

Non-iterative frequency estimation in the DFT magnitude domain

Ricardo Sousa and Aníbal Ferreira

Abstract—The accurate estimation of the frequency of sinusoids is a frequent problem in many signal processing problems including the real-time analysis of the singing voice. In this paper we rely on a single DFT magnitude spectrum in order to perform frequency estimation in a non-iterative way. Two new frequency estimation methods are derived that are matched to the time analysis window and that reduce the maximum absolute estimation error to about 0.1% of the bin width of the DFT. The performance of these methods is evaluated including the parabolic method as a reference, and considering the influence of noise. A combined model is proposed that offers higher noise robustness than that of a single model.

I. INTRODUCTION

A frequent problem encountered is many signal processing problems, including speech/audio coding, singing analysis, or automatic transcription of PCM music to symbolic music notation (*e.g.*, MIDI), is the accurate estimation of the frequency of sinusoids. Two specific application areas motivate our research: polyphonic pitch estimation and accurate analysis of the harmonic structure of the singing voice. In addition, we focus on real-time interactive applications which means processing delay and computational complexity are important issues. Frequency estimation of sinusoids may be performed using time domain techniques or frequency domain techniques. While the former are mainly based on the autocorrelation function, the latter are mainly based on the phase derivative of the DFT spectrum, on cepstral analysis, or on peak analysis of the DFT magnitude spectrum. For a survey of common techniques please see [1], [2]. In a separate paper, we discuss the relative performance of several frequency estimation algorithms in the context of harmonic signals [3].

Some pitch estimation techniques require the computation of two DFTs, others rely on multiple partials of a harmonic structure, and others require complex iterative [4] or analysis-by-synthesis procedures [5]. Our research follows the approach described in [6] where a single-step estimation procedure is carried out using a single DFT magnitude spectrum. This approach paves the way for the accurate frequency estimation of all individual sinusoids existing in a signal. Not only this facilitates a detailed analysis of the quality of an existing harmonic structure (*e.g.*, regarding the singing voice), but also facilitates multi-pitch analysis in the frequency domain. In addition, since it is important that fast frequency variations are captured in singing or speech, the analysis must be performed taking into consideration the duration of short voiced sounds which may be around 10 ms or even shorter. In this paper we presume the simplest possible analysis scenario: a signal $x(n)$ is multiplied by a window function $h(n)$, of length N , and the result is DFT transformed. The resulting discrete frequency representation $V(k)$ is used as input to an algorithm that identifies

all peaks corresponding to sinusoids in the magnitude spectrum, $|V(k)|$, and that delivers for each one the best frequency estimate, in a non-iterative way. This paper is structured as follows. In section II we present the analysis framework and detail how two new window-aware frequency estimation methods are derived. These methods consist of closed-form formulas that get inspiration from the results obtained in [6] and that are adapted to the rectangular window, and to the square root of a Hanning window (*i.e.*, a sine window). The popular frequency estimation method based on the ‘parabolic model’ is also described as a reference window-agnostic procedure that is used with the Hanning window. The relative noise robustness of the three methods is discussed. In section III a combined procedure is presented that uses two of the above methods in order to provide frequency estimates that are more robust to noise. Section IV concludes this paper with a discussion on the significance of the results and possible application areas.

II. ACCURATE FREQUENCY ESTIMATION

While the results presented in [6] presume the Odd-DFT, in this paper we presume the DFT since it is more frequently used in general frequency analysis. For simplicity but without loss of generality, we consider that $x(n)$ represents a sinusoidal signal of arbitrary frequency that is corrupted by additive Gaussian noise, $r(n)$:

$$x(n) = A \sin \left[\frac{2\pi}{N}(\ell + \Delta\ell)n + \phi \right] + r(n). \quad (1)$$

