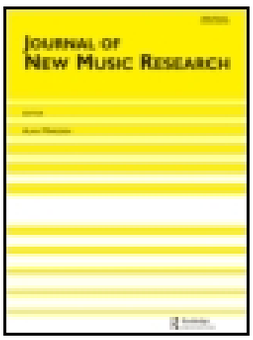

Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: http://www.tandfonline.com/loi/nnmr20

High-Accuracy Frequency Analysis of Harmonic Signals Using Improved Phase Difference Estimation and Window Switching

Hee-Suk Pang, Jun-seok Lim & Soonil Kwon

To cite this article: Hee-Suk Pang, Jun-seok Lim & Soonil Kwon (2017): High-Accuracy Frequency Analysis of Harmonic Signals Using Improved Phase Difference Estimation and Window Switching, Journal of New Music Research, DOI: 10.1080/09298215.2017.1356855

To link to this article: http://dx.doi.org/10.1080/09298215.2017.1356855

Journal of New Music Research, 2017
https://doi.org/10.1080/09298215.2017.1356855

Routledge
Taylor & Francis Group

High-Accuracy Frequency Analysis of Harmonic Signals Using Improved Phase Difference Estimation and Window Switching

Hee-Suk Pang[1], Jun-seok Lim[1] and Soonil Kwon[2]

[1]Department of Electrical Engineering, Sejong University, Seoul, Korea; [2]Department of Digital Contents, Sejong University, Seoul, Korea

(*Received 7 November 2016; accepted 10 July 2017*)

Abstract

Accurate frequency tracking of harmonic signals is an essential process for analysis of musical instrument and singing sounds. For fine frequency estimation, we propose two methods for enhancement of the phase difference estimation (PDE). The first method is improved PDE that maximises the signal level at the frequency estimation position, which is effective especially at low signal-to-noise ratio (SNR). The second method is window switching that minimises the effects of harmonics by suppressing their sidelobe levels, which is effective at high SNR. Window switching is also applicable to existing fine frequency estimation methods as well as the PDE and improved PDE. Experimental results and application examples show that the proposed methods show meaningful improvements over the PDE and are effective for high-accuracy frequency analysis. The proposed methods can be adopted as a module for sound analysis and processing tools and software, which are expected to contribute to the comprehensive analysis of musical signals.

Keywords: frequency tracking, fine frequency estimation, software, audio analysis, music analysis

1. Introduction

Accurate frequency tracking of harmonic signals is an essential process for analysis of musical instrument and singing sounds. Frequency tracking is usually composed of two steps: coarse frequency estimation and fine frequency estimation. Discrete Fourier transform (DFT)-based frequency estimation methods have often been used for that purpose, where the coarse frequency estimation of each harmonic is performed basically by finding the harmonics' spectral peaks in consideration of the harmonic structure. Since the DFT index at which the spectral peak occurs does not coincide with the true frequency, fine frequency estimation is applied to improve the accuracy of the frequency estimate. For example, zero-padding is used with the DFT for automatic music transcription (Cañadas Quesada, Ruiz Reyes, Vera Candeas, Carabias, & Maldonado, 2010; Ryynänen & Klapuri, 2008), where the frequency estimates should be accurate enough for note identification.

In many cases, the accuracy of frequency estimates at the note level is insufficient. Accurate pitch tracking is necessary for studies on traditional music, e.g. pitch scale studies (Akkoç, 2002; Bozkurt, 2008; Bozkurt, Yarman, Karaosmanoğlu, & Akkoç, 2009), automatic tonic identification (Gulati et al., 2014), raga recognition (Koduri, Gulati, Rao, & Serra, 2012), intonation analysis (Koduri, Ishwar, Serrà, & Serra, 2014), classification of melodic motifs (Rao et al., 2014), linking scores and audio recordings (Şentürk, Holzapfel, & Serra, 2014), and automatic flamenco transcription (Gómez & Bonada, 2013). Accurate frequency estimation is also essential for analysis of musical instrument sounds, e.g. instantaneous frequency measurements for the magnetic resonator piano (McPherson, 2010), frequency estimation of the overtones of gongs (Rapoport, Shatz, & Blass, 2008), pitch tracking of guqin music (Li & Leman, 2007), and frequency extraction of sho sounds for synthesis (Hikichi, Osaka, & Itakura, 2004). Such accuracy is also

Correspondence: Hee-Suk Pang, Department of Electrical Engineering, Sejong University, Neungdong-ro 209, Gwangjin-gu, Seoul 05006, Korea. E-mail: hspang@sejong.ac.kr

necessary for inharmonicity estimation for stringed instrument sounds (Chin & Berger, 2010; Galembo & Askenfelt, 1999; Rauhala, Lehtonen, & Välimäki, 2007), tuning of musical instruments (Bozkurt, 2012), estimation of tuning and inharmonicity from a Western perspective (Dixon, Mauch, & Tidhar, 2012), tonality estimation of sounds (Kulesza & Czyzewski, 2009), and speech-music discrimination (Shirazi & Ghaemmaghami, 2010). Several fine frequency estimation methods were compared for harpsichord tuning classification (Tidhar, Mauch, & Dixon, 2010).

DFT phase-based methods use phase information in the DFT for frequency estimation of a sinusoid. For example, they include the reassignment method (Flandrin, Auger, & Chassande-Mottin, 2003), the derivative estimation (Desainte-Catherine & Marchand, 2000), the trigonometric estimation (Lagrange, Marchand, & Rault, 2005), the arctan estimation (Betser, Collen, Richard, & David, 2006), and the phase difference estimation (PDE) (Brown & Puckette, 1993; Puckette & Brown, 1998). The PDE calculates the frequency of a sinusoid using the phase difference between consecutive DFTs and shows good performance among the DFT phase-based methods (Lagrange & Marchand, 2007). Due to its good performance, the PDE has been frequently used for analysis of audio and music signals. For example, it has been used for sinusoidal modelling of musical signals (Lagrange, Martins, Murdoch, & Tzanetakis, 2008), analysis and detection of vibrato tones (Pang, 2004; Pang & Yoon, 2005), and processing of musical sounds for cochlear implants (Zakis, McDermott, & Vandali, 2007).

Though the PDE shows good performance for fine frequency estimation, it is still necessary to improve its accuracy. This is because the frequency estimates of harmonics are primary data for the comprehensive analysis of musical signals and should be as accurate as possible. The improvement of the PDE has not been studied much in the literature. One approach was iterative frequency estimation which repeats the phase calculation and frequency estimation based on the PDE (Pang, Lim, Kwon, & Shin, 2012). We generalise and expand this approach to propose improved PDE which maximises the signal level at the frequency estimation position. We also propose window switching which minimises the effects of harmonics by suppressing their sidelobe levels. Since fine frequency estimation methods are used as a module for sound analysis and processing tools and software (Nunes, Esquef, & Biscainho, 2009; Sethares, Milne, Tiedje, Prechtl, & Plamondon, 2009; Six, Cornelis, & Leman, 2013), the proposed methods can be adopted as the module and used for accurate frequency tracking of musical signals. The proposed methods are relevant for monophonic signals, but are also applicable to polyphonic signals whose harmonics do not overlap with each other.

This paper is organised as follows. Section 2 presents a brief review of the PDE. In Section 3, we analyse the characteristics of the frequency estimation errors of the PDE and propose improved PDE and window switching. In Section 4, we analyse the performance of the proposed methods with experimental results. In Section 5, we present application examples of the proposed methods to the frequency tracking of musical instrument sounds. Section 6 provides concluding remarks.

2. Review of the phase difference estimation

We model a harmonic signal as

$$x[n] = \sum_{m=1}^{M} a_m \sin\left(2\pi \frac{f_m}{f_s} n + \theta_m\right) + z[n], \tag{1}$$

where M is the number of harmonics, a_m, f_m, and θ_m are the amplitude, frequency, and phase of the mth harmonic, f_s is the sampling frequency, and $z[n]$ is zero-mean white noise. In addition, $f_m = m \times f_0$, where f_0 is the fundamental frequency. The DFT of the signal is calculated as

$$X[k,h] = \sum_{n=0}^{N-1} x[n+h]w[n]e^{-j\frac{2\pi k}{N}n}, \quad k = 0, 1, ..., N-1, \tag{2}$$

where k is the DFT index, h is the sample index corresponding to the time at which the DFT is computed, and $w[n]$ is an N-point window function. Since the signal is composed of M harmonics, $|X[k,h]|$ is composed of $2 \times M$ spectral peaks whose locations correspond to $\pm f_m$ for $m = 1, 2, \ldots, M$.

Suppose that we estimate f_p which is the frequency of the pth harmonic. After coarse frequency estimation, we select the DFT index which is closest to the pth harmonic as

$$k_p = \arg \max_k |X[k,0]|. \tag{3}$$

The search range of k_p can be coarsely predetermined in consideration of the harmonic structure of the signal, e.g. the subharmonic summation (Hermes, 1988). Once k_p is determined, the frequency in Hz is estimated as

$$\hat{f}_p = \frac{f_s}{2\pi H}\left\{\angle\left(X[k_p, H]\right) - \angle\left(X[k_p, 0]\right)\right\}, \tag{4}$$

where $\angle(y)$ is the phase of y and H is a hop size which is the sample delay between two consecutive DFTs. A hop size of 1 ($H = 1$) has been primarily used to minimise the overall sample size and to prevent the ambiguity due to the phase unwrapping problem (Betser et al., 2006; Brown & Puckette, 1993; Lagrange & Marchand, 2007; Lagrange et al., 2008; Pang & Yoon, 2005). Therefore, a hop size of 1 will be used in the remainder of this paper. The frequency estimate in Hz is represented as

$$\hat{f}_p = \begin{cases} \frac{f_s}{2\pi}\{\angle(X[k_p,1]) - \angle(X[k_p,0])\}, & \text{if } \angle(X[k_p,1]) \geq \angle(X[k_p,0]), \\ \frac{f_s}{2\pi}\{\angle(X[k_p,1]) - \angle(X[k_p,0]) + 2\pi\}, & \text{otherwise,} \end{cases} \tag{5}$$

in consideration of the phase unwrapping problem.

