Time-frequency Analysis of Musical Rhythm

Xiaowen Cheng
Department of Mathematics; University of Minnesota–Twin Cities; Minneapolis, MN 55455
E-mail: chen1098@umn.edu, Tel. 612-625-2004, Fax. 612-626-2017

Jarod V. Hart
Department of Mathematics; University of Wisconsin–La Crosse; La Crosse, WI 54601
E-mail: hart.jaro@students.uwlax.edu, Tel. 608-785-6603, Fax. 608-785-6602

James S. Walker (Corresponding author)
Department of Mathematics; University of Wisconsin–Eau Claire; Eau Claire, WI 54702–4004
E-mail: walkerjs@uwec.edu, Tel. 715-836-4835, Fax. 715-836-2924

(Received 00 Month 200x; In final form 00 Month 200x)
We discuss a time-frequency analysis of musical rhythm and its relationship to melodic structure. The mathematical tools we employ are Gabor transforms (spectrograms) and the new technique of percussion scalagrams. We determine satisfactory parameters for computing percussion scalagrams as descriptors of musical tempo. We use our tools to display the hierarchical structure of both rhythm and melody within a variety of percussion and melodic performances. Our melodic-rhythmic analysis reveals objectively the multidimensional, multi-temporal nature of music.

**Keywords:** Music theory, time-frequency analysis, Gabor transforms, spectrograms, continuous wavelet transforms

**AMS 2000 Subject Classification Codes:** 42A99, 00A06, 00A69

**Introduction**

We shall use the mathematical techniques of Gabor transforms and continuous wavelet transforms to analyze the rhythmic structure of music and its interaction with melodic structure. This analysis reveals common features to these rhythmic and melodic structures. Most importantly, we shall find analogous hierarchical structures.

The paper is organized as follows. In Section 1 we summarize the mathematical method of Gabor transforms (spectrograms) and apply them to analyzing the melodic structure of music. Spectrograms provide a tool for visualizing the patterns of pitches within a musical passage; we illustrate this with an analysis of a Beatles melody. In Sections 2 and 3 we mathematically analyze the method of percussion scalograms, a new technique for determining rhythm and tempo introduced in [27]. We discuss its relationship to a continuous wavelet transform and deduce satisfactory parameters for displaying a percussion scalogram (in [27] this was done by trial and error). Our analysis of a drum solo shows how the percussion scalogram reveals the hierarchical structure within the rhythms of a percussion performance. In section 4 we introduce a new method, melodic-rhythmic analysis, that uses simultaneous displays of spectrograms and percussion scalograms to provide an objective representation of the multi-dimensional, multi-temporal nature of music. A brief concluding section provides some ideas for future research.

**Note:** musical passages, software for analyzing them, and video demonstrations of our examples, are available at the website:

http://www.uwec.edu/walkerjs/tfamr/

1 **Gabor transforms and Music**

We briefly review the widely employed method of Gabor transforms [12], also known as short-time Fourier transforms, or spectrograms, or sonograms. In [7, 8], Dörfler describes the fundamental aspects of using Gabor transforms for musical analysis. Two other sources for applications of short-time Fourier transforms are [17, 22]. There is also considerable mathematical background in [10, 11, 13], with musical applications in [9]. Using sonograms or spectrograms for analyzing the music of bird song is described in [14, 18, 21]. The theory of Gabor transforms is discussed in complete detail in [10, 11, 13], with focus on its discrete aspects in [1, 25]. However, to fix our notations for subsequent work, we briefly describe this theory.

The sound signals that we analyze are all digital, hence discrete, so we assume that a sound signal has the form \( \{f(t_k)\} \), for uniformly spaced values \( t_k = k\Delta t \) in a finite interval \([0, T]\). A Gabor transform of \( f \), with window function \( w \), is defined as follows. First, multiply \( \{f(t_k)\} \) by a sequence of shifted window functions \( \{w(t_k - \tau_{\ell})\}_{\ell=0}^{M} \), producing time localized subsignals, \( \{f(t_k)w(t_k - \tau_{\ell})\}_{\ell=0}^{M} \). Uniformly spaced time values, \( \{\tau_{\ell} = t_j\}_{\ell=0}^{M} \) are used for the shifts (\( j \) being a positive integer greater than 1). The windows \( \{w(t_k - \tau_{\ell})\}_{\ell=0}^{M} \) are all compactly supported and overlap each other. See Figure 1. The value of \( M \) is determined by the minimum number of windows needed
to cover $[0, T]$, as illustrated in Figure 1(b).

![Figure 1](image1.png)

Figure 1. (a) Signal. (b) Succession of shifted window functions. (c) Signal multiplied by middle window in (b); an FFT can now be applied to this windowed signal.

Second, because $w$ is compactly supported, we treat each subsignal $\{f(t_k)w(t_k - \tau_\ell)\}$ as a finite sequence and apply an FFT $\mathcal{F}$ to it. This yields the Gabor transform of $\{f(t_k)\}$:

$$\{\mathcal{F}\{f(t_k)w(t_k - \tau_\ell)\}\}_{\ell=0}^M.$$  \hspace{1cm} (1)

Note that because the values $t_k$ belong to a finite interval, we always extend our signal values beyond the interval’s endpoints by appending zeroes, hence the full supports of all windows are included.

