Investigation of Window Effects and the Accurate Estimation of Spectral Centroid

Venkata Krishna Rao M[1]

[1] Vidya Jyothi Institute of Technology

Received: 10 February 2015 Accepted: 1 March 2015 Published: 15 March 2015

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 lowlevel 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.

Index terms— spectral centroid, MPEG-7, sum of sine waves, band limited impulse train, STFT, peak detection.

1 Introduction

he 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.

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.

2 II.

3 Spectral Centroid

Mathematically, the spectral centroid of a continuous time signal y(t) is given by???? = ? ð ??"ð ??" ??(ð ??"ð ??")??ð ??"ð ??" ? 0 ? ??(ð ??"ð ??")??ð ??"ð ??" ? 0(**1**)

where ??(ð ??"ð ??") is the one-sided magnitude spectrum of the signal y(t).

The counter part of the discrete time signal y(n) is given by???? = ? ?? ? |??(??)| ???1 ??=0 ? |??(??)| ???1 ??=0(**2**)

where ??(??) 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 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

4 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 ??(??, ??) = 1 ???? ?? ?? ?? ?? (?? + ????) ??(??)?? ??? 2?? ???? ?? ???1 ??=0 ? 2 0 ? ?? ? ?? ? 1, 0 ? ?? ? ?? ? 1 (**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????(??) = ? ?? ? ??(??, ??) ???1 ??=0 ? ??(??, ??) ???1 ??=0 0 ? ?? ? ?? ? 1 IV.

5 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 onesided magnitude spectrum is computed from the FFT output.

The algorithm for computing the Spectral Centroid is given in figure **??**. 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.

6 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,

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.

7 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.

8 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

9 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. 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 (?) of this estimated spectral centroid vector is computed and given in the 3 rd 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. 2). For each tone, the standard deviation (?) is also computed. The estimation results of the proposed method for the same set of signals are given in the 5 th and the 6 th columns of table 2. This method exactly estimates the SC and hence both the mean (µ) and standard deviation (?) 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. 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 **??** 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.

