2010-01-01

Playing Technique and Violin Timbre: Detecting Bad Playing

Jane Charles

Technological University Dublin

Playing Technique and Violin Timbre:

Detecting Bad Playing

Jane Ada Charles

A thesis presented to the Dublin Institute of Technology,

Faculty of Engineering,

For the degree of

Doctor of Philosophy

2010

Research Supervisors: Dr. Derry Fitzgerald

Prof. Eugene Coyle

I certify that this thesis which I now submit for examination for the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for another award in any Institute or University.

The work reported on in this thesis conforms to the principles and requirements of the Institute's guidelines for ethics in research.

The Institute has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature _____

Date _____

Abstract

For centuries, luthiers have committed to working towards better understanding and improving the sound characteristics and playability of violins. With advances in technology and signal processing, studies attempting to define a violin's sound quality via physical characteristics and resonance patterns have ensued. Existing work has primarily focused on physical aspects reflecting an instrument's sound quality.

In the music information retrieval domain, advances have been made in areas such as instrument identification tasks. Although much research has been completed on finding suitable features from which musical instruments can be represented, little work has focused on the violin's complete timbre space and the effect a player has on the sound produced. This thesis specifically focuses on representing violin timbre such that a computer can detect the sound associated with a beginner from that of a professional standard player and detect typical beginner playing faults based on analysis of the waveform signal only. Work has been limited to nine playing faults considered by professional musicians to be typical of beginner violinists.

In order to achieve these goals, it was necessary to create a suitable dataset consisting of an equal number of beginner and professional standard legato note samples. Feature extraction was then carried out by taking features from the time, spectral and cepstral domains. Selected features were then used to represent the samples in a classifier based on their efficacy at reflecting change within the violin's timbre space. The dataset underwent the scrutiny of professional standard stringed instrument players via listening tests from which the target audience's perception was captured. This information was verified and normalised before use as *a priori* labels in the classifier. Based on different feature representations, classification of violin notes reflecting perceived sound quality is presented in this thesis. The results show that it is possible to get a computer to determine between beginner and professional standard player legato notes and to detect playing faults. This thesis involves a thorough understanding of violin playing, its perception, suitable analysis methods, feature extraction, representation and classification.

Acknowledgements

The completion of this thesis would not have been possible without the help and advice of many people. I wish to thank my supervisors Dr. Derry Fitzgerald and Prof. Eugene Coyle for their help. Undertaking this work would not have been possible without the support, humour and rapturous enthusiasm exuded by all the musicians who offered up their ears, training and skills for experimentation. Special thanks go to Lioba Petrie, Ailleen Kelleher, Grainne Hope, Sinead Hope and Karla Charles for helping me obtain a data set from which the research could take place. Also essential to this work was all the time, expertise, advice given to me by Owen Tighe, sound engineer. Without Owen's assistance, a master sound sample disk would not exist. Thank you. At one point during this research project I wanted to set up and observe Helmholtz motion on a violin. This was possible thanks to Finbar O'Meara and Ted Burke who helped source and set up the necessary equipment so that this could be carried out. Last but not least my family for their continued support.

Abbreviations

AC	autocorrelation
CK	spectral centroid kurtosis
CM	spectral centroid mean
CQT	constant Q transform
CQTH	constant Q transform harmonic bin content
CV	spectral centroid variance
dB	decibel
DCT	discrete cosine transform
DFT	discrete Fourier transform
DSP	digital signal processing
FFCV	four fold cross validation
FFT	fast Fourier transform
HMM	hidden Markov model
Hz	Hertz
KLT	Karhunen – Loève transform
k-NN	k-nearest neighbour
LOOCV	leave one out cross validation
MFCC0	Mel frequency cepstrum first coefficient
MFCC0M	Mel frequency cepstrum first coefficients mean
MFCC0S	Mel frequency cepstrum first coefficients skew
MFCC1S	Mel frequency cepstrum second coefficients skew
MFCC5	Mel frequency cepstrum sixth coefficient
MFCCs	Mel frequency cepstrum coefficients
MMO	Music Minus One

MMV	moving mean variance
MPO	Music Plus One
NMF	non-negative matrix factorisation
PSD	power spectral density
PSD190	power spectral density below 190Hz
RCC0	real cepstrum first coefficient
RCC1	real cepstrum second coefficient
RCC3	real cepstrum fourth coefficient
RCC5	real cepstrum sixth coefficient
RCCK	real cepstrum coefficients kurtosis
RCCM	real cepstrum coefficients mean
RCC	real cepstrum coefficients
RCCS	real cepstrum coefficients skew
RCCV	real cepstrum coefficients variance
RWC	Real World Computing music database
SCM	spectral contrast measure
SCM190	spectral contrast measure below 190Hz
SF	spectral flux
SFM	spectral flatness measure
SFMK	spectral flatness measure kurtosis
SFMM	spectral flatness measure mean
SFMS	spectral flatness measure skew
SFMV	spectral flatness measure variance
SOM	self-organising map
STFT	short-time Fourier transform
SVD	singular vector decomposition
TK	time domain kurtosis

TM	time domain mean
TS	time domain skew
TV	time domain variance

Playing Fault Abbreviations

BADE	playing fault poor finish to a note
BADS	playing fault poor start to a note
BB	playing fault bow bouncing
CR	playing fault crunching
INT	playing fault poor intonation
NV	playing fault nervousness
SE	playing fault sudden end to note
SK	playing fault skating
XN	playing fault extra note

TABLE OF CONTENTS

List of Figures

List of Tables

1 Introduction

Music, as a form of expression and of aural tradition, is a part of all cultures. From the sixth century, the monophonic liturgical chant of the Roman Catholic Church, known as Gregorian chant or plainsong, was passed on orally in Europe [Machlis90]. As the number of chants increased, a means of recording the different melodies was needed. By the 8th century, ascending and descending symbols, known as neumes, were written above the text suggesting the musical direction, but pitch and time could not be represented [*ibid.*]. Guido's four-line staff musical notation system was in use by the tenth century for monophonic chant. Polyphonic music only began to emerge towards the end of the Romanesque period (c. 1050-1150) [*ibid.*]. Following this, a gradual rise in the importance of instrumental and secular music began to evolve during the fourteenth century. The current five-line staff notation system became widespread by the 16th century [*ibid.*]. It has taken about eight centuries to develop a standardised Western music notation system and method of recording the primarily oral tradition.

As music has evolved over the centuries reflecting societal developments and changes, so too have the instruments, playing techniques, styles, how music is perceived, methods of recording the material and teaching methods. Focusing on the teaching of music, feedback and interaction with a tutor is central to a student's progress especially during the initial years. An important part of this process involves developing muscle memory or a link between hearing a sound and what it feels like to produce it under expert guidance. Refining this aural training is a lifetime's work and key to musical expression on many instruments, in particular bowed stringed instruments.

When learning to play a bowed stringed instrument, such as the violin, contact time between teacher and student is very important, but often limited. This has resulted in the development of a number of practice tools such as accompaniment only recordings, as available through the Suzuki Method publishers [Suzuki09], Music Minus One [MMO09] recordings and more interactive systems such as Music Plus One [MPO09]. For the struggling student, a home computer based tutoring system capable of analysing his or her violin playing and offering feedback could be of benefit. Prior to being able to develop such an interactive system, a thorough understanding of violin timbre, how it is produced and can be represented quantitatively, while still reflecting the qualitative

expressions used by musicians, is needed. For example, what features can be used to characterise a poorly played note versus that of a well executed one and can playing faults be reflected by a measurement.

Finding a set of features from which violin sound can be represented which correlates with violinists' perception provides a challenge. There is much to be gleaned given the lack of perceptual correlates of violin sound quality as well as quantitative analysis of the effect a player has on the sound he or she produces. Understanding, quantifying, representing and classifying the effect playing technique has on violin timbre involves finding workable guidelines for what is considered by professional standard violinists to be a good sound, to describe and determine typical playing faults and finding methods from which this information can be quantified.

The research aim of this thesis is to obtain sufficient quantitative understanding of the qualitative relationship between violin sound and playing technique from which it is possible to determine a beginner note from a professional standard one and to detect common beginner playing faults. Knowledge of signal analysis methods and violin playing technique are important in this work. The successful development of techniques capable of such discrimination is reliant on finding and establishing appropriate features as well as a suitable classification process. This thesis builds on existing work from many fields, from acoustics to signal processing, leading to a novel approach to further understanding the violin timbre space.

To complete this work, a suitable dataset is needed as existing datasets have no beginner note samples. A dataset consisting of equal numbers of professional standard legato and of beginner player notes with playing faults, which have been obtained under the same conditions, is required. From these recordings, detailed waveform analysis in the temporal, spectral and cepstral domains has been conducted in order to better understand the sonic effects of bad violin technique and ways of obtaining measures to represent this information. Listening tests are conducted to assign qualitative labels to the samples, to remove subjectivity and to see if any perceptual correlates between violin timbre measures and qualitative expressions used by musicians can be established.

In the rest of this chapter, Section 1.1 very briefly presents how sound is produced on a violin. Section 1.2 introduces violin playing technique, emphasising some of its difficulties. Many texts exist on violin playing such as [Auer80, Flesch00, Szigeti79], so violin playing technique will be kept to a minimum throughout this thesis and only

included as necessary. Current research relevant to this work is summarised in Section 1.3 after which, an overview of musical signal representations is presented in Section 1.4. The thesis is outlined in Section 1.5, and the original contributions in this thesis are presented in the last section of this chapter.

1.1 A Brief Introduction to the Violin and Violin Sound

The violin as it is known today was perfected in Cremona, Italy in the late 17th century, by the school of luthiers founded by Antonio Stradivarius (1644-1720) [Gill84] and is currently used in a wide range of musical endeavours ranging from symphonic, solo, chamber music, jazz, folk, popular to religious. For a detailed presentation of violin playing from the renowned violinist and pedagogue Leopold Auer, the reader is referred to [Auer80]. Throughout this text, parts of the violin are referred to and labelled images of the violin are given in Figure 1.1 and Figure 1.2, which have been taken from [Violin09].

Drawing a bow across the string correctly causes the violin body to resonate due to a complex system of different couplings. The excitation causes waves related to its length to propagate along the string. These vibrations pass through the bridge to the sound post and bass bar allowing the instrument's body to resonate. The treble frequency vibrations pass through the right foot of the bridge and sound post shown in Figure 1.2 and the lower or bass frequency vibrations go through the left foot of the bridge and pass along the bass bar. The brightest sound from a violin is produced when the bow is drawn across the string, parallel to the bridge, in line with the tops of the f-holes. Not pulling the bow in such a manner mostly results in poor sound, associated with weak or developing technique, such as that belonging to a beginner.

The sound post and bass bar transmit vibrations, allowing the whole instrument body to resonate. The pressure changes resulting from the resonating body cause the sound to come out of the f-holes. Many functions are associated with the f-holes, including providing the opening required for the main air resonance, which relies on the volume of air in the violin's body [Bissinger92]. The f-hole shape has evolved to let the bridge oscillate more freely in transferring the string vibrations through the instrument body and thereby creating a louder sound [McLennan03]. Should the movement of the bridge be restricted, for example, by fitting a mute which restricts the vibrations going through from the bridge, the sound quality and volume are effected.

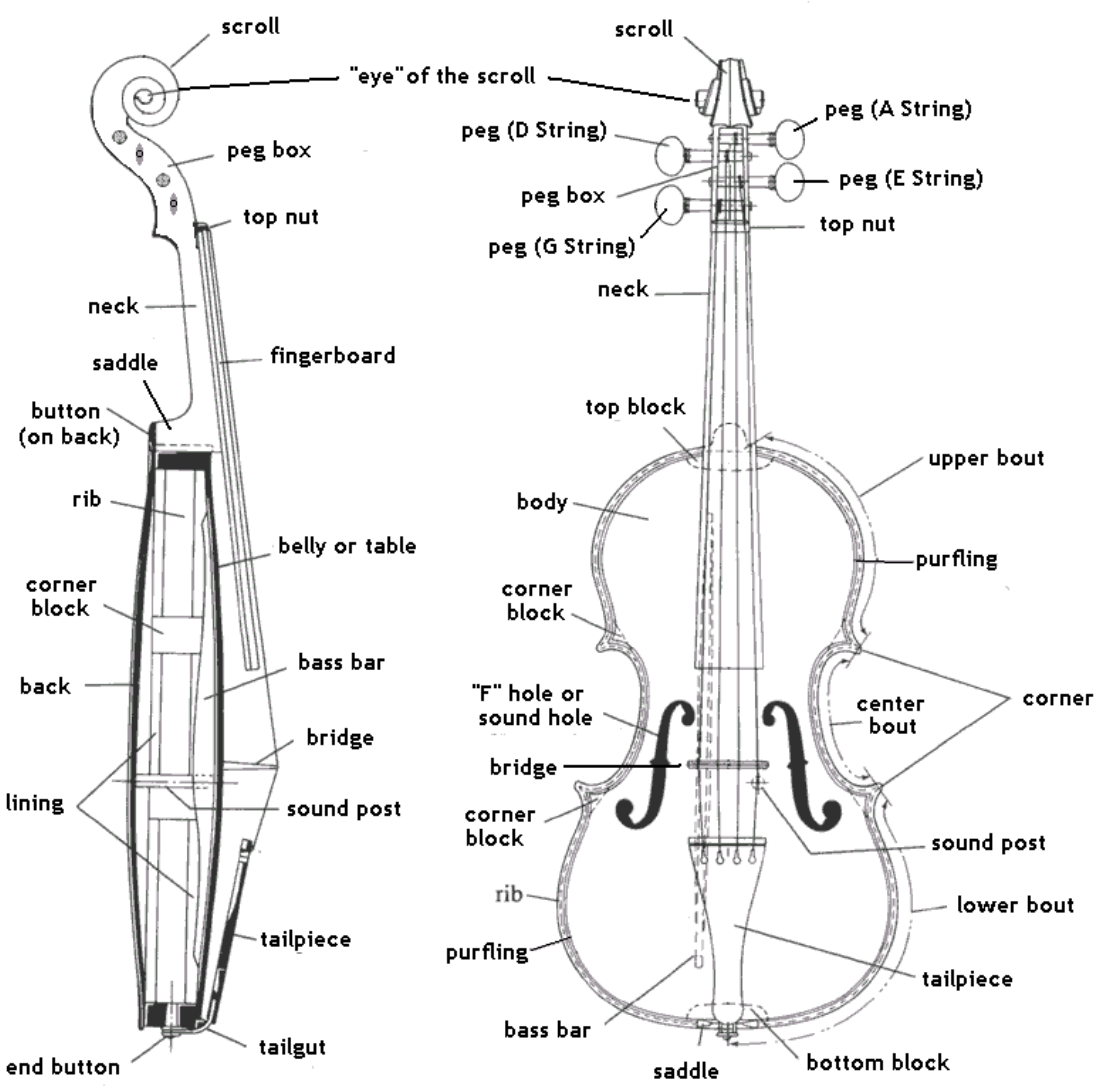

Figure 1.1: Violin parts.

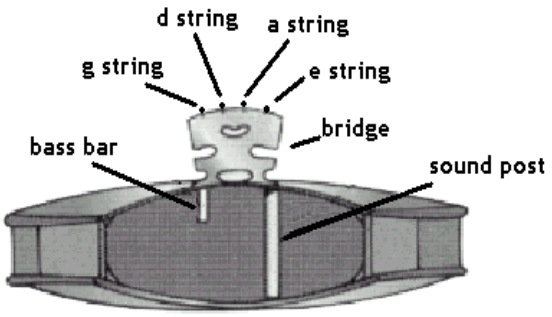

Figure 1.2: Internal parts of the violin.

The violin's shape has evolved to maximise sound output requiring the air volume to have the fundamental cavity or the f-hole resonance at about 260 to 290 Hz

[Hutchins97]. Bissinger researched the effects of area, shape and position of the f-holes on cavity mode frequencies and showed that f-hole shape has an important effect on the fundamental resonance, important in the radiation of acoustic energy or sound [Bissinger92]. The cavity resonance goes down in pitch should one f-hole be blocked [Hutchins90]. For a good sound to be produced on a violin, multiple complex resonances are excited [Bissinger98]. One text which covers violin physics in much greater detail is [Cremer84]. The next section details how violin sound is affected by playing technique.

1.2 Violin Playing Technique

The legato bow stroke is the basic bow stroke for all violinists and means “smoothly connected” [Jackson87:23]. To cite Auer, “legato is really the negation of angles in violin playing. It is the realizing of an ideal – the ideal of a smooth, round, continuous flow of tone.” It is the bow stroke which gives “the beautiful singing tone which is the normal tone of the instrument” [Auer80:32]. Once a beginner player has gained sufficient bow control to master the legato bow stroke, the player is ready to progress onto more challenging bowing patterns and more advanced bow strokes such as staccato, a “detached, disconnected” bow stroke [Jackson87:44] or martelé, “a sharply accented bowing” [*ibid.*:28], which involve much greater bow control.

A violinist alters the instrument’s timbre by changing his or her playing technique within the framework dictated by the instrument and bow. Drawing a straight bow seems simple enough to a non-player but a straight bow involves good posture, no muscle tension, a good bow hold, a loose wrist and keeping the violin still among other things. If a bow is not drawn as described previously, the sound quality suffers. The sound produced on a violin is a direct result of the player’s bow control, which influences how the instrument cavity resonates. The previous section detailed briefly how sound is produced assuming correct bowing technique. This section considers the effects of a player, a beginner in particular, on the sound produced, focusing on typical bowing faults.

Elements influencing bow control include the bow arm position, bow pressure and speed, bow angle, the location of the bow on the string and the straightness of the bow stroke. Some typical qualitative beginner player bowing faults include crunching, skating, nervousness, bow bouncing, extra note, sudden end, poor starts and ends to

notes. Too much pressure at the wrong place relative to bow speed causes the sound to crunch. Throughout this thesis the expression “crunch” or “crunches” refers to inappropriate force being applied to the string via the bow causing the sound to contain many more unwanted and unrelated frequencies. Not drawing a straight bow at the optimum place on the string results in a “whisping” or “skating” sound effect as the bow skids at an angle along the string. Player nervousness results in a non committed sound being produced and is caused by tension in the bow arm. Bow bouncing is also due to too much tension in the bow arm. Due to a lack of bow control, extra unplanned notes can be played. Unclean, gritty or crunchy beginnings and endings to notes usually occur until the player has mastered the bow hold, finger movement and smooth bow turns. A flexible, loose wrist and fingers are required to maintain sound quality. Bowing technique determines the attack strength, sound projection, harmonic content, timbre, pitch, instrument resonance and the length of the note. An overview of the relationship between bowing technique and sound is detailed in Table 1.1.

Issue	***Level***	***Result/effect on sound***
Bow hold	Too tight	Crunches, bumps, nasal, wobbling bow
Bow hold	Too loose	Light, unconvincing sound
Straight bow	Not straight	Skating sound, wobbling bow
Bow pressure	Too much	Crunches, wobbling bow
Bow pressure	Not enough	Sound lacks commitment, nervous.
Bow angle/hair contact	Too little	Playing on wood
Location of bow on string	Too close to bridge	Squeaks
Location of bow on string	Over fingerboard	Committed sound: airy, distant, dreamy. Not committed sound: nervous

Table 1.1: Relationships between playing technique and sound.

The variables reflecting bowing technique influence violin pitch as shown in Figure 1.3. The arrows pointing towards pitch in Figure 1.3 indicate player controlled variables and the outward pointing arrows show variables that are influenced by pitch. Although it is not marked on the diagram, none is independent.

Violin resonances are maximised by good playing technique. In this work, the evolution of a note is considered from a playing technique perspective only, thus avoiding the highly debatable concept of style. To Auer, “style in music … is the mode or method of presenting the art in question in a distinctive and intrinsically appropriate way” [Auer80:75]. Comparing and contrasting styles and interpretation constitutes a very large body of research and although it must be acknowledged, will not be covered in this thesis.

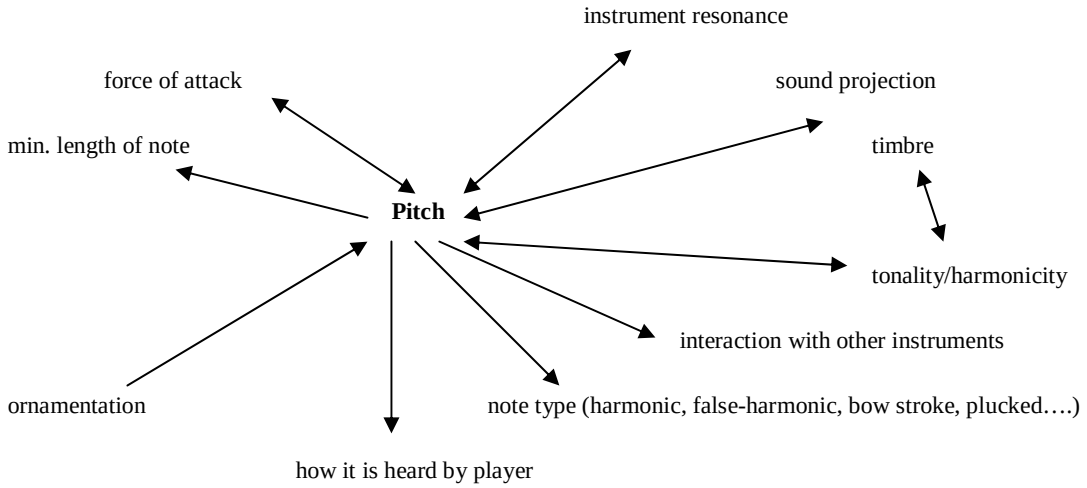

Figure 1.3: Elements influencing violin pitch.

Features through which sound quality and playing faults may be detected are sought. More specifically, characteristics or patterns of legato bow strokes, which are ideally independent of pitch and sample length, need to be considered in order to quantify legato sound. This involves finding a suitable representation of legato sound which reflects a violinist's perception. Through the quantitative and qualitative analysis of violin sound, a link between features and how musicians describe sound or playing characteristics is sought. Before detailing real violin sound representations, current research in the area is presented in the next section.

1.3 Current Research

Research influencing and inspiring this work comes from many areas including violin acoustics, music teaching methods and aids, music information retrieval, automatic accompaniment systems, speech recognition and player-instrument relationships. An overview is given in Figure 1.4. Certain teaching methods, such as the Suzuki Method [Suzuki73], place significant emphasis on listening, more so than more traditional methods. This "mother tongue" approach to teaching relies on the development of listening skills or "ear training" from the outset. The student does not learn to read music until they are proficient at playing the instrument and pieces of music. The basics of playing the instrument are mastered first and when the student is introduced to reading music, he or she learns by association. The Suzuki Method publishers have been making recordings of the repertoire and accompaniment only recordings available since the 1980s.

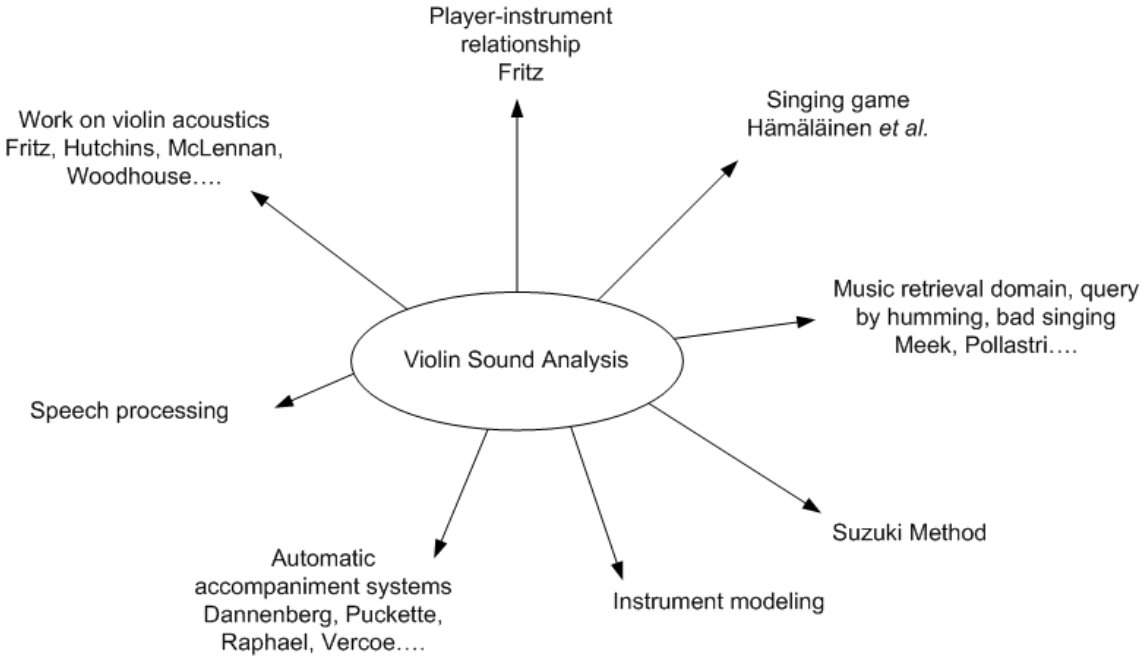

Figure 1.4: Relevant research domains.

A significant amount of work has been carried out on violin acoustics including finding perceptual correlates [Fritz06, Fritz07]. Much of this work though is focused on trying to emulate the old, Italian master violin makers, such as Stradivarius [Smithsonian09, Hutchins97], not on playing technique and its effects on sound. Although much work has been conducted on violin acoustics and the violin is the most uniform of the stringed instruments [Jansson97], much remains unknown. The complexities of how the violin resonates make it extremely difficult to develop a complete physical model. Work towards developing a physical model of a bowed violin string has been done [Serafin01] and Wilson has tried extracting violin performance information necessary to drive a digital waveguide model of a bowed violin [Wilson02]. With the violin, minute changes such as moving the sound post less than a millimetre greatly influence the instrument's sound [Molin90, McLennan01]. Such variables, of which there are many, need to be captured by a physical model. However, information relating to physical models for various wind instruments such as the trumpet, trombone, saxophone [Vergez06], oboe [Almeida04] and piano [Giordano04] has been published.

Many approaches used in the music domain originated from general signal and speech processing techniques. Of particular interest is research into the singing voice and developments which aim to alter its sound characteristics. This includes the potential for improving sound quality as the violin is the instrument that best mimics the human singing voice [Winkel67]. Pollastri draws attention to the need to develop

specific algorithms to deal with a singing voice [Pollastri02b]. Much work has gone into studying the singing voice from analysis and synthesis perspectives and Sundberg [Sundberg87] offers a thorough look at the peculiarities of the singing voice. Work on the analysis, synthesis and improvement of the singing voice is on going [Bonada01a, Bonada01b] and some work has been conducted on poor singing. Papers which consider poor singing quality, within the music information retrieval domain include [Meek02, Pollastri02a, Pollastri02b] and involve classification methods through query by humming or singing. Several plug-ins have been developed for improving or adding special effects to a recorded singing voice. One such example emulates the Louis Armstrong growl in an approach that allows a modal voice to be transformed into a growl voice by adding sub-harmonics in the frequency domain [Loscos04].

With current advances in signal processing and interactive computing, much more sophisticated systems and learning aids are now being developed. Such systems are of interest because the analysis of poor musical sound attributes is considered. Hämäläinen *et al.* developed a successful real-time singing aid in [Hämäläinen04], which describes the use of pitch-based control of a game character by the user's voice. A direct transfer of this approach to a violin, or another instrument aid would not be as successful. A singer is physically "free" to concentrate on a screen and able to react to it. Instrumentalists, especially beginners, need to be looking at what they are doing and looking elsewhere, i.e. at a screen, will disturb their position. For this reason, a system which offers feedback after the user has played his or her notes would be much more effective. This differs greatly in approach to the MMO CDs, which offer a variety of recordings to which the user plays the solo part or MPO, which is an interactive accompanying system [MPO09]. Automatic accompaniment systems have evolved greatly since MMO and many developments have occurred since the systems put forward by Vercoe [Vercoe85] and Dannenberg [Dannenberg85]. Raphael's MPO system reacts in real time to changes in the soloist's tempo allowing for musical expression [Raphael03] and delay has not been reported to be an issue by MPO's users [*ibid.*].

Meek and Birmingham put forward a Hidden Markov (HMM) based model for dealing with how similar a query is to a potential target within the area of music retrieval [Meek02]. This model was developed linking several elements together in the same model. These elements, as presented in [*ibid.*], are transposition, modulation, tempo, tempo change, non-cumulative local error, cumulative local error, insertions and

deletions. Essentially, it is a model which deals with pitch and tempo changes in depth made with the assumption of conditional independence between the representational elements. In [Shifrin03], the performance of a query-by-humming HMM is successfully tested on a large musical database.

Spectral features have been used for musical instrument timbre classification [Agostini01, Agostini03] as have cepstral and temporal features [Eronen00, Eronen01]. In these works, instruments including the violin are represented by multiple features. Features used for identifying individual instruments focus on good instrument sound and not on representing change within an instrument's timbre space. Little research has been carried out into the relationship between player and effect on instrument sound. Fritz investigated the relationship between clarinettist and sound produced depending on glottal and windpipe shape [Fritz04]. The classification of three common violin bow strokes has been done using data collected from an electric violin and a carbon bow to which sensors have been attached [Young08]. These works consider measurements obtained via sensors for good playing sound or technique only. There seems to be no work conducted on analysing poor violin playing technique. This thesis focuses on the effects of acoustic violin playing technique on sound and ways of detecting playing faults. Before presenting the thesis outline, suitable musical signal representations are presented in the next section.

1.4 Musical Signal Representations

This section illustrates different ways of representing musical signals in the time, frequency, time-frequency and cepstral domains. An acoustic signal is most commonly represented in the digital domain by its sampled waveform where each sample describes the signal's amplitude with respect to time. Time-frequency analysis allows the time at which signal frequencies are present to be identified and representations used throughout this thesis include the spectrogram and the Constant Q Transform (CQT). The cepstral domain is also useful for representing instrumental sound [Klapuri01, Eronen01] as the presence of periodicities in the signal are captured [Oppenheim89].

Starting with the time domain input signal $x(n)$, its frequency response $X(n)$ is obtained via the Discrete Fourier Transform (DFT), where N is the signal length in Equation 1.1 [*ibid.*]:

$$X(n) = \sum_{n=0}^{N-1} x(n) e^{-j(\frac{2\pi}{N})} \tag{1.1}$$

In Figure 1.5, the top image is the waveform of a professional standard legato note and a close up of the most significant part of its spectrum is shown in the bottom image, displaying the note's fundamental and harmonics. The clean execution of this note is reflected in its spectrum as few unrelated frequencies are present.

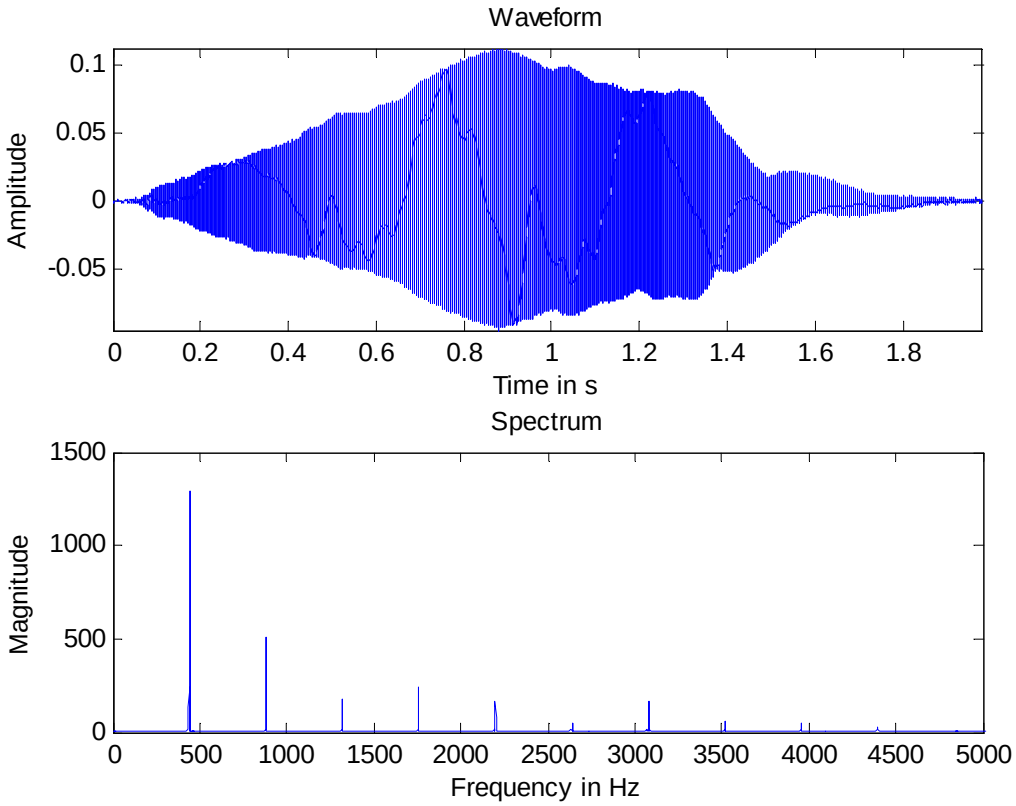

Figure 1.5: Waveform (top) and harmonic spectrum section (bottom) of a professional standard player legato A440 note.

Harmonic content throughout a note is better represented in time-frequency representations as changes in its frequency content with respect to time are illustrated. One widely used time-frequency representation is the short-time Fourier transform (STFT) based spectrogram whereby the data is presented via a succession of windowed DFTs. A Hamming window $w(m)$ is used in this work and the STFT is given by $X(n,k)$ in Equation 1.2, where k is the frequency bin, n the frame and N the signal length [*ibid.*]:

$$X(n,k) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(hn+m)w(m)e^{-\frac{2j\pi mk}{N}} \tag{1.2}$$

The signal's content is represented in terms of frequency versus time and can be used to give a temporal evolution of a note as can be seen in Figure 1.6. In this figure, the darkest lines are the note's harmonics. A 1024 point window with 50% overlap has been applied.

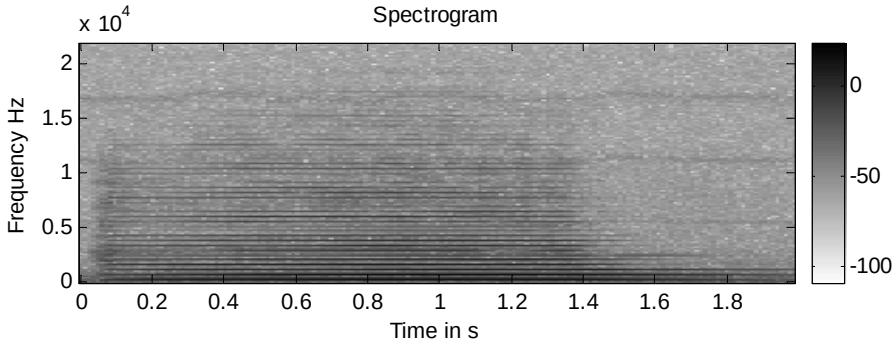

Figure 1.6: STFT based spectrogram of an A440 legato note.

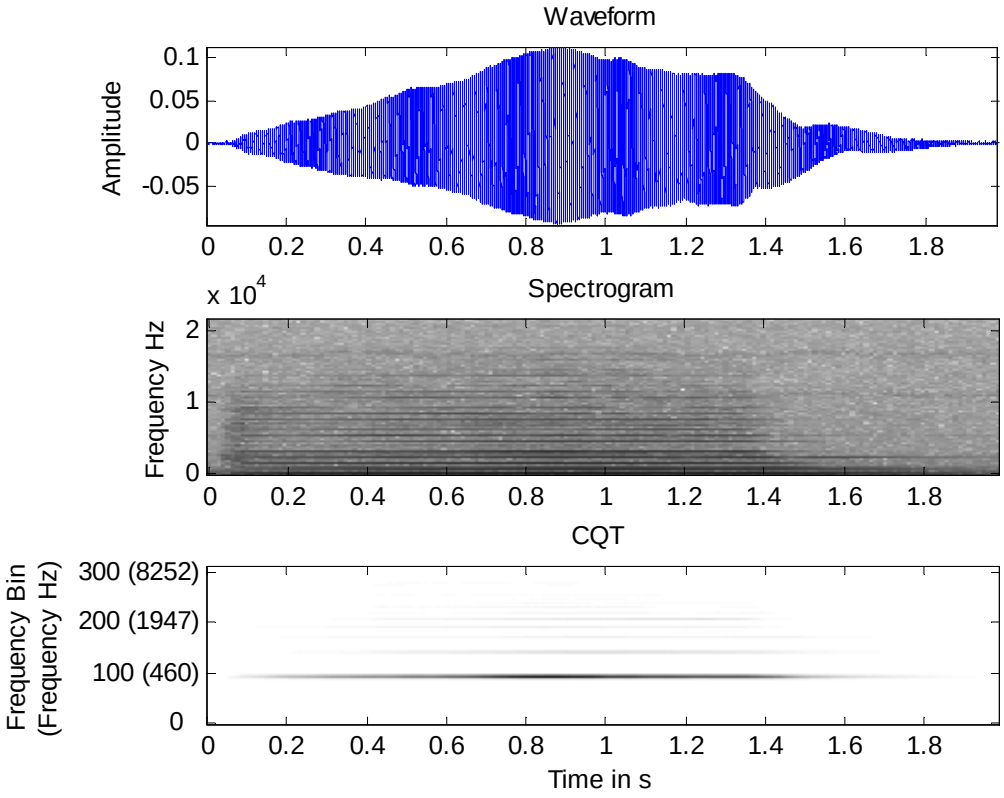

Figure 1.7: Signal representations: waveform (top), spectrogram (middle), CQT (bottom).

The CQT representation is effective for visualizing and exploiting information about the harmonic content of a note. In Figure 1.7, the waveform, spectrogram and CQT representations of a professional standard legato A440 violin note are illustrated. The Constant Q Transform (CQT) is a log-frequency scaled time-frequency representation of a signal [Brown91]. It differs from the DFT in that the ratio between centre frequency and frequency resolution remains constant making it suitable for the representation of musical signals by setting the frequency resolution to match that of equal temperament. In equal temperament, such as twelve-tone equal temperament, each octave is divided into 12 equal parts whereas linear spacing occurs in the DFT. To obtain the CQT of a signal, lower and upper frequency limits must be selected. A lower or "start" frequency limit of 110Hz which is sufficiently below the lowest note G (approximately 196Hz) on a violin tuned to A440 and an upper frequency limit of 10kHz are assigned. Eighth tone spacing is selected in this work over the more often used quartertone spacing [*ibid.*] to access more information in the beginner note samples. There are 48 eighth tone spaces in an octave and the centre frequencies are calculated from Equation 1.3 where b = 48, the number frequency bins per octave, *f*, the initial or previous frequency and c = 1, 2, 3…

$$f_c = f\ 2^{\frac{1}{b}} \tag{1.3}$$

After the centre frequencies f_c have been returned, the ratio between the centre frequency and bandwidth, represented by Q, is obtained through Equation 1.4:

$$Q = \frac{1}{2^{\frac{1}{b}} - 1} \tag{1.4}$$

The sampling frequency f_s is equal to 44.1kHz. Each frequency bin is estimated with a frequency dependent window length N_c, limited by a maximum window size and a frame size given by Equation 1.5:

$$N_c = Q\frac{f_s}{f_c} \tag{1.5}$$

The CQT is obtained from Equation 1.6, where $w[n,N_c]$ is a windowing function [*ibid.*] and Hamming window has been used as in [*ibid.*]:

$$X[c] = \frac{1}{N_c} \sum_{n=0}^{N_c-1} w[n, N_c]x[n]e^{-2\pi jQ\frac{n}{N_c}} \tag{1.6}$$

Cepstral analysis is used successfully in processing speech, seismic, biomedical, sonar signals, old acoustic recordings [Oppenheim89], music modeling [Logan00] and in instrument identification tasks [Brown01, Martin98]. The complex cepstrum of a signal is the inverse Fourier transform of the log spectrum shown below in Equation 1.7 [Oppenheim89]:

$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} [\ln|X(e^{j\omega})| + j * phaseX(e^{j\omega})]e^{j\omega n} d\omega \tag{1.7}$$

Although the log of any base may be used [Deller00], throughout this work the natural log has been applied.

The real cepstrum or cepstrum differs from the complex cepstrum in that it leaves out the signal's phase information and is given by Equation 1.8:

$$c_r[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|X(e^{j\omega})|e^{j\omega n} d\omega \tag{1.8}$$

The complex cepstrum need only be used in phase-sensitive applications such as vocoders, whereas the cepstrum is more often used in speech analysis and recognition systems. Due to its ability at detecting periodicities in the spectrum, the real cepstrum has numerous applications, including pitch detection, speech modeling, in digital filter design and machine diagnostics [Oppenheim89].

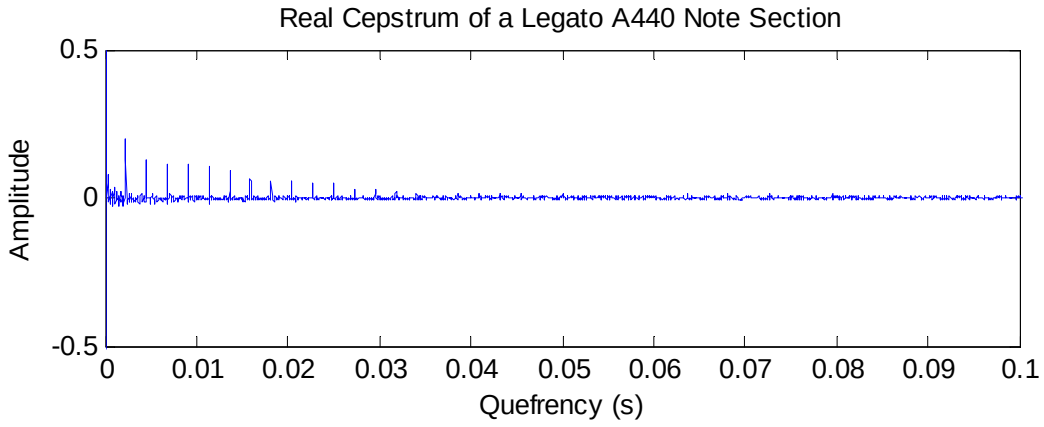

Figure 1.8: Real cepstrum representation section of a legato A440 violin note.

A section of the real cepstrum of a professional standard legato note is displayed in Figure 1.8 where the periodic nature of the sound is visible.

The representations illustrated in this section will be used for comparing, analyzing and extracting features from which violin samples may be represented. An outline of the thesis is given in the next section.

1.5 Thesis Outline

So far in this chapter, violin sound, playing technique, relevant research, musical signal representation and the aims of this thesis have been presented. Violin sound, its perception, production and analysis are presented in Chapter 2. Why, what and how the dataset was obtained as well as the listening tests carried out are explained in Chapter 3. In Chapter 4, the effects of violin playing technique on a note's waveform are detailed and common playing faults are presented. Temporal, spectral and cepstral analyses of the dataset are documented in Chapter 5, Chapter 6 and Chapter 7 respectively. In Chapter 8, features are selected according to their performance in their respective domains and the dataset is represented by a feature array which is then used in a *k*-nearest neighbour classifier using the *a priori* labels from the listening tests. Two tasks are tested for classification: one for determining beginner from professional standard playing and the second, fault detection. Selected successful detection feature combinations are then tested on new data. Conclusions drawn from the work completed and alternatives, strengths, weaknesses and possible further work are detailed in the final chapter. The following section briefly details the original contributions presented in this work.

1.6 Original Contributions

The gap in existing work relating to complex instrument signal analysis, such as violin sound, allows for further work in this area to be undertaken. Existing work including that conducted on violin acoustics, instrument modelling, musical instrument identification and classification, music information retrieval, automatic accompaniment systems bases its analyses on or towards good violin sound and playing. This thesis puts forward a novel approach to violin sound analysis by considering the effect a player has on the sound produced. It will be shown that correct detection of a beginner note from a professional standard legato one is possible with over 96% accuracy and that multiple

playing faults are detectable. This work presents how this is achieved through monitoring performance based on sound limits which are considered by professional players to be good and reflected through quantitative measures. These quantitative measures include standard features from the temporal, spectral and cepstral domains but also modifying some of these to focus on the frequency range below the lowest note on the violin, approximately 196Hz, which has not previously been done. Some excellent yet unexpected results from these features will be detailed, highlighting areas meriting further study. In particular, analysis within the time domain will be shown to be very effective, including features such as the waveform amplitude mean and the moving mean variance values which separate completely the beginner from the professional standard legato note samples in the dataset. In the spectral domain, specific CQT frequency bins below 196Hz which completely separate the dataset based on player standard will be displayed. The content present at some of these frequencies can be considered to reflect certain violin resonances as excited by different players. Taking the mean PSD and SCM present in the frequencies below 190Hz also perform very well at distinguishing between the dataset's different player types. The results, obtained from these features and displayed as modified and detailed in this work, will be shown to be statistically significant and have not previously been used to analyse violin sound. This thesis will show, for the first time, that a beginner player notes can be distinguished from professional standard legato ones and that multiple playing faults can be detected.

2 Perception and Analysis of Violin Timbre

As used in music, timbre refers to the characteristic sound/s created by a musical instrument. It is “a term describing the tonal quality of a sound” [Sadie01:25:478]. A thorough understanding of violin sound as well as finding suitable quantitative representations of violin sound is required to further violin timbre analysis. To a musician, a poor quality sound (or timbre) implies an unconvincing sound or a sound which contains audible playing faults. Such sounds are produced by poor technique or by not making a note sound well balanced or placed within its context. The latter is due to poor musicianship rather than to poor technique. Throughout this thesis, poor sound quality or timbre will refer to sounds affected by playing errors such as those associated with a beginner player while good quality sound will refer to well produced sounds. Neither of these terms in this text implies the standard of the recordings, nor the quality of the instrument used. In this chapter, how the human auditory system processes sound, paying particular attention to how a musician trains their hearing is briefly introduced as well as the relationships between pitch, timbre, Helmholtz motion and violin playing technique.

2.1 Hearing Sound and a Musician’s Training

Sounds exist as pressure differences in air which the human auditory system transfers into mechanical vibrations in the middle ear, liquid vibrations in the inner ear and finally as electrical impulses in the nerves leading to the brain. Audible sound is received as a sensation by the ear and passed to the brain where it is represented in the mind of the listener. The difficulty lies in understanding or trying to represent this type of “aural” imaging. A continuous frequency to place transformation takes place along the basilar membrane which has been represented in some work as non-linear frequency bands referred to as critical bands [Howard01, Noll93]. A brief outline of how the auditory system processes sound can be seen in Figure 2.1. For a detailed explanation of how the human auditory system functions, refer to [Moore82].

The variety of different types of sounds that the human auditory system deals with and the concept of collective perception are worth noting before focusing more specifically on violin timbre. There are certain sounds which immediately capture an

individual's attention such as warning signs. A well-known example would be the effect of scratching one's nails down a blackboard. Most people cringe at the sound and it has been traced back to our primate ancestry. Primates signal danger to the group by scratching their nails on trees in a downward motion [Schafer02]. The importance of this issue is the existence of sounds to which a large section of the human condition immediately react to, otherwise known as collective perception or learned response.

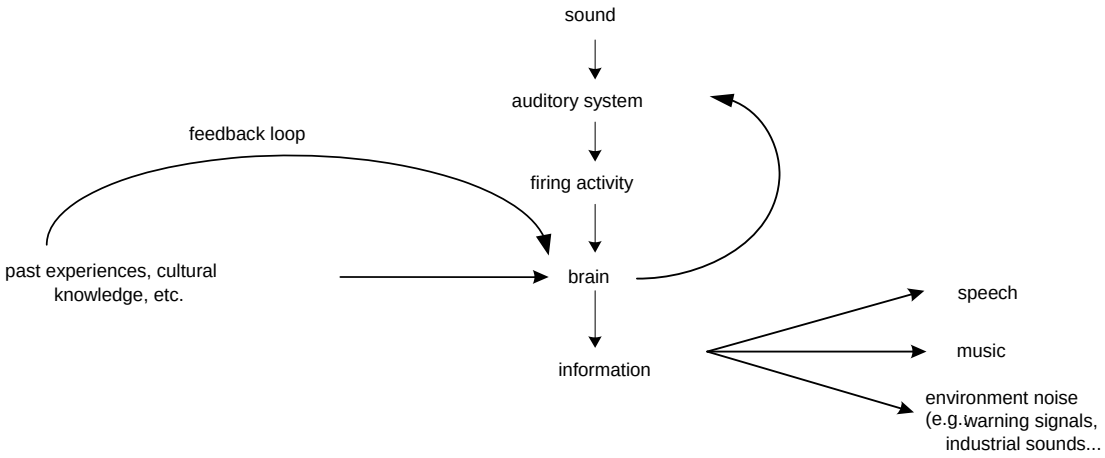

Figure 2.1: Processing sound via the human auditory system.

The human listener relies on a wide variety of information to help process sounds [Bregman90] and often manages to correctly interpret the information transmitted even if at times this information is faulty or unclear due to the presence of disturbance(s). A common example of this is listening to a person speak in a noisy environment, also known as "the cocktail party effect" and noted in 1953 by Cherry [Cherry53]. Similarly, a listener may have to compromise when listening to music. For example, if a wrong note is played, but the timbre is consistent, it is less evident to the listener, particularly in a fast noted passage. However, should a musician produce an unexpected "squeak" in the middle of a phrase it creates a temporary unpleasantness. Tone quality recognition is considered to be a type of pattern recognition and a skill a beginner violinist needs to develop along with the associated appropriate muscle memory.

Professional musicians highly develop their sensory perception through training and practice which can be thought of functioning as shown in Figure 2.2. Music is not just encoded in memory as an aural representation, musical memory is also encoded through other means, for example though fingering and bowing patterns, which require muscle memory. The development of these muscle memories is reliant on how the sound is perceived. This training allows for much more sensitive perception to develop.

Although musicians rely on multiple memory types, audition is the most important one, and the one which trains the muscle memories.

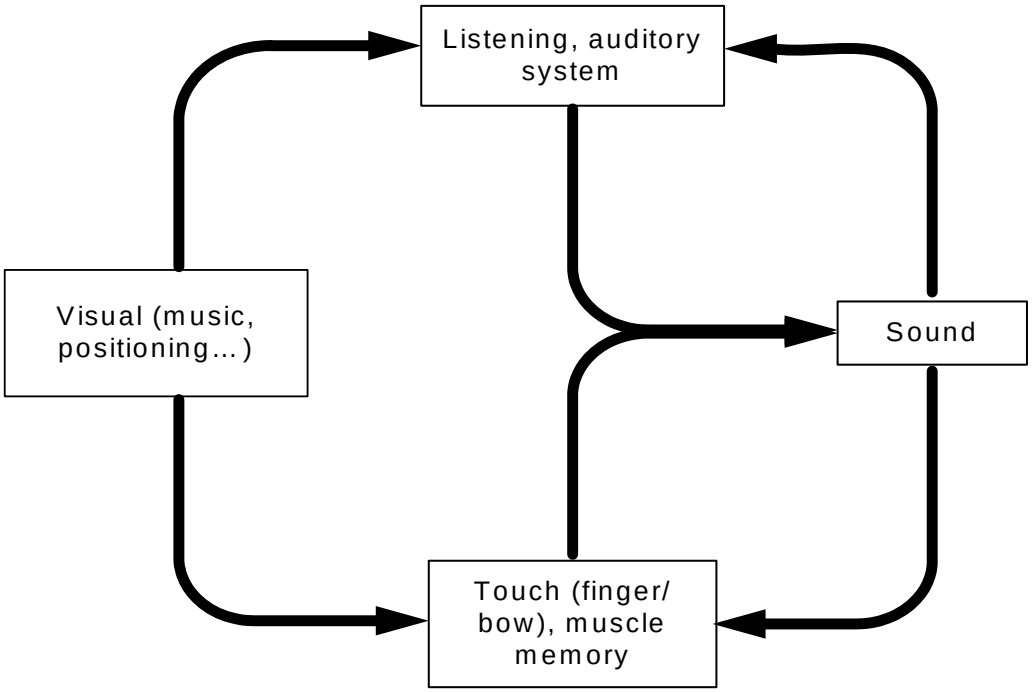

Figure 2.2: A musician's sensory system.

Musical information is transmitted not by independent sounds but rather through their relationships. A sound or a note means very little on its own. Every note in a phrase has its "weighting", just like words in a sentence. This is noticeably true for basic tasks such as hearing intervals, the distances between notes. Relatively few individuals have perfect pitch which is the ability to recognize the pitch name without a reference pitch being present. More importantly, music students are trained to listen and identify intervals, to develop relative pitch. Observations show that the mechanism for identifying intervals is independent of the ability to recognize pitch [Tuiguiane93]. The composer Hindemith (1895-1963) correctly pointed out that one must not think of music as a series of emitted tones but as a continuum [Hindemith40]. This is important as it indicates that humans rely on more than just pitch and tonality to understand music. Music psychologists sometimes refer to the mental "coding" that facilitates the perception and understanding of music [Narmour92]. The development of this mental coding in a violin student is important as it forms their understanding of pitch and timbre.

2.2 Pitch, Timbre and the Violinist

From *The New Grove Dictionary of Music and Musicians*, pitch is "the particular quality of a sound (e.g. an individual musical note) that fixes its position in the scale" [Sadie01:19:793]. Pitch is a perceptual attribute which is often used to describe a sound and timbre is what gives an instrument its characteristic sound. Schönberg captured the breadth of what is meant by timbre in stating that "tone colour is the large area of which pitch is one division" [Schönberg78]. This section considers pitch and timbre and how they specifically relate to violin sound and violinist, encompassing psychoacoustics and signal analysis as well as understanding how timbre perception has evolved.

John Puterbaugh's chronological timbre line gives an excellent overview of how the understanding of timbre perception has evolved over the centuries [Puterbaugh09]. Changes in instrument construction, musical style, genre and architecture are reflected by this. Already in 1758, Diderot and D'Alambert recorded that timbre was what differentiates types of sound [*ibid.*]. In the late 17th century, Hooke was able to show that pitch, as heard by a musician and measured frequency, are quite similar [*ibid.*]. In 1937, psychophysicists at Harvard began a series of investigations showing that the relationship between pitch and frequency is not one to one [*ibid.*]. This led to the search of a subjective scale of pitch and the subsequent Mel scale emerged. It is a scale judged by listeners where notes are of equal distance from one another and relates real frequency to perceived frequency [Stevens40]. Pitch and timbre are not independent. This is supported by the results of psychological investigations such as those completed by Miller and Carterette who researched the effects of pitch on the similarity of tones [Miller75]. Timbre is a multidimensional auditory attribute and there have been numerous attempts made, based on perceptual experiments using synthesised and recorded tones, to understand its underlying dimensionality. Grey worked on spatial solutions for representing timbral similarities between musical tones [Grey77]. The multidimensional scaling algorithm used in Grey's paper geometrically maps these subjective distance relationships.

Since the introduction of acoustical spectral analysis, it has become possible to observe the partial components of a sound. Harmonic sounds are periodic or approximately periodic sounds with a clear pitch salience and a spectral structure in which the main frequency components are evenly spaced. Each instrument has its characteristic vibration pattern and hence timbre, where certain harmonics are more

prominent while others are lacking. This is determined by the instrument body and modified by the player. For example, the violin generates many harmonics and produces a complex sound while its characteristic frequency resonances are the main reason why it sounds similar to the human voice [Winckel67]. Cleveland researched the spectral characteristics of timbre types [Cleveland77] and Grey did much work on the spectral fluctuation throughout the duration of a tone [Grey77]. The fluctuations are responsible for how timbre changes and the note evolves. Partials form an important part in creating timbre, as has been documented by Miller and Carterette [Miller75]. The amplitudes and frequencies of single partials of a sound spectrum can be changed greatly before a distortion of the tone colour is perceived. The psychologist Karl Stumpf (1848-1936), who is noted for his research on the psychology of music and tone *Tonpsychologie*, or tonal fusion theory and a major influence on Gestalt psychology, completed experiments demonstrating this by masking overtones by interference [Stumpf03]. Through signal analysis, it has been found that certain characteristics help create timbres, including formant prominence and location, pitch distribution, attack style and decay patterns of harmonics.

Attack style is of great perceptual importance in creating timbre as it contains important non-harmonic information which decay quickly but is characteristic of an instrument. It has been well documented that cutting the attack and decay transients of sounds leads to ambiguity [Miller75]. An instrumental sound of constant pitch and intensity loses its character to a certain extent if the attack is removed [Winckel67]. During a presentation at the *Sonorities Festival* at Queen's University, Bensa successfully demonstrated that it was close to impossible to hear any difference should the first 20ms of a piano note be dropped [Bensa04]. It was not noticeable at the lower frequencies, but to a highly trained ear it was only just about noticeable at higher frequencies when listening for it. Whether this holds for bowed stringed instruments remains to be seen, but seems unlikely given the results of Miller's work. A problem associated with its investigation is the lack of ability to determine consistently the attack period of a bowed stringed instrument note.

Violin pitch and timbre are not independent as bowing style influences harmonic content and pitch stability. Figure 2.3 illustrates the pitch fluctuations due to a fast bow attack compared to that of the steady pitch of a legato note on a violin open A string.

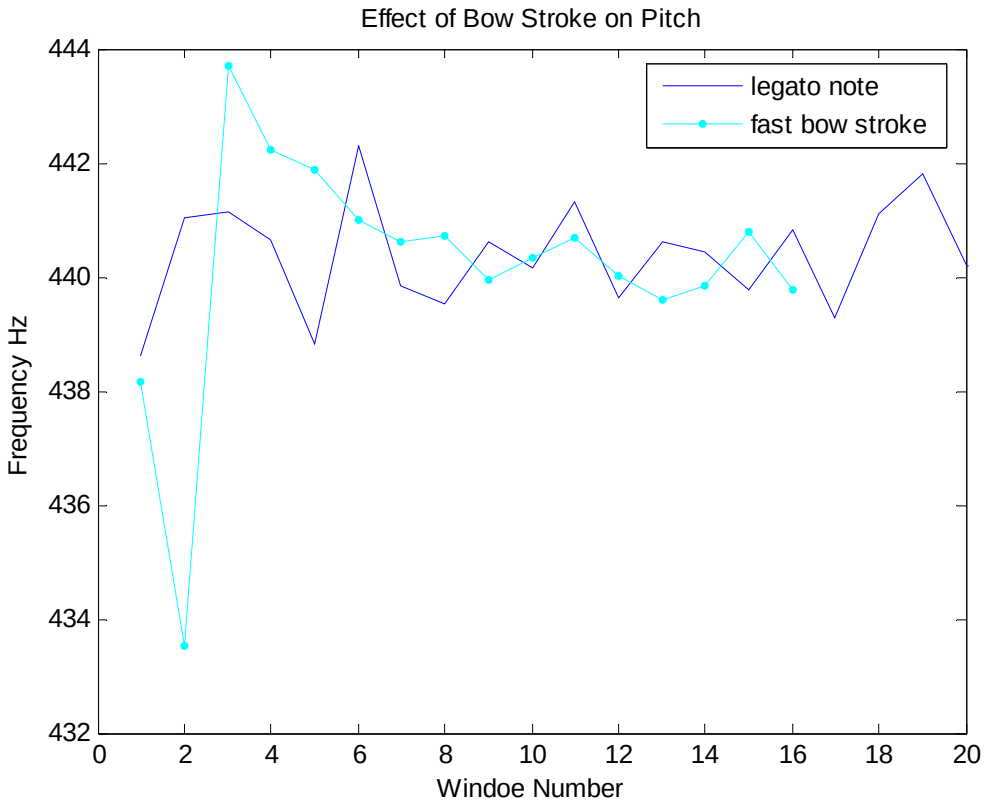

Figure 2.3: Effect of attack style on pitch.

The method applied for pitch detection is based on the distance between peaks obtained from the spectrum. As can be seen in Figure 2.3, the pitch of the note which was played with a much faster attack fluctuates by approximately 10Hz. This has been included to illustrate the difficulties associated with getting a computer to differentiate between acceptable pitch fluctuations and poor intonation for violin sound.

Another cause for acceptable pitch fluctuations is vibrato which is “the wavering effect of tone secured by rapid oscillation of a finger on the string which it stops” [Auer80:22]. Some research has been completed on the relationship between vibrato and pitch [d’Allessandro94, Brown96, Harrera98, Shonle80]. Detuning in the musical range is rendered less noticeable by vibrato, which causes pitch uncertainty. A psychoacoustics study confirming this common “musician’s knowledge” is presented in [Yoo98]. A very slight detuning within the spectral structure of sound seems to be evaluated as a positive sensation in the ear.

2.3 Violin Sound and Helmholtz Motion

Violin playing technique shapes the timbre produced and is dependent on the collective behaviour of several vibrations, which may be weakly or strongly coupled together. The

physics of the violin are very complex due to the very large number of variables influencing the sound. These variables range from the thickness of the wood used to the humidity of the air and are the constraints within which the player must work. Another group of variables are semi-fixed, i.e. choice of strings, quality of hair, type of rosin used etc. The next set of variables reflects how the instrument is played. For example, the relationship between the location and positioning of the bow on the string and sound produced. Central to this is how Helmholtz motion is established and maintained along a string by bowing technique.

When the bow is drawn correctly across the string, a rich harmonic spectrum can be maintained. On a vibrating string, Helmholtz observed a "V" shape moving along the string. This is known as the "Helmholtz Corner". When this "corner" reaches the bow, the friction switches from stick to slip. In an ideal situation, this cyclical switch between these different types of friction is known as Helmholtz motion. A simulation of this motion can be observed at Professor Joe Wolfe's web pages [Bows09]. This motion can be observed by using an oscilloscope and a strong magnet to induce a current along the string as described in [Woodhouse04]. This approach has been used by Wilson, in his work towards extracting violin performance information, specifically Helmholtz motion and how it is characterised by the speed at which the bow is drawn across the string [Wilson02]. Figure 1 in [*ibid.*] illustrates Helmholtz string displacement ranging from ideal to chaotic. Wilson's results prompted observing the effect bowing technique has on Helmholtz motion by emulating these experiments, the results of which are presented below.

The effect bowing technique has on Helmholtz motion is investigated through observing it in a set up similar to that detailed in [Woodhouse04]. The violin was placed on the table and oscilloscope connected to a string as shown in Figure 2.4. In this set up a nickel plated Neodymium magnet was used to induce a current along the string. Legato bow strokes as well as beginner bowing faults which have been described in the previous chapter, are emulated and the effect on the induced current observed. The limitations of the set up are that the violin cannot be played in its regular position, held under the chin and the data recording equipment, TiePie Engineering Handyscope HS4 version 2.85 [TiePie09], only lets segments of up to 2.5 seconds be recorded at a time. The results obtained are presented next.

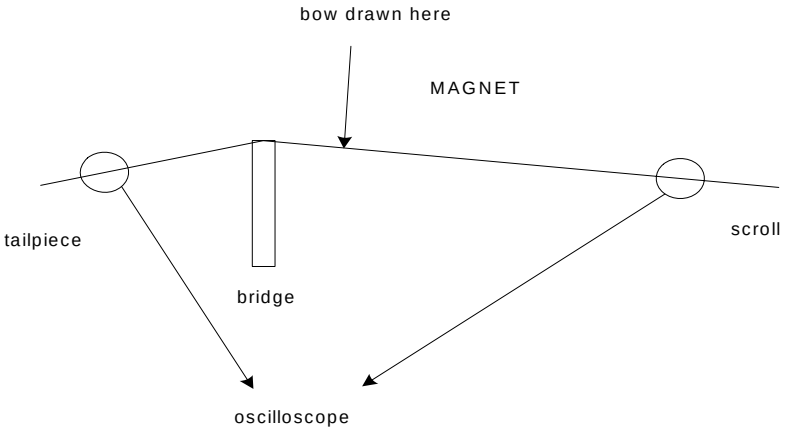

Figure 2.4: Set-up for observing Helmholtz motion.

2.3.1 Effects of Bowing Technique on Helmholtz Motion

In this section, the following figures display the effect of different bowing styles on the Helmholtz motion taken using the set up shown in Figure 2.4. According to Fletcher and Rossing, the characteristic Helmholtz waveform for a bowed string is a saw tooth waveform [Fletcher98]. In Figure 2.5, the effect a legato bow stroke has on the voltage readings is illustrated. The friction types switch evenly giving a regular shape but not quite the saw tooth waveform expected.

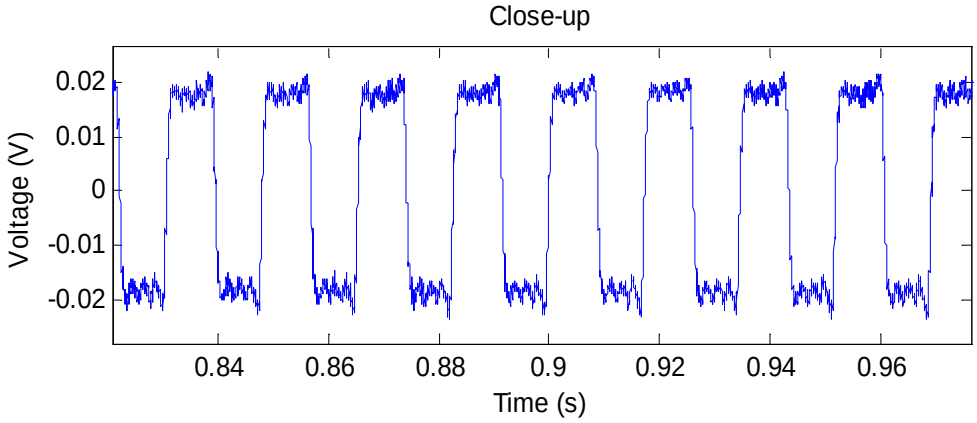

Figure 2.5: Helmholtz motion legato bow stoke.

In Figure 2.6, the voltage readings for a section of forced playing are given and no regular pattern is easily discernable. What has been termed by Fletcher and Rossing as “multiple fly back” motion is visible. This refers to the theoretical single return section of the Helmholtz motion being replaced by many “fly backs” with alternating signs.

