and

$$P(1/2) = 0, \quad \frac{dP}{d\theta}(1/2) = 0. \tag{5.70}$$

The equations in (5.69) correspond to Equations (3.33a) through (3.33c), while the equations in (5.70) correspond to Equations (3.32a) and (3.32b).

5.7 Gabor transforms and spectrograms

Our treatment of frequency analysis would not be complete without discussing the most up-to-date methods involving *dynamic* spectral analysis, analyzing how the spectrum of a signal evolves in time. This analysis is carried out with *Gabor transforms,* which are close relatives of wavelet transforms. Gabor transforms are widely employed in the fields of speech processing, communications theory, audio design, and music (where they are also referred to as either *Short Time Fourier Transforms* or *sonograms.*) In this section we provide the basic mathematical description of Gabor transforms, while in subsequent sections we illustrate their application to musical analysis, musical synthesis, and denoising.

Gabor transforms are particularly useful in analyzing musical signals. To motivate their definition we cannot do better than to quote from one of the founders of the theory, Jean Ville:

> If we consider a passage [of music] containing several measures (which is the least that is needed) and if a note, *la* for example, appears once in the passage, harmonic [Fourier] analysis will give us the corresponding frequency with a certain amplitude and a certain phase, without localizing the *la* in time. But it is obvious that there are moments during the passage when one does not hear the *la.* The [Fourier] representation is nevertheless mathematically correct because the phases of the notes near the *la* are arranged so as to destroy this note through interference when it is not heard and to reinforce it, also through interference, when it is heard; but if there is in this idea a cleverness that speaks well for mathematical analysis, one must not ignore the fact that it is also a distortion of reality; indeed when the *la* is not heard, the true reason is that *la* is not emitted. Thus it is desirable to look for a mixed definition of a signal of the sort advocated by Gabor: at each instance, a certain number of frequencies are present, giving volume and timbre to the sound as it is heard; each frequency is associated with a certain partition of time that defines the intervals during which the corresponding note is emitted.[3]

As an example of what Ville means, consider the signal defined by taking 8192 uniform samples of the following function

$$g(t) = e^{-400(t-0.2)^2} \sin 1024\pi t + e^{-400(t-0.5)^2} \cos 2048\pi t + e^{-400(t-0.7)^2} (\sin 512\pi t - \cos 3072\pi t)$$

[3] From [2], page 63.

