**FIGURE 6.2**

(a) 2-level Daub 9/7 wavelet transform of *Barbara* image with magnitudes of fluctuation values larger than 8 shown in white. (b) Similar display for the 2-level Daub 9/7 wavelet packet transform of *Barbara* image.

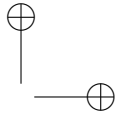
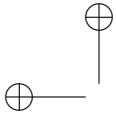
exhibits a more compact distribution of significant coefficients, hence a greater compression, than the 4-level wavelet transform.

6.2.1 Fingerprint compression

This example also gives us some further insight into the WSQ method of fingerprint compression. WSQ achieves *at least* 15:1 compression, without noticeable loss of detail, on all fingerprint images. It achieves such a remarkable result by applying a wavelet packet transform for which not every subimage is subjected to a further 1-level wavelet transform, but a large percentage of subimages are further transformed. To illustrate why this might be advantageous, consider the two transforms of the *Barbara* image in Figure 6.2. Notice that the vertical fluctuation \mathbf{v}^1 in the lower right quadrant of Figure 6.2(a) does not seem to be significantly compressed by applying another 1-level wavelet transform. Therefore, some advantage in compression might be obtained by *not* applying a 1-level wavelet transform to this subimage, while applying it to the other subimages. The wavelet packet transform used by WSQ is described in [6] and [7]. It should also be noted that JPEG 2000 Part 2 allows for wavelet packet transforms, see p. 606 of [7].

6.3 Continuous wavelet transforms

In this section and the next we shall describe the concept of a *continuous wavelet transform* (CWT), and how it can be approximated in a discrete form using a computer. We begin our discussion by describing one type of CWT, known as the *Mexican hat* CWT, which has been used extensively in seismic analysis. In the next section we turn to a second type of CWT, the Gabor CWT, which has many applications to analyzing audio signals. Although we do not have space for a thorough treatment of CWTs, we can nevertheless



introduce some of the essential ideas.

The notion of a CWT is founded upon many of the concepts that we introduced in our discussion of discrete wavelet analysis in Chapters 2 through 5, especially the ideas connected with discrete correlations and frequency analysis. A CWT provides a very redundant, but also very finely detailed, description of a signal in terms of both time and frequency. CWTs are particularly helpful in tackling problems involving signal identification and detection of hidden transients (hard to detect, short-lived elements of a signal).

To define a CWT we begin with a given function $\Psi(x)$, which is called the *analyzing wavelet*. For instance, if we define $\Psi(x)$ by

$$\Psi(x) = 2\pi w^{-1/2} [1 - 2\pi(x/w)^2] e^{-\pi(x/w)^2}, \quad w = 1/16, \quad (6.7)$$

then this analyzing wavelet is called a *Mexican hat wavelet*, with *width parameter* $w = 1/16$. See Figure 6.3(a).

It is possible to choose other values for w besides $1/16$, but this one example should suffice. By graphing the Mexican hat wavelet using different values of w , it is easy to see why w is called a width parameter. The smaller the value of w , the more the energy of $\Psi(x)$ is confined to a smaller interval of the x -axis.

The Mexican hat wavelet is not the only kind of analyzing wavelet. In the next section, we shall consider the Gabor wavelet, which is very advantageous for analyzing recordings of speech or music. We begin in this section with the Mexican hat wavelet because it allows us to more easily explain the concept of a CWT.

Given an analyzing wavelet $\Psi(x)$, a CWT of a discrete signal \mathbf{f} is defined by computing several correlations of this signal with discrete samplings of the functions $\Psi_s(x)$ defined by

$$\Psi_s(x) = \frac{1}{\sqrt{s}} \Psi\left(\frac{x}{s}\right), \quad s > 0. \quad (6.8)$$

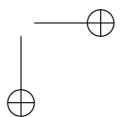
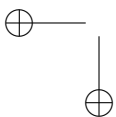
The parameter s is called a *scale parameter*. If we sample each signal $\Psi_s(x)$ at discrete time values t_1, t_2, \dots, t_N , where N is the length of \mathbf{f} , then we generate the discrete signals \mathbf{g}_s defined by

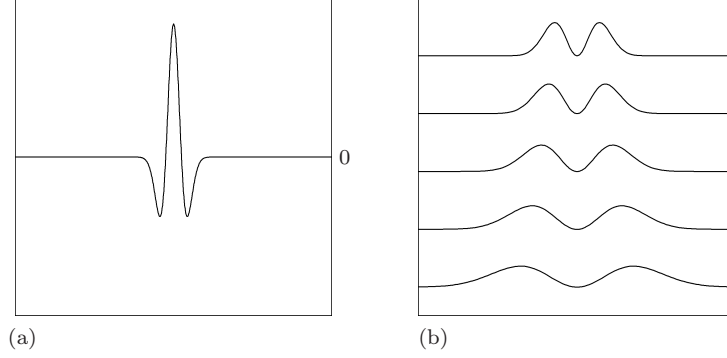
$$\mathbf{g}_s = (\Psi_s(t_1), \Psi_s(t_2), \dots, \Psi_s(t_N)).$$

A CWT of \mathbf{f} then consists of a collection of discrete correlations ($\mathbf{f} : \mathbf{g}_s$) over a finite collection of values of s . A common choice for these values is

$$s = 2^{-k/M}, \quad k = 0, 1, 2, \dots, I \cdot M$$

where the positive integer I is called the number of *octaves* and the positive integer M is called the number of *voices* per octave. For example, 8 octaves and 16 voices per octave is the default choice in FAWAV. Another popular choice is 6 octaves and 12 voices per octave. This latter choice of scales



**FIGURE 6.3**

(a) The Mexican hat wavelet, $w = 1/16$. (b) DFTs of discrete samplings of this wavelet for scales $s = 2^{-k/6}$, from $k = 0$ at the top, then $k = 2$, then $k = 4$, then $k = 6$, down to $k = 8$ at the bottom.

corresponds—based on the relationship between scales and frequencies that we describe below—to the scale of notes on a piano (also known as the *well-tempered scale*).

