COMPUTER RECOGNITION OF MUSICAL INSTRUMENTS: AN EXAMINATION OF WITHIN CLASS CLASSIFICATION

A THESIS SUBMITTED TO THE
SCHOOL OF COMPUTER SCIENCE AND
MATHEMATICS
VICTORIA UNIVERSITY
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Robert Moore
June 2007

Doctor of Philosophy Declaration

"I, Robert Moore, declare that the PhD thesis entitled 'Computer Recognition of Musical Instruments: An Examination of Within Class Classification', is no more than 100,000 words in length, exclusive of tables, figures, appendices, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work".

Signature: Date:

Acknowledgements

The work in this thesis was under taken during a scholarship in the school of Computer Science and Mathematics at Victoria University, and later in a continuing lecturer position in the same school.

Thanks are to my numerous supervisors. To Dr Charles Osborne, my initial co-supervisor, for the idea for the proposal (my interest in this topic has sustained me through difficult times); Tom Peachey, my initial supervisor, for hours of tuition on the background mathematics required to undertake this thesis; Associate Professor Pietro Cerone, my second supervisor, for his helpful advice, guidance and feedback through the bulk of this thesis; Dr Neil Diamond, my second co-supervisor, for his advice and support at difficult times and for the germ of the idea that enabled such good classification results; and Dr Greg Cain who gave much advice and encouragement,

I would also like to acknowledge the support and encouragement given by my co-students and colleagues in the school of Computer Science and Mathematics. In particular, Associate Professor Neil Barnett who read my work and gave me valuable feedback; Foster Hayward for his excellent work recording the music samples; Dr John Roumeliotis for help with Latex and general encouragement; Dr Jakub Szajman for his help with Latex; Dr Ian Kaminskyj (Monash Uni.) for stimulating discussions on the topic of 'musical instrument recognition'; and to Professor Judith Brown (Wellesley College, US) for making available her MATLAB code for the Constant Q Transform. I also owe a great debt to Olivia Calvert, who offered her time and expertise to play the violin tones that are a significant part of the data in this thesis.

I would also like to acknowledge my friends for their support and encouragement, which helped to sustain me over the years. In particular, David Jones, Frank Revell and my friends at Kensington-Flemington junior sports club.

Most of all, I would like to thank my family, partner Chris and children Cailean and Brighde, for their love, support and understanding. I would like to acknowledge the sacrifices they have made in terms of my time away from them.

To my children Cailean Moore and Brighde McGrath.

Abstract

This dissertation records the development of a process that enables within class classification of musical instruments. That is, a process that identifies a particular instrument of a given type - in this study four guitars and five violins.

In recent years there have been numerous studies where between class classification has been attempted, but there have been no attempts at within class classification. Since *timbre* is the quality/quantity that enables one musical sound to be differentiated from another, before any classification can take place, a means to measure and describe it in physical terms needs to be devised.

Towards this end, a review of musical timbre is presented which includes research into musical timbre from the work of Helmholtz through to the present. It also includes related work in speech recognition and musical instrument synthesis. The representation of timbre used in this study is influenced by the work of Hourdin and Charbonneau who used an adaption of multi-dimensional scaling, based on frequency analysis over time, to represent the evolution of each musical tone. A trajectory path, a plot of frequencies over time for each tone, was used to represent the evolution of each tone. This is achieved by taking a sequence of samples from the initial waveform and applying the discrete Fourier transform (DFT) or the constant Q transform (CQT) to achieve a frequency analysis of each data window.

The classification technique used, is based on statistical distance methods. Two sets of data were recorded for each of the guitars and violins in the study across the pitch range of each instrument type. In the classification trials, one set of data was used as reference tones, and the other set, as test tones. To measure the similarity of timbre for a pair of tones, the closeness of the two trajectory paths was measured. This was achieved by summing the squared distances between corresponding points along the trajectory paths. With four guitars, a 97% correct classification rate was achieved for tones of the same pitch (fundamental frequency), and for five violins, a 94% correct classification rate was achieved for tones of the same pitch.

The robustness of the classification system was tested by comparing a smaller portion of the whole tone, by comparing tones of differing pitch, and a number of other variations. It was found that classification of both guitars and violins was highly sensitive to pitch. The

classification rate fell away markedly when tones of different pitch were compared. Further investigation was done to examine the timbre of each instrument across the range of the instrument. This confirmed that the timbres of the guitar and violin are highly frequency dependent and suggested the presence of formants that is, certain fixed frequencies that are boosted when the tone contains harmonics at or near those frequencies.

Glossary

ASA: American standards association

FT: Fourier transform

DFT: discrete Fourier transform

FFT: fast Fourier transform

CQT: constant Q transform

MDS: multi-dimensional scaling

PCA: principal component anaylsis

PC: principal component

LDA: linear discriminant analysis

LD: linear discriminant

FAC: factorial analysis of correspondence

FIR: finite impulse response filter

sc: spectral centroid

f_o: fundamental frequency or first harmonic

Contents

Chapter 1

Introduction

1.1 Introduction to Musical Instrument Recognition

The problem of musical instrument recognition by computer has not been widely studied until recent years. Similarly, the use of computers in the study of musical timbre is a relatively recent phenomena. However, there is a considerable volume of work on the perception of musical timbre by humans dating back to the significant work of Helmholtz (1863) in the nineteenth century. Since the advent of high-powered computers, there has been considerable work done in the related areas of speech recognition and speaker recognition. The interest in this area is not surprising considering the obvious commercial applications of this work. In recent years, researchers in the instrument recognition area have increasingly borrowed from work in the speech recognition area (Kaminskyj & Materka 1995) and also from work on synthesis of musical instrument sounds.

In all of the work on instrument recognition to date, the main focus has been on the task of distinguishing between instruments of different types. There has been an implicit assumption that the timbre or sound quality of instruments of the one type is the same or similar. Our goal in this investigation is to show that, within a set of instruments of one type, there is considerable variation in timbre and that is possible to differentiate between and classify instruments within a class. In order to achieve this goal, a two stage classification system has been developed, which firstly represents the timbre of instrument tones in terms of time and frequency and then measures the closeness of timbre as defined by this representation for a pair of tones. We also attempt to highlight the physical features

that enable classification of instrument tones.

1.2 Sound Source Recognition Systems

Computer recognition of musical instruments can be considered to be a subset of the broader field of sound source identification. All sound source identification systems use some kind of pre-processing to highlight the physical features which best describe the sounds and enable discrimination between sounds of each type. These features are the input to the classification process which is generally statistically based or, alternatively, involves the use of neural networks.

Looking beyond musical instrument recognition, there are numerous examples of sound source identification systems that recognise sounds in a particular domain. Most of these systems are fairly crude since they are only required to discriminate between a small number of classes, each with markedly different sound properties. A system was developed (Nooralahiyan, Kirby & McKeown 1998) to recognise different types of motor vehicles from the engine and road noise as they passed a point on a highway. The system used a linear prediction algorithm feeding to a neural network. The vehicles were classified into one of four classes (car, van, truck or motor cycle). The system performed at an 84% success rate on the test data. The key discriminating features were not specifically specified.

Several systems have been constructed to discriminate between speech and music (for example, Scheirer & Slaney (1997)). The system developed by Scheirer & Slaney comprised of fifteen-second segments recorded from an FM radio station. Thirteen features were considered with four different classifiers. The best of these classifiers gave a correct classification rate of 94.2%.

A system that allows for a larger number of categories (Wold, Brum, Keislar & Wheaton 1996) such as laughter, female speech and telephone touch-tones was developed. A Gaussian model for each sound class was built based on the correlation between features such as loudness, pitch, brightness, bandwidth, and harmonicity. The Mahalanobis distance was used to find the best match between a sound sample and an item in a data-base of sounds.

In the animal domain, a system was developed to identify species of bat present in a certain habitat by classifying the bat echolocation calls (Jolly 1997). The data consisted

of echolocation calls for four species of bat in a given region. Features were extracted by first fitting a timefrequency curve to each observation. Statistical methods (Fisher's discriminant analysis) were used to perform the classification. The test data indicated an almost zero error rate with two species of bat but a significant error with the other two species (15% and 25%).

A common thread can be observed in the above source recognition systems. They mostly consist of a small number of classes with large differences in acoustical properties between the classes. As a consequence, only low level feature extraction is required for successful classification. Furthermore since, in each system, the features are specifically chosen to suit the particular problem, they may not be significant when applied to musical instrument recognition. A further point of difference is that, in general, these features are not time dependent.

The question of speaker identification has received some attention in recent years. In an example of this work (Reynolds 1995), mel-frequency cepstral coefficients (MFCC) were used as input to a statistical classification process. The system produced a 98% classification rate for 10 different speakers degrading to 83% for 113 speakers. The system was tested in a noise free environment suggesting it may not handle mixtures of sounds.

A innovative approach was tried in a system developed to recognise environmental sounds (Klassner 1996). A system based on the artificial intelligence methods used in visual scene analysis was used to match test sounds with 40 environmental sounds stored in a data bank. It is credited with a 59% correct classification rate.

1.3 Musical Instrument Recognition

In recent years there have been numerous attempts to build systems to recognise musical instruments. Most have operated on single isolated tones. Some have operated on musical phrases from monophonic(single instrument) real world recordings. The desired outcome of distinguishing between instruments in a natural real world multi-instrument context is at this stage too challenging.

In pioneering work, De Poli and his colleagues used self-organising neural networks in several different attempts to classify timbres from musical instruments (Cosi, Poli & Lauzzana

(1994), Poli & Tonella (1993), Poli & Prandoni (1997)). In each case they used single isolated tones, one from each instrument, each tone of the same frequency. Different methods of pre-processing were used: firstly, using features similar to those used by Grey (1977) who began with subjective judgements about the timbres of musical instruments and mapped these into three dimensional space (Poli & Tonella 1993); and secondly, borrowing from work done in speech recognition where auditory modelling techniques are used. In each case the set of recordings consisted of one tone from each instrument at each pitch.

There have been a number of other systems built using neural networks, all of similar scope, using isolated tones and one example of each. For example, Kaminskyj & Materka (1995), Feiten & Gunzel (1994), and Langmead (1995). In a typical example, Kaminskyj & Materka (1995) compared the classification abilities of a neural net system with that of the nearest neighbour technique. They achieved good results for the classification of four instruments (guitar, piano, marimba and accordian). The instruments chosen for classification all have very different acoustical properties. More recently, Kaminskyj & Czaszejko (2005) have again used neural networks, with a more sophisticated system, and achieved much improved results. The system used six features: cepstral coefficients, constant-Q transform, frequency spectrum, multi-dimensional scaling analysis trajectories, RMS amplitude envelopes, spectral centroid and vibrato. The system achieved a 93% correct recognition rate.

In more recent times, Kostek (2001,2002,2003) has investigated the use of neural networks, together with wavelet analysis, for instrument recognition. Wavelets were used for the parameterization process and Neural Networks, for the classification. Kostek also investigated which of the parameters in the new MPEG-7 standards are useful in defining musical timbre. Another investigator to use Wavelet analysis as a basis for the parameterization of musical sounds is Wieczorkowska (2001).

The early work in instrument recognition used statistical pattern-recognition techniques. For example, Bourne (1972) used spectral features as input to a Bayesian classifier. The training data consisted of 60 isolated tones from three instruments (clarinet, French horn and trumpet). The test data consisted of 15 tones (8 not used in training). In recent times traditional statistical pattern recognition techniques have received renewed attention. Fujinaga (1998) used spectral properties from 23 instruments as input to a nearest neighbour classifier to achieve a successful classification rate of about 50%. A system which attempts to model the human auditory system in the feature extraction stage was built by Martin

& Kim (1999). Temporal and spectral features were extracted using a log-lag 'correlogram' and fed to a Gaussian classifier employing Fisher's multiple discriminant analysis. The data set included 1023 isolated tones taken from 15 orchestral instruments over a full range of frequencies. In a later development of this system, Martin (1999) used a taxonomic hierarchy approach. Martin builds on the work of Bregman (1990) and Ellis (1996) who developed a perceptual representation of sound where the physical model reflects the human auditory system. Hierarchical approaches, using decision trees, have also been used by Wieczorkowska (1999) and Kaminskyj (2001).

In recent years statistical classification methods have received renewed attention, in particular, the Gaussian Mixture Model(GMM) first used in this domain by Brown (1999) who borrowed the technique from work in the speech domain. For example, Marques & Moreno (1999) compared the use of the GMM with support vector machines as classifiers. In the same work they compared the efficiency of cepstral features and linear prediction based features. They concluded that the choice of features was more important than the choice of classification method. More recently, Eggink & Brown (2003) used the GMM for instrument identification using spectral features together with a 'missing feature' approach which enabled them to classify two instruments playing concurrently. Essid, Richard & David (2005) used the GMM using data in a 'one-on-one' fashion. The features used were: auto-correlation coefficients, AM features,mel-cepstral coefficients, MPEG7 spectral flatness and constant Q transform together with "octave band signal intensities".

A common feature of most of the above systems is that they operated on isolated tones obtained from only one instrument of each type. Many used sample recordings produced by McGill University (Opolko & Wapnick 1987)) for use in music synthesis. A limitation of this data set is that it contains only one recording of each instrument at each frequency since this is all that is required for the purposes of music synthesis. This means that when training data and test data are selected then, there are no possible matches between tones played on the same instrument at the same frequency. Use of this data exclusively, does not aid the development of a system that will work successfully on real world data, that is, recordings of real musical performances where tones are run together in musical phrases.

It is only in more recent times that attempts have been made to build a system that can classify an instrument on the basis of a portion of a musical phrase rather than from an isolated tone. For example, Marques (1999) used mel-frequency cepstral coefficients as features feeding to a 9-way classifier. The correct classification rate was 72% on commercial

recordings dropping to 45% on non-professional recordings. Brown (1999) constructed a system which borrowed heavily from the techniques commonly used in speech recognition. This was motivated by a choice to attempt instrument identification from musical phrases rather than isolated tones and thereby increasing the similarity with the challenges of speech recognition. The system was tested as a two-way classifier (oboe and saxophone) giving a 94% correct classification rate with samples collected from commercial recordings and later as a four-way classifier (oboe, saxophone, flute & clarinet) giving an 84% correct classification rate. In further work, Brown, Houix & McAdams (2001) examined which features are most effective in instrument recognition. As before, techniques from the speech domain were used and musical phrases rather than isolated notes were used. The most effective features in the identification process were found to be cepstral coefficients, bin-to-bin differences of the constant-Q coefficients and autocorrelation coefficients (correct ID of between 79%-84%). After a comparison with human performance, they concluded that computers do as well as humans in identifying woodwind instruments. Other to have attempted classification based on musical phrases are Martin (1999), Eggink & Brown (2003), and Essid et al. (2005).

There has been some initial attempts at the difficult problem of separating/classifying instruments in recordings of polyphonic(multi-instrument) music. An interesting albeit unsuccessful project was carried out by Kostek & Wieczorkowska (1997) who attempted to separate the sounds of two musical instruments. Another attempt to classify instruments in a polyphonic (multi-instrument) context was made by Kashino & Murase (1999) using a template based time domain approach. This points towards projects of the future, which may be concerned with separation of parts from an improvised musical performance, instrument identification and then the transcription of different musical parts.

Finally, in her book, 'Perception-Based Data Processing in Acoustics', Kostek (2005) offers a comprehensive overview of the techniques available for feature extraction and classification in musical instrument classification.

1.4 Other Related Work

Other work of great relevance to this study is the large volume of work in the area of musical instrument timbre research and the related area of speech recognition which will occupy

our attention in the next chapter. This work from the time of Helmhotz, has been primarily concerned with how humans perceive and distinguish between the sounds of musical instruments. Mapping this work into the realm of machine recognition presents problems as there is not a one-to-one correspondence between the perceptual qualities of sound associated with human hearing and the physical quantities by which sound can objectively be described. For example, human responses are measured in subjective quantities such as pitch and loudness whereas the physical features are objective quantities such as frequency and power. The word 'timbre' itself is a perceptual quality with a somewhat unsatisfactory definition which we will discuss further in the next chapter.

Since the advent of high powered computers there has been renewed interest in the study of musical timbre for the purpose of musical instrument synthesis which can be thought of as the inverse process of instrument recognition. Particular instruments are analysed to determine the salient features that enable human recognition. These physical features are used to construct a synthetic version of the instrument. The likeness to the real instrument is then measured by a human listening panel.

1.5 Contribution of this Thesis to the Field of Computer Instrument Recognition

Within the body of work on computer recognition of musical instruments, there have been no attempts made to date, to compare the timbre of different instruments of the same type. Further, there have been no attempts to use a computer to distinguish between instruments of the same type. In contrast, in the field of timbre perception by humans, there have been a number of projects focusing on human assessment of instrument quality. For example, Caldersmith (1988) used a panel of human experts to make a quality assessment of four violins (three built by old 'masters' and the other, a new violin) in a 'blind' listening test.

In this thesis we use a frequency-time based trajectory path to quantitatively represent the timbre of musical tones. We develop a method to measure the closeness of the timbre for two instruments of the same type and use this distance measure to classify particular instrument tones from a set of four guitars and five violins.

We also attempt to determine which features, related to timbre, enable the classification

of instruments of the same type. We attempt to relate knowledge gained in instrument analysis and design to the classification process. This entails an in-depth analysis of the spectrum in the attack, steady-state or decay in order to highlight the distinguishing features of each instrument. A timbre analysis of this depth has not been attempted in previous work on computer recognition of musical instruments.

In the preliminary work for this thesis, we present a wide ranging review of timbre not previously seen in other work on instrument recognition. This review looks at timbre research over 150 years and encompasses work on human perception of musical sounds, speech recognition, synthesis of musical instruments, analysis of tone for musical instruments, as well as work directly related to computer recognition of musical instruments.

1.6 Planning an Instrument Recognition System

In planning a musical instrument recognition system, the first step is to understand the nature and complexities of a musical tone in order that the timbre can be represented in physical terms and in a form that is conducive to classification.

We begin at the the point where a musical tone can be described as a quasi-periodic signal with an approximately fixed pitch (Backus 1970). It is comprised of a complex set of mainly harmonic frequencies that vary with respect to frequency and intensity over time. We can divide musical tones into two categories: firstly, impulsive tones such as the guitar and piano which have a short attack period where the tone quickly evolves, followed by a lengthy decay period where the tone slowly attenuates; secondly, steady state tones such as the violin and clarinet where a short attack leads to an approximately steady-state period which terminates with a fast decay.

In determining how to best represent tones in a quantitative way, we can draw on the physical quantities used in previous work on instrument identification. This work has often been informed by research in the area of speech recognition (Brown 1999). However it is informative to look more widely to incorporate the ideas used in synthesis of musical instrument sounds and the large body of work on human perception of musical sounds (Bregman 1990, Ellis 1996).

The processing of highlighting salient features may be performed in the time domain (temporal), frequency domain (spectral) or a combination of both. Both domains are important in extracting the features since the sound of a musical tone can be described as a set of frequencies sounding together but changing in amplitude and pitch over time.

The information gained from temporal analysis, spectral analysis or combined time-frequency analysis can be used directly or can serve as a basis for extracting perceptually based features. These features may be derived from human perception such as vibrato, length of attack or 'brightness' of the sound. Data collected on instrument design and construction (Bachem (1955), Wolfe, Smith, Breilbeck & Stocker (1995) and Fletcher & Rossing (1998)) can provide useful information on the sound characteristics of particular instruments. Research into human perception of musical timbre (McAdams (1993), Hajda, Kendall, Carterette & Harshberger (1997)) provides useful information about the features humans use to discriminate between musical instruments. Work in the related area of music synthesis provides further information to assist with determining the features which characterise a particular instrument (for example, Beauchamp (1982), Grey (1977)). Some particularly interesting work in this area was done by Chowning (1973) who used frequency modulation to synthesise the sounds of various musical instruments. His work offers an alternative method for determining features which characterise an instrument.

Given that we have found a satisfactory way to represent the salient features of a musical tone, we then need to find a means of comparing tones in order to assign them to a class. The second stage of the system is the classifier which accepts the characteristic features of an unknown instrument and determines which class of instrument it is from. In most recent systems, sophisticated statistical classifiers have been used. Examples of some methods that have been used are Fisher's linear discriminant analysis (Martin 1999), nearest-neighbour technique (Kaminskyj 1999) and Gaussian mixture model (Brown 1999). These methodologies will be discussed in detail in chapter 3.

Finally, we need to assess the performance of the classifier and also try to determine the key discriminating features that enable classification. In chapter 5, we will investigate in some detail the tone production process and its effect on the characteristic features of the timbre for the instruments studied in this thesis.

Chapter 2

Features Defining Timbre

2.1 Introduction

It is generally agreed that the manner in which a sound is perceived by the human ear is dependent on the psycho-acoustical attributes pitch, loudness, duration and timbre. Pitch, loudness and duration are one-dimensional quantities that can be defined with a satisfactory degree of precision and correlate well respectively with the physical quantities frequency, power and time. On the other hand, timbre is yet to be satisfactorily defined and the nearest corresponding physical description is that of *wave form* (Seashore 1938). It is generally agreed that timbre is multi-dimensional (Plomp 1970). The problem of determining which physical features correlate with timbre is one that has interested investigators for at least one hundred and fifty years.

The Webster dictionary defines *timbre* as 'the characteristic quality of sound that distinguishes one voice or musical instrument from another or one vowel sound from another'. This definition is unsatisfactory in that it fails to take into account differences in pitch and loudness and any variations over time. Furthermore, there is an implicit suggestion that timbre is restricted to tones of a harmonic nature. A better definition is offered by the American Standards Association (*ASA:Psychoacoustical Terminology* 1973)- 'timbre is that attribute of auditory sensation in terms of which a listener can judge that two steady state complex tones having the same loudness and pitch are dissimilar'. Both definitions are uninformative in that they describe timbre in terms of functionality rather than specifying any properties associated with it. The question of relevance to this thesis is, what

is it that enables us to distinguish one musical tone from another and are these factors consistent with either of the above definitions of timbre?

In early research beginning with Helmholtz (1863) in the nineteenth century, it was thought that the timbre of a musical tone depended only on the spectral properties of the steady state portion of a tone. So, at this time, timbre was considered to be synonymous with the tonal quality of an instrument. Investigations from 1960 to the present (for example Clark, Luce, Abrams & Schlossberg (1963)) have established that the attack period of a musical tone often plays a significant role in differentiating between musical timbres. Given this fact, if we are to accept the definition of the American standards association as still being relevant, then we need to adjust our understanding of timbre to include the attack transient in addition to the steady state. There is still a general consensus that the decay section of a tone only plays a minor role in timbre differentiation (Clark et al. 1963). If we still consider the tonal quality of a voice or instrument to depend on the spectral properties of its steady state then, as a consequence of modifying our description of timbre, timbre and tonal quality are no longer synonymous.

Another important issue in understanding timbre is that it is dependent on pitch and loudness, a fact implicitly alluded to in the ASA definition and verified by numerous researchers (for example, Fletcher (1934)). These dependencies indicate the complexity of timbre and the consequent difficulty in differentiating between tones of differing loudness and pitch.

2.2 A Review of Timbre Research: 1860 to 1960

2.2.1 Early Years

Until the 1960's, it was a commonly held belief that the timbre of a sound was solely dependent on the harmonic structure of the steady state portion of that sound. This idea can be traced back to the literature in the first half of the nineteenth century (for example Willis (1830), Bindseil (1839), Seebeck (1849) as cited in Plomp (1970)). However, the first researcher to support the theory that timbre is related to harmonic structure with an experimental base was Helmholtz (1863) who, in his ground breaking work 'On the Sensations of Tone', dedicated the first six chapters to a treatment of timbre.

Helmholz showed experimentally that musical tones consist of a series of harmonics and that the human ear is able to distinguish a number of these harmonics individually. His work confirmed the theory proposed by Ohm that the human ear breaks down a complex tone into a number of component simple tones in the same way as the Fourier theorem mathematically decomposes a complex signal into component sine waves. Helmhotz explained that although the human ear does not consciously decompose a complex tone into its component parts, it does so unconsciously in the process of distinguishing between tones of different quality. He concluded that the timbre of a musical tone was dependent on its harmonic structure - that is, the harmonics present in the steady state of a musical tone and the corresponding amplitudes. These findings have become known as 'The Harmonic Structure Theory'. By relating the physical and perceptual differences in tones, Helmholz postulated the following set of rules for timbre (Helmholtz 1863):

1 Simple tones sound sweet and pleasant, without roughness, but dull at low frequencies;

2 Complex tones with moderately loud lower harmonics (up to 6th) sound more musical and rich than simple tones and sound sweet and pleasant if higher harmonics are not present;

3 Complex tones with strong harmonics beyond the 6th or 7th sound sharp, rough and penetrating;

4 Complex tones consisting of only odd harmonics sound hollow and if many harmonics are present sound nasal;

5 Predomination of the fundamental gives a full tone whereas a weak fundamental gives an empty sounding tone.

After establishing the dependence of timbre on harmonic structure, Helmholtz then investigated the relationship between timbre and the phase of component harmonics. He concluded that the relative phases of partials were not a factor in determining the timbre of a musical tone but added several qualifications. Firstly, that initially there was a slight change in timbre with change in phase of the harmonics but this quickly disappeared with time. Secondly, he conceded that, since the presence of harmonics beyond the 6th to 8th cause dissonance and roughness, there may be a phase effect for these harmonics. After this significant contribution by Helmhotz, there were few major advances in the understanding of timbre for about 100 years and most contributions were related to work in speech perception.

2.2.2 Timbre Research and Speech Perception

In research into vowel sounds, Willis (1830) showed experimentally that the timbre of each vowel sound corresponded to peaks in the harmonic structure at certain frequencies. He achieved this by constructing sound apparatus that could simulate the vowel sounds. His view was confirmed by the work of Helmholtz (1863). This relationship between vowel sounds and peaks in the harmonic structure was further investigated by Herman who called these peaks 'formants'. There were attempts to represent the timbral differences in vowel sounds by a two dimensional plot of the first two formants F1 and F2. Although useful, it was clear that two dimensions were not enough to adequately represent the timbral differences in vowel sounds and that at least three would be needed for a satisfactory representation (Pols, van de Kamp & Plomp 1969).

The observation that different vowel sounds corresponded to certain peaks (formants) in the frequency spectrum raised the question as to what happens to the peaks if the fundamental frequency of the vowel is raised. Most researchers (for example, Helmholtz (1863), Donders (1864), Grassman (1877) as cited in Plomp (1970)) were of the opinion that the formants remained constant thereby changing the shape of the spectrum ('The Fixed Pitch Theory'). The contrary view ('The Relative Pitch Theory') lost credibility with the invention of the phonograph which enabled researchers to demonstrate that vowel quality

was severely affected by increasing all frequencies by the same factor. However, Stumpf (1926) demonstrated that the relative pitch theory could not be totally dismissed. He showed that, if the fundamental frequency of a vowel was raised by a large amount then, the timbral quality was kept as similar as possible by raising the frequency of the formants by a small amount in the same direction.

This view was later confirmed by Slawson (1967) who found that when the fundamental frequency of a tone was raised by an octave the difference in the vowel quality was minimized if the lower two formants were raised by about 10% of the change in the fundamental frequency. The experiment indicated that vowel sounds were extremely sensitive to frequency shifts in the lower two formants. He found that shifts in the frequencies of higher formants had only a marginal effect on vowel quality. It should be noted that the average difference in pitch between the male and female voice is about an octave and there is a difference in the formants of about 10%, figures consistent with the findings of Slawson. In his experiments he used synthesized tones with the fundamental frequency decreasing in the latter part of the tone (inflection) to simulate actual speech as accurately as possible.

In an extension to the experiment designed to test the fixed pitch and relative pitch theories of timbre, a direct comparison was made between vowel sounds and musical tones. The musical tones where simulated with tones of fixed fundamental frequency (uninflected). In the experiment the fundamental frequency of all tones was raised by a factor of 1.5 (less than an octave). All formants in the spectral envelope were adjusted by a constant factor chosen to minimize the change in timbre (or vowel quality). It was found that the differences for both musical timbre and vowel quality were minimum when the formants were left unchanged (factor of 1.0). This result offers strong support for the fixed pitch theory; that is, the existence of a spectral envelope fixed on the frequency continuum.

Another issue raised as a consequence of research into speech recognition, was a discussion on whether simple tones also possess the attribute of timbre. In his classification of musical tones, Helmholtz (1863) implicitly recognised the frequency dependent timbre of simple tones. He noted that simple tones sounded dull at low frequencies and bright at high frequencies. Engel (1886) and Stumpf (1890) asserted that if complex tones have the attribute of timbre then it followed that so must simple tones have timbre. This view was supported by the findings of Grassmann (1877) and von Wesendonk (1909) (as cited in Plomp (1970)) that there was a resemblance between each of the vowel sounds and simple tones at specific frequencies. This can be explained by the correspondence of the single

tone with the frequency of the most significant formant for each of the vowel sounds.

Further support for the idea that the timbre of a simple tone is related to pitch is implicit in the theory proposed by Bachem (1955) to explain an aspect of absolute pitch (the ability of some musicians to name the pitch of a tone without an external reference). In order to explain the fact that in the execution of this skill some musicians occasionally name the correct note but the wrong octave, Bachem suggested that all notes of the same pitch name (eg. C3, C4 etc.) have the same *chroma*; that is a characteristic timbre. This implies some periodicity in the timbre of a tone so that the timbre of a simple tone is repeated at intervals of an octave. This finding was supported by a number of other researchers. For example, in a later work Shepard (1982) showed that the structure of pitch can be represented geometrically as a helix which reflects the partially periodic nature of pitch.

2.2.3 Timbre Research: Post-Helmholtz to 1960

Dating back to the earliest investigations, researchers have attempted to map the timbre of sound onto some kind of scale. Since timbre is a multidimensional quality it does not easily lend itself to this kind of representation. In an early attempt to represent timbre in this way, Stumpf (1890) attempted to describe timbre in terms of 20 semantic categories such as wide-narrow, smooth-rough, and round-sharp.

In a report on early work in this area from a research career spanning more than 40 years, Fletcher (1934) examined the correspondence between the perceptual characteristics of sound, namely: loudness, pitch and timbre, and their physical equivalents: intensity, frequency and the harmonic structure. Up to this time it was accepted that there was a one to one correspondence between the physical quantities and their perceptual equivalents. Fletcher showed that this was clearly not so. He found that pitch depends not only on the fundamental frequency but also on the harmonic structure and the intensity of the tone. He found that loudness depends not only on intensity but also on the fundamental frequency and the harmonic structure. And of particular importance to this thesis, was his finding that timbre depended not only on the harmonic structure but also on the fundamental frequency and the intensity of the tone. These were findings that would be re-assessed by a number of researchers in the 1960's (for example Clark, Robertson & Luce (1964)). As the title of his paper would suggest, Fletcher's work at this time is founded on an acceptance of the Helmholtz Harmonic Structure theory.

In a subsequent work on the relationship between the harmonic structure of a tone and its timbre, Lob(1941) investigated the effect of varying the intensity of one of eight harmonics in a complex tone. After equalising complex tones on the basis of pitch and loudness, Lichte (1941) compared complex tones in order to determine attributes which facilitated discrimination between tones; in other words to find attributes that defined the timbre of a tone. He suggested three attributes that assist in specifying timbre: *brightness* which is determined by the location of the mean of the energy distribution on the frequency continuum; *fullness*- depending on the ratio of odd to even numbered partials; and *roughness*- to describe a tone with consecutive high partials above the 6th . The author explained roughness in terms of a harshness in the tone caused by musically dissonant intervals between upper harmonics (This explanation could be relevant to the 'roughness' Helmholtz attributed to phase differences in upper partials). The fact that harshness can be moderated by removing some upper harmonics, and thereby altering the brightness and possibly the fullness, indicates some correlation between the three attributes. It could be that roughness and fullness are functions of one variable: the complexity of frequency ratios between the partials.

A precursor to later research in the 1970s is the work by Richardson (1954), in which the results of an investigation of transient tones of wind instruments suggest that the attack transient may be an important factor in timbre discrimination within and between musical instrument classes.

2.3 Research on timbre: 1960 onwards

2.3.1 1960-70: Understanding timbre through analysis, synthesis and evaluation

In the period up to 1960, the prevailing opinion was still that timbre was determined by the steady state portion of a tone. With the advent of more powerful computers in the 1960's and a subsequent interest in computer synthesis of musical sounds, there was an increased interest in the investigation of timbre. Clark et al. (1964) in the first of a series of investigations confirmed the existence of perceptual families of instruments. That is, instruments of a particular class tend to sound perceptually similar. They also found that, in most instrument families, instruments can be grouped into sub-families on the basis of their perceptual similarity. For example, in the strings: violin/viola and cello/double-bass ; in the brass: trumpet/trombone ; in the double-reeds: oboe/English-horn. The investigations were carried out by submitting real and synthesized tones to a listening group for identification. One doubt on the reliability of the findings is raised by the fact that some tones were raised in pitch by a factor of 2 or 4 with no indication that the spectral envelopes were adjusted appropriately. In an earlier and less sophisticated investigation of human instrument identification using live tones and tones heard through a PA system, Eagleson (1947) showed that confusion generally occurred within instrument families. However some confusion was found across instrument families, generally between instruments with similar pitch ranges. The confusion occurred more often with the degraded quality of the PA system. Some examples were: the violin incorrectly identified as a clarinet or flute; the alto saxophone incorrectly classified as a trombone on numerous occasions.

Most investigations in this period fell into one of two categories. Firstly, those that were directly concerned with the investigation of timbre often used a technique in which tones were altered in a variety of ways and then offered, interspersed with natural tones, to a listening panel for identification. By comparing identification rates for the natural and altered tones, it was possible to draw conclusions about the salient features of timbre. Secondly, investigations concerned with accurately synthesizing a particular instrument began with analysis of the natural tones in order to determine salient features, re-synthesis of the tones based upon the analysis, and an assessment of the closeness of the natural and re-synthesized tones based on trials with a listening panel. If a method of synthesis

is shown to produce a good likeness to the instrument it is modelling, then it can be deduced that the parameters used should be useful in instrument identification. The goal was to characterise the tones of musical instruments by as few parameters and principles as possible. The assumption is made that the auditory system processes only certain gross features of a musical tone and that, for reasons of economy, some of the detail in a tone can be omitted. The task of the researcher was to determine which features could be omitted without compromising the likeness to the original tone. A limitation with this type of investigation was that features found to be salient in describing the timbre of one instrument type may not necessarily be generalised to the timbre of other instruments. It should be noted that in all investigations of this period, where the tones were first recorded on magnetic tape, the phase response and hence the wave form of the tones is changed in some way. It was argued that this change does not, in general, affect human perception of the tones.

In the area of timbre research there was now some doubt about the validity of Helmhotz' 'Harmonic Structure Theory'. Researchers were beginning to test the proposition that timbre may depend on more than the average spectrum in the steady state. A particular area of interest at this time was the contribution of the attack transients to timbre. In an investigation into the relative importance of different parts of a tone, Clark et al. (1963) recorded tones from the full set of orchestral instruments , divided the tones up into segments and offered these to a listening panel for identification. The segments were as follows: full tone; attack transient; short steady-state only; long steady-state only. Using whole tone recognition as a reference, the results showed satisfactory recognition rates for the attack transient and the long steady-state segments but with some increase in between family confusion. For the short steady-state segments, the recognition rates were considerably impaired with a marked increase in within and between family confusion. In a follow up experiment, the attack transient was truncated and segments of different duration were offered to the listening panel for recognition. It was found that the attack segments were very resistant to shortening and identification rates were not noticeably reduced. In the final part of this experiment, segments from the decay transient of the instrument tones were offered to the panel for identification. Very poor recognition rates were obtained for all instruments. From these experiments it was concluded that both the attack transient and the steady state play important roles in timbre identification, each on its own providing sufficient information for identification, whereas the decay transient provided insufficient information for reliable identification.

In a subsequent investigation, Clark et al. (1964) set out to determine if the timbre of musical tones was dependent on loudness. Since it is known that the spectra of musical tones is dependent on intensity (the ratio of the amplitudes of high partials to low partials increases with intensity), it was anticipated that timbre would be dependent on loudness. The full set of non-percussive orchestra instruments were recorded at low, medium and high loudness levels (dynamic levels: pianissimo, mezzo-forte, and fortissimo), standardised for loudness and then offered to an expert listening panel for identification of dynamic levels. It was found that in general, the dynamic levels were very poorly identified. This unexpected result indicates that any variation in the spectra of the tones due to varying intensity levels was insufficient to significantly affect the timbre of the tones. Perhaps it is only in the decay period, where it is known that higher partials die away faster than the low partials (Saldana & Corso 1964), that timbre is significantly affected. This supposition is consistent with the results of Clark's previous investigation, outlined above (Clark et al. 1963), where the decay transient alone, gave very poor recognition rates.

Clark's co-worker Luce (1963), in further work in this area, suggested that timbre depends on the attack transient, modulation (in amplitude and frequency) in the steady state and on one or more formants. He suggested that, in all non-percussive instruments, the steady state spectra can be characterised by the formants. It was noted by the author that frequency was an important cue in identification, which is consistent with the assertion that timbre is frequency dependent.

Berger (1964) investigated the relative importance of attack, steady state spectrum and decay to the timbre of woodwind and brass instruments. Tones were altered to determine if recognition would be impaired by: removing the attack and decay; interchanging the attack and decay; and removing the partials to negate the effect of the steady state spectrum. He concluded that, although the attack served as an important cue in recognition, the steady state spectrum was a more significant cue. The lowest recognition rate was when all but the fundamental frequency was filtered out of the tone. It should be noted that this filtering process would also remove valuable information from the attack.

In an experiment with ten orchestral instruments, Saldana & Corso (1964) set out to investigate the relative importance of steady state spectrum, frequency, vibrato, transient motion (attack and decay), and steady state duration. By altering tones in certain ways and then submitting these tones together with the natural tones they were able to show that the initial transients contribute substantially in the identification process. They also noted

that the presence of vibrato in a tone was an aid to identification. Their analysis offered some possible explanation of how the initial transients might be important. It was noticed that the partials did not appear simultaneously but were staggered over time adhering strictly to the order lowest to highest. Further, it was observed that the onset times and temporal gradients were different for each instrument. The reverse was observed in the offset with partials disappearing one by one finally leaving just the fundamental. It was anticipated that the study would shed light on the relevance of the harmonic structure and formant theories. It is not clear to this writer why the authors expected their investigation to enlighten this debate and it is therefore not surprising that their findings on this aspect were equivocal. In their concluding remarks, the authors suggested that there was merit in the proposal by Seashore (1938) that tonal quality could best be described by two fundamental aspects: firstly *timbre* - based on spectrum at a certain time; and secondly *sonance* - the cumulative effect of changes in timbre, pitch and loudness. A further point in their discussion was that a further cue in identification could be the presence of inharmonic partials.

This suggestion was supported by the work of Fletcher and colleagues (Fletcher, Blackham & Stratton (1962), Fletcher (1964)) in their investigations of the quality of piano tones. Their project had the dual goal of developing an objective description of the quality of piano tones and at the same time develop synthetic piano tones equal in quality to those of a high quality instrument. Starting with the premise that the quality of piano tones depends on the steady state spectrum, pitch, loudness, decay and attack time, variation with time of partials as well as the mechanical noise of the tone production, analysis was carried out on sample tones and synthetic tones were produced. It was noted that partials below middle C (C4) were found to be inharmonic (a little sharp in pitch).

Fletcher and colleagues followed up their work on pianos with an examination of the tonal qualities of string instruments (Fletcher, Blackam & Geertsen 1965) and then with a more in-depth examination of the quality of violin vibrato tones (Fletcher & Sanders 1967). The same methodology as that in their work on pianos (analysis followed by synthesis) was used. It was found that the characteristics essential to describing the timbre of a string instrument were the harmonic structure in the steady state, the harmonic structure of the decay, the vibrato when present and the mechanical noise associated with the production of a tone. These findings contrast with other studies of the period in which the attack was rated as a significant cue and the decay of little significance (for example, Berger (1964);

Saldana & Corso (1964); Clark et al. (1964)). In modelling the attack for synthesis they considered the attack so insignificant that they assigned it the same harmonic structure as the steady state. The decay, which in a bowed string instrument takes the form of an after-ring following the removal of the bow, was found to be made up of the fundamental plus the first 3 or 4 partials. There were some interesting findings relating to the mechanical noise of tone production. The bow noise was analysed by drawing the bow across the bridge where it does not excite the strings and then modelled. It was noted that for tones below G4 (on the third string of a violin), the noise is barely audible and that for tones above G5 (first string on the violin), it is distinctly audible. This is due to the bow noise being masked by tones with a fundamental frequency of G5 or below. In a follow up investigation into the effect of vibrato on the quality of violin tones, four quantities were investigated and the following observations were made:

1 It was found that the pitch variation (but not frequency) was the same for all partials;

2 The intensity levels for the partials underwent periodic fluctuation but by significantly different amounts and with differing phase;

3 As a consequence of this phase difference, the harmonic structure was not constant with the relative amplitudes of the harmonics changing periodically at the vibrato rate;

4 There is a complex tone coming from open strings induced by a sympathetic vibration with the bowed string (Sympathetic Transient Tone). When vibrato is present, the STT varies in intensity at twice the rate of the vibrato.

Working at the same university as Fletcher and his co-researchers, Strong & Clark (1967*b*) carried out parallel investigations on wind instrument tones (brass and woodwind) with the goal of producing authentic sounding synthetic tones but with the most economic use of parameters and principles. They analysed tones and developed a mathematical model to represent the sound of each wind instrument. In the implementation of their model, partials for each instrument were controlled by a single spectral envelope, fixed for each instrument regardless of pitch frequency, and three temporal envelopes depending on the frequency of each particular partial. It was found that the spectral envelope was nearly independent of the loudness of the instrument tones. When tests with natural tones and the synthesised tones were conducted with a listening panel, the identification rate for natural tones was

85% and for the synthesised tones the rate was 66%. Given that a number of the confusions were intra-family, this indicated that the synthesised tones had a good likeness to the natural tones. In a follow up investigation, Strong & Clark (1967*a*) examined the relative importance of the temporal and spectral envelopes in the timbre of wind instruments. Tones were synthesised with the following changes: exchange of temporal and spectral envelopes; creation of artificial spectral envelopes; and by perturbation of the spectral envelopes. It was found that where the spectral envelope is unique, it predominates over the temporal envelope in aural significance but where the spectral envelope is not unique, then it is equal or subordinate to the temporal envelope in aural significance. Although the conclusions appear informative, it is not clear to this author what information might be gained by exchanging the temporal and spectral envelopes since changing two factors means it is difficult to attribute any change in timbre to one particular factor.

The method of analysis, synthesis and testing against natural tones was again used in an investigation by Freedman (1966,1968). His work was based on the premise that both the attack transient and the steady state played important parts in both the synthesis and the recognition of musical tones. The goal was to find a mathematical representation of a musical tone that used only a small number of parameters to specify it. The basis of the analysis was a mathematical model of the attack and the steady state based on Fourier's theorem but taking into account the time dependent nature of the frequencies. It included parameters such as the onset times of partials, attack rates and frequencies in the attack, total attack time and total decay time. The evaluation of parameters for each particular tone was achieved using a pair of integral transforms specifically developed for the attack and the steady state. The tones from violin, clarinet, bassoon, saxophone and trumpet were re-synthesised and submitted together with the natural tones to a listening panel for informal qualitative analysis. It was considered that the synthesised saxophone and trumpet tones were almost indistinguishable from the natural tones. This suggests that the techniques may be useful in instrument recognition. The informal nature of the evaluation makes any comparison with previous investigations of a similar type rather difficult.

Using similar methodology, Risset & Mathews (1969) explored the timbre of trumpet tones, concentrating on the time varying aspects. In their preliminary work they assessed the outcomes of previous investigations into timbre research by programming a synthesiser and

found that all descriptions tried were inadequate to generate synthetic tones with a satisfactory likeness to the natural tone. They pointed out that many past researchers, especially in the time of Helmholtz, had relatively primitive measuring instruments incapable of the resolution to produce more than spectral data averaged over time. Due to recent technological advances, the importance of temporal changes in timbre could now be demonstrated by computer analysis and synthesis. The authors seemed to be unaware of other relevant investigations previously cited in this thesis - for example, Berger (1964), Strong & Clark (1967*b*) & Strong & Clark (1967*a*). In their analysis they searched for formants, zeros in the spectrum, attack and decay rates and the variation of the spectrum with time. To take account of the time-variant nature of the tones, the authors used *pitchsynchronous analysis* to perform a running analysis of each tone. The technique is basically segmenting the wave into individual pitch periods and determining the spectrum for each by Fourier analysis. To provide maximum information about the attack and growth rate of partials, the processing was performed backwards in time. The authors synthesised the trumpet sound to their satisfaction from a set of parameters determined by their analysis. The parameters focused on the time variant nature of the tones. They determined a 'line segment function' for the growth and decay for each of 13 harmonics. In the steady state portion they determined the average spectrum plus the periodic and random fluctuations in pitch (vibrato and jitter). Study of a number of spectrographs showed that the spectrum followed a formant structure; that is, the spectrum varied with frequency following an approximately invariant spectral envelope. When synthesised tones were submitted to a listening test, auditors had difficulty in differentiating between the real and synthetic tones. However, the synthesis was rather complex so the authors attempted to develop a more efficient method. By varying parameters one by one, the importance of each to the timbre was assessed and tones were re-synthesised with a reduced number of parameters. The evaluation process indicated that the key parameters were: the spectrum based on the fixed spectral envelope (a formant at about 1500 Hz); the attack including total length, rate of growth for low order partials, rate of growth for high-order partials; and the high-rate quasi random frequency fluctuation. The simplified tones were submitted to a listening panel and gave results equivalent to the previous more complex tones. The authors suggested that the likeness to natural tones could be further improved by some attention to the decay, introduction of a function to control frequency, and random 'glides' to simulate intonation slips.

2.3.2 The relationship between phase and timbre

As mentioned earlier in this chapter, the relationship between phase and timbre was first investigated in the monumental work of Helmholtz (1863). He concluded that the relative phases of partials were not a factor in determining the timbre of a musical tone but added several qualifications. Firstly, that initially there was a slight change in timbre with change in phase of the harmonics but this quickly disappeared with time. Secondly, he conceded that, since the presence of harmonics beyond the 6th to 8th cause dissonance and roughness then, there may be a phase effect for these harmonics.

Although there was further investigation and discussion of the importance of phase to timbre (for example, Konig (1881)) in retrospect it can be said that the measuring instruments of the time were too primitive to allow any advances in experimentation and therefore understanding of the issue. Helmholtz' findings remained unquestioned for nearly one hundred years with little work being done on this aspect of timbre. Even the advent of electronic amplification and recording did little to advance thinking in this area. The considerable phase distortion in a musical signal caused by amplification and recording processes, together with the observation that there was no noticeable effect on the quality of reproduced speech and music, served to reinforce the prevailing view that phase effects on timbre can be regarded as negligible (Fletcher 1934).

