Dun et al. EURASIP Journal on Audio, Speech, and Music Processing (2017) 2017:8
DOI 10.1186/s13636-017-0105-5

EURASIP Journal on Audio, Speech, and Music Processing

RESEARCH Open Access

An MDCT domain three-point interpolation-based low-complexity frequency estimator

Yujie Dun*, Guizhong Liu and Xingsong Hou

Abstract

Signal frequency estimation is a problem of significance in many applications including audio signal processing. Compressed domain audio frequency estimators that directly use the modified discrete cosine transform (MDCT) coefficients are suitable for low-complexity audio applications. A new frequency estimation approach, which can obtain the estimated value from a simple combination of three MDCT coefficients, is proposed herein. It exploits the underlying relation among adjacent MDCT values and provides a general form of this type of estimators. The estimator manifests obvious computational advantages over other MDCT domain estimators and is suitable for high signal-to-noise ratio (SNR) conditions.

Keywords: Frequency estimation, MDCT, Audio, Low complexity

1 Introduction

Frequency estimation is a basic problem in signal processing research and has been widely used in various applications such as economics, meteorology, astronomy, industry, and consumer electronics [1]. In recent years, low-complexity frequency estimators, which are suitable for low-cost applications, have been proposed in addition to so-called high-resolution (or even super-resolution) frequency estimation techniques such as Pisarenko [2], MUSIC [3] and ESPRIT [4]. A typical class of the low-complexity algorithms operates in the frequency domain (via discrete Fourier transform, DFT) and uses several DFT bins to obtain the estimated value [5–8].

For audio signals, frequency estimation plays a crucial role in parametric audio processing, which has been reported in various applications such as synthesis [9, 10], recognition [11], enhancement [12], and frame-loss concealment [13, 14]. In particular, in audio coding, the following two major profiles in MPEG-4 audio coding are based on the sinusoidal analysis of an audio signal: HILN (Harmonic and Individual Lines plus Noise) [15] and SSC (SinuSoidal Coding) [16]. Using the low-complexity frequency estimator can effectively lower the resource requirement of the entire processing system, which is significant for massive amount multimedia data processing and portable ultra-low-power media devices. However, the aforementioned frequency estimation algorithms are not applicable for most low-cost audio applications.

Audio data that are used in most audio applications are stored and transmitted in compressed format, but the compression is not based on DFT. Thus, estimating the parameters of an audio signal, which includes the frequency estimation, is considerably complex. The time-domain signal samples should first be recovered from the compressed data before the estimation, but the recovery generally has a relatively high degree of computational complexity. For high-quality audio compression standards such as MPEG2/4 AAC, Dolby AC-3, WMA, and IETF Opus, the compression is conducted in the modified discrete cosine transform (MDCT) [17] domain, where an overlap of 50% between successive blocks and time domain alias cancellation (TDAC) are used to mitigate the block effect. To recover one block of the time samples, the inverse MDCT (IMDCT) of three successive blocks is required. Although the frequency estimation algorithm is simple, the IMDCT significantly adds the computational complexity during the recovery of the time domain samples.

To reduce the complexity, several approaches have been proposed. One is to directly calculate the DFT from MDCT with a fast algorithm [18], and the frequency estimation is performed with these DFT values. However, computing the DFT of every block requires

* Correspondence: dunyj@xjtu.edu.cn
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

the MDCT values of the corresponding block, previous block, and succeeding block, which causes an inevitable algorithm delay of one block. Another approach is to use the odd-DFT as an intermediate domain between the time domain and the MDCT domain. The frequency is estimated with the odd-DFT coefficients; then, the MDCT is obtained from the odd-DFT by a simple conversion [19–21]. Using the odd-DFT, the system complexity of an audio application can effectively be decreased, but this scheme is not fit for the applications that take the compressed audio as their input. Another approach is to directly estimate the frequency with the MDCT coefficients. With the analysis of the MDCT coefficients of a sinusoid [22], several MDCT domain estimators have been proposed in the last decade [23–25], which shows great convenience for the low-complexity implementation of an estimator. All estimators are based on the ratio of two coefficients using the mapping relationship between the frequency value and the coefficient ratio. Effective estimation is restricted in the monotone mapping region. However, in practice, the noise is unavoidable, which leads the estimation to the non-monotonic region and produces a wrong result.

The major objective of this paper is to propose a three-point interpolation-based estimator, which avoids the effect of non-monotonic mapping and further reduces the complexity of the MDCT domain frequency estimator to render a simple method for various applications. The contributions are summarized as follows: (i) derive an analytical expression of the MDCT of a single-tone sinusoid based on the sine window's centered DFT (CDFT); (ii) propose an MDCT domain three-point interpolation-based low-complexity approach for the signal frequency estimation problem. The proposed algorithm estimates the frequency from three MDCT bin values with only simple calculations and is significantly less complex than the existing methods. The method is effective for the sine window case and exhibits an estimation error lower than 1 Hz when the signal-to-noise ratio (SNR) is above 20 dB.

This paper is organized as follows. In Section 2, we provide the MDCT analysis of a sinusoid, which is the basis of the MDCT domain estimators. The proposed algorithm is presented in Section 3. In Section 4, the Monte-Carlo simulation results are shown and the complexity is analyzed. The conclusions are summarized in Section 5.

