Journal of New Music Research
2001, Vol. 22, No. 5, pp. ••–••

0271-3683/01/2205-367$16.00

Meter and Periodicity in Musical Performance

William A. Sethares[1] and Thomas W. Staley[2]

[1]Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706-1691 USA;
[2]Science and Technology Studies Program, Virginia Polytechnic Institute, Blacksburg, VA 24061

Abstract

This paper presents a psychoacoustically based method of data reduction motivated by the desire to analyze the rhythm of musical performances. The resulting information is then analyzed by the "Periodicity Transform" (which is based on a projection onto "periodic subspaces") to locate periodicities in the resulting data. These periodicities represent the rhythm at several levels, including the "pulse", the "measure", and larger structures such as musical "phrases." The implications (and limitations) of such automated grouping of rhythmic features is discussed. The method is applied to a number of musical examples, its output is compared to that of the Fourier Transform, and both are compared to a more traditional "musical" analysis of the rhythm. Unlike many methods of rhythm analysis, the techniques can be applied directly to the digitized performance (i.e., a soundfile) and do not require a musical score or a MIDI transcription. Several examples are presented that highlight both the strengths and weaknesses of the approach.

1 Introduction

Listeners can easily identify complex periodicities such as the rhythms that normally occur in musical performances, even though these periodicities may be distributed over several interleaved time scales. At the simplest level, the pulse is the basic unit of temporal structure, the foot tapping "beat". Such pulses are typically gathered together in performance into groupings that correspond to metered measures, and these groupings often cluster to form larger structures corresponding to musical "phrases". Such patterns of grouping and clustering can continue through many hierarchical levels, and many of these may be readily perceptible to an attentive listener.

Attempts to automatically identify the metric structure of musical pieces (Brown, 1993; Palmer & Krumhansl, 1990; Rosenthal, 1992; Steedman, 1997), to study rhythmic articulation (Repp, 1996), or to follow a score (Vantomme, 1995) often begin with a musical score or with a MIDI file transcription, and often are monophonic (Longuet-Higgins & Lee, 1984). This simplifies the rhythmic analysis in several ways. The pulse is inherent in the score, note onsets are clearly delineated, multiple voices cannot interact in unexpected ways, and the total amount of data to be analyzed is small compared to a CD-rate audio sampling of a performance of the same piece. Such simplification limits the applicability of the method to those pieces which are available in either standard musical notation or in MIDI data files. Since most pieces exist in audio recordings, it would be preferable to examine the metric structure directly from a sound recording. Besides the issue of availability, this would also provide a tool for those interested in comparing the rhythmic features of different performances of the same piece. Indeed, such a method is necessarily an analysis of the performance, and only incidentally of the underlying musical work.

This paper proposes a method of determining rhythmic structure from digitized audio that is based on the two ideas shown schematically in Figure 1. The first is a method of data reduction that filters the data into a number of channels corresponding to critical bands and then resamples so as to significantly decrease the amount of data. This is similar in spirit to periodicity detection models for pitch determination, and is described in detail in section 2. The second idea is to search the (reduced) sound data for periodicities. Originally, we exploited the Fourier Transform for this, but as the limitations of the Fourier analysis became apparent, we developed the Periodicity Transform (which is reviewed in section 3 and fully detailed in (Sethares & Staley, 1999)) that searches directly for periodicities (rather than frequencies, which are inverse periodicities). Thus this is more a method of signal processing at the level of sound waveforms than of symbol manipulation at the note level.

Accepted for publication: February 26, 2001.

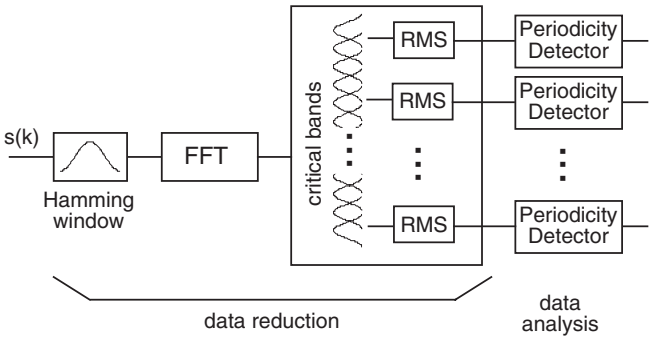

Fig. 1. The data reduction module partitions the sampled audio data into twenty-three 1/3-octave bands. The output is the energy in each band, sampled at an effective rate between 50 and 200 Hz. The data is then analyzed by searching for periodicities.

The rhythm finding approach is then examined in a number of cases in section 4, beginning with a simple polyrhythm and proceeding through "real" music. A significant limitation of the method is that it requires a steady tempo; pieces which change speed lack the kinds of periodicities that are easily detected. Accordingly, our focus is on music with a steady pulse. For example, the basic pulse, the 4/4 metric structure, and the grouping into phrases of the song "Jit Jive" by the (Bhundu Boys, 1988) are all clearly visible from the Periodicity Transform. Similarly, the pulse at 0.35 seconds, the 5/4 metric structure, and the larger phrasing at two and four bars are readily discernible in the analysis of Brubek's performance of "Take Five" (Dave Brubek Quartet, 1997). The method is also applied to a gamelan performance (Gamelan Gong Kebyar, 1988), indicating that the method is applicable cross culturally. The results are sensible, though we do not have the expertise to determine if they would agree with an Indonesian's assessment of the rhythmic structure of the piece. The final section discusses the relevant literature, suggests some possible extensions of the ideas, and concludes. An earlier version of this paper (Sethares, 1999) was presented at the FWO Research Society meeting on "Music and Timing Nets", held at the University of Ghent.

2 Psychoacoustically motivated data reduction

The kinds of periodicities associated with musical pulse, meter, and rhythm occur on time scales between tenths of a second and tens of seconds. The standard audio sampling rate of 44.1 KHz with its 22 KHz bandwidth contains significant redundancies, and some kind of data reduction is advantageous because it reduces the computational burden and because the method can be tailored to emphasize perceptually relevant aspects of the data.