At the time of Helmhotz the phenomenon of *beating* had already been observed. It could be induced with three simple tones of frequency $(n-1)p$, np, and $(n+1)p$ Hz (n an integer) with one tone slightly mistuned. It could be interpreted as a complex tone of three harmonics with the phase of one harmonic changing continuously relative to the other two. The beat sensation could be interpreted as a periodic change in timbre. A similar but less distinct beating was observed when **two** simple tones were tuned to the same frequency but with a small error. Further qualitatively based investigations of the phenomenon of beats by Mathes and Miller (1947), Zwicker (1952) and Goldtein (1967) (as cited in Plomp & Steeneken (1969)) provided evidence that the human ear was in fact sensitive to phase. In 1957, Lichlider developed a tone generator that could control both phase and amplitude for the first 16 harmonics. He noticed that a change in phase did produce a discernable change in timbre. He noted that changing the phase of the higher harmonics produced a stronger effect than changing the lower harmonics. The effect was most noticeable for tones of low fundamental frequency.

In the 1960's, when there was a renewed interest in timbre investigation, it was common knowledge that the phase of musical signals was perturbed by the recording process and that the phase of a musical tone was perturbed by reverberation in the room in which it sounded. However, it was believed that these phase effects had a negligible effect on the timbre of a musical tone. In a comprehensive study using Kruskal's multidimensional scaling techniques, Plomp & Steeneken (1969) quantified the effects of the phase pattern of harmonics upon timbre. He concluded that the phase pattern was a specific factor in timbre, independent of spectrum and loudness, but of lesser importance than the spectrum. Further, he found that the effects of phase pattern are more marked for lower fundamental frequencies and more audible for a continuously changing phase relation.

2.3.3 Post 1970: Time variant Features, Data Reduction, Speech Analysis

To examine the time variant characteristics of violin tones Beauchamp (1974) developed a mathematical model for the time variant spectra and applied Fourier analysis to a number of test tones to determine parameters. He found that the phase of partials in non-vibrato tones was locked together but for vibrato tones the phase of partials is shifted as their frequencies change. It was found that the amplitude (power) of partials correlated with the phase vibrato pattern. The growth and decay of partials in the attack and decay sections were modelled with linear equations. It was found that partials were inharmonic in both the attack and decay stages. The author suggested that the high level of some partials accompanied by phase fluctuations could be explained by their proximity to the frequencies of strong resonances in the instrument. These observations indicate the existence of formants in the frequency response of the violin.

In further work, Beauchamp (1982) used two time-variant parameters as a basis for synthesizing cornet and alto saxophone tones. A time variant spectrum analysis was performed on the digitalized signal by using sequential Fourier transforms with a moving window. The parameters used to synthesize the tones were the spectral centroid (brightness) and the amplitude. Since this synthesis system produced "quite good" results for the cornet and yet unsatisfactory results for the saxophone it would appear that the parameters of spectral centroid and amplitude may be helpful but not sufficient for instrument identification.

Although the efforts in synthesis in the late 1960's and early 1970's produced good results,

the analysis methods of hetrodyne filtering (Moorer 1973), phase vocoder (Moorer 1978) followed by re-synthesis (also known as *additive synthesis*) were inefficient in that they required large quantities of data from the analysis stage to produce good results in the re-synthesis. In this time numerous attempts were made to find ways of synthesizing a tone with only a fraction of the data but without a discernable difference in timbre. Non-linear synthesis techniques were being developed which gave satisfactory results in synthesis with only a fraction of the data. For example, Chowning (1973) developed his frequency modulation technique where the timbre was controlled by adjusting a simple modulation index.

Charbonneau (1981), following the lead from Chowning (1973), investigated three ways of achieving data reduction in synthesis without significant deterioration in timbre. Firstly, rather than using a time varying spectral envelope, he used the average spectral envelope normalized to a peak amplitude of one. Secondly, the frequency variation for all partials was based upon a function modelling the frequency variation in the fundamental frequency. Thirdly, start and finish times for the partials were replaced by two polynomial functions which depended on the harmonic number. Tones for sixteen orchestral instruments were synthesized in this way and offered to a listening panel for comparison with the original tones. The results indicated that no significant timbral information had been lost. Further, there seemed to be support for the findings of previous studies that the spectral envelope plays an important part in timbre recognition. Results also indicated that: of the three significant dimensions, amplitude was the most sensitive ; frequency variations are an important ingredient in timbre; and differences in start and end times for partials are not an important factor in timbre.

In another approach to data reduction in defining timbre, Pollard & Jansson (1982) describe what they call the tristimulus method where the data from the spectral envelope is compacted into three variables: the power from the fundamental frequency; the power from the second to fourth partials; and the power from the fifth to the nth partials where n is the highest significant partial. No objective assessment of its ability to adequately represent the timbre of a musical tone is attempted.

Sandell & Martens (1995) assert that the well used harmonic analysis and re-synthesis (additive synthesis) techniques dating back to the 1960's backed by modern technology and with effective data reduction techniques, offer more flexibility than the more modern sampling synthesizers currently in vogue. The large data set created by the progressive

harmonic analysis creates problems with the algorithms used for real time re-synthesis. They use principal component analysis as a means of reducing the data set with a minimal loss of information. They achieved almost identical re-synthesis (for the cello, trombone and clarinet) with a 40-70% reduction in data. The authors found that more principal components were required to achieve acceptable re-synthesis of violin and flute tones. This is due to the level of spectral flux (fluctuation and complexity) in these tones. They suggest that the number of principal components needed to represent 99% of the variance could serve as a measure of spectral flux. If principal component analysis can be used so successfully for analysis and re-synthesis then this suggests that analysis and data reduction by principal components could be a useful approach in classification.

In an attempt to assess the significance of the various parameters commonly used in additive synthesis, Kostek & Wieczorkowska (1997) applied statistical methods in order to investigate the degree of separation that could be achieved between instrument groups by the various parameters. In the steady-state, parameters considered were: normalized frequency; second and third tristimulus parameters (based on power when spectrum is divided into three equal bands); and sum of power from odd and even harmonics. Also a number of time dependent parameters from the attack related to the rise time and duration of particular harmonics were considered. The findings were somewhat equivocal but showed that steady-state parameters were not sufficient to differentiate between instrument types. In other words, recognition was optimized by the use of both the time variant attack and steady state parameters.

McAdams, Beauchamp & Meneguzzi (1999) investigated the importance of a number of temporal and spectral features by conducting spectral analysis of seven orchestral instruments and then re-synthesising them with six types of data simplification in various combinations. The instruments studied were: clarinet, flute, oboe, trumpet, violin, harpsichord, and marimba. The data simplification techniques were: amplitude variation smoothing; coherent variation of amplitudes over time; spectral envelope smoothing; forced harmonic frequency variation; frequency variation smoothing; and harmonic frequency flattening. Results indicated that the most salient features in timbre discrimination were: firstly, the shape of the spectral envelope (jagged vs smooth); and secondly the spectral flux (time variation of the normalised spectrum).

In this more recent period, understanding of timbre can again be informed by research in the speech area but also now by work in speaker recognition. Brown (1981) found

that certain parameters associated with the vowel sounds were most important in speaker recognition. These were: the mean fundamental frequency, formant mean, and formant band-width. These findings suggest that similar spectral features may be important in musical instrument recognition.

Poli & Prandoni (1997) borrowed from techniques used in the speech processing community in experiments which analyse timbre and algorithmically define timbre space. Their analysis involved the use of mel-frequency (log based) cepstral coefficients, a well used technique in speech analysis. Two approaches, using Kohenen neural networks and principal component analysis, were used for data reduction and production of a timbre space. Their methodology and findings will be be discussed in detail in the next section, however, the authors concluded that the steady state portion of a musical tone is the key determinant of tonal quality whereas the attack phase is important in instrument recognition.

Brown (1999), also influenced by work in the speech processing domain, used a similar timbre analysis method in her work in classifying wind instrument timbres. To highlight features, she used the constant Q transform to obtain log-frequency data and then calculated cepstral coefficients.

2.3.4 Multi-Dimensional Scaling Techniques

Dating back to the time of Helmhotz, researchers have been interested in representing timbre on some kind of multi-dimensional scale (for example, Stumpf(1890)). Pratt & Doak (1976) suggested a system where timbre could be described by subjective ratings on three scales: *dull-brilliant*, *cold-warm*, and *pure-rich*. This allowed a three dimensional graphical representation. von Bismarck (1974) investigated the use of 30 verbal scales to describe timbre. Using factor analysis, he showed that timbre could be almost completely defined with the four independent scales: dull-sharp, compact-scattered, full-empty, colorful-colorless. The most significant were dull-sharp (sharpness) and compact-scattered (compactness). Compactness was found to differentiate between complex harmonic sounds and noise whereas sharpness was found to correlate with the spectral centroid. The limitation of investigations of this type is that the categories to be investigated are chosen by trial and error rather than by any objective means.

In spite of the advances made in understanding timbre in the 1960's, there was an inherent problem in the research methodology used by most investigators in that there was no objective or systematic method of assessing the relative importance of the different components of timbre. They relied on the technique of varying components one by one and submitting a consequent synthesised sound to a listening test. In the 1970s the technique of multi-dimensional scaling, which had proved a useful analysis tool in other domains (for example, Shepard (1992a,1992b) and Kruskal (1964a,1964b)), was extensively applied to the analysis of sound - for example, Grey (1977). It provided researchers with a tool with which they could objectively evaluate the importance of particular features to the timbre of tones.

The purpose of multi-dimensional scaling is to look for hidden structure in a matrix of empirical data and to represent it as a geometric model. This can be applied to the study of timbre in two ways. Firstly, a data matrix is created storing dissimilarity ratings for each pair of tones in a set. This is based on subjective judgements by a panel of listeners. A triadic system is commonly used where they consider the tones in combinations of three and choose the two most similar tones and the two most dissimilar tones. Secondly, a scaling technique is then used to convert similarity and dissimilarity ratings between pairs of tones into distances. These distance measures are then used to create an n dimensional timbre space for a set of n tones. Each tone is represented as a point in n-dimensional space and the distances between points represent the dissimilarity between each pair of tones. A method of data reduction is employed where the aim is to explain most of the variation with just a few variables. This can be achieved by using principal component analysis (PCA) or factorial analysis of correspondence (FAC). The investigator attempts to interpret each dimension in the timbre space in terms of the physical attributes of the sound.

One of the first researchers to apply multi-dimensional scaling to the study of timbre was Plomp & Steeneken (1969). He used multi-dimensional scaling based on perceptual dissimilarity ratings to examine the relative effects of phase change and amplitude change on timbre. They concluded that the effect of varying the phase spectrum was small in comparison with the effect of varying the amplitude spectrum. Much subsequent use of the multi-dimensional scaling technique was made by Plomp and his colleagues in their work investigating the relationship between timbre and sound spectrum - this will be discussed later in this chapter.

In the first of a series of much cited investigations, Grey (1977) used multi-dimensional scaling to investigate the salient features controlling the timbre of 16 musical instrument tones. Beginning with a recorded set of 16 actual musical instrument tones, he analysed and re-synthesised the tones and then submitted the tones in combinations of two to a listening panel for a subjective assessment of their timbral similarity. The results were stored in the form of a similarity matrix for each listener and then processed used the multi-dimensional scaling algorithm. A representation in three dimensional space was found to be the most useful for interpreting the timbral space of the tones. A physical interpretation of each axes was made in terms of spectral energy and temporal features of the tones. The first dimension was interpreted in terms of the spectral energy distribution. On one extreme were tones with a narrow spectral bandwidth and a concentration of energy in the lower harmonics. On the other extreme were tones with a wide spectral bandwidth and a significant proportion of energy in the upper harmonics. The second dimension was interpreted as relating to the onset/offset patterns of partials within tones and, in particular, the synchronicity in attack and decay for upper harmonics. Related to this is the amount of spectral fluctuation in the steady state of tones. It was observed that along this axis there was a good separation of instruments along family lines. The exception to this was clustering together of the brass and the strings. The third dimension was interpreted as relating to the presence or absence of high frequency and most often inharmonic energy in the attack segment. A complete separation along family lines was gained by considering the second and third axes together. Certain exceptions to the clustering along family lines suggested that there are factors related to the attack that may override the tendency of instruments to cluster together. For example, the flute shares properties with the strings.

In a follow up investigation, Grey & Moorer (1977) used multi-dimensional scaling to investigate the perceptual differences between natural tones, complex synthesised tones and sythesised tones simplified in various ways. As in their previous investigation, they found a small perceptual difference between the original tones and the complex synthesised tones. Small but significant differences were found between the complex synthesised tones and tones synthesised with simplified time varying amplitude and frequency functions and tones synthesised with constant frequency. However, the largest perceptual difference was found between the tones synthesised without an attack and all other tones. These results suggest that the spectrum, fluctuations in the spectrum with time, fluctuations in fundamental frequency over time and the attack are all important in timbre and in timbre recognition. Of interest in this study was the finding that ‘precedent low-amplitude

inharmonicity' during the attack is an important feature of timbre.

In the third of this series of investigations, Grey & Gordon (1978) studied the effects of context on timbre discrimination. He used tones of three instruments in three different contexts: (1) isolated tones; (2) tones in a single voice melody; (3) tones in a multi-voice melody. The results were equivocal with discrimination for the clarinet and trumpet being best for isolated tones and discrimination for the bassoon being best in single voice melody passages. It can be concluded that for a set of instruments with marked spectral differences, discrimination may be best in a single voice melody and that for a set of instruments with temporal distinctions in attack or articulation, discrimination may be better for isolated tones. The difficulty in discriminating between tones in a single or multi-voice melody context may be explained by the masking effect on the attack of slightly overlapping tones. For multi-voice tones, where the resultant timbre is a combination of the timbres of the separate instruments, they came to the interesting conclusion that the timbre space location for the multi-voice tone was the vector sum of the timbre space position vectors for the individual component tones.

In the last of this series of investigations, Grey (1978) studied the effects of modification of the spectral energy distribution on timbre discrimination. They noted that previous studies such as Plomp (1970), Wessel (1973, 1974), Wedin & Goude (1972), Grey (1977a&b), Miller & Carterette (1975) had all uncovered at least one common aspect of tone in similarity structures for steady-state tones (attack removed), complete natural tones and synthetic tones. This common attribute is related to the spectral energy distribution (spectral centroid) of the tones. For steady state tones the spectral energy distribution is the sole attribute upon which judgements about perceptual relationships between tones can be made. For complete natural tones and synthetic tones, the spectral component is a single dimension in the perceptual similarity space. For complete natural tones, the interpretation of this axis is complex because of the number of factors influencing the energy distribution such as the bandwidth of the signal, the balance of energy in the low harmonics and the existence of upper formants. Grey and Gordons's investigation set out to test their interpretation of this axis by using the data from their previous investigations and modifying it in the following way. Half of the tones were grouped into pairs and the spectral envelopes exchanged. The complete set of modified and unmodified tones were then subjected to a similarity rating assessment by a listening panel. A multi-dimensional scaling process was then carried out on this data. For the axis under examination, the output supported

the interpretation given in previous studies, that is, the axis is related to the spectral distribution of the tones- a composite of the effects of bandwidth, balance of energy in the lower harmonics, and upper formant regions. The evidence for this was that the modified pairs exchanged positions on the axis compared to the previous study. Since, for each pair, the original bandwidths were maintained when the spectral envelopes were exchanged, it can be concluded that bandwidth is a less significant feature than the actual shape of the energy distribution.

In a 1991 study, Kendall & Carterette (1991) conducted a similarity study on wind instrument duos, that is, tones played simultaneously by two wind instruments where the timbre is a combination of the two individual timbres. Two of flute, oboe, alto saxophone, clarinet and trumpet were played together in all possible pairings (dyads) in unison and at an interval of major third. Human subjects rated the similarity of all possible pairs of dyads. Multi-dimensional scaling was performed resulting in a three dimensional timbre space. The first two dimensions were interpreted as *nasal* v *non-nasal* and *rich* v *brilliant*. A third, less stable dimension, was interpreted as *simple* v *complex*. An important finding was that the timbre space of a composite timbre dyad can be predicted from the vector sum of the positions of the individual component timbres. This finding is consistent with that of Grey (1977).

McAdams, Winsberg, Donnadieu, DeSoete & Krimpoff (1995) analysed and synthesised musical tones with an extended version of the multi-dimensional scaling algorithm (CLASCAL) using similarity ratings obtained from a panel of subjects: some with a high degree of musical training, some with a moderate background of musical training, and some with little musical training.

2.3.5 Multi-dimensional Scaling using Physical Features

In 1966, Plomp and his colleagues investigated the relationship between the timbre of Dutch vowel sounds and the frequency spectrum. They studied 15 vowel sounds with 10 speakers giving 150 observations. The average spectrum for each observation was determined by using n successive frequency passbands which enabled each instance of a vowel sound to be represented as a point in an n-dimensional spectral space. Using a technique that was essentially principal component analysis they were able to represent 84% of the variance in 4 dimensions and at the same time achieve good separation of the vowel sounds. They

found that the first two dimensions corresponded to the formant frequencies of the vowels.

In follow up work, Pols et al. (1969) carried out experiments to investigate the correlation between the perceptual and physical (spectral) space for 11 Dutch vowel sounds. An n-dimensional spectral space was obtained for the vowel sounds by the method used in their earlier work and then reduced to three dimensions by means of principal component analysis. A three dimensional perceptual space was obtained using a triadic comparison procedure and the multidimensional scaling techniques described previously. When the perceptual and physical representations were compared it was found that there was very strong correlation between each corresponding axes. The close correspondence between perceptual space and the physical space was a finding that had important implications for future work in the study of timbre.

In work similar to the above, Plomp (1970, 1976) investigated the relationship between timbre and average spectrum for musical tones. He applied the triadic comparison procedure together with multi-dimensional scaling techniques to create a timbre (perceptual) space and used filtering techniques to create a spectral (physical) space for a set of tones. In both cases each tone was represented by a single point in space. To compare the timbre space with the spectral space he used a canonical matching process developed by Cliff (1966) in which the first three factors (dimensions) were compared. A very strong correlation was found between the timbral and spectral spaces along the factors I, II, III. He concluded that there was excellent agreement between the timbre space and the spectral space indicating that differences in timbre can be predicted from differences in frequency spectrum. This investigation was a precursor to work in the 1990's on multi-dimensional scaling analysis of musical instrument spectra - for example, the work by Hourdin & Charbonneau (1997).

In another similar investigation, Poli & Prandoni (1997) borrowed from techniques used in the speech processing community in order to analyse timbre and algorithmically define timbre space. Their goal was to produce a 'sonoligical' model that included analysis methods most appropriate to the sound characteristics to be studied and an effective data reduction process. Experiments were conducted with 21 instrument sounds generated with a synthesiser. Time varying aspects were highlighted by using short overlapping data windows. The analysis involved the use of mel-frequency cepstral coefficients - a well used technique in speech analysis. Two approaches were used for data reduction and production of a timbre space. The first was a non-linear projection of the n-dimensional space onto a timbre space of reduced dimensions using Kohonen neural networks. The second was a linear projection

onto a space of lower dimensions using principal component analysis. (Note that their use of PCA differed from Plomp (1966) in that spectral data was collected from multiple windows whereas Plomp used the average spectrum.) In both transformations Euclidean distance was retained. The authors found that the mel-frequency cepstral coefficients were well suited to representing musical timbre. They found principal component analysis the most effective technique for creating timbre space. They found the first axis related to the spectral energy distribution (brightness), the second axis correlated with spectral energy across the whole frequency band for musical sounds, and the third correlated to energy in a narrow region of the spectrum around 700Hz. They concluded that the third principal component was a differentiating factor in the quality of musical timbre. Borrowing from the terminology of audio amplifiers, they referred to this characteristic as *presence*. They assert that these qualities of brightness and presence can easily be modified by instrument makers whereas temporal qualities found in the attack stage are fundamentally tied to the instrument structure. These temporal qualities are relatively constant within instrument classes and therefore offer key clues in instrument recognition. The authors conclude that the steady state portion of a musical tone is the key determinant of timbral quality whereas the attack phase is important in instrument recognition.

In an investigation focusing on both spectral and temporal aspects of timbre, Hourdin & Charbonneau (1997) take the general principles of multi-dimensional scaling (MDS) and apply them to a physical description of musical tones. The beginning point is a series of short data windows analysed with a hetrodyne filter to yield spectral-temporal data over the duration of each tone. This physical data is in contrast to the more subjective and perceptually based dissimilarity ratings used by Grey (1977). A further point of difference is that the standard MDS studies do not directly take account of temporal features. To represent the variance of the data in just a few variables the authors used factorial analysis of correspondences (FAC) in a similar way to principal component analysis. Each data point was then plotted in the geometric space created with the new variables. This enabled a dynamic representation of each tone via a closed discrete curve which started and finished with silence. To show that the shape of the curve was related to timbre they examined the effect of changes in duration and pitch on the shape of the curve by synthesising modified tones. It was not possible to examine the effect of changes in intensity since the filtering process normalised intensity. Results showed that changes in duration did not affect the trajectory for that instrument. However, by comparing tones at pitch C4 with tones at C3 artificially raised to C4, they showed that pitch did have a significant effect on

the trajectory. They concluded that the trajectory path for each tone incorporated both spectral and temporal features of a tone and gave a good representation of the timbre of that instrument - it represented a *tone signature*. The trajectory paths for a pair of tones therefore relate to the task of distinguishing between the timbre of those two tones. The authors attempted to interpret each of the first three principal components in terms of physical features in order to compare with the investigations by Grey (1977). The first principle component was thought to correspond to the energy level of the signal - a quantity not relevant to Grey's study. The second principle component was thought to correspond to band width - similar to the first axis in Grey's study. The third principle component was thought to correspond to balance of energy between the lower harmonics and the higher harmonics - an interpretation that did not accord with Grey's second axis. They concluded that spectral width was the most important feature of timbre. The interpretations of the first two principle components accord reasonably well those of Plomp, Pols & van de Geer (1966).

In an extension of the work of Hourdin & Charbonneau (1997), Kaminskyj (1999) set out to use physical data together with data reduction methods to produce a geometric space that would adequately represent the timbre of musical tones in just a few dimensions. As in the Hourdin, study he took a series of short data windows but used the constant Q transform (Brown 1990) to extract the spectral-temporal data. To achieve an initial reduction in data Kaminskyj chose to use only the the frequency bins corresponding to, or adjacent to, the first 20 harmonics. A reduction in dimensionality was achieved by then applying principal component analysis. The first PC was interpreted as corresponding to energy level and band width; the second as the sum of the energy in the odd harmonics; and the third seemed to correspond to energy in the fundamental and odd harmonics. His intention was to use the 3-dimensional trajectories from MDS as a feature in a system for instrument identification.

2.4 Implications of Timbre research for Instrument Recognition

This exploration of timbre has a two fold purpose: firstly, to try to define timbre as precisely as possible and secondly, to investigate what are the physical attributes that enable instrument recognition. At this stage we will accept the *ASA:Psychoacoustical Terminology* (1973) definition of timbre - that attribute by which we can judge that two sounds of the same loudness and pitch are dissimilar. This definition is pertinent to the task in musical instrument recognition where the task is to decide which of a group of musical sounds is least dissimilar to a given musical sound of unknown origin.

The problem with the above definition of timbre is that it is a psycho-acoustic definition relating to human perception whereas what is required is a well accepted precise physical definition which allows timbre to be objectively and empirically measured (Hajda et al. 1997). It should be noted that in contrast to pitch, loudness and duration, timbre does not have a physical equivalent other than the less than satisfactory term - 'wave form'. Since it is generally accepted that timbre is a multi-dimensional quantity/quality (Plomp 1970), we need to be able to define timbre in terms of a number of physical quantities in order to control or objectively evaluate it.

In tracing the path of timbre research we began with the work of Helmholtz who concluded that the timbre of a musical tone was dependent on the harmonic structure of the steady state portion of a tone. His concept of timbre made little allowance for time variant factors in the attack or steady state. Helmholtz investigated the effect of phase on timbre and concluded that it had a small but insignificant effect on timbre , a view that was supported by Plomp (1970) in comprehensive investigations more than than a century later. Helmholtz' harmonic structure theory was not questioned for nearly a century until Richardson (1954) suggested that the attack transient was an important cue in the instrument recognition. This view was supported by the work of Luce (1963), Clark et al. (1963), Berger (1964), Saldana & Corso (1964), Grey & Moorer (1977). There was no general agreement regarding the relative importance of the attack compared to the steady state and in fact the answer to this question probably varies depending on the instrument class. The importance of the attack was explained in terms of the different behavior for each successive partial for a particular instrument and that the onset times and temporal gradients (rate of growth) were different for each instrument. Investigations in general (for example, Berger (1964))

confirmed the commonly held belief that the decay was of little importance to timbre. If it is now accepted that the attack is an integral part of timbre then perhaps we can no longer use the terms 'tonal quality' and 'timbre' interchangeably. Although tonal quality has never been precisely defined, since the time of Helmholtz it has been used to refer to a judgement on the quality of a sound in the steady state. Tonal quality was considered to be dependent on the steady state spectrum.

In the 1960's and 1970's there was a growing interest in time varying aspects of timbre. In their investigation into trumpet tones, Risset & Mathews (1969) confirmed the importance of the attack transients but also found that the time varying aspects of the steady state were important - in particular the periodic and random fluctuations in pitch. Beauchamp (1974) highlighted the time variant nature of the spectrum in violin tones. He noted that recognition became easier if vibrato was present. Grey & Moorer (1977) came to similar conclusions that, after the attack and the average spectrum in the steady state, changes in the spectrum over time and fluctuations in the fundamental frequency were of importance. Although we have stated that phase is not a significant factor in timbre, change in phase over time is one factor that causes fluctuation in the spectrum.

Since the ASA definition of timbre refers to tones of equal loudness and pitch (frequency), it is important to discuss the relationship between timbre and loudness (intensity), and timbre and pitch. Since the spectra of musical instruments is dependent on intensity it might be expected that as a consequence timbre would be dependent on intensity. However Clark et al. (1964), Strong & Clark (1967*b*) found this not to be so. They found that changes in spectra due to variation in intensity caused no significant change in timbre, however, there seems to be considerable evidence that timbre is highly dependent on pitch. Investigations by Bachem (1955) and Shepard (1982) suggest that the timbre of a simple tone is repeated at intervals of an octave. Other researchers, for example Richardson (1954), Risset & Mathews (1969), Beauchamp (1974), found that timbre and, in particular, the spectra to be frequency dependent. They explained spectral fluctuations with frequency change in terms of formants in the frequency response of the instruments under consideration. The concept of formants dates back to early research in speech perception - for example Willis (1830), Helmholtz (1863). It was found that each vowel sound was characterised by certain peaks in the spectrum. These formants were of fixed frequency irrespective of the fundamental frequency of the vowel. Slawson (1967), provided conclusive evidence for the existence of formants in the timbre of both vowel sounds and musical sounds. It follows

that the sounds have a spectral envelope fixed on the frequency continuum and that as the fundamental frequency varies, the spectra varies in accordance with the spectral envelope. The existence of formants for non-percusive musical instruments was supported by the findings of Luce (1963), Risset & Mathews (1969), Beauchamp (1974). The formants were explained in terms of proximity to frequencies of strong resonance in the instruments. Sympathetic transient tones, where a string or a part of an instrument is excited and vibrates in 'sympathy' with the prevailing tone, have an influence on formants and enrich the timbre of tones at that frequency. The mean frequencies and the bandwidth of the formants were considered to be key factors in determining timbre (Plomp et al. (1966), Brown (1981)). The issue of formants for the violin and guitar will be discussed in detail in chapter five.

There have been many attempts to isolate quantities which may control or, in some way, measure timbre. Some investigators have given labels to certain aspects of timbre which can only be used as qualitative descripters and assigned in a subjective matter - for example, Stumpf (1890), Pratt & Doak (1976). Many investigations have involved a subjective choice of quantities and then an attempt to measure the perceived change in timbre corresponding to a change in a particular variable. In an early attempt to find quantitative attributes that defined timbre, Lichte (1941) proposed *brightness* which is determined by the location of the mean of the energy distribution on the frequency continuum (spectral centroid); *fullness* which depends on the ratio of odd to even numbered partials; and *roughness* which describes tones with consecutive high partials above the 6th (a qualitative rather than quantitative attribute). There have been a number of other parameters defined in order to summarise spectral properties. *Spectral bandwidth* has been commonly used (for example, Grey & Gordon (1978)). Another example is the *tristimulus method* of analysis proposed by Pollard & Jansson (1982), where the spectral energy is divided into three bands and the energy level for each band is determined. Since the the attack transient has been accepted as an important cue in recognition, much work has been done in analysing aspects of the attack, in particular the onset of partials. The most commonly used variables have been the the onset (beginning) time, and the rate of growth (gradient) (Saldana & Corso (1964), Freedman (1968)).

There are a number of features that assist in describing timbre where it is not the measurement of the quantity that is significant but the fact of its presence or absence - for the purpose of timbre analysis they can be considered boolean quantities. For example,

the presence of frequency fluctuation whether periodic (*vibrato*) or random (*jitter*), was found by numerous researchers to be an important cue in timbre identification (for example Saldana & Corso (1964), Fletcher et al. (1962), Risset & Mathews (1969), Charbonneau (1981), Grey (1977)). Amplitude fluctuation (usually referred to as amplitude modulation) was also considered to be an important quality of timbre although it is often correlated with frequency fluctuation. McAdams et al. (1999) coined the term *spectral flux* to refer to spectral variation over time due to both amplitude and frequency fluctuations. He also suggested that the degree of smoothness (jagged v smooth) in the spectral envelope was important in describing timbre.

The presence of inharmonic frequencies in the signal are also considered an important cue. These may take several forms:

- inharmonic frequencies in the attack - for example natural resonances in the body of an instrument (Fletcher & Rossing (1998), Grey (1976));
- inharmonic partials in string instruments such as guitar and piano (Saldano & Corso(1964), Fletcher (1962,1964), Beauchamp (1974), Grey (1976));
- and instrument noise generated in the production of the tone - for example bow noise in violin tones (Fletcher 1965), mechanical noise in saxophone tones and piano tones (Fletcher 1962).

Much of the advance in understanding of timbre has come from work in musical instrument synthesis - generally the additive synthesis method where a comprehensive spectral analysis over time is performed on actual musical instrument tones using such tools as the hetrodyne filter. The tones are then re-synthesised using some form of data reduction method. According to Charbonneau (1981), synthesis is perfect when indistinguishable from the original tone. The problem with research in this area, is that the the choice of parameters and methods of data reduction is generally subjective. The researchers in this domain are not looking to refine the definition of timbre but rather to achieve a better result in music synthesis.

The advent of multi-dimensional scaling (MDS) in timbre research, (for example, Plomp & Steeneken (1969), Grey (1977)), provided a more objective means of analysing timbre as well as a means to create an n-dimensional timbre space. Building on the work of Shepard (1962a,1962b) and Kruskal (1964a,1964b), they used subjective dissimilarity together

with the MDS algorithm to create an n-dimensional timbre space. Using associated data reduction methods they were able to reduce the timbre space to 3-dimensions. Much was learnt from the MDS investigations into timbre, especially the much cited series of investigations by Grey et al. (1977a,1977b,1978a,1978b). Each of the three axes in the timbre space was interpreted to yield valuable information about timbre as discussed above. These investigations, however, had two limitations in that, firstly they began with a subjective judgement of instrument dissimilarity, and secondly they did not represent time varying aspects of timbre. Plomp et al. (1966) used physical features , namely spectral data, to reveal information of the timbre of Dutch vowel sounds and musical tones. This investigation allowed for objective analysis in that it was founded on physical data rather than subjective dissimilarity judgements but again it did not take into account the time varying aspects of timbre. Hourdin & Charbonneau (1997), in an attempt to replicate the work of Grey (1977) but with physically based features, used time varying spectral data as a basis for an investigation. This allowed a *trajectory path* to be plotted for each musical tone based upon the time varying spectral data. They demonstrated that the trajectory path for each tone was a good representation of the timbre of that instrument - it represented a *tone signature*. So the task of assessing the closeness of the timbres for two musical tones now depended on the closeness of the two trajectory paths rather than the closeness of two points as in the standard MDS timbre space. Another important outcome was the authors finding that timbre is frequency dependent.

We note that there are variations in the way this timbre space can be constructed. Kaminskyj (1999) used principal component analysis rather than factor analysis as a means of data reduction which enabled the frequency-time data to be represented more succinctly than with factor analysis. Poli & Prandoni (1997) and Brown (1999) used cepstral coefficients as means of representing the time varying spectral data.

In conclusion, we summarise the properties of timbre upon which there is reasonable consensus. Timbre is a multi-dimensional quantity that depends on both the attack transient and the steady state portions of a tone. In the attack, the important aspects are the manner in which the partials evolve and the presence of inharmonic frequencies. In the steady state it is the shape of the frequency spectrum including fluctuations over time of both a periodic and random nature that are important. There is also agreement that the spectral envelope is highly dependent on fundamental frequency and to a much lesser extent dependent on loudness. It must be restated, however, that there is no consensus on

a precise operational or constitutive definition of timbre which allows it to be objectively and empirically measured (Hajda et al. 1997). In other words, there is no single parameter or combination of parameters which allow us to measure and assess timbre. There is good evidence, however, that the curve or trajectory path from the MDS process based on a physical analysis (short term spectral analysis over the the duration of the tone) gives an accurate representation of the timbre of a tone. If we can find a way to quantitatively assess the closeness of the trajectories of two tones then we may have a means of objectively measuring the closeness of the timbres and hence classifying musical instrument tones.

The next stage in this thesis is to outline a means by which the timbres of the instruments in this study, namely the guitar and violin, can be represented and further, to outline a means by which the similarity of the timbre for the instruments can be measured This similarity measure will then facilitate the classification of instruments within the violin and guitar classes.

Chapter 3

Theoretical Background of the Analysis

3.1 Introduction

3.1.1 The Nature of a Musical Tone

In comparing the timbre of musical tones, we begin at the point that a musical sound is a complex mixture of frequencies that vary in both a random and a systematic manner over time. In general, each tone has a beginning (*an attack*), a middle (*steady state*) and an end (*decay*). Instruments can be broadly divided into two classes: firstly, those with continuing tones - an attack followed by a steady state continuous tone and finally a short decay when the tone terminates (for example the violin) ; and secondly, those with impulsive tones - a fast rising attack followed by an exponential decay (for example the guitar and piano). Instruments such as the piano and, to a more limited extent, the guitar are characterised by an extended decay which approximates a steady state.

Each musical instrument has its own *signature* or defining features. In the time domain each musical sound can be represented by a complex sinusoidal wave graph. The attribute most visible in the time domain is the power envelope. For example in figure 3.1, a plot of the the sound wave shows the differing power envelopes for a violin and guitar. Kaminskyj & Materka (1995) used the power envelope as the sole discriminating feature to successfully discriminate between guitar, piano, marimba and accordion. For any musical tone, if a

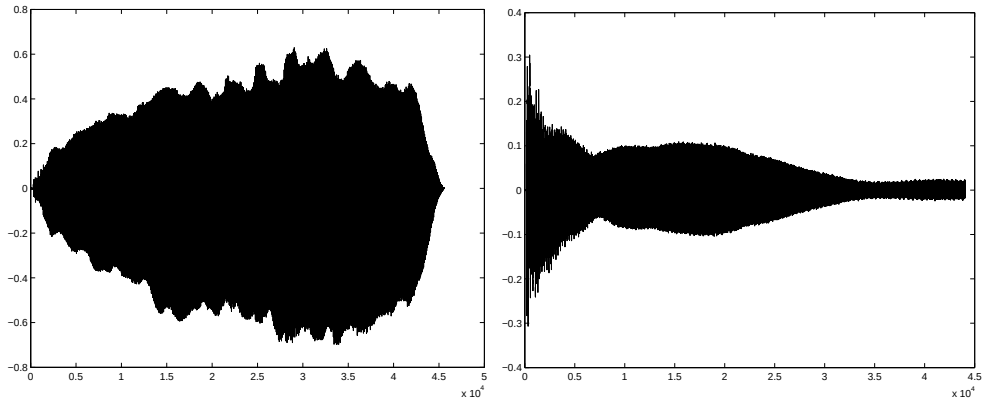

Figure 3.1: Wave envelope for a violin tone(left) and a guitar tone(right).

segment of the wave graph is enlarged we can see the periodic nature of the tone. For example, see the violin tone illustrated in figure 3.2.

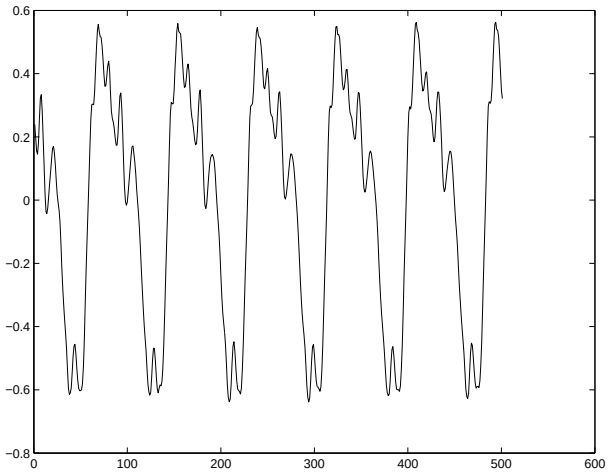

Figure 3.2: A magnified section of a violin wave showing the periodicity.

3.1.2 Mathematical Representation of a Musical Tone

A simple tone, characterised by a single constant frequency and a constant amplitude (power), can be represented mathematically by the periodic function,

$$g(t) = a \sin \omega t \tag{3.1}$$

where ω is the angular frequency of the tone and a is the amplitude of the function.

A complex musical tone can be considered to be a linear combination of a number of simple tones each with a different frequency and magnitude. For most musical instruments the components of a tone will vary in frequency and magnitude over the duration of the musical tone. However for the purposes of sampling and analysis we must assume that each component is constant in frequency and magnitude over a short interval of time. A complex tone, where no phase shift is assumed, can be expressed in the form,

$$x(t) = c_0 + a_1 \sin \omega_1 t + a_2 \sin \omega_2 t + a_3 \sin \omega_3 t + ... \tag{3.2}$$

where ω_k , for k a positive integer, represent the frequencies of the component tones. More generally, allowing for phase shift,

$$x(t) = c_0 + a_1 \sin(\omega_1 t + \varphi_1) + \alpha_2 \sin(\omega_2 t + \varphi_2) + \alpha_3 \sin(\omega_3 t + \varphi_3) + ... \tag{3.3}$$

or equivalently,

$$x(t) = c_0 + a_1 \cos \omega_1 t + b_1 \sin \omega_1 t + a_2 \cos \omega_2 t + b_2 \sin \omega_2 t + ... \tag{3.4}$$

which can more conveniently be expressed in complex form. Namely,

$$x(t) = \sum_{k=0}^{\infty} \alpha_k e^{j\omega_k t} \tag{3.5}$$

where $|\alpha_k|$ is related to the power of each component tone.

Musical tones for most instruments other than drums are primarily harmonic in nature. That is the steady state portion of a tone is comprised of *harmonics* (or *partials*) each with a frequency which is a multiple of the fundamental frequency. That is, harmonic tones take

the form $\omega_k = \omega_f k$ where k is a positive integer indicating the number of the harmonic and ω_f is the fundamental frequency (first harmonic). The equation becomes,

$$x(t) = \sum_{k=0}^{\infty} \alpha_k e^{j\omega_f kt} \tag{3.6}$$

The particular combination of harmonic tones gives a musical tone its *flavour* and makes a major contribution to the characteristic signature of each musical instrument.

All musical tones also include inharmonic components such as mechanical noise related to the mechanics of the instrument or natural resonances in an instrument, audible particularly in the transient attack stage. These inharmonic tones also help to define the characteristic signature of a musical instrument. Equation 3.6 can be modified to take account of the presence of both harmonic and inharmonic tones, namely,

$$x(t) = \sum_{k=0}^{\infty} \alpha_k e^{jk\omega_f t} + \sum_{l=0}^{\infty} \beta_l e^{j\omega_l t} \tag{3.7}$$

where ω_f is the fundamental frequency,
ω_l represents the frequencies of the inharmonic tones,
and $|\alpha_k|$, $|\beta_l|$ are the amplitudes of the respective harmonic and inharmonic tones.

In summary, for any tone from a musical instrument, the component tones will vary in both frequency and their amplitude over time. Consequently, a musical tone can only be considered to be periodic over a short window in time. It is the set of component tones present and their variation in both frequency and amplitude over a period of time that defines the characteristic signature of the tone.

3.1.3 Representing a Musical Tone in the Frequency Domain

The task of uncovering the secrets of each instrument's *signature* is best begun in the frequency domain. It is generally agreed that a greater quantity of relevant data can be more readily accessed if the data is transformed from the time domain and processed in the frequency domain (for example, Ifeacher & Jervis (1993)). The frequency spectrum of a musical tone over a short interval of time gives an accurate picture of the sound we are hearing at that time. For example, figure 3.3 shows a plot of the spectrum for a violin at a certain time during the steady state portion of the tone.

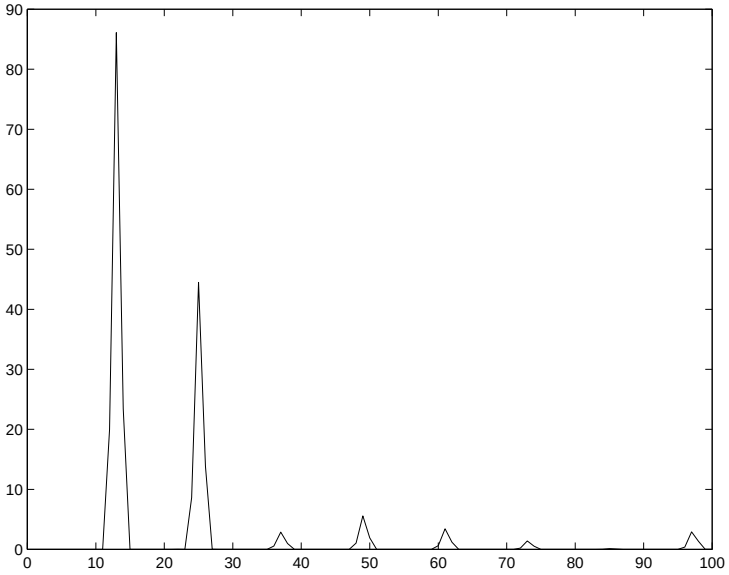

Figure 3.3: The spectrum of a violin tone over a short period of time.

Since any musical tone is constantly changing with time, to fully describe the timbre of a tone we need a sequence of 'snap shots' of the spectrum at short intervals throughout the duration of the tone. Each snap shot involves selecting a short window of the signal in the time domain and transforming the data into the frequency domain.

We can plot the frequency data from this series of snap shots over time on a 3-dimensional waterfall graph to show the evolution of a tone over time. In fact this gives us a good pictorial representation of the timbre of a complete musical tone. For example, figure 3.5 shows the waterfall plot for a guitar tone illustrating the changes in frequency and amplitude of each component over time.

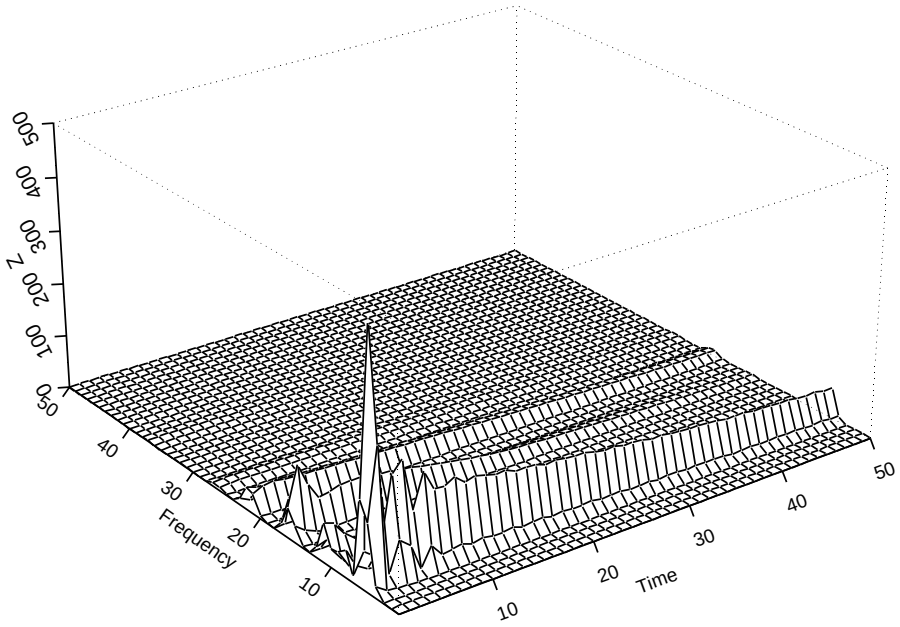

Figure 3.4: Waterfall plot showing the spectrum of a guitar tone over time.

In order to capture the evolution of a tone, snapshots of the wave form are taken via a moving window to give a picture of the tone development over a period of time. The data for each snapshot is transformed into the frequency domain using either the fast Fourier transform (FFT) or the constant Q transform (CQT) (Brown 1990). The scaling may be either linear in terms of frequency (FFT), or linear in terms of pitch (CQT) and hence logarithmic in terms of frequency. Performing an FFT on each window gives the average power for each frequency present in the tone over that short period of time. In this experiment a series of 50 overlapping windows of length 0.05 seconds was used. The data from each of the snapshots can be represented as an n-dimensional plot which illustrated the development of the tone over time. Each axis represents the power at a certain frequency determined by the *bin* frequencies in the FFT. The corresponding frequency for each successive axis increases incrementally within a predetermined range. Each window corresponds to an increment in time from the start of the wave data to the end and is represented as a point in n-dimensional space . The sequence of points defines the multi-dimensional scaling trajectory path which represents the evolution of the timbre of the tone over time (see MDS in chapter 2).

3.2 Time-Frequency Transform with the Discrete Fourier Transform

To trace the evolution of each frequency present in a tone over time, we need to accurately determine the frequency spectrum for each of the adjacent windows in the time domain.

In the process of recording each tone, the amplitude of the original analogue signal $x(t)$ is sampled at a given rate and saved in digital form. The sampling rate determines the highest frequency that can be preserved (Nyquist sampling theorem) (Ifeacher & Jervis 1993). To analyse the harmonic structure of a musical tone at a given time, a window over a short interval of time is considered. The width of the window determines the lowest frequency that can be processed and the degree of resolution.

To describe the waveform for a musical tone of the form given by equation 3.5 at a particular time, we assume periodicity for the duration of a window and determine the frequencies present and their corresponding magnitudes. In practical terms, we can find estimates of these values by the use of a series of narrow pass-band filters. Equivalently, we can determine the frequencies present by use of the discrete Fourier transform (DFT).

Applying the DFT to the data for a given window enables us to determine a Fourier series (Ifeacher & Jervis 1993) of the form,

$$x(t) = \sum_{k=0}^{N-1} c_k e^{j\omega_k t} \tag{3.8}$$

where ω_k is the angular frequency of the kth frequency bin,
N is the the window width in samples
and $|c_k|^2$ is the power at each frequency.