The purpose of computing all these correlations is that a *very finely detailed time-frequency analysis* of a signal can be carried out by making a judicious choice of the width parameter w and the number of octaves and voices. To see this, we observe that Formula (5.40) tells us that the DFTs of the correlations $(\mathbf{f} : \mathbf{g}_s)$ satisfy

$$(\mathbf{f} : \mathbf{g}_s) \xrightarrow{\mathcal{F}} \mathcal{F}\mathbf{f} \overline{\mathcal{F}\mathbf{g}_s}. \quad (6.9)$$

For a Mexican hat wavelet, $\mathcal{F}\mathbf{g}_s$ is real-valued; hence $\overline{\mathcal{F}\mathbf{g}_s} = \mathcal{F}\mathbf{g}_s$. Therefore Equation (6.9) becomes

$$(\mathbf{f} : \mathbf{g}_s) \xrightarrow{\mathcal{F}} \mathcal{F}\mathbf{f} \mathcal{F}\mathbf{g}_s. \quad (6.10)$$

Formula (6.10) is the basis for a very finely detailed time-frequency decomposition of a discrete signal \mathbf{f} . For example, in Figure 6.3(b) we show graphs of the DFTs $\mathcal{F}\mathbf{g}_s$ for the scale values $s = 2^{-k/6}$, with $k = 0, 2, 4, 6$, and 8 . These graphs show that when these DFTs are multiplied with the DFT of \mathbf{f} , they provide a decomposition of $\mathcal{F}\mathbf{f}$ into a succession of frequency bands. These successive bands overlap each other, thus providing a very redundant decomposition of the DFT of \mathbf{f} . Notice that the bands containing higher frequencies correspond to smaller scale values; *there is a reciprocal relationship between scale values and frequency values*. It is also important to note that there is a decomposition in time, due to the essentially finite width of the Mexican hat wavelet [see Figure 6.3(a)].

A couple of examples should help to clarify these points. The first example we shall consider is a test case designed to illustrate the connection between

a CWT and the time-location of frequencies in a signal. The second example is an illustration of how a CWT can be used for analyzing an ECG signal.

For our first example, we shall analyze a discrete signal \mathbf{f} , obtained from 2048 equally spaced samples of the following analog signal:

$$\begin{aligned} & \sin(40\pi x)e^{-100\pi(x-.2)^2} \\ & + [\sin(40\pi x) + 2\cos(160\pi x)]e^{-50\pi(x-.5)^2} \\ & + 2\sin(160\pi x)e^{-100\pi(x-.8)^2} \end{aligned} \quad (6.11)$$

over the interval $0 \leq x \leq 1$. See the top of Figure 6.4(a).

The signal in (6.11) consists of three terms. The first term contains a sine factor, $\sin(40\pi x)$, of frequency 20. Its other factor, $e^{-100\pi(x-.2)^2}$, serves as a damping factor which limits the energy of this term to a small interval centered on $x = 0.2$. This first term appears most prominently on the left-third of the graph at the top of Figure 6.4(a). Likewise, the third term contains a sine factor, $2\sin(160\pi x)$, of frequency 80, and this term appears most prominently on the right-third of the signal's graph. Notice that this frequency of 80 is four times as large as the first frequency of 20. Finally, the middle term

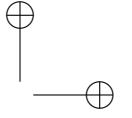
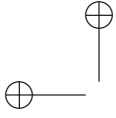
$$[\sin(40\pi x) + 2\cos(160\pi x)]e^{-50\pi(x-.5)^2}$$

has a factor containing both of these two frequencies, and can be observed most prominently within the middle of the signal's graph.

The CWT, also known as a *scalogram*, for this signal is shown at the bottom of Figure 6.4(a). The analyzing wavelet used to produce this CWT was a Mexican hat wavelet of width $1/16$, with scales ranging over 8 octaves and 16 voices. The labels on the right side of the figure indicate *reciprocals* of the scales used. Because of the reciprocal relationship between scale and frequency noted above, this reciprocal-scale axis can also be viewed as a frequency axis. Notice that the four most prominent portions of this scalogram are aligned directly below the three most prominent parts of the signal. Of equal importance is the fact that these four portions of the scalogram are centered on two reciprocal-scales, $1/s \approx 2^{2.2}$ and $1/s \approx 2^{4.2}$. The second reciprocal scale is four times larger than the first reciprocal scale, just as the frequency 80 is four times larger than the frequency 20. Bearing this fact in mind, and recalling the alignment of the prominent regions of the scalogram with the three parts of the signal, we can see that the CWT provides us with a *time-frequency portrait* of the signal.

We have shown that it is possible to correctly interpret the meaning of this scalogram; nevertheless, we can produce a much simpler and more easily interpretable scalogram for this test signal using a Gabor analyzing wavelet. See Figure 6.5(a). We shall discuss this Gabor scalogram in the next section.

Our second example uses a Mexican hat CWT for analyzing a signal containing several transient bursts, a simulated ECG signal first considered in Section 5.4. See the top of Figure 6.4(b). The bottom of Figure 6.4(b) is a



scalogram of this signal using a Mexican hat wavelet of width 2, over a range of 8 octaves and 16 voices. This scalogram shows how a Mexican hat wavelet can be used for detecting the onset and demise of each heartbeat. In particular, the aberrant, fourth heartbeat is singled out from the others by the longer vertical ridges extending upwards to the highest frequencies (at the eighth octave). Although this example is only a simulation, it does show the ease with which the Mexican hat CWT detects the presence of short-lived parts of a signal. Similar identifications of transient bursts are needed in seismology for the detection of earthquake tremors. Consequently, Mexican hat wavelets are widely used in seismology.

6.4 Gabor wavelets and speech analysis

In this section we describe Gabor wavelets, which are similar to the Mexican hat wavelets examined in the previous section, but provide a more powerful tool for analyzing speech and music. We shall first go over their definition, and then illustrate their use with some examples.

A *Gabor wavelet*, with width parameter w and frequency parameter ν , is the following analyzing wavelet:

$$\Psi(x) = w^{-1/2} e^{-\pi(x/w)^2} e^{i2\pi\nu x/w}. \quad (6.12)$$

This wavelet is complex valued. Its real part $\Psi_R(x)$ and imaginary part $\Psi_I(x)$ are

$$\Psi_R(x) = w^{-1/2} e^{-\pi(x/w)^2} \cos(2\pi\nu x/w), \quad (6.13a)$$

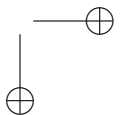
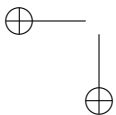
$$\Psi_I(x) = w^{-1/2} e^{-\pi(x/w)^2} \sin(2\pi\nu x/w). \quad (6.13b)$$

The width parameter w plays the same role as for the Mexican hat wavelet; it controls the width of the region over which most of the energy of $\Psi(x)$ is concentrated. The value ν/w is called the *base frequency* for a Gabor CWT.