Another important consideration in the PDE is the choice of a window function. It is well known that there is a trade-off between the mainlobe width and sidelobe attenuation of a window function (Harris, 1978; Nuttall, 1981). That is, a window function whose mainlobe is narrow suffers from the high sidelobe level, and a window function whose sidelobe level is low suffers from the wide mainlobe. For the PDE, a Hanning window has been primarily used due to its moderate mainlobe width and good sidelobe attenuation (Betser et al., 2006; Brown & Puckette, 1993; Lagrange & Marchand, 2007; Pang, 2004; Pang & Yoon, 2005). Therefore, a Hanning window will be used as the default window for the PDE.

3. The proposed methods

We first analysed the characteristics of the frequency estimation errors in the PDE. Specifically, we divided the causes of the errors into two cases, which are noise-dominant and harmonics-dominant cases. We propose two methods to minimise the frequency estimation errors for each case.

3.1 Error analysis of the phase difference estimation

The discrete-time Fourier transform (DTFT) of $x[n]$ is calculated as

$$X(\omega) = \sum_{n=0}^{N-1} x[n]w[n]e^{-j\frac{2\pi\omega}{N}n}, \quad \omega \in [0, N), \tag{6}$$

where ω represents the normalised frequency expressed in bin. Note that ω is a real number. The DTFT of the harmonic signal in Section 2 is represented as

$$X(\omega) = \sum_{m=1}^{M} \frac{a_m}{2j}\left[W(\omega - \omega_m)e^{j\theta_m} - W(\omega + \omega_m)e^{-j\theta_m}\right] + Z(\omega), \tag{7}$$

where $W(\omega)$ is the DTFT of $w[n]$, $\omega_m = N \times f_m/f_s$, a_m and θ_m are the amplitude and phase of the mth harmonic, and $Z(\omega)$ is the DTFT of $z[n]\times w[n]$. Suppose that we estimate f_p, which is the frequency of the pth harmonic, and then we rewrite $X(\omega)$ as

$$X(\omega) = X_T(\omega) + X_I(\omega) + Z(\omega), \tag{8}$$

where

$$X_T(\omega) = \frac{a_p}{2j}W(\omega - \omega_p)e^{j\theta_p}, \tag{9}$$

$$X_I(\omega) = \sum_{m=1, m\neq p}^{M} \frac{a_m}{2j}W(\omega - \omega_m)e^{j\theta_m} - \sum_{m=1}^{M} \frac{a_m}{2j}W(\omega + \omega_m)e^{-j\theta_m}. \tag{10}$$

$X_T(\omega)$ is the target component whose frequency corresponds to f_p. $X_I(\omega)$ is composed of the harmonics except the target component, which we refer to as the interference component. If ω is close to ω_p, $X_I(\omega)$ is composed of the sidelobes of the harmonics. $Z(\omega)$ is the noise component. Since the PDE calculates the frequency using the phase difference between consecutive DFTs, its frequency estimation errors are caused mainly by the interference and noise components.

In addition, since the DFT is a sampled version of the DTFT, the frequency estimation is performed at $\omega = k_p$ in the PDE. Since the DFT index k_p does not coincide with the true frequency ω_p, it can be represented as

$$\omega_p = k_p + \delta_p, \tag{11}$$

where $-.5 \leq \delta_p < .5$. $|X_T(k_p)|$ oscillates from $|a_p/2 \times W(0)|$ to $|a_p/2 \times W(\pm 1/2)|$ as a function of δ_p and reaches its maximum $|a_p/2 \times W(0)|$ at $\delta_p = 0$ $(\omega_p = k_p)$ and its minimum $|a_p/2 \times W(\pm 1/2)|$ at $\delta_p = \pm 1/2$ $(\omega_p = k_p \pm 1/2)$. Examples of $|X_T(\omega)|$, $|X_I(\omega)|$, and $|Z(\omega)|$ are shown in Figure 1, where $|X_T(\omega_p)|$ and $|X_T(k_p)|$ are also indicated.

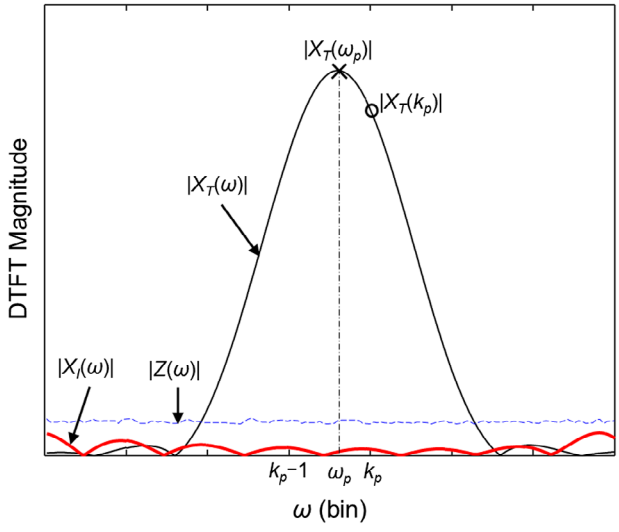

Fig. 1. Examples of $|X_T(\omega)|$, $|X_I(\omega)|$, and $|Z(\omega)|$.

For further analysis, we define the signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{a_p^2}{2\sigma^2}\right), \tag{12}$$

where a_p is the amplitude of the pth harmonic and σ is the standard deviation of the white noise. The SNR represents the power ratio of the target harmonic to the noise. We separate the causes of the frequency estimation errors into two cases in consideration of the SNR.

First, when the SNR is low, $|Z(\omega)| \gg |X_I(\omega)|$ and the frequency estimation errors are caused mainly by the noise component. Then the errors are inversely proportional to the ratio $|X_T(\omega)|/|Z(\omega)|$ at the frequency estimation position. That is, the errors are small when the ratio is high and are large when the ratio is low. For white noise, $|Z(\omega)|$ is supposed to be a constant, and its average magnitude can be calculated using the Parseval's theorem as

$$\bar{Z}^2 = E[|Z(\omega)|^2] = \sum_{n=0}^{N-1} z^2[n]w^2[n] = \sigma^2 \sum_{n=0}^{N-1} w^2[n], \tag{13}$$

which is determined by the window function. If $\omega_p = k_p$ ($\delta_p = 0$), $|X_T(k_p)|$ is maximal and thus the frequency estimation errors are minimal. Similarly, if $\omega_p = k_p \pm 1/2$ ($\delta_p = \pm 1/2$), $|X_T(k_p)|$ is minimal and thus the frequency estimation errors are maximal. Therefore, the frequency estimation errors of the PDE oscillate between their maxima and minima as the fundamental frequency increases.

Second, when the SNR is high, $|X_I(\omega)| \gg |Z(\omega)|$ and the frequency estimation errors are caused mainly by the interference component. Then the errors are inversely proportional to the ratio $|X_T(\omega)|/|X_I(\omega)|$ at the frequency estimation position. In consideration of the sidelobe characteristics of a Hanning window shown in Figure 2, the errors become large if the interfering harmonics are close to the target component and small if the interfering harmonics are distant from the target component. In our signal model, as the fundamental frequency increases, the interfering harmonics become more distant from the target component and the frequency estimation errors generally decrease.

To verify the error analysis of the PDE, we conducted a simulation using a signal composed of five harmonics of unity amplitudes in zero-mean white noise as

$$x[n] = \sum_{m=1}^{5} \sin\left(2\pi \frac{m \times f_0}{f_s} n + \theta_m\right) + z[n], \quad n = 0, 1, \ldots, N. \tag{14}$$

In the simulation, θ_m was a random variable in the range $[0, 2\pi)$, N was 1024, f_s was 48 kHz, and a Hanning window was applied in the computation of the DFT. In Figure 3, the root mean square errors (RMSEs) of $\hat{f}_0$ are depicted as a function of fundamental frequency f_0, where $\hat{f}_0$ was estimated for the fundamental using the PDE. The RMSE is defined as

$$\mathrm{RMSE} = \sqrt{E[(f - \hat{f})^2]}, \tag{15}$$

which represents the average error in the root mean square sense between the true frequency f and the frequency estimate $\hat{f}$. In the simulation, f_0 varied from 187.5 to 562.5 Hz, which corresponded to a range of 4–12 bin in the DFT. The minimum value of 4 bin was chosen to prevent the mainlobe of the fundamental from overlapping with that of the second harmonic. The results were average values over 10,000 independent trials for $\sigma = 1$ and $\sigma = .001$. In Figure 3, the results verify the error analysis of the PDE; the frequency estimation errors oscillate between their maxima and minima as f_0 increases at low SNR

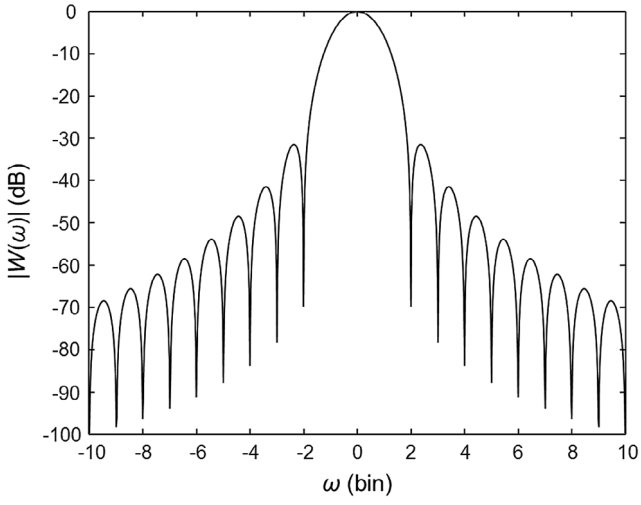

Fig. 2. DTFT magnitude of a Hanning window.

