STUDIES ON THE COMPOSITIONAL USE OF SPACE

Ircam Composer in Research Report 2012

Rama Gottfried

Composer
University of California, Berkeley
Center for New Music and Audio Technologies (CNMAT)
rama.gottfried@gmail.com

ABSTRACT

This report describes the results and reflections from my residency with Ircam's *Acoustic and Cognitive Spaces* research team during the year 2012. Central aspects of investigation were: the use of spatial projection system as an extension of electro-acoustic instrument design, the perceptual spatial identity of a source, the creation of a single object/body by a complex of sources, the interaction between a spatial sound object and its real or virtual surrounding space, and the empirical study of the application of these spatial organizations. Resulting from the study is a theoretical model in which the perceptual relationship between spatial organization and timbral/motivic content is defined, and integrated into 'spatio-timbral' objects. Several 'instrumental' approaches to working with Wave Field Synthesis (WFS) and Higher-Order Ambisonic (HOA) rendering systems are presented, and implemented in the étude *Fluoresce,* composed for the inauguration of the new WFS and HOA system in Ircam's Espace de Projection. In conclusion, future developments for the artistic manipulation of sound and space are proposed.

1. PROJECT BACKGROUND

As high resolution spatial audio rendering techniques are becoming increasingly refined and flexible [1, 2, 3, 4, 5], it is becoming increasingly important for composers and sound artists to develop critically aesthetic approaches to these techniques. As with any instrument, control and eventually virtuosity is accumulated though practice, repetition, trial and error. Likewise, as in any historical period, artists must fully know and control their material if there is to be any hope for meaningful aesthetic work. From the basis of this understanding, the proposal for my residency began with the idea that the spatial rendering processes of WFS and HOA could be thought of as fundamentally instrumental techniques, with their own unique characteristic possibilities, limitations and processual symbolic meanings. Understood as such, the areas of focus for my residency were to analyse the rendering systems from their inner computational mechanics to the perceptual experience of listening in physical space; to empirically test ideas of spatiality based on hypothetical formal designs and to develop new ideas of spatial form through the careful study of the experimental results. The final goal of my proposed study was to develop a perceptually relevant approach to spatial form as a primary element of composition.

In preparation for my period of artistic research at Ircam, I spent six months working with the 12 and 120 channel spherical loudspeaker arrays [18] at the Center for New Music and Audio Technologies (CNMAT), performing a series of empirical tests of compositional ideas (e.g. the perceptual limits of spatial cognition at various sonic grain durations, the effects of dynamic spherical harmonic manipulation, etc.). During this preparation period I also designed a method of notation using hierarchical labelling of vector objects in Adobe Illustrator, which could be exported as SVG format and read in MaxMSP as control data.

In addition to the invitation from Ircam to be part of the 2012 Musical Research Residency, I was the recipient of the University of California, Berkeley's 'George Ladd *Prix de Paris* Award for Compositional Excellence,' which facilitated my stay in Paris for the year 2012.

2. RESEARCH TOPICS

With the integration of spatial information as an active element of composition as a primary goal, I began the organization of research using a conceptual model based on instrument design. Similar to an acoustic instrument physically performed in a space, the perceived spatial identity of a virtual source can be characterized by its projective aspects: its morphology, location, and orientation; and by its reflective interaction with the surrounding space. These generalized active and reactive aspects of sound in space, and microscopically the instrumental aspects of activation and resonance, served as the foundation of my study, described in the following sections. Through the course of the residency I worked to develop various approaches to the design and realizations of these aspects of spatial identity, and to find expressively piquant elements to be manipulated through composition.

3. TESTS AND RESULTS

The experimental process was organised as a series of steps in which I first developed ideas based on a general topic and constructed prototype patches in MaxMSP. This was followed by a two step testing period, first in the studio and then in Ircam's Espace de Projection (Espro). At the conclusion of each testing period, I empirically reflected on the results, and used this as a basis for the next round of prototyping and testing. In this way, the ideas, tests and results described below where influenced and informed in chronological incrementation. Ultimately, the combined approaches were refined, and first implemented in the étude *Fluoresce* composed for the inauguration of the HOA/WFS system.

3.1. Directivity

The first series of tests focused on approaches to the use of directivity with the WFS system. Since natural directivity patterns vary with time and frequency [6, 7, 8], one early approach was to break a mono signal source into several frequency regions, each having a time varying directivity pattern whose shape could be controlled through streaming spherical harmonic weights, sent to the WFS rendering engine via OSC messages. There were several issues with this first approach, technically and perceptually.

Technically, there were issues with the SonicEmotion WFS system dealing with a high rate of OSC data flow. On the one hand, the SonicEmotion system that was used for directivity tests is capable of very CPU efficient manipulation of source directivity, however, after a few minutes of rapid flow of spherical harmonic coefficients over the network, the WFS server would become unresponsive, and then automatically restart itself, making the dynamic manipulation of radiation patterns unworkable.

In response to this message overload issue with the SonicEmotion system, an alternative approach to time varying directivity was tested. Instead of the direct approach where each frequency region has its own dynamic radiation pattern, the available 16 channels of rendering with the SonicEmotion system were treated as sequential points within the temporal variation of the pattern. A single sound source was sent through a matrix in MaxMSP and then crossfaded across the 16 directivity patterns. However, compared to natural radiation this is a significantly reduced resolution in both frequency and time domains, since the 16 channels were in static patterns and were used to control the directivity pattern of the full frequency range of the source, rather than varying with frequency as in natural sources. Time varying the radiation patterns for the full spectrum of the source was primarily perceived as a general filtration of the sound, similar to the Spat perceptual model of directivity, but was found too subtle to be a primary compositional element.

Perceptually, using a stacked set of frequency varying, temporally static directivity patterns, does create a sense of complexity and liveliness of the source within the real space. However, in the tests made with this approach (at Ircam and with CNMAT's spherical array), the result always sounds like a reproduction, rather than a naturally radiating source. This is due to the simple fact that the samples used for testing were either close mic recordings,

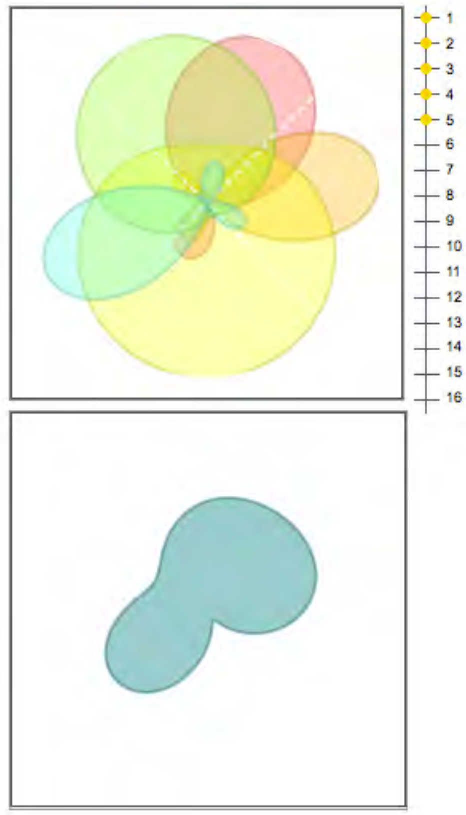


Figure 1. Sixteen channels stacked in one location, each with an idependent directivity pattern, was used for testing frequency dependent radiation patterns with the WFS system.

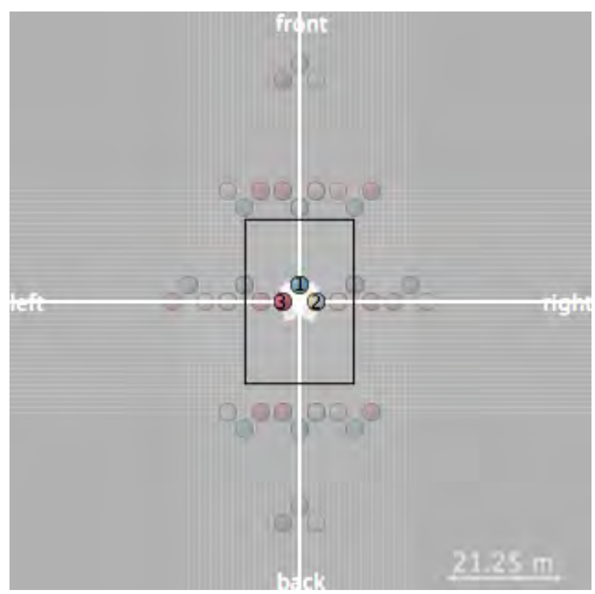


Figure 2. Shoebox / Image-source reflection model implemented in Ircam Spat.

or synthetic sounds, which tend to reduce the illusion of a 'real' source in the room.

