1 Time Scales of Music

The evolution of musical expression intertwines with the development of musical instruments. This was never more evident than in the twentieth century. Beginning with the gigantic Telharmonium synthesizer unveiled in 1906 (Weidenaar 1989, 1995), research ushered forth a steady stream of electrical and electronic instruments. These have irrevocably molded the musical landscape.

The most precise and flexible electronic music instrument ever conceived is the digital computer. As with the pipe organ, invented centuries earlier, the computer's power derives from its ability to emulate, or in scientific terms, to model phenomena. The models of the computer take the form of symbolic code. Thus it does not matter whether the phenomena being modeled exist outside the circuitry of the machine, or whether they are pure fantasy. This makes the computer an ideal testbed for the representation of musical structure on multiple time scales.

This chapter examines the time scales of music. Our main focus is the micro time scale and its interactions with other time scales. By including extreme time scales—the infinite and the infinitesimal—we situate musical time within the broadest possible context.

Time Scales of Music

Music theory has long recognized a temporal hierarchy of structure in music compositions. A central task of composition has always been the management of the interaction amongst structures on different time scales. Starting from the topmost layer and descending, one can dissect layers of structure, arriving at the bottom layer of individual notes.

This hierarchy, however, is incomplete. Above the level of an individual piece are the cultural time spans defining the oeuvre of a composer or a stylistic period. Beneath the level of the note lies another multilayered stratum, the microsonic hierarchy. Like the quantum world of quarks, leptons, gluons, and bosons, the microsonic hierarchy was long invisible. Modern tools let us view and manipulate the microsonic layers from which all acoustic phenomena emerge. Beyond these physical time scales, mathematics defines two ideal temporal boundaries—the infinite and the infinitesimal—which appear in the theory of musical signal processing.

Taking a comprehensive view, we distinguish nine time scales of music, starting from the longest:

1. *Infinite* The ideal time span of mathematical durations such as the infinite sine waves of classical Fourier analysis.
2. *Supra* A time scale beyond that of an individual composition and extending into months, years, decades, and centuries.
3. *Macro* The time scale of overall musical architecture or form, measured in minutes or hours, or in extreme cases, days.
4. *Meso* Divisions of form. Groupings of sound objects into hierarchies of phrase structures of various sizes, measured in minutes or seconds.
5. *Sound object* A basic unit of musical structure, generalizing the traditional concept of note to include complex and mutating sound events on a time scale ranging from a fraction of a second to several seconds.

6. *Micro* Sound particles on a time scale that extends down to the threshold of auditory perception (measured in thousandths of a second or milliseconds).
7. *Sample* The atomic level of digital audio systems: individual binary samples or numerical amplitude values, one following another at a fixed time interval. The period between samples is measured in millionths of a second (microseconds).
8. *Subsample* Fluctuations on a time scale too brief to be properly recorded or perceived, measured in billionths of a second (nanoseconds) or less.
9. *Infinitesimal* The ideal time span of mathematical durations such as the infinitely brief delta functions.

Figure 1.1 portrays the nine time scales of the time domain. Notice in the middle of the diagram, in the frequency column, a line indicating "Conscious time, the present (∼600 ms)." This line marks off Winckel's (1967) estimate of the "thickness of the present." The thickness extends to the line at the right indicating the physical NOW. This temporal interval constitutes an estimate of the accumulated lag time of the perceptual and cognitive mechanisms associated with hearing. Here is but one example of a disparity between *chronos*—physical time, and *tempus*—perceived time (Küpper 2000).

The rest of this chapter explains the characteristics of each time scale in turn. We will, of course, pay particular attention to the micro time scale.

Boundaries between Time Scales

As sound passes from one time scale to another it crosses perceptual boundaries. It seems to change quality. This is because human perception processes each time scale differently. Consider a simple sinusoid transposed to various time scales (1 µsec, 1 ms, 1 sec, 1 minute, 1 hour). The waveform is identical, but one would have difficulty classifying these auditory experiences in the same family.

In some cases the borders between time scales are demarcated clearly; ambiguous zones surround others. Training and culture condition perception of the time scales. To hear a flat pitch or a dragging beat, for example, is to detect a temporal anomaly on a micro scale that might not be noticed by other people.

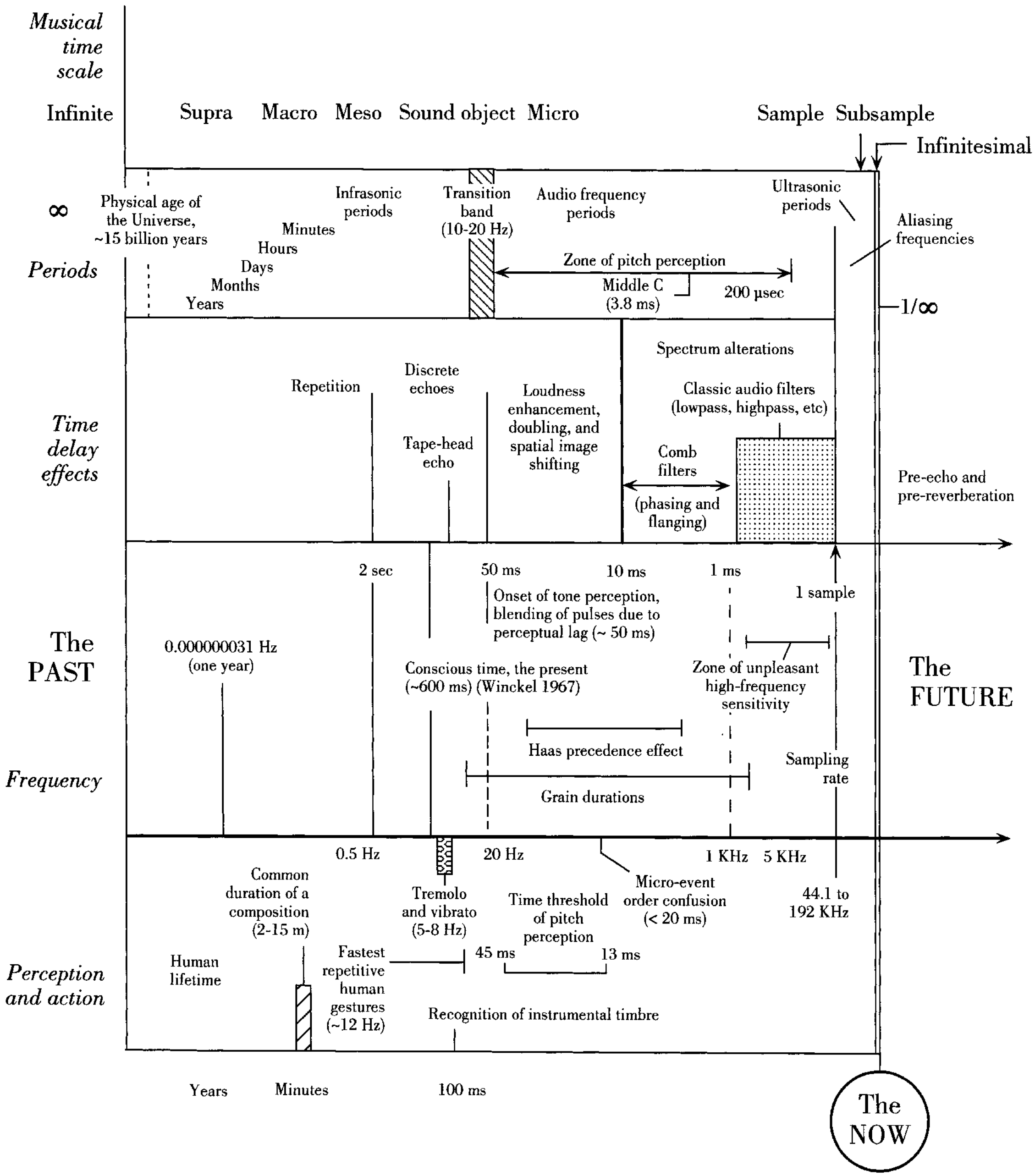

Figure 1.1 The time domain, segmented into periods, time delay effects, frequencies, and perception and action. Note that time intervals are not drawn to scale.

Digital audio systems, such as compact disc players, operate at a fixed sampling frequency. This makes it easy to distinguish the exact boundary separating the sample time scale from the subsample time scale. This boundary is the Nyquist frequency, or the sampling frequency divided by two. The effect of crossing this boundary is not always perceptible. In noisy sounds, aliased frequencies from the subsample time domain may mix unobtrusively with high frequencies in the sample time domain.

The border between certain other time scales is context-dependent. Between the sample and micro time scales, for example, is a region of transient events—too brief to evoke a sense of pitch but rich in timbral content. Between the micro and the object time scales is a stratum of brief events such as short staccato notes. Another zone of ambiguity is the border between the sound object and meso levels, exemplified by an evolving texture. A texture might contain a statistical distribution of micro events that are perceived as a unitary yet time-varying sound.

Time scales interlink. A given level encapsulates events on lower levels and is itself subsumed within higher time scales. Hence to operate on one level is to affect other levels. The interaction between time scales is not, however, a simple relation. Linear changes on a given time scale do not guarantee a perceptible effect on neighboring time scales.

Zones of Intensity and Frequency

Sound is an alternation in pressure, particle displacement, or particle velocity propagated in an elastic material. (Olson 1957)

Before we continue further, a brief discussion of acoustic terminology might be helpful. In scientific parlance—as opposed to popular usage—the word "sound" refers not only to phenomena in air responsible for the sensation of hearing but also "whatever else is governed by analogous physical principles" (Pierce 1994). Sound can be defined in a general sense as mechanical radiant energy that is transmitted by pressure waves in a material medium. Thus besides the airborne frequencies that our ears perceive, one may also speak of underwater sound, sound in solids, or structure-borne sound. Mechanical vibrations even take place on the atomic level, resulting in quantum units of sound energy called phonons. The term "acoustics" likewise is independent of air and of human perception. It is distinguished from *optics* in that it involves mechanical—rather than electromagnetic, wave motion.