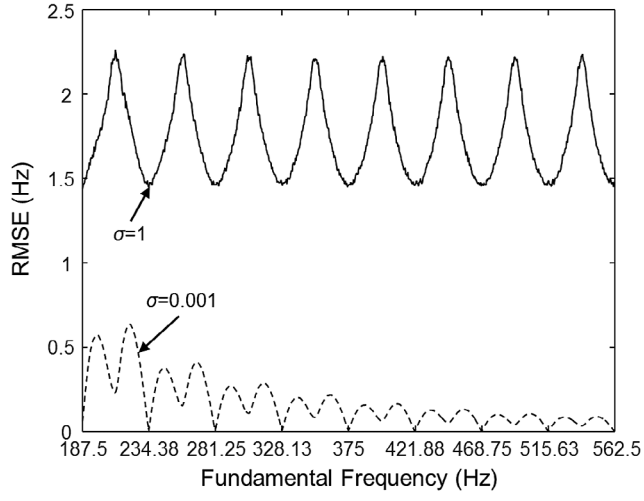

Fig. 3. Frequency estimation errors of the PDE as a function of fundamental frequency for the signal in Equation 14.

($\sigma = 1$) and decrease with fluctuations as f_0 increases at high SNR ($\sigma = .001$).

3.2 Improved phase difference estimation

When the SNR is low, the frequency estimation errors are caused mainly by the noise component. Accordingly, we use the signal-to-noise floor ratio (SNFR) defined as

$$\mathrm{SNFR} = 20\log_{10}\left(\frac{|X_T(\omega)|}{|Z(\omega)|}\right), \tag{16}$$

which represents the ratio of the target component to the noise component (Pang et al., 2012). The frequency estimation errors are minimised by maximising the SNFR at the frequency estimation position. For white noise, $|Z(\omega)|$ can be replaced by its average magnitude, and the SNFR is represented as

$$\begin{aligned}\mathrm{SNFR} &= 10\log_{10}\left(\frac{|X_T(\omega)|^2}{\bar{Z}^2}\right)\\ &= 10\log_{10}\left(\frac{|X_T(\omega)|^2}{\sigma^2\sum_{n=0}^{N-1}w^2[n]}\right).\end{aligned} \tag{17}$$

Then the frequency estimation errors are minimised by maximising $|X_T(\omega)|$ at the frequency estimation position. Since the frequency is estimated at $\omega = k_p$ in the PDE, $|X_T(k_p)|$ varies between $|a_p/2 \times W(0)|$ and $|a_p/2 \times W(\pm 1/2)|$. If we estimate the frequency at $\omega = \omega_p$ instead of at $\omega = k_p$, $|X_T(\omega_p)| = |a_p/2 \times W(0)|$ and the SNFR is maximised. The maximum of the SNFR was referred to as the peak signal-to-noise ratio (Zivanovic, Röbel, & Rodet, 2008), but we refer to it as the peak signal-to-noise floor ratio (PSNFR). An example of the PSNFR is shown in Figure 4.

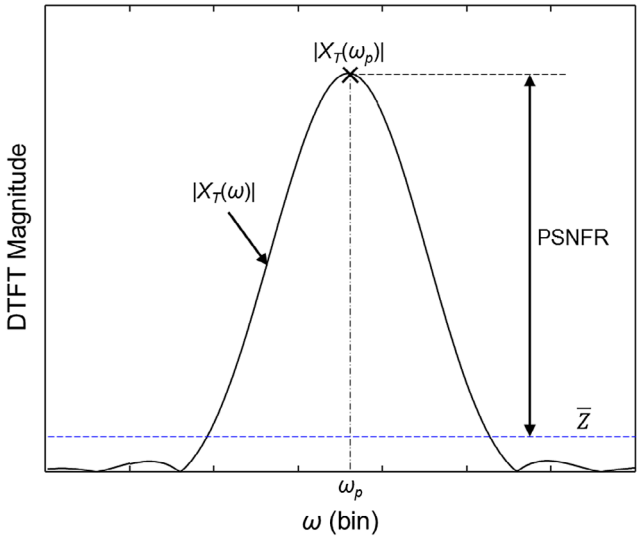

Fig. 4. The peak signal-to-noise floor ratio (PSNFR).

In practice, we cannot estimate the frequency at $\omega = \omega_p$ since ω_p is unknown. Instead, we calculate the frequency in Hz at $\omega = \hat{\omega}_p$, which is an estimate of ω_p, as

$$\hat{f}_p = \begin{cases} \frac{f_s}{2\pi}\{\angle(X(\hat{\omega}_p,1)) - \angle(X(\hat{\omega}_p,0))\}, & \text{if} \angle(X(\hat{\omega}_p,1)) \geq \angle(X(\hat{\omega}_p,0)), \\ \frac{f_s}{2\pi}\{\angle(X(\hat{\omega}_p,1)) - \angle(X(\hat{\omega}_p,0)) + 2\pi\}, & \text{otherwise}, \end{cases} \tag{18}$$

where $X(\omega, h)$ is the DTFT of $x[n + h]$. Whereas k_p in the PDE is an integer DFT index, $\hat{\omega}_p$ is a real-valued frequency expressed in bin.

The condition regarding $\hat{\omega}_p$ is that it should be sufficiently close to ω_p. Therefore, we can use any conventional frequency estimation methods so long as their estimate $\hat{\omega}_p$ is sufficiently accurate. Specifically, we take two methods as examples. The first method is to use the PDE for estimation of $\hat{\omega}_p$ as

$$\hat{\omega}_p = \begin{cases} \frac{N}{2\pi}\{\angle(X[k_p,1]) - \angle(X[k_p,0])\}, & \text{if} \angle(X[k_p,1]) \geq \angle(X[k_p,0]), \\ \frac{N}{2\pi}\{\angle(X[k_p,1]) - \angle(X[k_p,0]) + 2\pi\}, & \text{otherwise}. \end{cases} \tag{19}$$

Then, $\hat{f}_p$ is estimated using the PDE at $\omega = \hat{\omega}_p$.

The second method is to use zero-padding, which has been frequently used with a DFT. If we compose $x_{ZP}[n]$ by adding $(L-1) \times N$ zeros to the end of $x[n]$, its $L \times N$-point DFT becomes

$$X_{ZP}[k] = \sum_{n=0}^{LN-1} x_{ZP}[n]w[n]e^{-j\frac{2\pi k}{LN}n} = \sum_{n=0}^{N-1} x[n]w[n]e^{-j\frac{2\pi k}{LN}n}, \quad k = 0, 1, \ldots, LN-1. \tag{20}$$

Its DFT index closest to the pth spectral peak is selected as

$$k_{p,ZP} = \arg\max_k |X_{ZP}[k]|, \tag{21}$$

and $\hat{\omega}_p = k_{p,ZP}/L$. Since $X(\hat{\omega}_p) = X_{ZP}[k_{p,ZP}]$, the PDE calculates $\hat{f}_p$ directly using $X_{ZP}[k_{p,ZP}]$. Since $X_{ZP}[k]$ is a sampled version of $X(\omega)$ with an actual step size of $1/L$ bin, $\hat{\omega}_p$ approaches the true frequency ω_p as L increases. According to preliminary experiments, zero-padding was adequate for the purpose for $L \geq 4$. Therefore, we use zero-padding with $L = 4$.

Though improved PDE is effective at low SNR, it is also applicable at high SNR. Since $|X_T(\omega_p)| \geq |X_T(k_p)|$, improved PDE reduces the frequency estimation errors in the average sense compared to the PDE. The PDE and improved PDE have different patterns of frequency estimation errors since their frequency estimation positions are different.

3.3 Window switching

When the SNR is high, the frequency estimation errors are caused mainly by the interference component.

Accordingly, the frequency estimation errors are minimised by maximising the ratio $|X_T(\omega)|/|X_I(\omega)|$, which is represented as

$$\frac{|X_T(\omega)|}{|X_I(\omega)|} = \frac{\left|a_p W(\omega-\omega_p)\right|}{\left|\sum_{m=1,m\neq p}^{M} a_m W(\omega-\omega_m)e^{j\theta_m} - \sum_{m=1}^{M} a_m W(\omega+\omega_m)e^{-j\theta_m}\right|}. \quad (22)$$

If ω is close to ω_p, the numerator corresponds to the mainlobe of the target component and the denominator corresponds to the sidelobes of the harmonics except the target component. Therefore, the ratio is determined basically by the sidelobe attenuation of the window function. If the Hanning window is replaced by any window function whose sidelobe level is lower than that of the Hanning window, it will increase the ratio and thus reduce the frequency estimation errors. For example, we take a Blackman window whose DTFT magnitude is shown in Figure 5.