The Gabor transform that we employ uses a Blackman window defined by

$$w(t) = \begin{cases} 
0.42 + 0.5 \cos(2\pi t/\lambda) + 0.08 \cos(4\pi t/\lambda) & \text{for } |t| \leq \lambda/2 \\
0 & \text{for } |t| > \lambda/2 
\end{cases}$$

for a positive parameter $\lambda$ equalling the width of the window where the FFT is performed. The Fourier transform of the Blackman window is very nearly positive (negative values less than $10^{-4}$ in size), thus providing an effective substitute for a Gaussian function (which is well-known to have minimum time-frequency support). See Figure 2. Further evidence of the advantages of Blackman-windowing is provided in [2, Table II]. In Figure 2(b) we illustrate that for each windowing by $w(t_k - \tau_m)$ we finely partition the frequency axis into thin rectangular bands lying above the support of the window. This provides a thin rectangular partition of the (slightly smeared) spectrum of $f$ over the support of $w(t_k - \tau_m)$ for each $m$. The efficacy of these Gabor transforms is shown by how well they produce time-frequency portraits that accord well with our auditory perception, which is described in the vast literature on Gabor transforms that we briefly summarized above. We shall provide some new examples that illustrate the effectiveness of these Gabor transforms. For all of our examples, we used 1024 point FFTs, based on windows of support $\lesssim 1/8 \text{ sec}$ with a shift of $\Delta \tau \approx 0.008 \text{ sec}$. These time-values are short enough to capture the essential features of musical frequency change.

1.1 Spectrograms of musical passages

The spectrogram of a waveform for a musical passage provides a time-frequency portrait of the passage. As a very basic example, consider the spectrogram in Figure 3 of a piano version of the melody from the Lennon-McCartney song, *All Across the Universe.*\(^1\) The short rectangular structures in the spectrogram correspond to the notes of the

\(^1\)More complex examples are illustrated in [27] and in [3].
Time-frequency Analysis of Musical Rhythm

Figure 2. Left: Blackman window, $\lambda = 1$. Right: Time-frequency representation—the units along the horizontal are in seconds, along the vertical are in Hz—of three Blackman windows multiplied by the real part of the kernel $e^{i2\pi nk/N}$ of the FFT used in a Gabor transform, for three different frequency values $n$. Each horizontal bar accounts for 99.99% of the energy of the cosine-modulated Blackman window (Gabor atom) graphed below it.

piano playing the passage. This is clearly evident by listening to the passage as the spectrogram is traced (using for example the software AUDACITY [26, 28]).

To musically analyze this, and other passages, we employ the methods of time-frequency analysis introduced in [27]. Specifically, we employ the following principle.

**Multiresolution Principle.** _Music is a patterning of sound in the time-frequency plane. Analyze music by looking for repetition of patterns of time-frequency structures over multiple time scales, and multiple resolution levels. The time-scales involved are those at the limits of human perception (roughly between 0.1 sec to 100 sec)._ This principle is an encapsulation in time-frequency terms of Jackendoff and Lerdahl’s hierarchical theory of music [15, 16]. Pinker has given a succinct summary of their theory [20, pp. 532–533]:

Jackendoff and Lerdahl show how melodies are formed by sequences of pitches that are organized in three different ways, all at the same time... The first representation is a grouping structure. The listener feels that groups of notes hang together in motifs, which in turn are grouped into lines or sections, which are grouped into stanzas, movements, and pieces. This hierarchical tree is similar to a phrase structure of a sentence, and when the music has lyrics the two partly line up... The second representation is a metrical structure, repeating sequence of strong and weak beats... summed up in musical notation as the time signature... The third representation is a reducational structure. It dissects the melody into essential parts and ornaments. The ornaments are stripped off and the essential parts further dissected into even more essential parts and ornaments on them... we sense it when we recognize variations of a piece in classical music or jazz. The skeleton of the melody is conserved while the ornaments differ from variation to variation.

To see how this theory applies to the _All Across the Universe_ passage, consider the time-frequency structures labelled $A_1$, $A_2$, $A_3$, $A_4$ in the spectrogram at the top of Figure 3. Here we can see repetition of a basic pattern $A_1$, in the bass scale, repeated within shorter time-scales (and shifted up an octave in frequency to the treble scale) as the patterns $A_2$, $A_3$, and $A_4$. Notice also that the structures $A_2$, $A_3$, $A_4$ have the following overall pitch pattern:

$$A_2 \quad A_3 \quad A_4$$

which repeats, as groups of notes, the up/down pitch pattern of the individual notes within these structures (with the higher notes for $A_3$ equalling the lower notes for $A_2$ and $A_4$). This passage has a hierarchical pitch structure, as characterized in the first representation described by Pinker.
Listening to this passage as the spectrogram is traced verifies that our analysis captures the essence of the passage’s melodic structure. We shall further analyze the rhythmic structure of this passage in Section 4. The basis of our rhythmic analysis is the method of percussion scalograms.

2 Percussion scalograms

In this section we briefly review the method of scalograms (continuous wavelet transforms) and then discuss the method of percussion scalograms.

2.1 Scalograms

The theory of continuous wavelet transforms is well-established [4, 5, 19]. A CWT differs from a spectrogram in that it does not use translations of a window of fixed width, instead it uses translations of differently sized dilations of a window. These dilations induce a logarithmic division of the frequency axis. The discrete calculation of a CWT that we use is described in [1, Section 4]. We shall only briefly review the definition of the CWT in order to fix our notation. We then use it to analyze percussion.

