Music: Broken Symmetry, Geometry, and Complexity

Gary W. Don, Karyn K. Muir, Gordon B. Volk, James S. Walker

The relation between mathematics and music has a long and rich history, including: Pythagorean harmonic theory, fundamentals and overtones, frequency and pitch, and mathematical group theory in musical scores [7, 47, 56, 15]. This article is part of a special issue on the theme of mathematics, creativity, and the arts. We shall explore some of the ways that mathematics can aid in creativity and understanding artistic expression in the realm of the musical arts. In particular, we hope to provide some intriguing new insights on such questions as:

- *Does Louis Armstrong's voice sound like his trumpet?*
- *What do Ludwig van Beethoven, Benny Goodman, and Jimi Hendrix have in common?*
- *How does the brain fool us sometimes when listening to music? And how have composers used such illusions?*
- *How can mathematics help us create new music?*
- *Melody contains both pitch and rhythm. Is it possible to objectively describe their connection?*
- *Is it possible to objectively describe the complexity of musical rhythm?*

In discussing these and other questions, we shall outline the mathematical methods we use and provide some illustrative examples from a wide variety of music.

The paper is organized as follows. We first summarize the mathematical method of Gabor transforms (also known as short-time Fourier transforms, or spectrograms). This summary emphasizes the use of a discrete *Gabor frame* to perform the analysis. The section that follows illustrates the value of spectrograms in providing objective descriptions of musical performance and the geometric time-frequency structure of recorded musical sound. Our examples cover a wide range of musical genres and interpretation styles, including: Pavarotti singing an aria by Puccini [17], the 1982 Atlanta Symphony Orchestra recording of Copland's *Appalachian Spring* symphony [5], the 1950 Louis Armstrong recording of "La Vie en Rose" [64], the 1970 rock music introduction to "Layla" by Duane Allman and Eric Clapton [63], the 1968 Beatles' song "Blackbird" [11], and the Renaissance motet, "Non vos relinquam orphanos", by William Byrd [8]. We then discuss signal synthesis using dual Gabor frames, and illustrate how this synthesis can be used for processing recorded sound and creating new music. Then we turn to the method of continuous wavelet transforms and show how they can be used together with spectrograms for two applications: (1) zooming in on spectrograms to provide more detailed views and (2) producing objective time-frequency

Gary W. Don is professor of music at the University of Wisconsin-Eau Claire. His email address is dongw@uwec.edu.

Karyn K. Muir is a mathematics major at the State University of New York at Geneseo. Her email address is kkm5@geneseo.edu.

Gordon B. Volk is a mathematics major at the University of Wisconsin-Eau Claire. His email address is volkgb@uwec.edu.

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

portraits of melody and rhythm. The musical illustrations for these two applications are from a 1983 Aldo Ciccolini performance of Erik Satie's "Gymnopédie I" [81] and a 1961 Dave Brubeck jazz recording "Unsquare Dance" [94]. We conclude the paper with a quantitative, objective description of the complexity of rhythmic styles, combining ideas from music and information theory.

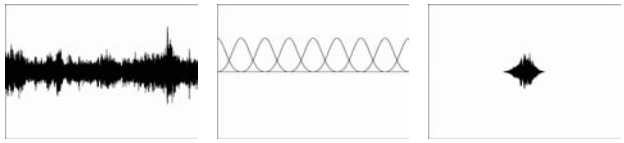
Discrete Gabor Transforms: Signal Analysis

We briefly review the widely employed method of Gabor transforms [53], 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* [27] — a book that still deserves close study. A more recent contribution is [62]. In [37, 38], Dörfler describes the fundamental mathematical aspects of using Gabor transforms for musical analysis. Other sources for theory and applications of short-time Fourier transforms include [3, 76, 19, 83, 65]. There is also considerable mathematical background in [50, 51, 55], with musical applications in [40]. Using sonograms or spectrograms for analyzing the music of birdsong is described in [61, 80, 67]. The theory of Gabor transforms is discussed in complete detail in [50, 51, 55] from the standpoint of function space theory. Our focus here, however, will be on its discrete aspects, as we are going to be processing digital recordings.

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, $\{\tau_\ell = t_{j\ell}\}_{\ell=0}^{M}$, 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. This yields the Gabor transform of $\{f(t_k)\}$:

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

We shall describe (1) more explicitly in a moment (see Remark 1 below). For now, 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 zeros; hence the full supports of all windows are included.



(a) (b) (c)

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

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

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

for a positive parameter λ equaling the width of the window where the FFT is performed. In Figure 1(b) we show a succession of these Blackman windows. Further background on why we use Blackman windows can be found in [20].

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. In this paper we shall provide many additional examples illustrating their efficacy.

Remark 1. To see how spectrograms display the frequency content of recorded sound, it helps to write the FFT $\mathcal{F}$ in (1) in a more explicit form. The FFT that we use is given, for even N, by the following mapping of a sequence of real numbers $\{a_m\}_{m=-N/2}^{m=N/2-1}$:

$$\{a_m\} \xrightarrow{\mathcal{F}} \left\{ A_\nu = \frac{1}{\sqrt{N}} \sum_{m=-N/2}^{N/2-1} a_m\, e^{-i2\pi m\nu/N} \right\}, \tag{2}$$

where ν is any integer. In applying $\mathcal{F}$ in (1), we make use of the fact that each Blackman window $w(t_k - \tau_\ell)$ is centered on $\tau_\ell = j\ell\Delta t$ and is 0 for t_k outside of its support, which runs from $t_k = (j\ell - N/2)\Delta t$ to $t_k = (j\ell + N/2)\Delta t$ and is 0 at $t_k = (j\ell \pm N/2)\Delta t$. So, for a given windowing specified by ℓ, the FFT $\mathcal{F}$ in (2) is applied to the vector $(a_m)_{m=-N/2}^{N/2-1}$ defined by $\left(f(t_k)w([k - j\ell]\Delta t)\right)_{k=j\ell-N/2}^{j\ell+N/2-1}$.

In (2), the variable ν corresponds to frequencies for the discrete complex exponentials $e^{-i2\pi m\nu/N}$ used in defining the FFT $\mathcal{F}$. For real-valued data, such as recorded sound, the FFT values A_ν satisfy the symmetry condition $A_{-\nu} = A_\nu^*$, where A_ν^* is the complex conjugate of A_ν. Hence, no significant information is gained with negative frequencies. Moreover, when the Gabor transform is displayed, the values of the Gabor transform are plotted as

squared magnitudes (we refer to such plots as *spectrograms*). There is perfect symmetry at ν and $-\nu$, so the negative frequency values are not displayed in the spectrograms.

Gabor Frames

When we discuss audio synthesis, it will be important to make use of the expression of Gabor transforms in terms of Gabor frames. An important seminal paper in this field is [92]. A comprehensive introduction can be found in [48]. We introduce Gabor frames here because they follow naturally from combining (1) and (2). If you prefer to see musical examples, then please skip ahead to the next section and return here when we discuss signal synthesis.

Making use of the description of the vector (a_m) given at the end of Remark 1, we can express (1) as

$$(3)\quad \frac{1}{\sqrt{N}} \sum_{k=k_0}^{k_1} f(k\Delta t)\, w([k-j\ell]\Delta t)\, e^{-i2\pi[k-j\ell]\nu/N}.$$

In (1), the values k_0 and k_1 are the lower and upper limits of k that are needed to extend the signal values beyond the endpoints of $[0,T]$. We have also made use of the fact that $w([k-j\ell]\Delta t) = 0$ for $k \le j\ell - N/2$ and for $k \ge j\ell + N/2$.

We now define the functions $\mathbf{G}_{\ell,\nu}(k)$ for $\nu = -N/2,\ldots,N/2-1$ and $\ell = 0,\ldots,M$ as

$$(4)\quad \mathbf{G}_{\ell,\nu}(k) = \frac{1}{\sqrt{N}}\, w([k-j\ell]\Delta t)\, e^{-i2\pi[k-j\ell]\nu/N}$$

and write $C_{\ell,\nu}$ for the Gabor transform values that are computed by performing the computation in (3):

$$(5)\quad C_{\ell,\nu} = \sum_{k=k_0}^{k_1} f(t_k)\, \mathbf{G}_{\ell,\nu}(k).$$

We have suppressed the dependence on j since it is a fixed value, specifying the amount of shifting for each successive window (via $\tau_\ell = t_{j\ell}$). It is set in advance and does not change during the analysis procedure using $\{\mathbf{G}_{\ell,\nu}\}$.

The significance of Equation (5) is that each Gabor transform value $C_{\ell,\nu}$ is expressed as an inner product with a vector $(\mathbf{G}_{\ell,\nu}(k))$. Hence, the entire arsenal of linear algebra can be brought to bear on the problems of analyzing and synthesizing with Gabor transforms. For instance, the vectors $\{\mathbf{G}_{\ell,\nu}\}$ are a discrete *frame.* The theory of frames is a well-established part of function space theory, beginning with the work of Duffin and Schaeffer on nonharmonic Fourier series [45, 99] through applications in wavelet theory [33, 18, 58] as well as Gabor analysis [50, 51, 55, 24].

To relate our discrete Gabor frame to this body of standard frame theory, and to provide an elementary condition for a discrete Gabor frame, we require that the windows satisfy

$$(6)\quad A \le \sum_{\ell=0}^{M} w^2(t_k - \tau_\ell) \le B$$

for two positive constants A and B (the *frame constants*). The constants A and B ensure *numerical stability,* including preventing overflow during analysis and synthesis. The inequalities in (6) obviously hold for our Blackman windows when they are overlapping as shown in Figure 1(b). Using (4) through (6), along with the Cauchy-Schwarz inequality, we obtain (for $K := k_1 - k_0 + 1$):

$$(7)\quad \sum_{\ell=0,\nu=-N/2}^{M,N/2-1} |C_{\ell,\nu}|^2 \le (KB)\ \|f\|^2,$$

where $\|f\|$ is the standard Euclidean norm: $\|f\|^2 = \sum_{k=k_0}^{k_1} |f(t_k)|^2$. When we consider Gabor transform synthesis later in the paper, we shall also find that

$$(8)\quad (A/K)\ \|f\|^2 \le \sum_{\ell=0,\nu=-N/2}^{M,N/2-1} |C_{\ell,\nu}|^2.$$

So we have (using $B_1 := A/K$ and $B_2 := KB$):

$$(9)\quad B_1\|f\|^2 \le \sum_{\ell=0,\nu=-N/2}^{M,N/2-1} |C_{\ell,\nu}|^2 \le B_2\|f\|^2,$$

a discrete version of standard *analysis operator* bounds in function space theory. By *analysis operator,* we mean the Gabor transform $\mathcal{G}$ defined by $\{f(t_k)\} \xrightarrow{\mathcal{G}} \{C_{\ell,\nu}\}$.

Remark 2. (1) It is not necessary to use a single fixed window w to perform Gabor analysis. For example, one could use windows $w_\ell(t_k - \tau_\ell)$ for each ℓ, provided $A \le \sum_{\ell=0}^{M} w_\ell^2(t_k - \tau_\ell) \le B$ is satisfied. Doing so allows for more flexibility in handling rapid events like drum strikes [100], [37, Chap. 3], [39]. (2) The time values $\{t_k\}$ do not need to be evenly spaced (this is called *nonuniform sampling*) [50, 51]. An application using nonuniform sampling is described in Example 12.

