In : S. L. Hopp, M. J. Owren and C. S. Evans (Editors)

Animal Acoustic Communication : Sound Analysis and Research Methods.

Springer-Verlag, Berlin, 1998.

Digital Signal Analysis, Editing, and Synthesis

K. BEEMAN

1 Introduction

Historically, the quantitative study of sound has been wedded to the development of sound-measurement technology. Researchers have routinely seized on and resourcefully adapted various technological tools, whether intended for sound analysis or not. Sabine (1900), for example, developed acoustical reverberation theory in an empty theater at Harvard University, using an organ pipe, a chronograph, and his own hearing to measure reverberant sound duration. Similarly, Brand (1934) characterized the time-varying frequency of birdsong by recording vocalizations on motion-picture film and measuring spatial line-density on the soundtrack. Successive milestones in sound-measurement technology — notably the microphone, the oscilloscope, and later the sound spectrograph — helped researchers to visualize and measure sounds but not to model them directly. Modeling of acoustic communication was instead typically performed indirectly via statistical analysis, comparison, and classification of individual measured sound features.

The development of digital computers has allowed researchers to extend and automate sound-measurement capabilities in many areas. However, the most important impact of this technology on bioacoustics may ultimately be the resulting accessibility of sophisticated mathematical tools for modeling sound structure, and the integration of these tools into the measurement process. The implications of this merger range from dramatic improvement in efficiency to the opportunity to conduct sophisticated interactive experiments in which the computer, for instance, presents stimuli, receives and analyzes resulting behavioral responses, and selects stimuli on this basis for the next testing cycle. Mathematical tools allow the investigator to measure and manipulate features ranging from discrete sound parameters, such as time and frequency maxima, to more comprehensive sound properties, such as time-varying amplitude and frequency and overall spectrographic similarity.

This chapter describes digital measurement and modeling techniques while providing a practical survey of the tools available to bioacousticians. Topics include measurement of detailed time and frequency parameters over precise intervals, real-time analysis and display of individual spectra and spectrograms, digital sound comparison and statistical classification, precise rearrangement of sound elements in the temporal domain, and sound synthesis based on natural sounds, mathematical functions, or both. Synthesis manipulations such as arithmetic operations, sound combination, time and frequency shifting and rescaling, frequency modulation, harmonic manipulation, and noise addition and removal are covered. Digital sound analysis technology has been applied to a wide variety of animal groups, notably primates (e.g., Owren and Bernacki 1988; Hauser 1991, 1992; Hauser and Schön-Ybarra 1994), anurans (e.g., Gerhardt 1989, 1992; Ryan et al.1992), birds (e.g., Nowicki and Capranica 1986; Nelson and Croner 1991; Suthers et al. 1994; Nelson et al. 1995), insects (e.g., Wells and Henry 1992; Stephen and Hartley 1995), and marine mammals (e.g., Clark 1982; Buck and Tyack 1993).

More advanced topics in signal analysis, measurement theory, and digital technology applicable to sound analysis are discussed in other works. These include mathematical signal analysis (Schafer 1975), engineering measurement theory (Randall 1977), digital hardware principles and guidelines (Stoddard 1990), and signal analysis and bioacoustic instrumentation (Rosen and Howell 1991). Digital sound analysis systems have been developed for various computers, including DEC minicomputer (Beeman 1987), IBM PC (Beeman 1989, 1996b), Amiga (Richard 1991), and Apple Macintosh (Charif et al. 1995).

2 Temporal and Spectral Measurements

Animals and humans respond to both the temporal and spectral structure of sounds. Biologically important sound attributes include temporal properties such as duration, repetition, and sequencing of sound elements, as well as spectral properties such as frequency, bandwidth, harmonic structure, and noisiness. Temporal properties can be measured from the *amplitude-time waveform*, which specifies acoustic pressure as a function of time, and the *amplitude envelope*, which specifies time-averaged acoustic intensity as a function of time. Spectral properties can be derived from the *power spectrum*, which specifies energy distribution as a function of frequency, and the *frequency-time spectrogram*, which specifies energy distribution as a function of both frequency and time. Significant information can be encoded in both the temporal and spectral domains of sound signals, and the following sections provide a survey of the digital techniques used to extract and analyze the associated acoustic properties.

Digital signal analysis typically begins with the measurement of basic time and frequency parameters. Important components include maximum and minimum values and the associated time and frequency coordinates. These extrema can be measured over the entire sound or, by restricting the analysis window to specific segments, over a succession of signal intervals. Temporal and spectral examples include, respectively, the peak level of an amplitude envelope and its time coordinate, and the amplitudes and frequency coordinates of successive peaks within a spectrum. These measurements can be compared statistically over an ensemble of signals to assess a wide variety of features. Important characteristics of the time waveform include its average signal level (which, if different from zero for an acoustic signal, may indicate instrumentation problems), and its *root-mean-square* (RMS) signal level (the square-root of the time-averaged squared amplitude), which is the standard measure of signal energy. Other parameters of interest that are available from spectrographic representations include the overall duration of a signal, its frequency range, and successive maxima and minima in its time-varying pitch contour.

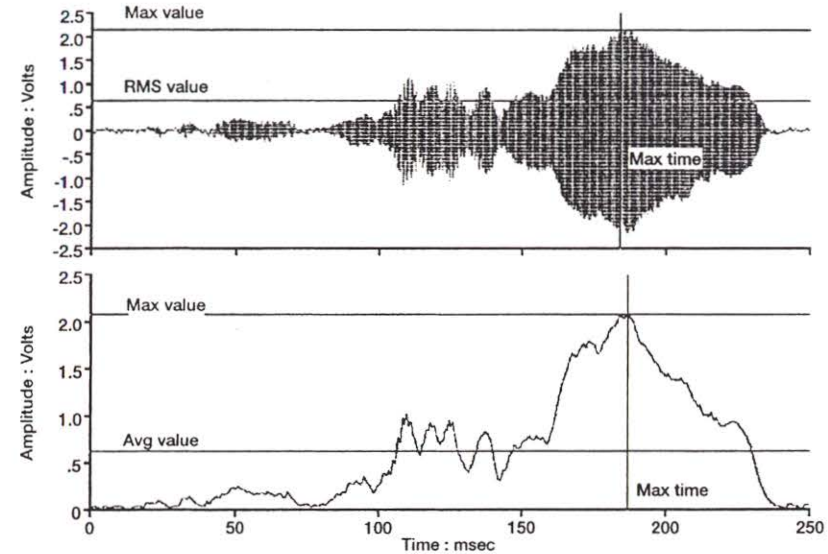

Fig. 1. Measuring duration, amplitude, and root-mean-square (RMS) level of a swamp sparrow (*Melospiza georgiana*) syllable from its waveform *(top panel)* and amplitude envelope *(bottom panel)*

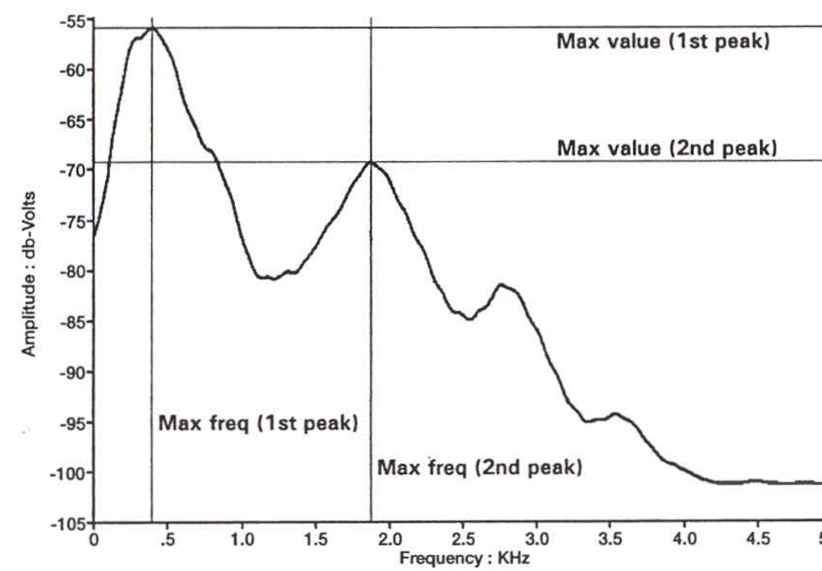

Fig. 2. Measuring the frequency and amplitude of major spectral peaks in a power spectrum

In a digital environment, measurements can be performed on individual notes, entire calls, or average signals derived from an ensemble of sounds (Beeman 1996b). Using the results of such measurements, the sounds can then be manipulated individually or in groups. For example, an ensemble of sounds can be normalized to the same sound energy level before playback, by dividing each sound by its own RMS value. Figure 1 illustrates the use of digital techniques to measure the duration, amplitude, and RMS value of a time waveform and its associated amplitude envelope, while Figure 2 shows frequency and amplitude measurements of peaks in a power spectrum.

3
Time-Varying Amplitude Analysis

3.1
Amplitude Envelopes

Temporal analysis is concerned with the location, duration, and pattern of sound events, such as the sequence and timing of notes within a song, songs within a calling sequence, or the pattern of variation in sound intensity (see Gerhardt, this Volume, for discussion of terminology used to distinguish acoustic units in the temporal domain).

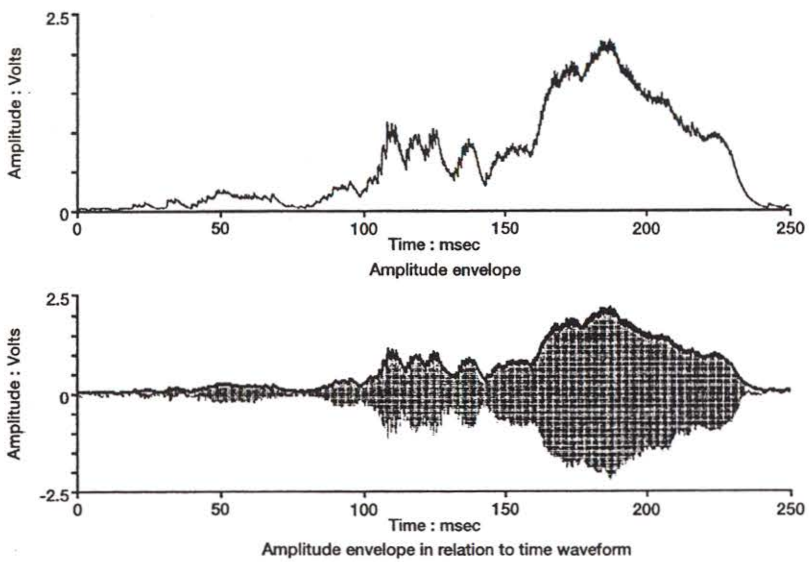

Fig. 3. The amplitude envelope (*top panel*) superimposed on the waveform (*bottom panel*) shown in Fig. 1

Although many parameters of interest can be measured directly from the time-amplitude waveform, it can be more efficient and objective to derive these from the corresponding amplitude envelope.

Visually, the *amplitude envelope* traces the outline of a waveform, as shown in Figure 3. Such tracking is achieved mathematically by jumping from peak to peak and bridging valleys of the signal, with the goal of retaining larger-scale and potentially meaningful amplitude variations while ignoring smaller-scale changes. Before computing the envelope, the time waveform is *rectified* by inverting all negative-going waveform components to positive-going ones. The tracking algorithm is designed to follow the rectified input signal when its amplitude increases, while decaying in accordance with a specified time constant (T) when the input level falls.

Discrimination of larger versus smaller-scale amplitude variations is determined by the decay-time constant used, where different values of T will cause the envelope to reflect signal features of different time scales. With a large T-value, longer, slower variations are tracked and retained while shorter variations are smoothed away. A small T-value, in contrast, produces an envelope that retains detailed aspects of amplitude modulation in the original signal. The optimal T-value for a particular signal type must be determined empirically, and should typically be both small enough to retain significant amplitude-modulation features and large enough to reveal the desired amount of overall envelope shape.

Amplitude envelopes can be used to measure duration, onset and offset patterns, amplitude modulation, intensity variation between notes, and time-varying intensity of waveforms. Envelopes can be used in synthesis to isolate and manipulate a sound's amplitude-time characteristics. Amplitude envelopes can also be used to quantitatively derive the average amplitude behavior of an ensemble of signals (by averaging the entire set of envelope functions), or to compare the similarity of amplitude behavior between notes (by cross-correlating the envelopes). For playback purposes, the amplitude character of a waveform can be removed by dividing the signal by its envelope function, producing a constant amplitude signal. The amplitude modulation of a signal envelope can be derived and analyzed mathematically by deriving a smoothed amplitude envelope (using a larger T) that is subtracted from the original envelope, leaving only the modulation function (see discussion below and in Beeman 1996b).

3.2
Gate Functions

While the amplitude envelope allows the researcher to measure the time-varying characteristics of a signal at various levels of time resolution, one is sometimes interested only in the overall temporal pattern of a sequence of sounds. A useful technique for temporal analysis at this level is to convert the amplitude envelope into a *gate function* (illustrated in Figure 4; also see Beeman 1996b). This continuous time function has only two values, used to distinguish signal events from silences. It is derived by comparing the amplitude envelope to a *threshold level*, which is normally set just above the signal's baseline noise level. The gate function is assigned a value of 1 wherever the signal exceeds the threshold and, zero elsewhere. It is usually also desirable to require the signal to be

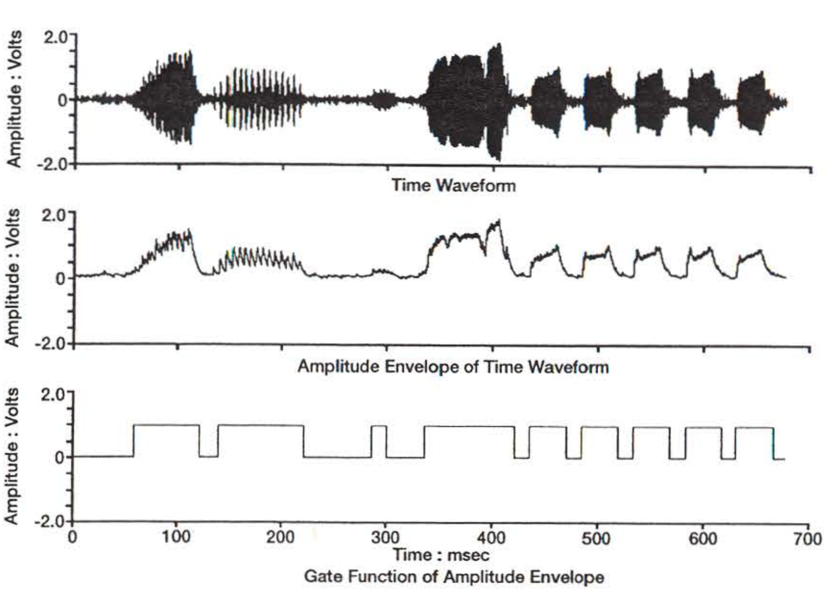

Fig. 4. Using a gate function to measure note duration and spacing. The time waveform (*top panel*) is converted to an amplitude envelope (*middle panel*) which is converted to the binary-valued gate function (*bottom panel*), indicating the presence and locations of sound segments

present or absent for some minimum interval before gate-function polarity can change. This allows the algorithm to respectively reject short noise bursts and bridge brief amplitude gaps within a single coherent event.