In this equation A represents the magnitude of the sinusoid, ℓ and $\Delta\ell$ represent respectively the integer part and the fractional part of the frequency of the sinusoid on a DFT frequency bin scale whose periodicity is N , and ϕ represents the phase of the sinusoid. Three windows of decreasing selectivity and increasing main-side lobe attenuation are considered in this study: the rectangular window

$$h_R(n) = 1, \quad n = 0, 1, \dots, N-1, \quad (2)$$

the sine window

$$h_S(n) = \sqrt{h_H(n)} = \sin \frac{\pi}{N}(n + 0.5), \quad n = 0, \dots, N-1 \quad (3)$$

and the Hanning window

$$h_H(n) = \frac{1}{2} \left[1 - \cos \frac{2\pi}{N}(n + 0.5) \right], \quad n = 0, \dots, N-1. \quad (4)$$

The frequency response (FreqResp) of these windows is illustrated in Fig. 1. This figure shows that the main lobe width of the rectangular, sine and Hanning window is $4\pi/N$, $6\pi/N$, and $8\pi/N$, respectively. This can also be seen as the minimum frequency separation between two sinusoids so that they can be resolved by adjacent DFT spectral lines (*i.e.*, so that they appear as individual peaks in the DFT magnitude spectrum). In this perspective, the rectangular window has the best selectivity and the Hanning window has the poorest selectivity. On the other hand, the larger the main lobe, the better the attenuation between the main lobe and the side lobes, a feature also referred to as ‘leakage’. In this perspective, the rectangular window has the largest leakage and the Hanning

This work was supported by the Portuguese Foundation for Science and Technology, an agency of the Portuguese Ministry for Science, Technology and Higher Education, under grant SFRH/BD/24811/2005

R. Sousa is with the Department of Electrical and Computer Engineering, University of Porto - Faculty of Engineering, Porto, Portugal, e-mail: sousa.ricardo@fe.up.pt

A. Ferreira is with the Department of Electrical and Computer Engineering, University of Porto - Faculty of Engineering, Porto, Portugal, e-mail: ajf@fe.up.pt

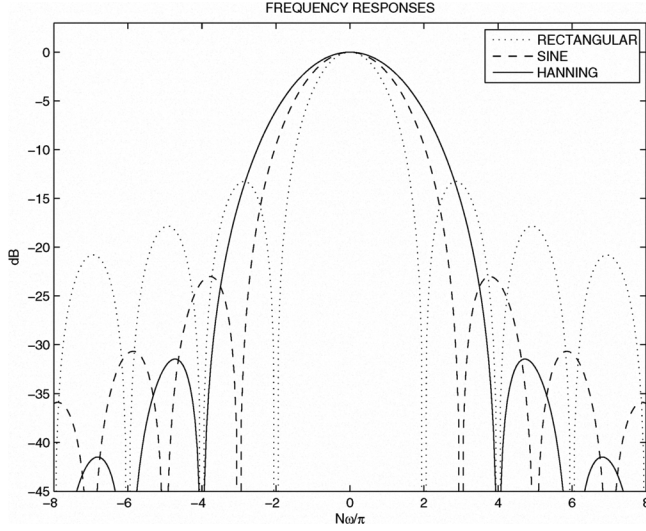


Fig. 1. Frequency response of the rectangular window (dotted), the sine window (dashed), and Hanning window (solid).

window has the smallest leakage. The lower the leakage, the lower the mutual influence between two resolved sinusoids in the DFT magnitude spectrum. These are therefore opposite aspects that are important in spectral analysis. In all cases, the estimation can be formulated in a simple way: given the input signal according to (1), find the values of ℓ and $\Delta\ell$ after the signal has been multiplied by one of the analysis windows (2-4) and transformed to the frequency domain using an N -point DFT, $V(k) = \sum_{n=0}^{N-1} h(n)x(n)e^{-j\frac{2\pi}{N}kn}$. The natural frequency resolution of the N -point DFT (or bin width) is $2\pi/N$. Thus, as a crude frequency estimation approach, a local maximum is identified at $k = \ell$, and a simple rounding makes that the estimated frequency is $\ell 2\pi/N$. In this worst case scenario, the maximum absolute estimation error is $0.5 * 100 = 50\%$ of the bin width.