It should be noted that window switching should be used only at high SNR. Since

$$\left|X_T(\omega_p)\right| = \left|\frac{a_p}{2}W(0)\right| = \left|\frac{a_p}{2}\sum_{n=0}^{N-1} w[n]\right|, \quad (23)$$

the PSNFR is represented as

$$\mathrm{PSNFR} = 10\log_{10}\left(\frac{a_p^2\left(\sum_{n=0}^{N-1} w[n]\right)^2}{4\sigma^2\sum_{n=0}^{N-1} w^2[n]}\right). \quad (24)$$

Using the definition of the SNR, it is also represented as

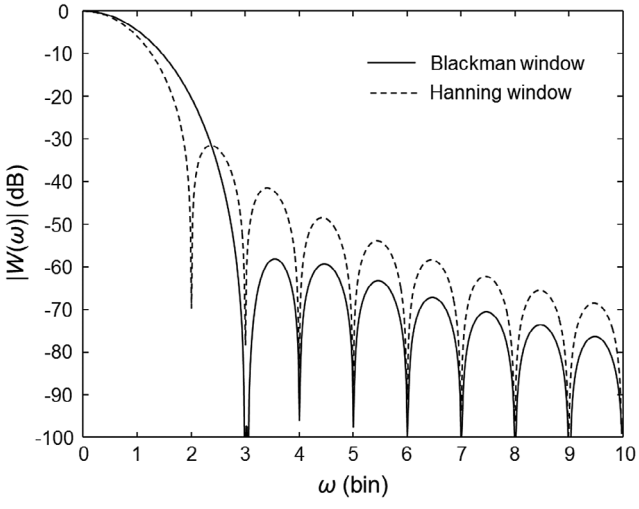

Fig. 5. DTFT magnitudes of Blackman and Hanning windows.

$$\mathrm{PSNFR} = \mathrm{SNR} + 10\log_{10}\left(\frac{\left(\sum_{n=0}^{N-1} w[n]\right)^2}{2\sum_{n=0}^{N-1} w^2[n]}\right), \quad (25)$$

which is a function of the window size N as well as the SNR. With some calculations (see the Appendix 1), it is represented for a Hanning window as

$$\mathrm{PSNFR} = \mathrm{SNR} + 10\log_{10}\left(\frac{N}{3}\right), \quad (26)$$

and for a Blackman window as

$$\mathrm{PSNFR} \approx \mathrm{SNR} + 10\log_{10}(.29\ N). \quad (27)$$

Since the PSNFR using a Blackman window is lower than that using a Hanning window, window switching degrades the performance of the PDE at low SNR. Therefore, window switching should be applied only at high SNR.

It should also be noted that, as described above, window switching is not based on any specific fine frequency estimation methods. Therefore, window switching is applicable to existing fine frequency estimation methods as well as the PDE and improved PDE.

4. Experimental results

Experiments have been performed for assessment of the proposed methods. The experimental set-up is first described and discussions on experimental results are given for each experiment.

4.1 Experimental set-up

In the simulation, the signal was composed of M harmonics of unity amplitudes in white noise as

$$x[n] = \sum_{m=1}^{M} \sin\left(2\pi\frac{m\times f_0}{f_s}n + \theta_m\right) + z[n], \quad n = 0,\ 1,\ldots,\ N, \quad (28)$$

where θ_m was a random variable in the range $[0,\ 2\pi)$. The sampling frequency f_s was 48 kHz and the window sizes N of 512, 1024, and 2048 were used in the DFTs. We estimated $\hat{f}_0$ for the fundamental using the proposed methods and PDE, and then calculated the RMSEs of $\hat{f}_0$. The fundamental frequency f_0 varied from 375 to 1125 Hz for n = 512, 187.5 to 562.5 Hz for N = 1024, and 93.75 to 281.25 Hz for N = 2048, which corresponded to a range of 4–12 bin in the DFTs. At each frequency, the results were average values over 10,000 independent trials.

Since the proposed methods are for fine frequency estimation, we supposed that the coarse frequency estimation was performed previously. Therefore, the DFT index closest to the fundamental, k_1, was found simply by searching for the maximum among DFT magnitude samples around the true frequency. Results are represented for $M = 5$ in Sections 4.2–4.4 and are represented as a function of M in Section 4.5.

4.2 Error analysis of improved phase difference estimation as a function of fundamental frequency

We analysed the frequency estimation errors of improved PDE with $\hat{\omega}_1$ estimated by the PDE and by zero-padding, which we refer to as PDE(PDE) and PDE(ZP), respectively. δ_1 is equal to $\omega_1 - k_1$, which is the difference expressed in bin between the true fundamental frequency and its corresponding DFT index.

At low SNR ($\sigma = 1$), while the frequency estimation errors of the PDE oscillate with their maxima at $\delta_1 = \pm 1/2$ and minima at $\delta_1 = 0$, the errors of improved PDE are relatively constant with small fluctuations as shown in Figure 6. It is worth noting that, when f_0 is low, the errors of improved PDE are relatively large due to the high sidelobe levels of the harmonics. When averaged over all the frequencies in Figure 6, the average frequency estimation errors of PDE(PDE) and PDE(ZP) are smaller than those of the PDE by 1.5 dB and by 1.4 dB. It is also meaningful to calculate the errors averaged over frequencies at which $\delta_1 = \pm 1/2$ in Figure 6, which we refer to as the maximum frequency estimation errors. The resulting maximum errors of PDE(PDE) and PDE(ZP) are smaller than those of the PDE by 3.0 to 3.2 dB and by 3.4 dB.

In addition, the probability density functions (PDFs) of the frequency estimation errors are depicted in consideration of δ_1 in Figure 7, which were calculated for $\sigma = 1$ and $N = 1024$. The PDFs of the PDE depend on δ_1. Specifically, for $.4 \le |\delta_1| \le .5$, i.e. ω_1 is close to $k_1 \pm 1/2$, the PDFs are relatively wide. In contrast, the PDFs of PDE (PDE) are relatively narrow irrespective of δ_1, which means that its probability of large frequency estimation errors is consistently lower than that of the PDE at low SNR. Though not depicted, the PDFs of PDE(ZP) resemble those of PDE(PDE).

At high SNR ($\sigma = .001$), the frequency estimation errors of improved PDE decrease with fluctuations as f_0 increases, but their error patterns are different from those of the PDE as shown in Figure 6. When averaged over all the frequencies in Figure 6, the errors of PDE(PDE) and PDE(ZP) are smaller than those of the PDE by .5 dB. For example, the average frequency estimation errors of the PDE, PDE(PDE), and PDE(ZP) are .2182, .2064, and .2069 Hz, respectively, which are calculated from the results in Figure 6(b).

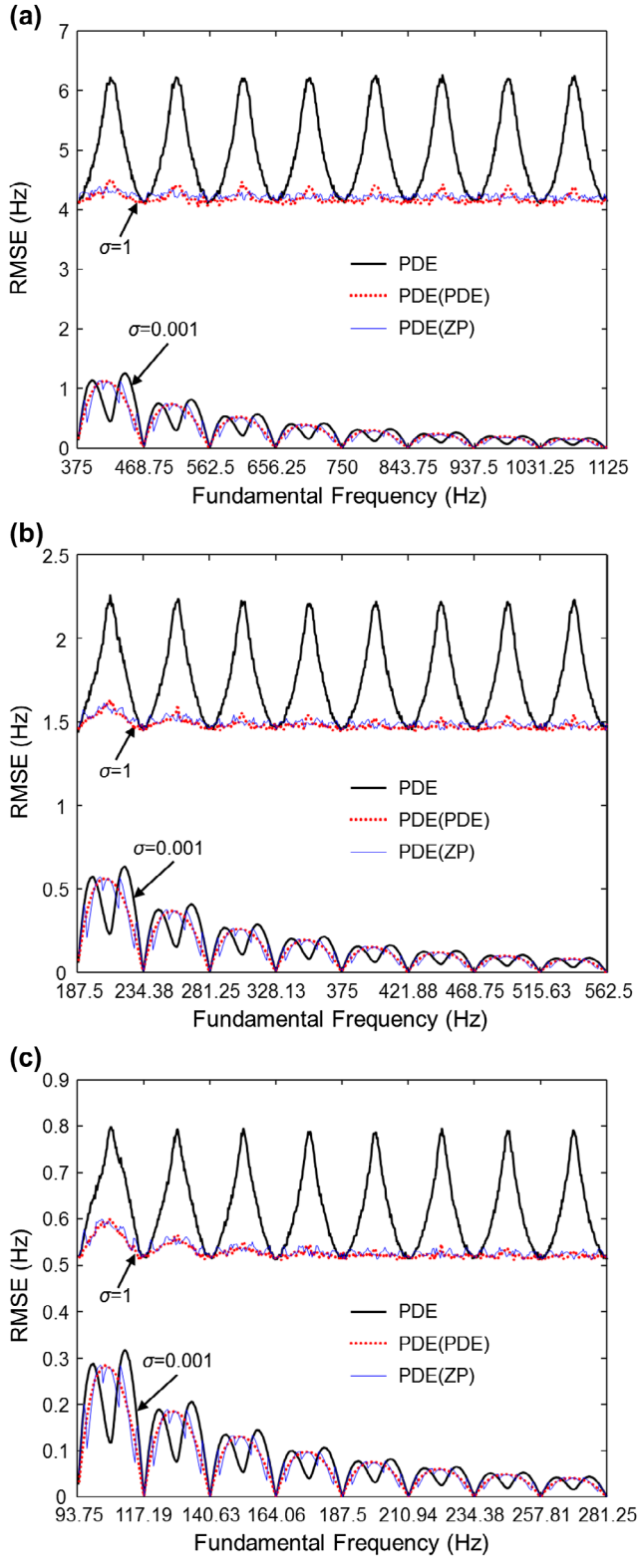

Fig. 6. Frequency estimation errors of the PDE and improved PDE as a function of fundamental frequency for the signal in Equation 28 ($M = 5$). (a) $N = 512$. (b) $N = 1024$. (c) $N = 2048$.

