Instrumental Radiation Patterns as Models for Corpus-Based Spatial Sound Synthesis: *Cosmologies* for Piano and 3D Electronics

Aaron Einbond[1]
[1]City, University of London
aaron.einbond@city.ac.uk

Jean Bresson[2,3]
[2]Ableton, Berlin
bresson@ircam.fr

Diemo Schwarz, Thibaut Carpentier[3]
[3]STMS lab IRCAM-CNRS-Sorbonne Université, Paris
{schwarz,carpentier}@ircam.fr

ABSTRACT

The Cosmologies project aims to situate the listener inside a virtual grand piano by enabling computer processes to learn from the spatial presence of the live instrument and performer. We propose novel techniques that leverage measurements of natural acoustic phenomena to inform spatial sound composition and synthesis. Measured radiation patterns of acoustic instruments are applied interactively in response to a live input to synthesize spatial forms in real time. We implement this with software tools for the first time connecting audio descriptor analysis and corpus-based synthesis to spatialization using Higher-Order Ambisonics and machine learning. The resulting musical work, Cosmologies *for piano and 3D electronics, explodes the space inside the grand piano out to the space of the concert hall, allowing the listener to experience its secret inner life.*

1. INTRODUCTION

How does a listener immediately know the difference between a live grand piano and a recording? One reason is the complex interaction between the piano's sound and the space that surrounds it. Research in the field of music perception points to the essential role of situated or embodied cognition in our listening experience: "cognition as an activity that is structured by the body situated in its environment" [1]. The Cosmologies project seeks to place the embodied presence of the instrument and its performer at the center of research and creation, using machine learning of audio features to decipher the intricate inter-dependencies of timbre and space that bring an instrument to life.

This work was framed in the context of an artistic research residency project focused on the development of computer-assisted composition tools connecting digital signal processing, spatial audio and machine learning. [1] Composer Aaron Einbond proposed an artistic work, *Cosmologies* for piano and 3D electronics, focused on the goal to situate the listener inside a virtual grand piano. How could this be executed, technically and artistically? In order to highlight the spatial presence of the acoustic instrument as it blends with interactive electronics, one of the main challenges in this project is to drive sound spatialization not by artificial symbolic processes, but to root it in natural acoustic phenomena.

The measurement and reproduction of acoustic properties in a musical work for piano and electronics requires careful choices for both composer and computer, leading us to explore novel computer-assisted composition and machine learning techniques to process and map large databases of acoustic data to synthesized audio material. Indeed, while applications of Artificial Intelligence (AI) research to audio are ubiquitous today, up to now they have rarely treated the embodied spatial presence of live instruments and performers. Most AI applications assume an audio output of stereo loudspeakers or headphones, flattening the listening experience. In contrast, one goal of our project is to use machine learning to "re-embody" computer synthesis in a 3D spatial environment.

[1] Work supported by the VERTIGO project as part of the STARTS program of the European Commission, based on technological elements from the Project OM7/om-spat: https://vertigo.starts.eu/calls/starts-residencies-call-3/projects/om-spat/detail

2. SPATIAL PRESENCE

Past research has elucidated the complex spatial radiation patterns of acoustic instruments [2]. Attempts have been made to reproduce instrumental models electronically using techniques of Higher-Order Ambisonics (HOA), but these were primarily in a research context rather than artistic creation [3]. In his previous work *Cartographies* for piano with two performers and electronics, Einbond was inspired by these studies to produce a live interactive response to the acoustic piano based on an analysis of its audio features to spatialize electronic sound in 43.4 channels in the Kubus of the Zentrum für Kunst und Medien (ZKM), Karlsruhe [4]. In *Cartographies* the piano is conceived not as a sonic abstraction, but instead as a physical object, prepared with foreign materials and a catalogue of lavalier and contact microphones, with which the two performers interact with their full bodies.

