White Paper presented at IRCAM Forum Workshops, February 2021

Spatial Sculpting: Computer-aided Composition of Spatial Sound Shapes and Textures

Marlon Schumacher*† Núria Giménez-Comas‡

> *The power of the computer to help us construct the internal architecture of sounds from first principles allows us to broaden the concept of composer to include the notion of sonic sculpture.*
>
> Trevor Wishart, *On Sonic Art*

Abstract

We report on our efforts to further elaborate and investigate the integration of sound synthesis algorithms with spatial sound rendering as part of the compositional process. The primary motivation for this collaborative, artistic research project focuses on the idea of creating "Spatial Sound Sculptures", i.e. not only the sensation of spatial sounds with volumetric properties (e.g. shape, size, etc.), but as a consequence, methods and tools for rendering sound layers and textures for active listening situations.

The *Spatial Sound Synthesis* framework for computer-aided composition (CAC), *OMPrisma*, was extended with new control tools and a refactored architecture for improved flexibililty and ergonomy. For handling the complexity of parameter specification, graphical interfaces for visualization and manipulation of spatial properties, such as directivities and distributions are presented. Aligning perceptual considerations with musical concepts, examples are presented considering aspects of both synthesis and auditory display (sensory-motor integration, listener engagement).

Aided by these new possibilities, a number of artworks have been created by the second author, including an architectural piece, an installation study, and an acousmatic piece. In these pieces, the synthesis of close-proximity sources (distinct sound objects) and distant textures (diffuse soundscapes) was realized by combining different spatialization systems (Ambisonics and Virtual Microphone Control) in an unorthodox loudspeaker configuration, reminiscent of a promenade-like setting. Audience members were invited to freely move within the space and actively explore the spatial sound compositions.

---

*Hochschule für Musik, Karlsruhe

†Centre for Interdisciplinary Research in Music Media and Technology, CIRMMT

‡Institut de Recherche et Coordination Acoustique/Musique, IRCAM

FIGURE 1: Prototype loudspeaker configuration @ *MUTProbe 2*, HfM Karlsruhe, September 2017.

Introduction

Composition of spatialization comes into play as soon as spatial attributes of sound become an integral part of the compositional thought. Computer-aided composition is a field which is traditionally concerned with providing frameworks of tools and interfaces for the generation and manipulation of discrete symbols representing domain-specific (musical) data. Drawing from earlier efforts of extending these frameworks by integrating multi-dimensional models for sound, space, and gesture [Schumacher, 2016], this project aims at further developing tools and methods for spatial composition of higher-level spatial attributes.

Since the renewed momentum of spatialization techniques following advances in signal processing and computing power, there has been an increased interest in the computer music community for conceiving spatialization beyond what is essentially the animation of point-like sources in a "scene". To name but a few: circumspectral [Torchia & Lippe, 2004, Kim-Boyle, 2008, Khosravi, 2014], timbral [Normandeau, 2009], granular [McLeran et al., 2008], and environmental/statistical [Verron et al., 2010]. Most commonly, current approaches consider spatialization rouhgly as a process of synthesizing perceptually distinct versus diffuse sound events surrounding the listener. In this project we were specifically interested in (computer-aided) composition of spatial sound sculpture, including its conceptualization, rendering and perception

This article is divided into two major parts: the first is dedicated to research and development (by the first author); the second to creative process and artworks (by the second author). We first discuss the notion of a "spatial sound sculpture", from which we draw definitions and criteria for lower-level concepts informing our development efforts (Section 1). We then discuss general design guidelines and present control tools and extensions for our spatialization framework (Section 2). Special consideration is given not only to the synthesis aspects, but – informed by perceptual considerations – qualities of the auditory display method (Section 3). We conclude the first part with a presentation of new extensions to the DSP-architecture and demonstrate an example application (Section 4). Next, we focus on the artistic part of this project. We begin with compositional motivations (Section 5) and their applications in real-world artistic works (Section 6). Following this, we move on to reporting the usage and challenges of the aforementioned developments in the creative process (Section 7), before finally detailing considerations and potential future work (Section 8).

1 Spatial Sound Sculpture

Approaching the notion of sculpture from a visual or fine arts understanding, the Merriam-Webster dictionary defines “sculpture” as

> a: The action or art of processing (as by carving, modeling, or welding) plastic or hard materials into works of art.
> b: (1) A work produced by sculpture, (2) a three-dimensional work of art (such as a statue).

In the domain of music and sonic arts, definitions are various. Harry Partch’s concept of *corporeal music*, for instance, names three core characteristics of sound sculpture: 1) physicality of sound, 2) totality of aesthetic experience of the artwork, and 3) audience engagement on a visceral level. According to Keylin, *movement* plays a prominent role, including: movement of the parts of the sculpture or the performer’s gesture; sound moving through space; or listeners moving through spatially arranged sound. The abundance of movement ensures a visceral, as opposed to merely intellectual, experience [Keylin, 2015].

Within the context of spatial sound synthesis, we transfer the notion of a plastic object as a virtual (perceptual) sonic entity, defined by its characteristic three-dimensional *shape*. Further, borrowing from computer graphics, the characteristic “materiality” of the object as a surface feature, i.e. a *texture* (such as in the design of rendered “wallpapers”). The visceral audience engagement is transferred as *dynamicism* and *movement* – of both sonic content and audience. A detailed discussion of the reasoning behind these conceptual mappings is beyond the scope of this article, thus we limit ourselves in the next section to the discussion of concepts necessary for realizing the metaphor of *spatial shapes* and *spatial textures*. Visceral audience engagement is addressed in Section 3.

1.1 Spatial Shapes

Let us first consider how to define and apply attributes usually associated with plastic objects, such as volume and shape, to a temporal phenomenon such as a pressure wave moving through air (sound). The spatial audio literature reveals substantial interest in the perception and rendering of such ideas, aiming at determining physical and signal-based correlates with perceptual attributes. Kendall [Kendall, 2010], for instance, divides the attributes of *spatial imagery* following the framework of Rumsey into *dimensional* and *immersive* attributes [Rumsey, 2002]. We will focus on the former, source-related attributes.

The most common technique for rendering sound sources with spatial extension is the multi-source approach, in which spatially displaced copies of the original sound source are generated and their signals artificially decorrelated in phase (this can be seen as an extension of classical stereo-widening techniques in audio recording and post-production [Cabrera, 2013]). The goal is to achieve a wider angular “spread” and to decrease the IACC (interaural cross correlation), which is perceptually correlated with apparent source width (ASW), i.e. the aurally perceived spatial extent of a source [Kendall, 1995, Potard & Burnett, 2004, Zotter et al., 2011, Schmele & Sayin, 2020]. In his classification of spectral sound shapes, Stuart, for instance, has defined a categorical set of multi-point

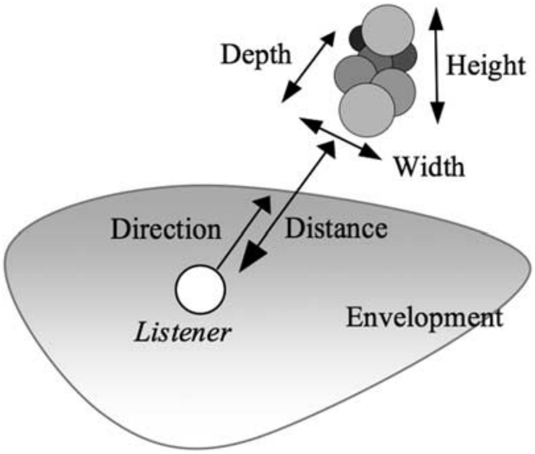


FIGURE 2: Illustration of attributes of spatial imagery. Source: [Kendall, 2010]

distributions [James, 2016a]. Drawing from the fields of composition theory, auditory scene analysis, and psychoacoustics, he transfers the broader notion of *de-composition* and *re-composition* (see also [Stefani & Lauke, 2010]) to the spatial domain, as a sound's reconstruction that "...can take not only the amplitude and phase of each frequency into consideration, but may also determine the spatial redistribution of each frequency."

