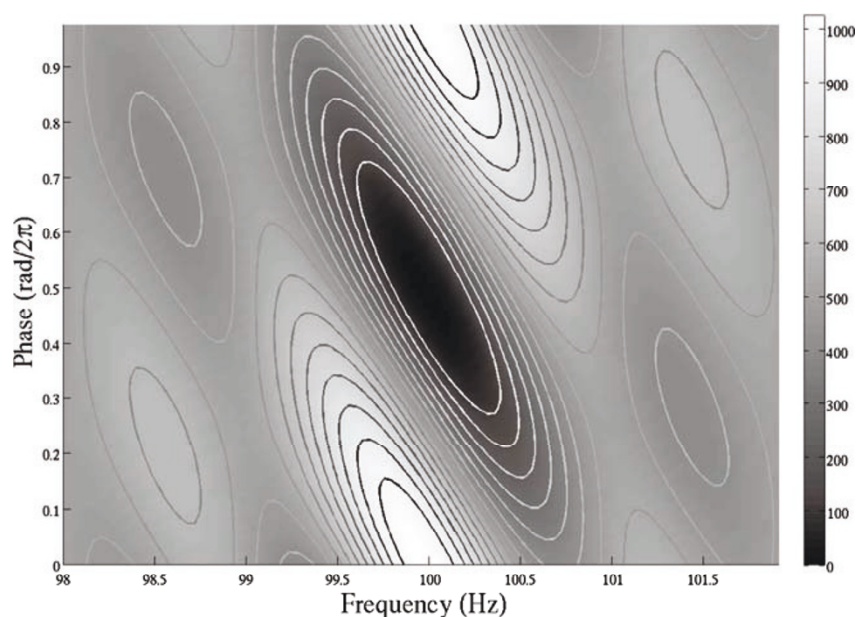


Figure 3 Distribution of the cost function.

method and the retardation method to obtain the following expressions:

$$\hat{f}_{m,p} = \hat{f}_{m,0} - \mu_{m,p} \frac{\partial F_{m,0,0}}{\partial f}, \quad (3)$$

$$\hat{\phi}_{m,p} = \hat{\phi}_{m,0} - \mu_{m,p} \frac{\partial F_{m,0,0}}{\partial \phi}, \quad (4)$$

where p is the operated number of the retardation methods for the frequency and the phase, and m is the number of iterations of the steepest descent method. We use the following shorthand

$$F_{m,p,q} = F(\hat{A}_{m,q}, \hat{f}_{m,p}, \hat{\phi}_{m,p}), \quad (5)$$

where q is the number of iterations of the retardation method. These variables are iterated as shown in Figure 4. In the above equations, $\mu_{m,p}$ is a weighting coefficient based on the retardation method and has a value between 0 and 1 to convert the cost functions calculated by recurrence formulas into a monotonically decreasing sequence [14-16]. In this article, we use this weighting coefficient as follows

$$\mu_{m,p+1} = 0.5\mu_{m,p}, \quad (6)$$

where $\mu_{m,1}$ is set to 1.

This series of calculations is repeated to cause $\hat{f}_{m,p}$ and $\hat{\phi}_{m,p}$ to converge with high accuracy until the following conditions occur:

$$F_{m,p,0} < ((1 - 0.5\mu_{m,p}) \cdot F_{m,0,0}). \quad (7)$$

The next step is the convergence of the amplitude.

2.3.2. Amplitude convergence

Here, A can be uniquely determined only if $\hat{f}_{m,p}$ and $\hat{\phi}_{m,p}$ are known, and the following formula is used to cause A to converge:

$$\hat{A}_{m,q} = \hat{A}_{m,0} - v_{m,q} \frac{\partial F_{m,p,0}}{\partial A} \quad (8)$$

Similarly, $\mu_{m,p}$ and $v_{m,q}$ are weighting coefficients based on the retardation method [14-16] and are given by

$$v_{m,q+1} = 0.5v_{m,q}, \quad (9)$$

with $v_{m,1} = 1$. This causes $\hat{A}_{m,q}$ to converge with a high degree of accuracy until

$$F_{m,p,q} < ((1 - 0.5v_{m,q}) \cdot F_{m,p,0}). \quad (10)$$

Then, $\hat{A}_{m+1,0}$, $\hat{f}_{m+1,0}$, and $\hat{\phi}_{m+1,0}$ are set to $\hat{A}_{m,q}$, $\hat{f}_{m,p}$, and $\hat{\phi}_{m,p}$, and q and p are reset to 1.

Next, the steepest descent method and the amplitude converging algorithm are recursed until the cost function becomes partially converged. Newton's method is then applied.

2.3.3. Newton's method

Although the steepest descent method causes values to converge over a comparatively wide range, a single series of operations cannot ensure sufficient accuracy. In order to achieve a highly accurate conversion, NHA uses Newton's method following the lower accuracy steepest descent method. The following recurrence formula is used for Newton's method:

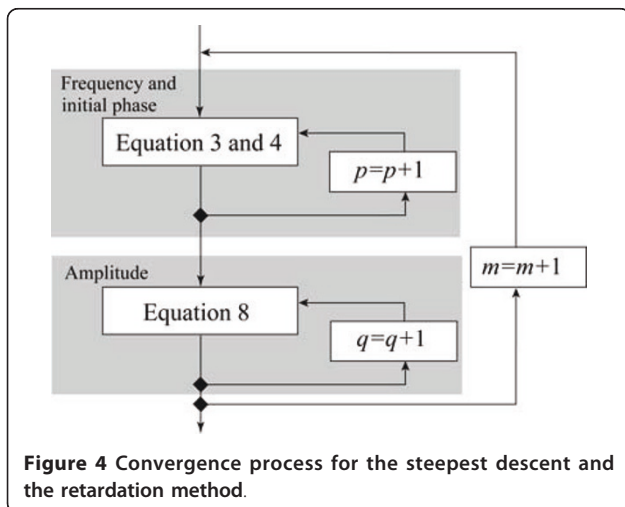
$$\hat{f}_{m,p} = \hat{f}_{m,0} - \frac{\mu_{m,p}}{J} \begin{vmatrix} \frac{\partial F_{m,0,0}}{\partial f} & \frac{\partial^2 F_{m,0,0}}{\partial f \partial \phi} \\ \frac{\partial^2 F_{m,0,0}}{\partial \phi \partial f} & \frac{\partial^2 F_{m,0,0}}{\partial \phi^2} \end{vmatrix}, \quad (11)$$

$$\hat{\phi}_{m,p} = \hat{\phi}_{m,0} - \frac{\mu_{m,p}}{J} \begin{vmatrix} \frac{\partial^2 F_{m,0,0}}{\partial f^2} & \frac{\partial F_{m,0,0}}{\partial f} \\ \frac{\partial^2 F_{m,0,0}}{\partial f \partial \phi} & \frac{\partial F_{m,0,0}}{\partial \phi} \end{vmatrix}, \quad (12)$$

where

$$J = \begin{vmatrix} \frac{\partial^2 F_{m,0,0}}{\partial f^2} & \frac{\partial^2 F_{m,0,0}}{\partial f \partial \phi} \\ \frac{\partial^2 F_{m,0,0}}{\partial \phi \partial f} & \frac{\partial^2 F_{m,0,0}}{\partial \phi^2} \end{vmatrix}, \quad (13)$$

and m is the number of iterations of Newton's method. In addition, $\mu_{m,p}$ is similarly obtained from Equation 6. This series of calculations is also repeated



to cause \hat{f}_m and $\hat{\phi}_m$ to converge accurately. After applying Equations 11 and 12, \hat{A}_m is made to converge by applying Equation 8 in the same manner as in the steepest descent method, and the series of calculations is repeated. The only difference is that the converging algorithm is repeated using Newton's method instead of the steepest descent method. Thus, the frequency parameters are estimated to a high degree of accuracy and at high speed by using a hybrid process combining the steepest descent and Newton's method.

2.3.4. Sequential reduction

Even for the case in which there are several sinusoidal waves, the spectral parameters can approximately be derived by sequential reduction. Here, $x(n)$ is expressed as the sum of K sinusoidal waves in the following manner:

$$x(n) = \sum_{k=1}^K \left\{ A_k \cos \left(2\pi \frac{f_k}{f_s} n + \phi_k \right) \right\}. \quad (14)$$

According to Parseval's theorem, the object signal frequency f_k and the model signal's frequency \hat{f} do not match, i.e., if

$$f_k \neq \hat{f}, \quad (15)$$

then

$$F(\hat{A}, \hat{f}, \hat{\phi}) = \hat{A}^2 + \sum_{k=1}^K \hat{A}_k^2. \quad (16)$$

In addition, if the pair of \hat{f} and $\hat{\phi}$ matches either f_k or ϕ_k , then

$$F(\hat{A}, \hat{f}, \hat{\phi}) = (\hat{A}^2 - A_j)^2 + \sum_{k=1, k \neq j}^K \hat{A}_k^2. \quad (17)$$

If both A_j and A match, then a frequency component of an estimated spectrum can completely be removed from an object signal. Therefore, the problem of acquiring an optimum solution is frequency independent and is applicable even to a signal consisting of several sinusoidal waves by sequential and individual estimation from the object signal. In other words, even when the object signal is a composite sinusoidal wave, several sinusoidal waves can be extracted by performing similar processing on sequential residual signals. If the frequencies of two spectra are adjacent to each other, the other spectrum generates another trough in the trough around the true value shown in Figure 3 and distorts the evaluation function. This may result in an error, as discussed later herein.

