Exploring Sound and Spatialisation Design on Speaker Arrays using Physical Modelling

Georgios Marentakis
Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz
marentakis@iem.at

David Pirrò
Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz
pirro@iem.at

ABSTRACT

In the course of the realization of the sound installation *Interstices*, questions pertaining to the auditory perception of location and extent and the spatial composition of the micro and macro structure of sound were explored in a *poietic* way. Physical modelling was re-interpreted as a framework to design the spatial and timbral appearance of sounds upon a set of distributed speaker array clusters. This explorative process lead to observations that helped formulating novel research questions within the context of psychoacoustics and auditory display.

1. INTRODUCTION

We present an investigation into designing spatialised sound output on speaker arrays. Of particular interest, are the perceptual impressions generated when sound is not rendered as a point source within a reproduction system, but is rather spatially distributed. We are also preoccupied with the perceptual effect of the simultaneous presentation and dynamic reorganization of such perceptual objects from speaker array surfaces, that are physically realized in space as opposed to surrounding the listener area. From this point of view, we attempt to question the way speaker arrays have been seen in the past.

Our work is inspired by observing how loudspeakers have been used in sound installation art. There, as opposed to traditional spatialisation practice, speakers are used freely, often made explicitly visible, in this way participating in the formation of the visual impression, or even more contributing to a physical sculpture. Such multichannel works examine not only the spatial nature of sound, but also rethink the way the loudspeakers are placed and integrated in space.

Quite commonly within this practice, sound material is distributed upon the speaker surface. Spatial auditory impressions of considerable interest emerge, in particular with respect to their extent and geometry. The formation of these impressions has largely remained outside scientific research, and the knowledge associated with their creation, has not been expressed outside the compositional domain, although it would be invaluable in the field of sonic interaction and in general audio display design. Translating this knowledge into the domain of scientific research is a non-trivial task.

We decided to engage in this research in a *poietic* way, embarking in a creative process aimed at the realization of *Interstices*, a multichannel sound installation. This approach enabled us to collect and maintain valuable practical and sensual experiences which we formalised in a set observations. These reflections inform both future compositional endeavours and the formulation of scientific questions that will be addressed in future studies.

Drawing from our previous scientific and artistic experiences and having at our disposal a flexible software framework to work with, we use physical modelling to explore sound spatialisation on speaker arrays. As we will explain, this framework is used as a tool to design and connect the microscopic and macroscopic structure of sound. We mention here, that we employ physical modelling not in order to imitate mechanical instruments or simulate the acoustics of sound propagation. Rather, we adopt these models with a "non-standard" approach within the context of sound synthesis and spatialisation, in order to realize compositional ideas of sound organization [1] on different scales within time and space.

2. BACKGROUND

2.1 Spatialisation & Speaker Arrays

Primary models for spatialisation have been trans-aural stereo, panning techniques such as vector-based amplitude panning [2] and distance-based panning [3], holographic techniques such as Ambisonics [4] and Wavefield synthesis [5], and simulations of acoustic environments such as ViMiC [6]. The techniques differ substantially on their approach towards sound spatialisation, however, their aim is the same: to provide the impression of sound originating from a location around the listener, given a fixed, discrete loudspeaker setup. This very constrain gives rise to a number of limitations. As the distance between the loudspeakers becomes bigger, the ability of spatialisation algorithms to provide realistic impressions of sounds originating along a continuum between the loudspeakers decreases. For holographic techniques, increasing loudspeaker spatial separation also deteriorates sound quality, as higher frequencies are not sufficiently well reconstructed, due to the well known phenomenon of spatial aliasing. Furthermore, the success of spatialisation algorithms is compromised by the requirement that a certain symmetry needs to be observed between the location of the listeners and this of the loudspeaker array. Listeners, outside the 'sweet spot', are facing deterioration of localization and sound quality.

