Frequency Estimation Performance of Several MDCT Domain Algorithms

Yujie Dun, Guizhong Liu
School of Electronic and Information Engineering,
Xi'an Jiaotong University, Xi'an, China, 710049
dunyj@mail.xjtu.edu.cn, liugz@mail.xjtu.edu.cn

ABSTRACT
The frequency estimation performance of several modified discrete cosine transform (MDCT) domain algorithms is investigated for single real sinusoidal signal in white Gaussian noise. Two discrete Fourier transform (DFT) domain algorithms are included to make comparison. It is demonstrated that most of the MDCT domain algorithms performs alike when the signal-to-noise ratio (SNR) is low. Only at sufficiently high SNR, these algorithms show their own characteristics. But compared to the DFT algorithms and the Cramer-Rao bound (CRB), there should be opportunity to develop a better estimator in the MDCT domain.

Categories and Subject Descriptors
H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing – *signal analysis, synthesis, and processing.*

General Terms
Algorithms.

Keywords
Frequency estimation, Modified discrete cosine transform, audio processing

1. INTRODUCTION
Frequency estimation is a fundamental problem in signal processing and has many applications in science and engineering, including spectrum estimation, array signal processing, radar signal processing, speech and audio processing and communications. There are many algorithms that have been proposed to solve this problem [1]-[5]. For audio signals that compressed with the modified discrete cosine transform (MDCT) [6], several algorithms that operate directly with the MDCT coefficients have been proposed during the last decade [7]-[9]. These algorithms are raised primarily to reduce the complexity that encountered when a traditional frequency estimation algorithm is applied to a compressed audio.

But frequency estimation in the MDCT domain is far more complex than the one in traditional DFT domain. In the DFT domain, a two step-estimation procedure, including a coarse search of the DFT magnitude peak to estimate the integer part of the digital frequency, and a fine search with various methods to find a frequency offset, is rather straightforward. But in the MDCT domain, the magnitude peak of the coefficients may varies due to the influence of the initial phase and cannot be used directly as the estimation value of the integer part. A more complicate analytical expression of the MDCT coefficient for a monophonic sinusoidal signal makes the estimation of the frequency offset an even harder task. We have found only three articles that report the frequency estimation methods directly operate with the MDCT coefficients. Merdjani [7] estimated the frequency with the Pseudo-spectrum [10] and the magnitude ratios. Zhu [8] proposed a simplified version with the knowledge of the ratio range. Zhang [9] extended such magnitude ratio based method to a more precise model, he also proposed a phase factor based method, and an iterative method to get precise estimation without noise.

Till now, the performance of these MDCT domain estimation algorithms under noisy condition, which is a common case in the applications, has not been reported yet. In this paper, we give the performance comparison of these algorithms, and demonstrate the behavior of these methods. We add two DFT domain algorithms [3],[5] and the Cramer-Rao bound(CRB) [1] as the references also, and show the results in the following part of this paper.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICIMCS'14, July 10–12, 2014, Xiamen, Fujian, China.
Copyright 2014 ACM 978-1-4503-2810-4/14/07 …$15.00.

2. DESCRIPTION OF THE ALGORITHMS
Here we will give the description of the frequency estimation algorithms that operate directly with the MDCT coefficients. At first the MDCT expression of a monophonic sinusoidal signal is given. Then the algorithms are described briefly.

2.1 MDCT ANALYSIS
For a monophonic sine wave signal with amplitude A, digital frequency f_n ,initial phase ϕ, and analysis length of $2N$,

$$x(n) = A\sin\left[\frac{2\pi f_n}{2N}n + \phi\right], \tag{1}$$

its MDCT transform is defined as

$$X(k) = \sqrt{\frac{2}{N}}\sum_{n=0}^{2N-1} x(n)h(n)\cos\left[\frac{\pi}{N}(n+\frac{1}{2}+\frac{N}{2})(k+\frac{1}{2})\right], \tag{2}$$

where $k = 0,1,\cdots,N-1$, $h(n)$ is the window used in the MDCT transform. Here we discuss the sine window case,

$$h(n) = \sin[\frac{\pi}{2N}(n+\frac{1}{2})], \quad n = 0,1,\cdots,2N-1. \tag{3}$$

