COMPUTER AIDED DESIGN OF FRICTION BRAKES AND CLUTCHES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÜLLÜ KIZILTAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF MECHANICAL ENGINERING

SEPTEMBER 1998

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Tayfur ÖZTÜRK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ediz PAYKOÇ
Head of the Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Bilgin KAFTANOĞLU
Co-Supervisor

Assoc. Prof. Dr. Tuna BALKAN
Supervisor

Examining Committee in Charge:

Prof. Dr. Metin AKKÖK (Chairman)

Assoc. Prof. Dr. Tuna BALKAN (Supervisor)

Prof. Dr. Bilgin KAFTANOĞLU (Co-Supervisor)

Prof. Dr. M.İlhan GÖKLER

Y. Müh. Aydın GÜLTEKİN

ABSTRACT

COMPUTER AIDED DESIGN OF FRICTION BRAKES AND CLUTCHES

Kızıltaş, Güllü

M.S., Department of Mechanical Engineering

Supervisor : Assoc. Prof. Dr. Tuna Balkan

Co-Supervisor: Prof. Dr. Bilgin Kaftanoğlu

September 1998, 146 pages

The present study aims to develop a computer program on a PC to analyze and optimally design certain type brakes and clutches. In a way it is an application of CAD methods to an analysis and design problem. The program is composed of a Visual User Interface developed in C++ Builder, a main part (source code) written in C++ language and numerical and graphical outputs.

At the same time, the program is tried to be developed in a modular approach. The aim is easy future development and the implementation of this study in a general power transmission design program. Therefore the mechanical analysis is carried out with an equivalent inertia model and certain assumptions underlie the thermal analysis.

The effect of the complex nature of the coefficient of friction is tried to be taken into account where its variation with respect to temperature could be found in the literature and is implemented into the analysis using numerical techniques. Constant deceleration and constant torque is assumed in the case of no data.

The thermal response for single, repeated and continued braking and clutching applications with a general thermal model is investigated and their temperature variation is displayed graphically. Optimum design of friction type brakes and clutches is tried to be carried out based on maximum torque and minimum volume criteria and various performance and geometrical constraints.

Keywords: Friction Type Brakes and Clutches, Coefficient of friction, Thermal Analysis, Optimum design

ÖZ

SÜRTÜNMELİ FREN VE KAVRAMALARIN BİLGİSAYAR DESTEKLİ TASARIMI

Kızıltaş, Güllü
Yüksek Lisans, Makina Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Tuna Balkan
Yardımcı Tez Yöneticisi : Prof. Dr. Bilgin Kaftanoğlu

Eylül 1998, 146 sayfa

Bu çalışmada, PC ortamında belli tip fren ve kavramaların analizi ve optimum tasarımına yönelik bir programın geliştirilmesi amaçlanmıştır. Bu bir bakıma bilgisayar destekli tasarım metodlarının bir analiz ve tasarım problemine uygulanmasıdır. Bu program, C++ Builder yazılım programıyla geliştirilmiş görsel kullanıcı ortamından, C++ programlama dilinde yazılmış olan ana program kodundan, sayısal ve grafiksel çıktılardan oluşmaktadır.

Bu program aynı zamanda modüler bir yaklaşımda geliştirilmeye çalışılmıştır. Amaç, bu çalışmanın ilerde kolaylıkla geliştirilebilmesi ve genel bir güç aktarım elemanlarının tasarımı pogramında uygulanabilmesidir. Bu nedenle mekanik analiz, eşdeğer kütlesel atalet momenti modeli kullanılarak yapılmış, ısısal analiz de belli kabullenmelere dayandırılmıştır.

Sürtünme katsayısının karmaşık yapısı, bunların literatürde bulunan sıcaklığa bağlı değişimiyle dikkate alınmaya çalışılmış, analizde sayısal teknikler kullanılarak

uygulanmıştır. Bu değişimin bulunmadığı durumlarda analiz düzgün yavaşlama ve sabit tork kuvveti kabullenmelerine dayandırılmıştır.

Genel bir ısı modeliyle , bir tek, tekrarlı ve sürekli fren ve kavrama uygulamalarının ısısal durumları incelenmiş ve sıcaklık değişimleri grafiksel olarak elde edilmiştir. Sürtünme tipi fren ve kavramaların değişik performans ve geometrik sınırlamalara dayalı maksimum tork kapasitesi ve minimum hacim kriterlerine göre optimum tasarımı da yapılmaya çalışılmıştır.

Anahtar kelimeler : Sürtünme Tipi Fren ve Kavramalar, Sürtünme Katsayısı, Isısal Analiz, Optimum Tasarım

To my family

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Assoc. Prof. Dr. Tuna Balkan and Co-Supervisor Prof. Dr. Bilgin Kaftanoğlu, for their guidance, supervision, and support throughout the study.

I am also indebted to Prof. Dr. Metin Akkök for his valuable comments.

Very special thanks go to rescue squad, Nuran Katırcı, Nilgün Fesçioğlu, Özgür Ünver, Ali Emre Turgut and Bülent Özer.

I would like to thank my friends Elif Ahçı, İsmail Tarhan, Tuba Okutucu, Hakan Filiz, Ahmet Erdemir, Umut Ali Yener, Ahmet Karademir, Serhan Açıkgöz and Cafer Baş for lighting a candle when it is dark.

Also, I am grateful to Ms. Esma Aydın for all her aid.

Finally, words are inadequate sometimes but, greatest thanks go to my family who shaped me with never-ending patience, love, and sacrifice.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

ALTERNATIVE TRANSFORM TECHNIQUES FOR MUSICAL SIGNAL PROCESSING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÜRHAN GÜNCE GÜRSEL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

DECEMBER 1998

Approval of the Graduate School of Natural and Applied Sciences.

Prof.Dr. Tayfur ÖZTÜRK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof.Dr. Fatih CANATAN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof.Dr. Semih BİLGEN
Supervisor

Examining Committee Members

Assoc.Prof.Dr. Güney GÖNENÇ (Chairman)

Prof.Dr. Semih BİLGEN

Prof.Dr. Neşe YALABIK

Assoc.Prof.Dr. Melek YÜCEL

Assoc.Prof.Dr. Mete BULUT

ABSTRACT

ALTERNATIVE TRANSFORM TECHNIQUES FOR MUSICAL SIGNAL PROCESSING

GÜRSEL, Gürhan Günce

Ms., Department of Electrical and Electronics Engineering

Supervisor: Prof.Dr. Semih Bilgen

December 1998, 73 pages

In this thesis, various transform techniques for musical signal processing are analyzed and applied on different musical instrument samples. These transforms are the Constant Q and Discrete Wavelet Transforms. An efficient algorithm for computation of the Constant Q Transform is investigated. The transforms are compared according to their performances on the same sound files. The comparison is based on complexities, frequency and time resolutions of each transform. From this comparison, advantages and disadvantages of the transforms are ascertained.

Keywords: Constant Q Transform, Wavelet, Discrete Wavelet Transform

ÖZ

MÜZİKSEL SİNYAL İŞLEMESİ İÇİN ALTERNATİF DÖNÜŞÜM YÖNTEMLERİ

GÜRSEL, Gürhan Günce

Yüksek Lisans,Elektrik ve Elektronik Mühendisliği

Tez Yöneticisi: Prof.Dr. Semih Bilgen

Aralık 1998, 73 sayfa

Bu tezde, müziksel ses işlemesi için çeşitli dönüşüm yöntemleri analiz edilmekte ve değişik müzik enstrümanı örnekleri üzerine uygulanmaktadır. Bu dönüşümler, Sabit Q ve Kesikli Dalgacık dönüşümleri olarak adlandırılmaktadır. Sabit Q Dönüşümünün hesaplanması için verimli bir algoritma da ayrıca incelenmektedir. Tüm dönüşümlerin aynı ses dosyaları üzerindeki performansları karşılaştırılmıştır. Bu karşılaştırma, her dönüşümün karmaşıklığı ile, frekans ve zaman çözünürlüğüne dayalıdır. Bu karşılaştırmadan, dönüşümlerin avantaj ve dezavantajları ortaya konulmaktadır.

Anahtar Kelimeler: Sabit Q Dönüşümü, Dalgacık, Kesikli Dalgacık Dönüşümü.

To my mother and father

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to Prof. Dr. Semih Bilgen for his guidance and insight throughout the thesis study. Thanks go to the ASELSAN A.Ş. for the proper working conditions. I wish to thank to all of my colleagues for their comments and suggestions. And, my special thanks to my dear sister, Gizem, for her understanding and support, which made the completion of this work possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DFT : Discrete Fourier Transform

FFT : Fast Fourier Transform

DWT : Discrete Wavelet Transform

CHAPTER 1

INTRODUCTION

1.1 Musical Sound

All sounds originate from a body in a state of vibration. For a musical sound to be produced, vibrations of the source must be regular, that is, the frequency must remain constant for an appreciable interval of time thus producing some semblance of a note. With this property, a musical sound is a complex sound which has at least one perceivable pitch and a set of time dependent partial tones variable both in intensity and frequency. Pitch depends on the length of time in which each single vibration is executed, or which comes to the same thing, on the number of vibrations completed in a given time. If a second is taken as the unit of time, then the pitch number (or frequency) of a tone is the number of vibrations which the particles of a sounding body perform in one second of time. In musical terminology, the frequency of a tone is called the pitch. Note is the position of the pitch in the musical scale. Loudness of a sound is the magnitude of the auditory sensation produced by the sound. Timbre of a tone is a multi-dimensional characteristic that distinguishes two notes played by different instruments which have no pitch difference and furthermore identifies the instrument which it is played on. In fact, timbre is the tonal spectrum, or the acoustic spectrum. In the consideration of timbre of a tone, the number, distribution, relative intensity of the partials, the inharmonic partials, the fundamental tone, and the total intensity are the physical characteristics which determine the quality. Any frequency component of a sound is a partial. Harmonic partials are called overtones. The number of partials present determines the richness of the tone. The beauty of a tone depends upon the location of the partials in the harmonic series. The relative intensity of the partials is another important factor in determining the quality of a tone. The conventional spectrum of a tone depicts the fundamentals and partials as a harmonic series. The fundamental tone of different voices and instruments varies over a large section of the audio spectrum. The total intensity of a tone also influences the timbre. The higher the intensity level, the greater the number of partials which are generated by the instrument ([1], [2], [3], [4], [5], [6]).

1.2 Characteristics of Musical Instruments

Each instrument produces a characteristic frequency spectrum that distinguishes it from other instruments. The timbre of the instrument defines this frequency spectrum. However, if one note of an instrument is examined, the frequency spectrum changes over time. The beginning of the note often contains very strange behavior in the frequency spectrum. Overtones may arbitrarily fluctuate violently. Frequencies other than overtones may also be prevalent, and are likely to be changing rapidly as well. This region of the note is called the attack. The fluctuations in the overtones and other frequencies are called transients. These transients quickly die away to leave the tone of the instrument. The part of the sound without transients is called the steady state portion of the note, where the initial conditions are irrelevant. In instruments capable of sustaining a note, a third portion of the note, the release, the decay exists. That is, the complete musical sound consists of a recognizable starting transient and attack followed by a steady-state portion, then finally a decay until the sound dies up. The growth and decay characteristics of individual instruments in the different classifications vary from individual instrument to instrument. The steady state portion does not appear in every instrument. For example, instruments setting strings into vibration by plucking or hitting, can exhibit no steady state and the sound starts to decay immediately after attack ([4], [6], [7]).

The duration of a tone is the length of time that a tone emitted by a musical instrument lasts or persists without an interruption or discontinuity in the sound output. The duration characteristics influence to some degree the pitch, loudness and the timbre of a tone. From the standpoint of duration of a tone, musical instruments may be classified as follows: fixed duration, variable but fixed maximum duration and unlimited duration. The emitted tones in guitar, mandolin, bells, piano are of fixed duration. Violin, viola, flute, accordion emits the tones in a variable but fixed maximum duration whereas organ and electrical organ emits the tones in unlimited duration [4].

1.3 Musical Sound Processing

Computerized processing of musical sound has long been a subject of research. A fundamental step in musical sound processing is the recognition of the time and frequency characteristics of the played sound. This is in all cases, the first stage leading to further investigations such as instrument recognition, score generation, melody and expression analysis. Discrete Fourier Transform, Hartley Transform and Constant Q Transform are used to transform of the musical signal from time to frequency domain ([6], [8], [9], [10], [11]). Wavelet Transform is a transformation from time to time-frequency domain [17]. Within these transforms, Constant Q and Wavelet Transforms are particularly important for musical sound processing. The Constant Q Transform is important because it is created according to the needs of musical sound processing. Wavelet Transform is important because it is relatively a new method compared with the other transforms and therefore it is open for future enhancements.

1.4 Scope of the Study

This work aims to comparatively study the effectiveness and efficiency of the Constant Q, Discrete Fourier and the Wavelet Transform techniques for the purposes of musical sound processing.

Chapter 2 is concerned with the definition and calculation of Constant Q Transform [8]. This transform is a calculation similar to the Discrete Fourier Transform and is mainly defined for musical applications. The spacing between two adjacent frequencies is defined as the frequency resolution. Frequency resolution of the Discrete Fourier transform is constant. Whereas for Constant Q Transform, resolution is variable depending on the frequency on which information is desired. In order to obtain a variable resolution, the window size, i.e. the number of samples in the window, is also variable and decreases with increasing frequency. Q corresponds to the quality factor of the transform, and is the constant ratio of center frequency to resolution at that frequency so that two adjacent notes in the musical scale played simultaneously can be resolved anywhere in the musical frequency range. In Chapter 2, the Discrete Fourier Transforms of sample sound signals are also presented so that the results of two transforms are compared. Also in Chapter 2, an efficient algorithm [9] for calculation of the Constant Q Transform is implemented. This method uses the Parseval Equation so that the transform is calculated on the frequency domain instead of time domain. This makes it possible to take full advantage of the computational efficiency of the Fast Fourier Transform. Also with the efficient algorithm, it is seen that the number of multiplications is reduced because the calculated values on the frequency domain are dropped if absolute values are less than that of an adjustable parameter. This yields fewer multiplications in the calculation of each component of the Constant Q Transform but adds an error into the transform. For different values of the parameter, the number of multiplications and the error added into the transform are calculated and an optimum value for the parameter is estimated.

In Chapter 3, the Wavelet Transform is studied. The goal of the Wavelet Transform is to give an informative, efficient and useful description of a function or signal. If the signal is represented as a function of time, wavelets provide efficient localization in both time and frequency. Instead of cosine waves as in the Fourier Transform, the building blocks are wavelets. These are small waves that start and stop, that have a finite length. All are generated from a mother wavelet. In order to obtain time-frequency information of a signal, Wavelet Transform analyzes any signal by using multiresolution analysis (MRA). It analyzes the signal at different frequencies with different resolutions. The transform is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. In order to understand how

Wavelet Transform analyzes the signal, mathematical definitions are given with a brief introduction. Building blocks of the wavelet system, that are the scaling and wavelet functions are described. The definition of the Discrete Wavelet Transform is presented in the latter part of Chapter 3. In Chapter 3, the scaling and wavelet function coefficients are proved to be equal to the impulse responses of the filters used during the calculation of the Discrete Wavelet Transform. Properties and algorithms of the transform are also represented in Chapter 3.

Chapter 4 considers applications of the Discrete Wavelet Transform. As described in Chapter 3, the calculated coefficients of the Discrete Wavelet Transform, are the data of a 3-D graph. That is, they carry both amplitude and frequency information as well as time information. This is the main difference between the Wavelet Transform and the Fourier Transform. The Fourier Transform yields information on how much each frequency exists in a signal. It provides no information on the time plane. This difference is illustrated in the first part of this chapter and both transforms are compared. A principle, similar to the Heisenberg's uncertainty principle states that the frequency and time information of a signal at some certain point on the time-frequency plane can not be known simultaneously. This is the problem of resolution both on time and frequency domain. Although the Wavelet Transform provides extra information on time, it loses an amount of information on frequency. The calculated coefficients divide the time-frequency plane into equal portions. If these portions are defined as the tiles, these tiles represent location of the corresponding coefficient on the time-frequency plane. These tiles are represented graphically in Chapter 4. Different wavelet functions are used for the calculation of the Discrete Wavelet Transform ([16], [17], [22]). These functions are Haar, Sinc, Spline and Daubechies with different orders. The results are compared in order to discover a mid-range wavelet system. Various different sound samples are analyzed using the selected wavelet system and the calculated coefficients of the Discrete Wavelet Transforms are illustrated in Chapter 4.

In Chapter 5, the bases for comparison are defined as the number of multiplications performed during the calculation, the frequency resolution and the time resolution of the transforms. From the number of multiplications, the complexities of the transforms are obtained. Results of applications of the three main techniques, the Discrete Fourier, the Constant Q and the Discrete Wavelet Transform, on different sound samples are compared and advantages and disadvantages are described.

Chapter 6 concludes the thesis.

CHAPTER 2

CONSTANT Q TRANSFORM

2.1 Definition

In musical applications, there is a need for information about the spectral components produced across the wide frequency range of a particular musical instrument. Musical composition is based on musical scales. Musical frequencies are geometrically distributed. Western music divides the octave into 12 steps or equal intervals, semitone intervals, called the chromatic scale or equal temperament scale. These steps or equal intervals are called tempered half tones. A semitone, or half tone, in the scale of equal temperament is the frequency ratio between any two tones whose frequency ratio is the twelfth root of two ([1], [4]).

The spacing between two adjacent frequencies is defined as the frequency resolution. The frequency resolution of Discrete Fourier Transform (DFT) is (also called the frequency bin of DFT) [10]:

$$\Delta\theta = 2\pi / N \tag{2.1}$$

The corresponding physical frequency resolution (or frequency bin), in hertz, is,

$$\delta f = S / N \tag{2.2}$$

where S is the sampling rate and N is the window size in samples. Since the sampling rate and the window size is constant, the frequency resolution of DFT is constant and therefore the frequency components calculated with DFT are separated with a constant frequency difference and a constant resolution.