Overall, the gate function is a powerful tool that can be applied on a variety of scales in both the time and frequency domains. Applied to temporal features, gate functions can be used to characterize durations of both individual notes and entire songs, producing objective groupings of successive sounds and silences. Temporal variation in repeated note sequences can be characterized by measuring the durations of notes and inter-note silences and collecting the results in two corresponding histograms. These histograms can then be compared among both species and individuals. The similarity between gate functions can also be quantified by calculating their cross-correlation values, for example, to characterize the temporal similarity among trill sequences from different individuals of a species. An example of the application of gate-function analysis to spectral parameters is detection of the bandwidth of frequency peaks.

4 Spectral Analysis

The power spectrum of a signal is most commonly derived by *Fourier transformation* of the time waveform, which is a mathematical integration technique that characterizes

the amplitude levels of the various frequencies that are present (see Clements, this Volume, for discussion of Fourier methods). The first digital Fourier technique was the *discrete Fourier transform* (DFT), while an efficient computational algorithm known as the *fast Fourier transform* (FFT) later made this technique routinely available in digital acoustic analysis systems.

4.1 Power Spectrum Features

In generating Fourier-based power spectra, the researcher must consider the time-versus-frequency-resolution trade-off, selection of a particular transform window, application of transform smoothing, and (where absolute energy levels are important) scaling the transform. Various aspects of these issues are discussed elsewhere in this Volume (see Clements, and Owren and Bernacki), and by Beeman (1996b), and only the most important points are summarized here.

The *bandwidth* of a signal is defined as its maximum frequency range, in other words, a signal containing frequencies up to 10 kHz has a 10-kHz bandwidth. The *time resolution* and *frequency resolution* of a spectral measurement express its degree of inherent temporal and spectral uncertainty, respectively. When Fourier transforms are used, resolution is determined solely by the physical duration of the transformed time signal (T), and does not depend on digital sampling rate except as the latter affects analysis duration. Time resolution is then T seconds, while frequency resolution is $1 / T$ Hz.

For measurement purposes, a silent interval can be added at the end of the signal before transforming in order to produce a smoother spectrum, a useful procedure referred to as *zero-padding*. Zero-padding will be beneficial up to an overall segment length of about 4 times that of the original signal (e.g., 1024 data points plus 3072 zeroes). Mathematically, zero-padding increases the density of computed points in the Fourier spectrum, thereby achieving better sampling of the true underlying spectrum. Nonetheless, maximum spectral resolution is limited by the length of the original signal without zero-padding.

In Fourier analysis, time and frequency resolution are mutually constrained by a trade-off known as the *time-bandwidth product*, meaning that the product of time resolution and frequency resolution is always equal to 1. As a result, improving resolution in one domain necessarily degrades resolution in the other. For example, a 100-msec temporal resolution means that the time of occurrence of any acoustic feature of interest can be pinpointed only within a 100-msec interval. The associated spectral resolution of 10 Hz implies that the frequency of any acoustic feature of interest can only be pinpointed within a 10-Hz interval. Improving temporal resolution to 10 msec dilates the bandwidth of spectral uncertainty to 100 Hz, while sharpening the latter to 5 Hz dilates time resolution to 200 msec. This relationship (also known as *temporal-spectral uncertainty*) is discussed and illustrated by Beecher (1988), and Gerhardt (this Volume).

Fourier analysis of a short segment of a continuous sound can produce mathematical artifacts due to the abruptness of the segment's onset and offset. To minimize these edge effects, a *window* function is usually applied to the signal before transformation to taper its beginning and ending amplitudes. A *rectangular window* provides no tapering, while a *Hanning window* varies sinusoidally between a beginning value of 0, a center value of 1, and an ending value of 0. The *Hamming window* is similar to the Hanning, but begins and ends on a value greater than zero. Transform windows differ in the manner and extent to which they blur the true underlying spectrum and the choice involves unavoidable trade-offs. The Hanning window, for instance, separates closely spaced harmonics or sidebands better than does the rectangular window, but also broadens and delocalizes the exact frequency of these bands. Overall, however, the Hanning window is the best choice for biological sound analysis (see also discussion by Clements, this Volume).

Digital power spectra are typically fuzzy, and it is usually desirable to *smooth* them in order to see the underlying contour more clearly. Smoothing is usually performed through the *running-average* technique, in which each point in the power spectrum is recomputed as the mean value of itself and a number of adjacent points (specified as the *width* of the smoothing window). Smoothing should be considered an integral follow-up to the calculation of a digital power spectrum. Optimal smoothing width depends on the scale of the features of interest. Larger smoothing widths, for instance, emphasize overall spectral shape but suppress fine frequency variation, such as closely spaced harmonics. This distinction is illustrated in Figure 5, which shows the power spectrum of a human vowel sound conditioned by two different smoothing widths. The first was smoothed with a 50-Hz window and shows the harmonic energy components produced by periodic vocal-fold vibration. The other spectrum was twice smoothed using a 150-Hz window. The effect is to average across individual harmonics, revealing an overall spectral shape. Frequency resolution of the transform before smoothing was 0.76 Hz, time resolution was 1.32 sec, and a Hanning window was used.

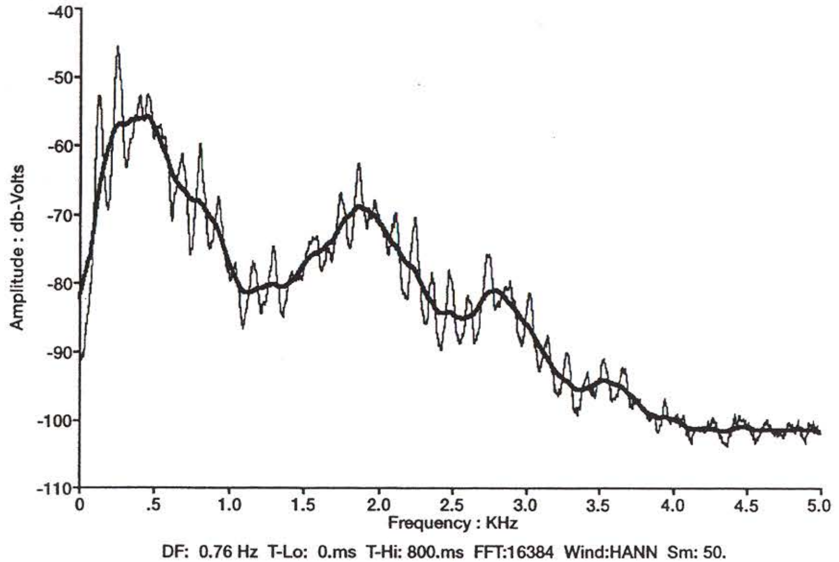

Fig. 5. Power spectra of a vowel sound from human speech illustrating the effects of smoothing window widths of 50 Hz (*light trace*) and 150 Hz (*bold trace*).

Often, the researcher needs only to measure relative energy levels within a power spectrum, or to compare energy levels between different spectra obtained under constant conditions. Sometimes, however, measurements of absolute sound levels are required. In these cases, spectral energy must be explicitly related to some physical quantity, such as pressure or acceleration. Such relationships are complex; more detailed discussions can be found in Randall (1977) and Beeman (1996b). In brief, power spectra should be displayed in decibel (dB) units relative to 1 Volt-RMS in the source time - signal, so that a 1-Volt-RMS sinewave signal is transformed to a nominal peak spectral level of 0 dB. This relationship is affected by a variety of factors, including spectral density, the transform window used, and the energy-versus-power units chosen for a particular application. Power spectra can be converted to physical units by applying the calibration factor of the measurement system (e.g., microphone, amplifier, recorder), which is the ratio of the physical input (such as sound pressure) to electrical output (Volts).

4.2 Measuring Similarity Among Power Spectra

As in the temporal domain, patterning in the frequency spectrum can represent a biologically significant aspect of an acoustic signaling process. It is therefore of interest to compare the power spectra derived from various sounds, for instance with respect to overall *sound similarity* and relative *frequency-shift* effects (in which signals show a similar spectral pattern but differ in the absolute frequency position of important spectral peaks). These characteristics can be quantified by cross-correlating pairs of power-

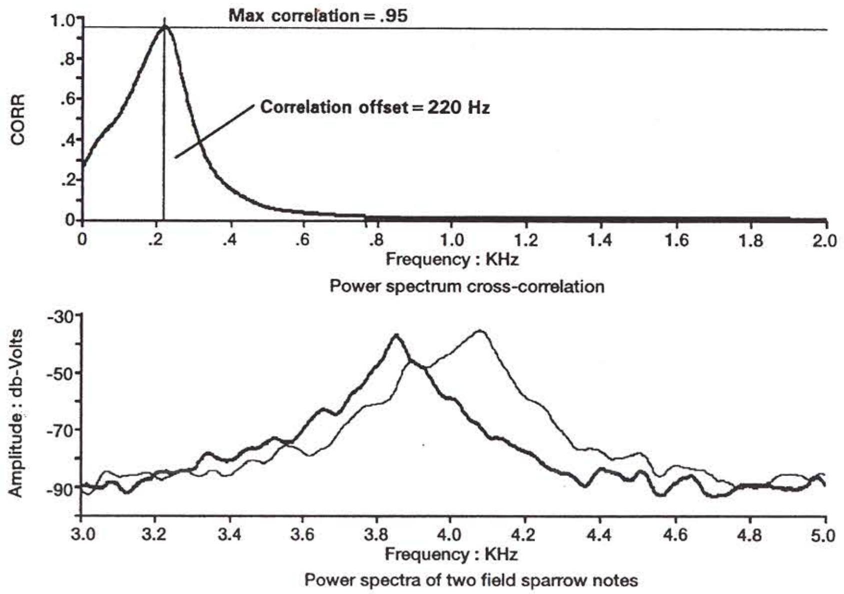

Fig. 6. The power spectra of two field sparrow (*Spizella pusilla*) notes (*bottom panel*), and their cross-correlation function (*top panel*), showing the maximum correlation value and its frequency offset.

spectrum functions (Beeman 1996b). This computation produces both an overall correlation value and mean frequency-shift for each pair, as illustrated in Figure 6. The results of pairwise similarity analysis over an entire data set can be collected in matrices and analyzed using multivariate statistical techniques like multidimensional scaling to reveal clustering tendencies. Two potential applications of this approach are analyzing the level of spectral similarity among individuals and species, and characterizing spectral changes occurring during ontogeny.

4.3 Other Spectral Analysis Techniques

Spectral analysis can also be conducted using techniques that are not based on the Fourier transform. One technique is *zero-crossing analysis* (ZCA; Greenewalt 1968; Dorrscheidt 1978; Staddon et al. 1978; Margoliash 1983). In this approach, the frequency content of a time-signal segment is calculated by measuring successive time intervals at which the waveform crosses the zero-Volt line, and computing the reciprocal of each such period. The principal advantage of ZCA is that time and frequency resolutions can be as much as ten times greater than in Fourier analysis, because there is no inherent time-frequency resolution trade-off involved. The principal disadvantage of ZCA is that signal components other than the fundamental frequency (e.g., higher harmonics and noise) strongly alter the zero-crossing points in the waveform, degrading the frequency measurements. Fourier transforms, in contrast, produce measurements based on multiple waveform cycles, mitigating the effects of noise, and are unaffected by the presence of higher-order harmonics. Thus, while ZCA avoids the time-bandwidth product limitation, it lacks the performance stability and noise-immunity associated with the Fourier transform.

A statistical technique for spectrum estimation called *linear predictive coding* (LPC; see Markel and Gray 1976, as well as Owren and Bernacki, and Rubin and Vatikiotis-Bateson, this Volume) is widely used for the spectral measurement, modeling, and synthesis of human speech and has also been applied to nonhuman subjects. This technique models the sound spectrum as a sum of spectral resonances, as in the human vocal tract, and consideration should be given to its appropriateness to sound structure and production physiology before application to other species (see Owren and Bernacki for detailed discussion of this issue).

5 Spectrographic Analysis

A frequency-time spectrogram expresses the time-varying spectral content of a signal. It displays energy in three dimensions - amplitude, frequency, and time - where the amplitude of a signal component is expressed visually as the darkness of an area displayed on axes representing time and frequency. Two spectrograms of the human speech utterance "dancing" are shown in Figure 7, produced using approximately the wide and narrow frequency bandwidths traditionally used in the study of acoustic pho-

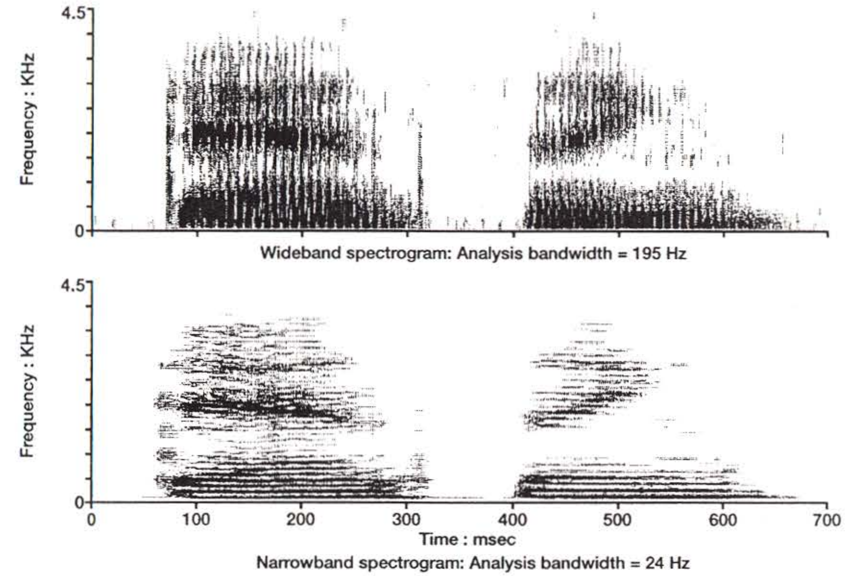

Fig. 7. Narrowband (24 Hz bandwidth, *bottom panel*) and wideband (195 Hz bandwidth, *top panel*) digital spectrograms of the human speech utterance "dancing"

netics. This analysis technique was made widely available with the invention of the analog sound spectrograph (Koenig et al. 1946), an instrument that also became known as the *sonagraph*, the product of which is correspondingly called a *sonagram*. Spectrographic analysis can be enormously useful in discerning spectral-temporal patterns, harmonic structure, and general similarity among sounds, and has historically been the preferred method for investigating bioacoustic signals.

Just as a digitized waveform represents a sampled, numerical version of a physical signal, a digital spectrogram is a sampled, numerical version of the traditional, analog spectrogram. In essence, a digital spectrogram is a two-dimensional, frequency-by-time matrix in which each cell value represents the intensity of a particular frequency component at a given time. Each column in this matrix represents the sound spectrum derived over a short time segment. A digital spectrogram is generated by stepping a short analysis window across the waveform and calculating the discrete Fourier transform of the "windowed" waveform segment at each step. The spectrograms shown in Figure 7 are digital, created using this stepping process, which is illustrated in Figure 8. The spectrogram is displayed on the computer screen by assigning numerical amplitude gradations either to a range of gray shades or to various colors. The matrix can also be shown in a three-dimensional plot, or *waterfall display*, which shows each power spectrum slightly offset from its neighbors on a diagonally oriented time axis.