Given a function $\Psi$, called the wavelet, the continuous wavelet transform $\mathcal{W}_\Psi[f]$ of a sound signal $f$ is defined as

$$\mathcal{W}_\Psi[f](\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t) \Psi\left(\frac{t - \tau}{s}\right) dt$$  (2)

for scale $s > 0$ and time-translation $\tau$. For the function $\Psi$ in the integrand of (2), the variable $s$ produces a dilation and the variable $\tau$ produces a translation.
We omit various technicalities concerning the types of functions $\Psi$ that are suitable as wavelets; see [4, 5, 19]. In [4, 6], Equation (2) is derived from a simple model for the response of our ear’s basilar membrane—which responds to frequencies on a logarithmic scale—to a sound stimulus $f$.

We now discretize Equation (2). First, we assume that the sound signal $f(t)$ is non-zero only over the time interval $[0, T]$. Hence (2) becomes

$$W_\Psi[f](\tau, s) = \frac{1}{\sqrt{s}} \int_0^T f(t) \Psi\left(\frac{t - \tau}{s}\right) dt.$$  

We then make a Riemann sum approximation to this last integral using $t_m = m \Delta t$, with uniform spacing $\Delta t = T/N$; and discretize the time variable $\tau$, using $\tau_k = k \Delta t$. This yields

$$W_\Psi[f](\tau_k, s) \approx \frac{T}{N \sqrt{s}} \sum_{m=0}^{N-1} f(t_m) \Psi([t_m - \tau_k]s^{-1}).$$  

The sum in (3) is a correlation of two discrete sequences. Given two $N$-point discrete sequences $\{f_k\}$ and $\{\Psi_k\}$, their correlation $\{(f : \Psi)_k\}$ is defined by

$$(f : \Psi)_k = \sum_{m=0}^{N-1} f_m \Psi_{m-k}.$$  

[Note: For the sum in (4) to make sense, the sequence $\{\Psi_k\}$ is periodically extended, via $\Psi_{-k} := \Psi_{N-k}$.]

Thus, Equations (3) and (4) show that the CWT, at each scale $s$, is approximated by a multiple of a discrete correlation of $\{f_k = f(t_k)\}$ and $\{\Psi_{s\cdot k} = s^{-1/2} \Psi(t_k s^{-1})\}$. These discrete correlations are computed over a range of discrete values of $s$, typically

$$s = 2^{-r/J}, \quad r = 0, 1, 2, \ldots, I \cdot J$$  

where the positive integer $I$ is called the number of octaves and the positive integer $J$ is called the number of voices per octave. For example, the choice of 6 octaves and 12 voices corresponds—based on the relationship between scales and frequencies described below—to the well-tempered scale used for pianos.

The CWTs that we use are based on Gabor wavelets. A Gabor wavelet, with width parameter $\omega$ and frequency parameter $\nu$, is defined as follows:

$$\Psi(t) = \omega^{-1/2} e^{-\pi (t/\omega)^2} e^{i2\pi \nu t/\omega}.$$  

Notice that the complex exponential $e^{i2\pi \nu t/\omega}$ has frequency $\nu/\omega$. We call $\nu/\omega$ the base frequency. It corresponds to the largest scale $s = 1$. The bell-shaped factor $\omega^{-1/2} e^{-\pi (t/\omega)^2}$ in (6) damps down the oscillations of $\Psi$, so that their amplitude is significant only within a finite region centered at $t = 0$. See Figures 5 and 6. Since the scale parameter $s$ is used in a reciprocal fashion in Equation (2), it follows that the reciprocal scale $1/s$ will control the frequency of oscillations of the function $s^{-1/2} \Psi(t/s)$ used in Equation (2). Thus, frequency is described in terms of the parameter $1/s$, which Equation (5) shows is logarithmically scaled. This point is carefully discussed in [1] and [25, Chap. 6], where Gabor scalograms are shown to provide a method of zooming in on selected regions of a spectrogram.
2.2 Pulse trains and percussion scalograms

The method of using Gabor scalograms for analyzing percussion sequences was introduced by Smith in [23], and described empirically in considerable detail in [24]. The method described by Smith involved pulse trains generated from the sound signal itself. Our method is based on the spectrogram of the signal, which reduces the number of samples and hence speeds up the computation, making it fast enough for real-time applications. (An alternative method based on an FFT of the whole signal, the phase vocoder, is described in [22].)

Our discussion will focus on a particular percussion sequence. This sequence is a passage from the song, Dance Around. Listening to this passage you will hear several groups of drum beats, along with some shifts in tempo. This passage illustrates the basic principles underlying our approach.

In Figure 4(a) we show the spectrogram of the Dance Around clip. This spectrogram is composed of a sequence of thick vertical segments, which we will call vertical swatches. Each vertical swatch corresponds to a percussive strike on a drum. These sharp strikes on drum heads excite a continuum of frequencies rather than a discrete tonal sequence of fundamentals and overtones. The rapid onset and decay of these strike sounds produces vertical swatches in the time-frequency plane.

![Figure 4](image)

Figure 4. Calculating a percussion scalogram for the Dance Around sound clip. (a) Spectrogram of sound waveform with its pulse train graphed below it. (b) Percussion scalogram and the sound clip waveform graphed above it. The dark region labeled by G corresponds to a collection of drum strikes that we hear as a group, and within that group are other subgroups over shorter time scales that are indicated by the splitting of group G into smaller dark blobs as one goes upwards in the percussion scalogram (those subgroups are also aurally perceptible).