A. The framework

Our framework presumes two equivalent perspectives. On one hand, the multiplication in the discrete domain of $x(n)$ and $h(n)$ corresponds to the convolution in the frequency domain of the Fourier transforms $X(e^{j\omega})$ and $H(e^{j\omega})$:

$$v(n) = x(n)h(n) \xrightarrow{\text{FT}} V(e^{j\omega}) = \frac{1}{2\pi} X(e^{j\omega}) * H(e^{j\omega}). \quad (5)$$

The desired DFT, $V(k)$, is then obtained by sampling $V(e^{j\omega})$ at $\omega = k2\pi/N$ for $k = 0, 1, \dots, N-1$. On the other hand, an equivalent perspective is obtained by computing $X(k)$, the DFT of $x(n)$, and $H(k)$, the DFT of $h(k)$. $V(k)$ results from the circular convolution between $X(k)$ and $H(k)$:

$$v(n) = x(n)h(n) \xrightarrow{\text{DFT}} V(k) = \frac{1}{N} X(k) \otimes H(k). \quad (6)$$

This last perspective is particularly convenient to highlight the fact that all three frequency estimation methods can be obtained using the DFT of the input signal and a subsequent circular convolution with the DFT of each window. This concept is illustrated in Fig. 2. For example, the DFT of the rectangular window is given by $H_R(k) = \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N}kn}$ whose result is N for $k = 0$ and 0 for $k = 1, 2, \dots, N-1$. Taking into consideration that $H_R(k)$ is periodic with period N , this means that $V(k) = \frac{1}{N} X(k) \otimes H_R(k) = X(k)$, a result that would be expected since windowing by the rectangular window does not modify the input signal. In the case of the Hanning window, $H_H(k) = \sum_{n=0}^{N-1} h_H(n)e^{-j\frac{2\pi}{N}kn}$

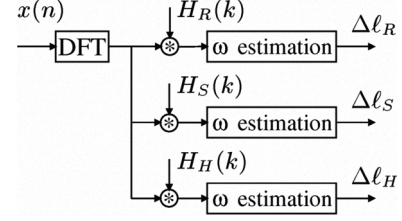


Fig. 2. Accurate frequency estimation using different windows.

which gives rise to only three coefficients different from zero in a period: $N/2$ for $k = 0$, and $-N/4$ for $k = 1$ and $k = N-1$. Thus, the circular convolution is simply obtained by running a 3-coefficient FIR filter on $X(k)$. A similar operation applies in the case of the sine window. Thus, as illustrated in Fig. 2, the basic computational cost of the three different frequency estimation algorithms consists of a single DFT (via an FFT) and two circular convolutions, one of which is quite modest computationally. In the following, each frequency estimation algorithm is derived and explained.

B. Rectangular window model

It can be easily shown that when $x(n) = A \sin(\omega_\ell n + \phi)$ with $\omega_\ell = \frac{2\pi}{N}(\ell + \Delta\ell)$, the sampling in the frequency domain of $V(e^{j\omega})$, the Fourier Transform of $v(n) = x(n)h_R(n)$, at $\omega = k\frac{2\pi}{N}$, $k = 0, 1, \dots, N-1$, leads to two N -periodic complex functions corresponding to two images of the FreqResp of the rectangular window on the positive and negative frequency axis. Taking the absolute value of $|V(k)|$ and ignoring the influence due to the 'negative' image of the FreqResp of the rectangular window, one gets the approximation

$$|V(k)| \approx \frac{A}{2} \left| \frac{\sin(k - \ell - \Delta\ell)\pi}{\sin(k - \ell - \Delta\ell)\pi/N} \right|. \quad (7)$$

