# Time-frequency Analysis of Musical Rhythm

*Xiaowen Cheng, Jarod V. Hart, and James S. Walker*

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 the hierarchical structure of rhythm. Hierarchical structure is common to rhythmic performances throughout the world's music. The work described here is interdisciplinary and experimental. We use mathematics to aid in the understanding of the structure of music, and have developed mathematical tools that (while not completely finished) have shown themselves to be useful for this musical analysis. We aim to explore ideas with this paper, to provoke thought, not to present completely finished work.

The paper is organized as follows. We first summarize the mathematical method of Gabor transforms (also known as short-time Fourier transforms, or spectrograms). Spectrograms provide a tool for visualizing the patterns of time-frequency structures within a musical passage. We then review the method of percussion scalograms, a new technique for analyzing rhythm introduced in [34]. After that, we show how percussion scalograms are used to analyze percussion passages and rhythm. We carry out four analyses of percussion passages from a variety of music styles (rock drumming, African drumming, and jazz drumming). We also explore three examples of the connection between rhythm and melody (a jazz piano piece, a Bach piano transcription, and a jazz orchestration). These examples provide empirical justification of our method. Finally, we explain how the parameters for percussion scalograms are chosen in order to provide a satisfactory display of the pulse trains that characterize a percussion passage (a key component of our method). A brief concluding section provides some ideas for future research.

*Xiaowen Cheng is a student of mathematics at the University of Minnesota–Twin Cities. Her email address is* chen1098@umn.edu.

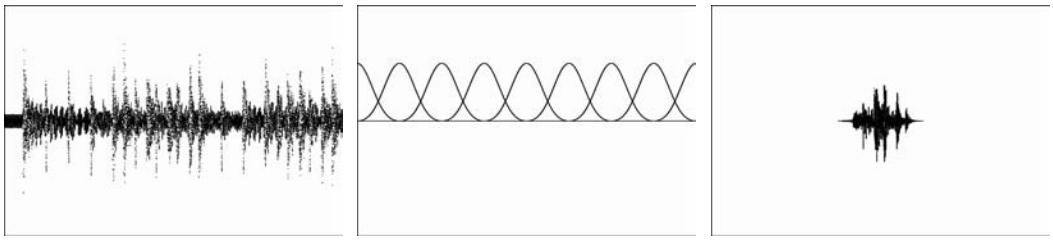
*Jarod V. Hart is a student of mathematics at the University of Kansas, Lawrence. His email address is* jhart@math.ku.edu.

*James S. Walker is professor of mathematics at the University of Wisconsin–Eau Claire. His email address is* walkerjs@uwec.edu.

## Gabor Transforms and Music

We briefly review the widely employed method of Gabor transforms [17], also known as short-time Fourier transforms, or spectrograms, or sonograms. The first comprehensive effort in employing spectrograms in musical analysis was Robert Cogan's masterpiece, *New Images of Musical Sound* [9]—a book that still deserves close study. In [12, 13], Dörfler describes the fundamental mathematical aspects of using Gabor transforms for musical analysis. Two other sources for applications of short-time Fourier transforms are [31, 25]. There is also considerable mathematical background in [15, 16, 19], with musical applications in [14]. Using sonograms or spectrograms for analyzing the music of bird song is described in [21, 30, 26]. The theory of Gabor transforms is discussed in complete detail in [15, 16, 19], with focus on its discrete aspects in [1, 34]. 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



(a) (b) (c)

**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.**

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, $\{\tau_\ell = t_{j\ell}\}_{\ell=0}^{M}$ 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).

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. (A good, brief explanation of how FFTs are used for frequency analysis can be found in [1].) This yields the Gabor transform of $\{f(t_k)\}$:

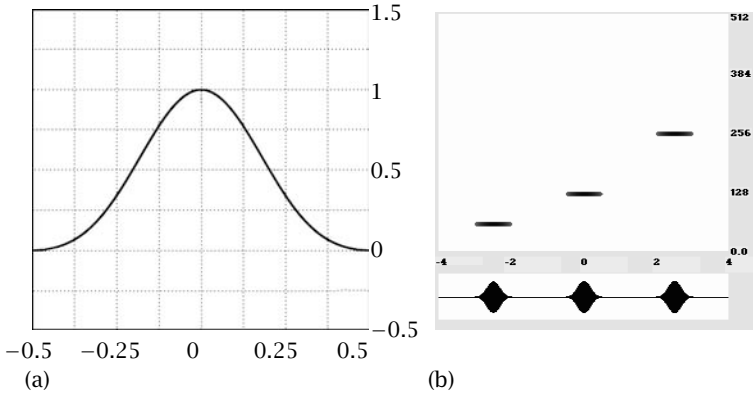
$$\{\mathcal{F}\{f(t_k)w(t_k - \tau_\ell)\}\}_{\ell=0}^{M}. \tag{1}$$

Note that because the values $t_k$ belong to the finite interval $[0, T]$, 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

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

for a positive parameter $\lambda$ equaling 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 [3, 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.



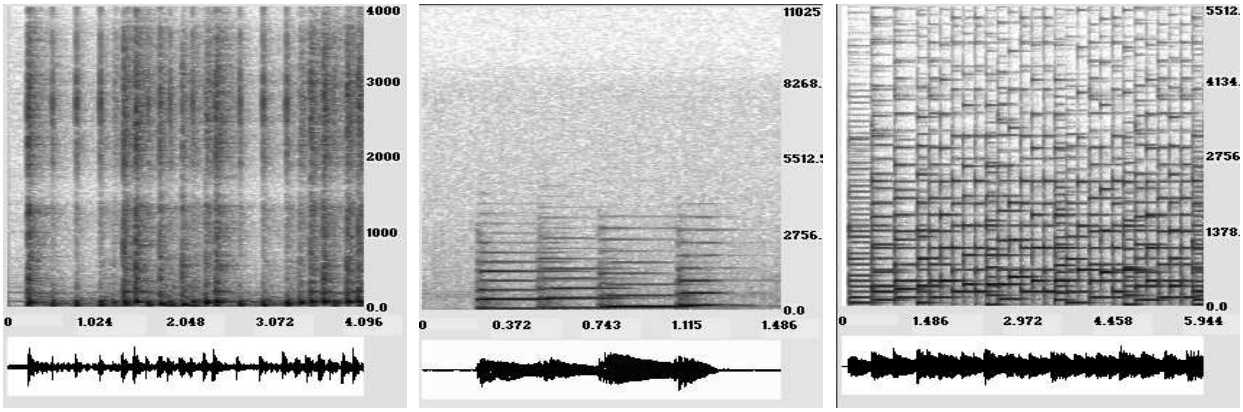

(a) (b)

**Figure 2. (a) Blackman window, $\lambda = 1$. Notice that it closely resembles the classic Gabor window—a bell curve described by a Gaussian exponential—but it has the advantage of compact support. (b) 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.**

It is interesting to listen to the sound created by the three Gabor atoms in Figure 2(b). You can watch a video of the spectrogram being traced out while the sound is played by going to the following webpage:

(2) `http://www.uwec.edu/walkerjs/TFAMRVideos/`

and selecting the video for *Gabor Atoms*. The sound of the atoms is of three successive pure




(a) Drum Clip (b) Piano scale notes (c) Bach melody

**Figure 3. Three spectrograms. (a) Spectrogram of a drum solo from a rock song. (b) Notes along a piano scale. (c) Spectrogram of a piano solo from a Bach melody.**

tones, on an ascending scale. The sound occurs precisely when the cursor crosses the thin dark bands in the spectrogram, and our aural perception of a constant pitch matches perfectly with the constant darkness of the thin bands. These Gabor atoms are, in fact, good examples of *individual notes*. Much better examples of notes, in fact, than the infinitely extending (both in past and future) sines and cosines used in classical Fourier analysis. Because they are good examples of pure tone notes, these Gabor atoms are excellent building blocks for music.

We shall provide some new examples that further illustrate the effectiveness of these Gabor transforms. For all of our examples, we used 1024 point FFTs, based on windows of support $\lesssim 1/8$ sec with a shift of $\Delta\tau \approx 0.008$ sec. These time-values are usually short enough to capture the essential features of musical frequency change.

In Figure 3 we show three basic examples of spectrograms of music. Part (a) of the figure shows a spectrogram of a clip from a rock drum solo. Notice that the spectrogram consists of dark vertical swatches; these swatches correspond to the striking of the drum, which can be verified by watching a video of the spectrogram (go to the website in (2) and select the video *Rock Drum Solo*). As the cursor traces over the spectrogram in the video, you will hear the sound of the drum strikes during the times when the cursor is crossing a vertical swatch. The reason why the spectrogram consists of these vertical swatches will be explained in the next section.

Part (b) of Figure 3 shows a spectrogram of a recording of four notes played on a piano scale. Here the spectrogram shows two features. Its main feature is a set of four sections consisting of groups of horizontal line segments placed vertically above each other. These vertical series of short horizontal segments are the fundamentals and overtones of the piano notes. There are also thin vertical swatches located at the beginning of each note. They are the percussive attacks of the notes (the piano is, in fact, classed as a percussive instrument).

