

Rhythm and Transforms, Perception and Mathematics

William A. Sethares

University of Wisconsin, Department of Electrical and Computer Engineering,
1415 Engineering Drive, Madison WI 53706
sethares@ece.wisc.edu

Abstract. People commonly respond to music by keeping time, tapping to the beat or swaying to the pulse. Underlying such ordinary motions is an act of perception that is not easily reproduced in a computer program or automated by machine. This paper outlines the flow of ideas in *Rhythm and Transforms* (Sethares 2007), which creates a device that can “tap its foot” along with the music. Such a “beat finding machine” (illustrated in Fig. 1) has implication for music theory, on the design of sound processing electronics such as musical synthesizers, on the uses of drum machines in recording and performance, and on special effects devices. The beat finder provides a concrete basis for a discussion of the relationship between the mind’s processing of temporal information and the mathematical techniques used to describe and understand regularities in data. Extensive sound examples (Sethares 2008) demonstrate beat-based signal processing techniques, methods of musical (re)composition, and new kinds of musicological analysis.

1 What Is Rhythm?

How can rhythm be described mathematically? How can it be detected automatically? People spontaneously clap in time with a piece of music, and can effortlessly internalize and understand rhythmic phenomena – but it is tricky to create a computer program that can keep time to the beat. Teaching the computer to synchronize to music requires both interesting mathematics and unusual kinds of signal processing.

There are many different ways to think about and notate rhythmic patterns. A variety of different notations, tablatures, conventions, and illustrations are used throughout *Rhythm and Transforms* to emphasize the distinction between symbolic notations (which accentuate high level information about a sound) and acoustical notations (which allow the sound to be recreated). Surveying the musics of the world shows many different ways of conceptualizing the use of rhythmic sound: for instance, the timelines of West Africa, the clave of Latin America (illustrated in Fig. 2), and the *tala* of India.

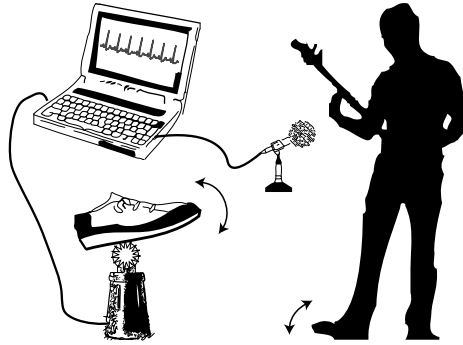


Fig. 1. A foot-tapping machine designed to mimic people’s ability to synchronize to complex rhythmic sound must “listen” to the sound, locate the underlying rhythmic pulse, anticipate when the next beat timepoint will occur, and then provide an output

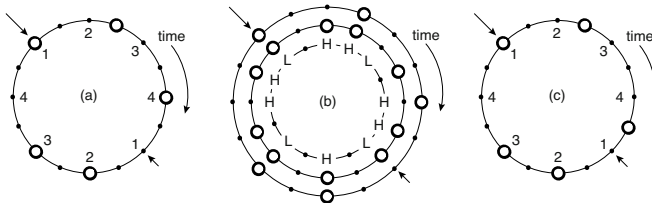


Fig. 2. The son clave rhythm is arranged in necklace notation; the 3-2 clave begins at the larger arrow while the 2-3 clave begins at the smaller arrow. (a) The beats of the two $\frac{4}{4}$ measures are indicated inside the circle along with the 16 timepoints that represent the tatum (short for “temporal atom,” the fastest pulsation present in the music) (b) repeats the basic clave in the outer circle and shows how various other rhythmic parts complement, augment, and can substitute for the straight clave pattern. The middle circle shows the *cáscara*. The inner circle shows a bell pattern with low (L) and high (H) bells. (c) shows the *guanquancó* (rumba) clave.

2 Auditory Perception

The auditory system is not simple. Underlying the awareness of rhythmic sounds are basic perceptual laws that govern the recognition of auditory boundaries, events, and successions. Research into the mechanisms of perception sheds light on the physical cues that inspire rhythmic patterns in the mind of the listener. These cues help distinguish features of the sound that are properties of the signal (such as amplitude and frequency) from those that are properties of the perceiving mind (such as loudness and pitch). Just as pitch is a perceptual correlate of frequency and loudness is a perceptual correlate of amplitude, the “beat” is a perceptual correlate. A major part of *Rhythm and Transforms* is the search for physically measurable correlates of the beat perception. Fig. 3 illustrates this idea.