Traditionally, array geometry has been tied to the spatialisation technique used. This led to arrays based on triangular or horizontal equidistant speaker arrangements for stereophony, spherical or semi-spherical designs for Ambisonics and most commonly linear speaker arrays for Wavefield Synthesis. In the field of acoustic radiation modelling, alternative designs have been realized, again tied to the specificities of the sound radiation model used for the reproduction of the sound field, most commonly using spherical harmonics [7].

In an application domain such as sonic interaction design or human computer interaction, this approach is limiting. This is because the constraints imposed by spatialisation algorithms cannot be easily fulfilled by the potential products they will be embedded into. It is therefore necessary, to research spatialisation techniques and speaker array designs, that work robustly within this new application domain, and reconsider how speaker arrays can be integrated into interactive objects.

2.1.1 The IEM modular speaker array system

Within this context, and with the goal to create a a platform for experimentation the IEM Modular Speaker Array system (Figure 1) was created. The system provides the opportunity to prototype and test quickly speaker array configurations. It uses affordable Class-D amplifiers and small, easy-to-mount foil speakers. The speakers can be attached onto aluminum rods at variable heights that can be freely fitted onto a wooden platform. In this way, it is possible to quickly create variable speaker setups to test different design, psychoacoustic or spatialisation hypotheses. While other attempts to create speaker arrays with small and medium sized speakers have been made [8], these do not offer the flexibility of this prototype.

Further, any speaker array system intended to be used for sound output and sound design needs to be controlled through a generalized software framework. Such a framework needs to encapsulate the need for presentation of objects and to provide feedback with respect to display state and user actions. Most current such frameworks for spatialisation control are heavily influenced from the game industry. Sounds are considered local and unitary objects, placed somewhere within the spatial reference system. While considerations such as source directivity, room modelling and so on are usually done, the sound microstructure rarely enters the spatialisation framework. In most of the cases, sound is represented as an anechoic recording or as the output of a synthesis algorithm. When one, however, considers that no source is point-like and that sound is created through the vibration of innumerable points on the surface of objects, this framework of interpreting sound within a spatial audio scene is limited. This is not only because of sound directivity aspects, but also because essential aspects of spatial audio perception are underrepresented within this model. For example, it is well known that the spatial distribution of the frequencies within a sound event gives rise to variable perceptions of auditory source width [9]. Control over the spatialisation of the constituent elements of sound is therefore necessary. Although a variety of methods could be used in this direction, in this work, we focus on using physical modelling.

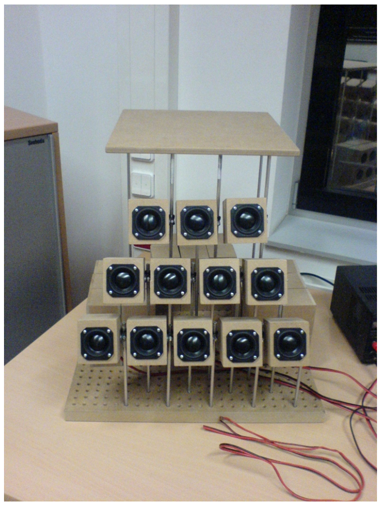

Figure 1. IEM modular speaker system used to create a side of a pyramid.

2.2 Physical Modelling

Physical modelling allows for the simulation of the mechanics and therefore the physical characteristics of acoustic instruments, producing rich and realistic sounds, while at the same time providing intuitive high level control of sound synthesis [10]. It has been widely used in conjunction with various motion tracking technologies to interactively produce sound [11, 12]. Many computational physical modelling frameworks have been developed employing different approaches such as waveguide models, modal synthesis (e.g. Modalys [13]) or particle-based models like CORDIS-ANIMA [14] or the TAO [15] frameworks.

