Detecting Meter in Recorded Music

Joseph E. Flannick, Rachel W. Hall, and Robert Kelly
Dept. of Math and C. S.
Saint Joseph's University
5600 City Avenue
Philadelphia, PA 19131, USA
E-mail: rhall@sju.edu

Abstract

Most pieces of popular dance music feature repeated patterns of rhythmic accents, or beats. We use the Discrete Fourier Transform and the Periodicity Transform (Sethares and Staley, 1999) to identify the primary rhythmic content of a piece of popular music. Before applying the transforms, we reduce the data by filtering out pitch. We use the data reduction method proposed by Scheirer (1998), which separates recorded music into bands of pitches, roughly half an octave each, and extracts the pattern of energy bursts in each band. After applying the DFT and PT, we find that the basic rhythmic structure, or meter, of the piece we analyzed is reflected in the relationship between periodic accents made by low- and high-pitched instruments. We have written MATLAB algorithms that implement these methods. Audio examples are available at `http://www.sju.edu/~rhall/Bridges`.

1. Introduction

Even to the untrained ear, it is quite apparent that mathematics is at play in music. As one delves deeper, one realizes that not only is math involved in music, but that there is an inextricable connection between the world of mathematics and every single element of music—whether it be the theory of sound waves, the physics of instruments, or the structure of musical rhythm. In this paper we demonstrate a method for detecting the underlying rhythmic structure of a piece of popular music.

Periodic phenomena occur in music at different levels. Musical instruments produce vibrations, which in turn create periodic variations in air pressure, producing sound. The limits of humanly audible sound are 20 to 20,000 Hz (cycles per second). The sounds produced by musical instruments are more complex than pure sine waves, but generally one frequency or range of frequencies predominates. *Pitch* is the relative highness or lowness of an audible sound, determined by the frequency of the vibration producing the sound. We use the word "pitch" to designate audible frequencies above 20 Hz. In popular music, especially dance music accompanied by drums, there are also heavy periodic rhythmic accents, produced by bursts of energy within an audible sound—that is, changes in the amplitude of the sound wave. Our project involves the study of these periodic accents, or beats. Pitch and beats operate on a different scale—pitches are typically measured in hundreds or thousands of cycles per second, while beats are measured in tens to hundreds of cycles per minute. However, both pitches and beats are periodic phenomena, and so we can borrow some of the traditional methods used to study pitch in our study of rhythm.

We use mathematical techniques to detect the underlying rhythmic organization, or meter, of a piece of recorded popular music. Meter is reflected in the strength of accents that are placed on each beat. The basic unit of time is the measure, which is subdivided into a number of equal beats, each of which may be further subdivided into half notes, quarter notes, and so on. Accents are used to mark the divisions and have a hierarchy stemming from the order in which they are created. Each time a division is made, the accents on newly created beats are weaker. Music contains many high and low sounds produced by many different instruments and voice. Despite all of this activity, our ears can almost always detect the meter of a musical work. Why is it that the meter stands out so easily? Our work sheds some light on this question.

2. Sampling and Data Reduction

When we hear music played the waveform is continuous. In order to produce a CD, which cannot hold an infinite amount of data, the continuous waveform must be sampled, meaning that a discrete set of values is taken from the original waveform at regular time intervals. The more samples that are taken per second, the more closely the discrete function resembles the continuous function. We assume that the sampling is sufficiently frequent that we don't lose much audible information by using the discrete approximation. Typically, the original waveform is sampled at 44.1 kHz (44,100 Hz), meaning that 44,100 sample points are taken per second.[1] Since a sampled sound is discrete, we can use MATLAB to analyze it—think of the sampled music as a very long vector. At this sampling rate, a one-minute song contains 2.646×10^6 samples! Performing any analysis on this many samples requires lengthy computations. However, we have established that meter is a low frequency component of music, so the pitch may be filtered out without compromising the rhythmic information.