Constant Q Transform is defined for musical applications. The aim of the transform is to analyze the signal with variable resolutions depending on the frequency on which information is desired. Since the musical frequencies are geometrically distributed, the resolution should be geometrically related to the frequency. If the spacing is semitone spacing, the frequency of the k^{th} spectral component is defined as [8]:

$$f_k = (2^{1/12})^k f_{min} \tag{2.3}$$

Similarly, if quartertone spacing is used, the frequency is calculated from:

$$f_k = (2^{1/24})^k f_{min} \tag{2.4}$$

The upper frequency value will be below the Nyquist frequency (half of the sampling frequency) and f_{min} is defined as the minimum frequency about which information is desired. The resolution at the k^{th} spectral component is the difference between the frequencies of the $k+1^{th}$ and the k^{th} spectral component. If semitone spacing is used for frequency spacing, the resolution can be calculated as:

$$\delta f_k = f_{k+1} - f_k \tag{2.5}$$

$$\delta f_k = (2^{1/12})^{k+1} f_{min} - (2^{1/12})^k f_{min} \tag{2.6}$$

$$\delta f_k = (2^{1/12} - 1)(2^{1/12})^k f_{min} \tag{2.7}$$

$$\delta f_k = (2^{1/12} - 1) f_k \cong 0.06 f_k \tag{2.8}$$

Similarly, if quartertone spacing is used, the resolution will be:

$$\delta f_k = (2^{1/24} - 1) f_k \cong 0.03 f_k \tag{2.9}$$

It can easily be seen that, resolution is not constant but rather, varies geometrically and depends on frequency. But for each spectral component, there is a constant ratio of frequency to resolution. This constant ratio is defined as the quality factor, Q (constant Q) of the transform [8]. From Equations (2.8) and (2.9), the Q values are calculated as:

$$Q = f_k / 0.06 f_k \cong 17 \tag{2.10}$$

$$Q = f_k / 0.03 f_k \cong 34 \tag{2.11}$$

for semitone and quartertone spacing, respectively. Since frequency resolution is a ratio of sampling rate to the number of samples in the window, in order to obtain variable resolution, the window size of the k^{th} spectral component should be variable because sampling rate is constant. If the number of samples in the window of the k^{th} spectral component is defined as N[k], it can be calculated by rewriting Equation (2.2) and using Equation (2.10):

$$N[k] = S/\delta f_k \tag{2.12}$$

$$N[k] = (S/f_k)Q \tag{2.13}$$

From the above equation, it can be seen that window size N[k] and frequency f_k are inversely proportional. This means that, the number of samples in the window desired for the minimum frequency is the maximum and it decreases with increasing frequency and it has its minimum value for the Nyquist frequency of the transform.

The "Constant Q Transform" is defined as [8]:

$$X[k] = \frac{1}{N[k]} \sum_{n=0}^{N[k]-1} w[k,n] x[n] \exp\{-j2\pi Qn/N[k]\} \tag{2.14}$$

The above equation is an expression for the k^{th} spectral component and is similar to DFT. The digital frequency for DFT is $2\pi k/N$, whereas in Equation (2.14), the digital frequency is $2\pi Q/N[k]$ for the k^{th} component. The window function, w[k,n], has the same shape for each component but its length is determined by N[k], therefore it is a function of k as well as n. For the window function, either Hamming or Hanning window can be used. The window function, w[k,n] is defined typically as [11]:

$$w[k,n] = \alpha - (1-\alpha)\cos(2\pi n / N[k]) \qquad (2.15)$$

α=25/46 for Hamming window,

α=23/46 for Hanning window.

Also in Equation (2.14) normalization is performed by dividing the sum by N[k] since the number of terms varies with k.

2.2 Calculation of the Constant Q Transform

To illustrate the calculation of the Constant Q Transform, sound files were generated synthetically. To these sound files, both the DFT and the Constant Q Transform were applied. For the generation of these sound files, a sampling rate of 11025 Hz was chosen. All of the sound files had ten partials with different fundamental frequencies. The fundamental frequencies were also chosen in order to correspond to different musical note frequencies.

Figure (2.1) and Figure (2.2) present the Constant Q Transforms of the sound files having a fundamental at G_4 (G_4=392 Hz) and $G\#_4$ ($G\#_4$=415 Hz) respectively. These two fundamental frequencies are selected because there is a semitone spacing between the frequencies. The values on the vertical axis are the magnitudes of the complex coefficients and are normalized by dividing with the window size. In order to obtain a better view, the frequency axes of these figures are logarithmically scaled. As seen from the figures, there are ten peaks. These ten peaks correspond to the fundamental frequency and the nine harmonics of the fundamental.

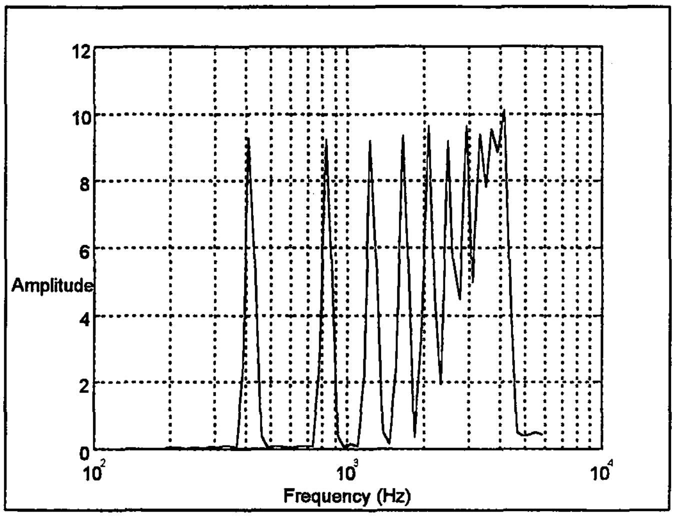

Figure (2.1) Constant Q Transform of sound with fundamental at 392 Hz having 9 harmonics

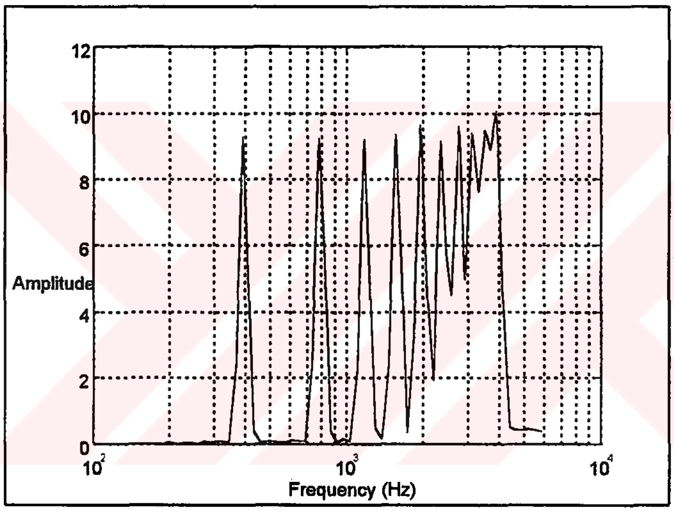

Figure (2.2) Constant Q Transform of sound with fundamental at 415 Hz having 9 harmonics

Figure (2.3) and Figure (2.4) are the Discrete Fourier Transform of the same sound files that are described in the preceding paragraph. These figures are obtained by taking 1024-point Fast Fourier Transform of the samples. For the defined sound samples, both of the transforms are applicable and all of the partials are resolved.

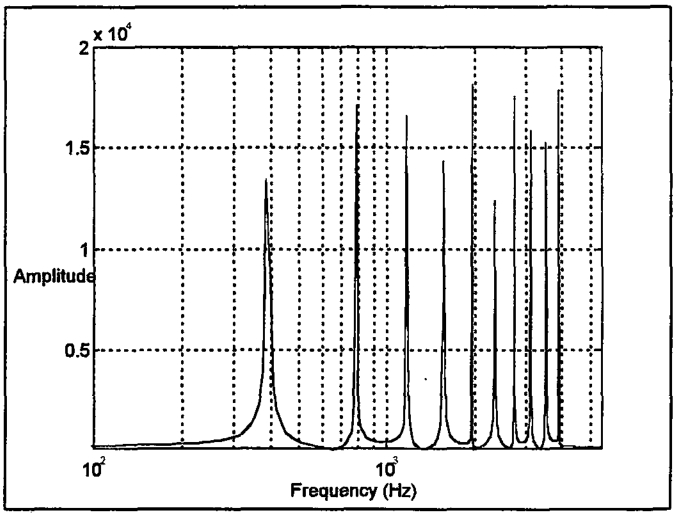

Figure (2.3) DFT of sound with fundamental at 392 Hz having 9 harmonics

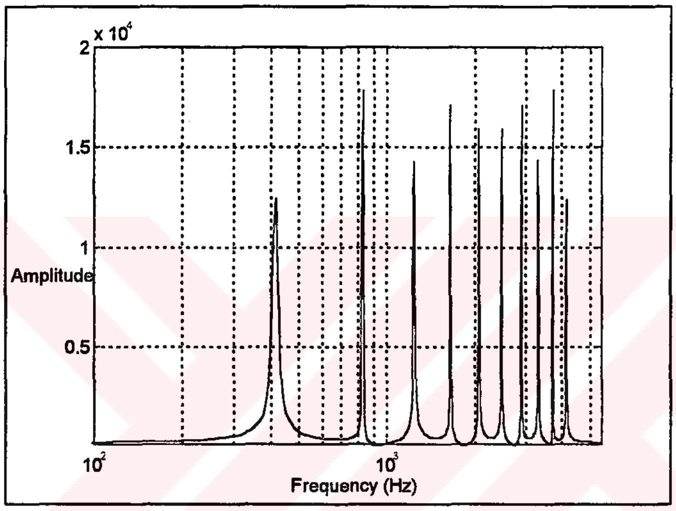

Figure (2.4) DFT of sound with fundamental at 415 Hz having 9 harmonics

Figure (2.5) illustrates the Constant Q Transform of the sound file which is generated from four harmonics of the fundamental frequencies at G_4 and $G\#_4$. Figure (2.6) is the Discrete Fourier transform of the same sound file. From Figure (2.5), it is seen that the first nine partials are resolved. The fourth harmonic of $G\#_4$ can not be resolved. In order to resolve this partial, the quality factor of the transform can be increased so that the number of samples in the window is increased. Also instead of semitone spacing, quartertone spacing can be used. For the DFT case, even though the partials are resolved, the two fundamentals are difficult to distinguish.

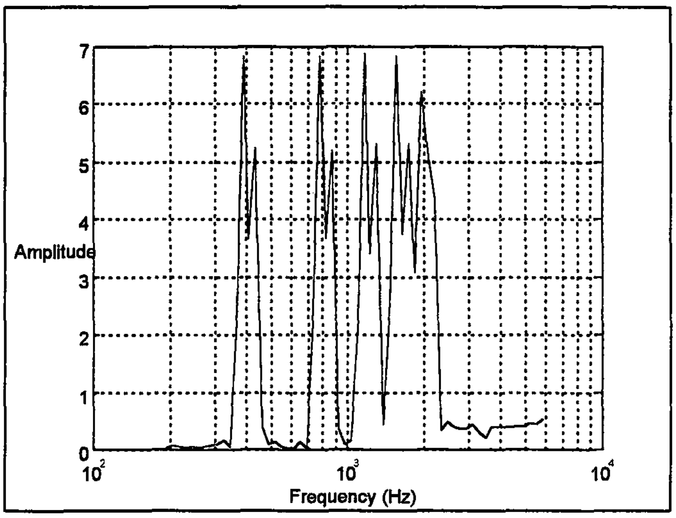

Figure (2.5) Constant Q Transform of sound with fundamentals at 392 Hz and 415 Hz having 4 harmonics of each fundamental

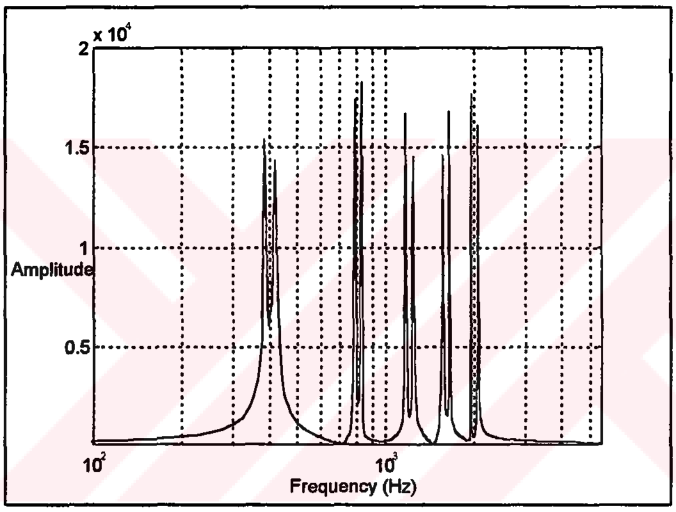

Figure (2.6) DFT of sound with fundamentals at 392 Hz and 415 Hz having 4 harmonics of each fundamental

A similar sound file is also generated from the harmonics of B2 (B2=123 Hz) and C3 (C3=139 Hz). For the calculation of Constant Q Transform of this sound file, the minimum frequency is decreased to 100 Hz. There is semitone spacing between the adjacent frequencies. Figure (2.7) presents this transform. All of the partials are resolved. Figure (2.8) is the DFT of this sound file. For this case, it is impossible to resolve the fundamental frequencies. Also there is a difficulty for the other harmonics. Therefore for this case, the DFT is not applicable.

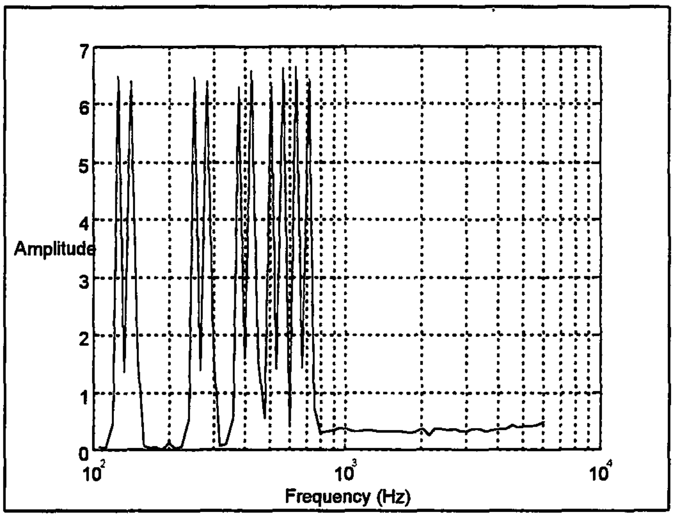

Figure (2.7) Constant Q Transform of sound with fundamentals at 123 Hz and 130 Hz having 4 harmonics of each fundamental

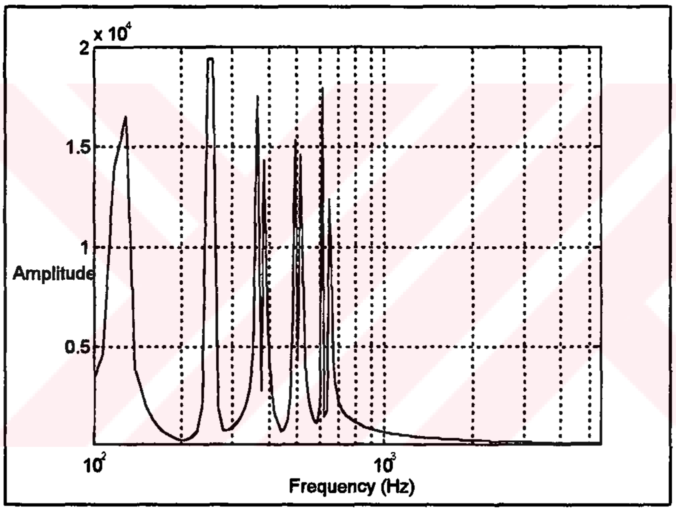

Figure (2.8) DFT of sound with fundamentals at 123 Hz and 130 Hz having 4 harmonics of each fundamental

2.3 Efficient Algorithm for the Constant Q Transform

Direct evaluation of Equation (2.11) is computationally inefficient. However, it can be shown that for any two discrete functions of time x[n] and y[n]:

$$\sum_{n=0}^{N-1} x[n]y^*[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]Y^*[k] \qquad (2.16)$$

where X[k] and Y[k] are the discrete Fourier transforms of x[n] and y[n], and Y*[k] is the complex conjugate of Y[k]. Equation (2.16) is a form of Parseval equation ([9], [12]).

A new form of Constant Q Transform is given as [9]:

$$X^{cq}[k_{cq}] = \frac{1}{N[k_{cq}]} \sum_{n=0}^{N[k_{cq}]-1} w[n,k_{cq}]x[n]e^{-jw_{k_{cq}}n} \qquad (2.17)$$

where $X^{cq}[k_{cq}]$ is the k_{cq} component of the constant Q transform. Here x[n] is a sampled function of time, and, for each value of k_{cq}, $w[n,k_{cq}]$ is a window function of length $N[k_{cq}]$. In Equation (2.17) [9]:

$$w_{k_{cq}} = \left(2^{(1/12)}\right)^{k_{cq}} w_{min} \qquad (2.18)$$

$$w_{min} = 2\pi f_{min}/S \qquad (2.19)$$

are defined to be the digital frequencies which are separated with semitone spacing. Also the digital frequencies can be calculated from:

$$w_{k_{cq}} = \frac{2\pi Q}{N[k]} \qquad (2.20)$$

If Equation (2.20) is substituted into Equation (2.18), it can be seen that both of the Constant Q Transform definitions are equivalent.