While the perceptual and aesthetic results of the crossfading directivity pattern tests were in my opinion unsuccessful — or perhaps merely inconclusive regarding the perceived presence of the virtual source and its interaction with the real space — in the process of testing, an unusual synthesis effect was discovered.

Using a MaxMSP [phasor~][1] to control the rate of triangular crossfades between 16 stacked directivity patterns, I conducted an experiment using this same [phasor~] as the phase of an oscillator [cos~], and used this as the source material for the directivity patterns. Unexpectedly, the characteristic of each directivity pattern acted as a kind of spatial phase filter of the original oscillator. Since each channel was phase synchronized to a fixed portion of the waveform, changes to the directivity pattern of a given layer caused a filtering of the waveform at the corresponding phase region. The result was an altering of the waveform, simultaneously spatialized based on that phase region/channel's directivity pattern, and throwing different angles of the oscillation in different angles around the room.

There are several other possible approaches to directivity which I have not yet tested but which could conceivably become compositionally interesting. One approach could be to use offline rendering based on high resolution directivity analyses of acoustic instruments applied to real and synthetic sources, rendering offline, which would allow much more control over the time and frequency varying aspects which are difficult to achieve in realtime with high resolution. Another approach which could be interesting to develop further compositionally is a simplified wind instrument model, where each fingered pitch (or other rubric for discrete segmentation) would be assigned to a unique directivity pattern, or set of time/frequency varying patterns. This approach has been studied, but has not been explored compositionally in depth.

3.2. Room / Reflection

In complement to the 'active' principles of sound radiation and propagation, the 'reactive' principle of geometric reflection and diffusion is a potentially fertile parameter for compositional manipulation. By rendering the characteristic reflection patterns of a room as a spatial model, it is possible to alter the perceived context of the source within a geometric model. This becomes particularly interesting with high resolution systems such as WFS and HOA.

As a first step towards the compositional integration of room geometry, we began by using the [spat.shoebox] object for calculating an image-source model for the early reflections of a typical rectangular room. The initial implementation was in Studio 5, using the 64 speaker WFS array for the direct source, and a six channel surrounding system for rendering the early reflections and late reverberation using VBAP. This did give the sources a bit of extra spatial context, but was in 2D and a very low resolution.

Later, the shoebox / image-source model approach was refined, using as reference Markus Noisternig and Thomas Musil's 2003 paper, '3D Binaural Sound Reproduction using a Virtual Ambisonic Approach'[9]. Using the block diagrams outlined in the paper, and referring to the accompanying Pd patch, the approach was ported to Spat and implemented using a poly~ based approach, and is now included in the Spat tutorials (tutorial 19, 'Image Sources Shoebox' in Spat version 4.5.5).

In the [poly~] implementation, the direct source is rendered at the highest HOA order possible for maximum spatial clarity and CPU load[2], with second order image-source reflections rendered at a lower HOA order. After some testing, attempting to further reduce CPU cost (due to the large number of sources in motion: 24 reflections for each direct source, with 25 sources per Max instance), I found that it was significantly less full sounding using only the first order reflections, and that it was preferable to keep the second order reflections, at the expense of reducing the HOA order

I found that perceptually the image-source model approach creates very natural sounding early reflection pattern, since they are physically correct, however there is still some work to be done in the current implementation to improve the texture of the late reverb, which tends to have a slightly audible oscillation in its decay pattern. There are several reasons for this. Since the number of delayed signals in a second order image source model is still not very high compared to a real room, the input to [spat.reverb~]'s feedback delay network is possibly more correlated than would be ideal. There is currently in place a normalized random diffusion matrix distributing the image sources across the input channels of the late reverb bus, however this random distribution may not be the most aesthetically pleasing technique. The [spat.reverb~] object is generally used in conjunction with [spat.cluster~] module as its input and which contains further de-correlation. Potentially, by including a cluster stage between the image-source bus and the late reverb stage, the quality of the late reverb might be improved. The late reverb in Noisternig and Musil's Pd implementation seems somewhat smoother than the Spat/Max version. While both are based on the developments by Jean-Marc Jot, internally there may be some slight differences concerning the unitary matrix included in the feedback delay network.

In summary, I found that the image-source model perceivably adds an extra reenforcement to a source's

1 MaxMSP objects are notated here with [square brackets].

2 7th order HOA was used, which is not the maximum for the Espro system, but was adjusted for the CPU load of the patch.

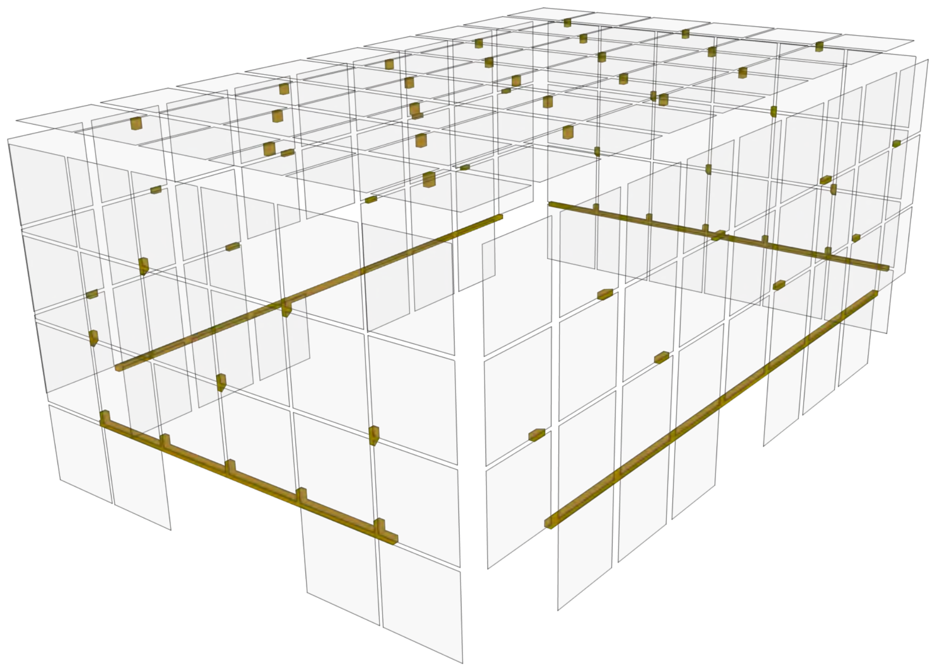


Figure 3. A 3D rendering of Ircam's Espace de Projection, by Markus Noisternig, showing the installation of WFS and HOA speaker arrays.

contextual placement in a geometric space. Using the rectangular shoebox model for image sources is an efficient, practical solution for a preliminary compositional integration of room geometry. What remains to be extensively developed is the compositional use of dynamic room shapes rendered in a high order 3D system. In particular, the challenge (for myself at least) is to know *which elements* of the model can be altered for maximum effect. Ideally, in the future there will be improved compositional / notation tools and strategies for describing geometrically and acoustically three dimensional environments as composed, responsive aspects of the sonic scene.

3.3. Instrument Design

Taken as a found object, an abstracted spatial rendering system can be observed as a process of computation, the mechanics of which an artist's fingers may sculpt at any variable. Each control point in a system has its own chain of events which are affected by its variation — or described in another way, the nuance of each element in the computational process can be traced from the resulting whole, just as an instrumental tone is the result of many physical operations.

Using this computational-instrumental parallel as an entry point, I began development on the Espro rendering systems by deconstructing the WFS and HOA algorithms, and worked to develop unique 'instrumental techniques' which could only be possible with these systems.

3.3.1.System Specific Techniques: WFS

WFS is rendered using tightly packed arrays of loudspeakers, with each rendered source having a delay and gain for each loudspeaker, based on the source's virtual location [7] (see figure 4). Taking this as a starting point, I was able to use Thibaut Carpentier's new [wfs.wfs] object (non-tilda) to extract the list of delays and gains used for 'correct' rendering of a WFS focused source (i.e. a synthesized sound source in front of the array), and make irrational changes to the values, blurring and upsetting the system. By suddenly introducing the original unaltered delay and gain values, the temporal blurring and flattening of the sounds against the speakers could transition to the sound jumping off of the speakers to the interior of the space, in what I have been calling 'focus delay.' Since the WFS system is composed of long flat arrays of loudspeakers, this 'focus delay' technique blurs the boundaries between the virtual and real by bringing the listener's awareness to the physicality of the speakers themselves, and then suddenly coalescing into a virtual physicality radiating from within the speaker array (provided the listener is within the effective region[7]).

