Global Journal of Researches in Engineering: J
General Engineering
Volume 15 Issue 4 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 Print ISSN:0975-5861

Investigation of Window Effects and the Accurate Estimation of Spectral Centroid

By Venkata Krishna Rao M
Vidya Jyothi Institute of Technology, India

Abstract- The spectral centroid is one of the useful low level features of a signal that was proposed for speech-music classification, speech recognition and musical instrument classification, and was also considered one of the low-level features to describe the audio content in MPEG-7 Content Description and Interface Standard. When the spectral centroid is computed from practical data, the estimate is different from the true expected theoretical value. Moreover, the behavior of the estimation error, when computed from finite length data i.e. from a short segment of signal would of high interest because most of the classification algorithms use dynamic features as the signals are nonstationary. In this paper, windowing effects on the spectral centroid estimation are investigated considering some well structured signals that appear frequently in speech and audio content. A novel algorithm is proposed to counter the window effects and better estimation of spectral centroid.

Keywords: spectral centroid, MPEG-7, sum of sine waves, band limited impulse train, STFT, peak detection.

GJRE-J Classification : *FOR Code: 291899p*

INVESTIGATIONOFWINDOWEFFECTSANDTHEACCURATEESTIMATIONOFSPECTRALCENTROID

Strictly as per the compliance and regulations of :

© 2015. Venkata Krishna Rao M. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http://creativecommons.org/licenses/by-nc/3.0/), permitting all non commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Investigation of Window Effects and the Accurate Estimation of Spectral Centroid

Venkata Krishna Rao M

Abstract- The spectral centroid is one of the useful low level features of a signal that was proposed for speech-music classification, speech recognition and musical instrument classification, and was also considered one of the low-level features to describe the audio content in MPEG-7 Content Description And Interface Standard. When the spectral centroid is computed from practical data, the estimate is different from the true expected theoretical value. Moreover, the behavior of the estimation error, when computed from finite length data i.e. from a short segment of signal would of high interest because most of the classification algorithms use dynamic features as the signals are nonstationary. In this paper, windowing effects on the spectral centroid estimation are investigated considering some well structured signals that appear frequently in speech and audio content. A novel algorithm is proposed to counter the window effects and better estimation of spectral centroid.

Keywords: Spectral Centroid, MPEG-7, Sum of Sine waves, Band Limited Impulse Train, STFT, Peak detection.

I. Introduction

The spectral centroid (SC) is one of the low level spectral domain features of a signal useful in signal classification or identification applications. The spectral centroid has been proposed by researchers in several applications like estimating the timbral brightness of music [1], for discriminating between the speech and the music [2,3,4], Speaker Recognition [5], Noisy Speech Recognition [6,7], Identification of Musical Instruments [8]. The spectral centroid was also incorporated as one of the Audio Low level features for audio content in MPEG-7 multimedia standard [9]. In [10], an AR(2) model based dynamic estimation of spectral centroid of a Narrowband Acoustic Doppler Volume Backscattering Signal was proposed.

The spectral centroid represents the "center of gravity" of the magnitude or power spectrum of a signal. Perceptually, the spectral centroid is a measure of the brightness of a sound. The unit of such a centroid would be the unit of frequency, Hz. Intuitively, the spectral centroid of a single tone signal is the frequency of the tone itself. Similarly, the spectral centroid of a signal having two equal amplitude real sinusoids is the mean frequency of two sinusoids.

Author: *Vidya Jyothi Institute of Technology, Hyderabad, India.*
e-mail: mvk_rao@hotmail.com

Mostly, the natural or real signals (e.g. speech, voice, audio, etc) are nonstationary in nature. Classification of such signals requires extraction of dynamic features that change with time. When spectral centroid is considered a promising feature, it is estimated dynamically from short segments of signal (one value of each segment), and the spectral centroid vector thus obtained for the entire signal becomes a feature vector for the classification system. The estimation of the spectral centroid from a short segment of signal data is a challenging task due to the windowing effects. In the literature, to the best of the knowledge of the author, .no systematic study results were reported on the finite data effects on the estimation of spectral centroid.

In this paper, a systematic study is carried out on the estimation of spectral centroid from finite data of different lengths. The windowing effects on the estimation error are investigated considering certain deterministic signals that appear frequently in speech and audio content. A novel algorithm is proposed to counter the finite window effects and for better estimation of spectral centroid. Well structured signals are used to make the bench marking easy, nevertheless the algorithm can be applied on any kind of real signals.

The remainder of the paper is organized as follows. The mathematical basics of spectral centroid are introduced in the section II. Short time fourier transform (STFT) for estimating the magnitude spectrum of the signal dynamically is presented in section III. The proposed algorithm along with the flowchart is discussed in section IV. Section V discusses the details of simulations and the test signals used in the simulations. Section VI presents the results and discussions on the findings. Finally conclusions on the research work are drawn in Section VII.

II. Spectral Centroid

Mathematically, the spectral centroid of a continuous time signal *y(t)* is given by

$$SC = \frac{\int_0^\infty f\, Y(f) df}{\int_0^\infty Y(f) df} \tag{1}$$

where $Y(f)$ is the one-sided magnitude spectrum of the signal *y(t)*.

© 2015 Global Journals Inc. (US)

The counter part of the discrete time signal *y(n)* is given by

$$SC = \frac{\sum_{n=0}^{N-1} n \cdot |Y(n)|}{\sum_{n=0}^{N-1} |Y(n)|} \tag{2}$$

where $Y(n)$ is the one-sided power spectrum of the signal *y(n).*

For example, the magnitude spectrum of a tone signal of unit amplitude and frequency F is an impulse at F Hz on the frequency axis. The spectral centroid of this signal is F Hz Itself. Similarly, the magnitude spectrum of a signal consisting of two tones of equal amplitude and frequencies F_1 and F_2 contains two equal amplitude impulses at F_1 Hz and F_2 Hz on the frequency axis. The spectral centroid of this signal is the mid frequency of F_1 and F_2 i.e. $(F_1 + F_2)/2$ Hz. If the amplitudes of two tones are not equal, then the spectral centroid is biased towards the higher amplitude tone. Figure 1 describes the centroid concept for several cases of F_1 and F_2 . The F_1 and F_2 values are selected as the integer multiples of 10.77Hz (44100Hz/4096) i.e. from the set {0, 10.77, 21.53, … , 5480.20, …, 11025, … , 11401.83 …, 22028.47, 22039.23, 44100/2) Hz, where 44100Hz is the sampling frequency of a CD quality audio signal. In each case, the sum of amplitudes is selected to be unity. This is to make the amplitude spectrum resemble a probability function. The figure 1(a) shows a sine wave of frequency 5840.20Hz and unity amplitude. Naturally the SC is also the same frequency 5840.20Hz. In figure 1(b) the signal consists of two sine waves of frequencies: 5840.20Hz and 11401.83Hz, and equal amplitude of 0.5. Here the SC is the mean of the two frequencies i.e. 8441.02Hz. In figure 1(c) the signal consists of two sine waves: 5840.20Hz (amp: 0.70) and 11401.83Hz (amp: 0.30). Here the SC (7256.69Hz) shifts towards the left from the mid (mean) value because the first sine wave amplitude is high. In figure 1(d) the signal consists of two sine waves: 5840.20Hz (amp: 0.15) and 11401.83Hz (amp: 0.85). In this case, the SC (10513.59Hz) shifts towards the right from the mid value. Because the second sine wave amplitude is high.