Parseval equation can be used to evaluate the Constant Q Transform as follows [9]. Letting

$$\frac{1}{N[k_{cq}]} w[n,k_{cq}] e^{-jw_{k_{cq}}n} = K^{*}_{temp}[n,k_{cq}] \qquad (2.21)$$

Equation(2.17) gives

$$X^{cq}[k_{cq}] = \sum_{n=0}^{N-1} x[n] K^{*}_{temp}[n,k_{cq}] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] K^{*}_{spec}[k,k_{cq}] \qquad (2.22)$$

where $K_{spec}[k,k_{cq}]$ is in the frequency domain and is called as the spectral kernels of the transformation. Also $K_{temp}[n,k_{cq}]$ is in the time domain and is called as the temporal kernels [9]. The spectral kernels are the Discrete Fourier Transform of the temporal kernels, that is

$$K_{spec}[k,k_{cq}] = \sum_{n=0}^{N-1} K_{temp}[n,k_{cq}] e^{-j2\pi kn/N} = \frac{1}{N[k_{cq}]} \sum_{n=0}^{N-1} w[n,k_{cq}] e^{jw_{k_{cq}}n} e^{-j2\pi kn/N} \qquad (2.23)$$

In Equation (2.22), the upper bound of the summation is given as N. Here N corresponds to a constant window size. Whereas the window size is defined as variable depending on the frequency of the spectral component in the calculation of Constant Q. Here the constant window size, (i.e. N) is set to a value that is nearest to the maximum window size of the transform (corresponding to the window size of the minimum desired frequency) and is a power of 2. This value should to be a power of 2 because FFT is used for the calculation of the frequency domain components. If the minimum frequency is 183 Hz, the sampling rate is 11025 Hz and the spacing is semitone then, the maximum window size is calculated as:

$$N[0] = \frac{S}{f_{min}} Q \Rightarrow N[0] = \frac{11025}{183} 17 \Rightarrow N[0] = 1024.18$$

Therefore, N is set to 1024 throughout the calculation of the transform for the above conditions. It is known that the window sizes (N[k] for the k^{th} spectral component) corresponding to the other frequencies are less than this constant value. The temporal kernels are defined to be symmetric around the center of the constant window. That is temporal kernels are zero outside the interval (N/2-N[k]/2,N/2+N[k]/2). In the given interval, they are calculated using Equation (2.23). The temporal kernels are constant unless the sampling rate, minimum frequency and frequency spacing are changing. Since the spectral kernels constitute the Discrete Fourier Transform of the temporal kernels, they are also constant under the same condition. Also the temporal kernels are conjugate symmetric, that is:

$$K_{temp}\left[n,k_{cq}\right]=\frac{1}{N\left[k_{cq}\right]}w\left[n,k_{cq}\right]e^{jw_{k_{cq}}n} \quad (2.24)$$

$$K^{*}_{temp}\left[-n,k_{cq}\right]=\frac{1}{N\left[k_{cq}\right]}w^{*}\left[-n,k_{cq}\right](e^{jw_{k_{cq}}(-n)})^{*} \quad (2.25)$$

$$w^{*}\left[-n,k_{cq}\right]=(\alpha-(1-\alpha)\cos(\frac{-2\pi n}{N\left[k_{cq}\right]}))^{*}=w\left[n,k_{cq}\right] \quad (2.26)$$

$$K^{*}_{temp}\left[-n,k_{cq}\right]=\frac{1}{N\left[k_{cq}\right]}w\left[n,k_{cq}\right]e^{jw_{k_{cq}}n}=K_{temp}\left[n,k_{cq}\right] \quad (2.27)$$

Since the temporal kernels are conjugate symmetric, the condition for a real Discrete Fourier Transform holds and the spectral components are real.

In order to use this efficient algorithm, the temporal kernels have to be calculated. Then by taking N-point FFT, where N is the maximum window size, of these kernels, the spectral kernels are evaluated. The spectral kernels are calculated once and can be used for the same conditions that is for the same sampling rate, the same minimum frequency and the same frequency spacing. The following five figures, (Figure (2.9)-(2-13)) present the spectral kernels for different k_{cq}^{th} spectral component. The spacing is semitone, the minimum is 183 Hz and the sampling rate 11025 Hz. The horizontal axis of these figures are FFT Bin Number. Since 1024 point FFT is taken, the first 512 of these are important but in order to display the spectral kernels in a meaningful way, different band of values are taken. The values of the spectral kernels are zero for the rest of the spectrum. It can be easily seen that, with increasing k_{cq} value, the band of the spectral kernel increases. Also the position of the band on the FFT Bin Number axis, shifts to right, to the high numbers. Figure (2.9), is the spectral kernel calculated for the maximum bin number $k_{cq\text{-}max}$, corresponding to the Nyquist frequency. For the calculation of this maximum bin number, Equation (2.18) can be rewritten as:

$$(2^{1/12})^{k_{cq}} = \frac{w_{k_{cq}}}{w_{min}}$$

$$k_{cq} = 12\frac{\log(w_{k_{cq}} / w_{min})}{\log(2)} \qquad (2.29)$$

The minimum digital frequency w_{min} is calculated from Equation (2.19) as:

$$w_{min} = \frac{2\pi(183)}{11025} = 0.105$$

The digital frequency of the Nyquist frequency is π, from Equation (2.29), $k_{cq\text{-}max}$ is calculated as:

$$k_{cq-max} = 12\frac{\log(\pi/0.105)}{\log(2)} \cong 59$$

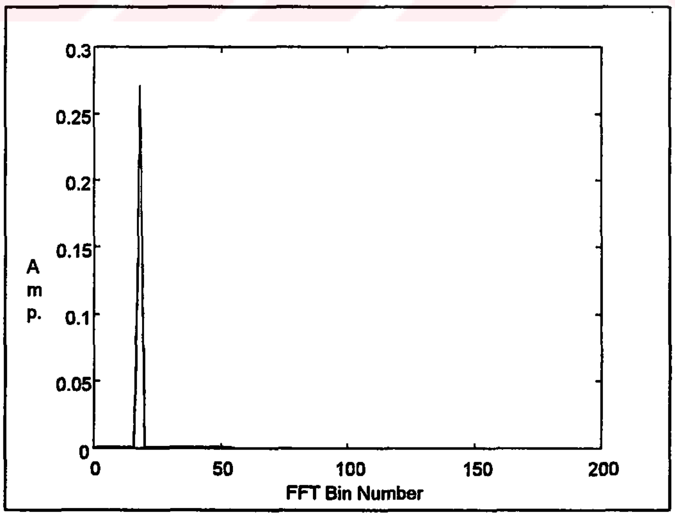

Figure (2.9) The Spectral Kernels for k_{cq}=1

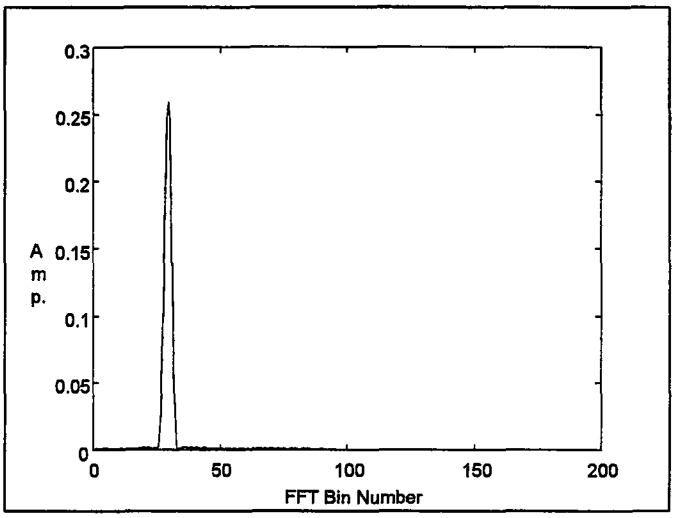

Figure (2.10) The Spectral Kernels for k_{cq}=10

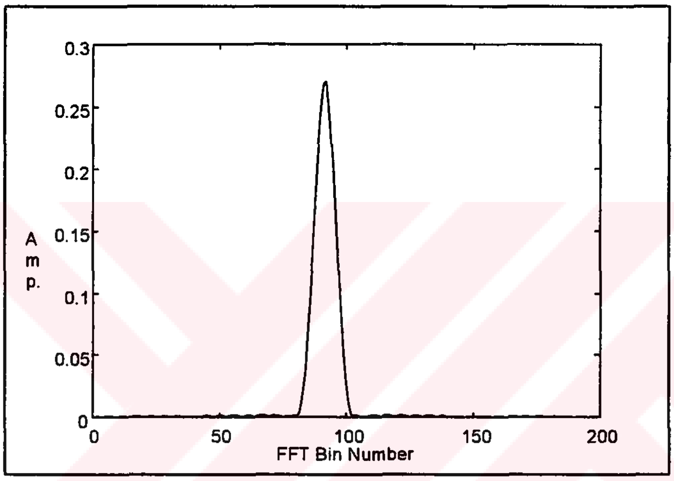

Figure (2.11) The Spectral Kernels for k_{cq}=30

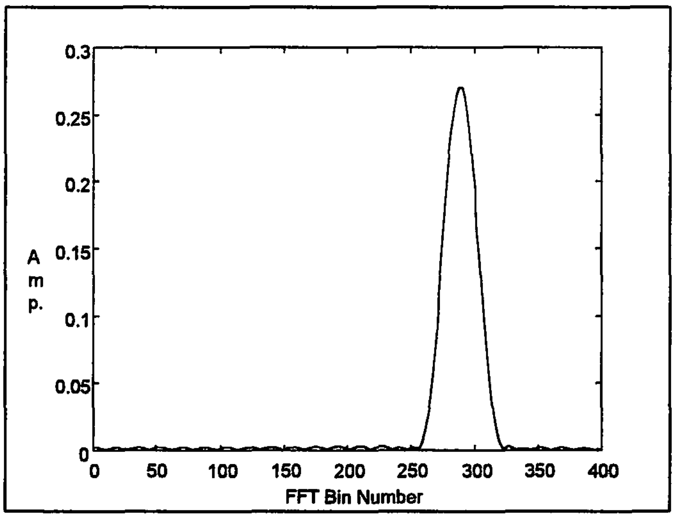

Figure (2.12) The Spectral Kernels for k_{cq}=50

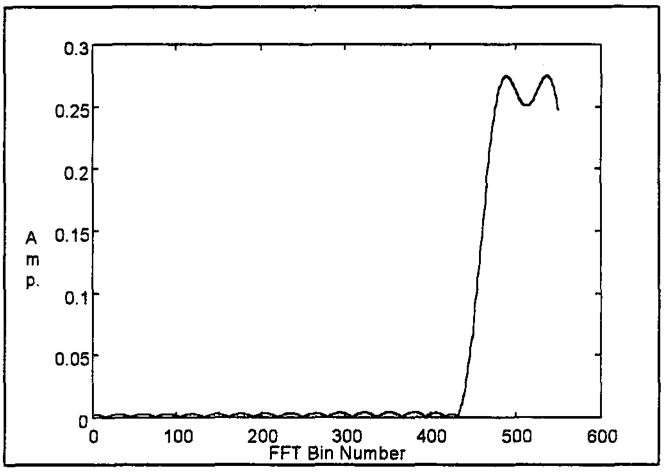

Figure (2.13) The Spectral Kernels for k_{cq}=59

In order to use the right side of Equation (2.22) for the calculation of Constant Q Transform, the spectral kernels are evaluated once or called from a previous calculation having the same condition. Then the FFT of the input sample have to be performed. By taking the summation, defined in Equation (2.22), $X^{cq}[k_{cq}]$ values are found.

Figure (2.14) and (2.15), present Constant Q Transforms calculated using Equation (2.22) on the sound file having 392 Hz fundamental frequency and nine harmonics of the fundamental. For this calculation, an adjustable parameter called MINVAL is used in order to select the spectral kernels, having absolute values greater than MINVAL. That is, if the absolute value of the spectral kernel is less than the value of MINVAL, the value of the kernel is set to zero. In other words, the value of the parameter MINVAL is a cutoff value for the spectral kernels. Table (2.1) shows the number of desired multiplications for the transform for different values of MINVAL. As seen from the table, the number of multiplications decreases suddenly from a small increase on the value of MINVAL. There is an error in dropping small values of kernels. The second column of the Table (2.1) is this estimated error. The error is calculated by summing the absolute values of the numbers which are dropped and dividing by the sum of the absolute values of the kernels.

Table (2.1) Number of multiplications performed and error estimated for different MINVAL values

MINVAL	ERROR	MULTIPLICATIONS
0	0	30720
0.03	0.0844	1521
0.05	0.0997	1369
0.07	0.1438	1241
0.1	0.1541	1106
0.12	0.2289	1007
0.15	0.3413	867
0.18	0.4332	742
0.2	0.4636	648
0.25	0.6020	360

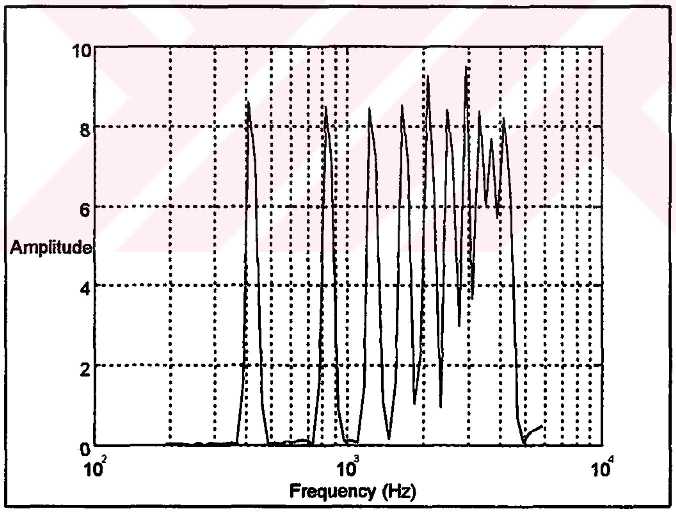

Figure (2.14) Constant Q Transform using efficient algorithm of sound with fundamental at 392 Hz having 9 harmonics (MINVAL=0)

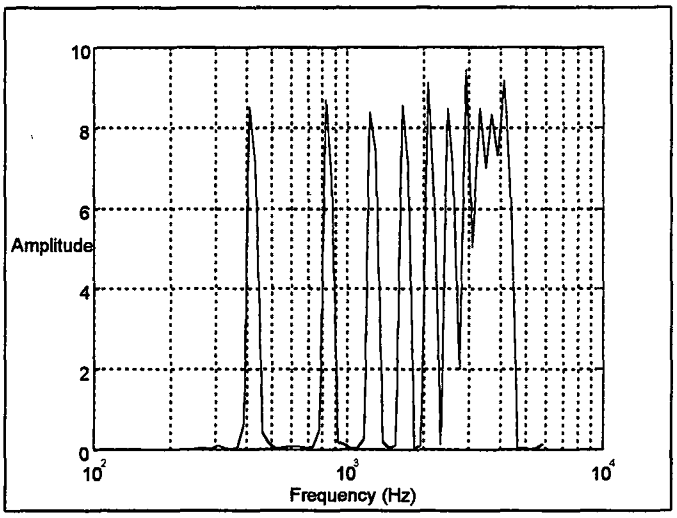

Figure (2.15) Constant Q Transform using efficient algorithm of sound with fundamental at 392 Hz having 9 harmonics (MINVAL=0.10)

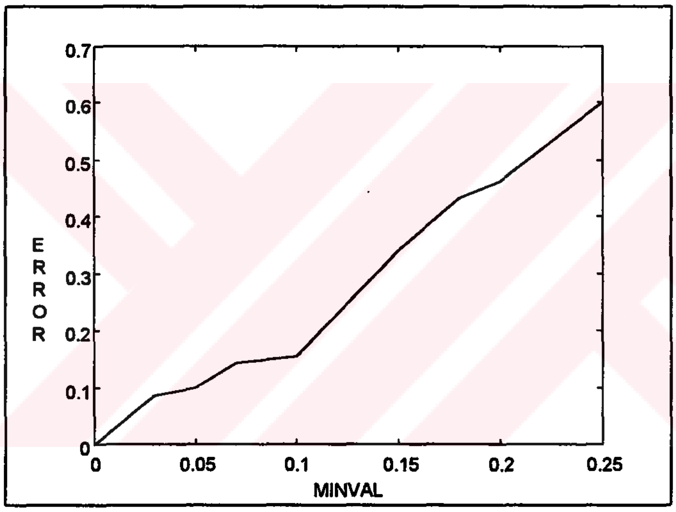

Figure (2.16) Error in dropping small values of kernels versus cutoff (MINVAL) value

Figure (2.16) is obtained from the first two columns of Table (2.1). From Table (2.1), by selecting the value of the parameter MINVAL as 0.1, there is an approximately 15% error for the calculation of the transform. With this error rate, there is still a constant pattern as seen in Figure (2.15). Also the number of complex multiplications diminishes from 30720 to 1106. Since the run-time of the algorithm depends heavily on the number of performed multiplications, this error ratio can be tolerated with this gain on the run-time of the algorithm.

2.4 Constant Q Transforms of Musical Instrument Samples

The following figures are the calculation of Constant Q Transform of the musical instrument sample files. The musical instruments are flute, saxophone, accordion, piano, guitar and org. The musical instruments are selected in order to represent the different types of instruments. All of the musical instrument sample files [13] are in the Waveform Audio File Format (.WAV), one of the standard multi-media file format of Windows. The Waveform Audio Files are made up of "chunks" ([14], [15]). In these chunks, different kinds of information are stored. One of the chunks, that is the format header chunk, is dedicated for the information of how the audio data is formatted in the file. Another chunk is for the audio data that is the raw data ([14], [15]). From the format header of the instrument sample files, it is known that all of the files are sampled at 44100 Hz. The number of channels of audio present in the file is one, that is the sounds are monaural. The raw data is sixteen-bit in signed integer audio data. For the calculations of the Constant Q Transform of the all of the sample files, the efficient algorithm is used. The value of the adjustable parameter MINVAL is set to 0.1. The audio data sets are downsampled to 22050 Hz by eliminating the even numbered samples. There is semitone spacing between the adjacent frequencies and the minimum frequency is selected as 183 Hz. From the Equation (2.13), the maximum window size can be calculated as:

$$N[0] = (S / f_{min})Q = (22050 / 183) * 17 \cong 2048$$

Therefore 2048 point FFT is used for the calculation of the spectral kernels from the temporal kernels. Since all of the musical instrument samples are obtained from 22050 Hz sampling rate and the same conditions exist, the spectral kernels are calculated once and are used for all transforms.