3.3.2.System Specific Techniques: HOA

A unique element of HOA rendering is the use of spherical harmonics to represent the distribution of acoustic energy on the surface of a sphere. This representation of spatial form is essentially a Fourier transform in the spatial domain, and as such, provides a context to apply spatial-frequency domain transformations.

My first instrumental test with HOA was to attempt manipulation directly in the spherical harmonic domain, using a point source encoded to HOA with [spat.pan~] and processed with spherical harmonic domain feedback delay (figure 5a). This is conceptually very potent, however the results tended to blur towards a predominantly omnidirectional distribution. The second prototype was an interpolation between hand coded spherical harmonic

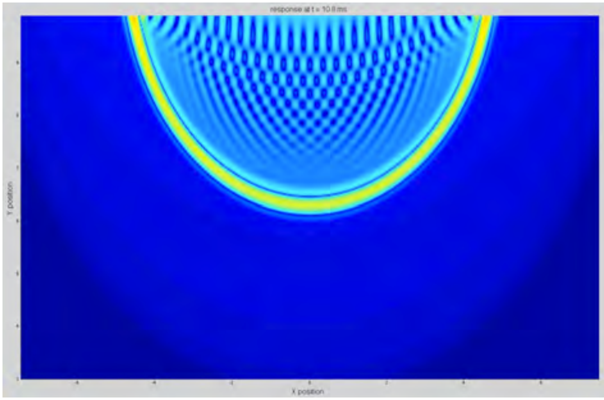


Figure 4a. A point source located behind the array

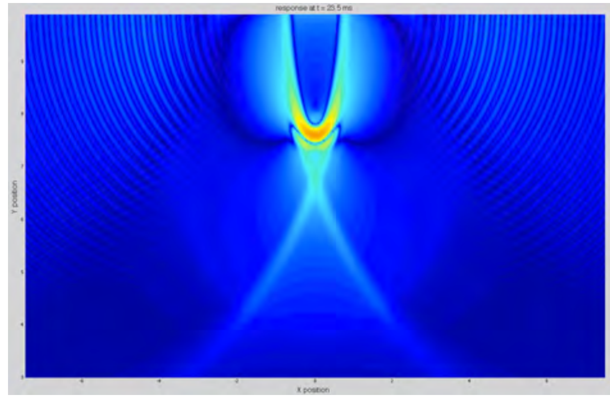


Figure 4b. A focused source located in front of the the array, with a narrow radiation pattern. Note the triangular region of linear phase.

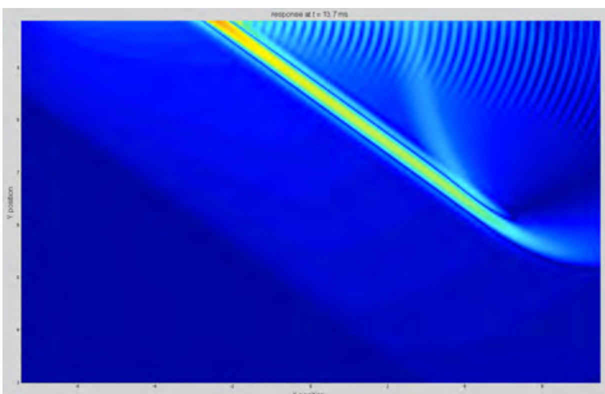


Figure 4c. A plane wave with an angled orientation

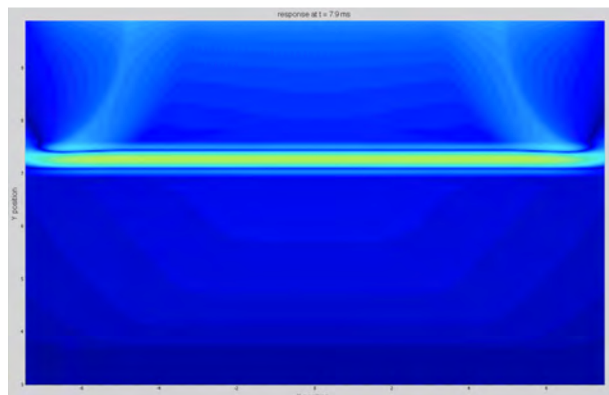


Figure 4d. A plane wave with a flat orientation

Figure 4. Wave Field Synthesis sound field propagation, generated by calculating the delay and gain for each speaker in a linear array, such that the listener perceives a natural sound field radiation pattern. This technique was deconstructed through the course of the study, and used to create effects of blur and focus at various proximities. WFS energy distribution illustrations by Markus Noisternig.

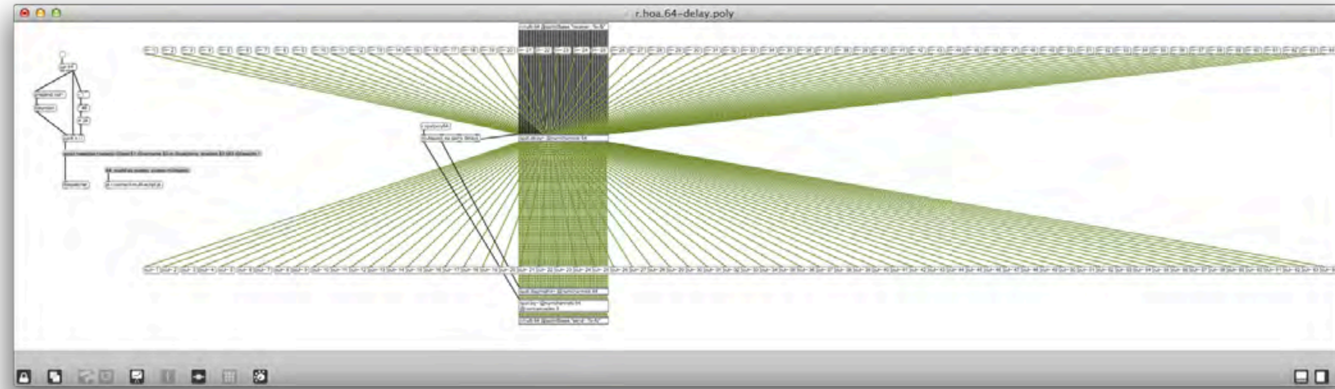


Figure 5a. Poly~ voice for spherical harmonic feedback delay

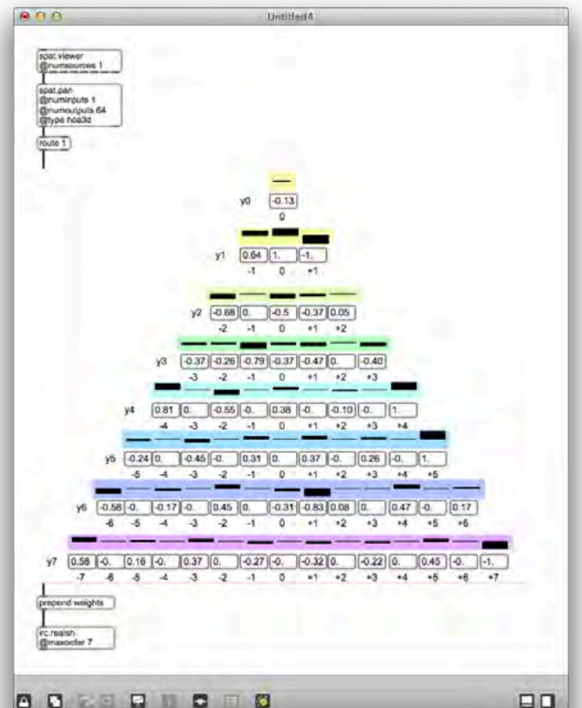


Figure 5b. Manual spherical harmonic manipulation interface.

Figure 5. Approaches investigated for designing spherical harmonic distributions for Higher Order Ambisonic rendering. After a period of investigation of direct manipulation of spherical harmonic coefficents (figure 5a and 5b), a higher-level approach was adopted, and deemed more perceptually viable (figures 5c and 5d).