Musical Examples of Gabor Analysis

We now discuss a number of examples of using spectrograms to analyze recorded music. Our goal is to show how spectrograms can be used to provide another dimension, a quantitative dimension, to understanding the artistry of some of the great performances in music. For each of the displayed spectrograms, you can view a video of the spectrogram being traced out as the music plays by going to this webpage:

(10) `http://www.uwec.edu/walkerjs/MBSGC/`

Viewing these videos is a real aid to understanding how the spectrograms capture important features of the music. The website also provides an online bibliography and links to the software

we used. Another site with advanced software for time-frequency analysis is [69].

As we describe these examples, we shall briefly indicate how they relate to the art of music and its creation. While spectrograms are extremely useful for quantitative analysis of performance—and performance itself is an act of artistic creation—we shall also point out some ways in which spectrograms can be of use in aiding the creative process of musical composition. When we finish discussing these examples, we will summarize all of these observations on musical creativity.

Example 1 (Pavarotti's vocals)**.** Our first example is a spectrogram analysis of a 1990 recording of the conclusion of Puccini's aria, "Nessun Dorma", featuring a solo by Luciano Pavarotti [17]. See Figure 2. This spectrogram shows the high-amplitude vibrato (oscillations of pitch) that Pavarotti achieves. Using the spectrogram, we can measure quantitatively the amplitudes in the vibrato. This illustrates an interesting *creative application* of spectrograms. Because they can be displayed in real time as the singer is performing, they can be used to help performers to analyze and *improve* their vibrato, both in amplitude and steadiness. One affordable package for real-time spectrogram display can be found at [1]. Notice also that Pavarotti is able to alter the formant structure (the selective amplification of different frequency bands) of the same word in the lyrics, which produces a clearly audible change in brightness (timbre) in the sound. Using spectrograms to study formants is usually the domain of linguistics [75]. But here we see that formants—through a real-time adaptation of selective frequency amplification (selective resonance)—play a role in the magnificent instrumentation that is possible with the human voice.[1]

Example 2 (Contrasting choral and solo voicings)**.** Our second example is another passage from the same recording of "Nessun Dorma" [17], containing a choral passage and the introduction of Pavarotti's concluding solo. See Figure 3. We can see in the spectrogram that there is a notable contrast between the sinuous, blurred vibrato of the chorus in the beginning of the passage versus the clearer vibrato of Pavarotti's solo voice, which is centered on elongated, single pitches. The blurring of the vibrato of the chorus is due to *reverberation* (persistent echoing). We can describe this with the following model:

$$f(t_k) = \sum_{j=0}^{J} g([k - jm]\Delta t)h(j), \tag{11}$$

[1] *There is some speculation that the human vocal apparatus actually evolved first in order to sing, rather than speak* [73].

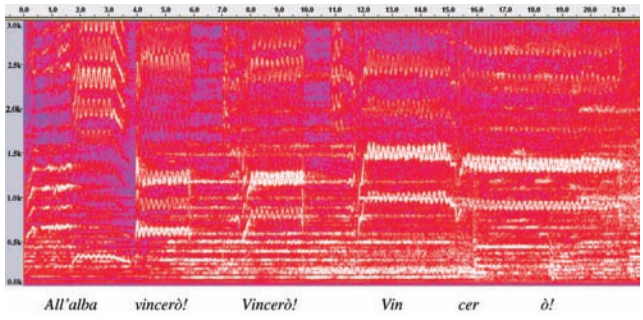


Figure 2. Spectrogram from a recording of "Nessun Dorma" with Luciano Pavarotti. Pavarotti's large-amplitude vibrato is clearly visible *and measurable.* The differing *formants* (selective amplification of different frequency bands) for different vocalizations of the lyrics (printed below the spectrogram) are also clearly displayed. Pavarotti changes the formants for the first two vocalizations of *vincerò,* and this is clearly audible as a change of brightness (timbre) of the sound. In addition, Pavarotti holds the final note, sung as an extended *ò,* with large, constant amplitude vibrato for more than 5 seconds, an amazing feat. (A video for this figure is at (10), a larger graphic is at [101].)

where m is a positive integer, g is the sound that is reverberating, and h is a damping function. The superposition of the slightly damped time-shifted versions of g creates the blurring effect, due to the closeness together in time of almost identical, shifted versions of g. Equation (11) is a discrete *convolution.* Audio engineers frequently employ a model like this, called *convolution reverb,* to simulate the effects of reverberation [44], [7, Sec. 16.7.2]. The function h, called the *impulse response,* is created once through digitally recording the reverberation of a very sharp sound (an impulse). Assuming linearity and shift-invariance properties of reverberation, the reverberation of other sounds is simulated digitally via Equation (11). We shall give an example of using convolution reverb for producing a new musical composition, "Sierpinski Round", in Example 13.

In contrast to the reverberation in the chorus, Pavarotti's vocals—which emerge at high volume from the midst of the chorus, an emotionally moving experience—are much more clearly defined in their vibrato, and his vibrato oscillates around constant pitches. The lack of reverberation in Pavarotti's vocals is probably due to a difference in the way his voice was recorded. The recording was done live at the ruins of the Baths of Caracalla in Rome. Pavarotti sang into a single microphone, while the multiple voices of the chorus were recorded by another microphone (or small set of microphones relative to the large number of

chorus members). Consequently, the recording of the chorus picks up the reverberation off of the walls of the ruins, while Pavarotti's single microphone records his initial voicing without much reverberation. The resounding emergence of Pavarotti's voice from the midst of the chorus is no doubt an artistic choice of Puccini to create a dramatic contrast between the individual and the collective. The blurring of the reverberation of the choral voices versus the clarity of the solo voice serves to further enhance the contrast. This enhanced contrast may also be a creative decision, and with methods such as convolution reverb it can be deliberately produced in creating digital recordings.

Example 3 (Does Louis Armstrong's voice sound like his trumpet?)**.** The classical pianist Edna Stern has answered this question very nicely, with her succinct description of the creative artistry of Louis Armstrong:

> The way he plays trumpet is very similar to his singing, it comes through in the way he is vibrating or marking the expressive moments in the music. Also the timing and the way he is building his phrases are both done in the same declamatory way as his singing. When he is singing, his vibratos are very similar to the trumpet vibratos and I think that there is a clarity in the way he punctuates his syllables, using the exploding ones in a sort of trumpet way. [89]

In Figure 4 we show two spectrograms of clips taken from Armstrong's 1950 recording of "La Vie en Rose" [64] that illustrate perfectly, in a quantitative way, what Stern is talking about. They contain a trumpet solo and vocals of Armstrong. All the vocals exhibit vibrato. The fundamentals for the trumpet notes are in a higher frequency range (around 500 Hz) than the vocals (around 200 Hz). The most prominent trumpet notes exhibiting vibrato are between 0.2 and 1.4 seconds, between 10.5 and 12.2 seconds, and between 14.0 and 15.3 seconds. It is interesting how Armstrong increases the amplitude of the vibrato as these notes progress. This is most evident for the three notes just cited. There is an interesting contrast between these trumpet notes compared with the constant frequency notes of other instruments in the passage (the horizontal bars in the spectrogram that correspond to bass notes at the lowest frequencies, as well as guitar and piano notes at higher frequencies). During the recording, Armstrong also plays notes with constant amplitude vibrato. For example, at the end of the recording, he sustains a note with constant amplitude

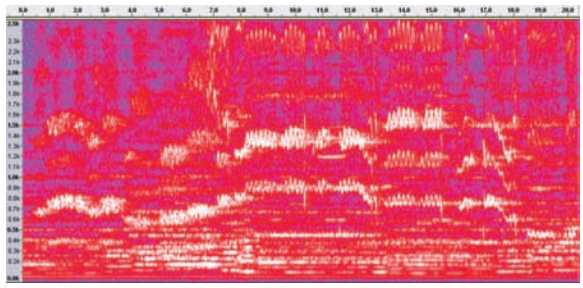


Figure 3. Spectrogram from a recording of "Nessun Dorma" with Luciano Pavarotti. In the first 7.5 seconds, there is a chorus of voices singing. Their lyrics produce several sinuous blurry curves, corresponding to fundamentals and overtones of male and female voices. This chorus is singing with vibrato, but the vibrato is obscured by the blurring (a convolution of reverberating sound). In contrast, Pavarotti's solo emerges from the chorus at about 7.0 seconds. His voicings are with much clearer vibrato and are centered on single, more elongated pitches. (A video for this figure is at (10), a larger graphic is at [101].)

vibrato for about six seconds. The corresponding spectrogram image is similar to the final note for *ò!* held by Pavarotti in Figure 2, so we do not display it.

The exploding syllables that Stern refers to are evident in the very brief, sharply sloped structures that initiate many of the vocals. A particularly prominent one occurs in the vocals spectrogram in Figure 4 at about 14.0 seconds. A similar explosive onset of a trumpet note occurs at about 8.8 seconds in the spectrogram containing the trumpet solo.

Louis Armstrong is well known for both his trumpet playing and his unique vocal style. Here we have shown, in a quantitative way, that it is pointless to try to separate these two aspects of his performance.

Example 4 (Dissonance in rock music)**.** Our next example is a recording of one of the most famous passages in rock music, the introduction to the 1970 recording of "Layla" by Derek and the Dominoes [63]. See Figure 5. The passage begins with the two guitars of Duane Allman and Eric Clapton playing in perfect synchronicity. At around 1.7 seconds, however, a buzzing distortion of the sound occurs. This distortion is due to a beating effect between closely spaced overtones. Although this dissonance may offend some ears, it is most certainly a conscious effect done by the musicians as a prelude to the intense pleading, with a very rough textured voice, of the singer later in the song. It is interesting to contrast this intense dissonance with the more subtle one invoked in "Gymnopédie I" (see Figure 15). When we discuss

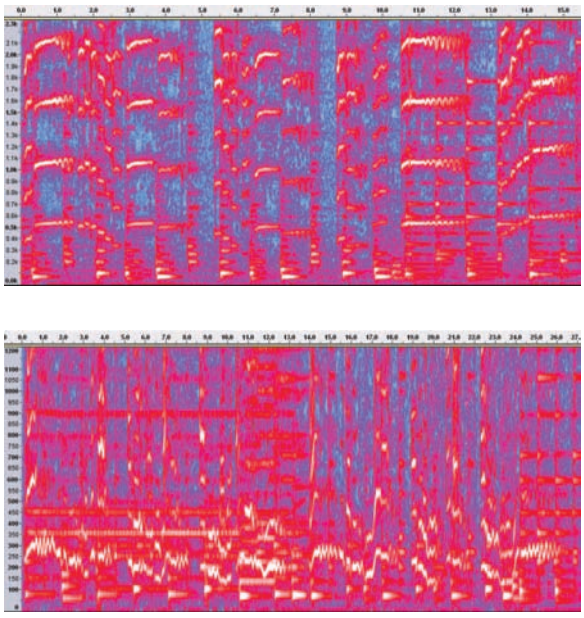


Figure 4. Top: Spectrogram of a trumpet solo by Louis Armstrong. Bottom: Spectrogram of Louis Armstrong vocals. (A video for this figure is at (10), a larger graphic is at [101].)

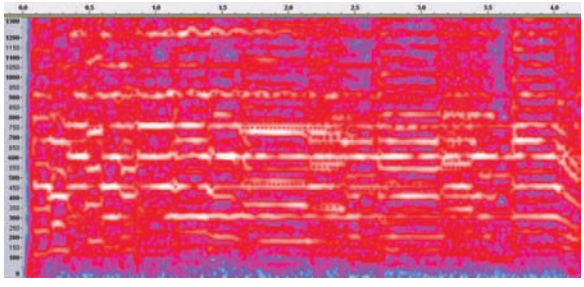


Figure 5. Spectrogram from a recording of "Layla" with Eric Clapton and Duane Allman on guitars. An important feature of the spectrogram is the clear indication of beating between different overtones. For example, for time between 1.7 and 2.2 seconds and frequencies around 450 Hz and 750 Hz, and also for time between 2.2 and 2.4 seconds and frequencies around 300 Hz, 550 Hz, and 700 Hz. The beating appears as sequences of dark points lying between horizontal overtone bands. It is clearly audible as a buzzing sound. (A video for this figure is at (10), a larger graphic is at [101].)

