Decorrelation as a By-Product of Granular Synthesis

Chris Rolfe,
Third Monk Software, Vancouver, Canada

Damián Keller
Center for Computer Research in Music and Acoustics, Stanford University, CA, USA
http://ccrma.stanford.edu/~dkeller

Rolfe, C., & Keller, D. (2000). Decorrelation as a by-product of granular synthesis. *Proceedings of the XIII Colloquium on Musical Informatics*. L'Aquila: AIMI.

Abstract

Several researchers (Kendall, 1995; Truax, 1992) have noted that decorrelation occurs as a by-product of granular synthesis (GS). Decorrelation between grain streams is responsible, for instance, for the unique stereo and panning effects that have been achieved under GS transformation. The correlation measure itself, however, is not generally explicit or variable within existing synthesis models. The following paper describes a systematic approach to granular decorrelation, relating individual parameters to their effect upon grain-to-grain, cross-channel (stream) and instance (event) signal correlation.

1 Introduction and Definitions

The cross-correlation measure of two signals is a significant predictor of many perceptual phenomena including spatial imagery [3], constructive and destructive interference [1], echo suppression [4] and externalization under headphone listening [2]. In multi-channel loudspeaker reproduction, inter-channel cross-correlation (ICCC) predicts the likelihood that a given signal will, all else being equal, suppress an earlier reflection arriving from a different direction (the precedence effect) [2].

The cross-correlation measure of two signals y1 and y2 is:

$$F(\partial t) = \lim 1/2\ T\ \sum y1(t) * y2(t+\Delta t)dt$$

To obtain a useful, single measure of cross-correlation, we take the peak value of the cross-correlation function normalized to a range -1.0 to 1.0.

Expressed statistically, the cross-correlation function is the covariance between two signals divided by the product of their standard deviations. In audio signal processing, it may be more familiar to describe cross-correlation as passing an input signal y1 through a weighted, moving-average filter, y2. The value, *k*, is adjusted for DC offset, RMS and group delay, and is presumed to provide a convenient measure of similarity between two sounds.

Decorrelation is defined here as *any* technique that reduces the absolute value of the cross-correlation measure between two signals y1 and y2 (not only allpass filters designed explicitly for that purpose).

For practical purposes, given an arbitrary function F(x), and excluding the trivial cases of identity F(x) -> y=x, delay F(x) -> y[t]=x[t-z] and phase inversion F(x) -> y=-x, the cross-correlation(x,y), is always $-1.0 > k < 1.0$, so we further restrict our definition of decorrelation to include only those techniques that permit some means of influencing the degree of correlation across the entire range for *k*, -1.0 to 1.0.

2 Importance of Correlation in Granular Synthesis

Granular synthesis (GS) of sampled sound, or, as it is increasingly known, time-scrambling granulation, is essentially a statistically-controlled mixing scheme that recombines thousands of short snippets (grains) into multiple channels of output. In such mixing systems, constructive and destructive interference is of paramount importance not only in determining subjective outcomes, but also in controlling overall output signal levels.

Furthermore, because grain rates usually lie on the boundary between audio and event rates, GS output may be best described as depending upon emergent perceptual properties, wherein correlation measures have been shown to be useful predictors. Correlation in GS successfully predicts chorusing

and echo, volume and diffuseness, and the robustness of spatial effects.

Most published granular synthesis (GS) models also allow for stochastic variation of control parameters, usually as a parameter range specification, as does our model. One reason for introducing randomness is to distribute or 'smear' artifacts caused by windowing and granulation both in the time-domain (pulsing, or beating) and the frequency-domain (amplitude modulation and comb-filtering).

For example, an amplitude modulation of a complex source signal by a 40 millisecond triangular window at a scanning ratio of 1:1 distributes the input signal's energy equally between the original spectrum and generated AM sidebands. The sound is subjectively mechanical because the artifacts are bound, perceptually, in the frequency domain. The AM sidebands are phase-correlated with the input signal, and thus perceived in toto as a spectral modification to the original input.

By focusing on those parameters to the GS model that help to decorrelate the signal, we can contrive strategies to avoid or minimize such artifacts, or, indeed, tailor the artifacts to our musical purposes.

Finally, although it is correct in one sense to describe decorrelation as a by-product of granulation, it is one of the most characteristic outcomes of GS processing, and thus deserving of closer examination.

3 (De-)correlation at Various Levels

In granulating sampled sound for various applications, it is useful to consider three related, but distinct, cross-correlation measures (described below). In characterizing each level, we will make reference to two phases of granulation, *analysis*, during which the choice of position within the input stream(s) is made, and *synthesis*, during which the source is manipulated and recombined into multiple grain streams. For reasons of computational efficiency, analysis in real-time GS systems tends to be limited to advancing a pointer according to a set of parameter controls. Particularly in real-time systems such as ours, we speak of decorrelation as a by-product of the synthesis stage, without implying that an actual correlation measure is to be calculated during run-time.