As $0.0 \leq \Delta\ell < 1.0$, the spectral lines at $k = \ell$ and $k = \ell + 1$ correspond to local maxima since only two spectral lines fit within the main lobe of the FreqResp of the rectangular window (whose width is $4\pi/N$). Using (7) we can compute

$$\frac{|V(\ell+1)|}{|V(\ell)|} \approx \frac{1}{\sin \frac{\pi}{N} \cot \frac{\pi}{N} \Delta\ell - \cos \frac{\pi}{N}} \quad (8)$$

from which we obtain

$$\Delta\ell \approx \frac{N}{\pi} \arctan \frac{\sin \frac{\pi}{N}}{\cos \frac{\pi}{N} + \frac{|V(\ell)|}{|V(\ell+1)|}}. \quad (9)$$

The accuracy of this estimate depends on ℓ and $\Delta\ell$ and is essentially independent of the magnitude and phase of the sinusoid. Fig. 3 depicts the maximum absolute estimation error relative to the bin width as a function of ℓ , when $\Delta\ell$ varies between 0.0 and 1.0 ($N = 512$). Regarding the rectangular window model (the other two models will be addressed in subsequent sections), while the maximum absolute estimation error can be as low as about 0.125%, this error increases, as expected, when ℓ approaches 0.0 or $N/2$ because of the influence due to the image of the FreqResp of the rectangular window on the negative frequency axis. The relatively slow variation of the error function when ℓ is low, can be explained by the significant leakage of the rectangular window. However, the estimation accuracy can be significantly better than that of other methods such as the parabolic model, as will be shown in section II-D.

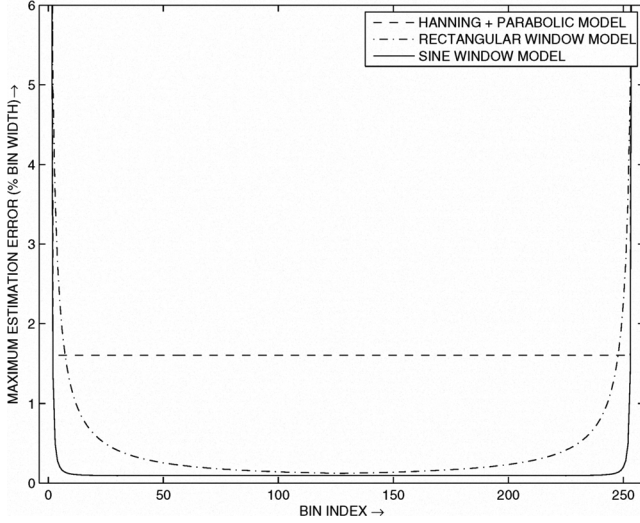


Fig. 3. Maximum estimation error of three frequency estimation methods as a function of the DFT spectral line ($N=512$).

C. Sine window model

The sine window model follows closely the results presented in [6]. However, these results have been derived for the Odd-DFT (whose sampled frequencies are given by $\frac{(2k+1)\pi}{N}$, $k = 0, 1, \dots, N-1$), and therefore have to be adapted to the DFT (whose sampled frequencies are given by $\frac{k2\pi}{N}$, $k = 0, 1, \dots, N-1$). Although different equations result, the same optimization parameters derived in [6] apply. As $x(n) = A \sin(\omega_\ell n + \phi)$, its Fourier Transform results as $X(e^{j\omega}) = -jA\pi\delta(\omega - \omega_\ell) + jA\pi\delta(\omega + \omega_\ell)$. Thus, according to (5), the Fourier Transform of the windowed input results as a modulation of $H_S(e^{j\omega})$ to $\omega = \omega_\ell$ and to $\omega = -\omega_\ell$. As $H_S(e^{j\omega})$ is not easily tractable analytically, the magnitude of its main lobe is approximated by $|H_S(e^{j\omega})| \approx |\cos \frac{N\omega}{6}|^{1/G}$, $|\omega| < 3\pi/N$. As illustrated in Fig. 4, at most three DFT spectral lines fit within the main lobe of the FreqResp of the sine window since its width is $6\pi/N$. Therefore, as illustrated in Fig. 4, two important