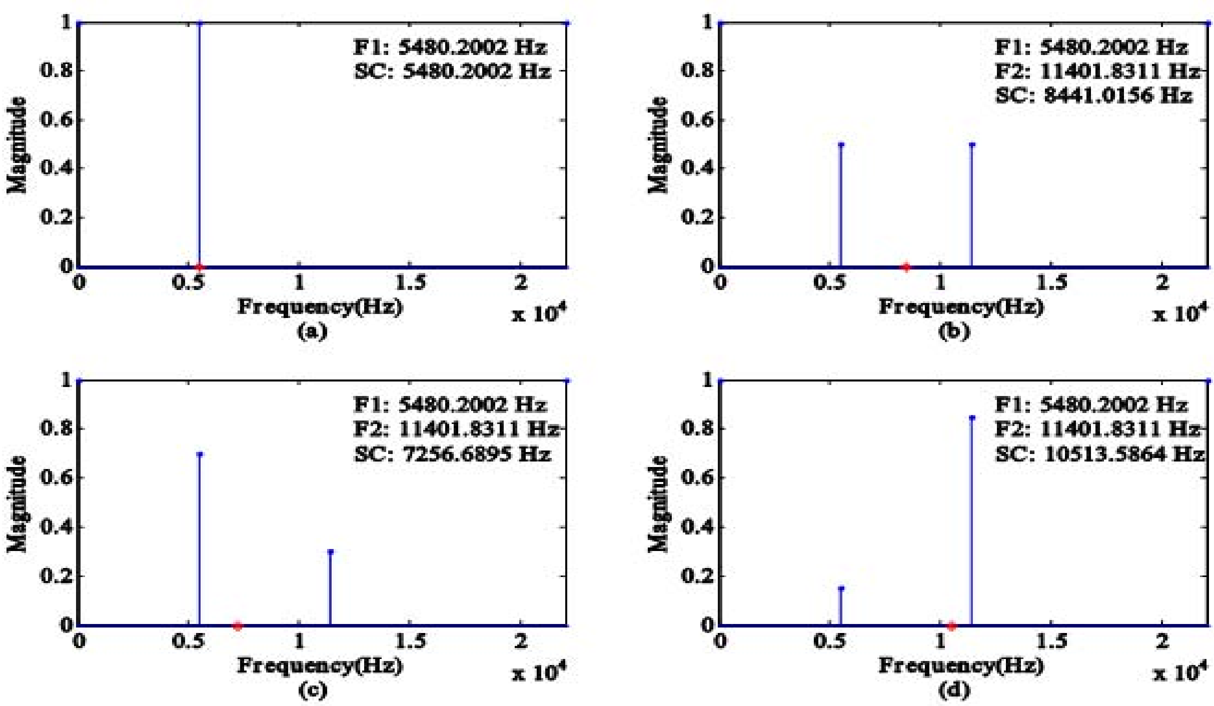

Fig. 1 : Description of Spectral Centroid. For cases of F_1 and F_2 are given in (a) through (d). In each case the sum of spectral amplitudes are selected to be unity. The spectral centroid in each case is shown as red colored star mark

III. Short Time Fourier Transform

When fourier transform is applied on short segments of data to dynamically analyze the signal, it is called short time fourier transform (STFT). To carry out the the short term analysis of a signal, the given signal *x(n)* is divided into overlapping frames of size *N*, each frame is weighed by a window function *w(k)*, typically a hamming or a hanning window and analyzed by using the Fourier Transform. A matrix is formed by arranging the short time fourier transform (STFT) coefficients as

© 2015 Global Journals Inc. (US)

columns and is popularly known as a spectrogram, given by

$$S(k,l) = \frac{1}{MW_n N}\left|\sum_{n=0}^{N-1} x\,(n+lM)\,w(n)e^{-j\frac{2\pi nk}{N}}\right|^2$$

$$0 \le k \le K-1, 0 \le l \le L-1 \qquad (3)$$

where k is the discrete frequency index, l is the time frame index, M is the hop size, K is the total number of bins of ones-sided STFT and L is the total number of frames. The spectral centroid is computed from the magnitude spectrum of each frame of signal, thus yielding a SC vector of length L., and is given by

$$SC(l) = \frac{\sum_{k=0}^{K-1} k \cdot S(k,l)}{\sum_{k=0}^{K-1} S(k,l)} \qquad 0 \le l \le L-1 \qquad (4)$$

IV. Proposed Algorithm for Spectral Centroid Estimation

The input signal data is segmented into overlapped frames of frame size (W) with 50% overlap i.e. with a hop size of W/2. For each frame, Short Time Fourier Transform (STFT) is computed using FFT algorithm with Nfft points between [0,Fs/2]. The one-sided magnitude spectrum is computed from the FFT output.

The algorithm for computing the Spectral Centroid is given in figure 2. When the steps in the dashed boxes *A*, *B* and *C* are eliminated, then the algorithm computes the spectral centroid using the equation (4) directly and it called the direct method here.

In the proposed method, a threshold *STH* is applied on the magnitude spectrum of each frame (operation: *A*) and a peak detection algorithm is applied on the spectral coefficients above the threshold (operation: *B*). Once the peaks are detected, magnitude spectrum is modified keeping only the peak values and making all other coefficients zero. The spectral centroid is then computed using this modified magnitude spectrum (operation: *C*). In this way the junk spectral coefficients (artifacts) which are produced due to finite data are get rid of from the computation process resulting in more accurate estimation of spectral centroid.

V. Simulations

The DFT spectrum is computed with 4096 points; thus for a sampling frequency of 44100Hz, the spectrum is computed with a resolution of /4096=10.76Hz and the frequency grid is (0, 10.77, 21.53, … , 11025, …, 22028.47, 22039.23, 22050)Hz.