The Fourier series generated by the DFT represents discrete frequencies from ω_0 to ω_{N-1}. The lowest frequency that can be detected is ω_0 which is determined by the window width. Each successive frequency in Fourier series will be a multiple of ω_0. That is, the discrete bin frequencies can be expressed as $\omega_k = k\omega_0$.

The DFT can be written more conveniently in terms of the frequency bin numbers, namely,

$$x(n) = \sum_{k=0}^{N-1} c_k e^{j(2\pi k)n/N} \quad where \quad c_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n) e^{-j(2\pi n)k/N}. \tag{3.9}$$

where n is time measured in samples.

It follows that the window size, N, also determines the frequency resolution Δf (width of each bin). If a window width is doubled then the resolution is improved by a factor of two such that the bin width and the lowest possible frequency is halved. There would, however, be no change in the highest detectable frequency since, as stated earlier, this is determined by the sampling rate of the original data. The choice of window width also relates to the ability of the process to represent the frequency components of a tone at a given time . By necessity, analysis of a window gives an average over time. Hence the larger the window the less able the process is to represent change over time. So we have two competing requirements here: the need for high resolution in the frequency spectrum and the need for short term frequency analysis. In musical tones the rate of change is sufficiently slow that the DFT can satisfy both needs. In other contexts, where the rate of change is faster, wavelet analysis is a more satisfactory tool.

An important consideration relating to the Nyquist frequency is the possible distortion in the frequency spectrum that can occur if there are frequencies present in the signal above the Nyquist cut-off frequency. This is due to the periodic nature of the DFT output and the resulting overlap or 'rollover' that occurs with frequencies above the Nyquist cut-off frequency. It is, therefore, essential that the musical tones are filtered to eliminate any frequency components above the Nyquist cut off.

It is clear that, since the DFT resolves the signal into discrete frequency bins, there will not always be an exact match between the frequencies in the signal and the available frequency bins in the DFT. Consequently, if the frequencies in the tone do not fall on or near the frequencies in the DFT, then there will be some distortion in the resulting spectral data. An alternative way of thinking of the problem is to begin with the idea that the DFT generates a sinusoidal series based upon the segment of the wave form in the window. If the segment is a periodic wave form that is commensurate with the window (that is, the window encases an exact number of cycles), then the DFT will merely reproduce that function. If the original wave form is not commensurate with the window (that is, the period does not evenly divide into the window width), then the DFT will produce a more complex sinusoidal function to create a new periodic function from the fragment of a sinusoidal function.

If, to analyse a tone, we perform a DFT on a data window of a predetermined fixed width, then many of the component frequencies in the tone will not be commensurate with the window size. Hence, in the resulting frequency spectrum, there will be considerable

'leakage' of frequency to other frequency bins and a degree of error in the spectral analysis.

3.2.1 An Example: the Effect of Window Size

Let us now consider some data and examine the effect of varying the window size. Data for a simple tone of 110Hz was obtained by sampling the sine function $y = \sin(2\pi 110t)$ at a rate of 11000Hz. We can determine that the period for this tone is 100 sample periods (data points). A DFT was performed on the window and the data plotted .

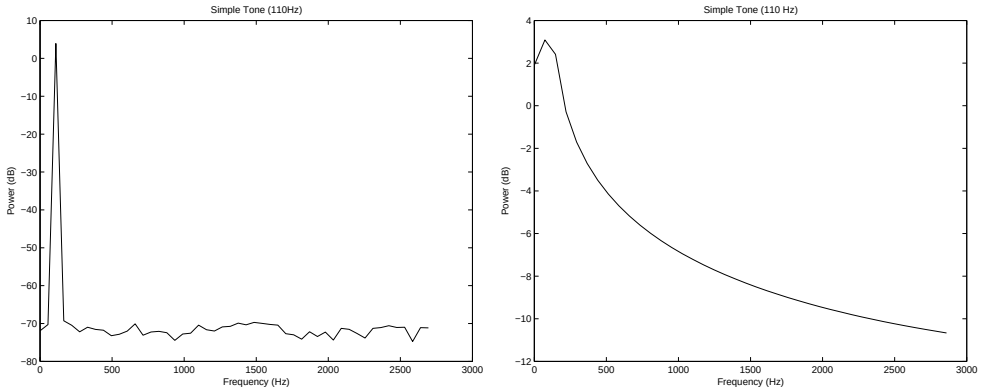

Figure 3.5: Two plots of frequency spectrum for a simple tone of 110 Hz (wave length 100 data points): (a) window length $N = 200$ - 2 periods of simple tone; (b) window length $N = 150$ -1.5 periods of the simple tone.

Figure 3.5(a) shows the simple tone as a distinct single spike when the window size is an exact multiple of the the period of the tone. Figure 3.5(b) shows the distorted spectrum when the window size is not an exact multiple of the period - a leakage of frequencies to neighbouring bins suggesting the presence of frequencies that are not actually present in the tone. These 'phantom ' frequencies will be passed to the classification stage thereby providing incorrect information about the harmonic structure of the tone. It can be shown that increased resolution reduces the overflow effect but this is not a satisfactory solution as the consequent larger window size compromises the ability of the DFT to measure changes in the spectrum over time.

3.2.2 A Closer Look at Overflow with the Fourier Transform

Let us examine more closely how a particular frequency will be represented in a frequency spectrum when the window period is not commensurate with the period of $x(t)$. For simplicity we will consider the continuous function and apply the continuous Fourier transform rather than sampling the function and applying the DFT. This approach will also enable us to generate a function in the frequency domain and analyse its behaviour (Moore & Cerone 1999).

Example 1: Consider the theoretical case of a known $x(t)$ comprised of a single frequency, for example $x(t) = \sin 3\pi t$ with a period of $\frac{2}{3}$ and a frequency of 1.5.

The Fourier series generated for $x(t)$ would take the form,

$$x(t) = \sum_{-\infty}^{\infty} c_n e^{\frac{j2\pi nt}{T}} \qquad where \qquad c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t).e^{\frac{-j2\pi nt}{T}} dt. \tag{3.10}$$

Firstly, let us take a window from $x(t)$ with period $T = 2$, such that the periods of $x(t)$ and the window are commensurate. The bin frequency in the spectrum will be 0.5. It can be shown that,

$$|c_n| = \begin{cases} \frac{1}{2} \;\; if \;\; n = 3 \;\; (ie.\ f = 1.5) \\ 0 \;\; elsewhere \end{cases} .$$

Example 2: Now let us take a window from $x(t)$ with period $T = 1$, such that the periods of $x(t)$ are not commensurate. The bin frequency in the spectrum will be 1. It can be shown that:,

$$|c_n| = \frac{1}{\pi} \left| \frac{4n}{(4n^2 - 9)} \right| .$$

We notice in figure 3.6(a), in the plot of $|c_n|$ for example 1, that their is a spike peak at $n = 3$ corresponding to the fundamental frequency of $x(t)$. In figure 3.6(b)(example 2) there is an overflow to frequency bins adjacent to the peak at $n = 1.5$. This is because the actual frequency of $x(t)$ falls between frequency bins $n = 1$ and $n = 2$. It can be concluded that the effect of non-commensurate periods is that, for the DFT, there will be an overflow to neighbouring bins. It can be shown that, except at low frequencies, the frequency does not impact greatly on the overflow to adjacent bins.

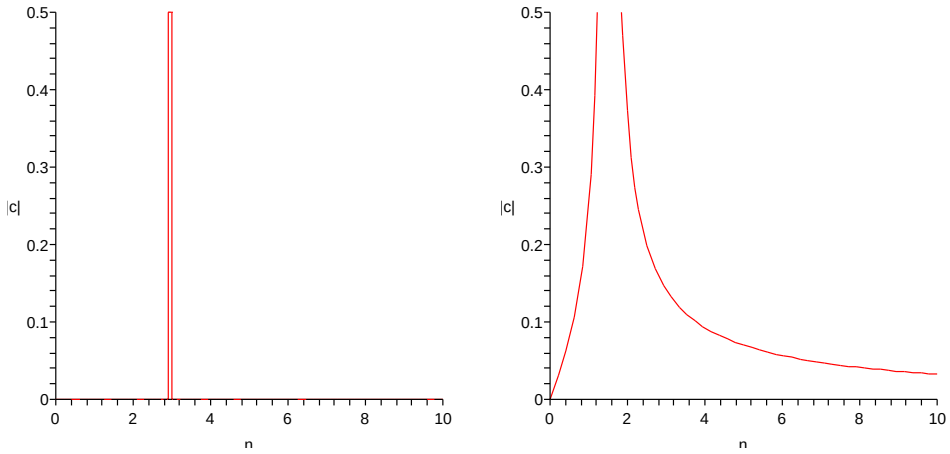

Figure 3.6: (a) A plot of frequency n versus $|c_n|$ when the period of the function is commensurate with the period of the window, (b) A plot of frequency n versus $|c_n|$when the period of the function is NOT commensurate with the period of the window.

3.2.3 Windowing: A Solution to Spectral Leakage

We have previously discussed how, for musical tones, frequency analysis over time can be achieved by taking a series of adjacent rectangular windows of data in the time domain and transforming them into the frequency domain using DFT. We have shown that any component frequencies which fall between the discrete frequencies of the DFT cannot be accurately represented. These component frequencies will be truncated and the energy shared with neighbouring frequency bins. At low frequencies, due to the closeness of adjacent harmonics, there will be distortion in neighbouring harmonics.

This problem of *spectral leakage* cannot be eliminated but the error can be reduced by using the windowing techniques developed in the signal processing domain (Ifeacher & Jervis 1993). In finite impulse response (FIR) filter design a windowing technique is used to minimise the distortion of through-put frequencies. The DFT operating on data of a finite length can be considered to be a bandpass filter with upper and lower frequency limits determined by the window width and the sampling rate. This window of data can be expressed mathematically by considering that we are effectively multiplying the amplitudes in the sampled section, x_0 to x_{N-1}, by unity while all other data values outside this section are multiplied by zero. The rectangular window function can be expressed as,

$$w(n) = \begin{cases} 1, & n = 0, 1, ...N-1 \\ 0 & elsewhere. \end{cases}$$

The sampled data values $x(n)$ can be expressed as the product of the window function $w(n)$ and the full set of data values $s(n)$,

$$x(n) = w(n)s(n). \tag{3.11}$$

This product in the time domain is equivalent to convolution in the frequency domain with the power of the nth frequency component given by,

$$X(\omega_n) = \sum_{k=-N}^{N} W(\omega_n - \omega_k)S(\omega_k), \tag{3.12}$$

where ω_n is the angular frequency of the nth frequency component and X, W, S refer to the transformed functions in the frequency domain.

As we have previously stated, in comparison with the other windows, the rectangular window introduces maximum spectral leakage into the spectrum in the form of side lobes for each frequency component of the signal (Ifeacher & Jervis 1993). A number of tapered window shapes have been developed in the signal processing domain to minimise the effect of the spurious side lobes. An unfortunate side effect of all these window types is a thickening of the main lobe which causes overflow into the adjacent lobes. Window types such as the Hamming, Hanning and Kaiser-Bessel each have advantages and disadvantages and a subjective decision must be made as to which window type is most appropriate in a particular application. The window chosen in our study is the Hanning window given by,

$$w(n) = \alpha + (1 - \alpha)cos(2\pi n/N), \tag{3.13}$$

where $\alpha = 1/2$ and $0 < n < N - 1$. The Hanning window has been successfully used in numerous studies of musical instrument timbre.

In summary, our analysis of a musical tone is performed by selecting adjacent or overlapping sections of the tone $x(n)$ in the time domain and then multiplying by the hamming window $w(n)$ to obtain our modified set of data values $v(n)$ given by,

$$v(n) = w(n)x(n). \tag{3.14}$$

The frequency spectrum for each sampled section is obtained by the DFT in the form

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} w(n)x(n)e^{-j(2\pi n)k/N} \tag{3.15}$$

The choice of window width is a trade off between resolution and the ability to represent the spectrum accurately at a particular point in time. A narrow window will have poor resolution for lower frequencies whereas a wider window will have an averaging effect on the spectrum and changes over time will not be well represented.

When the above is implemented in a Matlab program, the output is a matrix of data describing the evolution of the tone over time in terms of frequency. Each row is a vector containing the power/magnitude at discrete frequency values across a predetermined range.

3.2.4 Quantitative Description of Spectral Features

The amount of data in a frequency spectrum is large as is the number of variables and the relationships are complex. The spectrum contains a large amount of information describing the timbre of a tone but it is not readily accessible. It is useful to have quantitative measures that summarize spectral information and allow ready comparison between spectrum. The *spectral centroid* is one such measure. It is a measure of center for the power in relation to frequency. It is an indicator of the power and extent of the higher harmonics relative to the fundamental frequency. In perceptual terms it is an indicator of the *brightness* of a tone (also *richness* or *fullness*). It is important to define spectral centroid so that it is independent with respect to fundamental frequency in order to allow comparison between tones of differing fundamental frequency. To achieve this, a logarithmic frequency scale is used to give a linear scale in terms of musical intervals and moments are calculated about the fundamental frequency. One method of calculating the spectral centroid (sc) is given by the formula,

$$sc = \frac{\Sigma log_e(f(k)/f_0)P(k)}{\Sigma P(k)}, \tag{3.16}$$

where f_0 is the fundamental frequency, $f(k)$ is the frequency of the kth harmonic and $P(k)$ is the power of the kth harmonic.

In this study, the use of the spectral centroid has several important applications: firstly, for a particular tone, changes in spectra over time can be observed; and secondly, by considering the average spectral centroid, tones of different pitch from a particular instrument can be compared. For example, figure 3.7 shows a comparison of the spectral centroid for two independent sets of recordings for a particular guitar over the pitch range of the instrument. We can see that the irregular plot of spectral centroid versus frequency is highly reproducible with independent recordings.

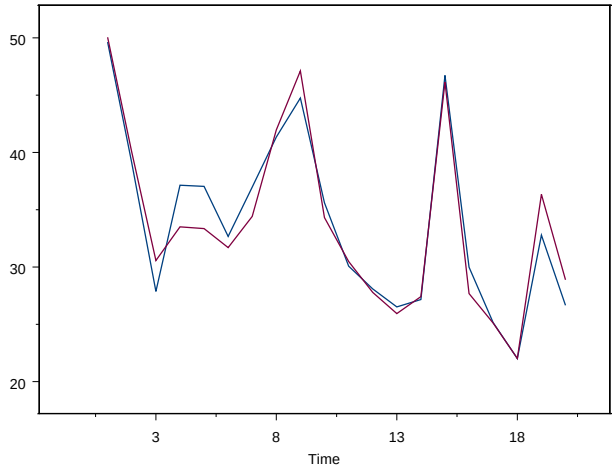

Figure 3.7: Average spectral centroid for two sets of independent recordings of guitar#4 across the frequency range of the instrument.

3.2.5 Time-Freq. Transform with the Constant Q Transform

In music related research, an issue that often arises is the fact that western music is based on a twelve tone scale which is geometrically spaced. When the FFT is used to transform musical data from the time domain into the frequency domain the output is linear in terms of frequency and exponential in terms of musical notes. In musical tasks such as note identification (pitch tracking) or instrument recognition it is often more convenient to have the musical scale (pitch) expressed on a linear basis and the corresponding frequencies expressed logarithmically. On a log frequency plot, the relative positions of the harmonics for a musical tone are the same irrespective of the fundamental frequency. Simply taking the logarithm of the FFT output will give an output in the desired logarithmic form but creates problems with resolution. The resolution at low frequencies will be poor and the resolution at higher frequencies will be greater than required. To address this problem, Brown (1990) developed the constant Q transform (CQT) which gives consistent resolution across the musical scale. This means that the resolution is geometrically related to the frequency. It follows that frequencies sampled by the discrete Fourier transform should be exponentially spaced to give constant resolution in terms of the musical scale. Brown suggested that the resolution should be set at quarter tone intervals in order to distinguish between musical tones at semi-tone intervals. This gives a variable resolution of $(2^{1/24} - 1) = 0.029$ times the frequency. In other words, for quarter note resolution we have a constant ratio of

frequency to resolution (constant Q) given by,

$$Q = f/\delta f = f/0.029f = 34.$$

Given that our spectral analysis is at intervals of a quarter tone, the frequency of the kth spectral component will be

$$f_k = (2^{1/24})^k f_{min},$$

where f will vary between f_{min} up to f_{max}. The minimum and maximum frequencies are set to cover the range where information is required with the proviso that the maximum does not exceed the Nyquist frequency. Given that for the FFT, the resolution (bandwidth) is given by the ratio of sampling rate to window size, to achieve the desired constant resolution the window size must vary inversely with the frequency. It follows that, to determine the spectral power at each frequency increment across the range, a separate application of the FFT with a varying window size will be required. Since the sampling rate is given by $S = 1/T$ where T is the sample time, the window length at frequency f_k is given by,

$$N(k) = S/\delta f_k = (S/f_k)Q. \tag{3.17}$$

Using the discrete Fourier transform, the kth spectral component of the constant Q transform can be written as,

$$X(k) = \sum_{n=0}^{N-1} w(n)x(n)e^{-j(2\pi n)k/N}. \tag{3.18}$$

Taking into account the constraints of the constant Q transform, the above equation becomes,

$$X(k) = \frac{1}{N(k)} \sum_{n=0}^{N(k)-1} w(k,n)x(n)e^{-j(2\pi n)Q/N(k)}, \tag{3.19}$$

where window length is $N(k) = N_{max}/(2^{1/24})^k$
and N_{max} is Q times the period of f_{min} in sample points.

3.3 Characteristic features and Classification

3.3.1 Representing a Musical Tone Using a Multi-Dimensional Scaling Trajectory Path

From the work of Hourdin & Charbonneau (1997), we have observed that a useful means to represent the evolution of a musical tone over time is as a series of n-dimensional plots in the frequency domain (multi-dimensional scaling trajectory path). Each axis represents the power at a certain frequency and where each point is plotted represents the set of frequencies present at a particular time. This is achieved by selecting a sequence of adjacent or overlapping data windows from the raw data (wave file) for a particular tone and transforming into the frequency domain. The result is a matrix of data containing information about the evolution of the tone over time in terms of the frequency. Each row of the matrix is a vector defining the physical attributes of the tone at a particular time in terms of power and frequency and can be plotted in n-dimensional space. Each variable corresponds to a certain frequency. The increment size and the range for frequency are predetermined in the transform. Plotting this sequence of points generates a trajectory path that traces the evolution of a particular tone over time.

To reduce the dimensionality of these n-dimensional plots and to uncover the structure of the timbre, principal component analysis or factor analysis can be applied. In this process the spacial location of the n-dimensional plots is essentially unchanged but the axes are rotated to represent the data with a reduced number of variables. The new variables no longer correspond to actual frequencies present in the tone but relate to frequency dependent physical features such as power and spectral centroid (brightness). We can give a reasonable representation of the trajectory path with a plot of the first three principal components as shown in figure 3.8.

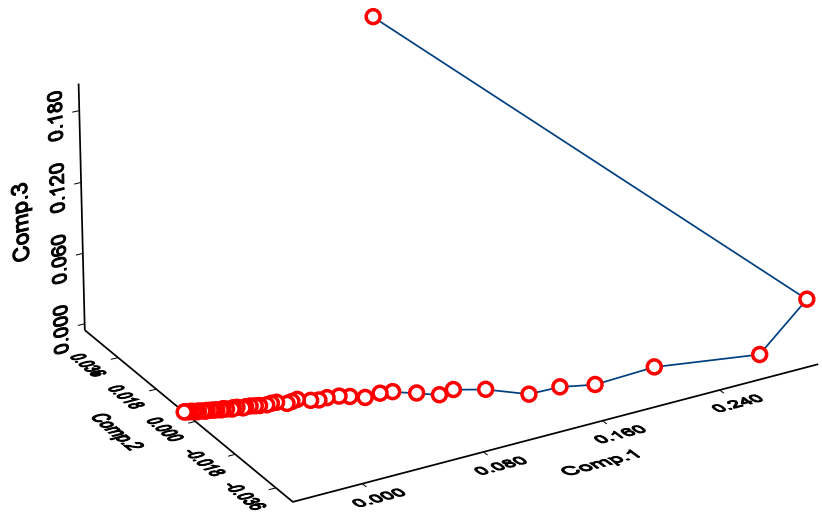

Figure 3.8: A trajectory path in 3D for a guitar tracing the evolution of a guitar tone over time.

3.3.2 More on Principal Component Analysis

The basic idea of principal component analysis (PCA) is to express a set of multi-dimensional data in terms of a few uncorrelated variables which are a linear combinations of the original variables. The two main objectives are: firstly, data reduction - the original variables are replaced by a smaller number of variables called principal components which represent most of the total variance in the data; and secondly, interpretation - the principle components often reveal relationships that could not be observed in the original data (Johnson & Wichern 1982).

The first principal component is a linear combination of the original variables that defines an axis with the maximum possible variance. Each successive principal component is uncorrelated with existing principal components and represents the largest remaining variance. The effect is that most of the variance in the system can be represented in just a few principal components. The principal components depend entirely on the covariance matrix of the system and do not require the system to follow a multivariate normal distribution.

To find the principal components of a given multivariate system with n variables denoted

as $\mathbf{x} = [x_1, x_2.....x_n]$, the first step is to find a linear function $\alpha_{\mathbf{1}}'\mathbf{x}$, in terms of the elements of $\mathbf{x}$, that has maximum variance and takes the form,

$$y_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_2 + ... + \alpha_{1n}x_n. \tag{3.20}$$

The next step is to find a second linear function $\alpha_2'\mathbf{x}$ with maximum variance that is uncorrelated with the first linear function and so on (Jolliffe 1986).

These linear functions can be obtained by using the covariance matrix Σ to obtain the eigenvalue-eigenvector pairs $((\lambda_1, \mathbf{e_1}), (\lambda_2, \mathbf{e_2}).....(\lambda_n, \mathbf{e_n}))$. The set of n principal components is denoted by $\mathbf{y} = [y_1, y_2....y_n]$ where $y_i = \mathbf{e_i'}.\mathbf{x}$. Note that the variance $Var(y_i) = \lambda_i$, where λ_i are the eigenvalues of the covariance matrix Σ (Johnson & Wichern 1982). In practice the origin is translated so that the mean along each principal component axis is zero. The direction (+ or -) of each principal component is arbitrary and depends on the algorithm used in the software implementing the PCA. Useful information about relationships between the original variables or the structure of the data is often revealed by studying the eigenvectors (loadings) of the principal components.

Since most of the variance for a given system is generally contained in the first few principal components, the data points in the system can be well described with a reduced number of variables. The number of principal components necessary to include a certain amount of the variance in the system can be determined from the eigenvalues which represent the variance of each principal component. It is important to note that it is not necessarily the higher variance PCs that give best separation of classes. We must bear in mind that by omitting lower variance PCs we may be throwing away information that is potentially valuable in the classification process. To avoid losing important data we should choose sufficient PCs so that a high percentage of the variance is retained.

Geometrically, the principal components represent a rotation of axes. The technique enables the data to be represented with a reduced number of variables but preserves the spacial location of the data. In the context of our experiment, the MDS trajectory paths representing the timbre of each sound remain unaltered in n-dimensional space but the axes are rotated around the points in order to assign most of the variance between data points to the first few axes (principal components).

Since principal components depend upon the scaling of the original variables, it is not uncommon to use the correlation matrix in preference to the covariance matrix. This

effectively re-scales all variables to have a variance of unity thereby giving equal weighting to all variables. However this is not appropriate in our study for reasons related to the fact that the variables represent the set of all possible frequencies that may be present in a musical tone. The mix of frequencies present in a tone and the relative power of each frequency are the key factors which determine the timbre of a musical tone. To use the correlation matrix instead of the covariance matrix would effectively equalize the power of all harmonics present and, in so doing, seriously impair the ability of the trajectory paths to represent differences in timbre. An unwanted side effect would be that some variables contain low level noise which would be amplified to the same power level as the harmonics.

It should be noted that the purpose of principal components is to spread the data along each PC on the basis of the whole data set and does not necessarily serve to separate data on a between class basis. Since this thesis is concerned with discriminating between classes it will be prudent to consider the techniques of linear discriminant analysis where the data is projected onto one or more linear discriminant axes for the purpose of dividing the whole data set on the basis of class.

3.3.3 More on Linear Discriminants

Given the task of separating two classes, linear discriminant analysis attempts to achieve this by determining a line or function that gives maximum separation of the means. More specifically it aims to find a linear combination **ax** of the variables that maximises the ratio of the between class variance to the within class variance, that is, maximising the ratio,

$$\frac{\mathbf{a}'B\mathbf{a}}{\mathbf{a}'W\mathbf{a}}$$

where W is the within-class covariance matrix and B is the between-classes covariance matrix (Venables & Ripley 1999). In a similar way to principal components, the first linear discriminant takes the form,

$$y_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + ... + \alpha_{1n}x_n. \tag{3.21}$$

If there are g classes then, given the number of observations is sufficient (ie. no. observations $p \geqslant g-1$), $g-1$ linear discriminants can be calculated, that is, there is one less linear discriminant function than the number of classes. Although linear discriminant analysis does not assume that the data is normally distributed, problems may be encountered when there is a significant departure from normality (Krzanowski 1988). It is not uncommon to replace the data set **x** with selected principal components in order to reduce the dimensionality of the problem. This is a processing sequence that has been successfully used in the image processing domain where high dimensionality is often a problem (Swets & Weng 1996). The question of how many PCs to use as the input data for calculating the linear discriminants is a complex one since we cannot assume that separation between classes will be in the direction of the high-variance PCs. Omitting low variance PCs could result in throwing away information that is potentially valuable in the classification process.

When the problem is to assign individual observations to an appropriate class, a discriminant function can be calculated to allocate each new observation to a class. However, this is not useful in the context of our study since a whole set of observations constitutes a single tone from a musical instrument. Distance methods, which we later describe, are more suited to the classification of musical tones since they enable their time dependent nature to be taken into account.

3.3.4 More on Multi-dimensional Scaling

Multi-dimensional scaling (MDS) was originally developed for use in the social science domain (for example, Kruskal & Hill (1964)). It is often based upon data where human subjects rate n objects for similarity. In the metric form of MDS, the closeness of each pair of objects is represented in numerical form. From these ratings a dissimilarity (or proximity) matrix is generated using the mean values of the similarity ratings. In the non-metric implementation the dissimilarity matrix is based on the rank-order of the inter-object dissimilarities. The aim is to uncover any structure or pattern that may be present in the dissimilarity data. The basis of analysis is the n dimensional spatial model where each point corresponds to one of the n objects. The indicator of dissimilarity for any two objects is the distance between the corresponding two points. A number of between point distance measures are used but the most common is the Euclidean distance. For n dimensions, the distance between object i and object j is given by

$$d_{ij} = \sum_{k=1}^{n} (x_{ik} - x_{jk})^2 \tag{3.22}$$

where $x_{i1}, ..., x_{in}$ and $x_{j1}, ..., x_{jn}$ are elements of vectors $\mathbf{x_i}$ and $\mathbf{x_j}$ respectively (Everitt & Dunn 1993).

In the same way as principal component analysis seeks to reduce the dimensionality of data using the covariance matrix, the multidimensional scaling technique seeks a spacial model where the distances between points are represented as accurately as possible with as few dimensions as possible. The algorithm moves the 'objects' around in space for a given dimensionality and checks how well the distances between objects can be reproduced for a particular configuration of axes. More precisely, it uses a function minimization algorithm that uses the eigenvalues of the dissimilarity matrix to give the best configuration in terms of goodness of fit (Statsoft 2003). The process is sometimes called metric scaling or principal co-ordinate analysis (Krzanowski 1988). As with principal component analysis, multidimensional scaling can indicate structure in the data that is not observable from the raw dissimilarity data. The multidimensional scaling technique was adapted for use in analysis of musical sounds by Grey (1977) who found the technique very fruitful in his studies of musical timbre.

In general, with MDS the beginning point is a dissimilarity matrix where each of n objects is rated on a similarity basis with each other object. Except in special cases (for example,

the distances between cities on a map) the ratings cannot be exactly represented in n dimensional space. An alternative approach is where each object is defined by n attributes or variables. In this case each object can be exactly represented by a point in n dimensional space. The dissimilarity between objects can be measured by the distance (usually Euclidean) between corresponding points. To reduce the dimensionality and to explore the structure of the data and expose meaningful relationships between the variables the technique of factor analysis or principal component analysis can be applied. Each of the axes, after the transformation, is called a principal component or a factor and corresponds to the most important features contained in the data.

As we described earlier, a similar approach was used in the 1990's for analysis of musical timbre (Hourdin & Charbonneau (1997), Kaminskyj (1999)) but using physical data based on spectral analysis of each tone. In a plot of this data, each point in space represented a spectral snapshot of the sound at a particular time. The technique allowed the changes in musical sound quality to be plotted over time and the timbre of a musical tone to be represented as a sequence of points in n dimensional space - a multidimensional scaling 'trajectory path'. By using the physical properties of a tone this approach provides an objective approach to measuring timbre in contrast to the subjectivity of data based on human judgements. To reduce the dimensionality of these n-dimensional plots and uncover the structure of the timbre principal component analysis or factor analysis was applied.

The underlying premise of this approach is that the n-dimensional trajectory plot of the evolution of a tone over time represents the timbre of the tone. It follows that if the n-dimensional trajectory paths of two tones directly correspond, then the timbre of the tones is identical and visa versa. See figure 3.9 for a plot of the MDS trajectory paths using the first three PC's of two guitar tones at the same pitch from the same instrument and for a plot of the trajectories of two guitar tones from different instruments. We can easily see that the pair of paths from the same instrument correspond more closely than the pair from different instruments. If we can find a way of measuring the closeness of the paths then this will provide a means for classifying tones based on information from the whole tone.

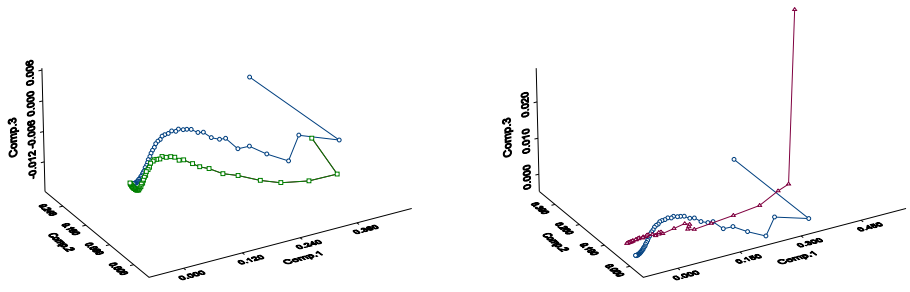

Figure 3.9: MDS trajectory paths for two guitar tones at the same pitch from the same guitar and from different guitars - tracing their evolution over time.

3.3.5 Classification of Musical Tones by Distance Methods

Given recorded tones from a set of four or five different instruments of the same type and a corresponding set of independent test tones from the same instruments, the fundamental task is to be able to identify the source of a given test tone. In order to compare the tones from a set of instruments, the data for each tone are first transformed into the frequency domain and principal components are found in order to express the MDS trajectory path in simplest terms. Recall that the spacial location of the MDS trajectory paths remain unaltered after principal components are calculated.

The method of classification is based on the premise that for two given tones, the closeness of MDS trajectory paths reflects the closeness of timbre. The most straight forward way to measure the closeness of the trajectory paths is to determine the squared Euclidean distance between corresponding points at each increment of time along the trajectory paths and then to sum these distances to give a measure of closeness for the two tones. Each trajectory path must be of the same duration (same number of data points) and should be synchronised so that the beginning point of each attack coincides.

The measure of closeness measured by the sum of Euclidean distances for corresponding points for two instruments A and B is given by (Statsoft 2003),

$$d_{AB} = \sqrt{\sum_{i=1}^{n} (x_A(t_i) - x_B(t_i))^2}. \tag{3.23}$$

Working with principal components, there are several variations possible in the way the

closeness/distance between corresponding points in the context of our study can be measured.

The sum of squared Euclidean distances can be used to place progressively greater weight on points that are at a greater distance apart. It is given by,

$$d_{AB} = \sum_{i=1}^{n} (x_A(t_i) - x_B(t_i))^2. \tag{3.24}$$

The *city-block* distance is the average distance across dimensions and gives less weight to outliers. The sum of city-block distances is given by,

$$d_{AB} = \sum_{i=1}^{n} |(x_A(t_i) - x_B(t_i))|. \tag{3.25}$$

If we wish to vary the weight placed on individual dimensions in a controlled way, we can use the *power-distance*. The sum of power-distances is given by,

$$d_{AB} = (\sum_{i=1}^{n} |x_A(t_i) - x_B(t_i)|^p)^{1/r}, \tag{3.26}$$

where r and p are user controlled parameters. Parameter p controls the weight placed on individual dimensions and r controls the weight placed on pairs of points separated by larger distances. Note that if r and p are equal to 2 then we have the Euclidean distance.

Each method described above in some way measures the sum of distances between the curves in their 'natural' space based on a given number of principal components. Recall that principal component analysis does not change the spacial location of points from the original n dimensional frequency space but merely rotates the axes. We also know that there is no certainty that the highest variance PC's will be the best discriminators between classes. For example, the feature that corresponds to the first PC may be one that is both common and consistent within classes. It is therefore prudent to attempt classification with spacial orientations modified in some way to highlight differences between classes.

We can try classification with a space defined by a number of linear discriminants. As described above, linear discriminant axes are determined on the basis of giving maximum separation between classes. One approach that we try is to use the component scores (linear

discriminants) from Fisher's linear discriminant analysis as described above. Although an essentially similar process to PCA, linear discriminant analysis chooses axes that best separate the means for each class rather than separating on the basis of the whole data set. There is some distortion of the MDS trajectories paths under LDA since the data points in this study are projected onto three or four dimensional space. ($k-1$ linear discriminant axes for k classes of data). Classification using linear discriminants would be expected to give similar results to PCA but may yield some improvement.

It may be useful to try an approach where a certain number of PCs are selected and given an equal weighting. This is achieved with a standardized multivariate normal distribution where the origin is set at the mean and the variance of each axis is adjusted to unity. The distance of each data point from the origin/mean is called the *Mahalanobis* distance and is the equivalent of a z-score in a uni-variate normal distribution.

Once we have decided on a method of measuring the closeness of timbre for two tones, then we need a means of measuring the reliability of our classification experiments where independent test tones are compared , one at a time, to a set of reference tones. The process is essentially a binomial distribution with an unknown but fixed probability p of correct classification in each trial (we are assuming that p is constant for all trials). In our experiments the proportion of trials with correct classification $\hat{p}$ gives an estimate of that probability. We can calculate confidence limits for the probability p of correct classification using the formula:

$$p = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{3.27}$$

where $\hat{p}$ is the proportion of correct classifications in n trials. For 95% confidence limits the z^* statistic is 1.960.

Chapter 4

Classification Experiments

4.1 Introduction and Data Collection

The aim in this investigation was to discriminate between the sounds of four guitars (expt.1) and five violins (expt.2). These instrument classes were chosen because of their wooden construction and the consequent variety of timbres associated with individual examples of these instruments. The goal was to devise a system that, given a test tone from one of the four guitars or five violins, enables us to determine on which instrument the tone was played. A secondary goal was to shed some light on which features of the tone contributed most to discrimination between the tones.

4.1.1 The Guitars

A set of four acoustic guitars was chosen for the experiment, each with a slightly different characteristic sound (according to the human ear). The instruments chosen were:

Guitar 1: classical guitar (Yairi),

Guitar 2: arch-top(f-hole) guitar (Maton-Alvarez),

Guitar 3: flat-top steel string guitar (Martin D28),

Guitar 4: flat-top steel string guitar (Epiphone Texan FT79).

Two sets of tones were recorded in the pitch range E2 to C5 based on the C major scale (82.4Hz to 523.2Hz). The notation is based on that of the *ASA:Psychoacoustical Terminology* (1973). The tones represent the full normal range of the instruments. Each tone was played and recorded twice in order to give independent data for use in classification. Tones were played fortissimo (loud) and a plectrum was used to produce the tones. All effort was made to keep the manner of tone production as consistent as possible, especially the position of the plectrum relative to the bridge of the guitar and the angle at which the strings were struck. During the recording, each tone was allowed to ring for at least 2 seconds and then mutted with the fingers.

4.1.2 The Violins

A set of five violins was chosen for the experiment, each with a slightly different characteristic sound (according to the human ear). The instruments chosen were:

Violin 1: cheap quality -learners violin,

Violin 2: C19th factory violin (very bright tone, sound post missing),

Violin 3: C19th German factory violin,

Violin 4: C19th French factory violin(mellow tone),

Violin 5: high quality C19th violin - *Perry of Dublin (1822).*

Two sets of tones were recorded in the pitch range G3 to G5 based on the C major scale (196.0Hz to 784.0Hz). The tones represented almost the full normal range of the instruments. Each tone was played and recorded twice with vibrato and twice without vibrato in order to give independent data for use in classification and to determine the importance of vibrato in the classification process.

Tones were were played fortissimo (loud) and a bow was used to produce the tones. All effort was made to keep the manner of tone production as consistent as possible, especially the position of the bow relative to the bridge of the violin and the angle at which the bow is held relative to the strings. During the recording each tone was bowed for at least 2 seconds.

4.1.3 Recording the Data

The recordings were made in a specially prepared 'quiet' room at Victoria University. The guitars were played by the author, Robert Moore. The violins were played by Olivia Calvert. The recording engineer was Foster Hayward. The system was as follows:

Microphones: 2x Shure KSM32-SL Studio Microphones,

Preamplifier: Behringer UltraGainPro,

Mixer: Behringer MX2642,

Recorder: SonyDTC-750 DAT.

To prepare the recordings for use in classification experiments, the tones were edited with *Cool Edit 2000* software. The data was divided such that each tone was saved as a separate file with a standardized length of 2 seconds leaving 0.01 seconds of silence before the beginning of the attack (silence is defined as power less than 38dB). The files were preserved in mono form at a sample rate of 22050Hz. In accordance with the Nyquist principle, frequencies above 11025Hz were filtered out to prevent aliasing. This allowed frequencies of up to 11025Hz to be preserved, hence including all significant higher harmonics.

4.2 Classification of Four Guitars

4.2.1 Experiment 1a: Whole tones, Same Pitch

In the first part of the experiment, tones that contained the maximum amount of data were used - whole tones containing both the attack transient and the extended decay (semi-steady-state) portions. Two sets of tones with equal fundamental frequency were compared across the range E2 to C5. In each trial a tone from one set served as the test tone and all tones in the other set served as the reference tones.

Set A - one tone at each pitch in the range from each of the four guitars.

Set B - a second independent recording at each pitch from each of the four guitars.

In order to compare the tones from each instrument, the data was first transformed into the frequency domain using either FFT or CQT in order to define the MDS trajectory path for each tone. Principal components were then calculated to reduce the dimensionality of the data and the first ten principal components (PC's) were used for the MDS trajectory paths. For all tones, the first ten PC's represented at least 98% of the variance for the data so little information was lost. The tones were then compared in pairs and a closeness of timbre measure was calculated. Note that several different measures of closeness were trialled (see appendix A) but the one that gave most satisfactory results was the Euclidean distance. Each tone in set A was compared in turn with each tone in set B to generate a distance matrix for the instrument tones at that fundamental frequency. For example, table 4.1 shows the distance measures for the four guitars at the pitch A2.

	Guitar1a	Guitar2a	Guitar3a	Guitar4a
Guitar1b	0.28	0.74	1.08	0.90
Guitar2b	0.78	0.04	1.34	1.18
Guitar3b	1.02	1.35	0.34	0.54
Guitar4b	0.85	1.15	0.61	0.14

Table 4.1: Distance matrix for two sets of tones from four guitars at pitch A2 - all tones are correctly classified.

Using the distance matrix illustrated in table 4.1, each instrument tone in set A can be compared to each tone in set B to find its closest match. Similarly, each tone in set B can

be compared to each tone in set A. The objective is to find the closest match for each of the eight independent tones from the four instruments. If the minimum value for a row or column is located in the leading diagonal, then the instrument tone is correctly classified. It follows that, if the leading diagonal contains the minimum value for each row and column then, all tones are correctly classified.

The method for pre-processing was varied to compare two methods of transforming data into the frequency domain and two different data reduction techniques.

Preprocessing by FFT followed by PCA

Using FFT for time-frequency transform and PCA (first ten PC's) for data reduction, 160 classification trials at 20 different frequencies were performed (see appendix B). The results matrix for the classifications are shown in table 4.2 below. The numbers in the leading diagonal represent correct classifications. We can see that a 97.5% correct classification rate was achieved. There was some confusion in the identification of guitar #3.

	Guitar1a	Guitar2a	Guitar3a	Guitar4a
Guitar1b	19	0	0	0
Guitar2b	0	20	0	0
Guitar3b	0	1	17	0
Guitar4b	1	0	2	20

Table 4.2: Number of matches for 160 trials at 20 different pitches according to the 'distance matrix' for two sets of tones from four guitars. This indicates a 97.5% correct classification rate.

To assess the reliability of the classification process we assume that the process follows a binomial distribution - 160 trials (8 classifications at each pitch) with a fixed but unknown probability p of a correct classification . Our experimental result of 97.5% correct classification can be considered to be an estimate of the true classification rate for this system. We can state with 95% confidence that the true classification rate is at least 95.1%.

Analysis of Physical Characteristics corresponding to each PC

In the previous chapter we saw that each of the principal components of a given multivariate system can be written in the form

$$y_i = e_{i1}x_1 + e_{i2}x_2 + e_{i3}x_3 + ... + e_{in}x_n, \tag{4.1}$$

where the coefficients are given by the eigenvectors $\mathbf{e_i}$. By examining the coefficients (loadings) from the eigenvectors and relating the x_i values to the harmonics in the instrument tone, useful information about the relationship between physical attributes of the timbre and each PC is often revealed. This is examined below.

Guitar-PC1

From figure 4.1, we can see that PC1 is a representation of the total power of the harmonics, a finding consistent with that of Hourdin & Charbonneau (1997) and Kaminskyj (1999). In particular, the first four harmonics were given significant weighting. There is some variation in weighting with pitch - at a lower pitch the second harmonic has the highest weighting but at higher pitch (eg. C4) the fundamental (first harmonic) has the highest weighting. This could be explained by the fact that the fundamental frequency tends to predominate in the spectrum at higher frequencies. We can also see that the frequencies corresponding to resonances in the body of the instruments, are given some weighting at higher fundamental frequencies (discussed further in chapter 5). Since all the FFT outputs for all tones are normalised, then PC1 essentially measures the differences in power envelopes over time. PC1 gave good results as a sole indicator of class - 86.8% correct classification. This is consistent with the findings of Kaminskyj (1999) who achieved good results with three different instruments using the power envelope as a sole indicator.

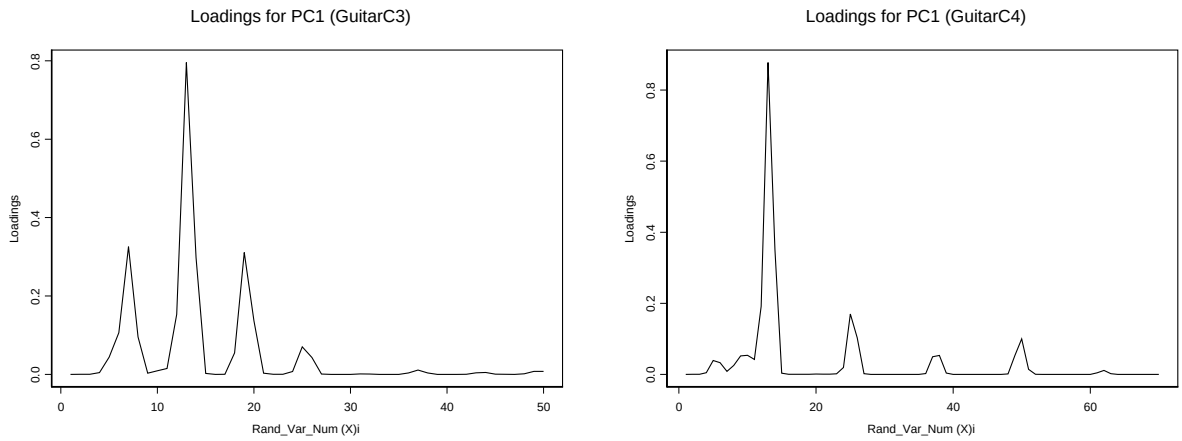

Figure 4.1: A plot of the loadings for each random variable for PC1 at C3 (fundamental at X_6) and C4 (fundamental at X_{12}).

Guitar-PC2

PC2 separates tones on the basis of the power of the fundamental compared to the sum of the power in the next few harmonics (see figure 4.2). Extremes of this PC will be a weak fundamental with other harmonics strong, and a strong fundamental with other harmonics weak. Essentially, PC2 is a measure of the spectral centroid and also relates to bandwidth. This is a similar interpretation to that of Grey (1977), Hourdin & Charbonneau (1997) and Kaminskyj (1999). The body resonances are a factor in this PC. At C3, they create a formant around the fundamental which is reflected in the heavy weighting of the fundamental. At C4, the body resonances are given a weighting in the same direction as the harmonics (discussed further in chapter 5).

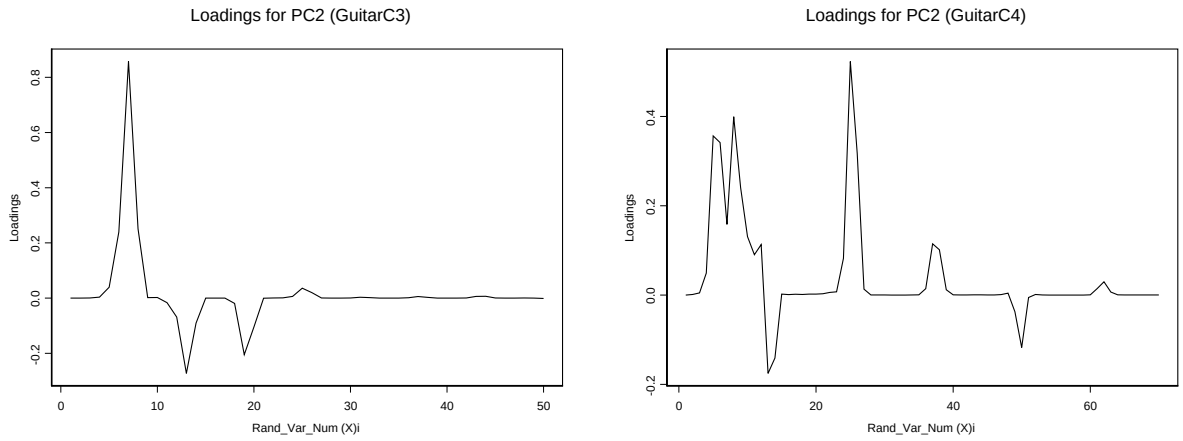

Figure 4.2: A plot of the loadings for each random variable for PC2 at C3 (fundamental at X_6) and C4 (fundamental at X_{12}).

