AN IMPROVED MDCT DOMAIN FREQUENCY ESTIMATION METHOD

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

ABSTRACT

In this work, an improved low complexity high precision method for frequency estimation in the Modified Discrete Cosine Transform (MDCT) domain is presented. The method is derived from the analytical expression of the MDCT coefficients of a sinusoid under symmetric windows. With this expression, frequency estimation under the sine and the Kaiser-Bessel Derived (KBD) windows can be made. This paper presents different procedures of the estimation under the sine and KBD windows according to their characteristics. The presented procedures have the ability to avoid estimation mistakes and large estimation errors. The method is tested to show great enhancement of the estimation precision when compared with the traditional methods, and can be used in compressed domain audio processing.

***Index Terms*—** frequency estimation, MDCT, audio coding

1. INTRODUCTION

Frequency estimation is a fundamental problem in signal processing, which is the basis of many applications that depend on the detected frequencies. In audio signal processing, accurate frequency estimation, especially for the tonal components, is very important. But neither the time-domain methods nor the frequency-domain methods are efficient because audio signals are usually stored and transmitted in compressed format instead of the time-domain samples or the DFT coefficients. Extra transform is required before the estimation. Most of the audio codecs such as MPEG 2/4 AAC, Dolby AC-3 and WMP perform the compression based on the MDCT (Modified Discrete Cosine Transform) coefficients [1]. Frequency estimation is not easy to do directly in the MDCT domain due to the phase-related feature of this transform. An Odd-DFT domain method[2] has been proposed for audio systems, but the conversion complexity is still high. Based on a simplified MDCT coefficient model[3], new low complexity methods [4][5] are proposed, but they are only fit for the estimation under the sine window. Recently, two low complexity frequency estimation methods which can be used under arbitrary window functions are presented in [6], but they are shown to have big estimation errors.

In this paper, we present an improved method to do the frequency estimation while keeping the similar complexity with [4]-[6]. The improvement is made especially for the estimation when the sine window or the Kaiser-Bessel Derived (KBD) window is used in the MDCT. The proposed method is shown to lower the estimation error effectively. The remainder of this paper is organized as follows: In the next section, Section 2, the expression of the MDCT coefficient under arbitrary window is given, after which the problem to be solved is stated. In Section 3, the improved estimation under the sine and KBD windows is presented. Then, in Section 4 some experimental results are shown. Section 5 concludes on the work.

2. PROBLEM STATEMENT

Single tonal component in an audio signal can be modeled as a sinusoid. Our frequency estimation method is based on the analytical expression of the MDCT coefficient of a sinusoidal signal.

For a sinusoid with amplitude A, digital frequency f_n and initial phase ϕ,

$$x(n) = A\sin(\frac{\pi f_n}{N} n + \phi),\ n = 0, 1, \cdots, 2N-1, \quad (1)$$

the MDCT transform is defined as

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

where N is the MDCT bin number, $k = 0, 1, \cdots, N-1$, $h(n)$ is the window function used in the MDCT transform [7]. When the window $h(n)$ is symmetric, (2) can be expressed as [6]

$$\begin{aligned} X(k) = {} & A\sqrt{N/2}\cos(\varphi_0 - \frac{3\pi}{2}k)\, A_H(k - f_n) \\ & + (-1)^k A\sqrt{N/2}\sin(\varphi_0 - \frac{3\pi}{2}k) A_H(-k - f_n - 1), \end{aligned} \quad (3)$$

where the phase factor φ_0 is

$$\varphi_0 = \frac{2N-1}{2N}\pi f_n - \frac{5}{4}\pi + \phi, \quad (4)$$

$A_H(\xi)$ is the real part of the $h(n)$'s Centered Discrete Fourier Transform (CDFT),

$$A_H(\ ) = \mathrm{Re}\{H(\ )\}$$
$$= \mathrm{Re}\left\{\frac{1}{N}\sum_{n=0}^{2N\ 1} h(n)e^{\ j\frac{2}{2N}(n+\ 0.5\ \ N)(+\ 0.5)}\right\}. \quad (5)$$

The expression given in (3) is the exact result of the MDCT coefficient for a given signal $x(n)$ under arbitrary symmetric window $h(n)$, but it is a little complicate to get a close form frequency estimator. Now we will simplify this expression to derive a simple form estimator.

