WHAT DOES MUSIC LOOK LIKE?

VOLUME 1: THE MANY FACES OF SWING

TABLE OF CONTENTS

Chapter 1 Sound, Music, Rhythm and Dance *1*

- Motivation, and cultural background of Swing. Exploring the basics of the Science of Sound. How this is used for analysis of music. Description of frequency spectrum and other primary visualization tools for analyzing time/frequency characteristics of music.

Rhythm *3*

- In standard European musical notation, also called tablature, time is counted based on a fixed "master" beat, and subdivision of the time between main beats by a factor of 2, giving the familiar whole note, half note, quarter note and so on. For convenience we call this style "Mozart-Bach" or MB notation. Many types of music do not follow this paradigm.

Music: Cultural Aspects *5*

- How learning the "right" way to play music left me seriously handicapped.

Music: Technical Aspects *6*

- Sound gets complicated as several instruments play simultaneously.

Time, Frequency and Sound Waves *8*

- Frequency and Pitch. Sine waves. Additive synthesis of complex waveforms from simple waveforms. Introduction to Fourier analysis.

Computer Analysis of Music *14*

- How to extract musical note features by selecting frequency ranges where the main part of an instrument's sound exists. Examples of the spectra of several instruments, and vocals. Spectral Graphs, Time Series Graphs, Diffdot Plots.

Aside: Learning a Caixa Batida, and Perception of Time *24*

- Playing Samba batucada is like standing between two freight trains passing each other at 70 miles per hour. The immense sound volume seems to affect the speed of auditory perception. Visual perception is apparently not affected.

Dance and Motion *26*

- How does a box headed computer nerd like me know whether the music is swinging or not? Get a clue: watch the Samba dancers. If they are moving, the music is swinging. If they are standing around talking about shopping, it ain't.

Mathematical Models of Swing *27*

- Railroads, trolley cars, horses etc. Kinematics in modeling character animation for computer graphics special FX. Spring-mass systems and differential equations.

Aside: More about Mozart-Bach Notation *29*

- There are many Brazilian rhythms which are written identically, but played differently. Although subtle, these details are amenable to technical analysis.

Chapter 2 Roots of Modern Swing Music ***34***

- Using the computer visualization graphs, we look at the rhythms from the Early Days of Swing. Cross reference with standard books like Time/Life series.

Ragtime *34*

- Rhythmic breakout from constraints of European Classical counting and subdivision by out-Mozarting Mozart.

Boogie-Woogie *34*

- When the club music didn't even pretend to be respectable.

Blues *34*

- An old tradition that is one of the roots of many music styles.

Gospel and Soul *34*

- Making sure that the Sinners weren't the only ones who had swing.

Dixieland and Swing *34*

- Swing makes a (fairly) graceful entrance into the parties of White America.

Chapter 3 Classic American Swing ***35***

- Apply the visualization tools to standard Jazz tunes

Louis, Duke and Benny *35*

- The Big Three of Swing

Big Band Sound *35*

- Glenn Miller et al, adapting Swing for mainstream America

Django Reinhardt and Alternative Jazz *35*

- Cross over between Jazz and more Folk like music

Tinsel Town *35*

- Swing tunes in the movies

After the War: R & B, Jungle Music *35*

- Look at Ray Charles, James Brown, Richie Valens et al.

Cowboy Swing *35*

- American Country & Western, Bluegrass etc

Motown *35*

- Not every tune from Motown had Swing, but the ones that did, really had it.

New Orleans *35*

- Although Chicago, Saint Louis and Kansas City would disagree, New Orleans really is the Heart and Soul of many forms of American music including Swing.

Chapter 4 Brazilian Swingbee ***36***

- Explanation of African Roots in Brazil. Candomble, Favelas, Folk music. Exploration of the standard forms of Swing in the Brazilian style.

Samba, Bossa Nova and Pagode *36*

- The classic Brazilian sounds. Many rhythms from these three similar styles can be found in American Rock and Roll and other familiar music.

Samba-Reggae, Samba-Afro, and Samba-Funk *36*

- Brazil joins Bob Marley and other Freedom Fighters who use drums, guitars and trumpets instead of guns in their struggle for equality.

O Nordeste *36*

- The Northeast area of Brazil, including Salvador, Recife, Pernambuco, Manaus, plays a role similar to New Orleans, Memphis and Nashville in American Music.

Desinhas *36*

- Brazilian music has many ornamental rhythms in its repertoire. In some ways, these correspond to clave in Cuban music. Clave is familiar to fans of American music as the “Bo Didley” beat. Brazilian desinhas are perhaps not as well known although they appear in many tunes. One example that may be familiar is “Touch Me” by the Doors.

California Brazil Camp *36*

- If you can take six months and go to Brazil to play batucada with some Escola de Samba in Rio de Janeiro or Salvador, do it! Otherwise if you want a top quality immersion experience in Brazilian music, dance and culture, go to California Brazil Camp for a week. Get your 20 hours a day of Brazilian culture and party.

Chapter 5 The Caribbean Stylee ***37***

- A look at many flavors of Caribbean music and their special swing styles.

Reggae *37*

- “tight and straight” describes Reggae, but there is nothing square about this music.

Calypso *37*

- I need to dig into the other Islands. Unique rhythms and swing feelings but I don’t know much about these yet.

Cubanismo *37*

- Havana and New Orleans were for years like the North and South poles of a magnet: inseparable, unified and yet radically different from each other.

Clave *37*

- The lopsided clave is a signature rhythm in Salsa and other Cuban styles. The basic clave is 3 equally spaced beats with a value of 1 1/2 eighth notes followed by 2 equally spaced beats with a value of 1 eighth note. Together they fill a bar of music, giving it two easily identified “sides”, and a tension/release characteristic. There are many variations of the clave. Beware of trying to count them with standard subdivision! This may lead you to play a particular clave incorrectly.

Chapter 6 Modern Pop Swing ***38***

It don't mean a thing (if it ain't got that Swing)

–– Duke Ellington and Irving Mills (1923)

Chapter 1 Sound, Music, Rhythm and Dance

Motivation, and cultural background of Swing. Exploring the basics of the Science of Sound. How this is used for analysis of music. Description of frequency spectrum and other primary visualization tools for analyzing time/frequency characteristics of music.

Louis Armstrong is famously quoted as saying "If I gotta tell you, then you're never gonna know" when some one asked him *"What is Jazz?"* He may have said something similar about Swing, but I haven't found such a quote yet. Nonetheless, like Jazz, people have traditionally considered swing music to be an intuitive and somewhat mysterious essence that gives a tune pep and vigor that isn't found in many other musical styles.

I am going to tell the story of my quest for Swing, and how a white guy from Kansas found rhythm. Parts of the story comes from computer science, parts come from my experience playing live music in California, New Orleans and Brazil. Equally important is large amounts of time I spent listening to recorded music and sometimes playing along with it. If you already know how to swing a tune, you may find my approach novel and interesting. You may gain some insight into things that you know intuitively but not analytically. If you are baffled by Swing, as I was, I hope that this story will help you understand how Swing works, and even to learn how to play music with an authentic swing style.

Ask a musician what makes music swing. You will probably hear that swing is a *feeling*, and maybe some information about counting or subdividing the beat. Commonly, triplet subdivision of the main beat is a feature in swing music, but this is not the entire story or else a waltz in 3/4 time, and 6/8 or 12/8 meter would inherently swing. Some of these do, and some don't. Swing is found in Jazz, Blues, Motown and many other styles of American music. Swing can also be found in music from other parts of the world: Brazil, Cuba, Jamaica, Trinidad, Africa, the Middle East. Is there some common thread that ties together all these seemingly separate musical traditions? Jazz is often thought to be the original source of swing, but Blues and Gospel both have swing that is undoubtedly as old as Jazz.

Since we know that classic Jazz is not the only form of swing, we want to extend the view of Swing to include all musical examples that might be considered to "swing" by some valid measure (e.g. the musicians or dancers think the music is swinging). I use an ad hoc cultural rather than technical definition for swing: it is a property of music *as played* that causes listeners to dance or otherwise move their bodies in an energetic rhythmic manner. This definition lets us consider a broad range of music, beyond well known forms like Jazz. It also allows us to distinguish between ***Swing*** and other types of rhythmic expression.

What is *rhythmic expression*? First, consider the difference between written music (data) and music as played by humans (information). Simply stated, a musician plays music differently from the exact rendition of the musical data as it might be played by a computer sequencer or other device based on a clockwork metronome. Rhythmic expression is a technique of modifying the timing of notes as played versus notes as written. Typically this gives a less mechanical feeling to the music. These modifications are often fairly subtle.

The sheet music is a guide to the content of the music performance, giving the sequence of musical notes, their pitches and (canonical) time durations for all the instruments which are playing the tune. Like a movie script, this written form gives you a lot of information about how things go together, but very little information about how a real person would actually render the information in its intended form: music that we listen to a*nd enjoy* as a break from more structured aspects of life. Just as an actor or actress gives us more in their performance than is written in the script (e.g. emotion, humor, irony), a musician also interprets the written music to give it depth and life.

While a musical score contains a fair amount of guidance for dynamics or loudness (*piano, forte, mezzo-forte* etc., plus the customary accents that happen on the downbeat of a measure), rhythmic expression information in sheet music is usually quite limited (*rubato, with a swing feel* etc.). Moreover, if you are unfamiliar with a piece, or musical style, the instructions for rhythmic expression are actually not much help.

New Scientist magazine reported that when *West Side Story* opened in London in 1959, the theatrical producers had a very difficult time finding a drummer who could play the score properly. Classical drummers could read the sheet music very well but didn't know the Jazzy style and couldn't swing the beat. Jazz and Pop drummers could play with the right swing feel but couldn't read music well enough to follow the complexities that Leonard Bernstein had put in the score.

Eventually a drummer was found who could adequately read sheet music *and* swing the beat. This episode points to a basic inadequacy of standard European music tablature – it is a system based on very specific and limited counting and subdivision, primarily dividing the beat by factors of two. As noted above, triplet subdivision is common in swing. There is no way to get a triplet by subdividing by 1/4, 1/8, 1/16 notes, although there is the special case of eighth or quarter note symbols being tied together into a triplet group. Still, this only gives one variation, and doesn't adequately represent the many varieties that Swing can take.

Some traditions do not write down the music at all. Some very excellent professional musicians don't read standard (European) music notation. Much of the music from these sources has excellent toe tapping swing. Indeed, the foundation for many of the familiar musical forms that swing comes from the drum and dance traditions of West Africa, brought to the Americas on slave ships, and kept alive for centuries by the cultural practice of Black people from Brazil to the Caribbean to the USA. Each place where these traditions survived developed unique but related versions of the common roots.

Before beginning the journey into the details of Swing, I need to review some technical aspects of sound. While there is substantial mathematics behind the theory, I will try to present the information in an accessible, non-scary (but accurate) manner. In the bibliography I also include pointers to more technical versions of the same information.

RHYTHM

In standard European musical notation, also called tablature, time is counted based on a fixed "master" beat, and subdivision of the time between main beats by a factor of 2, giving the familiar whole note, half note, quarter note and so on. For convenience we call this style "Mozart-Bach" or MB notation. Many types of music do not follow this paradigm.

When I was in New Orleans in 2001, I often rode the Saint Charles streetcar to and from the French Quarter out towards Tulane University, to stay at the Youth Hostel, and connect up with Casa Samba for practices, gigs, lessons with Jorge Alabe, costumes and the other activities that go into putting on a samba show. The streetcars are probably 100 years old which is *ancient* in the United States, and it occurred to me that some days I might be sitting on the exact same bench that Louis Armstrong or other great musician sat on when he was kid growing up in New Orleans. The cars run on old rails that are straight enough to be safe at low speeds, but overall have a lot of features and artifacts that make the streetcar shift and bounce constantly. One day I had one of those *Eureka!* moments going down the street with dozens of different clunks, clanks, rattles, snicks, squeaks and so on. I suddenly realized I was listening to *Samba batucada* emerging from the synchronized randomized rhythm that the streetcar was making. I had long known about the connection between riding on railroads and the beat in musical forms like the Blues, but here suddenly was a laid back, quick tempo, loose but utterly locked together rhythm with both great simplicity and elegant complexity that baffles the mind to decipher its pattern, a pattern that is obviously a product of human intention – on the streetcar, the accidental connections between mechanical parts of the vehicle system design, and in Samba, the intricate interlocking of several rhythmic parts, or *batidas*. *If only*, I thought, *I could play rhythm like this streetcar!*

In the forms of Swing that I've studied, some of the rhythmic structure may be subdivided by other than a factor of two. Generally, there are some timing features which can be divided in two, four, eight etc., and some that shouldn't be. Often a two bar or four bar phrase is obvious, but the master beat or *pulse* may not always be divided exactly in half or quarter. This yields a lopsided feel to the timing that contributes to the overall feeling of Swing. Additionally, some examples of pulse have three or five notes arranged in a non symmetric pattern giving a lopsided feel of a different sort. It seems that almost any feature of timing that is not "exact" in some arithmetic sense, or doesn't have a symmetrical structure based on factors of two, can help produce a Swing feel, *if it has repetitive or cyclic features*.

Although Swing has characteristics that are not found in MB notation and the metronome worldview, it is not typically less precise. Merely playing with sloppy timing rarely gives a feeling of Swing. For example, in the MB world, the timing between a series of pulse notes is a constant, say every 2/3 second with no variation. By contrast, swing pulse timing may be composed of two or more time intervals with different lengths, not necessarily very much difference, but perceptible either directly or intuitively. In the 1962 recording of *It Don't Mean a Thing (if it ain't got that Swing)* by Louis Armstrong and Duke Ellington, the range of pulse time intervals is between 610 and 640 milliseconds and the variations of timing have a somewhat consistent *short-long* style. The slight variations added to a larger pattern which is consistent adds an organic or vital quality to the experience, a kind of texture.

If the musician's goal is to produce a very precise or *tight* rhythm, there are also limits of human bodily precision. It is often more important for the rhythmic notes to consistently connect with each other than that they are exact in the metronomic sense.

Kim Atkinson[1] told me a story about some West African drummers he knows, and their attempts to use standard tablature form to notate some complex rhythms from their tribal tradition. The rhythms have elements of 4/4 and 6/8 meter, plus additional counting tricks and timing subtleties. The drummers transcribed their rhythms into both 4/4 and 6/8 and then the rhythmic data was played by a computer sequencer. Neither the 4/4 nor the 6/8 meter adequately captured the authentic count or feeling of the rhythms. The difficulty apparently is that the rhythms really are both and neither 4/4 and 6/8 at the same time, and then some extra tricks as well. This is a fairly common motif in West African drumming and its Brazilian, Cuban and other American or Caribbean descendants.

Jorge Alabe[2] generally breaks down the conceptual aspects of Brazilian Samba into the *pulse* and the *swinghee*. The pulse is closely related to the downbeat, backbeat and such major rhythmic landmarks. The swinghee, also called *balance* in Portuguese, is the rhythmic expression which distinguishes various examples of a *batida* or beating pattern. The batidas of various samba schools can be transcribed into tablature and some would be recognizably different from each other. Other batidas would look identical on paper, but when played they have distinct differences that may be quite subtle. Learning to distinguish between, say, *Mocidades* and *Mangueira* samba schools can be tricky to say the least, and just to learn that there ***is*** a difference has taken me a number of years of experience listening to and playing different examples of batucada.