Digital spectrograms can be used in different ways. First, they can be examined visually, using human judgment to discern patterns and relationships among sounds. Second, unlike the analog case, digital representations can be analyzed using mathematical tools to characterize acoustic properties of interest. This numeric approach allows the researcher to, for instance, measure onset and offset characteristics of notes or syllables both in the time and frequency domains, extract time-varying frequency contours, calculate an average, or template, representation from an ensemble of spectrograms, quantify spectrogram similarity through cross-correlation, and perform pattern recognition and sound classification using a variety of image analysis, similarity, and statistical techniques.

5.1 Spectrogram Generation

Essential considerations in spectrogram generation include time and frequency resolution, the number of transforms used, time *granularity*, and characteristics of the analysis window. As digital spectrograms are based on DFT calculation, the transform parameters used strongly influence the resulting representation. To interpret spectrograms accurately, it is essential to understand these parameters and their effects. These issues are discussed in Beeman (1996b) and only the most important points are reviewed here. Flanagan (1972) and Clements (this Volume) also provide brief mathematical descriptions of DFT-based spectrogram generation.

Time-frequency resolution is constrained by the time-bandwidth product, as discussed earlier. While finer time resolution reveals finer temporal features, the accompanying coarser frequency resolution blurs spectral features. Conversely, achieving finer frequency resolution of these features tends to obscure their temporal locations. These outcomes are clearly illustrated in the two spectrograms shown in Figure 7, which

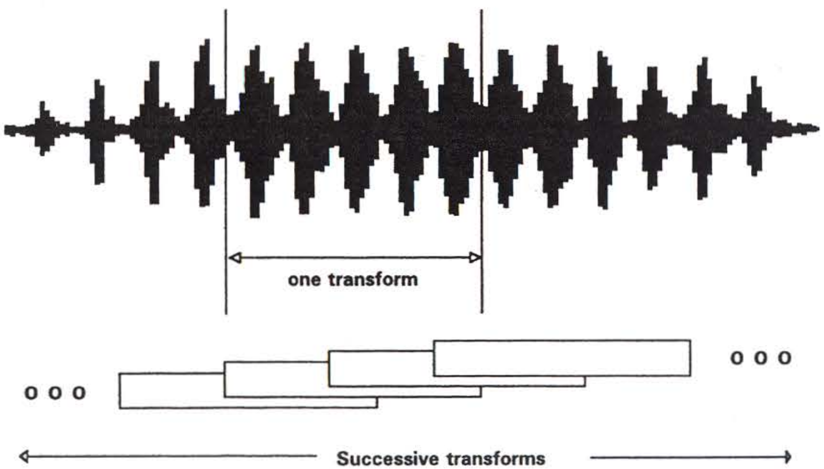

Fig. 8. Generation of a digital spectrogram by applying a transform window to successive segments of the time waveform. The user controls the length of the window and the degree of overlap between adjacent windows

are drawn respectively with wideband (195 Hz, 5 ms) and narrowband (24 Hz, 42 ms) resolutions. Optimum time and frequency resolutions are chosen on the basis of the signal features of interest and the relative importance of temporal and spectral behavior in the analysis (Beecher 1988). Spectrograms with high time and frequency resolution can be produced by non-Fourier-based techniques such as Wigner and Gabor transforms, which avoid the time-bandwidth limitation (Loughlin et al. 1993). These nonlinear techniques are promising but can introduce spurious spectral components when applied to non-tonal sounds, and should be used carefully.

Selecting the transform window for spectrographic representation involves essentially the same considerations that apply in calculating a power spectrum. Relative to a rectangular window, the Hanning window produces finer, higher-contrast spectrograms, and is recommended for general use. The Hanning window is useful in separating closely spaced signal harmonics or sidebands, which are blurred by a rectangular window. On the other hand, the Hanning window broadens and delocalizes the bands, making exact frequency determination more difficult.

5.2 Spectrogram Display

As noted earlier, digital spectrograms are displayed by converting the values associated with individual cells in the frequency-time matrix to gray shades or colors on the screen. The *intensity display range* of the spectrogram should be matched to the *dynamic range* of the signal. The display range is the range of screen display levels and varies from darker shades for higher amplitude levels to lighter shades for lower amplitude levels, with white representing the absence of energy. The dynamic range is the difference between the highest and the lowest numerical signal values in the matrix. By varying the display range, the user can emphasize features associated with different intensity levels in the spectrogram. For example, displaying the upper 40 dB of a spectrogram can highlight the signal by causing lower-amplitude, background noise components to lighten or disappear, while narrowing the range to the upper 20 dB may suppress even low-amplitude signal components, producing a *silhouette* spectrogram. Other traditional spectrograph features can also be emulated digitally, such as the *high-frequency shaping filter* used to enhance high-frequency components by increasing their relative amplitude.

In the traditional analog spectrogram, the number of transforms involved is effectively infinite, since the analyzer moves continuously in the temporal domain. In a digital spectrogram, a finite number of transforms are calculated, producing visible time-domain granularity. This granularity can be reduced by computing a larger number of transforms, which increases the processing time required. This trade-off can be ameliorated by interpolating additional columns and rows between the original ones in the spectrogram matrix during the display process. This method is quite effective in increasing the smoothness and visual resolution of the spectrogram display, but does not improve the underlying mathematical time and frequency resolutions.

5.3 Spectrogram Parameter Measurements

Computer-based spectrographic analysis typically involves a spectrogram screen display and a cursor used to measure features of interest. Results of such measurements are typically stored in a text file for further analysis. Some systems allow the user to navigate the spectrogram, displaying and measuring waveform and power-spectrum cross-sections directly. A further enhancement is the ability to view sound data continuously on a real-time, scrolling spectrographic display (Beeman 1991, 1996a), pausing the display as necessary to make screen measurements.

Computer-based spectrographic screen measurements are faster and more accurate than traditional approaches involving manual measuring instruments, but are still manual and thus do not represent a theoretical advance. Truly new approaches include numerical techniques which analyze the spectrogram matrix to measure sound parameter values, and image analysis techniques for the recognition of sound features. For example, time and frequency boundaries in the spectrogram can be extracted in a manner analogous to gate-function analysis by locating the coordinates at which matrix values exceed specified threshold levels. This technique can be used to characterize the temporal and spectral ranges of the signal, allowing automation of the editing and analysis process by identifying the temporal segments to be extracted, or the boundaries of specific frequency bands (see Figure 9).

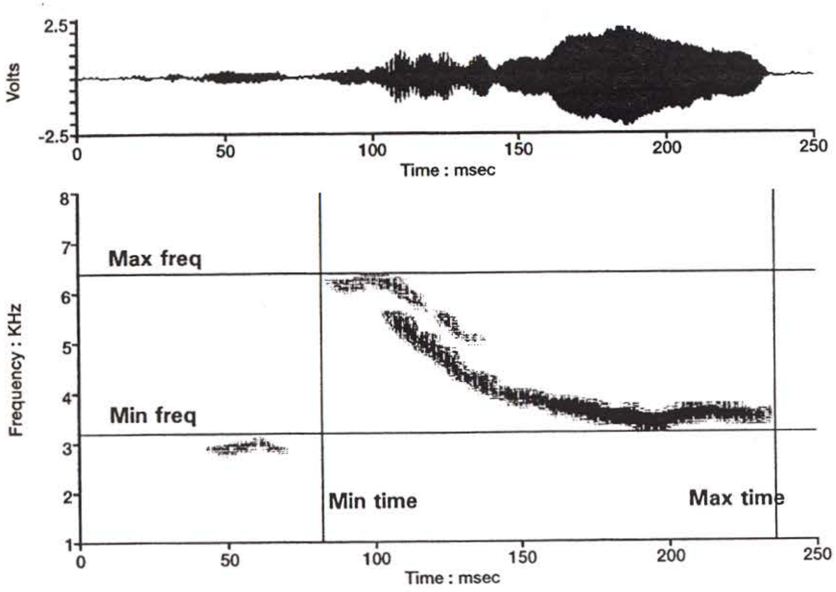

Fig. 9. The onset and offset times and minimum and maximum frequency coordinates of a swamp sparrow note (*top panel*) are measured from its spectrogram (*bottom panel*)

6 Classification of Naturally Occurring Animal Sounds

In the remainder of this chapter it will be useful to refer to specific sound-types based on their distinctive acoustic features. The terms used for these signals are drawn from the working nomenclature of bioacoustics researchers. Eight basic sound-types are delineated, reflecting both temporal and spectral structures. Each is briefly described below in the context of specific examples from various species, and illustrated by waveforms and spectrograms in Figure 10.

6.1 Properties of Ideal Signals

6.1.1 *Periodicity*

A brief explanation of the applicable, engineering-based terminology is also helpful (see Gerhardt, this Volume, for a more complete discussion). In this literature a distinction is made between periodic and non-periodic signals. *Periodic* signals are characterized by repeated amplitude cycles in the time-waveform. In a naturally occurring biological signal, such periodicity typically represents the repeated operation of a sound-producing mechanism like vibration of a membrane (e.g., syringeal sound production in birds; see Gaunt and Nowicki, this Volume) or in paired tissue structures (e.g., vocal folds underlying laryngeal sound production in mammals, including nonhuman primates and humans; see Rubin and Vatikiotis-Bateson, and Owren and Bernacki, this Volume).

Periodic sounds are classified as being either *simple* (sinusoidal or pure-tone), containing energy at a single frequency, or *complex*, containing energy at multiple frequencies. In the latter, spectral energy occurs in the *fundamental frequency* (corresponding to the basic rate of vibration in the underlying sound-producing structure), and its *harmonics* (spectral components occurring at integer multiples of the fundamental frequency). Periodic sounds are contrasted with *aperiodic* sounds whose amplitude variation does not repeat cyclically. Aperiodic sounds appear noisy, and their spectra cannot be expressed as a fundamental frequency and its harmonics. Aperiodic sounds can result from a single, sharp impulse like a click, or from inherently unpatterned, turbulent airflow. Note that these sound types are idealized in that no naturally occurring sound would be likely to be a pure tone, nor to repeat exactly from cycle to cycle.

6.1.2 *Amplitude Modulation*

Amplitude modulation (AM) refers to changes in the overall energy level of a signal occurring over time (see Stremler 1977). As used here, this term refers to any periodic variation in intensity about an average intensity level, or slow changes in intensity that may not be cyclical in nature. The frequency of the energy component whose intensity

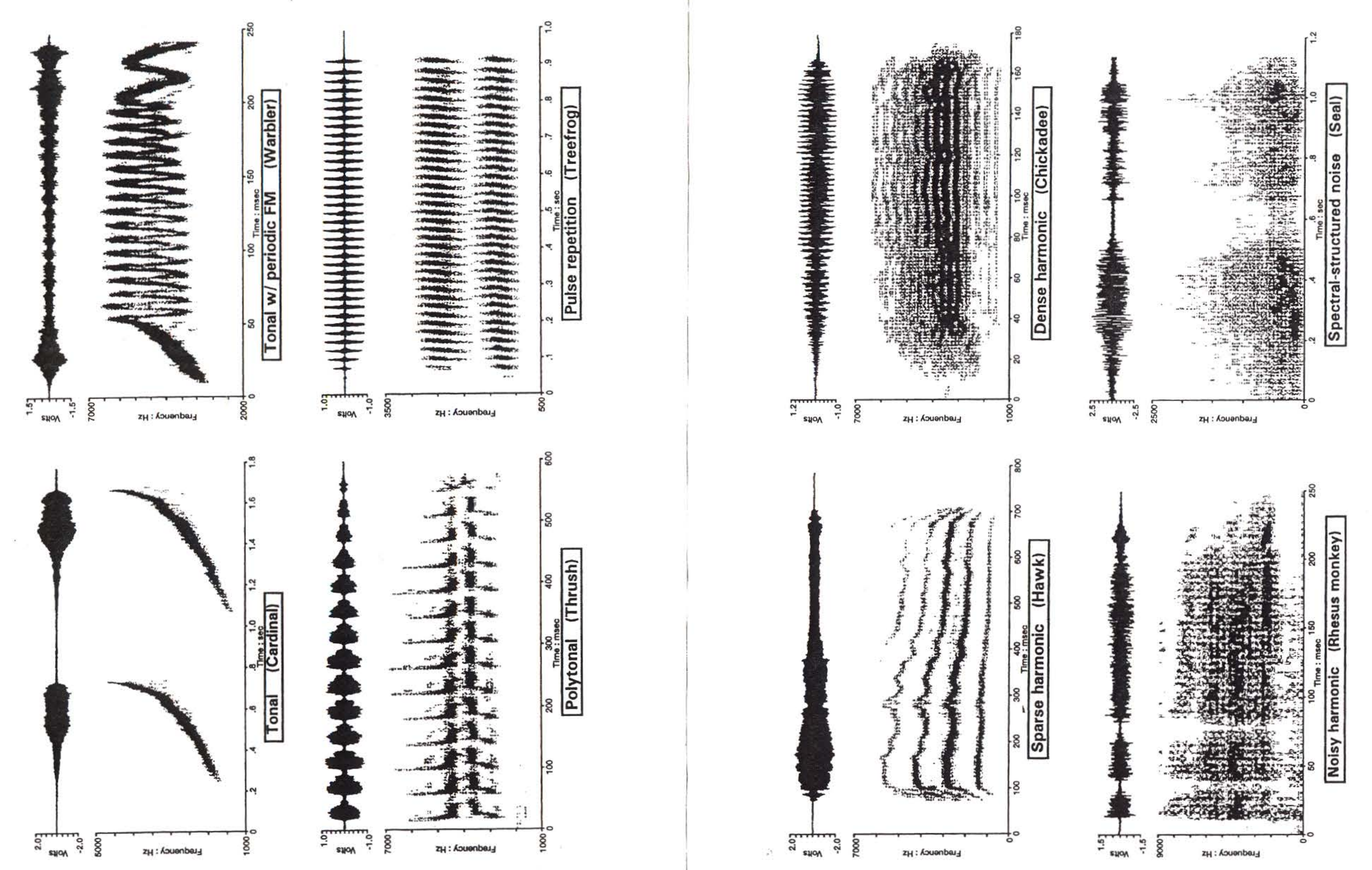

Fig. 10. Classification of eight basic sound types based on biologically relevant temporal and spectral properties. Representative waveforms (*top panels*) and spectrograms (*bottom panels*) are shown for each

is changing is called the *carrier frequency*, the rate of envelope variation is the *modulation frequency*, and the magnitude of variation is expressed by the *modulation index*. Thus in Figure 20, a 1000-Hz signal whose peak amplitude varies between 0.80 and 1.20 at a rate of 100 times per second has a carrier frequency of 1000 Hz, a modulation frequency of 100 Hz, and a modulation index of 0.20. AM patterning is readily produced and is often of functional significance in bioacoustic signals, for instance as an important component of sound character, or in providing clues about the physical nature and operation of the sound-producing mechanism.