The simplified analytical expression of the coefficient $X(k)$ is

$$\begin{aligned} X(k) = A\sqrt{N/2}\cdot\big[&M(k-f_n)\cdot\cos(\varphi-3\pi k/2) \\ &+(-1)^k M(-k-f_n-1)\cdot\sin(\varphi-3\pi k/2)\big], \end{aligned} \tag{4}$$

where $M(\xi)$ is the amplitude term

$$M(\xi)=\frac{\sin(\pi\xi)}{2N}\cdot\left(\frac{1}{\sin(\frac{\pi}{2N}\xi)}-\frac{1}{\sin(\frac{\pi}{2N}(\xi+1))}\right), \quad (5)$$

and φ is the modified phase term defined as

$$\varphi=\frac{2N-1}{2N}\pi f_n-\frac{5\pi}{4}+\phi. \quad (6)$$

The expression (4) and its simplified version of omitting the second part are the basis of the estimators proposed by Zhang [9]. In the meanwhile, a further approximation to (5) results the version used in [7],[8] as

$$X(k)=\frac{A\sqrt{2N}}{2\pi}\cdot\frac{\sin[\pi(k-f_n)]}{(k-f_n)(k-f_n+1)}\cos(\varphi-\frac{3\pi}{2}k). \quad (7)$$

2.2 ESTIMATION ALGORITHMS

Five frequency estimation algorithms presented in three articles [7]-[9] will be described here and be compared in the following section.

- Merdjani, Pseudo-Spectrum method [7],
- Zhu, rational approximation method [8],
- Zhang-Env, envelope function method [9],
- Zhang-Phs, phase factor method [9],
- Zhang-Itr, iterative method [9],

2.2.1 Merdjani, Pseudo-Spectrum method

The Pseudo-spectrum method splits the estimation to two parts, the integer part $f_0=\lfloor f_n\rfloor$ and the fractional part $\varepsilon=f_n-f_0$. The integer part f_0 is estimated with the peak index of the Pseudo-spectrum [10],

$$S(k)=\sqrt{X^2(k)+[X(k-1)-X(k+1)]^2}, \quad (8)$$

$$\hat{f}_0=\arg\max_k(S(k)). \quad (9)$$

The fractional part ε is estimated with the formula

$$\hat{\varepsilon}=\frac{3+\alpha-\sqrt{\alpha^2+14\alpha+1}}{2(1-\alpha)}, (\hat{\varepsilon}=0.5 \textit{ for } \alpha=1), \quad (10)$$

where α is the MDCT coefficient ratio,

$$\alpha=-X(\hat{f}_0-1)/X(\hat{f}_0+1). \quad (11)$$

A substitution formula together with a ratio β is used in case that $X(f_0+1)$ and $X(f_0-1)$ are both very small,

$$\hat{\varepsilon}=\frac{5+3\beta-\sqrt{\beta^2+62\beta+1}}{2(1-\beta)}, (\hat{\varepsilon}=0.5 \textit{ for } \beta=1) \quad (12)$$

$$\beta=X(\hat{f}_0-2)/X(\hat{f}_0+2). \quad (13)$$

2.2.2 Zhu, rational approximation method

The rational approximation method avoids the calculation of the Pseudo-Spectrum. This method makes a coarse estimate of the integer part first,

$$\hat{f}_0=\arg\max_k(abs(X(k))), \quad (14)$$

adjusts it according to the value range of α calculated via (11). Then the steps to estimate ε used in Merdjani's Pseudo-spectrum method are adopted. A complementary method is introduced to handle the case that α is not in the pre-defined ranges because the range is set according to the approximated formula as (7).

2.2.3 Zhang-Env, envelope function method

Zhang's envelope function method finds the round value of f_n, $p=\lfloor f_n+0.5\rfloor$ first, and calculate two ratios,

$$I_0=-X(p-2)/X(p),\ \ I_1=-X(p+1)/X(p-1). \quad (15)$$

One of the ratios should fall into the value range of a ratio, $R(\xi)$,

$$R(\xi)=M(\xi-1)/M(\xi+1),\ \ \xi\in(-1,-0.5]. \quad (16)$$

The values of $R(\xi)$ are calculated with $M(\xi)$ in form of (5) and stored in a look-up table. The estimation of the fractional part $\hat{\varepsilon}$ is obtained by searching the look-up table and interpolating. Here, no consideration on the very small MDCT bin values is made.

2.2.4 Zhang-Phs, phase factor Method

Zhang's phase factor method finds the round value p of f_n first. The fractional part $\varepsilon=f_n-p$ is estimated with the MDCT coefficients of two successive frames knowing the adjacent-frame has a phase increase factor of πf_n,

$$\sin(\pi f_n)=(U-V)\operatorname{sgn}(\sin\varphi_1\cos\varphi_1)\sqrt{\frac{1-V^2}{U^2-V^2}}\sqrt{\frac{1-U^2}{V^2-U^2}}, \quad (17)$$

where φ_1 and φ_2 are the phase factors as defined in (6), U and V are the phase ratios of the odd and even bin maximum positions, p_o and p_e (one of them is equal to p), respectively,

$$U=X_1(p_0)/X_0(p_0),\ V=X_1(p_e)/X_0(p_e), \quad (18)$$

where the $X_0(\bullet)$ and $X_1(\bullet)$ refer to coefficients of the former and the later frame, respectively.

