Thibaut Carpentier,* Natasha Barrett,† Rama Gottfried,** and Markus Noisternig*

*Unité mixte de recherche UMR9912:
STMS Lab IRCAM-CNRS-UPMC
Institut de Recherche et Coordination
Acoustique/Musique
1 place Igor Stravinsky, 75004 Paris, France
{thibaut.carpentier,
markus.noisternig}@ircam.fr
†Institute for Musicology, University
of Oslo
Postboks 1017 Blindern, 0315 Oslo,
Norway
nlb@natashabarrett.org
**Center for New Music and Audio
Technologies
Department of Music
University of California, Berkeley
1750 Arch Street, Berkeley, California
94709, USA
rama.gottfried@berkeley.edu

Holophonic Sound in IRCAM's Concert Hall: Technological and Aesthetic Practices

Abstract: This article presents a report on technological and aesthetic practices in the variable-acoustics performance hall, Espace de Projection, at the Institut de Recherche et Coordination Acoustique/Musique. The hall is surrounded by a 350-loudspeaker array for sound-field reproduction using holophonic approaches such as wave-field synthesis and higher-order Ambisonics. First we present the design and implementation of the audio system and discuss the challenges of both hardware and software architectures. This is followed by a discussion of spatial composition techniques, aesthetic approaches, and methodologies for composing computer music for high-density loudspeaker arrays, explored through the paradigmatic examples of pieces produced by two artist-in-research residencies.

A Space for Sound Projection

The Espace de Projection ("Espro") is the variable-acoustics performance hall of the Institut de Recherche et Coordination Acoustique/Musique (IRCAM). With a capacity of 250 to 350 seats, the space is used as a concert hall, a recording studio, and as a room for acoustic experiments. It was designed and built by Peutz et Associés with the intention of providing the largest possible variability with regard to form, volume, and acoustic properties (Peutz 1978). The walls and ceiling consist of rotatable prisms (called "periactes," after the Greek *periaktos*, see Figure 1) with three different material surfaces to absorb, reflect, or diffuse the incident

sound. Each of the 171 groups of three periactes can be controlled independently with a remote interface. To change the volume and shape of the room, three ceiling sections can be varied in height, and a roller curtain may be used to acoustically divide the spaces into subregions. The volume of the room is thus variable, with a floor space of 24 m × 15.5 m and a height in the range of 1.5 to 10.5 m. This spatial configurability allows independent modification of various acoustic characteristics, e.g., reverberation times from 0.6 sec to over 3 sec can be achieved.

Complementing the remarkable architectural flexibility of this room, a high-density loudspeaker array (HDLA) surrounding the inside of the hall has recently been installed (Noisternig, Carpentier, and Warusfel 2012a,b). The ambition for this new electroacoustic system is twofold: (1) to offer state-of-the-art equipment for experimental research

on spatial audio and auditory spatial cognition, and (2) to provide unique facilities for artistic performances integrating immersive audio.

The rapid increase in available computing power and the fast evolution of audio interfacing technologies have led to a new generation of sound-spatialization systems with a high number of playback channels. Several computer music institutions are now housing HDLAs with dozens or even hundreds of available channels (e.g., Zmölnig, Sontacchi, and Ritsch 2003; Alcorn, Corrigan, and Rebelo 2005; Ramakrishnan, Gossmann, and Brümmer 2006; Amatriain et al. 2007; Wilson and Harrison 2010), in which two main typologies can be observed: the acousmonium model and the sound-field reproduction approach. The primary purpose of acousmonium systems is for the diffusion of acousmatic works, performed live by controlling the level of signals routed to loudspeakers, which are considered as "instrumental sources." In contrast, the sound-field approach is designed to synthesize virtual sounds, abstracted from the arrangement of loudspeakers, and aims for the diffusion system to be a "transparent" rendering process. Without taking sides in aesthetic debates that may arise regarding these two categories, it was decided to focus on the latter approach for the newly installed HDLA in the Espro.

The following sections briefly review the theory of sound-field reproduction techniques and present the hardware and software designs for the Espro's HDLA.

Approaches to Reproduction of Spatial Sound

Spatial audio systems aim to synthesize virtual sound sources by controlling the auditory cues perceived by one or more listeners. This allows for the manipulation of perceived localization of virtual sounds as well as the associated reverberation effects. Three main approaches to sound reproduction can be used, which we refer to as "perceptual," "binaural," and "holophonic."

The perceptual approach synthesizes phantom sources emulating the main auditory cues (differences in interaural level and time), and relies on psychoacoustic properties of human hearing to produce the illusion of a sound emanating from a particular direction. This approach encompasses stereophony and its extensions such as vector base amplitude panning (Pulkki 1997), and is known to suffer from a limited "sweet spot."

The binaural approach, on the other hand, recreates the auditory cues by sampling the auditory space and does not rely on any perceptive or physical model. The localization cues are reproduced by spatial filtering with the head-related transfer functions of the listener. This technique is restricted to the reproduction over headphones, however.

More recently, the holophonic approach has generated significant interest in the HDLA community. This technique aims to synthesize physically correct acoustic wave fields, independent of any hearing-related model. Such holophonic reconstruction can be achieved over an extended area in two or three dimensions, and can thus be used to present spatial sound to a large audience. The generated sound field is determined by means of mathematical formalisms that solve the wave equation within the reproduction area. Wave-field synthesis (WFS) and higher-order Ambisonics (HOA) are based on boundary integral and modal formulations of the wave equation, respectively. These two technologies are intensively used for spatial sound processing in the Espro.

Wave-Field Synthesis

Wave-field synthesis uses the Kirchhoff-Helmholtz integral equations to determine the sound pressure at any point within a closed and source-free volume given the continuous distribution of secondary monopole and dipole sources on the volume's boundary surface. The strength of each monopole source is determined by the normal component of the particle velocity of the primary sources, and the strength of each dipole is determined by the pressure of the primary sources. Berkhout and colleagues have formulated several approximations (and associated corrective terms) that allow a practical implementation of the WFS principle (Berkhout 1988; Berkhout, de Vries, and Vogel 1993). In particular, they have shown that only the pressure distribution of secondary sources can be used, and that this can be achieved by linear arrays of transducers (instead of the surface arrays prescribed by the theory). In practice, this enables the synthesis and control of coherent sound fields within an area bounded by arrays of loudspeakers. Each loudspeaker is driven by the primary source signal, and delays and filters are applied to properly synthesize the wavefront according to the source position. There is an inherent limitation of WFS, however: The discretization of the array creates spatial aliasing and, as a result, holophonic reconstruction is only guaranteed below the frequency determined by the distance between loudspeakers.

Wave-field synthesis not only covers a large audience with a coherent sound field, but it also offers an interesting palette of spatial figures (Baalman 2007). Virtual sources can be synthesized anywhere "behind" the loudspeaker array and by creating a circular wavefront with the correct direction of arrival. For virtual sources infinitely far away, a planar wavefront is simulated. Finally, "focused" sources can be rendered in front of the array, yielding the impression of a point source inside the room.

Etienne Corteel (2007) showed that it is further possible to control the radiation characteristics of the virtual sources. Manipulating the aperture and orientation of the source allows for natural modification of the level and the sound color of the direct sound, as well as increased interaction with the listening environment.

Higher-Order Ambisonics

First introduced by Michael Gerzon in the 1970s, Ambisonics experienced an intensive surge of
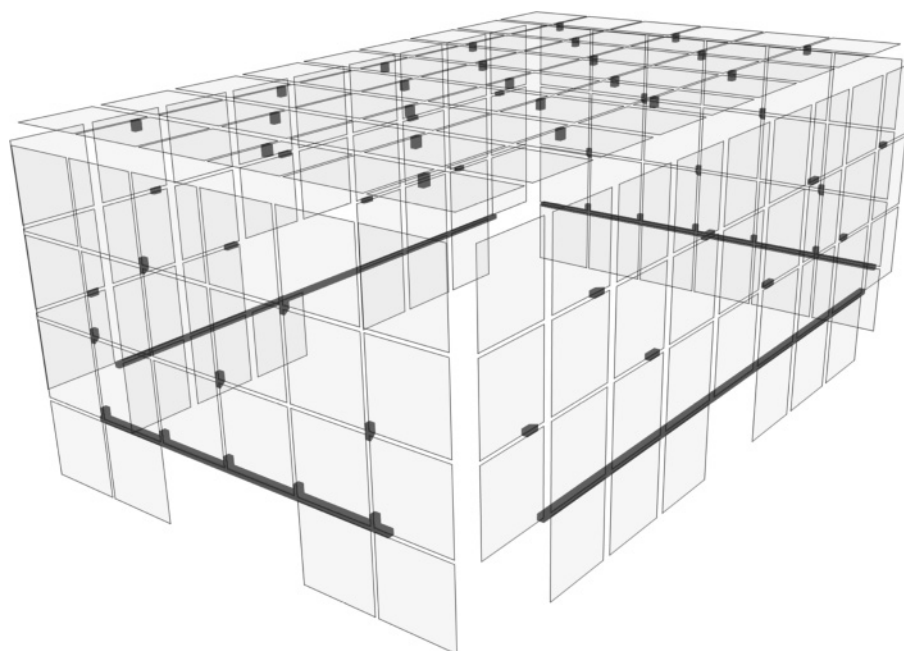
research activity after the introduction of higher-order representations by Jérôme Daniel (2001). High-order Ambisonics expands the 3-D sound field into a set of basis functions, the spherical harmonics. The expansion coefficients (also referred to as wave-spectral or modal coefficients) represent the sound pressure on a sphere and can be either synthesized or derived from spherical microphone-array recordings. These coefficients are used to encode the signals in the Ambisonics domain, and are independent of the reproduction system.