Our fundamental assumption is that the periodicities associated with rhythmically important events are due to periodic fluctuations of the energy within various frequency bands. (Scheirer, 1998) argues this by creating a "modulated noise" signal from the amplitude envelopes of the outputs of a collection of filterbanks which often elicits "the same rhythmic percept" as the original audio signal. The effect of the basilar membrane is commonly modeled as a collection of bandpass filters that divide the sound into (roughly) 1/3 octave regions over the audio range of 20 Hz to 20 KHz. Our proposed data reduction mimics this by calculating the RMS energy in each 1/3 octave band at an effective rate of between 50 and 200 Hz. This range was chosen because periodic events faster than 25 Hz (the bandwidth of the 50 Hz sampling) would likely be perceived as pitched rather than rhythmic. Moreover, such an FFT-based energy accumulation is functionally analogous to the energy accumulation of the lowpass filtering of the rectification nonlinearity associated with inner hair cell models (Patterson & Moore, 1986).

The scheme is diagrammed in Figure 1 where the signal $s(k)$ is passed through a Hamming window and then transformed by an FFT. The following block calculates the RMS energy in each of the 1/3 octave bands covering the audio range. This is accomplished by summing the appropriate terms in the magnitude spectrum. Two parameters that need to be specified are the size of the window (which must match the size of the FFT) and the amount of overlap between successive blocks (which defines the "effective" sampling rate). These choices are discussed in detail in section 4.1.2.

This data reduction method is similar in concept to common models of the auditory system such as those by (Patterson, 1986) and (Leman, 1995). The goal in such models is to explain aspects of auditory perception such as the pitch of the missing fundamental, pitch shift due to inharmonic components, repetition pitch, detection of the pitch of multiple tones sounding simultaneously, and the pitch of interrupted noise. The present goal is to drastically reduce the amount of data in a relevant way. Clearly, it is possible to replace the FFT's and critical band filters of Figure 1 with a decimation filter bank structure or with a bank of bandpass filters without significantly changing the results.

The output of the data reduction scheme is a matrix of values that represent the energy of the signal in each frequency band (the rows) at times indicated by the columns. Figure 2, for instance, shows a spectrogram-style contour plot of this matrix for the song "Jit Jive" by the (Bhundu Boys, 1988). Observe the regular pulses in the bottom rows, which correspond to the strong bass and drum, and the relatively dense information in the upper regions. Each row will be searched for periodicities and these periodicities can be combined into an image that mimics the rhythmic structure of the piece. The next section details the method used to extract periodicity information from this matrix.

3 Periodicity transforms

Periodicity Transforms can be used to decompose sequences into a sum of small-periodic sequences (whenever possible) by projecting onto the "periodic subspaces" ρ_p. As the name suggests, this decomposition is accomplished directly in

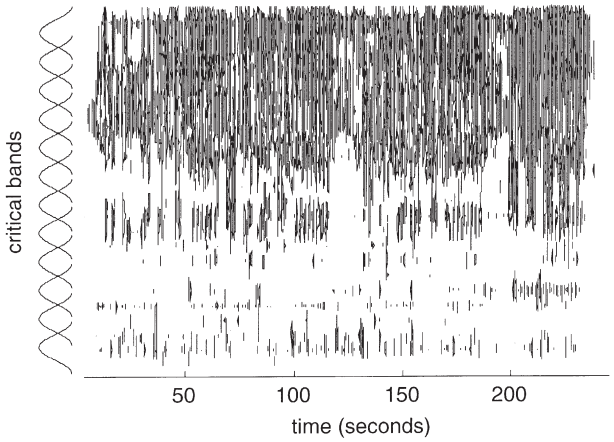

Fig. 2. The output of the data reduction module can be pictured in spectrogram-style where the energy in each critical band is indicated by the depth of the shading. Time evolves from left to right. This example shows the song "Jit Jive" by the Bhundu Boys. Each row of this data can then be examined for periodicities.

terms of small-periodic sequences and not in terms of frequency or scale, as do the Fourier and Wavelet Transforms (Strang & Nguyen, 1996). In consequence, the representation is linear-in-period, rather than linear-in-frequency or linear-in-scale. The Constant-Q Transforms of (Brown, 1991), which are linear in log-frequency, again use sinusoids as basis functions, and are fundamentally noninvertible.

Unlike most transforms, the set of basis vectors is not specified a priori, rather, Periodicity Transforms finds their own "best" set of basis elements. In this way, they are analogous to the approach of Karhunen-Loeve (Burl, 1989), which transforms a signal by projecting onto an orthogonal basis that is determined by the eigenvectors of the covariance matrix. In contrast, the periodic subspaces ρ_p lack orthogonality. This underlies the simplicity of representation possible when analyzing periodic sequences as well as its computational complexity. Like the process of autocorrelation, the Periodicity Transform is inherently a time-based operation, but it is not equivalent to the power spectrum. It is also conceptually related to "comb filtering", but it does not use the output of a single filter, rather, it uses the "prongs" of many "combs" to define periodic basis functions that together describe and/or decompose the signal. We have presented an in-depth investigation of the Periodicity Transform in (Sethares & Staley, 1999) and outline the method here.

3.1 Mathematical development

A sequence of real numbers $x(k)$ is called *p-periodic* if $x(k + p) = x(k)$ for all integers k. Let

ρ_p be the set of all p-periodic sequences, and
ρ be the set of all periodic sequences.

Observe that ρ_p is closed under addition since the sum of two sequences with period p is itself p-periodic. Similarly, ρ is closed under addition since the sum of x_1 with period p_1 and x_2 with period p_2 has period (at most) p_1p_2. Thus, with scalar multiplication defined in the usual way, both ρ_p and ρ form linear vector spaces, and ρ is equal to the span of the union of the ρ_p.

In order to project sequences in ρ onto ρ_p, consider the inner product from $\rho \times \rho$ into R defined by

$$\langle x, y \rangle = \lim_{k\to\infty} \frac{1}{2k+1} \sum_{i=-k}^{k} x(i)y(i) \tag{1}$$

for arbitrary elements x and y in ρ. This essentially finds the "average correlation" between the x and y over all possible periods. For the purposes of calculation, observe that if x is of period p_1 and y is of period p_2, then the sequence $x(i)y(i)$ is p_1p_2 periodic, and (1) is equal to the average over a single period, that is,

$$\langle x, y \rangle = \frac{1}{p_1 p_2} \sum_{i=0}^{p_1p_2-1} x(i)y(i). \tag{2}$$

The associated norm is

$$\|x\| = \sqrt{\langle x, x \rangle}. \tag{3}$$

As usual, the signals x and y in ρ are orthogonal if $\langle x, y \rangle = 0$, and two subspaces are orthogonal if every vector in one is orthogonal to every vector in the other. Unfortunately, no two subspaces ρ_p are orthogonal. In fact, they are not even linearly independent, since $\rho_1 \subset \rho_2$ for every p. More generally, $\rho_{np} \cap \rho_{mp} = \rho_p$ when n and m are mutually prime, which shows how the structure of the periodic subspaces reflects the structure of the integers.

