ICSV14
Cairns • Australia
9-12 July, 2007

14th International Congress on Sound & Vibration

WAVELET DESIGN FOR MUSIC PERCEPTION APPLIED TO INDIAN CLASSICAL MUSIC

Charu[1 *], V.G.Das[2 †], V.Prem Pyara[3 ‡], V.Prem Kumari[4 §]

[1]Jaypee Institute of Engineering and Technology, Raghogarh, India

[2]Dayalbagh Educational Institute, Agra, India

[3]Department of Electrical Engineering, Dayalbagh Educational Institute, Agra, India

[4]Department of Music, Dayalbagh Educational Institute, Agra, India

[*]pathak_charu@rediffmail.com
[†]admin@dei.ac.in
[‡]prempyara@hotmail.com
[§]dbei@vsnl.com

Abstract

Wavelet Transforms provide a time-frequency analysis with constant Q factor. Although these transforms require intensive computational effort, they are efficient in terms of accuracy. Here, a wavelet is designed for the estimation of frequency of audio signals. This wavelet provides a resolution of 1% which is comparable to human auditory system and is applied here to Indian Classical Music. For this style of music, the task of frequency estimation is comparatively difficult as the note frequencies are not fixed but vary at fixed ratios with respect to the base frequency chosen by the performer. The obtained results match the theoretically expected frequencies.

1. INTRODUCTION

Wavelet Transforms are a well known tool for multiresolution analysis. This tool gives the information of frequencies over a logarithmic distribution. Since the perception of human ear has a similar property, it is very useful in audio signal analysis. The per unit resolution obtained by this transform is the same both at high and low frequencies. This transform is evaluated as a function of time at different scale factors. The scale may be considered to be the inverse of the frequency. This gives a three dimensional view of frequency distribution. Two dimensions are time and scale. The amplitude forms the third dimension. The wavelet transforms have logarithmic distribution in frequency. Since the Q factor remains constant in the analysis, it complements the music signal analysis. Human ear has a perception of about 1%. This property of human ear thus percieves according to the logarithmic distribution of frequencies. Thus, a wavelet with a resolution of 1% is proposed and is used to extract the note frequencies in a music signal.This tool is also applicable for musical instrument recognition. Several researchers have applied wavelets for speech and music signal analysis [5,6,3,4,2,1]. A constant Q analysis of music signals can also be used for musical instrument recognition.

2. ESTIMATION OF MINIMUM QUALITY FACTOR FOR PERCEPTION IN INDIAN MUSIC

Human ear can perceive upto 1% resolution among frequencies. Thus, it is necessary to investigate the minimum value of Q factor required to maintain the quality in Indian Music. The scale of Indian Music divides an octave into seven swaras and further divides them into srutis. The swaras differ from one another by 2, 3 or 4 srutis. The interval of one sruti is expressed as frequency ratio. These srutis are not equal. The intervals are $\frac{81}{80}$, $\frac{25}{26}$, $\frac{256}{243}$. The smallest of the sruti ratios so identified is $\frac{81}{80}$. This is similar, as resolving $1.25Hz$ in $100Hz$. Thus, the desired resolution of a machine recognition system should be better than $1.25Hz$ at $100Hz$. The desired Q factor of the wavelet should be equal to 100.

3. DESIGN OF THE MOTHER WAVELET

For the music signals, the period of one tone varies from a fraction of one second to a few seconds. The wavelet width in time cannot exceed the signal width. Thus, we may consider a sinusoidal function truncated by a rectangular window so that the total width of the window does not exceed the period of one note. Suppose, we take N cycles of a sinusoidal function $\cos(\omega_0 t)$ truncated by a rectangular window and study its magnitude spectrum. The multiplication of the two functions in time domain causes convolution of their Fourier Transforms in frequency domain. Mathematically, the above functions may be expressed as follows. The sinusoidal function $f_1(t)$ may be taken as

$$f_1(t) = \cos(\omega_0 t). \tag{1}$$

. Its Fourier Transform $F_1(\omega)$ is given as

$$F_1(\omega) = \pi\{\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\}. \tag{2}$$