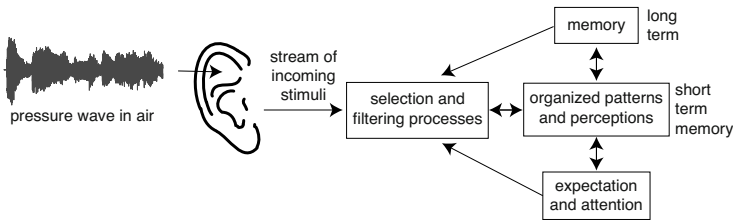


Fig. 3. Perception of sound is not a simple process; it begins with a physical waveform and may end with a high level cognitive insight (for example, understanding the meaning of a sound). There are constant interactions between long term memory, attention and expectation, and the kinds of patterns formed. There are also constant interactions between memory, attention, expectation, and the ways that the raw information is selected and filtered. The time span over which the short term memory organizes perceptions is called the *perceptual present*.

3 Transforms

Transforms model a signal as a collection of waveforms of a particular form: e.g., sinusoids for the Fourier transform, mother wavelets for the wavelet transforms, periodic basis functions for the periodicity transforms. All of these methods are united in their use of inner products as a basic measure of the similarity and dissimilarity between signals, and all may be applied (with suitable care) to problems of rhythmic identification. A transform must ultimately be judged by the

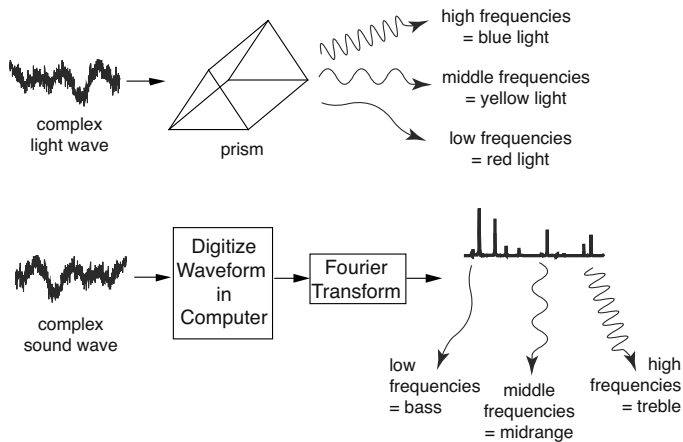


Fig. 4. Just as a prism separates light into its simple constituent elements (the colors of the rainbow), the Fourier Transform separates sound waves into simpler sine waves in the low (bass), middle (midrange), and high (treble) frequencies. Similarly, the auditory system transforms a pressure wave into a spatial array that corresponds to the various frequencies contained in the wave.

insight it provides and not solely by the elegance of its mathematics. Transforms and the various algorithms derived from them (for instance, the phase vocoder and short time Fourier transform) are mathematical operations that have no understanding of psychoacoustics or of the human perceptual apparatus. Thus a square wave may be decomposed into its appropriate harmonics by the Fourier transform irrespective of the time axis. It makes no difference whether the time scale is milliseconds (in which case we would hear pitch) or on the order of seconds (in which case we would hear rhythm). It is, therefore, necessary to explicitly embed psychoacoustical insights into the mathematics (Terhardt (1982) and Parncutt (1994) provide two well known examples) in order to make more practical and effective models. Mathematics is perceptually agnostic – it is only the interpretation of the mathematics that makes a psychoacoustic model. Fig. 4 presents one such interpretation.

4 Adaptive Oscillators

One way to model biological clocks is with oscillators that can adapt their period and phase to synchronize to external events. To be useful in the beat tracking problem, the oscillators must be able to synchronize to a large variety of possible input signals and they must be resilient to noises and disturbances. Clock models can be used to help understand how people process temporal information and the models are consistent with the importance of regular successions in perception. One simple situation is shown in Fig. 5.

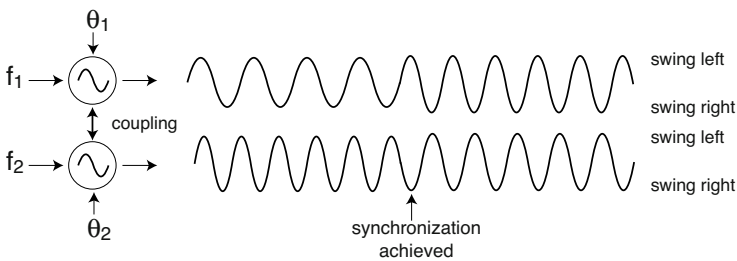


Fig. 5. When two oscillators are coupled together, their frequencies may influence each other. When the outputs synchronize in frequency and lock in phase, they are said to be entrained. The musicians represent one “oscillator” and the beat finding machine represents a second. When they synchronize, the machine has “found the beat.”