2.4. Accuracy of NHA

Among the techniques based on DFT, generalized harmonic analysis (GHA or Hirata's algorithm) is generally considered to have the highest accuracy [17-20].

According to these analyses, the frequency resolution depends on the frame length because one analysis window apparently has the length of several windows. However, the decomposition frequency has a finite length, and an object signal of any other frequency cannot be analyzed. Figure 5 shows the numbers of frequencies that can be analyzed by DFT and GHA at each frame length. Successful frequency analysis means that the number of spectra of the object signal matches the number of spectra after analysis, that is, if the frame length is unique, then DFT has N decomposition frequencies ($0, f_s/N, 2f_s/N, \dots, (N-1)f_s/N$ [Hz]). Compared to DFT of approximately half the data length, GHA is one order of magnitude more accurate. If the spectrum of the object signal is not in the group of the harmonic spectra, the group of harmonic spectra appears near the true frequency.

In order to verify the frequency resolution of NHA, we compared DFT and GHA experimentally, as shown in Figure 6. With the frame length set to 1 s (512 samples), we analyzed a single sinusoidal wave. By each technique, one sinusoidal wave was extracted, and the square of the error from the original signal was examined.

DFT exhibited low analytical accuracy except when the signals had frequencies that were integral multiples of the fundamental frequency. At frequencies above 1 Hz, GHA exhibited accuracies that were two to five orders of magnitude greater. At the same frequencies, NHA was 10 or more orders of magnitude more accurate than DFT. At frequencies below 1 Hz, DFT and GHA were equally accurate, but NHA was able to estimate the frequency

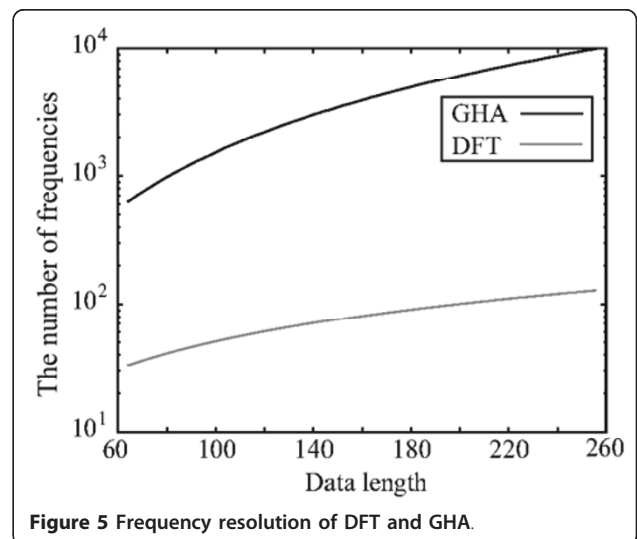


Figure 5 Frequency resolution of DFT and GHA.

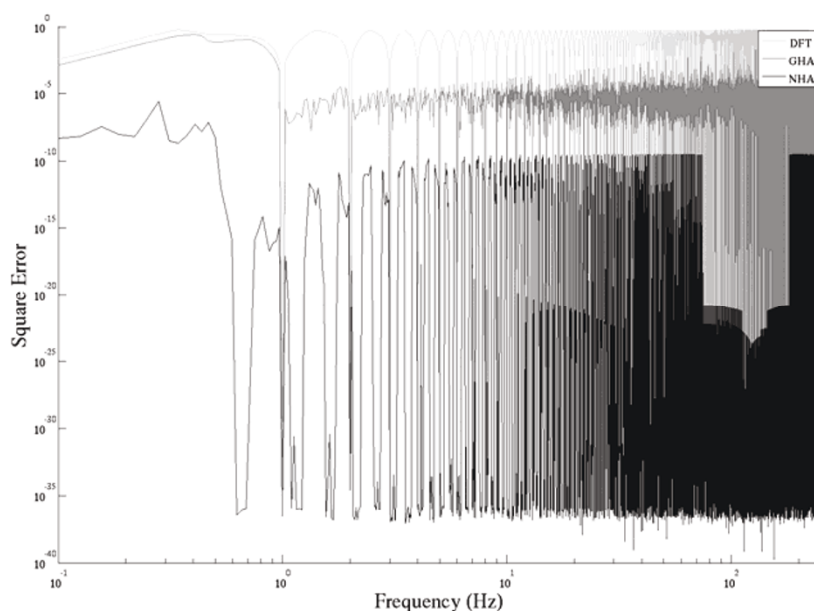


Figure 6 Square error (frame length: 512).

and other parameters correctly without being affected by the frame length. Thus, NHA was demonstrated to have an even greater analysis accuracy than GHA, which was developed from DFT.

Accurate estimation at frequencies below 1 Hz means that even object signals having periods longer than the frame length can accurately be analyzed. Therefore, it may be possible to accurately estimate the spectral structures of signals representing stock prices and other fluctuation factors.

Figures 7 and 8 show the square errors of two sinusoidal waves. A similar evaluation to that in Figure 6 was performed by adding another sinusoidal wave ($f = 0.6$ Hz) in order to determine whether both sinusoidal waves could be correctly extracted.

The ratio of the amplitudes of the two sinusoidal waves is 1:1 in Figure 7 and 1:10 in Figure 8. The latter is the sinusoidal wave ratio at $f = 0.6$ Hz. In both cases, the accuracy increases in the order of NHA, GHA, and DFT. If the two sinusoidal waves have similar amplitudes, the evaluation functions shown in Figure 3 interfere with each other, increasing the distortion, which results in a greater error than that when only one sinusoidal wave is used. As mentioned above, this tendency becomes more noticeable as the frequencies become closer to each other. However, the NHA error is less than the average, as compared to the errors of DFT and GHA.

