Granular Synthesis

Just as light energy can be viewed both in terms of wavelike properties and in terms of particulate properties (photons), so can sound. *Granular synthesis* builds up acoustic events from thousands of sound *grains*. A sound grain lasts a brief moment (typically 1 to 100 ms), which approaches the minimum perceivable event time for duration, frequency, and amplitude discrimination.

Granular representations are a useful way of viewing complex sound phenomena—as constellations of elementary units of energy, with each unit bounded in time and frequency. Such representations are common inside synthesis and signal-processing algorithms, although there are many different terms for similar phenomena. The "quantum" (Gabor 1946, 1947), "Gaussian elementary signal" (Helstrom 1966; Bastiaans 1980), "short-time segment" (Schroeder and Atal 1962), "short-time weighting function" (Flanagan 1972), "window" (Arfib 1991; Harris 1978; Nuttall 1981), "sliding window" (Bastiaans 1985), "window function pulse" (Bass and Goeddel 1981), "wavelet" (Kronland-Martinet and Grossmann 1991), "formant-wave-function" or "FOF" (Rodet 1980), "VOSIM pulse" (Kaegi and Tempelaars 1978), "wave packet" (Crawford 1968), "toneburst" (Blauert 1983; Pierce 1990), "tone pulse" (Whitfield 1978), and even the "tone pip" (Buser and Imbert 1992) can all be described as granular representations of musical signals.

The grain is an apt representation for sound because it combines time-domain information (starting time, duration, envelope shape, waveform shape) with frequency-domain information (the period of the waveform inside the grain, spectrum of the waveform). This stands in opposition to representations at the sample level that do not capture frequency-domain information, and abstract Fourier methods that presume that sounds are summations of infinitely long sinusoids.

Granular Synthesis: Background

Atomistic views of sound as "particles" can be traced to the origins of the scientific revolution. The Dutch scholar Isaac Beekman (1588–1637) proposed in 1616 a "corpuscular" theory of sound (Beekman 1604–1634; Cohen 1984). Beekman believed that any vibrating object, like a string, cuts the surrounding air into spherical corpuscules of air that are projected in all directions by the vibration. When these corpuscles impinge on the eardrum, Beekman theorized, we perceive sound. While this theory is not strictly true in a scientific sense, it paints a colorful metaphor for the perception of granular synthesis.

Centuries later, the notion of a *granular* or *quantum* approach to sound was proposed by the British physicist Dennis Gabor in a pair of brilliant papers that combined theoretical insights from quantum physics with practical experiments (1946, 1947). According to Gabor's theory, a granular representation could describe any sound. This hypothesis was verified mathematically by Bastiaans (1980, 1985). In the 1940s Gabor actually constructed a sound granulator based on a sprocketed optical recording system adapted from a film projector. He used this to make experiments in *time compression/expansion* with *pitch shifting*—changing the pitch of a sound without changing its duration, and vice versa. (See chapter 10 for a discussion of time compression and expansion with pitch-shifting.)

A granular representation is implicit in the *windowing* technique applied in the *short-time Fourier transform,* developed in the 1960s (Schroeder and Atal 1962; see chapter 13 and appendix A). The MIT cybernetician Norbert Wiener (1964) and the information theorist Abraham Moles (1968) also proposed granular representations for sound.

The composer Iannis Xenakis (1960) was the first to explicate a compositional theory for grains of sound. He began by adopting the following lemma: "All sound, even continuous musical variation, is conceived as an assemblage of a large number of elementary sounds adequately disposed in time. In the attack, body, and decline of a complex sound, thousands of pure sounds appear in a more or less short interval of time Δt." Xenakis created granular sounds using analog tone generators and tape splicing. These appear in the composition *Analogique A-B* for string orchestra and tape (1959). The composition is described in Xenakis (1992). (The score and tape are available from Editions Salabert.)

The author of this book developed the first computer-based implementations of granular synthesis in 1974 at the University of California, San Diego (Roads 1978c) and in 1981 at the Massachusetts Institute of Technology (Roads 1985g). The technique appears in several compositions, including *nscor* (1980, Wergo compact disc 2010-50), *Field* (1981, MIT Media Laboratory compact disc), and *Clang-tint* (Roads 1993b). Granular synthesis has been implemented in different ways, notably by the Canadian composer Barry Truax (1987, 1988, 1990a, b), as we discuss in more detail later.

Sonic Grains

An amplitude envelope shapes each grain. This envelope can vary in different implementations from a Gaussian bell-shaped curve to a simple

