SPATIO-OPERATIONAL SPECTRAL (S.O.S.) SYNTHESIS

David Topper[1], Matthew Burtner[1], Stefania Serafin[2]

VCCM[1], McIntire Department of Music, University of Virginia

CCRMA[2], Department of Music, Stanford University

email: topper@virginia.edu, mburtner@virginia.edu, serafin@ccrma.stanford.edu

Abstract

We propose an approach to digital audio effects using recombinant spatialization for signal processing. This technique, which we call Spatio-Operational Spectral Synthesis (SOS), relies on recent theories of auditory perception. The perceptual spatial phenomenon of objecthood is explored as an expressive musical tool.

1 Introduction

Spatial techniques in music composition have been in use since the 16th century [8]. These techniques, including the more recent practices of electroacoustic music, have relied on the projection of an audio object within a defined space.

Spatio-Operational Spectral Synthesis or SOS, is a signal processing technique based on recent psychoacoustic research. The literature on auditory perception offers many clues to the psychoperceptual interpretation of audio objecthood as a result of streaming theory [4]. Streaming describes audio objects as sequences displaying internal consistency or continuity [5]. Bregman has further defined a stream as, "a computational stage on the way to the full description of an auditory event. The stream serves the purpose of clustering related qualities ([1] p10)." Thus it becomes the primary defining factor of an acoustic object.

SOS breaks apart an existing algorithm (ie, Additive Synthesis, Physical Modeling Synthesis, etc.) into salient spectral components, with different components being routed to individual or groups of channels in a multichannel environment. Due to the inherent limitations of audition, the listener cannot readily decode the location of specific spectra, and at the same time can perceive the assembled signal. In this sense, the nature of the auditory object is altered by situating it on the threshold of streaming, between unity and multiplicity.

The "Theory of Indispensable Attributes" (TIA) proposed by Michael Kubovy [5] puts forth a framework for evaluating the most critical data the mind uses to process and identify objects. In the case of audio objects, TIA holds that pitch is an indispensable attribute of sound while location is not, simply put, because the perception of audio objects can not exist without pitch. His experiments have demonstrated that pitch is a descriminating factor the brain seems to use in distinguishing sonic objecthood, whereas space is not as critical.

Bregman notes that conditions can be altered to make localization easier or more difficult, so that, "conflicting cues can vote on the grouping of acoustic components and that the assessed spatial location gets a vote with the other cues. ([1] p305)": " Curious about how Kubovy's and Bregman's theories could be utilized for signal processing, we began applying spatial processing algorithms to spectral objects.

When spectral parameters are spatialized in a certain manner the components fuse and it is impossible to localize the sound, yet when they are spatialized differently the localization or movement is predominant over any type of spectral fusion. Creatively modulating between fusion and separation is where SOS comes into being. One of our main questions is this: if the mind does not treat location as indespensible, can SOS force the signal into an oscillation between unity and multiplicity by exploiting spatialization of the frequency domain?

The technique exploits what might be called a "Persistence of Audition" insofar as the listener is aware that auditory objects are moving, but not always completely aware of where or how. This level of spatial perception on the part of the listener can also be controlled by the composer with specific parameters.

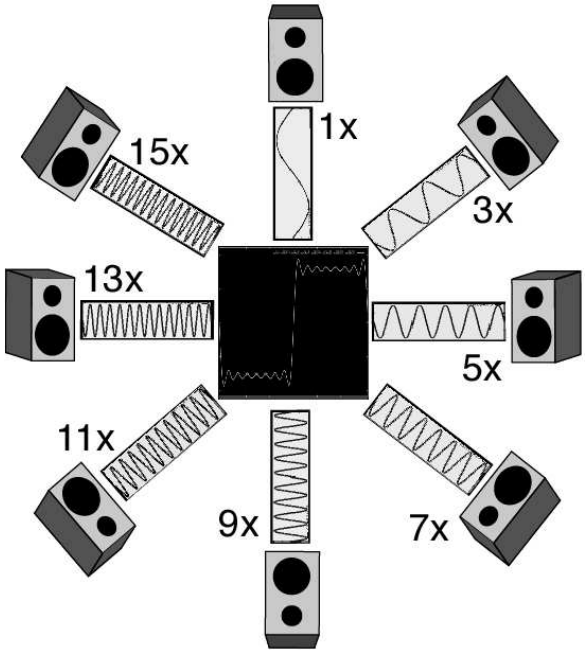

Figure 1. SOS Recombinant Principle

SOS is essentially a two-step operation. Step one consists of taking an existing synthesis algorithm and breaking it apart into logical components. Step two re-assembles the individual components generated in the previous step by applying various spatialization algorithms. Figure 1 illustrates the basic notion of SOS as demonstrated in the following example of a square wave.