2 MDCT analysis of sinusoids

2.1 Signal model of the estimation

Audio signals are commonly modeled as a combination of several sinusoidal frequency components, which can be expressed as

$$s_a(n) = \sum_{m=0}^{P-1} s_m(n) = \sum_{m=0}^{P-1} A_m \sin(2\pi f_m n + \phi_m), \quad (1)$$

where n is the signal index; P is the number of components; A_m, f_m, and $\varnothing_m$ are the amplitude, normalized frequency, and phase of each component $s_m(n)$, respectively. The problem of the audio signal parameter estimation is to obtain the values of each parameter set $\{A_m, f_m, \phi_m\}$ for $m = 0, 1, \ldots, P-1$. In general, the frequency estimation is the most important. These frequencies can be estimated together as most time domain methods do or estimated one by one as the frequency domain methods commonly do. When these components are well separated in the frequency scale, the estimation of each component in the frequency domain can be treated as the problem of estimating each single frequency component where all other components act as interference noise. Thus, the signal model may be simplified to a single-component model. In this paper, we concentrate on the frequency estimation of a single tone.

Given a discrete sinusoid, the single-tone signal is expressed as

$$s(n) = A \sin(2\pi f n + \phi), \quad (2)$$

where A, f, and $\varnothing$ are the magnitude, frequency, and initial phase of this sinusoid, respectively. Considering the noisy case, the observed signal is

$$x(n) = s(n) + w(n) = A \sin(2\pi f n + \phi) + w(n), \quad (3)$$

where $w(n)$ is generally assumed as the additive white Gaussian noise (AWGN) with zero mean and variance σ^2. The SNR is $A/(2\sigma^2)$.

To estimate the parameters in the MDCT domain, the signal $x(n)$ is framed by weighting a window function $h(n)$ of length $2N$, which satisfies the Princen-Bradley perfect reconstruction conditions [17], and converted to its N point MDCT coefficients,

$$X(k) = \sum_{n=0}^{2N-1} x(n)h(n) \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right], \quad (4)$$

where $k = 0, 1, \ldots, N-1$ is the MDCT bin index. The problem of MDCT domain frequency estimation is to estimate the value of f from MDCT coefficients $X(k)$. f is commonly expressed as

$$f = \frac{f_s}{2N} l = \frac{f_s}{2N} (l_0 + \delta), \quad (5)$$

where f_s is the sampling frequency, $l_0 \in \mathbf{Z}_0^+$, and $\delta \in [0,\ 1)$ is the integer and fractional part of the digital frequency l. Thus, the estimation of l is to obtain the values of l_0 and δ.

2.2 Generalized MDCT analysis

The MDCT analysis of a sinusoid is the basis of the frequency estimator in the MDCT domain. It exhibits the underlying relationship between the MDCT coefficients and the parameters of the sinusoidal signal. This relationship was first explored by Daudet [22] for the sine window case and generalized by Zhang [25] to other window cases. Here, we briefly describe the generalized MDCT analysis. The analysis is similar to that of [25], but the signal model uses Eq. (3).

Considering the noiseless case, the signal is shown in (2); the general form of the MDCT coefficient $X(k)$ of the signal with window $h(n)$ is the real part of an expression $Z(k)$ in the form of [25]

$$Z(k) = A\sqrt{\frac{N}{2}}\cdot e^{j\phi}\cdot\left[e^{-j\frac{3\pi}{2}(k+0.5)}H(k-l) + e^{j\frac{3\pi}{2}(k+0.5)}H(-k-l-1)\right], \tag{6}$$

where

$$\phi = \frac{2N-1}{2N}\pi l - \frac{\pi}{2} + \phi. \tag{7}$$

$H(\xi)$ is the centered discrete Fourier transform (CDFT) of a window function $h(n)$,

$$H(\xi) = \frac{1}{N}\sum_{n=0}^{2N-1} h(n)e^{-j\frac{2\pi}{2N}(n+0.5-N)(\xi+0.5)}, \tag{8}$$

where ξ is not restricted to integer. If $h(n)$ is even-symmetric (a common case in MDCT analysis), the values of its CDFT $H(\xi)$ are real. The MDCT coefficient of the signal in (2) is expressed as

$$X(k) = A\sqrt{\frac{N}{2}}\cdot\left[cos\left(\phi_0 - \frac{3\pi}{2}k\right)\cdot H(k-l) + (-1)^k\, sin\left(\phi_0 - \frac{3\pi}{2}k\right)H(-k-l-1)\right], \tag{9}$$

where ϕ_0 is defined as

$$\phi_0 = \phi - \frac{3\pi}{4} = \frac{2N-1}{2N}\pi l - \frac{5\pi}{4} + \phi. \tag{10}$$

Equation (9) provides the precise result of the MDCT coefficient for a given sinusoidal signal with an arbitrary symmetric window function case.

To build a simple relation between the sinusoidal frequency and the MDCT coefficients, we must simplify (9). Such simplification can be performed based on the features of the window and its CDFT $H(\xi)$. The window function has fast fading sidelobes, which makes the significant values of its CDFT coefficients appear only at approximately $\xi = 0$ [25]. For $k = 0, 1, \ldots, N-1$ and l far from 0 or N–1, only the first term in (9) is significant. Thus, the simplified expression of (9) is

$$X(k) = A\sqrt{\frac{N}{2}}\cdot cos\left(\phi_0 - \frac{3\pi}{2}k\right)\cdot H(k-l). \tag{11}$$

2.3 MDCT analysis for sine window case

The sine window is commonly used in audio signal processing and coding. The frequency estimator for the sine window case is important for practical applications. The analytical expression of the CDFT coefficient $H(\xi)$ for the sine window can be derived; thus, the analytical expression of the MDCT coefficient $X(k)$ can also be derived. The expression of $X(k)$ is the basis of the proposed three-point interpolation-based low-complexity frequency estimator.