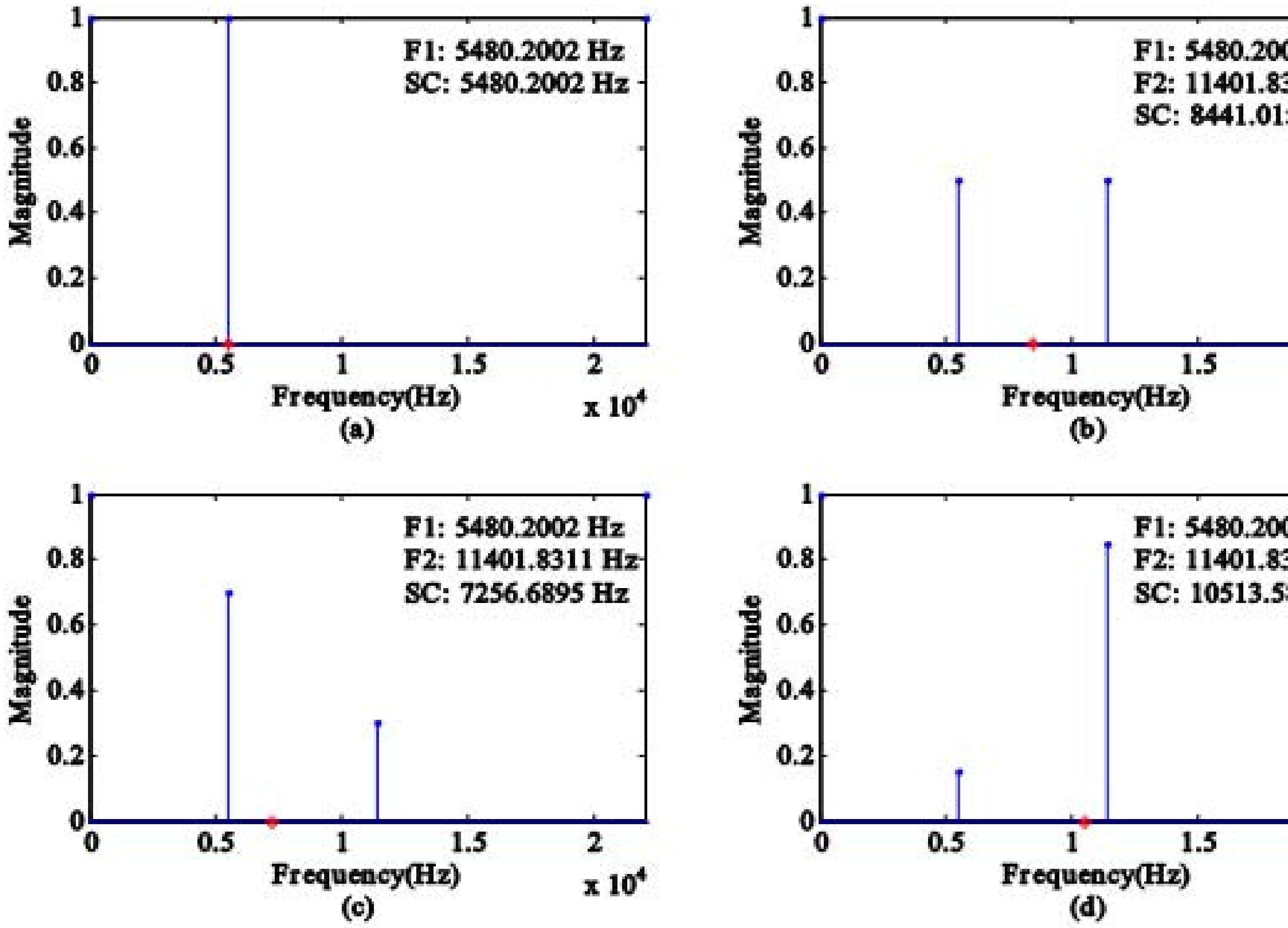
10 Conclusions

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

Figure 1:

more accurately compared to direct method for all the signals under consideration and for all window lengths.

[1] [2]



1

Figure 2: Fig. 1 :