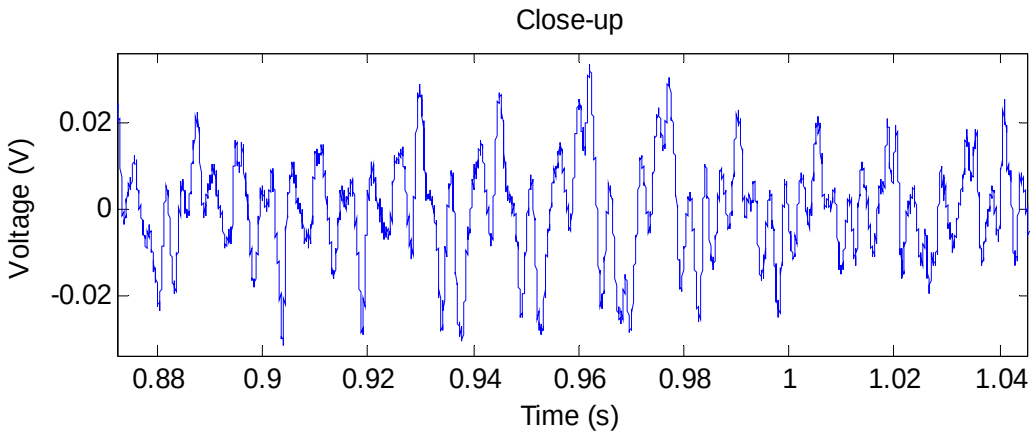

Figure 2.6: Helmholtz motion forced note section.

When the bow is drawn at an angle across the string and "skating" is emulated, the effect on voltage readings is illustrated in Figure 2.7.

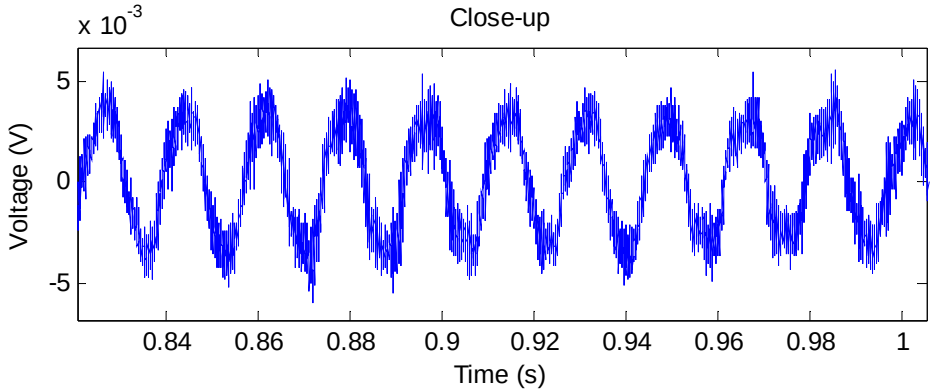

Figure 2.7: Helmholtz motion effect of emulated skating.

If not enough force is applied to the string, the bow cannot hold the string during the "stick" part of the cycle and the typical periodic Helmholtz motion does not develop. At the other end, should too much force be applied to the string, a slipping phase cannot be started consistently and therefore the cyclical stick-slip motion cannot be continued. Comparing the voltage readings between these three figures, the emulated skating bow stroke has the lowest voltage levels. Schelling estimated maximum and minimum Helmholtz motion limits [Askenfelt89], outside of which, the cyclical stick-slip motion cannot be maintained. This is relevant as below a certain level, the bow cannot hold the string during the sticking phase. Too much force causes the stick-slip oscillations to break down due to difficulty in starting the slipping phase.

From conducting this experiment, it can be shown that Helmholtz motion is effected by bow stroke style but good sound is not strictly limited to the ideal Helmholtz motion. This observation is supported by existing research and is best summed up by Sacksteder: “although many aspects of Helmholtz resonance have been observed, measured and calculated for centuries, there seems to be no clear consensus about how to relate it to the foundations of acoustics” [Sacksteder87]. As interesting as these results are, the approach is not practical for obtaining violin timbre features for use in this work because they cannot be easily extracted and used in a classifier. Even if a more user friendly set up were to be designed, the issue remains as any information obtained through this set up is in real-time and not compatible for use in a classifier using feature vectors. However, this is the first exploration of the effects of bad violin playing using such an approach.

2.4 Summary

The difficulties and complexities associated with understanding and representing the human auditory system, auditory perception, in particular musicians’ trained hearing and current research have been presented in this chapter. It focuses on musicians’ training and perception, the effect playing technique has on timbre, pitch and Helmholtz motion. Inducing a current along a bowed violin string allowed Helmholtz motion to be observed. The effects of legato bowing and emulated playing faults of forcing and skating have on the Helmholtz motion have been displayed. Emulating these faults shows that the conditions under which a cyclical stick-slip motion exists cannot be maintained due to poor bowing technique. Although informative, the results obtained from this Helmholtz motion study are not practical for extracting violin timbre features. To carry out the research aims of detecting overall violin note sound quality and playing faults, finding a suitable way of representing violin notes which reflect the change perceived by musicians is needed. This begins with considering the expressions used by musicians to describe violin sounds and finding note samples which can be linked to these expressions. The next chapter details the violin note dataset and the listening tests.

3 The Dataset and Listening Tests

In the previous chapter, research relating to sound analysis and psychoacoustics has been presented with some of the difficulties associated with defining pitch and timbre quantitatively. The aims set out in this thesis focus on how violin sound can be represented so that a system capable of differentiating between beginner and professional standard players as well as identifying playing faults is established. At the time of this work, no research existed on exploring the acoustic violin timbre space from a playing technique and sound analysis perspective. Existing work assumes good sound or that associated with a professional standard of violin playing from acoustics [Hutchins97, Fletcher98], to information retrieval [Wilson02] and timbre classification [Agostini01, Eronen00]. Information relating specifically to overall violin note quality and playing faults is sought. Apart from representative measures, qualitative expressions describing the sounds are needed as well as a means of linking the two together. For this to occur, a suitable dataset with qualitative labels must be acquired which meets the requirements set out by the thesis' aims. In this chapter, the dataset requirements are outlined, information about existing instrument samples and how the dataset was obtained are detailed. This is then followed by the listening tests run.

3.1 Dataset Requirements

The research aims, as set out in Chapter 1, involve establishing a system capable of identifying beginner note samples from professional standard legato ones as well as detecting playing faults. At the time of this research, no dataset consisting of beginner violin notes existed so one had to be made which includes corresponding professional standard legato note samples. Having a suitable dataset on its own is not sufficient and the opinions of professional string players regarding overall sound quality and descriptions of the samples are sought. Musicians often use qualitative or onomatopoeic terms to explain a desired effect or playing fault. Some such examples include crunching, where the player uses too much bow pressure and as a result the sound cracks and skating, where the bow is drawn across the string at an angle causing the bow to skid along the string. The question arises as to how to quantify such terms. This requires a dataset in which each sample has been assigned its appropriate qualitative

expression label(s). A dataset comprising of equal numbers of professional standard player and beginner player note samples is needed. The professional standard player note samples serve as a reference to which the beginner notes can be compared. The bow stroke used for these reference samples is legato, which is the bow stroke a violinist must master before progressing onto other strokes which require more bow control, such as staccato. It is appropriate to consider a robust system for fault detection and one that is not dependent on sample length or pitch as these two descriptors are different for most if not all samples. The ultimate aim is to find features which can be applied to the note independent of its length or pitch and which can be used for representing the violin timbre space to which qualitative expressions have been assigned through listening tests. Both beginner and professional standard player samples need to be collected in the same environment using the same equipment. This keeps the dataset as uniform as possible and by using the same instrument and set up, the number of variables has been reduced allowing the work to focus on playing characteristics by making the recordings more readily comparable.

3.2 Available Datasets

Having considered the dataset requirements as set out by the thesis aims, some information relating to existing instrumental sample collections is now given. Many commercial instrument sound sample libraries such as the Vienna Symphonic Libraries [VSL09] and samples from the London Symphony Orchestra Samples [LSO09] are available. Several professional standard recordings of different orchestral instruments are available free of charge from the Electronic Music Studios at the University of Iowa [UofI09]. The Real World Computing (RWC) Music Database also provides instrumental samples including violin samples [RWC09]. McGill University has produced a CD consisting of musical instrument samples which is available for purchase [McGill09]. As with seemingly all sample libraries easily available for download, individual violin legato notes are included but no beginner note samples are available. At the outset of this research, no database consisting of both professional legato and beginner violin note samples existed and the dataset requirements for this work are still not met by available instrument sample collections. In the following sections, the instrument used, the recording process and dataset samples are presented.

3.3 The Recording Set Up and Dataset Samples

The dataset was made using the best microphone available in a recording studio having a very dead acoustic. A Beyerdynamic M201TG dynamic microphone with hypercardioid polar pattern was used and placed as close as possible to the f-holes without disturbing the bow arm. The track was recorded onto DAT and saved as monophonic 16-bit, 44.1 kHz format wav samples. The same recording studio, set up, violin and bow were used for recording all samples in the dataset.

An old French violin was used with a modern, 60g well-balanced bow. It is a relatively large violin which speaks easily and evenly throughout its frequency range and has a big, clear sound. It is an instrument that a beginner is able to play easily. At the time of these recordings, the strings on the instrument were Thomastik Dominant Mittel for the G, D and A strings and a Pirastro Oliv E string.

The dataset made consists of 88 beginner note samples and 88 professional standard legato notes. Each sample contains one note only and the average sample length is 1.88s. The pitch range of the dataset is any note which is played in the first position, which is the lowest possible position on the violin, i.e. open G3 to B5, fourth finger on the E string. Two professional standard players and three beginner players recorded samples. The player breakdown for the beginner samples is as follows: 18 from player one, 19 from player two, and 51 from player three. For the professional standard legato notes, 44 samples were taken from each player. The next step involves labeling the dataset as perceived by professional musicians. Through these opinions, qualitative tags will be associated with each sample. The listening tests used to source these labels are detailed in the following section.

3.4 Listening Tests

The research aims are to find a system capable of detecting overall violin sound quality, i.e. professional standard versus beginner and playing faults. One important part of this involves finding suitable quantitative representations of violin sounds. Another part is to capture professional standard musicians' perceptions of violin sound quality and appropriate descriptions. The latter requires listening tests to be conducted which are designed to collect this information. Musicians often use qualitative expressions to describe sound. In an attempt to meet the research aims, these expressions need to be linked to one or more samples in the dataset. Once this information has been gathered, a

way of representing the samples by quantitative measures can be undertaken. The dataset is to provide a means by which the qualitative expressions and quantitative measures can be bridged together via the listening tests.

The listening tests target professional violinists in particular but to increase numbers, cellists and violists have also been included. Many of the sound faults, due to bowing technique on the violin have an equivalent on other bowed stringed instruments. The listening group consisted of 21 string players, 19 of whom are professional musicians, performing and teaching and the other two are violinists of professional standard of playing, but are not making a living as musicians. More specifically, the group consists of 11 violinists, one violist, four cellists, and five musicians who play both violin and viola.

Listening Test Terms Explained

1. **Terms associated with overall timbre quality:**
   (Listener to select only one)

very poor	→ significant playing fault/s dominates sound
poor	→ playing fault/s present in sample
reasonable	→ sound is predominantly good, contains a small playing fault
reasonable no fault	→ sound is good but there is room for improvement in the timbre.
good	→ no playing faults, good confident sound
excellent	→ note where instrument is perceived to resonate at its best

2. **What is meant by the terms associated with sound characteristics:**
   (Listener may select as many as appropriate)

Crunching	→ the sound breaks due to too much bow pressure occurring anytime within the duration of the note.
Skating/whisping/whistling	→ sound due to bow being on an angle and skidding down the string as opposed to going across it.
Uncommitted/nervous sound	→ player 'chickens' out, not enough pressure, poor contact with string, sound may be reasonable, but fluctuations in timbre or pitch can be perceived
Intonation problem	→ not in tune
Bouncing bow	→ poor bow control and tension in bow arm leading to the bow bouncing along the string (right hand vibrato).
Tips another note or string	
Ends too suddenly	
Poor finish to note	→ not clean finish due to lack of bow control, may include faults such as crunching
Poor start to note	→ not clean start due to lack of bow control, may be hesitant, may include faults such as crunching.
Good	→ no noticeable faults

Figure 3.1: Document explaining listening test terms.

Listening tests have been devised so that each sample has at least one qualitative expression, an overall quality grade of between 1, poor and 6, excellent as well as a beginner or professional player label. The outcome of these listening tests provide the *a priori* labels for the classification process from which perceptual correlates for violin timbre may be established. The listeners received no training, only a copy of the explanation of the terms and of the testing procedure steps, copies of which can be seen in Figure 3.1 and Figure 3.2 respectively.

Listening Test Instructions

1. **Listen to note once. Test progresses at your speed.**

2. **Column 2: Grade the overall sound quality between 1 and 6 where**
   1 - very poor
   2 - poor
   3 - reasonable
   4 – reasonable no fault
   5 - good
   6 - excellent
   NB: selecting good or excellent implies no tone fault; reasonable no fault, the sound is predominantly good with no distinct fault but could be better; reasonable implies that there is a disturbance in the tone at some point; poor has at least one fault and very poor contains multiple faults.

3. **Column 3: Please associate sound with a beginner or a professional player.**

4. **Column 4: Please tick as necessary the sound characteristics which best describe the note. NB: a sound may contain more than one of these characteristics.**
   □ crunching
   □ skating/whisping/whistling
   □ uncommitted/nervous sound
   □ intonation problem
   □ bouncing bow
   □ tips another note or string
   □ ends too suddenly
   □ poor finish to note
   □ poor start to note
   □ good (no noticeable faults)

5. **Column 4: Please add in any additional comments about the sound which you feel have been omitted by the previous sections in this final column.**

Figure 3.2: Listening test instructions document.

The speed at which the test progressed was controlled by the listener, but each sample can only be played once. AKG K240 “Monitor” headphones [AKG09] were used and samples were accessed and played through Matlab [Matlab04]. As soon as the listener activated the testing/listening program, a random play list was generated consisting of all samples from the dataset. The exact list for each listener only became

available at the end of the listening test. Each listener completed a box as shown in Figure 3.3 for every sample. After the test data had been collected, the consistency of the results was inspected, after which normalising this information provided the *a priori* labels required in the classification process.

As can be seen from Figure 3.3, the listeners could leave comments. The comments received from the listeners fall into two groups. The feedback either specified the approximate temporal location of the fault, i.e. slight crunch in middle, or related to the sound quality of individual samples where the listener felt that the existing descriptions were lacking in detail. When asked how they (the listeners) perceived a sound to be produced by a beginner player rather than a professional standard player, when no distinct faults are present, the replies all referred to either intonation, note texture, to the relative proportions of the note or to overall consistency or balance. The listeners were not given the option of replying "do not know" relating to the beginner or professional player choice and were deliberately forced into making a decision by the listening test format. Several listeners did point out that they genuinely found making this decision very difficult for three or four samples. Another point that was raised by the listeners is the specific case where a sound is marked as being "faultless", an overall quality rating of 4, but still is associated with a beginner player. The term "a good beginner sound" emerged. The listening tests provided much information which had to be checked for consistency, then normalised to create an average listener, a process which is presented in the next section.

#	Overall Quality	Beginner or Professional?	Sound Characteristics (please tick as necessary)	Any Additional Comments
			□ crunching anytime during note □ skating/whisping/whistling □ uncommitted/nervous sound □ intonation problem □ bouncing bow □ tips another note or string □ ends too suddenly □ poor finish to note □ poor start to note □ good	

Figure 3.3: Copy of listening test form for each sample.

3.5 The Average Listener

The main reason for running listening tests is to establish an average or normalized listener, a ground truth, for use as *a priori* labels in a classifier. Before creating a

normalized listener, the listeners' perception is verified for consistency. Consistency in this case involved checking that the range of grades obtained for each sample is acceptable, i.e. no one sample has grades returned of both 1 and 6 by the listeners. Should this happen, a mistake has been made or a problem exists with the test design or procedure indicating that new listening tests would be required. Fortunately, this was not the case. The range and mean grade for each sample are displayed in Figure 3.4 where the mean is shown by an asterisk.

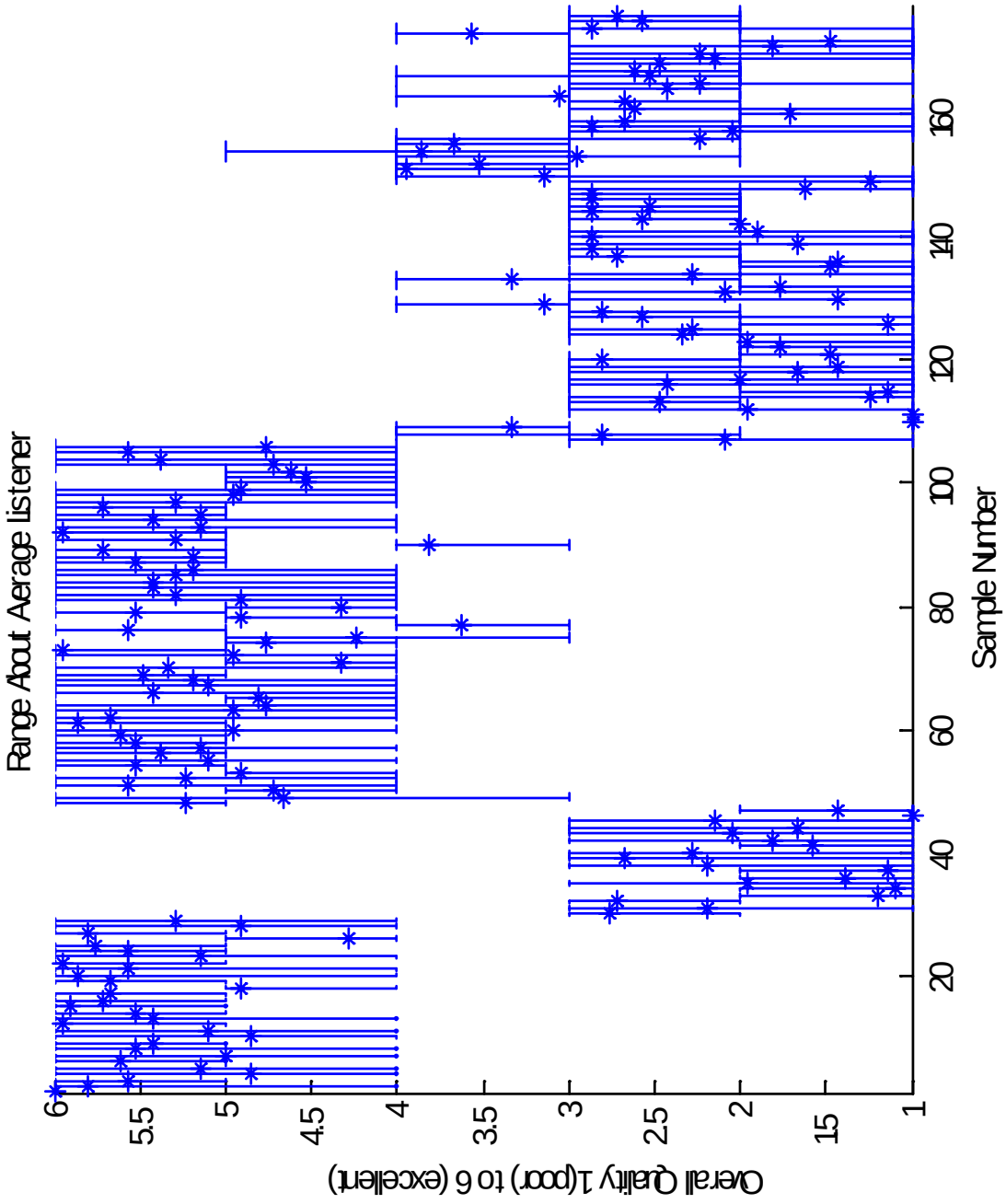

Figure 3.4: The listening group's overall sound quality grading range.

The sample numbers in this figure represent the beginner and professional standard legato note samples in alphabetical order. This is because the test sample order was randomised for each listener. Knowing this, allows for a clearer understanding of the grouping shown in Figure 3.4. The numbering of the dataset is as follows: samples one to 29, professional standard samples, samples 30 to 47, beginner notes, samples 48 to 106, professional standard notes and from sample 107 to the end, are beginner notes. Consistency, having been found to be acceptable, means that these results can be used to provide the *a priori* labels required in the classification process.

In addition to giving an overall sound quality grade to each sample, the listeners had to state whether the sample was produced by a beginner or by a professional player. Out of the 176 samples, 94 have been labelled "beginner" and the remainder "professional". There are six professional standard legato note samples which have been labelled as beginner which are detailed in Table 3.1. Of these samples, the two with the lowest quality grades are to be noted as the quantification of these samples is of particular interest comparatively to the other samples in the dataset.

Sample No.	***Grade***	***Beginner/professional?***	***Faults perceived?***
Legato 24	4.2857	Beginner	bow bouncing
Legato 47	4.3333	Beginner	none
Legato 50	4.2381	Beginner	none
Legato 52	**3.619**	**Beginner**	**none**
Legato 56	4.3333	Beginner	none
Legato 71	**3.8095**	**Beginner**	**none**

Table 3.1: Professional standard legato note samples labelled as "beginner".

For Task II, fault identification, consistency was also verified for fault presence in much the same way as for Task I and the results were found to be acceptable. From the listening tests, Table 3.2 gives the number of times each fault is recorded as having been perceived. The results shown are taken from the mean perception which has been obtained by summing up the number of times a fault is identified by all listeners in a sample and divided by the number of listeners. Based on these results if a majority of players had perceived the fault presence, the fault is considered to be present. Fault three, which is "nervousness", is the most prevalent fault in this dataset.

Fault	***No. of Samples Present***
crunch	33
skate/whistle	30
nervousness	57
intonation	30
bow bounce	16
extra note/sound	15
sudden end	30
poor start	24
poor end	37

Table 3.2: Breakdown of fault perceived presence in dataset.

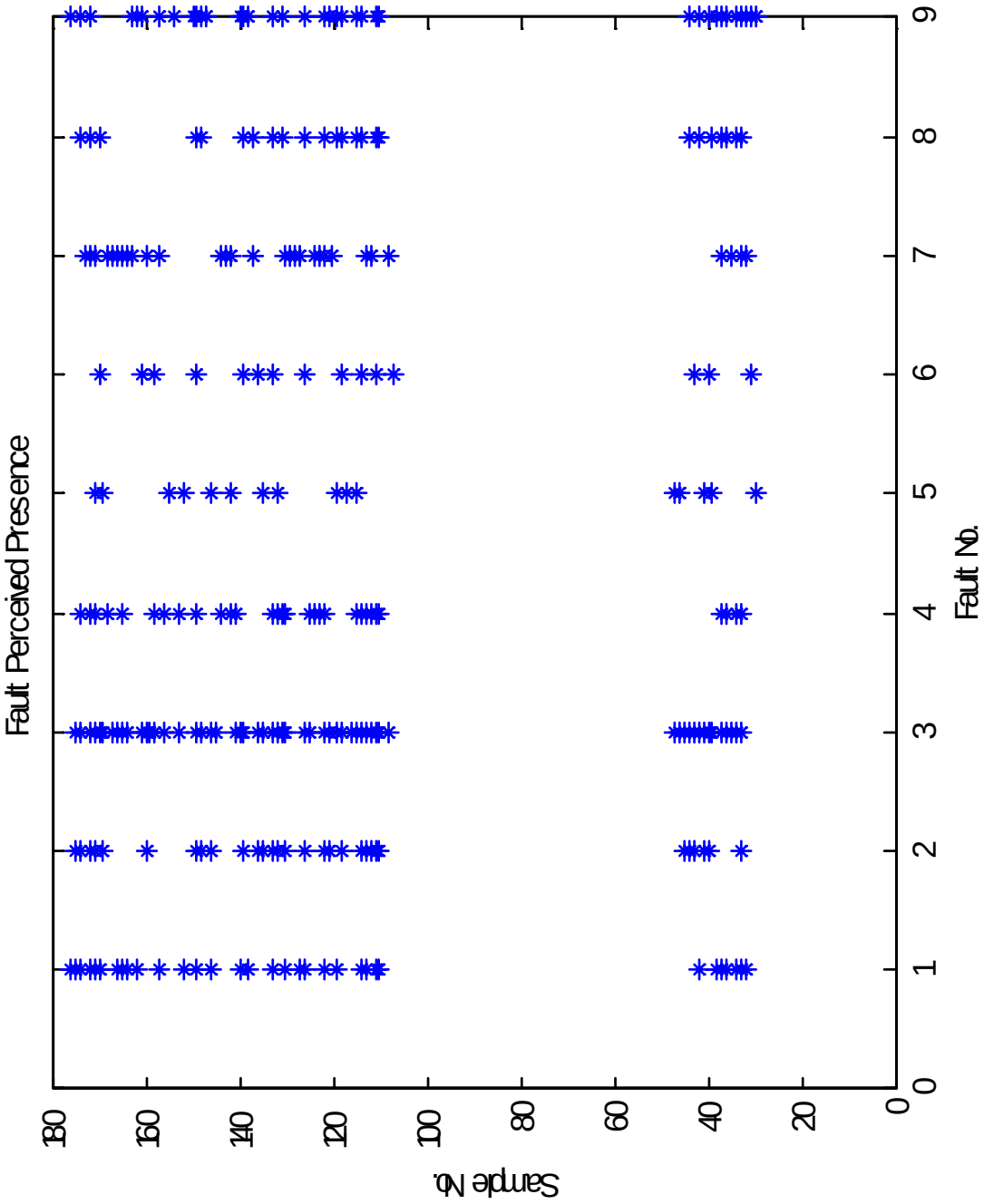

Figure 3.5: Mean perceived faults in all samples, same sample order as in Figure 3.4.

In most samples, more than one fault is present as can be seen in Figure 3.5 which shows perceived fault presence in all samples. Some interesting observations can be drawn from the results displayed in Figure 3.5. Faults bow bouncing and extra note have not been perceived as being present together in any sample. Given the descriptions assigned to these faults, this is good because sometimes a bow bounce can be perceived as an extra note and vice versa. In these recordings, it has to be one or the other, not both simultaneously. In this dataset, faults mostly appear together rather than

independently, as summarised in Table 3.3 where only 16 samples have been identified as having one fault.

As displayed in Table 3.3, playing faults crunch, skate, intonation and poor start have not been perceived as occurring independently. “Poor finish” has occurred on its own in three samples. Of the 33 samples which have crunching, 30 of them also contain sudden end, and/or poor start, and/or poor finish, supporting the previous statement. Crunching tends to occur more often at the starts and ends of notes. Skating always occurs with player nervousness in this dataset, but the reverse does not always hold. To better illustrate multiple fault occurrences, Table 3.4 gives the percentages of samples containing the two indicated playing faults.

Fault	***Perceived Independent Fault Occurrence***
crunch	0
skate/whistle	0
nervousness	3
intonation	0
bow bounce	3
extra note/sound	1
sudden end	6
poor start	0
poor end	3

Table 3.3: Perceived independent fault occurrence.

	CR	**SK**	**NV**	**INT**	**BB**	**XN**	**SE**	**BAD S**	**BAD E**
CR	100	45	75.76	48.48	12.12	18.18	39	48.48	67
SK	50	100	100	46.67	20	33.33	26.67	46.67	53.33
NV	43.86	52.63	100	43.86	19.3	22.81	26.32	40.35	43.86
INT	53.33	46.67	83.33	100	13.33	16.67	46.67	46.67	46.67
BB	30.77	37.5	69	30.77	100	0	12.5	19	19
XN	40	67	86.67	33.33	0	100	0	53.33	67
SE	43.33	26.67	50	46.67	6.67	0	100	16.67	23.33
BAD S	67	58	95.83	58	12.5	33.33	20.83	100	87.5
BAD E	59.46	43.24	67.57	37.83	8.11	27.03	18.92	46.67	100

Table 3.4: Percentages overlapping faults.

3.6 Summary

The dataset requirements needed to fulfill this thesis’ aims have been outlined and information relating to how the dataset was obtained including players, instrument, set up and recording process have been detailed in this chapter. Information relating to available violin note samples has also been included. Working towards the thesis’ aims relies on being able to establish a link between the qualitative and the quantitative descriptions of violin timbre. This involved listening tests through which, each sample was assigned a label linking it to one or more of the playing expressions used by musicians, given an overall sound quality grade between 1, for poor and 6, for excellent as well as a beginner or a professional player label. This allowed the subjectivity to be

removed and a sense of professional string player perception to be documented. Most importantly, an average listener has been established by first checking then normalising the listeners' perception of the dataset. These results are to be used as the *a priori* labels for use in the classification process. Now that qualitative labels have been assigned to each sample, methods of representing the dataset via quantitative measures will be presented. In the next chapter, the effect of violin playing technique on waveform and timbre is illustrated, after which, in subsequent chapters, suitable quantitative measures are sought for representing the dataset.

4 Effects of Violin Playing on Waveforms and Harmonic Content

So far in this thesis, background information pertaining to the thesis' aims of getting a computer to classify violin sound quality and to detect playing faults has been detailed. This includes the dataset and listening tests for the proposed tasks which have been presented in the previous chapter. The dataset consists of an equal number of beginner note and professional standard legato note samples over a range of pitches. In this chapter, the relationship between playing characteristics due to poor playing technique and their observed effect on waveforms is presented. As multiple playing faults are possible, this work has been limited to nine faults found in five main categories. These five fault categories reflect the main waveform disturbance patterns and locations observed. They are onsets, offsets, amplitude, unevenness, and asymmetry. The qualitative fault descriptions used in previous chapters are discussed in terms of playing technique and fault category in this chapter. In Section 4.1, each fault category is presented individually summarised in Section 4.2.

4.1 Main Playing Faults Categories

From visual inspection of the dataset's waveforms, much variability is observed within the waveforms produced by both professional and beginner violinists. The beginner note waveforms though show much greater variability. Standard waveform analysis can rarely be applied as identifying the different sections (attack, steady-state, decay) of a violin note from its waveform alone, is difficult. Beginner notes often have certain unwanted characteristics, some of which are visually discernable in the waveform. These characteristics have been grouped from visual inspection into five categories: onset, offset, amplitude, unevenness and asymmetry around the abscissa. The first four categories are directly related to bow control. The causes for the non-symmetry visible around the abscissa in certain samples are not known precisely but are most likely linked to the player's bow-technique, effecting the sound produced. As the playing faults being considered result from poor bowing, they are not independent. This means that more than one fault is often present at the same time. This section considers each

fault category individually and professional standard player note waveforms are contrasted with those belonging to beginner players. In each category, its associated qualitative expressions are detailed. For the actual classification tasks, the faults are named using their qualitative descriptions.

4.1.1 Onsets

Onset refers to the start of a musical note. The onset or attack is very important for establishing an instrument's timbre. How onset style effects a note's waveform and harmonic structure are presented below. A stringed instrument has different types of attacks, reflecting different string excitations. In Figure 4.1, the waveforms of three standard violin onsets are given. The sudden attack of a plucked note, the quick attack of a fast bow stroke and the gradual onset of a legato note are illustrated.

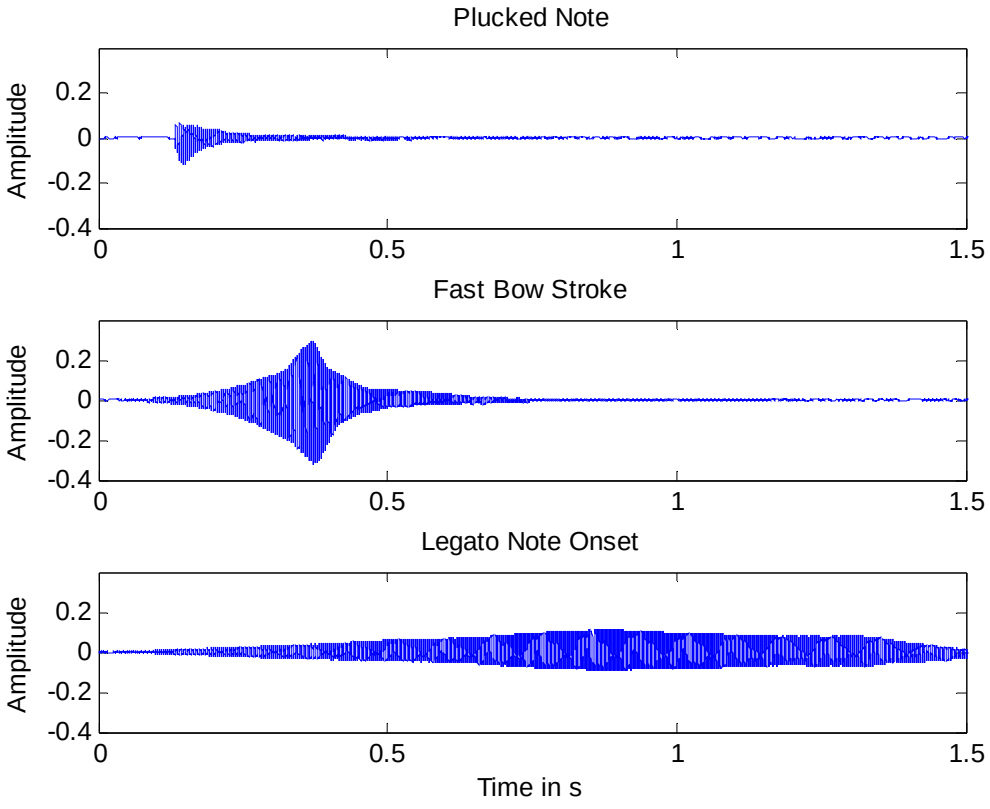

Figure 4.1: Three standard violin waveform onsets: plucked note (top), fast bow stoke (middle) and legato note (bottom).

Bowed stringed instruments have two types of onsets, separate and slurred notes. Separate means a bow change occurs and slurred implies that at least two notes are played in the same bow stroke. The effect of these different onsets on their waveforms is illustrated in Figure 4.2. Spectrograms have been included to help visualise the pitch

changes. Although the same pitch is played throughout the right hand images in Figure 4.2, the effect of changing the bow smoothly is observed. The notes played in the sample on the left are A3 B3 A3 B3 A3 B3.

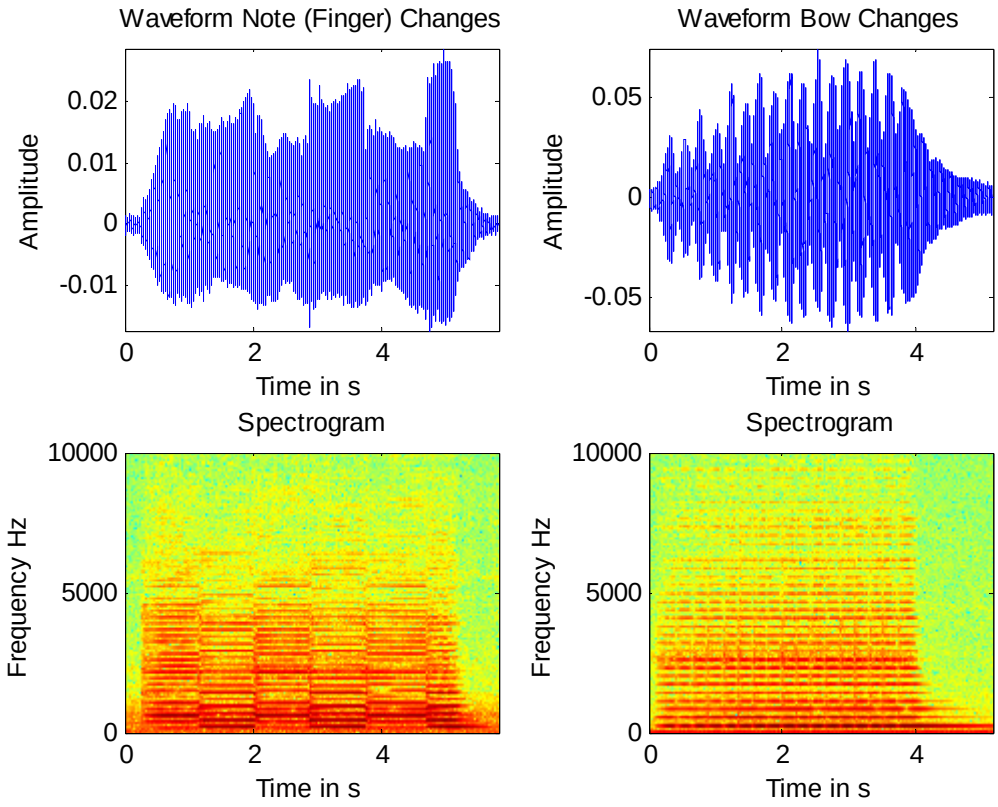

Figure 4.2: Note changes waveform and spectrogram (left) and bow changes waveform and spectrogram (right).

The waveform attack section of a professional standard legato note onset compared to that of a beginner's note displayed in Figure 4.3, illustrating the relative waveform smoothness of the legato note sample. A beginner violinist lacks the bow control necessary to achieve clean and precise onsets resulting in the note not being fully established. Too much bow pressure is often used which leads to "crunching". If not enough pressure is applied and the note is not started cleanly, there is unwanted noise present which is not specifically "crunching". This effect is referred to in this work as a poor start. The waveforms of various violin note onsets have been illustrated. The effect of different onsets have on the harmonic content is presented next.

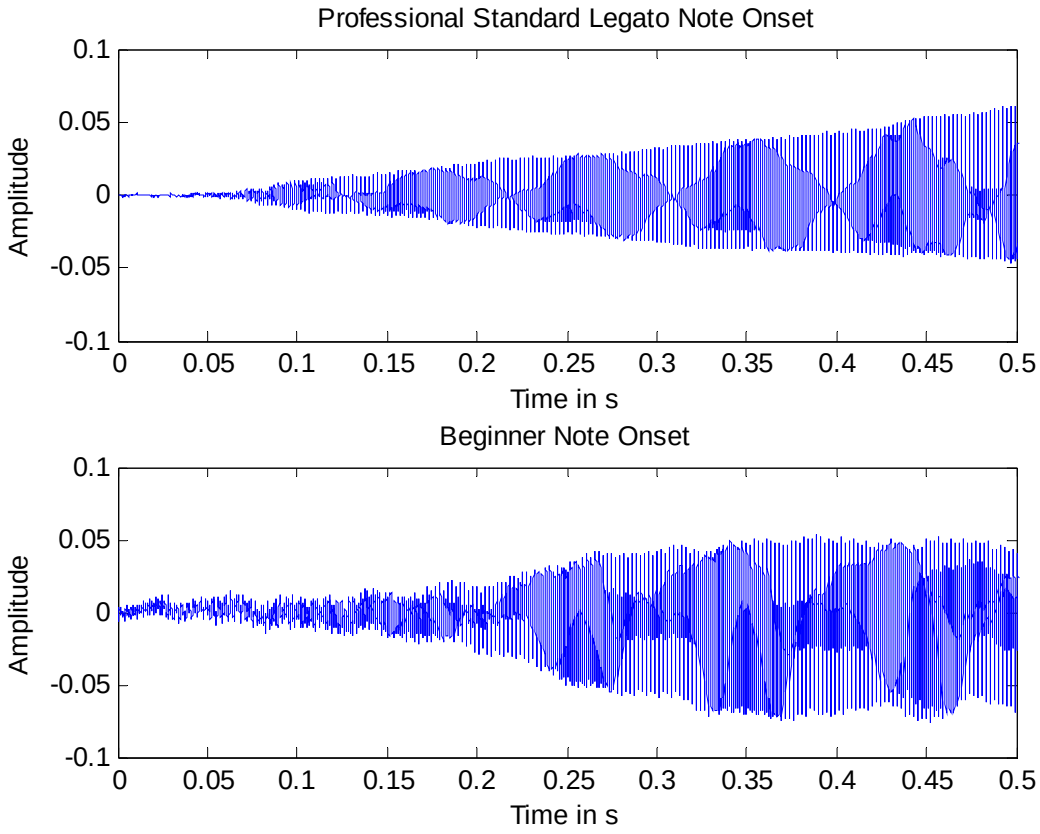

Figure 4.3: Professional standard legato (top) and beginner (bottom) note onsets.

The time-frequency representations of the samples illustrated in Figure 4.1 are given in the following two figures. Their spectrograms and CQT representations are displayed in Figure 4.4 and Figure 4.5 respectively. The spectrograms shown in Figure 4.4 are STFT based and have been obtained by using a 1024 point window Hamming with 50% overlap. The samples in this figure are not the same length, so only the section comprising of the first 1.5s of the professional standard legato note (bottom image) is used in Figure 4.4 and Figure 4.5. In the spectrograms of the fast bow stroke (middle image) and the legato note (bottom image), more harmonics are excited for longer resulting in very different timbres to each other and to the plucked note spectrogram. The evolution of harmonics over time reveals much about the sound. Richer harmonic content can be seen in the legato note sample compared to the fast bow stroke and plucked note samples. The plucked note has the most sudden attack (top image) and comparatively few harmonics are excited and those that are, dissipate immediately. A similar temporal evolution of the harmonics is observed in the CQT representations of these samples which are illustrated in Figure 4.5.

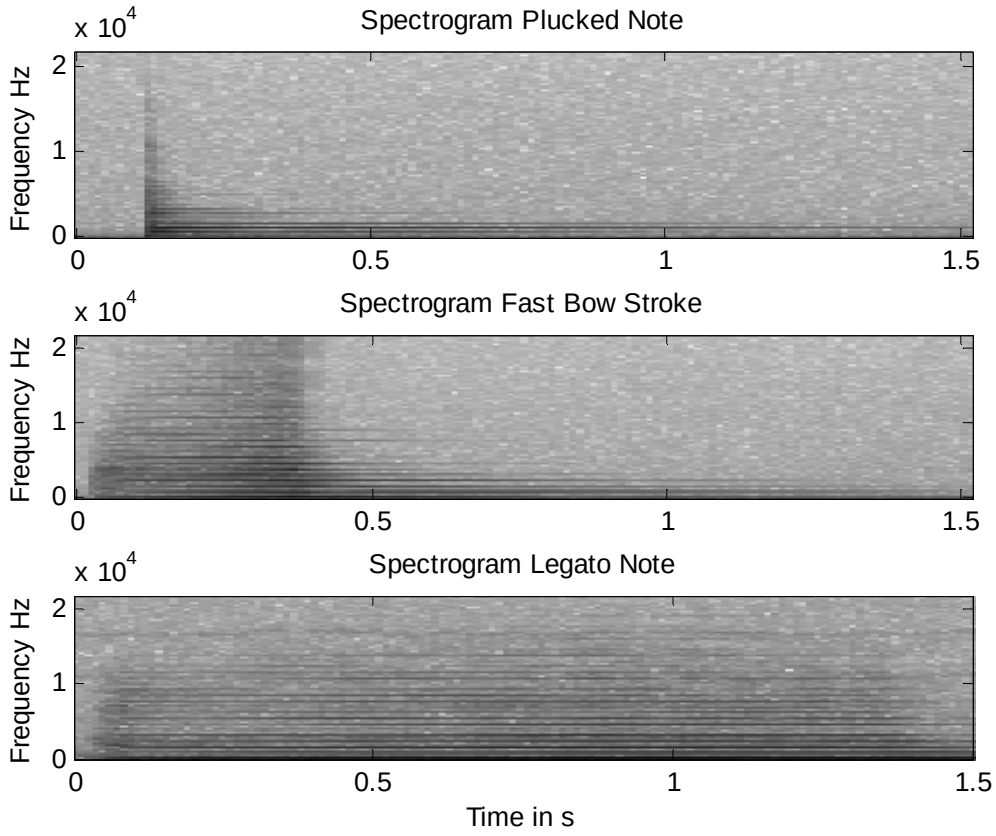

Figure 4.4: Spectrograms of Figure 4.1 waveforms.

From the qualitative terms used in Chapter 3, beginner player onsets are associated with crunching and poor start to the note. These are typical beginner onset faults and are caused by an inappropriate amount of bow pressure being applied. Beginner players often crunch at the start and end of notes. This crunching is due to stiffness in the bow arm and hand which does not allow bow pressure to fluctuate with respect to the bow's speed and position along the string. The arm should be relaxed and supple but firm, hanging from the shoulder, letting the shoulder muscles do all the work. Also linked to this stiffness problem are poor bow hold and lack of "feel" for the bow which is learned over time. The "feel" of the bow refers to the confidence and experience the player has with their technique to be able to alter effects such as pressure, angle and speed with smoothness and ease. These three elements are the right arm expression tools and they depend on a loose, relaxed bow arm. A poor start is not a clean start and has more to do with hesitation but does not go as far as crunching and is not easily visible on a waveform. To avoid crunching, Auer advises "hold[ing] the bow lightly, yet with sufficient firmness to be able to handle it with ease [and to resist from] bring[ing] out a big tone by pressing the bow on the strings" [Auer80:20]. To increase the tone, finger

pressure not arm-pressure should be applied “thus avoiding forcing the tone which otherwise grows rough” [Auer80:21], i.e. crunching.

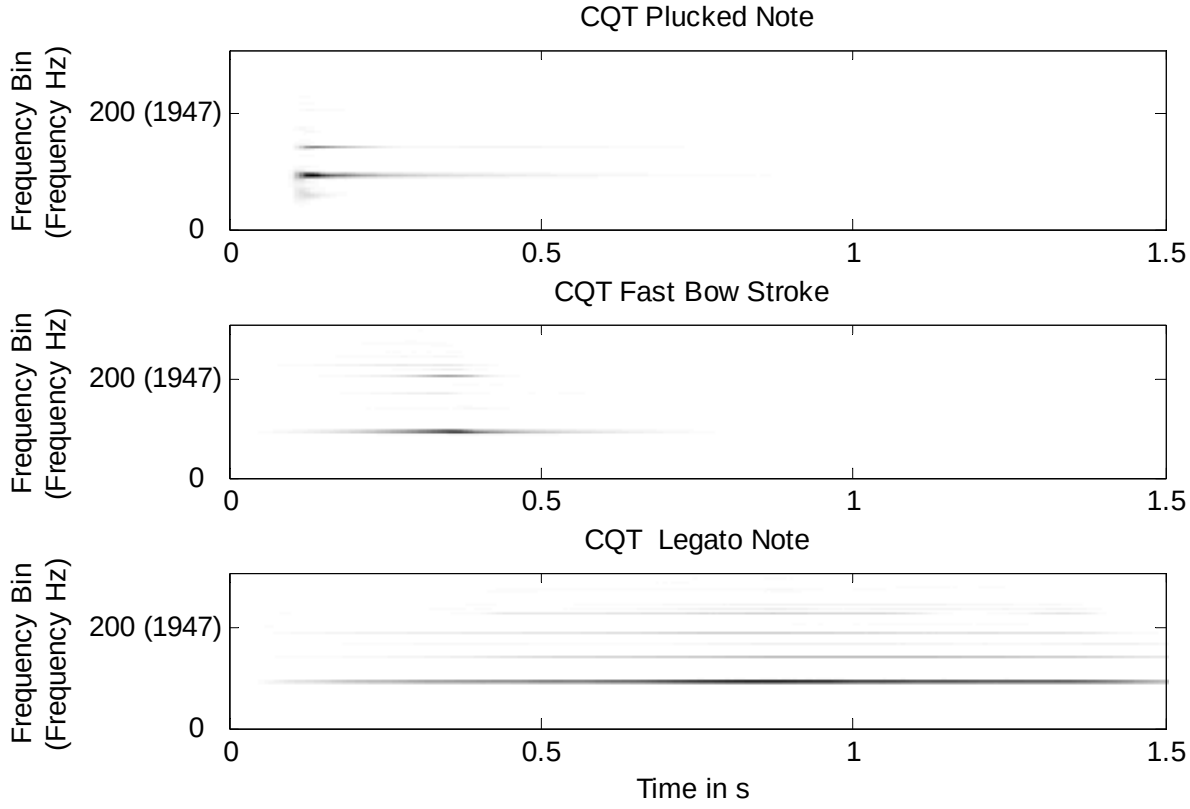

Figure 4.5: CQT representations of Figure 4.1 waveforms.

An example of the effect a beginner player’s onset crunching has on the waveform and CQT representations is illustrated in Figure 4.6. Crunching results in patchy harmonic evolution and unwanted frequencies, visible in the CQT representation up to about 0.45s in this figure. This beginner sample has been assigned, via the listening tests, an overall sound quality grade of 1.14 out of 6, where 6 is excellent. The qualitative fault terms associated with this sample are crunching, skating, nervousness, bad start and poor end. The comparative overall smoothness of a professional standard legato note waveform (top image) and its CQT representation (bottom image) are displayed in Figure 4.7. The harmonics evolve more evenly and consistently, reflecting a good onset and the note being established and maintained smoothly.

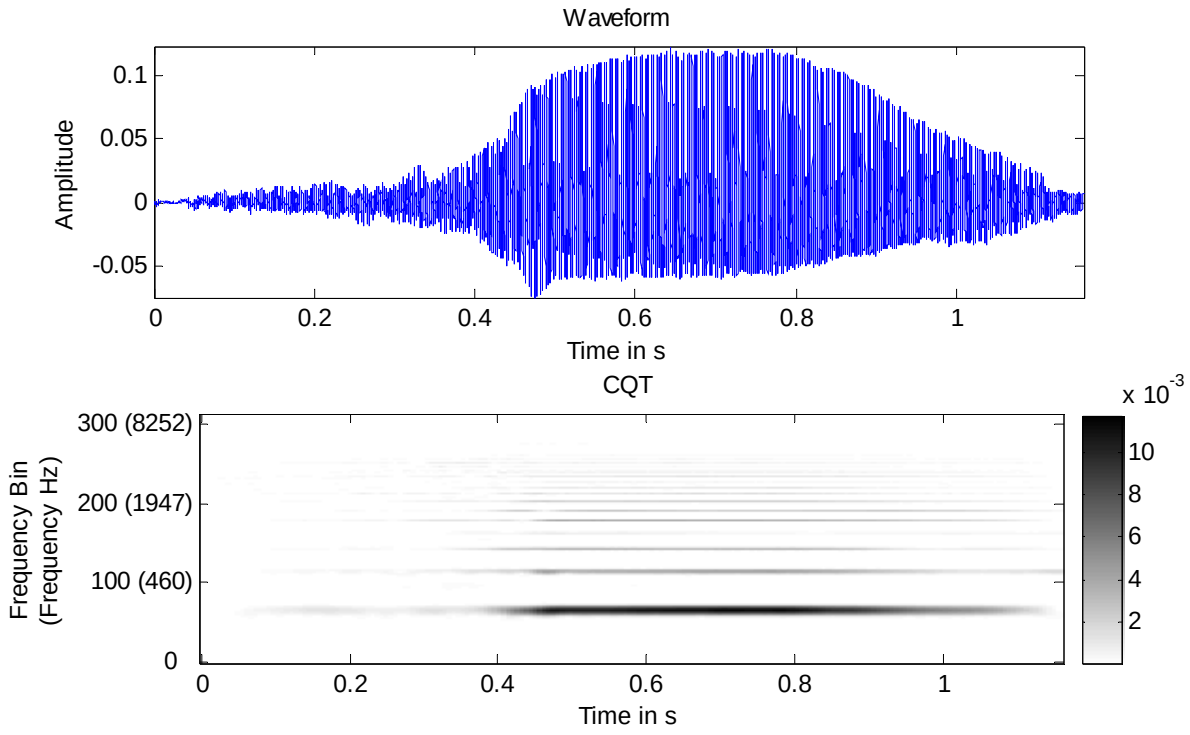

Figure 4.6: Beginner note waveform (top) and CQT representation (bottom) with crunching during onset.

To further contrast the difference in timbre between a professional standard legato note sample and that of a beginner, the spectra of two such samples are displayed in Figure 4.8. The same number of harmonics is present in both samples but the harmonics are much less developed in the beginner note sample's spectrum. From the listening tests, this beginner note sample contains crunching. Crunching causes the harmonics to spread out more and become less well defined as additional frequencies are present. The qualitative term crunching is associated with the presence of unwanted frequencies in the sound. Investigating the sonic properties of crunching further, the effect deliberate crunching or forcing has on the waveform, spectrum and CQT representations is illustrated in Figure 4.9. Much unwanted harmonic content is present and visible between the harmonic peaks in the sample's harmonic spectrum (middle image) and its waveform (top image) lacks smoothness. Rippling and inconsistency of harmonic development is visible in its CQT representation (bottom image), in particular during the onset period.

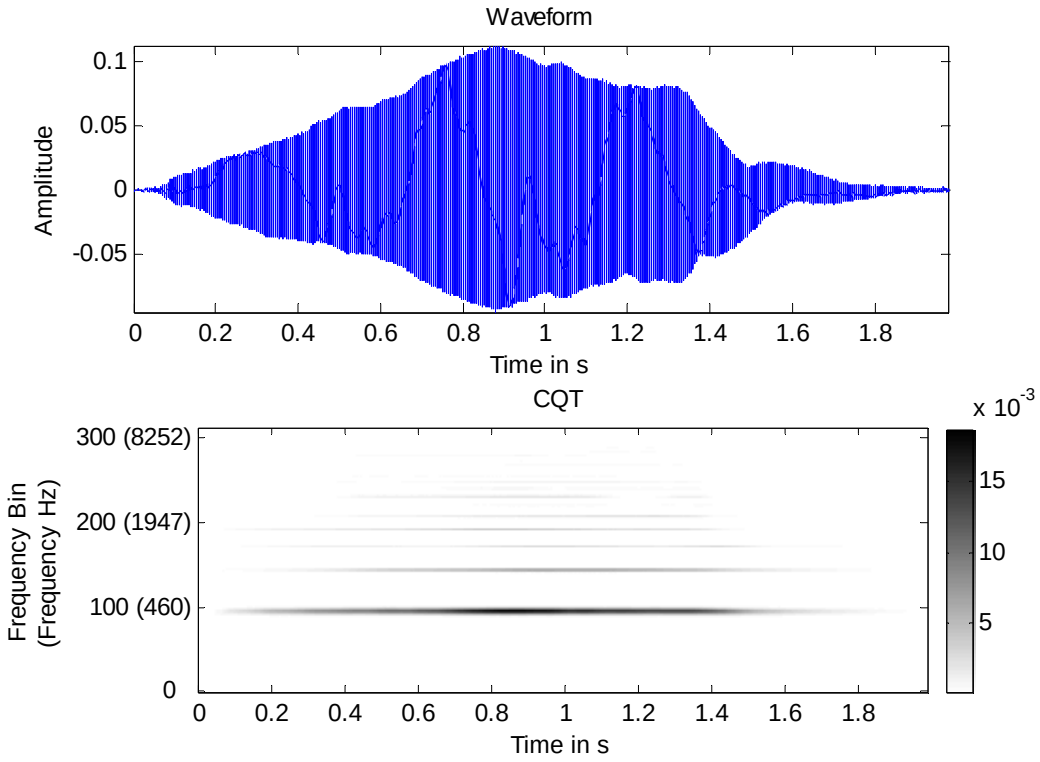

Figure 4.7: A professional standard legato note sample waveform (top) and its CQT representation (bottom).

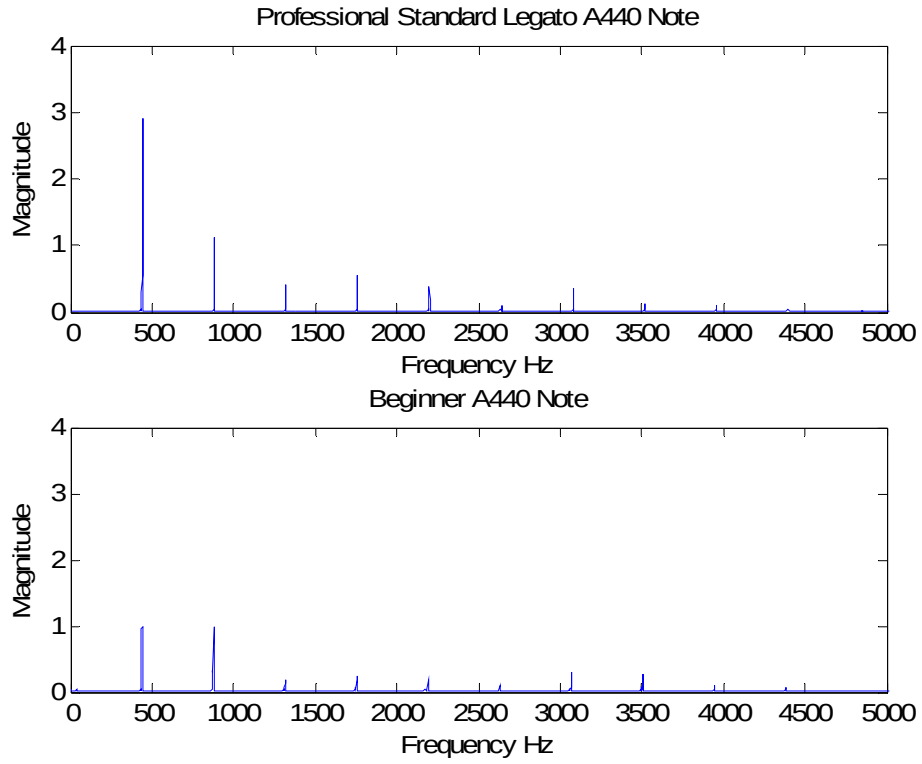

Figure 4.8: Spectra of a professional standard legato (top) and a beginner (bottom) A440 note samples.

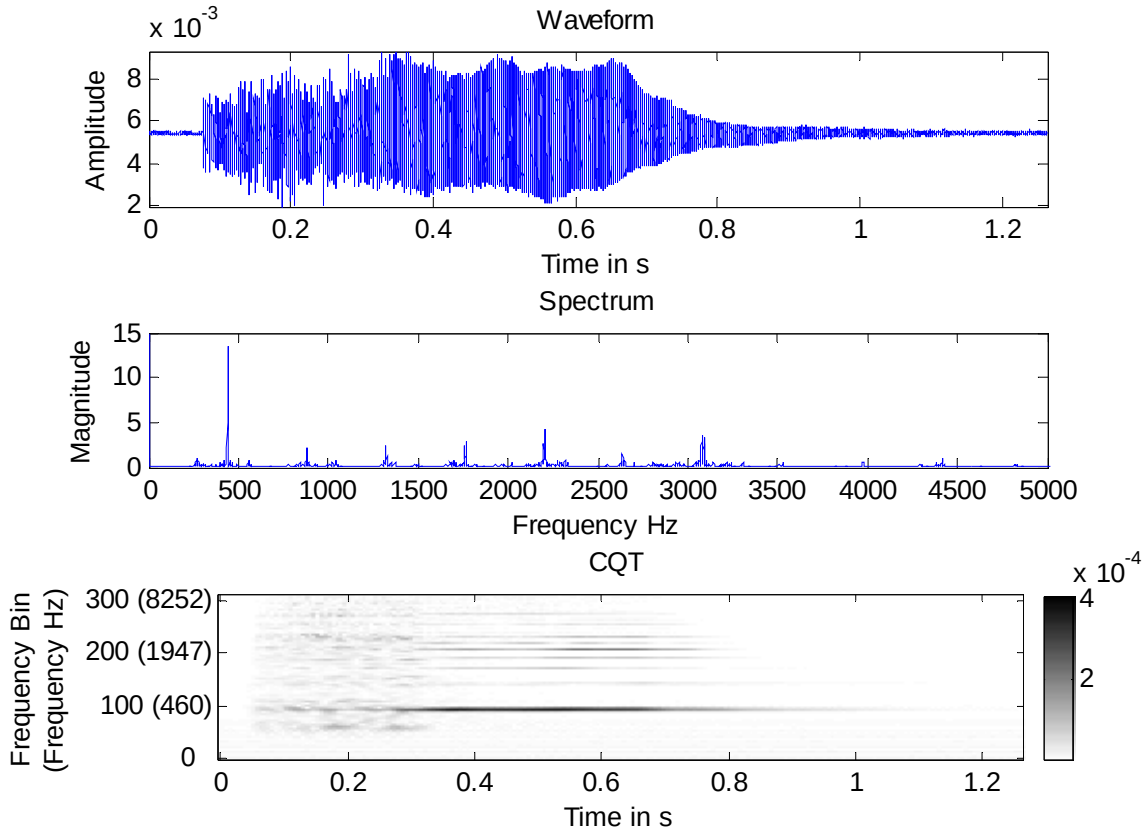

Figure 4.9: Effect of forcing on a note's waveform (top), harmonic structure (middle) and CQT representation (bottom).

Violin playing has many different onset styles which help create a note's harmonic spectrum and hence timbre, as has been illustrated in this section. The qualitative faults often associated with beginner note onsets include crunching and poor starts which are due to poor bow control which, in theory, can be detected based on the presence of unwanted frequencies which have been shown to be visible in the time-frequency representations. Offsets and the types of faults to which they are susceptible, are presented in the following section.

4.1.2 Offsets

Offset refers to the end of a musical note and the effect playing technique has on offsets is presented in this section. Onsets and offsets are similar in that both are susceptible to similar or equivalent types of qualitative faults. Just as for onsets, there are different offsets: full offsets and partial offsets. A full onset is where the note is allowed to resonate and fade away without any restriction or is ended deliberately by the player. This type of offset occurs at ends of phrases or before rests. A partial offset is when the

propulsion of a note played is continued into another note. This work focuses on full offsets only as work is being carried out on individual note samples.

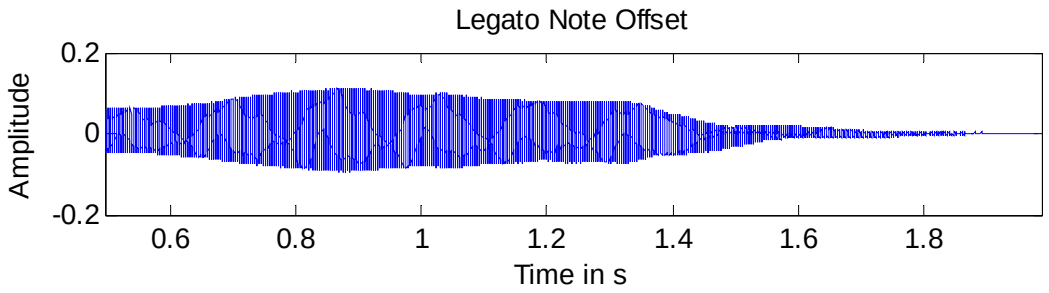

Figure 4.10: Legato note offset.

The waveforms of the offsets of a plucked note and a fast bow stroke are displayed in the top two images in Figure 4.1. The corresponding legato note offset is illustrated in Figure 4.10. The legato note sample has the most gradual offset.

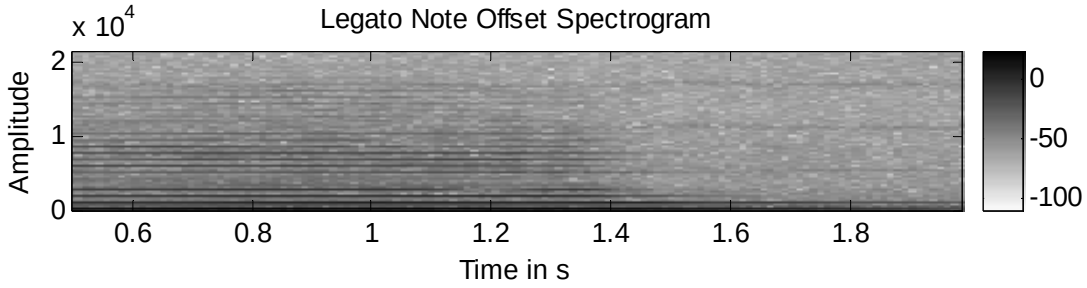

Figure 4.11: Spectrogram of legato note offset.

Bow speed and whether or not the note is let ring out effects the offset. The spectrograms and CQT representations of the plucked and fast bow sample offsets are shown in the top two images in Figure 4.4 and in Figure 4.5 respectively. The corresponding legato note offset spectrogram and CQT representations are displayed in Figure 4.11 and Figure 4.12.

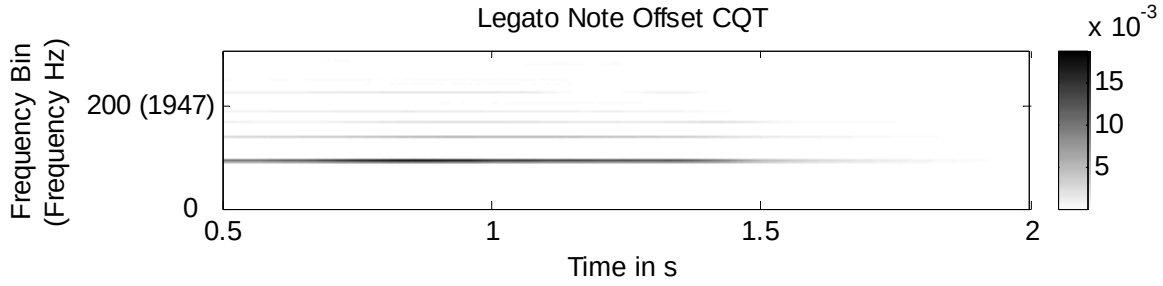

Figure 4.12: CQT representation of legato note offset.

The effect different offset styles have on their harmonic content is displayed in these time-frequency representations. The onset establishes a note's harmonic content and offset. In the plucked note and fast bow stroke samples, the string is released completely allowing the offset to fade out quickly and naturally. In the legato note sample, the bow is kept on the string and the note is let taper off.

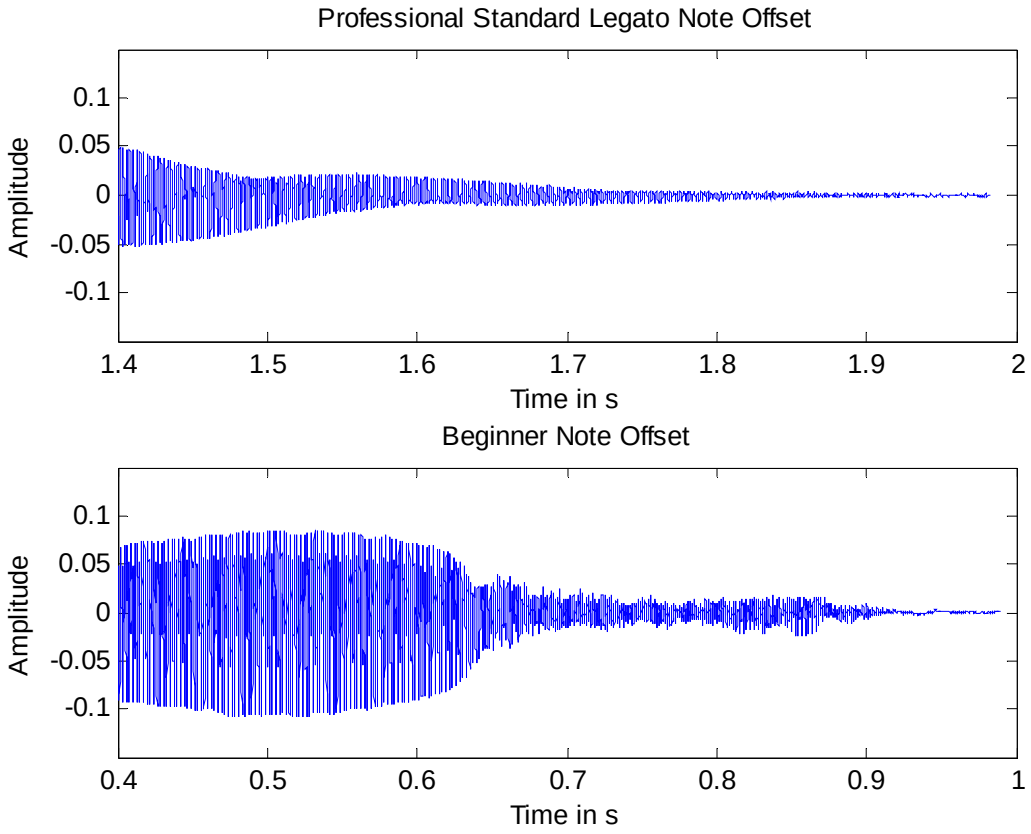

Figure 4.13: Professional standard legato (top) and beginner (bottom) note offsets.

