MORE THAN A FEELING—SOME TECHNICAL DETAILS OF SWING RHYTHM IN MUSIC

Kenneth A. Lindsay
tlafx, Ashland, Oregon 97520

and

Peter R. Nordquist
Southern Oregon University, Ashland, Oregon 97520

Introduction

If you ask a musician what makes music swing, he will reply that Swing is a feeling, and may mention counting or subdividing the beat. Commonly, in classic Jazz for example, triplet note subdivision is a feature in Swing music, but this is not the entire story. Otherwise a waltz (3/4 meter), or a 6/8 or 12/8 meter piece would inherently swing. Some pieces do, and some do not. There are also musical examples that one knows intuitively have Swing, but on close analysis do not appear to have triplet subdivision either as the main or only feature that contributes to the Swing. In this article, the authors presume that accent (differences in loudness between note events) also contributes to Swing, but thus far our research has focused solely on the timing aspects of swing rhythm. One aspect of Swing is interpreted to be the changes in the rhythmic structure around a solid and precise beat. It is the variations in that structure that are swinging.

"The real world of music as performed is more complex and interesting than the mechanistic world of sheet music."

Since classic Jazz is not the only representative of Swing, the authors want to extend the definition of Swing to include all musical styles that might be considered to "swing" by some valid metric, e.g., the musicians or dancers think the music is swinging. An ad hoc cultural definition rather than a technical definition is used to describe Swing: it is a property of music *as played* which causes listeners to dance or otherwise move their bodies in a cyclical, energetic, rhythmic manner. This definition allows consideration of a broader range of music than most prior research into Swing rhythm as well as to distinguish between Swing and other types of rhythmic expression. Rhythmic expression is the parent category of Swing, and includes many examples of differences in music as played compared to the strict metronomic timing that is specified in the written form of music. This rigid structural framework is referred to as *Mozart-Bach* or MB notation due to its historical origins. It is not implied that European Classical music is only played strictly by the metronome, however, this mind set is quite common in the training of musicians in the Western academic tradition. The real world of music as performed is more complex and interesting than the mechanistic world of sheet music, just as a movie or stage play performance has more depth and expression than is apparent by reading the script. This article will provide a short summary of prior computer science research into Swing rhythm and the analysis methods used will be briefly described. Finally, the fun stuff—a detailed technical analysis of the timing variation for a variety of styles of Swing music will be given.

Prior research

Cholakis (1995) cataloged an extensive set of Jazz drummers and analyzed the statistical nature of how each musician swung the beat in a different style by extracting the ratio of temporal intervals for notes *as played*. He claimed that this analysis allowed MIDI sequencer music with a more "human" feel to be produced. Gabriellson (1987 and 2000) observed that rhythmic variation is almost universal in music performance and reports that listeners generally prefer music played with rhythmic expression than music played strictly by the metronome. This phenomenon applies to popular music, European Classical music, and non-European traditions, such as African and Middle Eastern music. Waadelund (2004) has linked swing style to body movement, and used video recordings to study the *body english* of drummers in order to correlate their movements to the rhythmic style being played. Friberg and Sundstrom (1999 and 2002) extended Cholakis' swing ratio work. Guoyon (2005) developed computational signal processing techniques to change the swing feel in a music sample. Hamer (2000) puts a cultural slant on Friberg's and other's research, as does Birch (2003). Several software companies have products aimed at training musicians to understand and play various types of Swing.

In our research, extensive use was made of the standard *spectrogram*, i.e., the short time Fourier transform (STFT), for extracting the rhythms of different instruments. Fulop and Fitz (2006) describe a newly rediscovered form of the spectrogram which we consider to be a major advance in this information processing approach. The new spectrogram allows better time and frequency resolution for a given data set, and makes available information which is ignored in traditional uses of fast Fourier transforms (FFTs) and STFTs, such as instantaneous phase and frequency.

Many prior researchers have analyzed rhythm by using statistical analysis of note events in musical samples. This can be a useful technique, but we assert that the performance of music, whether by human or computer, is not a statistical process. Rather, each note event relates to other note events in very specific ways, and metaphors other than statistical analysis, such as symbolic relationships or local measures of specific timing between note events, should be used as appropriate. An obvious example is the hierarchical timing relations between repetitive groups of note events at different time scales. This gives rise to common musical features such as meter, beat, and subdivision. Statistical analysis can be

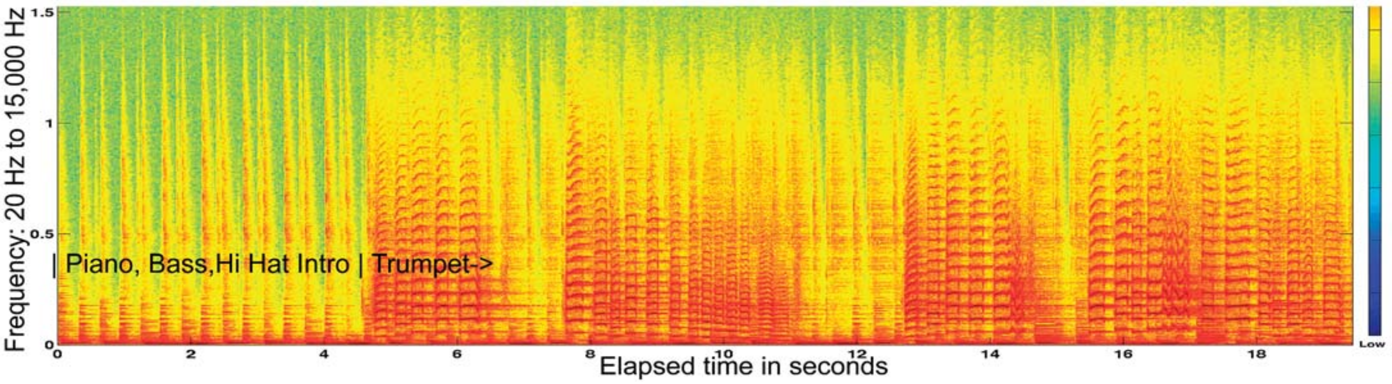

Fig 1. Spectrogram of the introduction to "It don't mean a thing (if it ain't got that swing)."

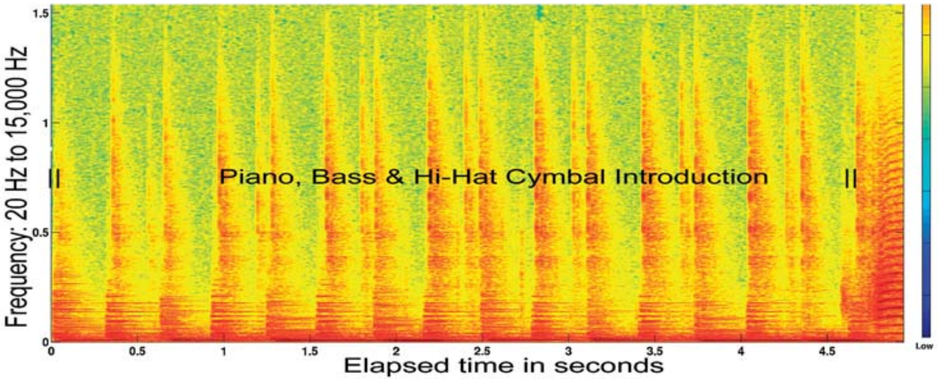

Fig 1a. Close-up of Fig. 1 showing the first 4.5 seconds (piano, bass and hi-hat cymbal) of the 19 second sample.

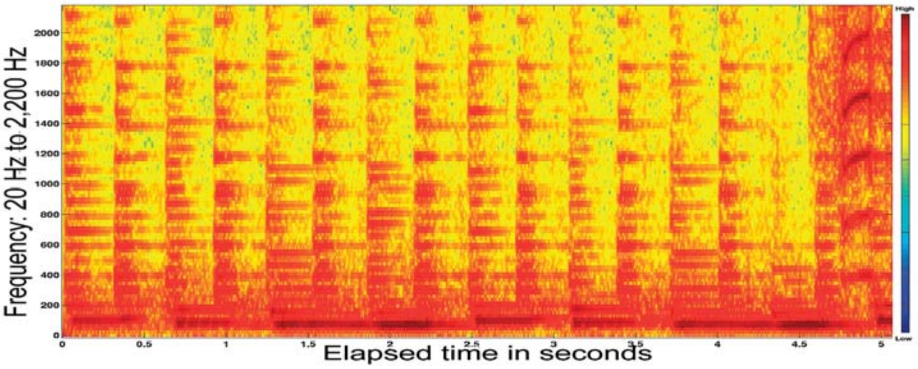

Fig 1b. Close-up showing piano and bass portion.

used for measuring the change of tempo or simple swing ratio, but it is a mistake to think that the meaning of the music is statistical. Rather, the meaning is in the specific details of the many complex forms of the Swing.