Accordingly, we can envisage the two following complementary methods, analogous to sound synthesis techniques: a subtractive (or "source-filter") approach, i.e. taking a rich source and "carving out" certain parts (usually by filtering energy in certain frequency regions of the spectrum), or an additive approach, i.e. "modelling" a sound by summation of individual (smaller-scale) sound components. We adopt this dichotomy to the rendering of "spatial shapes" as:

- **subtractive:** by "splitting" an omnidirectional source into multiple copies (e.g. filterbands), and assigning directivity patterns to them;
- **additive:** by modeling a spatial shape via superimposition of individual sound components with individual directivity patterns.

1.2 Spatial Textures

Amongst the plethora of sound synthesis applications, the notion of *sound texture* has received special attention. Schwarz, for instance, in his overview of sound texture techniques [Schwarz, 2011], borrows definitions from Saint-Arnaud and Popat:

> *A sound texture is like wallpaper: it can have local structure and randomness, but the characteristics of the fine structure must remain constant on the large scale. Source: [Saint-Arnaud & Popat, 1995], p. 294*

A list of five defining criteria for *sound texture* is established, loosely based around the concept of information content:

1. Sound textures are formed of basic sound elements, or atoms;
2. Atoms occur according to a higher-level pattern, which can be periodic, random, or both;
3. The high-level characteristics must remain the same over long time periods;
4. The high-level pattern must be completely exposed within a few seconds;
5. High-level randomness is also acceptable, as long as there are enough occurrences within the attention span to make a good example of the random properties.

Schwarz further states that in cases of music composition the term *texture* is often used negatively to exclude traditional concepts (such as pitch, rhythm, etc.). Indeed, musical concepts depending on the interaction of multiple parameters (such as texture, structure, timbre, entropy, etc.) however, seem to lack formal definitions and lend themselves more toward subjective interpretation and idiosyncratic definitions.

In his categorization of "Klangtypen neuer Musik" (sound-types of contemporary music), composer H. Lachenmann groups together *Farbklang* (timbre-sound), *Fluktuationsklang* (fluctuation-sound), and *Texturklang* (texture-sound) as stationary sounds with respectively increasing degrees of complexity; the difference being in the statistical qualities of their inner morphology. While the timbre-sound is characterized by inner stasis and the fluctuation-sound by periodic oscillations, the texture-sound exhibits somewhat more aperiodic qualities while also revealing a global statistical distribution or pattern. Its perception does not depend upon the global duration of unfolding inner relationships, but is rather experienced after grasping its higher-level statistical characteristics [Lachenmann, 1996].

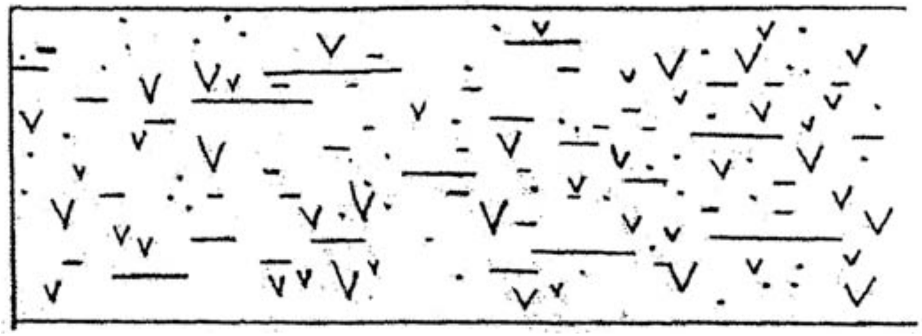


FIGURE 3: Schematic Drawing of a "Texturklang" [Lachenmann, 1996].

Taking these ideas further to the domain of spatial music, it might be sensible to take a look at theories of acousmatic music composition as spatial aspects are inherent to its performance practice, the loudspeaker diffusion. In his writings on "spectromorphology" and "spatiomorphology", composer D. Smalley defines *spatial texture* as a characteristic temporal quality, i.e. how global spatial characteristics are revealed over time [Smalley, 2001]. A distinction is made between *contiguous* and *non-contiguous* space: the former being, for example, continuous motion through space; the latter, for example as spatial discontinuity between successive events. Smalley further adds that those two categories are not mutually exclusive, but depend on the scale (global vs. local) of observation:

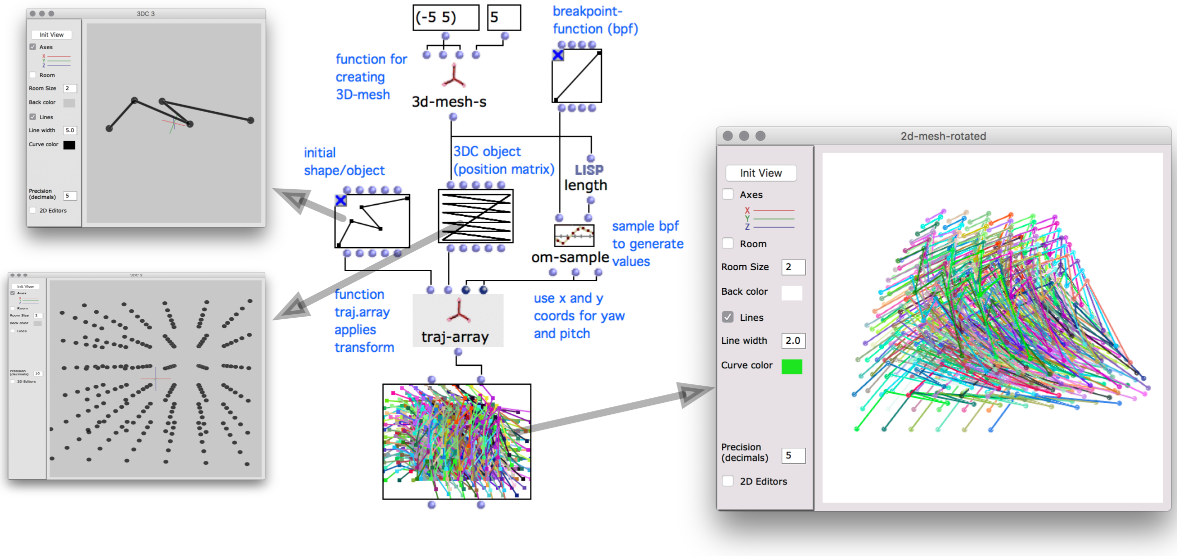
> *"...taken over a sufficiently long duration, the individual non-contiguous points of this texture, perceived as a whole, cover a contiguous space."*

We adopted these notions of spatial textures for our project and will later in this article demonstrate examples for realizing *contiguous* and *non-contiguous* spatial textures with our spatialization framework.

2 Control Strategies

Let us now discuss our developments for implementing the aforementioned concepts. Following the tradition of CAC systems being designed as adaptable and extensible environments for creative musical programming independent of personal aesthetics, our goal is to propose tools and interfaces which are versatile and adaptable, while ergonomic and intuitive as possible. Accordingly, rather than providing application-specific interfaces, we extend our toolset with functions for metaprogramming and manipulation of spatialization data as abstract geometrical objects.

Borrowing from Section 1.2, the notion of *sound texture* can be loosely defined as a morphology formed by smaller-scale objects (*atoms*), which are distributed and possibly transformed according to some higher-level pattern. Transferred to the spatial domain, i.e. atoms being spatial objects (such as 3D-vectors for trajectories, rotations, etc.), implementing this notion requires functionalities for representing and manipulating abstract geometrical data on multiple hierarchical levels. Figure 4 shows the new *traj-array* (trajectory-array) function, which accepts geometric objects (such as 3D-trajectories, point-distributions, shapes, etc.) as inputs, and computes new instances while applying global transforms (such as translations, rotations, etc.) controlled by parameters set via other high-level objects and specifications. Note, that this process can be applied recursively at different levels.



☐ Axes
☐ Room
Room Size 2
Back color
☑ Lines
Line width 2.0
Curve color 🟩
Precision (decimals) 5
☐ 2D Editors

FIGURE 4: Implementation of concepts of *texture* from Section 1.2.