Offsets, like onsets, cause bow control issues for beginner players, often resulting in both categories having very similar sound problems. Crunching and ending the note too quickly or poorly are qualitative fault descriptions which can be given to many beginner note offsets. An offset is dependent on the note's excitation as well as the extent of the bow's release of the string and is important in shaping a note's overall sound. How much a bow is released is determined by style and tempo. A fast tempo usually requires a full release, whereas a slower one, the note is tapered and the bow is kept on the string. Acceptable offset characteristics, regardless of style, require smoothness and the sound needs to die out at a reasonable speed, i.e. it should not be crunched or stopped suddenly as is apparent in some of the beginner note samples. One example of the

observable waveform differences between a professional standard legato and a beginner note offsets are displayed in Figure 4.13, where the waveform smoothness of the top image can be contrasted to that of the beginner note in the bottom image. This beginner sample has been associated with the qualitative expressions of crunching and a poor end via the listening tests and has an overall grade of 2.19 out of 6.

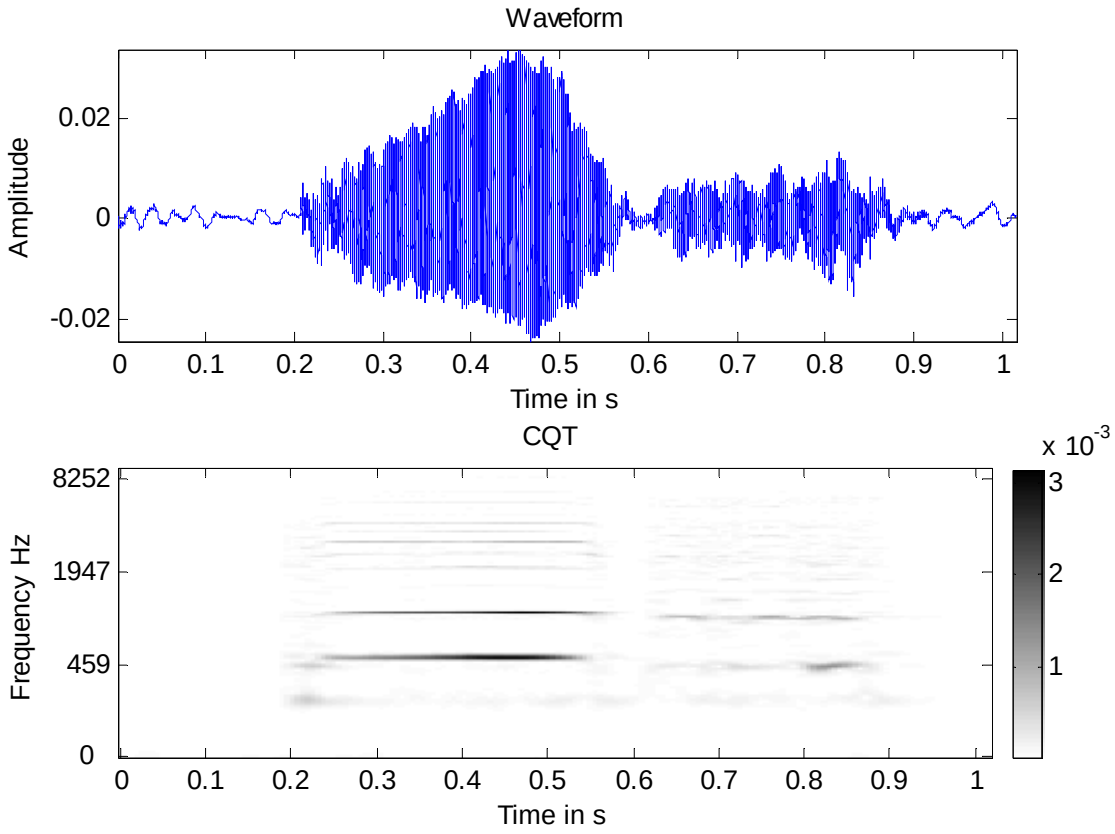

Figure 4.14: Waveform (top) and CQT representation (bottom) of beginner A440 note with crunching at start and end.

The following figure, Figure 4.14, illustrates the waveform and CQT representations of a beginner A440 sample which has been identified by the listeners as having crunching at the start and end of the note. The rippling effect is clearly visible in certain frequency bins reflecting the note's harmonic content in the CQT representation, at the start and more noticeably towards the end of the note. Additional, undesired frequency content, such as the two blotches below left of the first harmonic, which is the first dark line in the lower image in Figure 4.14, is present. The note played in this sample is C above A440, giving it a fundamental of approximately 515.75Hz which is bin 108 in the CQT representation. The two blotches are centred on frequency bins 62 and 98 which have centre frequencies of 62.29Hz and 452.89Hz respectively. These frequencies and

their neighbouring frequencies are not related the sample's fundamental frequency. Unwanted frequencies aside, the inconsistency of how the sample's harmonics are maintained is visible by the rippling effect along the frequency bins.

Playing faults such as crunching and poor end are often present in beginner note offsets and are similar to those occurring during the onset. They too are caused by poor bow control and stiffness in the bow arm. The next section details waveform amplitude and its associated qualitative faults.

4.1.3 Amplitude

Large and more sudden changes in amplitude are often observed in beginner note sample waveforms. Significant variation in amplitude levels is often visible in the waveforms of beginner note samples compared to those of professional standard legato ones as exemplified by Figure 4.15.

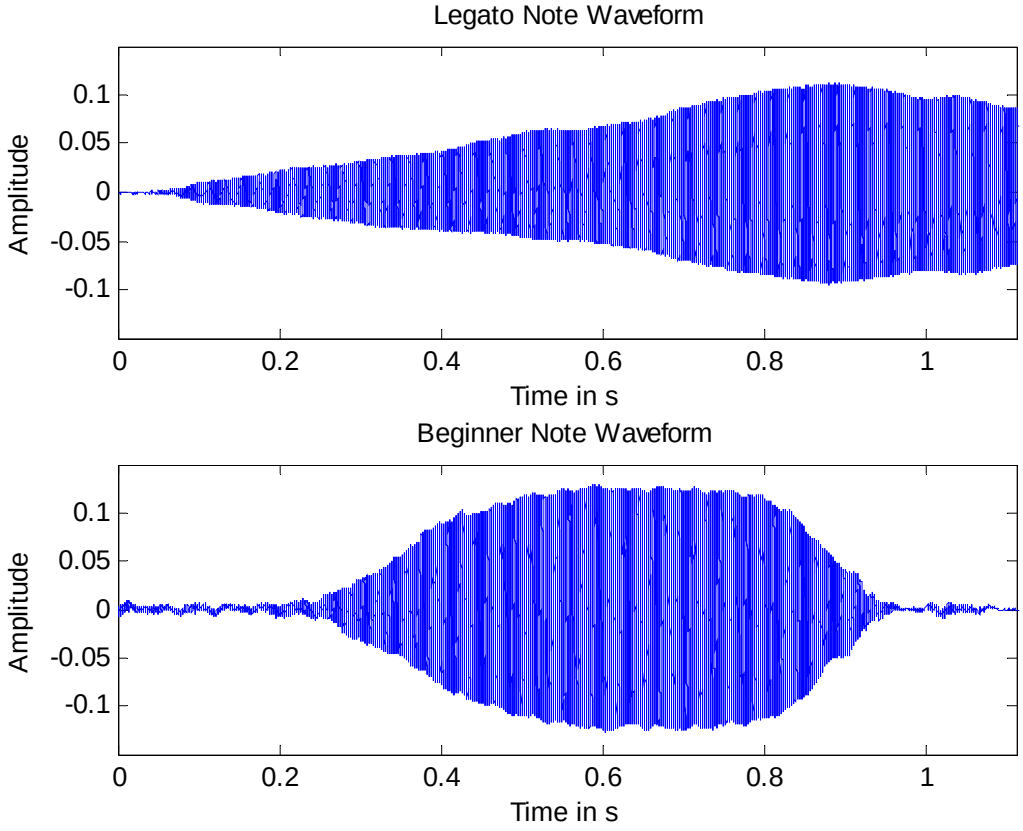

Figure 4.15: Waveforms of professional standard legato (top) and beginner (bottom) note samples.

Visually the most noticeable details about these two samples relate to amplitude and overall waveform smoothness. The professional standard legato note waveform

increases in amplitude relatively gradually, maintaining smoothness, whereas there is more noticeable “ripples” present along the beginner note’s waveform. The beginner note sample is reported to crunch at the beginning after which its waveform increases relatively suddenly in amplitude.

A beginner must learn the acceptable pressure range for drawing a bow along a string. Too much pressure gives rise to “crunching” and too little at the wrong angle results in a “whisping” or a “skating” effect in the sound. A bow hand that is too stiff results in the bow bouncing along the string, which occurs in the sample illustrated in Figure 4.16. Waveform amplitude is affected by bowing technique and typical qualitative faults reflected by changes in the waveform amplitude include crunching, skating, amplitude level changes and bow bouncing. Crunching and forcing result in spiky waveforms caused by sudden changes and jumps in waveform amplitude, which have an effect on a note’s harmonic content and evolution as illustrated in Figure 4.16.

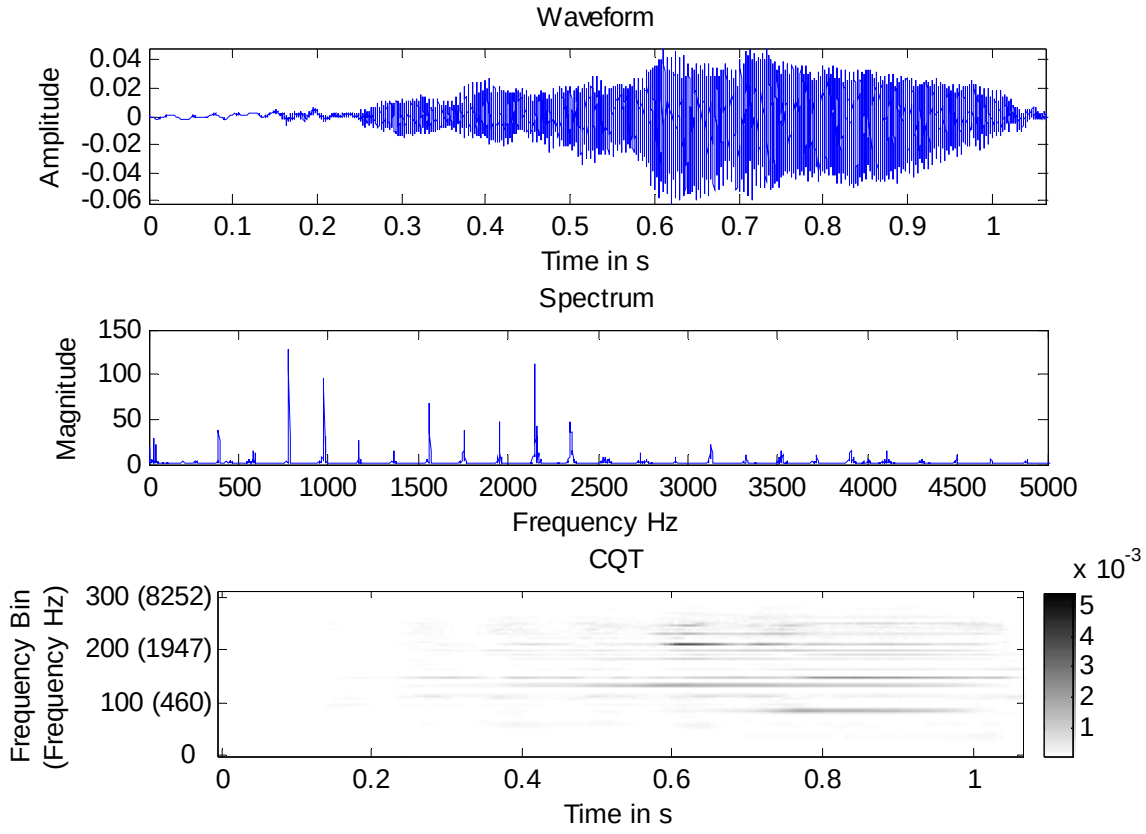

Figure 4.16: Beginner note sample with bow bounce.

All qualitative faults can be reflected to greater or lesser extent in the waveform amplitude, but it is difficult to link a specific fault with a certain characteristic as

playing faults rarely occur independently. In professional standard samples, certain characteristics which cause fluctuations in the amplitude are acceptable. The most common example is that of vibrato which is illustrated in Figure 4.22. Another example is tremolo whereby the player repeats very short bow strokes quickly, usually towards the tip of the bow as displayed in Figure 4.25. These are detailed in the section on acceptable waveforms, Section 4.1.6. Waveform unevenness is presented next.

4.1.4 Unevenness

Unevenness refers to the lack of smoothness in the waveform's shape. It differs from amplitude in that it focuses on the waveform variations in the time or abscissa direction, whereas amplitude refers to the changes in the ordinate direction. Both often occur in a same waveform. The top image in Figure 4.17 illustrates this effect and the lower image is that of a professional standard legato note for comparison, displaying relative waveform smoothness.

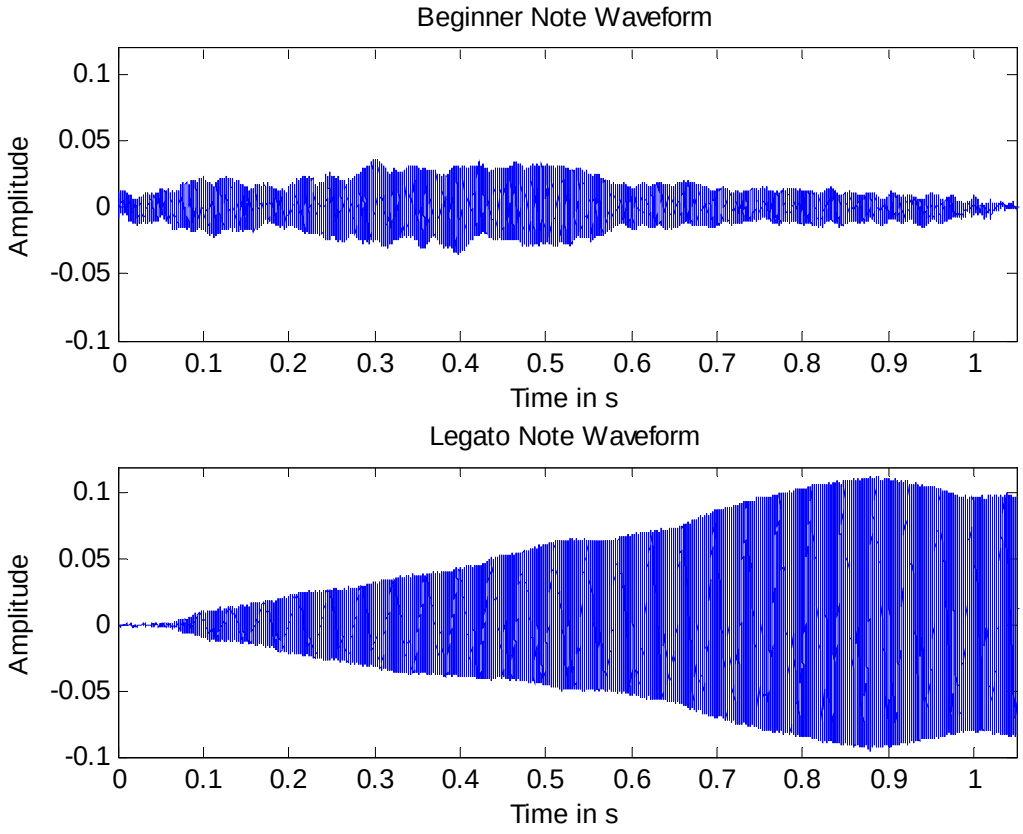

Figure 4.17: Beginner note waveform displaying waveform amplitude unevenness (top) contrasted with a professional standard legato note waveform (bottom).

The brightest, clearest sound is produced by pulling the bow across the section of the string, lined up with the f-holes, where a richer harmonic content is created. Should the bow go onto or close to the bridge, the sound squeaks. “In each and every stroke, the bow should move in a straight line running parallel with the bridge.” [Auer80:22]. A bow that is pulled across the string at an angle adds a “skating” or “whisping” effect to the sound which is often associated with uncertainty or nervousness. Informal listening tests suggest that unevenness is associated with the qualitative faults skating and nervousness.

The effect waveform unevenness has on its CQT representation and spectrum is displayed in Figure 4.18. Unevenness in violin playing impedes the harmonics from developing or evolving smoothly and consistently throughout a note, as illustrated in the CQT representation. In this sample, there is no clean start to the note as it takes some time before the note becomes established. In the spectrum, broader harmonics are visible. Unevenness is not the only playing fault which has additional frequencies around the harmonics present.

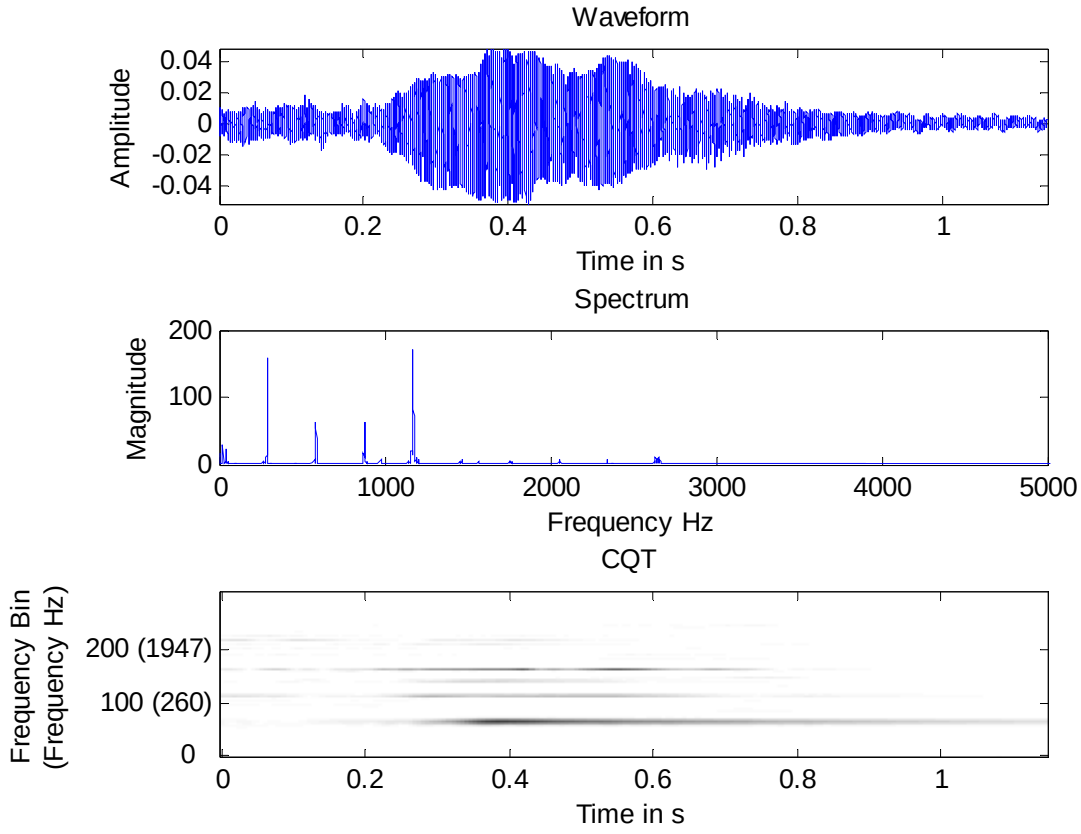

Figure 4.18: Waveform (top), CQT (middle) and spectrum (bottom) representations of a beginner D3 note sample displaying unevenness.

A committed sound has to do with exciting the right combinations of harmonics to the necessary levels to create a clear violin timbre. Figure 4.19 displays the harmonic spectrum of a committed legato (top image) to that of a beginner note sample reported to contain nervousness and skating (bottom image). Skating has a significant effect on the harmonic content, as can be seen in the lower image of Figure 4.19. From informal listening, the presence of skating in a sound is associated with reducing the magnitudes and clarity of the harmonics as the excitation of the string is not clean. This is due to the bow being drawn across the string at an angle, which brings in unwanted frequencies. Waveform asymmetry about the abscissa is presented next.

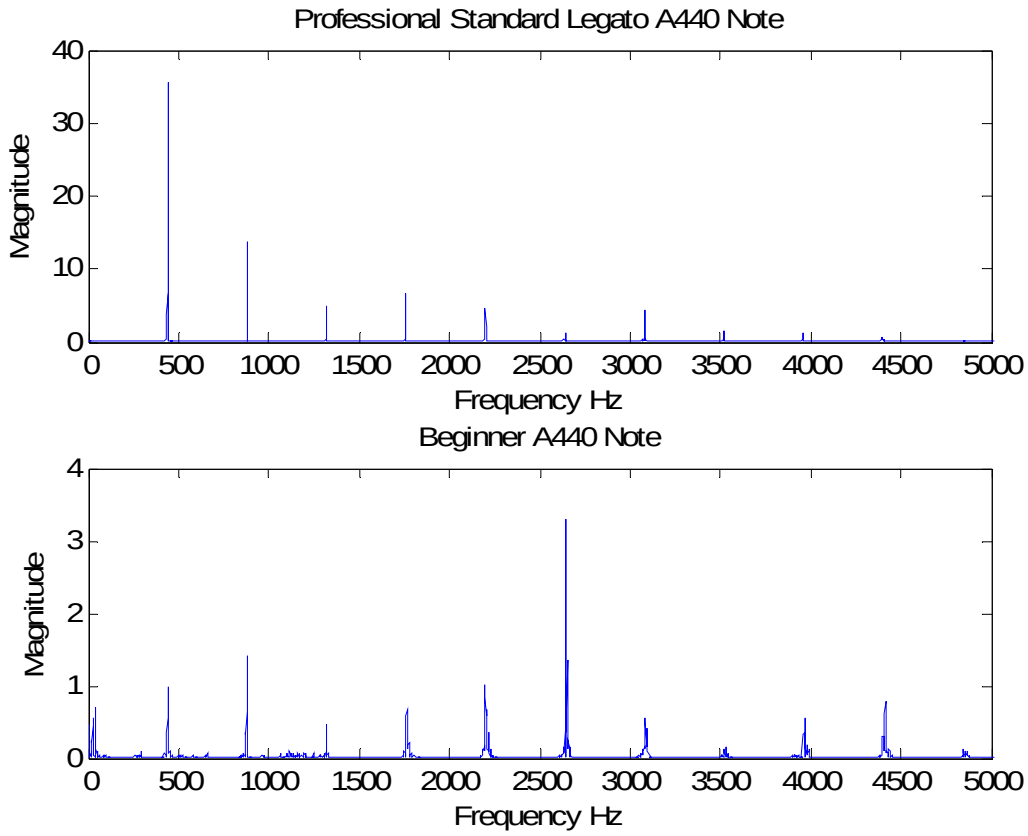

Figure 4.19: Spectra of legato professional standard note (top) and beginner note sample (bottom).

4.1.5 Asymmetry

Asymmetry refers to the unevenness about the abscissa and is best described by viewing effected samples. Most violin note samples' waveforms are effected to a certain extent

by asymmetry, but the most asymmetric ones in the dataset are from beginner note samples, such as the waveform of the sample displayed in Figure 4.20.

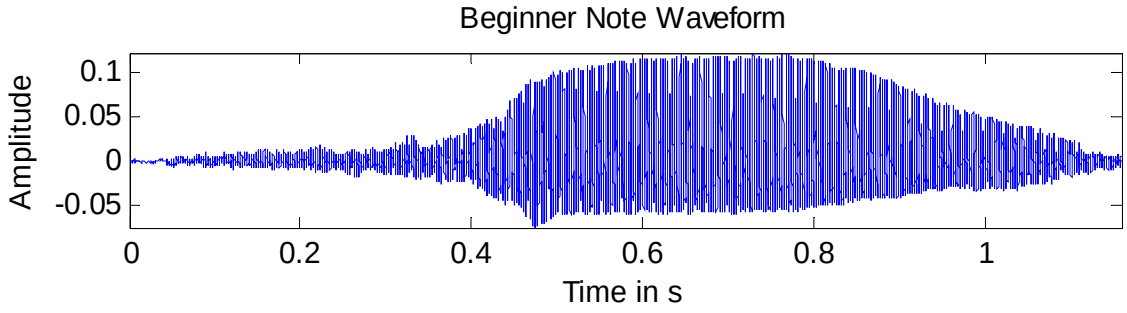

Figure 4.20: Beginner note waveform displaying asymmetry around the abscissa.

The link between asymmetry and sound quality is not clear. Multiple fault descriptions used tend to be associated with the most asymmetric waveforms, making linking these waveforms to a particular qualitative playing fault difficult. The waveforms of professional standard legato note samples can be asymmetric as displayed in previous figures, but not quite to the same extent as those belonging to some of the beginner sounds, such as the sample illustrated in Figure 4.20. Poor bowing technique has an effect on waveform symmetry. From the listening tests, the most asymmetric samples given as examples in this section all contain multiple faults. The qualitative terms which have been associated with these beginner samples include skating, nervousness, and sudden end to note.

Waveform asymmetry is reflected in the note's timbre, visible in the CQT representation and in its spectrum, displayed in Figure 4.21. Inconsistent, blotchy harmonic development is visible in the CQT representation and the spectrum has wider harmonic peaks. Waveform asymmetry is not unique to having these visible effects.

Five fault categories have been presented with their associated qualitative playing faults suggested by informal listening. No one qualitative expression can be consistently linked to a specific characteristic. The following section focuses on deliberate playing effects which cause acceptable changes to a sample's waveform and harmonic content.

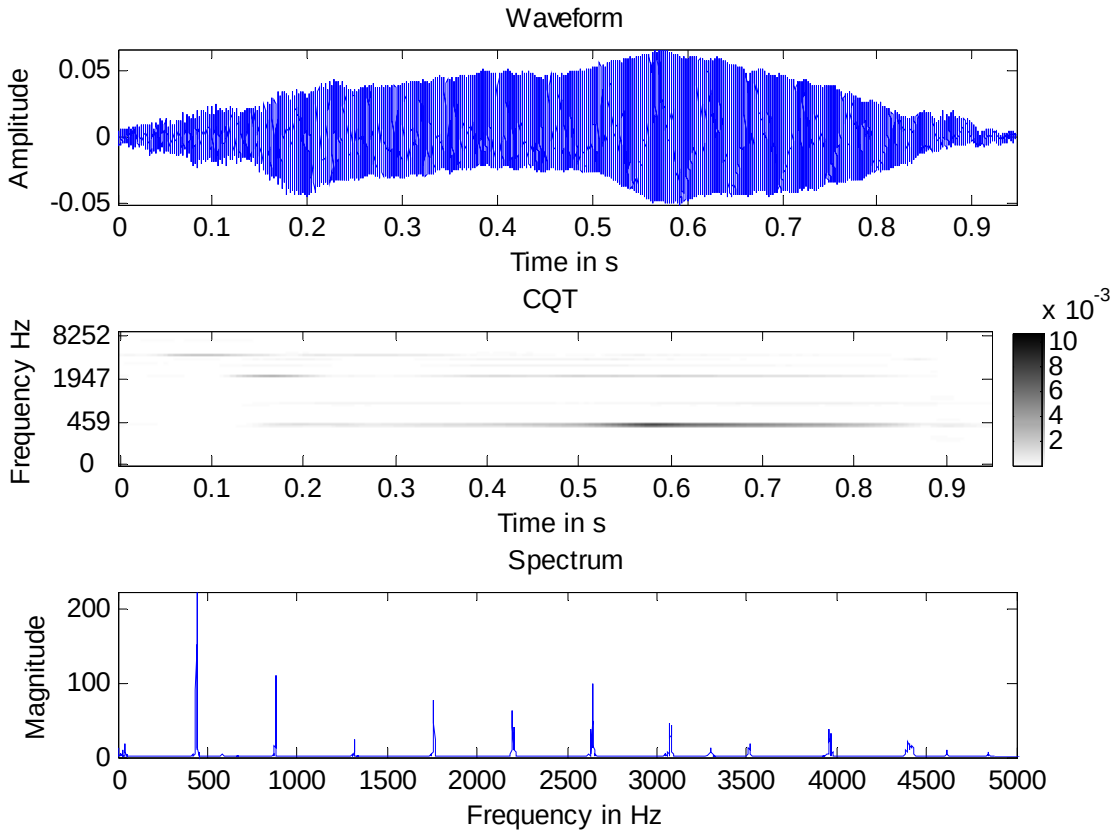

Figure 4.21: Waveform (top), CQT (middle) and spectrum (bottom) representations of a beginner A440 note sample displaying unevenness.

4.1.6 Acceptable Waveforms

In the previous sections, waveform categories have been tentatively associated with playing faults. Limitations and exceptions to these observations are considered in this section. Several sound colouring effects such as vibrato, tremolo as well as playing techniques, i.e. bow changes and note changes, are picked up in the waveform and can be quite similar to those of some of the undesired sounds. Bow and note changes have already been presented in relation to note onsets. This section describes the effects of vibrato and tremolo.

Vibrato refers to the embellishment of a note by adding tastefully what is effectively a small, local frequency modulation. Style influences vibrato, giving the player use of a range of vibratos. These include a very narrow, fast vibrato to a slower, wider version. From the technical side, a finger, hand, and full arm vibrato and combinations of these exist. The waveform of a sample with a standard, general use vibrato, which is a combination of finger, hand and arm vibratos, is illustrated in Figure 4.22 and displays amplitude modulation.

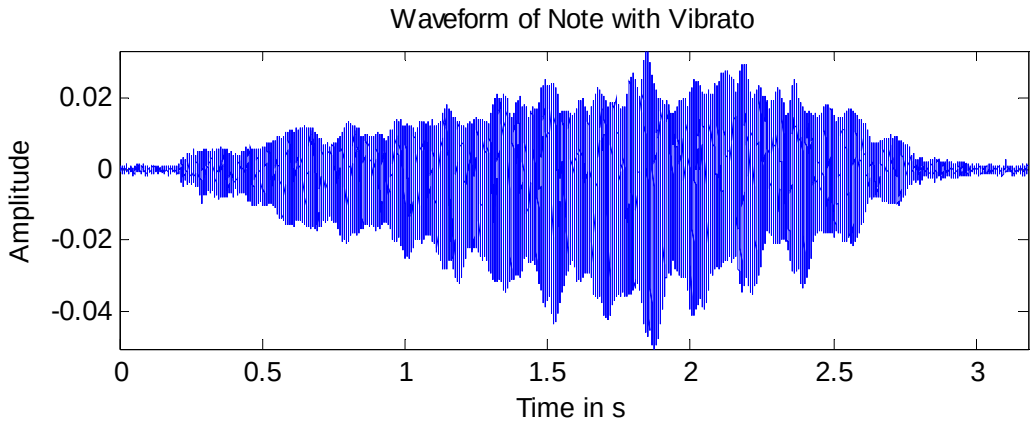

Figure 4.22: The effect of vibrato on a note's waveform.

Vibrato usually changes throughout the duration of the note, helping to bring out a specific note's place in a phrase by changing the colour or intensity. The first thorough study on vibrato was conducted by Seashore[1] [Winckel67]. Although some work has been completed on instrumental vibrato [d'Allessandro94, Brown96] and on violin vibrato [Mellody00], much more has been done relating to vibrato of a singing voice [Prame94, Prame97, Bonada03].

The waveform and spectrogram illustrating the effect vibrato has on these representations is displayed Figure 4.23. The note played is D4, third finger on the A string. To the left in the figure, the waveform of the note with a gentle vibrato is given and to the right, the waveform of the same note without vibrato. The approximately even fluctuations due to vibrato are visible on the left hand side of this figure. An acceptable vibrato on the violin has about a 5 to 10Hz rate [Dodge97]. This figure shows that vibrato on the violin has both an effect on frequency and amplitude. This sample was not taken in the recording studio the effect of vibrato is well displayed in this sample.

The effect of vibrato on the sample's spectrum is illustrated in Figure 4.24. As expected, the harmonic peaks in the spectrum of the note with vibrato are broader and not as well defined. Vibrato involves bringing in neighbouring frequencies to colour the sound thereby giving slightly wider harmonic peaks in its spectrum.

[1] Carl Emil Seashore (1866-1949): psychologist; Seashore Tests of Musical Ability.

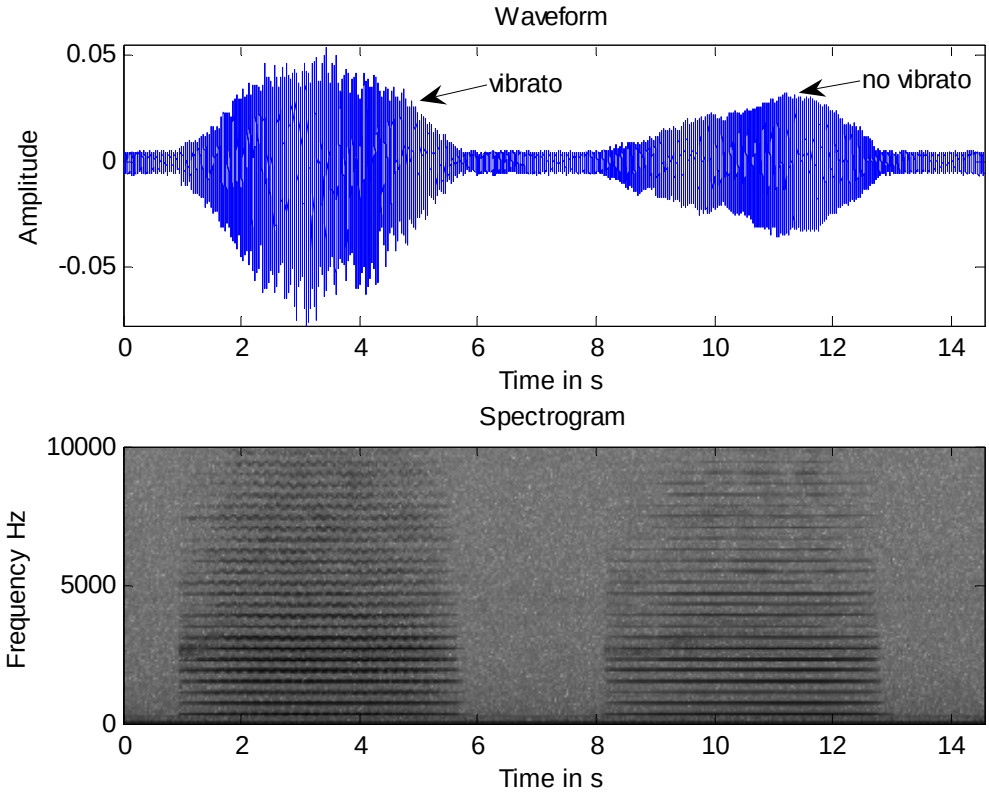

Figure 4.23: Effect of vibrato on a note's waveform (top) and spectrogram (bottom).

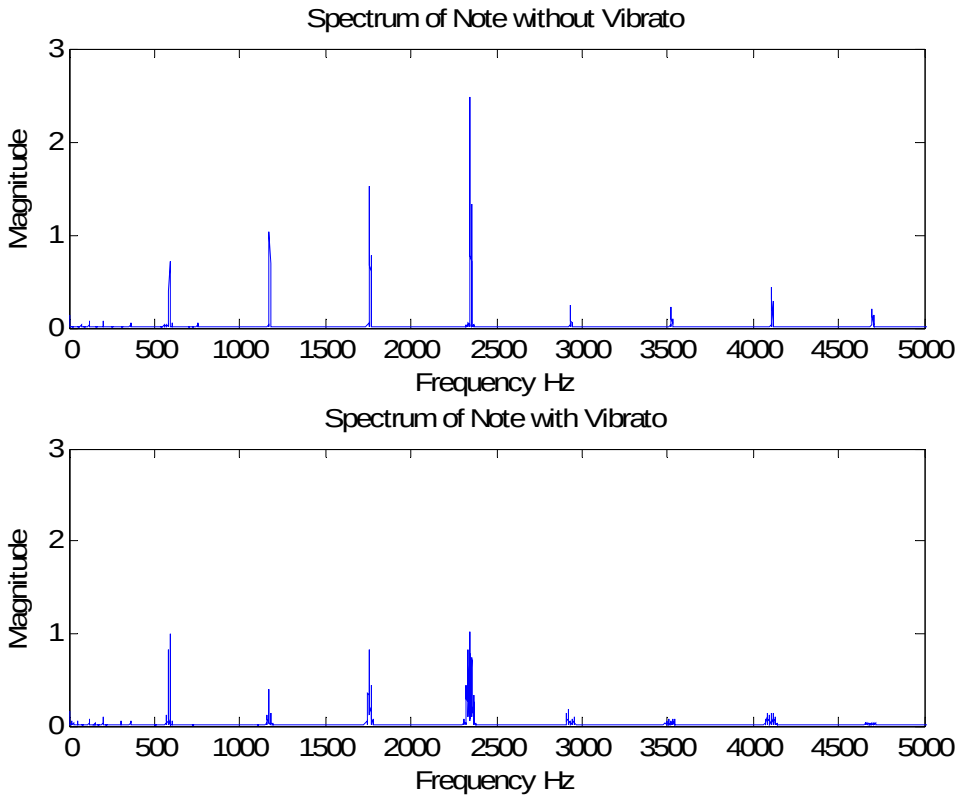

Figure 4.24: Spectra of a note without vibrato (top) and with vibrato (bottom).

Another effect is tremolo which “consists of very small, unaccented détaché bow strokes, usually performed near the bow-tip” [Jackson88:51]. The effect tremolo has on the waveform, spectrogram and CQT representations is depicted in Figure 4.25. Although the pitch remains constant for the duration of the sample, the quick bow changes add extra frequencies to the sound giving a shimmering effect which is reflected in the spectrogram and CQT representations. The presence of the additional frequencies is visible in the tremolo sample’s spectrum, displayed in Figure 4.26 where the harmonic are much wider.

This section on acceptable waveforms has been included to illustrate the difficulties and limits associated with determining good violin sound and playing faults. An awareness of the differences and proximities between playing faults and deliberate playing techniques is important.

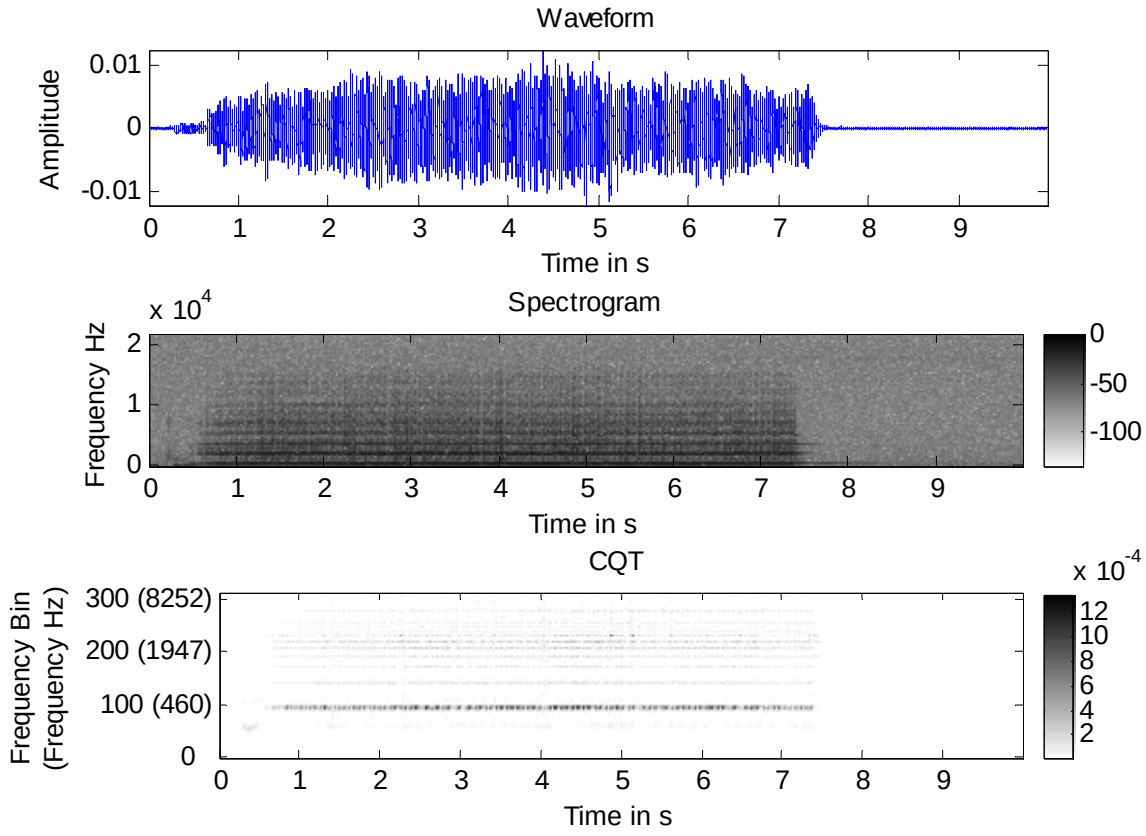

Figure 4.25: Waveform (top), spectrogram (middle) and CQT (bottom) representations of a tremolo sample.

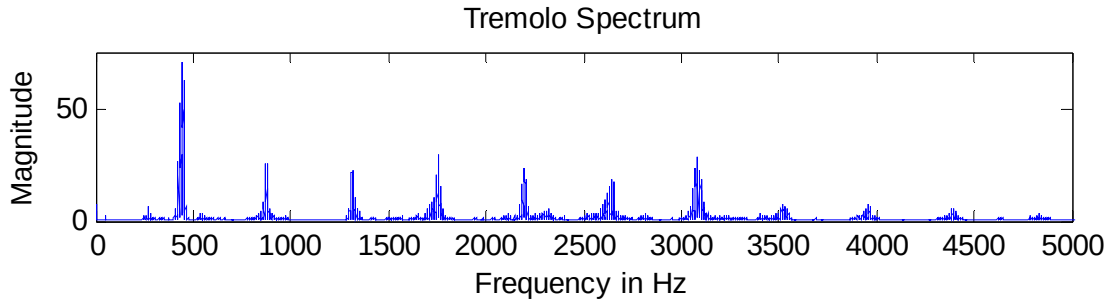

Figure 4.26: Spectrum of a tremolo sample.

4.2 Summary

Sound characteristics have been loosely linked to playing technique disturbances which result from poor bow control in this chapter. Correct and efficient bowing technique ensures a certain smoothness which is reflected in the waveform of the legato notes and which can be perceived in the sound. Acceptable waveform disturbances due to deliberate or desirable effects, such as changing bowing direction, vibrato and tremolo have been presented and should not be mixed up with playing faults. Bow strokes depend on the speed, pressure and the amount of hair in contact with the string and are reflected in the time, frequency and time-frequency representations. A beginner player's developing bowing technique is associated with faults, many of which are reflected in the waveform representation. Five main waveform categories have been illustrated and explained with the typical qualitative fault descriptions that are often associated with a beginner player. This chapter has outlined some of the boundaries required so that suitable feature detection can ensue. The following chapters present features which are used to represent violin sound in the time, spectral and cepstral domains.

5 Temporal Features

In Chapter 4, visual observations of violin note waveform characteristics typical of the dataset's samples have been presented and various sound characteristics have been tentatively linked to specific playing disturbances which result from poor bow control. Violin note waveforms contain much variability but less variability is observed in the waveforms of the professional standard legato notes. Comparatively, the beginner note waveforms tend to display more asymmetry and unevenness. These visible waveform characteristics in the beginner notes prompted a statistical approach to obtain possible suitable features for violin sound analysis. Statistics provide a way of getting a collection of quantitative features from which possible violin sound descriptors are sought. Four moments of first order statistics, mean, variance, skew and kurtosis are obtained and their usefulness within the context of comparing beginner to professional standard player notes is presented. Signal periodicity is presented as described by the autocorrelation coefficient. The analysis has been completed on the dataset's samples as well as on the forced note samples. The forced samples are notes where a professional player has emulated a beginner's crunching and has forced the sound even further. This chapter considers how and if any of qualitative expressions used in this text can be reflected or captured through statistical analysis.

5.1 First Moment: Mean

The mean is the arithmetic average of all values shown in Equation 5.1 [Stuart87] where N is the data length and i, the current sample number:

$$\overline{data} = \frac{1}{N}\sum_{i=1}^{N} data(i) \tag{5.1}$$

The waveform amplitude mean readings of the samples are displayed in Figure 5.1, where the professional standard legato note samples are shown in blue and the beginner ones are given in red. Figure 5.1 shows a very large, consistent gap between the mean of a beginner note and that of a professional standard player legato note. The scatter plot of the same results can be seen in Figure 5.2, where the professional standard legato note samples are much more tightly grouped together than the beginner ones.

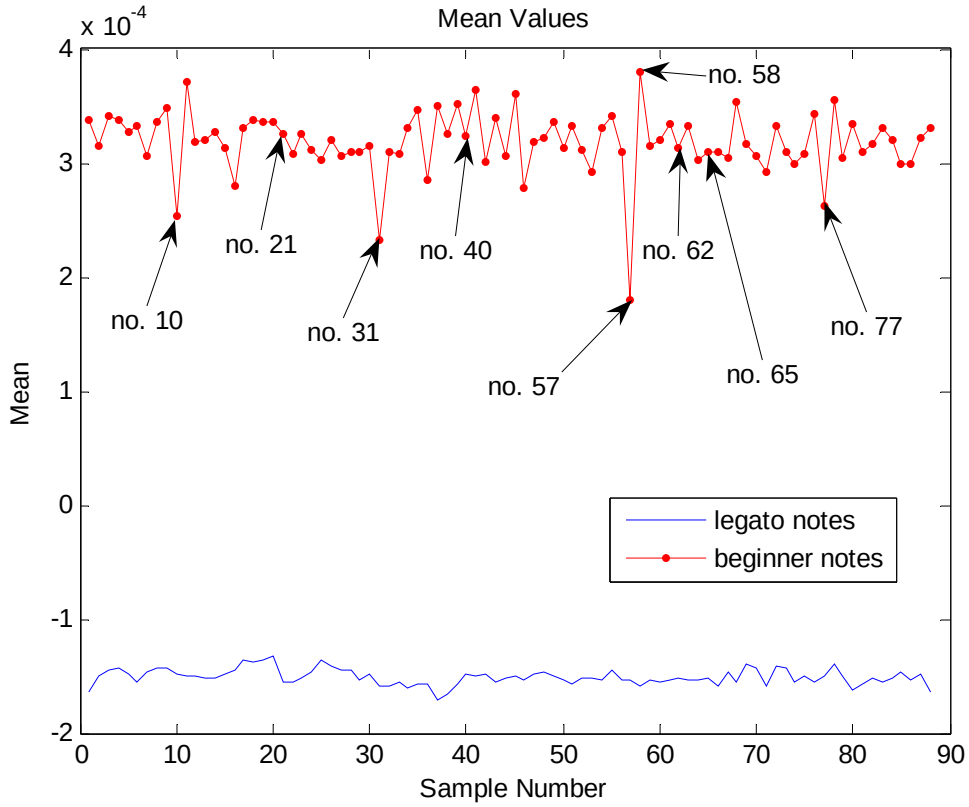

Figure 5.1: Waveform amplitude mean values of professional standard legato and beginner player note samples.

Sample No.	*Grade*	*Mean*	*Perceived faults?*	*Comments*
Beginner 10	2.67	$2.53*10^{-4}$	NV, BB, BADS	Low mean beginner sample
Beginner 21	**3.33**	**$3.26*10^{-4}$**	**None**	**No playing faults perceived**
Beginner 31	1.43	$2.32*10^{-4}$	CR, NV, BB, BADS, BADE	Low mean beginner sample
Beginner 40	**2.29**	**$3.24*10^{-4}$**	**None**	**No playing faults perceived**
Beginner 57	2.52	$1.81*10^{-4}$	NV	Lowest mean beginner sample
Beginner 58	2.86	$3.80*10^{-4}$	BADE	Highest mean beginner sample
Beginner 62	**3.95**	**$3.14*10^{-4}$**	**None**	**Best sounding beginner sample**
Beginner 65	**3.86**	**$3.10*10^{-4}$**	**BADE**	**2nd best sounding beginner sample**
Beginner 77	2.24	$2.63*10^{-4}$	CR, NV, INT, SE	Low mean beginner sample

Table 5.1: Waveform amplitude mean value sample information.

Based on this information, it is possible for a computer to detect a beginner from a professional standard legato note sample in this dataset. The divide between the sample groups is such that a classifier in this case is not required, a simple threshold value suffices. Another observation drawn from these results relates to the grouping patterns observed. The professional standard legato note samples are all grouped relatively tightly together while the beginner note sample readings are more varied. From visual inspection, the professional standard legato note sample waveforms are more consistent, smoother and have lower mean readings. As a reference, the mean professional standard legato sample mean is $-1.51*10^{-4}$ and the mean beginner sample mean is $3.18*10^{-4}$. Waveform asymmetry is reflected by these readings as the beginner note samples are

consistently averaging positive mean values that are further away from zero than the professional standard legato ones. To verify that the difference between the beginner and professional standard note samples, as represented by their waveform amplitude mean values, is statistically significant and not caused by sampling variability [Herzberg83], a t-test was run with a 0.01 significance level. The null hypothesis is rejected and a p-value of $6.6*10^{-119}$ is returned. The next step involves finding which qualitative feature(s) are reflected by the mean reading.

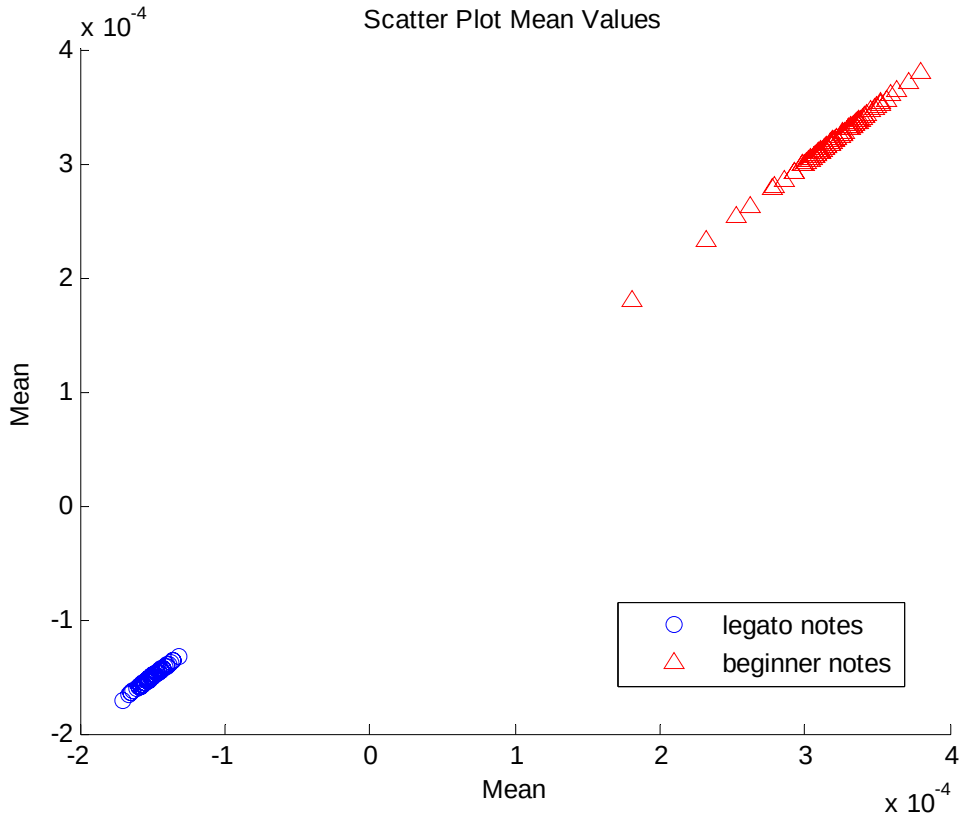

Figure 5.2: Scatter plot of waveform amplitude mean values.

Information relating to samples marked in Figure 5.1 is detailed in Table 5.1. From these results, high mean readings are associated with beginner sounds, but not all beginner sounds contain the same faults and some have even been perceived by the listeners to be of reasonable quality. From the listening tests, three beginner samples have not been associated with any qualitative playing faults. They are beginner samples 21, 40 and 62. Although they have been perceived as being faultless, this has not been reflected by a reduction in their mean values, nor in them having the best overall quality grades for beginner note samples. From the listening tests, the best sounding beginner samples are 62 and 65. The beginner note with the lowest mean value is from sample 57

which has an average sound quality score of 2.52 out of 6 and is reported to have nervousness in the sound. Samples 10, 31 and 77, which are the three next lowest beginner samples shown in Figure 5.1, also have low overall sound quality grades and contain multiple faults.

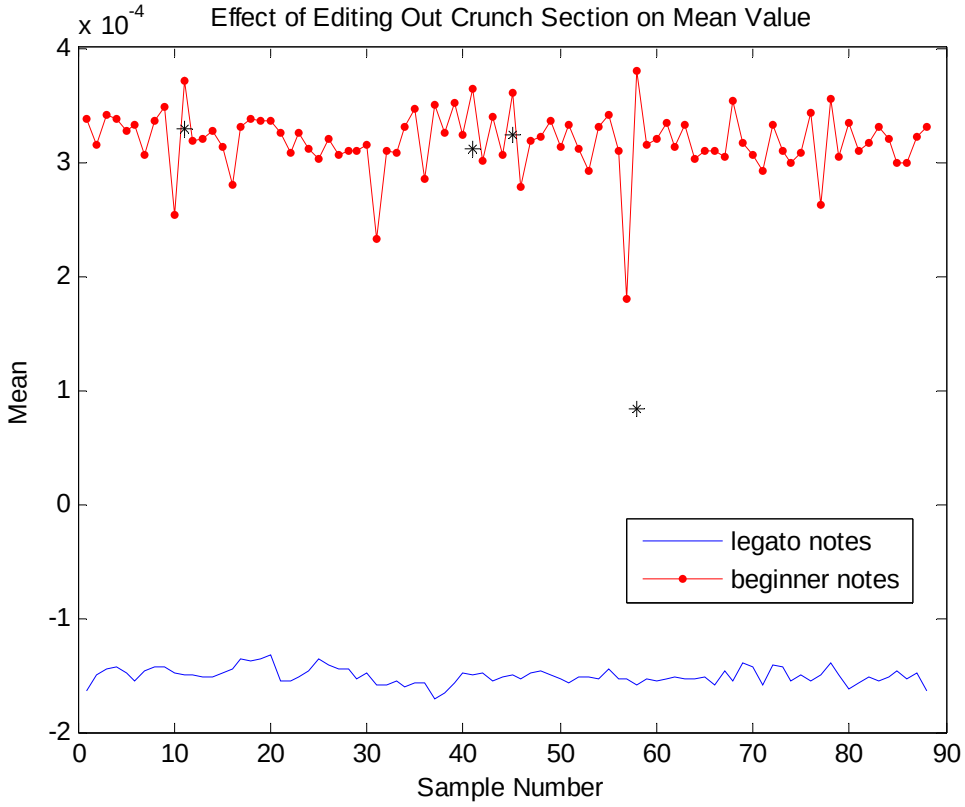

Figure 5.3: Effect of removing crunch sections from beginner samples on waveform amplitude mean value.

Although the waveform amplitude mean value differentiates effectively between the two different player types in the dataset, the overall sound quality perceived by the listeners is not directly reflected by this measure. As a means of further investigating this lack of relationship, beginner player crunching and professional standard player deliberate crunching and forcing are considered. Crunching is being investigated specifically as it is the one fault that professional standard players can emulate easily. It also tends to occur mostly at the starts and ends of notes and as a result, can often be easily edited out. The other typical beginner faults cannot be emulated by professional violinists as they have become too well conditioned. To better understand the relationship between crunching and the waveform amplitude mean value, the crunch sections of four samples which have been confirmed through the listening tests to

contain crunching, have been removed. The beginner samples, 11, 41, 45 and 58, have crunching at the starts and ends of their notes which have been edited out. The waveforms' means were then recalculated and compared to the original values. These recalculated mean values are marked in Figure 5.3 by the black asterisks. Trimming off the crunch sections at either end of a note reduces the mean value.

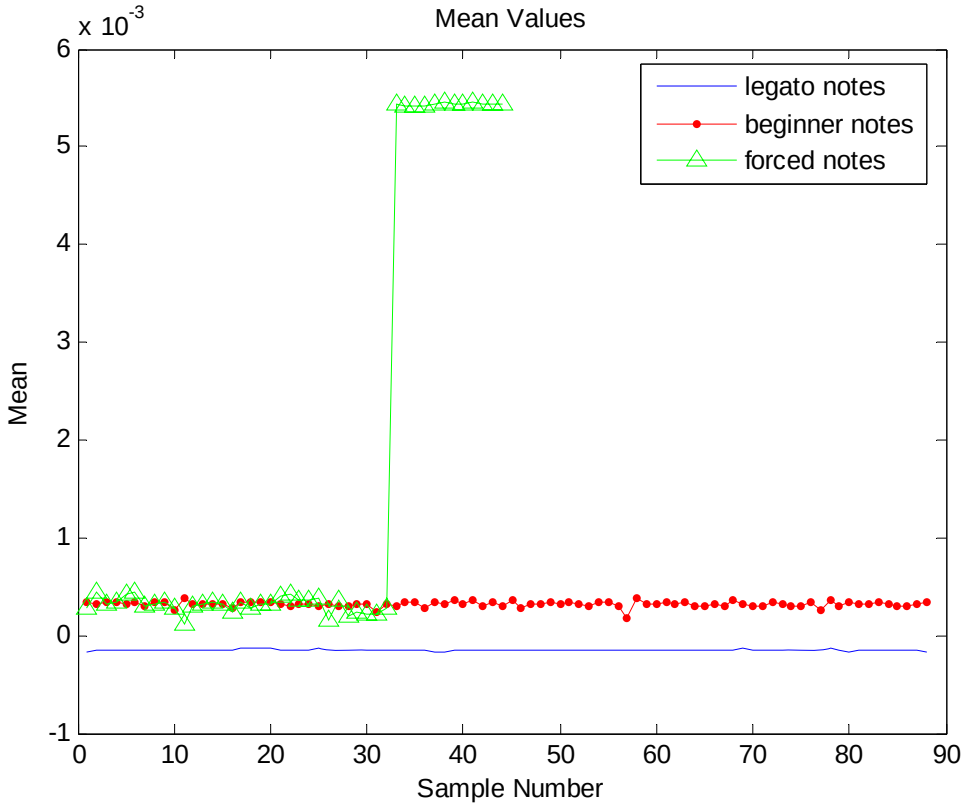

Figure 5.4: Effect of forced crunching on waveform amplitude mean values.

The effect of forced crunching on mean values is plotted in Figure 5.4 where the forced notes fall into two groups. Samples one to 32, which is the last point before the triangulated line jumps up to a much higher value, are recordings of a professional standard player crunching deliberately at the start and ends of the notes only. These waveform amplitude mean values are similar to those returned by the beginner note samples. This supports the link between bow crunching, mean reading and sound quality with the results matching those of the beginner note samples. From sample 33 onwards, the forced notes samples contain notes that are forcefully crunched for the duration of the note, which significantly raises the mean value.

So far, a high waveform amplitude mean reading is associated with emulated crunching and beginner sound. To see if this can be extended to include any other

standard bow strokes, the relationship between bow stroke style and mean reading is presented next. The type of bow stroke or articulation is of great importance as it includes information about bow pressure and speed as it initiates the note. The effects of different bow strokes on the waveform's amplitude mean are illustrated in Figure 5.5.

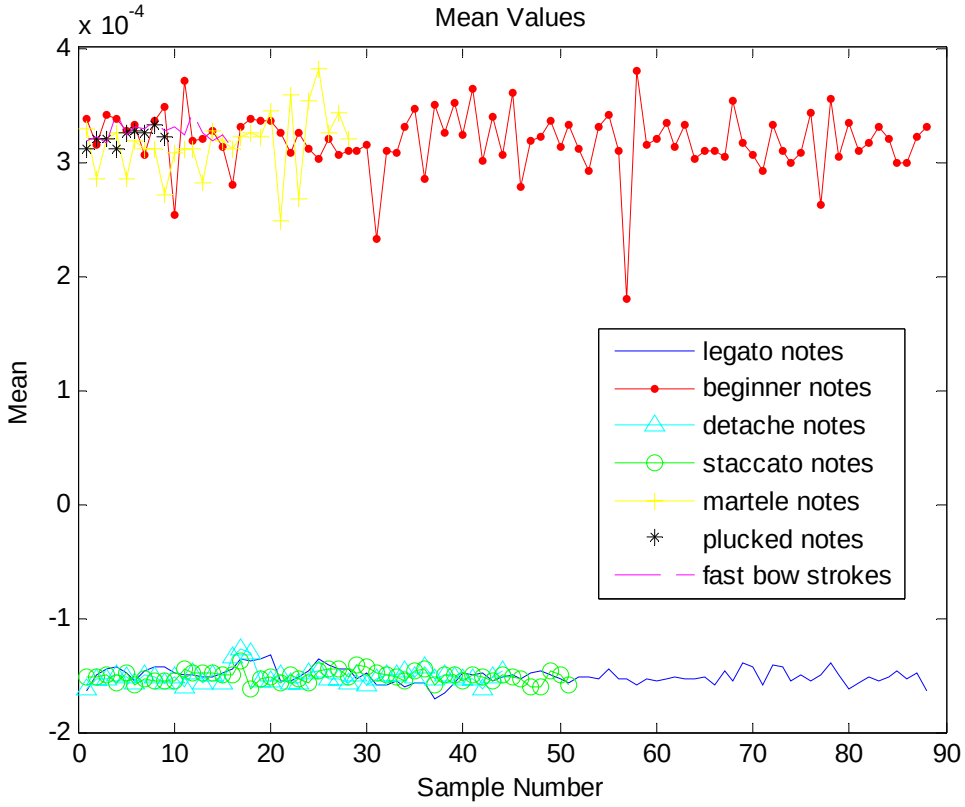

Figure 5.5: Effect of different bow stroke styles on waveform amplitude mean value.

In Figure 5.5, legato, staccato and detaché bow strokes all have much lower mean values. Plucked notes, martelé and fast bow strokes all have waveform amplitude means within the same region as those belonging to the beginner note samples. A plucked string is the most sudden and percussive of the attacks on the violin. Martelé is a "sharply accented staccato bowing" [Jackson88:28], having a strong prepared attack and a much faster bow stroke than legato. In these recordings, the staccato notes are produced by short bow strokes in the lower half of the bow. The fast bow strokes also have an accented, sudden start. What plucked notes, martelé and fast bow strokes have in common is the strength of their accented attacks. This provides an explanation for the high waveform amplitude mean readings for these samples. A high mean can be associated with a strong attack which suggests that in the beginner note samples, the players are pushing too hard. Further detailed work is required to support the effect of

bow stroke on mean reading focusing specifically on attack styles and bow pressure. As this work is concerned with determining professional standard legato from beginner note samples and fault detection, more detailed work contrasting various types of bow strokes will not be completed.

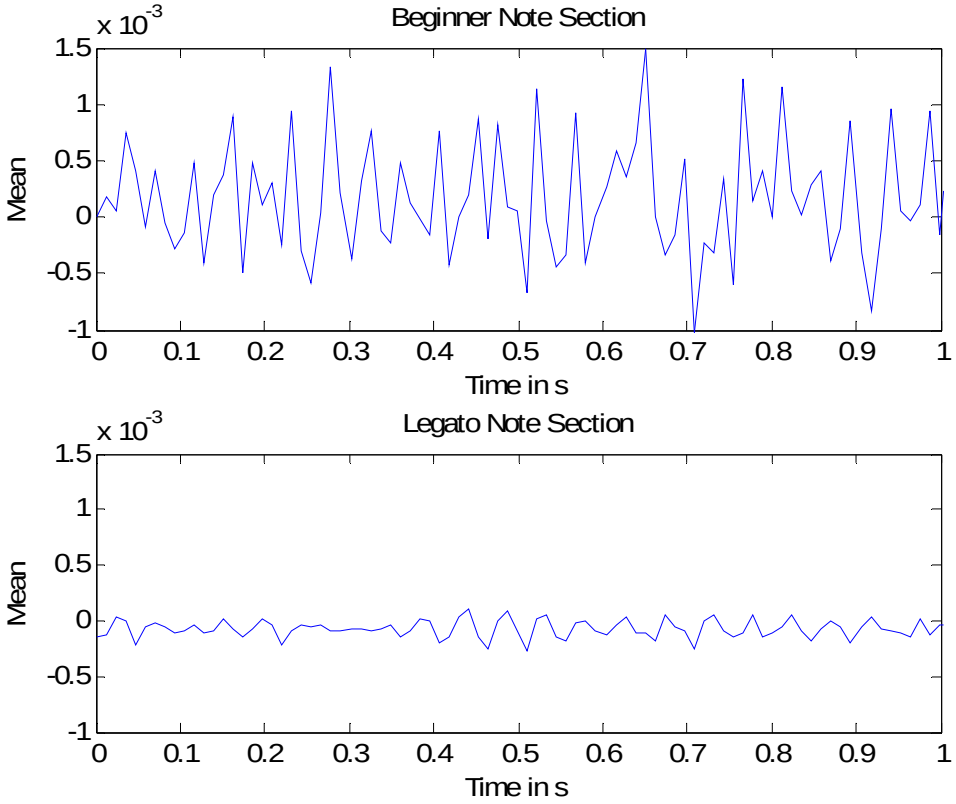

Figure 5.6: Moving mean sections of beginner (top) and professional standard legato (bottom) note samples.

The waveform amplitude mean results presented so far have been calculated on complete note samples. Next, the windowed mean or moving mean results are given. The motivation for applying a moving mean approach was to inspect at which point the sound quality changes if possible. The moving mean values of a typical section of a professional standard legato note sample and that of a beginner note are illustrated in Figure 5.6.