Our percussion scalogram method has the following two parts:

I. Pulse train generation. We generate a “pulse train,” a sequence of subintervals of 1-values and 0-values [see the graph at the bottom of Figure 4(a)]. The rectangular-shaped pulses in this pulse train correspond to sharp
onset and decay of transient bursts in the percussion signal graphed just above the pulse train. The widths of these pulses are approximately equal to the widths of the vertical swatches shown in the spectrogram. Most importantly, the location and duration of the intervals of 1-values corresponds to our hearing of the drum strikes, while the location and duration of the intervals of 0-values corresponds to the silences between the strikes. In Steps 1 and 2 of the method below we describe how this pulse train is generated.

II. Gabor CWT. We use a Gabor CWT to analyze the pulse train. This CWT calculation is performed in Step 3 of the method. The rationale for performing a CWT is that the pulse train is a step function analog of a sinusoidal of varying frequency. Because of this analogy between tempo of the pulses and frequency in sinusoidal curves, we employ a Gabor CWT for analysis. As an example, see the scalogram plotted in Figure 4(b). The thick vertical line segments at the top half of the scalogram correspond to the drum strikes, and these segments flow downward and connect together. Within the middle of the time-interval for the scalogram, these drum strike groups join together over three levels of hierarchy (see Figure 8). Listening to this passage, you can perceive each level of this hierarchy. This is a perfect example of Jackendoff and Lerdahl’s theory applied to rhythm.

Now that we have outlined the basis for the percussion scalogram method, we can list it in detail. The percussion scalogram method for analyzing percussive rhythm consists of the following three steps.

Percussion Scalogram Method

Step 1. Compute a signal which, at each time-coordinate for the Gabor transform, consists of averages of Gabor transform square-magnitudes lying within a frequency range consisting mostly of vertical swatches. For the time intervals corresponding to vertical swatches, this step produces higher square-magnitude values that lie above the mean of all square-magnitudes (because the mean is pulled down by the intervals of silence). For the Dance Around passage, the entire frequency range was used, as it consists entirely of vertical swatches corresponding to the percussive strikes. (When analyzing a percussive passage from a multi-instrument musical performance, we may have to isolate a particular frequency range that contains just the vertical swatches of the drum strikes. We illustrate this in Section 4.)

Step 2. Compute a signal that is 1 whenever the signal from Step 1 is larger than its mean and 0 otherwise. As the discussion in Step 1 shows, this produces a pulse train whose intervals of 1-values mark off the position and duration of the vertical swatches (hence of the drum strikes). Figure 4(a) illustrates this clearly.

Step 3. Compute a Gabor CWT of the pulse train signal from Step 2. This Gabor CWT provides an objective picture of the varying rhythms within a percussion performance.

We have already discussed the percussion scalogram in Figure 4(b). We shall continue this discussion, and provide several more examples of our method in Section 4. In each case, we find that a percussion scalogram allows us to finely analyze the rhythmic structure of percussion sequences, and some melodic sequences as well, using the selection of parameters that we now describe.

3 Choosing the Parameters for Percussion Scalograms

In this section, we describe how to choose the parameters for a percussion scalogram. Our main result is the following:

A satisfactory choice (satisfactory in the sense that both the average number of beats/sec are displayed and the individual beats are resolved) for the parameters of a percussion scalogram are (given the constraints of using positive integers for
the octaves I and voices M and using 256 total correlations):

\[ \omega = \frac{pT}{B}, \quad \nu = \frac{B}{pT}, \quad p = 4\sqrt{\pi} \]

\[ I = \left\lfloor \log_2 \left( \frac{p^2T^2}{2\delta B^2} \right) - \frac{3}{2} \right\rfloor, \quad M = \left\lfloor \frac{256}{I} \right\rfloor. \]  

(7)

The remainder of this section provides the rationale for this result. Those readers who are primarily interested in musical applications are encouraged to defer reading our discussion and turn instead to Section 4 where we discuss a variety of musical examples. These examples provide strong empirical support for our parameter choices in (7), in addition to the theoretical reasons that we now discuss.

3.1 Choosing the width and frequency parameters

We start with the pulse train described in the previous section; which we denote by \( \{P(t_m)\} \). This signal represents the percussion by \( P(t_m) = 1 \) during the strike of an instrument and \( P(t_m) = 0 \) at times when no instrument is being struck.

We use the Gabor wavelet in (6) to analyze the pulse signal. Since we are using the complex Gabor wavelet, we have both a real and imaginary part:

\[ \Psi_R(x) = \omega^{-\frac{1}{2}} e^{-\pi(x/\omega)^2} \cos \frac{2\pi\nu x}{\omega} \]  

(8)

\[ \Psi_I(x) = \omega^{-\frac{1}{2}} e^{-\pi(x/\omega)^2} \sin \frac{2\pi\nu x}{\omega}. \]  

(9)

These real and imaginary parts have the same envelope function:

\[ \Psi_E(x) = \omega^{-\frac{1}{2}} e^{-\pi(x/\omega)^2}. \]  

(10)

The width parameter \( \omega \) controls how quickly \( \Psi_E \) is dampened. For smaller values of \( \omega \) the function dampens more quickly. Also, \( \omega \) controls the magnitude of the wavelet function at \( x = 0 \). In fact, we have \( \Psi(0) = \omega^{-1/2} \).