In *Cosmologies*, this approach was developed further and augmented by data with rich acoustic spatial features that we obtained from two different origins: (1) a database of measured radiation patterns of acoustic instruments [5], and (2) recordings and amplification of prepared piano with a spherical microphone array, the mh acoustics 32-channel Eigenmike (EM32). [2] These two distinct spatial audio models are superposed in the live performance of *Cosmologies* to situate the listener in an immersive sonic space where it is intentionally difficult to differentiate the real and virtual sources.

[2] https://mhacoustics.com/products

2.1 Instrumental Radiation Data

We exploit a database of instrumental radiation patterns measured, analyzed, and freely distributed by the Technische Universität (TU) Berlin [5, 6]. This database includes radiation patterns for 41 orchestral and historical acoustic instruments recorded in the anechoic chamber of TU Berlin using a surrounding spherical array of 32 microphones (see Figure 1). The measured directivity patterns are represented in the spherical harmonic domain using 25 coefficients corresponding to 4^{th}-order harmonics. These complex-valued coefficients were first converted using Matlab to real-valued spherical harmonics following the Ambisonics ACN N3D convention [7] for compatibility with the software tools later used in this project. The database provides directional data for the first 10 partials of each note of each instrument at two dynamic levels, as well as data averaged over all partial frequencies within each of 31 third-octave bands. As our project is focused on complex, timbrally rich sounds produced by prepared piano, the radiation data for third-octave bands was used rather than single frequencies. Ambisonics coefficients were normalized to avoid drastic level changes and stored in one text file for each instrument for easy reference by the live performance patch. These ambisonic coefficients, or spatial weights, could then be applied to novel mono recordings, endowing them with the spatial radiation pattern of a chosen instrumental model.

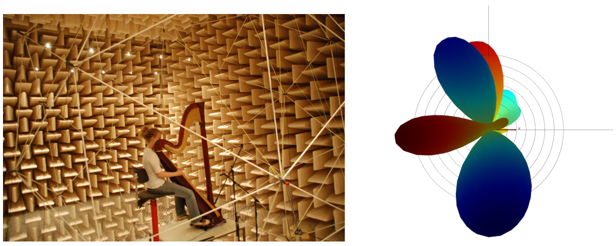

Figure 1: Radiation data recorded in the anechoic chamber of TU Berlin (from [5]) visualized with spat5.hoa.plot.

2.2 Ambisonic Recording and Live Amplification

Although it features a wide variety of acoustic instruments, the TU Berlin database does not include data for the piano. [3] Therefore, to complement the database, we made recordings of a Steinway D grand piano in IRCAM Studio 5. Here, the idea was not to measure the radiation characteristics of the instrument, but rather to record samples incorporating spatial attributes. During the recording session, Einbond created spatial gestures of objects and preparations including aluminum foil, knitting needles, scrub brush, wrapping paper, and superball. Recordings were made with the EM32 suspended upside-down over the instrument (see Figure 2). The session also included a 1^{st}-order-Ambisonics Sennheiser AMBEO VR microphone, which was not used in the final composition in favor of the richer spatial resolution of the EM32, as well as two DPA 4006 omni microphones, which were used to create a mono sample corpus. Some of the piano samples as well as field recordings produced with the EM32 were encoded into 4^{th}-order HOA files [8] and triggered during the live performance for decoding and playback over an ambisonic loudspeaker system. These were complemented by the same EM32 setup, encoded and decoded with 4^{th}-order HOA in real time, for live 3D amplification of the piano in the concert space.

Figure 2: Recording prepared piano with mh acoustics EM32, Sennheiser AMBEO VR, and DPA 4006 omni microphones.

3. SOFTWARE SYNERGIES

The computer music tools and methodologies developed for the project connect audio descriptor analysis and corpus-based synthesis with spatialization using HOA and machine learning techniques, as summarized in Figure 3. The project harnesses computer programs Max and OM#, as well as associated packages Spat5 [9] [4] and MuBu [10]. [5]

3.1 OM# and computer-assisted composition

The OM# computer-assisted composition environment [11] was used for pre-compositional prototyping including creating audio mosaics, testing spatialization approaches, and evaluating descriptor choices for machine learning.