Considering the fast fading characteristics of the window function, the second term in (3) can be omitted because $A_H(\ k\ \ f_n\ \ 1)$ is very small when f_n is far from 0 or N ($A_H(\xi)$ has significant value only near $\ = 2l\ N$, l is integer). (3) is simplified to

$$X(k) = A\sqrt{N/2}\ \cos(\ _0\ \ \frac{3}{2}k)\ A_H(k\ \ f_n). \quad (6)$$

Factor $\cos(\ _0\ \ 3\ \ k/2)$ has a period of 4 along with the bin index k as

$k =$ 0, 1, 2, 3, 4, 5, $\cdots$

cos , sin , cos , sin , cos +, sin , $\cdots$

Using the ratio of $X(k)$ with bin interval of 2 or 4, a simple relationship between the MDCT coefficients and f_n can be obtained as

$$= \frac{X(k\ \ 1)}{X(k+1)} = \frac{A_H(k\ \ 1\ \ f_n)}{A_H(k+1\ \ f_n)}, \quad (7)$$

$$= \frac{X(k\ \ 2)}{X(k+2)} = \frac{A_H(k\ \ 2\ \ f_n)}{A_H(k+2\ \ f_n)}. \quad (8)$$

Methods in [4]-[6] are all derived from such coefficient ratio or their combination. References [4] and [5] give an accurate expression between f_n and , in the sine window case with further approximation to $A_H(\xi)$. In [6], to adapt the method with other symmetric windows in the transform, a look-up table is used to express the relationship between f_n and the ratio α according to the formula

$$R(\ )\Big|_{\ =n} = \frac{A_H(n\ \ \ \ 1)}{A_H(n\ \ +\ 1)}, \quad = (k\ \ f_n). \quad (9)$$

In practice, the approximation from (3) to (6) may lead to estimation failure. When the value of $X(k)$ is big enough, the omission error can be ignored. But when $X(k)$ is too small due to the phase-modulation factor $\cos(\ _0\ \ 3\ \ k/2)$ or the side-lobe attenuation, the omission error is big and will lead to large estimation errors or even mistakes. Methods to avoid the very small bin values are required to improve the performance.

It is worth pointing out that, parameters such as amplitude and phase can also be estimated with (6) [4][5] after the frequency is obtained. Accurate frequency estimation is the most important. Our goal is to get a more accurate frequency estimator by proper selection of the involved MDCT bins and proper combination of the ratios.

3. PROPOSED ESTIMATION METHOD

Now let p be the closest integer to the digital frequency f_n, define two MDCT bin ratios,

$$I_0 = \frac{X(p\ \ 2)}{X(p)}, \quad I_1 = \frac{X(p+1)}{X(p-1)}. \quad (10)$$

For f_n varies from $p-0.5$ to $p+0.5$, I_0 corresponds to the envelope ratio $R(\xi)$ when ξ changes from -0.5 to -1.5, while I_1 corresponds to $1/R(\xi)$ when ξ changes from 0.5 to -0.5. Because $A_H(\xi)$ is even-symmetric along $\xi =\ \ 0.5$, I_0 and I_1 seems to be duplicated. But they have different phase-modulation factors, thus if one encounters very small bin values, the other is more reliable. So, to avoid large model error, unlike the memory saving manner used in [6], which only records $R(\xi)$ when $\xi\ \ [\ 1,\ \ 0.5]$, we record double number of the values with $\xi\ \ [\ 1.5,\ \ 0.5]$.

But unluckily, $R(\xi)$ is not monotonic in $(\ 1.5,\ \ 1.25)$ when the sine window is used, thus we define a new table

$$R(\ )\Big|_{\ =n} = \frac{A_H(n\ \ \ \ 2)}{A_H(n\ \ +\ 2)}, \qquad [\ 1.5,\ \ 0.5]. \quad (11)$$

This table is monotonic and is a supplementary method to avoid those very small bin values under the sine window.