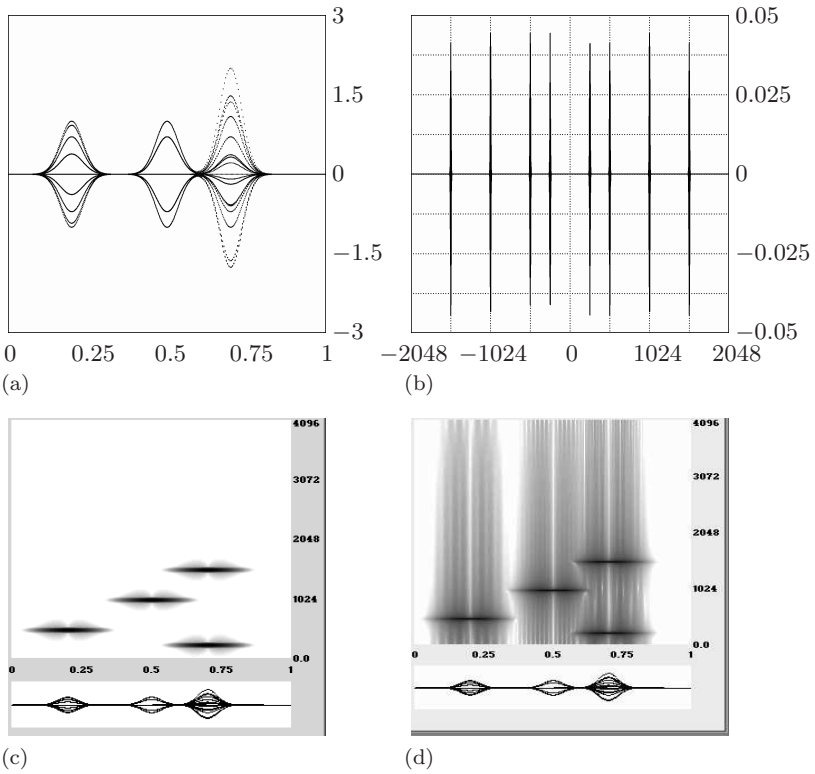


FIGURE 5.9
(a) Test signal. (b) DFT of test signal. (c) Blackman windowed spectrogram of test signal; time is in seconds along the horizontal, frequency is in Hz along the vertical. (d) Boxcar windowed spectrogram of test signal. The Blackman window produces a good spectrogram, while the Boxcar window produces a poor spectrogram.

over the time interval $[0, 1)$. When played over a computer sound system, this signal sounds like three flute-like notes. We have graphed this signal in Figure 5.9(a). The exponentials produce the "bump" shaped portions of this graph as they decrease rapidly to 0 when t moves away from 0.2, 0.5, or 0.7. The sines and cosines that multiply each exponential produce oscillations of frequency 512, 1024, 256, and 1536. In Figure 5.9(b) we show the DFT of this signal. Notice that these frequencies are identified by the positions of the spikes in the DFT. As alluded to by Ville, however, it is quite difficult to infer any information about the time locations of the emitted notes from the DFT. In contrast, consider Figure 5.9(c). In that figure, we have graphed a spectrogram (the magnitudes-squared of a Gabor transform) as advocated by Ville. The dark horizontal bars lie above the notes in the signal and at heights corresponding exactly to the frequencies of those notes. This spectrogram provides a dynamic time-frequency portrait of this musical signal.

We now give a precise definition of Gabor transforms and spectrograms. Suppose that our signal $\mathbf{f}$ consists of discrete samples of an analog signal $g(t)$. That is $f_k = g(t_k)$ for all time values t_k, which we take to be uniformly separated by the constant difference $\Delta t = t_{k+1} - t_k$. A *Gabor transform* of $\{g(t_k)\}$ is defined in two steps. For the first step, we multiply $\{g(t_k)\}$ by a sequence of shifted *window* functions $\{w(t_k - \tau_m)\}$, producing a sequence of time localized subsignals

$$\{g(t_k)w(t_k - \tau_m)\}_{m=1}^{M}.$$

Uniformly spaced time values $\{\tau_m\}_{m=1}^{M}$ are used for the shifts. The supports for the windows $\{w(t_k - \tau_m)\}$ are all of finite extent and overlap each other. See Figure 5.10. Two common window functions are Hanning and Blackman windows. The *Hanning window* w is defined by

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

where λ is a parameter that gives the length of the support of the Hanning window. See Figure 5.11(a). The *Blackman window* w is defined by

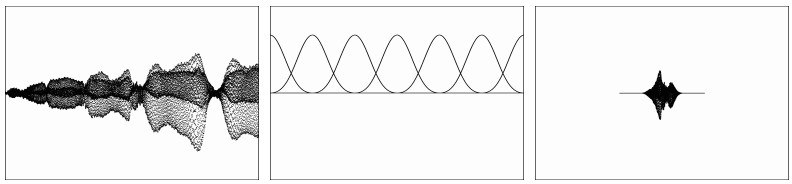
$$w(t) = \begin{cases} 0.42 + 0.5\cos(2\pi t/\lambda) + 0.08\cos(4\pi t/\lambda) & \text{for } |t| \le \lambda/2 \\ 0 & \text{for } |t| > \lambda/2 \end{cases}$$

where λ again equals the length of the support of the window. See Figure 5.11(b).

Now, for the second step. Because each subsignal $\{g(t_k)w(t_k - \tau_m)\}$ has a finite support, we can apply DFTs to them *over the length of their supports.* See Figure 5.10(c). Denoting these DFTs by $\mathcal{F}$, we then have the Gabor transform of $\{g(t_k)\}$:

$$\{\mathcal{F}\{g(t_k)w(t_k - \tau_m)\}\}_{m=1}^{M}. \tag{5.71}$$

These succession of shifted window functions $\{w(t_k - \tau_m)\}$ provide the *partitioning of time* referred to by Ville, and the DFTs provide the frequency analysis of the signal relative to this partition.