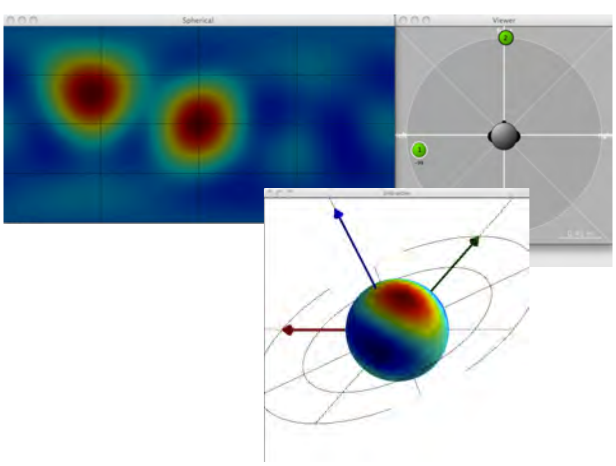


Figure 5c. Spherical harmonic design through the organization of point sources with Ircam Spat. Interfaces seen above are: irc.realshmap (top left), irc.realsh (bottom), and spat.viewer (top right).

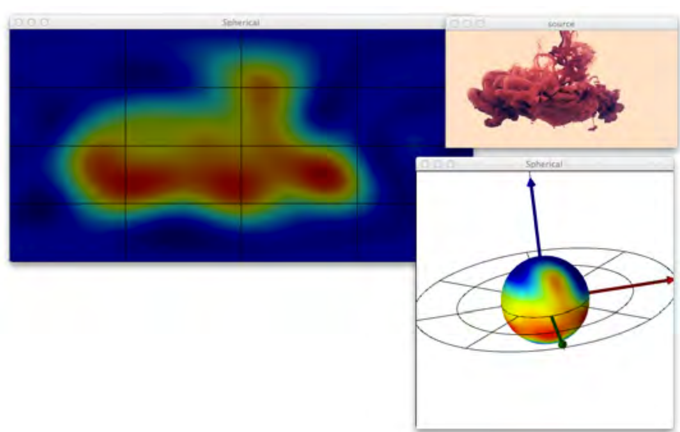


Figure 5d. Spherical harmonic design through the decomposition from Jitter pixel matrix to spherical harmonics, though irc.realshdecomposition developed during my residency. Interfaces seen above: irc.realshmap (top left), irc.realsh (bottom), and jit.window (top right).

coefficients in a triangular interface (figure 5b). Similar to the feedback delay technique, the result was a predominately omnidirectional blurring, and was difficult to maintain a sense of spatial form within the processing.

So far, I have found the most HOA unique, and perceptually successful spherical domain manipulations to be with very low order spherical harmonics. In particular, the HOA transformation tools developed by Noisternig and Carpentier for viewing and manipulating spherical harmonics ([spat.hoatransform], and [spat.hoatransform~]), are very efficient in their perceptual results. These tools allow the user to manipulate the weights of the spherical harmonic orders, which perceptually alters the sense of a source's physical volume and spatial complexity, and provide an interface for perceptual control of matrix rotations of the HOA coefficients (i.e. yaw, pitch, roll, etc.). These matrix transformations are very perceptually effective, and I would be interested to further investigate possibilities working with complex scenes and these transformations for various layers of spatial formation.

Working with the team, we developed additional new tools for drawing spatial energy distributions by hand or via generative video techniques, representing the resulting energy distribution as a pixel map which could then be decomposed into spherical harmonics and rendered via HOA (see figure 5). These tools aimed to develop methods for breaking out of the point based understanding of space into the more sculpturally flexible (see help patches for [irc.realshmap] and [irc.realshdecomposition] objects). As with the above spherical harmonic domain manipulations, there is a tendency towards omnidirectional blur, however the decomposition tool is potentially a very powerful tool for sonic sculpture.

In the future, I believe there is still a possibility to use the feedback delay approach in a more perceptually coherent manner, for example if each feedback delay became less and less focused, or the reverse, an 'HOA focus delay,' where the delay begins as omnidirectional (maybe at a reduced amplitude) and becomes increasingly focused.

3.3.3.Spatial multi-instrumental envelopes

To explore the idea of spatial rendering as an extension of musical instrument, it is helpful to define an instrument as containing active and reactive processes within the total instrumental network. In an acoustic instrument the instrumentalist is the active force which sets the reactive chain of sub-activations and reactions resulting in transient noise and resonance. Extending this to the design of 'spatial instruments,' we can think of the space as either containing the complete instrumental process within itself as a closed system (i.e. physical model), or think of the relationship between space and instrument as an open system, where the two instrumental processes (active and reactive) interact across embodiment boundaries. For example, in an open system the interactive point between processes could be at the point of activation (e.g. an embodied attack activating a disembodied resonance, or a disembodied activation of an acoustic instrumental resonance). Or alternatively, a cross-body interaction could be created within an intermediate stage of a given process. This is more difficult to implement, but hypothetically possible. For example, this could be a real ensemble of instruments performing the early reflections of a disembodied virtual source, followed by a virtual reverberation.

3.4. Object / Spatial Treatment

Collecting the above techniques for expressing spatial information through a given rendering system, it becomes possible to draw together a collection of points and lower order shapes (i.e. plane waves, lower order spherical harmonics) to create combined 'spatial objects' which may hypothetically be perceivable as cohesive forms. Used as expressive musical objects, the spatial identities of these collections, may be used to extend the meaning of purely aural phenomena, adding new physical articulation to the already disembodied medium of electro-acoustic sound.

3.4.1.Relative Source Relationships

Looking at how a traditional point-source might be used in a grouped context, it is important to understand how points are separated from each other. For example, the differences in distance from the listener will create a de-correlation of delays, gains, and spectra between points. Similarly, the level of discreteness vs. continuity between point locations in space will effect our perception of separation vs. connection between points based on their level of time and frequency domain continuity — essentially a product of stream segregation principles[10]. The WFS 'focus delay' described above could be seen as being a product of this source-to-group relationship.

In addition to the relative distances between points, the apparent size of point may also be manipulated. This is particularly effective using HOA spherical harmonic operations. It is also possible using WFS rendering, however this was left untested in my study. In the future this could be a potentially effective approach in WFS as well.

3.4.2.Source processing based on spatial location

One of the first tests of group organization I made was to create regions in the space with which a group of points organized in static formation would interact, theoretically re-emphasizing the relative nature of the points through their interaction with absolute regions in space. In the test, I used different processes, such as filters, distortions, or cross synthesis (supervp objects) with gain weights based on the source's proximity to a given region in the space. The result of this approach I found less than satisfying, since in effect it produced a simple fading in and out of a given process. Since the process was always located in the same location, the effect tended to be static spatially, paradoxically, since it was produced by a group of moving sources.

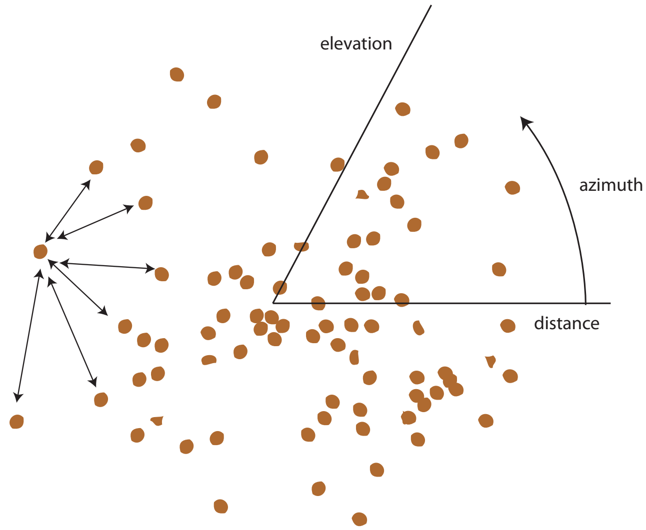


Figure 6. A snapshot of points generated by the "boids" flocking algorithm. Each point's movement is discribed in terms of generalized flock behaviour characteristics. The distribution of points was then used for mapping spatially relative synthesis parameters.

After further tests, I found that a more effective approach to this idea of processing based on spatial location is to work within the spat model itself. For example, the [spat.oper_] object provides a method of calculating the low-level filter weights for a given source based on the source's high-level 'aperture' and 'yaw' parameters. The result is a perceptual model for virtual source directivity, which in a group organisation may be used to describe the outline of a virtual shape. One application of this in my tests was to create a ring of points, with reduced apertures, and with the yaw values for all points facing the center of the circle. Then, moving the entire group (with the new [spat.group] object created for this study) it was possible to move the group of points off into the distance, and slowly bring the group over the head of the listener. As the group drew near, it was perceived as a conglomerate of points, but audibly faced away, and then as the group passed the listener, the listening point would enter the inside of the ring of points, moving into the group's 'radiation region' where the filtration would shift from low-shelf to high, and give the effect of increased separation between points surrounding the listener.