Part (c) of Figure 3 shows a spectrogram of a clip from a piano version of a famous Bach melody. This spectrogram is much more complex, rhythmically and melodically, than the first two passages. Its melodic complexity consists in its *polyphonic* nature: the vertical series of horizontal segments are due to three-note *chords* being played on the treble scale and also individual notes played as counterpoint on the bass scale.[1] (This contrasts with the single notes in the *monophonic* passage in (b).) We will analyze the rhythm of this Bach melody in Example 5 below.

## Scalograms, Percussion Scalograms, and Rhythm

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

### Scalograms

The theory of continuous wavelet transforms is well-established [10, 8, 27]. 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

[1] *The chord structure and counterpoint can be determined either by careful listening or by examining the score* [2].

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

$$\mathcal{W}_\Psi[f](\tau,s) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} f(t)\overline{\Psi\left(\frac{t-\tau}{s}\right)}\,dt \tag{3}$$

for *scale* $s > 0$ and *time-translation* $\tau$. For the function $\Psi$ in the integrand of (3), 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 [8, 10, 27]. In [8, 11], Equation (3) is derived from a simple analogy with the logarithmically structured response of our ear's basilar membrane to a sound stimulus $f$.

We now discretize Equation (3). First, we assume that the sound signal $f(t)$ is non-zero only over the time interval $[0, T]$. Hence (3) becomes

$$\mathcal{W}_\Psi[f](\tau,s) = \frac{1}{\sqrt{s}}\int_{0}^{T} f(t)\overline{\Psi\left(\frac{t-\tau}{s}\right)}\,dt.$$

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

$$\mathcal{W}_\Psi[f](\tau_k,s) \approx \frac{T}{N}\frac{1}{\sqrt{s}}\sum_{m=0}^{N-1} f(t_m)\overline{\Psi\left([t_m-\tau_k]s^{-1}\right)}. \tag{4}$$

The sum in (4) 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

$$(f:\Psi)_k = \sum_{m=0}^{N-1} f_m\overline{\Psi_{m-k}}. \tag{5}$$

(Note: For the sum in (5) to make sense, the sequence $\{\Psi_k\}$ is *periodically extended,* via $\Psi_{-k} := \Psi_{N-k}$.)

Thus, Equations (4) and (5) show that the CWT, at each scale $s$, is approximated by a multiple of a discrete correlation of $\{f_k = f(t_k)\}$ and $\{\Psi_k^s = s^{-1/2}\Psi(t_k s^{-1})\}$. These discrete correlations are computed over a range of discrete values of $s$, typically

$$s = 2^{-r/J}, \quad r = 0, 1, 2, \ldots, I\cdot J, \tag{6}$$

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 equal-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:

$$\Psi(t) = \omega^{-1/2}e^{-\pi(t/\omega)^2}e^{i2\pi\nu t/\omega}. \tag{7}$$

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 (7) damps down the oscillations of $\Psi$, so that their amplitude is significant only within a finite region centered at $t = 0$. See Figures 13 and 14. Since the scale parameter $s$ is used in a reciprocal fashion in Equation (3), 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 (3). Thus, frequency is described in terms of the parameter $1/s$, which Equation (6) shows is logarithmically scaled. This point is carefully discussed in [1] and [34, Chap. 6], where Gabor scalograms are shown to provide a method of zooming in on selected regions of a spectrogram.

### Pulse Trains and Percussion Scalograms

The method of using Gabor scalograms for analyzing percussion sequences was introduced by Smith in [32] and described empirically in considerable detail in [33]. 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 [31].)

Our discussion will focus on a particular percussion sequence. This sequence is a passage from the song, *Dance Around.* Go to the URL in (2) and select the video, *Dance Around percussion,* to hear this passage. 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. Because the rapid onset and decay of these sharp strikes produce approximate delta function pulses—and a delta function pulse has an FFT that consists of a constant value for all frequencies—it follows that these strike sounds produce vertical swatches in the time-frequency plane.

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

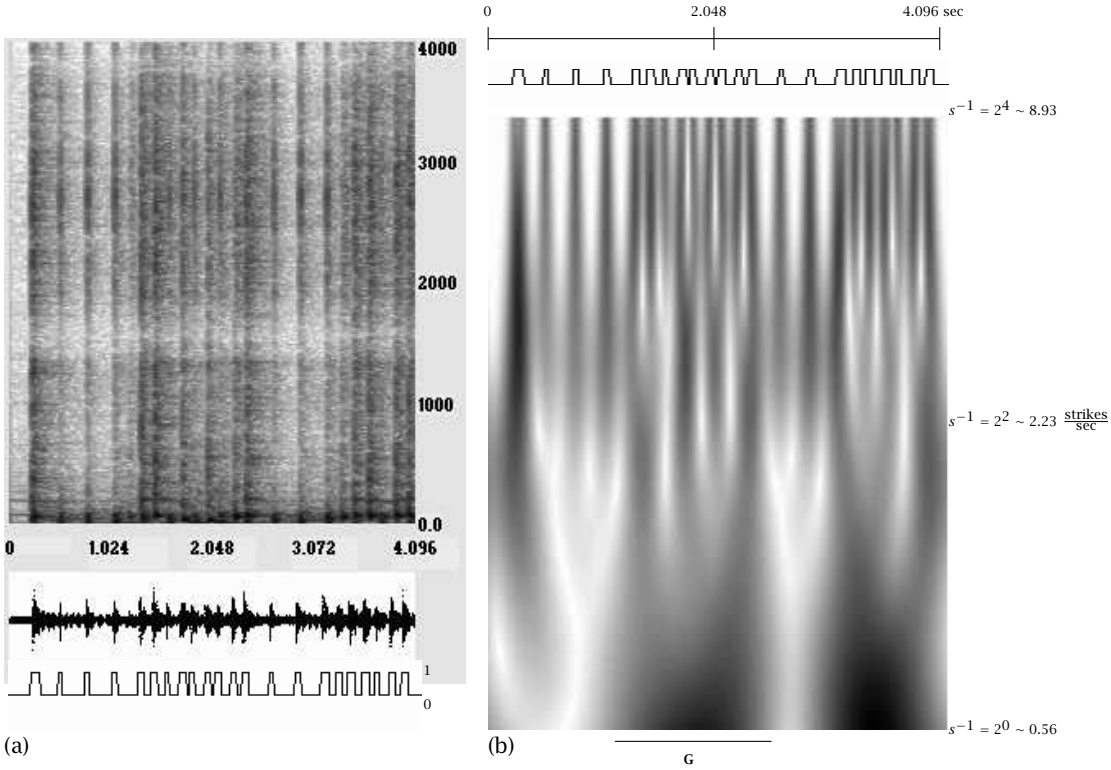



**Figure 4. Calculating a percussion scalogram for the *Dance Around* sound clip. (a) Spectrogram of sound waveform with its pulse train graphed below it. (b) Percussion scalogram and the pulse train graphed above it. The dark region labeled by G corresponds to a collection of drum strikes that we hear 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). See Figure 5 for a better view of G.**

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 Step 1 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 2 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 four levels of hierarchy (see Figure 5). Listening to this passage, you can perceive each level of this hierarchy.

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 two steps.

**Percussion Scalogram Method**

**Step 1.** Let $\{g(\tau_m, y_k)\}$ be the spectrogram image, like in Figure 4(a). Calculate the average $\bar{g}$ over all frequencies at each time-value $\tau_m$:

$$\bar{g}(\tau_m) = \frac{1}{P}\sum_{k=0}^{P-1} g(\tau_m, y_k), \tag{8}$$

(where $P$ is the total number of frequencies $y_k$),

and denote the average of $\bar{g}$ by $A$:

$$A = \frac{1}{M+1} \sum_{m=0}^{M} \bar{g}(\tau_m). \tag{9}$$

Then the pulse train $\{\mathcal{P}(\tau_m)\}$ is defined by

$$\mathcal{P}(\tau_m) = \mathbf{1}_{\{\tau_k : \bar{g}(\tau_k) > A\}}(\tau_m). \tag{10}$$

where $\mathbf{1}$ is the indicator function.[2] The values $\{\mathcal{P}(\tau_m)\}$ describe a pulse train whose intervals of 1-values mark off the position and duration of the vertical swatches (hence of the drum strikes). See Figure 6.

**Step 2.** *Compute a Gabor CWT of the pulse train signal* $\{\mathcal{P}(\tau_m)\}$ *from Step 1.* This Gabor CWT provides an objective picture of the varying rhythms within a percussion performance.