Analysis methods

The first objective in the analysis was to identify and separate the various instruments that have combined to play the musical selection. The basic tool that was used was the STFT, a standard digital signal processing (DSP) tool. First, an audio sample was divided into short time slices of a few milliseconds each. Then a window function (e.g., Hamming window) was applied to each slice to reduce aliasing effects. Finally, an FFT was used to obtain the frequency spectrum of each time slice. The choice of overlap between time slices determined the temporal resolution. This process yielded a view of how the sample's frequency content that was plotted in a spectrogram changed with time. Judicious selection of various frequency bands in the spectrogram distinguished one instrument's note events from those of the other instruments. In this way the rhythm and the Swing for each instrument was extracted.

Figure 1 illustrates a typical spectrogram image. The musical sample is the first 19 seconds of the piece *It Don't Mean a Thing (if it ain't got that swing)* that was recorded by Duke Ellington and Louis Armstrong in 1962. The first 4.5 seconds of the 19 second sample are dominated by a series of thin yellow/red spikes that are produced from the sound of the hi-hat cymbal. The remaining 14 seconds are dominated by Armstrong's trumpet solo. The introduction is expanded in Fig. 1a. In the low frequency portion of Fig. 1a there is a dense concentration of red that is produced by the piano and bass. Further expansion of the low frequencies (Fig. 1b) shows more details. To analyze the timing details, a high frequency band (7500 to 22,000 Hz) was chosen to isolate the hi-hat cymbal note events, and several low frequency bands that contain the piano (850 to 1020 Hz), (240 to 850 Hz), and bass (20 to 240 Hz). The objective was to identify and separate musical note events for each instrument, and to extract the relative timing details so that the rhythm could be specified directly from the recording, rather than approaching from the perspective of sheet music.

After selecting useful frequency bands in the spectrogram, time series graphs of the changing amplitude of each frequency band was created. Figure 1c illustrates the process of creating time series power graphs from the spectrogram. This was accomplished by adding the values for each separate time slice in each frequency band. The sum of a time slice in a frequency band was plotted as the Y value of the time series power graph for that frequency band at that point in time. The X value for the graph was set as the elapsed time (in the original recording) that corresponded to the time slice in the spectrogram. Values of the amplitude were obtained from the color in the spectrogram. Each color represented a value in the spectral data set, specifically, the amplitude coefficient of the Fourier component for a frequency. This process allowed the amplitude of each instrument in the ensemble to be isolated as a function of time.

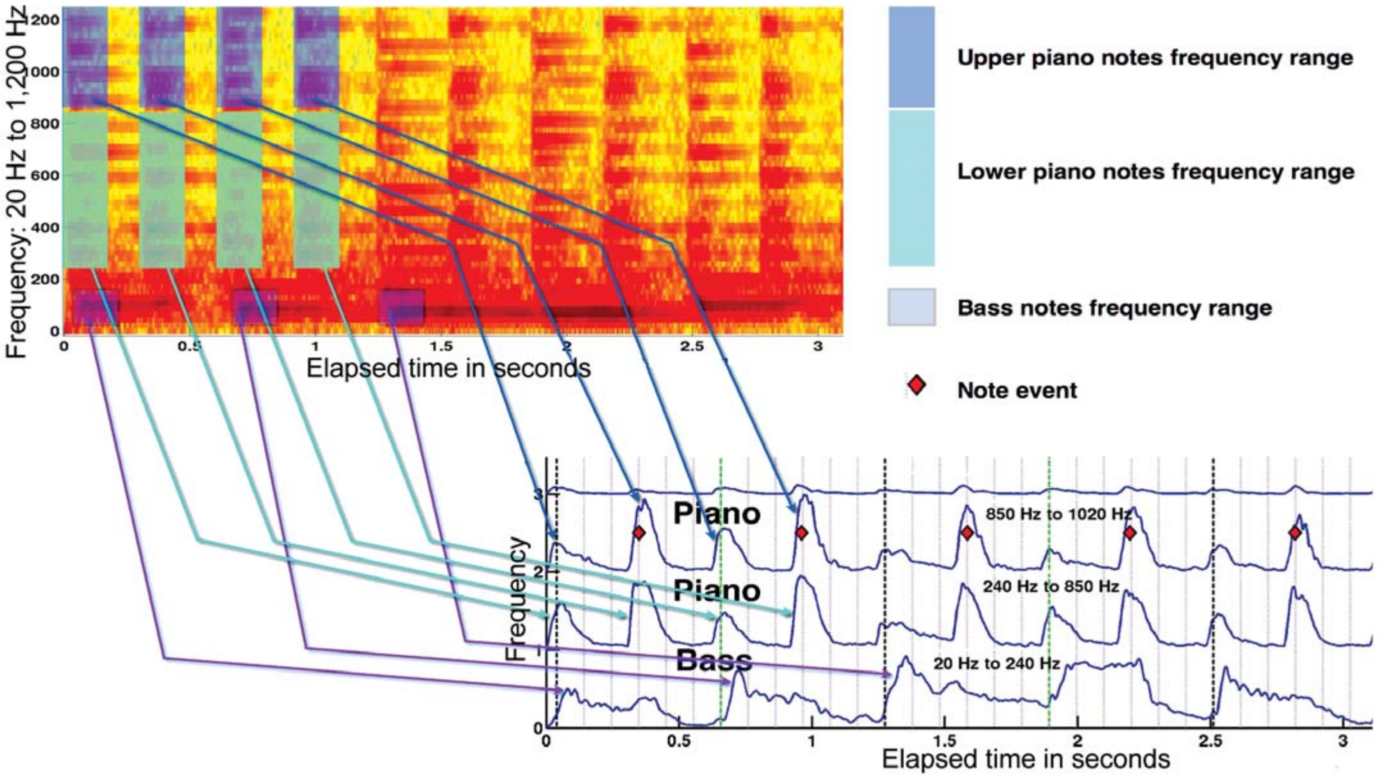

Fig1c.Construction of the CHKDOT plot formed from the frequency bands containing the piano (upper notes and lower notes) and the bass.

The next objective was to establish the elapsed time in the original recording for a "note event" (the "beat" when the note is played), i.e., the loudest time point in the vicinity of a sound pulse. The number of values in a time/frequency tile (one time slice for a frequency range) is typically between a few dozen and a few hundred, depending on the size of frequency range. The set of summed totals in each frequency band for all time slices was used to create the time series amplitude graph for that frequency range. There were the same number of points in the time series as there were in the time slices in the spectrogram simplifying time comparisons between the different types of plots. The several time series as played by each instrument or group of instruments that was generated from the chosen set of frequency bands were then stack plotted from low to high frequency as determined by a MATLAB script developed for this work. The plot and the program are called CHKDOT. The plot (see Fig. 1d) shows time aligned musical events for all frequency bands, as played by each instrument.

The computer code searched the CHKDOT waveforms for peaks representing the note events (peak amplitude in the frequency band). The *time* locations of these "note events" were extracted automatically by an algorithm that chose the point where the graph first turns back downwards immediately following a sharp vertical rise above some predetermined threshold level. This is illustrated in Fig. 1d where the note event in each peak is plotted as a red diamond. Although the algorithm is adequate, it is not ideal since in real music there may be numerous artifacts that can "mislead" the algorithm. This is due to the fact that not all note events are clear, sharp and precise. The notes played by the hi-hat in Fig. 1d can be seen to be precise, but the notes played by the piano are not. In the second and third graphs from the bottom there are collections of two, three or four small ripples at the top of some of the peaks. These may be caused by two, three or four fingers hitting the piano keys that are not precisely synchronized. Meticulous listening of the original recording can reveal the multiplicity of key note events in this frequency range. The first event in such a cluster was chosen as the keynote event. Addressing the question of what was the musician's intention, or whether the choice of note event time

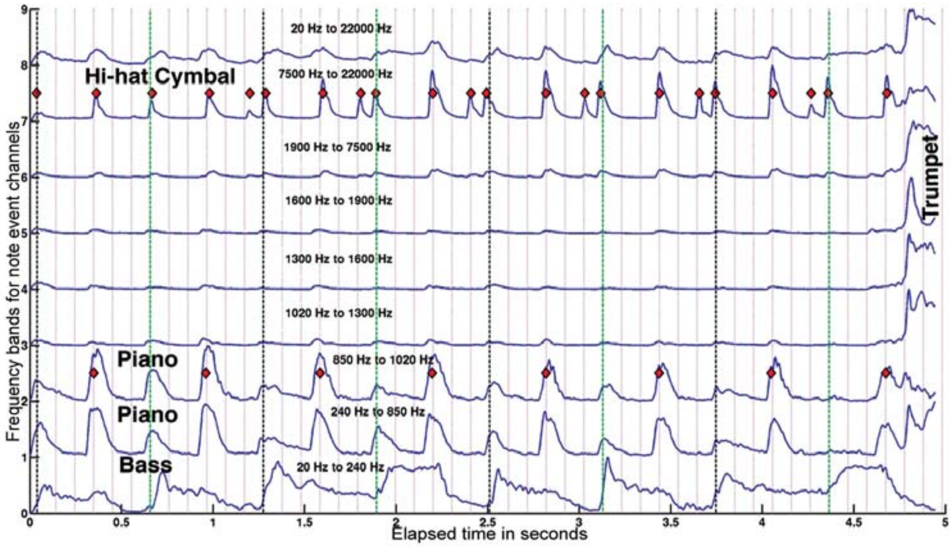

Fig 1d. CHKDOT plot of the Introduction to "It don't mean…."