2.2.5 Zhang-Itr, Iterative Method

For accurate estimation of the frequency, the precise model in (4) is used. But direct solution is hard to be obtained because this model is very complex. An iterative method is proposed to resolve this problem. The iterative method builds the equations related the normalized frequency f_n, the phase factor φ_n and the MDCT coefficients of bins near f_n. With an initial value $\tilde{f}_n$, the Newton-Raphson method is used to solve the equations.

In this method, a proper initial value of the digital frequency is needed to make the result converge quickly to the required accuracy. The envelope function method and the phase factor method can all be used to obtain this initial value. Thus the complexity of the iterative method is inevitable higher than others.

2.3 REFERENCE ALGORITHMS AND CRB

The Macleod algorithm and its unbiased counterpart are introduced also in our test as the comparative methods in the DFT domain.

- Macleod-biased, a linearized estimator [3],
- Macleod-unbiased, a nonlinear estimator [5].

The Cramer-Rao bound (CRB) for the frequency estimation of a complex single tone sinusoidal in white Gaussian noise(WGN) is

$$CRB=\frac{6f_s^2}{4\pi^2N(N^2-1)\cdot\mathrm{SNR}}, \quad (19)$$

where f_s is the sampling frequency, N is the sample number, SNR is the signal-to-noise ratio.

The formula (19) is valid for a complex sinusoidal in noise. For a real sinusoidal, the CRB is approximately two times of this value if the frequency (in hertz) is greater than about f_s / N.

3. EXPERIMENTAL RESULTS

In this part, we present the results of different tests. The tests have been designed to compare the performance of these MDCT domain algorithms according to the specific characteristics of the frequency estimation in MDCT domain for audio coding and processing. For all experiments, we used the following parameters, $A = 1$, $f_s = 44{,}100$ Hz, $N = 1024$. Other parameters were set differently in the specific test. We generated the complex sinusoidal signals with noise at first, then fed them to the DFT domain algorithms and their real part to the MDCT domain algorithms. For each signal, we set its length to cover 100 consecutive MDCT frames. For each test, we repeated the experiment 100 times. The $\Delta\xi$ used to build the look-up table with (16) is set to 2^{-11} (used in Zhang-Env).

3.1 Test 1 – varied frequencies without noise

The first test is designed to compare the estimation errors of the five algorithms without the present of noise. We must take the model error into account in this test. Excepting Zhang-Itr, the iterative method, the other four algorithms are derived from a simplified version by omitting the second term in (4) or further approximation after such omission, the closer the f_n is to 0 or N, the bigger the model error caused by this "second term" is. So we test two typical values of the integer part of f_n, 10 as a value close to 0, and 512 as the value far from 0 or N. We vary the fractional part from 0.05 to 0.95 with step size 0.05.

The results are given in Figure 1. The MSE of the Zhang-Itr method is below 10^{-20} in both cases, so we show the curves of the other four methods. One can observe that the estimators perform obviously better with f_n near 512. For both case, the estimators perform differently with different fractional part of the frequency. But in average, Zhang-Env is the worst. This is because this algorithm does not give an alternative manner to handle the too-small MDCT coefficient(s) that may be involved in the process. Merdjani and Zhu are almost the same because they use similar manner to do the estimate. The performance of the phase method varies greatly with the change of the fractional part.

3.2 Test 2 – varied frequencies with noise

This test is designed to investigate the estimation errors of the five MDCT domain algorithms under noisy condition. SNR is set to 80dB in this test. Here we use only 512 as the integer part of f_n. We again vary the fractional part from 0.05 to 0.95 with step size 0.05. The result is shown in Figure 2. The CRB for the frequency estimation of real single tone sinusoidal is also given (CRB-R, the bold black line). From the result we can see that, even at SNR of 80dB, none of the MSE of these algorithms is close to CRB. Again, a sophisticated MSE distribution is observed. With different fractional part of f_n, the relationship of these algorithms' MSE values is different, too. Zhang-Env still has the worst MSE scores. The Merdjani and Zhu still keep the similar performance in this test. But for the Zhang-Itr, Zhang-Phs and Merdjani(Zhu), nearly all possible relations are exhibited when the fractional part change from 0.05 to 0.95. So, the relation of these algorithms' MSEs presented in the following test has nearly no meaning. That is to say, we cannot conclude from the curves that one is better than another.