**Remarks**. **(a)** For the time intervals corresponding to vertical swatches, equations (8) and (9) produce values of $\bar{g}$ that lie above the average $A$ (because $A$ is pulled down by the intervals of silence). See Figure 6(a). For some signals, where the volume level is not relatively constant (louder passages interspersed with quieter passages) the total average $A$ will be too high (the quieter passages will not contribute to the pulse train). We should instead be computing local averages over several (but not all) time-values. We leave this as a goal for subsequent research. In a large number of cases, such as those discussed in this article, we have found that the method described above is adequate. **(b)** 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 other percussive passages, we may have to isolate a particular frequency range that contains just the vertical swatches of the drum strikes. We illustrate this later in the musical examples we describe (see the next section, "Examples of Rhythmic Analysis"). **(c)** We leave it as an exercise for the reader to show that the calculation of $\bar{g}(\tau_m)$ can actually be done in the time-domain using the data from the windowed signal values. (Hint: Use Parseval's theorem.) We chose to use the spectrogram values because of their ease of interpretation—especially when processing needs to be done, such as using only a particular frequency range. The spectrogram provides a lot of information to aid in the processing. **(d)** Some readers may wonder why we have computed a Gabor CWT in Step 2. Why not compute, say, a Haar CWT (which is based on a step function as wavelet)? We have found that a Haar CWT does provide essentially the same information as the *magnitudes* of the Gabor CWT (which is all we use in this article; using the phases of the complex-valued Gabor CWT is left for future research). However, the Haar CWT is more difficult to interpret, as shown in Figure 7.

[2] *The indicator function* $\mathbf{1}_S$ *for a set* $S$ *is defined by* $\mathbf{1}_S(t) = 1$ *when* $t \in S$ *and* $\mathbf{1}_S(t) = 0$ *when* $t \notin S$.

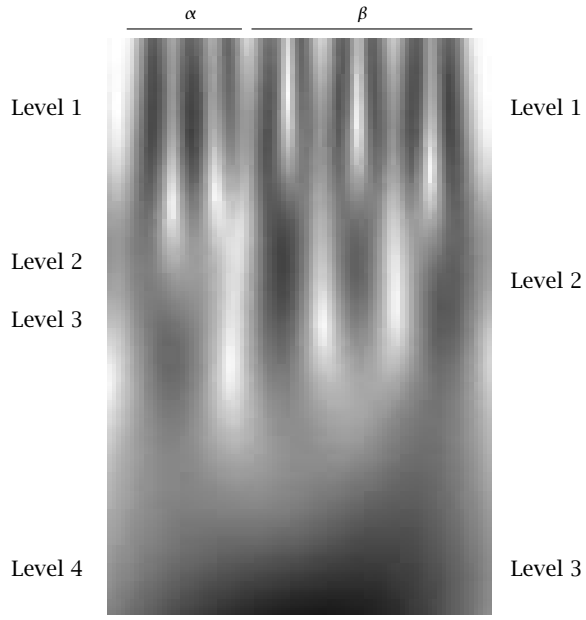



**Figure 5. A rhythm hierarchy, obtained from the region corresponding to G in Figure 4. The hierarchy has two parts, labeled $\alpha$ and $\beta$. In each part the top level, Level 1, comprises the individual strikes. These strikes merge at Level 2 into regions which correspond to double strikes and which are aurally perceptible as groupings of double strikes. Notice that the Level 2 regions for $\beta$ lie at positions of slightly increasing then decreasing strike-frequency as time proceeds; this is aurally perceptible when listening to the passage. There is also a Level 3 region for $\alpha$ that merges with the Level 2 regions for $\beta$ to comprise the largest group G.**

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 the next section. In each case, we find that a percussion scalogram allows us to finely analyze the rhythmic structure of percussion sequences.

## Examples of Rhythmic Analysis

As discussed in the previous section, a percussion scalogram allows us to perceive a hierarchal organization of the strikes in a percussion sequence. Hierarchical structures within music, especially within rhythmic passages and melodic contours, is a well-known phenomenon. For example, in an entertaining and thought-provoking book [24] with an excellent bibliography, *This Is Your Brain On Music,* Daniel Levitin says in regard to musical production (p. 154):

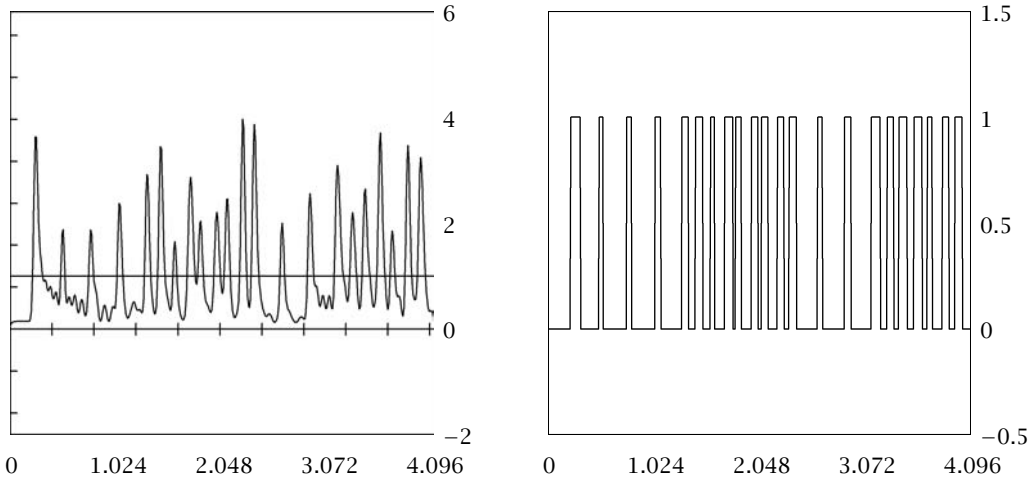


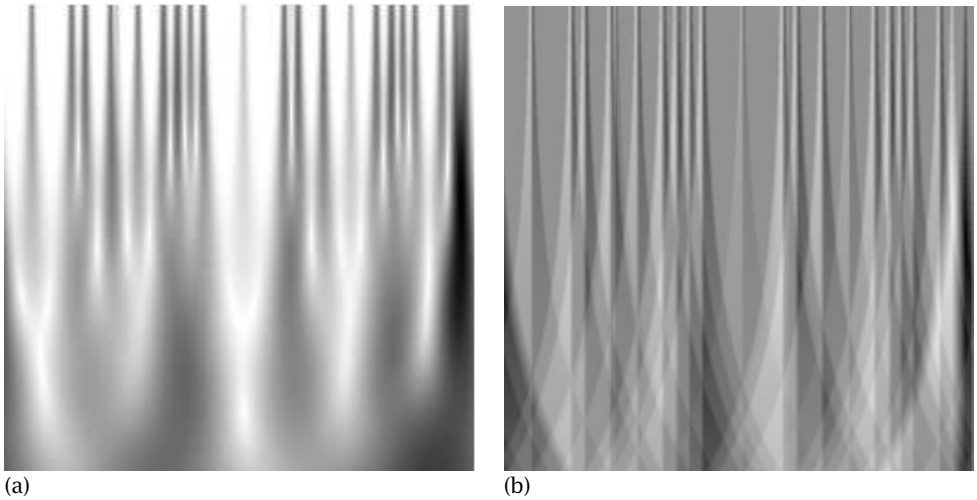

**Figure 6. Creation of a pulse train. On the left we show the graph of $\bar{g}$ from equation (8) for the spectrogram of the *Dance Around* sound clip [see Figure 4(a)], which we have normalized to have an average of $A = 1$. The horizontal line is the graph of the constant function 1. The pulse train, shown on the right, is then created by assigning the value 1 when the graph of $\bar{g}$ is larger than $A$, and 0 otherwise.**

> Our memory for music involves hierarchical encoding—not all words are equally salient, and not all parts of a musical piece hold equal status. We have certain entry points and exit points that correspond to specific phrases in the music…Experiments with musicians have confirmed this notion of hierarchical encoding in other ways. Most musicians cannot start playing a piece of music they know at any arbitrary location; musicians learn music according to a hierarchical phrase structure. Groups of notes form units of practice, these smaller units are combined into larger units, and ultimately into phrases; phrases are combined into structures such as verses and choruses of movements, and ultimately everything is strung together as a musical piece.

In a similar vein, related to musical theory, Steven Pinker summarizes the famous hierarchical theory of Jackendoff and Lerdahl [23, 22] in his fascinating book, *How The Mind Works* [28, 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, the repeating sequence of strong and weak beats that we count off as "ONE-two-THREE-four." The overall pattern is summed up in musical notation as the time signature…The third representation is a reductional 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.

In regard to the strong and weak beats referred to by Pinker, we observe that these are reflected by the relative thickness and darkness of the vertical segments in a percussion scalogram. For example, when listening to the *Dance Around* passage, 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.)



(a) (b)

**Figure 7. (a) Magnitudes of Gabor CWT of a pulse sequence. (b) Haar CWT of the same pulse sequence.**

With these descriptions of the hierarchical structure of music in mind, we now turn to representations of them within four different percussion sequences.

**Example 1: Rock Drumming**

In Figure 8 we show a percussion scalogram for a clip from a rock drum solo, which we have partially analyzed in the previous section. Here we complete our analysis by describing the hierarchy shown in the scalogram in a more formal, mathematical way, and then introducing the notion of *production rules* for the performance of the percussion sequence.

