A New Strategy for Control of Adaptive Digital Audio Effects: Linking Scores, Features and Variable Control

Donald S. Bosley

June 7th, 2013

Advisor: Dr. Kenneth Peacock

Submitted in partial fulfillment of the requirements for the
Master of Music in Music Technology
in the Department of Music and Performing Arts Professions
in The Steinhardt School
New York University

Contents

List of Figures

List of Tables

Acknowledgements.

I could thank a lot of people that helped me get to this point... The librarians who showed me how to use the card catalog as a kid; my mom for making sure I knew the library was there to begin with; my amazing family who generally admits to not understanding me but supports me nonetheless; the coaches, teachers, and professors who saw something in me despite my sometimes less than stellar behavior and academic achievements; and the drill instructors and Marines who showed me that the impossible, never is. Oh yeah, and Mr. Elliot who was totally right - the stuff usually does work best when it's plugged in and turned on.

I mostly want to thank my dad for all the hours in the garage, where I learned a lesson whose value far outweighs its simplicity, and that I always go back to, even in this paper - always use the right tool for the job. Unfortunately, I'm still not always so good at putting the tools back where they belong when I'm finished with them, but I'm still trying.

Abstract

An explicit mapping strategy is introduced which aims to correlate variations in performance which result from interpretations of symbolic musical scores, to the control of variables within adaptive effects(A-DAFx). The physics and electronics of electric guitars and violins are discussed along with markings from traditional notation. Recordings are made which include interpretations of these markings, and the resulting signals are analyzed in order to display the variations on the instruments' electrical outputs. Key features in variation are identified and then applied in a control scheme for various A-DAFx which utilize stringed instruments with polyphonic pickups as both the audio source and control stream. This strategy of linking notated markings with control features gives traditional performers an easily understood, predictable, repeatable, and thus musically useful method for controlling A-DAFx; a few examples of the final results are displayed.

Author's Note

While there are many strategies available for the mapping of A-DAFx, it seems to me that most have been developed and are presented with more emphasis on the technology itself rather than the musical materials these systems are designed to empower. When I first began attending electronic music events in the classical world, many performances left me wanting something more. Things like adaptive effects and interactive systems satisfied some of this desire, but in many instances there still seemed to be a gap between the electronic and more traditionally based acoustic musicians.

With a strong interest in the vast possibilities that music technology offers, I began to explore methods of how to fill the perceived gap. Traversing paper after paper, and book after book, I found that the relationship of the technology to music itself was generally left out of the studies. I still find this slightly disconcerting and it became my motivation in this research. With the ever growing specialization within technical fields, and the demanding nature of being a modern art musician, the odds of success at being a 'renaissance' person have been driven down enormously. In my experience, I have met many individuals who are very well versed in either side of the discussion, but generally have difficulty bridging the two.

For this reason a great deal of this paper covers a variety of richly interwoven topics. This document is fully self-contained, providing the basic elements necessary to understand each idea that is outlined within. In cases where there is a substantial body of related or more advanced literature, additional resources are suggested. The style of writing is constructed on the precept that a musician with limited technical knowledge can grasp the mechanics of signal processing. It simultaneously attempts to tailor to the engineer with limited traditional musical training.

The aim here is to serve the musical arts through technology and it is my hope that this document can facilitate a better understanding of the marriage of the two subjects for both parties, while offering a perspective on the use of the type of electronic music tools discussed. Musicians with more insight into technical matters, can express their creative vision in a clearer fashion to engineers and programmers based on informed thought. Developers who better grasp the nature of musical problems will have better insight into the design of tools which provide the desired musical functions. In essence, this document should serve to open a clearer dialogue in order to further the use of technology in musical endeavors.

- dB, April 2013

1 Introduction

Adaptive Digital Audio Effects (A-DAFx) represent an exciting class of methods for transforming digital audio. While analog effects units, such as guitar pedals, are still extremely popular, digital technology has offered the possibility of implementing existing effects cast from new molds, as well as new effects without precedent. A-DAFx are a specific category of audio transformation, which exploit the richness and flexibility of manipulation within digital audio. They do so by dynamically altering aspects of these transformations, often in real-time and using the audio content itself as the source of control for variations. A-DAFx are considered "content based transformations."

Imagine a device that constantly measures your distance from a television and adjusts the volume accordingly. In this instance we say that the volume (a variable), is controlled by distance (the feature), because it is assigned (mapped) to do so. **Feature extraction** allows a digital signal to be broken into usable lower level information, such as variations in loudness or estimated pitch. The process of assigning these features to the control of variables in a complex fashion is known as **mapping**.

Feature extraction should be executed so that it can provide what Kitahara[1] refers to as 'invertibility', or regenerating the original high-level content from its low-level representations. In some previous works regarding A-DAFx, features are assigned to control variables in their low-level state, mapping controls on a sonic or perceptual, rather than musical basis [2, 3, 4, 5, 6]. The major aim of this research is to develop a strategy which draws stronger connections between mappings of low-level features and the high-level content found in the activities associated with composing and performing instrumental music in conjunction with A-DAFx. In order to accomplish this aim, the strategy being explored here draws direct connections between a chain of events: a symbolic representation of music provided by a composer (notation), execution by a performer (interpretation), features extracted from digitized musical signals, and control or mapping(A-DAFx). By identifying the relationship between these events, conclusions can be drawn about effective mapping, driven more by musical content rather than just perceptual or sonic features.

The guiding principles in decision making are 1)that any system subsequently developed, would provide a performer with an easily understood, predictable, repeatable, and therefore musically useful method for controlling A-DAFx, 2)provides a composer with a strategy for directly connecting musical materials with A-DAFx, partially rooted in classical orchestration techniques, and 3)provides an audience with a meaningful link between performative acts and the resulting sound.

1.1 Motivation and Meaning

> *"Many are troubled by the apparent disparity between performative acts and the sounds resulting from them. Others give it no thought, having become accustomed to valuing only the sound, because that is all most recorded music delivers. Perhaps a new performance practice is developing with different values, or perhaps the troubling qualities of this disparity can help us reflect on our understanding of the value of live performance."*
>
> \- Jeffrey M. Morris [7]

Throughout the history of electronic music, audiences, composers and performers alike have questioned the validity of the performance practices which have developed within the genre. For many, there is a general sense that the bounds represented by the musical developments of the centuries prior have in some way been violated, especially where performances over loudspeakers or based on laptops aren't particularly interesting to watch [7, 8, 9]. Often times electronic components are generated in a cryptic manner, leaving an audience wondering what the performer is actually doing. For example, when a general purpose tool such as the laptop is the method of both sound generation and control, the performative act can appear no different than checking an email or using a word processor.

Some may object to the use of traditional instruments in new idioms, but there are advantages to using an acoustic instrument as the source for both the audio material and control information within an electronic music system [10, 11, 12, 13]. First, and most important, is that experienced musicians have developed an expertise in dealing with musical materials. It is important that musical systems, both acoustic and electronic, incorporate a deeper knowledge of structure and music-making strategies. Mastery is attained and the skills of the performer honed, by overcoming the associated challenges [14, 15]. Human performers add a visual component to live performance, and can help to create a bridge between the audience and electronics by providing cues through physical gesture which link to musical results.

Many studies have been conducted which analyze the physical aspects of performance, or 'gesture analysis', in order to create schemes which correlate physical action on the part of the performer to musical events [16, 17, 18, 19]. While an extremely exciting field, gesture analysis has a number of drawbacks. It generally requires expensive or customized sensory hardware. In many instances an additional layer of processing is required between the sensors and computer based audio effects, such as a micro-controller. Hardware must be carefully designed so as not to encumber the player's natural motion, or force a modification of technique. Many technologies, such as flex-sensors, capacitive-touch interfaces, and cameras can create widely-varied or unexpected output during live performance. This stems from changes in environmental conditions such as changes in ambient light, like the adjustment of stage lighting. When a sensor makes contact with a human body whose physiology changes during performance it is subject to moisture (sweat) and variations in temperature (body heat).

Score following, or linking performed musical events to a pre-existing score, can also be used to trigger extra-musical events at the desired locations in time. Score-following also has its merits in particular applications, but has major drawbacks for certain types of performers and performances. Score following requires an existing score, first in one of many digital forms, and second in a form readable by the player. There are many forms of improvisation which rely on no score, an incomplete score, or a score utilizing non-standard graphic notation. There are also many musicians with great technical prowess and a mastery of expressive materials, who are incapable of reading traditional notation.

A relatively popular solution to accessing rich sounds using existing acoustic instruments, has been the use of analysis-synthesis, mainly in the form of pitch-to-**MIDI (Musical Instruments Digital Interface)** converters. The latency, general quantization/reduction of expressive parameters, and the sometimes problematic pitch detection in these types of systems has limited their consistent widespread acceptance and use. Synthesizer controllers which use electric switching rather than analysis such as Electronic Wind Instruments (EWI's), or the SynthAxe, are attempts to offer performers access to keyboard-like control through familiar interfaces. In many cases, players complain that using either type of system forces them to change their technique, and in essence feel as though they are playing an unfamiliar instrument.

The major remaining option for creating varied and rich sounds lies in processing live instruments through audio effects. Acoustic instruments possess a natural set of complexities, and these complexities can be augmented, or reduced, using **digital signal processing (DSP)** techniques. Building upon techniques from the **music information retrieval (MIR)** and computer music communities, A-DAFx whose variations are derived directly from musical content can be designed. For all of the aforementioned reasons, an acoustic instrument will be the source of audio, and DSP techniques will be implemented to extract the features used for controlling variables in A-DAFx.

One assumption is that A-DAFx are 'intelligent' in cases that the source of control comes from a feature extracted from a performance [2]. What is rarely discussed is the definition of what intelligent actually implies. Is it an added level of complexity, a testament to a decision making process, or something that aids in the conveyance of musical and emotional content?

According to Kurzweil [20], "*Human intelligence is in many ways, derived from the nature of the human body itself, and the desire to solve problems and goals associated with it.*" Intelligence is arguably built into the performance, as the performer is using their body to solve the problem of musical communication. However, it is still necessary to develop and implement strategies which retain this intelligence during the conceptualization and construction of the electronic component, along with the link connecting the two. It is clear that A-DAFx require a map, and by definition a map generally requires a legend to make it useful for navigation. The goal of this thesis is to create a coherent strategy for creating such a legend, narrowing the application to a specific set of instrumental idioms, discussed below.

1.2 Choices of Instruments

Stringed instruments, specifically the electric guitar and electric violin, were chosen very carefully for two reasons in this study. The first reason deals directly with the popularity, including the amount of existing musical literature, pedagogical or technique texts, and extensive contexts in which these instruments are used. This wide user base might extend the impact of the research, where a study dealing generically with the issues of control in A-DAFx may not. Although many companies do not publish sales figures publicly, the electric guitar is likely the most commercially successful of all musical instruments in the Western World [21]. The violin possesses one of the richest and oldest catalogs of literature in not only Western Art Music (WAM), but it is also central in other genres such as world, folk, pop, bluegrass and jazz [22].

The second reason is what the electric versions of these instruments offer. They output little acoustic energy, allowing for near complete sound-source separation in performance, and isolation of the input signal for analysis and control. Microphone type and positioning can vary from performance to performance, however pickups never change position in relation to the strings. The removal of room reverberation, bleed from other instruments, and division of strings into individual signals via divided pickup technology, minimizes the variables in sound capture. Less variables in sound capture lead to less variables in analysis, allowing more consistent manipulation and control of the resulting digital signals.

1.3 Outline

This document is divided as follows : Section 2 is a primer on stringed instruments, focusing on the electric guitar and electric violin, which includes the physics and general methods of operation of their internal electronics. In Section 3 feature extraction will be introduced, through the explanation of the signal processing algorithms employed along with a basic explanation of their historical basis where relevant. Section 4 will outline traditional expressive markings associated with both instruments and provide clues as to how feature extraction might be applied to identify musical content linked with these markings. Section 5 will discuss the conversion of resulting features into control streams, and the overall mapping strategy, giving examples of a few A-DAFx which were developed off-line in MATLAB.[1] Section 6 offers final thoughts, draws conclusions and offers suggestions for future improvements and use.

Since a portion of the decision making for using electric guitar and violin lies in their popularity, it is assumed that out of a body of interested readers some may be unaware of certain technical (or musical) concepts. For this reason, a number of brief appendices are provided as a supplement. The author suggests scanning the list of appendices in the table of contents, and reviewing any concept that the reader is not familiar with.

2 Stringed Instruments

2.1 Short Overview of History

The electric guitar and violin are the result of a few centuries of evolution stemming from their acoustic counterparts. Early versions of the violin began to appear in the 1500's, having the design codified by Stradivari, and eventually retro-fittings by French makers in the eighteenth century [23]. The standard six string configuration of the guitar most widely used today, began to appear around 1770 with the earliest documented instrument being built in 1779 by Gaetano Vinaccia in Italy [24]. Numerous builders began to construct instruments in this form and by the year 1810, the instrument was in widespread use [25]. Like many technologies and tools, it was during the age of electrification that these instruments began to receive a major re-examination.

The first stringed instrument to use electromagnetic principles was created around 1890 by George Breed, a United States naval officer but was limited in its construction and did not permit strumming or picking [26]. In the decades following Lee Deforest's 1906 patent for the 'Audion', a breakthrough in vacuum tube technology, the size and nature of ensembles and performance spaces

[1] Although the applications are being developed offline, they employ the block methods found in low level languages like C/C++. The author has already began working on the conversion of these effects into real-time Virtual Studio Technology (VST) Plug-ins.

began to change [27].[2] Recording and wireless technology began to mature allowing music, in a certain sense, to transcend time and space. In many contexts the acoustic guitar and violin suffered from fundamental physical limitations in volume.

Early recording methods required that enough air pressure be supplied to drive a carving needle into the storage medium, and the guitar and violin were at a disadvantage [21]. To solve this problem, inventors began to experiment with methods of capturing and amplifying signals utilizing various devices from phonograph needles to telephone receivers. The first electric guitar appears to have been demonstrated in 1923 by Lloyd Loar of the Gibson Mandolin-Guitar Company, having the signal generated by a carbon-granule microphone sensing vibrations at the bridge. Subsequently in 1928, Victor Pfeil applied for the first patent on a solid body electric violin, using a similar set-up as Loar [21].

The greatest leap in pickup technology came from George Beauchamp in the mid-1930's, who developed the first electromagnetic pickup. The electric guitar reached its most recognizable state under three developers : Les Paul(born Lester Polfus), Leo Fender, and Paul Bigsby, who established various degrees of solid body instruments, preventing electromagnetic feedback between pickup systems and amplification units. Improvements and balances in design are still being sought though, from the use of "space-age" materials to rethinking of pickup technology. The electric violin also continues to evolve with various types of body shapes, pickup systems, and even variations in the number of strings.

2.2 The Physics of Strings

A brief review of the physics of these instruments is essential for two reasons. The first is their established physical properties provide context in understanding some of the physicality involved with performance idioms, therefore allowing insights into limitations and usage. The second reason is that since signals resulting from the use of these instruments will be examined, physics can provide models which help predict the outcome of variations within extracted features, and drive choices in signal processing algorithms that may yield the most useful results.

The central operating point of stringed instruments and the origin of all other function begins with the vibration of a string itself. A number of alternate methods of excitation exist, but the two most common cases of how strings are excited into motion are first by plucking with a finger, fingernail, or plectrum/pick; the second is with a bow. In both cases, there are generic underlying principles.

A vibrating string is stretched across two points, in the case of the guitar between the bridge and the nut, or alternately between the bridge and a fret when the string is depressed. The violin string is stretched across a bridge and nut, or what is sometimes referred to as a top nut. Where guitars generally have frets which subdivide the string length into fixed intervals, the violin can achieve an infinite number of subdivisions, or micro-tones, across the fingerboard.

A common analogy provided for the motion of a vibrating string utilizes a rope affixed at one end, and snapped or given an impulse at the other[28, 29]. The rope may maintain its shape fairly well as the waveform travels its length. When it hits the tied end, it will reflect back towards

[2] Although it was a *"breakthrough"*, it was based on the work of Nikola Tesla.

the initial snapping point but with the curve in the inverse direction. The fundamental frequency at which the theoretical rope (or a string) cycles, f_0, is defined by a number of physical factors. Eq.1 displays the fundamental frequency of the string stretched across two points, where $n =$ mode number, $L =$ length, $\mu =$ mass per unit length and T is equal to tension per unit length (i.e. pounds per square inch)[30]:

$$f_n = \left| \frac{n}{2L} \right| \sqrt{\frac{T}{\mu}} \tag{1}$$

Unlike the snapped rope where motion was initiated from one end, plucking or bowing generally occur at a point along the length of the string. When the plectrum or bow is used to apply force and vibrate the string, energy travels in both directions away from the point of displacement. The energy then reflects from the endpoints of the string back towards the point of displacement. Depending upon the direction and timing of travel, these waves will cause constructive and destructive interference at various points along the string's length.

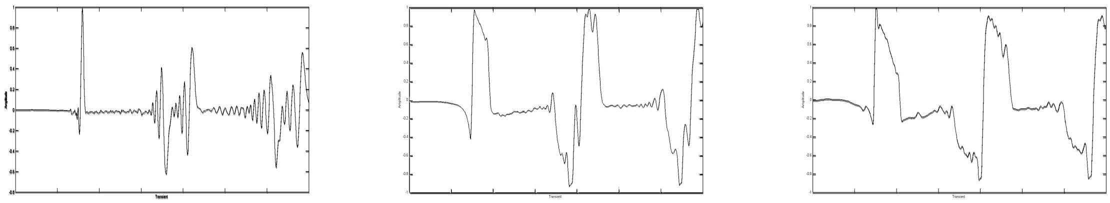
Standing waves occur when this reflected energy meets again at a central point. Points of destructive interference where displacement is at a minimum are called nodes. The points of maximum movement as the result of constructive interference are called anti-nodes. Anti-node displacement oscillates at the same frequency as the individual waves [28]. Anti-nodes combine to form the overall shape of the string's movement. Anti-nodes and nodes do not change position on the vibrating string and as a result they are called standing waves [30, 29, 28]. The normal modes of each string can be summed in order to give the overall pattern of vibration at a particular frequency [28].

The reason the vibrating string produces very little audible sound, is because it displaces a minimal amount of air in the process of vibration [30]. Acoustic stringed instruments contain resonators - bodies with hollow cavities of air, often referred to as sound boxes. The energy of the vibration is transmitted through the bridge into the body and causes the cavity to resonate. The electric guitar and violin do not require a resonating body as the pickups take their place, converting the strings' pattern of motion into electrical energy. The strings of an electric instrument tend to have more vibrating time or sustain because less energy is typically lost through the bridge into the body.

Guitar			Violin		
Note	String	f_0(Hz)	Note	String	f_0(Hz)
E4	1	329.63			
B3	2	264.94	E5	I	659.26
G3	3	196.00	A4	II	440.00
D3	4	146.83	D4	III	293.67
A2	5	110.00	G3	IV	196.00
E2	6	82.407			

Table 1: Fundamental Frequencies of Open Strings

The standard f_0's of the electric guitar and violin's strings, displayed in Table 1, are important to consider for analysis purposes. The human ear can typically detect frequencies between 20 and 20,000 Hz. As points of reference: The flickering of incandescent lights is around 60 Hz, the dial



(a) Near the Bridge **(b)** Middle **(c)** Near the Neck

Figure 1: Time Domain Analysis of Picking Position

tone heard on American Telephones is around 440 Hz, and a Dog Whistle is generally well over 23,000 Hz (23 kHz). All feature extraction, analysis, and effects functions are tuned based on the ranges of the instruments in question.

Table 1 uses pitch names along with a numeric designation to indicate what octave the pitch belongs to. Violin strings are identified by Roman Numerals, while guitar notation uses Arabic Numerals to designate strings because Roman Numerals are utilized in scores to designate the hands position along the fingerboard. Despite a wide variation in f_0, the guitar's strings generally have a uniform tension, giving them a somewhat homogeneous character (with variations to be discussed). The violin's strings on the other hand, have significant differences in tension with the high e requiring the most force per mass unit. From a sound quality perspective, the strings of the violin each have distinct character, often exploited for the elicitation of specific musical results.

2.3 Methods of Excitation

2.3.1 Plucking

The Karplus-Strong model, a classic signal processing algorithm for synthesizing plucked-string sounds, states that the action of exciting a string with a pluck or strike produces an initially irregular or noisy attack [31]. This portion of the signal is called the transient, which contains high frequency content and does not always have an identifiable cyclic repetition. As the energy dissipates the string begins to settle into a regular pattern of vibration called the steady state. Eventually, and often times rather quickly, the motion stabilizes such that a listener perceives a fixed pitch.

The position of where the plectrum excites the string is a key component in shaping what modes along the string are excited and their resulting amplitudes. Figure 1 shows how greatly the waveform can vary, based upon picking position. Equation 2 displays which of these modes are not excited where m is a series of positive whole numbers (1,2,3...), L is the length of the string, and d is the distance of the plucking point from the closest end of the string [30].

$$Modes\,not\,excited = m\left|\frac{L}{d}\right| \tag{2}$$

The modes not excited will be integer multiples m, of the ratio of L to d. Guitar strings are generally excited in this manner, and the technique known as *pizzicato* in the bowed string family indicates a pluck versus a bowing, designated in scores as *arco*.

2.3.2 Bowing

Bowing and picking have a few fundamental differences. A plucked note is excited once and energy instantly begins to dissipate. This results in a constant loss of amplitude, and the disappearance of upper harmonic content. A bow can continuously shape the amplitude of a string's vibration over the course of its evolution. There appear to be a larger number of techniques achievable with a bow than with traditional methods of string excitation for the guitar. The trade-off being that a guitarist's plucking/picking hand is free of a bow, allowing them more direct access to controls and varied combinations of plucking and fingering techniques during performance.

The violin bow is sectioned into areas known as the point, middle and frog. The frog of the bow is where the bow is held, and the adjustment of tension on the bow's hairs are made. According to Flesch [32], more divisions exist which are: the extreme point, "small" middle, and extreme frog. The four most important parameters in execution and control of bowing are bow speed, force (termed bow pressure by musicians), position (relative to the frog or tip), and distance from the bridge [32, 33, 34, 35].

All of these factors contribute to control, timbre, and dynamic range. It takes a minimum amount of speed and force in order to set a string in motion and keep it consistently vibrating. The position of the bow along the string's length greatly impacts the variety of force, and therefore range of techniques and pressure which can be applied. The bow's position relative to the bridge plays a role in shaping the timbre. The portion of the bow used, along with the tension of its hair, also determine the techniques available for use.