For Ambisonics playback, the encoded audio streams must be decoded with respect to the loudspeaker setup. In practical implementations, the wave spectrum is truncated to the maximum order possible for a set of loudspeakers. Decoding higher-order signals to a lower-order loudspeaker array may result in a blurring due to the reduction of spatial resolution.

High-order Ambisonics were later reformulated to remove the limitations implicit in the assumption of plane waves (Daniel 2003; Daniel and Moreau 2004). Near-field compensated HOA (NFC-HOA) attempts to correctly resynthesize curved wave fronts of virtual sources at different distances. The

practical realization of NFC-HOA requires the use of regularization functions, however, and this has an impact on the reproduced sound field, on auditory distance cues, and on relative distance perception as discussed by Favrot and Buchholz (2012).

HDLA in the Espro

This section discusses the software and hardware aspects of the HDLA in IRCAM's Espro and gives insight into the system architecture, system design, build, and implementation.

Hardware Implementation

Installed between the years 2008 and 2011, the HDLA in the Espro surrounds the hall with a 350-loudspeaker array, composed of four horizontal linear arrays mainly used for WFS and 2-D HOA, and a hemisphere array optimized for 3-D HOA (Noisternig, Carpentier, and Warusfel 2012a,b). The array structure is illustrated in Figure 2. The four horizontal arrays consist of a total of 280
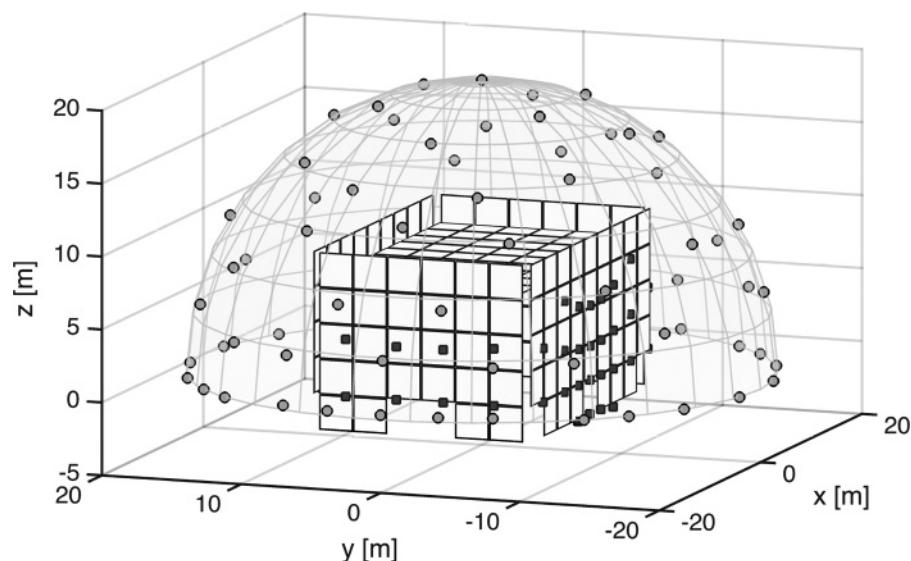
independently controlled loudspeakers (88 channels for the front array and 64 in each of the rest) at a height of 2.2 m. The distance between the speakers is 16 cm in the front and rear arrays, and speakers are separated by 29 cm in the side arrays. The spacing for the front and rear arrays corresponds to a spatial aliasing frequency of approximately 1 kHz, which is high enough to ensure a correct reproduction of the interaural phase-delay cues over the listening area. The choice of type and location of the loudspeakers was partially dictated by mechanical and safety constraints: The cabinets needed to be mounted between groups of periactes, and so were limited to an outer dimension of 20 cm in order not to hinder the movement of the periactes or the room's lighting bridges.

The Amadeus PMX-5 ES/A loudspeakers used for the array are compact coaxial speakers with a 5.25-in. neodymium woofer and a 1.5-in. diaphragm compression driver. Passive speakers are used for the side arrays and the power amplifiers (Lab.Gruppen C20:8X) are located in a plenum space behind the periactes. For full-band sound reproduction, audio signals below 100 Hz are sent via crossover to eight subwoofers (db-audio Q-series). The loudspeaker arrays are daisy-chained and fed with streams of digital audio using the EtherSound protocol.

The HOA array consists of 75 loudspeakers located on the ceiling and on the walls. The lower ring is formed of 26 speakers, of which 16 are integrated into the WFS side arrays. The geometrical layout corresponds to an upper hemisphere projected onto the walls of the cuboid-shaped room (see Figure 3), using digital compensation for propagation delays and amplitude levels to realign the system as a virtual hemisphere centered in the middle of the hall.

The spatial distribution of the loudspeakers was obtained by numerical simulations for different discretization schemes on the sphere to find an optimal distribution suited for HOA playback. Using the condition number of the HOA decoding matrix as a quality measure, Noisternig, Carpentier, and Warusfel (2012b) have shown that the array structure implemented is well suited for 3-D HOA reproduction up to ninth order.

For HOA pre- and postproduction, a 24-loudspeaker hemisphere array was installed in one of IRCAM's studios.

Spatial Audio Rendering

A cluster of five computers located in a sound-isolated room next to the Espro renders audio signals to the HDLA in real time (see Figure 4). In

Figure 4. Audio processing and signal routing structure. The numbers in brackets indicate the number of transmitted audio channels.
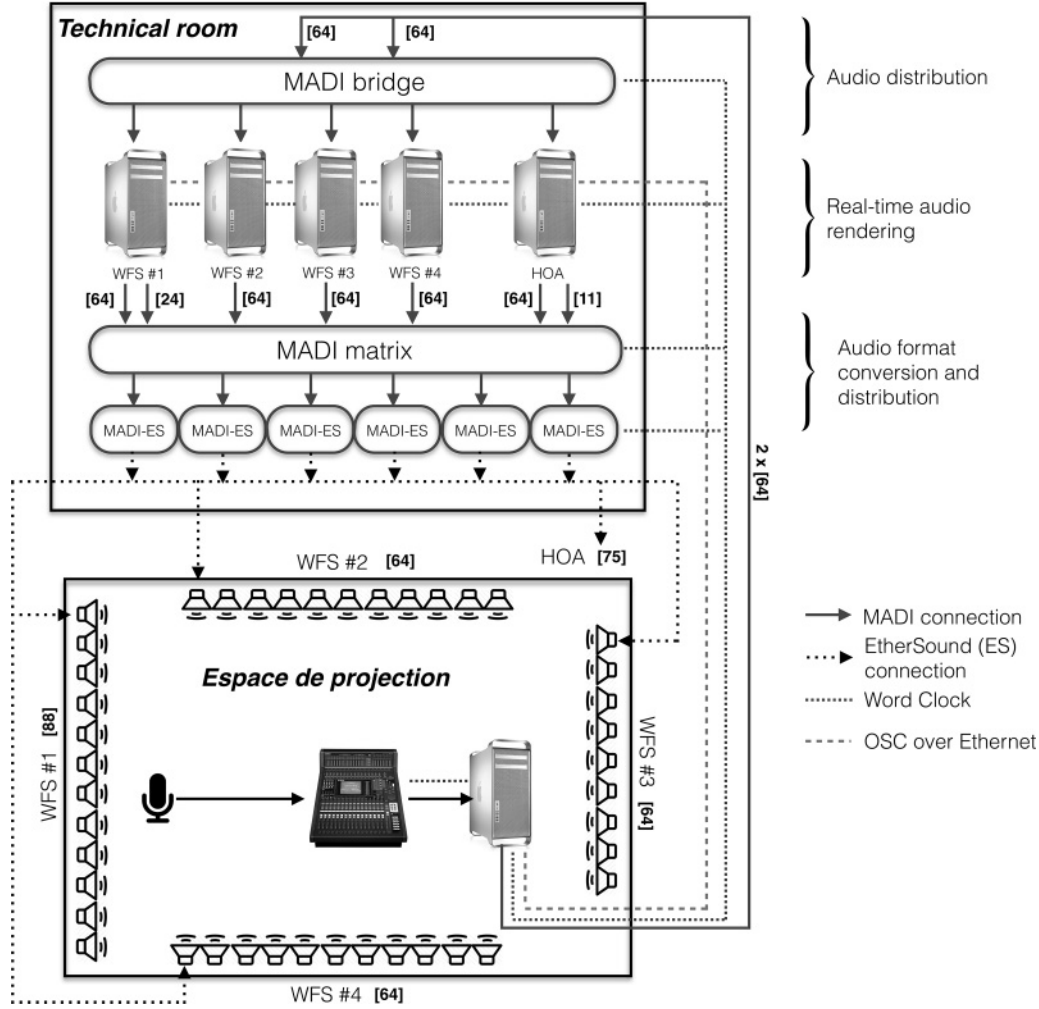
the basic configuration, each computer (Apple Mac Pros, each with a low-latency RME MADI interface) is allocated to a component array of the system (i.e., one computer for each WFS array, and one computer for the HOA array). Espro and the server room are connected by optical fiber cables, which are patched to a 8×64 I/O MADI bridge used to replicate and route up to 128 audio channels as input signals to each audio renderer. After processing, a 512×512 channel (8-port) MADI routing matrix distributes the audio output channels to the loudspeakers. This allows the combination of individual channels to any output destination and thus to use different audio systems simultaneously (e.g., WFS for two-dimensional and HOA for three-dimensional spatial sound reproduction). The overall round-trip latency can be as low as 1.5 msec (2×32 samples at 48 kHz). Word clock is used for synchronization of all devices. All processors are connected over a network for remote control, typically through the Open Sound Control protocol (Wright 2005).