Pagode is a folk music form in Brazil where many people at neighborhood and other gatherings play, sing or listen several times a week to a long list of well known standard

---

[1] Kim Atkinson is a professional drummer and drum teacher in the San Francisco Bay area.

[2] Jorge Alabe is a Master drummer from Rio de Janeiro, Brazil.

tunes. Starting from childhood, the early exposure gives Brazilians a natural knowledge of the music patterns and feelings without necessarily resorting to technical learning approaches like counting and subdividing the beat in the MB style. Indeed, many very excellent Brazilian musicians don't read music[3], and some have difficulty counting time in the European style. One time Mestre Beiçola tried to teach us *samba de roda* which has a tricky 3 against 4 feeling. He plays it very well, but learned by ear and so had difficulty counting it in MB style. He ended up explaining the feel by "hit here [right hand], here [left hand], wait a little bit, here [right hand].' I was still hung up on the tablature style of rhythmic subdivision that I had learned in piano lessons and music classes in school. The information that Beiçola had gained from his early exposure to a less mechanical approach to rhythm took several years to really sink into my brain, but eventually the knowledge base of drumming from a *listen and play* tradition versus a *count and subdivide* approach began to make sense to me.

MUSIC: CULTURAL ASPECTS

How learning the "right" way to play music left me seriously handicapped.

When I was a kid, my dad played Ragtime piano. I loved it, and thought it was really cool, although I didn't have that word in my vocabulary (I probably called it *neat-o!* or something like that). Ragtime had an exciting and interesting quality that I don't think I experienced in any other part of my exceedingly normal life. At one point when I was maybe eight or ten I announced that I wanted to learn to play the piano too. Unfortunately at this point my parents did the dutiful middle class thing, sought out a conventional[4] piano teacher and enrolled both me and my brother for a year of sheer torture, playing exercises and short pieces that I didn't like at all, the tedious bother of copying and transcribing short musical lines in different key signatures to learn all the keys, and other very standard tasks that are quite important to anyone who really wants to know about European music, or at least the mainstream technical aspects of it.

I didn't really want to do that. I didn't even realize what music theory and education was although I was completely immersed in the very process of doing it. It was sort of a forest and the trees effect, combined with developing linguistic descriptions of the world, emergence of symbolic consciousness and such standard growing up and learning stuff that we all go through in some form. What I *wanted* was to play the lively, bouncy energetic, somewhat rowdy key jangling that my dad did. What I got was the opposite: rigid structure, technical and mechanistic information about a topic that I could not, with effort, have

---

[3] e.g., Airto Moreira, a professional musician from southern Brazil who is a major innovator in American Jazz. Airto has played extensively with Miles Davis and other well known Jazz and Latin musicians.

[4] My mom assures me that the teacher they picked was the *least* conventional of all the ones they talked to.

cared less about. Subsequently, after the dismal piano lessons episode, I did little in the way of playing music for many years to come.

One thing I did that made the experience more pleasant was to play the exercises with what I now recognize as Swing, giving the simple mechanical rhythms a more bouncy feel by playing the customary short-long type of timing pattern. In my youthful enthusiasm, I showed my teacher how much more fun it was to play with the bouncy style. That was a BIG MISTAKE! The teacher conferred with my parents about how I was playing things *wrong* and how important it was to lock into the regular tick-tock of the metronome. Consequently, I was forced to play all the exercises v-e-e-r-y s-l-o-w-l-y so I would "learn" about regular note timing and comply with the "correct" approach. This was greatly exacerbated by the fact that when set to a slow tempo, our metronome became increasingly lopsided in its tick-tocks. It also had a certain amount of randomness at extremely slow tempos. Trying to match this sequence of beats proved difficult, and the more trouble I had, the metronome was set to a slower tempo, making matters worse. Despite the fairly obvious deficiency of the metronome, like a doctor, the machine was considered to be the expert and its directions regarding timing must be followed meticulously, an impossible task in this case.

This episode left me severely rhythm impaired for decades. After about five years of studying Brazilian percussion, I finally started to understand that my entire rhythmic knowledge base had been built on erroneous metaphors.

MUSIC: TECHNICAL ASPECTS

Sound gets complicated as several instruments play simultaneously.

Stu Fessant[5] has given me a lot of valuable exposure to musical knowledge. First was the point that Reggae rhythm often leaves the downbeat as silence, using this gap as a concrete feature in the music. Second was to assure me that the pulse in music is very steady, and that the rhythmic complexities are not random or mysterious, but that they are intimately connected to the pulse. Although this is a simple thing, and probably everyone in the world except me knew it, discovering the concept of the steady pulse was the beginning of my being able to understand rhythm. Prior to this point, I was baffled by how rhythm works. Maybe this was some long term passive-aggressive reaction to the family metronome.

Stu also gave me an insight into the nature of hearing. We were listening to some Jazz tune, I don't remember which one, and suddenly he said *Hey there's only 4 horns playing in this section. Usually there's five saxophones.* Stu was playing saxophone professionally with Walter Bridges Big Band at the time, whereas I considered myself fairly astute if I was able to distinguish between the sound of a trumpet and a saxophone, much less between sax and clari-

---

[5] Stu Fessant is a professional musician and producer in Portland, OR.

net, or trumpet and cornet. *Four horns instead of five playing?* How could he know that? The simple answer is *experience.*

As time went on and I listened and learned more, I began to discover the sorts of details that can be perceived which let me understand such subtleties. For me it boils down to listening carefully (and repeatedly), and paying close attention to details. An analysis of the intro to *Fever* recorded by Ray Charles will show what I mean. First I will talk about my research motivations, then some technical descriptions of the Science of sound.

Fever by Eddie Cooley and John Davenport is a classic R & B tune with a strong backbeat. Ray Charles' 2004 version with Natalie Cole is an extremely hip recording of a great song. While I was working on my Masters thesis research into Swing, I came across this recording and extracted some sample loops because I really like the feel of this version of the song. I didn't actually intend to use it as part of my analysis since I thought I should concentrate on Louis, Duke, Benny Goodman, Glenn Miller and other "greats" in the traditional Swing world. A diversion into R & B would be hard to justify, since listening to R & B often leads to Rock 'n Roll. I was in *academia* and everyone wanted me to act grown up. Rock 'n Roll wasn't even mentioned.

I was reading dozens of academic papers and Masters and PhD theses in the area of computer analysis of music, and the whole field seemed really scattered and incoherent to me. I had to make some sense of things so I could get on with some kind of legitimate research work of my own. I had heard that triplet eighth notes[6] were an important part of Swing, and some of the papers I read described technical analysis of this triplet subdivision feature, so I started with the goal to find these triplets.

Fortunately, I also had to contend with Brazilian *swinghee* as part of my research since that was my main interest in the first place. Nobody was paying me any money to go to graduate school, so I thought I would do whatever pleased me rather than follow some one else's agenda. But it's difficult to push through a Math and Computer Science thesis proposal by saying "Hey! I want to do *Samba* for my thesis" (much less Rock 'n Roll).

So I had to appear as if I was approaching Swing from a pretty straight point of view to sneak under the academic radar. I started with Duke and Louis and *It Don't Mean a Thing*. After a short time, I got my code working tolerably well, and discovered the existence of triplet notes in *It Don't Mean a Thing*. I also discovered the slight timing variation in the pulse of the beat, which to my knowledge had not yet been described in any of the research literature. Emboldened by the appearance of success, I moved onto the Brazilian *pandeiro* and discovered triplets in that rhythm, as well as a push-pull variation to the pulse, and a

---

[6] Chris Wood is the founder of the ***Samba Like it Hot!*** bateria in Ashland, Oregon..

very peculiar distribution of the other beats. See the section about MB notation at the end of this chapter for details on this topic.

I was on a roll! For each sample analyzed, I improved my code. I was limited to fairly simple mixes since neither my signal processing nor pattern recognition was very sophisticated. I had to work with music that had instruments with very distinct frequency ranges, such as bass drum and cymbal or snare drum. Trying to analyze a duet with two horns was definitely out of the question. Just for grins, and because I like the recording so well, I decided to analyze *Fever* mostly as an exercise for the code since I still didn't see how I was going to sneak R & B or Rock 'n Roll past the bouncer, oops I mean the thesis committee.

TIME, FREQUENCY AND SOUND WAVES

Frequency and Pitch. Sine waves. Additive synthesis of complex waveforms from simple waveforms. Introduction to Fourier analysis.

The sound we hear begins as vibrations in the outside world. Typical musical sounds start as vibrating strings, drum heads, hollow objects (horns, organ pipes), loudspeakers, human vocal tract and so on. The source object vibrations cause the air molecules to vibrate, and this air vibration is transmitted in all directions eventually reaching our ears. Then things get quite interesting. The vibrating air molecules create a complex system of echoes in the outer ear (pinna) and the shape and surface qualities of the ear canal bring the vibrations in to the ear drum which resonates in response to the air vibrations.

At this point the sound is converted from a multi-dimensional system to a one dimension system, i.e. 3 spatial directions {x, y, z}, are converted to a back and forth motion in the thin tissue of the eardrum. All of the complexities from the echoes, reflections and reverberation of the outside space (e.g. concert hall, restaurant or other room), plus the complexities added by the action of the pinna are encoded into mechanical motion of the eardrum, middle ear bones, and inner ear (cochlea). The motions of the eardrum and middle ear bones are probably not strictly one dimensional, but I have not found any research that investigates this detail of hearing. There is a general assumption in physics and audiology that the sound is one dimensional from the eardrum into the inner ear and brain. Since there are two ears, the term *binaural* is used. The brain somehow combines the information from both ears into a complex three dimensional form so we perceive sound direction, location and so on. The action of the cochlea is profoundly nonlinear and represents a far more complicated audio signal than one would see in a software application for sound editing.

The various types of motion of the tissues in the cochlea trigger reactions from nerve cells which encode the mechanical information into a tremendously multi-dimensional series of time correlated nerve impulses that travel into the audio cortex of our brain. In the audio cortex, the patterns of nerve impulses are organized into still other forms of information that eventually reaches our conscious mind. At this point we make connections with our symbolic knowledge and memories, and can decide things like "That's Louis Armstrong playing the trumpet." The entire process of human audio perception is almost instantaneous, typically being no more than a few milliseconds, although there may be some interesting exceptions.

The one dimensional audio signals are familiar to anyone who has edited music in software like GarageBand. For real world sound, the waveforms are very complex. Some sections of the waveform obviously differ from others, and these correspond to the musical features that we want to extract, such as drum beats. Separating the different types of waveforms from each other is a challenging task, but we have our ears to inspire our research work. This sophisticated arrangement of meat can create both our mental consciousness and the perception of sound that is so familiar. Being that similar arrangements of meat can be found in any bird or mammal, it is unlikely that the processing and pattern recognition uses scary symbolic math, as taught in school. With this in mind, I will describe some techniques that should be fairly simple if you grok the concept of the audio waveform.

Figure 1 shows a short section of a raw audio waveform, about 1/10 of a second – 100 milliseconds (thousandths of a second). This time interval is somewhat longer than a fingersnap and somewhat quicker than a short word like "to". During this time, existing harmonious sound from vocals and melodic instruments is suddenly drowned by a new note event which has a percussive attack from the guitar pick hitting the strings. Figure 1a shows 250 milliseconds (1/4 second) of the same note, and the onset event can be seen in context of the sustain tone following the onset. Figure 1b shows a closeup of the first half of figure 1, just at the very beginning of the transient which is the short, most complex little wiggle. You should carefully inspect the differences in complexity between the four sections of the sound: first fairly smooth (before the note), next a busy complicated section, and then a waveform whose complexity is somewhere in between the previous two. Finally, the waveform returns to its original form as the short burst of sound from the guitar fades, and the other sounds (which never disappeared) come to dominate the shape of the waveform again.

Audio waveforms represent the basic raw (unprocessed) data of musical audio. They are generally very complicated and obscure, difficult to directly use in a practical manner for extracting useful information from the music. A convenient technique to make the waveform complexity more tractable for pattern recognition is to break down the complex waveform into several simpler waveforms which add together to approximate the original audio waveform. This is no more tricky than saying that you can make 11 by adding 4 and 7. The only minor complication is that there are millions of numbers in an audio sample. For each number in the audio sample, we add together dozens perhaps hundreds of numbers, one from each simple waveform in the set of *Fourier* components. Figure 2 shows some *sine* waves which are often considered the “simplest” waveform, although some other waveforms are equally simple, such as the square wave or triangle wave, and could be used equally well.

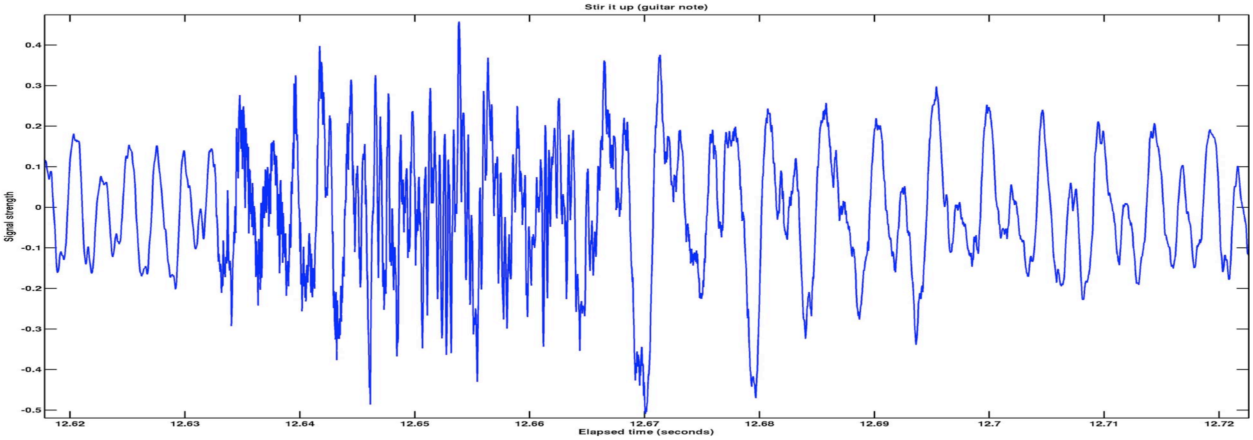

Figure 1. Audio waveform showing harmonic sound giving way to a guitar note event which has a dense **onset** *followed by a less chaotic* **sustain** *section (from **Stir it up** by Bob Marley and the Wailers)*

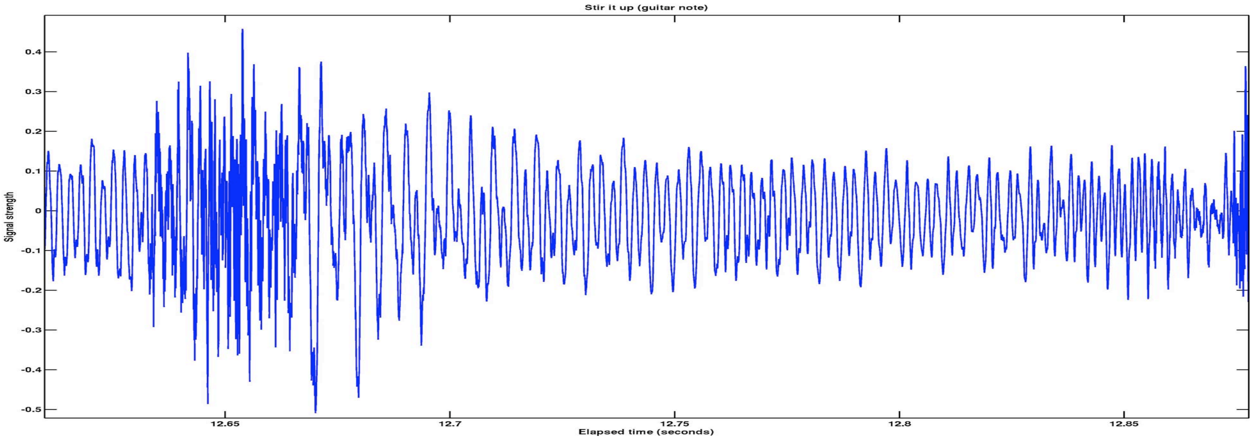

Figure 1a. Audio waveform zoomed out from figure 1. The onset transient is seen in context of the sustain section. The sustain following the onset (the guitar pick striking guitar strings) is several times longer than the onset. Another note event onset is visible on the right side of the graph.