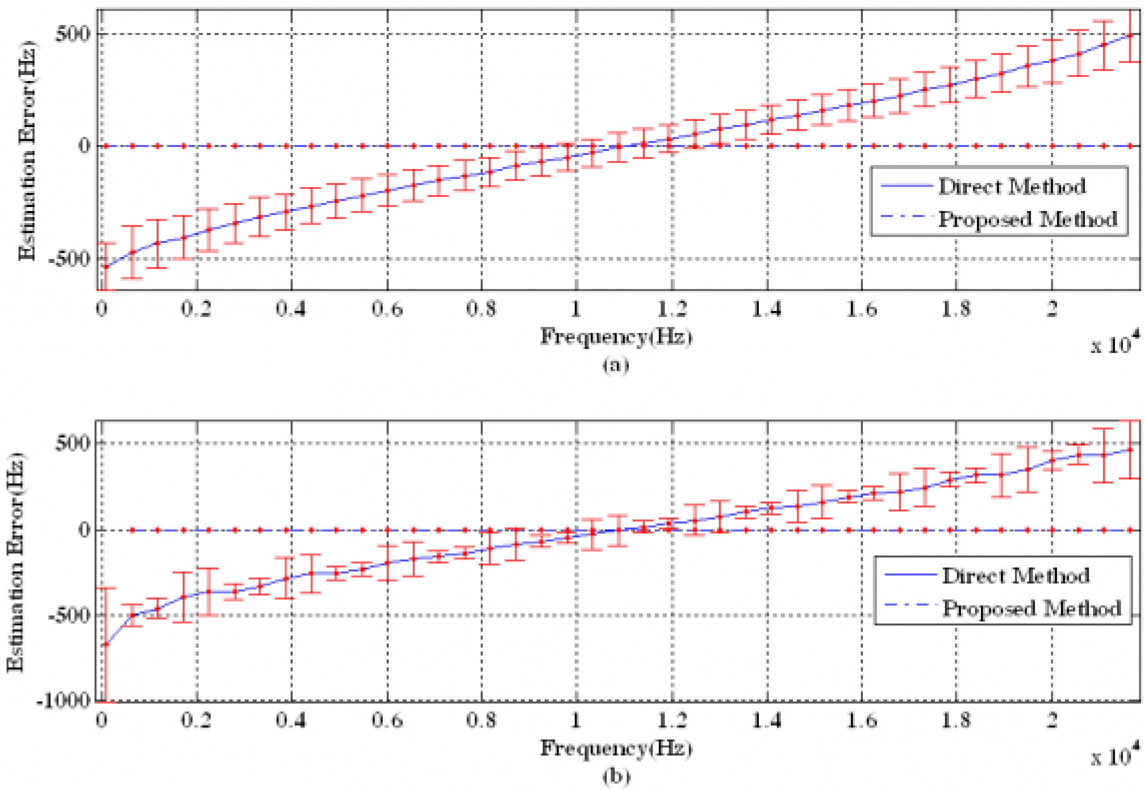
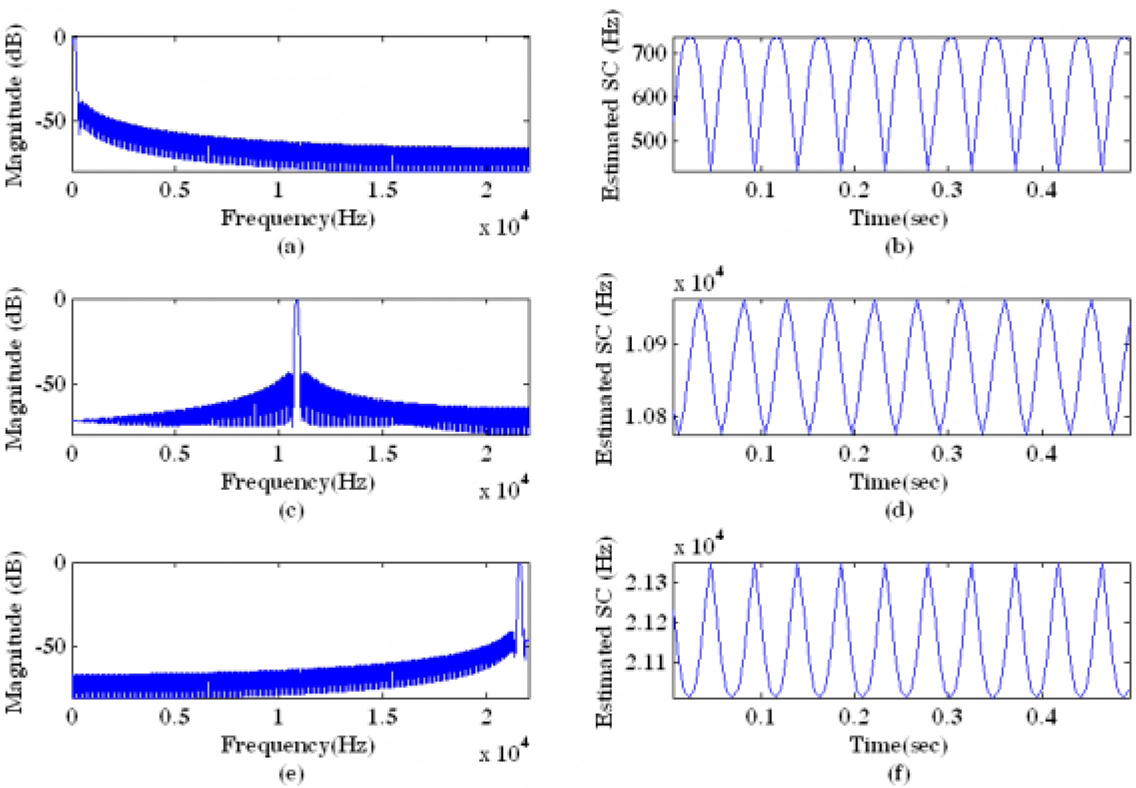
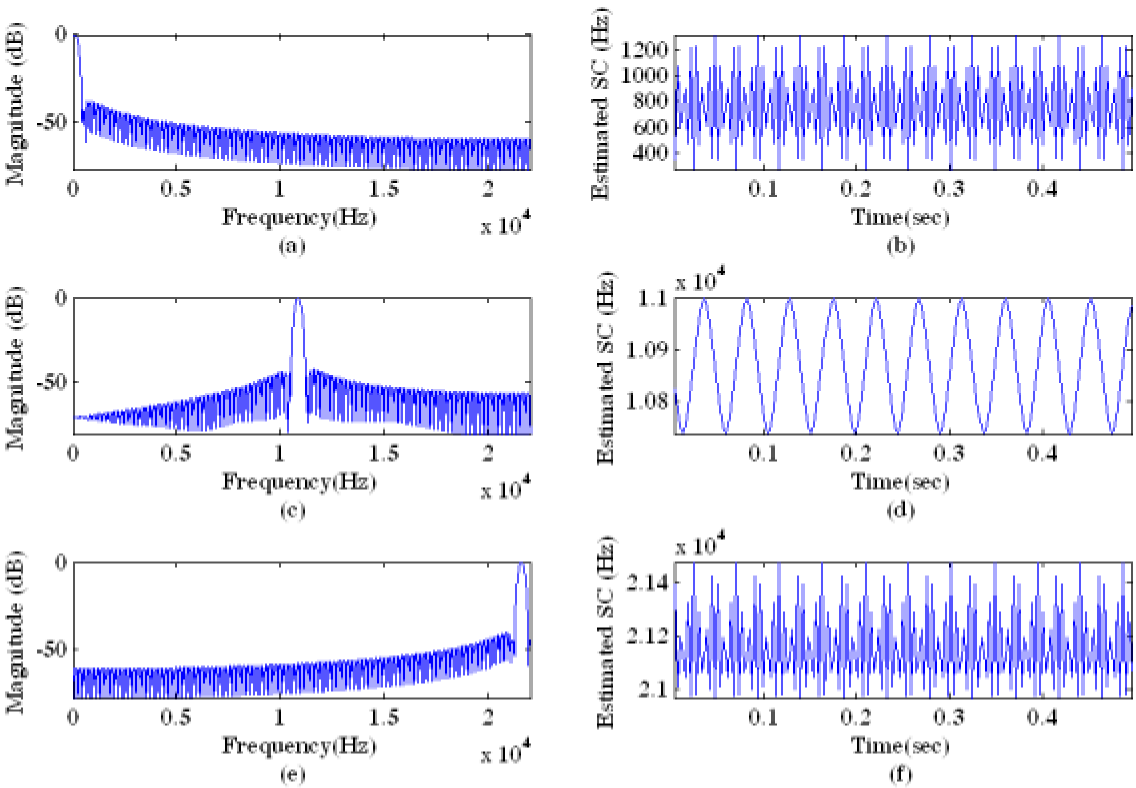