The sine window is defined as

$$h_{sin}(n) = sin\left[\frac{\pi}{2N}\left(n + \frac{1}{2}\right)\right], \tag{12}$$

where $n = 0, 1, \ldots, 2N-1$ has the identical length as the MDCT input data. The sine window is even-symmetric, and its CDFT is real-valued. Substituting (12) into (8) and simplifying, we obtain the following expression of the CDFT

$$H(\xi) = \frac{sin(\pi\xi)}{2N}\cdot\left(\frac{1}{sin\left(\frac{\pi}{2N}\xi\right)} - \frac{1}{sin\left(\frac{\pi}{2N}(\xi+1)\right)}\right). \tag{13}$$

For ξ near 0, which implies that the bin index k is near the digital frequency l, Eq. (13) can be approximated as

$$H(\xi) \approx \frac{1}{\pi}\cdot\frac{sin(\pi\xi)}{\xi(\xi+1)}. \tag{14}$$

Values at $\xi = \{0, -1\}$ are obtained using L'Hospital's rule. This approximation leads to an error less than 1.25×10^{-7}. Substituting (14) into (11), a simplified MDCT bin value $X(k)$ is obtained

$$X(k) = \frac{A}{\pi}\sqrt{\frac{N}{2}}\cdot\frac{sin[\pi(k-l)]}{(k-l)((k-l+1)}\cdot cos\left(\phi_0 - \frac{3\pi}{2}k\right). \tag{15}$$

This result is the basis of the proposed frequency estimator.

3 Proposed frequency estimator

3.1 General form

To obtain the estimator, we reform (15) as

$$X(k) = \frac{A}{\pi}\sqrt{\frac{N}{2}}sin(\pi l)\cdot\frac{1}{(k-l)((k-l+1)}\cdot(-1)^{(k+1)}cos\left(\phi_0 - \frac{3\pi}{2}k\right). \tag{16}$$

In (16), $X(k)$ is composed of three parts: a constant valued part $\frac{A}{\pi}\sqrt{\frac{N}{2}}sin(\pi l)$, a variable value part $\frac{1}{(k-l)(k-l+1)}$,

and a phase modulation factor $(-1)^{k+1}\cdot cos\left(\phi_0-\frac{3\pi}{2}k\right)$. The phase modulation factor has a period of 4 and can be listed as

$$k = \begin{matrix} 0, & 1, & 2, & 3, & 4, & \cdots \\ -cos\phi_0, & -sin\phi_0, & cos\phi_0, & sin\phi_0, & -cos\phi_0, & \cdots \end{matrix}$$

Thus, taking $M(k)=\frac{1}{X(k)}$, for a given k_0, denoting $M_- = M(k_0-2)$, $M_0 = M(k_0)$, and $M_+ = M(k_0+2)$, we construct a combination of these three values in the form of

$$\lambda = \frac{b_1M + b_2M_0 + b_3M_+}{a_1M + a_2M_0 + a_3M_+}, \tag{17}$$

where a_i and b_i ($i = 1, 2, 3$) are real-valued coefficients. Then, the constant part and phase modulation factor in (15) are canceled out, and only combinations of $(k-l)(k-l+1)$ remain. Defining $\delta_0 = l - k_0$ and substituting it into (17), we obtain

$$\lambda = \frac{B_2\delta_0^2 + B_1\delta_0 + B_0}{A_2\delta_0^2 + A_1\delta_0 + A_0}, \tag{18}$$

where

$$\begin{cases} A_2 = a_1-a_2+a_3 \\ A_1 = 3a_1+a_2-5a_3 \\ A_0 = 2a_1+6a_3 \end{cases}, \text{and} \begin{cases} B_2 = b_1-b_2+b_3 \\ B_1 = 3b_1+b_2-5b_3 \\ B_0 = 2b_1+6b_3 \end{cases}. \tag{19}$$

If the coefficients a_i and b_i are properly set, a simple relation between λ and δ_0 can be obtained and δ_0 can be estimated. For example, if we set $A_2 = A_1 = 0$ and $B_2 = B_0 = 0$ by properly selecting the coefficients a_i and b_i, then $\lambda = \delta_0 \cdot B_1/A_0$, B_1/A_0 is a constant determined by a_i and b_i. An estimation to δ_0 is $\lambda/(B_1/A_0)$. Thus, the frequency value $\hat{l}$ (we use $\hat{}$ to denote an estimated value) can be estimated by $\hat{l} = k_0 + \hat{\delta}_0$.

3.2 Proposed estimator

In the proposed estimator, k_0 is set to the index of the maximum MDCT magnitude $|X(k)|$. δ_0 is estimated using the following formula:

$$\delta_0 = \frac{3M_- + 2M_0 - M_+}{2M_- + 4M_0 + 2M_+}. \tag{20}$$

To simplify the computation, we convert formula (20) to a form that directly uses $X(k)$. For $i = -2, 0, 2$, denoting $X(k_0 + i)$ as X_-, X_0, and X_+, respectively, we obtain a new form of (20)

$$\delta_0 = \frac{3X_0X_+ + 2X_-X_+ - X_-X_0}{2(X_0X_+ + 2X_-X_+ + X_-X_0)}. \tag{21}$$

The key steps of the proposed estimator are summarized as follows:

(1) Find the bin index of the MDCT magnitude peak,

$$\hat{k}_0 = \arg\max k(|X(k)|). \tag{22}$$

(2) Estimate δ_0 with the MDCT values of X_-, X_0, and X_+ according to formula (21).