At first glance, the vibration of the bowed string at f_0 appears to be a circular, or even semi-sinusoidal motion, however even with crude equipment Heinrich Helmholtz proved in the mid-19th century that the motion of bowing forms two nearly straight lines with a sharp bend at the point of intersection [30, 33, 36]. This idealized pattern, similar to Figure 33b (App.C), is aptly named the "Helmholz Motion", and bowing at the midpoint between the bridge and fingering position is most likely to produce this idealized waveform [33].

As the bow is drawn in a perpendicular fashion across the string, a pattern of sticking and slipping occurs. The horse hair of the bow grips the string and pulls on it until slippage occurs and causes the string to snap back towards its original position. The string vibrates freely until the bow re-grips it, and the cycle is repeated for the duration of the bow being drawn across the string.

2.4 Internal Electronics

By plugging a standard 1/4" cable into an instrument's input, it can be connected to amplifiers, analog-to-digital converters, or signal processing units. Electrified instruments typically contain a number of internal components used to capture and subsequently shape this output. These systems fall into one of two categories, either active or passive. Active systems require a power source while passive systems can run using only the current passed through the cable into the instrument's jack. There are advantages and disadvantages present in each. Passive systems can only be subtractive in nature, while active can boost or add to signals. Due to more common use and simplicity, passive electronics will be discussed here.

2.4.1 Pickups

Pickups are the heart of capturing a string's motion, and generally come in three forms : electromagnetic, which are the most common type seen on the electric guitar; piezo which are generally used on acoustic guitars, and in the bridges of electric guitars and violins; and a relatively new breed of pickup which operates on light waves called the optical pickup. A complete overview of pickup technology, specifically with guitars as the focus can be found in [37]. The type of pickups, the materials used in their construction, how far they are set from the strings, and other factors contribute to form the overall sound an instrument produces. Magnetic pickups can also influence small variations in tuning/intonation based on their distance from the string.

Electromagnetic pickups consist of a wound coil surrounding a magnetic core. The vibration of the metal string, generally made of a blend of nickel and/or steel, causes a change in the magnetic flux through the core, resulting in the induction of an electrical signal into the coil [24]. In order to amplify the output of passive pickups, a device requires an input that uses a significantly high impedance (Hi-Z), typically at least 1MΩ or greater.

Pickups have a resonant frequency defined in Equation 3; the resonant frequency is important because it is a sharp peak in the amplitude response, with frequency response dropping immediately after. This filter topology is called a resonant low-pass filter and stems from the electrical properties of a pickups' construction. A pickup consists of coils causing a large inductance, L, measured in Henries, and a large capacitance C measured in Farads, created by the massive number of windings within the coil [21]. This plays a major role in defining the dynamic range of certain registers of an instrument.

$$f_{res} = \left(\frac{1}{2\pi}\right)\sqrt{\frac{1}{LC}} \tag{3}$$

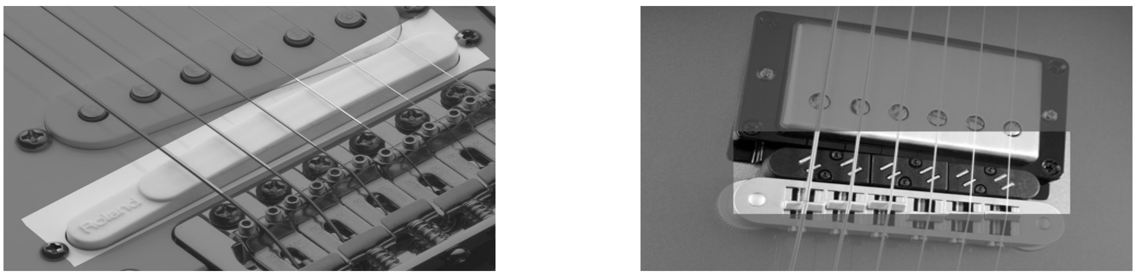
Electromagnetic pickups come in two main varieties, single coil and humbucking. The single coil pickup (Fig.2a, www.diystrat.blogspot.com) is the basis of many designs and is called such because there is only one set of windings around a single magnet, or set of magnets. A very specific variant of the single coil pickup is the Lace Sensor (Fig.2b, www.lacemusic.com) which contains thirty-four smaller, concentrated magnetic fields. With less concentrated pull exerted on each string, the sustain is even longer than with standard single-coil pickups. The third benefit is that the radiant field barriers included in the pickup work to reduce 60-cycle hum and other noise [38].

The classic humbucking pickup (Fig.2c, www.Gibson.com) was designed by Seth Lover for the Gibson Company in the late 1950's [38]. The humbucker consists of two single coils contained within one unit, generally housed under one cover. The two coils are wired together in-series, and magnetically out-of-phase, with pole pieces or magnets for each coil charged magnetically opposite. The north facing poles of one coil and the south facing poles of the other, give this type of pickup the ability to eliminate hum through phase cancellation, known as common-mode rejection [39].

Contact microphones which sense physical vibration of the string or a resonating portion of the instrument have been in existence for decades, however their lack of quality, and uncontrollable nature led to the development of the Piezo pickup. In the late 1960's companies such as Baldwin

(a) Single Coil **(b)** Lace Sensor **(c)** Humbucker

Figure 2: Electromagnetic Guitar Pickups



(a) Roland-Ready Fender Stratocaster **(b)** Gibson HD.6

Figure 3: Divided Pickups (Highlighted Portions of the Images)

and Barcus-Berry were working on designs, and by the 1980's companies such as L.R. Baggs and Larry Fishman began to produce very effective under the saddle pickups of this variety [38]. The under the saddle, or in the bridge configuration, is the most popular use of piezos in both models of electric and acoustic guitars and violins.

Piezoelectric materials are insulators with asymetric structures. The application of a force, or strain on one side of the material introduces a positive charge, and force on the opposite side produces a negative charge [21]. These changes in pressure on the surface of the material are sensed by electrodes and the degree of deformation controls the amount of charge induced. $\Delta V = \Delta Q/C$ expresses this relationship where ΔQ is the difference in charge, and C is the capacitance between the electrodes. This is typically in the range of between 100-1,000 pico Farads (pF).

The core of a number of items which will be discussed throughout the remainder of this document are centered around the concept of divided pickup systems, which are commonly referred to as **polyphonic pickups**. A divided pickup contains a discrete output for every magnetic pole piece or piezo element. In the case of many mass-produced polyphonic pickups such as the Roland GK series or Axon PU-100 which use six small humbuckers, there is one signal for each pair of pole pieces. Common divided pickups for guitar are referred to as "hexaphonic" or "hex" for short, because they output six individual audio streams. The divided pickup concept was also used by computer musician Max Mathews on a number of violin instruments he constructed, and has been the focus of a great deal of work by Miller Puckette in recent years [40, 41].

2.4.2 Controls

Beyond selection of wood types, pickups, and other material choices made during construction, the variation of an electric instrument's tone, also stems from its internal controls. There are a number of knobs and switches that generally appear on the surface of an electric instrument. These are controls which can change the values of any number of variables within the electronics. The typical control structure accesses the volume and tone.

The volume control is most often a variable resistor, known as a potentiometer (or pot), that has a logarithmic taper. This control balances the amount of signal that flows through the hot output of the jack, against the amount that is sent to ground. The volume control has no effect on the physical dynamics of the instrument, however it gives the user control over the amount of electrical signal which travels to connected equipment, directly effecting the maximum amplitude of the signal, and is commonly referred to as a gain control.

Tone Controls are tied to various processes of filtering, and the adjustment of a tone control sets the cutoff frequency f_{co}. Passive tone controls are generally simple low-pass or high-pass filters. Switching simply means changing from one state to another. In this context, switching functions can perform any number of operations. Some common practices have been developed including turning pickups on and off, selecting different combinations of pickups, and switching the phase of a pickup by 180°.

Pickups are distributed along the string length, and various combinations will sense the vibration of harmonic nodes along the string's length differently, thus altering the frequency content of the signal. Pickup selection is in many ways analogous to the position of excitation. For example, a pickup near the bridge will not capture as wide of a vibration as a pickup near the neck due to the position of nodes and anti-nodes along the string's length. The result is that the bridge pickup subjectively sounds thinner or tinnier, while the neck pickup sounds more mellow or dark.

3 Feature Extraction

Feature extraction is the method of distilling digital audio information for use in higher level musical representations that indicate ideas about the underlying musical content [42]. It is relatively easy for a trained musician or an avid listener to perceive the evolution of musical content. Computerized methods of analysis have improved vastly in recent decades, but concepts like pattern recognition and detection of specific types of events are resource intensive tasks for machines. Being a highly parallel device, the human brain has an advantage in segregating and simultaneously identifying disparate elements of musical events for individual processing and cognition [43]. Feature extraction attempts to mimic this segregation and common methods are discussed here.

3.1 Methods of Interpreting Feature Measurements

All features are initially measured as discrete values of varying types, relying on the sample as the building block. While absolute measures are one aspect of feature extraction, features can be defined by relationships through comparisons to fixed values, groupings of values, and also by variations over time. The two main categories used in this document and their sub-variants are discussed prior to introducing the actual feature extraction processes used. This should provide the reader with insight into possible usage.

3.1.1 Comparators and Variations of

Comparators are the simplest method of dealing with what their name implies - comparisons. In the analog domain they measure an input signal's voltage (V_{in}) and compare it to an established reference value (V_{ref}), most often designated as the threshold. While the comparator can operate on any type of electrical signal, it is only capable of outputting an on/off or high/low (binary) state based on the comparison. The logic for this operation would be "***if** Input > threshold value, **then** A, **else** "Input < threshold value, **do** B.*"

One drawback to a basic comparator is the possibility that while a signal is approaching the threshold (from either direction), its value can change rapidly around the threshold itself causing erratic behavior. The result is a rapid and often undesirable alternation between states. One way to deal with this is "switch debouncing", a common practice in analog and microcontroller applications [44, 45, 46]. A blackout period is established so that after the signal crosses the threshold, it cannot return to the alternate state until after that period. Filters can also be used to remove high frequency content, thus smoothing the signal.

Another method of dealing with erratic signals around the threshold is called the Schmitt trigger. This device acts like two comparators in parallel utilizing a set of independent V_{ref}'s known as trigger points [47]. While the output of the Schmitt trigger is still one of two binary states, the difference is that once threshold A is crossed, state A is not left until threshold B is crossed. The converse also stands true.

V_{ref} points do not have to be static either, they can be flexibly changed based on any number of variables. One common method of accomplishing this is **adaptive thresholding**, the general term used for altering the value of the threshold over time. This may be based upon the current behavior, a prediction about the incoming signal, or through analysis of the signal's past behavior.

While analog techniques require individual circuits for each V_{ref} in use, digital technology allows for a much greater flexibility in the comparison of values. Theoretically, an immense number of thresholds could be set, but in terms of practical use a system could be established which quantizes feature values for table look-up for ease of switching between a larger number of possible states.

3.1.2 Differentials/Velocity

While thresholding offers a relatively static V_{ref} or set thereof, measures of change over time can also be implemented. The first method for measuring these differences is to take the differential (Δ,δ), or the difference between measured values at two points. Differentials can be measured by various orders, or the differences over longer periods of time. For instance, $x[n] - x[n-1]$, would measure the difference between amplitude of the current input sample and the previous sample. Differences need not be calculated on a sample by sample basis; average values, such as RMS (Eq.21, App.B.1), where larger blocks of sample values are combined can be used. Differentials can be used to calculate derivatives, however equations dealing directly with velocity are less mathematically complex and offer similar results.

The velocity of an object, in this case a sampled value, is the change of its speed and direction over time [48]. The rate of change of speed is acceleration in cases of increase, or deceleration in cases of decrease. Velocity is such an important component in describing aspects of sound that it is the part of the MIDI standard, although it is often times simplified to describe only the attack portion of a note event. The basic formulas for calculation of velocity and its related measures as a function of time t, are as follows :

$$Velocity \quad v(t) = \frac{\Delta position}{\Delta time} \tag{4}$$

$$Speed \quad s(t) = |v(t)| = \left| \frac{\Delta position}{\Delta time} \right| \tag{5}$$

$$Acceleration \quad a(t) = \frac{\Delta velocity}{\Delta time} \tag{6}$$

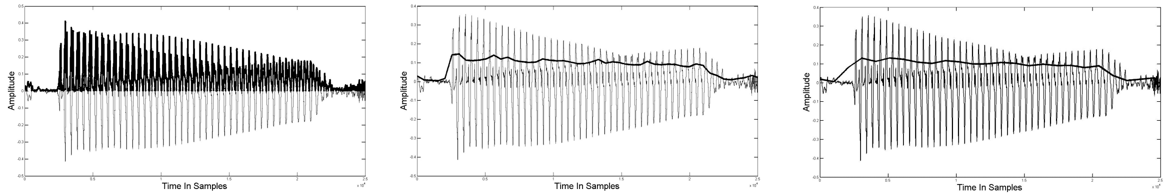
3.2 Amplitude/Envelope Following

Since analog and digital representations of sound are based upon amplitude measure, amplitude is in essence already extracted, and as a result it is possibly the easiest and most frequently utilized feature for many tasks. Amplitude and envelope following, terms used interchangeably in many contexts, attempt to trace the outline of the intensity of the input signal. The earliest known commercial device which packaged an amplitude follower in conjunction with an effect is the Electro-Harmonix Mutron. Released in 1972, The Mutron combines an envelope follower that alters the center frequency of a variable resonant filter based on the player's dynamics, giving an automated wah-wah type of effect without the use of a foot controller [10].

A typical envelope follower involves a pre-amplification stage which ensures that the signal is strong enough to be read properly by the subsequent analyzer stage. The next step will be to either half, or full wave rectify the signal. Half-wave rectification involves elimination of all portions of the waveform with a particular sign, in general the "-" side. Full wave rectification makes all samples positively signed. Various thresholds and discharge are then applied in order to create a usable control signal [49, 50, 51]. Early circuits employed strategies such as peak following which store an incoming voltage in a capacitor and more complex variations allow for close control of the charging and discharging time, allowing tight control of attack and release times [52].

Digital implementations can easily utilize blocks with RMS(EQ.21, App.B.1), peak values (EQ.20, App.B.1), or a moving average/mean. In all cases, the size of the blocks analyzed greatly affect the results achieved. With very low block or running times, there is a danger of imparting the timbre of the source into the feature's values, called **timbre stamping**. Figure 4 displays the trade-off between various block sizes used to calculate the RMS values; it becomes clear that too low of a value (Fig.4a) traces the signal's amplitude directly, while too large a value damages the temporal precision and smears attack and decay times (Fig.4c).

The strategy employed in this research fullwave rectifies the signal and finds the peak value (Eq.20) within a sliding window. The sliding window's values are linearly scaled so that the most recent sample values carry the greatest weight. This allows for accurate rising times and tracking of initial peaks, while minimizing falling times as a musical event terminates. The size of the window is adjusted based on the ratio of f_s to the lowest f_0 possible on the instrument, or respective string in the case of divided pickup use. In order to smooth ripple in the envelope a low pass filter is



(a) 1 Sample Window **(b)** 512 Sample Window **(c)** 1024 Sample Window

Figure 4: RMS Block Envelope Following

employed on the output of the envelope follower, with a cutoff frequency based on the highest f_0. At an f_s of $44.1kHz$, the highest lag in attack or decay is roughly three milliseconds for lower frequencies, while the applications with higher f_0's track very tightly to the envelope of the original signal.

3.3 Pitch Detection

> *"While it is easy for a trained human ear to determine pitch readily, it is a difficult task for an electronic circuit because the waveform from a guitar is rich in harmonics and is very unpredictable. The development of effective pitch-to-voltage converters has been the last hurdle in the production of the guitar synthesizer. The operating of a guitar synthesizer will depend entirely on the accuracy, speed and adaptability of the pitch-to-voltage converter."*
>
> - David Friend, 1977 to the 58th AES Conference [53]

In 1977, David Friend was the president of synthesizer giant ARP, who at the time occupied a large share of an ever growing market. In a memo to the board of executives, company founder Alan R. Perlman responded "*...it seems that we are planning to spend over 25% of our R & D money...to make a product...which is more likely to be a disaster than not*[54]." The resulting product, the ARP Avatar, was the first major breakthrough in commercial guitar synthesis. It was a failure both functionally and commercially, and contributed to the company's bankruptcy in 1981 [55]. What was learned still rings true to this day: pitch detection is a computationally costly and difficult process, and thirty-plus years later still filled with inaccuracies, especially in real-time applications.

Although methods such as the Auto-correlation Function (ACF) and the YIN algorithm have greatly improved pitch detection methods, they are still error prone, particularly susceptible to octave errors [56, 57]. Many well developed commercial pitch detection units boast tracking times of roughly fifteen milliseconds, but can be as high as forty milliseconds with lower frequencies. The preference in pitch detection for the A-DAFx applications (Sec.5.3) relies on a method of mitigating this type of analysis altogether in favor of generating periodicity using the instantaneous phase of an incoming signal.

In recent years Miller Puckette has developed a method for extracting the instantaneous phase of a waveform, which mitigates the latency and inaccuracies commonly associated with detection

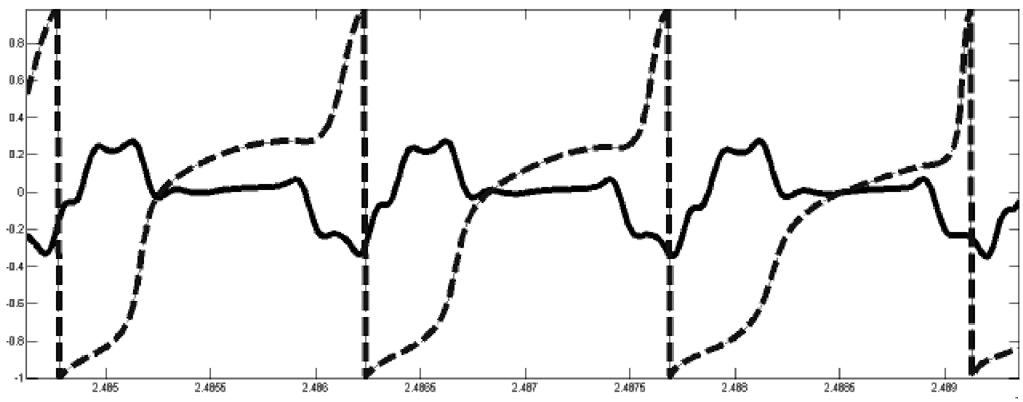


Figure 5: Three Cycles of A Guitar Waveform and Resulting Hilbert Output

of pitch [11, 41]. The method relies on two sets of bi-quadratic all-pass filters which simulate the Hilbert transform, a process placing a signal in phase quadrature with itself.[3] The use of a divided pickup mitigates problems with polyphonic analysis, isolating individual pitches, even when notes are sounded simultaneously. Pre-processing in the form of band-pass filtering for each string is applied based on respective f_0's for each string (Table 1) in order to bring the input signals as close to ideal sine waves as possible [58, 59].

The all-pass filters combine to give a phase shift of 90° between the signals using the following difference equations, where the first two equations represent the 'real' (cosine) component, and the second set produces the imaginary (sine).[4]

$$y[n] = 1(x[n]) + .2569(x[n-1]) - .2605(x[n-2]) - .2569(y[n-1]) + .2605(y[n-2]) \tag{7}$$

$$y[n] = 1(x[n]) - 1.868(x[n-1]) + .8706(x[n-2]) + 1.868(y[n-1]) - .8706(y[n-2]) \tag{8}$$

$$y[n] = 1(x[n]) - 1.946(x[n-1]) + .9465(x[n-2]) + 1.946(y[n-1]) - .9465(y[n-2]) \tag{9}$$

$$y[n] = 1(x[n]) - .8377(x[n-1]) + .0633(x[n-2]) + .8377(y[n-1]) - .0633(y[n-2]) \tag{10}$$

A signal's instantaneous phase and magnitude can be ascertained from the orthogonal set of real signals, on a sample-by-sample basis, using the same equations as those in extracting the phase and magnitude of frequency bins in the Fourier Analysis (Eqs. 24,25,App.C.1).[5] A sine wave input provides a ramp, or sawtooth, output, providing good indication of periodicity even with slightly more complex real-valued signals. The magnitude calculation can also be used in lieu of outlined envelope following techniques. Specific uses will be discussed in context.

[3] The difference equations were taken from Miller Puckette's Pure Data object called Hilbert

[4] The outputs are normalized by scaling for better results, but the transfer function could be modified such that the resulting change in magnitudes is already normalized.

[5] Many programming languages built-in arctangent functions output a wrapped two quadrant phase argument; it is important to use a four-quandrant unwrapped arctangent function for proper results.

3.4 Spectral Content

The result of increased complexity in waveforms in the time domain, is a more complex spectrum in the frequency domain. Breaking a complex signal into its simpler frequency components, by transforming a time domain series into the frequency domain, is referred to as **spectral decomposition**. A number of sub-processes have been developed based on these methods which are collectively referred to as **spectral analysis**. The information provided by these analyses are known as **spectral content**.

Spectral analysis and the information obtained from it, is in many ways the most useful and flexible tool-set in the process of feature extraction. It does however come with a great number of limitations and balances which must be struck depending upon the application. While there are a number of methods for spectral analysis, the most common applications used for periodic signals, are based around variations of the work of Baron de Fourier.[6] The following spectral analysis methods rely on the Fourier Transform and readers unfamiliar with the basics should refer to Appendix C.1.

3.4.1 Spectral Centroid

The spectral centroid is the "center of gravity" of the spectrum of a given sound and is correlated with the high-level feature defined as "brightness" [42]. Brightness is the relationship of the number and magnitude of upper partials a signal has in relation to its fundamental. Other words for brightness which relate to high frequency content commonly used are "brilliance" and "sizzle"[61]. The formula for the calculation of a spectral centroid is displayed in Equation 11, where k is the frequency bin number, and $|X[k]|$ is the corresponding magnitude spectrum. In order to save resources on redundant mathematical operations, bins above the Nyquist frequency are eliminated [62].

$$centroid = \frac{\sum\limits_{k=0}^{N/2} k \times |X(k)|}{\sum\limits_{k=0}^{N/2} |X(k)|} \tag{11}$$

3.4.2 Spectral Roll-Off

The spectral roll-off is defined as the frequency that is the cut-off below which eighty-five percent of the magnitude spectrum is concentrated [42].

$$\sum_{n=0}^{R_n-1} = 0.85 * \sum_{n=0}^{N-1} |X[k]|n \tag{12}$$

While the lower eighty-five percent seems like it could be a useful cutoff, that number is essentially arbitrary. The scaling coefficient or threshold, namely 0.85, accepted as the standard throughout the relevant literature, could be changed to meet various needs. A second point to be made, is

[6]Incidentally, while Fourier's work revolved around heat transfer, the work of his contemporaries Bernoulli, Euler, and d'Alembert dealt with the motion of vibrating strings [60]

that while roll-off is analogous to a low-pass concept, the process could be inverted so that the concentration above a given value would be the focus of implementation.

