Barry Truax
Department of Communication
Simon Fraser University
Burnaby, British Columbia V5A 1S6 Canada

Discovering Inner Complexity: Time Shifting and Transposition with a Real-time Granulation Technique

This article describes a digital signal-processing technique using short grains of sampled sound—generally less than 50 msec in duration—that extends the source sound's duration without altering its pitch. In addition, individual voices of the stretched sound may be transposed to a variety of harmonic frequencies. The psychoacoustic implications of the technique, such as the magnification of instantaneous resonances and the perception of increased volume, are discussed, as is compositional experience that links the inner complexity of the sound to the complexity of the external world.

Background

Since 1986, I have been working with the technique of granular synthesis (Roads 1978; 1988; 1991), and since 1987, with the granulation of sampled sound in real time (Truax 1988) using the programmable DMX-1000 digital signal processor (Wallraff 1979). Briefly, this technique produces complex sounds by the generation of high densities (100 to 2,000 events/sec) of small "grains" on the order of 10 to 50 msec duration. The content of the grain itself can be a fixed waveform, a simple FM timbre, or a sampled sound, with a hierarchy of control parameters directing the density, frequency range and temporal evolution of the synthesized sound textures. With sampled sound as a source, particularly rich textures may result from extremely small fragments of source material. Since 1989, the granulation technique has been applied to a process of stretching the sound in a manner called variable-rate time shifting. The technique either leaves the original pitch intact or transposes the sound in each grain by a different frequency ratio. The technique is similar to the time-shifting work reported by Jones and Parks (1988) except that the goal is to lengthen the sound, not shorten it. An implementation using the IRCAM Signal Processing Workstation has also been recently reported (Lippe 1993). In addition, the technique is designed to work in real time, unlike computationally-intensive methods such as the phase vocoder (Dolson 1986). Compositional experience using the technique has been particularly rewarding (Truax 1990b; 1992a), and my colleague and I are currently implementing the technique on a microprocessor-controlled circuit board that uses the Motorola DSP56001 digital signal-processing chip and MC68000 controller (Truax and Bartoo 1992).

Interpolating between Fixed and Continuous Sampling

The real-time program named GSAMX (granular sampling with the DMX-1000) implements a sampled-sound instrument for granular synthesis in which each grain consists of a short segment of sampled sound with specifiable duration and offset time from the beginning of the sound sample. The synthesis instrument consists of a bank of simple envelope generators with specifiable duration and delay (in milliseconds) between successive envelopes. Each generator produces a three-part linear envelope

Computer Music Journal, 18:2, pp. 38–48, Summer 1994

[*Editor's note:* Side 2 of the soundsheet that was included in *Computer Music Journal* 18:1 contained a collection of extended musical examples from the author that are intended to accompany this article.]

Figure 1. Addressing for the granulation of a fixed (stored) sample from within the memory of the DMX-1000.

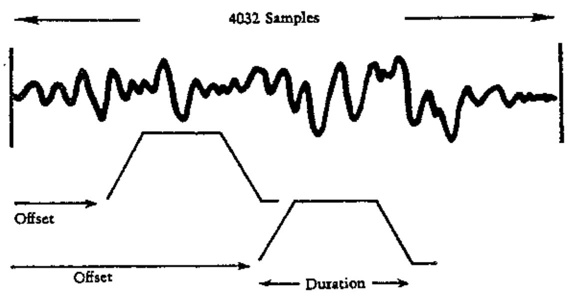
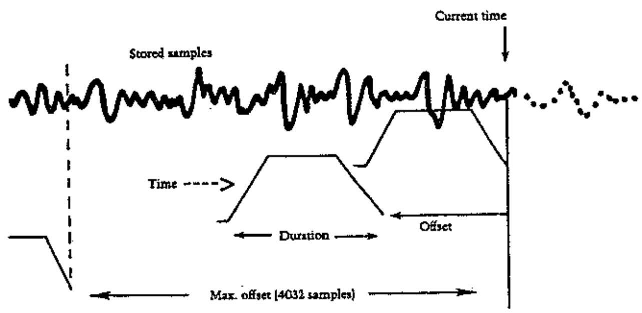


Figure 2. Addressing for the granulation of a continuous stream of input samples.



whose attack and decay portions are a specifiable fraction of the event duration (Truax 1988). Additional variables include the start sample (or offset) and the range over which this variable may be chosen. Up to 20 simultaneous streams or voices of this synthesis instrument are possible with the DMX-1000. The instrument is controlled by a scheduler program running on the Digital Equipment Corporation PDP Micro-11 host computer in which each grain is initiated and terminated under clock interrupts set at a 1-msec rate. The shorter the grain duration, the higher the overall density of grains per second (gps). The minimum grain duration that can be effectively controlled in real time by the DMX is 10 msec, hence densities of up to 2,000 gps can be achieved with the 20 simultaneous voices. At lower densities, however, grain durations may be as short as 2 msec.