Similarly, it seems hypothetically possible to create a negative spatial image based on a global gestalt of source directivities as points on a pseudo-surface, re-enforced perceptually by common-fate motion of the macro shape. In the future, I think this may be an interesting direction, as well as further integration with the source's filter parameters, for instance adding accentuated timbre changes based on spat's low-level, location relative filter calculations.

3.4.3. Geometric and stochastic source groups with delays and filters

Continuing from my findings in the first stage of experimentation on this idea of relative source relationships, I reflected that rather than timbrally defining a source's relationship to a group-object as being relative to a static physical location (i.e. the room), its timbre might rather be defined as a vector, where each source is relative to the other sources in the group as an abstracted macro-object. In this case, the physical / spatial rendering of the group could be 'transparently' handled by the rendering algorithm with the intrinsic coloration effects for the given rendering technology.

To test this approach, I created a matrix of sources each with an independent filter and delay, whose parameters were calculated based on the source's relative position in the group, centred around the OpenGL <0,0,0> reference point. For the initial 2D test with the WFS array in Studio 5, the value for each source was calculated as follows (in odot [o.expr] notation [11]):

/delay = ["delay", /id, abs(/xyz[[0]]) * abs(scale(/xyz[[1]], -2., 2., -3000., 3000.))];

and then:

/freq = scale(/delay[[2]], 0., 3000., 1000., 5000.);

Which defines the delay of point number </id> as being the absolute value of its x coordinate multiplied by the absolute value of its y coordinate scaled to be approximately 3 seconds at maximum x/y distance. The frequency of the filter was then calculated as a mapping of the delay value. Thus each source's delay and frequency increases as it moves away from the group's <0,0,0> coordinate.

This experiment proved much more successful than the previous. Where previously, the coordinate in space was mapped to different processes, here the parameters of a fixed process are mapped to the source's dynamic motion in space. The global motion of the group was controlled through manipulating spherical coordinates for the points, and then flattening them to 2D, which created wavelike motions. Additionally, I found that by varying the interpolation time of the delay (via [spat.tapout~], and later [spat.delay~]) I could accentuate the natural frequency shifts for each point. I believe the success of this approach is that the motion *and* the timbre variation for each source are controlled in terms of its place in the group — each reinforcing the other. An anecdotal proof of success was when one researcher visiting the experiment had to leave the room after a few minutes complaining that he was beginning to feel seasick.

Expanding on the geometric source group approach, I began to look for more dynamic methods for controlling many points which combine to form a macro-object. I began with simple particle systems, and soon settled on using the 'Boids' flocking algorithm developed in the 1980s by Craig Reynolds. The benefit of the Boids algorithm is the ability to simultaneously create situations

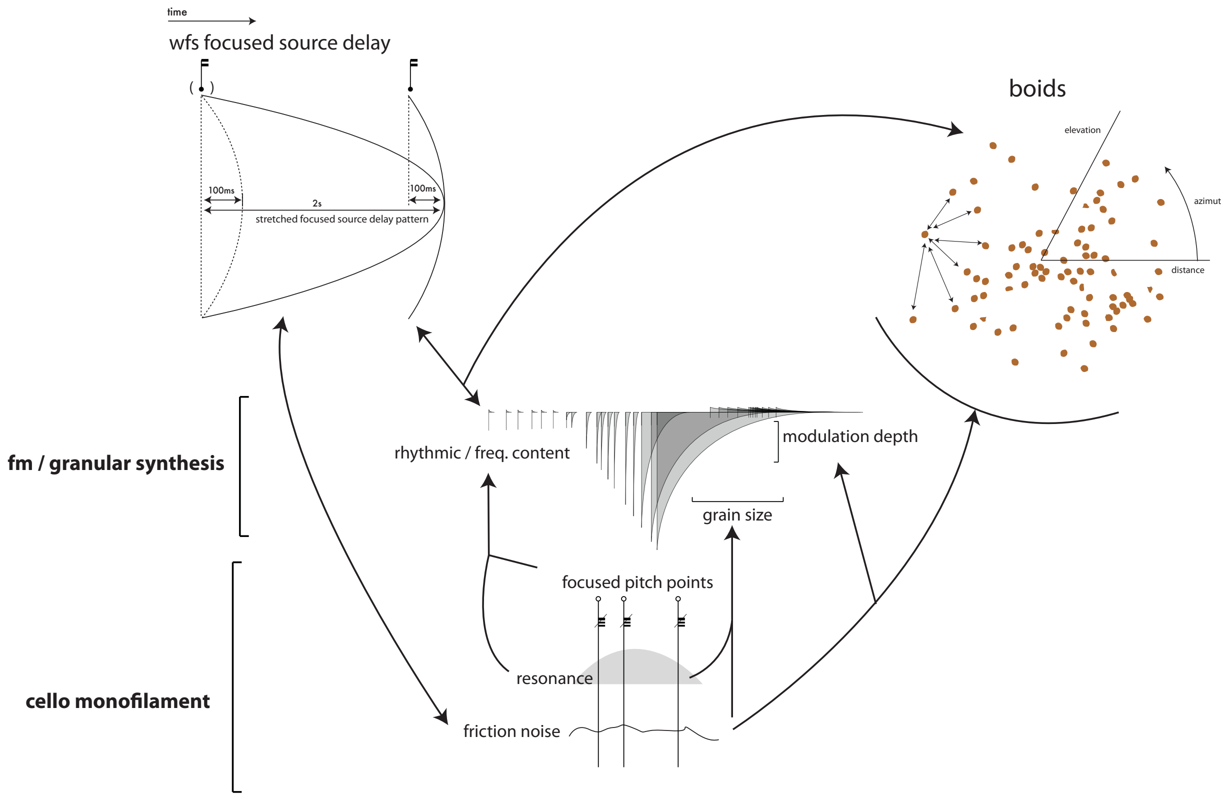
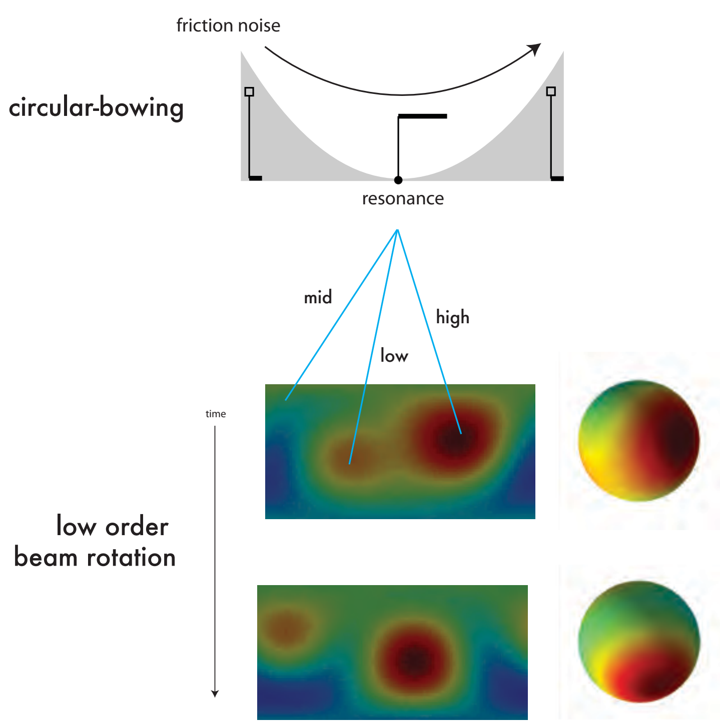


Figure 7. Relating electro-acoustic timbre and gesture to *spatia*l timbre and gesture in the étude *Fluoresce*. (above) The monofilament technique heard at the opening of the piece can be discribed spatially as having fore, mid, and background within a complex timbre field. Each layer of the texture/timbre was used as sculptural form, and used to unify qualitatively, in rhythmic and frequency domains, with the WFS 'focus delay' and 'Boids' spatial processes. (below) Circular bowing, was found to be texturally related to low order spherical harmonic rotations; used as a live treatment, the direct signal was split into three frequency regions using the [irc.gammatone~] filter, with three variably shaped spherical harmonic patterns.



where a large flock of points function always in terms of the others, while additionally having options for seemingly independent motion through attraction and repulsion parameters. Through this approach, I found I was able to create organic macro objects, while additionally accessing insect-like swarm behaviours.