3.4.3 Spectral Flux

Spectral flux is an example of a feature which takes direct advantage of differentials, and is a key component in other feature extraction tools such as onset detection (Sec.3.5). Spectral flux is the measure of change between adjacent frequency bins over time. This can be expressed by the equation [63]:

$$SF(n) = \sum_{k=0}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|) \quad where \quad H(x) = \frac{x + |x|}{2} \tag{13}$$

The equation displays the summation of the differences between magnitudes n at frequency bins k, from the previous values for n contained in the adjacent frequency bins k. In this particular instance the signal is half-wave rectified, eliminating values for x that fall below 0. The half-wave rectification may not always optimal over full-wave, and the particular instances will be discussed during application.

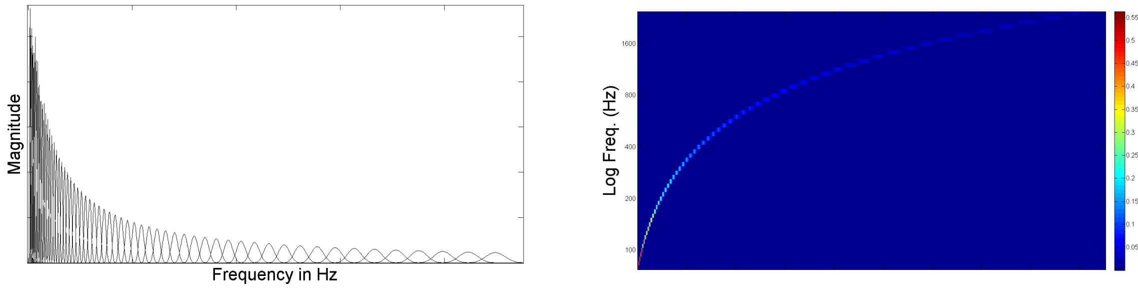
3.4.4 Chroma

Milton Babbitt began using set theory to comprehend, describe, and compose the serialist style of music pioneered by composer Arnold Schoenberg. His "*theory proceeded from a mathematical description of the familiar twelve-tone operations(transposition, inversion, retrograde)*"[64]. The normalized order used in these analyses, reduced pitches down to one pitch-class per note name regardless of the octave, and organized these collections of pitch classes into sets. Work by Shepard which began in the mid-1960s represented continuous pitch on a helix where all pitches with the same name were vertically in line, and their height dependent upon octave [65, 66, 67]. Chroma, as it has been applied in signal processing, stems from the work of these, along with other musicians.

Chroma vectors are a map of the twelve chromatic pitches used in Western Music, where height information is disregarded, and a single magnitude is assigned to each pitch class profile (PCP) [68, 1]. Using the strategy outlined by Peeters [69], the incoming signal is downsampled so that the frequency resolution of the FFT can be increased. Considering that the resolution of the FFT is $\frac{f_s}{N}$, a decrease in f_s provides an automatic increase in resolution if the same size N is maintained.

The next step is to take the FFT of the signal, and convert the linearly spaced bins into a logarithmic representation corresponding to the twelve tones of the equal-tempered scale. The is accomplished by multiplying the FFT by a bank of band-pass filers. The center frequencies of the filters, f_c are calculated using Equation 14, where β is the desired number of bins per octave, and k_{lf} is the quantity of center frequencies, from zero ($f_m in$), through $\beta\times$ the number of octaves.

$$f_c(k_{lf}) = f_{min} \times 2^{k_{lf}/\beta} \tag{14}$$



(a) Log Spaced Filter Bank

(b) Magnitude Response

Figure 6: Log Spaced Filter Bank Example

The minimum frequency, or f_{min}, is chosen by the lowest f_0 possible on the respective instruments.[7]

In both cases a span of five octaves was chosen, with a β value of twelve bins-per octave. Although 36-bin per octave strategies have been employed in other studies, twelve bins per octave were sufficient in the application [70].[8] Each filter in the bank was constructed using a wavetable lookup from a 4096-point representation of a hamming window shape. This allows for linear interpolation of each filter's scaling coefficients, rather than rounding the centers off to the nearest integer value for an existing bin. Other window shapes can be easily implemented varying the amount of overlap between bins. The interpolation method aids in keeping the analysis window size minimal by applying very specific weights to the respective bins, while increasing the accuracy of difficult to resolve lower frequencies.

Filters are spaced with their centers at f_c, and overlapping so that the center of the previous filter is always the start of the next. The overlap was constructed so that higher frequency halves of the bins are generally longer, accounting for the logarithmic stretching of intervals. The bins are then normalized by dividing by the sum of their magnitudes. The filter bank used in guitar processing is displayed in Figure 6a with its magnitude response in Figure 6b.

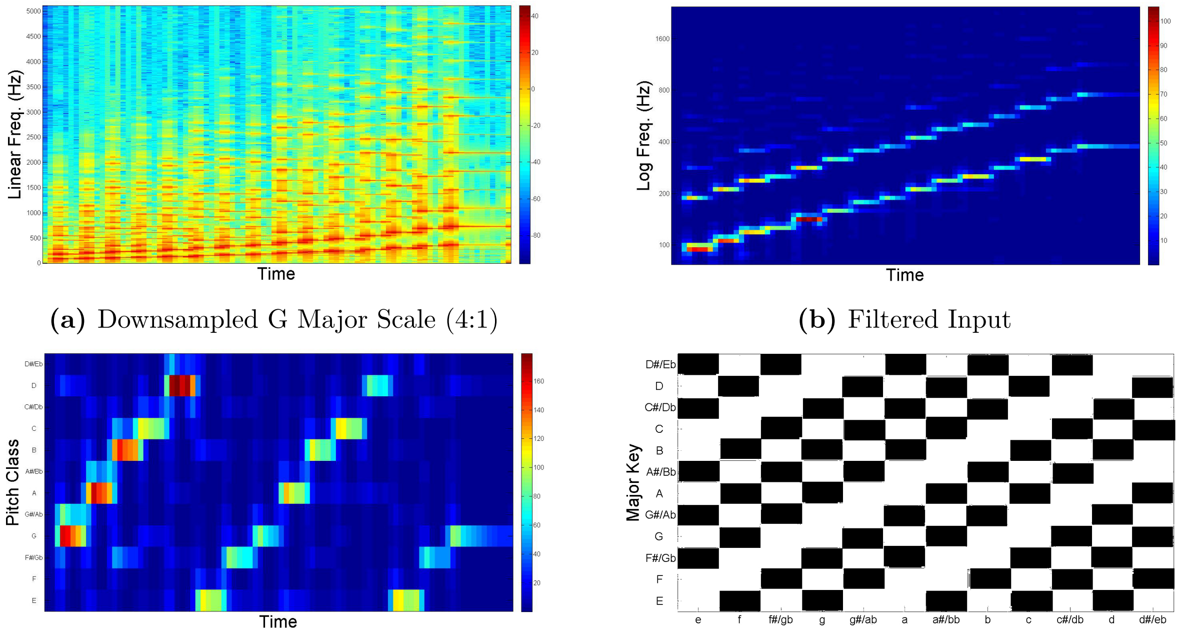
Using a modulo math operation with β (twelve in this case) as the operator, all bins of equivalent mod are summed together to form the twelve PCP's. The resulting chroma vector can then be multiplied by a template for matching to either key (scale), chord, or other harmonic structures. To illustrate the process, Figures 7a-7d are provided. In Figure 7a, a downsampled (4:1) input (a G major scale) is FFT'd using a hamming window. The input is then multiplied with the Filter Bank (Figs.6a,6b), to produce Figure 7b.

Figure 7b is then folded across octaves into the twelve PCPs, shown in Figure 7c. In this case the PCPs are summed across time and multiplied with a key matching template. The key matching template is constructed in a binary fashion using ones for where notes are members of the pitch class set associated with that key, and zero for where they are not. The maximum value of the product of the chroma vector and template indicate the highest correlation, and thus the key. The result in this case, is an output stating the music is in the key of G major - which is correct.

Chroma can provide very powerful information about harmonic content and its evolution, but

[7]In the implementation the vector of center frequencies is extended so that the filterbank has a full-size window at the ends. The half windows of the extra frequencies are then removed.

[8]If a musician tunes in a non-standard fashion, f_{min} can be adjusted accordingly.



(a) Downsampled G Major Scale (4:1)

(b) Filtered Input

(c) Folded Pitch Class Profiles

(d) Binary Key Template

Figure 7: The Process of Extracting Chroma Vectors

it has limitations. Chroma is highly dependent on the matching template, for instance different weights would need to be applied in order to identify harmonic content as a mode, rather than as a major key containing the same pitches classes. Inverted chords, and chords with extensions beyond the triad can be problematic depending on the template used.

The major balance which must be struck, is the trade-off between lag resulting from the time required for collecting enough samples to make a complete window, while retaining enough frequency resolution for the lower end of an instrument's register. In violin applications the window lengths were 182 milliseconds long, and guitar suffered from a trade-off between bleeding in the lower bins, or having good resolution with 371 millisecond windows. An $N/8$ hop-size was adopted to provide control information more frequently, however the information from the beginning of the window still approaches large fractions of a second. While less than a half of a second doesn't seem to be a great deal of latency in the real-world, in musical responsiveness, it can be a lifetime. For this reason, Chroma will be used in the applications for modifying A-DAFx parameters that don't require temporal precision.

3.5 Onset Detection

Onset detection is the identification of when a note event begins. While methods like amplitude following can indicate clearly when a signal goes above a threshold, it may not provide any indication about the beginning of a second musical event that is closely connected with the previous. Onset detection can be useful for identifying disparate events happening within the same musical space.

A myriad of onset detection methods exist based on spectral flux, changes in local energy, combinations of phase and energy, etc... [71, 72, 73, 74]. Onset detection methods are generally selected based upon the specific signals being dealt with. Methods will be examined from MIR literature for use with Pitched Percussive (PP), and Pitched Non-Percussive (PNP) sounds. The most common methods create an onset or novelty function from which peaks can be picked, and thus onset events detected. The examination of existing methods will be applied in the specific contexts of electric guitar and violin.[9]

A number of fairly recent papers (2011-2013) dealing directly with automatic guitar transcription [75, 76, 77, 78], and a fifth dealing with plucking point estimation (2005) [79], were examined in order to review methods of onset detection specifically tailored to guitar. According to Bello [72], "*in order to define a note-event, three parameters are essential: pitch, onset, duration.*" [75, 76, 77] make no mention of onsets, or temporal segmentation in general.

[78, 79] each briefly discuss their methodology and they have assumed that the guitar is a PP instrument. Their methodologies rely on the creation of a detection function based on spectral flux with additional weight given to High Frequency Content(HFC). HFC can be emphasized by high-pass filtering the incoming signal, or linearly weighting the power spectrum, as displayed in Equation 15 [72], where $N/2+1$ corresponds to $f_s/2$, and $X(k)$ is the k^{th} bin:

$$HFC = \sum_{k=2}^{\frac{N}{2}+1} (|X(k)|^2 \bullet k|) \tag{15}$$

There are serious problems with this strategy in a real-world scenario - namely that not every note played on a guitar is plucked/picked. The main issue is that this strategy overlooks the very common and idiomatic connection of multiple notes under one pick stroke (Sec.4.6.2). Without using longer windows (i.e. 4096 samples at 44.1 kHz), it becomes difficult to create a detection function that can be peak picked when the string is not re-excited.[10] As a result, the guitar should be treated as a PNP instrument in order to capture all possibilities.

Previous work has discussed dividing the spectrum into sub-bands to generate individually tailored detection functions[80]. The strategy outlined here uses two methods simultaneously : HFC in conjunction with a "pitch based" scheme borrowed loosely from [72, 81], which utilizes the previously described Hilbert function rather than traditional pitch detection (Sec.3.3) to detect all other onsets.

Upon input the signal is split into a high-pass filtered version, and band-passed input to the Hilbert function. The high-passed input is FFT'd using a short window (64 or 128 samples) with a hop size of N/2. The spectral flux is calculated, fullwave rectified, then median filtered. Median filters are unique in that they smooth the signal while preserving transient edges [52]. Rather than peak picking, when the function is above a threshold, a plucking event is determined to have occurred, and the assumption is the detection function will drop below the threshold prior to the next pluck event.

In order to detect periodicity with a greater resolution, the band-pass filtered signal is over-

[9]The author is aware of emerging uses of neural networks as methods for creating low-latency, accurate onset detection, however the algorithmic complexities are beyond the scope of this work.

[10]The author spent an extensive period testing various window sizes and types, overlap, smoothing, and peak picking algorithms.

sampled with rates based on the frequencies pertinent to the respective string. Zero crossings are counted to determine the periodicity of the Hilbert functions ramp, and each period's length is stored in a buffer. The value in the buffer is constantly compared to, and replaced by, the incoming measure of period. If there is a deviation (+/-) beyond a certain measure in the number of samples, a new onset is detected. If the maximum sample value within the period is not above a set threshold, no note event is considered to exist. The Hilbert based scheme can also be applied to violin for onset detection.

4 Examination of Expressive Markings

"By itself notation is not music; it is only the vehicle by which the composer indicates his ideas and wishes to the performer."

- Gardner Read [82]

Notation is a vehicle for transmission of musical ideas in much the same way that text is used as a printed version of verbal communication. Prior to audio recordings, notation was the only way to transfer musical ideas, other than performance itself. Interpretation of how both the composer and performer view the represented ideas means that notation on many fronts is not absolute. Composers have continually developed methods of controlling expression with the addition of new symbols, alterations of existing ones, and radically different ways to display events such as graphical notation [82, 83, 84]. Information in this section on markings has been collectively gathered from [85, 86, 82, 83, 87, 32, 35, 34, 88, 89, 90], along with scores from musical literature.

This section is the union of three areas in the originally described chain - it opens the doors of a methodical DSP-based examination of a performer's interpretation of traditional notation. During the course of the tests which follow, assumptions about aspects of music or performance techniques as they appear in signal processing literature are challenged. The results will be used to begin developing musically based, rather than perceptual or sonically based mappings. The author chose to use notation as a basis for visual representation and consistency, however the types of techniques and materials discussed are not relegated to use by only those who can read music.

Violin performance samples were provided by Scott Tixier, a Parisian violinist residing in Brooklyn, New York who is conversant in both the jazz and classical idioms.[11] Mr. Tixier typically uses a French Violin constructed in 1889, but for the samples he provided, utilized a Zeta Electric Violin which contains piezo elements in the bridge. Although Mr. Tixier is a very experienced violinist, it took him roughly an hour to feel comfortable with the slight change in scale length (distance from bridge to nut), different strings, and the bow used with the Zeta which is constructed from synthetic materials. The standard 1/4″ output was utilized with a Monster Cable connected through a direct box to a MOTU UltraLite MkIII USB 2.0 interface. Samples were recorded at an f_s of 44.1 kHz and quantized using 24 bits.

[11] In 2012, Mr. Tixier performed with Anthony Braxton, was featured on National Public Radio and released a solo album entitled *Brooklyn Bazaar*.

pppppp ppppp pppp ppp pp p mp mf f ff fff ffff

Figure 8: Dyanmics

The tempos used in the initial testing to be discussed were 60 and 120 beats per minute. The only subdivisions of the beat were whole, quarter and eighth notes. This makes the automated division or segmentation of the audio files extremely simple. Mr. Tixier played to a metronomic 'click track' in order to maintain consistency. Very little post alignment or editing was performed in order to preserve each element of the performer's interpretation of the markings. In all tests there is consistently only one variable. For example an A major scale was utilized over two octaves ascending and descending - the variable in most cases was the type of articulation or phrasing. Guitar testing was conducted predominantly by the author, using the standard electromagnetic pickup systems, and the nature of the specific instruments used in the respective testing will be identified.

4.1 Dynamic Markings

> *"...the degree of power of a 'f' depends on the tonal qualities of the player or of the instrument he [she] is using; on the discretion of those playing with him [her]; or on the acoustic properties of the room in which he[she] plays. The value of dynamic indications, is therefore always relative so far as the performer is concerned."*
>
> - Carl Flesch in The Art of Violin Playing, Book 2 [35]

The term dynamics comes from the Greek word for "power" and although this power can be measured empirically (amplitude), the aim of marked dynamics in a score are to provide a performer with indications of relative loudness or softness. A marking can provide information as to how loud an individual note should be, up to an entire movement or work [82]. Amplitude is an empirical measurement, loudness is perceptual, and dynamics are the composer's tool for informing a performer of how to make the first accurately represent the second.

In [4] it is assumed that "*the musical counterpart to loudness is called dynamics, and corresponds to a scale ranging from pianissimo(pp) to fortissimo (ff) with a 3-dB space between each dynamic level.*" This only allows for a range of six stepped grades of loudness, or an 18 dB range. When compared to something like the compact disc which has 96 dB of dynamic range, or the human ear which has a range of roughly 120, this is very limiting. According to Read [82], the dynamic scale began to expand further citing Tchaicovsky's use of *pppppp* and *ffff* in the Pathetique Symphony and 1812 Overture respectively.

Compared to other instruments the power of an acoustic guitar is limited. The guitar's intensity can only reach approximately the *mf* of a piano [89]. Depending upon the construction of a violin, and its qualities of projection, it has a greater range and possibility of intensity. However, this study isn't dealing with acoustic instruments; in both cases amplification is an integral portion of the performance system, allowing for this range to be greatly extended.

(a) Crescendo **(b)** Decrescendo

Figure 9: Crescendo/Descrescendo

A 3-dB width of each dynamic, an 18-dB range, still assumes some form of empirical measurement. Musically speaking dynamics in human performance will never be this perfect as illustrated by the quote from Flesch. Milton Babbitt was the first composer to utilize serialized dynamics in 1947's *Three Compositions for Piano No.1*, which attempted to force performers to interpret twelve grades of absolute dynamics [84]. The result was a demand for wider ranges than many acoustic instruments permitted, and proved to be an unrealistic practice.

The opposite of terraced or hard graded dynamics are gradual increases or decreases in intensity. These are generally indicated using wedge shapes which began to appear in notation in the late 17th Century [82]. The marking which indicates an increase in intensity is the crescendo, while a decrease is a decrescendo (See Fig.9), which is often interchanged with the term diminuendo. A number of variations exist which define how intense the difference should be when not flanked by dynamic markers, and also the rate of change.

To test true dynamic range the performer begins at a theoretical *pppp* and increases intensity up to a $ffff$, then decreases back down to *pppp*. A single pitch is repeated twenty-four times allowing for a number of gradations to occur with the focus solely on the dynamic range. Each exercise was conducted four times and all results were kept as a consideration of human performance capability.[12] Dynamic ranges were calculated by finding the peak amplitude (Eq.20) of the absolute values of the 24 notes in each exercise. In every case, up-bows resulted in the minimum dynamic range, and downbows resulted in the maximum. The complete results are as follows:

Note	f_0(Hz)	String	Num Reps.	Min(dB)	Max(dB)	Mean(dB)
G3	196.00	IV	4	15.6	20.5	18.6
D4	293.67	III	4	16.4	23.5	20.5
D4	293.67	IV	4	20.2	26.6	23.9
C5	523.25	II	3	19.2	27.3	23.6
A5	880.00	I	4	15.2	23.2	19.2

Table 2: Violin : Test Results of Individual Dynamics

Testing of dynamic range was conducted for guitar in the same fashion as the violin, except the pitches selected are one octave lower. A Fender American Standard electric guitar containing single-coil pickups was used, but realizing that the frequency response and position of pickups would play a part in the dynamic range, two separate tests were conducted on the same data set. The first set used the pickup closest to the bridge, and the second used the pickup closest to the

[12]The fourth test of C5 was thrown out due to uncontrollable harmonics of the 60 Hz Hum induced by the lights.

neck.

Note	f_0(Hz)	String	P-Up	Num Reps	Min(dB)	Max(dB)	Mean(dB)
G2	98.00	6	Neck	4	17.2	25.4	22.3
D3	146.84	5	Neck	4	11.3	20.1	16.7
D3	146.84	4	Neck	4	13.3	18.8	16.0
C4	261.63	2	Neck	4	11.4	19.2	14.8
A4	440.00	1	Neck	4	20.0	37.9	28.0
G2	98.00	6	Bridge	4	20.3	28.7	23.9
D3	146.84	5	Bridge	4	17.8	20.8	19.6
D3	146.84	4	Bridge	4	17.0	19.9	18.4
C4	261.63	2	Bridge	4	12.8	18.2	16.1
A4	440.00	1	Bridge	4	12.7	22.3	17.3

Table 3: Guitar : Test Results of Individual Dynamics

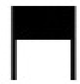
In the case of guitar testing, up-strokes brought down the dynamic range, supporting a claim that will be made in Section 4.2. No electronic means, such as gain controls, or an external volume pedal were used. The response of an electric instrument plugged into an amplifier, which is usually tailored to the frequency response and Hi-Z input, generally has a wider dynamic range than with the type of Lo-Z input on the MOTU audio interface. What is absolutely clear in the guitar tests is that the resonant notch of the electromagnetic pickup impacts the apparent volume of certain higher f_0's by enhancing or dampening their component frequencies unevenly.

The range of crescendos/decrescendos is slightly harder to calculate - at what point does the sound actually start and stop? When starting from the lightest attack like the *'niente'*, which means *"from nothing"* in Italian, it becomes a subjective judgment of where the sound truly begins. A realistic threshold must be established above the noise floor of the system in use. The same decision making must be implemented for diminishing sound as a note is released. The guitar lacks the ability to sustain continuously or crescendo a single note, and is therefore left out of this testing.

The results of these tests combine to bring the assumption that even a solo performer is capable of only six grades of 3-dB into question. It is legitimate to assume that some performers are not capable of consistently executing these dynamics at the same absolute levels, and that even professional players may not accomplish a wider dynamic range in every register of their instrument. However, given that electronic means of amplification and volume control are generally used with these instruments, combined with raw numbers from these tests, the stance taken is that experienced performers can achieve greater than an 18-dB range in a controlled fashion.

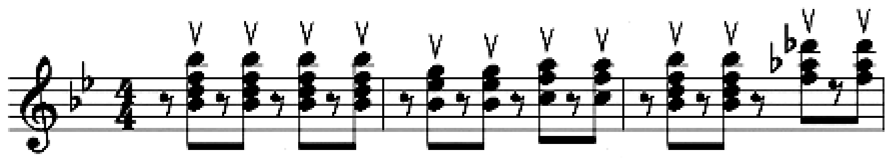
4.2 Direction of Excitation

Picking and bowing can happen in only two directions, thought of as up and down. While guitar picking is actually a vertical up and down motion, violin bowing is a lateral motion which varies depending upon the string or strings the performer is playing. As demonstrated in the previously discussed dynamic testing (Sec.4.1) both bowing and picking direction are known to have a large impact on the players ability to execute certain musical gestures with various amounts of force and



(a) Down Bow/Stroke

(b) Up Bow/Stroke



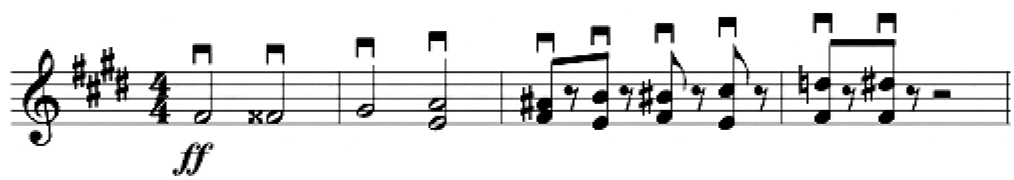
(c) Desmond Dekker, Guitar, *The Isrealites*

(d) Vln. I, mm.109, Tchaikovsky, Symphony No.6, 3rd mvmnt.