The first level, grain-to-grain correlation, is defined as the cross-correlation measure between successive grains within a single stream. Given a sine tone input to the process, for example, grain-to-grain correlation varies as a function of source synchronization, that is, measures how closely the phase of the grain envelope function aligns with the sine tone's frequency. Variation in correlation in this simple case is slight, affected mainly by boundary and base-frequency considerations.

In more complex examples, however, or in the case where input has been analyzed to form a pre-selected pool of grains rather than a continuously sampled sound event, or finally, when the grain duration is smaller than the audio rate boundary (< 20 msec), then grain-to-grain correlation becomes more important. In this sense, grain-to-grain correlation predominates during the analysis phase, and is significant primarily in matters of synchronization.

The second level, stream-to-stream correlation, is defined here as the cross-correlation between grain voices, or, *streams*, each stream representing an independently controllable channel to be mixed or routed to a given output(s).

Stream-to-stream correlation is important during the synthesis phase and has the most predictive value in terms of how multiple streams will be perceived after they are combined. It is here that we find the characteristic GS effects such as increased volume and diffuseness.

And thirdly, output instance correlation is stream-to-stream correlation extended to include non-contemporaneous GS output streams, either at the event level, or generated on different occasions.

It is important to mention decorrelated instances in order to emphasize the profound ecological difference between identical and similar events, and the general desirability of decorrelation in the electroacoustic domain. A wholly deterministic synthetic process, for example, produces identical output from a given parameter set; serendipity and variation is removed, and the resulting output is exactly repeatable. In contrast, most GS systems rely upon stochastic variation of parameters resulting in similar, but nearly always unique outputs.

Not all parameters to a GS model affect output instance correlation equally, however: amplitude may be varied stochastically at a grain rate, for instance, but does not decorrelate one instance from another. The effect of individual parameters and their role in decorrelation is considered below.

4 Synthesis Parameters

Our granular synthesis model presents the user with the following parameters, each controllable independently per stream:

advance_rate**:**	controls time-expansion or -compression;
delay:	initial sample delay;
delay_range:	varies sample delay;
mod_phase:	initial phase of amplitude modulator;
envelope:	amplitude modulator function

duration: grain duration (AM frequency);
duration_range: varies grain duration;
amplitude: grain amplitude.

Grain streams can be synchronous or quasi-synchronous according to the setting of delay_range parameter.

As granulation begins, each stream's read pointer is set to its initial delay +- delay_range. If no time-expansion or –compression is chosen, that is, if the advance rate is equal to the original sampling rate, then the process resembles a multi-tap delay line. Each delay tap, however, is also amplitude modulated by the grain envelope function, which in our case is usually a triangle or variable trapezoidal window with a peak equal to grain amplitude, and edges set to zero.

Each delay tap is updated at the end of each grain cycle according to a random value whose range is controlled by the delay_range parameter.

At the end of each cycle, the modulation function is zero, thus allowing us to move to a new value without encountering discontinuities.

The sum of the grain streams is described by the instantaneous impulse response shown below:

Careful selection of initial delays and phases allows for the creation of many traditional effects, such as comb-filtering and reverberation, although usually with the addition of a "beating" or amplitude modulation caused by the grain enveloping. This common side-effect of granulation, however, can be avoided simply by pairing grain streams such that, for each pair, the modulation functions combine to a constant amplitude [5]. Under this pairing scheme, the total number of taps available is equal to the number of grain streams divided by 2, since it requires two output streams to, in effect, cross-fade between successive delay values.

If all delay tap values are constant, then the process is time-invariant and determinate. Grain-to-grain correlation will depend upon the grain durations chosen and the nature of the input signal. Cross-correlations between paired grain streams will equal 1.0, and the cross-correlation between output instances will be also be constant at 1.0. The subjective impression when using short delays is that all streams tend to fuse into a single percept and location, and repetitions (instances) of the process result in exact duplicates.

When a stream's grain duration is allowed to wander by even a few samples, however, the cancellation of amplitude modulation within stream pairs quickly breaks down as the modulation phases within pairs are randomized. Desynchronizing the modulation phases, however, only decorrelates the output streams very slightly: amplitude modulation by any unipolar function, such as a triangle window, combines the input signal with generated sidebands and thus results in a highly correlated, if non-linearly distorted, output. Another way to look at it: as the number of streams approaches infinity, the combination of the modulation functions approaches a constant. For practical purposes, then, 24-64 streams is sufficient to distribute AM effects, although we note that amplitude variation does not, per se, decorrelate a signal (although in practice, quantization error can introduce some calculable, if not desirable, decorrelation).