Guitar-PC3

For PC3, there seems to be no general explanation of how the tones are separated. The loadings seem to vary with frequency. At C3, the key factor seems to be the strength of the second harmonic compared to the strength of the third whereas, at C4, the key factor seems to be the strength of the fundamental (first harmonic) compared to the second harmonic (see figure 4.3).

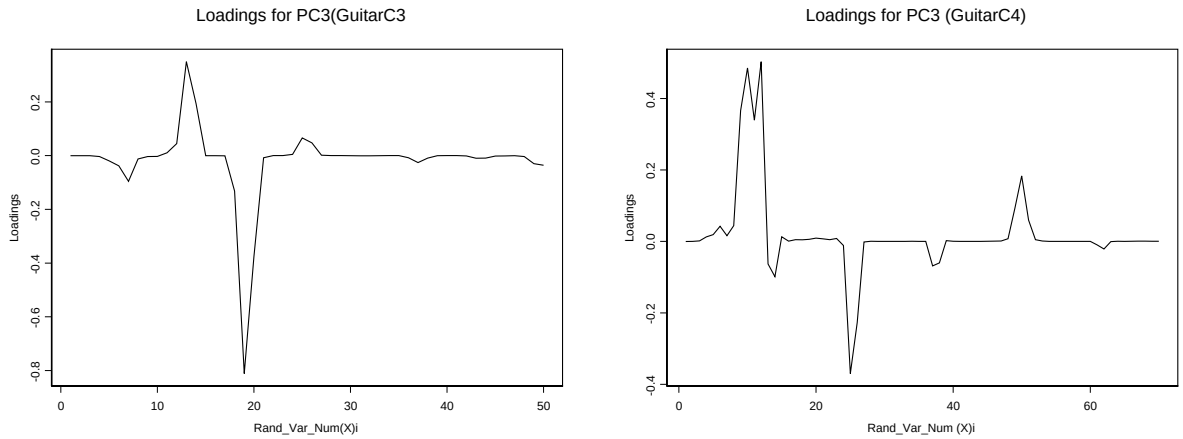

Figure 4.3: A plot of the loadings for each random variable for PC3 at C3 (fundamental at X_6) and C4 (fundamental at X_{12}).

Classifying with a Larger Reference Set

Since we actually have eight independent recordings at each pitch, the robustness of our classification system can be tested by merging the two sets A and B and classifying by withdrawing one tone at each pitch to be the test tone leaving the other seven as the reference group. This reduces the probability of a chance correct classification to $\frac{1}{7}$. Using this approach a correct classification rate of 94.4% was achieved. The true classification rate can be stated to be at least 90.8% at a 95% confidence level.

Contribution of Individual PC's to the Classification Process (FFT)

In order to determine the contribution of each principal component to the classification process based on FFT, we tried classifying the four guitars using just one principal component at a time. The table 4.3 below shows the correct classification rates.

Prin Comp No.	1	2	3	4	5	6	7	8	9	10
% Correct Class	86.8	84.3	82.5	66.2	72.5	60.0	59.4	56.9	56.0	53.1

Table 4.3: Percentage of correct matches for 160 trials at 20 different pitches according to the 'distance matrix' for two sets of tones from four guitars using just one PC from FFT data.

It can be seen that the first three principal components are the most reliable when used as the sole data source for classification. If we inspect plots of each of the the first three PC's for guitar tones at a particular pitch, in most cases we can visually identify the pairings for tones from the same guitar. For example, in figure 4.4 and figure 4.5, where each instrument is represented by 50 data points, the similarity between corresponding tones in set A and set B can easily be seen.

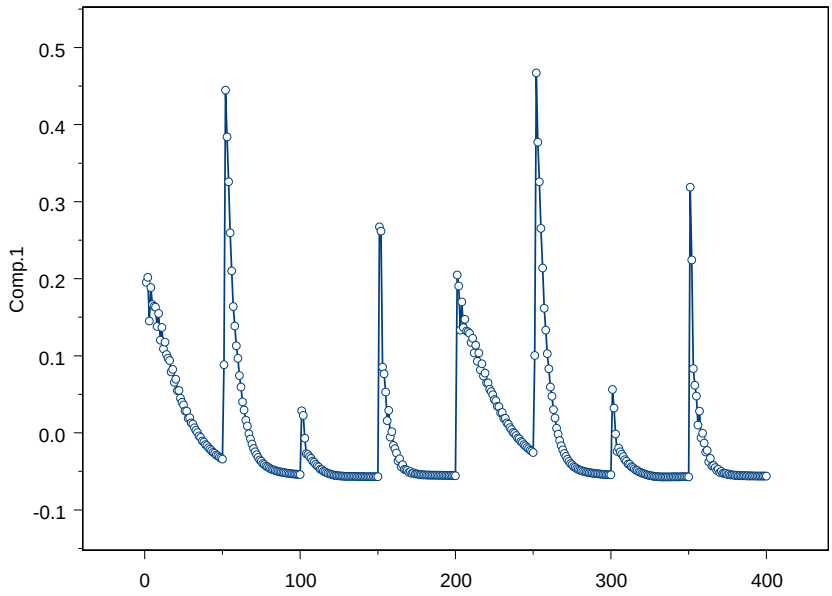

Figure 4.4: A plot of PC1 for each guitar in set A (200 pts) and set B (200 pts) tracing the evolution of a guitar tone over time.

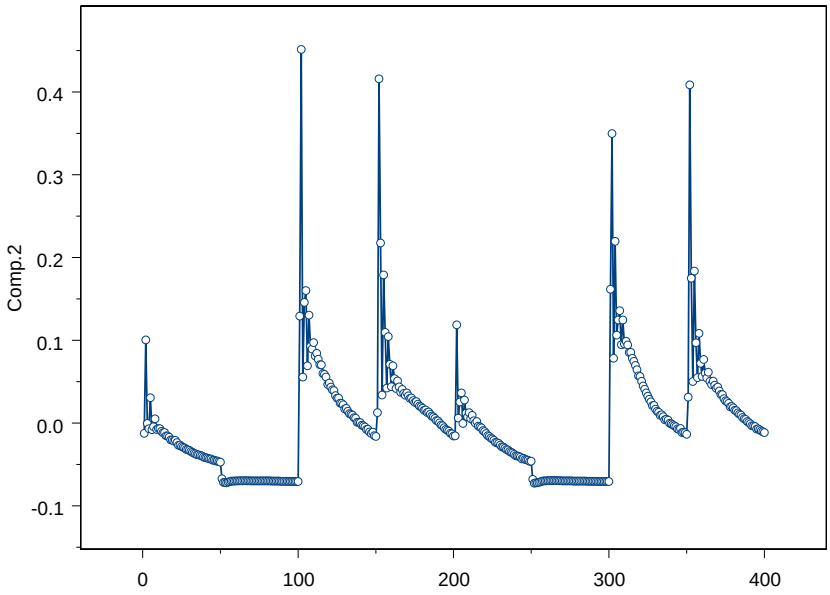

Figure 4.5: A plot of PC2 for each guitar in set A (200 pts) and set B (200 pts) tracing the evolution of a guitar tone over time.

In the light of this information, classification using the first five and the first three PC's was attempted. The correct classification rate dropped from 97.5% for ten PC's to 96.8% for five PC's and 93.1% with just three PC's. We can therefore conclude that each of the first 10 PC's makes a significant but decreasing contribution to the classification process. It should also be remembered that each PC axis is scaled according to the variance for that PC and that in principal component analysis each successive PC has a smaller variance. In this part of the experiment, ten PC's accounted for 99.9% of the variance, the first five approximately 98% and the first three approximately 95%. Hence, differences in values for the lower number PC's (eg PC1, PC2,...) have more influence over the position of a data point in Euclidean space than the higher number PC's.

Another approach to determining the number of PC's required is Kaiser's criteria (StatSciences 1993) - a criteria commonly used in principal component analysis. According to Kaiser's criteria, PC's with eigenvalues/variance less that the mean should be omitted. It should be noted that the criteria may give differing indications across the range of pitches in our data. If we consider the pitch E2, the eigen values suggest that we should use only the first three PC's and reject higher number PC's. At the pitch E3, the eigenvalues

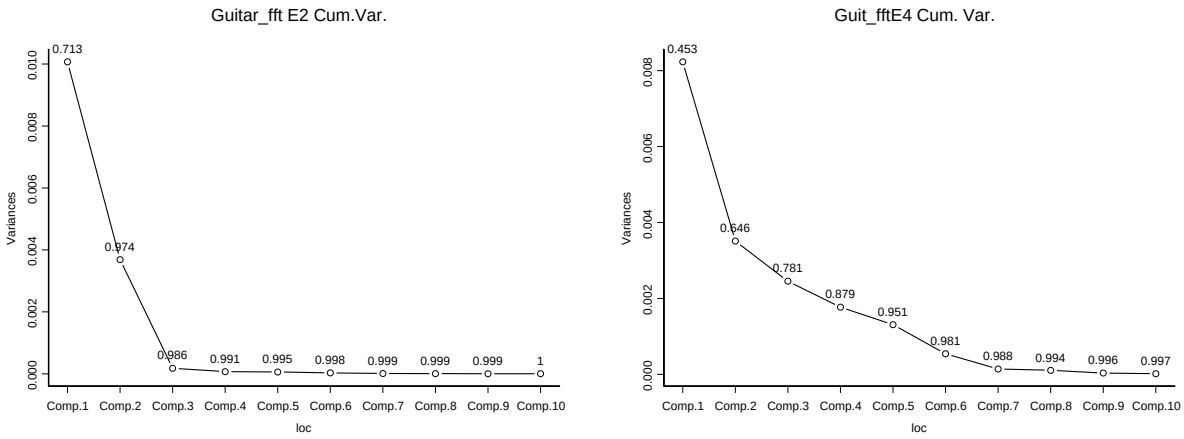

Figure 4.6: A plot of the cumulative variance from FFT pc's for guitars at E2(left) and E4(right).

suggest that we should use only the first four PC's. At the pitch E4, the eigen values suggest we use seven PC's. If we compare the cumulative percentage for three PC's at the three pitches, we get 98.6% at E2, 98.0% at E3 and 78.1% at E4 (see figure 4.7). The above data indicates that at higher pitches there is valuable information for classification in PC's four, five and six. This indicates that three PC's will give good classification results at lower pitches but reduced efficiency at higher pitches. The classification results for three and five PC's support this conclusion.

Classification based on Mahalanobis space (FFT)

In the light of the above discussion, we tried classification with an equal weighting for each PC. Each PC was normalised to achieve a unit variance. The spacial location of data points is altered under this transformation - the distance of each point from the origin after this transformation is referred to as the 'Mahalanobis' distance (Venables & Ripley 1999). We began by classifying with just one normalized principal component and then adding additional principal components one by one. We know that, in adding each additional principal component, we are increasing the information available to enable classification but at the same time the quality of the information is decreasing. It was hypothesized that the classification rate would initially improve as each successive PC was added but would then peak and subsequently decline as a large number of the higher PC's were included. Classification with this new space gave a correct classification rate of 86.8% for one PC, peaked at 95.0% for five PC's and then gradually declined as further PC's were added

until, with 100 PC's, the classification rate was 35.0% - a little better than random chance. The table 4.4 and figure 4.7 show the correct classification rates.

Number of PC's.	1	2	3	4	5	6	7	10	50	100
% Correct Class	86.8	90.6	91.3	93.8	95.0	93.8	87.5	83.1	51.3	35.0

Table 4.4: Percentage of correct matches for 160 trials at 20 different pitches using an increasing number of PC's from normalized FFT data (Mahalanobis Distance).

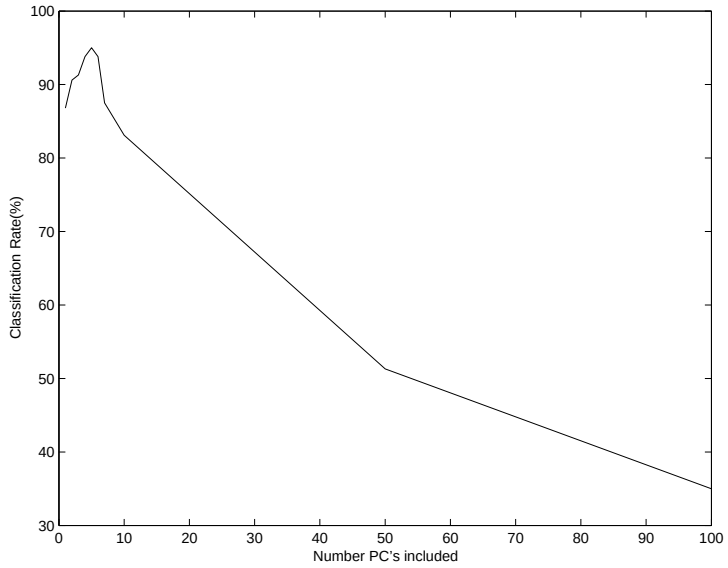

Figure 4.7: A plot of the classification rate versus the number of PC's included in Mahalanobis space -data based on FFT.

Classification based on the Mahalanobis space gave a best correct classification rate of 95.0% with five PC's. This result indicated that although the higher number PC's contribute to the classification process that best results are obtained with the lower number PC's more strongly weighted as in the Euclidean (natural) frequency space.

Preprocessing by CQT followed by PCA

In order to compare the FFT and the CQT as a means of time-frequency transform, the experiment was repeated with pre-processing by CQT. Shown below, in table 4.5, are the classification results for a system using CQT with data reduction by PCA. The correct classification rate with this system was initially 95.0%. We can state with 95% confidence that the true classification rate is at least 91.6% - a result that is marginally lower than that using FFT for the time-frequency transform. Note that there was some confusion with guitar#3 - the tone from set A was misclassified in 4 instances.

	Guitar1a	Guitar2a	Guitar3a	Guitar4a
Guitar1b	38	0	1	1
Guitar2b	0	40	1	0
Guitar3b	0	0	36	1
Guitar4b	0	0	2	40

Table 4.5: Number of matches for 160 trials at 20 different pitches according to the 'distance matrix' for two sets of tones from four guitars. This indicates a 95.0% correct classification rate.

In the next section, some variations intended to fine tune the classification process are documented and the outcomes are discussed.

Contribution of Individual PC's to the Classification Process(CQT)

In order to determine the contribution of each principal component to the classification process based on CQT, we tried classifying the four guitars using just one principal component at a time. The table 4.6 shows the correct classification rates.

Prin Comp No.	1	2	3	4	5	6	7	8	9	10
% Correct Class	84.3	76.3	75.6	68.1	49.3	64.3	55.6	53.1	48.8	57.5

Table 4.6: Percentage of correct matches for 160 trials at 20 different pitches according to the 'distance matrix' using just one PC from CQT data.

We can see that the first four principal components are the most reliable when used as the sole data source for classification. In the light of this information a few variations to the

number of principal components used for the classification process were tried. It was found that we could obtain a better result by reducing the dimensions in the frequency space. With just the first three principal components, we obtained a correct classification rate of of 96.8%. It can be stated with 95% confidence, that the true classification rate is at least 94.1%.

We can conclude that, with CQT, each of the first 10 PC's makes a significant but decreasing contribution to the classification process. The cumulative proportion of variance corresponding to 10, 5 and 3 PC's was 99%, 97.5% and 94%.

We observe that for, classification based upon FFT data, reducing the number of PC's from ten down to three resulted in a degradation in the classification performance, whereas, a similar variation in technique with CQT data resulted in an improved classification performance - it is not clear why this should be. In a subsequent section we will attempt to interpret the PCA data and link particular physical features with each principal component. This may shed some light on why the classification performance was improved with less principal components.

Classification based on Mahalanobis space (CQT)

We noted above that there is useful information for classification in the second and higher principal components but because of their smaller variance they are not highly weighted in the classification process. As with the FFT based data we will try equal weightings of the principal components using the Mahalanobis distance. We classify with just one normalized principal component and then add additional principal components one by one. It was hypothesized that the classification rate would initially improve as each successive PC was added but would then peak and then decline as a large number of the higher PC's were included. Classification with this new space gave a correct classification rate of 84.3% for one PC, peaked at 92.5% for three PC's and then gradually declined as further PC's were added until at 100 PC's the classification rate was 46.9% - still significantly better than random chance. The table 4.7 and the figure 4.8 show the correct classification rates.

Number of PC's.	1	2	3	4	5	10	50	100
% Correct Class	84.3	91.3	92.5	90.0	88.8	86.3	67.5	46.9

Table 4.7: Percentage of correct matches for 160 trials at 20 different pitches using an increasing number of PC's from normalized CQT data (Mahalanobis Distance).

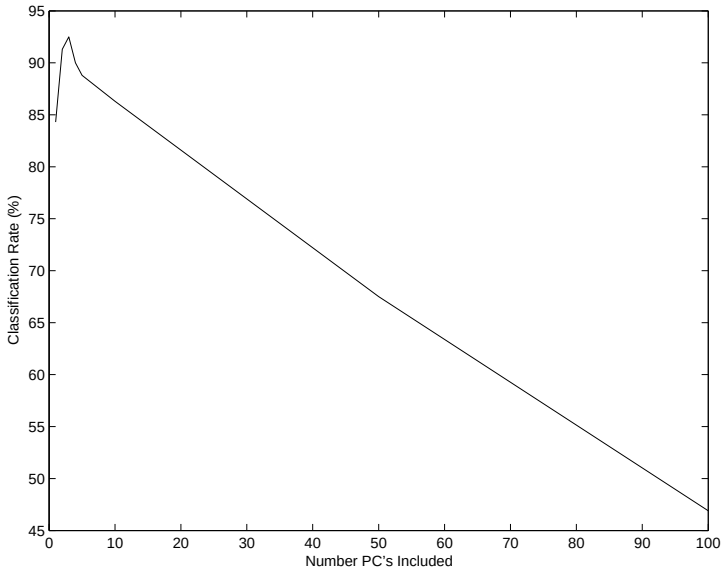

Figure 4.8: A plot of the classification rate versus the number of PC's included in Mahalanobis space - data based on CQT.

Classification based on the Mahalanobis space gave a best correct classification rate of 92.5% with three PC's. This result indicates that, for CQT data, although the higher number PC's contribute to the classification process, the best results are obtained with the lower number PC's more strongly weighted as in the Euclidean (natural) frequency space. These results show that, if the Mahalanobis distance is used for classification then, the number of PC's used is critical to the performance. We saw previously that classification based on Euclidean distance was not particularly sensitive to the number of PC's used.

Preprocessing by FFT/CQT followed by PCA with LDA

As discussed earlier, the purpose of rotation of axes in PCA is to maximise the variance along each axis, whereas the purpose of rotation of axes in LDA is to maximise the distance between means of each class. It was felt that rotation by LDA may result in improved separation of classes and hence give some improvement in the classification rate. The LDA

process was not robust enough to process the raw FFT data because of its high dimensionality (ie. 150 frequency variables). It was therefore necessary to use PCA as a data reduction tool and then use the LDA process to perform a second rotation. Shown in table 4.8 are the classification results for a system comprising time-frequency transformation by the FFT and data reduction by PCA (using 10 PC's) combined with rotation using LDA techniques. The correct classification rate obtained was a disappointing 54.3%.

	Guitar1a	Guitar2a	Guitar3a	Guitar4a
Guitar1b	20	9	5	6
Guitar2b	8	23	4	7
Guitar3b	7	8	11	11
Guitar4b	1	3	4	33

Table 4.8: Number of matches for 160 trials at 20 different pitches for two sets of tones from four guitars using three Linear Discriminants. This indicates a 54.3% correct classification rate.

This poor result may be related to the fact that the guitar is an impulsive instrument with a constantly changing power envelope and hence frequency spectrum. As a result, the MDS plots are not clustered around a mean in the frequency space as would be expected for an instrument with a pronounced steady state portion such as the violin. In terms of sound quality, since the MDS plots for a guitar are constantly changing, the mean value of points in the frequency space may not accurately represent its timbre. This suggests that separating the data on the basis of means (as in linear discriminants) may not be best for an impulsive instrument such as the guitar. The classification system was varied in an attempt to improve the performance. It was found that, by reducing the number of PC's to three, the correct classification rate was lifted to a respectable 91.9%. This is comparable to the 93.1% obtained with a single PCA rotation and just 3 PC's.

With pre-processing by CQT, PCA and then LDA, the correct classification rate was 95.0% using 3PC's - an identical result to that without LDA. Overall, the results show no improvement in the classification performance using linear discriminants.

4.2.2 Experiment 1b : Incomplete Tones, Same Pitch

In a natural situation where musical tones are often incomplete or overlapping, it would be desirable to be able to classify incomplete tones. To test this possibility, we attempted classification using only part of the tone to determine how this loss of information would affect classification performance. It was also hoped that this would shed some light on the salient features enabling classification.

Given that, in this experiment with FFT, a complete guitar tone is represented by a sequence of windows at 50 points in time, the first 10 points of the tone were pruned to exclude all information associated with the attack. The correct classification rate dropped to 89.3% indicating that valuable information for classification is contained in the attack.

To further explore the importance of information contained in the attack, we attempted to classify with incomplete tones containing only the attack and a small portion of the decay. Initially, we attempted classification with the first 10 data points and achieved a 96.8% correct classification rate. This result was somewhat higher than expected, indicating that there is enough information in the attack alone to produce good classification results. The attack was then cut to five data points and achieved 97.6% correct classification. Overall, this indicated a classification rate with a short attack nearly as high as with the complete tone. In the next chapter, we will investigate the attack in more detail in an attempt to discover the key features of the attack which allow discrimination between tones.

A number of other subsets of the data set were tried as a basis for classification, namely: progressively smaller subsets from the attack and subsets of 20 data points taken at a progressively later stage of the decay. The results are summarised below in table 4.9.

The table 4.9 shows that, given the data is synchronised with respect to the start of the tone, good classification results can be achieved with almost any portion of the tone. The best results, however, were obtained with segments from the attack. Even when the size of the subset was reduced to two data points, the classification rate exceeded that of the complete decay. We conclude that, for the guitar, the most valuable information for classification is contained in the attack but there is enough information in any portion of the tone to enable good classification. To further explore the robustness of classification with the decay, subsets of twenty data points were taken at various stages of the decay. From our review of timbre research in chapter 2, we expected that the power of the higher

Portion of Tone	% Class. Rate
whole tone	97.5
attack (10pts) t=1:10	96.5
attack (5pts)	97.6
attack (3pts)	94.3
attack (2pts)	93.7
decay (40pts,t=11:50)	86.9
decay (20pts,t=11:30)	83.8
decay (20pts,t=21:40)	87.4
decay (20pts,t=31:50)	85.6

Table 4.9: Classification rates with reduced information using a portion of the complete tone.

harmonics might diminish at a faster rate than the fundamental in guitar tones. This would result in a reduced amount of information being available for classification for subsets taken from later in the decay. A surprising result was obtained in that the classification rates did not decline for subsets taken from the later portion of the tone. In other words, similar classification rates were obtained for subsets at any portion of the decay. This suggested that we should examine the relationship of spectral centroid with time over the duration of the tone to determine empirically if there is any change in the amount of spectral data available as the tone decays. The plot is shown in figure 4.9 below.

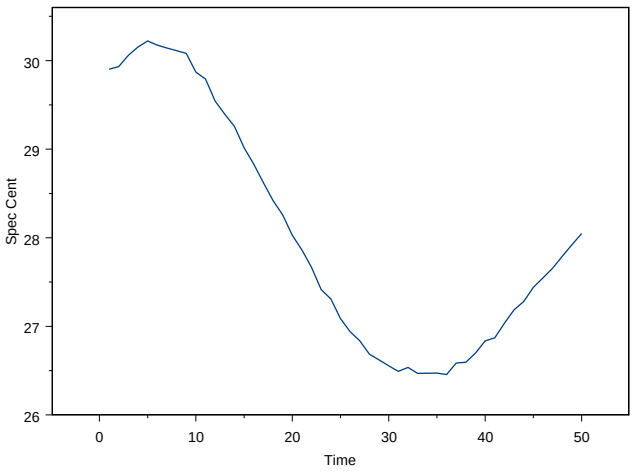

Figure 4.9: Plot of spectral centroid versus time for guitar#4 at pitch A3.

We can see that, overall, the spectral centroid decreases with time but not by a large amount. This confirms earlier findings that higher harmonics attenuate at a faster rate than the fundamental and lower harmonics. We conclude that the timbre of each guitar tone is slowly changing with time but that there is still ample spectral information available to enable classification at the same rate. The finding that the timbre of a guitar tone varies with time suggests that comparing two tones that are not synchronised with respect to the attack may present difficulties.

Since very good classification results were obtained with just the first two data points of the tone, further classification was attempted with just one data point at various points in the tone. This is essentially a single snapshot of the frequency spectrum of the tone at a particular time. The results are set out in table 4.10 below.

Tone Sample(1pt)	% Class. Rate
attack (t=1)	85.0
attack (t=2)	91.8
attack (t=5)	87.5
attack (t=10)	86.9
decay (t=20)	81.9
decay (t=30)	81.9
decay (t=40)	76.2
decay (t=50)	80.6
mean spec (20pts,t=10:30)	81.3

Table 4.10: Classification rates with data from just one window taken at different points in time across the complete tone.

We observe that relatively good classification results are obtained with just one data point. In the decay section the correct classification rate was roughly in accordance with that for the mean value (81.3%) taken over a period of 20 points. Note again that the correct classification rate does not degrade as the power of the harmonics attenuate with time (except at $t = 40$). It is interesting to note that there is a rough correlation between the classification rates in table 4.10 and spectral centroid in figure 4.9. As with the larger subsets of data points, the best classification results are obtained with points within the attack - the peak being 91.8% at $t = 2$.

4.2.3 Experiment 1c: Un-Synchronized Tones, Same Pitch

In the experiments discussed so far, we have compared tones that are synchronized with respect to time relative to the start of the attack. In real world situations it would be useful to be able to compare tones that are not synchronized, perhaps because the starting points of the tones are indeterminate. In this section we compare guitar tones (with pre-processing by FFT) which are offset by varying amounts. Table 4.11 shows the results.

Tone Portion-offset	% Class. Rate
attack(10pts,offset-5pts)	58.8
decay (20pts,offset-5pts)	73.8
decay (20pts,offset-10pts)	54.4
decay (20pts,offset-20pts)	39.4
mean spec(20pts,offset-10pts)	46.2
One point(t=10,t=20,offset-10pts)	43.7

Table 4.11: Classification rates when the two tones being compared are not synchronized with respect to the start of the attack.

We can see that the classification performance is seriously affected by offset in all cases. The attack is particularly sensitive to offset. This is perhaps explained due to the fast rate of change of the frequency spectrum in the attack. In the decay stage, where larger offsets are possible, we can see that the classification performance deteriorates as the amount of offset increases. Overall we can deduce that the physical characteristics of guitar tones vary significantly over time. This is partly explained by the fact, as we have shown above, that the spectral centroid varies over time. This indicates that the higher frequency harmonics attenuate at a faster rate than the fundamental and other lower frequency harmonics. A second factor in the reduced classification performance may be the difference in mean power of offset tones. Since guitar tones are impulsive with no steady state, the mean power is constantly changing and so the location of points in the frequency space is strongly time dependent. Perhaps the effects of non-synchronisation could be reduced by performing a 'normalisation' on the segments of each tone.

4.2.4 Experiment 1d: Whole Tones, Pitch a Step Apart

An important question to answer is the effect of pitch upon classification, that is, what happens when tones being compared are of a different fundamental frequency. This has implications for the question of whether timbre is frequency dependent (see chapter 2).

To answer this question, we attempted classification with reference tones that were one step different in pitch to the test tones. With the pitch of tones based on the C major scale this meant each pair of tones differed by a tone or a semi-tone.

Firstly, tones with pre-processing by FFT followed by PCA were classified. It was necessary to compensate for the fact that under FFT the bin location of each harmonic varies with change in the fundamental frequency. The tones were first separated into groups of common fundamental frequency and then PCA was performed separately on each group of tones. This problem does not arise with CQT since the fundamental can easily be assigned to a fixed location. The classification results after FFT and PCA are shown in table 4.12 below.

	Guitar1a	Guitar2a	Guitar3a	Guitar4a
Guitar1b	24	7	3	7
Guitar2b	2	25	10	1
Guitar3b	2	5	17	12
Guitar4b	7	5	6	19

Table 4.12: Number of matches for 152 trials at 20 different pitches according to the 'distance matrix' for two sets of tones from four guitars where set A and set B are a step apart in pitch (FFT). This indicates a 56% correct classification rate.

We repeated the experiment using CQT and PCA with results as follows in table 4.13.

	Guitar1a	Guitar2a	Guitar3a	Guitar4a
Guitar1b	20	13	1	3
Guitar2b	2	25	3	3
Guitar3b	4	3	23	13
Guitar4b	12	3	6	18

Table 4.13: Number of matches for 152 trials at 20 different pitches according to the 'distance matrix' for two sets of tones from four guitars (CQT). This indicates a 56.6% correct classification rate.

With a correct classification result of about 56% for both the FFT and CQT based data, it is clear that classification is very sensitive to fundamental frequency. The results suggest that the timbre of a guitar varies markedly with change in pitch/fundamental frequency, that is, although the tones are all identifiable to the human ear as from a guitar, the timbre varies significantly with change of pitch. We conclude that the timbre of a guitar is, in fact, a set of different pitch specific timbres that together make up its sound quality. The pitch specific nature of guitar tones, as well as a possible explanation in terms of formants, will be discussed in the next chapter.

To examine further the pitch specific nature of guitar timbre, we investigated the relationship between spectral centroid and frequency - the spectral centroid being the best single indicator of timbre over a short interval of time. The following graph, shown in figure 4.10, confirms that the timbre of each instrument varies in an irregular fashion as the fundamental frequency (pitch) changes.

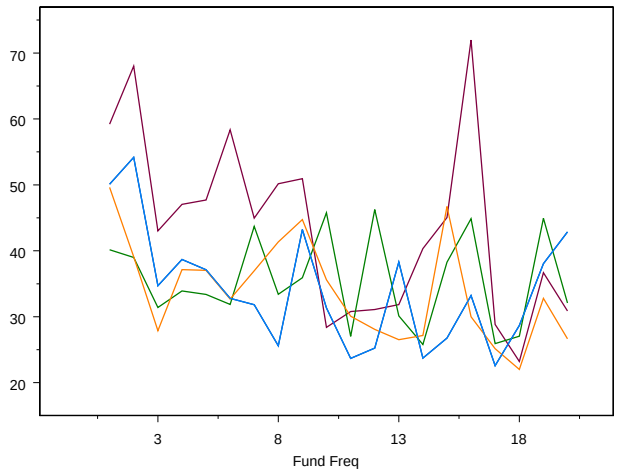

Figure 4.10: The spectral centroid for four guitars across the frequency range.

Another question that can be answered by examination of the spectral centroid, is how consistent is the tone at a given pitch from a given instrument. In other words, how well can the timbre of a tone from a given instrument at a certain pitch be reproduced over a number of performances. Our classification results, to date, would indicate that pairs of independent recordings from the same instrument, at the same pitch, are close in timbre and this can be confirmed by examination of the spectral centroid.

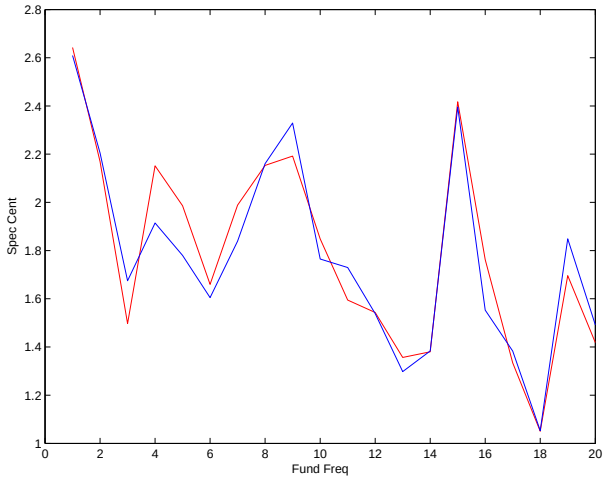

Figure 4.11: Spectral centroid for two independent sets of recordings for guitar#4 across its frequency range.

In figure 4.11 we see that the irregular plot of spectral centroid versus frequency is highly reproducible with independent recordings. These findings confirm that the timbre of each guitar is, in fact, a set of different pitch specific timbres that together make up the sound and that these timbres will be consistent over numerous performances.

4.2.5 Discussion and Conclusions for Experiment1: Classification of Four Guitars

The fundamental task in this experiment was to represent the timbral qualities of the four guitars in a quantitative form in order to classify given test tones from a set of reference tones. Using MDS trajectory paths as a representation of the timbre of each tone we achieved results which exceeded our expectations. The correct classification rate was 97.5% with FFT and 95.1% with CQT. It appears from these results that the trajectory paths in frequency space offer a very good representation of the timbre or characteristic 'signature' of each instrument across its pitch range. From examination of the PC loadings, it appears that features enabling classification are the shape of the power envelope, spectral features related to the spectral centroid and the body resonances present in the attack portion of guitar tones. To further improve the process, a number of variations were tried. Secondary goals were to attempt to uncover some secrets of guitar timbre.

To compare the performance of FFT and CQT as techniques for time-frequency transform, some variations in the number of principle components were considered. We found that, when the number of PC's was reduced to three, the performance of the FFT data reduced from 97.5% to 93.1% whereas the performance of the CQT data improved from 95.1% to 96.8%. One possible explanation for this could be that the CQT data is stored in a more compact way due to the logarithmic nature of the frequency scaling. This means there is more detailed information about the lower harmonics and less about the higher harmonics which are bunched together with CQT. Overall, we can say that the results with FFT were comparable to those with CQT.

In a variation to the standard use of principal components (weighted according to variance), the Mahalanobis distance was used where all PC's are given equal weighting. In this case, as we add more PC's, we add more information but of an increasingly unreliable nature and all of equal weighting. Our test involved beginning with one PC as a basis for classification and one by one adding more PC's. We found that, initially, the classification performance increased, peaking at 95.0% with five PC's for FFT and 92.5% with three PC's for CQT and then gradually decreasing until at 100 PC's the performance was little better than random selection. So overall, the variation produced slightly inferior results but with an increased level of complexity in the process. It is interesting to note that, with FFT data, there was useful information that enhanced classification in PC's four and five, whereas

the classification with CQT declined when PC's four and five were included. This suggests that, for CQT, there is less salient information contained in the higher numbered PC's. This may be related to the logarithmic nature of the CQT frequency scale which has a clustering effect on the higher harmonics in the spectrum, meaning that the data can be effectively represented by fewer PC's.

A second rotation of the axes was attempted using LDA after PCA - the data was represented as three linear discriminants. For guitar tones, this resulted in a significant reduction in the classification rate compared to that obtained with PCA only. A possible explanation for this is related to the percussive/impulsive nature of guitar tones. As a consequence of this impulsive nature, the power and hence the position in frequency space for each window of data is constantly changing with time. This means that there is no cluster of points in frequency space and hence the mean spectrum will be a poor representation of a guitar timbre. Since the LDA works on the basis of choosing linear discriminant axes that give maximum separation between the interclass means, it is not surprising that it is less effective than the PCA as a means of classifying guitar tones.

Further variations were attempted to test the robustness of the classification process. In a natural situation it may not always be possible to compare complete and synchronised tones. When partial (but still synchronised) guitar tones were offered for classification, some interesting results were obtained. With a portion of the attack, we achieved a correct classification rate of 97.6%, an almost identical result to that with a whole tone, and with a large portion from the decay, we obtained an 87% correct classification rate. These results indicate that valuable information for classification is contained in both the attack and the decay but also, that the attack is a more efficient classifier than the decay. These results support the findings of Clark et al. (1963) who found that, for a full set of orchestra instruments, both the attack and steady-state were important aspects of timbre. These conclusions differ from much earlier findings of Helmholtz (1863) and his followers, who concluded that timbre was dependent on the steady state spectrum. A significant difference between our work and these earlier works on timbre is that their timbral studies depended on a human listening panel whereas our work depends on machine recognition. Using a human listening panel means the limitations or characteristics of the human listening process must be taken into consideration.

Interestingly, we found that good results could be obtained with only a small portion of a tone. Good results (around 80% correct) could be obtained from a single window taken

from the decay (approximate steady-state for guitar). Even better results (around 90% correct) were obtained from a single window from the attack. This indicates that guitar tones are highly reproducible and, given that tones can be synchronised, classification is possible with a small portion of a tone.

Due to the impulsive nature of guitar tones and the fast rate of change of the spectrum over time, it was expected that classification rates would be severely affected when offset (unsynchronised) tones were compared. This prediction is supported by the plots of spectral centroid over time, shown earlier (figure 4.9), which illustrate the variable nature of the spectral centroid and consequently the spectral envelope. The graph suggests that the attack would be the portion of a tone that would be most sensitive to imperfect synchronisation. With an offset of just 5 data points or about 0.1 sec., the classification for a section of data including the attack, dropped to about 59% and for a section from the decay (approximately steady state), to about 74% - confirming our expectation. We found that further increase in the offset produced an even larger drop in the classification rate. The most significant factors that cause this sensitivity to offset are the variation over time in both the power envelope and the spectral centroid. These are problems that would have to be addressed in a real word recognition process. One possible solution could be to normalise the power in the extended decay section of a guitar tone. While this may produce a better classification result with incomplete tones, it is tampering with the timbre of the guitar tones in a major way - the power envelope is a significant attribute of the timbre of a guitar tone. Another possible solution to the problem is , given a pair of guitar tones for comparison, to match the pair of tones by essentially sliding one power envelope across the other until we have achieved the best match or, in other words, we have obtained a minimum value for the 'distance' measure.

Another important issue in comparing guitar tones is the importance of the tones being of the same pitch or fundamental frequency. When tones just one step different in pitch were compared, the classification rate dropped to about 56%. This suggested that the timbre of guitar tones is highly dependent on fundamental frequency. To further examine this question, the spectral centroid was plotted against frequency for several instruments. These graphs indicated that the average spectrum, and hence the timbre, is highly dependent on fundamental frequency and, in fact, the timbre for a guitar is a set of different but frequency related timbres (see figure 4.11). This finding suggests a possible area for improvement in other instrument recognition works (eg. Martin (1999), Kaminskyj (1999)) where pairs of

tones of similar but not equal fundamental frequency were compared.

In summary, we have shown that it is possible to obtain extremely high classification results for a set of guitars in a controlled experiment but that the classification performance depends on a few factors which would be very difficult to control under natural circumstances (eg. fundamental frequency and synchronisation of tones). We have also learnt that, for the guitar, there are important cues for classification in both the attack and the steady-state/decay and that the information in the steady-state/decay is more robust. The experiment has highlighted aspects of guitar timbre not previously taken into account in instrument classification. In particular, that the timbre of a particular tone varies significantly with time; that overall, guitar timbre is highly dependent on fundamental frequency; and that the timbre of a given guitar at a particular frequency is highly reproducible.

4.3 Classification of Five Violins

4.3.1 Experiment 2a : Whole tones, Same Pitch

Recall that the aim in this second experiment is to discriminate between the sound of five violins. Initially, test tones and reference tones that contained the maximum amount of data were used, that is, whole tones containing both the attack transient and the steady-state portions. Two sets of tones with equal fundamental frequency were compared across the range G3 to G5. In each trial, a tone from one set served as the test tone and all tones in the other set served as the reference tones.

Set A - one tone at each pitch in the range from each of the five violins.

Set B - a second independent recording at each pitch from each of the five violins.

In order to compare the tones, the data for each tone was first transformed into the frequency domain and then data reduction by PCA was carried out. The tones were then compared in pairs and a closeness/distance measure was calculated. Each tone in set A was compared with each tone in set B to generate a distance matrix for the instrument tones at that fundamental frequency. For example, the following table 4.14 shows the distance matrix for the five violins at the pitch A4 with no vibrato and preprocessing by FFT. If the minimum value for a row or column is located in the leading diagonal then the instrument tone is correctly classified.

	Violin1a	Violin2a	Violin3a	Violin4a	Violin5a
Violin1b	0.62	0.84	0.78	0.88	1.10
Violin2b	0.52	0.36	0.57	0.65	0.81
Violin3b	0.21	0.68	0.25	0.46	1.11
Violin4b	0.29	0.69	0.37	0.19	1.03
Violin5b	1.02	0.81	1.06	0.98	0.32

Table 4.14: Distance matrix for two sets of tones from five violins at pitch A4 - 8 out of 10 tones are correctly classified.

In this experiment, the preprocessing was varied to compare FFT and CQT as methods of transforming data into the frequency domain and two different data reduction techniques were tried.

Preprocessing by FFT followed by PCA

Classification for five violins, with tones played without vibrato using FFT for time-frequency transformation and PCA for data reduction, was attempted. The results are summarised in table 4.15 below. We can see that a 77.3% correct classification rate was achieved. There was some confusion in the identification of violin #1. We can state with 95% confidence that the true classification rate is at least 70.8%.

	Violin1a	Violin2a	Violin3a	Violin4a	Violin5a
Violin1b	19	1	3	1	5
Violin2b	4	24	2	0	2
Violin3b	7	0	22	2	4
Violin4b	1	0	1	27	0
Violin5b	1	0	0	0	24

Table 4.15: Number of matches for 150 trials at 15 different pitches according to the 'distance matrix' for two sets of tones from five violins using FFT. This indicates a 77.3% correct classification rate.

We repeated the classification with violin tones played with vibrato and no significant difference in correct classification rates was noticed - a correct classification rate of 75.3% was achieved.

Contribution of Individual PC's to the Classification Process (FFT)

In order to determine the contribution of each principal component to the classification process under FFT, we tried classifying the five violins using just one principal component at a time. The table 4.16 below shows the correct classification rates.

Prin Comp No.	1	2	3	4	5
% Correct Class	61.3	62.6	56.7	41.3	40.6

Table 4.16: Percentage of correct matches for 150 trials at 15 different pitches according to the 'distance matrix' for two sets of tones from five violins using just one PC from FFT data.

It can be seen that the first three PC's are the most reliable when used as the sole data

source for classification. PC's higher than three resulted in a poor classification performance. If we inspect plots for any of the the first three PC's for the full set of violin tones at a particular pitch, in most cases we can identify the pairings for tones played by the same violin (figure 4.12).

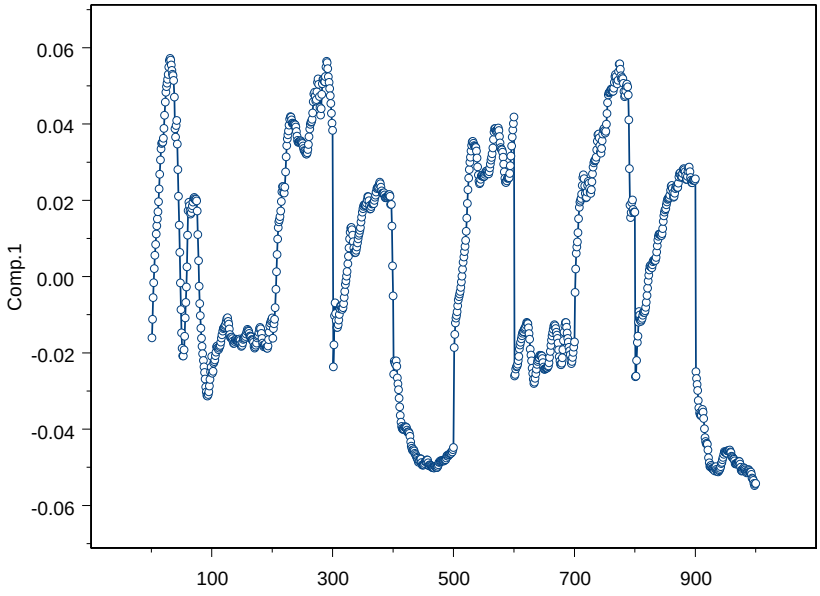

Figure 4.12: A plot of PC2 for each violin in set A and set B tracing the evolution of tones over time.

Since the first three PC's seem the most significant, classification with only three PC's was attempted yielding a classification rate of 72.7% compared to 77.3% for ten PC's. We can conclude, therefore, that each of the first 10 PC's make a decreasing contribution to the classification process. In this version of the experiment, ten PC's accounted for 99.9% of the variance, the first five approximately 98% and the first three approximately 95%. Hence, differences in values for the lower numbered PC's have more influence over the position of a data point in frequency space than the higher PC's.

Analysis of Physical Characteristics corresponding to each PC

Earlier in the chapter (equation 4.1), we showed that each principal component could be written in terms of the eigen values $\mathbf{e_i}$. By examining the coefficients (loadings) from the eigenvectors and relating the x_i values to the harmonics of the instrument tone, useful information about the relationship between physical attributes of the timbre and each PC is often revealed. This is examined below.

Violin-PC1

As with the guitar, the figure 4.13 shows that PC1 is a representation of the total power of the fundamental (first harmonic) and the other harmonics. In particular, the fundamental was given significant weighting. There is some variation with pitch - at a lower pitch the fundamental has the highest weighting with a reduced, but significant, weighting for the higher harmonics but at higher pitch (eg. D5) the fundamental is highly weighted and a very low weighting is given to the other harmonics. This could be explained by the fact that the fundamental frequency tends to predominate in the spectrum at higher frequencies. Since all the FFT outputs for all tones are normalised, then PC1 essentially measures the differences in the power of the fundamental frequency over time. As a sole indicator of of class, PC1 gave a 61.3% correct classification rate.

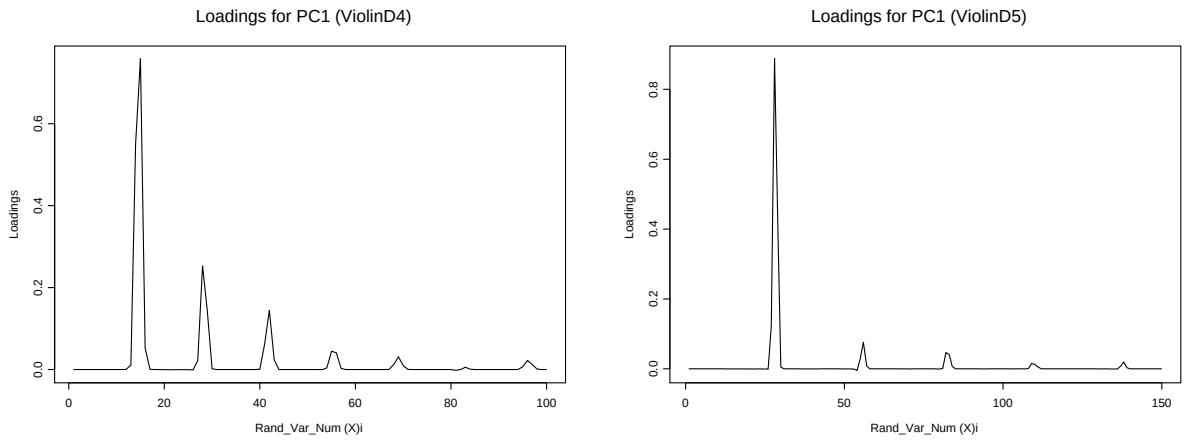

Figure 4.13: A plot of the loadings for each random variable for PC1 at D4 (fundamental at X_{14}) and D5 (fundamental at X_{28}).