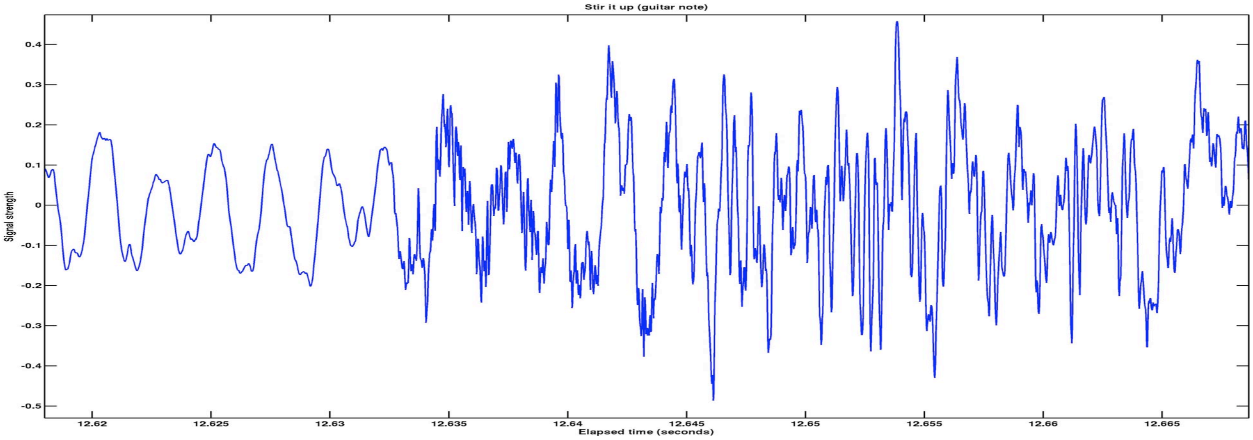

Figure 1b. Closeup of the guitar onset transient waveform. The waveform leading into the event is relatively less complex than the onset event (middle 1/3 of window), or the waveform following the onset.

Simple waveforms are primarily described by their *frequency* (how fast they change), *amplitude* (loudness) and *shape* (e.g. sine, square, triangle). In figure 2, sine waves are stacked from low frequency to higher frequencies. The waveform at the top of the plot is the sum of all seven sine waves. While the sine waves all have the same shape and amplitude, they differ in frequency. The composite waveform, shown in closeup in Figure 2a, has a fundamentally different quality from the sine waves: its shape is more complex than any of the simple waveforms. The shape of the top waveform in figure 2 is simply the result of adding all seven sine waves together, point by point. The additive complexity in this example is the basis for most of the analysis of musical waveforms.

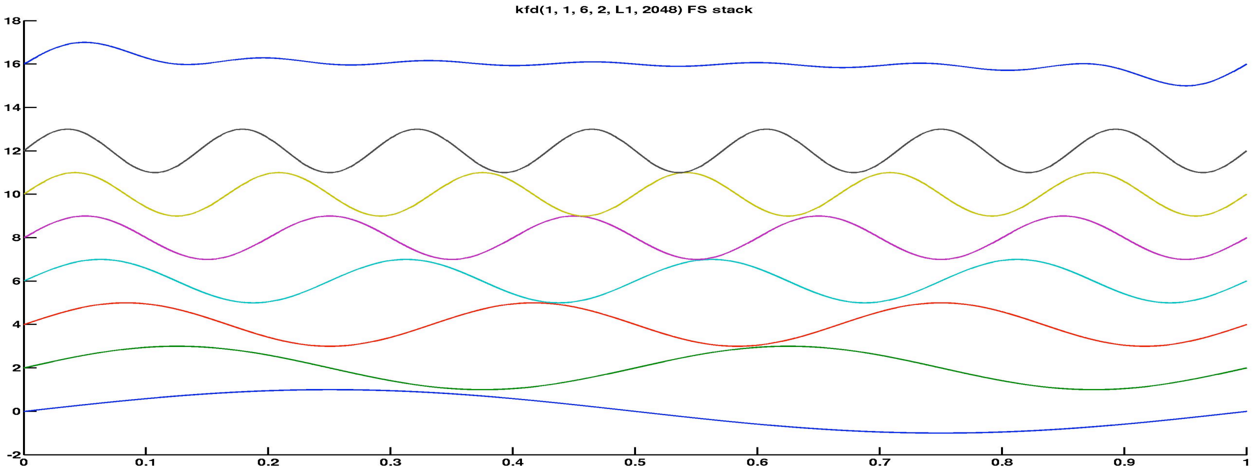

Figure 2. Several simple sine waves of different frequencies (below) and the composite waveform generated by adding all the simple waveforms together (above).

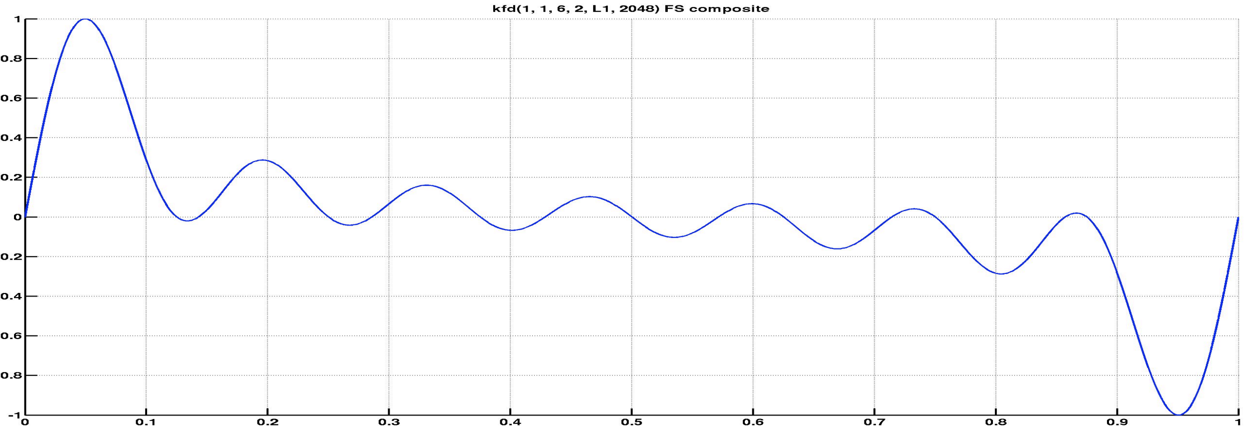

Figure 2a. Closeup of composite waveform from figure 2, showing more detail in the shape.

Figure 2b shows another additive waveform generated by a number of sine waves. Features exist at several scales. Like figure 2a, there is a single large feature spanning the entire figure. This low frequency feature would repeat if we zoomed out the view. Higher frequency details are visible as small scale wavelets superimposed on the overall waveform. Mid frequency waves also exist in the waveform, more visible perhaps by slightly defocusing your eyes, or by ignoring the detailed view of the local shape of the wave, scanning your eyes quickly from side to side, and letting the overall visual flow of the peaks and valleys of the waveform take over your perception. While they can be a bit tricky to see directly, these frequency features are actually present, both in the shape of the waveform, and in the component sine waves. Similarly, figure 2a shows a dominant downwards trend from left to right.

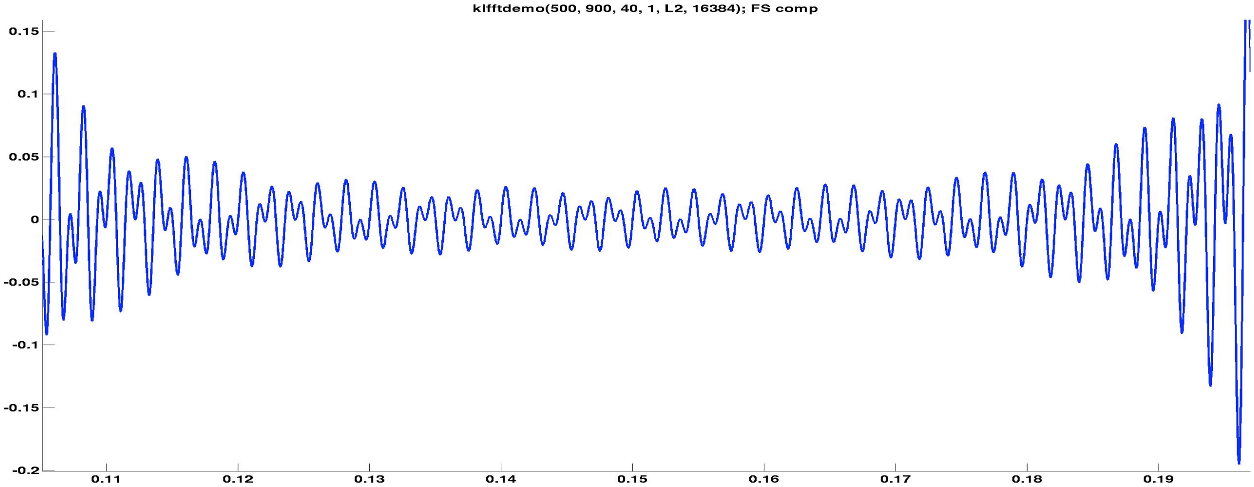

Figure 2b. A more complex waveform generated by adding sine waves together.

Compare the waveforms in figure 2b and figure 3. The first is completely synthetic, generated by adding sine waves together as described for figures 2 and 2a. Figure 3 is a real musical waveform, specifically a trumpet note played by Louis Armstrong in *It Don't Mean a Thing (if it ain't got that Swing)*. Although the two waveforms are fairly different from each other, nonetheless there are similarities which are easily observed by looking at the detailed shapes of the waveforms. Fourier analysis generates a set of sine waves which can be added together to mimic any part of the trumpet waveform. The concept is simple, the devil is in the details. If you are unfamiliar with the math behind this, but are willing to accept my word that a suitable set of sine waves can be found, then I won't make you go read a book on differential equations. You doubters can find the book in the bibliography and start studying.

Figures 3a and 3b show the same trumpet note over longer periods of time. The evolution of the sound is apparent. Initially, there are dominant mid frequencies (large spike shapes). After a few cycles, the waveforms become more ragged as small amplitude higher frequencies are added in. Finally in figure 3b, a dominating low frequency waveform emerges about halfway through the diagram. In all three images, the mid frequency spikes maintain a consistent presence. These are the sorts of details that we need to extract in order to analyze music.

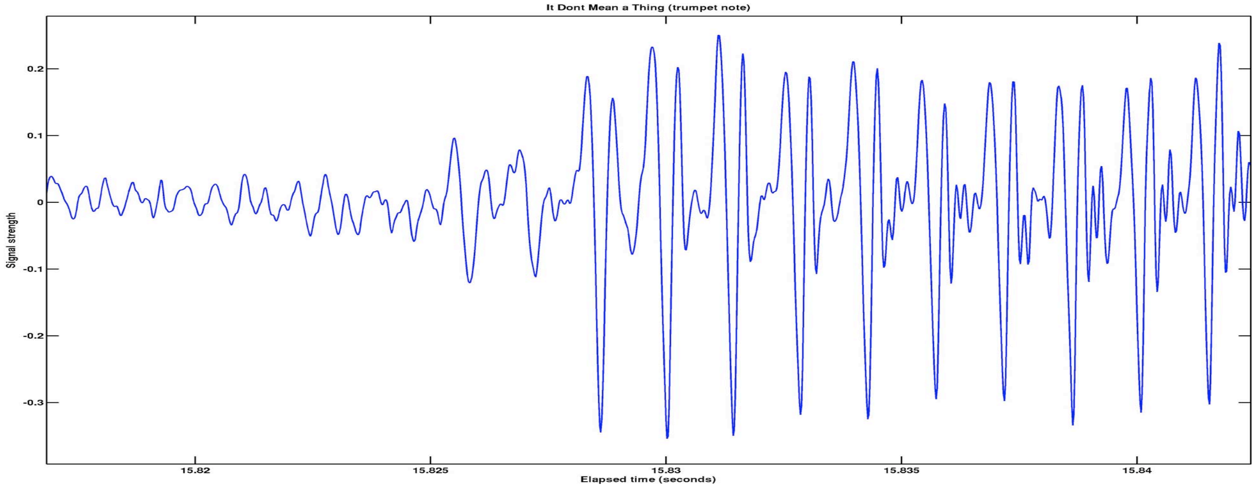

Figure 3. Close view of start of the waveform of Louis Armstrong playing a trumpet note.

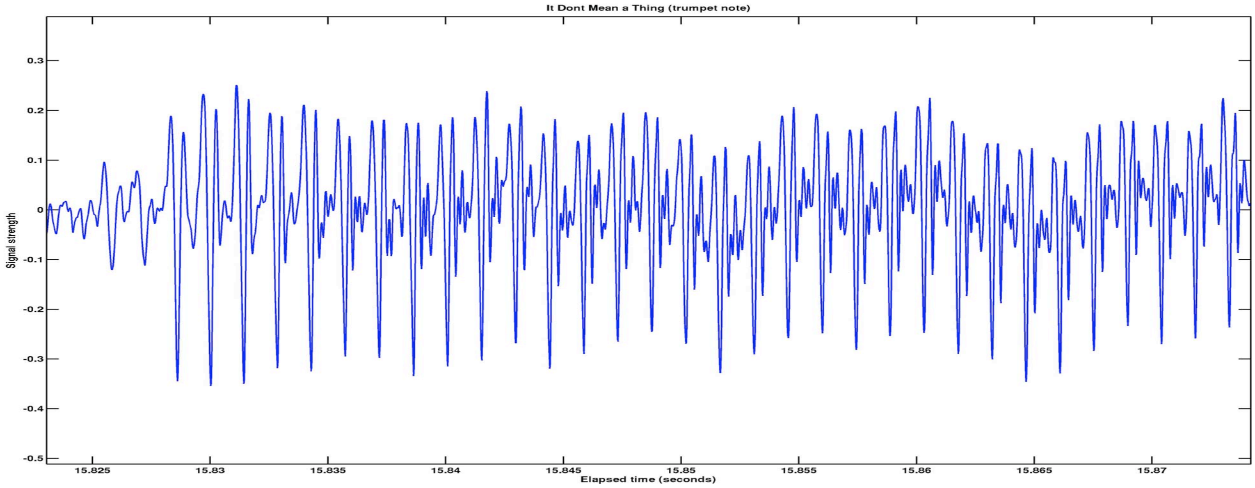

Figure 3a. Zoomed out view of trumpet note, showing emergence of high frequency waves.