3.1.1 Audio mosaicking

The score of *Cosmologies* was prepared using computer-assisted composition tools in OM# and Max to generate audio mosaics of prepared piano recordings and field recordings. As reported in previous research [12], these *target* recordings were segmented and mapped to sample segments in the pre-recorded prepared piano *corpus* with the most similar descriptor values. This process was prototyped using IAE [13], the *Interactive Audio Engine* of MuBu integrated in OM# [6] (see Figure 4). The IAE object calculates corpus and target segmentation and descriptor analyses and the resulting mosaic can be visualized and rendered with the IAE-container object. Based on these prototypes, further audio mosaics were produced in Max using the MuBu and bach packages [14], exported in MusicXML format, and subjectively edited in computer program Finale to produce the instrumental score to be interpreted by the live performer.

[3] This was due to the large physical weight and size of the instrument, which would have required too complicated an installation and too large a microphone sphere. (Stefan Weinzierl, personal communication.)

[4] https://forum.ircam.fr/projects/detail/spat/

[5] https://forum.ircam.fr/projects/detail/mubu

[6] https://github.com/cac-t-u-s/om-iae

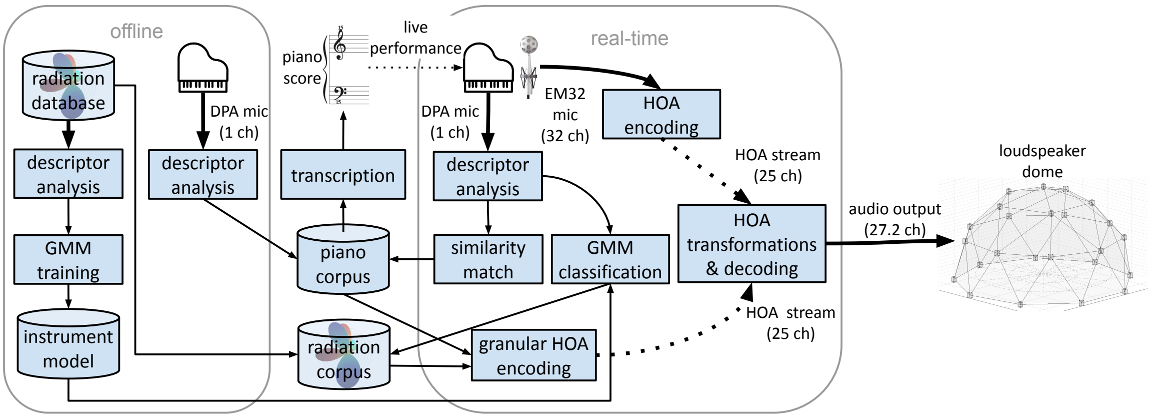

Figure 3: Offline and real-time processing steps in *Cosmologies*. Thin, thick, and dotted arrows represent data, audio, and HOA streams, respectively.