Violin-PC2

The second principal component appears to separate tones on the basis of the power of the fundamental (first harmonic) compared to the power in the second harmonic (see figure 4.14). Extremes of this PC will be a weak fundamental with a strong second harmonic and a strong fundamental with a weak first harmonic. PC2 is related to the spectral centroid.

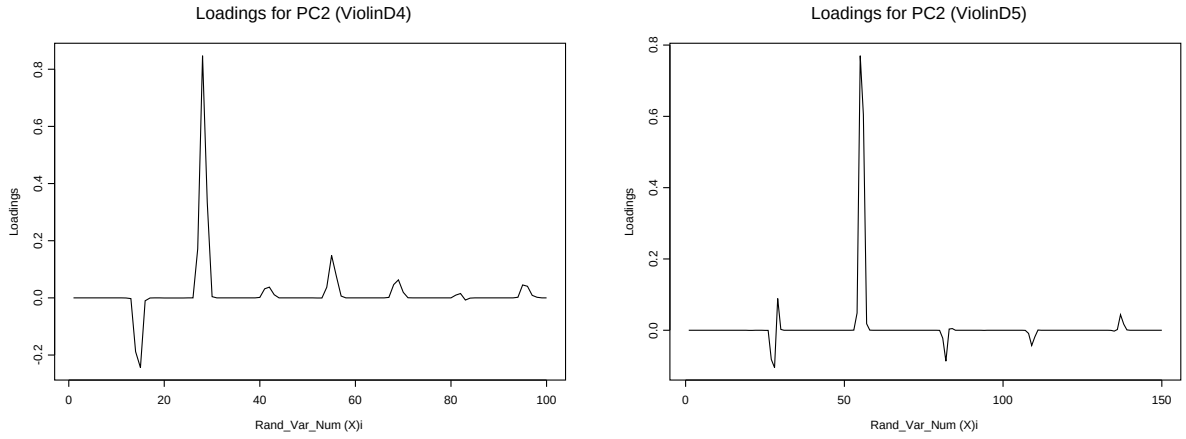

Figure 4.14: A plot of the loadings for each random variable for PC2 at D4 (fundamental at X_{14}) and D5 (fundamental at X_{28}).

Violin-PC3

PC3 appears to separate tones on the basis of the power of the fundamental compared to the power of the second and third harmonics (see figure 4.15). Extremes of this PC will be a weak fundamental with strong second and third harmonics, and a strong fundamental with weak second and third harmonics. PC3 is also related to the spectral centroid.

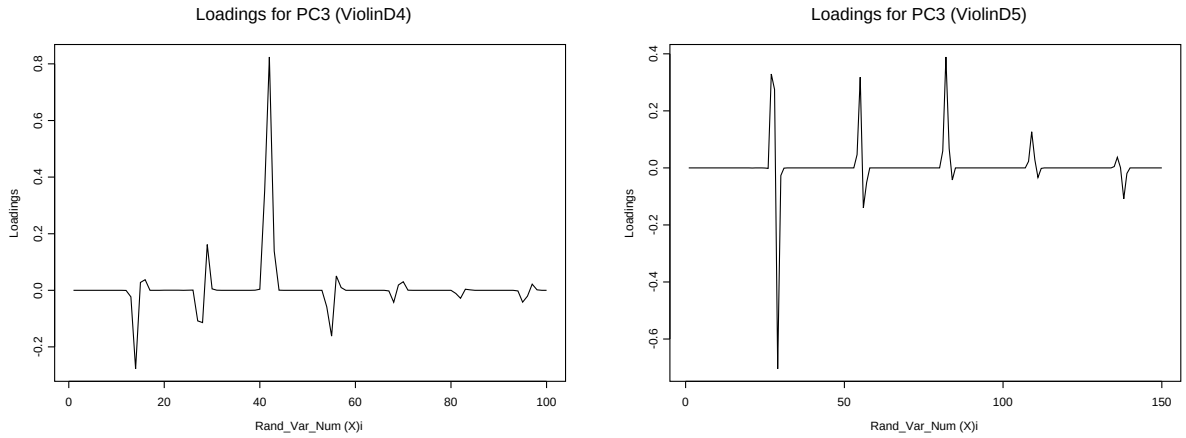

Figure 4.15: A plot of the loadings for each random variable for PC3 at D4 (fundamental at X_{14}) and D5 (fundamental at X_{28}).

Violin-PC4

With PC4, there seems to be no general explanation of how the tones are separated. The loadings seem to vary with frequency. At D4, the key factor seems to be the strength of the second harmonic compared to the sum of the higher harmonics, whereas, at D5, the key factor seems to be the strength of the third and higher harmonics (see figure 4.16). An interesting factor for PC4 is the loadings in the region of the fundamental, which are opposite in sign. This would make PC4 sensitive to differences in pitch between the tones.

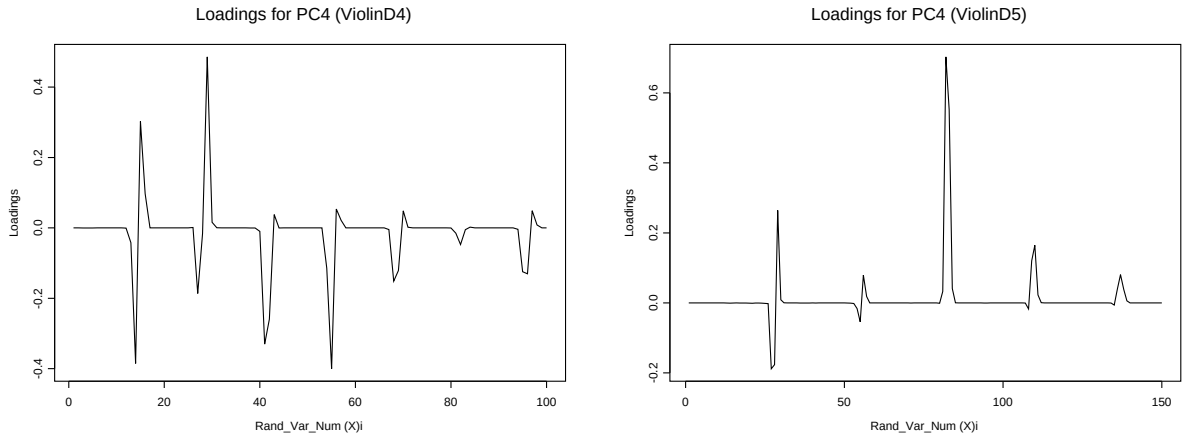

Figure 4.16: A plot of the loadings for each random variable for PC4 at D4 (fundamental at X_{14}) and D5 (fundamental at X_{28}).

Preprocessing by CQT followed by PCA

In order to compare classification based on the FFT frequency data with classification based on CQT data, the experiment was repeated with pre-processing by CQT. Shown below, in table 4.17, are the classification results for a system with time-frequency transform by CQT and data reduction by PCA. For violin tones with no vibrato, the correct classification rate was 92.7%. We can state with 95% confidence, that the true classification rate is at least 88.5%. - a result that is significantly higher than that using FFT for the time-frequency transform.

	Violin1a	Violin2a	Violin3a	Violin4a	Violin5a
Violin1b	26	1	1	1	1
Violin2b	0	26	0	1	1
Violin3b	2	0	30	0	1
Violin4b	1	0	0	28	0
Violin5b	1	0	0	0	29

Table 4.17: Number of matches for 150 trials at 15 different pitches according to the 'distance matrix' for two sets of tones from five violins(no vibrato)using CQT. This indicates a 92.7% correct classification rate.

We repeated the classification with violin tones played with vibrato and a slightly reduced correct classification rate of 87.3% was achieved. An explanation for the effects of vibrato could be that instruments with a continuous tone, such as the violin, tend to have frequency data that is tightly clustered in frequency space and the fluctuations in frequency that characterise vibrato would tend to increase the size of the cluster and thereby increase the variance. Given this result, and in the light of the previous results with FFT, we can conclude that classification with CQT is affected more by the presence of vibrato than classification with FFT. It is unclear why this should be so and warrants further investigation.

In the next section, some variations to the process are outlined and trialled and the outcomes discussed in an attempt to improve the classification rate and to explain why pre-processing by CQT produced better classification rates than pre-processing by FFT.

Contribution of Individual PC's to the Classification Process(CQT)

In order to determine the contribution of each principle component to the classification process using CQT and PCA, we tried classifying the five violins using just one principle component at a time. The table 4.18 shows the correct classification rates.

Prin Comp No.	1	2	3	4	5	6	7	8	9	10
% Correct Class	74.7	58.0	64.0	51.3	46.0	40.7	34.0	32.0	26.7	30.0

Table 4.18: Percentage of correct matches for 150 trials at 15 different pitches according to the 'closeness matrix' using just one PC from CQT data.

It can be seen that the first four principle components are the most reliable when used as the sole data source for classification, each producing a better than 50% correct classification rate . In the light of this information, a few variations to the number of principal components used for the classification process were tried. However, in contrast to the guitar experiments, we found that reducing the dimensions in the frequency space produced a slight decrease in classification performance. With the first five principal components, we obtained a correct classification rate of of 91.3% and with three principal components a correct classification rate of of 86.6% was achieved.

We can conclude that, with CQT, each of the first 10 PC's makes a significant but decreasing contribution to the classification process. The cumulative proportion of variance corresponding to 10, 5 and 3 PC's was approximately 99%, 93% and 80% respectively. An interesting fact observed here is that the proportion of variance for three and five principal components is frequency dependent - the proportion is higher in the lower range for the violin. For example, the proportion of variance for three principal components is 98% at G3, 80% at G4 and 77% at G5. This indicates that at higher fundamental frequencies there is an increased amount of information to assist classification in the higher number principal components.

Classification based on Mahalanobis space (CQT)

In the above discussion we noted that there is useful information for classification in the second and higher number principal components but because of their smaller variance they are not highly weighted in the classification process. As before, we tried classification using the Mahalanobis distance with all PC's equally weighted. We began by classifying with just one PC and then adding additional PC one by one. Recall that we hypothesized that the classification rate would initially improve as each successive PC was added and would then peak and then decline as a larger number of the higher PC's were included. Classification gave a correct classification rate of 74.7% for one PC, peaked at 92.0% for four PC's and then gradually declined as further PC's were added until at 100 PC's the classification rate was 28% - marginally better than random chance. The table 4.19 shows the correct classification rates.

Number of PC's.	1	2	3	4	5	6	7	10	20	50	100
% Correct Class	74.7	84.0	84.7	92.0	90.7	84.7	80.7	73.3	63.3	38.7	28.0

Table 4.19: Percentage of correct matches for 150 trials at 15 different pitches using n PC's from normalized CQT data (Mahalanobis Distance).

Preprocessing by FFT/CQT followed by PCA with LDA

It was felt that, after time-frequency transformation by the FFT and data reduction by PCA (using 10 PC's), rotation by linear discriminant analysis may result in improved separation of classes and hence give some improvement in classification rate. The correct classification rate obtained with this system was a 76.6% compared to 77.3% without LDA. With pre-processing by CQT, then PCA and LDA, the correct classification rate was 94.0% compared to 92.7% without LDA. Neither result shows any significant change in the classification performance using linear discriminants.

4.3.2 Experiment 2b: Incomplete Tones, Same Pitch

As with guitar tones, in a natural situation violin tones are often incomplete or overlapping. To simulate this situation we attempted classification of violins using incomplete tones to determine how this loss of information would effect classification performance. All tones were without vibrato with preprocessing by CQT.

Firstly, we investigated classification of violin tones with the attack portion removed leaving only the steady-state portion of the tone. Given that for preprocessing by CQT, a complete violin tone is represented by 100 sample points, the first 20 points of each tone were pruned to exclude all information associated with the attack. The correct classification rate increased from 92.7% to 94.0%, indicating that information contained in the attack is not essential to and, in fact, may hinder the classification of violins. When a small subset of only 10 points from the steady state was taken from violin tones, the classification rate dropped to 82.0%, indicating that perhaps a larger slice of the steady state is necessary to account for variation in the tone.

Investigating further, we then attempted to classify with incomplete tones containing only the attack and with a small portion of the steady state. Using the first 20 data points, a 73.3% correct classification rate was achieved. This result indicated that the information in the attack alone does not produce good classification results and is consistent with our finding that pruning the attack from a violin tone actually improves the classification rate.

4.3.3 Experiment 2c: Unsynchronized Tones, Same Pitch

If, in a real world performance situation, we have incomplete or overlapping tones, then achieving sychronisation of tones would be a difficulty. We therefore tested to determine what effect offset would have on the classification of violin tones. The significant finding was that off-setting the test tones with respect to the reference tones (10 points relative to the start of the tone) produced a correct classification rate of 92.7% - the same rate at whole tones in perfect synchronization. We conclude that offset tones have no effect on the classification of violin tones. This finding can be explained by the fact that, since violin tones are predominantly steady state and power is relatively constant, the MDS points in frequency space for violin tones tend to be tightly clustered together. Further, timbre for a violin tone does not vary as a function of time in the way that guitar tones and other percussive instruments do. We can verify this by examining the change in spectral centroid over time as shown in figure 4.17. We see that, if we exclude the attack and decay, the spectral centroid is relatively constant, showing random variation due to the bowing. In the light of the above discussion, we can conjecture that comparing the means of spectral data from the steady state of violin tones would give similar classification results. To explore this further, classification by comparing Euclidian distances between spectral means in frequency space for the two sets of violin tones was attempted. This yielded a 94.0% classification rate - virtually identical the 94.6% classification rate achieved for steady state tones with the attack removed.

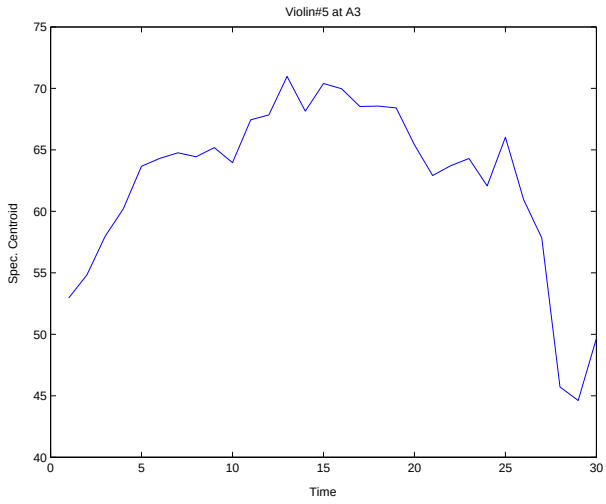

Figure 4.17: Spectral centroid versus time for violin#5 at pitch A3.

4.3.4 Experiment 2d: Whole Tones, Pitch a Step Apart

As with classification of guitars, an important question to investigate is the effect of pitch upon classification of violin tones, that is, what happens when tones being compared are of a different fundamental frequency. In other words, is the timbre of violin tones frequency dependent?

To investigate this issue we attempted to classify test tones by comparing them with reference tones that were of a different pitch (fundamental frequency). Test tones were compared to reference tones one step different in pitch, based on the C major scale, in other words, with pitch differing by a tone or a semi-tone.

Tones with pre-processing by CQT were classified and the correct classification rate was found to be 26.4% - approximately that of random chance. It is clear that classification of violin tones is very sensitive to changes in fundamental frequency. The results suggest that the timbre of a violin varies markedly with change in pitch/fundamental frequency. We conclude that, as with the guitar, the timbre of a violin is in fact a set of different pitch specific timbres that together make up the sound of that violin.

To verify this finding we investigated the relationship between spectral centroid and frequency - the spectral centroid being the best single indicator of sound quality over a short interval of time. The graph in figure 4.18 confirms that the timbre of each instrument varies in an irregular fashion over the pitch range of the violin. The reasons for this variation in timbre with frequency will be discussed in the chapter 5.

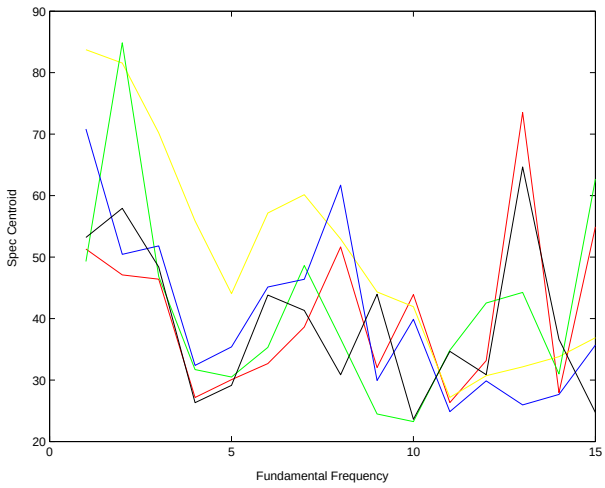

Figure 4.18: Spectral centroid for five violins across the frequency range.

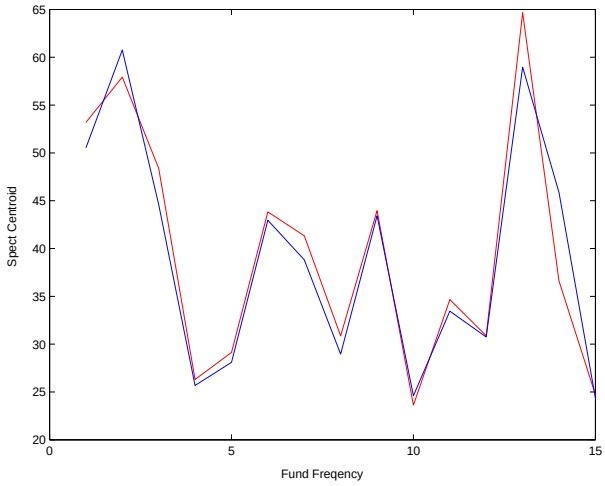

Figure 4.19: Spectral centroid for two independent sets of recordings for violin#1 across its frequency range.

Another question that can be answered by examination of the spectral centroid is how consistent is the tone at a given pitch from a given instrument, in other words, how well can the timbre of a tone from a given instrument at a certain pitch be reproduced over a number of performances. Our classification results to date would indicate that pairs of independent recordings from the same instrument at the same pitch are close in timbre but this can be confirmed by examination of the spectral centroid.

In figure 4.19 we can see that the irregular plot of spectral centroid versus frequency is highly reproducible with independent recordings. This is an important finding due the variation due to high degree of human input in the production of the tone. These findings confirm that the timbre of each violin is, in fact, a set of different pitch specific timbres that together make up the sound of that violin and that these timbres will be relatively consistent over repeated performances.

4.3.5 Discussion and Conclusions for Experiment2: Classification of Five Violins

As with the guitars, the fundamental task was to quantitatively represent the timbral qualities of the five violins in order to classify given test tones from a set of reference tones. Using MDS trajectory paths as a representation of the timbre of each tone, we achieved a correct classification rate of 77.3% with FFT and 92.7% with CQT. It appears from these results that the MDS trajectory paths in frequency space offer a very good representation of the timbre or characteristic signature of each violin at each pitch. Examination of the PC loadings seems to indicate that features enabling classification are the relative power of the fundamental, spectral features related to the spectral centroid and fluctuations in pitch.

To test the robustness of the classification process and gain some insight into the timbre of violins, a number of variations in the process were tried.

To compare the performance of FFT and CQT as techniques for time-frequency transform some variations in the number of principal components were considered. We found that when the number of PC's was reduced to three, the performance of the FFT data reduced from 77.3% to 72.7% and the performance of the CQT data reduced from 92.7% to 86.6% showing that, with both forms of preprocessing, less information produced a reduction in classification performance. An interesting observation with the violins was that the proportion of variance contained in three principal components reduced with increase in fundamental frequency. This indicates that violin tones of higher frequency are more complex to describe than tones of a lower frequency. Overall the classification rate with FFT was significantly lower to that with CQT. This contrasts with the findings with guitar tones. It is not clear why the CQT should be so superior in this context but a contributing factor could be the logarithmic nature of CQT which highlights the information on the lower harmonics and bunches the information in higher harmonics.

A variation in the standard use of principal components (weighted according to variance) was tried using the Malalanobis distance where all PC's are given equal weighting. Our test involved beginning with one PC as a basis for classification and, one by one, adding more PC's. We found that initially the classification performance increased, peaking at 92% with four PC's, and then gradually decreased until at 100 PC's the performance was little better than random guessing. So overall, the variation produced equivalent results

but with an increased level of complexity in the process.

The use of linear discriminants as a classification tool was again attempted using LDA after PCA providing a space based on three linear discriminants. For violin tones, this resulted in a small decrease in the classification rate (76.7%) with FFT and a small increase to 94% with CQT - overall there was no significant improvement.

In a natural situation it may not always be possible to compare complete and synchronised tones. When partial (but still synchronised) violin tones based on CQT were offered for classification, some interesting results were obtained. With a portion of the attack we achieved a correct classification rate of 73.3% - a significant decrease in the result for a whole tone, and with a large portion from the steady-state we obtained a 94.6% correct classification rate - a small increase in the rate for the whole tone. This indicates that valuable information for classification is contained in both the attack and the steady-state but, in contrast to the guitar, the steady-state is a more efficient classifier than the attack. We also found that from the mean spectrum for the steady state, a 94.0% correct classification was obtained. When only a small portion of the steady-state was considered for classification, performance was reduced but not markedly. For example, good results (around 80% correct) were obtained from a sequence of 10 windows taken from the steady state. In comparison with the guitar, violin tones were more difficult to classify with a small amount of data, indicating that there is a higher degree of random fluctuation in a violin spectrum.

When classification of violins was attempted with offset tones, a result that contrasted with guitar tones was obtained. With an offset of 10 data points, we obtained a correct classification rate of 92.8%, indicating that offset made no significant difference to classification with violin tones. This is most likely due to the the steady-state nature of violins and the result indicates a relatively low level of variation over time. This result is supported by the plot of spectral centroid over time shown earlier (figure 4.17), which illustrates the relatively stable nature of the spectral centroid in the steady-state.

Another important issue in comparing violin tones in a natural situation is the importance of the tones being of the same pitch or fundamental frequency. When tones just one step different in pitch were compared, the classification rate dropped to about 28%- marginally better than random chance. This suggested that the timbre of violin tones is highly dependent on fundamental frequency. This was confirmed by plots of the spectral centroid

which indicated that the average spectrum, and hence the timbre, is highly dependent on fundamental frequency and that, as with the guitar, the timbre for a particular violin is a set of different but frequency related timbres (see figure 4.11).

In summary, we have shown that it is possible to obtain extremely high classification results for a set of violins in a controlled experiment but the classification performance depends on a factors which would be very difficult to control in a real world situation (eg. pitch). We have also learnt that, for the violin, there are important cues for classification in both the attack and the steady state but the information in the steady-state is more reliable. The experiment has highlighted aspects of violin timbre which, although well known to researchers of violin timbre, have not been taken into account in instrument classification. In particular, that the timbre of a particular violin tone is relatively consistent over time and that violin timbre is highly dependent on fundamental frequency.

Chapter 5

Sound Production and Timbre

5.1 Sound Production in String Instruments

Note that the significant reference in this chapter is Fletcher & Rossing (1998).

5.1.1 Introduction

Each musical instrument type has its own unique way of generating sounds with a certain loudness, duration, pitch, and timbre. Our primary goal of classification is integrally linked to sound production and we investigate this with the guitar and violin. We are interested in the nature of timbre in these instruments, in particular, the factors influencing the timbre which include the pitch and time from excitation.

Although representing timbre is complex and challenging, the timbre at a particular time can be measured by the spectrum at that time and hence the spectral centroid is the best single indicator of timbre at a particular time. It was observed in the classification experiments described in the previous chapter, that the timbre of the guitar and violin is frequency dependent. By plotting the average spectral centroid (omitting the attack period) against pitch we can show how the timbre of an instrument varies significantly across its range. For example see the plot for the Martin guitar in figure 5.1. We also notice that the data graph is highly reproducible, reinforcing the view that variations in spectral centroid are predominantly due to changes in timbre. In this chapter, we attempt to explain these differences in timbre in terms of the physics of sound production for the

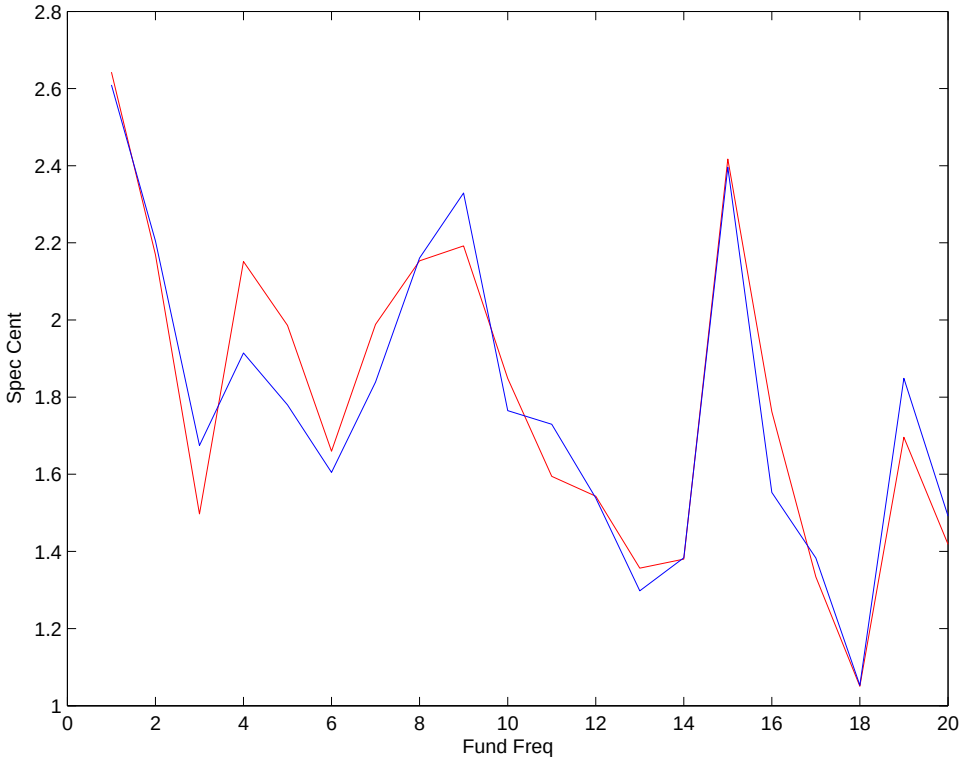

Figure 5.1: Spectral centroid for two independent sets of recordings of the Martin D28 guitar(#4) across its frequency range.

violin and guitar.

This variation in timbre with pitch for a guitar or violin is integrally tied to the means of construction and the properties of the body of the instrument. A guitar or violin can be thought of as a system of vibrating wooden plates or air cavities coupled together in various ways. Although each component in the system may behave as a simple vibrator, the behaviour of the coupled system is relatively complex. The sound is initiated when a string is stimulated to vibrate by either plucking or bowing. The vibration patterns in a taught string are easily predictable if the position of plucking or bowing is known. We can solve a differential equation to determine the relative amplitudes of each of the partials present. The loudness of sound generated in air by a vibrating string is negligible - for instance, consider the sound produced by a unamplified solid-body electric guitar. The sound from a violin or guitar is generated by the vibrating string interacting with the body of the instrument through the bridge (over which the strings are suspended) and in turn the vibrating parts of the instrument which stimulate sound waves in air. The body of each instrument will add its own unique flavour or musical signature to the vibrations of the string. We can think of the body of a stringed instrument as a kind of *acoustic transfer*

function (Wolfe et al. 1995) acting on a standardised input - a vibrating string excited repeatedly at the same point generates an invariant spectrum. In this way the instrument body determines the *frequency response* of the instrument.

In the case of a guitar, the player has little control of the sound once a string has been plucked. However, there are several variables in the plucking of the string which affect the timbre of the tone. The most important of these are: the position relative to the bridge at which the string is plucked, which determines the strength of particular harmonics in the string vibration ; the force with which the string is plucked, which determines the initial amplitude; and the angle of the initial string displacement relative to the bridge of the instrument, which is a factor in determining the rate of decay for the tone. In the case of the violin, the player is constantly interacting with the instrument during the period of production of the tone. The key variables relating to bowing which influence the timbre are: the position of the bow relative to the bridge; the speed of the bow; the force applied by the bow to the string; and the variation in the speed and force over time. In this study, since the aim is to compare the timbre of instruments, we have attempted to keep the human factor as consistent as possible.

5.1.2 Equation for a Vibrating String

The string vibrations associated with a guitar or violin are predominantly transverse with a small and relatively insignificant component of longitudinal vibrations. For the purposes of this investigation we will consider only the transverse vibration. The displacement of the string at any time will be the sum of the displacements for each of the component modes. To write the transverse wave equation for a vibrating string, we first consider a uniform string with linear density μ stretched to a tension of T Newtons. If the string is displaced from its equilibrium position and released, a vibrating motion is initiated. The displacement y of the string from its equilibrium position, at point x from one end, can be represented by the differential equation (Fletcher & Rossing 1998),

$$\frac{\partial^2 y}{\partial t^2} = \frac{T}{\mu}\frac{\partial^2 y}{\partial x^2} = c^2\frac{\partial^2 y}{\partial x^2}, \tag{5.1}$$

where $c^2 = T/\mu$.

Now let us consider the behaviour of a vibrating string at its end points. In the case of a

single pulse moving toward a fixed end, the pulse is reflected as a pulse of opposite phase, whereas, at a free end (free to move), the pulse is reflected as a pulse with the same phase. A solid body electric guitar is an example of a string instrument with fixed ends. For simplicity we will consider the string motion of a violin or a guitar as approximately that with fixed ends. In reality, however, the motion is a little more complex since there is small but significant movement at the bridge end - the bridge being mounted on a stiff but vibrating 'wooden plate'.

If a vibrating string is considered to have fixed ends, then the vibration takes the form of a standing wave. The string has normal modes of vibration given by (Fletcher & Rossing 1998),

$$y_n(x,t) = (A_n \sin \omega_n t + B_n \cos \omega_n t) \sin \frac{\omega_n x}{c}. \tag{5.2}$$

The general solution for the displacement of the string is the sum of all the modes,

$$y(x,t) = \sum_n (A_n \sin \omega_n t + B_n \cos \omega_n t) \sin k_n x. \tag{5.3}$$

5.1.3 Time and Frequency Analysis of a Plucked String

When string motion is initiated by plucking (or bowing), the resulting vibrations are a combination of the natural modes of vibration. The mix of modes and their relative strengths are determined by the position at which the string is plucked. If the string is struck at half its length, then the second harmonic/mode and multiples of this mode (even harmonics) will be suppressed leaving only the odd harmonics. Except at the moment the string is excited, when all modes are instantaneously in phase, the different modes will move in and out of phase due to their different frequencies. As a result the shape of the string will be a complex and continuously changing shape. It is possible, however, to represent the string displacement at a given position and time as a Fourier series with each term corresponding to a vibrational mode of given frequency (see equation 5.3). The total displacement of the string is represented as the sum of displacements due to each mode. The coefficients A_n and B_n in the Fourier series can be expressed as follows (Fletcher & Rossing 1998),

$$A_n = \frac{2}{\omega_n L}\int_0^L \dot{y}(x,0)\sin\frac{n\pi x}{L}dx, \tag{5.4}$$

$$B_n = \frac{2}{L}\int_0^L y(x,0)\sin\frac{n\pi x}{L}dx. \tag{5.5}$$

If the position of the initial excitation of the string is known, then the Fourier coefficients can be calculated and hence the amplitude of the vibrations for each of the modes can also be calculated. In acoustical terms, the amplitude of each mode is a measure of the power of each partial in a musical tone. From this data the frequency spectrum of a plucked string can be plotted.

5.1.4 A Frequency Analysis for a Vibrating Guitar String: A Particular Example

In this study, all guitar tones were produced by plucking the string at approximately the same position. This was found to be at a distance of approximately $\frac{1}{6}$ of the string length measuring from the bridge. As a consequence of plucking the string at this position, the 6th, 12th, 18th etc. harmonics will be suppressed. Note that this ratio will vary somewhat since the effective string length varies depending on the note being fretted.

Using equations (5.4) and (5.5), we can now calculate the Fourier coefficients for a guitar string plucked at $\frac{1}{6}$ of its length. If the string is considered to be vibrating at fixed ends, then these coefficients represent the frequency spectrum for the string vibration. As the string is released we can state the following initial conditions,

$$\dot{y}(x,0) = 0; \quad y(x,0) = \frac{6h}{L}x, \quad 0 \leqq x \leqq \frac{L}{6}, \quad y(x,0) = \frac{6h}{5}(1-\frac{x}{L}), \quad \frac{L}{6} \leqq x \leqq L. \tag{5.6}$$

From the first condition we get $A_n = 0$ and from the second,

$$B_n = \frac{2}{L}\int_0^{\frac{L}{6}} \frac{6h}{L}x\sin\frac{n\pi x}{L}dx + \frac{2}{L}\int_{\frac{L}{6}}^{L} \frac{6h}{5}(1-\frac{x}{L})\sin\frac{n\pi x}{L}dx, \tag{5.7}$$

to give

$$B_n = \frac{36h}{2n^2\pi^2} \sin \frac{n\pi}{6}. \tag{5.8}$$

Using equation (5.5) and our result (5.8), we list the relative amplitudes of the harmonics in table 5.1. Figure 5.2 provides the plot of the spectrum. Note that the sixth harmonic is absent.

Guitar	
Harmonic	Relative Amplitude
1	1
2	0.433
3	0.222
4	0.262
5	0.040
6	0
7	0.020
8	0.033
9	0.025
10	0.017

Table 5.1: Frequency analysis for a string plucked at $d = \frac{1}{6}$.

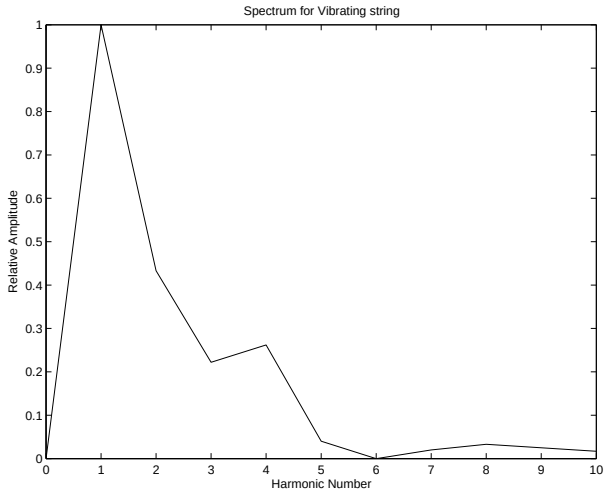

Figure 5.2: The spectrum for a vibrating string plucked at $d = \frac{1}{6}$.

It is important to note that we have assumed that, when a string is excited, the initial shape is two straight lines meeting at a point. In reality the initial shape is two straight lines joined by a relatively small smooth curve. This deviation from the ideal model is due partly to the inherent stiffness in the string and partly to the small but significant width of the finger, plectrum, hammer or bow initiating the string vibration. One consequence of these factors is to cut out the highest modes of vibration (above about $n = 40$ in a guitar string) (Benade 1990). Another consequence, for free vibrating strings (guitar and piano), is that the frequencies of the upper partials are higher than the expected multiples of the fundamental frequency. This effect is more marked in thicker strings due to their greater stiffness. In musical terms, the upper harmonics on the lower strings of a guitar will sound a little sharp, contributing to the distinctive timbre of the guitar. This effect is more marked in piano timbre where the strings, in general, are thicker than those of a guitar (Fletcher (1964); Beauchamp (1974)).

An important aspect of a vibrating guitar string is the damping that takes place due to: the stiffness in the string, straight motion at the end supports (particularly at the bridge) and the transfer of energy to other parts of the guitar. The exponential dissipation of energy in the string vibration is an important factor in defining the timbre of a guitar.

We have established that the sound produced by a guitar or violin is initiated by a relatively constant driving force in the form of a vibrating string. We now turn our attention to how the vibrating string interacts with the body of an instrument. We have previously stated that a guitar or violin body can be considered to be a transfer function, where the string vibration provides a relatively standard input into a complex vibrating system which possesses a characteristic frequency response. Further, knowing that we are starting with a common sound stimulus allows a reliable means of comparing the tonal qualities of each instrument. We can determine the effect that the body of each instrument, as a complex system of coupled vibrators, has on a common input stimulus.

5.1.5 The Motion of a Bowed Violin String

Although the motion of a bowed string appears to have a smooth curved shape, Helmholtz (1863) showed that in fact the string takes the shape of two straight lines with a sharp bend which moves from end to end along the curved path of the string once per cycle (ie. at the fundamental frequency).

The bow acting on the string can be described as a stick and slip action. The string is pulled at the velocity of the bow until the bond is broken by the sharp bend meeting the bow. The string vibrates freely until the string is again picked up by the bow due to the frictional force between the bow and the string.

The motion of a bowed vibrating string depends on the fact that the coefficient of friction of a weighted bow moving across a string is less than the coefficient of friction of a bow which is stationary relative to the string. If a bow is placed on the string with a certain downward force on the string and then moved in a direction perpendicular to the string, the bow will initially stick to the string, bending it until the restorative force from the tension in the string equals the static frictional force - at which time the string breaks away from the bow, vibrating in a semi-free fashion due to the reduced co-efficient of friction. At the moment in time when the velocities of the bow and the string are again equal, the frictional force increases again locking the string and the bow together. The cycle repeats giving a continuous stick-slip motion.

The harmonic nature of a bowed string is approximately the same as that of a naturally vibrating struck or plucked string. In the same way as a plucked string, if the position of the bow along the length of the string is a rational fraction $\frac{m}{n}$, then the n^{th} partial will be suppressed. The factors influencing the harmonic nature of a tone are a little more complex than for a free vibrating string - in addition to the bow position the downward pressure of the bow and the bow speed have an influence on the tone quality. These factors can produce *jitter*, which is random variation in the pitch of the vibrating string. In addition, an increase in bow pressure results in a slight flattening of the pitch.

5.1.6 The Guitar/Violin as a Vibrating System

A guitar or violin body can be described as a system of coupled vibrators that interact to amplify and colour, in a characteristic way, the sound initiated by the string vibrations. Although the sound is initiated by plucking or bowing the strings, this accounts for only a small portion of the radiated sound. The strings excite the bridge and the top plate which in turn excite the ribs (sides), back plate and air cavity. In this way the sound is radiated through the vibrating plates and the sound hole. High frequency sounds are predominantly radiated by the top plate, whereas, low frequency sounds are predominantly radiated by the back plate via the ribs and air cavity.

An important property of each of these vibrating parts is that each has a natural resonant frequency at which it will vibrate. In other words, when a vibrating string exerts a driving force on each of the parts of a guitar/violin body, it will initially begin resonating at its own natural frequency as well as at the frequency of the driving force - the vibrating string. If an instrument is to sound harmonic, then these inharmonic frequencies must be dampened in some way. This occurs through the mechanical resistance inherent in the body construction and the viscous drag of air in the air resonance cavity.

To explain the behaviour of each component part of a guitar or violin, consider a simple point mass driven by an external sinusoidal force $f(t) = F\cos\omega t$. (For the guitar or violin the driving force is the harmonic motion of a vibrating string). The vibrating mass is subject to a damping force which resists movement and attenuates the energy of the oscillator over time. The equation of motion for a damped oscillator is (Fletcher & Rossing 1998),

$$m\ddot{x} + R\dot{x} + Kx = F\cos\omega t. \tag{5.9}$$

Taking into account the damping, the resonant angular frequency of a damped oscillator will be,

$$\omega_d = \sqrt{\omega_0^2 - \alpha^2}, \tag{5.10}$$

where α is the damping coefficient. Notice that the natural resonant frequency of an oscillator is lowered somewhat by the damping.

As a reference for comparing the displacement amplitude of a vibrating part of an instrument, it is useful to define the static displacement, $x_s = F/K$, produced by a constant force F, where K is the stiffness of the vibrating part. At very low driving frequencies, the displacement amplitude of the oscillator will approach the static displacement. As a measure of amplification, it is useful to define

$$Q = \frac{x_0}{x_s}, \tag{5.11}$$

where Q is the ratio of displacement at resonance to the static displacement.

When the driving frequency equals that of the natural resonant frequency, $\omega = \omega_d$, then we have a resonance and the displacement amplitude reaches a peak. A resonance such as this boosts the output from a string instrument, colouring the sound quality accordingly. The amplification, and hence the magnitude of the displacement at resonance, is dependent on the damping coefficient. The stronger the damping on an oscillator, then the smaller the amplitude of the resonance.

For a guitar or violin there will be a resonance when the a harmonic of the vibrating string is close in frequency to the resonant frequency of any of the vibrating parts of the body. In the language of acoustics, we say that there is a formant at this frequency and, as a consequence, the output from the instrument at that frequency will be boosted.

When the external driving force is first applied to a vibrating instrument part, the motion is usually complex. At first, the oscillator vibrates at both its natural resonant frequency f_0 and the frequency of the driving force f. At this time the motion of the oscillator is a superposition of the component motions of the two vibration modes. If the oscillator is damped then, the vibrations at the natural resonant frequency will decay and the oscillator will be progressively forced to vibrate at the frequency of the driving force. In time, the oscillator will move through this transient state to an approximately steady state vibration at a frequency f.

The displacement of an oscillator driven by force $f(t) = Fcos\omega t$ can be shown to be,

$$x = Ae^{-\alpha t}\cos(\omega_d t + \phi) + \frac{F}{\omega Z}\sin(\omega t + \phi), \tag{5.12}$$

where A and ϕ are arbitrary constants, ω_d is the natural frequency of vibration for the damped oscillator, α is the damping coefficient and Z is the mechanical impedance of the oscillator.

So far, we have assumed that each vibrating part in an instrument body has just one mode of vibration. In reality, each part has potentially an infinite number of modes, although it is generally the simplest mode that predominates.

To explain the operation of the body of a guitar or violin, we consider vibration modes of the various body parts. The body of each instrument can be thought of as being made up of wooden plates of an approximately rectangular shape and an air cavity which approximates a Helmholtz resonator. The nature of the vibration in a rectangular plate is determined by the boundary conditions which may be either free, clamped (rigid), or supported (hinged). In reality, the plates in a guitar or violin are somewhere in between supported and clamped. The behaviour of a wooden plate differs from a regular plate in that wood is not a uniform material. Properties such as stiffness and elasticity vary according to which axis is being considered.

If we begin with a rectangular plate of thickness h and dimensions L_x and L_y, with simply supported edges, we can express the equation of motion as (Fletcher & Rossing 1998),

$$\frac{\partial^2 z}{\partial t^2} + \frac{Eh^2}{12\rho(1-\nu^2)}\nabla^4 z = 0. \tag{5.13}$$

This equation can be solved for the displacement amplitude giving,

$$Z = A \sin\frac{(m+1)\pi x}{L_x} \sin\frac{(n+1)\pi y}{L_y}, \tag{5.14}$$

where m and n are non-negative integers representing the vibrational modes (m, n). Note that $(0, 0)$ represents the fundamental mode of vibration.

Allowing for the effects of the direction of grain in a wooden plate, the corresponding vibration frequencies are given by,

$$f_{mn} = 0.453h[c_x(\frac{m+1}{L_x})^2 + c_y(\frac{n+1}{L_y})^2], \tag{5.15}$$

where c_x and c_y are constants depending on properties of the wood which, in turn, depend on the direction of the grain. These properties are: elasticity (Young's Modulus), density, and stiffness (Poisson Ratio).

Although this equation allows us to roughly estimate likely resonant frequencies for instrument parts such as the back plate and the top plates, the only accurate way to determine them is through empirical measurement. This is because parts such as the wooden sound board and the back are only approximately rectangular, having rounded corners, and are not regular in either density or stiffness. The predominant modes and their frequencies depend on shape and on the *bracing* used in the construction of the instrument, as well as variation in the properties of the wood. In fact, in the case of high quality instruments, makers adjust the resonant frequencies of the back and sound board during the construction process to produce the the most pleasing timbre.

Another important part of the system of vibrators that makes up a guitar or violin is the air cavity, which operates partly as a Helmholtz resonator. A Helmholtz resonator with a neck operates in a similar way to a piston vibrating freely in a cylinder whose motion, in turn, is analogous to a mass attached to a spring. The mass of air in the neck acts a piston and the large volume of air in the cavity serves as a spring.

In instruments such as the violin and the guitar, the body operates as a Helmholtz resonator without a neck. (The neckless resonator operates in a similar fashion to a resonator

with neck). The natural frequency of a neckless Helmholtz resonator with a large face surrounding the hole is,

$$F_0 = \frac{c}{2\pi}\sqrt{\frac{1.85a}{V}}, \tag{5.16}$$

where a is the radius of the opening and V is the volume.

In fact, the air cavity in a guitar does not operate purely as a Helmholtz resonator but also interacts with the vibrating wooden plates. It is the simple mode of the air cavity that creates the lowest resonance in a guitar or violin body.

Whilst having considered the vibrating parts of a guitar or violin in isolation, in fact, each body part acts as a part of a coupled system. If two vibrating parts with the same natural frequency ω_0 are coupled together, the coupled system will vibrate with two normal modes - one at a frequency equal to the natural frequency of the oscillators and a second at a higher frequency which depends on the strength of the coupling. In the more general case, when two vibrating parts with different frequencies are coupled together, two new modes of vibration will be created - the strength and frequencies will depend on the strength of the coupling.

5.1.7 Summary of Factors Influencing the Frequency Response of the Guitar and the Violin

The timbre of a guitar tone is initially influenced by the input from the vibrating string. The variables associated with the vibrating string are the position, angle and force with which the string is plucked. The frequency response of the guitar - the response to the vibrating string - is influenced by the natural resonances and anti-resonances associated with the system, which is the body of guitar. There is a consequential boosting of partials with frequencies near resonances (formants) and suppression of partials with frequencies near anti-resonances. The degree of dampening of the 'forced' vibrations in the body of the guitar is another distinguishing feature of timbre. It determines the length of the transition from attack to steady-state. A feature that distinguishes the guitar (and piano) from other instruments is slightly inharmonic partials (due to stiffness in thick strings). In particular, the upper harmonics on the thicker (lower pitch) strings will sound a little sharp.

The timbre of a violin tone is initially influenced by the input from the bowed string. The variables associated with the vibrating bowed string are the position, angle and the continuing downward force of the bow acting on the string. In a similar way to the guitar, the frequency response of the violin is influenced by the natural resonances and anti-resonances associated with the system.

The resonances, for both the violin and the guitar, are associated with the air cavity, the top plate, the back plate and the whole body as a coupled system. We can speculate, for both the violin and the guitar, whether the resonances and anti-resonances will create an invariant spectral envelope where the initial spectrum of the vibrating string will be modified in a predictable way depending on the fundamental frequency.