Because each grain has an attack and decay, there is no possibility of clicks or transients, depending on the portion of sampled sound being used. Moreover, when the grain streams are unsynchronized because each grain has a different duration or delay time between grains and when each grain starts at a different position within the sound sample, very complex textures can result from even a very simple source sound.

Initially, two contrasting approaches were developed within the program as to the treatment of the sampled sound: fixed-sample (with approximately 4k-words of stored samples) and continuous-sample input from disk at normal speed.

The fixed-sample option, shown in Figure 1, uses a short sequence of source material, up to 4,032 samples or around 150 to 170 msec of sound because of the limitation of 4k-words of on-board memory in the DMX-1000. The duration of grains used in granular synthesis is typically less than 150 msec, so the effect of the fixed sample size is to limit the variety of simultaneous "windows" that may be accessed from the sound material. The continuous sample version, shown in Figure 2, involves real-time granulation of sound directly from disk with the 4k-word memory acting as a short delay line or time window that may be tapped to furnish the various grains.

In the present system, sampled sound is stored in 5k-word blocks on a hard disk, where it can be played back with various signal-processing options. During playback and granulation, samples are looped; the length of the loop can be any number of blocks from 1 to 1024. As an alternative to looping, the user may specify particular segments of the material to be played on a specific keystroke or in a predetermined sequence, the effect being that of real-time editing. The start block number and number of blocks may also be "synchronized" to increment or decrement automatically at the end of each loop. Another option randomizes the start block of each segment within a specified range. Synthesis may occur at various speeds, resulting in sampling rates from 19 to 50 kHz. To avoid transients arising from sample discontinuities at the end of one segment and the beginning of the next, the user can request a short linear fade-out and fade-in, lasting a given number of samples or msec. With most material, 100 samples per fade is inaudible and avoids transients.

In the fixed-sample version discussed earlier, the samples may be stored in a file that can be used independently of the source disk. However, in the continuous version, samples are transferred from the disk to the DMX-1000 via a DMA interface and granulated during performance. All looping and processing is non-destructive of the source material. Additional options for mixing or filtering samples prior to granulation are also available.

Control Variables

The following are the five control variables available to the user that determine how successive grain parameters are calculated:

Average offset from the start and offset range
Average grain duration and duration range
Delay time between grains or grain density
Speed of output (this acts as a pitch/time transposition)
Total number of voices sounding (max 20), including the number of grain streams per stereo channel

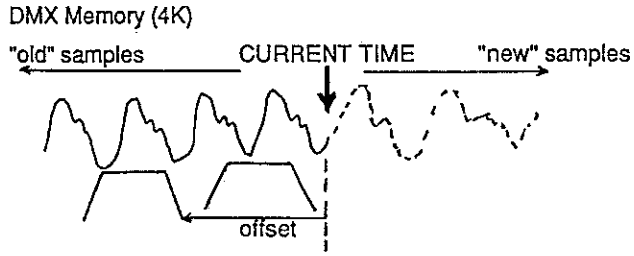
In the fixed-sample case, shown in Figure 1, the offset is the number of samples past the start of the source where the grain begins, whereas in the continuous or variable-rate mode, shown in Figure 2, it refers to how far back into the recent past of the sound sequence the grains are taken (similar to a delay line). Varying the offset from grain to grain by means of the offset range allows each grain to be different and results in a richer aural effect. The second variable, grain duration, is often in the range of 10 to 30 msec with granular synthesis, but with sampled sound it is more common to use 40- to 50-msec grains so that the timbral character of the original material is the least modified by audio rate effects created by shorter grains that extend the sound's bandwidth.

The choice of delay or density in the third variable corresponds to the distinction made by Roads (1991) between quasi-synchronous and asynchronous granulation, respectively. In the former case, the user controls the delay time between grains that is more or less similar, thereby creating periodic modulation effects at high densities. With asynchronous granulation, the user controls the more intuitive variable of grain density in grains per second. For a given density, the program calculates an average delay time based on the average grain duration and number of simultaneous grain streams and then chooses a random value for the delay of each grain between zero and twice the average value. If the user specifies too high a density, the average grain duration must be reduced. This modification of the quasi-synchronous model seems to approximate closely Roads's completely random scattering of grains that he terms asynchronous.