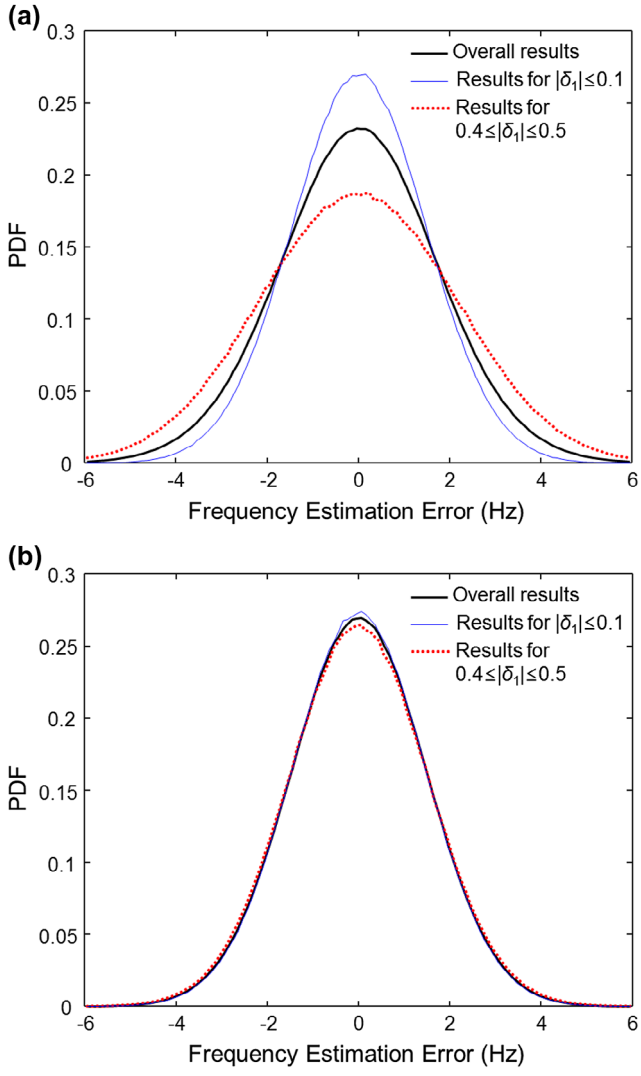

Fig. 7. PDFs of the frequency estimation errors ($M = 5$, $\sigma = 1$, and $N = 1024$). (a) PDE. (b) PDE(PDE).

4.3 Error analysis of window switching as a function of fundamental frequency

We refer to the PDE with window switching as PDE(WS) and PDE(PDE) with window switching as PDE(PDE,WS). At high SNR ($\sigma = .001$), the frequency estimation errors of PDE(WS) and PDE(PDE,WS) are much smaller than those of the PDE as shown in Figure 8, where the error patterns of PDE(PDE,WS) are different from those of the PDE. When averaged over all the frequencies in Figure 8, the average frequency estimation errors of PDE(WS) and PDE (PDE,WS) are smaller than those of the PDE by 9.5 dB and by 9.9 dB. At high SNR, window switching is more effective than improved PDE.

In addition, the PDFs of the frequency estimation errors are depicted in Figure 9, which were calculated for $\sigma = .001$ and $N = 1024$. The PDFs of PDE(WS) and PDE (PDE,WS) are relatively narrow, which means that their probabilities of large frequency estimation errors are consistently lower than that of the PDE at high SNR.

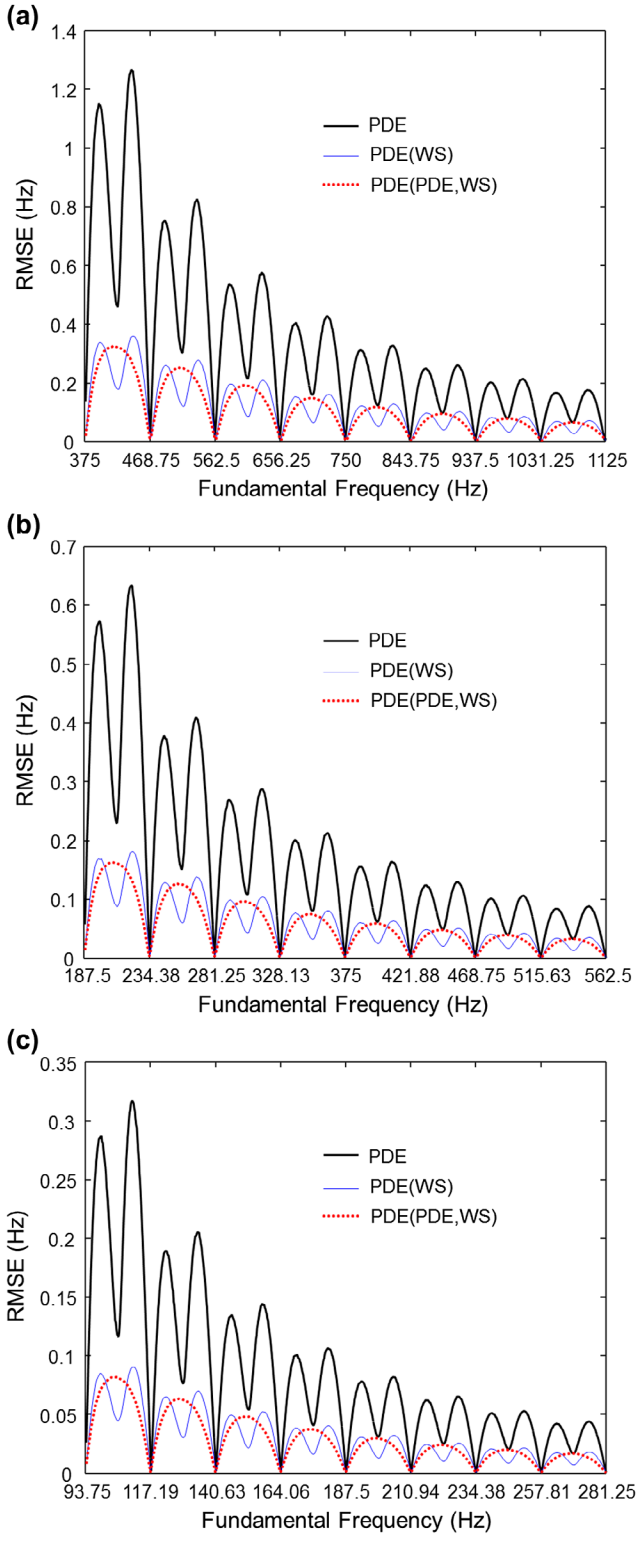

Fig. 8. Frequency estimation errors of the PDE, PDE(WS), and PDE(PDE,WS) as a function of fundamental frequency ($M = 5$ and $\sigma = .001$). (a) $N = 512$. (b) $N = 1024$. (c) $N = 2048$.

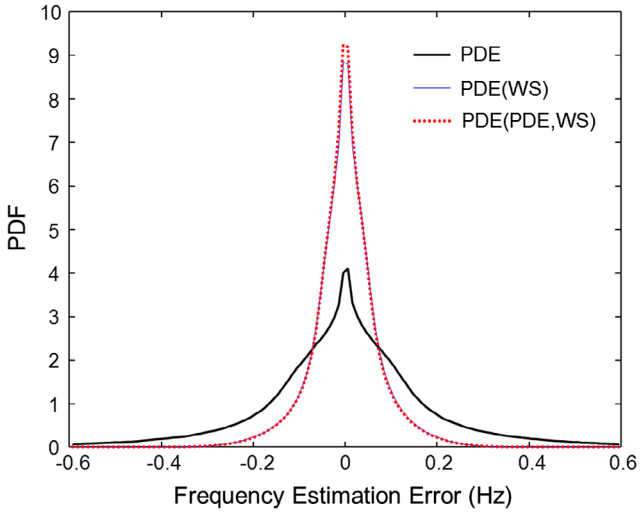

Fig. 9. PDFs of the frequency estimation errors ($M = 5$, $\sigma = .001$, and $N = 1024$).

4.4 Error analysis as a function of SNR

The frequency estimation errors are depicted as a function of SNR in Figure 10, where the results are average values over all the frequencies in the range [4, 12 bin]. In addition to the PDE, PDE(PDE), and PDE(PDE,WS), we calculated the RMSEs of the corrected quadratically interpolated fast Fourier transform (CQIFFT) (Abe & Smith, 2004), which is known to estimate the frequency of a sinusoid very accurately (Dixon et al., 2012; Tidhar et al., 2010). We also applied window switching to the CQIFFT, i.e. we used it either with a Hanning window or with a Blackman window. When it is used with a Blackman window, we refer to it as CQIFFT(WS). When it is used with a Hanning window, we refer to it as CQIFFT. For both cases, a zero-padding factor of 4 was used ($L = 4$).

The errors decrease as the SNR increases at low SNR and are almost constant at high SNR in Figure 10, where no single method is superior to the other methods at all SNR levels. At low SNR, the RMSEs of PDE(PDE) are the smallest. For example, the RMSEs of PDE(PDE), CQIFFT, and PDE are 1.0572, 1.2505, and 1.2577 Hz, respectively, at 0 dB SNR in Figure 10(b). At high SNR, the RMSEs of CQIFFT(WS) are the smallest and the RMSEs of PDE(PDE,WS) are the second smallest. For example, the RMSEs of CQIFFT(WS), PDE(PDE,WS), CQIFFT, PDE(PDE), and PDE are .0358, .0701, .0903, .2067, and .2184 Hz respectively at 50 dB SNR in Figure 10(b). At high SNR, window switching is effective in reducing the RMSEs of both CQIFFT and PDE(PDE). Though not depicted, PDE(ZP) and PDE(ZP,WS) provide comparable results to those of PDE(PDE) and PDE (PDE,WS).