The *traj-array* function is somewhat similar to the *jit.gl.multiple* object in Cycling 74's *Max* Software

Following a similar approach for volumetric modeling, we have used geometrical objects and algorithms for generating 3D-directivity patterns. Figure 5 shows the generation of a 3D directivity pattern as the superposition of eight directional "lobes"; the function *microphone* (*a*) computes classic microphone polar responses as 2D-objects (gain as a function of incidence angle) with a specific horizontal orientation (”yaw”) and "sharpness" Values for both parameters are randomly specified by the *om-random* functions. These 2D-objects are then converted into 3D-objects by rotation (”rolled” 180 degrees around the forward-facing axis, *c*). Each of them is rotated vertically (”pitch”, *e*) specified via a sampled breakpoint-function (*bpf*) object using the *om-sample* function (*d*). Since all spatial rendering classes in *OMPrisma* share the *SpatDIF* principle of *gracefully failing* [Bresson & Schumacher, 2011, Peters et al., 2013], these geometric specifications can be employed interchangeably and independently for any of the spatialization techniques implemented in the *OMPrisma* class library.

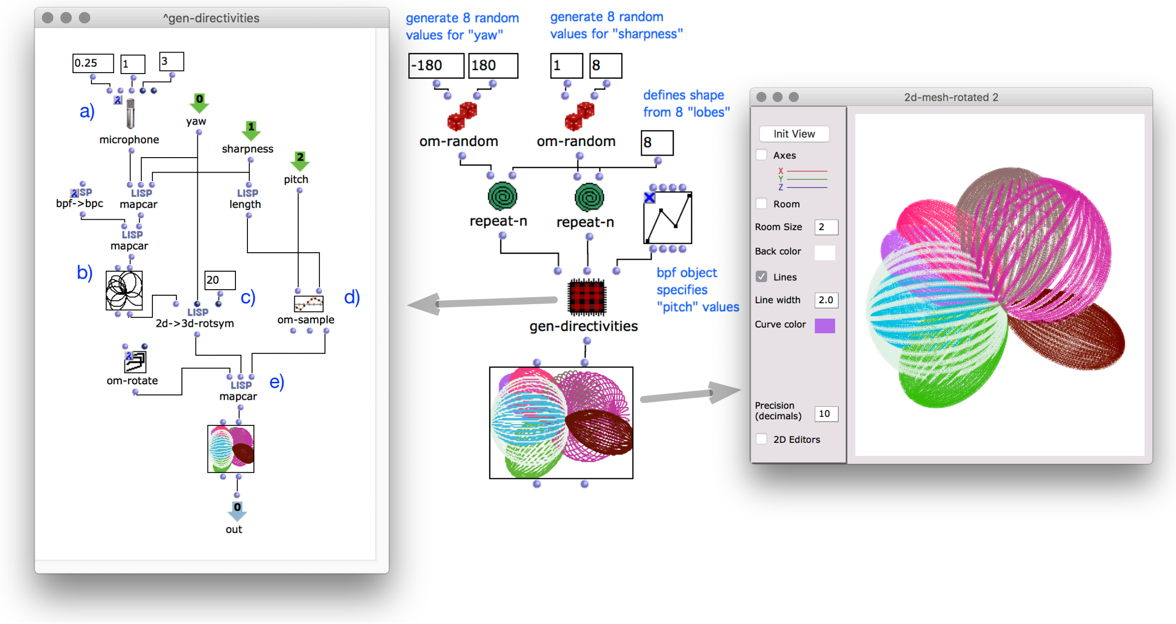


FIGURE 5: Implementation of concepts of *shape* from Section 1.1.

3 Auditory Display

As intriguing as the visual analogy of a material *sculpture* with spatial attributes like extension, shape, etc. may be, the distinct characteristics of the visual versus auditory perceptual modalities need to be carefully considered for, as well as the interplay of sound and spatial attributes is unsurprisingly far from orthogonal and might interfere with or create more or less coherente sensations [Senna et al., 2017]. Contributing further to this need, the perceptual system tends to gravitate towards redundancy for the grouping of sound events following perceptual heuristics [Bregman & Steiger, 1980, Bregman, 1994].

E.g. spatiotopic coding of the receptors on the retina (visual), vs. deconvolution of acoustic cues that arise at each ear and between the ears (auditory).

Keeping this in mind, consider dealing with the rendering of complex sound shapes and textures across an extended spatial area. Here, it becomes particularly important to not only focus on the synthesis aspects, i.e. how spatial sounds are rendered, but also how they are presented, i.e. aurally displayed to the listener. For three-dimensional listening situations, the visual analogy is confounded further when not all auditory cues (in particular for perceptually less stable dimensions, such as distance and depth [Coleman, 1962]) are equally-well provided. In fact, most common 3D-spatialization techniques (e.g. the family of amplitude-panning techniques, basic ambisonics, etc.) are based on ICLDs (interchannel level-/phase differences) rendered for a single spatial position and generally do not have a dedicated representation of distance or depth. Therefore, such 3D-spatialization might be more accurately described as a "flattened" projection (onto horizontal and vertical angular dimensions), comparable to the projection of graphical 3D-objects on a 2D-screen. This "reduced information" can create unstable images, defeating the synthesis efforts (analogous to optical illusions). See Figure 6 for a visual analogy.

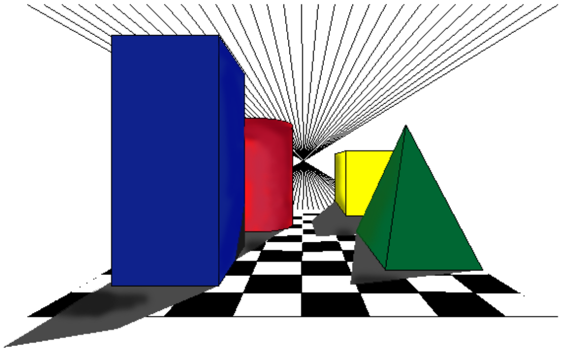


FIGURE 6: Visual Representation of 3D objects on a 2D-dispay. Source: [Wittek et al.,].

While much research on extended sources has focused on angular distributions or "spreading" (e.g. [Verron, 2009, James, 2016b, Pihlajamäki et al., 2014]), for our project (dealing with spatial textures and volumes) we found it essential to be able to render depth and distance.

3.1 Sensory-Motor Integration in Auditory Spatial Perception

Much of the classical literature on human hearing is based on experiments using test signals (such as noise bursts or sine tones) presented to stationary subjects, thereby isolating auditory cues. More recent findings in auditory neuroscience suggest however, that sensory-motor integration plays a significant role in auditory perception (e.g. integrating head position and binaural cues) and can aid in the deconvolution of dynamic spatial audio environments [Carlile & Leung, 2016, Genzel et al., 2018].

Take the example of looking into an aquarium versus being immersed within it: it seems intuitive, that the full-body engagement with the environment (perception-action

---

It is interesting to note, that in contrast to *distance*, which is a source-related attribute, *depth* is regarded as a perceptual attribute of the acoustical scene involving multiple cues [Rumsey, 2002].

link) would facilitate a more accurate mental representation. While *Enactivist* approaches to cognition have been mostly concerned with the visual sensory modality, there is neurophysical evidence that auditory motional cues, such as acoustic parallax, as well as enaction with the environment can provide a better sense of depth and should therefore aid in the perception and/or identification of spatial sound synthesis scenes [Yost, 2018].

Approaching the above tactically, the use of these attributes would require a spatialization and auditory display system to provide motional auditory cues (parallax), as well as locomotive, vestibular/proprioceptive cues, i.e. allowing listeners to move and actively change position and angle relative to sources. For a first assessment study, the first author designed an auditory display configuration with spatially displaced loudspeakers, allowing free listener movement between them. Notably, this configuration aligns well with the requirements of *sound sculpture* as stated by H. Partch (Section 1).

An interesting rendering technique for driving such a loudspeaker configuration is *Virtual Microphone Control* (ViMiC). In this technique, a virtual scene is "picked up" by an array of (possibly spaced) virtual microphones, each with its individual parameters (position, orientation, directivity pattern) [Braasch et al., 2008]. The virtual microphone signals can then be directly mapped to loudspeakers at the corresponding physical positions (therefore replacing a virtual "sink" with a physical output). ViMiC is also capable of simulating other spatialization concepts, such as wavefield-synthesis or ambisonics. In the case of our spatially displaced loudspeaker configuration we could think of a hybrid constellation, providing sparsely distributed "sample points" into a virtual sound scene. Through manipulation of microphone (sink) parameters (e.g. directivity patterns, gains, etc.), it also provides great flexibility for fine-tuning global characteristics of the spatialization process.