Figure 3:



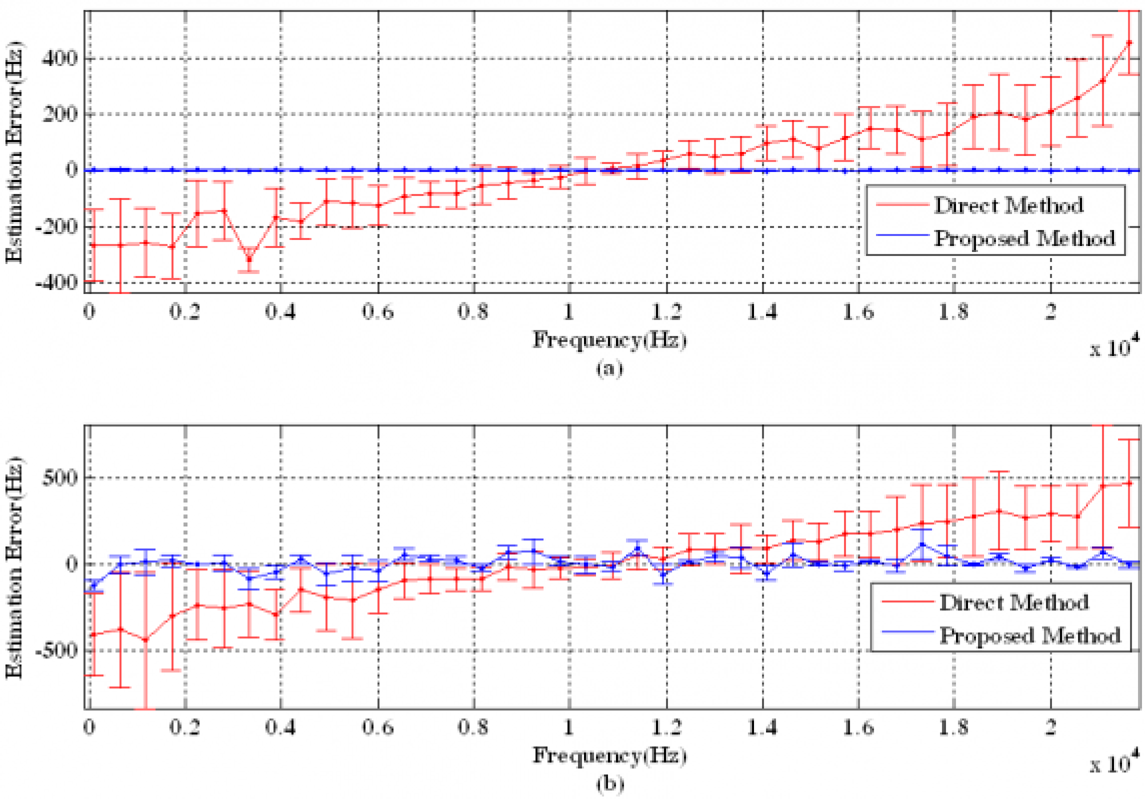
3

Figure 4: Fig. 3 :