6.1.3
Frequency Modulation

Frequency modulation (FM) refers to changes in the instantaneous frequency of a signal over time (see Stremler 1977). As used here, it refers to any periodic variation of instantaneous frequency about an average frequency value, or overall frequency changes that are not cyclical in nature. In a signal that varies in a periodic fashion, the center frequency of the modulated energy component is called the *carrier frequency,* the rate of variation is the *modulation frequency,* and the frequency magnitude of the variation is the *modulation depth.* Thus, a signal that varies between 980 and 1020 Hz at a rate of ten times per second has a carrier frequency of 1000 Hz, a modulation frequency of 10 Hz, and a modulation depth of 20 Hz.

6.1.4
Biologically Relevant Sound Types

In addition to the above terminology, bioacousticians have devised terms to express the various sound structures in naturally occurring communication signals. These are described here and illustrated in Figure 10 in order of increasing temporal and spectral complexity. *Tonal* sounds contain a single, *dominant* frequency component at each time instant, although frequency and amplitude may vary with time. A dominant frequency is one that is significantly higher in amplitude than any other frequency component present and is often the perceptually salient spectral component. Note, however, that the perceptual significance of any aspect of a bioacoustic signal can only be examined through explicit testing, as described in this Volume by Cynx and Clark, and Hopp and Morton, who examine laboratory- and field-based methods, respectively. Tonal sounds, which often resemble whistles, include many bird songs and signals produced by marine mammals (e.g., dolphins and beluga whales, *Delphinapterus leucas*). *Tonal sounds with periodic FM* are single-frequency sounds that show regular, often rapid pitch variation, whose auditory quality can resemble a harsh buzz rather than a whistle. Because they exhibit only one frequency at a time, they are considered tonal and are commonly observed among birds, especially warblers. The dominant frequency in both types of tonal sounds may represent a fundamental or a higher-level harmonic. For instance, Gaunt and Nowicki (this Volume) discuss how filtering in the vocal tract of songbirds can shape the sound by suppressing energy at the fundamental, leaving the dominant frequency at a higher harmonic.

Polytonal sounds contain two or more spectral components that are not harmonically related, representing the operation of two or more independent periodic sound-producing mechanisms. An example is the two-voice phenomenon of birdsong, also discussed by Gaunt and Nowicki (this Volume).

Pulse-repetition signals consist of a series of energy bursts whose acoustic structure may be constant or may vary between pulses. The pulses themselves may be tonal, show multiple frequency components, or exhibit broadband spectral energy, and may be amplitude- or frequency-modulated. These sounds are produced by a variety of species, including anurans, marine mammals, birds, and nonhuman primates. Their audible quality is often a repetitive buzz and can incorporate underlying pitch variation over the pulse sequence.

Sparse-harmonic sounds contain a relatively small number of harmonically related spectral components, and range in sound quality from a harsh whistle to a light rasp. *Dense-harmonic* signals contain a larger number of spectral components and sound more harsh or nasal than sparse-harmonic sounds. Examples include chickadee *dee* notes, humpback whale (*Megaptera novaeangliae*) cries, nonhuman primate calls, and human speech sounds in which vocal fold vibration (*voicing*) occurs.

Noisy-harmonic sounds consist of a combination of tonal or harmonic components with significant additional noise. These signals have a harsh auditory quality in proportion to their noise content. Examples include nonhuman primate vocalizations, and some calls produced by starlings and crows.

Spectrally structured noise consists of noise energy with one or more time-varying spectral peaks or other coherent spectral structure. Examples include nonhuman primate *screams* and *barks*, Mexican chickadee (*Parus sclateri*) calls, and fur seal vocalizations. Sounds like nonhuman-primate screams that show pronounced spectral peaks can give an audible impression of time-varying pitch. In other cases, the sound quality may be of a grunt, bark, or roar with little pitch impression.

7
Time-varying Frequency Analysis

Many sounds involve frequency variation over time, and this variation is often significant in the communication process. Frequency-time analysis involves the measurement and characterization of such variation, and is concerned with the duration, range, and patterning of spectral changes including, for example, maximum and minimum frequencies and the spectral patterns of different sound types. Tonal and harmonic sounds, in particular, can be characterized effectively by their time-varying dominant frequency, which will be referred to as the *spectral contour* of a signal (Beeman 1987, 1996b). Mathematically, the contour value at each instant is the frequency containing the most energy, and this can be visualized as the frequency "spine" of the sound's spectrogram. Depending on the sound, the spectral contour may represent the fundamental frequency or a higher harmonic.

Spectral contour analysis represents a powerful tool for characterizing tonal and harmonic signals. The resulting function can be measured, manipulated, and compared statistically in the same way as the sound's amplitude envelope. It is more comprehensive than individual sound parameters, and mathematically more tractable than the spectrogram as a whole. Once derived, spectral contours can be used to automatically extract and measure frequency-time features such as contour minima and maxima, rate of frequency change over time, frequency range, and signal duration. Contours can be used as the basis for similarity analysis, which can be advantageous for harmonic or noisy sounds that are less amenable to similarity measurement by cross-correlation of the entire spectrogram (discussed below).

7.1
Deriving Spectral Contours

Any means of determining the instantaneous dominant frequency of a sound can produce a spectral contour function. This section discusses several approaches. More extended discussion and additional references are provided by Beeman (1996b).

Dominant frequency extraction can be performed by determining the frequency at which each column of a digital spectrogram achieves maximum amplitude, then recording these frequency values as a continuous time function – a technique referred to here as *spectrogram contour detection* (SCD; see Beeman 1987, 1996b; Buck and Tyack 1993) and illustrated in Figure 11. Because SCD is based on Fourier analysis, it is inherently frequency-selective and contours can be extracted successfully from time signals containing both multiple harmonics and broadband noise. In fact, SCD is sufficiently immune to noise components that it has been used successfully to extract noise-free playback stimuli from field recordings (see below).

SCD may not function effectively on sounds whose dominant frequency jumps between harmonics [such as human speech, nonhuman primates, or the black-capped chickadee's (*Parus atricapillus*) *dee* note], or which contain noisy components with time-varying spectral peaks. For such sounds, the frequency values returned may jump between harmonics or momentary peaks in the noisy spectrum. In these cases, accurate frequency contours can nonetheless be obtained by *hand-drawing* over a frequency-magnified spectrogram display using a mouse or some other tracing method (Beeman, 1996b). Existing contours can be edited in the same manner for smoothing or error correction.

Zero-crossing analysis (ZCA), can also be used to derive spectral contours, and can yield contours with very high time and frequency resolution on tonal sound material. The limitation of ZCA is that it cannot be used effectively on sounds containing harmonic or noise energy. However, ZCA can be a useful technique, as noted with respect to sound synthesis below. *Hilbert transform analysis* (Dabelsteen and Pedersen 1985; Oppenheim and Schafer 1989) applies the mathematics of complex variable theory to decompose any tonal or harmonic time signal into the product of a frequency function $F(t)$ and an amplitude function $A(t)$. With tonal sounds, Hilbert analysis produces precise and useful results, but like ZCA, it degrades in the presence of harmonic and noise

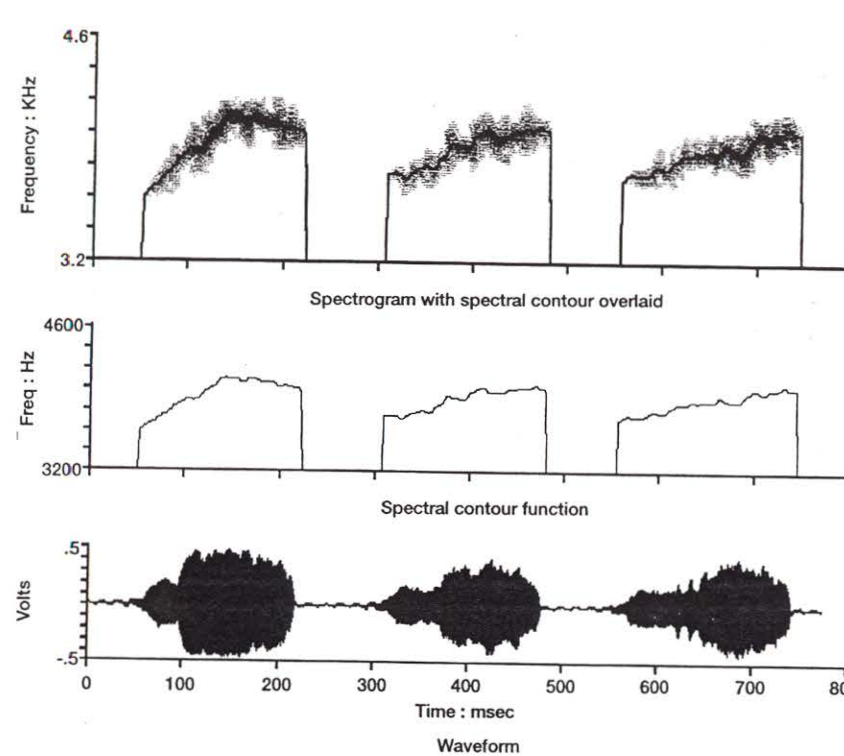

Fig. 11. Spectral contour detection applied to three swamp sparrow notes, showing spectrographic representations with the extracted pitch contours superimposed (*top panel*), the pitch contours alone (*middle panel*), and the waveforms (*bottom panel*).

energy. Like ZCA, Hilbert analysis is not the method of choice in most biological applications, but can perform well on tonal or nearly tonal material

7.2
Sound Similarity Comparison

Researchers have long sought to gauge the overall similarity of various sounds, for instance to relate signal acoustics to ontogeny, individual variation, geographical location, and species identity. Sound comparison has traditionally been performed either by comparing sounds on a qualitative basis (e.g., Borror 1965; Kroodsma 1974; Marler and Pickert 1984), or by reducing sounds to significant parameters that can be compared statistically (e.g., Nelson and Croner 1991). The first approach includes judgments involving auditory comparison of sounds and visual comparison of spectrograms, both of which can include bias related to human perceptual processing, and thereby lack objectivity, repeatability, and mathematical foundation. The second approach is better in these respects, but reducing a complex sound to a limited set of parameters may fail to capture all its salient aspects and parameter choice is subject to bias. Nonetheless, waveforms, power spectra, and spectrograms have all been used as sources for a wide variety of parameters.

Parameter-based sound similarity has been measured in a number of ways. Nelson (1992), for instance, used the simple Euclidean distance between parameter vectors (i.e., the square-root of the sum of the squared differences between corresponding parameter values) as a similarity metric. Parametric and nonparametric analyses of variance can be used to determine which features differ statistically across different sound types. Various multivariate statistical techniques (see Sparling and Williams 1978 for a review) have been applied as well, including discriminant-function analysis (Hauser 1991) and principal-components analysis (Martindale 1980; Clark 1982; Nelson and Croner 1991). These approaches have used the *proportion of variance explained* as a measure of the significance of various sound features and have classified sounds on the basis of statistical clustering on these features. Multidimensional scaling can also be used to portray the clustering tendencies within a dataset, for instance based on pairwise similarities between sounds (Clark et al 1987; Hauser 1991).

Because spectrograms provide an intuitive overview of the spectral and temporal characteristics of a sound, they are a desirable basis for quantitative similarity comparisons between signals. With digital spectrograms, similarity can be measured by comparing the complete numerical matrices underlying the visual representations. This approach has the advantage of being objective and repeatable and of including the whole sound rather than individual, measured parameters. Quantitative similarity can be measured by calculating the normalized covariance between two spectrograms at successive time-offsets (Clark et al. 1987). This process can be visualized as sliding two transparencies containing the spectrograms over each other along the time axis, and measuring the degree of overlap between the images at each point. The resulting correlation function, *R(T)*, is a sequence of correlation values representing spectrogram similarity as a function of time-offset. The peak value of *R(T)* quantifies the maximum similarity between the two spectrograms.

In this approach, the spectrograms are first normalized for overall amplitude and time-offset becomes a variable (*normalization* refers to the process of removing the variation in one sound feature from an ensemble of sounds, rendering the signals equivalent on that particular parameter). The maximum-similarity value is then independent of overall amplitude differences or temporal offset in the original digital signal files. Figure 12 shows a spectrogram-based cross-correlation comparison between two swamp sparrow (*Melospiza georgiana*) syllables, showing the maximum correlation value and time-offset. Because digital spectrograms are based on Fourier transforms, which have inherent immunity to spectral noise, the spectrogram-similarity technique can be used effectively on quite noisy signals. Furthermore, the comparison can be restricted to a specified bandwidth within the spectrograms, thereby excluding out-of-band extraneous noise.

In general, quantitative spectrogram-similarity comparison provides a sensitive and comprehensive measure for sounds that generally resemble each other. However, the

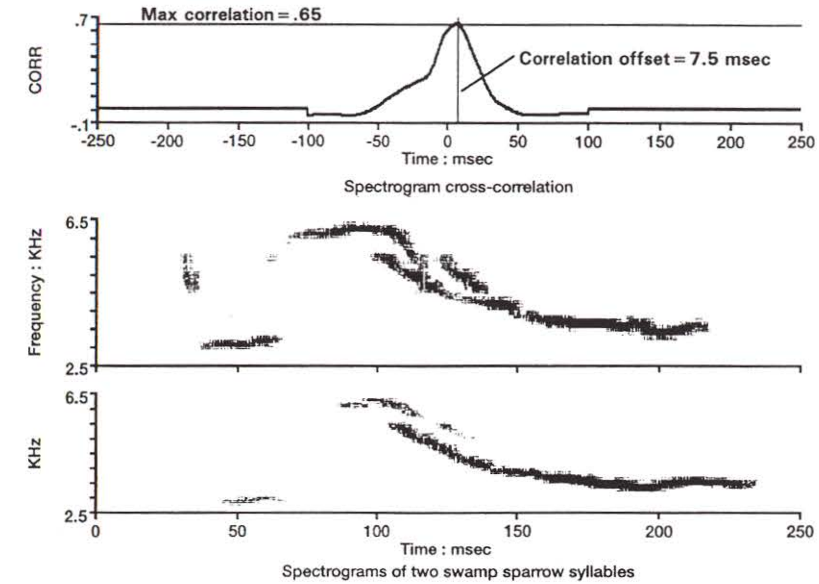

Fig. 12. Quantifying the overall similarity between two swamp sparrow syllables by cross-correlation (*top panel*) of their spectrograms (*middle* and *bottom panels*)

technique can have severe limitations, even for quite similar sounds, unless normalization is performed with respect to characteristics such as mean frequency, frequency range (maximum minus minimum frequency), duration, or harmonic interval. In the absence of normalization, variation in these features may produce low cross-correlation values for apparently similar sounds and mask biologically important similarities. Note, however, that removing a parameter's variation from the similarity calculation also removes that parameter's influence from the analysis. It is therefore important to study clustering tendencies in a dataset with various parameters both included and excluded, as well as specifically testing the behavioral importance of these parameters empirically.

Recently developed tools address some of the limitations noted above (Beeman 1996b). For example, mean frequency differences between individual signals in an ensemble can be removed by cross-correlating their power spectra and frequency shifting the associated spectrograms before cross-correlation. Similarly, differences in signal duration can be removed by uniform linear expansion of the time-scale of each spectrogram as needed before comparison. Alternatively, sounds can be compared on the basis of mathematical functions that are more comprehensive than extracted parameters, but more flexible than spectrograms. Such functions include amplitude-time envelopes, pulse-repetition patterns, spectral contours, and power spectra (Beeman

1996b). For example, differences in frequency range and harmonic interval can be overcome by reducing sounds to spectral contours, performing the necessary normalization, and then calculating similarity.