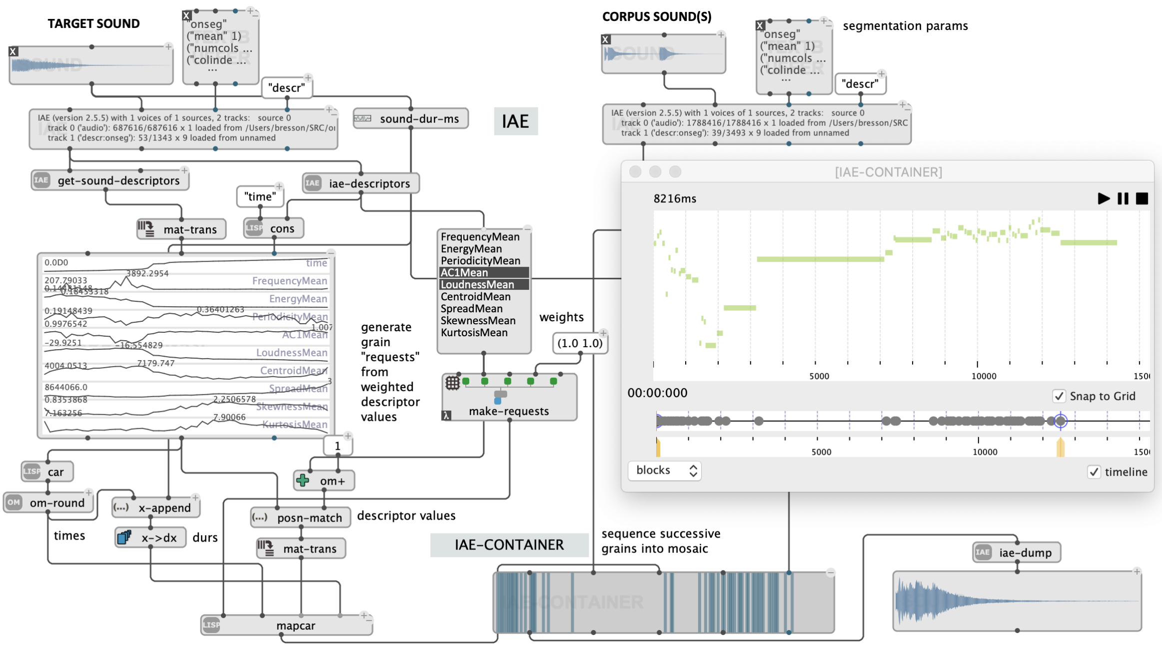

Figure 4: Audio mosaic with OM# and IAE: analyze multiple descriptors from a target (left), use them to select grains from a corpus (right), and synthesize the result.

3.1.2 Offline spatial rendering

The spatial electronics were prototyped using Spat modules integrated in OM# [15].[7] This allowed previews of spatial scenes to be generated and evaluated in binaural simulations, informing the design of the live electronic patch. We compared filtering a corpus sample into 31 separate frequency bands, each spatialized individually, versus spatializing the entire sample according to its loudest frequency band (see Figure 5). We eventually chose the latter approach, as the former had the effect of averaging out spatial differences to produce a more omni-directional result.

3.1.3 Machine learning with OMAI

The OMAI library [16] for OM# was used as well with IAE to perform preliminary experiments in the classification of samples by instrument based on their audio features. Using the IAE object, different descriptors from the IrcamDescriptors audio feature set [17] were compared as inputs to the *k-means* clustering algorithm, then samples from the training data set were tested to evaluate the accuracy of the classification. The best results were obtained with *Mel-Frequency Cepstrum Coefficients* (MFCC) and *Relative Specific Loudness* (RSL, a multi-band loudness curve comprising the loudness of a chosen number of frequency bands, each normalized by the total loudness).

3.2 MuBu for Max

Based on the OM# prototypes, similar processes were performed in real time in the *Cosmologies* concert patch programmed in Max 8 with the MuBu and Spat5 packages.

3.2.1 Audio descriptor analysis and corpus-based synthesis

As in the deferred-time case, audio mosaics are created in real time using using corpus-based synthesis techniques [18].

[7] https://github.com/cac-t-u-s/spat

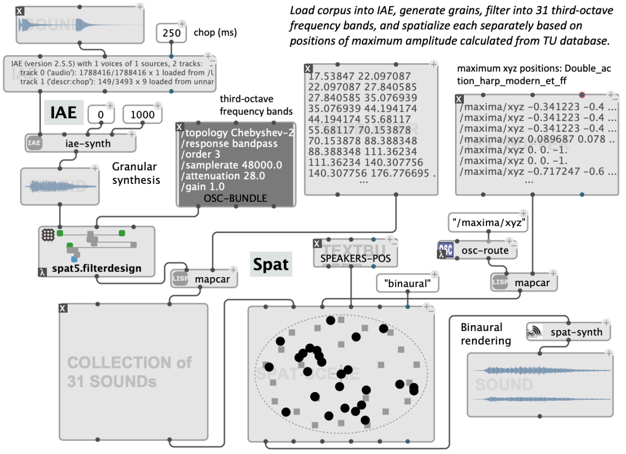

Figure 5: OM#: synthesis, band-filtering and spatialization of granular sounds using IAE and Spat.