3.2 Example of Non-contiguous Spatial Texture Synthesis

Let us now look at a first example for non-contiguos spatial texture synthesis. The *OpenMusic* patch illustrated in Figure 7 shows an *OMPrisma* class for spatial granular synthesis via *ViMiC (a)*. A graphical user interface allows the specification of an array of virtual microphones and respective parameters *(b)* (in red colour), the editor window is visible on the top right. The input parameters of the granular synthesis process, visible on the left-hand side (in blue colour) are used to set the parameter matrix of the ViMiC class (see [Agon et al., 2011] for detailed descriptions). In this example, 300 grains are extracted from a source soundfile and arranged in space and time according to periodically looped parameter lists. The specification of spatialization parameters is illustrated on the right hand side of the patch *(c)* (in green colour). Aligning with the definitions from Section 1.2, starting from an initial "spatial pattern" a number of deviated copies are generated and superimposed, therefore while exposing local variations, high-level characteristics (exposed within a few second) are retained. Note, that the parameter-lists have prime-number lengths, therefore the coinciding of parameter values (e.g. same sound at the same position) is unlikely, while the larger scale characteristics are preserved; this is also nicely represented graphically within the ViMiC class' *miniview* as sawtooth-shapes with different phase-lengths.

A more detailed discussion of design and implementation specifics is beyond the scope of this paper; the reader is referred to [Stuchlik, 2017] for further information.

Finally, the function *display-vimic-3d* visualizes source and microphone parameters in a *3DC*-object *(d)*, the editor is visible at the right bottom. The patch for this example and resulting sounds are available for download and individual exploration, see Section 8. The physical loudspeaker configuration used in the patch is shown in Figure 1.

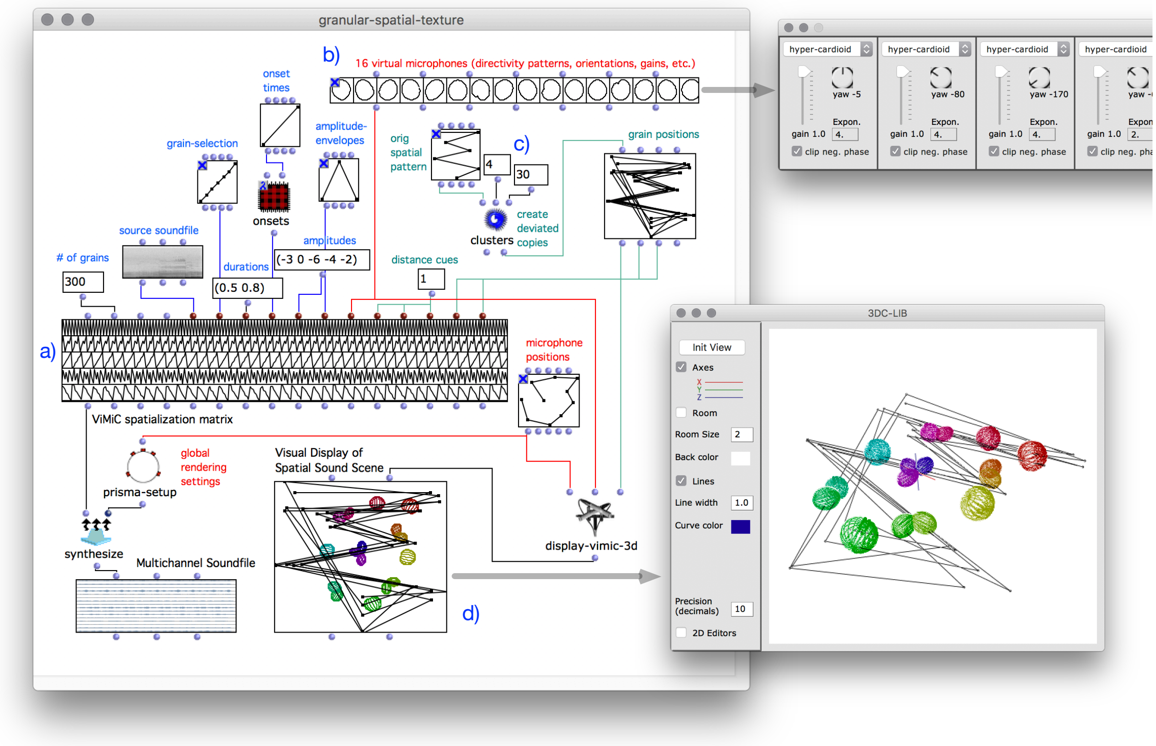
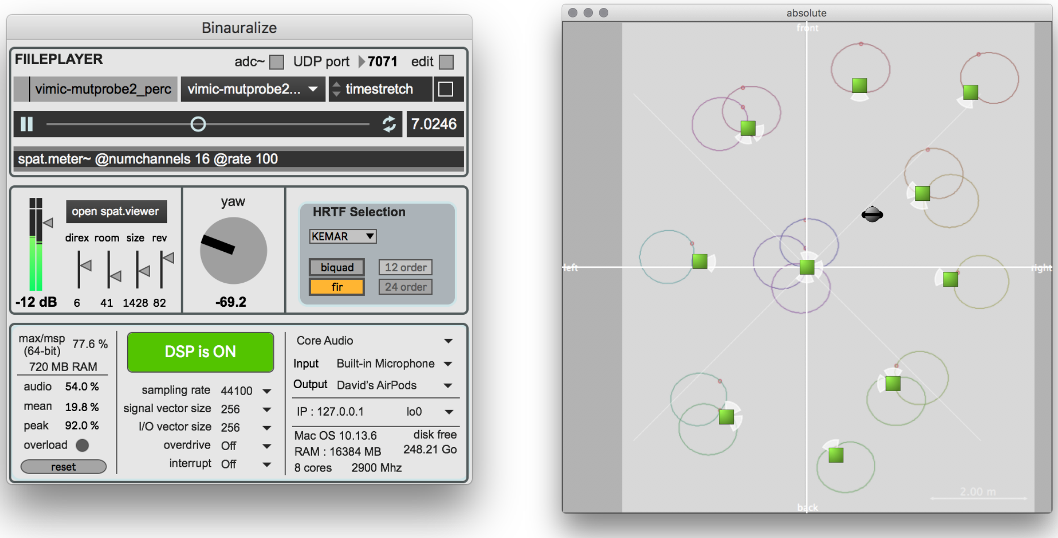


FIGURE 7: Non-contiguous Spatial Texture Synthesis with *OMPrisma* class *ViMiC.discrete*. Top-right: editor interface for specifying virtual microphone parameters. Bottom-right: 3D visualization of ViMiC configuration and source positions.

3.3 Virtualization of Auditory Display System

In order to be able to experiment with and develop arbitrary (spaced, on-stage, etc.) loudspeaker configurations without their physical presence, some tools have been implemented by the first author, allowing for (off-site) binaural rendering via loudspeaker virtualization using IRCAM's SPAT software [Carpentier, 2015]. The virtual microphones (representing speaker feeds), with their respective positions, orientations, and directivities, are transmitted from the composition environment and represented as sources with respective positions, orientations, and "apertures". The configuration used for the example in the previous section is shown in Figure 8b. The virtual "listener" can be moved freely within this virtual space (using mouse, or keyboard), therefore adapting the binaural rendering in real time. In addition to the physical loudspeaker configuration it offers optional adjustment of room characteristics using graphical user interfaces, shown in Figure 8a. The integration of generic head-tracking is part of future developments.



Binauralize

FIILEPLAYER

adc~ ☐ UDP port 7071 edit ☐

vimic-mutprobe2_perc

vimic-mutprobe2...

timestretch ☐

7.0246

spat.meter~ @numchannels 16 @rate 100

-12 dB

direx 6

room 41

size 1428

rev 82

yaw -69.2

HRTF Selection

KEMAR

max/msp (64-bit) 77.6 %

720 MB RAM

audio 54.0 %

mean 19.8 %

peak 92.0 %

overload ●

DSP is ON

sampling rate 44100

signal vector size 256

I/O vector size 256

overdrive Off

interrupt Off

Core Audio

Input Built-in Microphone

Output David's AirPods

IP : 127.0.0.1 lo0

Mac OS 10.13.6

RAM : 16384 MB

8 cores 2900 Mhz

disk free 248.21 Go

(A) Player/Settings Graphical User Interface.