3.2 Projection onto periodic subspaces

The primary reason for stating this problem in an inner product space is to exploit the projection theorem. Let $x \in \rho$ be arbitrary. Then a minimizing vector in ρ_p is an $x_p^* \in \rho_p$ such that

$$\|x - x_p^*\| \le \|x - x_p\| \quad \text{for all} \quad x_p \in \rho_p.$$

Thus x_p^* is the p-periodic vector "closest to" the original x. The projection theorem, which is stated here in slightly modified form, shows how x_p^* can be characterized as an orthogonal projection of x onto ρ_p.

Theorem 3.1 (The Projection Theorem) [Luenberger] *Let $x \in \rho$ be arbitrary. A necessary and sufficient condition that x_p^* be a minimizing vector in ρ_p is that the error $x - x_p^*$ be orthogonal to ρ_p.*

Since ρ_p is a finite (p-dimensional) subspace, x_p^* will in fact exist, and the projection theorem provides a way to calculate it. The optimal $x_p^* \in \rho_p$ can be expressed as a linear combination

$$x_p^* = \alpha_0 \delta_p^0 + \alpha_1 \delta_p^1 + \cdots + \alpha_{p-1} \delta_p^{p-1}$$

where the sequences δ_p^s for $s = 0, 1, 2, \ldots, p-1$ are the p-periodic orthogonal basis vectors

$$\delta_p^s(j) = \begin{cases} 1 & \text{if } (j-s) \bmod p = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

Let $x_{\tilde{N}}$ be the $\tilde{N}$-periodic sequence constructed from the first $\tilde{N} = p \lfloor N/p \rfloor$ elements of x, where N is the length of the data record and where $\lfloor n \rfloor$ represents the largest integer contained in n. Then the α_s can be calculated as

$$\alpha_s = \frac{1}{\lfloor N/p \rfloor} \sum_{n=0}^{\lfloor N/p \rfloor - 1} x_{\tilde{N}}(s + np). \tag{4}$$

Since there are p different values of s, the calculation of the complete projection x_p requires $\tilde{N} \approx N$ additions. A subroutine that carries out the needed calculations is available at our web site (MATLAB).

Let $\pi(x, \rho_p)$ represent the projection of x onto ρ_p. Then

$$\pi(x, \rho_p) = \sum_{s=0}^{p-1} \alpha_s \delta_p^s. \tag{5}$$

Clearly, when $x \in \rho_p$, $x = (x, \rho_p)$.

3.3 The Algorithms

Most standard transforms can be interpreted as projections onto suitable subspaces, and in most cases (such as the Fourier and Wavelet transforms), the subspaces are orthogonal. Such orthogonality implies that the projection onto one subspace is independent of the projection onto others. Thus a projection onto one sinusoidal basis function (in the Fourier Transform) is independent of the projections onto others, and the Fourier decomposition can proceed by projecting onto one subspace, subtracting out the projection, and repeating. Orthogonality guarantees that the order of projection is irrelevant. This is not true for projection onto nonorthogonal subspaces such as the periodic subspaces ρ_p. Thus the order in which the projections occur effects the decomposition, and the PT does not in general provide a unique representation. Once the succession of the projections is specified, however, then the answer is unique.

The Periodicity Transform searches for the best periodic characterization of the length N signal x. The underlying technique is to project x onto some periodic subspace giving $x_p = \pi(x, \rho_p)$, the closest p-periodic vector to x. This periodicity is then removed from x leaving the residual $\rho_p = x - x_p$ stripped of its p-periodicities. Both the projection x_p and the residual r_p may contain other periodicities, and so may be decomposed into other q-periodic components by further projection onto ρ_p. The trick in designing a useful algorithm is to provide a sensible criterion for choosing the order in which the successive p's and q's are chosen. The intended goal of the decomposition, the amount of computational resources available, and the measure of "goodness-of-fit" all influence the algorithm. The most useful of the algorithms are:

1. The "small to large" algorithm assumes a threshold T and calculates the projections $x_p = \pi(x, \rho_p)$ beginning with $p = 1$ and progressing through $p = N/2$. Whenever the projection contains at least T percent of the energy in x, then x_p is chosen as a basis element.
2. The " -best" algorithm maintains a list of the best periodicities and the corresponding basis elements. When a new (sub)periodicity is detected that removes more power from the signal than one currently on the list, the new one replaces the old, and the algorithm iterates.
3. The " -correlation" algorithm projects x onto all the periodic basis elements δ_p^s for all p and s, essentially measuring the correlation between x and the individual periodic basis elements. The p which contains the basis element with the largest (in absolute value) correlation is then used for the projection.
4. The "best-frequency" algorithm determines p by Fourier methods and then projects onto ρ_p.

Routines to calculate all of these variations are available on our website (MATLAB).

The "timing networks" of (Cariani, 1999) and (Leman et al., 1999) can be viewed as a (physiologically plausible) method of implementing the projections onto periodic subspaces, although there are some differences in the details of implementation. For instance, the cycles in the timing networks that integrate the data within each period have a characteristic time constant, while the actual projections proceed "all-at-once". The choice of a single period (the one with the largest magnitude) to represent the musical beat corresponds to the choice of using the " -Best" algorithm with = 1.

4 Examples

This section provides several examples where the data reduction technique coupled with the periodicity transforms provide reasonable explanations for certain levels of rhythmic structures. These begin with a simple three-against-two polyrhythm, which is explored in depth to show how various parameters of the model effect the results. The "effective sampling rate", which is a combination of the audio sampling rate and the offset or overlap parameter in the data reduction module, is an important parameter. A simple method of making a good choice is presented. This method is used throughout the remaining (more musical) examples. When the tempo of the performance is unsteady, the periodicity methods fail, highlighting the methods' reliance on a steady underlying pulse.