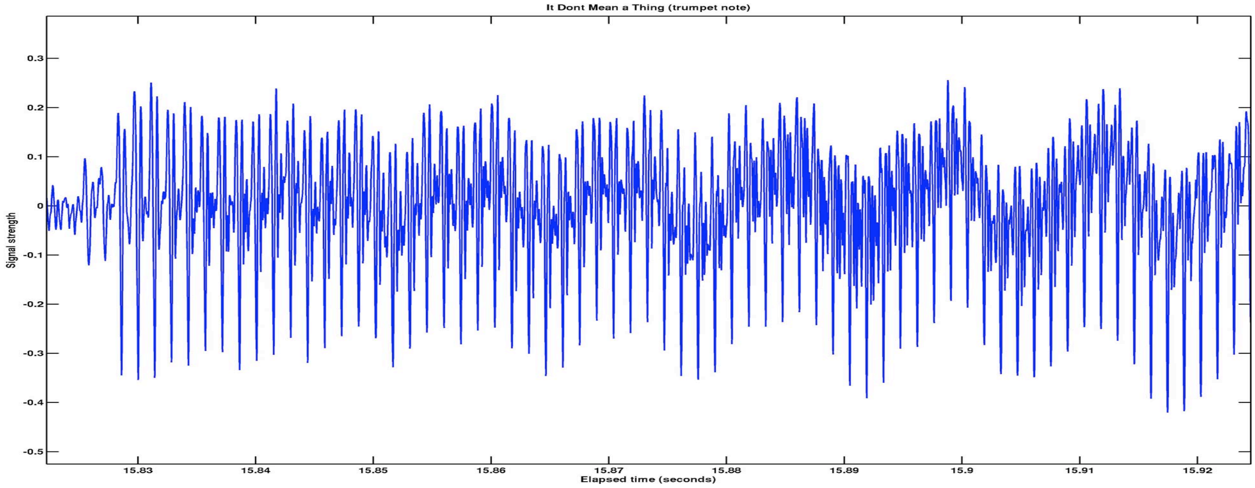

Figure 3b. Trumpet note zoomed even further out, initial 100 milliseconds (1/10 second). High frequency waves (brassy sound) and mid frequencies (spikes) have an underlying low frequency wave.

This exercise is intended to start you thinking about how waveform shape and complexity relates to the sound qualities that we hear, and how our ear and audio cortex in the brain manages to make sense of the huge amount of raw data streaming through. In fact, when we analyze sound, we don't use the straightforward approach of comparing shapes of audio waveforms. Instead we transform the raw data into other forms that are easier to handle, such as breaking down a complex waveform into a set of sine waves, which is the process of Fourier analysis. The result of the Fourier transformation of an audio waveform is called the *spectrum* of the waveform. The sine waves generated by Fourier analysis is the set of *harmonic components* for the original waveform. Just as an audio waveform can be seen as a long series of ordinary numbers which are plotted to yield the waveform, the set of harmonic Fourier components can be seen as a collection of simple objects which, taken together, give a direct view of the spectrum of the sound, just as the audio waveform is a direct view of the sound itself.

COMPUTER ANALYSIS OF MUSIC

How to extract musical note features by selecting frequency ranges where the main part of an instrument's sound exists. Examples of the spectra of several instruments, and vocals. Spectral Graphs, Time Series Graphs, Diffdot Plots.

The waveforms shown in the previous section can be broken down into component sine waves using Fourier analysis. The Fourier view is called the *frequency* domain, whereas the original audio waveforms represent the *time* domain. These are not the only two ways of looking at sound, but both are useful. Together they supply a set of tools for doing more interesting things such as finding where, in all this technical complexity of data, the simple intuitive human aspect that we call *swing* can be found. Figure 4 shows a musical sample that includes the main elemental features that we will look at throughout this book: rhythm, melody and *timbre* (the unique sound quality of an instrument or vocals).

As the sound of the music changes in time, the spectrum also changes. Looking at the way that the spectrum changes in time provides a fairly natural method of analysis to extract information features about each musical note event as it happens. Each note event is also associated with a time point in the audio sample, making it easy to describe the rhythm by listing the sequence of note events and the time differences between them.

The waveform graphs shown earlier are neither easy to work with, nor efficient forms to use for summarizing the audio characteristics of a musical sample. Figure 4 introduces a useful, compact visual form called a *spectrogram* which is a graph of the spectral changes over time for an audio sample. This one multi-colored image summarizes data which would take at least 32 waveform graphs like those presented earlier. The waveform graphs in turn, are more efficient efficient summaries of the long list of numbers stored in the computer that represent the sound. These numbers are sent from computer memory to the audio output port and then to speakers so you can hear them. There are over half a *million* numbers needed for the audio sample that generated figure 4. It would take around 1000 pages of text to list all the numbers that make up the musical data in this 14 second sample. This picture is worth quite a bit more than a thousand words.

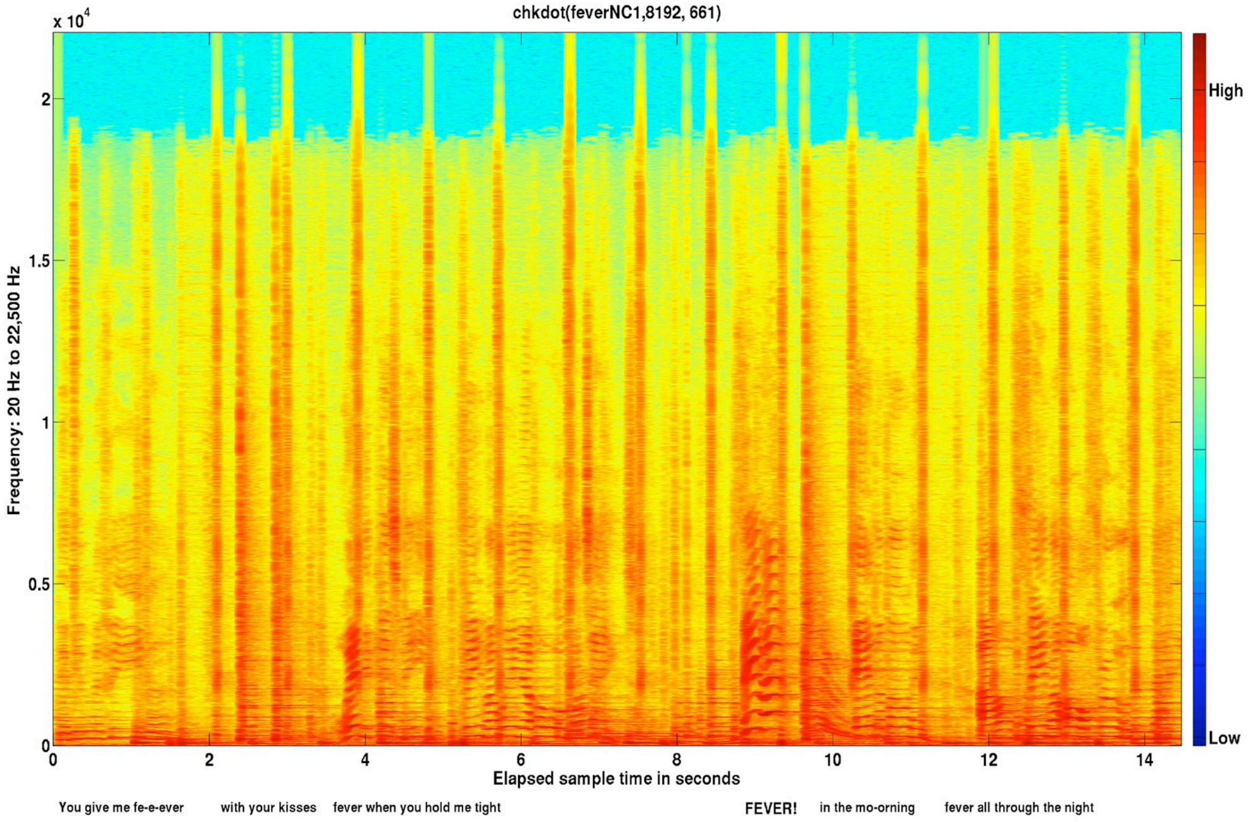

Figure 4. Spectrogram for Natalie Cole singing **Fever** *(2004).*

In figure 4, time is along the bottom and frequency is plotted vertically. The blue "sky" at the top is caused by audio recording limitations: typically the frequencies above 20 KHz (20,000 cycles per second) are not recorded because the "official" range of normal human hearing is from 20 Hz to 20 KHz. This viewpoint is not entirely correct, as evidenced by the fact that many professional digital recording facilities in the 1990's began using frequency ranges up to 192 KHz. Musicians and producers *could* hear the difference between the standard recording range and the extended high frequency range, even though audiology researchers have for decades measured the range of human hearing as being limited by 20 KHz. Nonetheless, there is more to the hearing process than meets the eye. More research is needed to try to find what subtle and interesting aspects of human hearing are involved in the perception of sound above the frequency range of normal hearing. Suffice to say for our purposes that such interesting, undiscovered aspects of human perception *do* exist.

Notice that there are several yellow/red spikes extending up into the blue sky in figure 4. Most of these are Ray Charles' fingersnaps, and a few are from conga beats. The excursions above 20 KHz actually exist in the real world, although they may not "actually" be recorded by the recording process. My current opinion is that these visual artifacts in the spectrogram are caused by limitations in the digital form of the musical audio data. These limits are related to deficiencies in information theory and the Shannon/Nyquist sampling theorem. Discussion of this topic is technically quite advanced, and I won't go into it in this book except as it might affect our real work: *finding ways to describe Swing*.

In the lower half of figure 4, there are several groups of correlated wavy red lines. These are Natalie Cole singing. I include the lyrics below the image so you can get a sense of the connection between the words and the audio features in the music sample. At the very bottom of the picture is a concentration of red which represents most of the sounds of the bass, drums, piano and so on. Fourier analysis has limitations which make this part of the spectrum difficult to resolve, somewhat like looking at Jupiter or Saturn through a toy telescope – you might see the moons and rings if you are lucky, but you won't get a very clear view of them. One of my current research interests is improving the resolution of the low frequencies in order to better distinguish the many sounds that make a complex music mix.

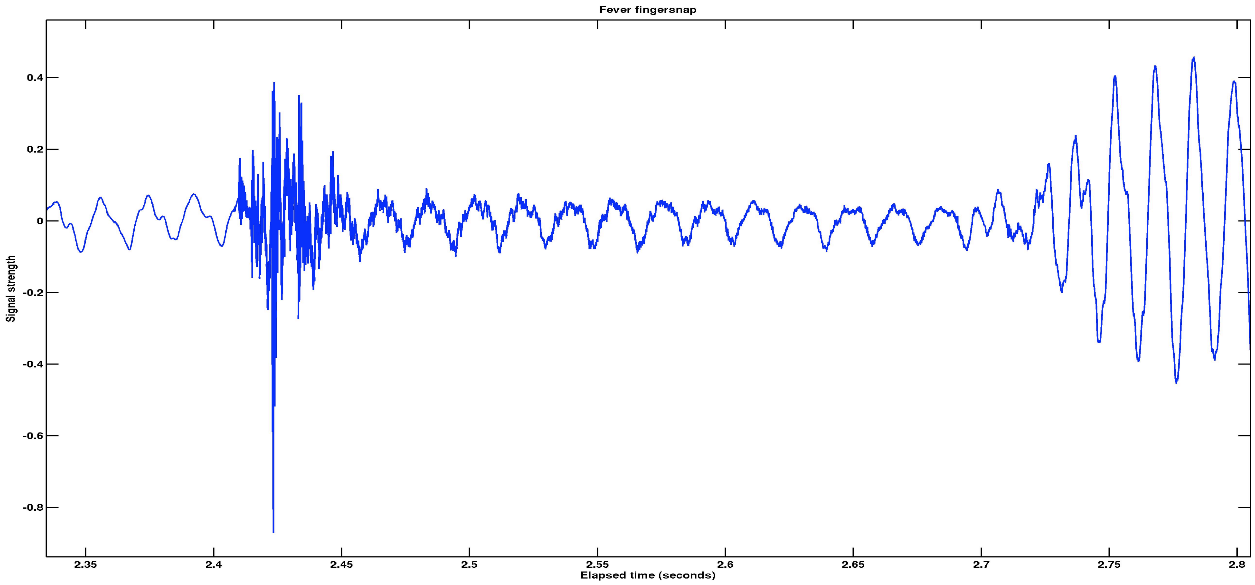

Figure 5. Ray Charles' fingersnap, followed by a conga beat.

Figure 5 shows the audio waveform for a fingersnap in the recording. The waveform to the left of the snap is primarily low frequencies from the bass guitar. The note at the right of the plot is a beat on the conga. The fingersnap itself is a very busy part of the waveform, indicating a preponderance of high frequencies. The fuzziness extending to the right of the main snap are echoes from the reverb in the mix, superimposed on the waveform of the bass guitar. In this graph, the differences between the three types of notes are quite obvious.

What is going on with the spectrum of this waveform? Figure 5a shows a spectrogram that has a conga beat, fingersnap and a second conga beat (with a different sound) close together. This is not the same part of the musical sample as figure 5, but the note events are equivalent. Figure 5a also has two note events which overlap slightly in time. This will show how to distinguish several simultaneous note events by their different spectral qualities. The conga beat on the left, being alone, is easy to distinguish. It is an *open* tone, and has a spectrum with a clearly structured set of parallel red lines that represent dominant harmonics in the set of Fourier coefficients. The red feature on the right is a composite of *two* note events: first the second conga note followed by a heavy mass of red which is the fingersnap. The second conga beat, a *closed slap,* has a different timbre from the first conga beat. The closed slap has less harmonic content, although there is some tone which is indicated by the broader, more separate swatches of red in its spectrum. There are fewer harmonics, and they smear into the spectrum of the fingersnap. This shows a limitation of Fourier analysis when analyzing complex sounds with multiple sources. Happily, this does not prevent us from finding the beats and extracting their timing so we can analyze the Swing.

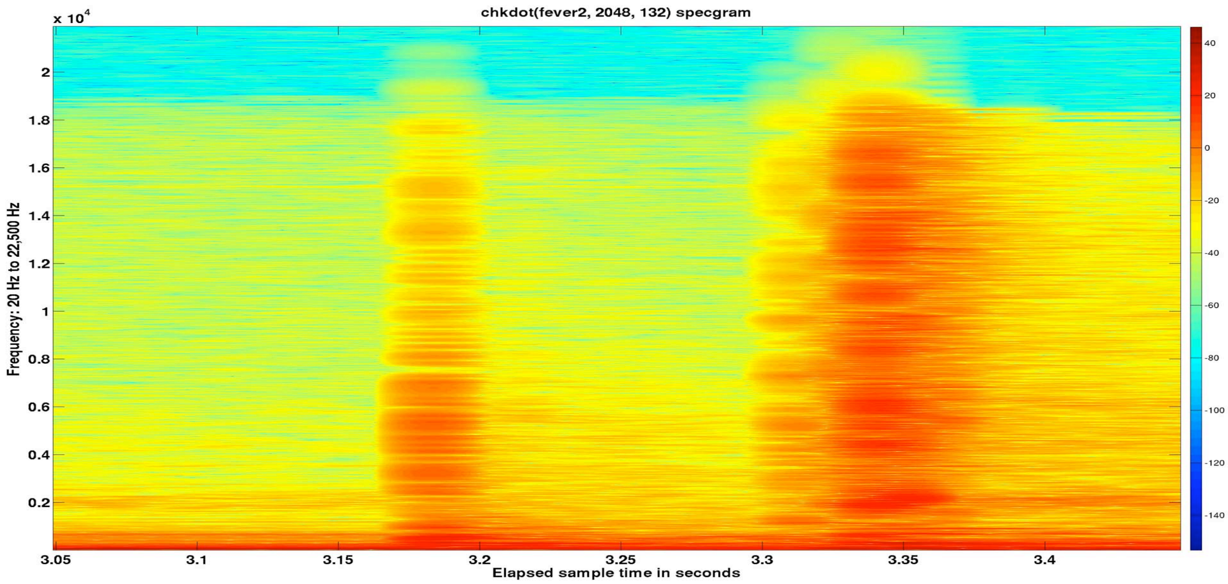

Figure 5a. Spectrogram of conga beat followed by a second conga beat and fingersnap.