We can see that there are five separate groupings of drum strikes in the scalogram in Figure 8:

| $A$ | $B$ | $C$ | $B'$ | $C'$ |
|---|---|---|---|---|
| (1-level) | (2-levels) | (4-levels) | (2-levels) | (4-levels) |

The separate hierarchies within these groupings can be symbolized in the following way. We use the notation $*\!\!\downarrow$ to symbolize a "whole note", a longer duration, more emphasized strike. And the notation $*\!\!\downarrow^{-}$ to symbolize a "half note", a shorter duration, less emphasized strike. This allows us to symbolize the different emphases in the rhythm. Furthermore, the underscore symbol _ will be used to denote a rest between strikes (notes). For example, $*\!\!\downarrow^{-}\_*\!\!\downarrow$ symbolizes a half note followed by a rest followed by a whole note. Using this notation, the strikes in Figure 8 are symbolized by

$$*\!\!\downarrow\_*\!\!\downarrow\_*\!\!\downarrow\_*\!\!\downarrow\_*\!\!\downarrow*\!\!\downarrow*\!\!\downarrow^{-}\ *\!\!\downarrow*\!\!\downarrow^{-}\ *\!\!\downarrow*\!\!\downarrow*\!\!\downarrow*\!\!\downarrow\_*\!\!\downarrow\_*\!\!\downarrow\_*\!\!\downarrow*\!\!\downarrow^{-}\ *\!\!\downarrow*\!\!\downarrow*\!\!\downarrow*\!\!\downarrow^{-}\ *\!\!\downarrow*\!\!\downarrow$$

This notation is essentially equivalent to the standard notation for drumming used in musical scores (for examples of this notation, see [29] and [5]). We have thus shown that percussion scalograms can be used to read off a musical score for the drumming from its recorded sound. This is important because percussion playing is often extemporaneous, hence there is a need for notating particularly important extemporaneous passages as an aid to their repetition by other performers.

There is, however, much more information in a percussion scalogram. We can also use parentheses to mark off the groupings of the notes into their hierarchies, as follows:

$$*\!\!\downarrow\_(*\!\!\downarrow\_*\!\!\downarrow\_*\!\!\downarrow)\_((*\!\!\downarrow(*\!\!\downarrow*\!\!\downarrow^{-}))(*\!\!\downarrow*\!\!\downarrow^{-})(*\!\!\downarrow*\!\!\downarrow)(*\!\!\downarrow*\!\!\downarrow))$$
$$\_(*\!\!\downarrow\_*\!\!\downarrow)\_((*\!\!\downarrow(*\!\!\downarrow^{-}*\!\!\downarrow)(*\!\!\downarrow*\!\!\downarrow^{-}))(*\!\!\downarrow*\!\!\downarrow))$$

The advantage of this notation over the previous one is that the hierarchical groupings of notes is indicated. *We believe this enhanced notation, along with the videos that we create of sound with percussion scalograms, provide an important tool for analyzing the performance of percussion sequences. For example, they may be useful in teaching performance technique (recall Levitin's discussion of how musicians learn to play musical passages) by adding two adjuncts, notation plus video, to aid the ear in perceiving subtle differences in performance technique.*

In addition to this symbolic notation for percussion passages, there is an even deeper (and somewhat controversial) notion of *production rules* for the generation of these percussion sequences (analogous to Chomsky's notion of "deep structure" in linguistics that generates, via production rules, the syntactical hierarchy of sentences: [22, Section 11.4] and [7, Sections 5.2, 5.3]). Examples of these rules in music are described in [22, pp. 283–285, and p. 280]. Rather than giving a complete mathematical description at this time, we will

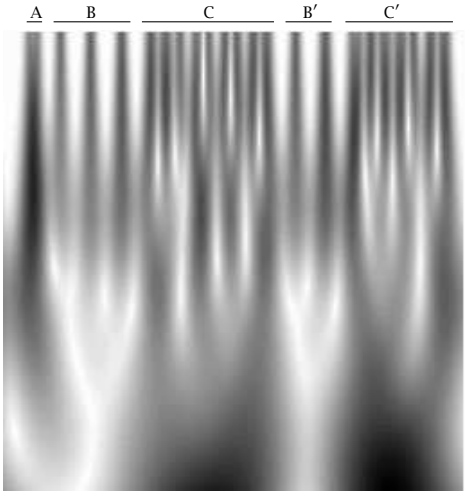



**Figure 8. Percussion scalogram for rock drumming. The labels are explained in Example 1. To view a video of the percussion scalogram being traced out along with the drumming sound, go to the URL in (2) and select the video for *Dance Around percussion.***

simply give a couple of examples. For instance, the grouping $B'$ in the passage is produced from the grouping $B$, by clipping two strikes off the end:

$$B' \leftarrow \mathsf{End}(B)$$

As another example, if we look at the starting notes in the groups $C$ and $C'$ defined by

$$\mathsf{Start}(C) := \text{𝅘𝅥}(\text{𝅘𝅥𝅮}\,\text{𝅘𝅥𝅮}) \quad \text{and} \quad \mathsf{Start}(C') := \text{𝅘𝅥}(\text{𝅘𝅥𝅮}\,\text{𝅘𝅥})$$

then $\mathsf{Start}(C')$ is produced from $\mathsf{Start}(C)$ by a modulation of emphases:

$$\mathsf{Start}(C') \leftarrow \mathsf{Modulation}(\mathsf{Start}(C))$$

In this paper we are only giving these two examples of production rules, in order to give a flavor of the idea. A more complete discussion is a topic for a future paper.

### Example 2: African Drumming

Our second example is a passage of African drumming, clipped from the beginning of the song *Welela* from an album by Miriam Makeba. In this case, the spectrogram of the passage, shown in Figure 9(a), has some horizontal banding at lower frequencies that adversely affect the percussion scalogram by raising the mean value of the spectrogram averages. Consequently, we used only values from the spectrogram that are above 1000 Hz to compute the percussion scalogram shown in Figure 9(b). By listening to the video referenced in the caption of Figure 9, you should find that this percussion scalogram does accurately capture the timing and grouping of the drum strikes in the passage.

This passage is quite interesting in that it is comprised of only 20 drum strikes, yet we shall see that it contains a wealth of complexity. First, we can see that there are seven separate groupings of drum strikes in the scalogram in Figure 9:

| *A* | *B* | *C* | *A* | *B* | *C* | *D* |
|---|---|---|---|---|---|---|
| (1-level) | (3-level) | (3-level) | (1-level) | (3-level) | (3-level) | (2-level) |

Notice the interweaving of different numbers of levels within this sequence of groups. Second, the drum strikes can be notated with hierarchical grouping as follows:

$$\text{𝅘𝅥𝅮}\,\_\_\_((\text{𝅘𝅥𝅮}\,\text{𝅘𝅥})\_(\text{𝅘𝅥}\,\_\,\text{𝅘𝅥𝅮}))\_((\text{𝅘𝅥𝅮}\,\text{𝅘𝅥})(\text{𝅘𝅥𝅮})(\text{𝅘𝅥}))$$

$$\_\_\_\text{𝅘𝅥𝅮}\,\_\_\_((\text{𝅘𝅥𝅮}\,\text{𝅘𝅥})\_(\text{𝅘𝅥}\,\_\,\text{𝅘𝅥𝅮}))\_((\text{𝅘𝅥𝅮}\,\text{𝅘𝅥})(\text{𝅘𝅥𝅮})(\text{𝅘𝅥}))\_(\text{𝅘𝅥𝅮}\,\text{𝅘𝅥})$$

This passage is interesting not only in terms of the complex hierarchical grouping of notes, but also because of the arrangement of the time intervals between notes. It is a well-known fact among musicians that the silences between notes are at least as important as the notes themselves. In this passage we have the following sequence of time-intervals between notes (1 representing a short rest, 2 representing a long rest, and 0 representing no rest):

$$2\,0\,1\,1\,1\,0\,0\,0\,2\,2\,0\,1\,1\,1\,0\,0\,0\,1\,0$$

which quantitatively describes the "staggered" sound of the drum passage. (The reader might find it interesting to compute the sequence of rests for Example 1, and verify that it is less staggered, with longer sequences of either 1's or 0's.)