One advantage that Gabor wavelets have when analyzing sound signals is that they contain factors of cosines and sines, as shown in (6.13a) and (6.13b). These cosine and sine factors allow the Gabor wavelets to create easily interpretable scalograms of those signals which are combinations of cosines and sines—the most common instances of such signals are recorded music and speech. We shall see this in a moment, but first we need to say a little more about the CWT defined by a Gabor analyzing wavelet.

Because a Gabor wavelet is complex valued, it produces a complex-valued CWT. For many signals, it is often sufficient to just examine the magnitudes of the Gabor CWT values. In particular, this is the case with the signals analyzed in the following examples.

For our first example, we use a Gabor wavelet with width 1 and frequency 5 for analyzing the signal in (6.11). The graph of this signal is shown at the top of Figure 6.5(a). As we discussed in the previous section, this signal consists



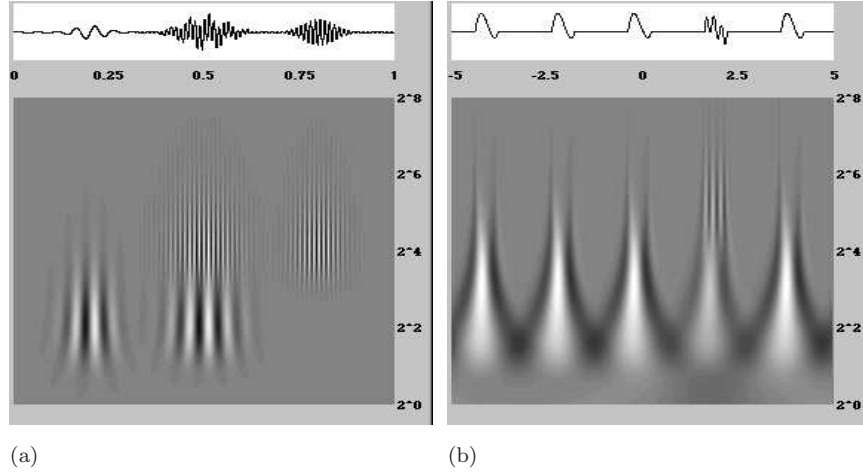


FIGURE 6.4

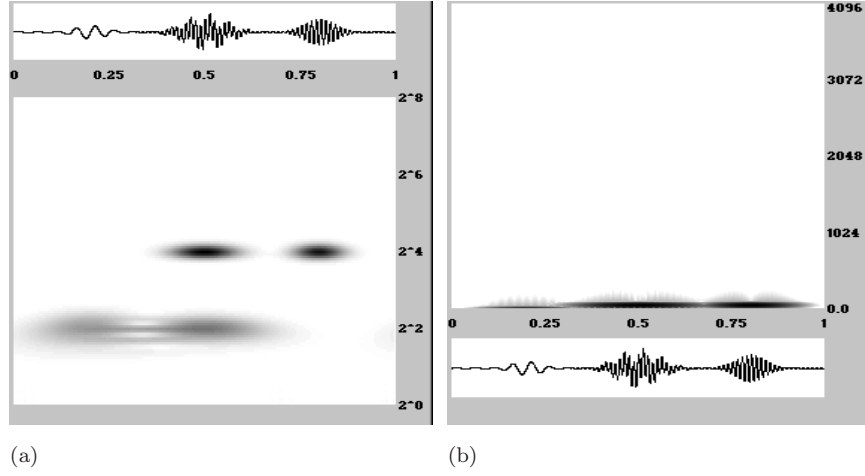
(a) Mexican hat CWT (scalogram) of a test signal with two main frequencies.
 (b) Mexican hat scalogram of simulated ECG signal. Whiter colors represent positive values, blacker values represent negative values, and the gray background represents zero values.

of three portions with associated frequencies of 20 and 80. The magnitudes for a Gabor scalogram of this signal, using 8 octaves and 16 voices, are graphed at the bottom of Figure 6.5(a). We see that this *magnitude-scalogram* consists of essentially just four prominent spots aligned directly below the three most prominent portions of the signal. These four spots are centered on the two reciprocal-scale values of 2^2 and 2^4 , which are in the same ratio as the two frequencies 20 and 80. [Notice also that the base frequency is $5/1 = 5$ and that $20 = 5 \cdot 2^2$, $80 = 5 \cdot 2^4$.] It is interesting to compare this scalogram with a spectrogram of the test signal shown in Figure 6.5(b). The spectrogram has all of its significant values crowded together at the lower frequencies; the scalogram is a zooming in on this lower range of frequencies, from 5 Hz to 1280 Hz along an octave-based scale.

It is interesting to compare Figures 6.4(a) and 6.5(a). The simplicity of Figure 6.5(a) makes it much easier to interpret. The reason that the Gabor CWT is so clean and simple is because, for the proper choices of width w and frequency ν , the test signal in (6.11) consists of terms that are identical in form to one of the functions in (6.13a) or (6.13b). Therefore, when a scale value s produces a function $\Phi_R(x/s)/\sqrt{s}$, or a function $\Phi_I(x/s)/\sqrt{s}$, having a form similar to one of the terms in (6.11), then the correlation $(\mathbf{f} : \mathbf{g}_s)$ in the CWT will have some high-magnitude values.

This first example might appear to be rather limited in scope. After all, how many signals encountered in the real world are so nicely put together as this test signal? Our next example, however, shows that a Gabor CWT performs equally well in analyzing a real signal: a speech signal.

In Figure 6.6(b) we show a Gabor magnitude-scalogram of a recording

**FIGURE 6.5**

(a) Magnitudes of Gabor scalogram of test signal. (b) Spectrogram of test signal. Darker regions denote larger magnitudes; lighter regions denote smaller magnitudes.

of the author saying the word *call*. The recorded signal, which is shown at the top of the figure, consist of three main portions. The first two portions correspond to the two sounds, *ca* and *ll*, that form the word *call*. The *ca* portion occupies a narrow area on the far left side of the *call* signal's graph, while the *ll* portion occupies a much larger area consisting of the middle half of the *call* signal's graph. The third portion lies at the right end of the signal and is a "clipping sound" that is the start of the consonant "b" that begins the word "back" (the *call* signal was clipped from a recording of the author speaking the phrase "you call back"). For comparison, we have also plotted the *call* signal's spectrogram in Figure 6.6(a). Again, as with our test signal, the scalogram is able to zoom in and better display the time-frequency structure of this speech signal.