The width parameter also affects the frequency of oscillations of the wavelet. As \( \omega \) is increased, the frequency of oscillations of the wavelet is decreased. See Figure 5.

![Figure 5](image-url)

Figure 5. Real parts of Gabor wavelet with frequency parameter \( \nu = 1 \) and width parameter \( \omega = 0.5, 1, \) and 2. For each graph, the horizontal range is \([-5, 5]\) and the vertical range is \([-2, 2]\).

The frequency parameter \( \nu \) is used to control the frequency of the wavelet within the envelope function. This
parameter has no effect on the envelope function, as shown in Equation (10). As $\nu$ is increased, the Gabor wavelet oscillates much more quickly. See Figure 6.

![Figure 6. Real parts of Gabor wavelet with width parameter $\omega = 1$ and frequency parameter $\nu = 0.5, 1, \text{and} 2$. For each graph, the horizontal range is $[-5, 5]$ and the vertical range is $[-2, 2]$.](image)

When using the Gabor wavelet to analyze music, correlations are computed using the Gabor wavelet with a scaling parameter $s$. For our pulse train $\mathcal{P}$ these correlations are denoted $(\mathcal{P} : \Psi_s)$ for $s = 2^{-r/M}$, $r = 0, 1, ..., IM$, and are defined by

$$
(\mathcal{P} : \Psi_s)(\tau_k) = \sum_{m=0}^{N-1} \mathcal{P}(t_m)s^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})}.
$$

(11)

Since $\{\mathcal{P}(t_m)\}$ is a binary signal, the terms of this sum will equal $s^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})}$ if $\mathcal{P}(t_m) = 1$, and 0 if $\mathcal{P}(t_m) = 0$. The values of $\tau_k$ represent the center of the Gabor wavelet being translated along the time axis. So for values of $t_m$ closer to $\tau_k$, $s^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})}$ will be larger in magnitude. Then, at values for $t_m$ where $\mathcal{P}(t_m) = 1$ and $t_m = \tau_k$, the corresponding term in the correlation sum will be

$$
\mathcal{P}(t_m)s^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})} = s^{-1/2}\overline{\Psi(0)} = \frac{1}{\sqrt{s}} \sqrt{\frac{2B}{pT}}
$$

which will represent the striking of an instrument. So as $s$ reaches its smallest values, near $s = 2^{-1}$, the correlations will have large magnitude values only near $\tau_k$, and where $\mathcal{P}(t_m) = 1$, i.e. at the beat of the instrument. This happens because small values of $s$ result in $2^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})}$ being dampened very quickly, so very little else other than the actual beats are detected by the Gabor CWT.

Detection of the rhythm and grouping of the percussion signal is accomplished by the larger values of $s$ that result in a slowly dampened Gabor wavelet. As the correlation sum moves to values such that $t_m \neq \tau_k$, the function $s^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})}$ is being dampened. But with the wavelet being dampened more slowly now, the values of $s^{-1/2}\overline{\Psi([t_m - \tau_k]s^{-1})}$ are larger near $t_m = \tau_k$ than they were before. Hence the $t_m$ values where $\mathcal{P}(t_m) = 1$ will result in summing more values of the wavelet that are significantly large. Therefore, any beat that is close to another beat will result in larger correlation values for larger values of $s$. Notice also that those values of $t_m$ where $\mathcal{P}(t_m) = 0$ that are close to $t_m$ values where $\mathcal{P}(t_m) = 1$ will result in summing across the lesser dampened Gabor wavelet values—our scalogram will thus be registering the grouping of closely spaced beats.

Now we need to choose the parameters $\omega$ and $\nu$ based on $\{\mathcal{P}(t_m)\}$ to obtain the desired shape for the Gabor wavelet. To choose these parameters for a specific percussion signal we will use an approximate measure of the average beats per second, and time between beats. The measure of average beats per second will be calculated by the total number of pulses in the signal, $B$, divided by the total time of the signal $T$. That is, the average beats per
second of the signal is \( B/T \). The approximate time between beats will then be the reciprocal of the average beats per second, \( T/B \). Then we let the width parameter \( \omega \) and the frequency parameter \( \nu \) be defined by

\[
\omega = \frac{pT}{2B}, \quad \nu = \frac{B}{pT}
\]

where \( p \) is a parameter used to control the resolution level of the Gabor wavelet. With these width and frequency parameters, the Gabor wavelet is

\[
\Psi(x) = \sqrt{\frac{2B}{pT}} e^{-\pi(2Bx/pT)^2} e^{i4\pi B^2/(p^2T^2)}.
\]

We want to detect beats that are within \( T/B \), the average time between beats, of each other. Likewise, we want separation of the beats that are not within \( T/B \) of each other. We accomplish this by inspecting the envelope function evaluated at \( x = T/B \),

\[
\Psi_E\left(\frac{T}{B}\right) = \sqrt{\frac{2B}{pT}} e^{-4\pi/p^2}.
\]

Now, the value of the enveloping function \( \Psi_E(T/B) \) can be written as a function of the parameter \( p \), call it \( M(p) \):

\[
M(p) = \sqrt{\frac{2B}{pT}} e^{-4\pi/p^2}.
\]

Remembering that \( T \) and \( B \) are constants determined by the percussion sound signal, the maximization of the magnitude of the wavelet at \( x = T/B \) becomes a simple one variable optimization problem. The first derivative of \( M(p) \) is

\[
M'(p) = \frac{16\pi - p^2}{2p^3 e^{-4\pi/p^2} \sqrt{pT/2B}}.
\]

Hence \( p = 4\sqrt{\pi} \) maximizes the value of the envelope function of the wavelet at \( x = T/B \), thus allowing us to detect beats within \( T/B \) of each other.