Now we have had more than one formula to estimate the digital frequency f_n. After systematically investigation on their performance, we get our final approach to do the estimation with selecting formulas carefully as the following.

(1) Locate the maxima of the MDCT coefficient $|X(k)|$, denote the position as p and the maximum value as $X_{\max}$.

(2) If $|X(p+1)\ \ X(p\ \ 1)|\ \ a_0\ \ X_{\max}$, where a_0 is used to examine the too small phase-modulation factor that makes current $X(p\pm1)$ fail to be used for any decision on the following process, go to (7) and enter a special process.

(3) If $|X(p+1)| > |X(p\ \ 1)|$, update p with $p = p+1$.

(4) Calculate I_0 and I_1.

If $I_0 > I_1$, use condition

$$= \left|\frac{X(p\ \ 1)}{X(p\ \ 2)\ \ X(p)}\right|. \quad (12)$$

Otherwise, use

$$= \left|\frac{X(p)}{X(p\ \ 1)\ \ X(p+\ 1)}\right|. \quad (13)$$

(5) For the KBD window, get I according to Table I, the estimation result is obtained via

$$\tilde{f}_n = \begin{cases} p\ \ 1\ \ R^{\ 1}(I), & if\ I=\ I_0 \\ p+1+R^{\ 1}(I), & if\ I = I_1 \end{cases}. \quad (14)$$

(6) For the sine window, get I according to Table I, the estimation result is obtained via

$$\tilde{f}_n = \begin{cases} p\ \ 1\ \ R^{\ 1}(I), & if\ I=\ I_0 \\ p+1+R^{\ 1}(I), & if\ I = I_1 \\ p\ \ 1\ \ R_2^{\ 1}(I), & if\ I=\ I_2 \\ p+1+R_2^{\ 1}(I), & if\ I = I_3 \end{cases}. \quad (15)$$

(7) Special process.

a. For the KBD window,

if $|X(p+2)| < |X(p-2)|$
  $p = p+1$
  calculate I_1, $\tilde{f}_n = p+1+R^{-1}(I_1)$
else
  calculate I_0, $\tilde{f}_n = p-1-R^{-1}(I_0)$
end

b. For the sine window,

if $|X(p+2)| > |X(p-2)|$
  $p = p+1$
  calculate I_1
  if $I_1 > I_{th}$
    $\tilde{f}_n = p+1+R^{-1}(I_1)$
  else
    calculate I_2, $\tilde{f}_n = p-1-R_2^{-1}(I_2)$
  end
else
  calculate I_0
  if $I_0 > I_{th}$
    $\tilde{f}_n = p-1-R^{-1}(I_0)$
  else
    calculate I_3, $\tilde{f}_n = p+1+R_2^{-1}(I_3)$
  end
end

The empirical values of a_0 and λ_0 is different for the two windows to get better results. In our experiment, when KBD window is used, $a_0 = 5 \times 10^{-5}$, $\lambda_0 = 3.139$. When the sine window is used, $a_0 = 0.02$, $\lambda_0 = 3.889$. I_{th} is set according to the ratio range of (7), $I_{th} = -0.0606$.

4. EXPERIMENTAL RESULTS

We will now present some experimental results. Compared to the envelope method[6], the complexity increasing of the proposed method is mainly from one comparison in step (2) and one division to get the condition λ. In the experiments to follow, we will show the precision improvements.

4.1. Estimation of the Fractional Part

The first test is to examine the estimation precision of the fraction part of the digital frequency, $\xi = f_n - \lfloor f_n \rfloor$. We have selected two typical values of 10 and 510 according to N=1024 to show different effects of the omission to (3). With ε= 0.005, we have generated a series of the sine wave signals. For each signal, the estimation is proceeded in 126 consecutive windows. The step $\Delta\xi$ is set to 2^{-13} for an appropriate precision and the look-up table size. We have compared the proposed method with the original envelope ratio method [6]. The Pseudo-spectrum method [4] has also been used as a reference when the MDCT is under the sine window. The method in [5] is reported to be worse than the one in [4] and is ignored in our test. The results of the maximum errors and the mean errors under different windows are given in Fig. 1. From the plot we can see that, the proposed method can decrease the estimation error markedly in all cases.