The rectangular window with the width equal to N time periods of the sinusoidal function is given by

$$f_2(t) = \text{rect}(\frac{t}{NT}) \tag{3}$$

and its Fourier Transform is

$$F_2(\omega) = NT\ \text{Sa}(\frac{\omega NT}{2}). \tag{4}$$

Here, $\text{Sa}(x)$ represents the sampling function of x and is defined as $\frac{\sin(x)}{x}$. Let, $f_3(t) = f_1(t).f_2(t)$. The Fourier Transform of $f_3(t)$ is given by

$$F_3(\omega) = \frac{1}{2\pi}\{F_1(\omega) * F_2(\omega)\} \tag{5}$$

or,

$$F_3(\omega) = \frac{NT}{2}\left\{\text{Sa}\left(\frac{NT}{2}(\omega - \omega_0)\right) + \text{Sa}\left(\frac{NT}{2}(\omega + \omega_0)\right)\right\}. \tag{6}$$

Substituting the value of T as $\frac{2\pi}{\omega_0}$ we get

$$F_3(\omega) = \frac{N\pi}{\omega_0}\left\{\text{Sa}\left(\frac{N\pi}{\omega_0}(\omega - \omega_0)\right) + \text{Sa}\left(\frac{N\pi}{\omega_0}(\omega + \omega_0)\right)\right\}. \tag{7}$$

The plot of $F_3(\omega)$ for $N = 10$ and $\omega = 150$ is shown in fig. 1. It shows that the peak

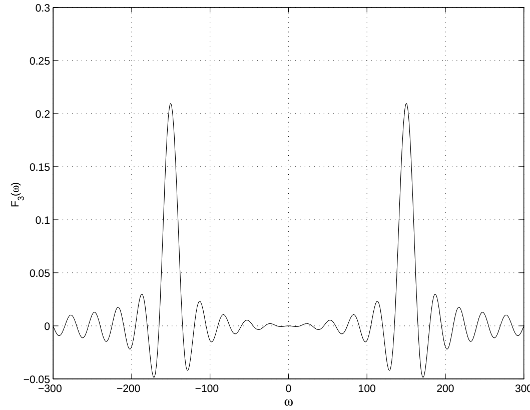

Figure 1. Spectrum of the sinusoidal wavelet

value occurs at $\omega = 150$. On increasing the value of N, the central lobe becomes narrow, thus increasing the quality factor. The central lobe has a width of $\frac{2\omega_0}{N}$ at its zero-crossing points. The 3dB width is approximately half of this value. That is $\Delta\omega = \frac{\omega_0}{N}$. Hence, the Q achieved by the signal is nearly equal to N. Here, N is the number of cycles of the sinusoidal function that are selected by the rectangular window. This gives us an idea of the number of cycles of the sine wave that should be taken into the wavelet so as to achieve a certain value of Q. Hence, in order to achieve a Q of 100, the wavelet must comprise of 100 cycles of sine wave. The total width of this wavelet will be $100T$, where T is the time period of one cycle. We wish to scan a range of frequencies from f_1 to f_2 in our analysis. Thus, the width of the wavelet should vary from $\frac{100}{f_1}$ to $\frac{100}{f_2}$. The maximum width of the Wavelet corresponds to the lowest frequency f_1, and is given by $\frac{100}{f_1}$. In case $f_1 = 100Hz$, the width comes out to be 1 second. This in turn requires that the singer must maintain the tone of the note steadily for at least 1 second. Though this type of wavelet can detect frequencies present in the signal with good accuracy, it also shows some spurious outputs simultaneously. This happens because of Gibb's phenomenon. Some other window function may be used in place of rectangular window, which would reduce this effect.

4. APPROPRIATE WINDOW

To select an appropriate window, the value of attenuation A of the first side lobe with respect to the central lobe, and corresponding Q are calculated for a few standard window functions. Let the width of the window function in each case be 100 cycles of the sinusoidal function. As the lowest frequency of interest is around 100Hz, the width of 100 cycles at this frequency equals to exactly 1 second. Some well known window functions are analyzed here. The duration of wavelet is chosen to be 1 second and the frequency of sinusoidal function f_0 equals to 100Hz. The attenuation ratio in dB is given by

$$A = 20\, log_{10}\left\{\frac{\mid F(\omega)_{max} \mid}{\mid F(\omega)_{sl} \mid}\right\} \tag{8}$$

where, $\mid F(\omega)_{max} \mid$ is the peak amplitude of main lobe and $\mid F(\omega)_{sl} \mid$ is the peak amplitude of the side lobe occurring next to main lobe. The quality factor Q is,

$$Q = \frac{f_0}{\Delta f} \tag{9}$$

Several window functions as Rectangular window, Hamming window, Gaussian window and Blackman window were analyzed for the $'Q'$ factor and relative attenuation of side lobe with respect to central lobe $'A'$. These values are listed in table 1. This analysis shows that, though the