Most productions in the Espro use the Spat software suite for real-time spatialization (Jot and Warusfel 1995). Developed at IRCAM, Spat is designed as a modular and scalable C++ library for sound spatialization and artificial reverberation,

and is primarily distributed as a set of external objects for Max/MSP. The software provides a unified framework for multichannel processing, covering a wide range of spatialization algorithms accessed through high-level perceptual control interfaces (Carpentier, Noisternig, and Warusfel 2015). In particular, 2-D and 3-D HOA encoding and decoding are supported without any limitation on the maximum Ambisonics order, as all algorithms are implemented with recursive formulae. Spat also provides various decoding strategies. Traditional decoding approaches (see Daniel 2001) consist in sampling the spherical harmonic excitation at the loudspeaker positions (sampling decoder) or matching the excitation modes of a continuous sound field to those of the loudspeakers (mode matching). Because these techniques do not perform satisfactorily with nonuniformly distributed loudspeaker arrays, Zotter, Pomberger, and Noisternig (2012) have proposed an "energy-preserving" Ambisonics decoder. This decoder avoids the problem of energy from virtual sources varying with the direction of arrival, even coping with regions of sparse loudspeaker coverage, while exhibiting a low directional deviation. This method has been found to be effective with respect to the Espro layout, as discussed later in this article. Furthermore, the Spat decoder module can perform dual-band decoding, with adjustable crossover frequency, and in-phase or max-r_E optimizations can be applied in each band (Daniel 2001). Spat provides numerous utilities for enhanced HOA processing and workflow, including transformations in the spatial Fourier domain (e.g., weighting the spherical harmonic components for blurring the spatial image or transformation matrices for efficiently rotating the 3-D sound scene); conversion between various normalization schemes (N3D, SN3D, FuMa, etc.); sorting of the HOA components according to different conventions (ACN, SID); and spherical microphone array encoding. Spat additionally supports NFC-HOA processing. We know that practical realizations of NFC filters require regularization to avoid strong low-frequency boost. In its current version, Spat only implements the regularization method proposed by Daniel and Moreau (2004), although Favrot and Buchholz (2012) later showed that a modified version of Tikhonov regularization

(MTReg) reproduces the sound field over a larger area and more accurately matches the pressure-field amplitude decay of a point source. The advantage of Daniel and Moreau's method is that efficient infinite impulse response filters exist for real-time implementation of NFC-HOA, which is not possible when using MTReg.

The Spat library also includes Max objects for real-time WFS rendering and supports the synthesis of directional and focused (i.e., in-front-of-the-array) sound sources. Wave-field synthesis filters depend not only on the array configuration, but also on the source position and directivity, and so a set of filters is precomputed for a dense grid of possible source positions to keep the CPU load as low as possible. The real-time rendering is processed in three stages: First, the directivity of the source is synthesized by linear combination of filters corresponding to elementary patterns (circular harmonics); then, delays and gains are used to generate the wavefront according to the position of the source; finally, the frequency response of the transducers of the array is equalized.

Working with HDLA and large multichannel streams comes with a certain complexity. Traditional audio-processing tools are frequently inadequate for such a workflow, so the Spat library provides several tools for dealing with common tasks, such as: file I/O (`spat.sfplay~` and `spat.sfrecord~` support playback and recording of audio files up to 250 channels, corresponding to the current limitation of Max/MSP, and they support the WAV RF64 format, which allows the usual 4-GB limit on file size to be exceeded); automatic time and level alignment of an arbitrary set of loudspeakers according to their geometrical arrangement; multichannel bus (send and receive) without patch cords; and binaural monitoring of any multichannel streams.

Aesthetic Use of the Espro HDLA

Since the installation of the HDLA, a number of artistic productions have used the system, covering a wide range of forms (electroacoustic concerts, theatrical plays, embodied generative music and dance

performances, interactive installations, video projections, etc.). In particular, the research by composers in IRCAM's Musical Research Residency Program presents an interesting perspective on the nature, methodologies, and prospects of computer music for HDLAs. Indeed, the research residency provides a unique context in which composers and scientists collaborate on common concerns, at the intersection of artistic creativity and scientific discovery. Such joint projects culminate in the creation of new art works, as well as the development of multipurpose tools that benefit not only their authors, but also the greater computer music community. In 2012, Natasha Barrett and Rama Gottfried undertook research residencies aimed at pioneering innovative techniques for spatial composition, challenging the possibilities of the hardware and software environments in the Espro, and eventually revealing the artistic potential of HDLAs. Although they followed very distinct—yet complementary—strategies and sought different musical aesthetics, the approaches and methodologies they have developed provide paradigmatic insight into the use of HDLAs for computer music. Barrett focused on the practical, perceptual, and compositional application of HOA while working on a pure acousmatic composition. Gottfried explored the use and creative misuse of WFS and HOA signal-processing techniques from the prospective of site-specific real-time electroacoustic performance practice. In the following two sections, they each present the main findings of their research residencies.

Natasha Barrett: Tangible Space and Hyperreality

My work investigated the compositional implications of a heightened perceptual impression of distance and spatial clarity, and how 3-D imagery can imbue abstract or surreal musical scenes with the spatial details idiomatic of real-world landscapes. I was particularly interested in musical manifestations of "tangible" space animated by realistic motion behavior, and how a coupling of spatial tangibility and sound identity may feed back to influence our perception of spatial features.

Compositional Motivation

These interests grew out of my previous Ambisonics compositions exploring spatiomusical structure. In my early work (Barrett 2002) I discovered Ambisonics techniques as more conducive to 3-D spatial concepts than panning or sound-diffusion performance, and from this period I developed a compositional approach to spatial structure independent of the loudspeakers. Musically, in this work I experimented with immersion, sound separation, moving trajectories, the projection of real-world spatial relationships, and the preservation of composed musical space after the work leaves the studio. Later work centered on spatial imagery and how the angular clarity of HOA challenges the composer to avoid monophonic points in space, while offering "enhanced reality" or "hyperreality" in acousmatic listening (Barrett 2010). The techniques I developed during the residency in 2012 extended this work with four research areas of enquiry.

1. Listening area: Can HOA reproduce stable spatial images over a large listening area, and what variations do listeners experience for different types of material depending on where they are seated?
2. Sound imagery: Does an HDLA change the approach to composing sound images?
3. Near-field compensated HOA: How successfully does NFC-HOA stabilize source location and enrich our perception of distance and depth?
4. Recorded fidelity in higher and lower orders: Do higher-order recordings more accurately reproduce real spatial environments than first-order recordings, and if so, what are the challenges for their use as compositional sources?

Although some answers can be found in mathematical models, these models are not designed to capture our perceptual experiences, nor can they address the context of musical materials. The complexity of the problem is such, however, that solutions can also be revealed through compositional investigation. The three-part composition *Hidden Values* is illustrative of my results.

After initially considering live performers, I soon realized that our visual perception, which affects our spatial hearing (such as the well-known ventriloquism effect; see Radeau and Bertelson 1974), would distract from both my own spatiomusical investigations as well as listeners' perception of the results. The work thus became an acousmatic composition, where an unhindered investigation of spatial sound in compositional practice could be achieved by removing visual causation cues. My work has since developed to explore tangible spatial images directly connected to human physical performance by way of precise motion tracking and data sonification in 3-D (Barrett 2015).

