SIAM REVIEW
Vol. 44, No. 3, pp. 457–476

Time-Frequency Analysis of Musical Instruments*

Jeremy F. Alm†
James S. Walker‡

Abstract. This paper describes several approaches to analyzing the frequency, or pitch, content of the sounds produced by musical instruments. The classic method, using Fourier analysis, identifies fundamentals and overtones of individual notes. A second method, using spectrograms, analyzes the changes in fundamentals and overtones over time as several notes are played. Spectrograms produce a time-frequency description of a musical passage. A third method, using scalograms, produces far more detailed time-frequency descriptions within the region of the time-frequency plane typically occupied by musical sounds. Scalograms allow one to zoom in on selected regions of the time-frequency plane in a more flexible manner than is possible with spectrograms, and they have a natural interpretation in terms of a musical scale.

All three of these techniques will be employed in analyzing music played on a piano, a flute, and a guitar. The two time-frequency methods, spectrograms and scalograms, will be shown to extend the classic Fourier approach, providing time-frequency portraits of the sounds produced by these instruments. Among other advantages, these time-frequency portraits seem to correlate well with our perceptions of the sounds produced by these instruments and of the differences between each instrument.

There are many additional applications of time-frequency methods, such as compression of audio and resolution of closely spaced spectral lines in spectroscopy. Brief discussions of these additional applications are included in the paper.

Key words. time-frequency analysis, spectrogram, scalogram, continuous wavelet transform, fast Fourier transform, Fourier series

AMS subject classifications. 65T50, 65T60, 42A01, 42A16, 94A12

PII. S0036144500382283

1. Introduction.

In this paper we shall describe several different approaches to analyzing the sound of musical instruments, ranging from the classic method of Fourier to the most up-to-date methods of dynamic spectra and wavelets. These methods will be applied to sounds produced from a piano, a flute, and a guitar. Although, with the possible exception of the spectroscopy example, the results in this paper are not new, nevertheless we hope that it will provide an enlightening discussion of the mathematical analysis of music.

The contents of the paper are as follows. In section 2 we review basic notions of pitch and frequency. The musical concepts here are *fundamentals* and *overtones.* Mathematically, these concepts are described via Fourier coefficients, and their role

*Received by the editors December 8, 2000; accepted for publication (in revised form) February 6, 2002; published electronically August 1, 2002.
http://www.siam.org/journals/sirev/44-3/38228.html

†Department of Mathematics, Iowa State University, Ames, IA 50011 (almjf@iastate.edu).
‡Department of Mathematics, University of Wisconsin–Eau Claire, Eau Claire, WI 54702–4004 (walkerjs@uwec.edu).

in producing sounds is modeled via Fourier series. Although Fourier series are an essential tool, they do have limitations; in particular, they are not effective at capturing abrupt changes in the frequency content of sounds. These abrupt changes occur, for instance, in transitions between individual notes. In section 3 we describe a modern method of *time-frequency analysis,* known as a spectrogram, which better handles changes in frequency content over time. Spectrograms provide a time-frequency portrait of musical sounds, gracefully handling the problem of describing quick transitions between notes. They provide a type of "fingerprint" of sounds from various instruments. These fingerprints allow us to distinguish one instrument from another. While spectrograms are a fine tool for many situations, they are not closely correlated with the frequencies (pitches) typically found on musical scales, and there are cases where this leads to problems. In section 4, we describe a method of time-frequency analysis, known as scalograms, which does correlate well with musical scale frequencies. Scalograms yield a powerful new approach, based on the mathematical theory of wavelets, which can solve problems lying beyond the scope of either Fourier series or spectrograms.

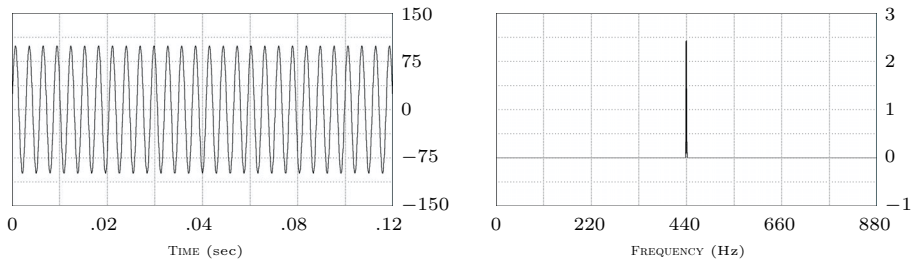
The figures in this paper were all generated with the software FAWAV, which can be downloaded for free from the website listed in [24]. That website also contains copies of the sound recordings of the instruments we discuss. Using FAWAV, these sound recordings can be both played and analyzed with the techniques described in this paper.

2. Pitch and Frequency—Fourier Series and FFTs. In this section we review some basic concepts in analyzing music: the concepts of pitch and frequency and their relation to Fourier coefficients. Since there are innumerable references for these concepts (e.g., [25], [26], [16], [18]), we shall aim for succinctness rather than thoroughness.

2.1. The Connection between Pitch and Frequency. As its name implies, a tuning fork is used for tuning the pitch of notes from musical instruments. The sound from a tuning fork can be recorded with an oscilloscope attached to a microphone. This will produce a graph similar to the one shown in Figure 2.1(a). The graph in Figure 2.1(a) was created by plotting the function $100 \sin 2\pi\nu t$, a sinusoid of frequency $\nu = 440$ cycles/sec (Hz). Played over the computer's sound system,[1] this finite segment of a sinusoid produces a tone identical to a tuning fork with a pitch of A_4 on the *well-tempered scale.* A pure tone having a single pitch is thus associated with a single frequency, in this case 440 Hz. In Figure 2.1(b) we show a computer calculation of the *Fourier spectrum* of the sinusoid in (a); the single peak at 440 Hz is clearly evident. The formulas used to generate this Fourier spectrum will be discussed below. For now, it is important to take note of its precise identification of the single frequency present in the pure tone.

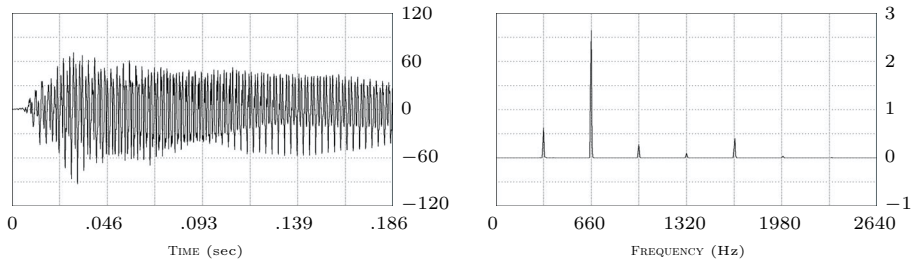
Unlike tuning forks, sounds from musical instruments are time-evolving superpositions of several pure tones, or sinusoidal waves. For example, in Figure 2.2(b) we show the Fourier spectrum of the piano note shown in Figure 2.2(a). In this spectrum, there are peaks located at the (approximate) frequencies 330 Hz, 660 Hz, 990 Hz, 1320 Hz, and 1620 Hz. Notice that these frequencies are all integral multiples of the smallest frequency, 330 Hz. This base frequency, 330 Hz, is called the *fundamental.* In this case, it is nearly equal to the standard frequency of 329.628 Hz for the E_4 note [16, p. 48]. The integral multiples of this fundamental are called *overtones.* In this case,

[1] Some further details on how computers play sound signals will be provided below.



(a) (b)

Fig. 2.1 *Fourier analysis of a pure tone.* (a) *Graph of a finite segment of a pure tone,* 440 *Hz.* (b) *Computer-calculated Fourier spectrum.*



(a) Piano note, E_4 (b) Spectrum of piano note

Fig. 2.2 *Fourier analysis of the piano note* E_4 *(E above middle C). (Note: The vertical scales of all spectra shown in this paper have been normalized to the same range.)*

the first overtone is 660 Hz, the second overtone is 990 Hz, and so on. We now turn to the mathematical theory that captures these musical notions of fundamentals and overtones, the theory of Fourier series.