5 Statistical Models

The search for rhythmic patterns can take many forms. Models of statistical periodicity do not presume that the signal itself is periodic; rather, they assume that there is a periodicity in the underlying statistical distributions. In some cases, the randomness is locked to a known periodic grid on which the statistics

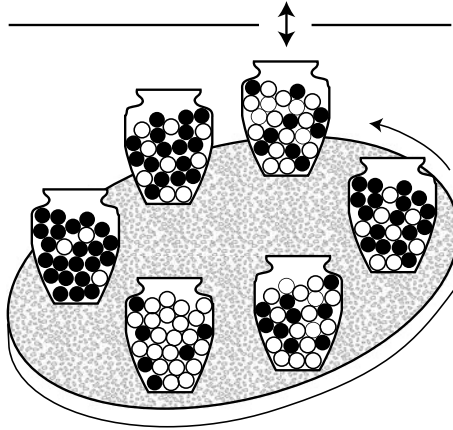


Fig. 6. The simplest useful model is a generalization of the “ball and urn” problem where a collection of urns are mounted on a carousel. Each time a ball is removed from one of the N urns (indicated by the arrow), the platform rotates, bringing a new urn into position. When N is unknown, it is necessary to infer both the percentage of balls in each urn and the number of urns (the periodicity) from the experiments. In terms of the periodicity-finding goals of beat tracking, inferring N is often more important than inferring the individual percentages of black or white balls.

are defined. In other cases, the random fluctuations may be synchronized to a grid with unknown period. In still other cases, the underlying rate or period of the repetition may itself change over time. The statistical methods relate the signal (for example, a musical performance) to the probability distribution of useful parameters such as the period and phase of a repetitive phenomenon. One simple model is shown in Fig. 6.

6 Automated Rhythm Analysis

Just as there are two kinds of notations for rhythmic phenomenon (the symbolic and the acoustical), there are two ways to approach the detection of rhythms; from a high level symbolic representation (such as an event list, musical score, or standard MIDI file) or from a acoustical representation such as a direct encoding in a .wav file. Both aspire to understand and decompose rhythmic phenomena, and both exploit a variety of technologies such as the transforms, adaptive oscillators, and statistical techniques. A preliminary discussion of the rhythmic parsing of symbolic sequences is then generalized by incorporating perceptually motivated feature vectors to create viable beat detection algorithms for audio. The performance of the various methods is compared in a variety of musical passages. A visual representation is shown in Fig. 7.

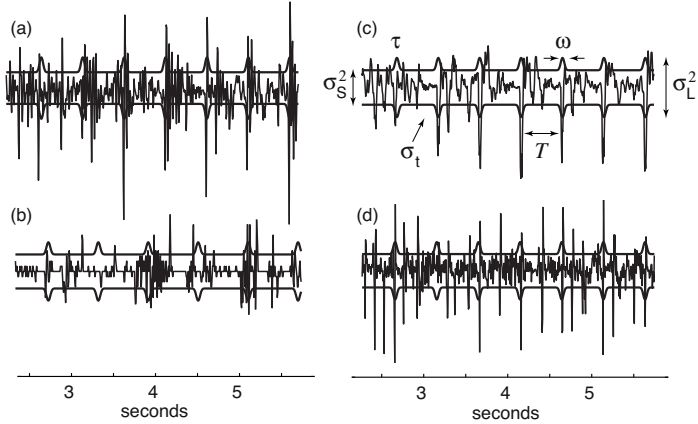


Fig. 7. A few seconds of four feature vectors of *Pieces of Africa* by the Kronos Quartet are shown. The estimated beat times (which correctly locate the pulse in cases (a), (c), and (d)) are indicated by the bumps in the curve σ_t that are superimposed over each vector. The three timing parameters T (period), τ (phase), and δT (change in period, not shown) are estimated from the feature vectors.