(a) (b) (c)

FIGURE 5.10
(a) Signal. (b) Succession of shifted window functions. (c) Signal multiplied by middle window in (b). A DFT can now be applied to this signal.

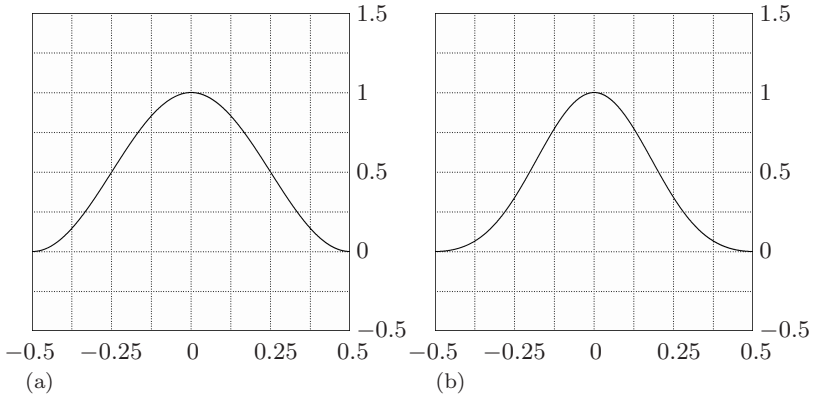


FIGURE 5.11
(a) Hanning window, $\lambda = 1$. (b) Blackman window, $\lambda = 1$.

When displaying a Gabor transform, it is standard practice to display a plot of its magnitude-squared values, with time along the horizontal axis, frequency along the vertical axis, and darker pixels representing higher square-magnitudes. We shall refer to such a plot as a *spectrogram.* The spectrogram in Figure 5.9(c) was obtained using a Blackman window. Spectrograms for other Blackman windowed Gabor transforms are given in the next section in our analysis of musical signals.

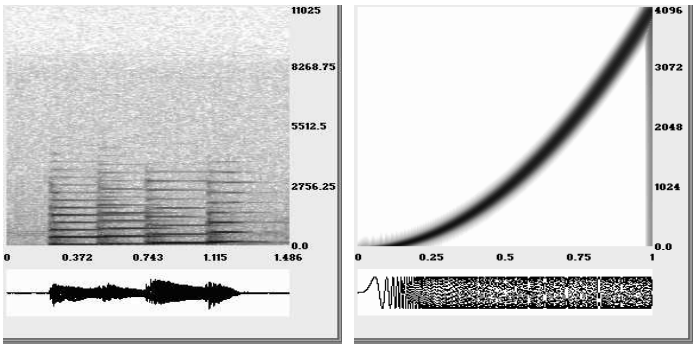
The reader may wonder about using a window that damps down to 0 at the ends, like the Blackman or Hanning window. Why not just use a rectangular window? For example, the window

$$w(t) = \begin{cases} 1 & \text{for } |t| \leq \lambda/2 \\ 0 & \text{for } |t| > \lambda/2 \end{cases}$$

which is called a *Boxcar* window (or rectangular window). The problem with using a Boxcar window is that its jump discontinuities at $\pm\lambda/2$ cause unacceptable distortions when multiplying the signal. For example, in Figure 5.9(d) we show a Boxcar windowed spectrogram of the test signal considered above. It is easy to see that the Blackman windowed spectrogram in Figure 5.9(c) provides a much better time-frequency description of the signal. The Boxcar windowed spectrogram shows far too much distortion. As seen, for example, in the long vertical swatches of gray pixels arising from the artificial discontinuities introduced by multiplying by the Boxcar window.

5.8 Musical analysis

Having introduced the basic definitions of Gabor transform theory in the previous section, we now turn to a beautiful application of it to the field of musical theory. We have chosen to analyze music in some detail because it



(a) (b)

FIGURE 5.12
(a) Spectrogram of several piano notes. (b) Spectrogram of artificial chirp.

lies at the intersection of art and science, and we hope to illustrate how these two domains can enhance rather than oppose each other.