8 Digital Sound Synthesis

Digital signal processing provides a wide variety of techniques for sound *synthesis*, which will be considered to include any significant creation or manipulation of sound material, whether based on natural or mathematically specified signals. General approaches include digital *editing*, *arithmetic manipulation*, and *generation* of sound, each of which provides the ability to create precise, repeatable, and incremental signal variations. Editing is manipulation of existing sound material by rearranging waveform segments in the time domain. Arithmetic manipulation includes both combining and altering signals through such operations as addition and multiplicative amplitude scaling using a specific function of interest. Sound generation – creating new waveforms – is a particularly important technique that can combine elements of both editing and arithmetic manipulation with de novo creation of segments or entire signals.

8.1 Editing

Sound editing, as used here, means rearranging existing signal components in time, without generating new physical waveforms. Ideally, a digital time-signal editing system should be visually based, since the process involves moving sound segments from one location to another. Beginning with source and target waveforms on the computer screen, the user graphically selects a signal segment and an editing operation to perform. The system should provide basic cut-and-paste operations, allowing individual signal components to be extracted, inserted, deleted, and concatenated. These processes are illustrated in Figure 13. Editing commands can be used to extract or reorder signals for analysis, or to alter signals for use in playback experiments, for instance in combination with synthesis techniques that can be used to alter the temporal and spectral characteristics of the signal.

Bioacousticians have long desired a digital *spectrogram editor* that would be capable of extracting, moving, or deleting specified segments in *both* the time and frequency domains. Such an editor would be able to shift a signal's energy components in the frequency domain, selectively erase or redraw particular harmonics, and the like (Zoloth et al. 1980), then generate the corresponding time signal. However, such capabilities have not developed far because the approach itself is fundamentally problematical. Frequency-based editing changes can cause severe mathematical artifacts in the corresponding time signals. For example, changes in frequency characteristics can produce amplitude modulation in the waveform, while many frequency-band deletions simply cannot be transformed back to the time domain. More practicable approaches can sometimes be used to produce the desired results, including spectral-filtering and

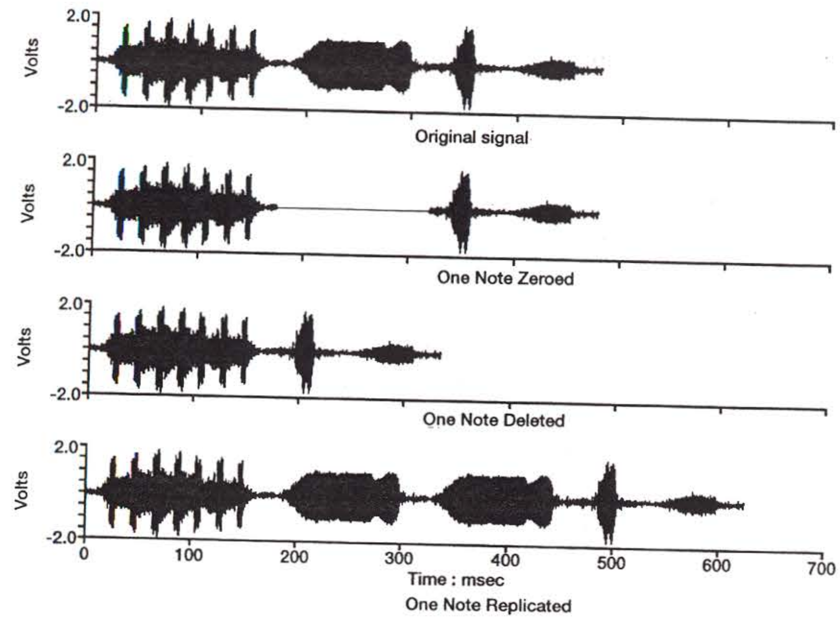

Fig. 13. Using digital signal editing to alter the waveform of a sparrow song (*top panel*) to set one note to zero (*top middle panel*), delete the resulting silent interval (*bottom middle panel*), or replicate that note (*bottom panel*)

waveform-conditioning techniques discussed in the following sections (also see Stoddard, this Volume).

8.2 Arithmetic Manipulation and Generation of Sound

The arithmetic manipulation of existing waveforms and the generation of new sounds are often combined in practice. In fact, many of the functions of the "mythical" spectrogram editor can be implemented using combinations of the techniques described below. Complex signal characteristics like amplitude-envelope shape, spectral composition, AM, FM, and phase relationships of individual components can all be modified with great precision. Furthermore, temporal and spectral features can be independently shifted and expanded. The functions used to alter or synthesize a given signal or signal element can be based on naturally occurring sounds, or created through graphical approaches like mouse-based drawing, or derived from mathematical functions.

8.3 Synthesis Models

There are many synthesis models currently in use. Each embodies different assumptions about sound structure and sound production, and as a result, different models synthesize different sound types better than others. The *tonal model*, which is described in depth below, represents sounds as frequency- and amplitude-modulated sinusoids. Because of its mathematical tractability, this model has received much attention in animal-related studies, and works well for a wide range of natural sound material, including whistles, rapid AM and FM buzzes, harmonic sounds, pulse-repetition signals, and some noisy-harmonic sounds.

Another approach models critical aspects of the physiological processes involved in the production of the sound. For instance, many aspects of speech production have been modeled in this fashion, as described by Rubin and Vatikiotis-Bateson (this Volume). One classic example of this type of *articulatory model* is the *Klatt synthesizer* (Klatt 1980), which has been extensively used in speech research. This software model passes voiced sound (based on periodic energy bursts designed to mimic vocal-fold vibration pulses) and unvoiced sound (noisy energy that mimics turbulent air flow) through a bank of parallel filters that simulates the physical resonances of the human vocal tract. The parameters of these filters are continuously varied, thereby representing the time-varying resonances resulting from the articulatory maneuvers of speech. The Klatt synthesizer has also been used with nonhuman primates to test the auditory processing of both human speech sounds (e.g., Hienz and Brady 1988; Sinnott 1989) and species-specific vocalizations (e.g., May et al. 1988; Hopp et al., 1992).

8.3.1 *Tonal Model*

A wide variety of bioacoustic signals can be modeled as tonal or harmonic, and as a result can be readily represented, manipulated, and synthesized (e.g., Greenewalt 1968; Dorrscheidt 1978; Margoliash 1983; Beeman 1987, 1996b). Such sounds are mathematically tractable and form versatile building blocks for more complex sounds. The tonal model provides techniques used to synthesize tonal sounds, pulse-repetition sounds, and harmonic sums of tonal sounds, and allows for extensive mathematical manipulation of amplitude, temporal, and spectral sound properties.

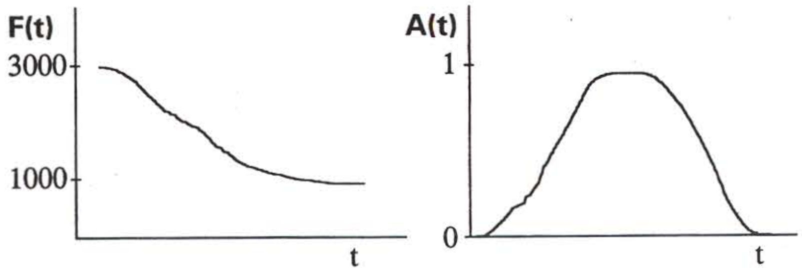

Fig. 14. *F(t)* and *A(t)* functions for a tonal sound that descends in frequency from 3000 to 1000 Hz, while first rising and then falling in amplitude

Due to the simplicity of its structure, a tonal sound can be completely represented by two time functions – a spectral contour, *F(t)*, which represents its time-varying frequency, and an amplitude envelope, *A(t)*, representing its time-varying intensity. As an example, the functions shown in Figure 14 describe a sound whose frequency descends from 3000 to 1000 Hz, and whose amplitude rises from 0 to a maximum level, remains steady, and then declines to 0.

F(t) and A(t) are used to synthesize sounds as follows. Both functions are represented as time signals, with values specifying instantaneous frequency (in Hz) and intensity (in Volts), respectively. To recover the original signal, F(*t*) is passed through a voltage-to-frequency sinewave generator, producing a constant amplitude signal of time-varying frequency F(t). This signal is then multiplied by A(*t*), giving it the appropriate amplitude characteristics. This process is schematically illustrated in Figure 15. Mathematically, the process is justified by Hilbert transform theory, which shows that any signal can be decomposed into frequency and amplitude functions, and then recovered from these functions without loss of information (Oppenheim and Schafer 1989). Because the frequency and amplitude functions are independent, frequency and amplitude features can be independently modified, allowing their communicative significance to be tested separately.

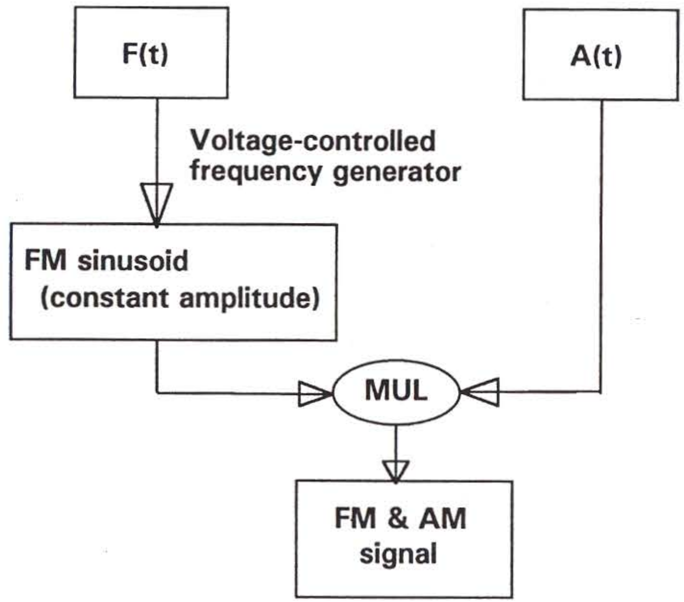

Fig. 15. How *F(t)* and *A(t)* functions are used in conjunction with a voltage-to-frequency sinewave generator to produce a synthetic, tonal sound. *MUL*= Multiply

8.4 Sources of F(t) and A(t) Functions

8.4.1 *Mathematically Based Functions*

F(t) and A(*t*) functions can be based on elementary mathematical functions such as *constants, linear ramps, exponential ramps, sinusoids,* and combinations of these. Figure 16 shows five different *F(t) functions*, together with the resulting time waveforms. The constant function produces a single, unchanging frequency component. Linear- and exponential-ramp functions produce signals that change in frequency at a constant rate (in Hz per second) and at an exponential rate (in octaves per second), respectively. A sinusoidal function produces a signal that changes frequency at a sinusoidal rate, for instance ten times per second. Custom functions allow the user to create a wide variety of FM signals through any combination of mathematical functions and natural sound material.

A(t) functions are created in a similar manner. Figure 17 shows *rectangular, trapezoidal taper, exponential taper, cosine taper,* and *sinusoidal* functions, and the resulting

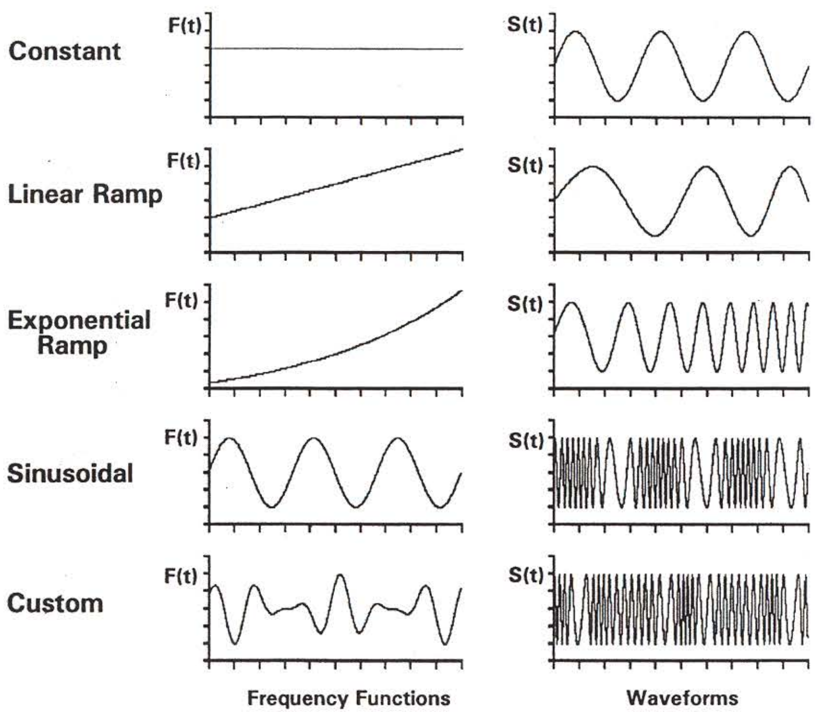

Fig. 16. Mathematical *F(t)* functions used in sound synthesis to produce frequency-modulated waveforms, *S(t)*

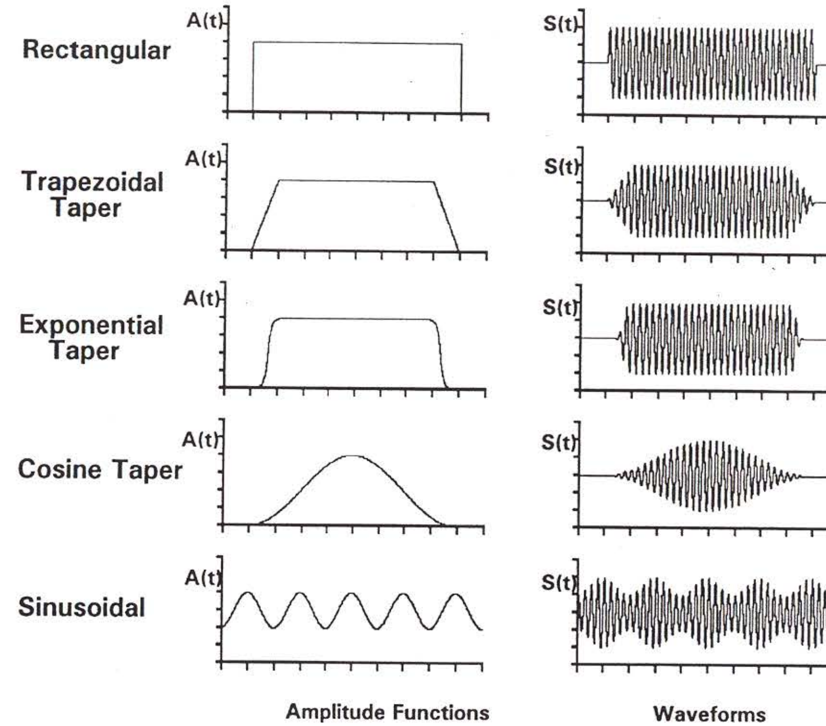

Fig. 17. Mathematical *A(t)* functions used in sound synthesis to produce amplitude-modulated waveforms, *S(t)*

time waveforms. The rectangular function produces a waveform whose envelope changes amplitude instantaneously at onset and offset, with constant amplitude during the signal. The trapezoidal taper function produces a waveform whose amplitude changes linearly between zero and its maximum value at onset and offset. The exponential taper function produces waveform onset and offsets whose amplitudes change exponentially. The cosine taper function produces a short, heavy taper, with minimum discernible onset or offset. Finally, sinusoidal amplitude modulation is produced by a sinusoidal function, and results in periodic variation in the waveform envelope. Amplitude functions can contribute to sound quality (as in amplitude modulation), or they can simply be used to taper the onsets and offsets of synthesized stimuli before experimental use. Overly sharp onsets and offsets (i.e., rising or falling to full amplitude within in a few milliseconds) generate perceptible pops on playback, in addition to injecting spurious spectral energy, and should be avoided.