3.5. Perceptual Considerations of Spatial Form

There were several psychoacoustic observations which arose through the experiments made during my residency. Two of the most interesting were the question of spatial form in relation to critical bandwidth and spectral fusion, and the relationship between spatial and sonic forms.

In the case of the first, I found in experimentation with resonance models, that two sources, no matter how far apart they were spatialized, when struck synchronously would cognitively be fused or separated *spatially* when the intervallic distance was greater than the critical bandwidth. This repeats findings by Bregman [10], but was still surprising given the distance between the sources in Espro. The effect was that when sources were within the critical bandwidth (up to ca. a minor third) they would appear to be a very large object; once the interval became larger, they would separate into distinct sources in space (even when struck synchronously).

The second important observation was that perceptually, auditory spatial form is tightly bound to the other spatial aspects of sound, such as pitch-height, foreground/background, and psychoacoustic binding of dynamic/timbral to distance cues. This became of particular interest to me, as described in the following discussion of my compositional process in the étude *Fluoresce*.

4. COMPOSITIONAL APPROACH: *FLUORESCE*

In the final period of my residency, I composed the étude *Fluoresce* for the inauguration of the new WFS / HOA system installed in Espace de Projection. The piece was conceived as a study miniature for cellist Séverine Ballon and Espro; and as a first step in exploring the aesthetic implications of techniques developed through the experimentation described above. Through the process of composing the piece I developed the concept of 'spatial-timbre' — an approach to spatial form coherently fusing with instrumental timbre through shared perceptual morphology, based on the understanding of spatial rendering systems and dynamic group organization as being fundamentally *instrumental* techniques.

4.1. Instrumental Timbre ⟷ Spatial Timbre

Just as complex timbre, when understood spatially, can be described in terms of focal plane — fore, mid, and background layers — complex spatial organization can be understood as timbre. Likewise, tracing a given sound, physical gesture, or spatial process to its irreducible components, there appears common qualitative characteristics which form a relational network, functioning across medium (see figure 7).

For example, the sound of a cello being played with thin pieces of monofilament (plastic fishing line) pressed against its strings, heard in the beginning of the piece, could be described as having several envelope layers of pitch, noise, and resonance (figure 7, top). Here, the pitch layer could be heard as forming tiny dots of pitch which jump out from behind the mid, and background layers of friction noise. Viewed in relationship to the 'instrumental' techniques developed for the spatial rendering systems, this monofilament technique becomes qualitatively related to the WFS 'focus delay' technique, where similarly perforated points jump out from a noisy mid-background texture. Similarly, the grain size used in the 'focus delay' and granular FM synthesis relate to the cello's short perforated pitch points and to the delay distribution in the 'Boids' dynamic group mapping.

Or in contrast, the circular bowing texture used in the second section of the piece, has a more fluid, larger scale gesture, transitioning in a gradient between noise and pitch. Qualitatively, the smoothly transforming nature of the circular bowing seems to relate to the lower order spherical harmonic transformation in the HOA system. (figure 7, bottom)

This act of locating a perceptual intersection between acoustic instrumental technique and spatial rendering was a moment of opening in a personal development of a compositional/aesthetic approach to spatiality. While there are many historical precedents of spatial compositional theories (notably the developments by Boulez, Stockhausen, Xenakis, Nono, Furrer, Nunes, Ablinger and many others [12]), I believe that this approach is unique in that it carries a compositional benefit of being grounded in a nuanced understanding of high resolution virtual spatial rendering as a fundamentally *instrumental* technique, the possibility of which is only now becoming viable for composers through new compositional tools being developed at Ircam and other leading centers for spatial audio research[13]. This approach of understanding spatial organization as instrumental timbre, allows the composer to relate the accumulated knowledge of performance practice and formal texture to new electro-acoustic parameters of space, such as perceived spatial density, presence, and the gestural agency of spatial objects.

4.2.Technical Realisation

Due to the number of channels being rendered, and the real-time computational demands of the 'Boids' spatialization/processing approach and early-reflection image-source model used, extensive testing was required in the composition of the piece to find an efficient system for distributing the computation across the available processors.

The setup involved a cluster of five dual-quad-core MacPro tower computers used as servers to handle the multichannel processing and WFS and HOA rendering, controlled by a sixth 'client' computer which handled the

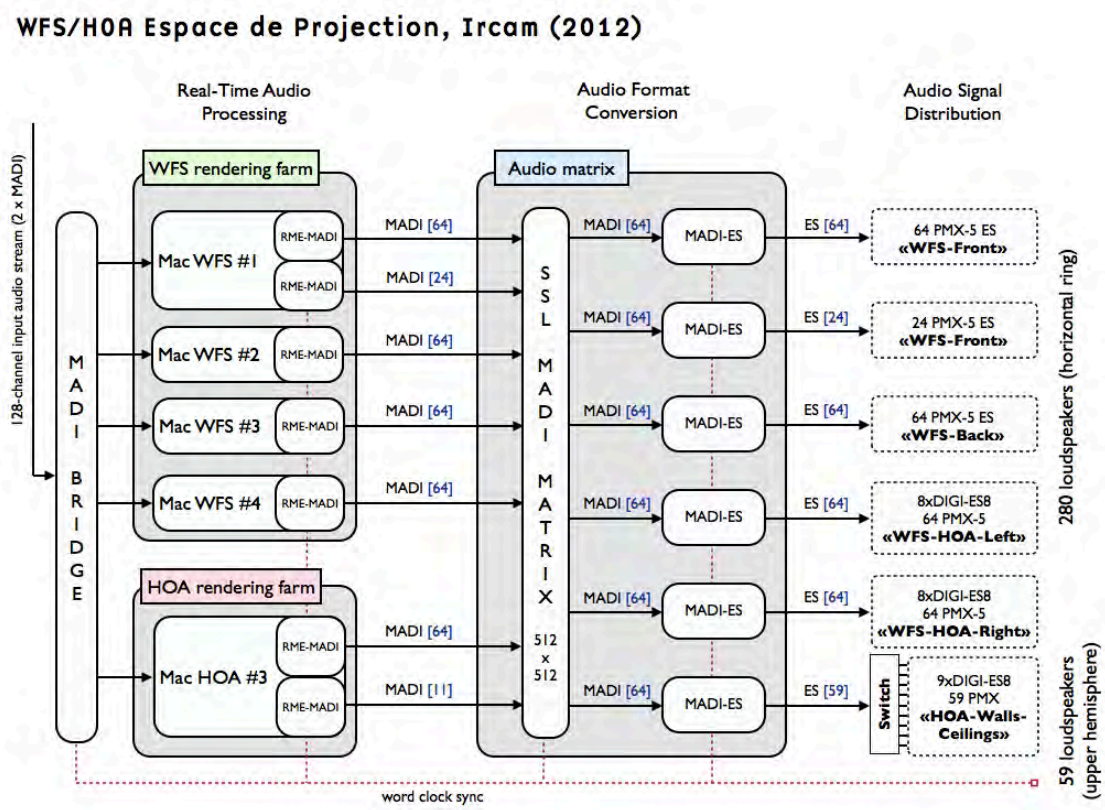


Figure 8a. Computer and audio routing system design for WFS and HOA systems in Ircam's Espace de Projection, designed by Markus Noisternig, Thibaut Carpentier, and Olivier Warusfel.