To analyze the *call* signal, we used a Gabor wavelet of width $1/8$ and frequency 10 (hence a base frequency of $10/(1/8) = 80$ Hz), with scales ranging over 4 octaves and 16 voices. The resulting magnitude-scalogram is composed of several regions. The largest region is a collection of several horizontal bands lying below the *ll* portion. We shall concentrate on analyzing this region.

This region below the *ll* portion consists of seven horizontal bands centered on the following approximate reciprocal-scale values:

$$2^{0.625}, 2^{1.625}, 2^{2.188}, 2^{2.625}, 2^{2.953}, 2^{2.97}, 2^{3.219}, 2^{3.625}. \quad (6.14)$$

If we divide each of these values by the smallest one, $2^{0.625}$, we get the following approximate ratios:

$$1, 2, 3, 4, 5, 6, 7, 8. \quad (6.15)$$

Since reciprocal-scale values correspond to frequencies, we can see that these

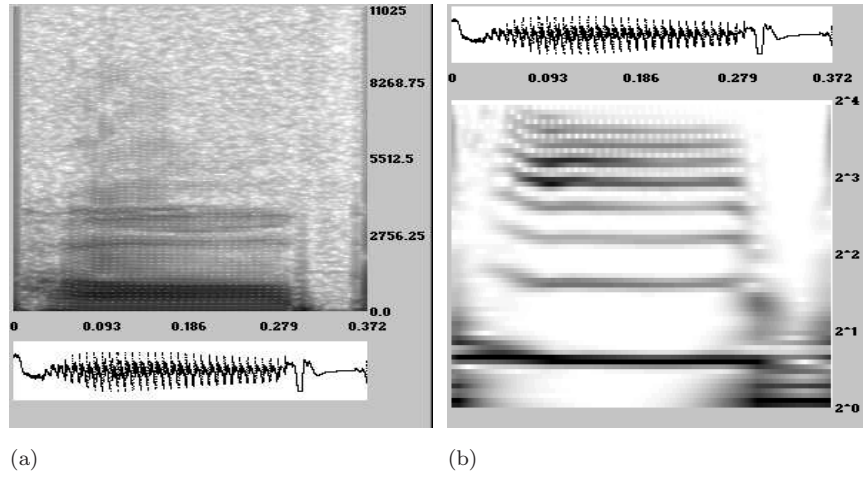


FIGURE 6.6

(a) Spectrogram of *call* sound. (b) Magnitudes of its Gabor scalogram.

bands correspond to frequencies on a harmonic (musical) scale. In fact, in Figure 6.7(b) we show a graph of the spectrum of a sound clip of the *ll* portion of the *call* signal. This spectrum shows that the frequencies of peak energy in the *ll* portion have the following approximate values:

$$124, 248, 372, 496, 620, 744, 868, 992. \quad (6.16)$$

Notice that these frequencies have the same ratios to the lowest frequency of 124 as the ratios in (6.15) [and that $124 \approx 80 \cdot 2^{0.625}$, \dots , $992 \approx 80 \cdot 2^{3.625}$].

This region of the scalogram illustrates an important property of many portions of speech signals, the property of *frequency banding*. These frequency bands are called *formants* in linguistics. All speakers, whether they are native

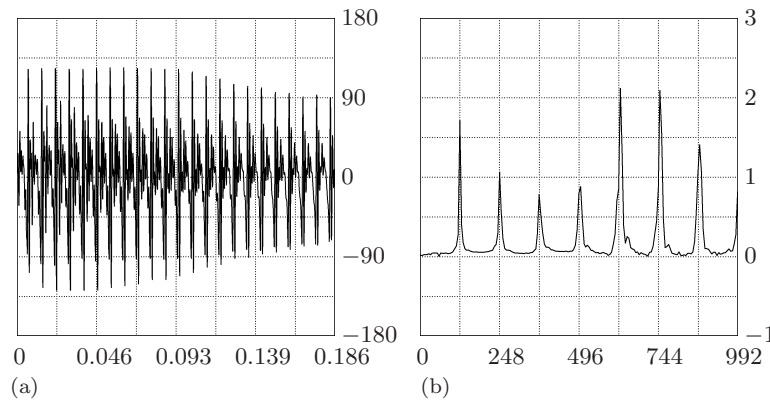


FIGURE 6.7

(a) A portion of the *ll* sound in the *call* signal; the horizontal axis is the time axis. (b) Spectrum of signal in (a); the horizontal axis is the frequency axis.

speakers of English or not, produce a sequence of such frequency bands for the *ll* portion of *call*. For some speakers, the bands are horizontal, while for other speakers the bands are curved. The *ll* sound is a fundamental unit of English speech, called a *phoneme*.

Notice that there are also curved formants directly preceding the straight formants of the *ll* sound; these formants describe the phoneme *aa* that immediately precedes the phoneme for *ll* in the word *call*. There are also two other structures in the scalogram that do not exhibit formants, these occur at the beginning and end of the recording. At the beginning there is a hard *c* sound corresponding to the beginning of the word *call*. In the scalogram we see that the portion at the far left is composed of a widely dispersed, almost continuous, range of frequencies without significant banding—this reflects the percussive nature of the hard consonant phoneme *c*. There is a similar continuous range of frequencies at the end of the *call* signal corresponding to the “clipping” sound that begins the consonant *b*.

This example shows what a powerful tool the Gabor CWT provides for analyzing a speech signal. We have used it to clearly distinguish the phonemes in the *call* sound, to understand the formant structure of its *ll* portion, and to determine that its consonants lack a formant structure.

Another application of these Gabor scalograms is that, when applied to recordings of different people saying *call*, they produce visibly different scalograms. These scalograms function as “fingerprints” for identifying different speakers. The ribbon structure of formants for the *ll* portion is displayed for all speakers, although tracing out different curves for different speakers.

6.4.1 Musical analysis: formants in song lyrics

We close this section by showing how Gabor scalograms, in concert with spectrograms, provide a powerful time-frequency approach to analyzing music. We already illustrated this once in Figure 5.15 (see p. 163). The image in Figure 5.15(b) is a Gabor scalogram that zooms in on the spectrogram in Figure 5.15(a) over a two-octave frequency range of 500 to 2000 Hz.

Another interesting musical example is shown in Figure 6.8. The image in Figure 6.8(a) is a Blackman windowed spectrogram of a passage from the song *Buenos Aires*. The dark curved ribbons are the formants of the singer’s voice as she sings the lyrics “you’ll be on me too.” In Figure 6.8(b) we show a scalogram that zooms in on a 3-octave range of the spectrogram spanning 200 to 1600 Hz.² The time-frequency analysis shown in these two images reveals aspects of the Multiresolution Principle at work: there are evident repetitions of time-frequency structures and substructures at different locations and with different sizes.