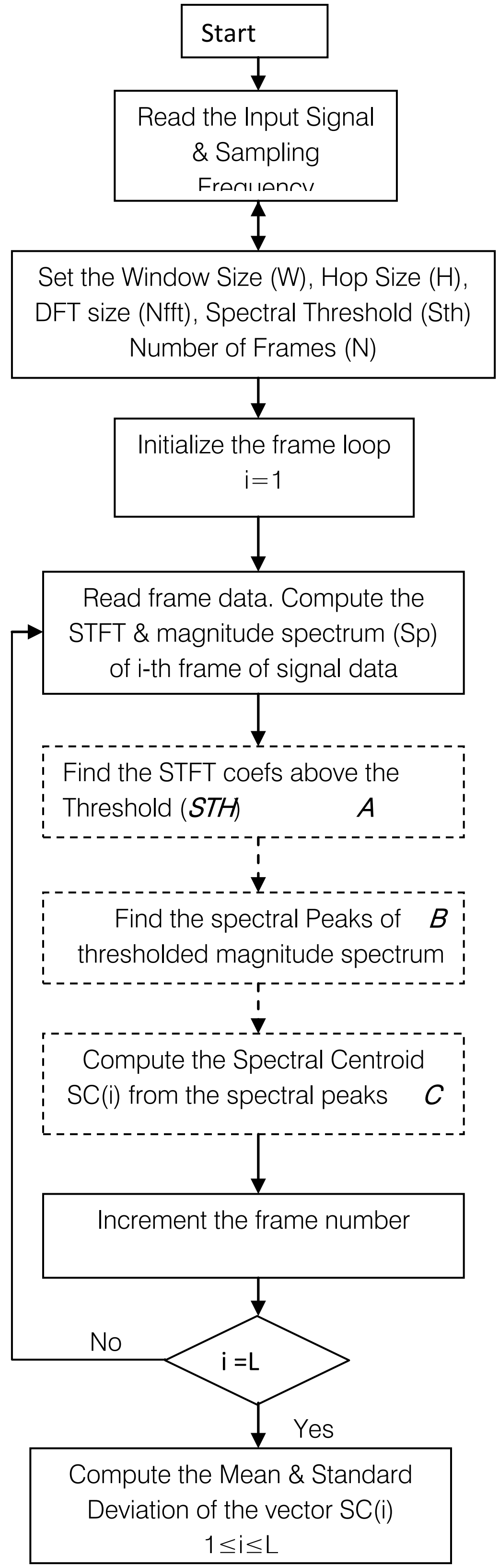

Fig. 2 : Flowchart of Proposed Algorithm for Spectral Centroid Estimation

© 2015 Global Journals Inc. (US)

The algorithm is tested on the three categories of simulated test signals:

- Tones
- Sum of Tones
- Band Limited Unit Impulse Trains

a) *Test Data Set:1 (Tones)*

In the first category, a set of 41 sine wave signals of frequencies: 96.9Hz, 635.23Hz, 1173.56Hz, … , 21091.77Hz, 21630.10Hz with a uniform spacing of 538.33Hz and random amplitudes in the range [0,1] are generated. These spot frequencies are selected so as to coincide with the DFT grid points on the frequency line (0 - Fs/2) i.e. 0Hz - 22050Hz, where Fs=44100Hz.

b) *Test Data Set:2 (Sum of Tones)*

In the second category, a sum of 5 or 10 or 50 sine waves of distinct frequencies are generated. In each case, the sine waves are separated with a uniform spacing of 10.76 Hz or 96.90Hz or 495.26Hz. These spacing are selected so as the generated frequencies coincide with the DFT grid points. In each set of 5 or 10 or 50 frequencies. the first frequency is taken from one of the 41 spot frequencies of the first category, the total number of composite signals generated under this category is 41 x 3 x 3 =369.

c) *Test Data Set:3 (Band Limited Unit Impulse Trains)*

In the third category, a set of Band Limited Unit Impulse Trains (BLUITs) each with a different fundamental frequency is generated. The frequencies of 41 sine waves of first category are used as fundamentals, thus we get 41 sets of BLUITs. The spectral envelope of each BLUIT can be constant (i.e. 0dB/Octave) or decay at a rate of 12dB/Octave. The Fundamental frequencies and number of harmonics in each BLUIT (=0.5 F_s/F_0) are given in the table 2. The number of harmonics for the BLUIT nos: 21-41 is one i.e. the fundamental itself and hence not considered in the simulations and hence are not listed in the table 2.
As there are two cases of dB/octave rates, a total of 2 x 20 = 40 BLUITs form this category of test signals are generated.

Thus the total data set comprises 450 (=41 + 369 + 40) differently structured test signals.

VI. Results

In this section, the results obtained by applying both the direct and proposed methods are presented. The performance comparison of both the methods is also given.

Table 1 : Frequencies of Band Limited Impulse Train used in evaluating the proposed algorithm

BLUIT no	Fundamental Frequency	Number of Harmonics
1	96.90	227
2	635.23	34
3	1173.56	18
4	1711.89	12
5	2250.22	9
6	2788.55	7
6	3326.88	6
7	3865.21	5
8	4403.54	5
9	4941.87	4
10	5480.20	4
11	6018.53	3
12	6556.86	3
13	7095.19	3
14	7633.52	2
15	8171.85	2
16	8710.18	2
17	9248.51	2
18	9786.84	2
19	10325.17	2
20	10863.50	2

The SC estimation results of Test Set-1 (Tones) signals of frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration (hamming window size is 512, Fs=44100Hz) for both direct and proposed methods are given in Table.2. Each row in the table 2 corresponds to the estimated SC vector of a particular tone frequency of duration 0.5 seconds of full length signal corresponding to a total of 22050 samples. Both the mean (μ) and standard deviation (σ) of this estimated spectral centroid vector is computed and given in the 3rd column of the table 2.

The estimated errors for direct method are large at both the lowest and the highest frequencies in the range. For the lowest (start) frequency the error is negative and for the highest (end) frequency it is positive. It means the direct method over estimates the SC at lower frequencies and under estimates at the higher frequencies. This is because of the fact that for lower frequencies, the spectral mass distribution on either side of the tone frequency is unevenly distributed and is more on the right (higher frequency) side.Hence, the estimated values shift towards the higher side of the frequency axis.

Similarly, for higher frequencies, the estimated values shift towards the lower side of the frequency axis. As the frequency of the tone is spanned from the lowest frequency (96.8994Hz) to the highest frequency 21630.1025Hz), the mean error (μ) reduces and becomes zero at the middle of the range i.e. at tone frequency approximately equal to Fs/4. At this frequency, the mean error changes its sign from negative to positive value, builds up and again reaches its maximum at the highest frequency (please see the 4th

© 2015 Global Journals Inc. (US)

column of the table 2). For each tone, the standard deviation (σ) is also computed.

The estimation results of the proposed method for the same set of signals are given in the 5th and the 6th columns of table 2. This method exactly estimates the SC and hence both the mean (μ) and standard deviation (σ) are zeros. The spectral threshold *STH* is chosen as the 0.02 fraction of the maximum value of the magnitude spectrum, which corresponds to about -14 dB down the peak value. This is approximately the side lobe level (SLL) of the spectrum of rectangular window. For other windows the SLL is always less than -13dB, though the main lobe width is more compared to that of a rectangular window, which anyway does not affect the peak detection process.

The estimation results of table 2 are also shown in figure 3(a) for both direct (solid line) and proposed (dashed line) methods are shown. For direct method, the RMS range of the estimated Centroid is marked as red vertical lines at each point. For the proposed method the estimated value is exactly equal to true value, hence the RMS range is zero. Thus no red vertical lines are seen on the dashed line. The figure (b) shows the similar results for window size is 256.