"Gymnopédie I", we will show how this contrast can be characterized quantitatively and how that could be of use in musical composition.

Example 5 (The geometry of *Appalachian Spring*)**.** The use of finite groups in analyzing the structure of notes and chords in musical scores is well established [56, 47, 15]. The work now extends even to the use of highly sophisticated methods from algebraic geometry [70] and orbifold topology [95]. Computer methods are being used as well [21, 12, 7].

One advantage that spectrograms have over analysis of scores, however, is that they provide a way for us to identify these group operations being performed over an extended period of time in a complex musical piece. In such cases, score analysis would be daunting, at least for those without extensive musical training. Spectrograms can help those without score-reading experience to see patterns, and it can help those who do have score-reading experience to connect these group operations to the patterns that they see in the score. Either way, it is valuable. Certainly a solid understanding of these group operations, and how master composers have used them, is an important aspect of learning to create new musical compositions.

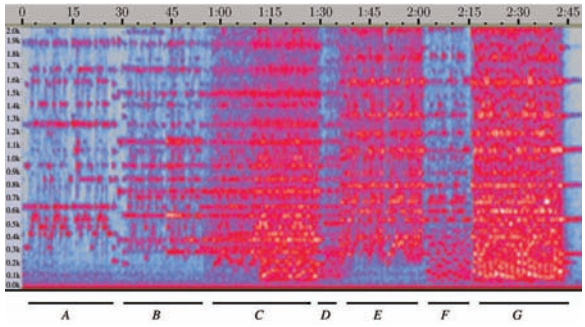


Figure 6. Spectrogram of a passage from *Appalachian Spring* by Aaron Copland. In the time intervals marked by A through G, there is a complex sequence of transpositions and time dilations, played by various instruments, of the basic melody introduced in A. (A video for this figure is at (10), a larger graphic is at [101].)

As an example of analyzing a spectrogram without score references, we look at Section 7 from an Atlanta Symphony recording of Aaron Copland's symphonic suite, *Appalachian Spring* [5]. See Figure 6. Copland uses group-theoretic operations in the time-frequency plane—transposition and dilation—in order to develop the basic melody of the Shaker hymn, "Simple Gifts".

In the spectrogram, within the time interval A and the frequency range of 300 to 600 Hz, there is the initial statement of the melody, played by woodwinds. There is also a transposition and dilation (time shortening) of part of the melody to a much higher frequency range (time range 0:15 to 0:17 and frequency range 900 Hz to 1200

Hz)—a sort of grace note effect, but using an entire melodic motive. In the time interval *B* and the frequency range of 200 to 500 Hz, there is a transposition down to stringed instruments, with a time dilation (shrinkage) of the melody, plus some interpolated notes so that the melody still plays for about 30 seconds. In the time interval *C*, there is a fugue-like development of the melody. It begins with another transposition from *A*'s melody to a lower frequency range than in *B*. Then at about time 1:12 there is an overlay of a transposition of *A*'s melody to a higher frequency range played by horns, and underlying it a very slow-moving time dilation (stretching) of a transposition down to a lower frequency range played by bass strings. In the time interval *D* there is a transition passage, played by woodwinds, of short motives pulled from the melody. The beginning of time interval *E*, from about time 1:37 to time 1:40, contains a rising horn crescendo that leads into a transposition of the melody from *A* to a higher frequency range (also played by horns). Within the time interval *F*, there is a transposition down again, played by woodwinds. Finally, the passage concludes with the time interval *G*, containing a combination of all the previous transpositions: played by strings, horns, and woodwinds at high volume, emphasized by a stately rhythmic pounding of a bass drum.[2] This passage from *Appalachian Spring* is a remarkable example of an extended development of a melodic theme.

Example 6 (What do Ludwig van Beethoven, Benny Goodman, and Jimi Hendrix have in common?)**.** The short answer, of course, is that they all created great music. In Figure 7 we show three spectrograms of recordings of short passages from their music. It is interesting to find the similarities and differences between these spectrograms and the music they reflect. In the Beethoven passage—which is a portion of his Piano Sonata in E (Opus 109) performed by David Añez Garcia [66, Movement 1, Measures 15–17]—we see a descending series of treble notes followed by an ascending series, reflecting an approximate mirror symmetry. The symmetry is broken, however, by an ascending set of bass notes (indicated by the arrow on the spectrogram). The mirror-symmetric pattern is also broken by the gentle rising treble notes trailing off to the right.

The spectrogram of the Goodman recording—a clip from a circa 1943 live broadcast of "Sing, Sing, Sing" [14][3]—also shows a similar, approximate mirror symmetry from time 3.5 seconds to 8.5 seconds, which is also broken by a gently rising scale trailing off to the right. In this case, Goodman is playing a clarinet and *bending the notes* as is commonly done in jazz style. The bending of the notes is clearly displayed in the spectrogram by the connected curved structures (which possess a symmetry of their own). This is a significant contrast to the discretely separated, constant harmonic piano notes in the Beethoven passage.

The spectrogram of the Hendrix passage—a clip from his 1968 recording of "All Along the Watchtower" [46]—exhibits a similar pattern to the other two, an approximately mirror-symmetrical descension and ascension of pitch, followed by a gently rising trailing off of melodic contour. Hendrix, however, illustrates a unique aspect of his music. Rather than using discrete notes, he instead uses his electric guitar to generate a *continuous flow of chords.* The chord progression is continuous rather than a set of discrete tones typically used in most Western music. It is interesting that the electronic sound produced here is surprisingly warm and soft, especially in the later "trailing off" portion. Perhaps this is due to Hendrix's synthesizing continuous chord transitions and vibrato ("wah-wah") within the blues scale, a remarkable achievement.

A more extended spectrogram analysis of the Beethoven piano sonata can be found in [27, pp. 49–56]. For more on jazz *vis-à-vis* classical music, see [16, pp. 106–132]. For incisive comments on Hendrix in particular, see [2, pp. 197 and 203].

Example 7 (Birdsong as a creative spark for music)**.** Birdsong has a long and distinguished history of providing a creative spark for musical composition. Donald Kroodsma, perhaps the world's leading authority on birdsong, has succinctly described this history:

> As early as 1240, the cuckoo's *cuckoo* song appears in human music, in minor thirds. The songs of skylarks, song thrushes, and nightingales debut in the early 1400s. The French composer Olivier Messiaen is my hero, as he championed birdsongs in his pieces. I love Respighi's *Pines of Rome,* too, as the songs of a nightingale accompany the orchestra…I learned recently, too, why I have always especially enjoyed the "Spring" concerto from

[2] *For those listeners familiar with the religious background of the Shaker hymn, the metaphor—external to the music—of the inheritance of the earth is unmistakable in this concluding portion.*

[3] *We hope our discussion of this piece will help to alleviate its undeserved obscurity. At the time of writing,* [14] *is still in print.*

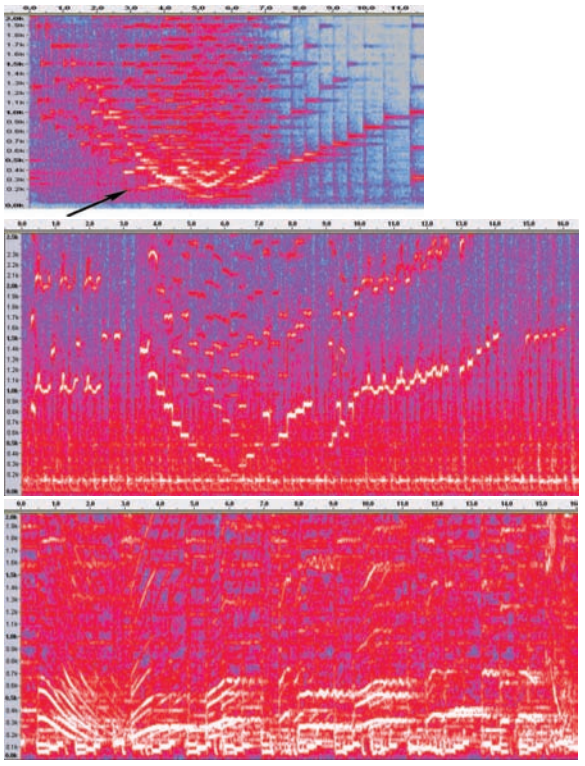


Figure 7. Top: Spectrogram from a recording of Beethoven's Piano Sonata in E (Opus 109). The arrow indicates an ascending bass scale, entering in contrast to the descending treble scale. Middle: Spectrogram from a Benny Goodman recording of "Sing, Sing, Sing". Bottom: Spectrogram from a Jimi Hendrix recording of "All Along the Watchtower". (A video for this figure is at (10), a larger graphic is at [101].)

> Vivaldi's *Four Seasons:* It is the birds and their songs that celebrate the return of spring in this concerto, and now I clearly hear their birdsongs in the brief, virtuoso flourishes by the solo violinist. [61, p. 275]

Kroodsma's book, from which this quote is taken, is full of spectrograms and their use in analyzing the musical qualities of birdsong. There is even a most amusing description on p. 274 of his researching which composers Java sparrows prefer and carrying out musical education with them!

Unfortunately, we do not have space here to do justice to the music of Messiaen. However, there is an excellent discussion, utilizing both spectrograms and scores, of the creative inspiration of birdsong for his music in Rothenberg's book [80]. The website for that book [98] also contains some samples of Messiaen's music and related birdsongs. One technique that Messiaen employed was to slow down the tempo of birdsong to increase its comprehensibility to human ears. He did this based on his own meticulous transcriptions into musical scores of songs he heard in the field. With Gabor transforms, this process can be done digitally with the recorded songs themselves; we will discuss how in Example 12. There is certainly much more to explore in Messiaen's music using the tools of spectrograms. The tools of percussion scalograms for rhythm analysis (which we discuss later) can be employed as well, since Messiaen incorporated the rhythmic styles of birdsong in his music [80, p. 198].

Besides Messiaen, Stravinsky also incorporated elements of birdsong into some of his compositions. In his classic work of musical theory, Mache devotes an entire chapter to studying, via spectrograms, the connections between the music of birds and elements of Stravinsky's music [67, Chap. 5]. Mache uses spectrograms as a kind of *generalized musical notation,* which gets around the very challenging problem of transcribing birdsong into standard musical notation. Mache also discusses other animal music-making in this chapter, which he entitles "Zoomusicology".

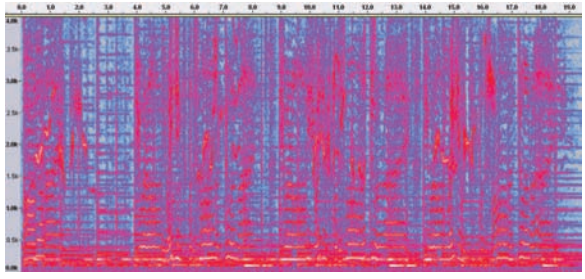


Figure 8. Spectrogram of Paul McCartney's duet with a songbird, from the Beatles' recording "Blackbird". (A video for this figure is at (10), a larger graphic is at [101].)