The moving mean results remain steadier and fluctuate less for the legato note sample compared to those taken from the beginner one. From the results obtained, it is not possible to show at which point the sound quality changes. Observing these results prompted looking into taking the variance of the moving mean waveform amplitude mean, the results of which are displayed in Figure 5.7. The moving average has been

taken using a window length of 1024 with 50% overlap. The values displayed in Figure 5.7 show a distinct gap between the results of the two player groups, indicating a feature which distinguishes well between the beginner and the professional standard legato note samples in the dataset. When first plotted, the moving mean variance (MMV) values made it difficult to observe the distinct sample grouping. This was due to five beginner samples whose values are outliers, far exceeding the highest beginner sample MMV value shown in Figure 5.7. These samples have been replaced with the beginner samples' mean MMV and are marked in Figure 5.7 by black asterisks. The actual MMV readings of these samples are listed in Table 5.2. Applying a t-test with a 0.01 significance level to the MMV results, the null hypothesis is rejected and a p-value of 9.9*10-5 returned. The differences displayed between the different player types as reflected by the waveform amplitude MMV values are statistically significant.

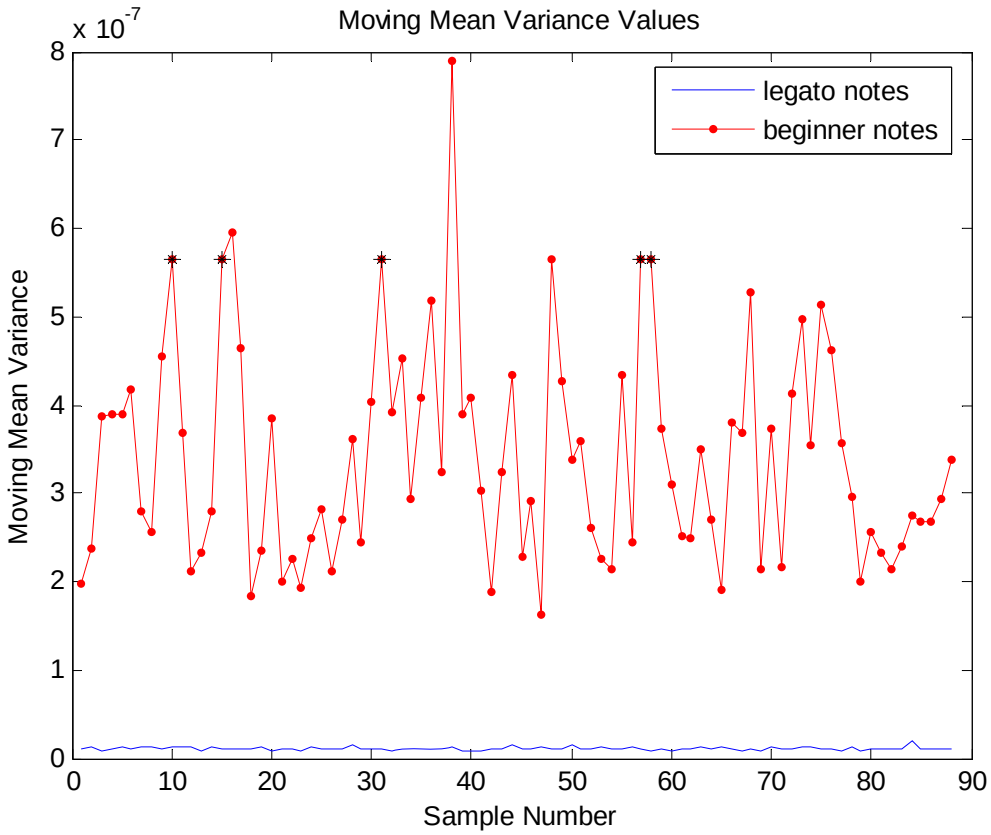

Figure 5.7: Moving mean variance values.

Sample	*Moving Mean Variance*	*Grade*
Beginner 10	$2.64*10^{-6}$	2.67
Beginner 15	$3.06*10^{-6}$	1.67
Beginner 31	$2.57*10^{-6}$	1.43
Beginner 57	$1.1*10^{-5}$	2.52
Beginner 58	$1.98*10^{-6}$	2.86

Table 5.2: Moving mean values of samples replaced with asterisks in Figure 5.7.

The waveform amplitude mean and its MMV, provide descriptors which separate effectively between the beginner and professional standard legato note samples in this dataset. The waveform amplitudes are much more consistent and symmetric for the professional standard legato note samples than they are for the beginner ones, which accounts for the visible gap between the mean readings between these two player groups. The MMV reflects the greater fluctuations present in beginner sample waveform amplitudes. Comparatively, the effect of forcing the sound on the MMV values is displayed in Figure 5.8.

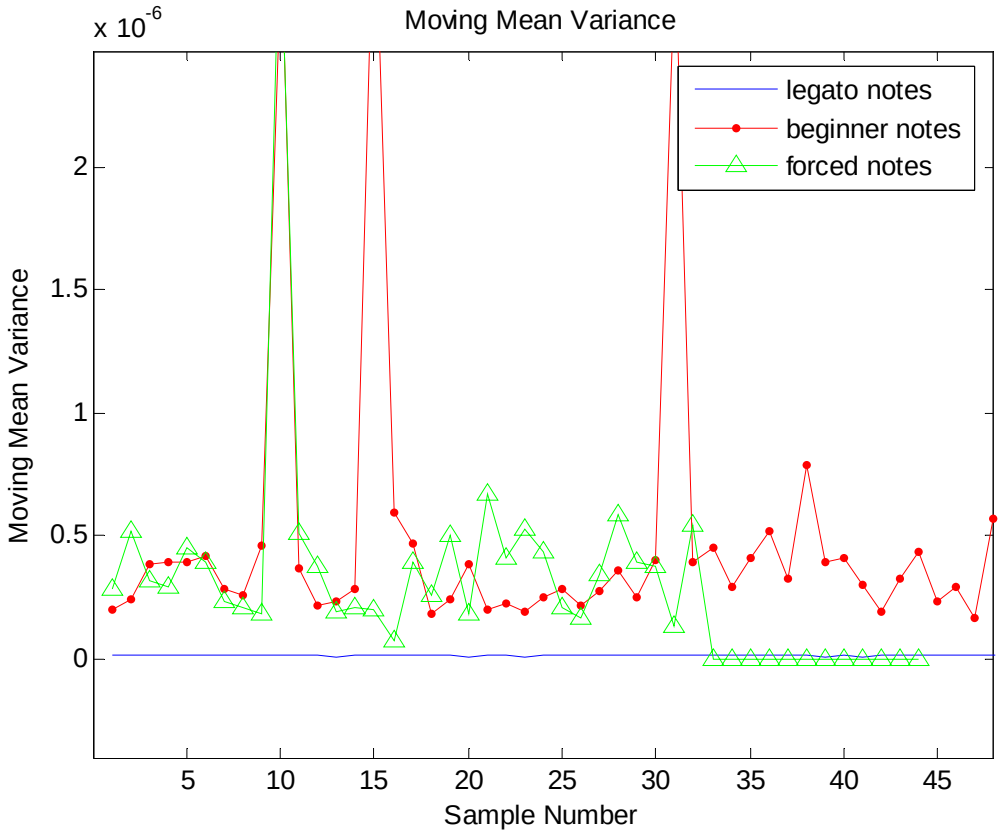

Figure 5.8: Waveform amplitude moving mean variance results professional standard legato, beginner and forced note samples.

Samples with emulated crunching at the starts and ends of notes return MMV readings similar to those belonging to the beginner note samples. When forcing is present throughout a sample, as in the forced note samples numbered 33 onwards, the MMV values drop and the readings match those returned by the legato note samples. MMV is a measure reflecting waveform amplitude consistency.

The waveform amplitude MMV separates the beginner from the professional standard legato note samples in the dataset effectively. Partial, rather than continued

crunching causes greater, more uneven changes in the waveform amplitude and is reflected by this measure. Any playing fault or bowing style which has this effect on the waveform amplitude is also reflected by the MMV, meaning that as a measure it does not exclusively reflect crunching, but playing faults and styles which cause the waveform's amplitude to change suddenly. The mean readings in the frequency domain are not as useful as they are pitch dependent. For this information to be comparable, all samples in the dataset would need to have the same note. This research focuses on violin sound discriminators that are both pitch and sample length independent. Moving variance values are presented in the section on waveform amplitude variance which is presented in the following section as a violin timbre feature.

5.2 Second Moment: Variance

Variance is a measure of the spread of the data and is given by Equation 5.2 [Stuart87]:

$$\text{var} = \left(\frac{1}{N-1} \sum_{i=1}^{N} (data(i) - \overline{data})^2 \right) \tag{5.2}$$

The waveform amplitude variance values of the beginner, professional standard legato and forced note samples are displayed in Figure 5.9.

In this figure, the professional standard legato note samples tend to have lower and more consistent waveform variance values. Comparatively, the beginner note samples' values are much less consistent. A beginner player does not have enough bow control to be able to produce consistent results, hence some very high variance values within these results. Using variance does not offer a particularly useful option for separating beginner and professional standard notes within this dataset but these values support what a trained listener would say about beginner sound compared to a good violin sound regarding consistency. From these readings, differing amounts of crunching and forcing return the lowest variance values. In particular, deliberate crunching throughout the duration of the note, which occurs from forced sample 33 onwards, returns the lowest readings and are more clearly shown in Figure 5.10. An explanation for this is that the waveform amplitude variance readings remain consistent for the continued crunching samples.

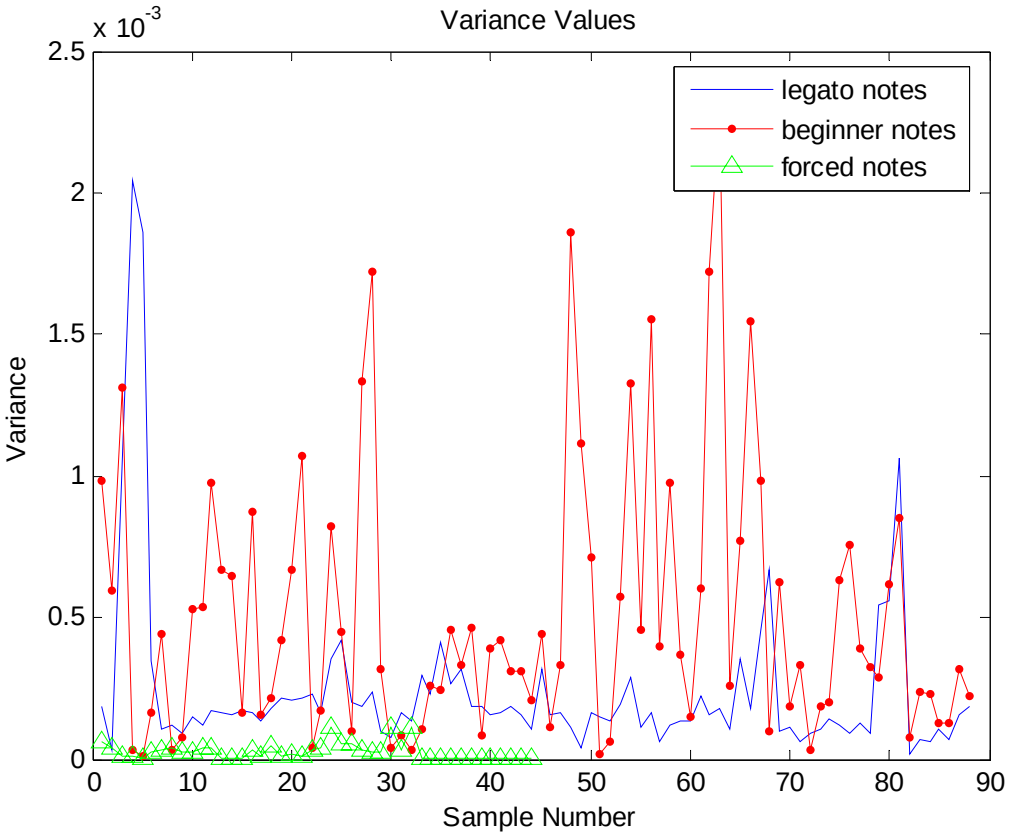

Figure 5.9: Waveform amplitude variance values for professional standard legato, beginner and forced note samples.

The beginner samples with the eight lowest variance readings are detailed in Table 5.3, seven of which have crunching. From these results, crunching alone is not a sufficient condition to lower a sample's waveform amplitude variance value. In the dataset, 33 beginner samples contain crunching according to the listening tests. Not all of these samples have low variance values. The waveform amplitude variance value reflects consistency or amplitude change throughout a sample.

Sample	*Grade*	*Variance*	*Crunching?*
Beginner 4	1.19	$0.14*10^{-4}$	yes
Beginner 5	1.1	$0.30*10^{-4}$	yes
Beginner 8	1.43	$0.36*10^{-4}$	yes
Beginner 22	1.24	$0.43*10^{-4}$	yes
Beginner 30	1.67	$0.41*10^{-4}$	yes
Beginner 32	2.81	$0.34*10^{-4}$	no
Beginner 51	2.57	$0.21*10^{-4}$	yes
Beginner 72	2.43	$0.34*10^{-4}$	yes

Table 5.3: Beginner samples with lowest waveform amplitude variance values.

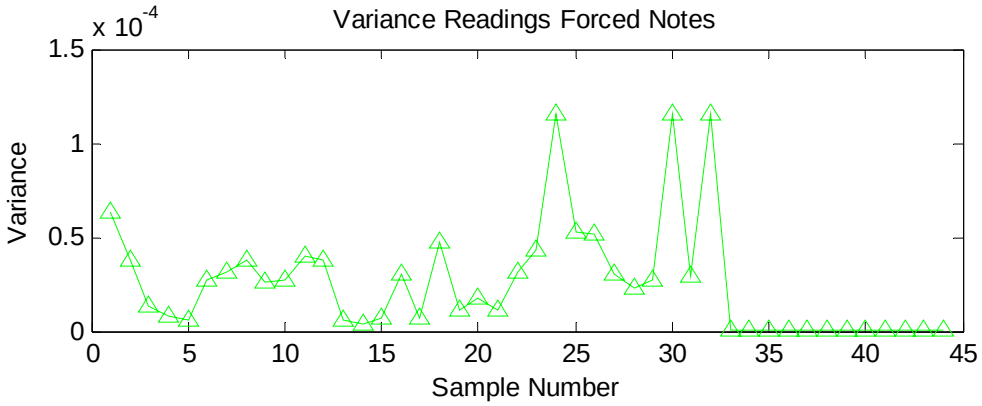

Figure 5.10: Variance readings of forced note samples.

The waveform amplitude moving variance is presented next. The moving variance readings of a professional standard legato and that of a beginner note are illustrated in Figure 5.11 and in Figure 5.12 respectively. Based on these results, finding a one-value measure for the dataset, independent of sample length, reflecting the different sample groups is not evident based on this information.

Waveform amplitude variance readings obtained for the dataset reflect a beginner player's inconsistency by returning varied results comparatively to those obtained from the professional standard legato note samples. Using the forced note samples allows the relationship between waveform amplitude and variance reading to be further tested. Results show that waveform amplitude consistency and not sound quality is reflected by this measure. These results support what is said about beginner playing in terms of consistency but are inconclusive regarding detecting overall violin sound quality within this dataset and cannot be easily correlated with any of the qualitative expressions used. The third moment, skew, is considered as a potential violin timbre feature and is presented in the next section.

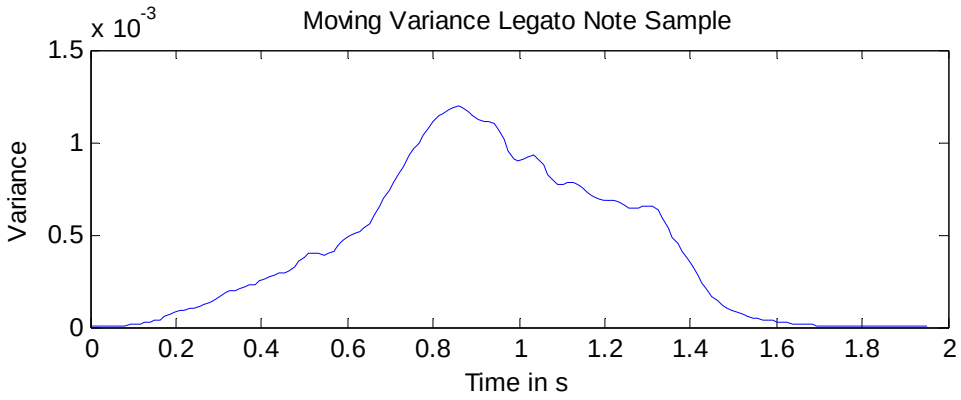

Figure 5.11: Moving variance results for a legato note sample.

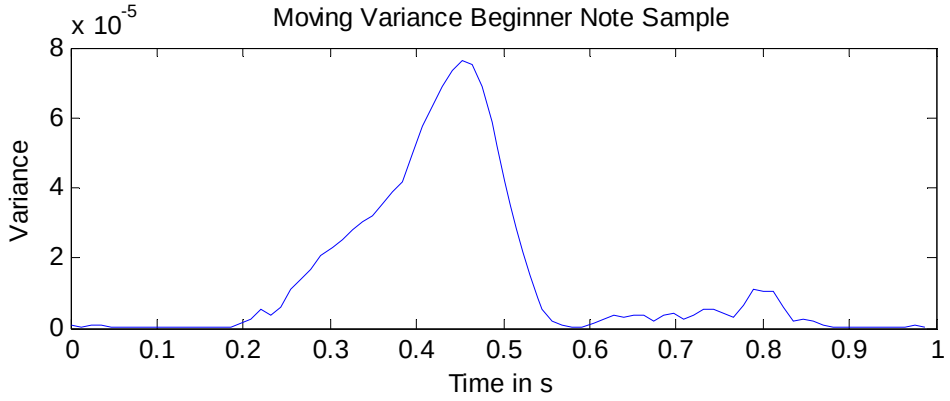

Figure 5.12: Moving variance results for a beginner note sample.

5.3 Third Moment: Skew

Skewness is a measure of symmetry. A normal distribution has a skewness value of zero and the more symmetric data is, the closer its skew value is to zero. Negative skew values indicate that the data is skewed to the left whereas positive values represent data skewed to the right. Skew is defined in Equation 5.3 where σ is the standard deviation [Stuart87]:

$$\text{Skew} = \frac{\sum_{i=1}^{N} (data(i) - \overline{data})^3}{(N-1)\sigma^3} \tag{5.3}$$

The waveform amplitude skew values for the dataset samples are plotted in Figure 5.12, where mostly overlapping results for both player sample groups are displayed.

The most prominent peaks in Figure 5.13 have been obtained from beginner note samples. A summary of how these samples have been perceived by the listeners is given in Table 5.4. All contain various faults and have overall sound quality grades below 3 from the listening tests. No similarity between faults perceived, overall sound quality grade and waveform amplitude skew value can be easily drawn, as reflected by the samples detailed in Table 5.4.

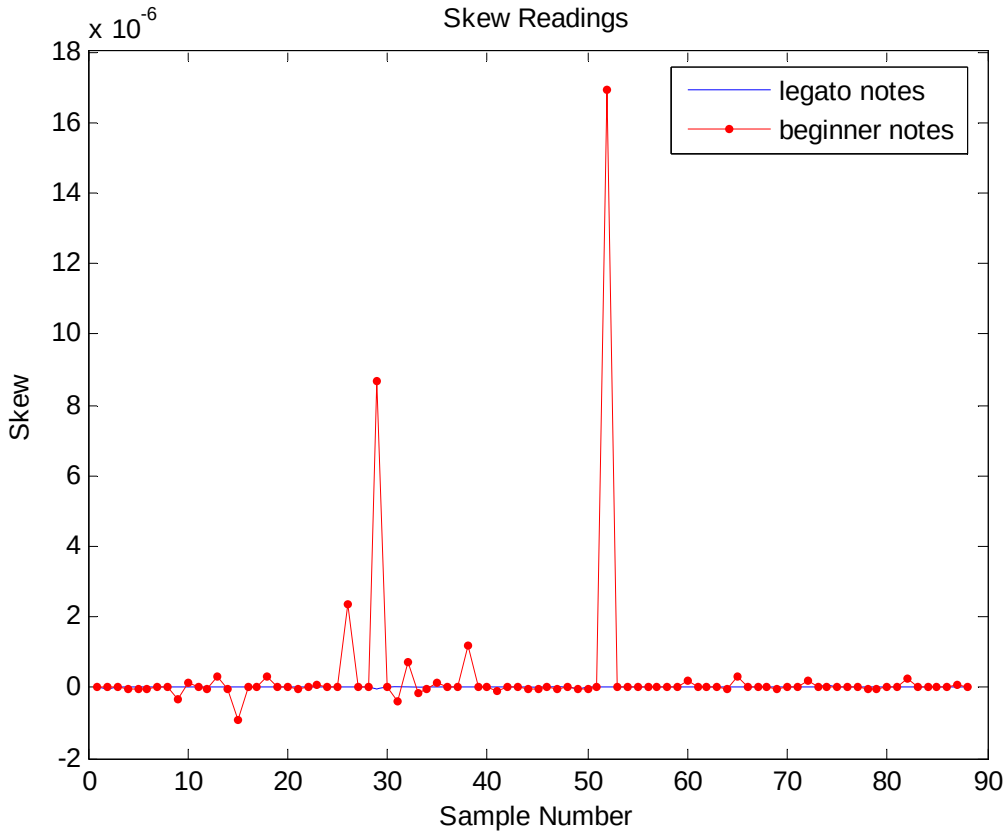

Figure 5.13: Waveform amplitude skew values for beginner and professional standard legato note samples.

Sample No.	*Quality Grade*	*Faults Present*
Beginner 26	1.24	NV, INT, BB, BADS, BADE
Beginner 29	2	SK, NV, XN, BADS, BADE
Beginner 38	1.14	CR, SE
Beginner 52	2.81	SE

Table 5.4: Information about prominent beginner note samples in Figure 5.13.

The legato note samples in Figure 5.13 provide results that, when on the same scale, give the impression of forming a straight line. To better view these results, they are plotted separately in Figure 5.14 where the difference in scale between these sample groups is noted.

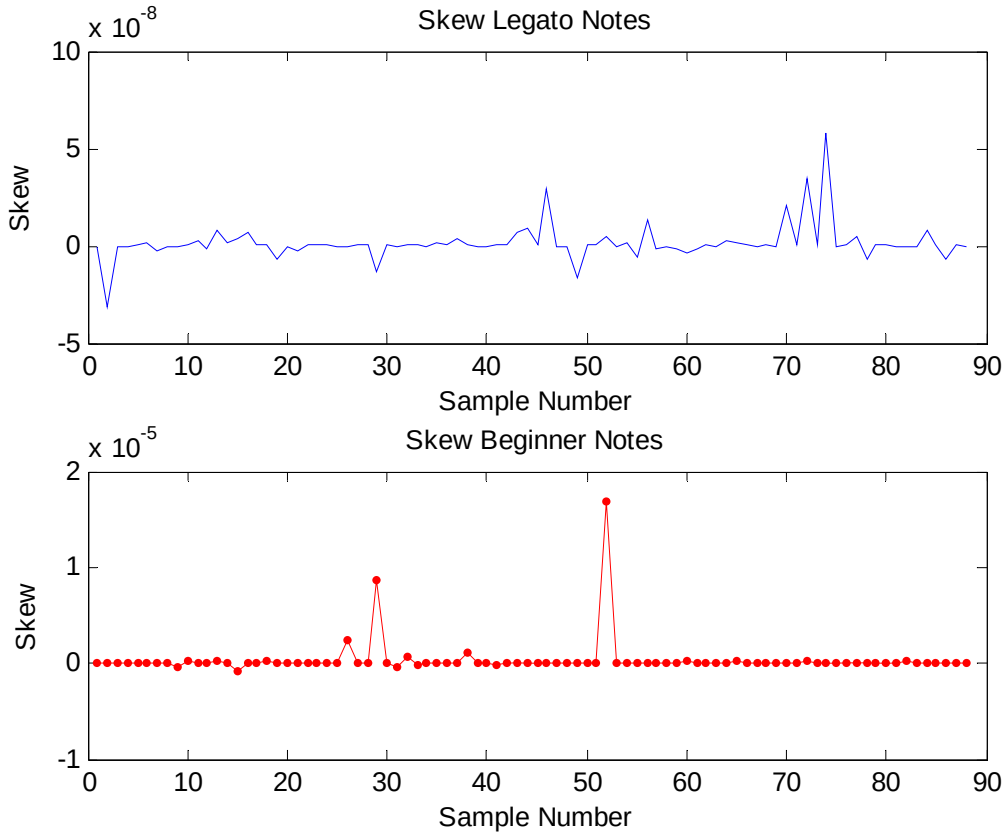

Figure 5.14: Skew values plotted of professional standard legato (top) and beginner (bottom) note samples.

From the analysis carried out on the dataset so far, using skew readings within the context of violin sound quality and playing fault detection and this dataset have not proven to be effective. No perceptual correlates at this point can be assigned either but waveform unevenness is picked up by the skew reading as illustrated in Figure 5.15. The waveforms of three beginner samples are displayed in this figure: the first has the highest positive skew reading, the second, has the skew value closest to zero, and the third one has the largest negative skew reading.

The larger the skew value in either a positive or negative direction is reflected by greater asymmetry in the waveforms. The closer the skew reading is to zero, the smoother the overall waveform. This makes sense given the definition of skew but how this can be linked to the qualitative expressions used in this thesis is not evident. Information relating to the samples shown in Figure 5.15 and a legato note sample with skew closest to zero are given in Table 5.5. The relationship between kurtosis and the dataset's violin note samples is presented next.

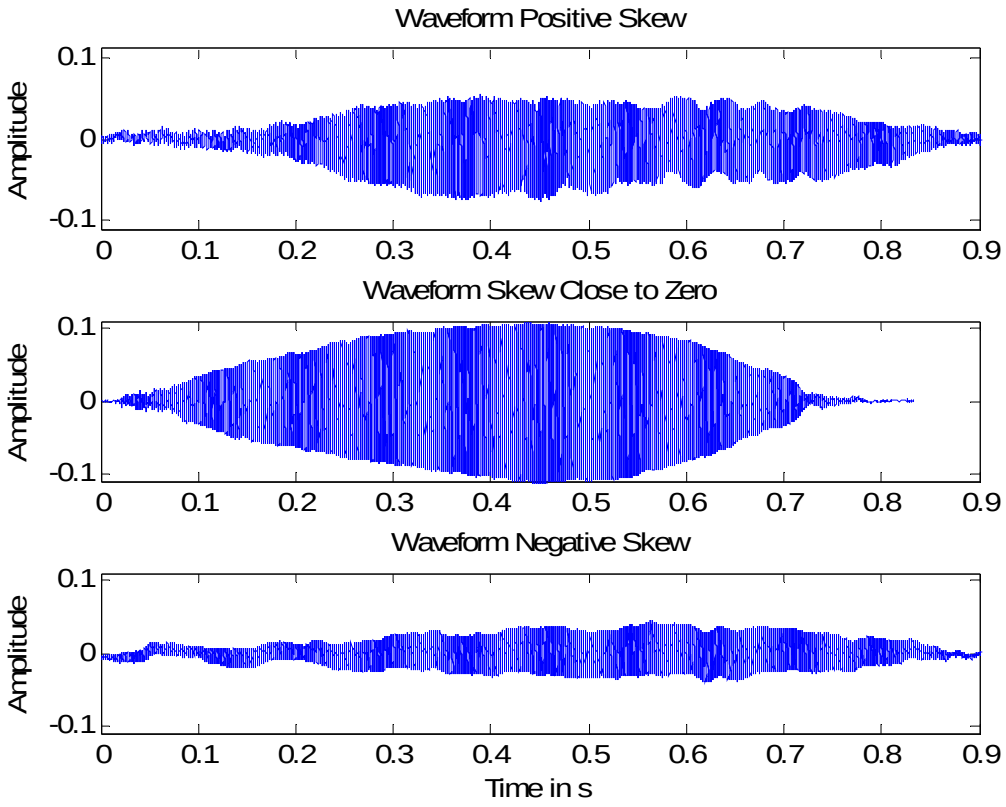

Figure 5.15: Waveforms of beginner samples with the highest positive skew (top), Skew closest to zero (middle) and the lowest negative skew (bottom).

Sample No.	*Skew*	*Grade*	*Perceived Faults?*
Beginner 38	$1.21*10^{-6}$	1.14	CR, SK, NV, XN, BADS, BADE
Beginner 56	$-2.90*10^{-12}$	2.86	INT
Beginner 33	$-1.55*10^{-7}$	1.48	SK, NV
Legato 79	$1.76*10^{-14}$	5.43	None

Table 5.5: Information about beginner samples shown in Figure 5.15.

5.4 Fourth Moment: Kurtosis

Kurtosis measures the data's "peakiness" compared to that of the normal distribution. A high kurtosis value is representative of data that has a distinct peak located close to the mean. A signal with a low kurtosis value tends to have a much flatter top around the mean rather than a sharp peak. Kurtosis values greater than three or positive values depending on the equation used, represent a peaked distribution (super Gaussian). Whereas values less than three or negative values indicate a flat distribution (sub Gaussian). Kurtosis is obtained from Equation 5.4 [Stuart87]:

$$kurtosis = \frac{\sum_{i=1}^{N} (data(i) - \overline{data})^4}{(N-1)\sigma^4} \tag{5.4}$$

From Equation 5.4, the normal distribution has a kurtosis reading of 3. The waveform amplitude kurtosis readings obtained for the dataset's beginner and professional standard player legato note samples are plotted in Figure 5.16.

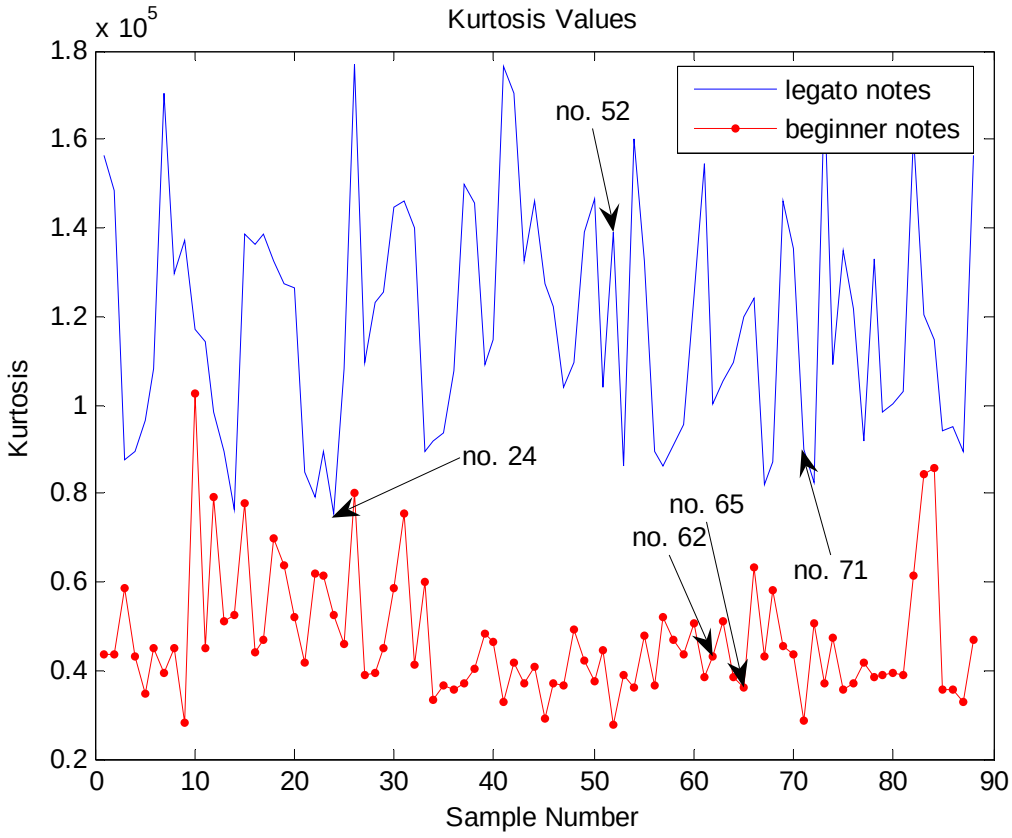

Figure 5.16: Waveform amplitude kurtosis values for beginner and professional standard player legato notes.

Sample No.	*Grade*	*Kurtosis*	*Beg or pro?*	*Faults perceived?*	*Comments*
Legato 24	4.2857	$7.5518*10^{+4}$	Beginner	BB	Pro perceived as beg
Legato 52	4.76	$1.39*10^{+5}$	Beginner	None	Worst professional
Beginner 62	3.95	$4.33*10^{+4}$	Beginner	None	Best beginner
Beginner 65	3.86	$3.61*10^{+4}$	Beginner	None	2nd best beginner
Legato 71	3.81	$9.01*10^{+4}$	Beginner	None	2nd worst professional

Table 5.6: Information about marked samples in Figure 5.16.

From the results displayed in Figure 5.16, both sample lists provide quite "peaky" or super-Gaussian results as all values returned are greater than 3. Although some overlapping values are present in this figure, separation between the two different player sample groups is good. The beginner note samples tend to have lower kurtosis values than the professional standard legato ones. From these results, the professional standard legato note samples have "peakier" distributions than the beginner ones. The samples with overlapping kurtosis values in Figure 5.16 are detailed in Table 5.6. What these

samples have in common, as captured by the peakiness of their distributions, has not been reflected by how they have been perceived by the listeners, nor by their overall sound quality grading. The mean waveform amplitude kurtosis reading for professional standard legato note samples is $1.19*10^{+5}$ and for the beginner samples, $4.7*10^{+5}$. The highest beginner note kurtosis values which overlap with some of the professional standard legato note kurtosis values, all contain different perceived playing faults and overall sound quality grades. From the listening tests, the two perceived best sounding beginner samples, samples 62 and 65, both return relatively low kurtosis values. From the listening tests, no evident link can be established between kurtosis reading and any of the qualitative playing expressions used in this text. The difference between the waveform amplitude kurtosis values for the dataset's beginner and professional standard legato note samples is statistically significant. The null hypothesis of a t-test with a 0.01 significance level is rejected and a p-value of $3.9*10^{-47}$ is returned. To further investigate kurtosis values within the violin timbre context, the effect of deliberately forced notes is presented next.

The effect of forcing the sound has on waveform amplitude kurtosis value is displayed in Figure 5.17. As has been observed previously with waveform amplitude mean values, the forced notes group splits into two sections when the kurtosis values are taken. The first 32 samples, which have emulated crunching at the starts and ends of all notes, have lower kurtosis values than the beginner note samples. Forcing throughout the note, as in forced note samples numbered 33 onwards is reflected by an increase in kurtosis value.

From the grouping patterns present in Figure 5.16 and Figure 5.17, the results show that kurtosis reflects sound quality. These kurtosis readings are sensitive to levels of poor sound quality. What can be considered as being the worst sounding samples, where the sound has been forced throughout the note, do not return the lowest kurtosis readings, implying a measure which reflects waveform consistency. The specific element of sound quality captured by the kurtosis value is not represented by the qualitative playing terms used in this text. From these results, kurtosis values differentiate between beginner and professional standard legato notes within the dataset effectively but do not directly reflect the qualitative expressions used in this thesis. Autocorrelation is presented in the following section.

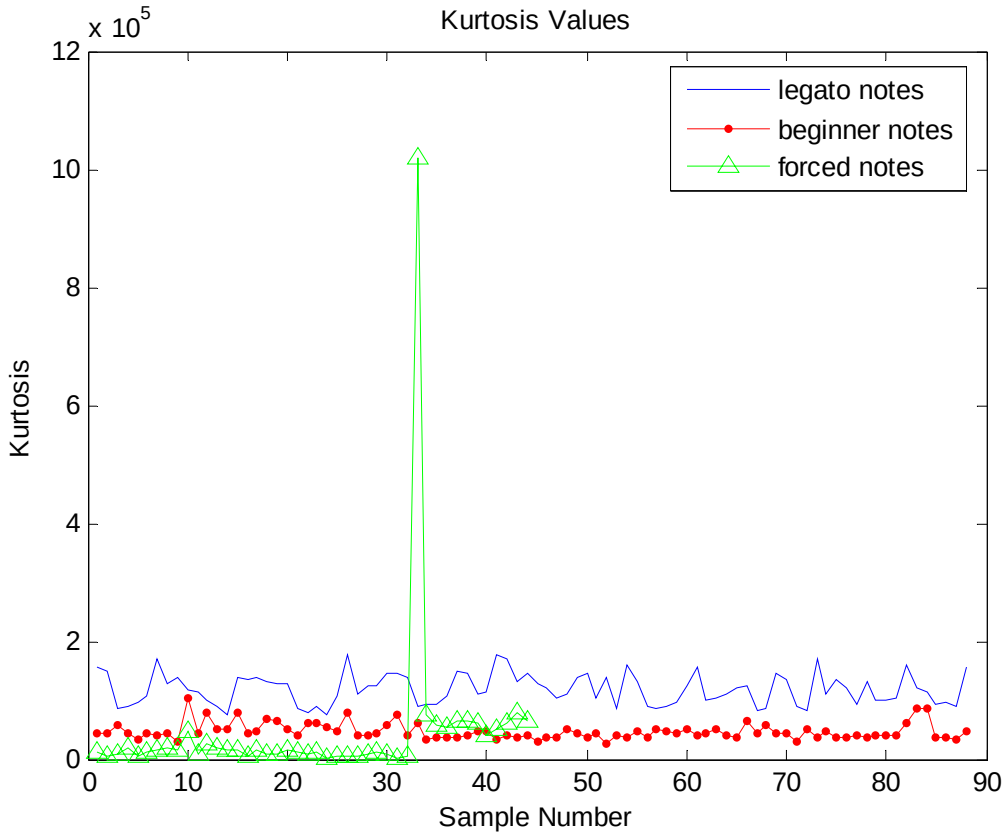

Figure 5.17: Kurtosis values for professional standard legato, beginner and forced note samples.

5.5 Autocorrelation

Autocorrelation is the correlation of a signal with itself and is useful for determining signal periodicity and is used as a first stage in some pitch detection systems [Oppenheim89]. It has also been used as a feature in musical instrument identification tasks [Martin98]. More specifically it quantifies the closeness of the amplitudes of two samples as a function, in this case, of their time separation and is given by Equation 5.5 [Jayant84]:

$$acf(k) = \frac{1}{N} \sum_{n=0}^{N-k-1} x(n)x(n+k) \tag{5.5}$$

The mean autocorrelation values of the dataset's samples are displayed in Figure 5.18.

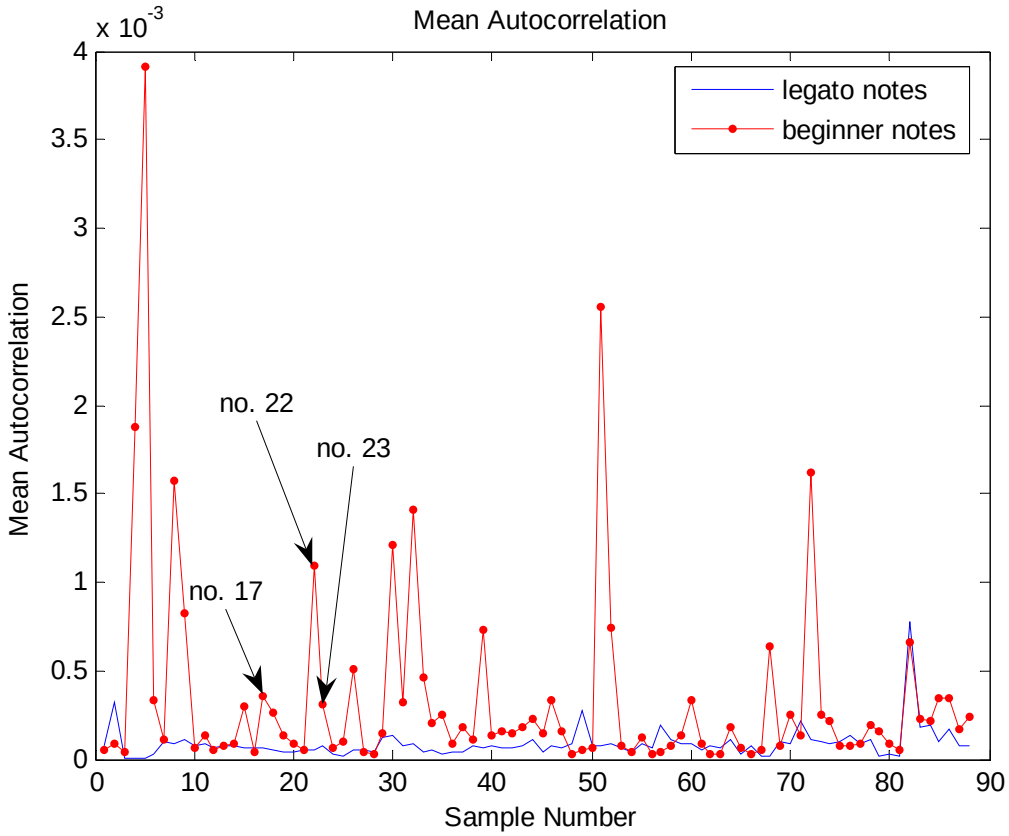

Figure 5.18: Mean autocorrelation values of beginner and professional standard legato note samples.

Beginner samples 17, 22 and 23, are recorded as being the worst sounding samples in the dataset, having overall quality grades of 1. These samples have not been observed to have the highest mean autocorrelation values. Samples having similar autocorrelation means do not have similar sound attributes. The professional standard legato note samples have mean autocorrelation values that are low and comparatively consistent to those obtained from the beginner note samples. A low autocorrelation mean implies little or gradual change in the signal with respect to time, as reflected by the legato note sample readings. The mean autocorrelation readings displayed in Figure 5.18 for the beginner note samples are varied. Some of these values are much more elevated, indicating sudden, jagged changes in the signal with respect to time, reflecting bowing problems more prevalent in beginner note samples. This is further illustrated by the mean autocorrelation values returned for the forced note samples, which are displayed in Figure 5.19. From forced note sample number 33 onwards, crunching is maintained for the duration of each note. This is reflected by a sudden increase in mean autocorrelation values, indicating much change in the signals. As a feature, the mean

autocorrelation does not discriminate effectively between the beginner and professional standard legato notes in this dataset but reflects the inconsistency associated with beginner playing. Although the autocorrelation mean results overlap, the professional standard legato note samples tend to have lower and much more consistent values than the beginner ones do.

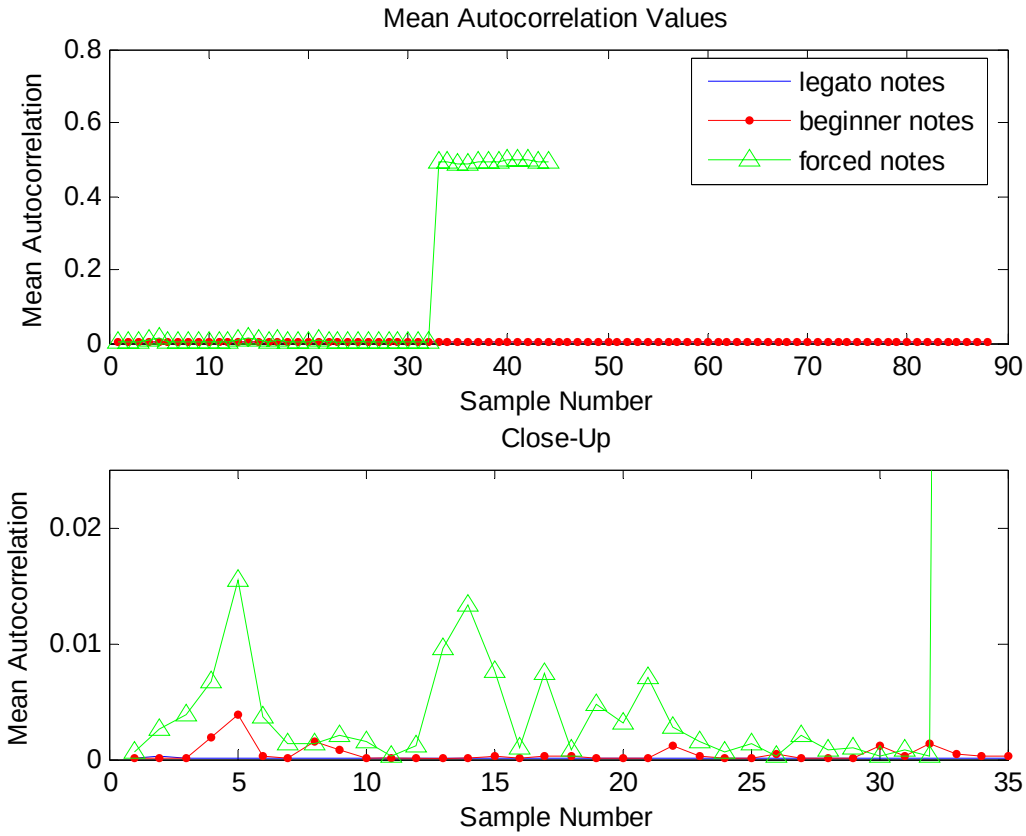

Figure 5.19: Mean autocorrelation values for professional standard legato, beginner and forced note samples (top) and close-up (bottom).

5.6 Summary

First order statistics and the mean autocorrelation have been applied to the dataset's samples and the results obtained have been presented in this chapter. The waveform amplitude mean and moving mean variance readings returned excellent results in terms of detecting the different player groups within the dataset used. Using some of these measures, such as the TM and the MMV, allows the beginner note samples to be separated with 100% accuracy from the legato professional standard ones in the dataset. The difference between the values returned for these features for the dataset's beginner and professional standard legato note samples have been shown to be statistically significant by rejecting the null hypotheses of the applied t-tests. Another feature, TK,

separates well the two player types in the dataset, with few overlapping sample values. An underlying pattern in the waveform amplitude variance results showed a greater range in variance readings for the beginner note samples than those for the professional standard legato ones. The variance readings reflect the inconsistency of the beginner note samples. On a note by note basis, this is of limited use as no direct link between sound quality and variance value has been established, but the underlying pattern is of interest. Much overlap is present in the skew readings for the dataset's samples, making it not a suitable feature for violin sound classification within the context of this research. However, true to its definition, skew readings reflect waveform asymmetry. Samples that are strongly positively or negatively skewed in this dataset have been associated with much more asymmetric waveforms. The samples having skew values close to zero have waveforms that are much more symmetric. The mean autocorrelation returned overlapping results which are of limited use towards these specific research aims but the underlying pattern reflects waveform consistency. An issue with consistency measures is that they do not quantify good from bad samples unless they are comparatively inconsistent.

The most significant features in this section based on the waveform amplitude are the TM, TK and MMV values. These measurements alone are not sufficient to meet the thesis' aims, but when combined with features from other domains, more robust results are expected. In the following chapter, Chapter 6, spectral domain features for representing violin timbre are presented.

6 Spectral Analysis

In Chapter 5, temporal analysis has been applied to the data with the aim of finding suitable violin timbre features. This chapter details spectral analysis to further understand and represent violin timbre. Spectral analysis permits the component frequencies present in a sound to be observed, giving insight into its harmonic structure and timbre. Among the features presented in this chapter are constant Q transform (CQT) based harmonic content strength features, spectral flux, spectral centroid, power spectrum, spectral flatness and spectral contrast measures. In this chapter, the efficacy of spectral features for representing the violin timbre space, discriminating between beginner and professional standard legato notes as well as fault detection within the dataset used are presented.

6.1 Constant Q Transform

The CQT as introduced by Brown in [Brown91] yields a log-frequency scaled time-frequency representation of the signal. As illustrated in Figure 6.1 and Figure 6.2, which show A440 notes played by a professional standard and a beginner player respectively, the CQT domain is effective for visualising and exploiting information about the harmonic content of a note due to the frequency resolution.

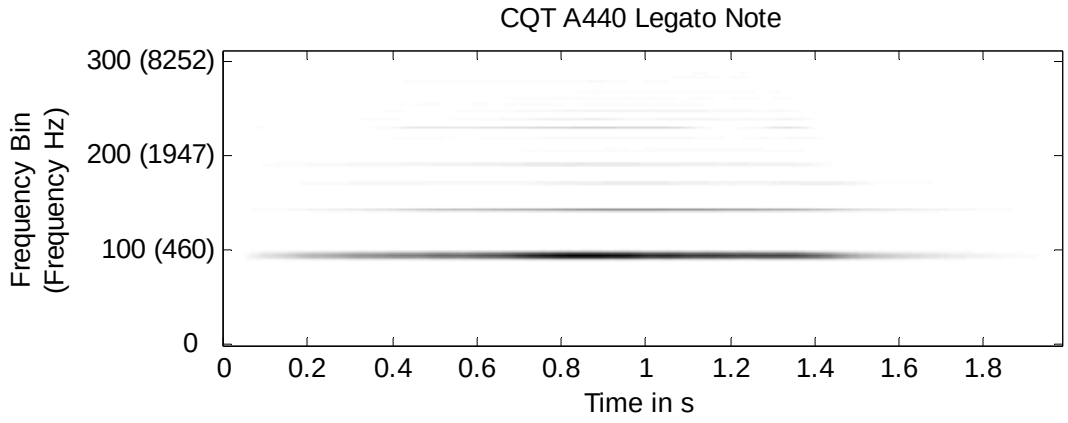

Figure 6.1: CQT of a professional standard legato A440 note.

Contrasting these figures, the beginner player's note is not as cleanly executed as the professional standard legato one. This is reflected by the presence of additional

frequencies unrelated to those of the actual note as well as the harmonics not being as well defined as those displayed in Figure 6.1. This gives rise to the visible blotching and rippling effect present in Figure 6.2, particularly from 0.65s onwards in this image.

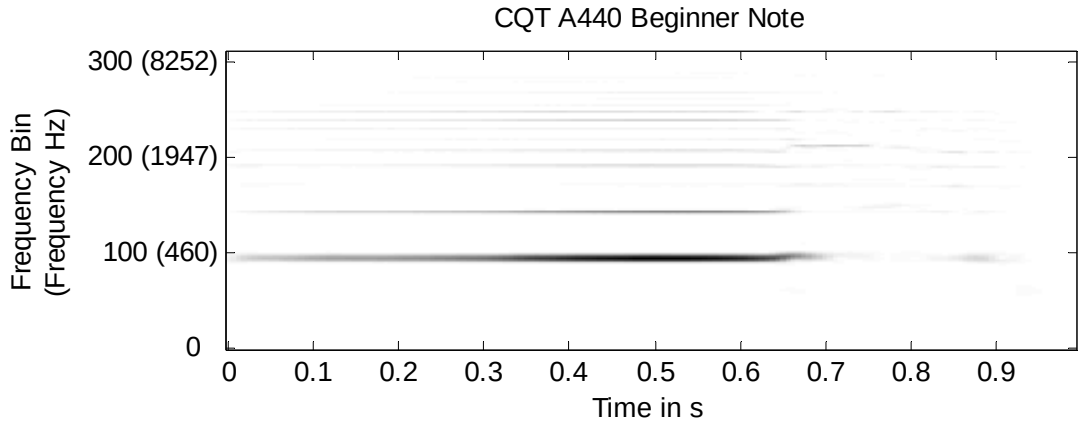

Figure 6.2: CQT of a beginner A440 note.

From observing the CQT representations of the dataset samples, differences between the beginner and professional standard legato note samples are visible. The CQT representation uses 312 frequency bins with the first frequency bin centre set at 110Hz which is well below the frequency of the lowest note on the violin, G3 (approximately 196Hz when tuned to A440). The frequency content present below this note may contain unwanted content reflecting playing quality. Focusing within this frequency range and recalling that eighth tone spacing has been used in this study means that bin 41 has a centre frequency of 196Hz and bin 40, that of 193.2Hz. The mean content of each sample for the first 40 individual frequency bins was taken and revealed some useful information. Separation between the dataset's professional standard legato and beginner note samples is achieved using this information. Nine out of these 40 frequency bins, labelled according to their centre frequencies, are detailed in Table 6.1, correctly group these two player types in the dataset.

Frequency Bin No.	*f_c (Hz)*
4	114.87
5	116.54
6	118.24
7	119.96
8	121.70
9	123.47
10	125.27
11	127.09
20	144.73

Table 6.1: CQT frequency bin centre frequencies which effectively group beginner and professional standard legato note samples.

Fewer than 10 overlapping samples are present in the results obtained from these frequency bins. To illustrate the level of separation between player groups, the mean frequency content from frequency bins four (f_c=115Hz), nine (f_c=123Hz) and 20 (f_c=145Hz) are plotted in Figure 6.3, where the beginner note samples are in red and the professional standard legato ones, in blue. The results returned for all these frequency bins are illustrated in Appendix A. Applying a t-test with 0.01 significance level to the results displayed in Figure 6.3, returned results that are statistically significant. The null hypothesis is rejected in all three cases with p-values of $1.1*10^{-78}$, $2.5*10^{-99}$ and $3.4*10^{-80}$ respectively.

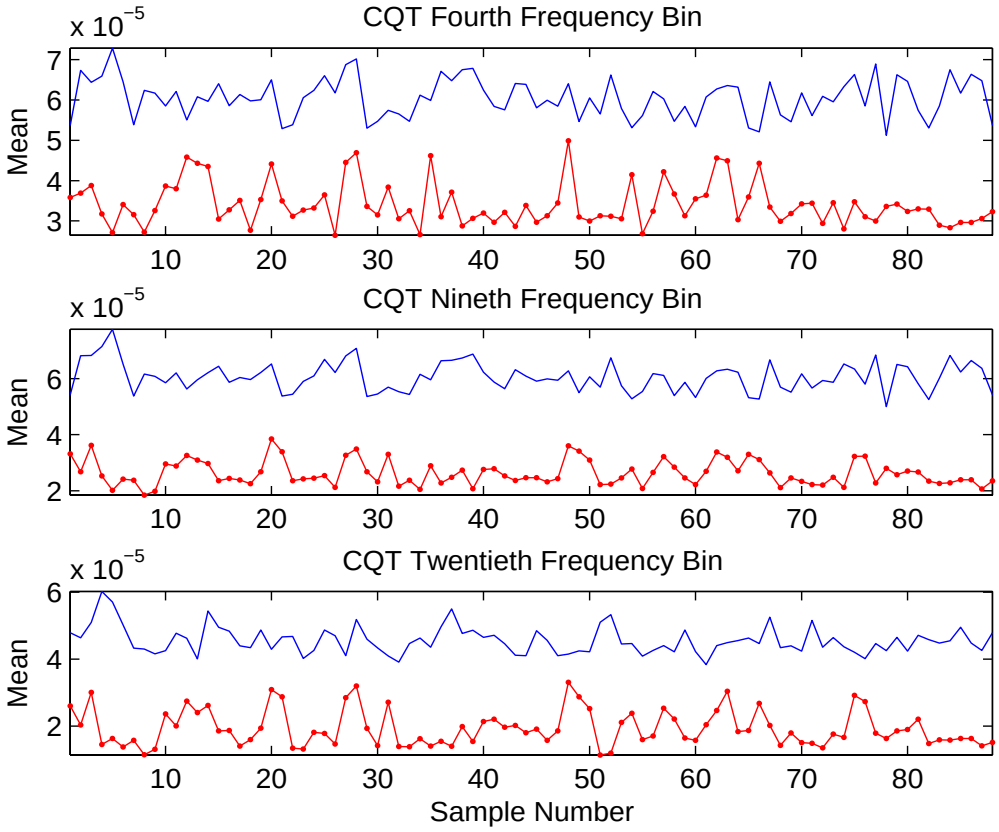

Figure 6.3: Mean frequency content from CQT bins four (f_c=115 Hz), nine (f_c=123 Hz) and twenty (f_c=145 Hz).

Excellent separation between the professional standard legato and beginner note samples within the frequency range 110Hz to 196Hz is displayed in Figure 6.3 and in Appendix A. The professional standard legato note samples have higher mean frequency content in these frequency bins than the beginner ones do. An explanation for this gap in frequency content between the beginner and professional standard legato note samples is the excitation of instrument resonances and modes. A violin's

fundamental cavity resonance is at approximately 260-290Hz [Hutchins97], the frequency range for a sub-harmonic is approximately 135-145Hz. Taking the mean frequency content of the twentieth frequency bin, which has a centre frequency of 145Hz, completely separates the beginner from the professional standard legato note samples in the dataset. A plausible explanation for these results is how the violin's fundamental cavity resonance at approximately 290Hz is excited by the different players. The professional standard legato note samples all have much higher frequency content in this bin than the beginner note samples do, as shown in the lower image in Figure 6.3. A professional standard player is expected to excite the frequencies associated with the fundamental resonance more. Consulting Marshall's work on violin modes in [Marshall85], the frequency content present in frequency bins six and seven, reflects modes 5 and 1 respectively. Mode 5, the first vertical cantilever of the fingerboard, is at 236.5Hz which is approximately twice the centre frequency of bin six. Mode 1 at 119.5Hz, is the vertical reflection of the tailpiece, is reflected by frequencies in bin seven. For further information on violin modes, refer to [Hutchins93]. The frequency content in the remaining bins may reflect specific violin modes too but this was not revealed in the material referred to. A professional standard player is expected to excite these modes more consistently and to a greater extent than a beginner player would. This accounts for the gap in average frequency content in these frequency bins. The frequency content present in nine CQT frequency bins with centre frequencies below the lowest note on the violin provide effective and statistically significant discriminators between the dataset's beginner and professional standard player legato note samples.

6.2 Spectral Flux

Spectral flux is the average correlation between amplitude spectra in adjacent windows [Scheirer96] where the amplitude spectrum is the magnitude of the DFT, $|X(n)|$. From the spectral flux, a "smoothness" factor is investigated from which harmonicity is represented. Scheirer and Slaney used spectral flux in an automatic discrimination system between music and speech. Music is reported to have a much higher rate than speech [*ibid.*]. Hawley applied the spectral flux for detecting harmonic continuity in music in his PhD thesis [Hawley93]. It has also been used as a feature in music information retrieval [Tzanetakis02].

Spectral flux is applied to the dataset with the idea of obtaining a possible crunch detection method, based on the understanding that crunching increases the number of unwanted frequencies present in a sample. The mean spectral flux values obtained for the dataset samples are displayed in Figure 6.4. Taking the spectral flux did not provide any insight into violin timbre or player fault detection due to the high number of overlapping values as displayed in Figure 6.4. For most of the samples in the dataset, change is not great enough from one window to the next to be picked up by this measure. To further inspect the relationship between sound quality and spectral flux reading, the forced note samples' spectral flux values are also depicted in this figure.

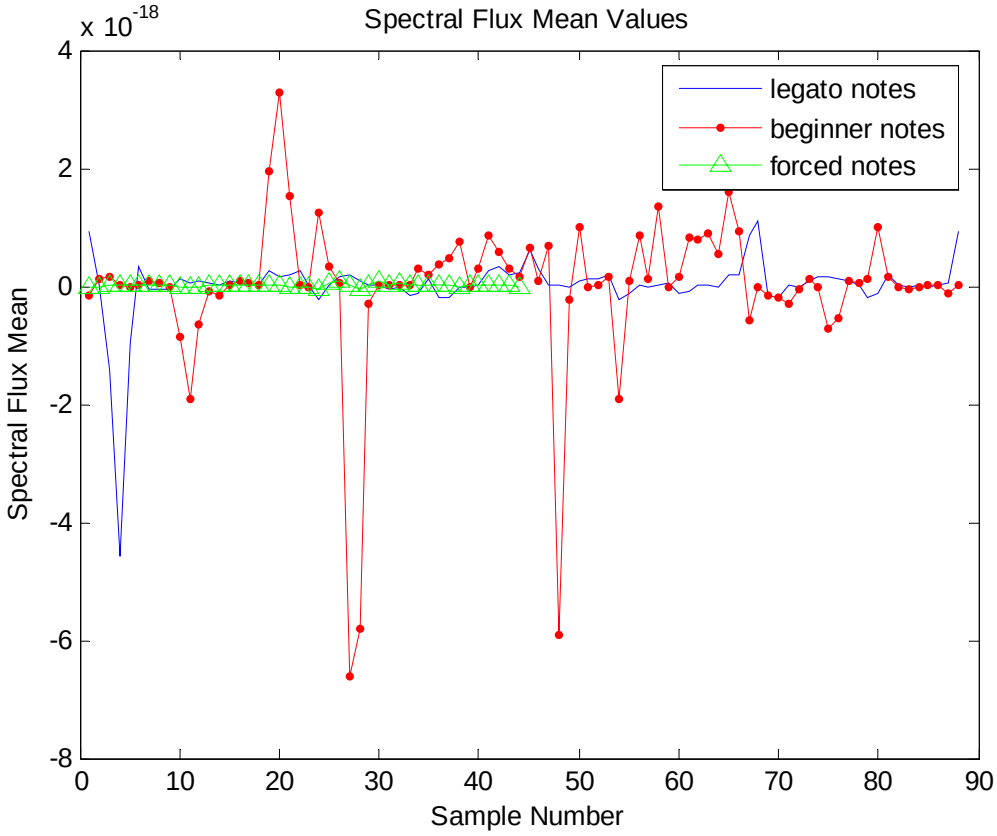

Figure 6.4: Average spectral flux for professional standard legato, beginner and forced note samples.

The forced note samples return the most consistent readings and have spectral flux closest to zero. By taking the average spectral flux of a sample, samples with harmonic content that changes little, return values closest to zero. This confirms spectral flux to be a consistency rather than a sound quality measure. Spectral flux reflects harmonic content consistency but is not effectively used to differentiate between the beginner and professional standard legato note samples in the dataset.

6.3 Spectral Centroid

The spectral centroid is defined by the ratio of the sums of the magnitudes multiplied by the relevant frequencies all divided by the sum of magnitudes and correlates strongly with the perceived brightness of a signal [Grey77]. It has been used in instrument identification tasks [Harrera00, Eronen01]. In this work the spectral centroid is applied to the violin's timbre space and its efficacy at representing sonic change is presented. Equation 6.1 is used to calculate the spectral centroid, where N is the length of the DFT, $|X(n)|$ is the magnitude of the DFT and $f(n)$ is the frequency at n [Beauchamp82]:

$$spectral\_centroid = \frac{\sum_{n=1}^{N-1} |X(n)| * f(n)}{\sum_{n=1}^{N-1} |X(n)|} \tag{6.1}$$

The waveform and spectral centroid of a professional standard legato note sample are displayed in Figure 6.5.

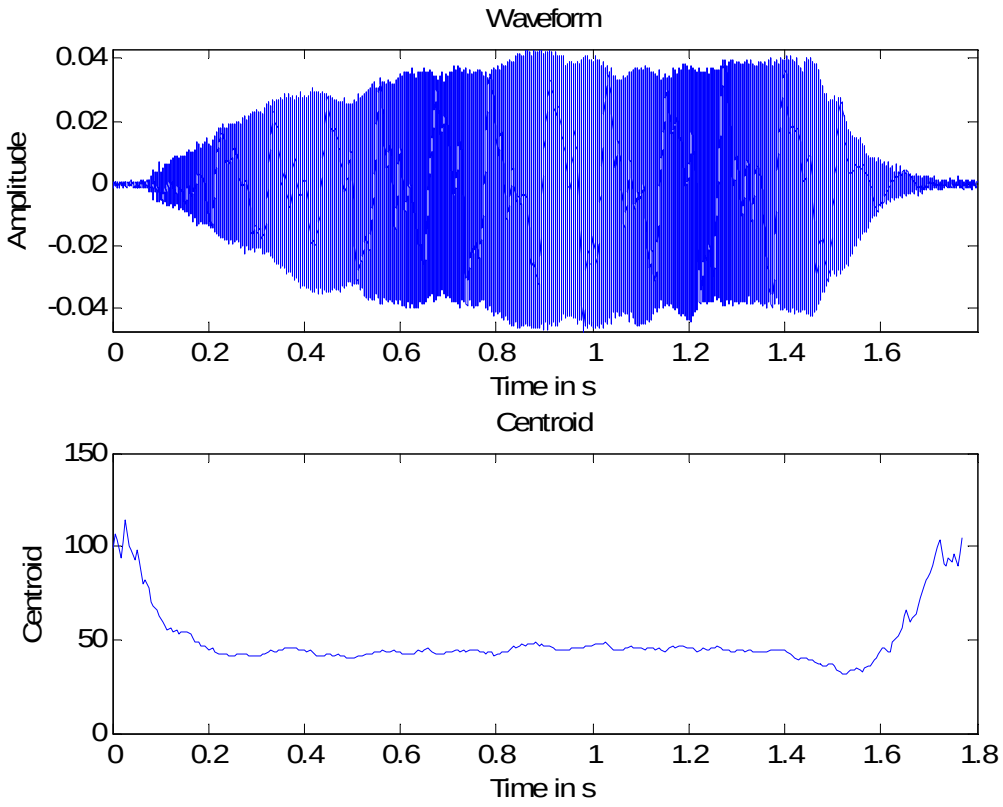

Figure 6.5: Waveform (top) and spectral centroid (bottom) of a professional standard legato note sample.

From visual inspection, using the centroid provides readings from which the waveform can be approximately split into regions (attack-steady-state-decay). After having observed the spectral centroids of the professional standard legato note samples, a similar pattern emerges. During the steady-state section of the note, where a sound is typically its most consistent, the readings drop and level out towards the middle of the note. As bow speed changes at either end of the note to prepare for a bow change, the spectral centroid increases. In the case of a reasonable sounding beginner note, the spectral centroid is less consistent, as illustrated in Figure 6.6.