7 Beat-Based Signal Processing

There is an old adage in signal processing: if something is known about a signal, use the knowledge. The ability to detect beat timepoints is information about the naturally occurring points of division within a musical signal and it makes sense to exploit these points when manipulating the sound. Signal processing techniques can be applied on a beat-by-beat basis or the beat can be used to control the parameters of a continuous process. Applications include beat-synchronized special effects, spectral mappings with harmonic and/or inharmonic destinations (as illustrated in Fig. 8), and a variety of sound manipulations that exploit the beat structure. Illustrative sound examples can be heard online (Sethares 2008).

There are two ways to exploit beat information. First, each beat interval may be manipulated individually and then the processed sounds may be rejoined. To the extent that the waveform between two beat locations represents a complete unit of sound, this is an ideal application for the Fourier transform since the beat interval is analogous to a single “period” of a repetitious wave. The processing may be any kind of filtering, modulation, or signal manipulation in either the time or frequency domain. For example, Fig. 9 shows the waveform of a song partitioned into beat-length segments by a series of envelopes. Each of the segments can be processed separately and then rejoined. Using envelopes that decay to zero at the start and end helps to smooth any discontinuities that may be introduced.

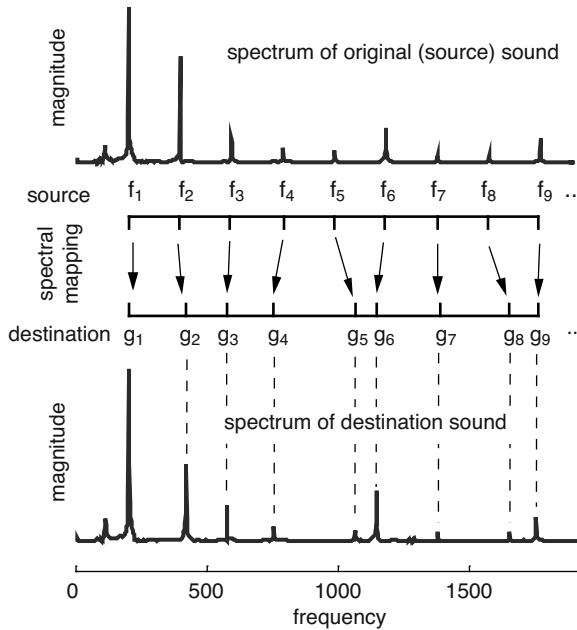


Fig. 8. In this schematic representation of a spectral mapping, a source spectrum with peaks at f_1, f_2, f_3, \dots is mapped into a destination spectrum with peaks specified at g_1, g_2, g_3, \dots . The spectrum of the original sound (the plot is taken from the G string of a guitar with fundamental at 194 Hz) is transformed by the spectral mapping for compatibility with the destination spectrum. The mapping changes the frequencies of the partials while preserving the energy in each partial, leaving the magnitudes approximately the same.

The second method uses beat locations to control a continuous process. For example, a resonant filter might sweep from low to high over each beat interval. The depth of a chorusing (or flanging) effect might change with each beat. The cutoff frequency of a lowpass filter might move at each beat boundary. There are several commercially available software plug-ins (for example, *Camelspace* and *SFXMachine*) that implement such tasks using the tempo specified by the audio sequencer; the performer implicitly implements the beat tracking.

Since certain portions of the beat interval may be more perceptually salient than others, these may be marked for special treatment. For example, time stretching by a large factor often smears the attack transients. Since the beat locations are known, so are the likely positions of these attacks. The stretching can be done nonuniformly: to stretch only a small amount in the vicinity of the start of the beat and to stretch a larger amount in the steady state portions between beat locations.

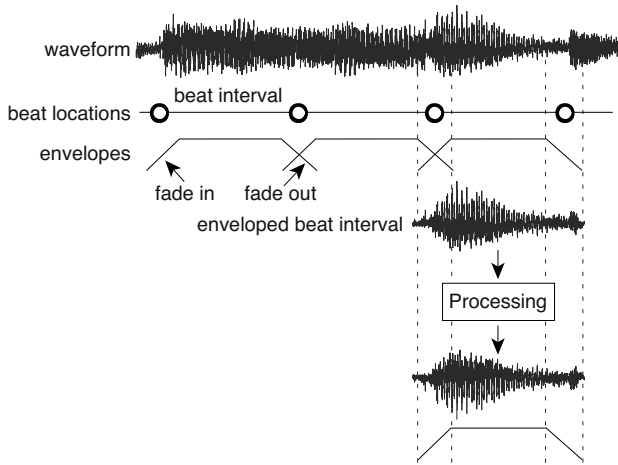


Fig. 9. A collection of windows separates the waveform into beat intervals, which can be processed independently. After processing, the intervals are windowed again to help reduce clicks and edge discontinuities. The final step (not shown) is to sum the intervals to create a continuous output.