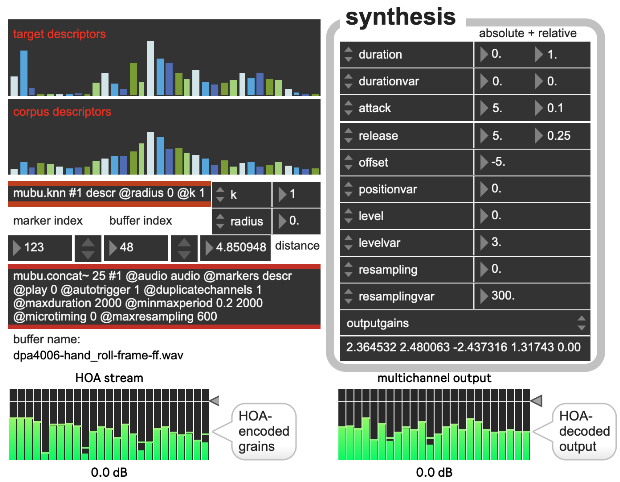

Figure 6: Descriptor analysis, selection, and synthesis of 4th-order HOA-encoded sample segment with mubu.concat∼.

MuBu's *onseg* onset detection algorithm segments an incoming target audio signal from the live input, in this case two DPA 4099 cardioid microphones positioned inside the piano. The complementary Max external objects pipo.ircamdescriptors and mubu.process allow both real- and deferred-time audio to be analyzed according to identical descriptors and parameter settings chosen among those available from IrcamDescriptors. [8] Following tests with OM#, the descriptors chosen for the performance patch were *loudness*, *spectral centroid*, and *RSL*. While RSL is associated with 24 Bark bands by default [17], we chose a custom set of frequency bands to match the 31 third-octave bands defined in the TU Berlin database [6]. Then the mubu.knn external maps each target segment to the segment in the pre-recorded corpus with the most similar descriptor values. During live performance, sounds from the large corpus of pre-recorded prepared piano samples are triggered and resynthesized in response. The resulting stream of sample segments is further processed through a list of granular synthesis parameters before spatial rendering (Figure 6).

3.2.2 *HOA encoding*

Live spatial encoding is performed through concatenative synthesis with the mubu.concat∼ module, which offers elegant possibilities for individual HOA encoding of each sample segment as it is output from the pre-recorded corpus. The chosen segment is duplicated into the required number of channels, in this case 25 for 4th-order HOA encoding. Then the *outputgains* parameter is used to scale each channel by a linear coefficient, in this case derived from the spherical harmonic coefficients of the TU Berlin database. Each resulting 25-channel HOA-encoded segment can then be added to the outgoing HOA stream for later HOA decoding. Thanks to the overlap-add algorithm, an arbitrary number of overlapping segments can be summed, each with its own spatial radiation pattern, situating the listener in the midst of dynamic spatial polyphony (see Figure 6).

