CHAPTER 1. INTRODUCTION

1.1 Purpose and Research Strategy

The purpose of the current work is to produce useful techniques for extracting and recognizing certain features in music. Our primary interest is in rhythm, distinguishing amongst various types of percussive note events, and characterizing these feature sets for determining what makes some music have a swing feeling, while other music does not.

1.1.1 Time and Frequency

Sound is typically described in terms of time and frequency. The human auditory system, like laboratory or recording studio devices, is stimulated by vibrations of air molecules against some type of transducer (eardrum, microphone) which converts the air vibrations into another form, such as electrical signals in circuitry, or nerve impulses, which are both electrical and chemical in nature. The patterns in circuitry are easily analyzed using Digital Signal Processing (DSP) techniques, and this information is a useful framework for understanding the details that gives the swing feel to music.

Information in the time domain (e.g. input audio stream) can be converted to information in the frequency domain (pitches, or tones). An event with a one millisecond (1/1000 of a second) repetition rate corresponds to a frequency of one thousand cycles per second (1000 Hz, or 1 Khz). The range of frequencies audible to the human ear is approximately 20 Hz to 20,000 Hz (20 KHz). Much of the information of interest for music and speech is in the range of 100 Hz to 5000 Hz.

A set of frequencies derived from a time domain input stream is called the *spectrum* of the input data. The time and frequency forms of information are mathematically equivalent, but often one form is more convenient than the other to use for a particular

purpose. The spectrum is closely related to how we perceive sounds. This is explored in chapter 4.

1.2 Musical Audio Events

Most popular music can be broadly described in terms of rhythm and pitch, which are encapsulated in *musical events*. Rhythm is the temporal relationships of musical events. Pitch is the simple frequency content of these events. Timbre is a complex variant of pitch, and is used to describe the *qualities* of the sound, enabling distinction between trumpet and piano for example, even if they play the same musical note, or pitch.

Not all changes in music are adequately described in terms of separate events. Many forms of music have important features that change smoothly from one set of frequencies to another, or that smoothly modulate the loudness or pitch of a note. These changes are often subtle and more difficult to analyze than sharp percussive events. We have focussed on recognizing and analyzing distinct percussive note events, but our techniques can be extended for analyzing these more subtle musical patterns. We believe these subtle changes are highly correlated with human emotional response to music and consider this an important area for future research.

For the current work we recognize musical events in terms of rapid changes (faster than fifty milliseconds) of pitch and power level[1]. These changes are generally complex rather than simple. They are derived either from broad portions of the frequency spectrum, or specific subsets of correlated frequencies.

1.2.1 Note Identification

A musical note event is characterized by a rapid power change (as short as one millisecond) in a set of frequencies, called ***onset*** or ***attack,*** followed by a longer period of mostly steady frequencies, generally called ***decay-sustain-release*** (DSR) collectively.

---

[1] Power level is essentially the same as amplitude for the purpose of this thesis. The typical definition of power is amplitude squared (amplitude times amplitude). The shape of the waveform changes somewhat, but the features we are extracting are about the same.

The DSR period for percussion events ranges from tens to hundreds of milliseconds. Onset is often associated with large changes in loudness or power of the audio signal. This large quick change is characteristic of most percussive note events, although there are exceptions, which we discuss later.

We have developed computer algorithms for extracting and identifying a variety of percussive note events. We visually analyze the events representing a musical sample in order to discern temporal relationships between notes. These temporal relationships are the fundamental nature of rhythm.

1.2.2 Rhythm

After several note events are identified in terms of pitch, onset time and duration, they can be represented as a time series and the next level of information extraction, rhythm, can be performed. Once we have an informational representation of the rhythm, we can use it to characterize the ***style*** of the music. In particular, we investigate ***swing*** vs. ***straight*** time. Swing rhythm is found in Jazz, Blues, and many other styles with African roots including Cuban, Brasilian, and Caribbean music. Straight time is typified by some classical European music styles. These are not the only two forms of rhythmic style. Many examples of music exist that have temporal variations that are not well classified as either swing or straight time.

As a convenience, we refer to straight time, especially musical tablature based on the standard European notation, as Mozart-Bach (MB) notation or time. This is not to say that Mozart and Bach did not employ rhythmic variation and expressiveness in their music, but merely that the standardization of pitch and timing notation can be traced to that era in European music (1700's).