The fourth variable is a simple speed control that causes the signal processor to run more slowly within a fairly small range, similar to a variable-speed tape recorder. The fifth variable allows the sound density to be reduced by switching voices off. With fewer voices of quasi-synchronous granulation, the minimum grain duration can be reduced to 2 msec. Stereo panning, either manually or cyclically activated and moving at various speeds, is achieved by controlling the number of simultaneous grain streams assigned to each output channel. Therefore, grain density, rather than amplitude level per channel, determines the perception of lateral position. The user may also specify a two-dimensional "trajectory" pattern to control spatial movement automatically. The trajectory is realized with grain density for the lateral dimension and overall amplitude level for the dimension of apparent depth.

One main reason for the dynamic quality of granulated sound lies in the possibility of both successive and simultaneous grains to have different parameter values, particularly when the grain streams are unsynchronized and fed to separate stereo output channels. It is often difficult to realize such independence of grains and grain streams with MIDI-controlled samplers, for instance. On the other hand, given the range of sound densities desired, the calculation of each grain's parameters must be very efficient. Deterministic control by predetermined values would be burdensome for both the user and the computer. The simplest solution, implemented in GSAMX, is to adopt a stochastic model in which the user specifies the average or minimum value of the control variable and a range within which individual parameter choices may be randomly made. The calculation of such values is performed by a background programming task, the values being used whenever a new grain is initiated. Experimentation with non-linear chaotic algorithms, such as the logistic and "gingerbread man" maps in which simple equations

Figure 3. Addressing for variable-rate granulation using the DMX-1000.



produce complex patterning, have also proved to be interesting (Truax 1990a). These maps are applied to the offset, duration, and delay parameters and are calculated when the foreground task initiates each grain because the current value depends on previous ones.

In addition to the basic control variables, each model has one additional variable that is specific to it. With the fixed-sample model, the user can control the number of voices sounding at transposed frequencies, and with the continuous-sample model, the amplitude of samples being fed back into the delay line. With the fixed-sample version, the option allows a certain number of voices in the instrument to send out samples at different frequencies by skipping or repeating samples. The duration of the grain is not affected by these different sample rates unless the end of the source material is reached. As a result, part of the sound texture may sound at a pitch above or below the rest of the material. With the continuous-sampling version, this option is replaced by a feedback control to allow the user to recirculate samples through the delay line. The continuous model also allows the memory to be "frozen" at particular moments, similar to the fixed-sample model. The freezing may be triggered by a user keystroke, by a peak amplitude, or at regular intervals, with or without the original signal being included in the output.

Work with these two approaches suggested the need for a method that would interpolate between them—that is, vary the rate at which new samples are introduced into the granulation process (hence the term "variable-rate" granulation). The desired effect is to be able to depart from the normal time flow of the continuous-sampling model in a manner that eventually approaches the "frozen" time of the fixed-sample model. Such interpolation preserves the ongoing development of forward time flow but combines it with the sense of magnification of the moment associated with the fixed version.

Variable-Rate Sampled Sound

In the variable-rate implementation, the key variable is the rate at which new sound samples enter the DMX-1000's memory from disk compared with the synthesized output, as shown in Figure 3. The "rate" of the time-shifted sound is defined as the ratio of "off" milliseconds to "on" milliseconds and is called the off:on ratio. Therefore, a ratio of 0:1 is normal speed because there is no "off" time, and a ratio of 99:1 results in 99 msec of no forward movement through the sample before there is a 1-msec shift forward, thereby producing a 100-fold time extension of the sample. However, since the grains are always taken from the current memory at one sample per calculated output frame, the frequency of the source material is not distorted, only the rate at which the user advances through it in a macro-level sense. This process has the effect that micro-level waveform patterns and macro-level temporal changes have been effectively separated.

If all of the sampled sound were simultaneously available in memory, each grain could begin at the current time position and extend through the subsequent samples (this is the case with the implementation using the Motorola DSP 56001 chip). With the DMX-1000, however, these "future" samples are not present in memory because they have not been delivered by the DMA interface from the disk. This difficulty is surmounted thanks to a psychoacoustic phenomenon that illustrates the quantum nature of the grain. During the "off" milliseconds when the contents of memory are frozen, the grains take their samples in the reverse direction. As long as the direction chosen remains the same throughout a grain that is less than 50 msec with a symmetrical envelope, there is no difference between forward and reverse in terms of the aural result. During the "on" milliseconds, new samples are added to the memory, thereby losing the old ones, and the grains may be taken in either the forward or reverse directions. This choice results in two options for the user. The user may decide that the sample direction will always be in reverse, thereby producing minimal spectral alteration, or that the sample direction will alternate (forward and reverse) during the "on" and "off" states, thereby "modulating" the sample at the micro level.