Besides the world of classical music, or art music, birdsong has been used in more popular music as well. A striking example is from the Beatles, who bridged the worlds of popular song and artistic music. At the end of their 1968 recording of "Blackbird" [11], Paul McCartney literally sings a duet with a blackbird. In Figure 8 we show a spectrogram of this portion of the recording. The blackbird's chirps lie in the upper portion of the time-frequency plane with sharply sloped rapid changes of pitch, while McCartney's vocals lie more in the lower portion (although his overtones often extend into the region where the blackbird is singing). In some places, for instance between 10.0 and 10.5 seconds, McCartney's rapid articulation is quite similar to the blackbird's, as we can see in the similar sharply sloped structures—for McCartney, most prominently between 250 and

1000 Hz, and for the blackbird, most prominently between 1500 and 3000 Hz—whereas in other places, there is a contrast between the longer time-scale rhythm to McCartney's vocals and the shorter time-scale rhythms of the blackbird's chirps.

Example 8 (Do our ears play tricks on us?)**.** In fact, our ears can fool us sometimes. The most famous auditory illusion is due to Shepard [85]. Shepard created electronic tones that seem to rise endlessly in pitch, even though they in fact ascend through just one octave. The tones can also be arranged to create an illusion of endless descent. The website [52] has a nice demonstration of Shepard tones. To hear examples showing that an illusion is actually occurring, go to the URL in (10) and select the topic Example 8: Audio Illusions.

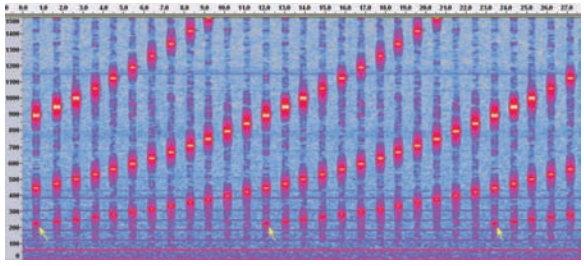


Figure 9. Spectrogram of Shepard's illusion of an endlessly rising series of tones, which in fact stay within just one octave. The arrows indicate where the illusion occurs. (A video for this figure is at (10), a larger graphic is at [101].)

Shepard's illusion is easily explained using spectrograms. Figure 9 practically speaks for itself. The illusion is due to our brain tracking the pitch contour of the fundamentals for the tones and expecting the next fundamental to be exactly one octave higher than where it occurs. The overtones of the electronic tones are all designed in an octave series—the fundamental multiplied by two, four, eight, etc.—to hide the drop in pitch.

Shepard's auditory illusion has been used subsequently by composers in electronic music, the most notable compositions being produced by Risset [7, Chap. 13]. The endlessly rising illusion can be used for expressing moments of infinite, transcendent joy, while the endlessly descending illusion can be used for invoking the opposite emotion.

The special arrangement of only octave spacings in the overtones does not lend itself well to traditional instruments. It is natural to ask, however, whether composers in the past have used illusions like the Shepard tones. Some examples by Bach have been proposed [60, p. 719]. Another

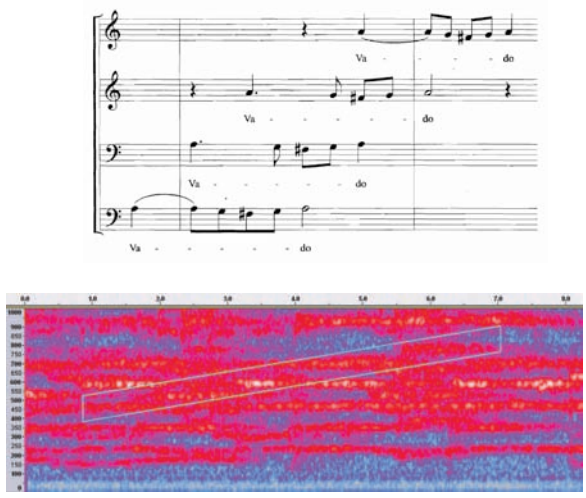


Figure 10. Top: Extract from the score of William Byrd's motet, "Non vos relinquam orphanos". Bottom: Spectrogram of a recording of the passage, with its structures aligned with the vocals in the score above. An ascending series of overtones lies within the parallelogram. (A video for this figure is at (10), a larger graphic is at [101].)

example is pointed out by Hodges in his delightful article on geometry and music [47, Chap. 6]. The illusion occurs in performances of a Renaissance motet by William Byrd, "Non vos relinquam orphanos" ("I will not leave you comfortless"). To describe the illusion, we quote from Hodges's article:

> Jesus is foretelling his ascension into heaven, *Vado* 'I am going'.
>
> The moment passes quickly …The *Vado* motif seems to move steadily upward through the voices, pointing to Jesus' own movement upwards to heaven. …In fact, the movement is not as steady as it sounds; at two of the repetitions there is no movement upwards. …the ear is deceived.

In Figure 10, we show the score and a spectrogram that illustrates the overtones of the voices at the four vocalizations of *Vado* in a 2003 Cambridge Singers recording of the motet [8]. There is clearly a similarity to the Shepard tone arrangement of overtones. As Hodges points out, the effect is subtle. Undoubtedly it is more noticeable when listening to the motet in a locale with long reverberation times contributing to the frequency tracking process in our brains (such as the Great Hall of University College School, London, where this recording was made). As a partial confirmation of this idea, we will amplify the indicated overtones in the

motet's spectrogram in Example 10. This is an example of audio synthesis with Gabor transforms, which we discuss in the next section.

Summary. The musical examples we have discussed were chosen to illustrate that spectrograms enhance our understanding of how great performing artists produce their music and to provide ideas for using Gabor analysis as an aid to creating new music. Our discussion of these examples produced at least these five ideas:

(1) *Perfecting vibrato.* Using spectrograms, either recorded or in real time, to analyze and perfect vibrato. This applies to vibrato in singing and also to vibrato in instrumental music. More generally, this applies to improvements in other musical techniques, such as holding a constant pitch.
(2) *Convolution reverb.* Spectrograms can be used to compare the results, in a quantitative way, of recordings made with different convolution reverb methods.
(3) *Analyzing dissonance.* The beating effect of dissonance can be quantified easily with either spectrograms or scalograms (as we describe in Example 14). This allows for evaluation of creative efforts to employ dissonance in music.
(4) *Visualizing geometric transformations of the time-frequency plane.* Such transformations are regularly used in musical compositions. Spectrograms provide an effective tool for analyzing the patterns of such transformations over longer time scales than score analysis facilitates. Such analyses are valuable for understanding the connection between these patterns and the music they reflect.
(5) *Zoomusicology.* Spectrograms provide a generalized musical notation for capturing the production of sonorities by other species, such as birds. This can yield new ideas for pitch alterations and rhythmic alterations (analyzed with the percussion scalogram method). Also, slowing down the tempo of birdsong, which we discuss later in Example 12, is an important technique (first applied by Messiaen). With Gabor transforms, this slowing down can be performed automatically (without tedious and complicated field transcriptions to standard musical notation).

Discrete Gabor Transforms: Signal Synthesis

The signal $\{f(t_k)\}$ can be reconstructed from its Gabor transform $\{C_{\ell,\nu}\}$. We shall briefly describe this reconstruction and show how it can be expressed in terms of synthesis with a *dual frame,* a frame dual to the Gabor frame $\{\mathbf{G}_{\ell,\nu}\}$. Following this discussion, we apply this synthesis method to audio processing and the creation of new electronic music.

First, we briefly sketch the reconstruction process. The FFT in (2) is invertible, via the formula:

$$a_m = \frac{1}{\sqrt{N}} \sum_{\nu=-N/2}^{N/2-1} A_\nu \, e^{i2\pi m\nu/N}. \tag{12}$$

Hence, by applying such FFT-inverses to the Gabor transform in (1), we obtain a set of discrete signals:

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

For each ℓ, we then multiply the ℓth signal by $\{w(t_k-\tau_\ell)\}$ and sum over ℓ, obtaining

$$\{\sum_{\ell=0}^{M} f(t_k)w^2(t_k-\tau_\ell)\} = \{f(t_k)\sum_{\ell=0}^{M} w^2(t_k-\tau_\ell)\}.$$

Multiplying the right side of this last equation by the values

$$\left[\sum_{\ell=0}^{M} w^2(t_k-\tau_\ell)\right]^{-1},$$

which by (6) are no larger than A^{-1}, we obtain our original signal values $\{f(t_k)\}$.

Synthesis Using Dual Frames

We now describe one way in which this reconstruction can be expressed via a dual frame. Alternative ways are described in [23, 25, 57].

We apply inverse FFTs to the Gabor transform values $\{C_{\ell,\nu}\}$ in (5), then multiply by $\{w([k-j\ell]\Delta t)\}$, and sum over ℓ to obtain:

$$\sum_{\ell=0}^{M}\sum_{\nu=-N/2}^{N/2-1} C_{\ell,\nu}\,\frac{w([k-j\ell]\Delta t)}{\sqrt{N}}\,e^{i2\pi(k-j\ell)\nu/N}. \tag{13}$$

We then divide by $\sum_{m=0}^{M} w^2([k-jm]\Delta t)$, obtaining

$$\sum_{\ell=0}^{M}\sum_{\nu=-N/2}^{N/2-1} C_{\ell,\nu}\,\frac{w[(k-j\ell)\Delta t]\,e^{i2\pi(k-j\ell)\nu/N}}{\sqrt{N}\sum_{m=0}^{M} w^2([k-jm]\Delta t)} = f(t_k).$$

Now, define the *dual Gabor frame* $\{\Gamma_{\ell,\nu}\}$ by

$$\Gamma_{\ell,\nu}(k) = \frac{w([(k-j\ell]\Delta t)\,e^{i2\pi(k-j\ell)\nu/N}}{\sqrt{N}\sum_{m=0}^{M} w^2([k-jm]\Delta t)}. \tag{14}$$

We leave it as an exercise for the reader to show that $\{\Gamma_{\ell,\nu}\}$ is, in fact, a discrete frame. Using (14), we obtain

$$f(t_k) = \sum_{\ell=0}^{M}\sum_{\nu=-N/2}^{N/2-1} C_{\ell,\nu}\,\Gamma_{\ell,\nu}(k). \tag{15}$$

Equation (15) is our synthesis of $\{f(t_k)\}$ using the dual frame $\{\Gamma_{\ell,\nu}\}$.

We note that by combining (15) with (6) and using the Cauchy-Schwarz inequality, we obtain Inequality (8).

Remark 3. When there is a high degree of overlapping of windows, then $\sqrt{N}\sum_{m=0}^{M} w^2([k-jm]\Delta t)$ is approximately a constant C, and we have $\Gamma_{\ell,\nu}(k) \approx \mathbf{G}^*_{\ell,\nu}(k)/C$. Hence, the dual Gabor frame is, modulo a constant factor and complex conjugation, approximately the same as the Gabor frame. In any case, (15) shows that we can reconstruct the original signal f as a linear combination of the vectors in the dual Gabor frame, each of which is supported within a single window of the form $w([k - j\ell]\Delta t)$. These last two statements reveal why we multiplied by the windowings before summing in (13).

The synthesis on the right side of (15) can be expressed independently of the Gabor transform. We shall use S to denote the *synthesis mapping*

$$(16)\quad \{B_{\ell,\nu}\} \xrightarrow{S} \left\{\sigma(t_k) = \sum_{\ell=0}^{M}\sum_{\nu=-N/2}^{N/2-1} B_{\ell,\nu}\,\Gamma_{\ell,\nu}(k)\right\}$$

based on the right side of (15), but now applied to an arbitrary matrix $\{B_{\ell,\nu}\}$.