Corresponding to this broad definition of sound is a very wide range of transient, chaotic, and periodic fluctuations, spanning frequencies that are both higher and lower than the human ear can perceive. The *audio* frequencies, traditionally said to span the range of about 20 Hz to 20 kHz are perceptible to the ear. The specific boundaries vary depending on the individual.

Vibrations at frequencies too low to be heard as continuous tones can be perceived by the ear as well as the body. These are the infrasonic impulses and vibrations, in the range below about 20 Hz. The infectious rhythms of the percussion instruments fall within this range.

Ultrasound includes the domain of high frequencies above the range of human audibility. The threshold of ultrasound varies according to the individual, their age, and the test conditions. Science and industry use ultrasonic techniques in a variety of applications, such as acoustic imaging (Quate 1998) and highly directional loudspeakers (Pompei 1998).

Some sounds are too soft to be perceived by the human ear, such as a caterpillar's delicate march across a leaf. This is the zone of *subsonic* intensities.

Other sounds are so loud that to perceive them directly is dangerous, since they are destructive to the human body. Sustained exposure to sound levels around 120 dB leads directly to pain and hearing loss. Above 130 dB, sound is felt by the exposed tissues of the body as a painful pressure wave (Pierce 1983). This dangerous zone extends to a range of destructive acoustic phenomena. The force of an explosion, for example, is an intense acoustic shock wave.

For lack of a better term, we call these *perisonic* intensities (from the Latin *periculos* meaning "dangerous"). The *audible* intensities fall between these two ranges. Figure 1.2 depicts the zones of sound intensity and frequency. The α zone in the center is where audio frequencies intersect with audible intensities, enabling hearing. Notice that the α zone is but a tiny fraction of a vast range of sonic phenomena.

Following this discussion of acoustical terms, let us proceed to the main theme of this chapter, the time scales of music.

Infinite Time Scale

Complex Fourier analysis regards the signal sub specie aeternitatis. (Gabor 1952)

The human experience of musical time is linked to the ticking clock. It is natural to ask: when did the clock begin to tick? Will it tick forever? At the

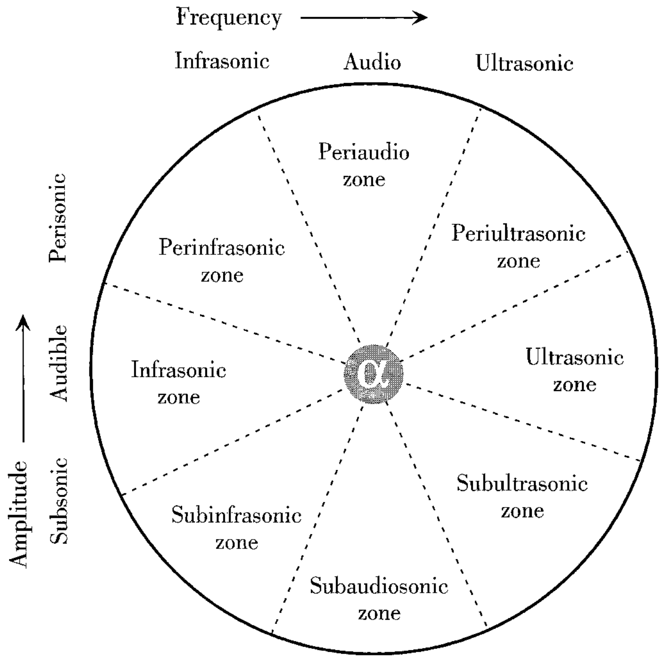

Figure 1.2 Zones of intensities and frequencies. Only the zone marked α is audible to the ear. This zone constitutes a tiny portion of the range of sound phenomena.

extreme upper boundary of all time scales is the mathematical concept of an infinite time span. This is a logical extension of the infinite series, a fundamental notion in mathematics. An infinite series is a sequence of numbers $u_1, u_2, u_3 \ldots$ arranged in a prescribed order and formed according to a particular rule. Consider this infinite series:

$$\sum_{i=1}^{\infty} u_i = u_1 + u_2 + u_3 + \cdots$$

This equation sums a set of numbers u_i, where the index i goes from 1 to ∞. What if each number u_i corresponded to a tick of a clock? This series would then define an infinite duration. This ideal is not so far removed from music as it may seem. The idea of infinite duration is implicit in the theory of Fourier analysis, which links the notion of frequency to sine waves of infinite duration. As chapter 6 shows, Fourier analysis has proven to be a useful tool in the analysis and transformation of musical sound.

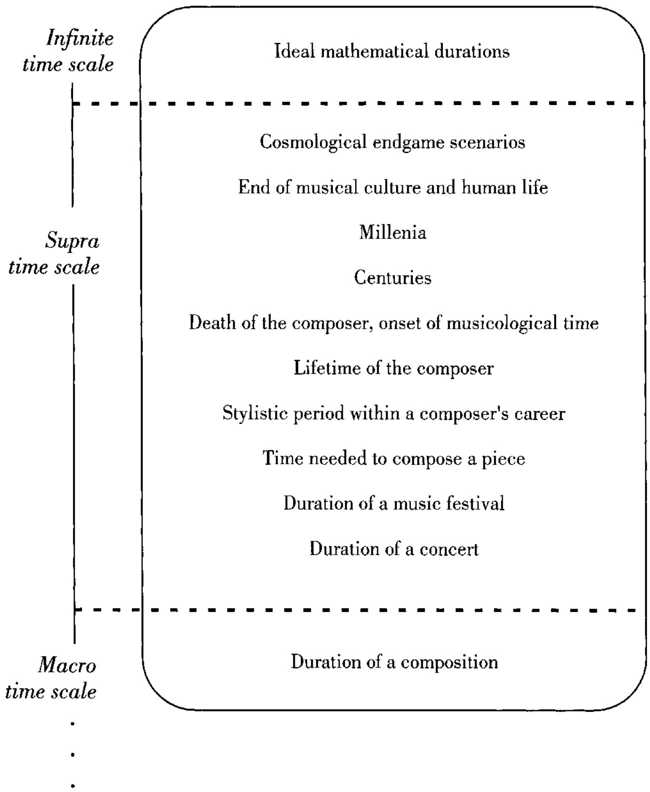

Figure 1.3 The scope of the supratemporal domain.

Supra Time Scale

The supra time scale spans the durations that are beyond those of an individual composition. It begins as the applause dies out after the longest compositions, and extends into weeks, months, years, decades, and beyond (figure 1.3). Concerts and festivals fall into this category. So do programs from music broadcasting stations, which may extend into years of more-or-less continuous emissions.

Musical cultures are constructed out of supratemporal bricks: the eras of instruments, of styles, of musicians, and of composers. Musical education takes years; cultural tastes evolve over decades. The perception and appreciation of a single composition may change several times within a century. The entire history of music transpires within the supratemporal scale, starting from the earliest known musical instrument, a Neanderthal flute dating back some 45,000 years (Whitehouse 1999).

Composition is itself a supratemporal activity. Its results last only a fraction of the time required for its creation. A composer may spend a year to complete a ten-minute piece. Even if the composer does not work every hour of every day, the ratio of 52,560 minutes passed for every 1 minute composed is still significant. What happens in this time? Certain composers design a complex strategy as prelude to the realization of a piece. The electronic music composer may spend considerable time in creating the sound materials of the work. Either of these tasks may entail the development of software. Virtually all composers spend time experimenting, playing with material in different combinations. Some of these experiments may result in fragments that are edited or discarded, to be replaced with new fragments. Thus it is inevitable that composers invest time pursuing dead ends, composing fragments that no one else will hear. This backtracking is not necessarily time wasted; it is part of an important feedback loop in which the composer refines the work. Finally we should mention documentation. While only a few composers document their labor, these documents may be valuable to those seeking a deeper understanding of a work and the compositional process that created it. Compare all this with the efficiency of the real-time improviser!

Some music spans beyond the lifetime of the individual who composed it, through published notation, recordings, and pedagogy. Yet the temporal reach of music is limited. Many compositions are performed only once. Scores, tapes, and discs disappear into storage, to be discarded sooner or later. Music-making presumably has always been part of the experience of *Homo sapiens*, who it is speculated came into being some 200,000 years ago. Few traces remain of anything musical older than a dozen centuries. Modern electronic instruments and recording media, too, are ephemeral. Will human musical vibrations somehow outlast the species that created them? Perhaps the last trace of human existence will be radio waves beamed into space, traveling vast distances before they dissolve into noise.

The upper boundary of time, as the concept is currently understood, is the age of the physical universe. Some scientists estimate it to be approximately fifteen billion years (Lederman and Scramm 1995). Cosmologists continue to debate how long the universe may expand. The latest scientific theories continue to twist the notion of time itself (see, for example, Kaku 1995; Arkani-Hamed et al. 2000).

Macro Time Scale

The macro level of musical time corresponds to the notion of form, and encompasses the overall architecture of a composition. It is generally measured in minutes. The upper limit of this time scale is exemplified by such marathon compositions as Richard Wagner's *Ring* cycle, the Japanese Kabuki theater, Jean-Claude Eloy's evening-long rituals, and Karlheinz Stockhausen's opera *Licht* (spanning seven days and nights). The literature of opera and contemporary music contains many examples of music on a time scale that exceeds two hours. Nonetheless, the vast majority of music compositions realized in the past century are less than a half-hour in duration. The average duration is probably in the range of a kilosecond (16 min 40 sec). Complete compositions lasting less than a hectosecond (1 min 40 sec) are rare.

Perception of the Macro Time Scale

Unless the musical form is described in advance of performance (through program notes, for example), listeners perceive the macro time scale in retrospect, through recollection. It is common knowledge that the remembrance of things past is subject to strong discontinuities and distortions. We cannot recall time as a linearly measured flow. As in everyday life, the perceived flow of musical time is linked to reference events or memories that are tagged with emotional significance.

Classical music (Bach, Mozart, Beethoven, etc.) places reference events at regular intervals (cadences, repetition) to periodically orient the listener within the framework of the form. Some popular music takes this to an extreme, reminding listeners repeatedly on a shorter time base.