If the first option is chosen, samples are always read in the same direction, namely reverse, and therefore the result is a pure time-shifting effect with little timbral alteration except that introduced as a result of the granulation process. Because the purpose is to produce musically interesting sound and not just a processed signal, up to 18 simultaneous superpositions of the grains with independent characteristics are normally used to give a sense of magnification to the sound. (At around a 30-kHz sampling rate, the DMX-1000 can produce 12 voices of variable-rate granulation, and the DSP implementation 32 voices, which, with a 2-msec grain, means a peak density of 16,000 gps). However, the study of micro-level pitch changes is of interest to researchers in linguistics and ethnomusicology, and a version with less enhancement may be appropriate in those contexts. For instance, pairs of grain streams may be synchronized such that the end of the steady state of one grain triggers the start of a grain in the other stream. Because the grain envelope is linear, the sum of the attack and decay produces a reasonably continuous output amplitude.

In the second option, the *n:n* ratios, such as 1:1, 2:2, 3:3, etc. (which are not equivalent), produce an interesting phase modulation effect because for equal amounts of time (1, 2, 3 msec, respectively), the grain goes backward through the sound sample, then forward through the exact same material, and so on. The series of ratios mentioned above produces a descending subharmonic series of phase modulation frequencies from 500 Hz. Likewise, ratios of 2:1, 4:2, and 8:4 combine a certain amount of phase modulation with the slowing down effect that larger ratios produce. Instead of a strong pitch component being added, however, these ratios produce a noisier broadband result.

In each case, the amount of time shift, described as the *time-extension factor* (TEF), can be calculated from the off:on ratio as follows:

$$\text{time-extension factor} = (\text{off ratio} + \text{on ratio})/(\text{on ratio}) \tag{1}$$

Therefore, the ratio 1:1 produces a TEF of 2 times normal speed, and a ratio of 999:1 produces a TEF of 1,000 times normal speed. This latter example means that 1 sec of sound can last more than 16 minutes! Because the TEF is proportional to the ratio, there is a strong intuitive relationship between them. One advantage of this approach is that there is no limit, other than one's patience to listen, to the amount of time stretching.

The off:on ratio may be typed in and stored as part of a preset—that is, a set of control variables that can be recalled with a single keystroke. Each component of the ratio may also be separately "synchronized" to allow the ratio or its components to be ramped, thereby continuously changing the TEF.

Automated Rate Control

Two types of automated control of the rate are also available. The first temporarily reverts to real-time when the maximum sample amplitude in a given disk block falls below a user-specified threshold value. This control acts as a kind of filter to skip over quiet parts of the sound stream that otherwise would be time extended along with the rest of the sound, thereby eliminating lengthy pauses. The second automated control correlates the rate to the maximum sample amplitude, thereby slowing down higher amplitude sounds and speeding up lower amplitude sections. The amount of rate variation depends on the maximum rate value that the user selects. This maximum value is implemented during the blocks with peak amplitude, with proportionately smaller values used during blocks with other amplitudes.

Depending on the length of the time window over which amplitude is assessed in order to perform this correlation, attack transients may be smeared by being part of the highest amplitude portion of the sound. Because the effect of longer time stretching is to lose the temporal character of the sound in order to enhance its spectral makeup, this loss of transients may be compositionally limiting. Therefore, a simple modification of the amplitude correlation that offsets it by one block or time window is useful. If the attack is preceded by relative silence, then it will be given a minimum stretch, whereas the following steady state will be stretched by the maximum amount.

Manual control of the rate via the presets has also proven to be extremely effective, even with rapidly changing material such as speech. With normal human reaction time, particular vowels or consonants can be elongated in the midst of a speech stream. At the moment, only manual, preset, ramped, and automated controls of the off:on ratio are available in the variable-rate version; scores and masks have not been implemented. The main reason for this is that such higher-level controls would be difficult to synchronize with the temporal pattern of the particular sound. Although it is occasionally compositionally interesting to impose a pattern onto a sound in an unsynchronized manner, it has been more useful to develop controls correlated to the sound material itself, such as the automated rate control just described.

Harmonizing

Once an independence of pitch and duration was achieved with the time-shifting technique, it seemed desirable to add simultaneous sample transposition. There are several standard approaches to this problem; the simplest one, harmonizing, is based either on skipping an integer number of samples (e.g., taking every second sample for a transposition an octave up, every third sample for the third harmonic, and so on) or on repeating samples (e.g., using each sample twice for a transposition an octave down). Computationally more complex approaches involve interpolation or using a non-integer sample increment, as with a digital oscillator. Typically, the user employs x bits for the integer part of the table lookup address and y bits as the fractional part, thereby simulating a wavetable of size $x + y$. Without the fractional part, one can only produce harmonically-related frequencies; with it, the frequency resolution is much better.