3. Extracting single sinusoidal waves

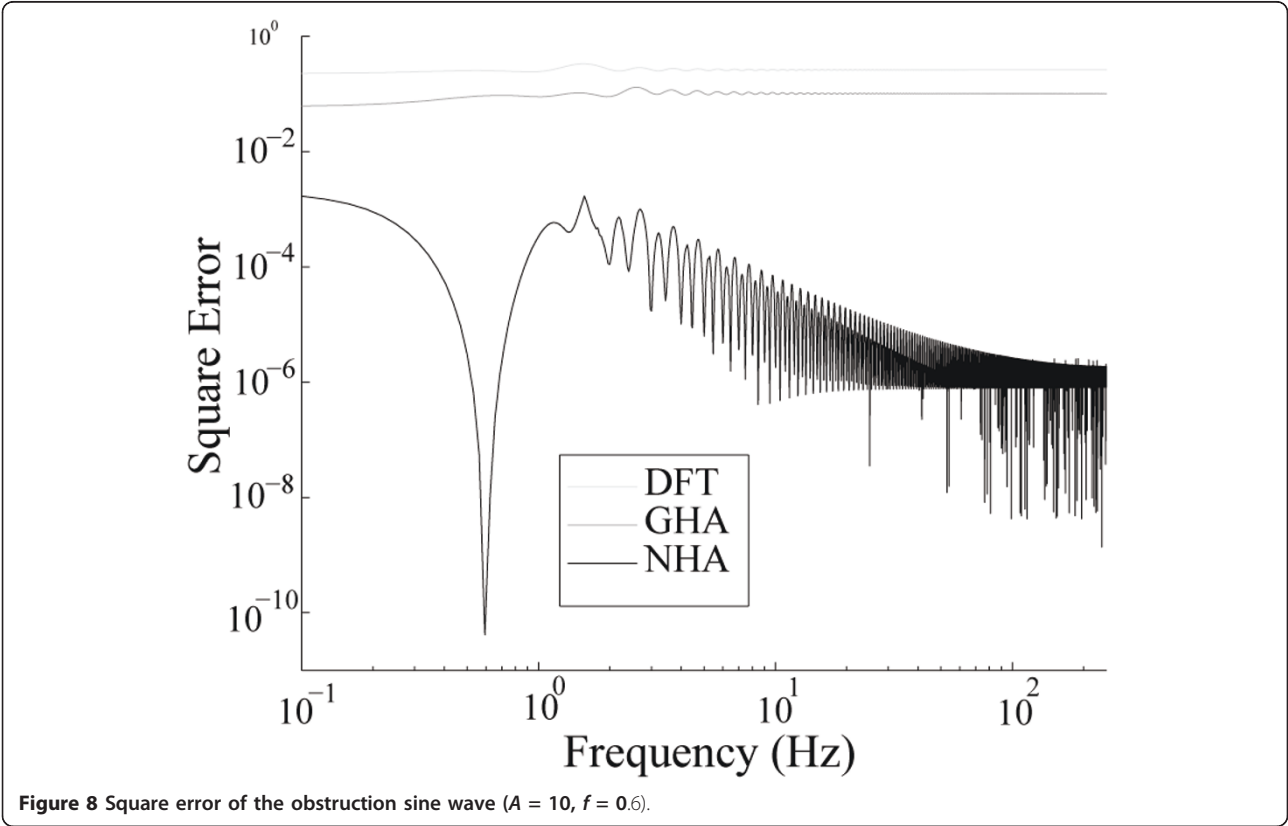
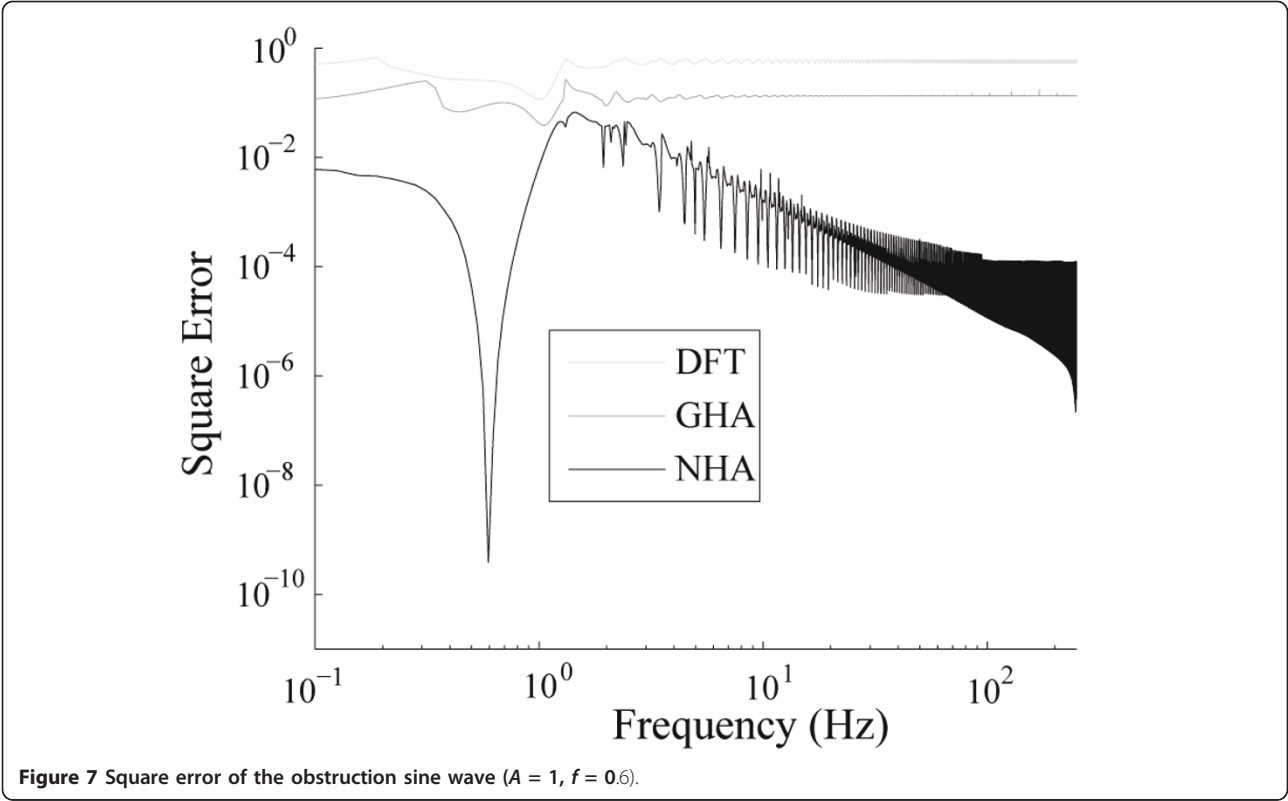
In this section, a quantitative comparison of the extraction accuracy and the calculation time of DFT and

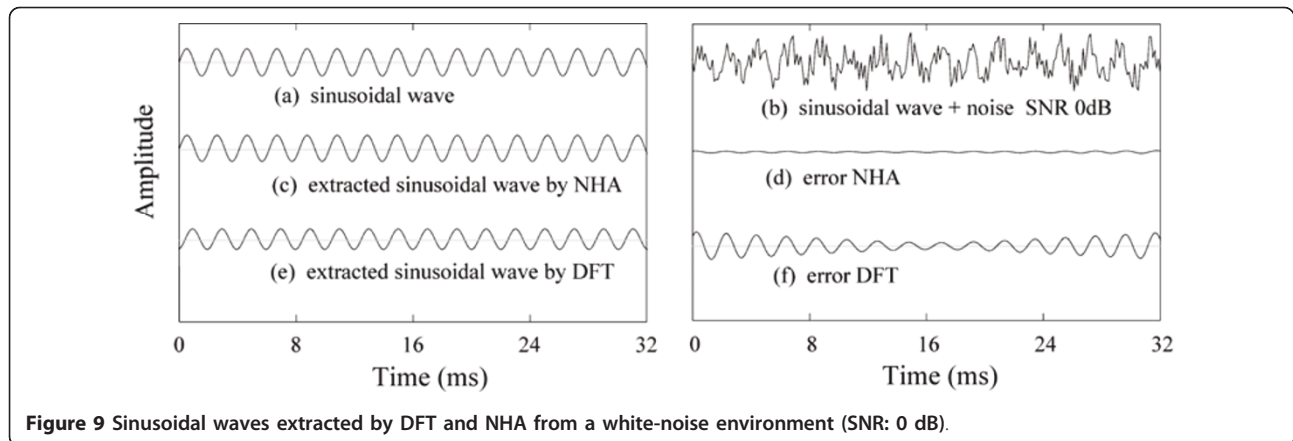
NHA is performed. A single sinusoidal wave in a noisy environment was used for the experiment. For each method, an optimum spectrum (closest to the target signal frequency) was selected and converted to a waveform for evaluation. For DFT, f is necessarily an integral multiple of the fundamental frequency. For the calculations, the frame length was set to 256, and the sampling frequency was set to 488 kHz. The sinusoidal wave was set to 488 Hz in order to investigate frequencies that DFT could not estimate.

Figure 9 shows the sinusoidal wave extracted by DFT and NHA from a white-noise environment in which the SNR was 0 dB, where (a) is the 488 Hz target signal and (b) is the added white noise signal.

Figure 9c, 9e are the signals detected by NHA and DFT, respectively, and (d) and (f) are the residual signals obtained by subtracting (c) and (e) from the target signal. This figure shows that NHA more accurately extracts the original signal. When noise is added to the signal, DFT produces errors if the frequency is not a multiple of the fundamental frequency. The output SNR was approximately 24 dB when NHA was used for extraction and approximately 4 dB when DFT was used. Thus, an improvement of approximately 20 dB was confirmed.

These calculations were performed using a personal computer (CPU: Intel Core i7-930@2.8 GHz, Memory: 6 GB). The time required for calculating a signal consisting of 256 samples by DFT and NHA are 2.8 and 12.0 ms, respectively. It is noted that DFT is calculated by the fastest FFT using a radix-2 number in this article.





For statistical verification at various target signal frequencies, an extraction experiment was conducted in which the frequency f and the initial phase ϕ of the target signal were varied 1,000 times in different noise environments using uniformly distributed random numbers. The range of f and ϕ was $0 < f < 4000$ and $-\pi < \phi < \pi$, respectively. In this case, the amplitude A was maintained constant. The input signal was generated by adding white noise to a single sinusoidal wave. Throughout the experiments, the input SNR was maintained in the range from -10 to +10 dB and was varied in 5-dB steps.

Figure 10 shows the results for a white-noise environment. The upper dotted line indicates the theoretical limit of recovery using DFT. This corresponds to the case in which the extracted spectrum could be converted back to a waveform with the original amplitude. As shown in Figure 10, NHA performed much better in white-noise environments. Because of the finite frequency resolution, recovery of a single spectrum using DFT was limited, particularly in a low-noise environment. Recovery using NHA

yielded results well above the theoretical limit of DFT and showed a linear improvement even in a low-noise environment, thus confirming the importance of improved frequency resolution.

4. Suppression of side-lobes

In this section, the ability of NHA to suppress side-lobes is discussed. A frequency analysis was performed on a waveform composed of four sinusoidal waves (see Table 1). Figure 11 shows the resulting waveform, and Figure 12 shows the frequency spectra of this waveform as determined by DFT (zero-padding indicates interpolation of the DFT) and NHA. In the case of DFT, side-lobes exist around the main-lobe because of the limited frequency resolution. In the case of NHA, a line spectrum that is similar to that of the original waveform is obtained, and no side-lobes are produced. Even spectral components that are weaker than the DFT side-lobes can be extracted, as shown in Figure 12c.

In a case such as that shown in Figure 13, in which the source spectrum is mixed with a noise spectrum, side-lobe suppression can lead to greater noise reduction. The black line indicates the signal source spectrum, and the gray line represents the noise signal spectrum.