(B) Bird's eye view with "listener" avatar.

FIGURE 8: Binauralize application. On the right: Loudspeaker configuration from 7.

4 Modular Framework Extension

OMPrisma [Schumacher & Bresson, 2010] is a library for the *OpenMusic* environment for computer-aided composition [Bresson et al.,]. The library provides high-level control interfaces and a rich collection of spatialization instruments, which wrap underlying DSP-code in the Csound language into *classes* with object-oriented features (such as polymorphism and multiple inheritance). The metaprogramming tools provided in its underlying environment (*OpenMusic*) allows for "merging" individual DSP-instruments, such as sound generating sources (synthesizers) and spatial sound renderers, into compound structures by re-writing underlying DSP-code and defining new classes just-in-time [Schumacher & Bresson, 2010]. This framework has been extended with a plugin-architecture, allowing for the definition of *processor classes* which are inserted between sound generating and spatial sound rendering classes. Processor classes can be daisy-chained, allowing for the design of *compound classes* with underlying (serial) dsp-topologies of arbitrary complexity. Users can therefore integrate the "design" of custom spatialization instruments as part of compositional considerations, rather than relying on preexisting ones (which might not always fit or match the required characteristics). From a workflow perspective, this architecture provides greater structure and flexibility, allowing quicker prototyping and reducing the cognitive load related to implementation-specific details of the underlying signal processing. While being a relatively new feature, the library of processor classes has already grown to include filters, modulators, decorrelators, resonators, to name a few examples.

4.1 Example of Contiguous Spatial Texture Synthesis

Let us now look at an example for synthesizing a contiguous spatial texture. Figure 9 shows an *OpenMusic* patch combining some of the aforementioned tools and functionalities. We follow propositions from [Potard & Burnett, 2004] for rendering sound source extent by decomposing the original source into several spatially distinct point sources. The original sound source is bandfiltered into 14 "components" (corresponding to perceptual (bark) frequency-bands) and each of the components' phase-spectrum is dynamically decorrelated for achieveing a perceptually "broad" sound image. This is achieved by daisy-chaining a sound generator class (*smpl-1*), two processor classes: an IIR-bandfilter (*reson.discrete*) and a second-order all-pass filter (*phaser.discrete*), finally, a spatial sound renderer (*vimic.continuous*). The spatialization parameters of this multi-point source are generated using the tools described in Section 2: an inital trajectory (*atom*) is copied onto spatial locations determined by a 2-dimensional mesh. Each of the 14 point-sources is assigned an individual trajectory, while the ensemble of trajectories form a larger-scale global pattern (visible at the bottom right).[5] It is worth noting, that although in this example the sound and spatial data are created in independent processes, they could easily be linked. In fact, certain relationships and possibilities might emerge explicitly due to embedding both processes in the same formalism (program).

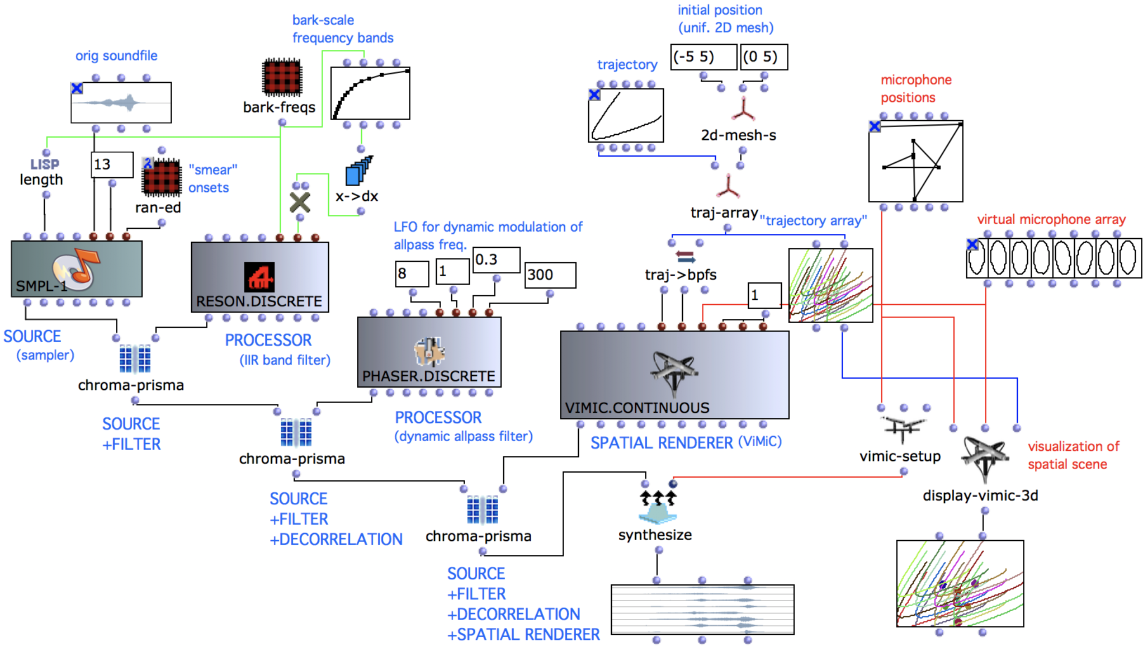


FIGURE 9: Example patch showing the class-merging extension for creating a *contiguous* spatial texture. Note, the chain of classes from top-left (smpl-1) to bottom-right (vimic.continuous).

---

[5] The patch for this example and resulting sounds are available for download and individual exploration, see Section 8.

5 Compositional Motivations

A number of artworks by the second author have been inspired by the idea of combining spatialization systems in an unorthodox loudspeaker setup in order to synthesize a broad spectrum of "spatial sound objects" – ranging from close-proximity sound sources to distant textures or ambiances – which the listener is free to individually explore and appreciate. The interest is to create not only the sensation of given sound sources projected into space, but multiple spatial sound morphologies around and within the audience that are increasing in density in different emerging sound layers. These artistic ideas have motivated the development of new tools in the *OM-Spat* library [Agger et al., 2017], as well as the connection of these new tools in *Max* to monitor and adapt the various synthetic textures by perceptual evaluation (e.g. by mixing and fine-tuning in different studios and venues, such as IRCAM's studio 5 (see Figure 14), studio 1, and the "Kubus" at ZKM, Karlsruhe [Ramakrishnan et al., 2006]).

In the early stages of this project, a new *Spatial Sound Synthesis* class was defined. This class implements a spectral sound model based on harmonic content (partials) and noise (residual) [Serra & Smith, 1990], which lends itself particularly well to transformation and mapping to spatial resynthesis algorithms.

Initial Experiments: First Artwork

For the piece *Back into Nothingness*, some of these ideas were elaborated for the electronics part of the poetic and spatial dramaturgy of the piece: resynthesis as a "sound metaphor", as well as first sketches for spatial granular synthesis.

The idea of a "sound metaphor" stems from the resynthesis of poetic "images" from the text; transforming data of concrete sounds and extending a more or less "blurred" sound image in space. This uses the idea of spectral spatialization, e.g. focalizing noisy parts of a sound in one part of space while expanding sinusoidal parts in different parallel trajectories. The composition created a dialog of two spaces: one on the scene with three speakers placed high up in different angles and orientations, and the other around the audience, as a surrounding (immersive) circle. These two spaces could interact and alternate at some dramaturgical moments. The scene speakers could also dialog with the actress and the choir placed on the scene.

In this dialogue, as well as in the resynthesis of concrete sounds, granular synthesis was employed for creating immersive "bloc textures" of sound. This included granular spatial synthesis via CataRT to create big sound masses [Schwarz, 2007], while specifiying position data depending on descriptor information of each grain. These masses, containing concrete sounds of rather different natures, were a parallel of "faits divers"; similar to abundant press of differing natures as an over-accumulation of information. The first experiments with synthetic soundscapes and "synthetic cloud textures" were inspired by the idea of the metaphor of "sound clouds" from the poetic text of Laure Gauthier [6].