Here, we focus on using particle-based physical models to formulate behaviour. We use single "objects" or particles that are connected together in a network, linked by the forces acting between them. This linking results in an interaction between the objects, as the behaviour of each of them depends on its relations to the rest. Composing and defining these forces eventually means to compose the behaviour and the dynamics at a local (object) and a global (network) level. A software framework designed to explore the design space and used to create our installation is described next.

2.2.1 Rattle

Rattle [1] is an efficient implementation of a mass-based physical modelling server, written in C. It allows for the

[1] Rattle is being developed and maintained by David Pirrò

rapid prototyping of particle-based physical models. Models can be defined by adding, placing or removing particles, defining their mass and linking them to each other using forces and attrition, and containing their movement using boundaries in a virtual space. *Rattle* samples the locations, displacements, velocities, accelerations and energies of the elements of a model at variable sample rates. This allows to work with different time scales, seamlessly joining or exchanging them within the same framework. Furthermore, *Rattle* can sample at audio rate, operating as a sound synthesis engine. At sub-audio rates *Rattle* can be used to control macroscopic behaviour of external objects, for instance for spatialisation control.

The forces, constraints, spatialisation algorithms as well as new functions linking the behaviour of the objects can be scripted using a text-based programming language using the C syntax. All parameters can be either specified in advance or adjusted in real-time, while the model is running. Different models can be run in parallel. Adjustments can be done through either an OSC [2] or a MIDI interface [3]. Finally, *Rattle* uses LLVM technology [4] to JIT-compile new functions and load them in real time into the running simulation, without the need to re-compile, enabling quick and flexible prototyping.

2.3 Approach

As mentioned earlier, we want to explore the perceptual impressions generated by the spatial rendering of sound microstructure. The notion of microstructure, is here used to refer to the elements used to construct a sound. These are of course variable, and depend on the representation used, could be bins in a spectrogram, grains in a granular synthesis context, or vibrations of independent particles within a physical model. Here we attempt to make some predictions with respect to the fate of these sound bits, when they are spatialised within reasonable proximity to each other. Had the physical output model been rendered from a single position in space, the microstructure of the percept would not be so easy to perceive due to masking. Spatialising the sound of the individual particles, allows listeners to focus on microstructure properties, in a similar way to what happens in the cocktail party effect [16]. As the role of spatial organization is secondary in auditory object formation [17], bits that are perceptually fused, based on frequency, modulation and timing constraints, will yield a perceptual object [18], whose spatial characteristics will be determined by the locations of the contributing sounds. Spatially proximate bits, that will not fuse perceptually, will be perceived within the object's area and depending on their timbre, they may or may not relate to the containing object. Within this framework several questions are worth asking. The first, is naturally addressing the geometry and spatial extent of the spatial impressions that can be created. The second, addresses dynamic aspects of such representations, in particular their robustness to translation and reshaping. Finally, it is worth considering how interference from other auditory objects will affect the generated spatial impression.

[2] Liblo: http://liblo.sourceforge.net/

[3] Portmidi http://sourceforge.net/apps/trac/portmedia/wiki/portmidi

[4] http://llvm.org/

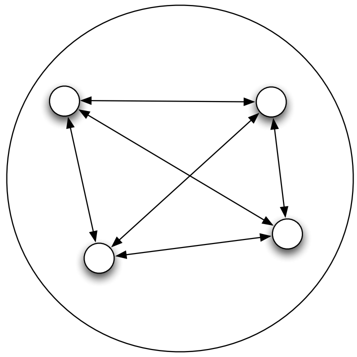

Figure 2. Schematic representation of elements and relationships in the the "High Frequency" model: the four masses construction the model interact with each other and are constrained into a sphere.

The creation of an installation, apart from its artistic value, offers the possibility to work in a systematic way towards the exploration of a design space. Importantly, it emphasizes the researcher's exposure to the materials, which is essential in the formation of research questions, that can be later explored in a systematic way.