2.2. Fourier Series. The classic mathematical theory for describing musical notes is that of Fourier series. Given a sound signal $f(t)$ (such as a musical note or chord) defined on the interval $[0, \Omega]$, its *Fourier series* is

$$c_0 + \sum_{n=1}^{\infty} \left\{ a_n \cos \frac{2\pi nt}{\Omega} + b_n \sin \frac{2\pi nt}{\Omega} \right\} \tag{2.1}$$

with *Fourier coefficients* c_0, a_n, b_n defined by

$$\begin{aligned} c_0 &= \frac{1}{\Omega} \int_0^{\Omega} f(t)\, dt, \\ a_n &= \frac{2}{\Omega} \int_0^{\Omega} f(t) \cos \frac{2\pi nt}{\Omega}\, dt, \quad n = 1, 2, 3, \ldots, \\ b_n &= \frac{2}{\Omega} \int_0^{\Omega} f(t) \sin \frac{2\pi nt}{\Omega}\, dt, \quad n = 1, 2, 3, \ldots. \end{aligned} \tag{2.2}$$

Each term $\{a_n \cos(2\pi nt/\Omega) + b_n \sin(2\pi nt/\Omega)\}$ has a fundamental period of Ω/n, and hence a frequency in time of n/Ω. Thus (2.1) is a superposition of waves of

frequencies $1/\Omega, 2/\Omega, 3/\Omega, \ldots$. (The constant c_0 represents a constant background level, corresponding to a constant air pressure level; see section 2.3.) These frequencies in the Fourier series are thus *integral multiples of a fundamental frequency* $1/\Omega$. It is well known that for a differentiable function $f(t)$, the Fourier series for $f(t)$ will converge to it at each point in $(0, \Omega)$. Furthermore, if $f(t)$ is square-integrable, then its Fourier series converges to it in mean-square.[2] This latter result will certainly apply to musical sound signals $f(t)$.

It is more convenient to rewrite (2.1) and (2.2) using complex exponentials. Via Euler's formulas,

$$e^{i\theta} = \cos\theta + i\sin\theta, \quad e^{-i\theta} = \cos\theta - i\sin\theta,$$
$$\cos\theta = \frac{1}{2}e^{i\theta} + \frac{1}{2}e^{-i\theta}, \quad \sin\theta = \frac{i}{2}e^{-i\theta} - \frac{i}{2}e^{i\theta},$$

it is a simple exercise in algebra to rewrite the Fourier series in (2.1) as

$$c_0 + \sum_{n=1}^{\infty} \left\{ c_n e^{i2\pi nt/\Omega} + c_{-n} e^{-i2\pi nt/\Omega} \right\} \tag{2.3}$$

with *complex Fourier coefficients*

$$c_n = \frac{1}{\Omega}\int_0^{\Omega} f(t) e^{-i2\pi nt/\Omega}\, dt, \quad n = 0, \pm 1, \pm 2, \ldots. \tag{2.4}$$

The Fourier series in (2.3) is identical, term for term, to the Fourier series in (2.1), and the complex Fourier coefficients are related to those defined in (2.2) by the formula $c_n = (a_n + ib_n)/2$. Since a sound signal f is real-valued, it follows from (2.4) that $c_{-n} = \overline{c_n}$. Consequently, the negatively indexed complex Fourier coefficients add no new significant information, since they are simply the complex conjugates of the positively indexed coefficients.

One of the most celebrated results of Fourier series is *Parseval's equality* [19, Chap. 4]:

$$\frac{1}{\Omega}\int_0^{\Omega} |f(t)|^2\, dt = |c_0|^2 + \sum_{n=1}^{\infty} \left\{ |c_n|^2 + |c_{-n}|^2 \right\}. \tag{2.5}$$

This equality has a beautiful physical interpretation. Define the *energy* of a function g over $[0, \Omega]$ to be $\int_0^{\Omega} |g(t)|^2\, dt$. Then it is easy to see that $\Omega \cdot |c_n|^2$ is the energy of the complex exponential $c_n e^{i2\pi nt/\Omega}$. Hence, (2.5) shows that *the energy of the sound signal f is equal to the sum of the energies of the complex exponentials (including the constant) that make up its Fourier series.* It is important to note that $|c_{-n}|^2 = |c_n|^2$, so Parseval's equality can also be written as

$$\frac{1}{\Omega}\int_0^{\Omega} |f(t)|^2\, dt = |c_0|^2 + \sum_{n=1}^{\infty} 2|c_n|^2. \tag{2.6}$$

[2] By convergence in mean-square, we mean that the *partial sums*

$$S_N(t) = c_0 + \sum_{n=1}^{N} \left\{ a_n \cos\frac{2\pi nt}{\Omega} + b_n \sin\frac{2\pi nt}{\Omega} \right\}$$

satisfy $\lim_{N\to\infty} \frac{1}{\Omega}\int_0^{\Omega} |f(t) - S_N(t)|^2\, dt = 0$. Further details can be found in [25, Chap. 2] or [19, Chap. 4].

Thus, the *Fourier series spectrum* $\{2|c_n|^2\}_{n\geq 1}$ completely captures the energies in the frequencies that make up the audio signal. Since $|c_0|^2$ is the energy of a constant background and is *inaudible*,[3] it is ignored here. It is this Fourier series spectrum that is calculated and displayed in Figures 2.1(b) and 2.2(b), and in every subsequent figure, it is referred to as a *spectrum.*

2.3. FFTs. Whether for good or ill, we have come to live in a digital world. The audio signals of piano notes discussed above were recorded in digital form and their Fourier spectra were computed digitally. The method of digitally computing Fourier spectra is widely referred to as the FFT (short for *fast Fourier transform*). An FFT provides an extremely efficient method for computing approximations to Fourier series coefficients; these approximations are called *DFTs* (short for *discrete Fourier transforms*). We shall briefly outline the ideas behind DFTs; for proofs and more details see [26], [4], [5], or [1].

A DFT is defined via Riemann sum approximations of the integrals in (2.4) for the Fourier coefficients. For a (large) positive integer N, let $t_k = k\Omega/N$ for $k = 0, 1, 2, \ldots, N-1$, and let $\Delta t = \Omega/N$. Then the nth Fourier coefficient c_n in (2.4) is approximated as follows:

$$\begin{aligned} c_n &\approx \frac{1}{\Omega}\sum_{k=0}^{N-1} f(t_k)e^{-i2\pi n t_k/\Omega}\,\Delta t \\ &= \frac{1}{N}\sum_{k=0}^{N-1} f(t_k)e^{-i2\pi nk/N}. \end{aligned}$$

The last quantity above is the DFT of the finite sequence of numbers $\{f(t_k)\}$. That is, we define the DFT of a sequence $\{f_k\}$ of N numbers by

$$F[n] = \frac{1}{N}\sum_{k=0}^{N-1} f_k\, e^{-i2\pi nk/N}. \tag{2.7}$$

The DFT is the sequence of numbers $\{F[n]\}$, and we see from the discussion above that the Fourier coefficients of a function f can be approximated by a DFT. In particular, the spectra $\{2|c_n|^2\}_{n\geq 1}$ shown in the figures above were obtained via DFT approximations $\{2|F[n]|^2\}_{n\geq 1}$ (where $f_k = f(t_k)$ for each k).