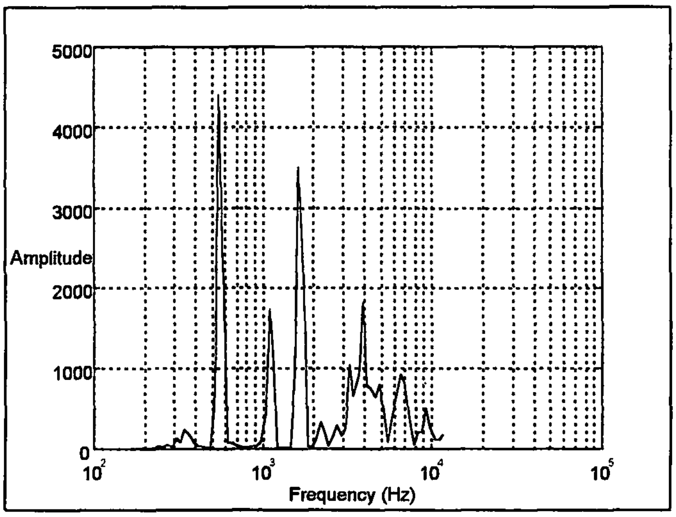

Figure (2.17) Constant Q Transform using efficient algorithm of guitar playing C5 (C5=523 Hz)

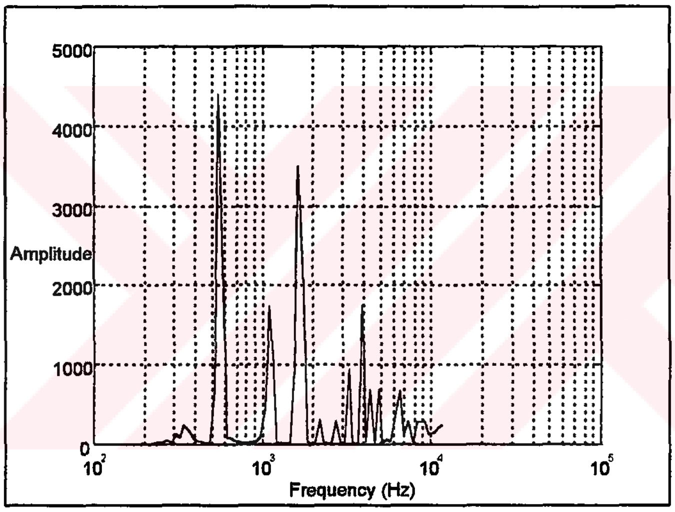

Figure (2.18) Constant Q Transform using efficient algorithm of guitar playing C5 (C5=523 Hz), the quality factor is doubled at frequency 1844.5 Hz

The above two figures represent the Constant Q Transform of the guitar playing C5. If these two figures are compared, it is seen that from Figure (2.17) that the resolution is insufficient for the high frequencies in the guitar spectrum. Therefore, Q is doubled near 1850 Hz. That is Q=34 for the frequencies greater than 1850 Hz. This frequency is selected because from Figure (2.17), it is clear that the harmonics near 2000 Hz are not resolved. With doubling the quality factor, the resolution is doubled. Therefore the harmonics near 2000 Hz are resolved, as seen in Figure (2.18).

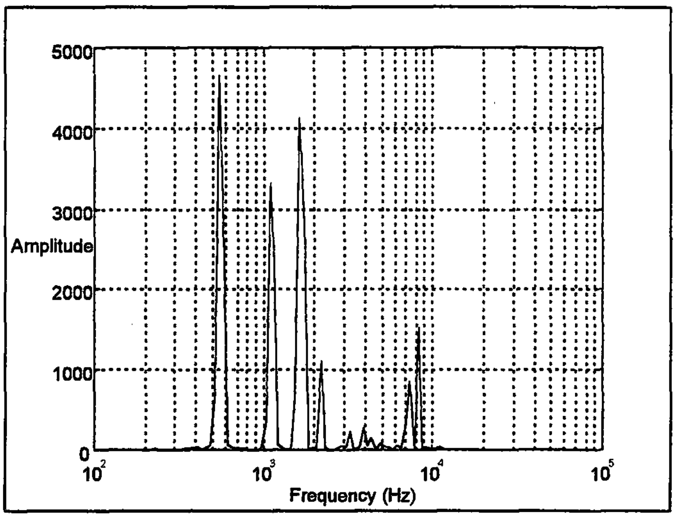

Figure (2.19) Constant Q Transform using efficient algorithm of piano playing C5 (C5=523 Hz), the quality factor is doubled at frequency 1844.5 Hz (MINVAL=0.1)

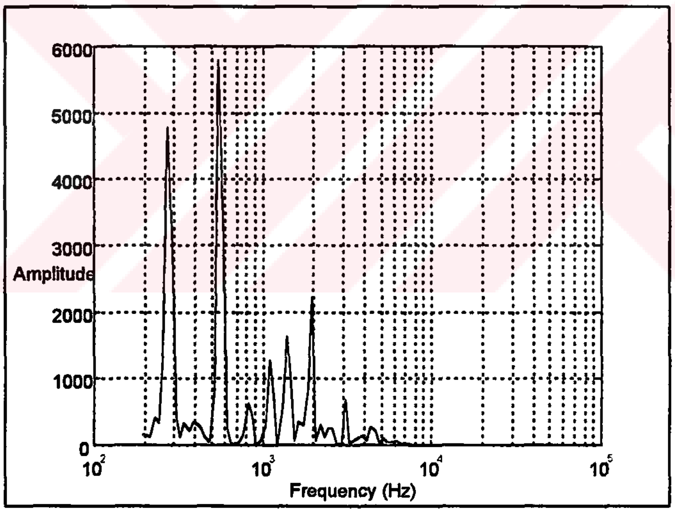

Figure (2.20) Constant Q Transform using efficient algorithm of piano playing C4 (C4=261 Hz), the quality factor is doubled at frequency 1844.5 Hz (MINVAL=0.1)

The above two figures illustrate the Constant Q Transform of piano playing two different tones C5 (C5=523 Hz) and C4 (C4=261 Hz) respectively. Nearly all of the overtones exist in the figures. This is the main characteristics of the string instruments. In the high frequency range, the amplitude of the harmonics decreases very rapidly.

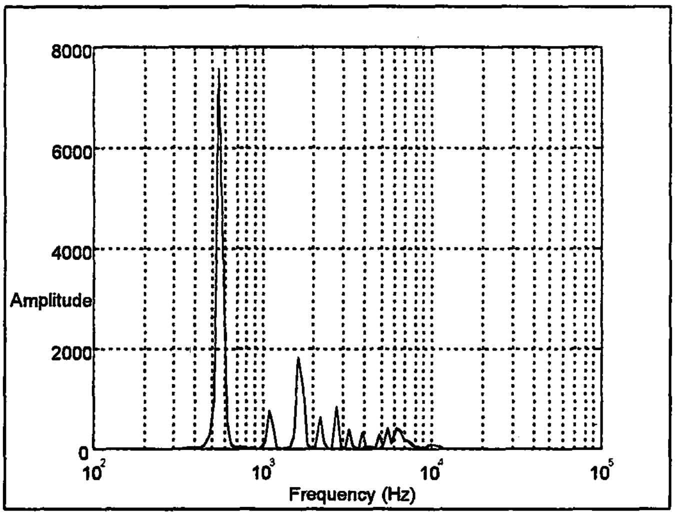

Figure (2.21) Constant Q Transform using efficient algorithm of accordion playing C5 (C5=523 Hz), the quality factor is doubled at frequency 1844.5 Hz

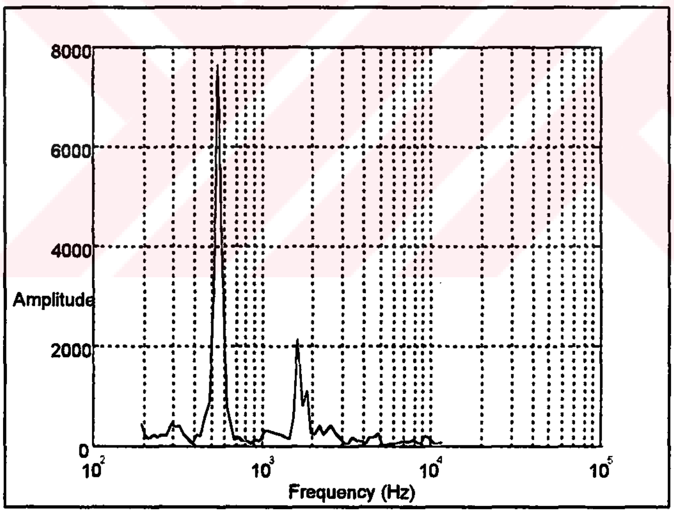

Figure (2.22) Constant Q Transform using efficient algorithm of flute playing C5 (C5=523 Hz), the quality factor is doubled at frequency 1844.5 Hz

The above two figures present the Constant Q Transform of accordion and flute, respectively. Both instruments are playing C5 (C5=523 Hz). For the accordion case, the spectrum is rich, that is the number of overtones is large. For the flute case, the large amount of sound power is in the fundamental with a small number of overtones. This is due to the fact that the flute has one of the purest tones of all musical instruments. Also the quality factor of the transform is doubled near 1844.5 Hz and the value of adjustable parameter, MINVAL is set as 0.1.

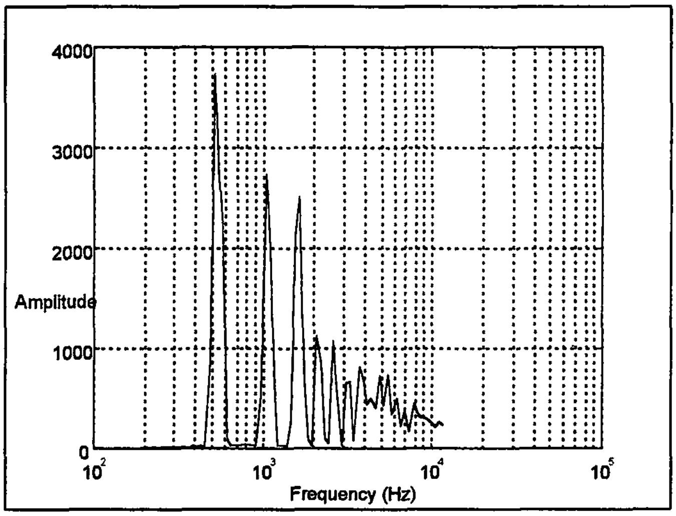

Figure (2.23) Constant Q Transform using efficient algorithm of saxophone playing C5 (C5=523 Hz), the quality factor is doubled at frequency 1844.5 Hz

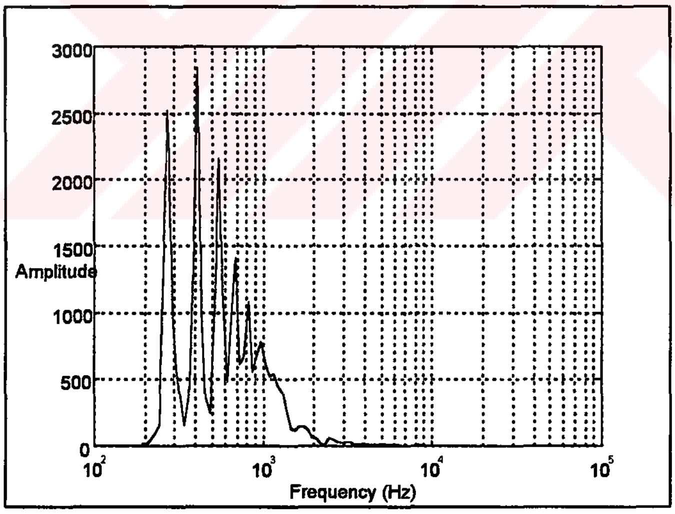

Figure (2.24) Constant Q Transform using efficient algorithm of saxophone playing C4 (C4=261 Hz), the quality factor is doubled at frequency 1844.5 Hz

The above two figures illustrate the Constant Q Transform of saxophone playing two different tones C5 and C4. If these two figures are compared, the amplitudes of the overtones are considerable. Even though it is seen that there is no fixed pattern in the overtone structure, the number of the overtones are nearly equal.

CHAPTER 3

WAVELET TRANSFORM

3.1 Introduction

A wave is usually defined as an oscillating function of time or space, such as a sinusoid. Fourier analysis is wave analysis. It expands signals or functions in terms of sinusoids (or, equivalently, complex exponentials) which has proven to be extremely valuable in mathematics, science, and engineering, especially for periodic, time-invariant, or stationary phenomena. The term wavelet indicates a "small wave" [16]. It has its energy concentrated in time to provide a tool for the analysis of transient, nonstationary, or time-varying phenomena. It still has the oscillating wavelike characteristic but also has the ability to allow simultaneous time and frequency analysis with a flexible mathematical foundation ([17], [18]).

The wavelet transform is a recent method of signal analysis and synthesis which represents signals in terms of wavelets. The wavelet transform decomposes an arbitrary function into a two parameter family of elementary wavelets that are obtained by shifts in time variable and by dilations that act both on the time and the frequency variables [18].

For the wavelet expansion, a two-parameter system is constructed such that:

$$f(t) = \sum_k \sum_j a_{j,k} \Psi_{j,k}(t) \tag{3.1}$$

where both j and k are integer indices and the $\Psi_{j,k}(t)$ are the wavelet expansion function (generating wavelet or mother wavelet) that usually form an orthogonal basis. The wavelet systems are generated from this wavelet by simple scaling and translation [17]. This can be formulated as:

$$\Psi_{j,k}(t) = 2^{j/2}\Psi(2^{j}t - k) \tag{3.2}$$

In this notation, k can be thought as the parameterization of time or space location and j as the frequency or scale. All wavelet systems satisfy the multiresolution conditions. This means that if a set of signals can be represented by a weighted sum of $\psi(t-k)$, then a larger set (including the original) can be represented by a weighted sum of $\psi(2t-k)$. In other words, if the basic expansion signals are made half as wide and translated in steps half as wide, they will represent a larger class of signals exactly or give a better approximation of any signal.

3.2 Mathematical Definitions

A function space "S", is defined as a linear vector space (finite or infinite dimensional) where the vectors are functions. On this space, the inner product of two functions is a scalar, obtained from two vectors, f(t) and g(t) by an integral. It is denoted as [17]:

$$a = \langle f(t), g(t) \rangle = \int f^{*}(t) g(t) dt \tag{3.3}$$

in which $f^{*}(t)$ denotes the complex conjugate of the function f(t). This inner product defines a norm or length of a vector which is denoted as [17]:

$$\|f\| = \sqrt{|\langle f, f \rangle|} \tag{3.4}$$

Two signals with non-zero norms are called orthogonal if their inner product is zero.

On the function space "S", any function, f(t), can be expressed as:

$$f(t) = \sum_{k} a_k \varphi_k(t) \qquad (3.5)$$

$\varphi_k(t)$ is called an expansion set for the space "S". If the representation, i.e. the set of $\{a_k\}$ is unique, the expansion set forms a basis. Alternatively, for a given expansion set or basis set, the space "S" consists of the set of all functions that can be expressed by $\varphi_k(t)$. This is called the span of the basis set [17].

A set of scaling functions, $\varphi_k(t)$ can be defined in terms of integer translates of the basis scaling function, $\varphi(t)$. This can be formulated as [17]:

$$\varphi_k(t) = \varphi(t-k) \qquad (3.6)$$

The function space can be defined as ϑ_0 and using the scaling function, a given function can be represented using Equation (3.5). In order to formulate the basic requirements of multiresolution analysis (MRA), a two dimensional family of functions, $\varphi_{j,k}(t)$ is generated from the basic scaling function $\varphi(t)$ by scaling and translation by:

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) \qquad (3.7)$$

whose span over k is ϑ_j. For $j>0$, the span can be larger since $\varphi_{j,k}(t)$ is narrower and is translated in smaller steps. It, therefore, can represent finer detail. For $j<0$, $\varphi_{j,k}(t)$ is wider and is translated in larger steps so these wider scaling functions can represent only coarse information and the space they span is smaller ([17], [18], [19]). Because of the definition of ϑ_j, the spaces have to satisfy a natural scaling condition. That is, if f(t) is in ϑ_j then f(2t) is in ϑ_{j+1} which insures that it belongs to the next space [17]. If we think $\varphi(t)$ is in ϑ_0 ,which means that $\varphi(t)$ is a function in the given space, it is also a member of the function space spanned by the $\varphi(2t)$.

This means that $\varphi(t)$ can be expressed in terms of a weighted sum of shifted $\varphi(2t)$ as [17]:

$$\varphi(t) = \sum_{n} h(n)\sqrt{2}\varphi(2t - n) \qquad (3.8)$$

This recursive equation is fundamental to the theory of the scaling function and is analogous to a different equation with coefficients h(n) and solutions $\varphi(t)$ that may or may not exist or be unique. The equation is called the refinement equation, the multiresolution analysis (MRA) equation, or the dilation equation ([17], [19], [20]).

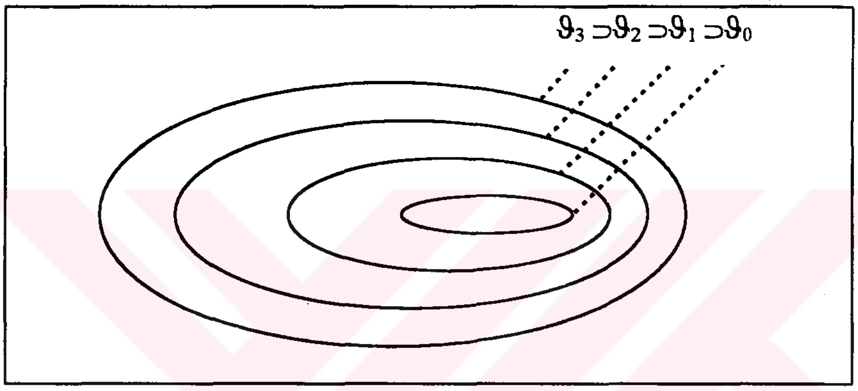

Figure (3.1) Vector Spaces Spanned by the Scaling Function

The important features of a signal can better be described or parameterized, not by using $\varphi_{j,k}(t)$ and increasing j to increase the size of the subspace spanned by the scaling functions, but by defining a slightly different set of functions $\Psi_{j,k}(t)$, i.e wavelet function, that spans the differences between the spaces spanned by the various scales of the scaling function [17]. Therefore the wavelet functions are orthogonal to the scaling functions; that is, if the span of the $\Psi_{j,k}(t)$ is W_j, and the span of the $\varphi(t)$ is ϑ_j then all members of ϑ_j are orthogonal to all members of W_j. That is [17]:

$$< \varphi_{j,k}(t), \Psi_{j,k}(t) >= \int \varphi_{j,k}(t)\Psi^{*}_{j,k}(t)dt = 0 \qquad (3.9)$$

for all j, k values. Since the wavelets reside in the space spanned by the next narrower scaling function, they can be represented by a weighted sum of shifted scaling function φ(2t) [17]:

$$\Psi(t) = \sum_n h_1(n)\sqrt{2}\varphi(2t - n) \qquad (3.10)$$

where $h_1(n)$ are defined as the wavelet function coefficients. Since orthogonality between φ(t) and Ψ(t) is required, there is a close relation between the wavelet function coefficients $h_1(n)$ and scaling function coefficients h(n). This relation is formulated as [17]:

$$h_1(n) = (-1)^n h(1 - n) \qquad (3.11)$$

For a finite even-length-N

$$h_1(n) = (-1)^n h(N - 1 - n) \qquad (3.12)$$

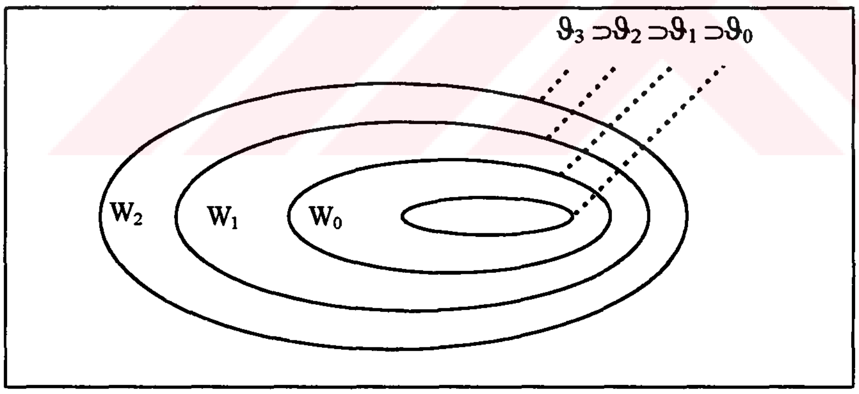

Figure (3.2) Scaling Function and Wavelet Vector Spaces

3.3 Discrete Wavelet Transform

Given a function f(t), its Discrete Wavelet Transform (DWT) is defined as [17]:

$$f(t) = \sum_{k} c_{j0}(k)\varphi_{j0,k}(t) + \sum_{k}\sum_{j=j0}^{\infty} d_j(k)\Psi_{j,k}(t) \qquad (3.13)$$

