ARCSYN: AN EXPRESSIVE AND EFFICIENT ADDITIVE SYNTHESIS ARCHITECTURE

David Romblom

Media Arts and Technology
University of California, Santa Barbara
romblom@gmail.com

ABSTRACT

Arcsyn is an architecture and control system for expressive additive synthesis; it provides satisfying dynamic behavior, compelling transients, and non-static tones. The complexity of additive synthesis is encapsulated within a musically sensitive control system inspired by instrument physics. Timbral information is represented in the frequency domain, and can be specified parametrically or made to use popular formats such as SDIF.

1. INTRODUCTION

Arcsyn is a software synthesis technique that addresses the shortcomings of electronic instruments when compared to acoustic instruments of the classical canon. It is conventional in the fact that it provides the performer with tools to create expressive melodic content, and that it in some ways acts like an acoustic instrument. It is more modern in the fact that the timbre can be arbitrarily defined, that timbre can “morph” to other arbitrary timbres, that it can be sequenced using “analog” control signals, and that it has novel and exotic modulation capabilities. Arcsyn is meant as a general synthesis method that begins where the subtractive paradigm meets its limitations. As opposed to more exotic synthesis techniques, it is an elegant representation of complexity and uses computational resources efficiently.

Using an Analysis/Synthesis package such as SPEAR [2] one can see a number of interesting aspects in the recording of a given note. In addition to expected features such as vibrato and weaker high frequencypartials, one also observes that all partials have random motion in amplitude and frequency, highly varied attack and release times, and are generally inharmonic during attacks. When thinking of the many notes that might make up a musical phrase, we are assisted by Strawn’s [6] summary of note to note pitch transitions.

Arcsyn uses steady-state spectral information and models the transitions and temporal variation. Synthesis is implemented as a bank of bandwidth-enhanced sinusoidal oscillators [1]. The steady-state spectral information can be specified parametrically, taken from an Analysis/Synthesis package such as SPEAR or Loris, or imported from an SDIF [9] library. This allows for both acoustic instrument emulation and instrument-like electronic tones.

For a given timbre, Arcsyn uses both ff and pp steady-state spectral information. One can interpolate freely between these two layers, allowing tones to swell or decay regardless of the onset dynamic. The ability to arbitrarily define the spectral envelope of the pp and ff layers allows non-trivial evolution in playing dynamic. The architecture can be easily extended to include the additional detail of multiple dynamic layers. Spectral morphing is achieved by adding a second, entirely distinct timbre with its own dynamic information. Here again, it is easy to extended the architecture to include multiple timbres. The third dimension of interpolation is pitch; new spectral information is specified at each well-tempered note. For acoustic emulation, pitch interpolation preserves instrument formants and yields “acoustically-correct” vibrato and glissandi. For electronic timbres, we gain instrument-like playing characteristics.

2. BEYOND SUBTRACTIVE SYNTHESIS

Subtractive synthesis is a pervasive synthesis architecture, where varied harmonic content is achieved by moving the cutoff frequency and resonance of a resonant lowpass filter. There are a small set of standard waveforms, the harmonic vocabulary of each waveform is predetermined by the filter’s sweep range. This vocabulary is not sufficiently large to be musically satisfying, though non-linear distortion [3] and waveform modulation do help the situation. Samplers are a close relative of the subtractive synthesizer that operate by playing back recordings “frozen” at a given pitch and dynamic. For acoustic instrument emulation, glissandos and vibrato are modeled by varying the playback speed; this method shifts the instrument’s formant resonance up and down resulting in a very unnatural sound. A preferable method, one closer to the physics of acoustic instruments, would preserve the instrument formants while the individual partials of the tone move under them. [7]

Different recordings are played back depending on the note-on velocity. This presents problems for sustained instruments where a performer may wish to start softly and then swell; they are limited to the harmonic content of the sample they first played. To address this, modern samplers cross-fade to a second recording, though this can sound unnatural due to phase misalignment and requires detailed, case-by-case control programming. Stasis in the tone is addressed by modulating the filter cutoff or oscillator pitch with a control signal. These
methods address the sound in bulk; there is no motion within the sound.

3. ARCSYN ADDITIVE OSCILLATOR

3.1. Oscillator Architecture

ARCsyn’s additive oscillator has three control signal inputs - pitch, dynamic level, and spectral morphing. The underlying synthesis architecture is a direct reflection of these dimensions. There is a database of steady-state spectral information, with every well-tempered note having a “slice” that contains both dynamic and spectral morphing information. Each corner of the “slice” consists of N steady-state partials, where N is defined by computational requirements.