It is interesting to compare Figures 5.15 and 6.8. The former is based on an instrumental Chinese folk melody and the latter is based on a modern,

²See the *Examples & Exercises* material at the Primer website for more details on how we created Figure 6.8(b).

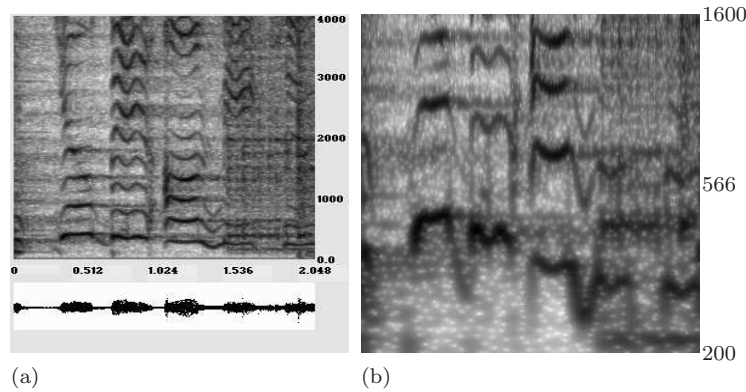


FIGURE 6.8

Time-frequency analysis of a passage from the song *Buenos Aires*. (a) Spectrogram. (b) Zooming in on three octaves of the frequency range of (a).

popular English song lyric. Notice that there is some similarity between the small angular structures: in the latter case, they are formants of the singer's voice; in the former case, they are "pitch excursions" of a Chinese stringed instrument that extends the power and range of human voice.

6.5 Percussion scalograms & musical rhythm

For our final discussion of the Primer, we describe how time-frequency methods can be used to analyze musical rhythm. In particular, we show how a new technique known as *percussion scalograms* makes use of both spectrograms and scalograms to analyze the multiple time-scales occurring within the rhythms of a percussion performance.

Our discussion will focus on two percussion sequences. The first sequence is an introductory passage from the song, *El Matador*, which is saved as the file `el_matador_percussion_clip.wav` at the Primer website. Listening to this file you will hear a relatively simple rhythm of drum beats, with one short shift in tempo, and with several whistle blowings as accompaniment. We shall use this passage to illustrate the basic principles underlying our approach. The second sequence, which will show the power of our method, is a complex Latin percussion passage that introduces the song *Buenos Aires*. This passage is saved in the file `Buenos Aires percussion clip.wav` at the Primer website. Listening to this file you will hear a richly structured percussion performance with several tempo shifts (as well as some high pitch background sound). Our percussion scalogram method will produce an objective description of these tempo shifts that accords well with our aural perception.

To derive our percussion scalogram method for analyzing drum rhythms, we consider the percussion sequence from the beginning of *El Matador*. In Figure 6.9(a) we show its spectrogram. This spectrogram is mostly composed

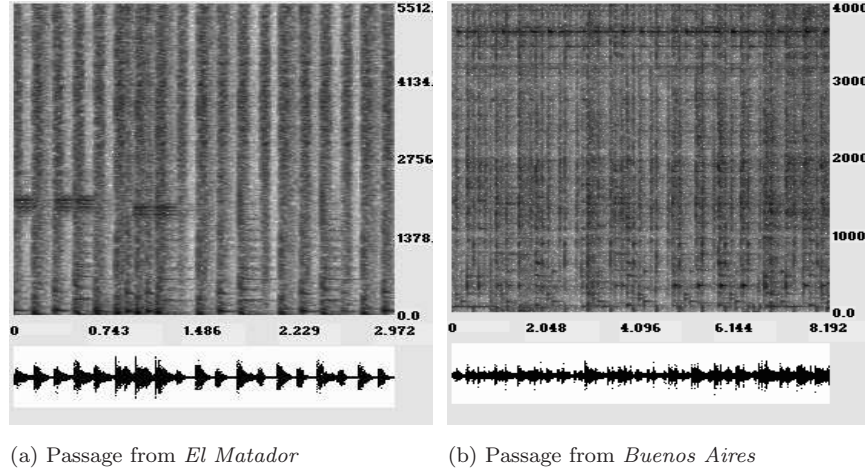


FIGURE 6.9
Spectrograms from two percussion sequences.

of a sequence of thick vertical segments, which we will call *vertical swatches*.³ Each vertical swatch corresponds to a percussive strike on a drum. These sharp strikes on drum heads excite a continuum of frequencies rather than a discrete tonal sequence of fundamentals and overtones. The rapid onset and decay of these strike sounds produces vertical swatches in the time-frequency plane. A more complex pattern of thinner vertical swatches can be seen in the spectrogram of the *Buenos Aires* percussion passage in Figure 6.9(b).

Our percussion scalogram method has the following two parts:

- I. *Pulse train generation.* We generate a “pulse train,” a sequence of alternating intervals of 1-values and 0-values (see the bottom graph in Figure 6.10). The location and duration of the intervals of 1-values corresponds to our hearing of the drum strikes, and the location and duration of the intervals of 0-values corresponds to the silences between the strikes. In Figure 6.10, the rectangular-shaped pulses correspond to sharp onset and decay of transient bursts in the percussion signal graphed just above the pulse train. The widths of these pulses are approximately equal to the widths of the vertical swatches shown in the spectrogram (we graphed only a portion of the spectrogram that omits the blotches from the whistle blowings, so as to isolate just the drum strikings). In Steps 1 and 2 of the method below we describe how this pulse train is generated.
- II. *Gabor CWT.* We use a Gabor CWT to analyze the pulse train. The rationale for doing this is that the pulse train is a step function analog of a sinusoidal of varying frequency. Because of this rough correlation between tempo of pulses in a pulse train and frequency in sinusoidal curves, we

³The whistle blowings correspond to three rectangular blotches at the left center of the spectrogram.

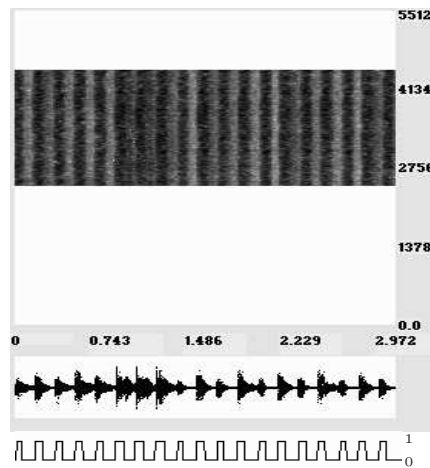


FIGURE 6.10
Pulse Train for the *El Matador* percussion sequence.

employ a Gabor CWT for analysis.⁴ For example, see Figure 6.11. The thick vertical line segments in the top half of the scalogram correspond to the drum strikes, and there is a connecting region at the bottom of three of the segments (the 6th, 7th, and 8th segments counting from the left). When listening to the passage we hear those three strikes as a group with a clearly defined tempo shift.⁵ This CWT calculation is performed in Step 3 of the method.