Figure 5b shows a different section of the spectrogram which has more note events. On the left is another open conga note, with its familiar harmonic lines. In the middle are four closely spaced beats from the snare drum. These have a less harmonic tone than the open conga beat, but some harmonic lines are visible in the spectrum. There is also more random looking scratchy stuff going on in these note events. This corresponds to the rattle of the snare coil on the bottom head of the drum, which vibrates briefly against the bottom head after a stick strikes the top head. Finally there is a fingersnap. This fingersnap also has a conga note buried in it, although it is difficult to be sure from looking at the spectrum.

The conga beat is more obvious when listening to the audio sample. In the spectrum, the main indicator of the conga note is the slight gaps (yellow bands) in the mass of red.

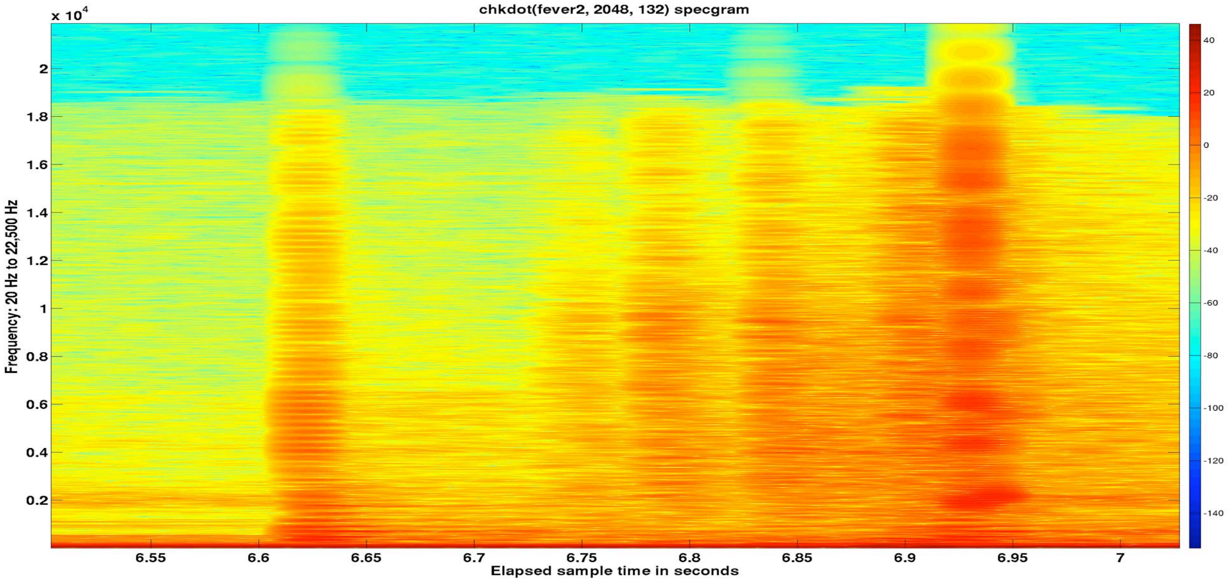

Figure 5b. Closeup of conga beat, snare drum roll, and fingersnap (6 beats).

Figure 5c shows a zoomed out view of figure 5b. Three fingersnap events are visible, as well as several conga notes, the snare drum roll, and other snare drum beats. This example shows how notes played by different instruments are typically not perfectly synchronized to the exact same time location. The fingersnaps in the middle and on the right also have a snare drum note associated with the beat time, but the snare drum is slightly ahead of the fingersnap. The note event exactly between the two fingersnaps is less red, indicating that the high frequencies are not as loud for this note event as for the fingersnap. The middle event is also a composite of two beats: one from the conga and one from the snare drum. In all three cases, you can see that the beats from different sources do not happen at exactly the same time. Keep in mind that this time difference is very short, approximately the length of time as the fingersnap itself. I'll say more about this in couple of pages.

In music played by a computer sequencer, all the beats are usually locked to the same time location. This detail is one of the differences that give music played by humans a less mechanical quality than music played by a computer.

Figure 5d shows a different view of the information from figure 5c. Figure 5d is similar to the time series plots of the waveforms which were shown earlier. The time range of figure 5d is twice as long as for 5c. Figure 5c maps to the first half of the graph in figure 5d. The top row of spikes with red diamonds is the graph of fingersnap events. The bottom row is conga events. On the top two rows of the plot are small bumps to the left of the second snap. These are the conga beat and snare drum roll shown in the previous spectrograms.

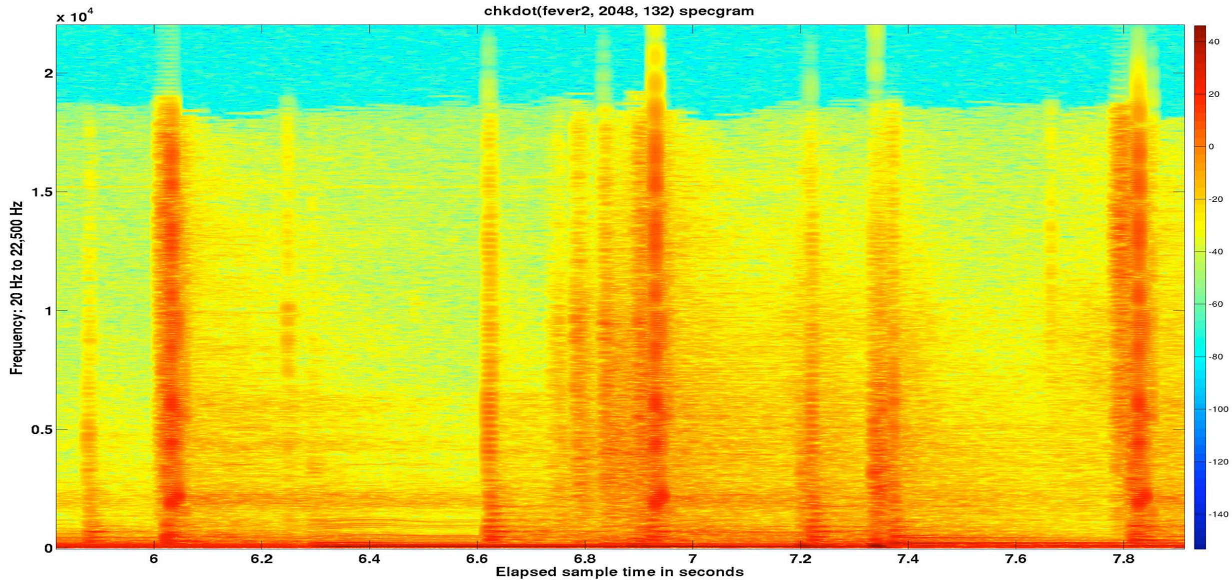

Figure 5c. Context for six beats.

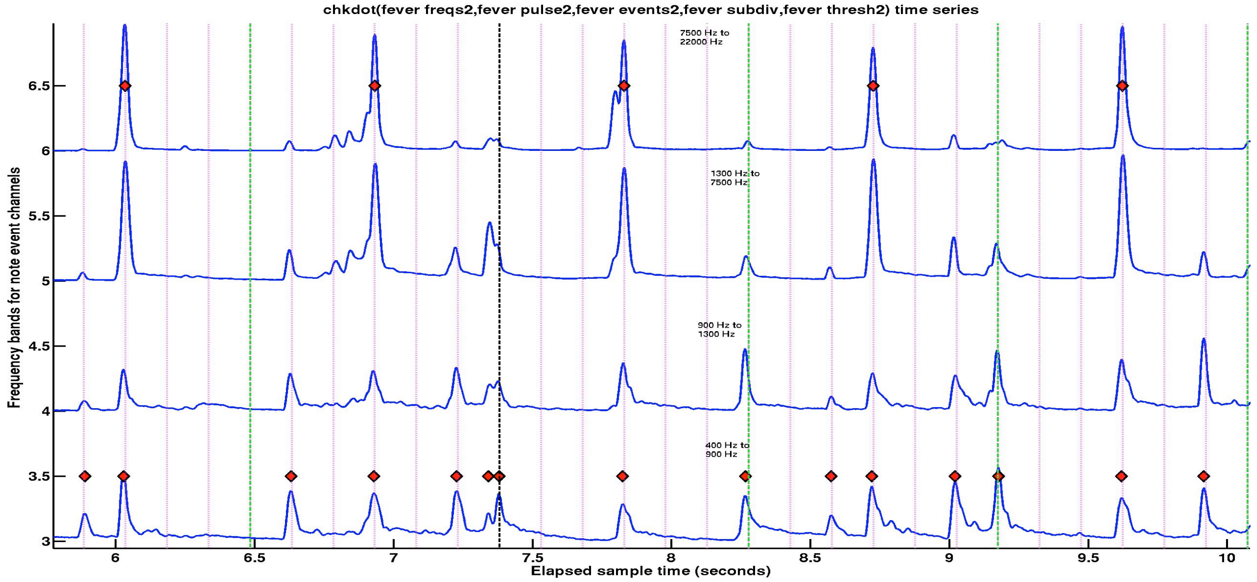

Figure 5d. Time series plot by frequency of six beat context, with "error".

Unlike the raw waveforms, the `chkdot` plot in figure 5d shows the change over time of the *power* or volume of sound for several different frequency ranges. The lower time series line is for the frequency range 400 Hz to 900 Hz, which is where the sound of the conga is the most identifiable. The top line is 7500 Hz to 22,000 Hz, which is where the fingersnaps are most identifiable. The higher frequencies also have a useful amount of information about the snare drum, as seen in the three little bumps to the left of the second fingersnap. The first bump is also mirrored in the lower frequencies because it is a conga note rather than snare drum. There is also a slight bump on the left side of the second fingersnap which is the third beat of the snare drum roll. The third fingersnap has a more noticeable bump to the left which is also a snare drum beat. The fourth and fifth fingersnaps have no such extra bumps, although in the audio sample, the snare drum beats are audible on the same beats as these two fingersnaps. On snap 2 and 3, the snare drum is very slightly faster than the beat of the fingersnap. On snap 4 and 5, the snare drum is locked onto the same time as the fingersnap. The event between the fingersnaps (at the black line) is the third "double" beat in figure 5c where the two note events (conga and snare) are not exactly synchronized.

What this means is that when the snare drum comes in, it pushes the beat a little for the first three beats, then lays into the groove set by the fingersnaps. This sort of technique is used in many tunes to give a little extra pep on certain beats. The loudness of the accent on a note event also influences the peppy feeling, but for the moment we are focusing strictly on the timing of the note events.

Figure 5e shows an extreme closeup of the snare drum rhythmic expression which is next to the middle fingersnap in figure 5d. The smaller middle bump in 5e is probably the conga. The slight bump on the right side of the fingersnap is apparently part of the snap sound since it appears in a very similar form on all `chkdot` fingersnap waveforms inspected. The bump on the far right is probably the bass drum. The time spanned by figure 5e is about 100 milliseconds, 1/10 of a second. Now you know how much time a fingersnap takes.

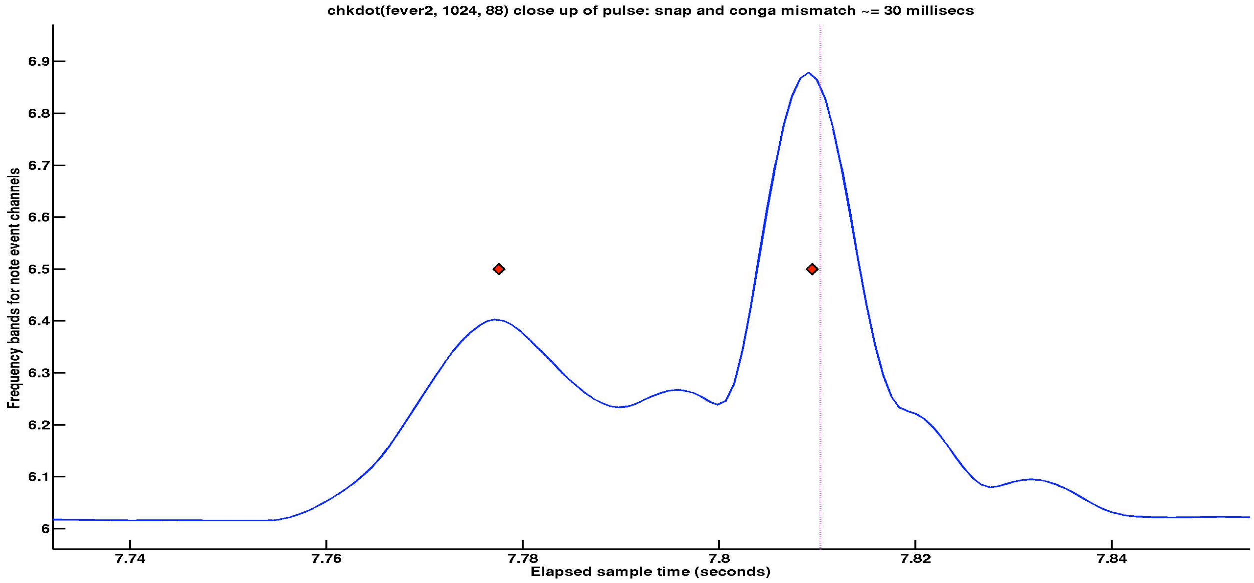

Figure 5e. 30 millisecond "push" by snare drum (left) with fingersnap (right).

The red diamonds in figure 5d and 5e mark the point in time (on the X axis) where the note event "happens." Obviously, by looking at the original waveforms or the spectrograms, note events (like everything in life) are not *instantaneous*. Rather, they are spread out over a certain amount of time, and most of them have a beginning, middle and end. Thus it is a bit arbitrary to chose what point on the `chkdot` waveform corresponds to what time our brains think is when the note event happened. Some people might engage in lengthy philosophical debate over this issue, but I just decided to pick the peak of the waveform until I know better. Knowing the times when all these red diamonds happened, we can now extract information about the actual rhythm. This takes us one step closer to Swing.

Look closely at the subdivision in figure 5d. This is marked by vertical black, green and pink lines. We know by listening to the recording that the fingersnaps are all on the *backbeats,* so we mark the downbeats with green and black lines so that the fingersnaps are exactly in the middle. Since we are looking for triplet notes, we divide the space between downbeats by 6. Standard MB subdivision would be limited to 2, 4, 8 or 16 (or 32 …). Since 2 sets of triplets equals 6 note events, the subdivision by 6 in the chkdot diagram gives us the triplet equivalent of the MB subdivision framework but not limited to multiples of 2. I had discovered this much when I analyzed *It Don't Mean a Thing*.

Here is the point where the strong backbeat of R & B turns into a variation of classic Jazz swing. The conga beats' time locations are exactly described by the subdivision by 6. If you look in figure 5d, you will see a conga beat on each of the pink lines except position #5.

For reference on counting: Pink line #1 is on the downbeat lines (black and green). Then the pink line numbers go 2, 3, 4, 5, and 6 being just before the next green or black line.