2 Initial Examples

In initial experiments testing SOS we used simple mathematical audio objects such as a square wave generated by summing together sinusoids having odd harmonics and inversely proportional amplitudes.

Formula (1) describes the basic formula used in this initial example:

$$x_{square}(t) = \sin(w_0 t) + 1/3 \sin(3w_0 t) + 1/5 \sin(5w_0 t) \ldots \quad (1)$$

In this experiment the first eight sine components of the additive synthesis square wave model were separated out and assigned to a specific speaker in an eight-channel speaker array. Although the square wave is spatially separated, summation of the complex object is accomplished by the mind of the listener (Figure 1).

Separation need not be completely discrete however. Any number of sinusoids can be used and animated in the space, sharing speakers. In a simple extension of this example sinusoids were used to generate a sawtooth wave as shown in Formula (2).

$$x_{saw}(t) = \sin(w_0 t) + 1/2 \sin(2w_0 t) + 1/3 \sin(3w_0 t) \ldots \quad (2)$$

When the sinusoids were played statically, in separate speakers, the ear can identify the weighting of the frequency spectrum between different speakers. For example, if the fundamental is placed directly in front of the listener and each subsequent partial is placed in the next speaker clockwise around the array, a slight weighting occurs in the right front of the array. The First Wavefront law would of course suggest this, but in actuality the blending of the sinusoids into a square wave is more perceptible than the sense of separation into components. In fact, the effect is so subtle that a less well-trained ear still hears a completely synthesized square wave when listening from the center of the space.

Animating each of the sinusoids in a consistent manner exhibits a first example of the SOS effect. By assigning each harmonic a circular path, delayed by one speaker location in relation to each preceding harmonic, the unity of the square wave was maintained but each partial also began to exhibit a separate identity. This of course is the result, in part, of phase and shifting (eg., circularly moving) amplitude weights. The mind of the listener, tries to fuse the components while also attempting to follow individual movement.

This simple example illustrates how the Precedence Effect can be confused so that the mind simultaneosly can cast conflicting cognitive votes for oneness and multiplicity in the frequency domain. This state of ambiguity, as a result of spatial modulation, is what we call the SOS effect.

We experimented with different rates of circular modulation of each sine component. Interestingly, each relationship was different but not necessarily more pronounced than the similar, delayed motion. Using the same, non-time-varying signal, a time-varying frequency effect can be achieved due to spatial modulation using only circular paths in the same direction. Figure 2 illustrates this type of movement.

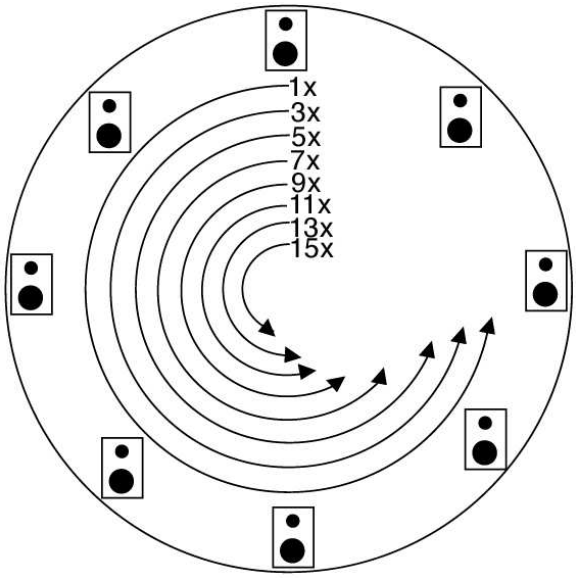

Figure 2. SOS with varying rate circular spatial path of the first eight partials of a square wave

An early example of spectral separation of this sort has been implemented in Roger Reynolds' composition, Archepelago (1983) for orchestra and electronics ([1] p296). In tests done at the IRCAM, Reynolds and Thiery Lancino divided the spectrum of an oboe between two speakers and added slight frequency modulation to each channel. If the FM were the same in both channels the sound synthesized, but if different FM were added to each channel, the sounds divided into two independent auditory objects.