A very succinct summary of the role of Gabor transforms in musical analysis is given by Monika Dörfler in her thesis:

> Diagrams resulting from time-frequency analysis ... can even be interpreted as a generalized musical notation.[4]

As an elaboration of this idea, we offer the following principle:

Multiresolution Principle. *Music is a patterning of sound in the time-frequency plane. Analyze music by looking for repetition of patterns of time-frequency structures over multiple time scales, and multiple resolution levels in the time-frequency plane.*

A simple illustration of this Multiresolution Principle is the structure of notes played on a piano. In Figure 5.12(a) we show a Blackman-windowed spectrogram of a succession of four piano notes. The division of the spectrogram into four sections (the four notes) along the time axis is evident in this figure. For each note, there are horizontal bars in the spectrogram that occur at integral multiples of a base frequency. For instance, in Figure 5.13(a) we show a graph of the magnitudes of a vertical slice of the Gabor transform values at a particular time, $t = 0.6$, corresponding to a point lying near the middle of the second note. This graph is the spectrum of the second piano note at $t = 0.6$. Notice that the main spikes in this spectrum occur at integral multiples of the base frequency 345 Hz. This base frequency is called the *fundamental* and the integral multiples of it are the *overtones*. The *first overtone* equals 2×345, the *second overtone* equals 3×345, and so on. Similarly, in Figure 5.13(b) we show the magnitudes of the Gabor transform at $t = 1.115$, which provide the spectrum for the fourth piano note at this new time value. For this spectrum,

[4]From [3], page xii.

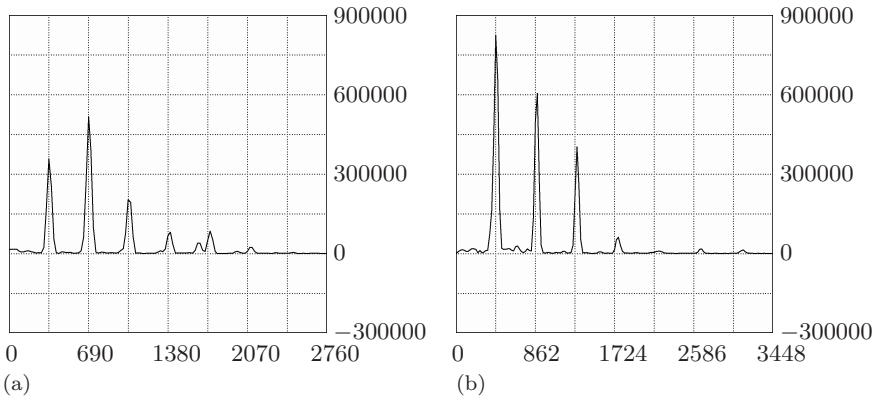


FIGURE 5.13
(a) Spectrum of piano note, $t = 0.6$. The fundamental is 345 Hz. (b) Spectrum of piano note, $t = 1.115$. The fundamental is 431 Hz.

the fundamental is 431 Hz, a higher pitch than the first spectrum. This patterning of structures in the time-frequency plane, horizontal bars at integral multiples of fundamental frequencies, is a basic feature of the tones produced by musical instruments, and has been given as an explanation of the pleasing effect of the notes produced by these instruments.[5]

For a second example, we look at a non-musical signal. In Figure 5.12(b) we show a spectrogram for an artificial signal known as a *chirp.* This chirp signal was generated from 8192 uniformly spaced samples $\{g(t_k)\}$ of the function

$$g(t) = \sin[8192(\pi/3)t^3]$$

over the interval $[0, 1]$. Chirp signals are used in Doppler radar tracking. Somewhat similar sounds are used by bats for echo navigation. When played on a computer sound system this chirp signal produces a kind of "alarm" sound of sharply rising pitch that most people would not classify as musical. Notice that there is very little if any "repetition of time-frequency structures at multiple time scales" as called for by our Multiresolution Principle.

5.8.1 Analysis of Stravinsky's Firebird Suite

We now turn to a more complex illustration of our Multiresolution Principle: an analysis of the famous ending passage of Stravinsky's *Firebird Suite.* Our discussion is based on one of the case studies in the article, "Music: a time-frequency approach" [4].