Implementation of the harmonizer approach would be simple were it not for the technique used to realize the time stretching, namely the fact that the software alternates between freezing the contents of memory for m milliseconds and advancing through the sample sequence for the next n milliseconds, the ratio of $m{:}n$ being the off:on ratio. During the periods in which the memory contents are "frozen," the user is forced to go backward through the stored samples to obtain the material for the grains. However, because the current time position may advance during the next millisecond, the rate at which the user steps through the samples needs to be higher during the "on" times to continue progressing backward at the same rate and thus maintain the original pitch without discontinuities.

For instance, with a current address marking the most recent sample received and a desired offset number of samples referring to a point in the past at which the grain is to start, the following equations show how to determine the address of the next sample (sample address) during the "off" and "on" times (note that some strategy is also required to keep the result within the memory address range):

$$\begin{aligned}\text{off mode: } & \text{current address} = \text{current address}\\ & \text{sample address} = \text{current address} - (\text{offset} + 1)\end{aligned} \tag{2}$$

$$\begin{aligned}\text{on mode: } & \text{current address} = \text{current address} + 1\\ & \text{sample address} = \text{current address} - (\text{offset} + 2)\end{aligned} \tag{3}$$

A simple harmonizing scheme to transpose the material to harmonic N would generalize the expression in brackets to (offset + N) for the off mode and (offset + N + 1) for the on mode. However, having only an upward transposition to a harmonic frequency is a severe limitation for the user, both in the absence of lower transpositions and the wide spacing of the first few harmonics. An alternative scheme, one that is implemented in GSAMX, chooses a harmonic series based on a fundamental achieved by dividing the untransposed frequency by a factor F. The case of $F = 4$ is particularly attractive because it allows a downward transposition of two octaves, such that the fourth harmonic is the original pitch, plus four transposition levels in the octave above the original (harmonics five to eight). This is illustrated in Figure 4. Equations 2 and 3, which determine the sample address, can be modified as follows to achieve harmonic N:

Figure 4. Harmonization scheme for transposition to the harmonic N = 4.

HARMONIZATION SCHEME

Harmonic Number		Relative Interval
1		2 octaves down
2		1 octave down
3		4th down
4	------->	normal pitch
5		3rd up
6		5th up
7		7th up (7:4)
8		1 octave up
9		
10		
11		
12		
13		
14		
15		
16		2 octaves up

off mode: sample address = current address – $(F * \text{offset} + N)/F$ (4)

on mode: sample address = current address – $(F * \text{offset} + N + F)/F$ (5)

Note that whereas F is a constant, N can vary with each grain or stream of grains, thereby allowing simultaneous transposition to different pitch levels in a multiple-voice implementation. The DMX-1000 version of this algorithm has a maximum of 15 simultaneous voices (10 at 30 kHz), each with its own transposition level and choice of stereo output channel. The results range from the expected chordal enhancement of the material with lower pitched material to interesting timbral enrichment with unpitched or high-frequency material.

A byproduct of this scheme is the combination of the octave downward transposition with the option of skipping every other sample at the input to the processor's memory. The skipping not only speeds up the material by a factor of two but also transposes it up an octave; when combined with the downward transposition, the material is returned to its original pitch but with the doubled tempo.

Psychoacoustic Implications

Gabor (1947) described the microscopic level of the grain as a quantum of sound whose parameters of frequency and time form a unit rectangle. If one "squeezes" the rectangle in the time domain and thereby shortens the grain duration, the frequency domain expands in compensation—that is, the bandwidth increases. This phenomenon is also known as the "law of uncertainty" in reference to Heisenberg's uncertainty principle, involving the position and momentum of an electron, and is most generally described as the inverse relation between frequency and time. For instance, in spectral analysis, longer time windows are required to define the low-frequency components accurately, and vice versa, the trade-off being between temporal and frequency resolution (Dolson 1986).

By linking frequency and time at the micro level, granulation makes it possible to treat the two variables independently at the macro level, as described here. Gabor was also aware of this application and performed rate-changing experiments using an adapted film projector (Roads 1991). However, at the macro level, the perceptual results of time stretching work on a similar inverse relationship—as a sound is progressively stretched, one is less aware of its temporal envelope and more aware of its timbral character. Ironically, with extreme stretching in time, a spectrum can be experienced psychoacoustically in the classical Fourier manner, namely as the sum of its spectral components! Brief acoustic events tend to be recognized non-analytically according to their overall loudness as well as their temporal and spectral envelopes, which are perceived as a gestalt pattern. With stretched sounds, one has time to refocus one's attention on the inner spectral character of a sound, which with natural sounds is amazingly complex and musically interesting. Therefore, transitions from the original to the stretched versions provide an interesting shift from one dominant percept to another.