Sound Materials and HOA

Ninth-order 3-D upper-equatorial, and twelfth-order 2-D horizontal, Ambisonics projections were the starting points for my tests. In previous work I had experienced how the inclusion of psychoacoustic parameters, rather than just the spherical harmonics representing the distribution of energy on the surface of a sphere, create an auditory presence "in space" (Barrett 2010). With this in mind, the `spat.oper` interface, which handles the correlation between psychoacoustic components of the spatial projection, was used to control both psychoacoustic and HOA encoding. Some factors of interest included distance-related attenuation of amplitude and emulation of air absorption for each source, a delay derived from source distance, the Doppler effect, and—for close sounds—a correct weighting of the omnidirectional component, giving the impression that the listener is immersed in the sound (similar to W-panning proposed by Menzies 2002). The reverberation model used by Spat was removed in favor of recorded environments and 3-D impulse response convolution.

Two classes of sounds were chosen for the tests: vocal sounds and a selection of percussion instruments portable enough to transport between recording locations of contrasting natural acoustics. Voiced sounds are interesting in that transformations in time, frequency, and amplitude may change the owner's identity, but rarely disguise a vocal source. When lacking other room or environmental cues, transformations are therefore more likely to influence our understanding of distance rather than identity. Percussion sounds are in many ways the opposite: Simple sound transformations may alter both sound identity and our assumptions of proximity.

A variety of common sound transformations were applied to these sources: spectral stretching and transposition, time-stretching, resampling and tape-transposition, brassage, and granulation. Temporal variation was controlled by three methods: algorithmic processes that gradually changed over time, parameters that were controlled by hand or by automation data, and the layering and mixing of many sounds to create new compound sound objects displaying a timbral temporal-morphology. The resulting sounds evolved over time scales from milliseconds to minutes, thereby creating fast, dynamic variations alluding to human physical gestures; producing spectrally spiraling simultaneously sounding layers; and generating continuities between static points, sound images, and immersive textures. Most importantly, temporal, amplitude, and timbral behavior was synchronized with spatialization encoded in HOA. This synchronization drew on how we perceive spatial sound in the real world. For example, to aid the illusion of a sound moving rapidly through the listening space, spatial proximity and motion speed would be allied with changes in spectral filtering, volume, and image spread. In contrast, densely granulated immersive fields would create a space, rather than defining an object in a space. In this way spatial and spectral development were embodied within each sound.

The Spat library, like most multichannel spatialization software, presents the user with a front-end that maps a monophonic sound to a point in space. It is therefore useful to describe three approaches to imaging:

1. Signal decorrelation can lead to a sound image appearing to change in size. Known as a standard technique in stereo processing, and similarly in HOA, by splitting a mono sound into multiple copies and partly decorrelating each signal, an image may be

controlled to expand and contract, or exist as a cloud of points occupying a specified azimuth and depth in relation to the listener.

2. Sound sources can be recorded with multiple microphones where each signal would occupy a location in the sound field. Similarities in each signal fuse into an image. I first summarized this technique in earlier work (Barrett 2010), but here developed a scalable approach for recording small acoustic instruments and larger, complex, indoor and outdoor environments.

3. Individual sound grains can be spatially encoded, including distance emulation. In granulation and brassage processes, grains can be defined uniquely and individually, globally by way of statistical control, or in a temporal transformation between different states (Barrett 2016).

Testing Materials and Decoder Types

The artistic nature of the project and the complexity of musically relevant materials rendered scientifically accurate decoder tests unrealistic. The study was, therefore, informal. Aside from testing for a decoder's ability to project the intended result, further assessments included whether sound gravitated to any single loudspeaker and how stable the image appeared across a zone spanning 50 percent of the loudspeaker array. All work was carried out with the Espro's variable acoustic set to the minimum reverberation time.

The preparation and composing work involved limited access to the Espro and was mainly carried out in the HOA Studio housing a fourth-order Ambisonics hemisphere, as well as in my home studio over a horizontal array. To feel compositionally comfortable, the workflow integrated Nuendo (a familiar digital audio workstation), the Spat library running in Max/MSP, and a way to seamlessly move between spaces by composing in ninth-order HOA while monitoring a decoding order appropriate for the array at hand. (The musical implications of a depreciated spatial monitoring are discussed in Barrett 2016.)

In the Espro, the optimal decoder parameters transpired to be the dual-band, in-phase, max-r_E, energy-preserving decoder at a crossover frequency between 200 Hz and 400 Hz, depending on the source material. Although lower than suggested in the literature (cf. Gerzon 1975), this frequency optimized the contrast between diffuse fields and point sources. The results for 2-D decoding in the Espro were the same as for 3-D, although the mode-matching decoder also performed particularly well. In the HOA studio, the 3-D mode-matching decoder was comparable to the energy-preserving method. The studio is equipped with a regular half-polyhedron over which energy preserving is comparable to mode matching when virtually "mirroring" the speaker array below the equator to improve the conditioning of the decoding matrix. The former offered a slight improvement in the 3-D impression of spatial images "in space" as independent from the loudspeakers.

Testing Decoder Orders

The ninth-order 3-D format contains 100 spherical harmonics, and the spatial encoding results in a 100-channel sound file, which requires significant resources for reading, writing, mixing, and mastering. It was therefore useful to test ninth-order 3-D in contrast to lower orders. I applied the same approach as that described in the previous section, and concluded through blind tests with colleagues (all experienced listeners), that seventh-order 3-D (a 64-channel format) produced the same impression as ninth-order 3-D. Although the angular blur for ninth order is lower than seventh order, in the Espro it is likely that room reflections and slight imperfections in loudspeaker locations and responses will add greater inaccuracies. In two dimensions, twelfth order offered an improvement over lower orders, especially for complex scenes, which may be due to our enhanced awareness of horizontal space and the lower impact of room reflections on a horizontal decoding. Fourth-order 3-D created noticeably blurry results, even though the decoding was tested using all loudspeakers as well as a more correct subset of 24 loudspeakers (audible artifacts are created when

using too many loudspeakers, as shown in a number of studies; see Zotter, Pomberger, and Noisternig 2010). In the HOA studio, however, fourth-order 3-D was experienced as similar to seventh-order 3-D in the Espro. As long as the listener is centrally located, the studio offers clear compositional advantages in terms of a dry and controlled environment. In transferring between studio and HDLA spaces, the most prominent differences were in the natural room reverberation and spatial dynamics, where a sound travelling 24 meters through the concert hall travels only 5 meters through the studio in the same time. In a traditional context of stereo sound diffusion, the performer would adjust speed of motion and volume dynamics appropriately for the expression of the music. In the Ambisonics context, we can consider two options.

First, real-time encoding and decoding, where composed sounds and composed spatial motion are stored separately as audio stems and control data. Spatial and volume data can be scaled for each space to ensure that the absolute distance travelled is constant between concert halls of different size. In my own work I tested this approach only to find that the technical complexity of the final composition was too high. If each sound image consisted of five to ten spatial points synchronized with a spectral-spatial transformation, and 100 sound images were mixed and montaged throughout the work, the scale of real-time resources is unrealistic, as are the organizational demands on the compositional process. We can easily see that there are clear advantages to the traditional compositional approach of continuously mixing and bouncing down to make new compound sound objects.

Alternately, we can regard the differences as a feature of the concert experience. Following in the acousmatic tradition, a cautious approach to artificial reverberation in the studio is coupled with the expectation that a concert space will add reverberation and energy to dynamic contrasts, such as attacks and rapid movements in space. We can further acknowledge that the studio is an artificial context, where in a concert the spatial motion will rescale, and changes relating to a "human scale" will be part of this transformation. Investigations into embodied cognition and gestural-sonorous objects

(Godøy 2006) underline my own experience that the musical expression is preserved, and the character of each concert space adds the excitement of a unique experience.

Testing Near-Field Encoding

Although WFS can synthesize virtual sources behind and in front of the loudspeaker array, it is interesting to investigate how successful NFC-HOA is at achieving similar results, even though we were aware of various limitations in the NFC implementation explained earlier in this article. In the twelfth-order 2-D decoding over 26 speakers, there was disagreement among listeners as to the proximity of the focused sources. For some listeners, the source appeared to hover inside the loudspeaker array, enhancing a sense of intimacy, yet there was no obvious difference between a distance of 3 or 6 meters (or 25 percent and 50 percent, respectively, of the radius of the Espro). Musically addressing absolute distance for focused sounds inside the loudspeaker array was thus unrealistic— and unsurprising, considering our poor ability to judge distance for real-world sound (Kolarik et al. 2016), such that it is more useful to consider distance contrast. For sources located outside the loudspeaker array, image stability was improved.

An unfortunate drawback of the NFC-HOA workflow is that the encoding and decoding stages are no longer independent: The loudspeaker array radius must be specified in the encoding. To address this problem, a radius adaptor tool, which is added to the decoding stage, was tested and proved to function successfully for compensations between 50 percent and 200 percent. In the studio there was no noticeable difference between HOA and NFC-HOA. Furthermore, the studio itself is already an intimate environment, where contrasts between near and distant will be less apparent than in the Espro.

The Composition of *Hidden Values*

Thematically, the composition explores drama and metaphor associated with three ancient inventions

familiar in our everyday lives: the umbrella, sight correction, and the lock, forming three movements, called *The Umbrella*, *Optical Tubes*, and *The Lock*.