[5]The similarity of these horizontal bands of overtones to similar banding in the time-frequency structure of speech (see Figure 6.6 and its discussion in section 6.4) is striking. Musical instruments have long been regarded as amplifiers and extenders of human voice, and this may be a partial explanation for music's ability to emotionally affect us.

In Figure 5.14 we show a spectrogram of a clip from the ending passage of the *Firebird Suite,* with labelling of several important time-frequency structures. The left half of the spectrogram is relatively simple. It corresponds to a horn playing notes of the main theme for the passage with a faint string background of constant tone. We have marked the parts of the spectrogram corresponding to these musical structures. The constant tonal background is represented by long line segments marked $\mathbf{B}$ and $\mathbf{B}_1$ which represent the fundamental and first overtone of the constant tonal background. The horn notes are represented by the structure $\mathbf{T}$ for the fundamentals, and structure $\mathbf{T}_1$ above it for the first overtones. Higher overtones of the horn notes are also visible above these two structures.

The right half of the spectrogram in Figure 5.14 is much more complex. It is introduced by a harp glissando, marked as structure $\mathbf{G}$. That structure consists of a series of closely spaced horizontal bars whose left endpoints trace a steeply rising curve. The notes of this glissando are repeated over a longer time scale—marked as structure $\widetilde{\mathbf{G}}$. This structure $\widetilde{\mathbf{G}}$ is a *prolongation* of the glissando $\mathbf{G}$; its relation to $\mathbf{G}$ is emphasized by faint harp notes included among the notes of $\widetilde{\mathbf{G}}$. Here we see a prime example of repetition at multiple time scales. Another example of such repetition is that, while the prolongation is played, the main theme is repeated by the string section of the orchestra at a higher pitch than the horn. This latter repetition is marked by $\mathbf{T}_1$, and above it $\mathbf{T}_2$, which comprise the fundamentals and first overtones of the string notes. Following those structures there is again a repetition of time-frequency structure: the second glissando $\mathbf{G}$ which is played by the string section of the orchestra. This second glissando introduces a repetition of the main theme at a new higher pitch played by a flute, along with orchestral accompaniment as a second prolongation of $\mathbf{G}$, a repetition (with some embellishment)

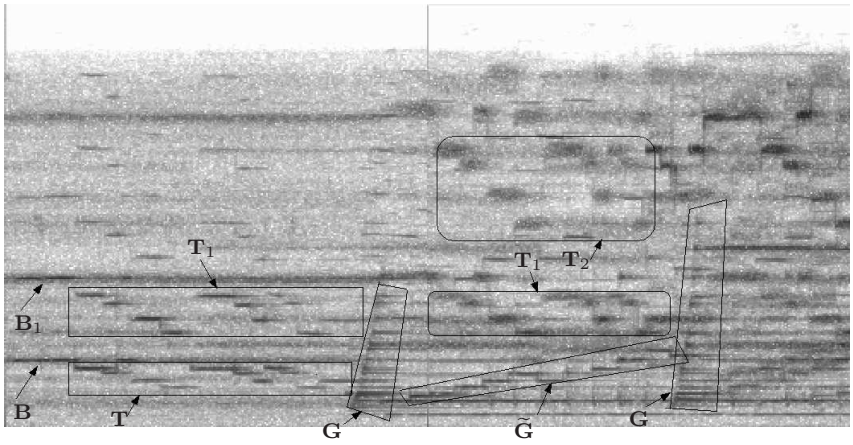
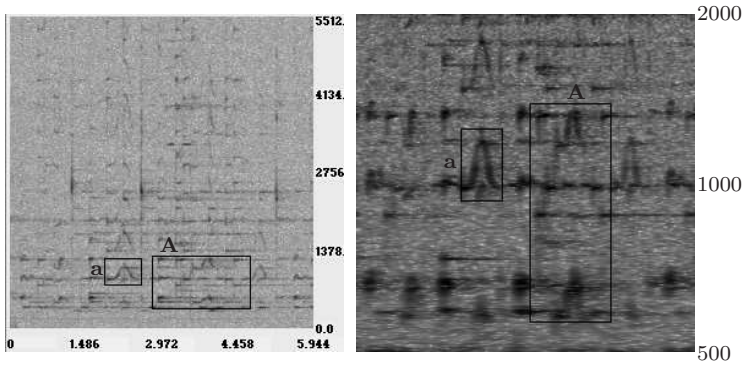


FIGURE 5.14
Spectrogram of passage from *Firebird suite* with important time-frequency structures marked (see text for explanation).