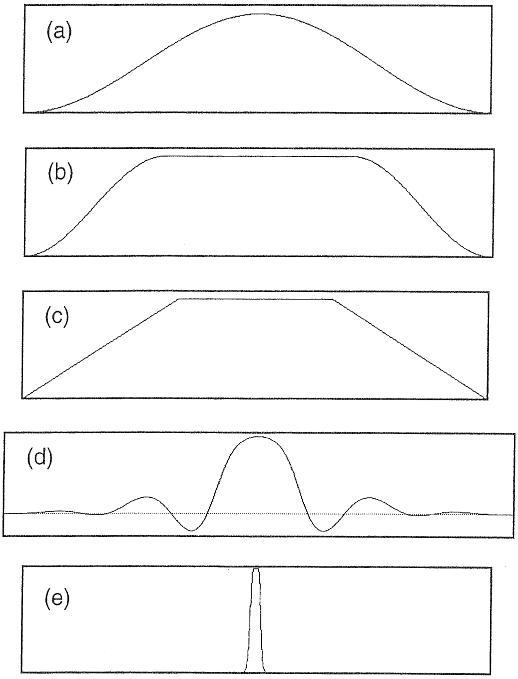

Figure 5.10 Grain envelopes. (*a*) Gaussian. (*b*) Quasi-Gaussian. (*c*) Three-stage linear. (*d*) Pulse. (*e*) Narrow impulse; this could be seen as equivalent to (*a*), but over a narrower timescale.

three-stage line-segment attack/sustain/decay (figure 5.10). The following equation defines a Gaussian curve $P(x)$:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where σ is the standard deviation (spread of the bell) and μ is the *mean* or center peak.

Figure 5.10b shows a *quasi-Gaussian* curve or *Tukey window* (Harris 1978), where the peak is extended over 30 to 50 percent of the duration of the grain. This shape has proved sonically effective (Roads 1985g).

Complicated envelopes like a band-limited pulse (figure 5.10d) create resonant grains that sound like woodblock taps in sparse textures when the grain duration is less than 100 ms. Narrow envelopes like figure 5.10e create crackling and popping textures when the total grain duration is less than 20 ms. As one would expect, sharp angles in the envelope cause strong side effects in the spectrum. These side effects are due to the convolution of the envelope's spectrum with that of the grain waveform. (See chapter 10 for an explanation of convolution.)

The grain duration can be constant, random, or it can vary in a frequency-dependent way. This means, for example, that we can assign shorter durations to high-frequency grains. A correspondence between grain frequency and grain duration is characteristic of the *wavelet* analysis/resynthesis, discussed later in this chapter and in chapter 13.

The waveform within the grain can be of two types: synthetic or sampled. Synthetic waveforms are typically sums of sinusoids scanned at a specified frequency. For sampled grains, one typically reads the waveform from a stipulated location in a stored sound file, with or without pitch-shifting.

Several parameters can be varied on a grain-by-grain basis, including the duration, envelope, frequency, location in sound file (for sampled grains), spatial location, and waveform (a wavetable for synthetic grains, or a file name or input channel for sampled grains). It is this grain-by-grain level of control that leads to the unique effects made possible by this method.

Grain Generator Instrument

Granular synthesis can be implemented with a simple synthesis instrument: a sine wave oscillator controlled by an envelope generator (figure 5.11). One could easily extend this instrument to allow a choice between several wavetable functions.

Despite the simplicity of the instrument, to generate even a plain, uncomplicated sound requires a massive amount of control data—up to thousands of parameters per second of sound. These parameters describe each grain: starting time, amplitude, etc. Since one does not want to have to specify each grain's parameters manually, a higher-level unit of organization is necessary. This unit of organization should automatically generate the thousands of individual grain specifications.

High-level Granular Organizations

The complexity of the sound generated by granular synthesis derives from the amount of control data fed to it. If n is the number of parameters for each grain, and d is the average grain density per second of sound, it takes $d \times n$ parameter values to specify one second. Since d typically varies between a few dozen and several thousand, it is clear that for the purposes of compositional control, a higher-level unit of organization for the grains is needed. The purpose of such a unit is to let composers stipulate large quantities of grains using just a few global parameters.

Existing granular synthesis methods can be classified into five types, according to the organization of the grains:

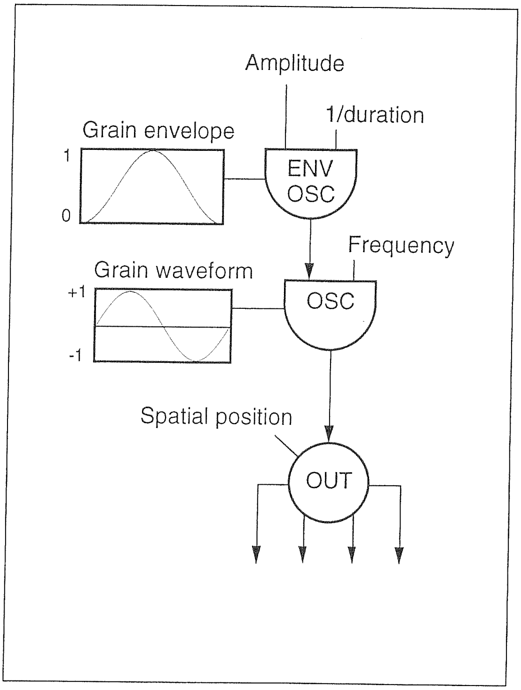

Figure 5.11 A simple granular synthesis instrument built from an envelope generator and an oscillator with multichannel output.

1. Fourier and wavelet grids
2. Pitch-synchronous overlapping streams
3. Quasi-synchronous streams
4. Asynchronous clouds
5. Time-granulated or sampled-sound streams, with overlapped, quasi-synchronous, or asynchronous playback

In the next sections we examine briefly each approach.