Musical Examples of Gabor Synthesis

We shall now illustrate Gabor synthesis with musical examples. There are two basic schemes. One scheme, which is used for creating new music, is to use the synthesis mapping S in (16). The input matrix $\{B_{\ell,\nu}\}$ is specified by the composer, and the output $\{\sigma(t_k)\}$ is the new, electronic audio. There is software available for synthesizing music in this way. A beautiful example is the MetaSynth program [71].

Another scheme, which is used for processing audio, can be diagrammed as follows (where P denotes some processing step):

$$(17)\quad \{f(t_k)\} \xrightarrow{G} \{C_{\ell,\nu}\} \xrightarrow{P} \{B_{\ell,\nu}\} \xrightarrow{S} \{\sigma(t_k)\}.$$

The end result $\{\sigma(t_k)\}$ is the processed audio.

Example 9 (Reversing figure and ground)**.** In [38], Dörfler mentions that one application of Gabor transforms would be to select a single instrument, say a horn, from a musical passage. We will illustrate this idea by amplifying the structure **H** shown in the spectrogram on the left of Figure 11, which is from a passage in a 1964 Boston Symphony Orchestra recording of Stravinsky's *Firebird Suite* [93]. The sound corresponding to this structure is a faint sequence of ascending harp notes.

To amplify just this portion of the sound, we multiply the Gabor transform values by a mask of quadratically increasing values from 3.7 to 4.1 within a narrow parallelogram containing **H** and value 1 outside the parallelogram[4] (see the right of Figure 11) and then perform the synthesis mapping S. Notice that the amplified structure **A** stands out more from the background of the remainder of the spectrogram (which is unaltered). Listening to the processed sound file, we hear a much greater emphasis on the harp notes than in the original, and the volume of the rest of the passage is unchanged. We have altered the figure-ground relationship of the music.

The modification we have performed in this example is a *joint time-frequency filtering* of the Gabor transform. With this example, we have touched on an important field of mathematical research known as *Gabor multipliers.* More details can be found in [9, 41, 49, 82, 97].

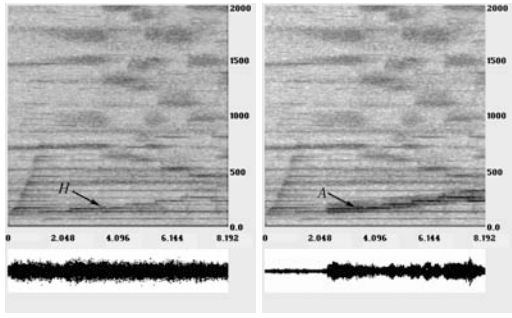


Figure 11. Left: spectrogram of portion of *Firebird Suite.* The structure to be selectively amplified is labeled H. Right: spectrogram with amplified structure labeled A. (A video for this figure is at (10), a larger graphic is at [101].)

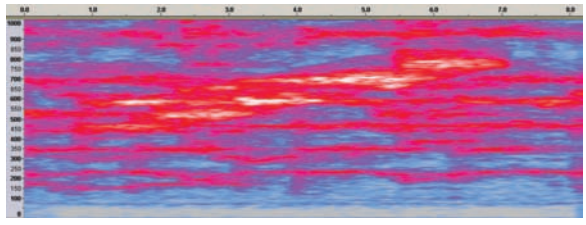


Figure 12. Selective amplification of a region in the spectrogram of a passage from a William Byrd motet, cf. Figure 10. (A video for this figure is at (10), a larger graphic is at [101].)

Example 10 (Amplifying Byrd's illusion)**.** A similar example of Gabor multipliers involves an amplification of a portion of the Gabor transform of the passage from the William Byrd motet considered in Example 8. We again amplify a parallelogram region of the Gabor transform (see Figure 10). This produces the spectrogram shown in Figure 12. After applying the synthesis mapping S to the modified transform, we are better able to hear the illusion. That is because we have selectively amplified the overtones of the vocals that our hearing is "tracking" when we hear the illusion.

Example 11 (Removing noise from audio)**.** One of the most important examples of audio processing with Gabor transforms is in audio denoising. For

[4] *The precise formula can be found at the website in* (10).

reasons of space, we cannot do justice to this vast field. Suffice it to mention that an excellent description of the flexibility of Gabor transform denoising, combining Gabor multipliers with sophisticated Bayesian methods, is given in Dörfler's thesis [37, Sec. 3.3]. An elementary introduction is given in [96, Sec. 5.10]. One advantage of Gabor transforms is that the compact support of the windows allows for *real-time denoising,* including adaptation to changing noise characteristics.

Example 12 (Slowing down sound)**.** In Example 7 we raised the issue of slowing down the tempo of birdsong. This could be done in two ways. The first way is to simply increase the size of the interval Δt between successive signal values when playing back the sound. This has the effect, however, of decreasing the frequencies of the pitches in the sound. For example, if Δt is doubled in size, then the frequencies of all the pitches are halved, and that lowers all the pitches by one octave. The website [98] has examples of this type of slowing down of birdsong.

Another approach to slowing down the birdsong, *while leaving the pitches unchanged,* is to make use of the Gabor transform $\{C_{\ell,\nu}\}$ of the sound. For example, to decrease the sound tempo by a factor of 2, we would first define a matrix $\{B_{\ell,\nu}\}$ as follows. We set $B_{2\ell,\nu} = C_{\ell,\nu}$ for each ℓ, and then interpolate between successive values $B_{2\ell,\nu}$ and $B_{2\ell+2,\nu}$ to define each value $B_{2\ell+1,\nu}$. Finally, we apply the synthesis mapping in (16) to the matrix $\{B_{\ell,\nu}\}$. The windows for this synthesis are all shifts of the original window w, centered at points spaced by the original increment $\Delta\tau$. Now, however, twice as many windowings are done (so as to extend twice as far in time). This method keeps the pitches at the same levels as the original sound, but they now last twice as long.

Slowing down by a factor of 2 is just the easiest method to describe. This latter approach can also be used to slow down (or speed up) the tempo of the sound signal by other factors. Both these methods can also be applied, either separately or in combination, to other digital sound files besides birdsong. For example, the MetaSynth program [71] allows for this sort of morphing of digital sound. It is interesting that the morphing we have described—which via Gabor frames can be done in a localized way on just parts of the signal—requires the extensions of the Gabor transform methodology described in Remark 2.

Example 13 (Granular synthesis of new music)**.** As mentioned above, composers can use the scheme in (16) to create new music. Some good examples can be found at the MetaSynth website [71]. However, we cannot resist showing two of our own compositions, which are really just motifs that still need to be set within larger compositions.

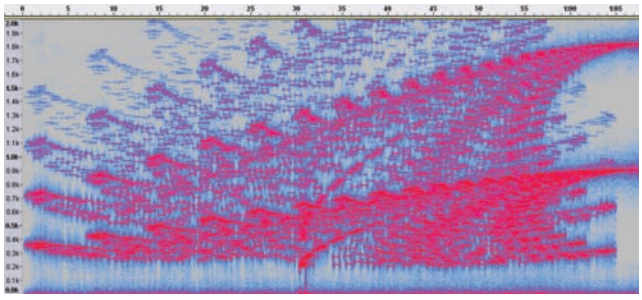


Figure 13. Spectrogram for the fractal music "Fern". (A video for this figure is at (10), a larger graphic is at [101].)

In Figure 13 we show the spectrogram of a granularly synthesized musical passage, "Fern" [35]. The method used for generating the grains $\{B_{\ell,\nu}\}$ was to use the iterated function system (IFS) created by Michael Barnsley for drawing a fractal image of a fern leaf [10, pp. 86–87]. Using the probabilities $p_1 = 0.01$, $p_2 = 0.85$, $p_3 = 0.07$, and $p_4 = 0.07$, one of the corresponding four affine transformations,

$$\mathcal{A}_i \begin{bmatrix} t \\ \nu \end{bmatrix} = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} t \\ \nu \end{bmatrix} + \begin{bmatrix} e_i \\ f_i \end{bmatrix}, \quad (i = 1, 2, 3, 4),$$

is applied to a pair of time-frequency coordinates (starting from an initial *seed* pair of coordinates). We shall not list all of the four affine transformations $\mathcal{A}_i$ here. They can be found in [10, Table 3.8.3 on p. 87]. It will suffice to list just one of them:

$$\mathcal{A}_3 \begin{bmatrix} t \\ \nu \end{bmatrix} = \begin{bmatrix} 0.2 & -0.26 \\ 0.23 & 0.22 \end{bmatrix} \begin{bmatrix} t \\ \nu \end{bmatrix} + \begin{bmatrix} 0 \\ 0.07 \end{bmatrix},$$

which is applied with probability $p_3 = 0.07$. Notice that $\mathcal{A}_3$ is not a member of the Euclidean group, since the matrix used for it is not orthogonal.

This IFS is used to draw points of the fern that begin the fundamentals of the notes. An instrumental spectrum generator was then used to generate overtones and time durations for the notes and to generate the actual sound file. Our iterations were done 2500 times using John Rahn's Common Lisp Kernel for music synthesis [78], which converts the generated fractal shape into a file of instrumental tone parameters determining initiation, duration, and pitch. This file of instrumental tone parameters was then converted into a digital audio file using the software C-Sound [30].

An interesting feature of the type of music shown in Figure 13 is that the notes contained in the piece do not exhibit the classical note symmetries (transpositions, reflections, etc.) that are all members of the Euclidean group for the time-frequency plane. That is because some of the affine transformations $\mathcal{A}_i$, such as $\mathcal{A}_3$, are not members of the Euclidean group. Perhaps these non-Euclidean operations on its notes are one source of the eerie quality of its sound.

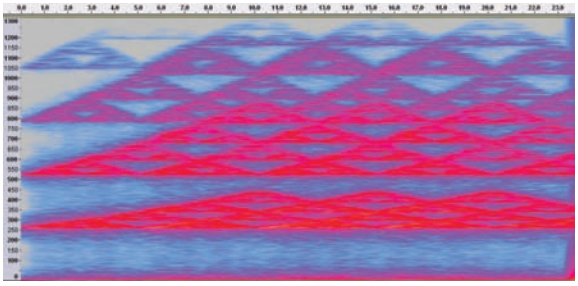


Figure 14. Spectrogram for the granular synthesis music, "Sierpinski Round". (A video for this figure is at (10), a larger graphic is at [101].)

As a second instance of granular synthesis, we applied Barnsley's IFS for the Sierpinski triangle [10, Table 3.8.1 on p. 86]. However, instead of generating only the grains for tones from Sierpinski's triangle, we also superimposed two shiftings in time of all such grains, thus producing three Sierpinski triangles superimposed on each other at different time positions. The spectrogram of the composition, "Sierpinski Round" [36], is shown in Figure 14. Besides having visual artistic value, it is interesting to listen to. It sounds like a ghostly chorus, with a bit of the Shepard rising-pitch-illusion effect.

We have only touched on the vast arena of granular synthesis, fractal music, and electronic music in general. More work is described in [7, 47, 74, 77, 28]. Although electronic music has been developing for about half a century, it is perhaps still in its infancy. Tonal music, based on harmony of one form or another, has been with us for millennia. In his masterful critical history of twentieth-century music, Alex Ross provides a balanced, nuanced view of the development of all forms of new music, including electronic music [79, Part III].

Summary. We have described several uses of Gabor transform synthesis for creating new music. These uses are (1) modifying the intensity of a specific instrument or set of overtones in a musical passage, (2) removing noise, (3) slowing down (or speeding up) the sound, (4) granular synthesis. Although this only scratches the surface of the use of Gabor transform synthesis, we did provide additional references in all of our examples.

Continuous Wavelet Transforms

In this section we briefly review the method of scalograms (continuous wavelet transforms) and then discuss the method of percussion scalograms. Both methods will be illustrated with musical examples.