This discovery caused no little amazement to me. Of course, a good musician probably would have listened to the music loop about ten times and then they could tell you what was going on, assuming they understand Swing. Classical musicians might have more trouble than pop or jazz musicians, but I haven't asked any. The upshot was that, having discovered triplet notes, I could now declare that *Fever* and by extension R & B can be called Swing.

To finish this introduction to my analysis process, I will introduce a form of notation which precisely describes the timing of the real musical note events that are detected in the `chkdot` diagrams. In the case of *Fever*, the conga and fingersnaps are very close to the "ideal" meter of MB notation, which has no push-pull to the pulse. *Fever* has a type of rhythmic style usually described as "tight." Many swing tunes are played with a push-pull on the pulse and/or other rhythmic expressions which may give them a "looser" rhythmic feel.

Since the fingersnaps mark the beat in such a regular fashion, it works best for my current computer code to use them as the pulse. The standard MB approach would see the downbeats as the pulse. The backbeat is just the opposite of the downbeat, so it really doesn't matter which is chosen, unless one is more convenient for some other reason (such as making the code easier to write: the bass guitar and drums play some standard downbeat notes, but sparsely, and the resulting rhythm is not regular enough to use as pulse).

Figure 6 introduces the `diffdot` diagram. Elapsed sample time is along the bottom axis, as in the spectrograms and `chkdot` diagrams. Vertically we display the time difference from one note to the next for the note events marked on each `chkdot` line graph. Events from the `chkdot` diagrams are transferred to the `diffdot` diagrams one for one. At the top of figure 6 are the fingersnap note events (green dots and lines). In the lower half of the diagram are the conga notes, and maybe a few snare drum notes (red dots). The pulse (fingersnaps) is designated by the light blue line across the top. The other horizontal blue lines correspond to subdivision of the pulse by 1/2, 1/3, 1/4, 1/6 and 1/8. Thus, note events which land on 1/2, 1/4 or 1/8 would be the standard MB half, quarter and eighth notes, and the 1/3 or 1/6 are the triplet notes.

Notice that most of the conga notes (red circles) land on either the 1/2, which is the backbeat (1/2 the pulse), or the 1/3 and 1/6 lines (triplets). There are no note events on the 1/4 line indicating that there are not any quarter note subdivisions to the conga rhythm. There are a few stray note events that don't fall on any of these "standard" non-MB subdivisions, and also some notes which are close to the subdivision lines but not exactly on. This is another indication of slight "errors" in the playing of the rhythm (or errors in detecting the exact time location of the note events). The slight deviations can also be deliberate rhythmic expression, as in the snare drum pushing the beat described earlier.. In other tunes I'll analyze later, there will be a wide variety of subdivision and patterns of note events that may not fall exactly on a "standard" time location This can be a loose rhythmic style or some other form of rhythmic expression.

`Diffdot` plots appear to be far more capable at describing the real features of rhythm than is MB notation. I am still thinking about ways to make them more useful, both for analyzing rhythm and for notating music as an alternative to tablature sheet music. Stay tuned for updates at my website www.tlafx.com .

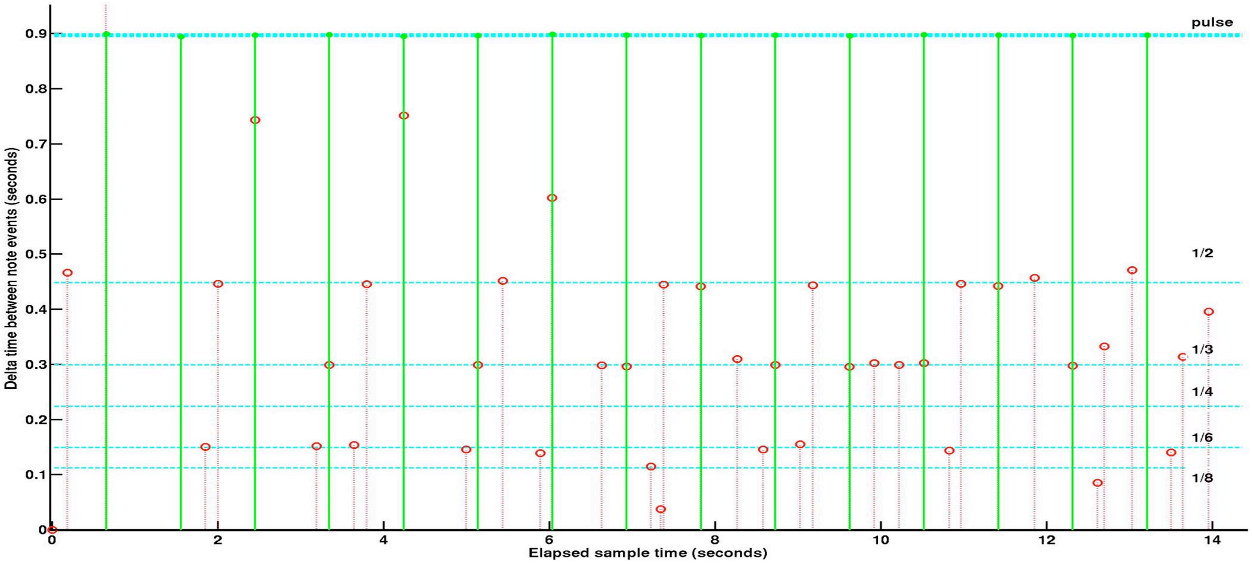

Figure 6. `Diffdot` *diagram of note events for* ***Fever****.*

If you look closely at figure 6 you can see that the pulse has more than one line. Since the pulse is not perfectly regular in most tunes, I adopted the notion of marking the quickest (min), longest (max) and average (mean) time difference between note events. I use the average as the "pulse" for subdividing. Figure 6a shows a closeup of the pulse for *Fever*. This is a direct measure of the variation in timing of Ray Charles' fingersnaps. As measured, the snaps are all within 5 milliseconds of each other, which means that they are each less than 2.5 milliseconds faster or slower than the "ideal" MB metronome beat. Since Ray was a musician and not a computer geek, I think it is safe to assume that he just recorded the track in realtime rather than setting up some kind of loop as many DJs or garage and basement music producers do these days. Similarly, we can assume the conga, drums etc. are played in realtime and not as computer loops.

It is not clear to me if Ray's fingersnaps are following some kind of intentional pattern, or if the slight variations in timing are basically random or stochastic in nature. Perhaps with improvements to my computer code I will be able to do a more detailed and accurate analysis of this recording. Until then, I will assume there is more here than I can measure. If need be, I will speculate (perhaps wildly) about what this particular Swing means.

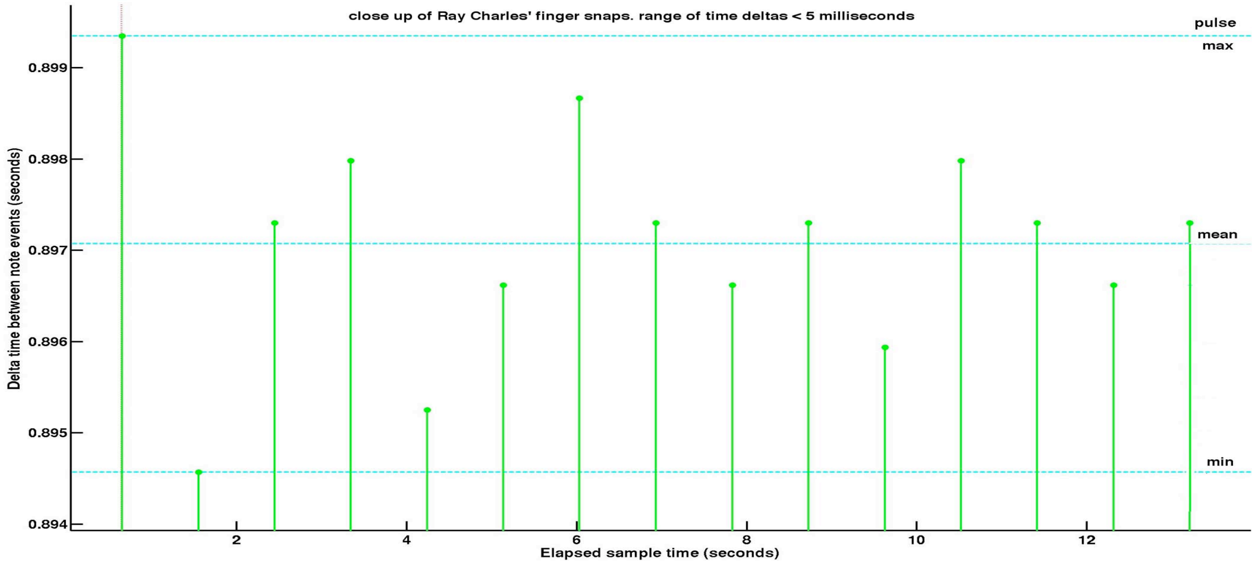

Figure 6a. Closeup of `diffdot` *pulse for* ***Fever***.

Finally, although it may not play a big role in this book, here is an example that shows more details of melody, and a comparison between a human voice (Natalie Cole) and a musical instrument. Figure 7 shows a fragment of the vocals from *Fever* on the left, and a section of Louis Armstrong's trumpet solo in *It Don't Mean a Thing* on the right.

The timbre (sound quality) is described by the patterns of correlated red lines in each image. Fluctuations in pitch in the melody correspond to the sets of red lines waving up and down in the image as time passes. The trumpet seems to be more constrained to specific frequencies (note pitch) than is the vocal, which make sense when we consider the physics of each sound producer. The trumpet has valves that change the length of the resonant cavity in discrete steps. This supports the playing of a limited set of pitches, with the possibility of bending the pitch slightly. The vocal is created by the elasticity of the throat which can take on a smooth range of resonant shapes, and hence smooth slides between pitch..

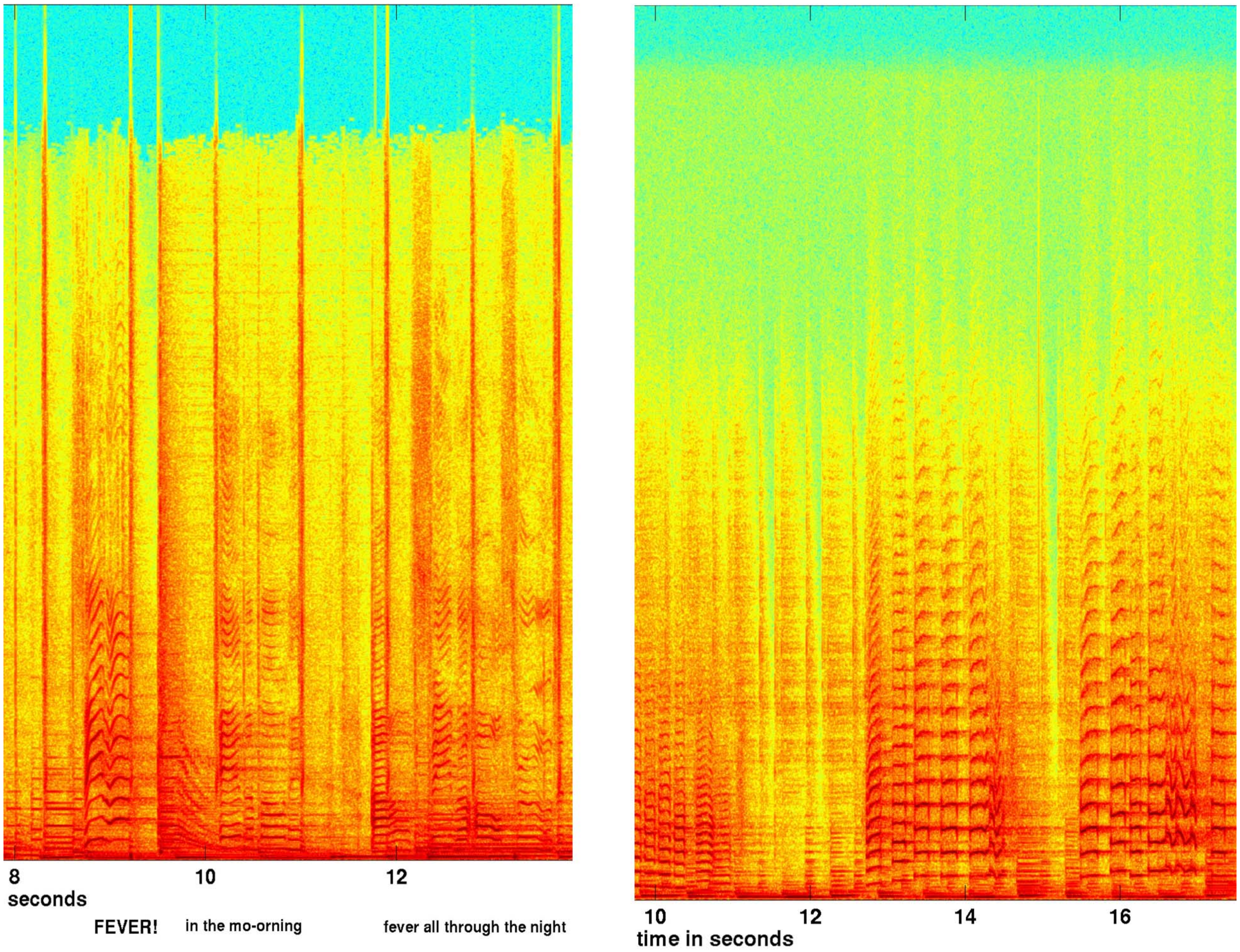

*Figure 7. Comparison of vocal music (left, Natalie Cole singing **Fever**) and instrumental (right, Louis Armstrong playing **It Don't Mean a Thing (if it ain't got that Swing)**. Melody, rhythm and timbre are all quite clear in the visual features of both of these images.*

Aside: Learning a Caixa Batida, and Perception of Time

Playing Samba batucada is like standing between two freight trains passing each other at 70 miles per hour. The immense sound volume seems to affect the speed of auditory perception. Visual perception is apparently not affected.

At California Brazil Camp 2005, I played under the direction of Marcio Peeter. The caixa (Brazilian snare drum) batida that Marcio taught is deceptively simple (see Figure 8), but I found it very difficult to play cleanly. Also, I experienced a unique perceptual phenomenon while learning this rhythm. On the first day of class, playing the rhythm itself and hearing the beats in time with the other drums was difficult. It was made more difficult because there was a temporal mismatch between my hearing a beat event, and watching the same event played by a fellow student (Topher) who knew the rhythm and played it well. Every time I heard the beat, I saw Topher's drumstick at the top of the hand motion rather than being down at the drum head. The sound of the beat, of course, is generated at the point in time and space when the stick hits the drum head. I was standing about 3 to 4 feet away from Topher which corresponds to about 3 milliseconds of delay from the propagation time of sound waves. The beats in the rhythmic pattern are separated by about 100 milliseconds – 8 beats per pattern and approximately 1 pattern per second. Thus the sound propagation delay was vastly quicker than the time between rhythmic events and contributed at most a negligible amount to my time perception experience.