(3) Finally, obtain the estimated value of l,

$$\hat{l} = \hat{k}_0 + \hat{\delta}_0. \tag{23}$$

It is noted that (20) is not the only formula to estimate δ_0; we have derived a set of such formulas; for example,

$$\delta_0 = \frac{3M_- - 14M_0 - M_+}{3M_- + 2M_0 - M_+} = \frac{6M_- - 12M_0 - 2M_+}{5M_- + 6M_0 + M_+} = \frac{24M_- - 8M_+}{17M_- + 30M_0 + 13M_+} = \cdots. \tag{24}$$

However, the coefficients in (20) are the most suitable for a simple calculation.

4 Results and discussion

4.1 Comparison benchmarks

Four reported MDCT domain estimators [23–26] and one simplified estimator were used as the performance comparison benchmarks. The four reported estimators are as follows:

- Merdjani [23], a method based on the analytical expression of the MDCT coefficient;
- Zhu [24], a computationally efficient version of [23];
- Zhang [25], an envelope-function-based method with a look-up table (the single-frame-based envelop method without iteration is used); and
- Dun [26], an improved version of the above envelope function method.

We have implemented the estimators of Merdjani and Zhu and obtained Zhang's from its author. Based on our previous work (Dun), we have noticed that all of these estimators involve conditional constructs, i.e., the specific algorithm is chosen according to one criterion or several criteria. The decision algorithm verifying the criteria and the conditional branch instructions selecting specific algorithm increases the complexity of the program flow especially for pipelined processing. Thus, in our verification tests, one additional benchmark, which is a simplified estimator derived from Merdjani [23], is used and labeled as "Simplified" in the following tests. This simplified estimator

Table 1 Comparison of the complexity

Estimators	Time complexity						Space complexity
	Addition	Multiplication	Division	Comparison	Square-root	Bit-shift	
Merdjani	6 + 2 N	4 + 2 N	3	4	1 + N	–	–
Zhu	8	3	5	5	1	–	–
Zhang	7	1	3	2	–	–	4096
Dun	10	1	5	5	–	–	6144
Simplified	5 + 2 N	2 + 2 N	2	–	1 + N	1	–
Proposed	5	3	1	–	–	3	–

has no conditional branch (similar to the proposed estimator), and the frequency is estimated by,

$$f = \frac{f_s}{2N}(k_0 + \delta) = \frac{f_s}{2N}\left(k_0 + \frac{3 + \alpha - \sqrt{\alpha^2 + 14\alpha + 1}}{2(1-\alpha)}\right), \quad (25)$$

where k_0 is the frequency bin that locates the maximum of the so-called pseudo-spectrum $S(k)$,

$$S(k) = \sqrt{X(k)^2 + [X(k-1) - X(k+1)]^2}, \quad (26)$$

$$k_0 = \arg\max k(S(k)), \quad (27)$$

and α is the ratio of two MDCT coefficients,

$$\alpha = -\frac{X(k_0 - 1)}{X(k_0 + 1)}. \quad (28)$$

4.2 Complexity comparison

4.2.1 General

Complexity refers to the resources that an executable program of the algorithm requires; it includes time complexity and space complexity. Here, the time complexity is compared by accounting the required operations to estimate the frequency, and the space complexity refers to the storage space size required by the algorithm.

To compare the time complexity, operations such as addition, multiplication, division, square root, comparison, and bit-shift are accounted for each algorithm. Most

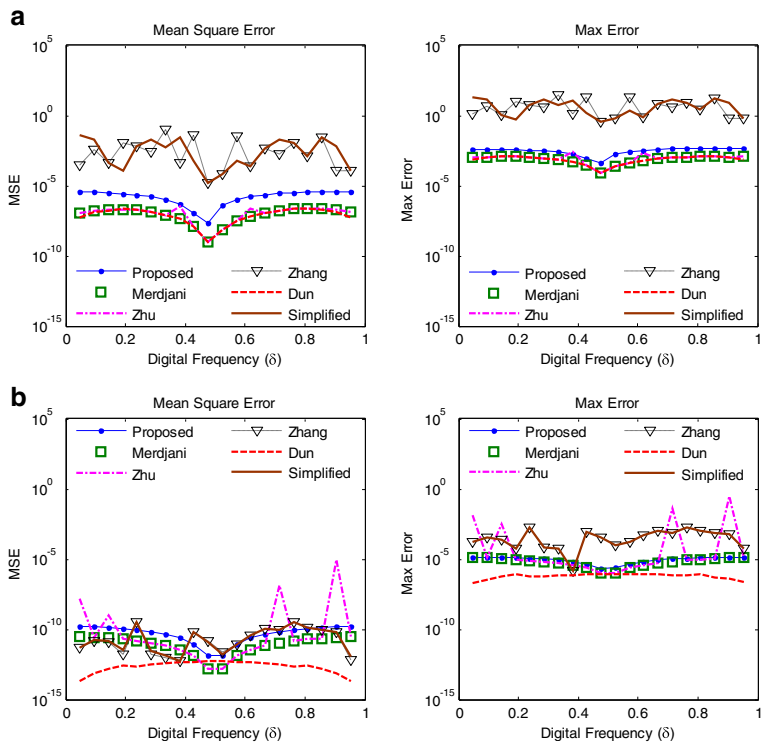

Fig. 1 MSE comparison of the estimators for single-tone sinusoidal input without noise. Two sets of the sinusoidal signal were used. The results of the two sets are presented in (**a**) and (**b**). We set $l_0 = 46$ for (**a**) and $l_0 = 510$ for (**b**). There were averages of 10,000 runs for each frequency set