Table 2 : Spectral Centroid of Test set-1 (Tones) signals estimated by direct and proposed methods

Tone no	True Spectral Centroid (Hz) (1)	Spectral Centroid (Estimated by Direct Method)(Hz) (2)	SC Est. Error (Direct Method) (Hz) (1) – (2)	Spectral Centroid (Estimated by Proposed Method) (Hz) (3)	SC Est. Error (Proposed Method) (Hz) (1) – (3)
1	96.8994	634.033 ± 103.2825	-537.13	96.8994 ± 0	0
2	635.2295	1107.9746 ± 114.8693	-472.75	635.2295 ± 0	0
3	1173.5596	1608.3049 ± 107.8578	-434.75	1173.5596 ± 0	0
4	1711.8896	2117.5429 ± 99.1608	-405.65	1711.8896 ± 0	0
5	2250.2197	2623.4037 ± 94.3929	-373.18	2250.2197 ± 0	0
6	788.5498	3131.4122 ± 89.3034	-342.86	2788.5498 ± 0	0
7	3326.8799	3642.626 ± 86.5037	-315.75	3326.8799 ± 0	0
8	3865.21	4158.9111 ± 81.3724	-293.7	3865.21 ± 0	0
9	4403.54	4669.3952 ± 79.7196	-265.86	4403.54 ± 0	0
10	4941.8701	5183.2597 ± 76.1674	-241.39	4941.8701 ± 0	0
11	5480.2002	5698.366 ± 75.0958	-218.17	5480.2002 ± 0	0
12	6018.5303	6217.31 ± 71.2696	-198.78	6018.5303 ± 0	0
13	6556.8604	6730.1438 ± 70.2138	-173.28	6556.8604 ± 0	0
14	7095.1904	7247.1104 ± 67.5403	-151.92	7095.1904 ± 0	0
15	7633.5205	7764.439 ± 66.7206	-130.92	7633.5205 ± 0	0
16	8171.8506	8283.9504 ± 64.5743	-112.1	8171.8506 ± 0	0
17	8710.1807	8797.6702 ± 64.3643	-87.49	8710.1807 ± 0	0
18	9248.5107	9316.0265 ± 63.4916	-67.52	9248.5107 ± 0	0
19	9786.8408	9834.3672 ± 63.363	-47.53	9786.8408 ± 0	0
20	10325.1709	10353.7776 ± 62.7943	-28.61	10325.1709 ± 0	0
21	10863.501	10867.8735 ± 62.924	-4.37	10863.501 ± 0	0
22	11401.8311	11387.1281 ± 63.1554	14.7	11401.8311 ± 0	0
23	11940.1611	11905.6919 ± 63.1063	34.47	11940.1611 ± 0	0
24	12478.4912	12424.4625 ± 63.3744	54.03	12478.4912 ± 0	0
25	13016.8213	12938.3795 ± 63.3481	78.44	13016.8213 ± 0	0
26	13555.1514	13457.8904 ± 64.7948	97.26	13555.1514 ± 0	0
27	14093.4814	13975.7026 ± 65.3201	117.78	14093.4814 ± 0	0
28	14631.8115	14493.3087 ± 67.2566	138.5	14631.8115 ± 0	0
29	15170.1416	15006.592 ± 67.9661	163.55	15170.1416 ± 0	0
30	15708.4717	15525.9341 ± 71.2219	182.54	15708.4717 ± 0	0
31	16246.8018	16042.2271 ± 72.4166	204.57	16246.8018 ± 0	0
32	16785.1318	16557.6751 ± 75.7913	227.46	16785.1318 ± 0	0
33	17323.4619	17069.2012 ± 76.9287	254.26	17323.4619 ± 0	0
34	17861.792	17586.7432 ± 81.2453	275.05	17861.792 ± 0	0
35	18400.1221	18099.7754 ± 83.0062	300.35	18400.1221 ± 0	0
36	18938.4521	18610.6458 ± 87.3431	327.81	18938.4521 ± 0	0
37	19476.7822	19118.961 ± 90.5547	357.82	19476.7822 ± 0	0
38	20015.1123	19632.1551 ± 97.5527	382.96	20015.1123 ± 0	0
39	20553.4424	20138.8757 ± 102.7702	414.57	20553.4424 ± 0	0
40	21091.7725	20639.2884 ± 109.3145	452.48	21091.7725 ± 0	0
41	21630.1025	21136.0545 ± 116.1857	494.05	21630.1025 ± 0	0

© 2015 Global Journals Inc. (US)

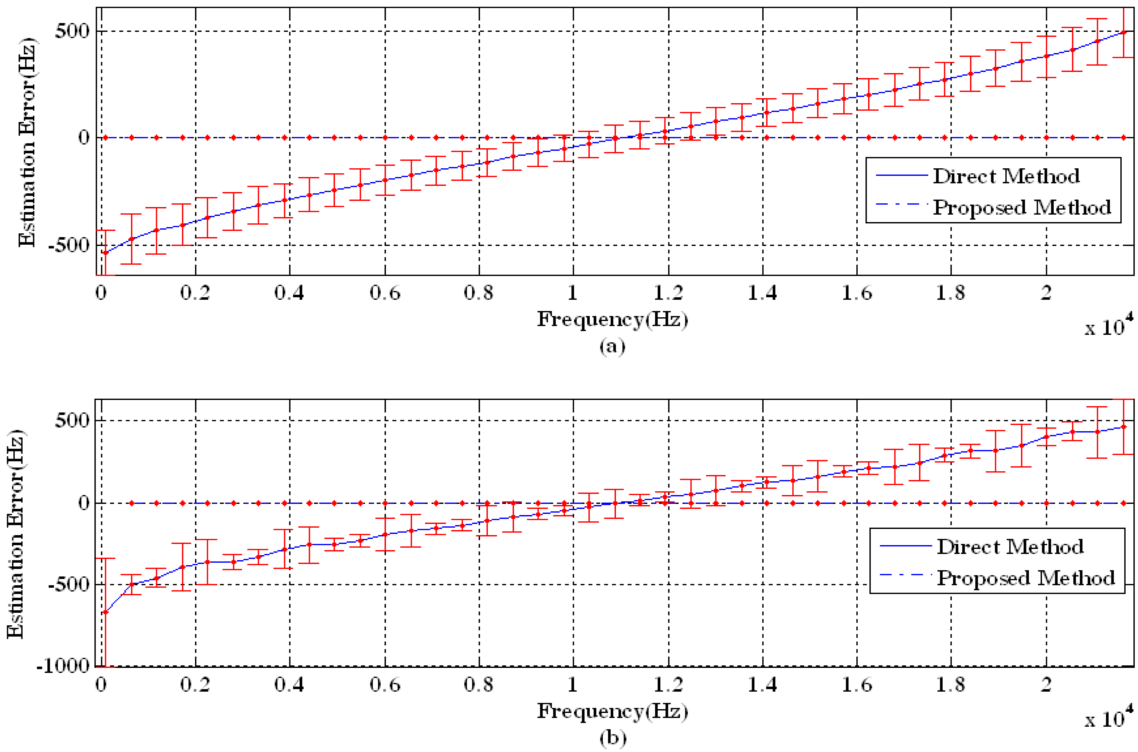

Fig. 3 : SC Estimation Error of Test set: 1 (tone) signals of frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration (a) for window size of 512). (b). for window size of 256