Based on the results in Figure 10, we first found the SNR values at which the RMSEs of PDE(PDE) are equal to

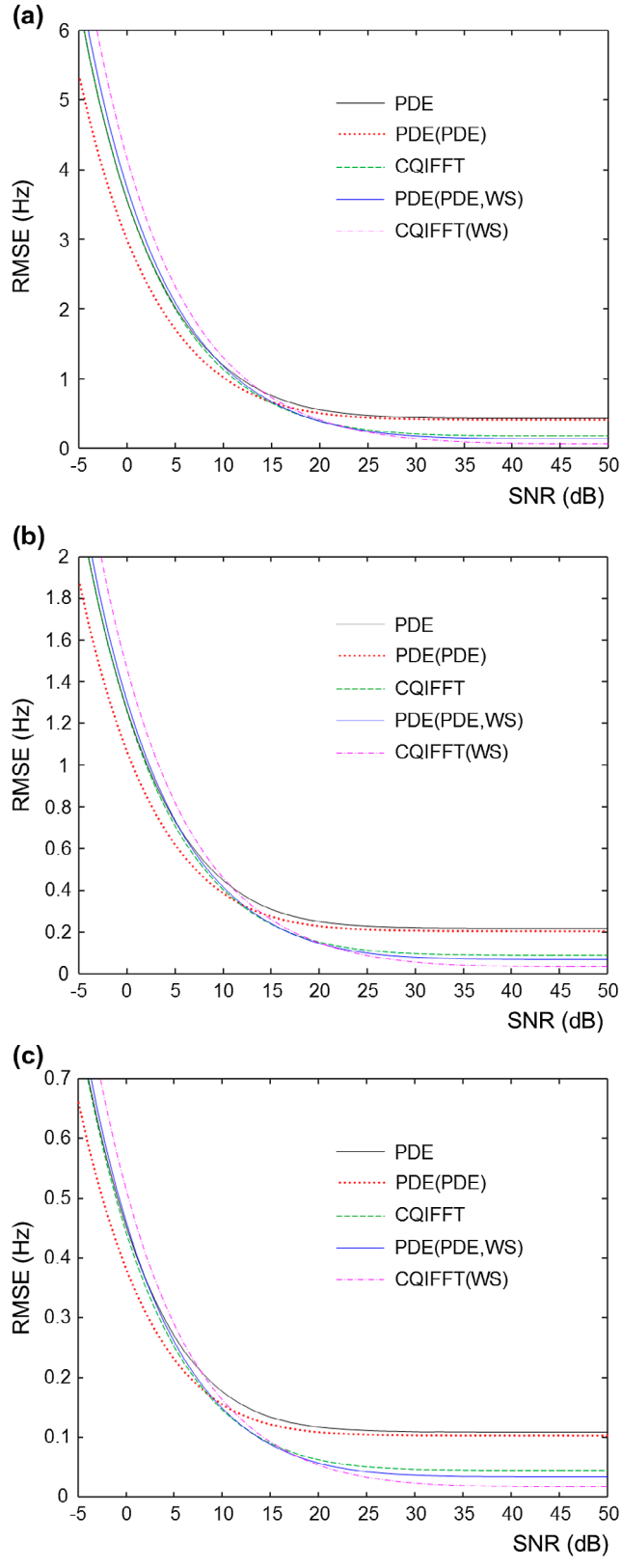

Fig. 10. Frequency estimation errors as a function of SNR ($M = 5$). (a) $N = 512$. (b) $N = 1024$. (c) $N = 2048$.

those of PDE(PDE,WS) and CQIFFT(WS) for $N = 512$, 1024, and 2048. Then, we performed additional experiments to find the corresponding SNR values for $N = 768$, 1536, and 3072. Finally, PSNFR values were calculated from the SNR values for a Hanning window, i.e. PSNFR = SNR + $10\log_{10}(N/3)$. The SNR and PSNFR values are listed in Table 1, where the PSNFR values are almost constant. If a PSNFR value using a Hanning window is larger than the values in Table 1, the RMSEs of CQIFFT (WS) or PDE(PDE,WS) are smaller than those of PDE (PDE) in our experiments. It should be noted that the results in Table 1 are valid for our signal model and experimental conditions. Further research is required to generalise the results for application to various types of musical sounds.

4.5 Error analysis as a function of the number of harmonics

The frequency estimation errors are depicted as a function of M in Figure 11, where the results are average values over all the frequencies in the range [4, 12 bin]. At low SNR ($\sigma = 1$), the errors of all the methods are almost constant irrespective of M. The results show that the noise component rather than the harmonics is dominant at low SNR. At high SNR ($\sigma = .001$), the errors of all the methods increase when M increases from 1 to 2 and then are almost constant irrespective of M. The results show that the harmonic whose sidelobe has the strongest effect on the fundamental is dominant at high SNR, which is the second harmonic in our signal model.

Table 1. SNR and PSNFR values at which the RMSEs of PDE (PDE) are equal to those of PDE(PDE,WS) and CQIFFT(WS). The PSNFR values were calculated from the SNR values for a Hanning window.

	PDE(PDE,WS)		CQIFFT(WS)	
N	SNR(dB)	PSNFR(dB)	SNR(dB)	PSNFR(dB)
512	15.2	37.5	17	39.3
768	13.4	37.5	15.2	39.3
1024	12.1	37.4	14	39.3
1536	10.4	37.5	12.2	39.3
2048	9.1	37.4	10.9	39.2
3072	7.3	37.4	9.2	39.3

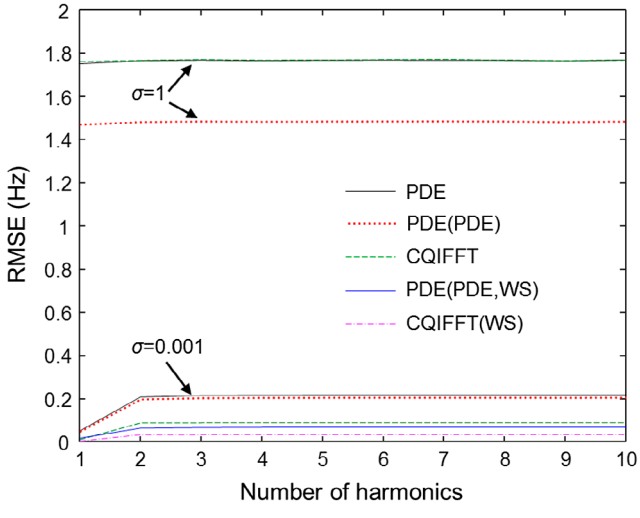

Fig. 11. Frequency estimation errors as a function of the number of harmonics M ($N = 1024$).

5. Application examples

We applied the proposed methods to the frequency tracking of musical instrument sounds. In all examples, we first performed a short-time Fourier transform (STFT) and performed coarse frequency estimation on a frame basis, which was based on manual detection of spectral peaks in the FFT magnitudes. Then we applied either PDE(PDE) or PDE(PDE,WS) in consideration of the values in Table 1. Specifically, we applied PDE(PDE,WS) if measured PSNFR values were sufficiently larger than the PSNFR vales in Table 1 for all the frames. Otherwise, we applied PDE(PDE).

The first example is the oboe sound 'Oboe.ff.A4.stereo.aif' whose pitch was A4, which was taken from the University of Iowa musical instrument samples (Fritts, 2012). The sound was in 24-bit format with a sampling frequency of 44.1 kHz, which was recorded in an anechoic chamber. The sound was recorded in mono, i.e. left and right channel sounds were identical, and it had duration of 1.57 s. We used a window size of 1024 samples and an overlapping size of 512 samples between frames in the STFT. Since the sound was clean as shown in Figure 12(a) and the PSNFRs for the fundamental and second harmonic were observed to be sufficiently high, we applied PDE (PDE,WS) and CQIFFT(WS). In Figure 12(b) and (c), the frequency tracking results for the fundamental and second harmonic are depicted for the PDE, PDE(PDE,WS), and CQIFFT(WS), where the frequency ranges were restricted to [436.5, 441.5 Hz] and [873, 883 Hz] to clearly compare the results. PDE(PDE,WS) and CQIFFT(WS) provide roughly comparable results to each other, which are slightly different from those of the PDE.

The second example is 'dev2_wdrums_inst_mix.wav' which was taken from the sixth community-based signal separation evaluation campaign (SiSEC, 2016). The sound was in 16-bit format with a sampling frequency of 16 kHz and had duration of 10 s. It was a mixture of a solo and two kinds of drum sounds, where both the drum sounds had wide bandwidths. We used a window size of 1024 samples and an overlapping size of 512 samples between frames in the STFT. According to preliminary analysis, while the fundamental and second harmonic were strong, the third and fourth harmonics were relatively weak and

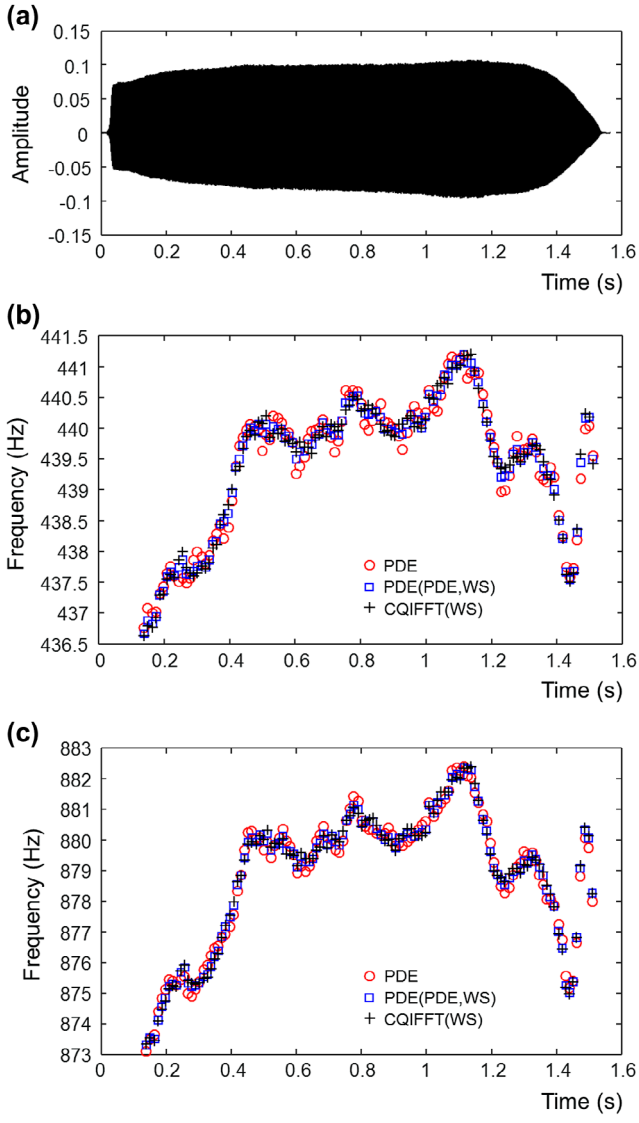

Fig. 12. Frequency tracking results of the oboe sound using the PDE, PDE(PDE,WS), and CQIFFT(WS). (a) Waveform. (b) Results for the fundamental. (c) Results for the second harmonic.

their PSNFRs were not always sufficiently high. We therefore applied PDE(PDE) to the frequency tracking of the first four harmonics of the solo sound, whose results are depicted in Figure 13(b). In the results, note changes are clearly observed for all the harmonics.