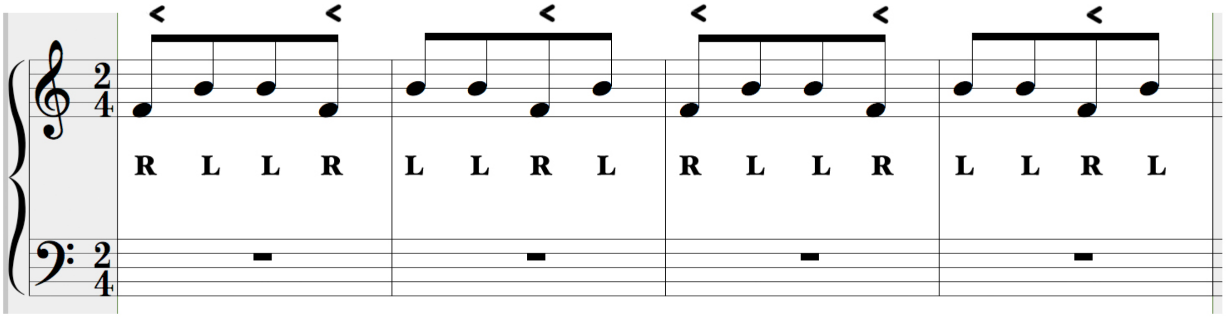

Figure 8. Straight notation for Ile Aye caixa batida. Note: the tablature is used for convenience only, to distinguish between two types of caixa beats, not to imply that the pitch of the beat is F or B.

Over the course of several days, my perception of these two sources of observing the beat events slowly became synchronized as I learned the rhythm better and became accustomed to the enormous sound volume of the bateria. For those unfamiliar with Brazilian batucada (Samba music played by an army of drummers, the bateria), it is at least as loud as standing between two train tracks with freight trains rushing by at high speed a few feet from your ears. Of course it is a good idea to wear earplugs. Eventually the visual and auditory inputs were closely synchronized in my perception. I read recently that the visual processing takes about 100 milliseconds from when an image is projected onto the retina until the visual cortex lights up with the neural activity corresponding to the action of perception. No doubt there is some delay with the audio cortex pathways as well [** look this up **] . Matching these two delays somehow was disrupted during my learning experience.

Part of the cause of my perceptual synchronization mismatch may be due to a particular physiological response in the hearing system. The front end audio processing of the inner ear develops noise canceling internal signals that can reduce the apparent sound volume in the audio cortex. This noise canceling effect is perhaps related to the tinnitus condition: e.g., I have a chronic high frequency internalized sound which is matched to the horizontal scan frequency of television: a legacy of watching too much TV too closely on a noisy set as a child. I also develop such ringing at other pitches after working for a few hours in front of a computer. The ringing sound is different, depending on which computer or other machine I am near for extended periods of time. These sounds generated by the ear have been measured by researchers using small microphones placed in the pinna (outer ear) and are known as *oto-acoustic* emissions. A similar effect in the visual system can produce a 3D effect from a single image (e.g. photograph) when one eye is covered with a dark lens while the other is not. This can be done by popping out one lens from a pair of sunglasses. The image entering the visual cortex through the dark lens is slightly delayed in the neural processing circuity compared to the clear eye, and the temporal mismatch tricks the visual system into interpreting the stimulus as three dimensional information. Both the visual and auditory artifacts are reduced over the course of time as the perceptual system is constantly exposed to the unusual input and adapts to the abnormal conditions.

DANCE AND MOTION

How does a box headed computer nerd like me know whether the music is swinging or not? Get a clue: watch the Samba dancers. If they are moving, the music is swinging. If they are standing around talking about shopping, it ain't.

Without bodily motion, music would be mere silence due to inaction. Thus an important question is *how does bodily movement directly influence auditory features in music?* In particular, how does body movement affect both the timing and dynamic expression of rhythm?

Obviously, the musician's body needs to move so as to produce the musical note(s) at the right time(s) in the rhythmic pattern. Although each musical part is well defined (typically), the rhythmic *feeling* is generally an emergent gestalt produced by several musicians. If each musician plays their part having no coordination with the other musical parts, then the music will likely have a clumsy or chaotic feel (this is intentional in some music). To achieve a unified musical feeling, the action process of each musician must incorporate feedback from listening to the other musicians. Additionally, there are inherent time lags in each musician's mind between the *intention* to play a note, the *impulse* which initiates the physical action, and the *delay* between the mental action framework and the point in time where the stick hits the drum. The note produced must then be heard by the musician who played it, and the sound and its timing are reincorporated into the overall perception of the rhythmic flow. All of this happens several times per second, for each note played. This description is a simplified view which I think is adequate for the current investigation.

Why do I hammer so hard on the details of something which is natural, intuitive and easy for a trained musician? There are two reasons. First, in any scientific inquiry, the *a priori* assumptions of the knowledge framework need to be explicitly stated. This helps validate the foundation on which the rest of the edifice is built, and exposes things which are wrong, ambiguous or otherwise suboptimal. The second reason is that my goal is to help those readers who are trying to learn to play Swing properly. This means more than just rendering the data from sheet music into note events in the acoustic world. A computer can do this, typically giving mechanical, robotic sounding music. The reasons for the differences between a human performance and a metronomically "correct" performance are often subtle if you don't know what they are and how to recognize them.

(Waadeland, 2004), (Gabrielsson, 1987) and others have investigated the connection between motion and swing rhythm. Waadeland presents video analysis of many drummers, comparing body motion with the rhythms they produce. Gabrielsson presents a collection of papers from the Third International Conference on Event Perception and Action sponsored by the Royal Swedish Music Academy, which includes a variety of research, opinions and conclusions about the relations between motion and rhythm. In all cases the connection between dynamics of bodily motion and production of swing or other rhythmic expression is well established. This should come as no surprise. The basic nature of any dynamical system is that it is extremely difficult to achieve perfect symmetry, and even the most meticulously crafted mechanical systems (e.g. Swiss watches with gears made from jewels) have a certain amount of lopsidedness to their action, even if it is measured in microseconds.

In Brazilian music, using this body english effect for producing Swing rhythm is almost universal. Indeed, musicians from many musical traditions move their bodies as part of their musical performance, whether they are a classical string quartet or the Gospel Choir in a revival Church. I am actually baffled that anyone would think that there is *not* a fundamental connection between motion and rhythm, but it has taken work by numerous researchers over several decades, as far back as (Seashore, 1938), for this very obvious effect to be accepted as a real phenomenon. This sort of resistance by the official voices of "Science" to what normal people consider common sense is surprisingly widespread.

MATHEMATICAL MODELS OF SWING

Railroads, trolley cars, horses etc. Kinematics in modeling character animation for computer graphics special FX. Spring-mass systems and differential equations.

What do a galloping horse, New Orleans trolley car and a good drummer have in common? They all produce a rhythm which we intuitively recognize as being somehow "natural". More technically, they can all be described by a mathematical technique of physics called "spring-mass" systems. Spring-mass systems are typically analyzed using *differential equations*, but don't panic. I'm not going to burden you with arcane formulae like undetermined coefficients, Laplace transforms and other techniques which I certainly found fairly difficult. Rather, I will just invoke something that's already familiar: Fourier series.

It's a well known principle in the special effects (FX) industry that the human eye is amazingly keen at distinguishing between realistic motion and artificial attempts to create natural motion using techniques like key-frame animation. Consequently, a large body of tools have been developed that use realistic physics and mathematics for defining the details of motion found in character movement, flames, smoke, explosions, flowing liquids and, to a lesser extent, motion paths in free 3D spaces, such as Spiderman swinging between buildings. Similar techniques would probably yield better swing feeling than what is typically available in music production software and drum machines. [** check out Michael's drum machine's swing settings !! **]. What little academic research I've found related to rhythmic expression has been primarily in the description and data recording stage of development, such as mentioned in the previous section, rather than research developing a mathematical theory about rhythm. [** do an exhaustive search for mathematical music models !! **]

The natural patterns of spring-mass systems are closely connected to Swing rhythm in music. Why is this? The basic answer is that both spring-mass systems and Swing rhythm have a very *bouncy* quality. This means they are predictable but not simple. The height of a bouncing ball typically diminishes in a well defined pattern, and doesn't suddenly jump to twice the height that you are expecting. On the other hand, describing the math model of the bouncing ball is fairly tricky, and generally uses differential equations. This is one of the principle reasons why using the *count and subdivide* approach of MB notation does not work very well for learning Swing. Counting and subdividing are essentially *arithmetic* in nature, and no matter how tricky the rhythms might be for any particular piece of sheet music, the fact remains that the rhythm is all based in simple mechanical notions like 2 + 2 = 4, and nothing more than this. Natural rhythms of spring-mass systems, however, belong to the field of *dynamical systems* which is quite different from arithmetic.

Intuitively, dynamical systems may actually be easier for many people to understand than are arithmetic and algebra. Like the movement patterns in special FX, the human brain has a great ability to recognize realistic versus artificial timing patterns. When you catch a frisbee, it is unlikely that you are doing any calculus in your head to predict its position and speed, although the frisbee's flight could be described by calculus and differential equations. Instead, you predict and catch a frisbee based on your previous experience of watching the flight of the frisbee. Similarly, when the music is swinging versus straight, recognizing this property is generally an intuitive perception rather than an analytical one.

Where does this intuition come from? Possibly there is something wired into our brains which directly facilitates the perception of "real" vs. "artificial", but to me it seems more likely that this is mostly a learned skill. The *ability* to learn such things may be wired in but the actual details of what is considered to be natural motion or rhythm is probably learned from the real world by example. This could be tested by large scale psychological studies of different social and cultural groups. I personally have a field study of one subject (myself), plus many conversations with other musicians, and observations of music being played with Swing or not. As I have told the story, I was raised on one side of the rhythm fence, and always yearned to be in the greener pastures of Swing music. There are a number of advantages to being in those swinging pastures – e.g. the music is more fun, and you have the chance to dance with many beautiful women, if you can, first, learn Swing, and second, translate the knowledge that's in your head to your feet and body. Both are difficult tasks for squares like me, but both are *possible.*

ASIDE: MORE ABOUT MOZART-BACH NOTATION

There are many Brazilian rhythms which are written identically, but played differently. Although subtle, these details are amenable to technical analysis.

Much of the batucada music in Brazil is not written down, but rather taught and played by example, and learned by ear and imitation. In my experience, trying to approach these rhythms from a notational viewpoint, with the standard mindset of MB subdivision, made it much more difficult to learn and play the rhythms authentically. In this section, I will show examples of the basic pandeiro and shaker rhythms, which are written the same but played differently. Using `diffdot` plots for analysis, the timing subtleties and differences are fairly obvious. Although this knowledge may not directly assist in learning and playing with *swinghee* (Brazilian swing), nonetheless, with this knowledge it becomes possible to break out of the square box.

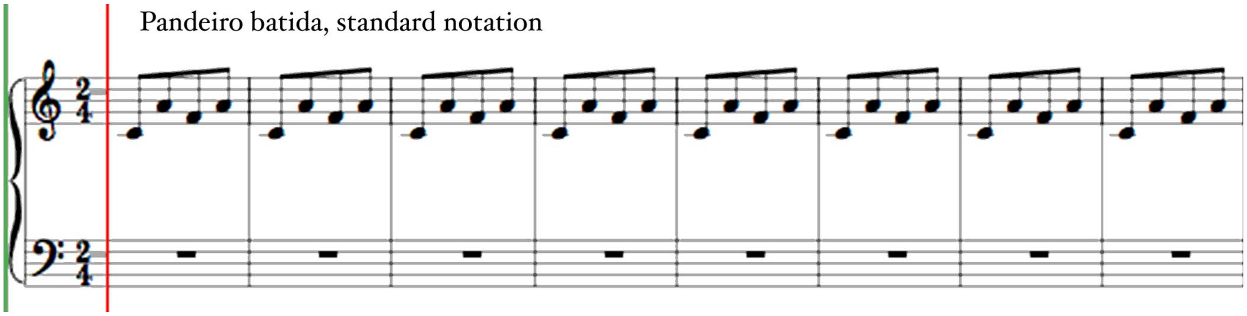

Figure 9 Tablature for the basic rhythm of pandeiro, shaker and other similar percussion parts. Note: the tablature notes indicate different types of beats, not different pitches of the instrument.

The pandeiro is the national instrument of Brazil. Basically it is like what Americans call the tambourine – a hand drum with a skin and jingles, about 10 to 12 inches in diameter. The pandeiro is played very differently from the tambourine, and has a variety of rhythms.

The basic pandeiro batida is taught as 4 beats, ostensibly evenly spaced in time. However, the rhythm is rarely *played* with evenly spaced beats. Figure 9 shows how the rhythm might be written in tablature. The note pitches (C, F and A) do not indicate different *pitches* for the beats, rather different *sounds.* The note events are ***one*** (thumb strikes pandeiro skin), ***ee*** (fingertips strike pandeiro skin near the top rim), ***and*** (palm heel hits lower rim) and ***uh*** (fingertips strike pandeiro skin near top, usually same as ***ee*** but not always). This basic cadence is played repeatedly, usually with slight variations of timing and accent. The ***one*** and ***uh*** are generally accented. A typical rhythmic feeling is like riding on a railroad.

Figure 9a shows the spectrogram for the pandeiro batida. Clearly, the note events are *not* evenly spaced in time. This gives the rhythm its *swinghee* (Brazilian swing).

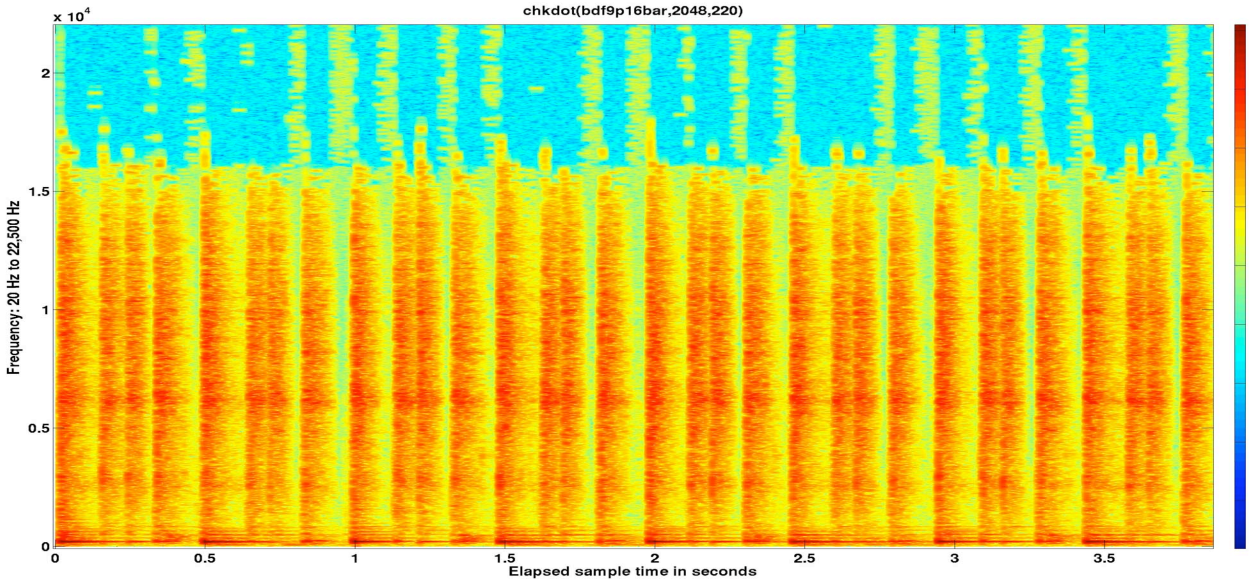

Figure 9a. Spectrogram for basic pandeiro batida.