(Bengsston, 1987) presents a perspective on the features and limitations of CCMN (Current Common Musical Notation), which is essentially what we are calling MB notation. He observes that learning by notation rather than by experience can impede a young

musician's ability to play authentically, due to the misinterpretation of the printed information which of course is only a guide to the musical data, not to the performance.

We distinguish between *rhythm* which means the temporal data of a set of musical notes (e.g., as written in tablature), and *rhythmic style* which is a form of expression that a human performer may use when playing the musical data. Style is generally indicated by a linguistic comment on the musical score such as *rubato* or *with a swing feel*. This is a form of meta-information meaningful to an educated performer who is familiar with the particular comment and style. Such linguistic comments are essentially useless to some one who is unfamiliar with either the music style or the meaning of the comment.

In chapter 5 we present a variation of standard MB notation that we believe is useful and informative for conveying the feeling of various swing styles, both to skilled musicians and beginners. Performance of music using variations of the temporal patterns as written in MB notation (1/4 notes, 1/8 notes and so on) is called *rhythmic expression*. Recently, computer algorithms have been developed that emulate the temporal variations as played by a human performer, i.e., swing and other rhythmic expression.

1.2.3 Related Work

We review a number of prior works in rhythm processing and music recognition. We also look at psychological research on human perception of music and time. We have discovered that reading old material, even if it is technically weak or obsolete, can be useful for several reasons. First, some of our own currently cherished dogmas about what is important may appear absurd to future researchers similarly to how we may consider the work of earlier researchers to be naive or ignorant. This can help produce an understanding of the evolution of knowledge in a complex technical area like music analysis, and can also facilitate the open mind that is essential for good scientific research.

Sometimes old insights or observations can lead to very useful ideas when put in a modern context. (Strawn, 1985) includes lengthy discussion about whether 12-bit encod-

ing of audio information is adequate for reproducing high fidelity music, and questions if 16-bit encoding is needed or a waste of compute resources. We see this as silly today, but it leads to the idea of analysis of compute costs for pattern recognition work, which may be quite practical using 12-bit audio. Moorer in (Strawn, 1985) includes very strong technical opinions based on the idea of exact frequency sets derived from Fourier Analysis, forgetting apparently that Fourier Analysis is merely a model for data and information, and should not be mistaken for the information *per se.* The frequency analysis strategy of the human ear is quite different from the results generated by Fourier Analysis, and this provides useful ideas for DSP and pattern recognition, like using instantaneous frequency metrics or nonharmonic Fourier series rather than standard Fourier series in the extraction of musical features. Ideas like these can lead to better quality algorithms, lower compute costs for similar results, or both. (Strawn, 1985) also reports that *vocoder* (voice encoder) techniques had improved substantially since the 1960's, primarily because the newer vocoders use phase information from the audio, whereas the earlier vocoders did not. Similarly, although our current work, and most or all of the research literature, ignores phase information available from spectral analysis, we believe that there is much useful potential in this discarded information. There is some evidence for believing the human ear takes advantage of phase information in separating and distinguishing information that comes from difference sources, even though they are completely blended in the input stream.

1.2.4 Psychology and Perception

Music is fundamentally a human experience. Physicists and psychologists have studied human perception of music since the 19th century. There is evidence for an underlying commonality of temporal perception in humans, and spontaneous production of rhythmic patterns that starts in early childhood. This work is useful in demonstrating, for example, that most commonly used tempos in music are within a temporal range that exists at a low level in the human perceptual apparatus, independent of music itself (Fraisse,

1982). Eventually this sort of research may shed light on human perception at the neurological level, working from the cognitive levels outwards toward the pattern recognition and data collection systems of the audio cortex. We think that percussion and rhythm sounds provide a simple and tractable approach to mapping the human auditory system from the outside in, much as flashing dots on a computer screen have proved useful for mapping the human visual cortex. In the 21st century, this is a practical field of research.

An important aspect of music that distinguishes it from other human symbol systems such as language, is the close connection between emotional response and the perception of musical performance. While it can be reduced to symbolic notation on a page, the essence of music is found in expressive live performances and perception of these. Much modern pop music is produced using robotic sequencing software, but this music does not evoke the complex emotional responses from the human information system that music performed by human beings does. The motion picture industry has a highly developed infrastructure to support the production of music that evokes these important and often subconscious viewer reactions. Rhythmic expressiveness is an important aspect of such music, and our technical analysis of rhythm might be used to help quantify features that are correlated with different emotional responses.