Besides being used for calculating DFTs, the finite set of discrete values $\{f(t_k)\}$ is also used for creating a digital sound file, playable by a computer sound system. A computer sound system is programmed to play a fixed number of equally spaced volume levels—typically either $256 = 2^8$ (8-bit sound) or $65536 = 2^{16}$ (16-bit sound). These volume levels correspond to air pressure levels that are fluctuating about a constant background level (hence there are negative as well as positive volume levels). By approximating the values of $\{f(t_k)\}$ by finite-precision numbers, the fluctuations of air pressure due to the sound signal are then describable by these volume levels. Because there are an integral number of equally spaced volume levels, they can be mapped to a computer file using integers as index values to the corresponding volume levels—these integers correspond to multiples of the computer's unit volume level. A computer's

[3] The reason that c_0 is inaudible is that our hearing responds linearly—via a resonance effect in the ear's basilar membrane—to fluctuations in air pressure caused by the sound signal. The constant c_0, since it does not fluctuate, is therefore inaudible.

sound system uses these integers to drive the speakers with time-varying volume levels, thus reconstructing a recorded sound signal.[4] For a more detailed discussion of the process just described, known technically as *analog-to-digital conversion,* see [21].

Two noteworthy properties of DFTs are: (1) they can be inverted, and (2) they preserve energy (up to a scale factor). The inversion formula for the DFT is

$$f_k = \sum_{n=0}^{N-1} F[n] e^{i2\pi nk/N} \tag{2.8}$$

and the conservation of energy property is

$$\frac{1}{N} \sum_{k=0}^{N-1} |f_k|^2 = \sum_{n=0}^{N-1} |F[n]|^2. \tag{2.9}$$

Exercise: Work out the connection between (2.9) and Parseval's equality (2.6).

Computer calculations of DFTs are done using a wide variety of algorithms that are all referred to as FFTs. Using FFTs, the computation of DFTs can now be done almost instantaneously. An important application of this rapid processing is in calculating spectrograms.

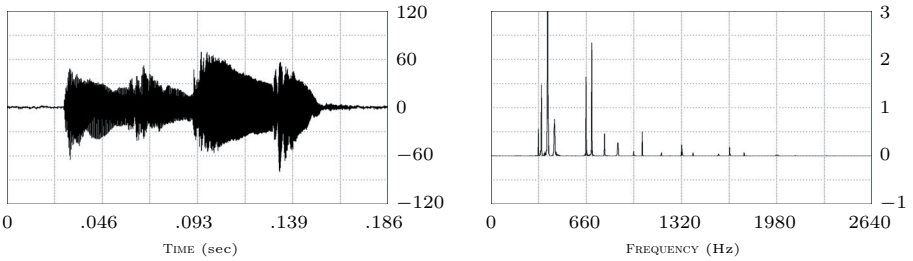
3. Time-Frequency Analysis—Spectrograms. While Fourier spectra do an excellent job of identifying the frequency content of individual notes, they are not as useful for analyzing several notes in a musical passage. For example, in Figure 3.1(a) we show a graph of a recording of a piano playing the notes E_4, F_4, G_4, and A_4. The spectrum from this musical passage is shown in Figure 3.1(b). Unlike the single note case, it is not as easy here to assign fundamentals and overtones; in fact, the spectrum in Figure 3.1(b) is a mixture of spectra from the individual notes.[5] The problem was succinctly analyzed by Ville [23], one of the founders of spectrogram analysis, as follows (translation from [14, p. 63]):

> If we consider a passage [of music] containing several measures... 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 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.... Thus it is desirable to look for a mixed definition of a signal... 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.

One way of implementing this "mixed definition of a signal" described by Ville is to compute *spectrograms,* which are a moving sequence of local spectra for the signal.

[4] Examples of such recordings can be found in the sound files available at the website [24]. Using FAWAV the reader can plot these sound files as graphs and play them as audio.

[5] There is an analogy here to absorption and emission spectra in molecular spectroscopy, the individual notes being analogous to spectra from individual atoms and the mixture of notes being analogous to compound spectra from many constituent atoms. At the end of the paper we give an example of resolving two closely spaced spectral lines, an important problem in spectroscopy.



(a) Several piano notes

(b) Spectrum of piano notes

Fig. 3.1 *Fourier analysis of a piano passage.*

In order to isolate the individual notes in the musical passage, the sound signal $f(t)$ is multiplied by a succession of time-windows (or simply *windows*): $\{w(t-\tau_m)\}$, $m = 1, 2, \ldots, M$. Each window $w(t-\tau_m)$ is equal to 1 in a time interval $(\tau_m - \epsilon, \tau_m + \epsilon)$ centered at τ_m and decreases smoothly down to 0 for $t < \tau_{m-1} + \delta$ and $t > \tau_{m+1} - \delta$. See Figure 3.2(b), where $M = 9$ windows are shown. As shown in this figure, we are assuming here that the values $\{\tau_m\}$ are separated by a uniform distance $\Delta\tau$, although this is not absolutely necessary. These windows also satisfy

$$\sum_{m=1}^{M} w(t-\tau_m) = 1 \tag{3.1}$$

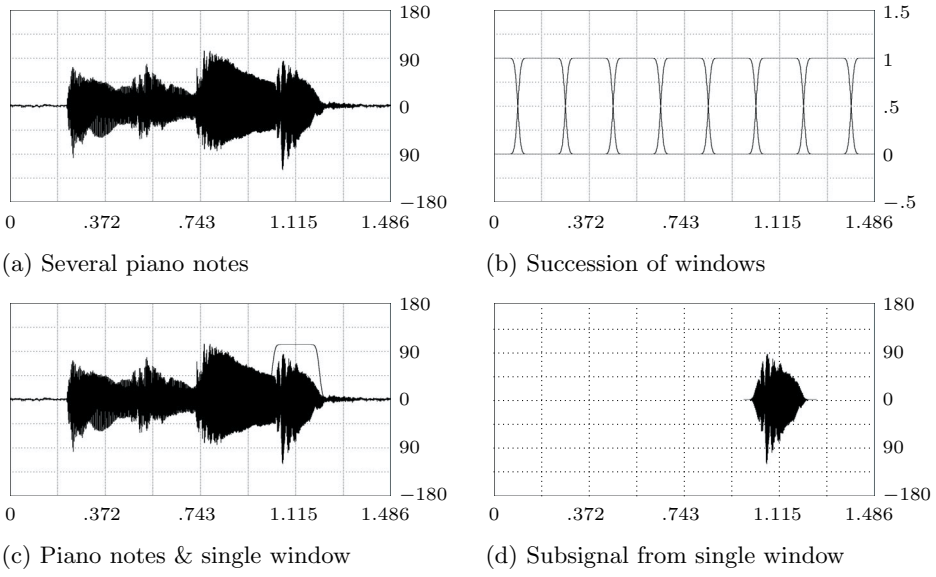
over the time interval $[0, \Omega]$. Multiplying both sides of (3.1) by $f(t)$ we see that

$$f(t) = \sum_{m=1}^{M} f(t)w(t-\tau_m),$$

so the sound signal $f(t)$ equals a sum of the subsignals $f(t)w(t-\tau_m)$, $m = 1, 2, \ldots, M$. Each subsignal $f(t)w(t-\tau_m)$ is nonzero only within the interval $[\tau_{m-1} + \delta, \tau_{m+1} - \delta]$ centered on τ_m. In Figures 3.2(c) and (d), we show the process of producing one subsignal $f(t)w(t-\tau_m)$.

Notice that in Figure 3.2(d) the subsignal $f(t)w(t-\tau_m)$ is shown as having a restricted domain, a domain of $[\tau_{m-1} + \delta, \tau_{m+1} - \delta]$. The domain of each subsignal $f(t)w(t-\tau_m)$ is restricted so that when an FFT is applied to the sequence $\{f(t_k)w(t_k - \tau_m)\}$, with points $t_k \in [\tau_{m-1}, \tau_{m+1}]$, then this FFT produces Fourier coefficients that are localized to the time interval $[\tau_{m-1} + \delta, \tau_{m+1} - \delta]$ for each m. This localization in time of Fourier coefficients constitutes the spectrogram solution of the problem of separating the spectra of the individual notes in the musical passage.