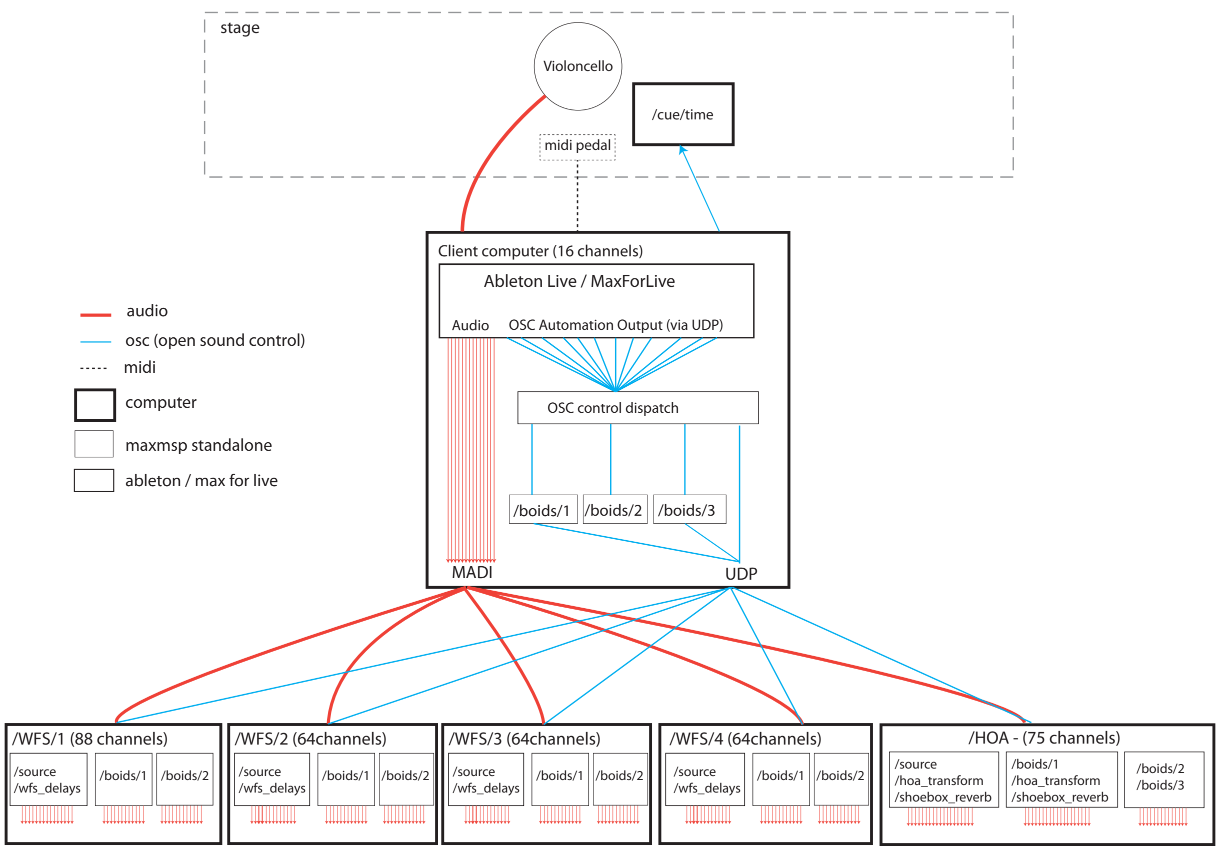


Figure 8b. *Fluoresce*, audio and data flow design. Five rendering computers (bottom row) controlled by the client computer (center) containing the Live/MaxForLive precomposed material, controlled in real time by the live input of the cello and foot pedal on stage (top). The client computer computes the necessary gains and delays to coordinate the four WFS arrays and HOA array. Each computer runs multiple instances of MaxMSP for effient distribution of computation needs.

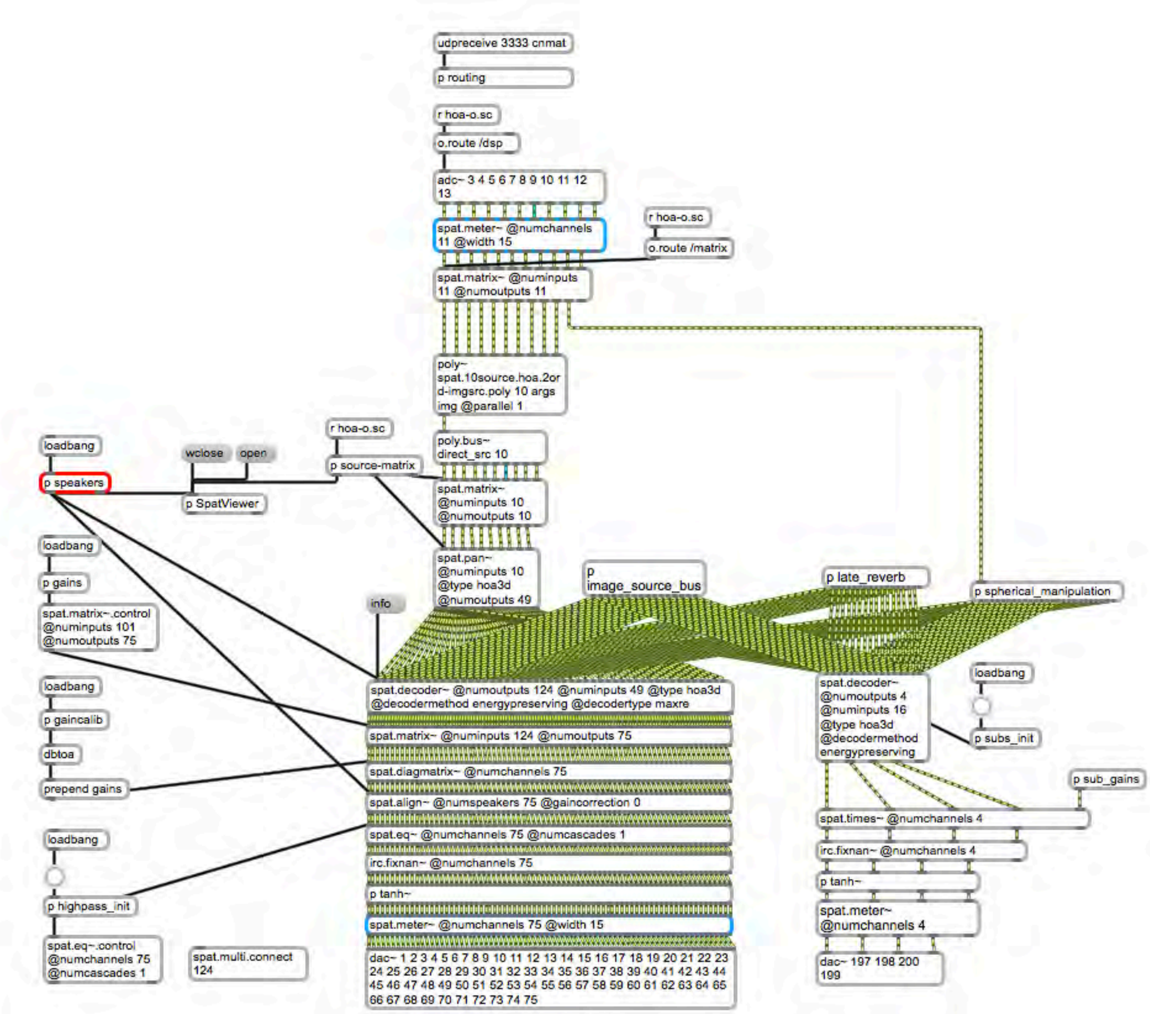


Figure 9. Screenshot of HOA MaxMSP server (1 of 3). Ten sources are rendered with image source model early reflections, and spat.reverb~ late reverberation, and low order spherical manipulation. All processes are controlled via OSC packets sent from the client computer (figure 8b).

central synthesis and message processing (Figure 8b).

The client computer was organized using Ableton/ MaxForLive as the primary control interface. The audio and MIDI input from the cello on stage was sent through a series of treatments and spectral followers (using the zsa.descriptor library developed by Emmanuel Jourdan and Mikhail Malt [14]) to trigger various processes in real-time (e.g. the WFS delay approach, see piece description for more details). Automation controls for the 'Boids' algorithm and spatial rendering were sent from Ableton via UDP to a central odot/OSC (open sound control protocol) [11] dispatch, which processed the messages, calculating the various filters, variables, and mappings for each source or flock, and relayed processed messages to the appropriate server. Each flock was additionally separated into parallel copies of MaxMSP which each had separate mappings.

Detailed in figure 8a is the audio channel path going to the five rendering servers via MADI Bridge and then its conversion to EtherSound as it is sent from the rendering servers to the speakers in the Espace de Projection. The MADI Bridge input allows a maximum of 128 channels potentially to be used by the HOA rendering system (or the front WFS 88 channel array). The WFS computers #2, #3, and #4 have only one MADI input, and so are limited to 64 channel IO.

In *Fluoresce*, the MADI Bridge was configured to mirror the same 64 channels to all systems, in this way I was able to use the client machine to adjust a source's distance attenuation, and transition between WFS arrays, and between the WFS array and the HOA system. The transition between WFS arrays was based on proximity — calculated on the client machine, which then sent the gain weights for each array server via OSC. The transition between the 2D WFS and 3D HOA system used two different approaches. The first, used in the 'boid' flocks, used a Z slice through the space as a hard switch between WFS and HOA. The second approach, developed by Markus Noisternig used a coefficient to smoothly scale the source out of the WFS and into the HOA system as it increased in elevation.