Figure 13a shows the case for DFT. The side-lobes of the source spectrum overlap the noise spectrum, making it difficult to estimate the amplitude. In addition, the phase information of the target signal is lost. If the side-lobes are removed, then the signal source cannot fully be recovered. On the other hand, the possibility of any overlap between

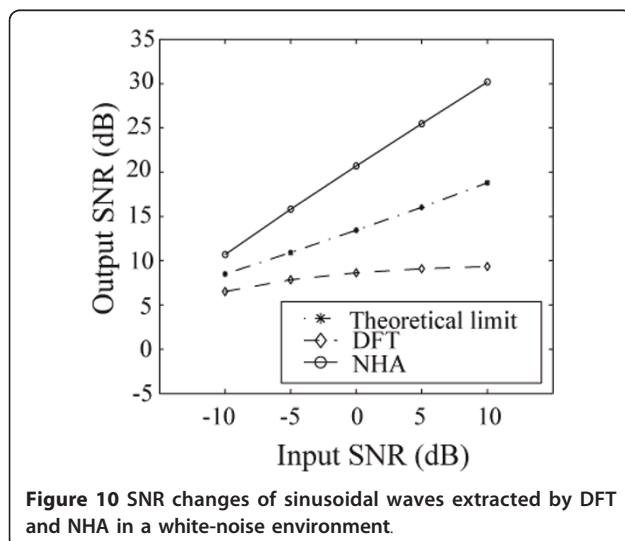


Table 1 Parameters of sinusoidal waves

Mark	Sinusoidal waves	
	Amplitude	Target frequency (Hz)
(a)	0.8	4.2
(b)	1	10.3
(c)	0.1	13.7
(d)	0.6	20.3

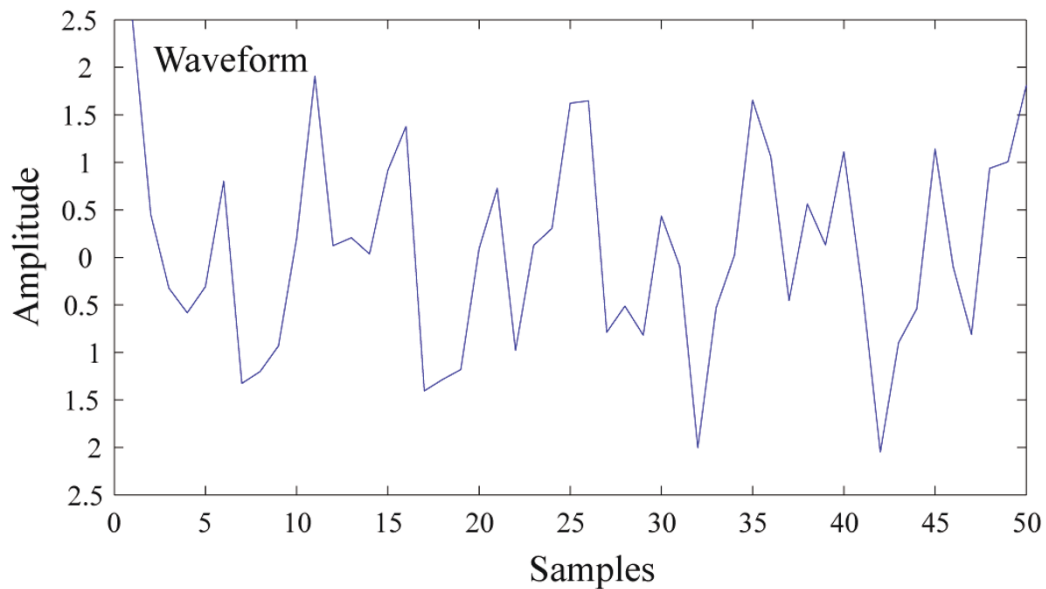


Figure 11 Composite wave synthesized by four sinusoidal waves.

the source and noise spectrum decreases because NHA is a high-frequency resolution analysis, as shown in Figure 13b. Therefore, there is a high possibility that the information contained in the source spectrum is isolated from the noise spectrum and can be recovered.

By DFT and NHA, we performed a frequency analysis on the part of the sound for which the input SNR of the white noise is 0 dB. Figure 14a is the original voice signal, and Figure 14b is the voice signal to which a noise was added. We removed noise by the SS method using DFT

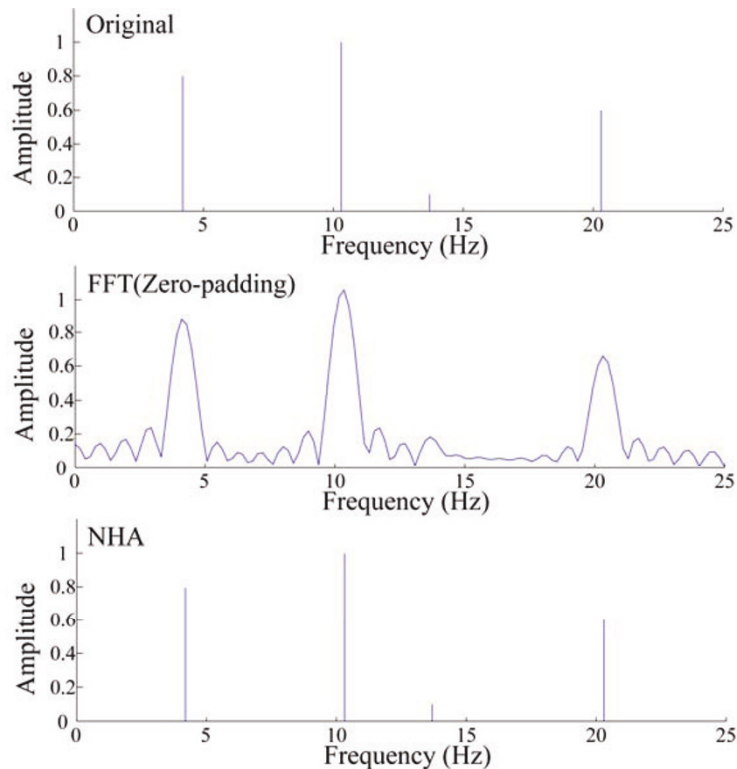


Figure 12 Frequency characteristics of four sinusoidal waves.

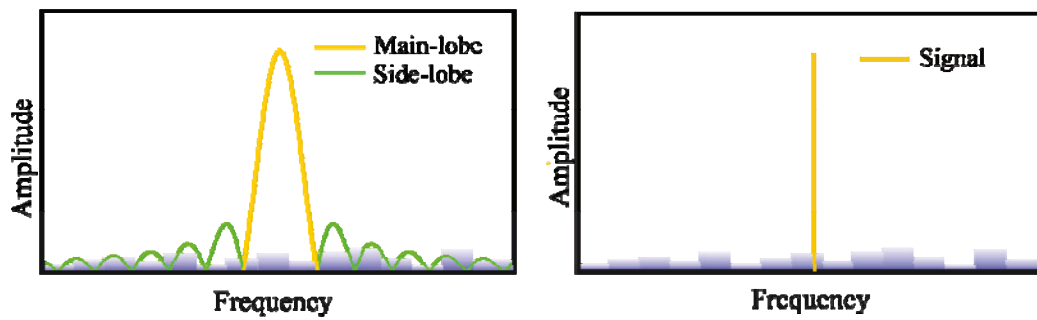


Figure 13 Spectrographs for a noise signal and a signal source. (a) low resolution, (b) high resolution.

and NHA; the results of which are described in Figure 14c, d, respectively. Figure 14e shows the variation of the output SNR by changing the threshold of the SS method. This figure shows that the maxima of output SNR using DFT and NHA are 9.1 and 17.4 dB, respectively. Therefore, the proposed technique using NHA is more useful in the noise reduction than that using DFT. In addition, it is important to appropriately determine the threshold for each noise because, as shown in Figure 14e, the output SNR changes significantly near the threshold to distinguish between signal and noise. One part of the output SNR using NHA is a straight line because small side lobes appear from the signal. However, NHA does not reveal the spectrum components of a sound in the side lobes. DFT is inferior to NHA because, in DFT, noise is mixed with the sound in the side lobes. Therefore, in NHA, the

threshold can be increased and the numerous noises can be suppressed, thereby improving the output SNR.

5. Constant threshold experiment

5.1. Experimental conditions for the constant threshold experiments

In order to investigate the relationship between the frequency resolution obtained by DTF, NHA, and the Ismo method [21,22], and the noise compression obtained by the SS method, we evaluate the results obtained by the segmental SNR method. In general, in the SS method, musical noises occur and affect the subjective evaluation. Although the spectral floor [23] has been proposed to eliminate these noises, in order to determine only the improvement in the results, we do not use this method in this study. In DFT, NHA, and the Ismo method,

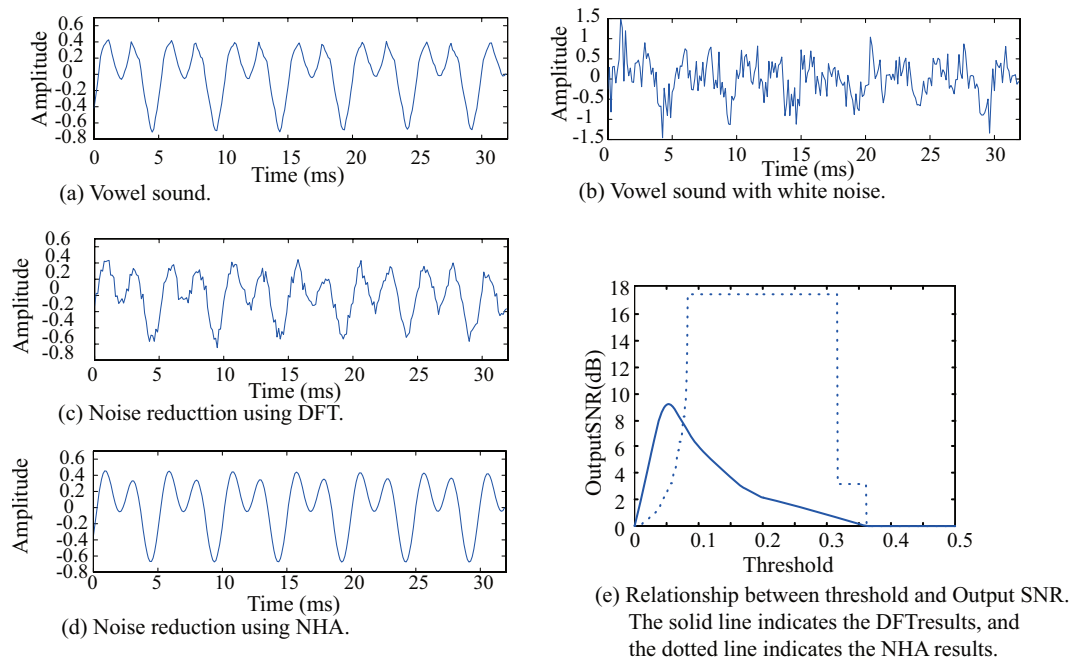


Figure 14 Noise reduction of the vowel sound.