To perform our exploration, we worked at three different levels of control. We refer to these as micro-scopic, meso-scopic and macro-scopic respectively. By microscopic, we refer to the sound generation process, in our case the control of the displacement of each particle within a physical model. By meso-scopic, we refer to the control of the location of each model particle, and by macro-scopic we refer to the control of the translation and rotation of the model as a whole. Each of the levels works at a different rate, ranging from sampling rate in terms of sound generation to much slower transformations at the macro-scopic level.

From an artistic perspective, the main interest lies in translating the formulation of the model's behaviour into the composition of sound in space. As we explained particle-based physical models offer the possibility to compose the relations between the single elements of an organic system. Eventually, when running the simulation, these elements will show a coherent behaviour according to the model dynamics. Musically it is interesting to understand how this behaviour affects sound and its spatial appearance. Depending on the relations governing the internal mechanics, these systems exhibit dynamics that lie within a continuum ranging from single organic "entities", to extended subspaces, or to a collection of disjoint atoms. Exploring this range of possibilities and making it subject to composition, is a central aim of our research. When this approach is applied in parallel to different time scales [19] of sound generation and spatialisation, distinct systems are used to define the choreography of sound in space.

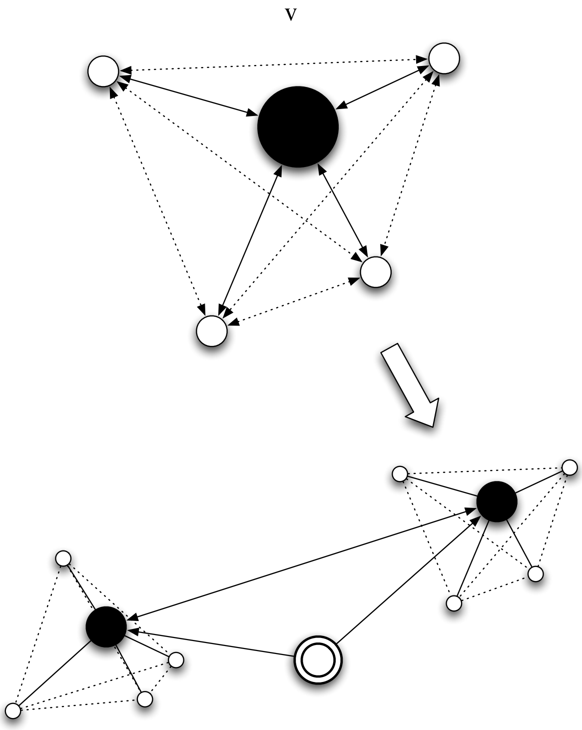

Figure 3. Top: “meso-scopic” model. Five masse interact with each other through gravitational forces. The black object is “bigger” and “heavier” then the other four. Bottom: “macro-scopic” model. The central fixed object acts on the black masses of the previous model attracting them. At the same time these objects repulse each other

3. *INTERSTICES*

Within the process of creating the installation we explored the three levels of control using the physical modelling approach described earlier. The installation took place in ESC Labor in Graz, between 12th and 21st of January 2012.

To realize the “micro-scopic” layer (see Figure 2), we used a network of four mutually interacting masses, whose movement was confined within a sphere by an elastic boundary. To produce sound we directly audifed the velocities of these masses. The weights of these particles’ masses and the magnitude of the forces connecting them to each other were chosen within a range such that their velocities changed with frequencies within the audible range. As a consequence the morphology of the sound output can be defined as a function of the weight, of the forces connecting the masses to each other and of the attrition acting on them. For instance, spring forces led to simple, relatively static harmonic spectra. Gravitational like forces instead produced more complex and inharmonic sounds with unstable time behaviour. When the forces were attractive, the sound exhibited clear pitches. Repulsive forces on the contrary, caused more impulsive, noisy sounds or bursts.