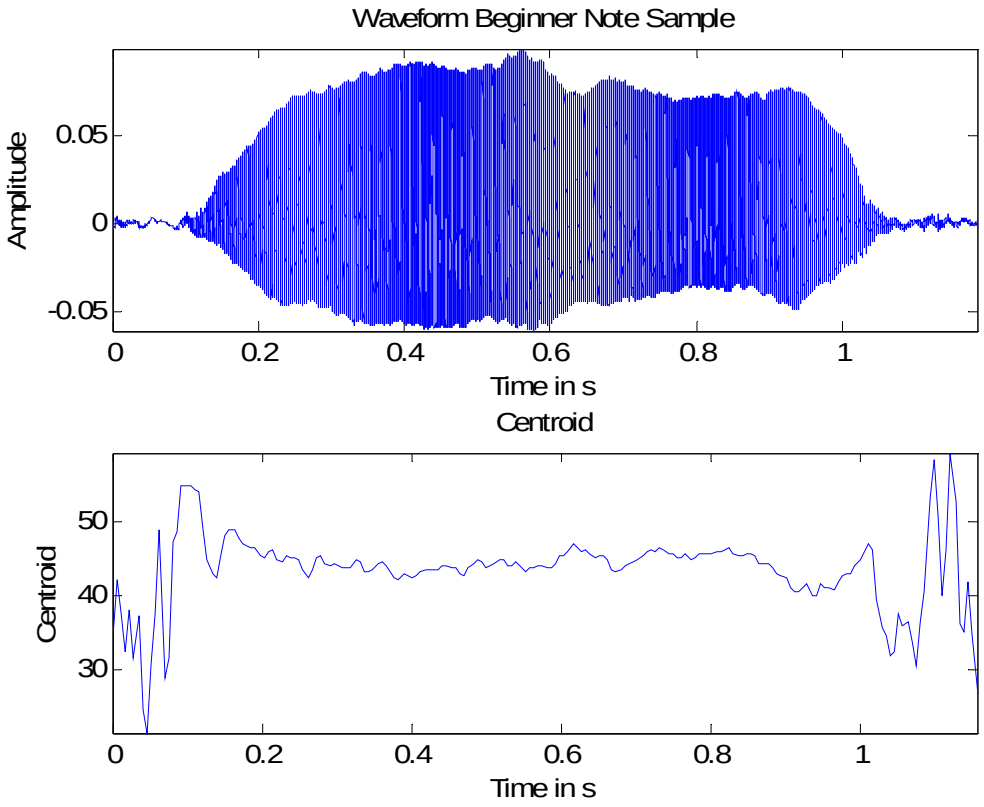

Figure 6.6: Waveform (top) and spectral centroid (bottom) of a reasonable sounding beginner note sample.

To capture the information reflected in these figures, the first order moments have been taken and are presented next. The mean spectral centroid values of beginner and professional standard legato note samples are displayed in Figure 6.7. Although much overlapping of results is visible in Figure 6.7, the beginner note samples tend to have lower centroid mean values. The “brightest” sounding samples in the dataset as reflected by this measure are some of the professional standard legato note samples. Referring to the lower images in Figure 6.5 and Figure 6.6, the spectral centroid mean

for these two sample groups is 41.9 and 43.6 respectively. The mean results for these quite different samples are close. The legato note samples tend to have more consistent steady-state region centroid values, but their onset and offset period readings are much higher. Taking the mean of such values effectively has a smoothing effect, returning similar readings for both player type samples in the dataset, despite their being quite perceptually different. The results obtained for certain samples are detailed in Table 6.2 including, based on the listening tests, the two worst sounding legato note samples, 52 and 71, and the two top sounding beginner note samples, 62 and 65. The three beginner note samples with overall sound quality grade of 1 are also detailed. The centroid variance readings are presented next.

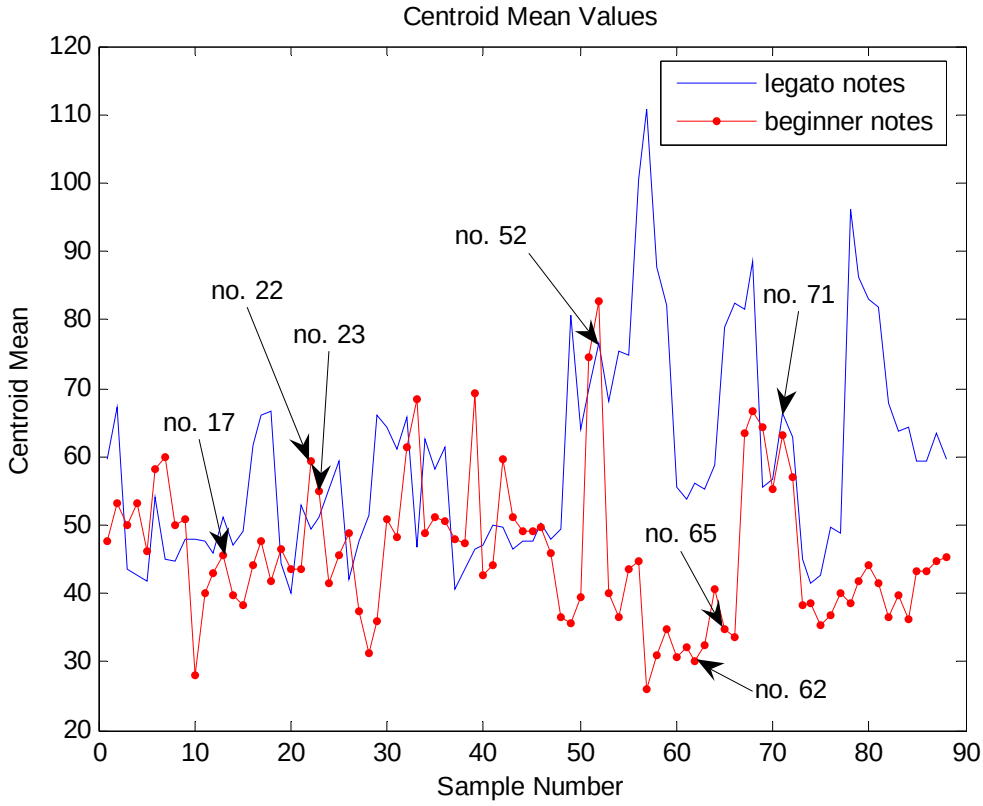

Figure 6.7: Mean centroid values for beginner and professional standard legato note samples.

Sample	*Grade*	*Centroid Mean*	*Beginner /Professional*	*Perceived faults?*	*Additional Information*
Beginner 17	1	47.51	Beginner	NV, BB	Perceived as worst beginner sample
Beginner 22	1	59.26	Beginner	CR, SK, NV, INT, BADS, BADE	Perceived as worst beginner sample
Beginner 23	1	55.04	Beginner	CR, SK, NV, INT, XN, BADS, BADE	Perceived as worst beginner sample
Legato 52	**3.62**	**76.49**	**Beginner**	**none**	**Perceived as worst pro sample**
Beginner 62	**3.95**	**30.12**	**Beginner**	**none**	**Perceived as best beginner sample**
Beginner 65	3.86	34.81	Beginner	BADE	Perceived as 2nd best beginner sample
Legato 71	**3.81**	**66.32**	**Beginner**	**none**	**Perceived as 2nd worst pro sample**

Table 6.2: Information about samples in Figure 6.7.

The spectral centroid variance values are displayed in Figure 6.8 and perform better than the centroid mean values at discriminating between the dataset's beginner and professional standard legato note samples. The beginner note samples tend to have lower centroid variance values than most of the legato note samples, implying less change in sound 'brightness' throughout these samples. The legato professional standard note samples have greater centroid variance values as the onsets and offsets have much higher centroid readings than those in the beginner note samples.

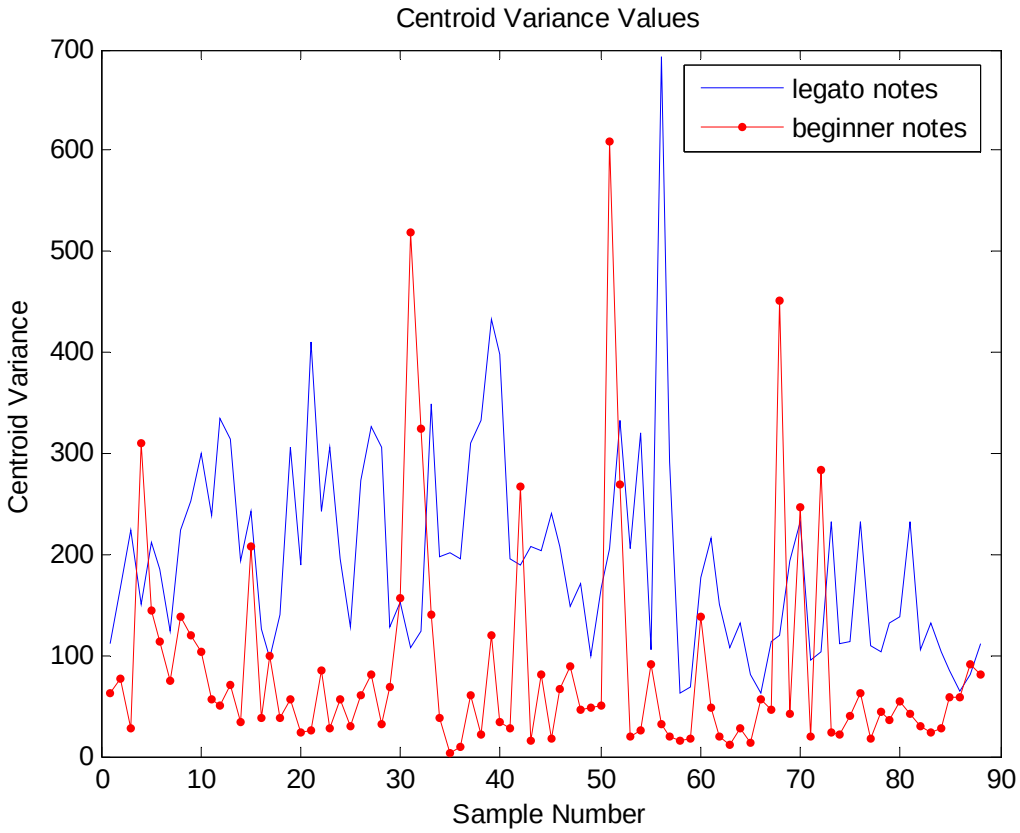

Figure 6.8: Centroid variance values beginner and professional standard legato note samples.

Sample No.	*Grade*	*Centroid Variance*	*Faults perceived?*
Beginner 4	1.19	310.18	CR, SK, NV, INT, SE, BADS, BADE
Beginner 15	1.67	207.22	SK, NV, SE, BADE
Beginner 31	1.43	519.76	CR, NV, BB, SE, BADE
Beginner 32	2.81	324.10	BADS
Beginner 42	1.43	267.27	SK, NV, BADS, SE, BADE
Beginner 51	2.57	609.00	CR, BADS
Beginner 52	2.81	268.22	BADS
Beginner 68	1.71	451.61	SK, NV, BADS
Beginner 70	2.67	246.90	CR,BADE
Beginner 72	2.43	283.45	CR, NV, BADS

Table 6.3: Beginner samples with highest centroid variance values in Figure 6.8.

The sound characteristics of samples with overlapping centroid variance values are of interest. The beginner note samples with the ten highest centroid variance values in

Figure 6.8 are detailed in Table 6.3. These samples have a wide range of sound quality grades. Neither the three worst sounding samples (17, 22 and 23) nor the two top sounding beginner samples (62 and 65) as perceived by the listeners, return centroid variance readings in this overlapping region. The two worst sounding professional standard legato note samples (52 and 71) do not have the lowest centroid variance readings among the legato note samples either. Perceived quality as captured via the listening tests is not reflected by the centroid variance readings.

In Figure 6.9, the centroid skew readings for the dataset are displayed. Although many of these values are very close, the beginner note samples tend to be more negatively skewed than the professional standard legato ones. The range of values obtained for the beginner note samples is much greater than those for the legato note samples. The differences in the centroid skew values do not separate between the two player types in the dataset but the inconsistency of beginner playing is reflected by this measure.

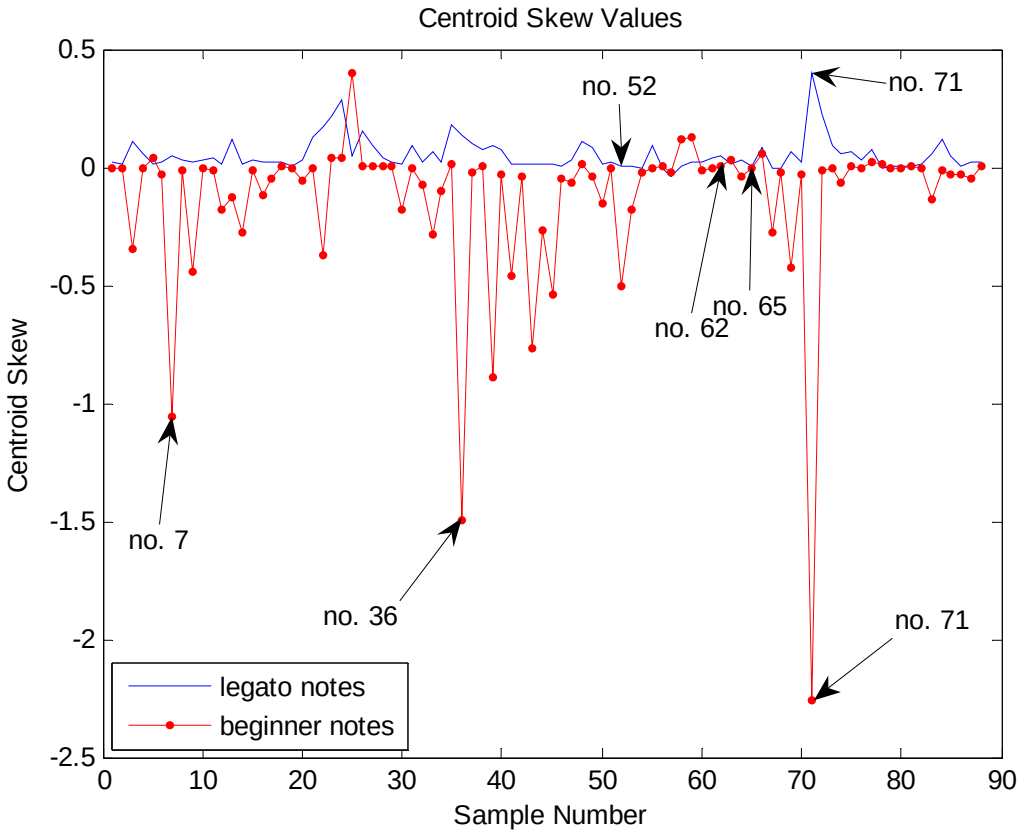

Figure 6.9: Centroid skew values for professional standard legato and beginner note samples.

Sample No.	*Grade*	*Centroid Skew*	*Faults Perceived?*
Beginner 7	1.81	-1.06	CR, NV, BADS, BADE
Beginner 36	2.33	-1.50	INT, SE
Beginner 71	3.05	-2.26	SE, BADE

Table 6.4: Three samples with lowest centroid skew values in Figure 6.9.

The three most negatively skewed centroid values are detailed in Table 6.4 and from the listening tests, these samples have a range of sound quality grades. From the listening tests, legato sample 71 is labelled as being the worst sounding legato note sample. This sample has the highest centroid skew value at 0.40. Legato sample 52, which also has a beginner player label, has a centroid skew value much closer to zero at $8.46*10^{-4}$. The two top sounding beginner samples 62 and 65 also have centroid skew values close to zero. From these results, it is difficult to assign a qualitative expression to these samples based on their centroid skew value only.

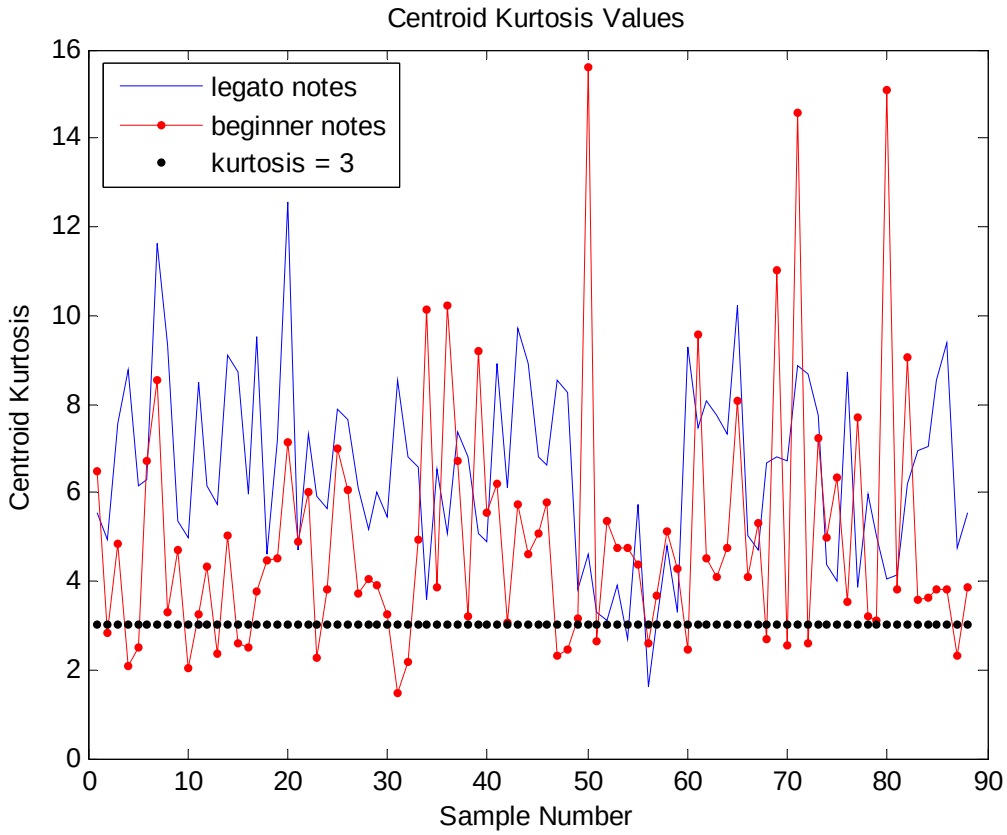

Figure 6.10: Spectral centroid kurtosis values for professional standard legato and beginner note samples.

Next, the spectral centroid kurtosis values for the dataset's samples are illustrated in Figure 6.10. In this figure, a dotted line indicates the normal distribution kurtosis value, 3. Above this line, the results are super-Gaussian and below, sub-Gaussian. Although

much overlap occurs between the centroid kurtosis values for the beginner and professional standard legato notes, most of the samples with the lower centroid kurtosis values are beginner samples. The beginner note samples have kurtosis values which cover a greater range than those for the legato note samples, reflecting less consistency.

Although existing research points to the spectral centroid as being a useful feature for instrumental identification tasks [Eronen01], it is of limited use for reflecting change within the violin's timbre space as represented by the dataset and listening tests used.

6.4 Power Spectral Density Estimation

Power Spectral Densities (PSDs) are power distribution estimates of a signal with respect to frequency [Jayant84]. PSD estimations have proven to be useful in many applications such as signal detection when the signal is hidden in wideband noise [Oppenheim99]. In this section, it is applied to the dataset and its efficacy at representing violin timbre and the qualitative descriptions used is presented.

Many application dependent methods exist for obtaining a PSD estimate. The periodogram is the simplest nonparametric method from which the PSD can be calculated and is based on getting the Fourier transform of fixed length signal segments. It is not regarded as being an accurate method due to bias effects and as a result does not provide a consistent estimate [*ibid.*]. For an illustrated example of why the periodogram, which uses a rectangular window, is not a consistent estimator, refer to [Kay88:66]. The periodogram can be improved by selecting an appropriate windowing function.

In this case, Welch's method, which is also a nonparametric method, uses a Hamming window and provides more consistent results [*ibid.*]. The PSD estimates of a professional standard legato note sample and that of a beginner are displayed in Figure 6.11 and Figure 6.12 respectively.

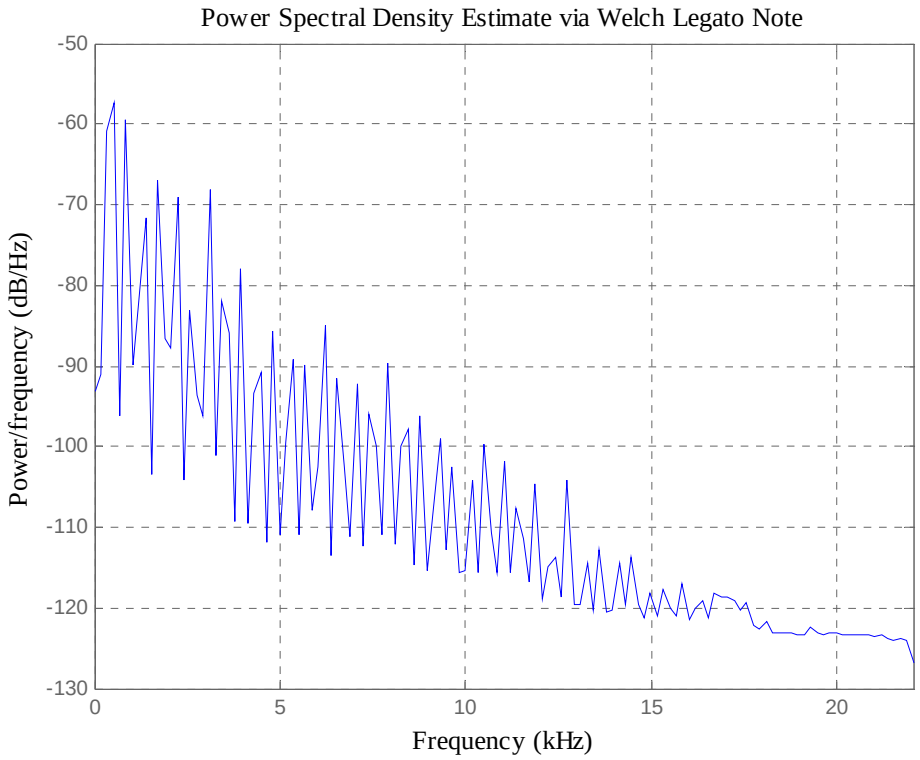

Figure 6.11: Power spectrum via Welch's method of a professional standard legato A440 note.

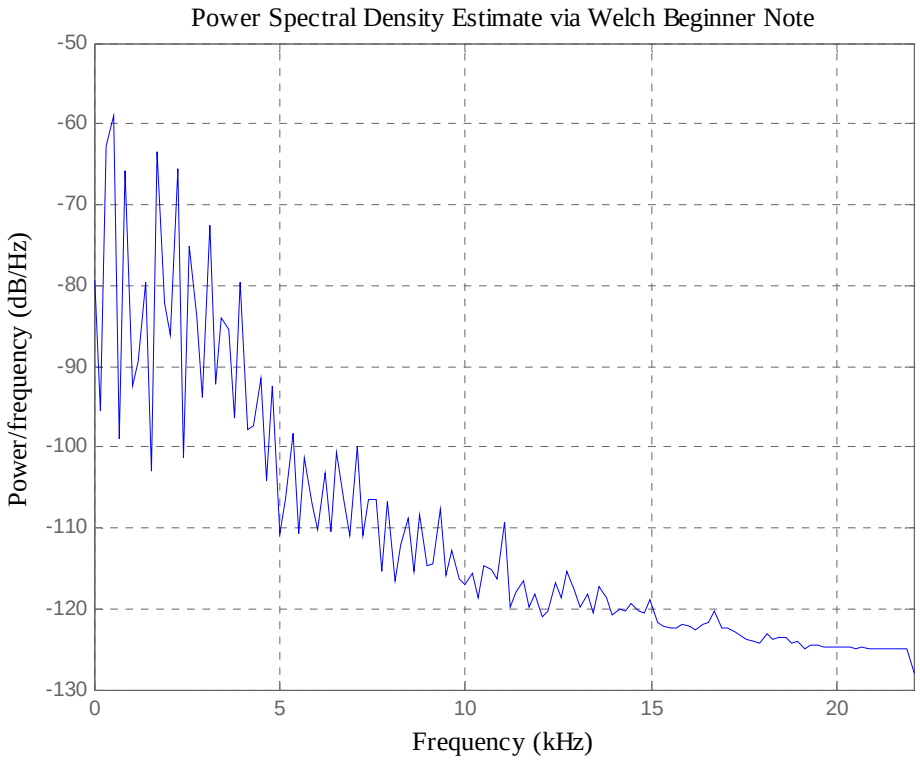

Figure 6.12: Power spectrum via Welch's method of a beginner A440 note.

From the differences visible in Welch's PSD representations of the beginner and professional standard legato note samples, features based on this information can be extracted and used to represent the data. The first PSD based feature, is the mean power of each sample. The results obtained for the dataset's samples are illustrated in Figure 6.13. It is expected that the beginner samples contain comparatively less power and less consistency than the professional standard legato note samples, due to the beginner players having less bow control and therefore not able to get the strength or consistency into the sound. This is not the case, as the results displayed in Figure 6.13 indicate. Apart from the professional standard legato note samples four and five in Figure 6.13, this group is more consistent in its mean PSD readings. There is nothing from the listening tests to indicate why these two samples have such elevated results comparatively to the rest of the professional standard legato note samples. The results obtained for the beginner note samples are much more varied and inconsistent. From the mean PSD readings displayed, overall sound quality is not reflected by this measure and neither are the qualitative expressions used. Applying first order statistics to the PSD data did not return any useful features either and have not been included.

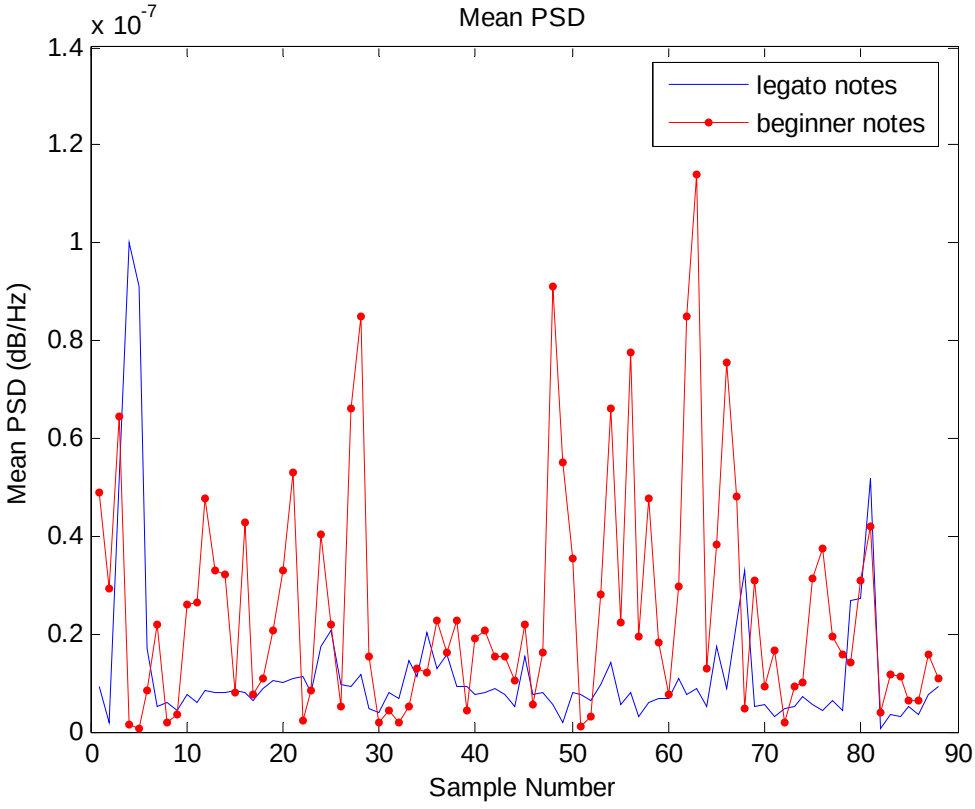

Figure 6.13: Mean power present in each sample based on Welch's PSD.

Recalling the results obtained from certain frequency bins within the 110-190Hz range in the CQT representations prompted looking at energy below the violin's frequency range. The second PSD based feature presented uses the frequency content present below the violin's lowest note. Beginner notes tend not to be as clear sounding due to bowing difficulties, i.e. scraping and crunching. Taking the power associated with the frequencies that fall below the violin's frequency range reflects this information. The mean power present below 190Hz (PSD190) in each sample is displayed in Figure 6.14. From this figure, beginner notes contain more power from the unwanted lower frequencies than the professional standard legato notes. The professional standard legato note samples are much more consistent in the amount of power present associated with the lower frequencies or "playing noise". The mean PSD below the violin's playing range can be used to represent violin timbre and detects beginner from professional standard legato notes in the dataset. The statistical significance of these results has been checked by running a t-test with a 0.01 significance level. The null hypothesis is rejected and a p-value of $1.96*10^{-5}$ is returned.

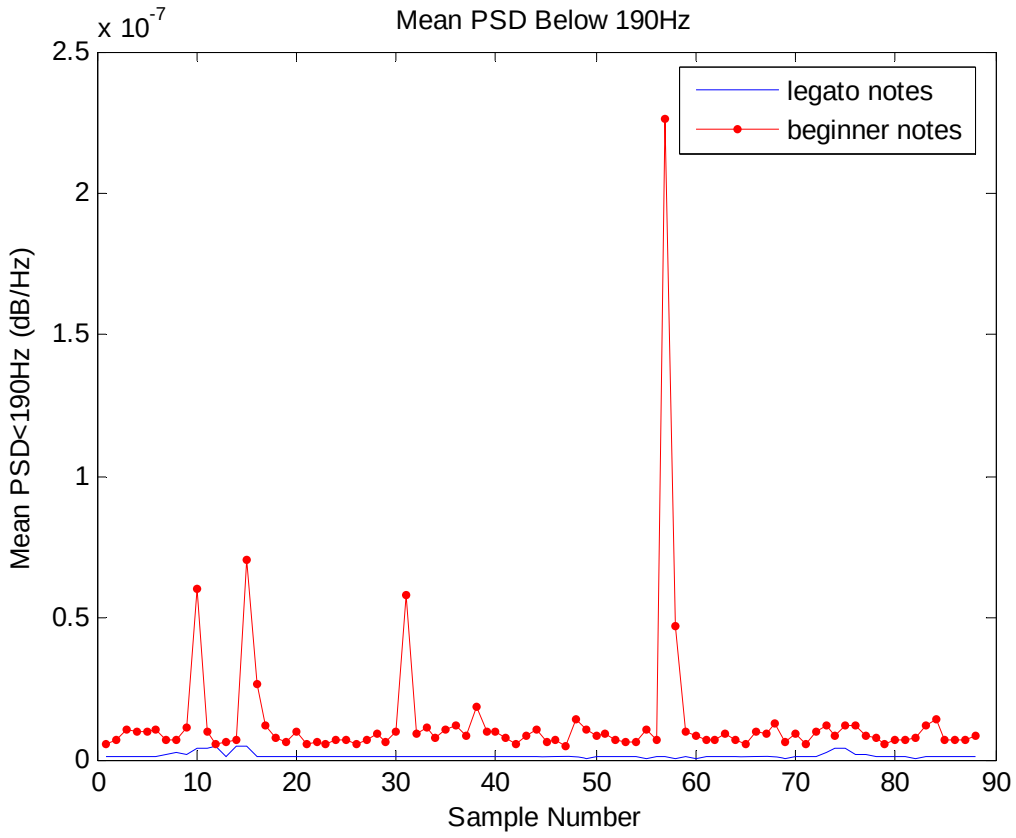

Figure 6.14: Mean PSD present below 190Hz.

Both the CQT and Welch's PSD based features rely on information obtained via the FFT. The PSD190 features take the mean of the power distribution with respect to frequency present up to 190Hz. The CQT, as used in this work, has been set with a start frequency of 110Hz and uses eighth tone spacing. The means of nine specific frequency bins with specific frequency centres below 190Hz serve well at grouping the dataset's samples according to player type. The mean CQT frequency bin numbers one to 39 means have also been taken and is displayed in Figure A4 in Appendix A. The two different player types are grouped accordingly, but with more overlapping samples. The results returned via the bin specific data is more accurate at distinguishing between the beginner and professional standard player note samples. Using the PSD190 to represent the data shows that the beginner note samples in the dataset tend to contain more power below 190Hz. The CQT frequency bin means indicate less frequency content present in the beginner than in the professional standard legato note samples around the selected centre frequencies. Both measures extract FFT based information from within the same frequency range, but one focuses on power distribution with respect to frequency and the other, on frequency content.

From these results, the values obtained from the PSD190 feature differentiate effectively between beginner and professional standard legato notes in the dataset and the difference in the values between the player types has been shown to be statistically significant via a t-test. The mean PSD reflects beginner player inconsistency but not the qualitative expressions used in this thesis.

6.5 Spectral Flatness Measure

The spectral flatness measure (SFM) is defined by the ratio of the geometric mean to the arithmetic mean of the power spectral density components in each critical band [Jayant84]. The steps taken to obtain the SFM readings are shown in Figure 6.15.

In theory, the readings obtained from the SFM give an indication of how noisy or how close to a pure sinusoid a signal is. As the level approaches 1, the signal is closer to white noise. The closer to zero the reading is, the closer the signal is to a pure sinusoid. Following this logic, the SFM can be used for crunch detection. Crunching has been shown to bring in additional unwanted frequencies into the sound, clearly visible in time-frequency representations. White noise by definition contains all frequencies and samples with crunching should have SFM readings above any samples with clear pitch salience but not as elevated as those with white noise. In this section, the usefulness of

the SFM readings for fault detection, specifically crunch detection, as well as the more general case of differentiating between professional standard legato and beginner notes is presented.

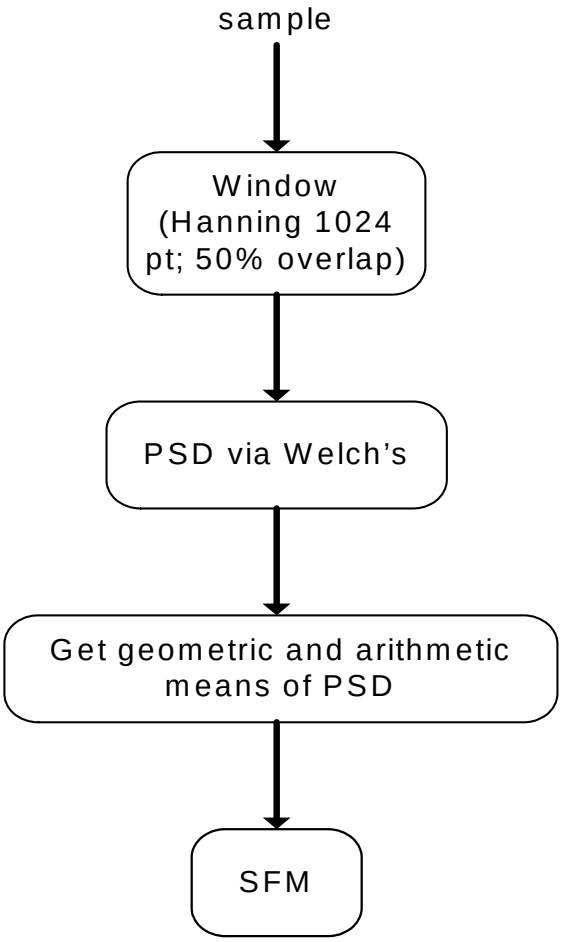

Figure 6.15: Steps taken to obtain the SFM.

The SFM values of a professional standard legato note sample section and that of a beginner note sample are illustrated in Figure 6.16. The beginner note sample in this figure has an overall grade of 2.8 and is reported to have nervousness and a poor start. Although only a section of the legato note sample is shown in Figure 6.16, the complete SFM readings for this sample are plotted in Figure 6.17.

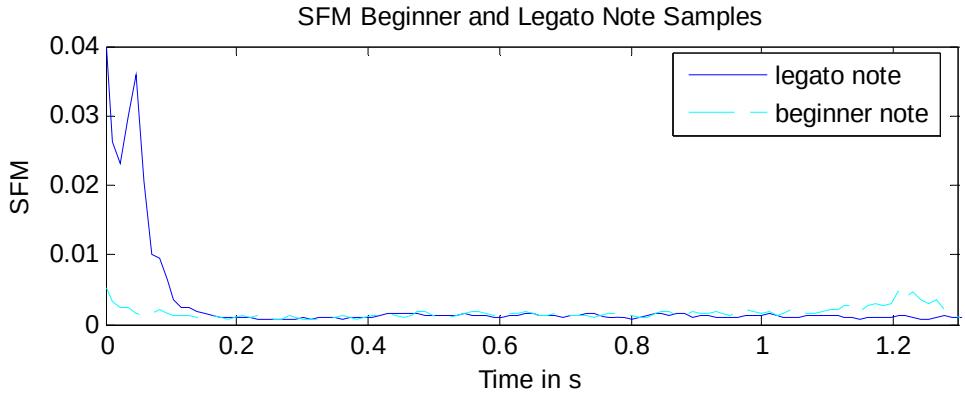

Figure 6.16: SFM values of a professional standard legato note and a beginner note.

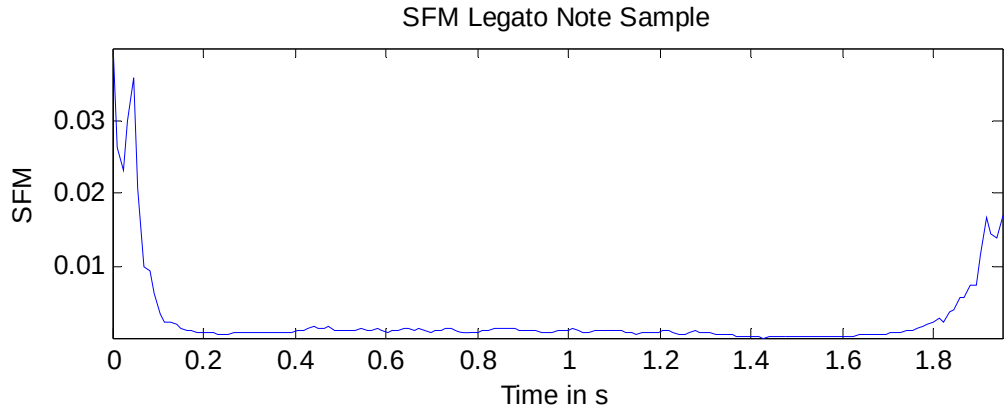

Figure 6.17: SFM of a professional standard legato note sample.

The steepest SFM reading changes occur at the beginning and end of the note and this pattern is repeated throughout the professional standard legato note samples. Reasonable sounding beginner samples start approaching this shape too. The bow pressure applied to the string is not kept the same throughout the duration of a note. The starts and ends of notes require more bow control than the middle section. These are also the regions where beginners typically “crunch”. The most bow pressure changes occur when the player in closest to either the tip (top of bow) or towards the frog (bottom of bow) and this is reflected in the SFM readings. The steady-state section of a professional standard legato note, where pressure is applied more consistently, the SFM readings flatten out and approach zero, reflecting pitch salience.

From the professional standard legato note SFM results, the attack-steady-state-decay sections become discernable. The SFM readings for all the professional standard legato note samples follow a similar shape to that displayed in Figure 6.17.

Comparatively, the beginner note samples' SFM values are less smooth as the note progresses and the readings are not as elevated at the starts and ends of notes. The SFM readings for poor beginner sounds are less smooth and are unreliable for observing the attack-steady-state-decay regions, as illustrated by the samples dispalyed in Figure 6.18 and in Figure 6.19. From a beginner note's SFM, it is more difficult to judge where the attack ends and at which point a steady-state is established. This is often due to the lack of a consistent steady-state being established and maintained which is a result of a poor attack.

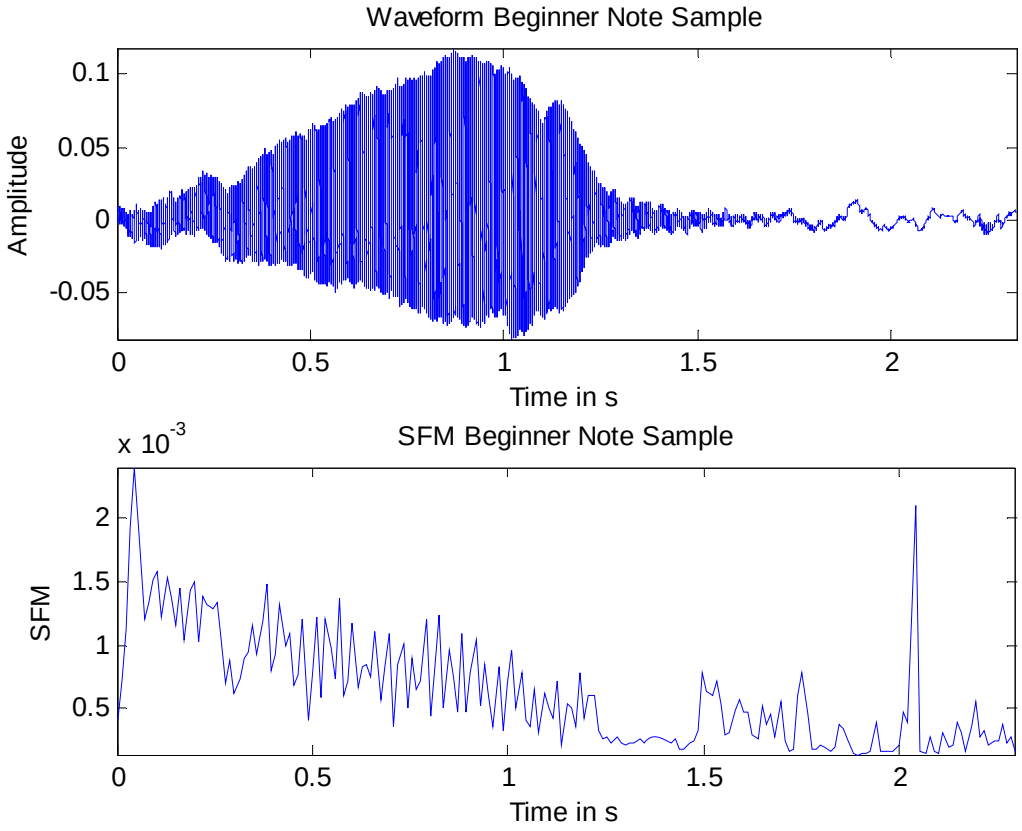

Figure 6.18: Beginner sample waveform (top) and SFM readings (bottom).

In Figure 6.18, the waveform and SFM values of a beginner note sample are illustrated. From the listening tests, this sample has an overall sound quality grade of 2.67 and contains several playing faults: crunching, nervousness, bow bouncing, poor start and end to note. The SFM readings show that a clean attack is not achieved and a brief almost steady-state section occurs from about 1.3s to 1.5s. As the player stops the bow at the end of the note, a short crunch is audible. This crunch coincides with the sharp peak which is visible in the right hand side of the lower image in Figure 6.18.

The waveform and SFM representations of another beginner note sample are displayed in Figure 6.19. From the listening tests, this sample has an overall grade of 1 and contains multiple faults: skating, nervousness and bow bouncing. The presence of these playing faults results in no consistency or steady-state being established and is reflected in the unevenness of the SFM readings. Next, the relationship between SFM reading and attack is presented.

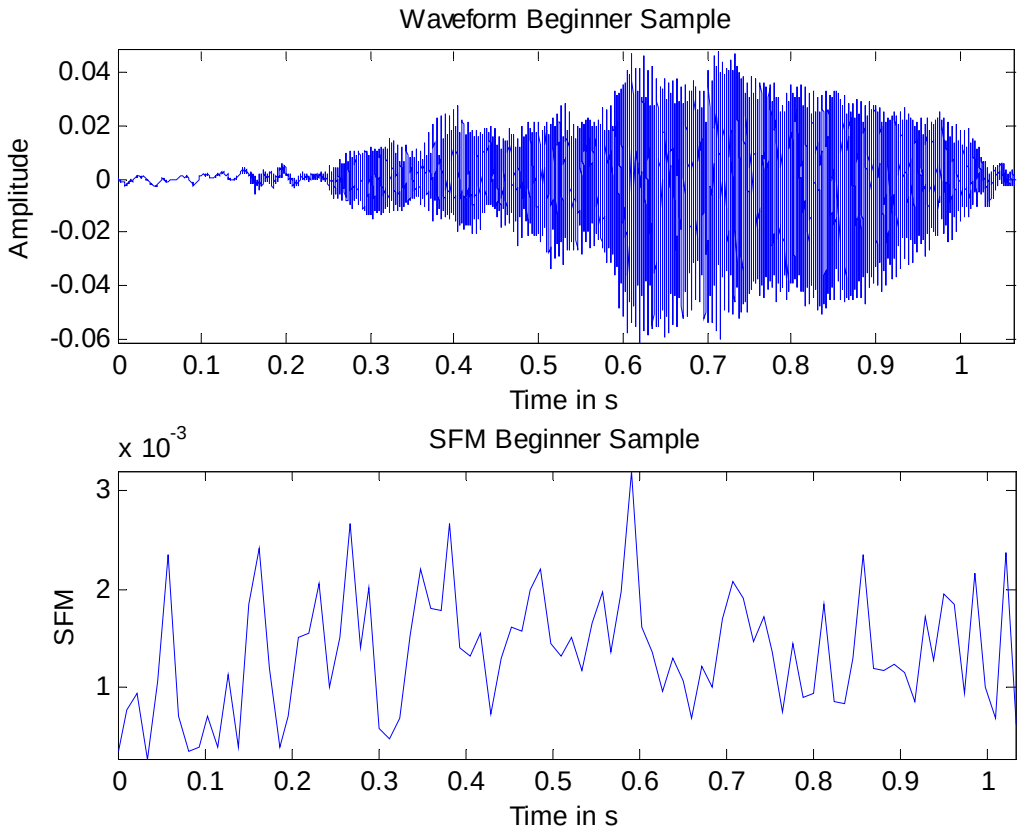

Figure 6.19: Waveform (top) and SFM (bottom) of a beginner note sample.

Starting with a fast bow stroke, the waveform and SFM readings of a fast bow stroke sample are illustrated in Figure 6.20. The bow stroke takes more time to settle down towards a steady-state than a legato note attack. This reflects the force applied to the string and the faster bow stroke causes greater string fluctuations which accounts for the jagged SFM readings. This figure has been included to show that SFM readings reflect not only pitch salience but overall playing too.

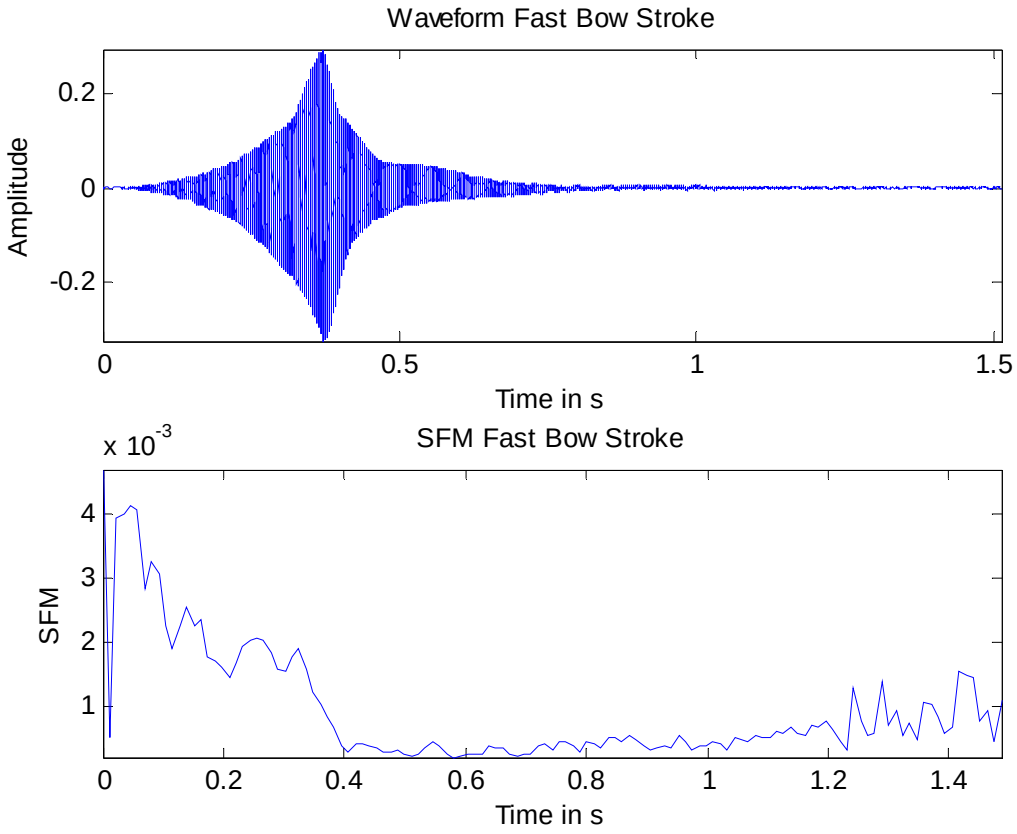

Figure 6.20: Waveform (top) and SFM (bottom) of a fast bow stroke.

The effect forcing has on the waveform and SFM is displayed in Figure 6.21. Forcing the sound to breaking point causes spikes to appear in the waveform amplitude and a sudden increase in the SFM readings as illustrated in this figure. The sound is forced then released many times, coinciding as the peaks appear and subside the lower image in Figure 6.21. The forced note samples are more extreme examples, but are useful in showing the effect forcing the sound has on the SFM readings which increase sharply and remain elevated and unsteady due to the extra frequencies present in the sound.

The more confident and clear sounding a legato bow stroke is, the lower its SFM reading. It will never reach zero as a violin is not capable of producing a pure sinusoid. The SFM has potential as a feature for monitoring the overall sound quality as the note progresses. To obtain SFM based features, first order statistics have been applied to the SFM readings and are presented next.

The SFM mean (SFMM) values for each sample are plotted in Figure 6.22. These results discriminate well between the two player groups in the dataset making it a useful feature for describing violin timbre within the dataset. The professional standard legato

note samples mostly have higher SFMM readings than the beginner samples do. The results do not reflect what was initially expected as a pure sinusoid returns a SFM reading of 0 and white noise, 1. The cause of the more elevated SFMM results for the legato notes has to do with the attack and offset SFM values. Although the attack is clean and accurate in the dataset's legato note samples, the SFM readings are always briefly much higher at the outset before falling, remaining steady and then rising again. This is not observed to the same extent in the beginner note samples, indicating that there is a problem with the attack and consequent establishment of the note. Beginner players tend not to have clean attacks and a steady-state is not always established as illustrated in Figure 6.18 and in Figure 6.19.

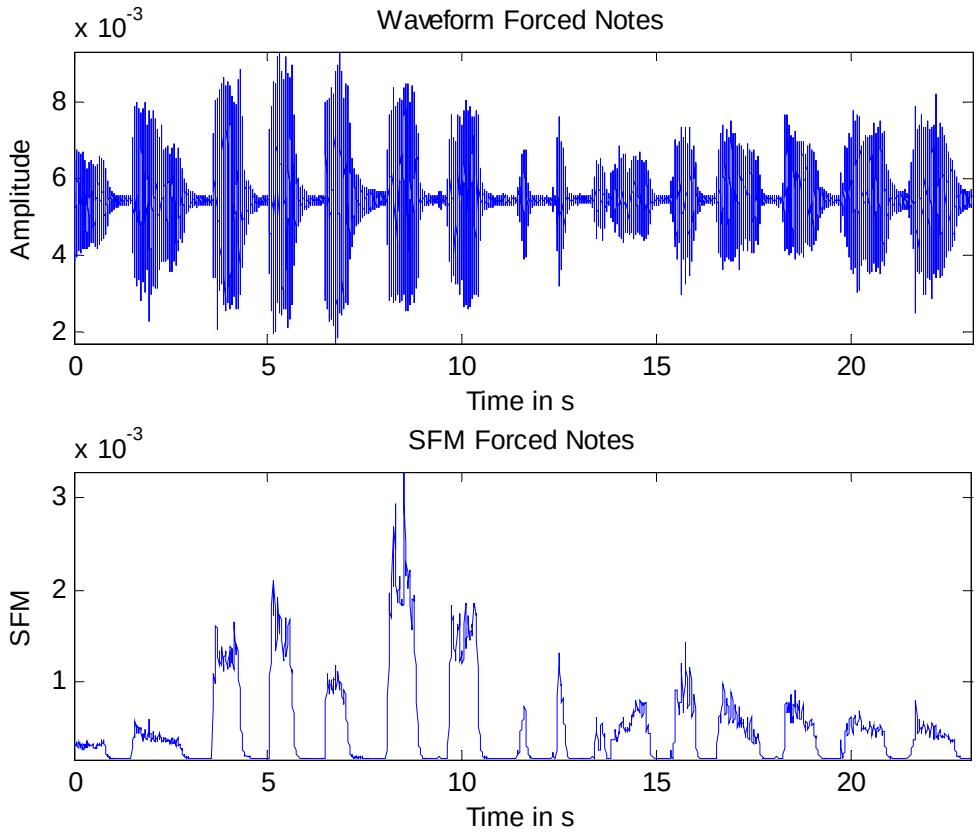

Figure 6.21: Forced crunching sample waveform (top) and SFM readings (bottom).

The two best sounding beginner note samples, 62 and 65, return low SFMM values. The two worst sounding legato note samples, 52 and 71, do not have low SFMM values comparatively. Having stated this, the poor sounding legato note samples still have overall quality grades that are higher than the beginner note samples. This makes linking used qualitative expressions to the samples according to the SFMM value

difficult. As a measure though, it allows the two sample types in the dataset to be accurately grouped based on player.

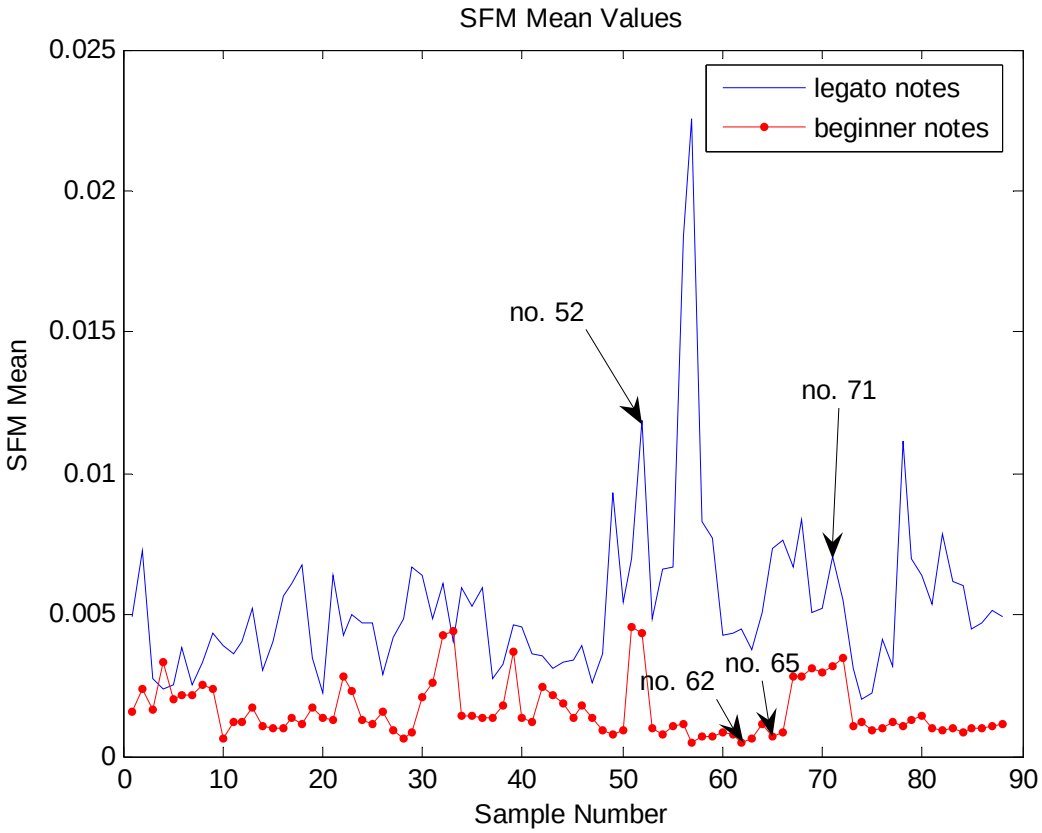

Figure 6.22: SFM mean readings for professional standard legato and beginner note samples.

Through observing the SFM result patterns for beginner and professional standard legato note samples, the range of SFM values returned is different for both player groups. This is reflected in the SFM variance (SFMV) readings which are displayed in Figure 6.23 where the dataset's samples are separated into two distinct groups, reflecting the different player types. The beginner notes have lower SFMV readings than the professional standard legato ones do, making the SFMV a good discriminator between these two player types. An explanation for these results is found by observing the SFM readings of all these samples. Professional standard legato note samples have low, smoother steady state sections but the starts and ends of all these samples have much higher SFM readings as illustrated by the sample in Figure 6.17. Beginner samples, although tending to return much more uneven SFM readings throughout the duration of the note, return SFM readings which are neither as high nor as low as those provided by the professional standard legato notes. The difference between these results

is statistically significant as the null hypothesis of a t-test with a 0.01 significance level is rejected. A p-value of $5.3*10^{-31}$ is returned. The performance of the SFM variance value at discriminating between beginner and professional standard legato note samples is not due to sample variability.

The two beginner note samples, 62 and 65, perceived to be the best sounding through the listening tests have low SFMV values and the two legato note samples, 52 and 71, perceived to be worst sounding have higher values. Although the SFMV separates effectively between the two player groups present in the dataset, a clear relationship is not established reflecting the expressions used in this text.

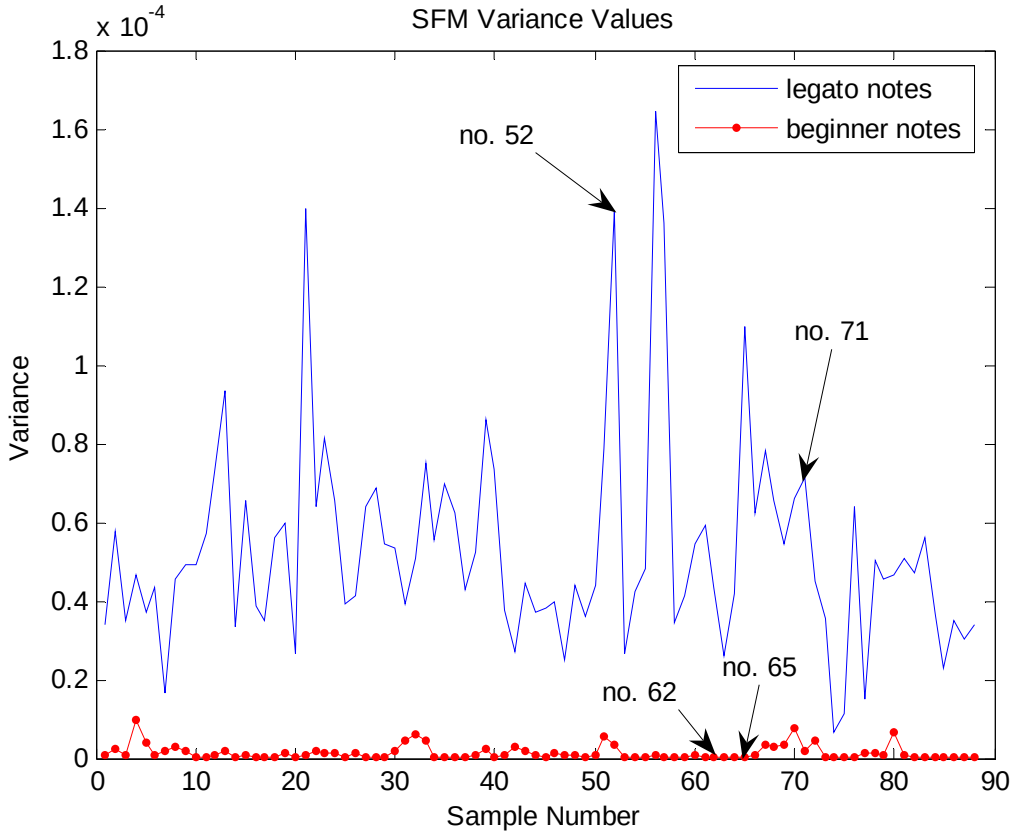

Figure 6.23: SFM variance readings for professional standard legato and beginner note samples.

The SFM skew (SFMS) readings display many overlapping results for the different player types in the dataset which makes it a much less useful feature for representing violin timbre compared to the previous measures. For this reason, a figure containing these results has not been included.

In Figure 6.24, the SFM kurtosis (SFMK) readings for the dataset samples are shown. All samples return very peaky results as all have kurtosis values greater than 3.

From these results the beginner note samples in the dataset tend to have lower SFMK readings then the professional standard legato ones. The two different player groups are not completely separable when represented by this measure, but an underlying pattern is present. Given the extent of overlap between these two groups, the SFMK is not the most suitable feature for distinguishing one player group from the other when applied to this dataset due to the number of overlapping readings.

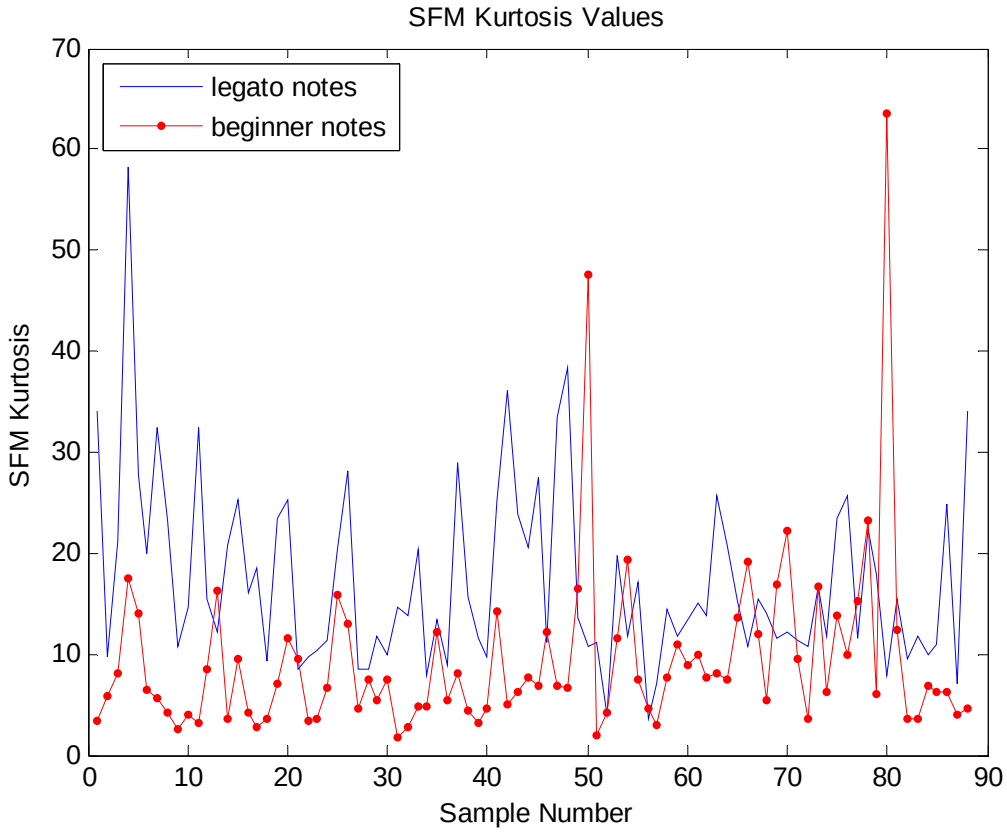

Figure 6.24: SFM kurtosis readings.

Although the SFMM and SFMV values accurately group the beginner from the professional standard legato note samples in this dataset, another application for the SFM is as a bow change detector. As the bow changes direction, the sound quality is altered. The amount by which it changes is a reflection of playing technique as well as musical style and bow stroke used. Smooth bow changes are what players strive for in legato bowing. Regardless of how smooth a bow change may be, the SFM reading is a sensitive measure to this change and rises a little creating a small peak in the plotted readings. Figure 6.25 shows the waveform and the SFM readings of a group of legato 16th notes. The peaks in the SFM readings line up with the exact point of bow change.

This figure illustrates how the SFM values indicate bow change onset detection for violin playing and can be extended to all bowed stringed instruments.

Comparing the smooth bow changes displayed in Figure 6.25 with that of a beginner player illustrated in Figure 6.26, shows that a beginner player's bow change is not as clean as that of a professional standard player. The beginner sample shown in Figure 6.26 does not quickly reach a steady-state nor does it return to one fast enough after the bow change. This results in jagged SFM values.

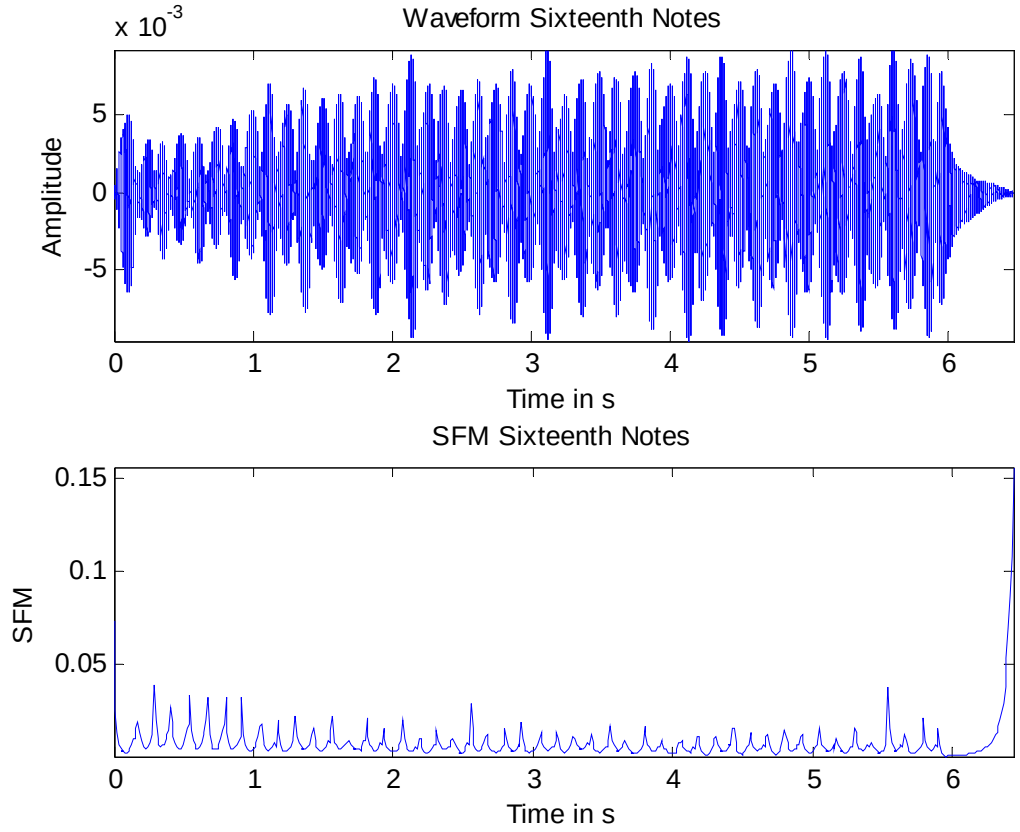

Figure 6.25: Waveform (top) and SFM readings (bottom) of a sample of bowed 16th notes.