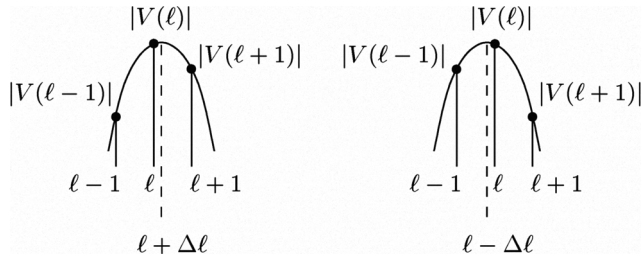


Fig. 4. Accurate estimation of $\Delta\ell$ using $|V(k)|$.

situations may occur: either $|V(\ell+1)| > |V(\ell-1)|$, which denotes that the exact frequency may be estimated as $\ell + \Delta\ell$; or $|V(\ell+1)| < |V(\ell-1)|$ in which case the exact frequency may be estimated as $\ell - \Delta\ell$. The estimation of $\Delta\ell$ must therefore deliver a number between 0.0 and 0.5. In the first case illustrated in Fig. 4, when $0.0 \leq \Delta\ell \leq \gamma$, improved estimation accuracy is obtained by using $Q = \frac{|V(\ell-1)|}{|V(\ell+1)|} = \frac{|H_R(\frac{2\pi}{N}(\Delta\ell+1))|}{|H_R(\frac{2\pi}{N}(1-\Delta\ell))|} \approx \left[\frac{\cos \frac{\pi}{3}(\Delta\ell+1)}{\cos \frac{\pi}{3}(1-\Delta\ell)} \right]^{\frac{1}{G}}$, from which we obtain

$$\Delta\ell \approx \frac{3}{\pi} \arctan \frac{1}{\sqrt{3}} \frac{1-Q^G}{1+Q^G}. \quad (10)$$

When $\gamma \leq \Delta\ell \leq 0.5$, improved estimation accuracy is obtained by using $S = \frac{|V(\ell+1)|}{|V(\ell)|} = \frac{|H_R(\frac{2\pi}{N}(1-\Delta\ell))|}{|H_R(\frac{2\pi}{N}\Delta\ell)|} \approx \left[\frac{\cos \frac{\pi}{3}(1-\Delta\ell)}{\cos \frac{\pi}{3}\Delta\ell} \right]^{1/F}$, from which we obtain

$$\Delta\ell \approx \frac{3}{\pi} \arctan \frac{2S^F - 1}{\sqrt{3}}. \quad (11)$$

The optimization of the estimation error in the *minmax* sense is performed as in [6] which specifies γ , G and F . In the second case illustrated in Fig. 4, the same expressions are obtained after redefining $Q = \frac{|V(\ell+1)|}{|V(\ell-1)|}$ and $S = \frac{|V(\ell-1)|}{|V(\ell)|}$. Fig. 3 depicts the maximum absolute estimation error relative to the bin width as a function of ℓ , when $\Delta\ell$ varies between 0.0 and 1.0 ($N = 512$). As in [6], the estimation error is consistently below 0.1% of the bin width. It can be noticed that when ℓ approaches 0 or $N/2$, the mutual influence between the two images of the FreqResp of the sine window (due to the two Diracs in the frequency domain and pertaining to the sinusoidal input), is quite minimal, which indicates that frequency estimates are quite reliable even when two sinusoids are only three DFT frequency lines apart (or $6\pi/N$). Other methods require at least 6 or 18 bins [1].

D. Hanning window and parabolic model

The parabolic model involves a parabola fitting the main lobe of the FreqResp of the analysis window on the logarithmic scale. As the top of the main lobe has a convex shape, a convenient model on a X-Y plane is given by $y = y_0 - m(x - x_0)^2$, where the unknown parameters are the horizontal shift (x_0), the vertical shift (y_0), and the convexity factor (m). Thus, three equations are required. As Fig. 4 suggests, $y_0 = A_{dB}$ is the unknown magnitude of the sinusoid, and $x_0 = \ell + \Delta\ell$ of which only $\Delta\ell$ is unknown. Three equations are thus readily obtained as $A_{dB}(\ell - 1) = A_{dB} - m(-\Delta\ell - 1)^2$, $A_{dB}(\ell) = A_{dB} - m(-\Delta\ell)^2$, and $A_{dB}(\ell + 1) = A_{dB} - m(1 - \Delta\ell)^2$. Solving for $\Delta\ell$ one obtains