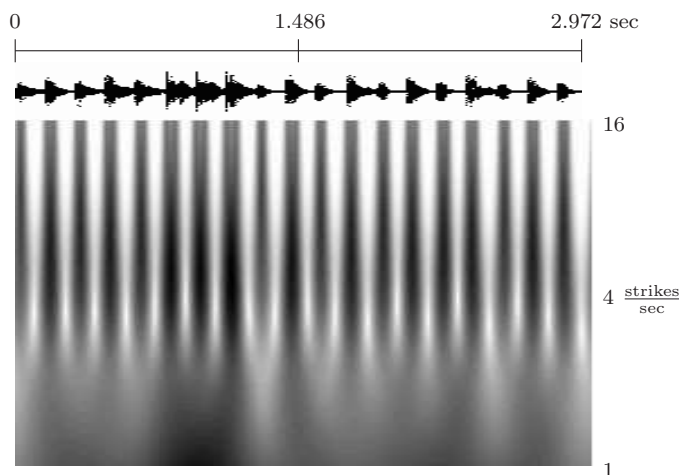
Now that we have outlined the basis for the percussion scalogram method, we can list it in detail. The percussion scalogram method for analyzing percussive rhythm consists of the following three steps.

Percussion Scalogram Method

Step 1. Compute a signal consisting of averages of the Gabor transform square-magnitudes for horizontal slices lying within a frequency range that consists mostly of vertical swatches. For the time intervals corresponding to vertical swatches in the spectrogram (for example, as shown in Figure 6.10) this step will produce higher square-magnitude values that lie above the mean of all square-magnitudes (because the mean is pulled down by the intervals of silence). For the *El Matador* sequence, the frequency range of 2500 to 4500 Hz was used, as it consists mostly

⁴In his thesis [18], Leigh M. Smith provides a thorough empirical study of the efficacy of using Gabor CWTs to analyze percussive pulse trains.

⁵An important feature of the FAWAV program used to generate this percussion scalogram is that the sound file can be played and a cursor will travel across the percussion scalogram, confirming our statements about the meaning of its features. More details on how to do this are given in the *Examples & Exercises* material at the Primer website.

**FIGURE 6.11**

Rhythmic analysis of *El Matador* percussion sequence. The percussion sequence is graphed on top, and below it is its percussion scalogram using 4 octaves, 64 voices, width 0.5, freq. 0.5 (obtained from the frequency range 2500 to 4500 Hz of its spectrogram).

of the vertical swatches corresponding to the percussive strikes. (Note: In a *purely percussive passage*, containing only drum strikes without any background sounds, the complete frequency range can be used.)

Step 2. Compute a signal that is 1 whenever the signal from Step 1 is larger than its mean and 0 otherwise. As the discussion in Step 1 shows, this will produce a pulse train whose intervals of 1-values mark off the position and duration of the vertical swatches (hence of the drum strikes). Figure 6.10 illustrates this clearly.

Step 3. Compute a Gabor CWT of the pulse train signal from Step 2. As we shall now discuss, this Gabor CWT provides an objective picture of the varying rhythms within a percussion performance.

We have already discussed the Gabor CWT shown in Figure 6.11 for the *El Matador* percussion sequence. The only points we want to add to that discussion are some details on the meanings of the parameters used for the Gabor CWT. To create the Gabor CWT in Figure 6.11, we used a width parameter of 0.5 and a frequency parameter of 0.5. That yields a base frequency of 1.0 strikes/sec. The range of 4 octaves that we used then gives an upper frequency value of 16 strikes/sec.⁶ Notice that the vertical bars in the scalogram in Figure 6.11 are centered on a frequency value of about 6 or 7 strikes/sec and that corresponds to the number of strikes that one detects

⁶These CWT parameter values were obtained empirically. An apriori (automatic) selection method is the subject of current research. See the preprint [22].

within any given 1 second interval; this illustrates that it is correct to interpret the vertical axis for the scalogram as an (octave-scaled) frequency axis of strikes/sec.

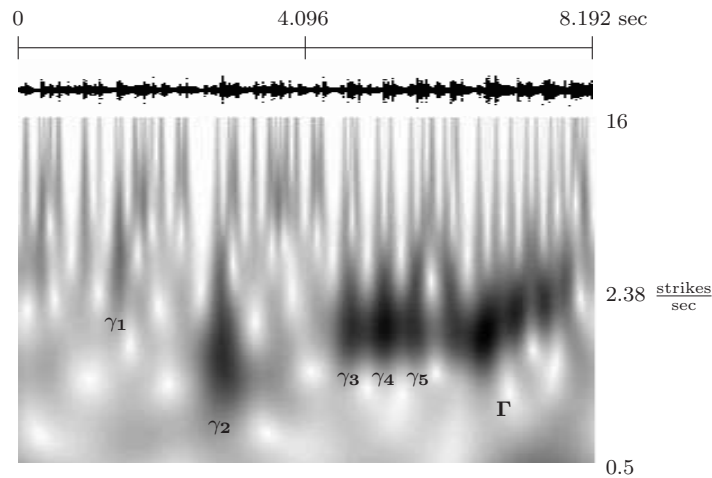
6.5.1 Analysis of a complex percussive rhythm

As an illustration of the power of our method, we use a percussion scalogram to analyze the complex rhythms of the opening percussion passage from the *Buenos Aires* song. See Figure 6.12. To isolate the percussive sounds, the drum strikes, from the rest of the sounds in the passage, we used a frequency range of 2000 to 3000 Hz in Step 1 of the percussion scalogram method. The parameters of the Gabor CWT are specified in the caption of Figure 6.12.