Scalograms

The theory of continuous wavelet transforms (CWTs) is well established [31, 26, 68]. 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 detail in [4, Section 4]. We shall only briefly review it here.

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

$$\mathcal{W}_\Psi[f](\tau,s) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} f(t)\overline{\Psi(\frac{t-\tau}{s})}\,dt \tag{18}$$

for *scale* $s > 0$ and *time translation* τ. For the function Ψ in the integrand of (18), the variable s produces a dilation and the variable τ produces a translation.

We omit various technicalities concerning the types of functions Ψ that are suitable as wavelets; see [26, 31, 68]. In [26, 32], Equation (18) 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 (18). First, we assume that the sound signal $f(t)$ is nonzero only over the time interval $[0, T]$. Hence (18) becomes

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

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 τ, using $\tau_k = k\Delta t$. This yields

$$\mathcal{W}_\Psi[f](\tau_k,s) \approx \frac{T}{N}\frac{1}{\sqrt{s}}\sum_{m=0}^{N-1} f(t_m)\overline{\Psi([t_m-\tau_k]s^{-1})}. \tag{19}$$

The sum in (19) 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

$$(f:\Psi)_k = \sum_{m=0}^{N-1} f_m\overline{\Psi_{m-k}}\,. \tag{20}$$

[Note: For the sum in (20) to make sense, the sequence $\{\Psi_k\}$ is *periodically extended,* via $\Psi_{-k} := \Psi_{N-k}$.]

Thus Equations (19) and (20) 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

$$s = 2^{-r/J}, \quad r = 0, 1, 2, \ldots, I\cdot J, \tag{21}$$

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.[5] A *Gabor wavelet,* with width parameter ω and frequency parameter ν, is defined as follows:

$$\Psi(t) = \omega^{-1/2} e^{-\pi(t/\omega)^2} e^{i2\pi\nu t/\omega}. \tag{22}$$

Notice that the complex exponential $e^{i2\pi\nu t/\omega}$ has frequency ν/ω. We call ν/ω the *base frequency.* It corresponds to the largest scale $s = 1$. The bell-shaped factor $\omega^{-1/2} e^{-\pi(t/\omega)^2}$ in (22) damps down the oscillations of Ψ, so that their amplitude is significant only within a finite region centered at $t = 0$. (This point is discussed further, with graphical illustrations, in [4, 20].) Because the scale parameter s is used in a reciprocal fashion in Equation (18), 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 (18). Thus frequency is described in terms of the parameter $1/s$, which Equation (21) shows is logarithmically scaled. This point is carefully discussed in [4] and [96, Chap. 6], where Gabor scalograms are shown to provide a method of zooming in on selected regions of a spectrogram. Here we shall provide just one new example.

Example 14 (Subtle dissonance in "Gymnopédie I")**.** On the bottom of Figure 15, we show a spectrogram of the beginning of a 1983 Aldo Ciccolini performance of Erik Satie's "Gymnopédie I" [81]. There is a very subtle dissonance of some of the overtones of the notes in the piece. To see this dissonance we calculated a Gabor scalogram for the first 5.944 seconds of the passage, using 1 octave and 256 voices, a width parameter of 0.1, and frequency parameter 65. The base frequency is therefore 650 Hz, and the maximum frequency is 1300 Hz. This scalogram is shown on the top of Figure 15. The scalogram has zoomed in on the spectrogram enough that we can see several regions where overtones are so closely spaced that beating occurs. This subtle beating is barely evident to our ears, contributing to the haunting quality of the tones in the piece.

Besides Satie, other composers have used subtle dissonance effects in overtones. Debussy was a master of such techniques, including using different types of scales to enhance the interplay of the overtones. The musical history of Ross [79, pp. 43–49] contains a nice synopsis of Debussy's use of overtones, along with the pentatonic scale of much Oriental music (Chinese, Vietnamese, and Javanese gamelan music). The article [34] explores the overtone structures of Debussy in depth and also examines the question of whether

[5] *A CWT with Gabor wavelet is closely related to the Stockwell transform* [91, 54, 43, 90].

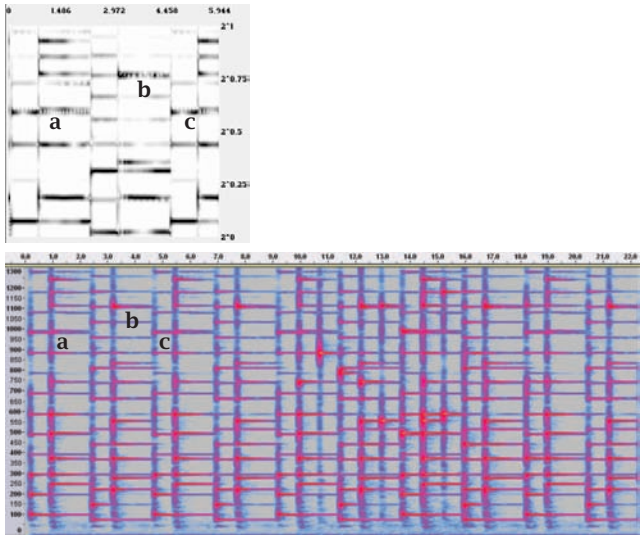


Figure 15. Bottom: Spectrogram of the beginning of "Gymnopédie I" by Erik Satie. Top: A zooming in on the first six seconds for frequencies between 650 Hz and 1300 Hz. There is beating at three places, marked a, b, c in both the scalogram and spectrogram. (A video for this figure is at (10), a larger graphic is at [101].)

the interplay of overtones would be enhanced if the system of *just intonation* were employed.

What Figures 15 and 5 illustrate is that we can use spectrograms and scalograms to produce a *quantitative measure* of the amount of beating in overtones. For example, *beating frequencies* can be easily read off from the spectrogram in Figure 5 and the scalogram in Figure 15. Furthermore, the intensity of the overtones that are "clashing" can be measured from the numerical data that those graphs are displaying. It is an interesting topic for future study: to measure these effects in the music of Debussy, for example, and compare the use of alternative systems of intonation.

The theory that beating in overtones creates a dissonant sound in music originated with Helmholtz [59], [47, Chap. 5]. Sethares's book [84] is an important reference relating Helmholtz's theory to a wide variety of different scales, including the pentatonic scale and Javanese gamelan music.

Percussion Scalograms

As described in [96, Chap. 6] and [20], scalograms can be used in conjunction with spectrograms to produce quantitative portraits of musical rhythm, called percussion scalograms. The theory behind percussion scalograms is described in detail in [20]. It is related to work of Smith [86, 87, 88]. Here we will only state the method. It consists of these two steps:

Step 1. Let $\{|C_{\ell,\nu}|^2\}$ be the spectrogram image. Calculate the average $\mu[|C|^2]$ over all frequencies at each time index ℓ:

$$\mu[|C|^2](\ell) = \frac{1}{N/2} \sum_{\nu=0}^{N/2-1} |C_{\ell,\nu}|^2, \tag{23}$$

and denote the time average of $\mu[|C|^2]$ by A:

$$A = \frac{1}{M+1} \sum_{\ell=0}^{M} \mu[|C|^2](\ell). \tag{24}$$

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

$$\mathcal{P}(\tau_\ell) = \mathbf{1}_{\{\tau_k : \mu[|C|^2](k) > A\}}(\tau_\ell), \tag{25}$$

where $\mathbf{1}_S$ is the indicator function for a set S ($\mathbf{1}_S(t) = 1$ when $t \in S$ and $\mathbf{1}_S(t) = 0$ when $t \notin S$). The values $\{\mathcal{P}(\tau_\ell)\}$ describe a pulse train whose intervals of 1-values mark off the position and duration of the percussive strikes.

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

When implementing this method, it is sometimes necessary to process the spectrogram by limiting its values to certain frequency bands (intervals of ν values), setting values of the spectrogram to 0 outside of such bands. We now show an example in which this leads to a precise analysis of the relationship between melody and rhythm. (A second example is discussed in [20, Example 6 on p. 355].)

Example 15 (Melody and rhythm in "Unsquare Dance")**.** In the 1961 Dave Brubeck Quartet's recording of "Unsquare Dance" [94], there is an amazing performance involving hand claps, piano notes, and bass notes all played in the unusual time signature of $\frac{7}{4}$. In Figure 16 we show our analysis of the melody and rhythm in a passage from "Unsquare Dance". We used three different frequency ranges from the spectrogram to isolate the different instruments from the passage. The passage 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 $\frac{7}{4}$ 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. In sum, the analysis shown in Figure 16 provides quantitative evidence for the "tightness" (rhythmic coherence) with which these musicians are performing.

Summary. We gave a couple of examples of the use of scalograms and percussion scalograms in quantitatively analyzing musical performance in the work of Satie and Brubeck. These examples show the value of these techniques for analyzing pitch, rhythm, and overtone structure. We indicated a topic of future research on the music of Debussy, which we hope will yield techniques that can be applied to composition in general.

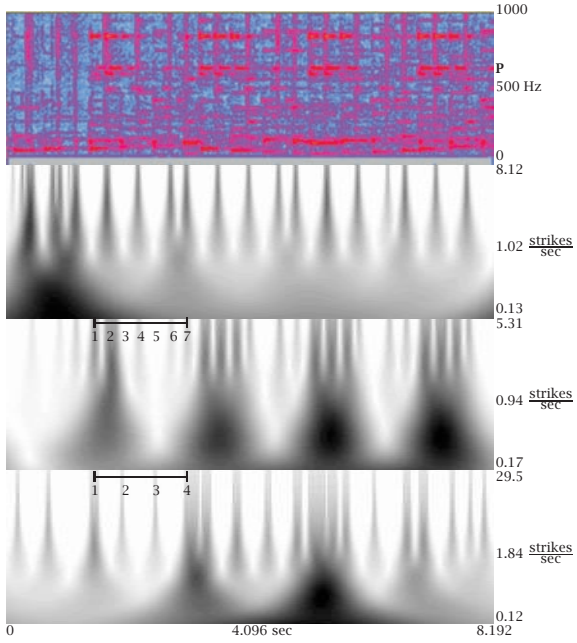


Figure 16. Melody and rhythm in a passage from "Unsquare Dance". *Top:* Spectrogram. P aligns with the piano notes. *Second from top:* Percussion scalogram of frequencies above 3000 Hz. Drumstick and hand clap percussion are emphasized. *Third from top:* Percussion scalogram of frequencies between 400 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 video for this figure is at (10), a larger graphic is at [101].)

Quantifying Rhythmic Complexity

Article [20] introduced sequences of rest lengths between notes (percussive strikes). See Figure 17. This was done because the rests between notes are at least as important as the notes themselves. As the classical pianist Artur Schnabel said: "The notes I handle no better than many pianists. But the pauses between the notes — ah, that is where the art resides."[6]

We will now attempt to quantify the complexity of a sequence of rests, using ideas from information theory, such as *entropy.* Ideas from

[6] *To hear Artur Schnabel perform, go to the URL in* (10) *and click on the link* Artur Schnabel: Beethoven's Moonlight Sonata.

information theory have been applied extensively to musical analysis. A major landmark in the field is the book by Meyer [72], which, although not providing explicit formulas and quantitative methods, nevertheless set the stage for future work. The field is now vast. One place to search for more material is the online bibliography of Downie [42], which, besides covering the important practical application of using information theory to retrieve musical works from databases, also includes many papers on the general topic of information theory and music. If there is anything new in our approach, it would be the use of percussion scalograms derived from spectrograms of recorded music of extemporaneous, or improvisational, performances (as opposed to score analysis). In any case, we felt this work to be sufficiently interesting to include here, even if it is only a brief introduction to the relations between information theory and music.