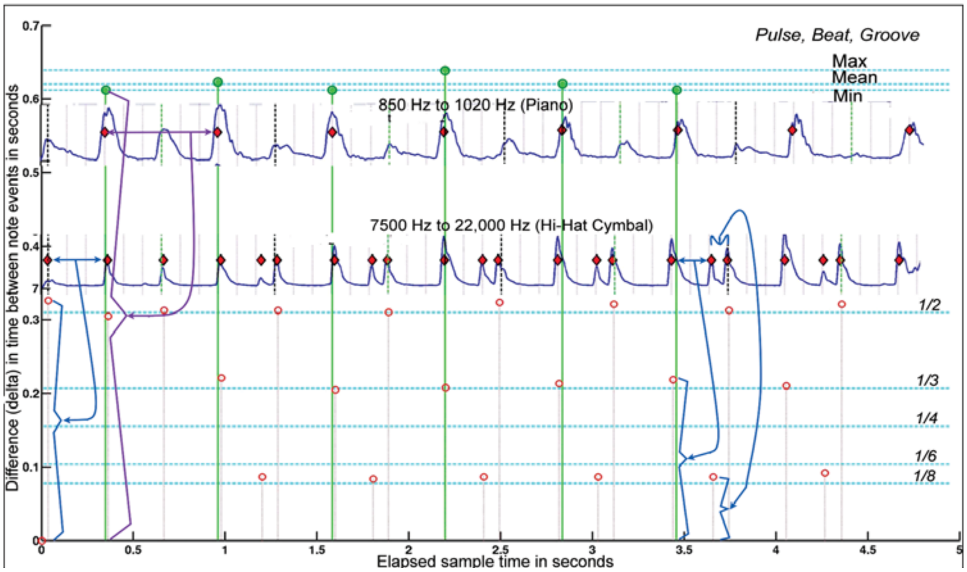

Fig 1e. Construction of the DIFFDOT plot formed from the note event time deltas.

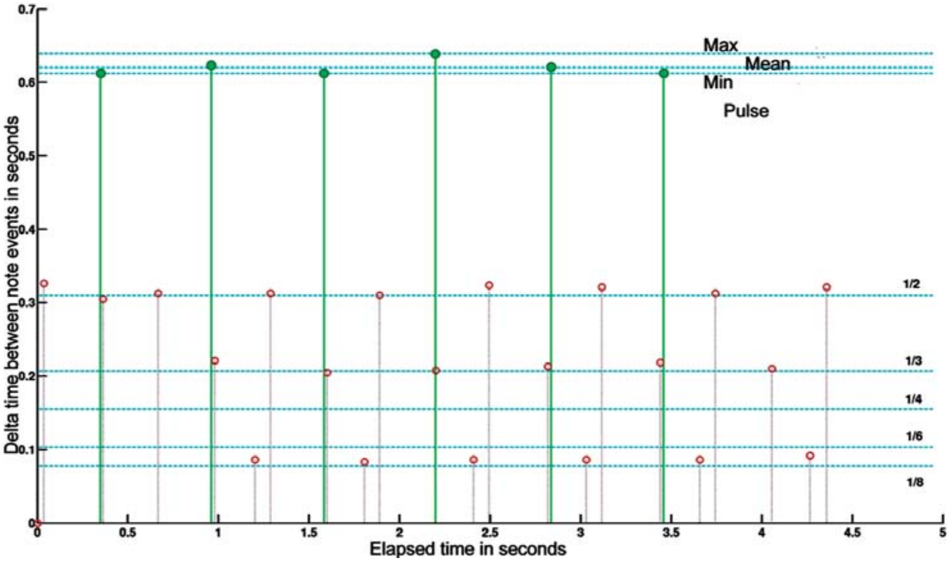

Fig 1f. DIFFDOT plot of the Introduction to "It don't mean…."

location for this study is identical to the *perceived* time location by a listener calls for further research. In this article, the focus is only on characterizing rhythmic timing and it is believed that our choice, while slightly arbitrary and ambiguous in some cases, is nonetheless reasonable for the current context.

The musical meter and subdivision were marked in a straightforward way on the CHKDOT plots. The black and green vertical lines delineate the start or downbeat of each musical measure. There are eight measures in the introduction. It can be seen that Armstrong, on trumpet, picked up his solo on the eighth measure. Figure 1d shows the breakdown for the hi-hat cymbal and piano/bass parts in *It Don't Mean a Thing*. The note events, marked by red diamonds, are placed along the invisible line in the center of the horizontal frequency band that contains the time series graph. In Fig. 1d there is one set of note events for the low frequency band (850 to 1020 Hz) and a second set of note events for the high frequency band (7500 to 22,000 Hz). Notice that some note events are sharp and distinct, e.g., the upper waveform—the hi-hat cymbal—while other time series waveforms have many jagged sections where the precise time of a note event becomes ambiguous, e.g., bottom three time series—piano upper, piano lower, and bass.

The less distinct waveforms, especially the bass, are spread out more in time than the sharp events, indicating that the attack envelope of the sound is slower for these events. These note events—actual time locations—may also be imprecise. Often, the piano and bass sounds obscure each other. Separating these overlapping note events would need more sophisticated signal processing techniques than are currently used in this study. Nonetheless, it is fairly easy to identify enough note events to specify the rhythmic timing details since there is much redundancy in the rhythm. These details are enough to reveal the Swing.

To mark the musical subdivision, note events that represent a *pulse* are first selected to use for the basic beat in the musical sample, such as the downbeat in a musical measure. This main beat can be subdivided in any convenient way, depending on the rhythm to be measured. Because triplets are a common timing feature in Swing, it was decided to subdivide the main beat by six. These subdivisions are marked with pink lines that provide for the location of a backbeat on the third pink line and triplets on the other pink lines in the same measure.

While triplets can be and often are marked in sheet music, the standard subdivision of "Mozart-Bach" (MB) notation is by factors of 2. This is one reason why notating Swing music is somewhat difficult—triplets do not fit naturally into a "subdivide by 2" metaphor. It will be demonstrated later that Swing can also contain subdivisions that are neither factors of 2 nor 3. The approach used in this study avoided the limitations of subdivision that are inherent to MB notation.

The actual note events in the recording are used to determine the musical meter and subdivision of the beat in the CHKDOT diagram (Fig. 1d). Essentially the reverse of playing a tune by reading sheet music, note information was extracted from the recording which could be used to *generate* sheet music. The pulse in Fig. 1d is marked by green and black vertical lines, which correspond to the downbeat of the measure in MB notation (a two measure phrase, one green and one black). Each musical measure was subdivided by six, looking for triplet notes of the classic Jazz Swing pattern and this subdivision was marked by using six pink lines in the CHKDOT diagram. The pink line exactly in the middle between a black and green line represents the time location of the *backbeat* of the rhythm. Thus it is observed that the piano/bass peaks are on the downbeat and backbeat, with diamond markers on certain backbeats in the third time series up from the bottom of the chart. These events were used to mark the pulse. The hi-hat cymbal note events in the time series at the top occur on the downbeat, backbeat and triplet pickup to the downbeat and backbeat. The triplet timing is indicated by note events on a pink line just ahead of a black or green line.

To analyze Swing rhythm it must be known when the beat occurred, the deviations of the beats of each instrument from their mean, and from each other. This is performed in MATLAB by a second program called DIFFDOT which extracts the time differences (time delta) between note events. Delta corresponds to the length of a musical note in MB notation—1/4 note, 1/2 note, 1/8 note etc. The pulse is used for the master time clock (whole note), and a time delta with length 1/2 of the pulse would be a half-note in MB notation, 1/4 of the pulse length would be a quarter-note and so on. Because the beat can be subdivided by any number that makes sense for a musical sample, triplets can be easily accommodated (divide by 3) or any other note time duration. Since the pulse note event timings have some variation, the minimum, maximum, and mean or average of the time differences are notated, and the mean value is used as the canonical pulse time to subdivide the beat.

Figure 1e shows the mapping process from a CHKDOT plot to a DIFFDOT plot. The two time series in Fig. 1d that were marked with note events—hi-hat cymbal and piano (upper)—are superimposed over the DIFFDOT plot, Fig. 1e, for the same time range. The elapsed time on the X axis is the same for both forms. A red diamond on the CHKDOT plot maps to a circle on the DIFFDOT plot—red circles for the hi-hat, and green circles for the piano. The X position of matching diamond/circle pairs is the same. The Y position of the circles indicates the time from that note event until the next note event in the set. Thus longer notes, such as the pulse, are at the top of the DIFFDOT plot, and shorter notes are in the lower half of the plot.