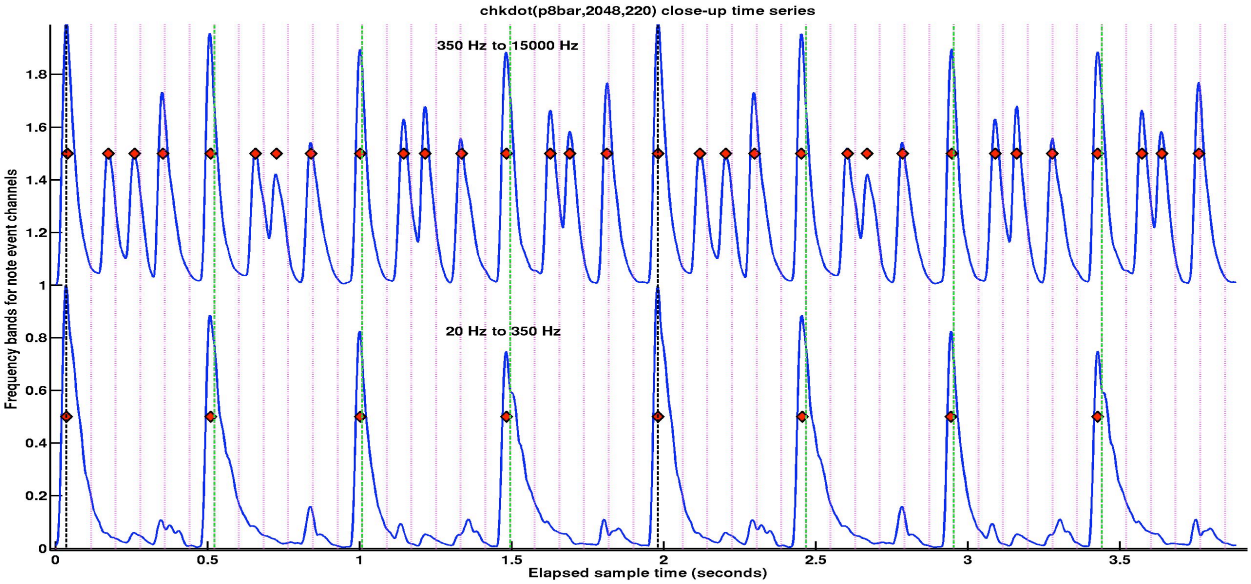

Figure 9b. `Chkdot` *diagram for pandeiro batida.*

The `chkdot` diagram is shown in figure 9b. The downbeat is the pulse and is easily identified by the low frequency *thump* of the thumb hitting the pandeiro skin. All four beats have jingle sound and so are picked out easily by choosing the high frequencies. The subdivision has four downbeats, each subdivided by 6. The ***uh*** beat lands exactly on a triplet pickup to the downbeats in all cases. This gives the rhythm a familiar Jazzy swing. The ***ee*** and ***and*** beats land on no obvious MB subdivision times, accentuating the lopsided feeling.

Other details of the swinghee are not particularly obvious in the `chkdot` diagram, but jump out quite clearly in the `diffdot` plot. First, notice the variations in the pulse timing. There is about a *6%* variation in the time deltas, and the beats seem to wander a bit, generally with a short-long pattern, but not always. In other recordings, the short-long pattern is more consistent than in this example, so we can conclude that this pandeiro player is going for a fairly loose feeling in the rhythm. Listening to the music sample doesn't give any particular feeling of sloppiness. It's more like the action of an old, well oiled, but slightly worn out machine – trolley cars jump to mind once again. The looseness is more typical of pagode than of Samba batucada, whose rhythmic style is generally quite tight.

Brazilian musicians play games with each other by pushing and pulling the beats of these rhythms, or by adding unpredictable accents and syncopations. When several excellent pandeiro players are playing this game, it may be called a *pancaderia* which would loosely translate into a fist fight or punching match. Of course, they are beating up on the pandeiros, not each other, and far from suffering, the rhythm takes on a lively jousting quality. No matter how far out one of the players may take the rhythm, they all know where the pulse is and easily come back to the same beat, leaving the audience breathless and impressed. The purpose is more to have fun than to show off.

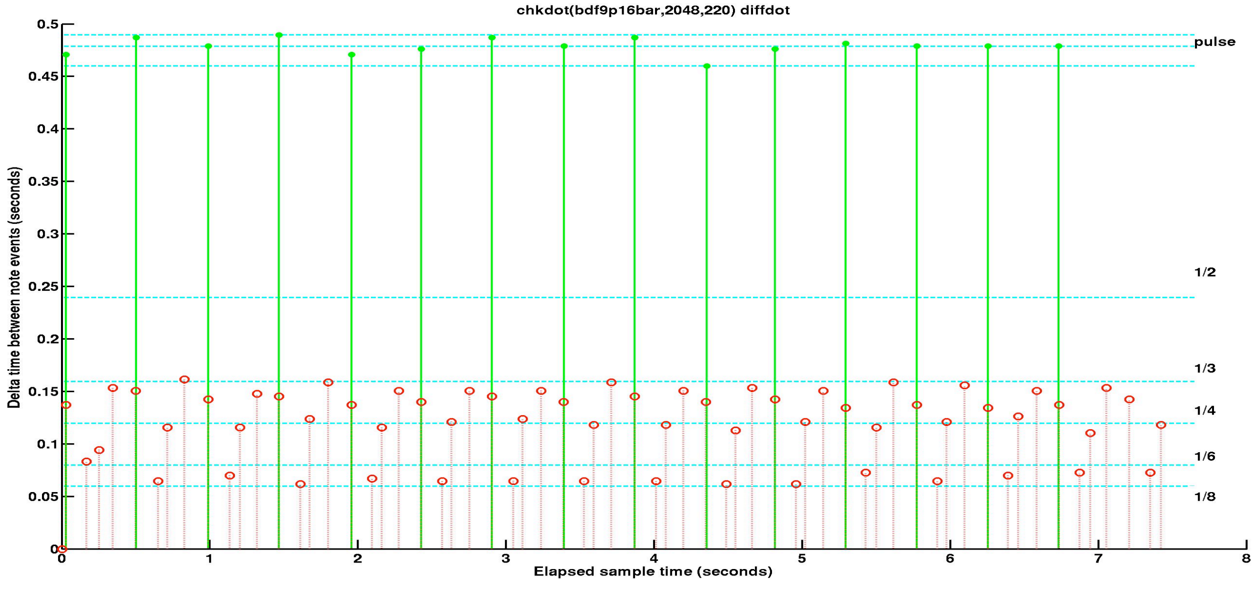

Figure 9c. `Diffdot` *diagram for pandeiro batida.*

In the red dots in figure 9c, notice that there is no explicit backbeat (1/2 the pulse). Other instruments may play a rhythmic part that emphasizes a backbeat feeling, but often the backbeat is absent in Brazilian music. If you scan the red dots individually from left to right you will see the pattern of shortening and lengthening of the beat timing, which is quite regular in its overall shape, with some slight variations. For instance, the beat following the green lines (pulse) is always the shortest, and the one preceding the green lines is generally closest to the triplet that we pointed out in the `chkdot` diagram. I find it quite interesting to note the consistent *presence* of a 1/4 note duration, which turns out to be the time delta between the ***and*** and ***uh*** beats. Although the *duration* of this beat is 1/4 note, since the ***ee*** and ***and*** beats don't happen on any standard MB subdivision, this would be very tricky to write in tablature notation and would require using a confusing collection of various rest glyphs that might be quite difficult to decipher. It would be aggravated by trying to string together the right sequence of 1/4, 1/8, 1/16, 1/32 etc. rests to approximate the 1/6 and 1/3 subdivisions of the triplet notes. A duration of 1/3 could be approximated by grouping 1/4, 1/16 and 1/64 rest glyphs, an onerous burden for the music reader to be sure.

A prototype of a notation form that might be more user friendly is shown in figure 10, which I call *swinghee notation*. The rhythm is written in the normal (inaccurate) form, and the rhythmic variations are indicated by shifting a ghost of the note symbol in both position and color to indicate intuitively to the user where the note event should be played in relation to the canonical MB beat. The second and third note events are shown leading the time where they "should" be played, and the fourth note event, being a triplet, has a 3 to indicate this special status, and is colored blue. This is not a finished form of the notation, and should be tested by real music students to evaluate its usefulness. Nonetheless, it can show the timing variations from the `chkdot` and `diffdot` diagrams quite accurately.

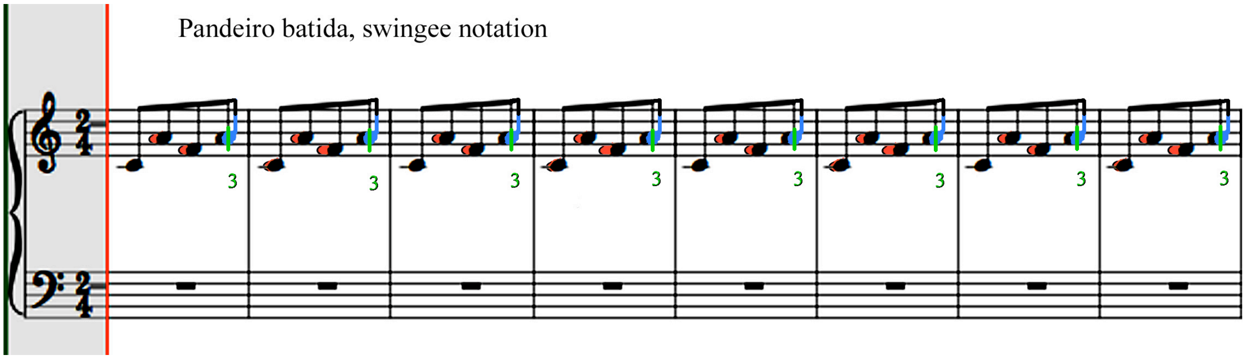

Figure 10. Swinghee notation for pandeiro batida.

Figure 11 and 11a show a rhythm which would be written identically to the pandeiro batida in figure 9, that is to say, an even quarter note subdivision. The recording has a surdo (Brazilian bass drum) and a shaker type instrument, the afoxe. I use the term *shuffle* to describe this kind of rhythm, due to the odd blend of precision and ambiguity in the actual time locations of the note events. Pandeiro beats, like most drums, are quite snappy and don't have this vague quality. The surdo pulse has a push-pull pattern. Generally the surdo plays a very tight pulse, so the push-pull in this case is far more regular than for the pandeiro. This is important in ensemble playing. where several musicians may be stretching their rhythm in different ways with timing variations that might lead to confusion without having a solid anchor such as the surdo pulse.

Like the pandeiro, the afoxe rhythm here has a strong triplet feel. The other notes of the rhythm are quite different from the pandeiro pattern. Since this is a sample of Samba batucada, the variations are played more regularly than the pandeiro part shown previously which is from a pagode style tune.

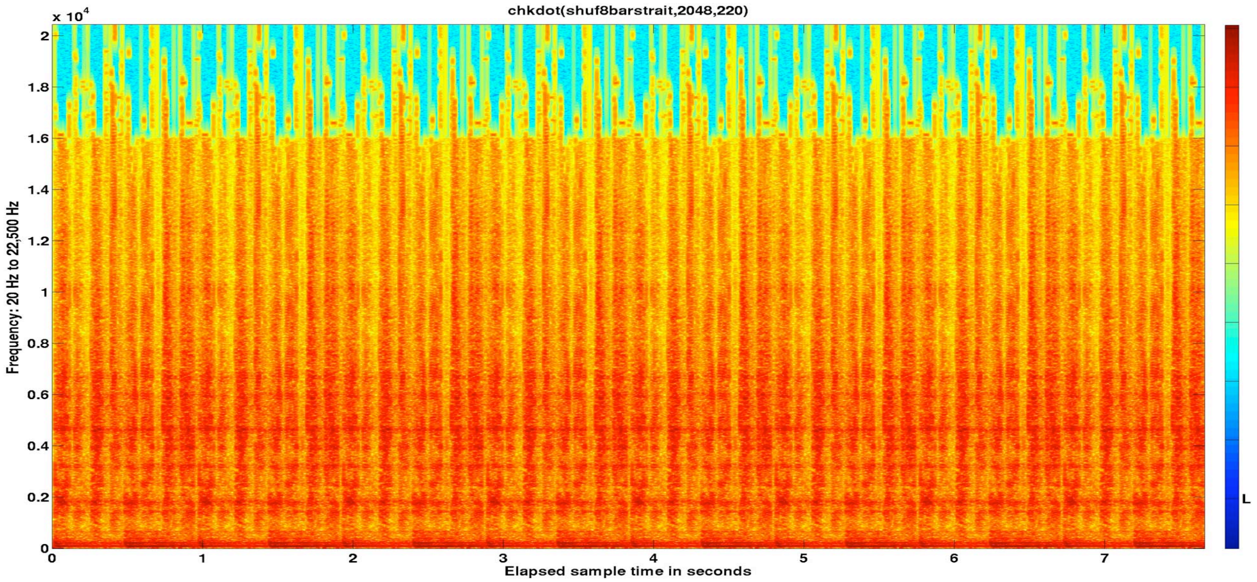

Figure 11. Spectrogram for afoxe and surdo shuffle rhythm.

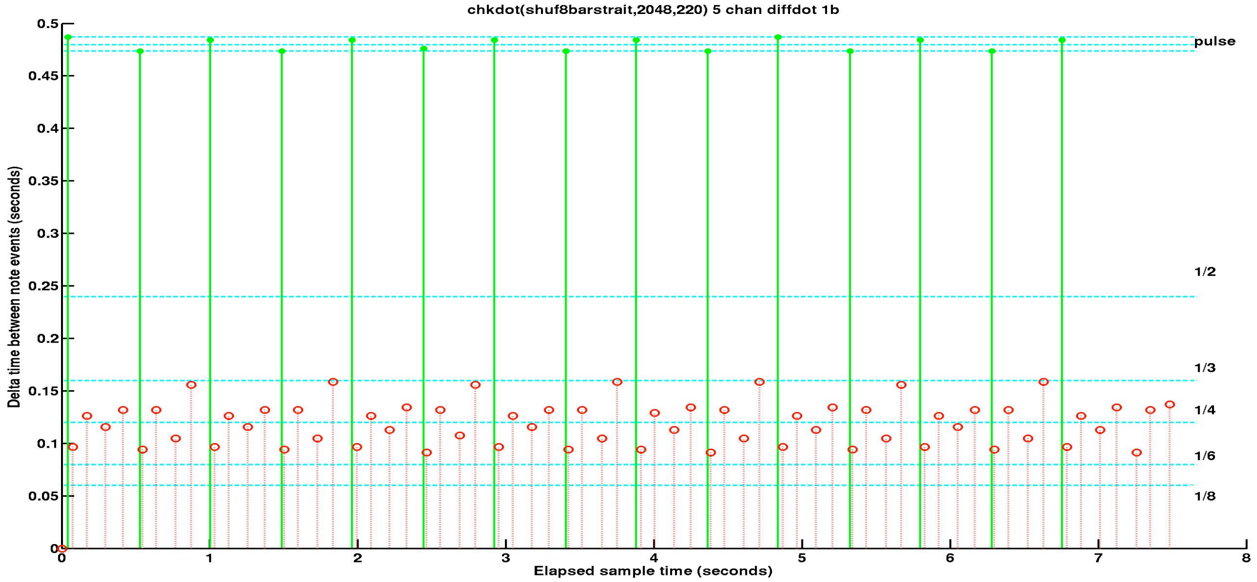

Figure 11a. `Diffdot` *diagram for shuffle rhythm.*