The third example is an excerpt from 'glock.wav' which was one of the test items used for the MPEG Surround standardisation (ISO/IEC, 2004). It was the 5.1-channel sound in 16-bit format with a sampling frequency of 44.1 kHz and consisted of bell and drum sounds. We used a window size of 2048 samples and an overlapping size of 1024 samples between frames in the STFT. According to preliminary analysis, the PSNFRs for some higher partials were not always sufficiently high. In addition, due to the decay characteristics of bell sounds, lower partials gradually become weak as time passes.

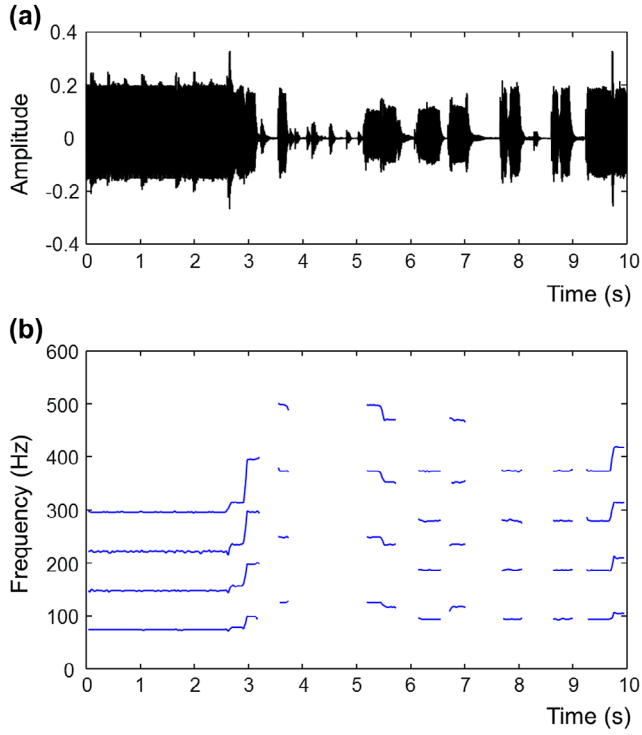

Fig. 13. Frequency tracking results of 'dev2_wdrums_inst_mix.wav' using PDE(PDE) (excerpt: left channel). (a) Waveform. (b) Results for the first four harmonics.

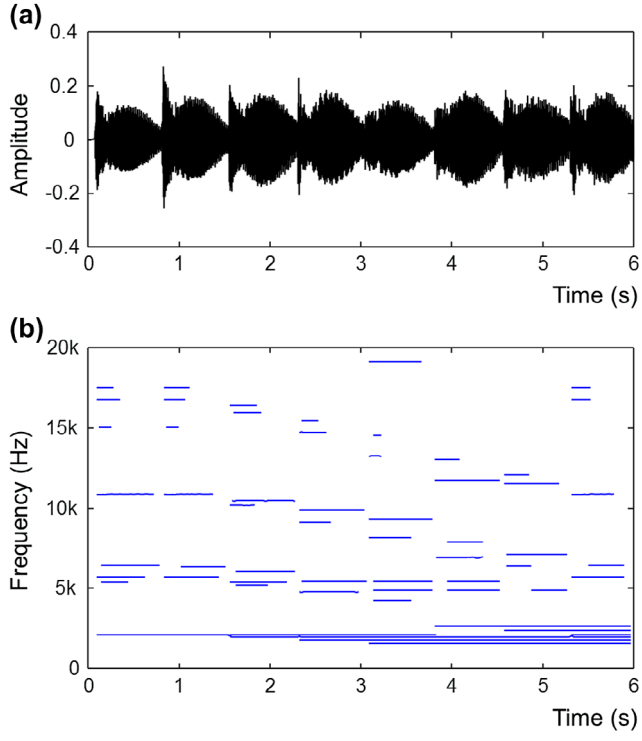

Fig. 14. Frequency tracking results of an excerpt from 'glock.wav' using PDE(PDE) (excerpt: .5 to 6.5 s, front right channel). (a) Waveform. (b) Results for the partials.

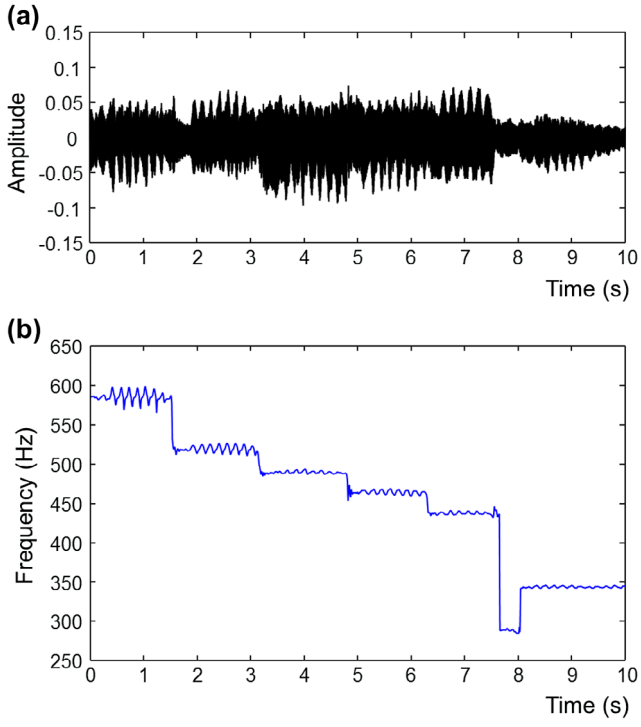

Fig. 15. Frequency tracking results of the flute sound using PDE (PDE,WS) (excerpt: 1:11.7–1:21.7, track 5, left channel). (a) Waveform. (b) Results for the fundamental.

In consideration of these facts, we applied PDE(PDE) to the frequency tracking of the partials of the bell sound, whose results are depicted in Figure 14(b). In the results, the inharmonic partials are observed and the lowest partials in the range [1.5, 2.7 kHz] are observed to last for several seconds.

The last example is a flute solo which was an excerpt from a CD (Menuhin, 1995). The sound was in 16-bit format with a sampling frequency of 44.1 kHz. As shown in Figure 15(a), the flute was played with vibrato. We used a window size of 2048 samples and an overlapping size of 512 samples between frames in the STFT. According to preliminary analysis, the PSNFR for the fundamental was sufficiently high. Therefore, we applied PDE(PDE,WS) to the frequency tracking of the fundamental, whose results are depicted in Figure 15(b). In the results, frequency vibrato is clearly observed.

6. Conclusion

We proposed improved PDE and window switching for high-accuracy frequency analysis of harmonic signals and showed their results from various viewpoints. Since high-accuracy frequency estimation methods are used as a module for sound analysis and processing tools and software, the proposed methods can be adopted as a new module for them. The proposed methods are expected to contribute to the comprehensive analysis of musical signals by providing the high-accuracy frequency tracking results of harmonic signals.

Acknowledgements

The authors thank the anonymous reviewers for their insightful and detailed comments that helped us to improve this paper.

ORCID

Hee-Suk Pang http://orcid.org/0000-0002-7108-8957

References

Abe, M., & Smith, J. O. (2004). *CQIFFT: Correcting bias in a sinusoidal parameter estimator based on quadratic interpolation of FFT magnitude peaks* (Technical Report No. STAN-M-117). Center for Computer Research in Music and Acoustics, Department of Music, Stanford University.

Akkoç, C. (2002). Non-deterministic scales used in traditional Turkish music. *Journal of New Music Research, 31*, 285–293.

Betser, M., Collen, P., Richard, G., & David, B. (2006). Review and discussion on classical STFT-based frequency estimators. *AES 120th convention* (Paper Number 6765). Paris, France.

Bozkurt, B. (2008). An automatic pitch analysis method for Turkish maqam music. *Journal of New Music Research, 37* (1), 1–13.

Bozkurt, B. (2012). A system for tuning instruments using recorded music instead of theory-based frequency presets. *Computer Music Journal, 36*, 43–56.

Bozkurt, B., Yarman, O., Karaosmanoğlu, M. K., & Akkoç, C. (2009). Weighing diverse theoretical models on Turkish maqam music against pitch measurements: A comparison of peaks automatically derived from frequency histograms with proposed scale tones. *Journal of New Music Research, 38*, 45–70.

Brown, J. C., & Puckette, M. S. (1993). A high resolution fundamental frequency determination based on phase changes of the Fourier transform. *Journal of the Acoustical Society of America, 94*, 662–667.