(a) (b)

FIGURE 5.15
Time-frequency analysis of a classical Chinese folk melody. (a) Spectrogram. (b) Zooming in on two octaves of the frequency range of (a). The marked time-frequency structures are explained in the text.

of the prolongation $\widetilde{\mathbf{G}}$. We have not marked these latter structures in the spectrogram, but they should be clearly visible to the reader.[6]

We encourage the reader to play the recording of the ending of the *Firebird Suite* which we have just analyzed. It is available as `firebird_clip2.wav` at the FAWAV website. An excellent way to play it is with the freeware AUDACITY, which can be downloaded from the website in [5]. AUDACITY allows you to trace out the spectrogram *as the music is playing.* Doing that will confirm the details of the analysis above, as well as providing a new appreciation of the beauty of the piece.

5.8.2 Analysis of a Chinese Folk Song

To illustrate the range of applicability of our Multiresolution Principle, we provide a brief analysis of a passage of classical Chinese folk music (available as `Chinese_Folk_Music.wav` at the book's website). In Figure 5.15(a), we show its spectrogram over a frequency range of 0 to 5512 Hz, with two important time-frequency structures labeled **a** and **A**. Structure **A** is an enlarged version of structure **a** created by repeating smaller scale versions of **a**. This is a perfect example of repetition of time-frequency structures at multiple time-scales. In (b) we show a zooming in on a 2-octave range of frequencies, from 500 to 2000 Hz, of the spectrogram in (a)—here we can see more clearly the repetition of patterns.[7] These time-frequency structures closely resemble the curved time-frequency bands occurring in speech, especially lyrics in song. An illustration

[6] One final observation on this piece: notice that the repetition of the main theme, $\mathbf{T}$, $\mathbf{T}_1$, $\mathbf{T}_2$, follows a rising arc in the time-frequency plane that repeats, over a longer time-scale, the rising arc of $\widetilde{\mathbf{G}}$. This is a perfect illustration of the Multiresolution Principle (and also of the hierarchical structure of music described in the quote from Pinker on p. 164).

[7] See the *Examples & Exercises* material at the Primer website for more details on how we created Figure 5.15(b).

from the lyrics of the song *Buenos Aires* is discussed on page 188, see especially Figure 6.8.

Of course, our Multiresolution Principle can only go so far in analyzing music. It represents an essential core of more elaborate principles formulated by Ray Jackendoff and Fred Lerdahl in their classic *Generative Theory of Tonal Music* [6]. Their theory is succinctly described by Steven Pinker as follows:

> Jackendoff and Lerdahl show how melodies are formed by sequences of pitches that are organized in three different ways, all at the same time... The first representation is a grouping structure. The listener feels that groups of notes hang together in motifs, which in turn are grouped into lines or sections, which are grouped into stanzas, movements, and pieces. This hierarchical tree is similar to a phrase structure of a sentence, and when the music has lyrics the two partly line up... The second representation is a metrical structure, the repeating sequence of strong and weak beats that we count off as "ONE-two-THREE-four." The overall pattern is summed up in musical notation as the time signature... The third representation is a reductional structure. It dissects the melody into essential parts and ornaments. The ornaments are stripped off and the essential parts further dissected into even more essential parts and ornaments on them... we sense it when we recognize variations of a piece in classical music or jazz. The skeleton of the melody is conserved while the ornaments differ from variation to variation.[8]

The similarity between the first and third representations described by Pinker and wavelet MRA is striking. It is why we referred to Gabor transforms as close relatives of wavelet transforms. Gabor transforms and spectrograms are used, together with the three representations described by Pinker, to provide a deeper analysis of musical structure in reference [4]. That article also extends Jackendoff and Lerdahl's theory to rhythmic aspects of music as well (we shall discuss this point further in Section 6.5).