Fourier/Wavelet Grids and Screens

Two related spectrum analysis techniques, the short-time Fourier transform (STFT) and the *wavelet transform,* take in a time-domain sound signal and measure its frequency content versus time. (Chapter 13 presents both techniques.) In effect, these methods associate each point in the analysis grid with a unit of time-frequency energy—a grain or wavelet (figure 5.12).

The STFT is well known and can be computed using the fast Fourier transform (Rabiner and Gold 1975). The "grain" in this case is a set of

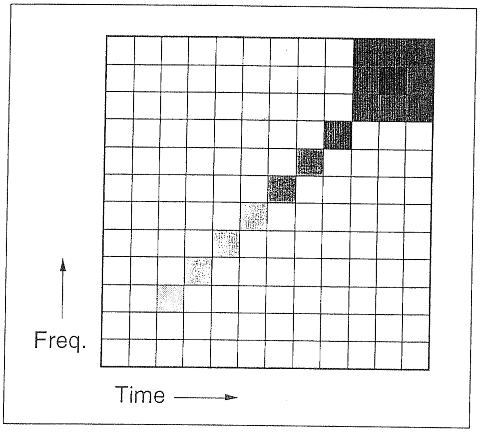

Figure 5.12 Fourier grid dividing the time domain and the frequency domain into bounded units. Each row represents a frequency channel, and each column indicates a period of time. The darkness in each square indicates the intensity in that time-frequency region. This example shows a sound that ascends in frequency and grows more intense. In the STFT the frequency grid is linear; in the wavelet transform it is typically logarithmic.

overlapping analysis windows within each of the *N* channels of the Fourier analyzer (the horizontal rows of figure 5.12). We can view the grains as if they were aligned on a two-dimensional time/frequency grid, where the intervals between the grid are equal. Arfib (1991) describes applications of the STFT in terms of granular operations.

The wavelet transform (Kronland-Martinet and Grossmann 1991) performs a similar operation, but the spacing of the analysis channels and the duration of the window (called the *analyzing wavelet*) is different from the STFT. In the STFT, the spacing between the channels on the frequency axis is linear, while in the wavelet transform it is logarithmic. That is, in the wavelet transform, the channel frequency interval (bandwidth) $\Delta f/f$ is constant. Also, in the STFT, the window duration is fixed, while in the wavelet transform it varies as a function of frequency. (See chapter 13 for more on wavelets.)

Both techniques permit analysis, transformation, and resynthesis, which make them potentially powerful musical tools for the manipulation of sampled sounds. The most obvious transformations using Fourier/wavelet grids involve stretching or shrinking the grid to effect time compression and expansion with pitch-shifting, that is, shifting pitch while keeping the duration the same, or vice versa.

Another grid-oriented conception, but not related to Fourier or wavelet analysis, is Xenakis's (1960, 1992) concept of *screens*. A screen is an amplitude-frequency grid on which grains are scattered. A synchronous sequence of screens (called a *book*) constitutes the evolution of a complex sound. Rather than starting from an analyzed sound, as in Fourier/wavelet grids, proposals for screen-based synthesis use generative algorithms to fill the screen with grains. Xenakis (1971, 1992) proposed scattering grains randomly into screens, then constructing new screens from set-theory operations—intersections, unions, complements, differences, among other operations:

> *Using all sorts of manipulations with these grain clusters, we can hope to produce not only the sounds of classical instruments and elastic bodies, and those sounds preferred in concrète music, but also sonic perturbations with evolutions unparalleled and unimaginable until now.*

Another screen-oriented proposal suggested that grain parameters could be derived from the interaction of cellular automata (Bowcott 1989).

Pitch Synchronous Granular Synthesis

Pitch synchronous granular synthesis (PSGS) is a technique designed for the generation of tones with one or more formant regions in their spectra (De Poli and Piccialli 1991). PSGS is a multistaged operation involving pitch detection, spectrum analysis and resynthesis, and impulse response-based filtering, technical processes that are described in later chapters; thus the description here is brief. (See De Poli and Piccialli 1991 for details.)

The first stage of the analysis is *pitch detection* (see chapter 12). Each pitch period is treated as a separate unit or grain. Spectrum analysis is performed on each grain. The system derives the *impulse response* of the spectrum and uses it to set the parameters for a resynthesis filter. (Chapter 10 discusses impulse response measurements.)

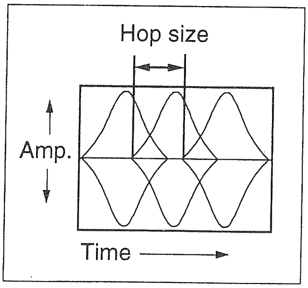

Figure 5.13 Stream of overlapped grains. The *hop size* is the delay between successive grains.

In resynthesis, a pulse train at the detected pitch period drives a bank of *finite impulse response* (FIR) filters. (FIR filters are discussed in chapter 10.) The output signal results from the excitation of the pulse train on the weighted sum of the impulse responses of all the filters. At each time frame, the system emits a grain that is overlapped and added with the previous grain to create a smoothly varying signal (figure 5.13). The implementation of PSGS by De Poli and Piccialli features several transformations that can create variations of the original sound. Later extensions allow separation of the quasi-harmonic part of the sound from the residual inharmonic part (Piccialli et al. 1992).