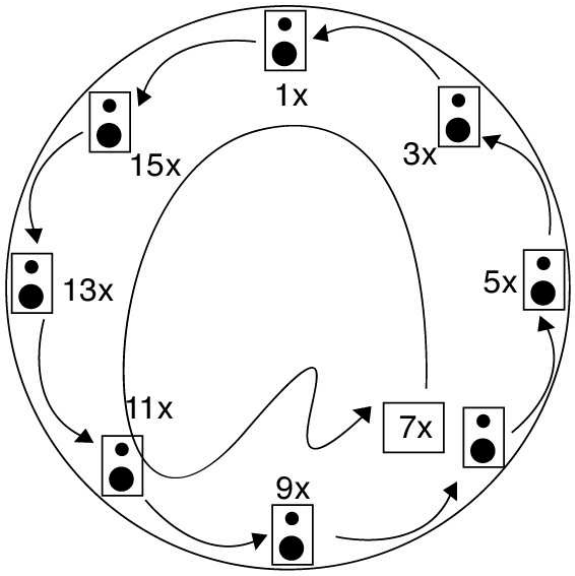

Figure 3. SOS with one partial moving against the others moving in a unified circular motion.

In our later tests, we noticed similar results to Reynolds and Lancino, even within the context of animated partials. By exaggerating the movement of one partial, either by increasing its rate of revolution, or assigning it a different path, the partial in question stood out and the SOS effect was somewhat reduced. By varying the amount of oscillation and specific paths of different partials, the SOS effect can be changed subtly.

3. Definitions of Spatial Archetypes for SOS

Any number of spatialization algorithms can be applied to the separated components' variables or audio stream. The types of spatialization employed by SOS can be thought of as having two attributes: motion and quality. A series of archetypal quality attributes were explored in a two dimensional environment.

Motion was divided into three categories:

1) static: no motion
2) smooth: a smooth transition between points
3) cut: a broken transition between points

Quality was divided into five archetypical forms:

1) circle: an object defines a circular pattern
2) jitter: an object wobbles around a point
3) across: an object moves between two speakers
4) spread: an object splits and spreads from one point to many points
5) random: an object jumps around the space between randomly varying points

These archetypes can be applied globally, to groups, or to individual channels. Each archetype has specific variables that can be used to emphasize or de-emphasize the SOS effect. Variables can also be mapped to trajectory or rate of change, defined by a time-varying function, or generated gesturally in real time.

4. Extended Examples

The following examples illustrate several different applications of SOS, describing how the experiments were conducted.

4.1 SOS processing using filter subband decomposition

The balance between frequency separation and sonic object animation became much more complicated when we attempted to apply our initial technique to an audio signal. Our initial tests assigned eight simple two pole IIR filter outputs to discrete speaker locations. Selection of the ration between the filters became a critical component in being able to achieve any effect at all. With filters set to frequencies that were not very strong in the underlying signal, the filters tended to blend together and sound as if some type of combined filtering were taking place rather than SOS. Similarly, when spatialization algorithms were applied with an improper filter weight, the underlying movement was more apparent than the separation.

We tested the filter technique with both white noise and live instrument (eg., Tenor Saxophone). The former of course offered much more flexibility with respect to frequency range and filter setup. The saxophone signal used, having the majority of its spectrum located between 150Hz and 1500Hz (with significant spectral energy up to approximately 8000Hz) suggested a filter/bandwidth weighting of: 32/5Hz, 65/15Hz 130/30Hz, 260/60Hz, 520/120Hz, 1000/240Hz, 2000/500Hz, 4000/1000Hz.