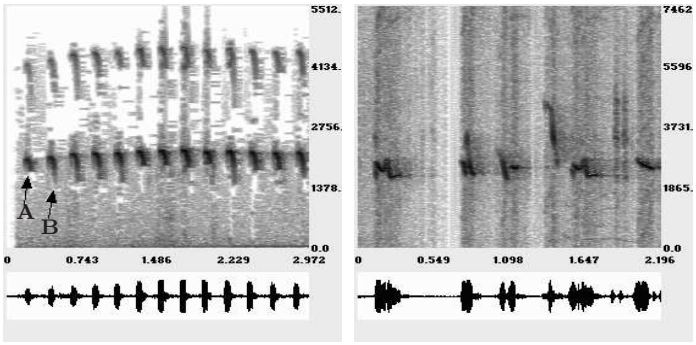
We conclude our treatment of musical analysis with a couple of non-human examples of music from the realm of bird song. Bird song has long been recognized for its musical qualities. In Figure 5.16 we show spectrograms of the songs of an oriole and an osprey. The song of the osprey illustrates our Multiresolution Principle very well. In fact, there seems to be a repetition of two basic structures, labeled **A** and **B** (along with overtones) in Figure 5.16(a). The osprey's song can be transcribed as

A B B B B A A A A B A A A

which reminds us of rhyming patterns in poetry.

The oriole song's spectrogram in Figure 5.16(b) is more richly structured. In contemplating it, we recall Dr. Dörfler's remark about *generalized musical*

[8] From [7], pages 532–533.



(a) Osprey (b) Oriole

FIGURE 5.16
Spectrograms of bird songs.

notation. We leave it as an exercise for the reader to break up this spectrogram into component chirps using our Multiresolution Principle.

Both of these bird songs are available at the book's website [8]. The recording of the oriole's song has intermittent noise that interferes with the clarity of the individual chirps in the song. We will show how to denoise this recording in Section 5.10. But first we need to describe how Gabor transforms are inverted.

5.9 Inverting Gabor transforms

In this section we describe how Gabor transforms can be inverted, and how this inversion can be used for musical synthesis. In the next section we shall use inversion of Gabor transforms for denoising audio.

In order to perform inversion, we shall assume that there are two positive constants A and B such that the window function w satisfies

$$A \leq \sum_{m=1}^{M} w^2(t_k - \tau_m) \leq B \tag{5.72}$$

for all time values t_k. These two inequalities are called the *frame conditions* for the window function. For all of the windows employed by FAWAV these frame conditions do hold. The second inequality implies that the process of computing a Gabor transform with the window function w is numerically stable (does not cause computing overflow) whenever B is not too large. The first inequality is needed to ensure that inversion is also numerically stable.

We now show how to invert a Gabor transform

$$\{\mathcal{F}\{g(t_k)w(t_k - \tau_m)\}\}_{m=1}^{M}$$

of a signal $\{g(t_k)\}$. By applying DFT-inverses we obtain a set of M subsignals

$$\{\{g(t_k)w(t_k - \tau_m)\}\}_{m=1}^{M}.$$

For each m, we then multiply the m^{th} subsignal by $\{w(t_k - \tau_m)\}$ and sum over m, obtaining

$$\{\sum_{m=1}^{M} g(t_k)w^2(t_k - \tau_m)\} = \{g(t_k) \sum_{m=1}^{M} w^2(t_k - \tau_m)\}.$$

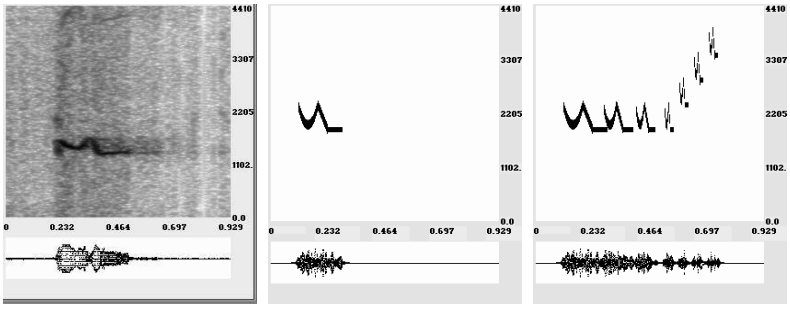
Multiplying the right side of this last equation by the values

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

which by the frame condition (5.72) are no larger than A^{-1}, we obtain our original signal values $\{g(t_k)\}$. Thus, we have inverted our Gabor transform.