F(t) and *A(t)* functions can be combined and transformed using various mathematical operations to generate sound features like harmonic complexes, modulation, and amplitude scaling. Because they are universal, objective, and readily described, synthetic sounds based on mathematical functions are widely used as stimuli for testing responses at the neurophysiological level. However, they are generally less successful as approximations of naturally occurring biological sounds.

8.4.2
Functions Derived from Natural Sounds

F(t) and *A(t)* functions can be derived directly from tonal and harmonic natural sounds, to allow the researcher to precisely modify frequency or amplitude features for testing (Margoliash 1983; Beeman 1987, 1996b). F(*t*) can be derived from the time waveform, using zero-crossing or Hilbert transform techniques, or from the spectrogram, using the SCD technique. A(*t*) can be derived from the amplitude envelope of the time waveform, or by Hilbert transform techniques. F(*t*) and A(*t*) functions of any shape can also be drawn by hand using the mouse.

While synthesis from mathematical functions involves generation and synthesis – *F(t)* and *A(t)* are generated then combined – synthesis from natural sounds begins with an analysis step, in which *F(t)* and *A(t)* are derived from the source sound. The process of analyzing *F(t)* and *A(t)* from natural sounds is summarized in the top panels of Figure 18.

9
Sound Manipulation and Generation Techniques

Once the *F(t)* and *A(t)* functions have been selected or derived, manipulation of an existing sound or generation of an entirely new signal can occur. The following sections describe these manipulations, includw'ng frequency shifting, time scaling, amplitude and frequency modulation, intensity manipulation, harmonic removal and rescaling, pulse repetition variation, phase adjustment, noise addition and removal, and the derivation of template sounds. Further details are described by Beeman (1996b). In the simple example shown in Figure 18, the synthesis manipulation consists of time-reversing the sound. Based on the original sound's spectrogram and time waveform (at the top of the figure), respectively, the sound is first analyzed into *F(t)* and *A(t)* functions. These functions are then reversed to produce the new, altered functions *F'(t)* and *A'(t)*. *F'(t)* is used to synthesize a constant-amplitude sinusoid representing the frequency variation, which is then multiplied by *A'(t)*. The spectrogram and waveform of the synthetic sound are shown at the bottom of the figure.

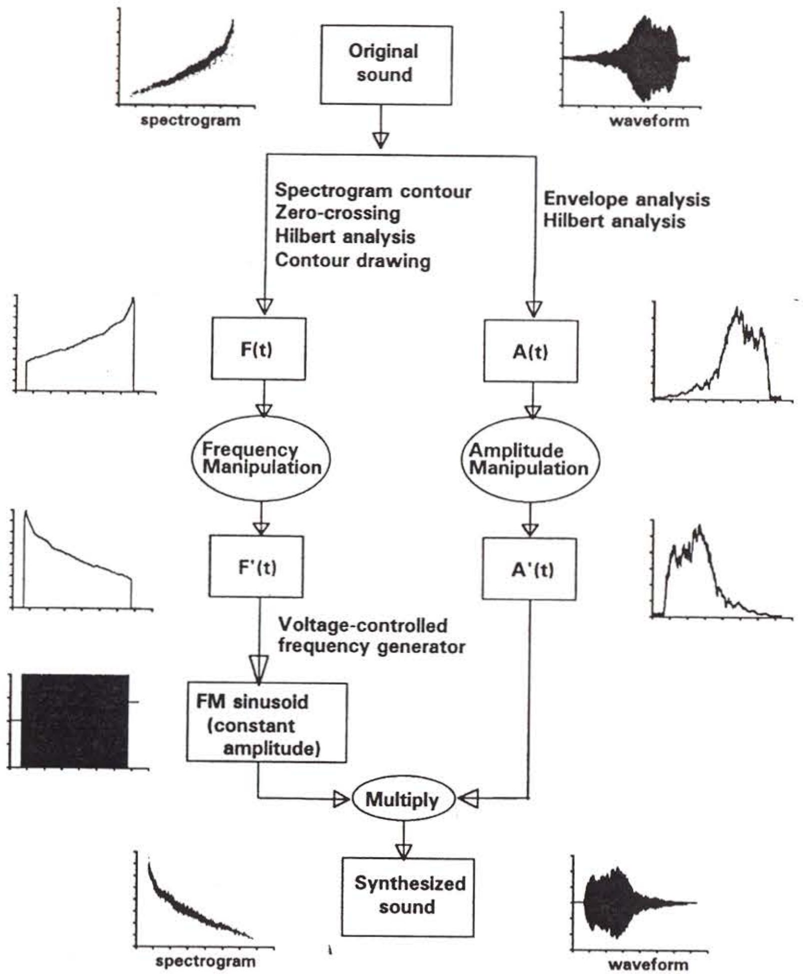

Fig. 18. The complete analysis-manipulation-synthesis process. *F(t)* and *A(t)* functions are extracted from a natural sound, reversed in time (in this example), and then resynthesized to produce a time-reversed synthetic version of the original signal

9.1
Duration Scaling

Sound duration can be uniformly expanded or compressed without altering frequency characteristics, by expanding or compressing both the *F(t)* and *A(t)* functions via linear interpolation. Attempting to alter temporal features by resampling the time waveform

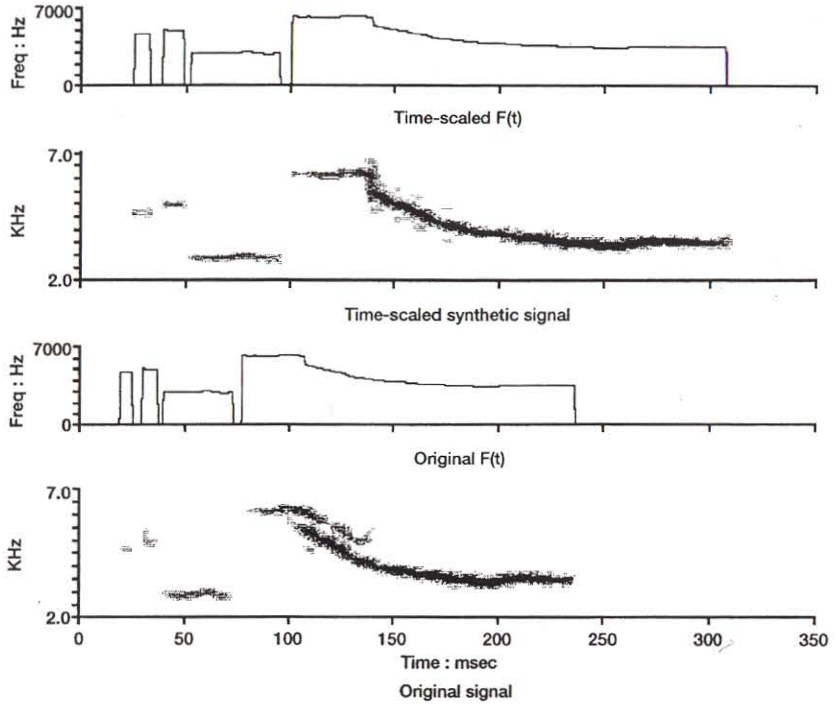

Fig. 19. A synthetic version of the swamp sparrow syllable shown in Fig. 1 is increased in duration by 30 % by time-expanding its $F(t)$ and $A(t)$ functions

itself would introduce severe artifacts. Instead, all manipulations are performed on $F(t)$ and $A(t)$, and artifacts are avoided. Figure 19 shows a swamp sparrow syllable (the same sound as shown in Figure 1) whose duration has been increased by 30 % by this method, without changing its frequency characteristics. This concept can be extended to non-uniform time expansion or compression, a technique sometimes called *dynamic time warping*, which involves a variable stretching or contracting process.

9.2 Amplitude Envelope Manipulations

A signal's amplitude envelope may include both short-term periodic and long-term non-periodic variation. The former can often be modeled sinusoidally, while the latter may represent signal onset or offset, phonetic emphasis, and other slowly varying effects. As illustrated in Figure 20, a signal's overall amplitude envelope, $A(t)$, can be mathematically modeled as the sum of a slowly varying envelope, $A_s(t)$ (a constant value of unity in the figure), and a rapidly varying modulation function, $A_m(t)$, so that $A(t) = A_{s(t)} + A_{m(t)}$. $A_{s(t)}$ is derived from $A(t)$ by smoothing, to remove the rapid modulation. The smoothing window should be wider than the modulation period, but short enough to retain the time-varying shape of $A_s(t)$. A 15-ms window, for example, might be appropriate for a 100-Hz (i.e., 10-msec period) modulation rate. $A_m(t)$ is then obtained by subtracting $A_s(t)$ from $A(t)$. $A_s(t)$ and $A_m(t)$ can then be analyzed independently.

Amplitude envelopes can be manipulated in other ways as well. As noted earlier, RMS energy can be equalized among a set of signals, for example to equalize the intensity levels of a series of playback signals created from natural sounds. Alternately, amplitude variation can be removed from a signal entirely by dividing the signal by its $A(t)$, leaving a signal of uniform amplitude but time-varying frequency. Note that this ma-

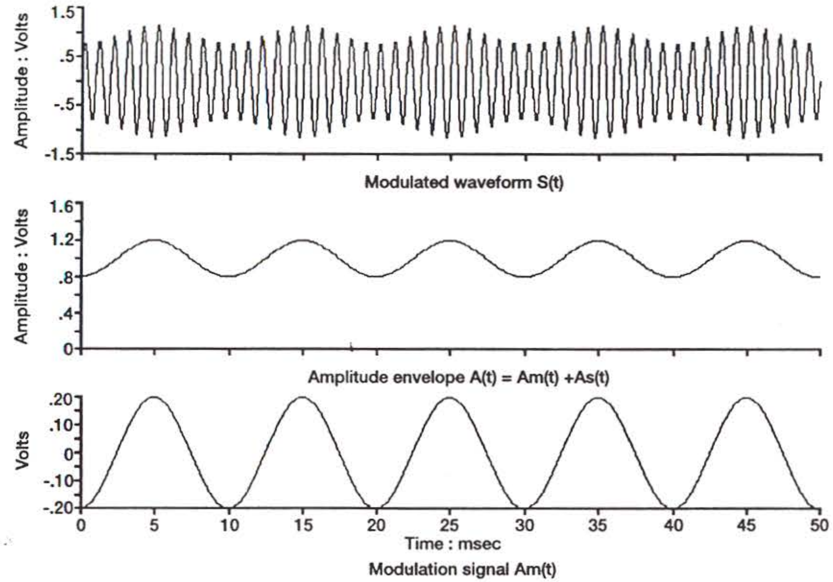

Fig. 20. An amplitude-modulated waveform [$S(t)$, *top panel*] and its complete amplitude envelope [$A(t)$, *middle panel*] and amplitude modulation function [$A_m(t)$, *bottom panel*)]

nipulation will amplify any low-level noise segments in the signal. Finally, all signals in an ensemble can be given the same amplitude envelope $A_{ref}(t)$, by multiplying each signal envelope $A_n(t)$ by the function $A_{ref}(t) / A_n(t)$.

9.3
Spectral Manipulations

9.3.1
Frequency Shifting and Scaling

Tonal synthesis allows spectral changes to be made without altering temporal or amplitude relationships, because $F(t)$ can be modified independently of $A(t)$ and other time characteristics. Uniform frequency shifting is performed by adding a constant positive or negative value to $F(t)$, while leaving $A(t)$ unchanged. Because the adjustment is made to the frequency function rather than the waveform or spectrogram, the output signal is free of artifacts (such as amplitude modulation) that have been a major shortcoming of these other approaches. Figure 21 shows a swamp sparrow syllable whose frequency has been uniformly raised by 500 Hz without altering overall duration. Frequency can be shifted logarithmically as well, for instance by multiplying $F(t)$ by a factor

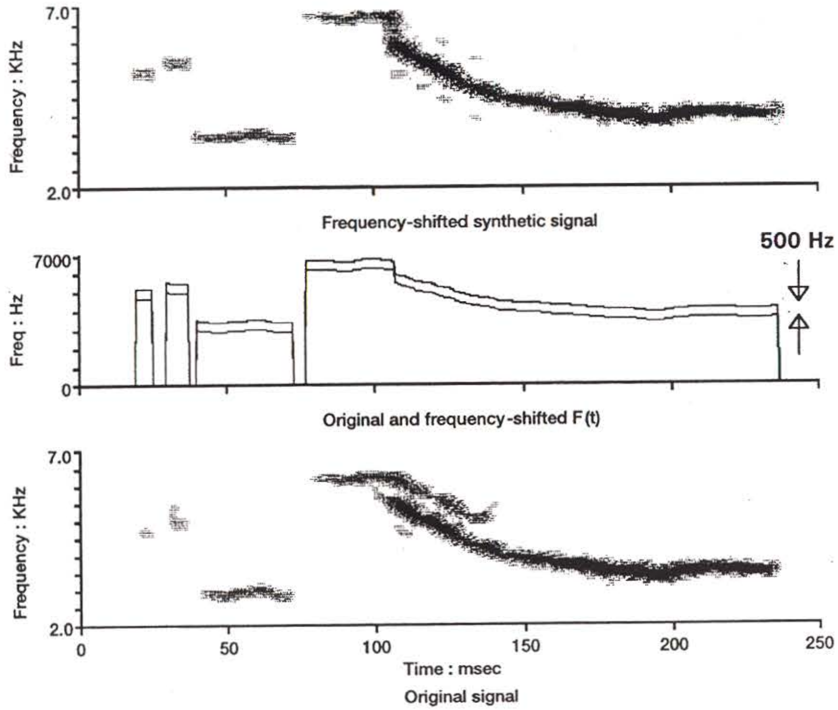

Fig. 21. A synthetic version of the swamp sparrow syllable shown in Fig. 1 is shifted upward in frequency by 500 Hz by adding 500 to its $F(t)$ function

of 2 to produce an increase of one octave. $F(t)$ can also be compressed or expanded in range to alter the signal's frequency variation about its mean frequency.

9.3.2
Frequency Modulation

Any FM characteristics modeled in an $F(t)$ function can be altered, replaced, or even removed. The general approach is to analyze the signal into two components, a slowly varying carrier frequency $F_c(t)$ and rapidly varying frequency modulation $F_m(t)$, so that $F(t) = F_c(t) + F_m(t)$, as described by Beeman (1996b). FM parameters are then determined from these functions. Figure 22 shows an Eastern phoebe (*Sayornis phoebe*) note analyzed in this manner into functions representing the total contour, $F(t)$, its carrier frequency, $F_c(t)$, and its frequency modulation, $F_m(t)$, about the carrier. One can then measure from $F_m(t)$ the signal's mean modulation frequency (about 86 Hz) and mean modulation depth (about 416 Hz). These functions can then be altered to produce a synthetic sound with altered characteristics. Independent changes to $F_c(t)$ and $F_m(t)$ can be used, for example, to increase or decrease the signal's overall frequency change by scaling $F_c(t)$, or to double the magnitude of modulation depth by multiplying $F_m(t)$ by a factor of 2. $F_c(t)$ and $F_m(t)$ are then recombined to produce a new $F(t)$.