Emotions are not generally considered a part of computer science, but of course they are an important part of the human information system. To this end, our work focuses on immediately practical techniques such as those that can let a computer serve as a technical tutor for humans to better learn, play and understand rhythmic complexities and subtleties, and thus better enjoy music. Music is sometimes used as part of medical therapy for emotional and psychological issues. Emotional well being (or lack of it) is a multi billion dollar industry in the modern world, and we suggest that learning to play and appreciate music better is a useful alternative to pills and therapy.

1.2.5 Computer Science in Music

Computers are used for many purposes related to music, but we focus primarily on two: computer analysis of music, and computer production of music. Practical production of music preceded detailed analysis techniques, but both are fairly mature now. We look at the state of the art in computer production of swing feeling in music, which is a feature available in some commercial software products. We present our research and results for computer analysis of music, and explore immediately practical applications to music production software.

The two information science techniques most commonly used for music analysis are DSP (digital signal processing) and pattern recognition. Various DSP techniques can render a stream of raw audio data (e.g., a format like CD audio, which is one or more channels of 16 bit integer data points sampled at 44,100 points/second) into a different format, such as a frequency spectrum, which is more useful for a particular purpose. We typically use the frequency spectrum of a music sample to identify note events from specific instruments. Pattern recognition, like DSP, is a field with many techniques. We currently use a few fairly simple pattern recognition techniques that are adequate for the current work, and we have also investigated more advanced techniques such as neural nets and statistical analysis.

The purpose of computer analysis of music can be conceptualized hierarchically. Starting with real world signals (musical audio data) we want to extract *information, knowledge, and understanding* about what is contained in the data. As humans we perform this parsing more or less automatically, but to create an information hierarchy in a computer we need to explicitly perform the data manipulation tasks. A typical scenario is that data is processed by DSP to extract information features, such as the frequencies found in the signal, temporal changes of power and frequency, phase relationships amongst the frequencies and so on. These information features are used to extract knowledge about the piece of music, e.g., what is the rhythm conveyed in a sequence of beats

by a particular percussion instrument, what is the fundamental pitch of the note played by the trumpet, how do the trumpet's overtones combine to produce the timbre or quality of sound (e.g. smooth and mellow, bright, punchy etc). The information features can be combined to generate a framework for knowledge about the music such as where are the main beats, what are the relationships of major and minor beats. Other researchers (Guoyon, Klapuri, Tzanetakis et al.) have used this sort of information to determine the meter, key, and time signature. Finally we can use this knowledge to answer the crucial question: *does this piece of music swing, or is it square?* Duke Ellington and other musical experts have expressed the notion that this is where the *meaning* resides. Meaning and *understanding* are higher level abstractions in the data, information, knowledge framework. Please note that these are not intended as rigid categories, but are merely a model for human conceptualization of music. This fuzzy classification methodology is often used in information science and artificial intelligence to describe knowledge of a system at various levels of abstraction.

While computers have been used for music production since the 1960's, their utility for recognizing patterns in music was not very practical until the 1990's. Some pattern recognition work was done in the 1970's and 1980's, but was limited to research labs. The development of digital music as a common form of distribution has led to great interest in automating the recognition of musical patterns for the practical purposes of searching musical databases using a symbol system appropriate to the salient native elements in music, and marketing of music based on similarity metrics accessible by such techniques. The most prevalent use of computers in music is the machinery which transforms digital data to acoustic form for listeners. In February 2006, Apple Computer celebrated selling one billion songs from its iTunes online music store. We report this as tangential information relevant to the current work, since people listening to music is what drives the expanding computer music market in all its forms.

1.3 Cultural Background: Swing vs. Straight Time

Many musicians, when asked about swing in music, will initially indicate that it is a ***feeling*** and follow this by a more technical detail such as "triplet eighth notes"[2] or "six against four rhythm"[3] or "there are as many kinds of swing as there are drummers"[4] and so forth. The underlying similarity is that swing music produces a different physical and emotional response in many people than do straight time performances. Swing is a desirable feature in music, indicated by the popularity of this style in a wide variety of musical genres during the 20th century. (Gabrielsson, 2000) reports that in his research listeners prefer music which has rhythmic expressiveness such as swing, and that they often react negatively to rhythms with completely uniform timing.