### Example 3: Jazz Drumming

In this example we consider a couple of cases of jazz drumming. In Figure 10(a) we show a percussion scalogram created from the drum solo at the beginning of the jazz classic, *Sing Sing Sing.* The tempo of this drumming is very fast. Our notation for this sequence was obtained from examining the percussion scalogram both as a picture and as the video sequence (referred to in the caption of the figure) is played. Here is the notated sequence:

**Very fast**

$$\text{𝅘𝅥}\,\_\_\_\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥𝅮}\,\_\,\text{𝅘𝅥}\,\_\,\text{𝅘𝅥𝅮}\,\_\,\text{𝅘𝅥}\,\text{𝅘𝅥}\_\,\text{𝅘𝅥}\,\_\_\_\,\text{𝅘𝅥}\,\_\_\_\,\text{𝅘𝅥}\,\_\,\text{𝅘𝅥𝅮}\,\_$$

$$\text{𝅘𝅥}\,\_\_\_\,\text{𝅘𝅥}\,\_\,\text{𝅘𝅥}\,\_\_\_\,\text{𝅘𝅥𝅮}\,\_\,\text{𝅘𝅥}\,\_\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥𝅮}\,\_\,\text{𝅘𝅥}\,\_\_\_\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥𝅮}\,\_\,\text{𝅘𝅥}$$

Our notation differs considerably from the notation given for this beginning drum solo in the original score [29], the first several notes being:

**Jungle Drum Swing**

$$\text{𝅘𝅥}\,\text{𝅘𝅥}\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥}\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥}\,\text{𝅘𝅥𝅮}\,\text{𝅘𝅥}\,\text{𝅘𝅥𝅮}$$

Setting aside its racist overtones, we observe that this tempo instruction is not terribly precise. We can see from comparing these two scores, that the drummer (Gene Krupa) is improvising the percussion (as is typical with jazz). Our percussion

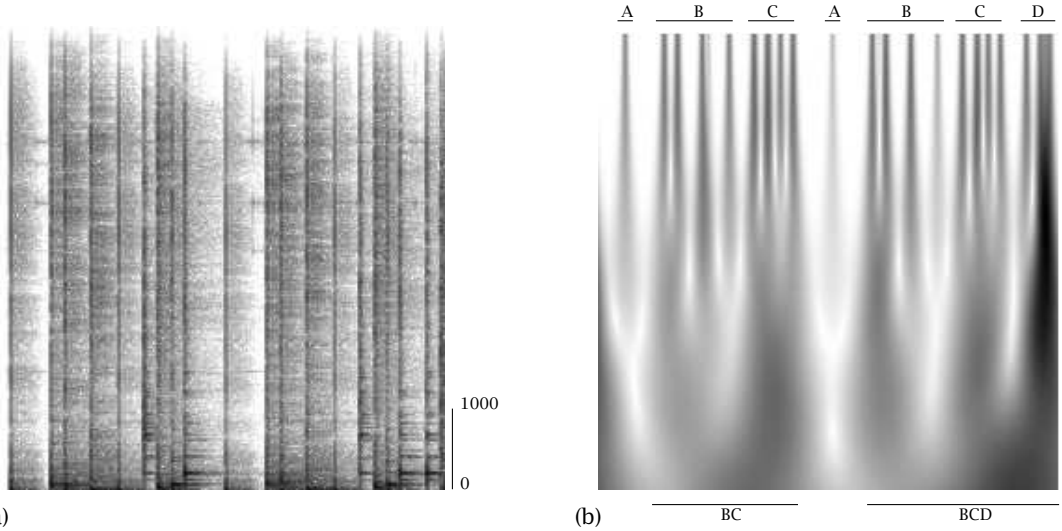



**Figure 9. (a) Spectrogram for African drumming. Between 0 and 1000 Hz, as marked on the right side of (a), there are a considerable number of horizontal line segments. Those segments adversely affect the percussion scalogram. Consequently only frequencies above 1000 Hz are used to create the percussion scalogram. (b) Percussion scalogram for African drumming, using frequencies above 1000 Hz. The labels are explained in Example 2. To view a video of the percussion scalogram being traced out along with the drumming sound, go to the URL in** (2) **and select the video for *Welela percussion.***

scalogram method allows us to derive a precise notation for Krupa's improvisation. We leave it as an exercise for the reader to notate the hierarchical structure of this drum passage, based on the percussion scalogram. From our notation above, we find that the pattern of rests in Krupa's playing has this structure:

2 0 1 1 1 0 1 2 2 1 1 2 1 2 1 1 0 0 0 1 2 0 1

Here, as with the African drumming, we see a staggered pattern of rests.

Our second example of jazz drumming is a clip of the beginning percussive passage from another jazz classic, *Unsquare Dance.* In the score for the piece [5], the following pattern of strikes (indicated as *hand clapping*)

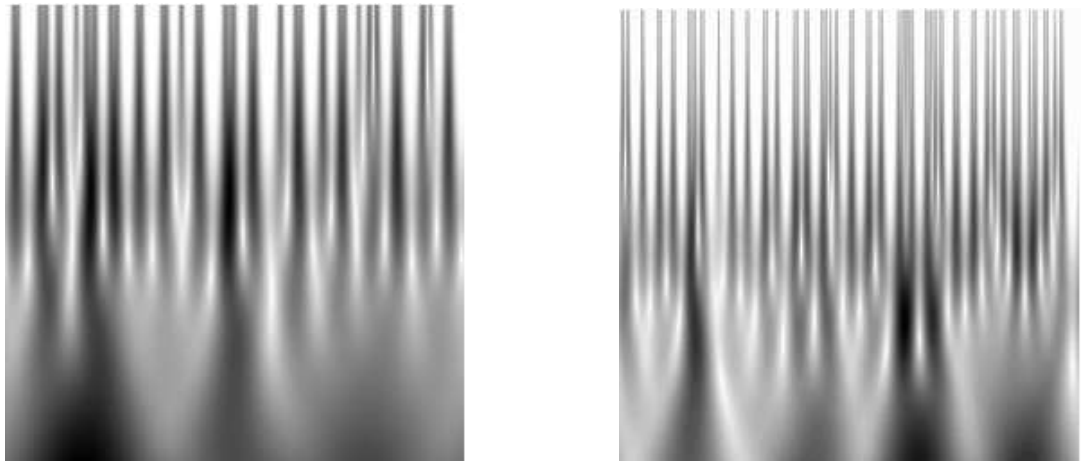
_ * _ * _ * *

is repeated in each measure (consistent with the 7/4 time signature). Listening to the passage as the video is played, we can hear this repeated series of "strikes" as groups of very fast individual strikings of drumsticks. The drummer (Joe Morello) is improvising on the notated score by replacing individual hand claps by these very rapid strikings of his drumsticks. It is noteworthy that, in many instances, the percussion scalogram is sensitive enough to record the timings of the individual drumstick strikings. The scalogram is thus able to reveal, in a visual representation, the double aspect to the rhythm: individual drum strikings within the larger groupings notated as hand claps in the original score.

These examples are meant to illustrate that the percussion scalogram method can provide useful musical analyses of drumming rhythms. Several more examples are given at the *Pictures of Music* website [6]. We now provide some examples of using both spectrograms and percussion scalograms to analyze both the melodic and rhythmic aspects of music. Because they are based on an assumption of intense pulsing in the musical signal due to percussion, which is only satisfied for some tonal instruments, percussion scalograms do not always provide accurate results for tonal instruments. However, when they do provide accurate results (a precise description of the timings of the notes), they reveal the rhythmic structure of the music (which is our goal). We now provide three examples of successful analyses of melody and rhythm.

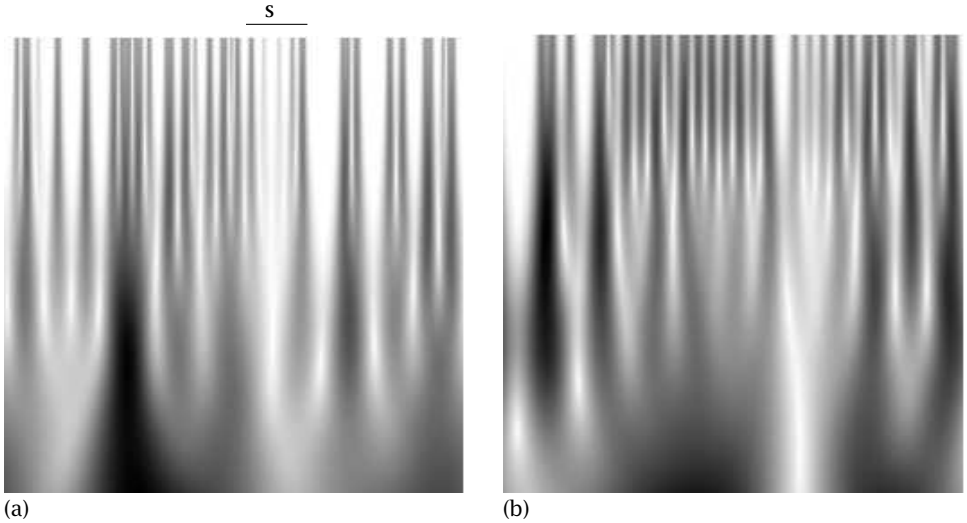
### Example 4: A Jazz Piano Melody

In Figure 11(a) we show a percussion scalogram of a recording of a jazz piano improvisation by Erroll Garner. It was captured from a live recording [18]. Since this is an improvisation, there is no musical score for the passage. Several aspects of the scalogram are clearly evident. First, we can see a staggered spacing of rests as in the African drumming in Example 2 and the jazz drumming in the *Sing Sing Sing* passage in Example 3. There



(a) (b)

**Figure 10. (a) Percussion scalogram for drum solo in *Sing Sing Sing* using frequencies above 1000 Hz. (b) Percussion scalogram for complex drum stick percussion in *Unsquare Dance* using frequencies above 2000 Hz. To view videos of these percussion scalograms being traced out along with the drumming sound, go to the URL in (2) and select the videos for *Sing Sing Sing percussion* or *Unsquare Dance percussion.***




(a) (b)

**Figure 11. (a) Percussion scalogram of a clip from an Erroll Garner jazz piano passage (using frequencies above 1800 Hz). To view a video of the percussion scalogram being traced out along with the piano playing, go to the URL in (2) and select the video for *Erroll Garner piano recording.* The label S indicates a syncopation in the melody. (b) Percussion scalogram from a clip of a piano interpretation of a Bach melody (using frequencies above 3000 Hz). To view a video of the percussion scalogram being traced out along with the piano sound, go to the URL in (2) and select the video for *Bach piano piece (scalogram).***

is also a syncopation in the melody, indicated by the interval marked **S** in Figure 11(a). By syncopation we mean an altered rhythm, "ONE-two-three-FOUR," rather than the more common "ONE-two-THREE-four." The percussion scalogram provides us with a visual representation of these effects, which is an aid to our listening comprehension. Although the percussion scalogram does not perform perfectly here (for example the last note in the sequence marked **S** is split in two at the top; the scalogram has detected the attack and the decay of the note), when viewed as a video the percussion scalogram does enable us to quickly identify the timing and hierarchical grouping of the notes (which would be much more difficult using only our ears).