---

[6] "Back into Nothingness" website: https://www.laure-gauthier.com/back-into-nothingness/

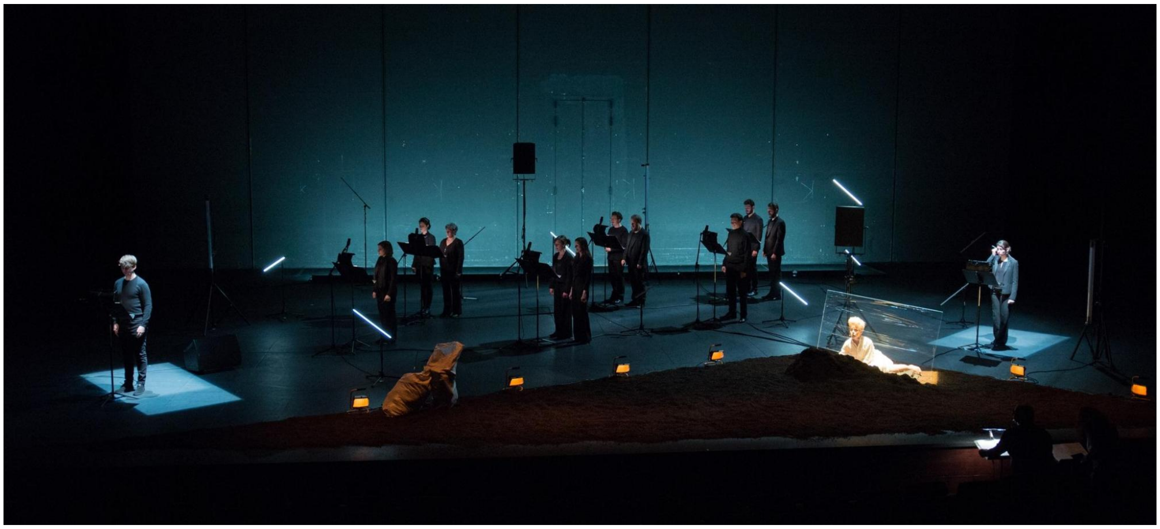


FIGURE 10: *Back into Nothingness*, scene by Giuseppe Frigeni at Theatre National Populaire of Villeurbanne.

6 Applications for Spatial Spectral Resynthesis

Initial stages of this research were dedicated to informal perceptual validations of our approaches, including applying spatial transformations. For example, this included examination of perceptions from single source to multiple sources depending on different parameters, such as divergence in time or spatial trajectory of partials of a resynthesized spectrum. Given the novel artistic possibilities for unorthodox loudspeaker configurations, and in continuation of the work of the "synthetic soundscape", the second author realized a number of installation studies during a residency at ZKM. These were further developed recently at studios at IRCAM, as described in the following sections.

6.1 Developing Ideas for an Installation

As part of this project, an installation study was created focusing on two primary ideas:

1. In promenade format - nearby sources
2. Synthetic soundscape in ambisonics format

These miniature compositions were dedicated to specific questions which are reflected in music installation studies. These include ideas such as: spectral spatialization, extended sources, density in space, different layers and textures, thereby always connecting space parameters to the beginning of a sound synthesis process. For the idea of spectral spatialization, the interfaces and tools described in Section 2 were employed to realize the idea of an "extended source". These interfaces allow a user e.g. to break up spectral contents (partials and residual, or deterministic and stochastic parts) in space by means of algorithmic trajectory generation tools: extension of a single trajectory, or multiple trajectories with the same morphology. Additionally, it was found very effictive to group sound components of complex or dense spectra and textures and distribute them in space in perceptually pertinent ways.

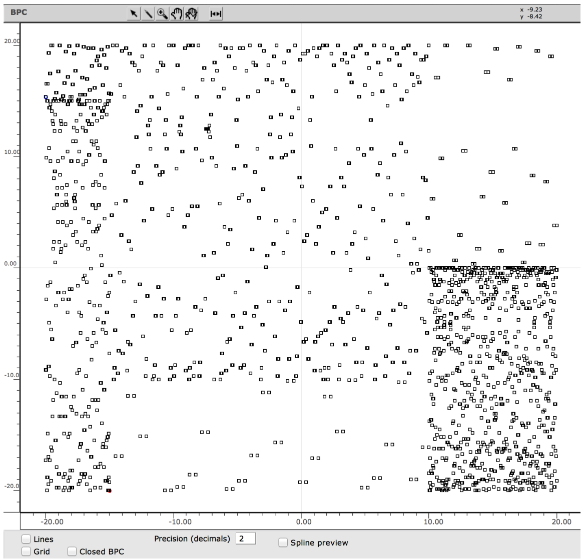


FIGURE 11: Sound areas of various densities (granulation layers at « between the leaves »).

A number of approaches were elaborated focusing on the concept of density, specifically for the installation studies: density in spectrum connected to spatial and granular textures with multiple densities in space. The interest being the control of synthetic textures (granular and additive) from a higher-level; the work on granular textures focused on densities of distributions in time and space, dividing space in different areas as shown in Figure 11. This technique can be related to the notion of *distribution styles* of spatial textures discussed by D. Smalley [Smalley, 2001]:

> *The distribution style of non-contiguous space can vary between the presentation of isolated spatial points, and scatterings over smaller or larger areas.*

In additive synthesis, the second author has worked in spatial and spectral density based on logistic map data (using the *Chaos* Library for *OpenMusic*)[7] to create clusters around harmonics. This allowed working at the thresholds between pitched sound and white noise, as a function of spectral density.

The shared control-interface and generality of the framework facilitates the combination of spatialization systems and speakers setups: ambisonics, ViMiC, and combining layers of nearby sources and far resonances. This installation format, with loudspeakers in the middle of the room, gives flexibility for "active listening": different auditory perspectives depending on visitor position and movement. In this configuration, the physical loudspeaker setup is reminiscent of "auditory windows" into virtual scenes, providing different perspectives and thus an individual narrative of the scene. The studies realized at ZKM are:

« between the leaves » systems combination (ViMiC and ambisonics)
« a l'interieur des cloches » systems combination (ViMiC and ambisonics)
« synthetic soundscape » just ambisonics

---

[7] https://github.com/openmusic-project/Chaos

In *a l'interieur des cloches*, bell sounds are resynthesized in space, trying to achieve spectral distribution in space and separating the noisy transient attacks (as *objects*) from sinusoidal resonances (as *background*) in space. Through this resynthesis, adding transformation of spectral content gives blurred images of polyphonic bells in space. This spatial sound synthesis was performed to the images of Dan Browne's "Nude descending (after Duchamp)".[8] Two of the installation studies have been revised and adapted at a second residency at IRCAM before a Forum presentation.

FIGURE 12: Loudspeaker configuration during research residency at ZKM's *Kubus*.

The study titled "between the leaves" in promenade format (with loudspeakers set up in the middle of the hall) deals with polyphonic textures in space (granular synthesis). This granulation has been sculpted by different density areas and filtering different layers to obtain mixed textures in space (opposite spectral areas and opposite densities in spatial regions). There is a dialog of 2 spaces systems-areas: 8 speakers as sources in the virtual surrounding space. This installation includes image fragments of Dan Browne's installation *An Island is land*.

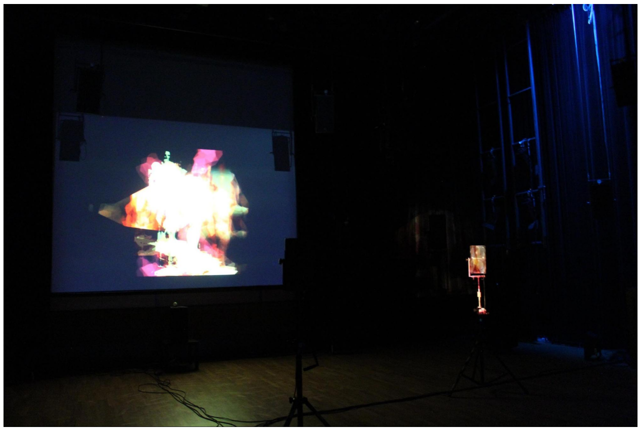


FIGURE 13: Installation at ZKM (Kubus), Dan Browne images.