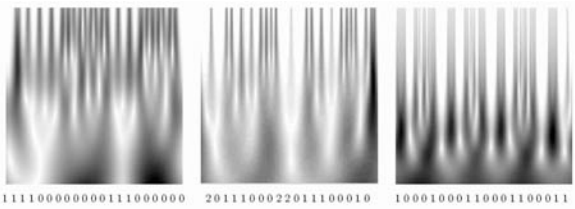


Figure 17. Percussion scalograms. Left: From "Dance Around". Middle: From "Welela". Right: From "Taxman" (bass notes only). Below each scalogram is the sequence of rests for the rhythmic sequence.

First, we review the basic notion of *entropy* from information theory [6, 29]. We assume that we have a stochastic source that emits sequences of symbols from a finite set $S = \{a_0, a_1, \ldots, a_n\}$ and p_k is the probability that the symbol a_k is emitted. The *entropy* E of the source is then measured in *bits* by the formula

$$\mathsf{E} = \sum_{k=0}^{n} p_k \log_2 \left(\frac{1}{p_k}\right). \tag{26}$$

The entropy E provides a greatest lower bound for the average number of bits per symbol that *any* uniquely decipherable encoding method could achieve in encoding the symbols from the source. In practice, a finite sequence of symbols would need to be encoded. By the law of large numbers, the *relative frequencies* of occurrence of the symbols $\{a_k\}$ will be converging to their probabilities, so the relative frequencies can be used in place of the theoretical probabilities $\{p_k\}$.

We will interpret an entropy value as indicating the magnitude of *complexity* of a finite sequence—the idea being that the higher the entropy, the more complex the sequence; hence the more bits it takes to encode the sequence. Using the term *complexity,* rather than *information,* allows us to avoid any connotations of *meaning* to our sequences of rests. Such connotations of *meaning* are generally irrelevant in information theory anyway—something being more complex has nothing to do with whether it is more or less meaningful.

We will now illustrate these ideas with a specific example of a rest sequence, the rest sequence from the "Dance Around" passage shown in Figure 17. The sequence is

(27) 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0.

For this sequence, we get relative frequencies of $p_0 = 12/14$ and $p_1 = 2/14$ for the rest symbols 0 and 1, respectively. Hence the entropy is $\mathsf{E} = 0.92$. We leave it as an exercise for the reader to check that for the rest sequence from the "Welela" passage shown in Figure 17, the entropy is $\mathsf{E} = 1.46$. These results seem to confirm our impression that the "Welela" sequence is more complex rhythmically than the "Dance Around" sequence.

However, there is more to it than that. In Table 1 we show our calculations of the entropy E for 10 different rhythmic passages from different genres of music. While the entries for E do seem to be consistent with the fact that African drumming is generally regarded as more rhythmically complex than rock drumming, we do notice that the value of E for "Dance Around" is fairly high. For instance, it is essentially the same as the "Taxman" sequence. But the "Taxman" sequence is

(28) 1 0 0 0 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1,

which seems to be more complex than the "Dance Around" sequence in (27). Moreover, using the entropy E would give exactly the same values of $\mathsf{E} = 1$ for the following two sequences:

(29) 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1

(30) 0 0 1 1 1 0 0 0 1 0 0 0 1 1 1 1

Table 1. Complexity Measures E and C for Rhythmic Sequences

Title*	*Genre*	E	C
Wipe Out	*Rock*	0.57	0.54
Dance Around	*Rock*	0.92	0.68
Toad	*Rock*	0.90	0.81
What'd I Say	*R & B*	1.22	0.97
Taxman	*Rock*	0.99	0.98
African Drumming 1	*African*	1.24	1.13
African Drumming 2	*African*	1.42	1.14
Welela	*African*	1.46	1.23
African Drumming 3	*African*	1.46	1.44
Sing, Sing, Sing	*Jazz*	1.52	1.47

*Audio files and discography for these sequences are at the URL in (10).

But it seems clear that the second sequence is more complex rhythmically than the first.

To solve this problem, we introduce a second notion of entropy. The entropy E is based on one type of modeling of a stochastic source: the *memoryless* or Markov-0 source model [29, 13]. We now consider another model, the Markov-1 source model. In the Markov-1 model, we assume that the probabilities of emitting any given symbol will depend on what the previous emitted symbol was. These probabilities will be the *transition probabilities* between *states,* a state being the value of a symbol. The Markov-1 entropy, which we will denote by C, is then calculated by an expected value over all states of the entropies for these transition probabilities. Rather than give a precise formula, it is probably easier to just give an example.

Consider the sequence (29). To do our calculation, we must assume an initial state. Because 0 is the most common rest value in rhythm, we will always assume that the initial state is 0 (this involves appending a dummy symbol 0 to the beginning of the sequence). We then get probabilities of $p_0 = 9/16$ and $p_1 = 7/16$ of being, respectively, in state 0 and state 1 *prior to emitting another symbol.* The transition probabilities are

State 0 ($p_0 = 9/16$)		State 1 ($p_1 = 7/16$)	
$0 \to 0$:	$\frac{8}{9}$	$1 \to 0$:	$\frac{0}{7}$
$0 \to 1$:	$\frac{1}{9}$	$1 \to 1$:	$\frac{7}{7}$

and we then calculate the Markov-1 entropy C by

$$\mathsf{C} = p_0\mathsf{E}_0 + p_1\mathsf{E}_1,$$

where E_0 and E_1 are calculated from the transition probabilities for state 0 and 1, respectively. This yields $\mathsf{C} = (9/16)(.503258) + (7/16)(\log_2 1) = 0.28$, whereas for the sequence (30), we obtain $\mathsf{C} = 0.89$. These Markov-1 entropy values are more consistent with the apparent complexity of the two sequences.

In Table 1 we show our calculations of the Markov-1 entropy C for the 10 different rhythmic passages. These results seem promising. The African drumming sequences are measured as the most complex. That includes the jazz sequence from "Sing, Sing, Sing", which imitates African rhythm—and is one of the most famous drum sequences in jazz history, perhaps because of its complexity. The rock drumming sequences are less complex using this measure.

Obviously we have worked with only a very small sample of music. We intend to continue assembling more data, but we do find the results of interest and hope that this discussion will spur some further work by others as well. Working on other measures of complexity, such as contextual entropy, is another research path. By *contextual entropy* we mean the calculation of entropy based on a non-Markov description of the source. Symbols will be produced with probabilities that depend on what *context* the symbol is in. One set of contexts would be the levels of hierarchy of the rhythm. As discussed in [20], there is a more complete notation for rhythm sequences that includes grouping the notes into hierarchies. The level of hierarchy a symbol belongs to might then be used as the context for that symbol. This idea has the advantage that it may be able to incorporate the fact that *production* of rhythm sequences is *not* a Markov process.[7]

We have done a couple of preliminary calculations with the "Dance Around" and "Welela" rhythmic hierarchies reported in [20, Examples 1 and 2]. We found that this new complexity measure, based on single (memoryless) frequencies of the note lengths as well as contextual (hierarchical) groupings, gives complexities of 1.12 for the "Dance Around" sequence and 1.58 for the "Welela" sequence. The two sequences are now closer in complexity value, and this may reflect the possibility that this new measure may be taking into account some of the speed variations in the "Dance Around" drumming (as discussed in the caption of Figure 5 in [20]). We mention these results only to pique the reader's interest. If our speculations are borne out, then we shall describe our findings in more detail in a subsequent paper.

Summary. We have described some complexity measures for musical rhythm and done an initial study of their value in quantifying the complexity in different rhythmic styles. We will continue working on the relation between entropy measures of complexity and musical rhythm, as that work has just begun.

Concluding Remarks

We have introduced some exciting aspects of the relations between mathematics and music. The methods of Gabor transforms and continuous wavelet transforms were shown to provide a wealth of quantitative methods for analyzing music and creating new music. Readers who wish to learn more can find ample resources in the References.

[7] *The reason is similar to what Chomsky describes for phrase structure in language* [22, p. 22]*: "in English, we can find a sequence* $a + S_1 + b$*, where there is a dependency between* a *and* b*, and we can select as* S_1 *another sequence containing* $c + S_2 + d$*, where there is dependency between* c *and* d*, then select as* S_2 *another sequence of this form, etc." The parallel between English phrases and musical phrases* (*with* $S_1, S_2, \ldots$ *as* bridges) *is obvious.*

Acknowledgements

This research was partially supported by the National Science Foundation Grant 144-2974/92974 (REU site SUREPAM program) and by a grant from the UW-Eau Claire Office of Research and Sponsored Programs. We thank William Sethares for his cogent review of an earlier draft of this paper. We are also grateful to Marie Taris for her perceptive editorial guidance throughout our preparation of this article. J.S.W., in particular, would like to thank Steven Krantz for all the encouragement and aid he has given him in pursuing the interdisciplinary study of mathematics and music.

References

[1] Acoustical Research Institute of the Austrian Academy of Sciences, STx software, available at http://www.kfs.oeaw.ac.at/content/section/16/392/lang,8859-1/.

[2] J. Adams, *Hallelujah Junction: Composing an American Life*, Farrar, Straus and Giroux, New York, 2008.

[3] J. B. Allen and L. R. Rabiner, A unified approach to short-time Fourier analysis and synthesis, *Proc. IEEE* **65** (1977), 1558–64.

[4] J. F. Alm and J. S. Walker, Time frequency analysis of musical instruments, *SIAM Review* **44** (2002), 457–76, available at http://www.uwec.edu/walkerjs/media/38228[1].pdf.

[5] *Appalachian Spring, Rodeo, Fanfare for the Common Man*, Atlanta Symphony Orchestra, cond. Louis Lane, Telarc 1982, Track 6: "Appalachian Spring (Suite)".

[6] R. B. Ash, *Information Theory*, Dover, New York, 1990.

[7] G. Assayag, H. Feichtinger, J.-F. Rodrigues (eds.), *Mathematics and Music: A Diderot Mathematical Forum*, Springer, New York, 2002.

[8] *Ave Verum Corpus—Motets and Anthems of William Byrd*, by Cambridge Singers, recorded at Great Hall of University College School, London, Collegium Records, 2003, Track 16: "Non Vos Relinquam".

[9] P. Balazs, Regular and Irregular Gabor Multiplier with Application to Psychoacoustic Masking, Ph.D. thesis, Univ. of Vienna, 2005.

[10] M. Barnsley, *Fractals Everywhere, Second Edition*, Academic Press, Cambridge, MA, 1993.

[11] *The Beatles* (The White Album), Capitol/EMI, 1968, Disc 1, Track 11: "Blackbird", Converted to mono by Audacity sound editor.

[12] J. W. Beauchamp, ed., *Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music,* Springer, 2007.

[13] T. C. Bell, J. G. Cleary, and I. H. Witten, *Text Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[14] *Benny Goodman Orchestra, Jazz Collector Edition*, Laser Light Digital, 1991, Track 13: "Sing, Sing, Sing", Circa 1943 live broadcast of Benny Goodman Orchestra.

[15] D. Benson, *Music: A Mathematical Offering*, Cambridge, 2006.

[16] L. Bernstein, *The Joy of Music*, Amadeus Press, Pompton Plains, NJ, 2004.

[17] *Carreras, Domingo, Pavarotti in concert*, Orchestra del Maggio musicale fiorentino and Orchestra del Teatro dell'Opera di Roma, conducted by Zubin Mehta. Decca, 1990, Track 12: "Nessun Dorma".

[18] P. Cassaza, The art of frame theory, *Taiwanese J. of Math.* **4** (2000), 129–201, available at http://www.math.nthu.edu.tw/~tjm/abstract/0006/tjm0006_1.pdf.