various window functions were chosen. In DFT and the Ismo method, a Hanning and a rectangular window functions were used. In NHA, only a rectangular window was used. In a previous study [11], the Ismo method applied a Hanning window at points at which the signal changed suddenly, and a rectangular window was applied at the other points. In this article, to consider frequency resolution, we use a Hanning window and a rectangular window separately in different experiments. The signal sources are musical sounds in the form of midi data (Do-Re-Mi, Für Elise) that are played by a YAMAHA XG WDM SoftSynthesizer for 2 s. Based on the findings of a previous study [11], the order of the filter used for the prediction of the Ismo method is less than one frame length, and the half-frame-length sections before and after the signal frame are extrapolated. Here, the frequency resolution of the Ismo method is theoretically twice that of DFT. In most cases considered herein, NHA is used to extract 512 spectra per frame. In addition, after subtracting the signals of 3/4-frame-length sections before and after the signal frame, we evaluate the result of the NHA to consider the overlap of the signal frames. We then determined whether the same tendency was observed for each method, for four window lengths of 256, 512, 1024, and 2048. Table 2 lists the experimental conditions.

5.2 Details of the methods used to obtain the amplitude-modified spectra

First, spectrum (A_k , f_k , ϕ_k), $X(k)$, and $X_{\text{ISM}}(k)$ are calculated by NHA, DFT, and the Ismo method, respectively. The previously estimated noise spectrum is then subtracted from the calculated spectrum. Output signal \hat{s}_{DFTsub} obtained by DFT using the SS method is as follows:

$$\hat{s}_{\text{DFTsub}}(n) = \Re \left[\text{IFFT} \left\{ |\hat{X}(k)| \exp(j\angle X(k)) \right\} \right] \quad k = 0, 1, 2, \dots, N-1 \quad (18)$$

$$|\hat{X}(k)| = \begin{cases} |X(k)| - \alpha |\hat{D}(k)| & \text{if } (|X(k)| - \alpha |\hat{D}(k)|) > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $|X(k)|$, k , and α denote the spectral amplitude, the spectral number, and the most suitable threshold of the input signal, respectively. In general, the SS method used in noise compression yields the most suitable output by adjusting the noise spectrum model by means of a subtraction factor [23]. However, we calculate the segmental SNR using a few suitable threshold values for each analysis method because it is predicted that the most suitable values of the variable used in noise compression differ depending of the analysis method. The obtained results confirm that the most suitable threshold values do differ depending on the analysis method. Consequently, we calculated the suitable values for each signal waveform and compared the analysis methods with the most suitable segmental SNR. For the case of white Gaussian noise, we use $|\hat{D}(k)|$ that is constant for k , because the power spectrum density is uniform in any frequency band. We select the most suitable value of α so that the segmental SNR becomes maximum by gradually increasing the segmental SNR from a small value and use the selected value of α in the experiments. For the case of pink noise, we use the noise model $|\hat{D}(k)|$ that varies linearly along frequency axis and select the most suitable value of α using the above-mentioned method. In this study, we also remove the noise by the spectrum extraction (SE) method based on the concept of high frequency resolution preventing spectrum mixture. In the SE method, the output signal of DFT \hat{s}_{DFTex} is given as

$$\hat{s}_{\text{DFTex}}(n) = \Re \left[\text{IFFT} \left\{ |\hat{X}(k)| \exp(j\angle X(k)) \right\} \right] \quad k = 0, 1, 2, \dots, N-1 \quad (19)$$

$$|\hat{X}(k)| = \begin{cases} |X(k)| & \text{if } (|X(k)| - \alpha |\hat{D}(k)|) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Substituting $X_{\text{ISM}}(k)$ obtained using the Ismo method for $X(k)$ in Equations 18 and 19, we calculate these equations in a similar manner and obtain the output \hat{s}_{ISMsub} by the SS method, and the output \hat{s}_{ISMex} by the SE method.

Table 2 Experimental conditions

Analysis method	DFT (rectangular), DFT (Hanning), Ismo (rectangular), Ismo(Hanning), NHA
Amplitude modification	Spectral extraction, SS
Sampling frequency	44.1 KHz
Length of Music	2 s
Frame length	256, 512, 1024, 2048
Shift length	(Frame length)/4
Added noise	White Gaussian noise, Pink noise
Input SNR (dB)	-10, -5, 0, 5, 10
Instrument of MIDI	Flute, Grand piano, Reed organ, Overdrive guitar, Trumpet
Music (midi)	Do-Re-Mi, For Elise
Software synthesizer	YAMAHA XG WDM SoftSynthesizer

As mentioned earlier, we investigated both $X(k)$ and $X_{ISM}(k)$ using a Hanning window and a rectangular window (α is optimally selected for each window function). The output signal \hat{s}_{NHAsub} of NHA obtained by the SS method is given by the following equation:

$$\hat{s}_{NHAsub}(n) = \sum_{k=0}^K \tilde{A}_k \cos\left(2\pi \frac{\hat{f}_k}{f_s} n + \hat{\phi}_k\right) \quad (20)$$

$$\tilde{A}_k = \begin{cases} (\hat{A}_k - 2\alpha|\hat{D}(f_k)|) & \text{if } (\hat{A}_k - 2\alpha|\hat{D}(f_k)|) > 0 \\ 0 & \text{otherwise} \end{cases}$$

and (A_k, f_k, ϕ_k) is the spectrum component obtained from the noise signal obtained by NHA. Here, α is doubled in order to be equal to $|X(k)|$. Similarly, the output signal \hat{s}_{NHAex} of NHA obtained by the SE method is as follows:

$$\hat{s}_{NHAex}(n) = \sum_{k=0}^K \tilde{A}_k \cos\left(2\pi \frac{\hat{f}_k}{f_s} n + \hat{\phi}_k\right), \quad \tilde{A}_k = \begin{cases} \hat{A}_k & \text{if } (\hat{A}_k - 2\alpha|\hat{D}(f_k)|) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

5.3. Results of the fixed-threshold experiment

The variation with respect to time of the output SNR for input signals in which white Gaussian noise is added to a grand piano sound source is shown in Figures 15, 16, and 17. In these figures, (a), (b), and (c) show the output SNRs obtained by the SE method, the SS method, and the time-waveform, respectively, for the original signal. The window length is 2048.

Compared to the SE method, the NHA, indicated by blue solid lines, provided the best results, followed by the Ismo method with a Hanning window, and DTF with a rectangular window provided the worst results. Similarly, compared to the SS method, NHA provided the best results, and DTF with a rectangular window provided the worst results.

For this sound source, the output SNR calculated by each method has a different magnitude, but these magnitudes change at approximately the same time and exhibit a similar trend.