### Example 5: A Bach Piano Transcription

As a simple contrast to the previous example, we briefly discuss the percussion scalogram shown in Figure 11(b), obtained from a piano interpretation of a Bach melody, *Jesu, Joy of Man's Desiring.* The sound recording used was created from a MIDI sequence. In contrast to the previous jazz piece, this classical piece shows no staggering of rests, and no syncopation. The hierarchy of groupings of notes is also more symmetrical than for the jazz piece. This hierarchy of notes, the rhythm of the passage, is easily discernable from this percussion scalogram, while it is not clearly evident from the score [2] (at least to untrained musicians).

### Example 6: A Jazz Orchestral Passage

For our final example, we analyze the spectrogram and percussion scalogram shown in Figure 12. They were obtained from a passage from a recording of the jazz orchestral classic, *Harlem Air Shaft,* by Duke Ellington. This passage is quite interesting in that it is comprised of only about 15 notes, yet we shall see that it contains a wealth of complexity. (We saw this in the African drum passage as well; perhaps we have an aspect of aesthetic theory here.) We now describe some of the elements comprising the rhythm and melody within this passage. It should be noted that, although there is a score for *Harlem Air Shaft,* that score is a complex orchestration that requires a large amount of musical expertise to interpret. Our spectrogram/percussion scalogram approach provides a more easily studied description of the melody and rhythm, including visual depictions of length and intensity of notes from several instruments playing simultaneously. *Most importantly, the spectrogram provides an objective description of recorded performances. It can be used to compare different performances in an objective way. Our percussion scalograms facilitate the same kind of objective comparison of the rhythm in performances.*

(1) *Reflection of Notes.* The passage contains a sequence of high pitched notes played by a slide trombone (wielded by the legendary "Tricky Sam" Nanton). This sequence divides into two groups of three, enclosed in the rectangles labeled **T** and $\mathcal{R}$**T** in the spectrogram shown at the top of Figure 12. The three notes within **T** are located at frequencies of approximately 855, 855, and 845 Hz. They are then reflected about the frequency 850, indicated by the line segment $\mathcal{M}$ between the two rectangles, to produce the three notes within $\mathcal{R}$**T** at frequencies of approximately 845, 845, and 855 Hz. The operation of reflection $\mathcal{R}$ about a specific pitch is a common, group-theoretical, operation employed in classical music [20].

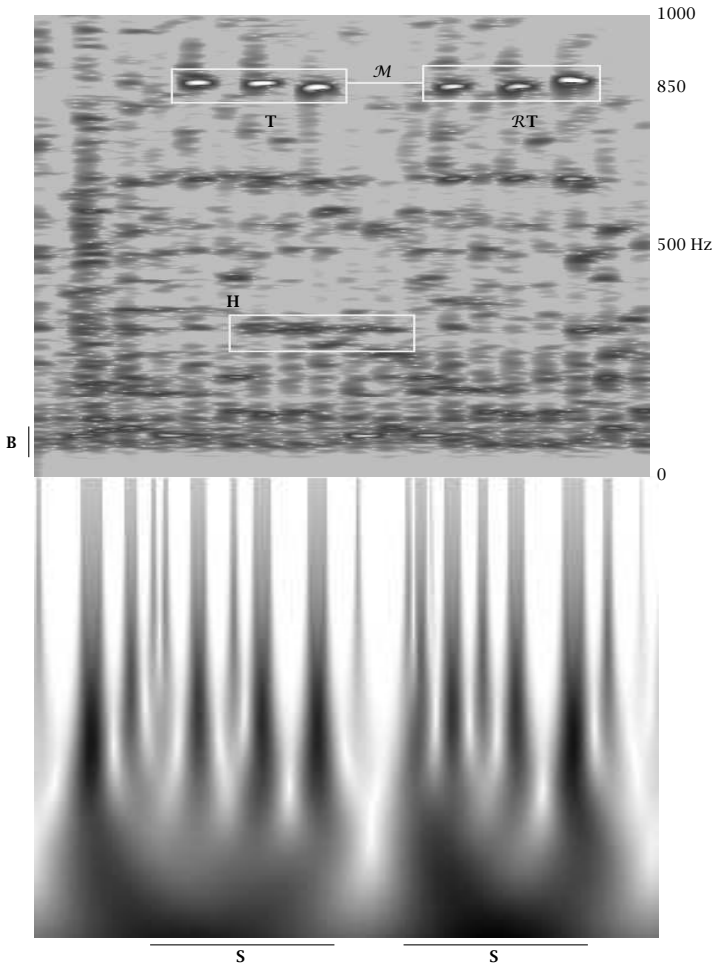



**Figure 12. Top: Spectrogram of a passage from a recording of *Harlem Air Shaft.* Bottom: Percussion scalogram of the same passage (using all frequencies). The boxed regions and labels are explained in the text. To view videos of this spectrogram and scalogram, go to the URL in** (2) **and select the links indicated by *Harlem Air Shaft.***

(2) *Micro-Tones.* The pitch interval described by going down in frequency from 855 to 845 Hz is

$$\log_2(855/845) = 0.017 \approx 1/48,$$

which is about 1/4 of the (logarithmic) half-tone change of 1/12 of an octave (on the 12-tone chromatic scale). Thus within **T** the pitch descends by an *eighth-tone.* Similarly, in $\mathcal{R}$**T** the pitch ascends at the end by an eighth-tone. These are *micro-tone* intervals, intervals that are not representable on the standard 12-tone scoring used in Western classical music. They are, in fact, half the interval of the quarter-tones that are a characteristic

of jazz music (based on its roots in African tonal scales [4]). Here Ellington *synthesizes a melodic characteristic of jazz (micro-tones) with one of classical music (reflection about a pitch level).*

(3) *Staggered Syncopation.* Viewing the video of the percussion scalogram shown in Figure 12, we note that there is a staggering of rests between notes, and we also observe a syncopation. This syncopation occurs when the slide trombone slides into and between the emphasized notes in the structures **T** and $\mathcal{R}$**T**. The pattern lying above the segments marked **S** in the figure is

one – TWO – three – FOUR – (rest) – FIVE

an unusual, *staggered syncopation,* of rhythm. (One thing we observe when viewing the video is that the percussion scalogram does capture the timings of most of the note attacks, although it does not perfectly reflect some muted horn notes—because the attacks of those notes are obscured by the higher volume trombone notes. Nevertheless, the scalogram provides an adequate description of the driving rhythm of the trombone notes, and the muted trumpet notes at the end of the passage.)

(4) *Hierarchies of Melody.* Within the passage there are several different types of melody, over different length time-scales. First, we note that there is the hierarchy of **T** and $\mathcal{R}$**T** at one level along with their combination into one long passage, linked melodically by a sequence of muted horn notes **H**. Notice that the pitch levels in **H** exhibit, over a shorter time-scale, the pitch pattern shown in the notes within **T** together with $\mathcal{R}$**T**. That is a type of hierarchical organization of melodic contour. There is a further hierarchical level (in terms of a longer time-scale) exhibited by the melodic contour of the bass notes (shown as a long sinusoidal arc within the region marked by **B**). Here the bassist, Jimmy Blanton, is using the bass as a plucked melodic instrument as well as providing a regular tempo for the other players. This is one of his major innovations for the bass violin in jazz instrumentation.

We can see from this analysis that this passage within just 6 seconds reveals a wealth of structure, including many features that are unique to jazz. Such mastery illustrates why Duke Ellington was one of the greatest composers of the twentieth century.

## Justification of the Percussion Scalogram Method

### Choosing the Width and Frequency Parameters

In this section we discuss how the parameters are chosen to provide a satisfactory display of a scalogram for a pulse train, which is the second step of the percussion scalogram method. The term *satisfactory* means that both the average number of pulses/sec (beats/sec) are displayed and the individual beats are resolved.