The estimation error follows a regular pattern for window size of 512 sample compared to the error for 256 sample window. This is due to the fact that the data has become too short to get a meaningful estimate. However, the error is almost symmetric around the middle frequency i.e. Fs/4. This symmetry would be disturbed if the window size is further reduced. The error becomes more for lower frequencies, as more number of cycles of the signal are not included in the short segment. So the window size is to be carefully selected based on the lowest frequency under consideration so that considerable number of signal cycles are included in the window. The figure 4 provides magnitude spectrum of a single frame of tone signals of frequencies: 96.8994 Hz, 10863.501Hz and 21630. 1025Hz (on the left side) and the corresponding estimated spectral centroid vectors (on the right side). The estimation errors (i.e. true SC - mean of estimated SC vector) are -537.13Hz, -4.37 (almost zero) and +494.05Hz for the three tone frequencies. Similar plots for window size of 256 samples are shown in figure 5.

© 2015 Global Journals Inc. (US)

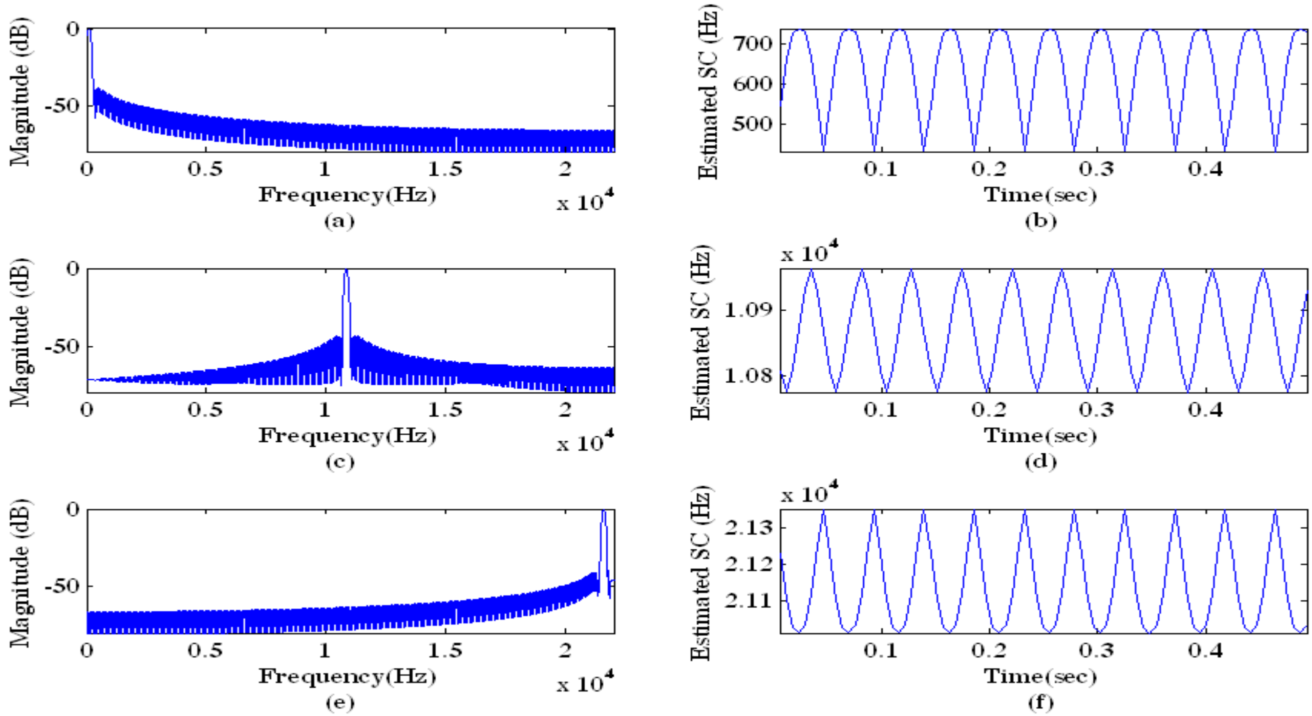

Fig. 4 : Magnitude spectrum of a single frame of tone signals of frequencies: 96.8994 Hz, 10863.501Hz and 21630.1025Hz on the left side (a), (c) and (e) for window length of 512 samples. Corresponding estimated spectral centroid vectors on the right side (b), (d) and (f)

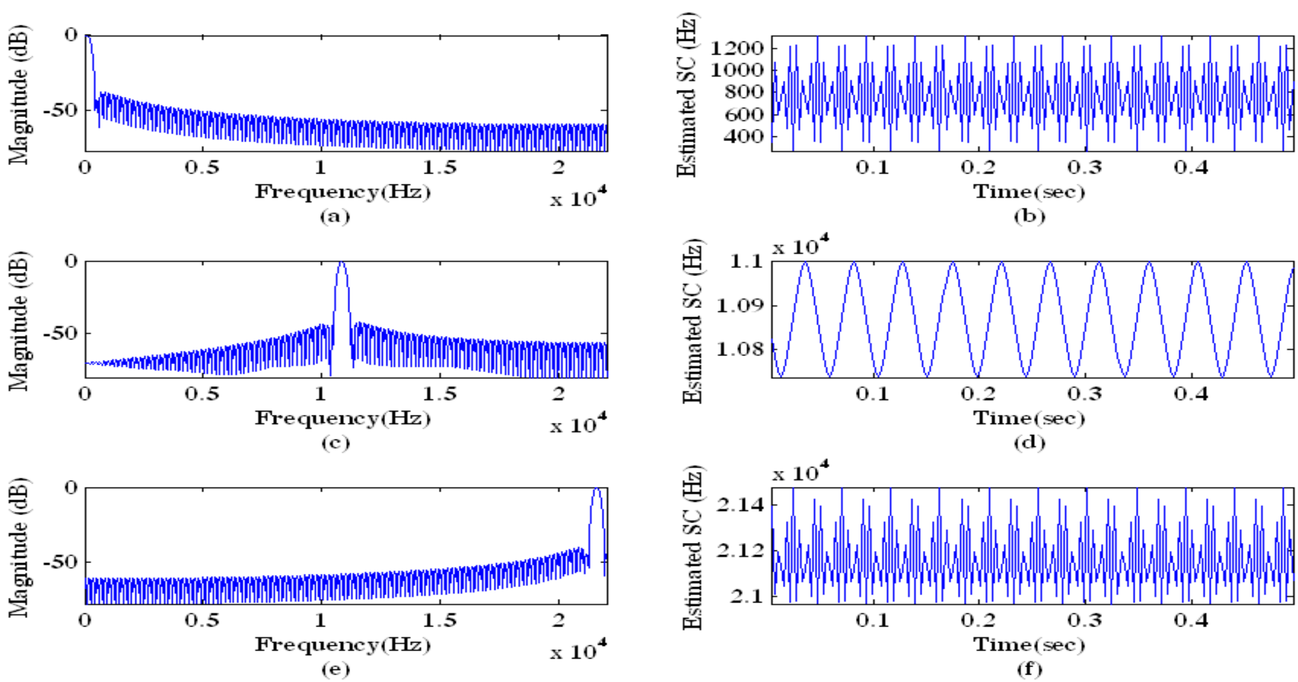

Fig. 5 : Magnitude spectrum of a single frame of tone signals of frequencies: 96.8994 Hz, 10863.501Hz and 21630.1025Hz on the left side (a), (c) and (e) for window length of 256 samples. Corresponding estimated spectral centroid vectors on the right side (b), (d) and (f)

