TIME-SHIFTING OF SAMPLED SOUND WITH A REAL-TIME GRANULATION TECHNIQUE

Barry Truax
Dept. of Communication & School of the Contemporary Arts
Simon Fraser University
Burnaby, B.C. CANADA V5A 1S6

ABSTRACT

Recent work (Truax, 1988) has shown that granular synthesis (Roads, 1978) may be implemented in real time using the microprogrammable DMX-1000 Digital Signal Processor (Wallraff, 1979) to produce complex sounds by the generation of high densities (e.g. 100 - 200 events/sec) of small 'grains' of 10-50 ms duration. Techniques involving additive synthesis, FM, and sampled sound for each grain have been implemented, with a hierarchy of control parameters directing the density, range and temporal evolution of the synthesized textures. This paper reports on a technique for variable rate time shifting of the material that leaves pitch intact. The technique is similar to that of Jones and Parks (1988) except that our goal is to lengthen the sound, not shorten it. In addition, it works in real time unlike computationally intensive methods such as the phase vocoder.

I. INTERPOLATING BETWEEN FIXED AND CONTINUOUS SAMPLING

The real-time program GSAMX implements an instrument for granular synthesis where 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 ms) between successive envelopes. Each generator produces a three-part linear envelope whose attack and decay are a specifiable fraction of the event duration. Additional variables include the start sample (or offset) and the range of this variable. Twenty simultaneous voices of this instrument are possible with the DMX-1000 and are controlled by a scheduler program on the host PDP Micro 11 where each grain is initiated and terminated during 1 ms clock interrupts. The shorter the grain duration, the higher the overall density of grains per second (gs). The minimum grain duration that can be controlled in real-time is 10 ms, hence densities of up to 2000 gps can be achieved with 20 simultaneous voices. When the grains are unsynchronized (i.e. of variable duration) and/or 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 to treat the sampled sound:

1) Fixed Sample (approx. 4K samples)
2) Continuous Sample Input from Disk at normal speed

The first option uses a fixed and rather short sequence of source material, namely 4032 samples or around 150-170 ms of sound, because of the 4K onboard memory in the DMX. The duration of grains used in granular synthesis is typically less than 150 ms, so the effect of the fixed sample size is to limit the variety of simultaneous 'windows' that may be accessed. The second version involves real-time granulation of continuous sound directly from disk with the 4K memory acting as a short delay line that may be tapped to furnish the various grains.

Work with these two approaches suggested the need for a method which would interpolate between them, that is, to vary the rate at which new samples were introduced into the granulation process. The desired effect of this 'variable rate' approach is to depart from the normal time flow of the continuous sampling model in a manner which would eventually

ICMC GLASGOW 1990 PROCEEDINGS
approach the 'frozen' time of the fixed sample model. Such an interpolation would preserve the sense of ongoing development of forward time flow but combine it with the sense of magnification of the moments associated with the fixed version.

II. VARIABLE RATE SAMPLED SOUND

In this implementation, the key variable is the ratio at which new sound samples enter the DMX's memory from disk (Fig. 1). 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 (i.e. no 'off' time), and a ratio of 99:1 results in 99 ms of no movement through the sample before there is a 1 ms shift forward, thereby producing a hundredfold reduction in speed and a time-extension of the sample. However, since the grains are always taken from the current contents of DMX memory at one consecutive sample per calculated output, the frequency of the source material is not distorted, only the rate at which one advances through it in a 'macro' sense. That is, micro-level waveform patterns and macro-level timbral changes have been effectively separated.

![DMX Memory Diagram](image)

Fig. 1. DMX memory during variable rate granulation.

During the 'off' milliseconds when the contents of memory are frozen, the grains take their samples in the reverse direction since at this time level there is no aural difference between forwards and reverse. During the 'on' milliseconds, new samples are added to the DMX memory (thereby losing old ones) and the grains may be taken in either the forwards or reverse directions. This choice produces the following options for the user:

1) Sample direction is always in reverse, thereby producing minimal spectral alteration.
2) Sample direction alternates (forwards/reverse) during the 'on' and 'off' states, thereby modulating the sample at the micro level.

In the first option, 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 which is introduced by the granulation process. Since our purpose is to produce musically interesting sound and not just a processed signal, we use 12-20 simultaneous superpositions of the grains each with independent characteristics to give increased volume and complexity to the sound.

In the second option, the N:R ratio, e.g. 1:1, 2:2, 3:3, etc. (which are not equivalents) produce an interesting phase modulation effect, because for equal amounts of time (1, 2, 3, ... ms respectively) the grain goes backwards through the sound sample, then forwards for the same duration, and so on. The effect is somewhat similar to a phase vocoder and the given series of ratios produces a decreasing subharmonic series of phase modulation frequencies from 500 Hz. Likewise, ratios of 2:1, 4:2, 8:4 combine a certain amount of phase modulation with the slowing down effect that larger ratios produce. However, instead of a strong pitch component being added, these ratios produce a noisier broad-band result.
In each case, the amount of time shift, described as the Time Extension Factor (TEF), can be calculated from the off ratio as follows:

\[
\text{Time Extension Factor} = \frac{(\text{Off Ratio} + \text{On Ratio})}{\text{On Ratio}}
\]

Therefore, the ratio 1:1 produces a TEF of 2 times normal speed. A ratio of 999:1 produces a TEF of 1000 times normal speed, meaning that one second of sound lasts over 16 minutes! Since there is a proportional relationship between the TEF and the ratio, a continuously changing TEF is possible by ramping the off part of the ratio compared to the on part.