3.1. Spectrograms for a Piano and a Flute. In Figure 3.3(a) we show a spectrogram for the sequence of piano notes E_4, F_4, G_4, and A_4. The sound signal is plotted at the bottom of the figure. It consists of 2^{15} values $\{f(t_k)\}$ at equally spaced points $\{t_k\}$ on the time interval $[0, 1.486)$. Above the sound signal is a plot of the FFT spectra $\{2|F_m[n]|^2\}$, $m = 1, 2, \ldots, M$, obtained from the M subsignals $\{f(t)w(t-\tau_m)\}$. The larger values of these spectra are displayed more darkly; the white regions represent values that are near zero in magnitude. The vertical scale on this figure is a frequency scale (in Hz), and the horizontal scale is a time scale (in sec). This spectrogram thus provides a description of the sound signal in the *time-frequency plane*.



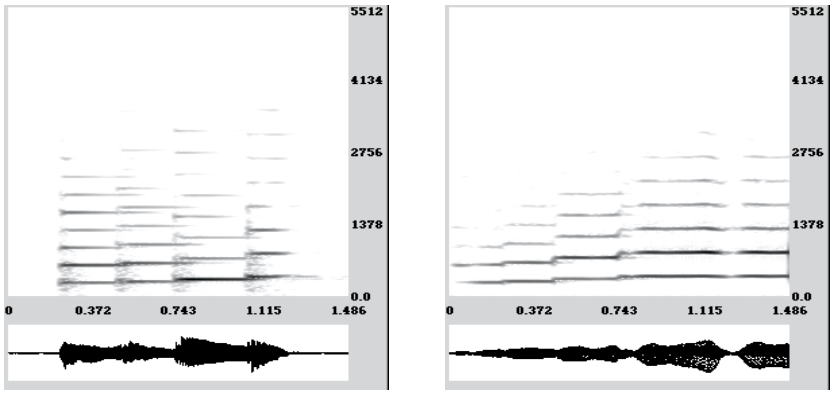
(a) Several piano notes

(b) Succession of windows

(c) Piano notes & single window

(d) Subsignal from single window

Fig. 3.2 *Components of a spectrogram.*



(a) Piano

(b) Flute

Fig. 3.3 *Two spectrograms.*

As can be seen clearly in Figure 3.3(a), the spectra for the individual notes are clearly separated in time. Similarly, in Figure 3.3(b), we can see a clear separation of the spectra for these same notes played on a flute. In both cases, the spectra for individual notes are clearly divided into groups of "line segments," lying above each note and corresponding to the fundamentals and overtones in each note.

It is interesting to compare these two spectrograms. There are clear differences between the "attack" and the "decay" of the spectral line segments for the notes played by the two instruments; these differences are visible in the two spectrograms. For the piano there is a very prominent attack—due to the striking of the piano

hammer on its strings—which is visible in the gray patches in its spectrogram near the beginning of the fundamental and first overtone line segments for each note. These gray patches (or "spectral smears") can be termed *anharmonic spectra,* since they are nonintegral multiples of the fundamental. (They arise from transient, nonlinear effects during the attack of each note.) There is also a longer decay for the piano notes—due to the slow damping down of the piano string vibrations—which is evident in the overlapping of the time intervals underlying each note's line segment. For the flute—where notes arise from a standing wave within the flute created by a gentle breath of the player and rapidly decay when this standing wave collapses at the end of the player's breath—the notes show little overlapping of spectral line segments. These line segments exhibit a much gentler attack and much more rapid decay than for the piano.[6] It is well known that, in addition to the harmonic structure of fundamentals and overtones, the precise features of attack and decay in notes are important factors in human perception of musical quality. This comparison of a piano with a flute illustrates how all of these features of musical notes can be quantitatively captured in the time-frequency portraits provided by spectrograms.

3.2. Inversion of Spectrograms. We have shown above that a spectrogram provides a useful method of *analysis*, i.e., splitting of the signal into *parts* and processing these parts, the parts being the subsignals $\{f(t)w(t-\tau_m)\}$ that are processed via FFTs. Complementing this analysis, there is also *synthesis,* whereby the signal is reconstructed from its constituent parts. This synthesis is an *inversion process* for spectrograms: recovering the original sound signal $f(t)$ from its spectrogram of FFTs $\{F_m[n]\}$, $m = 1, 2, \ldots, M$. (Actually, to be precise, the values of $f(t_k)$ at discrete points t_k are recovered.) One important application of this inversion process is to the compression of recorded music and speech.

In order to perform inversion, it is sufficient that the succession of windows $\{w(t-\tau_m)\}$ satisfies

$$A \leq \sum_{m=1}^{M} w(t-\tau_m) \leq B \tag{3.2}$$

for some positive constants A and B. The inequalities in (3.2) are a generalization of (3.1). In practice, it has been found that some windows w (such as Hamming or hanning windows [26, Chap. 4]), which satisfy (3.2), perform better for frequency identification than windows that are required to satisfy the more stringent condition (3.1).

Assuming that (3.2) holds, the largest possible constant A and the smallest possible constant B are called the *frame bounds* for the window w. We now prove that (3.2) suffices for inversion, and we assume that A and B are frame bounds. Provided B is not too large, performing the spectrogram analysis

$$\{f(t_k)\} \mapsto \{f(t_k)w(t_k-\tau_m)\}_{m=1}^{M} \mapsto \{F_m[n]\}_{m=1}^{M} \tag{3.3}$$

will be numerically stable. Furthermore, performing inverse FFTs on each of the

[6] With a flute, the player can generally only produce one note at a time (monophonic), while with a piano multiple notes can be played (polyphonic). The nonoverlapping of flute spectral line segments, and their partial overlap for a piano, are clearly in consonance with the different phonic capabilities of these two instruments. Furthermore, when multiple piano notes are played simultaneously (as in a chord, see section 4.5), there are significant problems in pitch detection, due to a large overlap of spectral line segments corresponding to different notes.

subsignal FFTs in (3.3) and adding the results together produces

$$\sum_{m=1}^{M} f(t_k)w(t_k - \tau_m) = f(t_k) \sum_{m=1}^{M} w(t_k - \tau_m). \tag{3.4}$$

Because of the inequality $A \leq \sum w(t_k - \tau_m)$, we have $1/\sum w(t_k - \tau_m) \leq A^{-1}$. Therefore, we can divide out the factor $\sum w(t_k - \tau_m)$ in (3.4) to obtain the discrete values $\{f(t_k)\}$. The constant A^{-1}, provided it is not too large, ensures the numerical stability of this division. Thus, (3.2) implies that spectrograms obtained from discrete values of a sound signal are invertible.

When an analog sound signal $f(t)$ is *band-limited*,[7] then this signal $f(t)$ can be recovered from the discrete values $\{f(t_k)\}$ as well [26, Chap. 5]. This last step is often unnecessary for digitally recorded sound because the only data available are the discrete values $\{f(t_k)\}$.

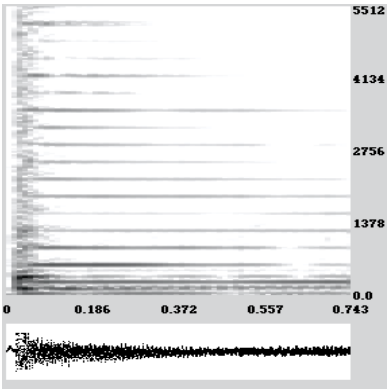
As an example of this inversion process for discrete data $\{f(t_k)\}$, if the spectrogram in Figure 3.3(a) is inverted, then the maximum error between the original sound data and the reconstructed values is less than 1.99×10^{-10} for all points. Moreover, if the reconstructed values are rounded to integers—because the original data in the computer sound file were integers—then the error is zero!