We followed an algorithm proposed by Eric D. Sheirer [8], and improved upon its implementation in MATLAB by Sethares and Staley [10]. Our algorithm separates recorded music into 21 frequency bands, of roughly half an octave each, and extracts the energy in each band. The output of this process is an *audio matrix*. The 21 rows of this matrix represent the changing energy in each band, and the columns provide the time dimension. The algorithm serves two functions: to strip the signal of pitch, while still preserving the rough relationships between high and low sounds, and to reduce the amount of data, thus making the DFT and PT computations shorter. Consider an arbitrary signal, s, sampled at 44.1 kHz. To begin, s is stripped to a mono signal (that is, a vector). The algorithm moves along s from beginning to end taking *windows*—vectors consisting of some fixed number of consecutive entries from s. We overlap our windows so that prominent frequencies are not split between windows. Filters are used to split each window into 21 frequency bands (think of the filters as a sort of prism). The energy in that band, defined to be the square root of the sum of the squared magnitudes of the Fourier coefficients, is then computed. The output is a column vector containing only 21 entries; the first row contains the energy in the lowest pitch band, and the last row contains the energy in the highest pitch band. This process of taking a windows continues until we reach the end of the signal. The end result is an audio matrix of size $21 \times (\textit{number of windows})$. Each row in the matrix represents the variations in energy in one pitch band for the duration of the piece. See Figure 1 for the image of an audio matrix.

3. The Discrete Fourier Transform

Once the data is reduced, we can apply the discrete Fourier transform (DFT), which decomposes the signal in each band into a sum of discrete sinusoids. A graph of the magnitudes of the coefficients of these sinusoids gives us information about integer frequencies in the signal; a spike at a particular frequency implies that the frequency is prominent in the signal. Although musical rhythm is rarely periodic, the pattern of accents in most popular dance music is "periodic enough" to be analyzed by the DFT, because popular music features repeated drum beats and is recorded to a metronome track that keeps the drummer perfectly in time. The DFT is typically applied to periodic signals, but, in practice, we can still gain relevant information from approximately periodic signals. Finally, we compare the DFTs of each frequency band to determine the meter of the piece.

3.1. Details of the DFT. Let's investigate discrete periodic functions of a fixed period N. Any discrete periodic function is of the form $f[n]$ where $n \in \mathbf{Z}$ and $f[n+N] = f[n]$ for some integer N, which is

[1]The reason for this high number of samples is the Nyquist Theorem, which states that a continuous waveform can be reconstructed from discrete samples as long as its frequency is less than half the sampling rate. Since the limit of audible sound is 20 kHz, we must sample at more than 40 kHz.

referred to as a period of f. We can write any N-periodic discrete function f in the form:

$$f[n] = \sum_{k=0}^{N-1} F[k]\, e^{2\pi ikn/N}, \text{ where } F[k] = \frac{1}{N} \sum_{n=0}^{N-1} f[n]\, e^{-2\pi ikn/N}.$$

The representation above is called the Discrete Fourier Transform (DFT).[2] The function $F[k]$ gives the coefficients of the sinusoids present in the musical sound. The magnitude $|F[k]| = (F[k]\overline{F[k]})^{1/2}$ of each coefficient is the strength of each frequency component.

3.1.1. Example. Let $f[n]$ be the discrete 4-periodic function defined by $f[0] = a$, $f[1] = b$, $f[2] = c$, $f[3] = d$, and $f[n+4] = f[n]$. Then $f[n] = F[0] + F[1]e^{\pi in/2} + F[2]e^{\pi in} + F[3]e^{3\pi in/2}$, where $F[0] = \frac{1}{4}(a+b+c+d)$, $F[1] = \frac{1}{4}(a+bi-c-di)$, $F[2] = \frac{1}{4}(a-b+c-d)$, and $F[3] = \frac{1}{4}(a-bi-c+di)$. Let's examine these coefficients more closely. We can see that $|F[1]| = |F[3]| = (1/4)((a-c)^2 + (b-d)^2)^{1/2}$ So, if f is 2-periodic (that is, $a = c$ and $b = d$), then $F[1] = F[3] = 0$. Likewise, if f is approximately 2-periodic, that is if $a \sim c$ and $b \sim d$, then $F[1]$ and $F[3]$ are relatively close to zero.