The results of Test set: 2 (sum of Tones) with a tone spacing of 200Hz are shown in figure 6 for (a). 512 sample window and (b) 256 sample window. The results of Test set: 2 with a tone spacing of 100Hz are shown in figures 7 for (a). 512 sample window and (b) 256 sample window. Similarly, figure 8 gives the results of Test set: 2 for a tone spacing of 500Hz for 512 and 256 sample windows.

© 2015 Global Journals Inc. (US)

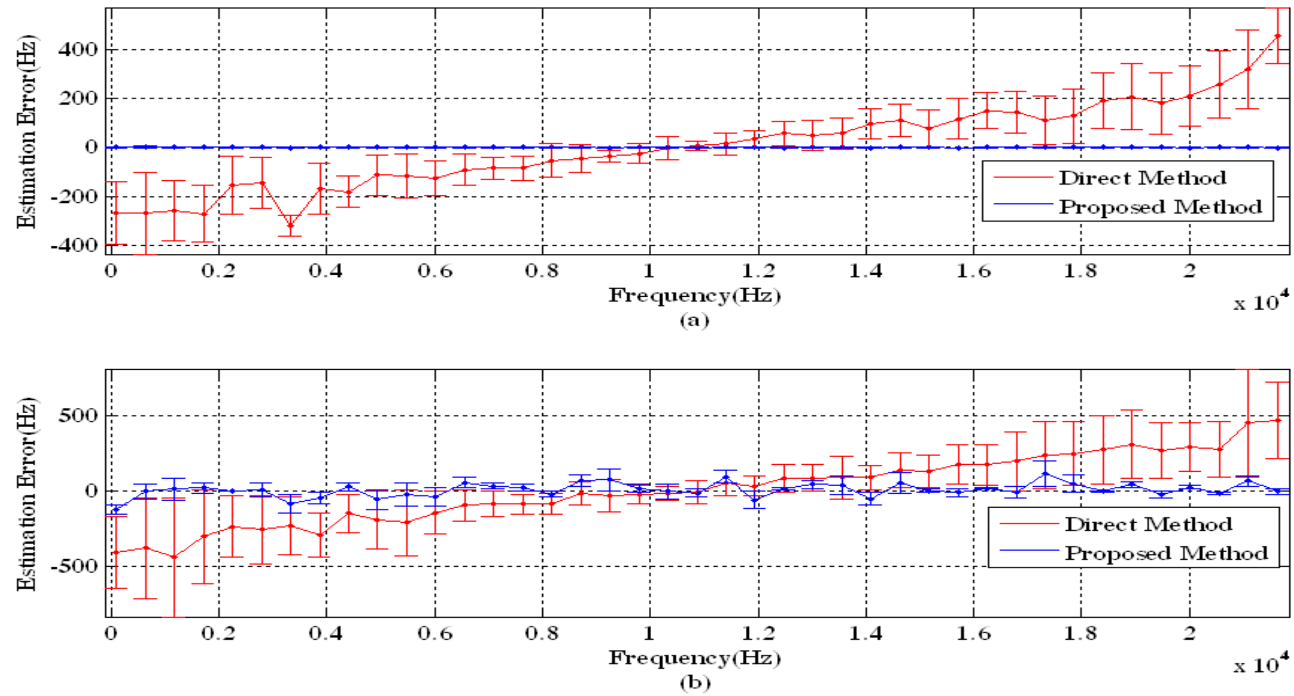

Fig. 6 : (a). SC Estimation Error of Test set: 2 (sum of tones with a frequency spacing of 200 Hz) signals of lowest frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration (window size is 512) for both direct (solid line) and proposed (dashed line) methods. (b). Same as (a) for window size is 256

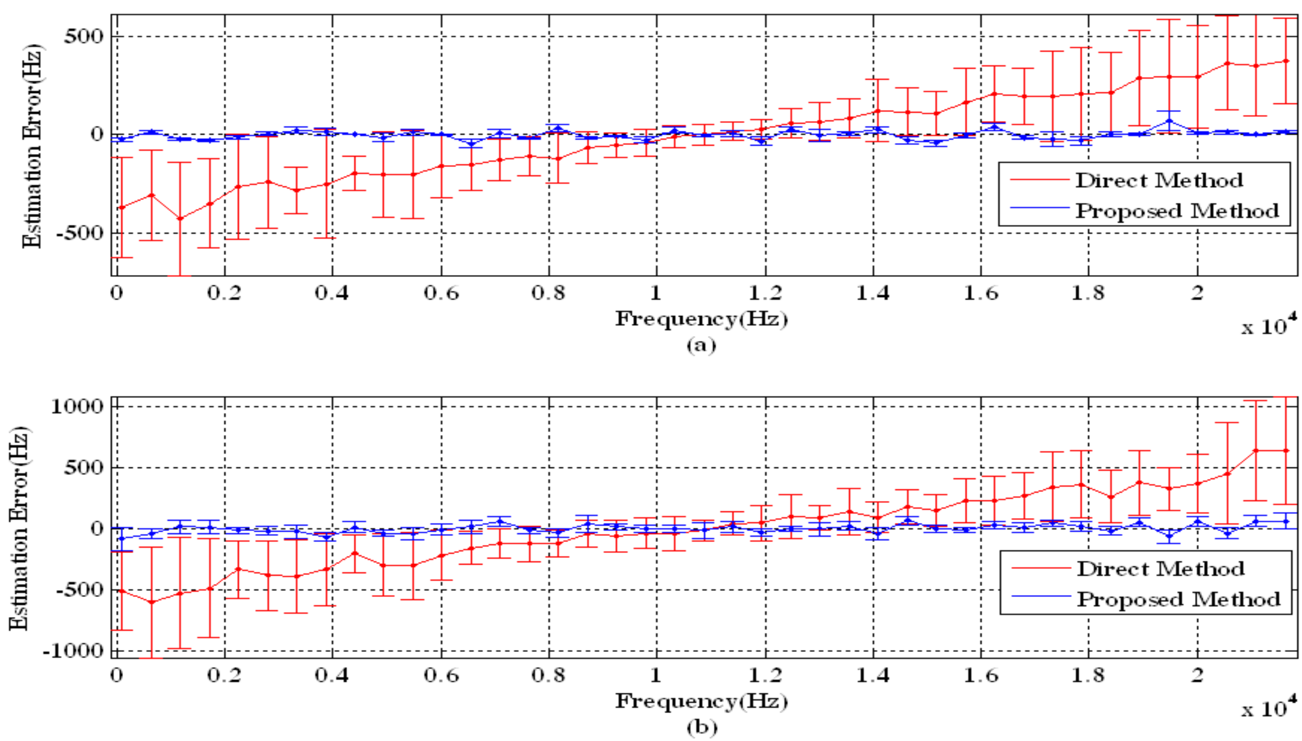

Fig. 7 : (a). SC Estimation Error of Test set: 2 (sum of tones with a frequency spacing of 100 Hz) signals of lowest frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration (window size is 512) for both direct (solid line) and proposed (dashed line) methods. (b). Same as (a) for window size is 256