$$\Delta\ell = \frac{A_{dB}(\ell + 1) - A_{dB}(\ell - 1)}{4A_{dB}(\ell) - 2A_{dB}(\ell + 1) - 2A_{dB}(\ell - 1)}. \quad (12)$$

The parabolic model requires that for all possible values of $\Delta\ell$, three DFT lines fit inside the main lobe of the FreqResp of the analysis window. Therefore, this implies that the parabolic model is not valid if used with the rectangular or the sine window. When the parabolic model is used with the Hanning window (4), the maximum absolute estimation error (relative to the bin width) as a function of ℓ , when $\Delta\ell$ varies between 0.0 and 1.0, results as depicted in Fig. 3 ($N = 512$). It can be seen that the best estimation error is about 1.6% of the bin width, which is significantly below the best performance of the rectangular or the sine window models. This fact is not surprising since the parabolic model is window agnostic. The estimation error due to the parabolic model is however quite stable even when ℓ approaches 0 or $N/2 - 1$. Reliable estimation requires that the frequency separation between two sinusoids is at least 4 DFT frequency lines (or $8\pi/N$).

E. Robustness to noise

In order to evaluate the relative noise robustness of the three frequency estimation models, additive Gaussian noise is added to the input sinusoid as in (1). The SNR varies between 0 dB and 70 dB in steps of 5 dB. For each SNR value and model, the average is computed of the maximum estimation error when $0 < \ell < N/2 - 1$. The results are depicted in Fig. 5. It can be seen that the rectangular model and the sine model perform always better than the parabolic model. On the other hand, and perhaps surprisingly, the rectangular model performs better than the sine model for SNRs below about 20

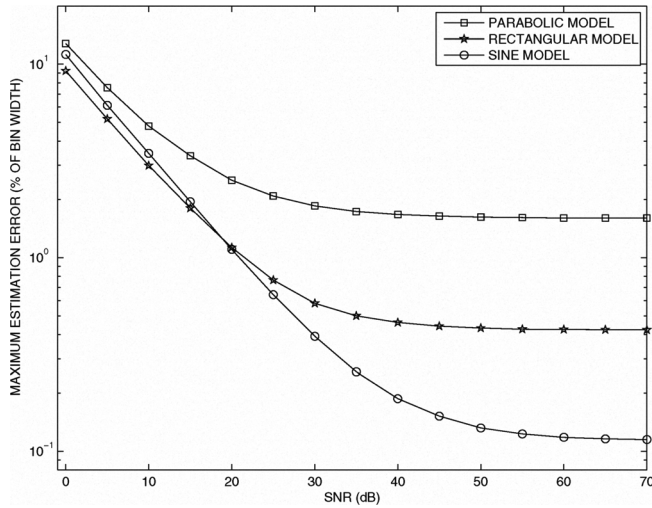


Fig. 5. Relative performance of the three estimation models.

dB. A possible explanation supporting this fact is that a low SNRs, the approximation model that has been used for the main lobe of the FreqResp of the sine window, and that has been optimized in the absence of noise, introduces more uncertainty than the exact model that has been derived for the main lobe of the FreqResp of the rectangular window.