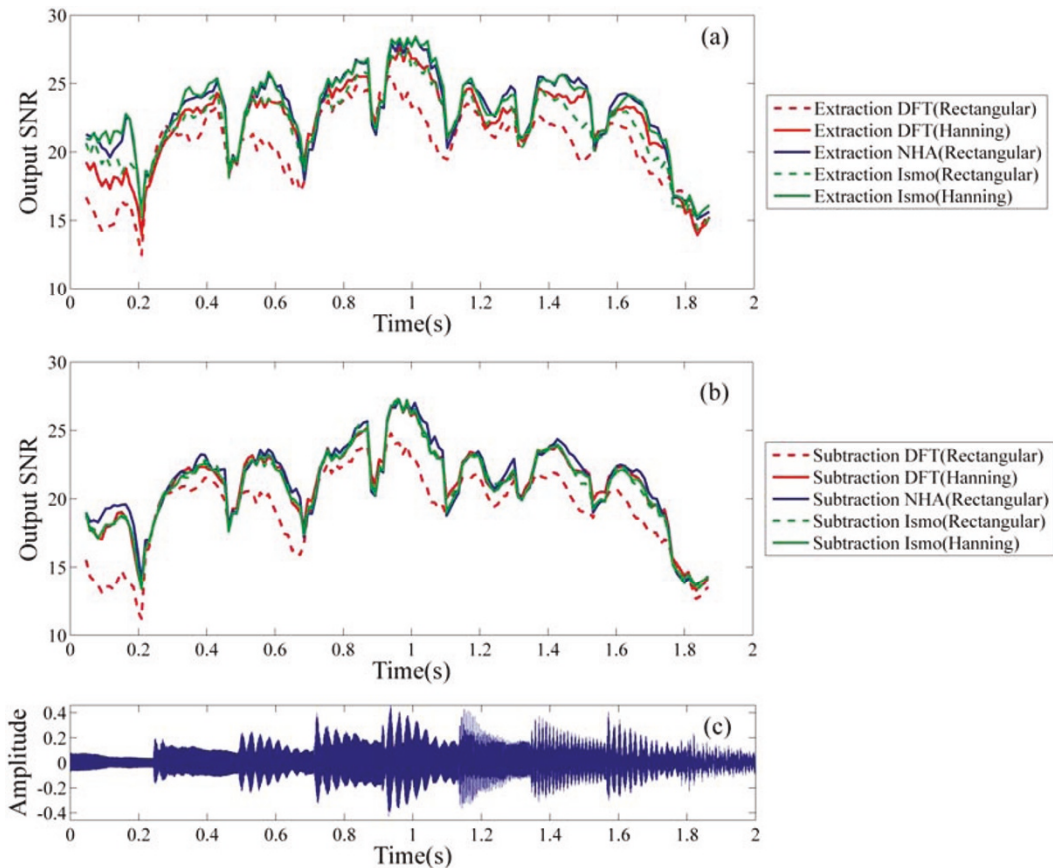


Figure 15 Change with respect to time in the output SNR of the signal source of a grand piano in a white Gaussian noise environment for which the input SNR is 0 dB. (a) SE method, (b) SS method, (c) signal source.

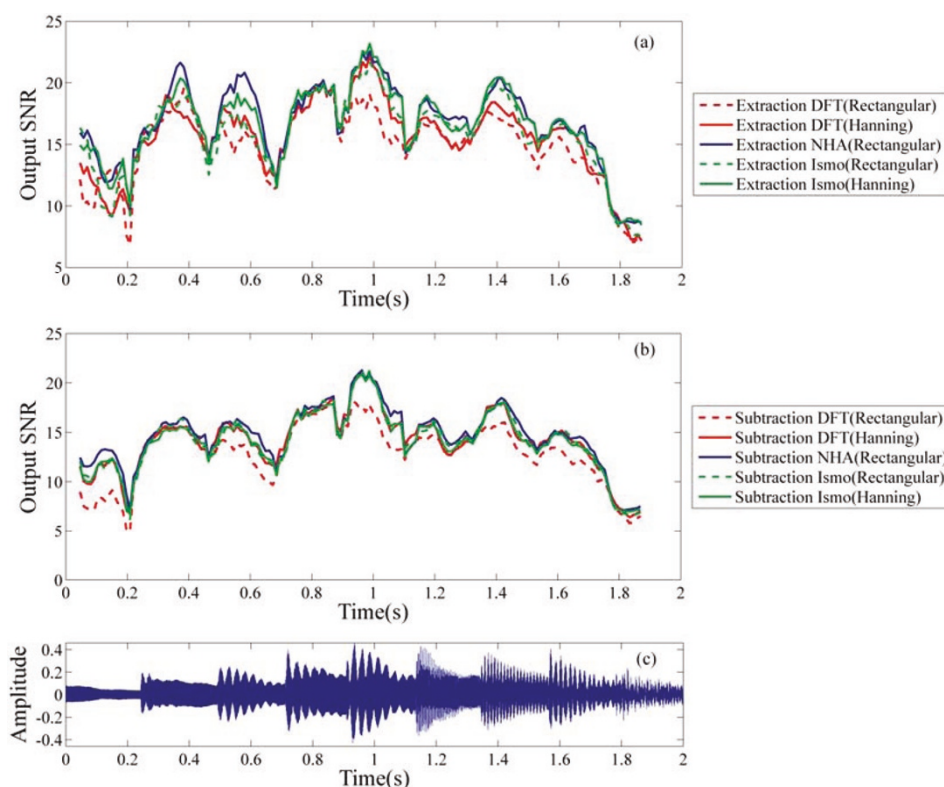


Figure 16 Change with respect to time in the output SNR of the signal source of a grand piano in a white Gaussian noise environment for which the input SNR is 10 dB. (a) SE method, (b) SS method, (c) signal source.

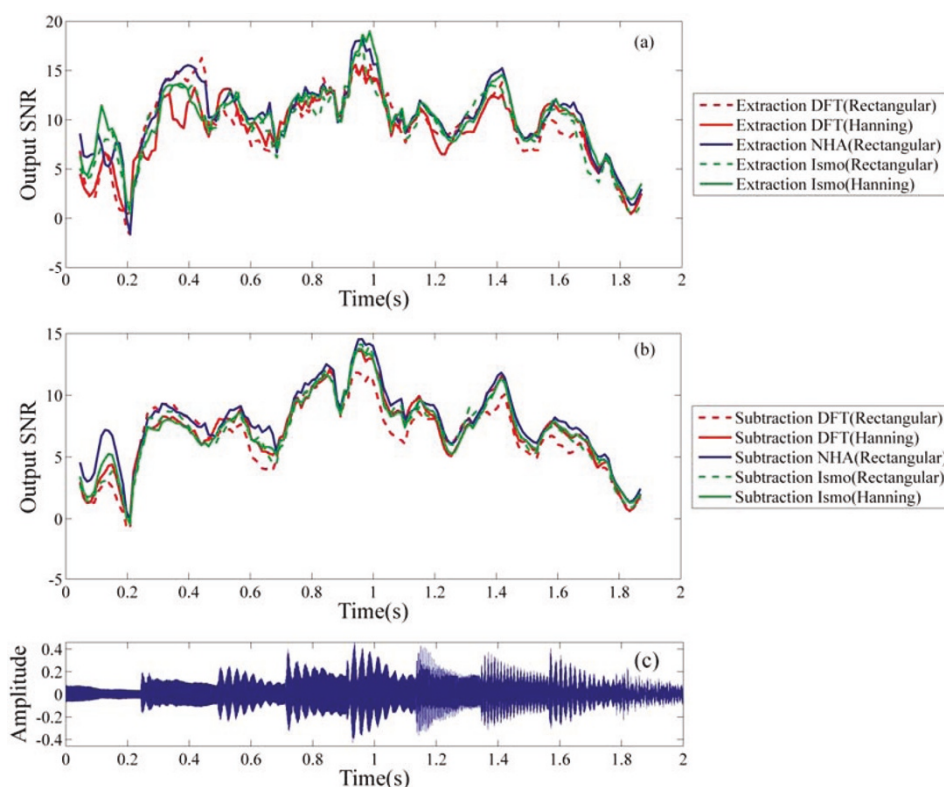


Figure 17 Change with respect to time in the output SNR of the signal source of a grand piano in a white Gaussian noise environment for which the input SNR is -10 dB. (a) SE method, (b) SS method, (c) signal source.

The results obtained for all of the analysis methods were poor during the periods of sudden changes in amplitude. In regions of stable amplitude, the high frequency resolution analysis methods that use a Hanning window function provided good results. Examples of signals for which a stable envelope was maintained are shown in Figures 18, 19, and 20.

The signal used here is stable and exhibits only a few changes in its envelope for both the SE and SS methods, as shown in Figures 18, 19, and 20. The calculated results for that signal were ranked in order of NHA, the Ismo method, and DFT. For the SE method, the Ismo method and NHA provided better results than DFT by approximately 5 and 3 dB, respectively, when the envelope changed markedly. For the SS method, the Ismo method and NHA provided better results than DFT by approximately 1.5 and 0.7 dB, respectively, when the envelope changed markedly. The results obtained by NHA may have been superior because the signal source spectrum was not dispersed and the frequency resolution was high. In addition, the results of the Ismo method are comparatively

good, in part because the prediction of the signal became easy.

Figure 21 shows the average segmental SNR for the music signal as obtained by ten noise reduction methods, which are the combinations of two noise subtraction methods and five frequency analysis methods in an analysis frame. Similar magnitude correlations appeared among the methods, even when the window length changed in Figure 21a-f. Similar results are observed for SNRs of 10, 0, and -10 dB.