Quasi-synchronous Granular Synthesis

Quasi-synchronous granular synthesis (QSGS) generates one or more *streams* of grains, one grain following another, with a variable delay period between the grains. The stream concept has the advantage of being straightforward and intuitive. Orton, Hunt, and Kirk (1991) developed a graphical interface for drawing stream trajectories as curved lines on a display screen.

Figure 5.14 shows a stream of five grains, each with a quasi-Gaussian envelope and a variable delay before the next grain. We say "quasi-synchronous" because the grains follow each other at more-or-less equal intervals. When the interval between successive grains is equal, the overall envelope of a stream of grains forms a periodic function. Since the envelope is periodic, the signal generated by QSGS can be analyzed as a case of *amplitude modulation* (AM). AM occurs when the shape of one signal (the *modulator*) determines the amplitude of another signal (the *carrier*). (See chapter 6 for more on modulation.) In this case the carrier is the waveform within the grain, and the modulator is the grain envelope.

From a signal-processing standpoint, we observe that for each sinusoidal component in the carrier, the periodic envelope function contributes a series of *sidebands* to the final spectrum. (Sidebands are additional frequency components above and below the frequency of the carrier.) The sidebands are separated from the carrier by a distance corresponding to the inverse of the period of the envelope function. For a stream of 20 ms grains following one after the other, the sidebands in the output spectrum are spaced at 50 Hz intervals. The shape of the grain envelope determines the precise amplitude of these sidebands.

The result created by the modulation effect of a periodic envelope is that of a *formant* surrounding the carrier frequency. That is, instead of a single line in the spectrum (denoting a single frequency), the spectrum looks like a sloping hill (denoting a group of frequencies around the carrier). QSGS

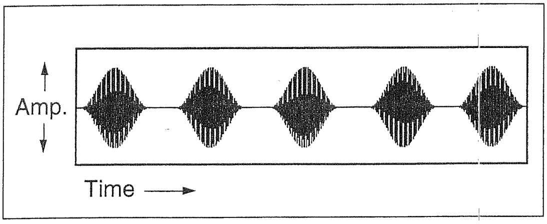

Figure 5.14 A stream of five 40 ms grains at 1.06 KHz with a Hanning envelope. In this case the delay period between the grains varies slightly.

is, in this sense, similar to the formant synthesis methods *VOSIM* (Kaegi and Tempelaars 1978) and *formant-wave-function* or FOF synthesis (Rodet 1980; Rodet, Potard, and Barrière 1984). (See chapter 7 for more on VOSIM and FOF synthesis.)

By combining several streams of quasi-synchronous grains in parallel (each stream creating its own formant around a separate frequency), the signal can simulate the resonances of the singing voice and acoustic instruments.

When the interval between the grains is irregular, as in figure 5.15, this leads to a controllable thickening of the sound texture through a "blurring" of the formant structure (Truax 1987, 1988). In its simplest form, the variable-delay method is similar to amplitude modulation (AM) using low-frequency colored noise as a modulator. (See chapter 6 for more on modulation.) In itself, this is not particularly interesting. The granular representation, however, lets us take this technique far beyond simple noise-modulated AM. In particular, we can simultaneously vary several other parameters on a grain-by-grain basis, such as grain waveform, amplitude, duration, and spatial location. On a more global level, we can also dynamically vary the density of grains per second to create a variety of striking effects.

Asynchronous Granular Synthesis

Asynchronous granular synthesis (AGS) gives the composer a precision spray jet for sound, where each dot in the spray is a sonic grain (Roads 1991). AGS scatters grains in a statistical manner over a specified duration within regions inscribed on the frequency-versus-time plane. These regions are called *clouds*—the units with which a composer works.

The composer specifies a cloud in terms of the following parameters, shown in figure 5.16.

1. Start time and duration of the cloud.

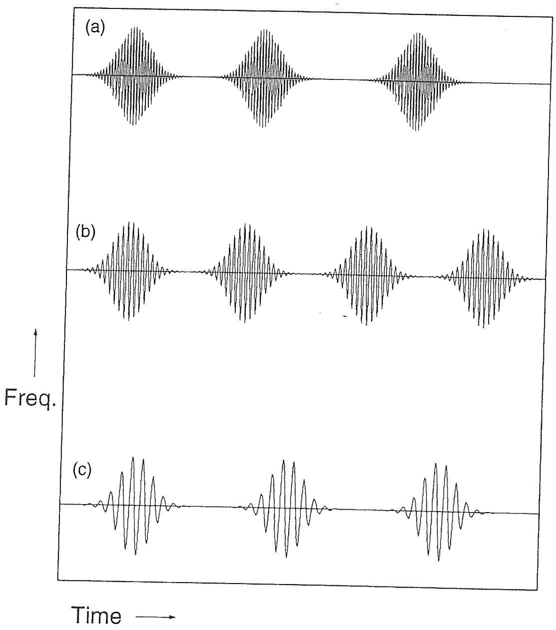

Figure 5.15 Schematic depiction of three streams in quasi-synchronous granular synthesis. The placement of a stream on the vertical axis indicates the grain frequency (i.e., the frequency of the waveform). The onset time between the grains is randomized.

