Mollspeak — Unreal Spatial Synthesis for a Narration

Olivier Pasquet
Goldsmiths, University of London
o.pasquet@gold.ac.uk

ABSTRACT

Mollspeak is a permanent immersive sound installation that explores the lives of eighteenth-century maidservants through an intimate portrayal of home life. The piece examines the cultural significance of domestic spaces and their impact on social dynamics. This article outlines the creative processes behind the piece and discusses the technical solutions used to realize it, including photogrammetry, physical model synthesis engines, and software for simulating surreal room acoustics. The sound piece showcases an innovative approach to immersive storytelling using abstract voice, and is part of a larger research endeavor in music-theatrical directions.

1. INTRODUCTION

Mollspeak is a permanent sound installation at the Museum of the Home in London, premiered on May 18th, 2021, and also aired as a 25-minute radio binaural version on October 12th 2021. The installation combines synthetic sounds, music generation, and acoustic architecture designed for the museum's building, featuring a text by Maria Fusco and voiced by actress Maxine Peake.

The project is part of a long-term artistic research exploring the connections between spatial representations of time and narration, stemming from the authors' backgrounds in theater, architecture, and music composition. Spatialization plays a crucial role in the piece, with all sounds and the composition's timeline generated and positioned in space and time using 3d scans of museum objects and a dedicated synthesis engine called *Dürer*.

The paper discusses how various tools and concepts come together to create a cohesive aesthetic for a series of pieces, showcasing an artistic research path that merges expertise, imagination, failures, and surprises.

2. SCENARIO

Maria Fusco's text reveals the pains, desires, and routines of eighteenth-century maidservants, drawing from extensive archival research and objects in the Museum's collections (fig. 1). *Mollspeak*, is a phrase employers coined to mock their servants' dialectical speech. Maria Fusco's original script mingles research with historic poetic form to give voice to a working class's story – reflecting on duties, desires and place in the world – to create a plainly spoken, yet lyrical portrait.

Exploring homes and home life requires intimacy and familiarity, which affects anthropological and cultural aspects. Subconscious memories shape our definition of home, with physical spaces and objects playing a crucial role. The sonic component of this piece explores the subconscious connections between objects and their spaces, leaving uncertainty about their origins. This ambiguity creates a fertile ground for composition.

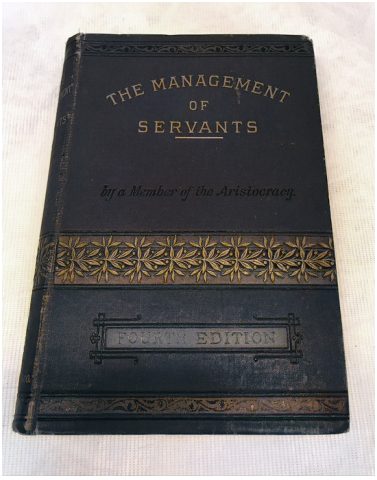

Figure 1. *The Management of Servants by a member of the aristocracy*, eighteenth-century.

3. USE OF THE MUSEUM'S COLLECTION

3.1 Photogrammetry of objects of the museum

The piece is based on classified objects from a museum, each representing a social aspect of British society. The resonant characteristics of each object were recorded and 3d scanned using photogrammetry to avoid physical contact with the fragile items (fig. 2). This allowed for the digital creation of physical models that could be transformed into unconventional forms later on.

3.2 Experiments using physical model synthesis engines

The initial concept involved using 3D scan data to generate resonances from quadrangle meshes [1, 2]. Tools like Modalys and Faust were used to apply finite element methods and create resonant models. However, the results only became musically interesting after applying transient design techniques, including compression, valve dis-

tortion, emphasizing mid-high frequency transients, and controlling low-frequency decay using specific plugins and tools [3, 4].

We prefers fast rhythms with sharp attacks to clearly understand musical structure. As a result, incorporating long-term resonances wasn't feasible. We realized that a less continuous approach was needed to maintain a pulsating style and thus prevent listener apathy [5].

Figure 2. Photogrammetry scans of horn beakers and a paper doll. Many of the museum objects were scanned. Some of them are not supposed to resonate, to make any sound...

3.3 Alternative symbolic strategies

We explored alternative interactions with 3D scans of museum objects, such as scratching or shaking, using tools like the Sound Design Toolkit (SDT) to create non-linear sounds [6] [1]. However, we opted for a more poetic approach, introducing temporal elements to sonic materials and imbuing objects with symbolic meaning to create metaphors and ostinati. This approach allowed for contrasting elements and meaningful comparisons.

Another strategy involved using actress Maxine Peake's voice in a musical manner, juxtaposing it with the text. The outcome was a blend of meaningful text and dadaistic composed phonemes derived from voice processing. Techniques using the yin fundamental frequency estimator included separating voiced and unvoiced elements of her voice, rhythmically playing them as percussive elements, and combining them to create textual meaning [7]. The text was often deconstructed and not read linearly, but Peake's declarative style added an emotional foundation [8].