Window	$'A'$ in dB	$'Q'$ at 3dB Frequencies
Rectangular	-13.4648	100
Hamming	-43.5729	77
Blackman	-59.4394	62.5

Table 1. Comparison factors as obtained through analysis of windows

Blackman Window function gives the maximum attenuation to the first undesirable side lobes, it is not the best window suitable for musical pitch analysis as it provides a low resolution. This low resolution is evident from the low value of Q or say large value of Δf.We may digress here to look at the side effect of the presence of side lobes in the frequency response of windowing functions.

4.1. Gibb's Phenomenon in Spectral Analysis

The presence of side lobes results in presence of unwanted frequencies at the central frequency. This undesirable effect may be minimized only by reducing the amplitude of the side lobes. Apart from this, the value of Δf obtained from the 3dB frequency values of the window also effect the accuracy. This is because a higher value of Δf implies a broader transition band which in turn reduces the resolution of the system. Thus, for a proper choice of the window, we need to look at both these aspects simultaneously. The degree of this simultaneous affect is estimated from the Fourier Transforms of windows. Fig. 2 shows this effect for Blackman

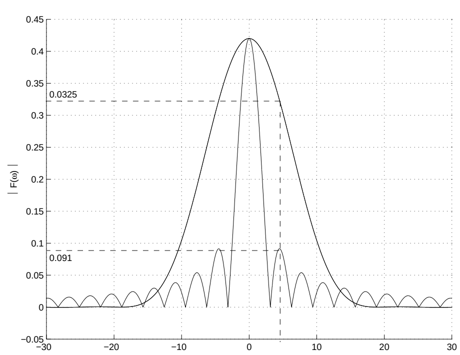

Figure 2. Spectrum of Blackman Window and a Rectangular Window

window and Rectangular Window. The amplitudes are normalized so that the central lobe peaks are equal in two cases. These plots show that the Blackman window increases the width of the response. That in turn implies a decrease in value of Q. The peak amplitude of the first side lobe in the rectangular case is 0.091. The response at this frequency in the case of Blackman window is approximately 0.325. This value is much higher than that for the rectangular window. This shows that Blackman Window reflects a higher magnitude of spurious components lying at this frequency compared to rectangular window.Similar results were obtained with Hamming window, Hanning window, Gaussian window etc. Here, the interest is that the side frequency attenuation be large. The rate of change of gain or attenuation does not matter. Hence, it may be concluded that a sine wave with a rectangular window gives better quality factor and consequently results in better resolution. In order to get a resolution of about 1%, the window should be of 100 cycles width.

5. MODELING OF THE PROPOSED MOTHER WAVELET

The mother wavelet developed in the section 4.1 may be written mathematically as,

$$\psi(t) = c\sin(2\pi t)\mathrm{rect}(\frac{t}{100}) \tag{10}$$

where, c is a constant. One of the conditions that a wavelet must satisfy is the condition of zero average. Mathematically, this is expressed as

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{11}$$

Since an integral number of cycles of a sine wave are taken in this wavelet, this condition is satisfied. The second condition is the condition of energy normalization. It is expressed as

$$\| \psi \| = 1 \tag{12}$$

The value of c in expression 10 to satisfy this requirement comes out to be $\frac{1}{\sqrt{50}}$. The function $\psi(t)$ is scaled and translated as

$$\psi_{s,\tau}(t) = \frac{c}{\sqrt{s}}\sin(\frac{2\pi(t-\tau)}{s})\mathrm{rect}(\frac{t-\tau}{100s}) \tag{13}$$

where $c = \frac{1}{\sqrt{50}}$. As we are interested in the amplitude of a spectral component of a certain frequency f embedded in $f(t)$, we shall use the frequency variable f in place of scaling factor s, by replacing s by $\frac{1}{f}$. Further, the value of c is also substituted and the expression for wavelet comes out as

$$\psi_{f,\tau}(t) = \sqrt{\frac{f}{50}}\sin(2\pi f(t-\tau))\mathrm{rect}(\frac{f(t-\tau)}{100}) \tag{14}$$

The wavelet transform is obtained by correlating the function with the signal as,

$$G(f,\tau) = \sqrt{\frac{f}{50}}\int_{\tau-\frac{50}{f}}^{\tau+\frac{50}{f}} f(t)\sin(2\pi f(t-\tau))dt \tag{15}$$