2. Grain duration (usually from 1 to 100 ms, but it can also vary above and below these bounds). The grain duration can be set to a constant, random within limits, derived from a curve, or it can vary as function of the frequency of the grain, where high-frequency grains have shorter envelopes.
3. Density of grains per second; for example, if the grain density is low, then only a few grains are scattered at random points within the cloud. If the grain density is high, grains overlap to create complex spectra. The density can vary over the duration of the cloud.
4. Bandwidth of the cloud, usually specified by two curves that form high- and low-frequency boundaries within which grains are scattered (*cumulus* clouds); alternatively, the frequency of the grains in a cloud can be restricted to a specific set of pitches (as in the *stratus* clouds).
5. Amplitude envelope of the cloud.
6. Waveform(s) within the grains; this is one of the most powerful cloud parameters. For example, each grain in a cloud can have a different waveform; waveforms can be synthetic or sampled.

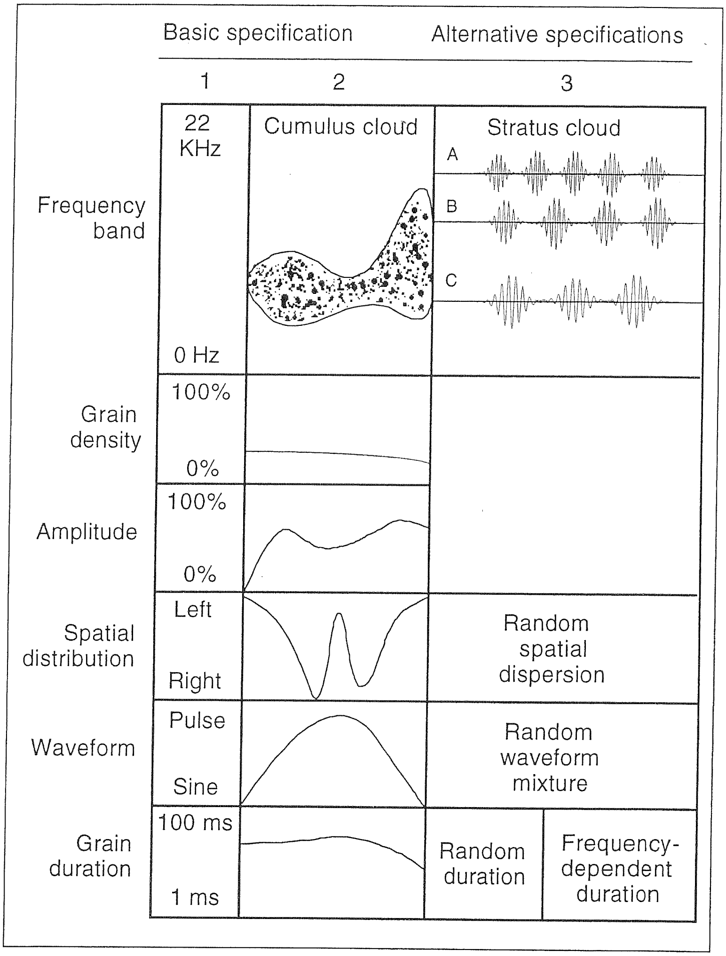

Figure 5.16 Pictorial representation of cloud parameters in asynchronous granular synthesis. The column labeled 1 shows the typical parameter ranges. Column 2 shows basic specifications for standard clouds. Column 3 shows alternative specifications for the frequency band, spatial distribution, waveform, and grain duration parameters.

7. Spatial dispersion of the grains in the cloud, where the number of output channels is specific to a given implementation.

By varying these seven parameters of AGS one can realize a wide range of effects. The rest of this section summarizes, in capsule form, the duration, waveform, frequency band, density, and spatial effects. The waveform and bandwidth parameters apply only to synthetic and not sampled grains. For a more detailed analysis of parametric effects in AGS, see Roads (1991).

As 5.16 shows, grain durations can be either constant (a straight line), variable, random between two limits, or frequency-dependent.

Grain duration changes the sonic texture of a cloud. Short durations lead to crackling, explosive sonorities, while longer durations create a much smoother impression. A profound law of signal processing comes into play in setting the grain duration: the shorter the duration of an event, the greater its bandwidth. Figure 5.17 demonstrates this law for three elementary signals.

Figure 5.18 shows the spectral effects of lowering the grain duration. Notice how the bandwidth expands dramatically as the grain duration shrinks.

Since the waveform can vary on a grain-by-grain basis, we can fill clouds with grains of a single waveform or multiple waveforms. A *monochrome* cloud uses a single waveform, for example, while a *polychrome* cloud contains a random mixture of several waveforms. A *transchrome* cloud mutates statistically from one waveform to another over the duration of the cloud.

For a cumulus cloud (figure 5.19a; see also figure 5.11, column 2), the generator scatters grains randomly within the upper and lower frequency bands. By narrowing these bands to a small interval we can generate pitched sounds. Various types of glissandi are easily achieved (figure 5.19b). An alternative specification is the stratus cloud (figure 5.19c; see also figure 5.11, column 3), where the grains are constrained to fall on a single pitch or specific pitches to create chords and pitch clusters.