The result is a seamless blend of narrative sound objects, musical metaphoric elements, fragmented words, and an intensely delivered text. The delivery is as forceful and insistent as the commands given to servants. Sound examples can be heard here:
`https://www.opasquet.fr/mollspeak/`.

3.4 Other alternative strategy with spatial synthesis

The tool *Dürer*, developed for this piece, is an audio synthesis and virtual acoustics tool for Rhino's Grasshopper, based on image source model room simulation (fig. 3). It pays homage to Albrecht Dürer and Vera Molnár [9]. It consists on representing time by space and displacement in the most straight forward way in order to create descriptive scores with a specific granular esthetics in the near future.

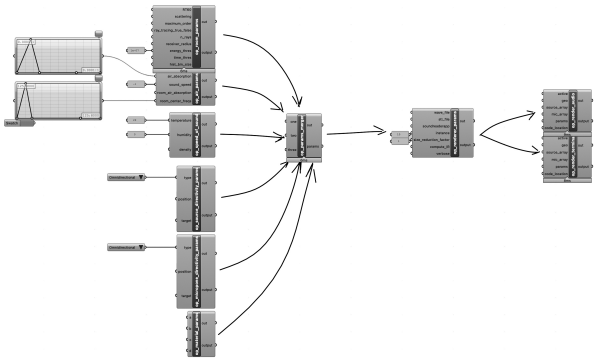

Figure 3. Components for *Dürer*'s renderer in Grasshooper's Rhino plugin. It consists on three parts: physical parameters (materials, directivities, scattering etc), renderer parameters (IR size, instances etc) and export (IR audio format, convoluted file etc).

The first development process began with a hack and the integration of the *Pyroomacoustics* Python library from EPFL, Lausanne, into Rhino [10, 11] [2]. The engine renders multichannel impulse responses either from a beam of microphones (beamforming) or a virtual Eigenmike, allowing manipulation of directivity, delays, and virtual ambisonics (fig. 4). For *Mollspeak*, Rhino imports all the 3D scans of objects within the museum, ensuring that meshes are closed. These meshs may undergo transformation and simplification to enhance rendering speed. Mesh reduction is a nuanced process that varies depending on specific requirements. We have developed a technique for mesh simplification, utilizing a genetic algorithm solver called Galapagos, which preserves reflections acceptably [12].

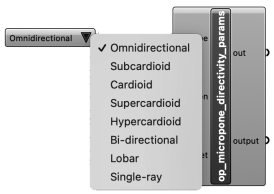

Figure 4. Available directivities for *Dürer*'s virtual microphones. "Single-ray" is not feasible in reality, which is precisely what we are interested in.

Dürer is primarily designed for unrealistic room acoustics. It can generate multichannel impulse responses and apply convolution to a set of sound files. It is also possible to create sound synthesis related to space using physically

[1] https://github.com/SkAT-VG/SDT

[2] https://github.com/SkAT-VG/SDT

unrealistic characteristics. A torus with specific face angles can produce intriguing click rhythms and oscillations, and the visual outcome closely correlates with the sound (fig. 7).
For example, one can play with:
- unattainable parameters such as room temperature, speed of sound, air absortion...
- amplifying materials positioned with precision to create oscillations (absortion >1.)
- RIR denoising by only synthesising a limited amount rays or bounces.
- an impractical number of virtual speakers or microphones.

The most interesting aspect, in conjunction with the piece's sharp aesthetics, is to capture the sequence and power of rays bouncing off mesh faces. Each object from the museum thus receives a unique characteristic derived from the topology of its surfaces. We voxelize the space, quantify distances and ray angles to get perceivable rthymical coherency (fig. 5). We also empirically adapt the number of rays, their initial directivity and intensity as if we were reconstructing acoustic directional characteristics of sources. In our specfic case, it has been found that it only makes sense to use a train of dirac impulses as sources to avoid any question about phases. Shortening the length of ray paths and setting mesh faces to be fully reflective creates a clear texture.

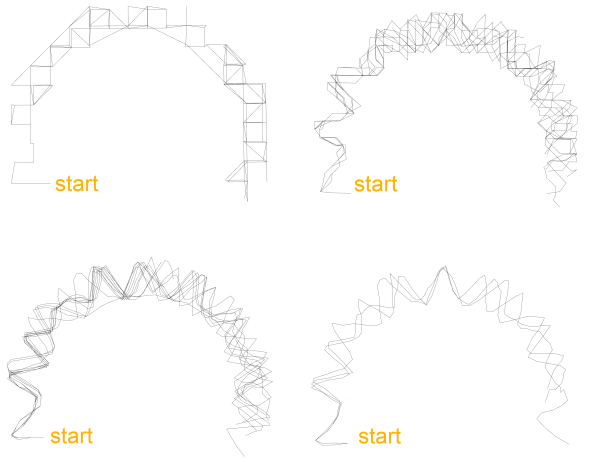

Figure 5. Four semi-torus used as meshes. The two top ones use a voxelized space. The two at the bottom have quantized ray angles and distances. They all use a small number of rays for the demonstration.

This concept remains easily comprehensible, both aurally and visually, when using a limited number of rays. However, employing a large number of rays enables a sense of continuity in the resulting RIR, effectively yielding rhythmic synthesis. The mesh's form and its topological coherence create temporal consistency and discontinuities from sample to sample, akin to resonance. For example, squeezing the object has a pronounced impact on the outcome (fig. 6).