The following examples are taken from a variety of sources with a steady tempo: dance music, jazz, and an excerpt from a Balinese gamelan piece. The method is able to discern the pulse and several "deeper" layers of rhythmic

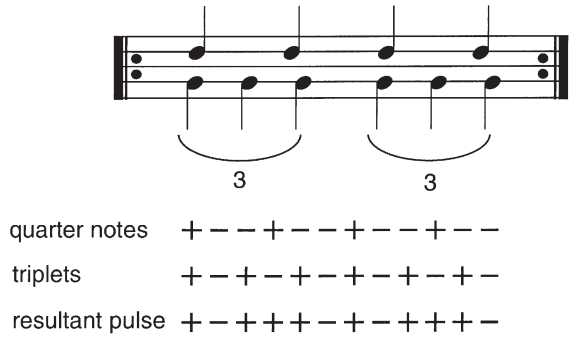

Fig. 3. The three vs. two polyrhythm written in standard musical notation. At a tempo of 120 beats per minute, two quarter notes occur in each second; three triplets occupy the same time span.

structure such as periodicities corresponding to measures and musical phrases.

4.1 Three against two polyrhythm

The three-against-two polyrhythm, which is notated in Figure 3, was played at a steady tempo of 120 beats per minute using two slightly different timbres, "wood block" and "sticks", for 15 seconds. This was recorded at 22.05 KHz and then downsampled using the data reduction technique of section 2 to an effective sampling rate of 140 Hz. The downsampled data was divided into 23 frequency bands to form the "audio matrix", which can be pictured as a spectrogram-style plot (such as Fig. 2) in which each row represents one of the frequency bands and each column represents a time instant of width 1/140 second. The rows can then be searched for periodicities.

4.1.1 Periodicity Transforms and the DFT

The standard signal processing approach to the search for periodicities is to use the DFT. This section compares the periodicity finding abilities of the DFT to that of the Periodicity Transforms for the simple Three against Two example. Figure 4 superimposes the magnitudes of the DFT of all 23 rows of the audio matrix. While not completely transparent, this plot can be meaningfully interpreted by observing that it contains two harmonic series, one based at 2 Hz (which represents the steady quarter note pulse) and the other at 3 Hz (which represents the steady triplet rhythm). These two "fundamentals" and their "harmonics" are emphasized by the upper and lower lattices which are superimposed on the plot. However, without the apriori knowledge that the spectrum of Figure 4 represents a three against two polyrhythm, this structure would not be obvious.

Applying the periodicity transforms to the rows of the audio matrix (in place of the DFT) leads to plots such as Figure 5. Here the Best Correlation method detects three periodicities, at $p = 72$ (which corresponds to the quarter note pulse), at $p = 48$ (which corresponds to the triplets), and at

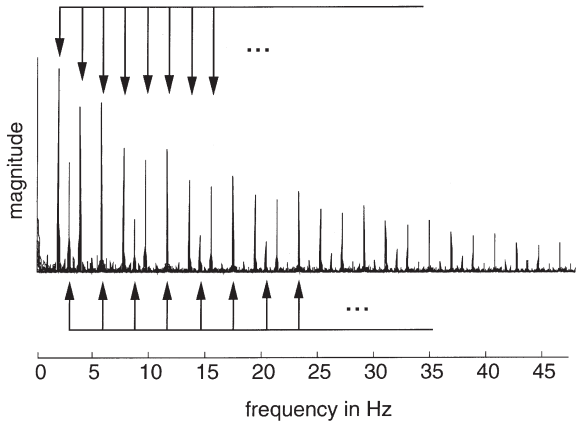

Fig. 4. The DFT's of the rows of the audio matrix for the Two vs. Three example are superimposed. The arrows of the upper ladder point to the "fundamental" at 2 Hz and its "harmonics" (which together represent the steady quarter note rhythm) while the arrows of the lower ladder point to the "fundamental" at 3 Hz and its "harmonics" (the steady triplet rhythm).

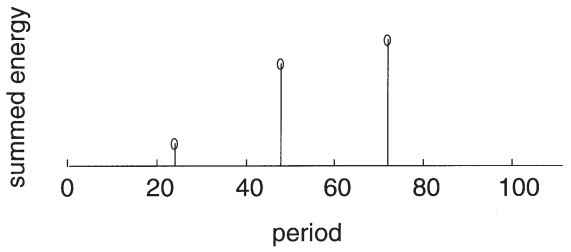

Fig. 5. The Best Correlation algorithm is used to detect periodicities in the Three vs. Two example. The horizontal axis is the period (each sample represents 1/140 second), and where the vertical axis shows the amount of energy detected at that period, summed over all the frequency bands. Periodicities are detected at $p = 72$ (the quarter notes), $p = 48$ (the triplets), and at $p = 24$ (the resultant "pulse" at 6 times per measure).

$p = 24$ (which corresponds to the speed of the resultant pulse, as in Figure 3). Clearly, it is much easier to interpret the periodicities in Figure 5 than the spectrum in Figure 4, and it was this kind of clarity that first inspired this project.

Two significant issues arise. First, how was the effective sampling rate chosen, and how crucial is this choice? Second, what happens when the underlying tempo (speed) of the performance varies? These are considered in the next two subsections.

4.1.2 Choice of effective sampling rate

While the Periodicity Transforms are, in general, robust to modest changes in the magnitudes of the data (Sethares & Staley, 1999), they are less robust to changes in the period. To see why, consider the simple case of a single detected periodicity at (say) $p = 10$. Suppose that the underlying signal were resampled (the effective sampling rate was changed) so

that the underlying periodicity now occurred "at" $p = 10.5$. Since the periodicity algorithms are designed to search for integer periodicities, they would detect a periodicity at two repetitions of $p = 10.5$ samples, that is, at $q = 21$ samples.

Since the effective sampling rate is

$$effsr = \frac{audio\ sampling\ rate}{offset},$$

nice clean pictures such as Figure 5 only occur for special values of the *offset*, the number of samples between the start of successive FFTs. For instance, the first periodicity analysis that we conducted of the Three against Two polyrhythm used an effective sampling rate of 132.43 Hz.[1] There were several strong periodicities detected, the largest occurring at 227 samples, which corresponds to a periodicity every 1.714 seconds. Since 227 a prime number, it is impossible for the periodicity algorithms to decompose it into sub-periods. By modifying the effective sampling rate so that (say) 240 samples (highly composite numbers, those with many factors, are particularly appropriate) occur within the same time span encourages the decomposition of this 1.714 second periodicity. Accordingly, we chose an effective sampling rate of 140 Hz, the rate actually used in the previous sections. To summarize this in a formula

$$new\ effsr = old\ effsr \frac{desired\ highly\ composite\ period}{detected\ period}. \qquad (6)$$

The other parameter that must be chosen is the width of the window and hence the time span of the FFT. Intuitively, this should correspond roughly to the length of an "instant", the smallest possible perceptible (rhythmic) event. Since events faster than 10 or 20 Hz are not percieved rhythmically, we chose to use 4 K FFTs, though this value did not appear to be critical. Thus the effective sampling rate can be adjusted independently of the window width, with the caveat that *effsr* should be less than half the window width. Values greater than this emphasize different segments of data differently.