We use a human rather than technical specification for swing: it is a property of musical performance that induces a more or less energetic rhythmic motion in listeners. This can be foot tapping, dancing, bouncing or swaying while seated or standing, or other participatory behavior. Both computer science and psychology researchers commonly use such a metric to define swing. We believe that this effect is an ancient piece of the human condition, and may predate the emergence of hominids. Geese for example, synchronize their wing flapping when flying in formation, as do horses running in an orderly herd (not a stampede). These can easily be analyzed in terms of system optimization. These synchronization effects, sometimes called *entrainment,* are similar to a group of musicians synchronizing to a leader, such as the drum master in Brasilian batucada.[5]

Another good metric for testing if a music sample swings is to make an audio loop of a short section of the piece and play the loop endlessly in a player like Quicktime. If listening to the loop becomes tedious, or begins to sound mechanical or repetitive after

---

[2] Chris Wood, professional musician, founder of "Samba Like it Hot!" bateria. Ashland, OR USA.

[3] Shawn Moore, musician. Ashland, OR.

[4] Statements similar to this are made by almost every musician surveyed about swing

[5] Todd Barton, professional composer and musician. Oregon Shakespeare Festival and SOU. Ashland, OR.

only a few repetitions, then it probably doesn't swing much. We found that we could listen to many of the analyzed loops (which have a good swing feeling) repeatedly and they did not become tiresome. We did not conduct any extensive survey of many listeners as a psychology researcher might do, but we believe this observation about tedium vs interest is well contained in the mainstream of music research (Gabrielsson, 2000).

Having a working definition of swing, we now ask the question, *where does swing come from?* Listening to the timing of a horse canter or human walking gives a good perceptual insight into swing: it is rooted in motion itself. Being found in animal motion other than human leads to the conclusion that swing predates language. A recording of my dog running shows strong resemblance to Brasilian pandeiro rhythm. Sounds generated by the motion of a vehicle such as a streetcar or railroad show how synchronized polyrhythms emerge as a natural result of the bouncing of the vehicle, and the asymmetric nature of the patterns suggests an origin of the swing style. These vehicles can be regarded as information systems, as surely as a database server is. Mathematical modeling of such dynamical systems is explored in chapter 6.

As modern music came to be dominated by sequencers and robotic rhythms like house or rave music, young listeners have not learned about temporal variation in the way that someone like Louis Armstrong may have learned it, riding on the New Orleans streetcars with their rhythmic but imprecise clackety sounds. This is a cultural loss, and our work shows how computers can be used to ameliorate this deficiency. We can provide technical feedback about the temporal patterns of swing. Mathematical models can generate timing variations for rhythmic modification in music production. These can be used by both music teachers and students to facilitate learning about swing.

1.3.1 Notes Inegales

European music has had temporal variations in music performance for many centuries. A style from the 17th and 18th centuries was called *Notes Inegales*, meaning une-

qual notes (i.e. note timings) in French. This belongs to the general category of rhythmic expressiveness in music. It is not clear whether the influence of this style had any direct effect on the development of modern swing such as American Jazz and Blues. Brasil had a strong influx of European music in the early 19th century because the Portuguese Emperor and his Court moved to Brasil when Napoleon invaded Portugal[6]. Certainly one can find strong influences from the European tradition that came down in Brasilian folk and formal music traditions. The main influences in Brasilian music are rooted in the African traditions, but there is also blending between European and African styles.

1.3.2 American, Brasilian and Other Types of Swing

Traditional American Swing is generally quick tempo and energetic, but we use a broader definition to easily include Samba, Reggae and other styles: swing is that quality in rhythmic performance that causes people to move with the music, whether consciously or unconsciously. The motivation is not to try to precisely categorize musical style, but rather to lay a foundation for finding similarities between music from different cultures, so a listener of one type of swing might find and enjoy other types, e.g., when searching a music database using a swing criterion. There is also the important phenomenon of syncretism of different cultures or traditions which synthesizes a new style by combining aspects of two or more extant styles. Samba-Reggae, Soukous and Hi-Life are popular styles that have evolved from combinations of other forms. We expect much new development of this sort of music in the 21st century as global travel and internet access broadens exposure to other cultures. Our work is useful for documenting similarities and differences between various types of swing style.