Subjective factors play into a distorted sense of time. Was the listener engaged in aesthetic appreciation of the work? Were they paying attention? What is their musical taste, their training? Were they preoccupied with stress and personal problems? A composition that we do not understand or like appears to expand in time as we experience it, yet vanishes almost immediately from memory.

The perception of time flow also depends on the objective nature of the musical materials. Repetition and a regular pulse tend to carry a work efficiently through time, while an unchanging, unbroken sound (or silence) reduces the flow to a crawl.

The ear's sensitivity to sound is limited in duration. Long continuous noises or regular sounds in the environment tend to disappear from consciousness and are noticed again only when they change abruptly or terminate.

Macroform

Just as musical time can be viewed in terms of a hierarchy of time scales, so it is possible to imagine musical structure as a tree in the mathematical sense. Mathematical trees are inverted, that is, the uppermost level is the root symbol, representing the entire work. The root branches into a layer of macrostructure encapsulating the major parts of the piece. This second level is the form: the arrangement of the major sections of the piece. Below the level of form is a syntactic hierarchy of branches representing mesostructures that expand into the terminal level of sound objects (Roads 1985d).

To parse a mathematical tree is straightforward. Yet one cannot parse a sophisticated musical composition as easily as a compiler parses a computer program. A compiler references an unambiguous formal grammar. By contrast, the grammar of music is ambiguous—subject to interpretation, and in a perpetual state of evolution. Compositions may contain overlapping elements (on various levels) that cannot be easily segmented. The musical hierarchy is often fractured. Indeed, this is an essential ingredient of its fascination.

Design of Macroform

The design of macroform takes one of two contrasting paths: *top-down* or *bottom-up*. A strict top-down approach considers macrostructure as a preconceived global plan or template whose details are filled in by later stages of composition. This corresponds to the traditional notion of form in classical music, wherein certain formal schemes have been used by composers as molds (Apel 1972). Music theory textbooks catalog the generic classical forms (Leichtentritt 1951) whose habitual use was called into question at the turn of the twentieth century. Claude Debussy, for example, discarded what he called "administrative forms" and replaced them with fluctuating mesostructures through a chain of associated variations. Since Debussy, composers have written a tremendous amount of music not based on classical forms. This music is full of local detail and eschews formal repetition. Such structures resist classification within the catalog of standard textbook forms. Thus while musical form has continued to evolve in practice in the past century, the acknowledged catalog of generic forms has hardly changed.

This is not to say that the use of preconceived forms has died away. The practice of top-down planning remains common in contemporary composition. Many composers predetermine the macrostructure of their pieces according to a more-or-less formal scheme before a single sound is composed.

By contrast, a strict bottom-up approach conceives of form as the result of a process of internal development provoked by interactions on lower levels of musical structure. This approach was articulated by Edgard Varèse (1971), who said, "Form is a result—the result of a process." In this view, macrostructure articulates processes of attraction and repulsion (for example, in the rhythmic and harmonic domains) unfolding on lower levels of structure.

Manuals on traditional composition offer myriad ways to project low-level structures into macrostructure:

> *Smaller forms may be expanded by means of external repetitions, sequences, extensions, liquidations and broadening of connectives. The number of parts may be increased by supplying codettas, episodes, etc. In such situations, derivatives of the basic motive are formulated into new thematic units.* (Schoenberg 1967)

Serial or germ-cell approaches to composition expand a series or a formula through permutation and combination into larger structures.

In the domain of computer music, a frequent technique for elaboration is to time-expand a sound fragment into an evolving sound mass. Here the unfolding of sonic microstructure rises to the temporal level of a harmonic progression.

A different bottom-up approach appears in the work of the conceptual and chance composers, following in the wake of John Cage. Cage (1973) often conceived of form as arising from a series of accidents—random or improvised events occurring on the sound object level. For Cage, form (and indeed sound) was a side-effect of a conceptual strategy. Such an approach often results in discontinuous changes in sound structure. This was not accidental; Cage disdained continuity in musical structure, always favoring juxtaposition:

> *Where people had felt the necessity to stick sounds together to make a continuity, we felt the necessity to get rid of the glue so that sounds would be themselves.* (Cage 1959)

For some, composition involves a mediation between the top-down and bottom-up approaches, between an abstract high-level conception and the concrete materials being developed on lower levels of musical time structure. This implies negotiation between a desire for orderly macrostructure and imperatives that emerge from the source material. Certain phrase structures cannot be encapsulated neatly within the box of a precut form. They mandate a container that conforms to their shape and weight.

The debate over the emergence of form is ancient. Musicologists have long argued whether, for example, a fugue is a template (form) or a process of variation. This debate echoes an ancient philosophical discourse pitting form against flux, dating back as far as the Greek philosopher Heraclitus. Ultimately, the dichotomy between form and process is an illusion, a failure of language to bind two aspects of the same concept into a unit. In computer science, the concept of *constraints* does away with this dichotomy (Sussman and Steele 1981). A form is constructed according to a set of relationships. A set of relationships implies a process of evaluation that results in a form.

Meso Time Scale

The mesostructural level groups sound objects into a quasi hierarchy of phrase structures of durations measured in seconds. This *local* as opposed to *global* time scale is extremely important in composition, for it is most often on the meso level that the sequences, combinations, and transmutations that constitute musical ideas unfold. Melodic, harmonic, and contrapuntal relations happen here, as do processes such as theme and variations, and many types of development, progression, and juxtaposition. Local rhythmic and metric patterns, too, unfold on this stratum.

Wishart (1994) called this level of structure the *sequence*. In the context of electronic music, he identified two properties of sequences: the *field* (the material, or set of elements used in the sequence), and the *order*. The field serves as a lexicon—the vocabulary of a piece of music. The order determines thematic relations—the grammar of a particular piece. As Wishart observed, the field and the order must be established quickly if they are to serve as the bearers of musical code. In traditional music, they are largely predetermined by cultural norms.

In electronic music, the meso layer presents timbre melodies, simultaneities (chord analogies), spatial interplay, and all manner of textural evolutions. Many of these processes are described and classified in Denis Smalley's interesting theory of *spectromorphology*—a taxonomy of sound gesture shapes (Smalley 1986, 1997).

Sound Masses, Textures, and Clouds

To the sequences and combinations of traditional music, we must add another principle of organization on the meso scale: the sound mass. Decades ago, Edgard Varèse predicted that the sounds introduced by electronic instruments would necessitate new organizing principles for mesostructure.

> *When new instruments will allow me to write music as I conceive it, taking the place of the linear counterpoint, the movement of sound masses, or shifting planes, will be clearly perceived. When these sound masses collide the phenomena of penetration or repulsion will seem to occur.* (Varèse 1962)

A trend toward shaping music through the global attributes of a sound mass began in the 1950s. One type of sound mass is a cluster of sustained frequencies that fuse into a solid block. In a certain style of sound mass composition, musical development unfolds as individual lines are added to or removed from this cluster. György Ligeti's *Volumina* for organ (1962) is a masterpiece of this style, and the composer has explored this approach in a number of other pieces, including *Atmosphères* (1961) and *Lux Aeterna* (1966).

Particles make possible another type of sound mass: statistical clouds of microevents (Xenakis 1960). Wishart (1994) ascribed two properties to cloud textures. As with sequences, their field is the set of elements used in the texture, which may be constant or evolving. Their second property is density, which stipulates the number of events within a given time period, from sparse scatterings to dense scintillations.

Cloud textures suggest a different approach to musical organization. In contrast to the combinatorial sequences of traditional meso structure, clouds encourage a process of statistical evolution. Within this evolution the composer can impose specific morphologies. Cloud evolutions can take place in the domain of amplitude (crescendi/decrescendi), internal tempo (accelerando/rallentando), density (increasing/decreasing), harmonicity (pitch/chord/cluster/noise, etc.), and spectrum (high/mid/low, etc.).

Xenakis's tape compositions *Concret PH* (1958), *Bohor I* (1962), and *Persepolis* (1971) feature dense, monolithic clouds, as do many of his works for traditional instruments. Stockhausen (1957) used statistical form-criteria as one component of his early composition technique. Since the 1960s, particle textures have appeared in numerous electroacoustic compositions, such as the remarkable *De natura sonorum* (1975) of Bernard Parmegiani.

Varèse spoke of the interpenetration of sound masses. The diaphanous nature of cloud structures makes this possible. A crossfade between two clouds results in a smooth mutation. Mesostructural processes such as disintegration and coalescence can be realized through manipulations of particle density (see chapter 6). Density determines the transparency of the material. An increase in density lifts a cloud into the foreground, while a decrease causes evaporation, dissolving a continuous sound band into a pointillist rhythm or vaporous background texture.

Cloud Taxonomy

To describe sound clouds precisely, we might refer to the taxonomy of cloud shapes in the atmosphere:

Cumulus well-defined cauliflower-shaped cottony clouds

Stratocumulus blurred by wind motion

Stratus a thin fragmented layer, often translucent

Nimbostratus a widespread gray or white sheet, opaque

Cirrus isolated sheets that develop in filaments or patches

In another realm, among the stars, outer space is filled with swirling clouds of cosmic raw material called *nebulae*.

> *The cosmos, like the sky on a turbulent summer day, is filled with clouds of different sizes, shapes, structures, and distances. Some are swelling cumulus, others light, wispy cirrus—all of them constantly changing colliding, forming, and evaporating.* (Kaler 1997)

Pulled by immense gravitational fields or blown by cosmic shockwaves, nebulae form in great variety: dark or glowing, amorphous or ring-shaped, constantly evolving in morphology. These forms, too, have musical analogies. Programs for sonographic synthesis (such as MetaSynth [Wenger and Spiegel 1999]), provide airbrush tools that let one spray sound particles on the time-frequency canvas. On the screen, the vertical dimension represents frequency, and the horizontal dimension represents time. The images can be blurred, fragmented, or separated into sheets. Depending on their density, they may be translucent or opaque. Displacement maps can warp the cloud into a circular or spiral shape on the time-frequency canvas. (See chapter 6 on sonographic transformation of sound.)