5.2 Analysis of Timbre for the Guitar and Violin

5.2.1 Introduction

In this chapter we will explore the changes in timbre with pitch for the particular guitars and violins used in this study. We will also explore in depth the natural resonances in the body of each guitar and violin that colour the frequency response of these instruments.

5.2.2 Analysis of Guitar Timbre

We have observed, in the classification experiments described in chapter 4, that the timbre of each guitar is frequency dependent. Using the spectral centroid as a single indicator of timbre, we have seen how the timbre varies across the frequency range of the instrument. In this section, we examine this variation in more depth.

We also show that the natural resonances in the body of a guitar can be observed in the attack period of a tone, but are quickly dampened so that the vibrating parts of a guitar are forced to vibrate at the various harmonic frequencies of the vibrating string. For example, see figure 5.3 showing the resonances in a guitar body. When these natural resonances approximately coincide with the harmonics of the vibrating string then the harmonics are strengthened (formants), colouring the sound accordingly.

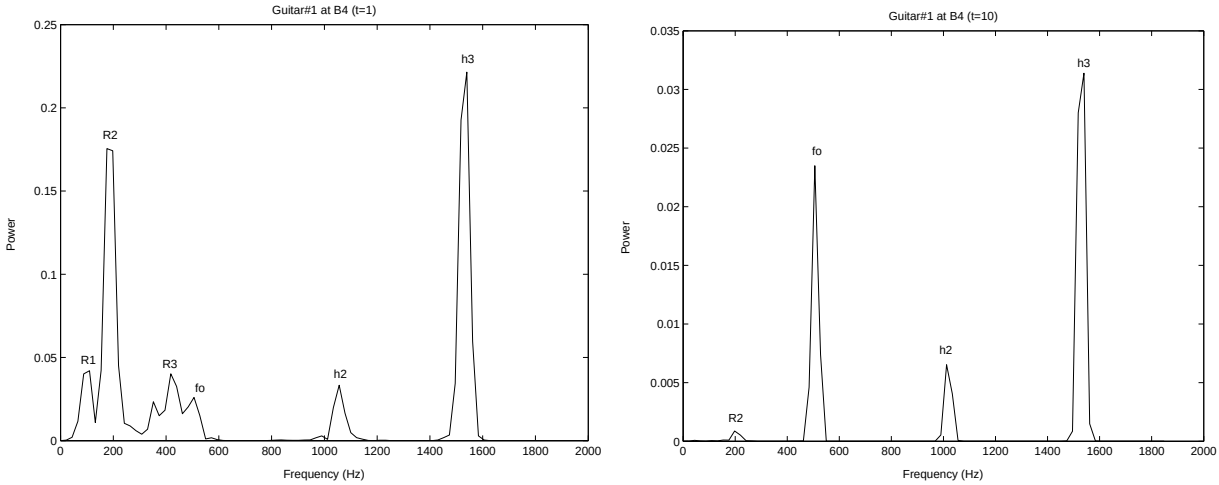

Figure 5.3: A plot of the spectra for the guitar#1 at B4, showing the natural resonances R_1, R_2, R_3 in the attack resolving to a harmonic spectrum in the decay.

These resonances correspond to the various modes of vibration associated with a guitar body. The significant modes can be described by the following notation (Fletcher & Rossing 1998).:

Top Plate: $(0,0)$, $(0,1)$, $(1,0)$, $(0,2)$, $(1,1)$, $(0,3)$, $(2,0)$, $(1,2)$

Back Plate: $(0,0)$, $(0,1)$, $(0,2)$, $(1,0)$, $(0,3)$, $(1,1)$, $(2,0)$, $(1,2)$

Air Cavity: A_0, A_1, A_2....

5.2.3 Guitar#1

The Attack: Natural Resonances

Analysing the spectra for the attack at pitches A2, A3 and B4 (figure 5.4) we can see that there are spikes indicating resonances at approximately 90Hz, 110Hz and 200Hz occurring consistently across the range of the guitar. There also appear to be minor resonances at around 340Hz and 420Hz. By reference to the work of Fletcher & Rossing (1998) who used sophisticated techniques to measure the modes of vibration for a range of different acoustic guitars, we can guess that the resonance at 110Hz is the primary mode of the air cavity (A_o) and the 200Hz is a resonance in the top-plate and/or the back-plate.

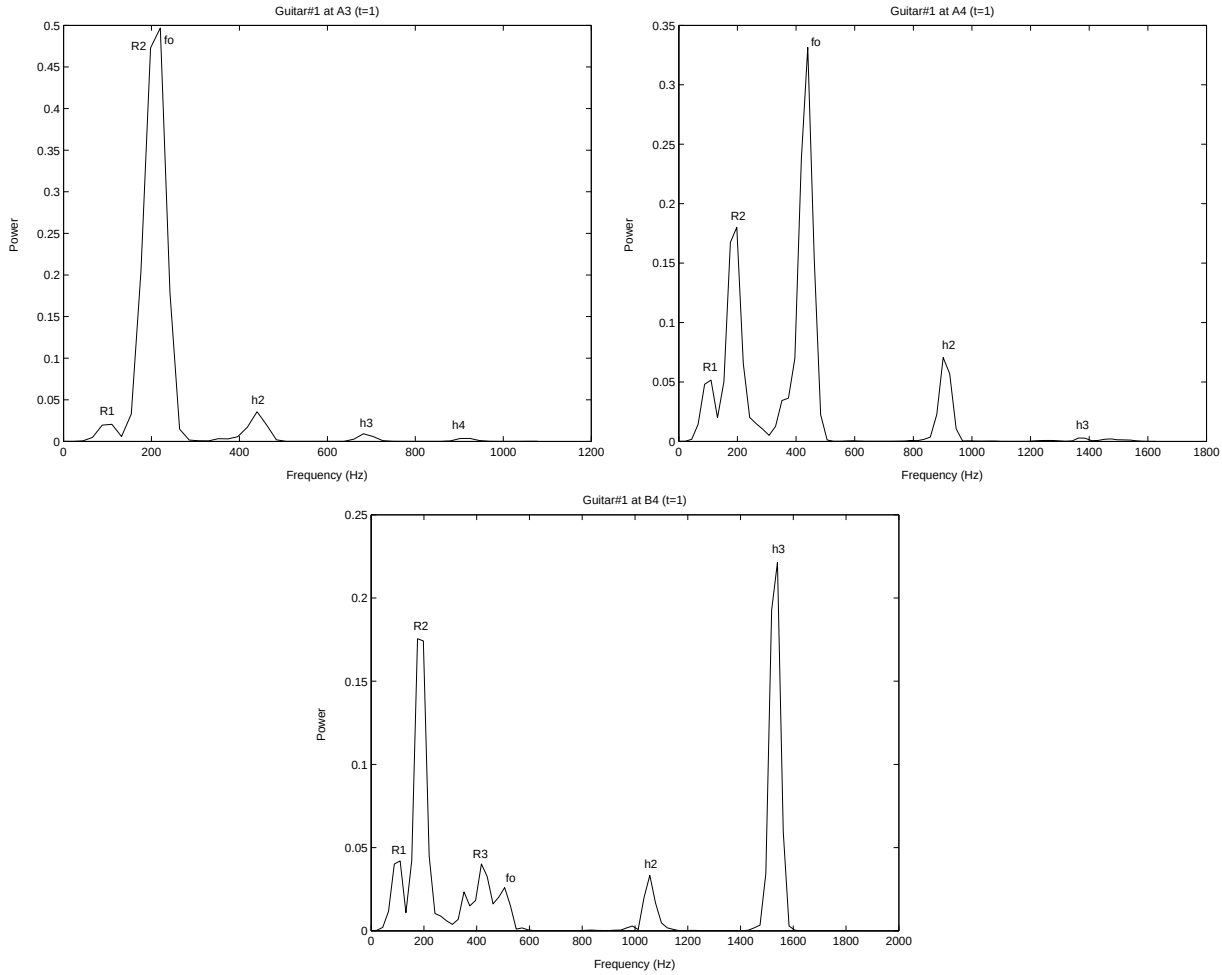

Figure 5.4: A plot of the spectra for the guitar#3 showing the resonances in the attack at A3, A4 and B4.

Steady-state/Decay

Examining figure 5.5, we see that this guitar displays a weak fundamental (first harmonic) at pitch A2; shows a balanced spectrum at A3; and at A4, the spectrum is dominated by the fundamental with other harmonics barely present. The spectrum at neighbouring tone B4 is a surprising contrast with A4, displaying a strong second harmonic and a dominant third. The spike at the third harmonic may be explained by the presence of a formant near this frequency.

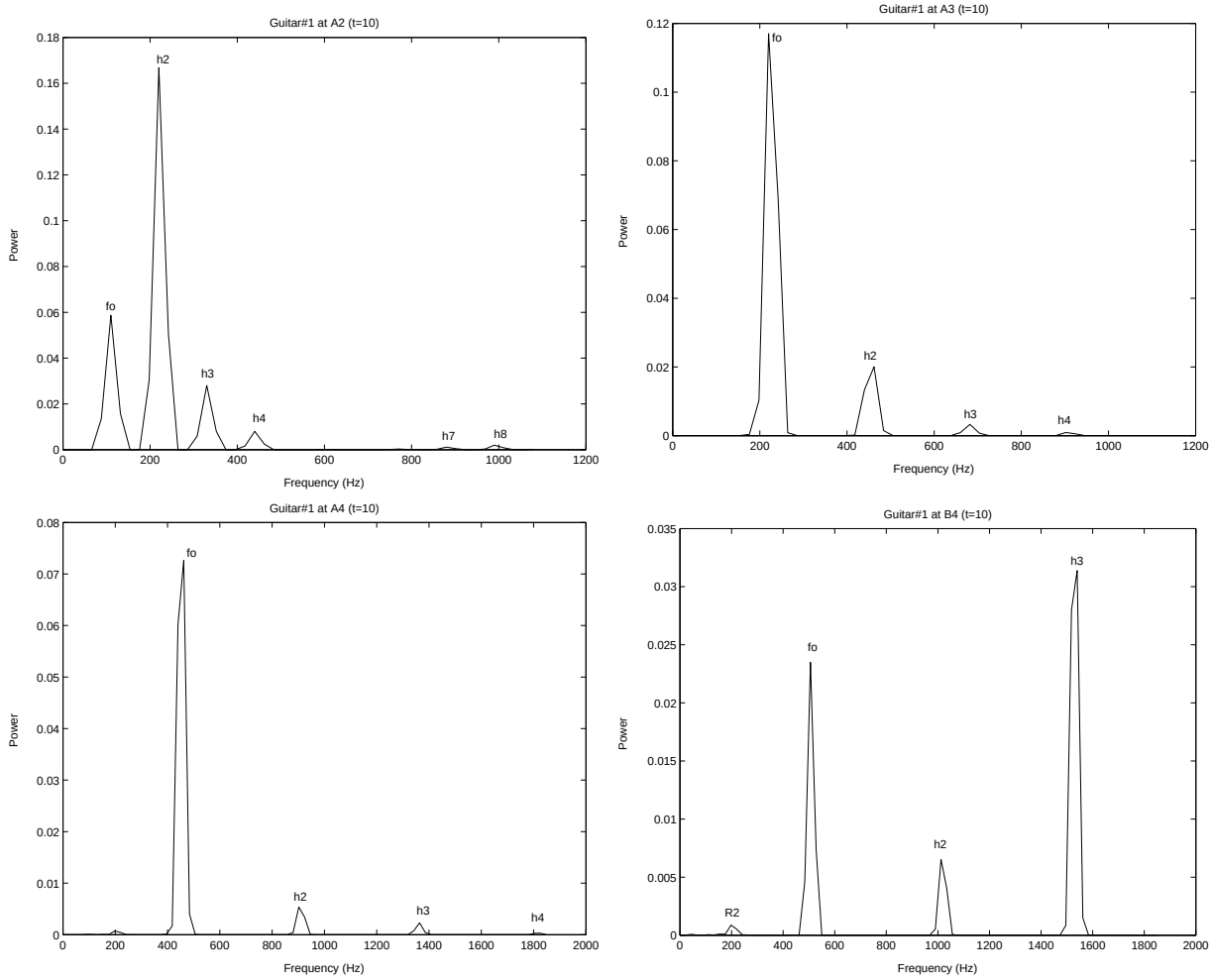

Figure 5.5: A plot of the spectra for the guitar#1 for the steady-state/decay at A2, A3, A4 and B4.

Overall, the timbre of this guitar is relatively consistent featuring a strong fundamental and few higher harmonics across the range of the instrument. However, it displays a weak fundamental at the lower end of its range and a spike in the third harmonic at B4.

5.2.4 Guitar#2

The Attack: Natural Resonances

Analysing the spectra for the attack at A2, A3 and G4 (figure 5.6), we can see that there is only one significant resonance, at approximately 220Hz, which occurs consistently across the range of the guitar. We can guess that the resonance at 220Hz is a resonance in the top-plate and/or the back-plate of the guitar body (Fletcher & Rossing 1998). The attack for this guitar also displays strong early growth in the harmonic frequencies.

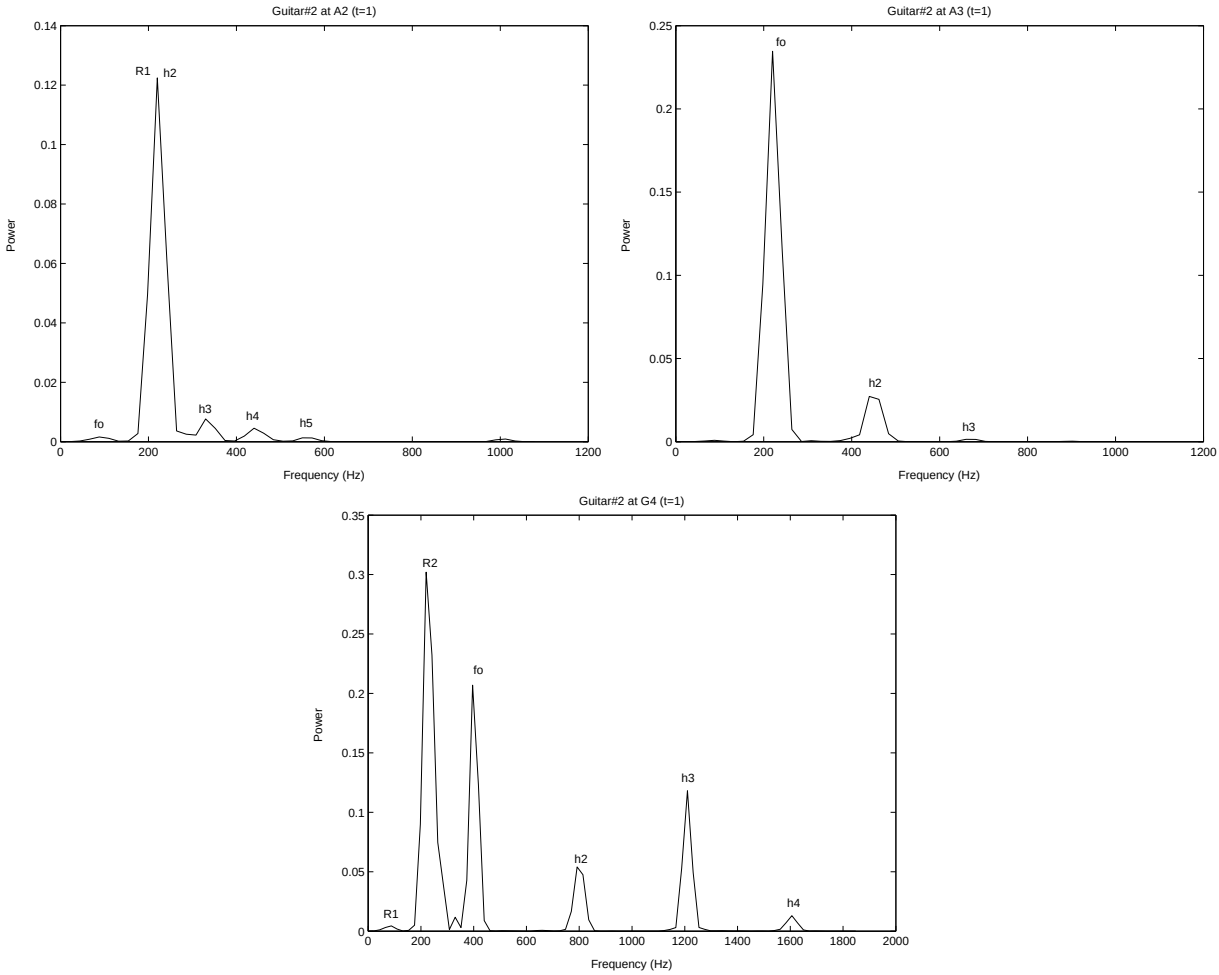

Figure 5.6: A plot of spectra for the guitar#2 showing the resonances in the attack at A2, A3 and G4.

Steady-state/Decay

Examining the spectra for this guitar (figure 5.7), we note that at A2, the fundamental is barely present and the dominant harmonic is the second at 220Hz. This may be explained by the resonance observed in the attack, creating a formant at this frequency. At A3, we see a rich and balanced spectrum, but at A4, the fundamental dominates with higher harmonics barely present. It is interesting that the neighbouring tones G4 and B4, on each side of A4, display a rich timbre with strong harmonics above the fundamental. The difference in timbre for the three adjacent tones can be explained by the presence of anti-resonances near the harmonics in A4. The timbre of this guitar is highly frequency dependent and does not vary in a systematic manner.

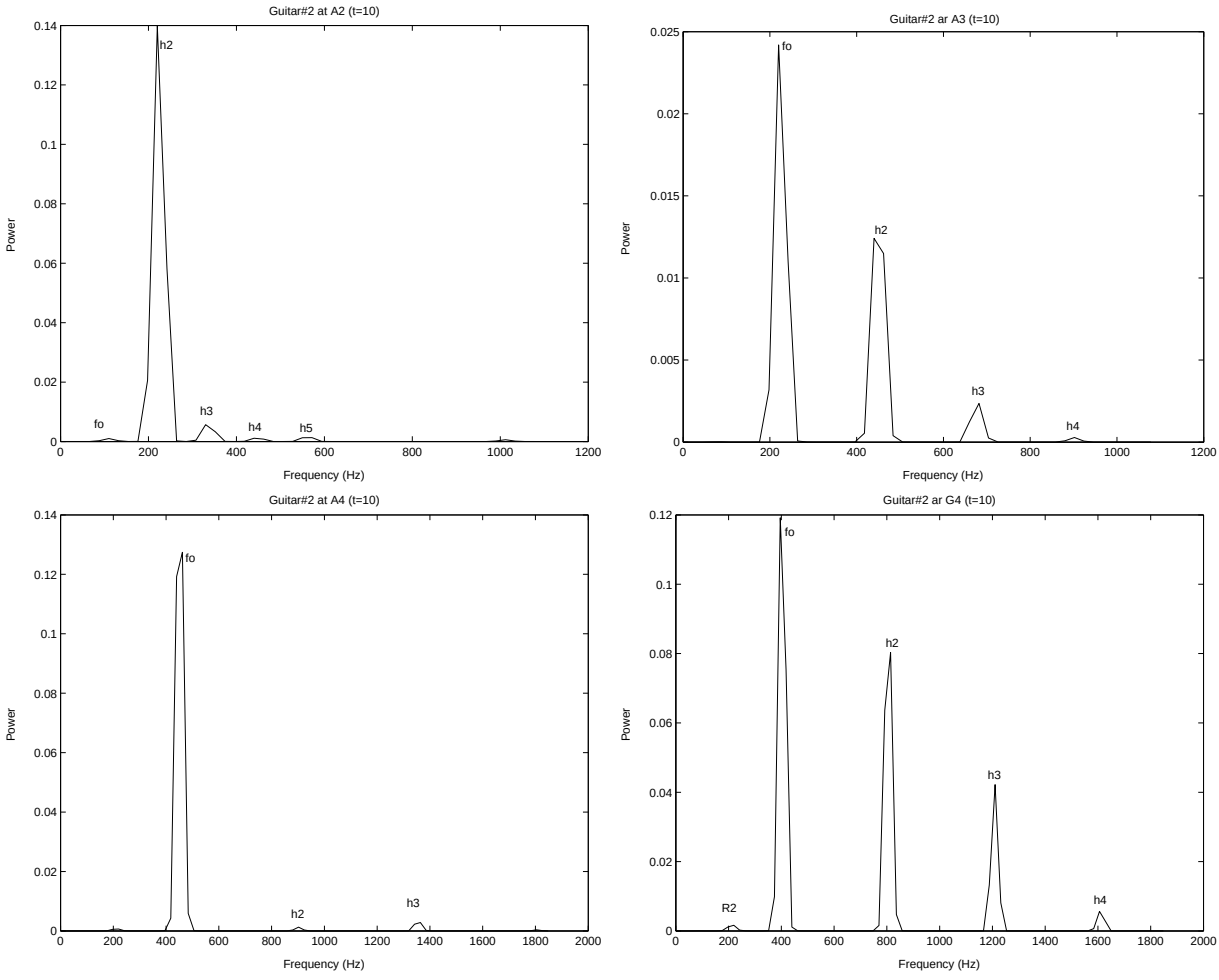

Figure 5.7: A plot of the spectra for the guitar#2 for the steady-state/decay period at A2, A3, A4 and G4.

Overall, guitar#2 displays a lack of resonance in the attack compared with the other guitars and its spectra features less harmonics above the fundamental. It displays a significant variation in timbre as illustrated by the spectra shown in figure 5.7. The rate of decay of its harmonics is fast relative to the other guitars. This instrument is an arch-top acoustic guitar of the type developed for use in orchestras before the days of amplification. The sound of these guitars is normally sharp and relatively loud, with a fast decay.

5.2.5 Guitar#3

The Attack: Natural Resonances

By observing the initial spectra at A2, A3 and G4 (figure 5.8), we can see that there are spikes indicating resonances at approximately 110Hz, 130Hz and 155Hz occurring consistently across the range of the guitar. There also appear to be minor resonances at around 220Hz and 350Hz. The same model guitar (Martin D28) is comprehensively tested in work by Fletcher & Rossing (1998). They found that that the fundamental mode of the air cavity resonated at 121Hz, the back plate at 161Hz and the top plate at 163Hz. These resonant frequencies accord reasonably well with with those observed in our work with a different example of the same model guitar.

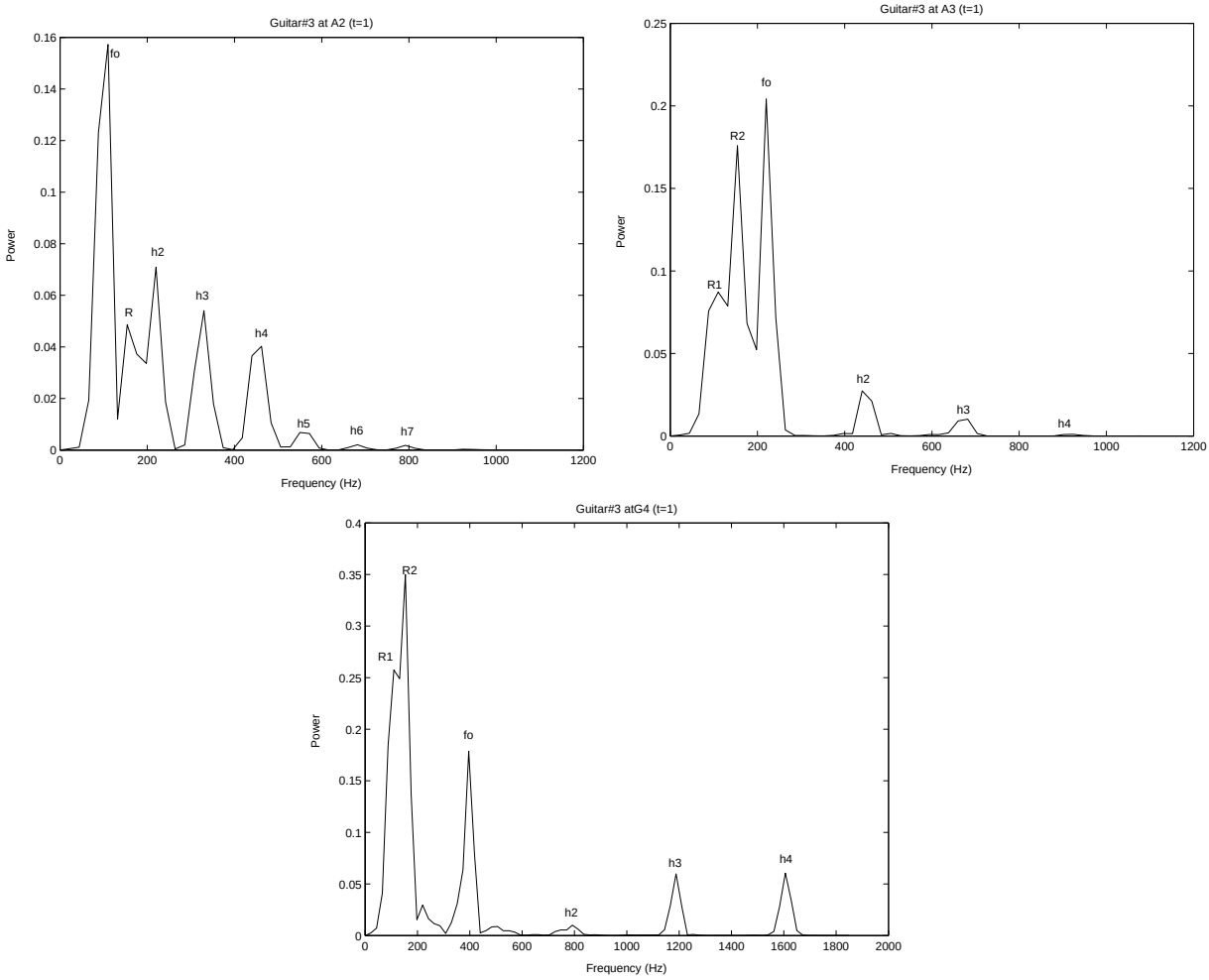

Figure 5.8: A plot of the spectra for the guitar#3 showing the resonances in the attack at A2, A3 and G4.

The Steady-state/Decay

Examining the spectra for this guitar we note that it displays a rich timbre across its range (with the exception of A4), as would be expected from a high quality instrument such as this (as defined by price and reputation). It is interesting to compare the spectra for the neighbouring tones G4 and A4 in the upper register of the guitar range. We observe that there are almost no harmonics above the fundamental at A4, whereas, at G4 there are significant harmonics. This difference in timbre may be explained by the presence of anti-resonances in the region of the harmonics for A4. This difference in timbre for two adjacent tones on the same string further strengthens the argument that the timbre of the guitar is highly frequency dependent and further, suggests that timbre does not vary in a systematic manner.

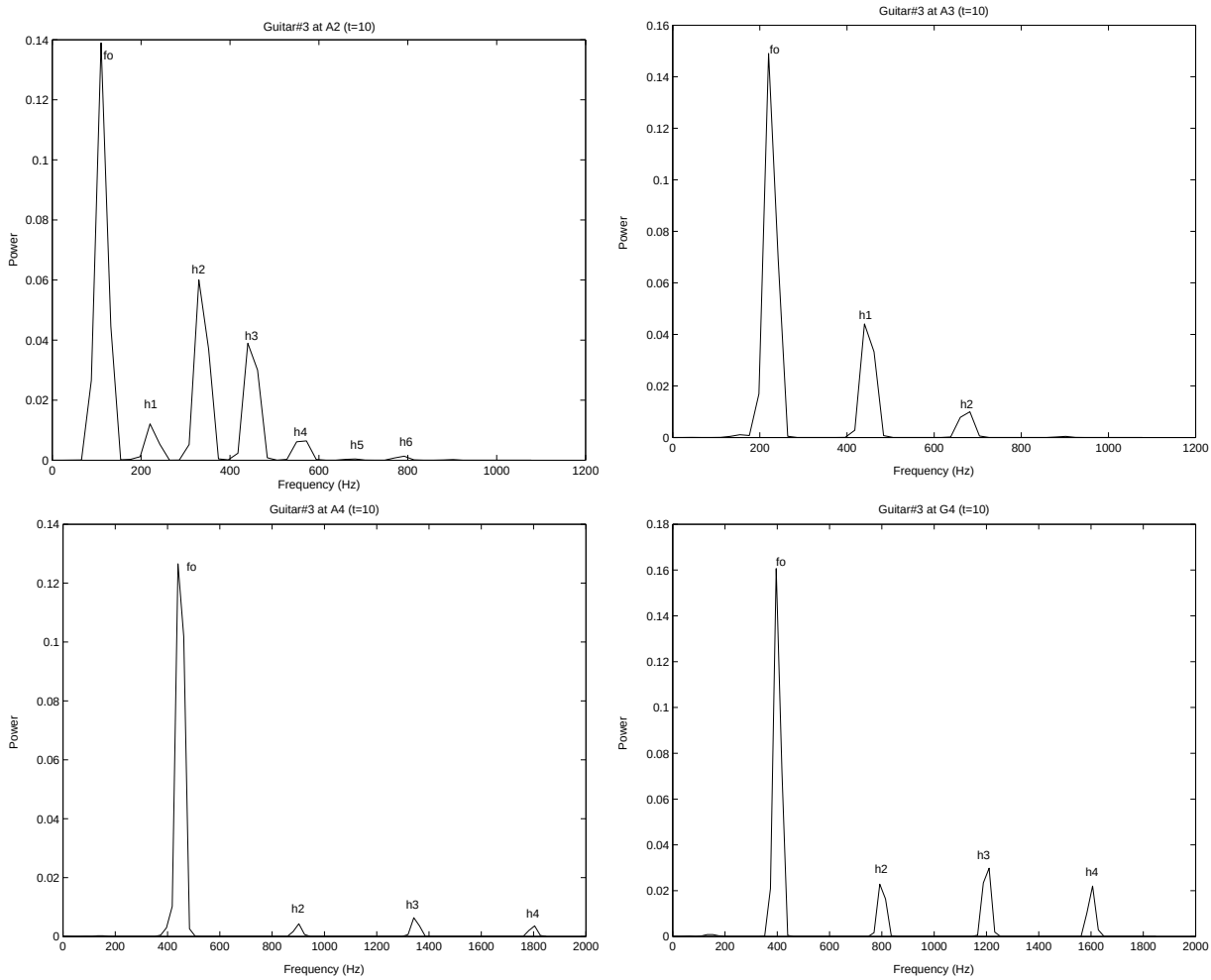

Figure 5.9: A plot of the spectra for the guitar#3 for the steady-state/decay at A2, A3, A4 and G4.

5.2.6 Guitar#4

The Attack: Natural Resonances

By observing the spectra at G4, A4 and B4 (figure 5.10), we can see that there are spikes indicating resonances at approximately 110Hz and 190Hz occurring consistently across the range of the guitar. These resonances likely correspond to the resonance of the air cavity ($A_o = 110Hz$) and top or back plate of the guitar ($190Hz$).

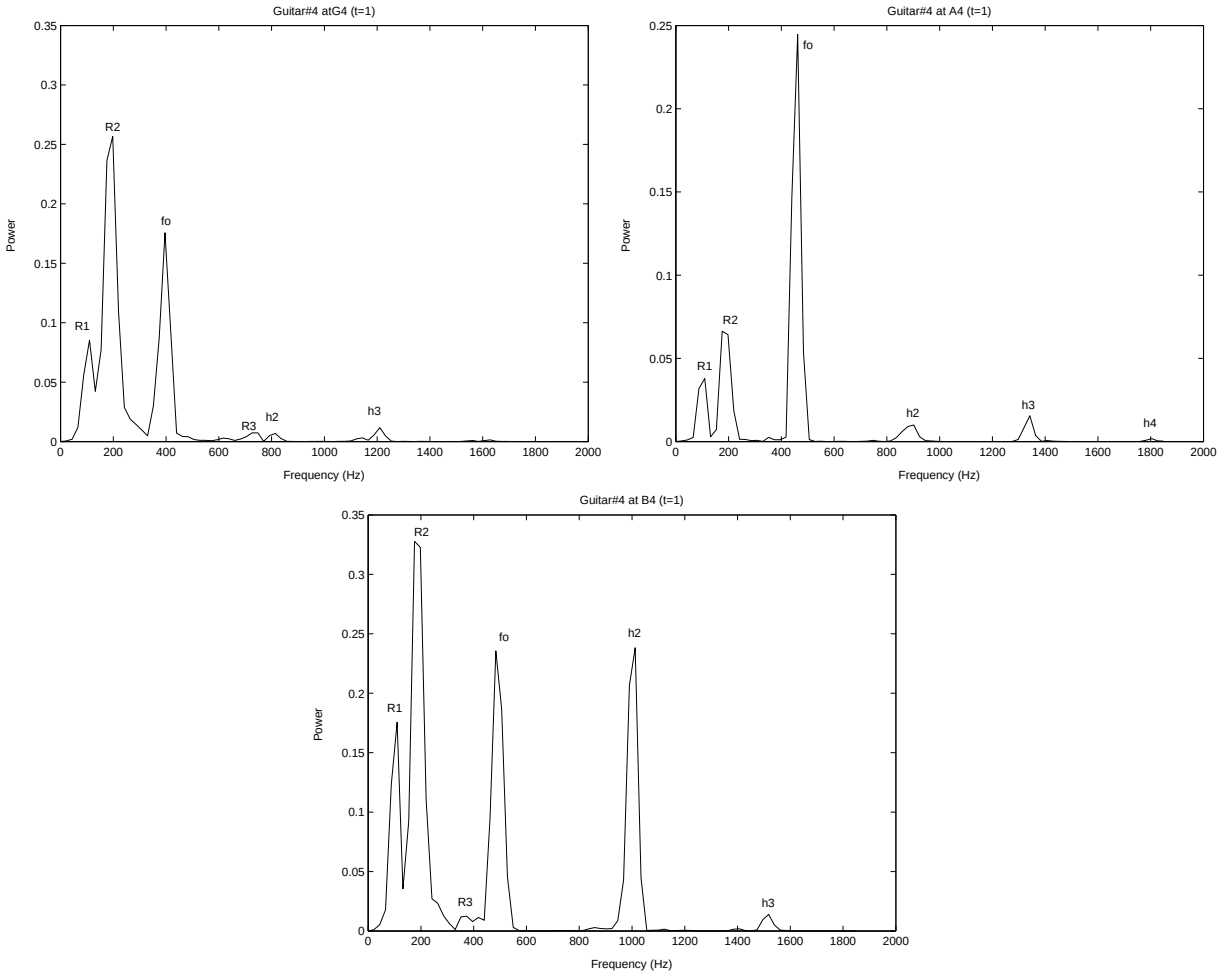

Figure 5.10: A plot of the spectra for the guitar#4 showing the resonances in the attack at G4, A4 and B4.

The Steady-State/Decay

For the decay/steady-state, this guitar displays very strong harmonics in the lower register, for example at A2, trending to moderate harmonics in the mid and higher register (see figure 5.11). It is interesting to note that at A2, there is a marked predominance of odd over even harmonics. Overall, this high quality guitar displays more consistency in its spectra, and a more regular trend across its frequency range, than the other instruments studied.

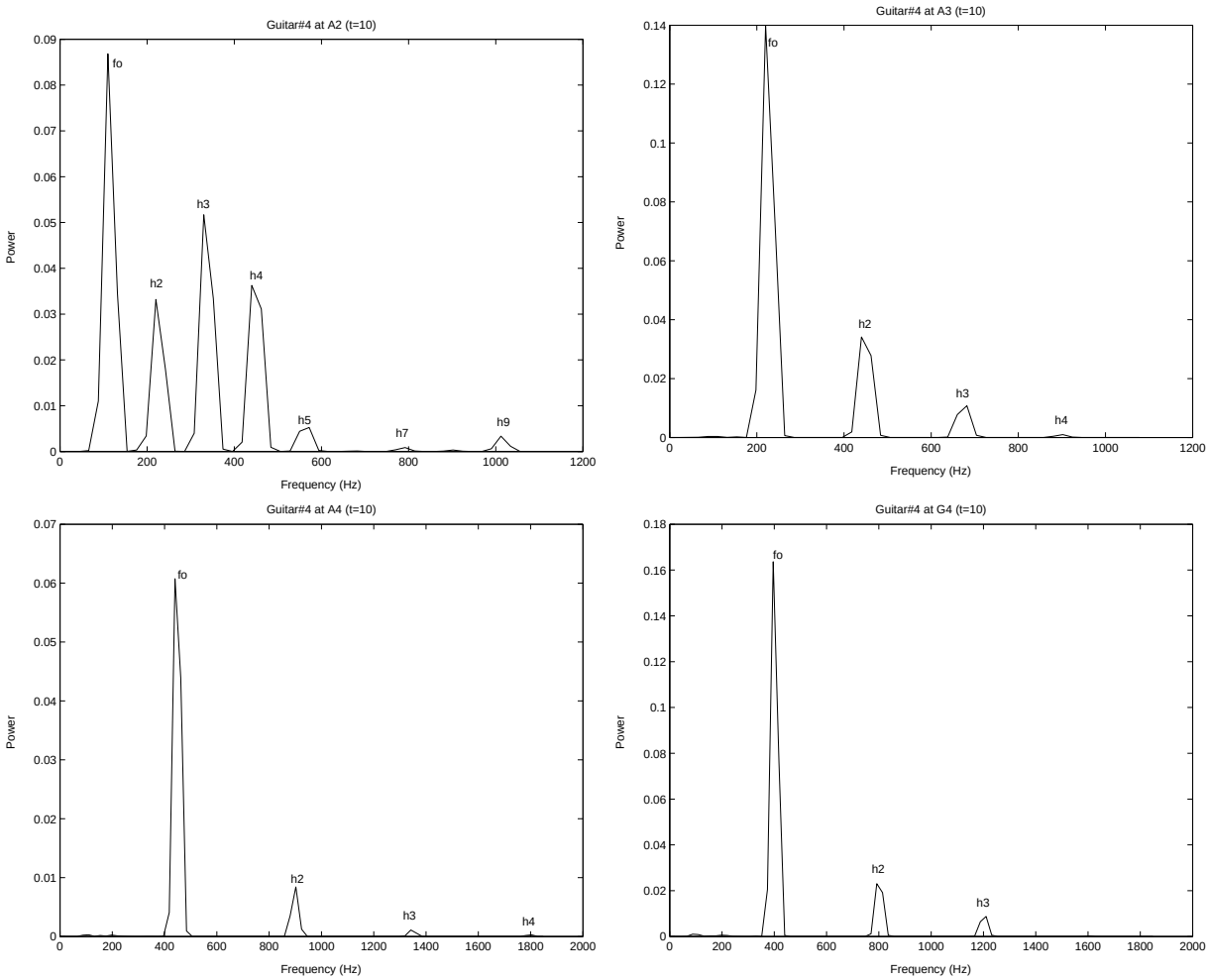

Figure 5.11: A plot of the spectra for the guitar#4 for the steady-state/decay at A2, A3, A4 and G4.

5.2.7 Summary of Guitar Timbre

Examination of the attack for each guitar showed the presence of distinctive resonances for each instrument, indicating why the information from the attack alone enabled good classification results to be obtained. The spectra for each of the guitars exhibited great variation across the frequency range, showing that it is difficult to define, in simple terms, the timbre of an instrument. We conclude that the timbre for a particular guitar must be defined as a set of timbres across the range of the instrument. Further, the variation in timbre between guitars, for a given pitch, reinforced our earlier finding that each guitar has its own unique 'signature'. The instrument with the most consistent timbre was guitar#4. It is noticeable, however, that there is a degree of similarity in the spectra between the four guitars. A general trend can be observed with strong harmonics in the lower register and weaker harmonics in the high register. All four guitars displayed weak harmonics at A3, suggesting a link to the design properties of all guitars. The two 'flat-top' guitars (#3 and #4) displayed a quite similar set of spectra.

We also observed that the rate of decay of the set of harmonics in a guitar tone is a factor in differentiating between the timbre of the instruments. We see in the following plot (figure 5.12) that the rate of decay for the harmonics in the classical guitar(#1) and the arch-top f-hole guitar(#2) were relatively fast. The two high quality flat top steel string guitars both showed a slower rate of decay. In musical terms, the tones exhibited long sustain.

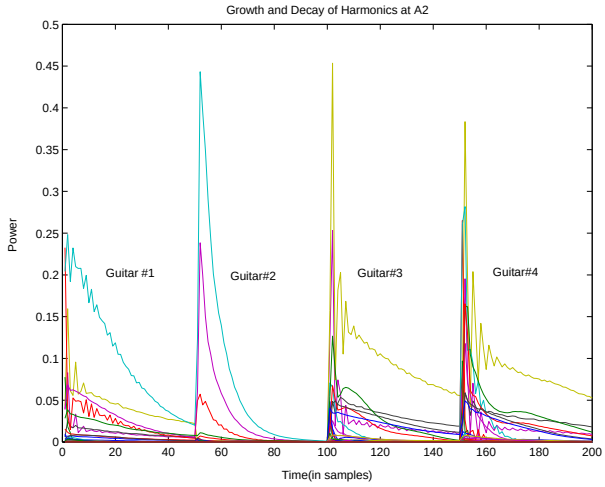

Figure 5.12: A plot showing the growth and decay of harmonics for each guitar over time.

5.2.8 Analysis of Violin Timbre

We have observed in the classification experiments described in chapter 4, that the timbre of each violin is frequency dependent. Using the spectral centroid as a single indicator of timbre, we observed that timbre varies across the frequency range of the instrument. We also showed that this graph is highly reproducible, indicating that there is a fixed/consistent timbre at each pitch across the range of each instrument.

In figure 5.13, a plot of a violin spectrum for the beginning of the attack ($t = 1$) shows the initial frequency response after the bow begins to move. It shows the growth of the fundamental and higher harmonics but more significant is the presence of numerous resonances in the body and air cavity of the violin. We observe in the second plot, ($t = 10$), that these resonances are transient and are quickly dampened to make way for a primarily harmonic spectrum.

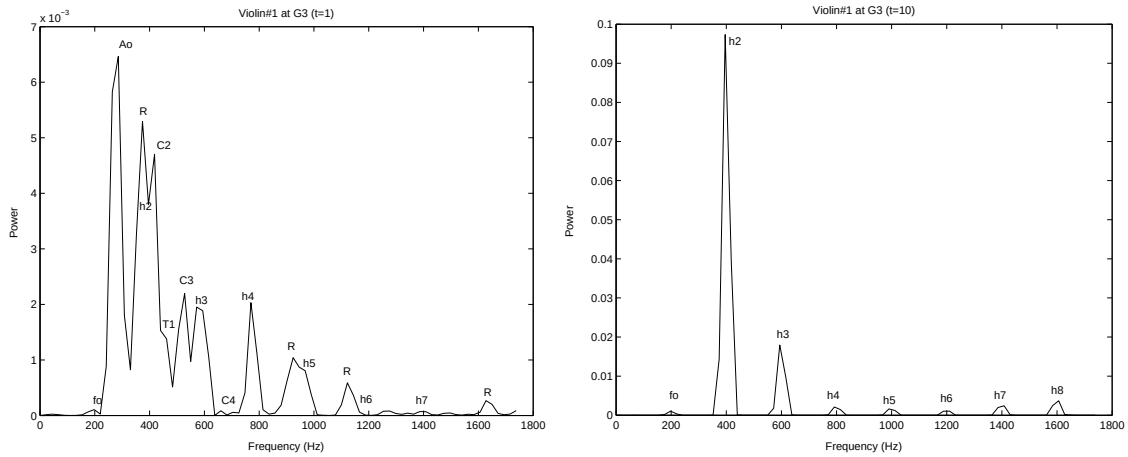

Figure 5.13: A plot of the spectra for violin#1 at G3, showing the resonances in the attack ($t = 1$) *resolving to a harmonic spectrum* ($t = 10$).

These resonances correspond to the various modes of vibration associated with a violin body. The significant modes can be described by the following notation:

Air Cavity: $A_0, A_1, A_2...$

Top Plate: $T_1, T_2, T_3...$

Whole Body: $C_1, C_2, C_3...$

The most significant modes in determining the timbre of a violin are: A_0, T_1, C_3, C_4. By referring to the review of research on violin acoustics (Fletcher & Rossing 1998), which includes analysis of resonances in high quality violins and also to the attack spectra of the violins in this study (figures 5.14 to 5.18), we can attempt to match the resonances found in the violins in this study to particular vibration modes of the violin (table 5.2).

	Violin#1	Violin#2	Violin#3	Violin#4	Violin#5
Ao	270	290	280	280	280
C2	370	420	345	415	350
T1	410	425	470	460	470
C3	520	560	500	560	500
C4	760	670	670	670	680
A2	920	840	880	820	830
A3	1120	1060	1120	1080	1160

Table 5.2: An attempt to match the resonances (in Hz) found in the initial spectra of each violin with the modes of vibration for a violin body.

The presence of these modes, their relative strength and the exact frequency will influence the formants and hence the frequency response of each violin. We note that, in this study, the resonances found in the violins are more uniform than those found in the guitars.

5.2.9 Violin#1

At the bottom of its range (G3), this violin displays a very weak fundamental (first harmonic), a very strong second harmonic, a significant third harmonic and very weak harmonics above that. At G4, the fundamental is strong, with a moderate second harmonic and weak higher harmonics. At G5, the upper register of the violin, there are surprisingly strong harmonics above the fundamental. Overall, the violin is characterised by weak harmonics above the fundamental in the lower and middle register. In terms of timbre, these physical features can be interpreted as a thin sound or a lack of brightness.