As seen in this example, the DFT identifies prominent frequencies in a signal. We graph the magnitudes of these coefficients to get a clear picture of the different frequencies present in the signal. Observe that if f is a real-valued function, $F[N-k] = \overline{F[k]}$, and hence $|F[k]| = |F[N-k]|$, so the graph of $|F[k]|$ is symmetric with respect to $k = N/2$, and therefore it is sufficient to graph the magnitudes of the first $N/2$ values of f (see Figure 2 for an example).

3.2. Analyzing Musical Rhythm Using the DFT. The DFT is a standard tool for analyzing pitch. We can also employ the capabilities of the DFT to analyze rhythm. By removing the pitch, we are left with the rhythmic components of the musical piece. When the DFT is applied to these components, much information about the rhythmic structure of the piece is revealed, including the relative strength of repeated beats in the song (see Section 5 for an example). However, the DFT has significant limitations in analyzing rhythm. It detects integer frequencies—but when studying rhythm, the period, and not the frequency, is significant. Moreover, the DFT makes it difficult to observe those periodic rhythmic structures, such as phrases, that are not as frequent as the beat. The Periodicity Transform, proposed by Sethares and Staley [9], addresses these limitations by searching for integer periods.

4. The Periodicity Transform

Let x be our signal. The idea behind Sethares and Staley's Periodicity Transform (PT) is to define a metric on the space of periodic vectors and find x^*, the closest periodic vector to x with respect to this metric. By subtracting x^* from x, we get a residual vector r. We then search for the closest periodic vector to r, subtract that vector from r, and the process is repeated. Finally, we have a decomposition of $x = x^* + r_1^* + r_2^* + \ldots$ into periodic vectors. Like the basis elements in the DFT, these periodic vectors give us an idea of the relative strengths of periodicities within x.

4.1. The space of p-periodic vectors. Recall that $x[k]$, $k \in \mathbf{Z}$ is p-periodic if $x[k+p] = x[k]$ for all p. Let $\mathcal{P}$ = all periodic vectors and let $\mathcal{P}_p$ = all p-periodic vectors. Notice that both $\mathcal{P}$ and $\mathcal{P}_p$ form vector spaces since they are both closed under addition and scalar multiplication.

We now need to define a basis vector for $\mathcal{P}_p$. The following sequence is a fitting choice:

$$\delta_p^s[i] = \begin{cases} 1, & \text{if } (i-s) = 0 \pmod p \\ 0, & \text{otherwise} \end{cases}$$

[2] Although upon first glance, the DFT equation may not appear to yield a periodic function, Euler's formula ($e^{i\theta} = \cos\theta + i\sin\theta$) can be used to rewrite it as a sum of sines and cosines. If f is a real-valued function, the imaginary parts cancel.

For example, $\delta_0^4 = \ldots, 1, 0, 0, 0, 1, 0, 0, 0 \ldots$. Note that δ_1^4, δ_2^4 and δ_3^4 will all just be shifts of δ_0^4.

Consider the following product:

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

for elements x, y in $\mathcal{P}$. We claim that this is an inner product on $\mathcal{P}$. The limit will always exist since if $x \in \mathcal{P}_{p_1}$ and $y \in \mathcal{P}_{p_2}$, $x[i]\, y[i] \in \mathcal{P}_{p_1 p_2}$ since it is now $p_1 p_2$-periodic. The inner product now becomes $\langle x, y \rangle = \frac{1}{p_1 p_2} \sum_{i=0}^{p_1 p_2 - 1} x[i]\, y[i]$ or the average of the $p_1 p_2$-periodic vector over a single period. We now have a way to measure distance: $||x|| = \langle x, x \rangle^{1/2}$.

Signals x and y in an inner product space 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. Notice, however, that no two periodic subspaces $\mathcal{P}_p$ are orthogonal since $\mathcal{P}_1 \subset \mathcal{P}_p$ for every p. Moreover, $\mathcal{P}_{np} \cap \mathcal{P}_{mp} = \mathcal{P}_p$ when n and m are mutually prime. As an example, take $\mathcal{P}_4$ and $\mathcal{P}_6$. If $x \in \mathcal{P}_4 \cap \mathcal{P}_6$, then $x \in \mathcal{P}_4$ and $x \in \mathcal{P}_6$. For this to be true, x must also be 2-periodic (indeed, $p = 2$ and $n = 2$, $m = 3$).