Sound Object Time Scale

The sound object time scale encompasses events of a duration associated with the elementary unit of composition in scores: the note. A note usually lasts from about 100 ms to several seconds, and is played by an instrument or sung by a vocalist. The concept of sound object extends this to allow any sound, from any source. The term *sound object* comes from Pierre Schaeffer, the pioneer of *musique concrète*. To him, the pure *objet sonore* was a sound whose origin a listener could not identify (Schaeffer 1959, 1977, p. 95). We take a broader view here. Any sound within stipulated temporal limits is a sound object. Xenakis (1989) referred to this as the "ministructural" time scale.

The Sensation of Tone

The sensation of tone—a sustained or continuous event of definite or indefinite pitch—occurs on the sound object time scale. The low-frequency boundary for the sensation of a continuous sound—as opposed to a fluttering succession of brief microsounds—has been estimated at anywhere from 8 Hz (Savart) to about 30 Hz. (As reference, the deepest sound in a typical orchestra is the open E of the contrabass at 41.25 Hz.) Helmholtz, the nineteenth century German acoustician, investigated this lower boundary.

> *In the first place it is necessary that the strength of the vibrations of the air for very low tones should be extremely greater than for high tones. The increase in strength ... is of especial consequence in the deepest tones. ... To discover the limit of the deepest tones it is necessary not only to produce very violent agitations in the air but to give these a simple pendular motion.* (Helmholtz 1885)

Helmholtz observed that a sense of continuity takes hold between 24 to 28 Hz, but that the impression of a definite pitch does not take hold until 40 Hz.

Pitch and tone are not the same thing. Acousticians speak of *complex tones* and *unpitched tones*. Any sound perceived as continuous is a tone. This can, for example include noise.

Between the sensation of a continuous tone and the sensation of metered rhythm stands a zone of ambiguity, an infrasonic frequency domain that is too slow to form a continuous tone but too fast for rhythmic definition. Thus continuous tone is a possible quality, but not a necessary property, of a sound object. Consider a relatively dense cloud of sonic grains with short silent gaps on the order of tens of milliseconds. Dozens of different sonic events occur per second, each unique and separated by a brief intervals of zero amplitude, yet such a cloud is perceived as a unitary event—a single sound object.

A sense of regular pulse and meter begins to occur from approximately 8 Hz down to 0.12 Hz and below (Fraisse 1982). Not coincidentally, it is in this rhythmically apprensible range that the most salient and expressive vibrato, tremolo, and spatial panning effects occur.

Homogeneous Notes versus Heterogeneous Sound Objects

The sound object time scale is the same as that of traditional notes. What distinguishes sound objects from notes? The note is the homogeneous brick of conventional music architecture. *Homogeneous* means that every note can be described by the same four properties:

- *pitch*, generally one of twelve equal-tempered pitch classes
- *timbre*, generally one of about twenty different instruments for a full orchestra, with two or three different attack types for each instrument
- *dynamic marking*, generally one of about ten different relative levels
- *duration*, generally between ~100 ms (slightly less than a thirty-second note at a tempo of 60 M.M.) to ~8 seconds (for two tied whole notes)

These properties are static, guaranteeing that, in theory, a note in one measure with a certain pitch, dynamic, and instrumental timbre is functionally equivalent to a note in another measure with the same three properties. The properties of a pair of notes can be compared on a side-by-side basis and a distance or interval can be calculated. The notions of equivalence and distance lead to the notion of *invariants*, or intervallic distances that are preserved across transformations.

Limiting material to a static homogeneous set allows abstraction and efficiency in musical language. It serves as the basis for operations such as transposition, orchestration and reduction, the algebra of tonal harmony and counterpoint, and the atonal and serial manipulations. In the past decade, the MIDI protocol has extended this homogeneity into the domain of electronic music through standardized note sequences that play on any synthesizer.

The merit of this homogeneous system is clear; highly elegant structures having been built with standard materials inherited from centuries past. But since the dawn of the twentieth century, a recurring aesthetic dream has been the expansion beyond a fixed set of homogeneous materials to a much larger superset of heterogeneous musical materials.

What we have said about the limitations of the European note concept does not necessarily apply to the musics of other cultures. Consider the shakuhachi music of Japan, or contemporary practice emerging from the advanced developments of jazz.

Heterogeneity means that two objects may not share common properties. Therefore their percept may be entirely different. Consider the following two examples. Sound A is a brief event constructed by passing analog diode noise through a time-varying bandpass filter and applying an exponentially decaying envelope to it. Sound B lasts eight seconds. It is constructed by granulating in multiple channels several resonant low-pitched strokes on an African slit drum, then reverberating the texture. Since the amplitudes and onset times of the grains vary, this creates a jittering sound mass. To compare A and B is like comparing apples and oranges. Their microstructures are different, and we can only understand them through the properties that they do not have in common.

Thus instead of homogeneous notes, we speak of *heterogeneous sound objects*. The notion of sound object generalizes the note concept in two ways:

1. It puts aside the restriction of a common set of properties in favor of a heterogeneous collection of properties. Some objects may not share common properties with other objects. Certain sound objects may function as unique singularities. Entire pieces may be constructed from nothing but such singularities.
2. It discards the notion of static, time-invariant properties in favor of time-varying properties (Roads 1985b).

Objects that do not share common properties may be separated into diverse classes. Each class will lend itself to different types of manipulation and musical organization. Certain sounds layer well, nearly any mixture of elongated sine waves with smooth envelopes for example. The same sounds organized in a sequence, however, rather quickly become boring. Other sounds, such as isolated impulses, are most effective when sparsely scattered onto a neutral sound canvas.

Transformations applied to objects in one class may not be effective in another class. For example, a time-stretching operation may work perfectly well on a pipe organ tone, preserving its identity and affecting only its duration. The same operation applied to the sound of burning embers will smear the crackling transients into a nondescript electronic blur.

In traditional western music, the possibilities for transition within a note are limited by the physical properties of the acoustic instrument as well as frozen by theory and style. Unlike notes, the properties of a sound object are free to vary over time. This opens up the possibility of complex sounds that can mutate from one state to another within a single musical event. In the case of synthesized sounds, an object may be controlled by multiple time-varying envelopes for pitch, amplitude, spatial position, and multiple determinants of timbre. These variations may take place over time scales much longer than those associated with conventional notes.

We can subdivide a sound object not only by its properties but also by its temporal states. These states are composable using synthesis tools that operate on the microtime scale. The micro states of a sound can also be decomposed and rearranged with tools such as time granulators and analysis-resynthesis software.

Sound Object Morphology

In music, as in other fields, the organization is conditioned by the material. (Schaeffer 1977, p. 680)

The desire to understand the enormous range of possible sound objects led Pierre Schaeffer to attempt to classify them, beginning in the early 1950s (Schaeffer and Moles 1952). Book V of his *Traité des objets musicaux* (1977), entitled *Morphologie and typologie des objets sonores* introduces the useful notion of *sound object morphology*—the comparison of the shape and evolution of sound objects. Schaeffer borrowed the term *morphology* from the sciences, where it refers to the study of form and structure (of organisms in biology, of word-elements in linguistics, of rocks in geology, etc.). Schaeffer diagrammed sound shape in three dimensions: the harmonic (spectrum), dynamic (amplitude), and melodic (pitch). He observed that the elements making up a complex sound can be perceived as either merged to form a *sound compound*, or remaining separate to form a *sound mixture*. His typology, or classification of sound objects into different groups, was based on acoustic morphological studies.

The idea of sound morphology remains central to the theory of electroacoustic music (Bayle 1993), in which the musical spotlight is often shone on the sound object level. In traditional composition, transitions function on the mesostructural level through the interplay of notes. In electroacoustic music, the morphology of an individual sound may play a structural role, and transitions can occur within an individual sound object. This ubiquity of mutation means that every sonic event is itself a potential transformation.

Micro Time Scale

The micro time scale is the main subject of this book. It embraces transient audio phenomena, a broad class of sounds that extends from the threshold of timbre perception (several hundred microseconds) up to the duration of short sound objects (~100 ms). It spans the boundary between the audio frequency range (approximately 20 Hz to 20 kHz) and the infrasonic frequency range (below 20 Hz). Neglected in the past owing to its inaccessibility, the microtime domain now stands at the forefront of compositional interest.

Microsound is ubiquitous in the natural world. Transient events unfold all around in the wild: a bird chirps, a twig breaks, a leaf crinkles. We may not take notice of microacoustical events until they occur en masse, triggering a global statistical percept. We experience the interactions of microsounds in the sound of a spray of water droplets on a rocky shore, the gurgling of a brook, the pitter-patter of rain, the crunching of gravel being walked upon, the snapping of burning embers, the humming of a swarm of bees, the hissing of rice grains poured into a bowl, and the crackling of ice melting. Recordings of dolphins reveal a language made up entirely of high-frequency clicking patterns.

One could explore the microsonic resources of any musical instrument in its momentary bursts and infrasonic flutterings, (a study of traditional instruments from this perspective has yet to be undertaken). Among unpitched percussion, we find microsounds in the angled rainstick, (shaken) small bells, (grinding) ratchet, (scraped) guiro, (jingling) tambourine, and the many varieties of rattles. Of course, the percussion roll—a granular stick technique—can be applied to any surface, pitched or unpitched.

In the literature of acoustics and signal processing, many terms refer to similar microsonic phenomena: *acoustic quantum*, *sonal atom*, *grain*, *glisson*, *grainlet*, *trainlet*, *Gaussian elementary signal*, *Gaussian pulse*, *short-time segment*, *sliding window*, *microarc*, *voicel*, *Coiflet*, *symmlet*, *Gabor atom*, *Gabor wavelet*, *gaborette*, *wavelet*, *chirplet*, *Liénard atom*, *FOF*, *FOG*, *wave packet*, *Vosim pulse*, *time-frequency atom*, *pulsar*, *waveset*, *impulse*, *toneburst*, *tone pip*, *acoustic pixel*, and *window function pulse* are just a few. These phenomena, viewed in their mathematical dual space—the frequency domain—take on a different set of names: *kernel*, *logon*, and *frame*, for example.