4.1.3 Unsteady pulses

To investigate the effect of unsteady pulses, the 'same' Three against Two polyrhythm was re-recorded, but the tempo was increased by 5% every eight beats. This resulted in an increase in tempo of more than 25% over the course of the 15 seconds. As expected, the periodicity transforms were unable to cope with this variation. Figure 6 shows the detected periodicities for the Small-to-Large, Best Correlation, and -Best algorithms. Each of the algorithms has its own peculiarities, but none of the periodicities removes a significant percentage of the energy from the signal. The Small-to-Large algorithm detects hundreds of different periods,

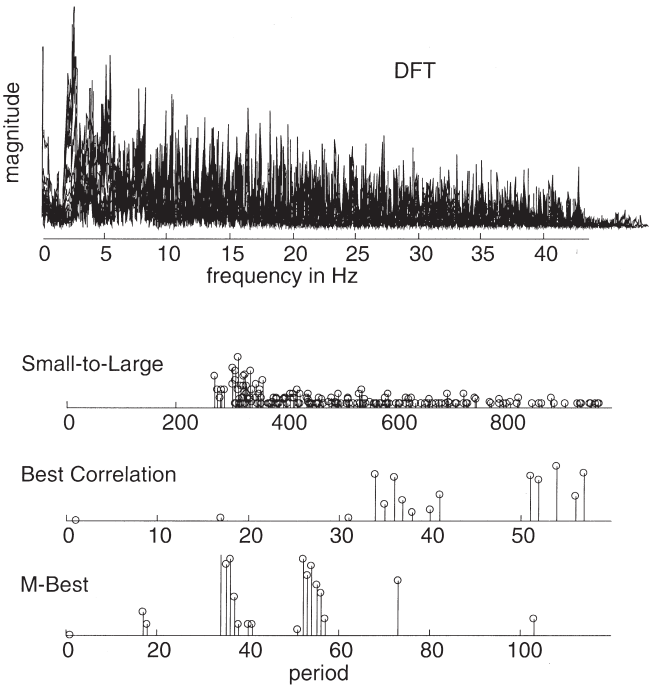

Fig. 6. The Three against Two example is performed with the tempo increasing 5% after every eight beats, for a total increase of about 25% over the 15 second duration. None of the periodicity algorithms detects periodicities that remove a significant fraction of the signal. The DFT also fails to reveal any significant structure.

most far longer than the "actual" signals of interest. Both the Best Correlation and -Best algorithms detect clumps of different periodicities between $33 < p < 40$ and $53 < p < 60$ which are in the midrange of the changing speed. One can view these clumps as decreasing in period, as should be expected from an increase in tempo, but it would be hard to look at these figures and to determine that the piece had sped up throughout. Similarly, the DFT (the top plot in Figure 6) is messy, giving little useful information.

This highlights what we consider to be the greatest limitation to the idea of detecting rhythm via periodicities, that rhythms can and do change their underlying pulse rate, whereas periodicities do not. We speculate that it may be possible to add a beat-tracking algorithm (perhaps similar to those of (Scheirer, 1998; Large Kolen, 1994), or (Goto & Muraoka)) to the "front end" of the periodicity method, and that this may allow a generalization of the present method to handle time variations more gracefully. Several possibilities are suggested in (Sethares, 1999).

4.2 Basis functions

Technically, the collection of all periodic subspaces forms a *frame* (Burrus et al., 1998), a more-than-complete spanning set. The Periodicity Transforms specify ways of sensibly handling the redundancy in this spanning set by exploiting some of the general properties of the periodic subspaces.

[1] This number coresponds to a FFT overlap of 333 samples.

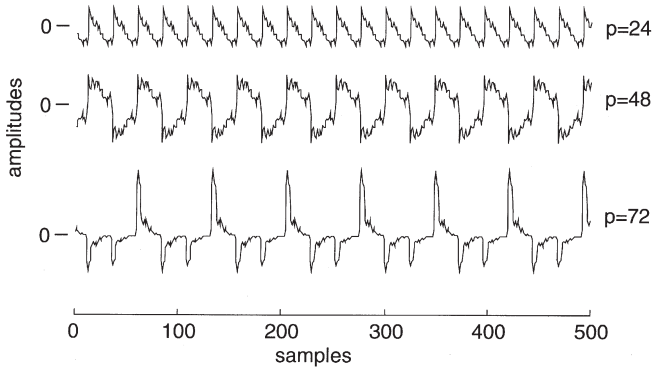

Fig. 7. The periodicity algorithms decompose the signal into periodic frame (basis) elements. Shown here are three frame elements from the "Best Correlation" analysis of the Three vs. Two example.

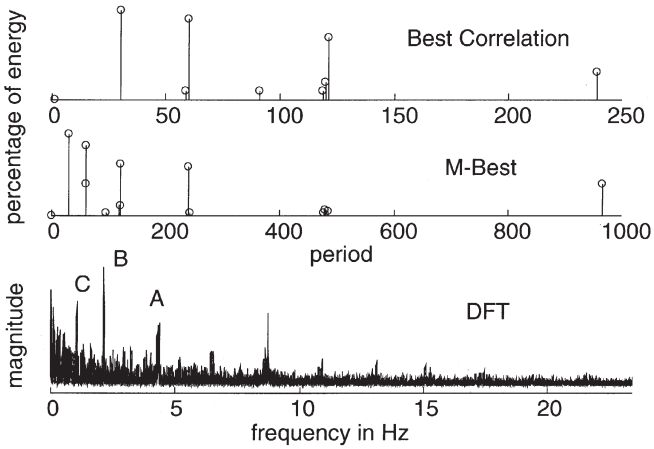

Fig. 8. Periodicity analysis of "Jit Jive" by the Bhundu Boys clearly reveals certain structural aspects of the performance. The major periodicities occur at 30 (which represents the pulse at 230 ms), 60 (two beats), 120 (the four beat measure which is 920 ms), 240 (the two measure phrase) and 960 (an eight bar phrase). The DFT shows three meaningful peaks: peak A represents the 230 ms pulse, and peak C represents the four beat measure. The intermediate peak B corresponds to the half measure.