The grain density combines with the bandwidth parameter to create various effects. Sparse densities, regardless of bandwidth, create pointillistic textures. At high grain densities, narrow frequency bands create pitched streams with formant spectra, while wide bands (an octave or more) generate massive blocks of sound.

Finally, in AGS as in all forms of granular synthesis, multichannel spatial distribution enhances granular texture. The spatial algorithm of a cloud can involve random scattering or panning effects over the duration of the cloud event.

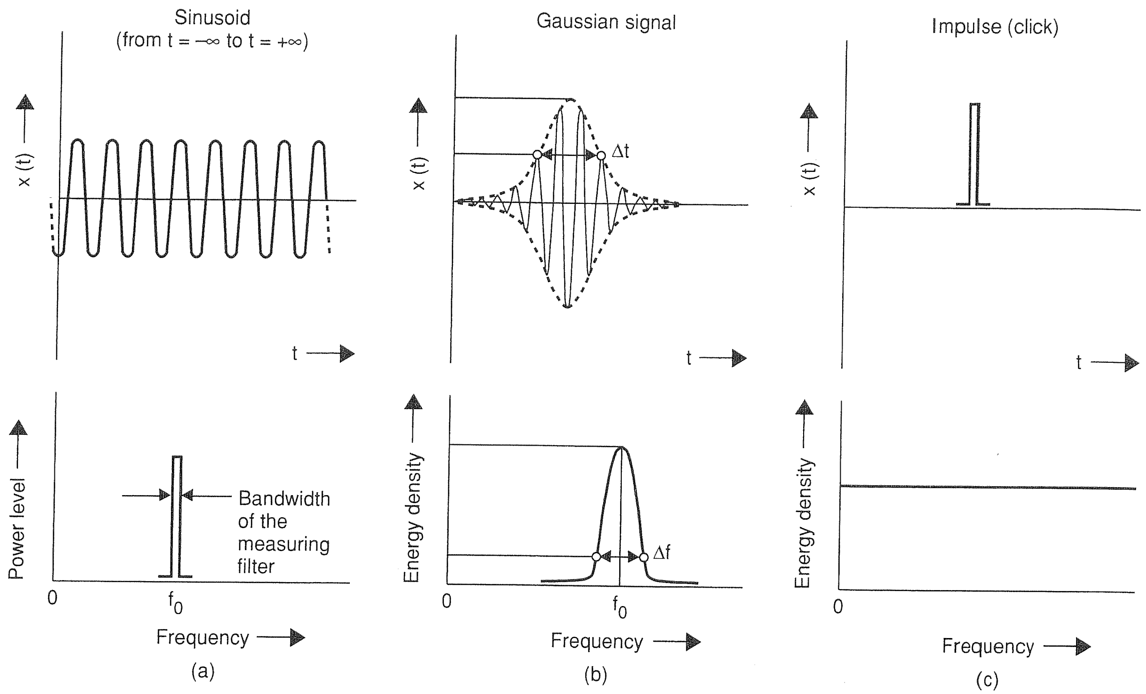

Figure 5.17 Time-domain functions (*top*) and spectra (*bottom*) of three elementary signals, after Blauert (1983). (*a*) Sine wave of infinite duration corresponds to a single line in the spectrum. (*b*) Gaussian grain and corresponding formant spectrum. (*c*) Brief impulse and corresponding infinite spectrum.

Time Granulation of Sampled Sounds

The *time granulation* of recorded (sampled) sounds feeds acoustic material into a kind of logical thrashing machine—delivering grains in a new ordering with a new microrhythm. That is, the *granulator* reads in a small part of a sampled sound (from a sound file or directly from an analog-to-digital converter) and applies an envelope to the portion read in. The order in which this grain is emitted (i.e., its delay) depends on the settings selected by the composer.

Time granulation takes three paths:

1. Granulation of a stored sound file, like a musical note, an animal sound, or a spoken text
2. Continuous real-time granulation of a given input sound or *time scrambling* (Truax 1987, 1988, 1990a, b)
3. Continuous real-time granulation of a given input sound with playback at variable time rate (Truax 1987, 1988, 1990a, b)