The SFM detects bow changes but note changes within a same bow stroke are not captured. In Figure 6.27, the waveform, spectrogram and SFM of note changes within the same bow stroke are displayed. The spectrogram is included to show the note changes. The notes played in this sample and G2 A3 G2 A3 G2 A3. The SFM does not detect note changes within the same bow stroke and consequently can only be used in one type of onset detection. This is logical as it is a power spectrum energy based measure and power spectrum energy changes are far greater when the bow changes direction than when a finger (note) changes. Additional frequencies are present when

the bow changes directions that do not occur when a finger is changed and are reflected in time-frequency representations. The presence of these extra frequencies causes the SFM readings to increase when the bow changes direction but not when the note changes within a bow stroke.

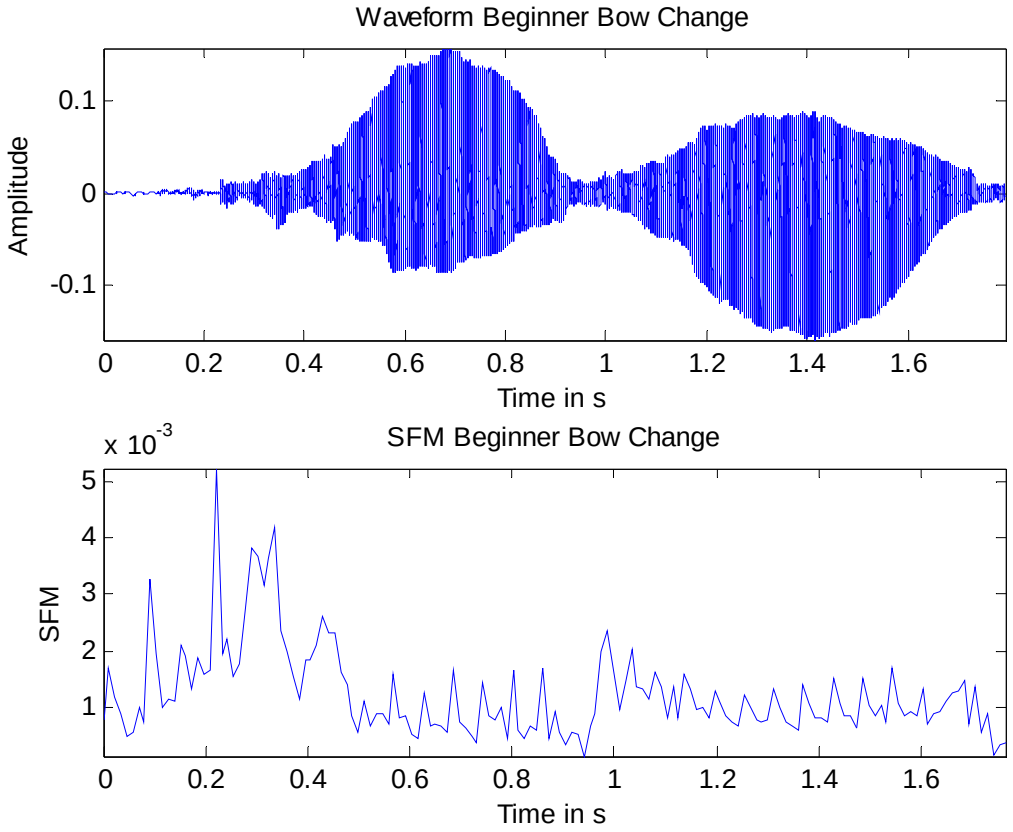

Figure 6.26: Example of beginner bow change waveform (top) and SFM readings (bottom).

The SFM is a useful measure for representing real violin sound from which many features have been obtained. It can be used to check the sound quality of a legato note as it progresses and for detecting bow change onsets. The results obtained from applying first order statistics to the SFM readings have shown that the SFMV values are the most effective at grouping separately the professionals standard legato note samples from the beginner ones. The SFMM values group correctly the dataset samples according to player, with some overlapping samples. More overlapping samples are visible when the SFMK values are plotted, but the underlying pattern is of interest as the beginner note samples tend to have lower values than most of the professional standard legato ones.

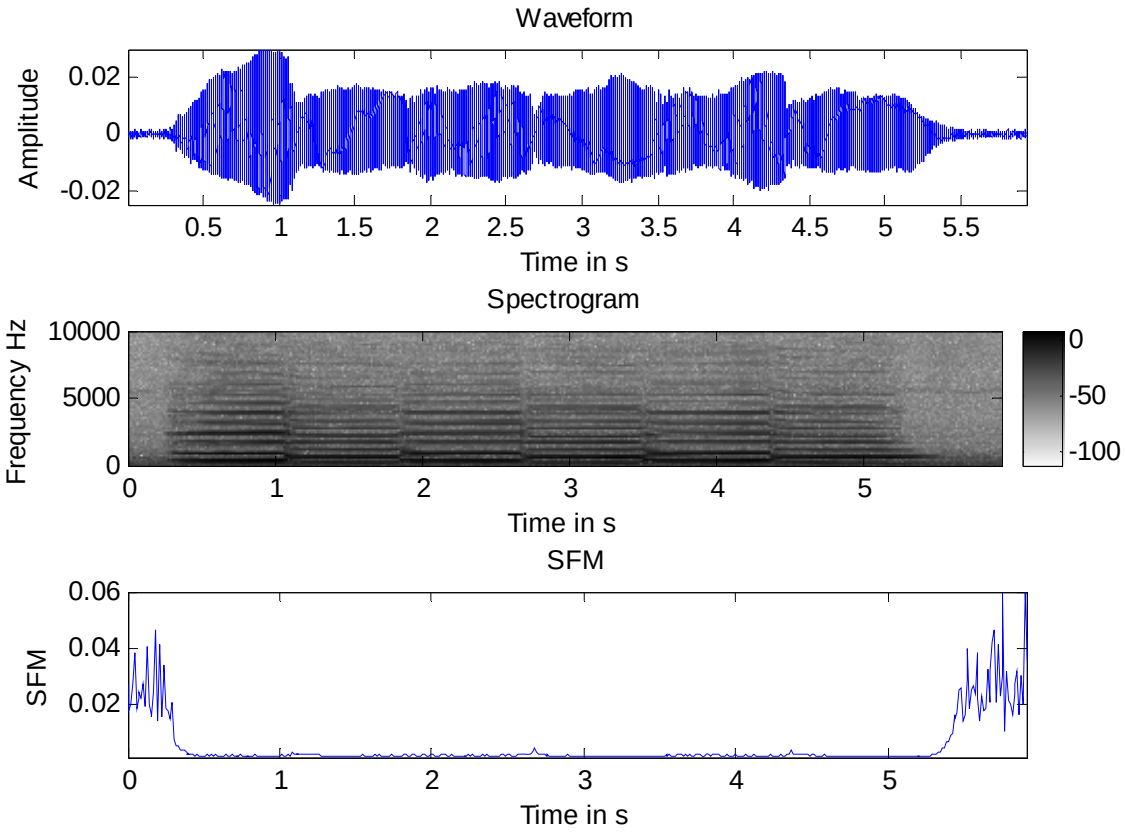

Figure 6.27: Note changes in the same bow stroke waveform (top), spectrogram (middle), SFM (bottom).

6.6 Spectral Contrast Measure

Jiang *et al.* put forward a filter based spectral contrast measure (SCM) feature in [Jiang02]. West *et al.* [West04] have also successfully used this feature in the automatic classification tasks of musical signals. As a feature it represents the spectral characteristics of music samples via the relative spectral distribution. It is selected as a violin timbre feature as it has been reported to be designed to give better results than the Mel Frequency Cepstral Coefficients (MFCCs) [*ibid.*]. It does this by considering the strength of spectral peaks and spectral valleys in each sub-band separately, reflecting the distribution of harmonic and non-harmonic components in a sample. The SCM is considered for its potential as a violin timbre quality detector in this section. The steps involved in extracting this feature are detailed in the afore mentioned papers and the steps applied are given in Figure 6.28. Jiang *et al.* used the KLT for the optimal reduction of the covariance between elements of the spectral contrast feature vector.

West *et al.* used the DCT to eliminate covariance in highly correlated data citing [Potkonjak97] for their choice.

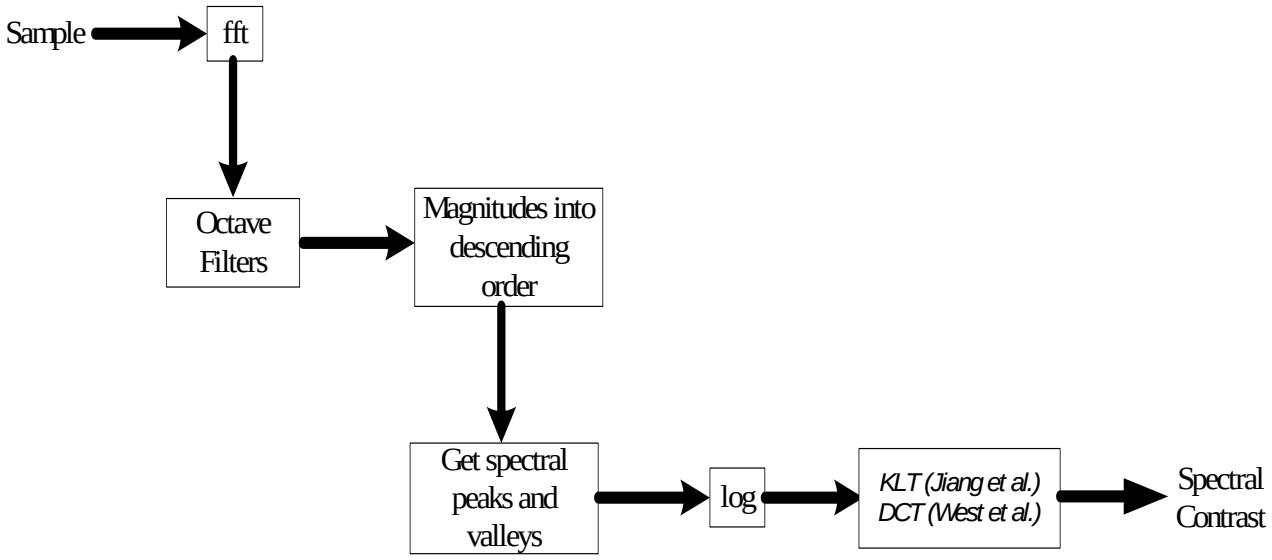

Figure 6.28: Spectral contrast steps used by Jiang *et al.* and West *et al.*

The data is sampled at 44.1 kHz. Eight filters are used to divide the frequency domain into sub-bands. The frequency ranges for the filters used are: 0-200Hz, 200-400Hz, 400-800Hz, 800-1600Hz, 1600-3200Hz, 3200-6400Hz, 6400-12800Hz, and 12800-25600Hz. The spectral magnitudes of each band are put into descending order according to magnitude. Equation 6.2 and Equation 6.3 are then applied to obtain estimates of the spectral peaks and spectral valleys [Jiang02]. In these equations, *i* is the index, *N* window size and α, the neighbourhood factor:

$$P_p = \ln\left(\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x_{p,i}\right) \tag{6.2}$$

$$V_v = \ln\left(\frac{1}{\alpha N}\sum_{i=1}^{\alpha N} x_{v,N-i+1}\right) \tag{6.3}$$

The inclusion of α, a neighbourhood factor, stabilises the feature by averaging the peaks and valleys within a small region. Jiang *et al.* found that varying α between 0.02 and 0.2 did not influence the performance significantly. In their implementation, α=0.02 was used. As a starting point in this work, α=0.02 was taken. Values ranging from α=0.01 to 0.25 in increments of 0.01 were also tried as well as values of α=0.3 to 0.9 in steps of 0.1. The spectral contrast of each sub-band is given by the difference between the peaks and valleys, Equation 6.4:

$$SpectralContrast_{sub-band} = P_{sub-band} - V_{sub-band} \quad (6.4)$$

A high SCM reading implies a signal having high peaks, low valleys and strong localized harmonic content. A low SCM reading represents a signal with less harmonic content. Results from all filters are plotted in Figure 6.29. The filter range represented is indicated on top of each image where the beginner samples are indicated by the dotted line and the professional standard legato ones, by the solid one.

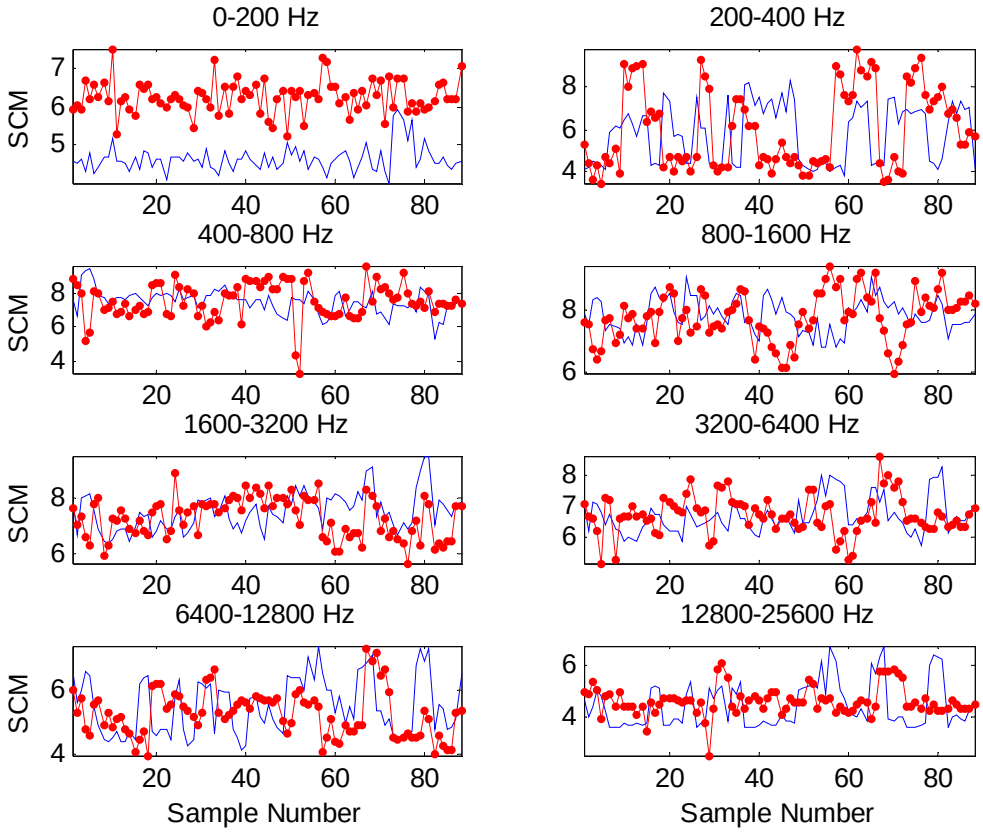

Figure 6.29: Spectral contrast results for all filters.

The SCM below 190Hz reflects similar information to that obtained by the PSD below 190Hz. The SCM, like the PSD and CQT features, is FFT based and uses both the harmonic and non-harmonic components. Although focusing on the frequency range below 190Hz, the spectral content is represented as the difference between the harmonic and non-harmonic components, as opposed to the power distribution and harmonic only content in the PSD and CQT based features. When applying the SCM to the dataset, the most interesting results are returned by the first filter, which is the frequency range below 200Hz. The lowest note on a violin tuned to A440 is the open G string which is associated with a frequency reading of approximately 196Hz. All the results for this

filter from the SCM with α ranging from 0.01 to 0.9 give very good separation between the professional standard legato and the beginner note samples in the dataset. Given that this range includes only the violin's lowest note and below, the content of this region is important. The statistical significance of these results has been verified by applying a t-test with a 0.01 significance level. The null hypothesis is rejected and a p-value of $1.67*10^{-64}$ has been returned.

To investigate this further, a series of filters focusing within this frequency range have been applied. The images displaying the spectral content below 190Hz, 120Hz, 85Hz and 75Hz as indicated on top of each image are displayed in Figure 6.30. A value of α=0.2 worked the best in terms of separating the two groups for the highest number of filters. The filters applied were <190Hz, <120Hz, <85Hz, <75Hz. Excellent separation between the two sample groups is visible until 90Hz at which point the groups are much closer together but the pattern remains discernable. The filters below 90Hz display overlapping frequency content levels. The frequency content levels for both sample groups are comparable indicating minimum or acceptable "noise" levels due to playing.

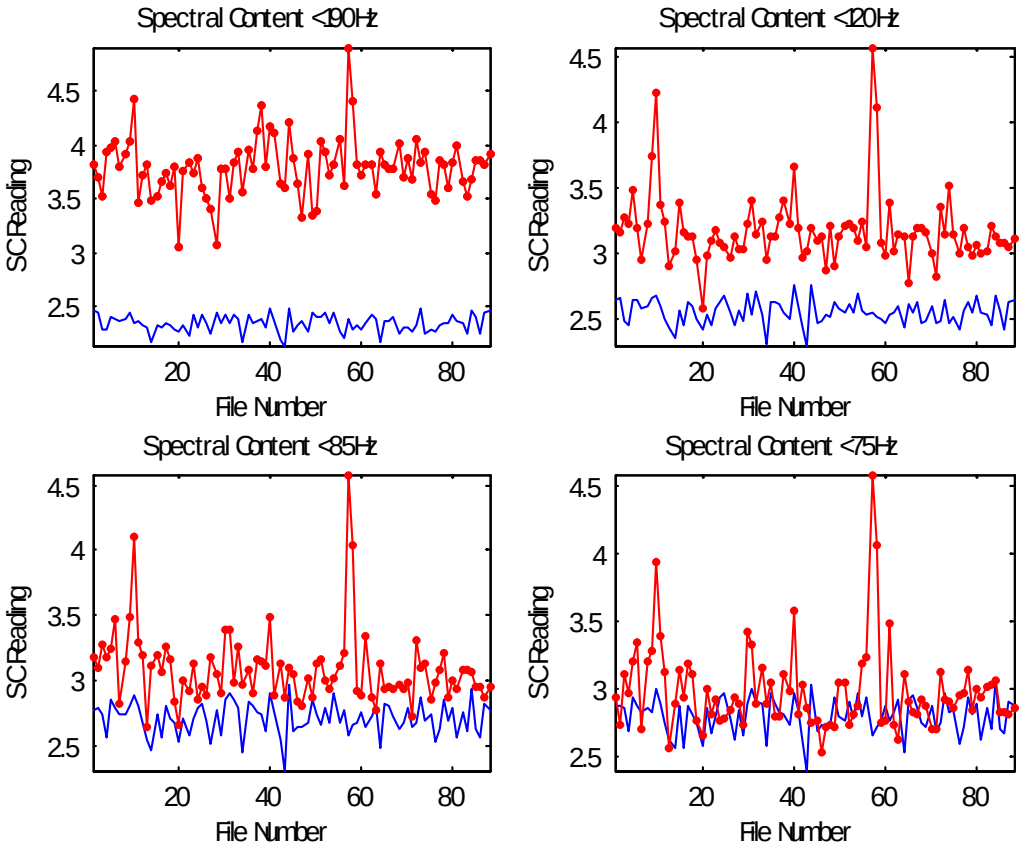

Figure 6.30: Spectral content <190Hz, <120Hz, <85Hz, <75Hz obtained via SCM.

The SCM provides useful representation of the violin note samples in the dataset. The results provided by filters below 200Hz are of particular use towards attaining the research aims.

6.7 Summary

Multiple spectral features have been shown in this chapter, some returning more useful results than others in relation to the research aims. The efficacy of features based on specific CQT frequency bin information, spectral flux, spectral centroid, PSD, SFM and SCM for representing violin timbre quality have been presented. Some of these features worked best on complete note samples whereas others, such as the SFM and spectral centroid are more useful when applied to windowed signals. Among the most successful spectral features for representing violin timbre in the dataset are nine specific CQT frequency bins, the spectral centroid mean, the PSD below 190Hz, SFMM, SFMV and SCM between 0 and 200Hz. These features all group the beginner and professional standard legato note samples correctly according to player type. The respective differences between the players' feature values have been shown to be statistically significant.

From the analyses presented so far, it is possible to extract features capable of discriminating between beginner and professional standard player legato notes in the dataset. From these spectral features, identifying individual playing faults has not been shown to be evident, nor has defining perceptual correlates for the violin timbre. When used in conjunction with other features from different domains, these detection results are expected to improve. Suitable features from the cepstral domain and how they are used to represent violin timbre is presented next in Chapter 7.

7 Cepstral Analysis

Temporal and spectral analyses described in the previous chapters have provided some useful descriptors for representing violin timbre within the context of this research. This chapter presents the efficacy of cepstral features at depicting violin timbre. Representing the signal through its cepstrum allows information about its periodicity to be obtained and is a standard technique for extracting pitch from speech signals [Youngberg79] and instrumental sounds [Klapuri04]. Other cepstral analysis applications include seismology, biomedical signals, and sonar signals [Oppenheim89]. The most effective cepstral analyses in the audio domain are based on the real and Mel cepstra [Deller00]. Features based on these cepstra are detailed in the following sections for their effectiveness at depicting violin sound quality and playing faults.

7.1 Real Cepstral Features

Cepstral features including statistical analysis of real cepstral coefficients (RCCs) and individual cepstral coefficients are presented. The RCCs provide a convenient way of modeling spectral information and are obtained by following the steps shown in Figure 7.1.

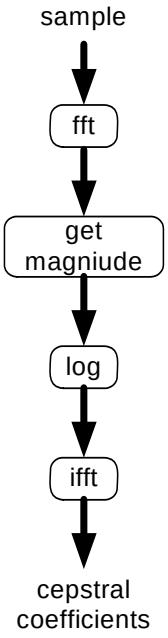

Figure 7.1: Steps for obtaining real cepstral coefficients.

RCCs have been successfully applied in instrument recognition tasks [Eronen01]. Their ability to provide suitable features to characterise the violin's timbre space within the context of the research aims is sought. The ability of each feature to separate the beginner note samples from the professional standard legato ones in the dataset is noted. Where small numbers of overlapping samples occur, they are identified and how they have been perceived by the average listener is included.

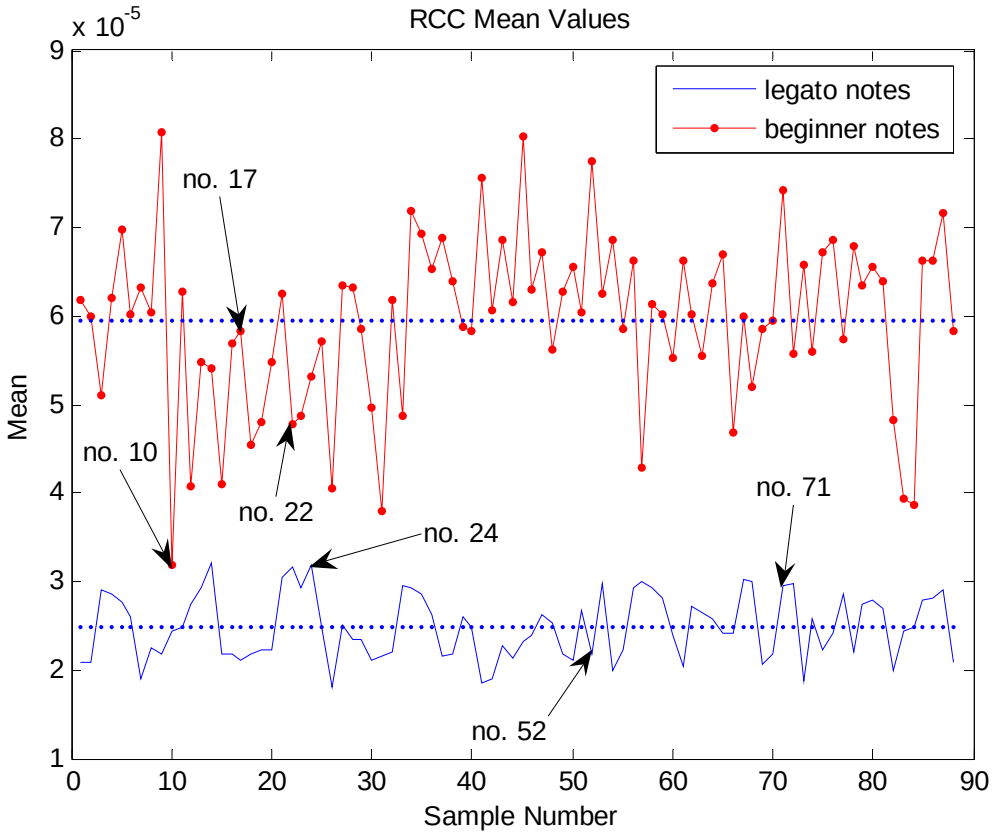

Figure 7.2: Real cepstral coefficients mean readings professional standard legato and beginner note samples.

First order statistics are applied to the RCCs and the RCC mean (RCCM) results are displayed in Figure 7.2. Through representing the dataset by this measure, two distinct sample groups are visible in this figure. Higher RCCM values for the beginner note samples in red are observed. These samples are much less consistent and have a wider value range than the legato professional standard ones which tend to have lower RCCM values, shown in blue. The mean RCCM for the professional standard legato note samples is $2.49*10^{-5}$ and $5.94*10^{-5}$ for the beginner note samples and are marked in Figure 7.2 by the dotted black lines. By running a t-test with a 0.01 significance level on the difference between the RCCM values for the beginner and professional standard

legato note samples, the results are statistically significant as the null hypothesis is rejected and a p-value of $4.6*10^{-57}$ is returned.

In Figure 7.2, the lowest beginner sample, sample ten, is of particular interest. On initial inspection of this figure, it was thought that this sample could be one of the better sounding, fault free beginner samples. From the listening tests, this beginner sample contains three faults: nervousness, bow bouncing and a poor start to the note and has an overall grade of 2.67 out of 6. Beginner samples 17, 22 and 23 are recorded as having the poorest overall sound quality, grade 1, yet these three samples do not have the highest nor the lowest RCCM values. The beginner sample returning the highest RCCM is sample nine, which has an overall grade of 2.19 and reported crunching and a poor finish. Other samples which are of interest are those whose RCCM readings are close. One such overlapping legato sample, number 24, has been perceived as a beginner note rather than a professional note. Listening to this sample, a slight bow bounce can be heard towards the end of the note. Its overall sound quality grade of 4.29 is higher than all the beginner note samples' grades. From these observations, the RCCM remains an effective discriminator between the two different player groups, but the qualitative expressions used in this text are difficult to associate with a specific RCCM value. To gain further insight into the relationship between the RCCM reading and violin timbre, the forced note samples are displayed in Figure 7.3.

The forced note samples provide results that are comparatively scattered to those obtained for either of the beginner and professional standard legato note samples. Exaggerated crunching or forcing does not push the RCCM readings clearly in any one particular direction as has been observed in, for example, the time domain waveform amplitude mean readings in Figure 5.4. A sound quality range exists within which the RCCM value is useful to differentiate between the dataset's beginner and professional standard legato note samples. It can be used as a more general feature in distinguishing effectively between the two different player groups in the dataset. The RCC variance (RCCV) values for the dataset are presented next.

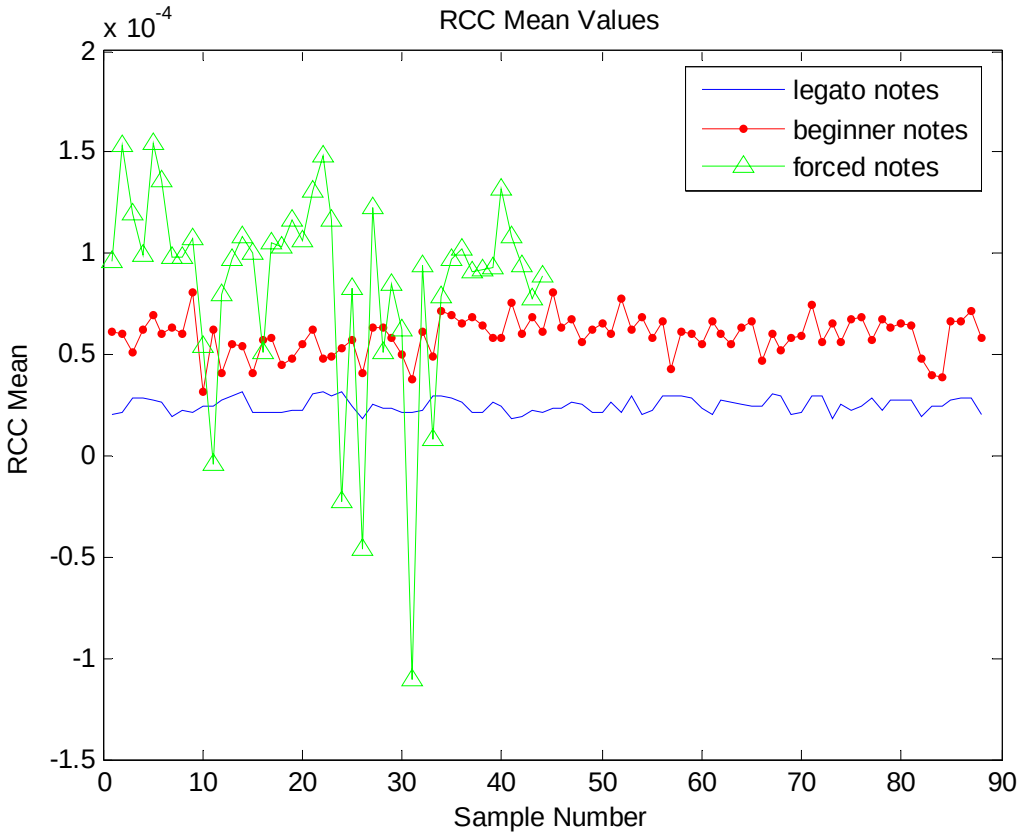

Figure 7.3: Real cepstral coefficients mean professional standard legato, beginner and forced note samples.

The dataset's RCCV values displayed in Figure 7.4 are low and more consistent for the professional standard legato note samples and tend to be higher and more varied for the beginner ones. This reflects a beginner player's inconsistency. As a measure, the RCCV discriminates well between the dataset's different player groups. The mean RCCV for the beginner note samples is $3.6*10^{-4}$ and for the professional standard legato notes, $8.92*10^{-5}$. These values are marked by the black dotted lines in Figure 7.4. The difference between the professional standard and beginner note samples' RCCV results are statistically significant when a t-test with a 0.01 significance level is applied. The null hypothesis is rejected and a p-value of $3.67*10^{-40}$ is returned. The beginner samples with lower RCCV readings contain multiple faults. Of the samples with overlapping RCCV values, legato sample 24 has a beginner player label. Neither the worst sounding professional samples nor the best sounding beginner note samples fall into this region. These results reflect what is associated with professional standard legato notes, consistency but none of the qualitative expressions used in this work is specifically reflected by this measure.

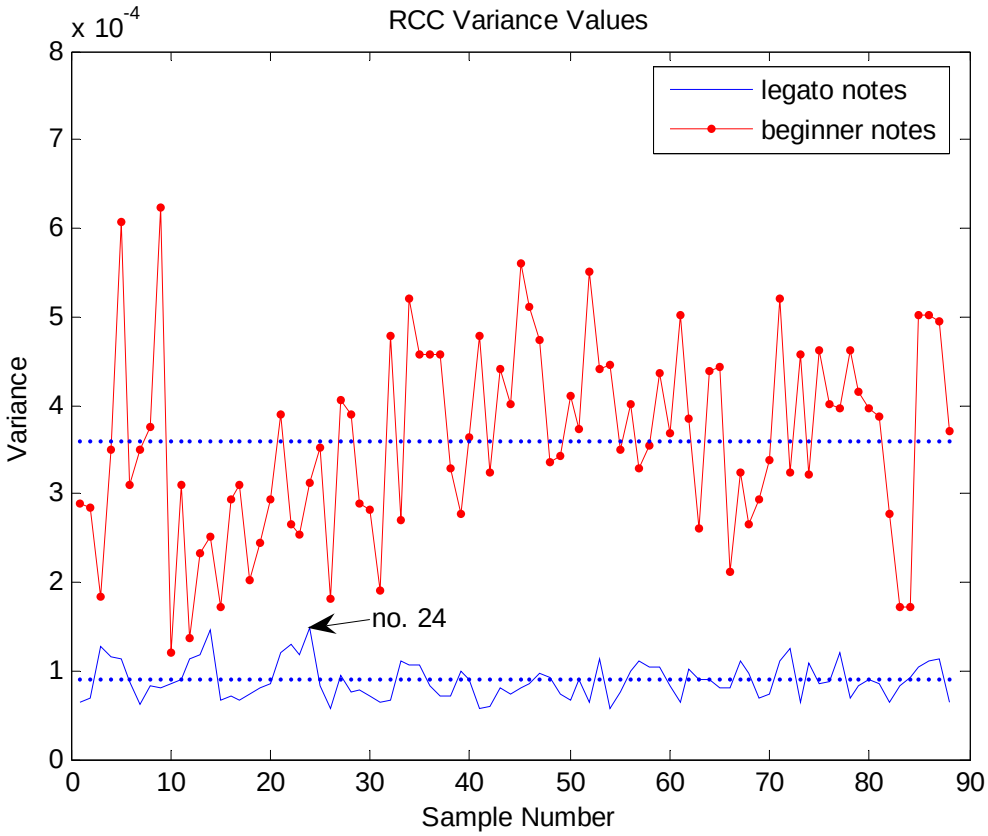

Figure 7.4: Real cepstral coefficients variance professional standard legato and beginner note samples.

Skewness is a measure of symmetry and both beginner and professional standard player legato note samples have overlapping positively RCC skew (RCCS) values and for this reason are not displayed. On closer inspection of these results, the beginner note samples have RCCS values that cover a wider range than those returned by the professional standard legato ones, revealing an underlying pattern confirming professional standard legato note consistency. There seems to be no sound quality relationship reflected by the skew value as these two player groups have been shown to be quite perceptually distinct based on the results of the listening tests. This indicates that no conclusion about violin timbre and this measure can be easily drawn.

Two distinct sample groups emerge when RCC kurtosis (RCCK) value is used to represent the dataset as displayed in Figure 7.5. The RCCK is a measure of the data's RCC "peakiness". All samples shown in Figure 7.5 have kurtosis readings well above 3, which are super-Gaussian results. The professional standard legato note samples tend to have higher RCCK values then the beginner ones with mean RCCK readings of $7.77*10^{+4}$ and $3.04*10^{+4}$ respectively. These mean values are marked in Figure 7.5 by

the dotted black lines. In this representation, the professional standard legato note samples provide more varied results than the beginner ones do.

The two worst sounding professional standard legato note samples from the listening tests, samples 52 and 71, have RCCK results which do not fall into the overlapping region. The lowest legato note sample in Figure 7.5 is sample 14 which has an overall sound quality grade of 5.19 out of 6. The professional standard legato note sample number 24, which has been labelled as a beginner note, appears within this overlapping region. The beginner samples with the highest RCCK values all contain multiple faults yet have similar RCCK values to those returned by some of the professional standard legato note samples. The two best sounding beginner samples from the listening tests, 62 and 65, have RCCK values close to the RCCK mean for the beginner player samples. Although the two groups are separated successfully by the RCCK value, the sensitivity of these readings does not correlate with the qualitative descriptions used in this text. The RCCK value is a coarse discriminator between the dataset's beginner and professional standard legato note samples.

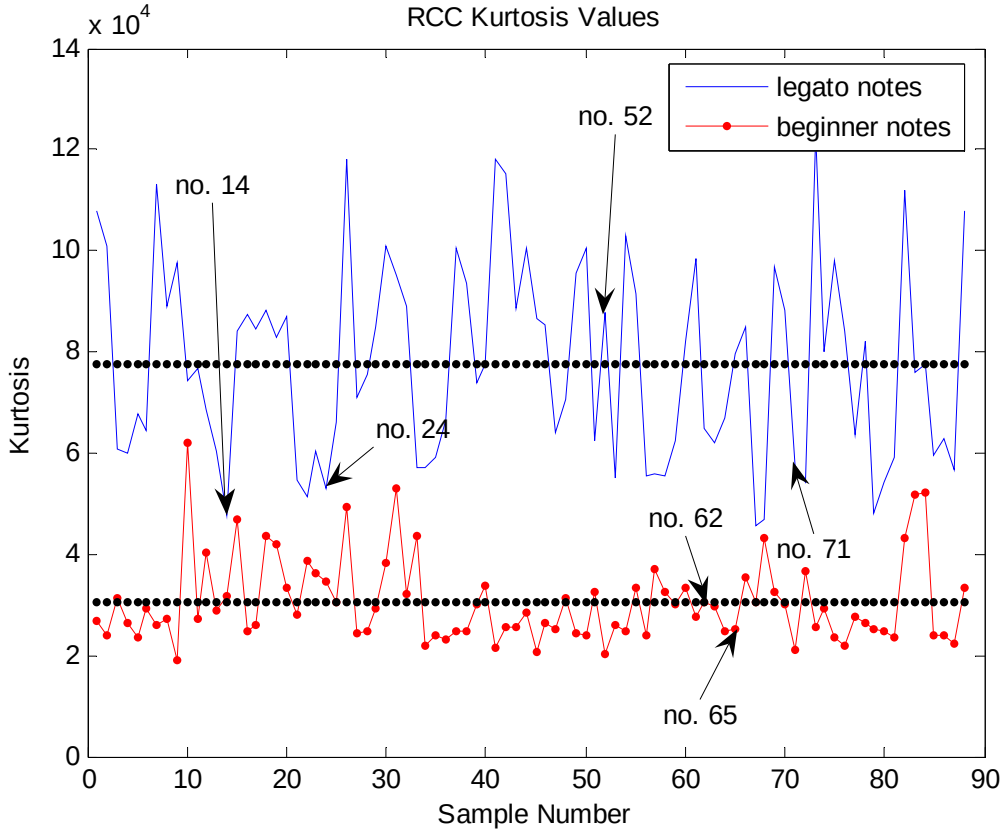

Figure 7.5: Real cepstral coefficients kurtosis readings professional standard legato and beginner note samples.

To investigate the sound quality and RCCK value relationship further, Figure 7.6 illustrates the effect of forcing the sound has on these readings. The forced notes' RCCK results fall into two groups, except for sample 33, which returns the highest RCCK value. The first section consists of samples where the professional standard player has emulated beginner crunching at the starts and ends of notes only. In the second part, crunching is maintained for the duration of the note. Knowing this, the relationship between amount of crunching and RCCK needs to be more clearly defined by grading the crunch quality and quantity. Regardless, the RCCK reading determines professional standard legato note samples from beginner ones in the dataset. When a t-test with a 0.01 significance level is applied to the difference between the beginner and professional standard legato note sample RCCK values, the null hypothesis is rejected, indicating that the results are statistically significant. The p-value returned is $2.2*10^{-41}$.

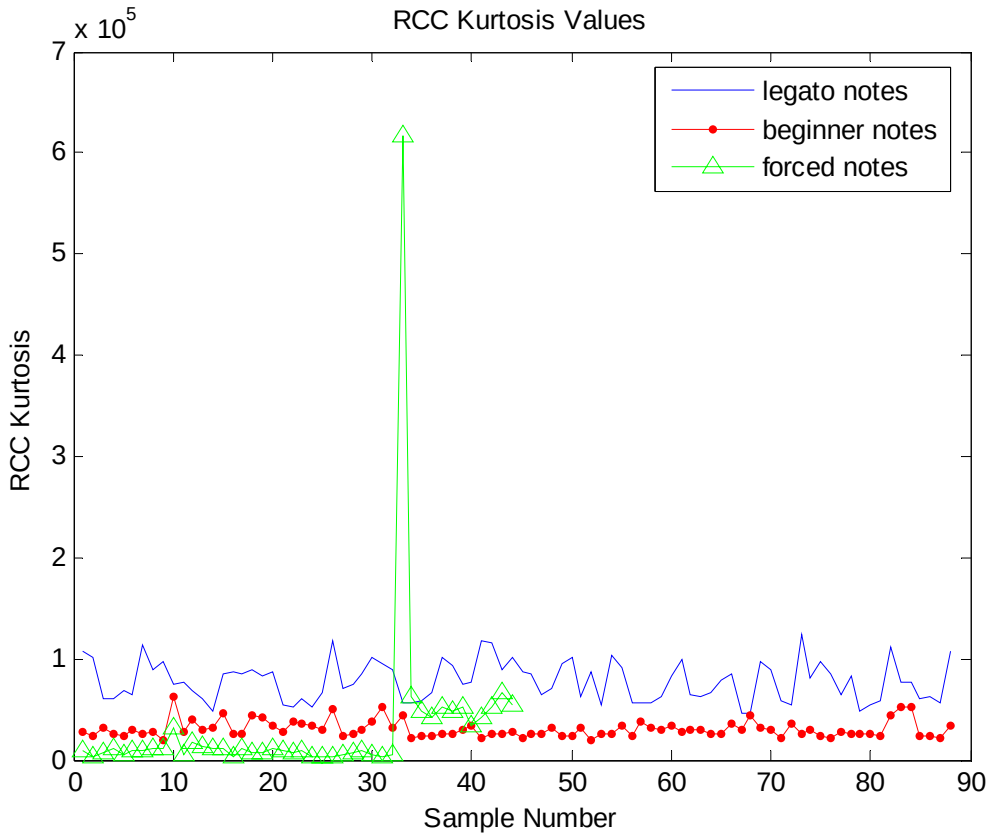

Figure 7.6: Real cepstral coefficients kurtosis readings professional standard legato, beginner and forced note samples.

The RCCM, RCCV, RCCS and RCCK readings describe violin timbre. All but the RCCS values effectively serve as coarse descriptors capable of differentiating between the professional standard player legato and beginner player note samples in the dataset.

The RCCS results do not separate the two different player groups but reflect an underlying pattern, that of the comparative greater inconsistency present in beginner note samples' results to those representing the professional standard legato ones. Unfortunately a link between these results and a specific sound characteristic has not been possible to establish with the dataset and qualitative sound descriptions used in this work. The effect of sound quality and individual RCCs is presented next.

According to [Hunt99], the most important information is available in the first 18 RCCs as coarse spectral shape is modeled by the lower RCCs. For this reason, inspection has been limited to these RCCs in this work. Of the 18 RCCs inspected, only three proved to be of use towards the research aims: the first, second and sixth RCCs. The first real cepstral coefficient (RCC0) is often used as a relative measure of cepstral energy and how it changes [Jayant84]. The RCC0 readings for the dataset's samples and the forced note samples are displayed in Figure 7.7.

Separation is good between the two different player types as professional standard legato note samples mostly have much higher RCC0 values comparatively to the beginner ones. From these results, energy levels are higher and more consistent for the dataset's professional standard legato note samples and lower with greater variance for the beginner ones. This fits with a beginner's playing being weaker and less consistent. Beginner players have less bow control and are less capable of producing a committed sound which is reflected by the RCC0 value, allowing this feature to be used as a discriminator between the beginner and professional standard legato note samples in the dataset. The effect forcing the sound has on the RCC0 value is also displayed in Figure 7.7, where most of the forced note samples have lower RCC0 values than the beginner ones. Forcing the note returns the most varied RCC0 values. Beginner samples three and 12 are the two highest beginner peaks with RCC0 values of -2.78 and -2.75 respectively. Both samples contain multiple playing faults. Out of the lowest scoring legato note samples, only one has been labelled by the listeners as being a beginner note, sample number 24. It has a RCC0 value of -3.07. From these results, the RCC0 results distinguish well between the two different player types present in the dataset.

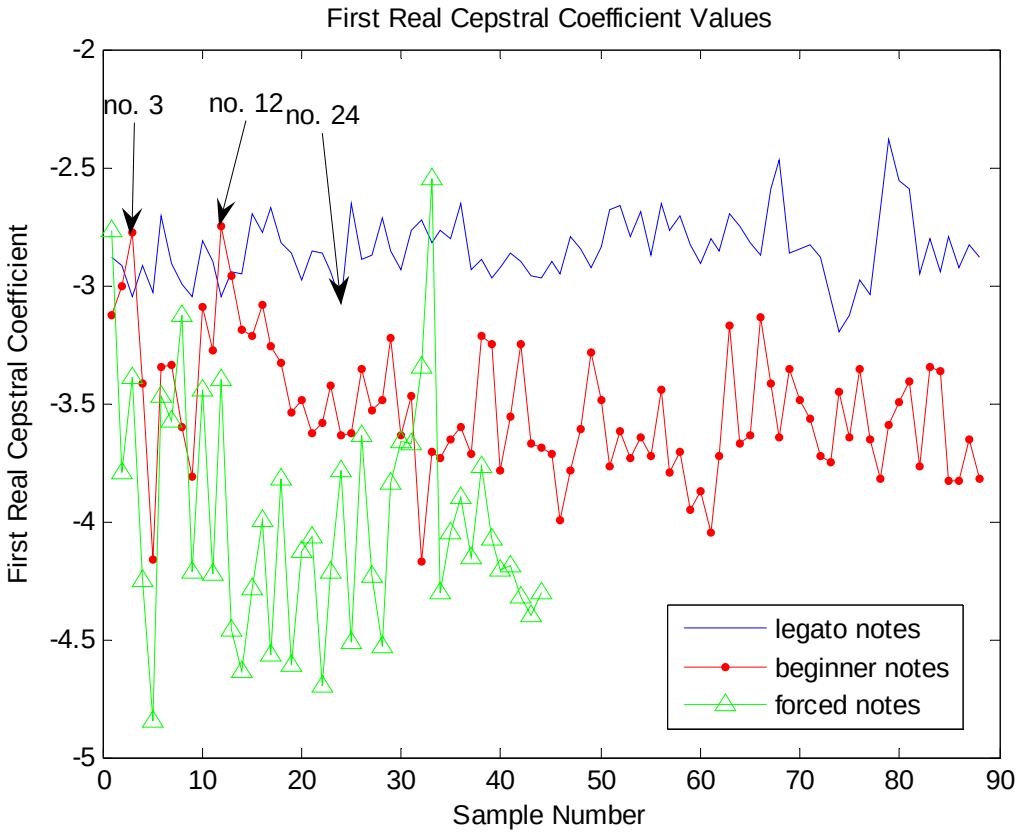

Figure 7.7: Real cepstral first coefficient readings professional standard legato, beginner and forced note samples.

The second cepstral coefficient (RCC1) represents the balance between the upper and lower halves of the spectrum [Hunt99]. It is a ratio of upper harmonic to lower harmonic presence. The RCC1 values for the dataset's samples and those for the forced note samples are plotted in Figure 7.8. The professional standard legato note samples mostly have lower RCC1 readings than the beginner ones. A significant amount of the beginner note RCC1 readings are above 1 implying that there are more upper harmonics present than lower ones in these samples. This supports the steelier, brasher, squeakier sound which is often associated with beginner playing.

In Figure 7.8, several beginner note samples' RCC1 readings overlap with those obtained from professional standard legato ones. The forced note samples return the most varied RCC1 values. Beginner samples with the lowest RCC1 values are reported to have playing faults. The two best sounding beginner samples, 62 and 65, have higher RCC1 values, but are still below 1. Beginner samples 17, 22 and 23, which have overall quality grades of 1, have high RCC1 values at 1.21, 1.14 and 1.18 respectively. The professional standard legato note sample with the highest RCC1 value, sample 14, has

an overall sound quality grade of 5.43 out of 6. The next two highest legato note RCC1 values are for samples 47 and 63 which have beginner player labels. These samples have overall sound quality grades of 4.33 and 3.62 respectively and no faults have been identified in either sample. Their RCC1 values are 0.83 and 0.82 respectively which are high compared to those returned by the other professional standard legato note samples. Critically listening to these two samples and focusing on the sound produced, sample 47 sounds a little “grainy” and lacks “body”, which are not specific faults in this work, whereas sample 71 lightly tips another note shortly after the start of the note. The mean RCC1 value for the legato professional standard note samples is 0.71. The beginner note samples with much lower RCC1 values all contain faults which the legato note samples with equivalent values do not possess. This implies that the relationship between sound quality as described in this text is not completely reflected by the RCC1 reading. Samples with emulated crunching at note starts and ends, return RCC1 values similar to those obtained by the beginner note samples. Prolonged crunching, which occurs from forced sample number 33 onwards, tends to lower the RCC1 readings.

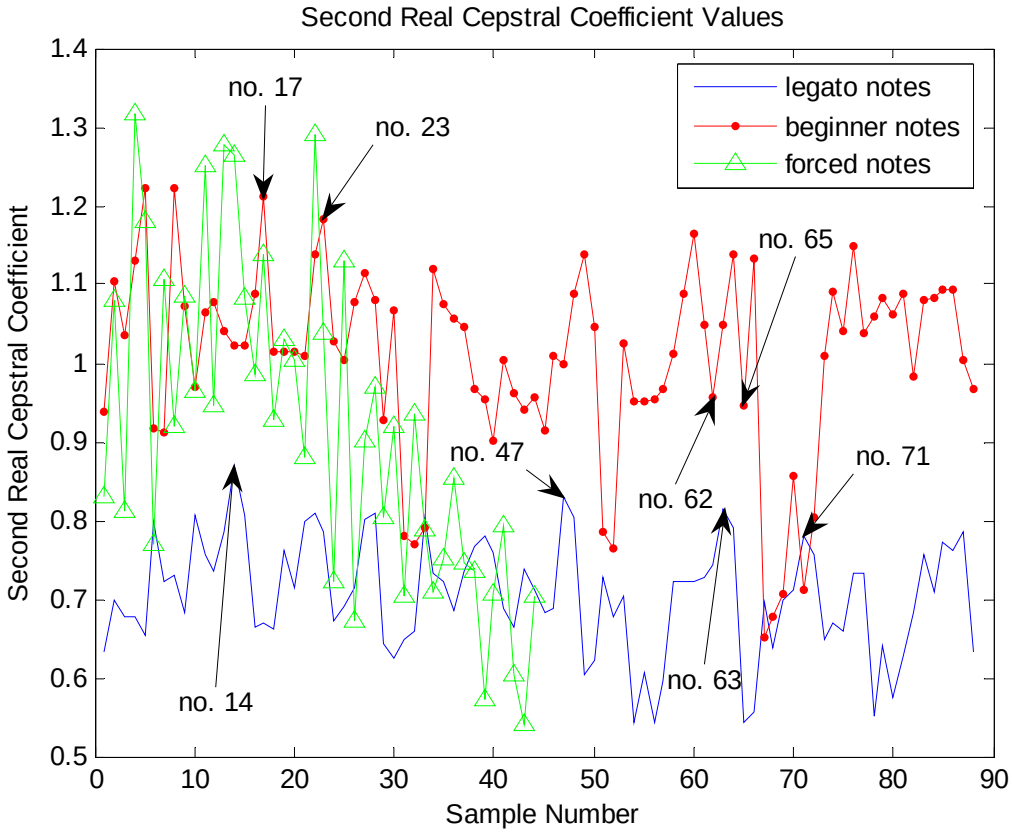

Figure 7.8: Real cepstral second coefficient values professional standard legato, beginner and forced note samples.

The RCC1 discriminates well between beginner and professional standard notes for the majority of samples in the dataset and can be used to describe violin timbre within the context set. The last RCC which represents the two player groups mostly separately is the sixth cepstral coefficient (RCC5), which is presented next.

The RCC5 values obtained for the dataset's samples and for the forced note samples are displayed in Figure 7.9, where the professional standard legato note samples tend to have comparatively higher values to the beginner and forced ones. The professional standard legato note sample with the lowest RCC5 value is sample 68 which has an overall sound quality grade of 5.52. The beginner note samples with the highest RCC5 values, samples 71, 44, 43 and 58, do not have the highest overall beginner sound quality grades. The top sounding beginner note samples, 62 and 65, have the sixth and 52nd highest RCC5 values respectively. The RCC5 values for many forced note samples are similar to those returned by the beginner ones. Based on these observations, a connection between a specific sound characteristic as described in this work and the RCC5 value is not evident, but this measure differentiates well between the dataset's beginner and professional standard legato note samples.

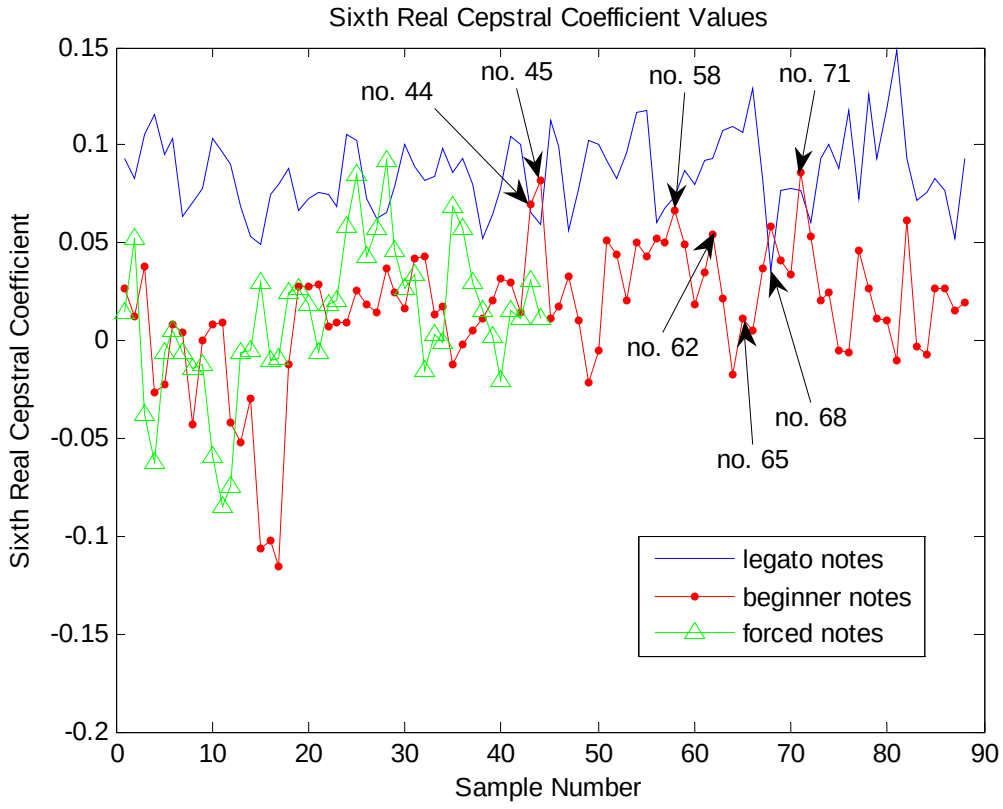

Figure 7.9: Real cepstrum sixth coefficient values professional standard legato, beginner and forced note samples.

The RCCM, RCCV and RCCK values as well as three individual cepstral coefficients, RCC0, RCC1 and RCC5 are coarse violin timbre descriptors within the context of this research and the dataset used. These six features correctly group the professional standard legato and beginner note samples in the dataset but none is correlated with the qualitative expressions used in this text. The Mel cepstrum and possible Mel cepstrum based features are presented in the next section.

7.2 Mel Cepstral Features

Developed by Stevens and Volkman, a Mel is a measure of perceived pitch of a tone [Deller00]. In the Mel frequency cepstrum, the data is converted into its Mel scale equivalent frequency before the discrete cosine transform is applied. On the Mel scale, frequencies below 1kHz are linear, which is within the human speaking range where the human and also the range within which the dataset samples' fundamentals fall. An approximation of the Mel scale conversion for f greater than 1kHz is achieved by applying Equation 7.1, which comes from an implementation in the HTK Toolkit where f is the frequency in Hz [Young95]:

$$mel(f) = 2595 \log_{10}(1 + \frac{f}{700}) \tag{7.1}$$

The Mel frequency cepstrum based features presented in this section include the Mel frequency cepstral coefficients (MFCCs) first order statistics and individual MFCCs.

First introduced by Davis and Mermelstein in a study on monosyllabic words [Davis80], MFCCs are widely used in feature extraction and music information retrieval algorithms [Logan01, Tzanetakis02]. MFCCs are obtained by taking the absolute STFT, converting to Mel frequency by grouping neighbouring frequency bins together into overlapping triangular bands with bandwidth according to the Mel scale and then by applying a DCT to its log. The first 12 MFCCs of a professional standard legato note sample and that of a beginner one are illustrated in Figure 7.10 have been obtained by applying the HTK toolkit approach which is available at [LABROSA]. A maximum frequency of 8kHz has been applied, 25ms window size and 10ms hopsize have been assigned. The filter bank used consists of 24 triangular filters with constant bandwidth up to 1kHz above which, constant Q applies. 40 Mel bands have been used.

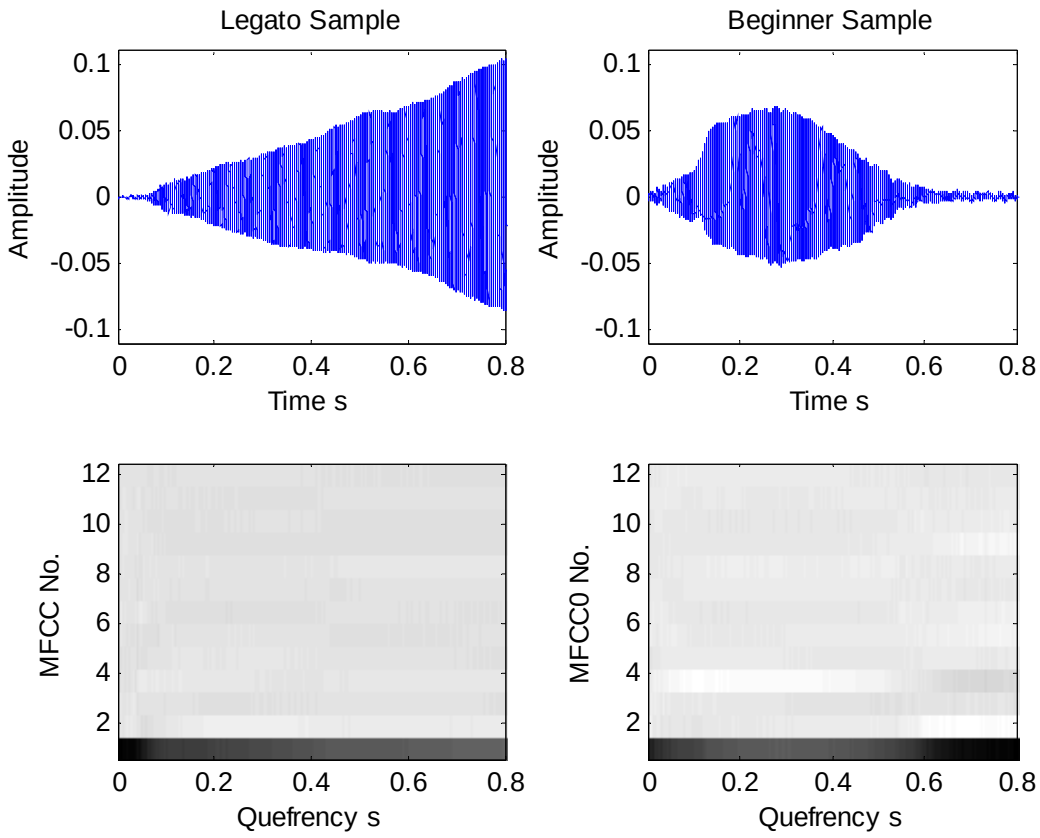

Figure 7.10: First 12 MFCCs of a professional standard A440l legato note sample (left) and of a beginner A440 note sample (right).

From the listening tests, the samples displayed in Figure 7.10 have overall quality grades of 5.52 and 1.76 respectively. The beginner note sample is reported to have skating, nervousness, intonation and bow bouncing as playing faults. From these images, based on fluctuations with respect to time, some MFCCs are more sensitive to changes within the violin timbre space than others. Also worth investigating is the change occurring within the first few frames of the MFCCs. From this, the difference in attack information between beginner and professional standard legato note samples is reflected.

First order statistics have been applied to the information returned within the first 18 MFCCs individually. In Figure 7.11, the first Mel cepstral coefficients mean (MFCC0M) values for the dataset samples are plotted. These values reflect energy content in a signal [Logan01]. The MFCC0M readings of the beginner samples display much greater variability than those belonging to the professional standard legato ones. This reflects greater player consistency present in the professional standard legato note samples than in the beginner ones. The remaining MFCCs have been inspected and only

the fourth Mel cepstral coefficients mean (MFCC3M) provided some useful information, the results of which are displayed in Figure 7.12.

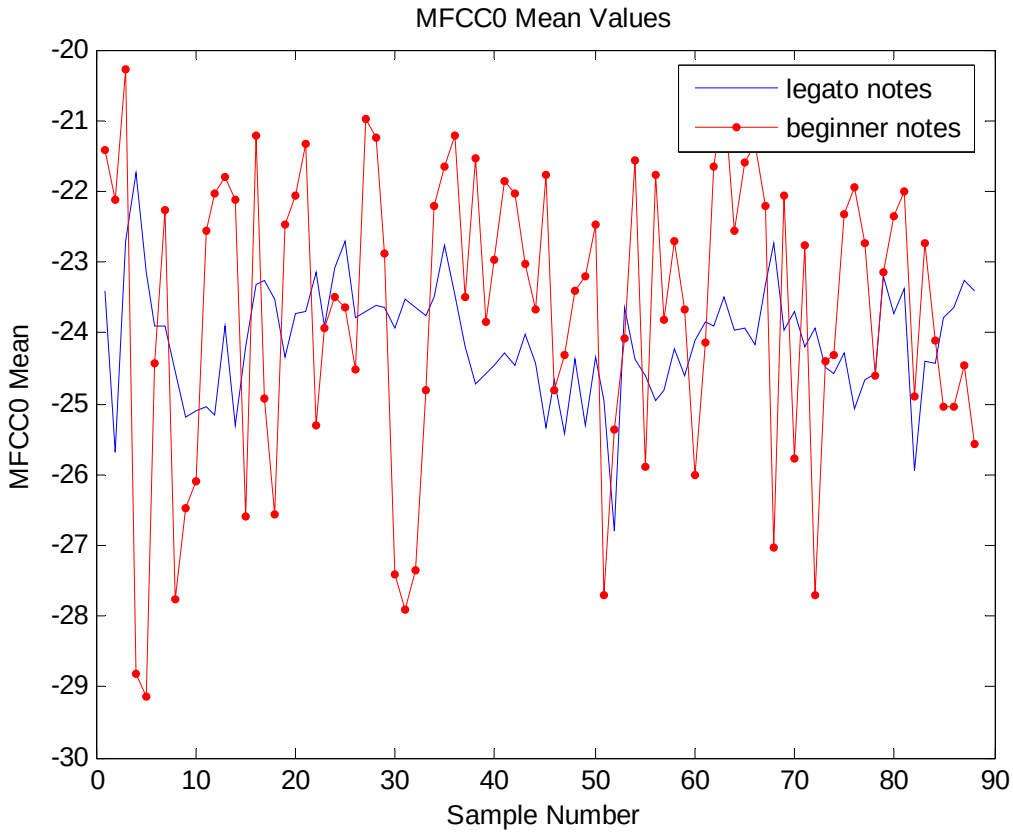

Figure 7.11: First Mel cepstral coefficient mean values professional standard legato and beginner note samples.

Taking the MFCC3M discriminates well between the two different player groups for part of the dataset only, as illustrated in Figure 7.12. From these results, the professional standard legato note samples tend to have mostly negative MFCC3M readings. Taking the MFCC3M proved to be partly effective at representing the dataset in the context of separating its two sample player groups. The beginner samples, having the poorest overall sound quality, return a wide range in their MFCC3M values whereas the legato note samples reflect greater consistency. The two best sounding beginner samples based on the listening tests, samples 62 and 65, have much lower MFCC3M values. These results alone do not facilitate drawing conclusions regarding the relationship between MFCC3M value and the expressions used in this text to describe violin timbre.

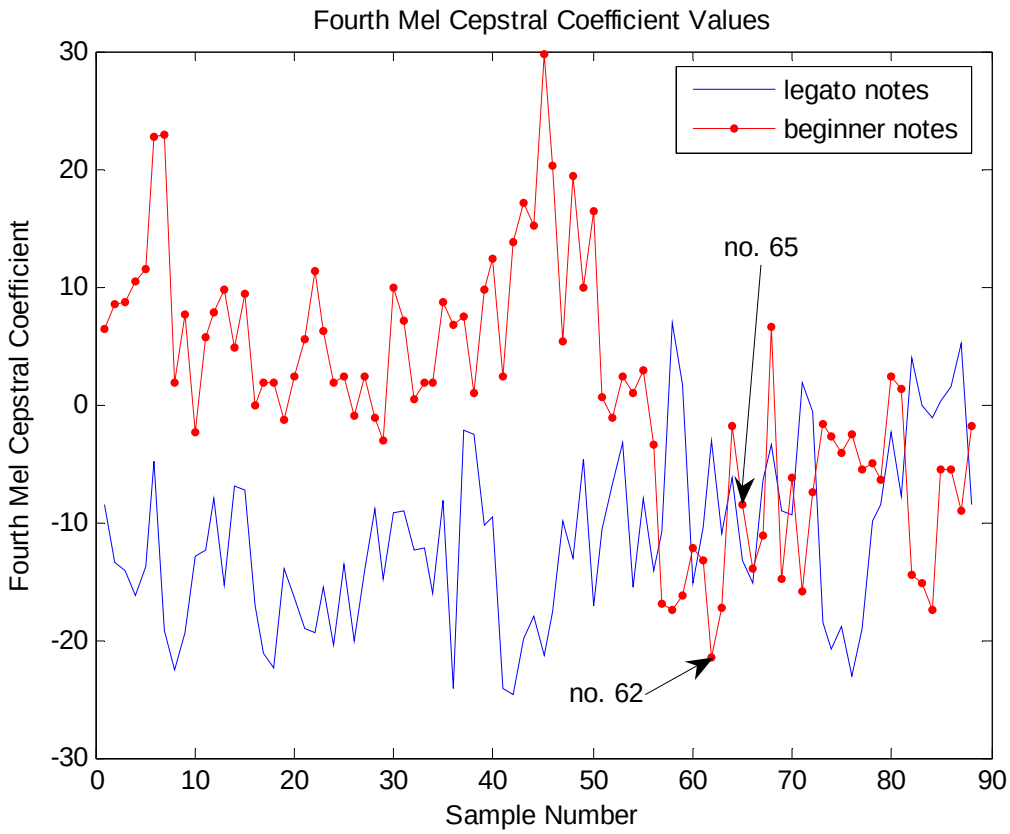

Figure 7.12: Mel cepstrum fourth coefficient mean values professional standard legato and beginner note samples.

The fluctuations in the MFCCs values throughout the samples have been observed and a measure which reflects this change is variance. The Mel cepstral coefficient variance (MFCCV) readings have been inspected and all proved to be ineffective at grouping separately the beginner from the professional standard legato note samples in the dataset. To illustrate this, only the MFCC0 variance values have been included and are depicted in Figure 7.13. In this figure, greater fluctuation is visible in the beginner note MFCC0 variance values then in the professional standard legato ones which remain much more consistent, falling within a smaller range. The comparative consistency of these samples is reflected by this measure although it does not group the different player types separately.