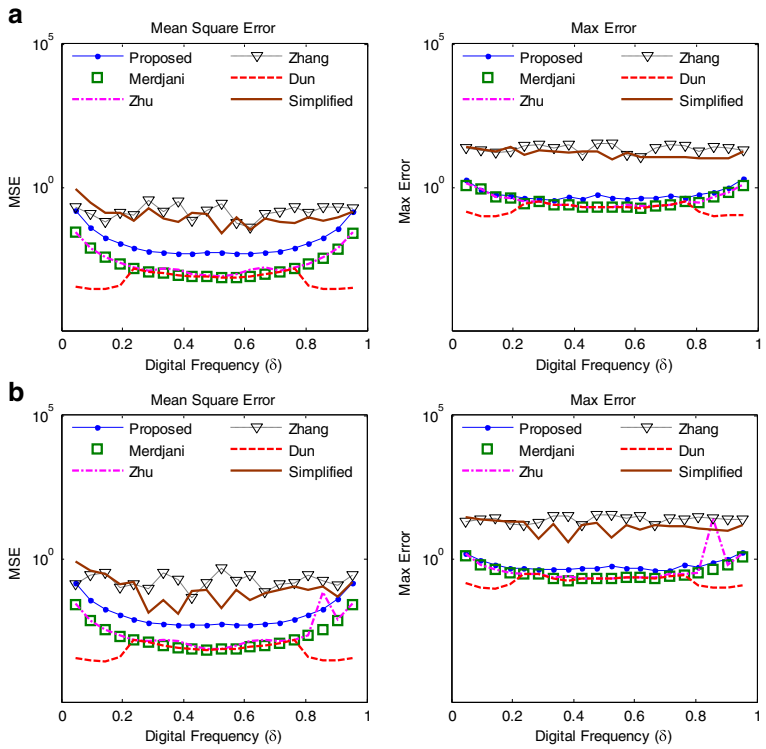

Fig. 2 MSE comparison of the estimators for single-tone sinusoidal input polluted by noise. Two sets of the sinusoidal signal were used. The results of the two sets are presented in (**a**) and (**b**). We set $l_0 = 46$ for (**a**) and $l_0 = 510$ for (**b**). SNR = 40 dB. There were averages of 10,000 runs for each frequency set

existing MDCT domain frequency estimation algorithms [23–26] consist of two steps: find the frequency bin k_0 that corresponds to the integer part l_0 and estimate the fractional part δ using a decision method. Note that finding the bin index of the peak location is a common step for all algorithms and the operations are identical, so the operations to find this peak are not included in the comparison.

To compare the space complexity, the required space size to store the look-up table is accounted. The required space to locate the variables and intermediate results is not included in the comparison.

4.2.2 The proposed estimator

According to the proposed frequency estimator in Section 3.2, with the bin index of the maximum $|X(k)|$, the operations to obtain the estimated value $\hat{l}$ is shown in (21), which includes three MDCT-coefficient-multiplications (X_X_0, X_0X_+, and X_X_+), three constant-coefficient-multiplications (with 3 and 2), four additions, and one division. A multiplication with numbers such as 2 and 3 is usually substituted by one bit-shift and addition. Thus, in practice, three multiplications, five additions, one division, and three bit-shifts are used. Neither additional information nor other operation is required.

4.2.3 Other MDCT domain estimators

First, all compared estimators find a peak location. [24–26] use other criteria after locating the initial maximum to obtain $\hat{l}_0$, whereas Merdjani [23] and the simplified estimator locate the maximum of pseudo-spectrum that is converted from MDCT spectrum. The use of a pseudo-spectrum helps to find the exact $\hat{l}_0$, but it also adds a certain amount of operations, which must be accounted in the comparison. Then, always with some decision algorithms (particularly in Zhu [24] and Dun [26]), the value of $\hat{\delta}$ is solved from a quadratic equation or computed from a look-up table with polynomial fitting.

We have compared the complexity of these methods as shown in Table 1. The given numbers are the typical values of every algorithm. The size of the look-up table relates to the step. The data in the table present how many values should be stored according to a step of 2^{-13} as reported in [26].

Table 1 shows that the proposed estimator only requires several addition, multiplication, and division operations aside from three bit-shift operations (the simplest operation among the list). Neither comparison nor saving space is required. Obviously, the proposed estimator has the lowest complexity. The simplified algorithm has a similar complexity with the proposed estimator if the calculation of $S(k)$ is not considered.