When analyzing a signal, the elements of the resulting frame are analogous to the basis elements of the DFT, though they consist of general periodic functions rather than sinusoids. Figure 7 shows three periodic frame elements detected by the Best Correlation algorithm during the analysis of the Three against Two example of Figure 5. One row of the audio matrix (in this case, row number 15, which corresponds to frequencies near 2 KHz) is represented as the sum of these three functions, plus an error that contains no discernible periodicities. This suggests why the periodicity representation is so much simpler than the corresponding DFT – all three of these are very different from sinusoids, which are the basis functions used by the DFT. Expressing these three periodic functions in terms of sinusoids is tantamount to locating the regular lattices of Figure 4.

Figure 7 also suggests that the periodicity frame representation may prove useful in the creation of "new" rhythmic patterns that mimic the structure of analyzed rhythms. To the extent that amplitude fluctuations within frequency bands correlate well with perceived rhythmic structure, these may be useful as a method of creating new rhythms modeled after (or derived from) old. In the simplest case, this could be accomplished by passing a noise process through the appropriate frequency filters and adjusting the amplitude over time as dictated by the frame elements.

4.3 "True Jit"

The first analysis of a complete performance is of the dance tune "Jit Jive" performed by the (Bhundu Boys, 1988). Though the artists are from Zimbabwe, the recording contains all the elements of a dance tune in the Western "pop" tradition, featuring singers, a horn section, electric guitar, bass and drums, all playing in a rock steady 4/4 beat. The data reduced audio matrix is shown in Figure 2. As suggested in section 4.1.2, we performed a preliminary analysis at a convenient effective sampling rate (in this case 100.23 Hz, an overlap factor of 440), and observed that there was a major periodicity at $p = 46$ samples. We chose to redo the analysis at a sampling rate so that this same time was divided into 60 samples. Using 6, this gave an effective sampling rate of 130.73 Hz. This was used to generate Figure 8, which compares the outputs of the Best Correlation algorithm, the *M*-Best algorithm, and the DFT.

In all cases, the transforms are conducted on each of the 23 rows of the audio matrix independently, and then the results are added together so as to summarize the analyses in a single graph. Thus the Figure labeled DFT is actually 23 DFTs superimposed, and the Figure labeled *M*-Best represents 23 independent periodicity analyses. The vertical axis for the DFT is thus (summed) magnitude, while the vertical axes on all the periodicity transform Figures is the amount of energy contained in the basis functions, normalized by the total energy in the signal. Hence it depicts the percentage of energy removed from the signals by the detected periodicities.

There are five major peaks in the DFT analysis, of which three are readily interpretable in terms of the song. The peak marked *A* represents the basic beat or pulse of the song which occurs at 230 ms (most audible in the incessant bass drum) while peak *C* describes the four beat measure. The intermediate peak *B* occurs at a rate that corresponds to two beats, or one half measure.

The periodicity analysis reveals even more of the structure of the performance. The major periodicity at 30 samples corresponds to the 230 ms pulse. The periodicities at 60 and 120 (present in both periodicity analyses) represent the two beat half note and the four beat measure. In addition, there is a significant periodicity detected at 240, which is the two bar phrase, and (by the *M*-Best algorithm) at the eight bar phrase, which is strongly maintained throughout the song.

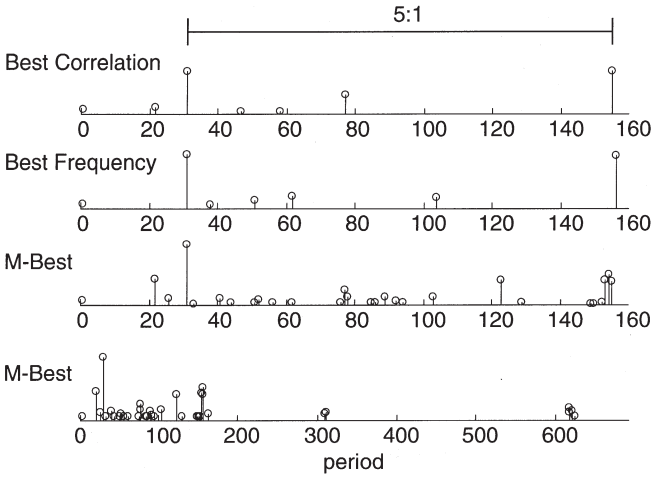

Fig. 9. Periodicity analysis of Brubeck's "Take Five" clearly reveals the five structure. The periodicity at 31 represents the beat, while the periodicity at (and near) 155 represents the five beat measure. The lower two plots show the periodicities detected by the -Best algorithm: the top is an expanded view of the bottom, which shows the larger periodicities near 310 (two measures) and near 620 (four measures). The piece is often notated in eight bar phrases.

Thus the periodicity transforms, in conjunction with the method of data reduction, allows an automated analysis of several levels of the rhythmic structure of the performance.

4.4 "Take Five"

Both of the previous pieces were rhythmically straightforward. Brubeck's "Take Five" (Dave Brubek Quartet, 1997) is not only in an uncommon time signature, but it also contains a drum solo and complex improvisations on both saxophone and piano. Figure 9 shows the periodicity analysis by the Best Correlation, Best Frequency, and -Best algorithms. All show the basic five to one structure (the period at 31 represents the beat, while the period at 155 corresponds to the five beat measure). In addition, the -Best algorithm finds periodicities at 310 (two measures) and at 620 (the four bar phrase). The piece would normally be notated in eight bar phrases, which is the length of both of the melodies. As is clear, there are many more spurious peaks detected in this analysis than in the previous two, and likely this is due to the added complexity of the performance. Nonetheless, the periodicity method has identified two of the major structural levels.

4.5 Gamelan Gong Kebyar

The final analysis is the Baris war dance, which is a standard piece in the Balinese gamelan tradition called (Gong Kebyar, 1988). The performance begins softly, and experiences several changes from soft to loud, from mellow to energetic, but it does maintain a steady underlying rhythmic pulse

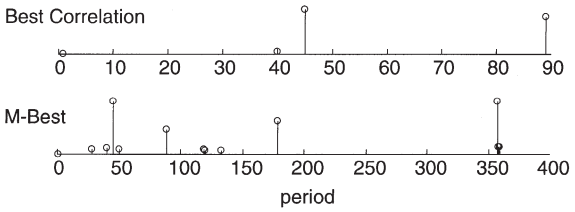

Fig. 10. Periodicity analysis of the Baris war dance, a piece in the Gong Kebyar style of Balinese gamelan, clearly reveals certain structural aspects of the performance. The major periodicities occur at 45 (which represents the pulse), 90, 180, and 360 (two, four, and eight times the beat). The largest of these is most prominent in the bass, and the periodicity at 360 is the rate at which the largest gong is struck.

throughout. This beat is alternately carried by the drum and the various kettle instruments (the bonangs and kenongs), and the bass gong is struck steadily after each eight beats.