As it can be seen from Equation (3.13), in order to calculate the Discrete Wavelet Transform of a function both scaling and wavelet function had to be defined. Also in Equation (3.13), choice of j_0 sets the coarsest scale whose space is spanned by $\varphi_{j0,k}(t)$, the rest is spanned by the wavelets which provide the high resolution details of the signal. The coefficients in Equation (3.13), i.e. wavelet expansion, completely describe the original signal and can be used in a way similar to Fourier series coefficients for analysis, description, approximation and filtering. If the wavelet system is orthogonal, these coefficients can be calculated by inner products of two real functions [17]:

$$c_j(k) = < f(t), \varphi_{j,k}(t) > = \int f(t)\varphi_{j,k}(t)dt \qquad (3.14)$$

$$d_j(k) = < f(t), \Psi_{j,k}(t) > = \int f(t)\Psi_{j,k}(t)dt \qquad (3.15)$$

If we use multiresolution analysis equation and scaling and translation of the scaling function, in order to rewrite Equation (3.14):

$$\varphi(t) = \sum_{n} h(n)\sqrt{2}\varphi(2t - n) \qquad (3.16)$$

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) \qquad (3.17)$$

$$\varphi(2^j t - k) = \sum_{n} h(n)\sqrt{2}\varphi(2(2^j t - k) - n) \qquad (3.18)$$

$$\varphi(2^j t - k) = \sum_{n} h(n)\sqrt{2}\varphi(2^{j+1} t - 2k - n) \qquad (3.19)$$

setting m=2k+n and rewriting the equation:

$$\varphi(2^j t - k) = \sum_m h(m-2k)\sqrt{2}\varphi(2^{j+1}t - m) \qquad (3.20)$$

Using the inner product for $c_j(k)$:

$$c_j(k) = \langle f(t), \varphi_{j,k}(t) \rangle = \int f(t) 2^{j/2} \varphi(2^j t - k) dt \qquad (3.21)$$

$$c_j(k) = \int f(t) 2^{j/2} \sum_m h(m-2k)\sqrt{2}\varphi(2^{j+1}t - m) dt \qquad (3.22)$$

$$c_j(k) = \sum_m h(m-2k) \int f(t) 2^{j+1/2} \varphi(2^{j+1}t - m) dt \qquad (3.23)$$

$$c_{j+1}(m) = \int f(t) 2^{j+1/2} \varphi(2^{j+1}t - m) dt \qquad (3.24)$$

$$c_j(m) = \sum_m h(m-2k) c_{j+1}(m) \qquad (3.25)$$

Similarly $d_j(m)$ can be calculated as [17]:

$$d_j(m) = \sum_m h_1(m-2k) c_{j+1}(m) \qquad (3.26)$$

Equation (3.25) and (3.26) state that, the scaling and wavelet coefficients at different levels of scale can be obtained by convolving the expansion coefficients at scale j by the time-reversed recursion coefficients h(-n) and $h_1(-n)$ then downsampling or decimating (taking every other term, the even terms) to give the expansion coefficients at the next level j-1. In other words, the scale-j coefficients are "filtered" by two FIR (Finite Impulse Response) digital filters with coefficients h(-n) and $h_1(-n)$ after which downsampling gives the next coarser scaling and wavelet coefficients. The Discrete Wavelet Transform employs two sets of functions, called scaling functions and wavelet functions, which are associated with lowpass and highpass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive highpass and lowpass filtering of the time domain signal ([17], [21], [22], [23]). In Discrete Wavelet Transform, lowpass filtering removes half of the frequencies, which can be interpreted as losing half of the information. Therefore, time resolution is halved after filtering operation. However, downsampling operation after filtering does not affect the resolution, since removing half of the spectral components from the signal makes half of samples redundant anyway but the scale will be doubled afterwards. That is to say that after downsampling operation, the frequency resolution is doubled because the frequency band of the signal now spans only half the previous frequency band. The above procedure, which is also known as the subband coding, can be repeated for further decomposition. At every level, filtering and downsampling will result in half the number of samples (and hence half the time resolution) and half the frequency band spanned (and hence double the frequency resolution) [13]. This procedure is described in Figure (3.3) with an example.

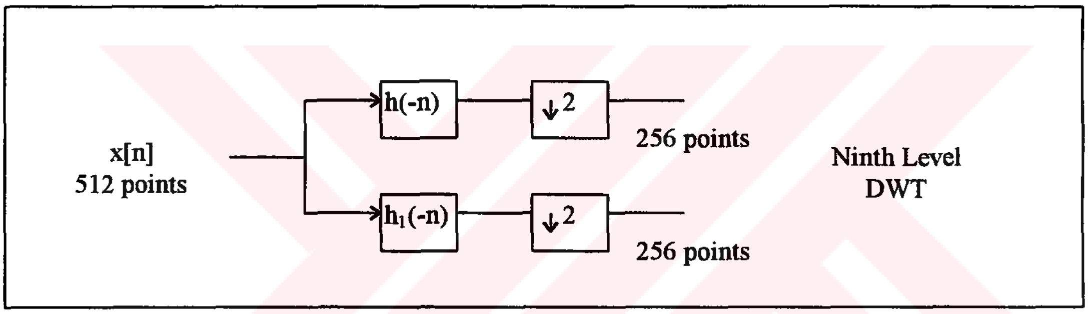

Figure (3.3) Ninth Level DWT Coefficients Calculation

In the given example, there is a signal with 512 sample points. Figure (3.3) shows the highest level of the Discrete Wavelet Transform calculation. In the same figure h(n) corresponds to the impulse response of the lowpass filter, similarly $h_1(n)$ corresponds to that of the highpass filter. These are the scaling function and wavelet function coefficients, respectively, and are calculated according to the selected scaling and wavelet functions for the calculation of the Discrete Wavelet Transform. The outputs of these filters are downsampled with a ratio of two. By this way, total number of coefficients at the outputs of the filters are equalized with number of samples at the inputs of the filters. This procedure continues until one coefficient is left. For calculation of the Discrete Wavelet Transform of a 512 sample example, there are nine levels of decomposition, each having half the number of the samples of the previous level. The Discrete Wavelet Transform of the original signal is then obtained by concatenating all coefficients starting from the first level of decomposition. The Discrete Wavelet Transform provides the same number of coefficients as the original signal.

3.4 Heisenberg Uncertainty Principle

Heisenberg Uncertainty Principle describes the fact that, a given signal x(t) can be localized simultaneously in time and frequency but it can not be concentrated simultaneously. The shorter-lived a function, the wider the band of frequencies given by its Fourier Transform; the narrower the band of frequencies of its Fourier Transform, the more the function is spread out in time [25]. In order to define the uncertainty principle, for a given function x(t), the Fourier Transform of it can be found from [12]:

$$X(f) = \int_{-\infty}^{\infty} e^{-i2\pi ft} x(t)dt \tag{3.27}$$

For every function x(t) (t a real number), such that

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = 1 \tag{3.28}$$

the product of the variance of t and the variance of f (the variable of X(f)) is at least $1/(16\pi^2)$ [25]:

$$\left(\int_{-\infty}^{\infty} (t-t_m)^2 |x(t)|^2 dt\right)\left(\int_{-\infty}^{\infty} (f-f_m)^2 |X(f)|^2 df\right) \geq \frac{1}{16\pi^2} \tag{3.29}$$

where t_m and f_m are the average values of the given functions. The left side of the inequality is the multiplication of two variances of the given functions. These variances measure to what extent t and f take values far from their average values t_m and f_m . Thus the more f is concentrated in a small window of time, the smaller the variance of t will be. If the signal is more spread out in time, the variance will be larger.

In order to apply Heisenberg Uncertainty principle in time-frequency representations, a plane in which time varies horizontally and frequency vertically, the time-frequency plane, is defined. This plane can be tiled with rectangles of size Δt and Δf. The width and the height of these rectangles are Δt and Δf where Δt represents the window of time; it is the standard deviation of time, and Δf represents the range of frequencies. Since the standard deviation is the square root of the variance, the rectangle areas can be calculated from Equation (3.29):

$$\Delta t * \Delta f \geq \frac{1}{4\pi} \tag{3.30}$$

Equation (3.30) describes the areas of rectangles in the time-frequency plane. If precise information about time is desired, a certain vagueness about frequency has to be accepted, similarly a precise information about frequency yields a certain vagueness about time [25].

CHAPTER 4

CALCULATION OF DISCRETE WAVELET TRANSFORM

4.1 Introduction

Each musical instrument produces a characteristic frequency spectrum. The frequency spectrum of the signal changes over time. Therefore time domain analysis for musical sound processing is important. For this analysis Fourier Transform can not be used because it provides only frequency information. How much of each frequency exists, is this information. Fourier Transform assumes that the frequencies in the spectrum exist at all times. That is, it assumes that the signal is stationary. Fourier Transforms of stationary and nonstationary signals having the same frequencies are the same. In order to show this, two sound samples are generated synthetically, one is stationary and the other is a nonstationary signal. Both of the samples have two frequencies, which are 500 Hz and 2000 Hz. Within the nonstationary signal, 500 Hz appears in the first half of the sampling duration, and then 2000 Hz appears for the rest of this duration. Whereas these frequencies are present at all times in the stationary signal. Figure (4.1) and (4.2) are the graphs of these two signals respectively. These signals are sampled at a rate of 11025 Hz. The figures show the first 1024 of these samples. Figure (4.3) and (4.4) are obtained by taking 1024-point FFT of these samples. In the spectrum of the nonstationary signal, there are some noises around the peaks this is due to sudden change between the frequencies. Except these noises, both spectra are similar. These figures reveal that 500 Hz and 2000 Hz exist in the signal.

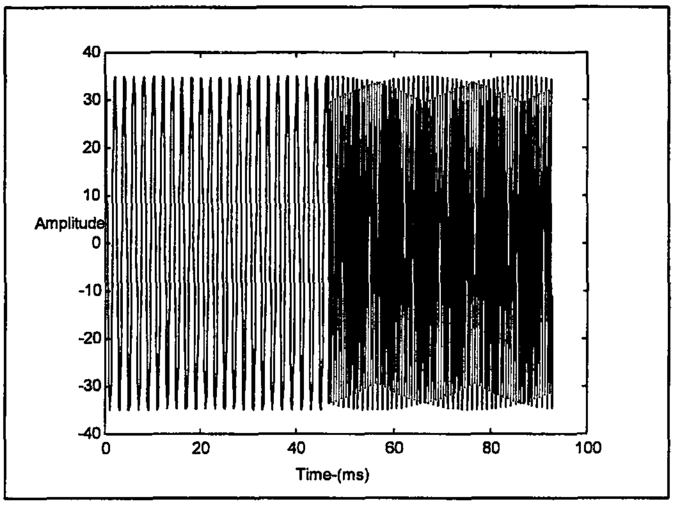

Figure (4.1) A nonstationary signal having 500Hz and 2000Hz

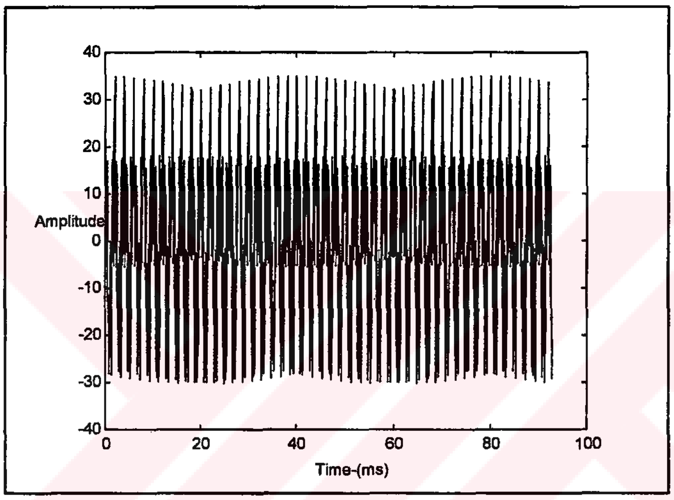

Figure (4.2) A stationary signal having 500Hz and 2000Hz

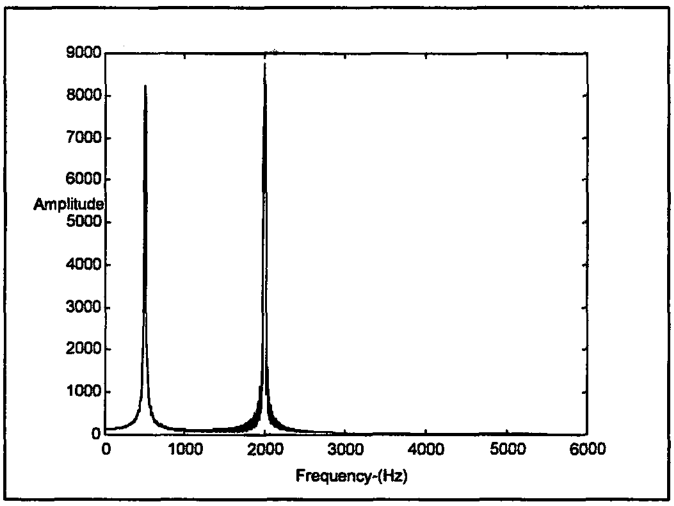

Figure (4.3) Spectrum of the nonstationary signal having 500Hz and 2000Hz

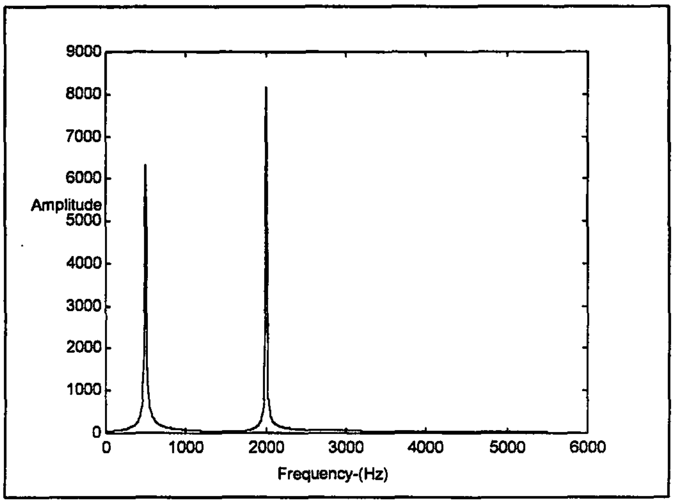

Figure (4.4) Spectrum of the stationary signal having 500Hz and 2000Hz

For time-frequency analysis of musical sound, Discrete Wavelet Transform can be used. A complete time-frequency information can not be obtained even by the Wavelet Transform. This is due to the resolution problem. The uncertainty principle originally found and formulated by Heisenberg, states that, the momentum and the position of a moving particle can not be known simultaneously [18]. Similarly for the signal processing case, frequency and time information of a signal at some certain point in time-frequency plane can not be known. In other words: what spectral component exists at any given time instant can not be known. The best that can be done is to investigate what spectral components exist at any given interval of time [18].

Calculation of the Discrete Wavelet Transform can be divided into different stages. The first stage is the selection of the scaling and wavelet functions that determine the wavelet system. The wavelet system refers both of the functions from here after. With selection of the wavelet system, the characteristics of both highpass and lowpass filters are selected because the wavelet system coefficients define these filters. These coefficients are the solutions of the multiresolution analysis equation (MRA) defined as Equation (3.8). These solutions can be found in any reference book on wavelet transform. During the second stage, the signal is passed through these filters at different levels. Filtering corresponds to the convolution of the signal with the impulse responses of these filters. The impulse response of the lowpass filter is the scaling function coefficients, defined as h(n) in Chapter 3 and similarly the wavelet function coefficients, defined as $h_1(n)$ in Chapter 3, and is that of the highpass filter. In order to obtain equal number of samples after convolution, the outputs of the filters are subsampled by two. The subsampled outputs of the highpass filter are the wavelet coefficients at that level. The subsampled outputs of the lowpass filter are the input coefficients for the next level. This process can continue until one coefficient is left but it can be stopped at any level. At any level throughout the transform, there are 2^n discrete wavelet coefficients where n corresponds to the level number. The last stage of the transform is display of the calculated coefficients. This stage is the most difficult part of the transform, because the calculated wavelet coefficients carry the information both on time and frequency at different resolution. As mentioned earlier in Section 3.4, time-frequency plane is partitioned into tiles by the Discrete Wavelet Transform coefficients $d_{j,k}$, where j defines the level, and k is the bin number at that level. Figure (4.5) is time-frequency plane, which is divided into tiles with a five level Discrete Wavelet Transform.

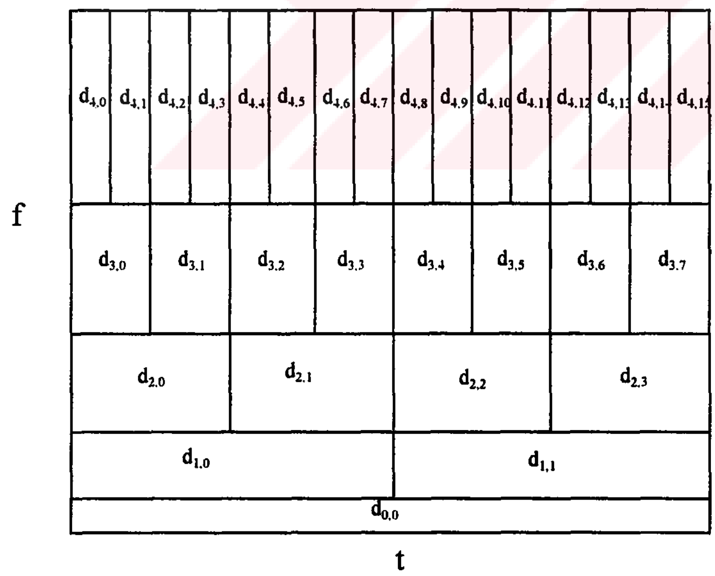

Figure (4.5) Time-frequency relation of DWT coefficients

If tiles are thought as boxes, it can be easily seen that although widths and heights of these boxes change, the area is constant. That is, each coefficient of the Discrete Wavelet Transform calculated at different levels, represents an equal portion of the time-frequency plane, but giving different proportions to time and frequency. Regardless of the dimensions of the boxes, the areas of all boxes are the same and determined by Heisenberg's inequality that is the areas of the boxes are lower bounded by $1/(4\pi)$, as described in Section 3.4. It is observed from Figure (4.5) that higher frequencies are better resolved in time, and lower frequencies are better resolved in frequency. This means that, a certain high frequency component can be located in time (with less relative error) than a low frequency component ([18], [19]).