Computational power, combined with geometric tools in Rhino and the described computational granular propagation, enables rapid construction and deconstruction of objects, transitioning seamlessly from continuous to discontinuous forms (fig. 7). We have explored various optimisation techniques for forms finding in order to generate specific rythms and clusters within the RIR. This unfortunately have not been further for *Mollspeak*.

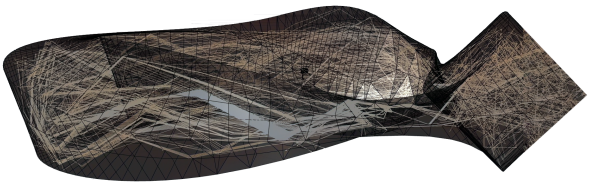

Figure 6. Distorted torus. By constraining and discretizing ray angles, parallelism and convergence emerge, yielding a construct showcasing intricate ray interactions and interferences.

Figure 7. Two "expensive donuts": the one on top is flat and will give complex early reflections, faces of the bottom one are set for a specific rhythm.

3.5 Radio version of the piece and download *Dürer*

It is possible to listen to the linear 25-minute radio stereo and binaural version of the piece online. *Dürer* requires the architecture design software Rhino 3d [3] .
`https://www.opasquet.fr/mollspeak/`
`https://github.com/opasquetdotfr/Duerer`

4. FUTURE DEVELOPMENTS

Several developments have been done after *Mollspeak* and more is to come:
Our recent works, *The Nudge* and *The Impossible Trilemma*, both premiered at CLab in Taipei, feature "uncontrolled" transformers of Bark text-to-speech model [4] . These pieces blend abstract narration with uncanny vocal effects, pushing the boundaries of music and meaning. They draw a

[3] https://www.rhino3d.com
[4] https://github.com/suno-ai/bark

direct inspiration from previous vocal works composed by Georges Aperghis [13].

Meanwhile, our ongoing project *Dürer*, initially used as a visual score generator for the collection of large-scale paper prints *Toi-1842b*, has evolved using an experimental multi-thread version of *Dürer* with Pytorch, a Python library initially used for machine learning [14] [5] . We also have the intension to use that spatial technique to explore latent spaces.

5. CONCLUSION

Mollspeak was a major step in stylistic research on our pieces using text and music. The deployment of diverse techniques, including those tailored specifically for that piece, allowed for experimentation and ultimately essential advances for ongoing long-term progress.

Dürer was built in the realm of generative forms and spatial-temporal representation. It serves as a personal artistic notation and generative tool for composition utilizing spatial computation [15]. We have described, with *Mollspeak*, a simple framework and individual workflow for a series of specific concepts that can ultimately be materialized into sound pieces, prints, or fabricated in 3D.

Acknowledgments

We would like to thank Maria Fusco, Sinead McCarthy, Christina Karpodini, Jan Schacher, Toro-Pérez Germán, Atau Tanaka, the Museum of the Home in London, and the entire team of ICST at ZHdK without whom *Mollspeak* would have never been realized.

6. REFERENCES

[1] R. Michon, "Faust-STK: A set of linear and nonlinear physical models for the Faust programming language," 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:4844055

[2] A. Stegaru, "Modalys Reference Manual Finite Element Objects," 2003.

[3] P. A. Tremblay, G. Roma, and O. Green, "Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit," *Computer Music Journal, vol. 45, no. 2, pp. 9–23*, 2021.

[4] G. Roma, O. Green, and P. A. Tremblay, "Stationary/transient audio separation using convolutional autoencoders," 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:186844125

[5] B. Riley, *L'Esprit de l'Oeil*, 2008.

[6] S. Baldan, S. D. Monache, and D. Rocchesso, "The Sound Design Toolkit," *SoftwareX*, vol. 6, pp. 255–260, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:55869488

[7] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music." *The Journal of the Acoustical Society of America*, vol. 111 4, pp. 1917–30, 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:1607434

[8] H.-T. Lehmann, "Postdramatic Theatre," 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:235209773

[9] V. Molnar, "Léonard de Vinci s'il eut eu un ordinateur..." 1981.

[10] D. Rutten, "Evolutionary Principles Applied to Problem Solving - Grasshopper," 2010.

[11] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 351–355, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:11855208

[12] D. Rutten, "Galapagos: On the Logic and Limitations of Generic Solvers," *Architectural Design*, vol. 83, pp. 132–135, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:62481370

[13] M. Woo, "L'interprétation musicale des phonèmes, des gestes et des images dans Machinations de Georges Aperghis," 2011.

[14] B. Rozemberczki, P. Scherer, Y. He, G. Panagopoulos, M. S. Astefanoaei, O. Kiss, F. Béres, N. Collignon, and R. Sarkar, "PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models," *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233289399

[15] A. Spicher, O. Michel, and J.-L. Giavitto, "Spatial Computing as Intensional Data Parallelism," *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop*, pp. 196–205, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:9328536

[5] https://pytorch.org