The ***swing ratio*** is a measure of slight but consistent time differences between pairs of successive "evenly" spaced note events as written in tablature. Use of this term goes back at least to (Cholakis, 1995), and has been investigated by (Anders & Sund-

---

[6] Harvey Weinappel, professional musician and musicologist, Los Angeles, CA USA.

strom, 1999), (Anders & Sundstrom, 2002) and (Birch, 2003). The ratio is obtained for a particular musical sample by statistical analysis of the patterns of "long" to "short" notes. This simple concept is very useful for analyzing American Swing and Jazz.

For example, given a drum score with a series of 1/8th notes on the cymbal, rather than play all notes evenly, a drummer might interpret the score by playing a *short-long-short-long* timing pattern, usually associated with an accent of the same count, e.g., all long notes have accents. A similar modification can be played inside a triplet pattern, as in the Blues. By crowding the downbeat and backbeat (typically the 1 or 3 of a 4 count measure) with a shorter note, or leaning away from the beat rather than into it, musicians can change the perceptual feel of a piece from how it would sound and feel if the rhythm was played evenly. We note that swing is also created by sources other than drums. Louis Armstrong and Benny Goodman express swing very clearly in their horn parts, as does Duke Ellington in his piano. Some research has been done into the swing that is created by the Gestaltic[7] performance of a group of musicians, called ***ensemble swing.*** Ensemble swing has been investigated by (Friberg & Sundstrom, 2002).

We have observed that the *dynamics* of an instrument's notes, especially the surdo in Brasilian music, contribute to swing by the timing of changes in loudness and timbre that would not be classified as note events, but rather are temporal elements inside a single note event. We have looked most closely at Brasilian swing, or *swingee (swing-ghee)*. We consistently find that swingee is substantially more temporally complex than can be expressed with a simple swing ratio, and we document some typical details in chapter 5.

1.3.3 Patterns of Temporal Variation

It is incorrect to regard swing time as less precise than straight time. We will show examples by Ray Charles, Paul Simon and others that demonstrate several aspects of the

---

[7] Gestalt comes from the work of Fritz Perls, who developed psychological theories and therapies intended to teach people to look at situations as a big picture (unified) rather than a lot of separate details. We use it to indicate a unified quality in music which is emergent when a group of musicians are all "in the groove."

precision (or looseness) of swing feeling. In traditional music lessons, students are admonished to play the notes in precise clockwork manner, in time with the metronome. This exactness usually entails counting to four using identical time differences. Swing music has an exact framework of this sort for defining the large scale structure of the rhythm. However the notes between the foundational downbeats will often occur at times other than the canonical quarter note locations. Classic American Swing and Jazz rely strongly on changing quarter note intervals to some form of triplets, and similarly for other factor of 2 subdivisions like eighth notes, sixteenth etc. In Brasilian swingee some notes are played at non MB locations and also on triplet subdivisions. We use the term triplet to describe any temporal subdivision which shows a factor of 3, typically in a 2 or 4 beat meter. This may or may not exactly coincide with standard music notation usage.

In learning Brasilian rhythms as well as in analyzing them technically, we find it useful to use the metaphor of ***rhythmic targets*** to anchor the music precisely in time. We then look at the relationship patterns between subdivision notes and the target(s) in order to accurately play a rhythmic pattern in the proper style, tempo and meter, as well as playing the correct rhythm (data) *per se*. To the best of our knowledge, the concept of rhythmic targets is presented in this paper for the first time.

No matter how well one plays the data, if it ain't got the swing, it don't mean a thing. This is far more than just a word trick. (Hamer, 2000) mentions how the 1959 London production of ***West Side Story*** stage play by Leonard Bernstein was stymied because of the difficulty of finding a drummer who could adequately play the rhythms in the intended jazzy style. This was caused by deficient music reading abilities of jazz drummers, and the inability of classically trained drummers to play the score correctly. Although the classical drummers could read the score perfectly well, they did not know how to play the rhythms in a swing style.