[19] D. Chen and S. Qian, *Joint Time-Frequency Analysis: Methods and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1996.

[20] X. Cheng, J. V. Hart, and J. S. Walker, Time-frequency analysis of musical rhythm, *Notices of the American Mathematical Society* **56** (2009), 344–60, available at http://www.ams.org/notices/200903/.

[21] E. Chew, Math & Music—The Perfect Match, *OR/MS Today* **35** (2008), 26–31, available at www-rcf.usc.edu/~echew/papers/ORMS_Today_2008/2008-orms_today_ec_lowres.pdf.

[22] N. Chomsky, *Syntactic Structures*, Mouton de Gruyter, New York, 2002.

[23] O. Christensen, Pairs of dual Gabor frame generators with compact support and desired frequency localization, *Appl. Comput. Harmon. Anal.* **20** (2006), 403–10.

[24] O. Christensen, *An Introduction to Frames and Riesz Bases,* Birkhäuser, Boston, 2003.

[25] O. Christensen and W. Sun, Explicitly given pairs of dual frames with compactly supported generators and applications to irregular B-splines, *J. Approx. Theory* **151** (2008), 155–63.

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

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

[28] *Computer Music Journal*, webpage: http://www.mitpressjournals.org/cmj/.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[30] C-Sound is available at http://www.csounds.com/.

[31] I. Daubechies, *Ten Lectures on Wavelets,* SIAM, Philadelphia, PA, 1992.

[32] 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, 527–46.

[33] I. Daubechies, A. Grossman, and Y. Meyer, Painless nonorthogonal expansions, *J. Math. Phys.* **27** (1986), 1271–83.

[34] G. W. Don, Brilliant colors provocatively mixed: Overtone structures in the music of Debussy, *Music Theory Spectrum* **23** (2001), 61–73.

[35] G. W. Don, "Fern", Unpublished recording, available at http://www.uwec.edu/walkerjs/MBSGC/fern.wav.

[36] G. W. Don and J. S. Walker, "Sierpinski Round", unpublished recording, available at http://www.uwec.edu/walkerjs/MBSGC/Sierpinski_Round.wav.

[37] M. Dörfler, Gabor Analysis for a Class of Signals Called Music, Ph.D. thesis, Univ. of Vienna, 2002.

[38] M. Dörfler, Time-frequency analysis for music signals—A mathematical approach, *J. of New Music Research* **30** (March 2001) .

[39] M. DÖRFLER and H. FEICHTINGER, Quilted Gabor families I: Reduced multi-Gabor frames, *Appl. Comput. Harmon. Anal.* **356** (2004), 2001–23, available at http://homepage.univie.ac.at/monika.doerfler/.
[40] 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.
[41] M. DÖRFLER and B. TORRÉSANI, Representation of operators in the time-frequency domain and generalized Gabor multipliers, to appear in *J. Fourier Anal. Appl.*
[42] J. S. DOWNIE, Music Information Retrieval online bibliography: http://www.music-ir.org/research_home.html.
[43] J. DU, M. W. WONG, and H. ZHU, Continuous and discrete inversion formulas for the Stockwell transform, *Integral Transforms Spec. Funct.* **18** (2007), 537–43.
[44] J. DUESENBERRY, Convolution reverb and beyond, *Electronic Musician,* April 1, 2005, available at http://emusician.com/mag/emusic_convolution_reverb_beyond/index.html.
[45] R. J. DUFFIN and A. C. SCHAEFFER, A class of nonharmonic Fourier series, *Trans. Amer. Math. Soc.* **72** (1952), 341–66.
[46] *Electric Ladyland,* Jimi Hendrix Experience, Reprise, 1968, Track 15: "All Along the Watchtower", converted to mono by AUDACITY sound editor.
[47] J. FAUVEL, R. FLOOD, and R. WILSON (eds.), *Music and Mathematics: From Pythagoras to Fractals*, Oxford, New York, 2003.
[48] H. G. FEICHTINGER, F. LUEF, and T. WERTHER, A Guided Tour from Linear Algebra to the Foundations of Gabor Analysis, in *Gabor and Wavelet Frames,* IMS Lecture Notes Series, **10** (2007), 1–49.
[49] H. G. FEICHTINGER and K. NOWAK, 2003, A First Survey of Gabor Multipliers, In reference [51], available at http://www.univie.ac.at/nuhag-php/home/feipub_db.php.
[50] H. G. FEICHTINGER and T. STROHMER (eds.), *Gabor Analysis and Algorithms: Theory and Applications,* Birkhäuser, Boston, MA, 1998.
[51] H. G. FEICHTINGER and T. STROHMER (eds.), *Advances in Gabor Analysis*, Birkhäuser, Boston, MA, 2002.
[52] S. FLINN, Shepard's Tones webpage: http://www.cs.ubc.ca/nest/imager/contributions/flinn/Illusions/ST/st.html.
[53] D. GABOR, Theory of communication, *Journal of the Institute for Electrical Engineers* **93** (1946), 873–80.
[54] P. C. GIBSON, M. P. LAMOUREUX, and G. F. MARGRAVE, Letter to the editor: Stockwell and wavelet transforms, *J. Fourier Anal. Appl.* **12** (2006), 713–21.
[55] K. GRÖCHENIG, *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston, MA, 2001.
[56] L. HARKLEROAD, *The Math Behind the Music,* Cambridge University Press, Cambridge, UK, 2006.
[57] D. M. HEALY JR. and S. LI, A parametric class of discrete Gabor expansions, *IEEE Trans. Signal Process.* **44** (1996), 201–11.
[58] C. E. HEIL and D. F. WALNUT, Continuous and discrete wavelet transforms, *SIAM Review* **31** (1989), 628–66, available at http://people.math.gatech.edu/~heil/papers/siam.pdf.
[59] H. HELMHOLTZ, *On the Sensations of Tone*, Dover, New York, 1954.
[60] D. R. HOFSTADTER, *Gödel, Escher, Bach: An Eternal Golden Braid,* Vintage, New York, 1980.
[61] D. KROODSMA, *The Singing Life of Birds,* Houghton-Mifflin, New York, 2005.
[62] J. LATARTARA, Pedagogic applications of Fourier analysis and spectrographs in the music theory classroom, *J. of Music Theory Pedagogy* **22** (2008), 61–90.
[63] *Layla and Other Assorted Love Songs,* Derek and the Dominoes, original recording remastered, Polydor, 1970, Track 13: "Layla", converted to mono by AUDACITY sound editor.
[64] *Louis Armstrong: Gold,* Hip-O Records, 1996, original June 26, 1950, recording remastered, Disc 2, Track 3: "La Vie en Rose".
[65] G. LOY, *Musimathics: The Mathematical Foundations of Music, Vol. 2,* MIT Press, Cambridge, MA, 2007.
[66] Ludwig van Beethoven, Piano Sonata in E (Opus 109), Movements 1 and 2, performed by Daniel Añez Garcia on January 25, 2007, at Jeneusses Musicales du Canada, available at http://www.youtube.com/watch?v=e5jgYHGoAg0&feature=related.
[67] F. B. MACHE, *Music, Myth and Nature, Contemporary Music Studies, Vol. 6,* Taylor & Francis, London, 1993.
[68] S. MALLAT, *A Wavelet Tour of Signal Processing, Second Edition,* Academic Press, San Diego, CA, 1999.
[69] MATLAB, Time-frequency toolboxes, available at http://www.univie.ac.at/nuhag-php/mmodule/.
[70] G. MAZZOLA, *The Topos of Music,* Birkhäuser, Basel, 2002.
[71] METASYNTH, music software, http://www.uisoftware.com/MetaSynth/.
[72] L. B. MEYER, *Music the Arts and Ideas: Patterns and Predictions in Twentieth-Century Culture,* University of Chicago Press, 1967.
[73] S. MITHEN, *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body,* Harvard University Press, 2006.
[74] *J. of New Music Research,* website: http://www.tandf.co.uk/journals/nnmr/.
[75] W. O'GRADY, M. DOBROVOLSKY, and M. ARNOFF, *Contemporary Linguistics, an Introduction,* St. Martin's Press, New York, 1993.
[76] L. R. RABINER and J. B. ALLEN, Short-time Fourier analysis techniques for FIR system identication and power spectrum estimation, *IEEE Trans. Acoust. Speech Signal Process.* **27** (1979), 182–92.
[77] *Perspectives of New Music,* website: http://www.perspectivesofnewmusic.org/.
[78] J. RAHN webpage: http://faculty.washington.edu/~jrahn/.
[79] A. ROSS, *The Rest Is Noise: Listening to the Twentieth Century,* Farrar, Straus and Giroux, New York, 2007.
[80] D. ROTHENBERG, *Why Birds Sing: A Journey into the Mystery of Bird Song,* Basic Books, New York, 2005.

[81] *Satie: L'Oeuvre pour piano, vol. 1,* Aldo Ciccolini, piano, recorded in 1983, EMI, 1987, Track 7: "Gymnopédie I".

[82] K. SCHNASS, Gabor Multipliers: A Self-contained Survey, master's thesis, Univ. of Vienna, 2004.

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

[84] W. SETHARES, *Tuning, Timbre, Spectrum, Scale*, Springer, London, 2005.

[85] R. N. SHEPARD, Circularity in judgements of relative pitch, *J. of the Acoustical Soc. of America* **36** (1964), 2346–53.

[86] L. M. SMITH, Modelling rhythm perception by continuous time-frequency analysis, *Proceedings of the 1996 International Computer Music Conference,* Hong Kong, 392–95.

[87] L. M. SMITH, A multiresolution time-frequency analysis and interpretation of musical rhythm, thesis, University of Western Australia, 2000.

[88] L. M. SMITH and H. HONING, Time-frequency representation of musical rhythm by continuous wavelets, *J. of Mathematics and Music* **2** (2008), 81–97.

[89] E. STERN, quoted from the *Pandalous* website, available at `http://www.pandalous.com/` using search topic: Louis Armstrong.

[90] R. G. STOCKWELL, A basis for efficient representation of the *S*-transform, *Digital Signal Processing,* **17** (2007), 371–93.

[91] R. G. STOCKWELL, L. MANSINHA, and R. P. LOWE, Localization of the complex spectrum: the *S* transform, *IEEE Trans. Signal Processing* **44** (1996), 998–1001.

[92] T. STROHMER, Numerical algorithms for discrete Gabor expansions, see [50, pp. 267–94].

[93] *Suite: The Firebird,* Igor Stravinsky, Boston Symphony Orchestra, 1964 recording, cond. Erich Leinsdorf, BMG Classics, 1991.

[94] *Time Further Out,* Dave Brubeck Quartet, SONY, 1961, Track 7: "Unsquare Dance", recorded on June 8, 1961.

[95] D. TYMOCZKO, The geometry of musical chords, *Science* **313** (2006), 72–4, available at `http://www.sciencemag.org/cgi/content/full/313/5783/72/DC1`.

[96] 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/MBSGC/PrimerExtracts/`.

[97] T. WERTHER ET AL., CPR Artifact Removal in Ventricular Fibrillation ECG Signals Using Gabor Multipliers, *IEEE Trans. Biomedical Engineering* **56** (2009), 320–27.

[98] *Why Birds Sing* website: `http://www.whybirdssing.com/`.

[99] R. M. YOUNG, *An Introduction to Nonharmonic Fourier Series,* Academic Press, New York, 1980.

[100] M. ZIBULSKI and Y. ZEEVI, Discrete multiwindow Gabor-type transforms, *IEEE Trans. Signal Processing* **45** (1997), 1428–42.

[101] Online version of figures: Available at `http://www.ams.org/noticespages/201001/2010010030/`.