With the wavelet function dampened sufficiently slowly, we know that the envelope function is sufficiently large. But the correlations are computed by taking the magnitude of the sum of the complex Gabor wavelet samples. Since the real and imaginary parts involve products with sines and cosines, there are intervals where the functions are negative. It is these adjacent negative regions, on each side of the main lobe of \( \Psi_R \), that allow for the separation of beats that are greater than \( T/B \) apart but less than \( 2T/B \) apart (if they are more than \( 2T/B \) apart, the dampening of \( \Psi_E \) produces low-magnitude correlations).

### 3.2 Width and Frequency for better display

There is one wrinkle to the analysis above. If the width and frequency parameters are set according to Equation (12), then at the lowest reciprocal-scale value \( 1/s = 1 \) the display of the percussion scalogram cuts off at the bottom, and it is difficult to perceive the scalogram’s features at this scale. To remedy that defect, when we display a percussion
scalogram we double the width in order to push down the lowest reciprocal-scale by one octave. Hence we use the following formulas

$$\omega = \frac{pT}{B}, \quad \nu = \frac{B}{pT}, \quad p = 4\sqrt{\pi}$$

(16)

for displaying our percussion scalograms.

### 3.3 Choosing Octaves and Voices

The variable $1/s$ along the vertical axis of a percussion scalogram [see Figure 4(b) for example] is related to frequency, but on a logarithmic scale. To find the actual frequency at any point along the vertical axis we compute the base frequency $\nu/\omega$ multiplied by the value of $1/s$. The value of $I$ determines the range of the vertical axis in a scalogram, i.e. how large $1/s$ is, and the value of $M$ determines how many correlations per octave we are computing for our scalogram.

In order to have a satisfactory percussion scalogram, we need the maximum wavelet frequency equal to the maximum pulse frequency. The scale variable $s$ satisfies $s = 2^{-k/M}$, where $k = 0, 1, \ldots, IM$. Hence the maximum $1/s$ we can use is calculated as follows:

$$\frac{1}{s} = 2^{IM/M} = 2^I.$$  

Now let $\delta$ be the minimum distances between pulses on a pulse train. By analogy of our pulse trains with sinusoidal curves, we postulate that the maximum pulse frequency should be one-half of $1/\delta$. Setting this maximum pulse frequency equal to the maximum wavelet frequency, we have

$$\frac{1}{2\delta} = \frac{\nu}{\omega} 2^I.$$  

(17)

Notice that both sides of (17) have units of beats/sec.

Using the equations for $\nu$ and $\omega$ in (16), we rewrite Equation (17) as

$$\frac{1}{2\delta} = \frac{\nu}{\omega} 2^I = \left(\frac{B}{pT}\right)^2 2^I.$$  

Solving for $I$ yields

$$I = \log_2 \left(\frac{p^2T^2}{\delta B^2}\right) - 1.$$  

(18)

Because $I$ is required to be a positive integer, we shall round down this exact value for $I$. Thus we set

$$I = \left\lfloor \log_2 \left(\frac{p^2T^2}{\delta B^2}\right) - \frac{3}{2} \right\rfloor.$$  

(19)

To illustrate the value of selecting $I$ per Equation (19), in Figure 7 we show three different scalograms for the
Dance Around percussion sequence. For this example, Equation (19) yields the value $I = 4$. Using this value, we find that the scalogram plotted in Figure 7(b) is able to detect the individual drum strikes and their groupings. If, however, we set $I$ too low, say $I = 3$ in Figure 7(a), then the scalogram does not display the timings of the individual drum beats very well. On the other hand, if $I$ is set too high, say $I = 5$ in Figure 7(c), then the scalogram is too finely resolved. In particular, at the top of the scalogram, for $1/s = 2^5$, we find that the scalogram is detecting the beginning and ending of each drum strike as separate events, which overestimates by a factor of 2 the number of strikes.

Having set the value of $I$, the value for $M$ can then be expressed as a simple inverse proportion, depending on the program’s capacity. For example, with FAWAV the number of correlations used in a scalogram is constrained to be no more than 256, in which case we set

$$M = \left\lfloor \frac{256}{I} \right\rfloor$$

and that concludes our rationale for satisfactorily choosing the parameters for percussion scalograms.

4 Examples of Analyzing Rhythm

In this section we shall apply the technique of percussion scalograms to various percussion passages and to some examples of the interaction of rhythm and melody. Our discussion will reveal objectively the multi-dimensional, multi-temporal nature of music.
4.1 Percussion passages

The percussion scalogram technique provides a tool for revealing the rhythmic structure of musical passages. Typically we find that the rhythm divides up into a sequence of groups of notes, or beats or strikes, that are organized into a hierarchy. For example, in Figure 8 we show such a rhythm hierarchy obtained from a portion of the Dance Around percussion scalogram in Figure 4(b). This rhythm hierarchy corresponds to a grouping of the drum strikes within the time interval corresponding to a group that we have labeled $G$. At the first level of this hierarchy we see individual dark bars that correspond to the individual drum strikes. These strikes merge at Level 2 of the hierarchy into larger regions that mark off the main tempo for the drumming. Finally, the Level 2 regions merge at Level 3 to comprise the largest group $G$, the complete interval for the tempo at the previous level. Levels 2 and 3 thus mark off the time signature for this rhythmic drumming.