Cañadas Quesada, F. J., Ruiz Reyes, N., Vera Candeas, P., Carabias, J. J., & Maldonado, S. (2010). A multiple-F0 estimation approach based on Gaussian spectral modelling for polyphonic music transcription. *Journal of New Music Research, 39*, 93–107.

Chin, S. H. L., & Berger, J. (2010). Analysis of pitch perception of inharmonicity in pipa strings using response surface methodology. *Journal of New Music Research, 39*, 63–73.

Desainte-Catherine, M., & Marchand, S. (2000). High-precision Fourier analysis of sounds using signal derivatives. *Journal of the Audio Engineering Society, 48*, 654–667.

Dixon, S., Mauch, M., & Tidhar, D. (2012). Estimation of harpsichord inharmonicity and temperament from musical recordings. *Journal of the Acoustical Society of America, 131*, 878–887.

Flandrin, P., Auger, F., & Chassande-Mottin, E. (2003). Time-frequency reassignment: From principles to algorithms. In A. Papandreou-Suppappola (Ed.), *Applications in time-frequency signal processing* (pp. 179–204). Boca Raton, FL: CRC Press.

Fritts, L. (2012). *University of Iowa musical instrument samples*. Retrieved from https://theremin.music.uiowa.edu

Galembo, A., & Askenfelt, A. (1999). Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano. *IEEE Transactions on Speech and Audio Processing, 7*, 197–203.

Gómez, E., & Bonada, J. (2013). Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing. *Computer Music Journal, 37*, 73–90.

Gulati, S., Bellur, A., Salamon, J., Ranjani, H. G., Ishwar, V., Murthy, H. A., & Serra, X. (2014). Automatic tonic identification in Indian art music: Approaches and evaluation. *Journal of New Music Research, 43*, 53–71.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE, 66*, 51–83.

Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. *Journal of the Acoustical Society of America, 83*, 257–264.

Hikichi, T., Osaka, N., & Itakura, F. (2004). Sho-So-In: Control of a physical model of the sho by means of automatic feature extraction from real sounds. *Journal of New Music Research, 33*, 355–365.

ISO/IEC. (2004). *Performance of MPEG Surround Technology* (ISO/IEC JTC 1/SC 29/WG 11/N7701).

Koduri, G. K., Gulati, S., Rao, P., & Serra, X. (2012). Rāga recognition based on pitch distribution methods. *Journal of New Music Research, 41*, 337–350.

Koduri, G. K., Ishwar, V., Serrà, J., & Serra, X. (2014). Intonation analysis of rāgas in carnatic music. *Journal of New Music Research, 43*, 72–93.

Kulesza, M., & Czyzewski, A. (2009). Tonality estimation and frequency tracking of modulated tonal components. *Journal of the Audio Engineering Society, 57*, 221–236.

Lagrange, M., & Marchand, S. (2007). Estimating the instantaneous frequency of sinusoidal components using phase-based methods. *Journal of the Audio Engineering Society, 55*, 385–399.

Lagrange, M., Marchand, S., & Rault, J.-B. (2005). Improving sinusoidal frequency estimation using a trigonometric approach, In *Proceedings of the International Conference on Digital Audio Effects* (pp. 110–115). Madrid, Spain.

Lagrange, M., Martins, L. G., Murdoch, J., & Tzanetakis, G. (2008). Normalized cuts for predominant melodic source separation. *IEEE Transactions on Audio, Speech, and Language Processing, 16*, 278–290.

Li, H., & Leman, M. (2007). A gesture-based typology of sliding-tones in guqin music. *Journal of New Music Research, 36*, 61–82.

McPherson, A. (2010). The magnetic resonator piano: Electronic augmentation of an acoustic grand piano. *Journal of New Music Research, 39*, 189–202.

Menuhin, Y. (1995). Instruments of the orchestra [CD]. EMI Classics.

Nunes, L., Esquef, P., & Biscainho, L. (2009). FlexSM: A flexible sinusoidal modeling system. *Journal of the Audio Engineering Society, 57*, 1042–1056.

Nuttall, A. H. (1981). Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 29*, 84–91.

Pang, H. S. (2004). On the use of the maximum likelihood estimation for analysis of vibrato tones. *Applied Acoustics, 65*, 101–107.

Pang, H. S., Lim, J. S., Kwon, O. J., & Shin, B. J. (2012). Iterative frequency estimation for accuracy improvement of three DFT phase-based methods. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E95-A*, 969–973.

Pang, H. S., & Yoon, D. H. (2005). Automatic detection of vibrato in monophonic music. *Pattern Recognition, 38*, 1135–1138.

Puckette, M. S., & Brown, J. C. (1998). Accuracy of frequency estimates using the phase vocoder. *IEEE Transactions on Speech and Audio Processing, 6*, 166–176.

Rao, P., Ross, J. C., Ganguli, K. K., Pandit, V., Ishwar, V., Bellur, A., & Murthy, H. A. (2014). Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research, 43*, 115–131.

Rapoport, E., Shatz, S., & Blass, N. (2008). Overtone spectra of gongs used in music therapy. *Journal of New Music Research, 37*, 37–60.

Rauhala, J., Lehtonen, H. M., & Välimäki, V. (2007). Fast automatic inharmonicity estimation algorithm. Journal of the Acoustical Society of America*, 121*(5), EL184-EL189.

Ryynänen, M. P., & Klapuri, A. P. (2008). Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal, 32*, 72–86.

Şentürk, S., Holzapfel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research, 43*, 34–52.

Sethares, W. A., Milne, A. J., Tiedje, S., Prechtl, A., & Plamondon, J. (2009). Spectral tools for dynamic tonality and audio morphing. *Computer Music Journal, 33*, 71–84.

Shirazi, J., & Ghaemmaghami, S. (2010). Improvement to speech-music discrimination using sinusoidal model based features. *Multimedia Tools and Applications, 50*, 415–435.

SiSEC. (2016). *Underdetermined-speech and music mixtures*. Retrieved from https://sisec.inria.fr/home/2016-underdetermined-speech-and-music-mixtures/

Six, J., Cornelis, O., & Leman, M. (2013). Tarsos, a modular platform for precise pitch analysis of western and non-western music. *Journal of New Music Research, 42*, 113–129.

Tidhar, D., Mauch, M., & Dixon, S. (2010). High precision frequency estimation for harpsichord tuning classification. In *Proceedings of the IEEE international conference on Acoustics, Speech, and Signal Processing* (pp. 61–64). Dallas, USA.

Zakis, J. A., McDermott, H. J., & Vandali, A. E. (2007). A fundamental frequency estimator for the real-time processing of musical sounds for cochlear implants. *Speech Communication, 49*, 113–122.

Zivanovic, M., Röbel, A., & Rodet, X. (2008). Adaptive threshold determination for spectral peak classification. *Computer Music Journal, 32*, 57–67.

Appendix 1

The PSNFR is represented as

$$\mathrm{PSNFR} = \mathrm{SNR} + 10\log_{10}\left(\frac{\left(\sum_{n=0}^{N-1} w[n]\right)^2}{2\sum_{n=0}^{N-1} w^2[n]}\right).$$

Suppose that a window function $w[n]$ can be represented as

$$w[n] = a_0 + a_1\cos\left(\frac{2\pi n}{N}\right) + a_2\cos\left(\frac{4\pi n}{N}\right),$$
$$n = 0, 1, \cdots, N-1.$$

Then,

$$\sum_{n=0}^{N-1} w[n] = \sum_{n=0}^{N-1}\left\{a_0 + a_1\cos\left(\frac{2\pi n}{N}\right) + a_2\cos\left(\frac{4\pi n}{N}\right)\right\}.$$

Since $\sum_{n=0}^{N-1}\cos\left(\frac{2\pi n}{N}\right) = \sum_{n=0}^{N-1}\cos\left(\frac{4\pi n}{N}\right) = 0$,

$$\sum_{n=0}^{N-1} w[n] = Na_0.$$

Now,

$$\sum_{n=0}^{N-1} w^2[n] = \sum_{n=0}^{N-1}\{a_0^2 + a_1^2\cos^2\left(\frac{2\pi n}{N}\right) + a_2^2\cos^2\left(\frac{4\pi n}{N}\right) + 2a_0a_1\cos\left(\frac{2\pi n}{N}\right)$$
$$+ 2a_0a_2\cos\left(\frac{4\pi n}{N}\right) + 2a_1a_2\cos\left(\frac{2\pi n}{N}\right)\cos\left(\frac{4\pi n}{N}\right)\}.$$

Since $\sum_{n=0}^{N-1}\cos^2\left(\frac{2\pi n}{N}\right) = \sum_{n=0}^{N-1}\cos^2\left(\frac{4\pi n}{N}\right) = \frac{N}{2}$ and

$\sum_{n=0}^{N-1}\cos\left(\frac{2\pi n}{N}\right)\cos\left(\frac{4\pi n}{N}\right) = 0$,

$$\sum_{n=0}^{N-1} w^2[n] = \frac{N}{2}\left(2a_0^2 + a_1^2 + a_2^2\right).$$

Then, the PSNFR is represented as

$$\mathrm{PSNFR} = \mathrm{SNR} + 10\log_{10}\left(\frac{Na_0^2}{2a_0^2 + a_1^2 + a_2^2}\right).$$

For a Hanning window, $a_0 = .5$, $a_1 = -.5$, and $a_2 = 0$ (Harris, 1978). Then,

$$\mathrm{PSNFR} = \mathrm{SNR} + 10\log_{10}\left(\frac{N}{3}\right).$$

For a Blackman window, $a_0 = .42$, $a_1 = -.5$, and $a_2 = .08$ (Harris, 1978). Then,

$$\mathrm{PSNFR} \approx \mathrm{SNR} + 10\log_{10}(.29\ N).$$