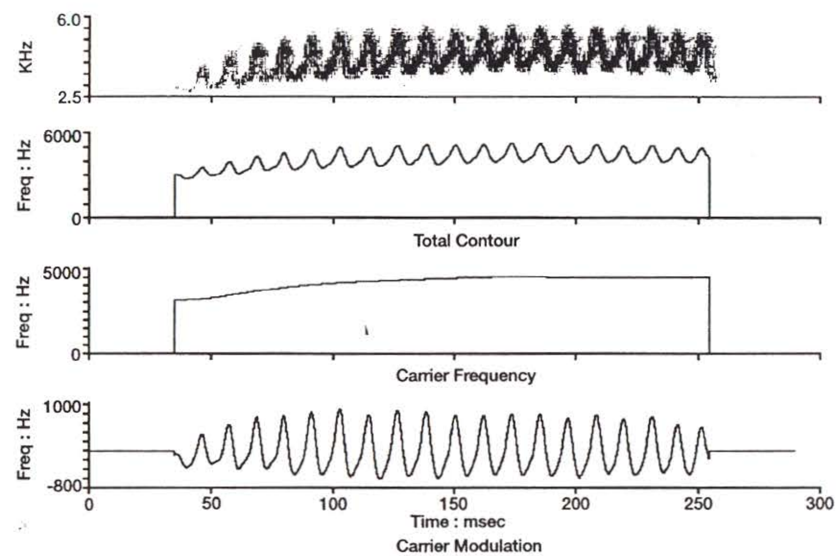

Fig. 22. The spectrogram of a frequency-modulated Eastern phoebe (*Sayornis phoebe*) note (*top panel*) is analyzed into its total spectral contour (*top middle panel*), which is further analyzed into carrier frequency (*bottom middle panel*) and carrier modulation (*bottom panel*)

9.4 Synthesis of Biological Sound Types

9.4.1 *Tonal and Polytonal Signals*

Tonal signals are the simplest to characterize and synthesize using the methods outlined in this chapter. Polytonal signals are more difficult to synthesize, and the choice of modeling approach depends on the original production process. Depending on the sound and the species, tones may be either added or multiplied during sound production. The former produces tonal sums, while the latter produces AM cross-products which can assume complex forms, including harmonics at intervals of the amplitude-modulation frequency, or sums or differences of the input frequencies. Pitch interactions within the polytonal sound can be subtle or quite dissonant, depending on the degree of coupling occurring between the sound sources and the frequencies involved. Polytonal sounds can display a great diversity and complexity of spectral characteristics and a corresponding variation in sound quality.

Synthesis approaches for these more complex signals depend on their structure. Polytonal sums can sometimes be analyzed by separating the voices via filtering or

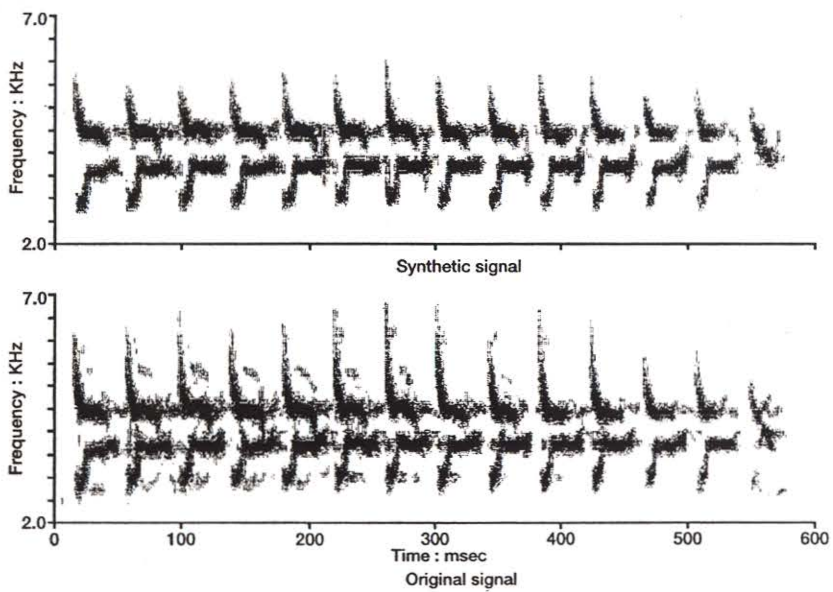

Fig. 23. A polytonal sound produced by a wood thrush (*Hylocichla mustelina*; *bottom panel*) is synthesized using polytonal synthesis (*top panel*)

spectrogram contour detection. However, the most complex polytonal sounds, such as some chickadee calls, may involve intermodulation between two harmonic voices, and therefore require modeling the modulation process itself in order to reproduce the complex spectral relationships. Figure 23 shows a segment of wood thrush (*Hylocichla mustelina)* song containing the sum of two distinct nonharmonic voices. Because the voices occupy different frequency ranges, they can be modeled by analyzing the sound into two independent sets of amplitude and frequency functions.

9.4.2 *Pulse-Repetition Signals*

Pulse-repetition sounds can also arise from a variety of production mechanisms and therefore vary widely in structure, and in the techniques used to synthesize them. One approach is to generate mathematical functions that are used singly or in combination to produce individual pulses, from which a sequence can be built through waveform editing. Pulses can be repeated at constant or variable intervals, and the sequence can be uniform, show patterned variation (such as stepped changes in frequency or intensity), or be randomly changing. The onset and offset amplitudes of each pulse should be tapered, as discussed earlier.

Alternatively, synthetic pulses can be derived from naturally occurring pulse-repetition sounds (such as anuran vocalizations). The pickerel frog (*Rana palustris)* call

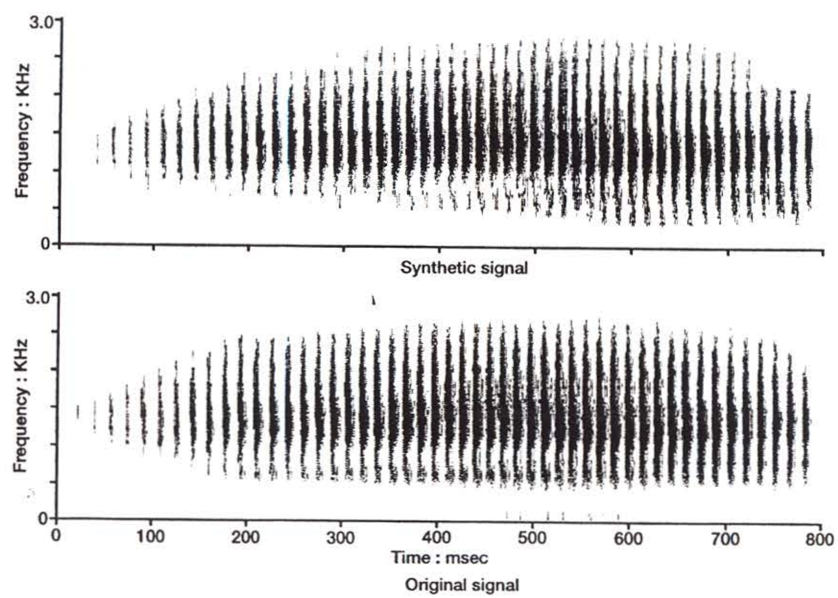

Fig. 24. A pulse-repetition call of a pickerel frog (*Rana palustris*; *bottom panel*) is synthesized using pulse repetition synthesis techniques (*top panel*)

shown in Figure 24, for example, can be modeled as a sequence of broadband FM pulses, further modulated by an AM pulse-repetition envelope. The aural impression of this sound reflects both the pitch characteristics of its FM behavior and the buzziness produced by the rapid pulse repetition. *F(t)* can be obtained through zero-crossing analysis, and *A(t)* through amplitude-envelope analysis. FM and pulse-repetition behavior are thus separated. Manipulations include altering or removing the FM component, compressing or expanding the pulse repetition rate, and inserting or rearranging individual pulses.

9.4.3
Harmonic Signals

Harmonic sounds can be generated by synthesizing individual harmonics using the tonal model, then adding these components together. Because individual harmonics are explicitly separated in this approach, they can be individually edited and weighted before being recombined. All harmonics can be synthesized from the fundamental contour, *F(t)*, referred to as $F_0(t)$. Figure 25 shows a chickadee *dee* note in its original form and in a synthetic version based on this technique.

Harmonic synthesis proceeds in five steps. First, the fundamental frequency contour, $F_0(t)$, is extracted using one of the techniques described above. If the energy is concentrated in the upper harmonics, leaving the fundamental undetectable, one of the visible harmonic contours, $F_n(t)$, can be extracted and then scaled by division to recover the fundamental. Second, higher harmonic frequency contours, $F_1(t), F_2(t)...,F_n(t)$, are derived from $F_0(t)$ by integer multiplication (i.e., 2, 3...,n + 1). Third, corresponding amplitude functions, $A_0(t), A_1(t)..., A_n(t)$, are derived individually from the fundamental and harmonic components of the natural sound. An efficient approach developed by Balaban (described in Beeman 1996b) derives $A_n(t)$ by automatically tracing the contour of the corresponding $F_n(t)$ through the spectrogram and reading successive amplitude values. Fourth, waveforms are generated for the individual harmonics, (i.e., $S_n(t) = A_n(t) * \sin [\, 2\pi F_n(t)\,]\, t$), by applying tonal synthesis to each pair, $F_0(t)$ and $A_0(t)$, $F_1(t)$ and $A_1(t)$, and so on. Finally, these waveforms are summed to produce the composite harmonic signal, $S(t) = S_0(t) + S_1(t) + ... + S_n(t)$.

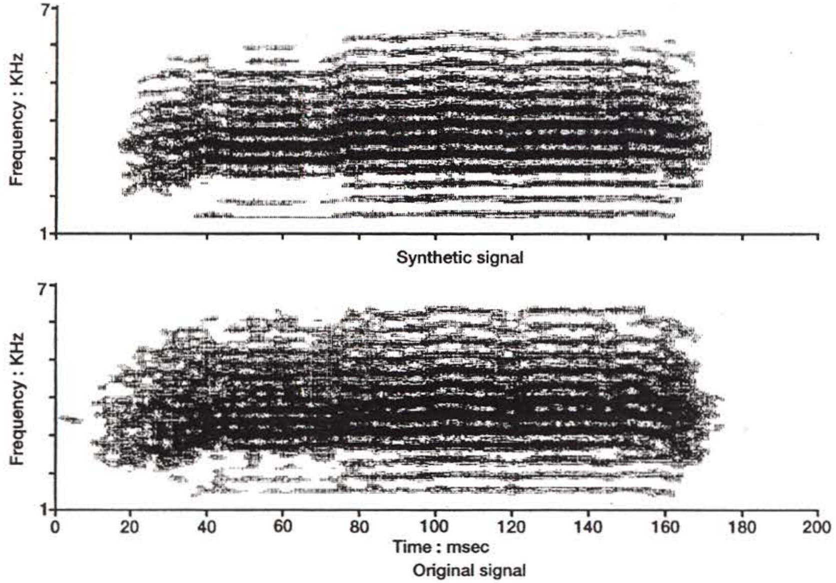

Fig. 25. A black-capped chickadee (*Parus atricapillus*) *dee* note (*bottom panel*) is synthesized using harmonic synthesis techniques (*top panel*)

Harmonic-synthesis techniques can be used to selectively remove or rescale the amplitude of individual harmonic components, in order to simplify a sound or test the significance of different harmonic components. An individual harmonic is rescaled by multiplying the corresponding $A_n(t)$ function by some factor and then resynthesizing. Similarly, an individual harmonic can be removed by omitting it from the harmonic sum.

9.4.4
Noisy Signals

Noisy signals can be also be synthesized, for instance by adding noise components to existing sounds, or by noise, summing noise bands to create spectral patterns. Noise signals can be based on uniform or Gaussian, both of which have roughly uniform spectral-energy distribution. When adding noise to an existing signal, the noise should be amplitude-scaled for a specific signal-to-noise ratio with respect to the signal. This is accomplished by separately measuring the RMS levels of signal and noise, and scaling these components accordingly. Noise energy also can be limited to a specific frequency band using a bandpass filter, and multiple noise bands can be combined to form a spectrally structured noisy sound. As before, careful amplitude-scaling of energy in different frequency bands can be used to explicitly simulate or alter the spectral energy distribution pattern of a natural sound (see also Owren and Bernacki, this Volume, for discussion of synthesis of noisy bioacoustic signals).

9.5
Miscellaneous Synthesis Topics

9.5.1
Template Sounds

Tonal synthesis techniques allow the user to derive and synthesize an average "template" sound from an ensemble of signals, for instance for morphological analysis or playback experiments (Beeman 1996b). The average sound is derived as follows (see Figure 26). *F(t)* and *A(t)* functions are extracted from each sound sample, then the

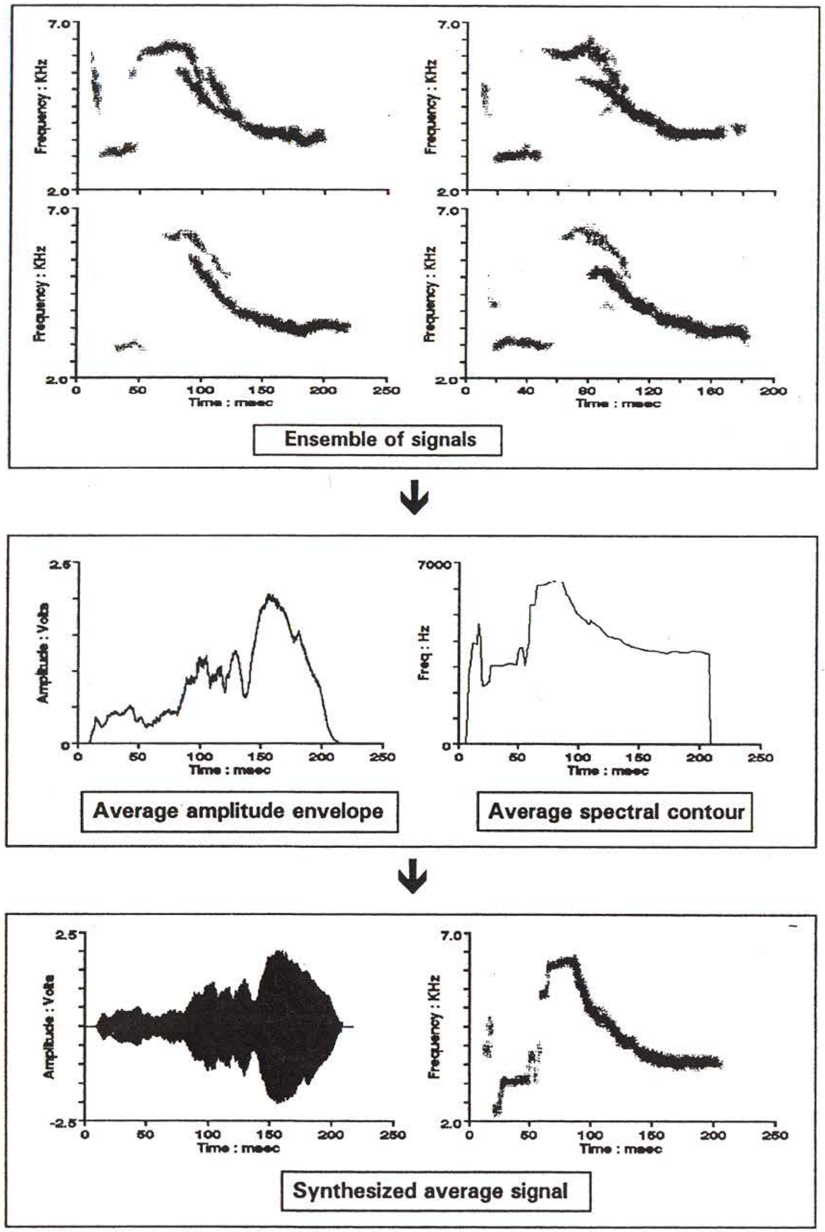

Fig. 26. Sound template derivation. Frequency and amplitude functions are derived from an ensemble of signals (*top panel*), to create averaged $A(t)$ and $F(t)$ functions (*middle panel*), from which a template sound is synthesized (*bottom panel*)

ensemble of functions is normalized in duration and aligned in time. The $F(t)$ and $A(t)$ functions are then averaged, yielding $F_{avg}(t)$ and $A_{avg}(t)$, which can be used to synthesize a single template waveform representing the average sound for the entire data set. Note that template sounds cannot be obtained by averaging time waveforms (whose random phase components would mutually cancel), from parameter measurements (which cannot be used to regenerate sounds), or from spectrogram averages (which cannot be inverse transformed to an artifact-free time signal).