During the preparation phase, the different spatial perceptions were examined independently as a function of the different sound morphologies. As a result, we identified a parameter space that yielded a satisfactory range of distinguishable timbres, ranging from harmonic to quasi-harmonic to transient. While the installation was running these parameters were gradually updated, causing a variation to the model’s internal state and in this way exploring the space of possibilities offered by the model. Two of these models that oscillated between the different states were run in parallel in the installation, yielding substantial timbral variation.

The “meso-scopic” layer (figure 3 top), was implemented using five particles connected through gravitational-like forces. The model was designed so that one of the objects acted as main attractor, keeping the other four orbiting around it. To achieve this, the attractor particle’s mass was substantially larger than the masses of the other four particles and the forces connecting the lighter objects to the attractor were substantially stronger than the ones connecting them to each other. Masses and forces were chosen such that the movements were significantly slower than in the previous model: the time needed for one of the smaller objects for a complete revolution was ca. 1 - 3 seconds. This model was also updated dynamically, changing the magnitude of the attractive forces and yielding variable spatial configurations with objects moving in a loose or tight way relative to each other, ending up concentrated or far away from each other. The location of each object in the meso-scopic model was used to define where each mass in the micro-scopic model would be spatialized. Therefore the qualities of their movement and their relative positions determine how localized or extended the sounds were projected by the loudspeaker array.

A similar approach was used to define movement at the macro-scopic level. Again a bigger mass, a fixed “sun”, was placed at the origin of the coordinate system. The two attractors of the meso-scopic level with gravitational forces revolved around this object. These, however, were mutually repulsed by similar “gravitational” forces so that the “meso-scopic” systems slowly revolved around this central sun, but remained most of the time well separated from each other and mixed only occasionally (Figure 3 bottom). Both “macro-scopic” and “meso-scopic” systems are constrained into a rectangular box with reflecting walls

To spatialise the movements, the 48 speaker array was used to create four speaker clusters each containing twelve speakers. Each speaker was represented in the macro and meso scopic system using a single point. A simplified Distance-Based Amplitude Panning algorithm (DBAP [3]) was used so that the sound of each of the micro-scopic model masses was rendered to the loudspeakers with an intensity that was determined based on its distance to each of the loudspeaker. To avoid an excessive “blurring” of the single sources, we parametrized this algorithm in such a way so that each sound could appear on maximum three loudspeaker at the same time.

4. DISCUSSION & CONCLUSIONS

We attempt to organize here our observations collected while realizing the installation and form questions that could be followed up in controlled experiments. Of particular importance to our investigation was to observe informally, *the*

Figure 4. Final distribution of the loudspeaker clusters in the ESC Labor space. Foto: Martin Rumori

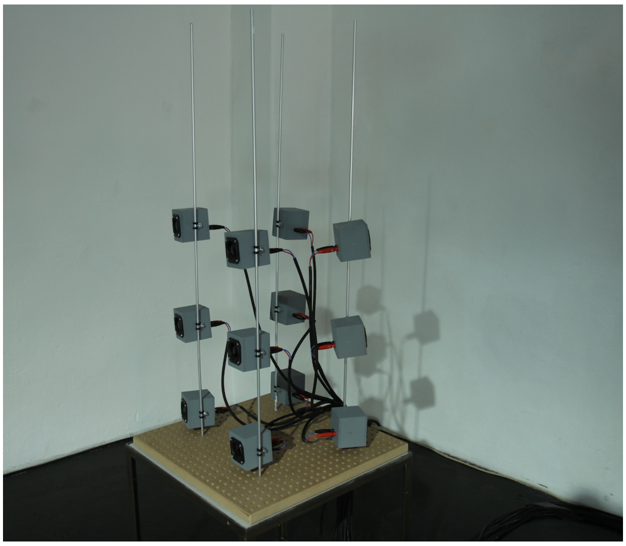

Figure 5. One of the loudspeaker clusters used in the sound installation *interstices*. Foto: Martin Rumori