The chief means of controlling decorrelation in the described GS implementation is to adjust the delay_range parameter in conjunction with the stream amplitudes. Varying a given stream delay by a random amount introduces phase-shifting causing the value k to vary dynamically from –1.0…1.0. The precise amount of decorrelation depends upon the relation between grain duration and source content, but can, with practice, be tuned by ear to the desired result.

Perceptually, introducing randomness into the delay taps creates a chorusing effect between streams. By controlling the distribution, amplitude and range of these delays, we exercise reasonably effective control over both the stream-to-stream and output instance correlations. Generally, the rule of thumb is that greater delay variation in more prominent (louder, earlier) streams increases output instance correlation.

The final parameter considered within our model is the advance rate. Advance rate controls the average grain hop as the input buffer is scanned. As Jones and Park [5] note, transparent (high-quality) time-compression and –expansion require that grain-to-grain correlation be maximized to maintain source phase synchronicity and thus to minimize boundary and overlap artifacts.

Because we selectively introduce random variation into our stream delays, however, we are in a sense forgoing the goal of transparency in our GS model in favour of a thickening or chorusing, and thus decorrelating effect. Transparent time-expansion requires maximizing correlation at all levels, best suited to one or two determinate streams, while we are interested equally in useful applications of decorrelation and multiple streams.

Generally, we avoid undue artifacts by selecting grain durations and envelopes by ear, rather than by calculation. As noted earlier, the grain-to-grain correlation is more important during the analysis than the synthesis phase of GS, which in a real-time system where efficiency is a concern.

Another parameter commonly found in GS applications but not considered as part of the above implementation, is pitch-shift:

Pitch-shifting is best counted as a form of stream pre-processing, for the practical reason that the

amount of decorrelation is strongly dependent upon the choice of algorithm: a simple, non-interpolating drop-sample technique, for example, introduces highly uncorrelated error into the signal, whereas more sophisticated algorithms may produce correspondingly less error. As with other uncorrelated errors, such as those caused by amplitude quantization or jitter, the error may *not* be in fact desirable, and is, in any case, not integral to granular synthesis.

5 Applications

5.1 Late-field Reverberation

Keeping in mind our impulse response above, we can model reverberation by decreasing the amplitude but increasing the delay range of later taps. As a result, streams with shorter term delays will be highly correlated with the original input signal, while those with longer delays will tend to be decorrelated. The result is an intuitive model of late-field reverberation, in which later signals are more diffuse. Two useful additions to this reverberation model include the introduction of a feedback mechanism to add IIR resonance, and varying the choice of modulation function to add control over spectral content.

5.2 Multi-channel Spatialization

Each decorrelated GS stream can also be routed to independent outputs, creating a unique type of spatialization. Merely routing copies of an input signal to multiple speakers would, owing to the precedence effect, create a phantom image at best, or collapse onto the speaker nearest the listener. Decorrelated grain streams, on the other hand, will tend to create a more complex spatial field. In conjunction with the stream-to-stream decorrelation described above for simulating late-field reverberation, rich spatial reverberators can be easily constructed from the relatively simply GS model described.

6 Conclusions

Cross-correlation measures have enormous predictive value when dealing with complex mixing operations, and in particular, when working at near-audio or audio rates. Our consideration of correlation measures above demonstrates that merely stochastically varying a parameter to a GS process may increase distortion or other effects but does not necessarily affect decorrelation. This has suggested several strategies for minimizing distortions and artifacts as well as articulated a basis for several techniques unique to GS. Additionally, because decorrelation in GS is achieved primarily through manipulation of time-delays, other parameters to the process can be fairly freely selected according to the purpose at hand.

7 Footnotes

[1] Augspurger, G., et al. 1989. Use of Stereo Synthesis to Reduce Subjective/Objective Interference Effects: The Perception of Comb Filtering, Part II. Preprint 2862, the 87th Convention of the Audio Engineering Society, October 1989, New York.

[2] Begault, Durand. 1994. 3D Sound for Virtual Reality and Multimedia. Boston: Academic Press.

[3] Kendall, Gary S. 1995. The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. Computer Music Journal, 19(4):71-87.

[4] Clifton, R.K. 1987. Breakdown of Echo Suppression in the Precedence Effect. Journal of the Acoustical Society of America 82:1834-1835.

[5] Jones, Douglas L. and Thomas W. Parks. 1988. Generation and Combination of Grains for Music Synthesis. Computer Music Journal 12(2):27-34.

[6] Truax, Barry. 1992. Discovering inner complexity: time-shifting and transposition with a real-time granulation technique. Computer Music Journal 18(2):38-48.