4.2. Projection onto p-periodic subspaces. The following result is stated and proved in [9].

Theorem 1 (Sethares and Staley) *Let $x \in \mathcal{P}$ be an arbitrary signal. A* minimizing vector *in $\mathcal{P}_p$ is an $x_p^* \in \mathcal{P}_p$ such that $||x - x_p^*|| \leq ||x - x_p||$ for all $x_p \in \mathcal{P}_p$. The vector x^* given by*

$$x^* = \alpha_0 \delta_p^0 + \alpha_1 \delta_p^1 + \ldots + \alpha_{p-1} \delta_p^{p-1},$$

where $\alpha_i = p\langle x, \delta_p^i \rangle$ for $0 \leq i \leq p-1$ is the unique minimizing vector in $\mathcal{P}_p$.

We will use the notation $\pi(x, \mathcal{P}_p)$ to represent the projection of x onto $\mathcal{P}_p$.

4.2.1. Example. Let $x = \ldots 1, 1, 0, 1, 1, 4, 0, 2, \ldots \in \mathcal{P}_8$. The projection of x onto $\mathcal{P}_2$ is the vector $x_2^* = \ldots \frac{1}{2}, 2, \frac{1}{2}, 2, \frac{1}{2}, 2, \frac{1}{2}, 2, \ldots$ and the residual is $r_2 = x - x_2^* = \ldots \frac{1}{2}, -1, -\frac{1}{2}, -1, \frac{1}{2}, 2, -\frac{1}{2}, 0, \ldots$ The projection of x onto $\mathcal{P}_4$ is $x_4^* = \ldots 1, \frac{5}{2}, 0, \frac{3}{2}, 1, \frac{5}{2}, 0, \frac{3}{2}, \ldots$, and the residual is $r_4 = x - x_4^* = \ldots 0, -\frac{3}{2}, 0, -\frac{1}{2}, 0, \frac{3}{2}, 0, \frac{1}{2}, \ldots$. Notice that projecting r_4 onto $\mathcal{P}_2$ gives the zero vector. This makes sense, because r_4 is the original signal with all 4-periodic subsignals removed. All 4-periodic signals are necessarily 2-periodic, and so $\pi(r_4, \mathcal{P}_2) = 0$. In fact, we have the following theorems, due to Sethares and Staley [9]:

Theorem 2 (Sethares and Staley) *Let $r_p = x - \pi(x, \mathcal{P}_p)$ be the residual after projecting x onto $\mathcal{P}_p$ and $r_{np} = x - \pi(x, \mathcal{P}_{np})$ be the residual after projecting x onto $\mathcal{P}_{np}$. Then $r_{np} = r_p - \pi(r_p, \mathcal{P}_{np})$.*

Theorem 3 (Sethares and Staley) *Let x be a periodic vector and p and n be positive integers. Then*

$$\pi(x, \mathcal{P}_p) = \pi(\pi(x, \mathcal{P}_p), \mathcal{P}_{np}) = \pi(\pi(x, \mathcal{P}_{np}), \mathcal{P}_p).$$

Corollary 1 (Sethares and Staley) *The projection of r_{np} onto $\mathcal{P}_p$ is the zero vector.*

Theorem 3 shows that the order of projection of a periodic vector x onto subspaces $\mathcal{P}_p$ and $\mathcal{P}_{np}$ does not matter, since $\pi(x, \mathcal{P}_{np})$ is an average over every npth entry in x.

It is advantageous at this point to take a step back and think about what it is we are actually doing here. When we project our signal x onto $\mathcal{P}_p$, we are stripping it of all its p-periodic components. However, the residual may still have other relevant periodicities, and so we should project this "new signal" onto other subspaces (perhaps $\mathcal{P}_q, \mathcal{P}_s \ldots$) to extract them as well.