As with the *El Matador* sequence, it helps to play the recording of the percussion passage and watch the cursor trace over the percussion scalogram. After listening a couple of times, and watching the cursor run along the top of the scalogram, you should find that the thin vertical strips at the top of the scalogram correspond to the individual drum strikes. What is even more interesting, however, is that several of these vertical strips bind together into larger blobs lower down on the frequency scale (for instance, the blobs above the labels γ_1 to γ_5 in the figure). If you listen again to the recording and watch the cursor as it passes these blobs γ_1 to γ_5 , you will notice that the strikes occur in groups that correspond precisely to these blobs. Furthermore, the blobs γ_3 to γ_5 are connected together, and one does perceive a larger time-scale grouping of percussion strikes over the time-interval covered by these three blobs. Finally, we note that there is another collection of blobs to the right of γ_5 . We have not labeled them, but we leave it to the reader to infer the connection between them and the shifting pattern of drum strikes in the recording. Notice, however, that these blobs appear to be linked to a larger region, labeled Γ , which provides an objective description of our aural perception of the further grouping of these collections of drum strikes.

6.5.2 Multiresolution Principle for rhythm

Our discussion of these two percussion sequences illustrates the fact that the Multiresolution Principle for tonal music that we introduced on p. 160—the patterning of time-frequency structures over multiple time-scales—also applies to rhythmic percussion. We can even see the three representations described by Pinker (p. 164) applying as well, if we substitute “strikes” for “notes.” For example, the single strikes are grouped into blobs, and some of these blobs are joined together into longer groups. This multiresolution time-frequency patterning, captured by our percussion scalograms, may be useful in characterizing different styles of percussion. But that is a subject for future research.

**FIGURE 6.12**

Rhythmic analysis of *Buenos Aires* percussion sequence. The percussion sequence is graphed on top, and below it is its percussion scalogram using 5 octaves, 51 voices, width 2, freq. 1 (obtained from the frequency range 2000 to 3000 Hz of its spectrogram). The labels are explained in the text.

6.6 Notes and References

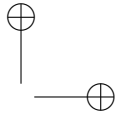
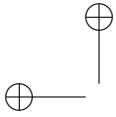
The best introductory material on wavelet packet transforms can be found in [2] and [3]. There is also a good discussion in [4]. A very thorough treatment of the subject is given in [5]. The relation between wavelet packet transforms and the WSQ method is described in [6], and the wavelet packet transform option allowed by JPEG 2000 Part 2 is described in [7].

Rigorous expositions of the complete theory of CWTs can be found in [8] and [9]. A more complete treatment of the discrete version described in this primer is given in [10].

For a discussion of the uses of the CWT for analysis of ECGs, see [11] and [12]. In addition to an excellent discussion of the topic, [12] also contains an extensive bibliography.

Applying Gabor CWTs to the detection of engine malfunctions in Japanese automobiles is described in [13]. An interesting relationship between CWTs and human hearing, with applications to speech analysis, is described in [14]. Background on formants and phonemes in linguistics can be found in [15].

Scalograms are used, in conjunction with spectrograms, to provide analysis of music and musical instruments in [16] and [17]. An empirical discussion of percussive rhythm and applications is given by Leigh Smith in [18]. Some other papers on related topics can be found on his webpage [19]. William Sethares has done profound work on computerized rhythm analysis [20], [21].



The preprint [22] describes an alternative approach to this same topic.

6.6.1 Additional References

We now provide references for some additional topics that extend the discussion in the text. These topics are (1) best basis algorithms, (2) local cosine series, (3) frames, (4) curvelets, (5) 3D wavelets, and (6) video compression.

Best basis algorithms. There are algorithms that choose which fluctuation subsignals to decompose further (by wavelet transform) according to some minimization of a *cost function*. An excellent introduction to such a best basis algorithm, and its application to the WSQ algorithm, can be found in David Walnut's book [23]. A complete discussion, by the discoverer of the technique, is in [5].

Local cosine series. The Gabor transform discussed in Chapter 5 has been criticized for its reliance on windows of constant size. A method which allows for windows of varying size, as well as using real-valued cosine functions, is the method of *local cosine series*. Local cosine series have proven to be quite useful in compression of signals. An elementary discussion can be found in Chapter 2 of [24]. See also the paper by Auscher et al in [25], and the treatment in [5]. A related area to local cosine series are the fields of *lapped orthogonal transforms* and *generalized lapped orthogonal transforms* (GenLOT). These fields are described in [26] to [30].

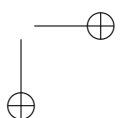
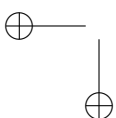
Frames. The inversion of Gabor transforms discussed in Chapter 5 provides one class of expansions of signals using *frames*. Frames have proven to be especially useful in denoising applications. See [31] to [35]. References [36] and [37] provide fundamental mathematical background.

Curvelets. David Donoho and his collaborators have done extensive work on a generalization of wavelets for image processing known as curvelets. Curvelets provide for efficient modelling of edges in images. See the papers [38] to [40], and also the website [41].

3D wavelets. Wim Sweldens and his collaborators have done a lot of work on wavelets and their generalizations in 3D. See [42] and [43], and the website [44]. Curvelets have also been adapted for 3D processing [45].

Video compression. A good synopsis of the basics of wavelet-based video compression can be found in [10]. Important work in the field is being done by Truong Nguyen's group, see [46] for many downloadable publications.

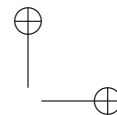
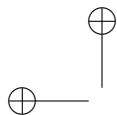
1. B. Burke. (1994). The Mathematical Microscope: waves, wavelets, and beyond. In *A Positron Named Priscilla, Scientific Discovery at the Frontier*, M. Bartusiak (Ed.), 196–235, National Academy Press.
2. M.V. Wickerhauser. (1993). Best-adapted Wavelet Packet Bases. In *Different Perspectives on Wavelets*, I. Daubechies (Ed.), AMS, Providence, RI, 155–172.