Let us consider the case when $f(t)$ is a signal of single frequency f , as

$$f(t) = A\sin(2\pi f_0 t - \phi) \tag{16}$$

Substituting this value of $f(t)$ in eq. 15, the convolution integral may be written as

$$G(f,\tau) = \sqrt{\frac{f}{50}}A\int_{\tau-\frac{50}{f}}^{\tau+\frac{50}{f}} \sin(2\pi f_0 t - \phi)\sin(2\pi f(t-\tau))dt \tag{17}$$

The value of $G(f,\tau)$ at $f = f_0$, may be evaluated from the above integral as,

$$G(f_0,\tau) = \left(\sqrt{\frac{50}{f_0}}A\right)\cos(2\pi f_0\tau - \phi) \tag{18}$$

The peak value of $G(f_0, \tau)$ as τ is varied over a range occurs at a value $\tau = \frac{\phi}{2\pi f_0}$ and its value is given by

$$G(f_0, \tau) \mid_{max} = \sqrt{\frac{50}{f_0}} A \tag{19}$$

Hence, to get the peak amplitude A of the spectral component of frequency f_0 from $G(f_0, \tau)$, we may compute the expression

$$A = \sqrt{\frac{f_0}{50}} G(f_0, \tau) \mid_{max} \tag{20}$$

The eq. 20 shows that the amplitudes reflected by the Wavelet transform, given by $G(f, \tau)$ do not represent the true amplitudes of the corresponding frequency components. To get the true amplitudes as a function of frequency one has to multiply the peak value of $G(f, \tau)$ by $\sqrt{\frac{f}{50}}$. When one is interested in amplitude spectrum only, we may combine eqs. 20 and 15 and evaluate the modified transform $g'(f, \tau)$ as

$$G'(f, \tau) = \frac{f}{50} \int_{\tau - \frac{50}{f}}^{\tau + \frac{50}{f}} f(t) \sin(2\pi\, f(t - \tau)) dt. \tag{21}$$

The corresponding Amplitude spectrum or the peak amplitude $A(f)$ shown as a function of f comes out as

$$A(f) = G'(f, \tau) \mid_{max} \tag{22}$$

6. APPLICATION TO INDIAN CLASSICAL MUSIC

A musical piece in Raga Yaman performed by Dr.S.B.Sharma, former Head, Dept. of Music, D.E.I. is selected. This raga has all the swaras of the octave in a sequence Sa, Re, Ga, Ma, Pa, Dha, Ni,$\dot{S}a$. $\dot{S}a$ denotes ‘Sa’ in next higher octave. A Tanpura is used as a drone to accompany the vocal sound. This sound sample is shown in fig. 3. This plot shows the amplitude variation

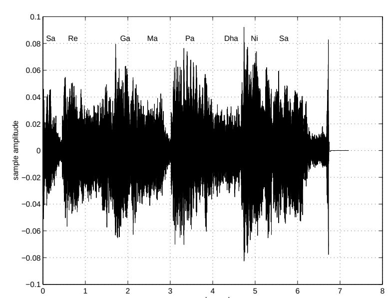

Figure 3. Test Signal

of the signal in time domain. The duration of the sample is about 15 seconds. The signal is sampled at the rate of 48000 samples per second using a standard Sound Card. The signal is divided into 20 bins each of 32768 samples. Each bin occupies a time of nearly $\frac{2}{3}^{rd}$ of a second and the number of samples in each bin is chosen to be a power of 2. A digital filter algorithm is used to cutoff unwanted higher frequencies. The desired range of the first octave being 240Hz to 480 Hz, the range of search of the pitch is limited to 100 Hz to 720 Hz. The Pass band limit of this low pass filter is kept at 720 Hz. The pass band starts at 1KHz. Since, now the signal contains only the frequencies upto 1KHz, the signal is downsampled by a factor of 20. Thus, the sampling rate of the signal has been compressed to 2400. This bandlimited signal is subdivided

into bins of 2048 samples each. The spectral analysis of each bin is carried out. The algorithms are coded in 'C' language and compiled on a UNIX operating system. The details of analysis of each method are given below.
In discrete time eqn 21 may be written as

$$G'(f,d) = \frac{f}{50N} \sum_{d-\frac{50N}{f}}^{d+\frac{50N}{f}} f(n\Delta\, t)\sin(2\pi\, f(n\Delta\, t - d\Delta\, t)). \tag{23}$$