Figure 1: Pictorial representation of the database.

The tone at any given time is determined by the value of the three control signals which specify an exact location within in the database. Interpolation is performed between the eight nearest spectral data locations, as well as the previous partial values. Note that this interpolation is best done in the logarithmic domain for musical character.

3.2. “Acoustically-Correct” Vibrato and Glissandi

One substantial benefit of using the spectral database architecture is that it intrinsically models acoustic vibrato and glissandos very well. This is a central problem in conventional sample based synthesis. Wakefield [7] writes: "Although the player creates mainly frequency modulation, it results in changes in amplitude that are crucial to the perception of vibrato….. The moving harmonics are boosted and depressed according to the resonances of the instrument body.”

A time-based sampler shifts pitch by speeding up or slowing down the playback speed. All partials are shifted up or down while maintaining the same amplitude, effectively shifting the instrument resonances. Maintaining a stationary formant requires that one “switch samples” as pitch changes. The Arcsyn oscillator does this well - for acoustic instrument emulation, the formant is encoded in the spectral information for each note. Changing pitch simply changes your coordinates within the database; if the samples are done correctly the formant remains stationary. For electronic timbres, the spectral data can be designed to change (or maintain) character throughout the tessitura of the instrument.

4. CONTROL SYSTEM: MODELED MOTION

4.1. Random Individual Partial Modulation

The SPEAR software package was used to look at Cello tones in all registers of the instrument. Vibrato is clearly evident in most tracks, being subtle in the lower partials and dramatic in the highest. Aside from this, one can generally state that there is a fair amount of random amplitude and frequency motion in the partial tracks. To model this, all partials in Arcsyn are individually modulated by a stochastic control signal. The user specifies the frequency of the update and the magnitude of modulation; it is applied individually (and asynchronously) to both the frequency and amplitude of all partials. Introducing this movement greatly improves the musical quality of the tone; it ranges from a subtle “life” to granular mayhem.

4.2. Transient Modulation

Partials are not perfectly harmonic during the attack portion of a sound, but gradually find their way to integer-multiple values as the attack portion gives way to the steady-state. This phenomenon is more pronounced at higher dynamic levels. Arcsyn models this by increasing the above random frequency modulation during note onset. This achieved cleanly by introducing a new sort of control signal - the derivative of the dynamic control signal. If this signal is changing quickly, partials will be pulled farther from their harmonic tracks. The user can control the duration and amount of the inharmonicity, as well as a set amount from each new note-on command. Note that this could also be used to modulate the noise content of the bandwidth-enhanced partials.

4.3. Partial Envelopes

The dynamic control signal is fed into the Arcsyn oscillator and specifies the position between the pp and ff layers. This signal can be driven by a continuous performance interface for sustained tones, or from any form of a control envelope for struck or percussive sounds.

Every partial has its own attack and release value; they track the dynamic control signal. When the dynamic signal is greater than the partial’s current level, it slews up based on the attack setting. When the input dynamic is less than the partial’s current dynamic, it slews down based on the release value. The overall temporal envelope is defined by the combination of the dynamic
control signal and the attack/release settings of all of the partials. The partials all have individual attack/release values and their time evolution is independent of the others. This feature allows for the emulation of phenomenon observed in acoustic instruments: the low partials of a bowed string stabilized more quickly during attacks and lasted longer in decays, whereas struck instruments may have highly transient upper partials. For synthetic timbres this allows an interesting specification of instrument attack and release characteristics.

4.4. “Acoustically-Correct” Pitch Transitions

Strawn [6] summarizes note to note pitch transitions for a wide variety of instruments:

1) There is some change in amplitude: the first note dies out, after which the second note starts up. This is true even for legato playing.
2) As the first note dies out, its spectrum “falls off” (that is, the higher order spectral components disappear); as the second note enters, its spectrum is enriched. There may be spectral cues in the attack of the second note which depend on a given playing style.
3) There is a change in pitch which falls closer to the attack of the second note than to the decay of the first.
4) There is some amount of time between the two notes. The decay of the first note does not coincide with the attack of the second.

Later he states: “there is no systematic difference between ascending and descending intervals for a given pitch, dynamic, and spectral morphing are always slewed from their current value to their target value. Initially, noise was added to these signals to control denormals on floating point processors. A happy accident, this has the nice property that it adds further motion to the sound. This is a different sound than the above random partial modulation. There, modulation is applied to all partials individually. Here, it is applied to all partials as a group. Taken to the extreme, it is a novel and particularly harsh effect.