We can see in the percussion scalogram for the whole passage, shown in Figure 4(b), that the entire drum solo is composed of a sequence of similar rhythm hierarchies. Such sequences of rhythm hierarchies are generic to percussion performances. For example, we show in Figure 9 the hierarchical structure of rhythm for four different drum performances chosen from a variety of different musical styles. What is most apparent in these percussion scalograms is this common feature, that they consist of arrangements of rhythm hierarchies. This illustrates perfectly the hierarchical theory of music expounded by Jackendoff and Lerdahl, especially the first representation described by Pinker in the quote above. The third representation is also illustrated by these examples. For instance, the individual drum strikes at the first level of the hierarchies are often functioning as ornamentations on the essential structures at the second level (which are providing the fundamental tempo or time-signature).

We will not spend time here analyzing these passages. We have analyzed the Dance Around passage over the course of the paper. The Buenos Aires passage has already been analyzed in [27] and in [25, Section 6.5]. We shall analyze the Unsquare Dance passage later in this section. The Bangora passage is left as an exercise for the reader.
Dance Around. Rock drumming. (All frequencies)

Buenos Aires. Latin percussion. (Frequencies between 2000 and 3000 Hz)

Unsquare Dance (Clip 1). Jazz drumming. (Frequencies above 3000 Hz)

Bangora. Indian folk drumming. (Doubling of frequencies below 600 Hz)

Figure 9. Four percussion scalograms. In parentheses we list the frequency data used from the Gabor transforms.

4.1.1 Perception of loudness. When listening to these passages one perceives, especially for the *Buenos Aires* passage and the *Bangora* passage, that the darker groups of strikes in the percussion scalograms seem to correlate
with loudness of the striking. This seems counterintuitive, since the pulse train consists only of 0’s and 1’s, which would not seem to reflect varying loudness. This phenomenon can be explained as follows. When a pulse is very long, that requires a more energetic striking of the drum, and this more energetic playing translates into a louder sound. The longer pulses correspond to darker spots lower down on the scalogram, and we hear these as louder sounds. (The other way that darker spots appear lower down is in grouping of several strikes. We do not hear them necessarily as louder individual sounds, but taken together they account for more energy than single, narrow pulses individually.)

4.2 Melodic-rhythmic analysis

We now use both spectrograms and percussion scalograms to analyze melodic pieces. We first discuss the Beatles melody that we analyzed earlier. This discussion illustrates how combining both spectrograms and percussion scalograms allows us to perform a new type of visualization of the music, which we call *melodic-rhythmic analysis*. We then discuss a lyrical passage from the Buenos Aires song and a passage from the jazz performance, Unsquare Dance, to further illustrate our method.

![Figure 10. Melodic-Rhythmic Analysis of Across the Universe.](image)

On top are the spectrograms from Clips 1 and 2, below them are their percussion scalograms constructed from Gabor transform frequency data above 2500 Hz. The structures marked $A_1$, $A_2$, $A_3$ are similarly labeled in Figure 3.
4.2.1 Melodic-rhythmic analysis of a Beatles melody. We analyze two clips from the All Across the Universe passage. We used only frequencies above 2500 Hz in their spectrograms to generate the percussion scalograms. We did this because the fundamentals of piano notes often overlap each other in time, while much higher pitched overtones dissipate more quickly and hence are more clearly separated in time. In Figure 10 we show spectrograms and percussion scalograms for the two clips. Clip 1 isolates the structure $A_1$ from Figure 3 and Clip 2 isolates the structures $A_2$ and $A_3$.

The percussion scalogram for Clip 1 is consistent with the up-down pitch pattern of the notes shown for $A_1$ within its spectrogram. We see a simple two-level hierarchy consisting of groupings of a darker strike followed by a lighter strike. This is consistent with the high-low note pattern of $A_1$ with the higher note slightly longer than the lower note, thus shown as a darker strike on the percussion scalogram. The consonance of these two representations of the music, the melody within the spectrogram and the rhythm within the percussion scalogram, creates the simple, mundane, effect of the music (like the ticking of a clock).

Clip 2 shows the more uplifting pattern of $A_2$ and $A_3$, as repetitions of $A_1$ at a higher pitch and over shorter time scales (with an overall up-down pattern at a higher hierarchical level). With this clip, the individual strikes for the lower notes are longer than for the higher notes, reversing the pattern for Clip 1. The percussion scalogram captures the slight variations in tempo that correspond to the variations in pitch in the spectrogram; the correspondence of these slight alternations in tempo and pitch give the passage its charm.

We observe that, because of the overlapping in time of two pairs of piano notes, they are not separated within the two longest width stripes of the percussion scalogram of Clip 2. Those two stripes each represent groupings of two piano notes, which we clearly perceive on listening to the piece. Although this is a defect of the percussion scalogram representation, it is important to note that our method of melodic-rhythmic analysis requires both the percussion scalogram and the spectrogram to be displayed together, and we see these notes clearly separated within the spectrogram.