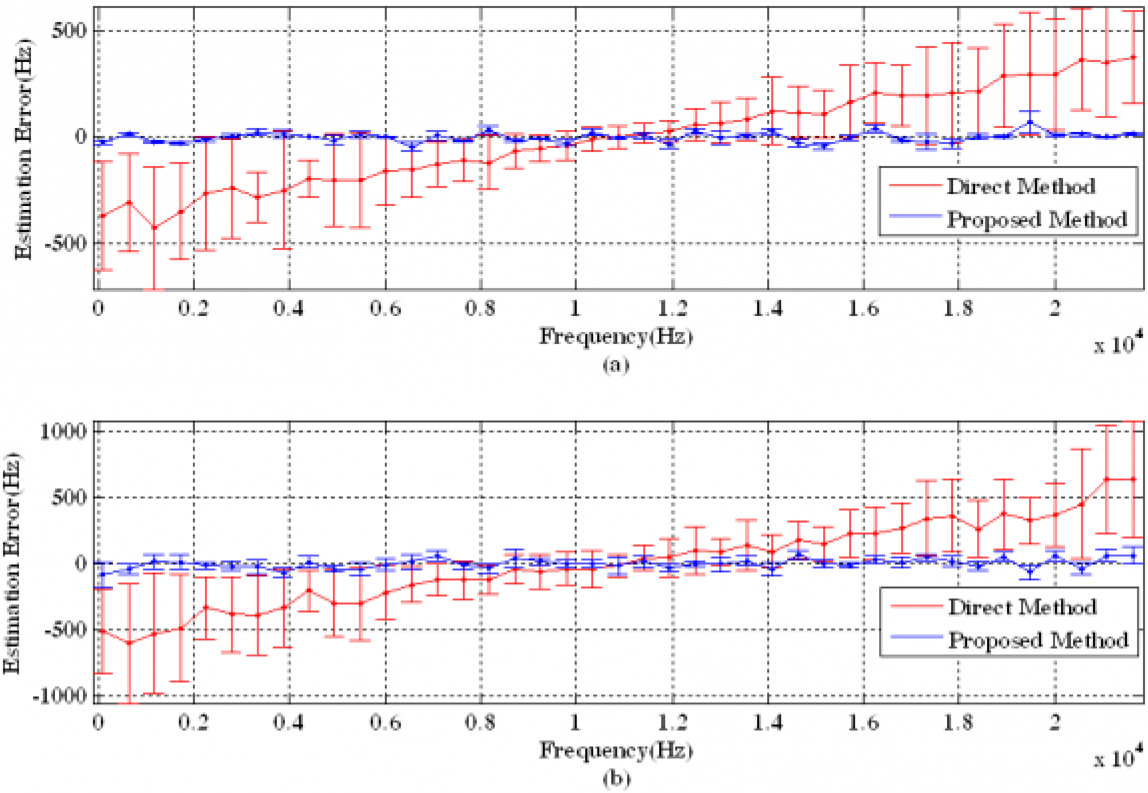
45

Figure 5: Fig. 4 :Fig. 5 :