In Figs.1e and 1f, the red circles are the hi-hat note events. Notice the first three red circles are fairly evenly timed on the backbeat (1/2) of the pulse. These three time deltas correspond to the first *four* diamonds in the corresponding time series graph. After four note events, the hi-hat starts to play triplet notes, clearly visible on the pink subdivision lines in the CHKDOT diagram (Fig. 1d) and transferred to the DIFFDOT diagram (Fig. 1f) onto the 1/3 and 1/6—lying between 1/6 and 1/8, really. These note events on the 1/3 and 1/6 lines of Figs. 1e and 1f are the time deltas between the swung notes in Figs. 1e and 1f, and the beats immediately before and after: i.e., 1/2 - 1/3 = 1/6. The slight imprecision of the note timings in this example indicate a somewhat loose rhythmic style for this recording. Later a recording which has a very tight rhythmic style will be analyzed. This is another aspect of the music performance that can be read directly from the DIFFDOT diagram.

Note events are essentially transferred one for one from the CHKDOT to the DIFFDOT plots. CHKDOT plots are more intuitive to read since they parallel standard musical notation. DIFFDOT plots may require careful inspection, but by looking at the spatial patterns it is possible to get an intuitive sense of how the Swing works.

In addition to the time differences between note events, the DIFFDOT plots can also show the *variations* in time locations of repetitive musical events extracted from the CHKDOT plots, such as pulse, backbeat and swung notes. This is not a feature which can be written in MB notation. The DIFFDOT plots also clearly show how on some beats two instruments may not be precisely synchronized—in some cases, the hi-hat plays slightly before the piano note event, and in other cases, the reverse is true. This can be read directly by looking to see whether the green line is to left or right of the red line for that particular time location. Only the beat in the center of the graph is exactly synchronized.

Note that the CHKDOT diagrams are a direct representation of standard MB subdivision and counting, albeit with more fine grained timing information included, whereas the DIFFDOT plots are a novel view of the same information, essentially looking at the “first difference” form of the original timing information.

To process each musical sample into a spectrogram, a short audio clip that is typically ten to twenty seconds long was used. These are edited to be played with seamless looping, such as in a QuickTime player, to listen to the rhythm very carefully for extended periods of time. While this is not strictly needed for the analysis, it was found that it can enhance greatly both enjoyment and understanding of the rhythms. Anomalies as short as five or ten milliseconds are sufficient to be perceptible as a break in the rhythmic flow, distinguishing them from editing artifacts that may cause an unnatural transition in the audio waveform, like a click or pop. For these reasons editing at zero crossing points in the audio waveform is desired although it may not be sufficient to avoid all artifacts that can be perceived either explicitly or intuitively by a well trained human ear.

Choice of frequency resolution and short time Fourier transform window overlap was constant for each processing run, but may differ for different samples. Sometimes a single sample was processed repeatedly, using several different choices of parameters. These results provided an interesting insight into the "Heisenberg Uncertainty" aspect of the time/frequency tradeoff that is inherent to Fourier analysis. A 2048 point fast Fourier transform (FFT) and three to ten millisecond time slice overlap are well suited to many samples. In some cases a time resolution as short as 0.5 milliseconds was used. Visual inspection of the spectrogram allowed a choice of the frequency bands most likely to distinguish musical notes played by various instruments. Sets of the possibly overlapping frequency bands are summed to obtain time series plots of the audio power in the several bands.

Musical samples

Analysis results for several Swing tunes are included in this article: *It Don't Mean a Thing (if it ain't got that Swing)* by Duke Ellington and Irving Mills, performed by Duke Ellington and Louis Armstrong (1962); *Graceland* by Paul Simon (1986); *Fever* by Eddie Cooley and John Davenport, performed by Ray Charles and Natalie Cole (2004); examples of Brazilian Samba *batucada* music from the CDs *Grupo Batuque Samba de Futebol* (2004) and *Os Ritmistas Brasileiros Batucada Fantastica* (1963/1998) by Luciano Perrone and Nilo Sergio.

It Don't Mean a Thing (if it ain't got that Swing) (Duke Ellington and Louis Armstrong 1962)

Figures 1, 1a and 1b show spectrograms of the introduction to *It Don't Mean a Thing (if it ain't got that Swing)*. Figure 1 is the overview of the 19.3 second sample, showing the entire Fourier spectrum up to 22,050 Hz (half of the CD sampling rate of 44,100 samples/sec). The main feature of the first few seconds of the spectrogram is the high frequencies produced by the hi-hat cymbal playing the classic *"tchzzz-tch-ta-tchzzz-tch-ta-tchzzz…"* Jazz swing rhythm. The low frequency bass and piano parts are shown in the lower portion of the plot as a thick red swath. Louis Armstrong's trumpet dominates the remainder of the sample, clearly revealing the harmonic structure, timing and pitches of the notes. Figures 1a and 1b show close-up views of the cymbal and piano/bass section. The inherent technical limitations of Fourier analysis are clear in the coarse low frequency resolution of the piano and bass data. Fulop and Fitz's (2006) reassigned spectrogram technique would have revealed much more useful information that is obscure in the current figures.

The rhythms played by the bass/piano and hi-hat are plotted in their corresponding frequency bands as shown in Fig. 1d as a time series. Note that events were marked at the power peaks of the waveforms, and their temporal locations were collected. This sample was analyzed using five millisecond temporal resolution that was sufficiently fine grained to measure accurately the timing of note events in this song. The time deltas are plotted in Figs. 1e and 1f with longer times at the top and shorter times at the bottom. The musical pulse note events, played by piano/bass, appear at the top, and the hi-hat syncopation is in the lower part of the figure. Notice in particular that the pulse is not uniform. Rather, it alternates between slight "pushes" and slight "pulls" on the beat, i.e., the notes are intentionally *not* played in a strict mechanical metronomic style. The longest, shortest and average pulse time deltas are marked with blue horizontal lines. The average delta time has been used as the canonical pulse clock tick. The backbeat (1/2 of the pulse) and swung note deltas (1/3 and 1/6 of the pulse) are more uniform than the pulse, indicating that these syncopated notes follow more closely the uniform timing paradigm of MB meter, although departing from the "divide by 2" metaphor. Keep in mind that the DIFFDOT plot is the time difference between notes, and should not be interpreted as mirroring standard musical tablature form. The CHKDOT plot *does* correspond to the subdivision representation of tablature.

Since this song must be regarded as one of the most fun-

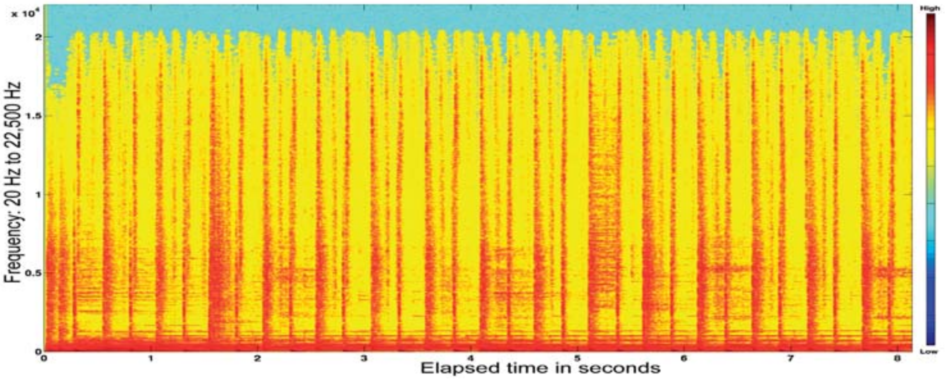

Figure 2. Spectrogram of "Graceland" intro.

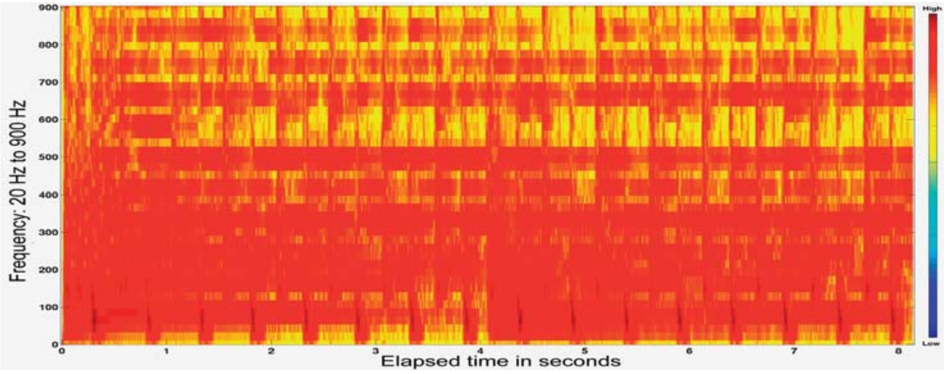

Fig 2a. Close-up showing bass and drums.

damental Swing tunes of all time, we conclude that the triplet subdivision which is clearly shown in Figs. 1d and 1f is an important feature of Swing style. What is new is the evidence of intentional time variation played in the basic pulse of the rhythm. This feature is examined more closely in subsequent examples. While a triplet subdivision can be reasonably written in MB notation, we are unaware of any similar notational device for indicating the variation of pulse timing.