5. ADDITIONS TO THE SPAT DISTRIBUTION

Many of the Max patches developed during my residency are now included with the Spat distribution, for example the Shoebox / Image-Source Model patch based on Noisternig and Musil's paper is now included as tutorial #19 in the spat release. In the process of studying the inner workings of the Spat library, I also re-implemented tutorial #9 "Patching Spat" with a more modularized approach, which is now included in the release. Additionally, I wrote a javascript tool, based on Nathanaël Lécaudé's Max ToolBox[15], for connecting large numbers of inlets and outlets in Max (which is not currently implemented in Max Toolbox — and avoids a large amount of tedious patching in massive multichannel situations) now included in the Spat release as [spat.multi.connect].

Other additions to the Spat distribution, are Spat/Jitter abstractions, for 3D viewing and matrix-based manipulations (see spat tutorial #20), and in the internal IRC library, the objects developed by Carpentier and Noisternig for viewing and decomposing spherical harmonics in map form (see [irc.realshdecomposition] and [irc.realshmap] in the IRC library). The image to spherical harmonic approach described above is included in the [irc.realshdecomposition] help patch.

6. FUTURE AESTHETIC APPLICATIONS

Perhaps the most important aspect which has been identified through my residency is the primary importance of empirical studies of spatial cognition; an approach exemplified by Spat's perceptual model for the source/room relationship. As I discovered in my personal research, the conceptual correlation between time and frequency domain morphologies with *spatial* morphologies is crucial to the successful handling of spatial composition. When working with high resolution systems such as WFS and HOA, we are challenged to find spatial forms which take full advantage of these systems' rendering processes as instrumental techniques themselves. Ideally, the instrument of physical space, the spatial rendering system, and electro-acoustic sonic morphology are treated as a compositionally unified entity, each representing stratified component processes of a single form.

This merging of electro-acoustic composition with spatial form exposes an area of underdevelopment in compositional theory and technique. How should the composer wishing to compose with space approach symbolically representing the relationship between sound and spatial form? Music notation is fundamentally two dimensional, with generally the horizontal axis representing time — how should spatial form be described in time? In most cases (and including *Fluoresce*), the spatial trajectory of a given point is *realized* in breakpoint envelope form, and played back by a digital work station. Even when conceived in a more three dimension way in connection with the sound that it shapes, spatial form is often re-interpreted to fit a pragmatic method of performance. And understandably so, the breakpoint envelope fits easily into the standardized methods for handling temporal form. However, to develop the potentials of spatial composition, qualitatively linking sonic morphology and high resolution spatial morphologies (potentially made of many grouped points), new computer-aided notational approaches need to be developed for representing this data as integrated. Drawing from data-visualization techniques, I believe that a higher-level of symbolic representation linking sonic and spatial form may be achieved. And consequently, provided the tools for symbolically visualizing this integration of sonic and spatial form, and process, an increased conceptual clarity will be fostered between physical space and sound, leading composers and sound artists to integrate these elements to a much higher degree than currently seen. Developments in this area are currently planned at Ircam and CNMAT in collaboration with Thibaut Carpentier (Ircam, Acoustic and Cognitive Spaces Team), Jean Bresson (Ircam, Musical Representations Team), and John MacCallum (CNMAT).

Pushing further the integration of frequency and time domain morphology with spatial morphology discussed above, there is a great potential for the development of the spatio-timbral approach to be applied in electro-acoustic ensemble music through the technique of multi-instrumental envelopes; macroscopic spatial/sonic objects constructed from microscopic instrumental grains of timbre. Much like the technique of concatenative synthesis, the feature descriptions of a sound object may be used orchestrationally to sculpt sound/space as a unified multiplicity of morphologies in the three perceptual axis of time, frequency and space. This approach aims achieve the perceptually coherent application of high resolution spatial rendering techniques described above in an electro-acoustic ensemble setting, and presents a newly articulated space of performance. However, due to the delicate nature of spatio-timbral form, this approach requires careful planning and a focused period of testing in order to effectively integrate the instrumental techniques with spatial rendering techniques. The central danger of composing with spatial audio is a perceptual flattening and diffusing of spatial form through the naïve combination of sound and spatial form. By unifying the timbral and spatial morphologies, it is possible to emphasize the desired virtual form.

Finally, one area I am particularly interested in pursuing is the use of physical space in combination with virtual spatial rendering systems. It is well known that our aural perception of space is highly influenced by our visual perception [16, 17], thus for artists interested in the manipulation of space and sound, there are vast possibilities for the integration of the physical into the compositional palette. I believe there is a huge potential for a viscerally expressive artistic language of spatial form merging sound, architecture, dance, object theater, and projection mapping through an extremely careful, qualitative integration of morphologies as described above.

This is a *perceptual* approach to the composition of temporally dynamic physical space; which in combination with advancements being made with high resolution spatial rendering systems, positions us at this time to explore an augmented expansion of compositional thinking many have dreamed about, but few have achieved.

7. ACKNOWLEDGMENTS

I would like to thank Olivier Warusfel, Markus Noisternig, and Thibaut Carpentier for their work, and insightful feedback throughout my residency. None of the developments described above could have been possible without their assistance and research. I would also like to thank Arshia Cont, Hugues Vinet and Frank Madlener, and the IRCAM staff for their support in facilitating this artistic research. The ideal of aesthetic integration of technological research through artistic production developed in my residency could not have been pursued in a more sympathetic environment.

8. REFERENCES

[1] Jot, Jean-Marc. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems.* 1999: 7:55-69. Springer-Verlag.

[2] Besson, Jean, and Marlon Schumacher. Compositional control of periphonic sound spatialization. Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, 2010.

[3] Peters, Nils. Towards an interchange format for spatial audio scenes. Proceedings of the 2008 International Computer Music Conference, 2008

[4] Schacher, Jan. Seven Years of ICST Ambisonics Tools for MAXMSP. Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, 2010.

[5] Baalman, M.A.J. 2007. "On Wave Field Synthesis and Electro-Acoustic Music, with a Particular Focus on the Reproduction of Arbitrarily Shaped Sound Sources." Ph.D. thesis, Technical University Berlin, Germany.

[6] Noisternig, Markus and Brian FG Katz. Reconstructing sound source directivity in virtual acoustic environments. Proceedings of the International Workshop on the Principals and Applications of Spatial Hearing, 2009.

[7] Corteel, Etienne. Synthesis of directional sources using wave field synthesis, possibilities, and limitations. EURASIP Journal on Advances in Signal Processing, 2007.

[8] Meyer, Jürgen. Directivity of the Bowed Stringed Instruments and Its Effect on Orchestral Sounds in Concert Halls. The Journal of the Acoustical Society of America. Volume 51, Issue 6B, pp. 1994-2009, 1972.

[9] Noisternig M., T. Musil, A. Sontacchi, and R. Höldrich, "3D binaural sound reproduction using a virtual ambisonic approach," Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003 IEEE International Symposium on, pp. 174–178, Jun. 2003.

[10] Bregman, A. 1990. Auditory Scene Analysis: The Perceptual Organization Of Sound. Cambridge, Massachusetts: MIT Press.

[11] MacCallum John, A. Freed, D. Wessel. "Agile Interface Development using OSC Expressions and Process Migration." Proceedings of the The International Conference on New Interfaces for Musical Expression, 2013.

[12] Harley, Maria Anna. "Space and Spatialization in Contemporary Music: History and Analysis, Ideas and Implementations." Dissertation. McGill University. Montreal. 1994.

[13] Peters, Nils. "Sweet [re]production: Developing sound spatialization tools for musical applications with emphasis on sweet spot and off-center perception." Dissertation. McGill University. Montreal. 2010.

[14] Malt Mikhail, Jourdan Emmanuel, "Real-Time Ises of Low Level Sound Descriptors as Event Detection Functions Using the Max/MSP Zsa.Descriptors Library", in *SBCM 2009,* Recife, Brazil, 2009.

[15] Lécaudé, Nathanaël. Max ToolBox. 2009. Online: http://code.google.com/p/maxtoolbox/wiki/Documentation

[16] Gardner, Mark. Proximity Image Effect in Sound Localization. Journal of the Acousitcal Society of America, 43:1243–1248, 1968.

[17] Vroomen, Jean and Beatrice de Gelder. Visual motion influences the contingent auditory motion after- effect. Psychological Sciecne, 14:357, 2003.

[18] Schmeder, Andrew. An Exploration of Design Parameters for Human-Interaction Systems with Compact Spherical Loudspeaker Arrays. Proceedings of the Ambisonics Symposium, Graz. 2009.