Sound Sources

Musical sketches were prepared for the two performers (soprano Evdokija Danajloska and percussionist Gilles Durot). An initial recording session took place in a dry studio, from which short sampled extracts were developed. A second set of compositional sketches were then recorded, this time in the sizable space of the foyer at the Centre Pompidou. Further materials were recorded outdoors. For all sessions, an Eigenmike (fourth-order 3-D sound field), a soundfield microphone (first-order 3-D sound field), close and far cardioid, and omnidirectional microphones were used in simultaneous and synchronized recordings for use in the spatial imaging technique described earlier.

Musical Development

Musical materials were developed in a similar way to the encoder and decoder test sounds. Notably for transformed percussion sources, the gesture and energy behind the articulation, rather than source identity, shaped the spatiomusical discourse. In addition, complex sound fields recorded by both the soundfield microphone and the Eigenmike furnished the music with complete 3-D environments that were a challenge to create by other means. Decoding first-order recordings over all 75 speakers confused the spatial image and reduced the size of the stable listening area, however. My solution was to use the Harpex decoder, a nonlinear method that separates simultaneous signals and renders the sound field using a panning method (Berge and Barrett 2010). It produced good angular separation for some sound fields, but proved less optimal for general immersive results. For the latter, turning defect into benefit, the blurry effect of lower-order sources decoded over high-order loudspeakers was used to project intentionally vague, immersive sound fields. Additionally, in-phase-only, rather than dual-band, decoding furthered this result, yet transpired in a trade-off between the immersive effect and an unstable space across the large, stable HOA listening area. Musically, however, as long as HOA added clarity and spatial reference, this instability proved a positive variation to each listening location.

When using Harpex to decode the first-order signal and the energy-preserving decoder for the fourth-order signal, the two microphones produced similarly clear source separation. There were absolute spatial and spectral differences, however, questioning which recording and decoding was closer to reality. In general, the Eigenmike produced a greater sense of "spaciousness." To enhance the sound field further, I developed a hybrid solution applying beam-forming techniques to extract sources and their locations in space, reencoding these extracted sources in seventh order, and playing the result simultaneously with the original 3-D recording. In this HOA layer I also incorporated other monophonic microphone signals, as well as musical sound transformations of the extracted sources (solving image distortion problems normally associated with transformations of 3-D recorded sources). The technique facilitated more accurate control over distance and proximity, combining spot microphones and NFC-HOA in contrast to, among other techniques, distant recordings and impulse response convolution.

Summary

The techniques I developed applied very-high-order Ambisonics to create a precise spatial discourse, extending the extremes of near and far, and integrating recorded 3-D sound fields in a hybrid approach to encoding, decoding, and composition. To achieve this, six types of encoded materials, organized into musical layers, were tied to six parallel decoders, summarized in Table 1. These technical considerations were integrated into my compositional workflow, an abstraction of which is summarized in Figure 5.

Rama Gottfried: Timbre Space and Spatial Timbre

Like the interval shift when overblowing on a clarinet, or Nyquist aliasing in digital frequency

Table 1. Compositional Integration of the Technical Workflow

HOA Format	Musical Materials	HDLA Result
Seventh-order 3-D (Full hemisphere)	Static points and moving gestures Images (multipoint shapes), moving, and static Granulated 360-degree fields Reencoded extractions from sound-field recordings	Precisely articulated 3-D space "Tangible" volumetric images High detail, immersive fields Articulations over real recordings
Twelfth-order 2-D (Horizontal only)	Static points and moving gestures Images (multipoint shapes), moving, and static	As seventh-order 3-D, but adds downward weight to the sound field
First-order 3-D (Full hemisphere)	Soundfield recordings and Harpex decoder Soundfield-recorded IRs and energy-preserving decoder	Panned source separation Immersive, imprecise blur
Fourth-order 3-D (Array subset)	Eigenmike recordings	Immersive real scene, low detail
NFC-HOA/WFS (Horizontal only)	Intimate sounds (timbre, volume, and identity); needs to accept "folding" into 2-D HOA layer when NFC or WFS is unavailable	Appearance of closer sources in front of the loudspeakers

modulation, the artifacts of a system's construction are its unique fingerprint (Cascone 2000). Thus, the design of an instrumental system can be seen as a form of composition that embodies certain possibilities of expressivity and musical structure (Chadabe 1984; Wessel 2006). Working to discover the unique possibilities of WFS and HOA as "instruments," my residency at IRCAM sought to develop a palette of aesthetic spatial material through exploring the use—and creative misuse—of the available tools and rendering processes.

System Specific Techniques: Wave-Field Synthesis

After exploring the traditional rendering approaches, I began experimenting with rational and irrational manipulation of the WFS delay and gain values. I found that by altering the WFS filter structure, the source location became "blurred," depending on how accurately the sound field was rendered. Scaling the delay times to as much as 10 sec produced an accelerating or decelerating wave of sound across the WFS array. Using short granular sounds, each speaker was articulated separately, which spatially flattened the sound to the location of the speakers and created a streaming, linear movement along the walls of the space.

The difference in delay structure between a standard point source, located behind the WFS array, and a "focused" source, located in front of the array, is particularly distinctive with this technique. When a source is behind the array, the delay times radiate naturally from the center of the sound source, but when the source is rendered in front of the array, a "time reversal" approach to delay is used, with the delays beginning at the edges of the array and sweeping towards the center of the virtual source location. Using the focus source-delay pattern, I found that by suddenly introducing the original, unaltered delay and gain values synchronized to the last delay of the series, the temporal blurring and flattening of the sounds against the loudspeakers could transition to the sound "jumping off" the speakers to the interior of the space. I called this phenomenon a "focus delay" (see Figure 6). This approach creates a dynamic shift in the listener's perception of location by first connecting the sound with the physicality of the speakers and then suddenly rendering the virtual focused source inside the room several meters in front of the array.

Figure 5. A hybridization of compositional and technical approaches in the workflow. Solid lines show the workflow within the composition and in the encoded domain; dashed lines show where the composition enters the decoding domain; and dotted lines indicate the exchange of synchronization commands and spatial control. At some point in the compositional process, the complexity of the spatial encoding exceeds practical constraints and the musical "chunk" or layer is rendered and inserted back into the workflow as a seventh-order HOA source.
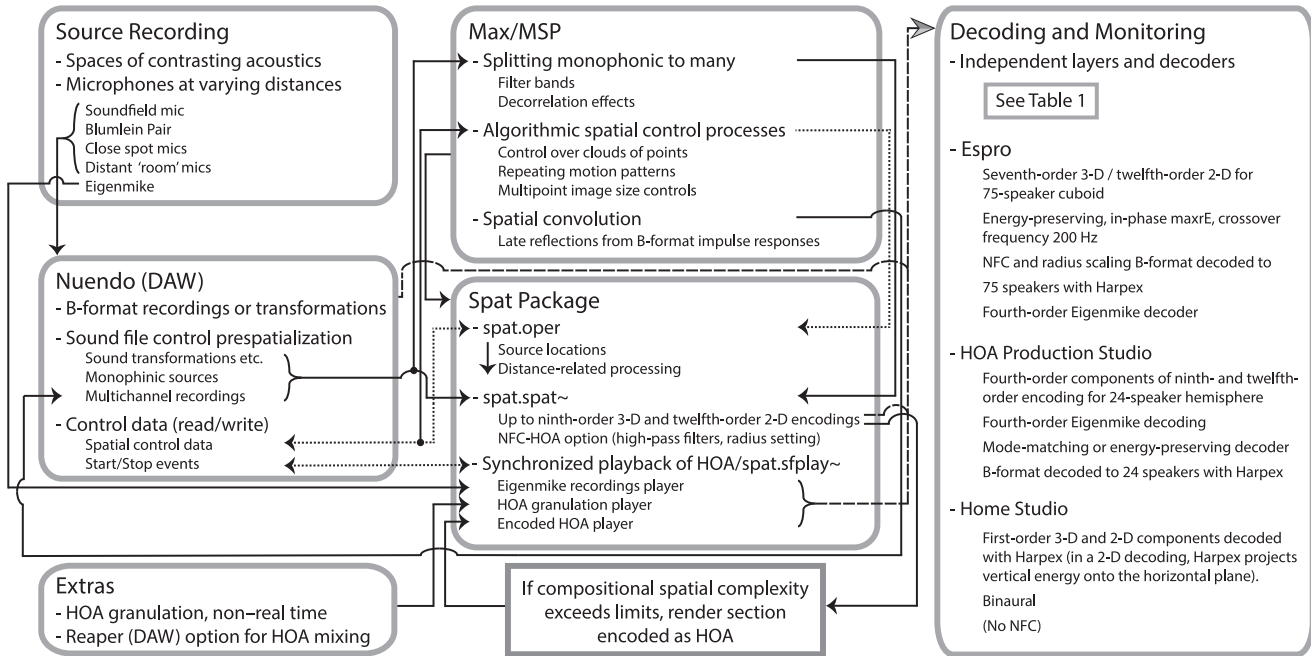
System Specific Techniques: High-Order Ambisonics

A unique element of rendering HOA is the use of spherical harmonics to represent spatial distribution. In my experiments, I found the most perceptually successful spherical domain manipulations was using very low-order spherical harmonics. One such processing approach was to break a sound source into multiple frequency layers, using a gammatone filter bank, and then to rotate each frequency region separately using low-order spherical harmonics. This approach was used in my composition *Fluoresce*, discussed later in this article. Used to spatially process a sustained complex string sound, the approach worked to separate, both sonically and spatially, the sound into its polyphonic components, while the unity of the sound would simultaneously bind the separated spatial components together perceptually.