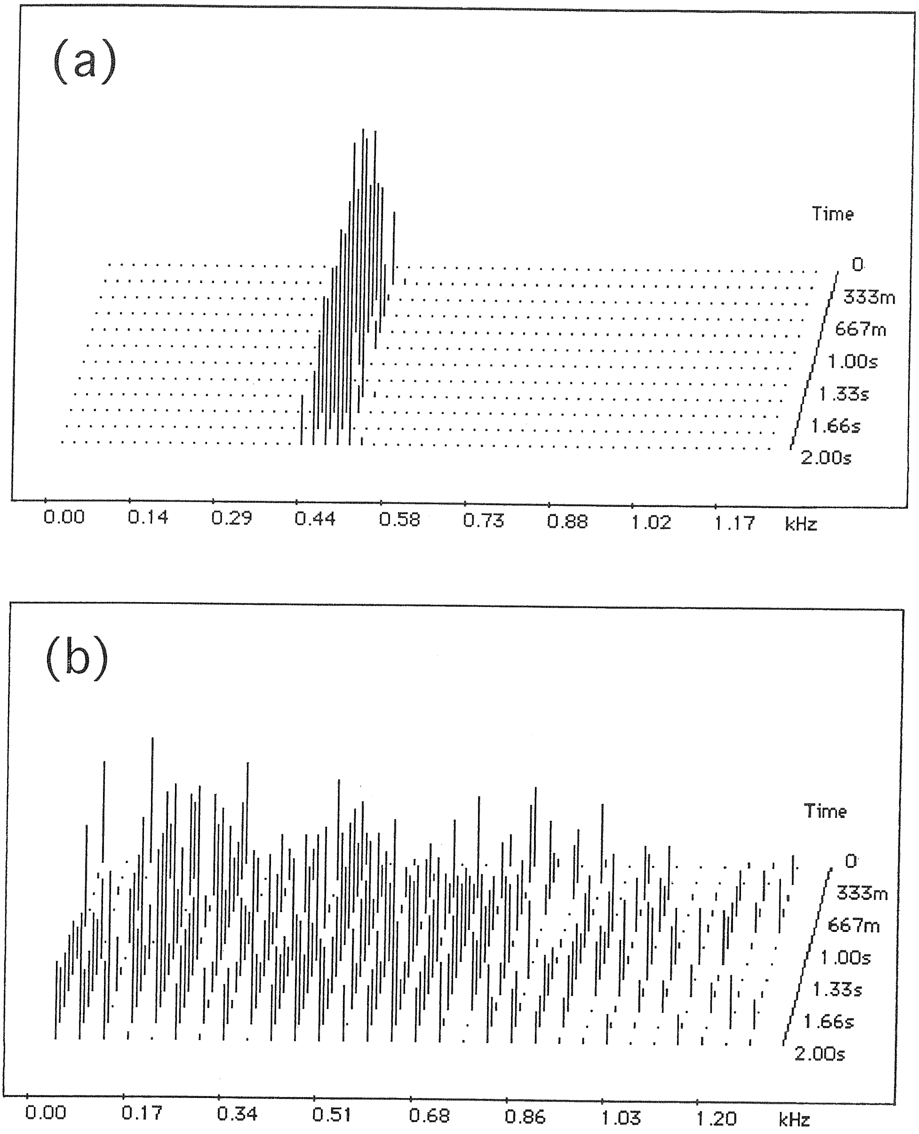

Figure 5.18 Spectral effect of the grain duration. (*a*) Spectrum of a cloud at a constant frequency of 500 Hz with 100 ms grains. Notice the formant region centered at 500 Hz. Time is plotted from back to front. (*b*) Spectrum of a cloud at a constant frequency of 500 Hz but with 1 ms grains. Notice the width of the spectrum.

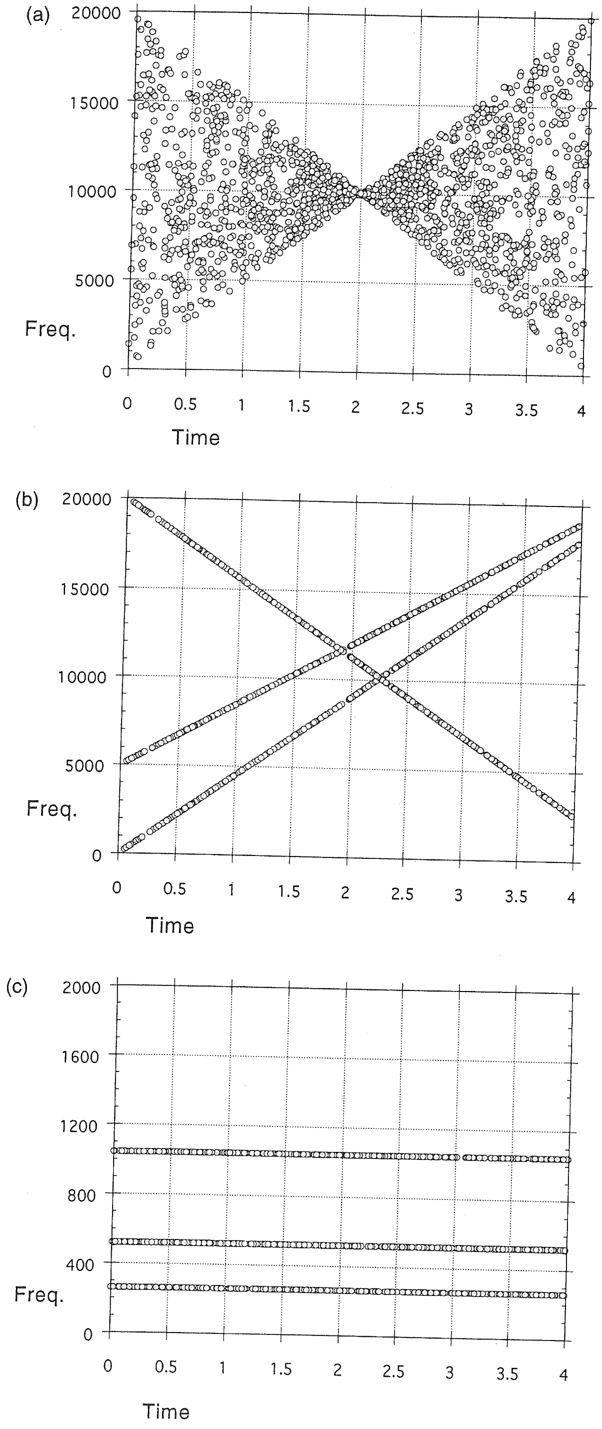

Figure 5.19 Cloud forms. (*a*) Cumulus. (*b*) Glissandi. (*c*) Stratus.