4.2 Example Scaling Functions and Wavelets

Below, four commonly used families of scaling and wavelet functions are defined. These are the Haar, Sinc, Spline and Daubechies. After they are defined, each function family is used to illustrate the calculation of the Discrete Wavelet Transform on the 500 Hz and 2000 Hz stationary and nonstationary signals, used in Section 4.1, above.

4.2.1 Haar Wavelet

Haar Wavelet System is the oldest and most basic of the wavelet systems that has most of the desired properties. The scaling function is a unit length step function (i.e. $\varphi(t)=1$ for $0<t<1$, and 0 otherwise)[17]. The unique coefficients are [17]:

$$h(n) = \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} \tag{4.1}$$

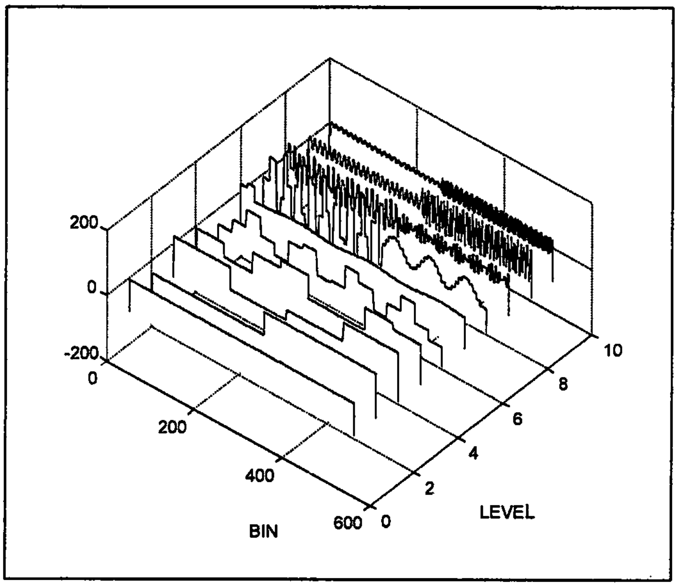

Figure (4.6) DWT of nonstationary signal using the Haar Wavelet System

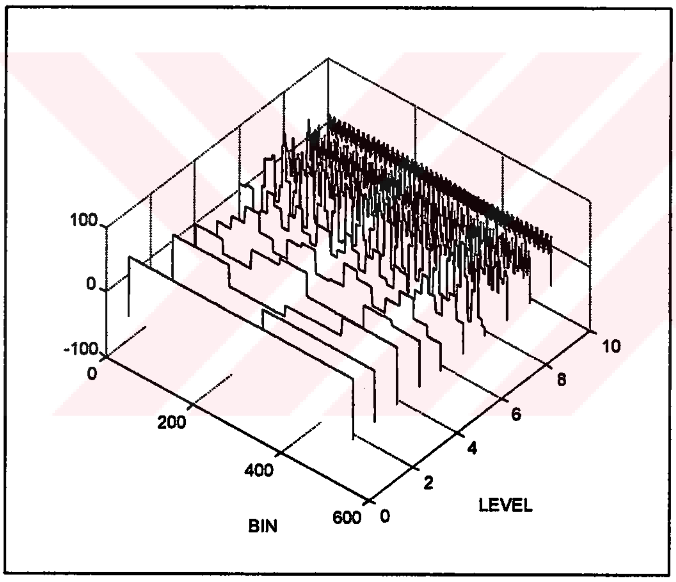

Figure (4.7) DWT of stationary signal using the Haar Wavelet System

The above two figures are Discrete Wavelet Transforms of the nonstationary and stationary signals using the Haar Wavelet System, respectively. These figures are 3-D representation of the transform where the axes are amplitude, bin and level. The bin axis corresponds to time axis and the level axis corresponds to that of frequency. For the calculation, 1024 points are sampled with a rate of 11025 Hz. If N is defined as upper value of the level of the Discrete Wavelet Transform, it is calculated from:

$$N = \log_2(1024) = 10$$

This means that, the operations of filtering and subsampling the outputs of the filters by two, had to be performed at most ten times. The highest level, that is the tenth level of the transform corresponds to a band of frequency between the Nyquist frequency (i.e. half of the sampling rate) and half of that frequency. This band is represented with half of the window size, in these examples 512-points. This level provides maximum information on time. The calculated coefficients at lower levels are represented with lower window sizes. Therefore if the calculated coefficients are expanded so that there are 512-points at each level, a matrix having a dimension of 10x512 is obtained. The row of this matrix, is the level and the column of it, is the bin number. The elements of this matrix are the calculated wavelet coefficients. From this matrix, 3-D graph of the transform can be obtained.

From the above figures, it is seen that, the coefficients between the seventh and the tenth level are meaningful. In order to obtain the frequency bands at each level of the transform, Figure (4.5) is analyzed again. Figure (4.5) is time-frequency plane of a five level DWT. That is frequency domain is divided into five unequal bands. From the Nyquist theorem, the maximum frequency information is half of the sampling rate. Therefore, frequency domain is between 0 Hz up to half of the sampling rate. Both of the lowpass and the highpass filters used during the transform, are half band type and therefore one removes all frequencies that are above half of the highest frequency of the level and the other removes all the frequencies that are below. Similarly, 1024-point DWT divides the frequency band into ten unequal bands. The levels and corresponding frequency bands are listed in Table (4.1) for a sampling rate of 11025 Hz.

Table (4.1) The Frequency Band of DWT for 1024 point and 11025 Hz sampling rate

Level	Frequency Band (Hz)
1	5.375-10.75
2	10.75-21.5
3	21.5-43
4	43-86.12
5	86.12-172.25
6	172.25-344.5
7	344.5-689
8	689-1378.125
9	1378.125-2756.25
10	2756.25-5512.5

4.2.2 Sinc Wavelet

The sinc basis set is formed by the sinc functions. The sinc function is defined as [17]:

$$\mathrm{sinc}(t) = \frac{\sin(t)}{t} \tag{4.2}$$

The scaling function is defined as [17]:

$$\mathrm{sinc}(Kt) = \sum_{n} \mathrm{sinc}(KTn)\mathrm{sinc}\left(\frac{\pi}{RT}t - \frac{\pi}{R}n\right) \tag{4.3}$$

Here the period T is ½ and the parameter K is defined as [17]:

$$K = \frac{\pi}{R} \tag{4.4}$$

Then the scaling coefficients are calculated from [17]:

$$h(n) = \mathrm{sinc}\left(\frac{\pi}{2R}n\right) \tag{4.5}$$

If R=1, then K=π and the scaling function generates an orthogonal wavelet system, where n defines the number of coefficients.

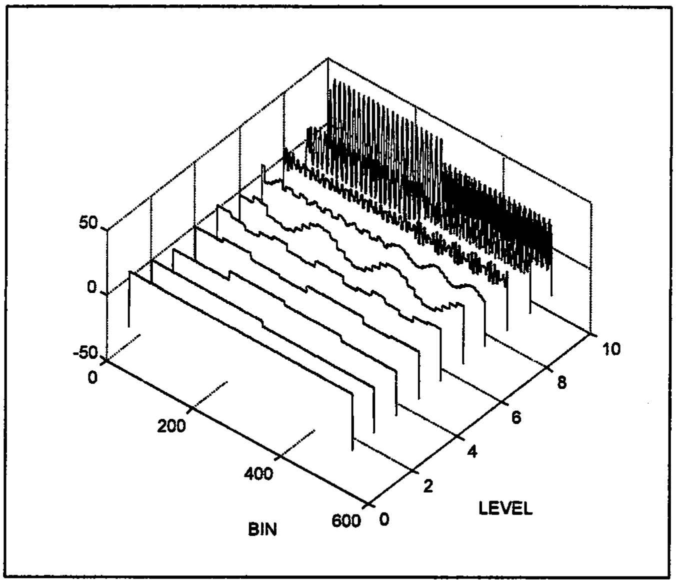

Figure (4.8) DWT of nonstationary signal using the Sinc Wavelet System with an order of 5 (n=5)

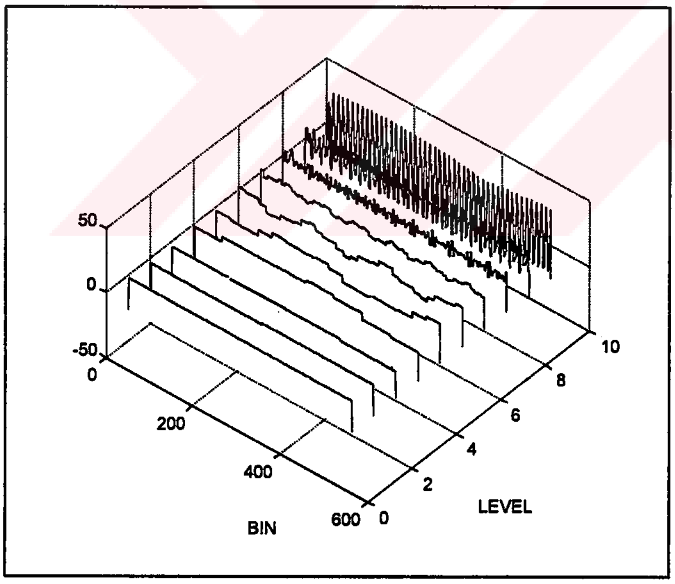

Figure (4.9) DWT of stationary signal using the Sinc Wavelet System with an order of 5 (n=5)

4.2.3 Spline and Battle-Lemarie Wavelet System

The triangle scaling function is a special case of a more general family of spline scaling functions. The scaling coefficient system

$$h(n) = \left\{ \frac{1}{2\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}, 0 \right\} \tag{4.6}$$

gives rise to the piecewise linear, continuous triangle scaling function [17].

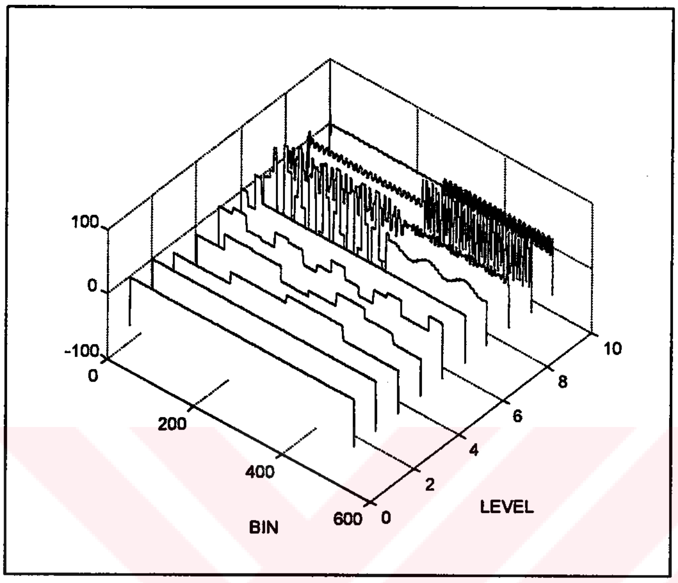

Figure (4.10) DWT of nonstationary signal using Spline Wavelet System (order 1)

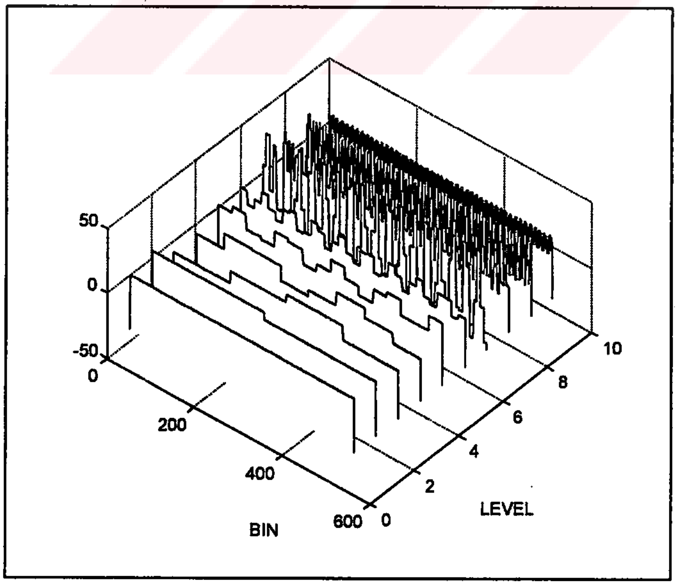

Figure (4.11) DWT of stationary signal using Spline Wavelet System (order 1)

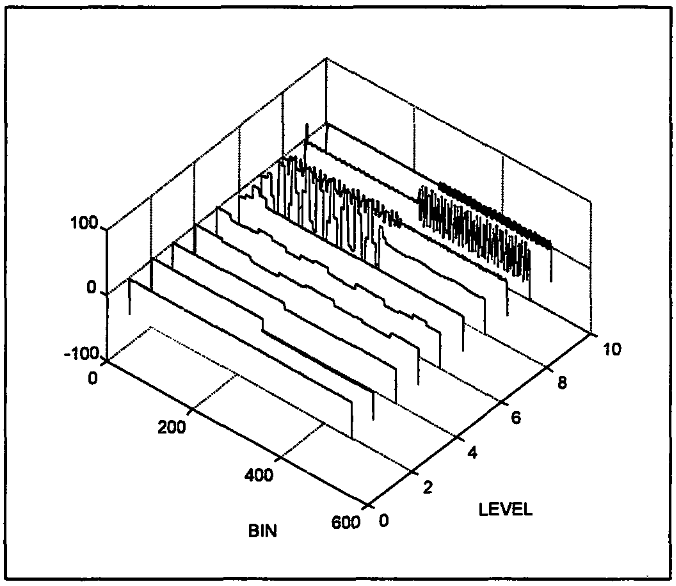

Figure (4.12) DWT of nonstationary signal using Spline Wavelet System (order 2)

4.2.4 Daubechies Wavelet System

Ingrid Daubechies constructed orthonormal wavelets with compact support [29]. The 'n' values in the succeeding figures are half of the number of the scaling and wavelet coefficients. Since these coefficients define the impulse responses of the filters in the Discrete Wavelet Transform, both filters are represented with 2*n coefficients [17].

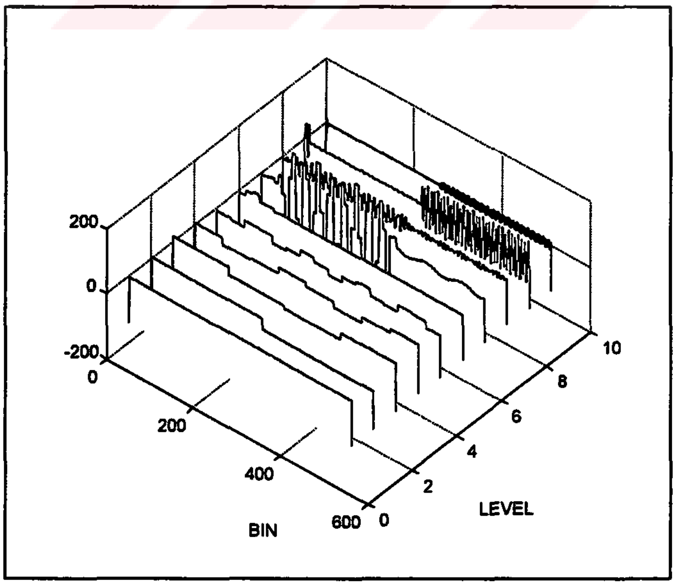

Figure (4.13) DWT of nonstationary signal, using Daubechies Wavelet System (n=5)

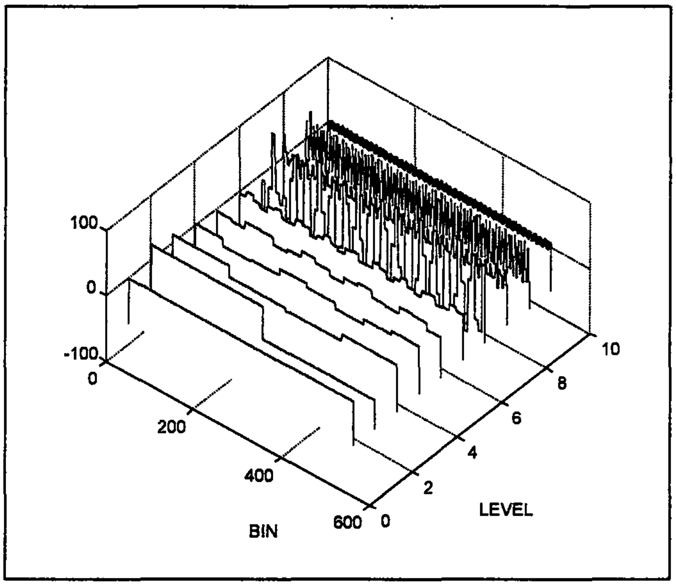

Figure (4.14) DWT of stationary signal, using Daubechies Wavelet System (n=5)

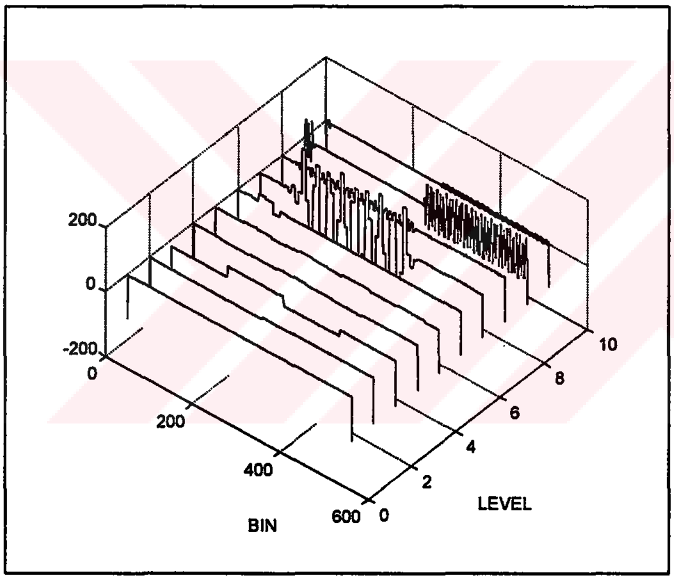

Figure (4.15) DWT of nonstationary signal, using Daubechies Wavelet System (n=10)

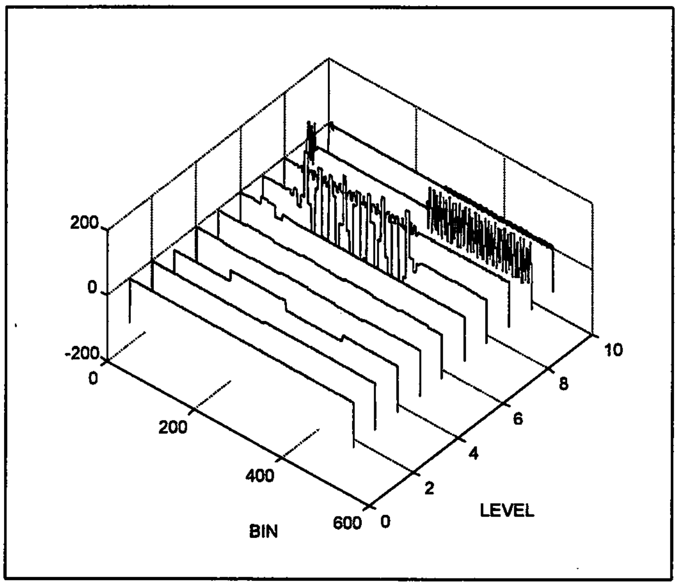

Figure (4.16) DWT of nonstationary signal, using Daubechies Wavelet System (n=20)

4.3 Comparison of the Wavelet Systems

In the previous part, four different wavelet systems are defined. These are the Haar, Sinc, Spline and Daubechies wavelet systems. In order to compare these systems, the Discrete Wavelet Transforms of stationary and nonstationary signals are taken by using the related wavelet systems, respectively. Also different orders from the defined wavelet system are taken and used to calculate the Discrete Wavelet Transform of the nonstationary signal in order to analyze differences on time domain from the results. From Table (4.1), it is found that 500 Hz and 2000 Hz are within a frequency band corresponding to level seven and nine of the transform. Since no other frequencies exist in these signals, the coefficients at the other levels are expected to be zero or near to zero.