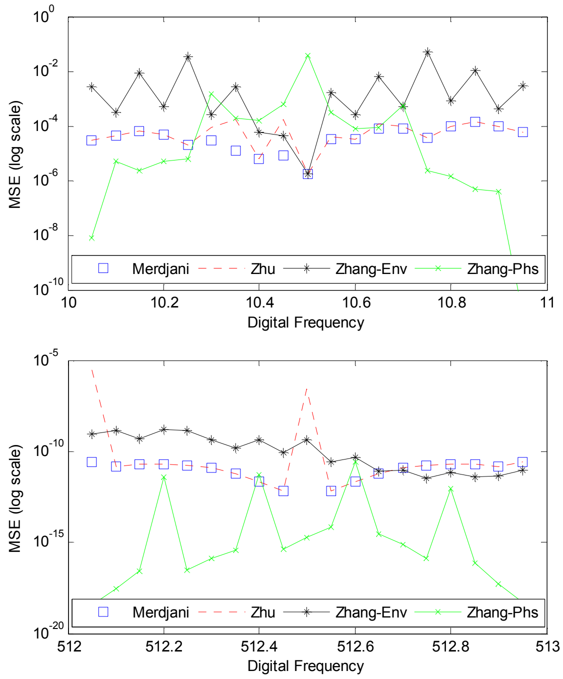

Figure 1. Comparison of MSE vs. f_n without noise.

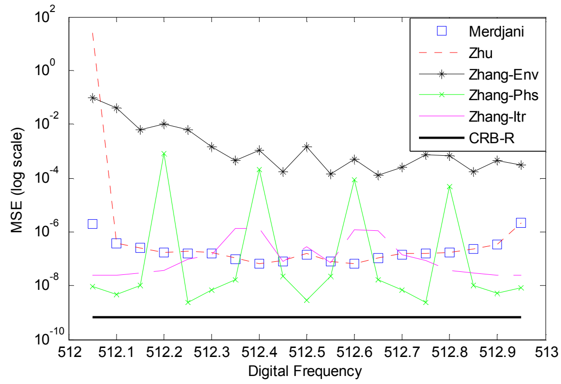

Figure 2. Comparison of MSE vs. f_n at SNR=80dB.

3.3 Test 3 – fixed frequency at different SNRs

This test is designed to investigate the MSE vary trends of these algorithms when the SNR increases from low to high. In this test, the two DFT domain frequency estimation algorithms are involved with their corresponding CRB (CRB-Z, the dashed bold black line). We select several f_n values, 10.15, 255.2 and 512.6. These values are selected with different integer part of the digital frequencies (one close to 0, one at the middle place and one far from either 0 or N) and different fractional values (we know from section 3.2 that, different fractional values may bring different relationship among the test algorithms). The results are given in Figure 3.

From the plot one can observe that, most of estimators follow the same trend, their MSEs and SNR have a linear relationship in log space similar to the CRB. When SNR is relatively high, some of