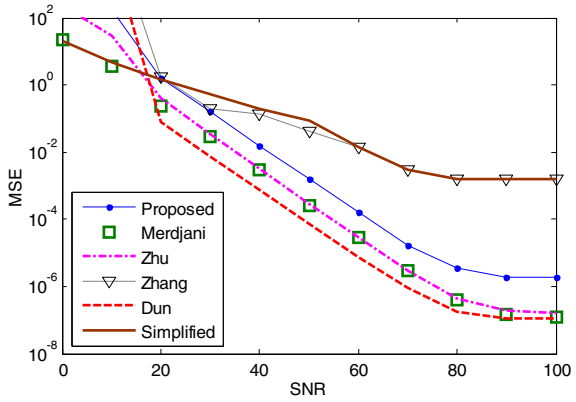

Fig. 3 MSE vs. SNR of different frequency estimators for single-tone sinusoidal input. There were averages of 10,000 runs for each SNR

Table 2 The description of the 12 MPEG mono sequences

Name	Time/s	Type
es01	10.73	Suzanne Vega
es02	8.7	Male speech, German
es03	7.6	Female speech, English
sc01	10.97	Haydn trumpet concert
sc02	12.73	Classical orchestral music
sc03	11.55	Contemporary pop music
si01	8	Harpsichord/cembalo
si02	7.73	Castanets
si03	27.89	Pitch pipe
sm01	11.15	Bagpipe
sm02	10.1	Glockenspiel
sm03	13.99	Plucked strings

4.3 Simulation results and discussion

Simulations were conducted to verify the proposed frequency estimator and compare with other estimators. Herein, the results for both noiseless and noise-polluted cases are presented.

In all simulations, parameters were set according to the audio applications. The block size and window length were set to $2N = 2048$, the sampling frequency was $f_s =$ 44.1 kHz, and the magnitude was $A = 1$. The initial phase ϕ was randomly generated in the range of $(-\pi,\ \pi)$, which obeyed the uniform distribution. The estimation error of the frequency value, i.e., $\varepsilon = \hat{f} - f$ in Hertz (Hz), where f is the sinusoidal frequency and $\hat{f}$ is the estimated value, was measured by the maximum value $\varepsilon_{\max}$ and mean square error (MSE). An MSE of 0 dB represents an error of 1 Hz.

With expression (5), we compared the precision of the estimators when δ varied from 0 to 1 with a step of 0.05. The signal frequency l partially decides the model error when simplifies the original form (9) to expression (11); therefore, two values, 46 and 510, were used for its integer part l_0 in this test. The value of 46 is a bin number that corresponds to approximately 1 kHz according to values of f_s and N. The value of 510 is approximately half of the MDCT bin index, which can minimize the interference caused by the negative frequency of a real-valued sinusoidal input. The results of the noise-free condition are shown in Fig. 1.

As expected, both MSE and maximum error are larger for all estimators when $l_0 = 46$. In this frequency domain, the proposed estimator exhibits a slightly larger MSE and maximum error compared to Merdjani, Zhu, and Dun's methods but significantly less than Zhang's method and the simplified method. In other words, although no conditional construct is used, the proposed estimator exhibits similar precision to the ones that have conditional branches, whereas other existing estimators significantly lose their accuracy. When $l_0 = 510$, the maximum error of the proposed estimator remains similar to other estimators that have conditional branches.

For both cases, the proposed estimator has a slightly larger MSE than the other branched method. The degradation in performance is mainly caused by the third coefficient. In [23, 24, 26], additional decisions are made to select the largest two values. In the proposed estimator, three values are required; neither decision algorithm nor conditional branch instruction is used. Thus, an ultra-low-complexity approach is obtained. Fortunately, the MSE remains near or below 10^{-10} for most frequencies.

Then, the corresponding test of the noise-polluted counterparts was performed. This test shows the performance of each estimator under the condition of noisy interference. For the frequency estimation of a real audio signal, the noise originates from other sound sources, environmental noise, and other frequency components of the audio signal. For multicomponent signals, the interference

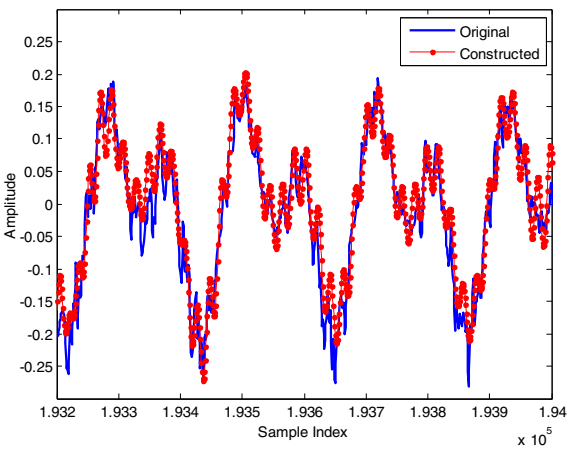

Fig. 4 Waveform comparison of the reconstructed audio and the original audio. The plot shows a detailed part of the es01 audio wave

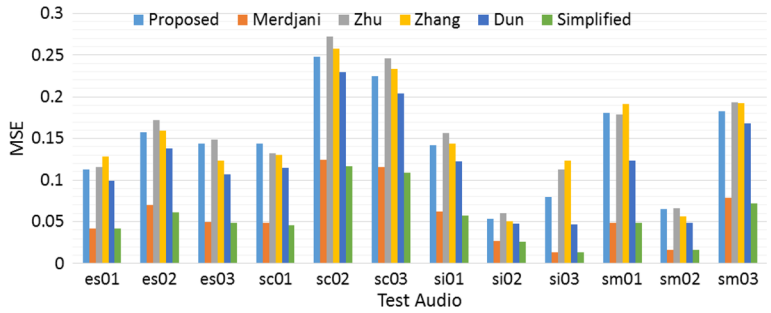

Fig. 5 Audio quality comparison by MSE. Using the original audio signals as references, MSE of each reconstructed audio is calculated

from other frequency components are a major source of noise. The corresponding results with noise of SNR = 40 are shown in Fig. 2. The precision of all estimators significantly degrades, and MSEs increase from less than 10^{-10} to greater than 10^{-3}. A level of 10^{-2} is shown for the proposed estimator, which corresponds to an error of 0.1 Hz.