To state our result we need to define several parameters. The number $T$ will stand for the time duration of the signal, while $B$ will denote the total number of pulses in the pulse train signal. We will use the positive parameter $p$ to scale the width $\omega$ and frequency $\nu$ defined by

$$\omega = \frac{pT}{2B}, \quad \nu = \frac{B}{pT}. \tag{11}$$

Notice that $\omega$ and $\nu$ are in a reciprocal relationship; this is in line with the reciprocal relation between time-scale and frequency that is used in wavelet analysis. Notice also that the quantity $B/T$ in $\nu$ is equal to the average number of pulses/sec. The best choice for the parameter $p$ in these formulas will be described below. Two further parameters are the number of octaves $I$ and voices $M$ used in the percussion scalogram. We shall see that these two parameters will depend on the value of $\delta$, the minimum length of a 0-interval (minimum space between two successive pulses).

Now that we have defined our parameters, we can state our main result:

*Given the constraints of using positive integers for the octaves $I$ and voices $M$ and using 256 total correlations, satisfactory choices for the parameters of a percussion scalogram are:*

$$\begin{gathered} \omega = \frac{pT}{B}, \quad \nu = \frac{B}{pT}, \quad p = 4\sqrt{\pi} \\ I = \left\lfloor \log_2\left(\frac{p^2T^2}{\delta B^2}\right) - \frac{3}{2} \right\rfloor, \quad M = \left\lfloor \frac{256}{I} \right\rfloor. \end{gathered} \tag{12}$$

The remainder of this section provides the rationale for this result (notice that the value of $\omega$ in (12) is twice the value given in (11); we shall explain why below). While this rationale may not be a completely rigorous proof, it does provide useful insights into how a Gabor CWT works with pulse trains, and it does provide us with the method used to produce the scalograms for the pulse trains shown in the section "Examples of Rhythmic Analysis" (and for the other examples given at the *Pictures of Music* website [6]). In fact, we have found that in every case, the method provides a satisfactory display of the scalogram for a pulse train from a musical passage (whether the pulse train is accurate is a separate question; we are still

working on extending its capabilities as described previously).

We shall denote a given pulse train by $\{\mathcal{P}(t_m)\}$. (Here we use $\{t_m = \tau_m\}$ as the time-values; this notational change is clarified by looking at equation (16).) This signal satisfies $\mathcal{P}(t_m) = 1$ during the duration of a beat and $\mathcal{P}(t_m) = 0$ when there is no beat. We use the Gabor wavelet in (7) to analyze the pulse signal. Since we are using the complex Gabor wavelet, we have both a real and imaginary part:

$$\Psi_{\Re}(t) = \omega^{-\frac{1}{2}} e^{-\pi(t/\omega)^2} \cos\frac{2\pi\nu t}{\omega} \tag{13}$$

$$\Psi_{\Im}(t) = \omega^{-\frac{1}{2}} e^{-\pi(t/\omega)^2} \sin\frac{2\pi\nu t}{\omega}. \tag{14}$$

These real and imaginary parts have the same envelope function:

$$\Psi_E(t) = \omega^{-\frac{1}{2}} e^{-\pi(t/\omega)^2}. \tag{15}$$

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 $t = 0$. 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 13.

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 (15). As $\nu$ is increased, the Gabor wavelet oscillates much more quickly. See Figure 14.

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

$$(\mathcal{P} : \Psi_s)(\tau_k) = \sum_{m=0}^{M} \mathcal{P}(t_m)\overline{s^{-1/2}\Psi([t_m - \tau_k]s^{-1})}. \tag{16}$$

Since $\{\mathcal{P}(t_m)\}$ is a binary signal, the terms of this sum will equal $\overline{s^{-1/2}\Psi([t_m - \tau_k]s^{-1})}$ 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 $\tau_k$, $\overline{s^{-1/2}\Psi([t_m - \tau_k]s^{-1})}$ 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

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

which will represent the striking of an instrument. So as $s$ reaches its smallest values, near $s = 2^{-I}$, 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 the function

$$\overline{s^{-1/2}\Psi([t_m - \tau_k]s^{-1})}$$

being dampened very quickly, so very little 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 $\overline{s^{-1/2}\Psi([t_m - \tau_k]s^{-1})}$ is being dampened. But with the wavelet being dampened more slowly now, the values of $\overline{s^{-1/2}\Psi([t_m - \tau_k]s^{-1})}$ 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 $B/T$ as our measure of the average beats per second. The average 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 (11), with parameter $p > 0$ used as a scaling factor. With these width and frequency parameters, the Gabor wavelet is

$$\Psi(t) = \sqrt{\frac{2B}{pT}}\, e^{-\pi(2Bt/pT)^2} e^{i4\pi B^2/(p^2T^2)}. \tag{17}$$

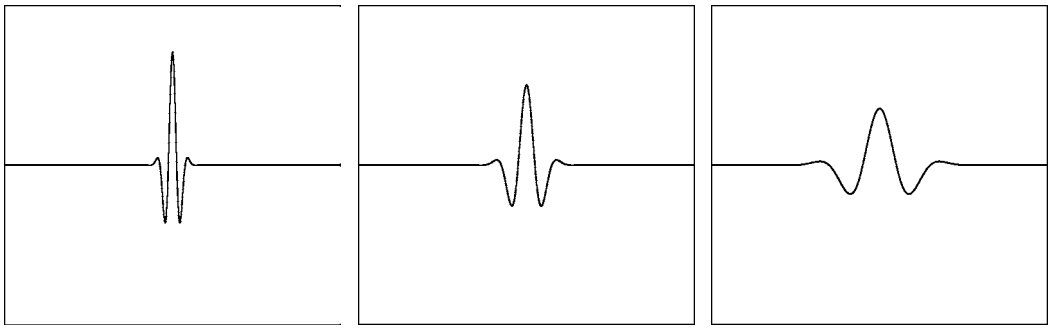
We want to detect beats that are within $T/B$, the average time between beats, of each other. Likewise, we want separation of the beats that are not within $T/B$ of each other. We accomplish this by inspecting the envelope function evaluated at $t = T/B$,

$$\Psi_E\left(\frac{T}{B}\right) = \sqrt{\frac{2B}{pT}}\, e^{-4\pi/p^2}. \tag{18}$$

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

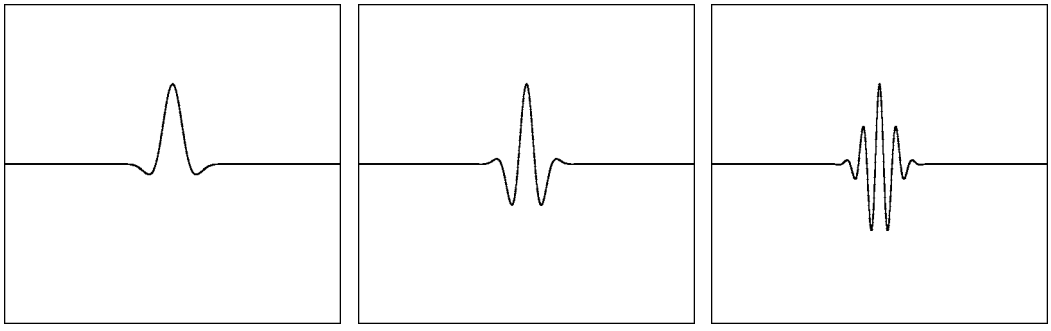
$$M(p) = \sqrt{\frac{2B}{pT}}\, e^{-4\pi/p^2}. \tag{19}$$

Remembering that $T$ and $B$ are constants determined by the percussion sound signal, the maximization of the magnitude of the wavelet



(a) $\omega = 0.5$ (b) $\omega = 1$ (c) $\omega = 2$

**Figure 13. 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 $[-2, 2]$.**



(a) $\nu = 0.5$ (b) $\nu = 1$ (c) $\nu = 2$

**Figure 14. Real parts of Gabor wavelet with width parameter $\omega = 1$ and frequency parameter $\nu = 0.5, 1$, and 2. For each graph, the horizontal range is $[-5, 5]$ and the vertical range is $[-2, 2]$.**

at $t = T/B$ 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 $t = T/B$, 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 wide. 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 $\Psi_{\Re}$, 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).

### 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 (11), 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

$$\omega = \frac{pT}{B}, \quad \nu = \frac{B}{pT}, \quad p = 4\sqrt{\pi} \tag{20}$$

for displaying our percussion scalograms.