Template sounds have several important applications. First, they produce an average exemplar for sound types that are too complex to be adequately characterized by measurements of sound parameters alone. Average morphological features of the sound can then be measured from the template version. Second, individual sounds from the data set can be compared to the template sound in order to measure their morphological similarity. Third, because template sounds are generated directly from amplitude and frequency functions, sound features can be altered systematically by manipulating these functions. The researcher can generate an array of test sounds of graduated similarity to the template in order to test, for instance, relationships between functional, perceptual, and acoustical similarity, the perceptual centrality of the template, the perceptual significance of various acoustic features in the template, and the locations and abruptness of functional and perceptual boundaries in relation to acoustic variation.

9.5.2
Noise Removal

Tonal-synthesis techniques can also be used to remove noise from tonal, and even some harmonic signals, by extracting relatively noise-free $F(t)$ and $A(t)$ functions from the natural sounds and using these to produce noise-free synthetic versions. This technique was developed by Nelson (described in Beeman 1996b), and has been used to clean up noisy field recordings for manipulation and subsequent playback. The success of this technique depends directly on the ability to extract accurate $F(t)$ and $A(t)$ representations of the underlying signal in the presence of noise, as well as the adequacy of the tonal model as a represention of the natural sound. This approach is effective because of the noise-insensitivity of the Fourier-transform-based SCD method for deriving $F(t)$, because noise occurring in $A(t)$ can be reduced by smoothing, and because minor artifacts in $F(t)$ and $A(t)$ can be corrected, if necessary, using mouse-drawing techniques. As illustrated in Figure 27, this noise-removal technique can produce quite dramatic results when applied to field-recordings.

10
Summary

Digital signal analysis technology adds many capabilities to the field of biological sound analysis. It provides rapid and flexible ways of visualizing sound signals and a reliable way of recording and storing them. More important, because digitized signals are numerical, digital technology allows researchers to use mathematical analysis and mod-

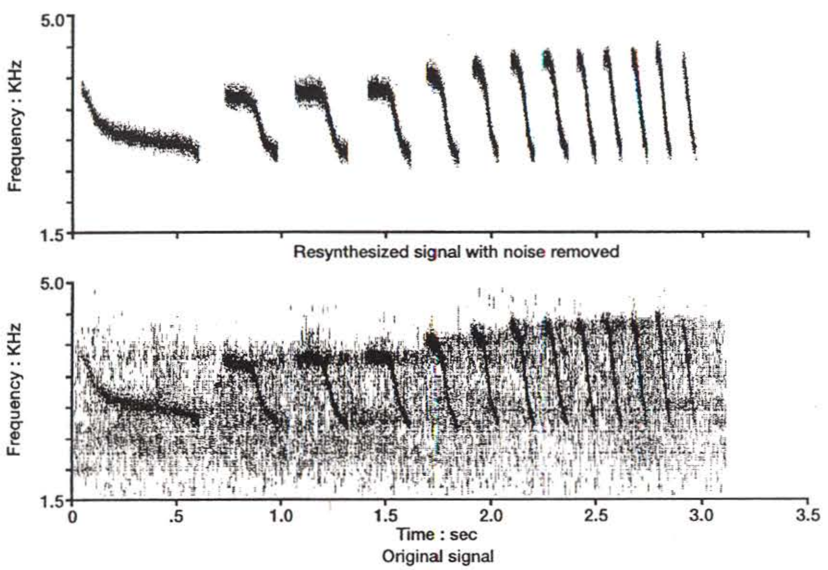

Fig. 27. A natural field sparrow song with a high level of background noise (*bottom panel*) is analyzed and resynthesized (*top panel*) using tonal-sound techniques for noise removal

sound analysis techniques made possible by digital signal analysis, and the underlying theory and practical considerations. Precise and automated measurements can be made on waveforms, power spectra, and spectrograms. Time signals can be converted either to amplitude envelopes to analyze their amplitude behavior, or to gate functions to visualize, measure, and compare their temporal patterning. Frequency spectra and frequency-time spectrograms of specific sound segments can be calculated digitally and measured to characterize sound features. Spectral analysis parameters such as time and frequency resolution, transform window, spectral smoothing, and physical scaling should be selected carefully for accurate analyses. Real-time spectrographic display is also possible.

Other digital techniques carry sound analysis beyond the scope of traditional analog tools. Researchers can digitally extract and calculate spectral contours, which are mathematical functions representing a sound's instantaneous time-varying frequency. Different contour extraction techniques, such as zero-crossing analysis, spectrogram contour detection, or Hilbert transform analysis, can be used depending on sound structure. Spectral contours can form the basis of both sound comparison and sound synthesis operations. Digital analysis also allows the researcher to compare sounds in a variety of ways, based on a variety of sound characteristics. Parameters representing isolated sound features can be measured, often automatically, and compared statistically for similarity. More advanced approaches allow researchers to compare sound functions used to represent a sound's entire behavior over one or more characteristics, such as amplitude envelopes, power spectra, spectral contours, or spectrograms. These

such as amplitude envelopes, power spectra, spectral contours, or spectrograms. These functions can be compared mathematically using cross-correlation techniques to measure quantitative similarity. An ensemble of sounds can be compared pairwise and analyzed statistically to estimate clustering tendencies, or a sequence of individual sounds can be compared to a sound template to measure their graded similarity.

Another important application of digital sound analysis is sound manipulation and synthesis. Digital editing tools can be used to rearrange existing sound material, and these replace traditional cut-and-splice tape techniques with faster and more accurate interactive screen tools. Digital sound generation techniques can create entirely new sounds from mathematical functions for use as auditory stimuli in psychophysical and neurophysiological research. Digital synthesis techniques can decompose natural sounds into their essential components, which can then be manipulated and recombined to form selectively altered sounds. In this manner, natural biological sounds can be varied in detail and then used in playback studies, to explore the behavioral significance of individual sound features. Eight natural sound types were described, along with techniques for synthesizing many of them, including tonal sounds, harmonic sounds, polytonal sounds, pulse repetition sounds, and noise, and combinations and sequences of these sounds. Much natural sound synthesis is based on the tonal sound model, which reduces sounds to one or more pairs of amplitude envelope and spectral contour functions, which can be independently manipulated to alter sound characteristics. Synthesis manipulations include frequency shifting, time scaling, amplitude and frequency modulation, intensity manipulation, harmonic removal and rescaling, pulse repetition variation, and noise addition and removal. Finally, digital analysis can be used to synthesize an average template sound from a sound ensemble, for feature measurement, similarity analysis, or perceptual testing.

References

Beecher MD (1988) Spectrographic analysis of animal vocalizations: implications of the "uncertainty" principle. Bioacoustics 1: 187–208
Beeman K (1987) SIGNAL V2.5 User's Guide [DEC PDP]. Engineering Design, Belmont, Massachusetts
Beeman K (1989) SIGNAL V1.3 User's Guide [IBM PC]. Engineering Design, Belmont, Massachusetts
Beeman K (1991) RTS V1.0 User's Guide [IBM PC]. Engineering Design, Belmont, Massachusetts
Beeman K (1996a) RTS V2.0 User's Guide [IBM PC]. Engineering Design, Belmont, Massachusetts
Beeman K (1996b) SIGNAL V3.0 User's Guide [IBM PC]. Engineering Design, Belmont, Massachusetts
Borror DJ (1965) Song variation in Maine song sparrows. Wilson Bull 77: 5–37
Brand AR (1934) A method for the intensive study of bird song. Auk 52: 40–52
Buck JR, Tyack PL (1993) A quantitative measure of similarity for *Tursiops truncatus* signature whistles. J Acoust Soc Am 94: 2497–2506
Charif RA, Mitchell S, Clark CW (1995) CANARY 1.2 users manual. Cornell Laboratory of Ornithology, Ithaca New York
Clark CW (1982) The acoustic repertoire of the southern right whale, a quantitative analysis. Anim Behav 30: 1060–1071
Clark CW, Marler P, Beeman K (1987) Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. Ethology 76: 101–115
Dabelsteen T, Pedersen SB (1985) A method for computerized modification of certain natural animal sounds for communication study purposes. Biol Cybern 52: 399–404
Dorrscheidt GJ (1978) A reversible method for sound analysis adapted to bio-acoustical experiments. Int J Bio Med Comput 9: 127–145
Flanagan JL (1972) Speech analysis, synthesis, and perception, 2nd edn. Springer Berlin Heidelberg New York

Gerhardt HC (1989) Acoustic pattern recognition in anuran amphibians. In: Dooling RJ, Hulse SH (eds) Comparative psychology of audition: Perceiving complex sounds. Erlbaum Assoc, Hillsdale, New Jersey, p 175
Gerhardt HC (1992) Multiple messages in acoustic signals. Sem Neurosci 4: 39–400
Greenewalt CH (1968) Bird Song: Acoustics and physiology. Smithsonian Institute Press, Washington, DC
Hauser MD (1991) Sources of acoustic variation in rhesus macaque vocalizations. Ethology 89: 29–46
Hauser MD (1992) Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: a learned mode of production? J Acoust Soc Am 91: 2175–2179
Hauser MD, Schön Ybarra M (1994) The role of lip configuration in monkey vocalizations: experiments using xylocaine as a nerve block. Brain Lang 46: 232–244
Hienz RD, Brady JV (1988) The acquisition of vowel discriminations by nonhuman primates. J Acoust Soc Am 84: 186–194.
Hopp SL, Sinnott JM, Owren MJ, Petersen MR (1992) Differential sensitivity of Japanese macaque (*Macaca fuscata*) and humans (*Homo sapiens*) to peak position along a synthetic coo call continuum. J Comp Physiol Psychol 106: 128–136
Klatt DH (1980) Software for a cascade/parallel formant synthesizer. J Acoust Soc Am 67: 971–995
Koenig W, Dunn HK, Lacy LY (1946) The sound spectrograph. J Acoust Soc Am 18: 19–49
Kroodsma DE (1974) Song learning, dialects and dispersal in Bewick's wren. Z Tierpsychol 35: 353–380
Loughlin PJ, Atlas LE, Pitton JW (1993) Advanced time-frequency representations for speech processing. In: Cooke M, Beet S, Crawford M (eds) Visual representations of speech signals. Wiley, New York, p 27
Margoliash D (1983) Song-specific neurons in the white-crowned sparrow. J Neurosci 3: 1039–1048
Markel JD, Gray AH (1976) Linear prediction of speech. Springer, Berlin Heidelberg New York
Marler P, Pickert R (1984) Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). Anim Behav 32: 673–689
Martindale S (1980) A numerical approach to the analysis of solitary vireo songs. Condor 82: 199–211
May B, Moody DB, Stebbins WC (1988) The significant features of Japanese macaque coo sounds: a psychophysical study. Anim Behav 36: 1432–1444
Nelson DA (1992) Song overproduction and selective attrition lead to song sharing in the field sparrow (*Spizella pusilla*). Behav Ecol Sociobiol 30: 415–424
Nelson DA, Croner LJ (1991) Song categories and their functions in the field sparrow (*Spizella pusilla*). Auk 108: 42–52
Nelson DA, Marler P, Palleroni A (1995) A comparative approach to vocal learning: intraspecific variation in the learning process. Anim Behav 50: 83–97
Nowicki S, Capranica RR (1986) Bilateral syringeal interaction in vocal production of an oscine bird sound. Science 231: 1297–1299
Oppenheim AV, Schafer RW (1989) Discrete-time signal processing. Prentice-Hall, Englewood Cliffs
Owren MJ, Bernacki RH (1988) The acoustic features of vervet monkey alarm calls. J Acoust Soc Am 83: 1927–1935
Randall R (1977) Application of B&K equipment to frequency analysis. Brel & Kjr, Nrum, Denmark
Richard JP (1991) Sound analysis and synthesis using an Amiga micro-computer. Bioacoustics 3: 45–60
Rosen S, Howell P (1991) Signals and systems for speech and hearing. Academic Press, New York
Ryan MJ, Perrill SA, Wilczynski W (1992) Auditory tuning and call frequency predict population-based mating preferences in the cricket frog, *Acris crepitans*. Am Natural 139: 1370–1383
Sabine WS (1900) Reverberation. In: Sabine WS (ed) Collected papers on acoustics. Dover, New York, p 3
Schafer RW (1975) Digital representation of speech signals. Proc IEEE 63:662–677
Sinnott JM (1989) Detection and discrimination of synthetic English vowels by old world monkeys (*Cercopithecus*, *Macaca*) and humans. J Acoust Soc Am 86: 557–565
Sparling DW, Williams JD (1978) Multivariate analysis of avian vocalizations. J Theor Biol 74: 83–107
Staddon JER, McGeorge LW, Bruce RA, Klein FF (1978) A simple method for the rapid analysis of animal sounds. Z Tierpsychol 48: 306–330
Stephen RO, Hartley JC (1995) Sound production in crickets. J Exp Biol 198:2139–2152
Stoddard PK (1990) Audio computers – theory of operation and guidelines for selection of systems and components. Bioacoustics 2: 217–239
Stremler FG (1977) Introduction to communication systems. Addison-Wesley, Reading, Massachusetts
Suthers RA, Goller F, Hartley RS (1994) Motor dynamics of song production by mimic thrushes. J Neurobiol 25: 917–936
Wells MM, Henry CS (1992) Behavioral responses of green lacewings (Neuroptera: Chrysopidae) to synthetic mating songs. Anim Behav 44: 641–652
Zoloth S, Dooling RJ, Miller R, Peters S (1980) A minicomputer system for the synthesis of animal vocalisations. Z Tierpsychol 54: 151–162