[8] See the *CaMu* (CataRT-MuBu) tutorial library: https://forum.ircam.fr/projects/detail/catart-mubu

3.2.3 *Machine learning with GMM*

The spatial radiation pattern for each segment to be output is selected through a machine learning algorithm that matches the live audio input descriptors to one of the instruments from the TU Berlin database and then selects the best-fitting register and dynamic for that instrument. A subset of 21 modern instruments was chosen from the 41 instruments of the database, focusing on instruments offering data in a wider range of frequency bands in order to allow for more varied possibilities for spatial mapping. The resulting database of 1788 monophonic samples was used as a training set for a classification task using a Gaussian mixture model (GMM) with the external object mubu.gmm. The database was further chopped into 100 ms segments and the GMM model was trained using the same 31-band RSL descriptors described above, with parameter values chosen with the goal of a tightly-fitting classification (see Figure 7). The resulting GMM model could then be conveniently saved in JSON format and used in performance without the need for the original sound files of the training data set. The model's *likelihood window*—how many consecutive results it awaits before changing classification—was varied dynamically during performance for dramatic effect, ranging from 5 for a more stable classification to 1 for the most rapid and volatile changes. Once the GMM algorithm classifies the live audio input by instrument, a specific radiation pattern of that instrument is further selected according to register, based on RSL values, and dynamic, based on loudness.

A control test of samples from the training set against the GMM classification model revealed high accuracy. However, as the final classification task involves matching prepared piano samples to a list of 21 instruments not including piano, there is no objective "ground truth" against which to judge the accuracy of the matches. Yet subjectively, the resulting classification showed a strong correlation of registral and timbral characteristics: for example, a superball drawn along low piano strings was mapped to tuba, contrabass, and trombone while knitting needles on the high strings (see Figure 8) were mapped to violin, flute, and oboe.

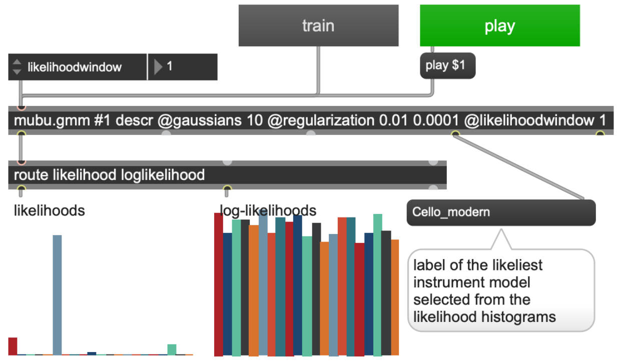

Figure 7: Instrument classification by Gaussian mixture model using mubu.gmm for the segment shown in Figure. 6.

Figure 8: Pianist Alvise Sinivia's performance captured by the EM32 microphone. Photo: Quentin Chevrier.

4. MUSICAL RESULTS

Preparations for the live premiere performance of *Cosmologies*, as well as two related musical realizations, presented opportunities to evaluate and refine the musical results.

4.1 Ambisonic Order and CPU

During rehearsals we encountered occasional audio glitches attributed to high CPU usage. The prepared piano sample corpus used for corpus-based synthesis mostly contained short segments (ca. 100 ms); however, a few longer segments (up to 2 seconds) were also included. When rendered with 4^{th}-order Ambisonics, these longer segments were responsible for spikes in CPU usage, due to the implementation details of mubu.concat~ wherein a grain is rendered completely to 25 (4^{th}-order) output channels when triggered. An effective solution was to create a second instance of mubu.concat~ set to output in four channels for 1^{st}-order Ambisonics encoding, used only during the specific sections of the piece where longer segments were needed. This resolved the audio glitches and the results remained perceptually convincing, as these long segments were mostly from the piano's low register where they tended to be mapped to more omni-directional radiation patterns.

4.2 Ambisonic Transformations and Decoding

Spatial transformations were applied to each HOA stream in order to tune its spatial rendering. With the EM32 microphone oriented upside-down (see Figures 2 and 8) we chose to map the "top" of the microphone, pointed down toward the piano strings, up to the top of the loudspeaker dome, to ensure the most sonorous part of the instrument was projected to the center of the dome. This required the sound-field to be transformed so that the left-right orientation of the piano was properly preserved, achieved by rotation and mirroring applied in the HOA domain. A further spatial *warping* effect [19] was applied specifically to the synthesized HOA stream in order to focus the 3D image toward the live piano and performer on stage, and away from the back of the concert hall, using spat5.hoa.warp~. Additional light equalization, compression, and reverb effects were applied in order to blend electronic and instrumental sound and avoid feedback from the 3D microphone. These operations were achieved using tools from the Spat5 package [9] before decoding the ambisonic stream with spat5.hoa.decoder~. During the premiere performance, two separate computers routed over a Dante audio network were used, respectively, to synthesize and decode the HOA streams to a 27.2-loudspeaker ambisonic dome installed in the Grande Salle of Centre Georges Pompidou, Paris. [9]