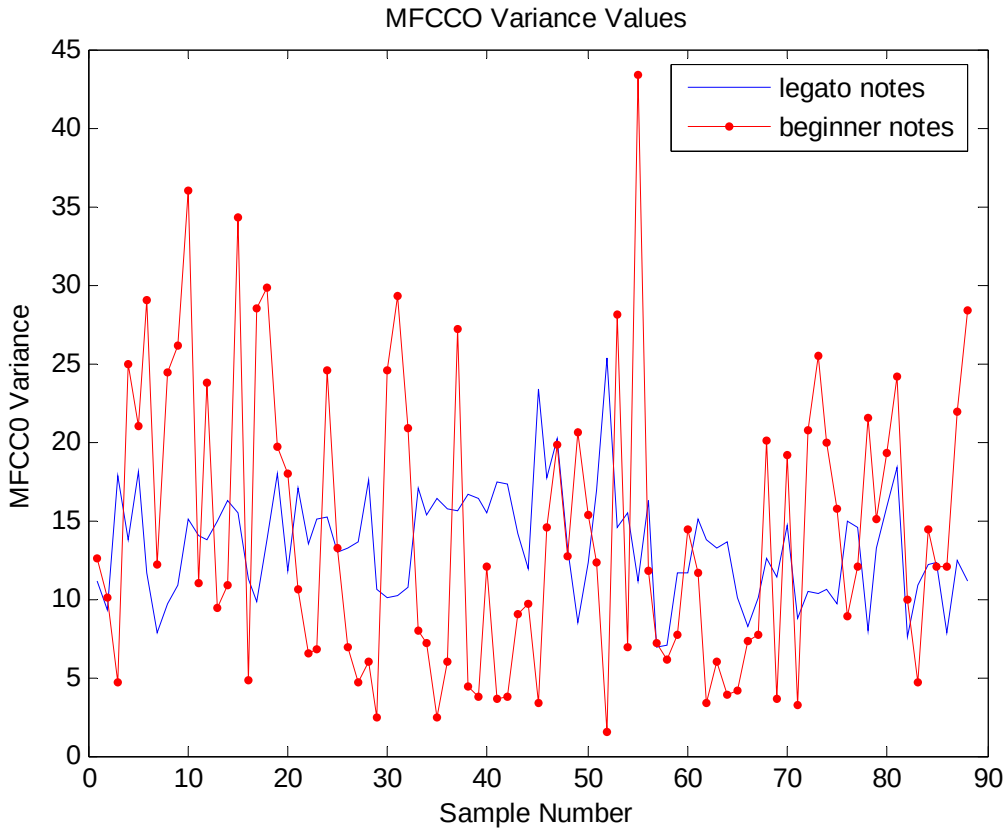

Figure 7.13: First Mel cepstral coefficient variance values professional standard legato and beginner note samples.

In Figure 7.14, the MFCC0 skew readings for the dataset's samples and those for the forced note samples are illustrated. Skew is a measure of asymmetry and the consistency of the legato note samples is reflected by this measure. The beginner note samples mostly have greater negative skew and the professional standard legato note samples' MFCC0 skew readings are more consistent and closer to zero. These readings all have small negative skew values where the lowest value is -0.0935 and the highest is -0.0037. Only four beginner note samples have MFCC0 values with positive skew, the remainder tend to be strongly negatively skewed. The two best sounding beginner note samples from the listening tests, samples 62 and 65, are negatively skewed. The three beginner samples with overall sound quality grades of 1, samples 17, 22 and 23, return varied MFCC0 values as marked in Figure 7.14. Beginner sample number 71 returns the lowest MFCC0 skew reading. This sample has not been perceived to contain any playing faults and has an overall sound quality grade of 3.8 out of 6. The professional standard legato note samples' comparative consistency is depicted but perceived sound quality as captured through the listening tests is not reflected by this measure. Emulated

crunching at the starts and ends of notes, as demonstrated by the forced note samples up to number 33, return varied skewed readings. Once the forcing or crunching is consistent, i.e. for the duration of the note, the skew readings stabilise and remain closer to zero as can be seen from forced sample 33 onwards. Skew reflects symmetry, which in this case, reflects consistency throughout a note. Consistently poor as well as consistently good sound is captured by this measure, as confirmed by the forced note samples.

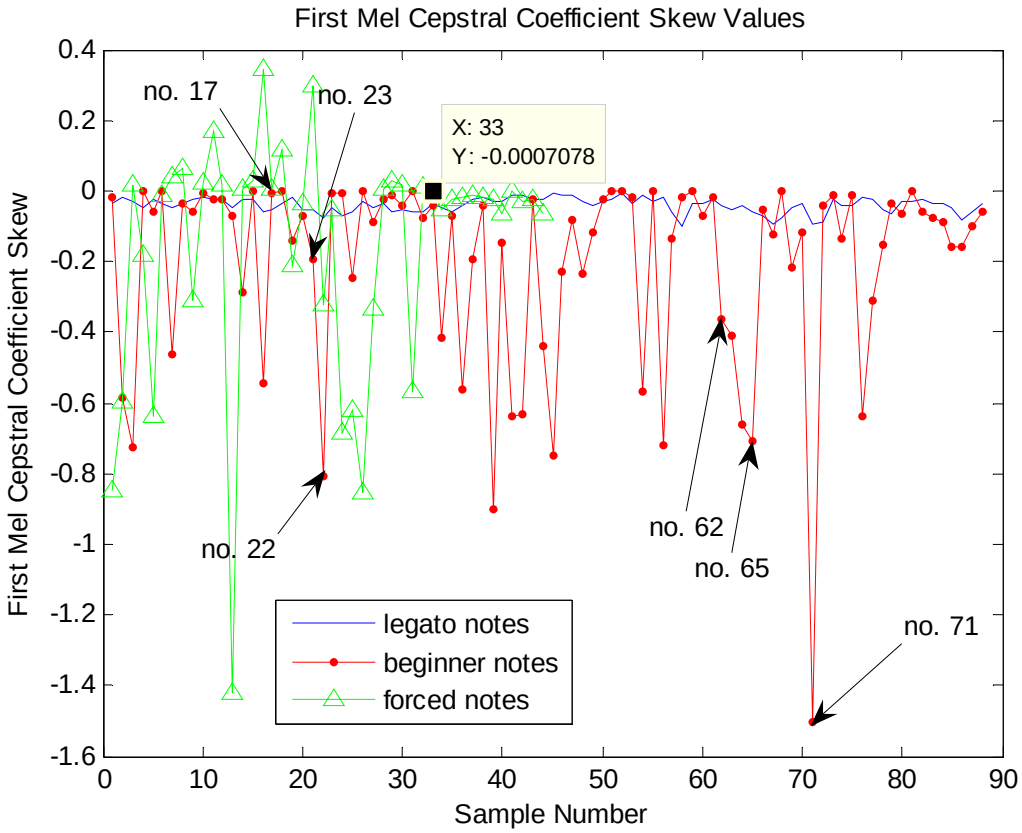

Figure 7.14: Mel cepstrum first coefficient skew values.

The results obtained from applying kurtosis to the dataset's samples MFCC0 values are displayed in Figure 7.15. Although this measure does not group the samples separately according to player, the underlying pattern is important. The professional standard legato note samples return results that fall within a smaller range compared to those representing the beginner ones, reflecting consistency. The MFCC0 variance and skew results also have similar underlying patterns, all indicating more consistent energy content in the professional standard legato note samples than in the beginner player ones.

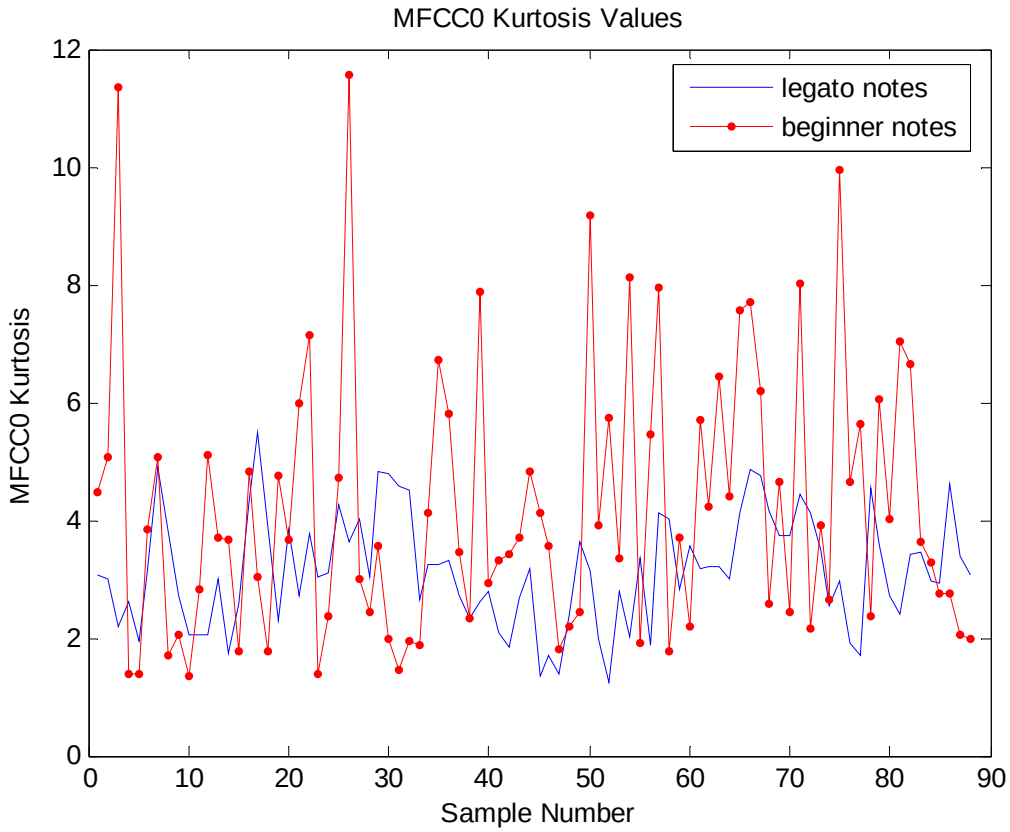

Figure 7.15: MFCCO kurtosis values professional standard legato and beginner note samples.

As the MFCC0 represents energy, how it fluctuates is important, in particular during a note's onset as this establishes the note's timbre. Focusing on these changes within the first 0.087s of each signal, the MFCC0 mean of this section of each sample is taken and plotted in Figure 7.16. In this figure, the professional standard legato note onsets as represented by the first ten frames MFCC0 mean, are much more consistent and controlled whereas the beginner note sample readings cover a much wider range of values.

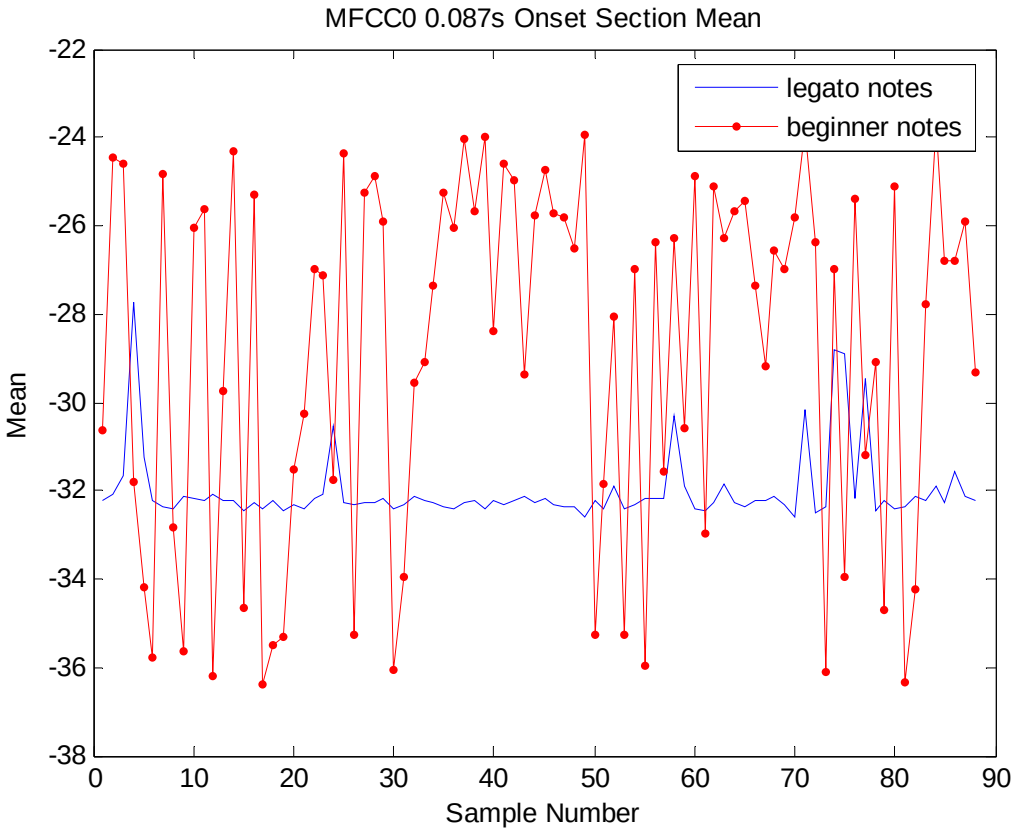

Figure 7.16: MFCC0 mean first 0.087s section of a professional standard legato and beginner note samples.

The information present within the first 0.087s period of the first 18 MFCCs of the dataset's samples was inspected and only MFCC2 revealed results of interest. These are displayed in Figure 7.17. Based on the results presented, the dataset's beginner note samples tend to have lower MFCC2 first 0.087s mean values with much greater variability. From the listening tests, beginner samples 17, 22 and 23 all have overall sound quality grade 1 and samples 62 and 65 have been perceived as being the best sounding beginner note samples in the dataset. Samples 56 and 71 have been labelled as being the two worst sounding professional standard legato note samples. These samples all have MFCC2 first 0.087s mean attack section readings that are not grouped in any particular manner, making the association between this measure and any of the qualitative expressions used in this thesis difficult. When used in conjunction with other features, the MFCC2 onset values may assist in correctly determining violin timbre and playing quality characteristics.

The Mel cepstrum has provided multiple features which describe violin timbre. First order statistics have been applied to each of the first 18 MFCCs and some have proved

to be suitable violin timbre discriminators for the dataset's samples. The Mel cepstrum provides many timbre descriptors but the MFCC0, MFCC2 and MFCC3 provide results which more specifically represent violin timbre quality in the dataset used. The MFCC0M, MFCC0V, MFCC0S, MFCC0K, MFCC0 first 0.087s mean and MFCC2 first 0.087s mean reflect the consistency of the professional standard legato note samples comparatively to that of the beginner ones, but correlating any of these measures with the expressions used is not evident.

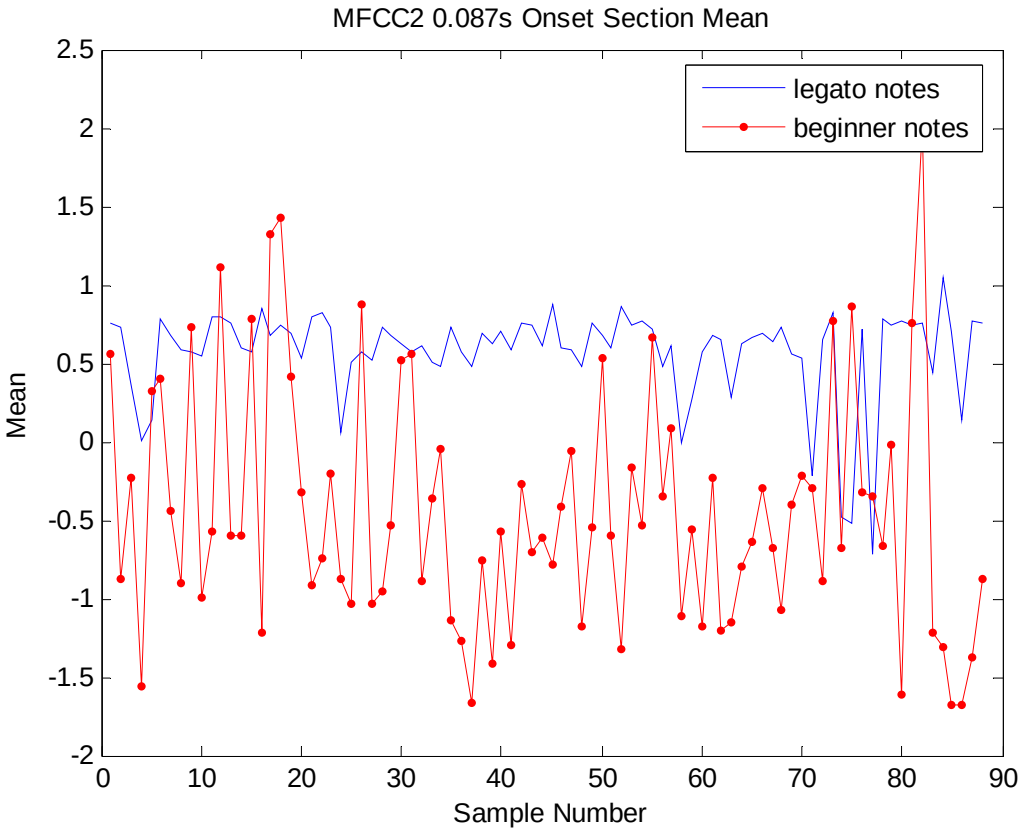

Figure 7.17: MFCC2 mean first 0.087s professional standard legato and beginner note samples.

7.3 Summary

Many cepstral features have been applied to represent violin timbre. More specifically, features have been found which distinguish between the professional standard legato and beginner note samples in the dataset from both the real and Mel cepstra. More specifically, in the real cepstrum, the RCCM, RCCV, RCCK, RCC0, RCC1 measures and from the Mel cepstrum, the MFCC0, MFCC2 and MFCC3 values provide some useful results for representing violin timbre change in the dataset. The results confirm

that the real cepstrum provides more useful results than the Mel cepstrum, supporting what has been concluded elsewhere [Deller00]. Although some of the features presented performed better at detecting professional standard legato notes from the beginner ones, others effectively detected sound consistency within the sample groups. Regardless of the not very evident relationship between sound descriptions and cepstral measurements, these features remain effective coarse descriptors for the violin timbre space. Features from these cepstra describe violin timbre and will be considered for use in the classifier. This stated though, features obtained from the spectral and time domains display greater potential for further work on violin timbre representations for analysis.

The time, spectral, and cepstral domains all provide useful violin sound descriptors. Some of these features perform better than others at distinguishing beginner note samples from professional standard legato ones in the dataset. Their individual and combined effectiveness at the detection tasks will be tested in the following chapter via a classifier.

8 Classification

So far in this work, quantitative and qualitative aspects of violin sound have been presented. The use of features from the temporal, spectral and cepstral domains has been examined, returning features capable of representing violin timbre. Listening tests have been run allowing qualitative labels to be assigned to the dataset's samples reflecting stringed instrument musicians' perception. The information gleaned so far merges in this chapter, providing a classification system for violin notes. Classification is the general term given to organizing or grouping similar data together according to selected characteristics or some common feature and is the approach taken in this work to test the representative features. Grouping data together based on similar patterns or descriptive features allows a class label to be associated with the group. The most significant aims of classification relate to data simplification and prediction, increasing the efficiency of tasks such as information retrieval [Gordon99]. Violin timbre features from the temporal, spectral and cepstral domains and the *a priori* labels obtained from the listening tests are used for classifying violin notes. The aim is to provide objective and stable classification for the subjective nature of violin sounds.

In this chapter the classification of violin notes based on overall sound quality and individual playing faults is presented. First, the classification steps are detailed and then the classifier is tested. Two tasks are put to the classifier: the detection of beginner from professional standard player legato note and individual playing fault detection. The outcomes' ability to generalise is then tested on new data and conclusions are drawn.

8.1 Classification Procedure

The classification steps applied to the violin timbre tasks are detailed in this section. The dataset is represented as an $n \times f$ array where n is the number of samples and f, the number of features used. From the dataset, suitable cluster centres are obtained via Jain and Dubes' k-means clustering algorithm [Jain88] and the *a priori* labels come from the listening tests. A k-nearest neighbour (k-NN) classifier is then used, the labels compared and classifier accuracy obtained, as shown in Figure 8.1.

The features initially selected to represent the violin note samples are based on visual inspection of their ability to separate the dataset's samples into two distinct

player groups within their respective domains. In all, 33 features are used to represent the dataset's samples. The playing fault descriptions used in this work are recalled in Table 8.1 along with the abbreviations used throughout this chapter.

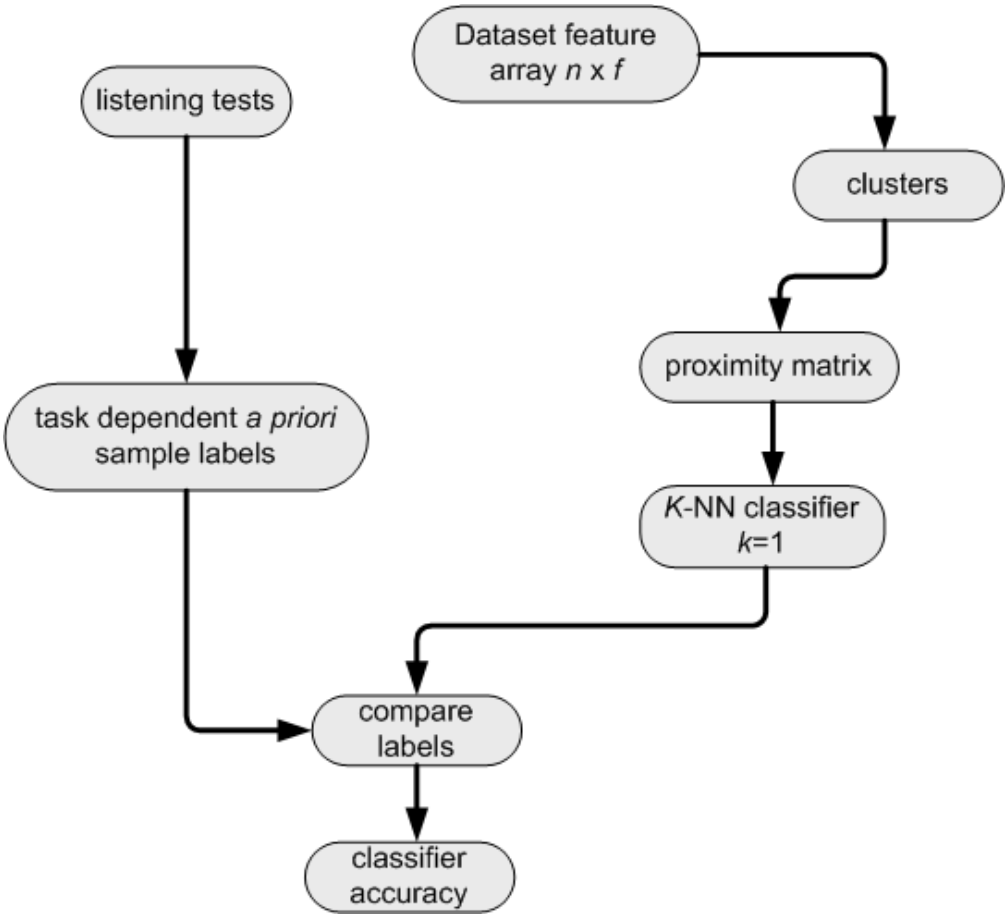

Figure 8.1: Classification steps.

Number	***Fault Name***
Fault 1	crunching (CR)
Fault 2	skating (SK)
Fault 3	nervousness (NV)
Fault 4	intonation (INT)
Fault 5	bow bouncing (BB)
Fault 6	extra note (XN)
Fault 7	sudden end to note (SE)
Fault 8	poor start to note (BADS)
Fault 9	poor finish to note (BADE)

Table 8.1: Fault descriptions.

The task dependent *a priori* labels have been obtained via the listening tests detailed in Chapter 3. There are two groups of labels: one reflecting the overall sound quality and the other, for individual faults present. Each listener evaluated the overall sound quality of all the samples by giving a grade between 1 (very poor) and 6 (excellent). To represent professional or beginner samples in the first task and existence or non-existence of a playing fault, only two clusters are required. Class labels of '1' for

professional player or no fault present and ‘2’ for beginner or fault present are used to reflect the listeners’ perception. This was done by finding all the samples which had an overall sound quality grade of 5 or above and re-labelling them as ‘1’ and the remaining samples as ‘2’. Grading level 5 was taken and not 4 because only the good to excellent sounds should be classified as professional sounds and not those with quality perceived as being “reasonable”. The dataset consists of 88 beginner notes and 88 professional standard legato good notes. 82 of the 176 samples have been perceived as being good and consequently have label ‘1’ and the remaining 94 have label ‘2’. Using the information obtained about fault perception, labels were assigned according as to whether a fault had been perceived or not. Samples perceived as having a specific fault have been labelled with ‘2’ for that fault and ‘1’, for the fault not having been perceived. These fault labels are stored in a 176 x 9 array, the order of faults is as shown in Table 8.1. Playing faults rarely occur in isolation and most of the beginner player samples contain more than one fault. After having obtained the dataset’s *a priori* labels, the next stage involves finding suitable clusters.

Clustering is an exclusive, intrinsic, partitional classification method and is the most common form of unsupervised learning and often used as the first stage of a classification process [Jain88]. Clustering techniques are used to find centres which reflect the distribution of data points [Bishop95]. For the first task, two clusters are sought: one for poorer quality or beginner violin sounds and another for professional standard legato notes. For the fault identification task, the clusters are for presence and absence of a fault. Although clusters are inferred from the data it is possible to influence the outcome by, for example, the choice of distance measure used [Duda73]. The distance measure represents the relationship between pairs of points or vectors belonging to the sample space and is important in any automatic procedure which attempts to mimic human perception for identifying clusters. The most commonly used distance measures include the Euclidean, Minkowski and Canberra metrics [Krzanowski95]. *K*-means clustering is one of the most often used clustering methods because of its simplicity and robustness. It converges well with the Euclidean distance, which has been selected for use in this work and is given in Equation 8.1 [Jain88]:

$$Euclidean\_Dist = \frac{1}{N}\left(\sum_{n=1}^{N}(A(n)-B(n))^2\right)^{1/2} \qquad (8.1)$$

The *k*-means clustering code, taken from the Somtoolbox [SOM] from Kohonen's work on self-organising maps [Kohonen90], uses the iterative partitional clustering algorithm put forward by Jain and Dubes, a description of which is in [Jain88:96-7]. An advantage of this algorithm is that it automatically assigns items to clusters. The disadvantages are that the number of clusters must be pre-selected and that all items are forced into a cluster, making it sensitive to outliers. The squared Euclidean distance metric is used which is computationally faster for clustering than the Euclidean distance [*ibid.*] shown in Equation 8.1. The clustering algorithm remains unaffected by this change as it is a partitional clustering method as opposed to a hierarchical one. Clusters obtained from non-negative matrix factorisation (NMF) and singular vector decomposition (SVD) have also been investigated. Significantly better results have been achieved using the *k*-means clustering algorithm comparatively to other clustering methods investigated. Only the *k*-means clustering results are presented in this work.

Running the *k*-means clustering algorithm provides the prototype vectors used in the *k*-NN classifier. For the first task, two cluster centres are needed. The "beginner" and the "professional" clusters centres become the f x 1 prototype vectors, where f is the number of features used. For the fault identification task, clusters are formed according to the presence or absence of a particular fault based on the listening tests. Prior to use in the classifier, these prototype vectors were checked by comparing their values with the means of all samples for each feature associated with its respective cluster to check for convergence. The algorithm converged well and no alterations had to be made. From the listening tests, clusters based on perceived presence of each fault were also used to inspect the existence of perceptual correlates for the violin's timbre space.

A proximity matrix is calculated using the squared Euclidean measure between the prototypes and the feature vector array. This matrix is inputted into the *k*-NN classifier, to which class labels are assigned. These labels are then compared with the *a priori* labels to obtain the classifier accuracy reading. The *k*-NN rule classifies a sample by assigning it the label which is most often associated with its *k*-nearest samples. When k=1, it is a special case of a *k*-NN classifier where every sample is assigned to the class of the nearest cluster. In practice, k=1 is often used [*ibid.*], as it is in this work, as the dataset size is not too large. Prior to detailing the classification results, cross-validation methods are briefly presented.

8.2 Cross-Validation

Cross-validation techniques are methods for detecting and preventing classifier over-fitting, used for checking classifier accuracy estimation and generalisation potential. Classifier accuracy is the probability of correctly labelling a randomly selected sample. Cross-validation serves as an accuracy consistency measure allowing these results to be generalised and subsequently applied to another dataset. For estimating the accuracy of a classifier, an estimation method with low bias and low variance is best.

Rather than running a classifier on the entire dataset, cross-validation involves putting the dataset samples in a random order after which, a portion of the dataset is put aside as a training set and the remaining samples are used for testing. Two well established cross-validation techniques are *n*-folds and leave-one-out cross-validation (LOOCV). In *n*-fold cross-validation, the dataset is put into *n* equal sections where *n-1* sections are used for training and the remaining section is used for testing. The sections are rotated and the means of the results of the *n* classifications are taken. In LOOCV, as the name implies, each sample is removed one at a time and used for testing and the rest of the samples are used for training. This makes LOOCV an almost unbiased method but high variance can be a problem which can lead to unreliable estimates [Efron83]. From a purely practical perspective, LOOCV is computationally intensive and is better used on smaller datasets and also why *n*-fold cross-validation is favoured in this work. Four-fold cross-validation has been selected.

In four-fold cross validation, the randomly ordered samples are divided into four equal parts. Randomising the dataset prior to splitting it up into four equal parts reduces the possibility of biasing the cross-validation. Three quarters of the dataset are used for training and the remaining quarter for testing the classifier. The sections are rotated so that each quarter is used as the test set once. The results obtained for each section are compared and the differences between test and train sets are an indication of feature combination suitability for the detection task. The mean readings are taken from all four folds and used for analysis. Four-fold cross validation has been applied to both tasks and to all possible feature combinations, the results of which are presented in the following section.

8.3 Classification Results

A large number of features and feature combinations have been tested in the classifier. A total of 33 individual features and combinations of, have been used to represent the dataset's samples. Task I is the detection of professional standard legato notes from poorer sound quality ones, such as those associated with a beginner violinist and Task II is playing fault detection. The results obtained, via four fold cross-validation for both tasks are presented and conclusions are drawn. Four-fold cross-validation involves obtaining the mean accuracy results for every possible feature combination. The smaller the error between the training and testing sets, the better the associated feature or feature combination suits the classification task. All results obtained for each combination without repetition will not be shown in the text as it amounts to $\binom{n}{r}$ combinations where n is the total number of features and r, the selected number of features. The features selected to represent the dataset's samples were based on their ability or lack of to group the samples according to beginner and professional standard player in their respective domains. Of the 33 features used to represent the samples, six completely separate the dataset's samples accurately based on player type when applied directly to the data. These features are the TM, MMV, CQTH9, PSD190, SFMV and the SCM190. Good separation between player groups is also provided through representing the data by taking the TK, SFMM, RCCM, RCCV, RCC0, RCC1, RCC5 and the MFCC3 values. When using these last representations, less than ten samples' values overlap. A further ten features including the TV, AC, CV, SFMK, MFCC0M, MFCC0V, MFCC0S and the MFCC0K, return values which reflect an underlying pattern of interest. Although the readings overlap, these features' values confirm musicians' perception about the two different players groups in the dataset, i.e. the relative inconsistency of beginner playing. The remaining features are not effective at differentiating between the different player groups in the dataset as their values completely overlap, making the different player groups indistinguishable from each other. The classification results obtained for both tasks are detailed next.

8.3.1 Task I Results

In this section, a summary of the results obtained for Task I is presented. To start with, the monothetic results returned for determining beginner from professional standard

legato note samples via four-fold cross-validation are displayed in Table 8.2. The results which are of greatest interest are those returning the highest detection rates with no or very small difference between the testing and training set results as this reflects feature combination suitability for the selected task. The best performing features for this task via the classifier are the TM, MMV and CQTH9 features, which have returned 97% detection accuracy. Although these three features have performed well at grouping the beginner from the professional standard legato note samples correctly in their respective domains, not all features perform as efficiently at the same task via the classifier, as can be seen by the results displayed in Table 8.2[2].

No.	*Feature*	*Train %*	*Test %*
1	**TM**	**97**	**97**
2	TV	52	44
3	TS	52	52
4	TK	90	89
5	**MMV**	**97**	**97**
6	**CQTH9**	**97**	**97**
7	PSD	52	53
8	PSD190	52	51
9	SFM	58	58
10	SFMM	63	60
11	SFMV	92	92
12	SFMS	52	52
13	SCM190	92	92
14	CM	67	65
15	CV	70	69
16	CK	63	64
17	RCCM	91	91
18	RCCV	90	90
19	RCCS	92	92
20	RCCK	86	89
21	RCC0	88	88
22	RCC1	90	90
23	RCC2	72	73
24	RCC3	75	77
25	RCC5	88	86
26	RCC12	62	63
27	RCC27	59	63
28	MFCC0M	58	53
29	MFCC1M	50	51
30	MFCC3M	75	74
31	MFCC0K	60	55
32	SF	48	48
33	AC	50	50

Table 8.2: Monothetic classification results for Task I.

Of the remaining three features which successfully distinguished between the two different player groups in the dataset, SFMV, SCM190 and PSD190, two performed slightly less accurately and the third one performed poorly in the classifier. The SFMV and SCM190 returned lower detection results of approximately 92% accuracy even though when applied directly to the data, the results are 100% correct. The PSD190 performed poorly in the classifier, returning detection results around 50%. An explanation for this last poor result is displayed in Figure 8.2, which shows the distance

[2] All of these features and how well they represent the dataset's samples have been presented or discussed in the text except for RCC12 and RCC27. How these two features represent the dataset's samples are displayed in Appendix C.

between the two cluster centres and the dataset's samples as represented by their PSD190 values. Many of the beginner note samples, when represented by their PSD190 values, have very similar distances from the cluster centres than the professional standard legato ones. For this reason, the PSD190 does not function well in the classifier. This measure serves better as a threshold decision surface. Some of the other poor classifier performances can be attributed to having cluster centres which are too close together. This accounts for the incorrect assignment of a high proportion of samples, thereby returning classifier detection at approximately 50%.

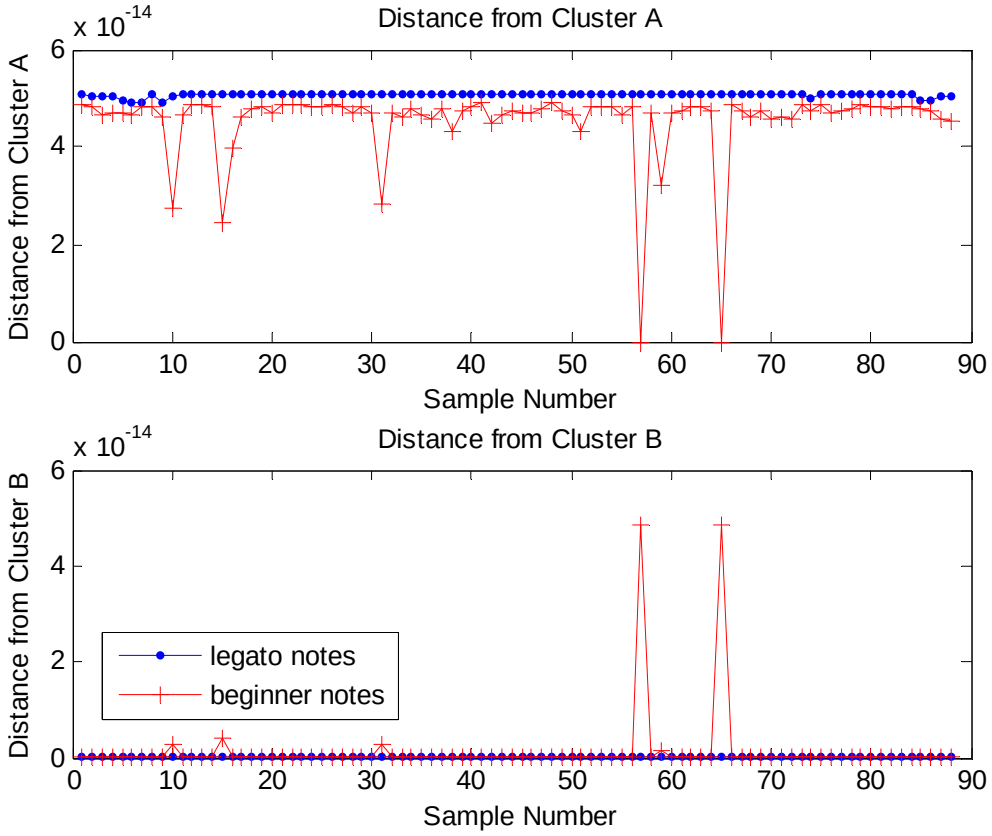

Figure 8.2: Distance between cluster centres and dataset samples' PSD190 values.

The best detection accuracy result for any feature used in the classifier, for the detection of beginner note samples from professional standard legato ones, is 97%. As stated previously, the features returning this result, the TM, MMV and CQTH9, separate the data according to player type with 100% accuracy when applied directly to the data. The slightly lower detection result returned when these features are used in a classifier can be linked to a small number of outlier values. There will always be a risk of classifier sensitivity to outliers when samples being tested are represented by only one feature. Although the monothetic results display much information about the dataset's

samples and cluster centres formed as represented by the different features, an improvement on 97% detection accuracy is sought. This may be achieved by decreasing cluster sensitivity to outliers by using more than one feature to represent the samples.

Before presenting the results returned when multiple features are used to represent the dataset's samples, the other features which have returned high detection rates for Task I are noted. These include the TK, SCM190, RCCM, RCCV, RCCS, RCCK, RCC0, RCC1, RCC5 and SFMV. In all, 13 features return detection results above 86% for this task. Seven of the top performing results have been returned by features that are based on the real cepstrum.

A summary of the feature combinations returning the best results, 97% detection accuracy, is displayed in Table 8.3. Where numerous feature combinations have been returned, they are detailed in the indicated tables in Appendix B. Using more than eight features causes the results to continue to drop in accuracy and for this reason feature combinations using more than eight features are not displayed. In Table 8.3, the leftmost column gives the number of features used. The next two columns list the train and test set results obtained. The fourth column states the number of feature combinations achieving these results and in the rightmost column, the features are listed.

Through observing the feature combinations which have returned the best detection results for Task I, a pattern is revealed. The same features are regularly returned, giving rise to a cumulative feature effect, i.e. the successful features using lesser numbers of features exist within those using a greater number of features. This indicates much repetition, hence redundancy, within the best performing feature combinations. It also highlights the most useful features for the detection of beginner from professional standard legato note samples.

No	*Train%*	*Test%*	*No. Combinations*	*Features*
1	97	97	3	TM; MMV; CQTH9
2	97	97	7	TM,CQTH9; TM,PSD190; CQTH9,RCCM; TM,MMV; TM,RCCM; TM,SFMV; MMV,CQTH9
3	97	97	27	Table B1
4	97	97	32	Table B2
5	97	97	27	Table B3
6	97	97	10	Table B4
7	97	97	1	TM,CQTH9,SF,SFMV, RCCM,RCCV,PSD190
8	95	95	568	-

Table 8.3: Feature combinations returning the best detection results Task I.

From Table 8.3, features 1, TM, 5, MMV and 6, CQTH9, are the three most significant features and are also the ones which performed well individually in the classifier. From these results, Task I is achieved by using any of these three features only. Observing all the successful feature combinations returned, at least one of these

features is present in each combination. The presence of any one of these three features makes the other features effectively redundant. The only features in the successful Task I features displayed in the above table and in those in Appendix B, are the TM, CQTH9, PSD190, RCCM, SFMV, MMV and SF all of which perform well when applied directly to the data except for the SF.

When more than eight features are used to represent the dataset's samples, the detection accuracy results drop to 95% and a much greater number of successful feature combinations is returned. The same seven features important in the previous feature combinations remain dominant in these combinations as well. As the number of features increases further, the accuracy continues to drop and a gap emerges between the train and test set results, indicating that the feature combinations with more features are not as well suited for the detection task as those using a lesser number of features. How well the same features and feature combinations perform at fault detection is detailed next.

8.3.2 Task II Results

Fault detection via four-fold cross-validation is presented in this section. The complete monothetic results obtained for fault detection are displayed in Table B5 and in Table B6 in Appendix B. From these results, all the playing faults have been detected. Feature suitability is based on the strength of the detection results, the closeness of the train and test set results and if any other faults have been detected by a given feature combination. The smaller the gap is between the train and test set results, the more suitable a feature is for detecting a specific playing fault. What is observed in these tables is that, should a feature perform well at detecting one playing fault, it tends to be effective at detecting many other playing faults too. This complicates individual feature detection.

Fault	*Train%*	*Test%*	*Feature*	*Other Faults Detected*
CR	80	83	SF	All others except for NV
SK	82	81	SF	All others except for NV
NV	75 75	72 73	RCC1 RCC5	at least a 5% gap
INT	83	83	SF	All others except for NV
BB	**90**	**91**	**SF**	**All others except for NV**
XN	**88**	**88**	**SF**	**All others except for NV**
SE	82	81	SF	All others except for NV
BADS	85	87	SF	All others except for NV
BADE	78	80	SF	All others except for NV

Table 8.4: Individual playing fault detection monothetic results.

A summary of the best detection results obtained for each playing fault based on the monothetic results is displayed in Table 8.4. The leftmost column lists the playing fault, the detection results of which are given in the following two columns. The fourth column lists the feature used and the last column lists any other playing faults detected.

From the results displayed in Table 8.4, playing faults bow bouncing and extra note achieve the highest detection accuracy readings at 90% and 88% on their training and 91% and 88% on their testing sets respectively. Both playing faults are detected by the same feature, SF. From these results, feature 32, SF, detects all playing faults within approximately 80% to 90% detection accuracy except for player nervousness. This feature returns the poorest detection results for nervousness at approximately 68% detection accuracy, which is at least 10% lower than that obtained for crunching. The results for bow bouncing and extra note are close after which, a drop of at least 6% exists before any other playing fault is detected. Features 22 and 25, which are RCC1 and RCC5 respectively, detect player nervousness. These features also detect other playing faults but there is a drop of at least 5% between its detection and that of another fault.

The proximity between the detection results for these playing faults is not entirely unexpected given that multiple playing faults tend to occur simultaneously in the dataset's beginner note samples. This is well illustrated by the data collected from the listening tests. More specifically, Figure 3.5 which illustrates the perceived playing fault presence in each sample. Furthermore, Table 3.3 lists the perceived independent fault occurrence for the dataset's samples, which is comparatively very low to the overall fault presence. Information about the proportion of the overlapping perceived playing faults is displayed in Table 3.4. Based on the information collected about the dataset through the listening tests, identifying faults individually and independently was not expected to be evident. The classifier is identifying the presence of multiple playing faults together, as too have the listeners. From these initial results, the same features detect multiple playing faults. The fault detection results obtained by using feature combinations are presented next.

From the monothetic fault detection results, bow bouncing and extra note are the two most readily detectable playing faults. When more than one feature is used to represent the data, bow bouncing and extra note also returned the highest detection results using the same feature combinations. Information relating to the detection of these playing faults is presented in Table 8.5. They are presented together to highlight the proximity of their detection results. The feature combinations displayed in this table are those that returned the highest detection rates, whose train and test results are the closest and which have a workable gap between the fault detection rate and that the

other playing faults. Where feature combinations are numerous, they are listed in Appendix B.

From the results displayed, bow bouncing and extra note are detected simultaneously. The proximity between the results obtained for detecting bow bouncing and extra note is well displayed in Table 8.5. These two playing faults are similar in that bow bouncing can be considered as a form of extra note as, as the bow bounces along the string, the effect can be thought of as little additional notes. The results obtained point to combining these two playing faults and renaming them under a common name for detection purposes. Taking a closer look at these results shows that if the highest detection rates are returned for bow bouncing, the next fault detected is always extra note. Although these two faults are detected within a close proximity of each other, a much larger gap exists after this, one of approximately 5% for some and 10% for the ones listed in the last four lines in Table 8.5 before another fault is detected as indicated in parentheses in Table 8.5.

Fault	*Train%*	*Test%*	*No. Features*	*No. Combinations*	*Features*	*Gap%*
BB	90	89	3	1	TM,SFMM,AC	+0.4
XN	90	91	3	1	TM,SFMM,AC	+4.7
BB	88	86	3	6	Table B7	+0.6
XN	87	90	3	6	Table B7	+4.9
BB	90	90	4	1	TM,RCCV,SFMV,AC	+0.2
XN	90	93	4	1	TM,RCCV,SFMV,AC	+5.1
BB	90	90	5	1	MMV,RCCM,RCCV,SFM,AC	+0.6
XN	90	93	5	1	MMV,RCCM,RCCV,SFM,AC	+4.7
BB	88	86	6	8	Table B8	+0.6
XN	87	90	6	8	Table B8	+4.5
BB	88	86	7	4	Table B9	+0.6
XN	87	90	7	4	Table B9	+4.5
BB	88	86	8	4	Table B10	+0.6
XN	87	90	8	4	Table B10	+4.5
BB	85	85	3	3	TV,SF,PSD; TV,PSD,RCCS; SF,PSD,RCC2	+2.8 (+10)
BB	85	85	4	1	PSD,RCCS,RCC2,RCC27	+3.2 (+10)
BB	85	85	5	1	TV,TS,PSD,RCCS,RCC12	+2.8 (+10)
BB	85	85	6	2	TV,TS,SF,PSD,RCCS,RCC2; TS,SF,PSD,RCC2,RCC12,RCC27	+3.2 (+10)

Table 8.5: Bow bouncing and extra note detection results.

Regarding the features present in the successful feature combinations, two points are noted. The first, a cumulative feature effect, as displayed in the Task I results, is not present in these fault detection feature combinations. This makes tracking the presence of a specific feature difficult. The second point is an observation about the type of features that are present. Features that have been labelled as poor performing features in this text are more prevalent in the fault detection combinations. The features which fall into this category include AC, TV, TS, SF, SFM, PSD, RCC3, RCC12 and RCC27. Fault detection is dependent on the presence of such features whereas the Task I results require features that perform well at discriminating between the two different player types in their respective domains for a successful classification outcome. The results

show that bow bouncing and extra note have been detected simultaneously. Nervousness is also readily detectable, although to a lower detection accuracy level, the results of which are detailed next.

Nervousness is detected to approximately 73% detection accuracy through using many feature combinations which are displayed in Table 8.6. Where numerous, the feature combinations are given in Appendix B. Although detection results do not exceed 74% accuracy, a gap of at least 7% to other faults exists when three, five, six, seven and eight features are used. From these results, feature seven, RCC1, is the most prominent feature, present in all the feature combinations of interest. Features six and seven, RCC0 and RCC1 respectively, are also present in the successful feature combinations listed in Table 8.6. When detecting nervousness, increasing the number of features used does not increase the detection rates but the gap between the detection of player nervousness and that of any other fault widens, which improves feature combination suitability for the given task, noting that the displayed drop in accuracy is small. This is shown by the 12, 13 and 14 feature combinations listed in the table below. Taking the difference between the training and testing set results as well as the detection proximity to other faults into account means that using more features is better for detecting nervousness within these conditions. Nervousness is best detected to approximately 72% accuracy by using the 12, 13 and 14 feature combinations. These feature combinations returned the closest train and test set results as well as providing a gap of at least 11% between detecting faults.

Train%	*Test%*	*No. Features*	*No. Combinations*	*Features*	*Gap%*
74	76	3	1	RCCM,RCC0,RCC1	+7%
74	76	5	1	MMV,RCC0,RCC1,SFMV,SFMS	+7%
74	76	6	7	Table B11	+7%
74	76	7	2	TM,MMV,RCCV,RCC0,RCC1,RCC3,SFMS; MMV,RCCM,RCC0,RCC1,SFM,SFMM,SFMV	+7%
74	76	8	2	TM,MMV,RCCV,RCC0,RCC1,RCC3,SFMV,AC; TM,RCCM,RCC0,RCC1,SFMM,SFMV,SFMS,AC	+7%
73	71	12	1	TM,TK,CQTH9,PSD190,SFMM,SFMV,SCM190,RCCV,RCC0,RCC1,RCC5,CV	+12
72	71	13	1	TM,TK,CQTH9,SFMM,SFMV,SCM190,RCCM,RCCV,RCC0,RCC1,RCC5,CV,AC	+12
72	71	14	1	TM,TK,CQTH9,PSD190,SFMM,SFMV,SCM190,RCCV,RCC0,RCC1,RCC5,CV,AC	+11

Table 8.6: Player nervousness detection.

A total of 33 features and multiple feature combinations have been used to represent the data in the classifier. The features and feature combinations have been selected based on their individual features' ability to group, according to player type, the dataset's samples. For Task I, 97% accuracy has been achieved through many different feature combinations. From the Task II results, all playing faults are detected, but only nervousness is detected in a way that can be considered to be independent of the other

faults. In other words, detecting player nervousness with a sufficient gap before another playing fault is detected. Bow bouncing and extra note are the faults with the highest detection levels returned by the same feature combinations, making these two playing faults simultaneously identifiable. The detection results of these two playing faults point to renaming them under a common name for detection purposes.

Through using different feature representations of the data, the relationship between cluster centres and detection task has been observed. Regarding feature choice for Task I, three points are important when using a *k*-NN classifier. Firstly, how the feature performs directly on the data, in this case at grouping the samples according to player type. Secondly, that the different player samples do not return similar distances between their representations and cluster centres. And finally, the proximity of the clusters to each other, as determined by the feature used. From the results returned, a pattern in the detection results emerges reflecting feature choice. For Task I, the 'good' performing features are necessary and for fault detection, the inclusion of what is referred to in this text as 'poor' performing features improves the detection results of playing faults. For task I, increasing the number of features has not further facilitated overall fault detection nor the detection of individual faults as reflected by the results detailed. Relying on one feature only for either detection task makes the results less robust and more sensitive to outliers or erroneous data. For cluster stability reasons, using more than one feature is favoured. In the following section, feature combinations for both detection tasks are tested on new data. Although the results presented have been obtained from a four-fold cross-validation and therefore should generalise, testing the feature combinations on new data further tests the generality of the results.

8.4 Testing New Data

From the results obtained from the classifier via four-fold cross-validation in the previous section, multiple suitable feature combinations have been returned for both detection tasks. In this section, some of these feature combinations are tested on new data. This serves as a further check on the feature combinations' ability to generalise. Prior to testing on new data, a summary of the feature combinations returned from both tasks deemed to be of interest, is given.

From the Task I results, where present, the same features have been shown to be important for the detection of beginner and professional standard legato note samples. Feature combinations have returned detection results of 97% accuracy using one to

seven features. The results show much redundancy in that the combinations with more features include the successful combinations with lesser numbers of features. Rather than testing all of the feature combinations individually for Task I, three feature combinations have been selected and are displayed in Table 8.7. These feature combinations have constituent features that all perform well at differentiating between beginner and professional standard legato note samples in the dataset when applied directly to the data. They consist of the most significant features present in the other successful feature combinations too.

nf	*Important Feature Combinations for Task I*
5	TM, MMV, RCCM, RCCV,SFMV
6	TM, CQTH9, PSD190, SFMV, RCCM, RCCV
7	TM, MMV, RCCM, RCCV,SFMV, PSD190, CQTH9

Table 8.7: Task I feature combinations.

Developing playing, i.e. that which is associated with a beginner player, is detected much more readily and robustly than identifying individual playing faults resulting in fewer combinations. The feature combinations for fault detection tend to require a greater number of features than those for Task I. A cumulative feature effect, as observed in the Task I results, is not present to the same extent. The feature combinations needed for fault detection are listed in Table 8.8.

nf	*Fault*	*Features*
3	NV	RCCM, RCC0, RCC1
3	BB&XN	TM, SFMM, AC
4	BB&XN	TM, RCCV, SFMV, AC
4	BB	PSD, RCCS, RCC2, RCC27
5	BB	TV, TS, SF, PSD, RCCS
5	NV	MMV, RCC0, RCC1, SFMV, SFMS
5	BB&XN	MMV,RCCM,RCCV,SFM,AC
12	NV	TM,TK,CQTH9,PSD190,SFMM,SFMV,SCM190,RCCV,RCC0,RCC1,RCC5,CV
13	NV	TM,TK,CQTH9,SFMM,SFMV,SCM190,RCCM,RCCV,RCC0,RCC1,RCC5,CV,AC
14	NV	TM,TK,CQTH9,PSD190,SFMM,SFMV,SCM190,RCCV,RCC0,RCC1,RCC5,CV,AC

Table 8.8: Prominent fault detection feature combinations.

There are two ways of testing new data: threshold values and comparing clusters. Threshold values have been obtained from the trial dataset results based on how well a feature performs at separating the two player groups in the dataset. This works well where the dataset samples are grouped distinctly. Depending on the features returned from the classifier, applying threshold values is not always possible. In the clustering test, representative features are obtained from the new data sample based on the successful feature combinations from the classifier and the distances from the clusters provided by the original dataset are compared. Should a sample be positioned closer to the beginner cluster, it gets a beginner note label and likewise for the professional legato label. Both of these testing methods can have their sensitivity increased or decreased by altering the number of conditions met. The conditions referred to are determined by the

features selected. For example, a sample may meet the professional standard for four out of the five features but depending on the level set, this sample can be labelled as a beginner or as a professional. This allows a certain flexibility to the system by being able to compensate for occasional erroneous data values. Of the features presented in this work, relatively few discriminate with 100% accuracy the two different player groups. Many more group data correctly but have some overlapping values. So a best of x out of y feature values approach allows the testing sensitivity to be altered. To check their performance, these testing methods have been run on the dataset and the labels compared with those assigned to the samples through the listening tests, the results of which are displayed in Table 8.9. The five features used are TM, MMV, RCCM, RCCV and SFMV.

Method	***Initial Sensitivity (4 out of 5)***	***Increased Sensitivity (all 5)***
Threshold Method	163 out of 176 labels same as listeners	170 out of 176 labels same as listeners
	6 pro std leg note samples incorrectly labelled	5 pro std leg & 1 beginner note samples incorrectly labelled
Clustering Method	163 out of 176 labels same as listeners	170 out of 176 labels same as listeners
	13 pro std leg note samples incorrectly labelled	**All labels correct**

Table 8.9: Testing methods based on five features for Task I.

Prior to testing new data, the test data is presented. The legato note samples have been downloaded from the University of Iowa's Electronic Music Studio's Musical Instrument Samples [UofI]. These note samples are comparable to the professional standard legato note samples in the dataset. Beginner note samples have been obtained from different sources. One group consists of samples which had been obtained in as similar a way as possible to the original data at a later date, using the same players. This group is referred to as Begtest. The other samples have been obtained from two young students using a Sony monophonic microphone and recorded directly into Cool Edit Pro [CoolEditPro98]. The recordings of these players were taken in the living rooms of their homes. All rooms are quiet and have good practicing acoustics, i.e. not live. One student plays a half size violin and the other, a full size. Both students have been playing a few years and are capable of producing good sound and were asked to play open notes and scales with separate bows, stopping between notes. These two players are better than the beginner players used to make the dataset. Many of the notes produced by these students have good sound quality and few are dominated by playing faults. So returning only beginner labels for these samples is not expected. Before testing the new data, the samples were labelled by a professional violinist.

A summary of the new samples' results obtained when using the same five features are displayed in Table 8.10. The five features used in this testing procedure are the TM,

MMV, RCCM, RCCV and SFMV. Using the clustering method is favourable to the threshold approach as it yields more accurate results. To contrast the results and to check the sensitivity level set, the test was run on the dataset samples. In this case, the outcome of each feature goes towards determining the label assigned and this level can be altered. In the results displayed, at least four features need to be labelled professional player for the sample to get this label.

Sample List	*Player*	*Violin Size*	*Type*	*No. Samples*	*Task I Outcome*
U of I	Professional	Full	Legato	93	All labelled professional
Begtest	3 beginners	Full	Legato	121	All labelled beginner
Student_1	Student	Full	Legato	97	16 beg and 81 pro (listener: 25 beg and 72 good)
Student_2	Student	1/2	Legato	49	10 beg and 39 pro (listener: 11 beg and 38 good)
Dataset	5 players	Full	Legato	176	Correct labels; 6 different to listening test labels

Table 8.10: Test data samples based on at least four out of five features.

The six dataset samples which have different labels to those assigned by the listening tests are all professional standard legato notes which have been given beginner player labels. The Begtest and U of I samples have all been labelled correctly. In the student samples, the listener and cluster labels mostly overlap but the violinist listener picked up more subtleties in the notes judged which formed their decision, based on the feedback returned. The inclusion of a professional violinist's opinion serves as a guide. The student samples are mostly good and comparatively to the dataset's beginner note samples, only small faults are present. These faults do not dominate the sound samples. With the aim of improving label accuracy, the number of features is kept the same but the test sensitivity is changed by altering the conditions met. These conditions are now determined by the outcomes of all five features and the results are displayed in Table 8.11. A professional standard label is assigned should all features return values which fall within an acceptable distance from the dataset's professional standard legato note samples' cluster.

Sample List	*Player*	*Violin Size*	*Type*	*No. Samples*	*Task I Outcome*
U of I	Professional	Full	Legato	93	All labelled professional
Begtest	3 beginners	Full	Legato	121	All labelled beginner
Student_1	Student	Full	Legato	97	All labelled beginner (listener: 25 beg and 72 good)
Student_2	Student	½	Legato	43	All labelled beginner (listener: 11 beg and 80 good)
Dataset	5 players	Full	Legato	176	99 beg and 77 pro; 13 different to listening test labels

Table 8.11: Test data samples based on five features with maximum sensitivity.

Increasing the test sensitivity level now assigns beginner labels to all the student samples and to some of the dataset's legato note samples. The results in the previous table are better for Task I. Further labelling improvements are tested by using more features. The results using six features are given in Table 8.12 and reveal labelling accuracy levels which fall between those provided in the previous two tables. The six

features used in this testing procedure are the TM, MMV, RCCM, RCCV, SFMV and PSD190. In these results, the labelling depends on all six features.

Sample List	*Player*	*Violin Size*	*Type*	*No. Samples*	*Task I Outcome*
U of I	Professional	Full	Legato	93	All labelled professional
Begtest	3 beginners	Full	Legato	121	All labelled beginner
Studen_1	Student	Full	Legato	97	87 beg and 10 pro (listener: 25 beg and 72 good)
Student_2	Student	1/2	Legato	43	All labelled beginner (listener: 11 beg and 80 good)
Dataset	5 players	Full	Legato	176	99 beg and 77 pro; 13 different to listening test labels

Table 8.12: Test data samples based on six features with maximum sensitivity.

The test samples which are most like the dataset samples, U of I and Begtest, are correctly labelled. Listening to the Student_1 samples which have been labelled as professional player, nine have good sound quality but the tenth does not have acceptable sound quality throughout the duration of the note. The violinist has labelled these nine samples as good but not the tenth. An explanation for this discrepancy is that Student_1 is a strong player and some features, such as TM, are influenced by force and not only good quality strength[3]. The labelling returned is more severe than desired in these results based on how the dataset has been labelled, indicating a need to alter the test conditions set. Reducing the test sensitivity improves the labelling while using the same feature list, as displayed by the results given in Table 8.13. In these results, at least five feature values out of six need to fall within an acceptable proximity to the dataset's professional standard legato note samples' values to be given a professional player label.

Sample List	*Player*	*Violin Size*	*Type*	*No. Samples*	*Task I Outcome*
U of I	Professional	Full	Legato	93	All labelled professional
Begtest	3 beginners	Full	Legato	121	All labelled beginner
Student_1	Student	Full	Legato	97	17 beg and 80 pro (listener: 25 beg and 72 good)
Student_2	Student	1/2	Legato	43	10 beg and 39 pro (listener: 11 beg and 38 good)
Dataset	5 players	Full	Legato	176	All labelled correctly; 6 different to listening test labels

Table 8.13: Test data samples based on six features with decreased sensitivity.

The test setup used to obtain the results displayed in Table 8.13 correctly labels all the dataset samples. Compared to the labels obtained via the listening tests, six professional standard legato note samples have different labels. These are the six legato samples that the listeners have labelled as beginner player. The number of labels assigned by the testing procedure for the student samples is close to those given by the violinist but the sample labels need to be inspected. All but two of the 17 Student_1 samples which have been labelled as beginner player by the test have been given fault descriptions by the violinist. Two of Student_2's beginner player labels as determined by the test conflict with the descriptions given by the violinist and are reported as being

[3] See Figure D1 in Appendix D which displays the TM values for these different sample groups.

of good quality. The results obtained when the number of features is increased to seven are presented next.

The seven features used in the feature combination tested are the TM, MMV, RCCM, RCCV, SFMV, PSD190 and CQTH9. The results obtained when using the test settings that return the dataset samples labelled correctly are given in Table 8.14. The U of I and Begtest samples have all been labelled correctly. All of Student_2's and most of Student_1's samples are labelled beginner which is more severe labelling than that given by the violinist. Many of the students' samples are of acceptable quality. To obtain these results, at least six of the seven features require values close to those of the dataset's professional standard legato note samples' cluster centre to get a professional player label. Taking the labelling based on all seven features, which is not displayed, return results that are even more severe, labelling eleven of the dataset's legato note samples as beginner player.

Sample List	*Player*	*Violin Size*	*Type*	*No. Samples*	*Task I Outcome*
U of I	Professional	Full	Legato	93	All labelled professional
Begtest	3 beginners	Full	Legato	121	All labelled beginner
Student_1	Student	Full	Legato	97	87 beg and 10 pro (listener: 25 beg and 72 good)
Student_2	Student	½	Legato	43	All labelled beginner (listener: 11 beg and 38 good)
Dataset	5 players	Full	Legato	176	All labelled correctly; 6 different to listening test labels

Table 8.14: Test data samples based on seven features with reduced sensitivity.

Sample List	*Player*	*Violin Size*	*Type*	*No. Samples*	*Task I Outcome*
U of I	Professional	Full	Legato	93	All labelled professional
Begtest	3 beginners	Full	Legato	121	All labelled beginner
Student_1	Student	Full	Legato	97	17 beg and 80 pro (listener: 25 beg and 72 good)
Student_2	Student	½	Legato	43	10 beg and 39 (listener: 11 beg and 38 good)
Dataset	5 players	Full	Legato	176	86 beg and 90 pro; 8 different to listening test labels

Table 8.15: Test data samples based on seven features with further reduced sensitivity.

The labelling results obtained once the test conditions have been reduced further are displayed in Table 8.15. In these results, at least five of the seven features must return values which fall within proximity to those representing the dataset's professional standard legato note samples. The U of I legato note samples retain their professional player labels and two label changes have occurred in the dataset samples. Two beginner note samples have professional player labels. Listening to these two samples, their sound quality is reasonable and not dominated by playing faults. The labelling for the students' samples is better aligned with the labels given by the violinist. From the results displayed, the optimum results use only five or six out of the seven features.

Although using one to twelve features for Task I is possible based on the classification results returned, using five and six feature combinations is favoured. Most of these features, when applied directly to the dataset, group the different player types

accurately, but some samples have overlapping values. By increasing the number of features used, should a sample overlap in one domain, it may not in the others, allowing for more robust labelling. Detecting beginner from professional standard violin notes is best achieved by a computer using five and six feature combinations consisting of the TM, MMV, RCCM, RCCV, SFMV and PSD190. What is important in achieving an acceptable outcome is setting the test sensitivity correctly based on the number of features used and how the dataset samples or a control sample set is labelled.

Fault detection feature combinations are tested in a similar manner. Before these results are presented, it should be noted that the beginner student samples used as test data are of a higher standard than the dataset's beginner player samples, as a result of labelling. Not all samples contain playing faults and the ones that do, the faults are often not severe, i.e. the note is not dominated by playing faults. From the classification results, many possible combinations can be used in theory for specific fault detection. However, when tested, the combinations using the higher number of features performed better as a result of labelling and only these are presented in this section.

Fault	*% correct labelling*
CR	81
SK	83
NV	68
INT	83
BB	91
XN	91
SE	83
BADS	86
BADE	79

Table 8.16: Fault detection dataset labels.

Fault	*Begtest*	*Uofl*	*Student_1*	*Student_2*
NV	99%	5%	20%	19%

Table 8.17: Test data player nervousness detection.

The fault detection feature combinations selected for identifying playing faults were first run on the dataset samples to check accuracy before running them on the new data. The first feature combination tested consists of 13 features: TM, TK, CQTH9, SFMM, SFMV, SCM190, RCCM, RCCV, RCC0, RCC1, RCC5, CV and MCC3M. The percentage of labels matching the listening test ones for all faults are displayed in Table 8.16.

From the classification results, this feature combination detects player nervousness to approximately 73% accuracy. Applying this combination directly to the dataset detects all the playing faults with extra note and bow bouncing returning the highest detection results and nervousness, the lowest. When tested on the new data, the results

obtained are displayed in Table 8.17. The values displayed are the percentages of samples with the selected fault label.

Listening to the samples with the assigned fault labels, these samples do contain faults but not uniquely the fault selected for detection, as displayed by the Begtest samples. This is acceptable as faults tend to occur together, as they do in the dataset used. The U of I samples though do not have any playing faults yet eight samples have been incorrectly labelled as having playing faults. Both Student_1 and Student_2 are better players than those who provided the dataset's beginner note samples. This accounts for the lower nervousness or fault detection in these samples. These samples are not perfect but none is dominated by any one particular playing fault. Based on these results, this feature combination functions better at a more general fault detection level rather than specifically for detecting nervousness.

Fault detection has not been as successfully achieved as the detection of beginner versus professional player task. The samples returned do contain playing faults but not specifically the fault set out to be detected. There are several possible explanations for this. To begin with, the detection results returned by the classifier were lower for fault detection than those for Task I. Successful fault detection feature combinations detected multiple playing faults and the largest detection level gap between any two faults was approximately 10%. Violin playing faults are more often than not linked to each other, making the presence and identification of independent faults difficult. Although the results returned had been obtained through a four-fold cross-validation classification method, testing on new data permitted to further check the generality of the feature combinations and specific detection tasks. The feature combinations for Task I have been shown to generalise well.

8.5 Summary

A 1-NN classifier with *k*-means clusters has been used via four-fold cross-validation to detect beginner note samples from professional standard legato ones and for detecting playing faults. Multiple features and feature combinations have been tested and compared. The effect of timbre descriptor choice on classification outcome has been investigated and certain feature selections have been shown to be more advantageous than others when used in the classifier. Successful feature combinations were then tested on new data from a variety of sources to check their ability to generalise. The

outcomes of these tests show that the feature combinations selected generalise well for Task I. Playing faults have also been detected. Individual fault detection though is more difficult and has not always possible under the conditions set.