Two related phenomena are commonly experienced with time shifting: the emergence of momentary resonances that are often quite vocal in character and the perception of increased volume, as distinct from mere loudness. The first phenomenon is described elsewhere as the emergence of "inner voices" (Truax 1992a)—that suggested archetypal imagery that inspired my works *Pacific* and *Dominion*. Particularly surprising was the discovery of these voices, resembling a distant choir singing vowels, in the sound of ocean waves. The only explanation of the effect that I can provide is that momentary resonances that normally are too fleeting and non-repetitive to be identified become audible by being prolonged and reinforced with multiple overlays. An analogy might be made to the microscope, which allows minute spatial patterns to be perceived by magnifying them. In general, time stretching is a unique way to bring out the inner complexity of a sound. Speech and other nonhuman utterances are also quite complex in this regard; most mechanical sounds and practically all sounds that are electronic in origin are much less so.

To explain the perception of increased volume in the sound, one can go back to the gestalt concept of volume as "the perceived magnitude of a sound," which tends to increase with spectral richness (or resonance), reverberation, duration, and of course intensity. This concept was current in early psychoacoustics (Seashore 1938) and can be found as late as 1967 in Olson (p. 260), by which time it was rapidly being replaced by the stimulus-response paradigm of loudness based on summation of sine tone components for pitched material, and narrow band noise components with aperiodic sounds (Kryter 1959). Unfortunately, this correlation of loudness to spectral intensity, just as timbre is related to Fourier components, continues to be repeated in most introductions to the subject in books on electroacoustic music and audio production and then is promptly ignored in practice, presumably because such theories are of little practical use to the composer.

In an effort to stimulate renewed interest in a complex, multi-parameter concept such as volume, I have proposed a working model whose dimensions are spectral richness, time, and "temporal density" that refers to the temporal spacing of independent spectral components, such as multiple sound sources and phase-shifted or time-delayed events (Truax 1992c). Time stretching contributes to all three dimensions, hence the perception of greatly increased volume. The overlay of simultaneous grains enhances spectral richness, the lack of synchronization between simultaneous grain streams adds to the temporal density, and the extended duration occurs along the time axis.

Compositional Experience

Composing with real-time granular sound (Truax 1990b) has not only opened up a new sonic world, but has also challenged some very fundamental ideas about what composition is. Whereas instrumental music models assume the note as the smallest compositional unit, granular synthesis works at the micro level of the grain. Composition means working within the sound as much as it does creating larger structural units. In fact, with this technique, sound and structure are extremely closely intertwined. The conventional distinctions—found even in computer music systems—between score and orchestra, or in MIDI between note commands and synthesizer patches, are obliterated in a more integrated, even organic process. Moreover, the issue of compositional control that has already been challenged by the use of aleatoric processes must be rethought in terms of the complex interaction of parallel processes found in a real-time granular synthesis system. Deterministic and linear thinking are clearly inappropriate, if not impossible; the composer is constantly being challenged by new concepts of sound and its organization, and if for no other reason than that, the technique may resist widespread commercialization.

The first two works based on the granulation of sampled sound, *The Wings of Nike* (1987), for computer images by Theo Goldberg and two soundtracks, and *Tongues of Angels* (1988), for oboe d'amore, English horn, and four soundtracks, use very short, fixed samples of recorded material. In the first work, these samples are male and female phonemes, and in the second piece the samples are derived from the live instruments. Despite the brevity of the source material, very rich textures and complex rhythmic patterns can be obtained. The pitch and timbre of the resulting sound are determined by the source material unless the grain duration is too short and a broad-band spectrum results. However, the overlay of up to 20 simultaneous versions of such sound per stereo pair of tracks, each with its own variations, produces a "magnification" of the original sound, as well as introducing the possibility of gradual or rapid movement through its micro-level characteristics.

The degree of magnification involved can be appreciated when it is realized that three of the four movements of *The Wings of Nike*, lasting approximately 12 min, were derived from only two phonemes, each about 170 msec long! The stereo tape is a mixdown from an eight-track original that includes four stereo pairs of the granular material; therefore, the vertical densities of sound are around 80 at any one moment, and the horizontal densities range from quite sparse to 8,000 events per second at the very end.