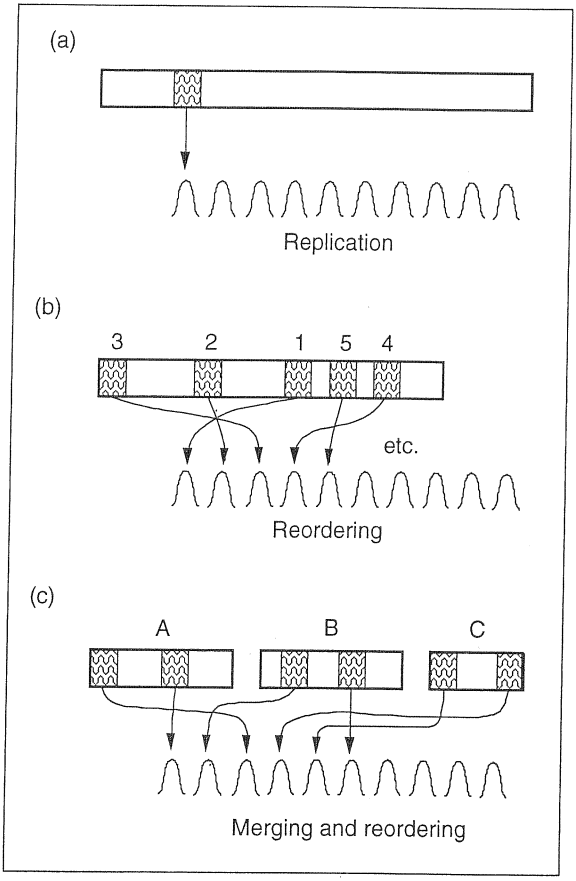

Figure 5.20 Three approaches to time granulation from stored sound files. (*a*) One grain is extracted and turned into a "roll." (*b*) Grains are randomly extracted from a sound file and recordered. (*c*) Grains are randomly chosen from different sound files and reordered. The grains need not be strictly sequential and may overlap.

The first case is the most flexible since one can extract grains from the file in any order. For example, one can extract a single large grain from a snare drum and clone a periodic sequence of hundreds of grains to create a snare drum roll (figure 5.20a). Alternatively the grain generator can sample randomly grains from longer file, such as speech or several notes, thus reordering them (figure 5.20b). An extension of this technique is to randomly sample several sound files and interweave their grains to create multicolored textures (figure 5.20c). These interwoven sound fabrics vary widely depending on the pitch and timbre of the individual grains used within them.

Case (2) above involves real-time granulation of continuous sound with the computer acting as a *delay line* or *window* that can be tapped to furnish the various grains. (See the description of delay lines and taps in chapter

10.) In this case the spectral side effects of the granulation distort and enrich the sound in a controllable way.

Case (3) resembles case (2) except that the playback rate can be varied by a parameter that controls the speed at which synthesis advances through the samples. The playback can vary from normal speed to a slowed-down rate in which a single sample is repeated over and over again. Hence this method can be thought of as an interpolation between case (1) and case (2).

Assessment of Granular Synthesis

Granular synthesis constitutes a diverse body of techniques that share only the concept of sonic grains. The granular representation is purely internal in Fourier and wavelet analysis, hidden from users. Indeed, a technical goal of these methods is creating the illusion of continuous, analog-like signal processing. A granular sonority appears only in pathological distortions such as too large a hop size in overlap-add resynthesis (see chapter 13). The pitch-synchronous analysis/resynthesis of A. Piccialli and his colleagues makes the granular representation more explicit. Techniques like quasi-synchronous granular synthesis (as developed by B. Truax) have been implemented on a variety of platforms.

Asynchronous granular synthesis (AGS) has proven valuable in modeling sounds that would be difficult to describe using earlier techniques. AGS sprays sonic grains into cloudlike formations across the audio spectrum. The result is often a particulated sound complex that can act as a foil to smoother, more sterile sounds emitted by digital oscillators. Time-varying combinations of clouds lead to dramatic effects such as evaporation, coalescence, and mutations created by crossfading overlapping clouds. A striking analogy exists between these processes and those created in the visual domain by *particle synthesis* (Reeves 1983). Particle synthesis has been used to create fire, water, clouds, fog, and grasslike textures, which are analogous to some of the audio effects possible with AGS (crackling fire, water gurgling, windy gusts, explosions). Finally, in combination with time granulation and convolution (Roads 1993a), granular methods are evolving from pure synthesis techniques to sound transformation applications.

Subtractive Synthesis

Subtractive synthesis implies the use of *filters* to shape the spectrum of a source sound. As the source signal passes through a filter, the filter boosts