One application of spectrogram inversion is to the *compression of audio signals.* After discarding (setting to zero) all the values in a spectrogram with magnitudes below a threshold value, the inversion procedure creates an approximation of the original signal that uses significantly less data than the original signal. That is because the thresholded spectrogram can be greatly compressed by removing the large amount of redundancy created by all of the zero values arising from thresholding. In Figures 3.3(a) and (b), for example, all of the white pixels shown in the spectrograms stand for zero values and that implies considerable redundancy for those spectrograms. Some of the best results in audio compression are based on sophisticated generalizations of this spectrogram technique; these techniques are called *lapped orthogonal transform coding* or *local cosine coding* [13], [12], [28], [22]. Such audio compression techniques underlie the real-time audio players and audio download sites available on the Internet.

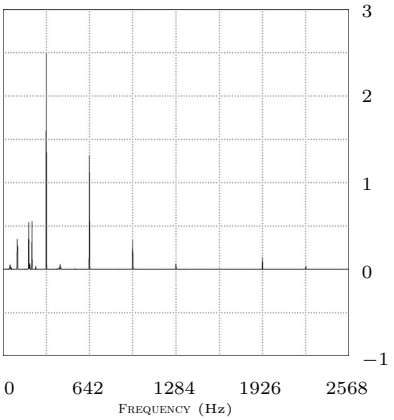
4. Musical Scales and the Time-Frequency Plane—Scalograms. Although spectrograms are profoundly useful, they do suffer from one drawback. They display frequencies on a uniform scale, whereas musical scales such as the well-tempered scale are based on a logarithmic scale for frequencies. We shall describe below how such a logarithmic scale is related to human hearing and how it leads to a new type of time-frequency analysis.

To illustrate how a uniform scale of frequencies can lead to problems, consider in Figure 4.1(a) the spectrogram of the note E_4 played on a guitar. In this spectrogram there are a number of spectral line segments crowded together at the lower end of the frequency scale. These line segments correspond to the lower frequency peaks in the Fourier spectrum for the note in Figure 4.1(b). This spectrum shows a fundamental at 321 Hz—which is close to the standard frequency for E_4 of 329.628 Hz—and overtones at 642, 963, ..., 2247 Hz. However, there are also peaks in the spectrum at lower frequencies of 104, 191, and 215 Hz. It is these frequencies that are crowded together in the spectral lines at the bottom of the spectrogram. These frequencies are (approximately) integral divisors of some of the overtones listed above, since $215 \approx 642/3$,

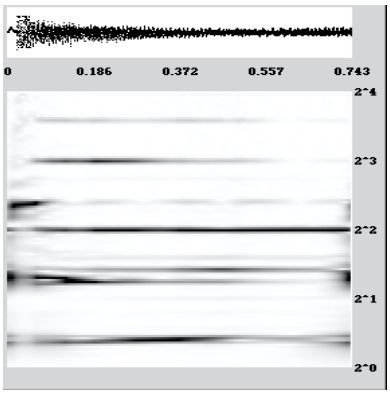
[7] A signal is called band-limited if its spectrum is zero-valued outside some finite range of frequencies.



(a) Guitar note spectrogram



(b) Spectrum



(c) Guitar note scalogram

Fig. 4.1 *Three analyses of a guitar note.*

$191 \approx 961/5$, and $104 \approx 321/3$. It may be that they are "undertones," resulting from body cavity resonances in the guitar.

In any case, what is needed is a technique of mathematically "zooming in" on these lower frequencies. The "scalogram" shown in Figure 4.1(c) provides this zooming property. As we shall explain below, the vertical scale on this scalogram consists of multiples of a base frequency of 80 Hz, viz., $80 \cdot 2^0 = 80$ Hz, $80 \cdot 2^1 = 160$ Hz, $80 \cdot 2^2 = 320$ Hz, $80 \cdot 2^3 = 640$ Hz, and $80 \cdot 2^4 = 1280$ Hz. This is a logarithmic scale of frequencies, *octaves,* as in the well-tempered scale. Notice that there are line segments in this scalogram that correspond to the undertone peaks in the spectrum for the guitar note.

We shall now describe how scalograms are computed. This is done via a method known as the *continuous wavelet transform* (CWT). The CWT differs from the spectrogram approach 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. Just as spectrograms are based on a discretization of Fourier coefficients, scalograms are also based on a discretization, a discretization of the continuous wavelet transform.

We now define the CWT. Given a function g, called the *wavelet*, the continuous wavelet transform $\mathcal{W}_g[f]$ of a sound signal f is defined as[8]

$$\mathcal{W}_g[f](\tau, s) = \frac{1}{s}\int_{-\infty}^{\infty} f(t)\overline{g\left(\frac{t-\tau}{s}\right)}\,dt \tag{4.1}$$

for *scale* $s > 0$ and *time-translation* τ. For the function g in the integrand of (4.1), the variable s produces a dilation and the variable τ produces a translation.

We omit various technical details concerning the types of functions g that are suitable as wavelets; the interested reader can consult [6] or [7]. It is shown in [6], and in [8], that (4.1) can be derived from a simple model for the response of our ear's basilar membrane—which responds to frequencies on a logarithmic scale—to an incoming sound stimulus f.

We now show how to discretize the integral in (4.1). First, we assume that the sound signal $f(t)$ is nonzero only over the time interval $[0, \Omega]$. Hence (4.1) reduces to

$$\mathcal{W}_g[f](\tau, s) = \frac{1}{s}\int_{0}^{\Omega} f(t)\overline{g\left(\frac{t-\tau}{s}\right)}\,dt.$$

Second, as we did for Fourier coefficients, we make a Riemann sum approximation to this last integral using $t_m = m\Delta t$, with a uniform spacing $\Delta t = \Omega/N$; and we also discretize the time variable τ, using $\tau_k = k\Delta t$. This yields

$$\mathcal{W}_g[f](k\Delta t, s) \approx \frac{\Omega}{N}\frac{1}{s}\sum_{m=0}^{N-1} f(m\Delta t)\overline{g\left(\frac{m-k}{s}\Delta t\right)}. \tag{4.2}$$

The sum in (4.2) is a correlation of two discrete sequences. Given two N-point discrete sequences $\{f_k\}$ and $\{g_k\}$, their *correlation* $\{(f : g)_k\}$ is the sequence defined by

$$(f : g)_k = \sum_{m=0}^{N-1} f_m \overline{g_{m-k}}\,. \tag{4.3}$$

[8] It is more common to use $1/\sqrt{s}$ in front of the integral in (4.1). The definition above causes no essential changes in the mathematics but simplifies some subsequent formulas.

(Note: In order for the sum in (4.3) to make sense, the sequence $\{g_k\}$ is *periodically extended* via $g_{-k} := g_{N-k}$.)

Thus, (4.2) and (4.3) show that the CWT, at each scale s, is approximated by a multiple of a discrete correlation of $\{f_k = f(k\Delta t)\}$ and $\{g_k^s = s^{-1}g(s^{-1}k\Delta t)\}$. These discrete correlations are computed over a range of discrete values of s, typically

$$s_p = 2^{-p/J}, \quad p = 0, 1, 2, \ldots, I \cdot J, \tag{4.4}$$

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 that we describe below—to the well-tempered scale used for pianos.