Perception of Microsound

Microevents last only a very short time, near to the threshold of auditory perception. Much scientific study has gone into the perception of microevents. Human hearing mechanisms, however, intertwine with brain functions, cognition, and emotion, and are not completely understood. Certain facts are clear.

One cannot speak of a single time frame, or a *time constant* for the auditory system (Gordon 1996). Our hearing mechanisms involve many different agents, each of which operates on its own time scale (see figure 1.1). The brain integrates signals sent by various hearing agents into a coherent auditory picture. Ear-brain mechanisms process high and low frequencies differently. Keeping high frequencies constant, while inducing phase shifts in lower frequencies, causes listeners to hear a different timbre.

Determining the temporal limits of perception has long engaged psychoacousticians (Doughty and Garner 1947; Buser and Imbert 1992; Meyer-Eppler 1959; Winckel 1967; Whitfield 1978). The pioneer of sound quanta, Dennis Gabor, suggested that at least two mechanisms are at work in microevent detection: one that isolates events, and another that ascertains their pitch. Human beings need time to process audio signals. Our hearing mechanisms impose minimum time thresholds in order to establish a firm sense of the identity and properties of a microevent.

In their important book *Audition* (1992), Buser and Imbert summarize a large number of experiments with transitory audio phenomena. The general result from these experiments is that below 200 ms, many aspects of auditory perception change character and different modes of hearing come into play. The next sections discuss microtemporal perception.

Microtemporal Intensity Perception

In the zone of low amplitude, short sounds must be greater in intensity than longer sounds to be perceptible. This increase is about +20 dB for tone pips of 1 ms over those of 100 ms duration. (A tone pip is a sinusoidal burst with a quasi-rectangular envelope.) In general, subjective loudness diminishes with shrinking durations below 200 ms.

Microtemporal Fusion and Fission

In dense portions of the Milky Way, stellar images appear to overlap, giving the effect of a near-continuous sheet of light ... The effect is a grand illusion. In reality ... the nightime sky is remarkably empty. Of the volume of space only 1 part in 10^{21} [one part in a quintillion] *is filled with stars.* (Kaler 1997)

Circuitry can measure time and recognize pulse patterns at tempi in the range of a gigahertz. Human hearing is more limited. If one impulse follows less than 200 ms after another, the onset of the first impulse will tend to mask the second, a time-lag phenomenon known as *forward masking*, which contributes to the illusion that we call a *continuous tone*.

The sensation of tone happens when human perception reaches attentional limits where microevents occur too quickly in succession to be heard as discrete events. The auditory system, which is nonlinear, reorganizes these events into a group. For example, a series of impulsions at about 20 Hz fuse into a continuous tone. When a fast sequence of pitched tones merges into a continuous "ripple," the auditory system is unable to successfully track its rhythm. Instead, it simplifies the situation by interpreting the sound as a continuous texture. The opposite effect, *tone fission*, occurs when the fundamental frequency of a tone descends into the infrasonic frequencies.

The theory of *auditory streams* (McAdams and Bregman 1979) aims to explain the perception of melodic lines. An example of a streaming law is: the faster a melodic sequence plays, the smaller the pitch interval needed to split it into two separately perceived "streams." One can observe a family of streaming effects between two alternating tones A and B. These effects range from *coherence* (the tones A and B form a single percept), to *roll* (A dominates B), to *masking* (B is no longer perceived).

The theory of auditory streaming was an attempt to create a psychoacoustic basis for contrapuntal music. A fundamental assumption of this research was that "several musical dimensions, such as timbre, attack and decay transients, and tempo are often not specified exactly by the composer and are controlled by the performer" (McAdams and Bregman 1979). In the domain of electronic music, such assumptions may not be valid.

Microtemporal Silence Perception

The ear is quite sensitive to intermittencies within pure sine waves, especially in the middle range of frequencies. A 20 ms fluctuation in a 600 Hz sine wave, consisting of a 6.5 ms fade out, a 7 ms silent interval, and a 6.5 ms fade in, breaks the tone in two, like a double articulation. A 4 ms interruption, consisting of a 1 ms fade out, a 2 ms silent interval, and a 1 ms fade in, sounds like a transient pop has been superimposed on the sine wave.

Intermittencies are not as noticeable in complex tones. A 4 ms interruption is not perceptible in pink noise, although a 20 ms interruption is.

In intermediate tones, between a sine and noise, microtemporal gaps less than 10 ms sound like momentary fluctuations in amplitude or less noticeable transient pops.

Microtemporal Pitch Perception

Studies by Meyer-Eppler show that pitch recognition time is dependent on frequency, with the greatest pitch sensitivity in the mid-frequency range between 1000 and 2000 Hz, as the following table (cited in Butler 1992) indicates.

Frequency in Hz	100	500	1000	5000
Minimum duration in ms	45	26	14	18

Doughty and Garner (1947) divided the mechanism of pitch perception into two regions. Above about 1 kHz, they estimated, a tone must last at least 10 ms to be heard as pitched. Below 1 kHz, at least two to three cycles of the tone are needed.

Microtemporal Auditory Acuity

We feel impelled to ascribe a temporal arrangement to our experiences. If β is later than α and γ is later than β, then γ is also later than α. At first sight it appears obvious to assume that a temporal arrangement of events exists which agrees with the temporal arrangement of experiences. This was done unconsciously until skeptical doubts made themselves felt. For example, the order of experiences in time obtained by acoustical means can differ from the temporal order gained visually . . . (Einstein 1952)

Green (1971) suggested that temporal auditory acuity (the ability of the ear to detect discrete events and to discern their order) extends down to durations as short as 1 ms. Listeners hear microevents that are less than about 2 ms in duration as a click, but we can still change the waveform and frequency of these events to vary the timbre of the click. Even shorter events (in the range of microseconds) can be distinguished on the basis of amplitude, timbre, and spatial position.

Microtemporal Preattentive Perception

When a person glimpses the face of a famous actor, sniffs a favorite food, or hears the voice of a friend, recognition is instant. Within a fraction of a second after the eyes, nose, ears, tongue or skin is stimulated, one knows the object is familiar and whether it is desirable or dangerous. How does such recognition, which psychologists call preattentive perception, *happen so accurately and quickly, even when the stimuli are complex and the context in which they arise varies?* (Freeman 1991)

One of the most important measurements in engineering is the response of a system to a unit impulse. It should not be surprising to learn that auditory neuroscientists have sought a similar type of measurement for the auditory system. The impulse response equivalents in the auditory system are the *auditory evoked potentials*, which follow stimulation by tone pips and clicks.

The first response in the auditory nerve occurs about 1.5 ms after the initial stimulus of a click, which falls within the realm of *preattentive perception* (Freeman 1995). The mechanisms of preattentive perception perform a rapid analysis by an array of neurons, combining this with past experience into a wave packet in its physical form, or a percept in its behavioral form. The neural activities sustaining preattentive perception take place in the cerebral cortex. Sensory stimuli are preanalyzed in both the pulse and wave modes in intermediate stations of the brain. As Freeman noted, in the visual system complex operations such as adaptation, range compression, contrast enhancement, and motion detection take place in the retina and lower brain. Sensory stimuli activate *feature extractor* neurons that recognize specific characteristics. Comparable operations have been described for the auditory cortex: the final responses to a click occur some 300 ms later, in the medial geniculate body of the thalamus in the brain (Buser and Imbert 1992).

Microtemporal Subliminal Perception

Finally, we should mention *subliminal perception*, or perception without awareness. Psychological studies have tested the influence of brief auditory stimuli on various cognitive tasks. In most studies these take the form of verbal hints to some task asked of the listener. Some evidence of influence has been shown, but the results are not clear-cut. Part of the problem is theoretical: how does subliminal perception work? According to a cognitive theory of Reder and Gordon (1997), for a concept to be in conscious awareness, its activation must be above a certain threshold. Magnitude of activation is partly a function of the exposure duration of the stimulus. A subliminal microevent raises the activation of the corresponding element, but not enough to reach the threshold. The brain's "production rules" cannot fire without the elements passing threshold, but a subliminal microevent can raise the current activation level of an element enough to make it easier to fire a production rule later.

The musical implications are, potentially, significant. If the subliminal hints are not fragments of words but rather musical cues (to pitch, timbre, spatial position, or intensity) then we can embed such events at pivotal instants, knowing that they will contribute to a percept without the listener necessarily being aware of their presence. Indeed this is one of the most interesting dimensions of microsound, the way that subliminal or barely perceptible variations in the properties of a collection of microevents—their onset time, duration, frequency, waveform, envelope, spatial position, and amplitude—lead to different aesthetic perceptions.

Viewing and Manipulating the Microtime Level

Microevents touch the extreme time limits of human perception and performance. In order to examine and manipulate these events fluidly, we need digital audio "microscopes"—software and hardware that can magnify the micro time scale so that we can operate on it.

For the serious researcher, the most precise strategy for accessing the micro time scale is through computer programming. Beginning in 1974, my research was made possible by access to computers equipped with compiler software and audio converters. Until recently, writing one's own programs was the only possible approach to microsound synthesis and transformation.

Many musicians want to be able to manipulate this domain without the total immersion experience that is the lifestyle of software engineering. Fortunately, the importance of the micro time scale is beginning to be recognized. Any sound editor with a zoom function that proceeds down to the sample level can view and manipulate sound microstructure (figure 1.4).

Programs such as our Cloud Generator (Roads and Alexander 1995), offer high-level controls in the micro time domain (see appendix A). Cloud Generator's interface directly manipulates the process of particle emission, controlling the flow of many particles in an evolving cloud. Our more recent PulsarGenerator, described in chapter 4, is another example of a synthetic particle generator.

The perceived result of particle synthesis emerges out of the interaction of parameter evolutions on a micro scale. It takes a certain amount of training to learn how operations in the micro domain translate to acoustic perceptions on higher levels. The grain duration parameter in granular synthesis, for example, has a strong effect on the perceived spectrum of the texture.