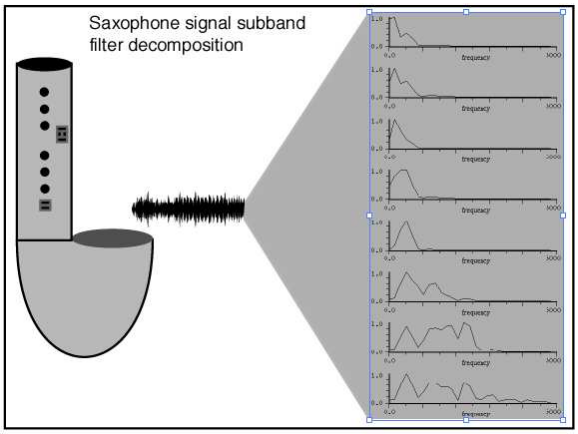

Figure 4: Saxophone signal subband filter decomposition for SOS.

4.2 SOS Processing of Physical Models

A more complicated example of SOS involves separating the modes or filter output of a physical model and applying individual spatial processing to each component.

Tests were done with a bowed string algorithm [10] in which bow friction was separated from the string sound. The second involved a physical model singing bowl [11] with the modes divided into different audio streams.

4.2.1 Bowed String Physical Model Parameter Separation

In the first experiment with physical models, we separated thefriction and the velocity waveform of a bowed string as shown in figure 5.

Digital waveguide models of bowed strings calculate the frictional force at the bow point by solving the couplingbetween the bow and the string. Once this coupling is solved, the outgoing waves propagating toward the bridge can becalculated as: vob = vin + f*Y/2, where Y is the admittance of the string, f is the frictional force and vob are the outgoingvelocity toward the bridge and incoming velocity from the nut respectively.

The output velocity at the bridge, vob, is the one that, given an appropriate combination of parameters, allows to obtain the so-called Helmholtz

motion, i.e. the ideal motion of a bowed string. In our SOS example, we are interested inseparating vob into its two components, i.e. the friction force and the incoming velocity from the nut.

The friction force f, scaled by the admittance factor, and the incoming nut velocity are sent to two differentchannels, as figure 5 shows.

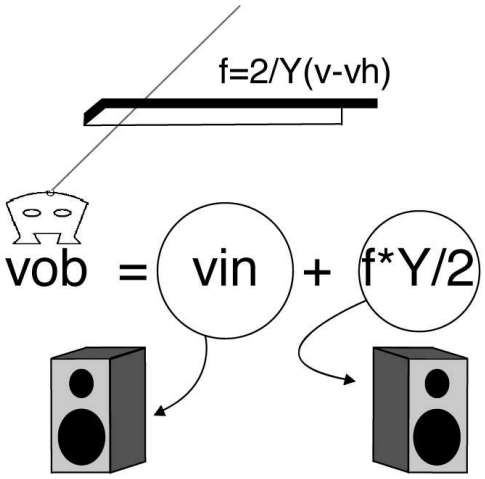

Figure 5: Bowing friction and velocity separated intodifferent channels.

By placing the components in different speakers, the two were easily identified as separate objects. Played through the same speaker however, they were fused into a single object.

Because the underlying model is one of an instrument with a great degree of gestural control simply changing a few parameters and routing them through an SOS spatialization algorithm is generally not a believable way to control the string model. As has been shown in earlier work [3,4] the bowed string physical model benefits greatly from careful controller interaction including haptics and detailed multi-parametric control. In the experiments we conducted, the components became distinct too easily to give satisfyingresults. The use of a gestural controller such as the Peavy PC1600x multislider improved the results due to the ability tocreate more interesting and differentiated control parameters.

4.2.2 Singing Bowl Physical Model Modal Separation

The physical model of the singing bowl proved to be an idiomatic instrument for SOS processing. The bowl model allows each of eight resonant modes to be controlled independently by user input, and processed separately on output. We explored possibilities of spatial processing of the modes of the bowl as an application of SOS.

The bowl was first played back with each mode of the system routed to a different speaker. Even without any spatial processing outside of separation, the emission of the bowl as a multi-modal spatialization algorithm gives good results. As different modes of the bowl changed according to the characteristics of the equation, the listener had an almost impossible time discerning between the "complete bowl" and the individual components.