Figure 10: Direction Symbols

control.

Down bowing/picking or down strokes are considered to be the stronger of the two directions possible, allowing a greater application of force. Heavy metal and punk rock music often execute a series of down strokes in order to emphasize ‘power’ and a classic example in string literature is Stravinsky’s use of repeated downbows in *The Rite of Spring.* The opposite example is in Reggea and Ska music where up or weak beats are emphasized, employing syncopation, by using up-strokes. Examples of up and down bowing from the literature are given in Figures 10d [85] and 10c.

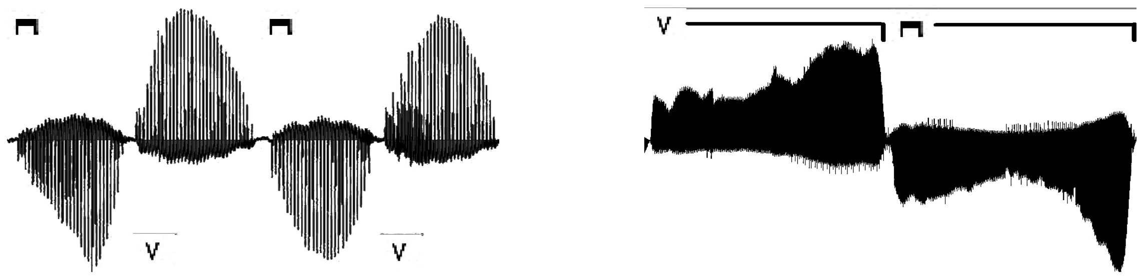
4.2.1 Direction of Bowing

In time domain waveform representations, bowing direction is easily seen by the concentration of negative or positive energy, or bias of the signal. Returning to the physics of how a bowed string is displaced, the time domain waveform demonstrates the difference between the slipping point with a peak, and the concentrated energy of the sticking phase in the region of opposite sign.

Figure 11 displays the typical behavior seen in all bowing tests, including a set conducted on the electric guitar. It may be fast and reliable to detect bow direction without any external hardware. In Section 5.4.1 a Signed Envelope Follower(SEV) is proposed. Where traditional amplitude/envelope following applications use absolute amplitude of a signal, the SEV is only concerned with the sign value of the average energy.

4.2.2 Direction of Picking/Plucking

While bowing direction is a recognizable feature using a relatively simple observation, this does not hold true for the direction of picking in the electric guitar. In the examination of the violin, a piezoelectric element is the transducer under scrutiny. While piezos are based on direct contact



(a) Individual Notes **(b)** Slurred Notes

Figure 11: Waveform of Down/Upbows

and displacement(Sec.2.4.1), electromagnetic pickups detect the motion of a string in free space. The problem this creates is that the during the induction of motion, the coil ‘sees’ vibration not only vertically, but horizontally as well, homogenizing all types of vibration into a single signal. The second issue, is that even if picking direction were easily detectable, the steady state waveform of the plucked strings has a tendency to distribute energy equally in both the positive and negative region of the waveform.

While humbucking, lace sensor, single coil, and piezos were used to attempt to ascertain picking direction, none were successful. The direction a pizzaccatto was executed with, consistently created a single peak that corresponded to the direction of pluck. What was determined at this point deals with the configuration of the pickup in question. In all of the cases attempted with the electric guitar, the string either hovers over, or makes contact with the top of the transducer. The electric violin’s strings rest in between two piezo elements, likely of balanced, but opposite charge. This accounts for the fact that the displacement is more easily seen in a simple bi-directional manner.

4.3 Position of Excitation

As displayed in Equation 2, the modes of a vibrating string are affected by the position from which they are excited. Composers and performers are familiar with the types of sounds that are yielded by these specific positions of excitation. Exciting the string near, and even over the finger or fretboard, is referred to as **sul tasto**. The resulting sound is subjectively described as *sweet*(*dolce*) or *round*. Excitation near the bridge is called **sul ponticello** and when utilizing the bow its effect is generally described as *glassy*, *eerie*, or *harsh*. When a guitar string is plucked near the bridge it is described as *clanky*, or *metallic*.

Generally speaking, performers excite the string at some point midway between the bridge and beginning of the finger/fretboard, marked as **norm.**, indicating normal position. Plucking position has a great effect on the spectral evolution as well as overtones contained within a signal. Bowing position doesn’t drastically change the harmonic structure, but as will be demonstrated it has a great impact on how rapidly the spectrum changes.

4.3.1 Bow Position

Guettler states in [33] that “*the bowing has less spectral influence than one would imagine*”, attributing the change in sound to the bow pressure required to set the string in motion near the

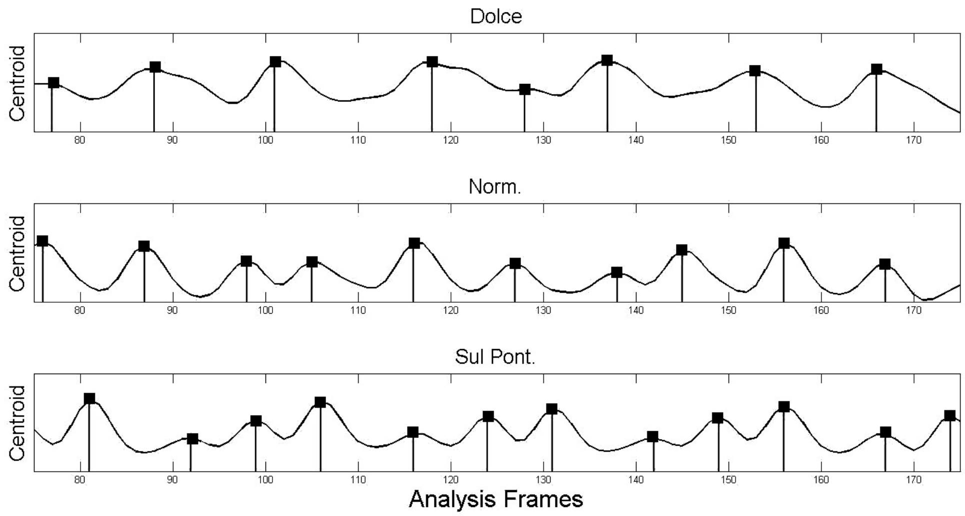


Figure 12: Moving Centroid Based on Bow Position

bridge. Looking at a dataset of spectrograms, it becomes clear why this assumption might be made with the eyes. However, following the ears, this makes little sense. This also does not seem to make much sense in conjunction with the model which physics provides.

To put bowing position to the test, the performer was asked to play five notes at a dynamic of mf in three different positions: directly over the bridge(*sul pont.*); in the Helmholtz region of the string (*norm.*); and at a location marked one inch from the edge of the fingerboard (*dolce*). Each note was played with both up and down bows two times.

What was observed is that the static spectrum contains some additional high frequency content as the bow is brought towards the bridge, but the average spectral centroid and apparent flux are not greatly altered in terms of where their centers lie. The difference in timbre results however, from the way in which these spectral features rapidly change in time.

The periodicity of spectral flux and centroid are very different depending on bow positioning with much longer time in cycles observed the further away from the bridge the player was. As an example refer to Figure 12 which displays 100 analysis frames (1.16 seconds) of the centroid of an open G string (196.00 Hz/G3). Each frame consists of 1024 samples, with a hop size of 512, where the results are low-pass filtered forward and backward to preserve phase information regardless of frequency [91]. Peaks were detected and plotted to augment clarity in seeing the varied cyclic behaivor.

It becomes clear that the rate of change in the centroid is different among all three bow placements, being slowest and most narrow in dolce, while the fastest and greatest change in peaks resulted from sul ponticello. Patterns of deviation based on position were not consistent between notes and this is the direct result of fingering a note changing the string length. The point to be drawn here is that bowing position does in fact change the spectrum, not necessarily in terms of overtone structure, but certainly the spectral envelope and interaction of the respective harmonics. The easily accessible varied motion can be mapped as a control similar to a low-frequency oscillator.

4.3.2 Picking/Plucking Position

To test the change in spectral content through change in picking position a Godin LGX-SA with two Humbucking pickups was used. The excitation of modes in guitar is dependent on two factors: the picking position and the pickup selection. The 6th string, E2, was marked at distances from the bridge at 1/3rd, 1/5th/, 1/10th and 1/36th of the total length. Although the scale of the instrument was known, the string was measured from saddle to nut to account for a change in distance caused by adjusting the intonation of the string.[13]

Each position was plucked ten times and allowed to sustain for three seconds (40 plucks). The same test was repeated three times using the bridge, middle, and neck pickups independently (120 plucks). The results were analyzed using the spectral centroid and roll-off as their measure. The prediction that can be made based on the model of the plucked string, is that the centroid and roll-off will initially be much greater than the fundamental and drop rapidly as the vibration settles into a steady state. With the consideration of modes not excited, the greatest high frequency content should come from plucks close to the bridge.

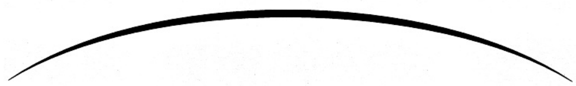
Uniform plucks were attempted not only in position, but also in dynamic, however human error can occur. The repetition allows for averaging and helps to eliminate some inconsistencies in playing. In order to analyze the data in a uniform matter, the groups of ten plucks underwent some pre-analysis processing. The onset portion of the note was detected and samples before the attack were eliminated. This aligns the peaks and troughs of the waveform in time. The samples were then normalized so that all of the maximum peaks had an amplitude of one. To keep the number of analysis windows consistent, all notes were then truncated to a length of two seconds. The process turned out to be so accurate that when looking at a zoomed-in time domain waveform of all ten notes overlaid, it was difficult to tell how many examples were actually used - the appearance was one thickened waveform.

The behavior seen is that the spectroid and roll-off increase as the bridge is approached. With the plucking point at a distance of roughly 1.5 cm from the bridge, the first mode not excited will be the thirty-sixth! That means there is a great deal of high frequency energy contained in the signal and as a result the values for centroid and roll-off are driven extremely high during the excitation portion of a note event, but the energy dissipates even more quickly than in other positions.

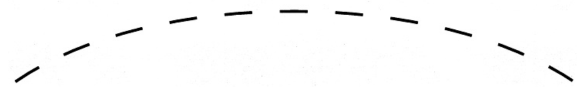
Playing the same note on multiple strings must also yield very different spectral results. By playing the same note on different strings, the same f_0 is excited, but there will be significant differences in string length, bearing in mind that the guitar's strings require a fairly uniform tension per unit of mass.

To judge the relationship of playing the same note on multiple strings, the note E4 (329.63 Hz) was selected as it is executable on five out of six of the guitar's strings. A fixed distance of 11.5 cm from saddle was used as the point of excitation in every case. Only the neck pickup was used in order to maintain consistency across tests. In all, ten plucks were executed on five notes (50 plucks), and the same alignment and normalization procedure from the previous test was applied. In the case of fretting a capo was utilized to keep 'fretting-hand' pressure consistent, and all the

[13] In order to 'tune' the interval relationships between frets, often times the string length must be adjusted by moving the saddles on the bridge forward or backwards.



(a) Phrase / Slur



(b) Broken Phrase Marking

(c) Slurs: Vln.II, mm.78, Wagner, *Prelude* from *Tristan and Isolde*

Figure 13: Phrase and Slur Markings

focus placed on regularity and accuracy of placement in picking.

String	Fret	"Length"	1st Mode Not Excited
5	19	22.1 cm	1.92
4	14	29.2 cm	2.53
3	9	38.6 cm	3.35
2	5	48.7 cm	4.23
1	0	54.8 cm	4.76

Table 4: Guitar : Modes Not Excited (E4 (329.63 Hz) on Strings 5:1)

The first factor to consider is what modes will not be excited; the first mode not calculated, along with the measure of length that fretting shortens the string to, are displayed in Table 4. One might note that the 54.8 cm of the first string is longer than the established scale length of the instrument. As in the previous test, this is also due to the adjustment of the length of the string for the purposes of intonation. As to be expected, when the string length was shortened, driving the value for the first mode not excited down, the high frequency content present in the onset greatly increases. The longer the string length used for the same pitch, the longer it takes for the centroid and roll-off to settle into steady values.

4.4 Slurring and Phrase Markings

A musical phrase is a set of successive notes that are grouped together, and phrase markings indicate how connected this content should be. There are two basic markings which indicate phrasing and while musically they accomplish the same thing, physically they do not. Traditionally the phrase marking, or slur displayed in Figure 13a indicates the connection of notes without break, called legato. Detaché is the opposite of legato, with varying degrees of how separated the notes actually are. On the violin this indicates that all notes should be connected under one bowing, as demonstrated in Figure 13c [92].

Slurs on the guitar indicate connection of notes with one pick stroke. It can also indicate the sustain of a pitch on one string while new pitches are introduced on others. Methods which guitarists generally employ to connect notes under one pick stroke are pull-offs and hammer-ons, which will be discussed in Section 4.6.2, and various types of arppeggiation where new notes are played while others are sustained.

The broken phrase marking displayed in Figure 13b indicates the connection of musical content where physicality may limit actual slurring or phrasing. The broken phrase marking is used in things like extended phrases, string skipping, or string changes in the middle of a musical idea where the composer wants to indicate musical connection.

There are two special cases of phrase connection which happen with the guitar. The first involves arpeggiation, and sweep arpeggios. Arpeggios are simply broken chords, or converting a harmonic relationship into a melodic one. It is very easy on guitar to sustain notes on one string while exciting another. Although these will not always be thought of as phrases musically, in many cases they can be. Sweep arpeggios take advantage of a series of pick strokes in a uniform direction but across strings; while the notes are heard as individually sounded pitches, when executed properly they appear to be connected in a legato manner.

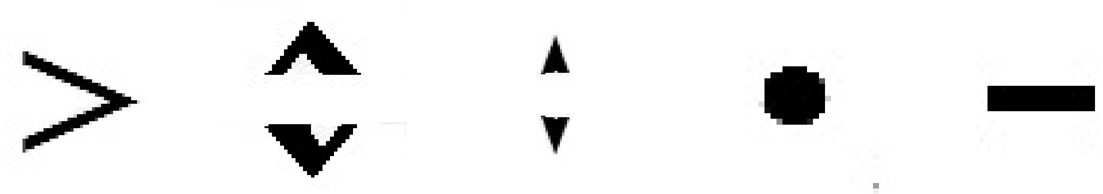
The second special case of legato playing is the technique known as tapping. Instead of the picking hand being used to strum or pluck, the fingers or plectrum are used to excite notes directly on the fretboard in conjunction with the fingering hand. This allows for the production of a very rapid, legato series of notes. The technique was made popular by many guitarists of the late 1970's and early 1980's, notably Randy Rhoads and Eddie Van Halen, and generally exploits only three fingers and arpeggio patterns. Electric guitarists, such as Stanley Jordan and Guthrie Govan, have expanded this to include up to eight fingers, asymmetrical scale runs, and full-on two handed counterpoint in the style of keyboard instruments. Both sweep arpeggios and a right-hand tap are used in Guitar A-DAFx example number one (Sec. 5.3).

4.5 Bowing Techniques Based on Accents and Phrasing

According to Rosenbaum [93], "*Accents are important to musical expression because they add variety and can even indicate higher level information, such as the induction of beat or metric structure.*" In Figure 15 [94], beats one and three are emphasized through the use of such accents. The catalog of accent markings has been expanded drastically since the turn of the 20th century and the result is that there are many of them. Rather than examine 'all' of them, a handful of the most common accent markings and their uses will be discussed in context. The other issue is that there are generally multiple techniques applied to the same type of marking, and consequently various names for the same types of techniques. The terms most common among the reviewed literature are what will be used in this section.

The accent markings discussed fall into two basic categorizations, pressure and percussive, which effect the intensity and length of a musical event respectively. Common pressure accents are pictured in Figures 14a, 14b, and 14c while Figures 14d and 14e deal with note length, or percussive excitation. The staccato indicates a truncation of the note length, or a percussive attack, and the tenuto marking means to carry a note through its full duration, even if it is not connected via a phrase marking to the next note. The following tests were conducted on eleven common bowing techniques.[14] The main feature examined was the amplitude envelope of the technique in question

[14] Additional tests were performed but the results did not provide conclusive evidence as to how the techniques



(a) Martelé **(b)** Marcato **(c)** Staccatissimo **(d)** Staccato **(e)** Tenuto

Figure 14: Basic Accent Markings

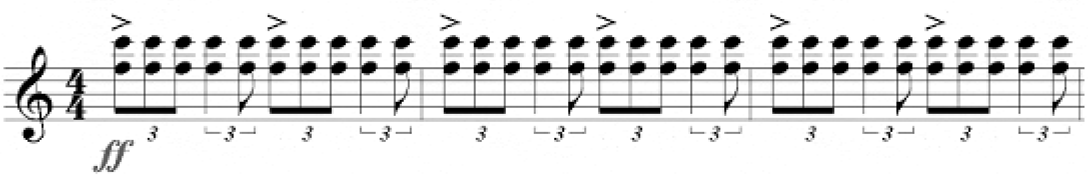


Figure 15: Martelé: Vln. I (Full Section), mm.1, Debussy, *Fêtes* from *Nocturnes for Orchestra*

and the results will be discussed based on groupings by similarity, termed for these purposes as 'shapes'.

On/Off String	Technique	Number of Notes	'Shape'
Off	Controlled Spiccato	120	Double Peak
Off	Ricochet	240	Sharp/Double Peak
Off	Volante	240	Sharp/Double Peak
On	Marcato	120	Ramp/Bell
On	Martele	120	Ramp/Bell
On	Slurred Staccato	240	Triangular
On	Staccato	240	Triangular
On	Detache Moyen	240	Bell
On	Grand Detache	240	Narrow Bell
On	Petite Detache	240	Bell
On	Loure	240	Narrow Bell

Table 5: Tests Conducted on Bowing Techniques

Martelé(Fig.14a), also known as Martellato, means 'hammer bow.' The initial portion of the attack is emphasized by literally hammering, or applying a striking force along with the bowing direction. While it takes longer to excite the string than other methods, once the energy of the bow is distributed through the string, a powerful attack is heard. Applying even more force, and considered the strongest attack is the Marcato(Fig.14b), also referred to as a tree or rooftop accent. Occasionally, the Staccatissimo is used in place of the Marcato in order to alleviate confusion between Marcato and the symbol for an up-bow (Fig.(10b). An excerpt from Brahms' *First Symphony* using the Staccatissimo is displayed in Figure 16 [95], and a combination of Martelé with staccato is displayed Figure 17 in an excerpt from Stravinsky's *Renard* [96].

The Marcato and Martelé were each executed 120 times and the information about them aver-

could be utilized effectively in conjunction with the signal processing techniques, or without being redundant in generating control information.

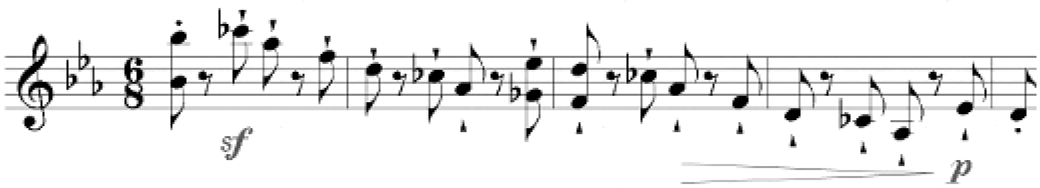


Figure 16: Staccatissimo : Vln. II, mm.97, Brahms *Symphony No.1*

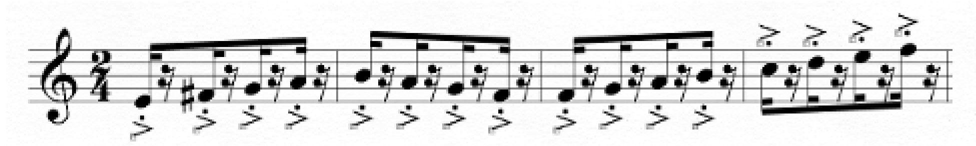
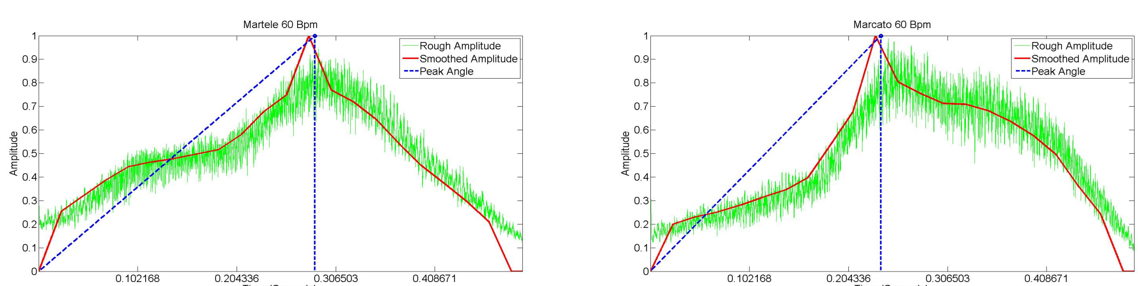



Figure 17: Martelé w/Staccato: Vln. I, mm.48, Stravinsky, *Renard, Allegro*

aged in order to obtain a general sense of the amplitude envelope, the velocity (angle) of attack, and if there were any anomalies preset in the signal that otherwise would not be considered. The hammering motion can be clearly seen in the slow attack which when as stated before, dissipates to produce a strong peak. Notice the difference in the attack slopes or curves between the two bowings; although they reach their peak amplitude in roughly the same amount of time, the Marcato has a much deeper up-slope as a result of a harder impact on the string. Because the results are normalized, it is difficult to see the absolute difference between the two strokes, however the shape of execution of these markings is quite clear.

The next set of bowings do not necessarily have a marking associated with them, but are usually indicated by the use of their name above the portion of the score where the composer would like the performer to begin employing them. Ricochet, Volante, and Controlled Spiccato, are all off-the-string methods which employ a type of bouncing of the bow on and off of the string. An example of Richochet, also called *jeté* in French, is displayed in Figure 19 from Stravinsky's *Firebird*[97].