This situation is no different from other well-known synthesis techniques. *Frequency modulation synthesis*, for example, is controlled by parameters such as carrier-to-modulator ratios and modulation indexes, neither of which are direct terms of the desired spectrum. Similarly, *physical modeling synthesis* is controlled by manipulating the parameters that describe the parts of a virtual instrument (size, shape, material, coupling, applied force, etc.), and not the sound.

One can imagine a musical interface in which a musician specifies the desired sonic result in a musically descriptive language which would then be translated

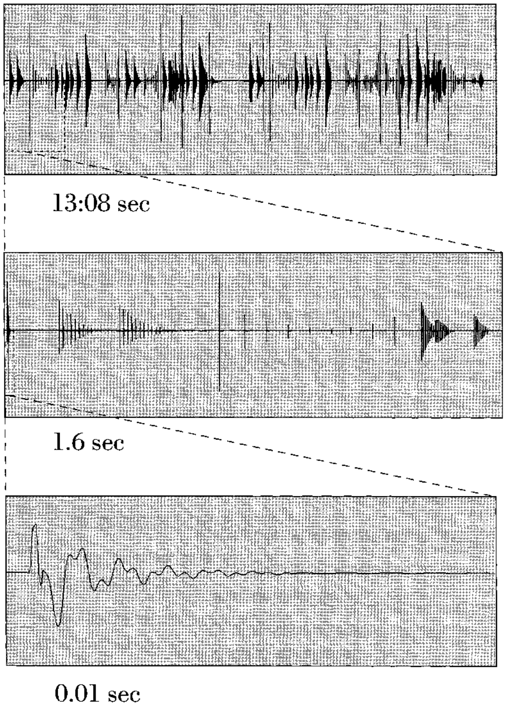

Figure 1.4 Viewing the micro time scale via zooming. The top picture is the waveform of a sonic gesture constructed from sound particles. It lasts 13.05 seconds. The middle image is a result of zooming in to a part of the top waveform (indicated by the dotted lines) lasting 1.5 seconds. The bottom image is a microtemporal portrait of a 10 millisecond fragment at the beginning of the top waveform (indicated by the dotted lines).

into particle parameters and rendered into sound. An alternative would be to specify an example: "Make me a sound like this (soundfile), but with less vibrato." This is a challenging task of *parameter estimation*, since the system would have to interpret how to approximate a desired result. For more on the problems of parameter estimation in synthesis see Roads (1996).

Do the Particles Really Exist?

In the 1940s, the physicist Dennis Gabor made the assertion that all sound—even continuous tones—can be considered as a succession of elementary particles of acoustic energy. (Chapter 2 summarizes this theory.) The question then arises: do sound particles really exist, or are they merely a theoretical construction? In certain sounds, such as the taps of a slow drum roll, the individual particles are directly perceivable. In other sounds, we can prove the existence of a granular layer through logical argument.

Consider the whole number 5. This quantity may be seen as a sum of subquantities, for example $1+1+1+1+1$, or $2+3$, or $4+1$, and so on. If we take away one of the subquantities, the sum no longer is 5. Similarly, a continuous tone may be considered as a sum of subquantities—as a sequence of overlapping grains. The grains may be of arbitrary sizes. If we remove any grain, the signal is no longer the same. So clearly the grains exist, and we need all of them in order to constitute a complex signal. This argument can be extended to explain the decomposition of a sound into any one of an infinite collection of orthogonal functions, such as wavelets with different basis functions, Walsh functions, Gabor grains, and so on.

This logic, though, becomes tenuous if it is used to posit the preexistence (in an ideal Platonic realm) of all possible decompositions within a whole. For example, do the slices of a cake preexist, waiting to be articulated? The philosophy of mathematics is littered with such questions (Castonguay 1972, 1973). Fortunately it is not our task here to try to assay their significance.

Heterogeneity in Sound Particles

The concept of heterogeneity or diversity of sound materials, which we have already discussed in the context of the sound object time scale, also applies to other time scales. Many techniques that we use to generate sound particles assign to each particle a unique identity, a precise frequency, waveform, duration, amplitude morphology, and spatial position, which then distinguishes it from every other particle. Just as certain sound objects may function as singularities, so may certain sound particles.

Sampled Time Scale

Below the level of microtime stands the sampled time scale (figure 1.5). The electronic clock that drives the sampling process establishes a time grid. The spacing of this grid determines the temporal precision of the digital audio medium. The samples follow one another at a fixed time interval of $1/f_S$, where f_S is the sampling frequency. When $f_S = 44.1$ kHz (the compact disc rate), the samples follow one another every 22.675 millionths of a second (μsec).

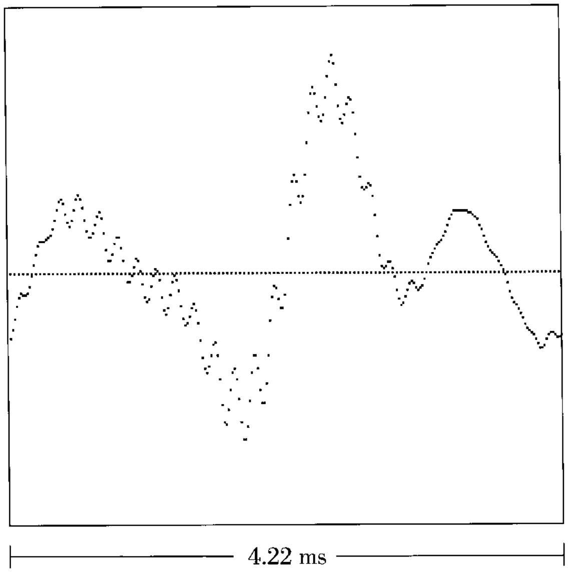

Figure 1.5 Sample points in a digital waveform. Here are 191 points spanning a 4.22 ms time interval. The sampling rate is 44.1 kHz.

The atom of the sample time scale is the *unit impulse*, the discrete-time counterpart of the continuous-time Dirac delta function. All samples should be considered as time-and-amplitude-transposed (delayed and scaled) instances of the unit impulse.

The interval of one sample period borders near the edge of human audio perception. With a good audio system one can detect the presence of an individual high-amplitude sample inserted into a silent stream of zero-valued samples. Like a single pixel on a computer screen, an individual sample offers little. Its amplitude and spatial position can be discerned, but it transmits no sense of timbre and pitch. Only when chained into sequences of hundreds do samples float up to the threshold of timbral significance. And still longer sequences of thousands of samples are required to represent pitched tones.

Sound Composition with Individual Sample Points

Users of digital audio systems rarely attempt to deal with individual sample points, which, indeed, only a few programs for sound composition manipulate directly. Two of these are G. M. Koenig's Sound Synthesis Program (SSP) and

Herbert Brün's Sawdust program, both developed in the late 1970s. Koenig and Brün emerged from the Cologne school of serial composition, in which the interplay between macro- and microtime was a central aesthetic theme (Stockhausen 1957; Koenig 1959; Maconie 1989). Brün wrote:

For some time now it has become possible to use a combination of analog and digital computers and converters for the analysis and synthesis of sound. As such a system will store or transmit information at the rate of 40,000 samples per second, even the most complex waveforms in the audio-frequency range can be scanned and registered or be recorded on audio tape. This . . . allows, at last, the composition of timbre, instead of with timbre. In a sense, one may call it a continuation of much which has been done in the electronic music studio, only on a different scale. The composer has the possibility of extending his compositional control down to elements of sound lasting only 1/20,000 of a second. (Brun 1970)

Koenig's and Brün's synthesis programs were conceptually similar. Both represented a pure and radical approach to sound composition. Users of these programs stipulated sets of individual time and amplitude points, where each set was in a separate file. They then specified logical operations such as *linking*, *mingling*, and *merging*, to map from a time-point set to an amplitude-point set in order to construct a skeleton of a waveform fragment. Since these points were relatively sparse compared to the number of samples needed to make a continuous sound, the software performed a linear interpolation to connect intermediate amplitude values between the stipulated points. This interpolation, as it were, fleshed out the skeleton. The composer could then manipulate the waveform fragments using logical set theory operations to construct larger and larger waveforms, in a process of hierarchical construction.

Koenig was explicit about his desire to escape from the traditional computer-generated sounds:

My intention was to go away from the classical instrumental definitions of sound in terms of loudness, pitch, and duration and so on, because then you could refer to musical elements which are not necessarily the elements of the language of today. To explore a new field of sound possibilities I thought it best to close the classical descriptions of sound and open up an experimental field in which you would really have to start again. (Roads 1978b)

Iannis Xenakis proposed a related approach (Xenakis 1992; Hoffmann 1994, 1996, 1997). This involves the application of *sieve theory* to the amplitude and time dimensions of a sound synthesis process. As in his Gendyn program, the idea is to construct waveforms from fragments. Each fragment is bounded by two breakpoints. Between the breakpoints, the rest of the waveform is filled in by interpolation. Whereas in Gendyn the breakpoints are calculated from a nonlinear stochastic algorithm, in sieve theory the breakpoints would be calculated according to a partitioning algorithm based on sieved amplitude and time dimensions.

Assessment of Sound Composition with Samples

To compose music by means of logical operations on samples is a daunting task. Individual samples are subsymbolic—perceptually indistinguishable from one another. It is intrinsically difficult to string together samples into meaningful music symbols. Operations borrowed from set theory and formal logic do not take into account the samples' acoustical significance. As Koenig's statement above makes clear, to compose intentionally a graceful melodic figure, a smooth transition, a cloud of particles, or a polyphonic texture requires extraordinary effort, due to the absence of acoustically relevant parameters for building higher-level sound structures. Users of sample-based synthesis programs must be willing to submit to the synthesis algorithm, to abandon local control, and be satisfied with the knowledge that the sound was composed according to a logical process. Only a few composers took up interest in this approach, and there has not been a great deal of experimentation along these lines since the 1970s.