These rhythmic elements are clearly reflected in the periodicity analysis. Figure 10 shows the pulse at period 45, and other periodicities at two, four, and eight times the pulse. As mentioned above, the large gong strikes each eight beats throughout. Such regular punctuation appears to be a fairly generic character of much of the Gong Kebyar style (Sorrell, 1990). Thus the periodicity analysis is applicable cross culturally, though we do not have the expertise to conduct in depth analyses of pieces within the gamelan traditions.

5 Discussion

The analysis of musical rhythms is a complex task. As noted in (Longuet-Higgins & Lee, 1984), "even the most innocuous sequences of note values permit an unlimited number of rhythmic interpretations". Most proposals for rhythm finding algorithms use "interonset" intervals as their basic elements (for instance (Desain, 1992; Goto & Muraoka)), which presupposes the accurate detection of note onsets and offsets. Our method, by ignoring the "notes" of the piece and focusing on the finding and grouping of pulses, bypasses (or ignores) this element of rhythmic interpretation. This is both a strength and a weakness. Without a score, the detection of "notes" is a nontrivial task, and errors such as missing a note (or falsely detecting a note that is not actually present) can seriously bias the detected rhythm. Since our method does nothing analogous to detecting notes, it cannot make such mistakes. The price, of course, is that the explanatory power of a note-based approach remains unexploited.

Steedman (1977) begins by stating, "When people beat time to a melody, they perform an act of musical understanding analogous to the understanding of an utterance of their native language. In the sound they hear, the notes that fall on the beat are not explicitly marked . . ." He then concludes that his proposed meter analysis program (which analyzes a "list of notes"), "is intended to constitute a

psychological theory of our perception of meter in melody." We believe that a psychological theory of the perception of meter need not operate at the level of notes. Indeed, people directly perceive sound waves, and the act of detecting notes, pulses, and meters are acts of cognition. To the extent that our method is capable of finding pulses and meters within certain musical performances, the "note" and "interonset" levels of interpretation are not a necessary component of rhythmic detection. This reinforces the argument in (Scheirer, 1998) where metric perceptions are retained by a signal containing only noisy pulses derived from the amplitude envelope of audio passed through a collection of critical band filters.

There is a long history of study of the various features of a sound segment that influence the perceived rhythm. As early as 1911 (Woodrow, 1911), investigated the role of pitch in rhythm and this same concern is an active area of research today (Singh, 1997). Though our method does not attempt to decode pitches (which are closely tied to a note level representation) it is not insensitive to frequency information. Indeed, one of the features of the psychoacoustic data reduction is that the data in each of the "critical bands" is processed independently. This allows the timbre (or spectrum) of the sounds to directly influence the search for appropriate periodicities, in a way that is lost if only interonset interval encoding is used.

The pulse finding technique of (Goto & Muraoka) is subtitled "a real-time beat tracking system for audio signals." The goal is to track "the temporal position of quarter notes," and the approach uses a multi-agent expert system. Each agent predicts (using a windowed autocorrelation) when the next beat will occur based on onset time vectors that are derived from a frequency decomposition. Unfortunately, the system uses a significant amount of high level information, including a requirement that the piece be in 4/4 time and that the tempo is constrained to be between 61 and 120 beats per minute. Within these limitations, the results are striking, and the system can work in real time. It cannot, however, provide information about large scale structures such as measures and phrases, since such knowledge is assumed a priori. Similarly, the score following paradigm of (Vantomme, 1995) includes a matching algorithm that detects note events and then locates them within a stored score. The most straightforward way to locate the pulse using the periodicity approach is to pick the most powerful periodicity[2] within the window of 250 ms to 1.0 second (perhaps weighted by a distribution determined by the "style" of music as in (van Noorden & Moelants, 1999)), which is the range over which most musical pulse rates occur. If there are two (roughly equal) periodicities within this range, then either would be a good candidate for the beat. It would be an interesting experiment to see if such a simple method can reproduce results such as in (Handel, 1984) where ambiguity is reported in tapping experiments.

[2] That removes the most power from the audio signal.

While our work has been stated in terms of identifying periodicities that correspond to musical pulses and phrases given a complete performance, there is no reason why the approach cannot be used in a real time system. Indeed, because the algorithm looks for periodicities directly, it is straightforward to operate the algorithm on a short time scale, which can be lengthened as more data becomes available. While operating on only a small segment of the data will undoubtedly involve noisier estimates of the periods, the detected periodicities can be readily projected into the future to make predictions of the expected behavior of the signal. This is reminiscent of the "expectancy" approach of (Desain, 1992) (which operates at the note level using interonset intervals) and is related to the ideas of (Jones, 1989) for whom "attention itself is a dynamic, many leveled affair based on nested internal rhythms." Viewing the detected periodicities as internal rhythms (which are "nested" whenever the periods are commensurate) provides a possible framework in which to formalize some of Jones' ideas.

Brown (1993) approached the determination of meter from a similar philosophical point of view, using an autocorrelation method on note level data derived from musical scores. While this may be different in detail, the autocorrelation is closely related to the power spectral density, that is, to the magnitude of the Fourier Transform. It is therefore not surprising that the results of applying the autocorrelation closely match the kinds of results we achieved using the DFT. Indeed, Brown's graphs look quite similar to our Figure 4. Analogously, wavelet transforms can be applied to the determination of meter (Todd et al., 1999), though the issues surrounding the interpretation of wavelet amplitude and phase diagrams have not been fully explored.

6 Summary and concluding remarks

Periodicity Transforms are designed to locate periodicities within a data set by projecting onto a set of (nonorthogonal) periodic subspaces. This can be applied to musical performances after an appropriate method of data reduction which decomposes the sound into the energy in each of the critical bands. Several examples showed that the method can detect periodicities that correspond to the pulse, the measure, and even larger units such as phrases. The method is not restricted to certain time signatures, meters, or rhythms, but it does require that the tempo be steady, and finding ways to accommodate tempo variations is an important area for further study. This reinforces the argument that rhythmic aspects of a musical piece can be determined using only periodic fluctuations of energy within the critical bands, and without the explicit consideration of notes or interonset intervals.