An interesting application of inversion is to the synthesis of musical passages. We now provide one illustration of such a synthesis. In Figure 5.17(a) we show a spectrogram of an oriole chirp, extracted from the beginning of the oriole's song that we discussed in the previous section. We used this spectrogram as our model for synthesizing an artificial bird chirp. That synthesis was begun by plotting the function described in the formula file `synthetic_oriole_whistle.uf2` using the plotting procedure of the spectrogram tool in FAWAV. This produced a parabolic segment, connected to a steeply sloping line segment, connected in turn to a horizontal line segment. See Figure 5.17(b). We then applied Gabor transform inversion to produce an audio signal, which we saved as `synthetic_bird_chirp.wav` at [8]. When played, it does indeed sound like a bird chirp.

As an isolated sound there is not much musicality to a single bird chirp. That follows from our Multiresolution Principle for musical structure in the time-frequency plane. There is simply not enough of a pattern in the time-frequency structure of this artificial bird chirp for us to perceive it as music. To create a more musical signal, an artificial bird song, we used this artificial bird chirp as a template for creating a pattern of repeating structures in the time-frequency plane. See Fig. 5.17(c). To produce this graph



(a) Oriole chirp (b) Synthetic bird chirp (c) Synthetic bird song

FIGURE 5.17
Spectrograms illustrating musical synthesis.

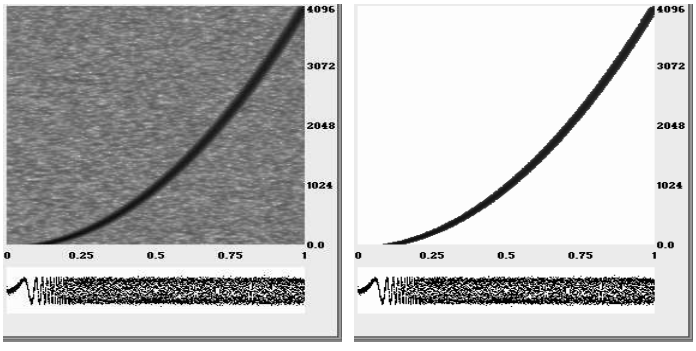
we repeated the time-frequency structure in Fig. 5.17(b) twice more at successively shrunken time-scales, followed by four more repetitions at a much shorter time-scale and translated upward in frequency. Our Multiresolution Principle predicts that a signal with such a time-frequency graph will sound musical. Indeed, the signal obtained by Gabor transform inversion (saved as `a_synthetic_bird_song.wav` at [8]) does sound like a bird song.

Of course, synthesizing bird song is only the beginning; although we would like to emphasize that bird song has inspired composers in the past. In the Notes and References section, we provide more details on the relation between musical composition and bird song, including pointers to some lovely avian-inspired music on the Internet. We also supply a reference to software for synthesizing beautiful electronic music from Gabor transforms.

5.10 Gabor transforms and denoising

An important application of inversion of Gabor transforms is noise removal. The basic idea is similar to wavelet denoising: apply thresholding to remove noisy transform values while preserving signal-dominated transform values. In this section we provide a brief introduction to this vast field. We discuss a couple of illustrative examples, and then turn to some examples of denoising real audio.

The essential idea behind threshold denoising of Gabor transforms can be illustrated with spectrograms. For example, in Figure 5.18(a) we show a spectrogram for a Gabor transform of the artificial chirp signal after random Gaussian noise has been added. Compare it with the one for the uncontaminated signal shown in Figure 5.12(b). As you can see, the noise appears as a background of gray pixels and there is a parabolic arc of signal-dominated transform values that clearly stands out from this background. In Figure 5.18(b) we show the spectrogram for the Gabor transform obtained by the thresholding process referenced below. Applying our inverse process to this latter



(a) (b)

FIGURE 5.18
(a) Noisy chirp spectrogram. (b) Thresholded spectrogram [c.f. Figure 5.12(b)].