extent to which distributing micro-structure on the speaker array, can provide the perception of a 'body of sound'. The term 'body of sound' is used here to describe the auditory perception of a bounded spatial area occupied by a coherent auditory object. We found convenient, in the course of preparing the installation, to refer to this perception based on the relation between the activated area on the speaker array, and the perceived extent of the auditory event. We used questions such as: *is the auditory extent bigger when the speaker array surface is increasing?* We experimented with combinations of harmonic, quasi-harmonic and transient vibrations from each model mass and found that our perception depended heavily on the sound material. Harmonic material provided the impression of a sound filling the room rather than emitted from the activated speaker array surface. Quasi-harmonic and transient material yielded a better match to the radiating area of the speaker array and the spatial boundaries of the auditory event were relatively easy to notice, and in general corresponded to the array size, more sharply defined in the horizontal, rather than the vertical dimension. Overall, the perception of auditory extent was more stable for sparse wide arrays, where speakers were separated in space. Rendering onto a dense speaker lattice, resulted in reduction of the spatial extent, in particular for transient and quasi-transient stimuli. The role of room reflections depended on the sound material used and the location of the arrays. For harmonic material, early reflections merged with perceptual object and may have contributed to the impression of sound filling the room [20]. For quasi-harmonic and transient material, depending on the speaker location, quite often reflections were perceived as segregated from the object and did not sufficiently merge.

We then turned our attention to the perception of the spatial distribution of energy within the spatial boundaries defining the auditory object's perception. We found that this was quite possible. The internal spatial clarity was nevertheless limited, when the array surface was compressed. We tried to identify whether the spatial energy was uniformly distributed within the perceptual object, or perceptual centres of gravity were observed. Based on our informal observations, both situations appeared, depending both on the sound material used and on the geometry of the speaker array. The exact workings behind this perception are still somewhat unclear to us, but we could formulate the question that needs to be answered here: *how do we get from the perception of spatial distribution of sound energy across the speaker area to the creation of perceptual centres of gravity within the area?* The answer is not as straightforward as using moving from decorrelated to correlated sound material, as [21] observed multiple centres of gravity also when distributing decorrelated noise signals upon a horizontal speaker array.

A third question that emerged is whether *'bodies of sound' can move or be dynamically reshaped?*. The aforementioned phenomena were stable and reproducible when the vibration of the model particles was statically distributed on the speaker array. The percept was however stable only for very slow translational movements. Upon quick movements it would collapse. This could be related to the time required by the auditory system to integrate the spatialisation information from the contributing elements, which is known to be relatively large [22]. In addition, the stability of the percept was also heavily influenced by the presence of a second perceptual object of different timbral quality. Interference was observed in the formation of the auditory percept by other objects.

Spatialising sounds in the space between the speakers was not as effective as we have hoped. Upon spatialisation, the perception of a single body of sound emerging by the contributions of the micro-elements was hard to create and it was much more common to observe multiple acoustical centres. Spatialisation at least in common practice, results in the reproduction of correlated signals. Inadvertently therefore, it forces the creation of perceptual centres that limits the diffuseness of the percept.

The modelling approach followed here is methodologically interesting, as the model can be analysed and reproduced thus repeating the experiment and the parameter space can be explored without changing the model. Usually, compositional approaches tend to be "true" only for one specific situation, which one cannot easily understand because their validity is limited to the specific situation. The model has parts and subparts that can be combined or change independently. It allows therefore to "isolate" specific features to be revisited and explored in a more focused way but also translated in different contexts. Furthermore, the use of this semi-formalised exploration procedure helped to define the following research questions that need to be addressed in more controlled studies: 1. how can we create the perception of 'bodies of sound' within a bounded spatial area occupied by a coherent auditory object? 2. how do we get from the perception of spatial distribution of sound energy across the speaker area to the creation of perceptual centres of gravity within the area? 3. what determines the outcome of overlapping 'bodies of sound'? 4. Can spatialisation algorithms in their present form be used to create such impressions?