Acknowledgments

The authors would like to thank Ian Dobson for periodic conversations about rhythmic matters.

References

Benjamin, W.E. (1984). A theory of musical meter. *Music Perception*, 355–413.

Bhundu Boys (1988). *True Jit*, Mango Records CCD 9812.

Brown, J. (1991). Calculation of a constant Q spectral transform. *J. Acoustical Society of America*, *89*, 425–434.

Brown, J. (1993). Determination of the meter of musical scores by autocorrelation. *J. Acoustical Society of America*, *94*(4), 1953–1957.

Dave Brubek Quartet (1997). *Time Out*, Sony/Columbia, CK65122.

Burl, J.B. (1989). Estimating the basis functions of the Karhunen-Loeve transform. *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 37, No. 1 (pp. 99–105) Jan. 1989.

Burrus, C.S., Gopinath, R.A., & Guo, H. (1998). *Wavelets and Wavelet Transforms*. Prentice Hall.

Cariani, P. (1999). Neural computations in the time domain. *Proc. of the FWO Research Society on the Foundations of Music*, "Music and Timing Networks", Ghent, Belgium.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception*, *9*(4), 439–454.

Gamelan Gong Kebyar (1988). *Music from the Morning of the World*, Elektra/Nonesuch 979196-2.

Goto, M. & Muraoka, Y. Beat tracking based on multiple-agent architecture. ICMAS-96, 103–110.

Handel, S. & Oshinsky, J.S. (1981). The meter of syncopated auditory polyrhythms. *Perception and Psychophysics*, *30*(1), 1–9.

Handel, S. & Lawson, G.R. (1983). The contextual nature of rhythmic interpretation. *Perception and Psychophysics*, *34*(2), 103–120.

Handel, S. (1984). Using polyrhythms to study rhythm. *Music Perception*, *1*(4), 465–484.

Jones, M.R. & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

Large, E.W. & Kolen, J.F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*, 177–208.

Leman, M. (1995). *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*. Berlin, Heidelberg: Springer-Verlag.

Leman, M., Tanghe, K., Moelants, D., & Carreras, F. (1999). Analysis of music using timing networks with memory: implementations and preliminary results. *Proc. of the FWO Research Society on the Foundations of Music*, "Music and Timing Networks", Ghent, Belgium, Oct. 1999.

Longuet-Higgins, H.C. & Lee, C.S. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, *1*(4), 424–441.

Luenberger, D.G. (1968). *Optimization by Vector Space Methods*. NY, John Wiley and Sons, Inc.

Martens, J.P. (1982). A new theory for multitone masking. *Journal of the Acoustical Society of America*, *72*(2), pp. 397–405.

van Noorden, L. & Moelants, D. (1999). Resonance in the perception of musical pulse. *J. New Music Research*, *28*(1).

Palmer, C. & Krumhansl, C. (1990). Mental representations for musical meter. *J. Exp. Psychology: Human Perceptual Performance*, *16*, 728–741.

Patterson, R.D. & Moore, B.C.J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. In: B.C.J. Moore (Ed.), *Frequency Selectivity in Hearing*, London: Academic Press.

Povel, D.J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440.

Repp, B.H. (1996). Patterns of note onset asynchronies in expressive piano performance. *J. Acoustical Society of America*, *100*(6), 3017–3030.

Rosenthal, D. (1992). Emulation of rhythm perception. *Computer Music Journal*, *16*(1).

Scheirer, E.D. (1997). Using musical knowledge to extract expressive performance information from audio recordings. In: H. Okuno & D. Rosenthal (Eds.), *Readings in Computational Auditory Scene Analysis*. Lawrence Erlbaum, NJ.

Scheirer, E.D. (1998). Tempo and beat analysis of acoustic musical signals. *J. Acoustical Society of America*, *103*(1), 588–601.

Sethares, W.A. (1999). Automatic detection of meter and periodicity in musical performance. *Proc. of the FWO Research Society on the Foundations of Music*, "Music and Timing Networks", Ghent, Belgium, Oct. 1999.

Sethares, W.A. (1997). *Tuning, Timbre, Spectrum, Scale*, Springer-Verlag.

Sethares, W.A. & Staley, T. (1999). The periodicity transform. *IEEE Trans. Signal Processing*, *47*(11).

Singh, P. (1997). The role of timbre, pitch, and loudness changes in determining perceived metrical structure. *J. Acoustical Society of America*, *101*(5) Pt 2, 3167.

Sorrell, N. (1990). *A Guide to the Gamelan*. London: Faber and Faber Ltd.

Smith, K.C. & Cuddy, L.L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *J. Experimental Psychology*, *15*(3), 457–471.

Steedman, M.J. (1977). The perception of musical rhythm and metre. *Perception*, *6*, 555–569.

Strang, G. & Nguyen, T. (1996). *Wavelets and filter banks*, Wellesley College Press: Wellesley, MA.

Todd, N.P.M., O'Boyle, D.J., & Lee, C.S. (1999). A sensory-motor theory of rhythm, time perception and beat induction. *J. of New Music Research*, *28*(1), 5–28.

Vantomme, J.D. (1995). Score following by temporal pattern. *Computer Music Journal*, *19*(3), 50–59.

MATLAB routines for the calculation of the Periodicity Transforms are available at our website http://eceserv0.ece.wisc.edu/'sethares/

Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review*, *11*.

Author Biographies

Tom Staley is a Ph.D. candidate in Science & Technology Studies at Virginia Polytechnic Institute, where he also serves as an instructor in Materials Science and Engineering and the Program in Humanities, Science, and Technology. In some of his spare time, he composes music, builds experimental instruments, and investigates topics in music theory.

William A. Sethares received the B.A. degree in mathematics from Brandeis University, Waltham, MA and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY.

He is currently Associate Professor in the Department of Electrical and Computer Engineering at the University of Wisconsin, where he teaches courses in signal processing, acoustics, and communications. He finds it fascinating that synchronization and clocking issues in communication receivers involve many of the same issues as beat tracking in musical analysis. His research interests include adaptation and learning in signal processing, communications, and acoustics, and is author of *Tuning*, *Timbre*, *Spectrum*, *Scale* (Springer).