The classic swing ***riff*** might be described by a verbal pattern like

tzzzhhhh, *tch-ta-****tzzzhhhh,*** *tch-ta-****tzzzhhhh,*** *tch-ta-****tzzzhhhh,*** *tch-ta-****tzzzhhhh***, ...

where bold font is the accented downbeat, and the time lapse between between elements of the pattern is not the same for all note events. In fact, this is the hi-hat cymbal rhythm in Duke Ellington's ***It Don't Mean a Thing if it Ain't Got That Swing.*** Most people who have listened to much American music have heard rhythmic patterns like this, and we don't render it in musical tablature because the point here is for you, the reader, to remember what this rhythm sounds and feels like. We intend this as a transmission of non symbolic information. If written in MB notation, the beats would indicate that the meter is in 4/4 time, but the ***feel*** of the rhythm as played includes a triplet timing, especially between the pickup beat (penultimate) and the accented (final, bolded) beat in each repetition of the pattern. Even though the rhythm has a triplet feel, it has no similarity to 3/4 time signature music like waltz, or 12/8 blues with their foundational count of 3. Accurately mapping these timing variations to clear temporal locations in the music is the essence of our research to characterize the swing feel.

Swing also appears in certain kinds of dance. Tap dancing clearly distinguishes between straight and swing rhythm. Most tap dance music is in either 4/4 or 2/4 time. Straight tapping is done on the canonical MB beats corresponding to quarter note type subdivision, including very fast 1/6th and 1/32nd notes. Swung beats in tap are counted with a subdivision of 3 inside the 2 or 4 meter, e.g.*uh* ***one*** *and, uh* ***two*** *and, uh ...* .[8]

1.4 Information Science and DSP Techniques

The main branches of information science that are useful in this work are signal processing (DSP) and pattern recognition. We use DSP techniques to transform audio data into a form where temporal and spectral features are more accessible. Using spectral

---

[8] Jim Giancarlo, professional choreographer and dance teacher. Artistic Director of Oregon Cabaret Theatre, Dance Instructor at Southern Oregon University. Ashland, OR USA

information we extract the note events together with their timing information. Timing information is the basis for recognizing swing, and for classifying rhythmic patterns.

The computer music research literature mentions using many DSP techniques including fast Fourier transform (FFT), short time Fourier transform (STFT, a variation of FFT), wavelets, zero crossings, frequency filtering, sub-band processing, principle and independent components analysis (PCA and ICA), and various statistical methods. We have primarily investigated wavelets, zero crossings and STFT. We find STFT to be the most practical for our work. Filtering and sub-band processing look quite promising and practical, but time limits prevented us from investigating them in depth.

The STFT produces a *spectrogram* that is a visual guide to the moment by moment changes in frequency content of the audio sample. The length of the FFT is crucial for producing waveforms that are easily parsed for note events. The FFT acts as a kind of smoothing filter for the complex and rapid changes in the audio input stream. Longer FFTs smooth more, and short ones smooth less. We find that short FFTs (less than 1024 samples) are generally not useful because the waveforms generated from these sequences are not smooth enough to reliably recognize note events. This is similar to the problem of looking directly at the waveform of the raw audio signal and trying to recognize patterns: there is too much activity in the waveform. Detecting note events is less a problem for clearly separated individual events, but for most music, several instruments are contributing to the audio signal at any particular time. Separating these mixed sources requires specialized techniques which we do not currently use, as well as high resolution of the time and frequency data. We use fairly high resolution in both time (1 to 10 milliseconds) and frequency (10 to 50 Hz).

1.4.1 Fast Fourier Transform (FFT)

Fourier analysis is a mathematical technique that transforms raw audio data in the time domain to a set of frequencies in the audio spectrum, or frequency domain. The fre-

quency domain form of the information is very practical for our work. In DSP, the Cooley-Tukey FFT algorithm is a commonly used algorithm that efficiently computes the spectrum of the input data. The FFT approximates the theoretical resolution of a continuous Fourier transform. The FFT is several orders of magnitude faster than the continuous transform, and also much faster than other discrete Fourier transforms. The primary trick of the Cooley-Tukey algorithm is to take advantage of certain symmetries in the Fourier transform. The data in the time domain is multiplied by complex exponential functions (essentially, sines and cosines of different frequencies) as part of the Fourier transform. The complex exponential functions can be composed by using other complex exponentials of different frequencies, much the same way that the number 1/4 can be factored as 1/2 x 1/2. By organizing these factorizations properly and re-using some of the exponentials many times rather than recomputing them each time they are needed, the compute cost of the FFT algorithm is greatly reduced. (Brigham, 1974 ; Elliot & Rao, 1982)

1.4.2 Pattern Recognition

We use the spectra extracted by DSP to identify different musical instruments by their tonal (frequency) content. The general strategy is to extract short, simple features from large, complex data sets. These time/frequency features are good for identifying a note event, such as an increase of power level in a frequency range during a time interval. This is quite useful for identification of percussion and drums. To recognize these patterns, we mostly use thresholding techniques, based on the spectral power density curves, which are plots of power vs time in a well chosen frequency range. We also use the first and second derivatives of these waveforms. These techniques are very practical but have limitations, especially for complex music samples such as several instruments playing at once, or melodic instruments with complex spectra. For these more challenging musical samples, we plan to use neural net approaches in the next stage of this work, since they are computationally efficient pattern recognizers with great adaptability.