III. COMBINED RECTANGULAR-SINE WINDOW APPROACH

In order to benefit from the good performance of the sine window model at high SNRs, and from the good performance of the rectangular window model at low SNRs, a combined rectangular-sine window model is devised by computing frequency estimates using both models. If the estimation error due to the sine window model is $|\Delta\ell - \Delta\ell_S| < err_S$, and that due to the rectangular window model is $|\Delta\ell - \Delta\ell_R| < err_R$, then $|\Delta\ell_S - \Delta\ell_R| < err_S + err_R$. Therefore, by evaluating the difference between the frequency estimates, a simple test of consistency can be obtained as

$$|\Delta\ell_S - \Delta\ell_R| < err. \quad (13)$$

If the test is met, both estimates are considered consistent and the final frequency estimate is that due to the sine window model. When the test is not met, the estimates are considered inconsistent and the final frequency estimate is obtained by taking the average of both estimates. Simulation results have been carried out and a suitable value of $err = 0.01$ (or 1%) has been found. The performance of the combined model is depicted in Fig. 6. It can be seen that while for high SNRs the combined model performs quite similarly to the sine model, for low SNR the combined model performs even slightly better than the rectangular model. These results confirm the pertinence of the approach represented in Fig. 2 because gains are obtained, particular at low SNR, by combining accurate frequency estimates that are computed independently.

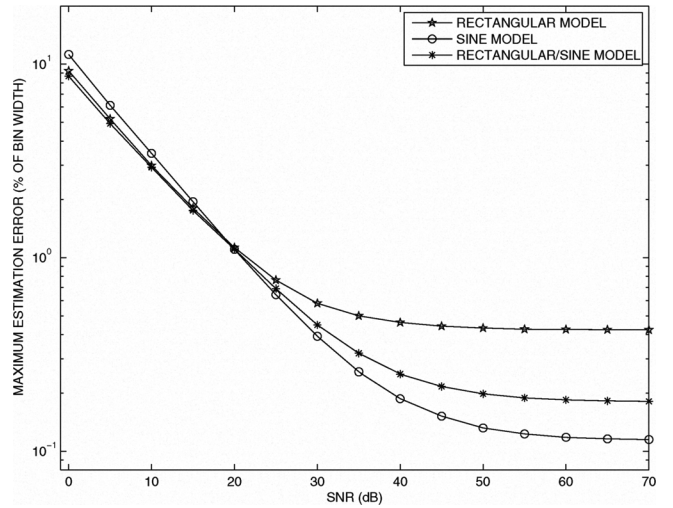


Fig. 6. Estimation performance of the combined model.

IV. CONCLUSION

In this paper we have derived and evaluated two frequency estimation methods that rely on a single DFT magnitude spectrum, a circular convolution operation, and a simple non-iterative frequency estimation algorithm. These methods are adapted to analysis windows offering a different trade-off of selectivity and leakage, and allowing to reduce the estimation error to less than 0.1% of the DFT bin width. The influence of noise in the estimation process has been evaluated and an estimation procedure combining two estimation methods has been proposed that is more robust to noise. The results presented in this paper are particularly suitable for real-time applications such as visual-feedback of the singing voice where detailed features of fast events (*e.g.*, *vibrato*) have to be accurately captured.

REFERENCES

- [1] J. Schoukens, R. Pintelon, and H. Van Hamme, "The interpolated fast Fourier transform: a comparative study," *IEEE Transactions on Instrumentation and Measurement*, vol. 41, no. 2, pp. 226–232, April 1992.
- [2] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *5th Int. Conf. on Digital Audio Effects (DAFx-02)*, 2002, pp. 51–58.
- [3] Anibal J. S. Ferreira and Ricardo Sousa, "DFT-based frequency estimation under harmonic interference," in *ISCCP 2010*, 2010, submitted for publication.
- [4] H. Renders, J. Schoukens, and G. Vilain, "High-accuracy spectrum analysis of sampled discrete frequency signals by analytical leakage compensation," *IEEE Transactions on Instrumentation and Measurement*, vol. 33, no. 4, pp. 287–292, December 1984.
- [5] Anssi Klapuri and Manuel Davy, Eds., *Signal Processing Methods for Music Transcription*, Springer, 2006, Part III: multiple fundamental frequency analysis.
- [6] Anibal Ferreira and Deepen Sinha, "Accurate and robust frequency estimation in the ODFT domain," in *2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2005, pp. 203–206.