---

[8] Website for D. Browne, Nude descending (after Duchamp)

Synthetic soundscape is a study specifically dedicated to densities in space and in spectrum. The second author worked on using spatial layers, grouping components in synthesis extended sources between almost white noise and pitched sound ("nuages sonores"), and dealing with connections from deferred time to real-time performance (or real-time mixing stage, listening space). This piece developed the idea of a synthetic soundscape inspired from natural sounds, but entirely done by synthesis. It also develops some metaphoric ideas, such as the idea of "synthetic clouds". Binaural sound examples can be found here: https://soundcloud.com/nuriagimenez/sets/synthetic-soundscape-binaural-excerpts.

7 Technical Challenges During the Creative Process

It has been essential to work in different studios and spaces and to exchange ideas with the EAC Team of IRCAM (Espaces Acoustiques et Cognitifs) and the team of engineers and researchers at ZKM. The work and tests in smaller-sized spaces with close speaker proximity, like ZKM's mini dome, as well as studio 1 and studio 5 of IRCAM, has been complementary to the work in a big space, such as the Kubus of ZKM. Changing studios and spaces has provided opportunities and required flexibility for the tools and formats used, e.g. in the work with HOA (encoded files) at times requiring local adaptations and fine-tunings, such as ambisonics order reduction, conversions from 3D to 2D, etc.

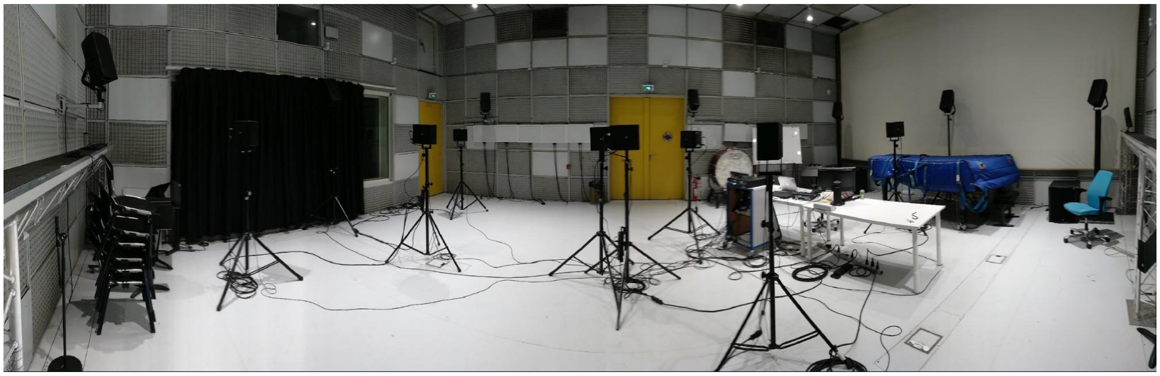


FIGURE 14: Loudspeaker configuration during research residency at IRCAM's *Studio 5*.

Tools for Deferred and Real-Time Applications

Complementary interfaces for deferred and real-time applications have been developed as part of this project. For the piece *Synthetic Soundscape* and the work with layers, our exchanges with Jean Bresson and Thibaut Carpentier on integration possibilites of the library *OM-Spat* were very important; while working in deferred time, the library provides possibilities for visualisation of the spatial parameter morphologies in the temporal dimension and in the connection between deferred time "sculpture" and real-time layers.

8 Outlook

The presented project has further explored the notion of *Spatial Sound Synthesis* from engineering and compositional perspectives. We presented a modular extension to the spatialization framework *OMPrisma* for combining sound synthesizers, sound processors and spatialization engines as part of the creative process. Inspired by existing concepts of *sound sculpture* (and accordingly, *shape* and *texture*), this research has motivated a number of artworks and vice versa, compositional motivations have also inspired some directions of the research.

From a perceptual research viewpoint, the developed tools might help to learn and better understand how separate dimensions interact and give rise to unexpected sound sensations. Developing a model for auditory attributes characteristic to sound sculpture (such as shape/volume, texture, "material", etc.) is a non-trivial task, as perceptual and signal/physical dimensions are incongruent and can interfere with one another. With these considerations in mind, our recent development efforts are hoped to open possibilities for compositional practice not only for integrating spatialization as an integral musical parameter, but further for an emancipation to the design of spatialization instruments and higher-level spatialization approaches in the sense of an "orchestration of the space".

Following this initial work, there are many ways and approaches on how the new framework and tools can be employed, suggesting several avenues for further development. From an artistic perspective, a number of potential improvements for installations are envisioned: For instance, the integration of visual aspects and the possibility of real-time interaction between space, synthesis and image are envisaged for new versions or future installations. On the spatial synthesis side, possibilities and methods for adapting the studies to WFS or other systems. Working on a new "Musique-Fiction" work for IRCAM, the second author is currently using some of these tools to resynthesize "natural" immersive soundscapes in 3D. She is also considering new research, such as extensions with real-time approaches and exchanges on possibilities offered by the *Antescollider* software [Fernandez et al., 2021].

The libraries *OMPrisma* and the *Binauralize* software are open source and freely available along with documentation and examples, including the rendered multichannel soundfiles and example patches shown in Figures 7 and 9.[9]

Acknowledgments

We would like to thank all of the people involved in this project, most notably: Jean Bresson, Marco Stroppa, Thibaut Carpentier, Markus Noisternig and the whole Acoustic and Cognitive Spaces team at Ircam, Greg Beller, Paola Palumbo and all the team of the IRCAM Forum, Götz Dipper, Benjamin Miller, Ludger Brümmer and the team of ZKM.

---

[9] **https://sourceforge.net/projects/omprisma/**
https://github.com/marleynoe/Prisma-Binauralize

Special thanks from the first author go out to Núria Giménez-Comas and the colleagues at IRCAM for support and collaboration during a period of illness. Finally, thanks to J. Dickhaus for editing an initial manuscript of this paper.

References

[Agger et al., 2017] Agger, S., Bresson, J., & Carpentier, T. (2017). Landschaften – Visualization, Control and Processing of Sounds in 3D Spaces. In *ICMC 2017* https://hal.archives-ouvertes.fr/hal-01567629/document.

[Agon et al., 2011] Agon, C., Bresson, J., & Stroppa, M. (2011). OMChroma: Compositional Control of Sound Synthesis. *Computer Music Journal*, *35*(2), 67–83.

[Braasch et al., 2008] Braasch, J., Peters, N., & Valente, D. L. (2008). A Loudspeaker-Based Projection Technique for Spatial Music Applications Using Virtual Microphone Control. *Computer Music Journal*, *32*(3), 55–71, https://doi.org/10.1162/comj.2008.32.3.55 http://dx.doi.org/10.1162/comj.2008.32.3.55.

[Bregman, 1994] Bregman, A. S. (1994). *Auditory Scene Analysis - The Perceptual Organization of Sound*. MIT Press.

[Bregman & Steiger, 1980] Bregman, A. S. & Steiger, H. (1980). Auditory streaming and vertical localization: Interdependence of "what" and "where" decisions in audition. *Attention, Perception, & Psychophysics*, *28*(6), 539–546 http://www.springerlink.com/index/H202352611324Q32.pdf.

[Bresson et al.,] Bresson, J., Bouche, D., Carpentier, T., Schwarz, D., & Garcia, J. Next-generation Computer-aided Composition Environment: A New Implementation of Open-Music . In *ICMC 2017* https://hal.archives-ouvertes.fr/hal-01567619/document.

[Bresson & Schumacher, 2011] Bresson, J. & Schumacher, M. (2011). Representation and interchange of sound spatialization data for compositional applications. In *International Computer Music Conference* Huddersfield, United Kingdom.

[Cabrera, 2013] Cabrera, A. (2013). Pseudo-stereo Techniques. (pp. 1–6).

[Carlile & Leung, 2016] Carlile, S. & Leung, J. (2016). The Perception of Auditory Motion. *Trends in Hearing*, *20*(1), 1–19, https://doi.org/10.1177/2331216516644254 http://journals.sagepub.com/doi/10.1177/2331216516644254.

[Carpentier, 2015] Carpentier, T. (2015). RÉCENTS DÉVELOPPEMENTS DU SPATIALISATEUR. In *Journées d'Informatique Musicale* (pp. 1–8). Montréal, Canada.