Examining the results of testing in Figure 20, it is clear that each method results in two distinct peaks in amplitude. The first peak can be attributed to the bow initially hitting the string, and the second peak as the bow leaves the string. Open string execution in the data set, versus fingered notes tends to drive the value of the second peak up because an open string vibrates longer and more freely than a closed one, even after bow pressure has been removed.



(a) Martelé Envelope

(b) Marcato Envelope

Figure 18: Ramp/Bell Shapes

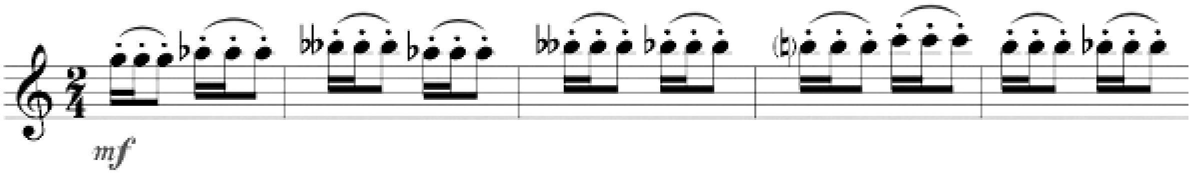
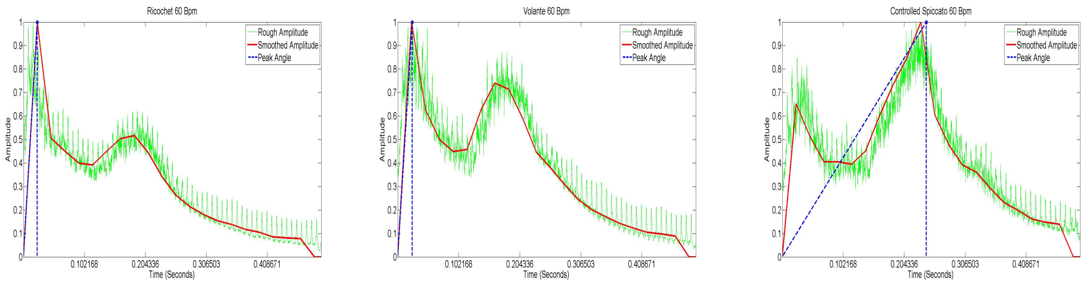


Figure 19: Jeté: Vln.II, mm.49, Stravinsky, *Infernal Dance of All Kastchei's Sunjects*:*The Firebird(1910)*



(a) Ricochet Envelope **(b)** Volante Envelope **(c)** Controlled Spiccato Env.

Figure 20: Sharp and Double Peak Shapes

The Staccato marking (Fig.14d) indicates many things, but a standard use is to designate that the time of a note event should be cut in half. An example of staccato is presented in Figure 21 from Haydn's *String Quartet No.4 in D Major* [98]. If executed in conjunction with no other marking, as displayed in Figure 22b, the Staccato should produce a somewhat triangular shape with an even attack and decay time. The Slurred Staccato combines a phrase marking or slur with the Staccato marking indicated underneath the slur itself. This tells the performer to bow in the same direction, but the bow movement is ceased in between notes. Because the technique calls for the bow to be left on the string during these stops, it causes an abrupt decay.

The last set of bowing techniques require the player to keep the bow on the string while detaching the notes. While the same type of bow stroke is required in all cases, the portion of the bow used dictates the final shape of the envelope. The three techniques are Detache Moyen, at the frog, Grand Detache using the more central region of the bow, and Petite Detache which utilizes the point or tip of the bow. The amount of force available at the frog versus the tip and the amount of bow length the player has available may cause them to conserve on bow speed, depending upon the tempo of the music. The contribution of available length and power give these bowing techniques their character. The sampled amplitude values do not attack or decay because the notes are connected to adjacent notes.

Loure, which can be indicated using a phrase, a set of tenuto markings, or simply the word

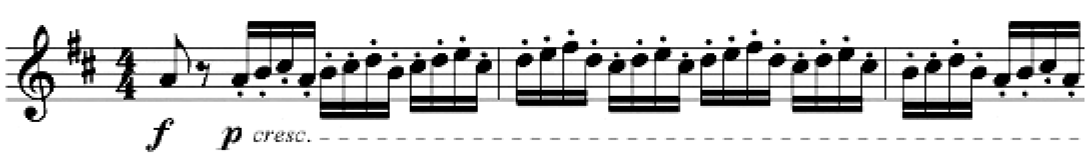
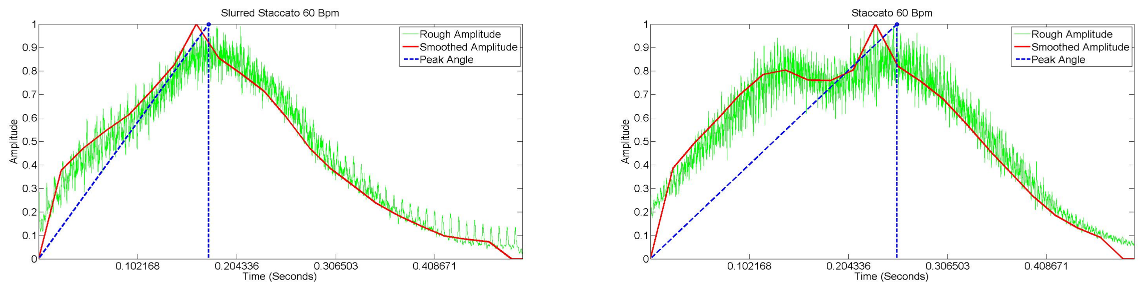
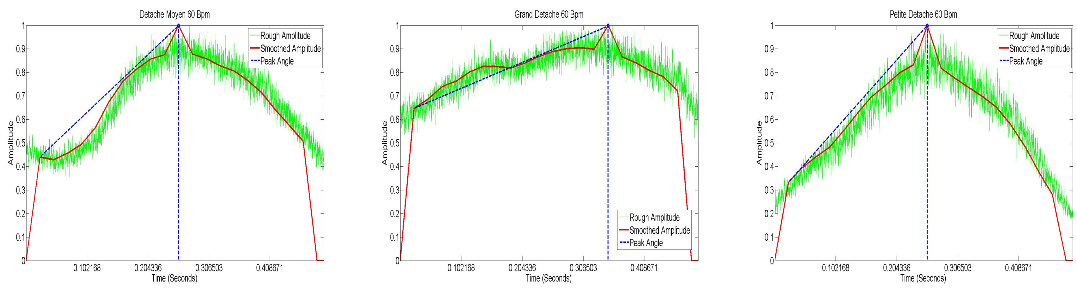


Figure 21: Staccato : Vln. I, mm. 43, Joseph Haydn, *Presto Scherzando* from *String Quartet No. 4*



(a) Slurred Staccato Envelope **(b)** Staccato Envelope

Figure 22: Triangular Shapes



(a) Detaché Moyen Envelope **(b)** Grand Detaché **(c)** Petite DetachéEnvelope

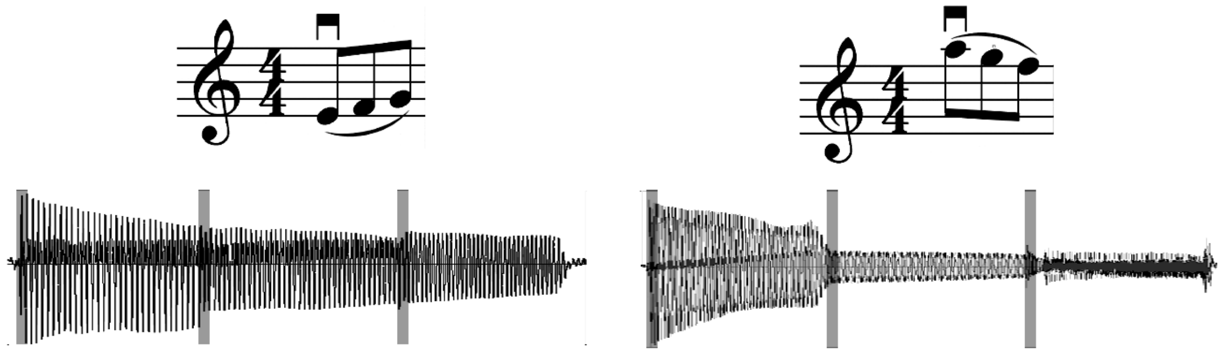
Figure 23: On Bow Detaché

itself, is also an on string legato bowing technique. It was omitted from display because there isn't really an attack or decay when connected notes are executed properly. It should also be noted that Figure 23a appears to wrap around from end to beginning; this is likely the result of human error during recording where the performer is consistently slightly ahead of or behind the metronomic pulse.

The unique characteristics of each accent can offer composers and performers an important and flexible toolset for control. By using velocity and its related measurements (Sec.3.1.2), the characteristic angles of attack and decay can be exploited for various uses in control. Combining onset detection with velocity, the attack information could determine some static value for the remainder of the duration of that note, or a continuous measure could be applied. The applications are only limited by the imagination of the programmer of the A-DAFx.

4.6 Picking Techniques Based on Accents and Phrasing

Although bowing possesses a greater variety of organic methods for string excitation than the plectrum, the freedom of the picking hand and its proximity to the strings allows for a rich body of alternate techniques as well. These include, using the palm of the hand on the strings, the right hand on the fingerboard, and individual fingers in conjunction with the pick itself. Techniques discussed here will center around those based on the interpretation of accent markings.



(a) Hammer-On **(b)** Pull-Off

Figure 24: Hammer-On and Pull-Off Analysis

4.6.1 Staccato

Staccato notes on guitar can be executed using one of two main techniques: palmuting using the picking hand, or releasing the pressure of the finger which is depressing the note, causing it to cease vibration. Palmuting involves placing the palm of the picking hand against the strings while picking in order to deaden their vibration, creating an increasingly percussive attack and dampened sustain. Palmutes can also be indicated with *P.M.* and a broken line above the notes to be executed with the technique. Various placements of the hand along the string's length, as well as varying pressure produce a seemingly infinite combination of possibilities.

While a staccato marking indicates a truncation of length, the difference between the two types of execution discussed have fundamental differences. The palmute provides less control as to the decay of the note, where the removal of the finger can be precisely timed. The result is that consistency in palmuting requires significantly more control on the part of the player. The evolution of spectral content and amplitude of the two techniques are also very different, as seen in the spectral flux and roll-off.

150 8th notes at 120 beats per minute of both palmuting and removal of the fingering hand for dampening were played with various pitches. As in previous testing, onsets are detected and the notes are peak aligned, with the peak value set to one. The results are quite simple. Palmuting rapidly reduces high frequency content, although it is still present during the attack and first few cycles of the note. Muting using the fingering hand resembles the relationships displayed in regularly plucked notes with a sharp drop-off in the spectral roll-off as the finger is being removed.

4.6.2 Hammer-Ons VS. Pull-Offs

Hammer-ons are phrased together by an initial pickstroke, followed by using the fretting hand to percussively activate subsequent notes on higher frets. Pull-offs are executed by exciting the highest pitch, and using the fretting hand to activate subsequent pitches on lower frets. As can be seen in Figure 24a, energy is fairly well maintained as new notes (vertical bars), are executed. Pull-offs, an example given in Figure 24b, tend to cause energy to dissipate more rapidly, and there is less regularity in the initial steady state portion of the new notes.

The speed of execution greatly affects the ability to maintain the energy caused by excitation. In all cases the centroid decreases relative to f_0 rapidly, and never rises again after the initial picking stroke. Roll-off typically diminishes more quickly with pull-offs than hammer-ons.

4.6.3 Pick Pressure

The hardness and method of execution have an influence not only on the amplitude, but the spectral evolution of the signal. Extremely hard picking, almost counter-intuitively, produces a longer attack time to peak as the energy in the string takes time to settle into a steady state. In considering the way an electric guitar responds to picking pressure, sometimes it is best to use the volume control in order to manage certain aspects of dynamics.

5 Application : A-DAFx

"Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?"

- Edward Lorenz[99]

American mathematician and scientist Edward Lorenz is best known for his work in Chaos Theory, and the impact of this statement displays that while the flapping of the butterfly's wings is a minimal "turbulence" in the given system, in this case weather, it can lead to unforeseeable consequences in another seemingly unrelated place - a tornado. What is important is to understand that in the evolution of any procedure, musical or otherwise, the initial conditions or subtle variations that occur during the process can greatly affect the outcome of events taking place at other points in time and space [100]. In the context of A-DAFx, variety of control assignment allows for an equally wide variance in the transmission of musical ideas. The richness of possibilities is astounding, however they are also the core of the problem - how does one sort out all of this available information into a usable form?

While the concept of A-DAFX is a more recent phenomenon beginning to appear heavily in literature after the year 2000, adaptive effects have been in existence for decades [2, 4, 6, 3]. Most early analog synthesizers allowed interconnection between various voltage controllers. Effects units such as the Digitech 2112 and 2120 released in the 1990's included assignable Low Frequency Oscillators and Dynamic Followers for automating controls.

Early implementations often provided one type of control per variable. This is a simple, intuitive and direct method of control, however it lacks much complexity. A map by its nature, implies a complex set of directions, and in this discussion can mean assigning one control to multiple variables, or multiple controls to one variable. The amount of power provided by modern computers is what makes this possible.

The mapping strategy developed here is not concerned with creating inherent complexity for the sake of complexity. To reiterate the goal earlier stated, the research was intended to create a coherent strategy (legend) for mapping in A-DAFx, by identifying relationships across a chain of events. The following mapping examples represent a simplified, but final link in the chain, aiming to maintain consistency with the guiding principles outlined in the introduction.

5.1 Control Streams : Assignment and Manipulation

In order to alter the value of variables in A-DAFx, there needs to be a control, and because these controls generally flow in a continuous fashion, they are referred to as control streams. The first step in creating a control is to define where it comes from, namely what feature or features will be used as the source of the values. These values may not be usable inputs for the variable.

For instance, a control may require a value between .001 and 16 for use in setting the rate of a Low Frequency Oscillator (LFO). If f_0 of the incoming audio from a guitar was to be the control, the values could potentially range from approximately 83 to 1320 Hz depending on the number of frets and tuning of the instrument. Treating these incoming values as a function, the processes of normalization, scaling, offset, and warping can be used to bring those values into a usable range. Combinations of these mathematical operations, as well as recursive algorithms, can produce a wide variety of results for creating a useful control stream.

5.1.1 Functions

A function is a rule which can only possess one output for each input and in accordance assigns a unique output in the set Y to each "input" from the input set consisting of X [48]. Certain functions cannot handle certain inputs, for instance the square root of negative numbers in something such as $f(x) = \sqrt{x}$, versus $f(x) = \sqrt{x^2}$ which could handle numbers with both signs. Piece-wise functions can be established which handle various inputs depending on their range, and the function's definition. Mathematically speaking, any feature can be considered by a static function or one which evolves over time.

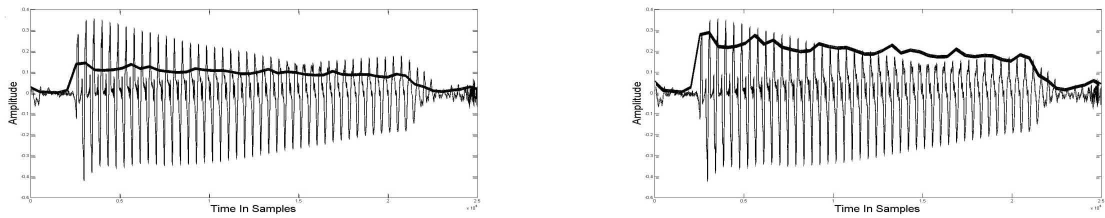
Normalization places a set of values into a specified and uniform range regardless of the range and nature of the input itself. This is most often done by placing all values between either 0 and 1, or between -1 and 1. This can be accomplished by using Equation 16. If the input is all positive values the results will be be values from 0 to 1, while inputs with signed values will range from -1 to 1.

$$norm(x[n] : x[n - M]) = x[M + N - 1] \div \max |x[M + N - 1]| \tag{16}$$

Scaling allows the user to alter the bounds of the values. Scaling operations are generally executed through multiplications and/or divisions. (See example Fig.25) Offset allows the range to be placed in a different region by setting a new maximum or minimum value. Offset is generally an addition or subtraction. Warping, a process often used in synthesis to shape waves, involves functions and alters data based on a function outside of the scale that the data originally came from. Common examples of warping are converting linear values to logarithmic scales, or vice versa.

5.1.2 Feedforward vs. Feedback

All effects can be categorized in three ways : feedforward, feedback, or a combination of the two. Feedforward, or finite systems, use input samples and then discard them permanently. Feedback, or systems with infinite responses, send their output back to some portion of their input or processing



(a) Envelope Following (Fig.4b) **(b)** Envelope Scaled 2X

Figure 25: Example of Useful Scaling

Figure 26: Conversion Cable and "Break-out" Box

chain. When a system calls the output of its own function, or when a function calls itself, this is known as recursion [101, 14]. The two can be combined to create varied results.

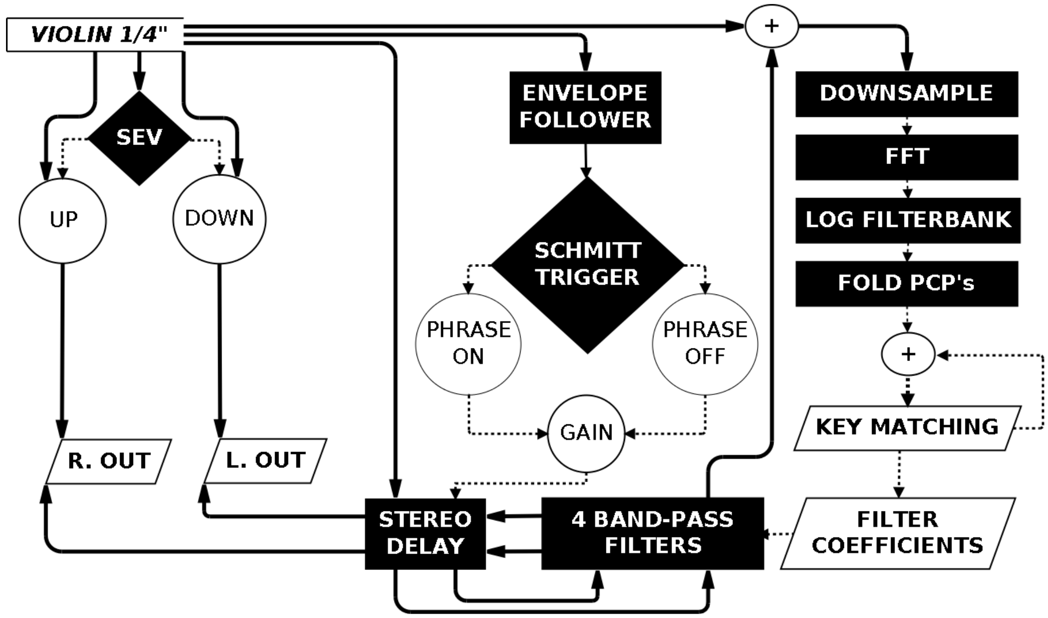
5.2 Monophonic vs. Polyphonic Analysis

For the purpose of feature assignment there are two levels of signal processing that occur. The first is considered monophonic and comes from the 1/4" output of the instruments in use; the term monophonic is used because all of the energy from each string is consolidated into one signal. 'Polyphonic' is used to indicate the isolation of each string as an independent signal. In many commercial devices which used divided pickups these modes are called 'mono' and 'poly'. In order to isolate the polyphonic signals, the author constructed a cable which converts the standard 13-pin polyphonic guitar cable to a common 9-pin connector, and a "break-out box" (Fig.26), which outputs each string on an individual 1/4" cable.

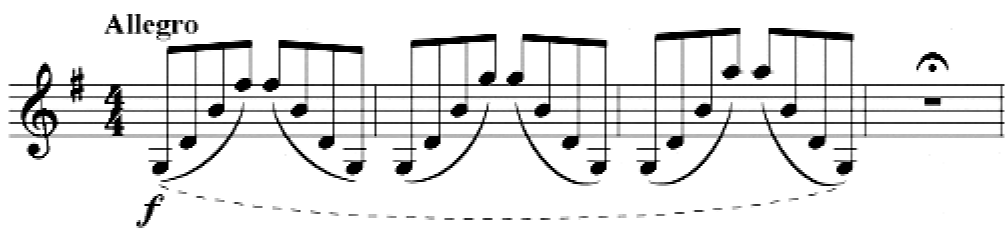
5.3 Effects Examples

The following examples were generated off-line in MATLAB.[15] In order to display as many of the features in a few short examples as possible, the "kitchen sink" approach is used, where everything

[15]While the code was written in a software designed to efficiently process matrices of data and pre-defined functions,the block structure and buffering commonly used in lower-level languages such as C/C++ was employed for the sake of future portability.



(a) Violin A-DAFx No.1 Flow Chart



(b) Applicable Musical Example

Figure 27: Violin A-DAFx No.1

but the kitchen sink is thrown in. While this may not be practical or musically desirable in some situations, for the purposes of illustration it is valuable. Each example seeks to exploit a unique combination of features, mapped to a unique set of A-DAFx, and in one case, generate a psuedo-synthesized version of the input. Additional concepts for use in other contexts are also discussed to display the flexibility and assure that techniques for the mappings displayed go beyond the uses in the developed effects, and satisfy needs in line with the initial guiding principles.

5.4 Violin A-DAFx No.1

Violin A-DAFx No.1 uses the $1/4''$ output of the electric violin and consists of a progression of chords which reside in a single key, with various periods of rest in between phrases. The sequence of arpeggios executed use the *arpeggiando* method of bowing, which requires the performer to slur notes across strings. Each set of arpeggios is up and down-bowed and the bowing direction dictates the position of the dry audio in the stereo field through the use of the Signed Envelope Follower described below. This directly correlates bowing direction to panning with a performer almost pointing to the direction of audio output.

A dotted phrase marker indicates connection between arpeggios and phrase detection is implemented by passing an envelope followed signal through a Schmitt Trigger to indicate on/off states. Chroma vectors are calculated for each frame of audio and the data is stored to maintain a running average, with recent frames weighted more heavily. All audio is fed into a stereo delay (which is commonly compared to the echo in a canyon) whose feedback gain is controlled by whether phrases are on or off. When no playing is occurring, gain is set to a high rate, keeping the feedback loop going infinitely. The feedback loop of the stereo delay contains a set of four band-pass filters whose center frequencies are set to emphasize octave relationships of the determined tonic. The default key is C, but will rapidly change with any incoming signal. The output of the band-pass filters also influence key selection as they are summed with the input prior to the Chroma process. If a key is detected and no further audio input is given, the feedback loop and filter combination would eventually produce four sine waves at octave intervals of the tonic pitch.