Figure 21a-c shows the results for input SNRs of 10, 0, and -10 dB, respectively, in a white Gaussian noise environment. Based on the results, the average segmental SNR obtained by NHA is the highest for the SE method, followed by the Ismo method using a Hanning window. For the SS method, the average segmental SNR obtained by NHA is high compared to other techniques. Unlike in a previous study [11], the improvement in precision by the Ismo method for the SS method could not be confirmed in the present experiment. However, the higher values are thought to have been obtained using transient

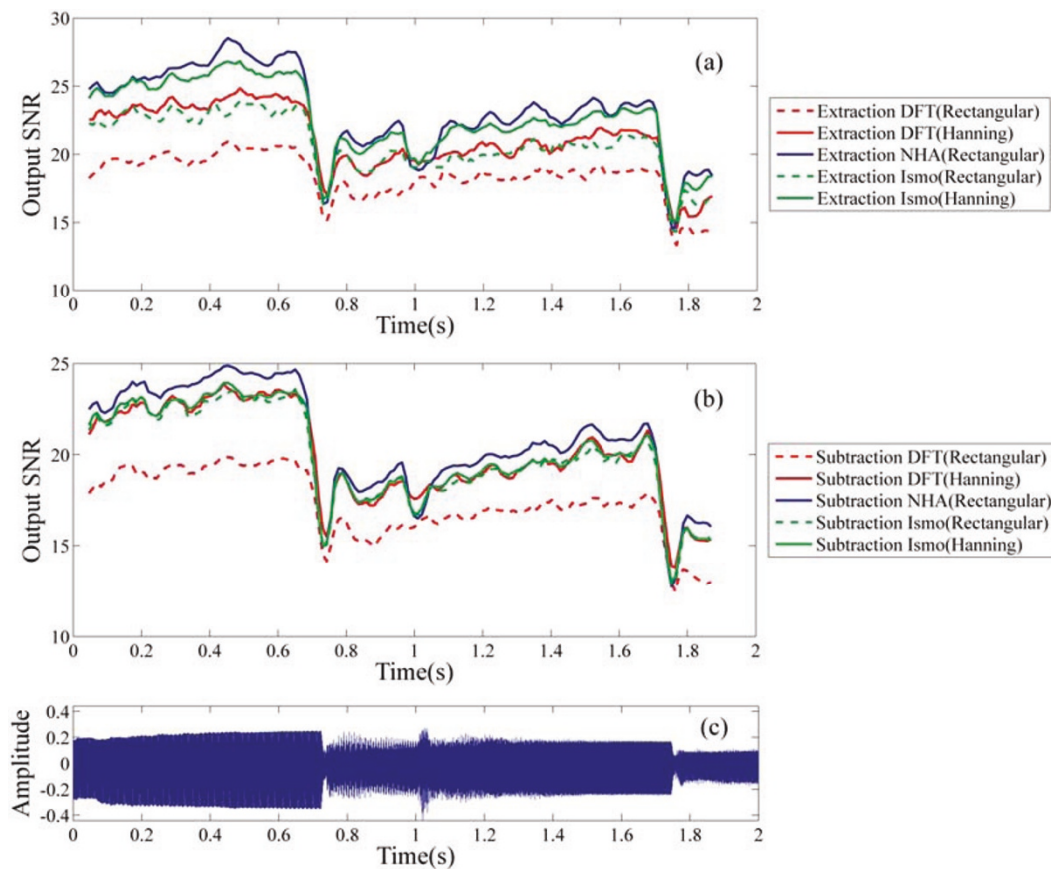


Figure 18 Change with respect to time in the output SNR of the signal source of a reed organ in a white Gaussian noise environment for which the input SNR is 10 dB. (a) SE method, (b) SS method, (c) signal source.

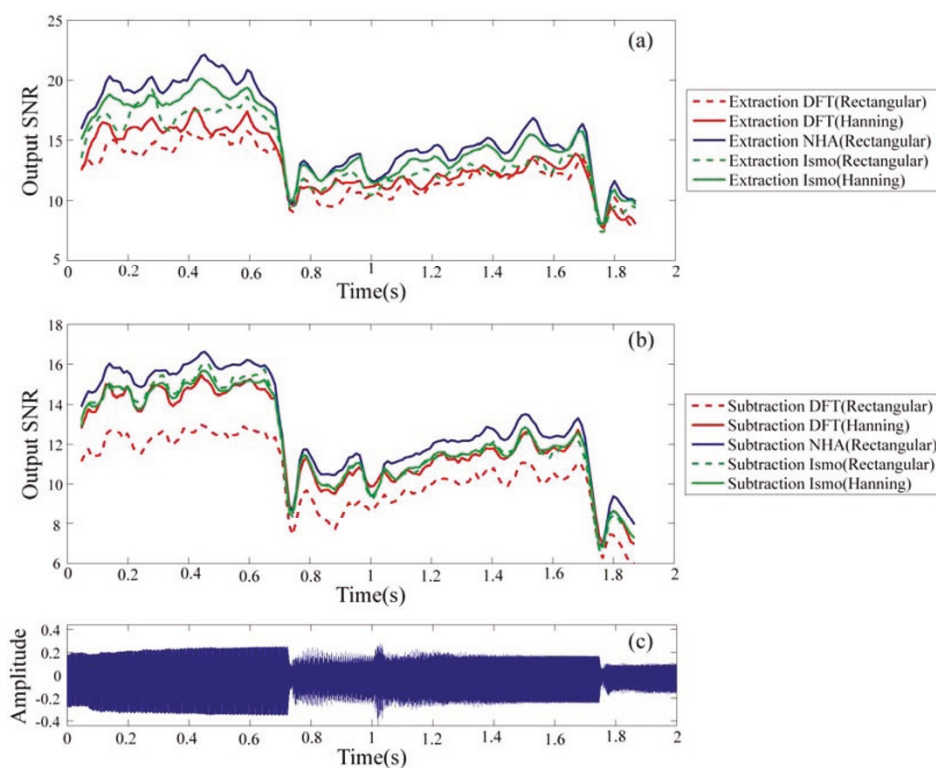


Figure 19 Change with respect to time in the output SNR of the signal source of a reed organ in a white Gaussian noise environment for which the input SNR is 0 dB. (a) SE method, (b) SS method, (c) signal source.

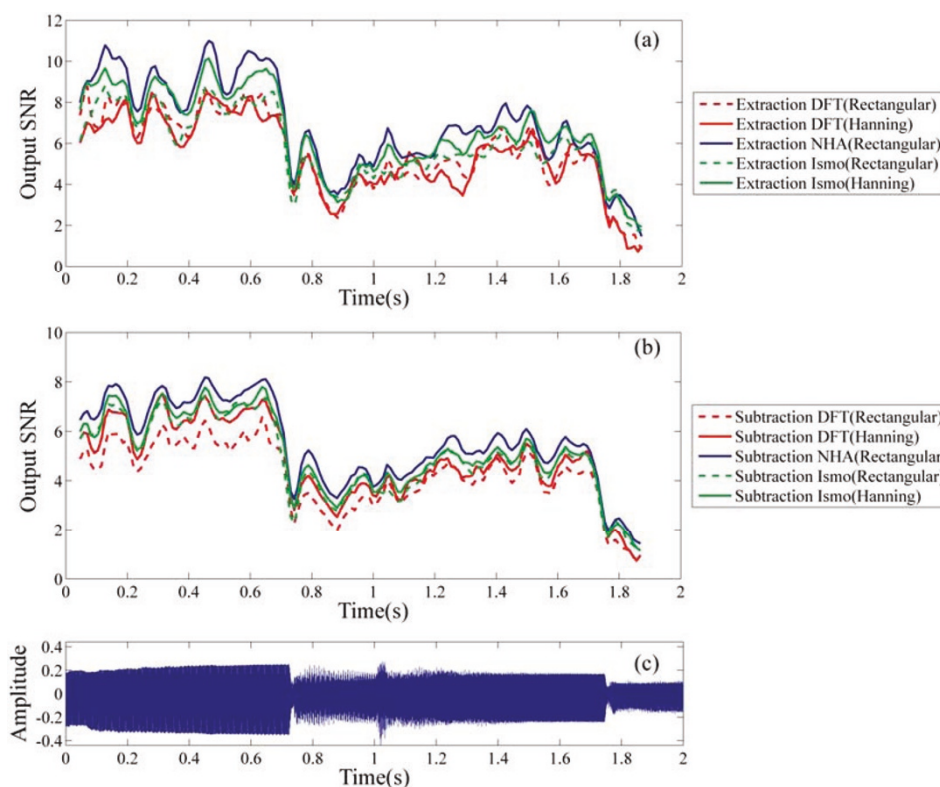


Figure 20 Change with respect to time in the output SNR of the signal source of a reed organ in a white Gaussian noise environment for which the input SNR is -10 dB. (a) SE method, (b) SS method, (c) signal source.