4.1. Gabor Wavelets. The CWTs that we use in this paper are based on Gabor wavelets. A *Gabor wavelet,* with width parameter w and frequency parameter η, is defined as follows:

$$g(t) = w^{-1}e^{-\pi(t/w)^2}e^{i2\pi\eta t/w}. \tag{4.5}$$

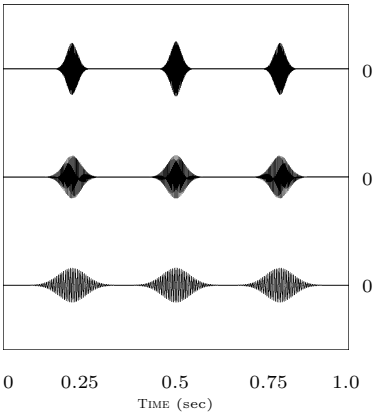
Notice that the complex exponential $e^{i2\pi\eta t/w}$ has a frequency of $\nu_0 = \eta/w$. The other exponential $w^{-1}e^{-\pi(t/w)^2}$ is a "bell-shaped" curve, a normal probability distribution with mean 0 and standard deviation $w/\sqrt{2\pi}$. This bell-shaped factor in (4.5) damps down the oscillations of g, so that their amplitude is significant only within a finite region centered at $t = 0$. In fact, if we use a rule of thumb from statistics that significant probabilities are confined to within plus or minus three standard deviations of the mean, then significant probability will result from g only within an interval of width $2.4w$ centered at 0.

To see how the CWT based on a Gabor wavelet produces a time-frequency analysis of a sound signal, let's consider the example of a Gabor wavelet of width $w = 0.25$ and frequency parameter $\eta = 20$. The complex exponential factor in (4.5) then has a frequency of $\nu_0 = 80$. In Figure 4.2 we show the components of a CWT based on this Gabor wavelet, and its application to a test signal. In Figure 4.2(a) we show plots of nine different functions $s^{-1}g(s^{-1}[t-\tau])$ for scale values $s = 2^{-p/6}$, $p = 8, 12, 16$, and translations $\tau = 0.2, 0.5, 0.8$. The top row shows three functions all having the same scale value $s = 2^{-16/6}$, the middle row shows three functions all having the same scale value $s = 2^{-12/6}$, and the bottom row shows three functions all having the same scale value $s = 2^{-8/6}$. These dilations and translations of this Gabor wavelet have significant values only in small intervals centered on $t = \tau$, with widths of $2.4sw \approx 0.38w, 0.60w, 0.95w$ for the top to bottom rows, respectively. *It follows that significant values for* (4.1)*,* (4.2)*, or* (4.3) *will only occur for values of the sound signal that are found within these small intervals.* This is how the CWT localizes its analysis of sound signals to small portions of time.

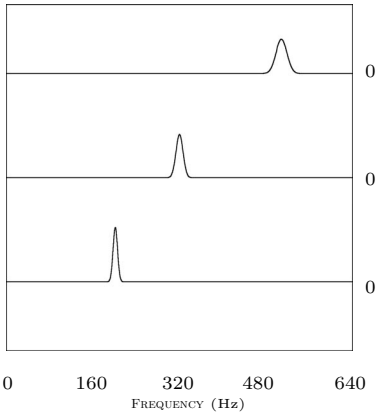
The CWT also localizes its analysis of sound signals to small portions of frequency. To see this, we observe that an FFT of (4.3) produces (see [1], [26], or [27])

$$(f : g)_k \mapsto F[n]\overline{G[n]}. \tag{4.6}$$

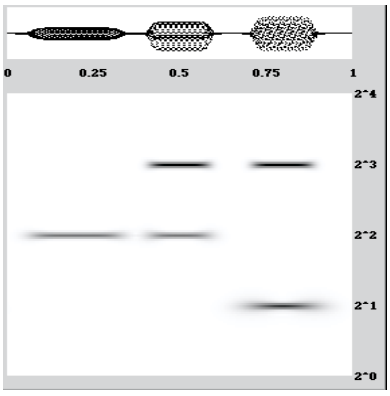
Formula (4.6) implies that the FFT of the correlation on the right side of (4.2) consists of FFT values for $\{f(t_k)\}$ multiplied by complex conjugates of FFT values for $\{g_k^s = s^{-1}g(s^{-1}t_k)\}$. In Figure 4.2(b) we show the magnitudes $|G^s[n]|$ of the FFTs of these sequences $\{g_k^s\}$ for the three scale values $s = 2^{-p/6}$, $p = 8, 12, 16$. These magnitudes



(a) Gabor wavelets



(b) Transform magnitudes



(c) Scalogram of a test signal

Fig. 4.2 *Gabor wavelet analysis.*

lie on bell-shaped curves of finite widths centered on specific frequencies. For instance, for $s = 2^{-12/6} = 2^{-2}$, the bell-shaped curve in the middle of Figure 4.2 is centered on the frequency $\nu = 320$ Hz. This frequency of 320 Hz satisfies $320 = 80s^{-1}$. The bell-shaped curve centered at $\nu = 320$ Hz can be shown to have a standard deviation of $(sw)^{-1}/\sqrt{2\pi} = 4w^{-1}/\sqrt{2\pi}$, which determines the visible width of the significant values of the bell-shaped curve in the figure. Similarly, for $s = 2^{-8/6}$ and $s = 2^{-16/6}$, there are bell-shaped curves centered on the frequencies $80s^{-1}$, which equal 201.587 Hz and 507.968 Hz, respectively. Thus, we see that the values of s in (4.4) induce a collection of bell-shaped curves centered on frequency values ν that lie along a logarithmic scale of values proportional to s^{-1}. For the particular Gabor wavelet we have been discussing, with frequency $\nu_0 = 80$ Hz, this is a logarithmic scale of frequencies that are multiples of $\nu_0 = 80$. These frequencies satisfy $\nu = \nu_0 s^{-1} = 80s^{-1}$. *It follows that significant values for the discretized CWT magnitudes* $|(f : g^s)_k|$ *will only occur for frequencies that are found within small intervals determined by the width of a bell-shaped curve centered on the frequency* $\nu_0 s^{-1}$. This is how the CWT localizes its analysis of sound signals to small portions along a logarithmic scale on the frequency axis.[9]

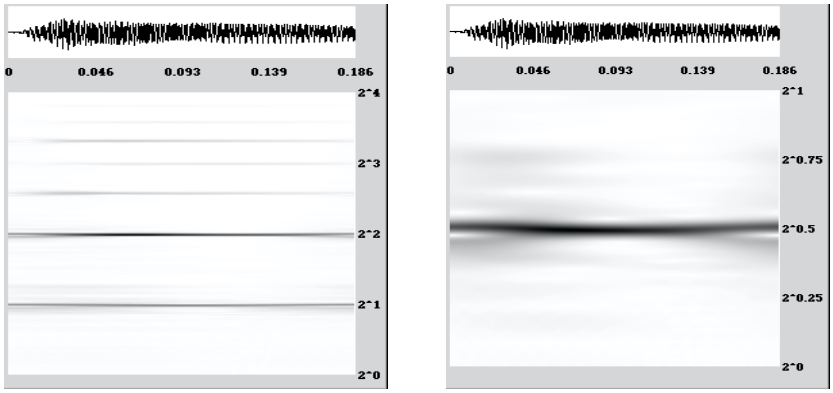
We have seen in the discussion so far that the magnitudes of a (discretized) CWT are localized within time and frequency. These magnitudes of a CWT are called the *scalogram* of the sound signal. Before we discuss scalograms for sounds from musical instruments, it may help to first examine a scalogram of a test signal. This example should illustrate that scalograms do provide time-frequency portraits of signals. The test signal is

$$\begin{aligned} &\sin(2\pi\nu_1 t)e^{-\pi[(t-0.2)/0.1]^{10}} \\ &\quad + [\,\sin(2\pi\nu_1 t) + 2\cos(2\pi\nu_2 t)\,]\, e^{-\pi[(t-0.5)/0.1]^{10}} \\ &\quad + [\,2\sin(2\pi\nu_2 t) - \cos(2\pi\nu_3 t)\,]\, e^{-\pi[(t-0.8)/0.1]^{10}}, \end{aligned} \tag{4.7}$$

where $\nu_1 = 320$, $\nu_2 = 640$, and $\nu_3 = 160$. A graph of this signal over the time interval $[0, 1]$ appears at the top of Figure 4.2(c) with its scalogram—using the Gabor wavelet of base frequency $\nu_0 = 80$ described above—graphed below it. As with spectrograms, the larger magnitudes in the scalogram are graphed more darkly. The time values are listed along the horizontal, while reciprocal scale values, $s^{-1} = 2^0, 2^1, 2^2, 2^3, 2^4$, are listed along the vertical.