The Max/MSP implementation of the singing bowl model offers 32 separate input controls. In the examples, changing several of the parameters allowed for an even greater expressive control. When any level of control was applied to individual parameters of the bowl, the SOS effect was enhanced. Simply applying amplitude modulation to independent channels also augmented the effect.

A strong sense of "interiority" results from the spatialized bowl. It is unique in our examples in creating a sense of "place," or a notion of "body" enveloping the listener. This example has been discussed in greater detail by the authors [2].

5. Implementation

SOS has been implemented both in MAX/MSP and RTcmix [7] on both Mac and PC/Linux hardware. The Linux implementations utilized the PAWN and SPAWN systems [9]. Figure 6 illustrates the SOS Control Interface in Max/MSP, allowing real time, prerecorded or graphic control over eight independent channels.

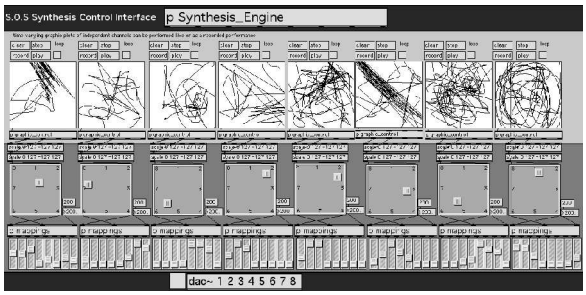

Figure 6. SOS control interface in Max/MSP.

6. Future Directions

Current SOS research has been done primarily in a two dimensional environment. Exploring a three dimensional environment will increase the effect of spatialization algorithms and offer a greater means of separation for various models (ie, 3D waveguides).

So far, only the authors who agreed on the results have performed listening tests. Future work consists of testing more subjects, in order to see if the segregation of the synthesis algorithms is performed in the same way by human listeners.

Much of the psychoacoustic research that inspired SOS also looks at the related phenomenon of audio streaming, in sequential segregation. In addition to exploring SOS based on "spectral" separation, it would be interesting to explore sequential stream separation and granular synthesis.

References

[1] A. S. Bregman. *Auditory Scene Analysis: the perceptual organization of sound.* MIT Press, Cambridge, MA, 1999.

[2] M. Burtner, S. Serafin, and D. Topper. "Real-time spatial processing and transformations of a singing bowl." *Proceedings of DAFX (Digital Audio Effects Conference),* Hamburg, Germany. 2002

[3] M. Burtner, and S. Serafin. "The Exbow Metasax: compositional applications of bowed string physical models using instrument controller substitution." *Journal of New Music Research,* vol. 22 num. 5. Swets & Zeitlinger, Lisse, The Netherlands. 2002.

[4] M. Burtner, S. Modrian, C. Nichols, and S. Serafin. "Expressive controllers for string physical models."

[5] M. Kubovy, D. V. Valkenburg. "Auditory and Visual Objects*," Cognition.* 80, p97-126. 2001.

[6] S. McAdams, and A. Bregman. "Hearing Musical Streams." *Computer Music Journal.* vol. 3 num. 4. CA., 1979.

[7] B. Garton, and D. Topper. "RTcmix -- Using CMIX in Real Time," Proc. of *International Computer Music Conference* (ICMC), Thesalonika, Greece, 1997.

[8] C. Roads. *The Computer Music Tutorial.* 1996, MIT Press, Cambridge, MA, 1996.

[9] D. Topper. "PAWN and SPAWN (Portable and Semi Portable Audio Workstation)." Proc. of *International Computer Music Conference* (ICMC), Berlin, Germany., 2001.

[10] S. Serafin, J.O. Smith III, and J. Woodhouse. "An investigation of the impact of torsion waves and friction characteristics on the playability of virtual bowed strings." IEEE Workshop on signal Processing to Audio and Acoustics. New Paltz, NY, 1999.

[11] S. Serafin, J.O. Smith III, and C. Wilkerson. "Modeling Bowl Resonators Using Digital Waveguide Networks." Proc. of DAFX (Digital Audio Effects Conference, Hamburg, Germany. 2002.