The frequency being the inverse of scaling function, the frequency of the wavelet is varied from 100Hz to 1000Hz. The correlated output has a cyclic behavior. The peaks of this cyclic

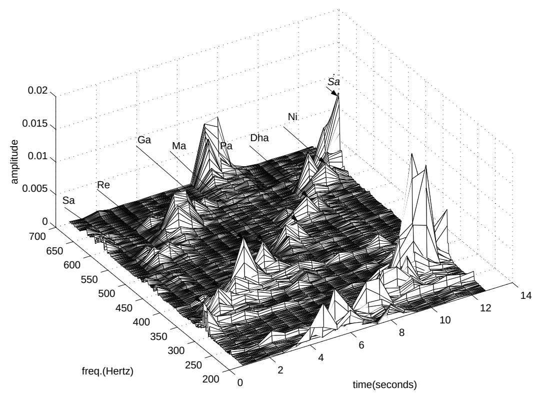

Figure 4. Identification of notes in time, frequency and amplitude using Wavelet Transforms

output are plotted in a three dimensional plot. The frequency of 'Sa'(base frequency) was found to be equal to 314Hz. The drone gave three frequency outputs that are present throughout. The amplitudes of these tones are low. These frequencies are 314Hz(Sa), 471Hz(Pa) and 628Hz($\dot{S}a$). The frequency range of 200Hz to 700Hz only is shown in fig. 4. The plot has x-axis as time, y-axis as frequency and z-axis represents the amplitude of the correlated function. The plot shows eight regions of octave clearly. The frequency chart of different notes according to their standard ratios as defined in the Indian Musical scale (Shadja Grama) with the base frequency of 'Sa' at $314Hz$ and the frequencies observed in the analysis by the three methods is given in table 2. The similar analysis was also carried out using Fast Fourier Transforms. These results can also be compared from the same table. It shows that the actual results obtained approximately match the ratios defined on a standard Indian musical scale and the sinusoidal wavelet gives more accurate results as compared to the Fourier Transforms.

7. CONCLUSIONS

The desired value of quality factor for the analysis of Indian Classical Music has been estimated and is found to be approximately equal to 100. A sinusoidal wavelet of 100 cycles truncated by a rectangular window is found to be most appropriate to achieve a $'Q'$ factor of 100. The notes in a vocal music signal were identified using the proposed technique. The frequencies have a good match with the theoretically estimated value.The same analysis was carried out using FFT and it was found that the Sinusoidal Wavelet Transforms gives more accurate results.

Note	Ratio w.r.t note 'Sa'	Expected frequency (Hz)	Observed frequency(Hz)	
			FFT	WT
Sa	1	314.00	314	314
Re	$\frac{9}{8}$	353.25	353	352
Ga	$\frac{10}{8}$	392.50	394	392
Ma	$\frac{4}{3}$	418.66	415	414
Pa	$\frac{3}{2}$	471.00	468	469
Dha	$\frac{27}{16}$	529.87	522	521
Ni	$\frac{15}{8}$	588.75	586	586
$\dot{Sa}$	2	628.00	625	627

Table 2. Comparison of theoretical estimates and observed frequencies

REFERENCES

[1] Michael and Unser, "On the Optimality of Ideal Filters for Pyramid and Wavelet Signal Approximation", *IEEE Transactions on Signal Processing* **25**,vol.51, 3591-3596(1993).

[2] J.Sadowsky, "Investigation of signal characteristics Using continuous Wavelet Transforms", *John Hopkins APL Technical Digest* **25**,vol.17, no.33 1996.

[3] J.C.Dovicchi, "Using New and Stable Wavelet coefficients in Instruments Sound Analysis ", *Electronic Musicological Review* **25**,vol.6, /March 2001.

[4] J.Fitch and W.Shabana, *A Wavelet based Pitch Detector for Music Signals*,jff@mathas.bac.ac.uk

[5] C.Wendt and A.Petropulu, "Pitch Determination and Speech Segmentation using Discrete Wavelet Transform", *IEEE Symposium on Circuits and systems*, vol.2,45-48, 1996.

[6] J.C.Dovicchi, G.A.Carrijo and L.Lima,"Music Based Co-efficients for Wavelet Transforms: A contibution to Music signal Analysis", *Proceedings of* 108^{th} *Convention of Audio Engineering Society, (Paris)*February 2000