The test signal in (4.7) has three terms. The first term contains a sine factor, $\sin(2\pi\nu_1 t)$, of frequency $\nu_1 = 320$. Its other factor, $e^{-\pi[(t-0.2)/0.1]^{10}}$, limits the significant extent of this term to a small interval centered on $t = 0.2$. This first term appears most prominently on the left third of the graph at the top of Figure 4.2(c). In the scalogram we can see a thin line segment lying directly below this left third of the signal, and this line segment is centered on the reciprocal scale value $s^{-1} = 2^2$ on the vertical axis. As we showed above, $s^{-1} = 2^2$ corresponds to a frequency $\nu = 320$ Hz. Thus the scalogram has produced a time-frequency portrait of the first term in the test signal, marking off a spectral line segment lying along the time axis in the same position as the significant values for this first term and lying along the frequency axis at the position of 320 Hz, which matches the frequency $\nu_0 = 320$ Hz. Similarly, there are spectral line segments for the remaining terms in the series and they mark off the locations, in time and frequency, of the significant values for these two terms.

[9] Alternative methods of time-frequency analysis, based on generalizing Gabor wavelets, have also been investigated. See, e.g., [2], [15], and [9].



(a) Piano note scalogram (b) Zoom on portion of (a)

Fig. 4.3 *Two scalograms.*

Exercise: Use FAWAV [24] to produce a spectrogram of this same test signal, and compare it with the scalogram.

For this test signal, we have seen that a Gabor wavelet scalogram provides an excellent time-frequency portrait. We shall now examine how well Gabor scalograms perform in analyzing the sounds from musical instruments.

4.2. Scalograms for a Guitar and Piano. We now discuss some scalogram analyses of musical sounds. In Figure 4.1(c) we have shown a scalogram of a guitar note. This scalogram, which spans four octaves using 32 voices per octave, was created using a Gabor wavelet with width parameter $w = 0.25$ and frequency parameter $\eta = 20$. Therefore, the base frequency ν_0, corresponding to $s^{-1} = 2^0$, is 80 Hz. The reciprocal scale values along the vertical axis in the scalogram are 2^0, 2^1, 2^2, 2^3, and 2^4, which correspond to frequencies of 80, 160, 320, 640, and 1280 Hz, respectively. By comparing the Fourier spectrum in Figure 4.1(b) with this scalogram, we can see that the spectral line segments in the scalogram are closely matched with the peaks in the Fourier spectrum. For instance, the two closely spaced peaks at approximately 191 and 215 Hz are clearly identified in terms of closely spaced spectral line segments in the scalogram, directly below the most intense spectral line segment at 320 Hz.

It is clear from Figures 4.1(a) and 4.1(c) that the scalogram provides a much more detailed time-frequency portrait of the guitar note than the spectrogram. The scalogram represents a mathematical zooming in on the portion of the time-frequency plane crowded together in the lower quarter of the spectrogram.

As another example of this zooming-in feature of scalograms, we show in Figure 4.3 two scalograms of a piano note. This piano note is the same as the one shown in Figure 2.2(a), a recording of the note E_4. The scalogram, which spans four octaves using 64 voices per octave, shown in Figure 4.3(a), was obtained using a Gabor wavelet having width parameter $w = 0.25$ and frequency parameter $\eta = 41.25$. These parameter values were chosen so that the fundamentals and overtones of the piano note are centered on the reciprocal scale values $s^{-1} = 2^1$, 2^2, and 2^3 shown in Figure 4.3(a). The base value on the reciprocal scale axis, $s^{-1} = 2^0$, corresponds to the frequency $\nu_0 = 165$ Hz, hence $s^{-1} = 2^1$ corresponds to the fundamental $\nu = 330$

Hz, and $s^{-1} = 2^2$, 2^3 correspond, respectively, to the overtones $\nu = 660$ Hz, 1320 Hz. Comparing the scalogram in Figure 4.3(a) with the Fourier spectrum in Figure 2.2(b), we see that the spectral line segments in the scalogram match up precisely with the peaks in the spectrum.

Comparing the scalograms of the guitar and piano notes in Figures 4.1(c) and 4.3(a), we can see how much "cleaner" the piano scalogram is. This surely corresponds to our auditory sense of the piano note as "purer" than the guitar note.

In Figure 4.3(b) we show how a scalogram can be used to zoom in on one spectral line segment in the time-frequency plane. For this scalogram, only 1 octave was used, but 256 voices were employed. The Gabor wavelet for this scalogram has a width parameter of $w = 0.125$ and frequency parameter of $\eta = 29.025$. Hence the reciprocal scale value of $s^{-1} = 2^{0.5}$ corresponds to the frequency $\nu_0 s^{-1} \approx 330$ Hz. Consequently, this scalogram provides a zooming in on the single spectral line segment at 330 Hz for the piano note. One interesting feature of this scalogram is the evident bending of this single spectral line segment. It is not centered on the single frequency of 330 Hz throughout the entire recording, but is slightly higher in pitch at the beginning and end. Whether our hearing can detect this slight variance in pitch is an interesting question.

4.3. Separating Closely Spaced Spectral Lines. For our final illustration of scalogram analysis, we discuss an example that has applications beyond music. In Figure 4.4(a) we show a plot of a sum of two sinusoids, $\sin 2\pi\nu_1 x + \sin 2\pi\nu_2 x$, with frequencies $\nu_1 = 60$ Hz and $\nu_2 = 59.2$ Hz. The graph was generated using 512 points, equally spaced over the interval $[0, 1)$. In Figure 4.4(b) we show the FFT spectrum for this signal. Notice that the two frequencies ν_1 and ν_2 are not clearly separated in this spectrum. In fact, there is just one peak in the spectrum, located at 59.0 Hz. There is also a subsidiary peak at 60 Hz, which appears as a "shoulder" in the graph shown in Figure 4.4(b). This lack of separation is reminiscent of the problem of resolving closely spaced spectral lines in spectroscopy, an important and fundamental problem in that field [3], [20].

The spectrogram in Figure 4.4(c) also does not reveal that there are two frequencies in the signal. Figure 4.4(d), however, shows that a scalogram has resolved the two separate frequencies. This particular scalogram spans 1 octave with 256 voices and was created using a Gabor wavelet having a width parameter $w = 1$ and frequency parameter $\eta = 42.4264 \approx 60/\sqrt{2}$. For a time interval extending from about 0.4 to 0.8, we can see a pair of separate spectral curves corresponding to the two frequencies. These bands are spread out in width and are *not* centered on reciprocal scales precisely corresponding to the frequencies of 59.2 and 60 Hz. It would be an interesting research problem to develop a method for accurately specifying these frequencies based on the scalogram data. One approach might be to locate the inner edges of these spectral bands and use their location along the reciprocal scale axis to identify separate frequencies. Edges would be defined, as in image processing, by a sharp transition from low to high magnitudes. In a scalogram this would mark a contrast line separating light and dark regions along the spectral bands. With this idea in mind, we found that the inner border along the top spectral band in Figure 4.4(c) lies at $s^{-1} \approx 2^{0.5}$, which corresponds to a frequency of $\nu = 60$ Hz. The inner border along the bottom spectral band lies at $s^{-1} \approx 2^{0.48}$, which corresponds to approximately $\nu = 59.2$ Hz. Where there is one spectral band, at the left and right sides of this scalogram, we instead determined the location of the central maximum along the single band, thus obtaining a reciprocal scale value of $s^{-1} \approx 2^{0.49}$, which