[Coleman, 1962] Coleman, P. D. (1962). Failure to Localize the Source Distance of an Unfamiliar Sound. *Journal of the Acoustical Society of America*, *34*(3), 345–346, https://doi.org/10.1121/1.1928121 http://asa.scitation.org/doi/10.1121/1.1928121.

[Fernandez et al., 2021] Fernandez, J. M., Giavitto, J.-L., & Donat-Bouillud, P. (2021). AntesCollider: Control and Signal Processing in the Same Score. (pp. 1–7).

[Genzel et al., 2018] Genzel, D., Schutte, M., Brimijoin, W. O., MacNeilage, P. R., & Wiegrebe, L. (2018). Psychophysical evidence for auditory motion parallax. *Proceedings of the National Academy of Sciences*, *115*(16), 4264–4269, https://doi.org/10.1073/pnas.1712058115 http://www.pnas.org/lookup/doi/10.1073/pnas.1712058115.

[James, 2016a] James, S. (2016a). *A CLASSIFICATION OF MULTI-POINT SPECTRAL SOUND SHAPES*. Technical report https://www.stuartgjames.com/uploads/1/7/4/5/17453311/2016jamesacmc.pdf.

[James, 2016b] James, S. (2016b). Multi-Point Nonlinear Spatial Distribution of Effects across the Soundfield. (pp. 1–3). https://www.stuartgjames.com/uploads/1/7/4/5/17453311/2016jamesicmc.pdf.

[Kendall, 1995] Kendall, G. (1995). The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. *Computer Music Journal*, *19*(4), 71–87 http://www.jstor.org/stable/10.2307/3680992.

[Kendall, 2010] Kendall, G. S. (2010). Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music. *Organised Sound*, *15*(03), 228–238, https://doi.org/10.1017/S1355771810000336.

[Keylin, 2015] Keylin, V. (2015). Corporeality of music and sound sculpture. *Organised Sound*, *20*(2), 182–190 http://en.dada-avis.net/wp-content/uploads/2020/02/corporeality_of_music_and_sound_sculpture.pdf.

[Khosravi, 2014] Khosravi, P. (2014). Circumspectral Sound Diffusion with Csound. (pp. 1–11).

[Kim-Boyle, 2008] Kim-Boyle, D. (2008). Spectral Spatialization - an Overview. In *International Computer Music Conference* Belfast, Ireland http://classes.berklee.edu/mbierylo/ICMC08/defevent/papers/cr1549.pdf.

[Lachenmann, 1996] Lachenmann, H. (1996). *Musik als existentielle Erfahrung*.

[McLeran et al., 2008] McLeran, A., Roads, C., Sturm, B., & Shynk, J. (2008). Granular Sound Spatialization Using Dictionary-Based Methods. In *Sound and Music Computing Conference* Berlin, Germany.

[Normandeau, 2009] Normandeau, R. (2009). Timbre Spatialisation: The medium is the space. *Organised Sound*, *14*(03), 277–285, https://doi.org/10.1017/S1355771809990094.

[Peters et al., 2013] Peters, N., Lossius, T., & Schacher, J. C. (2013). The Spatial Sound Description Interchange Format: Principles, Specification, and Examples. *Computer Music Journal*, *37*(1), 11–22, https://doi.org/10.1162/COMJ http://www.mitpressjournals.org/doi/abs/10.1162/COMJ_a_00167.

[Pihlajamäki et al., 2014] Pihlajamäki, T., Santala, O., & Pulkki, V. (2014). Synthesis of Spatially Extended Virtual Sources with Time-Frequency Decomposition of Mono Signals. *Journal of the Audio Engineering Society*, *62*(7), 1–18.

[Potard & Burnett, 2004] Potard, G. & Burnett, I. (2004). Control and Measurement of Apparent Sound Source Width and Its Applications to Sonification and Virtual Auditory Displays. (pp. 1–7).

[Ramakrishnan et al., 2006] Ramakrishnan, C., Goßmann, J., & Brümmer, L. (2006). The ZKM Klangdom. In *New Interfaces for Musical Expression* (pp. 140–143). Paris, France http://portal.acm.org/citation.cfm?id=1142215.1142250.

[Rumsey, 2002] Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society*, *50*(9), 651–666 http://www.ece.uvic.ca/~peterd/30605/rumseyattributes.pdf.

[Saint-Arnaud & Popat, 1995] Saint-Arnaud, N. & Popat, K. (1995). Analysis and Synthesis of Sound Textures. *Readings in Computational Auditory Scene Analysis*, (pp. 125–131).

[Schmele & Sayin, 2020] Schmele, T. & Sayin, U. (2020). Controlling The Apparent Source Size In Ambisonics Using Decorrelation Filters. (pp. 1–7).

[Schumacher, 2016] Schumacher, M. (2016). *A Framework for Computer-Aided Composition of Space, Gesture, and Sound. Conception, Design, and Applications.* PhD thesis, McGill University http://digitool.library.mcgill.ca/R/-?func=dbin-jump-full&object_id=143671&silo_library=GEN01.

[Schumacher & Bresson, 2010] Schumacher, M. & Bresson, J. (2010). Spatial Sound Synthesis in Computer-Aided Composition. *Organised Sound*, *15*(03), 271–289, https://doi.org/10.1017/S1355771810000300 http://www.idmil.org/software/omprisma.

[Schwarz, 2007] Schwarz, D. (2007). Corpus-Based Concatenative Synthesis. *Signal Processing Magazine* http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4117932.

[Schwarz, 2011] Schwarz, D. (2011). State of the Art in Sound Texture Synthesis. In *International Conference on Digital Audio Effects* Paris, France http://recherche.ircam.fr/pub/dafx11/Papers/30_e.pdf.

[Senna et al., 2017] Senna, I., Parise, C. V., & Ernst, M. O. (2017). Modulation frequency as a cue for auditory speed perception. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1858), 20170673–7, https://doi.org/10.1098/rspb.2017.0673 https://royalsocietypublishing.org/doi/10.1098/rspb.2017.0673.

[Serra & Smith, 1990] Serra, X. & Smith, J. (1990). Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, *14*(4), 12–24 http://www.jstor.org/stable/3680788.

[Smalley, 2001] Smalley, D. (2001). Spectromorphology: explaining sound-shapes. *Organised Sound* http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=76428.

[Stefani & Lauke, 2010] Stefani, E. & Lauke, K. (2010). Music, Space and Theatre: Site-specific approaches to multichannel spatialisation. *Organised Sound*, *15*(03), 251–259, https://doi.org/10.1017/S1355771810000270.

[Stuchlik, 2017] Stuchlik, J. (2017). *Virtuelle Raumakustik als modularer Ansatz, basierend auf physikalischen, perzeptuellen und signalbasierten Verfahren*. PhD thesis, Karlsruhe.

[Torchia & Lippe, 2004] Torchia, R. & Lippe, C. (2004). Techniques for Multi-Channel Real-Time Spatial Distribution Using Frequency-Domain Processing. In *Conference on New Interfaces for Musical Expression* (pp. 116–119). Hamamatsu, Japan http://portal.acm.org/citation.cfm?id=1085884.1085910.

[Verron, 2009] Verron, C. (2009). Spatialized Synthesis of Noisy Environmental Sounds. (pp. 1–17).

[Verron et al., 2010] Verron, C., Aramaki, M., Kronland-Martinet, R., & Pallone, G. (2010). A 3-D immersive synthesizer for environmental sounds. *Audio, Speech, and Language Processing, IEEE Transactions on*, *18*(6), 1550–1561.

[Wittek et al.,] Wittek, H., Kerber, S., Rumsey, F., & Theile, G. Spatial perception in wave field synthesis rendered sound fields: Distance of real and virtual nearby sources. *aes.org* http://www.aes.org/e-lib/browse.cfm?elib=12711.

[Yost, 2018] Yost, W. A. (2018). Auditory motion parallax. *Proceedings of the National Academy of Sciences*, *115*(16), 3998–4000, https://doi.org/10.1073/pnas.1803547115 http://www.pnas.org/lookup/doi/10.1073/pnas.1803547115.

[Zotter et al., 2011] Zotter, F., Frank, M., Marentakis, G., & Sontacchi, A. (2011). Phantom Source Widening with Deterministic Frequency Dependent Time Delays. In *International Conference on Digital Audio Effects*.