5.4.1 Bowing Direction Identification : Signed Envelope Following (SEV)

The detection of bowing direction relies upon a type of amplitude following, but one without precedent in the researched literature. What is proposed and implemented is named a Signed Envelope Follower (SEV). The SEV seeks to exploit the difference in mathematical sign between the peak of the waveform which occurs during the slipping phase, and the concentration of energy around the zero crossing and of the opposite sign during the sticking phase. An important note is that this technique is useless in a case where significant DC offset is present in the signal.

Instead of using a strategy such as that outlined in Section 3.2, the absolute value operation is eliminated. This leaves the signs of the samples in tact. To accentuate the slip versus sticking phases, a form of scaling is applied known as expansion. Taking the input values up to some exponent expands the absolute differences between the largest and smallest values, thus making the already strong slipping peaks stronger in relation to the sticking samples. Exponents of an even order cannot be used though, because the product of two negative numbers is positive; therefore an odd number is chosen and in this case the power of three was sufficient. In order to ascertain the bows direction, a comparator is implemented with 0 as the threshold. If A_{SEV} is greater than 0, the stroke is an up-bow, and less than 0 is a down.[16]

$$A_{SEV}\{x[n]\} = \frac{1}{N}(x[M]^3 + ... + x[M+N-1]^3) \tag{17}$$

Bowing direction can be easily mapped as a binary switch. Although not utilized in this example, the binary states can also be mapped so that they have uneven weight. If these values are accumulated and averaged over time the results can provide a very dynamic system for the performer to use. For instance, down strokes are valued at $+1$ while an up stroke is valued at -5, and a running average of these numbers is utilized as a control stream. This balances the output in certain situations, or unbalances them in other making the variety, or evolution over time greater.

[16]The phase of the output of the specific pickup system could reverse these results.

5.4.2 Phrasing/Slurring

Phrases are extremely important for the connection of shorter musical events, and are often used by composers and performers alike to indicate information about larger musical structure. Even when changing bowing direction, accomplished violinists are capable of make a seamless transition, giving the appearance of connection. Phrases are easily identified by the amplitude following strategy described in Section 3.2. If notes are truly connected, the envelope will stay above a given threshold. In this particular effect, the Schmitt Trigger is applied in order to ensure false starts or endings, such as accidental scraping or electrical hum/noise, do not falsely trigger the gain adjustment in the delay effect. This would prematurely cut-off the possibility of the infinite feedback loop described above.

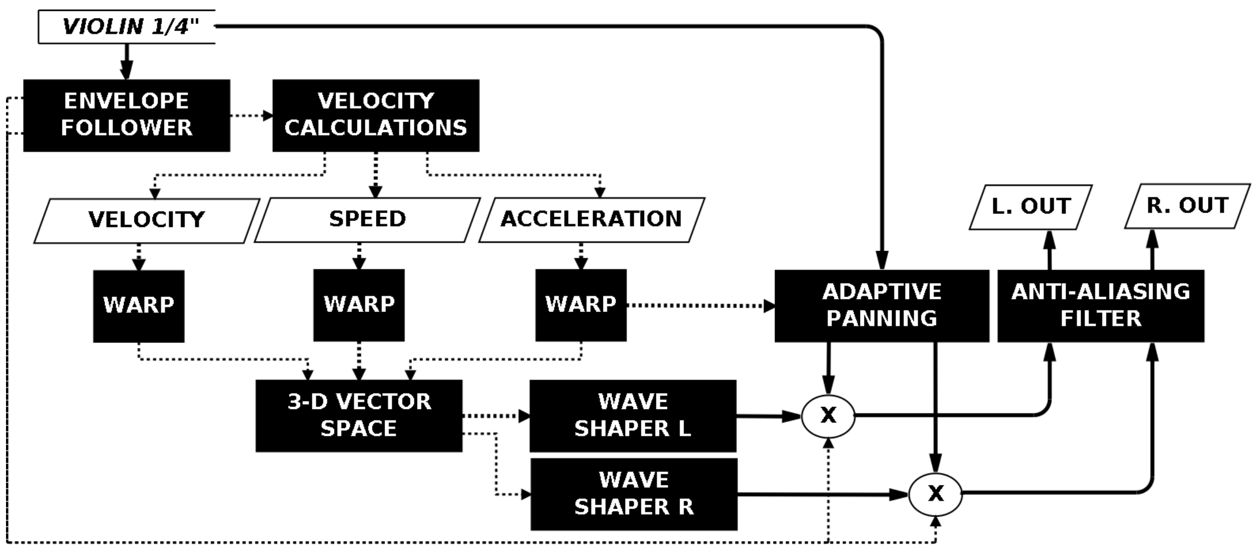
A signal above V_{ref1} indicates a phrase beginning, and the phrase has not concluded until the threshold drops below V_{ref2}. It is possible to combine onset, or pitch detection to determine if a phrase is really just an extended single note, but in this scenario was an unnecessary use of computing resources. A simple scheme for tying notes together as if they are part of the same phrase even when not connected is to first analyze the bowing direction utilizing SEV, and observe if the sign has changed or remains the same.

The length of phrases could be used as an effective control parameter as well. When V_{ref1} is crossed, a control stream could be set in motion on a given trajectory, or moving through a parameter space based on a combination with some secondary element. The longer the phrase, the greater the deviation of the control, providing a sonic indication about phrase length, and possibly how that phrase fits into a work structurally.

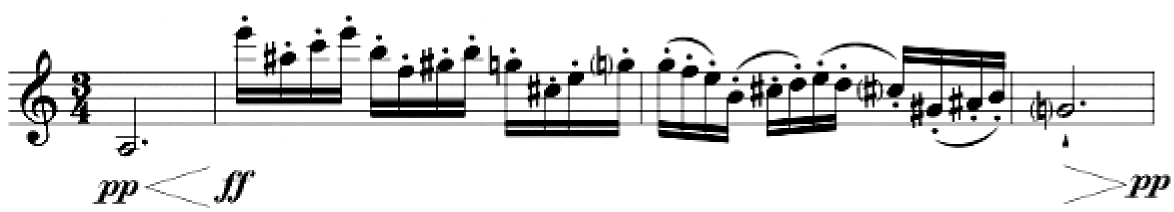
5.4.3 Other Uses of Chroma

Chroma has not been used extensively in real-time applications, and it is currently unrepresented as an object in the popular real-time libraries for feature extraction provided by Chuck, Max/MSP, Pure Data, or Aubio. This is unusual given that most of these software packages contain objects or code for the FFT/IFFT, Spectral Flux, Centroid, and Roll-Off. While the example provided here is relatively simple, the effect's design supports the harmonic content associated with the incoming signal. For tonal music, the sonic effect is subtle, but striking.

The template matching function can use any template, and obviously its assignment can be mapped in any reasonable fashion. An alternate thought is to create a delay with filters that tune to components in the chroma's running history rather than utilize any template matching. Patch changes could be assigned based on key change. For improvisational purposes, a scheme which tunes a bank of oscillators based on some number of the strongest PCP's could provide an interaction which steers the performer towards new harmonic centers. The opposite concept could be applied as well, where synthesized versions of the weakest PCP's or those the performer hasn't explored are generated, forcing them to either maintain the possible dissonances, or gravitate towards new harmonic content.



(a) Violin A-DAFx No.2 Flow Chart



(b) Applicable Musical Example

Figure 28: Violin A-DAFx No.2

5.5 Violin A-DAFx No.2

In this example, the input is altered using waveshaping. Waveshaping is a common method in synthesis and effects applications of taking simple signals and making them more complex. A common example is distortion or overdrive in guitar amplifiers, which squares off the waveform enhancing the upper harmonics of the signal.

The waveshaping is controlled in a fairly complex manner, with three variable dimensions controlled by calculations of velocity (Sec. 3.1.2). Panning is also controlled by acceleration, and instead of having the signal project from only one channel the signal can move seamlessly between any position in the stereo field. The use of space is a quality often exploited by composers in electronic music composition[102, 14]. An anti-aliasing filter is applied at the end of the signal chain to remove any ultra-high frequency content, and the envelope follower serves as a side-chained gate for the spurious noise caused by waveshaping in between note events.

5.5.1 Velocity Calculations: Crescendo, Decrescendo, Articulation Shapes

What is interesting about this effect, is the lack of absolutes in feature measurement - everything is relative. This may be less controllable to a performer in the beginning, but once mastered could provide a meaningful and lively animation of the incoming signal. The interplay between resulting sound and performance choices is also an exciting prospect. All of the adaptation is based upon the three calculations for velocity displayed in Equations 4, 5, and 6; these are velocity, speed, and acceleration respectively. The measures are taken from the output of the envelope follower previously described.

A number of steps are taken to process the velocity calculations to ensure their usefulness. The first is that rather than taking the first order differential, which would be a logical step, a buffer of previous envelope follower values is stored so that a greater order difference can be calculated. This aids in smoothing the resulting signals and giving precedence to the change instated by large values from the envelope follower. Because the current envelope value is being compared against a historic one, there is no latency in relation to the performer's current input.

The velocity measurements are then warped in order to give more weight to lower values. With values ranging from zero to one, and negative one to one, the cube root of the current value is calculated, which results in a non-linear input to output. Speed is easily warped, as all of the values are positive, however velocity and acceleration require piece-wise functions which perform the calculation based on the incoming values sign. The final values are low-pass filtered for additional smoothing, where they might cause otherwise erratic behavior in the resulting processing. Scaling can also be applied to the values depending upon the desired effect.

The results of the velocity measures are telling about the types of articulations and techniques in use. Therefore the the composer/performer can exploit the various techniques examined in Section 5 to create consistent, or inconsistent, variations in the resulting waveshaping. The velocity measures can provide values which link directly to the performers interpretation and execution of musical events, in particular bowing and pressure accent techniques.

An added feature of using difference measures, rather than static envelopes is their reaction to crescendos and decrescendos. Depending on the shape of either, the acceleration will remain constant providing a steady set of control values to the assigned variable, while speed and velocity will also remain fairly constant. What is provided is a constant timbre or waveshaping, against a changing amplitude.

5.5.2 Wave-Shaping Functions

Waveshaping is a relatively simple process with ideal signals, like the output from a function generator whose wavefroms retain their ideal shape regardless of frequency. Real-world signals are not generally so forgiving in creating a consistent waveshaping effect. For this reason a number of methods were experimented with. The final waveshaping algorithm is based on the visualization of a three-dimensional parameter space. The simple explanation of the assignment is that velocity controls the balance between even and odd ordered waveshaping functions (x). Acceleration controls the gain of the incoming signal with the possibility of phase reversal in the case of deceleration (y). Last, speed controls the overall gain scalars of the respective polynomials performing the waveshaping (z).

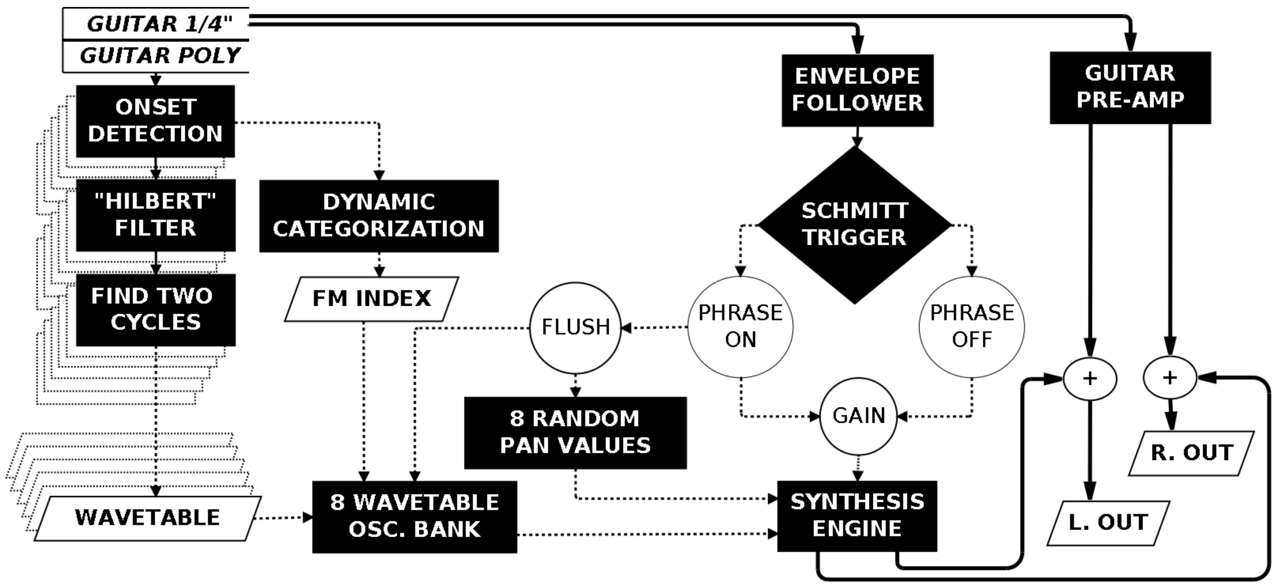
The result is a rich variety of timbres, which are directly dependent upon the performers interpretations of the expressive markings. Alternatively, the envelope follower could be used in a more traditional manner, directly mapping amplitude to one or more of these parameters, providing a shifting position in the parameter space with more connection to dynamics than bowing techniques.

5.6 Guitar A-DAFx No.1 : "Sustain Pedal with a Twist"

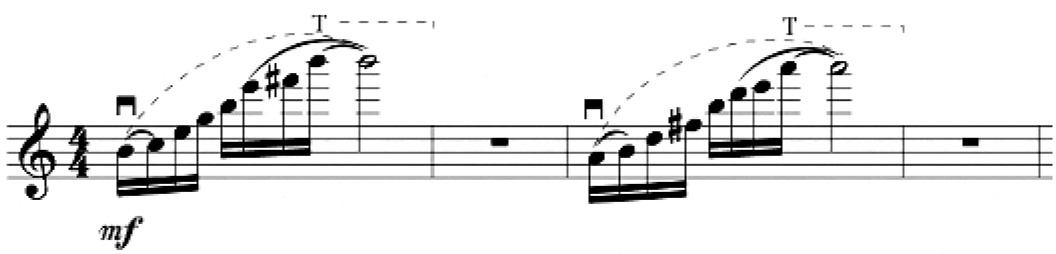
This effect utilizes both 1/4" and polyphonic pickup outputs. The actual sound or tone associated with the guitar generally comes in combination with an amplifier, and in this case the tone from the magnetic pickups is enhanced using a tube pre-amplifier. The sound is then blended with an effect akin to the sustain pedal of a piano. All notes connected under a phrase are 'held' until the next phrase begins. This is accomplished by detecting one cycle of the waveform of an onset detected note event, and storing the values into a table for cyclic playback using wavetable synthesis. The 'twist' comes in that dynamic categorization is used to determine a modulation index for the table using John Chowning's classic FM synthesis technique [103].

5.6.1 Dynamic Categorization

It has been reiterated multiple times throughout this document that dynamics are relative, and amplitude is an empirical measure. With that thought in mind it would be slightly absurd to try to create a control which directly correlates exact dynamic markings to amplitude measures. If amplitude input is considered in a terraced fashion, then amplitude can be followed directly to classify dynamic regions. Considering the results of dynamic testing, along with integration of a volume controls, eight regions seems reasonable, along with one 'dead' region for no activity. The process is relatively simple : when onset detection identifies the beginning of a new note event, the maximum absolute amplitude value from that block of samples is used for the comparison. The FM



(a) Flow Chart



(b) Applicable Musical Example

Figure 29: Guitar A-DAFx No.1

Index value is selected based on the categorization and stored along with the wavetable information to be described below.

Another interesting concept would be to map the dynamic categorization to a table which applies gain scalars to the input. The state with the highest value would remain at unity gain, while lower dynamics' gains could be scaled downward with the softest notes being attenuated the most. This would create a set of "hyper-dynamics", thus expanding the usable dynamic range of the instrument. This is different than a traditional expander/compressor in that it operates in a stepped fashion, rather then with a fixed curve for input to output scaling. The control could also remain un-quantized as well, using the raw maximum value for that block of samples as a direct control stream.

5.6.2 Wavetable Synthesis

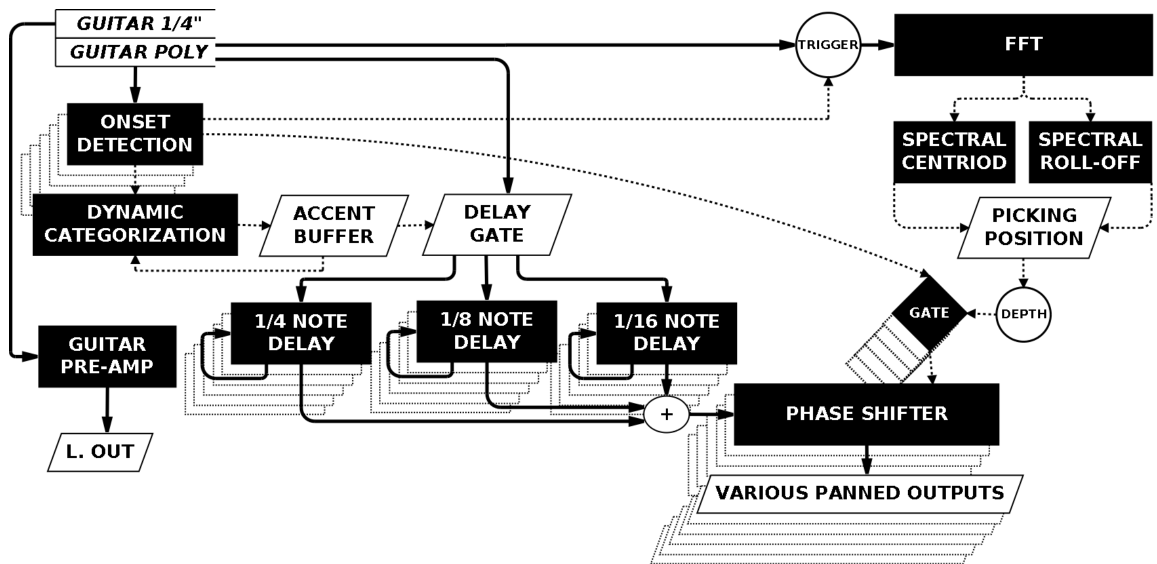
Puckette's strategy for the use of Hilbert drives classic Music N [102] style wavetables, giving synthesized outputs without the latency of pitch analysis windows. The approach here is slightly different. When an onset is detected Hilbert is 'activated' and the first complete cycle of the appropriately oversampled waveform is captured and stored in an array of wavetables. The wavetables are then accessed so that the actual guitar's waveform is cycled, and all pitches in a connected phrase are stored, with a maximum polyphony of eight voices. Information about the original sample rate is also stored for proper table read-out.

If this count is exceeded tables are replaced with more recent wavetables. The notes are sustained until a new phrase is detected causing all information in the wavetable to be flushed out. Complexity is added to the tones using FM synthesis where a carrier frequency is calculated based on the number of samples in the wavetable, and the index determined by the process discussed above. The individual synthesized outputs are randomly panned throughout the stereo field.

One of David Friend's original arguments against systems that preceded the ARP Avatar (Sec. 3.3), was their inability to drive an oscillator or sustain a note beyond the length of the guitar's actual decay [53]. Using this method of real-time wavetable creation, allows for a type of sample and hold of the waveform, and provides the possibility of the type of sustain offered by synthesizers.

5.7 Guitar A-DAFX No.2

The final example of A-DAFx uses accents and picking position as controls. Dynamic categorization is stored into a buffer and the relationship to previous values determines whether a note is an accent or not. Based on emphasis or de-emphasis of the note, a gate routes the note through various delays based on subdivisions of a pre-determined tempo. Picking position is approximated and controls the depth of a phase shifter. The sequence of events happens independently on every string, and while a dry signal from the magnetic pickups is panned to the far left channel, the polyphonic pickup output is spread across the stereo field.



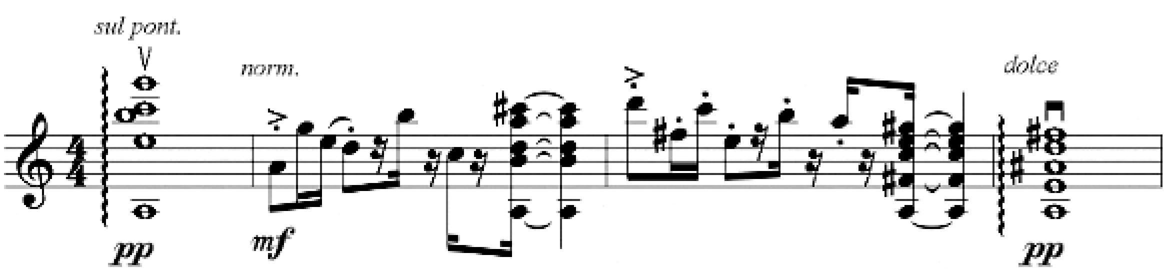
(a) Guitar A-DAFX No.2 Flow Chart

(b) Applicable Musical Example

Figure 30: Guitar A-DAFx No.2

5.7.1 Accents

Because accents, like dynamics, imply some relationship to another value, they present problems for classification in real time. The dynamic categorization used in the previous examples, while effective, disregards the nature of the function of certain types of accents - metric induction being a major one. Accents can provide extremely useful information for the listener about metric structure, and this portion of the effect seeks to exploit the use of accent to provide the listener with precisely that.

This example accumulates dynamic categorizations, and watches for differences between the running average. A note event that is stronger than the average values is considered an accent, and therefore emphasis. The gate feeds the audio input to the delay whose value is equal to the 1/4 note, creating a pulse. Average dynamic range notes are set to the eight note delay, whose gain is set lower than the 1/4 note delay, with slightly higher feedback. These notes provide 'harmonic support', and a slight connection to the first subdivision of the beat, further reinforcing the meter. Notes that are de-emphasized are run to the 1/16th note delay which has a much lower output gain, but a much higher feedback level, causing the de-accented notes to become more 'atmosheric', almost reverberant, components of the final sound.

5.7.2 Picking Position (Approximated)

It would be extremely computationally expensive, or rely on extremely long windows to determine the exact modes present in a waveform, but examining the data from picking position testing (Sec. 4.3.2), there is a clear indicator as to the the type of upper partials present in the signal in relation to the fundamental. The relationship of the ratio of the centroid to the roll-off are used as the determination of picking position.

With plucked guitar, the first harmonic tends to be the concentration of energy, with the second harmonic being initially stronger in very rare cases. The centroid tends to hover around or between between these harmonics, depending upon the resolution of the FFT. The roll-off tends to be much higher at the onset of a note and the higher the modes activated the greater the distance between the centroid and roll-off.