I also worked to develop a new tool for drawing distributions of spatial energy, both by hand and using generative video techniques, thus representing the energy distribution as a pixel map that could then be decomposed into spherical harmonics and rendered via HOA (see Figure 7). These tools aimed to develop methods for breaking out of the point-based understanding of space into something more sculpturally flexible, and the result could be described as kind of spatial texture relative to video input and HOA resolution.

Grouping and Relative-Source Relationships

Collecting these techniques I had developed, I began exploring the grouping of high-resolution point sources and lower-order shapes to create dynamic "spatial objects" that could be perceived as cohesive forms.

One successful early experiment in source grouping was to define a multisource spatial object using the Spat libraries' perceptual-directivity model to stabilize the relative movement of the group by orienting a perimeter of sources towards the center of a geometric shape. With the listener outside the shape, the form was perceived as being audibly located at a distance from the listener, heard as predominately

Figure 6. The "focus delay"
technique, combining
scaled and unscaled WFS
delay times.
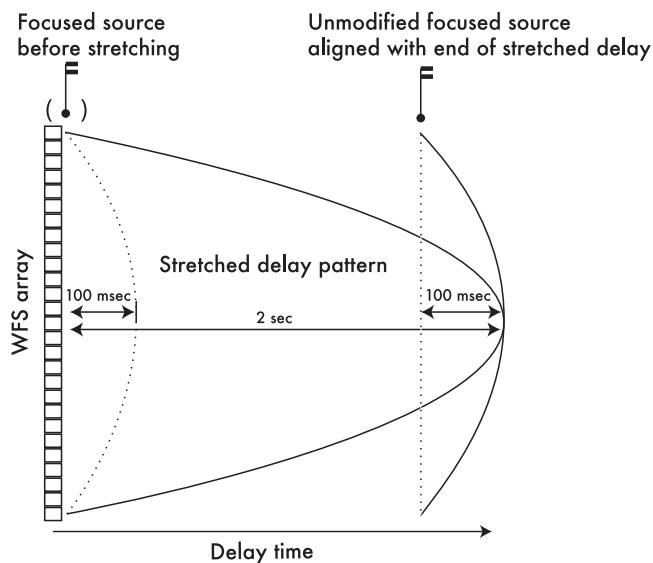
early reflections and late reverberation with attenuated direct signal. Then, as the listener entered into the ring of points, the directivity became orientated towards the listener, increasing the level of the direct signal path, producing an increased separation between points, and creating a distinct sense of being inside the shape. After the success of this experiment I began extending the approach, using perceptual-model signal processing to define stable group "entities."

In another experiment, I created a matrix of sources in a two-dimensional square in the plane of the WFS array, with each source having independent delay and bandpass filter. The delay time and center frequency were calculated based on each source's relative position in the group. Activating the matrix with a monophonic audio signal, the sound energy would propagate each point based on its delay time and filter frequency. Wavelike group movement was then created by offsetting the matrix point positions with 3-D spherical coordinates and projecting them onto the 2-D plane of the WFS array. Additionally, by varying the interpolation time of the delay, I found I could accentuate the natural frequency shifts for each point and increase the perceptual continuity of the group as a whole. Expanding on this processing approach for geometric source groups, I later used the "boids" flocking algorithm (Reynolds 1987)

to create group movements in 3-D with a similar approach to signal processing based on relative spatial position.

Timbre Space ⇆ Spatial Timbre

In the process of developing the "instrumental" spatial rendering techniques, it became clear that the programming required to work with HDLA systems has the side effect of separating spatial form from its sonic content. This is problematic, because the morphology of a sound itself has a strong influence on our perception of spatial rendering, although this separation of concerns also creates an opportunity to think through the relationship between sound and spatial rendering.

In his seminal work on "timbre space," David Wessel (1979) showed that by spatially representing the perceptual characteristics of timbre, stable relationships could be formed and used to create syntactic organizations of timbre. David Huron (1989) used a similar spatial representation to describe musical texture in terms of onset synchrony and semblant motion, in what he calls "texture space." In the spatial domain, a similar representation has been developed by Gary Kendall (2007) to describe perceptual grouping based on the number and correlation of sources.

While working with cello material developed with Séverine Ballon for my WFS- and HOA-based étude *Fluoresce*, I realized that there seemed to be an affinity between certain types of instrumental timbre and spatial processing morphologies. Complex timbres created by noise-influenced instrumental techniques tend to have a widely distributed frequency spectrum, where the low range can be below 20 Hz, blurring the boundary between instrumental timbre and contrapuntal texture. Similarly, a spatial-processing design can be thought of qualitatively as a "spatial timbre" or a spatiotemporal texture, where a point source or plane wave might correspond to a monophonic line, a flocking algorithm might imply polyphonic organization, and vice versa. From this prospective, Huron's texture space can be used as a common description method between sonic and spatial timbres and textures. For example, I

Figure 7. Map and
spherical representation of
energy distribution based
on video data
decomposition in
third-order HOA.
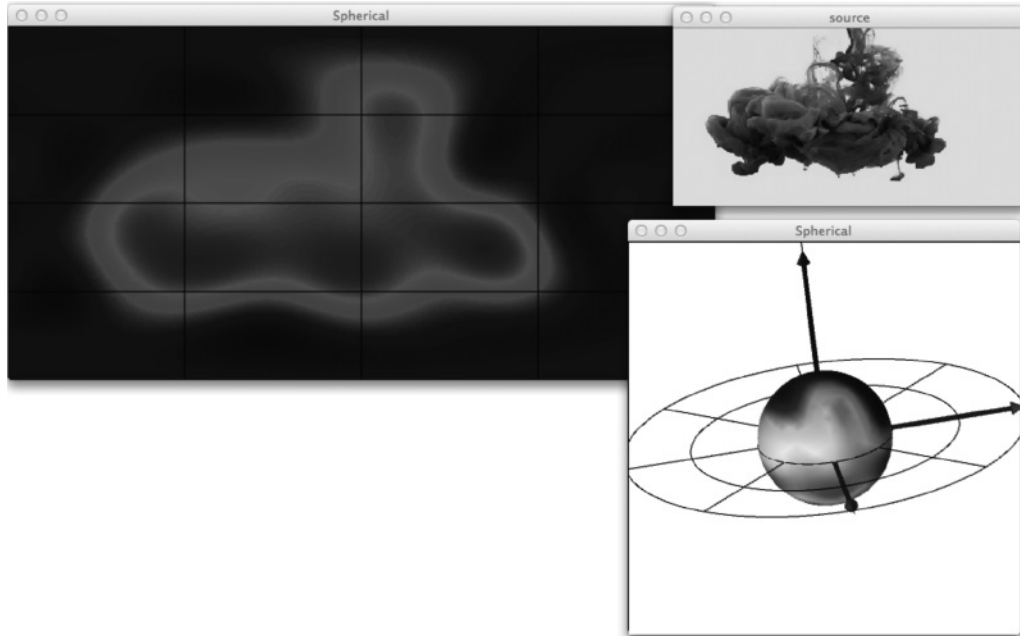
found that circular bowing on the cello (a wavering sound between noise and pitch) mapped well to the low-order spherical-harmonics technique described earlier, since both textures have a high degree of onset synchrony and slowly moving variance in what Huron calls "semblant motion." On the other hand, noisier granular textures mapped well to the boids approach, where both textures have a low degree of onset synchrony and semblant motion.

Technical Realization of *Fluoresce*

Because of the number of rendered channels and the real-time computational demands for spatial processing, extensive testing was required while prototyping the system to find an efficient solution for distributing the computation across the available processors. Figure 8 illustrates the audio and data flow in the system. A single instance of Max/MSP was capable of rendering approximately 25 sources in WFS or 25 sources in seventh-order Ambisonics, each with 24 early reflections encoded at third order (600 sources in total). By running three instances of Max per server machine, I was able to use parallel

CPU cores to handle the processing expense of the HOA image-source model, WFS "focus delays," low-order spherical-harmonics gammatone filters, and three flocks of "Boids," with the total CPU usage of each server averaging around 70 percent.

There was a distinct perceptual difference between sources rendered on the tightly packed WFS array versus the HOA array, which is more evenly distributed in the space. The physical presence of the WFS array and its higher spatial resolution gave sources an added clarity and almost tactile articulation in comparison to the seventh-order HOA. To blend HOA and WFS rendering, sources were cross-faded between the two systems, based on location. Sources located horizontally inside the perimeter of the room and vertically within 5 meters of the WFS array were rendered as focused sources. Reverberation and sources outside the WFS focused source region were handled by the 3-D HOA array.