8 Musical Composition and Recomposition

The beats of a single piece may be rearranged and reorganized to create new structures and rhythmic patterns including the creation of beat-based “variations



Fig. 10. This mosaic of Scot Joplin (created from many smaller pictures) presents a visual analog of an audio collage: a piece is deconstructed into beats, and then reconstructed by reordering the beats. A series of sound examples available on the website (Sethares 2008) demonstrate this.

on a theme.” For example, it is easy to remove every fourth beat. The effect is to change a piece in 4/4 time into 3/4, as is demonstrated by transforming Scott Joplin’s “Maple Leaf Rag” into the “Maple Leaf Waltz,” which can be heard on the author’s website (Sethares 2008). Similarly, two pieces may be merged in a time-synchronous manner to create hybrid rhythmic textures that inherit tonal qualities from both. See Fig. 10.

9 Musical Analysis via Feature Scores

Traditional musical analysis often focuses on the use of note-based musical scores. Since scores only exist for a small subset of the world’s music, it is helpful to be able to analyze performances directly, to probe both the symbolic and the acoustical levels. For example, Figure 11 displays a skeletal tempo score that shows how time evolves in several different performances of the Maple Leaf Rag. More generally, Banelos (2005) details several psychoacoustically motivated feature scores that are particularly useful in an analysis of Alban Berg’s Violin Concerto, subtitled *Dem Andenken eines Engels*, that merges standard analytical techniques with new feature scores in an elegant and insightful way. By conducting analyses in a beat-synchronous manner, it is possible to track changes in a number of psychoacoustically significant musical variables. This allows the automatic extraction of new kinds of symbolic feature scores directly from the performances.

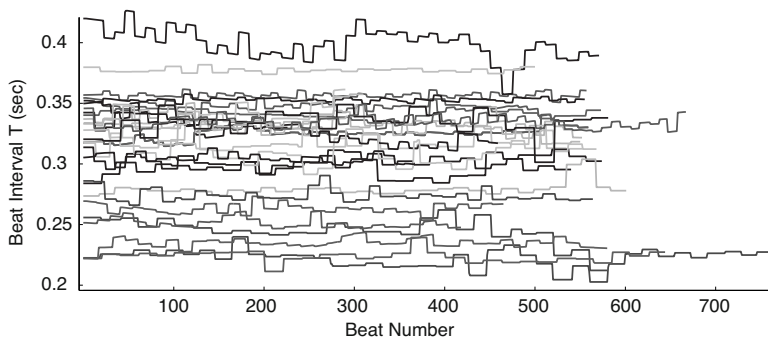


Fig. 11. A *tempo score* is a plot of the duration of each beat vs. the beat number; it shows how the tempo changes over time. In this plot, 29 performances of the *Maple Leaf Rag* are played in a variety of tempos ranging from $T = 0.22$ to $T = 0.4$ sec per beat. The plot shows how the tempo of each performance varies over time.

10 Conclusions

The ability to decompose a piece into its primitive beat-elements is a surprisingly powerful technique for musical analysis, for musical composition (such as

beat-synchronous sound collages) and for audio signal processing (where the beat boundaries provide a natural partitioning of the signal). *Rhythm and Transforms* (Sethares 2007) contrasts two ways of understanding temporal regularities in the world around us: directly via perception and indirectly via mathematical analysis. “Rhythm” alludes to the perceptual apparatus that allows people to effortlessly observe and understand rhythmic phenomena while “transforms” evokes the mathematical tools used to detect regularities and to study patterns. The book develops a variety of such applications and provides a wealth of sound examples (Sethares 2008) that concretely demonstrate the efficacy and the limitations of the techniques.

References

- Banuelos, D.: Beyond the Spectrum of Music, DMA Thesis, University of Wisconsin (2005)
- Parncutt, R.: A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception* 11(4) (1994)
- Sethares, W.: *Rhythm and Transforms*. Springer, Heidelberg (2007)
- Sethares, W.: Sound examples accompanying this article can be heard (2008), <http://eceserv0.ece.wisc.edu/~sethares/RT.html>
- Terhardt, E., Stoll, G., Seewann, M.: Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America* 71, 679–688 (1982)