When an onset is detected a trigger is set off and a buffer begins to fill up. By using the trigger,the FFT for centroid and roll-off calculations are only used when necessary, saving computational resources for other tasks. Once N samples are collected, the FFT is calculated and the ratio of the roll of to centroid is determined to create the new value for the phase shifter's depth. The gate is then closed until a new onset from the respective string is detected.

6 Improvements in Testing, Future Applications and Conclusion

6.1 Improvements in Testing

Although the use of a limited number of performers per instrument is legitimized by the fact that it is an 'expert opinion' at work, it may be useful to incorporate more players into a blanket study and see how the results differ from performer to performer on the same instrument. It also may be useful to analyze differences between specific instruments and pickup systems. As mentioned in the description of the testing, Scott Tixier does not typically play the instrument used in the study; working with violinists who have more experience with electric specific instruments would increase the validity of measurements.

Although analysis was made easier and more consistent by the use of directly outputting the signal from an electric instrument, there was a backlash. The signals coming from pickups in electric instruments can be effected by stray electrical signals, in particular lights. Unfortunately for the violin testing, although it was recorded in a professionally constructed iso-booth, the lighting inside induced a hum at certain points throughout testing. A direct box was introduced in the signal chain in order to try to cancel this hum, but because it was not directly induced by a power supply or the system itself, the counter effect proved to be negligible. In some of the spectral analysis this hum appears as relatively small, but additional peaks at multiples of roughly 60 Hz, the frequency of an alternating current waveform. It also made measuring certain low amplitude signals difficult as the amplitude of the hum itself was as strong or stronger than the input at a few rare, but noticeable points.

6.2 Future Applications

6.2.1 Larger Musical Structures, Data Accumulation, and Interaction

What was presented throughout this document are ties to simplified local events. Overarching musical structure is extremely important in connecting shorter musical events. Strategies should be developed in which effects are not only tied to local events, but also to larger structural elements of musical forms. This could be realized through patch changes between various A-DAFx for different structural elements of a work, or accumulating data to shape the control of variables over the course of a musical work.

One of the most exciting prospects of A-DAFx in general, is the concept of storing previous control stream information for later use. Essentially, A-DAFx can be built so that the adaptations themselves adapt; this can provide an added layer of complexity and interest for both audiences and performers alike. It also gives the electronic system a sense of personality like that seen in human performers. While data from the same performance can provide impetus for altering interpretations of music during that performance alone, data can be accumulated, modeled, and developed over many performances. Systems using machine learning techniques offer the greatest possibility for success in this type of application.

6.2.2 Additional Techniques : Trills, Tremolos, Vibrato, Pizzicatto, and Harmonics

There are a number of techniques that result in micro-variations in pitch, or changing pitch at an interval less than a half step (two adjacent keys on a piano) which were not discussed. Trills are the rapid succession of two notes, and Tremolos are trills at an interval greater than the musical interval of a second. While these are two distinct pitches, some performers have the ability to execute them so rapidly that they appear to the ear as some intertwined note event. Vibrato is applied to a sustained pitch and depending upon the width, changes the character of the pitch.

Bowed string players almost always use some form of vibrato as the default. Composers indicate lack of vibrato by writing '*sensa vib.*' or '*sensa vibrato*' above the score when no vibrato is desired.

All of these techniques are idiomatic to both the guitar and violin and the author is well aware of their uses. Data was collected for trills and vibrato but was not fully analyzed due to time constraints. Data was also collected for a number of variations of the Pizzicato technique including snaps and the "Bartok pizz."

Harmonics, which are a method of 'activating' nodes along the string by lightly touching them, produce very different timbres than standard fingered pitches. This is a common practice seen in many aspects of string literature. Although the usefulness of identifying harmonics versus standard fingered pitches may be of significant value in a performance system, the requirements for doing so represent a unique challenge that could be the subject of a large study in itself.

6.2.3 Use with Other Instruments

The methodology applied in this study, could be used in the case of other instruments. The obvious extension would be to apply the analysis techniques to other stringed instruments in the bowed family such as the cello or double bass, and plucked strings such as the banjo or mandolin. The isolation of the signal provided by electronic pickup systems could also be applied to instruments using modified or commercially available hardware. For instance, a variety of mutes exist for brass and wind instruments which dampen their acoustic output at the bell, and capture sound electrically through a transducer. While these mutes alter the players ability to execute certain musical gestures, and tend to limit range, they provide the type of isolated input discussed in Section 1.2.

6.2.4 Unexplored Existing Work

In the provided A-DAFx examples, those reliant on tempo were given a pre-determined tempo, and the subsequent realizations made were accomplished by the performer recording to a click track to ensure accuracy. Automatic beat or metric identification such as outlined by [104, 105], could be appropriated for adaptive time based effects. A popular example of this is the manual entering of a "tap tempo" seen in many commercial delay units, so that regeneration of the same events through feedback are linked to a specific time interval, or pulse. The studies mentioned provide methods in order to automate this process. Accent information is generally a strong indicator of meter, and the time intervals between identified accents could be linked to beat induction as well.

Another interesting study that was found too late in the process of this work, was a time domain based method for detection of guitar plucking position [79]. The algorithm appears to be very accurate and its results will be tested in future work. The strategy in [79] appears to be much more robust, and faster than the spectral methods utilized in this document, which in reality offer relatively limited resolution for the determination of plucking position.

6.3 "Packaging" and Evaluation

Due to time constraints and the amount of work and emphasis placed on careful examination of markings and development of the strategy in this document, extensive evaluation was not possible. The author's aim for future use is to package some of the elements in two types of VST (Virtual Studio Technology) Plug-Ins, for use with widely available Digital Audio Workstations (DAWs). DAWs are commonly used software packages utilized by both enthusiasts and professionals, and in both commercial and more arts based applications. Almost all DAWs support VST technology, and so do some graphical programming environments, with MAX/MSP being the prime example.

The author will be constructing two types of VSTs based on this research in the near future. The first will simply use feature extraction techniques to create MIDI data for controlling existing effects packages within the aforementioned DAWs. The second type will be combination packages, encapsulating both effects and feature extraction, resembling the types of A-DAFx outlined in the previous section.

By packaging these materials in VST format, the author can then distribute them via electronic mail and make them available for download through his personal website. By putting the work into the hands of composers and performers from various backgrounds, feedback can be gathered about the merits (or lack thereof) of the work's usefulness. By running the VSTs on various machines, at different sampling rates, and with a variety of instruments, the true nature of the pitfalls and drawbacks to the algorithms and strategy can be brought to light from a broader perspective.

6.4 Conclusion

What has been discussed throughout this document outlines a strategy which aims to link low-level features with the activities associated with instrumental composition and performance in the context of A-DAFx. The process of gathering musical materials and examining them carefully from a pure DSP standpoint lead to a few contributions. A handful of assumptions from existing signal processing literature were challenged both scientifically, and aesthetically, and the results in some cases provide information that bring the validity of the original claims into question. The Signed Envelope Follower, although extremely simple, is a unique application of this DSP, which resulted directly from this careful examination.

The overall process of examination provides a good base from which to delve more deeply into the links between the types of expressive markings and control of A-DAFx from both a technical and musical standpoint. The suggestions for creative applications in these A-DAFx, may provide impetus for subsequent exploration into these connections as well. Whether all techniques outlined are musically meaningful or successful will be determined through the mentioned evaluation, but the concept of developing a legend for the map, at the very least, was clearly demonstrated in the context of the electric guitar and violin. Most important of all, the overall work provides a cohesive strategy, based on examination of musical materials rather than purely sonic or perceptual frameworks, from which composers and performers can consider how to develop their own usage of A-DAFx.

A Introduction to Sound and Music

A.1 Representations of Sound and Music

"Imagine a player in San Francisco... You might, at home one evening in 1994, listen to music that was recorded in San Francisco in 1952. You might then wonder: how did the sound travel so far through space and time?"

- Joel Chadabe [106]

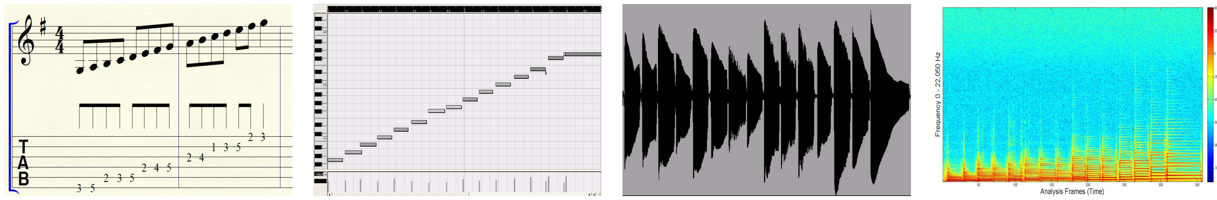
Music and/or audio information, can be represented in a number of ways and Chadabe's question serves as an excellent starting point to illustrate this concept. A composer may have written some musical ideas in the form of notation on a piece of paper; because these ideas are visual abstract representations, they are called symbolic. Utilizing either an instrument or voice, a performer interprets these markings and begins to make sounds. Air molecules are displaced and radiate in order to form gradations in air pressure. In this case, the sound energy from displacement vibrates the element of a microphone (transducer) which coverts the gradations into varying voltages. The varying voltage is then stored, which in 1952 would have been onto some analog medium, most likely a magnetic tape.

Eventually, the tape would be converted from the continuous analog medium, to a digital representation of discrete, sampled numeric values, to be stored on a compact disc. In modern times the digital audio could be stored to any number of compressed audio formats for distribution via the internet or portable digital media. At the listener's home, the digital representation is converted back into a continuous analog s signal where it is sent to an amplifier and drives some form of either headphone or loudspeaker. The sound has terminated at the ear drums of the listener.[17]

What has just been demonstrated is that musical information can be stored or encoded in various formats with different contextual meanings. A printed score, air pressure, magnetically arranged particles of iron oxide on a plastic tape, and the binary numbers in a digital audio system all have different meanings in terms of their formatting and use. What is similar is that all are useful methods of encoding musical events. When this information is transferred and interpreted by a listener, what is heard at a high-level is considered musical information. A number of different representations are specifically discussed throughout this paper including:

- Mathematical (Formulas, Functions, Definitions)
- Digital Signals (MIDI files, Digitized Audio Formats)
- Time Domain (Waveforms)
- Frequency Domain (Spectral Content)
- Symbolic (Musical scores, Lead sheets, Tablatures, Piano rolls)

[17]The alternate to re-conversion of an analog voltage is a class D amplifier which outputs a rapid series of digitally generated pulses.



(a) Notation + 'Tab' **(b)** MIDI Piano Roll **(c)** Digital Waveform **(d)** Spectrogram

Figure 31: Two Octaves of a G Major Scale : Various Representations

A.2 Musical and Sound Terminology

> *"Music has four essential elements: rhythm, melody, harmony and tone color...It is their combined effect-the seemingly inextricable web of sound that they form - with which listeners are concerned for the most part."*
>
> - Aaron Copland [107]

Rhythm is a general term used to refer to the temporal aspect of music, where beat is the basic pulse, and meter is the pattern of beats which can be subdivided into smaller values [108]. Melody and Harmony deal with pitch, which is the perceptual measure of how a listener is able to recognize the difference between various frequencies - frequency being an empirical measurement dealing with the number of times an event occurs within a given time space. In musical and audio applications Hertz(Hz), or the number of cycles or repetitions per second, is the standard unit for describing frequency. The question asked when distinguishing relative frequencies or pitches is *"which one was higher or lower?"*[84] **Melody** is the succession or changes in pitch through time, while **harmony** deals with the simultaneous sounding of multiple pitches. Pitch is a subjective measure, often times affected by factors including loudness and timbre.

The last element is **tone color**, often referred to as **timbre**, and is the least well understood of all the elements of music, but becomes central to many aspects of this study. Timbre in many cases, particularly through the 1960's, was negatively defined, meaning by what it is *not* rather than what it *is* [109, 29, 102]. Timbre is "*not pitch, not loudness, and not duration.*"[29] Timbre allows a listener to distinguish tones from different sources even when the pitch and loudness are identical [91]. Another explanation of timbre is directly implied in the term 'tone color'. In visual perception, color is a mixture of various bands or wavelengths of light, and timbre can be thought of similarly [29]. The building blocks of timbre, called harmonics, overtones, or partials are analogous to bands of light and contribute to form the overall makeup of a sound.[18] These individual sound components form what is referred to as the "spectrum" of a sound.

Human perception of sound, and subsequently musical events, are understood by their changes over time in frequency and amplitudes [109]. As such, all properties of sound can be explicitly described by their evolution in three dimensions which are time, spectrum, and intensity. Time can be thought of as the channel through which the other two dimensions flow and change. The

[18]It should be noted that the eye only detects about one octave of change in visible wavelength, while many humans can perceive up to ten octaves of range in sound.

spectral dimension deals the component frequencies which combine to form more complex sounds, while intensity deals with the amplitudes of either specific, or all frequencies.

Amplitude and magnitude are absolute measures of intensity, though psychoacoustic studies have shown that the brain interprets absolute intensity on a relative basis in time, and in relation to the intensity of concurrent sound components [30]. At a high-level, intensity can refer to musical dynamics, which can be the loudness of individual notes, up to the loudness of an entire musical passage. Intensity, even of simultaneous events, can be reduced to one overall value which evolves in time called the envelope. In simplest terms, an envelope contains or scales the amplitude of a sound. The way in which the amplitude of individual components of a sound evolve over time is generally referred to as the **spectral envelope**.

A.3 The Sine Wave

A sinusoid is the simplest form of motion or oscillation and is the cornerstone of many concepts used throughout this document. A sinusoid, also referred to as a sine or sine wave, can be defined by Eq.18 [41].[19]

$$x[n] = a\cos(\omega n + \phi) \quad OR \quad x[n] = a\sin(\omega n + \phi) \tag{18}$$

Typically the result of this equation would be represented by a continuous function $f(x)$, but because a sampled version is being discussed, n represents a discrete point in time. The amplitude of the waveform, is represented by a scalar a and determines how far the peaks and valleys of the wave fall beyond 0.

While the frequency is commonly represented by the unit Hz, signal processing and physics typically represent harmonic motion by angular frequency or angular velocity, ω, expressed in radians. Conversion from Hz to ω is a matter of multiplying by 2π, and multiplying ω by $\frac{\pi}{180}$ to achieve the inverse. The cycle of the waveform can be thought of circularly, where as one traces the outline of the circle, an angle to the current position can be measured. Therefore one complete cycle of a waveform has positioning known to be between 0 and 360°, or 0 to 2π radians. The current position within a periodic cycle is known as the phase angle, argument, or simply phase. When phase is bound between 0 and 2π, it is wrapped. It is possible to measure unwrapped phase, which can provide clues as to absolute, and not just angular changes over time.

The sine wave does not always have to begin at zero degrees/radians, and ϕ represents the initial phase, or phase offset value. Cosine waves are often used in lieu or conjunction with sines in order to represent the motion of soundwaves. Although sine and cosine waves have identical shapes, they are separated by a phase offset of 90° or 1/4 of a cycle as shown in Fig.32. This constant separation of 90° is known as phase quadrature, or an orthogonal relationship.

A.4 Decibels (dB)

Relative measures of amplitude and loudness are generally displayed in terms of decibels(dB) [41]. Decibels are a useful representation in applications where an extreme or wide range of values

[19] Although sine and cosine are displayed as equal, and in many ways similar, it should be noted that their mathematical properties are not in fact identical.

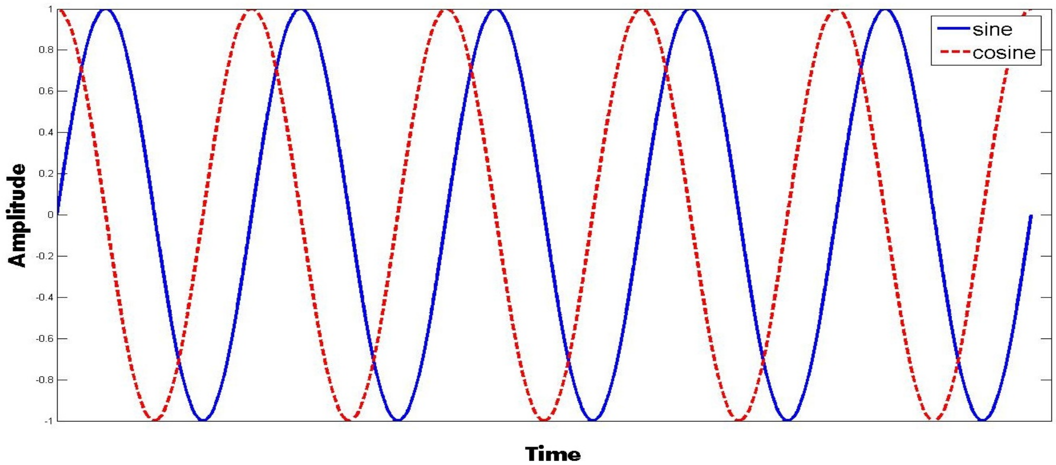


Figure 32: Full Amplitude Sin and Cos Waves

need to be viewed side by side. As an example the difference between the loudest and quietest sound detectable by the human ear approaches a 10,000,000 fold difference. By using a logarithmic representation of a ratio, dB becomes a useful tool in comparing these widely varied numeric sets. As a note, an increase in amplitude by a factor of two corresponds to an increase of 6.02 decibels, doubling power is an increase of 3.01 dB.

$$dB = 20 \cdot \log_{10}(a/a_0), \quad a_0 = 10^{-5} = 0.00001, \quad maxa = 1 \tag{19}$$

B Time Domain

Real-world signals are analog, meaning continuous. Digital signals are discrete representations, and are intervallic rather than continous. The time domain represents the changing values for a series of numbers over time. The process of converting analog to digital signals, and vice versa, occrs in the time domain via the sampling process. Besides a string of numbers, there are other ways to represent the time domain. Many digital audio software packages and workstations incorporate graphical representations of these amplitude changes such as the waveform displayed in Fig.31c. When looking at many samples simultaneously, time domain waveforms can provide indications about the nature of a sound(Fig.31c); zooming in, they can display the value of individual samples and the cyclic waveforms that comprise pitched sounds (Fig.33a).

B.1 Digital Signals and Sampling

Digital signals and the subsequent processes which are used to manipulate them are based on the concept of sampling. Whether it be measuring the changes in temperature over the course of a day, or the changes in amplitude of an electrical signal over the course of a second, the sampling process allows the measurement, storage, comparison, and manipulation of discrete values. The sampling of audio involves converting a continuous analog voltage into binary numbers at a fixed time interval known as the sampling period. The number of times per second that the period occurs is known as the sampling rate(f_s). The value of a sample is the amplitude of the signal at the time of measure. Individual samples are generally referred to as $x[n]$ for the current sample being fed into a system, and $y[n]$ for the current output.

Modulation is the process wherein one electrical signal acts upon another, and sampling is an example of this. Simple modulation processes involve two signals referred to as a carrier and a modulator. In the case of sampling the carrier ideally is a series of pulses that are infintismally short, or "delta kronecker" pulses, and happen regularly at the defined (f_s). The modulator is an incoming analog voltage from a device such as a pickup or microphone. As a result, the modulator changes the amplitude of the carrier, in an Amplitude Modulation(AM), specifically a Pulse-Amplitude Modulation(PAM). The changes in the value are measured and stored as binary numbers.

This conversion of an analog voltage to a binary number is known as the Analog-to-Digital Conversion (ADC) process. The measurements are quantized to a value dictated by the number of bits the system is designed to handle. When stored, these values can later be accessed for manipulation and use; the storage period is based on the application. In many DSP applications minimal input to output time is required. The measure of the amount of time it takes go from an input $x[n]$, to output $y[n]$, is known as **latency**. The reduction of latency increases temporal precision, which allows greater control by the user[14].

The regularity in timing, or clocking, of the sampling period is important because it also gives a reference point from which to convert a digital signal back to the analog domain. This process is known as Digital-to-Analog Conversion(DAC). Theoretically, during DAC, the signal sampled during ADC can be completely reconstructed, provided f_s is at least two times the highest frequency component contained within the signal; the cut-off frequency a system is capable of sampling is known as the Nyquist Frequency.

In order to ensure an incoming signal meets the criteria of falling below the Nyquist frequency, a low-pass filter is used which removes high-frequency content from the signal. The cut-off in theory of this filter would be 1/2 of f_s, and frequencies above that allowed into the system cause what is called aliasing, or foldover. The frequencies above Nyquist fold back into the sampled audio and cause false information in the data. For a complete discussion of both conversion processes, refer to [14, 56, 110].

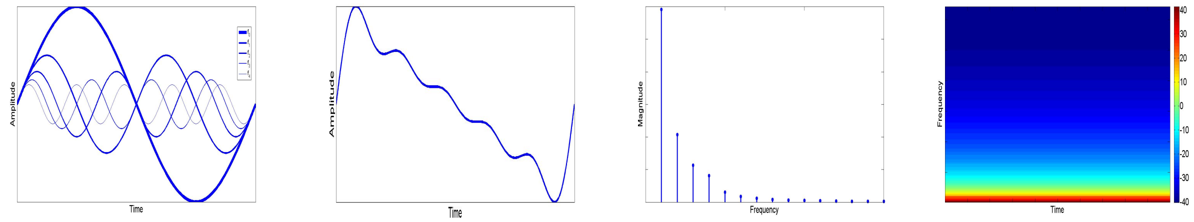
Most digital audio applications employ numeric systems which place the value of a sample between -1 and 1. Two important measures of these amplitudes are peak and average. Peak amplitude is the largest absolute value taken over the course of a given block of samples, expressed as:

$$A_{peak}\{x[n]\} = max|x[n]| \quad P\{x[n]\} = \frac{1}{N}(|x[M]|^2 + ... + |x[M+N-1]|^2) \tag{20}$$

while the root mean squared (RMS) is the average value over a block of samples expressed by:

$$A_{RMS}\{x[n]\} = \sqrt{P\{x[n]\}}, \quad P\{x[n]\} = \frac{1}{N}(|x[M]|^2 + ... + |x[M+N-1]|^2) \tag{21}$$

These blocks of samples can also be referred to as a window.



(a) First Five Partials **(b)** Sum of Sines **(c)** Stem Plot **(d)** Magnitude Spectrum

Figure 33: Sawtooth : Sum Of Sines

C The Frequency Domain

The frequency domain consists of a set of amplitude (and often phase) values for specific frequencies within a given time space. Spectrograms are a common form of visually representing the intensity of frequencies and their changes over time. The spectrogram in Figure 33d shows frequencies on the y-axis, time on the x-axis, and intensity by color, with corresponding values indicated by the color bar to the right. 3-dimensional plots which display intensity by height are also popular representations of changes in frequency content over time.