Subsample Time Scale

A digital audio system represents waveforms as a stream of individual samples that follow one another at a fixed time interval ($1/f_S$, where f_S is the sampling frequency). The subsample time scale supports fluctuations that occur in less than two sampling periods. Hence this time scale spans a range of minuscule durations measured in nanoseconds and extending down to the realm of infinitesimal intervals.

To stipulate a sampling frequency is to fix a strict threshold between a subsample and the sample time scale. Frequencies above this threshold—the *Nyquist frequency* (by definition: $f_S/2$)—cannot be represented properly by a digital audio system. For the standard compact disc sampling rate of 44.1 kHz, the Nyquist frequency is 22.05 kHz. This means that any wave fluctuation shorter than two samples, or 45 μsec, is relegated to the subsample domain. The 96 kHz sampling rate standard reduces this interval to 20.8 μsec.

The subsample time scale encompasses an enormous range of phenomena. Here we present five classes of subsample phenomena, from the real and perceptible to the ideal and imperceptible: aliased artefacts, ultrasounds, atomic sounds, and the Planck interval.

Aliased Artefacts

In comparison with the class of all time intervals, the class of perceptible audio periods spans relatively large time intervals. In a digital audio system, the sample period is a threshold separating all signal fluctuations into two classes: those whose frequencies are low enough to be accurately recorded and those whose frequencies are too high to be accurately recorded. Because a frequency is too high to be recorded does not mean that it is invisible to the digital recorder. On the contrary, subsample fluctuations, according to the theorem of Nyquist (1928), record as aliased artefacts. Specifically, if the input frequency is higher than half the sampling frequency, then:

aliased frequency = sampling frequency − input frequency

Thus if the sampling rate is 44.1 kHz, an input frequency of 30 kHz is reflected down to the audible 11.1 kHz. Digital recorders must, therefore, attempt to filter out all subsample fluctuations in order to eliminate the distortion caused by aliased artefacts.

The design of antialiasing filters has improved in the past decade. Current compact disc recordings are effectively immune from aliasing distortion. But the removal of all information above 22.05 kHz poses problems. Many people hear detail (referred to as *air*) in the region above 20 kHz (Koenig 1899; Neve 1992). Rigorous scientific experiments have confirmed the effects, from both physiological and subjective viewpoints, of sounds above 22 kHz (Oohashi et al. 1991; Oohashi et al. 1993). Furthermore, partials in the ultrasonic region interact, resulting in audible subharmonics and air. When the antialiasing filter removes these ultrasonic interactions, the recording loses detail.

Aliasing remains a pernicious problem in sound synthesis. The lack of *frequency headroom* in the compact disc standard rate of 44.1 kHz opens the door to aliasing from within the synthesis algorithm. Even common waveforms cause aliasing when extended beyond a narrow frequency range. Consider these cases of aliasing in synthesis:

1. A band-limited square wave made from sixteen odd-harmonic components causes aliasing at fundamental frequencies greater than 760 Hz.

2. An additive synthesis instrument with thirty-two harmonic partials generates aliased components if the fundamental is higher than 689 Hz (approximately E5).
3. The partials of a sampled piano tone A-sharp2 (116 Hz) alias when the tone is transposed an octave and a fifth to F4 (349 Hz).
4. A sinusoidal frequency modulation instrument with a carrier-to-modulator ratio of 1:2 and a fundamental frequency of 1000 Hz aliases if the modulation index exceeds 7. If either the carrier or modulator is a non-sinusoidal waveform then the modulation index must typically remain less than 1.

As a consequence of these hard limits, synthesis instruments require preventative measures in order to eliminate aliasing distortion. Commercial instruments filter their waveforms and limit their fundamental frequency range. In experimental software instruments, we must introduce tests and constrain the choice of waveforms above certain frequencies.

The compact disc sampling rate of 44.1 kHz rate is too low for high-fidelity music synthesis applications. Fortunately, converters operating at 96 kHz are becoming popular, and sampling rates up to 192 kHz also are available.

Ultrasonic Loudspeakers

Even inaudible energy in the ultrasonic frequency range can be harnessed for audio use. New loudspeakers have been developed on the basis of *acoustical heterodyning* (American Technology Corporation 1998; Pompei 1998). This principle is based on a phenomenon observed by Helmholtz. When two sound sources are positioned relatively closely together and are of a sufficiently high amplitude, two new tones appear: one lower and one higher than either of the original tones. The two new combination tones correspond to the sum and the difference of the two original tones. For example, if one were to emit 90 kHz and 91 kHz into the air, with sufficient energy, one would produce the sum (181 kHz) and the difference (1 kHz), the latter being in the range of human hearing. Reporting that he could also hear summation tones (whose frequency is the sum, rather than the difference, of the two fundamental tones), Helmholtz argued that the phenomenon had to result from a nonlinearity of air molecules. Air molecules begin to behave nonlinearly (to *heterodyne*) as amplitude increases. Thus, a form of acoustical heterodyning is realized by creating difference frequencies from higher frequency waves. In air, the effect works in such a way that if an ultrasonic carrier is increased in amplitude, a difference frequency is created. Concurrently, the unused sum frequency diminishes in loudness as the carrier's frequency increases. In other words, the major portion of the ultrasonic energy transfers to the audible difference frequency.

Unlike regular loudspeakers, acoustical heterodyning loudspeakers project energy in a collimated sound beam, analogous to the beam of light from a flashlight. One can direct an ultrasonic emitter toward a wall and the listener will perceive the sound as coming from a spot on that wall. For a direct sound beam, a listener standing anywhere in an acoustical environment is able to point to the loudspeaker as the source.

Atomic Sound: Phonons and Polarons

As early as 1907, Albert Einstein predicted that ultrasonic vibration could occur on the scale of atomic structure (Cochran 1973). The atoms in crystals, he theorized, take the form of a regular lattice. A one-dimensional lattice resembles the physical model of a taut string—a collection of masses linked by springs. Such a model may be generalized to other structures, for example three-dimensional lattices. Lattices can be induced to vibrate ultrasonically, subjected to the proper force, turning them into high-frequency oscillators. This energy is not continuous, however, but is quantized by atomic structure into units that Einstein called phonons, by analogy to photons—the quantum units of light. It was not until 1913 that regular lattices were verified experimentally as being the atomic structure of crystals. Scientists determined that the frequency of vibration depends on the mass of the atoms and the nature of the interatomic forces. Thus the lower the atomic weight, the higher the frequency of the oscillator (Stevenson and Moore 1967). Ultrasonic devices can generate frequencies in the trillions of cycles per second.

Complex sound phenomena occur when phononic energy collides with other phonons or other atomic particles. When the sources of excitation are multiple or the atomic structure irregular, phonons propagate in cloud-like swarms called *polarons* (Pines 1963). Optical energy sources can induce or interfere with mechanical vibrations. Thus optical photons can scatter acoustic phonons. For example, laser-induced lattice vibrations can change the index of refraction in a crystal, which changes its electromagnetic properties. On a microscopic scale, optical, mechanical, and electromagnetic quanta are interlinked as elementary excitations.

Laser-induced phonic sound focuses the beams from two lasers with a small wavelength difference onto a crystal surface. The difference in wavelength causes interference, or beating. The crystal surface shrinks and expands as this oscillation of intensity causes periodic heating. This generates a wave that propagates through the medium. The frequency of this sound is typically in the gigahertz range, with a wavelength of the order of 1 micron. Because of the small dimensions of the heated spot on the surface, the wave in the crystal has the shape of a directional beam. These sound beams can be used as probes, for example, to determine the internal features of semiconductor crystals, and to detect faults in their structure.

One of the most important properties of laser-induced phononic sound is that it can be made *coherent* (the wave trains are phase-aligned), as well as monochromatic and directional. This makes possible such applications as *acoustic holography* (the visualization of acoustic phenomena by laser light). Today the study of phononic vibrations is an active field, finding applications in *surface acoustic wave* (SAW) filters, waveguides, and condensed matter physics.

At the Physical Limits: The Planck Time Interval

Sound objects can be subdivided into grains, and grains into samples. How far can this subdivision of time continue? Hawking and Penrose (1996) have suggested that time in the physical universe is not infinitely divisible. Specifically, that no signal fluctuation can be faster than the quantum changes of state in subatomic particles, which occur at close to the *Planck scale*. The Planck scale stands at the extreme limit of the known physical world, where current concepts of space, time, and matter break down, where the four forces unify. It is the exceedingly small distance, related to an infinitesimal time span and extremely high energy, that emerges when the fundamental constants for gravitational attraction, the velocity of light, and quantum mechanics join (Hawking and Penrose 1996).

How much time does it take light to cross the Planck scale? Light takes about 3.3 nanoseconds (3.3×10^{-10}) to traverse 1 meter. The *Planck time interval* is the time it takes light to traverse the Planck scale. Up until recently, the Planck scale was thought to be 10^{-33} meter. An important new theory puts the figure at a much larger 10^{-19} meter (Arkani-Hamed et al. 2000). Here, the Planck time interval is 3.3×10^{-28} seconds, a tiny time interval. One could call the Plank time interval a kind of "sampling rate of the universe," since no signal fluctuation can occur in less than the Planck interval.