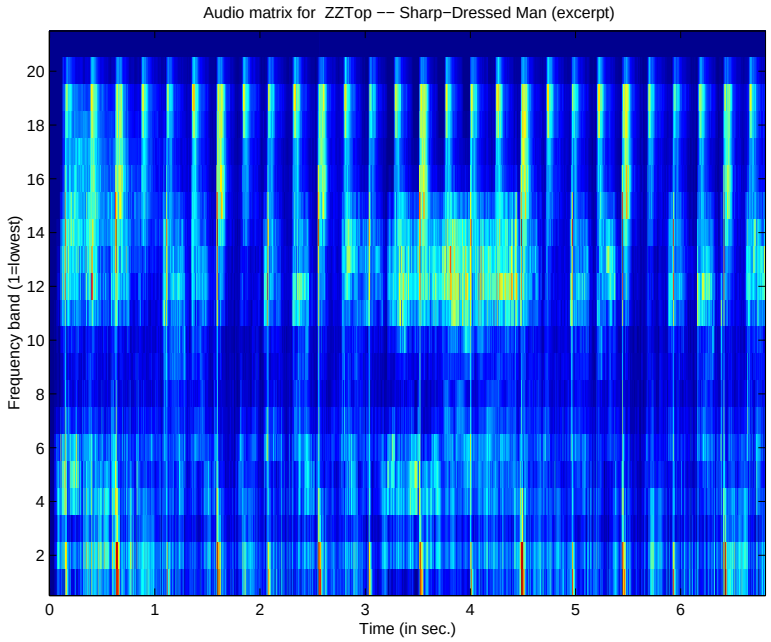

Figure 1: *Image of the audio matrix for ZZ Top's "Sharp Dressed Man"*

4.3. Nonuniqueness. Before going on, it is necessary to consider the nonuniqueness of this projection. We have seen above that in some cases (precisely, when the period of one subspace divides the period of the other), the order of projection does not matter. This is not true in general. While the DFT deals with orthogonal subspaces, the periodic subspaces $\mathcal{P}_p$ are not orthogonal to each other. Therefore, the representation of an arbitrary signal s as a linear combination of the basis elements is not unique. Furthermore, there is not a unique order to choose projection onto periodic subspaces, since different orders may yield different results.

4.4. Algorithms. At the heart of the PT is its ability to choose among these subspaces and determine the most relevant order in which to project. Sethares and Staley have proposed the Small-to-Large algorithm in [9]; just as its name suggests, the this algorithm scans a signal for relevant periodicities beginning at $p = 2$ and continuing up to larger ones. If the percent of the total energy removed by projection onto $\mathcal{P}_{p_i}$ is greater than a given threshold, the projection is carried out. Otherwise, that periodic space $\mathcal{P}_{p_i}$ is skipped and projection onto $\mathcal{P}_{p_{i+1}}$ is attempted. Observe that a "Large-to-Small" algorithm would be useless. Using the results of Corollary 1, if we first project a signal onto a subspace $\mathcal{P}_{np}$, the residual will not contain any of the smaller periodicities which are its divisors, p. This would yield misleading data. Sethares and Staley propose three additional algorithms; we used the Small-to-Large algorithm in our calculations primarily because it was the one that required the least amount of time to run.

5. Analysis of ZZ Top's "Sharp Dressed Man."

ZZ Top's "Sharp Dressed Man" (Audio Example 1) has a constant heavy rhythm throughout the song. Thus, we felt that this would be a good choice for analysis. The prominent beat of the song is introduced immediately when the song begins. Figure 1 is an image of the audio matrix that was created for the first 7 seconds of the song. The vertical axis corresponds to the 21 pitch bands of the audio matrix; each pitch band spans roughly half an octave, with the first band representing the lowest pitches. The horizontal axis represents time. The image is color-coded in rainbow order depending on the energy in a particular pitch band; red indicates high energy while blue indicates low energy.