The first work to use the time-stretching technique was a mixed-media performance piece for children and adults called *Beauty and the Beast* (1989), a collaboration with Theo Goldberg's computer graphic images. The work also includes a soloist using English horn and oboe d'amore (Lawrence Cherney, who commissioned the work with the assistance of the Canada Council), who acts as the storyteller using his instrument. The narrative text of the story is embedded within the computer graphics as well as heard as verbal dialogue on the accompanying tape. This dialogue proved to be effective source material for variable-rate granulation and, except for some use of the instrumental sounds for continuity during interludes, was the only source material needed to create the soundscape that accompanies the graphics. The compact nature of speech, incorporating as it does many acoustic elements (e.g., pitch, noise, formants) in a short space of time, makes it a particularly rich source material for time extension.

My recent work, *Song of Songs* (1992), for the same combination of elements as *Beauty*, uses male and female voices reciting the Biblical text of the title, plus environmental recordings of bird song from France, a stream and crackling fire from British Columbia, and cicadas, crickets, and a monk singing along with a monastery bell from Italy. Time shifting is used to modify the rhythm of the spoken text subtly and make it more songlike and to prolong the sounds into sustained timbral textures, frequently accompanied by multiple pitch shifting implemented with the harmonizing technique described earlier. This enrichment allows a more complex timbral construction to be derived from the original, and this is emphasized even further by the time shifting. Moreover, the amount of stretching was modified during the recording of the environmental soundtracks in response to others already present, thereby creating a constant interaction of all the material and further blurring the distinction between voice and environment. This sense of merging of sonic elements is consistent with the extended metaphor of the original text, which compares the Beloved to the richness of the landscape and its fruits. The melodies of the live instrumental part are derived either from the tempo and pitch inflections of the spoken text or from the traditional Hebrew cantillation on the *Song of Songs*, which at the end of the piece is intertwined with the monk's song from the Christian tradition. Time-shifted granulation allows the traditional boundaries between speech, music, and the soundscape to be blurred.

Time shifting of environmental sound is the main technique in three other recent works *Pacific* (1990), *Dominion* (1991), and *Basilica* (1992). In *Pacific*, one sequence of sounds is used for each of four movements. The materials are recordings of Canadian West Coast environmental sounds, namely ocean waves on the west coast of Vancouver Island, boat horns in Vancouver harbor on New Year's Eve, Vancouver harbor ambience with seagulls, and the Dragon Dance in Vancouver's Chinatown celebrating the Chinese New Year. In *Dominion*, the materials are recordings of Canadian "soundmarks," such as bells, whistles, foghorns, cannons, etc., as recorded by the World Soundscape Project during a cross-country tour in 1973. These materials are presented in an east to west direction with at least one sound from each province, suggesting a journey "from sea to sea," linked by the sound of the whistle of the transcontinental train, the railroad whose completion forged the founding of the country. The work is divided into four sections, each depicting a region of the country and starting with a unique soundmark that signals high noon (the noonday gun in St. John's, Newfoundland, the Westminster chime and hour bell from the Peace Tower in Ottawa, a noon siren from a small town in Alberta, and the "O Canada" horn sounded daily in Vancouver). The 12 strokes of the Ottawa bell appear in counterpoint with the bells of a representative of the other founding culture, the Basilica in Quebec City, whose sounds were processed further in the tape solo piece *Basilica*. The attack portion of each sound signal in the piece is minimally stretched, thereby preserving its recognizability, but the remainder of the sound is often prolonged by a significant amount that allows the listener to hear its inner musical character, the pitches of which form the basis of the twelve instrumental parts that are embedded within the tape sounds. In this work, the time stretching not only brings out the inherent musical character of the source material but also gives the listener time to allow memories and associations of these sounds to surface.

In *Basilica*, the three bells are heard at their original pitch, as well as an octave lower and a twelfth higher, but all of these versions are stretched, often to more than 20 times their original duration. The extended versions allow the listener to hear out the inner harmonics inside the bells, and in moving inside the sound, it seems as if the listener is entering the large volume of the church itself. The bell formants can easily be confused with those of a choir, and two-thirds of the way through the piece, a repetitive sequence of momentary bell spectra is heard, each element transposed down an octave and prolonged to resemble the melody of a chant. The piece ends similarly to the decelerando of the original bell sequence except that the effect is simulated by progressive time stretching of a single bell with additional harmonic pitches.

Works by other composers who have used these techniques with sampled sound are *Valley Flow* by Denis Smalley, *Crow* by John Rimmer, *Ocean of Ages Revealed* by Wende Bartley, *Birth/Rebirth Bearing Me* by Susan Frykberg, *Bronze Wound* by Chris Rolfe, and *Essai du vide, Schweigen* by Agostino Di Scipio.