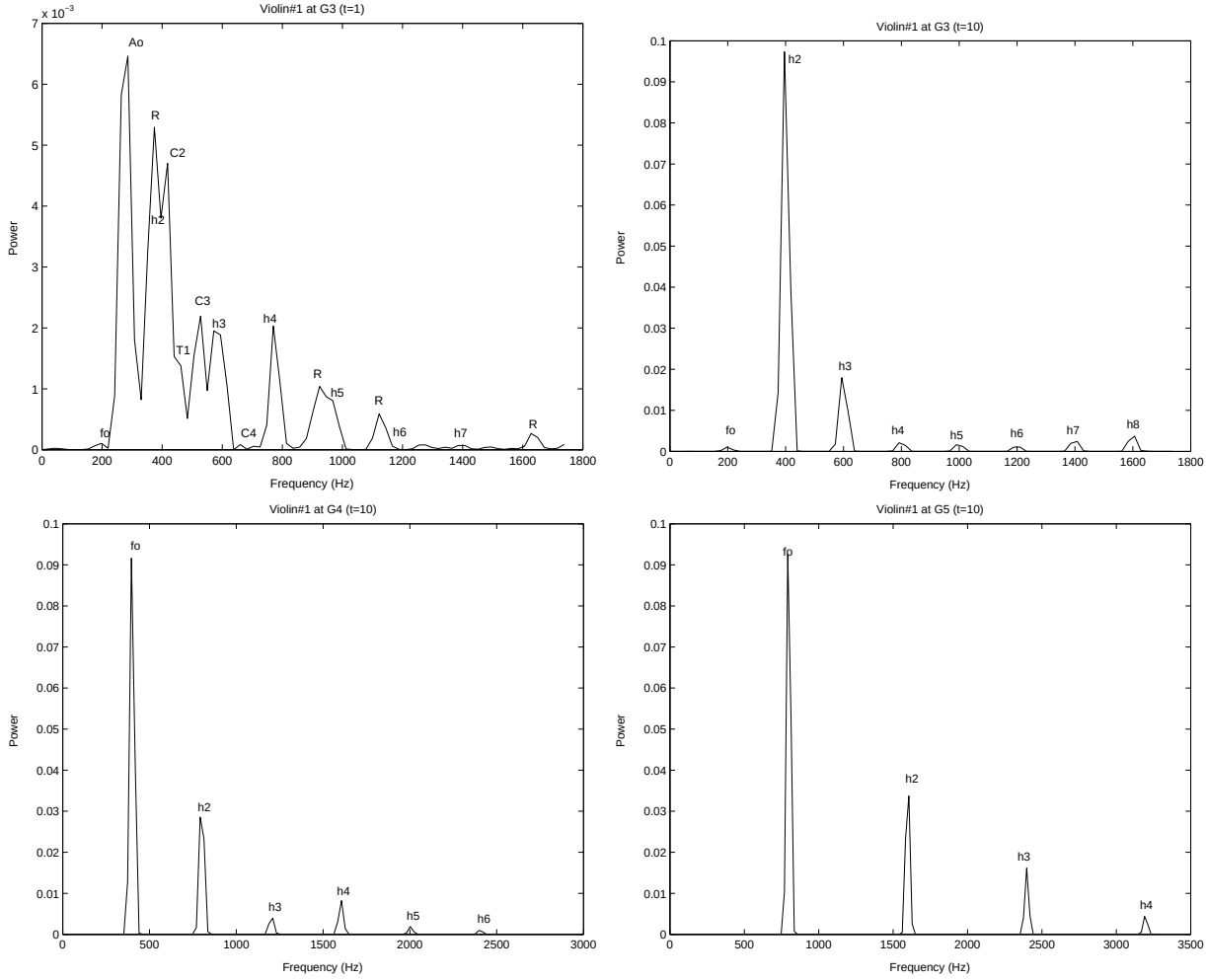

Figure 5.14: A plot of the spectra for the violin#1 for the attack at G3 and the steady state at G3, G4 and G5.

5.2.10 Violin#2

At G3, this violin features a weak fundamental, a dominant fourth harmonic, has significant third and sixth harmonics and all other harmonics are weak. At G4 the spectrum features a weak fundamental, dominant second and fourth harmonics, a moderate third harmonic and all other harmonics are weak. At G5, the fundamental (800Hz) is predominant, with a significant third harmonic. All other harmonics are weak or absent. In summary, there appears to be a formant at about 800Hz since the spectra at G3, G4, G5 are all strong in this region. The spectrum appears very unbalanced across the range of the instrument - a possible factor in this may be the fact that this violin had its sound post missing thereby limiting the coupling of body parts.

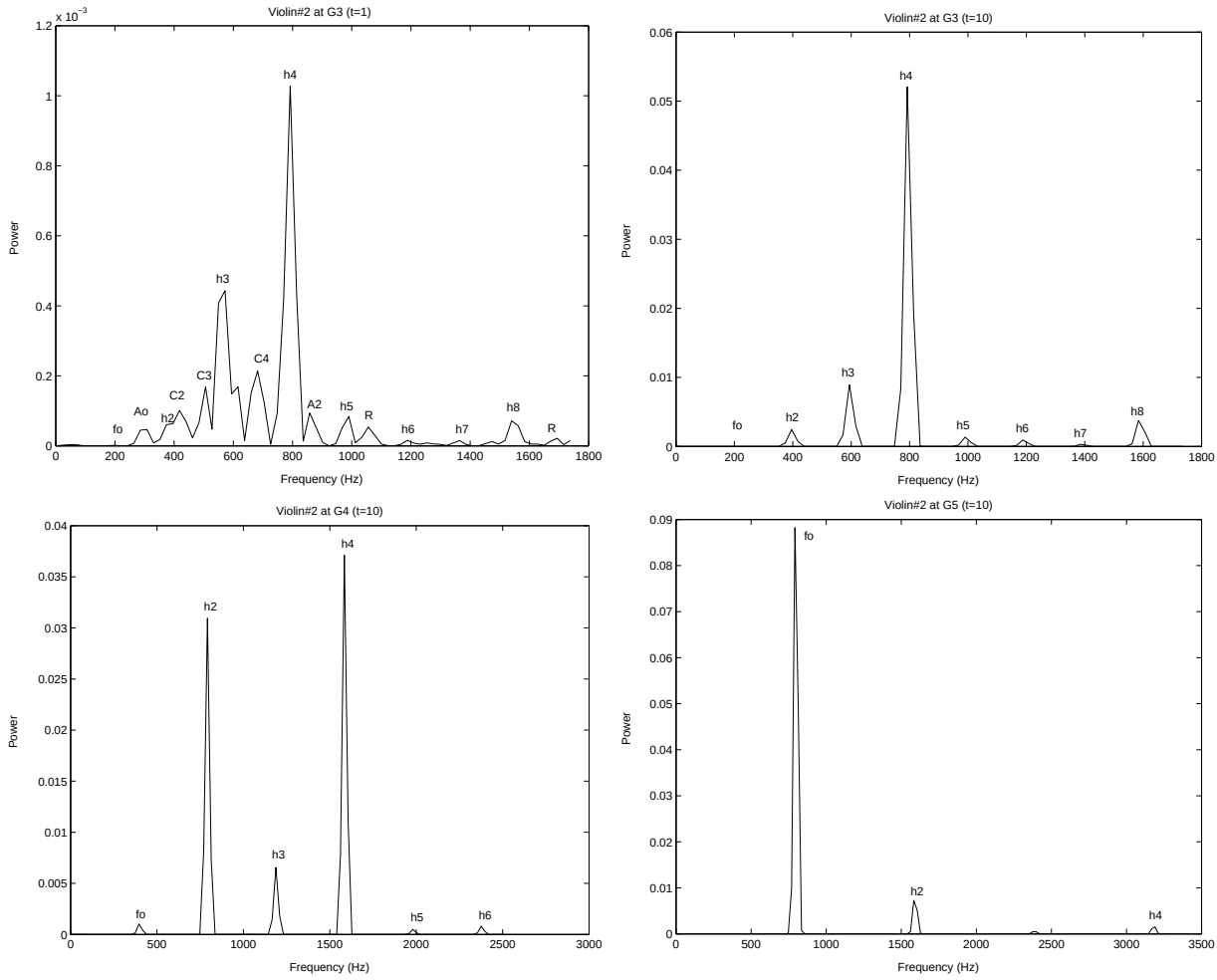

Figure 5.15: A plot of the spectra for the violin#2 for the attack at G3 and the steady state at G3, G4 and G5.

5.2.11 Violin#3

At G3, violin#3 features a weak fundamental, a dominant second harmonic, has significant third and sixth harmonics and all other harmonics are weak. At G4, the fundamental is strong and all other harmonics are weak. At G5, the second harmonic is stronger than the fundamental, suggesting a formant at about $1600Hz$, and also displays a very strong fourth harmonic. In summary, the timbre appears fairly thin (weak harmonics) in the low and mid range of the instrument but richer and very bright (high spectral centroid) at the high end of the range. Overall, the spectra varies markedly across the range of the instrument.

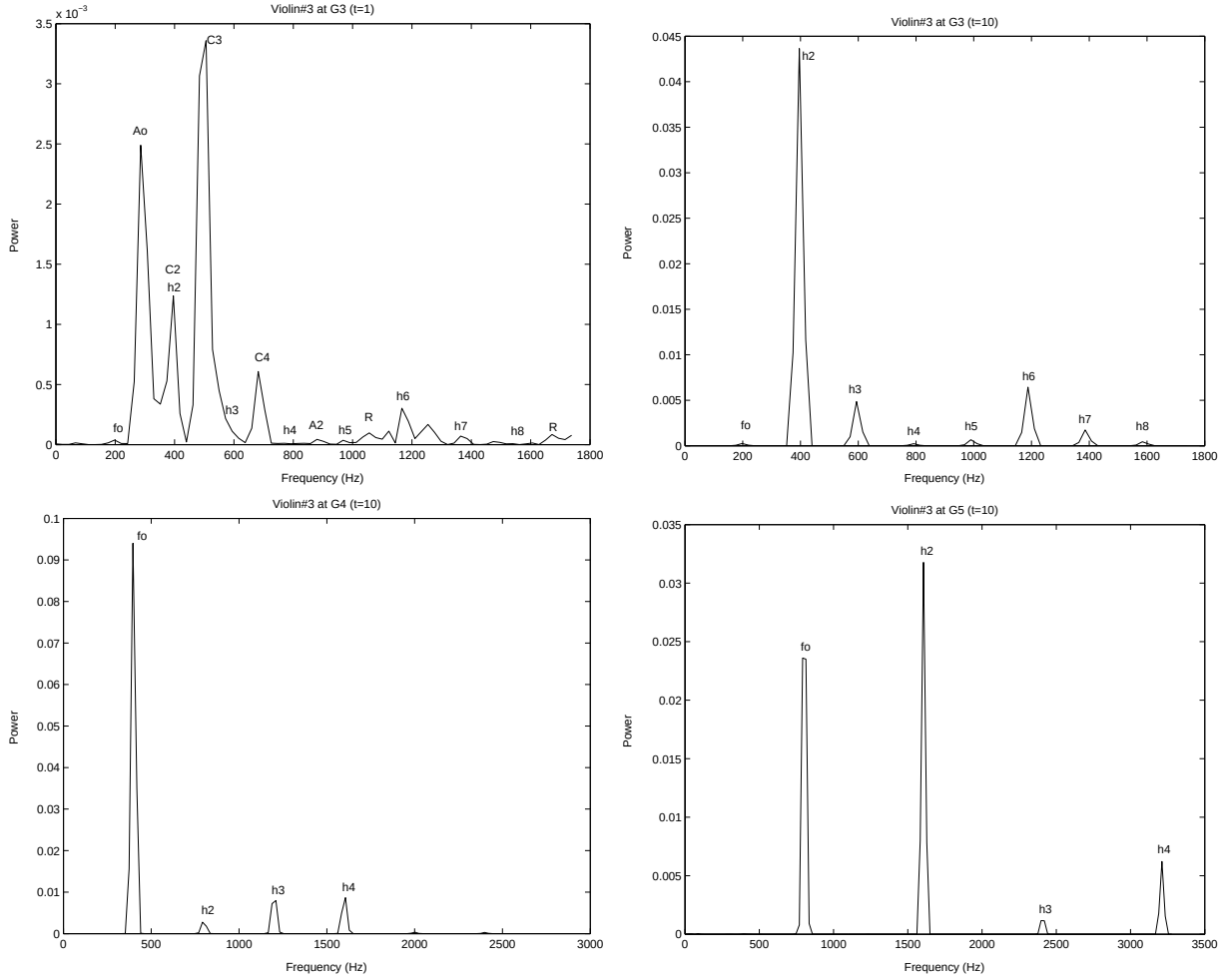

Figure 5.16: A plot of the spectra for the violin#3 for the attack at G3 and the steady state at G3, G4 and G5.

5.2.12 Violin#4

At G3, violin#4 has no significant fundamental but otherwise displays a balanced spectrum with strong upper harmonics. We note that the even harmonics predominate. At G4, the fundamental is again week - the second harmonic is stronger than the fundamental and the fourth harmonic stronger again. At G5, the fundamental is strong and all other harmonics are significant but not strong. In summary, in the lower and middle register, the timbre is weak in the fundamental and strong in the higher harmonics. In musical terms, this means that the violin lacks power or strength. The spectra is particularly unbalanced in the mid range. The spectra varies markedly across the range of the instrument.

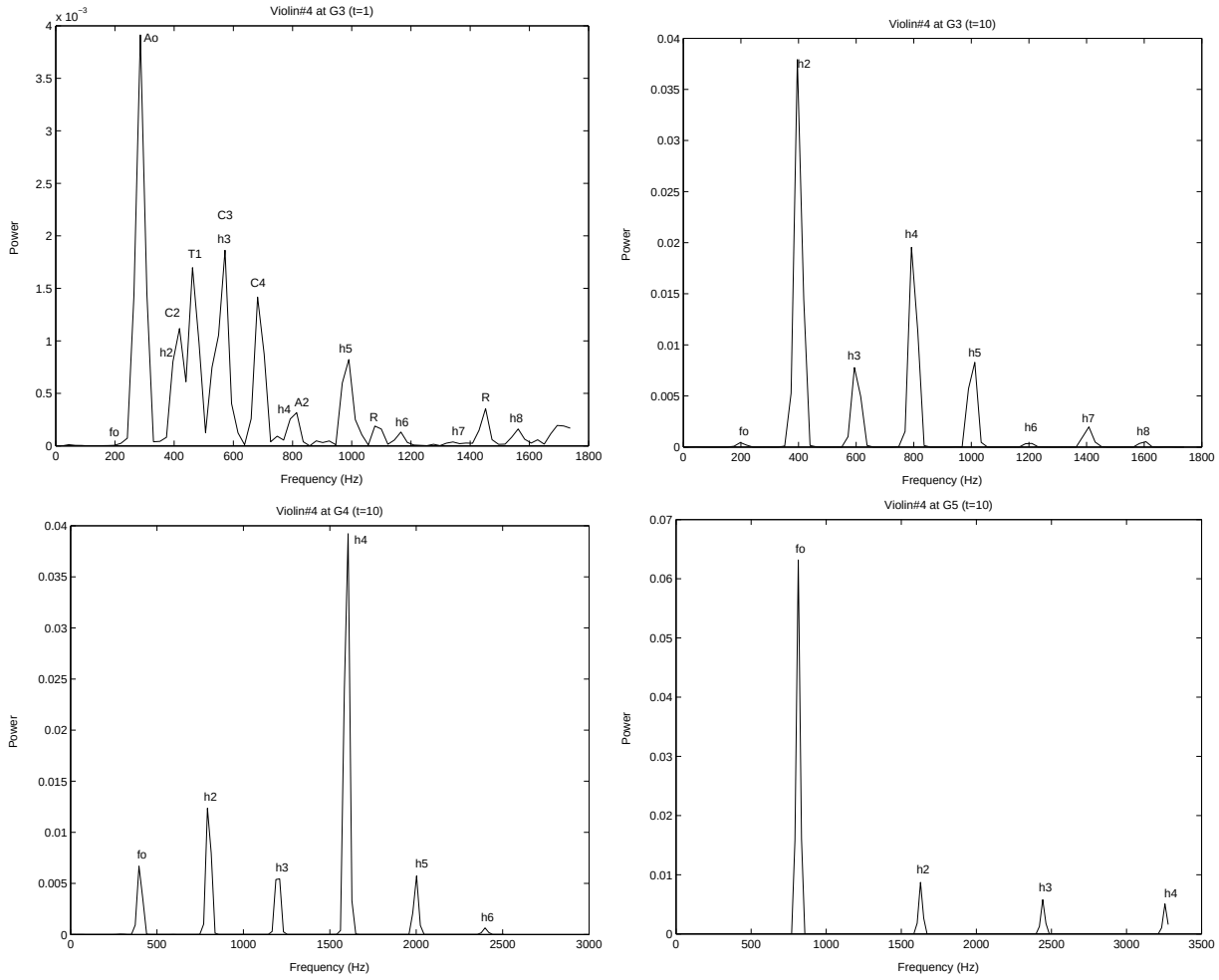

Figure 5.17: A plot of the spectra for the violin#4 for the attack at G3 and the steady state at G3, G4 and G5.

5.2.13 Violin#5

At G3, violin#5 has no significant fundamental harmonic but, other that this, displays a balanced spectrum but with relatively weak harmonics. At G4, the spectrum features a strong fundamental and second harmonic and a moderate third harmonic. Other harmonics are weak. At G5, the fundamental is strong with almost no higher harmonics. In summary, the timbre appears mellow and full in the low and mid range but thin at the high end of the range. This is the 'highest quality' instrument in the study, being an orchestra quality instrument (on a basis of human evaluation).

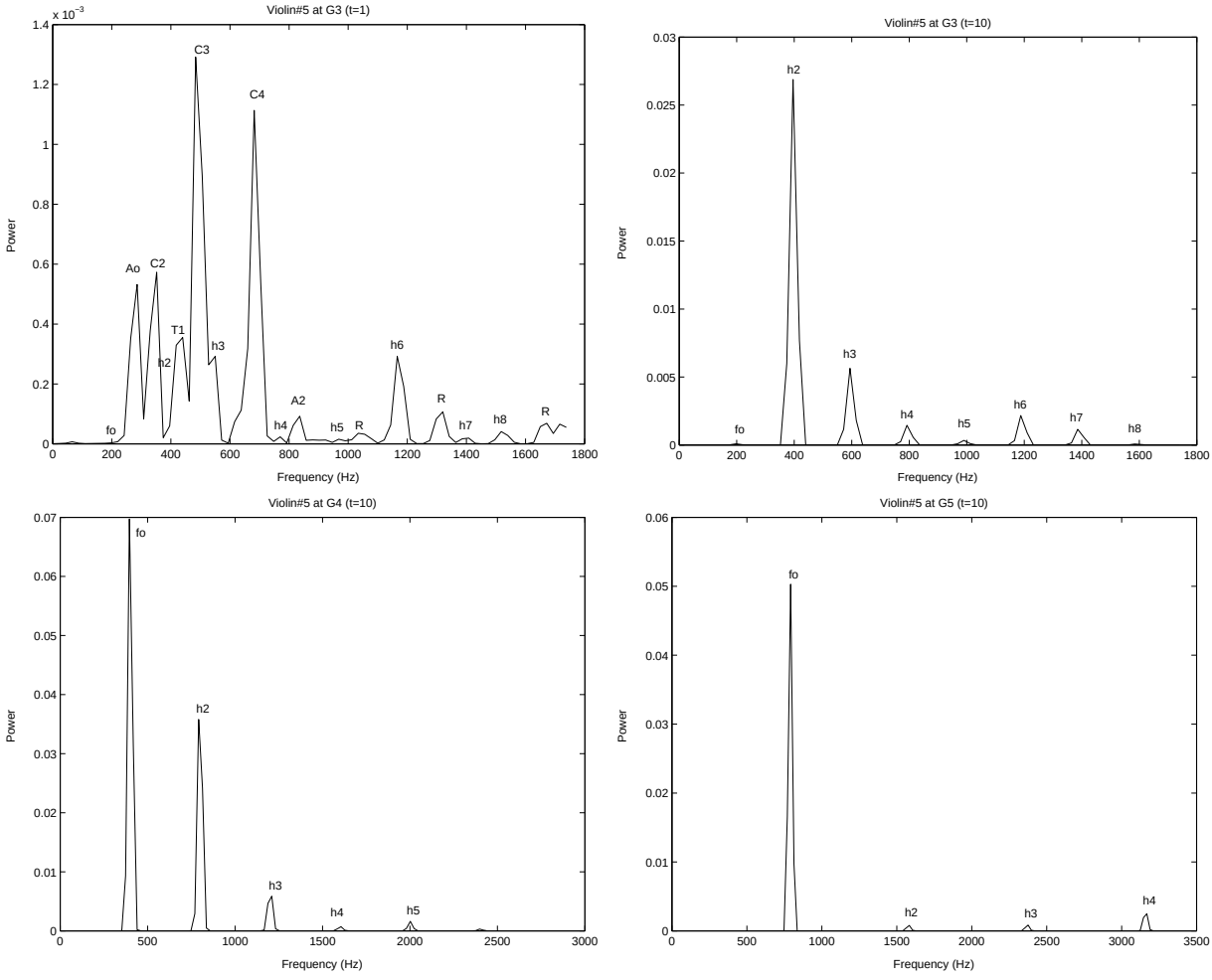

Figure 5.18: A plot of the spectra for the violin#5 for the attack at G3 and the steady state at G3, G4 and G5.

5.2.14 Summary of Violin Timbre

Examination of the attack for each violin (table 5.2) showed a degree of similarity in the resonances between the violins compared to the guitars. This explains why classification results, with only the attack, proved less satisfactory. In the steady state, each of the violins displayed considerable variation in timbre across the frequency range of the instrument. The degree of variation between violins at a given pitch reinforced our earlier finding, that each violin has its own unique 'signature'. The instrument with most consistent timbre was violin#5 - the only high quality violin in the study. Overall, this exploration confirmed that the timbre for a particular violin is best defined by the set of timbres across the frequency range.

Chapter 6

Conclusions and Further Work

6.1 Initial Goals, Summary of Process, Results and Conclusions

In this thesis, the primary goals have been to demonstrate that, within the instrument classes of guitar and violin, there is considerable variation in timbre and then, to attempt to differentiate between and classify, instruments within an instrument class.

To achieve these goals, a two stage process was set up. Beginning with a series of tones recorded for each of the four guitars and five violins across the range of the instruments, a frequency-time analysis was performed. From this data, a trajectory path was defined for each tone from each instrument in n-dimensional frequency space. Each trajectory path traced the frequencies present in the tone and their strength over the duration of the tone.

For each tone, a trajectory path to represent the timbre was determined in multidimensional frequency space. This was achieved using both FFT and CQT - the outputs were normalised with respect to power. Using these trajectory paths, we were able to represent the set of frequencies that characterised the development of each tone. Principal components were determined to simplify the classification process and to reveal information about the physical features enabling separation of the tones.

The second stage of the process involved comparing unknown test tones, one by one, with a set of reference tones from the complete set of instruments in that class. The closeness/distance measure used was the sum of the squared distances between the trajectory

paths over a period of time. From the resulting matrix of distance measures, the best match between each trial tone and the reference tones was determined.

The classification process proved very robust for the set of guitars provided that tones of the same pitch/fundamental frequency were compared. The best classification results were obtained using the whole tone, but similar results were obtained with either the attack alone or the decay alone. This suggests that there is sufficient information in either the attack or the decay to enable good classification. In further examination of the timbre of the guitars, it was observed that each guitar had distinctive resonances present in the attack period.

It was found that the classification results were reduced significantly if the tones were not synchronised with respect to time, that is, matched from the beginning of the attack period. This finding can be attributed to the impulsive nature of guitar tones and suggests the power envelope is an important factor in identification of guitars.

The fact that identification rates declined markedly when tones of similar but different pitch were compared, suggests that the timbre of guitars is highly pitch/frequency dependent. This conclusion is supported by reference to plots of the spectral centroid for each instrument, which showed large variation over the range of each instrument. We conclude that the timbre of a particular guitar is, in fact, a set of differing timbres across the pitch range of the instrument. Further investigation of the spectra of each guitar, across the frequency range of the instrument, showed the presence of formants, which corresponded to resonances in the body or air cavity of each instrument.

Classification with the violin proved more difficult than with the guitar. This is most likely because violin tones are less reproducible than guitar tones. In contrast to guitar tones, violin tones require a high degree of ongoing human input to produce the tone (pitching, vibrato and bowing).

With the violin, it was found that the steady state portion of the tone produced the best classification results with the attack alone producing less satisfactory results. Examination of the timbre for each of the violins in our trials showed that the resonances present in the attack period were less distinctive than for the guitar.

Violin classification did not prove to be sensitive to non-synchronisation between tones. This can be explained by the relative steady state nature of violin tones compared to guitar

tones. Consistent with this finding was the high classification rate achieved with spectral data averaged across the duration of the tone.

As with the guitar, violin classification proved highly sensitive to pitch indicating that violin timbre, like the guitar, is highly pitch dependent. This finding was reinforced by reference to plots of the spectral centroid over time for each of the violins in the study and by examination of the spectra across the range of each instrument.

6.2 Contribution of this Thesis to Instrument Recognition

Within the body of work on computer instrument recognition, this thesis is the first to compare the timbre of instruments of the same type in a quantitative way for the purpose of classification, or to classify instruments of the same type. In this thesis we have successfully classified four different guitars and five different violins. We have highlighted the importance of certain physical features related to the timbre of the guitar and violin which are well known in the field of instrument research, but not often taken into account in musical instrument classification. In particular, we have shown that each example of a guitar or violin has its own unique sound quality, different enough to allow classification within the instrument class. We have shown that the timbre of all instruments studied in this thesis is pitch dependent, that it varies in a non-regular way across the range of the instrument. Lastly, we have provided a broad ranging review of timbre over 150 years, not seen in other works in this field, that has informed the classification process in this thesis.

6.3 Implications of Findings on Other Research and Further Work

All of the previous work on instrument identification has been done with between class identification (instruments of different types). The findings in this study have implications for work in this area. In particular, the finding that the timbre of the guitar and violin are strongly frequency dependent. Many of the between class studies have been limited by the data which was available. For example, the sample tones from McGill university,

used by many researchers, are comprised of only one recording of each instrument at each pitch. This means that, for any set of reference tones at a given pitch, there are no independent test tones available. Consequently, in the classification process, tones of similar but different pitch must be compared. As we have shown, this will reduce the probability of correct classification.

With respect to between class instrument classification experiments, the findings of this study suggest the wisdom of first sorting instruments on the basis of impulsive or steady state tones. This approach was successfully used by Martin (1999) in his hierarchical approach.

A study by Brown (1999), in which she compared tones from oboe and saxophone, used some techniques which could readily be adapted for further work in within class classification. To represent time related spectral data, Brown used clusters calculated by the k-means algorithm. This approach could be used for within class classification and compared, as a basis for classification, to the trajectory path used in this study. In the feature extraction stage, Brown used cepstral coefficients (determined from spectral data) to encode the frequency/time related features. This approach, which worked well in her study, could be used in within class classification and compared with the trajectory paths used in this study.

Alternatively, work has been done on instrument recognition using wavelets for the feature representation (eg. Kostek (2003)). This approach could be tried for within class classification to determine if it better represents the time dependent nature of musical tones.

Bibliography

ASA:Psychoacoustical Terminology (1973), American National Standards Institute, New York.

Bachem, A. (1955), 'Absolute pitch', *The Journal of the Acoustical Society of America* **27**(6), 1180–1185.

Backus, J. (1970), *The Acoustical Foundations of Music*, John Murray, London.

Beauchamp, J. (1974), 'Time variant spectra of violin tones', *The Journal of the Acoustical Society of America* **56**(3), 995–1004.

Beauchamp, J. (1982), 'Synthesis by Spectral Amplitude and Brightness Matching of Analysed Musical Instrument Tones', *Journal of Audio Engineers Society* **30**(6).

Benade, A. (1990), *Fundamentals of Musical Acoustics*, Dover, New York.

Berger, K. (1964), 'Some factors in the recognition of timbre', *The Journal of the Acoustical Society of America* **36**(10), 1888–1891.

Bourne, J. B. (1972), Musical timbre recognition based on a model of the auditory system, Master's thesis, M.I.T., Cambridge, Massachusetts.

Bregman, A. S. (1990), *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press (1989), Cambridge, Mass, USA.

Brown, J. (1990), 'Calculation of a Constant Q Spectral Transform', *The Journal of the Acoustical Society of America* **89**(1), 425–434.

Brown, J. (1999), 'Computer Identification of Musical Instruments Using Pattern Recognition With Cepstral Coefficients as Features', *The Journal of the Acoustical Society of America* **105**(3), 1933–1941.

Brown, J., Houix, O. & McAdams, S. (2001), 'Feature Dependence in the Automatic Identification of Musical Woodwind Instruments', *The Journal of the Acoustical Society of America* **109**(3), 1064–1072.

Brown, R. (1981), 'An Experimental Study of the Relative Importance of Acoustic Parameters for Auditory Speaker Recognition', *Language and Speech* **24**(4), 295–310.

Caldersmith, G. (1988), 'Why are Old Violins Superior', *American Luthier* **14**, 12.

Charbonneau, G. R. (1981), *The Music Machine*, 'Timbre and the Perceptual Effects of Three Types of Data Reduction', MIT Press (1989), Massachussetts.

Chowning, J. (1973), 'The Synthesis of Complex Audio Spectra by Means of Frequency Modulation', *Journal of The Audio Engineering Society* **10**, 526–534.

Clark, M., Luce, D., Abrams, R. & Schlossberg, H. (1963), 'Preliminary experiments on the aural significance of parts of tones of orchestral instruments and on choral tones', *Journal of the Audio Engineer* **11**(1), 45–54.

Clark, M., Robertson, P. & Luce, D. (1964), 'A Preliminary Experiment on Perceptual Basis for Musical Instrument Families', *Journal of Audio Engineering Society* **12**(3), 199–203.

Cliff, N. (1966), 'Orthogonal Rotation to Congruence', *Psychometrika* **31**, 33–34.

Cosi, P., Poli, G. D. & Lauzzana, G. (1994), 'Auditory Modelling and Self Organizing Neural Networks for Timbre Classification', *Journal of New Music Research* **23**, 71–97.

Eagleson, H. (1947), 'Identification of Musical Instruments When Heard Directly and Over a Public-Address System', *The Journal of the Acoustical Society of America* **19**(2), 338–342.

Eggink, J. & Brown, G. J. (2003), Application of Missing Feature Theory to the Recognition of Musical Instruments in Polyphonic Audio, *in* 'Proceedings of International Symposium on Music Information Retrieval', ISMIR, Baltimore, USA.

Ellis, D. P. W. (1996), Prediction-driven Computational Auditory Scene Analysis, PhD thesis, M.I.T, Cambridge, Massachusetts.

Essid, S., Richard, G. & David, B. (2005), Instrument Recognition in Polyphonic Music, *in* 'Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing', ICASSP, Philadelphia, USA, pp. 245–248.

Essid, S., Richard, G. & David, B. (2006), 'Musical Instrument Recognition by Pairwise Classification Strategies', *IEEE Transactions on Audio, Speech and Language Processing* **14**(4), 1401–1412.

Everitt, B. S. & Dunn, G. (1993), *Applied Multivariate Data Analysis*, Edward Arnold, London, England.

Feiten, B. & Gunzel, S. (1994), 'Automatic Indexing of a Sound Database Using Self-Organising Neural Nets.', *Computer Music Journal* **18**(3), 53–65.

Fletcher, H. (1934), 'Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency and the Overtone Structure', *The Journal of the Acoustical Society of America* **6**(2), 59–69.

Fletcher, H. (1964), 'Normal Vibration Frequencies of a Stiff Piano String', *The Journal of the Acoustical Society of America* **36**(1), 203–209.

Fletcher, H., Blackam, E. & Geertsen, O. (1965), 'Quality of Violin, Viola, 'Cello and Bass-Viol Tones. 1', *The Journal of the Acoustical Society of America* **37**(5), 851–863.

Fletcher, H., Blackham, E. & Stratton, R. (1962), 'Quality of Piano Tones', *The Journal of the Acoustical Society of America* **34**(6), 749–761.

Fletcher, H. & Sanders, L. (1967), 'Quality of Violin Vibrato Tones', *The Journal of the Acoustical Society of America* **41**(6), 1534–1544.

Fletcher, N. H. & Rossing, T. D. (1998), *The Physics of Musical Instruments*, Springer-Verlag, New York.

Freedman, M. D. (1966), 'Analysis of Musical Instrument Tones', *The Journal of the Acoustical Society of America* **41**(4), 793–806.

Freedman, M. D. (1968), 'A Method for Analyzing Musical Tones', *Journal of the Audio Engineering Society* **16**(4), 419–425.

Fujinaga, I. (1998), Machine Recognition of Timbre Using Steady-State Tone of Acoustic Musical Instruments, *in* 'Proceedings of the 1998 International Computer Music Conference', p. 207.

Grey, J. (1977), 'Multidimensional Perceptual Scaling of Musical Timbres', *The Journal of the Acoustical Society of America* **61**(5), 1270–1277.

Grey, J. (1978), 'Timbre Discrimination in Musical Patterns', *The Journal of the Acoustical Society of America* **64**(2), 467–472.

Grey, J. & Gordon, J. (1978), 'Effects of Spectral Modifications on Musical Timbres', *The Journal of the Acoustical Society of America* **63**(5), 1493–1500.

Grey, J. & Moorer, J. (1977), 'Perceptual Evaluations of Synthesized Musical Instrument Tones', *The Journal of the Acoustical Society of America* **62**(2), 454–462.

Hajda, J., Kendall, R. A., Carterette, E. C. & Harshberger, M. L. (1997), Methological Issues in Timbre Research, *in* I. D. anf J. Sloboda, ed., 'Perception and Cognition of Music', Psychology Press, East Essex, UK, pp. 253–307.

Helmholtz, H. (1863), *On the Sensations of Sound*, Dover (republished 1954), New York.

Hourdin, C. & Charbonneau, G. (1997), 'A Multidimensional Scaling Analysis of Musical Instruments' Time-Varying Spectra', *Computer Music Journal* **21**(2), 40–55.

Ifeacher, E. C. & Jervis, B. W. (1993), *Digital Signal Processing: A Practical Approach*, Addison-Wesley, Workingham, England.

Johnson, R. A. & Wichern, D. W. (1982), *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, New Jersey.

Jolliffe, I. T. (1986), *Principal Components Analysis*, Springer-Verlag, New York.

Jolly, S. E. (1997), Analysis of bat echolation calls recorded by anabat bat detectors, Master's thesis, Victoria University, Melbourne, Australia.

Kaminskyj, I. (1999), Multidimensional Scaling Analysis of Musical Instrument Sounds Spectra, *in* 'Proc. ACMA Conf 1999', ACMA, Wellington, NZ, pp. 36–39.

Kaminskyj, I. (2001), 'Multi-feature Musical Instrument Sound Classifier', *Mikropolyphonie WWW Journal* **5**.

Kaminskyj, I. & Czaszejko, T. (2005), 'Automatic Recognition of Isolated Monophonic Musical Instrument Sounds Using kNNC', *Journal of Intelligent Systems* **24**, 199–221.

Kaminskyj, I. & Materka, A. (1995), Automatic Source Identification of Monophonic Musical Instrument Sounds, *in* 'Proc. IEEE Int. Conf. Neural Networks,Vol1 1995', IEEE, Perth, Aust, pp. 189–194.

Kashino, K. & Murase, H. (1999), 'A Sound Source Identification System for Ensemble Music Based on Template Adaption and Music Stream Extraction', *Speech Communication* **27**.

Kendall, R. A. & Carterette, E. C. (1991), 'Perceptual Scaling of Simultaneous Wind Instrument Timbres', *Music Perception* **8**(4), 369–404.

Klassner, F. I. (1996), Data Reprocessing in Signal Understanding Systems., PhD thesis, University of Massachusetts, Amherst.

Kostek, B. (2001), 'Soft Computing-Based Automatic Recognition of Musical Instrument Classes', *J. ITC Sangreet Reseach Academy* **15**(October), 6–32.

Kostek, B. (2002), Automatic Classification of Musical Instruments Sounds Based on Wavelets and Neural Networks, *in* 'Proc. 6th IASTED Intern.Conf,, Artificial Intelligence and Soft Computing', IASTED, Banff,Calgary,Canada, pp. 407–412.

Kostek, B. (2003), Musical Sound Parameters Revisited, *in* 'Music Acoustics Conference', Stockholm, Sweden, pp. 623–626.

Kostek, B. (2005), *Perception-Based Data Processing in Acoustics*, Vol. 3 of *Studies in Computational Intelligence*, 1st edn, Springer, Berlin/Heidelburg, chapter 3, pp. 39–185.

Kostek, B. & Wieczorkowska, A. (1997), 'Parametric Representation of Musical Sounds', *Archives of Acoustics* **22**(1), 3–26.

Kruskal, J. (1964), 'Nonmetric Multidimensional Scaling: A numerical Method', *Psychometrika* **29**(2), 115–129.

Kruskal, J. & Hill, M. (1964), 'Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hyposesis', *Psychometrika* **29**(1), 1–27.

Krzanowski, W. J. (1988), *Principles of Multivariate Analysis*, Clarendon Press, Oxford.

Langmead, C. J. (1995), A theoretical model of timbre perception based on morphological representations of time-varying spectra, Master's thesis, Dartmouth College.

Lichte, W. (1941), 'Attributes of Complex Tones', *Journal of Experimental Psychology* **28**(6), 455–480.

Luce, D. (1963), Physical Correlates of Nonpercussive Musical Instrument Tones, PhD thesis, MIT, Cambridge, MA.

Marques, J. (1999), Sound source recognition: A theory and computational model, Master's thesis, M.I.T, Cambridge, Massachusetts.

Marques, J. & Moreno, P. J. (1999), A study of musical instrument classification using gausssian mixture models and support vector machines, CRL Technical Report Series CRL/4, Cambridge Research Laboratory, Cambridge, Mass, USA.

Martin, K. D. (1999), An Automatic Annotation System for Audio Data Containing Music, PhD thesis, M.I.T, Cambridge, Massachusetts.

Martin, K. & Kim, Y. E. (1999), Musical Instrument Identification: A Pattern-Recognition Approach, *in* '136th meeting of Acoustical Society of America', ASA, pp. 1–12.

McAdams, S. (1993), *Recognition of Sound Source Events*, Oxford University Press, UK, pp. 146–198.

McAdams, S., Beauchamp, J. W. & Meneguzzi, S. (1999), 'Discrimination of Musical Instrument Sounds Resynthesised with Simplified Specrotemporal Parameters', *Journal of Acoustical Society of America* **105**(2), 882–897.

McAdams, S., Winsberg, S., Donnadieu, S., DeSoete, G. & Krimpoff, J. (1995), 'Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes', *Psychological Research* **58**, 177–192.

Moore, R. A. & Cerone, P. (1999), Automatic Musical Instrument Identification, *in* J. Singh, ed., 'SCIRF'99: Collaborative Research Forum', VUT, Melbourne, Australia, pp. 133–137.

Moorer, J. (1973), 'The Heytrodyne Filter as a Tool for Analysis of Transient Waveforms', *Memo AIM208 Stanford California, Stanford AI Lab* .

Moorer, J. (1978), 'The Use of the Phase Vocoder in Computer Music Applications', *Journal of the Audio Engineering Society* **26**, 42–45.

Nooralahiyan, A. Y., Kirby, H. R. & McKeown, D. (1998), 'Vehicle Classification by Acoustic Signature', *Mathematical Computer Modelling* **27**.

Opolko, F. & Wapnick, J. (1987), *McGill University Master Samples [Compact Disc]*, McGill University, Montreal, Quebec.

Plomp, R. (1970), 'Timbre as a Multidimensional Attribute of Complex Tones', *Proc. of Int. Symp. on Frequency Analysis and Periodicity Detection in Hearing* pp. 397–414.

Plomp, R. (1976), *Aspects of Tone Sensation: A Psychophysical Study*, Academic Press, London.

Plomp, R., Pols, L. & van de Geer, J. (1966), 'Dimensional Analysis of Vowel Spectra', *The Journal of the Acoustical Society of America* **41**(3), 707–712.

Plomp, R. & Steeneken, H. (1969), 'Effect of Phase on Complex Tones', *The Journal of the Acoustical Society of America* **46**(2), 409–421.

Poli, G. D. & Prandoni, P. (1997), 'Sonological Models for Timbre Characterization', *Journal of New Music Research* **26**, 170–197.

Poli, G. D. & Tonella, P. (1993), Self-organising Neural Networks and Grey's Timbre Space, *in* 'Proceedings of the 1993 International Computer Music Conference', I.C.M.A., San Francisco, USA, pp. 441–444.

Pollard, H. F. & Jansson, E. V. (1982), 'A Tristimulus Method for the Specification of Musical Timbre', *Acustica* **51**, 162–171.

Pols, L., van de Kamp, L. & Plomp, R. (1969), 'Perceptual and Physical Space of Vowel Sounds', *The Journal of the Acoustical Society of America* **46**(2), 458–467.

Pratt, R. L. & Doak, P. E. (1976), 'A Subjective Rating Scale for Timbre', *Journal of Sound Vibration* **45**, 317.

Reynolds, D. A. (1995), 'Speaker Identification and Verification Using Gaussian Mixture Speaker Models', *Speech Communication* **17**, 91–108.

Richardson, E. (1954), 'The Transient Tones of Wind Instruments', *The Journal of the Acoustical Society of America* **26**(6), 960–962.

Risset, J. C. & Mathews, M. V. (1969), 'Analysis of Musical Instrument Tones', *Physics Today* **22**(2), 23–30.

Saldana, E. L. & Corso, J. F. (1964), 'Timbre Cues and the Identification of Musical Instruments', *The Journal of the Acoustic Society of America* **36**, 2021–2026.

Sandell, G. J. & Martens, W. L. (1995), 'Perceptual Evaluation of Principal-Component-Based Synthesis of Musical Timbres', *Journal of the Audio Engineering Society* **43**(12), 1013–1028.

Scheirer, E. D. & Slaney, M. (1997), Construction and Evaluation of Robust Multifeature Speach/Music Discriminator, *in* 'Proceedings of the 1997 IEEE International Conference on Acoustics, Speech and Signal Processing.', Munich, pp. 253–307.

Seashore, C. E. (1938), *Psychology of Music*, McGraw-Hill, New York.

Shepard, R. N. (1962*a*), 'The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function.I.', *Pychometrika* **27**(2), 125–140.

Shepard, R. N. (1962*b*), 'The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function.II.', *Pychometrika* **27**(3), 219–246.

Shepard, R. N. (1982), 'Geometrical Approximations to the Structure of Musical Pitch', *Psychological Review* **89**(4), 305–333.

Slawson, A. (1967), 'Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency', *The Journal of the Acoustical Society of America* **43**(1), 87–101.

Stat-Sciences (1993), *S-PLUS Guide to Statistical and Mathematical Analysis*, StatSci, a division of MathSoft Inc., Seatle.

Statsoft (2003), *On Line Textbook*, Statsoft, http://www.statsoft.com.

Strong, W. & Clark, M. (1967*a*), 'Perturbations of Synthetic Orchestral Wind-Instrument Tones', *The Journal of the Acoustical Society of America* **41**(2), 277–285.

Strong, W. & Clark, M. (1967*b*), 'Synthesis of Wind-Instrument Tones', *The Journal of the Acoustical Society of America* **41**(1), 39–52.

Swets, D. & Weng, J. (1996), 'Using Discriminant Eigenfeatures for Image Retrieval', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(8), 831–836.

Venables, W. N. & Ripley, B. D. (1999), *Modern Applied Statistics with S-Plus*, Springer, New York.

von Bismarck, G. (1974), 'Timbre of steady sounds: a factorial investigation of its verbal attributes', *Acustica* **30**, 146.

Wieczorkowska, A. (1999), Classification of Musical Instrument Sounds using Decision Trees, *in* 'Proceedings of ISSEM'99.', Technical University of Gdansk, Gdansk, Poland, pp. 225–230.

Wieczorkowska, A. (2001), 'Musical Classification Based on Wavelet Analysis', *Fundamenta Informaticae* **47**(1).

Wold, E., Brum, T., Keislar, D. & Wheaton, J. (1996), 'Content Based Classification, Search and Retrieval of Audio', *IEEE Multimedia* pp. 27–36.

Wolfe, J., Smith, J., Breilbeck, G. & Stocker, F. (1995), 'A System for Real Time Measurement of Acoustic Transfer Functions', *Acoustics Australia* **23**, 1–19.

Appendix A

Developing the Classification Technique

A.1 Preliminary Trials

A.1.1 Introduction

Tones of equal fundamental frequency were compared in the range E2 to C5.

SetA - one tone at each pitch from each of the four guitars.

SetB - a second independent recording at each pitch from each of the four guitars.

A.1.2 Test1: (distM1)

An algorithm was written to find the squared Euclidean distance between the MDS curves at each increment in time and then sum the distances to give a measure of the closeness of the curves. The co-ordinates of each point on the curve were defined by the scores for each of the first ten principle components ie. points plotted in 10 dimensional space. The sum of the squared distances for any two instruments tones was considered to be a measure of the closeness of timbre.

A closeness/distance matrix was assembled comparing each tone in set A with each tone

set B to determine the best match. A success was considered to be a match between a particular instrument tone in set A and the corresponding independent tone in set B. Across the range E2 to C5 a 70% correct classification rate was achieved.

On examination of graphs of each principle component, it was observed that for certain tones the corresponding PC's from set A and set B were mirror images of each other. Due to the fact that in principle components the direction of each PC is arbitrary, in some pairs the software had reversed the pc direction for one of a corresponding pair.

A.1.3 Trial 2: (distM4B)

To correct the reflection problem in corresponding PC's mentioned above, the distance algorithm was altered to first check each pair of PC's for direction and adjust where necessary. The experiment was re-run for the complete data set and the correct classification rate improved to 87%.

A.1.4 Trial 3: (distM4D)

In previous trials the distance algorithm has used the the sum of squared distances as a measure of closeness - this is akin to measuring the variance. An alternative approach is to sum the absolute distances between each pair of PC curves. This is akin to numerical integration or finding the area between the pair of curves. This variation in technique was tried and made little difference to the classification process giving an 85% correct classification rate.

A.1.5 Trial 4: (distM4E)

The attack period of a guitar tone as shown on each PC is relatively large in magnitude and hence any error at this point is more significant in a measure of closeness than error at a later time in the evolution of the tone. Taking this into account, the distance algorithm was modified to attempt synchronize the two curves being compared before calculating closeness. The algorithm was modified accordingly and a re-run on the classification process produced a correct classification rate of 87% - equal to the best previous result in trial 2.

A.1.6 Trial 5: Combining Data Sets

Up to this stage, we have treated the data at each fundamental frequency as two separate sets - setA and setB. Principal component analysis has been performed on each set independently and the scores from the PCA on each set were then used for classification. We hypothesised that perhaps the classification would be more robust if the PC scores for all data were calculated in exactly the same manner. To investigate this question all data for each fundamental frequency was combined and PCA was then performed. This approach produced a marked improvement in classification giving a 96.8% correct classification rate. Surprisingly when the distance algorithm that checks and corrects for the direction of each PC was used, the correct classification rate dropped marginally to 95.6%. Although no explanation can be offered for this , further investigation did show that when PCA is performed once on the combined data set, the instance of corresponding PC's having opposite directions was reduced.

A.1.7 Trial 6: Giving equal weighting to all PC's

As has been previously stated the primary purpose of PCA is not to separate data on the basis of class but on a basis of variance for the whole data set. Each PC can often be matched to some physical feature of the sound. It is therefore possible that the PC with the largest variance (PC1) could correspond to some feature that is relatively consistent for all the instruments and may not therefore be the best discriminator between classes. In the normal output from PCA, each successive PC is weighted in accordance with the variance for that PC. It may therefore be informative to vary the weighting on each PC.

In this trial we gave equal weighting to each of the first 10 PC's. This is done by dividing the scores for each PC by the variance for that PC. Up until now the PCA has not altered the location or shape of our frequency based MDS plot in space but merely rotated the axes and translated the origin. However this is not so in this trial. When the PC's are weighted in the way described above, the location and the shape of the curve is altered. In an n-dimensional plot the distance from the origin to the plot for each point in time is called the Mahalanobis distance.

When the distance algorithm was altered in accordance with above the classification rate dropped to 86.6%, a decrease of 10% from our best performance.