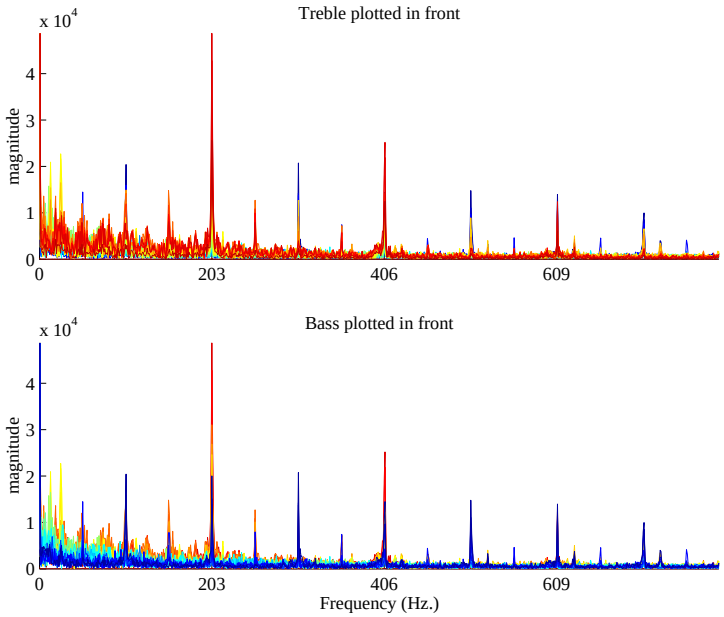

Figure 2: *DFT of data reduced "Sharp Dressed Man"*

In order to verify that we have not lost important rhythmic information through data reduction, we wrote an algorithm to output the audio matrix into a sound file using filtered white noise to fill in the rhythm bands (Audio Example 2). The drum beats were well represented; in addition, the voice was somewhat preserved. Other elements of the original song, such as the guitar, were almost completely lost. By creating the audio matrix, we managed to identify the rhythm with a much smaller amount of data, while still preserving some of the pitch information. The advantage of the 21-band audio matrix can be heard in Audio Example 3, which is the result of collapsing all the audio information into one band (rather than 21) and thus losing all information about pitch. Although the primary beat is quite audible, one cannot hear the relationship between the high and low bands that are a prominent feature of the rhythm.

5.1. DFT Analysis. The DFT reveals the frequency of the periodic bursts of energy in each pitch band. Figure 2 is a plot of the magnitudes of the DFTs of each row of the audio matrix superimposed. The colors correspond to the pitch bands of the audio matrix, with red representing the highest band. The height of a spike in the graph shows the relative prominence of each beat frequency within that band; we see a red spike at 220 on the x-axis, caused by a steady beat occurring 220 times (roughly 4 times a second) in the highest band. This enforces what we saw in Figure 1: a periodic high-energy burst in the upper pitches. This prominent spike in the DFT graph is the hi-hat cymbals. This is not the basic beat of the song; when we listen to the song, we tap our foot along with the bass drum. The blue spike at around 101 beats per minute is the best candidate for the primary beat. The frequency of this spike is half that of the prominent spike; that is, the high beat occurs twice for every low beat. We also see a spike in the middle bands at one-fourth the frequency of the bass drum, giving us a good candidate for the measure.

5.2. PT Analysis. Since we wish to detect integer periods, we first resample the song so that one beat corresponds to 12 samples (we chose 12 as a highly divisible number). Figure 3 shows the magnitudes of the residuals of projections of each row of the audio matrix onto the periodic subspaces $\mathcal{P}_p$; dark color indicates small residuals—in other words, subspaces that are close to x. We see that the 12-sample beat predominates in the high pitch bands, while the middle and low bands show either a 24-sample beat or a 96 ($= 4 \times 24$) sample measure. This relationship occurs because the song is in duple meter: all the divisions of a measure are by powers of two. Our implementation of the Small-to-Large a algorithm confirms this also. Figure 4 shows the magnitudes of the vectors resulting from the Small-to-Large decomposition of the signal. Again,

we see the numbers 12, 24, and 96 appearing as prominent periods.

6. Conclusion

We have discussed a few methods for quickly and efficiently detecting rhythm. Not only do our algorithms detect the primary beat, but they also give clues about the meter, which is revealed in the hierarchy of repeated accents and in the relationship between rhythms in the bass and the treble. In popular music, we are able to detect the meter of a particular work. However, to extend these methods to music without a metronomic beat would require additional processing, such as a beat tracking algorithm.

References

[1] William E. Boyce and Richard C. DiPrima. *Elementary Differential Equations and Boundary Value Problems*. Wiley, 2001.

[2] Joseph E. Flannick. *Rhythm Detection in Recorded Music*. Departmental honors thesis, under the direction of Rachel W. Hall and Adlai Waksman. Published at `http://www.sju.edu/~rhall/Rhythms/joe.pdf`. Saint Joseph's University, 2003.