Figure (4.6) and (4.7) illustrate the Discrete Wavelet Transforms of the nonstationary and stationary signals, using the Haar Wavelet System, respectively. From the DWT of the stationary signal, it is seen that these frequencies exist at all times. From the DWT of the nonstationary signal, it is seen that separation of these frequencies in time is perfect. But from the frequency point of view, although the coefficients at the levels seven and nine have the maximum values, the coefficients at the eighth and tenth level are also meaningful. DWTs of the signals using the Haar Wavelet System give rise to error on the lower levels because there is some information below the seventh level.

Figure (4.8) and Figure (4.9) represent the Discrete Wavelet Transforms of the defined signals, using the Sinc Wavelet System with an order of 5. It is seen from these figures that, the Sinc Wavelet System provides information neither on the time domain nor on the frequency.

Figure (4.10) and Figure (4.11) represent the Discrete Wavelet Transforms of the defined signal using the first order Spline Wavelet System. Figure (4.12) is the graph of DWT of the nonstationary signal using the second order Spline Wavelet System. It is seen from these three figures that, the calculated coefficients are near zero for the levels lower than seven. Also the magnitudes of calculated coefficients at the eighth and tenth level are lower than the corresponding level coefficients of the calculation based on the Haar Wavelet System. Therefore, the second order Spline Wavelet System provides the best frequency information if the Haar and both of the Spline Wavelet Systems are compared. But from the comparison of time information provided by these systems, the Haar Wavelet System has still the perfect information.

From the calculations using the different orders of the Daubechies Wavelet System, it is seen that the frequency information provided increases with increasing transform orders. Figure (4.16) is the most meaningful figure from the frequency point of view, because the coefficients at the lower levels are near zero. The magnitudes of the coefficients at the eighth and tenth level are dropped by an appreciable amount. But the time localization of the frequencies are worse for this system, compared with all other systems.

In all, a wavelet system that is seen to provide best time and frequency information together can not be found. A gain on time information corresponds to a lose on frequency information. Therefore a "mid-range" wavelet system had to be searched. Spline Wavelet System of order 2 and Daubechies Wavelet System of order 10 can be selected as candidate from the calculations on stationary and nonstationary signals. But in order to observe the performances on instrument samples, also the other wavelet systems are applied to the same samples. Evaluation of these attempts and comparison from the results will be presented below.

4.4 Discrete Wavelet Transforms of Musical Instrument Samples

The musical instrument sample files used throughout this part, are selected from the files defined in Chapter2. For the time-frequency analysis, first 16384 integer samples are taken from these files. In order to obtain 22050 Hz sampling rate, the even numbered samples are dropped, leading to 8192 samples for the calculation of Discrete Wavelet Transform. Table (4.2) defines the frequency bands corresponding to the levels of the transform. The total number of samples leads to a 13-level DWT. The frequency bands corresponding to the first and second level are small, near 0, therefore these values are not calculated. Also for calculation of Constant Q Transforms of these files, the minimum frequency is selected as 183 Hz. This frequency is in the level eight in Table (4.2). Therefore, the calculations are stopped at that level. The Daubechies Wavelet System of order 10 is applied to all of the instrument samples. Also the other wavelet systems are applied and their results are compared with results from DWT using the Daubechies Wavelet System.

Figure (4.17) is the figure of the first 8192 samples of saxophone playing C5 (C5=525 Hz). Figure (4.18) is the DWT of this sample file. Instead of bin axis, in Figure (4.18) there is time axis. The values on time axis are calculated from bin numbers of the calculated coefficients. The level axis and the related frequencies are in Table (4.2). If both figures are examined, with DWT the time information of the signal is preserved. Variations of the frequencies in time can be well defined but the exact values of the frequencies can not. If saxophone sample signal is analyzed by using Figure(4.17), gray lines correspond to lower frequencies and darker ones to higher frequencies. If the lower frequency of the signal is taken as the fundamental frequency, as it is the true case, the variation of the fundamental is represented at the ninth level of the DWT. Also since the first harmonic of this fundamental falls in the band of level ten, the tenth level of the transform represents this time variation of the first harmonic. Figure (4.19) illustrates the DWT of the same sample file using the second order Spline Wavelet System.

Table (4.2) The Frequency Band of DWT for 8192 point and 22050 Hz sampling rate

Level	Frequency Band (Hz)
1	*
2	*
3	5.375-10.75
4	10.75-21.5
5	21.5-43
6	43-86.12
7	86.12-172.25
8	172.25-344.5
9	344.5-689
10	689-1378.125
11	1378.125-2756.25
12	2756.25-5512.5
13	5512.5-11025

Figure (4.20) is the figure of the first 8192 samples of guitar playing C5 (C5=525 Hz). For the plucked-string instruments, the decay time of the emitted tone is relatively long and these instruments do not produce a steady-state output [4]. Figure (4.21) illustrates the DWT of the guitar sample. The fundamental frequency is at the same level. The magnitude of the fundamental frequency has a peak at the attack part of the signal and decreases during the decay part of the signal. At the other levels of the transform, there appear the harmonics of the fundamental frequency. The harmonics also have a maximum magnitude at the attack part of the signal. These values decrease also during the decay part. For the time after 300 ms, there are only the fundamental and the first harmonic of the signal. Figure (4.22) represents the DWT of the same file using the Haar Wavelet System. The coefficients at the eighth level are not zero for this calculation.

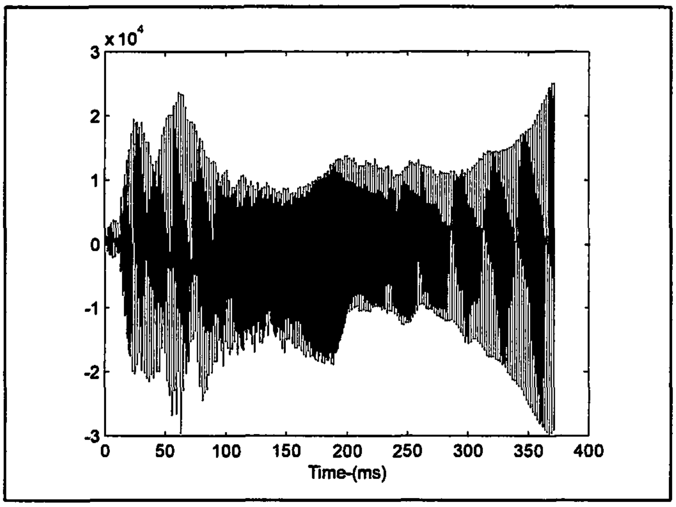

Figure (4.17) The sample of saxophone playing C5

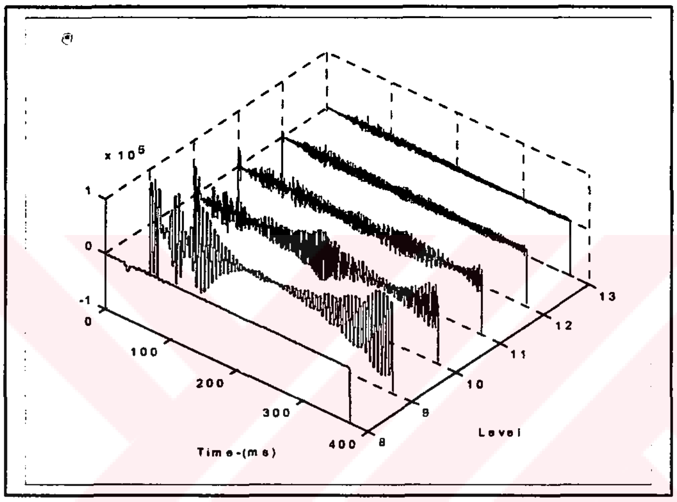

Figure (4.18) DWT of saxophone playing C5, using Daubechies Wavelet System (n=10)

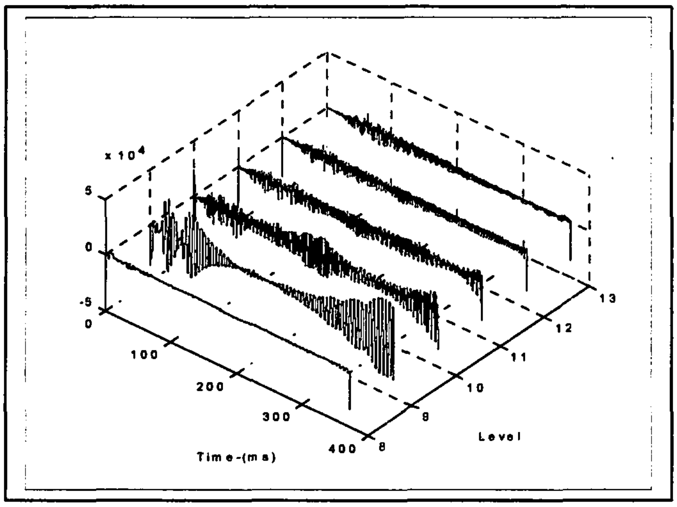

Figure (4.19) DWT of saxophone playing C5, using Spline Wavelet System (order 2)

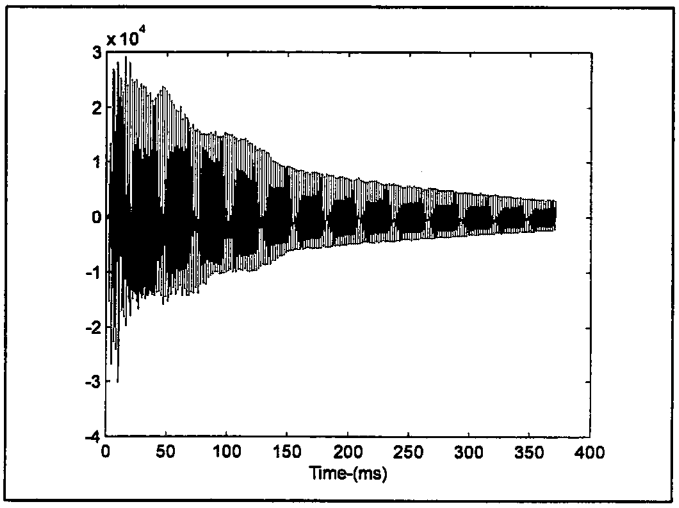

Figure (4.20) The sample of guitar playing C5

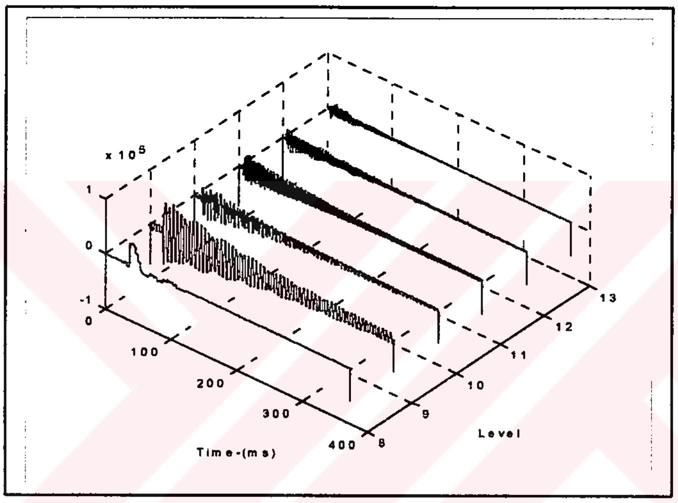

Figure (4.21) DWT of guitar playing C5, using Daubechies Wavelet System (n=10)

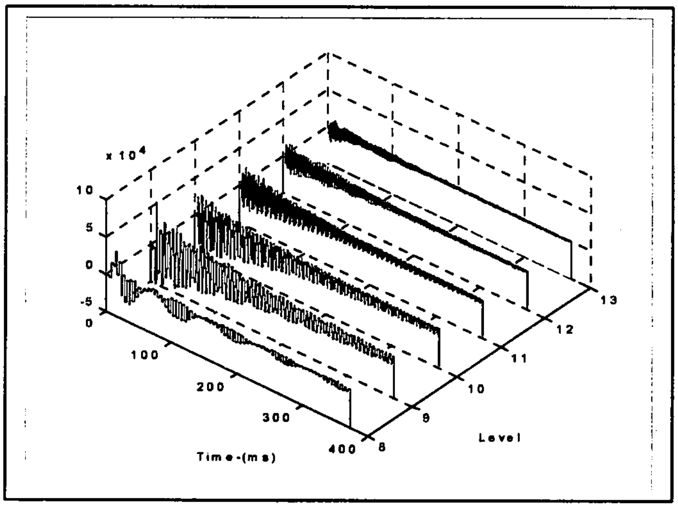

Figure (4.22) DWT of guitar playing C5, using Haar Wavelet System

Figure (4.23) presents the first 8192 samples of piano playing C5 (C5=525 Hz). For the struck-string instruments, the decay time of the emitted tone is longer than the plucked-string instruments and also this kind of instruments do not produce a steady-state output [4]. Figure (4.24) illustrates the DWT of the piano sample. The attack and the decay part of the signal can be extracted from the DWT of the instrument sample signal. The decay time of the emitted tone from piano is longer than that of the guitar. It is seen that during the decay time, the fundamental frequency and the first two harmonics of the fundamental subsist. Figure (4.25) illustrates the DWT calculation of the same file using the Sinc Wavelet System of order 5.

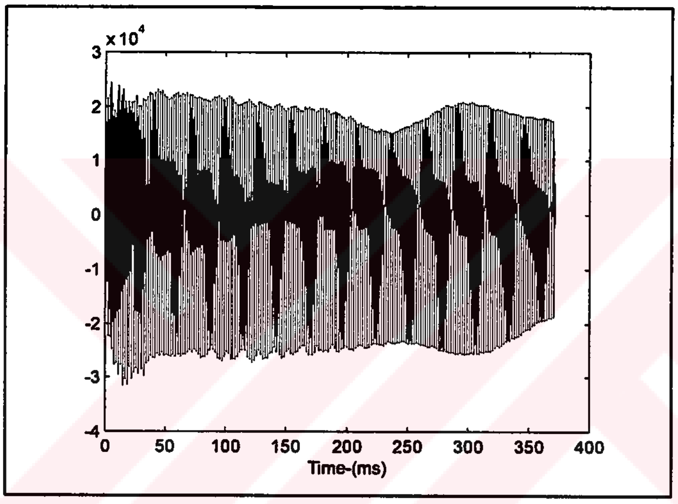

Figure (4.23) The sample of piano playing C5

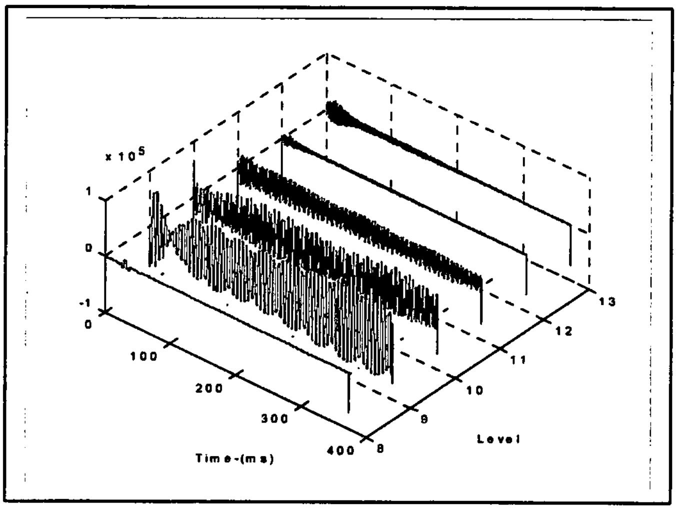

Figure (4.24) DWT of piano playing C5, using Daubechies Wavelet System (n=10)

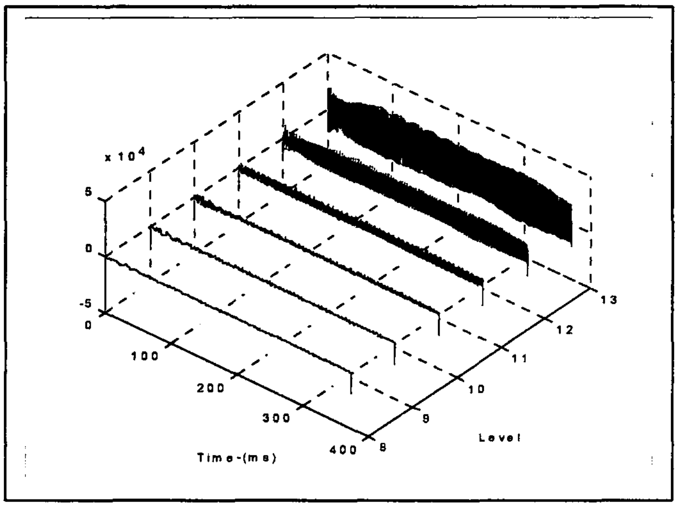

Figure (4.25) DWT of piano playing C5, using the Sinc Wavelet System (n=5)

CHAPTER 5

DISCUSSION

All of the codes written in this thesis have been developed separately with MATLAB and C++. The codes developed on MATLAB have been used because MATLAB is designed to cover the requirements of signal processing. Results and graphics obtained from MATLAB Notebook Suite have been included in the related chapters.