Being able to look at signals in the frequency domain is important because most signals, and particularly those with any aural interest, are not sine waves. They are a combination of sine waves of varying frequencies, amplitudes and initial phases. The lowest frequency of oscillation is referred to as the fundamental, designated as f_0. The sine waves related to f_0 are referred to by different terms including: harmonics which are integer related multiples of f_0, overtones which are frequencies above the fundamental, and partials which are the same type of series but include f_0.

Acoustic instruments produce either sounds that have harmonic partials whose frequency relationships are based on a set of integer multiples, or inharmonically related which implies almost any other type of relationship. While most vibrating bodies do not produce perfect integer relationships, they are still close enough to be heard as and considered harmonic. As an example, take the first five partials of a sawtooth shaped waveform (f_0 plus the first four harmonics/overtones).[20] The summation formula, represents the addition (Σ), of sinusoidal partials n as follows:

$$sawtooth = \sum_{n=1}^{n=5} \frac{1}{n} \sin(n \bullet f_0) \tag{22}$$

Figure 33a displays the five partials individually, while Figure 33b displays the resulting waveshape when the sines are summed as in Eq.22. While the first two representations are time domain waveforms, Figures 33c and 33d, display information in the frequency domain. Figure 33c shows the intensity, or magnitudes, of the individual partials with a linear relationship. Figure 33d is a spectrogram indicating the intensity on a logarithmic scale and location of frequencies within the spectrum.

[20]The sawtooth becomes a crucial waveform in the discussion of violin bowing (Section 2.3.2)

C.1 DFT/STFT/FFT

In his 1822 treatise *Théorie analytique de la Chaleur* (The Analytic Theory of Heat), Baron de Fourier provided the ground work for the proof that any periodic function could be represented as an expansion of its individual harmonic components [60]. The basic assumptions of Fourier's theorem are that the signal being decomposed is periodic and infinite, and the resulting expansion from this type of decomposition is known as a **Fourier's series**. The summation, which bears similarity to previously seen formulas, is mathematically expressed as:[30]

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t) \tag{23}$$

The sampled signals of digital audio are neither infinite nor continuous, and in many cases not perfectly periodic. Rather than use a Fourier's series, the Discrete Fourier Transform(DFT) must be applied. A type of DFT, The Short-Time Fourier Transform (STFT), allows slices of time (smaller than ∞), known as windows, to be taken for analysis. The choice of window involves three factors : length, type, and overlap. Windowing functions amplitude scale the time domain set according to the window's shape and prevent erroneous content caused by discontinuities at the edges of the STFT frame. Although windows can help in reducing extraneous frequency content, there is still some cost in the generation of another type of added frequency information, known as **spectral leakage**, or **side lobes**.

The resolution of the DFT is based on f_s, and the size of the window used in the STFT, which is generally designated as N. The result of the mathematical operation is a set of frequency bins, k, containing complex numbers in the form $x+jy$ [21] from which the intial phases and magnitudes of the respective bins can be derived. Complex numbers are used in lieu of sine and cosine combinations because it reduces the complexity of the algebra and calculations involved in subsequent processes. The magnitude of bin k can be obtained by:

$$Mag_{rect} = \sqrt{x^2 + y^2} \quad OR \quad Mag_{polar} = |x + jy| \tag{24}$$

The phase (θ) of bin k by:

$$\theta = \arctan \frac{y}{x} \tag{25}$$

Frequency bins are equally distributed from 0 Hz to f_s at the interval of $\frac{f_s}{N}$. The first bin, 0 Hz is referred to as the DC Component, because direct current (DC) only travels in one path, unlike AC which oscillates. The DC component contributes to offset or bias, which is the distance from "center" the waveform lies. The information contained from bin $k_{N/2+1}$ through bin k_N is redundant; the magnitudes and phases above are a mirror image of the lower bins with the Nyquist bin ($k_{N/2}$) as the point of symmetry.

As any scientist will tell you, there is an amount of uncertainty contained in every measurement or calculation. The uncertainty with the STFT deals with the balance between the resolution of frequency versus time. Longer windows allow for finer grades in frequency resolution, but time

[21] j is used for the $\sqrt{-1}$ in lieu of i which denotes current in electrical engineering

localization which is the ability to pinpoint when those frequencies occur, is less absolute [109]. Frequency and time resolution are inversely proportional and an increase in one, always results in a decrease in the other.

The DFT is relatively slow and requires approximately N^2 multiplications, or four additions per complex number. The Fast Fourier Transform (FFT) is an algorithm which relies on a power of two window length, and speeds up the process through the elimination of redundant computations requiring $N \times \log_2(N)$ operations [14]. In the event that a non-power of two window size is used, the concatenation of zeros to the end of the time domain series, or **zero-padding**, solves the issue. Zero-padding a window or frame can also increase the apparent frequency resolution of an analysis.[22] Depending upon the window size of the FFT, the increase in calculation speed can be immense, and this is critical if techniques are to be applied in real-time scenarios. The size of N also plays a major part in the minimum latency of a real-time system.

Utilizing the Inverse Fourier Transform(IFT), or Inverse Fast Fourier Transform(IFFT), the frequency domain information can be used to reconstruct the original time domain series. Processing in various forms can also be inserted in between the FFT and IFFT processes. While it is said that the reconstruction is identical to the original, there will always be minimal errors associated with round off, and the precision of calculation. Spectral analysis alone is a powerful tool, but a number of important sub-processes have been developed which center around it. More complete details, and an additional list of sources on the inner workings of the DFT and FFT are available in Curtis Roads *The Computer Music Tutorial* [14].

D Filtering

A filter simply stated is something that removes a given member of a set. Filters in audio remove or boost frequency components, generally by mathematical operations in the time domain. Filters are defined by three general characteristics : type, cut-off or center frequency, and order. Filter types generally employed are low-pass, high-pass, all-pass and band-pass. Low-pass filters remove high frequency content and allow lower frequencies to pass. Low pass filters are also referred to as averaging, or mean filters. High-pass filters utilize differences in sample values and block lower frequencies. All-pass filters do not affect the magnitude of frequency components and are typically used in to create specific phase shifts in signals. Band-pass filters allow a given set of frequencies to pass.

The cut-off frequency applies to low and high pass filters, and is defined as f_{co}. Band-pass filters are defined by a center frequency, and the width of the frequency band allowed to pass is generally defined by Q, which stands for quality factor. A collection of filters, typically a set of band-pass filters, is referred to as a filter bank. The order of a filter represents the amount of reduction imposed on the frequencies being removed from the signal. For each order, a three decibel(-3db) drop per octave occurs in the band above/below f_{co}. For a more in-depth overview of filter mechanics, additional types and specifications see [14, 52].

[22] Zero-padding relies on interpolation, therefore it does not increase **actual** frequency resolution.

References

[1] Tetsuro Kitahara. *Advances in Music Information Retrieval*, chapter Mid-Level Representations of Musical Audio Signals for Music Information Retrieval, pages 65–91. Springer-Verlag, Berlin, 2010.

[2] Vincent Verfaille, Udo Zolzer, and Daniel Arfib. Adaptive digital audio effects (a-dafx) : A new class of sound transformations. In *IEEE Transactions on Audio, Speech, and Language Processing*, pages 1–15, 2006.

[3] Martin J. Morrell and Joshua D. Reiss. Dynamic panner : An adaptive digital audio effect for spatial audio. In *Audio Engineering Society 127tth Convention*, New York, NY, October 9-12 2009.

[4] Vincent Verfaille, Marcelo M. Wanderley, and Philippe Depalle. Mapping strategies for gestural and adaptive control of digital audio effects. *Journal of New Music Research*, X(Y):zz–zz, 2006.

[5] *High-Level Semantic Metadata for the Control of Multitrack Adaptive Digital Audio Effects*, San Francisco, USA, October 2012.

[6] Vincent Verfaille and Philippe Depalle. Adaptive effects based on stft, using a source-filter model. In *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx'04)*, pages 296–301, Naples, Italy, October 5-8 2004.

[7] Jeffrey M.Morris. *Computer Music Modeling and Retrieval : Genesis of Meaning in Sound and Music*, chapter Ontological Substance and Meaning in Electroacoustic Music, pages 216–226. Springer-Verklag, 2009.

[8] Gordon Mumma. *The Development and Practice of Electronic Music*, chapter Live Electronic Music, pages 286–343. Prentice-Hall, Inc., 1975.

[9] Kim Cascone. Grain, sequence, system : Three levels of reception in the performance of laptop music. *Contemporary Music Review*, 22(4):101–104, 2003.

[10] Cornelius Poepel and Roger B. Dannenberg. Audio signal driven sound synthesis. In *Proceedings of the 2005 International Computer Music Conference*, pages 391–394, 2005.

[11] Miller S. Puckette. Patch for guitar. In *On-line Proceedings, PD Convention*, August 13 2007. available : http://pure-data.ca/en/.

[12] Cornelius Poepel. *Interactive Multimedia Music Technologies*, chapter IX: Driving Sound Synthesis with a Live Audio Signal. Information Science Reference, 2008.

[13] Miller S. Puckette. Grafting synthesis patches onto live musical instruments. In *Proceedings, International Computer Music Conference*, pages 250–253, 2010.

[14] Curtis Roads. *The Computer Music Tutorial*. M.I.T. Press, 1996.

[15] Johnn Eaton. This is an instrument. *Contemporary Music Review*, 18(4, part 3):21–24, 2003.

[16] Diana Young. The hyperbow controller: Real-time dynamics measurement of violin performance. In *In Proceedings of the Conference on New Instruments for Musical Expression, NIME*, Dublin, Ireland, 2002.

[17] Chad Peiper, David Warden, and Guy Garnett. An interface for real-time classification of articulations produced by violin bowing. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2003.

[18] Emmanuel Fléty, Nicolas Leroy, Jean-Christophe Ravarini, and Frédéric Bevilacqua. Versatile sensor acquisition system utilizing network technology. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2004.

[19] Frédéric Bevilacqua Nicolas H. Rasamimanana, Emmanuel Fléty. Gesture analysis of violin bow strokes. IRCAM.

[20] Ray Kurzweil. *The Age of Spiritual Machines : When Computers Exceed Human Intelligence*. Penguin Books, 1999.

[21] Colin Gough. *The Science of String Instruments*, chapter Electric Guitar and Violin, pages 393–415. Springer, 2010.

[22] Julie Lyonn Lieberman. *Alternative Strings : The New Curriculum*. Amadeus Press, 2004.

[23] Joseph Curtin and Thomas D. Rossing. *The Science of String Instruments*, chapter 13 : Violin. Springer, 2010.

[24] Thomas D. Rossing and Graham Caldersmith. *The Science of String Instruments*, chapter Guitars and Lutes. Springer, 2010.

[25] Tony Bacon. *Electric Guitar : The Illustrated Encyclopedia*. Backbeat Books, 2000.

[26] Nick Freeth and Charles Alexander. *The Electric Guitar*. Salamander Books Limited, 2000.

[27] Joel Chadabe. *Electric Sound : The Past and Promise of Electronic Music*. Prentice Hall, 1997.

[28] Thomas D. Rossing. *The Science of String Instruments*, chapter Plucked Strings. Springer, 2010.

[29] Gareth Loy. *Musimathics : The Mathematical Foundations of Music*, volume 1. MIT Press, 2006.

[30] David M. Howard and Jamie A.S. Angus. *Acoustics and Psychoacoustics : Fourth Edition*. Focal Press, 2009.

[31] Kevin Karplus and Alex Strong. Digital synthesis of plucked strings and drum timbres. *Computer Music Journal*, 7(2):43–55, Summer 1983.

[32] Carl Flesch. *The Art of Violin Playing : Book One.* Carl Fischer, 2000.

[33] Knut Guettler. *The Science of String Instruments*, chapter 16 : Bows, Strings, and Bowing, pages 279–299. Springer, 2010.

[34] Elizabeth A. H. Green. *Orchestral Bowings and Routines.* Ann Arbor Publishers, 2nd edition, 1962.

[35] Carl Flesch. *The Art of Violin Playing : Book Two, Artistic Realization and Instruction.* Carl Fischer, 2000.

[36] Thomas D. Rossing and Roger J. Hanson. *The Science of String Instruments*, chapter 12 : Bowed Strings, pages 197–208. Springer, 2010.

[37] Dave Hunter. *The Guitar Pickup Handbook : The Start of Your Sound.* Backbeat Books, 2008.

[38] Dan Erlewine. *Guitar Player Repair Guide : How to Set Up, Maintain, and Repair Electronics and Acoustics.* Backbeat Books, 2nd edition, 1994.

[39] Earl Gates. *Introduction to Electronics.* Delmar : Cengage Learning, 6th edition, 2012.

[40] Enda Bates, Dermot Furlong, and Donnacha Dennehy. Adapting polyphonic pickup technology for spatial music performance. International Computer Music Conference, 2008.

[41] Miller Puckette. *The Theory and Technique of Electronic Music.* World Scientific Press, 2007.

[42] George Tzanetakis. *Music Data Mining*, chapter Audio Feature Extraction. CRC Press, 2012.

[43] Daniel J. Levitin. *This is Your Brain on Music : The Science of Human Obsession.* Penguin Group, 375 Hudson Street, New York, New York, 2006.

[44] Forest M. Mims III. Engineers mini-notebook : 555 timer ic circuits. Siliconcepts, 1984.

[45] Hassan Parchizadeh and Branislav Vuksanovic. *PIC Projects : A Practical Approach.* John Wiley and Sons, LTD, 2009.

[46] Dan O'Sullivan and Tom Igoe. *Physical Computing : Sensing and Controlling the Physical World with Computers.* Course Technology, Cengage Learning, 2004.

[47] Albert Paul Malvino. *Semiconductor Circuit Approximations : An Introduction to Transistors and Integrated Circuits.* McGraw-Hill, Inc., fourth edition edition, 1985.

[48] Monty J. Strauss, Gerald R. Bradley, and Karl J. Smith. *Calculus.* Prentice Hall, 3rd edition edition, 2002.

[49] Forest M. Mims III. Engineers mini-notebook : Op amp ic circuits. Siliconcepts, 1985.

[50] Craig Anderton. *Electronic Projects for Musicians.* Amsco Publications, 1980.

[51] Brice Ward. *Electronic Music Circuit Guidebook*. Tab Books, 1st edition, 1975.

[52] Will Pirkle. *Designing Audio Effect Plug-Ins in C++: With Digital Audio Signal Processing Theory*. Focal Press, 2010.

[53] David Friend. An integrated guitar synthesizer for live performance. In *58th Audio Engineering Society Convention*. Audio Engineers Society, November 4-7 1977.

[54] Mark Vail. *Keyboard Presents : Vintage Synthesizers*. Miller Freeman Books, 2nd edition, 2000.

[55] Trevor Pinch and Frank Trocco. *Analog Days : The Invention and Impact of the Moog Synthesizer*. Harvard University Press, 2002.

[56] Hal Chamberlain. *Musical Applications of Microprocessors*. Hayden Books, 1985.

[57] Hideki Kawahara Alain de Cheveigné. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, April 2002.

[58] Nigel Gough. Guitar-to-midi system. *Everyday Practical Electronics*, 38(10):38–43, October 2009.

[59] Mamoru Ishikawa, Takeshi Matsuda, and Michael Cohen. Guitar-to-midi interface: Guitar tones to midi notes conversion requiring no additional pickups. In *Audio Engineering Society 129th Convention*, San Francisco , CA, November 4-7 2010.

[60] H.S. Carslaw. *Introduction to the Theory of Fourier's Series and Integrals*. Dover Publications, Inc., 3rd edition, 1930.

[61] Bruce Bartlett and Jenny Bartlett. *Practical Recording Techniques : The Step by Step Approach to Professional Audio Recording*. Focal Press, 3rd edition, 2002.

[62] D. Arfib F.Keiler and Udo Zolzer. *DAFX - Digital Audio Effects*, chapter Source Filter Processing, pages 299–372. John Wiley and Sons, LTD, 2002.

[63] Simon Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digiatl Audio Effects (DAFx-06)*. Austrian Research Institute for Artificial Intelligence, September 18-20 2006.

[64] Michiel Schuijer. *Analyzing Atonal Music: Pitch-Class Set Theory and Its Contexts*. University of Rochester Press, 2008.

[65] Roger N. Shepard. Circularity in judgements of relative pitch. *Jouranl of the Acoustical Society of America*, 36:2346–2353, 1964.

[66] Roger N. Shepard. Approximation to uniform gradients of generalization by monotone transformations of scale. In D.I. Mostofsky, editor, *Stimulus Generalization*, pages 94–110. Standford University Press, Stanford, California, 1965.

[67] Roger N. Shepard. *Stuctural Representations of Musical Pitch*, chapter 11, pages 344–390. Academic Press, Inc., Orlando, Florida, 1982.

[68] Matthias Mauch and Simon Dixon. Approximate note transcription for the improved identification of difficult chords. In *Proceedings of the The International Society for Music Information Retreival*, pages 135–140, 2010.

[69] Geoffroy Peeters. Chroma-based estimation of musical key from audio-signal analysis. In *Proceedings of the The International Society for Music Information Retreival*, page 1, 2006.

[70] Juan P. Bello and Jeremy Pickens. A robust mid-level representation for harmonic content in music signals. In *Proceedings of the 6th International Conference on Music Information Retrieval*, London, UK, September 2005.

[71] Juan Palbo Bello et al. A tutorial on onset detection in musical signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, September 2005.

[72] Juan P. Bello, Giuliano Monti, and Mark Sandler. Techniques for automatic music transcription. Department of Electronic Engineering, King's College London, Strand, London.

[73] Juan Pablo Bello and Mark Sandler. Phase-based note onset detection for music signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 441–444. IEEE, April 2003.

[74] Mike Davies Juan P. Bello, Chris Duxbury and Mark Sandler. On the use of phase and energy for musical onset detection in the complex domain. In *IEEE Signal Processing Letters*, pages 553–556. IEEE, June 2004.

[75] Xander Fiss and Andres Kwasinski. Automatic real-time electric guitar audio transcription. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 373–376. IEEE, May 2011.

[76] Tuomas Virtanen Fawad Mazhar, Toni Heittola and Jukka Holm. Automatic scoring of guitar chords. In *45th International Convention of the Audio Engineering Society*, Helsinki, Finland, March 1-4 2012.

[77] Lorenzo J. Tardón Ana M. Barbancho, Anssi Klapuri and Isabel Barbancho. Automatic transcription of guitar chords and fingering from audio. In *IEEE Transactions on Audio, Speech, and Language Processing*, pages 915–921. March 2012.

[78] Lance Alcabasa and Nelson Marcos. Automatic guitar music transcription. In *2012 International Conference on Advanced Computer Science Applications and Technologies*, pages 197–202. IEEE, 2013.

[79] Jaakko Siiskonen Henri Penttinen and Vesa Välimäki. Acoustic guitar plucking point estimation in real-time. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 209–212. IEEE, March 2005.

[80] Bingjun Zhang Olaf Schleusing and Ye Wang. onset detection in pitched non-percussive music using warping-compensated correlation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 117–120. IEEE, March-April 2008.

[81] Nick Collins. Using a pitch detector for onset detection. University of Cambridge, Centre for Music and Science, 2005.

[82] Gardner Read. *Music Notation : A Manual of Modern Practice.* Crescendo, 2nd edition, 1979.

[83] Kurt Stone. *Music Notation in the Twentieth Century: A Practical Guidebook.* W.W.Norton and Company, 1980.

[84] Stephen Kostka. *Materials and Techniques of Twentieth-Century Music.* Prentice Hall, 2nd edition, 1999.

[85] Samuel Adler. *The Study of Orchestration.* W.W. Norton and Company, Inc., 3rd edition, 2002.

[86] Kent Wheeler Kennan. *The Technique of Orchestration.* Prentice-Hall, 1952.

[87] Christopher Parkening in Collaboration with Jack Marshall and David Brandon. *The Christopher Parkening Guitar Method, Vol 1 : The Art and Technique of Classical Guitar.* Hal Leonard, 1999.

[88] Anthonly Glise. *Classical Guitar Pedagogy : A Handbook for Teachers (Books I-III).* Mel Bay, 1997.

[89] Hector Quine. *Guitar Technique : Intermediate to Advanced.* Oxford University Press, 1990.

[90] William Leavitt. *A Modern Method for Guitar : Complete*, volume 1-3. Berklee Press, 1999.

[91] Meinard Muller. *Information Retrieval for Music and Motion.* Springer-Verlag, 2007.

[92] Richard Wagner. *Tristan and Isolde : Partitura.* Konemann Music, Budapest, Hungary, 1994.

[93] David A. Rosenbaum. *Human Motor Control.* Academic Press, Inc., 1991.

[94] Claude Debussy. *Nocturnes for Orchestra in Full Score : Dover Miniature Scores.* Dover Publications, Inc., 2005.

[95] Hans Gál, editor. *Johannes Brahms Complete Symphonies : In Full Orchestral Score.* Dover Publications, New York, NY, 1974.

[96] Igor Stravinsky. *Pribaoutki, Renard and Ragtime for Eleven Instruments.* Dover Publications, Inc., Mineola, New York, 2000.

[97] Igor Stravinsky. *The Firebird : Original 1910 Version.* Dover Publications, Inc., Mineola, NY, 2000.

[98] Wilhelm Altmann, editor. *Joseph Haydn : String Quartets, Op.22 and 33, Complete*. Dover Publications, Inc., Mineola, NY, 1985.

[99] Edward Lorenz. Predictability : Does the flap of a butterfly's wings in brazil set off a tornado in texas? In *139th Annual Meeting of the American Association of Science*. American Association for the Advancement of Science, December 29th 1979.

[100] Gerhard Nierhaus. *Algorithmic Composition : Paradigms of Automated Music Generation*. SpringerWein, 2009.

[101] Brian W. Kernighan and Dennis M. Ritchie. *The C Programming Language*. Prentice Hall PTR, second edition edition, 1988.

[102] Max Mathews. *The Technology of Computer Music*. M.I.T. Press, 1979.

[103] John M. Chowning. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7), 1973.

[104] Audio Engineering Society. *Audio Effects for Real-Time Performance Using Beat Tracking*, 122nd Convention, Vienna, Austria, May 2007.

[105] A.J. Eronen A.J. Klapuri and J.T. Astola. Analysis of the meter of acoustic musical signals. In *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 14, pages 342–355. IEEE, January 2006.

[106] Joel Chadabe. Basic sound and acoustics 1. Handout, October 2011.

[107] Aaron Copland. *What to Listen for in Music*. New American Library, 2002.

[108] Stefan Kostka and Dorothy Payne. *Tonal Harmony : With an Introduction to Twentieth-Century Music*. McGraw Hill, 4th edition, 2000.

[109] Mark Ballora. *Essentials of Music Technology*. Prentice Hall, 2003.

[110] Ken C. Pohlmann. *Principles of Digital Audio*. McGraw-Hill, 5th edition, 2005.