3. R.R. Coifman, M.V. Wickerhauser. (1993). Wavelets and Adapted Waveform Analysis. A Toolkit for Signal Processing and Numerical Analysis. In *Different Perspectives on Wavelets*, I. Daubechies (Ed.), AMS, Providence, RI, 119–154.
4. R.R. Coifman and M.V. Wickerhauser. (1994). Wavelets and Adapted Waveform Analysis. In *Wavelets. Mathematics and Applications*, J. Benedetto, M. Frazier (Eds.), CRC Press, Boca Raton, FL, 399–424.
5. M.V. Wickerhauser. (1994). *Adapted Wavelet Analysis from Theory to Software*. A.K. Peters, Wellesley, MA, 1994.
6. J.N. Bradley, C.M. Brislawn, T. Hopper. (1993). The FBI Wavelet/Scalar Quantization Standard for gray-scale fingerprint image compression. *SPIE*, Vol. 1961, *Visual Information Processing II (1993)*, 293–304.
7. D.S. Taubman and M.W. Marcellin. (2002). JPEG2000: Image compression fundamentals, standards and practice. Kluwer, Boston, MA.
8. I. Daubechies. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA.
9. A.K. Louis, P. Maaß, A. Rieder. (1997). *Wavelets, Theory and Applications*. Wiley, New York, NY.
10. S. Mallat. (1999). *A Wavelet Tour of Signal Processing. Second Edition*. Academic Press, New York, NY.
11. L. Senhadji, L. Thoraval, G. Carrault. (1996). Continuous Wavelet Transform: ECG Recognition Based on Phase and Modulus Representations and Hidden Markov Models. In *Wavelets in Medicine and Biology*, A. Aldroubi, M. Unser (Eds.), CRC Press, Boca Raton, FL, 439–464.
12. P.S. Addison. (2005). Wavelet transforms and the ECG: a review. *Physiol. Meas.*, Vol. 26, R155–R199.
13. M. Kobayashi. (1996). Listening for Defects: Wavelet-Based Acoustical Signal Processing in Japan. *SIAM News*, Vol. 29, No. 2.
14. I. Daubechies and S. Maes. (1996). A Nonlinear Squeezing of the Continuous Wavelet Transform Based on Auditory Nerve Models. In *Wavelets in Medicine and Biology*, A. Aldroubi, M. Unser (Eds.), CRC Press, Boca Raton, FL, 527–546.
15. W. O’Grady, M. Dobrovolsky, M. Arnoff. (1993). *Contemporary Linguistics, An Introduction*. St. Martins Press, New York.
16. J.S. Walker and G.W. Don. (2006). Music: a time-frequency approach. Submitted. Available at <http://www.uwec.edu/walkerjs/media/TFAM.pdf>
17. J.F. Alm and J.S. Walker. (2002). Time-frequency analysis of musical instruments. *SIAM Review*, Vol. 44, 457–476.
18. L.M. Smith (2000). A multiresolution time-frequency analysis and interpretation of musical rhythm. Thesis, University of Western Australia.
19. L.M. Smith’s webpage: <http://staff.science.uva.nl/~lsmith/>
20. W. Sethares. (2007). *Rhythm and Transforms*. Springer, New York, NY.
21. W. Sethares. (2007). Rhythm and Transforms. An extended abstract of his plenary address to the Mathematics and Computation in Music conference in Berlin, May 18, 2007. Available at

<http://www.mcm2007.info/pdf/fri1-sethares.pdf>

22. X. Cheng, J.V. Hart, J.S. Walker. (2007). Time-frequency analysis of musical rhythm. Preprint. Available at
<http://www.uwec.edu/walkerjs/media/TFAMR.pdf>
23. D.F. Walnut. (2002). *An Introduction to Wavelet Analysis*. Birkhäuser, Boston, MA.
24. E. Hernandez, G. Weiss. (1996). *A First Course on Wavelets*. CRC Press, Boca Raton, FL.
25. C.K. Chui (Ed.) (1992). *Wavelets: a tutorial in theory and applications*. Academic Press, Boston, MA.
26. H.S. Malvar, D.H. Staelin. (1989). The LOT: transform coding without blocking effects. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, 553-559.
27. T.Q. Nguyen. (1992). A class of generalized cosine-modulated filter bank. *Proceedings 1992 IEEE International Symposium on Circuits and Systems*, Vol. 2, 943-946.
28. R.L. de Queiroz, T.Q. Nguyen, K.R. Rao. (1996). The GenLOT: generalized linear-phase lapped orthogonal transform. *IEEE Transactions on Signal Processing*, Vol. 44, 497-507.
29. S. Orintara, P. Heller, T. Tran, and T. Nguyen. (2001). Lattice structure for regular paraunitary linear-phase filterbanks and M-band orthogonal symmetric wavelets. *IEEE Transactions on Signal Processing*, Vol. 49, 2659-2672.
30. Y.-J. Chen, S. Orintara, K. Amaratunga. (2005). Dyadic-based factorization for regular paraunitary filter banks and M-band orthogonal wavelets with structural vanishing moments. *IEEE Transactions on Signal Processing*, Vol. 53, 193-207.
31. R. Coifman and Y. Zeevi, Eds. (1988). *Signal and Image Representation in Combined Spaces*. Wavelet Analysis and Applications, Vol. 7, Academic Press, Boston, MA.
32. I. Daubechies. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA.
33. H. Feichtinger and T. Strohmer, Eds. (1998). *Gabor Analysis and Algorithms*. Birkhäuser, Boston, MA.
34. H. Feichtinger and T. Strohmer, Eds. (2002). *Advances in Gabor Analysis*. Birkhäuser, Boston, MA.
35. R. Young. (1980). *An Introduction to Nonharmonic Fourier Series*. Academic Press, New York, NY.
36. P.G. Casazza. (2000). The Art of Frame Theory. *Taiwanese J. of Mathematics*, Vol. 4, 129-201.
37. B.D. Johnson. (2002). *Wavelets: generalized quasi-affine and oversampled-affine frames*. Thesis, Washington University in St. Louis.
38. D. Donoho and A.G. Flesia. (2001). Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics? *Network Computations in Neural Systems*, Vol. 12, 371-393.



- 39. D. Donoho and E. Candes. (2005). Continuous Curvelet Transform I: Resolution of Wavefront Set. *Applied and Computational Harmonic Analysis*, Vol. 19, 162–197.
- 40. D. Donoho and E. Candes. (2005). Continuous Curvelet Transform II: Discretization and Frames. *Appl. Comput. Harmon. Anal.* Vol. 19, 198–222.
- 41. Curvelet website: <http://www.curvelet.org/papers.html>
- 42. A. Khodakovsky, P. Schröder, and W. Sweldens. (2000). Progressive Geometry Compression. *SIGGRAPH 2000*, 271–278.
- 43. I. Guskov, W. Sweldens, and P. Schröder. (1999). Multiresolution Signal Processing for Meshes. *SIGGRAPH 1999*, 325–334.
- 44. W. Swelden’s papers: netlib.bell-labs.com/cm/ms/who/wim/papers/
- 45. L. Ying, L. Demanet, E. J. Candes. (2005). 3D Discrete Curvelet Transform. *Proceedings of SPIE—Volume 5914, Wavelets XI*, M. Papadakis, A.F. Laine, M.A. Unser (Eds.).
- 46. UCSD Video Processing page: <http://videoprocessing.ucsd.edu/>