4.6. Bandwidth-Enhanced Partials

Most instruments have components of both deterministic and stochastic sound. Purely sinusoidal representations see noise as many short, highly modulated sine waves that begin and end quickly. In Analysis/Synthesis contexts, this method can result in “wormy” transformations. [1] In the context of Arcsyn, the jittery sinusoids would not fit nicely into an otherwise clean architecture. Serra and Smith, 1990, represent the stochastic noise component with a spectrally shaped noise sample. [5] Fitz later introduced the concept of the Bandwidth Enhanced Partial where each partial has both a sinusoidal and stochastic component. This is more convenient than having a separate spectrally shaped noise sample, and was used in his software package Loris. [1] For similar reasons, bandwidth-enhanced partials fit cleanly into the Arcsyn architecture, and allow for individual adjustment of every partial. Spectral information for different dynamic levels, timbres, or pitches can have wildly varying noise content.

5. THE PROTOTYPE, EVALUATION, AND FUTURE WORK

5.1. The Prototype

The first incarnation of Arcsyn is a VST instrument. The underlying synthesis engine is an oscillator bank of bandwidth-enhanced partials, all control and audio rate signals are continuously smoothed for artifact-free sound. The partial database is parameterized synthetically, though it is clear that future work will include spectral samples from libraries such as SDIF and/or synthetic timbres specified in MATLAB. For now, the parameterization uses simple mathematical equations that roughly model the physics of stringed instruments, while allowing for some electronic timbres. The user can specify two instrument formats, each defined by three variables: dB gain, frequency, and frequency width (Q.) To insure complicated dynamic behavior instrument formants apply only to the ff layer, while a second user interface control specifies how much softer the pp layer is. String stiffness is approximated by stretch-tuning of the partials.

5.2. Evaluation and Future Work

In practice, Arcsyn behaves differently than a subtractive synthesizer. Tones created with Arcsyn have internal motion; when playing a single, sustained note, the
modulation of individual partials gives a clearly perceptible “life” to the sound. Vibrato sounds natural, and dramatic pitch bends maintain an appropriate, natural-sounding character. Spectral morphing is a welcome relief from the sound of swept lowpass filters, though the effect is highly dependent upon the spectral content of the two sides.

The current parameterization of the database requires more detail. It is difficult to get dramatically different sounds using only two instrument formants. For both acoustic emulation and purely synthetic tones, it is clear that a tremendous improvement could be made with offline specification of the databases. For acoustic emulation, this would be the steady-state spectral information of acoustic instruments generated from Analysis/Synthesis packages or pulled from SDIF libraries. Exotic electronic timbres could be specified in MATLAB. One can speculate that the increased detail in the steady-state spectral information would be directly proportional to the interest and nuance in the tone. An additional benefit would be the simplification of the user interface; instrument specification would then be a drop-down menu of pre-prepared databases.

There is no particular reason to limit the instrument to only two dynamic levels, or two spectral morphing sides. Given enough computational resources, Arcsyn can have tremendously detailed dynamic behavior, and spectrally morph between multitudes of timbres.

With improved waveform specification, the modeling of transients would likely require further detail. With further refinement, the resulting synthesizer could have the detailed tone of a sampler with playability approaching that of an acoustic instrument.

6. SUMMARY

The Arcsyn architecture allows for interesting dynamic response by specifying disparate spectral envelopes for the pp and ff dynamic layers. Interpolation in pitch can preserve an instrument’s formant structure, and as a result has pleasant properties for vibrato and glissandi. Spectral morphing brings one of the highlights of Analysis/Synthesis to the world of the performance synthesizer.

Frequency-domain timbral information can be specified parametrically or made to use popular formats such as SDIF. The possibility of spectral morphing between acoustic and electronic timbres is musically interesting. The control system mimics phenomenon observed in acoustic instruments. Transients are modeled by increasing inharmonicity using the derivative of the dynamic control signal; the same signal can also be used to add noise to the bandwidth-enhanced partials. All partials have their own attack and release values and track the dynamic control signal. Non-static tones are achieved by modulating the amplitude and frequency of every partial individually, and note-to-note transitions follow the observations laid out by Strawn [6]. Further motion in the tone is achieved by modulating all control signals.

7. ACKNOWLEDGMENTS

This work was done at the University of California, Santa Barbara, in the Media Arts and Technology program. I would like to thank Stephen Travis Pope, Curtis Roads, and my fellow students for their feedback in this project. I would also like to thank my colleagues at the Sennheiser DSP Research Lab for their more recent input, as well as their particularly bad jokes.

8. REFERENCES