A test of MSE vs. SNR was also conducted. In this test, l_0 was set to 46, which corresponded to approximately 1 kHz; δ was set to be randomly uniformly distributed in (0, 1). The results are shown in Fig. 3. Basically, for SNR higher than 20 dB, the MSEs of the proposed estimator are less than 1 Hz. The maximum sidelobe level of the sine window is −23 dB; thus, for two frequency components, a distance greater than one and a half bin guarantees that the interference is less than −23 dB. According to the parameter settings, this 1.5 bin distance corresponds to 32.3 Hz frequency offset, which is similar to the frequency difference of two music notes: C1 (261.6 Hz) to D1 (293.7 Hz). But in practice, the distance between the notes of a chord is greater than this value. Thus, the proposed estimator is suitable for the low-complexity frequency estimation at such high SNR situation.

4.4 Evaluation with real audio signals

In this part, the proposed algorithm is evaluated with real audio signals. After estimating the major components of an audio signal with sinusoidal model parameters (frequency, amplitude, and phase), the signal is reconstructed by the estimated components. The performances of the various methods are evaluated by comparing the original and the reconstructed signals.

In general, the major components of an audio signal are obtained by the following steps: firstly, finding the largest peak in the spectrum and estimating single-tone parameters from it; secondly, subtracting this estimated tone from the spectrum. These two steps are repeated until all major tones are estimated. This procedure is recommended in multiple component estimation algorithms because it enables detection of any tones that are initially masked by leakage from nearby large peaks.

In specific, the frequency of each component is estimated firstly; then, the amplitude and phase are estimated with the method given in Merdjani [23]. The proposed algorithm and the five benchmarks are used to get the estimated frequencies. To make comparison in a uniform framework, the components of an audio signal are estimated in the same order by all of the algorithms.

The test has been conducted with audio set that is used in the verification test of MPEG audio, which contains 12 mono audio files as listed in Table 2. With a sampling frequency of 48 kHz and frame length of 1024, each frame lasts about 21.3 ms. Maximum component number of 30 and minimum residual energy of 10^{-4} are

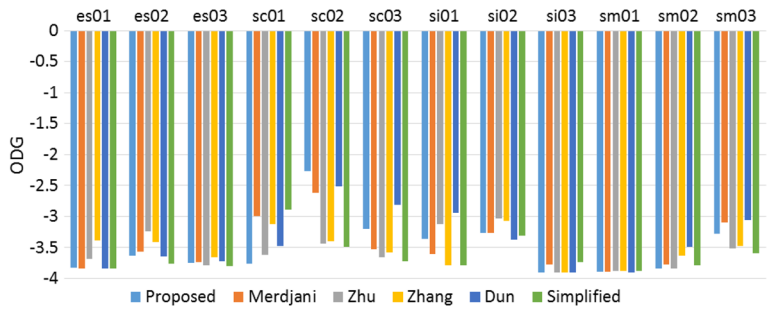

Fig. 6 Audio quality comparison by ODG. Using the original audio signals as references, ODG of each reconstructed audio is evaluated by software PQevalAudio

used as criteria to stop component extraction of a frame. An overlap of 50% is used between subsequent frames both in MDCT analysis and in waveform reconstruction. Figure 4 presents a detailed part of the reconstructed signal of "es01" when the proposed frequency estimation algorithm is used, and compares it with the original signal. It can be observed that the reconstructed waveform is almost the same with the original audio.

To evaluate the performance of the proposed algorithm, not only the errors between the original and the reconstructed signals are compared but also the audio qualities of the reconstructed signals are measured. The errors are compared by using MSE between the original and the reconstructed audio signals, and the result is plotted in Fig. 5. The audio quality is evaluated by using formal objective test with PQevalAudio software, which is used for perceptual evaluation of audio quality (PEAQ) specified in ITU BS.1387-1. The Objective Difference Grade (ODG), which has a range from 0 to −4, is used to indicate the audio quality. A score of 0 means no perceptible difference compared with a reference audio, and a score of −4 means that apparent performance degradation can be perceived. The test results are shown in Fig. 6.

The results of Figs. 5 and 6 show that the performance of the reconstructed audio signal remains similar to other estimators except the two most complexed ones although the proposed algorithm reduces the complexity greatly. The proposed algorithm avoids the spectrum conversion (from MDCT to pseudo-spectrum) used in Merdjani [23] and the simplified algorithm so that the algorithm complexity is irrelevant to the frame length N (as shown in Table 1, typical frame length of audio signal is 1024, 512, or so). At the same time, the proposed algorithm avoids the conditional constructs, which is beneficial to the speed of a frequency estimator in pipelined processor.

5 Conclusions

A low-complexity frequency estimator that operates with three MDCT coefficients and only several simple calculations is proposed in this paper. The analytical expression of the MDCT coefficients, which is the basis of the proposed estimator, is also presented. The proposed estimator shows a great reduction in complexity compared to other MDCT domain estimators and provides a good complexity/performance tradeoff. Without using conditional branch instructions, this estimator is especially fit for pipelined operators.