4.3 Further musical realizations: Cosmologies II–III

The tools and materials of the Cosmologies project were applied to two further modular compositions that may be performed together or separately from *Cosmologies* for piano and 3D electronics. *Cosmologies II* is an interactive sound installation that was presented before the premiere performance of *Cosmologies* as the audience entered the concert hall. The gains of the DPA 4099 microphones positioned in the piano are turned up to capture the ambient sound of the audience and trigger short grains from the prepared piano corpus spatialized with their selected radiation models. This allows the audience to experience the 3D electronics while free to move within the space and the piano is silent, in contrast to the live performance where the pianist is in motion and the audience is stationary.

Cosmologies III for fixed 3D electronics was created during a later residency in the ZKM Kubus, in which corpus-based synthesis techniques are combined with a wider selection of prepared piano samples and field recordings made with the EM32. It may be performed alone or directly following *Cosmologies*, after the live performer has left the stage, his presence remaining only as a shadow behind the spatial audio recordings. The work was premiered during an online streaming performance in a binaural version prepared using spat5.virtualspeakers~ based on the spatial positions of the 43.4-channel loudspeaker dome of the Kubus, allowing the listener to experience the 3D electronics over headphones as intimate virtual chamber music. [10]

[9] Video and binaural mix of the premiere performance available here: https://youtu.be/jKIWLwPrun4

[10] *Cosmologies III* binaural streaming version available here: https://youtu.be/sooNxK6oQ4c?t=14298

5. CONCLUSIONS AND FUTURE DIRECTIONS

Cosmologies was the first use of the EM32 microphone in live electroacoustic performance and simultaneously the first fully realized project involving the new OM# environment and its connection to Spat and MuBu. It represents a best-case scenario for art–science collaboration, where the synergy of technological research and artistic vision helps both artist and researchers to advance their goals in a direction that neither could have accomplished alone.

The two spatialization approaches described here present dual models for reproducing spatial acoustic phenomena: the EM32 receives sounds diffused inside the piano around the microphone, and projects these sounds toward the listener situated inside the loudspeaker dome, as if the listener is inside a larger-than-life grand piano. In contrast, the radiation patterns of the TU database were recorded through a sphere of microphones surrounding the instruments, and are turned inside-out to become diffusion patterns when applied to samples synthesized and rendered in the loudspeaker dome, as if the listener were seated in the same position as the instrumental performer in a perfectly reflective room. This spherical "eversion," exchanging the perspectives of the performer and listener, befits the title *Cosmologies*.

An alternative strategy to reproduce instrumental radiation patterns could be to project sound through a compact 3D loudspeaker such as the IKO spherical loudspeaker array [20]. Rather than immersing the listener inside a virtual grand piano, this technology could give the impression of a virtual acoustic source situated in the space with the listener. The framework developed through this project could be applied readily to such a hardware system to model the spatial presence of acoustic instruments in a different way.

Further goals include the evaluation and comparison of machine learning algorithms applied directly to Ambisonics coefficients, rather than through a GMM classification model. Finally, the authors plan to develop the software tools for the project into a repository of tutorial patches for OM# and Max for public release.

Acknowledgments

Thanks to Olivier Warusfel, Markus Noisternig, Riccardo Borghesi, Anders Vinjar, Stefan Weinzierl, Tânia Moreira, Louise Enjalbert, Greg Beller, Hugues Vinet, Manuel Poletti, Luca Bagnoli, Clément Cerles, and Alvise Sinivia.