TABLE I
DECISION ON THE FORMULAS

λ	KBD		Sine window	
	I_0>=I_1	I_0<I_1	I_0>=I_1	I_0<I_1
< λ_0[a]	I_0	I_1	I_0	I_1
>= λ_0[a]	I_1	I_0	if I_1> I_{th} I_1 else $I_2 = \frac{X(p-3)}{X(p+1)}$	if I_0> I_{th} I_0 else $I_3 = \frac{X(p+2)}{X(p-2)}$

[a] λ_0 can be different for the sine and KBD window.

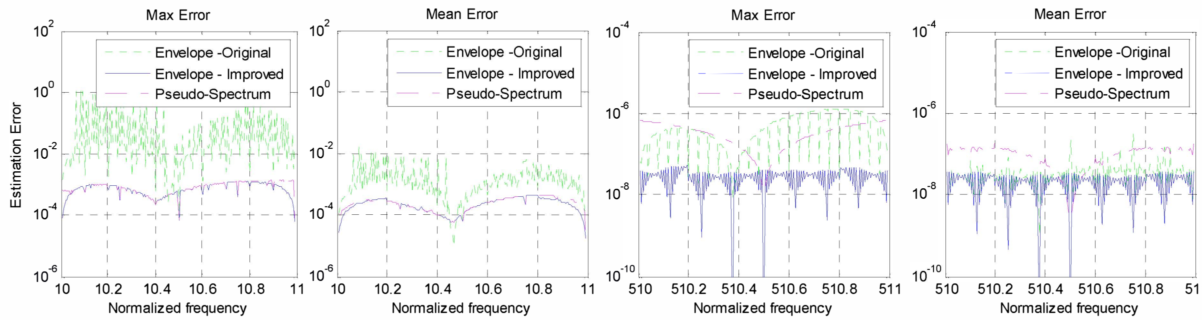
When the sine window is used, the proposed method can successively avoid the big errors. Even for $\lfloor f_n \rfloor = 10$, the maximum error is less than 10^{-3}. The proposed method is shown to have lower error than the Pseudo-spectrum method when ε near 0 or 1. For $\lfloor f_n \rfloor = 510$, the estimation errors are even lower. The precision of the proposed method is shown to be limited by the precision of the look-up table (obvious periodic oscillations), but this method still performs the best. The Pseudo-spectrum method has the least advantage with this frequency test set because its model error is obviously bigger now.

When the KBD window is used in the MDCT, the proposed method can restrict the maximum errors to even lower for $\lfloor f_n \rfloor = 10$. It is because the fast fading side lobes of the KBD window brings less interference to the estimation model than the sine window do. But for $\lfloor f_n \rfloor = 510$, this benefit can be ignored, and the model itself determines the estimation precision. The relatively flatten curve of the $R(\xi)$ under the KBD window gives a more coarse estimation result than under the sine window.

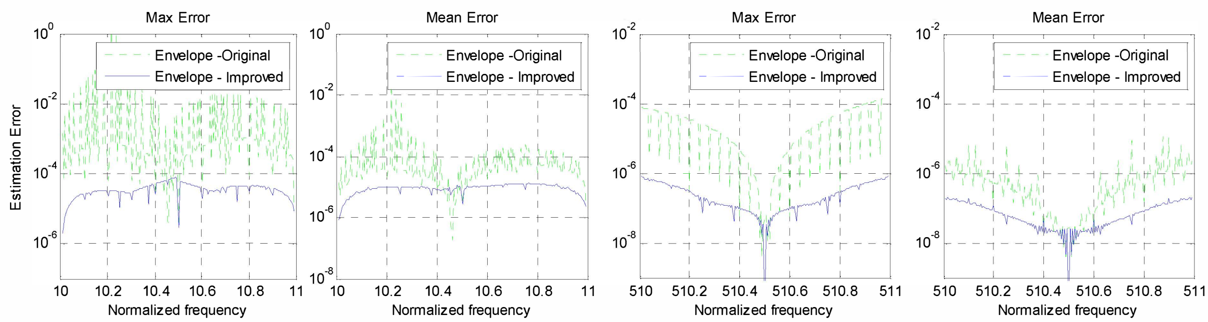
4.2. Estimation of the Frequencies in Practice