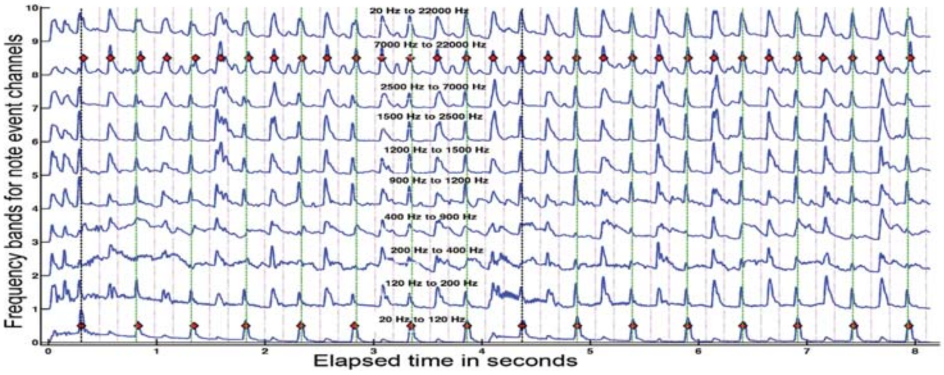

Fig 2b. Ten frequency band note events.

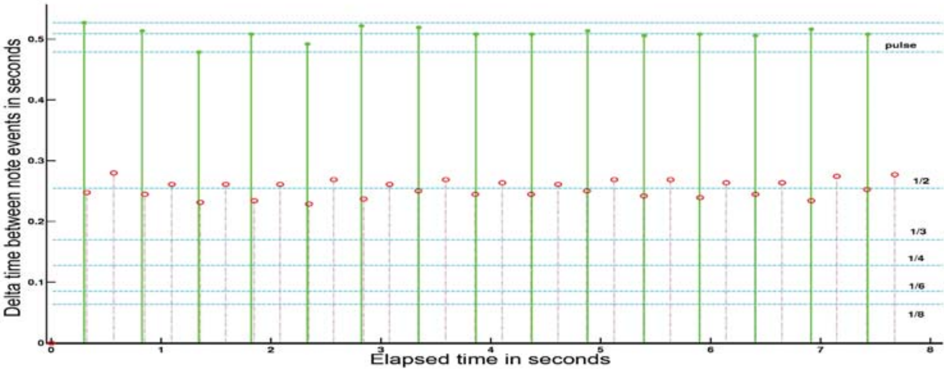

Fig 2c. DIFFDOT plot showing time variations of drum/bass pulse and rhythm guitar.

Graceland (Paul Simon 1986)

Graceland by Paul Simon is a pop tune that mimics the feeling of riding on a railroad. A prominent rhythmic feature is the song's strong backbeat, but without any great sense of the classical Jazz Swing feel. Nonetheless *Graceland* elicits a very bouncy bodily response. Figure 2 shows a spectrogram of the full audio sample, while Fig. 2a shows a close-up of the bass and drum parts. To the experienced eye, the backbeat rhythm in the low frequencies is clearer in this sample than it is in the selection *It Don't Mean a Thing*.

Figure 2b shows the time series plots of note events for ten frequency bands. The bass drum part marks the pulse in the bottom time series, including both downbeat and backbeat. The secondary note events are extracted from the high frequencies of the attack envelope of the electric guitar strumming. A triplet subdivision in the CHKDOT plots was used to look for Swing. Surprisingly, all note events were represented better by a divide by two subdivision scheme—hence half of the electric guitar notes fall between the triplet subdivision lines. The DIFFDOT plot, Fig. 2c, revealed the Swing feel for this song. Both the pulse and the rhythm guitar show a repetitive pattern of pushing and pulling the time locations of their note events. There is a substantial amount of variance to the time variations, especially in the beginning of the pulse, which indicates a short term tempo fluctuation. The rhythm guitar is much more consistent in the short/long variations of note timing, similar to the pulse of *It Don't Mean a Thing*. There is no evidence of any triplet subdivision in note timing variations. The variance of time deltas gives this song a fairly loose feel, but no sense of rhythmic sloppiness, due to the consistent repetitive pattern of time variations.

Fever (Ray Charles 2004)

Fever is a classic Rhythm and Blues (R&B) song with a strong backbeat. Ray Charles' 2004 version is played in a very tight, straight rhythmic style. Despite almost clockwork precision, this song is never boring and led to a secondary defining feature of Swing (beyond inducing body movement). A 14 second loop made from this recording

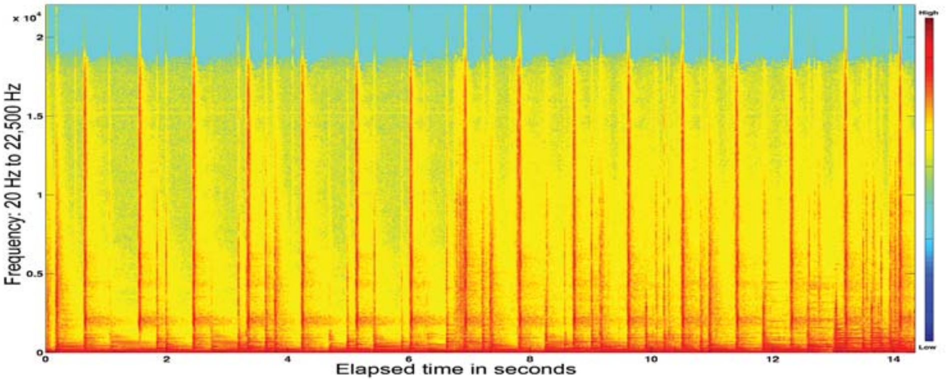

Fig 3. Spectrogram for "Fever."

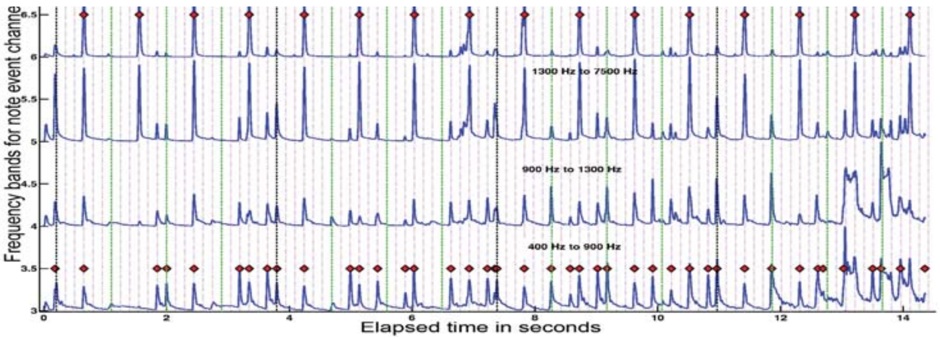

Fig 3a. Time series event plot for "Fever."

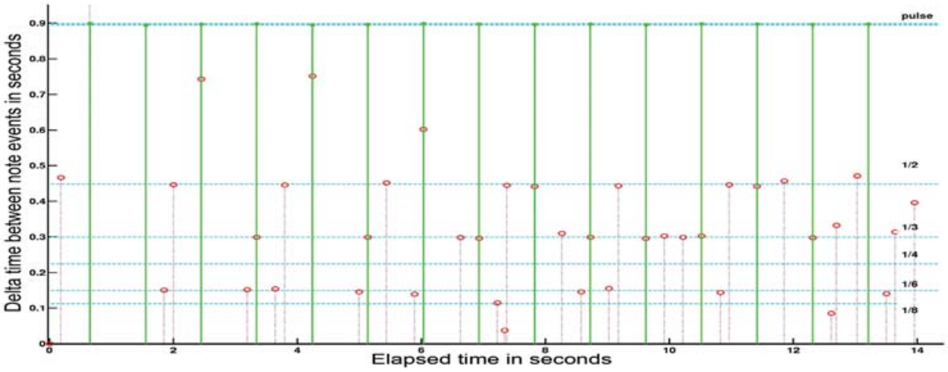

Fig 3b. DIFFDOT plot for "Fever."