Acknowledgments

We would like to thank the IEM, Institute of Electronic Music and Acoustic in Graz for supporting the realization of the Modular speaker array system and of the sound installation *Interstices*. Further we would like to thank Sebastian Blamberger for designing and building the amplifiers and loudspeakers and Winfried Ritsch for technical support. Finally we would like thank the "ESC im Labor" gallery for the invitation to present our work.

5. REFERENCES

[1] L. Doebereiner, "Models of constructed sound: Non-standard synthesis as an aesthetic perspective," in *Computer Music Journal*, vol. 35, no. 3. Massachusetts Institute of Technology, Fall 2011, pp. 28–39.

[2] V. Pulkki, "Spatial sound generation and perception using amplitude panning techniques," 2001.

[3] T. Lossius, P. Baltazar, and T. de la Hogue, "Dbap - distance-based amplitude panning," *Proceedings of the 2009 International Computer Music Conference*, 2009.

[4] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[5] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[6] J. Braasch, N. Peters, and D. L. Valente, "A loudspeaker-based projection technique for spatial music applications using virtual microphone control," *Comput. Music J.*, vol. 32, no. 3, pp. 55–71, Sep. 2008.

[7] F. Zotter, "Analysis and synthesis of sound radiation with spherical arrays," Ph.D. dissertation, University of Music and Performance Arts Graz, 2009.

[8] D. Beer, S. Mauer, S. Brix, and J. Peissig, "Flat panel loudspeakers consisting of an array of miniature transducers," in *126th Convention of the AES*. AES, 2009.

[9] F. Zotter, M. Frank, G. Marentakis, and A. Sontacchi, "Phantom source widening with deterministic frequency dependent delays," in *Proceedings of DAFX 2011*, 2011.

[10] C. Roads, A. Piccialli, G. D. Poli, and S. T. Pope, Eds., *Musical Signal Processing*. Lisse, The Netherlands, The Netherlands: Swets & Zeitlinger, 1997.

[11] D. M. Howard and S. Rimell, "Real-time gesture-controlled physical modelling music synthesis with tactile feedback," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1001 – 1006, 2004.

[12] M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," *Proceedings of the IEEE 2004*, vol. 92, no. 4, pp. 632 – 644, April 2004.

[13] G. Eckel, F. Iovino, and R. Caussé, "Sound synthesis by physical modelling with modalys," in *Proceedings of the International Symposium on Musical Acoustics*, July 1995, pp. 479 – 482.

[14] C. Cadoz, A. Luciani, and J. L. Florens, "Cordis-aniama: Modeling and simulation system for sound and image synthesis - the general formalism," *Computer Music Journal*, vol. 17, no. 1, pp. 19 – 29, Spring 1993.

[15] M. PEARSON, "Tao: a physical modelling system and related issues," *Organised Sound*, vol. 1, no. 1, pp. 43–50, 1996.

[16] A. Ihlefeld and Shinn-Cunningham, "Spatial release from energetic and informaional masking in a selective speech identification task," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4369–4379, 2008.

[17] C. Darwin and R. Hukin, "Perceptual segragation of a harmonic from a vowel by inter-aural time difference and frequency proximity," *The Journal of the Acoustical Society of America*, vol. 102, pp. 2316–2324, 1997.

[18] A. Bregman, *Auditory Scene Analysis: The perceptual organization of sound*. MIT Press, 1990.

[19] C. Cadoz, "The physical model as metaphor for musical creation'pico.. tera', a piece entirely generated by physical model." in *Proceedings of the International Computer Music Conference*, 2002, pp. 305 – 312.

[20] W. M. Hartmann, "Localization of sound in rooms," *The Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1380–1391, 1983.

[21] O. Santala and V. Pulkki, "Directional perception of distributed sound sources," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1522–1530, 2011.

[22] D. W. Chandler and D. W. Grantham, "Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1624–1636, 1992.