To compare the statistic data listed in the reference, we have carried out the same test as [4]-[6] have done. A set of 2090 sinusoids with random frequency distributed between 215 Hz and 4321 Hz obeying exponential law are used. Initial phases of these sinusoids are set randomly obeying uniform distribution on the interval between 0 and 2π. The errors are measured in percentage of halftones [8]. Results of the mean error μ, variance σ and the maximum error are presented in Table II.

From the result, we can see an obvious error decrease when the proposed method is used. It has even lower mean error and the variance than the Pseudo-spectrum method has (data extracted from [4]) in the sine window case. The better overall precision under the KBD window is because most of the test frequencies are concentrated around the low frequency part, where the estimation under the KBD window has apparent advantage.



(a) Estimation errors under the sine window



(b) Estimation errors under the KBD window

Fig. 1. Estimation errors, N=1024. , A=1. ϕ=0, $\Delta\xi=2^{-13}$. $\Delta\varepsilon$ =0.005.

TABLE II
FREQUENCY ERROR IN PERCENTAGE OF HALF TONES

method		σ	max
SINE WINDOW			
Pseudo-spectrum	0.005	0.013	0.167
Envelop ratio	0.0171	0.1439	3.6895
Improved Envelope ratio	0.0041	0.0124	0.1821
KBD WINDOW			
Envelope ratio	0.0011	0.0120	0.4160
Improved Envelope ratio	0.0002	0.0006	0.0080

5. CONCLUSION

In this paper, an improved low complexity high precision frequency estimation method has been presented. It is based on an in-depth analysis on the approximation of the MDCT coefficient expression and a systematic investigation on the estimation model and the formulas. The improved method avoids the mistakes caused by the approximation and the large errors brought by the inadequate formulas. Simulation tests show that the method can indeed enhance the estimation precision greatly. The method provides a way to carry out high precision estimation under the sine and KBD windows, which guarantees the accuracy of the following estimation on amplitude and phase in the MDCT domain.

6. REFERENCES

[1] H. Malvar, "Lapped transforms for efficient transform/subband coding." *IEEE Trans. Acoust., Speech, Signal Process.,* vol.38, no.6, pp.969-978, Jun. 1990.

[2] A.J.S. Ferreira. "Accurate estimation in the ODFT domain of the frequency, phase and magnitude of stationary sinusoids." In *2001 IEEE workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 21-24 2001, pp.47-50.

[3] L. Daudet and S. Merdjani. "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction." *IEEE Trans. Speech Audio Process.*, vol. SAP-12, no. 3, pp. 302-312, May 2004.

[4] S. Merdjani and L. Daudet. "Direct estimation of frequency from MDCT-encoded files." In *Proc. 6th Int. Conf. Digital Audio Effects* (*DAFx-03*), Sep.2003, pp.1-4.

[5] M. Zhu, W. Zheng, D. Li, and M. Zhang, "An accurate low complexity algorithm for frequency estimation in MDCT domain," *IEEE Trans. Consum. Electron.*, vol. CE-54, no. 3, pp. 1022–1028, Aug. 2008.

[6] S. Zhang, W. Dou, and H. Yang. "MDCT Sinusoidal Analysis for Audio Signals Analysis and Processing." *IEEE Trans. Audio, Speech, Language Process.*, vol. ASL-21, no. 7, pp. 1403-1414. Jul. 2013

[7] J. Princen and A. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1153–1161, Oct. 1986.

[8] F. Keiler, and S. Marchand. "Survey on extraction of sinusoids in stationary sounds." in *Proc. 7th Int. Conf. Digital Audio Effects* (*DAFx-02*), Sep. 2002, pp. 51–58.