### Representation in Spatial Composition

Composing for HDLA systems such as WFS and HOA, we are challenged to find spatial forms and

Carpentier et al. **29**

Figure 8. Audio and data flow design for the composition Fluoresce. Five rendering computers (bottom row) controlled by the client computer containing precomposed material in Ableton Live and Max for Live (center), controlled in real time by live input from on-stage cello and MIDI pedal (left). The client computer calculates the necessary gains and delays to coordinate the four WFS arrays and the HOA array.
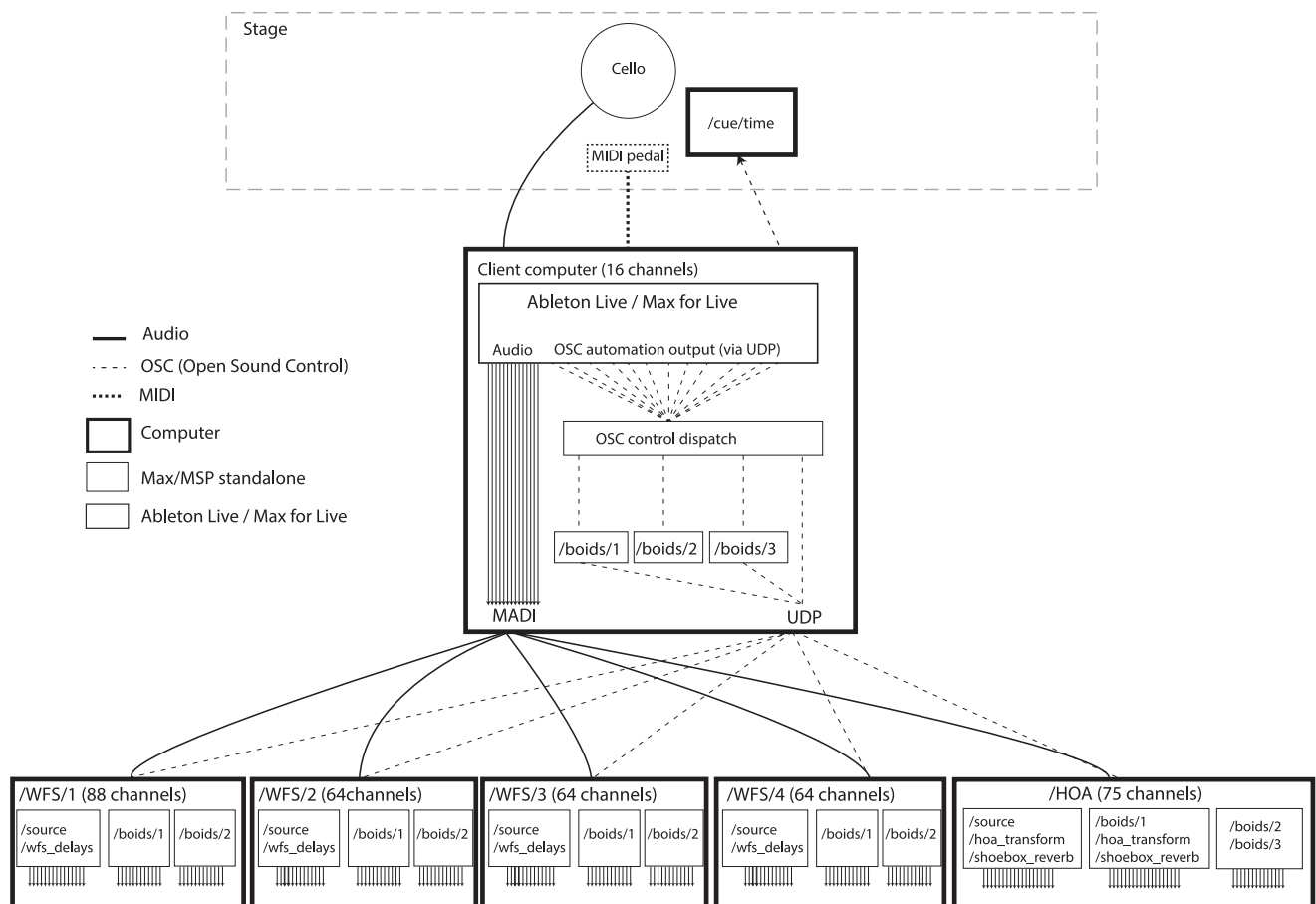
processes that make full use of the high-resolution rendering techniques available. As discussed earlier, correlating timbral and spatial morphologies is an important consideration in creating a perceptually coherent spatial scene, yet the technical requirements of composing for spatial-rendering systems tend to separate sound and spatial form. This exposes a question that deserves further consideration: How should a composer wishing to compose with space approach the symbolic representation of the relationship between sound and spatial form? Spatial movement is authored predominantly in the context of digital audio workstations, where spatial movement is typically represented as a single source's trajectory and with break-point automation curves used to control separate envelopes for each

parameter (e.g., azimuth, elevation, and distance). Because of this, the composer must very often think in terms of individual parameters rather than a complete 3-D spatial representation, in addition to the aforementioned separation between sound and spatial form. The development of digital score-authoring tools with more flexibility may assist in working through these challenges by providing a platform to graphically think through the relationship between sonic and spatial morphologies.

Like a blank piece of paper, a composition environment should provide the author an open space to develop symbolic representations capable of communicating the details of form and content to the reader, who interprets the symbols. The reader could be a human performer, musicologist,

or computational rendering system. An ideal digital notation system should address all three of these contexts. Some experimental progress towards a "blank page" notation system was achieved during my residency at IRCAM, using the Adobe Illustrator graphic-design environment and the Scalable Vector Graphics file format as a human-readable and data-rich notation (see Gottfried 2015).

Conclusion and Prospects

In this article we detail the implementation of the HDLA for the 3-D sound reproduction system installed in IRCAM's Espro, a variable-acoustics concert hall dedicated to both science and art. Through two artist residency reports we explored the affordances and challenges that were faced when working with such a system for music composition and performance art, and we discussed directions for future scientific and aesthetic research.

HDLAs are capable of rendering high-resolution spatial experiences far beyond the possibilities of traditional diffusion systems. By utilizing holophonic approaches such as WFS and HOA, HDLAs provide a means to expand the optimal listening area to include large audiences; and certain techniques, such as WFS focused sources and NFC-HOA, deliver high-resolution spatial images with enhanced depth perception, placing the sound and its behavior in a intimate relation to the listener. Recent production and experimental research at IRCAM and other HDLA locations suggest many potential creative applications of these technologies, and the growing interest among artists and institutions attest to the importance of further development.

Composition Challenges and Perceptual Research

As the use of HDLAs increases, we are challenged to master and apply rendering technologies in cognitively coherent ways, fusing spatial and sonic form. Advances in spatial signal processing now allow us to easily control large numbers of virtual sources, plane waves, and manipulate HOA microphone recordings, all in real time. Concep-tually, these advances encourage us to move away from "sounds as points" and towards "sounds as images," an enhanced sense of physical realism that expands compositional possibilities into the sculptural domain. Similarly, we are led towards a concern for the perceptual interdependency between spatial and sonic form. Although low-level computational approaches to HDLA already provide a high degree of control flexibility, composition for HDLAs could greatly benefit from more high-level authoring tools that take the psychoacoustic and multimodal context surrounding listener perception into account.

Technical Challenges

On the implementation side, there remain several areas that deserve further development. Rendering with NFC-HOA has the potential to create WFS-like proximity effects, although there remain unsolved problems of numerical stability in rendering the full range of the near-to-distant continuum. There are further questions regarding the perception of distance in context.

There are also challenges in the dissemination and portability of works between HDLA venues. Spaces housing HDLAs greatly differ in their dimensions, acoustic conditions, and technical configurations. Artists must therefore adapt their work to each listening situation. Although several technological and compositional strategies have been reported (see Peters, Marentakis, and McAdams 2011), there is yet no clear procedure for presenting computer music in a variety of HDLA systems. Future studies investigating effects of the acoustic environment on the reproduction quality of HDLAs may improve these issues, in particular taking cognitive aspects of listener perception (i.e., the interaction with visual or proprioceptive cues) into account.

There is also a clear need for compositional environments and audio production tools of greater flexibility while working with HDLAs. High-quality and high-resolution spatialization techniques, such as WFS and HOA, largely exceed the number of bus channels typically available in common DAWs and quickly expend the computational resources of any

personal computer. To accommodate the needs of high-quality spatial-rendering concepts within the DAW paradigm, the bus architecture needs to be extended to allow for significantly higher numbers of output channels. In the real-time paradigm, parallel computing should be exploited more effectively for handling intensive digital signal processing with large channel counts.