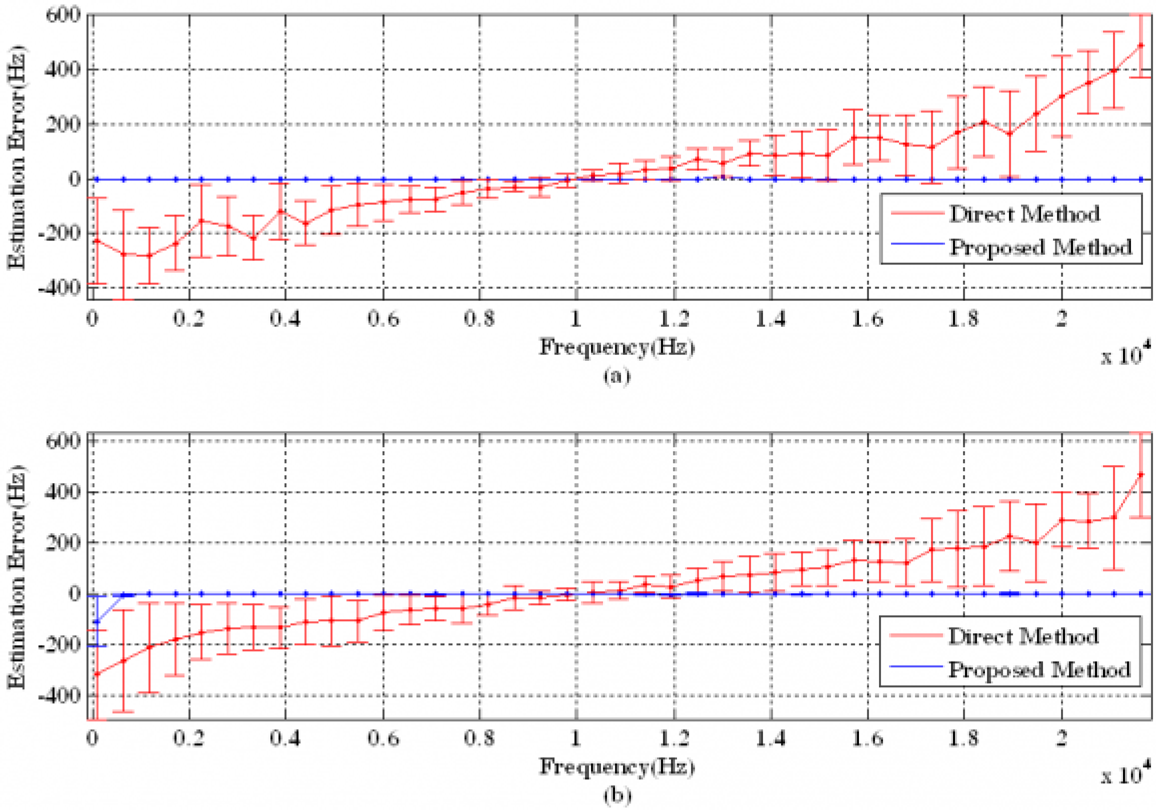
678

Figure 6: Fig. 6 :Fig. 7 :Fig. 8 :



9

Figure 7: Fig. 9 :



910

Figure 8: Figure 9 :Fig. 10 :

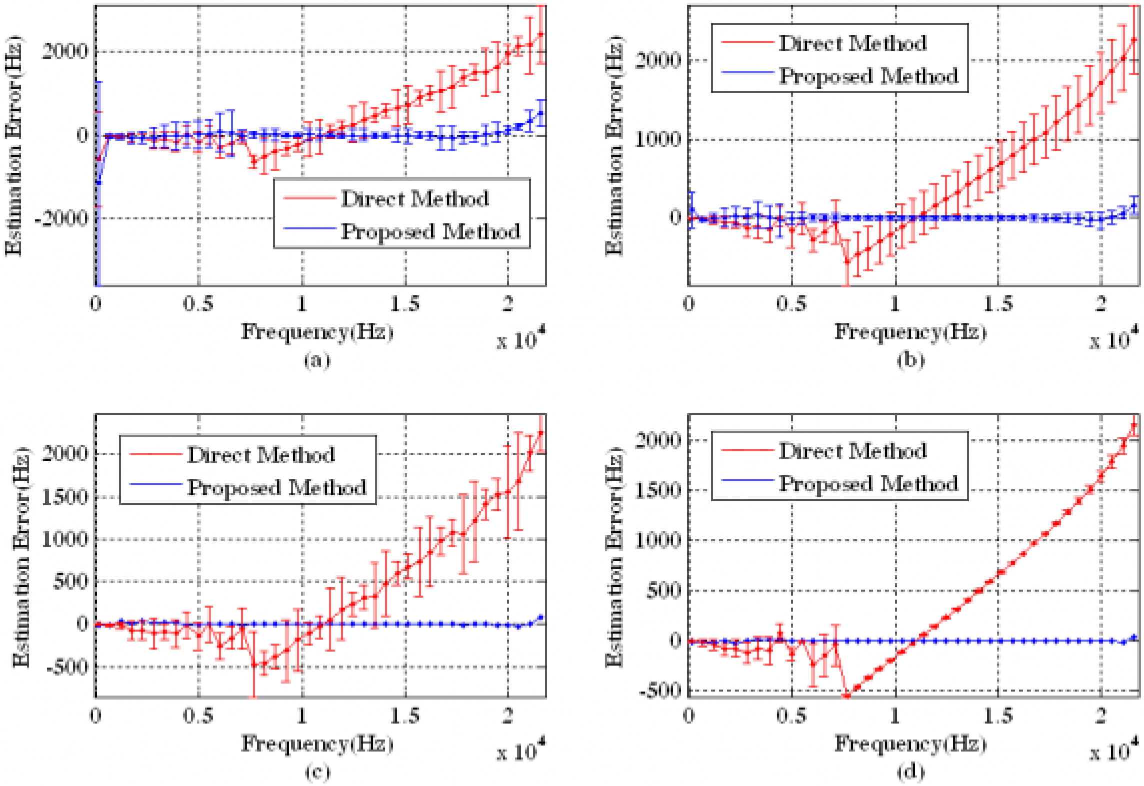


Figure 9:

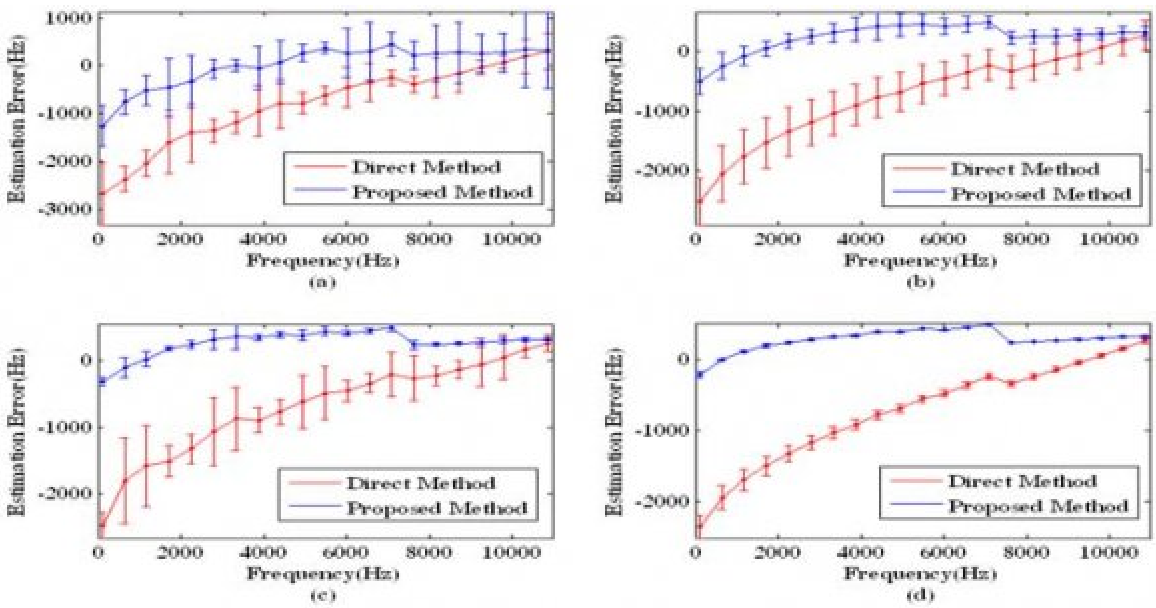


Figure 10:

Figure 11:

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

VI.

Figure 12:

1

Figure 13: Table 1 :