© 2015 Global Journals Inc. (US)

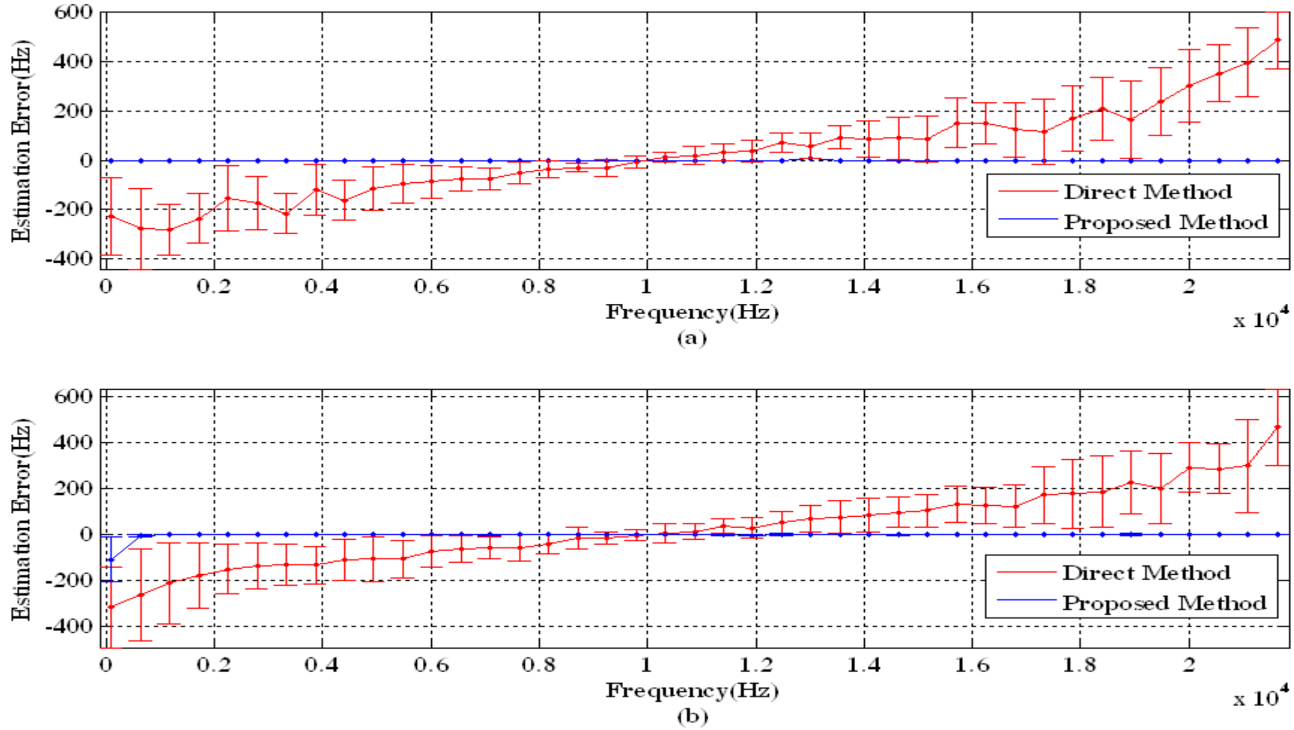

Fig. 8 : *(a).* SC Estimation Error of Test set: 2 (sum of tones with a frequency spacing of 500 Hz) signals of lowest frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration (window size is 512) for both direct (solid line) and proposed (dashed line) methods. *(b).* Same as (a) for window size 256

The results say that the estimation using the proposed is always better than that of the direct method. The accuracy is extremely well for larger spacing of tone frequencies, the reason being the better separation of. spectral peaks.

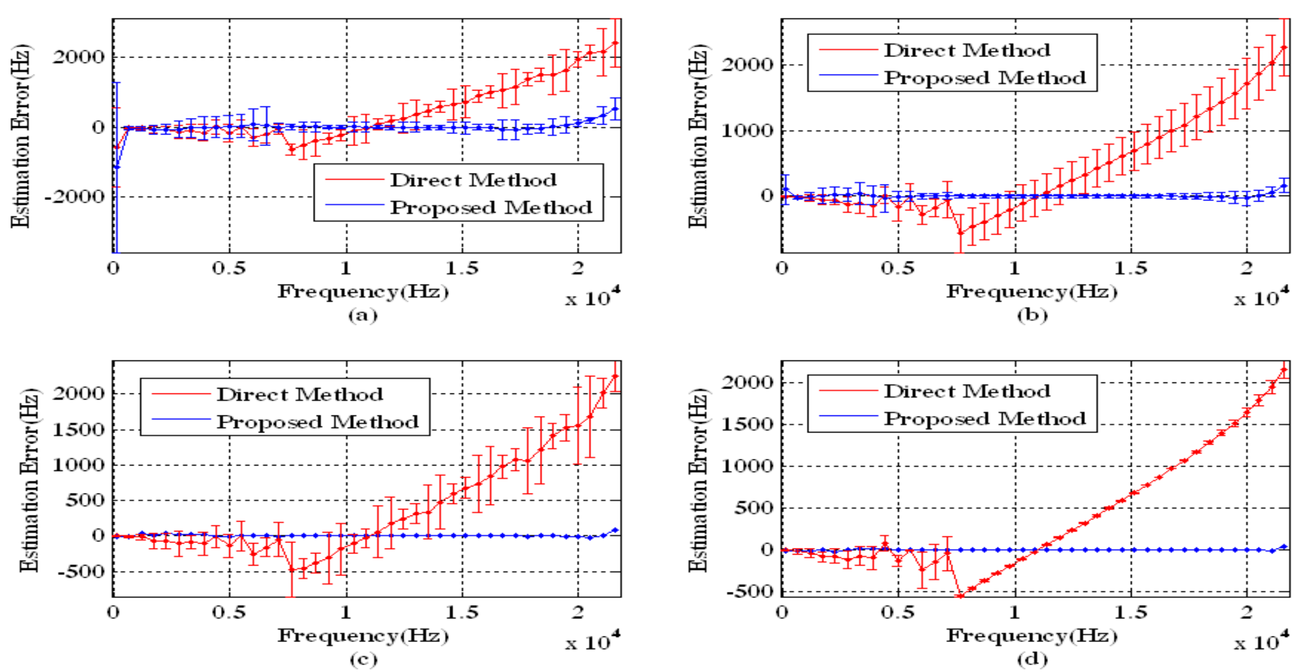

Fig. 9 : SC Estimation Error of Test set: 3 (BLUITs with a fundamental frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration; spectral slope 0 dB/Octave) for both direct (red line) and proposed (blue line) methods for (a). 256 sample window (b). 512 sample window (c). 768 sample window (d). 1024 sample window