II. 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 which 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 which the user selects. This maximum value is implemented during the blocks with peak amplitude, with proportionately smaller values during blocks with other amplitudes.

However, manual control of the rate via presets has also proved effective, even with rapid material such as speech. With normal human reaction time, particular vowels or consonants can be elongated by activating a preset in the midst of a speech stream. At the moment, only manual, preset, ramp and automated controls of the off/on ratio are available in the variable rate version; that is, scores, masks and trajectories have not been implemented as they are for fixed samples.

III. COMPOSITIONAL EXPERIENCE

Composing with real-time granular sound (Trax, 1990) has both opened up a new sonic world and challenged some very fundamental ideas about the nature of composition. 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, in this technique sound and structure are extremely closely intertwined. The conventional distinctions, found even in computer music systems, of score and orchestra, or MIDI note commands and arbitrary synthesizer patches, are obliterated in a more integrated 'organic' process. Moreover, the issue of compositional control which has been challenged by the use of aleatoric processes must be rethought in terms of the complex interaction of parallel processes found in real-time granular synthesis. 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, the technique may resist commercialization.

My recent works with variable rate granulations have proved to be an equally radical departure from previous experience. The first is a mixed-media performance piece for both children and adults called Beauty and The Beast (1989), a collaboration with Tito Goldberg and his computer graphic images. The work also includes an English horn and cello performer who acts as the storyteller using these instruments. The narrative text of the story is embedded within the computer graphics as well as heard as verbal dialogue on the accompanying tape.

The dialogue of the story, along with some instrumental sounds during the interludes, is the only source material used to create the soundscape that accompanies the graphics. The
compact nature of speech, incorporating many acoustic elements (e.g. pitch, noise, formants) in a short space of time, makes it a particularly rich sound material for time extension. The material has been transformed using three typical ranges of time extension (short, medium, and long) to create the following compositional elements and their associated ratios:

1) Transformed speech (1:1 to 5:1)
2) Musical/treatment of speech (5:1 to 20:1)
3) Environmental speech ambience (20:1 to 999:1)

In the first case, speech remains comprehensible but the timbre is altered to various degrees. The voice of Beauty, for instance, is often modulated to add various pitches using the N:N ratios, or is time extended with no modulation for expressive effect. The Beast voice is created using ratios such as 3:1 and 5:1 to add noisy modulations. With the longer ratios in the second case, comprehensibility may be lost in favour of extended phonemes and their spectrally rich timbres. A surprising result of the stretching of inflected speech is that the normally fleeting pitches resemble the sung voice when time extended. Although the percussive quality of the consonants is lost, the voiced phonemes become strikingly melodic in the time-extended versions.

Finally, extreme time extensions of 50, 200 or even 1000 times slower than normal speed prolong the sound past anything humanly possible and often suggest the textured sounds of natural ambiances. The story of Beauty and The Beast can be understood as a projection of Beauty's psychic state, and therefore it seems appropriate that the sonic environment in which she finds herself is quite literally an extension of her own voice. Although the most extreme extension (Beast) appears fearful to her at first, the story documents her eventual acceptance of it. In the music, this is symbolized by a gradual removal of the magic of the time extension.

From the point of view of purely timbral concerns, the time stretching technique increases the perceived volume or magnitude of a sound without necessarily altering its pitch or loudness. This is also the case with Pacific (1990) which is based on time-extended environmental sounds such as waves, hoofbeats, seagulls and percussion sounds. First, there is the increase of spectral richness by the superposition of 12-18 versions of the source per stereo pair of tracks. Such overlays intensify bands of spectral energy whether those of formants or noise elements. Secondly, the simultaneous voices are not phase coherent with respect to each other. This temporal independence (in the range of phasing and reverberation effects, viz. less than 50 ms) also results in the composite sound seeming to have greater volume. Finally, the variable rate technique adds a third dimension to the perceived magnitude, namely time extension: Spectra that are normally brief instants in time can now occupy virtually any duration. Speech is given a larger-than-life volume until it becomes the equivalent of substantial environmental sounds. Paradoxically, the stretched environmental sounds of Pacific become quite vocal in character. In the process of moving inside the sound, we find like Alice in Wonderland that it has expanded until we are surrounded by its magnified image.

References:


Truax, B. 1990. "Composing with real-time granular sound." Perspectives of New Music, 28(2).


ICMC GLASGOW 1990 PROCEEDINGS

107