4.2.2 Melodic-rhythmic analysis of a lyrical passage. In the melodic-rhythmic analysis of the passage from Buenos Aires shown in Figure 11 we can see the formants of the singer’s voice in the spectrogram. The percussion scalogram shown below the spectrogram was obtained through first processing the Gabor transform by increasing its values by 25% for time values greater than 2.048. This equalized the volume level of the singer’s voice (she sang the beginning lyrics at a higher volume). Listening to the recording play as a cursor moves across the percussion scalogram, we perceive that the rhythm hierarchies of the percussion scalogram align closely (although not perfectly) with the hierarchies of the words of the singer’s lyrical phrasing—a nice example of what Pinker describes as “This hierarchical tree is similar to a phrase structure of a sentence, and when the music has lyrics the two partly line up.”

4.2.3 Melodic-rhythmic analysis of a jazz performance. In Figure 12 we show our melodic-rhythmic analysis of a clip of a jazz recording, Unsquare Dance. This clip has two instrumental parts, drumstick strikings and bass notes. We have separated these parts into two percussion scalograms using different frequency ranges of the Gabor transform. The drumstick strikings exhibit a very complex tempo that appears to reflect different time-signatures than the time-signature exhibited by the bass notes. We can even observe four levels of hierarchy within parts of this complex tempo (see the hierarchy labeled $F$ in Figure 12)—a testimony to the virtuosity of the drummer.

Following the passage recorded in Clip 1, there is a transition to a performance involving hand claps, piano notes, and bass notes. In Figure 13 we show our melodic-rhythmic analysis of this second passage. We used three different frequency ranges from the Gabor transform to isolate the different instruments from the performance. The clip begins with a transition from rapid drumstick strikings to hand clappings when the piano enters. The rhythm of
the hand clappings plus piano notes has a 7/4 time signature. Notice that the bass notes are playing with a simple repetition of 4 beats that helps the other musicians play within this unusual time signature. The scalogram that isolates these bass notes (the bass scalogram at the bottom of Figure 13) may not be clearly perceived to correspond to the bass playing in Clip 2, due to the faintness of the bass notes. To show that this percussion scalogram does capture the timing and duration of the bass notes, we processed the Gabor transform of Clip 2. We multiplied Gabor transform values for frequencies less than 400 Hz by a factor of 20 (and left higher frequency data unchanged), and then performed an inverse Gabor transform ([19, 27], [25, Chap. 5]) to create a new sound signal. This processed sound signal has a percussion scalogram (using all frequency data) that is identical to the one at the bottom of Figure 13. When it is played along with the percussion scalogram, one perceives the exact correspondence between the scalogram and the bass notes.

4.2.4 Syncopation. Our melodic-rhythmic analysis of Clip 2 from Unsquare Dance shows that the piano performance exhibits some syncopation. We can see this syncopation by comparing the piano notes in the spectrogram (in line with P at around 600 Hz), with the emphases they are receiving in the piano percussion scalogram. There is less emphasis, thinness in the scalogram, on the concluding piano notes that are in line with the concluding, emphasized, fourth notes played by the bass (which are thicker in the bass scalogram).
4.2.5 Remarks on nearly simultaneous notes. As with Clip 2 of the All Across the Universe Passage, we find for Clip 2 of Unsquare Dance that there are a few groups of nearly simultaneous piano notes unresolved within their percussion scalogram. But we see them clearly resolved within the spectrogram. We reiterate that our melodic-rhythmic analysis method requires analyzing both the spectrogram and percussion scalograms. It also is worth noting that resolving these notes within the percussion scalograms would preclude the time signature analysis that we made in Figure 13.

Conclusion

We have shown how spectrograms and percussion scalograms provide a multi-dimensional, multi-temporal, description of music. The new method of musical theory, melodic-rhythmic analysis, introduced here will be applied in subsequent papers to a wide variety of music. Other research would involve applying our methods to real-time critiques of musical performances and to rhythmic/melodic processing of recordings.
Figure 13. Melodic-Rhythmic Analysis of Unsquare Dance (Clip 2). Top: Spectrogram. P aligns with the piano notes. Second from Top: Percussion scalogram of frequencies above 3000 Hz. Drum stick and hand clap percussion are emphasized. Third from Top: Percussion scalogram of frequencies between 400 and 3000 Hz. Piano notes are emphasized. Bottom: Percussion scalogram of frequencies below 400 Hz. Bass notes are emphasized. Notice that the hand clapping interlaces with the piano notes—7 beats to a measure of 4 (marked off by the bass notes)—a time signature of 7/4.

Acknowledgement

This research was supported by the National Science Foundation Grant 144-2974/92974 (REU site SUREPAM program).
References


http://www.uwec.edu/walkerjs/media/38228[1].pdf


http://www.uwec.edu/walkerjs/PicturesOfMusic/


http://www.uwec.edu/walkerjs/primer/Ch5extract.pdf

and

http://www.uwec.edu/walkerjs/primer/Ch6extract.pdf


http://www.uwec.edu/walkerjs/tfamr/DocAudFW.pdf


http://www.uwec.edu/walkerjs/media/TFAM.pdf

[28] Website for paper: http://www.uwec.edu/walkerjs/tfamr/