The technique of granular time stretching provides a unique way to experience the inner structure of timbre, hence to reveal its deeper imagery (Truax 1992a). For instance, each movement of *Pacific*, as well as the other works mentioned, is based on the imagery inherent in the environmental sound used as its material. Moreover, in the composition of each work, a metaphor is established that connects the sound to a deeper sense of cultural symbolism. All of this symbolism is designed to involve the listener strongly in the musical process by presenting a larger-than-life image of sounds that are strangely familiar. Whereas a simple collage of the material would provoke recognition of it only as sound effects, and typical sampler looping or concatenation would present the sounds more abstractly and devoid of context, the time-stretching technique draws the listener into the sound and evokes its inherent imagery and associations. The process results in what is described elsewhere as a music of complexity (Truax 1992a; 1992b), a music that is strongly contextualized, in contrast to music composed according to the dominant paradigm, in which sounds are related only to each other, thereby creating completely abstract works of art. The aim is to relate the inner complexity of the sound to the outer complexity of the real world, such that the two are integrated.

Conclusion

The complexity and dynamic quality of granulated sampled sound makes it an attractive alternative to methods based on the looping and transposition of such sounds. Moreover, the basic unit, or "quantum," of the grain is a potentially more flexible building block for treating sampled sound, particularly because the amplitude envelope of the grain avoids transient clicks when extracting and combining arbitrary sample segments. When granular synthesis is used to produce time-extended textures, it has no resemblance to instrumental and other note-based music; instead, the acoustic result often brings out the inner character of environmental sound.

However it is used, granular synthesis is clearly situated in a different psychoacoustic domain than that occupied by most computer music, which commonly is based in instrumental music concepts. By separating the micro level from the macro level in terms of sound features and allowing pitch and time warping without the limitations inherent in Fourier-based approaches, the techniques introduced here create a unique sound world and suggest new ways in which the music made with it can be related to the external world.

References

Dolson, M. 1986. "The Phase Vocoder: A Tutorial." *Computer Music Journal* 10(4): 14–27.

Gabor, D. 1947. "Acoustical Quanta and the Theory of Hearing." *Nature* 159(4044): 591–594.

Jones, D., and T. Parks. 1988. "Generation and Combination of Grains for Music Synthesis." *Computer Music Journal* 12(2): 27–34.

Kryter, K. 1959. "Scaling Human Reactions to the Sound from Aircraft." *Journal of the Acoustical Society of America* 31(11): 1415–1429.

Lippe, C. 1993. "A Musical Application of Real-Time Granular Sampling Using the IRCAM Signal Processing Workstation." *Proceedings of the 1993 International Computer Music Conference*. San Francisco: International Computer Music Association.

Olson, H. 1967. *Music, Physics and Engineering*. New York: Dover.

Roads, C. 1978. "Automated Granular Synthesis of Sound." *Computer Music Journal* 2(2): 61–62.

Roads, C. 1988. "Introduction to Granular Synthesis." *Computer Music Journal* 12(2): 11–13.

Roads, C. 1991. "Asynchronous granular synthesis." In G. De Poli, A. Piccialli, and C. Roads, eds. *Representations of Musical Signals*. Cambridge, Massachusetts: MIT Press.

Seashore, C. 1938. *Psychology of Music*. New York: McGraw-Hill.

Truax, B. 1988. "Real-Time Granular Synthesis with a Digital Signal Processor." *Computer Music Journal* 12(2): 14–26.

Truax, B. 1990a. "Chaotic Non-linear Systems and Digital Synthesis: An Exploratory Study." *Proceedings of the 1990 International Computer Music Conference*. San Francisco: International Computer Music Association.

Truax, B. 1990b. "Composing with Real-time Granular Sound." *Perspectives of New Music* 28(2): 120–134.

Truax, B. 1992a. "Composing with Time-Shifted Environmental Sound." *Leonardo Music Journal* 2(1).

Truax, B. 1992b. "Electroacoustic Music: The Inner and Outer World." In J. Paynter, et al., eds. *Companion to Contemporary Musical Thought*. London: Routledge.

Truax, B. 1992c. "Musical Creativity and Complexity at the Threshold of the 21st Century." *Interface* 21(1): 29–42.

Truax, B., and T. Bartoo. 1992. "The Electroacoustic Composer's Workstation." *Proceedings of the 1992 International Computer Music Conference*. San Francisco: International Computer Music Association.

Wallraff, D. 1979. "The DMX-1000 Signal Processing Computer." *Computer Music Journal* 3(4): 44–49.