1.5 Structure of this Thesis Document

Our work in computer music analysis has been somewhat broad ranging rather than tightly focussed on a specific narrow topic. (Plomp, 2002) has recommended that researchers not become mired in technical details to such extent that they risk seeing only trees and not the forest. The details are important of course, but so is the big picture. We present information about both small scale and large scale views of the complex topic of human perception of music, and the associated computer analysis of the musical data.

For readers who have limited time to spend or who are not keenly interested in excessive technical details about Information Science applied to computer music analysis, we recommend reading section 3.1 first which is a practical introduction to our technical approach, and then skipping directly to chapter 5 which presents the main body of our results. After this exposure, the reader may be interested in looking more closely at the technical details of signal processing and pattern recognition. The appendices on Brasilian music and the psychophysics of human hearing may also be of general interest.

In chapter 2 we survey some of the research in the field of computer analysis and recognition of music, especially swing research. We note that one of the principle differences between our work and all other research we have read is that we focus exclusively on music as a set of distinct events, whereas most or all of other research takes a statistical approach to music analysis. We believe strongly that analyzing musical events individually rather than *in toto* is a very important paradigm because this is the primary way that humans produce and consume music. Statistical and gestaltic analysis also has useful application in understanding music, but we do not work much with this paradigm. We also note the connection between swing rhythm and bodily motion which has been investigated by many researchers including (Gabrielsson, 1987) and (Waadeland, 2004).

Chapter 3 describes the variety of DSP techniques we have investigated, and in particular includes detailed descriptions of our FFT work.

In chapter 4 we describe our pattern recognition techniques which are useful but not particularly sophisticated. We also survey some pattern recognition techniques used by other researchers in the computer music analysis field.

Chapter 5 presents the main body of our original work, which is detailed analysis of several specific examples of different genres of music. We present results that use a much finer grained model of time than do most researchers in this field. We have found strong evidence that temporal granularity should be no more than 5 to 10 milliseconds for adequate understanding of critical details of rhythmic timing. Most other researchers use 10 to 20 milliseconds as the lower limit of their temporal subdivision. In particular, we present evidence that a highly experienced musician such as Ray Charles has temporal perception which has less than 5 millisecond granularity. We also present evidence that ensemble swing depends at least in part on interactions between musicians with the ability to perceive and manipulate time differences in this range of 5 to 15 milliseconds. We also present examples of alternative musical notation which gives a quantitative guide to playing swing rhythms authentically, and a technique for automated generation of swing timing variations.

In chapter 6 we present ideas for closely related future work, including some of the deficiencies we have found in Fourier analysis.

In the appendices we present some peripheral material which is germane to our broad view that parsing musical information should not be restricted to a purely computer data processing model. This includes observations about our own experience learning and playing music, information from professional musicians, the code for our algorithm, a discography of the music we have investigated, some information about Brasilian culture focusing on music and dance, and a brief description of some of the standard knowledge of the workings of the human auditory perceptual system.

In appendix E, we focus on the front-end parts of the hearing system such as the ear and cochlea because these are directly analogous to DSP extraction of information from digital audio data. We note that there are several parallel mechanisms for transforming sound vibrations into the neural patterns which enter the human brain and that eventually become our conscious perception of sound events. We make special note of the fundamental *nonlinear* qualities of the human audio data acquisition system, in contrast to analysis using computers which is predominantly based in linear mathematics. We also consider human factors related to music perception. Psychology research has produced a large body of knowledge about intrinsic properties of the human mind and its natural inclination towards producing rhythmic patterns. The human feelings and knowledge triggered by music are also very interesting but we do not pursue them deeply, deferring to the vast literature on neuroscience which is beyond the scope of this thesis. The connection between music and human emotion has been noted and investigated in both psychological and musicological literature. We find the emotional aspect quite interesting, and include some opinions based on our experiences performing and listening to music, but not as part of the main thrust of the current work.