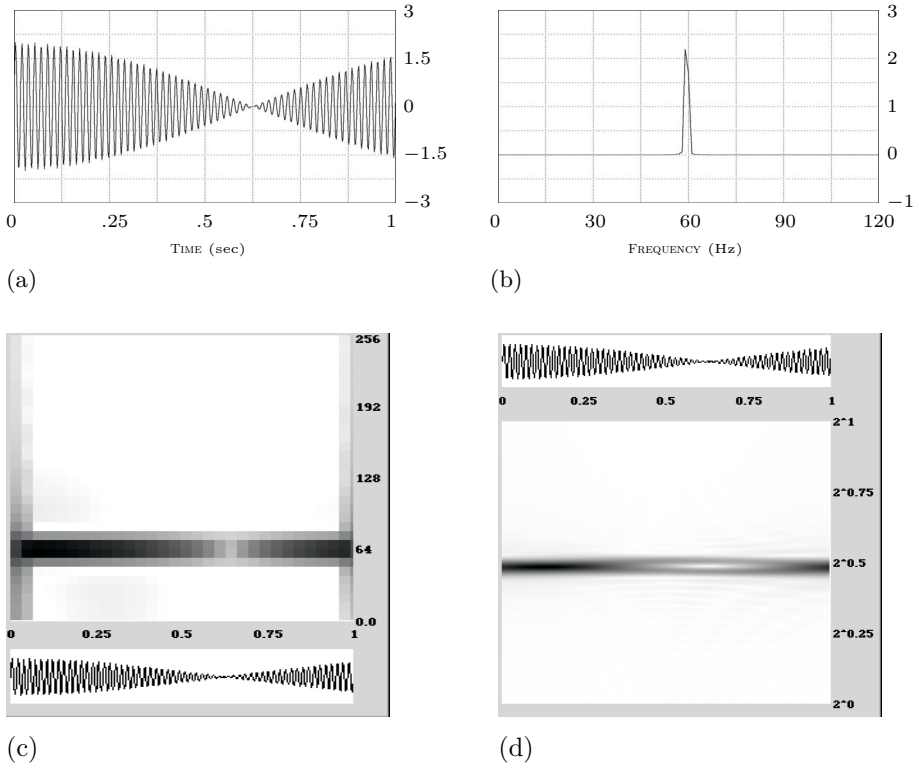


Fig. 4.4 (a) *Sum of two sinusoids,* 59.2 *and* 60 *Hz.* (b) *FFT Spectrum.* (c) *Spectrogram.* (d) *Scalogram.*

corresponds to a frequency of $\nu = 59.6$ Hz. Applications of these methods to more complicated spectra is a task for future research.

Exercise: Explain why the scalogram in Figure 4.4(d) shows a single spectral line segment on the left and right ends of the time interval $[0, 1]$, centered on a reciprocal scale value that corresponds to the frequency 59.6 Hz.

4.4. Inversion of Scalograms. In certain cases it is possible to invert scalograms. We shall not dwell on this point, since the mathematics is similar to what we described in section 3.2 on the inversion of spectrograms.

If we let $q_k = (f : g)_k$ and $\tilde{g}^s_k = g^s_{-k}$, then $(q : \tilde{g}^s)_k$ has an FFT satisfying (see [1], [26], or [27])

$$(q : \tilde{g}^s)_k \mapsto F[n]|G^s[n]|^2. \tag{4.8}$$

Summing the FFTs in (4.8) for $s = s_p$ in the range in (4.4), and assuming that

$$A \le \sum_{p=0}^{I \cdot J} |G^{s_p}[n]|^2 \le B \tag{4.9}$$

for some positive constants A and B, we can then divide out $\sum |G^{s_p}[n]|^2$ to obtain the FFT $\{F[n]\}$ of the sequence $\{f_k\}$. Hence inversion can be performed if (4.9) holds. More details can be found in [12, Chap. IV].

4.5. Further Connections to Music. In this section we established the connection between reciprocal scale and frequency, which yields a time-frequency interpretation of scalograms. This is the correct approach when discussing Gabor wavelets. Flandrin [10, p. 210], however, pointed out a broader interpretation:

> Although this time-frequency interpretation is of current use in wavelet analysis and scalograms, it is important to observe that it can be restrictive and does not always support the most pertinent point of view. In particular, this happens when we ... allow spectra [for the wavelet] with several "humps." Then they cannot simply be attached to a *single* frequency. Rather, they must be associated with the *proportions* between the frequencies. Explaining this briefly in terms of music, the wavelet transform looks more like an analysis by *chords* rather than *notes,* and this renders the *scale* more meaningful than the *frequency.*

An interesting, nonmathematical application of such a chordal analysis of music is lucidly and succinctly described in [17, pp. 528–538]. Further details, demanding more thorough musical knowledge, can be found in [11].

5. Conclusion. In this paper we have examined three different approaches to time and frequency analysis: Fourier spectra, spectrograms, and scalograms. Fourier spectra identify spectral peaks within *entire* musical signals, while spectrograms and scalograms provide two different ways of capturing the time-frequency content of a musical signal. Some basic applications to music were discussed, and an application to spectroscopy was illustrated as well.

REFERENCES

[1] R.N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed., McGraw-Hill, New York, 2000.

[2] R.N. Bracewell, *Adaptive chirplet representation of signals on time-frequency plane*, Elec. Lett., 27 (1991), pp. 1159–1161.

[3] G.L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Lecture Notes in Statist. 48, Springer-Verlag, Berlin, 1988.

[4] W.L. Briggs and V.E. Henson, *The DFT. An Owner's Manual for the Discrete Fourier Transform*, SIAM, Philadelphia, 1995.

[5] E.O. Brigham, *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

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

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

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

[9] H. Feichtinger and T. Strohmer, eds., *Gabor Analysis and Algorithms: Theory and Applications*, Birkhäuser, Boston, 1998.

[10] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, CA, 1999.

[11] R. Jackendoff and F. Lerdahl, *Generative Theory of Tonal Music*, MIT Press, Cambridge, MA, 1983.

[12] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1998.

[13] H.S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Norwood, 1992.

[14] Y. Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.

[15] D. Mihovilović and R.N. Bracewell, *Whistler analysis in the time-frequency plane using chirplets*, J. Geophysical Research, 97 (1992), pp. 17,199–17,204.

[16] H.F. Olson, *Music, Physics and Engineering*, 2nd ed., Dover, New York, 1967.

[17] S. Pinker, *How the Mind Works*, Norton, New York, 1997.

[18] J.G. Roederer, *Introduction to the Physics and Psychophysics of Music*, 3rd ed., Springer-Verlag, New York, 1995.

[19] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1986.

[20] D.S. Sivia, *Data Analysis. A Bayesian Tutorial*, Oxford University Press, Oxford, 1996.
[21] K. Steiglitz, *A Digital Signal Processing Primer*, Addison-Wesley, Menlo Park, CA, 1996.
[22] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Boston, 1996.
[23] J. Ville, *Théorie et applications de la notion de signal analytique*, Câbles et Transmissions, 2 A, 1948, pp. 61–74.
[24] J.S. Walker, Fawav software, available online from http://www.uwec.edu/academic/curric/walkerjs/TFAMI/.
[25] J.S. Walker, *Fourier Analysis*, Oxford University Press, New York, 1988.
[26] J.S. Walker, *Fast Fourier Transforms*, 2nd ed., CRC Press, Boca Raton, FL, 1996.
[27] J.S. Walker, *A Primer on Wavelets and Their Scientific Applications*, CRC Press, Boca Raton, FL, 1999.
[28] M.V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A. K. Peters, Wellesley, MA, 1994.