Tone	True Spectral	Spectral Centroid	SC Est. Error	Spectral Centroid	SC Est. Error	Year 2015
no	Centroid (Hz)	(Estimated by Direct Method)(Hz)	(Direct Method) (Hz)	(Estimated by Proposed Method)	(Proposed Method) (Hz)	35
1 2	(1) 96.8994	(2) 634.033 ± 103.2825	(1) -(2)	(Hz) (3) 96.8994	(1)	( )
3 4	635.2295	1107.9746 ± 114.8693	-537.13	± 0 635.2295	-	Vol-
5 6	1173.5596	1608.3049 ± 107.8578	-472.75	± 0 1173.5596	(3)	ume
7 8	1711.8896	2117.5429 ± 99.1608	-434.75	± 0 1711.8896	0 0	XV
9	2250.2197	2623.4037 ± 94.3929	-405.65	± 0 2250.2197	0 0	Is-
10	788.5498	3131.4122 ± 89.3034	-373.18	± 0 2788.5498	0 0	sue
11	3326.8799	3642.626 ± 86.5037	-342.86	± 0 3326.8799	0 0	IV
12	3865.21 4403.54	4158.9111 ± 81.3724	-315.75	± 0 3865.21 ±	0 0	Ver-
13	4941.8701	4669.3952 ± 79.7196	-293.7	0 4403.54 ± 0	0 0	sion
14	5480.2002	5183.2597 ± 76.1674	-265.86	4941.8701 ± 0	0 0	I
	6018.5303	5698.366 ± 75.0958	-241.39	5480.2002 ± 0		J
	6556.8604	6217.31 ± 71.2696	-218.17	6018.5303 ± 0		
	7095.1904	6730.1438 ± 70.2138	-198.78	6556.8604 ± 0		
		7247.1104 ± 67.5403	-173.28 -151.92	7095.1904 ± 0		
15	7633.5205	7764.439 ± 66.7206	-130.92	7633.5205 ± 0	0 0	of
16	8171.8506	8283.9504 ± 64.5743	-112.1	8171.8506 ± 0	0 0	Re-
17	8710.1807	8797.6702 ± 64.3643	-87.49	8710.1807 ± 0	0 0	searches
18	9248.5107	9316.0265 ± 63.4916	-67.52	9248.5107 ± 0	0 0	in
19	9786.8408	9834.3672 ± 63.363	-47.53	9786.8408 ± 0	0 0	En-
20	10325.1709	10353.7776 ± 62.7943	-28.61	10325.1709 ± 0	0	gi-
21	10863.501	10867.8735 ± 62.924	-4.37 14.7	10863.501 ± 0		neer-
22	11401.8311	11387.1281 ± 63.1554	34.47	11401.8311 ± 0		ing
23	11940.1611	11905.6919 ± 63.1063	54.03	11940.1611 ± 0		
24	12478.4912	12424.4625 ± 63.3744	78.44	12478.4912 ± 0		
25	13016.8213	12938.3795 ± 63.3481		13016.8213 ± 0		
26	13555.1514	13457.8904 ± 64.7948	97.26	13555.1514 ± 0	0 0	Global
27	14093.4814	13975.7026 ± 65.3201	117.78	14093.4814 ± 0	0 0	Jour-
28	14631.8115	14493.3087 ± 67.2566	138.5	14631.8115 ± 0	0 0	nal
29	15170.1416	15006.592 ± 67.9661	163.55	15170.1416 ± 0		
30	15708.4717	15525.9341 ± 71.2219	182.54	15708.4717 ± 0		
31	16246.8018	16042.2271 ± 72.4166	204.57	16246.8018 ± 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	

Figure 14: Table 2 :

.1 Global Journals Inc. (US) Guidelines Handbook 2015

www.GlobalJournals.org

[Scheier and Slaney ()] 'Construction and evaluation of a robust multifeature speech/music discriminator'. E Scheier , M Slaney . *Proc. IEEE ICASSP*, (IEEE ICASSP) 1997.

[Wold et al. ()] 'Content-based classification, search, and retrieval of audio'. E Wold , T Blum , D Keislar , J Wheaton . *IEEE Multimedia Mag* 1996. 3 p. .

[Schubert and Wolfe ()] 'Does Timbral Brightness Scale with Frequency and Spectral Centroid?'. Emery Schubert , Joe Wolfe . *Acta Acustica United With Acustica* 2006. 92 p. .

[Al (ed.) ()] *Introduction to MPEG-7*, Al . B. S. Manjunath (ed.) 2002. Wiley. (1st edition)

[Jia Min Karen Kua and Al (2010)] 'Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition'. Jia Min Karen Kua , Al . *The Speaker and Language Recognition Workshop*, (Brno, Czech Republic) 28 June -1 July 2010. p. .

[Tao and Al (2002)] 'Narrowband Acoustic Doppler Volume Backscattering Signal-Part II: Spectral Centroid Estimation'. Xiao-Jiao Tao , Al . *IEEE Transactions On Signal Processing* November 2002. 50 (11) p. .

[Chen and Al (2004)] 'Recognition of Noisy Speech Using Dynamic Spectral Subband Centroids'. Jingdong Chen , Al . *IEEE Signal Processing Letters* February 2004. 11 (2) p. .

[Gajic´ and Paliwal (2006)] 'Robust Speech Recognition in Noisy Environments Based on Subband Spectral Centroid Histograms'. Bojana Gajic´ , Kuldip K Paliwal . *IEEE Transactions On Audio, Speech, And Language Processing* March 2006. 14 (2) p. .

[Chandwadkar and Sutaone (2013)] 'Selecting Proper Features and Classifiers for Accurate Identification of Musical Instruments'. M Chandwadkar , M S Sutaone . *International Journal of Machine Learning and Computing* April 2013. 3 (2) p. .

[Peeters et al. ()] 'Toward automatic music audio summary generation from signal analysis'. G Peeters , A L Burthe , X Rodet . *Proceedings of the Third International Conference on Music Information Retrieval*, (the Third International Conference on Music Information RetrievalParis, France) 2002. p. .