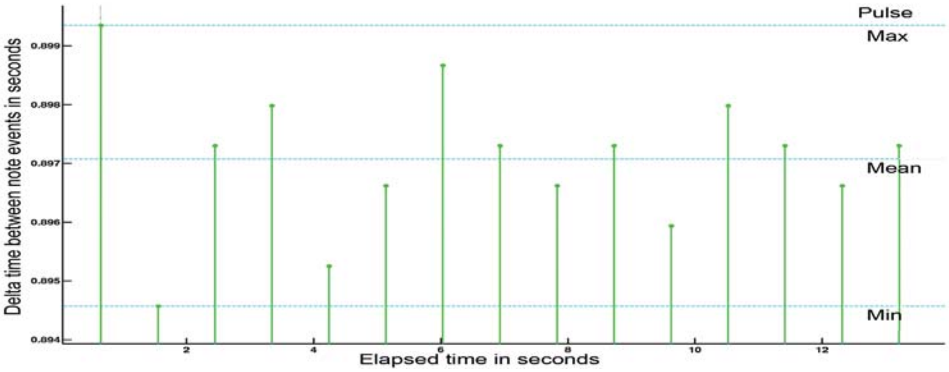

Fig 3c. Close-up DIFFDOT plot for the pulse note events in "Fever."

could play endlessly and after more than an hour, it still sounded incredibly fresh. *A sample which becomes perceptually tedious after only a few repetitions almost certainly does not Swing.*

Beneath the excellent musicianship, there exists a strong triplet element to the rhythm. The conga drum plays around the backbeat which is marked precisely by Ray Charles' finger snaps. About half of the conga note events are on triplet pickup beats before either the downbeat or the backbeat, with a few on triplets following these main beats. Unlike *Graceland* or *It Don't Mean a Thing*, this sample shows virtually no rhythmic looseness. The conga, drums, finger snaps, and bass guitar are synchronized with each other to a precision of better than 15 milliseconds in almost all cases. Contrast this precision with *Graceland's* consistent variations of 50 to 80 milliseconds, and *It Don't Mean a Thing's* somewhat random looking variations in the 30 to 40 millisecond range. These are details that distinguish between loose and tight rhythmic styles.

Figures 3, 3a and 3b show the familiar set of spectrogram, time series event plots, and DIFFDOT diagram. Subdivision of the meter in the time series plot is a four beat pulse phrase with six subdivisions of the pulse. Thus, the downbeat, backbeat and triplet temporal locations are marked by vertical lines. Finger snaps and conga drum beats land exactly on these time ticks. The precise clusters of note events on the pulse, backbeat, and triplet time lines in the DIFFDOT plot are evident. There is a general absence of note events on the quarter note line, just as there was in *It Don't Mean a Thing*.

A very remarkable aspect of this recording can be seen in the close-up DIFFDOT plot (Fig. 3c) showing only the pulse of Ray Charles snapping his fingers on the backbeat. It is obvious from the normal DIFFDOT plots of *Graceland* and *It Don't Mean a Thing* that the variations in the pulse event time deltas are much greater on those two samples than on *Fever*. The close-up shows that Ray Charles finger snap time deltas are less than 5 milliseconds. This means that the deviation from the canonical MB metronome beat times is less than +/- 2.5 milliseconds. Given the tight rhythmic style of this recording, and the fact that Ray Charles was one of the 20th century's best musicians, we believe this DIFFDOT plot represents an important data point regarding the limits of human time perception and physical action.

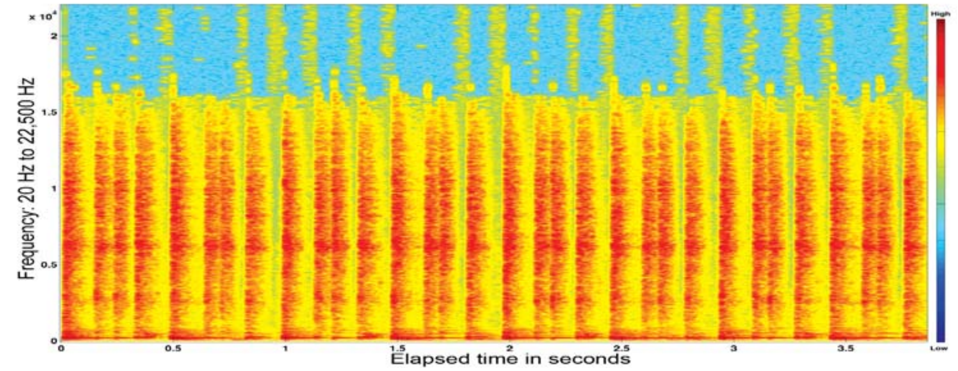

Fig 4. Spectrogram of pandeiro batida.

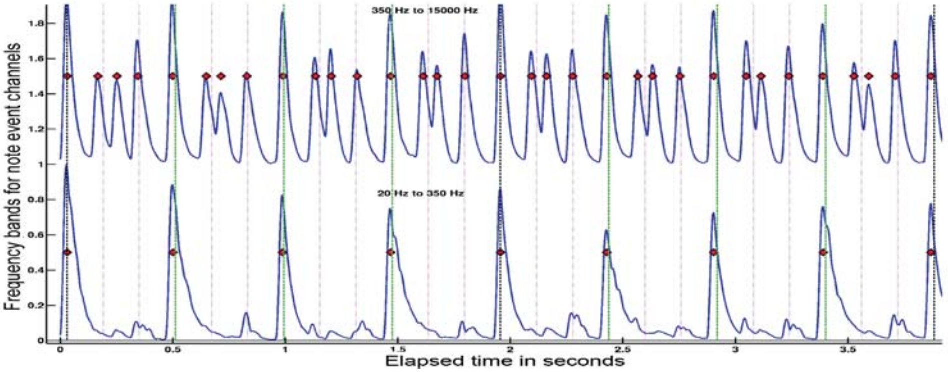

Fig 4a. Time series plot of pandeiro batida.

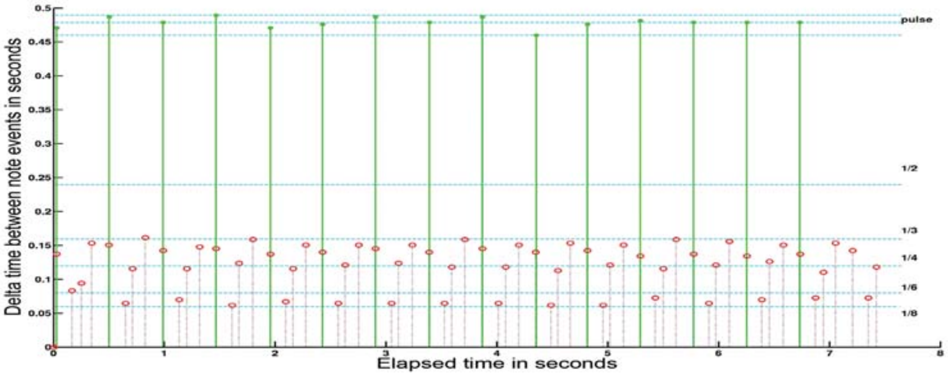

Fig 4b. Time series plot of pandeiro batida.

Brazilian swinghee

Swing may include complex rhythmic patterns, but it can also be found in very simple rhythms. This is well illustrated in a basic Brazilian rhythm, the "pandeiro batida," literally "beating pattern of the pandeiro." The pandeiro is the national instrument of Brazil and is approximately the same as a tambourine in American music. The tambourine is also found in many other musical traditions, but the Brazilian pandeiro has several playing styles that are unique. The basic pandeiro batida is a simple 1-2-3-4 pattern played continuously with slight temporal and accent variations that denote which phrase of a larger pattern is being played. This pandeiro batida is invariably taught as straight time: *one-ee-and-uh* played with thumb (*one*), fingertips (*ee*), palm heel (*and*), fingertips (*uh*), over and over. This batida is both taught and written as a succession of evenly spaced quarter or eighth notes, but playing in Brazilian swing style (called *swinghee*, or *balance* in Portuguese) is far removed from even-spacing.

The spectrogram in Fig. 4 clearly shows the basic simplicity of this rhythm, and also illustrates how the beats are not played with even timing despite being written as equal notes. The time series plot in Fig. 4a has the pulse in the lower frequency band which is the thumb hitting the pandeiro skin causing a low thump. All four notes appear in the upper frequencies which are caused by the metallic jingles of the pandeiro. The "uh" note is consistently played on a nearly exact triplet pickup to the pulse. This classic Jazz feature is certainly part of the swinghee feeling. A four beat pulse with three subdivisions per pulse, that gives vertical lines on the exact triplet note time locations is illustrated in Fig. 4a.

The second and third notes (*ee* and *and*) are played in two very odd locations in the first half of each measure. Neither of these is played on a triplet, quarter or eighth note location, and there are slight time variations between repetitions of the basic batida, indicating two sides of the larger phrase. The pattern of these time variations is consistent, since the DIFFDOT plot (Fig. 4b) clearly shows the Swing pattern as a repeating waveform with variation, rather than some kind of random pattern. The plot shows a complete absence of a backbeat (1/2 of pulse) and the consistent presence of a note time interval of 1/4 of the pulse. This is the time delta between the *ee* and *and* notes, and would be a standard quarter note in MB notation *if* the time location of the note events were on the canonical quarter note subdivision of the meter, which is not the case. An accurate rendering of this rhythm (as played) in MB notation would need a convoluted pattern of multiple rest and note glyphs of various lengths (e.g., 1/4 plus 1/32 plus 1/64, or 1/2 minus 1/3 etc, all very problematic for the music reader) to capture the actual timing of the notes as played. The DIFFDOT diagram shows these non-standard note time durations in a very natural fashion as subdivisions of the pulse. The DIFFDOT pulse shows the familiar push/pull on the canonical downbeat time locations, although in the time series plot, this is a subtle feature.

There are a variety of swinghee styles used to play the basic pandeiro batida. As in American music, there are probably as many styles of Swing as there are drummers or pandeiro players. Brazilian swinghee clearly has a very different feature set than American Swing, even in this simple example.