[3] Rachel W. Hall and Krešimir Josić. "The Mathematics of Musical Instruments." *The American Mathematical Monthly*, vol. 108, April 2001.

[4] Simon Haykin and Barry Van Veen. *Signals and Systems*. Wiley, 1999.

[5] Robert Kelly. *Mathematics of Musical Rhythm*. Departmental honors thesis, under the direction of Rachel W. Hall. Published at `http://www.sju.edu/~rhall/Rhythms/bobby.pdf`. Saint Joseph's University, 2002.

[6] David W. Kammler. *A First Course in Fourier Analysis*. Prentice Hall, 2000.

[7] D. Rosenthal. "Emulation of Rhythm Perception." *Computer Music Journal*, vol. 16, no. 1, Spring 1992.

[8] Eric D. Scheirer. "Tempo and Beat Analysis of Acoustic Musical Signals." *Journal of the Acoustical Society of America*, vol. 103, no. 1, January 1998.

[9] William A. Sethares and Thomas W. Staley. "Periodicity Transforms." *Transactions on Signal Processing*, vol. 47, no. 11, November 1999.

[10] William A. Sethares and Thomas W. Staley. "Meter and Periodicity in Musical Performance." preprint, 2001.

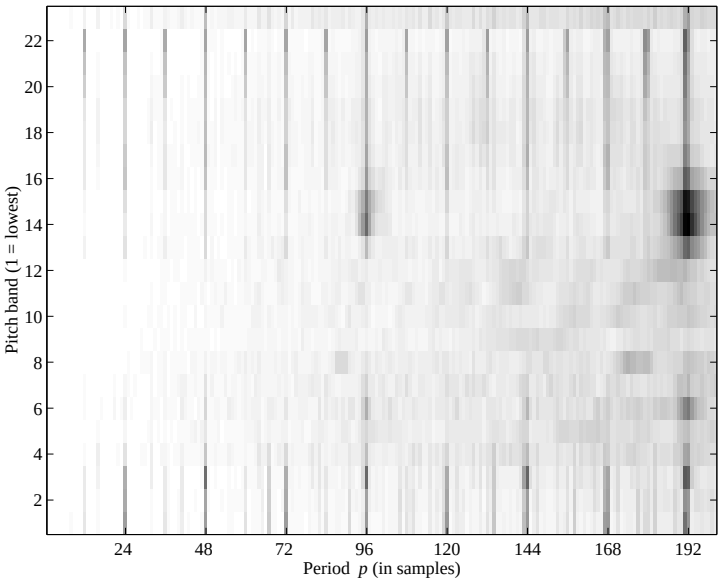

Figure 3: *Magnitudes of residuals of the audio matrix projected onto periodic subspaces*

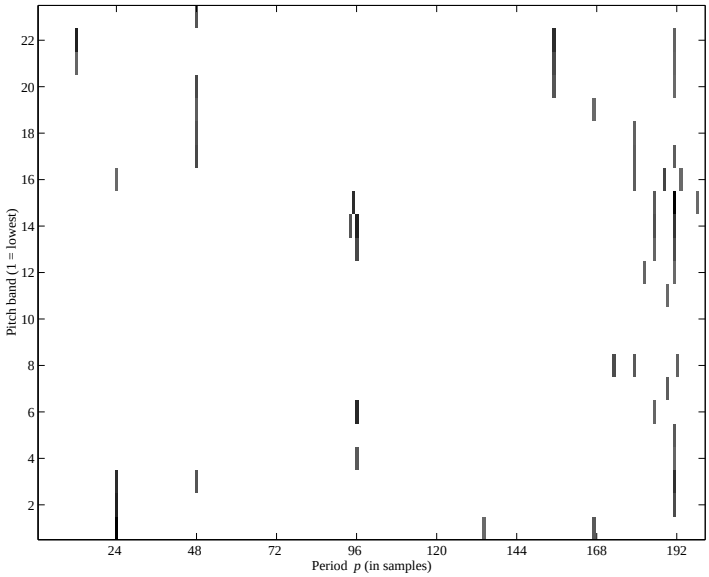

Figure 4: *Magnitudes of vectors in Small-to-Large decomposition of the audio matrix*