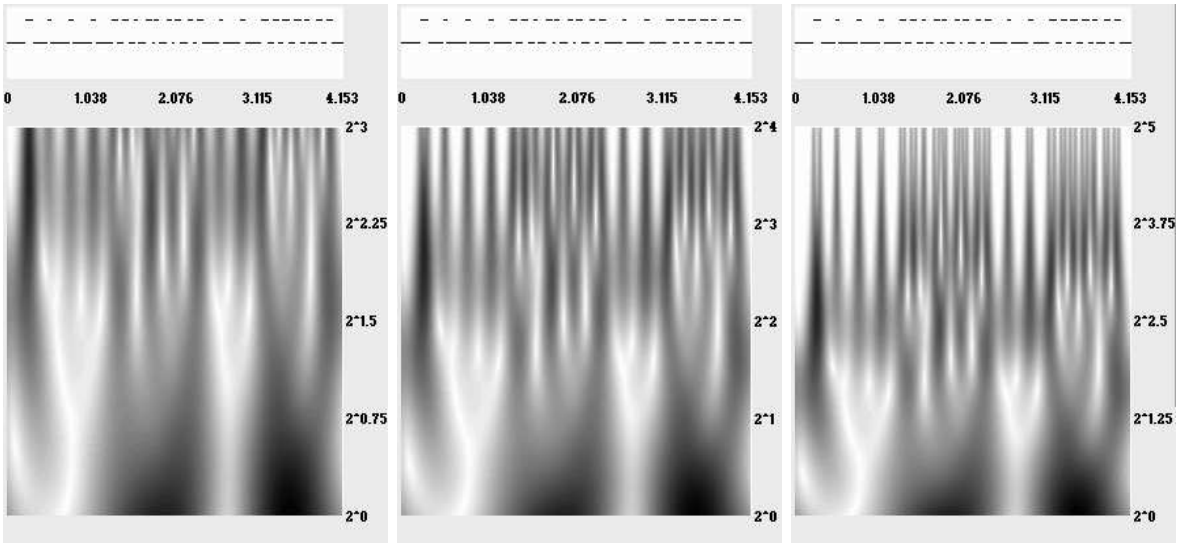
### Choosing Octaves and Voices

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




(a) $I = 3$ (b) $I = 4$ (c) $I = 5$

**Figure 15. Examples of percussion scalograms using different values of $I$, the number of octaves, for the *Dance Around* percussion passage. Graph (b) uses the value of $I = 4$ calculated from Equation (23).**

maximum pulse frequency should be one-half of $1/\delta$. Setting this maximum pulse frequency equal to the maximum wavelet frequency, we have

$$\frac{1}{2\delta} = \frac{\nu}{\omega}\, 2^I. \tag{21}$$

Notice that both sides of (21) have units of beats/sec.

Using the equations for $\nu$ and $\omega$ in (20), we rewrite Equation (21) as

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

Solving for $I$ yields

$$I = \log_2\left(\frac{p^2T^2}{\delta B^2}\right) - 1. \tag{22}$$

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

$$I = \left\lfloor \log_2\left(\frac{p^2T^2}{\delta B^2}\right) - \frac{3}{2} \right\rfloor. \tag{23}$$

To illustrate the value of selecting $I$ per Equation (23), in Figure 15 we show three different scalograms for the *Dance Around* percussion sequence. For this example, Equation (23) yields the value $I = 4$. Using this value, we find that the scalogram plotted in Figure 15(b) is able to detect the individual drum strikes and their groupings. If, however, we set $I$ too low, say $I = 3$ in Figure 15(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 15(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 [35] the number of correlations used in a scalogram is constrained to be no more than 256, in which case we set

$$M = \left\lfloor \frac{256}{I} \right\rfloor \tag{24}$$

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

## Conclusion

In this paper we have described the way in which spectrograms and percussion scalograms can be used for analyzing musical rhythm and melody. While percussion scalograms work fairly effectively on brief percussion passages, more research is needed to improve their performance on a wider variety of music (especially when the volume is highly variable). We only briefly introduced the use of spectrograms for analyzing melody and its hierarchical structure; more examples are discussed in [34] and at the website [6]. Our discussion

showed how percussion scalograms could be used to distinguish some styles of drumming, but much more work remains to be done. Further research is also needed on using local averages, instead of the global average $A$ that we employed, and on determining what additional information can be gleaned from the phases of the Gabor CWTs.


## Acknowledgment

This research was supported by the National Science Foundation Grant 144-2974/92974 (REU site SUREPAM program). We also thank William Sethares for his helpful comments on an earlier version of this paper.


## References


[1] J. F. ALM and J. S. WALKER, Time-frequency analysis of musical instruments, *SIAM Review* **44** (2002), 457–476. Available at http://www.uwec.edu/walkerjs/media/38228[1].pdf.

[2] J. S. BACH, (arranged and adapted by M. Scott), *Jesu, Joy of Man's Desiring*. A free version of the part of the score used for the clip analyzed here can be obtained from http://www.musicnotes.com/sheetmusic/mtd.asp?ppn=MN0026859

[3] P. BALAZS et al., Double preconditioning for Gabor frames, *IEEE Trans. on Signal Processing* **54** (2006), 4597–4610.

[4] L. BERNSTEIN, The world of jazz, *The Joy of Music,* Amadeus Press, Pompton Plains, NJ, 2004, pp. 106–131

[5] D. BRUBECK, *Unsquare Dance*. The beginning of the score, containing the hand-clapping introduction, can be viewed for free at http://www.musicnotes.com/sheetmusic/mtd.asp?ppn=MN0042913.

[6] X. CHENG, et al., *Making Pictures of Music: New Images and Videos of Musical Sound.* Available at http://www.uwec.edu/walkerjs/PicturesOfMusic/. Examples of more analyses of rhythmic percussion can be found at http://www.uwec.edu/walkerjs/PicturesOfMusic/MusicalRhythm.htm.

[7] N. CHOMSKY, *Syntactic Structures*, Mouton de Gruyter, New York, 2002.

[8] C. K. CHUI, *Wavelets: A Mathematical Tool for Signal Analysis*, SIAM, Philadelphia, PA, 1997.

[9] R. COGAN, *New Images of Musical Sound*, Harvard University Press, Cambridge, MA, 1984.

[10] I. DAUBECHIES, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

[11] I. DAUBECHIES and S. MAES, A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models, in *Wavelets in Medicine and Biology*, CRC Press, Boca Raton, FL, 1996, pp. 527–546.

[12] M. DÖRFLER, *Gabor analysis for a class of signals called music*, dissertation, University of Vienna, 2002.

[13] ———, Time-frequency analysis for music signals—A mathematical approach, *Journal of New Music Research* **30**, No. 1, March 2001.

[14] M. DÖRFLER and H. G. FEICHTINGER, Quantitative description of expression in performance of music, using Gabor representations, *Proceedings of the Diderot Forum on Mathematics and Music,* 1999, Vienna.

[15] H. FEICHTINGER and T. STROHMER, eds., *Gabor Analysis and Algorithms*, Birkhäuser, Boston, MA, 1998.

[16] ———, eds. *Advances in Gabor Analysis,* Birkhäuser, Boston, MA, 2002.

[17] D. GABOR, Theory of communication, *Journal of the Institute for Electrical Engineers* **93** (1946), 873–880.

[18] Erroll Garner recording from http://www.youtube.com/watch?v=H0-ukGgAEUo.

[19] K. GRÖCHENIG, *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston, MA, 2001.

[20] L. HARKLEROAD, *The Math Behind the Music*, Cambridge University Press, Cambridge, UK, 2006.

[21] D. KROODSMA, *The Singing Life of Birds*, Houghton-Mifflin, NY, 2005.

[22] F. LERDAHL and R. JACKENDOFF, *A Generative Theory of Tonal Music*, MIT Press, Cambridge, MA, 1983.

[23] F. LERDAHL and R. JACKENDOFF, An overview of hierarchical structure in music, in *Machine Models of Music* (1992), S. Schwanauer and D. Levitt, eds., MIT Press, Cambridge, MA, 289–312.

[24] D. LEVITIN, *This Is Your Brain On Music*, Dutton, New York, 2006.

[25] G. LOY, *Musimathics: The Mathematical Foundations of Music*, Vol. 2, MIT Press, Cambridge, MA, 2007.

[26] F. B. MACHE, *Music, Myth and Nature, Contemporary Music Studies*, Vol. 6., Taylor & Francis, London, 1993.

[27] S. MALLAT, *A Wavelet Tour of Signal Processing*, Second Edition, Academic Press, San Diego, CA, 1999.

[28] S. PINKER, *How the Mind Works*, Norton, NY, 1997.

[29] L. PRIMA, *Sing Sing Sing*. The initial part of the score, containing the drum solo introduction, can be viewed for free at http://www.musicnotes.com/sheetmusic/scorch.asp?ppn=SC0005700.

[30] D. ROTHENBERG, *Why Birds Sing: A Journey into the Mystery of Bird Song*, Basic Books, NY, 2005.

[31] W. SETHARES, *Rhythm and Transforms*, Springer, New York, NY, 2007.

[32] L. M. SMITH, Modelling rhythm perception by continuous time-frequency analysis, *Proceedings of the 1996 International Computer Music Conference,* Hong Kong, 392–395.

[33] ———, A multiresolution time-frequency analysis and interpretation of musical rhythm, thesis, University of Western Australia, 2000.

[34] J. S. WALKER, *A Primer on Wavelets and their Scientific Applications*, Second Edition, Chapman & Hall/CRC Press, Boca Raton, FL, 2008. Material referenced from Chapters 5 and 6 is available at http://www.uwec.edu/walkerjs/primer/Ch5extract.pdf; http://www.uwec.edu/walkerjs/primer/Ch6extract.pdf.

[35] ———, Document on using AUDACITY and FAWAV. Available at http://www.uwec.edu/walkerjs/tfamr/DocAudFW.pdf.

[36] J. S. WALKER and G. W. DON, Music: A time-frequency approach, preprint, available at http://www.uwec.edu/walkerjs/media/TFAM.pdf.