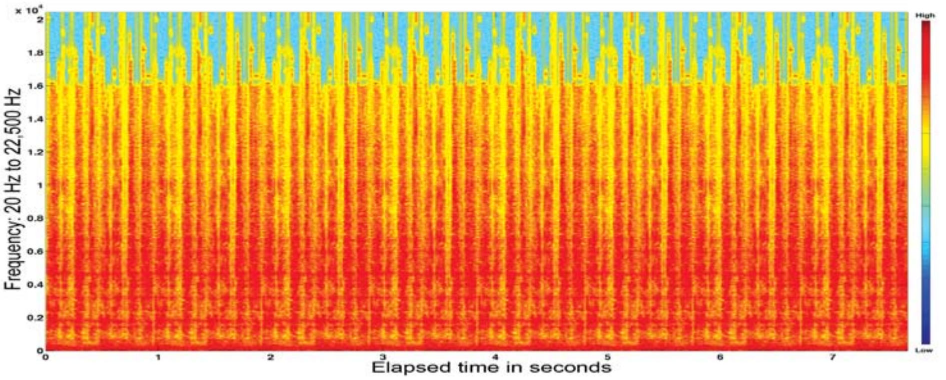

Fig 5. Spectrogram of shuffle rhythm.

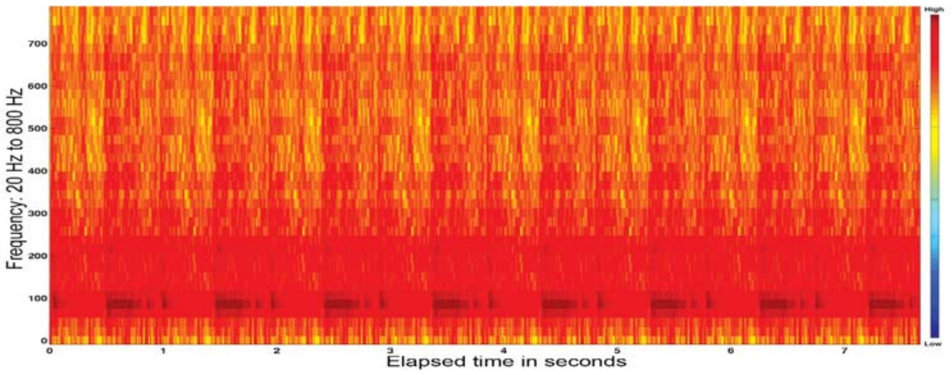

Fig 5a. Close-up of low frequencies.

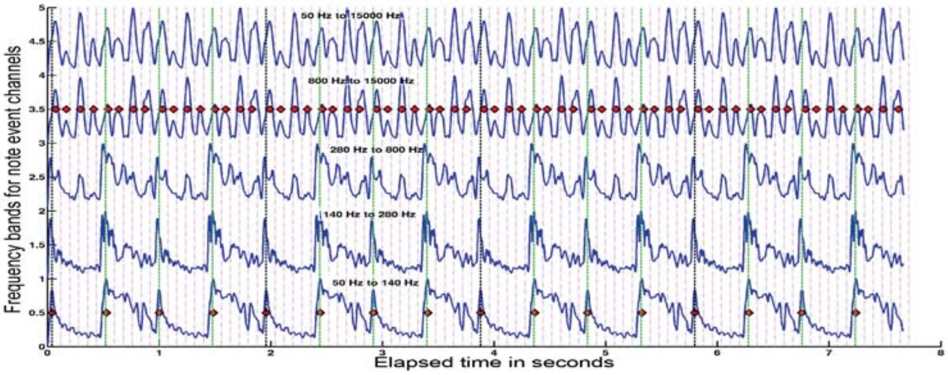

Fig 5b. Time series of shuffle rhythm.

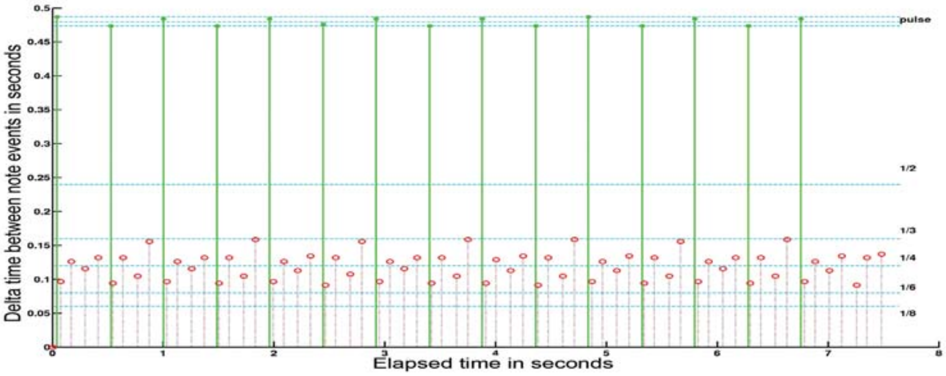

Fig 5c. DIFFDOT plot of shuffle rhythm.

Brazilian swinghee 2: Shuffle

Many percussion and drum note events have a very sharp and precise onset, making them easy to identify by our approach while others have a much less precise sound. The term "shuffle" is used to describe a wide range of Swing rhythms played in this style. Shakers, brushes on a snare drum or hi-hat cymbal, afoxe and guiro are all examples of shuffle instruments. Single note events from these instruments can be identified, but overall there is a feeling of blurring and blending of each note event into the next. The meter and subdivision of the rhythm is defined by the loudness peaks which are identifiable but somewhat temporally ambiguous events. Shuffle is an odd combination of vagueness and precision, difficult to describe with language.

Identification of note events is more difficult for these less precise musical events, and marking the onset time locations precisely can be subject to interpretation of how the rhythm feels. The peak power was chosen to be used as the location of the note event, although perceptually there is some activity happening before the peak, unlike most percussion sounds with their fast onset. The standard Brazilian ganza (shaker) rhythm usually has a noticeable snap that precedes the downbeat and this snap is fairly sharp, but the remaining notes are more blurry. The snap gives a precise anchor to the rhythm which makes the blurry parts sound well integrated to the ensemble Swing, rather than sounding as if played carelessly. Overall, this contributes to the flowing feel that many Brazilian songs have.

The spectrogram in Figs. 5 and 5a shows a shuffle pattern played by a surdo (Brazilian bass drum) and an afoxe (gourd instrument with a stick scraping across it). The audio spectrum is quite diffuse, although note events can be identified. The time series plot, Fig. 5b shows considerable complexity of the waveforms in all frequency bands due to the instrument's timbre—a stick scraping across the grooves in the gourd producing many correlated closely spaced clicking sounds which spread across many frequencies in a fairly non-harmonic fashion. The pulse in the low frequency is played by the surdo. The DIFFDOT plot, Fig. 5c shows swinghee timing variations in both the pulse and secondary events tracks. Like *Graceland*, the two rhythms are closely connected but playing the Swing in different ways.

Ensemble swing in Brazilian swinghee

In this section an example of complex interaction between two instruments will be examined in some detail. The pandeiro plays a duet with a tamborim, a small Brazilian hand drum generally hit with a stick. The *tamborim* plays many of the most complex rhythms in a Samba. The basic rhythms are often difficult, and the interpretive timing is very fine-grained and precise, typically 10 to 20 millisecond excursions from canonical beat locations.

Figure 6 shows the pulse played by the pandeiro, and the *desinha* (design, a Brazilian term for complex rhythmic ornament) played by the tamborim. In the upper plot when the tamborim starts playing, it is not at the standard beginning of the batida. Instead the drummer plays a variation on a portion of the second half of the entire tamborim phrase, which leads into the downbeat. The downbeat is indicated by the green marker at time location 1700, except that there is a further variation—it is not the primary downbeat but the offbeat, so the tamborim is playing on the opposite side from the pandeiro. It is very common in Brazilian music for some

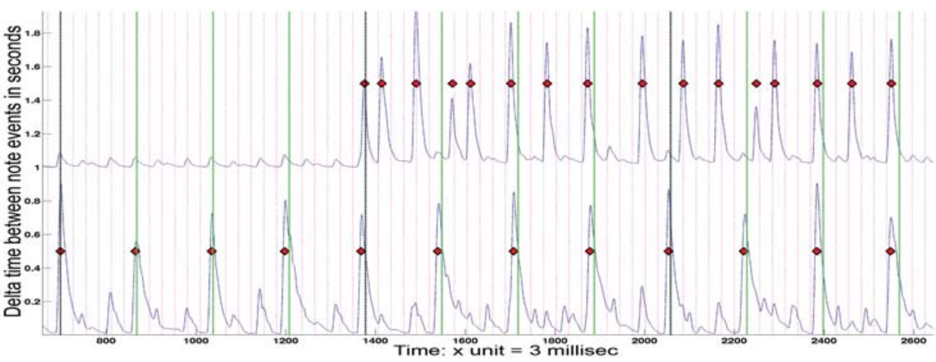

Fig 6. Pandeiro pulse and tamborim desinha.