If the flow of time stutters in discrete quanta corresponding to fundamental physical constants, this poses an interesting conundrum, recognized by Iannis Xenakis:

Isn't time simply an epiphenomenal notion of a deeper reality?... The equations of Lorentz-Fitzgerald and Einstein link space and time because of the limited velocity of light. From this it follows that time is not absolute ... It "takes time" to go from one point to another, even if that time depends on moving frames of reference relative to the observer. There is no instantaneous jump from one point to another in space, much less spatial ubiquity—that is, the simultaneous presence of an event or object everywhere in space. To the contrary, one posits the notion of displacement. Within a local reference frame, what does displacement signify? If the notion of displacement were more fundamental than that of time, one could reduce all macro and micro cosmic transformations to weak chains of displacement. Consequently ... if we were to adhere to quantum mechanics and its implications, we would perhaps be forced to admit the notion of quantified space and its corollary, quantified time. But what could a quantified time and space signify, a time and space in which contiguity would be abolished. What would the pavement of the universe be if there were gaps between the paving stones, inaccessible and filled with nothing? (Xenakis 1989)

Infinitesimal Time Scale

Besides the infinite-duration sinusoids of Fourier theory, mathematics has created other ideal, infinite-precision boundary quantities. One class of ideal phenomena that appears in the theory of signal processing is the mathematical impulse or delta (∂) function. Delta functions represent infinitely brief intervals of time. The most important is the *Dirac delta function*, formulated for the theory of quantum mechanics. Imagine the time signal shown in figure 1.6a, a narrow pulse of height $1/b$ and width b, centered on $t = 0$. This pulse, $x(t)$, is zero at all times $|t| > b/2$. For any nonzero value of b, the integral of $x(t)$ is unity. Imagine that b shrinks to a duration of 0. Physically this means that the pulse's height grows and the interval of integration (the pulse's duration) becomes very narrow. The limit of $x(t)$ as $b \to 0$ is shown in figure 1.6b. This shows that the pulse becomes an infinitely high spike of zero width, indicated as $\partial(t)$, the Dirac delta function. The two significant properties of the ∂ function are: (1) it is zero everywhere except at one point, and (2) it is infinite in amplitude at this point, but approaches infinity in such a way that its integral is unity—a curious object!

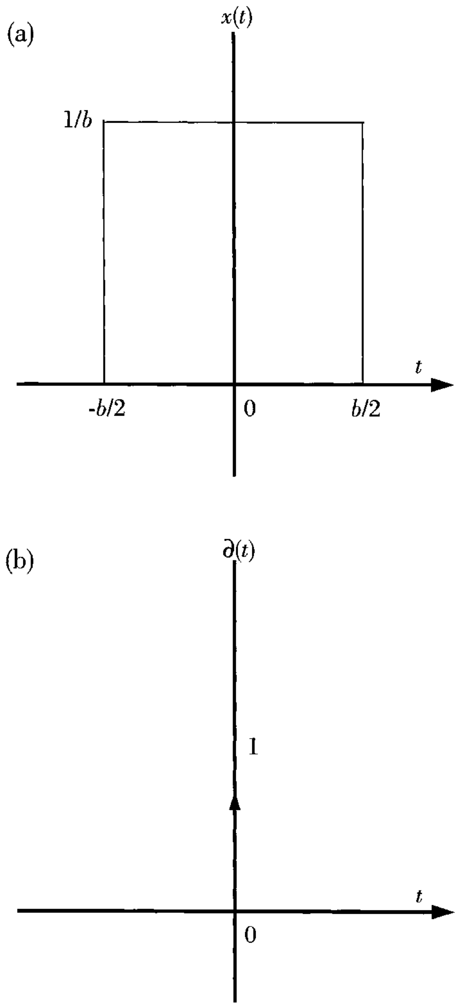

Figure 1.6 Comparison of a pulse and the Dirac delta function. (a) A narrow pulse of height $1/b$ and width b, centered on $t = 0$. (b) The Dirac delta function.

The main application of the ∂ function in signal processing is to bolster the mathematical explanation of the process of sampling. When a ∂ function occurs inside an integral, the value of the integral is determined by finding the location of the impulse and then evaluating the integrand at that location. Since the ∂ is infinitely brief, this is equivalent to sampling the function being integrated. Another interesting property of the ∂ function is that its Fourier transform,

$$|e^{-j2\pi ft}| = 1$$

for any real value of t. In other words, the spectrum of an infinitely brief impulse is infinite (Nahin 1996).

We see here a profound law of signal processing, which we will encounter repeatedly in this thesis, that duration and spectrum are complementary quantities. In particular, the shorter a signal is, the broader is its spectrum. Later we will see that one can characterize various signal transformations by how they respond to the ∂ function and its discrete counterpart, the unit impulse.

The older *Kronecker delta* is an integer-valued ideal impulse function. It is defined by the properties

$$\partial_{m,n} = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$$

The delta functions are defined over a continuous and infinite domain. The section on *aliased artefacts* examines similar functions in the discrete sampled domain.

Outside Time Music

Musical structure can exist, in a sense, "outside" of time (Xenakis 1971, 1992). By this, we mean abstract structuring principles whose definition does not imply a temporal order. A scale, for example, is independent of how a composer uses it in time. Myriad *precompositional* strategies, and databases of material could also be said to be outside time.

A further example of an outside time structure is a musical instrument. The layout of keys on a piano gives no hint of the order in which they will be played. *Aleatoric* compositions of the 1950s and 1960s, which left various parameters, including the sequence of events to chance, were also outside time structures.

Today we see installations and virtual environments in which sounds occur in an order that depends on the path of the person interacting with the system. In all of these cases, selecting and ordering the material places it in time.

The Size of Sounds

Sounds form in the physical medium of air—a gaseous form of matter. Thus, sound waves need space to form. Just as sounds exist on different time scales, so they take shape on different scales of space. Every sound has a three-dimensional shape and size, which is its diffusion or dispersion pattern over time. Since the wavelength of a high frequency sound is short, high frequencies form in small spaces. A low frequency waveform needs several meters to unfold. The temporal and the spatial morphologies of a sound intertwine. A sound's duration, frequency, amplitude, and pattern of radiation from its source all contribute to its physical form, as does the space in which the sound manifests.

The duration of a sound is an important determinant of physical shape, especially in the open air. A long-duration sound is long in spatial extent, spanning the entire distance from the source to the point at which its energy is completely absorbed. Short-duration sounds, on the contrary, are thin in spatial extent, disappearing from their point of origin quickly. The wave of a short-duration sound occupies a thin band of air, although the fluctuations that it carries may travel great distances if it is loud enough.

Today we have accurate measurements of the speed of sound waves in a variety of media (Pierce 1994). The accepted value for the speed of sound in dry air is 331.5 meters/second. Thus a 20 Hz acoustical wave requires no less than 16.5 meters (54.13 feet) to unfold without obstruction. Obstructions such as walls cause the wave to reflect back on itself, creating phase cancellation effects. A high-frequency waveform at 20 kHz has a period of only 1/20,000th of a second. This takes only 1.65 cm to form. The ear is very sensitive to the time of arrival of sounds from different spatial positions. Thus, even a minor difference in the distance of the listener from two separate sources will skew the spatial images.

The most important determinant of a sound's size is its amplitude. Very loud sounds (such as atmospheric thunder and other explosions) travel far. As they travel, the air gradually absorbs the high frequencies, so that only the low frequencies reach great distances.

Summary

> *Particle physics seeks to find a simple and orderly pattern to the behavior of matter on the atomic and subatomic level. To this end, large particle accelerators are built, acting like giant microscopes that zoom down through the atom ... Astronomers build equally complex devices—telescopes and observatories. These gather data from distant clusters of galaxies, all the way out to the rim of the cosmos ... We are seeing here a convergence between particle physics and cosmology. The instruments, and even the stated objectives, are different, but the languages draw closer. The laws of nature that control and order the microscopic world, and those that determined the creation and evolution of the universe, ... are beginning to look identical.* (Lederman and Schramm 1995)

Projecting time horizontally, and amplitude vertically, the concept of nil duration corresponds to a zero-dimensional point on the time-amplitude plane. This point zero is mute: no flux of energy can occur in the absence of a time window. In that ideal world experienced only by the gods of mathematics, the delta function $\partial(t)$ breaks the monotony with an instantaneous impulse that is born and dies within the most infinitesimal window beyond point zero.

Our mundane digital domain is a discrete approximation to the ideal realm of infinitesimal time. In the digital domain, the smallest event has a duration equivalent to the period of the sampling frequency. This sound atom, the sample period, is the grid that quantizes all time values in an audio signal. Any curve inscribed on the amplitude-versus-time plane must synchronize to this grid. Individual samples remain subsymbolic. Like the woven threads of canvas holding paint in place, their presence is a necessity, even if we can see them only in the aggregate.

As the window of time expands, there is a possibility for chaotic fluctuation, periodic repetition, echoes, tone, noise, and measured silence. Each additional instant of time accrues new possibilities.

Microsonic particles can be likened to molecules built from atomic samples. To view this level of detail, we rely on the tools of sound analysis and display. Under this scrutiny, remarkable patterns emerge and we gain new insight into sound structure. These images show the hidden morphologies of elementary sound molecules (figure 1.7).

Molecular materials alter the terrain of composition. Pliant globules can be molded into arbitrary object morphologies. The presence of mutating sound objects suggests a fluid approach to compositional mesostructure, spawning rivulets, streams, and clouds as well as discrete events. The package for all these

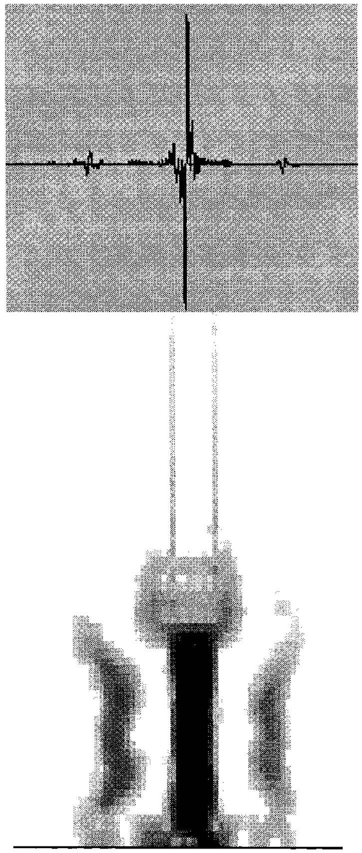

Figure 1.7 Image of a grain in the time-domain (top) and its frequency-domain counterpart (bottom).

musical structures, the macroform, can be tailored with high flexibility and precision in a sound mixing program.

It is necessary to see music over a broad range of time scales, from the infinitesimal to the supra scale (Christensen 1996). Not all musicians are prepared to view musical time from such a comprehensive perspective, however, and it may well take decades for this perspective to filter into our general musical vocabulary.