A.1.8 Trial 7: Taking logs of FFT input data

As previously stated, a potential problem with an impulsive instrument such as a guitar is that the large data values (power) in the attack stage and the consequent difference values between the the two curves are given a heavy weighting

It was therefore thought that it may be informative to repeat the process using the logs of FFT to smooth the curve and hence reduce the weighting towards the attack. It was found that classification with the log of FFT gave a correct classification rate of 93%, a marginally reduced performance compared to the unaltered FFT data.

A.1.9 Trial 8: Using LDA scores in place of PCA scores (distM4Blda)

Although PCA and LDA are essentially similar processes, LDA chooses axes that best separate the means for each class rather than in the case of PCA on the basis of maximum variance for all the data combined. We could hypothesise that the LDA with its emphasis on separating classes may classify a little better than PCA with its emphasis on separating features. It was found that in using LDA the large dimensionality of each observation presented a problem in implementation. The problem was overcome by first applying PCA and taking the first 15 PC's (representing more than 99% of the variance) as the input for LDA. The scores from the first 10 linear discriminant axes for each of the observations/windows were used as the basis for classification. The trial initially resulted in a disappointing decrease in correct classification rate from 95% to 54%. More trials were performed with LDA at a later stage.

A.1.10 Summary of Trials

We have found that although there would appear to be weaknesses with the MDS based technique used to classify in this experiment, none of the variations tried except one, resulted an any significant improvement in classification performance. In fact most variations had a slightly negative influence on the classification rate. The exception was trial 5 where combining both data sets before performing the FFT gave a marked improvement in correct classification rate (87% to 96%).

Appendix B

Computer Programs

B.1 Sample MATLAB Programs

B.1.1 Function to Generate FFT Data for a Single Instrument

%fft2aRawfn.m - function for data extraction and fft analysis on moving window (takes 50 overlapping windows of size 1024 from start of file, assembles matrix of fft outputs)

function normfftdat=fft2aRawfn(file);

y=WAVREAD(file);
fileMark=1; %start of window
count=1; %count number of windows
nfft=1024; %window size
Fs=22050; %sampling frequency

while count ¡= 50
x=y(fileMark:fileMark+(nfft-1)); %select window
w= hanning(nfft); %hanning ordinates
pxx=abs(fft(w.x)).ˆ 2; %periodogram/pow spec-no div'n
px=pxx(1:150);
% px=pxx(1:(nfft/4)); % takes first quarter of mag spec

```
if fileMark==1
fftdat=px';
else
fftdat=[fftdat; px'];
end % end if


fileMark = fileMark+nfft/2;
count=count+1;
end % end while


normfftdat=fftdat/norm(fftdat);
```

B.1.2 Program to Assemble matrix of FFT Data for 4 Guitars

```
%file:fftDataset.m
%author:R.Moore
%program to assemble matrix of fft data set
%last modified 11/02/02
%call by fftDataset


instr=input('Input file prefix in quotes: ');
start=input('Input first file number: ');
finish=input('Input last file number: ');
outfile=input('Output file name in quotes: ');


for num=start:finish
file=[instr,int2str(num)];
%file='guitG2_1';
fftdat=fft2aRawfn(file);
if num ==start
fftdataset=fftdat;
else
fftdataset=[fftdataset;fftdat];
end
end
%fftdataset=1;
dlmwrite(outfile, fftdataset);
```

B.1.3 Function to Assemble Matrix of Features Over Time from FFT (inc. Fund Freq, Spec Cent)

```
%file:fft3Featfn_t.m
%author:R.Moore
%function for data extraction and fft analysis on moving window
%assembles matrix of features (inc rsc) extracted from fft measures sc and other feats over time


function featmatrix=fft3Featfn_t(file, ffHz);


y=WAVREAD(file); %bypass to debug
%y=WAVREAD('guitA2-4'); %trial


start=10000;
%start of window x=y(start:start+1023); %select window
nfft=1024; %window size
Fs=22050; %sampling frequency
w= hanning(nfft); %hanning ordinates


for(j=1:50) % repeat sc calc 20 times
pxx=abs(fft(w.*x)).^ 2; %periodogram/pow spec-no div'n
px=pxx(1:(nfft/2)); % takes first half of mag spec
n=[0:(nfft/2-1)]'; % counts from zero
pxn=[px,n]; %mag, freq bin
[pow,k]=sort(px); % sorts by power
A =[pow,k-1]; % counts from 0 gives power & bin no.


%maxp=pow(nfft/2); %to find ff
%ffHz3=vgramfn(x);
%ffHz=349.23; %bypass vgram to debug
ffbin=round(ffHz*nfft/Fs); %bin corresponding to ffHz
```

```
ffr=pxn((ffbin+1)-1:(ffbin+1)+1, 1:2); %ff region -(index=bin+1)


Psumm=sortrows(ffr,[1]); %sort rows by power
partials=Psumm(3,1:2); %writes first partial
ff=Psumm(3,2); %bin corresponding to ff
powff=Psumm(3,1); %power of ff


%to separate first ten partials
last=ff+1; % initialize last to ff/ ref for next search
for i=1:9 %start inner for loop
s=[pxn(last+ff-1,:);
pxn(last+ff,:);
pxn(last+ff+1,:)];%next peak
ss=sortrows(s,[1]); %sort by power
max=ss(3,1:2); %max power and bin loc next partial
partials=[partials;max];
last=max(2)+1; %loc of last partial
end % end inner for loop


ffHz2=(px(ff)*(ff-1)+px(ff+1)*ff+px(ff+2)*(ff+1))/(px(ff)+px(ff+1)+px(ff+2))*Fs/nfft;
% fundamental frequency (centroid) /not curently used


sc=specCentfn(partials, Fs,nfft); %spectral centroid
rsc=sc/ff; %rel spec centroid - takes account of fund freq
stdsc = specCentfn2(partials);

f=partials(:,2);
E=partials(:,1);
E=E/maxp;
stdpartials=[E,f];


odd=E(1)+E(3)+E(5)+E(7)+E(9);
```

```
even=E(2)+E(4)+E(6)+E(8)+E(10);
oer=odd/even; % odd/even partials ratio


%feat=[E',ffHz,sc,rsc,oer] %features outputted to matrix


feat=[ffHz,ffHz2,rsc,stdsc,oer];


if(j==1) featmatrix=feat;
else featmatrix=[featmatrix;feat];
end %end if


start=start+nfft/2;
x=y(start:start+1023); %select window
end % end outer for-loop


featmatrix=featmatrix(:,3);%spectral centroid
```

B.1.4 Function to Assemble Mean Features from FFT

```
%file:fft3Featfn.m
%function for data extraction and fft analysis on moving window assembles vector of mean features extracted from fft


function meanfeat=fft3Featfn(file, ffHz);


y=WAVREAD(file); %bypass to debug
%y=WAVREAD('guitA2_4');%trial


start=10000; %start of first window
x=y(start:start+1023); %select window
nfft=1024; %window size
Fs=22050; %sampling frequency
w= hanning(nfft); %hanning ordinates


for(j=1:5) % repeat sc calc 5 times and take mean


pxx=abs(fft(w.x)).^ 2; %periodogram/pow spec-no div'n
px=pxx(1:(nfft/2)); % takes first half of mag spec
n=[0:(nfft/2-1)]'; % counts from zero
pxn=[px,n]; %mag, freq bin
[pow,k]=sort(px); % sorts by power
A=[pow,k-1]; % counts from 0/ gives power& bin
no. maxp=pow(nfft/2); %to find ff


%ffHz3=vgramfn(x);
%ffHz=349.23; %bypass vgram to debug
ffbin=round(ffHz*nfft/Fs); %bin corresponding to ffHz
ffr=pxn((ffbin+1)-1:(ffbin+1)+1, 1:2); %ff region -(index=bin+1)
```

```
Psumm=sortrows(ffr,[1]); %sort rows by power
partials=Psumm(3,1:2); %writes first partial
ff=Psumm(3,2); %bin corresponding to ff
powff=Psumm(3,1); %power of ff
%to separate first ten partials
last=ff+1; % initialize last to ff/ ref for next search


for i=1:9
s=[pxn(last+ff-1,:);pxn(last+ff,:);
pxn(last+ff+1,:)];%next peak
ss=sortrows(s,[1]); %sort by power
max=ss(3,1:2); %max power and bin loc next partial
partials=[partials;max];
last=max(2)+1; %loc of last partial
end


ffHz2=(px(ff)*(ff-1)+px(ff+1)*ff+px(ff+2)*(ff+1))/(px(ff)+px(ff+1)+px(ff+2))*Fs/nfft; %
fundamental frequency (centroid) /not curently used
sc=specCentfn(partials, Fs,nfft); %spectral centroid


rsc=sc/ff; %rel spec centroid - takes account of fund freq


stdsc = specCentfn2(partials);


f=partials(:,2);


E=partials(:,1);
E=E/maxp; stdpartials=[E,f];
```

```
odd=E(1)+E(3)+E(5)+E(7)+E(9); even=E(2)+E(4)+E(6)+E(8)+E(10);


oer=odd/even; % odd/even partials ratio


%feat=[E',ffHz,sc,rsc,oer] %features outputted to matrix
feat=[ffHz,ffHz2,rsc,stdsc,oer];


if(j==1) sumfeat=feat; else sumfeat=sumfeat+feat;
end %end if


start=start+nfft;
x=y(start:start+1023); %select window
end % end outer for-loop
meanfeat=sumfeat/5;
```

B.1.5 Function to Determine Spectral Centroid V1

```
%file:specCentfn.m
%author:R.Moore
%function to calculate standardized spectral centroid
% function call: specCentfn(partials)
%last edited 8/8/00


function sc=specCentfn(partials,Fs,nfft)


f=partials(:,2); P=partials(:,1);
sumM=0;
sumE=0;


for j=1:10
M=log(f(j)*Fs/nfft)*P(j);
E=P(j);
sumM=sumM+M;
sumE=sumE+E;
end


sc=exp(sumM/sumE); %spectral centroid in Hz
```

B.1.6 Function to Determine Specral Centroid V2

```
%file:specCentfn2.m
%author:R.Moore
%function to calculate standardized spectral centroid - method 2
% function call specCentfn2(partials)
%last edited 21/5/04


function stdsc=specCentfn2(partials)


n=[1:10]; P=partials(:,1);


sumM=0; sumE=0;


for j=1:10
M= n(j)*P(j);
E=P(j);
sumM=sumM+M;
sumE=sumE+E;
end


stdsc=(sumM/sumE); %standardized spectral centroid
```

B.1.7 Function to Assemble Matrix of CQT Data

```
%file:cqtDatafn.m
%uses Judith Brown functions 'genlgftkern', 'logft'

function output =cqtDatafn(instr,minfreq)


%set file details
start=1; finish=4;
%finish=5;


%set cqt variables


%minfreq= 82.407; %E2
%minfreq=174.6141; %G3
freqrat=1.0293022366; %1/4 tone


SR=22050;
maxfreq=SR/2; % not needed here
nfreqs=125; %Nyquist allows up to 125 @ minfreq=C4


windsizmax=2205;
hopsiz=300; %depends on f
graphit=0;


%calculate kernels


[kerncos,kernsin,freqs]=genlgftkern2(minfreq,freqrat,SR,nfreqs,windsizmax);


%read files one by one, calc cqt and append
```

```
for num=start:finish
file=[instr,int2str(num)];
%file='guitA3_4';
infile=WAVREAD(file);
[cq,t]=logft2(infile,hopsiz,kerncos,kernsin,windsizmax,nfreqs,SR);


cqt=cq(1:100,:);
normcqt=cqt/norm(cqt);


if num==start
cqtdataset=normcqt;
else
cqtdataset=[cqtdataset;normcqt];
end %end if
end % end for


output=cqtdataset;
```

B.2 Sample S-PLUS Programs

B.2.1 Function to Generate Distance Matrix

%call: distM4()

```
function(M) {
A <- M[1:200, 1:10]
B <- M[201:400, 1:10]

A1 <- A[1:50, ]
A2 <- A[51:100, ]
A3 <- A[101:150, ]
A4 <- A[151:200, ]
B1 <- B[1:50, ]
B2 <- B[51:100, ]
B3 <- B[101:150, ]
B4 <- B[151:200, ]

D <- matrix(0, 4, 4)
D[1, 1] <- dist1(A1, B1)
D[2, 1] <- dist1(A2, B1)
D[3, 1] <- dist1(A3, B1)
D[4, 1] <- dist1(A4, B1)
D[1, 2] <- dist1(A1, B2)
D[2, 2] <- dist1(A2, B2)
D[3, 2] <- dist1(A3, B2)
D[4, 2] <- dist1(A4, B2)
D[1, 3] <- dist1(A1, B3)
D[2, 3] <- dist1(A2, B3)
D[3, 3] <- dist1(A3, B3)
D[4, 3] <- dist1(A4, B3)
```

```
D[1, 4] < − dist1(A1, B4)
D[2, 4] < − dist1(A2, B4)
D[3, 4] < − dist1(A3, B4)
D[4, 4] < − dist1(A4, B4)
D
}
```

B.2.2 Function to Generate Distance Measure for two Instrument Tones

```
%call:dist1()


function(A, B) {
dist < − sqrt(sum(diag((A - B) %*% t(A - B))))
dist
}
```

Appendix C

Classification Results

C.1 Sample Distance Matices

Introduction

Included here is a selection of the distance matrices used in the classification process and the results of selected experiments. Two independent sets (A and B) of tones were compared across the pitch range for the guitar or violin. In each trial a tone from one set served as the test tone and all tones in the other set served as the reference tones. A correct classification occurs when the minimum value for a row or column is located in the leading diagonal.

C.1.1 Guitar Classification Using FFT

The distance matrices for four guitars using frequency-time data from FFT at pitches E2 to C5. Principal components were calculated and the first 10 PC's were used for the distance measures which were calculated using S-Plus (function'distM4'). Number of correct classifications were 156 from 160 trials (ie. 96.8%).

E2

	[1]	[2]	[3]	[4]
[1]	0.2319940	1.187113	0.3768364	0.2447211
[2]	1.1003648	0.259409	1.2931657	1.1104294
[3]	0.4798613	1.277109	0.2805694	0.4969961
[4]	0.2080784	1.139894	0.3651746	0.1062601

F2

	[1]	[2]	[3]	[4]
[1]	0.1215395	1.1427199	0.8980091	0.1651841
[2]	1.1499215	0.1712395	1.2192679	1.1656157
[3]	0.9206376	1.1762229	0.1933357	0.8858378
[4]	0.1921681	1.1496477	0.8894856	0.1314987

G2

	[1]	[2]	[3]	[4]
[1]	0.5811746	0.9584789	0.8696964	0.6066331
[2]	0.7094026	0.2560353	1.1804481	0.8074793
[3]	0.9795277	1.2386231	0.1723373	0.8024536
[4]	0.6689060	1.0153503	0.6852877	0.4150098

A2

	[1]	[2]	[3]	[4]
[1]	0.2797463	0.74409205	1.0779319	0.9019699
[2]	0.7772716	0.04997068	1.3487135	1.1833507
[3]	1.0236725	1.34763891	0.3355063	0.5371169
[4]	0.8500524	1.15397641	0.6075554	0.1459451

B2

	[1]	[2]	[3]	[4]
[1]	0.3998061	1.1035475	0.7138612	0.7437991
[2]	1.0416344	0.1734862	1.2197539	1.0026432
[3]	0.9090673	1.2917401	0.3963267	0.6234610
[4]	0.8318766	1.0162285	0.5640550	0.2115422

C3

	[1]	[2]	[3]	[4]
[1]	0.4901866	0.8700417	0.5704888	0.4970346
[2]	0.9385530	0.1024969	1.1799433	0.9893180
[3]	0.7344667	1.1331468	0.2783726	0.4051390
[4]	0.8273379	0.9171991	0.5385290	0.1624677

D3

	[1]	[2]	[3]	[4]
[1]	0.2265167	0.78498878	0.9635440	0.4067492
[2]	0.9591649	0.09693694	1.3604841	0.9970119
[3]	0.8598672	1.34680965	0.1677166	0.7825251
[4]	0.4226288	0.98483521	0.9140040	0.1376924

E3

	[1]	[2]	[3]	[4]
[1]	0.2640291	0.9903046	0.5846098	0.4481244
[2]	0.8802363	0.1348289	1.2703901	1.0906002
[3]	0.7714983	1.2971181	0.3129826	0.6682987
[4]	0.3762466	1.0658153	0.6386831	0.1242995

F3

	[1]	[2]	[3]	[4]
[1]	0.1545519	0.8710265	0.8843922	0.22553454
[2]	1.1019369	0.7343026	1.3583235	1.28237782
[3]	0.8670005	1.0476209	0.3476952	0.78586015
[4]	0.3101296	0.9935321	0.8864480	0.06751669

G3

	[1]	[2]	[3]	[4]
[1]	0.2048437	0.7413322	1.0842917	0.3110680
[2]	0.7564973	0.1738877	0.8770049	0.7073078
[3]	1.0928550	0.9625377	0.2383212	1.0810231
[4]	0.2821920	0.6431698	0.9994312	0.1332935

A3

	[1]	[2]	[3]	[4]
[1]	0.2051571	0.72416741	0.6929542	0.2186534
[2]	0.7869228	0.04090166	0.7137660	0.6097414
[3]	0.6451316	0.71605755	0.1778828	0.5305303
[4]	0.3017581	0.67786065	0.4270778	0.2043677

B3

	[1]	[2]	[3]	[4]
[1]	0.1609956	0.6298947	0.7246812	0.5118025
[2]	0.5570759	0.4244704	0.7591812	0.7204792
[3]	0.9336881	0.9844434	0.7234000	0.9198810
[4]	0.6708996	0.7901702	0.5106739	0.1507351

C4

	[1]	[2]	[3]	[4]
[1]	0.2220049	0.5419665	0.8622290	0.75755678
[2]	0.5535028	0.1809283	0.6923015	0.59500281
[3]	0.9958473	0.9119454	0.4580650	0.56429340
[4]	0.7079907	0.6306554	0.3723372	0.08019991

D4

	[1]	[2]	[3]	[4]
[1]	0.3611829	0.8116973	0.7587637	0.7645457
[2]	0.9401807	0.2158417	1.0612436	1.1554640
[3]	0.9931437	1.1393083	0.3217885	0.6965407
[4]	0.9462564	1.1363679	0.5727186	0.1091557

E4

	[1]	[2]	[3]	[4]
[1]	0.3973635	0.8660928	1.0360925	0.8851977
[2]	0.8900554	0.3522013	1.1982738	1.0928920
[3]	0.8017098	1.0977907	0.6732479	0.9832617
[4]	0.9139007	1.1384922	1.0733547	0.2501486

F4

	[1]	[2]	[3]	[4]
[1]	0.2230268	1.1240683	0.5005151	0.6415994
[2]	1.1101398	0.1760666	1.2043424	1.2024020
[3]	0.7969958	1.2962800	0.4609191	0.7712763
[4]	0.6826414	1.2318482	0.6231090	0.1764509

G4

	[1]	[2]	[3]	[4]
[1]	0.2676255	0.8022803	0.7212153	0.7540517
[2]	0.7855664	0.3171616	0.7644349	0.7819213
[3]	0.9550272	0.9467047	0.4678587	0.5801286
[4]	0.8242643	0.7930910	0.3792537	0.1551839

A4

	[1]	[2]	[3]	[4]
[1]	0.3206377	0.5899172	0.5324959	0.9926802
[2]	0.6409816	0.1314841	0.3029405	0.8743280
[3]	0.6993430	0.3572406	0.2242173	0.8315839
[4]	0.9891817	0.8501657	0.7581776	0.1488051

B4

	[1]	[2]	[3]	[4]
[1]	0.3247657	1.2523707	1.3435241	1.332984
[2]	1.1523111	0.1926999	0.7237007	1.330823
[3]	1.1878686	0.3615108	0.6086151	1.163076
[4]	1.3481936	1.1693635	0.9571680	0.370960

C5

	[1]	[2]	[3]	[4]
[1]	0.2544934	1.2348999	0.8951743	1.0460315
[2]	0.9521344	0.2760065	0.4744690	0.6419466
[3]	0.9498425	0.4227938	0.3625883	0.5292383
[4]	1.0300379	0.5711882	0.6866555	0.3079862

C.1.2 Classification of Guitar with Mahalanobis Distance FFT with 5PC's - (95%)

E2

	[1]	[2]	[3]	[4]
[1]	1.123216	2.129202	1.9353641	1.6019028
[2]	2.220678	1.160501	2.5045294	1.9688877
[3]	2.645517	2.591780	0.9307266	2.4471445
[4]	1.027980	1.984751	2.1307787	0.6186311

F2

	[1]	[2]	[3]	[4]
[1]	0.3502109	2.2985202	2.1306701	0.4597746
[2]	2.0930872	0.6807972	2.7358601	2.1248698
[3]	2.2659944	2.8570859	0.5869667	2.0852123
[4]	0.4538875	2.3388144	2.1379810	0.2151131

G2

	[1]	[2]	[3]	[4]
[1]	1.227043	2.357588	1.8280726	1.3642339
[2]	2.099244	1.022564	2.4607332	2.1503887
[3]	1.916495	2.375433	0.3746071	1.8921123
[4]	1.123820	2.247765	1.7058182	0.9862724

A2

	[1]	[2]	[3]	[4]
[1]	1.102081	1.47120419	2.287605	1.5310303
[2]	1.486998	0.07663873	2.609986	2.2928667
[3]	2.088673	2.57999305	1.332857	1.4569732
[4]	1.437222	2.30510862	1.747217	0.5044043

B2

	[1]	[2]	[3]	[4]
[1]	0.9902272	2.4055380	1.7498137	1.8085272
[2]	2.3038720	0.5236963	2.4699941	1.9878323
[3]	2.2182690	2.9172222	0.7963895	1.3379670
[4]	1.9080839	2.2484797	1.0344988	0.4800907

C3

	[1]	[2]	[3]	[4]
[1]	1.571699	2.1778699	1.3778015	1.6466893
[2]	2.568147	0.3295985	2.5597320	2.4189639
[3]	1.814417	2.5068411	0.8482146	0.8532405
[4]	2.235167	1.9896964	1.1310151	0.5341030

D3

	[1]	[2]	[3]	[4]
[1]	0.7300203	1.8522527	1.8391259	1.2446140
[2]	2.5611112	0.4222982	2.6501559	2.4595412
[3]	1.8703064	2.5464689	0.7802779	1.1955269
[4]	1.4834723	2.4276149	1.8847895	0.6211937

E3

	[1]	[2]	[3]	[4]
[1]	0.5048864	2.4966052	1.811430	1.1391315
[2]	2.2385189	0.8852635	2.488748	2.1220932
[3]	2.8174107	2.7267402	1.252777	2.2189076
[4]	1.4226160	2.1414276	1.608472	0.3149319

F3

	[1]	[2]	[3]	[4]
[1]	0.2802825	2.256477	1.661674	0.49257183
[2]	2.2511855	2.150806	2.968204	2.50379359
[3]	2.3755755	2.358605	1.035726	1.96825859
[4]	0.6691850	2.244016	1.487522	0.09540547

G3

	[1]	[2]	[3]	[4]
[1]	0.3090461	1.3256008	2.448626	0.7709182
[2]	1.3353586	0.3096276	2.103869	1.6617359
[3]	3.0438620	2.8150436	1.574847	2.7227833
[4]	0.9308558	1.6720115	2.344661	0.3187669

A3

	[1]	[2]	[3]	[4]
[1]	0.7852855	1.7619368	2.572520	1.3551019
[2]	2.1066341	0.1397395	1.870108	1.0944421
[3]	2.8842719	2.3792073	1.272055	2.2692013
[4]	1.4820743	1.3089609	1.449799	0.7529744

B3

	[1]	[2]	[3]	[4]
[1]	0.4160973	1.5297751	1.620356	1.5345645
[2]	1.8856395	0.7770637	1.455261	2.0581742
[3]	2.5055833	2.5967267	1.604528	2.5049705
[4]	2.0179936	2.2509711	1.676838	0.4234145

C4

	[1]	[2]	[3]	[4]
[1]	0.556023	2.2091266	2.051034	1.9072477
[2]	2.473296	0.3399701	1.695988	1.6771684
[3]	2.890823	2.5175749	1.211063	1.8887992
[4]	1.865642	1.8440655	1.242325	0.2102612

D4

	[1]	[2]	[3]	[4]
[1]	0.6922711	1.8729633	1.676091	1.5186333
[2]	2.3564790	0.3587894	1.990424	2.2601285
[3]	2.6816718	2.5705538	1.114327	2.4870395
[4]	1.8888602	2.1780502	1.863793	0.2356145

E4

	[1]	[2]	[3]	[4]
[1]	0.6095648	1.7306197	2.200228	1.7730864
[2]	1.7321945	0.5274805	2.349239	2.2075068
[3]	1.4921621	2.1949036	1.180948	1.8380809
[4]	1.8753387	2.2729815	2.356397	0.4458459

F4

	[1]	[2]	[3]	[4]
[1]	0.4894334	1.8440754	1.382550	1.6372917
[2]	1.9825525	0.2323955	1.822889	2.4179023
[3]	2.3613298	2.4003000	1.051415	2.2322101
[4]	1.7175611	2.5300963	1.853027	0.3509333

G4

	[1]	[2]	[3]	[4]
[1]	0.5939677	2.0728091	1.373649	1.4421545
[2]	1.6421361	0.6605863	2.300663	2.5734000
[3]	2.3222484	2.7558535	1.393014	2.2259996
[4]	1.5520650	2.5154459	1.492083	0.3075501

A4

	[1]	[2]	[3]	[4]
[1]	1.158715	2.1440062	1.8876829	1.8698107
[2]	2.748690	0.1781174	1.7810607	1.7917095
[3]	3.047437	2.2729846	1.0300569	1.3610788
[4]	2.603671	1.7688671	0.9551264	0.3759293

B4

	[1]	[2]	[3]	[4]
[1]	0.4468671	2.2673283	2.2477621	2.2000094
[2]	2.0801012	0.4023021	1.1923947	2.4591343
[3]	2.3335797	0.6199725	0.9068139	2.3259909
[4]	2.1899674	2.0588684	1.7271116	0.6161328

C5

	[1]	[2]	[3]	[4]
[1]	0.6930304	2.4339333	1.765685	2.230117
[2]	1.9547079	0.5942702	1.752624	2.533066
[3]	1.9284650	1.4422644	0.804180	1.700889
[4]	2.0309115	1.9951552	2.188000	1.266630

C.1.3 Classification of Guitar with CQT - (95.0%)

E2

	[1]	[2]	[3]	[4]
[1]	0.2732115	0.88206817	0.4165612	0.38401536
[2]	0.7816325	0.04295658	0.8485119	0.72203415
[3]	0.5300666	0.82893040	0.1776054	0.35764381
[4]	0.3621503	0.73922287	0.2555453	0.09715011

F2

	[1]	[2]	[3]	[4]
[1]	0.2210501	0.77708324	0.6612229	0.40260343
[2]	0.7626073	0.07561036	1.0152343	0.90249406
[3]	0.6955713	0.97050829	0.1512161	0.58362208
[4]	0.4361732	0.88070507	0.6068594	0.07288036

G2

	[1]	[2]	[3]	[4]
[1]	0.4291102	0.8172109	0.4377411	0.3495084
[2]	0.5674613	0.1337142	0.9088496	0.7760367
[3]	0.5110728	0.9086767	0.1695562	0.4632343
[4]	0.4615326	0.8576797	0.4888014	0.1926622

A2

	[1]	[2]	[3]	[4]
[1]	0.3578937	0.80861990	0.6890344	0.63901788
[2]	0.7512427	0.05112309	1.1817874	1.09013997
[3]	0.6626674	1.18428680	0.1834759	0.21892366
[4]	0.5753494	1.08175425	0.2616937	0.09139808

B2

	[1]	[2]	[3]	[4]
[1]	0.4156265	0.98964613	0.4513634	0.5356578
[2]	0.8680591	0.09861392	1.0337097	0.9179335
[3]	0.5020616	1.05026128	0.2354104	0.3269371
[4]	0.4873293	0.86862699	0.3338521	0.1105733

C3

	[1]	[2]	[3]	[4]
[1]	0.3916996	0.9787936	0.3894526	0.44270911
[2]	0.9225889	0.0545268	1.0744128	0.99549505
[3]	0.3475554	0.9962310	0.1888426	0.23598171
[4]	0.4192474	0.9545195	0.2956976	0.07853036

D3

	[1]	[2]	[3]	[4]
[1]	0.1439543	0.9119927	0.6870198	0.29147333
[2]	0.9778739	0.0441631	1.1314357	0.93722380
[3]	0.6271578	1.1250641	0.1757966	0.57873955
[4]	0.3207890	0.9301796	0.6698290	0.07690307

E3

	[1]	[2]	[3]	[4]
[1]	0.2816260	0.7465308	0.4638111	0.6025778
[2]	0.6235635	0.2236632	0.8621233	0.6476728
[3]	0.6286682	0.8410926	0.2886355	0.6423770
[4]	0.4648509	0.6300282	0.6053210	0.1227376

F3

	[1]	[2]	[3]	[4]
[1]	0.1397838	0.5954831	0.8412344	0.32525113
[2]	0.7518691	0.4984058	1.0014857	0.90813092
[3]	0.8912456	0.8155090	0.3203573	0.73100349
[4]	0.3593506	0.6632348	0.8011779	0.08429882

G3

	[1]	[2]	[3]	[4]
[1]	0.1407514	0.36005411	0.8090160	0.3001175
[2]	0.4424359	0.09523127	0.6715639	0.3865349
[3]	0.7706864	0.68289111	0.2487134	0.6824943
[4]	0.3702985	0.34571572	0.6086963	0.1176678

A3

	[1]	[2]	[3]	[4]
[1]	0.1743851	0.47671907	0.4592802	0.3196321
[2]	0.5112284	0.05394871	0.5831291	0.4781856
[3]	0.4414654	0.53559991	0.2047418	0.2746063
[4]	0.3394564	0.48758925	0.2430116	0.1575967

B3

	[1]	[2]	[3]	[4]
[1]	0.2136010	0.44765007	0.4974197	0.3743201
[2]	0.4433370	0.07470339	0.5138350	0.5768920
[3]	0.6355100	0.62207982	0.4018124	0.6097650
[4]	0.5340875	0.55797941	0.3482908	0.1557078

C4

	[1]	[2]	[3]	[4]
[1]	0.1376412	0.56462435	0.6875730	0.69456949
[2]	0.5112294	0.07029958	0.4942026	0.47755942
[3]	0.6968317	0.49064853	0.2603278	0.33363597
[4]	0.6215526	0.44752459	0.3927331	0.09135529

D4

	[1]	[2]	[3]	[4]
[1]	0.4255410	0.6179510	0.3796455	0.34581500
[2]	0.6507359	0.3071149	0.6697870	0.74449614
[3]	0.6341174	0.6849768	0.2333232	0.46827251
[4]	0.6258615	0.7096756	0.3957079	0.05521967

E4

	[1]	[2]	[3]	[4]
[1]	0.3018097	0.4519238	0.6156795	0.7169508
[2]	0.5996662	0.3334322	0.6931655	0.5836173
[3]	0.5809481	0.5222926	0.3197796	0.6644195
[4]	0.7231389	0.6127656	0.7175005	0.1251830

F4

	[1]	[2]	[3]	[4]
[1]	0.09909854	0.6686584	0.4198178	0.4471655
[2]	0.67249167	0.2623220	0.6330506	0.7146946
[3]	0.49249634	0.5758146	0.2932199	0.4718245
[4]	0.48468147	0.7965283	0.5255932	0.1715654

G4

	[1]	[2]	[3]	[4]
[1]	0.2104847	0.5285133	0.6362334	0.6084811
[2]	0.5229177	0.1908519	0.6095748	0.5828509
[3]	0.6675685	0.5598990	0.3200945	0.3665988
[4]	0.6476520	0.5740591	0.2528197	0.2552362

A4

	[1]	[2]	[3]	[4]
[1]	0.3504103	0.5439655	0.5080288	0.9041523
[2]	0.5419018	0.1318877	0.3905614	0.7139527
[3]	0.6476667	0.4307336	0.3155127	0.5922652
[4]	0.9199729	0.7368461	0.5814655	0.1065080

B4

	[1]	[2]	[3]	[4]
[1]	0.2663596	0.9681744	1.0575523	1.0619435
[2]	0.8572537	0.2315251	0.5329120	0.9700302
[3]	0.9113892	0.3920278	0.4251666	0.7109919
[4]	1.1220006	0.8346485	0.6078778	0.3590512

C5

	[1]	[2]	[3]	[4]
[1]	0.3074404	0.8901655	0.7373966	0.8850676
[2]	0.6170481	0.1871822	0.2678120	0.3552661
[3]	0.6903935	0.3884146	0.3193042	0.3026091
[4]	0.8063681	0.4293515	0.4339502	0.2419678

C.1.4 Classification of Violin with FFT -(77.3%)

G3

	[1]	[2]	[3]	[4]	[5]
[1]	0.4244898	1.3633799	0.3483462	0.8953346	0.3881846
[2]	1.3318024	0.2041590	1.3580192	0.9003136	1.2882202
[3]	0.3148347	1.3745894	0.2011640	0.9577649	0.2797303
[4]	0.9334877	0.8539042	0.9905004	0.3437792	0.9148461
[5]	0.2164783	1.2784491	0.2564902	0.8505269	0.2094217

A3

	[1]	[2]	[3]	[4]	[5]
[1]	0.2427167	1.3782860	0.4771657	0.5634069	0.8208285
[2]	1.3821100	0.4890708	1.3541011	1.3212186	1.2848557
[3]	1.1310957	1.2439342	0.8819555	0.8662244	1.1190655
[4]	0.7524380	1.3684117	0.8551102	0.7592852	1.0463783
[5]	0.7030062	1.3051299	0.8185025	0.7677301	1.0044562

B3

	[1]	[2]	[3]	[4]	[5]
[1]	0.4493159	0.8404943	0.8579574	0.8069439	0.8253792
[2]	0.5820436	0.5392474	1.2654790	0.9978826	1.2665865
[3]	1.1030451	1.2787218	0.4164319	1.0620777	0.3768479
[4]	0.9558114	1.1743678	0.7677862	0.4566571	0.7525723
[5]	1.0672212	1.2863405	0.5117508	1.0375620	0.3956330

C4

	[1]	[2]	[3]	[4]	[5]
[1]	0.2936371	1.2086167	0.2843521	0.4451607	0.2879332
[2]	1.1159708	0.2045846	1.0895365	0.9654015	1.3268983
[3]	0.2597257	1.0440192	0.1890231	0.2829125	0.5520376
[4]	0.2541801	0.9785608	0.2068350	0.2332015	0.5703478
[5]	0.4883321	1.3268100	0.4882979	0.6362150	0.1981513

D4

	[1]	[2]	[3]	[4]	[5]
[1]	0.2741258	0.5847961	0.4882469	0.4962136	0.2621528
[2]	0.6299202	0.5046645	0.6536486	0.5861079	0.6230427
[3]	0.2423084	0.5088358	0.4150094	0.4591894	0.2667602
[4]	0.3953241	0.5552329	0.4914548	0.2993266	0.4116700
[5]	0.3024431	0.5896681	0.5029034	0.5230593	0.2589066

E4

	[1]	[2]	[3]	[4]	[5]
[1]	0.2400370	1.0130936	0.3938184	0.6844494	0.5174689
[2]	1.0124308	0.5373681	0.9269132	0.9373402	0.8290975
[3]	0.2497841	0.8497509	0.2117127	0.6060438	0.4037716
[4]	0.7528287	0.7219213	0.7952137	0.3760536	0.7525413
[5]	0.3804473	0.8257211	0.5257156	0.7403901	0.2972561

F4

	[1]	[2]	[3]	[4]	[5]
[1]	0.1963502	0.5394186	0.9820033	0.9612411	0.9365196
[2]	0.4789660	0.2964859	1.1904656	1.1722297	1.1783830
[3]	1.0840887	1.2174649	0.4094090	0.4570108	1.1013005
[4]	0.8963960	1.1012503	0.4886084	0.4227436	0.7210416
[5]	0.9313533	1.1709451	0.9298421	0.8967035	0.4356002

G4

	[1]	[2]	[3]	[4]	[5]
[1]	0.5642636	1.0481236	0.6308800	1.015089	0.5493134
[2]	0.9580596	0.7580146	1.3707995	1.150891	1.1992773
[3]	0.8681506	1.3762195	0.2063575	1.038794	0.5373498
[4]	0.9919384	1.1553105	1.0719212	0.589123	1.0849518
[5]	0.6644670	1.1087682	0.4777657	1.062306	0.3734342

A4

	[1]	[2]	[3]	[4]	[5]
[1]	0.6199036	0.8406557	0.7824645	0.8876169	1.0990010
[2]	0.5239703	0.3635380	0.5691503	0.6517999	0.8089888
[3]	0.2131895	0.6773568	0.2459856	0.4595659	1.1055343
[4]	0.2908316	0.6902743	0.3699766	0.1923457	1.0287187
[5]	1.0222210	0.8089618	1.0574120	0.9767709	0.3163135

B4

	[1]	[2]	[3]	[4]	[5]
[1]	0.4675367	0.6589661	0.5099577	0.8978782	0.5397549
[2]	0.4885337	0.5487174	0.4709700	0.8232318	0.4591952
[3]	0.3515287	0.7247634	0.1176820	1.0073826	0.1655481
[4]	0.9464685	0.8029487	1.0084675	0.1805072	0.9923369
[5]	0.4055689	0.7345062	0.2815513	1.0206061	0.2681566

C5

	[1]	[2]	[3]	[4]	[5]
[1]	0.1216183	0.2305659	0.3946158	0.4113510	0.7856277
[2]	0.5530706	0.6173527	0.6829443	0.6718291	0.9257088
[3]	0.4105535	0.5326046	0.3359625	0.2593944	0.6246043
[4]	0.4634132	0.6374290	0.5207043	0.2092895	0.7279921
[5]	0.7328691	0.7818247	0.5185713	0.7447176	0.3973358

D5

	[1]	[2]	[3]	[4]	[5]
[1]	0.3076478	0.8175342	1.0202109	0.8071964	0.6372484
[2]	1.0110534	0.4048773	1.0052784	0.5648855	0.6984619
[3]	0.8975495	0.9891132	0.3542339	0.9282849	0.7789779
[4]	0.8738502	0.5270483	0.8295629	0.2268470	0.3177045
[5]	0.6056022	0.6425327	0.7118294	0.5070649	0.2263492

E5

	[1]	[2]	[3]	[4]	[5]
[1]	0.5873613	1.0329114	0.9465996	0.9672957	0.8389223
[2]	0.8885601	0.6984367	0.5395630	0.6314176	1.2567849
[3]	0.8594184	0.8142536	0.1907329	0.3743153	1.1913914
[4]	0.9160338	0.8596652	0.4034308	0.3361453	1.2657798
[5]	0.7650719	1.2558573	1.0650645	1.1560492	0.5506876

F5

	[1]	[2]	[3]	[4]	[5]
[1]	0.2352812	0.5206546	0.3606697	0.4711524	1.0675668
[2]	0.4802422	0.5099229	0.6335562	0.6928904	0.8514085
[3]	0.3515943	0.6504946	0.2233743	0.3553288	1.0647774
[4]	0.6127940	0.8927457	0.3448226	0.2574624	1.1381271
[5]	1.2738275	1.0792715	1.2541755	1.2475513	1.0165912

G5

	[1]	[2]	[3]	[4]	[5]
[1]	0.8371442	1.158661	1.0274137	1.0110309	0.6880680
[2]	0.8101630	1.006626	0.9195206	1.2444316	0.8279214
[3]	0.9543927	1.213167	1.0478965	0.9102140	0.7978463
[4]	1.3260596	1.361356	1.4512926	0.3143299	1.2595851
[5]	0.8793232	1.244385	1.2431058	1.3049499	0.7873669

C.1.5 Classification of Violin with CQT - steady state only. (94%)

G3

	[1]	[2]	[3]	[4]	[5]
[1]	0.2834410	1.0338346	0.2586078	0.6897016	0.355369
[2]	0.9784523	0.1632418	1.0363673	0.6493284	0.9091341
[3]	0.3269203	1.0668750	0.2180708	0.7531321	0.3724153
[4]	0.6645471	0.5964353	0.7104964	0.1998018	0.5920792
[5]	0.3179743	0.9043346	0.3641842	0.6072879	0.1828352

A3

	[1]	[2]	[3]	[4]	[5]
[1]	0.1645610	1.0884661	0.4644390	0.5126680	0.6537164
[2]	1.1294036	0.4059235	1.0117056	0.9702968	0.9853881
[3]	0.7162229	0.8764215	0.4205473	0.5387421	0.6984695
[4]	0.5078928	0.9669903	0.4967766	0.4966581	0.6374916
[5]	0.5265694	0.8134343	0.4880583	0.5195911	0.5650537

B3

	[1]	[2]	[3]	[4]	[5]
[1]	0.2661252	0.7582530	0.7563485	0.5920479	0.6404801
[2]	0.6844779	0.4166734	1.0152018	0.7185519	0.9928544
[3]	0.8417674	0.9910966	0.4298792	0.8031561	0.2390397
[4]	0.6679381	0.8258100	0.7077930	0.2598101	0.6348675
[5]	0.7737678	0.9900808	0.5403922	0.7829239	0.1928940

C4

	[1]	[2]	[3]	[4]	[5]
[1]	0.2181339	1.0219942	0.2326498	0.3280178	0.2266973
[2]	0.9537115	0.1182949	0.9298804	0.8013237	1.1288916
[3]	0.2453241	0.8948793	0.2012954	0.2447543	0.4261718
[4]	0.2570287	0.7956654	0.2390831	0.1878408	0.4719604
[5]	0.3992677	1.1172817	0.3645142	0.4471758	0.1169588

D4

	[1]	[2]	[3]	[4]	[5]
[1]	0.1621505	0.4873282	0.3188550	0.3712144	0.3062155
[2]	0.4444449	0.3541527	0.4514171	0.4899355	0.4414231
[3]	0.2399916	0.4091500	0.2247437	0.3485222	0.3178865
[4]	0.3414725	0.5233523	0.4015520	0.2055511	0.3819299
[5]	0.2979448	0.4589472	0.3847446	0.4128306	0.1915263

E4

	[1]	[2]	[3]	[4]	[5]
[1]	0.2365753	0.7101581	0.2970319	0.5247294	0.3758171
[2]	0.6193341	0.3740182	0.6687207	0.5689562	0.5840021
[3]	0.3231471	0.6412837	0.1855710	0.4519831	0.3236940
[4]	0.6049888	0.5049923	0.5881612	0.1943262	0.5847720
[5]	0.3220131	0.5940973	0.4504970	0.5463601	0.2939347

F4

	[1]	[2]	[3]	[4]	[5]
[1]	0.1521683	0.5195343	0.6096694	0.6158404	0.6841566
[2]	0.4890409	0.2173483	0.7846581	0.7562538	0.7754323
[3]	0.6803389	0.8382490	0.2325730	0.3822397	0.8098940
[4]	0.5706984	0.6836985	0.3705777	0.2453514	0.6278166
[5]	0.6334406	0.7231605	0.6529112	0.5965012	0.4470963

G4

	[1]	[2]	[3]	[4]	[5]
[1]	0.2935994	0.7315626	0.5802254	0.6550588	0.4234762
[2]	0.6765667	0.4769321	1.0180259	0.7561087	0.8492650
[3]	0.5741145	1.0029151	0.2085915	0.6937096	0.4793751
[4]	0.6346706	0.7114095	0.7522231	0.3753263	0.6744801
[5]	0.5309674	0.7835681	0.5524998	0.6779885	0.2447162

A4

	[1]	[2]	[3]	[4]	[5]
[1]	0.4916486	0.7229967	0.6320845	0.7165342	0.7910750
[2]	0.5423291	0.2518057	0.6126246	0.5841486	0.5176021
[3]	0.2811047	0.6891088	0.1775162	0.4831481	0.8724570
[4]	0.3749316	0.5961340	0.4585114	0.1486139	0.6696188
[5]	0.7440084	0.5560859	0.8434947	0.6625169	0.1563159

B4

	[1]	[2]	[3]	[4]	[5]
[1]	0.3507546	0.4757550	0.6135950	0.6782782	0.6383972
[2]	0.4521907	0.2953904	0.6119206	0.5309660	0.5808921
[3]	0.5154841	0.6915963	0.1339405	0.7894258	0.3075768
[4]	0.7161031	0.5391746	0.7887525	0.1252276	0.7506598
[5]	0.5979247	0.6740690	0.3029576	0.8106107	0.1888304

C5

	[1]	[2]	[3]	[4]	[5]
[1]	0.1066922	0.3429112	0.4625464	0.5208523	0.6838964
[2]	0.4565825	0.4061422	0.5766791	0.5481655	0.7927264
[3]	0.4997276	0.6632566	0.2871054	0.3781787	0.4622854
[4]	0.5365487	0.6684509	0.4950741	0.1198800	0.6021539
[5]	0.6184106	0.7368466	0.3238383	0.6337803	0.2454301

D5

	[1]	[2]	[3]	[4]	[5]
[1]	0.3492241	0.5522946	0.7447129	0.5955394	0.5256636
[2]	0.6699285	0.5543438	0.7676196	0.4754589	0.5189880
[3]	0.6925926	0.7684706	0.2532784	0.6371748	0.5376839
[4]	0.7197097	0.6473408	0.5337245	0.1964602	0.2409786
[5]	0.6304245	0.6060065	0.4522517	0.3062915	0.1202116

E5

	[1]	[2]	[3]	[4]	[5]
[1]	0.2702135	0.7932695	0.7364954	0.8764528	0.7318095
[2]	0.8085388	0.3843244	0.5617414	0.4812838	0.9248525
[3]	0.7354551	0.6220393	0.2043946	0.4283484	0.8040302
[4]	0.9215669	0.5561141	0.4909916	0.1900484	0.9479112
[5]	0.7245602	0.9016567	0.6822648	0.8728619	0.3148789

F5

	[1]	[2]	[3]	[4]	[5]
[1]	0.2112032	0.5227486	0.3786740	0.4675998	0.7969170
[2]	0.4781492	0.3408056	0.4928397	0.5415399	0.5181931
[3]	0.3966925	0.5885273	0.2017385	0.3455525	0.7081890
[4]	0.5438087	0.7225432	0.3056275	0.1666702	0.7581330
[5]	0.9083524	0.6545617	0.8157778	0.8023986	0.5971825

G5

	[1]	[2]	[3]	[4]	[5]
[1]	0.4721527	0.7959820	0.5503625	0.7756037	0.7122392
[2]	0.5594920	0.6328890	0.6154926	0.8259146	0.5605284
[3]	0.5909936	0.7312898	0.5383239	0.6953228	0.7791225
[4]	0.8314695	0.6038175	0.8341111	0.2390122	0.9591482
[5]	0.6385918	0.8799616	0.8365639	1.0615542	0.5142308