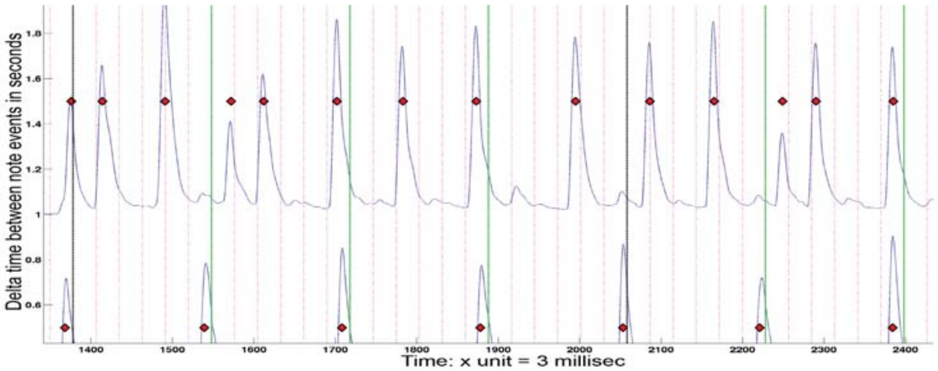

Fig 6a. Close-up showing micro-timing.

two phrase batidas to be played with the two phrases swapped. This is analogous to the 3-2 clave and 2-3 clave style in Cuban music. Swapping the sides gives a different feel, usually more syncopated if the unfamiliar variant is played.

The tamborim batida is very syncopated even when played straight. The "standard" place to start the basic tamborim batida is at note event #6 in Figs. 6 and 6a at time location 1700, very slightly ahead of the beat. Many batidas have beats played ahead of the standard subdivision beat, and/or also slightly ahead of, or behind the note events of other instruments. In this example, the pandeiro plays about 30 milliseconds ahead of the standard downbeat at this temporal location, and the tamborim plays about 15 milliseconds ahead of the pandeiro. This technique is used to give a push to the feeling of the rhythm by both instruments. A few beats on either side of the 1700 point, both instruments play notes exactly on a standard subdivision. The feeling of this pattern is consistent throughout the sample which is several minutes long. Figure 6a shows a closer view of the micro-timing.

The 15 to 30 millisecond time variations are on the order of 1/64 or 1/32 notes at the typical quick Samba tempo of 140 beats per minute. We believe from our experience with Brazilian music that the musicians are playing these timing variations entirely by intuition and experience, rather than explicitly subdividing the beat in the moment, i.e., by feeling rather than analysis. We have found that analytical understanding has substantially improved our ability to play and hear these rhythms, but that in the performance too much analysis actually impedes our ability to play the groove well.
Looking at the two sets of three evenly spaced notes starting at 1700 and 2000, observe that the first and third beats are slightly ahead of the standard subdivision. These beats push the rhythm slightly and give a somewhat more energetic feeling to the music than if they are played straight. In this case, these two tamborim note events are also accented, further emphasizing the push to the rhythm at these two time points. The combination of time push and accent are caused by the tamborim player putting a little extra "juice" into the rhythm for these note events. Waadelund (2004) has studied the relation between this type of "body english" and the rhythms played by drummers on drum kits. The investigation of the relation between motion and rhythm started in the early 20th century. Seashore (1938) and Gabrielsson (1987) both include a variety of reports, insights and opinions about this phenomenon. In this example, the tamborim plays the first beat right on top of the pandeiro on the "real" downbeat, instead of playing at the "standard" temporal location for the note. This portion of the batida starts its repetition at the ninth event location (time 2000, triplet pickup to downbeat), just before the main downbeat, marked by the black line at time 2050. You can see that the first beat ordinarily is on the triplet pickup to the downbeat, and the next two beats are almost exactly evenly spaced on the subsequent triplet time points. The slight variations from playing exactly on temporal locations that correspond to a standard subdivision are part of the swinghee style. While one might think that this is rhythmic looseness similar to the *Graceland* example, generally Brazilians play these slight temporal variations quite precisely, consistently and intentionally.

Conclusions and future work

Swing is a far more complex part of the musical landscape than reported previously in the academic computer science literature. The authors have analyzed Swing rhythms in American, Jamaican, and Brazilian music. Some of these are simple enough to allow a complete assessment of the musical features that give rise to Swing feeling. Others point in the direction of subtle complexities that require improvements to the pattern recognition and signal processing techniques to characterize fully the Swing details described in this article. There are many other musical styles which have Swing characteristics including Cuban, Middle Eastern, African, Funk, and Hip Hop. Our analysis results clearly point to a basic inadequacy of standard Euro-American musical notation to annotate swing rhythm styles. Comments and observations from professional musicians agree with this notational limitation. For the purposes of musical analysis in the context of music information retrieval (MIR), the authors feel that it is more fruitful to omit most attempts to render a musical performance as tablature. It would be more practical and accurate to maintain the information in a form which is close to the actual audio data, and the information features that can be extracted from such recordings.**AT**

References for further reading:

Birch, Alisdair MacRae (**2003**). "It Don't Mean a Thing If It Ain't Got that Swing," Just Jazz Guitar Magazine, August 2003. Online: http://www.alisdair.com/educator/musicarticles.html.

Cholakis, Ernest (**1995**). *Jazz Swing Drummers Groove Analysis.* Numerical Sound. Online: http://www.numericalsound.com.

Friberg, A., and Sundstrom, J. (**1999**). "Jazz drummers' swing ratio in relation to tempo," J. Acoust. Soc. Am. **105**, 1330(A) (1999).

Friberg, A., and Sundstrom, J. (**2002**). "Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern," Music Perception **19**(3), 333-349.

Fulop, Sean A., and Fitz, Kelly (**2006**). "A Spectrogram for the twenty-first century," Acoustics Today **2**(3) 26–33.

Fulop, Sean A., and Fitz, Kelly (**2006**). "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," J. Acoust. Soc. Am. **119**(1), 360–371.

Gabrielsson, A., ed. (**1987**). *Action and Perception in Rhythm and Music*. Papers given at a Symposium in the Third International Conference on Event Perception and Action. Royal Swedish Academy of Music, #55. Stockholm, Sweden.

Gabrielsson, A. (**2000**). "Timing in Music Performance and its Relation to Music Experience," in *Generative Process in Music* edited by J. A. Sloboda (Clarendon Press, Oxford).

Guoyon, F. (**2005**). "A Computational Approach to Rhythm Description," Ph.D. thesis. University of Barcelona, Spain.
Hamer, M. (**2000**). "It don't mean a thing if it ain't got that swing. But what is swing?," New Scientist **2270**, 48.

Seashore, C. E. (**1938/1967**) *Psychology of Music* (Dover Publications, Inc. New York).

Sloboda, J. A.(ed.). (**2000**) *Generative Process in Music*. (Clarendon Press, Oxford, UK).

Waadelund, C. H. (**2004**). *Spectral Properties of Rhythm Performance* (Norwegian University of Science and Technology, Trondheim).

Online resources:

http://www.makemusic.com/practice_tools.aspx
http://www.majorthird.com/ http://flat5software.com/
http://www.red-sweater.com/clarion/
http://fastrabbitsoftware.com/eartraining.htm
http://homepage.mac.com/ronmiller2/RonsSite/software.html
http://www.lpeters.de/ http://www.cope.dk/
http://www.midomi.com/ http://www.pandora.com/
http://www.musicip.com/ http://www.wizoo.com/ (original maker of Darbuka and Latigo software, now owned by digidesign.com)
http://www.tlafx.com/

Discography:

Louis Armstrong and Duke Ellington - *Louis Armstrong meets Duke Ellington* (1962)
Paul Simon - *Graceland* (1986)
Ray Charles - *Genius Loves Company* (2004)
Bob Marley - *Legend* (1990)
Grupo Batuque - *Samba de Futebol* (2004)
Luciano Perrone e Nilo Sergio - *Os Ritmistas Brasileiros Batucada Fantastica* (1963/1998)

Ken Lindsay is an Information Science researcher with tlafx in Ashland, Oregon. His current work includes extracting previously unseen information from music and biophysical signals. Previously he worked for 8 years at NASA Ames Research Center in the Neuro Engineering Lab, developing real-time 3D graphics for flight simulators. In other incarnations he has worked in hardware and software engineering, theatre, film and radio. A serious student of Brazilian music for over 10 years, he has performed in the San Francisco Bay area, New Orleans, and Rio de Janeiro. He holds an MS in Math and Computer Science from Southern Oregon University.

Pete Nordquist is an assistant professor in the Department of Computer Science at Southern Oregon University (SOU). He holds an MS in Computer Science and Engineering from the Oregon Graduate Institute and an MM in Choral Conducting from the University of Missouri Kansas City Conservatory of Music. He worked for 14 years at Intel Corporation in various software development groups and has taught computer science at SOU for the past five years. His musical involvement includes having sung with the Kansas City Chorale and the Oregon Repertory Singers. He currently sings with the SOU Chamber Choir, the Rogue Valley Chorale, and serves as rehearsal assistant for the Rogue Valley Youth Ensemble.