Funding

This research was supported in part by the National Natural Science Foundation of China under Grants NSFC61173110, NSFC61373113, NSFC61372091, NSFC61671365 and NSFC U1531141.

Authors' contributions

YD was responsible for proposing the algorithm and drafting the manuscript. GL and XH provided the comments on the verification tests and the drafts. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

6 Publisher's Note

Received: 8 September 2016 Accepted: 29 March 2017
Published online: 04 April 2017

References

1. P Stoica, RL Moses, *Spectral analysis of signals* (Pearson/Prentice Hall, Upper Saddle River, 2005)
2. VF Pisarenko, The retrieval of harmonics from a covariance function. Geophys. J. Int. **33**(3), 347–366 (1973)
3. RO Schmidt, Multiple emitter location and signal parameter estimation. Antennas and Propagation IEEE Transactions on **34**(3), 276–280 (1986)
4. R Roy, T Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques. Acoustics, Speech and Signal Processing IEEE Transactions on **37**(7), 984–995 (1989)
5. BG Quinn, Estimating frequency by interpolation using Fourier coefficients. Signal Processing, IEEE Transactions on **42**(5), 1264–1268 (1994)
6. MD Macleod, Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones. Signal Processing, IEEE Transactions on **46**(1), 141–148 (1998)
7. E Jacobsen, P Kootsookos, Fast, accurate frequency estimators [DSP Tips & Tricks]. Signal Processing Magazine, IEEE **24**(3), 123–125 (2007)
8. C Candan, Analysis and further improvement of fine resolution frequency estimation method from three DFT samples. Signal Processing Letters, IEEE **20**(9), 913–916 (2013)
9. H Kawahara, I Masuda-Katsuse, A De Cheveigne, Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. Speech Comm. **27**(3), 187–207 (1999)
10. EB George, MJ Smith, Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. Speech and Audio Processing, IEEE Transactions on **5**(5), 389–406 (1997)
11. A. Eronen, and A. Klapuri, Musical instrument recognition using cepstral coefficients and temporal features. (Acoustics, Speech, and Signal Processing, ICASSP'00. 2000 IEEE International Conference on, Istanbul, 2000), pp. II753-II756 vol. 2
12. DPN Rodríguez, JA Apolinário, LWP Biscainho, Audio authenticity: detecting ENF discontinuity with high precision phase analysis. Information Forensics and Security, IEEE Transactions on **5**(3), 534–543 (2010)
13. S.-U. Ryu, and K. Rose, An mdct domain frame-loss concealment technique for mpeg advanced audio coding. (Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, Honolulu, 2007), pp. I-273-I-276
14. M.-Y. Zhu, N. Chen, X.-Q. Yu, and W.-G. Wan, Packet Loss Concealment for compressed audio stream using sinusoidal frequency estimation. (Multimedia and Expo (ICME), 2010 IEEE International Conference on, Suntec City, 2010), pp. 316–321
15. H. Purnhagen, and N. Meine, HILN—the MPEG-4 parametric audio coding tools. (Circuits and Systems, The 2000 IEEE International Symposium on, Geneva, 2000), pp. 201–204
16. A. C. Den Brinker, J. Breebaart, P. Ekstrand, J. Engdegård, F. Henn, K. Kjörling, W. Oomen, and H. Purnhagen, An overview of the coding standard MPEG-4 audio amendments 1 and 2: HE-AAC, SSC, and HE-AAC v2, EURASIP Journal on Audio, Speech, and Music Processing. 2009(3(2009)
17. JP Princen, AB Bradley, Analysis/synthesis filter bank design based on time domain aliasing cancellation. Acoustics, Speech and Signal Processing, IEEE Transactions on **34**(5), 1153–1161 (1986)
18. S Zhang, L Girin, Fast and accurate direct MDCT to DFT conversion with arbitrary window functions. Audio, Speech, and Language Processing, IEEE Transactions on **21**(3), 567–578 (2013)

19. AJS Ferreira, *Accurate estimation in the ODFT domain of the frequency, phase and magnitude of stationary sinusoids* (Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop, New Platz, 2001), pp. 47–50
20. A. J. Ferreira, and D. Sinha, Accurate and robust frequency estimation in the ODFT domain. (Applications of Signal Processing to Audio and Acoustics, 2005 IEEE Workshop on New Paltz, NY, 2005), pp. 16–19
21. Y Dun, G Liu, A fine-resolution frequency estimator in the odd-DFT domain. IEEE Signal Processing Letters **22**(12), 2489–2493 (2015)
22. L Daudet, M Sandler, MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction. Speech and Audio Processing, IEEE Transactions on **12**(3), 302–312 (2004)
23. S Merdjani, L Daudet, *Direct estimation of frequency from MDCT-encoded files* (Proceedings of the 6th International Conference on Digital Audio Effects, London, 2003), pp. 8–11
24. M-Y Zhu, W Zheng, D-X Li, M Zhang, An accurate low complexity algorithm for frequency estimation in MDCT domain. IEEE Trans. Consum. Electron. **54**(3), 1022–1028 (2008)
25. S Zhang, W Dou, H Yang, MDCT sinusoidal analysis for audio signals analysis and processing. Audio, Speech, and Language Processing, IEEE Transactions on **21**(7), 1403–1414 (2013)
26. Y Dun, G Liu, *An improved MDCT domain frequency estimation method* ((Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference, Xi'an, 2014), pp. 120–123