6. REFERENCES

[1] V. Iyer, "Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music," *Music Perception*, vol. 19, no. 3, pp. 387–414, 2016.

[2] F. Zotter, "Analysis and Synthesis of Sound-Radiation with Spherical Arrays," Ph.D. dissertation, IEM, University of Music and Performing Arts, Graz, 2009.

[3] M. Noisternig, F. Zotter, and B. F. G. Katz, "Reconstructing Sound Source Directivity in Virtual Acoustic Environments," in *Principles and Applications of Spatial Hearing*. World Scientific, 2011, pp. 357–373.

[4] A. Einbond, "Mapping the Klangdom Live: Cartographies for piano with two performers and electronics," *Comput. Music J.*, vol. 41, no. 1, pp. 61–75, 2017.

[5] S. Weinzierl, M. Vorländer, G. Behler, F. Brinkmann, H. von Coler, E. Detzner, J. Krämer, A. Lindau, M. Pollow, F. Schulz, and N. R. Shabtai, "A Database of Anechoic Microphone Array Measurements of Musical Instruments," 2017. [Online]. Available: http://dx.doi.org/10.14279/depositonce-5861.2

[6] N. R. Shabtai, G. Behler, M. Vorländer, and S. Weinzierl, "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *J. Acoust. Soc. Am.*, vol. 141, no. 2, pp. 1246–1256, 2017.

[7] T. Carpentier, "Normalization schemes in Ambisonics: does it matter?" in *Proc. 142nd AES Conv.*, Berlin, 2017.

[8] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of Spherical Microphone," in *Proc. 120th AES Conv.*, Paris, 2006.

[9] T. Carpentier, "A new implementation of Spat in Max," in *Proc. SMC*, Limassol, 2018.

[10] N. Schnell, A. Röbel, D. Schwarz, G. Peeters, and R. Borghesi, "MuBu & Friends–Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP," in *Proc. ICMC*, Montreal, 2009.

[11] J. Bresson, D. Bouche, T. Carpentier, D. Schwarz, and J. Garcia, "Next-generation Computer-aided Composition Environment: A New Implementation of OpenMusic," in *Proc. ICMC*, Shanghai, 2017.

[12] A. Einbond, D. Schwarz, and J. Bresson, "Corpus-Based Transcription as an Approach to the Compositional Control of Timbre," in *Proc. ICMC*, Montreal, 2009.

[13] N. Schnell, D. Schwarz, R. Cahen, and V. Zappi, "IAE & IAEOU," in *Topophonie research project: Audiographic cluster navigation (2009-2012)*. ENSCI – Les Ateliers / Paris Design Lab, 2012, pp. 50–51.

[14] A. Agostini and D. Ghisi, "A Max Library for Musical Notation and Computer-Aided Composition," *Comput. Music J.*, vol. 39, no. 2, pp. 11–27, 2015.

[15] J. Garcia, T. Carpentier, and J. Bresson, "Interactive-Compositional Authoring of Sound Spatialization," *J. New Music Res.*, vol. 46, pp. 1–13, 2016.

[16] A. Vinjar and J. Bresson, "OM-AI: A Toolkit to Support AI-Based Computer-Assisted Composition Workflows in OpenMusic," in *Proc. SMC*, Malaga, 2019.

[17] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," IRCAM, Tech. Rep., 2004.

[18] D. Schwarz, "Corpus-based Concatenative Synthesis," *IEEE Sig. Proc. Mag.*, vol. 24, no. 2, pp. 92–104, 2007.

[19] H. Pomberger and F. Zotter, "Warping of 3D Ambisonic Recordings," in *Proc. Ambisonics Symp.*, Lexington, 2011.

[20] F. Zotter, M. Zaunschirm, M. Frank, and M. Kronlachner, "A Beamformer to Play with Wall Reflections: The Icosahedral Loudspeaker," *Comput. Music J.*, vol. 41, no. 3, pp. 50 – 68, 2017.