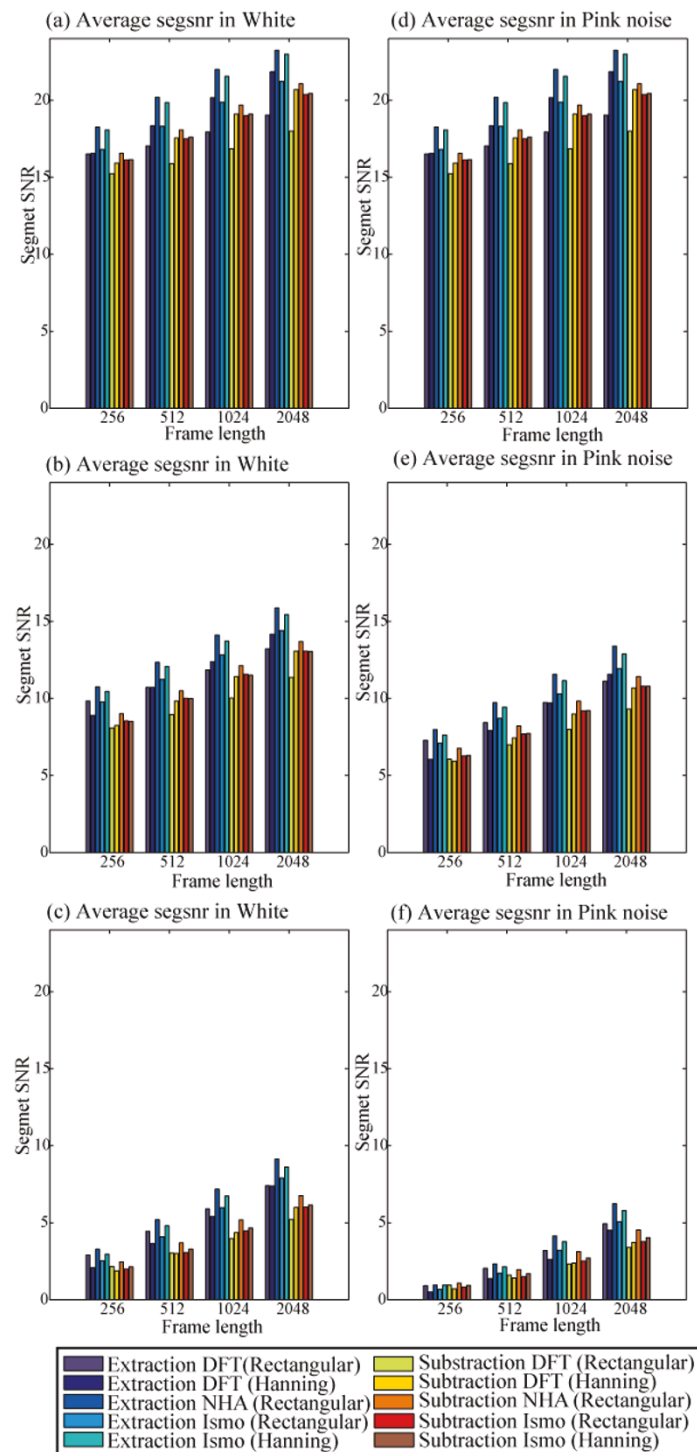


Figure 21 Average segmental SNR of a white Gaussian noise and a pink noise environment.

detection [21]. In this study, the threshold is chosen so that the segmental SNR is maximized each time the segmental SNR is calculated. The Ismo method is thought to be well suited to real applications (e.g., threshold

decision method that considers either human hearing [8] or musical noise [23]) and provides good affinity. Figure 21d-f shows the results for input SNRs of 10, 0, and -10 dB, respectively, in a pink noise environment. In this

case, the best NHA results were obtained using either the SE method or the SS method. Moreover, the combination of the Ismo method and a Hanning window provide good results compared to DFT by the SE method.

6. Summary

Previous studies have confirmed that the precision of the noise suppression is improved by increased frequency resolution for quality enhancement of sound to a previously existing recording. In this study, we demonstrate that NHA provides high frequency resolution by suppressing the influence of the window length. The limit to the precision improvement of noise suppression by NHA is examined. Since a frequency spectrum using NHA is not affected by the window length at the time of frequency conversion, the frequency resolution width is regarded as theoretically infinitesimal.

We added white Gaussian noise and pink noise to a music signal and performed experiments to examine the effects of noise suppression by the basic SS method. Segmental SNR was used to evaluate the effectiveness of noise suppression through a fixed-threshold experiment, and NHA and the conventional SS method were compared. The precision of the noise suppression obtained by NHA was confirmed to be better than that obtained by the conventional method. A similar magnitude correlation was confirmed to appear among the methods even if the window length changed. In addition, the improvement in precision of noise suppression by high frequency resolution was confirmed when the envelope was stable. Based on these results, an improvement in noise suppression precision, as compared to that provided by the conventional method, can be expected in various applications by incorporating NHA with a theoretically infinitesimal frequency resolution.

In this study, we attempt only to re-master the old music sources. Therefore, the main noise sources are usually generated by the old recording device and the deterioration of the recording media as pulsive noise and white noise. We do not assume noise encountered in a noisy environment, such as a subway or a roadside.

It may be feasible to apply the proposed technique to sound sources of daily conversations. It appears that we can recover enough even if a noise is mixed because the vowel sound is a periodic signal over a short time period. However, in the frequency analysis of the consonant, the calculation using NHA is approximately equivalent to the calculation using FFT.

In addition, we examined a pink noise as a representative colored noise. Other steady noises can be reduced in the same manner if the outline of the power spectrum is known. However, it appears that we must incorporate new methods other than the proposed method, and the new methods must be dynamically devised

because the characteristic of an unsteady noise must be predicted.

At this stage, we have not incorporated the proposed method into the embedded system or the portable device because the proposed method is several times longer than the calculation time of DFT (equivalent to the fastest FFT using a radix-2 number in this article). The high-speed SS method appears to be advantageous if the application is for the research of the speech recognition in the daily conversations. Although the calculation time is increased, the proposed technique will be effective if used in an application that requires high precision. We believe that the defects of the proposed method are best left for consideration in a future study if the proposed method is applied to a portable product or the research of speech recognition.

Acknowledgements

This work was supported by Grants-in-Aid for Challenging Exploratory Research, MEXT(No.23650110).

Competing interests

The authors declare that they have no competing interests.

Received: 27 June 2011 Accepted: 21 September 2011

Published: 21 September 2011

References

1. SF Boll, Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans Acoust Speech, Signal Process ASSP.* **27**(2), 113–120 (1979). doi:10.1109/TASSP.1979.1163209
2. CT Lin, Single-channel speech enhancement in variable noise-level environment. *IEEE Trans Syst Man Cybernet A.* **33**(1), 137–143 (2003)
3. SD Kamath, PC Loizou, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, in *Proceedings of the ICASSP*, pp. 4164–4167 (2002)
4. Z Goh, KC Tan, BTG Tan, Postprocessing method for suppressing musical noise generated by spectral subtraction. *IEEE Trans Speech Audio Process.* **6**, 287–292 (1998). doi:10.1109/89.668822
5. K Sorensen, S Andersen, Speech enhancement with natural sounding residual noise based on connected time-frequency speech presence regions. *EURASIP J Appl Signal Process.* **18**, 2954–2964 (2005)
6. IY Soon, SN Koh, Speech enhancement using 2-D Fourier transform. *IEEE Trans Speech Audio Process.* **11**, 717–724 (2003). doi:10.1109/TSA.2003.816063
7. H Ding, IY Soon, SN Koh, CK Yeo, A spectral filtering method based on hybrid wiener filters for speech enhancement. *Speech Commun.* **51**, 259–267 (2009). doi:10.1016/j.specom.2008.09.003
8. N Virag, Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Trans Speech Audio Process.* **7**(2), 126–137 (1999). doi:10.1109/89.748118
9. R Udrea, N Vizireanu, S Ciochina, An improved spectral subtraction method for speech enhancement using a perceptual weighting filter. *Digital Signal Process.* **18**(4), 581–587 (2008). doi:10.1016/j.dsp.2007.08.002
10. I Kauppinen, K Roth, Improved noise reduction in audio signals using spectral resolution enhancement with time-domain signal extrapolation. *IEEE Trans Speech Audio Process.* **13**, 1210–1216 (2005)
11. S Hirobayashi, F Ito, T Yoshizawa, T Yamabuchi, Estimation of the frequency of non-stationary signals by the steepest descent method, in *Proceedings of the Fourth Asia-Pacific Conference of Industrial Engineering and Management Systems*, pp. 788–791 (2002)
12. EB George, MUT Smith, Analysis-by-synthesis/overlap add sinusoidal modeling applied to the analysis and synthesis of musical tones. *J Audio Eng Soc.* **125**(40), 497–516 (1992)

13. EB George, MJT Smith, Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. *IEEE Trans Speech Audio Process.* **5**(5), 398–406 (1997)
14. JW Turkey, AE Beaton, The fitting of power series, meaning polynomials, illustrated on band-spectroscopic-data. *Technometrics.* **16**, 189–192 (1974). doi:10.2307/1267938
15. JM Chambers, *Computational Methods for Data Analysis*. Wiley, New York (1977)
16. PE Gill, W Murray, Quasi-Newton methods for unconstrained optimization. *J Inst Math Appl.* **9**, 91–108 (1972). doi:10.1093/imamat/9.1.91
17. T Terada, et al, Non-stationary waveform analysis and synthesis using generalized harmonic analysis, in *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 429–432 (1994)
18. N Wiener, in *The Fourier Integral and Certain of Its Applications*, (Dover Publications, Inc., New York, 1958), pp. 158–199
19. T Muraoka, S Kiri, Y Kamiya, Fast algorithm for generalized harmonic analysis (GHA), in *The 47th IEEE International Midwest Symposium on Circuit and Systems*, pp. 153–156 (2004)
20. Y Hirata, Non-harmonic Fourier analysis available for detecting very low-frequency components. *J Sound Vib.* **287**(3), 611–613 (2005)
21. I Kauppinen, K Roth, An adaptive technique for modeling audio signals, in *Proceedings of the 4th International Conference on Digital Audio Effects (DAFx-01)*, (Limerick, Ireland, 2001), pp. 1–4
22. I Kauppinen, K Roth, Audio signal extrapolation—theory and applications, in *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, (Hamburg, Germany, 2002), pp. 105–110
23. M Berouti, R Schwartz, J Makhoul, Enhancement of speech corrupted by acoustic noise, in *Proc IEEE ICASSP'79*, pp. 208–211 (April 1979)

doi:10.1186/1687-4722-2011-426794

Cite this article as: Yoshizawa et al.: Noise reduction for periodic signals using high-resolution frequency analysis. *EURASIP Journal on Audio, Speech, and Music Processing* 2011 **2011**:5.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com