In the first chapters, DFT and Constant Q Transforms were applied to the same sound files. Based on a comparison of the results, it was shown that the Constant Q Transform was able to fulfill musical sound processing requirements to a better extent than DFT. DFT was not as effective because musical sound frequencies are distributed geometrically and DFT supplies constant frequency resolution for all of the frequencies. Constant Q Transform can be used as it supplies a constant pattern to recognize fundamental frequency from the frequency spectrum. Another result obtained from the calculation of Constant Q Transform is that, although Constant Q Transform supplies the above advantages, its calculation is inefficient. For this reason, a study was carried out to find out a method like Fast Fourier Transform for faster calculation of Constant Q Transform. But as sampling window is variable for each frequency bin, this could not be implemented. Instead, an efficient algorithm for the calculation of Constant Q Transform was attempted. This efficient algorithm is based on Parseval Equation. This fast method basically moves the calculation from time domain to frequency domain. Therefore, it was shown that, a calculation done with the same frequency spacing, same sampling rate and same minimum desired frequency can be used for another transform calculation. Further, the efficiency of the transform was enhanced with a parameter whose value can be controlled. With the value of this parameter, a specific amount of error was added to the transformation, and based on the desired calculation efficiency and tolerance for this error, a default value was determined for the related parameter.

In order to compare the run-times of two Constant Q methods and DFT, the number of complex multiplications that are performed for the calculation are taken as a basis. For calculation of Fourier transform, FFT is used. If N is the width of the sample window, the complexity of FFT is defined as $O(N\log_2 N)$, similarly the complexity of DFT is defined as $O(N^2)$ [24]. In order to define the complexities of two methods for the calculation of Constant Q Transforms, the following series of steps are performed. If semitone spacing is presumed between adjacent frequencies, the window size desired for the calculation of the k^{th} bin is defined in Equation (2.13) as:

$$N[k] = (S/f_k)Q \tag{5.1}$$

Similarly for the $(k+1)^{th}$ bin:

$$N[k+1] = (S/f_{k+1})Q \tag{5.2}$$

The ratio of these two window sizes is evaluated as:

$$\frac{N[k+1]}{N[k]} = \frac{f_k}{f_{k+1}} \tag{5.3}$$

There is semitone spacing between the frequencies, therefore the frequency of the k^{th} bin is calculated from Equation (2.3):

$$f_k = (2^{1/12})^k f_{min} \tag{5.4}$$

If Equation (5.3) is rewritten using Equation (5.4):

$$\frac{N[k+1]}{N[k]} = \frac{(2^{1/12})^k f_{min}}{(2^{1/12})^{k+1} f_{min}} \Rightarrow N[k+1] = N[k]\frac{1}{(2^{1/12})} \tag{5.5}$$

If the bin number of the minimum frequency of the transform is designated as 0 then:

$$N[1] = N[0]\frac{1}{(2^{1/12})} \tag{5.6}$$

$$N[2] = N[1]\frac{1}{(2^{1/12})} = N[0]\frac{1}{(2^{1/12})^2} \tag{5.7}$$

$$N[k] = N[0]\frac{1}{(2^{1/12})^k} \tag{5.8}$$

Since the window size corresponds to the number of complex multiplications for each bin number, the total multiplications performed for transform calculation is the sum of all of these. That is:

$$\text{Number of multiplications} = N[0] + N[1] + N[2] + \ldots\ldots + N[k] \tag{5.9}$$

$$= N[0] + N[0]\frac{1}{2^{1/12}} + N[0]\frac{1}{(2^{1/12})^2} + \ldots\ldots + N[0]\frac{1}{(2^{1/12})^k} \tag{5.10}$$

$$= N[0]\sum_{n=0}^{k}\frac{1}{(2^{1/12})^n} \tag{5.11}$$

The power series are expanded as [26]:

$$\sum_{n=0}^{k} x^n = \frac{1-x^{k+1}}{1-x} \tag{5.12}$$

Similarly for infinite sum power series expansion [26]:

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \tag{5.13}$$

Using Equation (5.12) and taking N[0] as N, Equation (5.11) can be altered as:

$$\text{Number of multiplications} = N\sum_{n=0}^{k} \frac{1}{(2^{1/12})^n} = N\frac{1-(\frac{1}{2^{1/12}})^{k+1}}{1-(\frac{1}{2^{1/12}})} \tag{5.14}$$

The highest value for the bin number corresponds to Nyquist frequency of the transform, therefore depends on the sampling rate. For 1024 point Constant Q Transform with 11025 Hz sampling rate, the upper index of the summation is found as 60. For calculation of the complexity of Constant Q Transform, Equation (5.14) can be simplified. If instead of finite series expansion, infinite summation can be applied in Equation (5.11), a new equation can be obtained as:

$$\text{Number of multiplications} = N\sum_{n=0}^{k} \frac{1}{(2^{1/12})^n} = N\frac{1}{1-(\frac{1}{2^{1/12}})} \cong 17N \tag{5.15}$$

Equation (5.15) can be defined as the complexity of the transform. In the same way if quartertone spacing is used, the complexity of the transform can be obtained by replacing power of two with 1/24 instead of 1/12. This is calculated as 34*N. The quality factors (Q) of the transform for the above conditions are the same values respectively (From Equation (2.10) and (2.11)). This is by no chance because the expression in right part of Equation (5.15) is the quality factor of the transform.

Therefore the complexity of Constant Q Transform can be evaluated from:

$$\text{Complexity of Constant } Q = O(QN) \tag{5.16}$$

In the calculation of Constant Q Transform using efficient algorithm, FFT of input signal is taken. Furthermore according to an error rate, some multiplications are desired for the calculation. The amount of these is tolerable, therefore the complexity of the efficient algorithm is the same as that of FFT. If the two methods on calculation of Constant Q Transform are compared, it is seen that for 1024 point calculation and semitone spacing, the efficient algorithm runs approximately twice as faster the direct calculation of Constant Q Transform ($Q/\log_2 1024=17/10$). If quartertone spacing is used, this ratio is doubled and the efficient algorithm leads to an increase by a factor of roughly 3.5 in computational efficiency.

The definitions and properties of the Wavelet Transform are described in Chapter 3 and different wavelet systems are described and results of calculations of Discrete Wavelet Transform are presented in Chapter 4. The first reason why Discrete Wavelet Transform is applied in a part of this thesis is its newly developed concept, nearly having a background of 10-15 years, there are lots of mathematicians and engineers working on it and therefore it is open for future enhancements. The second reason is that it provides time-frequency information at the same time. Since scientists from different disciplines have worked on wavelet systems, their representation of transform and notations differ. This makes the transform difficult to realize at the outset. Although this can be considered as a disadvantage of the transform, it is obvious that this will disappear in time.

From the investigation of the calculation of Discrete Wavelet Transforms of musical instrument samples, it is observed that, the frequency resolution of DWT is the worst compared with that of Constant Q and DFT. The frequency bands corresponding to each level of the transform are geometrically distributed. Initially this has been thought as an advantage of DWT on musical sound processing, but it is realized soon that more than one harmonic of fundamental frequency fall into the upper levels, because frequency bands increase so much at the higher levels of the transform. From the results, the fundamental frequency and first two or three harmonics are resolved. But from this amount of information the complete frequency spectrum or the timbre of a musical instrument can not be determined. On the other hand, it must be noted that each instrument produces some information on time domain. The duration of a note, attack, decay and steady-state characteristics can be given as examples to this information. From the results of calculation of DWT, it is seen that, the above examples can be described. Attack, decay and steady-state parts of the musical instrument samples are easily observed in the time analysis of the transforms.

For calculation of the complexity of DWT, a similar work that is performed for calculation of the complexity of Constant Q Transform, is done. During calculation of DWT, successive convolution and downsampling operations are performed. With downsampling operation, sample size of the input signal is halved at each level of the transform. Therefore sample size at level k is half of that at level k+1. At any level, two convolutions are performed. One for highpass filter and other for lowpass filter. Complexity of the convolution operation is directly related to sample size of the input signal and to number of coefficients of the filter. If c is defined as number of coefficients of the filter, and N represents sample size of the signal, then complexity of convolution can be calculated as N*c. At each level of the transform, the convolution operation is applied two times, therefore complexity of DWT at a level can be defined as 2*N*c, where N defines the sample size of the input signal at a level. The process can be continued until one coefficient remains. Therefore the complexity of DWT can be formulated as:

$$\text{Complexity of DWT} = 2cN\sum_{n=0}^{\log_2 N}\left(\frac{1}{2}\right)^n \cong 2cN\frac{1}{1-\frac{1}{2}} \cong 4cN = O(cN) \quad (5.17)$$

For the expansion of the power series, Equation (5.13) is used. The complexity of DWT can not be directly compared with the complexity of FFT, because calculations both in FFT and Constant Q Transform are complex, whereas DWT calculations are all real. It is seen that the complexity of DWT depends on the sample size and the filter coefficients linearly. Also the Discrete Wavelet Transform can be interrupted at any level, the coefficients up to that level represent the DWT of the signal. But although it is seen that DWT have some advantages at higher sample sizes compared with FFT, it must be recalled that the results of DWT are slightly more complex to implement. This is due to the fact that the calculated wavelet coefficients carry information on time-frequency domain. On MATLAB or on similar mathematical programs, implementation of the results of DWT can be performed easily. Also the calculation of DWT on this kind of programs is faster than any of defined transforms in this study.

CHAPTER 6

CONCLUSIONS

In this thesis, alternative transform techniques on musical signal processing were described and performed on some musical instrument samples and the results are compared.

First, the Constant Q Transform was described in Chapter 2. Need for this kind of transform was illustrated on synthetically generated samples. It was observed from these illustrations that for musical signal processing, the frequency resolution of DFT was not good. This was due to constant frequency resolution of DFT. For musical signal processing applications, transform should provide a geometrical frequency resolution depending on frequency of interest. Constant Q Transform provides this kind of frequency resolution. An efficient algorithm proposed in the literature for Constant Q Transform computation was implemented. This algorithm is based on evaluation of the transform in frequency domain, instead of time domain. Parseval equation defines this. The algorithm is thought to be efficient because it enables the use of FFT for the frequency components. The spectral kernel values, defined in Chapter 2, were small enough so that some of these values could be dropped during the calculation of the transform. The calculation was repeated with different values of an adjustable parameter which defined the lower value of acceptable spectral kernel values. It was observed that increasing the value of this parameter, error added into the calculated coefficients increased but number of multiplications performed were decreased. Therefore a value for this parameter was searched yielding minimum multiplications and acceptable error on the calculated coefficients.

Chapters 3 and 4 were reserved for the Wavelet Transform. In Chapter 3, the definitions of the Wavelet Transform were presented. In Chapter 4, the calculation of the Discrete Wavelet Transform was presented using different wavelet systems. Performances of these wavelet systems were compared according to the results of DWT calculation. The Discrete Wavelet Transforms of the musical instrument samples were illustrated in Chapter 4.

In Chapter 5, all of the transforms described in this thesis were compared. The comparison was based on the complexities, time and frequency resolutions of the transforms. The complexities of the transforms were analyzed in this chapter. It was observed that the complexity of the direct calculation of the Constant Q Transform linearly depends on the window size and the quality factor of the transform. Complexity of the efficient algorithm was found to be equal to that of FFT. It was determined that if there was quartertone spacing between adjacent frequencies, the efficient algorithm lead to an increase of a factor 3.5 in computational efficiency. Similarly this factor of increase was observed to be equal to half of 3.5 for semitone spacing. Complexity of the Discrete Wavelet Transform was also found to be linearly depending on the sample size. But it was also discovered that there was a linear relation between complexity of the transform and the number of coefficients of the filter. It was observed that since the Haar filters were represented with only two coefficients, it was the fastest compared with other wavelet systems whereas it provided the worst frequency information.

From all of the results of the transforms, it was seen that Constant Q Transform provided the finest frequency information and its calculated coefficients mapped to musical frequencies in better way. Therefore for the applications where there is a need for recognition of fundamental frequency, or musical instrument timbre, this transform can be used which yields better information. It resolves semitone spaced frequencies which is the true case for musical frequencies. From the complexities of these transforms, it was observed that the Constant Q Transform was worst for lower window sizes because the quality factor of the transform, Q is greater than logarithm of the window size. From this point of view, efficient algorithm can be used which has a better complexity than that of the direct calculation. The same sound samples were used for the calculation of the Discrete Wavelet Transform. From these results, it was seen that the frequency resolution of the Discrete Wavelet Transform was the worst because the Discrete Wavelet Transform did not provide the exact values of the frequencies rather it provided a frequency band. Therefore it can not be used in pitch recognition applications. But from time-domain analysis of the Discrete Wavelet Transform, it was observed that the variations of the partials could be analyzed. These variations can be used in order to recognize the temporal characteristics of the sound generated by different musical instruments.

As a future study, the developing wavelet systems can be applied for the calculations of the Discrete Wavelet Transform. Within these systems, a best basis can be searched for musical signal processing. Depending on the results of the Discrete Wavelet Transforms, musical instrument recognition based on temporal characteristics, can be implemented. The Constant Q Transform can be repeated for different window sizes in a sample file, so that it can be implemented in order to provide time information.

REFERENCES

[1] J. S. Rigden, *Physics and the Sound of Music*, John Wiley & Sons New York, 1977.

[2] H. L. F. Helmholtz, *On the Sensations of Tone* , Dover Publications Inc. New York, 1954.

[3] H. Lowery, *A Guide to Musical Acoustics*, Dennis Dobson Ltd. London, 1965.

[4] H. F. Olson, *Music, Physics & Engineering*, Dover Publications Inc. New York, 1967.

[5] M. Ö. İzmirli, *Computer Analysis of Musical Sound*, Master Thesis, Department of Electrical Electronics Engineering, M.E.T.U. Ankara, 1989.

[6] E. Germen, *Recognition and Transcription of One Part Melody with Computer*, Master Thesis, Department of Electrical Electronics Engineering, M.E.T.U. Ankara, 1991.

[7] J. A. Moorer, "Signal Processing Aspects of Computer Music: A Survey", *Proc. of the IEEE*, vol. 65, no. 8, pp. 1108-1137, 1977.

[8] J. C. Brown, "Calculation of a Constant Q Spectral Transform", *J. Acoust. Soc. Am.*, vol. 89, no. 1, pp. 425-434, 1991.

[9] J. C. Brown, M. S. Puckette, "An Efficient Algorithm for the Calculation of a Constant Q Transform" *J. Acoust. Soc. Am.*,vol. 92, no. 5 pp. 2698-2701, 1992.

[10] S. J. Orfanidis, *Introduction to Signal Processing*, Prentice-Hall Inc. New Jersey, 1996.

[11] J. G. Proakis, D. G. Monolakis, *Digital Signal Processing Principles, Algorithms and Application*, Prentice-Hall Inc., New Jersey 1996.

[12] A. V. Oppenheim, A. S. Willsky, *Signals and Systems*, Prentice Hall Inc., New Jersey, 1983.

[13] "Instrument Samples", *http://www.maz-sound.com/instruments.html*, 1998.

[14] D. Cross, "Reading and Writing Wav Files", *http://intersrv.com/~dcross/wavio.html*, 08 July 1998.

[15] B. Gunter, *File Formats Handbook,* International Thompson Computer Press, Boston, 1995.

[16] C. K. Chui, *An Introduction to Wavelets*, Academic Press Inc.,New York, 1992.

[17] C. S. Burrus, R. A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall Inc., New Jersey, 1998.

[18] R. Polikar, "Multiresolution Analysis and The Continuous Wavelet Transform" *http://www1.iastate.edu/~rpolikar/wavelets/wtpart3.html*, 06 June 1996.

[19] C. K. Chui, *Wavelets:*A *Theory and Applications*, Academic Press, Inc., New York, 1992.

[20] Y. Meyer, *Wavelets Algorithms and Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 1993.

[22] R. Polikar, "Multiresolution Analysis and The Discrete Wavelet Transform" *http://www1.iastate.edu/~rpolikar/wavelets/wtpart4.html*, 13 June 1997.

[23] Y. Meyer, *Wavelets and Operators*, Press Syndicate of the University of Cambridge, New York, 1992.

[24] P. M. Embree, B. Kimble, *C Language Algorithms for Digital Signal Processing*, Prentice-Hall Inc., New Jersey, 1991.

[25] B. B. Hubbard, *The World According to Wavelets*, AKPeters Ltd., Wellessey, 1995.

[26] R A. Silvermann, *Calculus with Analytic Geometry*, Prentice Hall Inc., New Jersey, 1985.

[27] G. Strang, "Wavelets" *American Scientist*, vol. 82, no. 3, pp. 250-255 1994.

[28] C. K. Chui, *Wavelets: A Mathematical Tool for Signal Analysis*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[29] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[30] R. Polikar, "The Fourier Transform and The Short Term Fourier Transform" *http://www1.iastate.edu/~rpolikar/wavelets/wtpart2.html*, 06 June 1996.

[31] "Standard Musical Note Frequencies", *http://www.seathree.demon.co.uk/contents.html*, 02 August 1998.

[32] A. Graps, "Wavelet Page", *http://www.amara.com/current/wavelet.html*, 10 June 1998.

[33] J. C. Brown, M. S. Puckette, "A High Resolution Fundamental Frequency Determination Based on Phase Changes of the Fourier Transform", *J. Acoust. Soc. Am.*, vol. 94, no. 2, pp.662-667, 1993.

[34] M. B. Ruskai, G. Beylkin, R. Coifman, *Wavelets and Their Applications*, Janes and Bartlett Publishers, Boston, 1992.