© 2015 Global Journals Inc. (US)

Figure 9: shows the estimation results for Test set: 3 (BLUITs) with a fundamental frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration and spectral slope of 0 dB/Octave) for window sizes of 256, 512 768 and 1024 samples. Again results are extremely well for proposed method compared to those of the direct method, while the direct method fails even for larger window sizes. In figure 10, the estimation errors for Test set: 3 (BLUITs) signals of spectral slope of –12dB/Octave are shown for window sizes of 256, 512, 768 and 1024 samples. It can be observed that in all cases, mean error drastically low compared to that of direct method. More over, as the window length increases, the standard deviation of estimation error reduces faster for the proposed method compared to that of the direct method. (first two lines are rearranged properly)

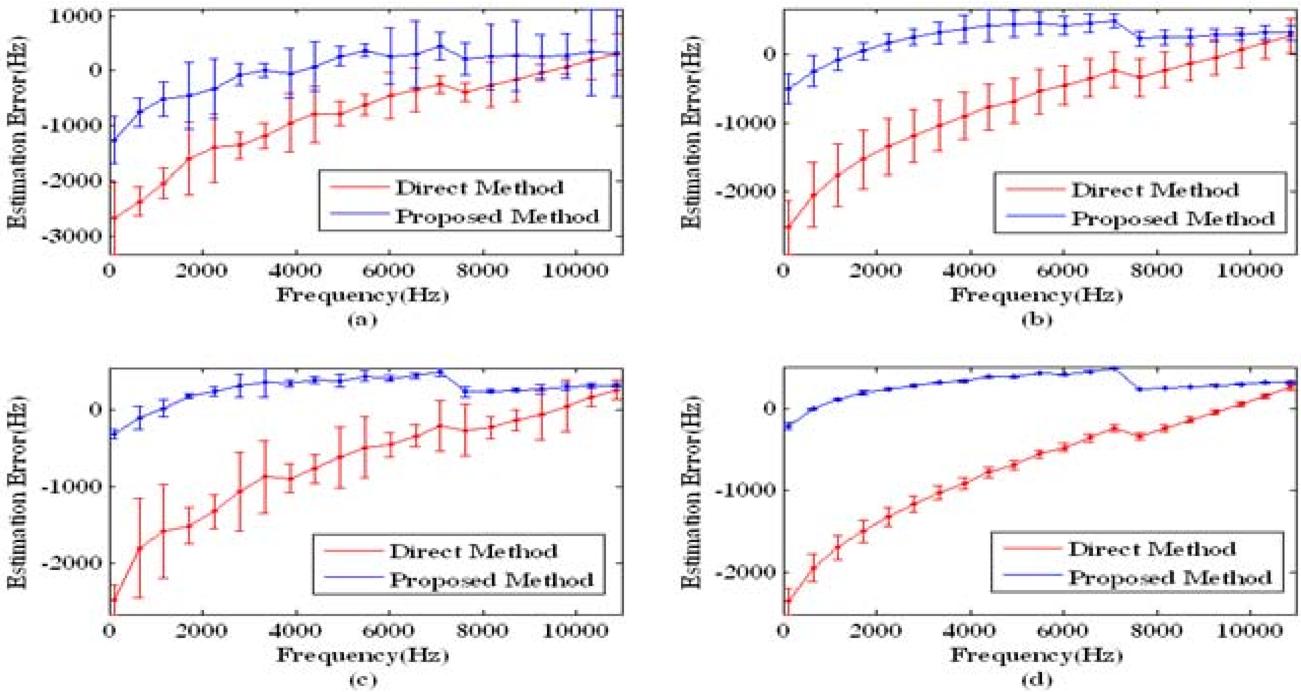

Fig. 10 : SC Estimation Error of Test set: 3 (BLUITs with a fundamental frequency spanning from 96.8994 Hz to 21630.1025Hz of 0.5 sec duration; spectral slope -12 dB/Octave) for (a). 256 sample window (b). 512 sample window (c). 768 sample window (d). 1024 sample window

VII. Conclusions

In this paper, windowing effects on the spectral centroid estimation are investigated considering three types of well structured signals: Tones, Sum of Tones and Band Limited Unit Impulse Trains. These test signals are considered because they appear frequently in speech and audio content. The spectral centroid is estimated using two methods: (1). the direct method using the equation 4. (2). The proposed method that uses threshold and peak detection on the magnitude spectrum. The proposed algorithm is shown to estimate the spectral centroid more accurately compared to direct method for all the signals under consideration and for all window lengths.

References Références Referencias

1. Emery Schubert and Joe Wolfe, "Does Timbral Brightness Scale with Frequency and Spectral Centroid?", Vol. 92, Acta Acustica United With Acustica, pp.820 – 825, 2006.
2. E. Scheier and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," Proc. IEEE ICASSP, 1997.
3. E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio," IEEE Multimedia Mag., vol. 3, pp. 27–36, Fall 1996.
4. Peeters, G., Burthe, A. L. and Rodet, X., "Toward automatic music audio summary generation from signal analysis", Proceedings of the Third International Conference on Music Information Retrieval, pp. 94-100, 2002, Paris, France.
5. Jia Min Karen Kua et. Al., "Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition", The Speaker and Language Recognition Workshop, 28 June – 1 July 2010, pp.34-39, Brno, Czech Republic.
6. Jingdong Chen, et. Al., "Recognition of Noisy Speech Using Dynamic Spectral Subband Centroids", IEEE Signal Processing Letters, Vol. 11, No. 2, pp. 258-261., February 2004.

© 2015 Global Journals Inc. (US)

7. Bojana Gajic´ and Kuldip K. Paliwal, "Robust Speech Recognition in Noisy Environments Based on Subband Spectral Centroid Histograms", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 14, No. 2, pp.600-608, March 2006.
8. M. Chandwadkar and M. S. Sutaone, "Selecting Proper Features and Classifiers for Accurate Identification of Musical Instruments", International Journal of Machine Learning and Computing, Vol. 3, No. 2, pp.172-175, April 2013.
9. B. S. Manjunath (editor), et. Al., "Introduction to MPEG-7", Wiley, 1st edition, 2002.
10. Xiao- Jiao Tao et. Al., "Narrowband Acoustic Doppler Volume Backscattering Signal–Part II: Spectral Centroid Estimation", IEEE Transactions On Signal Processing, Vol. 50, NO. 11, pp.2656-2660, November 2002.

© 2015 Global Journals Inc. (US)

Global Journals Inc. (US) Guidelines Handbook 2015

www.GlobalJournals.org