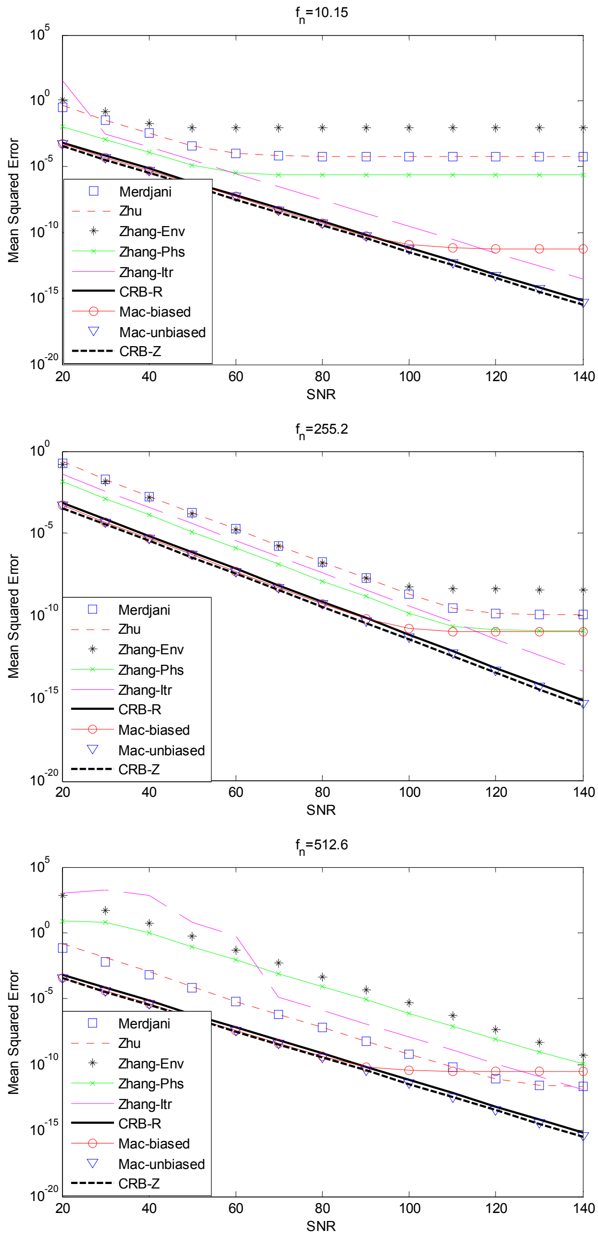

Figure 3. Comparison of MSE vs. SNR

them no longer follow the CRB because of the model error. The Zhang-Itr, the iterative method, although exhibits excellent MSE score in test 1, is very sensitive to noise at low SNR, and the estimation result may have great bias (indicates a wrong result). The performance of Zhang-Phs has a constant performance for a fixed fractional value when the SNR varies. So its MSE depends on the fractional value. None of the MSE curves of these MDCT domain frequency estimators is close to the CRB as the MSE curves of the DFT domain frequency estimators do.

4. CONCLUSIONS

The frequency estimation performance of several MDCT domain algorithms has been investigated for real sinusoids in noise, and compared with the CRB and two DFT domain algorithms with a complex sinusoids counterpart. It has been demonstrated that the MDCT domain approaches have the similar appearance as the DFT domain estimators, but their lines are a little far from the CRB, which means that these estimators are not the optimized ones. Potential improvements can be made in the future work. At the same time, the consistent behavior under noise condition implies that there is no obvious difference among these MDCT domain algorithms when used in real audio signal processing, because the noise is inevitable. The ideal no-noise performance difference may not be exhibited in practice.

It is worth to be noticed that, although the iterative algorithm is the most complex one, it has no advantage compared to other low complexity implementations under noise condition. The study for a new MDCT domain algorithm that can estimate the frequency with higher precision under noisy condition is worthwhile.

5. REFERENCES

[1] Rife, D. C., and Boorstyn, R. R. 1974. Single tone parameter estimation from discrete-time observations. *Information Theory, IEEE Trans. on*, 20, 5 (Sep. 1974), 591-598.

[2] Quinn, B.G. 1994. Estimating frequency by interpolation using Fourier coefficients. *Signal Processing, IEEE Trans. on* , 42, 5 (May 1994), 1264-1268,

[3] Macleod, M. D. 1998. Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones. *Signal Processing, IEEE Trans. on*, 46, 1 (Jan. 1998), 141-148.

[4] Jacobsen, E., and Kootsookos, P. 2007. Fast, accurate frequency estimators. *IEEE Signal Process. Mag*, 24, 3, (May 2007), 123-125.

[5] Liao, J, and Lo, S. 2014. Analytical solutions for frequency estimators by interpolation of DFT coefficients. *Signal Process.* Vol.100 (Jul. 2014), 93-100.

[6] Princen, J., and Bradley, A. 1986. Analysis/synthesis filter bank design based on time domain aliasing cancellation. *Acoustics, Speech and Signal Processing, IEEE Trans. on*, 34, 5 (Oct. 1986), 1153-1161.

[7] Merdjani, S., and Daudet, L. 2003. Direct estimation of frequency from MDCT-encoded files. *In Proceedings of the 6th International Conference on Digital Audio Effects*. (London, UK, September 8-11, 2003)

[8] Zhu, M. Y., Zheng, W., Li, D. X., and Zhang, M. 2008. An accurate low complexity algorithm for frequency estimation in MDCT domain. *Consumer Electronics, IEEE Transactions on*, 54, 3 (Aug. 2008), 1022-1028.

[9] Zhang, S., Dou, W., and Yang, H. 2013. MDCT Sinusoidal Analysis for Audio Signals Analysis and Processing. *Audio, Speech, Language Processing .IEEE Trans. on*, 21, 7 (Jul. 2013), 1403-1414

[10] Daudet, L., and Sandler, M. 2004. MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction. *Speech and Audio Processing, IEEE Trans. on*, 12, 3 (May, 2004), 302-312.