Detection accuracy of 97% for Task I has been returned via four-fold cross-validation. Within the numerous feature combinations achieving 97% detection, much redundancy is present. From the multiple feature combinations returning 97% detection accuracy for Task I, much feature repetition is present, illustrating a cumulative feature effect. Increasing the number of features above eight did not return any improvement in the detection results. The three most significant individual features for Task I via the classifier are the TM, MMV and CQTH9, all of which performed well at grouping the dataset's samples according to player type when applied directly to the data. The other important features for this task included the RCCM, PSD190 and the SFMV. Features that distinguished accurately the different player groups in the dataset did not always perform well in the classifier. PSD190 is one such feature. The advantage of using more than one feature in a feature combination is cluster stability. The more features that are used in defining a cluster centre, the less susceptible the classification process becomes to an erroneous or an outlying reading for any one feature. From the classification results returned by the various feature combinations, the effect on cluster design has been observed. Regarding feature choice for Task I, three points are important. Firstly, how the feature performs directly on the data, i.e. in this case, how effective it is at grouping the data according to player type. Secondly, the distances between the samples' representative values and the clusters and lastly, the proximity of the two clusters to each other. The closer the clusters are together or have similar distances for a majority of samples, the likelihood of incorrect labelling greatly increases. Although the classification results had been obtained through a cross-validation method, which in theory allows the outcomes to generalise, the successful feature combinations' generality was further tested on new data, which had been obtained from a variety of different sources. The results have been shown to generalise well for Task I. In certain instances, the computer has been shown to return more accurate labelling than the listeners.

Playing faults have been detected but isolating all specific faults was not possible. Bow bouncing and extra note have been detected simultaneously via multiple feature combinations and nervousness proved to be readily detectable too. From the features used, should one feature perform well at detecting a specific playing fault, it tended to

perform well at detecting the other faults too. In finding suitable fault detection feature combinations, attention had to be paid not only to detection accuracy, but also to the space between the fault's detection and that of another one. Feature selection based on the lack of ability to split the dataset into beginner and professional groups in their respective domains, is reflected favourably in the fault detection results. Mixed feature combinations as well as those using the poorer performing features returned the best results for fault detection within the dataset. From the results returned, specific fault detection remains difficult. This is due to playing faults mostly occurring together as reflected by the information collected via the listening tests. The combinations presented serve better at general fault detection rather than detecting any specific playing fault. In the following chapter, the research presented in this thesis is summarised and conclusions are drawn.

9 Conclusions

The research undertaken to investigate violin timbre has been presented in this thesis through the novel approach of analysing the relationship between violinist and sound produced. The aims of developing a way through which a system can differentiate between professional standard legato notes and those associated with a beginner player as well as detecting playing faults based on the information available in the waveform signal only were set. Work completed included inspecting, analysing and detecting nine main beginner playing faults: crunching, skating, player nervousness, intonation, bow bouncing, extra note, sudden end to note, poor start to note and poor finish to notes. Qualitative and quantitative analyses of violin timbre were required to glean a better understanding of how to best represent it. The approach taken has focused specifically on comparing typical beginner player notes to those played by professional standard violinists. Given the lack of existing research in the area, the work encompasses a range of topics including violin sound perception, signal analysis and representation, classification and testing.

The main steps in this work included the creation of a suitable dataset to facilitate establishing a link between the qualitative expressions and quantitative measures were established via listening tests. Through quantitative analysis, the dataset's samples have been represented by multiple features from the temporal, spectral and cepstral domains. The listening tests assigned qualitative labels to the dataset reflecting musicians' perception. Playing faults were defined and their presence or absence confirmed through the listening tests taken by professional standard string players. Information sought from the listeners included grading the overall sound quality, determining fault presence and labelling each sample as a beginner or as a professional player note. The listeners' consistency was verified by inspecting the range and means of values returned for each note sample before normalising the results to create an "average listener", which provided the *a priori* labels for use in the classifier.

The quantitative analyses included sourcing suitable features to represent the data effectively, capable of capturing the change in timbre due to the player. Much existing work on instrumental sound contrasts one instrument's timbre against that of another one and not typically for use within a specific timbre space. Numerous standard features

from the time, spectral and cepstral domains have been investigated for their efficacy at representing violin timbre within the dataset with varying levels of success.

The efficacy of each feature at grouping the different player types within the dataset separately is of great importance. Observing these standard representations for the dataset's samples led to inspecting the frequency region below the violin's lowest notes, G3 at 196Hz. This involved looking at the CQT frequency bins below 190Hz. Of these, nine frequency bins provided representation grouping the dataset's samples based on player type. The PSD below 190Hz was also effective at displaying the different player types in the dataset. It has been shown that the beginner samples contain more power in the unwanted frequencies, those below the violin's playing range, than the professional standard ones. Taking the SCM within the same frequency range also provided good results. The SCM190 results displayed a higher value for the beginner note samples than for the professional standard ones. Modifying these features to focus on the frequency content below 190Hz has returned useful results and has not previously been done in violin sound analyses.

Of all the features tested, a small number of features completely separate the beginner from the professional standard legato note samples in the dataset. They are the TM, MMV, SFMV, SCM190 and the mean CQT frequency bin content of nine specific bins below the lowest note. A further 12 features separate well the two player groups with less than ten samples overlap. These features come from the time, spectral and cepstral domains and include the TK, CV, PSD190, SFMM, SFMK, RCCM, RCCV, RCCK, RCC0, RCC1, RCC5 and MFCC3M. There is also a further category of results in which all samples overlap but there is an underlying pattern which reflects relevant information. In these representations, the beginner sample values cover a much wider range than the professional standard legato ones do, thus illustrating what musicians often say about beginner player violin notes, that they are inconsistent. This can be seen in the AC, MFCC0M, MFCC0V, MFCC0S, MFCC0K and in the MFCC0 and MFCC2 mean values of the first 0.87 seconds of the dataset's samples. The values obtained for the remaining features overlap, making distinguishing between the different player types in the dataset based on these representations not possible.

After sourcing features, different classification methods were considered. A *k*-means NN classifier was selected as it is a robust and effective classification method. Classification via a 1-NN classifier with *k*-means clusters using 33 different features and various feature combinations. All features and feature combinations were tested using

four-fold cross-validation resulting in some excellent results. The ability of these feature combinations to generalise was further tested on new data.

It has been shown that a computer can differentiate effectively between beginner and professional standard legato violin note samples with 97% accuracy. The results returned have shown that it is possible to achieve Task I by using one feature only. Basing a detection task on only one feature makes it sensitive to erroneous data and outliers, recalling that relatively few features have separated the dataset's samples based on player type with full accuracy. Using more than one feature to represent the data allows greater detection accuracy. Should a sample get an incorrect label from one feature, the values used to represent it based on the other features will rectify the situation, making for more robust detection. From the results obtained, improvements on this level of accuracy were not achieved by altering the feature choice or number. Where present, the same features have been returned indicating much redundancy in the successful feature combinations. Although four-fold cross-validation has been used to obtain the results, the ability to generalise was further tested on new data. A testing system was set up requiring five to seven features. The test sensitivity was set based on the number of conditions applied, which are determined by the feature combinations. The features used include the TM, MMV, RCCM, RCCV, SFMV, PSD190 and CQTH9. When the appropriate test sensitivity level had been selected, the results returned were found to be good indicating that the results generalise. Beginner notes have been successfully identified from the professional standard legato notes using at least five out of the seven features.

Task II, playing fault detection proved to be a much greater challenge than detecting beginner from professional standard legato notes. Faults have been detected, but individual fault identification has been shown to be much more difficult. The choice of feature combination plays an important role in fault detection as confirmed by all the results returned. Increasing the number of and using better performing features was shown to be less effective for fault detection than for determining professional standard from beginner notes. The presence of poor performing features in the feature combinations have proven to be better for Task II than using all the best performing ones. The poor performing features are those that did not differentiate between the dataset's different player types when applied directly to the data. Although using poor performing features returned the weakest Task I results, these features are beneficial for fault detection. The feature combinations returned for individual fault detection when

tested on new data, did not specifically detect the given fault but detected playing faults in general. Given that playing faults are rarely present in isolation in the dataset or in reality, these results are not unexpected.

The advantages of testing various features and feature combinations have allowed classifier performance to be confirmed in terms of consistency but more importantly, have helped to better understand the relationship between cluster design, detection task and the dataset's samples. For Task I, three points are important in cluster definition. The first, the features used perform well at differentiation between the two different player groups. Secondly, that a majority of samples do not return similar distances to the same clusters and finally, the presence of a suitable gap between the two cluster centres. Observing the most important features across all feature combinations tested for Task I revealed that the same features have been returned when present.

Task II requires the data to be represented by a greater number of features. These features do not need to meet the same conditions as those for Task I. A "cumulative feature effect" or feature redundancy is not observed in the fault detection results to the same extent as in the Task I results. In fault detection, the polythetic clusters tend to be quite different one from another with the feature combination significantly changing once another feature is added. This makes tracing a particular feature through the fault detection results difficult.

Improvements to the classification results were implemented by making the feature choice more diverse as well as increasing the number of features used. These changes did not return detection rates above 97% for Task I but certain changes permitted playing fault detection results to improve. Several approaches can be taken to improve classification results of which feature number and selection is just one. Changes can be made at different stages of the classification process from feature choice to dataset design. The most compatible with the results presented, is to find and use suitable features which more readily capture the subtly changing violin timbre. One possible example would be location specific features which quantify, for example, onset characteristics or pitch salience which is dependent on the steady-state section. These features might then facilitate the detection of certain location prevalent faults such as poor starts and finishes to notes. The success of such features though, is dependent on being able to determine effectively the attack, steady-state and decay regions of all violin note waveforms.

Alternatives to using new features include specific dataset design and the re-labelling of playing fault descriptions. It has been shown that playing faults rarely occur in isolation as beginners tend not to produce sound faults independently. This makes data collection a long and arduous process should one opt to use samples with only one fault present and for this reason improbable. Another way to improve detection results is to re-label the playing faults. Certain playing fault descriptions are quite similar and overlap as has been shown by the results obtained, which point to re-labelling the bow bouncing and extra note playing faults together.

An alternative to re-labelling the descriptions used which could be made to facilitate the testing procedure with the aim of improving accuracy, could be the use of fault gradations. Fault gradation could allow for more detailed information on how the dataset is perceived and would allow some of the comments made by the listeners to be taken into consideration. Including the feedback from the musicians would make the listening tests more detailed and precise. The listening tests could be redesigned, run and used in the classifier as it stands, but the dataset is likely to be too limited from which it could be difficult to obtain conclusive results. The down side of applying fault gradation is that a much larger dataset is needed; one including many more samples exhibiting varying levels of faults. With a larger dataset, listening test testing time and subsequently listener concentration and focus become issues. As presented, the listeners had to allow about an hour to complete the listening tests. Listener concentration at times appeared to fluctuate, which is why extending the testing time length or test size in one sitting is not advisable. Further work in this area needs to strike a balance between listener concentration and test size. Alternatively, fault specific test sets could be set up for example, a crunch test, whereby the listeners are asked only to state whether the selected fault can be perceived or not. This would involve a long, thorough dataset collection process which would not be evident as multiple playing faults are typically present in the majority of beginner note samples.

Various ways of improving the detection results have been suggested so far ranging from finding and using new features to redesigning the dataset and listening tests have been proposed. Another approach considers defining the distance from a cluster centre to which a label applies. In the work presented, both the professional standard legato and beginner note clusters have been inferred by the dataset as defined by the features used. Several possible methods of improving the results based on cluster design are proposed. One way is to decrease the acceptable professional standard player region.

This can be done by adding an intermediate or good beginner player cluster. This would allow samples which fall in the middle range between the two clusters to be labelled as beginner instead of professional, in theory leading to greater accuracy. This allows for the acceptable professional standard region to be reduced and that of the beginner one to be increased. An alternative to this, which may improve sample labelling, is to create a system whereby only the professional standard note clusters are defined by the dataset and let the beginner ones be determined by the user. This assumes that the user is a beginner violinist but more importantly allows the test sensitivity to be adaptable to suit the user. A record of this progress based on beginner cluster movement over a period of time can be kept. The beginner note samples as set by the dataset can be used as a default setting.

The research presented in this thesis can be easily modified and extended so that it can be used on other bowed stringed instruments. Apart from further applications and uses within the music information retrieval and analysis domains, more general uses include use within speech and language analysis. More specifically, taking a similar approach could be used in the development of articulation and language pronunciation tools by applying a more sonic rather than phoneme based analysis approach.

References

[Agostini01] G. Agostini, M. Longari, E. Pollastri, "Musical instrument timbres classification with spectral features", *IEEE Fourth Workshop Multimedia Signal Processing*, pp.97-102, 3-5 October 2001.

[Agositini03] G. Agostini, M. Longari, E. Pollastri, "Musical instrument timbres classification with spectral features", EURASIP *Journal on Applied Signal Processing* 2003:1, 1-11.

[AKG09] AKG "Monitor" Headphones product information [Online]. Available: http://www.akg.com/site/products/powerslave,id,429,pid,429,nodeid,2,_language,EN.html [Accessed: 12/12/09]

[d'Allessandro94] C. d'Allessandro, M. Castellongo, "The pitch of short-duration vibrato tones", *Journal of the Acoustical Society of America* 95(3):1617-30, 1994.

[Almeida04] A. Almeida, C. Vergez, R. Causse, X. Rodet, "Physical model of an oboe: comparison with experiments", *International Symposium on Musical Acoustics*, Nara, Japan, April 2004.

[Askenfelt89] A. Askenfelt, "Measurement of the bowing parameters in violin playing", *Journal of the Acoustical Society of America* 86(2):503-516, 1989.

[Auer80] L. Auer, *Violin Playing As I Teach It*, Dover Publications Inc., New York, 1980.

[Beauchamp82] J. W. Beauchamp, "Synthesis by spectral amplitude and brightness: matching analyzed musical sounds", *Journal of Audio Engineering Society* 30(6), pp. 396-406, 1982.

[Bensa04] J. Bensa, "Analysis synthesis of piano sounds using physical and signal models", workshop Physical Modeling: Future Directions at SARC 28/04/04

[Bishop95] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[Bissinger92] G. Bissinger, "Effect of f-hole shape, area and position of violin cavity modes below 2kHz", *Catgut Acoustical Society Journal*, 2:2(Series II), November 1992.

[Bissinger98] G. Bissinger, G. Gearhart, "A standardized qualitative violin evaluation procedure", *Catgut Acoustical Society Journal*, 3:6(series II):44-45, 1998.

[Bonada01a] J. Bonada, A. Loscos, P. Cano, X. Serra, "Spectral approach to the modeling of the singing voice", *Proceedings of the 111th Audio Engineering Society Convention*, New York, USA, 2001.

[Bonada01b] J. Bonada, O. Celma, A. Loscos, J. Ortolà, X. Serra, "Singing voice synthesis combining excitation plus resonance and sinusoidal plus residua models", *International Computer Music Conference* (ICMC), Havana, Cuba, 2001.

[Bonada03] J. Bonada, A. Lascos, "Sample-based singing voice synthesizers by spectral concatenation", *Stockholm Music Acoustics Conference* (SMAC03), Stockholm, Sweden 2003.

[Bows09] The bowed string [online]. Available: www.phys.unw.edu.au/jw/Bows.html, [Accessed: 10/12/2009].

[Bregman90] A. S. Bregman, *Auditory Scene Analysis: the Perceptual Organisation of Sound*, MIT Press, Cambridge, 1990.

[Brown91] J. C. Brown, "Calculation of a constant Q spectral transform", *Journal of the Acoustical Society of America*, 89, pp. 425-434, 1991.

[Brown96] J. C. Brown, K. V. Vaughn, "Pitch center of stringed instrument vibrato tones", *Journal of the Acoustical Society of America* 100(3):1728-35, 1996.

[Brown01] J. C. Brown

[Cherry53] E. C. Cherry, "Some experiments on the recognition of speech with one and two ears", *Journal of the Acoustical Society of America* 25(5):975-979, 1953.

[Cleveland77] T. F. Cleveland, "Acoustic properties of voice timbre types and their influence on voice classification", *Journal of the Acoustical Society of America*, 61:1622-1629, 1977.

[CoolEditPro98] Cool Edit Pro, Version 1.2, Syntrillium Software Corporation, 1998.

[Cremer84] L. Cremer, *The Physics of the Violin*, translated by T. Allen, MIT, London, 1984.

[Dannenberg85] R. B. Dannenberg, "An on-line algorithm for real time accompaniment", *Proceedings of the International Computer Music Conference* (ICMC), Paris, France, 1985.

[Davis80] S. B. Davis, P. Mermelstein, "Comparison of parametric representaions for monosyllabic word recognition in continuously spoken sentences", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28(4):357-366,1980.

[Deller00] J. R. Deller, J. H. L. Hansen, J. G. Proakis, *Discrete-Time Processing of Speech Signals*, IEEE Press, John Wiley & Sons Inc., 2000.

[Dodge97] C. Dodge, T. A. Jerse, *Computer Music: Synthesis, Composition and Performance*, Schirmer Books, New York, 1997.

[Duda73] R. O. Duda, *Pattern Classification and Scene Analysis*, Wiley, London, 1973.

[Efron83] B. Efron, "Estimating the error rate of a prediction rule: improvement on cross validation", *Journal of the American Statistical Association*, 78(382):316:331, 1983.

[Eronen00] A. Eronen, A. Klapuri, "Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2000.

[Eronen01] A. Eronen, "Comparison of features for musical instrument recognition", *Proceedings of IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, New Paltz, New York, 2001.

[Flesch00] C. Flesch, *The Art of Violin Playing*, NY, 2000.

[Fletcher98] N. Fletcher, T. D. Rossing, *The Physics of Musical Instruments*, Springer, London, 1998.

[Fritz04] C. Fritz, "La clarinette et le clarinettiste: influence du conduit vocal sur la production du son", PhD thesis, L'université de Paris 6 and the University of New South Wales, December 2004.

[Fritz06] C. Fritz, I. Cross, B. C. J. Moore, J. Woodhouse, "Perceptual correlates of violin acoustics", *Proceedings of the 9th International Conference of Music Perception and Cognition* (ICMPC), Bologna, 22-26 Aug. 2006 (pre-publication copy).

[Fritz07] C. Fritz, I. Cross, B. C. J. Moore, J. Woodhouse, "Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin", *Journal of the Acoustical Society of America*, 122(6),2007.

[Gill84] D. Gill (ed.), *The Book of the Violin*, Rizzolli International Publications, Inc., New York, 1984.

[Giordano04] N. Giordano, M. Jiang, "Physical modelling of the piano", *European Journal of Applied Signal Processing*, 7:926-933 2004.

[Gordon99] M. Gordon, P. Pathak, "Finding information on the World Wide Web: the retrieval effectiveness of search engines", *Information Processing and Management*, 35:141-180, 1999.

[Grey77] J. M. Grey, "Multi-dimensional perceptual scaling of musical timbres", *Journal of the Acoustical Society of America*, 61:5:1270-1277, May 1977.

[Hämäläinen04] P. Hämäläinen, T. Mäki-Patola, V. Pulkki, M.Airas, "Musical computer games played by singing", *Proceedings of the 7th International Conference on Digital Audio Effects* (DAFx-04), Naples, Oct. 5-8, 2004.

[Harrera98] P. Harrera, J. Bonada, "Vibrato extraction and parameterization in the spectral modeling synthesis framework", 1998.

[Harrera00] P. Harrera, X. Amatriain, E.Batlle, X. Serra, "Towards instrument segmentation for music content description: a critical view of instrument classification" (accessed 14/05/03 at http://ciir.cs.umass.edu/music2000/papers/herrera_abs.pdf)

[Hawley93] M. Hawley, "Structured Sound", PhD thesis, MIT, 1993.

[Herzberg83] P. A. Herzberg, *Principles of Statistics*, Wiley, New York, 1983.

[Hindemith40] P. Hindemith *The Craft of Musical Composition*, trans. Schott's, Mainz, 1940.

[Howard01] D. M. Howard, J. Angus, *Acoustics and Psychoacoustics*, 2nd edition, Focal Press, Oxford, 2001.

[Hunt99] M. J. Hunt, "Spectral signal processing for ASR", *Proceedings Automatic Speech Recognition and Understanding*, 1999.

[Hutchins90] C. M. Hutchins, *Journal of the Acoustical Society of America* 87(1), pp. 392-397, 1990.

[Hutchins93] C. M. Hutchins, "Mode tuning for the violin maker", *Journal of the Catgut Acoustical Society*, 2:4:5-9, 1993.

[Hutchin97] C. M. Hutchins (ed.), V. Benade (ass. Ed.), Research Papers in Violin Acoustics 1975-1993, Vols. I & II, Acoustical Society of America, Woodbury NY, 1997.

[Jackson87] B. G. Jackson, J. Berman, K. Sarch, *The A.S.T.A. Dictionary of Bowing Terms for String Instruments*, American String Teachers Association, 3rd edition, Tichenor Publishing Group, Bloomington, Indiana, 1987.

[Jain88] A.K. Jain, R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, 1988.

[Jain89] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989.

[Jansson97] E. V. Jansson "Admittance measurements of twenty-five high quality violins", *Acustica*, 83:337-341, 1997.

[Jayant84] N. S. Jayant, P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs NJ, 1984.

[Jiang02] Jiang, D-N, Lu, L., Zheng, H-J, Tao, J-H., Cai, L-H, "Music type classification by spectral contrast feature", *Proceedings IEEE International Conference on Multimedia and Expo*, Vol. 1:113:116, 2002.

[Kay88] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[Klapuri01] A. Klapuri, T. Virtanen, A. Eronen, J. Seppanen, "Automatic transcription of musical recordings", Consistent and Reliable Acoustic Cues Workshop, CRAC-01, Aalborg, Denmark, 2001.

[Klapuri04] A. Klapuri, “Signal processing methods for the automatic transcription of music”, PhD thesis, Tampere University of Technology, March 2004.

[Kohonen90] T. Kohonen, “The self-organizing map”, *Proceedings of the IEEE* 78:9:1464-1480, September 1990.

[Krzanowski94] K. J. Krzanowski, F. C. H. Marriott, *Kendall’s Library of Statistics 1: Multivariate Analysis*, Arnold, 1994.

[Krzanowski95] K. J. Krzanowski, F. C. H. Marriott, *Kendall’s Library of Statistics 2: Multivariate Analysis* Part 2, Arnold, 1995.

[LABROSA] Lab Rosa [Online]. Available: http://labrosa.ee.columbia.edu/ [Accessed: 12/10/09]

[Logan00] B. Logan, “Mel cepstral coefficients for music modeling”, 2000.

[Logan01] B. Logan, A. Salomon, “A music similarity function based on signal analysis”, *Proceedings IEEE International Conference on Multimedia and Expo*, 2001.

[Loscos04] A. Loscos, J. Bonada, “Emulating rough and growl voice in spectral domain”, *Proceedings of the 7th International Conference on Digital Audio Effects* (DAFX04), Naples, Italy, Oct. 5-8, 2004.

[LSO09] London Symphony Orchestra samples. [online] http://www.notionmusic.com/products/notion3/sounds.html [accessed: 12/10/09]

[Machlis90] J. Machlis, *The Enjoyment of Music*, 6th edition (standard version), Norton & Co., New York, 1990.

[Marshall85] K. D. Marshall, “Modal analysis of a violin”, *Journal of the Acoustical Society of America*, 77(2), Feb. 1985.

[Martin98] K. D. Martin, Y. E. Kim, “Musical instrument identification: a pattern-recognition approach”, *136th Meeting Acoustical Society of America*, October 1998.

[Matlab04] Matlab 7, Version 7.0.0.19920 (R14), 2004.

[McLennan01] J. E. McLennan, “The soundpost in the violin. Part II: The effect of soundpost position on peak resonance and sound output”, *Journal of the Australian Association of Musical Instrument Makers* XX(1), March 2001.

[McLennan03] J. E. McLennan, “The function of f-holes in the violin”, *Journal of the Australian Association of Musical Instrument Makers* XXII(3), September 2003.

[McGill09] McGill University Master Samples [Online]. Available: http://www.music.mcgill.ca/resources/mums/html/ [Accessed: 12/10/09]

[Meek02] C. Meek, W. Birmingham, “Johnny can’t sing: a comprehensive error for sung music queries”, University of Michigan, Advanced Technologies Laboratory, *Proceedings of the International Computer Music Conference*, 2002.

[Mellody00] M. Mellody, G. H. Wakefield, “Time-frequency characteristics of violin vibrato: modal distribution and synthesis”, *Journal of the Acoustical Society of America*, 107(1):598-611, 2000.

[Miller75] J. R. Miller, E. C. Carterette, “Perceptual space for musical structures”, *Journal of the Acoustical Society of America*, 58:3:711-720, September 1975.

[Molin90] N.-E. Molin, A. O. Wahlin, E. V Jansson, “Transient wave response of the violin body” (Letters to the Editor), *Journal of the Acoustical Society of America*, 88:5:2479:2481, November 1990.

[MMO09] Music Minus One [Online]. Available: www.musicminusone.com/ [Accessed: 12/10/09]

[MPO09] Music Plus One [Online]. Available: http://xavier.informatics.indiana.edu/~raphael/music_plus_one/ [Accessed: 12/10/09]

[Moore82] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 2nd edition, Academic Press, London, 1982.

[Narmour92] E. Narmour, *The Analysis and Cognition of Melodic Complexity: The Implication – Realization Model*, University of Chicago Press, 1992.

[Noll93] P. Noll, "Wideband speech and audio coding", *IEEE Communications Magazine*, November 1993.

[Oppenheim89] A. V. Oppenheim, R. W. Schafer *Discrete-Time Signal Processing*, Prentice-Hall Inc., New Jersey, 1989.

[Oppenheim99] A. V. Oppenheim, R. W. Schafer *Discrete-Time Signal Processing*, 2nd edition, Prentice-Hall Inc., New Jersey, 1999.

[Pollastri02a] E. Pollastri, "A pitch tracking system dedicated to process singing voice for music retrieval", IEEE 2002.

[Pollastri02b] E. Pollastri, "Some considerations about processing singing voice for music retrieval", *Proceedings of the 3rd International Conference on Music Information Retrieval* (ISMIR), 2002.

[Potkonjak97] M. Potkonjak, K. Kim, R. Karri, "Methodology for behavioural synthesis-based algorithm-level space exploration: DCT case study", Design Automation Conference, 1997.

[Prame94] E. Prame, "Measurements of the vibrato rate of ten singers", *Journal of the Acoustical Society of America* 96(4):1979-84, 1994.

[Prame97] E. Prame, "Vibrato extent and intonation in professional western lyric singing" *Journal of the Acoustical Society of America* 102(1):616-21, 1997.

[Puterbaugh09] J. Puterbaugh, Timbre Time Line [Online]. Available: http://www.music.princeton.edu/~john/timbretimeline.htm [Accessed: 10/12/09].

[Raphael03] C. Raphael, "Orchestral musical accompaniment from synthesized audio", *Proceedings of the International Computer Music Conference*, 2003.

[RWC09] Real World Computing [Online]. Available: http://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i.html [Accessed: 10/12/09].

[Sacksteder87] R. Sacksteder, "How well do we understand Helmholtz Resonance?", *Journal of the Catgut Acoustical Society*, No. 48, November 1987.

[Sadie01] S. Sadie (ed.), The New Grove Dictionary of Music and Musicians, 2nd Ed., Macmillan Publishers Limited, 2001.

[Schafer02] M. Schafer speaking at Journées Design Sonore à Paris, 20-21 March, 2002, general information at http://confs.loa.espci.fr/DS2002/

[Scheirer96] E. Scheirer, M. Slaney, "Construction and evaluation of a robust multi-feature speech and signal processing", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1997.

[Schönberg78] A. Schönberg, *Theory of Harmony*, trans. R. E. Carter, Faber and Faber, London, 1978 (originally published in 1922).

[Serafin01] S. Serafin, M. Burtner, C. Nichols, "Expressive controllers for bowed string physical models", *Proceedings of the 3rd International Conference on Digital Audio Effects*, Limerick, 2001.

[Shifrin03] J. Shifrin, W. Birmingham, "Effectiveness of HMM-based retrieval on large data bases", *Proceedings of the 4th International Conference on Music Information Retrieval*, 2003.

[Shonle80] J. I. Shonle, K. E. Horan, "The pitch of vibrato tones", *Journal of the Acoustical Society of America* 67(1):246-52, 1980.

[Smithsonian09] Digital Stradivari: computer models of violins reveal master luthier's techniques [Online]. Available: http://smithsonianscience.org/2009/11/digital-

stradivari-computer-models-of-violins-reveal-the-master-luthiers-secrets/ [Accessed: 12/10/09]

[SOM] The SOM Toolbox [Online]. Available: http://www.cis.hut.fi/somtoolbox/ [Accessed: 12/10/09]

[Stevens40] S. Stevens, J. Volkman, "The relationship of pitch to frequency", *American Journal of Psychology*, 53:329, 1940.

[Stuart87] A. Stuart, J. K. Ord, *Kendall's Advanced Theory of Statistics: Distribution Theory* (Volume I), 5th Edition, Charles Griffin and Co. Ltd. London, 1987.

[Stuart91] A. Stuart, J. K. Ord, *Kendall's Advanced Theory of Statistics: Classical Inference and Relationship* (Volume II), Edward Arnold, London, 1991.

[Stumpf03] K. Stumpf, *Tonpsychologie*, MA: Adamant Media Corporation, Boston, 2003 (originally published 1883 and 1890).

[Sundberg87] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, Dekalb, IL, 1987.

[Suzuki73] S. Suzuki, *The Suzuki Concept: An Introduction to a Successful Method for Early Music Education*, Diablo, Berkeley, 1973.

[Suzuki09] The Suzuki Method [Online]. Available: http://suzukiassociation/teachers/twinkler/ [Accessed: 10/12/09]

[Szigeti79] J. Szegeti, *Szegeti on the Violin*, Dover Publishers, First Edition, NY, 1979.

[Tanguiane93] A. S. Tanguiane, *Artificial Perception and Music Recognition* Lecture Notes in AI 746, Springer-Verlag, 1993.

[TiePie09] TiePie Handyscope HS4 instrument description [Online]. Available: http://www.tiepie.com/uk/products/External_Instruments/USB_Oscilloscope/Handyscope_HS4.html [Accessed: 10/12/09]

[Tzanetakis02] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals", *IEEE Transactions on Speech and Audio*, 10:5: 293-302, July 2002.

[UofI09] The University of Iowa Electronic Music Studios Musical Instrument Samples [Online]. Available: http://theremin.music.uiowa.edu/MIS.html [Accessed: 12/10/09]

[Vercoe85] B. Vercoe, M. Puckette, "Synthetic rehearsal: training the synthetic performer", *Proceedings International Computer Music Conference*, 1985.

[Vergez06] C. Vergez, P. Tisserand, "The BRASS project, from physical models to virtual instruments: playability", *Lecture Notes in Computer Science*, Vol. 2006 No. 3902, May 2006.

[Violin09] Violin parts [Online]. Available: www.violinstudent.com/violinmap.html [accessed: 12/10/09]

[VSL09] Vienna Symphony Library [Online]. Available: http://www.ilio.com/ [Accessed: 12/10/09].

[West04] K. West, S. Cox, "Features and classifiers for the automatic classification of musical audio signals", *Proceedings of the 5th International Conference on Music Information Retrieval* (ISMIR), 2004.

[Wilson02] R. S. Wilson, "First Steps Towards Violin Performance Extractions Using Genetic Programming" pp. 253-62 in J. R. Koza (ed.) *Genetic Algorithms and Genetic Programming at Stanford 2002*, Stanford, California, 2002.

[Winckel67] F. Winckel, *Music, Sound and Sensation: A Modern Exposition*, Dover, NY, 1967.

[Woodhouse04] Why is the violin so hard to play? [Online]. J. Woodhouse, P.M. Galluzzo, "Why is the violin so hard to play?" *Plus* 31 October 2004. Available: http://plus.maths.org/issue31/features/woodhouse/index.html [Accessed: 01/02/10].

[Yoo98] L. Yoo, D. S. Sullivan Jr., S. Moore, I. Fujinaga, "The effect of vibrato on response time in determining the pitch relationship of violin tones", *Proceedings of the 2nd International Conference of Music Perception and Cognition* (ICMPC), 1998.

[Young95] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, *The HTK Book (for HTK Version 3.4)*, Dec. 1995, revised Dec. 2006.

[Young08] D. Young, "Classification of common violin bowing technique using gestural data from a playable measurement system", *Conference on New Interfaces for Musical Expressions* (NIME08), Genoa, 2008.

[Youngberg79] J. Youngberg, "Rate/pitch modification of speech using the constant Q transform", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), Vol. 4:748-751, 1979.

Author's Publications

[Charles09] Charles, J., Fitzgerald, D., Coyle, E. "Violin sound classification", 17th Telecommunications Forum TELFOR 2009, Belgrade, Serbia, November 24-26, 2009.

[Charles08] Charles, J., Fitzgerald, D., Coyle, E. "Violin sound quality detection", Irish Systems and Signals Conference, NUI Galway, June 18-19, 2008.

[Charles06] Charles, J., Fitzgerald, D., Coyle, E. "Quantifying real violin sound", DMRN, Goldsmith College, London, July 22-24, 2006.

[Charles06] Charles, J., Fitzgerald, D., Coyle, E. "Violin timbre space features", Irish Signals and Systems Conference, Dublin Institute of Technology, 28-30 June 2006.

[Charles05] Charles, J., Fitzgerald, D., Coyle, E. "The violin timbre space", IT&T Annual Conference, Maritime Institute, Cork Institute of Technology, Cork, 26-27 October 2005.

[Charles05] Charles, J., Fitzgerald, D., Coyle, E. "Development of a computer based violin teaching aid, ViTool", Audio Engineering Society, Barcelona, May 28-31, 2005.

[Charles04] Charles, J., Fitzgerald, D., Coyle, E. "Towards a computer assisted violin teaching aid", International Symposium on Psychology and Music Education, PME04, Padua, Italy, Nov. 29-30, 2004.

Appendix A: CQT Frequency Bin Content

The CQT mean content from frequency bins four, five and six are illustrated in Figure A1, showing higher mean values for the professional standard legato notes than for the beginner ones.

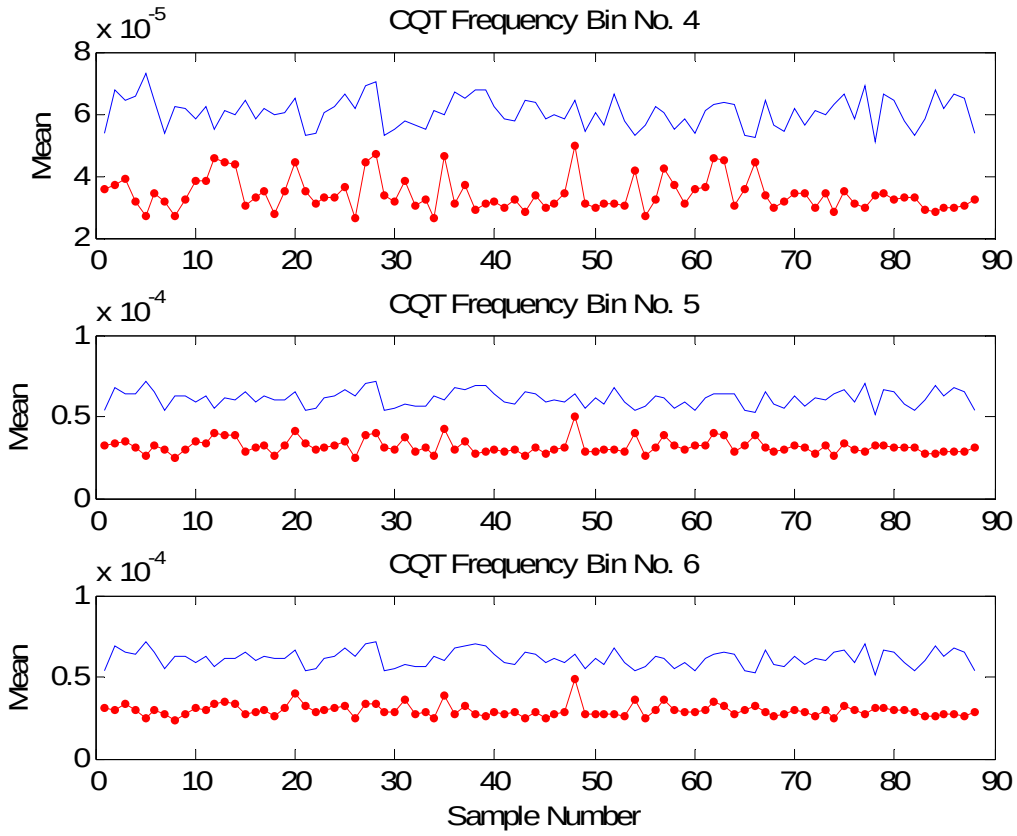

Figure A 1: Mean content CQT frequency bin numbers four (top), five (middle) and six (bottom).

Figure A2 and Figure A3 display respectively the mean CQT content from frequency bins seven, eight and nine and frequency bins ten, eleven and twenty.

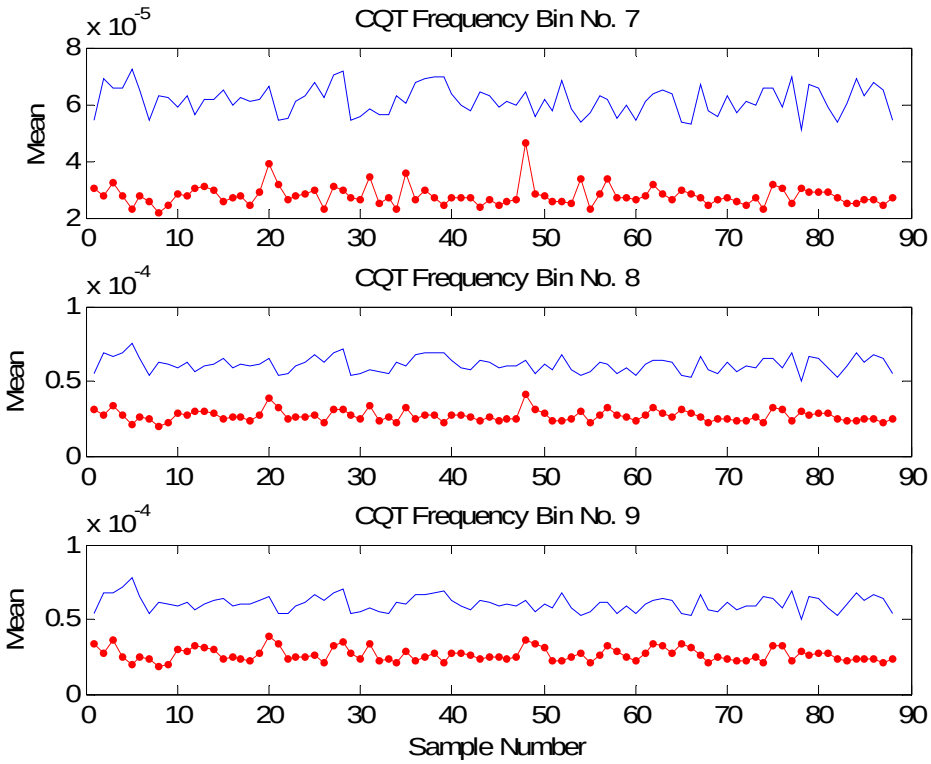

Figure A 2: Mean content CQT frequency bin numbers seven (top), eight (middle) and nine (bottom).

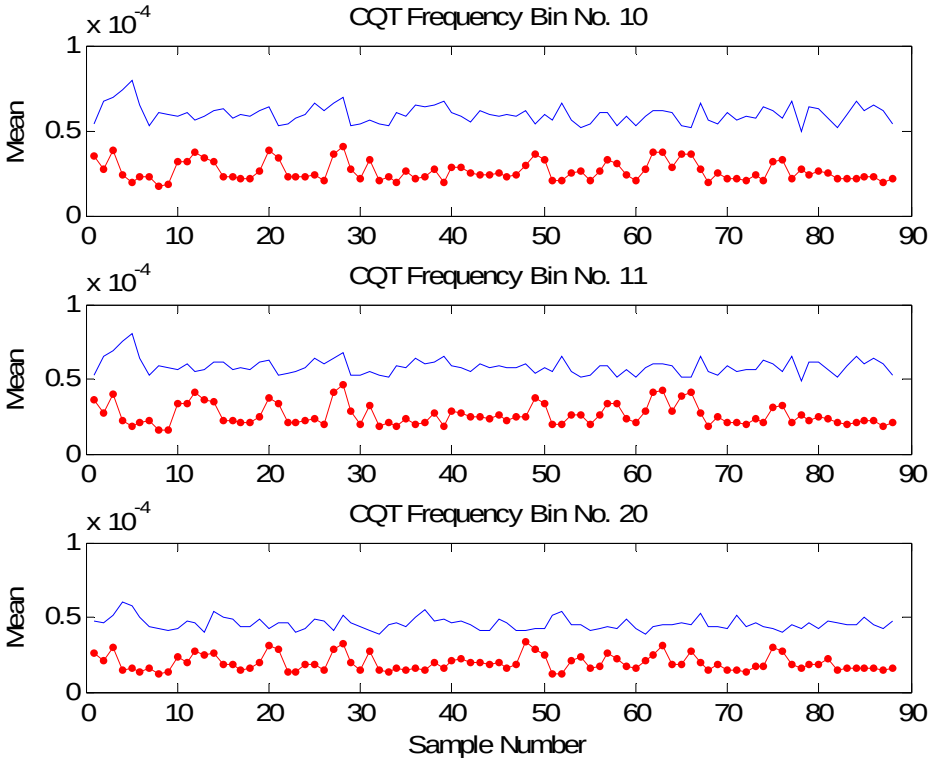

Figure A 3: Mean content CQT frequency bin numbers ten (top), eleven (middle) and twenty (bottom).

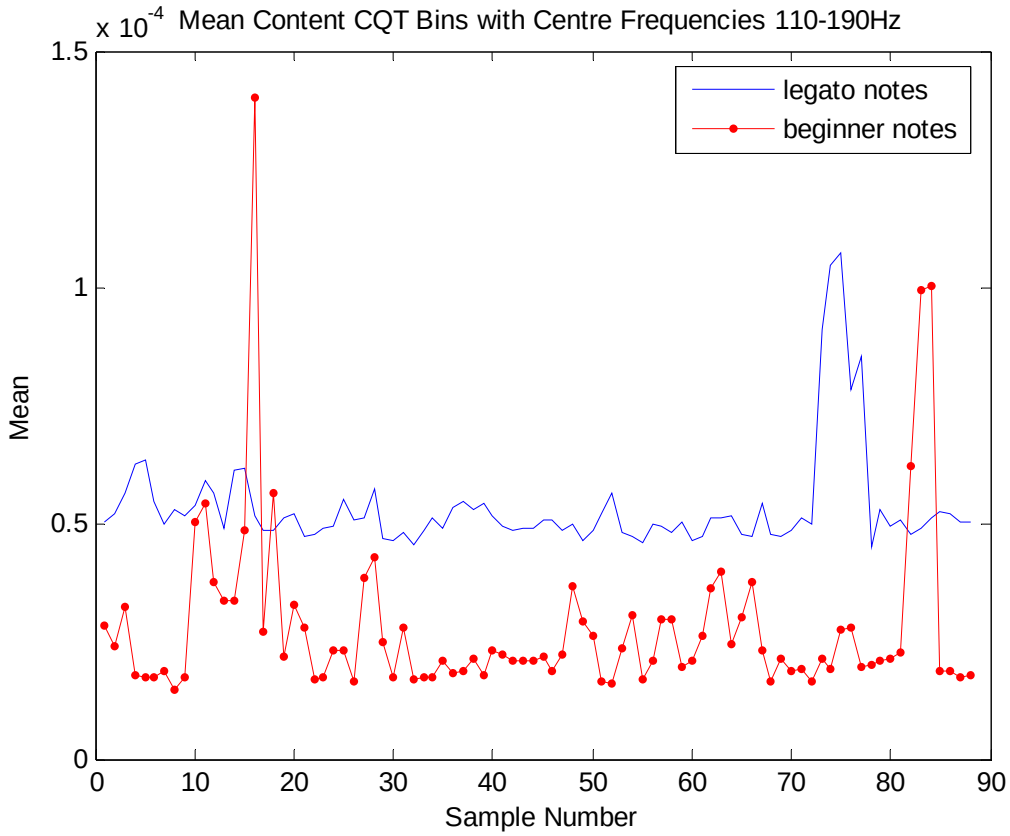

Figure A 4: Mean content CQT frequency bin numbers 1 to 39 (110-190Hz).

Appendix B: Feature Combinations

In this section, the successful feature combinations too numerous to list in Chapter 8 are displayed.

Features	*Train %*	*Test %*
TM,CQTH9,PSD190	97	97
TM,CQTH9,SFMV	97	97
TM,CQTH9,RCCM	97	97
TM,CQTH9,RCCV	97	97
TM,PSD190,SFMV	97	97
TM,PSD190,RCCM	97	97
TM,PSD190,RCCV	97	97
TM,SFMV,RCCM	97	97
TM,SFMV,RCCV	97	97
TM,RCCM,RCCV	97	97
CQTH9,PSD190,RCCM	97	97
CQTH9,SFMV,RCCM	97	97
TM,MMV,RCCM	97	97
TM,MMV,RCCV	97	97
TM,MMV,SFMV	97	97
MMV,RCCV,SFMV	97	97
TM,MMV,PSD190	97	97
TM,MMV,CQTH9	97	97
MMV,RCCM,CQTH9	97	97
MMV,PSD190,CQTH9	97	97
TM,CQTH9,SF	97	97
TM,SF,SFMV	97	97
TM,SF,RCCM	97	97
TM,SF,RCCV	97	97
TM,SF,PSD190	97	97
CQTH9,SF,RCCM	97	97
CQTH9,SF,PSD190	97	97

Table B 1: Best three feature combinations from Table 8.3.

Features	*Train %*	*Test %*
TM,CQTH9,PSD190,SFMV	97	97
TM,CQTH9,PSD190,RCCM	97	97
TM,CQTH9,PSD190,RCCV	97	97
TM,CQTH9,SFMV,RCCM	97	97
TM,CQTH9,SFMV,RCCV	97	97
TM,CQTH9,RCCM,RCCV	97	97
TM,PSD190,SFMV,RCCM	97	97
TM,PSD190,SFMV,RCCV	97	97
TM,PSD190,RCCM,RCCV	97	97
TM,SFMV,RCCM,RCCV	97	97
CQTH9,PSD190,SFMV,RCCM	97	97
TM,MMV,RCCM,RCCV	97	97
TM,MMV,RCCM,SFMV	97	97
TM,MMV,RCCV,SFMV	97	97
TM,MMV,RCCM,PSD190	97	97
TM,MMV,RCCM,CQTH9	97	97
TM,MMV,RCCV,PSD190	97	97
TM,MMV,RCCV,CQTH9	97	97
TM,MMV,SFMV,PSD190	97	97
TM,MMV,PSD190,CQTH9	97	97
MMV,RCCM,SFMV,CQTH9	97	97
MMV,RCCM,PSD190,CQTH9	97	97
TM,CQTH9,SF,SFMV	97	97
TM,CQTH9,SF,RCCM	97	97
TM,CQTH9,SF,RCCV	97	97
TM,CQTH9,SF,PSD190	97	97
TM,SF,SFMV,RCCM	97	97
TM,SF,SFMV,RCCV	97	97
TM,SF,SFMV,PSD190	97	97
TM,SF,RCCM,RCCV	97	97
TM,SF,RCCM,PSD190	97	97
TM,SF,RCCV, PSD190	97	97

Table B 2: Best four feature combinations from Table 8.3.

Features	*Train %*	*Test %*
TM,CQTH9,PSD190,SFMV,RCCM	97	97
TM,CQTH9,PSD190,SFMV,RCCV	97	97
TM,CQTH9,PSD190,RCCM,RCCV	97	97
TM,CQTH9,SFMV,RCCM,RCCV	97	97
TM,PSD190,SFMV,RCCM,RCCV	97	97
TM,MMV,RCCM,RCCV,SFMV	97	97
TM,MMV,RCCM,RCCV,CQTH9	97	97
TM,MMV,RCCM,SFMV,PSD190	97	97
TM,MMV,RCCM,SFMV,CQTH9	97	97
TM,MMV,RCCM,PSD190,CQTH9	97	97
TM,MMV,RCCV,PSD190,CQTH9	97	97
TM,MMV,RCCV,PDD190,CQTH9	97	97
TM,MMV,SFMV,PSD190,CQTH9	97	97
MMV,RCCM,SFMV,PSD190,CQTH9	97	97
TM,CQTH9,SF,SFMV,RCCV	97	97
TM,CQTH9,SF,SFMV,PSD190	97	97
TM,CQTH9,SF,RCCM,RCCV	97	97
TM,CQTH9,SF,RCCM,PSD190	97	97
TM,CQTH9,SF,RCCV,PSD190	97	97
TM,CQTH9,SFMV,RCCM,RCCV	97	97
TM,CQTH9,SFMV,RCCM,PSD190	97	97
TM,CQTH9,SFMV,RCCV,PSD190	97	97
TM,SF,SFMV,RCCM,RCCV	97	97
TM,SF,SFMV,RCCM,PSD190	97	97
TM,SF,SFMV,RCCM,PSD190	97	97
TM,SF,RCCM,RCCV,PSD190	97	97
CQTH9,SF,SFMV,RCCM,PSD190	97	97

Table B 3: Best five feature combinations from Table 8.3.

Features	*Train %*	*Test %*
TM,MMV,RCCM,RCCV,SFMV,PSD190	97	97
TM,MMV,RCCM,RCCV,SFMV,CQTH9	97	97
TM,MMV,RCCM,RCCV,PSD190,CQTH9	97	97
TM,MMV,RCCM,SFMV,PSD190,CQTH9	97	97
TM,MMV,RCCV,SFMV,PSD190,CQTH9	97	97
TM,CQTH9,SF,SFMV,RCCM,RCCV	97	97
TM,CQTH9,SF,SFMV,RCCM,PSD190	97	97
TM,CQTH9,SF,SFMV,RCCV,PSD190	97	97
TM,CQTH9,SF,RCCM,RCCV,PSD190	97	97
TM,SF,SFMV,RCCM,RCCV,PSD190	97	97

Table B 4: Best six feature combinations from Table 8.3.

In Table B5 and Table B6, the monothetic results, via four-fold cross-validation, for playing fault detection are displayed.

f	*CRtrain %*	*CRtest %*	*SKtrain %*	*SKtest %*	*NVtrain %*	*NVtest %*	*INTtrain %*	*INTtest %*	*BBtrain %*	*BBtest %*
1	63	63	63	65	74	73	61	61	56	59
2	51	52	53	44	53	43	53	45	50	48
3	20	18	16	19	32	30	19	18	10	12
4	56	55	54	56	66	65	53	55	48	50
5	73	77	79	80	69	73	77	78	83	84
6	63	63	63	65	74	73	61	61	56	59
7	51	52	53	44	53	42	52	45	50	49
8	21	19	18	19	31	29	18	18	9	13
9	27	24	25	27	39	36	26	24	18	20
10	28	26	26	26	39	35	26	23	19	19
11	58	58	56	58	69	68	55	57	50	53
12	22	21	18	20	33	31	19	19	12	16
13	62	63	61	64	71	72	58	60	55	59
14	44	45	41	43	51	52	42	44	37	41
15	38	38	37	40	47	45	40	40	31	35
16	43	40	42	40	49	46	37	36	38	39
17	64	62	63	64	72	71	64	64	57	60
18	53	53	50	53	63	62	50	49	44	46
19	65	63	63	65	72	72	65	65	58	60
20	53	53	51	53	63	62	50	51	44	46
21	67	66	63	65	72	71	68	69	60	60

22	66	65	67	69	75	72	67	68	59	63
23	41	39	41	44	53	52	40	40	34	37
24	67	70	65	66	69	70	65	65	67	70
25	68	66	68	69	75	73	67	66	62	65
26	58	61	60	63	60	64	62	63	63	62
27	38	40	36	41	45	47	34	36	34	39
28	58	60	59	55	55	56	57	60	59	64
29	45	41	48	45	44	44	47	43	46	43
30	57	53	61	61	65	63	62	58	58	60
31	62	63	63	64	59	57	62	65	64	64
32	80	83	82	81	69	71	83	83	90	88
33	18	16	18	20	31	27	17	17	10	13

Table B 5: Monothetic fault detection results for crunch, skate, nervousness, intonation and bow bouncing.

f	*NXtrain %*	*XNtest %*	*SEtrain %*	*SEtest %*	*BADStrain %*	*BADStest %*	*BADEtrain %*	*BADEtest %*
1	57	56	60	61	62	61	65	66
2	52	52	53	47	53	49	51	48
3	9	7	18	18	14	13	21	21
4	49	48	53	54	53	52	58	59
5	83	88	75	74	80	83	74	76
6	57	56	60	61	62	61	65	66
7	52	53	52	46	53	49	51	49
8	11	8	19	19	15	13	22	21
9	19	16	24	23	23	21	30	29
10	20	16	25	25	24	20	30	28
11	51	49	55	57	55	54	59	60
12	13	11	18	19	18	17	24	23
13	56	55	60	62	59	60	62	64
14	36	32	40	44	41	40	44	43
15	33	33	35	38	37	38	40	41
16	38	33	41	38	41	36	42	39
17	60	59	64	65	63	60	66	65
18	46	43	49	49	49	47	55	55
19	61	59	65	66	64	61	67	66
20	46	44	50	50	50	48	56	56
21	61	63	67	68	65	64	67	69
22	63	61	62	63	67	66	66	65
23	34	33	38	40	39	38	44	45
24	62	61	67	66	65	67	65	70
25	65	64	63	63	66	64	70	67
26	61	63	63	63	59	61	56	59
27	32	33	35	38	35	39	38	46
28	60	61	58	61	58	60	56	58
29	48	49	49	48	46	43	48	44
30	59	57	59	56	62	61	61	58
31	65	69	61	60	64	64	61	65
32	88	91	82	81	85	87	78	80
33	11	9	17	18	14	12	21	20

Table B 6: Monothetic fault detection results for extra note, sudden end, bad start and bad end to note.

From Table 8.5 feature combinations for fault detection using three, six, seven and eight features are in Table B7, Table B8, Table B9, Table B10 and Table B8 respectively.

Features	*Fault*
TM,RCCM,SFMS	BB & XN
TM,SFMM,SFMS	BB & XN
MMV,SFM,SFMS	BB & XN
RCCM,RCCV,SFMS	BB & XN
RCC3,SFM,SFMS	BB & XN
SFM,SFMV,SFMS	BB & XN

Table B 7: Best three feature combinations detecting bow bouncing and extra note from Table 8.5.

Features	*Fault*
TM,MMV,RCCM,RCC3,SFMM,SFMS	BB & XN
TM,MMV,RCCM,RCC3,SFMS,AC	BB & XN
TM,MMV,RCCM,SFMV,SFMS,AC	BB & XN
TM,RCCV,SFM,SFMM,SFMS,AC	BB & XN
MMV,RCCM,RCCV,SFM,SFMS,AC	BB & XN
MMV,RCCV,RCC3,SFMM,SFMV,SFMS	BB & XN
RCCM,RCCV,SFM,SFMM,SFMV,SFMS	BB & XN
RCC3,SFM,SFMM,SFMV,SFMS,AC	BB & XN

Table B 8: Best six feature combinations detecting bow bouncing and extra note from Table 8.5.

Features	*Fault*
TM,MMV,RCC3,SFM,SFMM,SFMS,AC	BB & XN
TM,MMV,RCC3,SFM,SFMV,SFMS,AC	BB & XN
TM,RCCV,RCC3,SFMM,SFMV,SFMS,AC	BB & XN
MMV,RCCM,RCCV,SFM,SFMM,SFMS,AC	BB & XN

Table B 9: Best seven feature combinations detecting bow bouncing and extra note from Table 8.5.

Features	*Fault*
TM,RCCM,RCCV,RCC3,SFM,SFMM,SFMS,AC	BB & XN
TM,RCCM,RCCV,RCC3,SFMM,SFMV,SFMS,AC	BB & XN
TM,RCCM,RCC3,SFM,SFMM,SFMV,SFMS,AC	BB & XN
MMV,RCCM,RCCV,RCC3,SFMM,SFMV,SFMS,AC	BB & XN

Table B 10: Best eight feature combinations detecting bow bouncing and extra note from Table 8.5.

Features	*Fault*
TM,MMV,RCCM,RCC0,RCC1,SFMM	NV
TM,RCCM,RCCV,RCC0,RCC1,SFMM	NV
TM,RCCM,RCC0,RCC1,SFMM,SFMS	NV
TM,RCCV,RCC0,RCC1,SFM,AC	NV
TM,RCC0,RCC1,RCC3,SFM,SFMS	NV
RCCM,RCC0,RCC1,SFM,SFMV,AC	NV
RCCV,RCC0,RCC1,SFM,SFMM,SFMV	NV

Table B 11: Nervousness detection feature combinations using six features from Table 8.6.

Appendix C: Further Real Cepstral Coefficients

The thirteenth and twenty-eighth real cepstral coefficients for the dataset's samples are displayed in Figure C1 and in Figure C2 respectively. These figures illustrate that neither coefficient serves well as a discriminator between the beginner and professional standard legato note samples in the dataset.

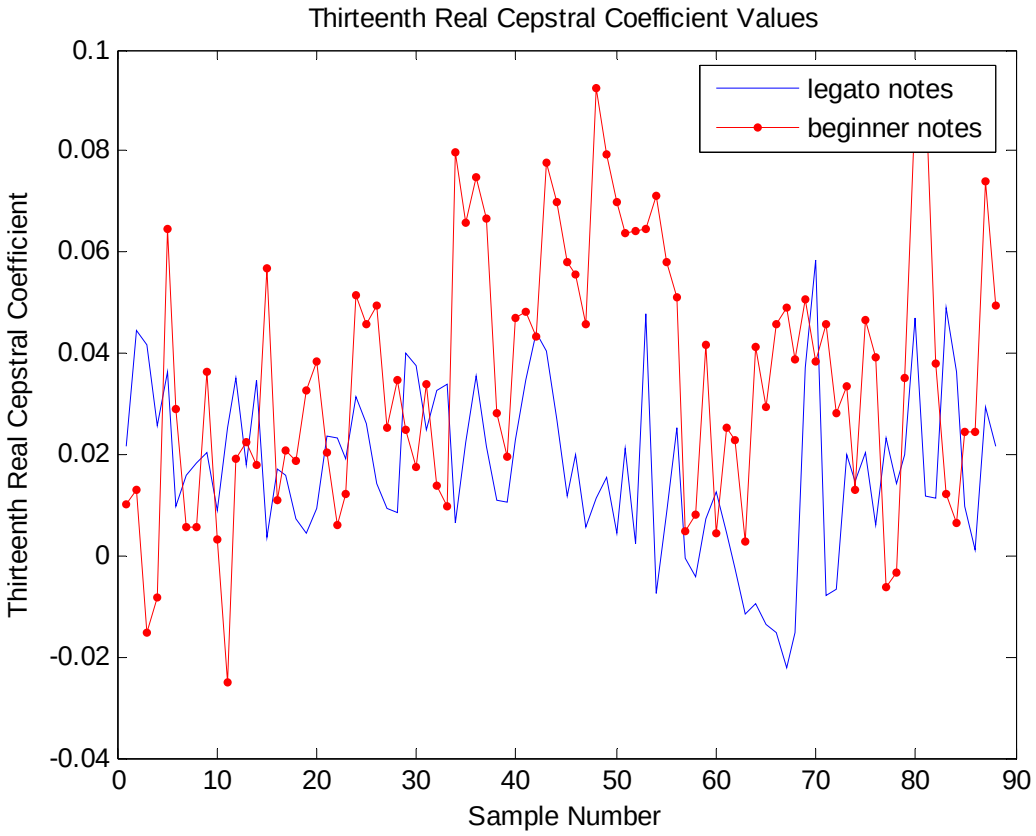

Figure C 1: Thirteenth real cepstral coefficient values.

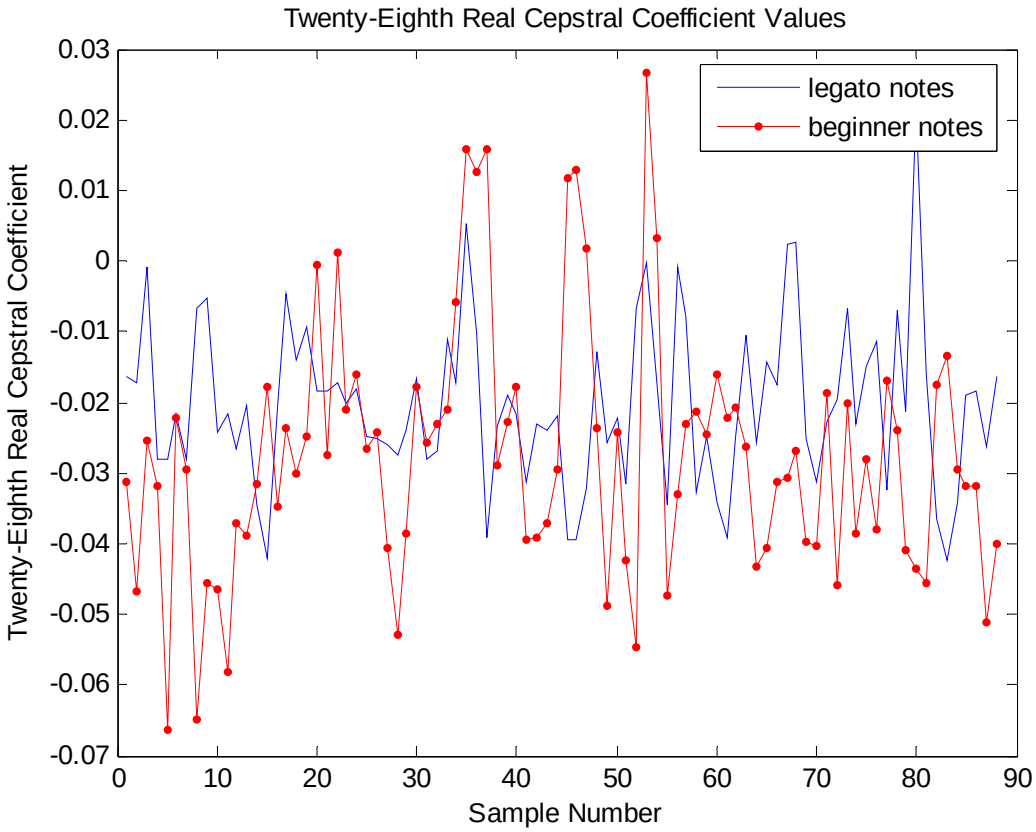

Figure C 2: Twenty-eighth real cepstral coefficient values.

Appendix D: Waveform Amplitude Mean and New Data

Figure D1 displays the waveform amplitude mean values for the new data. It had been included to highlight the difference in mean values between the dataset's professional standard legato note samples and those reflecting the Student_1 and Student_2 samples.

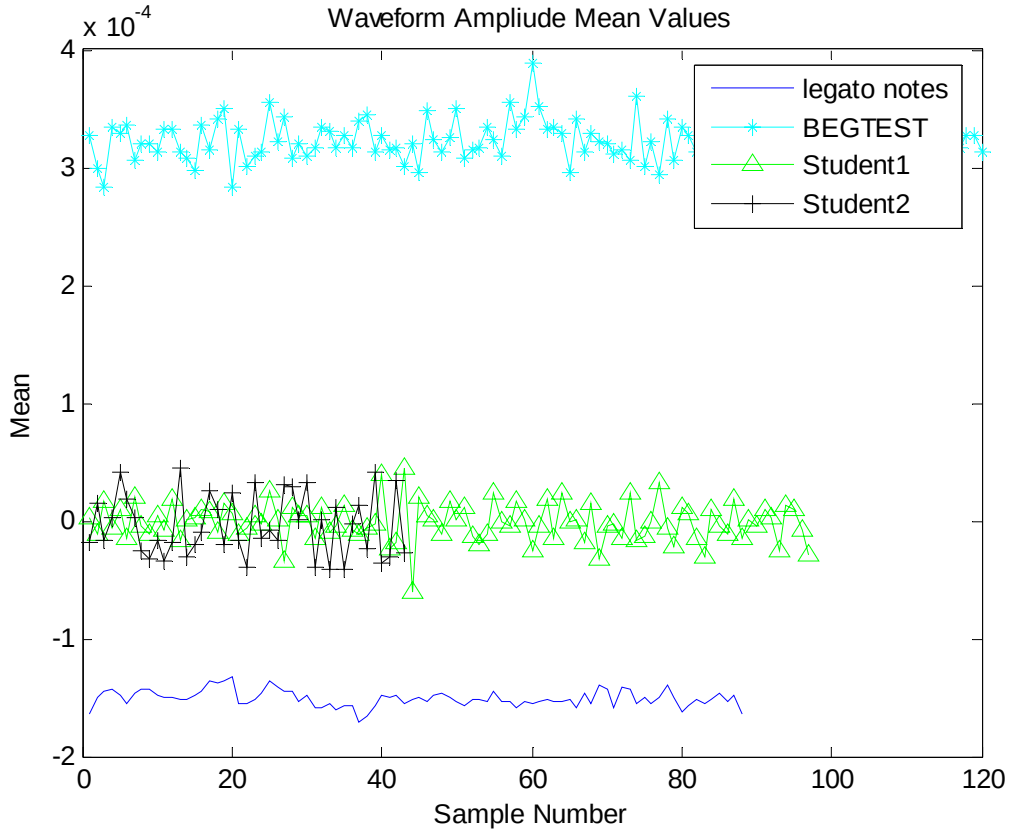

Figure D 1: Waveform amplitude mean values for different sample player groups.

Appendix E: CD Contents

Data Folder	→	Beginner	contains all beginner note samples used
	→	Professional	contains all professional standard legato note samples used

Code Folder	→	cep.m	% real cepstral coefficients
		cepm.m	% Mel cepstral coefficients
		centstats.m	% 1st order statistics from spectral centroid
		CQTfbinj.m	% this is Judith Brown's code modified for eighth tone spacing rather than quarter tone spacing
		first20CQTfreqbins.m	% 1st 20 CQT frequency bin content means; uses Judith Brown's CQT code above, see http://www.wellesley.edu/Physics/brown/
		getMELonsets.m	% Mel cepstra for onset period
		getwavstats.m	% 1st order statistics from waveforms
		getrccstats.m	% 1st order statistics from real cepstral coefficients; also use for Mel cepstral coefficients statistics – select as wanted
		movavgw.m	% waveform moving average statistics
		scm.m	% spectral contrast measure
		sfmstats.m	% 1st order statistics from spectral flatness measure
		specflux.m	% waveform spectral flux