Preservation, Dissemination, and Exchange of Ideas

The issue of portability is also strongly connected to media format specifications. The conventional medium of channel-based audio files appears ill-suited for HDLAs and massively multichannel works. To accommodate the needs of spatial audio production, new descriptive formats are necessary. Several container formats have been recently proposed and standardized. For example, the Audio Definition Model (ITU-R BS.2076) is an open standard that supports object-based coding (i.e., audio tracks associated with descriptive metadata) as well as HOA storage. A flexible media format such as this could help by facilitating an increased exchange of spatial audio data and eventually fostering broader adoption of HDLAs and dissemination of works of art.

Similarly, new notation schemes and symbolic representation of spatial attributes may help in documenting conceptual approaches to spatial sonic form and context. Although descriptions of spatial design will never replace the actual experience of HDLA rendering, an increased presence of score-based symbolic representation could potentially provide a means to communicate the author's intention on paper for future scholarly and artistic study. As the preservation and exchange of ideas and approaches to spatial composition increases, new ideas are stimulated for yet-unthought-of uses for HDLAs.

Acknowledgments

References

Alcorn, M., C. Corrigan, and P. Rebelo. 2005. "SARC: Studio Report." In *Proceedings of the International Computer Music Conference*, pp. 519–522.

Amatriain, X., et al. 2007. "Immersive Audio and Music in the Allosphere." In *Proceedings of the International Computer Music Conference*, pp. 276–283.

Baalman, M. A. 2007. "On Wave Field Synthesis and Electro-Acoustic Music: State of the Art 2007." In *Proceedings of the International Computer Music Conference*, pp. 457–464.

Barrett, N. 2002. "Spatio-Musical Composition Strategies." *Organised Sound* 7(3):313–323.

Barrett, N. 2010. "*Kernel Expansion*: A Three-Dimensional Spatial Composition Combining Different Ambisonics Spatialisation Techniques." In *Proceedings of the International Symposium on Ambisonics and Spherical Acoustics*. Available online at ambisonics10.ircam.fr/drupal/files/proceedings/presentations/O17_10.pdf. Accessed January 2016.

Barrett, N. 2015. "Creating Tangible Spatial-Musical Images from Physical Performance Gestures." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 191–194.

Barrett, N. 2016. "A Musical Journey towards Permanent High-Density Loudspeaker Arrays." *Computer Music Journal* 40(4):1–12.

Berge, S., and N. Barrett. 2010. "High Angular Resolution Planewave Expansion." In *Proceedings of the International Symposium on Ambisonics and Spherical Acoustics*. Available online at ambisonics10.ircam.fr/drupal/files/proceedings/presentations/O15_18.pdf. Accessed September 2016.

Berkhout, A. 1988. "A Holographic Approach to Acoustic Control." *Journal of the Audio Engineering Society* 36(12):977–995.

Berkhout, A., D. de Vries, and P. Vogel. 1993. "Acoustic Control by Wave Field Synthesis." *Journal of the Acoustical Society of America* 93(5):2764–2778.

Carpentier, T., M. Noisternig, and O. Warusfel. 2015. "Twenty Years of IRCAM Spat: Looking Back, Looking Forward." In *Proceedings of the International Computer Music Conference*, pp. 270–277.

Cascone, K. 2000. "The Aesthetics of Failure: 'Post-Digital' Tendencies in Contemporary Computer Music." *Computer Music Journal* 24(4):12–18.

Chadabe, J. 1984. "Interactive Composing: An Overview." *Computer Music Journal* 8(1):22–27.

Corteel, E. 2007. "Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities, and Limitations." *EURASIP Journal on Advances in Signal Processing* 2007(1). Available online at link.springer.com/content/pdf/10.1155%2F2007%2F90509.pdf. Accessed September 2016.

Daniel, J. 2001. "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia." Ph.D. dissertation, Université de Paris VI.

Daniel, J. 2003. "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format." In *Proceedings of the AES International Conference: Signal Processing in Audio Recording and Reproduction*. Available online at www.aes.org/e-lib/browse.cfm?elib=12321. Accessed September 2016.

Daniel, J., and S. Moreau. 2004. "Further Study of Sound Field Coding with Higher Order Ambisonics." In *Proceedings of the 116th AES Convention*. Available online at www.aes.org/e-lib/browse.cfm?elib=12789 (subscription required). Accessed September 2016.

Favrot, S., and J. Buchholz. 2012. "Reproduction of Nearby Sound Sources Using Higher-Order Ambisonics with Practical Loudspeaker Arrays." *Acta Acustica United with Acustica* 98:48–60.

Gerzon, M. 1975. "Ambisonics: Part Two, Studio Techniques." *Studio Sound* 17(8):24–26.

Godøy, R. I. 2006. "Gestural-Sonorous Objects: Embodied Extensions of Schaeffer's Conceptual Apparatus." *Organised Sound* 11(2):149–157.

Gottfried, R. 2015. "SVG to OSC Transcoding as a Platform for Notational Praxis and Electronic Performance." In *Proceedings of the International Conference on Technologies for Music Notation and Representation*, pp. 154–161.

Huron, D. 1989. "Characterizing Musical Textures." In *Proceedings of the International Computer Music Conference*, pp. 131–134.

Jot, J.-M., and O. Warusfel. 1995. "A Real-Time Spatial Sound Processor for Music and Virtual Reality Applications." In *Proceedings of the International Computer Music Conference*, pp. 294–295.

Kendall, G. S. 2007. "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata." In *Proceedings of the International Computer Music Conference*, pp. 63–68.

Kolarik, A. J., et al. 2016. "Auditory Distance Perception in Humans: A Review of Cues, Development, Neuronal Bases, and Effects of Sensory Loss." *Attention, Perception, and Psychophysics* 78(2):373–395.

Menzies, D. 2002. "W-Panning and O-Format, Tools for Object Spatialization." In *Proceedings of the AES International Conference: Virtual, Synthetic, and Entertainment Audio*. Available online at www.aes.org/e-lib/browse.cfm?elib=11151 (subscription required). Accessed September 2016.

Noisternig, M., T. Carpentier, and O. Warusfel. 2012a. "Dispositif de spatialisation sonore 3D à l'Espace de Projection de l'IRCAM: Un Réseau de 345 haut-parleurs pour une restitution par WFS et HOA." *Acoustique et Techniques* 71:30–39.

Noisternig, M., T. Carpentier, and O. Warusfel. 2012b. "Espro 2.0: Implementation of a Surrounding 350-Loudspeaker Array for Sound Field Reproduction." In *Proceedings of the AES UK Conference: Spatial Audio in Today's 3D World*. Available online at www.aes.org/e-lib/browse.cfm?elib=18121 (subscription required). Accessed September 2016.

Peters, N., G. Marentakis, and S. McAdams. 2011. "Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative Analysis." *Computer Music Journal* 35(1):10–27.

Peutz, V. M. A. 1978. "The Variable Acoustics of the 'Espace de projection' of IRCAM (Paris)." In *Proceedings of the 59th AES Convention*. Available online at www.aes.org/e-lib/browse.cfm?elib=3044 (subscription required). Accessed September 2016.

Pulkki, V. 1997. "Virtual Sound Source Positioning Using Vector Base Amplitude Panning." *Journal of the Audio Engineering Society* 45(6):456–466.

Radeau, M., and P. Bertelson. 1974. "The After-Effects of Ventriloquism." *Quarterly Journal of Experimental Psychology* 26(1):63–71.

Ramakrishnan, C., J. Gossmann, and L. Brümmer. 2006. "The ZKM Klangdom." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 140–143.

Reynolds, C. W. 1987. "Flocks, Herds, and Schools: A Distributed Behavioral Model." *ACM SIGGRAPH Computer Graphics* 21(4):25–34.

Wessel, D. 1979. "Timbre Space as a Musical Control Structure." *Computer Music Journal* 3(2):45–52.

Wessel, D. 2006. "An enactive approach to computer music performance." In *Actes des Rencontres musicales pluridisciplinaires: Le Feedback dans la création musicale*, pp. 93–98.

Wilson, S., and J. Harrison. 2010. "Rethinking the BEAST: Recent Developments in Multichannel Composition at Birmingham ElectroAcoustic Sound Theatre." *Organised Sound* 15(3):239–250.

Wright, M. 2005. "Open Sound Control: An Enabling Technology for Musical Networking." *Organised Sound* 10(3):193–200.

Zmölnig, J. M., A. Sontacchi, and W. Ritsch. 2003. "The IEM-Cube, A Periphonic Re-/Production System." In *Proceedings of the AES International Conference on Multichannel Audio: The New Reality*. Available online at www.aes.org/e-lib/browse.cfm?elib=12295 (subscription required). Accessed September 2016.

Zotter, F., H. Pomberger, and M. Noisternig. 2010. "Ambisonic Decoding with and without Mode-Matching: A Case Study Using the Hemisphere." In *Proceedings of the International Symposium on Ambisonics and Spherical Acoustics*. Available online at www.researchgate.net/publication/228669508. Accessed September 2016.

Zotter, F., H. Pomberger, and M. Noisternig. 2012. "Energy-Preserving Ambisonic Decoding." *Acta Acustica United with Acustica* 98:37–47.